CONSTRUCTAL THEORY OF DESIGN IN ENGINEERING AND NATURE

by

Adrian BEJAN

Original scientific paper
UDC: 65.011:612.087.1
BIBLID: 0354-9836, *10* (2006), 1, 9-18

This is a brief introduction to an engineering theory on the origin and generation of geometric form in all flow systems: the animate, the inanimate and the engineered. The theory is named constructal, and is based on the thought that it is natural for currents to construct for themselves in time paths of greater flow access. It is shown that this process of flow path optimization can be reasoned on the basis of principle: the maximization of global performance subject to finite-size constraints. One example is the generation of tree-shaped flow patterns, as paths of least resistance between one point (source, sink) and an infinity of points (area, volume), as in the circulatory, respiratory and nervous systems. Another is the generation of regular spacings in heat generating volumes, such as swarms of honeybees. The optimized tree-flow geometries account for allometric laws, e. g., the relationship between the total tube contact area and the body size, the proportionality between metabolic rate and body size raised to the power 3/4, the proportionality between breathing and heartbeating times and body size raised to the power 1/4, and the proportionality between the cruising speed of flying bodies (insects, birds, airplanes) and body mass raised to the power 1/6. The optimized flow structures constitute robust designs, and robustness improves as the complexity of the system increases. Flow architectures that are more efficient look more natural.

Key words: *constructal theory, tree networks, self-organization, self-optimization, allometric laws*

Optimal distribution of imperfection

In this paper I draw attention to an engineering theory on the generation of patterns of "self-organization" in natural flow systems [1-3]. The work of many engineers is in the area of design – the search for better configurations (physical structures) subject to various constraints (*e. g*., materials, volume, weight, cost), where "better" is defined by the objective (function, purpose) of the design. The goodness of the design is measured globally, for example, in terms of the energy conversion efficiency of a power generation plant, or the number of miles per gallon of fuel in the performance of a new automobile. The entire external and internal flow structure of the system contributes to this final, global figure of merit. The flow structure represents geometric form (pattern, design). It is the means, or the mechanism by which the system (initially a black box) achieves its stated objective under constraints.

The natural patterns and flow structures that are being recognized today as phenomena of self-organization and self-optimization also represent geometric form. The round cross-section of a blood vessel, earth worm or subterranean river is one such form, the circle. Another geometric form that is found everywhere in natural flow systems is the tree-shaped network, or the dendritic flow pattern. A tree flow connects one point (source, sink) to an infinity of points (area, volume), as in the lungs, vascularized tissues, river basins and lightning. It is important to note that these geometric forms unite all the flow systems, the animate with the inanimate. The search for a law that covers natural self-organization and self-optimization is not restricted to living systems.

The same flow patterns also occur in engineered systems, where they result (they are deduced) from the objective and constraints principle reviewed in the first paragraph. For example, it was shown recently that the tree-shaped flow path can be deduced by minimizing the resistance to flow between a point and a finite-size volume. This development is reviewed in the section on constructal trees. The same principle of flow resistance minimization subject to constraints generates other engineered shapes, such as the round duct, the aerodynamic and hydrodynamic shapes of airborne and seaborne bodies, and the characteristic internal spacings of spaces bathed by streams (see section on optimal internal spacings). The principle-based generation of these features of flow topology has been named *constructal design*.

It is an important coincidence that the geometric forms deduced from principle in engineering look like the geometric forms that occur in large number in nature. The thought that the same kind of principle (objective, constraints) accounts for the generation of flow geometry in natural flow systems is *constructal theory*. The time arrow of theory is illustrated in fig. 1: the principle is invoked first, and the deduced (predicted) form is compared with observations of nature later. This time arrow is the opposite of, and should not be confused with the time arrow of biomimetics (the method of copying from nature). If the flow geometries predicted based on principle are supported by observations of natural flows, then the principle is a law: the constructal law.

