

## Effect on scaling of heat removal requirements in three-dimensional systems

HALDUN M. OZAKTAS†, HAKAN OKSUZOGLU‡,  
R. F. W. PEASE§ and JOSEPH W. GOODMAN§

A first-order discussion of convective heat removal from a hypothetical 3-dimensional computing system is presented. A textbook treatment indicates that our heat removal capability can be characterized by a quantity  $Q$ , the amount of power we can remove per unit cross-section. Thus the minimum length of the system is proportional to the square root of the total power dissipated. We predict  $Q \sim 10^4 \text{ W cm}^{-2}$  assuming an applied pressure difference of 1 atm, a maximum temperature rise of 100 K and the use of water as the coolant fluid.

### 1. Introduction

It is desirable to pack the elements of a computing system as closely as possible so as to minimize signal delay. Wiring and heat removal requirements both set bounds to the maximum achievable density (Pence and Krusius 1987).

In two dimensions, wiring requirements dictate that the linear extent of a system of  $N$  elements (for instance gates) grow as  $\propto N^q$  where  $1/2 \leq q \leq 1$  is a measure of the connectivity of the system (Feuer 1982, Bakoglu 1990, Ozaktas and Goodman 1991). Most logic circuits are known to exhibit  $q \sim 0.6$  whereas, for instance, neural networks may exhibit  $q \sim 1$ . If we assume the power dissipation per element is constant and the amount of power we can remove per unit area is specified, heat removal requirements dictate that the linear extent of the system grow as  $\propto N^{1/2}$ . Thus, unless  $q = 1/2$ , wiring requirements will surpass heat removal requirements with increasing  $N$ . Larger values of  $q$  enable greater parallelism in data transfer, but result in larger layout area and delays, so that a detailed analysis is necessary to determine the optimal value of  $q$  resulting in a system with optimal properties (Bakoglu 1990).

In three dimensions, wiring requirements dictate that the linear extent of the system grow as  $\propto N^{q/2}$  where  $2/3 \leq q \leq 1$  (Ozaktas and Goodman 1991). Here we show that heat removal again dictates a growth rate of  $\propto N^{1/2}$ . Thus, for large  $N$ , the choice of  $q$  has little if any effect on the resulting system size. Since smaller values of  $q$  will not reduce system size and delays, we might as well employ high values of  $q$ , increasing parallelism and connectivity. Thus it will be more beneficial to employ highly interconnected approaches in large scale 3-dimensional computing systems.

The above arguments can be modified if the power dissipation per element is not constant. However, it should be evident that 3-dimensional systems exhibit a

---

Received 10 February 1992; revision accepted 19 May 1992.

†Department of Electrical Engineering, Bilkent University, 06533 Bilkent, Ankara, Turkey.

‡Department of Aeronautical and Astronautical Engineering, Stanford University, Stanford, CA 94305, U.S.A.

§Department of Electrical Engineering, Stanford University, Stanford, CA 94305, U.S.A.

