OPTIMAL ASPECT RATIOS FOR MINIMAL-AREA STANDARD-CELL INTEGRATED CIRCUITS

Eric Rosenberg

Eric Rosenberg *is a distinguished member of technical staff in the Data Networking Decisions and Design Systems Department at AT&T Bell Laboratories in Lincroft, New Jersey. He manages the design and development of PC-based tools that help AT&T's technical sales-support staff configure and analyze the performance of local-area networks and distributed computing workgroups. Mr. Rosenberg joined the company in 1979. He has a B.A. in mathematics from Oberlin College and a Ph.D. in operations research from Stanford University.*

What aspect ratio yields minimal area for a standard-cell integrated circuit (IC) that has a fixed number of cells and set of interconnections? Do square ICs really have minimal area? Our deterministic mathematical model provides answers and can help IC designers who want to vary the aspect ratio of a standard-cell block incorporated in a custom IC. Through a series of layout experiments, we show the model to be very accurate (the average error is 2.0 percent). We show that a cubic equation must be solved to find the aspect ratio that yields the theoretical minimal area. Key ingredients in our analysis are the maximum vertical and horizontal densities, which we empirically show to be linear functions of the number of rows. The major conclusion of this work is that square ICs do yield minimal area. But even over a wide range of aspect ratios (i.e., nonsquare shapes), little increase in area is incurred.

Introduction

In this paper, we construct an area-prediction model to study the problem of choosing the aspect ratio that minimizes the area of a standard-cell IC. Minimizing area is important for reducing fabrication costs and improving performance and is part of *floorplanning*, the stage in IC design where the designer decides the shapes and relative positions of all the major components of a custom IC. One motivation for our work is the lack of a published theoretical or experimental justification for the "rule of thumb" that ICs should be square.

During floorplanning, the designer may want to vary the aspect ratio of a standard-cell block that will be combined with components such as memory or programmed logic arrays (PLAs). So, it is also important to know how much of a penalty in area will be incurred when the design deviates from the optimal aspect ratio. (The penalty could mean that the IC has to be larger than planned, has much unused or wasted space, or needs "longer" wires—which

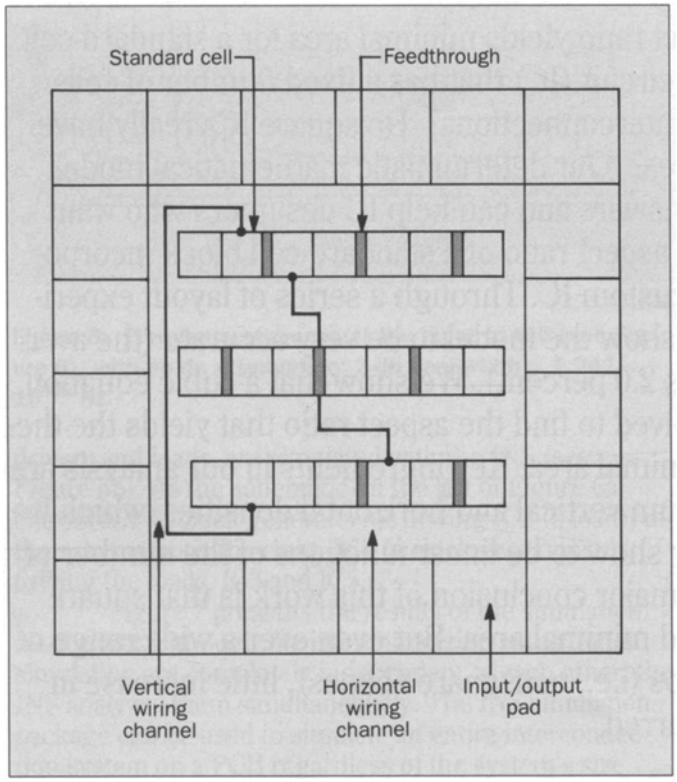

Figure 1. A standard-cell integrated circuit.

could affect performance, increase fabrication costs, or exceed fabrication technology limits.) If a significant penalty is incurred, then this should be considered in the floorplanning problem.

