Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

P.M.B. Vitányi

Signal propagation delay, wire length distribution and the efficiency of VLSI circuits

Department of Computer Science Report CS-R8412 July

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

SIGNAL PROPAGATION DELAY, WIRE LENGTH DISTRIBUTION AND THE EFFICIENCY OF VLSI CIRCUITS

PAUL M.B. VITANYI

Centre for Mathematics and Computer Science, Amsterdam

Using sound electronic principles, a signal propagation delay logarithmic in the length of a wire can be attained in VLSI circuits only if all wires in the layout have the same aspect ratio. This results in a penalty in surface area of the order of the square of the length of the wire. Thus, the global complexity of a VLSI circuit is affected. In particular, the complexity becomes very layout dependent. This effect will be truly pronounced in the emerging wafer scale integration technology. Simple theoretical considerations and experimental study of actual circuits have shown elsewhere that the wire length distribution $f(i)$ for the layout of integrated logic chips, in particular for VLSI, tend to satisfy $f(i) = c / i^{\lambda}$ $(1 \leqslant i \leqslant L)$ and $f(i) \approx 0$ $(i > L)$. There are such wire-length distributions for which the logarithmic delay assumption entails at least an exponential increase in area over the constant wire width assumption for any layout with that distribution. Consequently, the wires need to get so much longer to achieve logarithmic delay that the absolute propagation delay turns out to be not improved over an original linear or quadratic propagation delay. Taking into account also the fact that the wide wires have to be actually placed in a layout, the logaritmic delay requirement may in some cases not be implementable at all, apart from the fact that even if it could it would give no improvement in absolute time.

1980 MATHEMATICS SUBJECT CLASSIFICATION: 68C25, 94C99

1982 CR CATEGORIES: B.7.0, F.2.3

KEYWORDS & PHRASES: very large scale integrated circuits (VLSI), wafer scale integration, logarithmic signal propagation delay, electronic principles, driving long wires, wire aspect ratio, wire length distributions in VLSI layouts, Rent's Rule, layout area, computational complexity, efficiency

NOTE: This report will be submitted for publication elsewhere.

Report CS-R8412

Centre for Mathematics and Computer Science

P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1. Introduction

In a realistic complexity theory for the *very large scale integrated (VLSI)* electronic switching circuits, which now become possible, the treatment of signal propagation delay in long wires plays a crucial part. This seems truly the case on the level of the emerging *wafer scale integration.* In this technology in effect chips with 4 inch diameter are manufactured. On a scale like that, the signal propagation delay in long wires becomes a major factor. Synchronization requirements slow down the computation to a clocked switching time in the order of the delay in the longest wire. Such delays can be reduced by change of feature sizes or choice of material. The latter course is taken by the military sensitive investigation of very high speed integrated circuits (VHSIC). For instance, the use of semi-insulating Gallium-Arsenide and Silicon-on-Sapphire insulating substrate technologies allow switching times in the subnanosecond range [EWZL, YYC, TR]. New devices are afoot like the up-and-coming high-electron-mobility transistor (HEMT) [MS]. Historically, metals like aluminium and gold have been used in MOS-IC's (metal on silicon). With the advent of silicon-gate MOS technology, polysilicon has been extensively used to form gate electrodes and interconnections. Now such metals as tungsten, molybdenum, titanium, tantalum and their silicides are considered for replacing/complementing polysilicon [SM]. The quest for speed on chips is no luxury; high speed is absolutely required by many real-time applications.

