Evaluation of Transistor Densities for Submicronic CMOS Technologies

F. Moraes[1], L. Torres[2], M.Robert[2], D.Auvergne[2]

[1]PUC-RS - Av. Ipiranga, 6681 – Prédio 16
90619-900, Porto Alegre, Brazil
e-mail: moraes@andros.inf.pucrs.br

[2]LIRMM, 161 rue ADA
34392 – Montpellier – Cedex 5 - France
{torres,robert,auvergne}@lirmm.fr

Abstract

The transistor density is one of the parameters to be considered for an optimal use of CMOS process. Therefore, layout strategies have to be evaluated through metrics considering all the involved parameters. The objective of this paper is to study the real transistor density available for a given technology at the cell and circuit level, from the design rules. This study has three main consequences: (*i*) to evaluate the quality of the layout synthesis CAD tools in terms of area; (*ii*) to give a feedback to the logic synthesis step, allowing an accurate area prediction from the gate level abstraction; and (*iii*) to determine a transistor density roadmap. We will compare the predicted densities, for different design rules, with real layouts generated with macro-cell generators and for standard-cell libraries.

1. Introduction

The availability of submicronic process with multi-layers for routing opens the way for defining new optimal layout strategies. The routing, shielding and spacing of multi-level signals must conserve signal integrity. All these considerations must be used in defining layout strategies and style of implementation.

Layouts can be synthesized using methods like full-custom design, standard-cells, gate-arrays, or automatic layout generation. Layouts using full-custom design are extremely dense, however due to the high time to market this approach is practically not used anymore. The time to market constraints and circuits complexity require the use CAD tools to reduce the design time. The standard-cell approach is currently the solution to synthesize random logic blocks, using a library of full-custom cells automatically placed and routed. The automatic layout synthesis offers some advantages over standard-cells, such as easier technology migration and transistor sizing, paying a higher cost in area, as shown later.

In order to fill the gap left by the current absence of metrics to evaluate layout densities and to predict the tendencies of future processes, this paper proposes a method to estimate area. The basis of this method is a regular layout style, associated to assumptions on required routing area and transistors sizes.

This paper is organized as follows. Section 2 presents some characteristics of the layout style used to estimate transistor densities. Section 3 presents the method to estimate area. Section 4 compare the densities obtained from layout synthesis and standard-cells to that defined in the previous section. Finally, we present our conclusions.

2. Layout Styles

In this section is presented the layout style which will be used for density estimation.

Uehara and van Cleemput defined the basic leaf cell layout style in 1981, called ***linear-matrix*** [1]. In the linear-matrix style, all transistors are placed into two parallel diffusion rows. The main cost function is the diffusion gap minimization, aiming at area and side-wall capacitance reduction. A cell with no diffusion gaps has minimal area and minimal parasitic capacitance values. Another example of layout style is the ***gate-matrix*** [2]. The gate-matrix style has a great number of diffusion gaps and consequently poorer performance due to the associated parasitic capacitances.

THEDAP system [3] presents a linear-matrix approach, with power/ground rails routed in the first metal layer between, instead of outside, the diffusion rows. The main advantages of this style are: (*i*) the possibility to create large transistors, with smaller impact in the silicon area (the transistor can be extended in the routing channel); (*ii*) the polysilicon length is reduced (as it has poor conductivity, it is important to avoid wires in this layer); (*iii*) the routing is simplified.

In submicronic technologies, cell and routing delay are comparable, justifying new cell synthesis and routing approaches. Several routing layers and stacked vias are also available, which increases the need to change the traditional concepts of layout style and routing. Nowadays, some technologies are using silicides over polysilicon (polycides) and diffusion, increasing its conductivity, allowing local wiring with polysilicon. Figure 1 presents a linear-matrix style adapted for submicronic technologies.