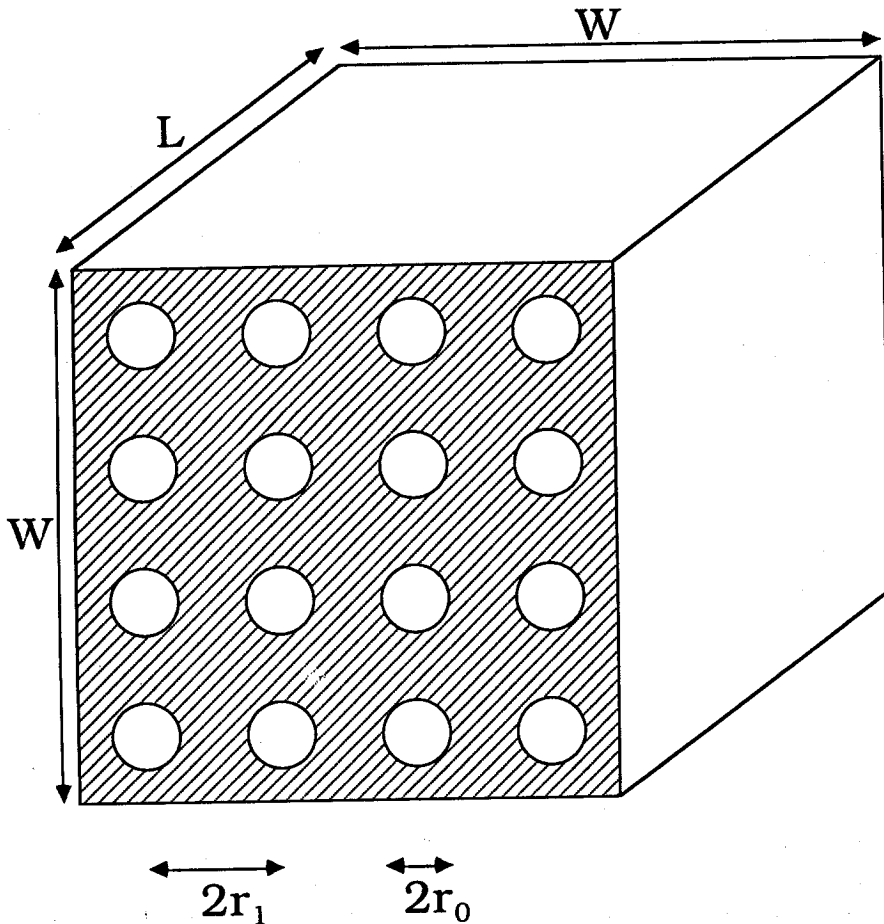
tendency to be heat removal limited, so that highly interconnected approaches can be employed with little effect on system size and delays.

## 2. Analysis

Let  $\mathcal{P}$  denote the total power dissipated in a square prism of volume  $W \times W \times L$  with  $L \leq W$  (see the Figure). We keep  $L/W$  fixed throughout our discussion. The total power  $\mathcal{P}$  is dissipated uniformly in the solid volume between the tubes. A coolant fluid flows through  $W^2/4r_1^2$  tubes of diameter  $2r_0$  and axial separation  $2r_1$ . So as to ensure that the volume of the tubes does not exceed a certain fraction of the total volume  $W^2L$ , we will require that  $\eta = r_0/r_1 \leq \eta_{\max} < 1$ .

Our purpose is to determine the minimum value of  $W$  at which we can successfully remove the dissipated power  $\mathcal{P}$ . To illustrate the general principles, several textbook (Holman 1981) assumptions will be employed:

- (a) Laminar flow of an incompressible fluid,



Heat removal from a square prism via fluid convection.

- (b) Temperature gradients do not affect fluid flow and the pressure is uniform at any cross-section, decreasing linearly in the axial direction,
- (c) The flow is fully developed, i.e. invariant in the axial direction (Holman 1981),
- (d) Ignore edge effects,
- (e) Based on similar arguments as in (Keyes 1987), heat conduction is significant only in the transverse plane (thus, the heat flux entering from the walls of the tube is uniform along the axial direction, resulting in a constant temperature gradient in the axial direction),
- (f) Material parameters are constant.

Our analysis is modelled on that of Tuckerman and Pease (1981) and Tuckerman (1984). The two dominating components of thermal resistance are those due to the heating of the fluid ( $\Theta_{\text{cal}} = \Delta T_{\text{cal}}/\mathcal{P}$ ) and due to the transfer of heat at the solid-fluid interface ( $\Theta_{\text{conv}} = \Delta T_{\text{conv}}/\mathcal{P}$ ).  $\Theta_{\text{cal}}$  may be expressed as  $1/\rho C_p f$  where  $\rho C_p$  denotes the volumetric heat capacity of the fluid and  $f$  is the total fluid flow rate. The flow rate through a single tube is given by  $\pi r_0^4 \Delta P / 8 \mu L$  where  $\Delta P$  is the applied pressure difference and  $\mu$  is the viscosity of the fluid (Holman 1981). Multiplying this with the number of tubes to find  $f$ , we obtain

$$\Theta_{\text{cal}} = \frac{32}{\pi} \frac{1}{\eta^4} \frac{\mu}{\rho C_p \Delta P} \frac{L}{r_1^2 W^2} \quad (1)$$

