Global Wiring by Simulated Annealing

MARIO P. VECCHI, MEMBER, IEEE, AND SCOTT KIRKPATRICK, MEMBER, IEEE

Abstract—**Simulated annealing, a new general-purpose method of multivariate optimization, is applied to global wire routing for both idealized (synthetic) and actual designs of realistic size and complexity. Since the simulated annealing results are better than those obtained by conventional methods we use them as a standard against which to compare several sequential or greedy strategies commonly employed in automatic wiring programs.**

I. INTRODUCTION

A GENERAL framework for dealing with large-scale optimization problems, the "simulated annealing" method, has recently been introduced by Kirkpatrick, Gelatt, and Vecchi [1]. The method is intended for problems with very many degrees of freedom and an objective function which combines conflicting goals. Such an objective function (Fig. 1) may have many local minima (typically of order exp (N), where N is the number of degrees of freedom), and the problem of finding the absolute lowest minimum can in simple cases be shown to be *NP* complete. In practice, however, one needs a good solution, not the absolute minimum, and assurance that there are no solutions significantly better than the one found. Heuristic methods are used to find good solutions with acceptable computational cost.

Simulated annealing is a family of heuristic optimization techniques, derived by a natural analogy with the statistical physics of random systems. The analogy is developed in detail and the method is applied to several problems arising in automated design of computers in [1]. The traveling salesman problem is also discussed in that work. This paper concentrates on the global wiring, or "rough routing" [2] problem which arises in automatic wiring of integrated circuits or their higher level packages.

We first summarize the simulated annealing approach, then sketch the slightly stylized version of global wiring which will be studied. For further details of the method, see [1]; for additional background on the problem, see [2] or [3]. A quantitative objective function for global wiring is then introduced. Traditional methods as well as the annealing approach are applied to two wiring problems:

1) An ensemble of randomly generated wire lists simulating actual designs.

2) The wire list generated by 98 chips placed on a second level package in a modern computer.

Manuscript received September 13, 1982; revised May 20, 1983.

M. P. Vecchi was with IBM Thomas Watson Research Center, Yorktown Heights, NY. He is now with the Physics Center at the Insitutto Venezolano de Investigaciones Cientificas (IVIC), Apartado 1827, Caracus 1010A, Venezuela.

S. Kirkpatrick is with IBM Thomas Watson Research Center, Yorktown Heights, NY 10598.

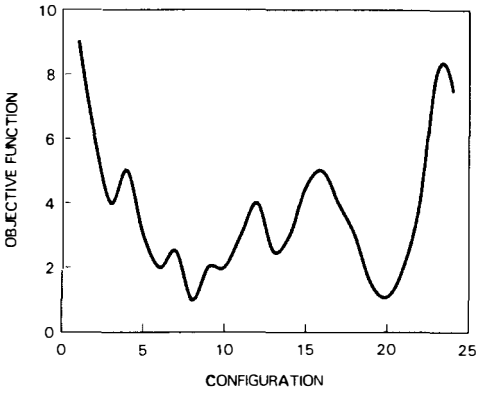

Fig. 1. Hypothetical objective function for a multivariate system with conflicting goals.

We compare the results of the different methods by several appropriate figures of merit as well as by estimating overflows, or wires which fail to fit into the intended package. Finally we consider the assignment of wires to planes of a multilevel package, including this as a part of the global wiring problem.

The simulated annealing approach consists of the following series of identifications between the many parameter problem and a hypothetical fluid consisting of many interacting atoms:

fluid	*optimization problem*
internal energy	objective function
atomic positions	parameters
cool into stable, low energy state	find a near optimal configuration.

To bring the fluid into a low energy state (as, for example, in growing large single crystals), the most effective procedure is careful annealing. First one melts the fluid, then lowers the temperature slowly, spending a long time at temperatures near the freezing point to allow defects to anneal out of the growing crystals, then cools the crystals more rapidly to bring the atoms to rest. The same sequence can be followed in optimization by introducing a pseudo-temperature, which is just a control parameter in the same units as the objective function. Several techniques exist for computer simulation of the motion of the atoms of a fluid in equilibrium at a given temperature. The Metropolis Monte Carlo method [4], [1] is particularly easy to extend to the optimization context. This is a probabilistic algorithm. In each step of the algorithm, one atom is moved. The new configuration is accepted with probability one if its energy is less then before, with probability exp ($-\Delta E/T$) if the energy is greater than before. This choice of acceptance probability for the uphill steps ensures that in

0028-0070/83/1000-0215$01.00

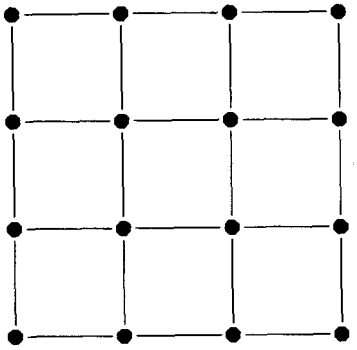

Fig. 2. Model grid representing the system to be wired. The nodes represent the sources (and sinks) of wires, and the links correspond to the channels through which the wires run.

the sample of configurations generated the probability of encountering low energy states is enhanced by a factor exp $(-E/T)$. Thus the lowest energy states, which typically have weight exp $(-N)$ in the whole configuration space, can still dominate the population of configurations generated as $T \to 0$.

The annealing process produces a metastable state of the fluid, not necessarily a true ground state or global optimum. In [1] it is argued that almost all metastable states reached by a given amount of annealing effort are of comparable quality, and there is no configuration significantly better than the solutions found by annealing, at least for the situations with conflicting constraints and very many parameters, where heuristic methods are required.

