Connectivity Models for Optoelectronic Computing Systems

Haldun M. Ozaktas

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

Abstract. Rent's rule and related concepts of connectivity such as dimensionality, line-length distributions, and separators have found great use in fundamental studies of different interconnection media, including superconductors and optics, as well as the study of optoelectronic computing systems. In this paper generalizations for systems for which the Rent exponent is not constant throughout the interconnection hierarchy are provided. The origin of Rent's rule is stressed as resulting from the embedding of a high-dimensional information flow graph to two- or three-dimensional physical space. The applicability of these traditionally solid-wire-based concepts to free-space optically interconnected systems is discussed.

1 Connectivity, Dimensionality, and Rent's Rule

The importance of wiring models has long been recognized and they have been used not only for design purposes but also for the fundamental study of interconnections and communication in computing. A central and ubiquitous concept appearing in such contexts is the connectivity of a circuit graph or computer network. Connectedness has always been a central concept in mathematical graph theory [1–3], whose extensions play a central role in graph layout [4]. The purpose of these concepts is to quantify the communication requirements in computer circuits. This paper aims to discuss several concepts related to the connectivity of circuits (such as Rent's rule, dimensionality, line-length distributions, and separators), provide certain extensions, and briefly discuss some of their applications. We first discuss generalizations for systems for which the Rent exponent is not constant throughout the interconnection hierarchy, and present a number of related results. Special emphasis is given to the role of discontinuities and the origin of Rent's rule. The applicability of these concepts to free-space optically interconnected systems and the role of Rent's rule in fundamental studies of different interconnection media, including superconductors and optics, are briefly reviewed.

Graph layout deals with the problem of how to situate the nodes and edges of an abstract graph in physical space. Optimal graph layout [5] is in general an NP-complete problem [6]. However, if a hierarchical decomposition of a graph is provided, this graph can be laid out following relatively simple algorithms. A

J. Rolim et al. (Eds.): IPDPS 2000 Workshops, LNCS 1800, pp. 1072-1088, 2000.

hierarchical decomposition of a graph consisting of N nodes and the associated decomposition function $k(\bar{N})$ are obtained as follows: First, we remove $k(N)$ edges in order to disconnect the graph into 2 subgraphs, each of approximately $N/2$ nodes. Roughly speaking, we try to do this by removing as few edges as possible. We repeat this procedure for the subgraphs thus created. The subgraphs will in general require differing numbers of edges to be removed from them to be disconnected into subsubgraphs of $\simeq N/2^2$ nodes each. We denote the largest of these numbers as $k(N/2)$. Continuing in this manner until the subgraphs consist of a single node each, we obtain the function $k(\bar{N})$, the (worst case) number of edges removed during decomposition of subgraphs of $\bar{N}$ nodes. Once such a decomposition is found, it is possible to lay out the graph in the intuitively obvious manner by working upwards [7, 4, 6]. Whereas one can always find some such decomposition, finding the decomposition that leads to a layout with some optimal property (such as minimal area) is not a trivial problem. We will assume that we agree on a particular decomposition obtained by some heuristic method.

Now, let us define the *connectivity* $p(\bar{N})$ and *dimensionality* $n(\bar{N})$ associated with the hierarchical level of a decomposition involving subgraphs of $\bar{N}$ nodes by [8]

$$p(\bar{N}) = \frac{n(\bar{N}) \Leftrightarrow 1}{n(\bar{N})} = \log_2 \frac{k(\bar{N})}{k(\bar{N}/2)}. \tag{1}$$

In general the defined quantities satisfy $0 \leq p(\bar{N}) \leq 1$ and $1 \leq n(\bar{N}) \leq \infty$. It is possible to find many examples of graphs for which the values of $k(\bar{N})$ and $p(\bar{N})$ for different values of $\bar{N}$ are totally erratic and have no correlation whatsoever. However, both computer circuits and natural systems are observed to exhibit varying degrees of continuity of the functions $k(\bar{N})$ and $p(\bar{N})$.

Let the geometric derivative $\tilde{f}$ of a function f at the point x be defined analogous to the usual arithmetic derivative:

$$\tilde{f}(x) = \lim_{\nu \to 1^+} \log_\nu \frac{f(\nu x)}{f(x)}. \tag{2}$$

If $k(\bar{N})$ is slowly varying, we may pretend that it is a continuous function and write

$$p(\bar{N}) = \frac{n(\bar{N}) \Leftrightarrow 1}{n(\bar{N})} = \tilde{k}(\bar{N}), \tag{3}$$

which may be inverted as

$$k(\bar{N}) = k(1) \exp\left(\int_1^{\bar{N}} \frac{p(\bar{N}')}{\bar{N}'} \, d\bar{N}' \right), \tag{4}$$

where $\bar{N}'$ is a dummy variable.

Of course, since $k(\bar{N})$ is actually a function of a discrete variable, we cannot actually let $\nu \to 1$. The smallest meaningful value of ν in our context is 2. Hence the geometric derivative should be interpreted in the same sense as we interpret the common derivative in the form of a finite difference for discrete functions.

For our definition to make sense, $k(\bar{N})$ must be a slowly varying function. As already noted, this is indeed observed over large variations of $\bar{N}$ in both computer circuits and natural systems. In fact, in many cases it is found that $p(\bar{N})$ and $n(\bar{N})$ are approximately constant over a large range of $\bar{N}$. Such systems are said to exhibit self similarity [9,10], or scale invariance.

Assuming that $p(\bar{N}) = p = \text{constant}$, we find from equation 4 that $k(\bar{N}) = k(1)\bar{N}^p$. Apart from a constant coefficient, this is nothing but Rent's rule [11, 12] which gives the number of graph edges $k(\bar{N})$ emanating from partitions of computer circuits containing $\bar{N}$ nodes (such as the number of pinouts of an integrated circuit package containing $\bar{N}$ gates). $k(1)$ is interpreted as the average number of edges per node and p is referred to as the Rent exponent.