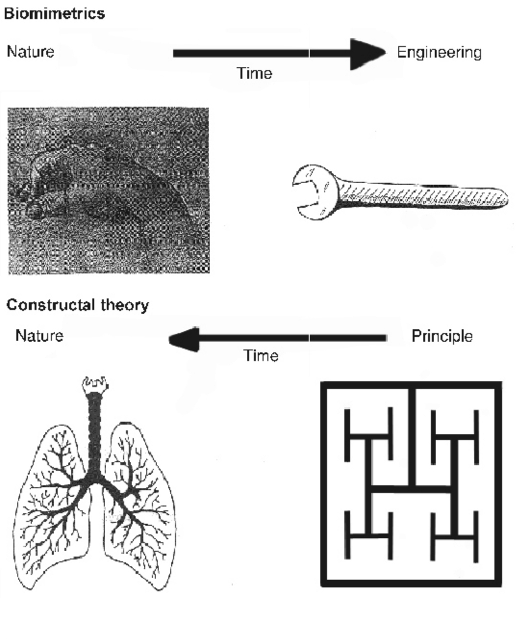

Figure 1. Proceeding against method, in time: the constructal theory of flow shape and structure in nature, against empiricism and biomimetics [1]

It is useful to review the thermodynamic meaning of the minimization of resistance to flow, or the improvement of performance. Currents (fluid, heat, electricity) that flow by overcoming resistances represent thermodynamic losses – the destruction (dissipation) of useful energy, the generation of entropy, or thermodynamic irreversibility [3]. The maximization of thermodynamic performance at the global level calls for the identification of features that allow the system to fulfill its functions while performing at the highest level possible. We seek flow architectures that operate least irreversibly. Architectures are destined to remain imperfect, because of finite-size constraints. Flows must always overcome resistances. The challenge then is to do the best possible under the specified constraints. This is accomplished by spreading the imperfections (the resistances) through the system in ways that are beneficial in the global sense. Optimal spreading constitutes the generation of physical structure – the actual being of the flow system.

Optimal distribution of imperfection is the mechanism that generates geometric form (construction, configuration) in flow systems. In constructal design, geometric form is the key unknown. Architecture is deduced, not assumed.

Constructal internal spacings

Internal spacings of certain, fine-tuned sizes occur in many engineered and natural flow systems. As an engineering example, consider a volume filled with heat-generating electronics, and think of cooling this assembly by blowing through it a stream of cold air. The global thermal resistance of the electronics package is the ratio between the maximum excess temperature registered in the hot spots ($T_{\max} - T_{\min}$) and the total rate of heat generation in the package (q). The entrance temperature of the cold air is $T_{\min}$. Desirable are designs with more components and circuitry installed in a given volume: desirable is a larger q. This can be accomplished by increasing the ceiling temperature $T_{\max}$ (usually limited by the design of the electronics), and by decreasing the global resistance $(T_{\max} - T_{\min})/q$. The latter is influenced strongly by the flow architecture.

We expect the existence of an optimal architecture based on the following reasoning. When the cooling channels (spacings) are wide and few, the total heat transfer surface is small, the global thermal resistance is high, and each heat-generating component is overheated. When the spacings are small and numerous, the cooling cannot flow easily through the package. The resistance is again high, because q is removed only by raising the operating temperature of the entire volume, *i .e.*, by increasing the outlet temperature of the coolant that manages to seep through. An optimal spacing size exists somewhere between the two extremes. In this configuration the flow volume invested in each channel is used to the maximum. Very compact and successful formulas for optimal spacings for cooling volumes have been developed based on this tradeoff [4, 5].

Natural flow systems also develop internal spacings. Bees control the hot-spot temperatures by opening vertical channels through which the air cools the swarm [1]. Inanimate flow systems also exhibit characteristic internal spacings. Cracks form at surprisingly regular intervals in volumetrically shrinking solids (*e. g.*, cracking mud). The crack size and the spacing between cracks has been deduced from the same analysis of constrained minimization of global resistance as the spacings for volumes filled with electronics [1]. The predicted spacing decreases as the wind speed increases, which is in agreement with observations.

