The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2177-5

ICTP Latin-American Basic Course on FPGA Design for Scientific Instrumentation

15 - 31 March 2010

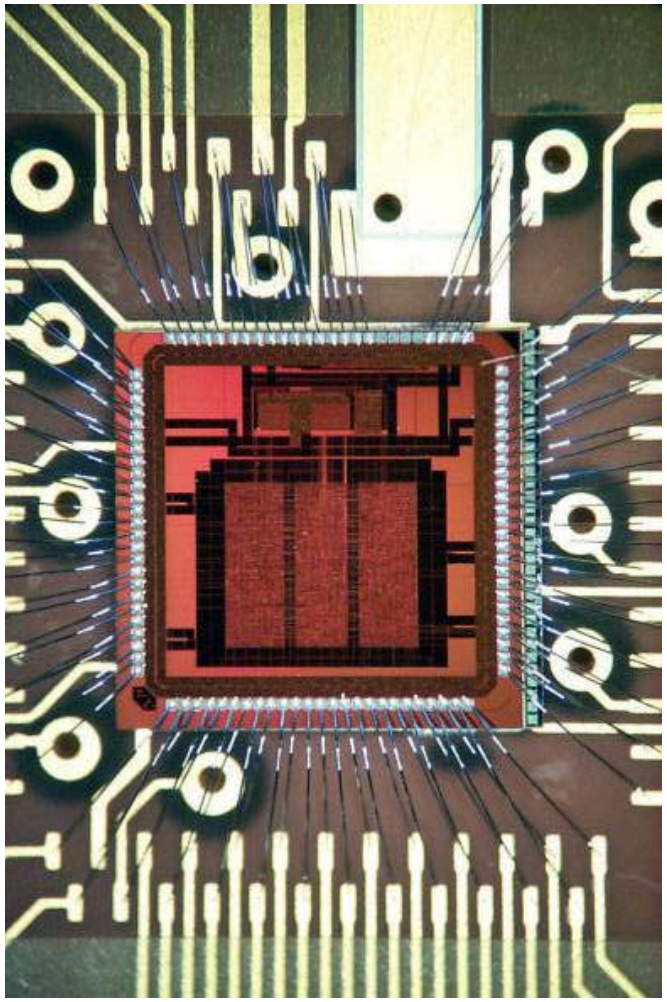
Scaling

MOREIRA Paulo Rodrigues S.

CERN
Geneva
Switzerland

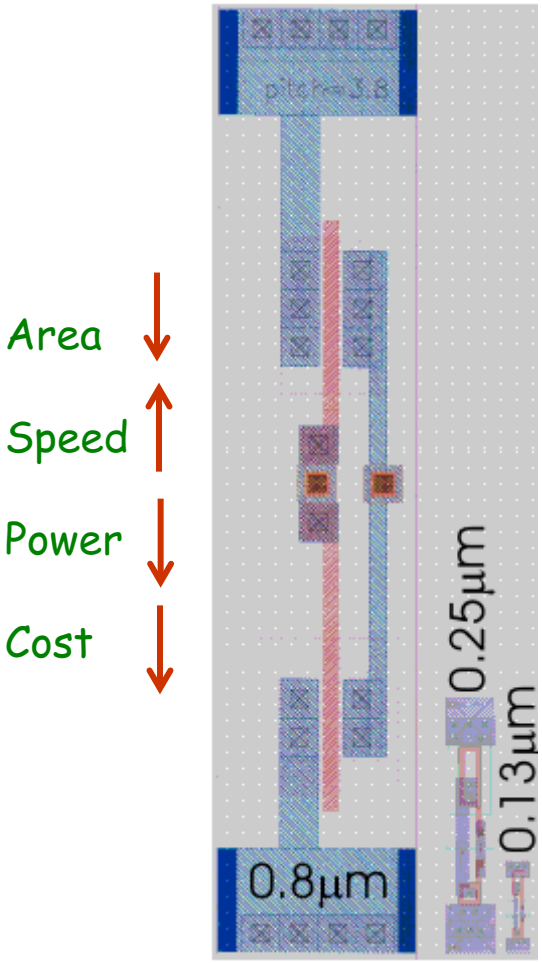
Outline

- Introduction
- Transistors
- The CMOS inverter
- Technology
- Scaling
  - Scaling objectives
  - Scaling variables
  - Scaling consequences
- Gates
- Sequential circuits
- Storage elements
- Phase-Locked Loops
- Example



Technology scaling

- Technology scaling has a threefold objective:
  - Increase the transistor density
  - Reduce the gate delay
  - Reduce the power consumption
- At present, between two technology generations, the objectives are:
  - Doubling of the transistor density;
  - Reduction of the gate delay by 30% (43% increase in frequency);
  - Reduction of the power by 50% (at 43% increase in frequency);