In this paper we report a series of experiments, comparing the results of various heuristic methods, including simulated annealing, when used to produce optimal global wiring.

II. Objective Function: Hamiltonian Analogy

In focussing on wiring, we shall assume that the logical specification of the design is fixed, and all components have been assigned physical locations. We shall also assume that nets grouping more than two pins have been "ordered," that is, broken into individual connections between a specific pair or pins. The problem is now specified by a "from-to list" of sources and sinks of wire.

The usual procedure, given an ordering, is first to construct a "global" or estimated coarse scale routing for each connection from which the ultimate detailed wiring can be completed. Package technologies and structured image chips have prearranged areas of fixed capacity for the wires. For the global routing to be successful, it must not call for wire densities which exceed this capacity.

We model the global routing problem (and even simple cases of detailed embedding) by lumping all actual pin positions into a regular grid of N_x by N_y points which are treated as the sources and sinks of all connections. The wires are then to be routed along the links which connect adjacent grid points. This is a reasonable model for structured packages. Adding blockages to the model by prefilling some of the links allows us to incorporate some of the effects of pre-placed components or macros found in custom technologies. This model is sketched in Fig. 2.

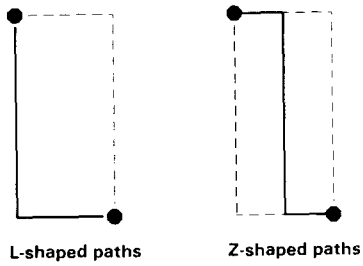

Fig. 3. Schematic of the two types of wire paths considered. For L-shaped paths (with only one bend) there are two possible states. For Z-shaped paths (with one or two bends) the possible states are as many as the length of the connection (in grid units).

Multilevel package technologies are becoming increasingly common. Typically, one level of such a package consists of a pair of planes, with horizontal lengths of wire on one plane, vertical wires on the other, and connections between the pair of planes made through via holes each time a wire turns. We shall initially discuss global routing in which all connections are considered as if on a single level, and subsequently generalize the algorithm to also perform optimal assignment of wires to levels.

We can characterize our simplified problem by several parameters. The total number of links available, $\mathfrak{N}$, is

$$\mathfrak{N} = (N_x - 1)N_y + N_x(N_y - 1). \tag{1}$$

Each wire has a minimum Manhattan length required (sum of the x and y distances between source and sink). If W is the total of these Manhattan lengths, the average number of wires per link required, ρ is

$$\rho = W/\mathfrak{N} \tag{2}$$

and the average length, λ, of each connection is

$$\lambda = W/M \tag{3}$$

where M is the number of connections. The methods of this paper consider only rearrangements which keep the length of each wire minimum, so W, ρ, and λ are constants for each choice of the M connections.

When the two points of a connection lie on a straight line, there will be only one minimum length path available. For every problem we will have M_f of these fixed connections. Each of the remaining M_r connections can be routed along at least two different paths. We can restrict the paths to having only one bend, and we refer to them as L-shaped paths. By allowing one or two bends in the path, we obtain what we call Z-shaped paths (Fig. 3). Vias between planes, required by bends, are costly and a source of reliability problems, so we shall not consider paths with more than two bends in this rough routing analysis.

Our wiring problem now reduces to choosing paths for the M_r routable connections in such a way that the likelihood of "overflows," wires which do not fit into the eventual detailed package, is minimized. This means that we seek the most uniform possible distribution of wires plus preexisting blockages. An objective function, F, which rewards the most balanced arrangement, is obtained by calculating the square of the numbers of wires on each link of the package, and summing the re-

sults over links,

$$F = \sum_{\nu=1} m_\nu^2 \tag{4}$$

where m_ν is the number of wires (including blockages) on the ν link. We can convert F into a dimensionless figure of merit, or quality measure, Q, by appropriate normalization

$$Q = A(\rho) F \tag{5}$$

choosing $A(\rho)$ such that $Q = 1$ will be a lower bound at each value of ρ.

In the regime of large wire density (i.e., $\rho >> 1$), the discrete nature of m_ν will be negligible, and the optimum configuration corresponds as closely as possible to $m_\nu = \rho$, for all ν. We choose, therefore,

$$A(\rho) = \mathfrak{N}, \quad \rho > 1 \tag{6}$$

in order to have a limiting value of $Q = 1$ for large wire densities.

In the regime of small wire density (i.e., $\rho < 1$), the discrete nature of m_ν is a dominant effect. This regime corresponds to detailed wiring, in which the constraint is that no more than one wire be assigned to a channel. In this case, the optimum solution corresponds to values of $m_\nu = 0, 1$. Therefore, we choose the normalization factor to be

$$A(\rho) = \sum_{\nu}^{\mathfrak{N}} m_\nu = W, \quad \rho < 1 \tag{7}$$

in order to again have a limiting value of $Q = 1$. Notice that at the point $\rho = 1$ there is no discontinuity in the choice of the normalization factor.

As we will see later, the factor Q is a good indicator of the quality of the wiring configuration obtained. Not all of the relevant information is contained in Q, of course, but a smaller value of Q is always an indication of a better solution to the wiring problem.

For the L-shaped moves, F has a relatively simple form. Let $\epsilon_{i\nu} = +1$ along the links which connection i uses in one orientation, -1 for the other orientation and 0 otherwise. Let $a_{i\nu}$ be 1 if the ith connection can run through the νth link in either of its two positions, and 0 otherwise. Note that $a_{i\nu}$ is just $\epsilon_{i\nu}^2$. Then if $\mu_i = \pm 1$, we obtain for the number of wires along the νth link,

$$m_\nu = \sum_i a_{i\nu}(\epsilon_{i\nu}\mu_i + 1)/2 + m_\nu(0) \tag{8}$$

where $m_\nu(0)$ is the contribution from straight wires, which cannot move without increasing their length, and blockages.