Although numerous papers on wiring-space estimation have appeared, none appear to consider the problem of finding the optimal aspect ratio. This paper addresses that problem and is organized as follows: First, we review related literature. Then, we formulate the problem as an optimization problem and discuss the theoretical solution. Finally, we discuss the empirical results.

Related Literature

The related literature considers the general problem of estimating the wiring space or total wire length.

Sutherland and Oestreicher studied[1] the area required for printed circuit board (PCB) wiring, with varying aspect ratios. They concluded that boards should be square. Their paper appears to be the only published theoretical analysis of when "square" is optimal. However, Sutherland and Oestreicher assume that:

- The maximum vertical and horizontal wire densities are equal. The *maximum vertical density* is defined as follows: For each imaginary horizontal line that bisects the IC, count the vertical wires that cross this line. The *maximum vertical density* is the maximum of this count, as the imaginary horizontal line ranges from the top of the IC to the bottom of the IC. The *maximum horizontal density* is defined similarly, but using imaginary vertical lines and horizontal wires.
- These wire densities are equal to the value pN, where p is the average number of active pins per IC and N is the number of ICs.

As our empirical results will show, both assumptions are invalid for standard-cell ICs.

Donath estimates[2] the total wire length of a square array of cells. His analysis models the effect of a clustered and hierarchical placement. El Gamal calculates[3] the maximum wire area per cell in an infinite array of cells, with some assumptions about interconnection wiring.

Heller, Mikhail, and Donath estimate[4] total wire length using a stochastic model. In their model, the connections are defined along a line that is then folded into a square array. Sastry and Parker calculate[5] the dimensions of routing channels and other wireability measures in a stochastic model with an N by N array of cells. Schmidt calculates[6] densities and total wire length using a different analytic model.

For standard-cell ICs, Kurdahi and Parker proposed[7] a stochastic area-estimation model in which a one-dimensional layout is folded into an array with a variable

number of rows. Closed-form solutions are not presented and numerical solution may be required. Kurdahi and Parker also observed, from empirical studies, that square chips have minimal area. But they do not provide theoretical justification for what they term a *rule of thumb*.

Mathematical Formulation and Solution

We assume a rectangular IC with N cells, arranged in an array of C columns and R rows. Thus, $N = R \times C$. Let each of the N cells have height, H, and width, W. (This uniformity assumption is not restrictive for large N.) We assume that the IC has unidirectional routing; i.e., a layer is used either only for horizontal routing or only for vertical routing. For simplicity, we assume one vertical and one horizontal layer (this concept can easily be extended). Figure 1 illustrates a standard-cell IC.

Let D_H and D_V be the maximum horizontal and vertical densities. (Anticipating one of our important empirical results, we mention now that D_H and D_V are functions of R). Let α be the number of horizontal wiring tracks that can be routed under or over each cell, and β be the number of vertical wiring tracks that can be routed under or over a cell. (See Figure 2.)

Thus, with one-level metal (i.e., two routing layers also used in the polycell rows), $\alpha = 0$ and β is equal to the average number of feedthroughs (FTs) per cell. With two-level metal (i.e., three routing layers), we have $\alpha = 0$ and β equal to the over-the-cell wiring capacity on the second metal layer. The case $\alpha > 0$ would represent a third metal layer in which horizontal wires could run over the cells.

Because $\alpha = 0$ for both one-level and two-level metal layouts, with these technologies we will always require extra horizontal wiring space. For one-level metal, we will also generally require extra vertical wiring space.

Let Q_H be the horizontal path width plus clearance, and Q_V be the vertical path width plus clearance.