In the computational models rampant in the literature, the eloquently justified assumptions concerning signal propagation delay in long wires range from constant delay (irrespective of the wire length) [Th1, BK1, BK2, Vu, Sa], via logarithmic- [PRS, Ra, Th2, TR] and linear delay [CM, BPP] to signal propagation delay which is square in the length of wires [BPP, CM, Se]. In all of these models the width (and thickness) of wires is assumed to be a unit, depending on the minimal feature width of the underlying technology. Some of these models are more reasonable than others, while all of them may cover selected real life situations. Most papers cite [MR1, MC] as justification. However, recently [MR2] emphasized once again, that the constant delay assumption cannot be held in general at all, while the logarithmic delay assumption is incompatible with the constant wire width asumption. To achieve a propagation delay logarithmic in the length of the wire, by using a hierarchy of drivers [MR1, MC, MR2], electronic considerations show that all such wires need have the same ratio between width and length, that is, *aspect ratio.* Thus, in multilayered chips now being manufactured, the connect wires are grouped in metal layers according to length and in a layer with a group of longer wires those wires are proportionally wider and thicker. The surface area occupied by long wires has to be proportional to the squares of their lengths in order to obtain logarithmic propagation delay. This fact influences heavily the resultant complexity theory for VLSI computation models assuming logarithmic propagation delay; in particular it makes the complexity more layout sensitive than previous assumptions. These considerations form the subject matter of the present investigation.

The discussion in [MR1, MC, Section 8.5] apparently has created confusion; in particular it has propagated the belief that logarithmic propagation delay is reconcilable with approximately unit width wires in a VLSI layout. To dispell that misunderstanding we first recapitulate the pertinent electronic facts in the spirit of [MR2], but more precisely. See also [SM]. It has been shown [Do], by a simple theoretical derivation and an experimental study of actual chip layouts, that the wire length distribution in layouts

tends to satisfy $f(i) = c / i^{\lambda}$ $(1 \leq i \leq L)$ and $f(i) \approx 0$ $(L < i)$. In Section 3 we investigate the effects of the logarithmic delay assumption (with attending constant aspect ratio for wires) on layouts with these wire length distributions. While leaving free the actual topology, placement and routing in the layouts, just consideration of the wire length distribution alone shows that for some distributions the logarithmic delay assumption necessitates an exponential increase in layout area, which in practice implies exponentially long connect wires. Thus the gain in greater relative propagation speed is completely lost by the greater distances the signals have to travel.

Related work. In a companion paper [Vi] we investigate the impact of the signal propagation delay, and in particular the logarithmic delay assumption, on different circuit topologies with the same function. It appears that a circuit topology which is optimal under one delay assumption may be suboptimal under another. Systolic search trees or fast permutation networks like the Cube-Connected Cycles [PV] can be laid out to perform superior under constant or logarithmic delay assuming constant width wires. However, under logarithmic signal propagation delay with attending constant aspect ratio for the wires, they are for every layout inferior to the naive layout for a systolic Mesh network. With a constant aspect ratio for the wires, *every* layout for a complete N-node binary tree is shown to require area nonlinear in N, so *also* the H-tree layout of [MR1, MC]. This may be of independent interest. For an extensive discussion on how the propagation delay varies as a function of the wire length, and how this influences overall circuit performance with respect to pipelined computations, see [PRS]. Apart from the fact that we require a constant aspect ratio for wires to obtain logarithmic propagation delay, we also need to attach *drivers* to drive the long wires. In [MR1, MR2, MC] it is shown that such a driver needs an area proportional to the length of the wire. In [Ra] the worst-case effect of insertion of such drivers in a fixed layout is investigated under the assumption that all wires have the same unit wire width before and after insertion.

2. Electronic basics

The time it takes a minimum transistor to drive a wire of length L, width W and thickness H can be estimated as follows. The wire is assumed to have distance D_l to neighbouring layers and D_w to other wires in the same layer. If W_0 is the minimal width of a wire in the current technology, then the minimal transistor, consisting of a wire crossing, occupies area W_0^2. The total time T to drive a wire is approximated by:

$$T \approx (R_t + R_w)\, C_w \quad , \tag{1}$$

where R_t is the resistance of the minimum transistor, R_w the resistance of the wire and C_w its capacitance. The total time T can be thought of as the sum of the time T_d needed to drive a zero resistance wire of capacitance C_w, and the time $R_w\, C_w$ needed to transport the oppropriate charge from a zero resistance source. Since the resistance of a wire is proportional to its length and inversely proportional to its cross section we have:

$$R_w = \rho_w \frac{L}{WH} \quad , \tag{2}$$

where ρ_w is the resistivity of the considered wire material. The capacitance of a wire is

inversely proportional to the distance of its neighbouring wires and layers, and proportional to the area of the side facing that neighbouring layer or wire:

$$C_w = \epsilon_w L \left(\frac{H}{D_w} + \frac{W}{D_l} \right) \tag{3}$$

where ϵ_w is a proportional constant consisting of the product of the permittivity of free space and the dielectric constant of the insulating material (usually SiO_2). Thus,

$$R_w C_w = \rho_w \epsilon_w \frac{L^2}{WH} \left(\frac{H}{D_w} + \frac{W}{D_l} \right) . \tag{4}$$

