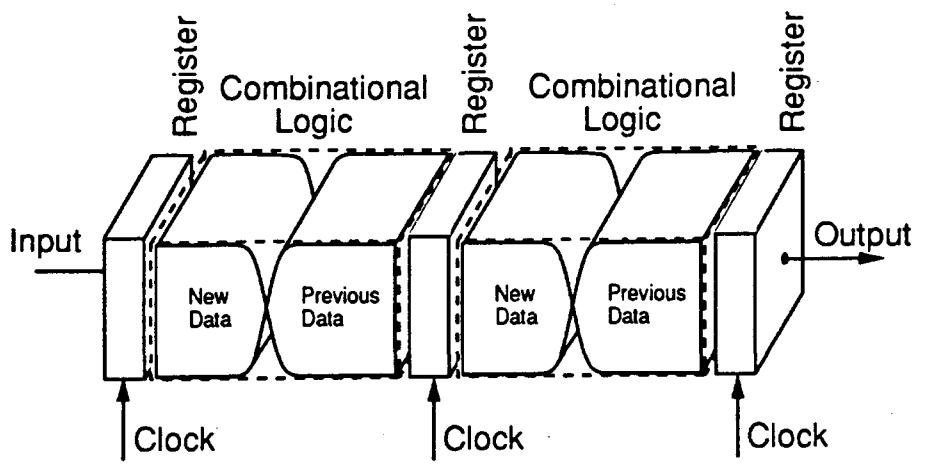
Pushing the Limits of CMOS Technology: A Wave-Pipelined Multiplier

Fabian Klass, Michael Flynn, and Ad J. van de Goor
Stanford University
Delft University

Conventional Pipelining

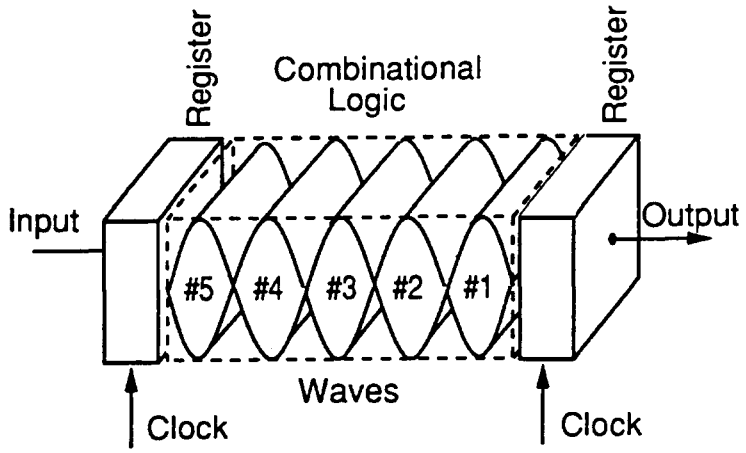
New data are applied after previous data are available:



Pipeline rate is limited by the maximum delay of one logic section + the clocking overhead

Wave Pipelining

New data are applied before previous data are available:



Pipeline rate is limited by the difference between maximum and minimum delay accross the logic section + the clocking overhead

Advantages of Wave Pipelining

- Achieve very high pipeline rates, approaching the physical speed limit of the technology
- Increase pipeline rate without significant latency increase
- Minimize clocking overhead
- Use fewer registers thus saving area
- Reduce clock distribution problems

Disadvantages of Wave Pipelining

- Require path delay equalization to maximize pipeline rate
- Clock signals must be synchronized with the waves
- Process and temperature variations are more critical than in conventional systems
- Difficult to handle at system level

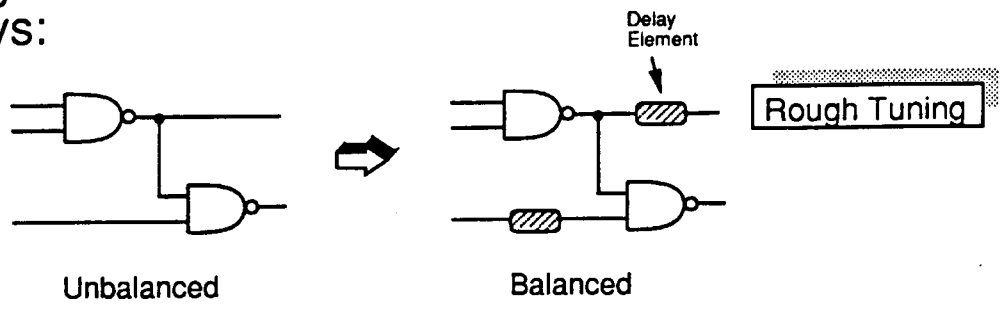
Delay Variations

Sources of delay variation in digital circuits :