Constructal trees

The first trees deduced entirely from principle emerged in 1996 as solutions in the maximization of flow access between a point and an area, with application to traffic patterns [6] and the cooling of small scale electronics [7]. The constructal law was stated as follows: "*For a finite-size open system to persist in time (to survive) it must evolve in such a way that it provides easier and easier access to the currents that flow through it.*" The flow path was constructed in a sequence of steps that started with the smallest building block (elemental area) and continued in time with larger building blocks (assemblies or constructs). The mode of transport with the highest resistivity (slow flow, diffusion, walking, and high cost) was placed at the smallest scales, filling completely the smallest elements. Modes of transport with successively lower resistivity (fast flow, streams, channels, vehicles, and low cost) were placed in the larger constructs, where they were used to connect the area-point or volume-point flows integrated over the constituents. The geometry of each building block was optimized for area-point access. The constructal architecture that emerged was a tree in which every geometric detail is a result – the tree, as a geometric form deduced from a principle.

The start of the construction is illustrated in figs. 2 and 3 for a two-dimensional volume that generates heat at a uniform rate per unit volume, q''' [W/m^3]. The volume must be cooled from one point. The smallest building block is the elemental rectangle H_0 L_0, the size of which is fixed. The generated heat current is fixed, $q_0 = q'''H_0L_0W$, where W is the third dimension of the elemental volume. This heat current escapes through one point. Two conductive materials inhabit the element. The low conductivity material (k_0, white) generates heat, and fills most of the element. The high-conductivity

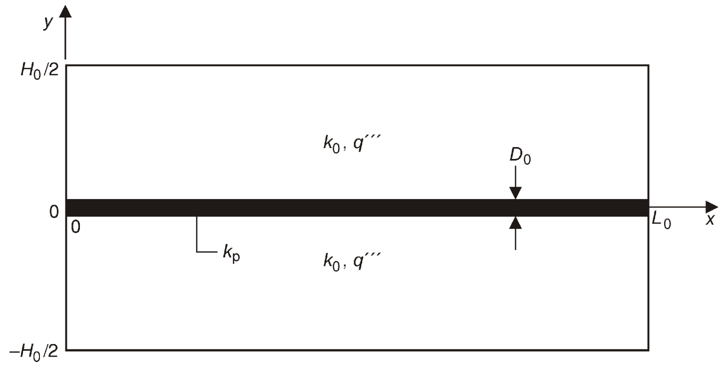

Figure 2. Elemental volume with uniform volumetric heat generation rate and high-conductivity insert along its axis of symmetry [7]

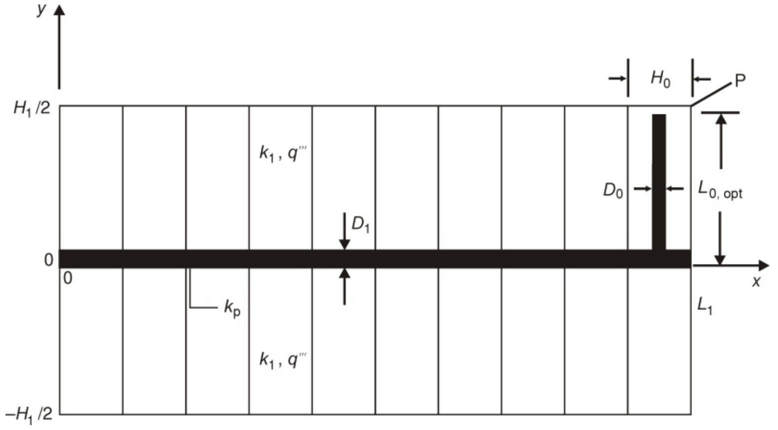

Figure 3. The first construct: a large number of elemental volumes connected to a central high-conductivity path [7]

material (k_p, black) is a small insert that guides q_0 out of the element. The shape of the element (H_0/L_0) is free to vary.

