

The optimal electromagnetic carrier frequency balancing structural and metrical information densities with respect to heat removal requirements

Haldun M. Ozaktas

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

and

Joseph W. Goodman

Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA

Received 3 February 1992; revised manuscript received 5 June 1992

The use of higher electromagnetic carrier frequencies for communication in a computing environment results in both increased spatial information density and larger available modulation bandwidth. However, assuming that the communication energies are dissipated, the heat that must be removed from unit volume per unit time increases quickly with higher frequencies, resulting in a maximum useful frequency based on our limited ability to remove heat. We show that this frequency is relatively insensitive to system specific parameters and estimate its order of magnitude to lie near the infrared and visible bands of the spectrum.

1. Introduction

Of the four basic forces in nature, only the electromagnetic force is effective on a scale comparable to biological organisms. The human visual and nervous systems are essentially based on electromagnetic interactions. Most man made computing systems rely on electromagnetic phenomena for both regenerative nonlinearities and communication among their elements. The nonlinear operations may be based on electronic interaction.

Electromagnetic wave propagation is a most basic means of information transmission. In this work we will consider the use of modulated electro-magnetic carrier waves to establish communication between the elements of a computing system. We assume that the signals are guided to their destinations with dielectric media only (including lenses, holograms, waveguides etc.), the use of conductors is excluded from consideration.

The elements of our computing system may be relatively simple switching devices or relatively com-

plex processing elements. In any event we will assume them to be very small in size. We would like our overall system to be as compact as possible. Being able to handle vast amounts of information in a small volume is not only a merit in itself, as exemplified by the human eye, but also results in smaller speed-of-light limited communication delays between distant elements of our system. Spatial information density and heat removal are two major physical considerations which will limit how densely we can pack the elements of our computing system. These considerations are intimately tied together through the carrier frequency. Increasing the carrier frequency improves spatial information density, however also increases the amount of heat we must remove per unit cross section and time. Since our heat removal ability is limited, there exists an optimal carrier frequency resulting in smallest possible system size and delay.

In the next section we discuss spatial information density and heat removal. In the third section we introduce a simple model computing system and show

that the optimal carrier wavelength is insensitive to system specific parameters (such as the number of elements, number of connections per element and interconnection topology) and lies near the infrared and visible bands of the spectrum. Finally, we briefly mention the importance of this band of frequencies in other areas of physical science.

Owing to the general nature of this study, our analysis is necessarily approximate. We have also preferred to leave out certain geometrical factors of the order of unity for simplicity and generality.

2. Physical mechanisms

In this section we discuss the two major physical mechanisms which will limit how closely we can pack the elements of our system.

2.1. Spatial information density

We will assume that information transfer takes place along each independent spatial channel in the form of binary digital pulses impressed on a sinusoidal electromagnetic carrier of wavelength λ . It is a well known property of wave propagation that a cross sectional area of at least $\sim \lambda^2$ must be allocated per each independent spatial channel. A more general formulation states that the total volume allocated for communication in a system with total interconnection length l_{total} must at least be [1]

$$\sim \lambda^2 l_{\text{total}}. \quad (1)$$

Thus we see that the spatial (or *structural* [2]) information density can be increased by reducing the carrier wavelength λ , or in other words by increasing the frequency f .

2.2. Energy dissipation and heat removal

The energy of a single photon of electromagnetic radiation is given by $hf = hc/\lambda$ where h is Planck's constant and c is the speed of light. With increasing frequencies, the increasing energy of a single photon will require larger energies to maintain reliable communication, leading to a decrease in the so called *metrical* [3] information density. In general, based on statistical considerations $\vartheta \gg 1$ photons will be re-

quired per transmitted bit. The human eye, under optimum conditions, can detect as little as 100 photons per second [4]. If it is assumed that the eye can still differentiate events spaced about 100 ms apart under these conditions, this corresponds to 10 photons per bit. Properly designed shot-noise limited systems may require ~ 100 – 1000 photons, depending on the error rate we are willing to tolerate. We will follow Smith [5] in taking $\vartheta = 10^3$. The actual communication energies involved in practical systems may be quite larger than $E = \vartheta hc/\lambda$ due to various forms of overhead.

The energy associated with each switching or regeneration event can usually be decreased with decreasing temperatures [6,7]. Thus for sufficiently low temperatures and operating voltages, the communication energies will dominate electronic switching and regeneration energies in the elements, so that the latter can be ignored. We are assuming that these energies are irreversibly dissipated.

Two remarks are appropriate at this point. First, we note that in principle, the energy E associated with each transmitted bit may be reduced down to $E \approx hB$, where B is the bit transmission rate [8,9]. Second, we are assuming that the energy associated with each transmitted bit of information is dissipated. It has been argued by many authors that this is not a fundamental necessity [10–13]. Thus, our assumptions do not correspond to ultimate physical limitations, but rather to an idealization of existing dissipative short-noise limited communication systems.

