Pergamon

Int. J. Heat Mass Transfer. Vol. 40, No. 4, pp. 799–816, 1997

0017–9310/97 $17.00 + 0.00

PII: S0017-9310(96)00175-5

Constructal-theory network of conducting paths for cooling a heat generating volume

ADRIAN BEJAN

Department of Mechanical Engineering and Materials Science, Duke University, Box 90300, Durham, NC 27708-0300, U.S.A.

(*Received* 3 *April* 1996 *and in final form* 7 *May* 1996)

Abstract—This paper develops a solution to the fundamental problem of how to collect and 'channel' to one point the heat generated volumetrically in a low conductivity volume of given size. The amount of high conductivity material that is available for building channels (high conductivity paths) through the volume is fixed. The total heat generation rate is also fixed. The solution is obtained as a sequence of optimization and organization steps. The sequence has a definite time direction, it begins with the smallest building block (elemental system) and proceeds toward larger building blocks (assemblies). Optimized in each assembly are the shape of the assembly and the width of the newest high conductivity path. It is shown that the paths form a tree-like network, in which every single geometric detail is determined theoretically. Furthermore, the tree network cannot be determined theoretically when the time direction is reversed, from large elements toward smaller elements. It is also shown that the present theory has far reaching implications in physics, biology and mathematics. Copyright © 1996 Elsevier Science Ltd.

1. THE VOLUME-TO-POINT ACCESS PROBLEM

This paper is about one of those fundamental problems that suddenly appear 'obvious', but only after considerable technological progress has been made on pushing the frontier. The technology in this case is the cooling of electronics (components and packages), where the objective is to install a maximum amount of electronics (heat generation) in a fixed volume in such a way that the maximum temperature does not exceed a certain level. The work that has been done to devise cooling techniques to meet this objective is enormous and is continuing at an accelerated pace [1–7]. In brief, most of the cooling techniques that are in use today rely on convection or conjugate convection and conduction, where the coolant is either a single phase fluid or one that boils.

The frontier is being pushed in the direction of smaller and smaller package dimensions. There comes a point where miniaturization makes convection cooling impractical, because the ducts through which the coolant must flow take up too much space. The only way to channel the generated heat out of the electronic package is by conduction. This conduction path will have to be very effective (of high thermal conductivity, k_p), so that the temperature difference between the hot spot (the heart of the package) and the heat sink (on the side of the package) will not exceed a certain value.

Conduction paths also take up space. Designs with fewer and smaller paths are better suited for the miniaturization evolution. The fundamental problem addressed in this paper is this:

> consider a finite-size volume in which heat is being generated at every point and which is cooled through a small patch (heat sink) located on its boundary. A finite amount of high conductivity (k_p) material is available. Determine the optimal distribution of k_p material through the given volume such that the highest temperature is minimized.

I will show that the solution to this problem is astonishingly simple and beautiful, with far reaching implications in physics, mathematics and the natural evolution of living systems.

2. THE SMALLEST (ELEMENTAL) SYSTEM

We begin with recognizing the function of the electronic package or assembly of packages: the generation of heat at a certain rate (q) in a fixed volume (V). The heat generation rate per unit volume is $q''' = q/V$ and for simplicity, we assume that q''' is constant (i.e. q is distributed uniformly), although this assumption can easily be abandoned in follow-up studies of the fundamental problem.

The fraction of the volume V that is occupied by all the high conductivity paths is V_p. This fraction is also fixed, although, as noted already, a smaller ratio V_p/V is better for miniaturization. Again, for simplicity, we assume that $V_p \ll V$. The thermal conductivity of the conducting paths is constant (k_p) and much larger than the thermal conductivity of the electronic material (k_0) that occupies the rest of the volume.

Finally, since the fundamental problem leads

NOMENCLATURE

A_i	area (profile) of heat generating material
A_{pi}	area (profile) of high conductivity path
D_i	width of high conductivity path
H_i	height of the A_i area
i	order of the assembly
j	branching stage, Fig. 10
k_p	path thermal conductivity
k_0	thermal conductivity of heat generating material
L_i	length of high conductivity path
n_i	number of assemblies of order $(i-1)$ built into the assembly of order i
N_j	number of paths after the branching stage j
q	total heat transfer rate
q'''	heat generation rate per unit volume
T	temperature
V	volume
V_p	volume of high conductivity material
W	dimension perpendicular to Fig. 1
x, y	cartesian coordinates, Fig. 1.

Greek symbols

ΔT	temperature difference
λ	Lagrange multiplier
ϕ_i	porosity
Φ	function.

Subscripts

a	assembly
i	level (order) of the assembly
j	branching stage
min	minimum
opt	optimal.

eventually to a problem of geometric optimization [5], the easiest and most visible way to present the optimization principles is in a plane. For this we make the assumption that the heat generation and temperature fields are two-dimensional.

The most basic function of any portion of the conducting path is to be 'in touch' with the electronic material that generates heat volumetrically. We arrive in this way at a problem of optimal allocation of conducting path length to volume of heat generating material, or vice versa. An extremely important observation is this: the allocation cannot be made at infinitesimal scales throughout V, because the conducting paths must be of finite length so that they can be interconnected to channel the total q to one side of V, where the heat sink is located. There is only one option, namely:

(i) to optimize the allocation of conductive path to one subsystem (elemental volume) at a time;
(ii) to optimize the manner in which the elemental volumes are assembled and their k_p paths are interconnected.

This introductory discussion explains why we base the analysis on the two-dimensional elemental volume model shown in Fig. 1. This elemental volume is finite and fixed (H_0L_0W), where W is the dimension in the direction perpendicular to the plane (x, y). It is a small part of the volume V. The dimensions H_0 and L_0 may vary, while their product $A_0 = H_0L_0$ remains fixed. Indeed, the chief unknown in this first (generic) problem is the shape of the elemental volume (or the geometric aspect ratio H_0/L_0) that maximizes the access of the generated heat ($q'''H_0L_0W$) to the lone conducting path that services the elemental volume.

Symmetry and the requirement to minimize the maximum temperatures are the reasons why the conducting path is placed on one of the axes of symmetry of the elemental volume. By definition, the volume is said to be 'elemental' because its entire heat generation rate is channeled to the exterior through one end of its allocated conducting path. In other words, with the exception of the origin (0, 0), the boundary of the rectangular area $H_0 \times L_0$ is adiabatic.

The analysis that will be repeated at several levels of complexity in this paper consists of minimizing the maximum temperature difference between a point inside the volume and the heat sink point (0, 0). In the first application of this analysis (Fig. 1), we assume that the shape of the rectangular area is sufficiently slender

$$H_0 \ll L_0 \tag{1}$$

such that the conduction through the heat generating material (k_0) can be assumed to be oriented in the y direction. The validity domain of assumption (1) will be established later in equation (14). The temperature field is two-dimensional because the heat that is collected by the k_p material at $y = 0$ is conducted in the x direction toward the origin (0, 0). Integrating the equation for steady conduction with uniform heat generation in the k_0 material (e.g. Ref. [8], p. 47),

$$\frac{\partial^2 T}{\partial y^2} + \frac{q'''}{k_0} = 0 \tag{2}$$

subject to the boundary conditions

$$\frac{\partial T}{\partial y} = 0 \quad \text{at } y = \frac{H_0}{2} \tag{3}$$

$$T = T_0(x) \quad \text{at } y = 0 \tag{4}$$

we obtain

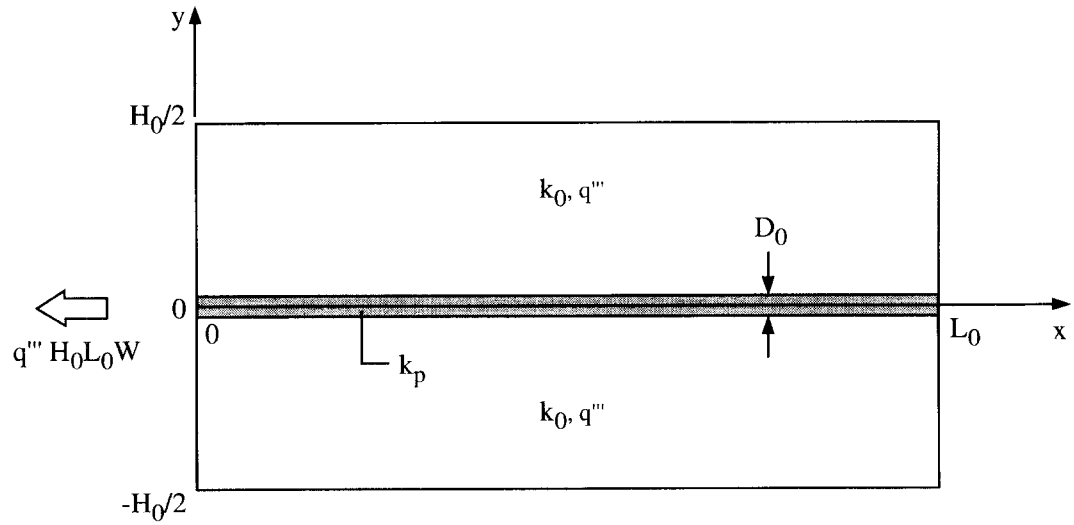


Fig. 1. Slender elemental volume with volumetric heat generation and one high-conductivity path along its axis of symmetry.

$$T(x,y)=\frac{q'''}{2k_0}(H_0y-y^2)+T_0(x). \tag{5}$$

The conduction problem along the x axis is governed by the fin equation (e.g. ref. [8], p. 63),

$$k_pD_0\frac{d^2T_0}{dx^2}+q'''H_0=0 \tag{6}$$

where $q'''H_0$ accounts for the rate at which the generated heat is being collected by the high conductivity path. Integrating equation (6) subject to

$$\frac{dT_0}{dx}=0 \quad \text{at } x=L_0 \tag{7}$$

$$T_0=T(0,0) \quad \text{at } x=0 \tag{8}$$

and substituting $T_0(x)$ into equation (5), we obtain

$$T(x,y)-T(0,0)=\frac{q'''}{2k_0}(H_0y-y^2)+\frac{q'''H_0}{k_pD_0}\left(L_0x-\frac{x^2}{2}\right). \tag{9}$$

This expression is valid strictly for $y>0$. The corresponding solution for $y<0$ can be obtained by replacing H_0 with $-H_0$.

