

L1. Intro & Motivation

Lecture : 95 slides (22+73)

1.1 Options for implementing a digital system

	Est.	today	sign.	future
1 Non-prog. std. parts	early 1960	low	none	
2 Micro/digital signal processor	early 1970	high	Steady	
3 FPGA	late 1980	fair		
4 Semi-Custom ASIC	early 1980	fair		
5 Full-Custom cell based ASIC	early 1980	fair		
5s Full-Custom hand layout ASIC	idem	low	declining	
6 Platform IC	early 2010	moderate		(Rising)

- Many alternative choices
- Platform IC: e.g. Zynq combination of
 - instr. set programmable
 - electrically configurable
 - hardwired circuit blocks

1.2 Models of industrial cooperationStandard parts

- + no HDL involved
- + mix COTS parts
- poor density
- hard-wired circuits

Program-controlled Processors

- + no HDL involved
- + unlimited agility
- dep. on third parties for more profound sols.

FPGA

- + no masks and wafer cycles
- + no Sachverständige tools
- more demanding than SW
- poor usage of silicon with large volume

Semi-cust. ASIC

- + circuit off.
- + large vol. prod.
- many contracts & agreements
- little agility
- + more optimized
- + even larger eco in scale
- very demanding development
- long turnaround

cell, full-cust. ASIC

- + supply chain
- wafers
- testing
- packaging

1.3 Costs of integrated circuitsYou-recurring costs

- circuit design
- sign-off
- + prop. right
- + prototypes
- + qualification

Recurring cost

- supply chain
- wafers
- testing
- packaging

$$C = \frac{C_0}{\#cls} + C_1$$

c : cost per unit
 $Hcls$: N working chks produced
 C_0 : you-recurring cost
 C_1 : recurring cost

$$C_{prod} = \frac{C_{wrk} + C_{wp}}{\#prod}$$

C_{prod} : cost of 1 functioning die
 C_{wrk} : raw wafer cost
 C_{wp} : wafer processing
 H_{prod} : defect-free per wafer

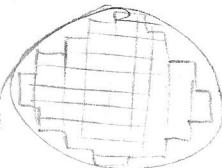
$$C_i = C_{prod} + C_{test} + C_{pack}$$

$$H_{wafer} \approx \frac{\pi}{4} \left(\frac{dw}{2} - \sqrt{A_d} \right)^2$$

H_{wafer} : N dies per wafer

dw : wafer diameter

A_d : die area



! As die area gets bigger, prob. of suffering from defect increases!

Fabrication yield: $y_f = \frac{\# good}{\# many} \approx (1 + \frac{D \cdot t_c}{A_d})^{-t}$

$y_f \leq 1$
negative binomial model

D: defect density $\approx 0.004 \mu m^{-2}$ 90nm
improves with process maturity
t: cluster factor (\approx to # lithographic steps)
 ≈ 4.0 for 90nm

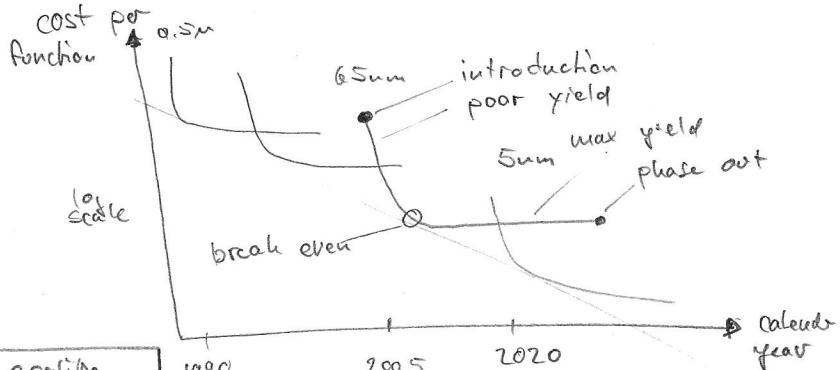
small chks \rightarrow cost upprop. to die size
complex chks \rightarrow expenses increase highly propotional

A_d: die area of actual layout structures

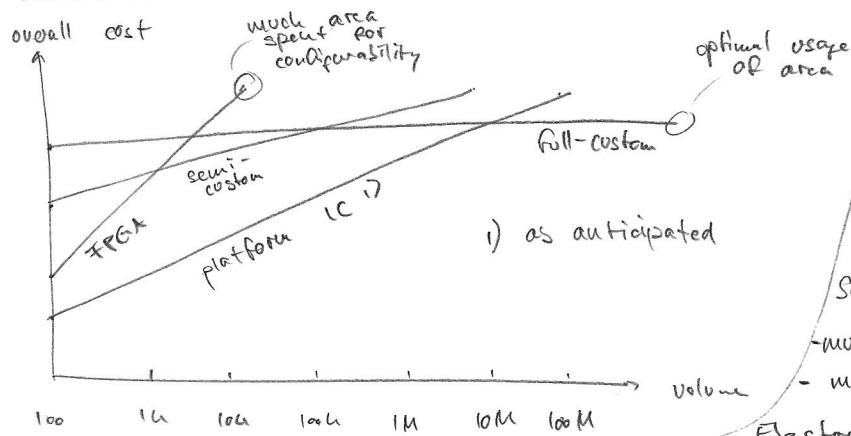
yield is limited by:
 a) fabrication defects
 b) unpredictable parameter variations (since 90nm gen.)
 - MOSFET V_{th} - leakage currents - metal thickness
 too high \rightarrow limit speed too low \rightarrow high leakage currents

The VLSI learning curve

Industry migrates to next process when the savings from fabricating in denser process compensate for more expensive masks and wafer processing.



Pace of cost reduction and technology scaling likely to slow down unless a breakthrough in photolithography eliminates multipatterning



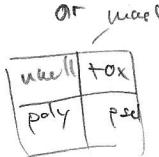
1.4 Fabrication avenues for small quantities

Problem: initial costs too high, boundaries require ≈ 1000 wafers per year min.
FPL/CPW fill gap but have - no support for analog or circuit - limited in package - no edge competition



Solutions: Mask sharing

- multiple different chips on 1 wafer
- multiple layers per single mask



Electron Beam Lithography

+ No masks E-Scan draws layout pattern
- limited throughput

Hardwired FPGA/structured ASIC
Semi-custom replacement for FPGA

+ No masks E-Scan draws layout pattern
- limited throughput

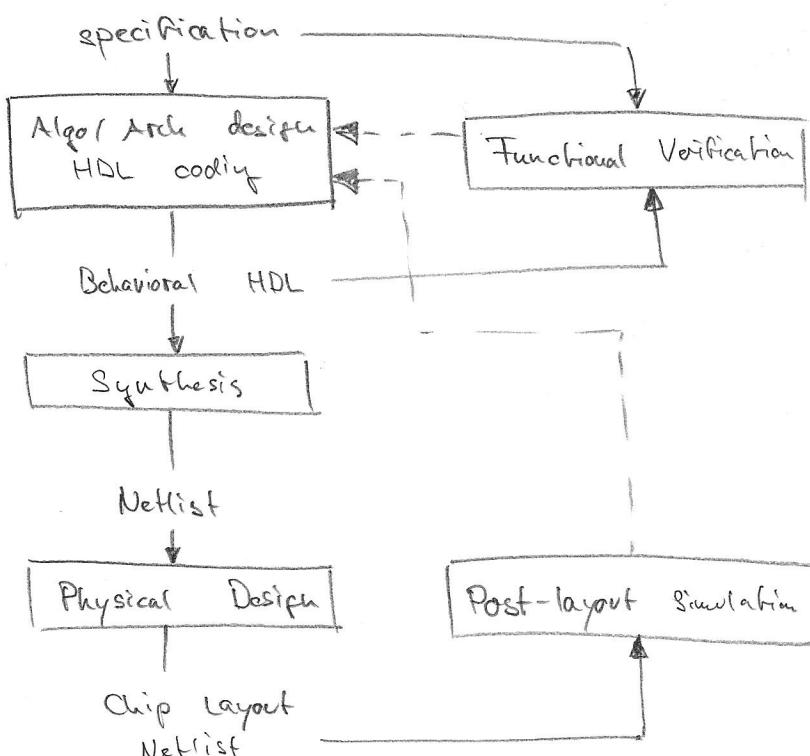
1.5 Making a choice

ASIC { pro: ↓ assembly cost ↑ reliability ↓ space requirements ↑ performance ↑ energy efficiency
con: ↓ flexibility ↓ long turnaround • large volume production • highly specialized designers

Reason 1: enable advanced products that are impossible to make otherwise (space, speed, energy, integrated sensors, reliability)

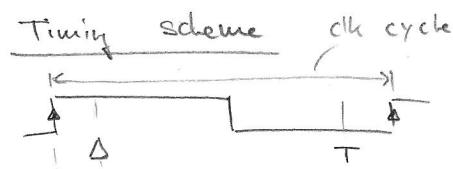
Reason 2: lower the recurring costs in comparison to alternative solutions

E0. Simulation & Synthesis flow for ASIC designer



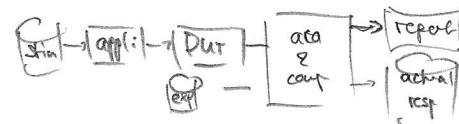
Functional Verification with Questa Sim

- design is compiled & optimized to simulate faster



Δ: stimuli application
T: response acquisition

Testbench structure



hard coded
TB

file-based
TB

RTL Synthesis

takes RTL description and generates functionally equivalent gate-level netlist. This netlist contains only instantiations of standard-cells (AND, FF, ...) and connections. These cells are provided by foundry as standard cell library.

