

Randomized Routing in Gate-Arrays

Prabhakar Raghavan

Clark D. Thompson

Computer Science Division

573 Evans Hall

U.C. Berkeley, CA 94720

ABSTRACT

A stochastic model for analyzing the performance of randomized algorithms for routing gate-arrays is presented; our model is based on an empirical observation known as Rent's Rule. Using the model, we analyze the space requirement of a wiring technique that only uses one-bend routes. We show how the technique can be extended to a case where several wiring layers are available, with near-optimal saving in area. As a by-product, we obtain a random procedure for converting a two-layer gate-array routing to a many-layer routing while reducing area efficiently. Within our model, we also show that the one-bend strategy is sub-optimal in terms of space requirement. However, we also show that the optimal strategy is not significantly superior to the random one-bend strategy.

September 27, 1984

Supported by National Science Foundation through its Computer Engineering Program Grant # ECS-8406408 and by Semiconductor Research Corporation Grant SRC 1-44247-52055 (Prabhakar Raghavan).



Randomized Routing in Gate-Arrays

Prabhakar Raghavan
Clark D. Thompson

Computer Science Division
573 Evans Hall
U.C. Berkeley, CA 94720

1. Introduction

Gate-arrays are finding increasing use as a framework for semi-custom VLSI design [1]. A gate-array is an $n \times n$ array of gates (Fig. 1a) on a chip; a large number of such chips are fabricated by a manufacturer. A customer who wishes to build a logic circuit decides on a mapping of the gates in her circuit onto the gates in the array. She specifies the interconnections to be made between the gates in the array in order to realize the circuit she has in mind (Fig. 1b). The manufacturer then makes the necessary connections using some routing algorithm, producing a wired chip that might look like Fig. 1c. Throughout this paper, we will consider circuits consisting entirely of *two-terminal nets*, i.e. each wire in the circuit connects exactly two gates.

Considerable effort has been directed towards two classic problems concerning gate-arrays:

1. How much space must be allowed between the gates on an array so as to be able to route a given circuit? In our model, each gate is represented by a square as in Fig. 1a. Each wire is assumed to take up a unit space; the number of wires passing over a square (gate) is the *channel-width* associated with the gate. The channel width is a quantity that must be decided beforehand by the manufacturer, without knowing the exact pattern of interconnections specified by the customer.
2. Given a placement of the components of a circuit on the gates in an array, how best does one go about making the specified connections? This is an instance of the *routing problem*. Fig. 1c is a solution to the instance depicted by Fig. 1b. Notice that all the routes in this solution are wires having only one bend; such a solution is called a *one-bend routing*.

Donath *et al.* first studied the former problem and introduced the use of *Rent's Rule*, an empirical observation concerning the average number of connections that leave a planar region on the array that includes a collection of gates. A heuristic approach to channel-width estimation was developed in [5]. Feuer [3] showed the relation between Rent's Rule and the distribution of wire-lengths. El Gamal [4] provided a stochastic approach to estimating channel widths by considering wires as random walks. Karp *et al.* [7] recently showed that in the worst case (i.e. when the customer chooses an interconnection pattern that forces large channel widths), the

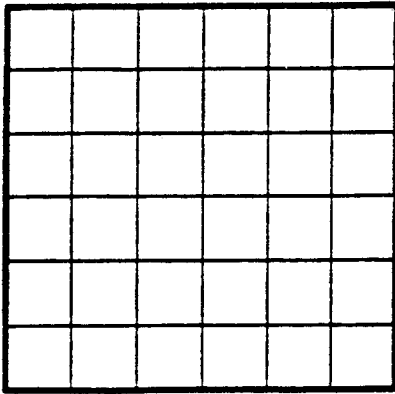


Fig 1a

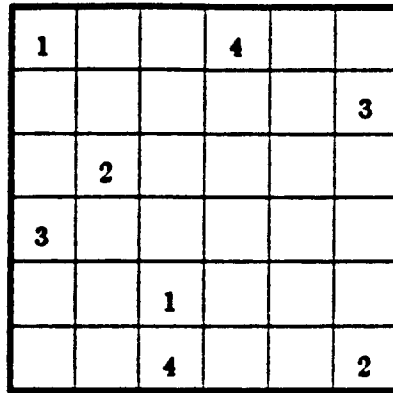


Fig 1b

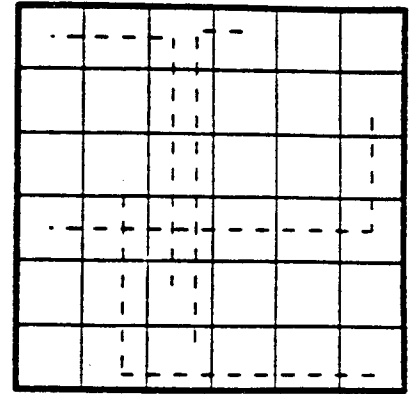


Fig 1c

routing algorithm need only consider one-bend routings.

In practice, however, the customer is hardly an adversary; the worst case is thus not very realistic (as it happens, it is extremely pessimistic). Although the placement problem in its various forms is known to be *NP-Complete* [6], it is reasonable to assume that the customer has made an intelligent attempt at placing the gates in her circuit on the array in such a fashion as to try and avoid having long wires, congestion, etc. It is thus necessary to make some assumption about the kind of interconnection patterns requested by the customer. Rent's Rule is such an assumption.