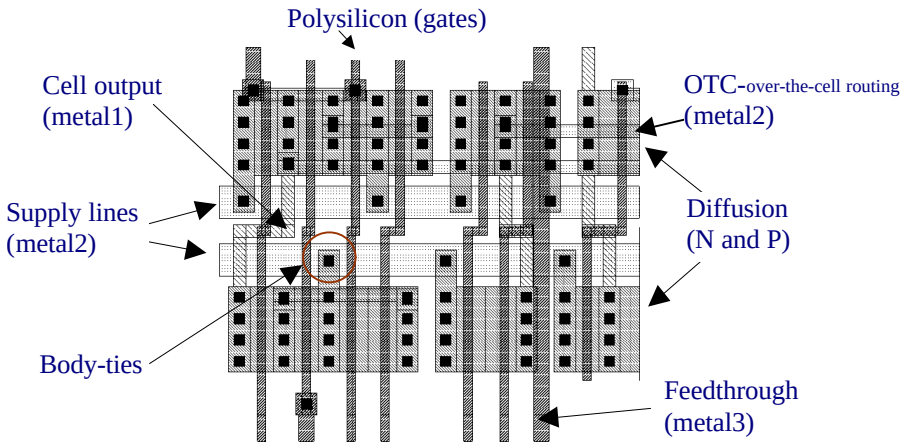


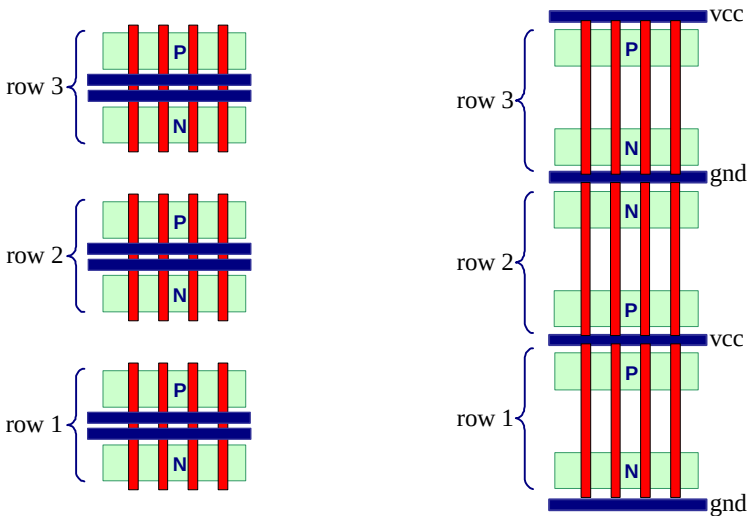
Figure 1 – Linear-matrix style, modified to use three metal layers and stacked contacts

The main features of this style are: over-the-cell routing (placing stacked contacts over drains/sources), supply lines between transistors (with body-ties placed over supply lines), four layers for routing between rows (vertical polysilicon/metal3 and horizontal metal1/metal2), gridless routing, metal3 over cells connecting nets at non adjacent rows. If silicides are used over diffusion, it is possible to minimize the number of contacts in drain/source regions, reducing the constraints imposed to the OTC routing.

3. Densities

Different silicon foundries present different characteristics, each one claiming the higher density. Our approach consists in determining if these density measures are reliable, and defining the methods to evaluate them from basic design rules.

The target layout style is the multi-row linear-matrix, currently used by layout generators like TROPIC [4] and LAS [5]. Cells are placed into parallel rows, similarly to the cell-based approach. Two main topologies can be used for the multi-row approach (Figure 2): channel-based and abutted rows, depending on the adopted strategy for supply routing. The channel-based approach considered is the one with supply lines placed between transistors, instead at the top and bottom sides of the cell [3].