Conversely, by taking Rent's rule as a starting point, however allowing the exponent to be a function $p(\bar{N})$ of $\bar{N}$, we can derive equation 4 by working up the hierarchy. (Readers wishing to skip this reverse derivation given below may move directly to the paragraph following equation 9.) We consider a system with a total of N primitive elements and express N in the form

$$N = N_m = \prod_{i=1}^{m} \nu_i, \tag{5}$$

i.e., we have ν_m groups of ν_{m-1} groups of ... of ν_1 primitive elements. We are assuming all subgroups of any group to be identical (it is of course possible to go one step further and remove this restriction as well). The ν_i are sufficiently small so that the connectivity requirements within each level of the hierarchy can be assumed constant. The subtotals at any level are similarly expressed as

$$N_j = \prod_{i=1}^{j} \nu_i, \qquad 1 \leq j \leq m, \tag{6}$$

with $N_0 = 1$. If we let i and j approach continuous variables we can write this as

$$N(j) = \exp\left(\int_1^j \ln \nu(i)\, di\right). \tag{7}$$

Let it be the case that the ν_j subgroups forming one of ν_{j+1} groups have connectivity requirements characterized by a Rent exponent of p_j. The number of edges k_j emanating from each of the ν_{j+1} groups (each containing ν_j subgroups) is

$$k_j = k_{j-1}\nu_j^{p_j} = k_0 \prod_{i=1}^{j} \nu_i^{p_i}, \tag{8}$$

where k_0 is the number of edges of the primitive elements. In continuous form we have

$$k(j) = k(0) \exp\left(\int_1^j p(i) \ln \nu(i)\, di\right). \tag{9}$$

Now, it is possible to combine this with equation 7 to eliminate $\nu(i)$ to obtain equation 4, completing the derivation.

The dimensionality and Rent exponent will depend not only on the graph, but also on the layout of the graph. However, since there must be some layout of the graph which results in the smallest exponent, this smallest exponent may be considered intrinsic to the graph and representative of the intrinsic information flow requirements of the computational problem. The minimum information flow requirements can be quantified for several relatively structured and simple problems, such as sorting, fast Fourier transforms, etc. [4]. However, it should not be forgotten that there may be no efficient method of finding the layout resulting in the smallest exponent, so that it may not be possible to determine this intrinsic minimum Rent exponent.

We now briefly discuss the concept of *separators*. A graph of N nodes is said to have an $S(\bar{N})$ separator (or to be $S(\bar{N})$ separable) if the graph can be disconnected into two (roughly equal) subgraphs by removal of $S(N)$ edges and if the subgraphs thus created are also $S(\bar{N})$ separable [4]. Although we do not go into the details, we note that a graph with connectivity function $p(\bar{N})$ has a separator of the form $S(\bar{N}) \propto \bar{N}^{\max[p(\bar{N})]}$ where the maximum is taken over the whole domain of $\bar{N}$. Separators play a central role in combinatoric approaches to graph layout [7, 4], sometimes referred to as area-volume complexity theory. (An alternative way of describing the communication requirements of graphs is based on what are called *bifurcators* [6].)

Thus, we see that both Rent's rule and separators of the form $S(\bar{N}) \propto \bar{N}^p$ are special cases of the more general formalism we have introduced. Apart from minor technicalities involved in their definition, all are essentially equivalent when $p(\bar{N}) =$ constant. In general, $p(\bar{N})$ and $n(\bar{N})$ will be functions of $\bar{N}$.

The dimensionality $n(\bar{N})$ defined in equation 3 is a fractal dimension [10, 13–15]. Fractal dimensions of natural systems, just as those of computer circuits, may also vary as we ascend or descend the hierarchical structure of a system. In any event, Rent's rule, fractal geometry and separators are tied together by the notions of self similarity, scale invariance, and continuity in the relationships between the volume-like (number of nodes) and surface-like (number of edges) quantities.

To clarify this point, we offer the following explanation of why the quantity defined as $n = 1/(1 \Leftrightarrow p)$ is referred to as a "dimension." The perimeter of a square region is proportional to the $1/2$ power of its area. The surface area of a cube is proportional to the $2/3$ power of its volume. In general, the hyperarea enclosing a hyperregion of e dimensions is proportional to the $(e \Leftrightarrow 1)/e$ power of its hypervolume. Let us now make an analogy between "hyperarea" $\Leftrightarrow$ "number of graph edges emanating from a region," and "hypervolume" $\Leftrightarrow$ "number of nodes in the region." According to Rent's rule, the number of edges emanating from the region is proportional to the pth power of the number of nodes in the region. Thus, it makes sense to speak of the quantity n defined by the relation $p = (n \Leftrightarrow 1)/n$ as a "dimension." Note that in general n need not be an integer.