Rent's Rule states that a region of the array containing C gates with b connections to each has an average of $b.C^p$ connections leaving the region. The Rent exponent p lies in the interval $[0.5, 1]$, and characterizes the type of logic being implemented; highly serial logic, for instance, would typically have a low value of p . In this form, the rule is difficult to utilize to obtain interesting results; we therefore use the connection distribution function derived by Feuer. Feuer showed that $q(r)$, the probability of the existence of a wire between two gates separated by a (Manhattan) distance r is given by

$$q(r) = a \cdot r^{-(1+2p)}$$

where a is a normalizing factor.

In the next section we introduce the model we will use, and state some straightforward implications of the model. Section 3 contains general results for randomized routing; for the two-layer case our results take on a form similar to those of El Gamal. We then show that with k layers, random assignment is close to optimal. This is important in the light of the use of 6-8 layers in recently fabricated chips and proposals for constructing chips with even more layers of interconnect. In section 4, we show that random one-bend routing is a good means of reducing the wiring-space required for a high probability of success in wiring. In doing so, we study the 'optimal' randomized routing strategy possible under our model, making comparisons with the

random one-bend strategy as appropriate.

3. The Stochastic Model

We now formally present the model we will use to prove our results.

- (1) A *two-dimensional array* is an infinite array of gates in two dimensions. Each gate is a square of fixed width W (measured in wire-widths). A *gate-array* is a square region of $n \times n$ gates in the two-dimensional array.
- (2) In an instance of the *interconnection problem*, each gate in the two-dimensional array emits one wire; the probability that a gate at a Manhattan distance r is the destination of the wire is $q(r)$. The $q(r)$ function is defined as

$$q(r) = a \cdot r^{-1+2p} \quad (2.1)$$

for $1 \leq r \leq L_{\max}$, where L_{\max} is the maximum wire length for the gate-array. Thus for a square $n \times n$ array, $L_{\max} \leq 2n$ (We assume that every wire follows a *minimum distance* route to its destination). This definition automatically provides us with a distribution on the wire-lengths. Note that we have restricted our attention to two-terminal nets; we will not consider multi-terminal nets in this paper.

Note that in our model, only one wire leaves a gate, whereas several (or none) arrive at it. We interpret a departing wire as a fan-in lead; we are thus restricting ourselves to the case where fan-in is one. All our results on channel-width scale up by a factor of b when there are b fan-in leads to each gate.

- (3) To begin with, we will allow the use of two layers of metallization for interconnection. We will generalize this to k layers in the latter half of section 3. We assume further that one of the two layers will be used for horizontal wire runs, and the other for vertical runs.
- (4) One horizontal and one vertical channel are associated with each square. A channel is thus a $W \times W$ region on one of the two layers. A vertical channel associated with a gate G is one that is used to accommodate all wires passing over G in the vertical direction; a horizontal channel is defined similarly. All channels through which a wire passes are charged one *track* for that wire; this includes the gates in which a wire begins/terminates (if a wire emitted by a gate leaves it in the vertical direction, the vertical channel is charged). The number of tracks required by a channel for an instance of the interconnection problem is the *channel space requirement* (physically, the width) of that channel. The maximum channel space requirement taken over all the channels in the gate-array is the *space requirement* of that instance of the interconnection problem.

- (5) A solution to an instance of the interconnection problem is an assignment of routes to wires such that the space requirement under that assignment does not exceed W .

In this paper, we study the performance of randomized algorithms that solve the interconnection problem. A *randomized routing strategy* is one that flips coins or casts dice to decide the choice of route for a particular wire; no attempt is made to adapt to the congestion prevailing on the chip at the time a wire is routed.

Each wire is thus routed independently of all others. Further, the performance of the strategy depends on the particular sequence of coin-flips in a given instance; our analysis averages the algorithm's performance over all coin-flip sequences. If a solution exists for a given instance of the interconnection problem under that strategy, then there exists a sequence of coin-flips (and therefore a non-zero probability) that the solution is found. An example of a randomized routing strategy is the *random one-bend strategy*. Each wire is routed according to one of the two possible one-bend (L-shaped) routes; either of them is equally likely.

To begin with, we will study the problem from the following standpoint. Given a randomized routing strategy, we wish to find a width W such that with probability $1-\epsilon$ the space requirement of a random instance of the interconnection problem does not exceed W , for some positive constant ϵ .

3. Routing Strategies with a High Probability of Success

In this section, the fundamental issue we will address is the prediction of the gate-width W . Our prediction is probabilistically defined: the probability of the channel space requirement of any channel exceeding W is bounded above by $\frac{\epsilon}{2n^2}$, where ϵ is a positive constant. Since there are $2n^2$ channels in the gate-array, we are then assured of being able to route the particular instance of the interconnection problem with probability at least equal to $1-\epsilon$.

3.1. The Two-Layer Case

El Gamal provides a similar result under a slightly different model. His results may be modified to obtain the ones here. Our motivation is to provide a general framework to facilitate the analysis of the k -layer case that follows, and to lay the foundation for studying a class of random routing schemes.

We will for the moment concentrate on the randomized one-bend strategy. Thus each net is assigned one of two "L-bend" (one-turn) routings picked at random, independently of the assignments of all other nets. In the next section, we will compare the one-bend strategy with more general strategies that make use of other routing patterns. In our analysis, we make use of the following powerful theorem due to Bernstein [10]. M denotes means and σ deviations.

