

EECS251B

Advanced Digital Circuits and Systems

Lecture 9 – Modern Technology and Chiplets

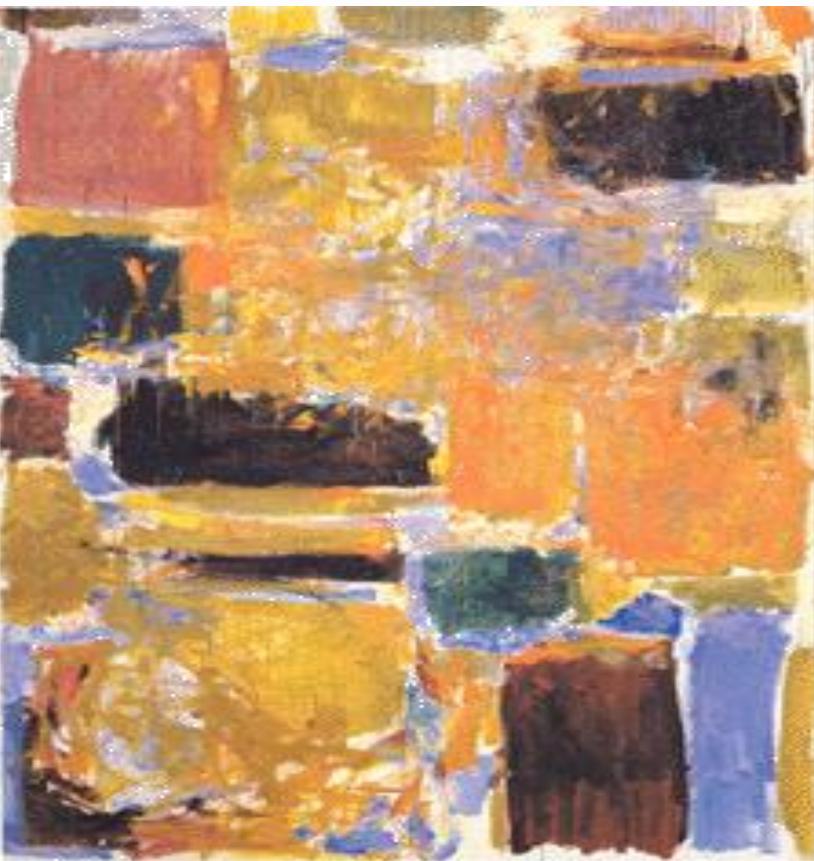
Vladimir Stojanović

Tuesdays and Thursdays 9:30-11am

Cory 521

Recap

- Technology affects circuit design
 - Optimized for standard cell, SRAM density
 - Recent scaling not uniform per layer
- Lithography restricts layer orientation, length quantization
 - Favors layout regularity
 - Has implications on variability
- FinFETs add more restrictions (width quantization)



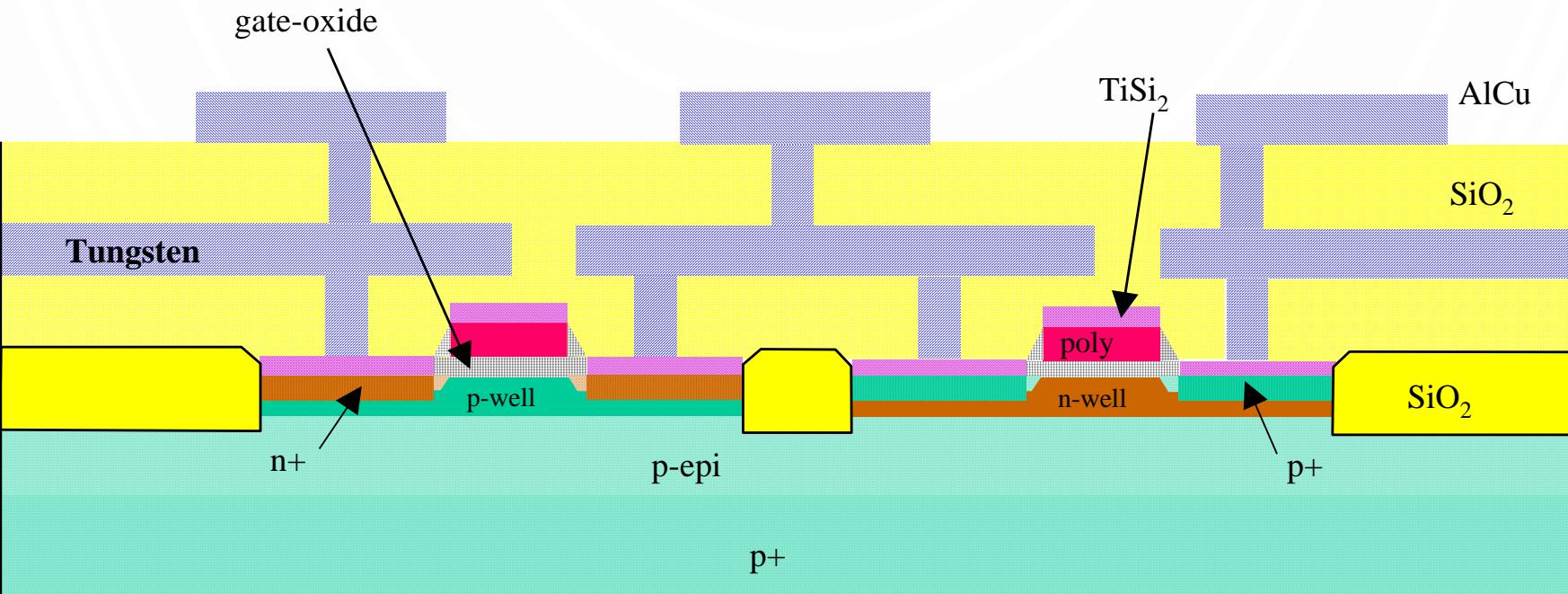
Modern Bulk/finFET/FDSOI processes

Some of the Process Features (Designer's Perspective)

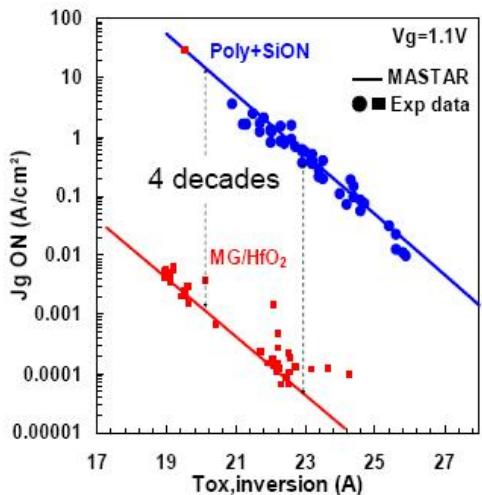
1. Shallow-trench isolation
2. High-k/Metal-gate technology
3. Strained silicon
4. Thin-body devices (28nm, and beyond)
5. Copper interconnects with low-k dielectrics

1. Shallow Trench Isolation

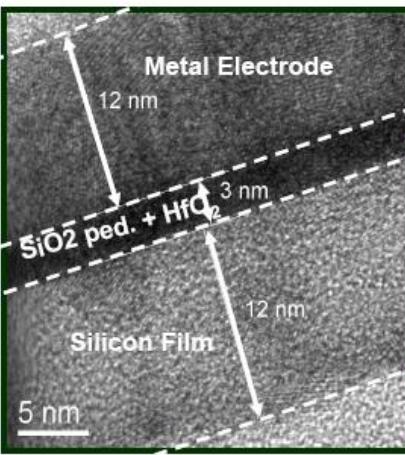
- Less space needed for isolation
- Some impact on stress (STI expansion can affect mobility)



