A General-Purpose High-Density Sea-of-Gates Architecture

R.J.H. Koopman and H.G. Kerkhoff

MESA Research Institute, University of Twente

P.O.Box 217, 7500 AE Enschede, The Netherlands

Abstract - This paper describes an efficient and flexible CMOS sea-of-gates architecture for digital applications. This architecture supports the design of integrated circuits at all important physical design areas such as performance, implementation and wiring. After a detailed discussion of the three architectural sub-levels performance, implementation and wiring, the sea-of-gates architecture will be presented. The functionality of this architecture will be illustrated by describing the design of some benchmark circuits.

1. Introduction

The sea-of-gates design style is based on the use of a prefabricated silicon chip. This chip, also called sea-of-gates master array, contains a regular array of transistors without defining wiring areas in advance. Within this array the dimensions, the locations and the arrangement of the array transistors are fixed. Wiring is put on top of the transistors to personalise the integrated circuit. Using the sea-of-gates design style has several advantages such as reducing the time and cost to produce the integrated circuit. The main disadvantage of this design style is that physical design is bound by the fixed sea-of-gates architecture.

Over the past years, many sea-of-gates architectures have been presented [1-6]. The goal of all these architectures is, in one way or another, to increase the flexibility and efficiency of the physical design process. Architectural features implemented to improve this flexibility and efficiency of the physical design process have been, among others:

- optimising towards maximum connectability and transparency [1-3]
- using a common-gate architecture to increase memory density [4]
- adding extra nMOS devices to support dynamic and memory structures [1-3,5]
- supporting a two-dimensional design approach [1-3,6].

The growing complexity of VLSI will lead to some major architecture characteristics of future VLSI systems:

- within one single chip design more distinct functions including analog functions
- the distinction between various application areas will become fuzzy
- increase of wire density.

To implement these complex integrated circuits with the use of a sea-of-gates array, more fundamental features must be incorporated in the sea-of-gates architecture. In this paper a sea-of-gates architecture will be presented which is based on these expected characteristics of future VLSI systems. This architecture supports physical design at all important design areas such as performance, implementation and wiring.

2. Architectural Considerations

In this section some items will be discussed concerning the sea-of-gates architecture. These have been used to design the architecture. The architecture will be divided into three different architectural aspects: performance, implementation and wiring.

2.1 Performance

The primary performance parameters of integrated circuit design are:

- satisfying speed requirements
- minimising power dissipation
- minimising silicon area.

In the following sections these performance parameters will be discussed. The results will subsequently be used to design the sea-of-gates architecture.

2.1.1 Speed

The speed of a circuit will be affected by several capacitances. These capacitances are:

- the parasitic transistor capacitance
- the intra-cell wiring capacitance
- the inter-cell wiring capacitance
- the transistor gate capacitance of fanout gates.

The total capacitance at the output node of a gate consists of two different contributions, the transistor gate capacitance of fanout gates and the inter-cell wiring capacitance. First, the influence of the transistor gate capacitance of fanout gates at the output node on speed will be investigated. Simulation results will be given to show the influence of this capacitance on speed. Figure 1 shows the propagation delay of a CMOS 2-input nand-gate ($W_p = (11.5/6.5) \cdot W_n$) as function of the transistor width. Inverters are used as fanout gates.

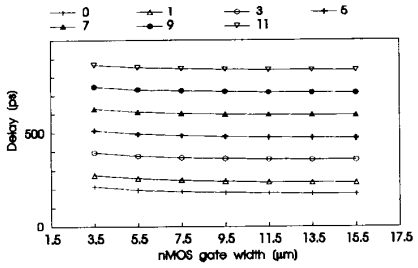

Figure 1. Relation between propagation delay and transistor width if only the fanout transistor gate capacitance is taken into account. (The # of fanout gates is the parameter.)

If the propagation delay is corrected for intrinsic gate delay and plotted as function of fanout, the results of Figure 2 are achieved.