THEOREM (Bernstein): If $\Psi_1, \Psi_2, \dots, \Psi_n$ are completely independent random variables such that $E(\Psi_k) = M_k$, $\sigma(\Psi_k) = \sigma_k$ exist and $|\Psi_k - M_k| \leq K$ ($k = 1, 2, \dots, n$), then for $\Psi = \Psi_1 + \Psi_2 + \dots + \Psi_n$ we have

$$P(|\Psi - M| \geq \mu \cdot \sigma) \leq 2 \exp \left[- \frac{\mu^2}{2 \left[1 + \frac{\mu K}{2\sigma} \right]^2} \right] \quad (3.1)$$

where

$$M = \sum_{k=1}^n M_k$$

and

$$\sigma = \left[\sum_{k=1}^n \sigma_k^2 \right]^{1/2}$$

where μ is a positive number such that $\mu \leq \frac{\sigma}{K}$.

Let us focus our attention on any one channel in the array. In our case, each of the random variables Ψ_i is an indicator for the event that the wire emitted by some gate in the array passes through the channel under consideration. The sum of these variables is then the channel space requirement of that channel. It is easy to see that the average value of Ψ is one-half the average wire-length; this holds under any choice of radially-symmetric wire-length distribution. Moreover, the passage of a wire emitted by some gate through the channel of interest can be viewed as a Bernoulli trial. Since we then have a series of independent Bernoulli trials,

$$\sigma_{(2)}^2 \leq \bar{V}_{(2)} \quad (3.2)$$

Here the subscript (2) denotes the fact that we are considering two-layer routing, and is not to be confused with the Ψ_2 of Bernstein's Theorem. This is a very loose bound; we will consider tighter ones in the next section. Applying Bernstein's Theorem, we find that the following choice of gate-width ensures that no channel has channel space requirement exceeding $W_{(2)}$ with probability of more than $\frac{\epsilon}{2n^2}$.

$$W_{(2)} = \bar{V}_{(2)} + \sigma_{(2)} \cdot \ln \frac{2n^2}{\epsilon} \quad (3.3)$$

The above result is independent of the wire-length distribution. For the wire-length distribution derived by Feuer, $\bar{V}_{(2)}$ grows as n^{2p-1} , so that the latter term (involving standard deviation) is asymptotically smaller.

3.2. The k-Layer Case

The above analysis may readily be extended to k-layer routing. We study this problem from two standpoints. First, given an instance of the interconnection problem, we wish to estimate the gate-width $W_{(k)}$ such that a solution can be found with probability at least equal to $1-\epsilon$. Second, given a two-layer routing of some instance of the interconnection problem, we wish to convert it into a k-layer routing so that chip area is shrunk to the extent possible. We begin by giving a brief account of our strategy for wiring gate-arrays on k layers, followed by the randomized procedures for accomplishing each of the above goals.

3.2.1. A General Methodology for k-Layer Routing

We assume k is even. Our strategy for k-layer routing consists of partitioning the k layers available into $\frac{k}{2}$ groups of two layers each. If the layers be numbered $1, 2, \dots, k$, then we group layers $2i-1$ and $2i$ together, $1 \leq i \leq \frac{k}{2}$. The odd-numbered layers $1, 3, \dots, \frac{k}{2} - 1$ will be used for horizontal wiring runs, and the even-numbered layers for vertical runs of wiring.

A net will be embedded entirely in one of the above groups; there is thus no "communication" between groups. We are now left with a case where each gate has $\frac{k}{2}$ horizontal and $\frac{k}{2}$ vertical channels over it; we wish to estimate the k -layer gate-width $W_{(k)}$ such that none of these is congested. Notice that there are now kn^2 channels in the gate-array.

Thompson [11] first noted that in going from two to k layers, we cannot shrink each dimension by more than a factor of k ; we thus cannot hope to achieve a reduction in area by more than a factor of k^2 . This observation stems from the fact that if we actually did shrink both dimensions by more than a factor of k , the resultant k-layer embedding could be "unwrapped" to a new two-layer embedding smaller than the original one (the details of this are fairly straightforward).

3.2.2. A Randomised Strategy for k-layer routing

We now extend the randomized one-bend strategy mentioned earlier to k layers. For each net, instead of flipping a coin to decide between one of two L-shaped routes, we now cast a k -faced unbiased die to decide between one of the k possible L-shaped routes (for a wire traveling "northeast" we have $\frac{k}{2}$ routes that first go north and then east, and $\frac{k}{2}$ that go east first and then north). To ensure a $1-\epsilon$ probability of success, we must make sure that each channel has a probability of congestion that does not exceed $\frac{\epsilon}{kn^2}$. Since the probability that any group (of

wire-layers) be used to route a wire is $\frac{1}{k}$, the average value of Ψ is now down by a factor of $\frac{2}{k}$; let us denote this by $\bar{\Psi}_{(k)}$. The variance (denoted by σ_k^2) is still upper bounded by this mean, so that $\sigma_{(k)} \approx \sigma_{(2)} \cdot \left(\frac{2}{k}\right)^{1/2}$. The required gate-width (by Bernstein's Theorem) is then

$$W_{(k)} = \bar{\Psi}_{(k)} + \sigma_{(k)} \cdot \ln \frac{kn^2}{\epsilon} \quad (3.4)$$

To put these results in perspective, we first recall that the gate-width $W_{(2)}$ derived for the two-layer case exceeded the average by only an asymptotically small quantity. We are now able to shrink each dimension by a factor of almost $\frac{k}{2}$.

We should of course bear in mind that for the small values of k currently in use, a $\frac{k}{2}$ saving does not appear very attractive; it is conceivable that for a specific instance, a good design engineer could (albeit laboriously) trim the dimensions down by close to a factor of k . However, proposals for more layers are being investigated both theoretically [8] and experimentally [2], and as n and k grow larger, we believe that manual layer assignment can do little better than random assignment. In essence, the law of large numbers can be put to advantage as designs grow large.