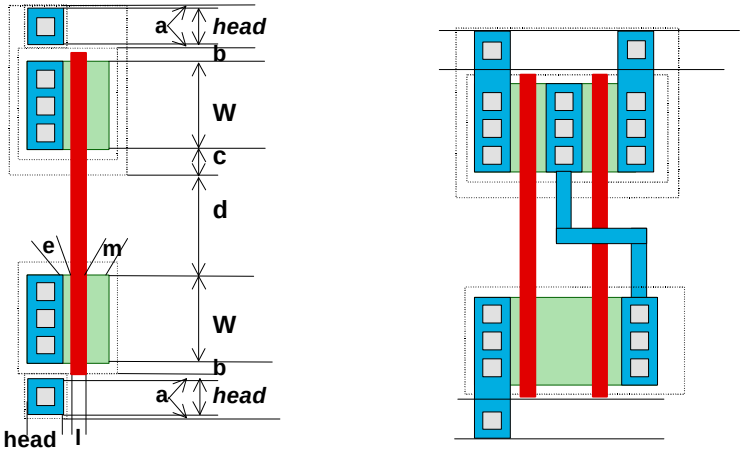


(a) Channel-based (b) Abutted rows

Figure 2 – Multi-row approach

Without routing, the final area of a circuit is not only the transistor surface. The layout height is defined by the minimal distance between 'N' and 'P' transistors and the placement of the body-ties (contacts to the substrate). The layout width is defined by the contacts to the drains and sources and the minimal diffusion area. As our objective is to evaluate the maximum transistor density, we choose the abutted-row topology, since in this approach the supply lines are not duplicated.

For an N-well process, the layout for a transistor pair (like an inverter) is presented in Figure 3a. In this layout, the connection in the drain region is implemented with contacts and by abutment in the source region. In this way, the next transistor pair uses the same area than the first one, as shown in the NAND gate (Figure 3b).



(a) Transistor pair (b) NAND gate

Figure 3 – Linear-matrix layouts

The area of a linear-matrix cell, without routing, is a function of the following design rules:

- **a**: spacing between diffusion and implant, for body-ties;
- **b**: spacing between diffusion and implant, for transistors;
- **c**: spacing between P diffusion and N well;
- **d**: spacing between N diffusion and N well;
- **e**: diffusion contact spacing to polysilicon gate minus overlap of contact;
- **l**: transistor length (defines the technology);
- **m**: maximum of (spacing between two polysilicon gates) and (½.head + e);
- **head**: contact width plus 2 times the overlap of contact;
- **width**: average transistor width in the circuit.

For the abutted-row approach, the transistor density (tr/mm^2), for *n* rows is:

- **Transistor number**: $tr\# = 2.n$, (a transistor pair per row);
- **Layout height**:
  $height = n.(c + d + 2.w) + (n-1).(head + 2.a + 2.b) + 2.(head + 2.a + b)$
  $height = n.(c + d + 2.w + head + 2.a + 2.b) + head + 2.a$
- **Layout width**: $width = head + e + l + m$

$$density = \frac{Tr\#}{height * width} \quad \textbf{(1)}$$

The previous equation gives the *optimal* density for the linear-matrix style. Two important points must be analyzed, to properly evaluate densities: routing area and transistor size.

The routing area, for design rules with two metal layers, corresponds to 50% of the chip area. For design rules with three metal layers, it is possible to aggressively use the over-the-cell routing [6] to reduce the routing area. There is no consensus in how much area is used in these processes for routing. However, if we consider that it is possible to reduce to half the routing tracks, we can assume the routing area in designs with three metal layers equal to 25% of the chip area. The approach "zero routing footprint" [7] (no routing area) is possible because the cells are designed to be transparent to several routing tracks, increasing in this way the cell height. When more than three metal layers are available for routing, they must be used to connect different blocks, not inside them (transparency over blocks).

The circuit can be sized for smaller area (and power), for smaller delay, or for some trade-off between these functions [8][9]. Figure 4 shows the relationship between delay and transistor width.