Synthesis steps

Analyze

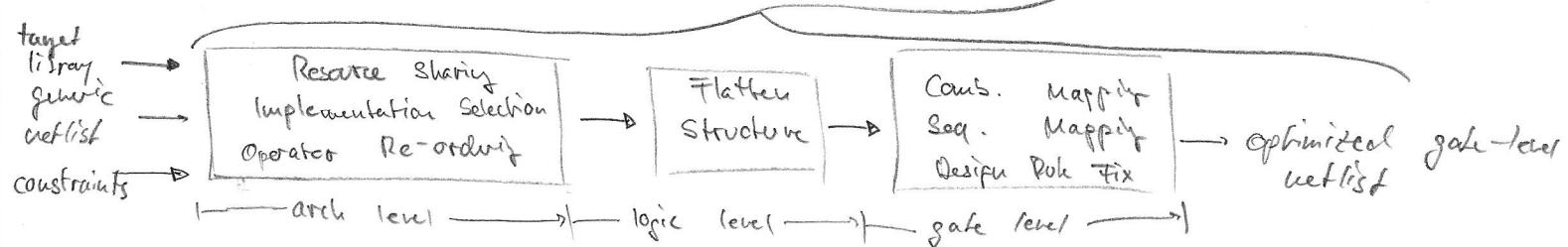
converts HDL to intermediate format

Elaborate

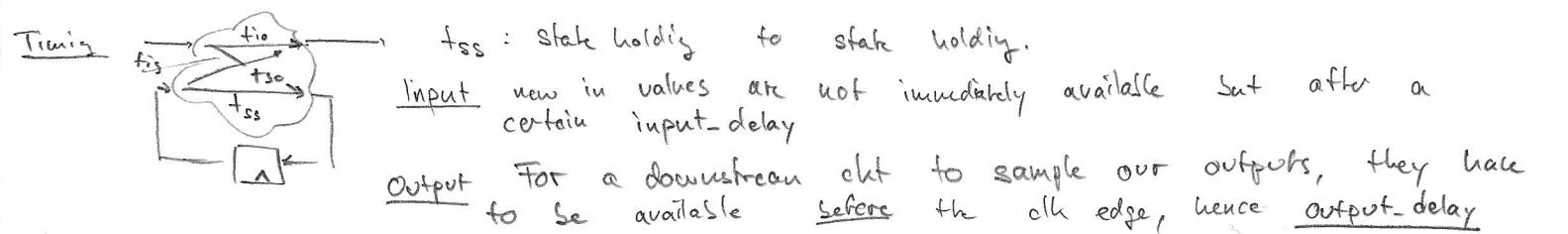
- set parameters to final value
- instantiate modules

Compile

- mapping to technology



Area \wedge GE (gate equivalent) = single two-input ND2 gate (TSMC65 : 1GE = 1.82 μ m²)



Input new values are not immediately available but after a certain input-delay

Output For a downstream clk to sample our outputs, they have to be available before the clk edge, hence output-delay

Drivers have limited strength and rise/fall time. Specify them to get correct timing reports.

Loads outputs have to drive capacitive loads. Specify with set-load

L2. Packaging and Interfaces Lecture: 46 slides

2.1 Why packaging? - Protect against ^{+ env. hazards} mechanical stress - provide elec. connection
- carry away thermal power - part handling

2.2 Parameters - Number of pins - Physical size - convection to solder
- thermal resistance / capacitance

- Voltage levels: same voltage for 0/1. Most are standard CMOS. From 5V to 3.3, 2.5, 1.8V
Lower voltage \rightarrow faster but more noise High voltage: slower but less noise
- Signal to chip:
 - I limit in speed: speed of light \rightarrow 10mm : 0.03μs \approx 30 GHz
 - II parasitics (R, L, C) slow things down
 - III switching signals requires energy
- Supply voltage delivery:
 - Limit in max current through wire - current peaks on clk toggle
 - Modern chips need a lot of current
- Outside chip: Large loads (pF), tracks, long distance, more exposed \rightarrow noise
- Inside chip: Small loads, thin oxides, short distances, controlled environment
- I/O driver: These transistors are LARGE, multiple stages from pin to core

2.3 ESD electrostatic discharge 0.1..10kV generate peak currents of several A.

(Precautions: controlled humidity, ESD workbenches, ...)

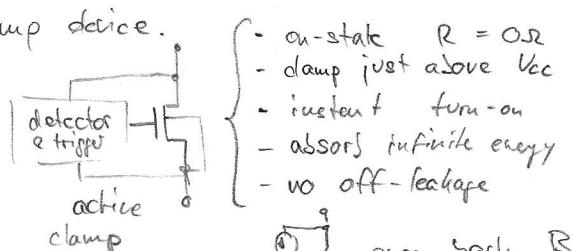
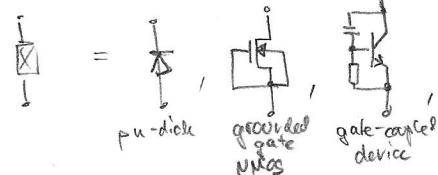
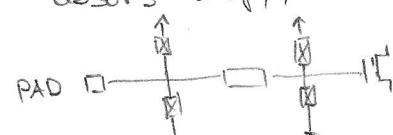
On-chip: 2.3.1 on-chip ESD protection

Traditional: absorption on every pad \rightarrow costly

Key idea: An ESD protection network must provide a low impedance path for ESD currents to discharge.

Pervasive: pn-diodes on every pad divert currents to supply rails. Shared power clamps

absorbs energy, active ESD detection triggers clamp device.

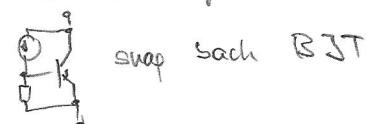


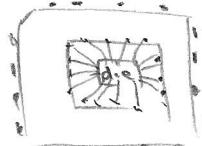
today used: input primary
input secondary
power rail clamp
output

pn-diode snap-BJT
active pn-diodes

}

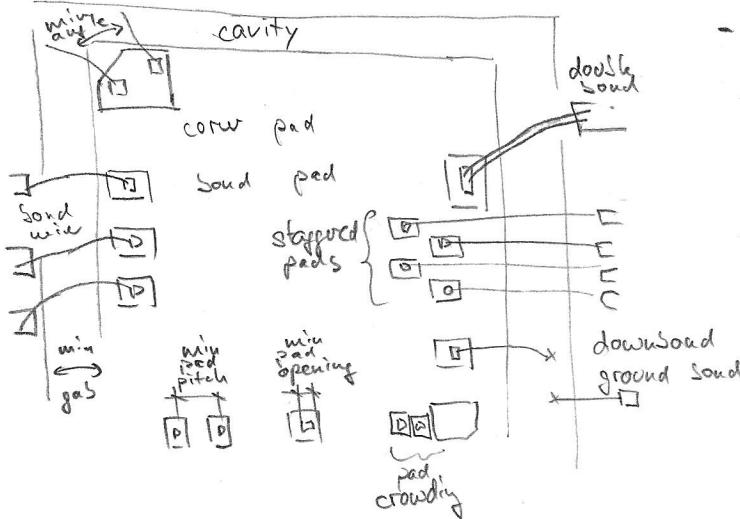
} move A efficient
control task





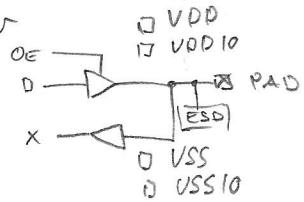
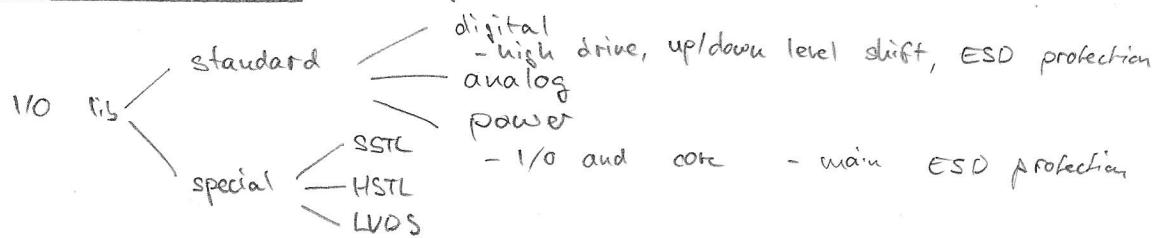
2.3 Packaging manufactured dies

1. Blue die into cavity
 2. Use bonding wires to connect
 3. seal package
- double/triple bonds reduce inductance

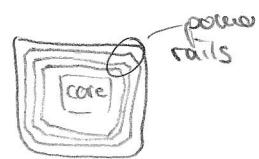


- thin wires are fragile
 - have higher resistance
 - allow smaller pads
 -
- power: ~ 14 µm diameter

2.4 I/O Cells Complete macro cells for I/O, and power



2.5 Power



~ 30% of all pins are power

- Simultaneous Switching Outputs SSO is a metric describing the period of time during which the switching starts/stops.

$$V_{drop} = L_{pH} \cdot \frac{di}{dt} \rightarrow \text{more VDD pins reduce noise}$$

- Do's:
- Many caps
 - As much power pins as possible
 - Separate supply for IO/core
 - Short wires
 - Low signal rate

A 48W chip @ 1.2V → 40A supply! ~10mΩ per wire → 4'000 power bond wires required
 SSO: Simultaneous Switching Outputs: Period SSW: Switching starts/ends → describes how many power pins required

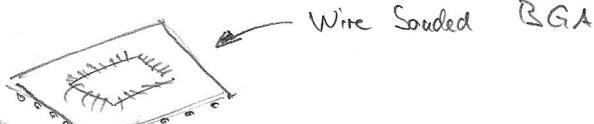
2.6 BGA : Ball Grid Array

PCB layout is difficult. If outer rows are I/O and inner power, flip ou.