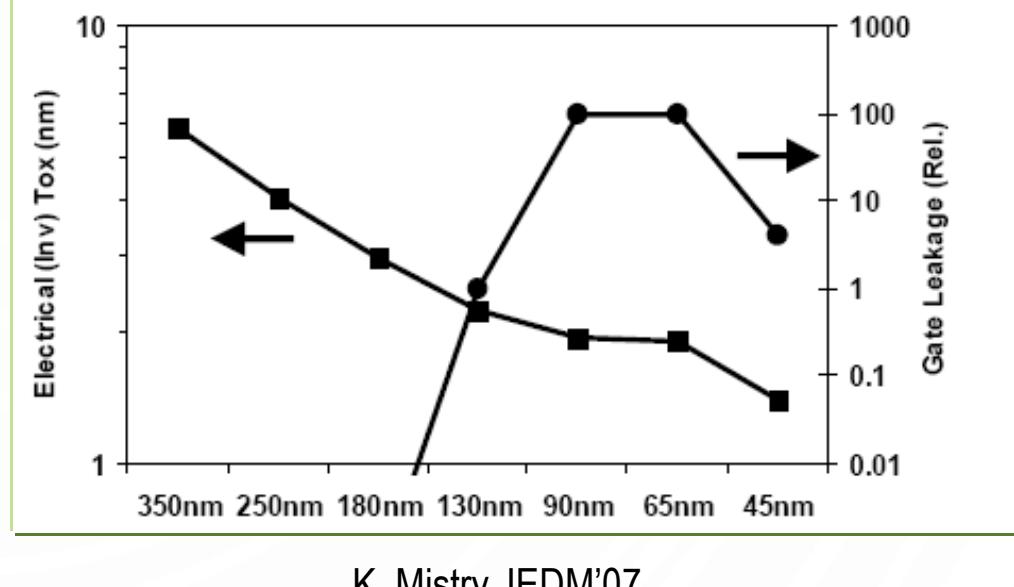
2. Hi-k/Metal gate



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

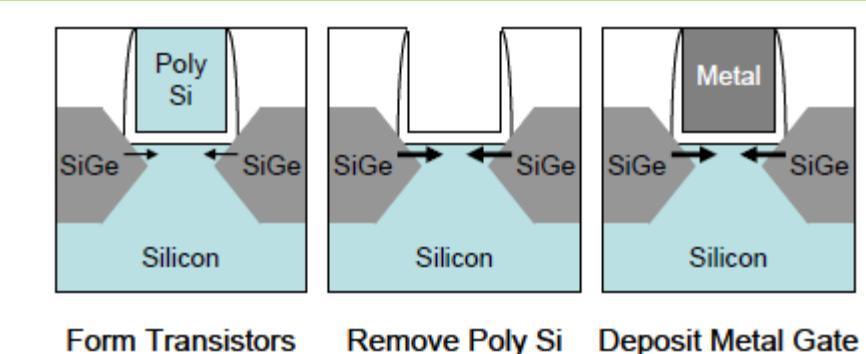


TEM X-section of HighK /Metal Gate Stack on FDSOI



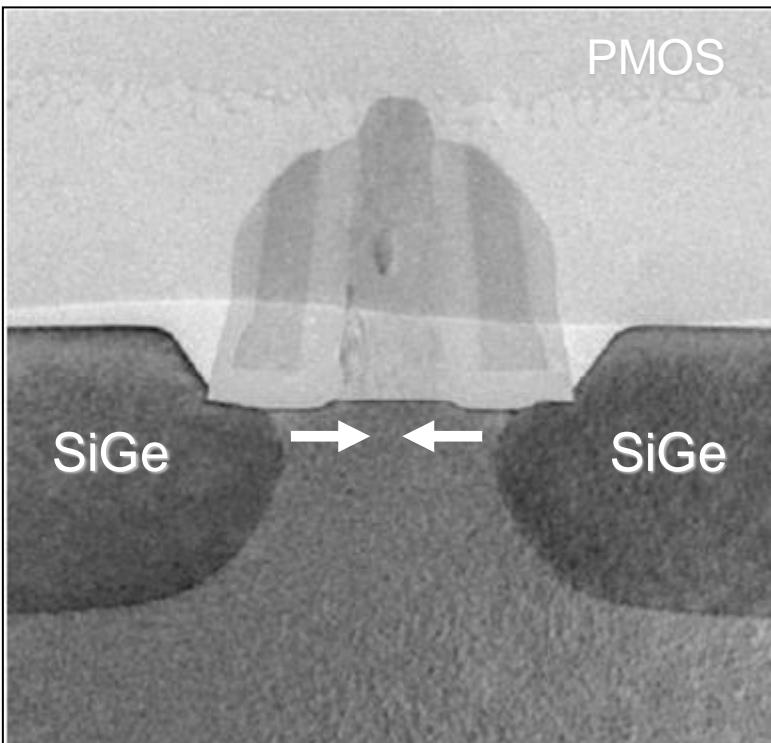
K. Mistry, IEDM'07

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

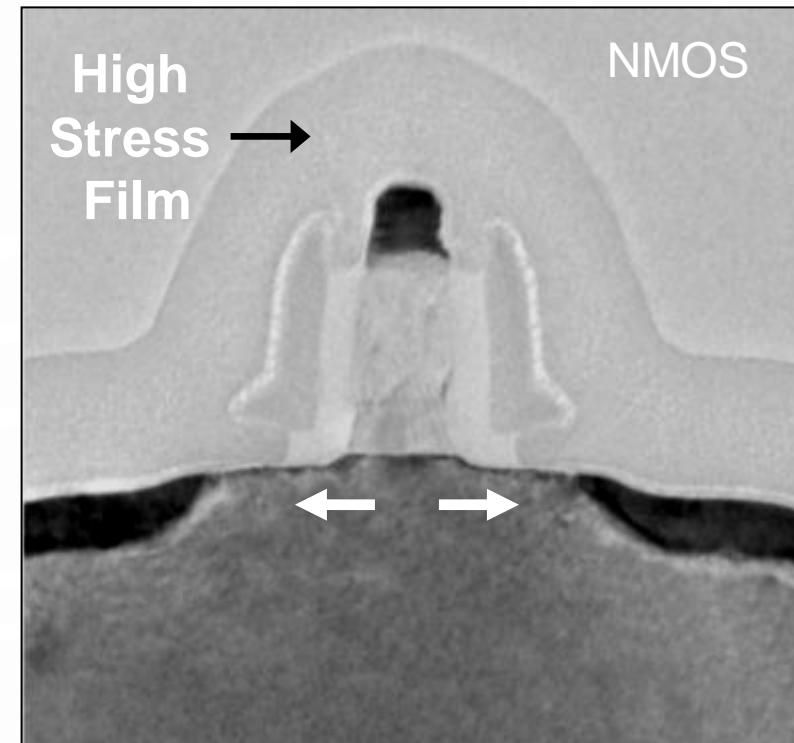


S. Natarajan, IEDM'08

3. Strained Silicon



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



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

Intel's Strained Si Numbers

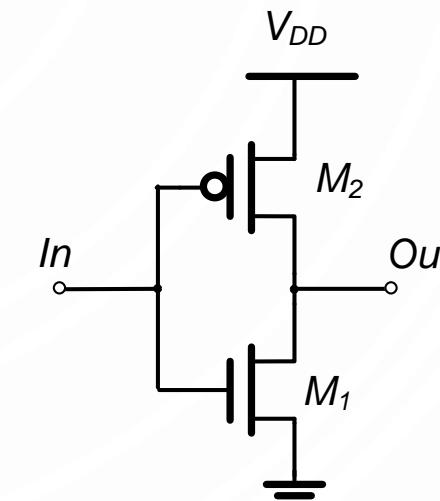
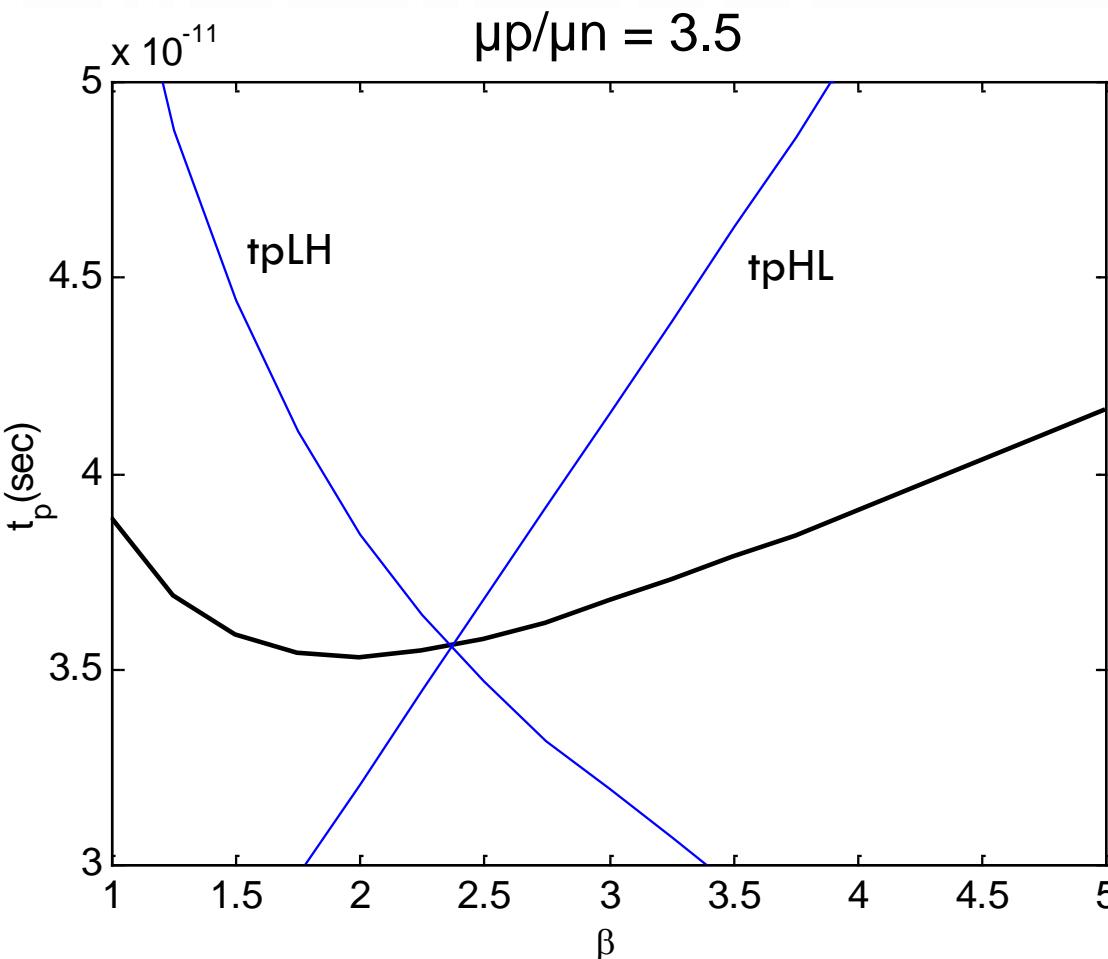
Performance gains:

	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

- $\beta = W_p / W_n$

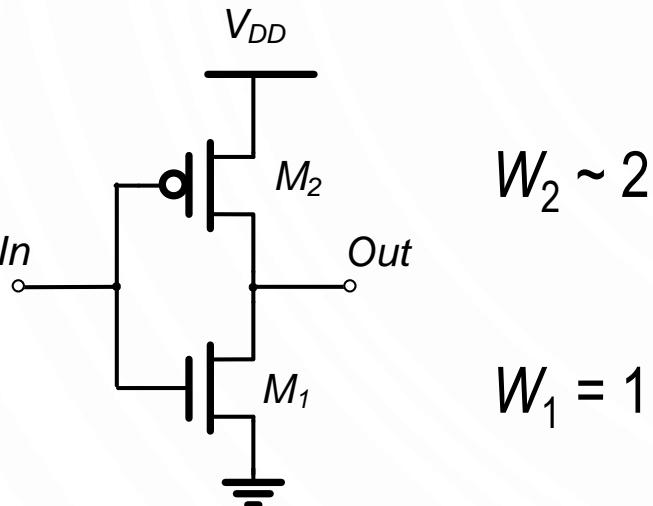


$W_2 \sim 2$

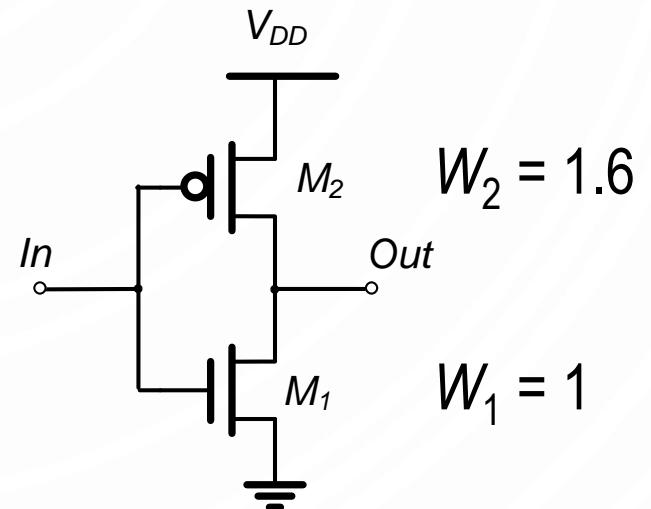
$W_1 = 1$

Strained Silicon: Implications on Sizing

- No strain

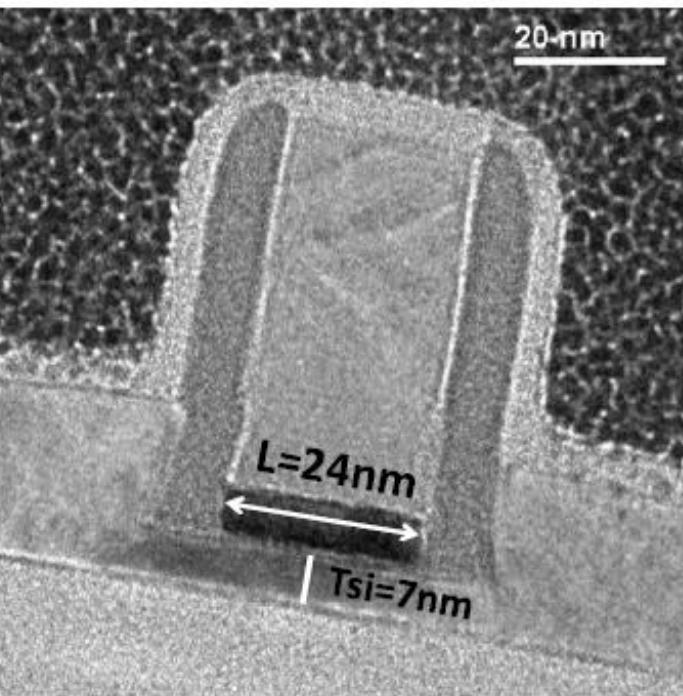


- Strained Si



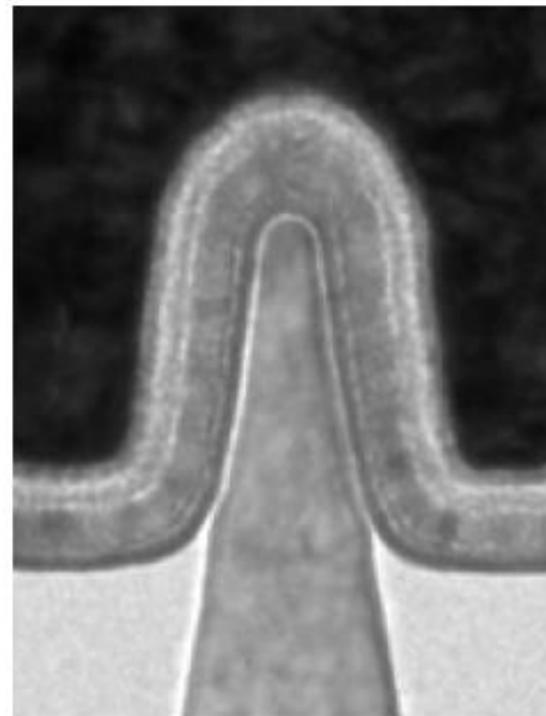
5. Thin-Body Devices

- 28nm FDSOI



N. Planes, VLSI'2012

- 22/14nm finFET

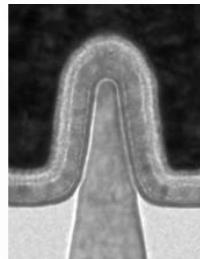


C. Auth, VLSI'2012

5. FinFETs

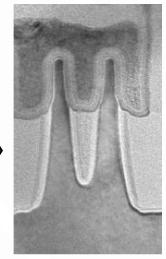
- FinFET scaling

22/20nm



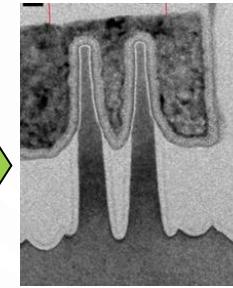
Intel,
IEDM'12

16/14nm

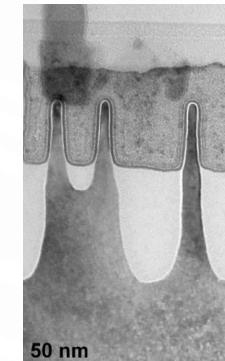


Intel,
VLSI'14

10nm

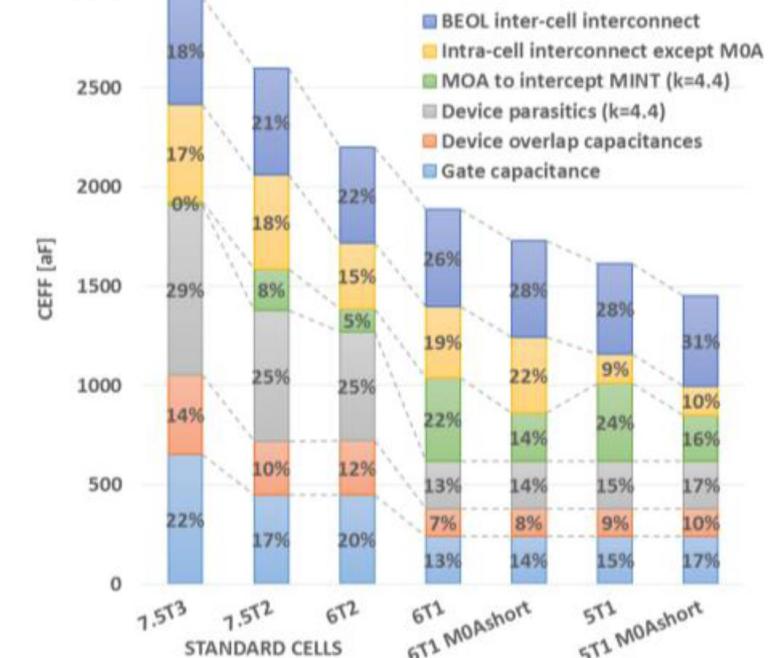
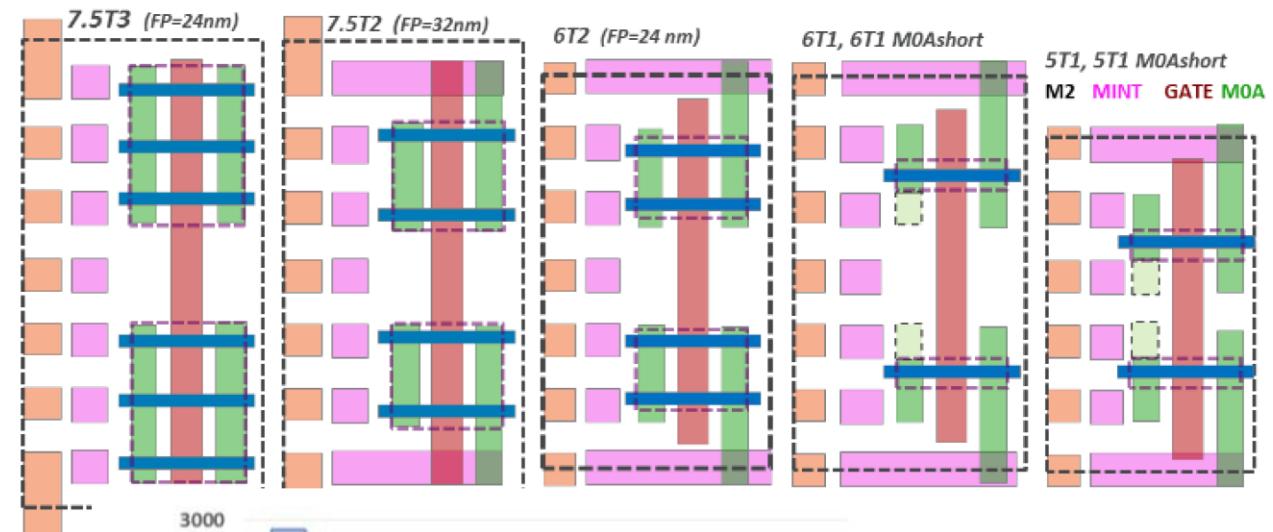


Intel,
IEDM'17



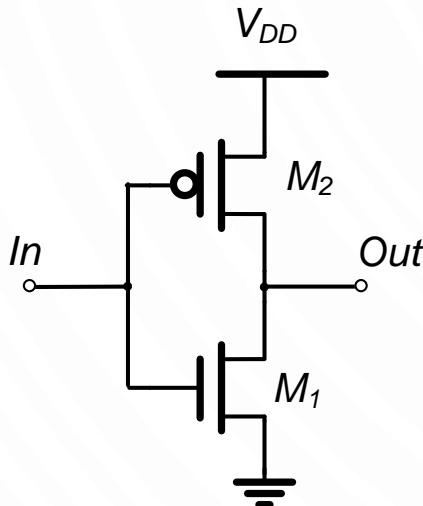
- N-P spacing

- Track scaling (MP different than FP)

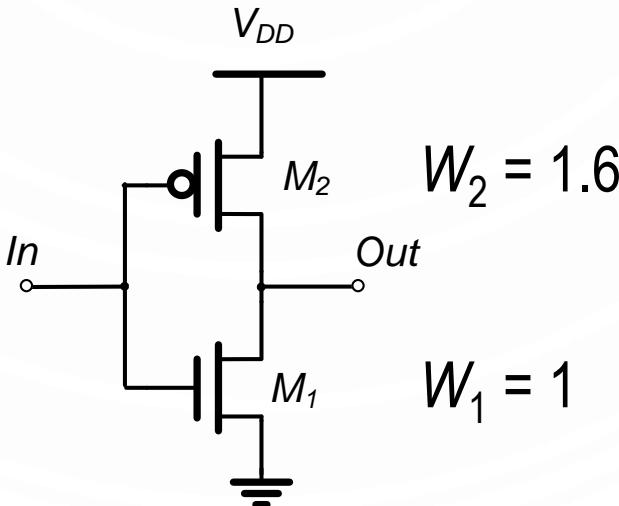


FinFETs and gate P/N sizing

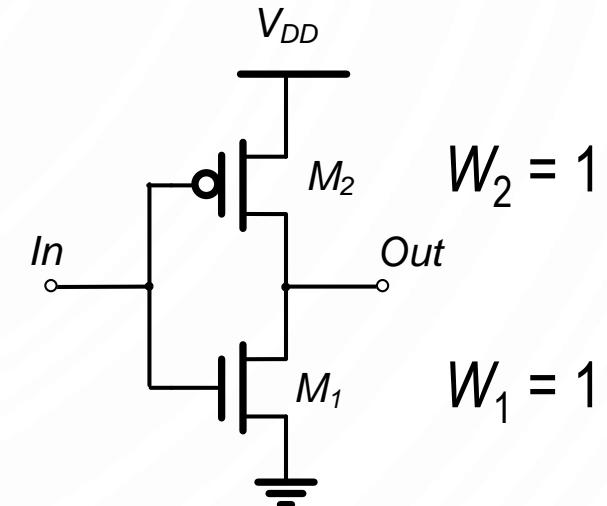
- The use of strain closes the gap between N and P on currents to $\sim 1:1$
- No strain
- Strained planar Si
- FinFET



$$W_2 \sim 2 \quad W_1 = 1$$



$$W_2 = 1.6 \quad W_1 = 1$$



5. FDSOI



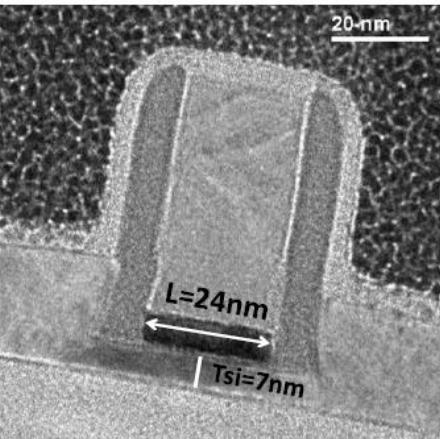
28FDSOI (STMicroelectronics)

28FD-SOI (Samsung)

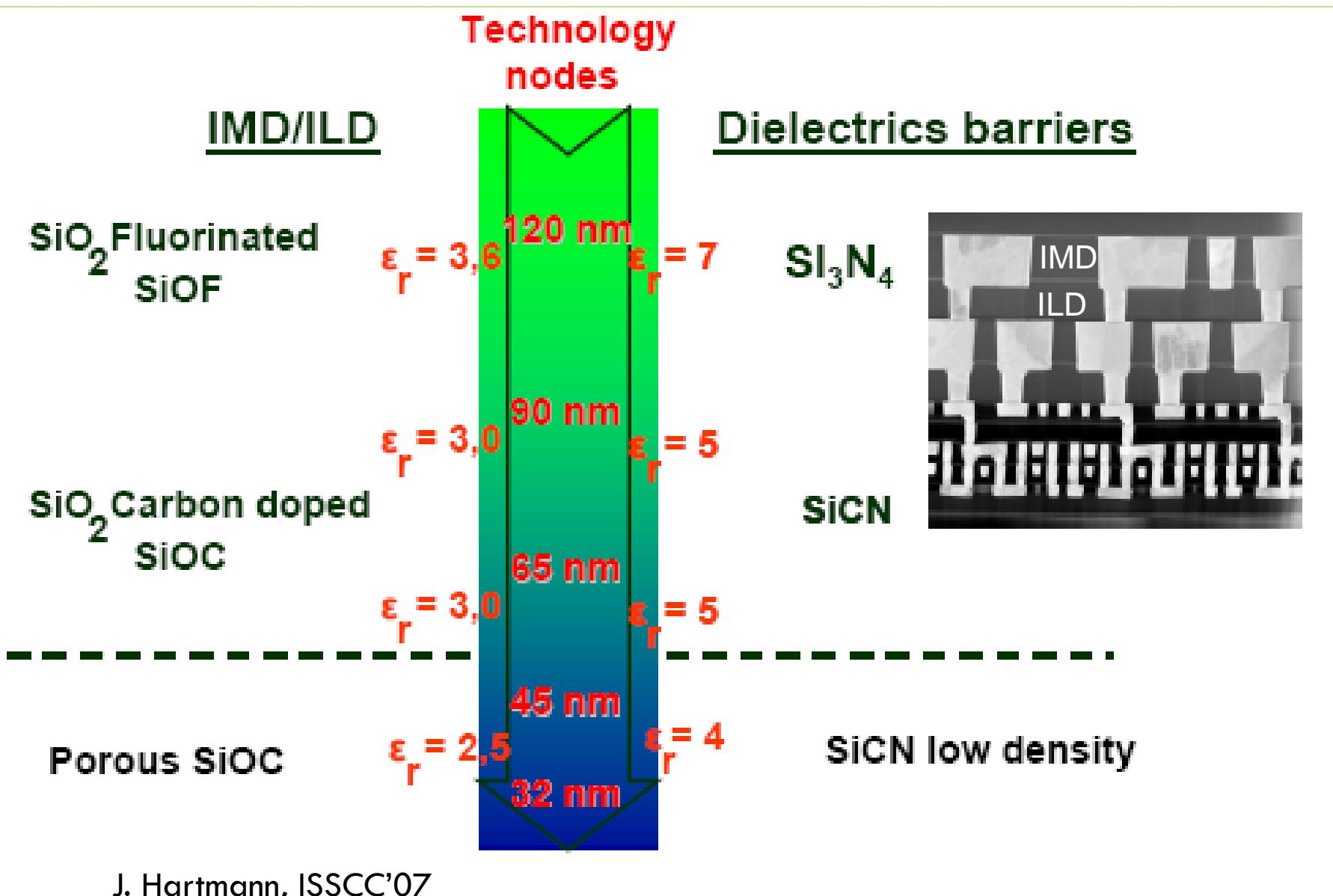
22FDX (GLOBALFOUNDRIES)

12FDX (GLOBALFOUNDRIES)

18FDS (Samsung)

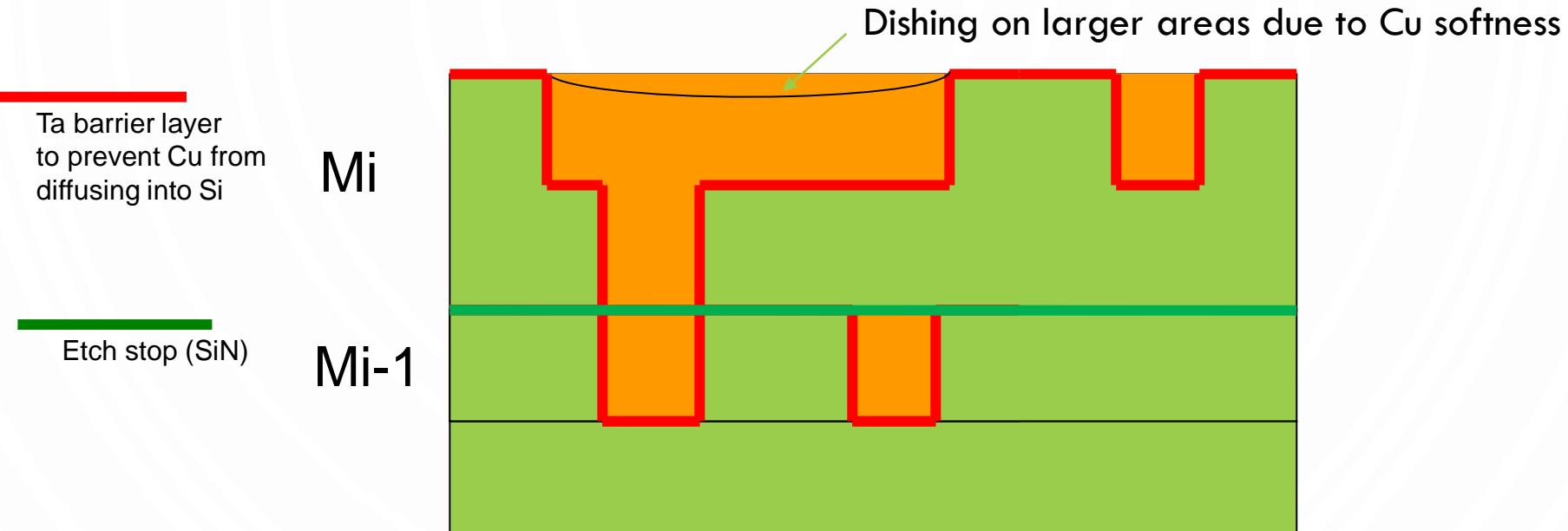


5. Interconnect – low-K dielectrics



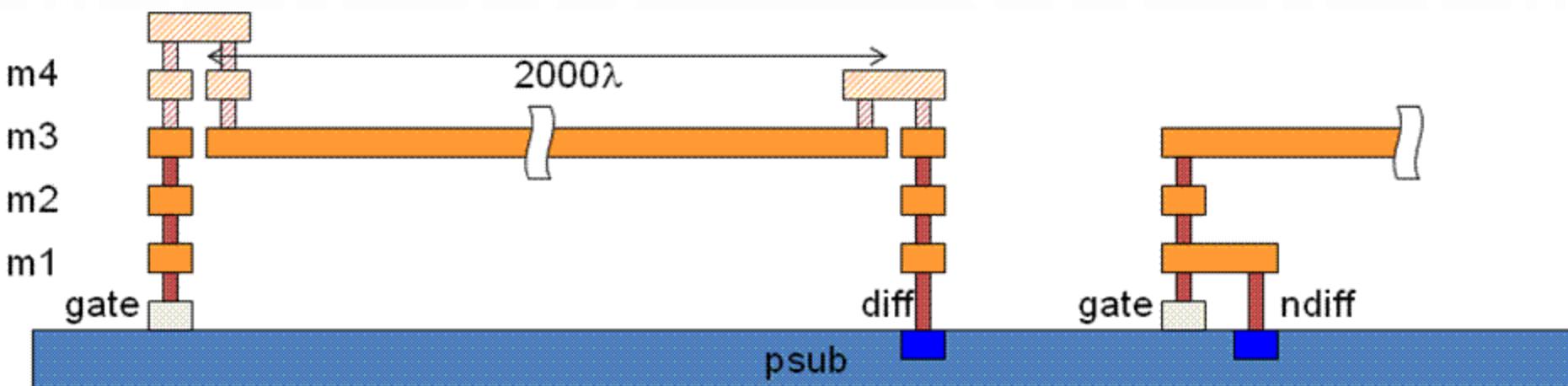
Interconnect: Chemical Mechanical Polishing (CMP)

