EECS251B : Advanced Digital Circuits and Systems

Lecture 10 – Technology Features

Borivoje Nikolić

February 2, 2024, EE Times: Tenstorrent Engineers Talk Open-Sourced Bare-Metal Stack.

Tenstorrent, the Jim Keller–led AI chip and IP startup, is making available and open-sourcing its bare-metal software stack, which allows access to the hardware at the lowest level, Tenstorrent senior fellow Jasmina Vasiljevic told EE Times. The startup also recently showed EE Times a large language model (LLM)—Falcon-40B—up and running on its 32-chip Galaxy system, and it has made hardware available to purchase for the first time in the form of evaluation kits for its first-generation Grayskull chips.

https://www.reuters.com/technology/asmls-next-chip-challenge-rollout-its-new-350-mln-high-na-euv-machine-2024-02-09/

Berkeley UNIVERSITY OF CALIFORNIA
BY NC SA

Project options

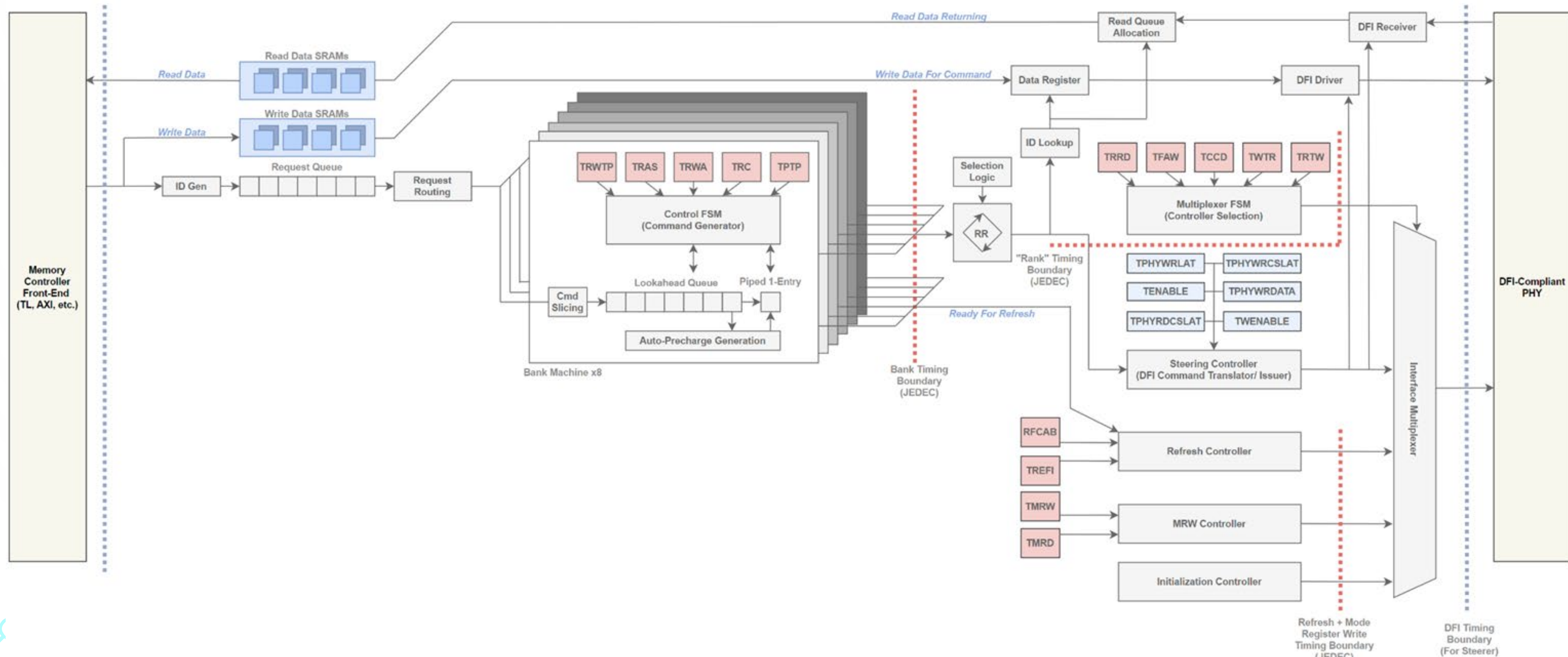
Develop complete modules, appropriate for SkyWater 130nm technology:

- LPDDR4/4X
- USB 2.0

Integrate in Chipyard

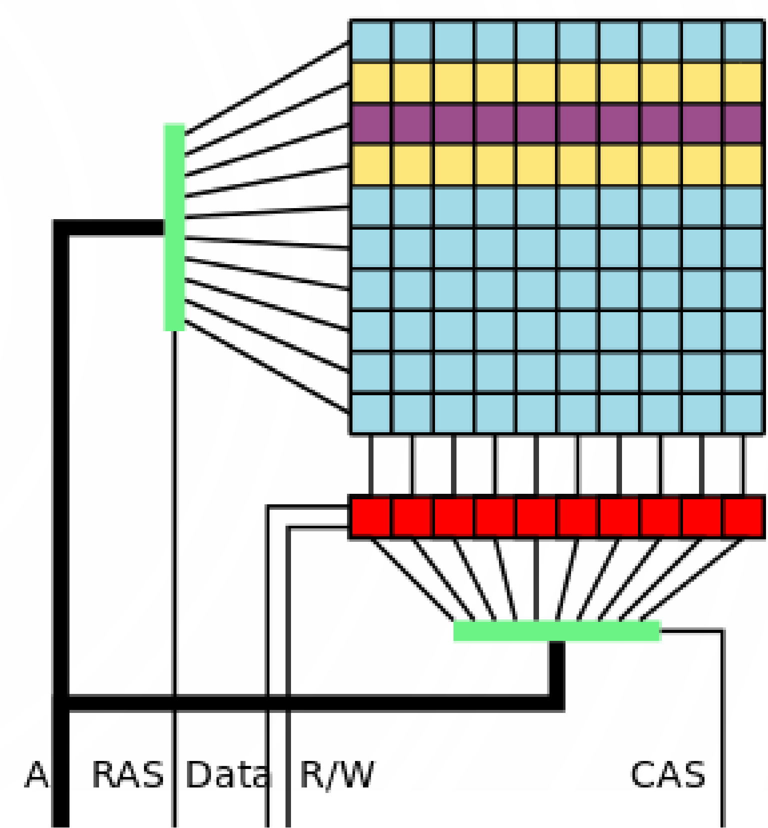
LPDDR4x

- Targeting slowest speed (800 MHz) in Sky 130
- Barebones memory controller RTL already implemented in Chipyard



LPDDR4x Digital

- Thorough verification
  - Timing, spec compliance, additional features
- Extensions:
  - Memory scheduling algorithms
  - Rowhammer exploit counter measures



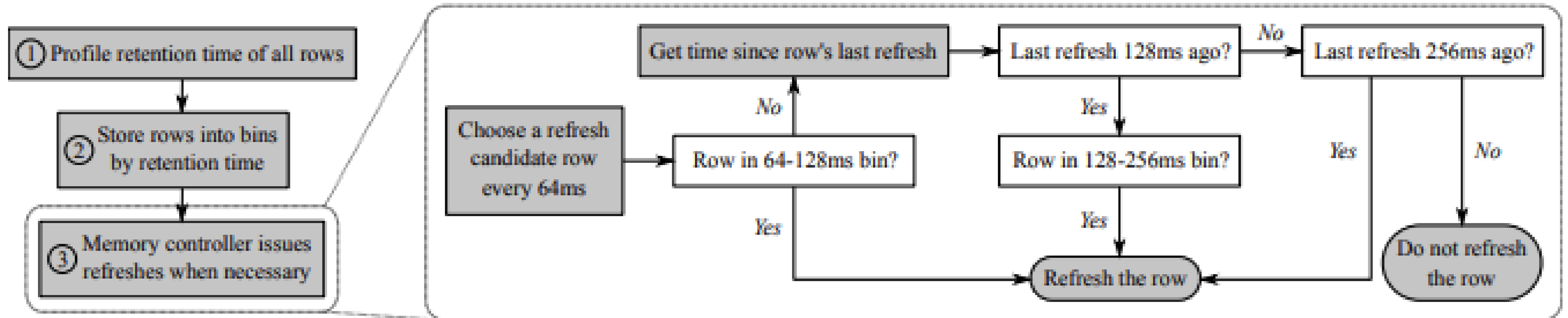
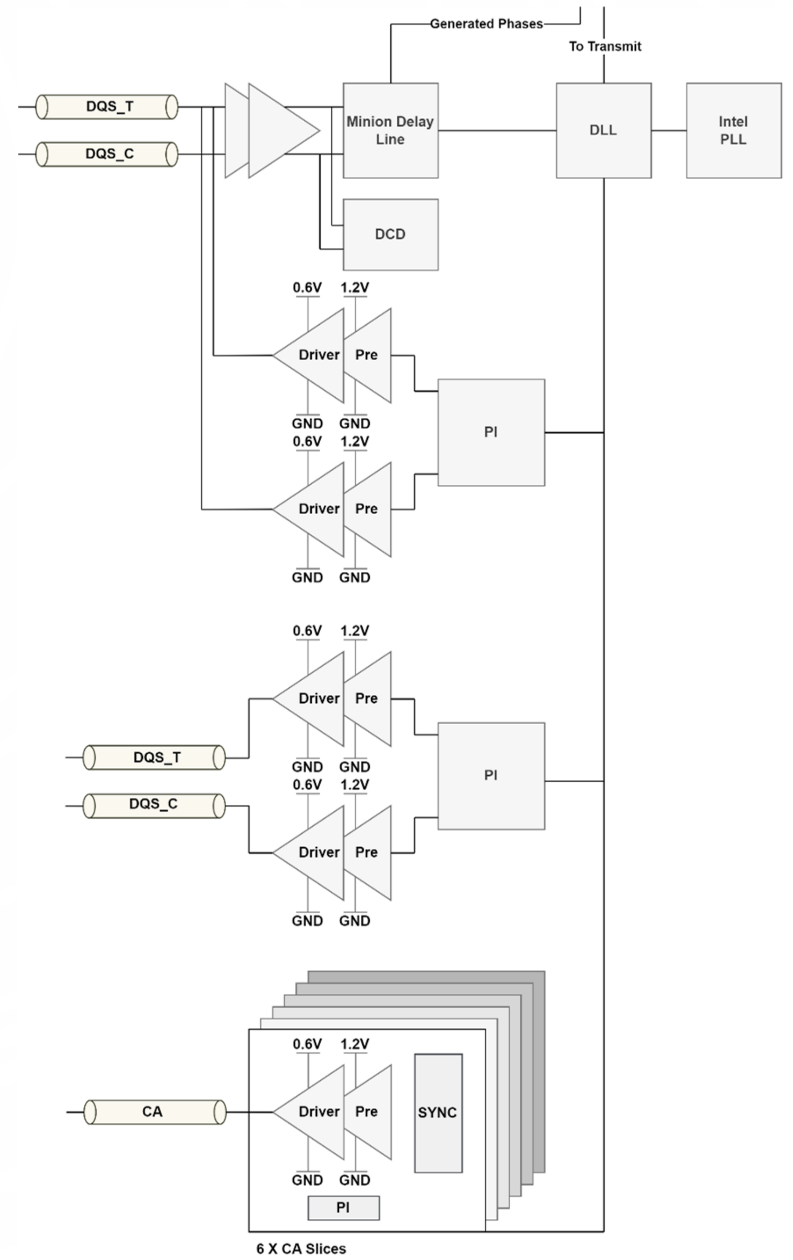