In conclusion, equation (9) shows that the maximum temperature in Fig. 1 occurs at the corners situated farthest from the origin, namely at $x=L_0$ and $y=\pm H_0/2$. We call this maximum temperature difference ΔT_0,

$$\Delta T_0=\frac{q'''}{2k_0}\left(\frac{H_0}{2}\right)^2+\frac{q'''H_0}{k_pD_0}\frac{L_0^2}{2} \tag{10}$$

and nondimensionalize it by noting that the area ($A_0=H_0L_0$) and the volumetric heat generation rate q''' are fixed,

$$\frac{\Delta T_0}{q'''H_0L_0/k_0}=\frac{1}{8}\times\frac{H_0}{L_0}+\frac{k_0H_0}{2k_pD_0}\times\frac{L_0}{H_0}. \tag{11}$$

On the right side, the ratio D_0/H_0 is a manufacturing constant accounting for the proportion in which high-conductivity material is sandwiched with blades of the original material at the elemental level. Equation (11) shows that ΔT_0 can be minimized with respect to the shape of the elemental system (H_0/L_0). The optimization results are

$$\left(\frac{H_0}{L_0}\right)_{\text{opt}}=2\left(\frac{k_0H_0}{k_pD_0}\right)^{1/2} \tag{12}$$

$$\frac{\Delta T_{0,\min}}{q'''H_0L_0/k_0}=\frac{1}{2}\left(\frac{k_0H_0}{k_pD_0}\right)^{1/2}. \tag{13}$$

They are valid when the elemental system is slender. In view of equation (12), this condition means that the conductivity ratio k_p/k_0 should be much greater than 1, such that

$$\frac{k_p}{k_0}\gg\frac{H_0}{D_0}\gg 1. \tag{14}$$

Two additional features of the geometric optimum (12) and (13) are worth recording at this time. First, under this condition the maximum difference along the x axis (from $x=L_0$ to $x=0$) is exactly the same as the temperature drop from the hot spot ($x=L_0, y=H_0/2$) to its shadow on the x axis,

$$T(L_0,0)-T(0,0)=T(L_0,H_0/2)-T(L_0,0)=\tfrac{1}{2}\Delta T_{0,\min}. \tag{15}$$

In other words, the minimized maximum temperature difference ($\Delta T_{0,\min}$) is divided exactly in half by the bend ($x=L_0, y=0$) in the path from the hot spot ($x=L_0, y=H_0/2$) to the heat sink ($x=0, y=0$).

What we have here is a *principle of equipartition* at the geometric optimum: the temperature drop through the k_0 material equals the temperature drop through the k_p material.

The second important feature of the geometric optimum can be seen by combing equations (12) and (13):

$$\Delta T_{0,\min} = \frac{q''' H_0^2}{4k_0}. \tag{16}$$

This means that at the elemental level, the excess temperature decreases as H_0^2, hence the incentive for manufacturing the smallest possible elemental system. The significance of equations (15) and (16) is considerably greater, as we will see in equations (22) and (56).

3. THE FIRST ASSEMBLY

To review the progress made so far, we started with the operational constraint (q''') and the dimensions of the smallest element ($L_0 H_0, D_0$) and determined the optimal shape of the element, or the optimal length of the high-conductivity path allocated to the element. The heat generated by the elemental volume is ducted as a single current through the origin (0, 0) of Fig. 1. There are many elemental heat currents of this kind. The next problem is how to connect and lead these currents out of the greater volume V.

One way to connect the elemental heat currents is shown in Fig. 2. A large number of optimized elemental volumes ($H_0, L_{0,\text{opt}}$) are lined up on both sides of new high-conductivity path of width D_1, such that their heat currents are collected by the new path. One elemental volume is highlighted in the upper right-hand corner of the new (and fixed) rectangular area $H_1 \times L_1$. The outer boundary of this area is adiabatic except for the D_1 patch over the origin, through which the collected heat current ($q''' H_1 L_1 W$) is led to the outside. The problem is how to shape the $H_1 \times L_1$ assembly, or how many elemental volumes to connect together in this way, such that the maximum temperature difference between a point in the assembly and the origin (0, 0) is minimum.

Figure 2 was drawn intentionally to look like Fig. 1, so that we may see at a glance that the optimization of the $H_1 \times L_1$ shape is the same problem as in Section 2. To start with, the conduction on both sides of the x axis of Fig. 2 is oriented in the y direction because of the paths of width D_0 that serve each elemental volume. This time, i.e. at this first level of assembly, the y direction on heat conduction is man-made, it does not hinge on a slenderness assumption like equation (1). In a volume averaged sense, the composite material situated above and below $y = 0$ in Fig. 2 behaves the same as the k_0 material of Fig. 1, except that its effective (vertical) thermal conductivity (k_1) is different. By using an energy conservation argument, it is easy to see that

$$k_1 = k_p \frac{D_0}{H_0}. \tag{17}$$

This equation illustrates the constitution of the $H_1 \times L_1$ system, which is an assembly of many smaller systems of shape $H_0 \times L_{0,\text{opt}}$. The second equation that describes this process of organization (assembly) is purely geometric:

$$H_1 = 2L_{0,\text{opt}}. \tag{18}$$

The temperature distribution over the $H_1 \times L_1$ face can be determined by repeating the analysis shown between equations (2)–(9). The maximum temperature difference (ΔT_1) occurs again between the farthest corner ($L_1, H_1/2$) and the origin (0, 0),

$$\frac{\Delta T_1}{q''' H_1 L_1 / k_1} = \frac{1}{8} \times \frac{H_1}{L_1} + \frac{k_1 H_1}{2k_p D_1} \times \frac{L_1}{H_1} \tag{19}$$

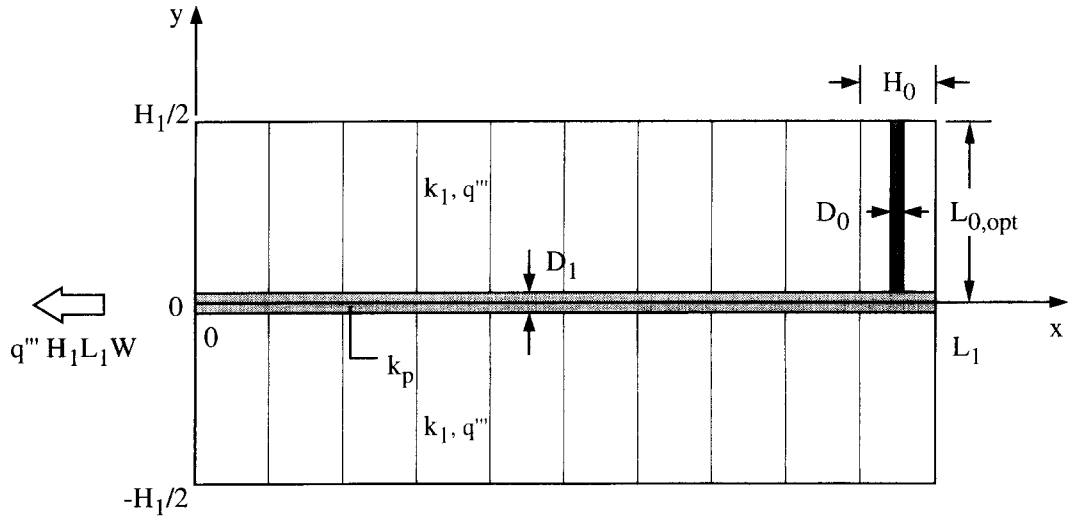


Fig. 2. The first assembly: a large number of elemental volumes (Fig. 1) connected to a central high-conductivity path.

and has a minimum with respect to shape (or geometric aspect ratio H_1/L_1), namely

$$\left(\frac{H_1}{L_1}\right)_{\text{opt}} = 2\left(\frac{k_1}{k_p}\frac{H_1}{D_1}\right)^{1/2} \tag{20}$$

$$\frac{\Delta T_{1,\min}}{q''' H_1 L_1/k_1} = \frac{1}{2}\left(\frac{k_1 H_1}{k_p D_1}\right)^{1/2}. \tag{21}$$