Summing the m_ν^2 gives

$$F = \sum_{i,j} J_{ij}\mu_i\mu_j + \sum_i h_i\mu_i + \text{constants} \tag{9}$$

which has the form of the interaction energy for a random magnetic alloy or spin glass, like those discussed in [1]. The "random field," h_i, felt by each moveable connection reflects the difference, on the average, between the congestion associated with the two possible paths:

$$h_i = \sum_\nu \epsilon_{i\nu}\left(2m_\nu(0) + \sum_j a_{j\nu}\right). \tag{10}$$

The interaction between two wires is proportional to the number of links on which the two nets can overlap, its sign depending on their orientation conventions:

$$J_{ij} = \sum_\nu \epsilon_{i\nu}\epsilon_{j\nu}/4. \tag{11}$$

Both J_{ij} and h_i vanish, on average, so it is the fluctuations in the terms which make up F which will control the nature of the states which minimize F.

We have not tried to exhibit a functional form for the objective function with Z-moves allowed, but simply calculate it by first constructing the actual amounts of wire found along each link. For a connection of minimum Manhattan length n grid units, there are n possible Z-moves. We label these moves by a variable α_i which takes on integer values $\alpha_i = 1, \cdots, n$, identifying $\alpha_i = 1, \cdots, n_x$ with Z-moves in which the wire goes α_i steps horizontally before turning into the vertical (or y-) direction, and $\alpha_i = n_x + 1, \cdots, n$ with the Z-shaped configurations in which the wire goes vertically then horizontally (see Fig. 3). The L-shaped configurations $\mu_i = \pm 1$ are identified with $\alpha_i = n_x, n$. We can represent the variable α_i (or the two-state variable μ_i) by n (or 2) positions on the unit circle. Each move generated in our Monte Carlo trials corresponds to rotating the angle, either clockwise or counterclockwise, to the next possible position.

It is clear that, because of the larger number of available states for each connection, the computational effort required to obtain a good solution for Z-shaped paths will be larger than for L-shaped paths. On the other hand, we also expect to reach a better solution to our wiring problem by allowing Z-shaped paths. Most of the results to be presented in this paper for Z-shaped paths were obtained as follows: first, a careful Monte Carlo annealing from high temperatures was done using only L-shaped paths. We then used this configuration as a starting point for another careful Monte Carlo annealing using Z-shaped paths, but only from low temperatures. The final configurations obtained for Z-shaped paths in this way were of comparable quality to the results obtained from the more time-consuming annealing from high temperatures using the full set of Z-shaped paths.

III. Monte Carlo Optimization Results

The placement of the wire end points, or "pins," on the grid is an important consideration, since the final wiring configuration will depend on the pin positions. We now describe a placement model for which an extensive study was made.

We placed the pins on the grid at random. The average length of the connections, λ, was controlled by restricting the two end points of each wire to lie within an L by L square. Both the separation of the points within the square, and then the position of the pairs of points on the grid, are chosen at random for each connection. The wire density, ρ, is determined by the total number of wires placed on the grid, M, and their average length, λ. Our study was extended to examples with up to $M = 10\,000$ wires on the grid. We choose a grid size

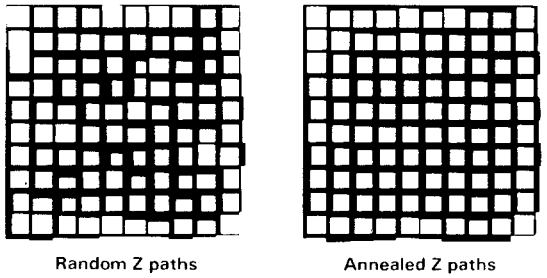

Fig. 4. Wire crossings corresponding to 750 connections with $L = 3$ placed at random on a grid of 11-by-11 nodes. The thickness of the lines are proportional to the number of wires crossing each link. Choosing each path at random (left), the maximum crossing is 20 wires/link, whereas after the optimization (right) the maximum crossing is reduced to 12 wires/link.

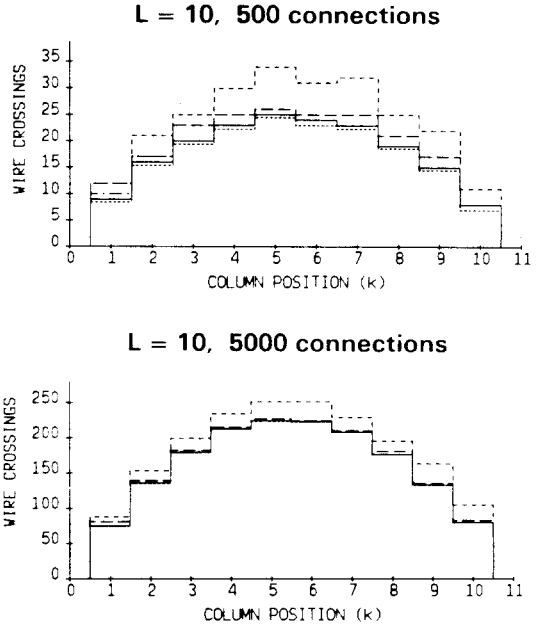

Fig. 5. Plot of the maximum number of wire crossings in the x-direction on each column of the 11-by-11 grid for two examples with $L = 10$. The average crossings are also drawn (dotted line). The different lines correspond to different configurations: random paths, short dash; quick freeze, dot-dash; annealed L-paths, long dash; annealed Z-paths, solid line.