Assuming (reasonably) that an unbiased k -faced die can be implemented by means of $\Theta(\log_2 k)$ unbiased coin flips, the k -layer random one-bend strategy requires about $\log_2 k$ times as much computation as the two-layer strategy.

3.2.3. Converting Two-Layer Routings to k -Layer Routings

The above strategy for k -layer routing can be viewed as a means of converting any given two-layer (or k' -layer for $k' < k$) routing to a k -layer routing. Given a two-layer routing $R_{(2)}$ produced by any algorithm, we cast a $\frac{k}{2}$ -faced die for each of the nets in $R_{(2)}$. The outcome of this die decides which of the $\frac{k}{2}$ layer-groups the net will use in the k -layer routing $R_{(k)}$.

There is no restriction on the style of routing in $R_{(2)}$; the nets could have several bends and, more importantly, several terminals (recall that in the preceding subsections we were constrained to two-terminal nets). We assume, reasonably, that a net does not pass through the same channel twice in $R_{(2)}$. As before, $R_{(k)}$ is smaller than $R_{(2)}$ by almost a factor of $\frac{k}{2}$ in each dimension. If $W_{(2)}$ were the gate-width in $R_{(2)}$, then the k -layer width that will suffice to restrict the probability of failure to ϵ is given by

$$W_{(k)} = \frac{2}{k} \cdot W_{(2)} + \left[\frac{2}{k} \cdot W_{(2)} \right]^{1/2} \cdot \ln \frac{kn^2}{\epsilon} \quad (3.5)$$

4. On the Relative Merits of Various Randomized Strategies

In the first part of the previous section, we considered the random one-bend strategy on two layers. We used the mean of the random variable Ψ to bound the variance of the number of wires in a channel. In this section, we will investigate the tightness of this bound. We only consider two-layer routings here, although similar results can be established for k layers.

There are two reasons to examine the actual variance. In equation (3.3), the two-layer channel-width $W_{(2)}$ is expressed as the sum of an "average" term $\bar{\Psi}_{(2)}$ and a "standard deviation" term $\sigma_{(2)} \ln\left(\frac{2n^2}{\epsilon}\right)$. Although the latter term is asymptotically smaller, it is quite large for present values of n . To illustrate this, consider a case where $p = \frac{2}{3}$, $n = 30$ and we desire $\epsilon = 0.1$. These are very typical values, and the standard-deviation term in $W_{(2)}$ (about 20) actually exceeds the average term (about 3). It will thus require several generations of gate-array development before they become large enough for the first term to dominate.

The second reason becomes relevant in the light of the first. In Section 3, we used the bound $\sigma_{(2)}^2 \leq \bar{\Psi}_{(2)}$ for randomized one-bend routing. In fact, this bound holds irrespective of the routing strategy used. The first (average) term of equation (3.3) is independent of the routing strategy (for all minimum distance strategies), but we can attempt to find a strategy that minimizes the standard deviation and thus the second term. For $p = \frac{2}{3}$ and $\epsilon = 0.1$ as before, such a reduction would become particularly important at values of n around 200 (for larger values of p , at smaller values of n). We would therefore like to know whether there exists an optimal strategy whose variance is asymptotically smaller than that of the one-bend strategy.

In this section, we will show that the one-bend strategy is sub-optimal in this respect, but that the optimal strategy cannot do much better. We begin by formalizing the notion of a *uniform routing strategy*; informally, this is the class of random routing strategies where the choice of paths available to a wire (and the probabilities assigned to each path) is independent of the gate it emanates from. We do not restrict ourselves to the Feuer wire-length distribution here; consider a general distribution where the probability that a wire be of length r is $p(r)$.

Figure 2 shows a typical gate G that we will study. C is a *Manhattan Circle* of radius L_{\max} ; it contains all gates at a Manhattan (or L_1) distance of L_{\max} or less from G . The wire emitted by each of these gates can potentially pass through either channel over G (we are only considering minimum-distance routings). Define a co-ordinate system $[i, j]$ where i and j are integers denoting respectively the x - and y - displacements of each gate from G . The gates inside the circle satisfy

$$|i| + |j| \leq L_{\max} \quad (4.1)$$

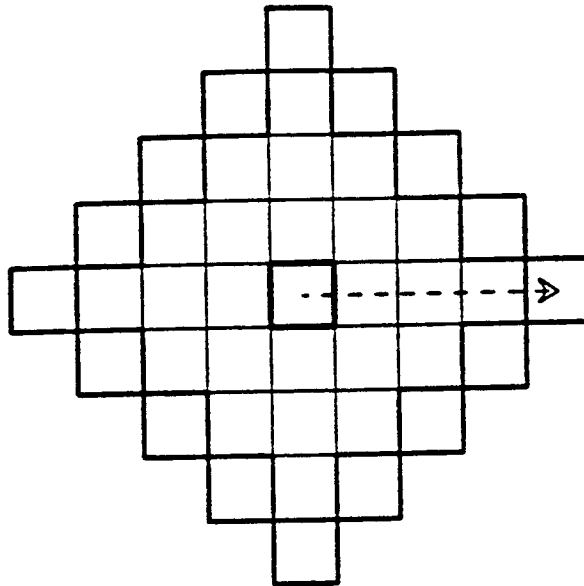


Fig 2

For each such gate, we define P_{ij} to be the probability that the wire emitted by gate $[i, j]$ uses up a track in either the vertical or the horizontal channel of G (the notion of "using up a track" is defined in assumption 4 in section 2). We place some additional restrictions on the P_{ij} (and therefore routing strategies) we will allow.