Figure 4: RAIDR operation

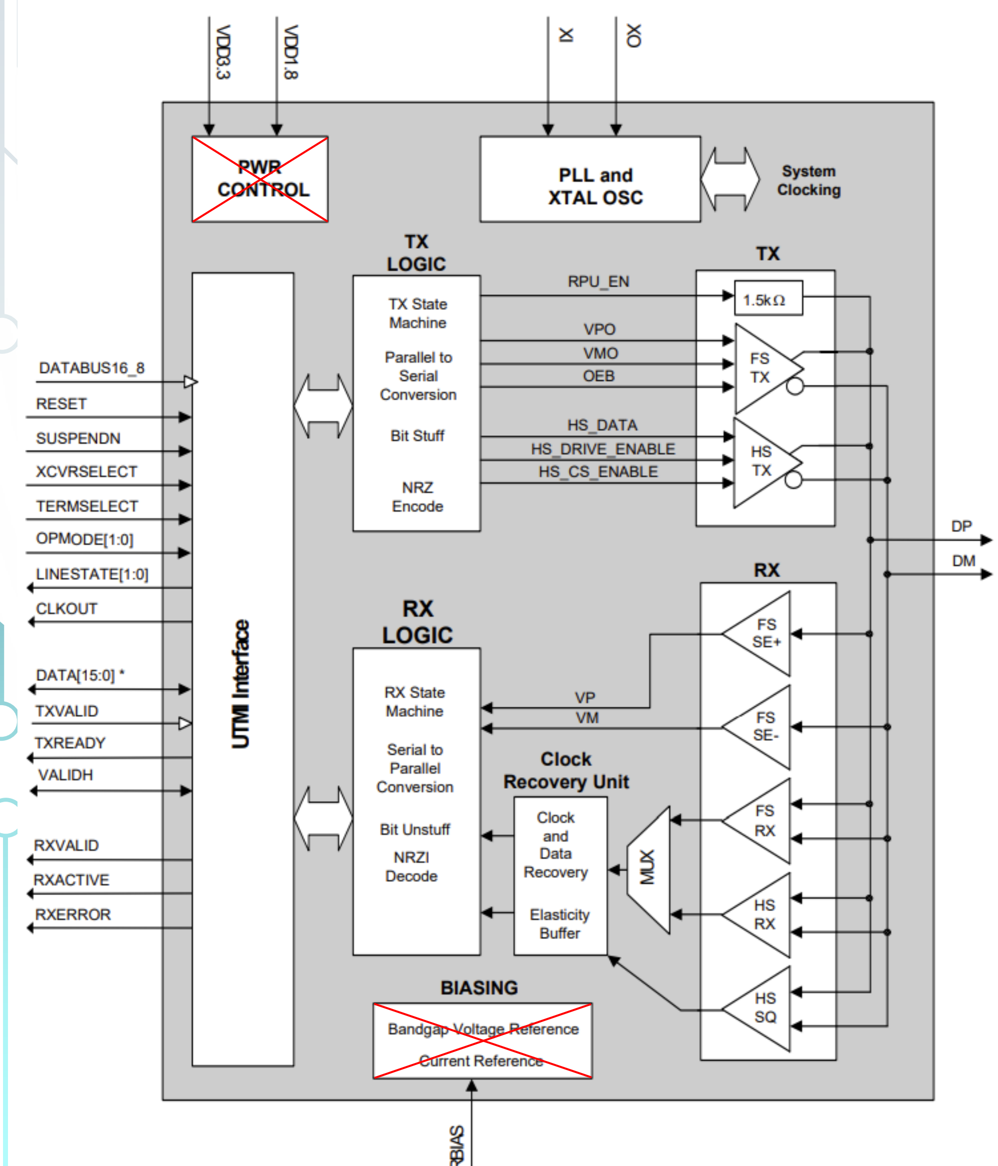
LPDDR4x Mixed-Signal

2x 8-Lane Byte Groups

0.6V
1.2V
Driver
Pre
GND
GND
Sync
PI
DQ
0.6V
VREF
GND
Sync
PI
LANE CIRCUITRY
Generated Phases
To Transmit



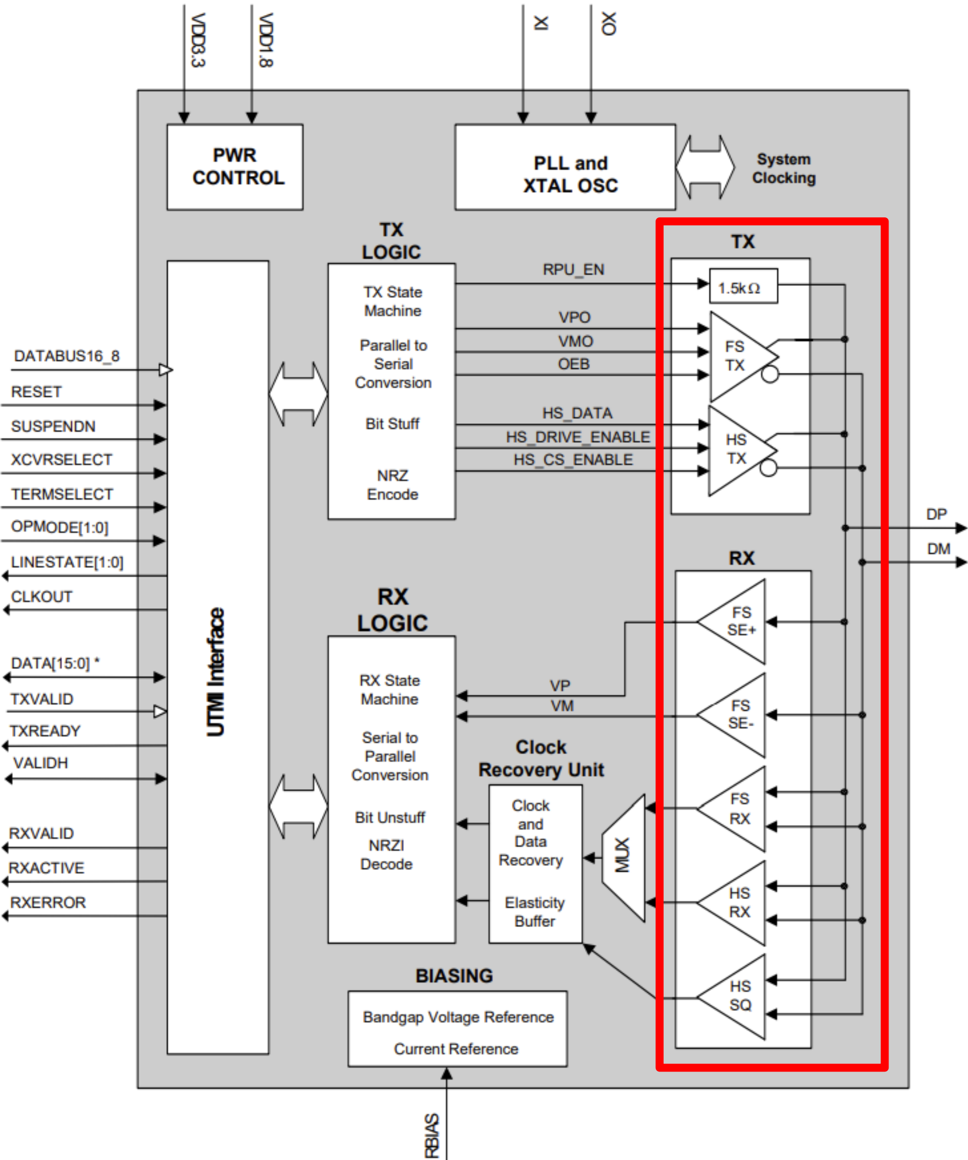
USB2 PHY High-Level Overview (in progress)



- USB2 operates at 12Mbps (“full speed”) and 480Mbps (“high speed”)
- USB organization provides implementation specs: https://usb.org/sites/default/files/usb_20_20211008.zip
- Can consult the many USB2 PHY standalone chips and IP blocks for implementation details and ideas

Example commercial USB2 PHY chip: https://ww1.microchip.com/downloads/en/DeviceDoc/00002142A.pdf

USB2 TX/RX Circuits



ices

o.

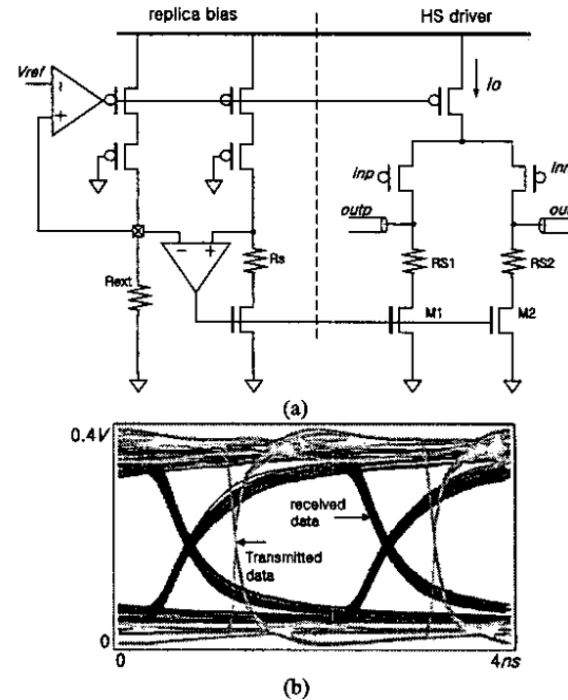


Fig. 4. (a) HS current steering output driver with replica bias. (b) Simulated eye patterns at transmitter and receiver ends