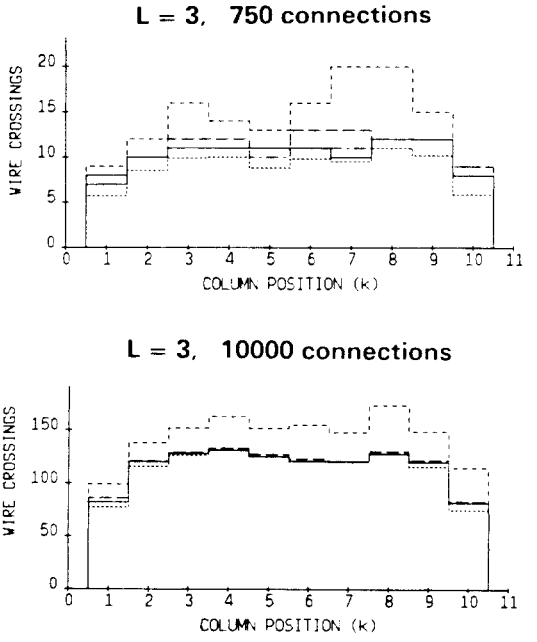

Fig. 6. Same as Fig. 5, but for two examples with $L = 3$.

of $N_x = N_y = 11$ points, and two choices of limiting connection sizes: $L = 10$ and $L = 3$. The average length per connection is $\lambda \cong 7.2$ for $L = 10$ and $\lambda \cong 2.7$ for $L = 3$, measured in units of the distance between adjacent points on the grid.

The improvements achieved by optimization are clearly seen in Fig. 4. The more uniform distribution of wires after annealing represents a better wiring configuration. A more detailed understanding can be obtained from Figs. 5 and 6, where we present the spatial distribution of wire crossings over each of the $(N_x - 1)$ columns of links in the x-direction. We also looked at the $(N_y - 1)$ rows of links in the y-direction, but since our model system is isotropic we only discuss the x-direction results. The best possible solution would be to obtain in each column the average number of wire crossings over its N_y links. We see that the maximum crossings in each column approach this average value as the optimization procedure is applied. The best results are obtained for the annealed Z-paths. One also sees that as the density of wires, ρ, increases, the optimized results are closer to the average values, both for long nets (Fig. 5) and for short nets (Fig. 6). This effect can be understood by considering the statistical fluctuations in wire crossings over the links on the grid, as we will discuss in more detail later. Notice also that there are substantial variations of the wire density over the grid as a consequence of edge effects.

In order to quantify the influence of edge effects, it is useful to define a function $g(k)$ as the probability that a wire will cross a link on column k. Starting from the probability of putting a pin on a given point of the grid (on a line across the channel in the x-direction) given by

$$r = \frac{1}{N_x} . \tag{12}$$

We can then write $g(k)$ as

$$g(k) = B \sum_{j=k_1}^{k} r \left[\sum_{i=j+1}^{k_2} r \right] \tag{13}$$

where the summation limits are given by

$$k_1 = \max \{1, (k + 1 - L)\} \tag{14}$$

and

$$k_2 = \min \{N_x, (j + L)\} \tag{15}$$

in order to keep the two points within the L-by-L square, as explained above. The normalization factor B is included for convenience in order to make

$$\sum_{k=1}^{N_x - 1} g(k) = 1. \tag{16}$$

Analogous expressions can be written for the distribution of wires in the y-direction. The distribution function $g(k)$ will be

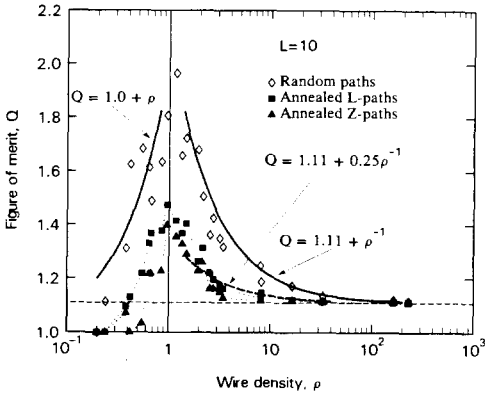

Fig. 7. Wire density dependence of the figure of merit Q for connections with $L = 10$ on a grid of 11-by-11 nodes.

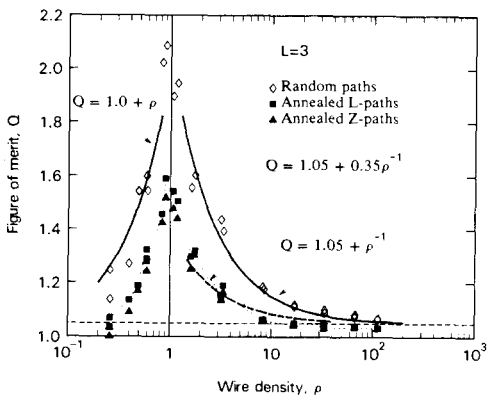

Fig. 8. Same as Fig. 7, but for connections with $L = 3$.

a function of the size of the grid (N_x, N_y) and of the dimensions of the L-by-L square. The function $g(k)$ can now be used to calculate the expected average of wire crossings per link in the channels and $g(k)$ accurately predicts the 5 different values for $L = 10$, and the 3 different values for $L = 3$.