2.7 Advanced Shift

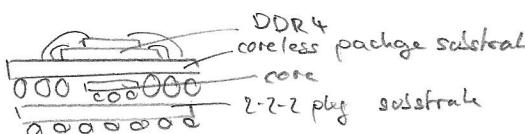
2.50 integration:
dies glued together



2.8 Misc - Slow signals → noise → Schmitt trigger

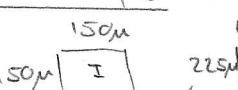
- LVDS for less noise more speed
- Need to make arch. opt. to reduce pin out

Interposers

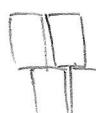


E1. Interfaces & Chip I/O

E1.1 Pad shapes Example: $I/Os = N_{IO} = 100$ power $N_{PC} = 6$ $N_{Pi} = 6$

Pads: 

Chip 1: $A_c = 0.5 \text{ mm}^2$ Pads II $\rightarrow A_{chip} = 10.56 \text{ mm}^2$
 chip 2: $A_c = 30 \text{ mm}^2$ pads I $\rightarrow A_{chip} = 35 \text{ mm}^2$

$\begin{cases} A_{chip} \gg A_{core} & \text{pad-limited} \\ \text{else} & \text{core-limited} \end{cases}$ staggered:  \rightarrow twice as many pads

E1.2 Off-chip interfaces I2C, I2S, SPI, STAG, UART, CAN, GPIO, MIPI-CSI, USB, PCIe

E1.3 On-chip interfaces

APB: Advanced Peripheral Bus

- simple
- unpipelined
- transfers take ≥ 2 cyc.

AHB: Advanced High-Performance Bus

- more complex
- doesn't scale well
- e.g. microcontrollers
- multi master, shared, 1 active master at time
- data rate limited by wiring

AXI: Advanced Extensible Interface, High-perf, High-speed

- separate addr/ctl/data phase
- out of order
- qualified support
- point to point
- can be pipelined
- bursts
- interleaved

AXI-Lite: 32/64/128, no bursts, no out-of-order transactions

AXI-Stream: No address, point to point

E1.4 Package types



E1.5 Reset Timing asynch reset deassertion can lead to different register states

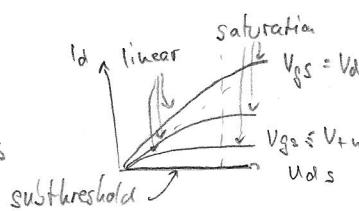
- synchronize reset to clk - balanced reset tree

L3. Cell based design & memories 97 slides3.0 Transistor

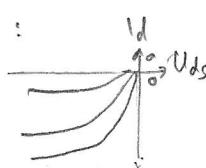
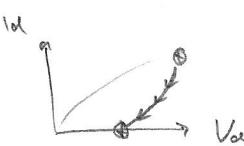
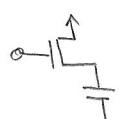
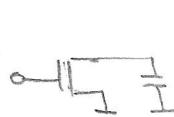
MOS : voltage-controlled current source. Simple to mfc, low leakage.
input is capacitive.

$$10 \frac{A}{W} \frac{W}{L}$$

n-channel: D
 $G \rightarrow H_s$

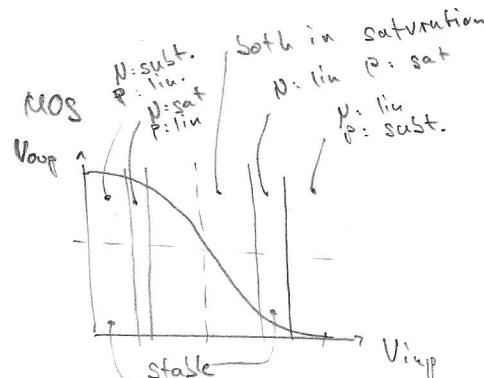
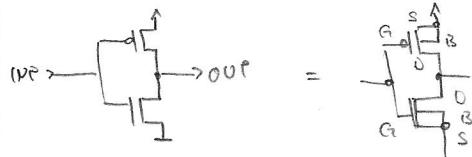
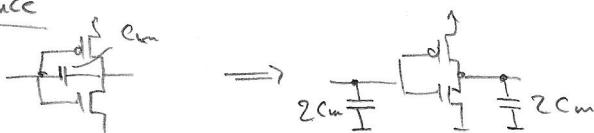


p-channel : D
 $G \rightarrow L_s$

MOS as switch

! Don't use NMOS as high side switch. V_{gs} diminishes as charging progresses !

	turn on	acts as	on-state conductance
NMOS $G \rightarrow H_s$	$V_{gs} - V_{th} > 0$	poor pull-up good pull-down	$G_{on} \propto W/L$
PMOS $G \rightarrow L_s$	$V_{gs} - V_{th} < 0$	good pull-up poor pull-down	$G_{on} \propto \frac{3}{2} \cdot W/L$ ← worse conductor

3.1 Combinational cells3.1.1 CMOS Invertercomplementary MOSMiller Capacitance

more undesirable than other parasitics

Propagation delay

$$t_{pd} \propto \frac{C_h V_{dd}}{(V_{dd} - V_{th})^{\alpha}}$$

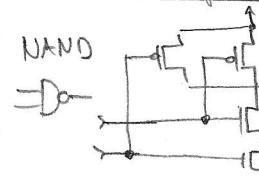
$C_h \approx C + 2C_m$ load cap.

V_{dd} supply

V_{th} threshold voltage

+ velocity sat. index $\alpha \approx 1.5 \pm 0.2$, $\alpha \approx 1$

+ $\alpha \approx 1$ for $V_{dd} \approx V_{th}$

3.1.1 Simple CMOS gates

dual:
parallel \leftrightarrow serial

- $1n + 1p$ for every argument in Boolean eqn.
- output never left floating
- rail-to-rail output

- no input can make path $V_{dd} - V_{ss}$

- no PDs in steady state, except leakage

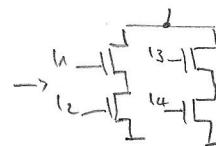
autagonistic: a set of n and p net

A subc is ratioless if the sizes and drive str. of its transistors do not affect its logical func.

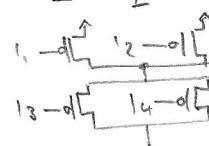
3.1.2 Composite or Complex gates

$$\text{OUP} = \overline{(I_1 \wedge I_2) \vee (I_3 \wedge I_4)}$$

1. construct u-network



2. add dual p-network



Guideline: no more than
3 MOSTET in series

Structural duality between P/N is

- a sufficient req. for fully complementary
- not always necessary to obtain electrical

static CMOS gate
autocomplementation

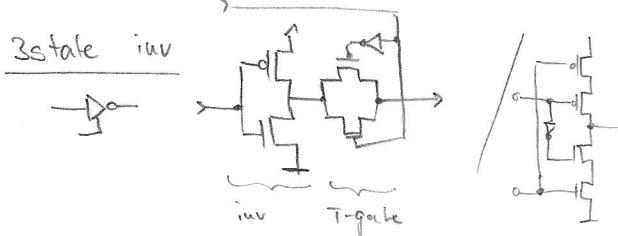
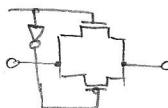
3.1.3 Gates with high Z capability

- no amplification / level restoration

- no GND / supply

- no in/out → bidirectional

Transmission gate or T-gate:



3.1.4 Parity Gates

$\Rightarrow D$
 $\oplus D$
 $\otimes D$

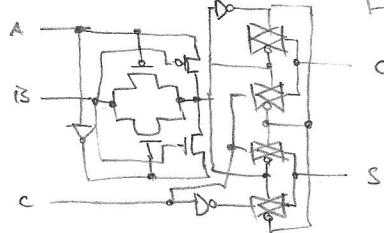
16 trans

10 trans

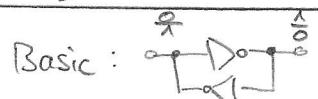
6 trans

3.1.5 Full Adder

T-gate adder is a low-power alternative,
no level restoration. Best when combined with
level-restoring adders



3.2 Bistable Cells (Lecture 4 starts here)



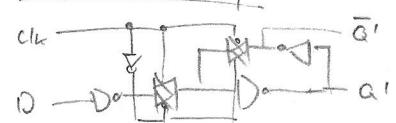
- two stable states with positive feedback
- combinational loop

Problems:

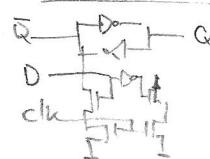
- How to write?
- Reading could influence cell content (especially with low power)

3.2.1 Latches

Switched Memory



Jain's latch

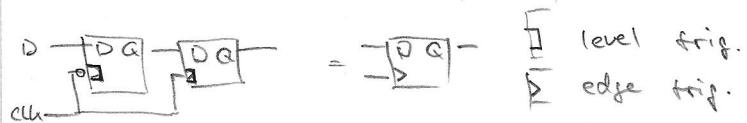


selectively pull-down
with stronger NOS

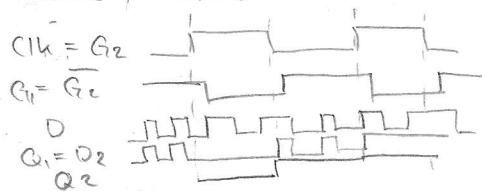
Important: sometimes output suffers
from required if T-gates are used
else you get memory loss
from backword signal propagation

3.2.3 Single-edge-triggered FF

Obtained from two latches in series with common clocks: master-slave



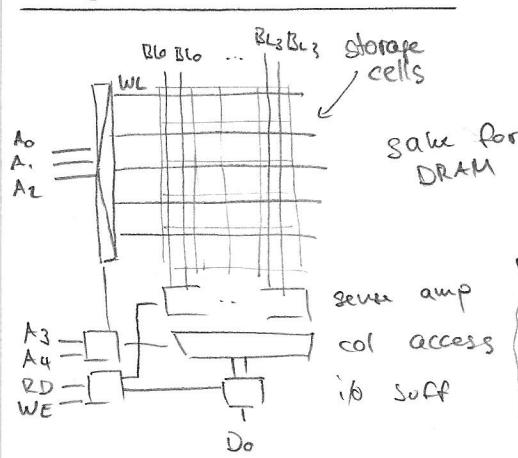
level trig.
edge trig.