Technology scaling

- How is scaling achieved?
  - All the device dimensions (lateral and vertical) are reduced by $1/\alpha$
  - Concentration densities are increased by α
  - Device voltages reduced by $1/\alpha$ (not in all scaling methods)
  - Typically $1/\alpha = 0.7$ (30% reduction in the dimensions)

Technology scaling

- The **scaling variables** are:
  - Supply voltage: $V_{dd} \rightarrow V_{dd} / \alpha$
  - Gate length: $L \rightarrow L / \alpha$
  - Gate width: $W \rightarrow W / \alpha$
  - Gate-oxide thickness: $t_{ox} \rightarrow t_{ox} / \alpha$
  - Junction depth: $X_j \rightarrow X_j / \alpha$
  - Substrate doping: $N_A \rightarrow N_A \times \alpha$

  This is called **constant field** scaling because the electric field across the gate-oxide does not change when the technology is scaled

  If the power supply voltage is maintained constant the scaling is called **constant voltage**. In this case, the electric field across the gate-oxide increases as the technology is scaled down.

Due to gate-oxide breakdown, below 0.8μm only "constant field" scaling is used.

Scaling consequences

Some consequences of 30% scaling in the constant field regime (α = 1.43, $1/\alpha$ = 0.7):

- Device/die area:

$$W \times L \rightarrow (1/\alpha)^2 = 0.49$$

  - In practice, microprocessor die size grows about 25% per technology generation! This is a result of added functionality.

- Transistor density:

$$(\text{unit area}) / (W \times L) \rightarrow \alpha^2 = 2.04$$

  - In practice, memory density has been scaling as expected.

Scaling consequences

- Gate capacitance:

$$W \times L / t_{ox} \rightarrow 1/\alpha = 0.7$$

- Drain current:

$$(W/L) \times (V^2/t_{ox}) \rightarrow 1/\alpha = 0.7$$

- Gate delay:

$$(C \times V) / I \rightarrow 1/\alpha = 0.7$$
$$\text{Frequency} \rightarrow \alpha = 1.43$$

  - In practice, microprocessor frequency has doubled every technology generation (2 to 3 years)! This faster increase rate is due to super-pipelined architectures ("less gates per clock cycle")

Scaling consequences

- Power:

$$C \times V^2 \times f \rightarrow (1/\alpha)^2 = 0.49$$

- Power density:

$$1/t_{ox} \times V^2 \times f \rightarrow 1$$

  - In practice, the power density has been increasing faster than foreseen by the simple scaling theory. This is due to the faster them foreseen increase in frequency

Interconnects scaling

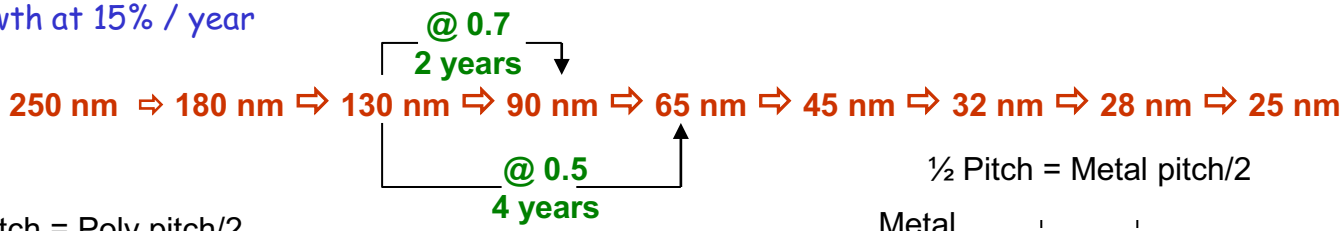
- Interconnects scaling:
  - Higher densities are only possible if the interconnects also scale.
  - Reduced width → increased resistance
  - Denser interconnects → higher capacitance
  - To account for increased parasitics and integration complexity **more interconnection layers** are added:
    - thinner and tighter layers → local interconnections
    - thicker and sparser layers → global interconnections and power