We can now calculate the expected limiting value of the parameter Q is the regime $\rho >> 1$ when edge effects are explicitly considered. For a grid with $N_x = N_y = N$, we can rewrite (5) using $g(k)$ as

$$\tilde{Q} = \mathfrak{N}\frac{2N\sum_{k=1}^{N-1}[(N-1)\rho g(k)]^2}{\left[2N\sum_{k=1}^{N-1}(N-1)\rho g(k)\right]^2} \tag{17}$$

or

$$\tilde{Q} = (N-1)\sum_{k=1}^{N-1}[g(k)]^2. \tag{18}$$

For the grid of $N_x = N_y = 11$, we then obtain

$$\tilde{Q}(L = 10) = 1.11 \tag{19}$$

and

$$\tilde{Q}(L = 3) = 1.05. \tag{20}$$

As we shall see, these limiting values of the parameter Q are in good agreement with the obtained results.

The behavior of the figure of merit Q as a function of ρ is shown in Fig. 7 for $L = 10$ and in Fig. 8 for $L = 3$. It is seen that Q has a maximum at $\rho \cong 1$, and that it then decreases both for $\rho << 1$ and for $\rho >> 1$.

The results for random paths can be used to understand the statistical behavior of the system. If we use as a starting point (17) and we assume that the fluctuations in wire crossings for each link k can be expressed as

$$\Delta_k \cong \sqrt{(N-1)\rho g(k)} \tag{21}$$

we can arrive at an approximate expression for Q as a function of ρ in the high-density regime:

$$Q \cong (N-1)\sum_{k=1}^{N-1}[g(k)]^2 + \frac{1}{\rho}, \qquad \rho >> 1. \tag{22}$$

This equation predicts that Q will tend to its minimum value (given in (18) for our model placement) as $\rho \to \infty$, with a $1/\rho$ dependence. Furthermore, as $\rho \to 1$, $Q \sim 2$. As one can see from both Figs. 7 and 8, (22) agrees quite well with the results for random paths, both in the $1/\rho$ dependence and the limiting value of Q at $\rho >> 1$.

To understand the statistical behavior of Q in the dilute regime ($\rho << 1$), we write the parameter Q from (5) as

$$Q = \frac{\sum_{n=1}^{W} n^2 P_W(n)}{\sum_{n=1}^{W} nP_W(n)} \tag{23}$$

where $P_W(n)$ is the probability of placing n wires on one of the $\mathfrak{N}$ links of the grid when the total wire length is W.

We can then arrive at

$$Q = 1 + \rho, \qquad \rho << 1. \tag{24}$$

This last expression predicts a linear dependence of Q on ρ in the dilute regime, with a value $Q \sim 2$ as $\rho \to 1$. We see from Figs. 7 and 8 that (24) is a reasonable fit to the results for random paths both for $L = 10$ and for $L = 3$.

Having analyzed the behavior of the figure to merit Q for random wire paths, we can make the following general remarks:

1) For $\rho << 1$, Q depends linearly on ρ. In the very dilute regime, we have many available links per wire, and hence $Q \to 1$ as the probability of overlapping becomes very small.

2) For $\rho >> 1$, Q has a $(1/\rho)$ dependence determined by the fluctuations in the density. At very high densities, the low of large numbers makes the fluctuations relatively less important. The finite size of the grid, with the corresponding non-uniform distribution of wires, determines the limiting value of $Q \gtrsim 1$ as $\rho \to \infty$.

3) The maximum value of Q occurs at $\rho \sim 1$, and it is expected to be $Q \sim 2$.

Let us now look at the values of Q obtained after the Monte Carlo annealing procedure. For both $L = 10$ and $L = 3$, the Monte Carlo optimization produces significant improvements in the values of Q. In the dilute regime (i.e., $\rho < 1$) the Monte Carlo annealing resulted in near-perfect (i.e., $Q \to 1.0$) solutions for Z-paths at $\rho \lesssim 0.3$, as seen both in Fig. 7 and Fig. 8.

In the high-density regime, we see that for $L = 10$ (Fig. 7) we

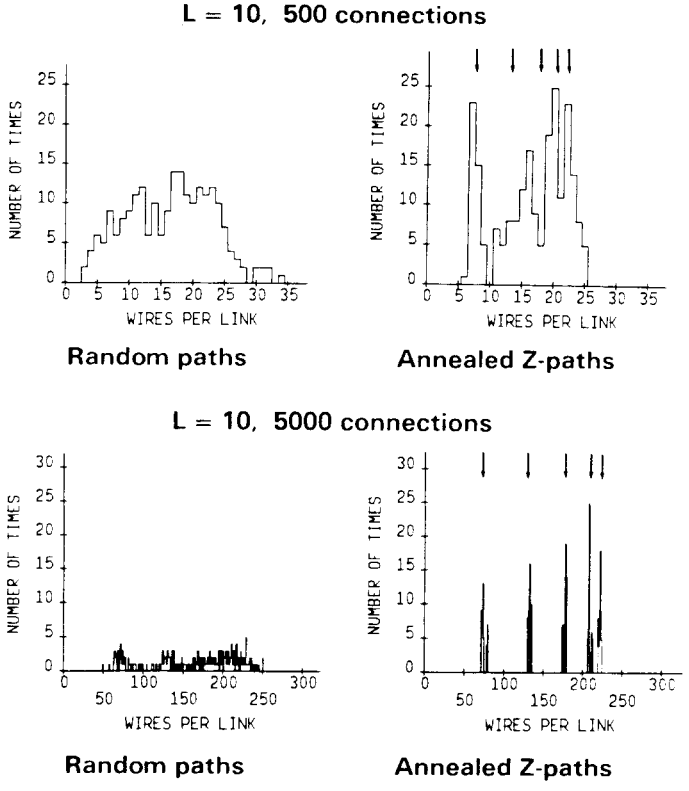

Fig. 9. Histograms of the number of wires/link for the two examples with $L = 10$ presented in Fig. 5. The arrows indicate the values predicted by considering edge effects using the function $g(k)$, as discussed in the text.