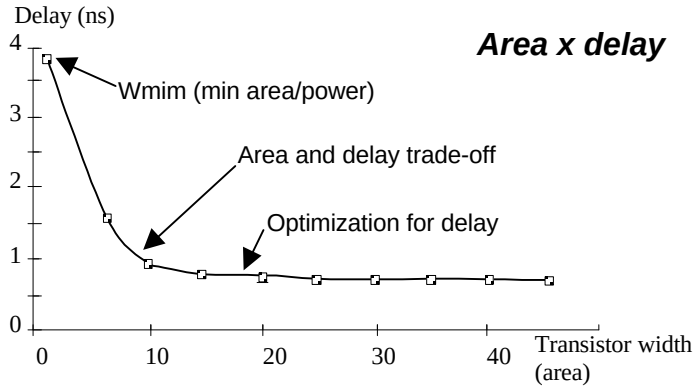


Figure 4 – Area and delay trade-off

Sizing transistors at the minimum width results in smaller area and "power", with delay controlled by loads (diffusion and routing capacitance). For larger width (w>16μm) the delay is almost independent of the load variations and is dominated only by the diffusion capacitance, proportional to the transistor width. Intermediate solution can be obtained optimizing the delay with locally reasonably sized transistors (area-delay trade-off). Currently, standard-cells are sized for delay optimization, since the load supplied by the cells is unknown. Some libraries have different templates for each cell, one for each optimization cost function. The price is the cost to manage and maintain such libraries. This constraint does not exist for layout synthesis tools: each cell can be sized according to the output load.

Simulating several circuits, with different design rules, we may set the sizing for area with transistors sized to the minimum width (*Wmin*), trade-off area-delay to *8Wmin* and optimization for delay to *16Wmin*. The tables 1 to 3 show the transistor density for 3 design rules [10]: ECPD07 (l=0.8 μm), ECDM05 (l=0.6 μm) and ECAT05 (l=0.5 μm). The first column corresponds to the three sizing solutions. The column "maximum density" corresponds to the densities obtained from equation 1; the third and fourth columns take into account the routing area when two and three metal layers are available for routing (maximum density multiplied by 0.5 and 0.75 respectively).

	Max. Density (tr/mm^2) *Diffusion at the minimum distance*	Density for two metal layers (tr/mm^2)	Density for three metal layers (tr/mm^2)
Area (Wmin)	32393	16197	24295
Area-Delay (8xWmin)	16592	**8296**	12444
Delay (16xWmin)	10653	5327	7990

a=0.4, b=1.0, c=2.4, d=4.8, e=0.2, l=0.8, m=1.2, head=2.0, Wmin = 1 (unites in μm*)*

Table 1- Densities (tr/mm^2) for the ECPD07 process, two metal layers

	Max. Density (tr/mm^2) *Diffusion at the minimum distance*	Density for two metal layers (tr/mm^2)	Density for three metal layers (tr/mm^2)
Area (Wmin)	57589	28795	43192
Area-Delay (8xWmin)	29497	**14748**	22123
Delay (16xWmin)	18039	9469	14204

a=0.3, b=0.75, c=1.8, d=3.6, e=0.15, l=0.6, m=0.9, head=1.5, Wmin = 0.75 (unites in μm*)*

Table 2- Densities (tr/mm^2) for the ECDM05 process, two metal layers

	Max. Density (tr/mm^2) *Diffusion at the minimum distance*	Density for two metal layers (tr/mm^2)	Density for three metal layers (tr/mm^2)
Area (Wmin)	85561	42781	64171
Area-Delay (8xWmin)	41775	20888	**31332**
Delay (16xWmin)	26349	13180	19769

a=0.3, b=0.6, c=1.5, d=3.0, e=0.1, l=0.5, m=0.7, head=1.2, Wmin = 0.7 μm *(unites in* μm*)*