Temperatures rise all over the elemental rectangle in order to drive q_0 out of the system. The highest temperatures (T_{max}) occur in the two P corners, fig. 3. The lowest temperature (T_{min}) occurs at the q_0 exit (0, 0). The global thermal resistance of the element is $(T_{max} - T_{min})/q_0$. When the element is very shallow (small H_0/L_0) the global resistance is high because the resistance along the k_p blade is high. When the element is tall (large H_0/L_0), the global resistance is again high, this time due to the high resistance to vertical conduction through the k_0 material. There exists an intermediate, optimal shape H_0/L_0, such that the global resistance is minimum. This is the configuration where the k_p and k resistances are balanced, *i. e.*, where the imperfections have been distributed optimally, and where the result is geometric form.

The same opportunity to optimize geometry exists at larger scales. The next, larger volume is an assembly – a "first construct" – of optimized elemental volumes. Figure 3 was drawn intentionally to look like fig. 2, to suggest that the optimization of the new shape (H_l/L_l) is the same problem as the H_0/L_0 optimization. When the total volume of k_p material in the first construct is fixed, there is also an optimal ratio D_1/D_0, *i. e.*, an optimal allocation of k_p material between the D_0 and D_1 blades.

Next, a "second construct" is formed by assembling a number of first constructs. The optimal external aspect ratio turns out to be $H_2/L_2 = 2$. The ratios of k_p-blade thicknesses D_2/D_1 and D_1/D_0 can also be optimized. Here D_2 is the thickness of the central (thickest) high-conductivity insert.

The construction is continued to higher orders of assembly, until the structured composite (k_0k_p) covers the given space. Starting with the second construct, the construction settles into a pattern of pairing, where the integer 2 is an optimization result. The left frame of fig. 4 shows the optimized fourth construct, where the horizontal striations represent the D_0 blades of the elemental volumes. The right frame of fig. 4 shows the optimized eight construct, however, the high-conductivity blades in constructs earlier than the fourth construct are not shown. This frame suggests that at high enough levels of assembly most visible are the high-conductivity channels, which form a tree. The details, or the start of the construction, are visible on the left side of the figure. The smallest (elemental) scale is finite, known, and fixed. The dichotomous structure shown in the right frame is not a fractal because it does not continue *ad infinitum* toward smaller scales [3, 8]. In other words, the left frame cannot be generated by blowing up an internal region of the right frame.

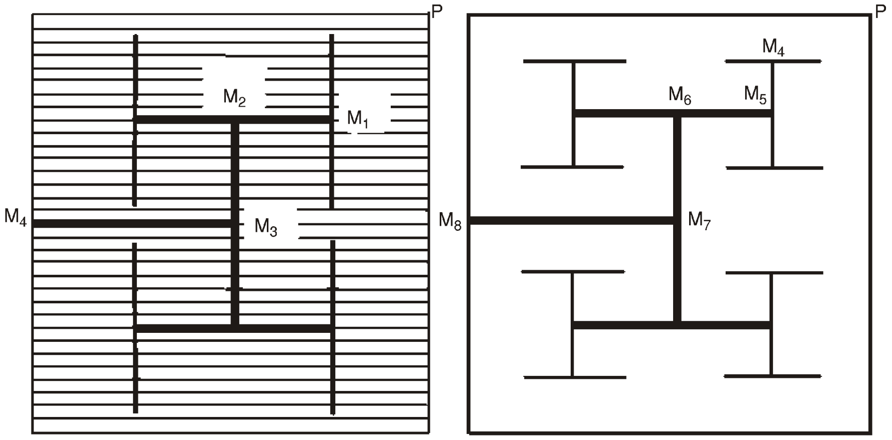

Figure 4. The optimized fourth construct (left) and eighth construct (right) for cooling a two-dimensional heat generating domain with one point heat sink [7]

The heat transfer language used in the description of the heat trees does not detract from the fundamental value of the construction. The high and low thermal conductivities are analogous to the two fluid flow regimes found in trees that carry fluid (lungs, vascularized tissues, river basins), namely, low resistivity inducts, and high resistivity due to diffusion in the surrounding spaces. The trees of traffic patterns have low resistivity (vehicles, high speed) along the streets, and high resistivity (walking, low speed) over the off-the-street areas. These and other analogies are pursued in the literature reviewed in [1].