$\Theta_{\text{conv}}$  may be expressed as  $1/hs$  where  $s$  denotes the total internal surface area of the tubes and  $h$  is the heat transfer coefficient.  $h$  is given by  $\text{Nu} \kappa / D$  where  $\kappa$  is the conductivity of the fluid and  $D$  is the hydraulic diameter, equal to its geometrical diameter for a tube (Holman 1981). The Nusselt number  $\text{Nu}$  is chosen as its steady state value of  $48/11$ , based on similar arguments by Tuckerman and Pease (1981) and Tuckerman (1984). Thus:

$$\Theta_{\text{conv}} = \frac{11}{12\pi} \frac{1}{\kappa} \frac{r_1^2}{L W^2} \quad (2)$$

Note that it is optimal to set  $\eta = \eta_{\text{max}}$ . Then the value of  $r_1$  minimizing  $\Theta = \Theta_{\text{cal}} + \Theta_{\text{conv}}$  and the resulting thermal resistance is found to be

$$r_0^4 = (\eta_{\text{max}} r_1)^4 = 34.9 \frac{\mu \kappa}{\rho C_p \Delta P} L^2 \quad (3)$$

$$\Theta = 2\Theta_{\text{cal}} = \frac{3.45}{\eta_{\text{max}}^2} \left( \frac{\mu}{\rho C_p \kappa \Delta P} \right)^{1/2} \frac{1}{W^2} \quad (4)$$

Note that the optimum value of  $r_1$  depends only on  $L$  whereas the minimum value of  $\Theta$  depends only on  $W$ . The resulting temperature rise is  $\Delta T = \Delta T_{\text{cal}} + \Delta T_{\text{conv}} = 2\Delta T_{\text{cal}} = \Theta \mathcal{P}$ . We can now define the more intrinsic quantity  $Q$  by the relation  $QW^2 = \mathcal{P}$ , so that  $Q = \Delta T / \Theta W^2$ . If the maximum allowed temperature rise is specified, we may express  $Q$  as

$$Q = 0.29 \eta_{\max}^2 \left( \frac{\rho C_s \kappa \Delta P}{\mu} \right)^{1/2} \Delta T \quad (5)$$

Once  $W = (\mathcal{P}/Q)^{1/2}$  is calculated, we can go back to (3) to calculate the optimum value of  $r_0$  for a given  $\mathcal{P}$ .

Choosing  $r_0^2 \propto L$  keeps the hydraulic resistance of the tubes and the mean flow velocity  $v$  through the tubes constant. It can be easily shown that  $v = r_0^2 \Delta P / 8 \mu L = 0.74 (\kappa \Delta P / \rho C_s \mu)^{1/2}$ , independent of  $W$  and  $\mathcal{P}$ . Likewise the mean flow velocity through the system  $v_s = (\pi r_0^2 / 4 r_1^2) v$  is given by  $0.58 \eta_{\max}^2 (\kappa \Delta P / \rho C_s \mu)^{1/2}$ , in terms of which we may write the intuitively appealing  $Q = \rho C_s v_s \Delta T_{\text{cal}} = \rho C_s v_s \Delta T / 2$ , consistent with (5).

We finally calculate the viscous power dissipation  $\mathcal{P}_p = \Delta P f$  associated with the fluid flow. Using the previously derived expression for  $f$  and expressing everything in terms of  $\mathcal{P}$ , we obtain  $\mathcal{P}_p = (2 \Delta P / \rho C_s \Delta T) \mathcal{P}$ , an exceedingly simple result.  $\mathcal{P}_p$  is proportional to the power dissipated by the devices. We will later show that for typical values,  $\mathcal{P}_p \ll \mathcal{P}$ .