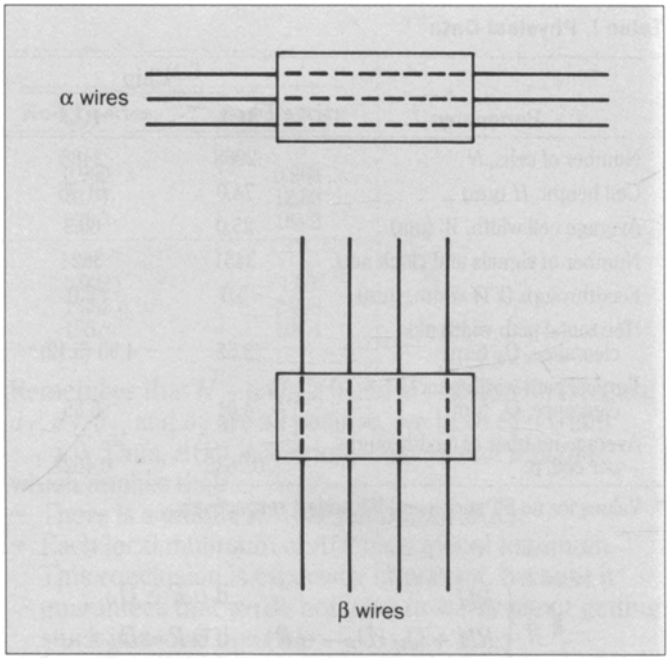

Figure 2. Wiring capacities under or over a cell; α and β are the number of horizontal and vertical wiring tracks, respectively.

Note that $H - \alpha Q_H \geq 0$ because—by definition—the horizontal wiring capacity over a cell, multiplied by the path width plus clearance, must not exceed the height of the cell. Similarly, $W - \beta Q_V \geq 0$.

Thus, with R rows and C columns, αR horizontal wiring tracks and βC vertical wiring tracks are available. If $\alpha R < D_H$, then we need $D_H - \alpha R$ extra horizontal wiring tracks, each with a path width plus clearance of Q_H. Similarly, if $\beta C < D_V$, then we need $D_V - \beta C$ extra vertical wiring tracks, each with a path width plus clearance of Q_V.

Let h and w be the height and width, respectively, of the IC. Thus:

Table I. Physical Data

Parameter	Chip IC1	Chip IC2
Number of cells, N	2968	2488
Cell height, H (μm)	74.0	61.75
Average cell width, W (μm)	25.0	60.3
Number of signals and clock nets	3451	3624
Feedthrough (FT) width, (μm)	5.0	7.0
Horizontal path width plus clearance, Q_H (μm)	3.52	4.90 (5.12)*
Vertical path width plus clearance, Q_V (μm)	4.01	5.55
Average number of feedthroughs per cell, β	0.2635	0.4622

* Values for no-FT series and FT series, respectively.

$$h = \begin{cases} RH & \text{if } \alpha R \geq D_H \\ RH + Q_H (D_H - \alpha R) & \text{if } \alpha R < D_H \end{cases}$$

$$w = \begin{cases} CW & \text{if } \beta C \geq D_V \\ CW + Q_V (D_V - \beta C) & \text{if } \beta C < D_V \end{cases}$$

So:

$$h = RH + \max\{0,\ Q_H (D_H - \alpha R)\} \quad (1)$$

$$w = CW + \max\{0,\ Q_V (D_V - \beta C)\} \quad (2)$$

The area minimization problem (AMP) is then:

$$\text{minimize} \quad hw$$

$$\text{subject to:} \quad RC = N$$

$$\text{for} \quad R,\ C \text{ positive integers} \quad \text{(AMP)}$$

Because N is large for ICs, we will ignore the integer restrictions on R and C and treat them as continuous variables.