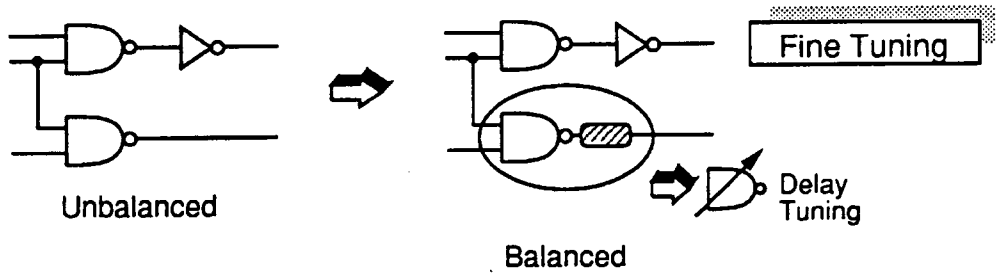
- Difference in path lengths (gate level)
- Different input patterns (transistor level)
- Coupling capacitance effects
- Power supply changes (noise)
- Fabrication process variations
- Temperature changes

Path Equalization Techniques

Delay elements are inserted to increase short path delays:

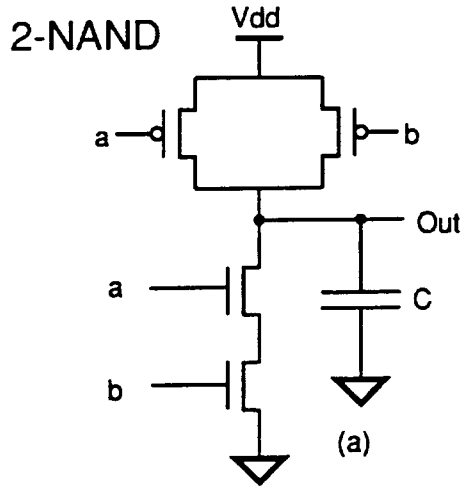


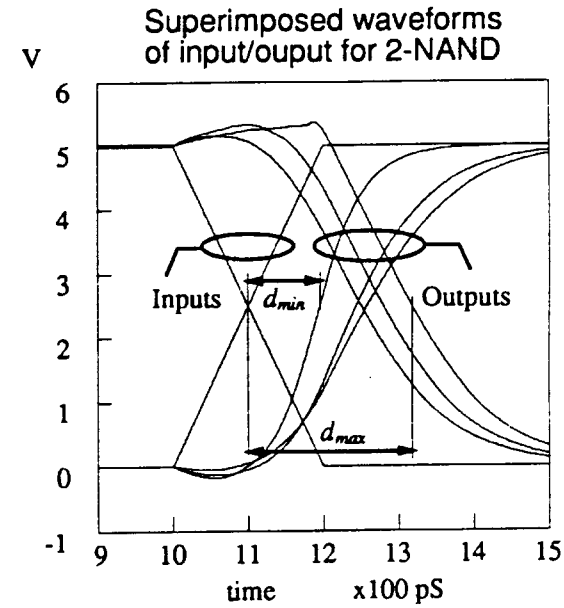
Gate delays are tuned to equalize path delays:



Delay in CMOS

The delay of a CMOS gate varies substantially with different input patterns:

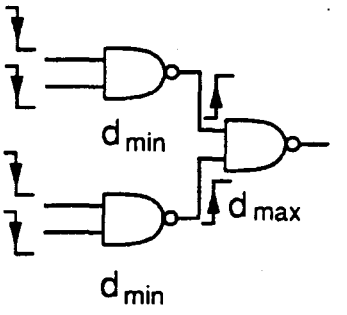




CMOS Wave Pipelining

By using proper design constraints, e.g., 2-NAND and inverters, and path equalization techniques, the overall delay variation of a CMOS circuit is relatively small, between 10-20%, despite substantial delay variations of individual gates of about 60%.