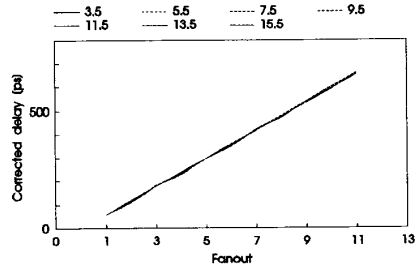

Figure 2. Propagation delay corrected for intrinsic gate delay plotted as function of fanout. (The nMOS transistor width is the parameter.)

0-7803-0510-8/92$03.00 ©1992IEEE

The main conclusion that can be drawn from Figure 2 is that a large fanout does not require large-sized transistors in the gate driving the fanout.
Next, the influence of a combination of inter-cell wiring capacitance and fanout transistor gate capacitance will be discussed. To show the influence of this combination, the following experiments have been carried out. The minimal transistor width has been estimated that will be necessary to set the propagation delay within 5% (Table 1) and 10% (Table 2) of a reference propagation delay depending on inter-cell wiring capacitance and fanout. Again a 2-input nand-gate has been used with inverters as fanout gates. The reference delay is the delay of a nand-gate without inter-cell wiring capacitance but with fanout (same fanout that apply to column in question). The nMOS transistor has a width W_n = 15.5 µm. This large width has been chosen to minimise influence of intrinsic gate delay.

Table 1. Required nMOS transistor width to set the delay within 5% of the reference gate delay ($W_p = (11.5/6.5)\cdot W_n$).
(left column: inter-cell capacitance; uppermost row: fanout)

	#1	#3	#5	#7	#9	#11
5fF	>10.5	8.5	7.5	5.5	5.5	4.5
10fF	"	10.5	9.5	7.5	6.5	5.5
15fF	"	>10.5	>10.5	8.5	7.5	7.5
20fF	"	"	"	10.5	9.5	8.5
25fF	"	"	"	>10.5	10.5	9.5
30fF	"	"	"	"	>10.5	10.5
50fF	"	"	"	"	"	>10.5
75fF	"	"	"	"	"	"
100fF	"	"	"	"	"	"

Table 2. Required nMOS transistor width to set the delay within 10% of the reference gate delay ($W_p = (11.5/6.5)\cdot W_n$).
(left column: inter-cell capacitance; uppermost row: fanout)

	#1	#3	#5	#7	#9	#11
5fF	7.5	5.5	4.5	3.5	3.5	3.5
10fF	9.5	6.5	5.5	4.5	3.5	3.5
15fF	>10.5	8.5	6.5	5.5	4.5	4.5
20fF	"	10.5	7.5	6.5	5.5	4.5
25fF	"	>10.5	8.5	7.5	6.5	5.5
30fF	"	"	9.5	7.5	6.5	5.5
50fF	"	"	>10.5	>10.5	9.5	8.5
75fF	"	"	"	"	>10.5	>10.5
100fF	"	"	"	"	"	"

Some interesting conclusions can be drawn when looking at the tabulated results. First, for a large fanout delay is mainly determined by capacitance of fanout gates. Small-sized transistors can be used in these situations. Second, in situations with a small fanout, delay is dominated by intra-cell and inter-cell capacitances. Third, in situations where inter-cell capacitance dominates the total capacitance at the output node, this inter-cell capacitance determines the delay. To improve speed in the second and third cases, very large-sized transistors have to be used. However, several other methods also exist to reduce delay without using very large-sized transistors:

- placing transistors in parallel
- enlarging the drive capability of the output stage
- inserting buffers to drive large inter-cell capacitances.

The strength of placing transistors in parallel will be shown by presenting the results of some experiments. Table 3 shows the minimum required nMOS transistor width to set the delay of a 2-input nand-gate, implemented with single transistors, within 5% of the delay of a 2-input nand-gate implemented with parallel transistors. Both nand-gates are loaded with one inverter and inter-cell capacitance as indicated in Table 3. The widths of the nMOS and pMOS transistors in the nand-gate with parallel transistors are: W_n=6.5 µm and W_p=11.5 µm.