Table II. Area Estimation Error

Chip	No-FT series Rows	No-FT series Error (%)	FT series Rows	FT series Error (%)
IC1	10	5.62	9	0.41
	24	-3.80	23	-3.65
	36	-2.72	36	0.40
	49	1.80	49	0.62
IC2	16	1.41	16	7.03
	25	-1.37	25	0.61
	34	-0.29	34	-1.47
	49	0.90	49	-0.68

Because $C = N/R$, problem AMP can be simplified to yield the problem AMP1 in the single unknown R:

$$\underset{R \geq 0}{\text{minimize}} \left[RH + \max\{0,\ Q_H (D_H - \alpha R)\} \right] \times \left[\frac{NW}{R} + \max\left\{ 0,\ Q_V \left[D_V - \frac{\beta N}{R} \right] \right\} \right] \quad \text{(AMP1)}$$

One of the surprising results of our experiments (discussed in detail below) is that both D_H and D_V are linear functions of R. That is, there exist *positive* constants a_1, a_2, b_1, and b_2 such that:

$$D_H \approx a_1 R + a_2$$

and

$$D_V \approx b_1 R + b_2$$

(The approximation is quite good.) Henceforth, we will write $D_H(R)$ instead of D_H, and $D_V(R)$ instead of D_V.

Given these expressions for $D_H(R)$ and $D_V(R)$, assume that $D_H(R^*) > \alpha R^*$ and $D_V(R^*) > \beta C^*$ for that R^* and C^* that minimize the IC area (where $C^* = N/R^*$). (R^* and C^* are the optimal number of rows and columns,

Table III. Linear Fit of Horizontal and Vertical Densities

	IC1		IC2	
	No-FT series	FT series	No-FT series	FT series
Horizontal density				
corr(R,H)	0.998	0.999	0.999	0.998
slope(H)	76.83	28.61	43.77	12.03
intercept(H)	339.6	186.5	238.5	184.2
Vertical density				
corr(R,V)	0.986	0.950	0.923	1.00
slope(V)	2.031	3.041	1.652	2.278
intercept(V)	231.8	135.55	155.8	109.4

respectively.) Then, the optimal R^* can be calculated as follows.

If we substitute $C = N/R$ in equation (2) and multiply equations (1) and (2), then the IC area, which we denote by $A(R)$, is given by:

$$A(R) = [\, RH + Q_H(a_1R + a_2 - \alpha R)\,] \times \left[\frac{NW}{R} + Q_V\left(b_1R + b_2 - \frac{\beta N}{R}\right)\right]$$

$$= c_0R^2 + c_1R + c_0 + c_{-1}R^{-1} \qquad (3)$$

where:

$$c_2 = b_1Q_V[H + (a_1 - \alpha)Q_H]$$

$$c_1 = Q_VHb_2 + Q_HQ_V[(a_1 - \alpha)b_2 + a_2b_1]$$

$$c_0 = NHW + Q_HNW(a_1 - \alpha) - Q_VH\beta N + Q_HQ_V[a_2b_2 - (a_1 - \alpha)\beta N]$$

$$c_{-1} = Q_HNa_2(W - \beta Q_V)$$

Remember that $H - \alpha Q_H \geq 0$ and $W - \beta Q_V \geq 0$. Because a_1, a_2, b_1, and b_2 are all positive, we have $c_2 > 0$ and $c_{-1} \geq 0$. Thus, $A(R)$ is a strictly convex function of R, which implies that:

- There is a unique R^* that minimizes $A(R)$.
- Each local minimum of $A(R)$ is a global minimum. This conclusion is especially important, because it guarantees that we do not have to worry about getting stuck in a local optima.

To minimize $A(R)$, we would take the derivative and set it equal to zero. This requires that we know the constants a_1, a_2, b_1, and b_2 and that we perform the difficult task of solving a cubic equation.

Fortunately, in the next section, we show empirically that it is not necessary to minimize $A(R)$. A nearly optimal solution can be obtained simply by choosing R such that the IC is square.

Experimental Results

To determine the accuracy of the area-prediction model, we performed two series of one-metal-layer layouts with two standard-cell ICs, denoted by IC1 and IC2:

- IC1 consists of cells designed for high performance. The average cell size is about that of a four-input NAND gate.
- IC2 uses an older technology. The cells are designed to balance performance and area. The average cell size for IC2 is about that of a seven input AND-OR-INVERT gate.