Now, let us assume that a graph with $n(\bar{N}) = n =$ constant is laid out in e-dimensional Euclidean space according to the divide-and-conquer layout algorithm (i.e., as intuitively suggested by its decomposition) [7, 4]. (e is often

$= 2$ but always ≤ 3.) Such a layout will internally satisfy Rent's rule. Donath [16] and Feuer [17] had shown that such a layout has a distribution $g(r)$ of line lengths of the form

$$g(r) \propto r^{-\frac{e}{n}-1}, \qquad r \leq \sqrt{e}N^{\frac{1}{e}}, \tag{10}$$

where r denotes line lengths in units of node-to-node grid spacing of the layout. (We assume the nodes are situated on a regular Cartesian grid.) The relationship between such inverse-power-law distributions and fractal concepts was discussed by Mandelbrot [18–20], closing the circle. Using the above distribution, or by combinatoric methods, we can show that when $n > e$ the average connection length $\bar{r}$ of such a layout of N elements is given by [9]

$$\bar{r} = \kappa(n,e)N^{\frac{1}{e}-\frac{1}{n}}, \tag{11}$$

where $\kappa(n,e)$ is a coefficient of the order of unity. The accuracy of this expression requires that $N^{1/e-1/n} \gg 1$. This result has a simple interpretation. The average connection length is simply the ratio of the linear extent $N^{1/e}$ of the system in e-dimensional space to the linear extent $N^{1/n}$ in n-dimensional space. The node-to-node grid spacing necessary to lay out a graph of dimensionality n is given by $\propto \bar{r}^{1/(e-1)}\lambda \approx N^{(n-e)/ne(e-1)}\lambda$, where λ is the line-to-line spacing of whatever interconnection technology is being used [21–23]. Thus, when $n > e$, the area (or volume) per node grows with N. This has been referred to as *space dilation* [24]. Examples of graphs with well-defined structures which exhibit large values of n are hypercubes, butterflies, and shuffle-exchange graphs. It is also easily verified that the given definition of dimension is consistent with that for multidimensional meshes [25].

Before closing this section, we discuss two further results regarding the calculation of the average and total connection lengths of a layout. Readers wishing to skip these may directly go to the next section.

First, we discuss the invariance of the total connection length under different grain-size viewpoints. One might view a computing system as a collection of its most primitive elements, for instance gates or transistors. Alternately, one might prefer to view it as a collection of higher-order elements, such as chips or processors which are simply taken as black boxes with a certain number of pinouts. Both viewpoints are perfectly valid, however one must interpret the average connection length with care so as to maintain consistency. The average connection length will be higher when calculated with reference to the higher-order picture, as compared to the lower-order picture. This is because the shorter interconnections inside the black boxes are not being taken into account while computing the average. Let us quantify this situation by considering N/N_1 black boxes (blocks) with N_1 primitive elements in each, representing a simple partitioning of the total N elements. Let the grid spacing of the black boxes be d_1 and that of the primitive elements be d_0. In an e-dimensional space we have $d_1 = N_1^{1/e}d_0$. Let us consider $n =$ constant $> e$. We let $\bar{\ell}$ denote the average interconnection length in real units, as opposed to $\bar{r}$ which is the average connection length in grid units. ℓ_{total} will denote the total interconnection length. Furthermore,

assume $N_1^{1/e-1/n} \gg 1$ and $(N/N_1)^{1/e-1/n} \gg 1$. The fine-grain picture yields

$$\bar{\ell} = \bar{r} d_0 = \kappa N^{\frac{1}{e}-\frac{1}{n}} d_0, \tag{12}$$

whereas the large-grain picture yields

$$\bar{\ell} = \kappa (N/N_1)^{\frac{1}{e}-\frac{1}{n}} d_1 = \kappa N^{\frac{1}{e}-\frac{1}{n}} N_1^{\frac{1}{n}} d_0, \tag{13}$$

from which we see that indeed the large-grain picture yields a larger value of $\bar{\ell}$ (unless $n \to \infty$). However, in calculating the total system size, it is not $\bar{\ell}$ but rather ℓ_{total} that is the significant quantity. For the fine grain picture we have

$$\ell_{\text{total}} = k_0 N \bar{\ell} = k_0 \kappa N^{p+\frac{1}{e}} d_0, \tag{14}$$

since $p = 1 \Leftrightarrow 1/n$. For the large-grain picture we use the number of pinouts as given by $k_1 = k_0 N_1^p = k_0 N_1^{1-1/n}$, so that

$$\ell_{\text{total}} = k_1 (N/N_1) \bar{\ell} = k_0 \kappa N^{p+\frac{1}{e}} d_0, \tag{15}$$

identical to what we found with the fine grain picture. So whatever way we choose to look at it we always will end up calculating the total system size consistently.

This result means that we can ignore the contributions of the local interconnections in calculating the total area (or volume) required for wiring. The longer interconnections, although much fewer in number, constitute most of the wiring volume. The total interblock wiring length is, once again

$$k_0 N_1^{1-\frac{1}{n}} (N/N_1) \kappa (N/N_1)^{\frac{1}{e}-\frac{1}{n}} d_1, \tag{16}$$

whereas the total local wiring length is

$$(N/N_1) k_0 N_1 \kappa N_1^{\frac{1}{e}-\frac{1}{n}} d_0. \tag{17}$$

The ratio of the former to the latter is

$$\frac{\text{Total interblock wire length}}{\text{Total local wire length}} = (N/N_1)^{\frac{1}{e}-\frac{1}{n}}, \tag{18}$$

which we had assumed to be $\gg 1$ from the beginning. When n is bounded away from e and when N_1 and N/N_1 are large, it is the higher level of the interconnection hierarchy that limits how dense the elements can be packed. This conclusion can also be traced to the fact that the integrand $rg(r)$ in the first moment integral of $g(r)$ decays slower than $1/r$ when $n > e$.