Cu interconnect: Dual damascene process



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

Interconnect: Antenna rules



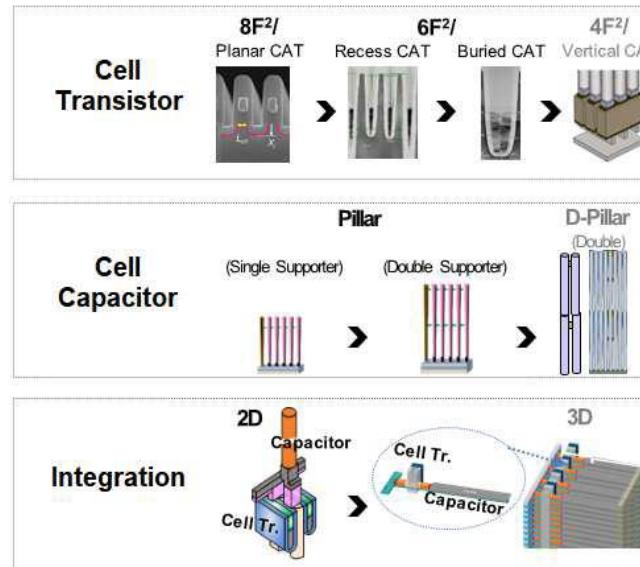
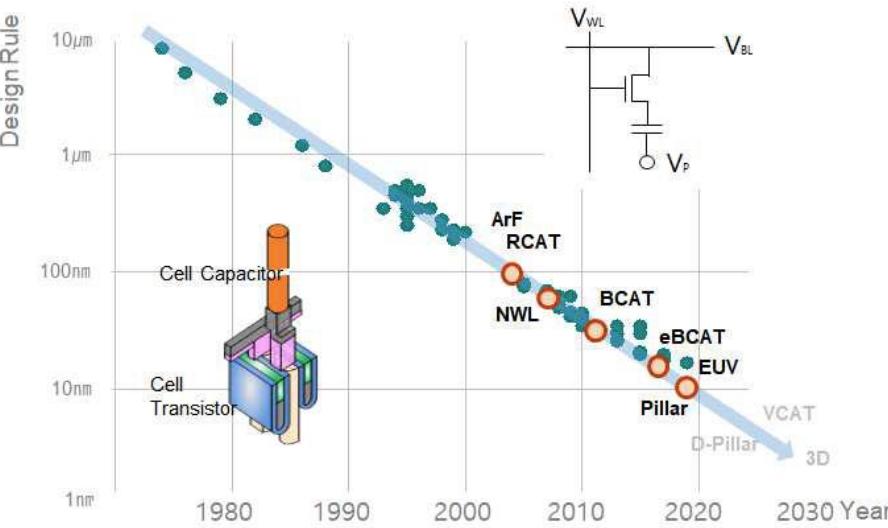
Bridging keeps gate away
from long metals until they
drain through the diffusion

Node diodes are inactive during
chip operation (reverse-biased p/n);
let charge leak away harmlessly

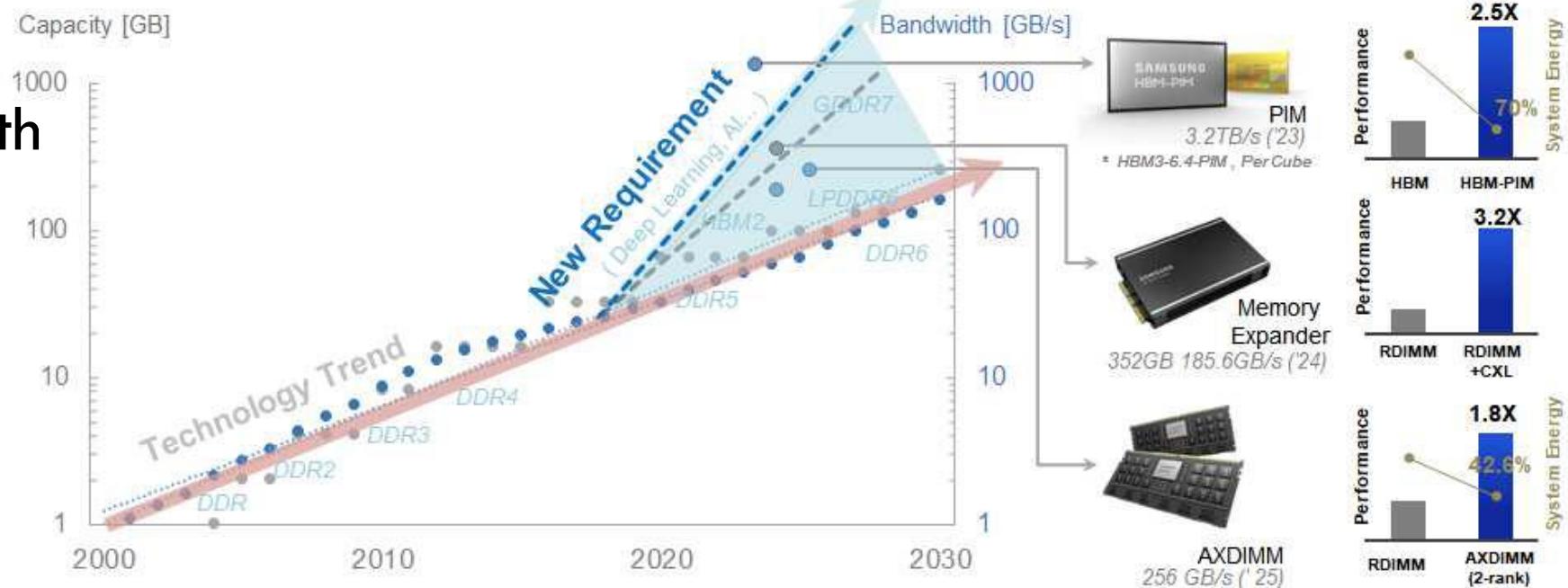
source: vlsi-expert.com

- 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



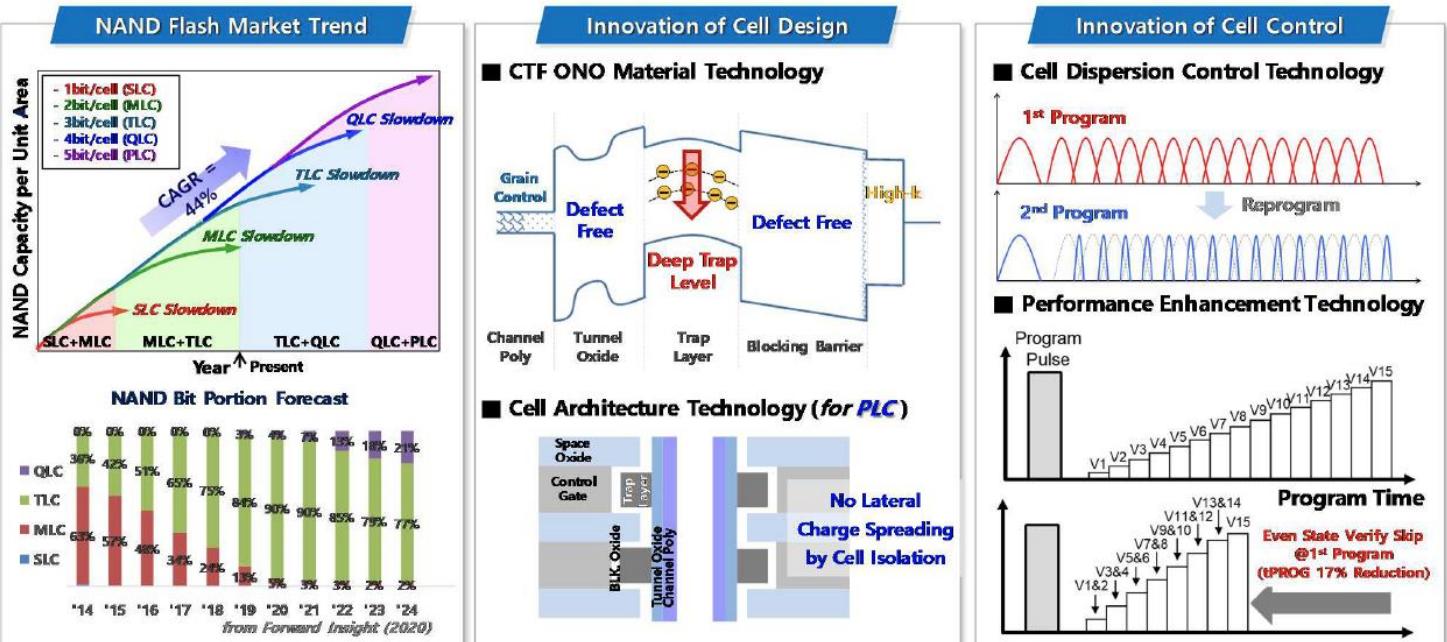
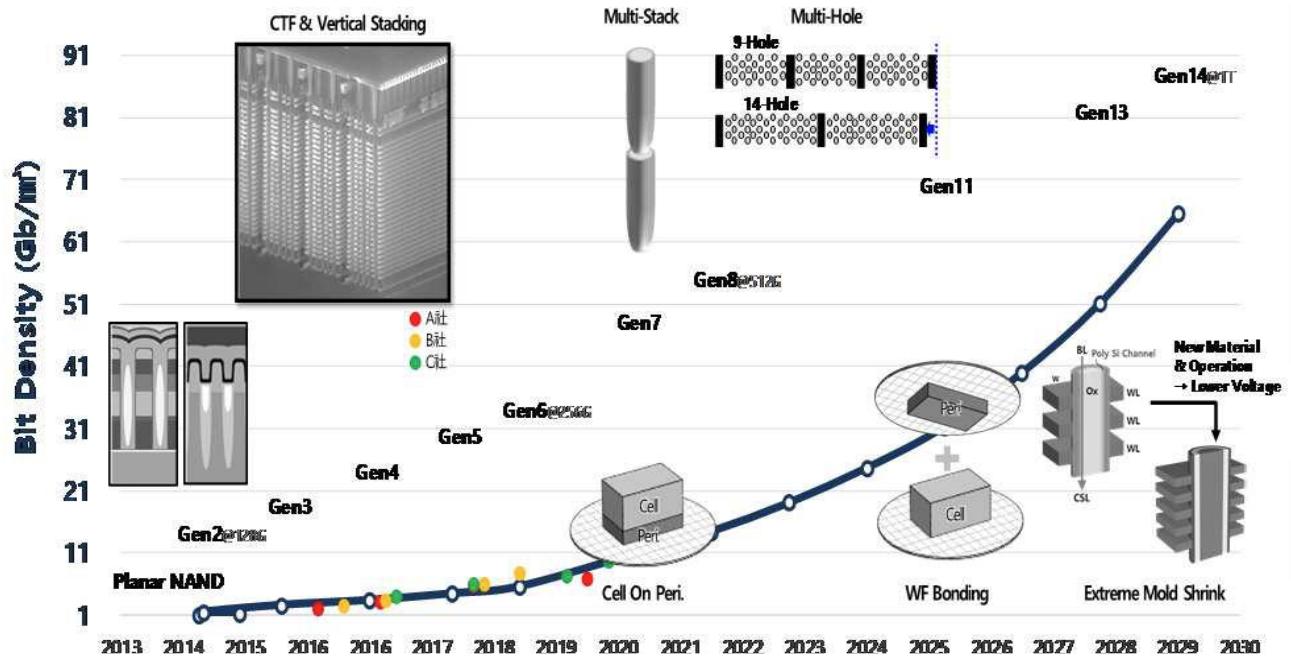
- DRAM capacity/bandwidth