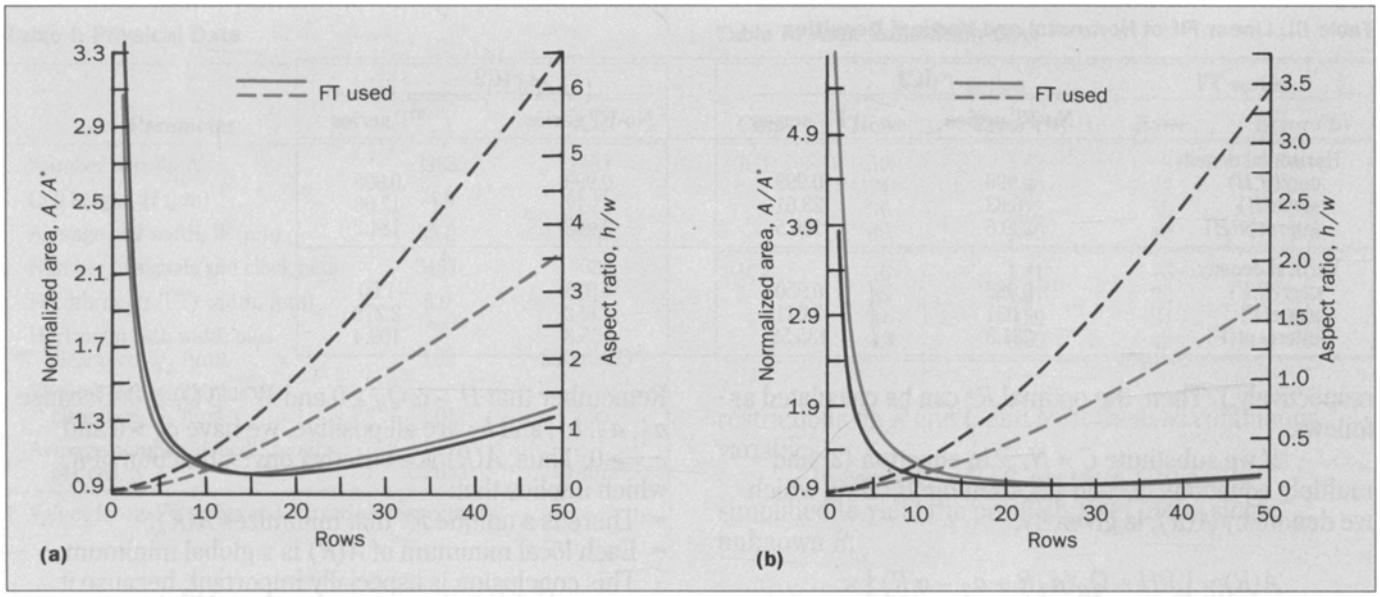

Figure 3. Normalized IC area (A/A^*), solid curves, and aspect ratio (h/w), dashed curves, for two integrated circuits: (a) IC1 and (b) IC2. "FT used" means these layouts used feedthroughs for routing.

In the first series of layouts, we prohibited routing from using feedthrough cells. In the second series, feedthroughs were used. (For convenience, we will refer to the first series of layouts as *no-FT* and the second series of layouts as *FT*.)

In each series (i.e., no-FT and FT), four layouts—each with a different number of rows—were done for each chip. Because the two series of layouts represent different routing approaches we could examine the robustness of the model and of the linear expressions for $D_H(R)$ and $D_V(R)$.

The *Easy LTX2* option of the LTX2 system requires minimum user intervention, so we used it to obtain a "good" layout. (LTX2 is the second generation of the LTX automatic layout system,[8] developed and in use at AT&T Bell Laboratories.) A good layout includes:

- Mincut placement—assigns each cell to a specific position in a row, to minimize horizontal and vertical densities.
- Loose routing—calculates the sequence of horizontal and vertical channels used for each interconnection.
- Placement improvement—improves the placement using the loose-routing results.
- Fine routing—calculates detailed positions of all wires.
- Compaction—removes unused space in the routing regions of the IC.