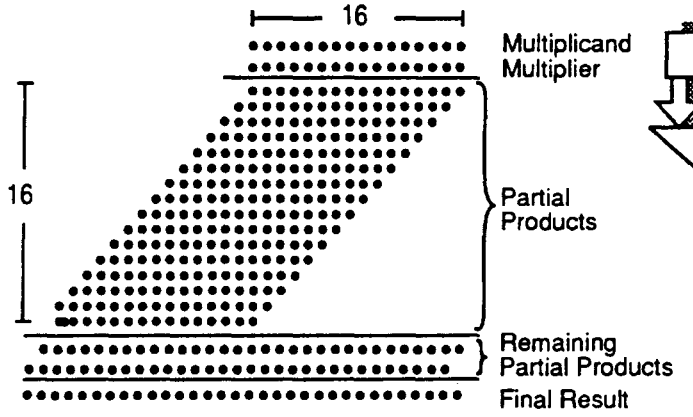
Example



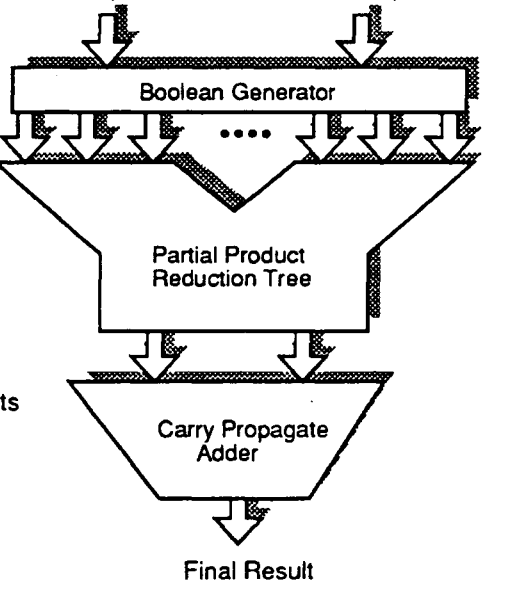
Delay variations cancel out, as if all gates have average delay

16x16-bit Multiplier

Algorithm

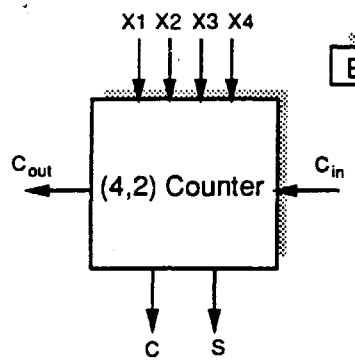


Architecture



(4,2) Counter

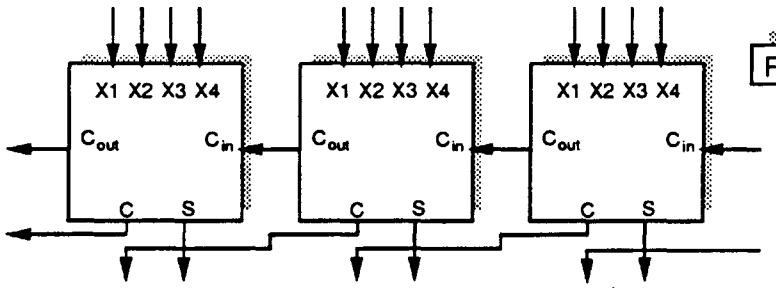
It is the building block of the partial-product tree:



Block Diagram

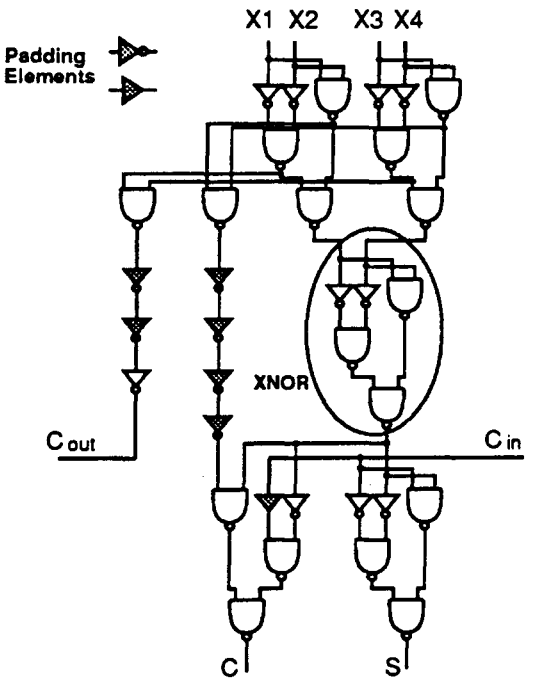
C_{in}
X1
X2
X3
+ X4
C S
C_{out}

Logic Function



Row of (4,2) counters

(4,2) Counter Implementation



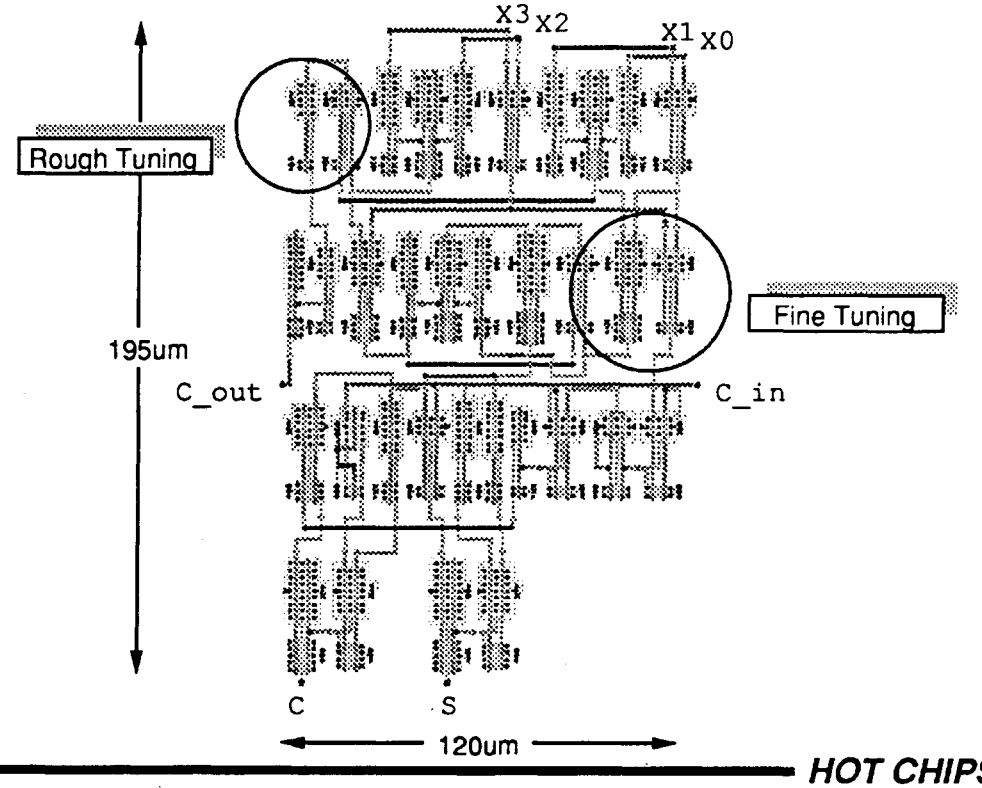
Minimum Delay Test Vectors				Maximum Delay Test Vectors			
0t	0t	0s	1t	0t	1s	1s	0t
0s	1t	0t	0t	1s	0t	0t	1s
0t	0t	1t	1s	1t	1t	0t	0s
1t	0s	0t	0t	1t	1t	0s	0t
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•

	C	S
Maximum Delay	1.46ns	1.46ns
Minimum Delay	1.19ns	1.28ns
Average Delay	1.33ns	1.38ns
Max-Min Delay	0.28ns	0.18ns

0s (1s) : steady 0 (1)