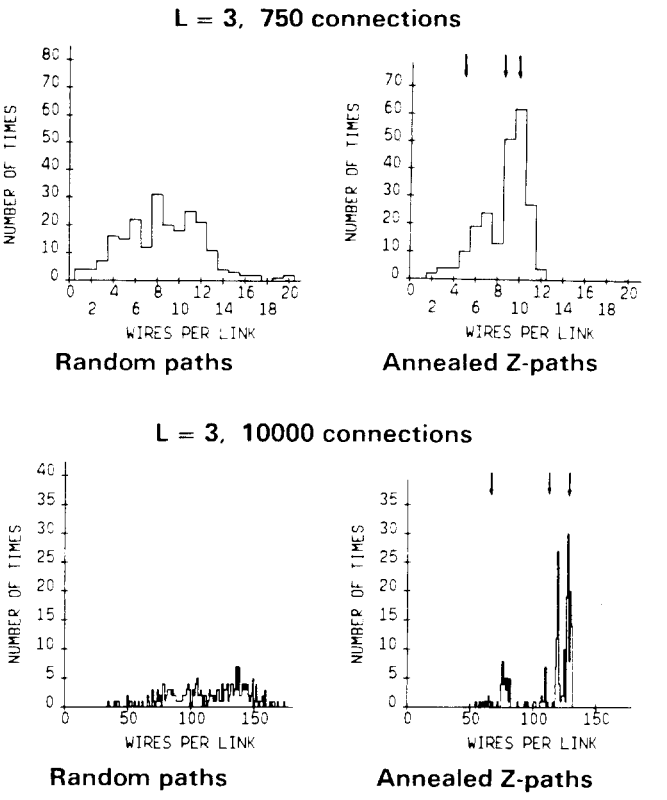

Fig. 10. Same as Fig. 9, but for the two examples with $L = 3$ presented in Fig. 6.

are within 1 percent of the best possible value of Q for $\rho \gtrsim 10$ (Monte Carlo Z-paths). In contrast, the random-path values of Q do not approach within 1 percent of the best possible value until $\rho \gtrsim 100$. As shown in Fig. 7, the annealed results can also be approximately fitted by a $1/\rho$ dependence, but with a multiplying factor (~ 0.25 for $L = 10$) that reflects the diminished amplitude of the fluctuations produced by the optimization. For $L = 3$ (Fig. 8), Q is within 1 percent of the best possible value for $\rho \gtrsim 8$ for both Monte Carlo L-paths and Z-paths, whereas for the random paths values this does not occur until $\rho \gtrsim 90$. Notice again that the $1/\rho$ fit for the annealed results includes a multiplying factor (~ 0.35 for $L = 3$).

Another way to see the changes in wire crossings for different configurations is to make histograms of the number of wires per link on the grid. These plots are shown in Figs. 9 and 10 for random paths and for the Z-shaped paths after Monte Carlo. We are presenting the same examples discussed in relation to Figs. 5 and 6. For random paths, the number of wires per link are distributed evenly over a wide range, and there is a rather long tail towards large values. The effect of the Monte Carlo annealing is to cut off this tail and to reduce significantly the maximum number of wires per link on the grid. The data for $L = 10$ shown in Fig. 9 indicates that the maximum is reduced from 34 to 25 for the $\rho = 16.4$ example (top), and from 251 to 224 for the $\rho = 163.3$ example (bottom). From Fig. 10 we see that for $L = 3$ the maximum number of wires per link is reduced from 20 to 12 for the $\rho = 8.6$ example (top), and from 174 to 131 for the $\rho = 113.7$ example (bottom).

An interesting feature of the histograms for the annealed Z-paths is that the data tends to concentrate in groups or bands. This effect is most evident in Fig. 9 (bottom), but it can also be seen in the other plots shown. One can understand this behavior by remembering the importance of edge effects. Using the values of $g(k)$, one expects that edge effects will produce five different values for $L = 10$ and three different values for $L = 3$. Fig. 9 (bottom) shows that one indeed obtains five sharp bands for $L = 10$, and the number of wires per link calculated using $g(k)$ agrees well with the positions of the peaks. The data shown in Fig. 9 (top) indicates that for $\rho = 16.4$ the final configuration obtained is not as close to the limiting solution as found for $\rho = 163.3$. This observation is confined by looking at Fig. 5, where one sees that for $\rho = 16.4$ the Z-paths solution has not approached the average values as closely as for the $\rho = 163.3$ example. One expects this difference in veiw of the importance of density fluctuations as $\rho \to 1$, as discussed in reference to (22).

The histograms for $L = 3$ shown in Fig. 10 indicate that there are larger fluctuations in wire density than for the $L = 10$ case. As we can see from Fig. 10 (bottom), even for $\rho = 113.7$ the three expected bands are broader. If we look again at Fig. 6, we see that the average number of crossings in each channel has larger fluctuations than for the $L = 10$ examples, and for the $L = 3$ cases we do not approach as closely the statistical limit predicted by $g(k)$.