Although these layouts are not necessarily production quality, they provide an excellent starting point should further optimization be desired. (Further optimization should not affect our conclusions.)

Table I gives the physical data for IC1 and IC2; i.e., number of cells, cell height, average cell width, number of signal and clock nets (i.e., interconnections that carry data and timing information, respectively), feedthrough width, Q_H, Q_V, and β. Note that Q_H was

Table IV. R^* and Range of Rows for Given Increase in Area

Chip	R^*	Range of rows for % increase in area				
		2.5%	5%	10%	15%	20%
IC1 (no-FT series)	15	11-21	9-24	7-29	6-34	6-38
IC1 (FT series)	17	12-23	10-27	8-32	7-37	6-41
IC2 (no-FT series)	22	15-32	13-37	10-45	9-49	8-49
IC2 (FT series)	28	21-39	18-44	15-49	13-49	11-49

NOTE: R^* is the optimal number of rows.

Table V. $(h/w)^*$ and Range of Aspect Ratios for Given Increase in Area

Chip	$(h/w)^*$	Range of aspect ratios for % increase in area				
		2.5%	5%	10%	15%	20%
IC1 (no-FT series)	1.06	0.64-1.81	0.46-2.24	0.31-3.00	0.24-3.78	0.24-4.46
IC1 (FT series)	0.70	0.40-1.14	0.29-1.47	0.21-1.90	0.17-2.36	0.13-2.72
IC2 (no-FT series)	0.92	0.48-1.71	0.38-2.17	0.25-2.95	0.21-3.36	0.17-3.36
IC2 (FT series)	0.73	0.43-1.15	0.34-1.38	0.26-1.62	0.20-1.62	0.16-1.62

NOTE: $(h/w)^*$ is the optimal aspect ratio.

obtained empirically by dividing the total widths of the horizontal channels by the total number of horizontal wires. Q_V was obtained similarly but using vertical channels and wires. For IC2, Q_H is 4.90 when feedthroughs are not used, and 5.12 when feedthroughs are used. For IC1, the values are both 3.52. The number of rows used in each layout appears in the second column of Table II.

Once the 16 layouts were done, a quick hand plot indicated that the maximum horizontal and vertical densities (obtainable from LTX2) were linear in R. Then, we used a curve-fitting program[9] to obtain the coefficients and the correlation. This data is given in Table III. Here, corr(R,H) is the correlation (Pearson's r value) of the fit of R and the observed values of D_H; using the above notation, slope$(H) = a_1$ and intercept$(H) = a_2$. Similarly, corr(R,V) is the correlation of the fit of R and the observed values of D_V, and slope$(V) = b_1$ and intercept$(V) = b_2$.

From the correlations, it is evident that the linear fit is quite good. But because the coefficients are different for the different runs, we cannot hope to predict these values for a new IC for which a layout is desired. To us, the value of these coefficients is in validating the area-prediction expression, $A(R)$, and in gaining insights on how both the area and aspect ratio change as a function of R.

To determine the accuracy of the area-prediction model, we compared the actual areas obtained in the layouts to the areas predicted by equation (3). For a fair comparison, some corrections to the actual areas were necessary:

- For small values of R, the IC would become pad-limited in the vertical direction, which causes unused space to be added in the horizontal channels. Similarly, for large R values, the IC would become pad-limited in the horizontal direction, which produces unused space in the vertical channels. We subtracted these unused areas.
- For the no-FT series, we subtracted the extra width caused by the unused feedthroughs.

Table II gives the error in area. The average error is 2.0 percent. This error is nearly entirely due to the linear fit of the densities not being exact.

Figures 3a and 3b graph the normalized area and aspect ratio (h/w) as functions of the number of rows. The normalized area is defined as $A(R)/A(R^*)$, where $A(R^*) = \min_R A(R)$.