Memory, robustness, natural look

The tree-shaped flows derived from the principle of global optimization of performance subject to constraints have several additional features that are worth stressing.

One is that the constituents assembled in each new construct were previously optimized. In other words, earlier results of optimization are memorized and compounded into progressively more complex structures.

Another feature is that the constructed trees look more natural if their freedom to provide easier access to currents is expanded. In figs. 2-4, it was assumed for simplicity that the high-conductivity blades form right angles. When this assumption is not made, *i. e.*, when each angle of confluence becomes a degree of freedom in the design, each angle can be optimized. The resulting structure has optimally inclined tributaries that point away from the root of the stem. These structures look more like the bronchial ramifications and the rivulets observed in nature.

The third feature is the robustness of the optimized flow architectures. For example, we found numerically that the optimization of angles yields a decrease of only 4 percent in the global resistance of a first construct [1]. Trees that look markedly different have nearly the same global performance. Further numerical optimization results show that robustness improves as the complexity of the tree structure increases [9].

The robustness of the tree design sheds light on why natural tree flows are never identical geometrically. They do not have to be, if the maximization of global performance is their guiding principle. The ways in which the details of natural trees may differ from case to case are without number because, unlike in the constructions exhibited in this article, the number of degrees-of-freedom of the emerging geometric form is not constrained. Local details differ from case to case because of unknown and unpredictable local features such as the heterogeneity of the natural flow medium, and the history and lack of uniformity of the volumetric flow rate that is distributed over the system. Marvelous illustrations of this element of "chance" are provided by seemingly irregular river drainage basins all over the world. The point is that the global performance and structure type (the tree) are predictable, and the principle that takes the system to this level of performance is deterministic.

The constructal law of geometric form generation: maximized flow access for currents, or maximized performance

The theoretical line reviewed in this article started with the thought that it is natural for currents to construct for themselves paths of maximum access. This led to the designed tree-shaped flows with which constructal theory began. Minimal resistance also means minimal irreversibility in systems with specified internal currents. In this way the constructal principle covers every flow system that strives to achieve better performance – (a) natural inanimate systems, such as the river basin and the cracking mud, (b) natural animate systems, such as all the animals and their respiratory and circulatory networks, and (c) engineered systems. The latter are our own extensions, and they improve our performance and chance of survival as individuals and societies.

Systems (c) are special cases of the much larger class (b), and for this reason the constructal principle means that to engineer is natural, or that nature engineers. This point is pressed with vigor by the data assembled in fig. 5. The theoretical flying speed for minimal fuel (or food) consumption is proportional to the flying mass raised to the power 1/6. This curve is well established in the biology and engineering literatures [10, 11], and was also derived based on constructal theory ([1], pp. 234-242). The flying data on insects, birds and aircraft support it convincingly.