K.Kim, IEDM,'21

Flash Scaling

- Density and architecture scaling



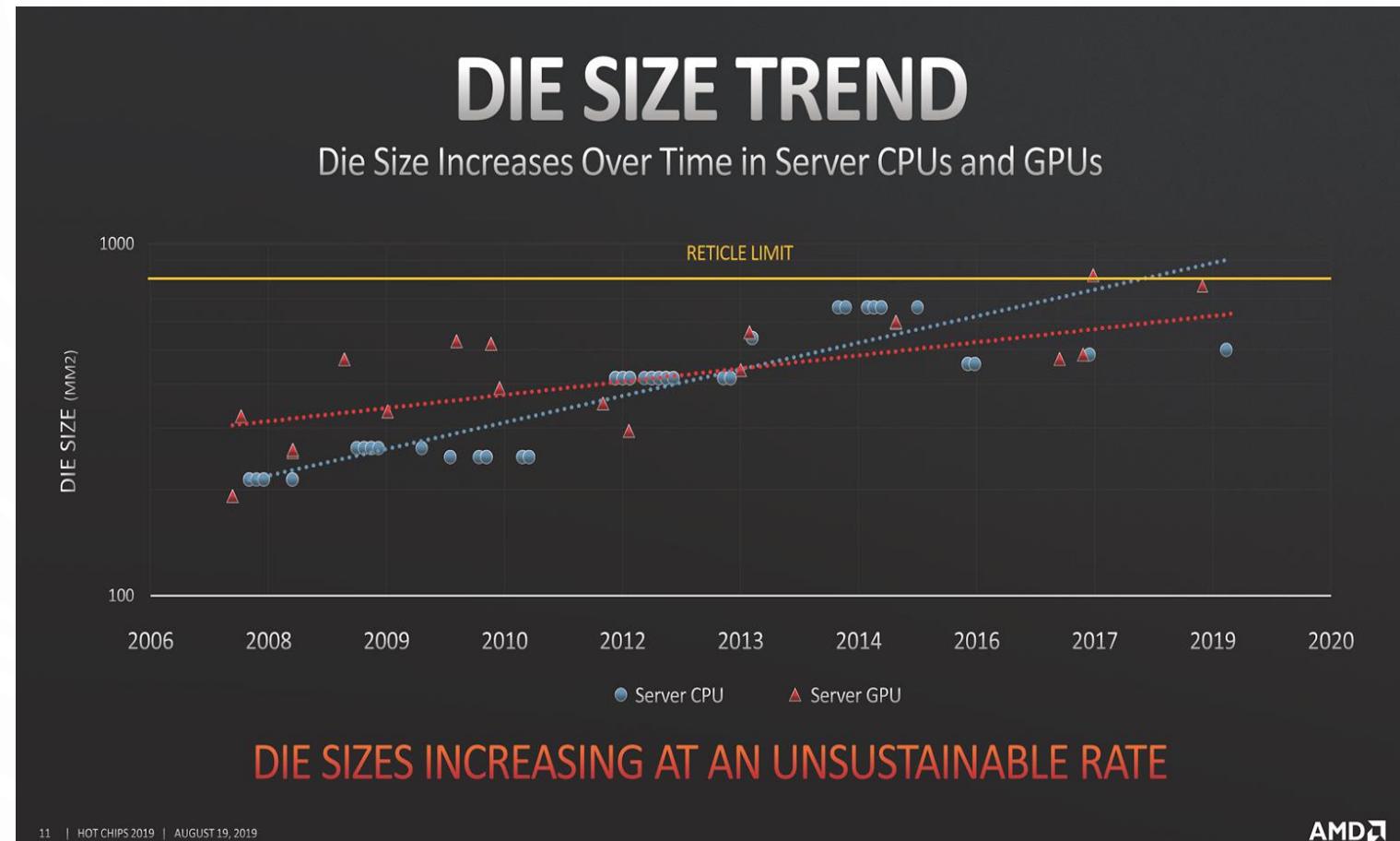
K.Kim, IEDM,'21



Chiplets

Die Size Trend

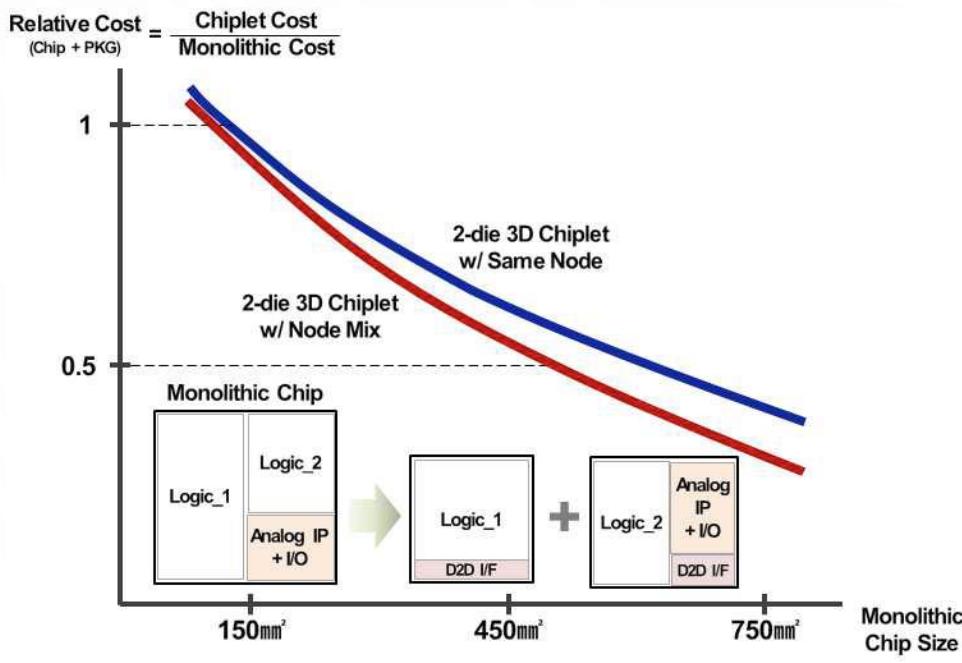
- To increase functionality and performance, die sizes have been increasing
 - Yield, cost tradeoffs



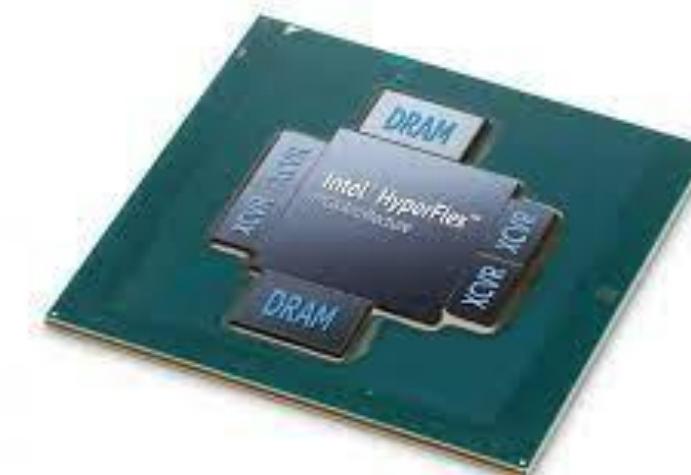
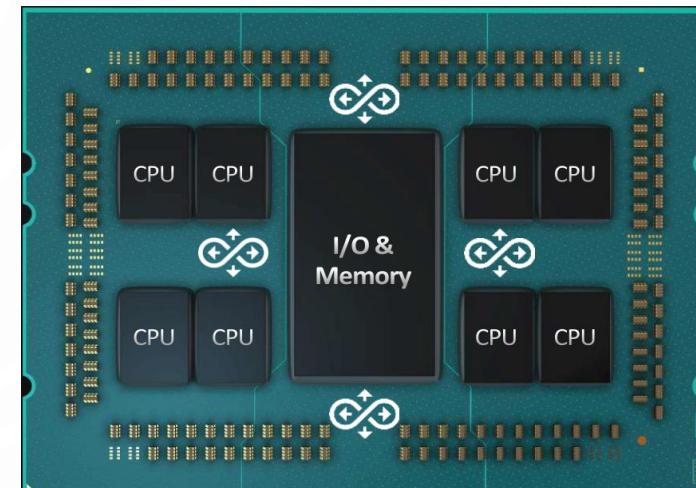
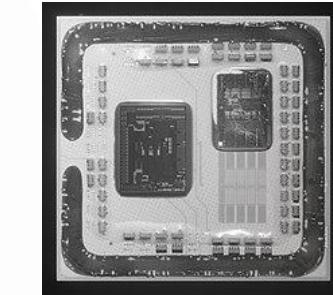
L. Su, HotChips'19 Keynote

Migration to Chiplets

- Split the product into multiple dies
 - Same or mixed technologies
- Increase functionality, performance
@ lower cost
- Mix technologies