3.3 On-chip memories

Types: Static (SRAM), Dynamic (DRAM)
latch and master-slave flip-flop

Static: state preserved \Leftrightarrow if voltage supplied

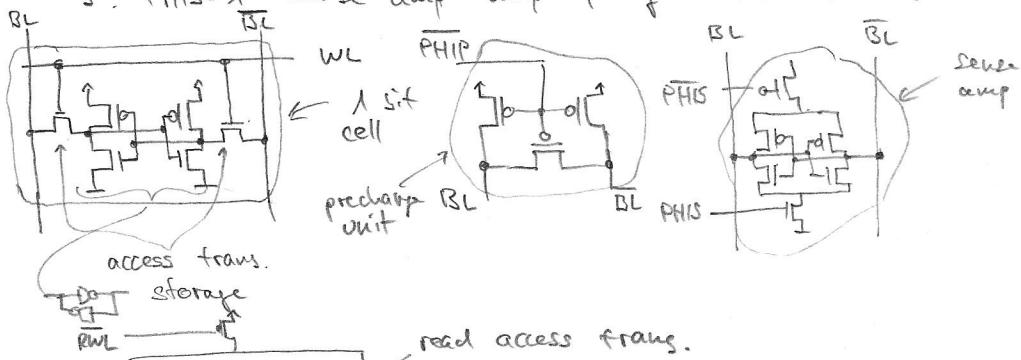
Dynamic: charge on a capacitance represents state \rightarrow periodic refresh required

3.3.1 Static RAM

BL: bit line
WL: word line

Read problem: long bit lines = large C
 \rightarrow cells would need high drive, sat need to be small. \rightarrow precharge and sense amp help.

1. PHIP=1 \rightarrow BL are precharged to $\frac{Vdd}{2}$
2. PHIP=0, activate BL, cells start polling to data state
3. PHIS=1 sense amp help polling BL/BL to this level



More robust cell: Q-trans. SRAM

- + readout separate MOSFET, relaxed sizing & improved stability
- + Vdd > 0.3V
- \sim PMOS access trans.
- larger area!

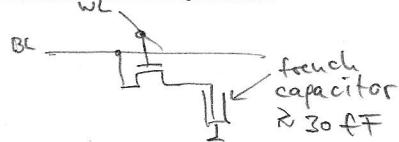
- Readout is destructive

- bit line cap is $\sim 5\text{-}15x$ larger than storage cap

$$\Delta U_{BL} = \pm \frac{Vdd}{2} \frac{C_{cell}}{C_{BL} + C_{cell}} \approx \pm \frac{Vdd}{2} \frac{1}{(5..15)+1} \approx \pm 125\text{-}47 \text{ mV} @ 1.5V_{dd}$$

- Sense amp automatically performs writesback on read

- V_{th} is high for lower leakage
- refresh every $\sim 64\text{-}256 \mu\text{s}$

3.3.2 Dynamic RAM

- Readout is destructive

- bit line cap is $\sim 5\text{-}15x$ larger than storage cap

$$\Delta U_{BL} = \pm \frac{Vdd}{2} \frac{C_{cell}}{C_{BL} + C_{cell}} \approx \pm \frac{Vdd}{2} \frac{1}{(5..15)+1} \approx \pm 125\text{-}47 \text{ mV} @ 1.5V_{dd}$$

- Sense amp automatically performs writesback on read

3.4 Electrical contraptions

3.4.1 Snapper



Prevents voltage of Estate node from drifting away while not being driven.

3.4.2 Schmitt-trigger



3.4.3 Tie-off

Pomacentry tie a cell's
input to O/I

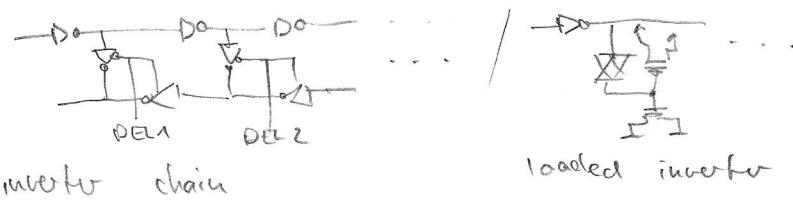
3.4.4 Fillcap

MOSCOW

80nm: 1.2 fF 100Ω

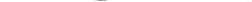
- can also be used as spare cell to change design w/o modifying Si mask, only metal

3.4.6 Dig. adj. delay line



3.5 Pitfalls

3state. - make sure, bus is always driven, e.g by
 ◦ snapper ◦ pull-up (unfavoured) ◦ centralized bus access ctrl

I-gate  drive conflict  low drive straight  back prop

Unsafe code

- may depend on external circumstances
- must make assumptions about env.

E2. Intro to Unions

2.1 Terminology

Module — | \cong VHDL entity, contains a level of design hierarchy

std cell — building block that impl. logic fun
is placed in core area

Macro cell → Larger, complex building block, e.g. RAN

Pad → Sudden block that connects to ping

Cell — Building block from library, e.g AND-X6... Instance — Specific copy of Building Blk

Row — | All std cells have same height and Terminal-HO of std cell
are aligned in rows

2.2 Clip overview

- routinely is done on several metal layers, on each layer either vertically or horizontally

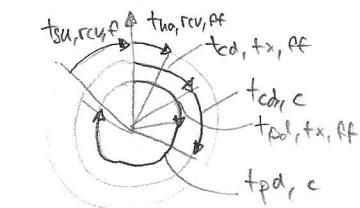
- Unicellular organisms can resize cells depending

- Cell syntax : MX12-X4-A8TL

UX12: type, 2 in MUL
X4: drive strength

A8: library, ARM 8-track
TL: threshold, low for high speed

5.2.1 Single Edge triggered one-phase clocking



positive skew \rightarrow more time for comb. logic. skew is lower bounded

$$-t_{sk} \leq T_{clk} - (t_{pd,tx,ff} + t_{pd,c} + t_{su,rx,ff})$$

$$T_{clk} \geq t_{ip} + \max|t_{sk}| \approx t_{id,ff} + \max(t_c) + \max|t_{sk}|$$

the shortest path upper-bounds skew

$$\max|t_{sk}| \leq \min(t_{ca,ff} + t_{cd,c} - t_{uo,ff}) = t_{sp}$$

Shift registers and scan paths are especially vulnerable to (positive) skew.

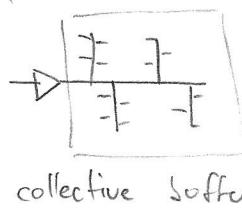
5.3 How to fix skew

- + fast clock ramps (low rise time)
- + distribute clock with same length and buffers
- + wires as short as possible
- + clock driver close to center
- + use low R, C upper metals

$$\text{current peak } \frac{I_{in}}{\text{fanout}} \times 2 \frac{C_{in,Udd}}{\text{frach}} = 2 \frac{10\mu\text{A} \cdot 2\text{V}}{100\text{ps}} \approx 2.0\text{A}$$

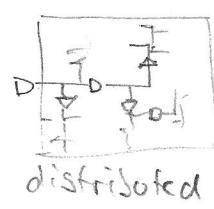
+ clk delay doesn't matter as long as it's balanced

- o t_{pd} is quadratic in #sections. with repeater buffer it is linear



- + easy
- larger power dissipation
- current crowding
- demanding routing

collective buffer



distributed

- + lower distribution delay
- + easier routing
- + clock gating possible

5.4 How to improve timing

	I/O timing	data valid win	$t_{so,in}$
input	friendly unfriendly	narrow wide	$t_{so,in}$ small large
output	friendly unfriendly	data valid win wide narrow	$t_{pd,out}$ small large

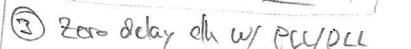
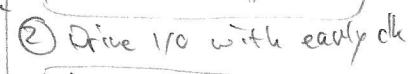
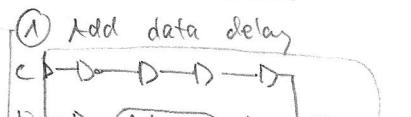
clk distribution delay can render a chip useless.
keeping it small is important

- + registers on input and output of chip helps. Input reg must not suffer from clk dist delay.

- + Adding artificial delay to data inputs is a way of improving timing

Ideal clk distribution: feedback loop with PCL/DLL

clocking tricks



E4. Design Review

active with rising edge
 active low of falling edge

- input
- output
- shift
- (+) odd
- (-) sos
- (X) awl
- (W) shift

comb clkt

state machine

complex sos

L6: Asynchronous Signals

Lecture: 47 slides

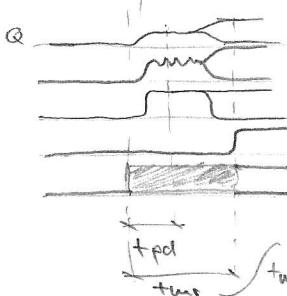
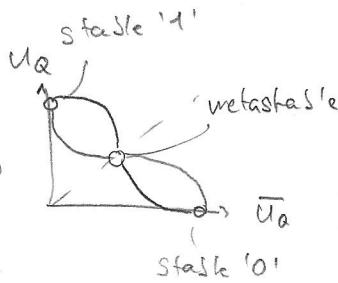


result:



more propagation delay

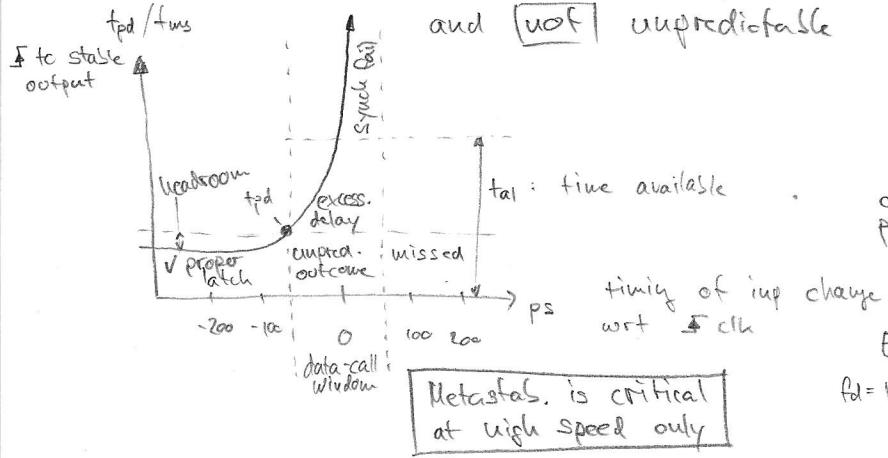
new equilibrium state
loss data retention of
two inv. amplifiers



no upper bound for t_{pd}
t_{met}: metastab. resolution time

Timing violation leads to propagation delay increase

Metastability is a problem bcs of unpredictable delay and NOT unpredictable logic outcome.