In each graph, we show curves for the no-FT and FT series. (Normalized areas are computed independently for the two series.) The striking feature of these graphs is that the normalized-area curves are quite flat for $10 \le R \le 49$, although the aspect ratio rises sharply as R nears 49. Tables IV and V quantify this.

In Table IV, we give the R^* (i.e., optimal number

of rows) and then identify the range of rows that would yield a value of $A(R)/A(R^*)$ within specified percentages of the minimum value, 1, of this ratio. Note that, even for a very modest penalty (5.0 percent), a wide range of values of R could be used. (We must remember that the model itself has an average error of 2.0 percent, as noted above.)

Table V shows the optimal aspect ratio and the range of aspect ratios allowable for specified percentage increases in normalized area. Although the aspect ratio that corresponds to $A(R^*)$ is not always 1.0, the range for a 2.5-percent area penalty always includes the value 1.

The model is accurate to within about 2 percent. So, for these ICs and technologies, we conclude that—if we make the chip square—we are achieving minimal area (to within the model accuracy). Moreover, we may deviate substantially from an aspect ratio of 1 and still incur very little penalty in area.

Summary

We have derived a new mathematical model for predicting the area of a standard-cell IC. Experiments with two ICs and two routing technologies show that the model predicts area with an average error of only 2 percent. Using the model, we concluded that choosing a square chip does minimize area. Also, if a nonsquare shape is desired (e.g., for floorplanning purposes), even deviating substantially from square causes little increase in area.

We recognize that we have not proved (in a rigorous mathematical sense) that square is optimal. Such a proof—if we assume that both $D_H(R)$ and $D_V(R)$ are linear in R—would require that we minimize $A(R)$ as a function of a_1, a_2, b_1, and b_2 and show that, whatever the values of these coefficients, the optimal shape is always square.

Acknowledgments

Bob Kossey, J. T. Li, and Dave Potter gave generously of their time in helping me run these experiments. Numerous stimulating discussions on standard-cell layout were held with them and also with Chi-Chang Liaw and Carl Seaquist. Bob Kossey and Chi-Chang also provided valuable comments on this paper.

References

1. I. E. Sutherland and D. Oestreicher, "How Big Should a Printed Circuit Board Be?," *IEEE Transactions on Computers*, Vol. C-22, 1972, pp. 537-542.
2. W. E. Donath, "Placement and Average Interconnection Lengths of Computer Logic," *IEEE Transactions on Circuits and Systems*, Vol. CAS-26, 1979, pp. 272-277.
3. A. A. El Gamal, "Two-Dimensional Stochastic Model for Interconnections in Master Slice Integrated Circuits," *IEEE Transactions on Circuits and Systems*, Vol. CAS-28, 1981, pp. 127-138.
4. W. R. Heller, W. F. Mikhail, and W. E. Donath, "Prediction of Wiring Space Requirements for LSI," *Journal of Design Automation and Fault Tolerant Computing*, Vol. 2, 1978, pp. 117-144.
5. S. Sastry and A. C. Parker, "Stochastic Models for Wireability Analysis of Gate Arrays," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. CAD-5, 1986, pp. 52-65.
6. D. C. Schmidt, "Circuit Pack Parameter Estimation Using Rent's Rule," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. CAD-1, 1982, pp. 186-192.
7. F. J. Kurdahi and A. C. Parker, "PLEST: A Program for Area Estimation of VLSI Integrated Circuits," *Proceedings of the 23rd Design Automation Conference*, June 29-July 2, 1986, Las Vegas, Nevada, IEEE, 1986, pp. 467-473.
8. G. Persky, D. N. Deutsch, and D. G. Schweikert, "LTX—A Minicomputer-Based System for Automated LSI Layout," *Journal of Design Automation and Fault Tolerant Computing*, Vol. 1, 1977, pp. 217-255.
9. *MathCAD: The Engineer's Scratch Pad* (software and user's manual), Mathsoft, Inc., Cambridge, Massachusetts, 1987.

(Manuscript received March 15, 1989)