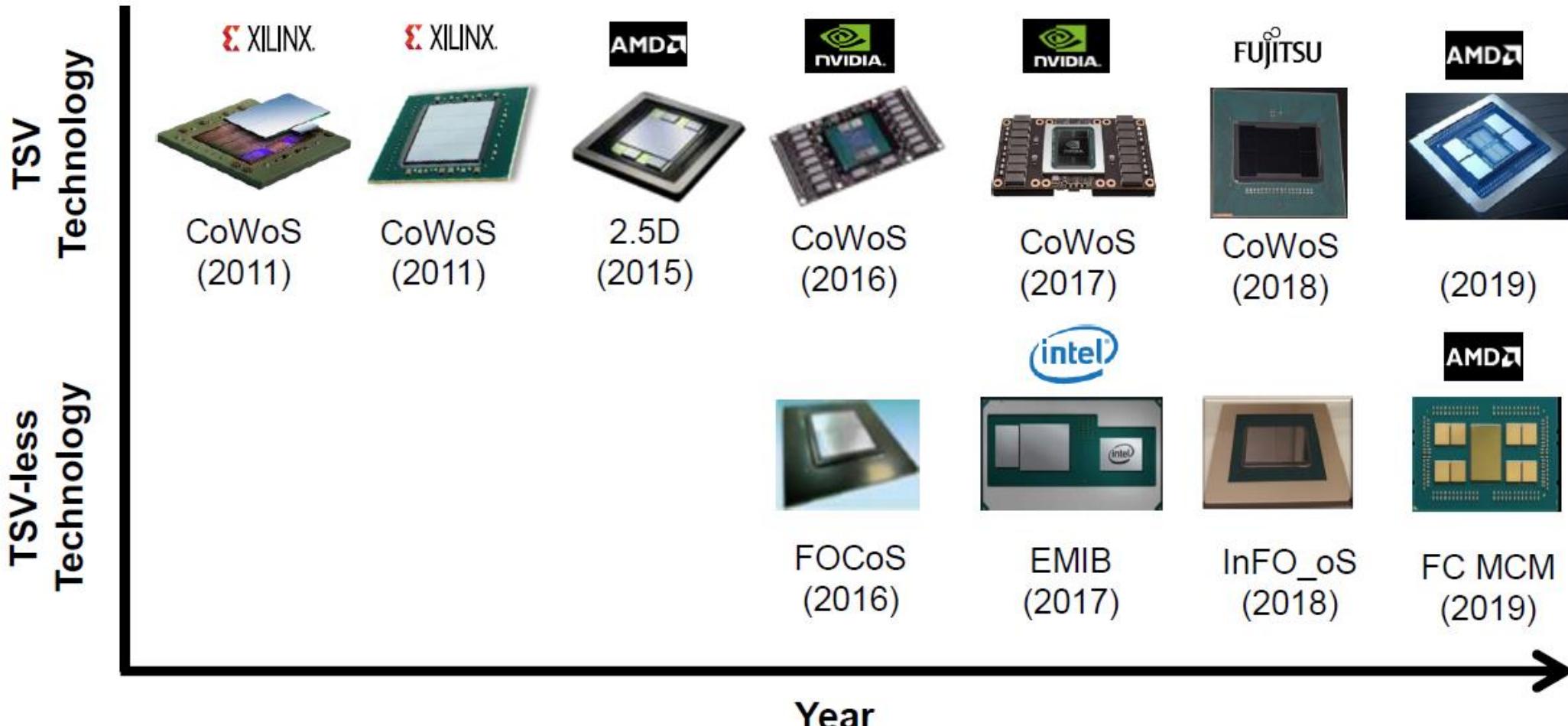


Plot from K.Kim, IEDM'21



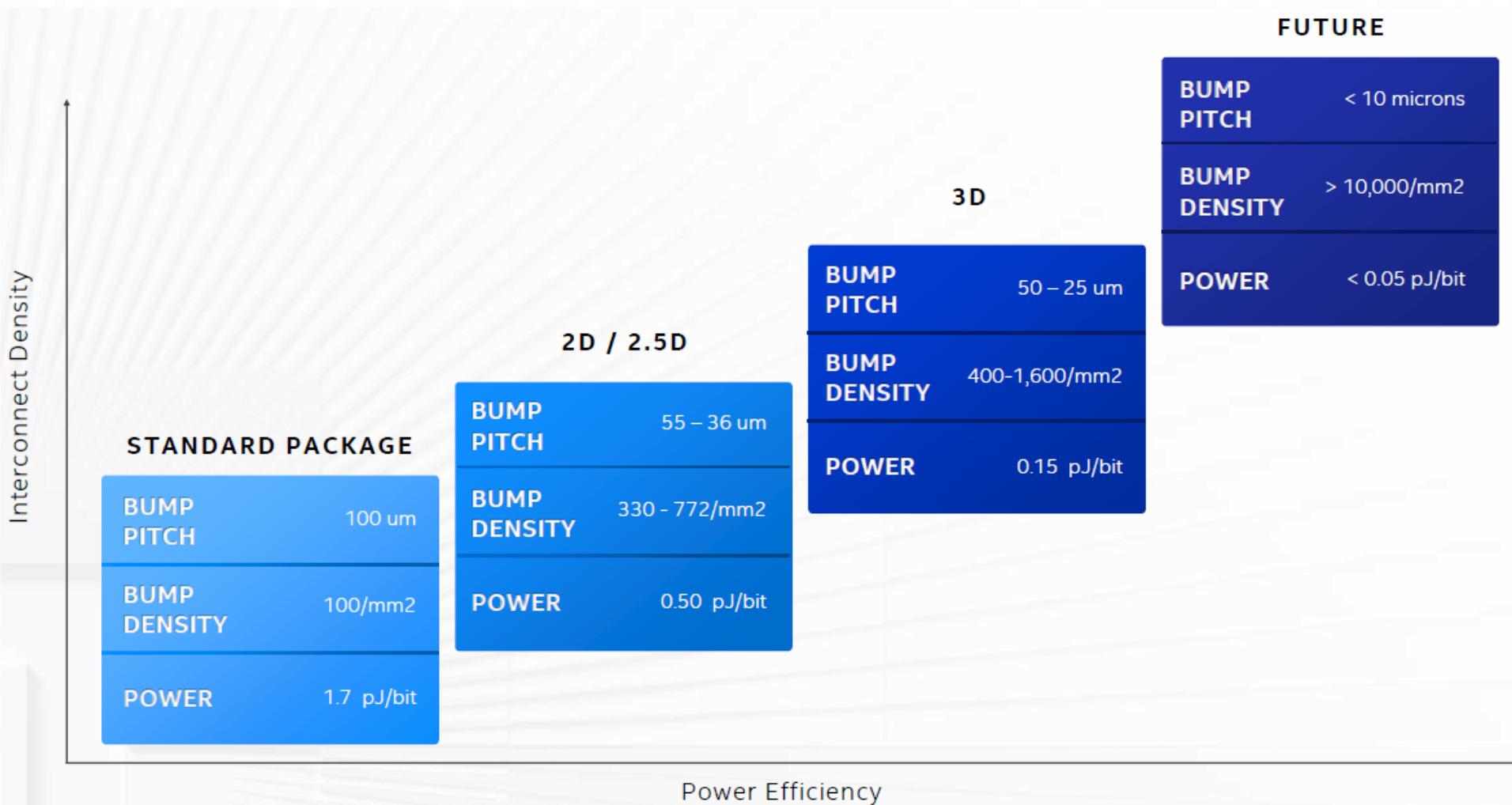
2D and 2.5D Chiplet Interfaces

- High-density interfaces have been evolving over the past decade



Interconnect Density Scaling

- Bump density and BW/edge or BW/area



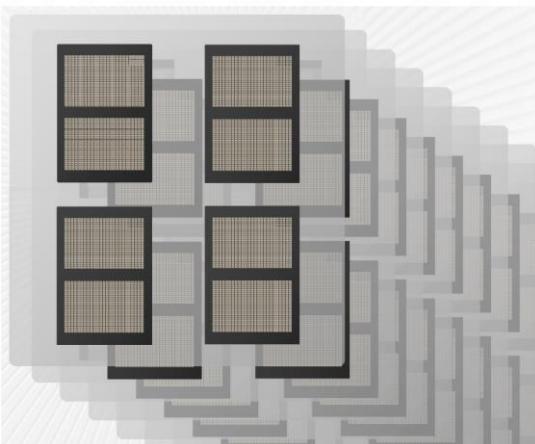
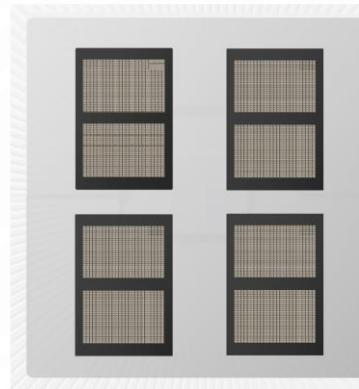
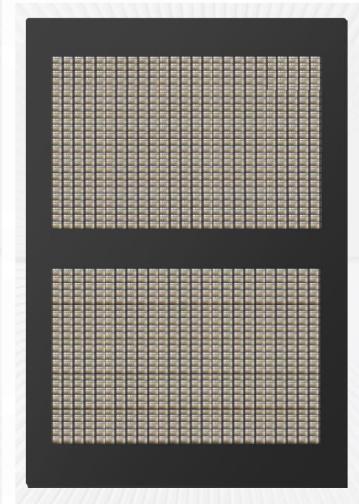
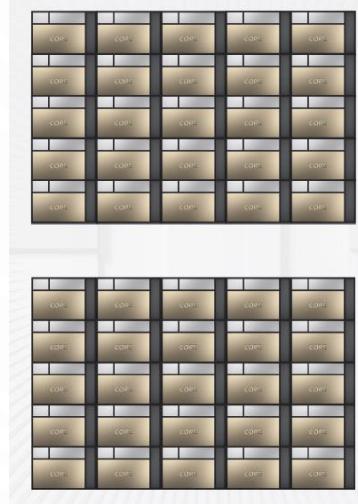
Adapted from R. Koduri keynote, Hot Chips 2020

Scaling is scale-out ... Getting to 1M cores/system

On-chip integration: 25x
Tech-scaling: 50x
Packaging: 2x
Scale-out: 400x



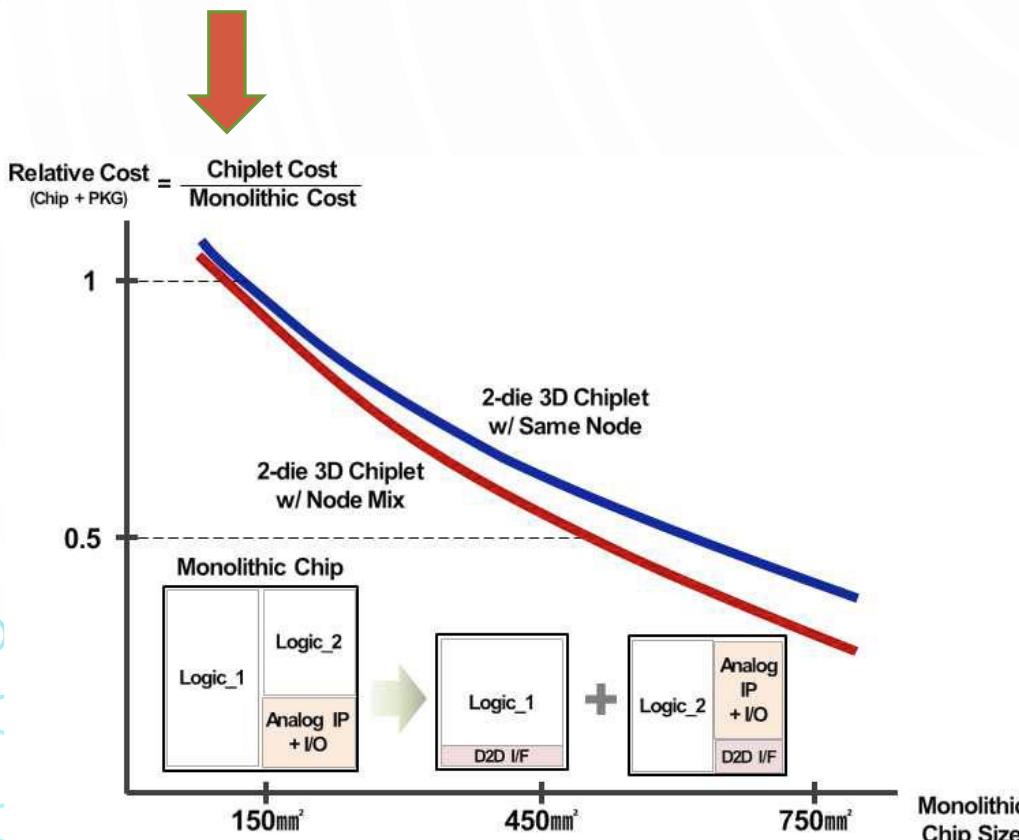
Today's core



Adapted from R. Koduri keynote, Hot Chips 2020

Some Open Issues

- High-value (e.g. hyperscale) products are driving the chiplet technology
 - What about sub-150mm² dies?



Cost of disintegration:

- AIB 1.0: 12mm² in 16nm @ \$0.1/mm² = \$1.2 on each side (50k mm² on a \$5k 12-in wafer; 3nm wafers are \$20k – chiplet interface is 2 x \$4.8)
- Substrate cost: >\$10 (could be >\$100)
- Test escape losses

Sum: \$25+ (but can be >\$100)

Chiplets are not for free

Can they offset the NRI costs?
Make medium volume ASICs affordable?

Summary

- FinFET and FDSOI processes deployed now
 - Expected to be replaced by nanosheets
- Lithography and manufacturing restrict design rules
 - Need to be aware of implications on design
 - EUV entering production
- More changes coming: forksheets, buried power rails, chiplets – 2.5D and 3D
 - Plurality of interconnect standards