- (1) *Reflexivity*: The wire emanating from G passes through either channel over the gate at $[i, j]$ with probability P_{ij} .
- (2) *Symmetry*: We ensure that the strategy behaves in the same manner in all directions by imposing the following restriction:

$$P_{ij} = P_{-i, j}$$

It follows from these two restrictions that

$$P_{ij} = P_{-i, j} = P_{i, -j} = P_{-i, -j} \tag{4.2}$$

The following key lemma applies to all uniform random routing strategies.

LEMMA 4.1: Every uniform random routing strategy satisfies the equality:

$$\sum_{|i|+|j|=r} P_{ij} = P(r), \quad 1 \leq r \leq L_{\max} \tag{4.3}$$

where $P(r)$ depends only on the distance r and not on the strategy itself.

PROOF: The proof is simple, and follows from the reflexivity assumption that P_{ij} is the same as the probability that the wire emitted by G passes through the gate at $[i, j]$. Each term in the summation then becomes equal to the probability that the wire emitted by G passes through one of the gates at distance r from it. The left-hand side is thus the probability that any wire is of length $\geq r$; we denote this probability by $P(r)$. ■

In fact, we can extend the above argument to the following observation:

$$\sum_{|i|+|j| \leq L_{\max}} P_{ij} = \sum_{r=1}^{L_{\max}} \sum_{|i|+|j|=r} P_{ij} = L \quad (4.3a)$$

where L denotes the average wire length.

We impose an additional restriction to ensure that the behavior of the strategy is the same in the vertical and horizontal directions. Let P_{Hij} be the probability that the wire emitted by G passes through the horizontal channel over the gate at $[i,j]$; P_{Vij} is defined similarly for the vertical channel over $[i,j]$. The following restriction ensures uniformity in the vertical and horizontal channel space requirements.

(3) For all $1 \leq r \leq L_{\max}$, the values P_{Hij} and P_{Vij} must satisfy:

$$\sum_{|i|+|j|=r} P_{Hij} = \sum_{|i|+|j|=r} P_{Vij} \quad (4.4a)$$

and

$$\sum_{|i|+|j|=r} (P_{Hij})^2 = \sum_{|i|+|j|=r} (P_{Vij})^2 \quad (4.4b)$$

We note that

$$\sigma^2 = \sum_{|i|+|j| \leq L_{\max}} P_{Hij} (1 - P_{Hij}) = \sum_{|i|+|j| \leq L_{\max}} P_{Hij} - \sum_{|i|+|j| \leq L_{\max}} (P_{Hij})^2 \quad (4.5)$$

where σ^2 denotes the variance of the channel space requirement for the random routing strategy. In the right-hand sides of both the above equations, the first term is the same; a good strategy thus is one that maximizes the second. We note in passing that, by this criterion, a good strategy is one that has a few relatively large values of P_{ij} , while the rest are very small.

The *Hit-graph* $\Gamma(S,E)$ of a routing strategy is a directed graph where each node $v \in S$ represents one of the gates in the circle C . Thus, $|S| = 4 \cdot L_{\max} + 1$. Each node $u \in S$ has co-ordinates $[i,j]$ which are the co-ordinates of the gate it represents. An edge (u,v) is in E for $u,v \in S$ if and only if the following two conditions are satisfied: (1) If $u = [i,j]$ then $v = [i,j+1]$ or $v = [i+1,j]$; (2) For any wire emitted by G which passes through v (henceforth referred to as "the wire"), there is a non-zero probability that it passed through u . In addition, we assign to each node and each edge in Γ a real value in the interval $[0,1]$ as follows: (1) The value a_{uv} for each edge $(u,v) \in E$ is the probability that the wire will pass through u on its way to v (the wire does not necessarily terminate at v). This probability depends on both the wire-length distribution and the choices made by the routing strategy. (2) The value P_{ij} (alternatively, P_u) for each node $u = [i,j] \in S$ is the probability that the wire occupies a track in either channel over gate $u = [i,j]$. In general, this includes two terms - the probability p_{ij} that the wire terminates in $[i,j]$, plus the probability that it passes through $[i,j]$ en route to some other gate. As before, this

takes into account the effects of both the wire-length distribution and the routing strategy. Note that by reflexivity, the usage of P_{ij} here is consistent with the usage in equation (4.3).

We now list some properties of the hit-graph Γ that characterizes a routing strategy. The node representing the gate G is clearly a "root" of the graph in the sense that it is the only node that has no edges entering it. The first of the two conditions to be satisfied for the existence of an edge ensures that edges can only exist between nodes that represent adjacent gates (two gates are adjacent if the L_1 distance between them equals one). Moreover, the edges are all directed away from G ; since every gate must be reachable from G , there must exist a directed path in Γ from G to every other node in Γ . The first condition also induces a partition on the nodes of the graph - two nodes lie in the same equivalence class if they have the same L_1 norm. We will call these equivalence classes *levels*; henceforth, we will speak of two gates as being on the same level if they both have the same L_1 norm. It is evident that Γ is a sub-graph of the directed two-dimensional rectilinear lattice with the origin at $[0,0]$. The set of directed paths from G to a gate at $[i,j]$ is precisely the set of possible routes in which the wire can be embedded by the routing strategy if and when it passes through or terminates at $[i,j]$.