The dissipated energy must be removed from the system. Whatever the modality (conduction, convection or radiation), heat transfer can only take place *through* a surface (flux conservation). This is most easily visualized by considering the flow of a fluid coolant through our system, which is assumed to occupy a cubic volume (fig. 1). Thus, as discussed at length in ref. [14], our ability to remove heat from a three-dimensional system (which for the purpose of this paper is assumed to be cubic in shape) can be characterized by a quantity Q , the amount of power we can remove per unit cross sectional area of the cubic system. (For instance, we would be able to remove QL^2 of power from a cube with edge length L .) The value of Q is estimated as [14]

$$Q \approx 0.29 (\rho C_s \kappa \Delta P / \mu)^{1/2} \Delta T, \quad (2)$$

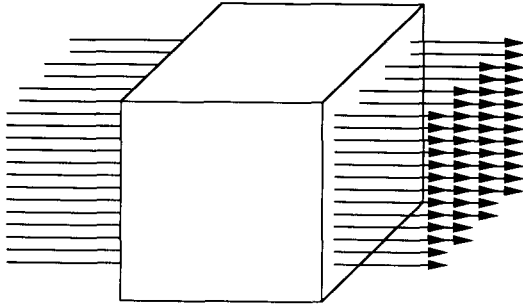


Fig. 1. Heat removal via fluid convection. The maximum amount of heat we may remove per unit time is proportional to the cross sectional area of the system.

where ρ , C_s , κ and μ denote the density, heat capacity, thermal conductivity and viscosity of the fluid coolant, respectively. ΔT is the maximum allowed temperature rise and ΔP is the applied hydraulic difference across the channels of fluid flow. An upper limit $\Delta P \leq \rho C_s \Delta T / 2$ is set by the necessity to limit viscous power dissipation. Thus, the maximum value of Q is found as

$$Q \approx 0.2 (\rho^2 C_s^2 \kappa / \mu)^{1/2} \Delta T^{3/2}. \quad (3)$$

According to ref. [15], water fares best in terms of maximizing the above expression. With $\Delta T = 100$ K, we find $Q \sim 10^6$ W/cm², a rather optimistic estimate several orders of magnitude beyond what is currently achieved.

Ultimately, regardless of the specifics of our heat removal system, how much we can increase Q is limited by material parameters which are in turn related to atomic constants. This, even allowing for advances in materials; we do not expect that the value of Q can be further improved beyond a few more orders of magnitude, if we are to construct our processing systems from solid state materials under 'earthly' conditions. Whereas such a conclusion is too imprecise to have any engineering value, it will suffice for our purpose.

3. Model computing system

Our model computing system is to be constructed by establishing a prespecified pattern of connections among an array of $N^{1/3} \times N^{1/3} \times N^{1/3}$ elements laid

out on a regular cubic grid with as yet unspecified lattice constant d . The system is confined in a cubic box of volume Nd^3 . Let k denote the average number of connections per element and \bar{r} denote the average length of the connections in units of grid spacing. Ignoring numerical factors of the order of unity, we have $1 \leq \bar{r} \leq N^{1/3}$. The total connection length in real units is $l_{\text{total}} = kN\bar{r}d$, since there are a total of kN connections. Using eq. (1), we find that a volume of at least $\sim \lambda^2 kN\bar{r}d$ is necessary to establish these connections. We must choose d so that the system volume Nd^3 exceeds this minimum required volume plus the volume of the elements. Assuming the volume occupied by the elements to be negligible, we find that d must satisfy

$$d^2 \geq \lambda^2 k \bar{r}. \quad (4)$$

The total amount of power dissipated is $kNEB$ where $E = \vartheta \hbar c / \lambda$ and B is the rate at which bits of data are being emitted into each connection. We again refer to fig. 1 and assume that we are able to remove Q of power through unit cross sectional area. Since the cross sectional area of our system is $N^{2/3}d^2$, the total amount of power we can remove is $QN^{2/3}d^2$. Requiring that this be greater than the total power dissipated, we find

$$d^2 \geq \frac{kN^{1/3} \vartheta \hbar c B}{\lambda Q}. \quad (5)$$

We would like to minimize d both for sake of maintaining a compact system and – when it is a limiting factor – minimizing speed-of-light limited communication delays. The delay across the extent of the system is given by

$$\tau = N^{1/3} d / c. \quad (6)$$

Thus, given N and B , one may choose λ so as to minimize d and τ . Equating the right hand sides of eqs. (4) and (5) and solving for λ we obtain, with $\vartheta = 10^3$ and $Q = 100$ W/cm²,

$$\lambda = (ch/Q)^{1/3} \vartheta^{1/3} B^{1/3} \{1, N^{1/9}\} \\ \sim 0.5 B^{1/3} \{1, N^{1/9}\} \mu\text{m}, \quad (7)$$

where B is in Gbit/s, and the notation $\{1, N^{1/9}\}$ means that any value between the two extremes is possible. Most present day computers operate at rates of 10–1000 Mbit/s. Switches operating at ~ 100