Once ℓ_{total} is obtained, calculation of the system linear extent $N^{1/e} d_0$ is easy. We simply equate the total available area (volume) to the total wire area (volume) [26]

$$N d_0^e = \ell_{\text{total}} \lambda^{e-1}, \tag{19}$$

where λ is the line spacing. Thus, within a wiring inefficiency factor we obtain

$$N^{\frac{1}{e}} d_0 = (k_0 \kappa N^p)^{\frac{1}{(e-1)}} \lambda. \tag{20}$$

Next, we derive an expression for the layout area for a system with arbitrary, possibly discontinuous $k(\bar{N})$ in two dimensions. So as to simplify the representation of the results, we will restrict ourselves to $p(\bar{N}) > 1/2$. First, consider a group of ν_1 primitive elements. This group can be laid out with linear extent $d_1 = k_0\kappa_1\nu_1^{p_1}\lambda = k_1\kappa_1\lambda$ where κ_1 is the coefficient corresponding to p_1. Thus, the total system linear extent must be at least $(N/\nu_1)^{1/2} = (N/N_1)^{1/2}$ times the extent of this group, where in general N_j is given by equation 6. Now consider a supergroup of ν_2 such groups. Taking $\lambda = 1$, the linear extent of this supergroup satisfies

$$\max(k_2\kappa_2, \nu_2^{\frac{1}{2}} d_1) \quad \leq \quad d_2 \quad \leq \quad k_2\kappa_2 + \nu_2^{\frac{1}{2}} d_1. \tag{21}$$

In general,

$$\max(k_j\kappa_j, \nu_j^{\frac{1}{2}} d_{j-1}) \quad \leq \quad d_j \quad \leq \quad k_j\kappa_j + \nu_j^{\frac{1}{2}} d_{j-1}. \tag{22}$$

$k_j\kappa_j$ is the wiring requirement obtained at the jth level. $\nu_j^{1/2} d_{j-1}$ is the requirement inherited from lower levels. The maximum and summation represent best case and worst case assumptions on how these requirements interact. Taking $d_0 = 1$, and expanding the recursion on both sides leads to the following bounds for the linear extent d_m of the complete system:

$$\max \left[k_i\kappa_i \left(\frac{N}{N_i}\right)^{\frac{1}{2}} \right]_{i=1}^{m} \quad \leq \quad d_m \quad \leq \quad \sum_{i=1}^{m} k_i\kappa_i \left(\frac{N}{N_i}\right)^{\frac{1}{2}}. \tag{23}$$

The right hand side of the above can be at most m times greater than the left hand side. Since $m \leq \log_2 N$, the system linear extent is given by the left hand side within this logarithmic factor. In continuous form, the linear extent is given by

$$N^{\frac{1}{2}} \max \left[\frac{k(\bar{N})\kappa(\bar{N})}{\bar{N}^{\frac{1}{2}}} \right]_{\bar{N}}, \tag{24}$$

where $\kappa(\bar{N}) \sim 1$ is only weakly dependent on $\bar{N}$.

If $k(\bar{N})$ is slowly varying, it may be expressed by equation 4 over the whole range of $\bar{N}$. Now if $p(\bar{N}) > 1/2$ as we have assumed, it is possible to show that $k(\bar{N})$ grows faster than $\bar{N}^{1/2}$ so that the expression in the square brackets above is maximized for $\bar{N} = N$. Thus, the system linear extent is given by

$$k(N)\kappa(N) \sim k(N). \tag{25}$$

That is, the system size is set by the highest level of interconnections if $k(\bar{N})$ is a slowly varying function and $p(\bar{N}) > 1/2$. This implies that the choice of interconnection technology for the highest level is the most critical.

2 Discontinuities and the Origin of Rent's Rule

Whereas it is observed that the function $k(N)$ exhibits considerable continuity over large variation of N, it is also observed that it occasionally exhibits sharp

discontinuities. In other words, it no longer becomes possible to predict the value of the function $k(N)$ for certain N by knowing its values at nearby N. For instance, in the context of Rent's rule, it may not be possible to predict the number of pinouts of a VLSI chip by observing its internal structure, or vice versa [13]. However, this does not imply that Rent's rule (in its generalized form, as given by equation 4) is useless. Consider a multiprocessor computer. Rent's rule may be used to predict the wiring requirements internal to each of the processors. It may also be used for similar purposes for the interconnection network among the processors. In fact, the Rent exponent may even be similar in both cases. However, the function $k(N)$ may exhibit a steep discontinuity (often downward), as illustrated in figure 1 [8]. As is usually the case, a finite number of discontinuities in an otherwise smooth function need not inhibit us from piecewise application of our analytical expressions. Such discontinuities are often associated with the *self-completeness* of a functional unit [12, 13]. Similar examples may be found in nature. For instance, mammalian brains seem to satisfy $n > 3$ (i.e. $p > 2/3$), since the volume per neuron has been found to be greater in species with larger numbers of neurons [27]. The human brain has 10^{11} neurons each making about 1000 connections [28]. Thus, we would expect at least $1000(10^{11})^{2/3} \sim 10^{10}$ "pinouts." However, we have only about 10^6 fibers in the optic nerve and 10^8 fibers in the *corpus callosum*.

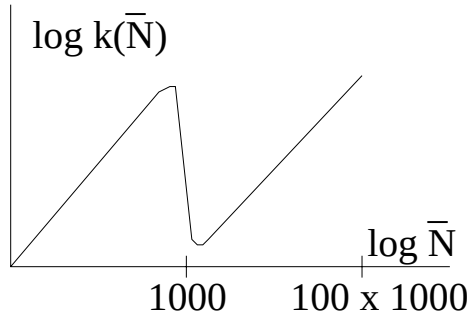

Fig. 1. $k(\bar{N})$ for a system of $N = 100 \times 1000$ primitive elements consisting of 100 processors of 1000 elements each. The number of "pinouts" of the processors bears no relationship to their internal structure. Equation 4 may be used directly for the range $1 < \bar{N} < 1000$, and with a shift of origin for the range $1000 < \bar{N} < 100 \times 1000$.