The application of simulated annealing to real systems has produced good results. Mapping the pin positions from the board or module to be wired onto our model grid permits the utilization of our programs in a straightforward manner. As

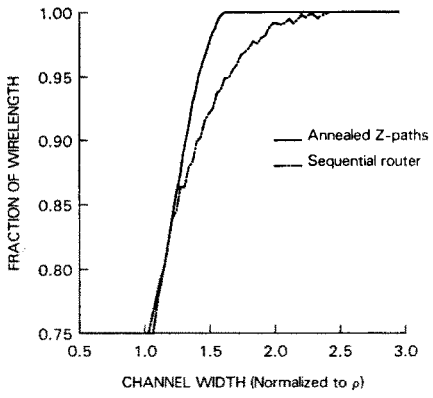

Fig. 11. Fraction of wirelength successfully routed as a function of channel capacity for the TCM example on an 10-by-10 grid. The abscissa has been normalized in units of $\rho = 64.35$.

an example, we took one TCM module[1] with 98 chips (arranged on 100 chip sites) and considered wiring the signal carrying planes. We had a total of 3792 connections after breaking up the nets into pairs of pins connected by a single wire.

On the global scale, we mapped the TCM pins onto a model grid with 10×10 nodes, obtaining an average wire density $\rho = 64.4$ wires/link with an average connection length $\lambda = 1.3$ grid units. Optimization reduced considerably the congestion of wires, and the maximum number of crossings/link decreased from 173 (random paths) to 105 (annealed Z-paths). In this TCM there is a large variation in wire density, determined by design choices, making the center of the module the most congested region. From the definition of the figure of merit Q (5), we can calculate its limiting value $\tilde{Q}$ using the average crossings over the columns (x-direction) and rows (y-direction), as discussed in reference to Figs. 5 and 6. For the TCM module, we obtain

$$\tilde{Q}(\text{TCM}) = 1.133. \tag{25}$$

The random paths configuration has $Q = 1.234$, and the best annealed results (using Z-paths) yield $Q = 1.146$, indicating that our optimization arrived at a solution very close to the limit given by (25). Assigning a channel capacity of 105 wires/link (corresponding to the worst-case crossings for annealed Z-paths), the occupancy in the most congested region of the module is in excess of 95 percent. Because of the large variation of wire density, the occupancy towards the edges of the module is much smaller. Nevertheless, further reductions in wiring requirements, beyond our optimization, cannot be obtained without changing the placement of the 98 chips on the module.

The TCM example can be used to compare the simulated annealing results with sequential wiring schemes. We implemented a simplified example of a typical production "maze-runner." First setting a limit channel capacity for successful wiring, we take one connection at a time (in a random sequence), examine all possible Z-paths, and accept the first successful route. Those connections that cannot be wired are the overflows corresponding to the channel capacity chosen. To compare the sequential router results with the simulated annealing, we search the nets in the optimized configuration in a random order and each net with overloads is removed until no more overflows are detected. All nets routed by either method are routed at their minimum length.

This comparison is shown in Fig. 11 for the TCM example. Notice that the channel capacity for 100 percent successful wiring is approximately 2.5ρ for the sequential router, and only about 1.5ρ for the annealed configuration. The much sharper edge of the annealed distribution is reminiscent of a system at low temperatures, which is in fact the result of the Monte Carlo annealing optimization procedure. Optimized routing yields statistical distributions that are more predictable than sequential, one-wire-at-a-time results, and hence provide useful data to be used for wireability predictions.

IV. Level Assignment

The final wiring of a system includes the assignment of wires to each of the several planes or levels that make up the overall package. We will use the TCM discussed earlier as an example to illustrate the usefulness of simulated annealing for optimization of level assignment.

We start by mapping the pin positions of the 3792 nets in the TCM on a grid of 170×170 nodes in order to match the actual package coordinates. The module is wired using 16 planes, 8 for horizontal wires and 8 for vertical wires. To simulate this system, we assign each net to one of 8 available levels represented by 8 grids with x- and y-direction wires. The maximum link occupancy for successful wiring is 1 wire/link on each of the 170×170 grids representing the wiring levels.

Using minimum length Z-shaped paths with the objective function given by (4), a Monte Carlo annealing is then performed following a three-step procedure:

a) The wires are uniformly distributed on a single 170×170 grid (i.e., no level assignment) in the manner described in the previous sections.
b) Keeping the Z-shaped paths obtained in (a) unchanged, we anneal from a random level-assignment configuration (i.e., $T = \infty$). The trial moves consist of shifting a wire from its level to any other level.
c) The $T = 0$ configuration obtained in (b) is again annealed, from *low* temperatures, using moves that change both the level assignments and the paths of the nets.

This optimization procedure makes use of the flexibility inherent in simulated annealing to define different sets of elementary moves and different ranges of temperatures in the annealing schedules to considerably reduce the overall computational effort in obtaining a good solution.

The maximum number of wire crossings per link at the end of step (a) is 8. The total wire length (on the 170×170 grid) is $W = 197689$, with an average length per net of $\lambda = 52.13$. After the Monte Carlo annealing for level assignment without changing the shapes of the nets (step (b)), the wires are distributed uniformly among the 8 available levels: the wire length per level has a maximum of 25486 and a minimum of 23770, to be compared with the average per level of 24711.1. The number of nets per level is also evenly distributed. There are, however, a number of overflows: of the 57460 links on each

[1] Thermal Conduction Module is a high-performance LSI package[5] used in the IBM 3081 Processor, holding up to 127 chips.

of the grids, an average of 89.50 links/level are overloaded with 2 wires.