Table 3- Densities (tr/mm^2) for the ECAT05 process, three metal layers

The difference between ECPD07 and ECDM05 processes is a shrinking factor 0.75, i.e. all design rules of the ECPD07 process are multiplied by 0.75 to obtain ECDM05. Consequently, for the same transistor number, the area was reduced by 0.56 (0.75^2) and the density increased by 1.78 ($1/0.75^2$). Applying a shrinking factor 0.7 in the ECAT05 process gives the density of the 0.35μm CMOS process. Prediction for the 0.25 μm and the 0.18 μm can be obtained in the same way. With a shrinking factor 0.7, the transistor density is multiplied by 2 ($1/0.7^2$). From this, we can extrapolate the densities for different processes (table 4), using as reference the density with three metal layers for routing, sizing area-delay optimization, from the process ECAT05.

Process (μm)	Transistor Density	Gate Density
l=0.8	8296	2.1
l=0.6	14748	3.7
l=0.5	31332	7.8
l=0.35	62644	15.7
l=0.25	125328	31.3
l=0.18	250656	62.7

Table 4- Roadmap for densities, gate density in kgates/mm^2 (one equivalent gate is equal to 4 transistors)

These data can be used to:

- Evaluate the quality of layout synthesis tools. Comparing the layouts obtained from a layout generator to the values obtained from the tables, area used for routing can be measured.
- Predict the area to be used by a circuit, from the number of equivalent gates and the design rules. This information can be very useful for logic synthesis tools to evaluate the area for a given cost function.
- Analyze the trends in terms of transistor area and transistor count for the next generation of integrated circuits. For example, applying twice a shrinking factor 0.7 over the 0.18 μm process

results in transistors with l=0.08 μm, with 1,000,000 transistors per square millimeter. Moreover, if it is possible to have an IC with 10 cm^2, we will have a circuit with 1 billion transistors!

4. Results

Table 5 shows the densities obtained from different benchmarks implemented with layout synthesis tools (LAS and TROPIC). The main characteristic of these tools is the absence of pre-characterized cells, with on-the-fly cell generation. The columns "Tr#" and "Rows#" indicate the number of transistors in each circuit and the number of rows used to implement the final layouts respectively. The columns LAS and TROPIC2 give the densities for layouts synthesized with two metal layers for routing. TROPIC3 is the TROPIC version for routing with three metal layers [11]. The circuits were generated using the process ECPD07, with transistors sized for the area-delay trade-off.

Circuit	Tr#	Rows#	LAS	TROPIC2	TROPIC3
Adder	28	2	4780	5942	12111
Addergate	40	2	5598	6020	11176
Alu	260	4	5812	5294	9061
Alugate	432	4	6050	5405	8592
Rip	448	5	5961	5610	10207
Cla	528	5	5614	5498	9146
Hdb3	570	4	4903	6516	7758
Mult6	972	7	5950	5779	7994
Mult2	4512	16	4879	4080	7057
Average transistor density:			***5505***	***5571***	***9233***
Estimated transistor density (from table 1)			*8296*	*8296*	*12444*

Table 5- Transistor densities (tr/mm^2), process ECPD07, w=8μm, l=0.8μm

We observe a more important contribution of the routing area than expected comparing the densities obtained from layout synthesis with the predicted ones. As mentioned before, the routing area should be 50% and 25% for 2 and 3 metal layers respectively. We got 66% and 44%. This difference can be explained by several reasons, at the layout level:

- over-the-cell routing is not used in the two metal process;
- stacked contacts are not allowed (two-metals process);
- the great number of contacts in each drain/source prevents routing over the transistors;
- large supply lines increase the circuit height;
- as the area of the block corresponds to its bounding box, rows with different length will produce empty spaces in the layout;
- the cells are separated by diffusion gaps, increasing the circuit width;
- the length of the cells will increase when diffusion gaps are necessary;
- the compactor used to translate the symbolic description into layout also plays an important role, for example, if jogs are automatically inserted the final area will be reduced.