This suggests a signal propagation time quadratic in L. However, the resistance R_t, of the minimum transistor, dominates in (1) for the magnitudes of L under consideration (smaller than, say, 1 meter). We can decrease that term by fitting a larger driver transistor to the wire. This transistor, in its turn, must be driven by the minimal transistor. Iterating this scheme, cf. [MC], we obtain a sequence of transistors, of which each next one is a factor α larger than the preceding one. The final transistor in the sequence should be large enough to drive the wire in a sufficiently short time. The time to drive a driver with capacitance C_2 by a driver with smaller capacitance C_1 is given by [MC]:

$$\tau \frac{C_2}{C_1} , \tag{5}$$

where τ is the time it takes a minimal transistor to charge the gate of another minimal transistor. If C_t is the capacitance of the minimal transistor then we require:

$$\#\,\text{drivers} = \log_\alpha \frac{C_w}{C_t} , \tag{6}$$

taking $T_d = \#\,\text{drivers}\,\tau\,\alpha$ time to charge the wire if it had no resistance. The capacitance of the minimum transistor is given by

$$C_t = \epsilon_t \frac{W_0^2}{D_0} , \tag{7}$$

where D_0 is the thickness of the gate insulator and ϵ_t is the product of the permittivity of free space and the dielectric constant of the gate insulator. Thus we can drive a zero resistance wire of capacitance C_w through a sequence of #drivers in time:

$$T_d = \alpha \tau \log_\alpha \frac{\epsilon_w D_0 L}{\epsilon_t W_0^2} \left(\frac{H}{D_w} + \frac{W}{D_l} \right) \tag{8}$$

From (1), (4) and (8) we obtain an expression for T.

$$T \approx \alpha \tau \log_\alpha \frac{\epsilon_w D_0 L}{\epsilon_t W_0^2} \left(\frac{H}{D_w} + \frac{W}{D_l} \right) + \rho_w \epsilon_w \frac{L^2}{WH} \left(\frac{H}{D_w} + \frac{W}{D_l} \right) \tag{9}$$

It is therefore clear that the signal propagation time heavily depends on the various dimensions and materials involved in the chip. In [MR2] it was observed that by keeping the derivatives, with respect to L, of the two terms in the righthand side of (9) balanced:

$$\frac{\alpha\tau}{L\ln\alpha} \approx 2\rho_w \epsilon_w \frac{L}{WH}\left(\frac{H}{D_w}+\frac{W}{D_l}\right) , \tag{10}$$

T grows logarithmic in L. Viz., from (9) we obtain by assumption of equality (10):

$$T \approx \frac{\alpha\tau}{\ln\alpha}\left\{\ln\left[\frac{\epsilon_w D_0 L}{\epsilon_t W_0^2}\left(\frac{H}{D_w}+\frac{W}{D_l}\right)\right]+\tfrac{1}{2}\right\} . \tag{11}$$

Having the minimum transistor drive the total wire outright, we obtain from (1), (4), (5) and (7):

$$\begin{aligned} T &\approx \tau\frac{C_w}{C_t} + C_w R_w \\ &= \left(\frac{\tau D_0}{\epsilon_t W_0^2}+\rho_w\frac{L}{WH}\right)\epsilon_w L\left(\frac{H}{D_w}+\frac{W}{D_l}\right) . \end{aligned} \tag{12}$$

The above shows that we can reduce the signal propagation time by employing new materials with more favorable characteristics, such as insulating substrates like Gallium-Arsenide and Silicon-on-Sapphire technologies [YYC] and alternative metal (W, Mo, Ti, Ta) or their silicides to replace or complement polysilicon [SM]. We can also change the size of the wires and the interwire- and interlevel separation. Thus, for the long interconnect wires usually extra layers with wider metal wires are used.