These results can be expressed in terms of the properties and dimensions of the optimized elemental volume by using equations (12), (17) and (18),

$$\Delta T_{1,\min} = \frac{q''' H_0^2}{4k_0} \tag{22}$$

$$\frac{L_{1,\text{opt}}}{H_0} = \frac{1}{2}\left(\frac{k_p}{k_0}\right)^{1/4}\left(\frac{D_0}{H_0}\right)^{1/4}\left(\frac{D_1}{H_0}\right)^{1/2}. \tag{23}$$

Comparing equations (22) and (16), we reach the surprising conclusion that the optimized assembly sustains a maximum temperature difference that equals the maximum temperature difference of the optimized elemental volume. Furthermore, by analogy with equation (15), it can be shown that $\Delta T_{1,\min}$ is divided in half by the 'bend' in the high-conductivity path, namely the point $(L_{1,\text{opt}}, 0)$ in Fig. 2. Finally, the optimal number of elemental volumes that must be grouped together is

$$n_{1,\text{opt}} = 2\frac{L_{1,\text{opt}}}{H_0} = \left(\frac{k_p D_1^2}{k_0 D_0 H_0}\right)^{1/4}. \tag{24}$$

The optimized assembly contains an additional degree of freedom that is represented by the path width D_1. The selection of D_1 must be done in relation to D_0, because the high-conductivity material is presumably in short supply. The total amount of k_p material is represented by the area

$$A_{p,1} = D_1 L_{1,\text{opt}} + n_{1,\text{opt}} D_0 L_{0,\text{opt}}. \tag{25}$$

This material constraint can be nondimensionalized by using the overall size of the assembly $(H_1 L_{1,\text{opt}})$, which is fixed,

$$\phi_1 = \frac{A_{p,1}}{H_1 L_{1,\text{opt}}}. \tag{26}$$

Equation (26) is a definition for the 'porosity' ($\phi_1 \ll 1$) of the assembly, if we view the assembly as a medium (k_0) with fissures filled by k_p material. Combining equation (25) with equations (12) and (23) we rewrite the constraint as

$$\phi_1 = \left(\frac{k_0}{k_p}\right)^{1/2}\frac{D_1}{H_0}\left(\frac{D_0}{H_0}\right)^{-1/2} + \frac{D_0}{H_0}. \tag{27}$$

Let us optimize D_1 such that the overall thermal resistance of the assembly is minimum subject to the constraint (27). The temperature difference is given by equation (22), however, a more accurate estimate can be made by recalling the equipartition principle (15) and that $\Delta T_{1,\min} = \Delta T_{0,\min}$. With reference to Fig. 2, we note that the temperature difference from the farthest corner $(L_1, H_1/2)$ to the shut end of the D_1 line [the point $(L_1, 0)$] is $\Delta T_{0,\min}$ and the difference from $(L_1, 0)$ to $(0, 0)$ is $\frac{1}{2}\Delta T_{1,\min}$. The total temperature difference of the first assembly,

$$\Delta T_{a1} = \Delta T_{0,\min} + \tfrac{1}{2}\Delta T_{1,\min} \tag{28}$$

can be nondimensionalized as the thermal resistance

$$\frac{\Delta T_{a1}}{q''' H_1 L_1/k_0} = \frac{3}{4}\left(\frac{k_0}{k_p}\right)^{3/4}\left(\frac{D_1}{H_0}\right)^{-1/2}\left(\frac{D_0}{H_0}\right)^{-1/4}. \tag{29}$$

Minimizing this expression subject to constraint (27) we obtain

$$\phi_1 = 2\frac{D_0}{H_0} \tag{30}$$

$$\frac{D_{1,\text{opt}}}{H_0} = \left(\frac{k_p}{k_0}\right)^{1/2}\left(\frac{1}{2}\phi_1\right)^{3/2} = \left(\frac{k_p}{k_0}\right)^{1/2}\left(\frac{D_0}{H_0}\right)^{3/2}. \tag{31}$$

In other words, when the high-conductivity material is scarce, the new path must be wider than the elemental paths. We conclude on a purely theoretical basis then that the path width must have an optimal magnification factor

$$\left(\frac{D_1}{D_0}\right)_{\text{opt}} = \left(\frac{D_0 k_p}{H_0 k_0}\right)^{1/2} \gg 1. \tag{32}$$

The corresponding minimum of the overall temperature difference of the first assembly is

$$\Delta T_{a1,\min} = \frac{3}{8}\frac{q''' H_0^2}{k_0}. \tag{33}$$

This result is very interesting (and important in practice) because it is nearly the same as the temperature difference at the elemental level [equation (16)]. Without the network of high-conductivity paths, the excess temperature of the first assembly would have been of the order $q''' H_1^2/k_0$ (see, for example, ref. [8], p. 49). In other words, the optimal use of k_p-paths decreases the excess temperature of the first assembly by a factor of order $(H_0/H_1)^2 \sim k_0 H_0/(k_p D_0) \ll 1$.

The optimal path widths (30) and (31) can be substituted into the earlier results of this section, to determine the geometry of a first assembly that has been optimized twice (with respect to the $H_1 \times L_1$ shape and the allocation of finite k_p material). For example, the optimal number of elemental systems (24) and the length of the D_1 path (23) become

$$n_{1,\text{opt}} = \left(\frac{k_p D_0}{k_0 H_0}\right)^{1/2} \gg 1 \tag{34}$$

$$L_{1,\text{opt}} = \frac{1}{2}\left(\frac{k_p}{k_0} H_0 D_0\right)^{1/2}. \tag{35}$$

The astonishing geometric feature that emerges out of these results [specifically, out of equations (35), (18) and (12)] is that the optimal aspect ratio of the first assembly is constant and exactly equal to the integer 2:

$$\left(\frac{H_1}{L_1}\right)_{\text{opt}} = 2. \tag{36}$$

This ratio is independent of the type of conducting materials (k_p/k_0) and the proportion in which they are built into the assembly (ϕ_1). It means that the optimal rectangular shape $H_1 \times L_1$ is such that a square of side L_1 forms on either side of the D_1 path (x axis). Another way to interpret this result is by combining equation (36) with $H_1 = 2L_{0,\text{opt}}$ to obtain $(L_1/L_0)_{\text{opt}} = 1$.

4. THE SECOND ASSEMBLY

The way to connect the D_1 paths into assemblies of larger scales is now clear. Figure 3(a) shows the second assembly, which is composed of a large number (n_2) of first assemblies aligned on both sides of a path of width D_2. This new assembly covers the area $H_2 \times L_2$, which is fixed. The first assembly optimized in the preceding section is highlighted in the upper right-hand corner of Fig. 3(a), where $H_2 = 2L_{1,\text{opt}}$ (note the squares) and $k_2 = k_p D_1/H_1$.

The optimization of the second assembly consists of the same steps as the optimization of the first assembly. For brevity, we review this sequence by listing only the key results. The optimal aspect ratio and corresponding maximum excess temperature are

$$\left(\frac{H_2}{L_2}\right)_{\text{opt}} = 2\left(\frac{k_2 H_2}{k_p D_2}\right)^{1/2} \tag{37}$$

$$\frac{\Delta T_{2,\min}}{q''' H_2 L_2/k_2} = \frac{1}{2}\left(\frac{k_2 H_2}{k_p D_2}\right)^{1/2}. \tag{38}$$

It can be shown that $\Delta T_{2,\min} = \Delta T_{1,\min} = \Delta T_{0,\min}$ and that the equipartition principle (15) holds here as well. In place of equation (23) we obtain

$$\frac{L_{2,\text{opt}}}{H_0} = \frac{1}{2}\left(\frac{k_p}{k_0}\right)^{3/8}\left(\frac{D_0}{H_0}\right)^{1/8}\left(\frac{D_1}{H_0}\right)^{-1/4}\left(\frac{D_2}{H_0}\right)^{1/2}. \tag{39}$$

The optimal number of first assemblies connected into the second assembly is

$$n_{2,\text{opt}} = \left(\frac{k_p}{k_0}\right)^{-1/8}\left(\frac{D_0}{H_0}\right)^{-3/8}\left(\frac{D_1}{H_0}\right)^{-1/4}\left(\frac{D_2}{H_0}\right)^{1/2}. \tag{40}$$

The optimization with respect to the second degree of freedom (D_2) begins with estimating the overall maximum temperature difference associated with the second assembly [see the discussion that preceded equation (29)], namely $\Delta T_{a2} = \Delta T_{0,\min} + \frac{1}{2}\Delta T_{1,\min} + \frac{1}{2}\Delta T_{2,\min}$, or

$$\frac{\Delta T_{a2}}{q''' H_2 L_2/k_0} = \left(\frac{k_0}{k_p}\right)^{5/8}\left(\frac{D_0}{H_0}\right)^{1/8}\left(\frac{D_1}{H_0}\right)^{-1/4}\left(\frac{D_2}{H_0}\right)^{-1/2}. \tag{41}$$

The path width D_1, however, was optimized in equation (31), therefore equation (41) becomes

$$\frac{\Delta T_{a2}}{q''' H_2 L_2/k_0} = \left(\frac{k_0}{k_p}\right)^{3/4}\left(\frac{D_2}{H_0}\right)^{-1/2}\left(\frac{D_0}{H_0}\right)^{1/4}. \tag{42}$$