$$t_{MSE} = \frac{e^{K_1 t_{al}}}{K_1 f_{clk} f_d}$$

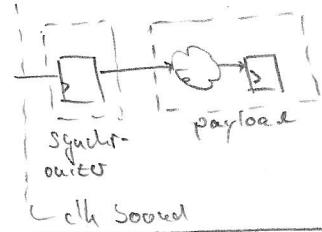
t_{MSE}: mean time SW errors
K₁, K₂: FF characteristics
f_{clk}: clock rate
f_d: data rate
tal: time available

$$tal = T_{clk} - \max(t_{pd}, t_{su}, t_x, t_{ff})$$

Ex: Virtex II @ 100 MHz $\rightarrow t_{MSE} \approx 3 \cdot 10^{52}$ years
f_d = 10 MHz @ 200 MHz $\rightarrow t_{MSE} = 4 \text{ s}$

plesiochronous: same freq.
but floating phase

Rule of thumb: Synch fail infrequent provided $tal \geq 3 \cdot t_{pd}$

single-stage synchronizer

$$tal = T_{clk} - \max(t_{pd}, t_{su})$$

Two-stage synchronizer

Almost entire clk period avail.
to recover from metastability

$$tal = T_{clk} - t_{su}$$

RZ & NRZ

synchronizer do's / don'ts

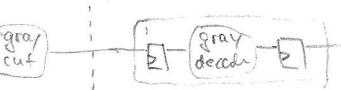
- + select FF w/ good metastab. recovery
- + remove comb delay from synchronizer
- + drive synchr. with fast switch clk
- + free synchr. from unnecessary load

Popular Synchronization schemesGuidelines

- Partition into as few clk domains as possible
- transition where SW is brief
- avoid clks that tend to fail in catastrophic manner

1. Unit distance coding

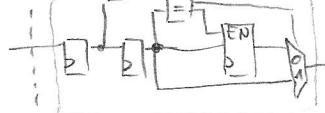
Any 2 adjacent codewords differ in exactly 1 bit



application must be static for this

2. Supposes jumbled data

Compare Subsequent words and discard those in transit



- + works with arb data seq.
- + simple
- High consumer clk! Double sample
- latency
- + could be SW implemented

3. Handshake

Avoid sampling while they might be changing
transfers get initiated, requested
→ concluded, acknowledged

- + strict event sequence
- + prevent data loss
- + modular design
- producer & consumer must be static
- FSM

4. Fifo

FIFO at clk boundary

- ? Asynch clk design ?
- o gray coded pointers

+ clks indep. from each other

+ handle unequal prod/con rate

- complex
- latency

E5. Timing in Back-End Flow

The delay of a digital gate depends on {current driving capability of the cell
capacitive load that is being driven}

PVT variation in timing due to process voltage and temperature.

Typical $\frac{1.2V}{25^\circ C}$ Worst $\frac{1.05V}{125^\circ C}$ Best $\frac{1.32V}{-40^\circ C}$

Timing reports WNS worst negative slack: Slack of most critical path.

WNS < 0 → timing violated

DRU Design rule violation (electrical, cap, transition, length)

Timing constraints stored in sdc file for $rg2reg$, $in2reg$, $rg2out$, ...
output loads

Optimize Libraries can optimize netlist by resizing gates, adding buffers, replicate logic
or even re-synth parts of critical path.

Op modes can set different constraints for e.g. scan chain mode...

MMAC Multi-Mode Multi-Clock analysis. Optimize clk at a set of conditions.
more complex setup.

Design Phases

↓
preCTS - after cell placement - trial route (early global route) estimates routing delay
- clocks considered ideal - upper bound fan-out

↓
postCTS - clock tree is synthesized and rooted - clock insertion delay is considered

↓
postRoute - clk is completely routed - optimization is harder

↓
SignOff - completely finished

Hold time can be easily fixed but not after mfc! Slower clock doesn't work.

Fixing by adding buffers or delay lines.

I/O timing out pads are slow bcs they drive high-Z loads. Multicycle paths say
that the pad changes only every Nth cycle. If this is not an option, you can add
latency by subtracting multiples of T_{ck} to offset timing.

L7. Power Estim & Low-Power Design Lecture: 64 slides

Efficiency is defined as energy per computation or per processed data item

$$E_{CP} = P \cdot T_{CP}$$

Problem with ...

- ... high speed : heat
- ... low power : battery drain by switching activity
- ... low activity : leakage

7.1 Energy dissip. in CMOS

- delay D \rightarrow makes node qnty larger

(charge transfer) Switching of capacitive loads

$$E_{CH} = \frac{1}{2} C_{L} V_{DD}^2 \text{ per clk cycle : } C_L = C_{Lq} + C_{Ld}$$

Node activity how many times a node switches per comp. avg. over many

DFF random inp: $a_n = 1/2 = \alpha_n$ $n \rightarrow$ node

$$E_{CH} = V_{DD}^2 \sum_k \frac{\alpha_n}{2} C_L \quad \alpha_n: \text{collect from simulation}$$

FIR: 0.3 CCDchip: 0.26 ...
Node activity is distributed uniformly! LSB toggles node

Crossover current flow through MOS during switch

$$E_{CR,k} = \frac{f_k B}{2 \cdot I_2} (V_{DD} - 2V_{TH})^3 \cdot t_{Rise} \approx f_k \frac{B}{2} C_L V_{DD}^2$$

B : gain f_k : fitting par. ≤ 0.2

- less relevant in low V_{DD} regimes

Static load DC path due to Resistors

- only in pseudo CMOS
- pull up/dn

$$E_{RL,k} = \frac{V_{DD}^2}{R_k} f_k T_{CP} \quad R_k: \text{resistance}$$

$I_{DS,OFF}$ Subthreshold conduction in FET that is turned off
 $I_{DS,REV}$ Reverse-biased drain-bulk & source-bulk junctions
 $I_{SS,REV}$ Reverse-biased well-well or well-substrate junctions
 $I_{G,TON}$ Electron tunneling through gate dielectric

Leakage

- subthreshold leakage $I_{DS,OFF}$
- gate tunneling/leakage $I_{G,TON}$

$$I_{LH} \propto \frac{I_{DS,OFF}}{SW} \sum_g W_g + 10A \left(\frac{V_{G,OFF} - V_{TH}}{S} \right) \sum_g W_g$$

Δ : subfr. current per width G n gates

W_g : width S : MOS subfr. slope, $70\text{--}120\text{mV/dec}$

\rightarrow leakage becomes a wide spread concern

$\rightarrow I_{DS,OFF}$ growing with temperature

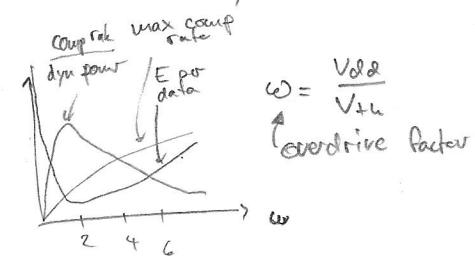
- minimizing V_{DD} , α_n , C_L and node count k reduces both E_{CH} & E_{CR}
- most important single factor is V_{DD}
- Overall power grows with freq. $\frac{1}{f_{Rise}}$ energy per op decreases \rightarrow efficiency increases!

V_{DD} voltage scaling

• good scs $E_{CP} \propto V_{DD}^2$

• further reductions of V_{TH} and V_{DD} get frustrated by subthreshold conduction and its critical dependence on threshold voltage $I_{LH}(V_{TH})$

\rightarrow managing leakage is crucial for high eff. devices



7.2 How to improve energy eff.

- ① keep track where spent & improve op. cond.
 - speed \rightarrow minimize $\frac{V_{DD}}{V_{TH}}$ activity \rightarrow maximize V_{TH}
 - power \rightarrow minimum V_{DD}

speed power
 \downarrow
 activity \downarrow
 sub-th \downarrow
 V_{TH}

DVFS: dynamic voltage/freq. scaling

- ② Cut irrelevant switching
 - local data \rightarrow less compute
 - no DRAM \rightarrow avoid idle/polling

- clk gating
- silence inputs to large subcts

- ③ Cut parasitic effects
 - don't go off-chip \rightarrow no polls \rightarrow do not oversize buffers

- ④ Use few low- V_{TH} MOSFETs
 - longer MOSFETs with higher V_{TH}
 - variable threshold logic

E6. Power Analysis

dynamic	{ charge/discharge crossover }	{ only during transit inside cell }	$P_{dyn} = P_{out} + P_{ext}$
static	{ driving resistive loads leakage }	{ always crossover }	$P_{ext} = f_{op} \frac{1}{2} C_{ext} V_{dd}^2$

SDT are exported from Itronics, contain info about interconnect and cell delays.
VCD value change dump, can extract node activity

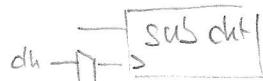
types of power analysis

stimuli based

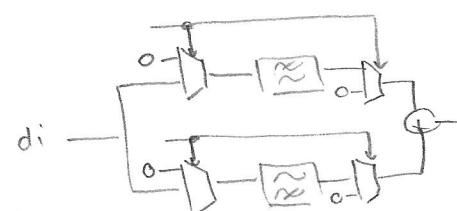
Modelsim \rightarrow VCD \rightarrow Itronics
with toggle data from simulation
Modelsim uses netlist from Itronics

- + Statistical
- + Global activity: Set for all internal nodes (pads consume a lot)
- + Input activity: Define & only on inputs

Architectural changes to save power



clock gating turns entire clock off



disable switching in FIR filters
if output not required

L8. Parasitic Effects in IC Design

Lecture: 35 slides

A NOS source faces different loads:

- R resistive load
- C capacitive
- L inductive
- G nonideal insulators

- Electromigration
too much current destroys metal connections

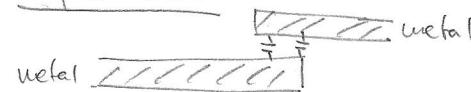
Resistance

+ larger X-section \rightarrow lower R + used conductors: Cu, Al, Poly Silicon

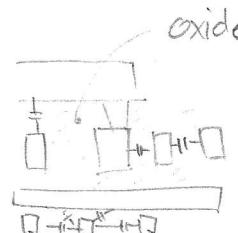
$$\square = - \square = 1 \cdot R_D$$

$$III = - \square = 3R_D$$

Capacitance



X-Section:



everywhere!