3. The expected wire length distribution for VLSI layouts, logarithmic signal propagation delay and layout area

In the sequel we make the following assumptions about the area for a VLSI layout. This area is expressed in A area units. The area unit is the square of the basic length unit which is the feature width of the underlying technology. This feature width is currently $4{\cdot}10^{-6}$ - $10{\cdot}10^{-6}$ meter, and is expected to decrease in the near future to $0.5{\cdot}10^{-6}$ - $4{\cdot}10^{-6}$ meter [SM, TR, YYC]. Assumptions like below need to be made in particular with respect to lower bound arguments.

1. The area is taken to be the area of the smallest convex region enclosing the layout times the worst-case amount of cross-over c.
2. The worst-case amount of cross-over $c>0$ is the least c such that no unit circle encloses points of more than c different edges (wires) or nodes (components or transistors).

Under the weak assumptions above, we consider the worst-case 'area $\times$ cross-over' product rather than the area. Effectively, we consider '3-dimensional' layered chips.

Let $f:\mathbb{N}\to\mathbb{N}$, connected with a VLSI layout, be a function which yields the number $f(i)$ of wires of length i in the design.

Every VLSI layout must have a constant bounded fan-in and fan-out of wires for the components (transistors). If the chip area is A, then the average maximal wire length $L_{\max}$ can be estimated by the statistical formula [SM] $L_{\max}= KA^g$ with K and g

constants which, by rule of thumb, can be set to ½ each. A reasonable assumption therefore is that the maximal wire length on a chip does not exceed

$$L_{\max} = \sqrt{A} \quad . \tag{13}$$

Consequently, the amount of wires in the layout is given by

$$\#\text{ wires} = \sum_{i=1}^{\sqrt{A}} f(i) \quad . \tag{14}$$

We now identify a common class of wire-length distributions for VLSI layouts. Firstly, it is argued that the requirement of logarithmic propagation delay favors such distributions. Secondly, other studies have shown such distributions to be likely for VLSI layouts on both theoretical and empirical grounds.

Under assumption (10) we can obtain a logarithmic signal propagation delay by, all other things being equal, maintaining:

$$L^2\left(\frac{1}{W D_w} + \frac{1}{H D_l}\right) = \text{constant} \quad , \tag{15}$$

rather than by just keeping L^2 proportional to WH as in [MR2]. Keeping the interwire distance proportional to the wire width, and the interlayer distance proportional to the wire height, we observe that if W, H and L are kept in proportion the desired logarithmic propagation delay is attained. (Note that we cannot reach this effect by keeping the wire width the same but using very 'tall' wires or vice versa.) The *aspect ratio* of a wire is the quotient of its width and length. To obtain a logarithmic signal propagation delay we thus need a fixed constant aspect ratio for all wires in the layout. In designing a high speed layout we therefore need to install drivers to drive the long wires and to design all wires with constant aspect ratio. The area taken by such a driver is linear in the length of the wire [MR2]: the minimal transistor occupies area W_0^2, the next driver in the sequence area αW_0^2, and so on ($\log_\alpha l$ terms for an l-length wire). The total driver area for an l-length wire becomes $W_0^2(l-1)/(\alpha-1)$. This area is required at the lowest silicon layer of the chip; the long interconnect wires are executed in the upper metal layers. If we double the length and width of the chip then the length of the longest wires and the area of their drivers doubles too. The area of the lowest layer, however, is quadrupled and can therefore accommodate at least double the amount of drivers. This allows us to add a new layer to place still longer wires. These longer wires come on higher levels where the wires are wider. If the wires on a level $k+1$ are β longer, wider and taller than the wires on level k then the maximal amount of wires N_k on level k satisfies $N_k = N_0\beta^{-2k}$. This suggests that the number of wires $f(i)$ of length i should decrease at least as fast as $N_0 i^{-2}$. However, in actual chip layouts the number of long wires may decrease less fast than this inverse square of the wire length; empirical wire length distributions $f(i) = \lfloor c2^{-\lambda} \rfloor$ $(1 \leq i \leq L_{\max})$ and $f(i) \approx 0$ $(i > L_{\max})$ with $1.5 < \lambda < 2$ have been reported [Do]. To achieve logarithmic propagation delay we can estimate and bound the layout area occupied by the fattened wires as follows. Let C be the amount of area of the layout occupied by *non-wire components* such as transistors. Assuming that C is also the order of magnitude of the number of basic components like transistors or logic gates in the cicuit we can reason as follows. Since the wires only serve to connect components we have $C \in O(\#\text{ wires})$ in a connected layout. The