In the context of microelectronic packaging, a quote from C. A. Neugebauer offers some insight as to why such discontinuities are observed: "Since the I/O capacity (of the chip carrier) is exceeded, a significant number of chips can be interconnected only if the pin/gate ratio can be drastically reduced, normally well below that predicted by Rent's rule. Rent's rule can be broken at any level of integration. The microprocessor chip is an example of the breaking of Rent's rule in its original form for gate arrays on the chip level. Being able to delay the breaking of Rent's rule until a much higher level is always an advantage

because it preserves many parallel data paths even at very high levels of integration, and thus offers higher systems performance and greater architectural flexibility." [29] The breaking of Rent's rule seems to be a technological necessity, and undesirable from a systems viewpoint. We will later discuss studies which indicate that superconducting or optical interconnections may allow the maintainment of a large dimensionality and Rent exponent throughout higher levels of the hierarchy.

The origin of Rent's rule has intrigued many researchers. Donath had shown that Rent's rule is a consequence of the hierarchical nature of the logic design process [30, 31]. Some have viewed it merely as an empirical observation obtained from an examination of existing circuits. Others have suggested that it is as natural as the branching of trees or the human lung (a consequence of their growth process), or that it represents the adaptation of computer circuits to serve the needs of percolation of information. Fractal concepts have been quite successful in describing natural phenomena. However, it is often more challenging to explain why fractal forms come up so often. Why do computer circuits lend to such a description? One suspects that fractal forms may exhibit certain optimal properties. For instance, bitonic (divide-and-conquer) algorithms can be viewed as elementary fractal forms. Is it possible to postulate general principles (such as the principle of least action in mechanics) regarding optimal information flow or computation that would lead to an inverse-power-law distribution of line lengths (a constant fractal dimension)? Mandelbrot has postulated maximum entropy principles to predict the observed inverse-power-law distribution of word frequencies (linguistics) [19] and monetary income (economics) [20]. Christie has pursued the idea that the wires in a computing system should obey Fermi-Dirac statistics, based on the observation that the wires are indistinguishable (any two wires of same length can be exchanged) and that they obey an exclusion principle (only one wire need connect two points) [32, 33]. Keyes [27] has shown how the number of distinct ways one can wire up an array of elements increases with average wire length. In [34] we showed that the number of distinct ways one can "wire up" an optical interconnection system increases similarly with a fundamental quantity known as the space-bandwidth product of the optical system, and thus the average interconnection length.

The author finds the following viewpoint especially illuminating. At the microscopic level, all information processing involves the distributed manipulation and back-and-forth transfer of pieces of information. There is a certain requirement on the amount of information that must flow or percolate depending on the particular problem we are trying to solve. This requirement can be embodied in an information flow graph. The dimensionality of this graph can then be taken as a measure of the information flow requirements of the problem. For some problems which require little transfer of information, this dimension may be small. For others, it may be large. When the dimensionality associated with the problem exceeds the dimensions of the physical space in which we construct our circuits (often 2 but at most 3), we are faced with the problem of embedding a higher-dimensional graph into a lower-dimensional space. This is what leads

to Rent's rule: the fact that we try to solve problems with inherently higher dimensionality of information flow than the two- or three-dimensional physical spaces we build our computers in.

Several structured problems, such as sorting and discrete Fourier transforming, are known to have global information flow requirements leading to separators which are $\propto \bar{N}$, corresponding to large dimensions and nearly unity Rent exponents. The dimensionality associated with general purpose computing may also be presumed to be large. In any event, it certainly seems that quite a fraction of interesting problems have dimensions higher than two or three, so that the space dilation effect associated with Rent's rule is expected.

Despite these considerations, Rent's rule may not apply to a particular circuit we examine. The challenges involved in dealing with greater numbers of interconnections may lead designers to reduce the number of physical ports and channels, and to shift the "communication burden" to other levels of the computational hierarchy [35]. Careful examination often reveals that the price of reducing the number of wires is often paid in terms of computation time, intermediated by techniques such as multiplexing or breaking the transfer of information into multiple steps. Clever schemes can reduce the number of wires that are apparently needed, but these often essentially amount to reorganizing the processing of information in such a way that the same information is indirectly sent in several pieces or different times. Ultimately, a certain flow and redistribution of information must take place before the problem is solved.

Several levels of graphs can come between the $n \gg 1$ dimensional graph characterizing the information flow requirements of the problem to be solved, and the $e \leq 3$ dimensional physical space. These graphs correspond to different levels of the computational hierarchy, ranging from the abstract description of the problem to the concrete physical circuits. The dimensionality of these graphs provide a stepwise transition from n dimensions to e dimensions (figure 2). Level transitions involving large steps (steep slopes) are where the greatest implementation burden is felt. For line a in figure 2, this burden is felt at the relatively concrete level, and for line c at the relatively abstract level. The burden is more uniformly spread for line b. Shifting the burden from one level to the others may be beneficial because of the different physical and technological limitations associated with each level. Techniques such as algorithm redesign, multiplexing, parallelism, use of different kinds of local or global interconnection networks, use of alternative interconnection technologies such as optics, can be used to this end. Better understanding and deliberate exploitation of these concepts and techniques may be expected to translate into practical improvements.

A particular question that may be posed in this context is whether the burden should lean primarily towards the software domain or primarily towards the hardware domain. An embodiment of the first option may be a nearest-neighbor connected mesh-type computer in which the physical interconnect problem is minimized. Global flows of information are realized indirectly as pieces of information propagate from one neighbor to the next. The second option, in contrast, might rely on direct transfer of information through dedicated global lines which

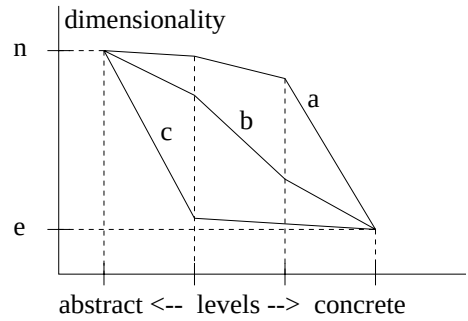

Fig. 2. The dimensionality of graphs corresponding to different levels for a hypothetical system with four levels.

result in heavy physical interconnect congestion. Although determination of the proper balance between these two extremes is in general a very complex issue, it has been addressed in a specific context in [36]. The conclusion is that use of direct global lines is more beneficial than simulating the same information flow on a locally connected system. This conclusion assumes the use of optical lines to overcome the severe limitations associated with resistive interconnections.