Thus, the layout synthesis tools must improve the efficiency of the layout style, placement and routing, in order to achieve the estimated densities, when routing is considered.

Table 6 shows the densities directly evaluated on standard-cell layout (processes ECPD07 and ECAT05.). The densities for all cells from these libraries are 12600 and 54000 tr/mm^2 respectively. These density values are not different from the estimated solution for area with routing (16197 and 64171 tr/mm^2, tables 1 and 3). This is possible because the cells of a library are handcrafted (full-custom design), with no pre-defined position for transistor placement and no preferential direction for the routing layers. At the cell level, the effect of routing is quite minimized.

CELL	TR#	Density
INV	2	3008
NAND2	4	5906
AOI22	8	11363
AND4	10	12449
DFF	29	14441

CELL	TR#	Density
an02d0	6	46875
an13d1	10	44643
an04d2	12	53571
aor21d4	14	48611
nr02d4	16	71429
aoi222d2	18	56250
lachq1	20	48077
aoi2222d2	24	57692
dfnrb1	26	42763
labhb1	28	51471
xr03d4	28	58333
dfcrn2	30	49342
denrq1	34	53125
jkbrb2	46	53241
sdbrb2	52	56034

Table 6- Standard-cell densities, process ECPD07 and ECAT05

Table 7 shows the densities obtained for circuits generated with standard-cells, using place and route tools. The routing area for these circuits corresponds to 56% of the total area, close to the predicted values.

Circuit	Tr#	Gates#	Density
Registre	242	60	6454
Mesure	3545	861	7082
Valid	5993	1498	7758
Top	6417	16	7627
Average transistor density:			***7230***

Table 7 - Density (tr/mm^2) for circuits generated from standard-cells, process ECPD07

The advantage of standard-cells over automatic layout generation consists in the possibility of optimizing complex functions at the cell level, such as flip-flops and adders. For example, the circuit "valid" has 35% of the transistors used to implement flip-flops. When flip-flops are generated using the linear matrix style, all transistors must be placed horizontally and the internal connections must respect the preferential directions for routing. On the other hand, when flip-flops are designed by hand, transistors can be freely placed, with no restrictions for the internal routing.

To automate the creation of cell layouts competitive with handcrafted quality layouts is a non-trivial problem. To date, the real complexity of the cell synthesis process has been the main obstacle to its commercial success [12]. Besides the problem of spending more area than standard-cells, the automatic layout synthesis is a very attractive solution to synthesize random logic blocks. Four main reasons can be advanced for this:

- The number of cells available into a library is quite small when compared to the possibilities given by using *complex gates*. The number of different functions that can be implemented with complex gates having 4 or less serially connected transistors is 3503 [13]. As showed in [14], it is possible to reduce the overall number of transistors up to 40%, reducing in this way area, delay and power.
- Technology independent synthesis is vital in supporting rapidly advancing processes [15]. Layout generation permits an *easy technology migration*, allowing to quickly change from a set of design rules to another.
- Transistors can be individually sized according to the designer constraints. One interesting feature of the layout synthesis is the possibility to adjust the delay of the critical path(s), downsizing the transistors outside the critical path, in order to achieve *power-delay optimization*.
- The cost for cell and library maintenance is very expensive, with a *virtual library* concept (no pre-characterized cells) this cost is suppressed.

As mentioned before, the density for automatic generated blocks can be improved by adding some new features to the layout style. Examples of features are: smaller number of contacts at drain/sources (if silicides are used); stacked vias to connect non-adjacent layers (e.g. diffusion and metal2); polysilicon wires for small distances (again if silicides are used); routing over-the-cells and over-the-transistors; rows with equal length; three metal layers for routing; layout compactor adapted to the layout style.