components are assumed to have at most a limited t connections to attach wires, which we suppose to account also for the *fan-in* and *fan-out* of the interconnect wires. Therefore $C \in \Omega(\#\text{ wires})$ and consequently $C \in \Theta(\#\text{ wires})$. Since we are primarily interested in order of magnitude in the sequel, we are justified to use C interchangeably for the amount of area occupied by the non-wire components, the number of non-wire components and the number of wires. The maximal area occupied by the wires (and interwire distances) under (13) and (15) is bounded by the available area:

$$\sum_{i=1}^{\sqrt{A}} f(i)\, ai^2 \approx A - C \quad , \tag{16}$$

where a is the constant quotient of width and length (the aspect ratio) of the connect wires as required by (15). Using a simple theoretical argument and an experimental study of actual layouts [Do] has developed the following wire length distribution relationship:

$$f(i) = \lfloor ci^{-\lambda} \rfloor \ (1 \leqslant i \leqslant L_{\max}) \text{ and } f(i) \approx 0 \ (i > L_{\max}) \quad , \tag{17}$$

for a *normalization* constant c yet to be chosen. Here $L_{\max}$ is a constant related to the size of the array (rectangular chip) and the adequacy of the placement; and λ is a constant characteristic of the logic. Equation (17) is derived using "Rent's Rule" which states that the average number of terminals per complex of C elements (in units, modules, cards, gates etc.) is tC^p, where t is the number of connections per individual element and p is the Rent constant characteristic of the logic complex. The analysis goes by dividing a square array of cells into 4 equal square arrays recursively down until the individual areas are the individual elements of the original logic. On each level of the recursion the number of connections crossing boundary lines is determined using Rent's rule. This shows that $\lambda \approx 3-2p$. In [Do] experimental results are given for some actual layouts placed using a hierarchical placement program: layouts for high-speed logic were p was found to be 0.75 and a layout for a hand calculator chip with $p=0.59$. For $1 \leqslant \lambda < 3$, equation (17) is of the form of the Pareto-Levy distribution; similar laws occur in contexts like word frequencies, noise in transmission channels etc. For additional discussion see [Do]. Let furthermore the network be connected, so the maximal amount of area units C available to place the components is not greater than the number of wires plus 1. From (16) and (17) we can estimate the maximal figure for the normalization constant c. For $\lambda \neq 3$:

$$c \approx \frac{(A-C)(3-\lambda)}{a(A^{(3-\lambda)/2}-1)} \quad , \tag{18a}$$

and for $\lambda = 3$,

$$c \approx \frac{2(A-C)}{a \log A} \quad . \tag{18b}$$

Consequently, for $\lambda \neq 1$ & $\lambda \neq 3$ by (14):

$$C \approx \sum_{i=1}^{\sqrt{A}} f(i) \approx \frac{(A-C)(3-\lambda)(A^{(1-\lambda)/2}-1)}{a(1-\lambda)(A^{(3-\lambda)/2}-1)} \quad . \tag{19a}$$

and for $\lambda=3$,

$$C \approx \frac{(A-C)(A-1)}{aA \log A} \quad . \tag{19b}$$

For $\lambda=1$,

$$C \approx \sum_{i=1}^{\sqrt{A}} f(i) \approx \frac{A-C}{a(A-1)} \log A \quad . \tag{19c}$$

(Note: for $\lambda<1$ we obtain $c<1$, resulting in $f(i)\approx 0$ also for small i, and C a small constant.) From the above analysis follows:

Lemma 1. *Let $f(i)$ be the wire length distribution function of a chip layout with area A. Let the signal propagation delay be logarithmic and let therefore all wires in the layout have the same aspect ratio. If $f(i) = \lfloor c / i \rfloor$ for some constant c then the total number of wires, and similarly the total number of gates and transistors, is* $O(\log A)$.