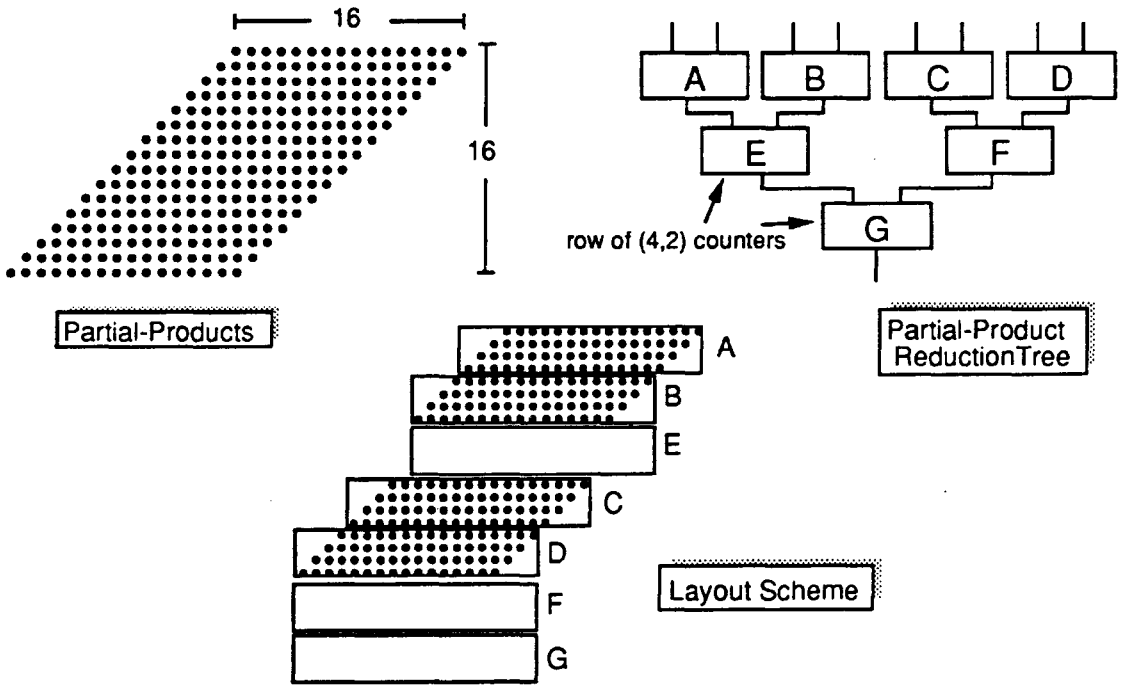
0t (1t) : transition 1-0 (0-1)

(4,2) Counter Layout



HOT CHIPS V

Partial-Product Reduction Tree

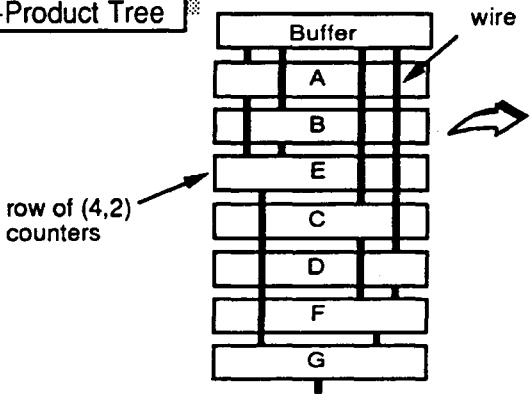


HOT CHIPS V

Wiring Delays

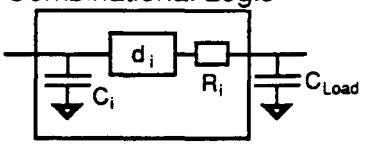
The output capacitance of a cell in the tree is dominated by the capacitance of the interconnection wires

Partial-Product Tree



Delay Model

Combinational Logic



$C_{Load} = C_{wire} + C_{logic}$

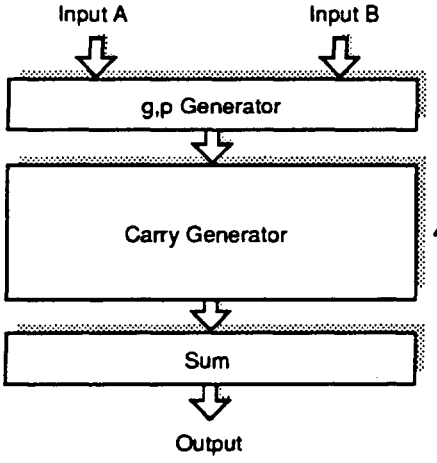
Delay at stage i:

$$D_i = \sum d_i + \sum R_i \, (C_{wire}(i,i+1) + C_{logic}(i+1))$$

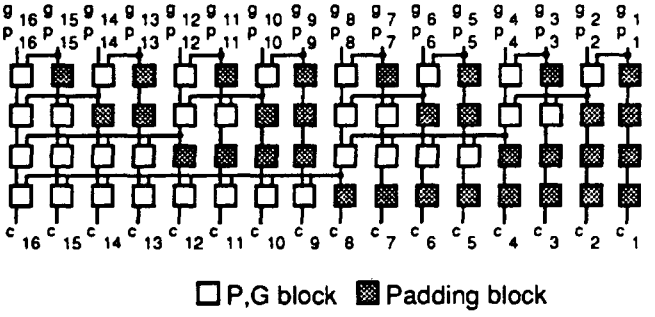
The overall delay of the tree is balanced since all stages are equal and the sum of the lengths of the interconnection wires is the same along any path.

16-bit Carry Propagate Adder

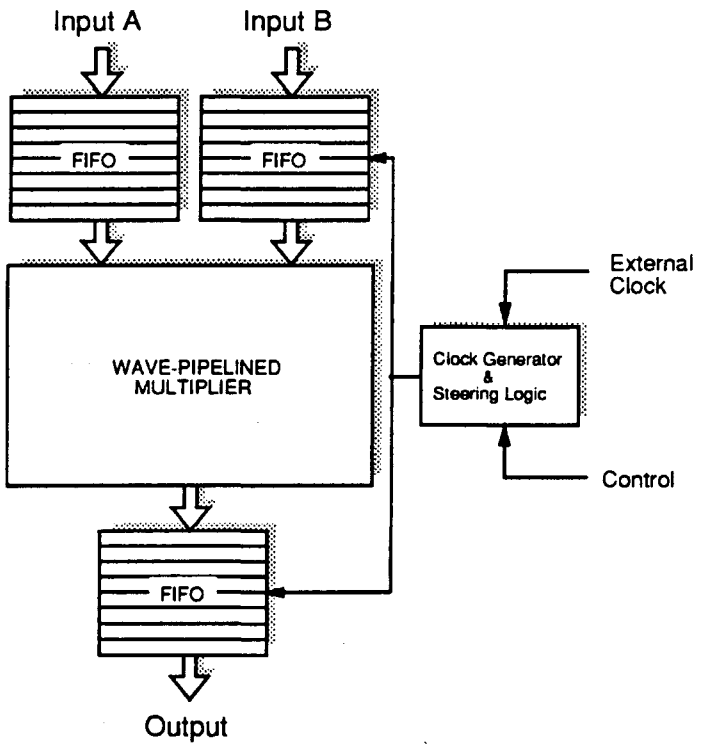
Block Diagram



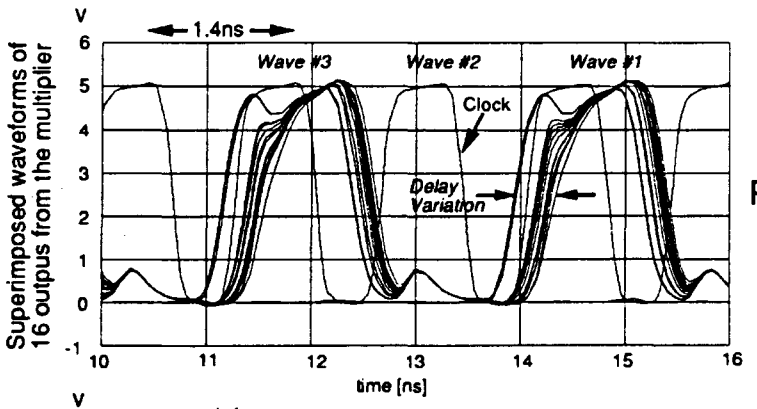
Parallel Carry Generator



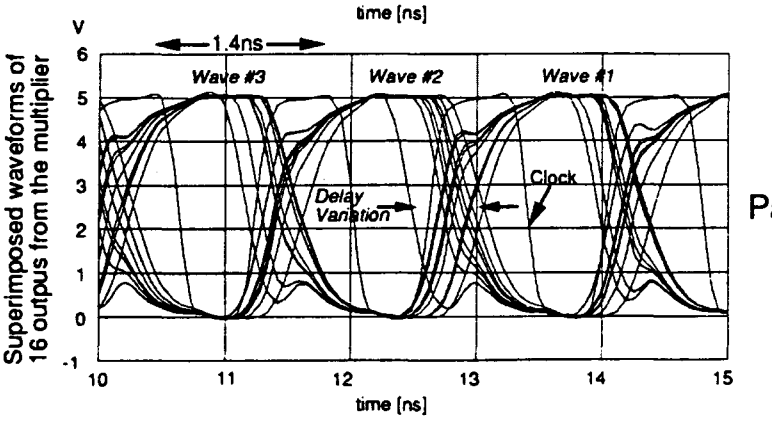
Test Chip



Simulation Results

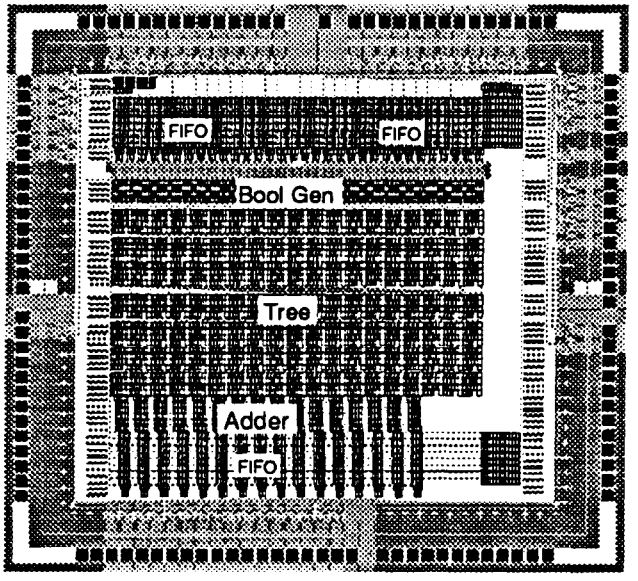


Pattern #1



Pattern #2

Chip Layout