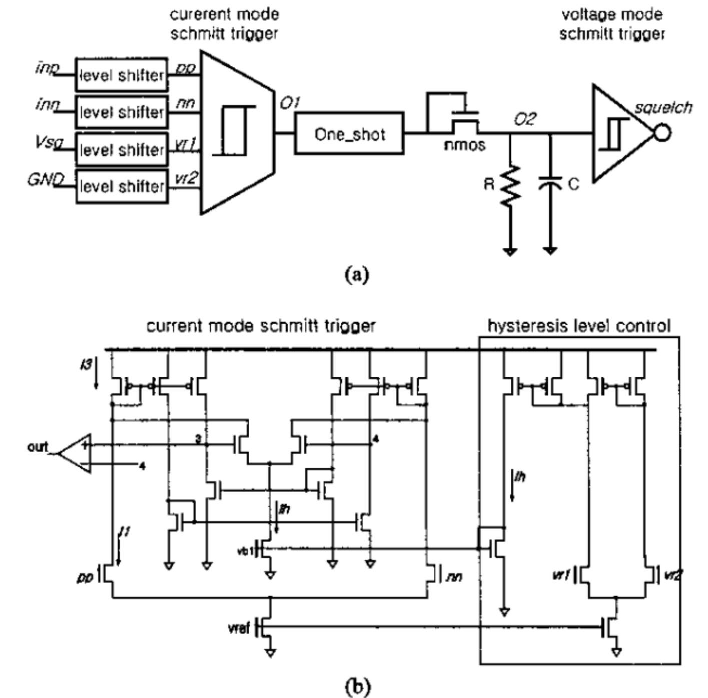
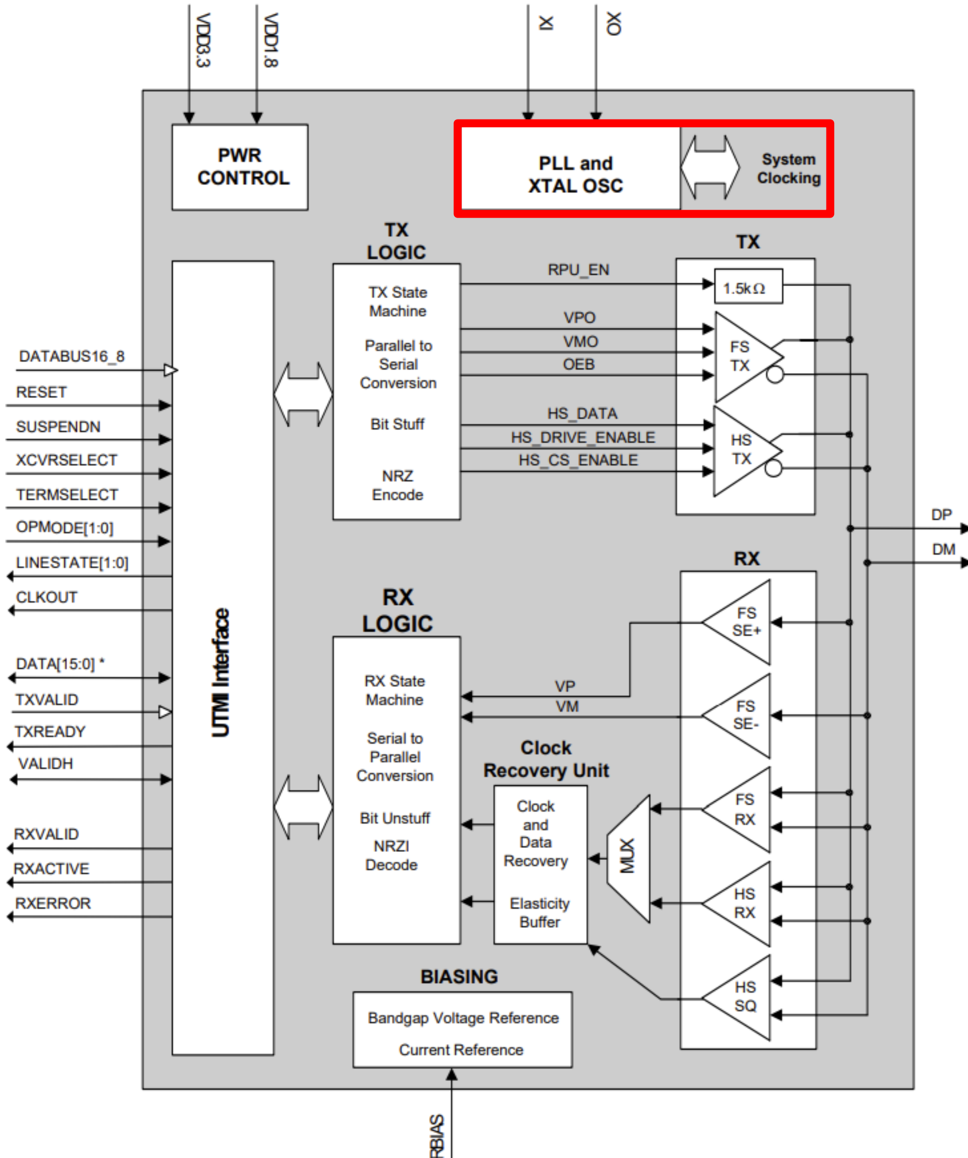


Fig. 5. (a)Transmission envelope detector (b)current-mode schmitt trigger with squelch level control

https://ieeexplore.ieee.org/abstract/document/1241532

Example commercial USB2 PHY chip:
https://ww1.microchip.com/downloads/en/DeviceDoc/00002142A.pdf

USB2 Clocking



Example commercial USB2 PHY chip:
https://ww1.microchip.com/downloads/en/DeviceDoc/00002142A.pdf

- 5-Phase 480MHz PLL for HS datarate
- Goal is to fully synthesize and P&R the PLL using digital tools
- Article below does it in 65nm CMOS – may need some improvements for 130nm

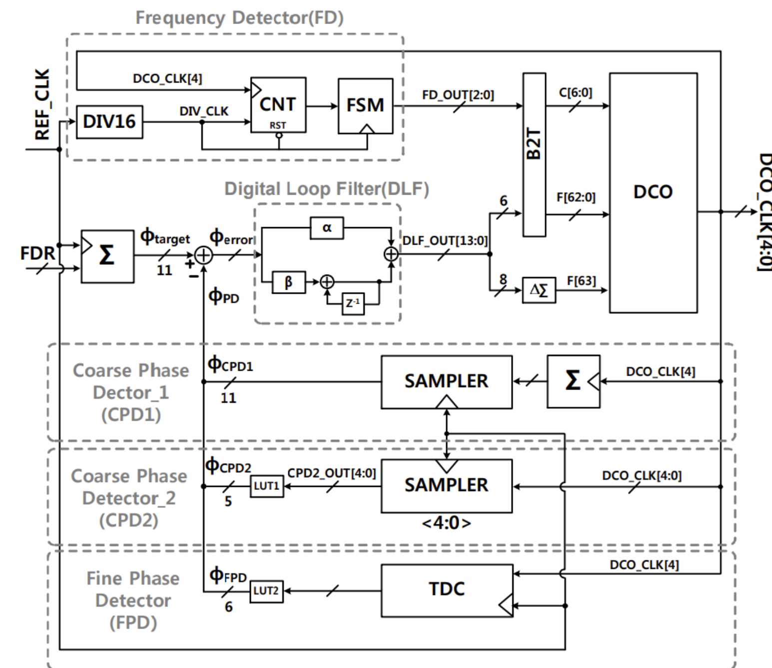
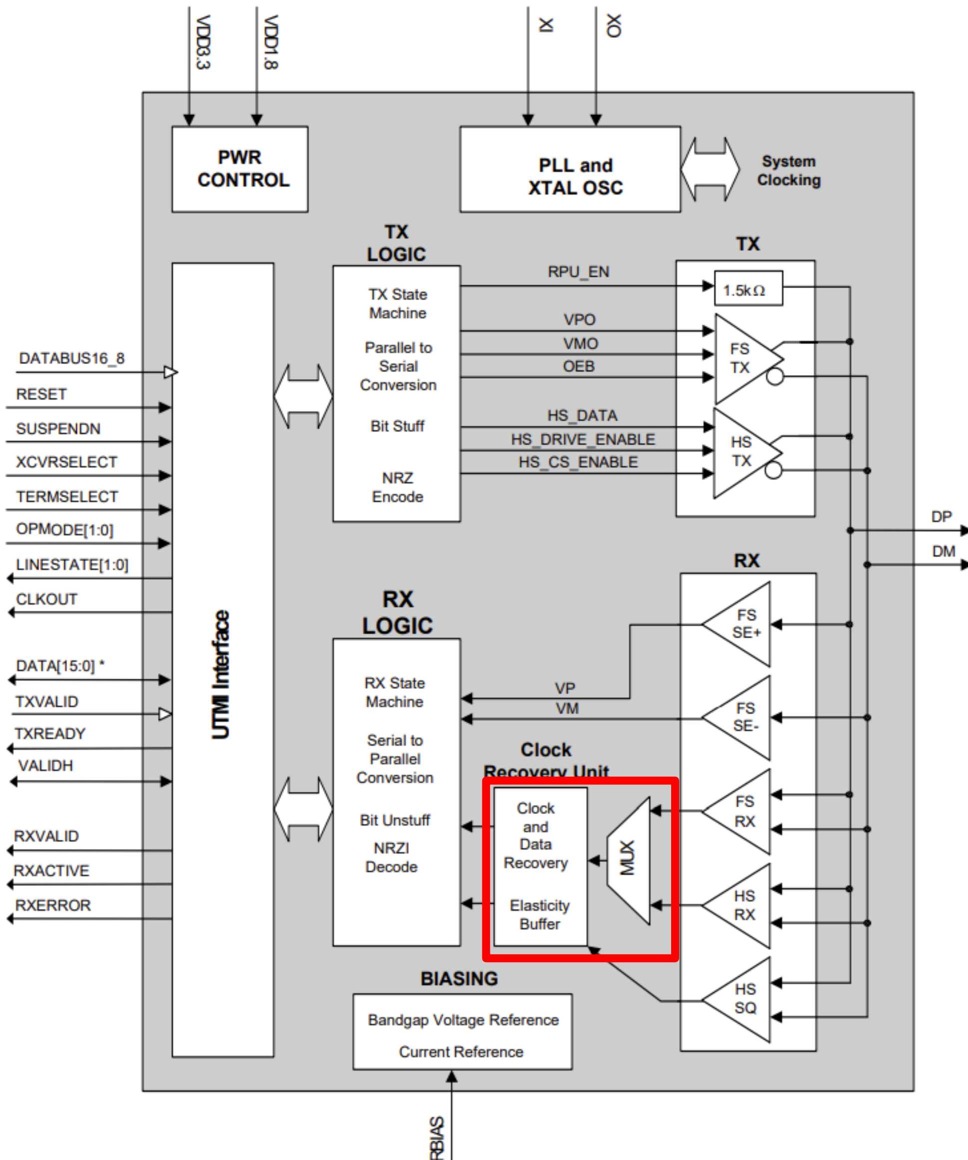


Fig. 1. Proposed all-synthesizable 5-phase PLL.

https://www.researchgate.net/profile/Juno-Sim/publication/305418056_All-Synthesizable_5-Phase_Phase-Locked_Loop_for_USB20/links/5cd22348299bf14d957e7529/All-Synthesizable-5-Phase-Phase-Locked-Loop-for-USB20.pdf

USB2 Blind Oversampling Data Recovery



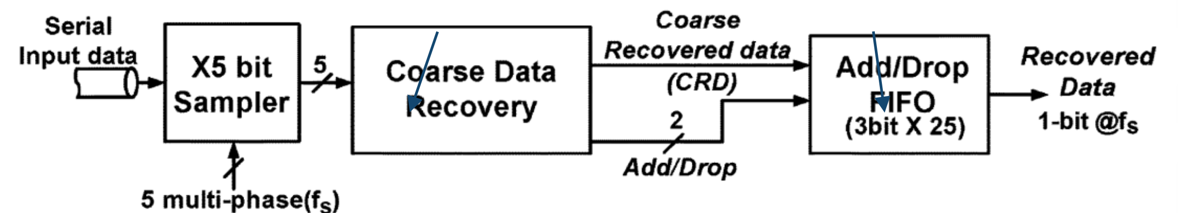
Example commercial USB2 PHY chip:
https://ww1.microchip.com/downloads/en/DeviceDoc/00002142A.pdf

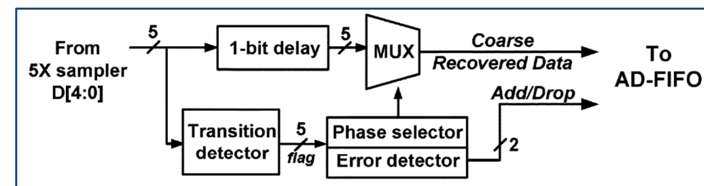
Data recovery using 5x blind oversampling, no clock recovery needed

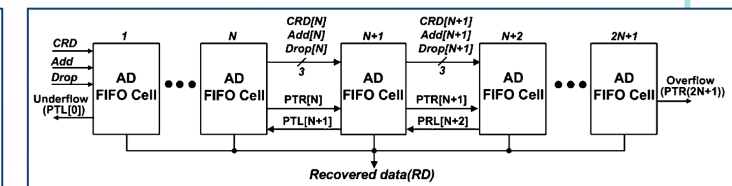
Must vote on sampled data and synchronize the incoming bitstream with onboard PLL

Verification has to cover cases where onboard PLL is faster and slower than incoming data