The similarities between this expression and equation (29) are worth noting. Finally, the high-conductivity material constraint for the second assembly, $A_{p2} = D_2 L_{2,\text{opt}} + n_{2,\text{opt}} A_{p1}$, can be cast in a form similar to equation (27),

$$\phi_2 = \frac{A_{p2}}{H_2 L_2} = \left(\frac{k_0}{k_p}\right)^{1/2}\frac{D_2}{H_0}\left(\frac{D_0}{H_0}\right)^{-1/2} + 2\frac{D_0}{H_0}. \tag{43}$$

The optimal path width D_2 that minimizes the resistance (42) subject to the constraint (43) is, after using equation (31),

$$D_{2,\text{opt}} = 2\left(\frac{k_p}{k_0}\right)^{1/2} H_0^{-1/2} D_0^{3/2} = 2D_{1,\text{opt}}. \tag{44}$$

The important implication of the D_2-optimization stage that we just completed is that equation (37) becomes $(H_2/L_2)_{\text{opt}} = 2^{1/2}$ and equation (40) now reads $n_{2,\text{opt}} = 2^{1/2}$. Such a small $n_{2,\text{opt}}$ value means that the assumed (large n_2) second assembly sketched in Fig. 3(a) falls definitely on one side of the optimum with respect to D_2 and that the best second assembly is the one that contains the smallest number ($n_2 = 2$) of first assemblies. This conclusion stands behind the second assembly constructed in Fig. 3(b): this geometry is yet to be optimized with respect to D_2, because equation (44) is not accurate since n_2 is not large.

According to equation (36), when we put two first assemblies together we obtain a square of side H_1, in other words

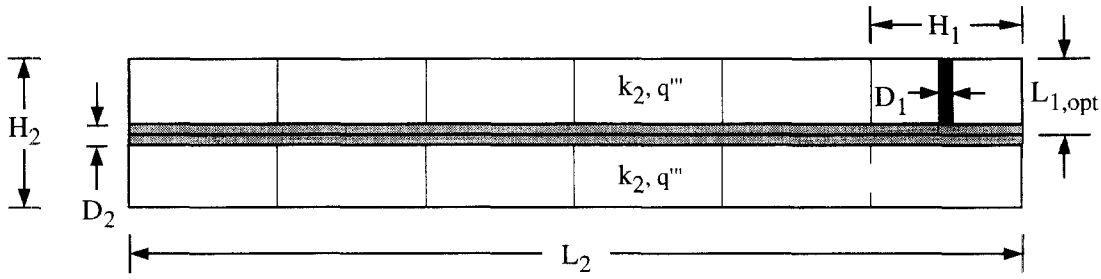
$$H_2 = H_1. \tag{45}$$

The two D_1 paths meet in the center of the square and join into a path of width D_2 and length

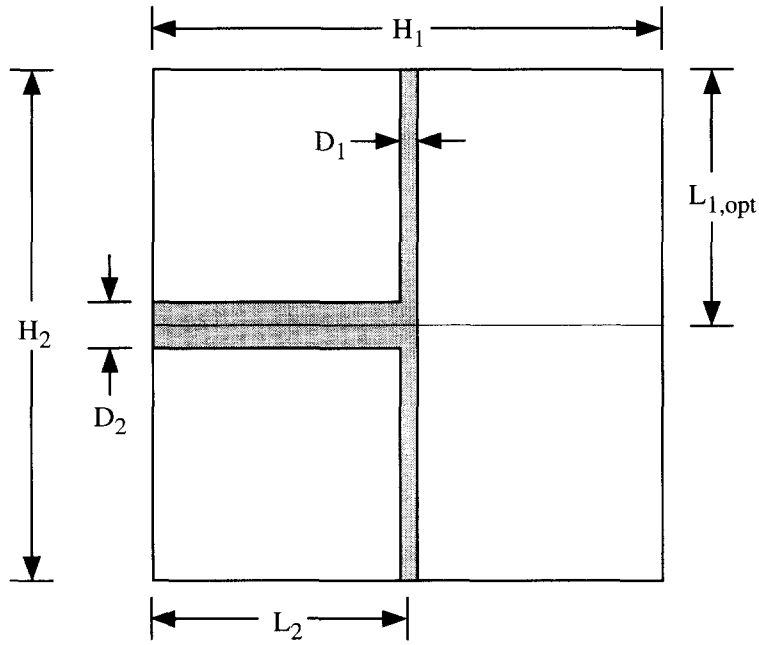
$$L_2 = \tfrac{1}{2}H_1. \tag{46}$$

Note that the D_2 path need not extend all the way across the large square, because the right side of the square is adiabatic. This material saving feature was not necessary in Figs. 2 and 3(a), because in those cases the assumed number of tributaries (D_0 and, respectively, D_1) was very large.

The total temperature drop along the D_2 path, namely ΔT_{D_2}, can be estimated by writing that the total heat current collected from the pair of first assemblies ($q''' H_1^2$) is the same as the conduction heat current conserved along the D_2 path, namely $k_p D_2 \Delta T_{D_2}/L_2$. Next, the overall excess temperature sustained by the second assembly ($\Delta T_{a2} = \Delta T_{D_2} +$



(a)



(b)

Fig. 3. (a) Second assembly containing a large number of first assemblies; (b) the best second assembly: a square made up of two optimized first assemblies.

$\Delta T_{a1,\min}$) can be put in a resistance form that shows explicitly the effect of the new degree of freedom (D_2),

$$\frac{\Delta T_{a2}}{q''' H_1^2/k_0} = \frac{1}{2}\left(\frac{k_0}{k_p}\right)^{1/2}\left(\frac{D_2}{H_0}\right)^{-1}\left(\frac{D_0}{H_0}\right)^{1/2} + \frac{3}{8}\frac{k_0}{k_p}\left(\frac{D_0}{H_0}\right)^{-1}. \tag{47}$$

The amount of high-conductivity material present in the second assembly (Fig. 3(b)) is proportional to the area $A_{p2} = D_2 L_2 + 2A_{p1}$. This quantity can be nondimensionalized as $\phi_2 = A_{p2}/H_2^2$, which, after some algebra, becomes

$$\phi_2 = \frac{1}{2}\left(\frac{k_0}{k_p}\right)^{1/2}\frac{D_2}{H_0}\left(\frac{D_0}{H_0}\right)^{-1/2} + 2\frac{D_0}{H_0}. \tag{48}$$

Minimizing the resistance (47) with respect D_2, subject to equation (48), we obtain

$$\frac{D_{2,\text{opt}}}{H_0} = \frac{4}{3^{1/2}}\left(\frac{k_p}{k_0}\right)^{1/2}\left(\frac{D_0}{H_0}\right)^{3/2} \tag{49}$$

and the ϕ_2 and $\Delta T_{a2,\min}$ expressions shown on the $i = 2$ line of Table 1.

We conclude that the excess temperature of the optimized second assembly continues to be of the same order as in the first assembly, equation (33), and the elemental system, equation (16). Another interesting conclusion is reached by dividing equations (49) and (31): in going from the first assembly to the second assembly, the width of the high-conductivity

Table 1. The main parameters of the first six optimized assemblies

Assembly level i	Assembly height factor H_i/H_{i-1}	Assembly shape[a]	Path width factor D_i/D_{i-1}	Path length factor L_i/L_{i-1}	$\frac{\Delta T_{ai,\min}}{q'''H_0^2/k_0}$	$\phi_i \frac{H_0}{D_0}$
0		r			$\frac{1}{4}$	1
1	$\left(\frac{D_0k_p}{H_0k_0}\right)^{1/2} \gg 1$	R	$\left(\frac{D_0k_p}{H_0k_0}\right)^{1/2} \gg 1$	1	$\frac{3}{8}$	2
2	1	S	$\frac{4}{3^{1/2}} = 2.31$	1	$\frac{3}{8}\left(1+\frac{1}{3^{1/2}}\right)$	$2\left(1+\frac{1}{3^{1/2}}\right)$
3	2	R	2	1	$\frac{3}{8}\left(1+\frac{2}{3^{1/2}}\right)$	$2\left(1+\frac{2}{3^{1/2}}\right)$
4	1	S	2	2	$\frac{3}{8}\left(1+\frac{4}{3^{1/2}}\right)$	$2\left(1+\frac{4}{3^{1/2}}\right)$
5	2	R	2	1	$\frac{3}{8}\left(1+\frac{6}{3^{1/2}}\right)$	$2\left(1+\frac{6}{3^{1/2}}\right)$
6	1	S	2	2	$\frac{3}{8}\left(1+\frac{10}{3^{1/2}}\right)$	$2\left(1+\frac{10}{3^{1/2}}\right)$

[a]S = square; R = rectangle (aspect ratio = 2); r = rectangle (aspect ratio $\ll$ 1).

path is magnified by a constant factor, $(D_2/D_1)_{\text{opt}} = 4/3^{1/2} = 2.31$. This result is nearly the same as in equation (44), which as noted earlier, is inaccurate because n_2 is not large.

5. THE THIRD ASSEMBLY

To construct the best third assembly we perform a double optimization (assembly shape, path width) of a system containing a large number of optimized second assemblies. The procedure was illustrated in the preceding section, by using Fig. 3(a). This approach leads again to the conclusion that the best third assembly is the one with the smallest number of constituents, namely two. This construction is shown in Fig. 4.