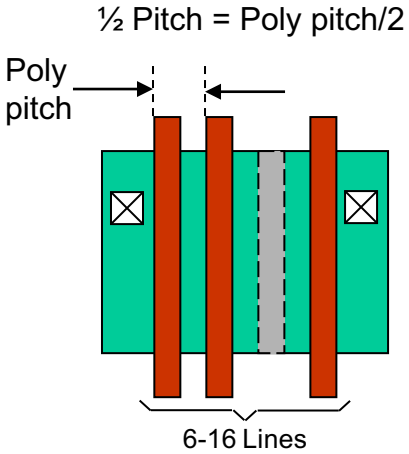
Scaling table

Parameter	Constant Field	Constant Voltage	
Supply voltage (V_{dd})	$1/\alpha$	1	Scaling Variables
Length (L)	$1/\alpha$	$1/\alpha$	
Width (W)	$1/\alpha$	$1/\alpha$	
Gate-oxide thickness (t_{ox})	$1/\alpha$	$1/\alpha$	
Junction depth (X_j)	$1/\alpha$	$1/\alpha$	
Substrate doping (N_A)	α	α	
Electric field across gate oxide (E)	1	α	Device Repercussion
Depletion layer thickness	$1/\alpha$	$1/\alpha$	
Gate area (Die area)	$1/\alpha^2$	$1/\alpha^2$	
Gate capacitance (load) (C)	$1/\alpha$	$1/\alpha$	
Drain-current (I_{dss})	$1/\alpha$	α	
Transconductance (g_m)	1	α	
Gate delay	$1/\alpha$	$1/\alpha^2$	Circuit Repercussion
Current density	α	α^3	
DC & Dynamic power dissipation	$1/\alpha^2$	α	
Power density	1	α^3	
Power-Delay product	$1/\alpha^3$	$1/\alpha$	

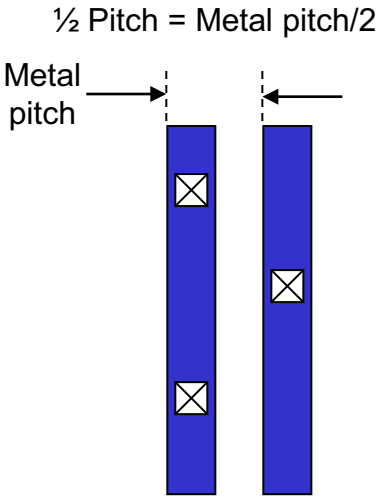
2008 and beyond …

- International Technology Roadmap For Semiconductors (ITRS – 2007)
- Forecast from the semiconductor industry with a 15 year perspective:
  - Near-term: 2007 – 2015
  - Long-term: 2016 – 2022.
- The forecast is done in terms of 1st year of production:
  - Product shipment first exceeds 10K units/month (using production tooling)
- A near-term scaling ratio of @ 0.7 is assumed
- These scaling trends will allow the electronics market to growth at 15% / year





Typical flash un-contacted poly



Typical DRAM/MPU/ASIC
Stagger-contact Metal Bit Line

2008 and beyond …

ITRS Road Map, 2007 edition:	2008	2010	2012	2022	(year of first production)
DRAM ½ pitch (nm)	57	45	36	11	
μP M1 ½ pitch (nm)	59	45	36	11	
Flash Poly ½ pitch (nm)	45	36	28	9	
Gate length, printed (nm)	38	30	24	7.5	
Gate length, physical (nm)	23	18	14	4.5	
DRAM:					
Bits/chip (Gbits)	2.15	4.2	4.29	68.72	
Chip size (mm^2)	74	93	59	93	
Gbits/cm^2	2.9	4.62	7.33	73.85	
Flash:					
Bits/chip (Gbits) SLC	8.59	17.18	17.18	274.88	
Chip size (mm^2)	101.80	128.26	128.26	128.26	
Gbits/cm^2	8.44	13.40	21.30	214.0	
Bits/chip (Gbits) MLC [2 bits/cell]	17.18	34.36	34.36	549.76	
Chip size (mm^2)	101.80	128.26	80.80	128.26	
Gbits/cm^2	16.90	26.80	42.50	429.0	
Bits/chip (Gbits) MLC [4 bits/cell]	34.36	68.72	68.72	1099.51	
Chip size (mm^2)	101.80	128.26	80.80	128.26	
Gbits/cm^2	33.8	53.6	85.1	857.0	
μP (high performance):					
Transistors/chip (Millions)	1106	2212	2212	35391	
Chip size (mm^2)	246	310	195	310	
Transistors/cm^2 (M/cm^2)	449	714	1133	11416	
Total pads	3072	3072	3072	3072	(66.7% allocated for power and ground)

IEEE Spectrum, July 1999

Special report: "The 100-million transistor IC"

2008 and beyond …

ITRS Road Map, 2007 edition:	2008	2010	2012	2022 (year of first production)
Performance:				
On-Chip clock (GHz)	5.0	5.9	6.8	14
Chip-to-board (GHz)				
Wiring levels (maximum)	12	12	12	15
Wiring levels (minimum)				
Power supply:				
Vdd (V):	1.0	1.0	0.9	0.65
Maximum allowable power (W)	198	198	198	198 (With heat sink)
Lithography:				
Field size – area (mm^2)	858	858	858	858
Wafer diameter (mm)	300	300	300/450	450