Article below does it in 180nm CMOS

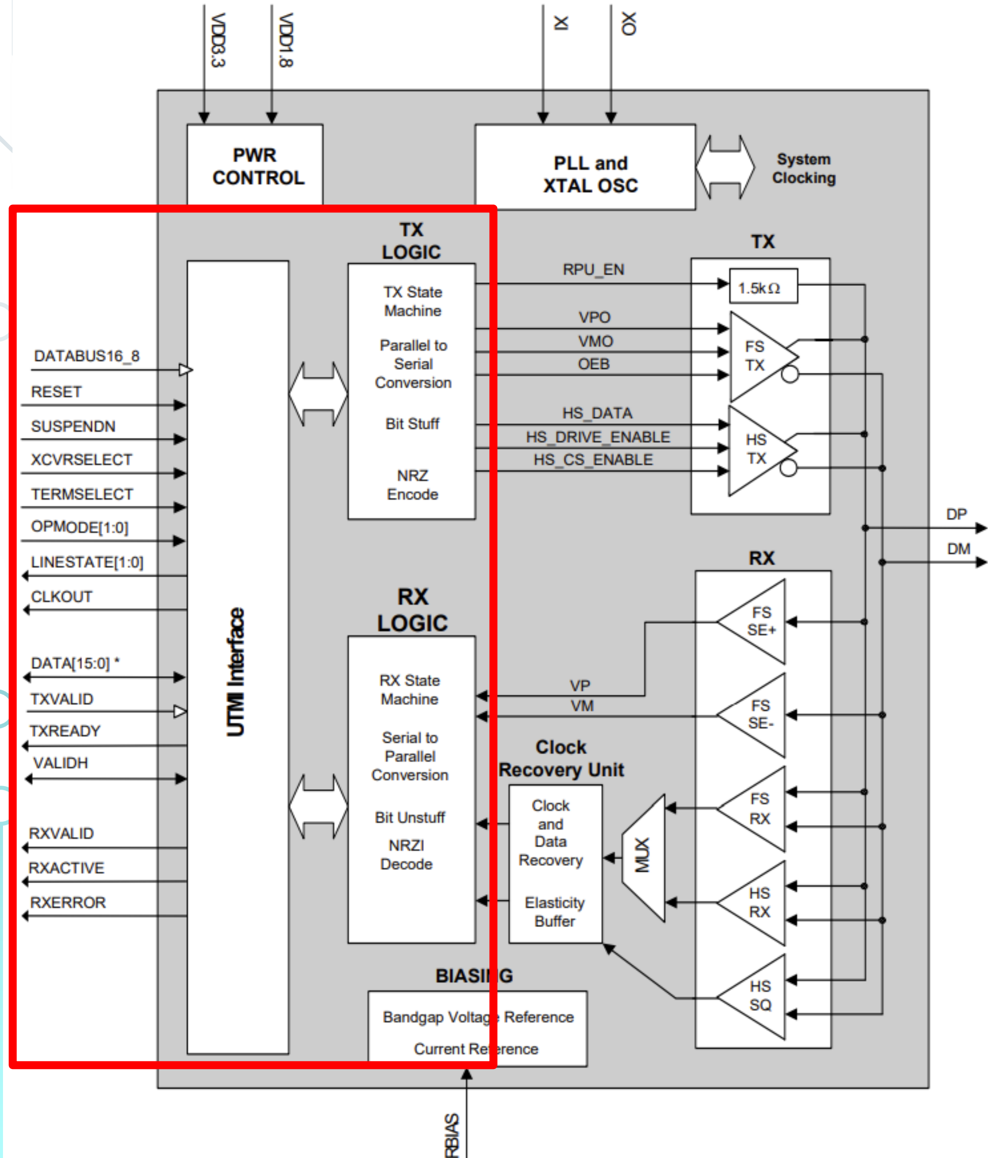






https://ieeexplore.ieee.org/abstract/document/4415527/

USB2 Digital Controller



- Various FSMs, encoders/decoders, and serializers/deserializers needed to interface between CPU and USB AMS circuits
- USB2 PHYs use a standard UTMI interface (USB 2.0 Transceiver Macrocell Interface) to CPU (high pin count, intended for ASIC implementation)
- ULMI interface (UTMI Low Pin Count Interface) reduces the pin count for external PHY chips

Example commercial USB2 PHY chip:
https://ww1.microchip.com/downloads/en/DeviceDoc/00002142A.pdf

Announcements

- Lab 4 this week
  - Lab 4b coming up
- Homework 2 will be posted this week
- No Lecture next Tuesday, February 19 (ISSCC)

Recap

- Lithography restricts layer orientation, length quantization
  - Favors layout regularity
  - Has implications on variability
- FinFETs add more restrictions (width quantization)

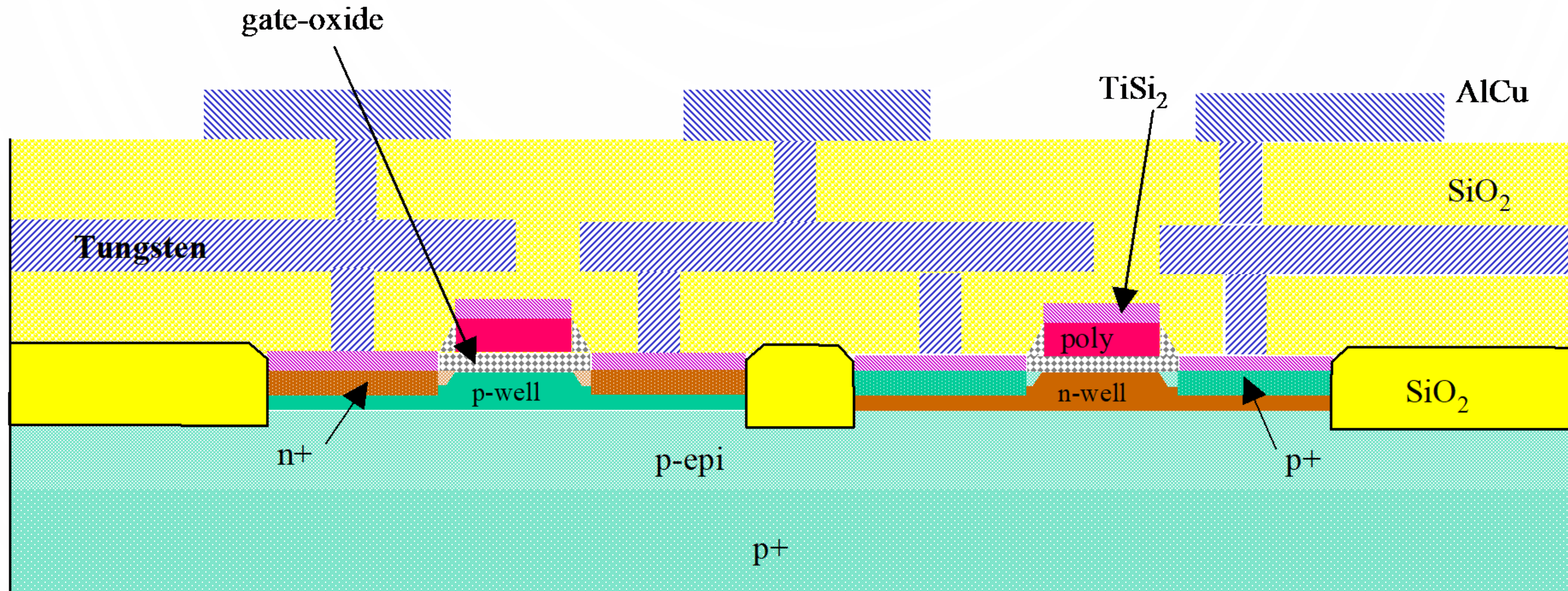
Modern Bulk/finFET/FDSOI processes

Some of the Process Features (Designer's Perspective)

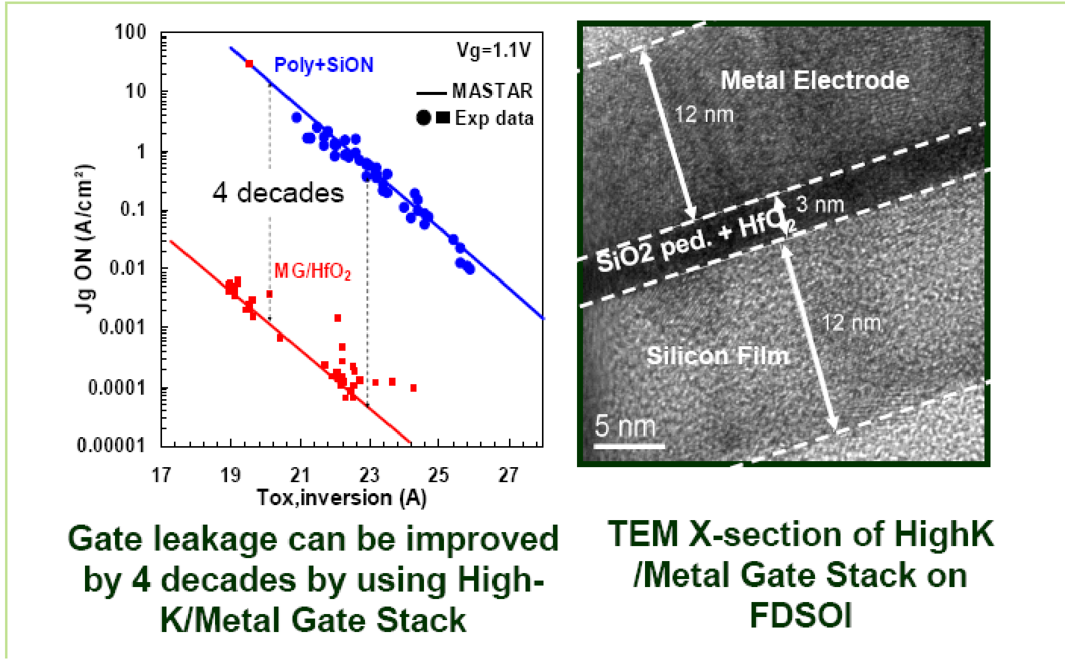
1. Shallow-trench isolation
2. High-k/Metal-gate technology
3. Strained silicon
4. Thin-body devices (28nm, and beyond)
5. Copper interconnects with low-k dielectrics

1. Shallow Trench Isolation

- Less space needed for isolation
- Some impact on stress (STI expansion can affect mobility)