We have ignored the effects of conduction in the solid medium in which the circuits are embedded. There will be no heat transfer through the boundaries of a cell  $2r_1 \times 2r_1 \times L$  enclosing each tube (except near the edges of our system). If we pretend this cell is a cylinder of diameter  $2r_1$ , conduction in the solid may be accounted for by replacing  $1/\kappa \rightarrow 1/\kappa + 12\pi\lambda(\eta)/11\kappa_s$ , where  $\kappa_s$  is the conductivity of the solid and  $\lambda(\eta) = (\ln(1/\eta^2) - (1 - \eta^2))/4(1 - \eta^2)$ . Again the largest possible value of  $\eta$  is preferred, for instance,  $\lambda(0.5) = 0.21$ . Thus the effects of solid conduction may be ignored with little error if  $\kappa_s > 10\kappa$  or so. (The value of  $\kappa_s$  should be taken as an appropriate average value for the particular materials and structures—including wiring—involved.)

We also assumed uniform power dissipation throughout the solid. If instead the power  $\mathcal{P}_d$  associated with each device is dissipated within a small radius of  $r_d$ , an additional temperature rise of  $\sim 3\mathcal{P}_d / 8\pi\kappa_s r_d$  would be observed. If  $\mathcal{P}_d = 1 \text{ mW}$ ,  $r_d \geq 0.1 \mu\text{m}$  and  $\kappa_s \sim 150 \text{ Wm}^{-1} \text{ K}^{-1}$ , we find that this temperature rise does not exceed  $10^\circ$  and is usually acceptable.

### 3. Numerical example

Assume  $L/W = 1$ ,  $\rho = 10^3 \text{ kg m}^{-3}$ ,  $C_s = 5 \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1}$ ,  $\kappa = 0.5 \text{ Wm}^{-1} \text{ K}^{-1}$ ,  $\mu = 10^{-3} \text{ kg m}^{-1} \text{ s}^{-1}$  (corresponding to water),  $\Delta P = 10^5 \text{ kg m}^{-1} \text{ s}^{-2}$  and  $\dagger \Delta T = 100 \text{ K}$ . We take  $\eta_{\max} = 0.5$  so that the volume occupied by the pipes is less than 25% of the total system volume. Then:

$$r_0^2 = 1.87 \times 10^{-7} L = 1.78 \times 10^{-11} \mathcal{P}^{1/2} \quad (6)$$

$$\Theta = 8.73 \times 10^{-7} \frac{1}{W^2} = 96.0 \frac{1}{\mathcal{P}} \quad (7)$$

$$Q = 1.1 \times 10^8 \quad (8)$$

$$v = 2.34 \quad (9)$$

$$\mathcal{P}_p = 4 \times 10^{-4} \mathcal{P} \ll \mathcal{P} \quad (10)$$

$\dagger 100 \text{ K}$  temperature rise has been picked for arithmetic convenience. In practice, a smaller value can be employed to avoid phase transitions, which are not considered in this paper.

where everything is in SI units. 10 kW can be removed per square centimetre. For a total power of 1 kW we find  $W = 3$  mm,  $r_0 = 24$   $\mu\text{m}$  and  $\Theta = 0.096$   $\text{K W}^{-1}$ . For 1 MW, we find  $W = 95$  mm,  $r_0 = 133$   $\mu\text{m}$  and  $\Theta = 9.6 \times 10^{-5}$   $\text{K W}^{-1}$ . A megawatt can be removed from a litre, quite larger than what was previously thought possible. We also note that the viscous power dissipation is negligible in comparison to the device power dissipation.