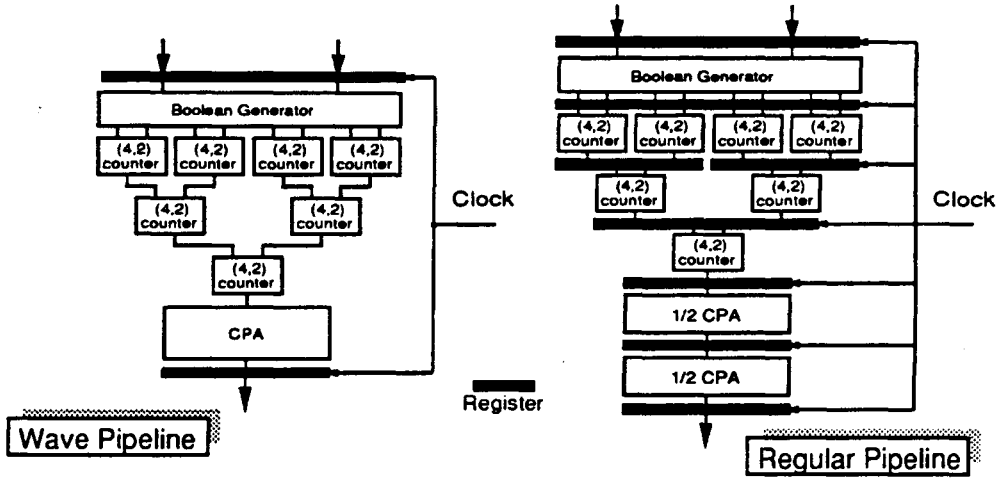
4.6mm

4.96mm

Poly width	1.0um
Contact 1	0.8um
Metal 1	1.0um
Metal 2	1.0um
Metal 3	3.6um

Technology Parameters

Wave Pipeline versus Regular Pipeline



Pipeline	Latency	Min Clock	Speed-up	Area	Latency x Clock
No	10.1ns	10.1ns	1.0x	A	$102ns^2$
Wave	10.2ns	1.45ns	7.0x	1.21A	$14.8ns^2$
Regular	16.2ns	2.7ns	3.7x	1.33A	$43.7ns^2$

Conclusions

Significant performance increase can be achieved in CMOS by using the technique of wave pipelining

Wave pipelining can achieve a product Latency x Cycle-Time about 2x better than Regular Pipelining