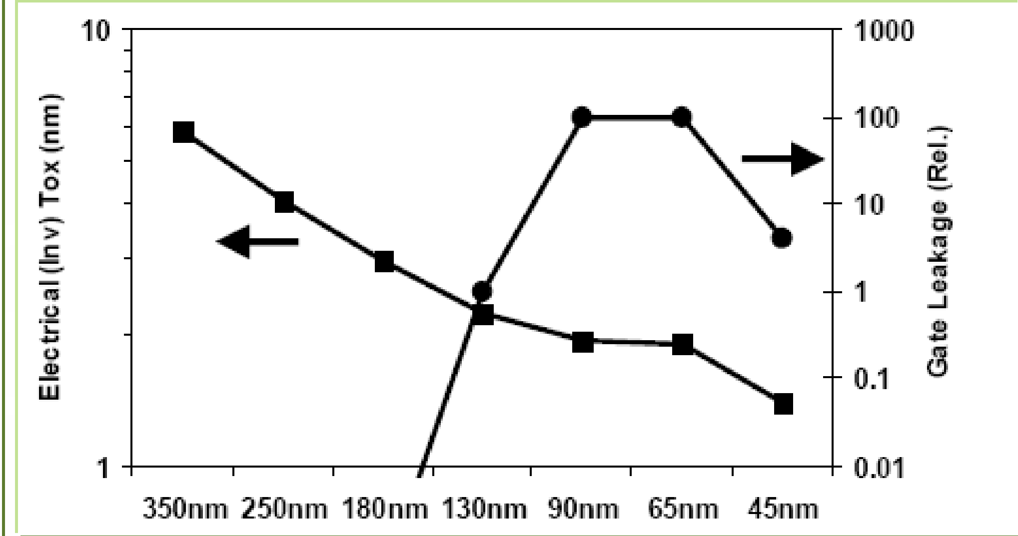


2. Hi-k/Metal gate



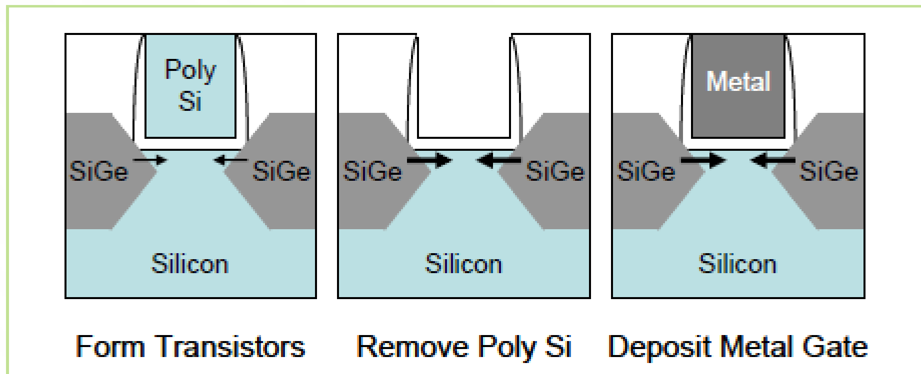
Gate leakage can be improved by 4 decades by using High-K/Metal Gate Stack

TEM X-section of HighK /Metal Gate Stack on FDSOI



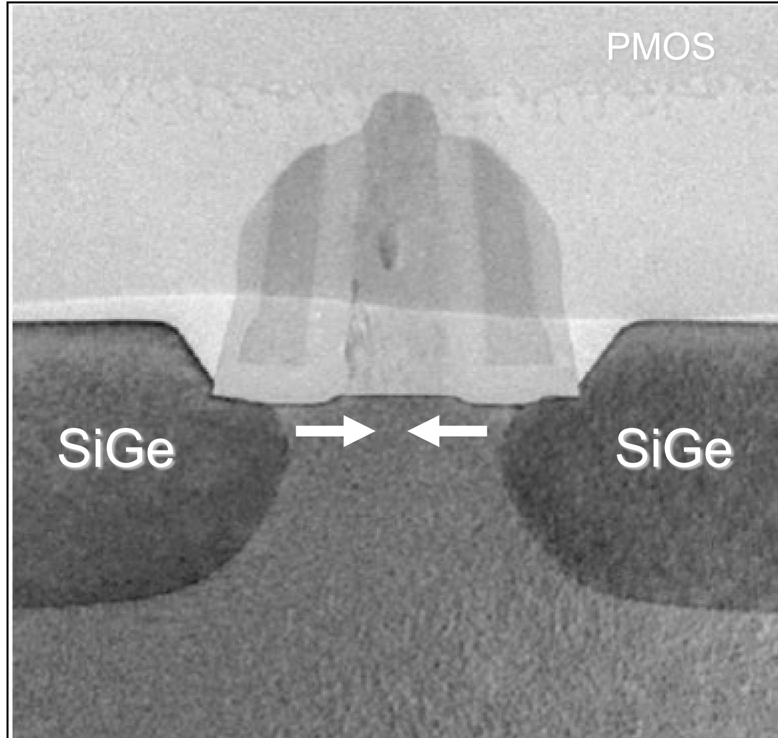
K. Mistry, IEDM'07

Replacement gate technology (Intel) – early version at 45nm



S. Natarajan, IEDM'08

3. Strained Silicon



Compressive channel strain
30% drive current increase
in 90nm CMOS

Tensile channel strain
10% drive current increase
in 90nm CMOS

Intel

Intel's Strained Si Numbers

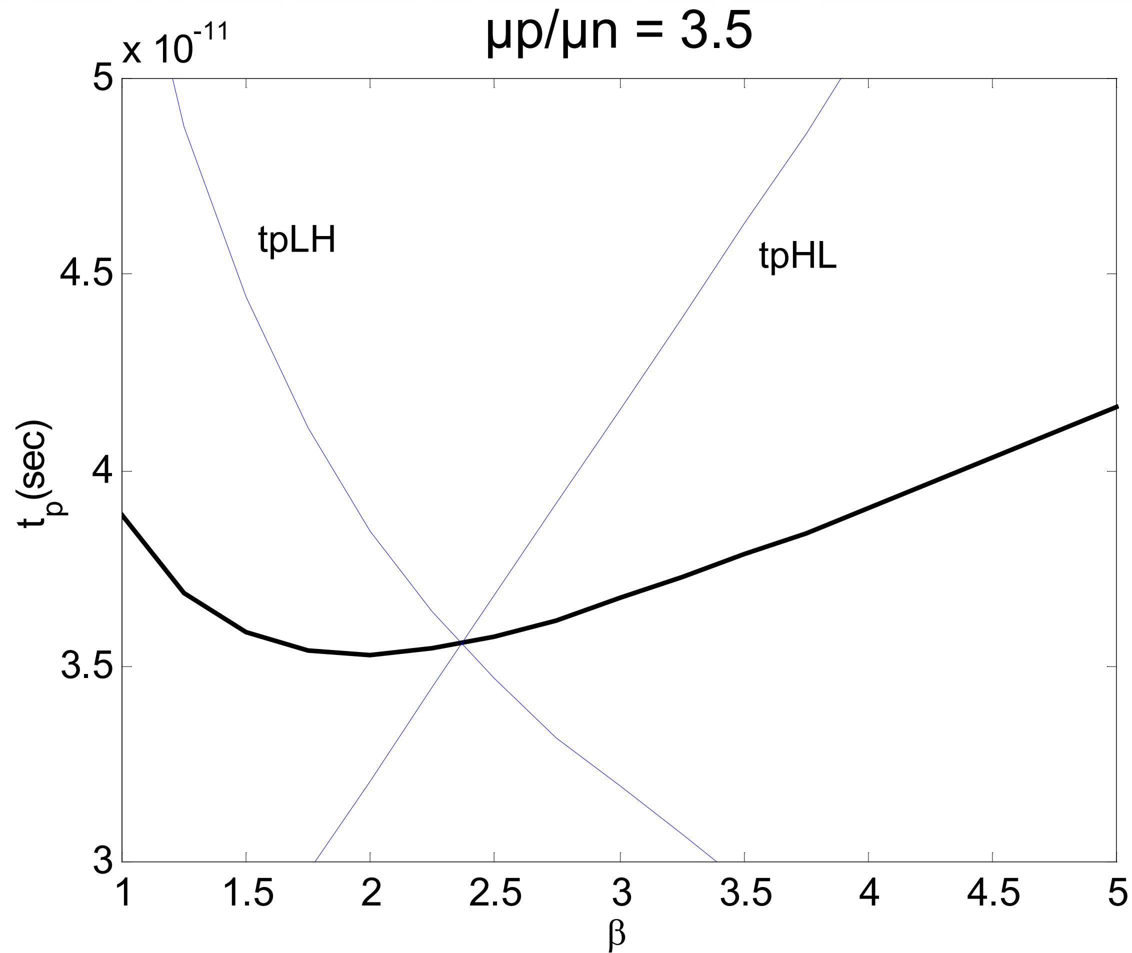
Performance gains:

	90 nm		65 nm	
	NMOS	PMOS	NMOS	PMOS
μ	20%	55%	35%	90%
IDSAT	10%	30%	18%	50%
IDLIN	10%	55%	18%	80%

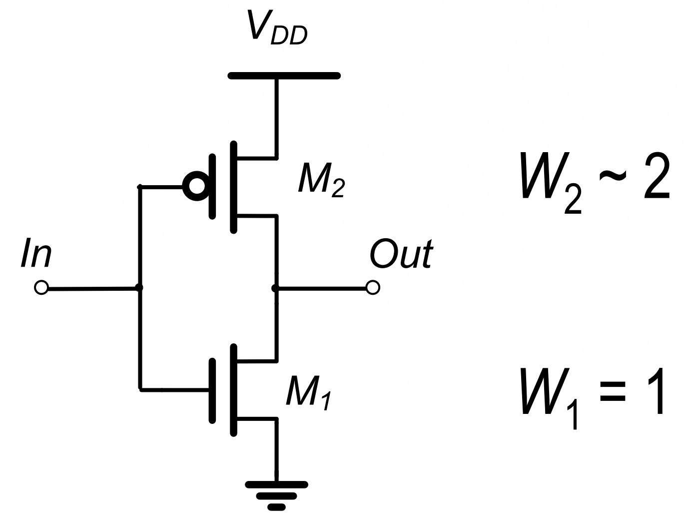
S. Thompson, VLSI'06 Tutorial

β-Ratio

- β=Wp/Wn



From Rabaey et al, DIC, 2nd ed, for 250nm CMOS



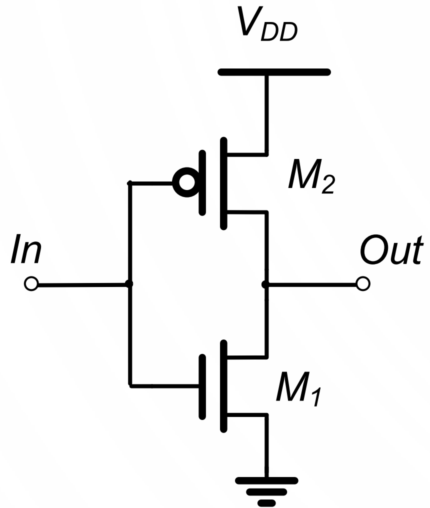
$W_2 \sim 2$

$W_1 = 1$

Strained Silicon: Implications on Sizing

- No strain

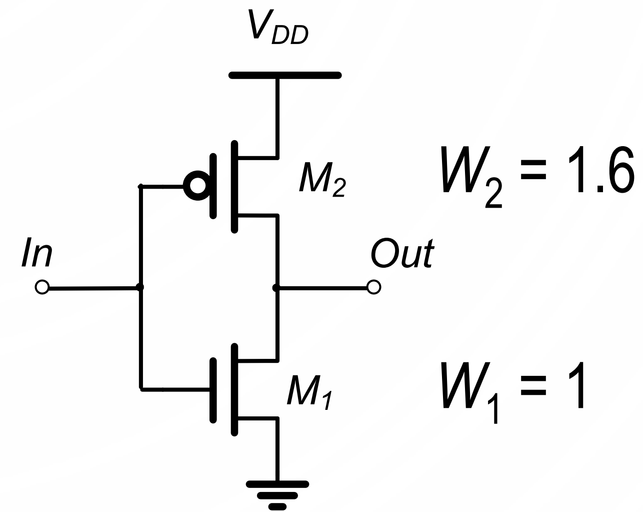
(e.g. 130nm)