Table 3. Required nMOS transistor width to set delay within 5% of reference delay ($W_p = (11.5/6.5)\cdot W_n$).
(first row: inter-cell capacitance [fF]; second row: gate width [µm])

10	15	20	25	30	50	75	100
73	40	30	26	23	18.5	16.5	15.5

The results of this section show that small-sized transistors can be used to meet the speed requirements.

2.1.2 Power Consumption

The switching or dynamic power dissipation consists of two components, the short-circuit dissipation and the charge-discharge dissipation.

Short-Circuit Dissipation

The CMOS inverter circuit will be used to discuss the dynamic power. During a transition at the input of the inverter there will be a period of time in which $V_{ss} + |V_{t,n}| < V_i(t) < V_{dd} - |V_{t,p}|$, where $|V_{t,n}|$, $|V_{t,p}|$ and $V_i(t)$ are the nMOS threshold voltage, the pMOS threshold voltage and the input voltage respectively. In this period both the nMOS transistor and the pMOS transistor will conduct, causing a conducting short-circuit path from supply to ground.
Suppose the input voltage $V_i(t)$ changes from V_{dd} to V_{ss}. During this transition the pMOS transistor will conduct the short-circuit current and the charge current to charge the capacitance at the output node. The nMOS transistor will only conduct the short-circuit current. Therefore, the analysis will concentrate on the current through the nMOS transistor. During this transition the nMOS transistor will operate in the regions as tabulated in Table 4. The short-circuit current is also tabulated in Table 4.

Table 4. Short-circuit current through nMOS transistor.
(V_o : output voltage; β_n : transistor gain factor)

voltage relations	region of operation of nMOST	short-circuit current $I_{scc}(t)$
$V_i(t) > V_{dd} - \|V_{t,p}\|$	linear	0
$\|V_{t,n}\| < V_i - V_o$	linear	$\beta_n\left((V_i(t) - \|V_{t,n}\|)\cdot V_o(t) - \frac{1}{2}\cdot V_o(t)^2\right)$
$\|V_{t,n}\| \geq V_i - V_o$	saturation	$\frac{1}{2}\cdot\beta_n(V_i(t) - \|V_{t,n}\|)^2$
$V_i(t) < V_{ss} + \|V_{t,n}\|$	cut-off	0

The short-circuit current and, therefore, the short-circuit dissipation depends besides circuit and process characteristics, on transistor dimensions $(W/L)_n$ and $(W/L)_p$ [7]. Reducing these transistor dimensions will reduce the short-circuit power dissipation.

Charge-Discharge Dissipation

During a transition at the output node from '0' to '1', a charge $C_L\cdot V_{dd}$ is placed on the output node, with C_L being the total capacitance at the output node. A corresponding charge current flows from the power-supply through the p-transistor to the output node. When the output node switches from '1' to '0', this charge $C_L\cdot V_{dd}$ is transferred to ground. A corresponding discharge current flows from the output node to the ground node through the n-transistor. The net result is a transfer of charge $Q_L = V_{dd}\cdot C_L$ from power-supply to ground. If the output switches at a frequency f Hz, then a charge $f\cdot Q_L$ is transferred every second, and the average current from V_{dd} to V_{ss} is $f\cdot V_{dd}\cdot C_L$. The average power dissipation is $f\cdot V_{dd}^2\cdot C_L$. The total capacitance C_L at the output node and can be split in:

- the inter-cell wiring capacitance C_w
- the total fanout transistor gate capacitance C_g

C_g can be written as:

$$C_g = C_{g\square}\cdot N(W_nL_n + W_pL_p)$$

with $C_{g\square}$ being the average gate capacitance per unit area of the

transistor channel and $N(W_nL_n + W_pL_p)$ the total area occupied by the nMOS and pMOS transistors of fanout gates [8]. Using these expressions the charge-discharge power dissipation can be written as:

$$f V_{dd}^2 \cdot (C_w + C_{g\square} \cdot N(W_nL_n + W_pL_p))$$

This expression shows that reducing the transistor dimensions will also reduce the charge-discharge power dissipation.