5. Conclusion

The contribution of this paper is the definition of the method to estimate silicon area to implement layouts, starting from the number of transistors and the layout design rules. Routing area was considered as a constant percentage of layout area, 50% and 25% for two and three metal layers respectively. Three sizing possibilities were considered: area, area-delay and delay optimization. The data obtained from this method can be used to evaluate layout synthesis tools, to estimate area in logic synthesis tools and to predict the densities of new processes. The difficulty for CAD developers is to know how far the heuristic algorithms for cell synthesis, placement and routing are from the optimal. The values presented can be used as an upper bound for transistor densities, driving heuristic algorithms to give results near to predicted values. Considering the number of routing layers available in actual processes (6 or 7), great effort must be given in defining strategy for local and inter-block connection and signal propagation. This strategy must support fabrication constraints as well as signal integrity preservation.

6. References

[1] T.Uehara, W.van Cleemput. "*Optimal Layout of CMOS Functional Arrays*". IEEE Transactions on Computers, Vol. C-30, No. 5, May 1981, pp. 305-312.

[2] A.D.Lopez, H.S.Law. "*A Dense Gate-Matrix Layout for MOS VLSI*". IEEE Transactions on Electron Devices, Vol. ED-27, No. 8, August 1980, pp. 1671-1675.

[3] C.Hwang; Y.Hsieh.; Y.Lin; Y.Hsu. "*An efficient layout style for two-metal CMOS leaf cells and its automatic synthesis*". IEEE Transactions on CAD, Vol. 12, No 3, March 93, pp. 410-423.

[4] F.Moraes, N.Azemard, M.Robert, D. Auvergne. "*Flexible macrocell layout generator*". 4th ACM/SIGDA Physical Design Workshop, Los Angeles, 1993, pp. 105-116.

[5] CADENCE. "*Virtuoso Layout Synthesizer - LAS - User Guide*". CADENCE™ Version 4.2, October 1991.

[6] J.Kim; S.M.Kang. "*A New Triple-Layer OTC Channel Router*". IEEE Transactions on CAD, Vol. 15, no. 9, September 1996, pp. 1059-1070.

[7] N.Sherwani. " *Algorithm for VLSI physical design automation*". Kluwer Academic Publisher, 538p. 1995.

[8] S.S.Sapatnekar; S.Kang. "*Design Automation for Timing-Driven Layout Synthesis*". Kluwer Academic Publisher, 296p. 1993.

[9] S.Turgis, N.Azemard, D.Auvergne. "*Design and selection on buffers for minimum power-delay product*". ED&TC'96: The European Design & Test Conference, Paris, France, March 11-14, 1996, pp. 224-228.

[10] ATMEL-ES2. ECAT05 and ECPD07 design-rules, 1998.

[11] F.Moraes; R.Reis; F.Lima. "*An Efficient Layout Style for Three-Metal CMOS Macro-Cells*". VLSI'97, Gramado (Brazil), pp. 415-426.

[12] M.Lefebvre; D.Marple; C.Sechen. "*The Future of Custom Cell Generation in Physical Synthesis*". DAC'97.

[13] E.Detjens; G.Gannot; R.Rudell; S.Vincentelli; A.Wang. "*Technology mapping in MIS*". ICCAD, Santa Clara, 1987, pp. 116-119.

[14] A.Reis; M.Robert; D.Auvergne; R.Reis. "*Associating CMOS Transistors with BDD Arcs for Technology Mapping*". Electronic Letters, Vol. 31, No 14, July 1995.

[15] M.Guruswamy; R.L.Maziasz; D.Dulitz; S.Raman; V.Chiluvuri; A.Fernandez; L.G.Jones. "*CELLERITY: A Fully Automatic Layout Synthesis System for Standard Cell Libraries*". DAC'97.

7. Aknowledgement

We gratefully acknowledge the discussion with engineers of ATMEL ES2 who have contributed to complete data given in this paper.

The author Fernando Moraes would like to gratefully acknowledge the continued support of the "Conselho Nacional de Desenvolvimento Científico e Tecnológico" (**CNPq**), Brazil, to this work. Research grant 522939/96-1.