$W_2 \sim 2$

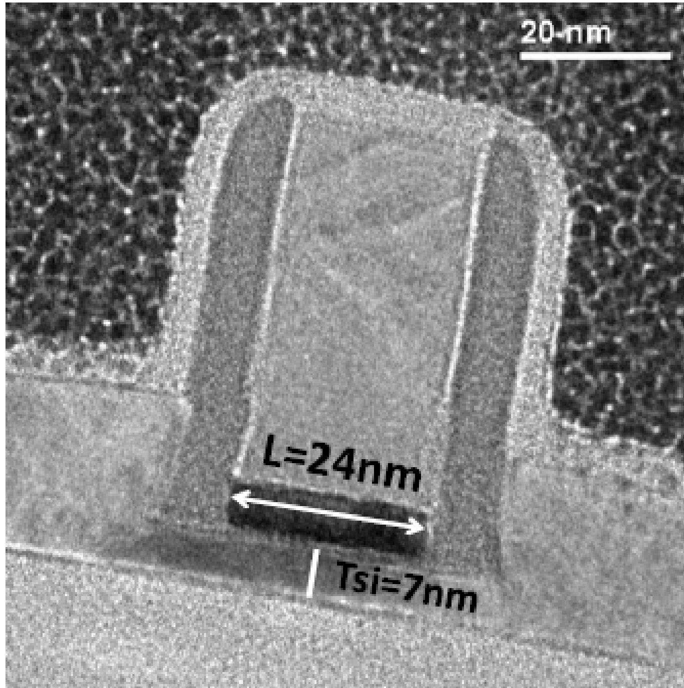
$W_1 = 1$

- Strained Si

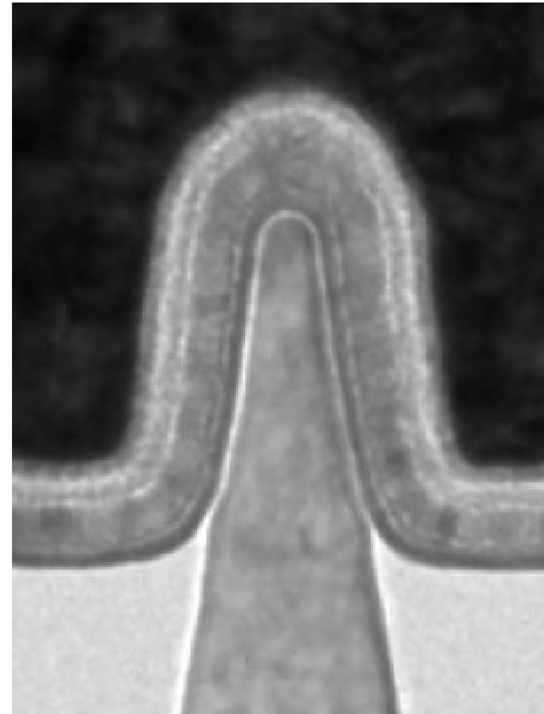


5. Thin-Body Devices

- 28nm FDSOI
- 22/14nm finFET



N. Planes, VLSI'2012

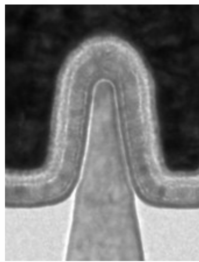


C. Auth, VLSI'2012

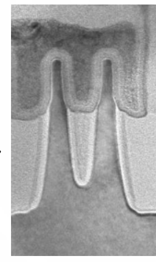
5. FinFETs

- FinFET scaling

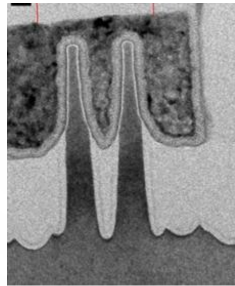
22/20nm **16/14nm** **10nm**



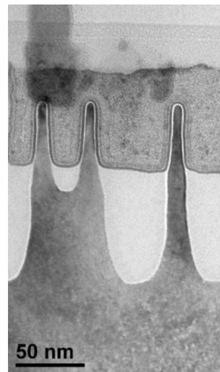
Intel, IEDM'12



Intel, VLSI'14

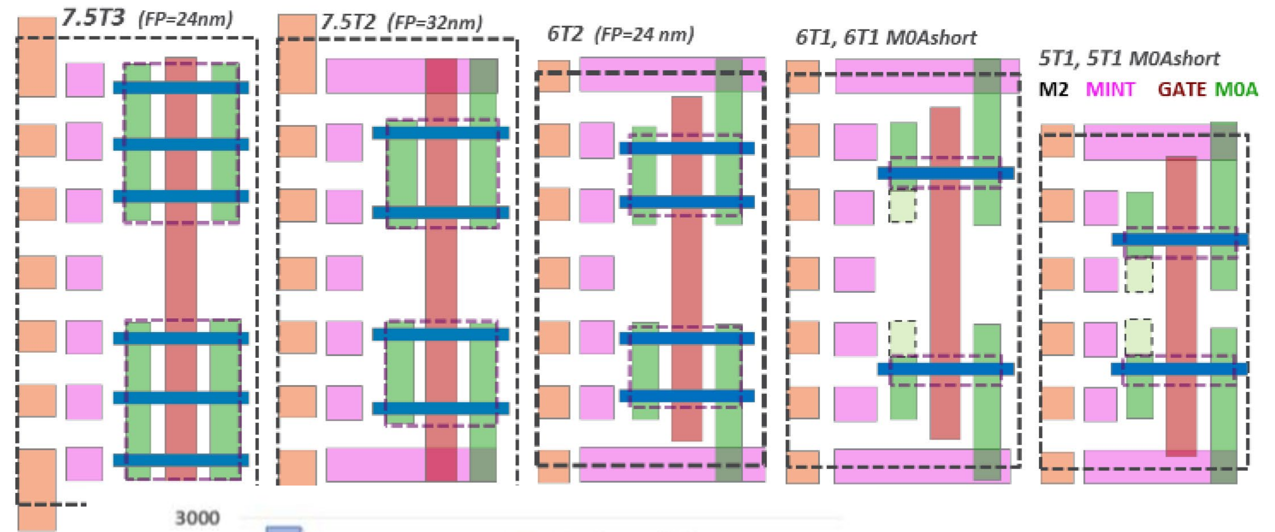


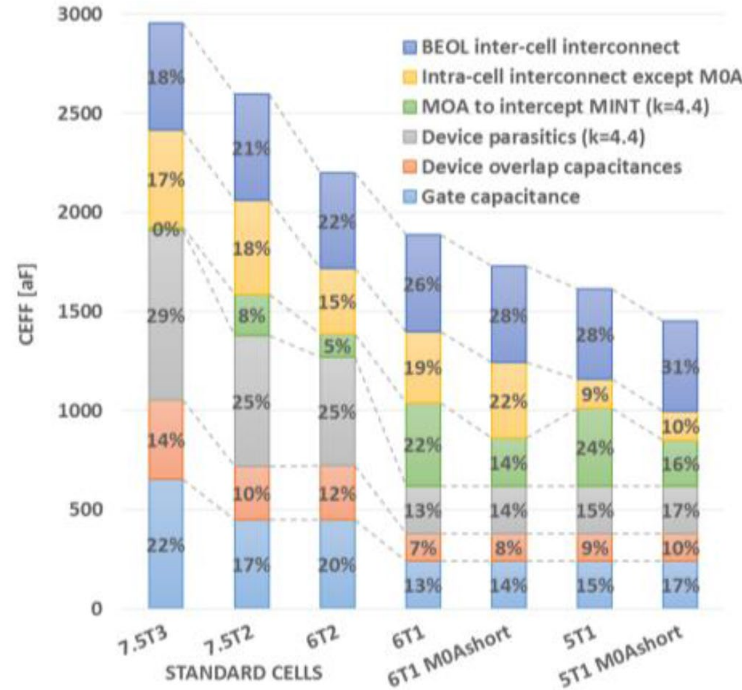
Intel, IEDM'17



- N-P spacing

- Track scaling (MP different than FP)



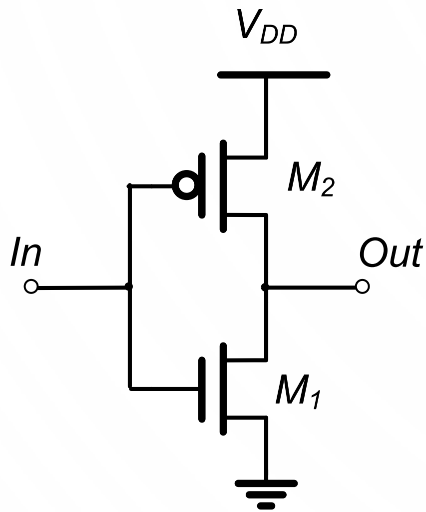


Garcia Bardon, IEDM'16

FinFETs and gate P/N sizing

- The use of strain closes the gap between N and P on currents to ~1:1

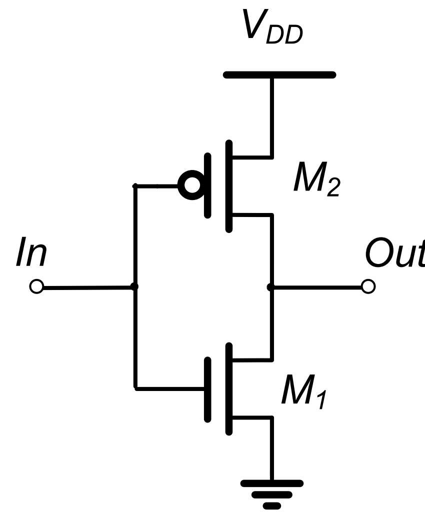
- No strain



$W_2 \sim 2$

$W_1 = 1$

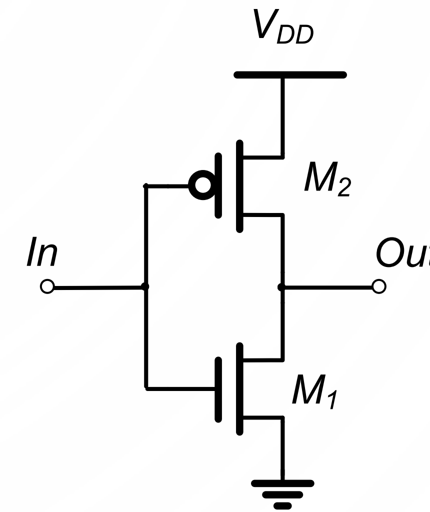
- Strained planar Si



$W_2 = 1.6$

$W_1 = 1$

- FinFET



$W_2 = 1$

$W_1 = 1$

5. FDSOI

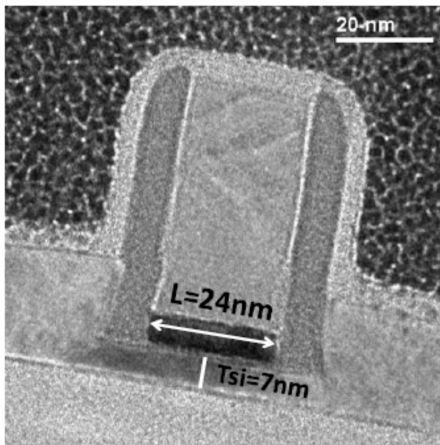
28FDSOI (STMicroelectronics)

28FD-SOI (Samsung)