- small R leads to large C
- large sidewalls

Solve
 ignore: not feasible for λ_{min}
 estimate: statistically, not very accurate
 calculate: field solvers, takes time

solve: larger wires

power distribution
 Electromigration: reliability issue, occurs gradually. too much current carries metal away
 "e wind pushes ions down wind" \rightarrow voids occur, extrusions make shorts

Al: $J \in 5..10 \text{ A}/\mu\text{m}^2$
 Cu: $J \in 25..50 \text{ A}/\mu\text{m}^2$ max!

IR drop static: causes V drop at gates \rightarrow they work slower

$$U_g(t) = U_{gj}(t) + L_j \frac{dI(t)}{dt}$$

stat
dyn

Supply pins to mitigate IR drop

+ thicker wires

supply droop voltage change in GND reference
causes wrong bit.



E8. Power Distribution

Inuous Rail Analysis. Shows current density, ground source, supply drop.

IR analysis threshold for minimum voltage on V_{DD} = IR drop in Inuous
 heat map shows IR drop must be provided: 1.2 V min core 1.08
 $\rightarrow V_{DD}$ thresh = 1.14 V (half)

L9. Physical Design Automation Lecture: 49 slides

9.1 Synthesis includes:
 - RTL to gate-level descr. (AND, FF, ...)
 - Verification for correctness
 - Map to standard cells

9.2 Partitioning

- Decompose large design into components
- each piece of manageable size
- minimize connections between subs
- HDL hierarchy is taken into account
- e.g. Kernighan-Lin algo

constraints:

- num. terminals in each sub
- area of each part
- number of partitions
- critical path should not cut boundaries!

- top-down (take netlist and split) or bottom-up (single cells and connect)

9.3 Floorplanning

- output of partition used for floorplanning
- in: blocks with shape, blocks w/o shape (approx area), netlist for block connections
- out: location for all blocks
- 10s and 10 constraints

- objective: min area, less wires, max. routability

- constraints: shapes, areas, pin locs, aspect ratio

9.4 Placement

- in: set of fixed modules, netlist
- out: best position for each module
- cost: wire len, routability, hotspots
- split into global and detailed/local placement
- simulated annealing algo

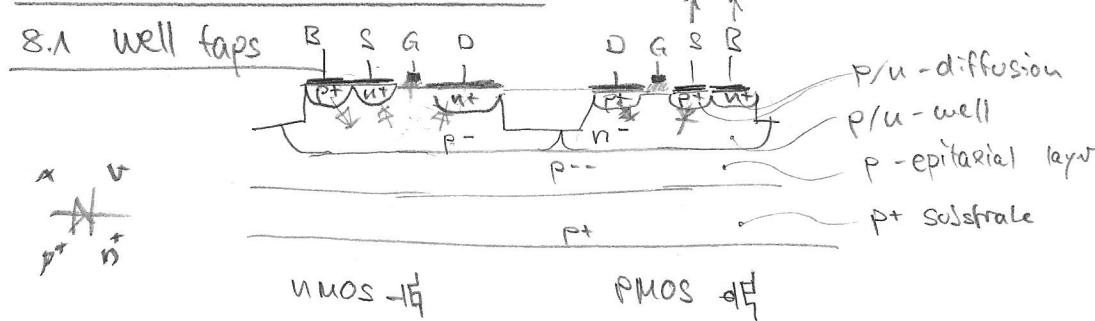
9.5 Routing

- connect std. cells with wires
- in: cell locs, netlist
- out: layout of wires
- two-step: global, local
- global: loose routes and allowed regions
- obj: 100% conn, min area, min wires
- constr: #layers, design rules, timing, crosstalk, process variations
- dock: arrive simultaneously, minimum delay
- H-tree with buffers
- power: low resistance

special blocks:

Future: ML in P&R

E8. Place & Route



Bulk of trans. needs to be connected to GND (NMOS p-well) or Vdd (PMOS, n-well)

insert well taps in a regular interval

8.2 Placement

problem is NP-complete. Can help by specifying scan chain.

Options: Congestion effort, timing driven, Module Plan, Scan Connection, Maximum density

8.2.1 Replace tie cells that force I/O on pin. can have many of these

8.2.3 Optimize run time Design to get first DRVs and fix them

8.3 Clk tree insertion

Parameters: Transition time, insertion delay, skew
↳ rise/fall time ↳ driver to endpoint ↳ difference in arrival time

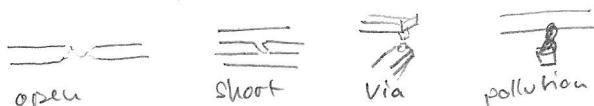
possible to define skew groups, different (macro) cells have different requirements

8.4 Signal routing

again, launch opt Design - post route after this step

L10. VLSI Test Lecture: 52 slides

10.1 Fault defects



Testing through ATE (automated IC test equipment)

criteria:

$$D_L = \frac{\# \text{defective parts sold}}{\# \text{total parts sold}}$$

defect level automatic < 2ppm
customer 100...200ppm

$$D_L = 1 - Y_F^{1-F_C}$$

assuming stuck-at faults

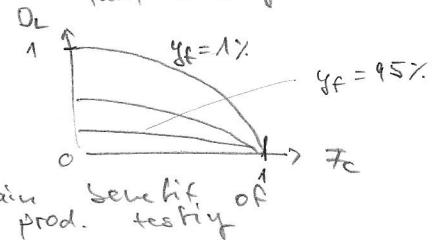
$$Y_F = \frac{\# \text{good parts shipped}}{\# \text{parts manufactured}}$$

fabrication yield
new nodes (10%...)
old, experienced 90-95%

$$\begin{cases} \text{for } F_C \rightarrow 1, D_L \rightarrow 0 \\ \text{for any } Y_F \end{cases}$$

$$F_C = \frac{\# \text{faults detectable}}{\# \text{faults possible}}$$

fault coverage



10.2 Stuck-at model

any defect is either S_{A1} (D_{11}) or S_{A0} (D_{10}) stuck at 1 or stuck at 0, assume only 1 fault at a time

ATPG : automatic test pattern generator

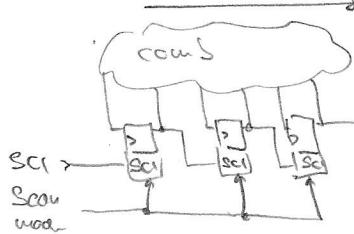
so that it propagates to observation point

10.3 Design for Test

Detecting a defect set correct input implies
can be done only via
be able to observe regular package pins

10.3 Design for Test (cont.)

Goal of DFT: Provide observability and controllability

Scan Testing¹⁾

daisy chain
all FF with
scan chain

1. Shift pattern in
2. run for 1 clock
3. shift out
4. compare

Efficient, elegant & popular

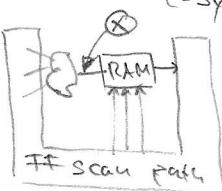
BIST

built in
self test
generate stim &
do resp. check on
chip

EAT Testing

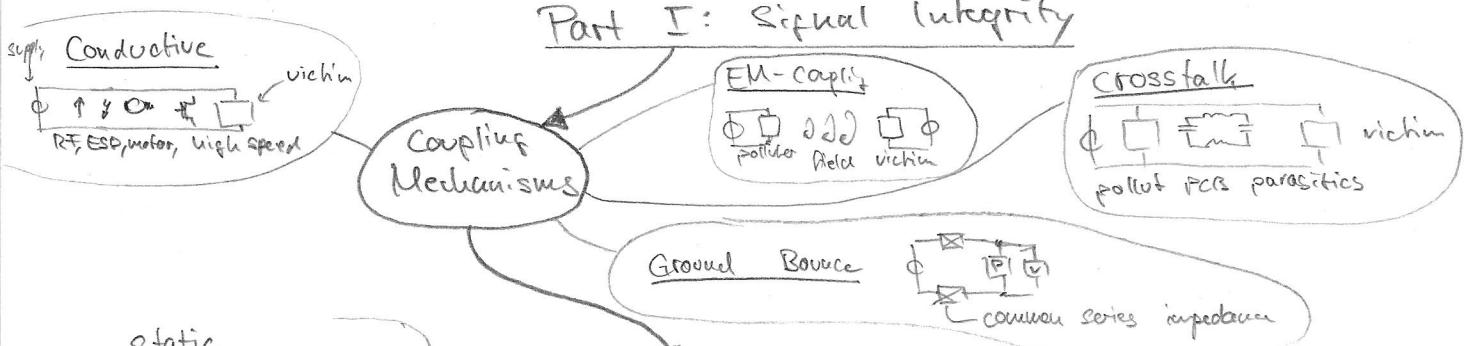
- 1) also: - check if reg chain works
- check if reset works
- 2) caveats: - each clk domain needs a chain
- requires special FF
- bypass clk gates - Bidir I/O caution!

this RAM is not fully testable bcs no control over
input of RAM \otimes . If defect in \otimes we don't know.
Insert \square at \otimes and add 'em to scan chain to fix.