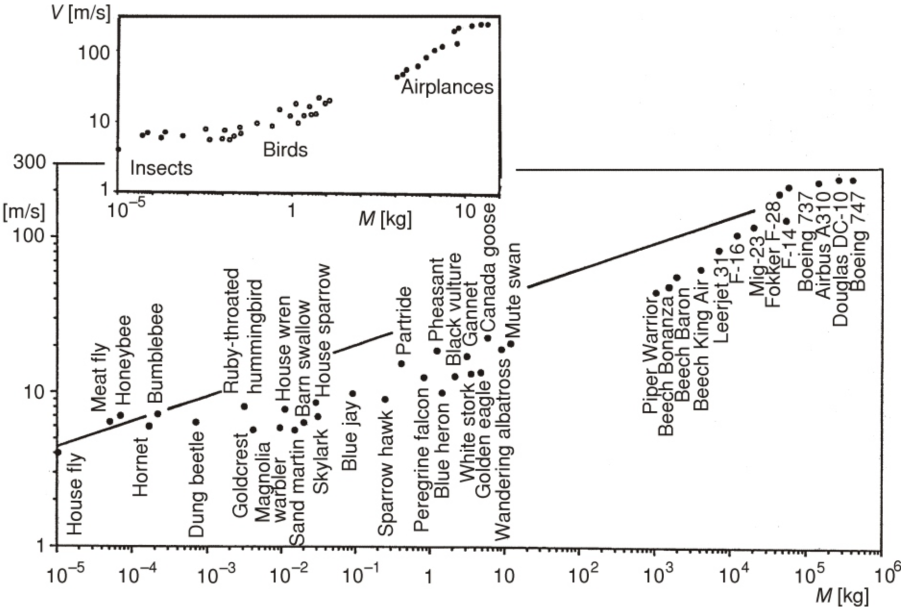

Figure 5. The flying speeds of insects, birds, and airplanes, and the theoretical speed for optimal thermodynamic performance [1]

Classes (a) and (b) are one and the same from the constructal point of view. The principle proclaims the existence of optimal and many near-optimal designs of flow architectures under constraints. In biology this goal is known as the “fittest” (or its equivalent: “the flow structure that survives”). The constructal law places this goal on a solid theoretical and universal foundation. All the systems that flow and have the ability to morph their configurations under constraints, will progress in steps of geometric form toward better performance – conglomerates of flows that flow more easily. They survive because they change, *i. e.*, they project themselves into the future by flowing through beneficially altered configurations.

The constructal law is a selfstanding principle distinct from the second law of thermodynamics. Consider an isolated thermodynamic system consisting of two parts. At time $t = 0$ the two parts are not in equilibrium. The second law proclaims the existence of the equilibrium state, were the total entropy of the system is maximal. According to the second law, the equilibrium state occurs after $t = 0$. The interior of the system is not described: it is a black box.

As an addition to this part of nature, the constructal law recognizes that beginning with $t = 0$ the flow represents a nonequilibrium system, internal currents flow between the two parts of the system, and currents always seek, find and use paths of least resistance. The currents endow the interior of the system with geometry, or structure – the flow paths of the matter that moves faster, on the background of the portions that move more slowly, by diffusion.

The generation of geometric form makes a flow system become a more efficient spreader of matter. The constructal law is the principle of geometry generation, in the way that the second law is the principle of entropy generation. The observation of geometric form is a test of constructal action, in the same way that the calculation of entropy generation is a test for second-law action.

References

[1] Bejan, A., Shape and Structure, from Engineering to Nature, Cambridge University Press, Cambridge, UK, 2000

[2] Poirier, H., A Theory Explains the Intelligence of Nature (in French), *Science & Vie,* (2003), 1034, pp. 44-63

[3] Bejan, A., Advanced Engineering Thermodynamics, 2nd ed., John Wiley and Sons, New York, USA, 1997

[4] ***, Air Cooling Technology for Electronic Equipment (Eds. S. J. Kim, S. W. Lee), CRC Press, Boca Raton, FL, USA, 1996

[5] Bejan, A., Convection Heat Transfer, 2nd ed., John Wiley and Sons, New York, USA, 1995

[6] Bejan, A., Street Network Theory of Organization in Nature, *J. Adv. Transportation,* (1996), 30, pp. 85-107

[7] Bejan, A., Constructal-Theory Network of Conducting Paths for Cooling a Heat Generation Volume, *Int. J. Heat Mass Transfer,* (1997), 40, pp. 799-816

[8] Avnir, D., Biham, O., Lidar, D., Malcai, O., Is the Geometry of Nature Fractal?, *Science*, (1998), 279, pp. 39-40

[9] Wechsatol, W., Lorente, S., Bejan, A., Tree-Shaped Insulated Designs for the Uniform Distribution of Hot Water Over an Area, *Int. J. Heat Mass Transfer,* (2001), 44, pp. 3111-3123

[10] Pennycuick, C. J., Animal Flight, Edward Arnold, London, UK, 1972

[11] Tennekes, H., The Simple Science of Flight, MIT Press, Cambridge, MA, USA, 1996

Author's address:

A. Bejan

Duke University, Department of Mechanical Engineering
and Materials Science
Durham, NC, 27708-0300, USA

E-mail: dalford@duke.edu

Paper submitted: December 2, 2004
Paper revised: January 19, 2005
Paper accepted: February 10, 2005