Contexts in which the nature of the problem to be solved does require global information flows, but only at a relatively low rate, may result in poor utilization of dedicated global lines, which nevertheless contribute significantly to system area or volume. This situation can be especially common with optical interconnections which can exhibit very high bandwidths which are difficult to saturate. For this reason, techniques have been developed for organizing information flow such that distinct pairs of transmitters and receivers can share common high-bandwidth channels to make the most of the area or volume invested in them [37].

3 Free-Space Optical Interconnections

The concepts discussed in this paper are immediately applicable to three-dimensional layouts [38–40], including those based on optical waveguides or fibers. However, the extension of results originally developed for "solid wires" to free-space optics, which can offer much higher density than waveguides and fibers, is not immediate.

Since optical beams can readily pass through each other, it has been suggested that optical interconnections may not be subject to area-volume estimation techniques developed for solid wires. However, proper accounting for the effects of diffraction leads to the conclusion that from a global perspective, optical interconnections can also be treated as if they were solid lines for the purpose of area and volume estimation, so that most of the concepts discussed in this paper are applicable to free-space optical systems as well.

This conclusion is based on the following result [41]: *The minimum total communication volume required for an optical system whose total interconnection length is* ℓ_{total} *is given by* $\ell_{\text{total}}\lambda^2$. This result is stated globally; it does not imply that each optical channel individually has cross-sectional area λ^2, but only that the total volume must satisfy this minimum. Indeed some channels may have larger cross-sectional areas but share the same extent of space with other channels which pass through them. The bottom line is that even with the greatest possible amount of overlap and space sharing, the global result is as if each channel required a cross-sectional area of λ^2, as if they were solid wires. If the average connection length in grid units is given by $\bar{r} = \kappa N^{p-2/3}$ as before, then the minimum grid spacing d must satisfy $Nd^3 = Nk\bar{r}d\lambda^2$, leading to a minimum system linear extent of $N^{1/3}d = (k\kappa N^p)^{1/2}\lambda$, just as would be predicted for solid wires of width λ (equation 20 with $e = 3$ and the subscript 0 suppressed) [42].

In many optical systems, the devices are restricted to lie on a plane, rather than being able to occupy a three-dimensional grid. Although in general these systems are subject to the same results, certain special considerations apply [43–46].

The above does not imply that there is no difference between optical and electrical interconnections. Optical interconnections allow the realization of three-dimensional layouts. Optical beams can pass through each other, making routing easier. Furthermore, the linewidth and energy dissipation for optical interconnections is comparatively smaller for longer lines. (This latter advantage is also shared by superconducting lines.)

4 Fundamental Studies of Interconnections

Rent's rule and associated line-length distributions have been of great value in fundamental studies of integrated systems [47–50]. Two considerations are fundamental in determining the minimum layout size and thus the signal delay: interconnection density and heat removal [51–54]. Both considerations are interrelated since, for instance, the energy dissipation on a line also depends on its length, which in turn depends on the grid spacing, which in turn depends on both the total interconnection length and the total power dissipated. The complex interplay between the microscopic and macroscopic parameters of the system must be simultaneously analyzed. Rent's rule and line-length distributions are indispensable to this end. However, it is necessary to complement these tools with physically accurate models of interconnection media. Such analytical models for normally conducting, repeatered, superconducting, and optical interconnections which take into account the skin effect, both unterminated and terminated lines, optimization of repeater configurations, superconducting penetration depth and critical current densities, optical diffraction, and similar effects have been developed in [43, 44] and subsequently applied to determine the limitations of these interconnection media and their relative strengths and weaknesses [43, 44, 40, 55–57, 36, 58]. Treating inverse signal delay S and bandwidth B as performance

parameters, these studies characterize systems with N elements by surfaces of physical possibility in S-B-N space, which are to be compared with surfaces of algorithmic necessity in the same space.

This approach has allowed comparative studies of different interconnection media to move beyond comparisons of isolated electrical and optical lines, to evaluation of the effects of their different characteristics at the system level. These studies clearly show the benefit of optical and superconducting interconnections for larger systems. One of the most striking results obtained is that there is an absolute bound on the total rate of information that can be swapped from one side of an electrically connected system to the other, and that this bound is independent of scaling. Such a bound does not exist for optics and superconductors [43, 59].

An interesting extension is to allow the longer lines in a system to be of greater width to keep their RC delays within bounds. Use of the calculus of variations has shown that the widths of lines should be chosen proportional to the cube root of their length for two-dimensional layouts and to the fourth root of their length for three-dimensional layouts [60]. Staircase approximations to these analytical expressions can serve as practical design guidelines.

These studies have also been extended to determine how electrical and optical interconnections can be used together. It is generally accepted that optics is favorable for the longer lines in a system whereas the shorter lines should be electrical. Results based on comparisons of isolated lines may not be of direct relevance in a system context. The proper question to ask is not "Beyond what length must optical interconnections be used?", but "Beyond how many logic elements must optical interconnections be used?". Studies have determined that optical interconnections should take over around the level of 10^4-10^6 elements [61–63].