The geometric shape of the smallest third assembly is fixed ($H_3 = 2H_2$). The length of the D_3 path is fixed as well: $L_3 = H_2/2$. The lone degree of freedom is the path width D_3, which can be optimized in relation to the other path widths (D_2, D_1, D_0), such that the excess temperature associated with the third assembly is minimum. The temperature drop along the D_3 path is $\Delta T_{D_3} = q'''H_1^2/k_pD_3$. The overall excess temperature of the third assembly is $\Delta T_{a3} = \Delta T_{D_3} + \Delta T_{a2,\min}$, which leads to the resistance formula

$$\frac{\Delta T_{a3}}{q'''H_3H_2/k_0} = \frac{1}{2}\left(\frac{k_0}{k_p}\right)^{1/2}\left(\frac{D_3}{H_0}\right)^{-1}\left(\frac{D_0}{H_0}\right)^{1/2} + \frac{3}{16}(1+3^{-1/2})\frac{k_0}{k_p}\left(\frac{D_0}{H_0}\right)^{-1}. \quad (50)$$

The total amount of k_p material is proportional to $A_{p3} = D_3L_3 + 2A_{p2}$, which can be expressed as

$$\phi_3 = \frac{A_{p3}}{H_3H_2} = \frac{1}{4}\left(\frac{k_0}{k_p}\right)^{1/2}\frac{D_2}{H_0}\left(\frac{D_0}{H_0}\right)^{-1/2} + 2(1+3^{-1/2})\frac{D_0}{H_0}. \quad (51)$$

The results of the resistance minimization analysis are

$$\frac{D_{3,\text{opt}}}{H_0} = \frac{8}{3^{1/2}}\left(\frac{k_p}{k_0}\right)^{1/2}\left(\frac{D_0}{H_0}\right)^{3/2} \quad (52)$$

and the ϕ_3 and $\Delta T_{a3,\min}$ expressions listed in Table 1. These parameters have the same scales as in the optimized second and first assemblies. The sizes (i.e. the numerical coefficients) exhibit moderate increases from one assembly level to the next. Particularly interesting is that this time the path width doubles, $D_{3,\text{opt}} = 2D_{2,\text{opt}}$.

6. THE FOURTH ASSEMBLY AND HIGHER-ORDER ASSEMBLIES

The fourth assembly is the last one to require a separate analysis, because it is still unlike any of its smaller constituents and consequently, its optimal parameters cannot be deduced based on recurrence formulas. One construction optimization rule that emerged beginning with the second assembly is that each new assembly must contain two optimized assemblies of the immediately smaller size. This pairing rule is used now in Fig. 5, which shows that the best fourth assembly is composed of only two optimized third assemblies.

The analytical steps for determining $D_{4,\text{opt}}$ and

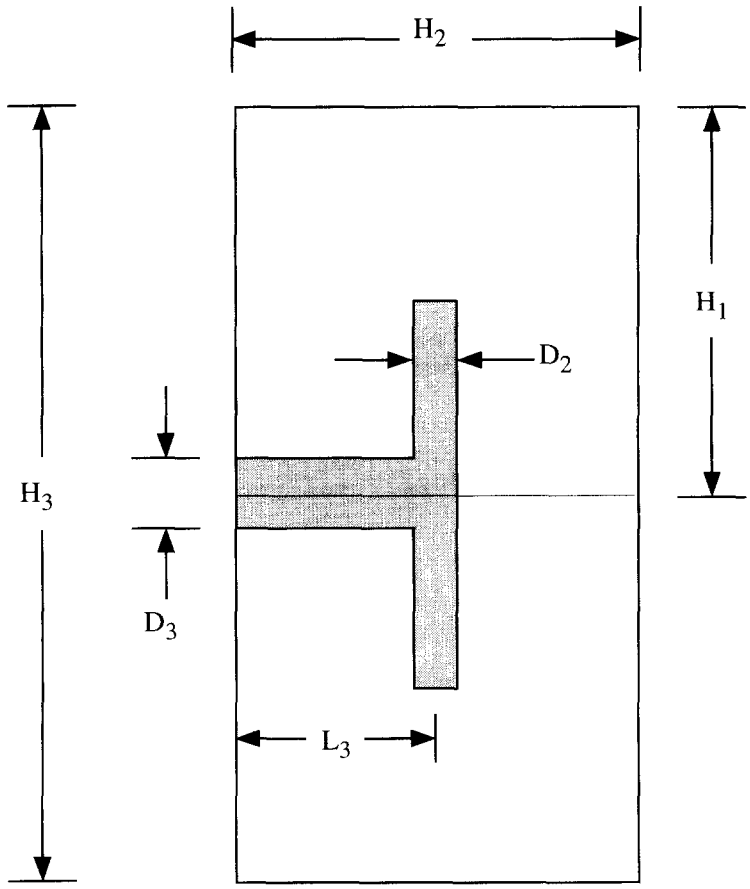


Fig. 4. The best third assembly : a rectangle made up of two optimized second assemblies.

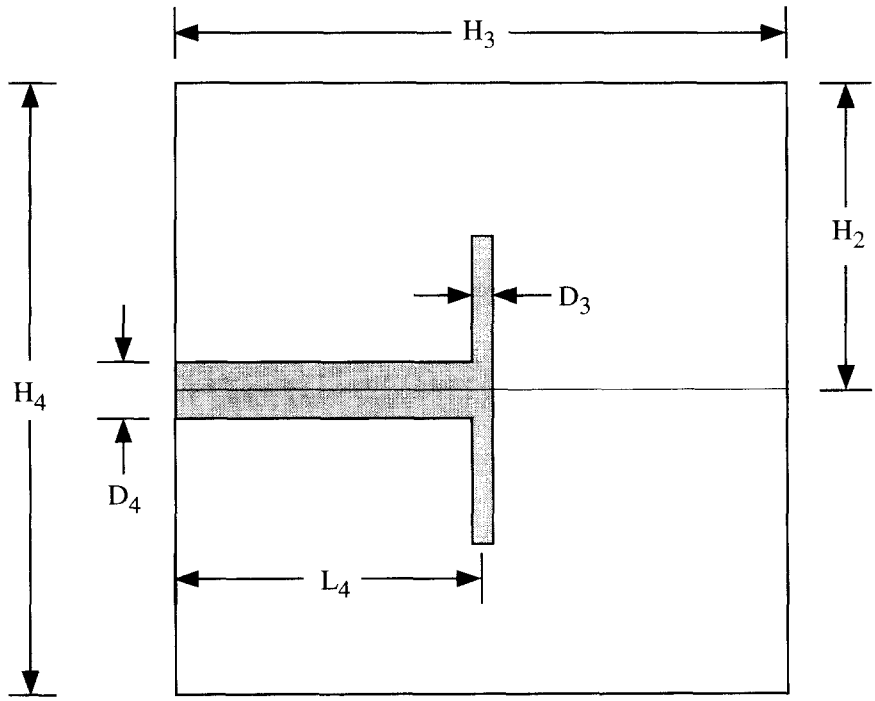


Fig. 5. The best fourth assembly : a square made up of two optimized third assemblies.

$\Delta T_{a4,\min}$ are the same as in the preceding section. For brevity, we report only the results, which are listed on the $i = 4$ line of Table 1. The fourth assembly shows that the doubling of the path width establishes itself as an optimization rule, such that for higher-order assemblies we have

$$D_{i,\text{opt}} = \frac{2^i}{3^{1/2}} D_0 \left(\frac{D_0 k_p}{H_0 k_0}\right)^{1/2} \quad (i \geq 2). \tag{53}$$

Table 1 also shows the optimization results for the fifth and sixth assemblies, which were developed step by step (with pencil and paper) as in the preceding sections. These results show that the path length settles into a pattern of doubling after two consecutive assemblies. The path length of higher-order assemblies is given by

$$L_i = 2^{(i/2)-m}\left(\frac{k_p}{k_0} H_0 D_0\right)^{1/2} \quad (i \geq 3) \tag{54}$$

where $m = 2$ when i is even and $m = 5/2$ when i is odd. A similar doubling pattern is exhibited by the height of the optimized assembly,

$$H_i = 2^{(i/2)-p} H_0 \left(\frac{D_0 k_p}{H_0 k_0}\right)^{1/2} \quad (i \geq 2) \tag{55}$$

where $p = 1$ when i is even and $p = 1/2$ when i is odd. The optimized assemblies coalesce (two at a time) in an alternating sequence of rectangles (e.g. Fig. 4) and squares (e.g. Fig. 5).

The performance of higher-order assemblies is described analytically by

$$\Delta T_{ai,\min} = \frac{3}{16} \frac{\phi_i}{D_0/H_0} \frac{q''' H_0^2}{k_0}. \tag{56}$$

By using equations (53)–(55) we can derive the recurrence relation for ϕ_i,

$$\phi_i = \phi_{i-1} + \frac{2^{(i-2r)/2}}{3^{1/2}} \frac{D_0}{H_0} \quad (i \geq 3) \tag{57}$$

where $r = 0$ when i is even and $r = 1/2$ when i is odd. In particular, when i is odd, equation (57) yields

$$\phi_i = \left[2 + \frac{8}{3^{1/2}} \times \frac{2^{(i-1)/2} - 1}{i-3}\right] \frac{D_0}{H_0} \quad (i = 5, 7, \ldots). \tag{58}$$

The assemblies of order two and higher are not optimal in a mathematical sense. They are close enough to being optimal, as equations (44) and (49) have shown. In any case, as building blocks, they are *the best that fit together.*

Figure 6(a) is a more detailed presentation of the optimized fourth assembly (Fig. 5), showing this time all the constituents (smaller assemblies) including the elemental volumes, which are shown as striations. The number of such striations is not a constant: it varies according to parameter $2(k_0 H_0/k_p D_0)^{1/2}$, cf. equation (12). The optimized width of the high conductivity paths increases as the order of the assembly increases.