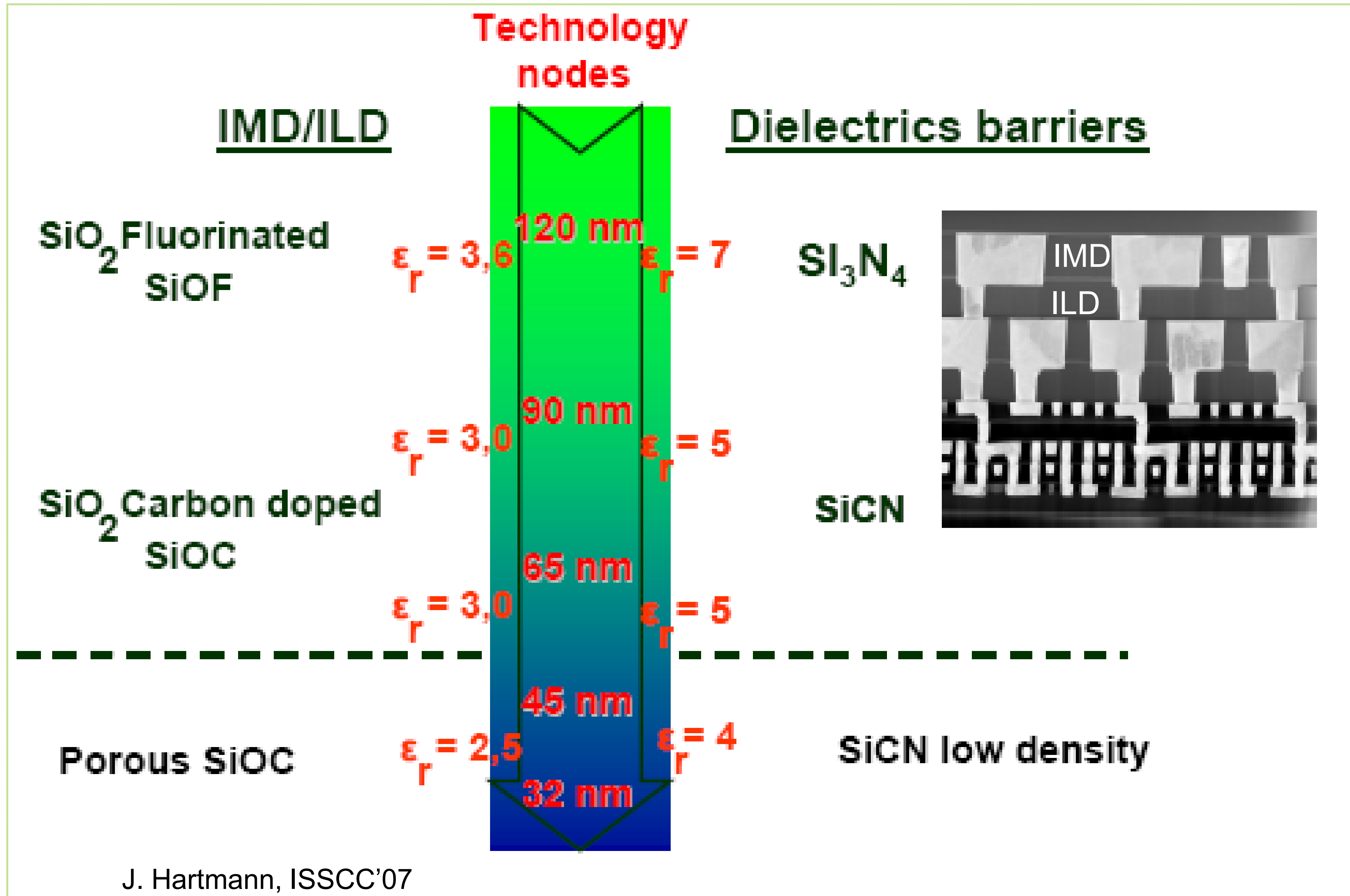
22FDX (GLOBALFOUNDRIES)

12FDX (GLOBALFOUNDRIES)

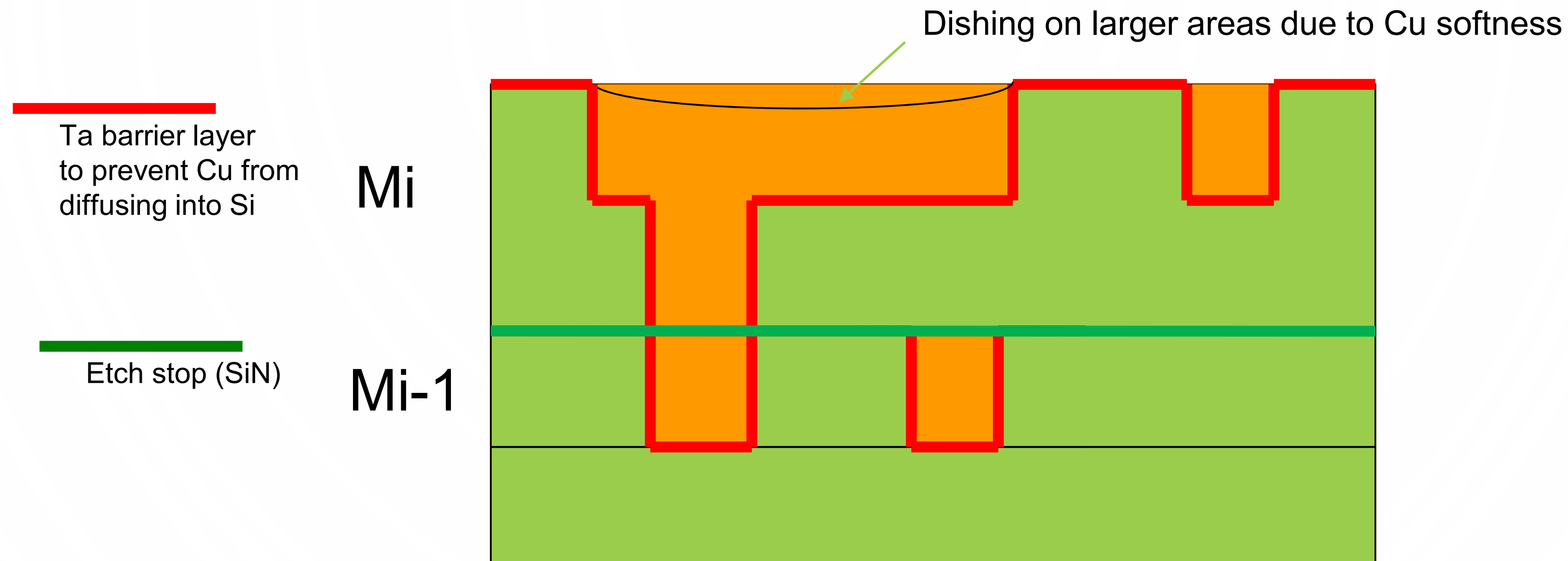
18FDS (Samsung)



5. Interconnect – low-K dielectrics

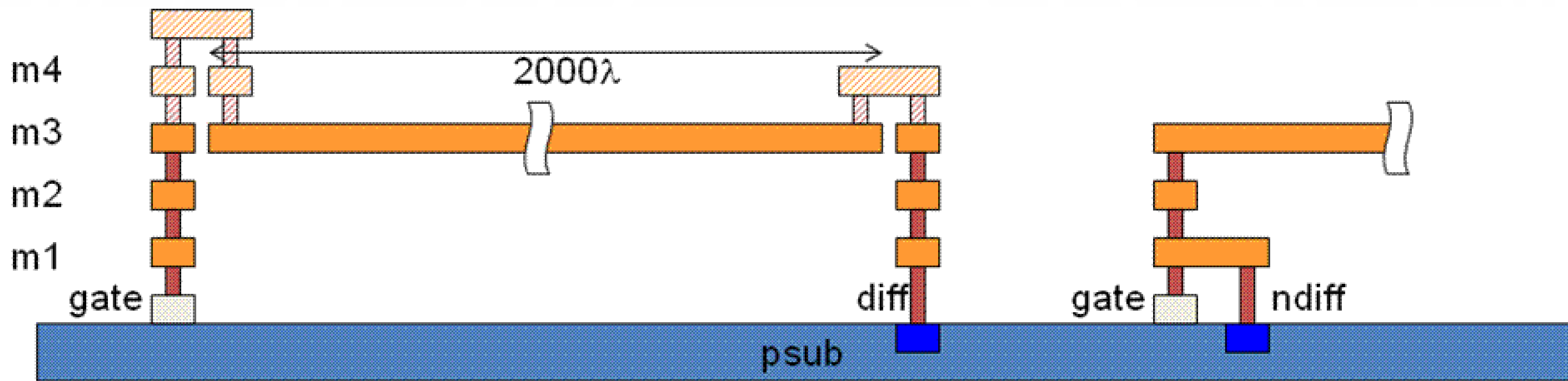


Interconnect: Chemical Mechanical Polishing (CMP)



- Metal density rules (20%-80%) (nowdays much tighter)
- Slotting rules
- Also: Antenna rules

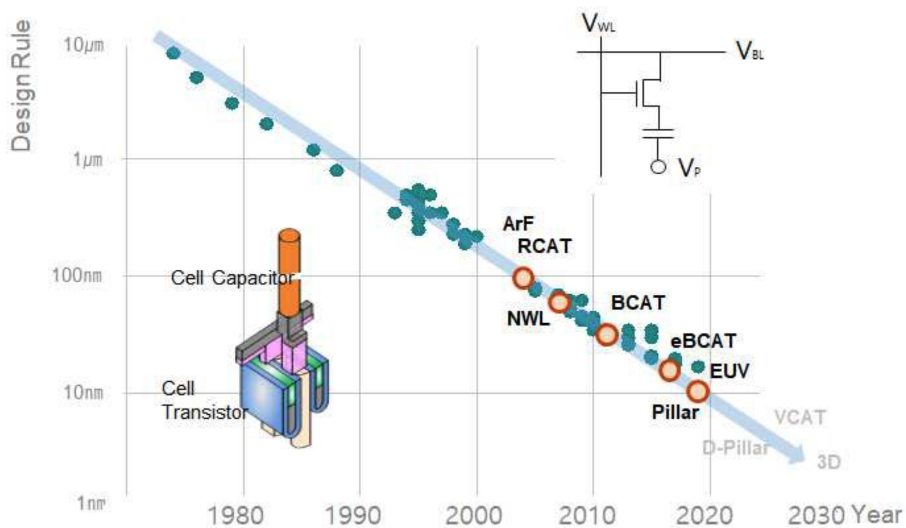
Interconnect: Antenna rules

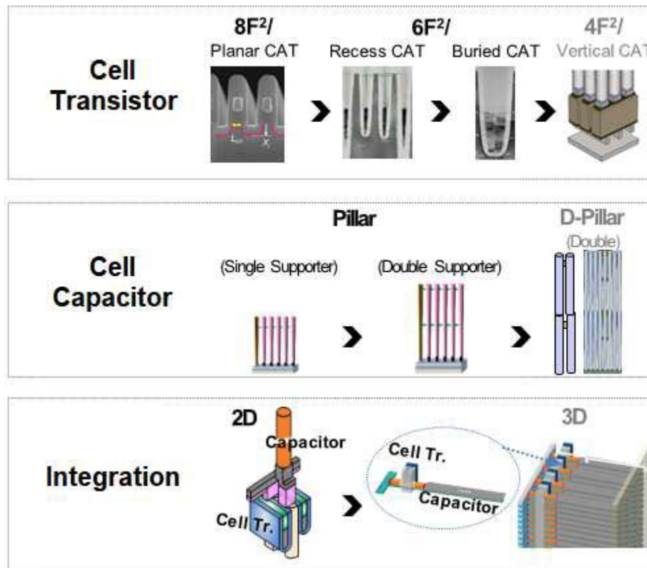


source: vlsi-expert.com

- Caused by charge accumulated on the metal wire during plasma etch
- Formulated as max wire area contacting the gate of certain area
- Design solutions
  - Jumper insertion – break signal wires and route to upper metal layers
  - Dummy transistors - addition of extra gates reduces the gate to wire cap ratio
  - Embedded protection diode (reverse bias)
  - Diode insertion after P&R