The values p_{ij} and a_w have to satisfy some conservation rules. Let $F(u)$ denote the set of predecessors of u in Γ ; namely, the set of nodes $\{w : (w,u) \in E\}$. Likewise, let $C(u) = \{v : (u,v) \in E\}$ be the successors of u . The probability that the wire occupies a track in $u = [i,j]$ is then

$$P_{ij} = \sum_{w \in F(u)} a_w = p_{ij} + \sum_{v \in C(u)} a_w \quad (4.6)$$

for all $|i| + |j| > 0$.

We are now ready to prove our results. In doing so, we make repeated use of the following fact:

Let Δ , p_x and p_y be non-negative real numbers, where $0 < \Delta \leq p_y \leq p_x$. Then

$$p_x^2 + p_y^2 < (p_x + \Delta)^2 + (p_y - \Delta)^2 \quad (4.7)$$

This suggests the following strategy for converting routing strategies into new ones that are at least as good. Whenever two probabilities must add up to some constant (typically, due to a conservation rule), we can augment the sum of their squares by increasing (if possible) the larger one at the expense of the smaller. As noted in equation (4.3), this would lead to an overall reduction in variance.

The following lemma states an important property of the optimal random routing strategy.

LEMMA 4.2: The hit-graph of the optimal strategy is a (spanning) tree.

PROOF: The proof depends on showing that a hit-graph Γ which is not a tree can always be improved upon in the variance-minimization sense. Let there exist two node-disjoint directed paths from $u = [i, j]$ to $v = [k, m]$, where $|i| + |j| \leq |k| + |m| \leq L_{\max}$. Moreover, let $[k, m]$ be the node of lowest level for which this is true. Clearly $[i, j]$ is then the only node that has two node-disjoint directed paths to $[k, m]$, and no node of lower level has two incoming edges. The lengths of both paths are equal to the difference $|k| + |m| - |i| - |j| = L$. Let the two paths be $[u = u_{10}, u_{11}, u_{12}, \dots, u_{1, L-1}, u_L = v]$ and $[u = u_{20}, u_{21}, u_{22}, \dots, u_{2, L-1}, u_L = v]$. Without loss of generality, let

$$\sum_{q=1}^{L-1} P_{u_q, v} \geq \sum_{q=1}^{L-1} P_{u_{2q}, v} \quad (4.8)$$

We now show how edge $(u_{2, L-1}, v)$ can be eliminated to give a new hit-graph that is at least as good as Γ . Let the value associated with the edges $(u_q, u_{t, q+1})$ be a_q , for $t \in \{1, 2\}$ and $0 \leq q \leq L-1$.

$a_{2, L-1}$ is the probability that the wire enters v through $u_{2, L-1}$; the wire has clearly passed through all the nodes u, u_{21}, u_{22}, \dots . Let

$$P'_{u_q} = P_{u_q} - a_{2, L-1}, \quad 1 \leq q \leq L-1 \quad (4.9a)$$

and

$$a'_{u_q} = a_{u_q} - a_{2, L-1}, \quad 0 \leq q \leq L-1 \quad (4.9b)$$

be the new values assigned to nodes and edges along the second path. Correspondingly, for the other path, let

$$P'_{u_q} = P_{u_q} + a_{2, L-1}, \quad 1 \leq q \leq L-1 \quad (4.10a)$$

and

$$a'_{u_q} = a_{u_q} + a_{2, L-1}, \quad 0 \leq q \leq L-1 \quad (4.10b)$$

be the new values assigned to nodes and edges along the first path.

The resultant graph (let us call it Γ') satisfies all the properties of a hit-graph, including conservation. The value $a_{2, L-1}$ has been made zero, signifying that the wire will never pass through $u_{2, L-1}$ in going to v ; in essence, the edge $(u_{2, L-1}, v)$ can be deleted. Most importantly, it represents an improvement in variance in the sense of equation (4.5); the inequality is in fact a strong one since $a_{2, L-1}$ is by definition non-zero. Repeated application of the above procedure reduces Γ to a hit-graph Γ' which is a tree with lower variance than Γ . That it is a spanning tree follows from the fact that every node in Γ' is reachable from $[0, 0]$. ■

It is worthwhile to stop and reflect on the significance of this result. For any Γ which is a tree, we may speak of the *descendants* $D(u)$ of a node $u = [i, j]$ without ambiguity. One property that is immediately evident for any Γ which is a tree is

$$P_u = p_U + \sum_{w \in D(u)} a_{uw} \quad (4.11a)$$

where

$$\sum_{w \in D(u)} a_{uw} = \sum_{[k, m] \in D(u)} p_{km} \quad (4.11b)$$

What does a tree hit-graph mean physically? The fact that Γ is a tree implies that there is exactly one directed path to every node in Γ from the root G . This in turn implies that if the wire terminates in some gate (node) $[i, j]$, there is a unique path along which it is laid out, and a unique set of intervening gates it will pass through. In essence, we leave the randomization to the connection distribution function; additional randomization by the algorithm adds to the variance of the channel space requirement. In practice, however, we cannot expect a customer to provide a "well-randomized" instance of the interconnection problem; so the algorithm must in fact perform some randomization. It is possible to produce "pathological" problem instances which can in general be wired successfully, but cause a fixed-route strategy (tree hit-graph) to fail. Such instances could be solved by an algorithm that does some randomization of its own. Nevertheless, out of purely theoretical interest, we now exhibit a fixed-route interconnection strategy that outperforms the random one-bend strategy.