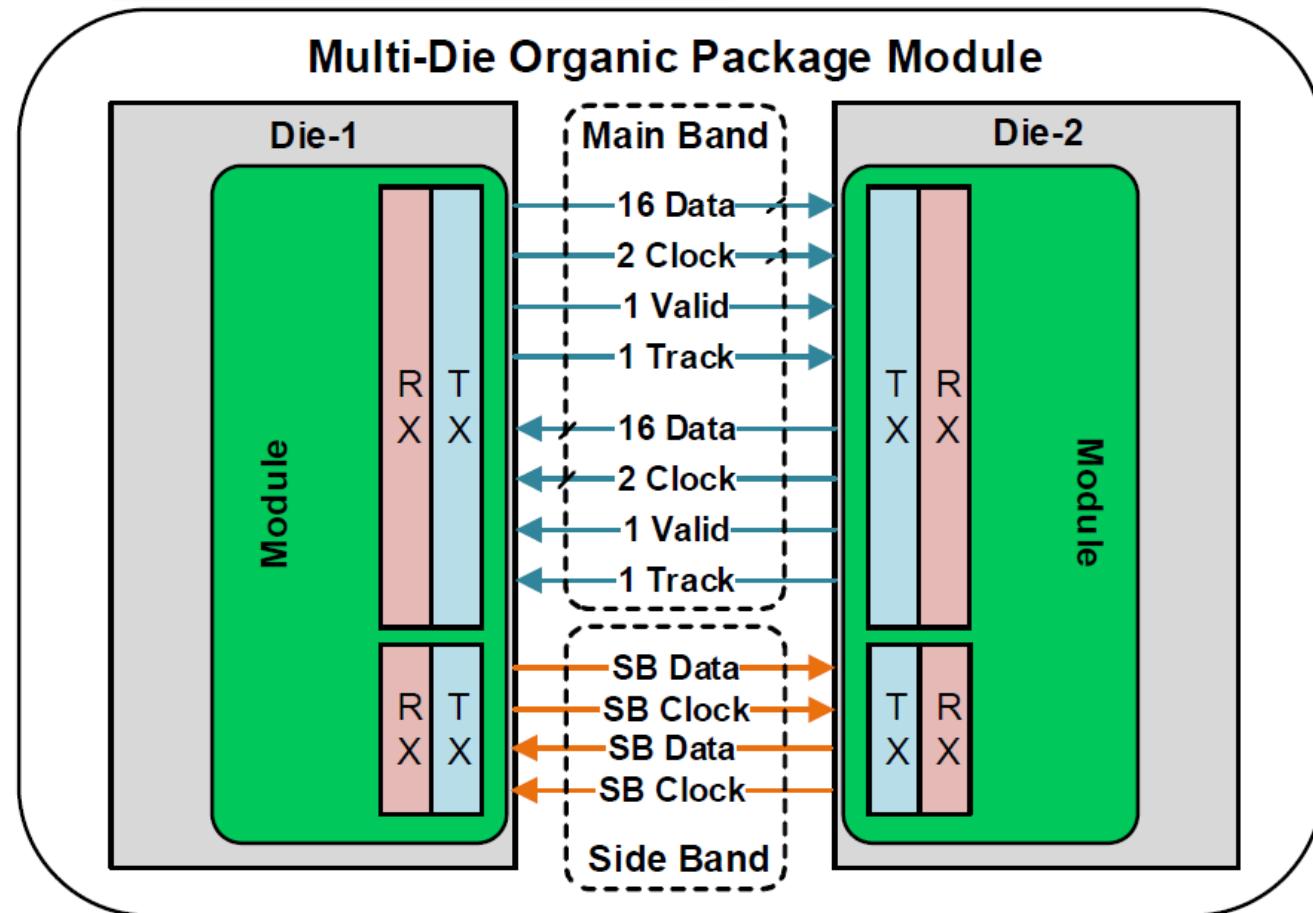


Universal Chiplet Interconnect Express (UCle) Overview

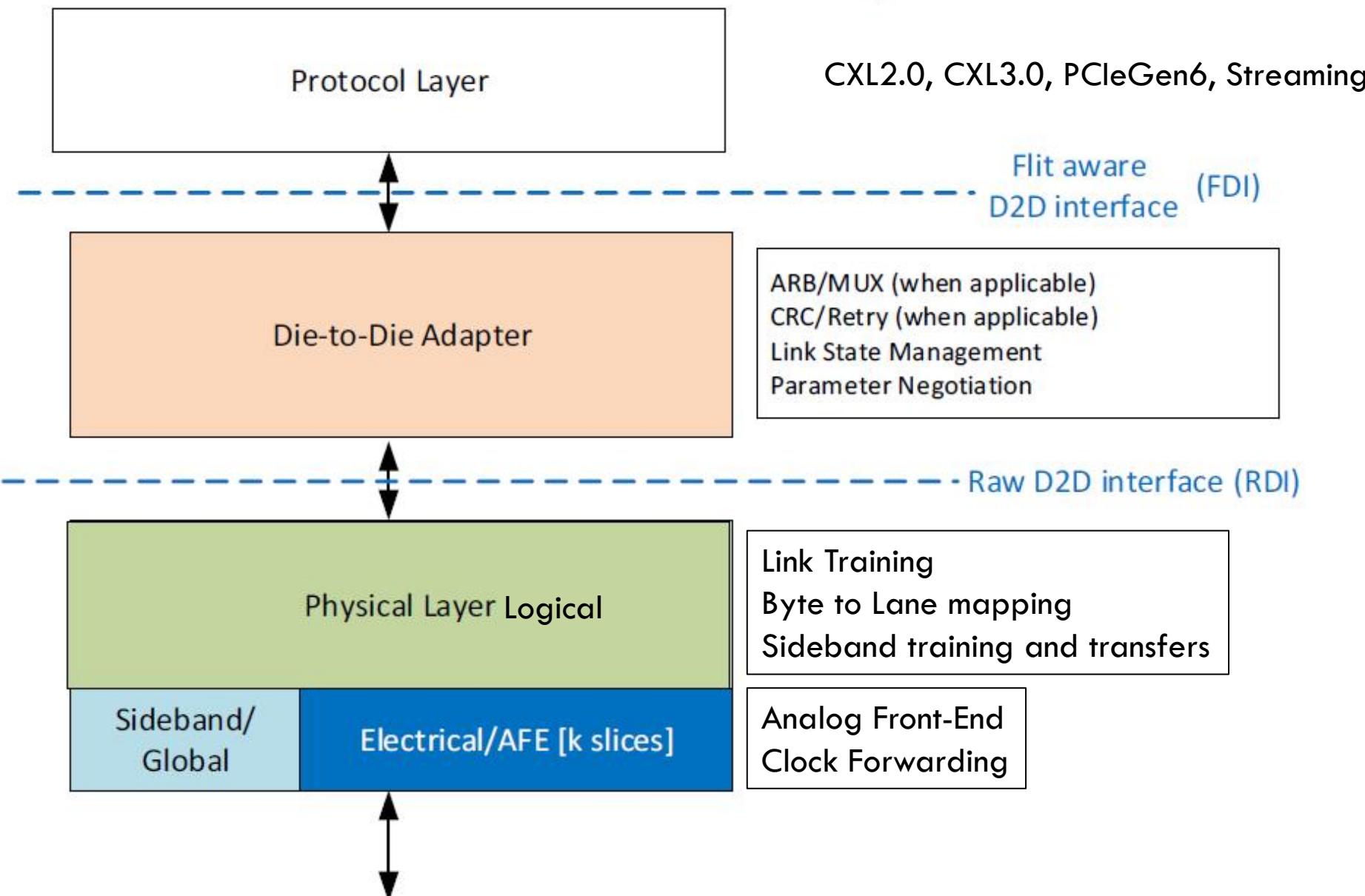
- Electrical Physical Layer

UCle: Standard Package Module

- Key attributes of electrical specification include:
 - 4, 8, 12, **16**, 24 and 32 GT/s data rates
 - Advanced and **Standard** package interconnects
 - Clock and power gating mechanisms
 - Single-ended unidirectional data signaling
 - DC coupled point-to-point interconnect
 - Forwarded clock for transmit jitter tracking
 - Matched length interconnect design within a module
 - Tx driver strength control and unterminated Rx for Advanced Package
 - Tx termination and data rate and channel-reach-dependent Rx termination for Standard Package
- Multiple modules can be integrated on an SoC
- Low speed sideband bus for initialization, Link training and configuration reads/writes
 - Sideband consists of a single-ended sideband data Lane and single-ended sideband Clock lane in both directions (transmit and receive)

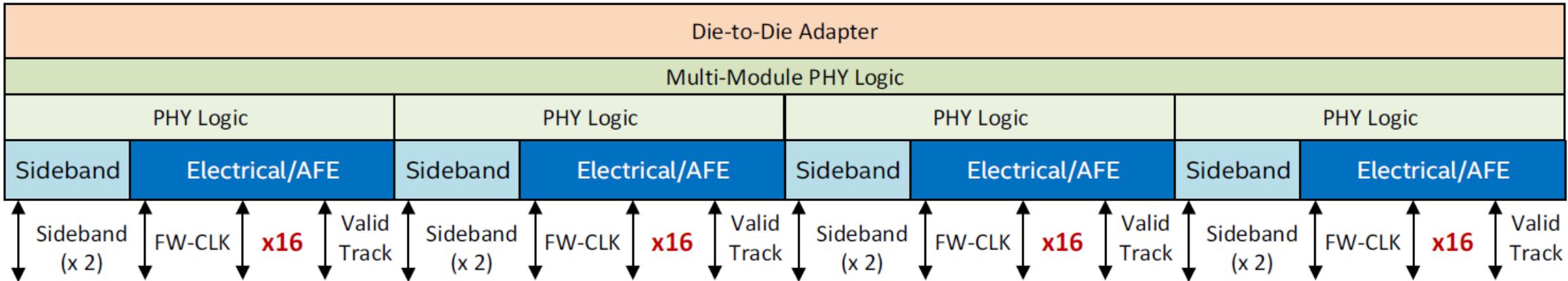


UCle: Interface Partitioning



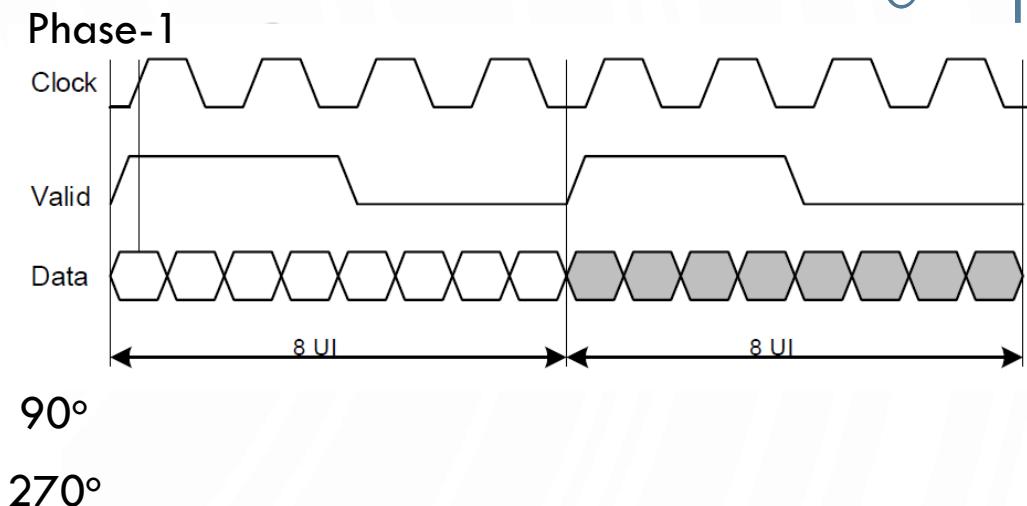
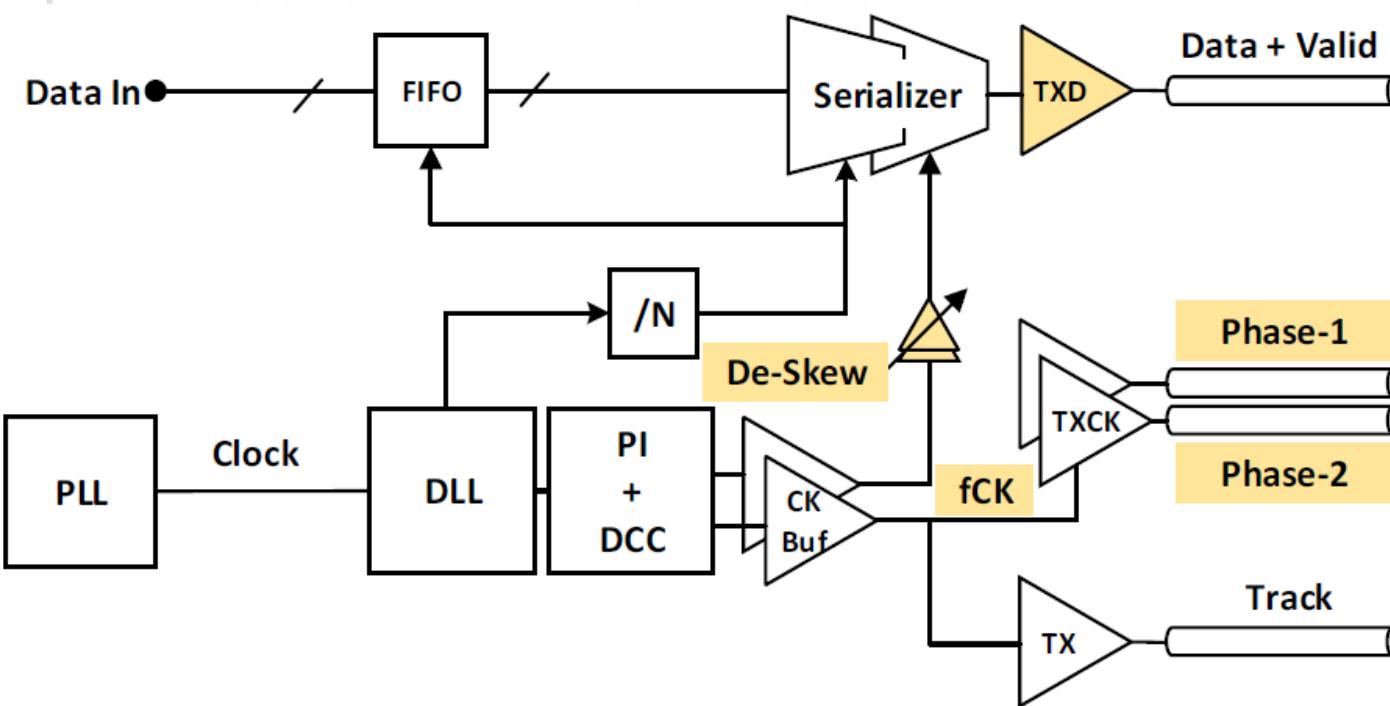
UCle: Multi-Module Configurations

Example 4-module interface configuration



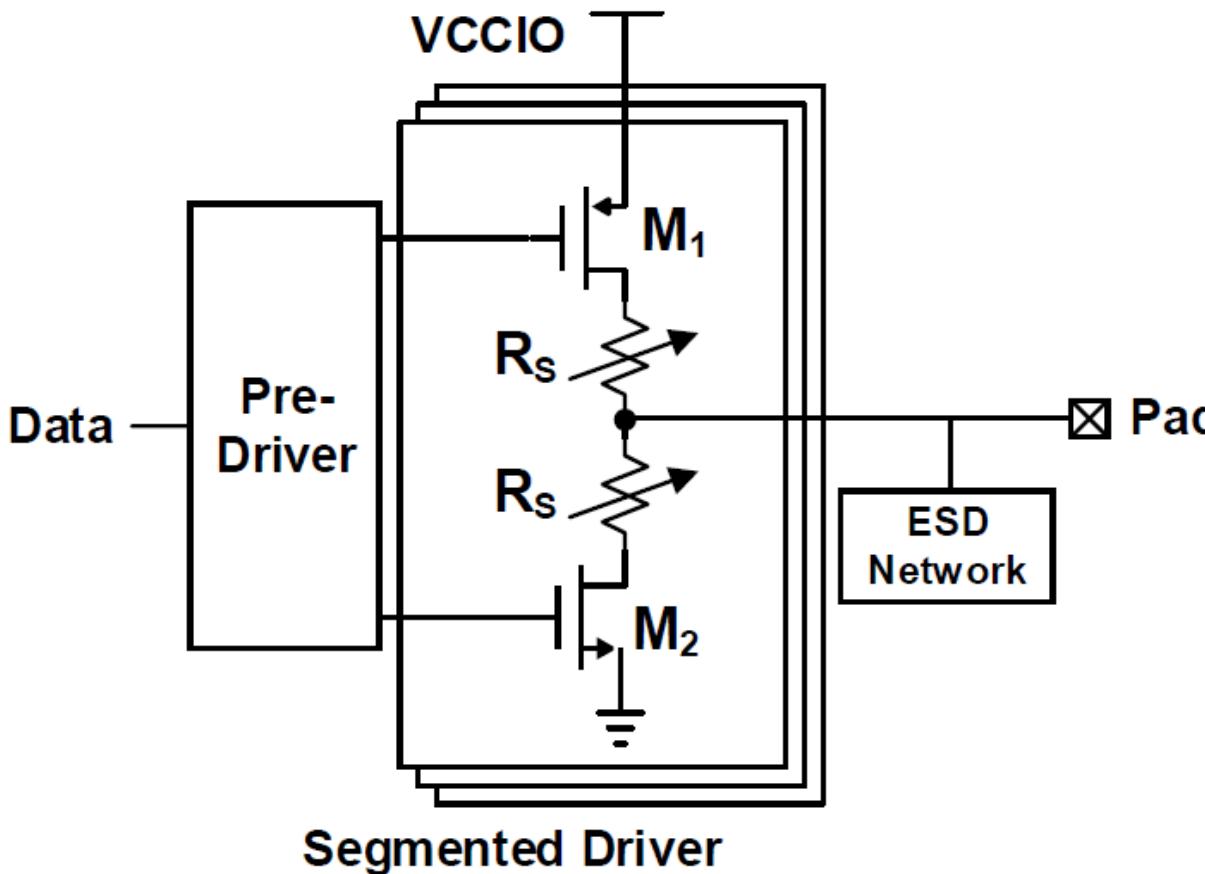
- 1, 2 and 4 modules allowed per interface