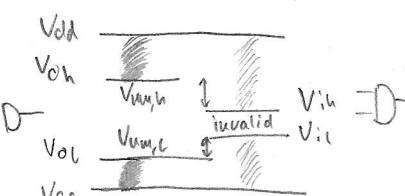
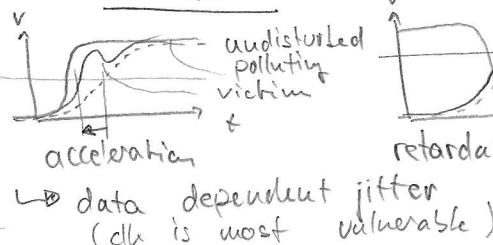
example

scan chain: 4x FF, 1x SIT RAM, #a=6, #d=16 \rightarrow #sc = a lot

speed up \circ add RAM scan FF to start of chain \quad for partial scan tests
 \circ Isolate RAM with its own scan test

III. Signal Integrity & Aging Lecture: $34+31=65$ slidesPart I: Signal Integritystatic

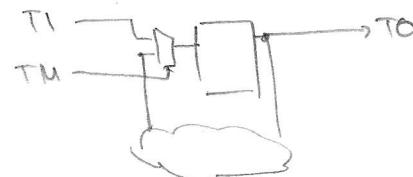
logic only works if
- keep 0/1 apart
- restore level at output

Impact on Signalstransient

$$\text{noise margin } V_{nm} = \min(V_{nm,u}, V_{nm,e})$$

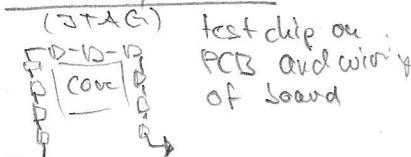
$$\text{SVCMOS} \rightarrow 1.4V, \text{QAVCMOS} \rightarrow 0.22V$$

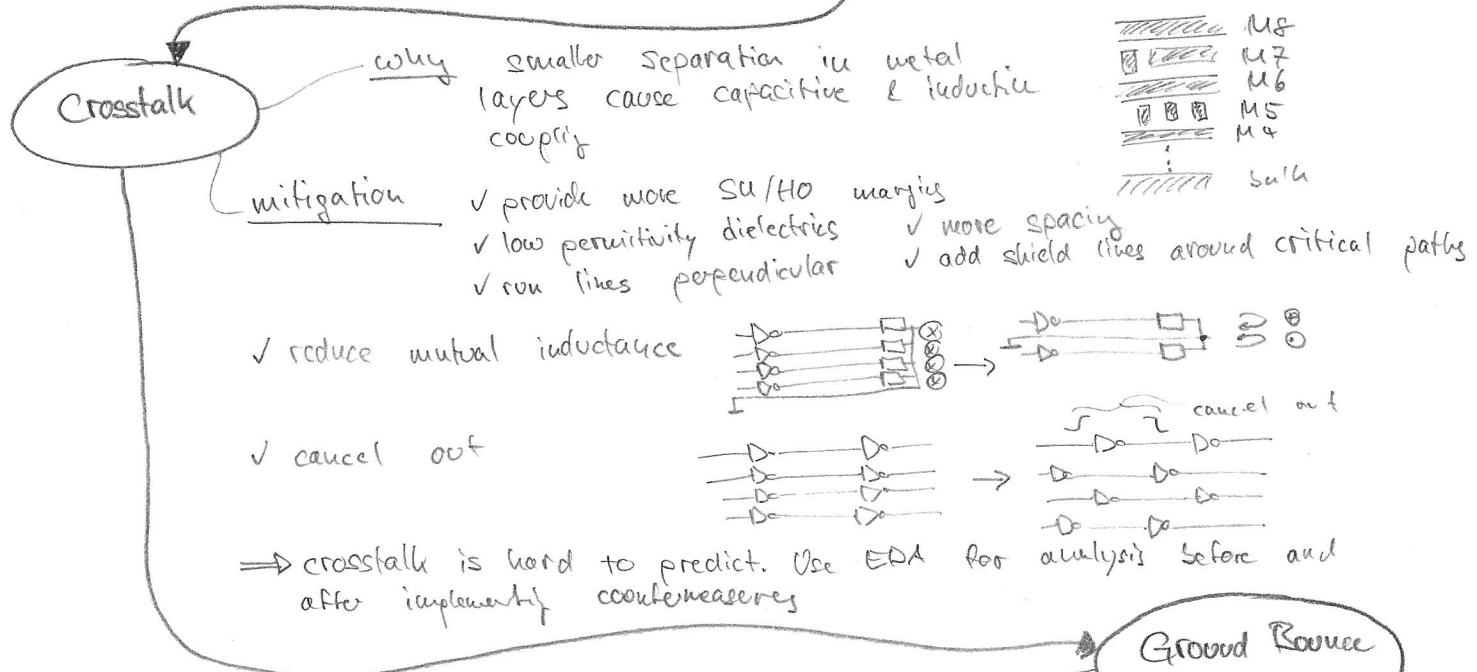
make part of clk
available on poly pins
for testing

Block isolationIddq monitoring

most faults cause non-zero supply current after settling

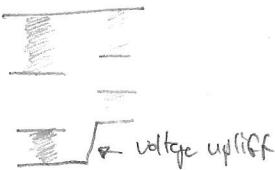
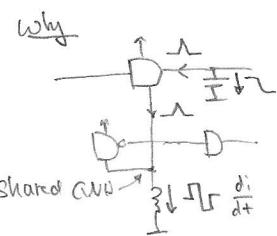
- must be done at slow clk
- divide clk into multiple domains for separate monitoring

Boundary scan (JTAG)



MS
M7
M6
M5
M4

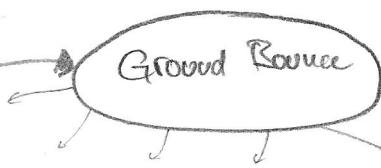
su/h



impact

- extreme ampl.
- ↳ loss of data stored
- ↳ uncontrolled oscillations
- strong ampl.
- ↳ extra glitching
- moderate ampl.
- ↳ depresses switch speed
- ↳ susceptible to external circumstances

Ground Bounce



mitigate

- ✓ decouple core voltage from pad drivers
- ✓ separate path for transient and stationary drivers
- I drives for transient
- II drives for static
- ✓ many power pins
- ✓ LVDS
- ✓ don't switch all at same time (=staggered outputs)
- ✓ bypass caps on multiple levels
 - o external o cavity o die

Conclusion

- current spikes are in GHz range
- design countermeasures

Part II : Aging

Problems of IC

Mechanical

- edges prone to crack
- bond wires can break/shear due to vibrations

Packaging problems

- mech. stress
- humidity

Thermal

- soldering stresses plg
- heat produced by IC → melts

Electrical

- ESD
- Electromigration
- latch-up (parasitics trigger avalanche currents)

MOS problems

- El. migration
- ESD

- ✓ seal ring protects from delamination, stops cracks
- ✓ pad window opening, min pitch s.t. wires can bond properly
- ✓ reflow guidelines to reduce thermal stress

latch-up ✓ proper substrate biasing needed everywhere

✓ lower Vdd not vulnerable

Oxide breakdown: defects in oxide cause short. trapped charges in oxide
 Hot carrier injection: source drain current charges can end up in oxide → V_{th} increased, slower switching
 Bias temperature instab.: even at no current, V_{gs} can cause charges to migrate into Gate Oxide

Conclusion

→ aging will happen → gradual process

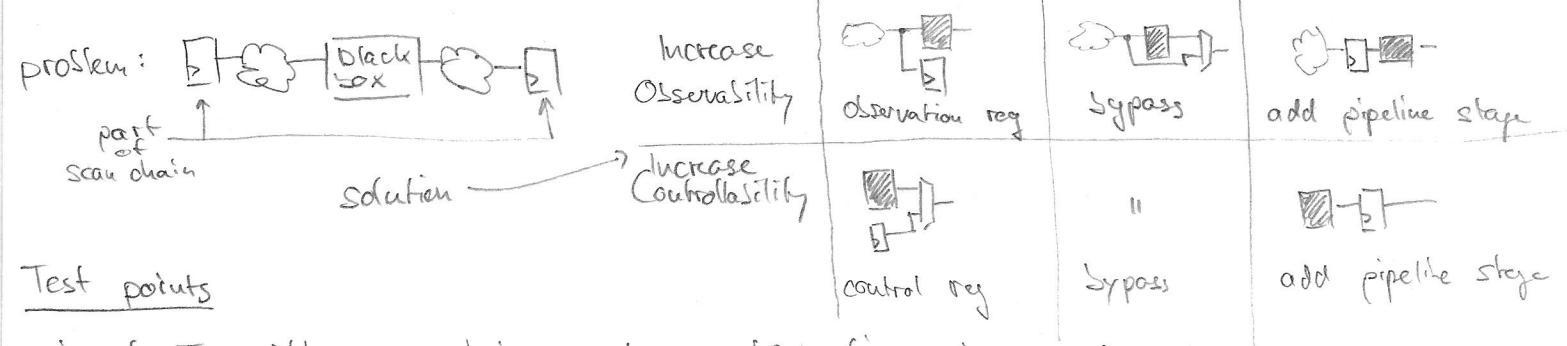
→ Smaller \downarrow more vulnerable → turn off unused parts, reduce Vdd, activity, frequency

E10. Scan Insertion

- Scan chain
- full/partial scan - needs additional pins \rightarrow change in HDL
 - is inserted during synth before or after design opt
 - ATPG used to generate test patterns, can be compressed to reduce #cycles