Since the number of components bounds the number of bits manipulated in each computation step, Lemma 1 tells us that this number is very much layout dependent, and depends in particular on the distribution of wire lengths. Consequently, even if the area A is polynomial in the binary size N of a problem, under a layout and signal propagation delay as in the lemma, the execution time, and also the period for systolic computation (cf. [BPP, Sa, Th1]), will be $\Omega(N)$ since it takes at least $N / \log N$ stages just to scan all N bits and for each such stage the delay in the longest wires is $\Omega(\log N)$. In some designs [Vi], however, the number of long wires decreases faster with the wire length. For such f the series in (16) converges also faster, and the maximal number of wires, and similarly the maximal number of components, may rise to order A (the area expressed in area units), under the logarithmic propagation delay requirement notwithstanding the attendant constant aspect ratio for wires.

The constant wire width assumption. For comparison we give the analogous analysis with above under the constant wire width assumption. Then equations (13) - (14) stay the same but equation (16) becomes

$$\sum_{i=1}^{\sqrt{A}} f(i)\, i \approx A - C \quad . \tag{20}$$

Thus, for $f(i) = \lfloor ci^{-\lambda} \rfloor$ $(1 \leqslant i \leqslant \sqrt{A})$ and $f(i) \approx 0$ $(i > \sqrt{A})$ and with A, C and c as above we obtain the following relations. For $\lambda=1$:

$$c \approx \frac{A-C}{\sqrt{A}-1} \tag{21}$$

$$C \approx \frac{(A-C)\log A}{2(\sqrt{A}-1)} \quad .$$

For $\lambda \neq 1$ & $\lambda \neq 2$:

$$c \approx \frac{(2-\lambda)(A-C)}{A^{(2-\lambda)/2}-1} \tag{22}$$

$$C \approx \frac{(2-\lambda)(A-C)(A^{(1-\lambda)/2}-1)}{(1-\lambda)(A^{(2-\lambda)/2}-1)} .$$

For $\lambda=2$:

$$c \approx \frac{2(A-C)}{\log A} \qquad (23)$$

$$C \approx \frac{2(A-C)(\sqrt{A}-1)}{\sqrt{A}\log A} .$$

(Note: for $\lambda<0$ we obtain $c<1$.) For $\lambda>0$ we have $C \in \Omega(\sqrt{A})$. Thus:

Lemma 2. *Let $f(i)$ be the wire length distribution function of a chip layout with area A under the constant wire width assumption. If $f(i) = \lfloor c / i^{\lambda} \rfloor$ $(1 \leqslant i \leqslant \sqrt{A})$ and $f(i) \approx 0$ $(i > \sqrt{A})$ for a constant c then the maximal feasible number of wires in the layout, and similarly the maximal number of gates and transistors in the layout, is $\Omega(\sqrt{A})$ for all $\lambda \geqslant 0$.*

Recall that the quotient of the length and width of a wire is its *aspect ratio*. By the previous analysis, considering just the wire length distribution while leaving free the actual circuit topology, placement and routing in the layouts, attaining a logarithmic signal propagation delay by changing constant wire width to constant aspect ratio for all wires in a layout can carry a surprisingly severe penalty.

Theorem. *Let the original layout area be A and the original amount of wires in the layout be C. For the wire length distribution $f(i) = \lfloor ci^{-1} \rfloor$ for $1 \leqslant i \leqslant \sqrt{A}$ and $f(i) \approx 0$ for $i > \sqrt{A}$, the change from constant wire width to wires with a constant aspect ratio has the following effect.*

(i) *Retaining the original amount of wires C and the original wire length distribution relative to the layout area, that is, $f(i) = \lfloor c' i^{-1} \rfloor$ for $1 \leqslant i \leqslant \sqrt{A'}$ and $f(i) \approx 0$ for $i > \sqrt{A'}$ with the normalization constant c' set to its maximal value, exponentially increases the required area A' over the original area A.*