This body of work has demonstrated that inverse-power-law type line-length distributions are very suitable for such studies. This is because distributions which decay faster, such as an exponential distribution, effectively behave like fully local distributions in which connections do not reach out beyond a bounded number of neighbors. Such layouts are essentially similar to nearest-neighbor connected layouts, and are already covered by Rent's rule when we choose $n = e$. On the other hand, for any layout in which the number of connections per element is bounded, the behavior is at worst similar to that described by a Rent exponent of unity. Thus, although all systems may not exhibit a precise inverse-power-law distribution of line lengths, Rent's rule is nevertheless sufficient to represent the range of general interest.

5 Conclusion

We believe that many criticisms of Rent's rule are a result of not allowing the Rent exponent and dimensionality to vary as we ascend the hierarchy and a failure to recognize discontinuities. It seems that in most cases of practical interest, the decomposition function $k(\bar{N})$ is piecewise smooth with a finite number of

discontinuities. The role of discontinuities in an otherwise smooth decomposition function, and whether it is beneficial to construct systems in the form of a hierarchy of functionally complete entities, are less understood issues. Is it functionally desirable to construct systems that way, or do physical and technical limitations force us to?

Parts of this work appeared in or were adapted from [8].

References

1. B. Bollobas. *Graph Theory: An Introductory Course.* Springer, Berlin, 1979.
2. G. Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Wellesley, Massachusetts, 1986.
3. H. N. V. Temperley. *Graph Theory and Applications.* Ellis Horwood Ltd., Chichester, 1981.
4. J. D. Ullman. *Computational Aspects of VLSI.* Computer Science Press, Rockville, Maryland, 1984.
5. T. C. Hu and E. S. Kuh. *VLSI Circuit Layout: Theory and Design.* IEEE Press, New York, 1985.
6. S. N. Bhatt and F. T. Leighton. A framework for solving VLSI layout problems. *J Computer System Sciences*, 28:300–343, 1984.
7. C. E. Leiserson. *Area-Efficient VLSI Computation.* The MIT Press, Cambridge, Massachusetts, 1983.
8. H. M. Ozaktas. Paradigms of connectivity for computer circuits and networks. *Optical Engineering*, 31:1563–1567, 1992.
9. W. E. Donath. Placement and average interconnection lengths of computer logic. *IEEE Trans Circuits Systems*, 26:272–277, 1979.
10. L. Pietronero. Fractals in physics: Introductory concepts. In S. Lundqvist, N. H. March, and M. P. Tosi, eds., *Order and Chaos in Nonlinear Physical Systems.* Plenum, New York, 1988.
11. B. S. Landman and R. L. Russo. On a pin versus block relationship for partitions of logic graphs. *IEEE Trans Computers*, 20:1469–1479, 1971.
12. R. L. Russo. On the tradeoff between logic performance and circuit-to-pin ratio for LSI. *IEEE Trans Computers*, 21:147–153, 1972.
13. D. K. Ferry. Interconnection lengths and VLSI. *IEEE Circuits Devices Mag*, pages 39–42, July 1985.
14. B. B. Mandelbrot. *Fractals: Form, Chance and Dimension.* W. H. Freeman, San Francisco, 1977.
15. P. Christie, J. E. Cotter, and A. M. Barrett. Design and simulation of optically interconnected computer systems. In *Interconnection of High Speed and High Frequency Devices and Systems, Proc SPIE*, 947:19–24, 1989.
16. W. E. Donath. Wire length distribution for placements of computer logic. *IBM J Research Development*, 25:152–155, 1981.
17. M. Feuer. Connectivity of random logic. *IEEE Trans Computers*, 31:29–33, 1982.
18. B. B. Mandelbrot. *The Fractal Geometry of Nature.* W. H. Freeman, New York, 1983.
19. B. B. Mandelbrot. Information theory and psycholinguistics: A theory of word frequencies. In P. F. Lazarsfeld and N. W. Henry, eds., *Readings in Mathematical Social Science.* The MIT press, Cambridge, Massachusetts, 1968.

20. B. B. Mandelbrot. The Pareto-Levy law and the distribution of income. *Int Economic Review*, 1:79–106, 1960.
21. I. E. Sutherland and D. Oestreicher. How big should a printed circuit board be? *IEEE Trans Computers*, 22:537–542, 1973.
22. W. R. Heller, W. F. Mikhail, and W. E. Donath. Prediction of wiring space requirements for LSI. *J Design Automation Fault Tolerant Computing*, 2:117–144, 1978.
23. A. El Gamal. Two-dimensional stochastic model for interconnections in master slice integrated circuits. *IEEE Trans Circuits Systems*, 28:127–134, 1981.
24. A. C. Hartmann and J. D. Ullman. Model categories for theories of parallel systems. In G. J. Lipovski and M. Malek, eds,, *Parallel Computing: Theory and Experience.* Wiley, New York, 1986.
25. W. J. Dally. *A VLSI Architecture for concurrent data structures.* Kluwer, Norwell, Massachusetts, 1987.
26. R. W. Keyes. The wire-limited logic chip. *IEEE J Solid State Circuits*, 17:1232–1233, 1982.
27. R. W. Keyes. Communication in computation. *Int J Theoretical Physics*, 21:263–273, 1982.
28. R. F. Thompson. *The Brain.* W. H. Freeman and Company, New York, 1985.
29. C. A. Neugebauer. Unpublished manuscript.
30. W. E. Donath. Stochastic model of the computer logic design process. Tech Rep RC 3136, IBM T. J. Watson Research Center, Yorktown Heights, New York, 1970.
31. W. E. Donath. Equivalence of memory to 'random logic'. *IBM J Research Development*, 18:401–407, 1974.
32. P. Christie and S. B. Styer. Fractal description of computer interconnection distributions. In *Microelectronic Interconnects and Packaging: System and Process Integration, Proc SPIE*, 1390, 1990.
33. P. Christie. Clouds, computers and complexity. In S. K. Tewksbury, ed., *Frontiers of Computing Systems Research, Volume 2*, pages 197–238. Plenum, New York, 1991.
34. H. M. Ozaktas, K.-H. Brenner, and A. W. Lohmann. Interpretation of the space-bandwidth product as the entropy of distinct connection patterns in multifacet optical interconnection architectures. *J Optical Society America A*, 10:418–422, 1993.
35. H. M. Ozaktas. Levels of abstraction in computing systems and optical interconnection technology. In P. Berthomé and A. Ferreira, eds., *Optical Interconnections and Parallel Processing: Trends at the Interface*, chapter 1. Kluwer, Dordrecht, The Netherlands, 1998.
36. H. M. Ozaktas and J. W. Goodman. Comparison of local and global computation and its implications for the role of optical interconnections in future nanoelectronic systems. *Optics Communications*, 100:247–258, 1993.
37. H. M. Ozaktas and J. W. Goodman. Organization of information flow in computation for efficient utilization of high information flux communication media. *Optics Communications*, 89:178–182, 1992.
38. A. L. Rosenberg. Three-dimensional VLSI: a case study. *J Assoc Computing Machinery.* 30:397–416, 1983.
39. F. T. Leighton and A. L. Rosenberg. Three-dimensional circuit layouts. *J Computer System Sciences*, 15:793–813, 1986.
40. H. M. Ozaktas and M. F. Erden. Comparison of fully three-dimensional optical, normally conducting, and superconducting interconnections. In *2nd Workshop on Optics and Computer Science*, April 1, 1997, Geneva. Submitted to *Applied Optics.*