Gbit/s have been built. How large can N be? The connection machine [16] has $\sim 10^5$ elements. Some computers may have 10^8 transistors or more. The human brain has 10^{11} neurons [17]. Although both B and N may vary over a large range, the optimal value of λ is quite insensitive to this variation.

In most cases, the rate B at which pulses of information are emitted into each connection will be related to the cycle time of the system, determined by the worst case signal delay τ , as given by ^{#1} eq. (6). This approach was taken by Keyes [6]. Thus let us write $B = \beta/\tau$ where β is a constant. Most authors take $\beta = 1$ without discussion, however there is no reason why β should not be larger (i.e. we may allow pipelining). This time, using eqs. (4), (5) with equality, eqs. (6), (7) and $B = \beta/\tau$ we might solve for the optimal value of λ as

$$\lambda = (c^2 h / Q)^{1/4} \vartheta^{1/4} \frac{\beta^{1/4}}{k^{1/8} \{1, N^{1/8}\}} \\ \sim 15 \frac{\beta^{1/4}}{k^{1/8} \{1, N^{1/8}\}} \mu\text{m}. \quad (8)$$

The optimal wavelength is very insensitive to both how it is calculated and to the various parameters. Equation (8) is plotted in fig. 2 for $\bar{r} = 1$ and $\bar{r} = N^{1/3}$ with Q as a parameter.

4. Analysis

In this article, we considered the use of a full three-dimensional layout which may not always be possible to realize. Our analysis may be repeated for a fully two-dimensional planar layout, for which the optimal wavelength is found to be of the order of $\sim 0.1 \mu\text{m}$. The optimal wavelength is smaller in this case because a fully two-dimensional layout is much more restrictive in terms of providing communication. The optimal wavelength is also somewhat more strongly dependent on the system parameters in two-dimensions. For layouts lying between these two extremes, which combine realizability and flexibility in

providing communication, the optimal wavelength will lie between this value and those found above for full three-dimensional layouts.

Apart from system parameters with limited effect, the optimal carrier wavelength is given essentially by $\lambda \sim (c^2 h / Q)^{1/4}$ which is equal to a few μm for $Q = 100 \text{ W/cm}^2$. Human eyes, as well as those of many other living beings, operate at wavelengths around $0.5 \mu\text{m}$. Even bees, despite severe diffraction problems, have developed compound eyes enabling them to operate at only slightly higher frequencies [18]. There exist 'windows' at visible frequencies in the absorption spectrum of water [19] and the spectral distribution of solar radiation arriving at the earth [20,21]. Many electronic energy levels, in particular semiconductor bandgaps, correspond to visible and infrared frequencies, enabling the construction of efficient sources and detectors at these frequencies [22]. Although it would be unwarranted for us to draw any conclusions from these facts, one is tempted to think that they are not unrelated. After all, Q is intimately related to atomic constants through material parameters.

5. Conclusion

We considered the use of electromagnetic radiation for communication among the elements of a computing system. In general, increasing the frequency of radiation enables higher information densities which potentially offer smaller system size and communication delays. Even when delay is not an issue, the ability to handle vast amounts of information in a compact volume is desirable in itself. Increasing the carrier frequency f also increases the energy per transmitted bit. Thus, assuming these energies are dissipated, the amount of heat that must be removed from unit volume per unit time quickly increases. The maximum removable heat pertaining to 'earthly' conditions may be estimated based on physical considerations. After a certain frequency, we can no longer benefit from the high information and packing densities offered because we will fail to meet the heat removal requirements.

We have shown that there exists an optimum carrier frequency balancing information density and heat removal imposed bounds on element spacing result-

^{#1} Here we are implicitly assuming that the length of the longest interconnection is of the order of the linear extent of the system. We should also note that τ may not be of direct significance in some cases, such as a nearest neighbor connected system.