The *Clockwise Pinwheel Graph* Q_{n+} is a hit-graph defined as follows:

- (1) There exist directed edges $([0, j], [0, j+1]), \forall 0 \leq j \leq n-1$.
- (2) In addition, $\forall 0 \leq i \leq n$ and $1 \leq j \leq n-i-1$, there are edges directed from $[i, j]$ to $[i+1, j]$.
- (3) For $0 \leq i, j, k, m \leq n$, if there is an edge $([i, j], [k, m])$ there also exist edges $([-i, -j], [-k, -m]), ([j, -i], [m, -k]),$ and $([-j, i], [-m, k])$.
- (4) Since Q_{n+} is a tree, the values P_U and a_{uw} have unique values that follow from the structure of the graph in an obvious fashion.

The third condition, informally, makes four copies of the first quadrant by clockwise rotation and replication in the other quadrants. The structure of the graph explains the name; Fig 3 illustrates Q_{n+} . The *Counterclockwise Pinwheel Graph* Q_{n-} may be defined similarly. We now perform the variance computation for the Clockwise Pinwheel graph Q_{n+} ; notice that at each level r , although every gate has a non-zero P_U , many of the P_{HKJ} or P_{VKJ} are zero. We compute the variance of the vertical channel space requirement by considering only the P_{VKJ} values in Q_{n+} (The Counterclockwise Pinwheel Graph Q_{n-} has the same variance).

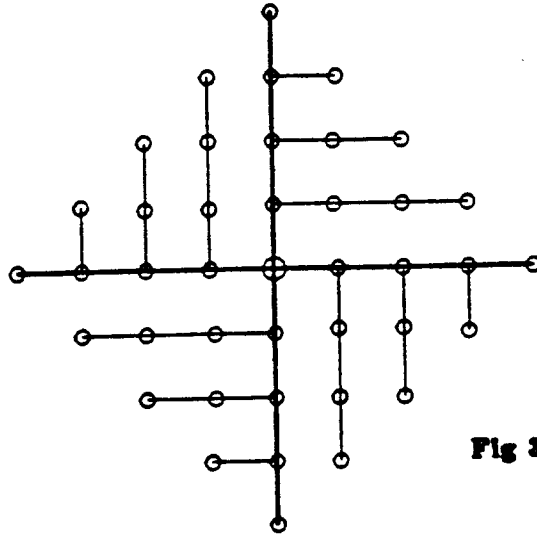


Fig 3

We evaluate the contribution $\sum_{|i|+|j|=r} P_{V_{ij}}^2, \forall 1 \leq r \leq n$. Recall that $P(r) = \sum_{k=1}^n p(k)$.

We then have

$$P_{V_{r,0}} = P_{V_{0,r}} = \sum_{k=1}^n \frac{P(k)}{4r}$$

The gates $[0,r]$ and $[0,-r]$ are counted as gates with $P_{V_{ij}} = 0$; this is because the number of bends in a wire is exactly one under the pinwheel strategy, so that each wire takes up one track in one of the channels of all gates it passes over except the gate over which it bends. The wire uses up a track in both the vertical and the horizontal channels over the gate over which it executes its turn. The fact that there is only one bend per wire means that there is an average of only one extra track in each channel over each gate when averaged over all gates, due to this phenomenon. We therefore use the convention that a gate has non-zero $P_{V_{ij}}$ if the wire from G enters it vertically; a similar convention is used for $P_{H_{ij}}$.

In addition, there are $2(r-1)$ gates each of which has $P_{V_{ij}} = \frac{P(r)}{4r}$. From these, it is clear that for the Clockwise Pinwheel graph Q_{n+} has

$$\sum_{|i|+|j|=r} P_{V_{ij}}^2 = \frac{1}{8r^2} \left[\sum_{k=1}^n P(k) \right]^2 + (r-1) \frac{P(r)}{8r^2} \quad (4.12)$$

A similar computation can be performed for the random one-bend strategy to arrive at:

$$\sum_{|i|+|j|=r} P_{V_{ij}}^2 = \frac{1}{8r^2} \left[\sum_{k=1}^n P(k) \right]^2 + (r-1) \frac{P(r)}{16r^2} \quad (4.13)$$

It is evident that the sum of the squares of P_{ij} for the clockwise pinwheel strategy exceeds the corresponding figure for the random one-bend strategy. The variance of the clockwise

pinwheel strategy (and therefore the space requirement) is thus lower than that of the random one-bend strategy. It is worth noting that the hit-graph for the random one-bend strategy is in fact the superposition of a clockwise pinwheel graph with a counterclockwise pinwheel graph. It is thus possible to generalize our 'unbiased' random one-bend strategy into a family of biased strategies where the route for an interconnection is chosen from the 'clockwise' pinwheel graph with some probability $1-\epsilon_0$, and from the 'counterclockwise' pinwheel graph with probability ϵ_0 , for some positive $\epsilon_0 < 1$. As long as $\epsilon_0 \neq 0.5$, the variance of such a biased random one-bend strategy is lower than that of the unbiased version.