Figure 6(b) is a sketch of only the high-order end of the assembling sequence. Once again, every feature of this drawing was the result of a deterministic process of optimization and organization, c.f. Table 1. In this drawing the k_p-path widths double as the assembly size increases. The finer details of the heat generating volume are left to the assemblies of lower order. For example, if the square shaded in the upper-right corner of Fig. 6(b) is one of the optimized fourth assemblies of Fig. 6(a), then the largest square of Fig. 6(b) represents the optimized eighth assembly.

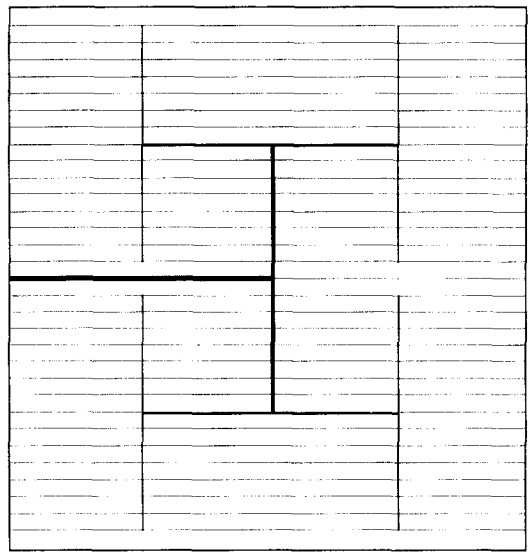
This point is made more clearly by the color print shown in Fig. 7. The largest square is the optimized sixth assembly. The heat generating material is shown in blue. Two colors are used for the high conductivity paths (first yellow and then red) to draw attention to the fact that the first two levels of assembly have construction rules (magnification factors) that differ from those of the higher-order assemblies.

The present analysis was set in the limit of $k_p/k_0 \gg 1$, $\phi_i \ll 1$ and paths at right angles, because its focus was on the fundamental 'cooling access' problem of Section 1, not on the design and manufacture of an actual device. Our objective was to demonstrate a deterministic principle of optimization and organization. These limiting assumptions can be relaxed in future studies of applications.

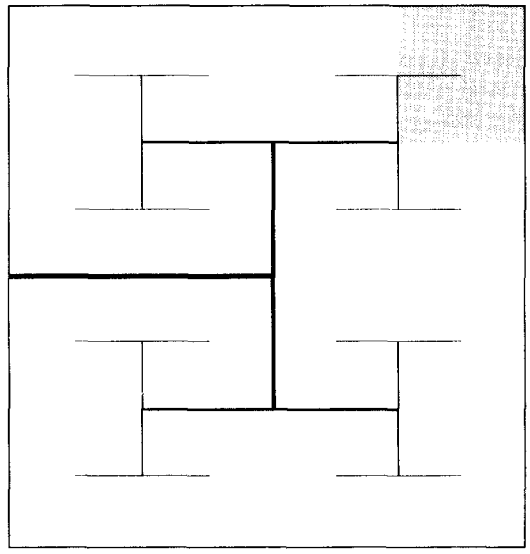
7. THE IMPORTANCE OF THE DIRECTION OF TIME

In Figs. 6(a) and (b) we have constructed, building block on building block, the network that maximizes the thermal conductance between a large heat generating volume and a spot on its boundary. The new aspect of this construction that every single step was determined based on theory. From the shape of the elemental volume (L_0/H_0) to the path width of the largest assembly (D_n), we relied on the minimization of the maximum temperature difference subject to volume constraints. This completely deterministic approach gave us the optimal shape of each assembly, the optimal number and orientation of constituents in each new assembly, and the optimal dimensions of each new high-conductivity path. The construction began with the smallest element, in which heat was generated volumetrically, and proceeded step-by-step (eventually through pairing) toward assemblies of larger sizes.

To appreciate how much is new in this theory, it is important to note that one portion of the network pattern (namely, only the portion formed by the



(a)



(b)

Fig. 6. (a) The complete structure of the fourth assembly, showing the optimized third, second and first assemblies and the striations associated with each elemental volume; (b) the optimized network formed by the higher-order assemblies.

higher-order assemblies, Fig. 6(b)) is not new. It was first proposed in physiology as a heuristic model for the circulatory system [9], where it was known empirically that each tube is followed by two smaller tubes, i.e. each tube undergoes bifurcation. It was also known that the tube diameter must decrease by a constant factor ($2^{-1/3}$) during each bifurcation: this result was derived based on flow resistance minimization [10] and is the only theory-based notion present in the algorithms used to reconstruct the pulmonary tree or other tree-shaped networks that appear in nature (trees, roots, leaves, river basins, deltas, lightning, lungs, nervous system, vascular system). The description of these geometric constructions has been made popular through the advent of fractal geometry: in fact, a two-dimensional version of Cohn's branching fluid network (similar to Fig. 6(b)) appeared in the books by Mandelbrot [11] and Pri-

gogine [12], where it is presented heuristically as a 'model of the lung'. These authors cannot (and do not) say anything predictive or deterministic *en route* to the algorithm used in making the drawing. Artists who design patterns for bathroom floor tiles enjoy the same creative freedom.

Man's ability to anticipate the branching fluid network on a purely theoretical basis remained limited to one result: the constant factor for diameter reduction during branching [which, upon closer scrutiny, turns out to have been empirical not theoretical, see equations (65) and (66)]. The construction steps that were left to be determined theoretically are the number of new (smaller) ducts formed during branching (why two branches and not six?), the relation between the branch length and the length of the original (larger) duct, and the position of the smaller branches relative to larger branches, i.e. the manner in which the network fills the volume. Another extremely important aspect that awaited an explanation is why the theoretical diameter reduction factor ($2^{-1/3}$) fails to describe the sizes of the smallest ducts. In other words, why does the heuristic construction of an algorithm-based network break down at a sufficiently small scale (as in Fig. 6)? What is that small scale? Or, using the terminology of fractal geometry, what is 'the inner cut-off', and why must it exist?

A close examination of Sections 2–6 will show that purely theoretical answers have been given to all these questions, of course, by using the heat generating volume problem, not the fluid flow resistance problem. The question is: why have the previous investigators failed? To see the answer let us rethink the heat generating volume problem by relying on the approach that has been used in the past for the bronchial tree and other fractal drawings.

There is a fundamental difference between the approach followed in the past and the approach proposed in this paper (Sections 2–6). In all the theoretical studies of 'branching' fluid networks, the network was first seen and accepted, and then it was broken down repeatedly (e.g. through 'bifurcation') beginning with the largest duct and proceeding towards smaller scales. It was this 'fracturing' point of view that made the natural fluid networks ideal examples of 'fractal' geometric constructions. The point of view exercised in the present paper is precisely the opposite: the network was constructed from the smallest elements by using optimized building blocks (e.g. through 'pairing') and proceeded toward larger scales.

In Fig. 8 we re-examine the heat generating volume problem by using the older approach favored in physiology. The volume V is filled by an existing network of high-conductivity paths. The network is seen as a sequence of n branching stages ($j = 1, 2, \ldots, n$) which proceeds in the direction of smaller scales. At each branching stage, the unknown dimensions of each path are D_j and L_j. The total number of paths of size (D_j, L_j) is also unknown and is labeled N_j. The starting (zeroth) path of the network is characterized by D_0, L_0 and $N_0 = 1$. The directions in which the smaller branches 'invade' the volume V are unknown.

The conservation of the total heat current (q_0) at each branching stage requires that

$$q_0 = N_j q_j \tag{59}$$

where q_j is the heat current through one branch of the jth order,

$$q_j = k_p D_j W \frac{\Delta T_j}{L_j} \tag{60}$$

and ΔT_j is the temperature difference across the jth branching stage. Combining equations (59) and (60) and summing up we obtain the total temperature difference sustained by the network,

$$\Delta T = \sum_{j=0}^{n} \Delta T_j = \frac{q_0}{k_p W} \sum_{j=0}^{n} \frac{L_j}{N_j D_j}. \tag{61}$$

We are interested in minimizing ΔT, or the ratio $\Delta T/q_0$, subject to the material constraint

$$A_p = \sum_{j=0}^{n} A_{pj} = \sum_{j=0}^{n} N_j L_j D_j. \tag{62}$$

The optimization problem (61) and (62) is equivalent to finding the extremum of the function

$$\Phi = \sum_{j=0}^{n} \left(\frac{L_j}{N_j D_j} + \lambda N_j L_j D_j \right) \tag{63}$$

where λ is a Lagrange multiplier. The function Φ has extrema with respect to each product of type $N_j D_j$, but not with respect to the lengths L_j. The optimized products are equal to the same constant, $(N_j D_j)_{\text{opt}} = \lambda^{-1/2}$. The constant is obtained by using the A_p constraint (62):

$$(N_j D_j)_{\text{opt}} = A_p \left(\sum_{j=0}^{n} L_j \right)^{-1}. \tag{64}$$

Equation (64) represents the most that the old optimisation approach can produce. Now, if we recognize *empirically* that each path undergoes bifurcation,

$$\frac{N_{j+1}}{N_j} = 2 \tag{65}$$

then equation (64) means that each path width must shrink by a factor of 1/2 from one stage to the next, smaller stage,

$$\left(\frac{D_{j+1}}{D_j} \right)_{\text{opt}} = \frac{1}{2}. \tag{66}$$

Equation (66) anticipates only the large-size portion ($i \geqslant 3$) of the path width magnification sequence determined theoretically and summarized in Table 1. It is important to keep in mind that, unlike the $i \geqslant 3$ results of Table 1, equation (66) is empirical because equation (65) is empirical.