41. H. M. Ozaktas and J. W. Goodman. Lower bound for the communication volume required for an optically interconnected array of points. *J Optical Society America A*, 7:2100–2106, 1990.
42. H. M. Ozaktas, Y. Amitai, and J. W. Goodman. A three dimensional optical interconnection architecture with minimal growth rate of system size. *Optics Communications*, 85:1–4, 1991.
43. H. M. Ozaktas and J. W. Goodman. The limitations of interconnections in providing communication between an array of points. In S. K. Tewksbury, ed., *Frontiers of Computing Systems Research, Volume 2*, pages 61–124. Plenum, New York, 1991.
44. H. M. Ozaktas. *A Physical Approach to Communication Limits in Computation.* PhD thesis, Stanford University, California, 1991.
45. H. M. Ozaktas, Y. Amitai, and J. W. Goodman. Comparison of system size for some optical interconnection architectures and the folded multi-facet architecture. *Optics Communications*, 82:225–228, 1991.
46. H. M. Ozaktas and D. Mendlovic. Multi-stage optical interconnection architectures with least possible growth of system size. *Optics Letters*, 18:296–298, 1993.
47. R. W. Keyes. *The Physics of VLSI Systems.* Addison-Wesley, Reading, Massachusetts, 1987.
48. R. W. Keyes. Fundamental limits in digital information processing. *Proc IEEE*, 69:267–278, 1981.
49. R. W. Keyes. The evolution of digital electronics towards VLSI. *IEEE Trans Electron Devices*, 26:271–279, 1979.
50. H. B. Bakoglu. *Circuits, Interconnections and Packaging for VLSI.* Addison-Wesley, Reading, Massachusetts, 1990.
51. H. M. Ozaktas, H. Oksuzoglu, R. F. W. Pease, and J. W. Goodman. Effect on scaling of heat removal requirements in three-dimensional systems. *Int J Electronics*, 73:1227–1232, 1992.
52. W. Nakayama. On the accomodation of coolant flow paths in high density packaging. *IEEE Trans Components, Hybrids, Manufacturing Technology*, 13:1040–1049, 1990.
53. W. Nakayama. Heat-transfer engineering in systems integration—outlook for closer coupling of thermal and electrical designs of computers. *IEEE Trans Components, Packaging, Manufacturing Technology, Part A*, 18:818–826, 1995.
54. A. Masaki. Electrical resistance as a limiting factor for high performance computer packaging. *IEEE Circuits Devices Mag*, pages 22–26, May 1989.
55. H. M. Ozaktas. Fundamentals of optical interconnections—a review. In *Proc Fourth Int Conf Massively Parallel Processing Using Optical Interconnections*, pages 184–189, IEEE Computer Society, Los Alamitos, California, 1997. (Invited paper, June 22–24, 1997, Montreal.)
56. H. M. Ozaktas. Toward an optimal foundation architecture for optoelectronic computing. Part I. Regularly interconnected device planes. *Applied Optics*, 36:5682–5696, 1997.
57. H. M. Ozaktas. Toward an optimal foundation architecture for optoelectronic computing. Part II. Physical construction and application platforms. *Applied Optics*, 36:5697–5705, 1997.
58. H. M. Ozaktas and J. W. Goodman. The optimal electromagnetic carrier frequency balancing structural and metrical information densities with respect to heat removal requirements. *Optics Communications*, 94:13–18, 1992.
59. D. A. B. Miller and H. M. Ozaktas. Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of the system architecture. *J Parallel Distributed Computing*, 41:42–52, 1997.

60. H. M. Ozaktas and J. W. Goodman. Optimal linewidth distribution minimizing average signal delay for RC limited circuits. *Int J Electronics*, 74:407–410, 1993.
61. H. M. Ozaktas and J. W. Goodman. Elements of a hybrid interconnection theory. *Applied Optics*, 33:2968–2987, 1994.
62. H. M. Ozaktas and J. W. Goodman. Implications of interconnection theory for optical digital computing. *Applied Optics*, 31:5559–5567, 1992.
63. A. V. Krishnamoorthy, P. J. Marchand, F. E. Kiamilev, and S. C. Esener. Grain-size considerations for optoelectronic multistage interconnection networks. *Applied Optics*, 31:5480–5507, 1992.