2.1.3 Area

The area of a circuit is strongly related to the area of the transistors. The area occupied by a transistor is given by:

$$A_t = (F + W) \cdot (D_s + D_d + L)$$

where F accounts for area overhead due to the polysilicon gate extension and power-supply or ground facilities, W is the transistor gate width, D_s is the source extension, D_d is the drain extension and L is the transistor gate length.

The raw transistor density is related to the transistor area by:

$$N_t = 1/A_t$$

To be able to feature a transistor densities as high as those defined by the used technology, the transistor layout parameters W, L, D_s and D_d should be reduced to increase N_t. Reducing D_s and D_d will also reduce parasitic transistor capacitances.

2.2 Implementation

During the metallisation phase the circuits are constructed. A wide range of logic styles, circuit structures and clocking strategies are available to implement the required function. All these logic styles and structures require specific device ratios and counts. Providing several device ratios and counts within the sea-of-gates architecture will have two effects. First, logic styles and circuit structures will benefit if required device ratios and counts are exactly available. Second, in situations where not all the available devices will be used, the overall circuit density will be reduced. With respect to this implementation it is very important that the sea-of-gates architecture provide a well-balanced trade-off between specialism and density.

2.3 Wiring

Wiring will be split in five levels:

- connectability of transistors, the number of physical locations to contact transistor terminals
- transparency of intra-cell wiring, the space on top of and next to transistors available for intra-cell wiring
- connectability of gate terminals, the accessibility of inputs and outputs of logic gates
- transparency of inter-cell wiring, the space available for inter-cell wiring
- power wiring, the facilities for power routing.

The main characteristics concerning wiring are the number of horizontal and vertical available wiring tracks of a core-cell. These number of tracks have impact on all wiring levels. Several strategies have been used to estimate these number of tracks: concentrating on intra-cell requirements [5,9], on intra-cell and inter-cell requirements [4,10] and on optimising towards maximum connectability and transparency characteristics [1].

In situations requiring more wiring space, wiring channels must be inserted to create the required wiring space. In [3] it has been shown that a small granularity will result in more efficient designs. Standard-cell and full-custom design are very efficient with respect to wiring because needed area for wiring can exactly be adjusted. This is because these design styles possess the availability of a granularity of one.

3. Sea-of-Gates Architecture

In this section the resulting sea-of-gates architecture will be described. The architecture is based on reducing the power consumption and suppressing redundant transparency, connectability and specialised devices. Furthermore, the architecture must possess the flexibility to be able to meet efficiently specific speed, device and wiring demands.

A high raw transistor density is a first requirement to create the necessary flexibility. As a consequence the transistors will be small-sized. This is also advantageous to reduce power consumption. Figure 3 shows the resulting architecture.

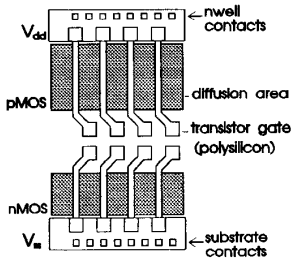

Figure 3. Sea-of-gates architecture.

The core-cell consists of one nMOS and one pMOS transistor. Adjacent core-cells (transistors) share common diffusion areas to minimise area. To be able to make contact between the first metal layer and the diffusion area of a transistor, at least space for one contact has to be available above the transistor diffusion area. To switch from the first metal layer to the second metal layer without using stacked contacts, space for a via between these two layers must also be available above the diffusion area. If there is no available area for this via, the via must be placed above the polysilicon transistor gate contact or the diffusion area of a nearby transistor. Both options are not desirable because they block the usage of a transistor. So the transistor gate has to be at least as wide as the space required for both these two contacts. These non-minimum sized transistors also show a better speed behaviour with respect to intra-cell capacitance. To adjust for the electrical difference between a p-channel and an n-channel transistor, the width of the pMOS has been made one contact position larger.