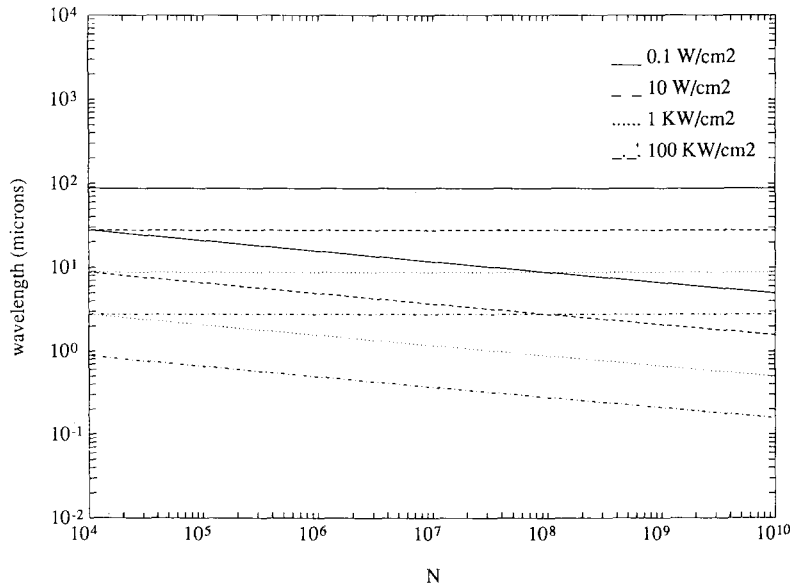


Fig. 2. The optimal carrier wavelength for full three-dimensional layouts. Equation (8) is plotted with Q as a parameter. We take $\vartheta = 10^3$ and $\beta^{1/4}/k^{1/8} = 1$. The horizontal lines correspond to $\bar{r} = 1$ and the slanted lines correspond to $\bar{r} = N^{1/3}$.

ing in smallest possible system size. This frequency is found to depend on the speed of light, which simply relates the frequency to the minimum resolvable dimension λ , on Planck's constant, which relates the frequency to the minimum resolvable energy, on our heat removal ability as quantified by Q and also on other system parameters. Since the optimal value of f is only weakly dependent on all parameters, it was possible to obtain a system independent estimate of $f \sim 10^{14}$ which corresponds to the infrared and visible bands.

Needless to say, care must be exercised in arriving at any practical conclusions from our results which have been derived for an idealized computing system limited by simple physical considerations only.

Acknowledgements

The first author acknowledges the support of the Alexander von Humboldt Foundation through a Post Doctoral Research Fellowship.

References

- [1] H.M. Ozaktas and J.W. Goodman, *J. Opt. Soc. Am. A* 7 (1990) 2100.
- [2] J.T. Winthrop, *J. Opt. Soc. Am.* 61 (1971) 15.
- [3] D. Gabor, in: *Progress in optics*, Vol. 1, ed. E. Wolf (North-Holland, Amsterdam, 1961) ch. 4.
- [4] A.E. Siegman, *Lasers* (University Science Books, Mill Valley, 1986).
- [5] P.W. Smith, *Bell Syst. Tech. J.* 61 (1982) 1975.
- [6] R.W. Keyes, *Science* 168 (1970) 796.
- [7] R.W. Keyes, *Proc. IEEE* 69 (1981) 267.
- [8] J.B. Pendry, *J. Phys. A* 16 (1983) 2161.
- [9] F.T.S. Yu, *Optics and information theory* (Wiley, New York, 1976).
- [10] C.H. Bennett and R. Landauer, *Scientific American* July (1985) p. 48–56.
- [11] R. Landauer, *Intern. J. Theor. Phys.* 21 (1982) 283.
- [12] E. Fredkin and T. Toffoli, *Intern. J. Theor. Phys.* 21 (1982) 219.
- [13] H.J. Caulfield and J. Shamir, *Appl. Optics* 28 (1989) 2184.
- [14] H.M. Ozaktas, *A physical approach to communication limits in computation*, PhD thesis, Stanford University, Stanford, 1991.
- [15] D.B. Tuckerman, *Heat-transfer microstructures for integrated circuits*. PhD thesis, Stanford University, Stanford, 1984.
- [16] W. Daniel Hillis, *The Connection Machine* (The MIT Press, Cambridge, 1985).
- [17] R.F. Thompson, *The Brain* (Freeman, New York, 1985).

- [18] R.P. Feynman, R.B. Leighton and M. Sands, *The Feynman lectures on physics*, Vol. 1 (Addison-Wesley, Reading, 1965).
- [19] W.L. Wolfe and G.J. Zissis, eds., *The infrared handbook* (Office of Naval Research, Department of the Navy, Washington D.C., 1978).
- [20] C.W. Allen, *Astrophysical quantities*, third edition (The Athlone Press, University of London, London, 1976).
- [21] J.P. Holman, *Heat transfer*, fifth edition (Mc-Graw Hill, New York, 1981).
- [22] D.A.B. Miller, *Optics Lett.* 14 (1989) 146.