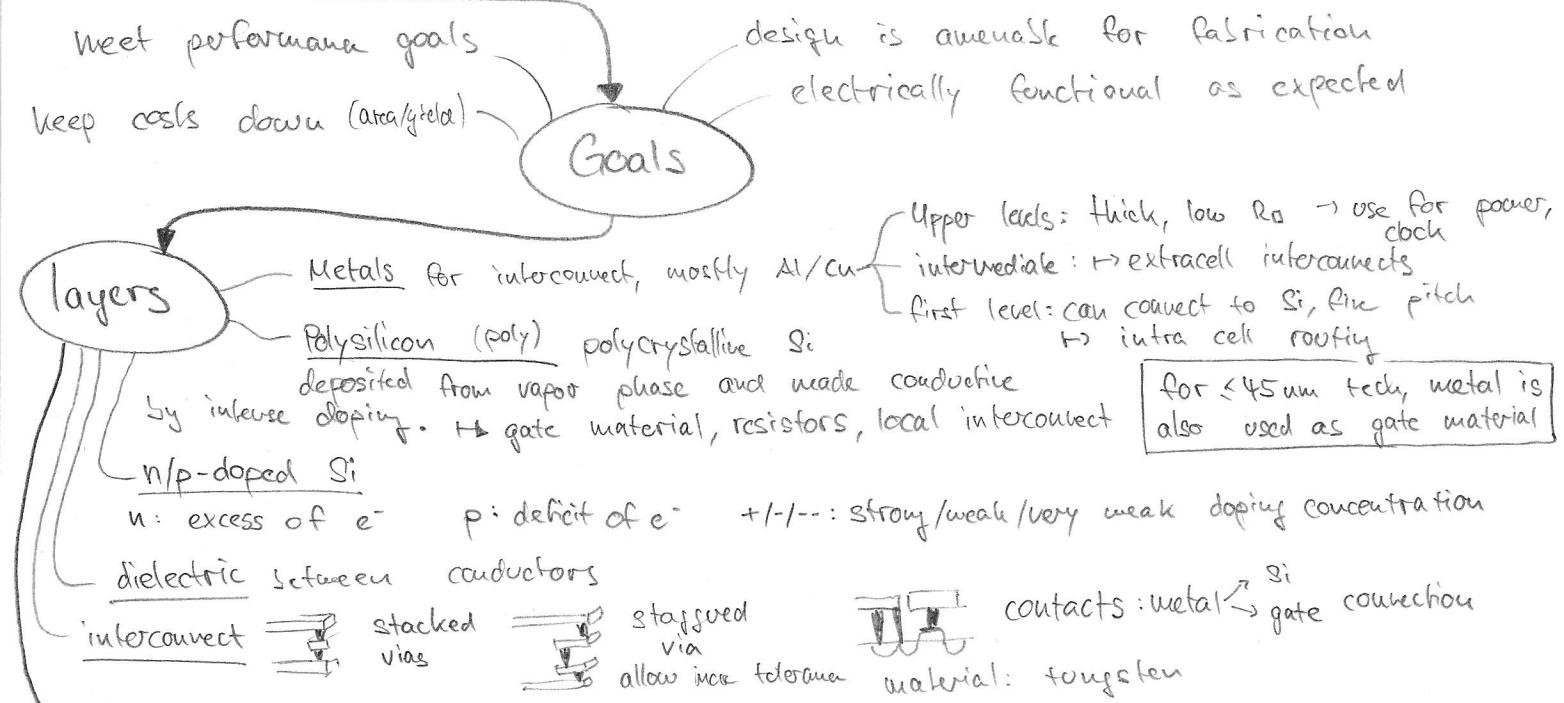
partial scan does not include all registers. does not save significant area, better test all

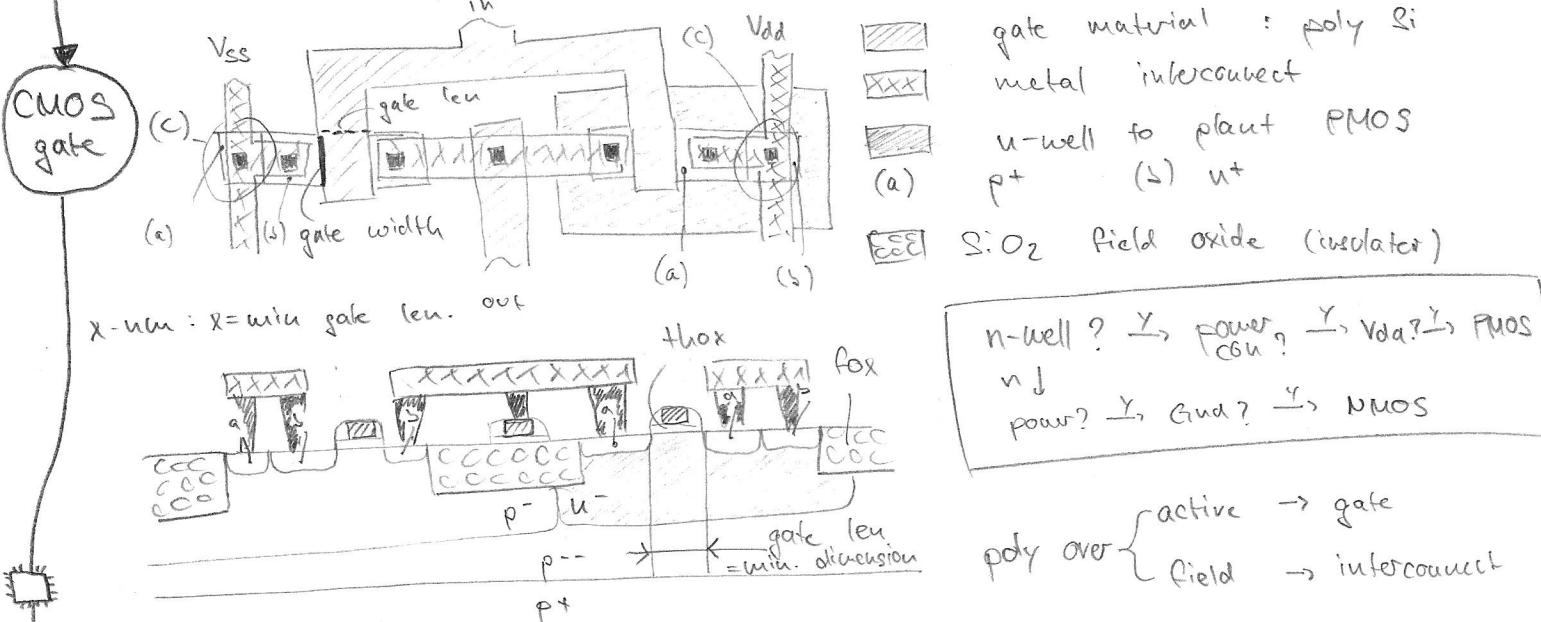
Issues with Macrocells they don't always support scan chaining \rightarrow use isolation

Test points

insert FF with scan chain where observation is required.

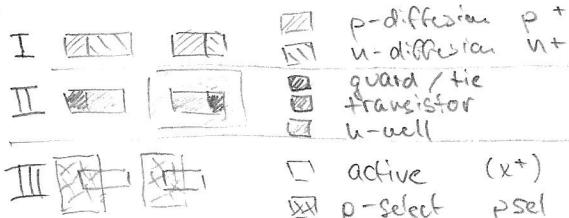
⚠ make sure they are not being optimized out by synth ⚡

L12. Physical Design Lecture: Si slides



(c) Body ties to prevent parasitic diodes

diffusion area coding styles



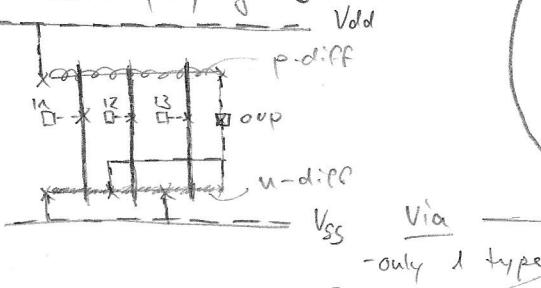
- requires pos proc. to get masks
- named after doping type & concentration
- named after role
- post proc. required
- represent masks
- used only in nfc

objective
 - min area
 - max perf.
 - max yield
 - less parasitics
 - min design effort

proven patterns

Horizontal: n/p diff stripes
 power/gnd rail

Vertical: poly gates



density chip is polished after each layer. If uneven metal densities, wells can form

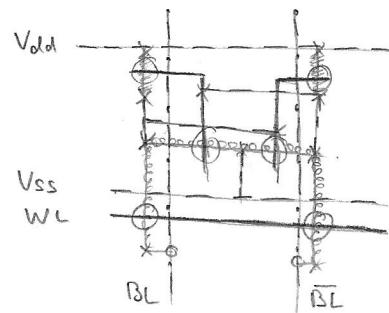
antenna during nfc, free charges accumulate on top metal layers and could cause oxide breakdown. ① reduce exposed metal
 ② antenna diode



layout

example: 6T SRAM cell

manual only if - high prod. vol.
 - library dev. - analog - sensors



— gate material
 -- M1
 --- M2
~~---~~ p-diffusion
~~---~~ n-diff

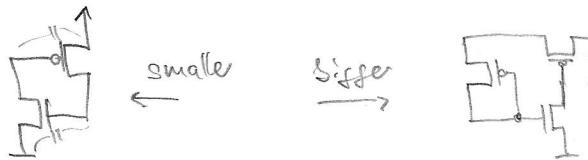
layout rules
 why?
 - tolerances
 - Al spacing
 - ...

min size, width, space, enclosure
 extension
 slotting
 limits mech. stress when tracks are too wide

restrict track widths/pitches to values that work well
 place on periodic grid
 Single MOSFET orientation
 avoid seats in gate
 use multiple vias/contacts

E11. Chip FinishFiller cells

free space around std cells is filled with them. Small ones have no purpose but bigger for capacitance. Now utilization is 100% after filling.

Verification

- DRC: Design Rule Check
- LVS: Layout versus schematic
- process antenna: long wire to gate and unconnected for several steps. These need to be removed or antenna diodes added
- verify well taps
- summary reports: cross-check #macro cells, #pads, area, max fanout
- routing utilization: check how crowded stuff is

export

- lvs.v netlist incl. physical & filler, use for LVS
- v netlist w/o physical, use for post layout sim
- .sdf.gz delay info
- .gds.gz layout for fab

Violations

- re-route can fix some
- ECO (engineering change order): delete partial nets and re-route
 - ↳ allows even small RTC changes to save time
 - ↳ place some unused gates, with ECO can add them (e.g. inact signal)
 - ↳ export netlist, make changes, load modified with ECO command

E12. Physical Verification

Use Calibre Designer to load GDS2

- can only edit top layer of hierarchy & layer

- use it to run DRC

- checks for antennas as well

- use to generate SPICE netlist for LVS

can have YN data on multi hier. levels
they get mapped into 1 YN mask

Why should there be DRC/LVS violations?

- manual changes (logo, pads, corners, PCM (process control monitors))
- error in tool/library
- invokers has abstract view of std cells and no knowledge over every bit
- boundaries specify DRC tool - Second opinion from different tool is good