To reduce the height of the core-cells and thereby improving the granularity, the outside polysilicon gate contacts of the nMOS and pMOS transistors are located under the first metal V_{ss} and V_{dd} rails. These contacts can only be used to connect the transistors to the power-supply or ground for isolation purposes. A further reduction of the core-cell height has been achieved by sharing first metal power and ground rails between rows. As a consequence, adjacent core-cell rows will be mirrored. The n-well and substrate contacts are situated under these V_{dd} and V_{ss} rails.

Table 5 shows some architecture characteristics. To estimate the logic density, a 2-input nand-gate has been used. This gate occupies 2.66 core-cells including the use of isolation transistors.

Table 5. Architecture characteristics (1.0µm double-metal layer CMOS process).

core-cell dimensions	39.75µm x 5µm
width nMOS	6.5µm
width pMOS	11.5µm
length nMOS, pMOS	1µm
logic density	1.89k gates/mm^2
ROM density	6.7k cells/mm^2
sRAM density	628 cells/mm^2

4. Circuit Design Examples

To show the functionality of this sea-of-gates architecture, the design of some benchmark circuits will be described. These circuits are:

- static random access memory [11],
- systolic array circuit for multiplication of matrices,
- arithmetic and logical unit [11,12].

In all examples a 1.0 µm double-metal layer CMOS process has been used for the design of these circuits. No stacked contacts and vias are

allowed within this process.

4.1 Static Random Access Memory

The 6-transistor sRAM CMOS memory-cell is a very powerful example to test the wirability of the architecture. This because of the huge amount of wire segments (feedback connection, access transistor connection, word and bit lines) needed to implement a memory-cell. Furthermore, word and bit lines within one cell must be designed such, that these lines are connected by abutment within the memory array.
The sea-of-gates architecture allows several memory architectures to be implemented, all with the same density. This is very important because the memory architecture will not be determined by the sea-of-gates architecture, but by the demands of surrounding on-chip functions.
Figure 4 shows the possible memory-cell designs. The basic 6-transistor memory cell occupies two transistor rows. Word and bit lines are indicated in the figure.

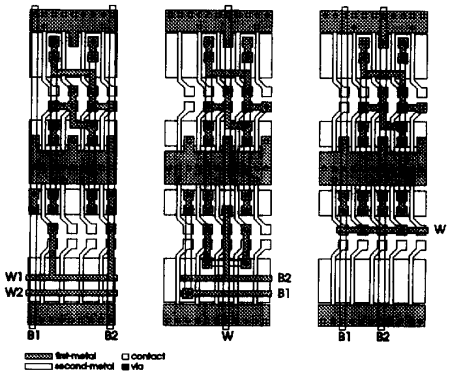

Figure 4. Memory-cell layouts.

4.2 Systolic Array Circuit for Multiplication of Matrices

This circuit performs the multiplication of two 2x2-matrices. The systolic array has been build using a basic processor element (BPE) several times with local connections between these BPEs. A BPE consists of four master/slave flip-flops and an arithmetic circuit to perform the multiplication and the addition. A two-phase nonoverlapping clock has been used. The two master clocks are distributed with local buffers to generate the complementary clock signals. The BPE has been designed to provide the necessary feed through tracks of data and clock lines. The area of the systolic array (7 BPEs) is 0.1 mm^2 and has a complexity of 150 logic gates (logic gate: a 2-input nand-gate consisting of 4 MOS transistors). The transistor utilisation of this circuit is 82% (inclusion of necessary isolation transistors) and 58% without any isolation transistor. Figure 5 shows the layout of a BPE.