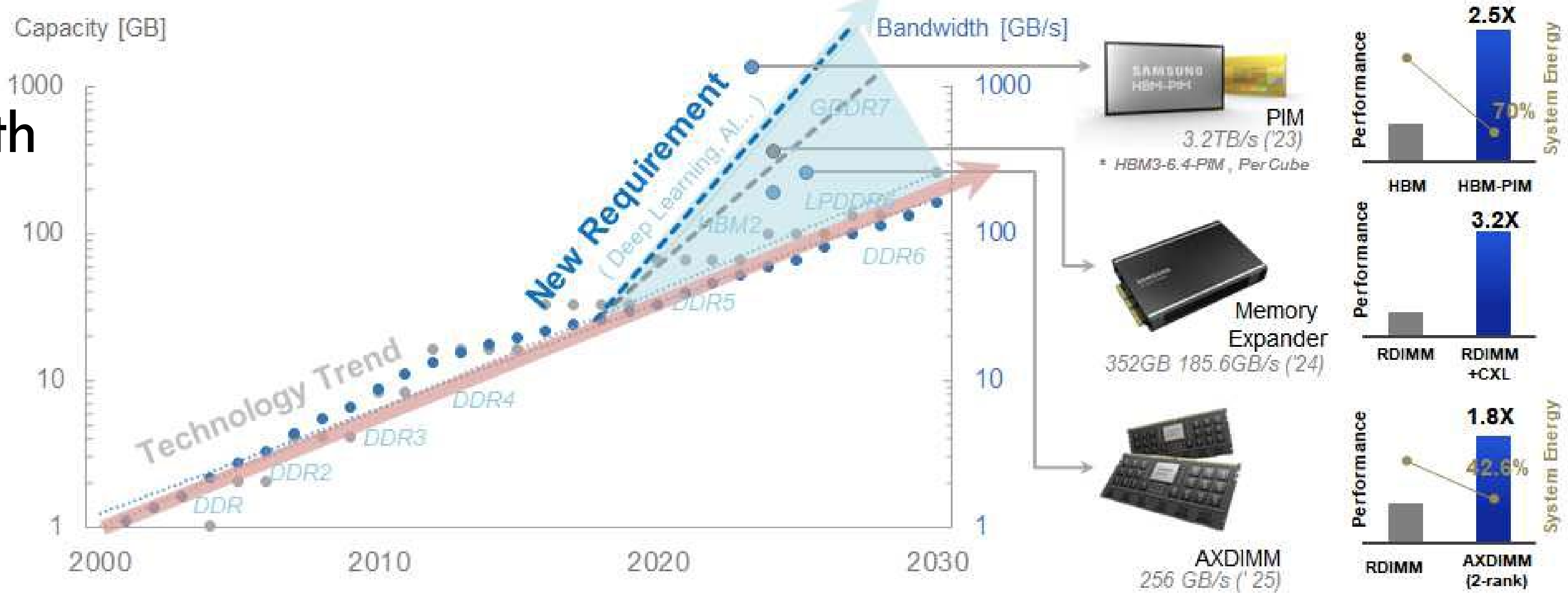
DRAM Scaling





- DRAM density scaling:
  - Transistor
  - Cap
  - Integration

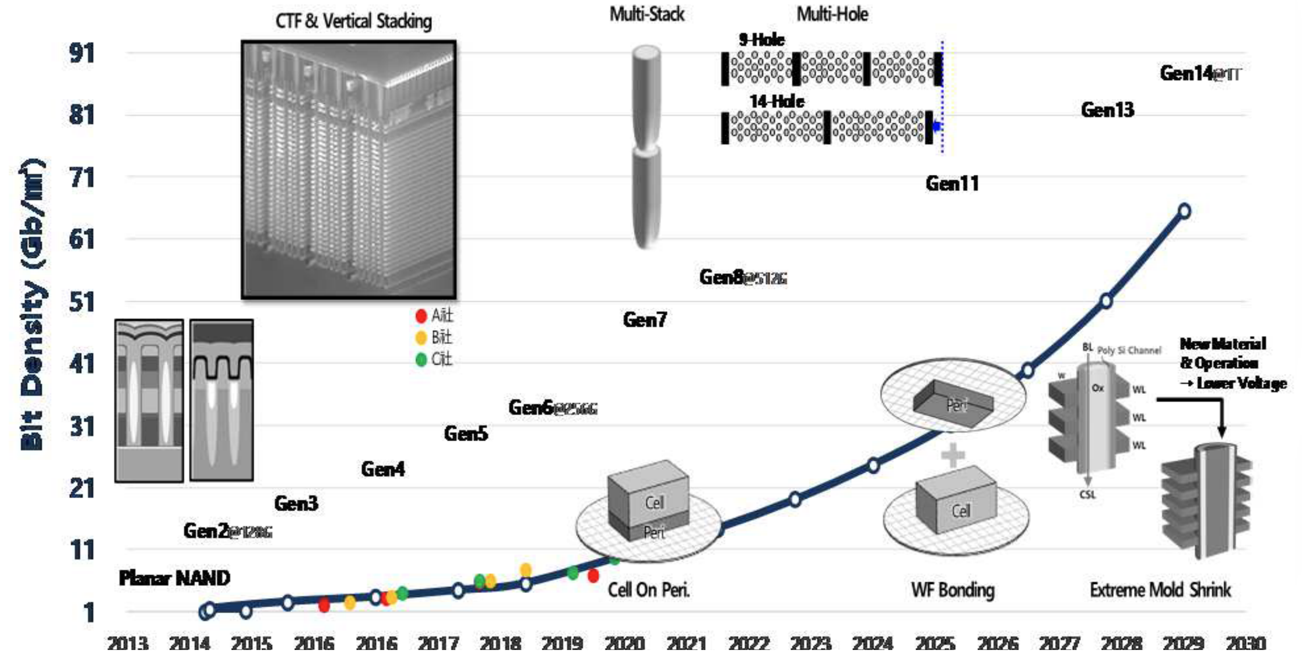
- DRAM capacity/bandwidth

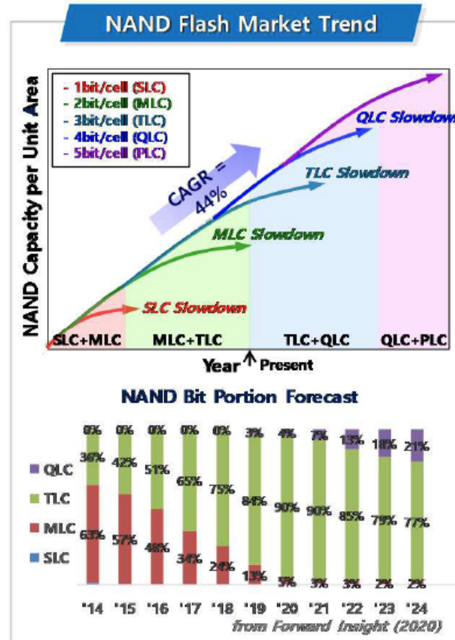


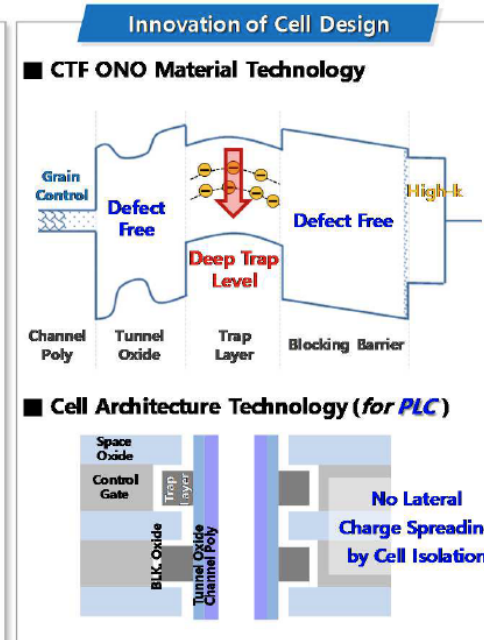
K.Kim, IEDM,'21

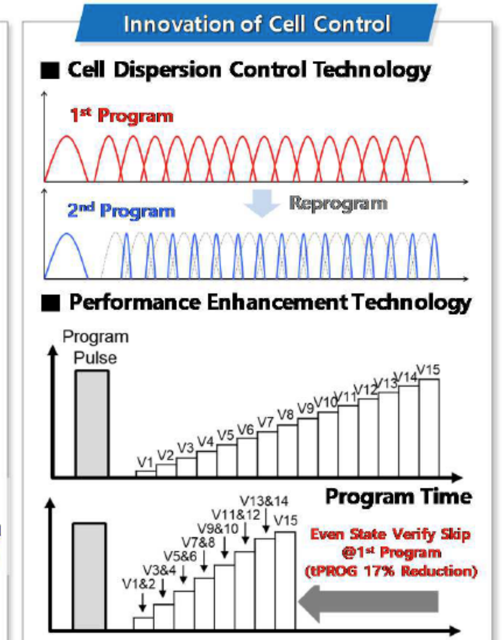
Flash Scaling

- Density and architecture scaling









K.Kim, IEDM,'21

MOS Transistor and Gate Delay Models

Modeling Goals

- Models that traverse design hierarchy
- Start with transistor models
- Gate delay models
- Use models to time the design
- Modeling variability

- Based on 251A, approach
  - Start simple
  - Increase accuracy, when needed

Device Models

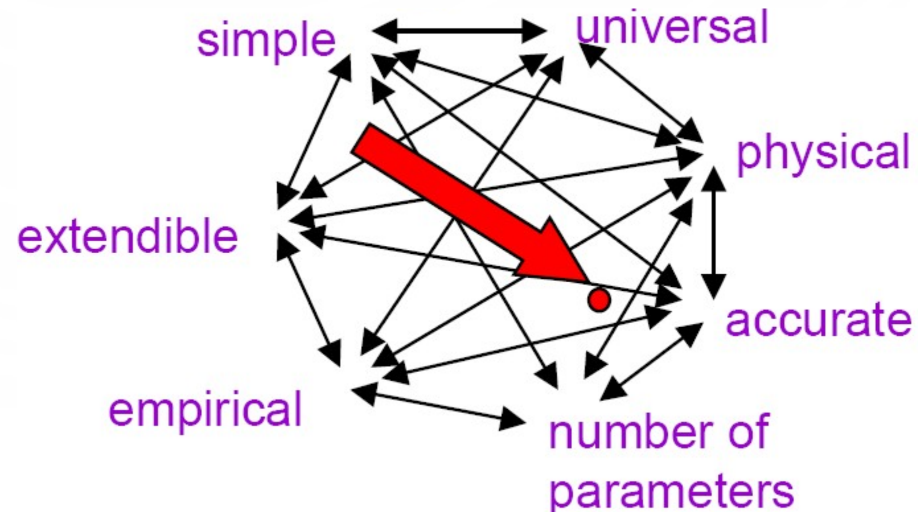
- Transistor models
  - I-V characteristics
  - C-V characteristics
- Interconnect models
  - R, C, L
  - Covered in EE240A

Transistor Modeling

- Different levels:
  - Hand analysis
  - Computer-aided analysis (e.g. Matlab, Python, Excel,…)
  - Switch-level simulation (some flavors of 'fast Spice')
  - Circuit simulation (Hspice)
- These levels have different requirements in complexity, accuracy and speed of computation
- We are primarily interested in delay and energy modeling, rather than current modeling
- But we have to start from the currents…

Transistor Modeling

- DC
  - Accurate I-V equations
  - Well behaved conductance for convergence (not necessarily accurate)
- Transient
  - Accurate I-V and Q-V equations
  - Accurate first derivatives for convergence
  - Conductance, as in DC
- Physical vs. empirical



from BSIM group

Transistor I-V Modeling

- BSIM
  - Superthreshold and subthreshold models
  - Need smoothening between two regions
- EKV/PSP
  - One continuous model based on channel surface potential

Goal for Next Week

- Develop velocity-saturated model for I_{on} and apply it to sizing and delay calculation
  - Similar approach as in 251A, just use an analytical model

Next Lecture

- Transistor models

CC BY NC SA