Let us check two major assumptions. The Reynolds number  $Re = v\rho D/\mu$  should be less than 2100 for laminar flow (Holman 1981). This leads to the condition  $\mathcal{P} < 128$  MW which would allow  $10^{11}$  circuits each dissipating 1 mW in a cube of edge length of about a metre. The fluid may be considered to be fully developed when the distance  $x$  from the entrance of the tube satisfies  $x/(D Re Pr) \geq 0.02$ , provided that the Prandtl number  $Pr = \mu C_p/\kappa > 5$  or so (Tuckerman 1984). Upon substitution we obtain  $x \geq 0.35 L$ , so that the velocity and temperature profiles are well developed over a greater portion of the distance along the tube. This is not a coincidence, but a direct outcome of the optimization procedure (Tuckerman 1984).

In 1981 Tuckerman and Pease experimentally demonstrated the removal of 790 W from a 1 cm  $\times$  1 cm surface using cooling fins about 0.04 cm in height. This corresponds to 790 W/(1 cm  $\times$  0.04 cm)  $\approx 20$   $\text{kW cm}^{-2}$  of power being removed per unit area along the direction of fluid flow, in reasonable agreement with our predictions. (The factor of two discrepancy can be traced down to the different parameter values employed.) In fact, if we imagine that we stack 25 such assemblies on top of each other, we obtain a system which essentially resembles that shown in the Figure. The only major difference is that the channels are narrow slits, instead of circular tubes. Provided their total area is always the same fraction of  $W^2$ , the use of alternate cross-sectional shapes for the channels alter the general results of this paper only by geometrical factors close to unity. Whereas a general proof seems difficult, it is possible to show that for narrow slits extending along the full width of our system, the value of  $Q$  is within 10% of what has been calculated for circular tubes.

#### 4. Conclusions

We have considered the problem of heat removal from a square prism of volume  $W \times W \times L$  in which a total power  $\mathcal{P}$  is uniformly dissipated. We showed that choosing the number and cross-sectional area of the tubes proportional to their length was the optimal solution. With this choice, we found that a heat-removal-imposed lower bound to the system linear extent  $W$  can be written as  $W^2 \geq \mathcal{P}/Q$  where  $Q$  is a function of the material parameters of the coolant fluid, the applied pressure difference  $\Delta P$  and the maximum allowed temperature rise  $\Delta T$ . Thus  $Q$  is conveniently interpreted as the maximum amount of power we can remove per unit cross-section, in analogy with the two-dimensional case where it is customary to specify the maximum amount of power we can remove per unit area (Keyes 1987). With  $\Delta P = 1$  atm,  $\Delta T = 100$  K and assuming water is the coolant fluid, we have estimated that  $Q \sim 10$   $\text{kW cm}^{-2}$ .

#### REFERENCES

- BAKOGLU, H. B., 1990, *Circuits, Interconnections and Packaging for VLSI* (Reading, Mass: Addison-Wesley).
- FEUER, M., 1982, Connectivity of random logic. *I.E.E.E. Transactions on Computers*, **31**, 29-33.

- HOLMAN, J. P., 1981, *Heat Transfer*, fifth edition (New York: McGraw-Hill).
- KEYES, R. W., 1987, *The Physics of VLSI Systems* (Reading, Mass: Addison-Wesley).
- OZAKTAS, H. M., and GOODMAN, J. W., 1991, The limitations of interconnections in providing communication between an array of points. *Frontiers of Computing Systems Research*, edited by S. K. Tewksbury, Vol. 2 (New York: Plenum), pp. 61-124.
- PENCE, W. E., and KRUSIUS, J. P., 1987, The fundamental limits for electronic packaging and systems. *I.E.E.E. Transactions on Components, Hybrids and Manufacturing Technology*, 10, 176-183.
- TUCKERMAN, D. B., 1984, Heat-transfer microstructures for integrated circuits. Ph.D. thesis, Stanford University, Stanford, California.
- TUCKERMAN, D. B., and PEASE, R. F. W., 1981, High performance heat sinking for VLSI. *I.E.E.E. Electron Device Letters*, 2, 126-129.