At this juncture, it is worth investigating the seriousness of the variance problem. It is clear that our class of strategies admits many that are superior to the unbiased random one-bend strategy, even though they may be less likely to find a solution for some bad instances of the problem. Specifically, we address the following question: can there exist an optimal hit-graph that is asymptotically superior to any of the above strategies in the minimum variance sense? We return to equation (4.5) for the answer.

$$\sigma^2 = \sum_{|i|+|j| \leq L_{\max}} P_{H\bar{H}A} - \sum_{|i|+|j| \leq L_{\max}} (P_{H\bar{H}A})^2 \quad (4.5)$$

$$\geq \sum_{\bar{r}} \frac{P(r)}{2} - \sum_{\bar{r}} \left(\frac{P(r)}{2}\right)^2 \quad (4.14)$$

since the second term on the right-hand side of (4.5) is upper-bounded by the second term in (4.14). Consider $P(r) = h \cdot r^{-\delta}$, where h is a constant and δ is positive. This corresponds to the wire-length distribution function $p(r)$ decreasing faster than $\frac{1}{r}$. The Feuer wire-length distribution with $p < 1$ is such a distribution. Using integrals to approximate sums, we find that the first term in (4.14) grows as $L_{\max}^{1-\delta}$, while the second grows no faster than $L_{\max}^{1-2\delta}$, the latter is thus always asymptotically smaller. From this it is clear that our bound on the variance in equation (3.2) is tight. It is therefore clear that the optimal strategy, whatever form it takes, cannot improve very much on a sub-optimal one like the random one-bend.

5. Conclusions and Caveats

In conclusion, we would like to make some remarks on the validity of our results, particularly in the light of the fact that some of them hinge on asymptotic comparisons between the terms in the gate-width expression. The weakest link in the analysis, perhaps, is the applicability of Rent's Rule. Although this analytical representation of locality has been borne out in practice [5], there are several reasons to suspect its continued validity.

Notice that Rent's Rule says nothing about logic circuits *by themselves*; it only speaks of circuits that have been embedded on a plane (p.c.b., chip, etc.). The rule thus assumes something

about the placement of logic components on a plane. It is necessary to view this from two points:

- (1) Can Rent's Rule be assumed to hold for large n ? Since GRAPH PARTITION and other problems closely related to logic component placement are known to be *NP-Complete*, it is not clear that very large circuits can be "successfully" placed so as to comply with Rent's Rule.
- (2) Can locality in logic circuits be characterized without reference to their embedding on planes? This would free us from the difficulty mentioned above. We do not have to assume that a circuit embedding has a certain intrinsic locality. Moreover, a successful characterization might lead us to provably good heuristics for the GRAPH PARTITION problem applied to 'circuit-like' graphs.

Two other aspects of our model are worth noting while evaluating our results.

- (1) Is it clear that as designs grow very large, minimum distance routings are in general the best? Note that a randomized strategy that permitted longer routes would suffer from a larger leading term in the limit equation (3.3).
- (2) Our analysis depends heavily on a uniformity assumption, but practical circuits will certainly exhibit varying behavior at different parts of the physical chip. Gates near the edges will have their wire emission patterns polarized towards the middle, for instance. From an algorithmic point of view, it is not clear whether uniform routing strategies are the right approach.

These doubts notwithstanding, we believe that the basic principle delineated by the results in Section 3 is that as designs grow very large, the law of large numbers makes an attractive case for randomized algorithms. The result in Section 4 derives much of its significance from the simplicity of one-bend routes, as also the fact that wires following such routes have only one via in them. The minimization of the number of vias in a wire/layout has long been considered a desirable feature in VLSI layouts [9]. Moreover, we used the fact that if the number of bends (vias) in a wire is a constant (either in the absolute or the average sense), gates over which a wire bends do not pay a heavy penalty; the number of additional tracks in each channel is a constant. Finally, the need to consider the second term in channel-width estimation once more highlights the fact that asymptotic results can be misleading in VLSI.

6. Acknowledgements

The authors would like to thank Dick Karp for motivating this work and furnishing several key references. This work also benefited from fruitful discussions with Alberto Sangiovanni-Vincentelli, Umesh Vazirani and Jean Walrand.

References

1. R. Beresford, "Evaluating Gate-Array Technologies," *VLSI DESIGN*, vol. V, no. 1, pp. 48-52, Jan '84.
2. R. D. Etchells, J. Grinberg, and G. R. Nudd, "Development of a three-dimensional circuit integration technology and computer architecture," *Soc. Photographic and Instrumentation Engineers*, vol. 282, pp. 64-72.
3. Michael Feuer, "Connectivity of Random Logic," *IEEE Transactions on Computers*, vol. C-31, no. 1, pp. 29-33, January 1982.
4. Abbas El Gamal, "Two-dimensional stochastic model for interconnections in master-slice integrated circuits," *IEEE Trans. Circuits and Systems*, vol. CAS-28, pp. 127-137, Feb 81.
5. W. Heller, W. Mikhail, and W. Donath, "Prediction of wiring space requirements for LSI," *Design Automation and Fault-Tolerant Computing*, pp. 117-144, 1978.
6. D. S. Johnson, "The NP-completeness column: An ongoing guide," *Journal of Algorithms*, vol. 3, pp. 381-395, 1983.
7. R. M. Karp, F. T. Leighton, R. L. Rivest, C. D. Thompson, U. Vazirani, and V. Vazirani, "Global Wire Routing in Two-Dimensional Arrays," *Proc. 24th Annual Symp. on Foundations of Computer Science*, pp. 453-459, October 1983.
8. Frank Thomson Leighton and Arnold L. Rosenberg, "Three-Dimensional Circuit Layouts," Technical Report TR84-08, Microelectronics Center of North Carolina, March 1984.
9. C. Mead and L. Conway, *Introduction to VLSI Systems*, Addison-Wesley, 1980.
10. A. Rényi, "Probability Theory," *North-Holland Series in Applied Mathematics and Mechanics*, vol. 10, North-Holland Publishing Company, Amsterdam, 1970.
11. C. D. Thompson, "A Complexity Theory for VLSI," Ph. D. Dissertation, CMU-CS-80-140, Computer Science Dept., Carnegie-Mellon University, August 1980.