The final configuration is obtained following the Monte Carlo annealing (of step (c)), and the improvements are noticeable. As expected, the wire length per level is still close to the average, with a maximum of 25349 and a minimum of 23674. The overflows, on the other hand, are fewer: on the average only 6.25 links/level are overloaded with 2 wires.

Another possible approach is to route the wires sequentially, one net at a time. We implemented a simple sequential router, with level assignment, as follows: starting from empty grids, we put as many nets as possible on the first level by searching all Z-shaped paths for each net and accepting the first route with no overloaded links. Nets that are not successfully routed are used for the second level, and so on. The nets are chosen in a random sequence. At the end, there will be a number of nets that were not successfully imbedded in any available level, corresponding to overflows that have to be manually embedded.

Applying our sequential router to the TCM with 8 levels, we were left with a total of 162 net overflows, corresponding to a wire length of 10621. The distribution of wire length successfully routed is not uniform: the first level is the most crowded with a wire length of 31753, and the last level only contains 12410 units of wire. Another consequence of our sequential routing procedure is that the shorter nets tend to be embedded first: the average length of the nets in the first level is 39.25, and it increases to 70.51 for the nets in the last level.

In order to compare the sequantial router's results with the Monte Carlo annealing, we search the nets in the annealed configurations in a random order and each net with one or more overloaded links is removed until there are no more overloads. Using the configuration at the end of step (b), we detected a total of 350 net overflows, which compares unfavorably with the sequential router's results. On the other hand, the final annealed configuration (at the end of step (c)) results in only 27 net overflows for the TCM with 8 wiring levels, which is considerably better than the sequential router's solution.

The problem of detailed wiring with level assignments falls in the dilute regime (i.e., $\rho \sim 0.5$ for each level), and one expects that sequential routing schemes will perform quite well. We have, in fact, seen this point in the previous discussion. Nonetheless, the simulated annealing optimization can produce even better solutions, not only in terms of a more balanced distribution of wires among the different levels, but also in terms of a reduced number of final overflows.

V. Conclusions

Monte Carlo simulated annealing has been shown to be an effective algorithm for global wiring. The quality of the solutions has been quantified and a detailed analysis has been given, taking into account both fluctuations in link occupancies and spatial inhomogeneities of pin positions.

A real system, the 3081 module, was used as an example to illustrate both global wiring optimization and detailed wiring with level assignment. Simulated annealing has been shown to be an effective algorithm in both cases.

The programs for our study were written in Fortran and run on an IBM VM/370 System with a 3033 Processor. The CPU time required for obtaining good solutions is modest. For example, the global wiring optimization for 3000 nets on the 11 by 11 grid requires about 2 s. For global wiring, the computational effort increases linearly with the number of nets, and quadratically with the grid size.

We have seen that the pseudo-temperature is a control parameter for which one can gain intuitive understanding. Hence, the required annealing schedules can be easily constructed and automated. Generalizations of the method, to include other wire shapes, channel or via restrictions, different objective functions, etc., can be implemented and may become useful design tools. Furthermore, simulated annealing provides the means to obtain statistically meaningful distributions to be used for wireability projections.

Acknowledgment

We would like to thank our colleagues Dan Gelatt, Ralph Linsker, Jerry Hickson, and John Cooper for their help and stimulation during the development of this work.

References

[1] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671-680, 1983.
[2] J. Soukup, *Proc. IEEE*, vol. 69, pp. 1281-1304, 1981, and references given therein.
[3] M. A. Breuer, editor, "Design automation of digital systems," Englewood Cliffs, NJ: Prentice-Hall, 1972; D. Hightower, Proc. Design Automation Workshop, pp. 1-24, 1969.
[4] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, *J. Chem. Phys.*, vol. 21, p. 1087, 1953. For a recent review, see K. Binder, editor," *The Monte Carlo Method in Statistical Physics.* New York: Springer Verlag, 1978.
[5] A. J. Blodgett, *IEEE Proc. Elec. Comp. Conf.*, New York, pp. 283-285, 1980. A. J. Blodgett and D. R. Barbour, *IBM J. Res. Dev.*, vol. 26, p. 30, 1982.

*

Mario P. Vecchi (S'69-M'69) was born in Caracus, Venezuela, on March 20, 1948. He received the B.S. degree from Cornell University, New York, in June 1969, and then the S.M., E.E., and Ph.D degrees from Massachusetts Institute of Technology, Cambridge, in 1972, 1974, and 1975, respectively, all in electrical engineering.

Since February 1975 he has been in the Physics Center at the Instituto Venezolano de Investigaciones Cientificas (IVIC), where he is the Head of the Semiconductors Laboratory. He is also Professor at the Universidad Metropolitana in Caracas. Dr. Vecchi was at the IBM Watson Research Center, Yorktown Heights, NY, on a sabbatical leave from September 1981 to September 1982.

He is member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, Phi Kappa Phi, American Physical Society, and Colegio de Ingenieros de Venezuela.

*

Scott Kirkpatrick (M'83) was born in Wilmington, DE, in 1941. He received the AB degree from Princeton University in 1963 and the Ph.D degree from Harvard University, Cambridge, MA, in 1969, both in physics.

He held a Post-Doctoral fellowship at the University of Chicago in 1970 and 1971, and joined the IBM Research Division in 1971. Since that time he has been a Research Staff Member at the T. J. Watson Research Center, Yorktown Heights, NY. His research has been principally in the physics of amorphous or disordered materials, and has recently shifted to problems in design automation and optimization. He is presently Manager, VLSI Design, in the Experimental Systems Department at the Yorktown Heights laboratory.

Dr. Kirkpatrick is a Fellow of the American Physical Society.