(ii) *Retaining the original area A and the original wire length distribution, that is, $f(i) = \lfloor c' i^{-1} \rfloor$ for $1 \leqslant i \leqslant \sqrt{A}$ and $f(i) \approx 0$ for $i > \sqrt{A}$ with the normalization constant c' set to its maximum value, reduces the maximal amount of wires in the layout, and therefore the proportionate amount of useful components like gates and transistors, to $C' \in O(\log C)$.*

(iii) *Retaining the original amount of wires C (or logic components) and the original area A requires a thorough change of wire length distribution to $f'(i) = \lfloor c' i^{-\lambda} \rfloor$ for $1 \leqslant i \leqslant \sqrt{A}$ and $f(i) \approx 0$ for $i > \sqrt{A}$ and the normalization constant c' set to its maximum value. Each $\lambda \geqslant 2+\epsilon$ for some small $\epsilon>0$ depending only on A and a suffices. For a given network topology this entails a placement and routing of the layout which may well be impossible in many cases.*

Proof. Since we assume the circuit to be connected we have $A > A - C > A/2$ in

the various equations. We also assume $A \gg 1$.

(i). Equate expression (21) for C with expression (19c) for C, with A' substituted for A in the latter. This yields $\log A' \in \Omega(\sqrt{A})$.

(ii). Substitute C' for C in equation (21) and express C' in terms of C by eliminating A from the resulting equation and (19c).

(iii). Equate expression (21) for C with expression (19a) for C (expressions (19b) and (19c) contradict (21)). The terms $(A - C)$ on both sides cancel each other. Solving λ yields $\lambda = 2 + \epsilon(A, a) > 2$ with $\epsilon(A, a) \to 0$ for $A \to \infty$ and a constant. Every distribution with exponent equal or larger than this λ suffices. □

We observe that in case (i) of the Theorem the wires get so long that the logarithmic propagation delay turns out to yield about the same absolute time delay as in the original wires. In case (ii) of the Theorem matters are probably as bad because the bit capacity of the chip has been logarithmically reduced. Finally, in case (iii) of the Theorem the subject circuit topology may not have a layout with the required wire length distribution.

For values of the exponent $\lambda > 1$ in the wire length distribution the analog of the Theorem holds with polynomial relationships in cases (i) and (ii). For $\lambda > 3$, the the number of long wires decreases so fast with the wire length that it suffices that $A' \in \Theta(A)$ in (i), $C' \in \Theta(C)$ in (ii) and nearly the same wire length distribution function suffices in (iii). However, $\lambda > 3$ implies a negative Rent constant p since $\lambda \approx 3 - 2p$ in [Do], and therefore layouts which do not satisfy Rent's Rule. The reader is invited to analyse the relations for different values of λ. We look at one more case, the wire length distribution with $\lambda = 2$, which is interesting because it is associated with hierarchical designs. For $\lambda = 2$, the different parts of the Theorem yield the following:

(i) $A' \in \Omega(A^2 / \log^2 A)$.

(ii) $C' \in O(\sqrt{C \log C})$.

(iii) This requires a change to a new distribution function $f'(i) = \lfloor c' i^{-\lambda} \rfloor$ $(1 \leqslant i \leqslant \sqrt{A})$ and $f'(i) \approx 0$ $(i > \sqrt{A})$ with the new normalization constant c' set to its maximum value. Each $\lambda \geqslant 3$ suffices.

Network topologies, which can be realized with constant width wires, may not be realizable at all on a multilevel Manhattan grid geometry with all wires having the same aspect ratio. Even for network topologies which do have a layout with a constant aspect ratio for the wires, the Theorem shows that the increase in area can be so much that the amplification in length of the wires will nullify (or worse) the increase in speed due to a change from linear or square propagation delay to a logarithmic one. It therefore appears that but circuits with proper topologies, for which there are layouts with the considered wire length distributions with relative large λ, are proper candidates for an improvement of speed by a logarithmic signal propagation delay. This matter is treated in [Vi].