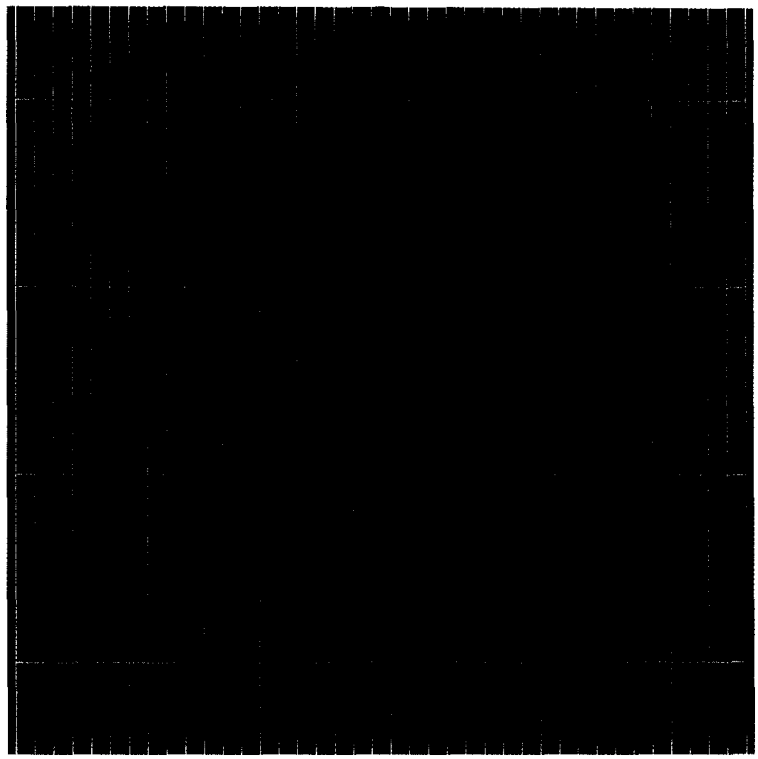


Fig. 7. The optimized sixth assembly and its complete internal structure, showing chromatically the change in the rules of assembly in going from the second assembly to the third.

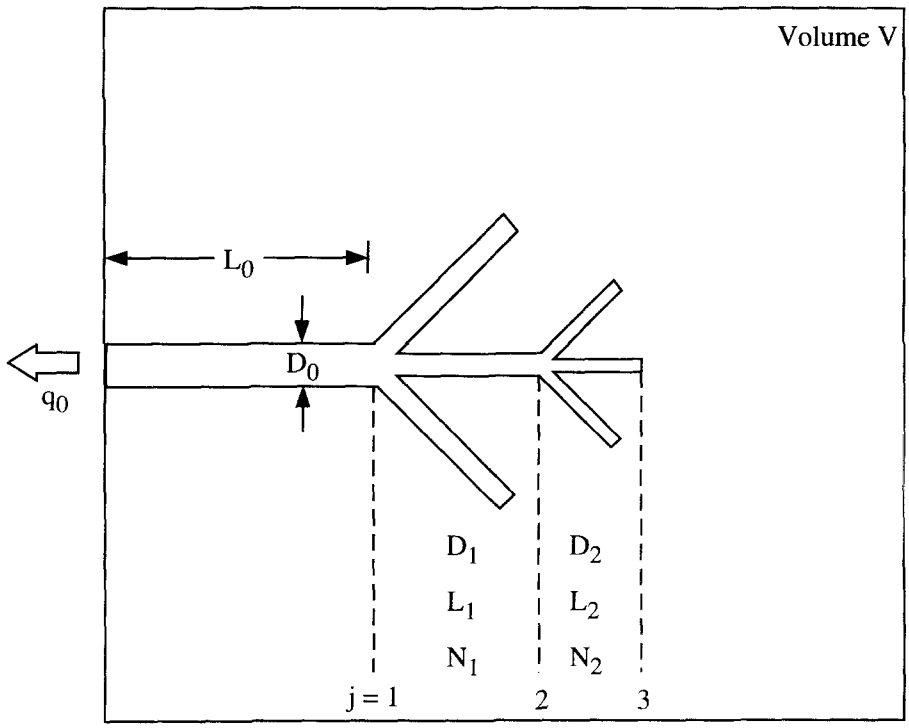


Fig. 8. An assumed (postulated) network of branching high-conductivity paths.

Equation (66) is the present (heat generating volume) equivalent of the diameter reduction factor $2^{-1/3}$ known in the fluid mechanics of the lung and circulatory system. The analysis that led to equation (66), i.e. the deconstructionist approach that consisted of breaking down each path into a new generation of more numerous paths, showed very simply why in fluid networks this approach cannot advance theoretically beyond the ability to predict the tube sizes when the tube numbers are accepted empirically.

What distinguishes the present (construction) approach from the approached used in physiology is the direction of time. The network can be anticipated in a purely deterministic manner only when the optimization and organization proceeds from small building blocks toward larger building blocks. This time direction was recognized and used for the first time in the repetitive construction of a turbulent flow as a sequence of geometrically similar building blocks, which are the repeated manifestation of the buckling phenomenon [13, 14]. The analogy between the geometrically similar construction in turbulence [13, 14] and the present construction (Figs. 1–7) is complete: in a turbulent edifice the geometric similarly breaks down at the smallest scale, where the first buckling block (eddy) is preceded by a smaller block covered volumetrically by viscous diffusion (as in Fig. 1), namely, the laminar tip of any shear (e.g. ref. [13], p. 86).

It is a remarkable coincidence that this idea, the assemblies of geometrically similar blocks, came at roughly the same time ($\sim$1980) as fractal geometry, when it was certainly unrelated to work that was being done in mathematics. In view of the growth experienced by both fractal geometry and buckling flows, it is even more amazing that the two fields have not intersected until now. The predictive power of the geometric approach of the buckling theory of turbulent flow was reviewed most recently in ref. [15].

8. OPTIMAL ACCESS, TIME, LIFE AND THERMODYNAMICS

There is an even more fundamental link between the optimized conglomerate we have constructed in Table 1 and Fig. 6, and the natural occurrences of similar structures, both living and not living. The optimization principle that rules the design of each building block, and the manner in which an optimal number of such blocks are assembled into the next (larger) block is the optimization of access, or the minimization of resistance. In the volumetric heat generation problem, the objective has been to minimize the temperature difference between each element and the next assembly. The resulting structure is such that the total heat transfer rate (q) is driven to the exterior by the smallest temperature difference ($\Delta T_{\min}$). The entropy generation rate associated with heat transfer through the volume V is proportional to the product $q\Delta T_{\min}$ (e.g. ref. [13], p. 99): since q is fixed by the electrical function and design of the circuits installed in V, the minimized excess temperature means that the entropy generation rate is minimum as well.

The optimal structure (Table 1, Fig. 6) represents

the easiest way in which the generated heat can get out of the system. This structure is a multiparameter thermodynamic optimum, which was derived subject to several global constraints: fixed overall volume, fixed total amount of high-conductivity material and specified function (q). It is as if the system (V) was forced to organize itself internally in order to cope with the constraints imposed from the outside. To cope means to survive, i.e. to persist in time. For if a better cooling path (or assembly of paths) can be designed, then that design will be favored, and will outlive the design developed in this paper.

This thermodynamic connection is just one example of the general observation [13, 16] that throughout the field and technologies of heat transfer augmentation, designs evolve in the direction of less resistance to both heat and fluid flow, or less entropy generation. It is this 'evolution', the history of new (better) designs that replace older ones, that measures the passing of time and brings time into thermodynamics. See also the history of heat engines and power plants, for example, Figs. 2.1 and 8.1 in ref. [17]: these evolve to persist in time because they do so along with us. They are an integral part, the modern part, of our own development.

Another example is provided by the fluid network [9, 10], where the constant diameter reduction factor corresponds to the minimum flow resistance. Furthermore, we know that the entropy generation rate due to fluid flow is proportional to the total flow rate ($\dot{m}$) and the total pressure drop (ΔP) experienced by the flow (e.g. ref. [13], p. 37). In the lungs and the circulatory system, the flow rates are fixed by the metabolic rate. The minimization of the flow resistance, making the path easier to travel, leads again to entropy generation minimization.

Similar observations hold for the networks seen in river basins and deltas. The overall size (a surface this time) is fixed and the total flow rate is dictated by rainfall. On the ground, the water has the option to seep (as a sheet) over or through the wet soil, in the direction of the lowest point of the collection area. Instead, the river basin develops its own network of 'high-conductivity paths', rivers, tributaries, creeks and rivulets, to collect the rain water as seepage only at the smallest (elemental) scale.