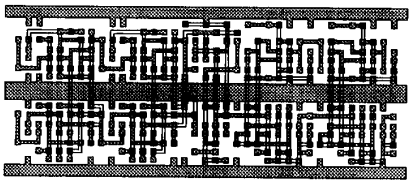

Figure 5. Basic processor element layout.

4.3 Arithmetic and Logic Unit

The 74F382 Arithmetic Logic Unit has been used to test a MSI-level circuit implementation of basic arithmetic and logic functions. The ALU has a complexity of 149 logic gates and the layout of the ALU is shown in Figure 6. This circuit layout shows a characteristic feature of circuit implementation. The required wiring area varies within the design. At places with a lot of interconnect lines, wiring boxes have been allocated. The area of the ALU is 0.13 mm^2 and the transistor utilisation is 59% (inclusion of necessary isolation transistors) and 45% without any isolation transistor.

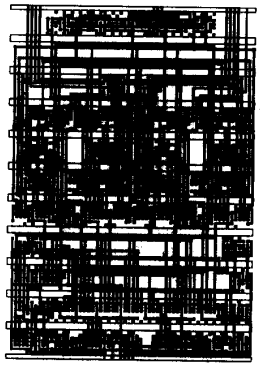

Figure 6. ALU layout (metal1 and metal2 only).

Conclusions

An improved CMOS general-purpose sea-of-gates architecture for digital applications has been presented. This architecture has been based on flexible and efficient physical design of complex VLSI systems. The basic requirements for flexible and efficient physical design have been discussed. The results have been used to design the sea-of-gates architecture. Benchmark circuits have been designed to show the functionality of the architecture.

Acknowledgements

This work has been carried out with the support of the Dutch Innovative Research Program (IOP IC-TEL 88.038).

References

[1] M.A.Beunder et al., "New Directions in Semicustom Arrays", IEEE Journal of Solid-State Circuits, vol.23, no.3, pp.728-735, June 1988.
[2] M.A.Beunder, "Design and Analysis of Semi-Custom Architectures", Ph.D. Thesis, Institute for Microelectronics Stttutgart, Germany, December 1989.
[3] G.Roos et al., "On Structured Gate Forest VLSI Design", Microprocessing and Microprogramming 27, pp.785-792, 1989.
[4] H.J.M.Veendrick et al., "An Efficient and Flexible Architecture for High-Density Gate Arrays", IEEE Journal of Solid-State Circuits, vol.25, no.5, pp.1153-1157, October 1990.
[5] P.Duchene and M.J.Declercq, "A Highly Flexible Sea-of-Gates Structure for Digital and Analog Applications", IEEE Journal of Solid-State Circuits, vol.24, no.3, pp.576-584, June 1989.
[6] R.J.H.Koopman et al., "Digital CMOS Sea-of-Gates Core Cells and Master Images", European Solid-State Circuits Conference, Grenoble, France, pp.245-248, September 1990.
[7] H.J.M.Veendrick, "Short-Circuit Dissipation of Static CMOS Circuitry and Its Impact on the Design of Buffer Circuits", IEEE Journal of Solid-State Circuits, vol.sc-19, no.4, pp.468-473, August 1984.
[8] A.Mukherjee, Introduction to nMOS and CMOS VLSI Systems Design, Prentice/Hall International, Inc., 1986.
[9] W.A.M.Van Noije and G.J.Declerck, "Advanced CMOS Gate Array Architecture Combining "Gate Isolation" and Programmable Routing Channels", IEEE Journal of Solid-State Circuits, vol.sc-20, no.2, pp.469-480, April 1985.
[10] D.W.Harberts et al., "Design for Routability of a High-Density Gate Array", IEEE ICCD, 1990.
[11] D.D.Hill and B.Preas, "Benchmarks for Cell Synthesis", IEEE Design and Automation Conference, pp.317-320, 1990.
[12] V.P.LaBuda, "Jedec Simplifies Selection of BiMOS, CMOS Gate Arrays", Electronic Design, March 17, pp.107-115, 1988.