The inspiration to investigate the subjects in this paper and [Vi] was supplied by a lecture by Martin Rem on the subject of [MR2] in the 1980/1981 Computer Science Colloquium on "Complexity and Algorithms" at the (former) *Mathematisch Centrum.*

REFERENCES

BPP Bilardi, G., M. Pracchi & F.P. Preparata, A critique and an appraisal of VLSI models of computation. In: Kung, H.T., B. Sproull & G. Steele (Eds.), *VLSI Systems and Computations.* Springer verlag, Berlin, 1981, 81-88.

BK1 Brent, R.P. & H.T. Kung, The chip complexity of binary arithmetic, Proc. 12th ACM Symposium on Theory of Computing, 1980, 190-200.

BK2 Brent, R.P. & H.T. Kung, The area-time complexity of binary multiplications, *J. ACM* **28** , 521-534.

CM Chazelle, B. & L. Monier, A model of computation for VLSI with related complexity results, Proc. 13th ACM Symposium on Theory of Computing, 1981, 318-325.

Do Donat, W.E., Wire length distribution for placement of computer logic, *IBM J. Res. Develop.* **25** (1981) 152 - 155.

EWZLEden, R., B.M. Welch, R. Zucca & S.I. Long, The prospects for ultra-high speed VLSI GaAs digital logic, *IEEE Journal of Solid-State Circuits,* SC-14 (1982) 221 - 239.

MC Mead, C. & L. Conway, *Introduction to VLSI Systems.* Addisson-Wesley, 1980.

MR1 Mead, C. & M. Rem, Cost and Performance of VLSI Computing structures, *IEEE J. on Solid State Circuits,* Vol. SC-14, 1979, 455 - 462.

MR2 Mead, C. & M. Rem, Minimum propagation delays in VLSI, *IEEE J. on Solid State Circuits,* vol. SC-17 (1982) 773 - 775. Correction: *Ibid,* (1984) 162.

MS Morkoc, H., & P.M. Solomon, The HEMT: a superfast transistor, *IEEE Spectrum,* Februari 1984, 28 - 35.

PRS Paterson, M.S., W.L. Ruzzo & L. Snyder, Bounds on the minimax edge length for complete binary trees. 11th ACM Symposium on Theory of Computing, 1981, 293 - 299.

PV Preparata, F., and J. Vuillemin, The Cube-Connected Cycles: a versatile network for parallel computation, Proceedings 20th IEEE Annual Symposium on Foundations of Computation, 1979, 140 - 147.

Ra Ramachandran, V., On driving many long lines in a VLSI layout, Proceedings 23rd IEEE Symposium on Foundations of Computation, 1982, 369 - 378.

SM Saraswat, K.C. & F. Mohammadi, Effect of scaling of interconnections on the time delay of VLSI circuits, *IEEE J. of Solid State Circuits,* vol. SC-17, 275-280, 1982.

Sa Savage, J., Planar circuit complexity and performance of VLSI algorithms, Proc. CMU Conference on VLSI Systems and Computations (H.T. Kung et. al. Eds.), Computer Science Press, 1981, 61-68.

Se Seitz, C.L., Ensemble architectures for VLSI - A survey and taxonomy, Proc. of MIT Conference on Advanced Research in VLSI (P. Penfield, Jr., Ed.), Artech House, 1982, 130-132.

Th1 Thompson, C.D., *A Complexity Theory for VLSI.* Ph.D. Thesis, Dept. of Computer Science, Carnegie-Mellon University, 1980.

Th2 Thompson, C.D., Fourier transforms in VLSI, *IEEE Transactions on Computing,* (1983)

TR Thompson, C.D., & P. Raghavan, On estimating the performance of VLSI circuits. Tech. Rep. UCB/CSD 84/138, Computer Science Division (EECS), University of California, Berkeley, September 1983.

Vi Vitányi, P.M.B., Signal propagation delay, circuit topology and the efficiency of VLSI circuits. In preparation.

Vu Vuillemin, J., A combinatorial limit to the computing power of VLSI circuits, Proc. 21th IEEE Symposium on the Foundations of Computer Science, 1980, 294-300.

YYC Yuan, H.-T., Y.-T. Lin & S.-Y. Chiang, Properties of silicon, sapphire, and semi-insulating gallium arsenide substrates, *IEEE J. of Solid State Circuits,* Vol. SC-17, 1982, 269-274.