The river basin problem turned inside out is the river delta. The water flow rate is constrained, enters a finite geographical area through one point, distributes itself through a network of low-resistance channels and finally seeps, with high flow resistance, into the ground. This final stage occurs over areas (stagnant pools) defined by the length of the smallest channels. Superimposed on this 'point to finite area flow', there is the related problem of distributing (again, with minimum resistance) a fraction of the original stream to the coastal portion of the perimeter of the delta area.

Lightning is a phenomenon very similar to the river basin, except that it is more obviously a one-shot process (more obviously than the river or the torrent). The electric charge that is initially distributed volumetrically in a finite air volume is conducted to one point on the volume boundary (e.g. the roof of a house), first, by electric diffusion from air to the smallest branches of the lightning 'tree' and then by unidirectional electric conduction along ionized paths of much higher conductivity, which at high levels of assembly become large enough to be visible.

The commonality of these phenomena is much too obvious to be overlooked. It was noted in the past [10] and most recently (empirically) in fractal geometry, where it was visualized (simulated) based on repeated fracturing (algorithms) that had to be assumed and truncated. The origin of such algorithms was left to the explanation that the broken pieces (or building blocks, from the reverse point of view of this paper) are the fruits of a process of self-optimization and self-organization [12]. The present paper places a purely deterministic approach behind the work 'self', the search for the easiest path (least resistance) when global constraints (current, flow rate, size) are imposed.

If we limit the discussion to examples of living systems (lungs, circulatory systems, nervous systems, trees, roots, leaves), it is quite acceptable to end with the conclusion that such phenomena are common because they are the end result of a long running process of 'natural selection'. A lot has been written about natural selection and the impact that thermodynamic efficiency has on survival. In fact, to refer to living systems as complex power plants has become routine. The tendency of living systems to become optimized in every building block and to develop optimal associations of such building blocks has not been explained: it has been abandoned to the notion that it is imprinted in the genetic code of the organism.

If this is so, then what genetic code might be responsible for the development of equivalent structures in such nonliving systems as rivers and lightning? What genetic code is responsible for a man-made network such as Table 1? Certainly not mine, because although highly educated, neither of my parents knew heat transfer (by the way, thermodynamics was not needed in Sections 1–7). Indeed, whose genetic code is responsible for the societal 'trees' that connect us, for all the electronic circuits, telephone lines, air lines, assembly lines, alleys, streets, highways and elevator shafts in multistory buildings?

There is no difference between the living and the nonliving when it comes to the behavior when faced with the opportunity to find a more direct route subject to global constraints, for example, the opportunity of getting from here to there in an easier (faster) manner. If living systems can be viewed as engines in competition for better thermodynamic performance, then physical systems too can be viewed as living entities in competition for survival. This analogy is purely empirical: we have a very large body of case-by-case observations indicating that certain designs (living and nonliving) evolve and persist in time, while

others do not. Now we know the particular feature (optimal access, least resistance, entropy generation minimization or EGM [13, 16]) that sets each surviving design apart, but we have no theoretical basis on which to expect that the design that persists in time is the one that has this particular feature. This body of empirical evidence forms the basis for a new law of nature that can be summarized as follows:

> 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 (global) currents that flow through it.

This 'fourth law' brings life and time explicitly into thermodynamics and creates a new bridge between physics and biology.

9. FUNDAMENTAL OPTIMA

The present theory was the result of an accident, the coincidence that I reviewed the explosive growth of the field of entropy generation minimization (or EGM) [16] exactly when I was reviewing my geometric optimization method for cooling electronic systems [5]. I was struck less by the power of these methods than by the obvious fact that they have been tried only on a few classes of engineering systems, i.e. on systems with purpose. While preparing my paper for the symposium honoring Chancellor Chang-Lin Tien [18], I saw that fundamental optima deserve much more attention from all of us, engineers and nonengineers. It is these compact and beautiful results, these jewels, that speak of the 'purpose' in what we do and what interests us. They speak of us. This meant that one direction in which the methods of thermodynamic and geometric optimization deserved to be extended was that of living systems.

Another coincidence, a very useful one, is that prior to the theoretical line taken in this paper I was unaware of the enormous amount of work done on tree networks [9, 10], fractal geometry [11] and self-optimization [12]. This coincidence was useful because otherwise I too might have gotten the direction of time wrong. To repeat the conclusion of Section 7, what makes the present optimization and organization method deterministic is the time arrow from small to large. Confluence yes, branching no. Pairing yes, bifurcation no. Construction yes, fracturing no. If 'fractal' is an appropriate Latin-based word† for breaking things [11], i.e. for the opposite of the direction in which natural systems evolve, then the appropriate word for the geometry and evolution of optimized and organized finite-size systems is *constructal*.‡

Because of the potentially deceiving 'electronics cooling' title of this paper, we need to be clear about what its contents represent. This line of inquiry is less about cooling techniques than it is about method, a completely deterministic theory of optimized and organized systems that evolve in time. It is about predicting natural patterns that, until now, could not be predicted. This method is a purely deterministic step that had been ruled impossible by contemporary physics and mathematics.

While writing these closing comments I discovered that the fundamental problem formulated in Section 1 is related to what in mathematics is known as the classical Steiner problem [19]. According to Bern and Graham's recent review [20], 'the solution to this problem has eluded the fastest computers and the sharpest mathematical minds' and its solutions 'defy analysis'. The present paper provides a purely theoretical and analytical alternative worthy of attention. The challenge was not to connect 'many points' to a single point—computers will accomplish this task some day. The challenge was to connect an *infinite* number of points (a finite area, or volume) to a single point. To accomplish this we used *two regimes*, volumetric diffusion (k_0) and channeled conduction (k_p), by placing the *slower* regime (k_0) at the volumetric level.

Acknowledgements—This work was supported by the National Science Foundation. Figures 1 8 and 10 were drawn by Kathy Vickers. Figure 9 was drawn by Gustavo A. Ledezma.

REFERENCES

1. Peterson, G. P. and Ortega, A., Thermal control of electronic equipment and devices. *Advances in Heat Transfer*, 1990, **20**, 181–314.
2. Aung, W., ed., *Cooling Technology for Electronic Equipment*. Hemisphere, New York, 1988.
3. Li, W., Kakac, S., Hatay, F. F. and Oskay, R., Experimental study of unsteady forced convection in a duct with and without arrays of block-like electronic components. *Wärme- und Stoffübertragung*, 1993, **28**, 69–79.
4. Kim, S. H. and Anand, N. K., Laminar developing flow and heat transfer between a series of parallel plates with surface mounted discrete heat sources. *International Journal of Heat and Mass Transfer*, 1994, **37**, 2231–2244.
5. Bejan, A., Geometric optimization of cooling techniques. In *Air Cooling Technology for Electronic Equipment*, eds S. J. Kim and S. W. Lee. CRC Press, Boca Raton, FL, 1996, Chap. 1.
6. Guglielmini, G., Nannei, E. and Tanda, G., Natural convection and radiation heat transfer from staggered vertical fins. *International Journal of Heat and Mass Transfer*, 1987, **30**, 1941–1948.
7. Stanescu, G., Fowler, A. J. and Bejan, A., The optimal spacing of cylinders in free-stream cross-flow forced convection. *International Journal of Heat and Mass Transfer*, 1996, **39**, 311–317.
8. Bejan, A., *Heat Transfer*. Wiley, New York, 1993.
9. Cohn, D. L., Optimal systems: I. The vascular system. *Bulletin of Mathematical Biophysics*, 1954, **16**, 59–74.
10. Thompson, D'A. W., *On Growth and Form*. Cambridge University Press, Cambridge, 1942.
11. Mandelbrot, B. B., *The Fractal Geometry of Nature*. W. H. Freeman, New York, 1982.

† From the Latin verb *frangere* (to break), which survives unchanged in both Italian and Romanian.

‡ From the Latin verb *construĕre* (to build), which survives as *construire* in both Italian and Romanian.

12. Prigogine, I., *From Being to Becoming*. W. H. Freeman, San Francisco, 1980.
13. Bejan, A., *Entropy Generation through Heat and Fluid Flow*. Wiley, New York, 1982, Chap. 4.
14. Bejan, A., *Convection Heat Transfer*. Wiley, New York, 1984, p. 287.
15. Bejan, A., *Convection Heat Transfer*. 2nd edn, Wiley, New York, 1995, Chaps 6–9.
16. Bejan, A., *Entropy Generation Minimization*. CRC Press, Boca Raton, FL, 1996.
17. Bejan, A., *Advanced Engineering Thermodynamics*. Wiley, New York, 1988.
18. Bejan, A., Fundamental optima in thermal science. Proceedings of the Symposium on Thermal Science and Engineering in Honor of Chancellor Chang-Lin Tien. Berkeley, CA, November 14, 1995; this paper will appear in the *International Journal of Mechanical Engineering Education*, 1997, **25**.
19. Courant, R. and Robbins, H., *What is Mathematics?* Oxford University Press, London, 1941.
20. Bern, M. W. and Graham, R. L., The shortest-network problem. *Scientific American*, 1989, January, 84–89.