UCle: Transmitter

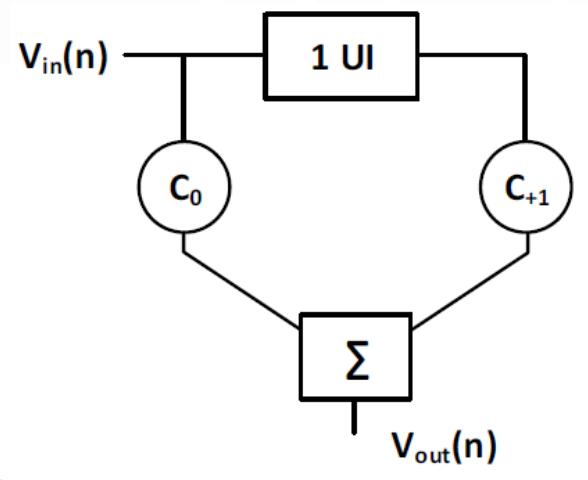


- Valid signal
 - Used to gate the clock distribution to all data Lanes to enable fast idle exit and entry
 - Valid framing
 - Tx for Valid signal the same as regular data Tx
- Track signal
 - Can be used for PHY to compensate for slow changing variables such as voltage or temperature
 - Unidirectional signal similar to a data bit
 - Transmitter sends a copy of Phase-1 of the clock signal when requested over the sideband by the Receiver.

UCle: Transmitter Driver

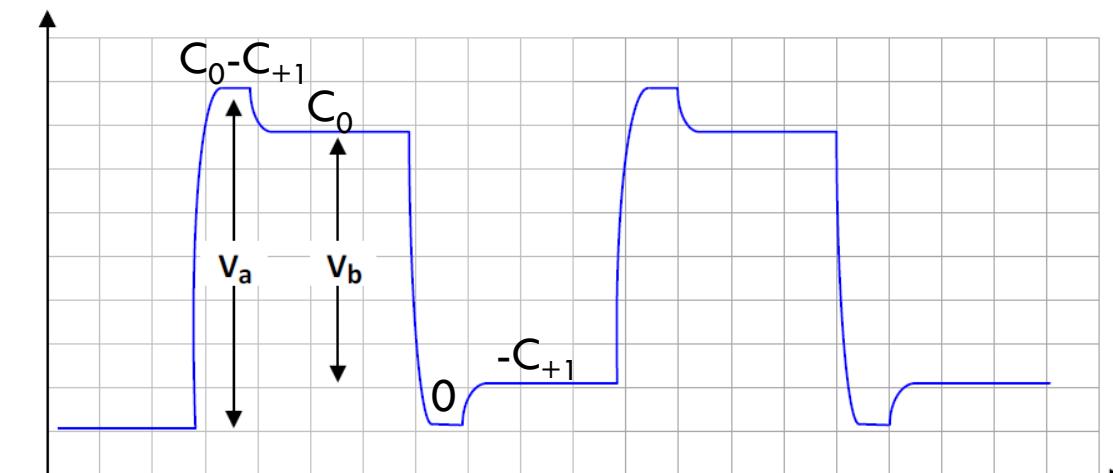


- Control loop or training to adjust output impedance to compensate for PVT variations
- Must Hi-Z in low power state



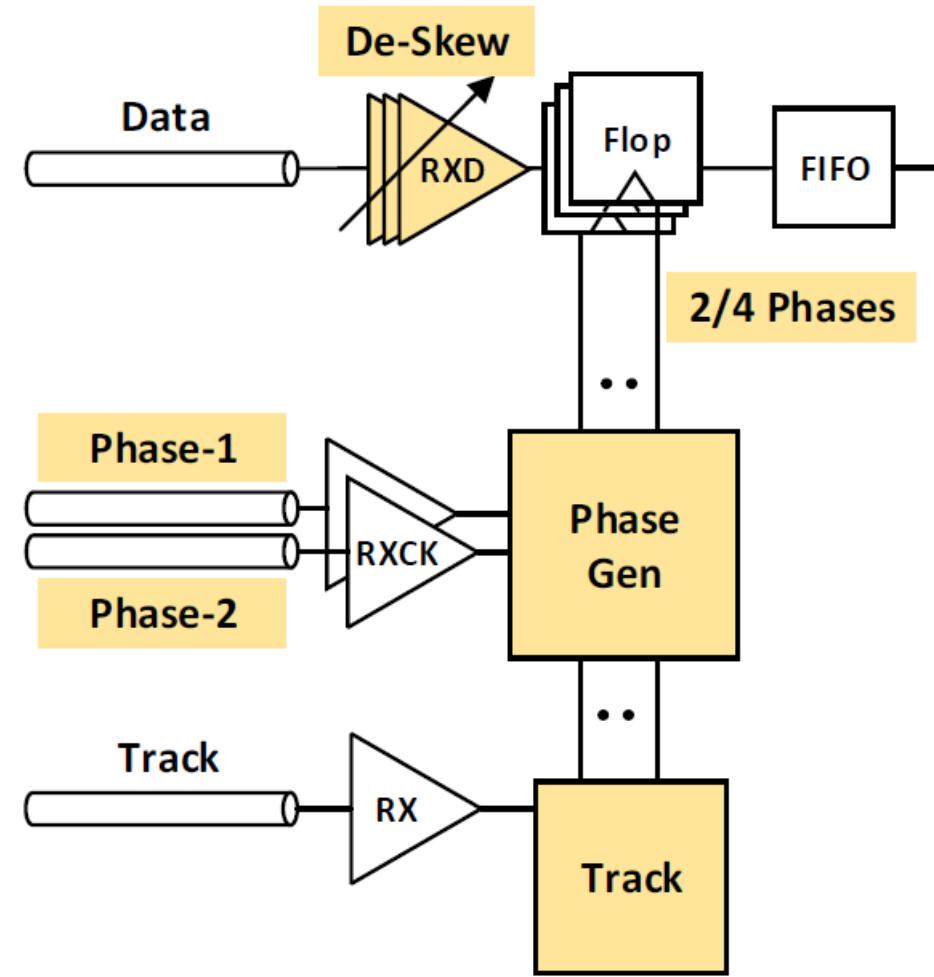
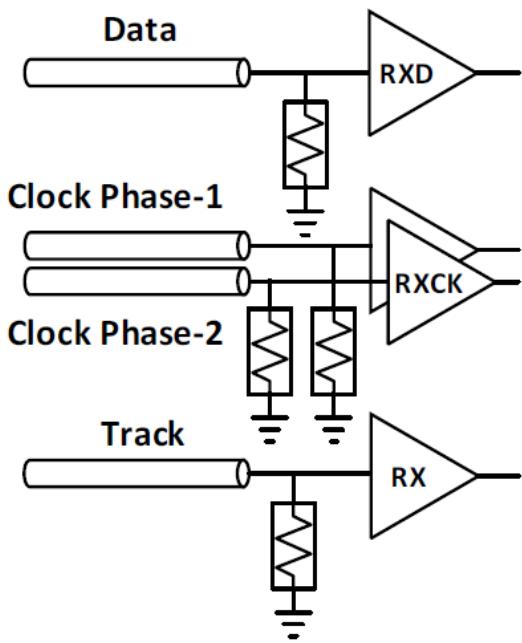
$$V_{out}(n) = C_0 V_{in}(n) + C_{+1} V_{in}(n-1)$$

$$|C_0| + |C_{+1}| = 1$$



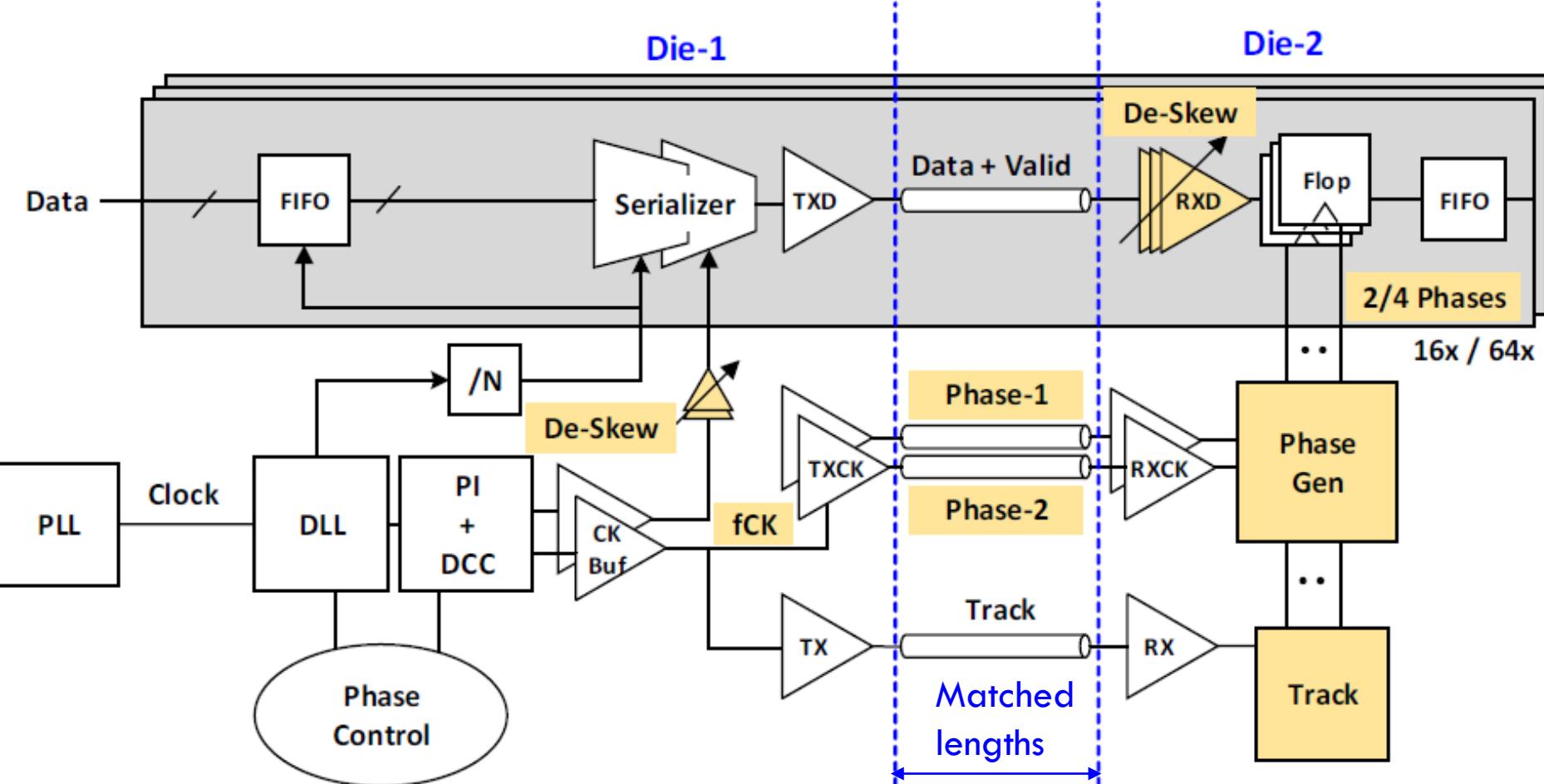
Transmit de-emphasis
 $V_b/V_a = (C_0 + C_{+1})/(C_0 - C_{+1})$

UCle: Receiver



- Received clock is used to sample the incoming data
 - Receiver must match the delays between the clock path and the data/valid path to the sampler to minimize the impact of power supply noise induced jitter
- Data Receivers implemented as 2-way or 4-way interleaved
 - For 4-way interleaved implementation the Receiver needs to generate required phases internally from the two phase of the forwarded clock, which may require duty cycle correction capability on the Receiver

UCle Electrical Physical Layer – putting it all together

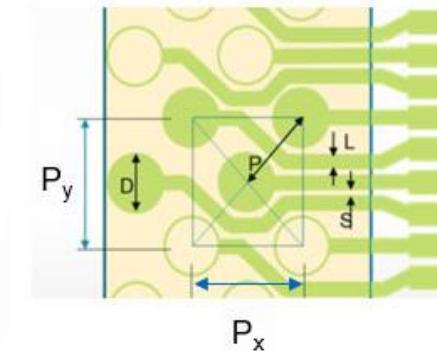
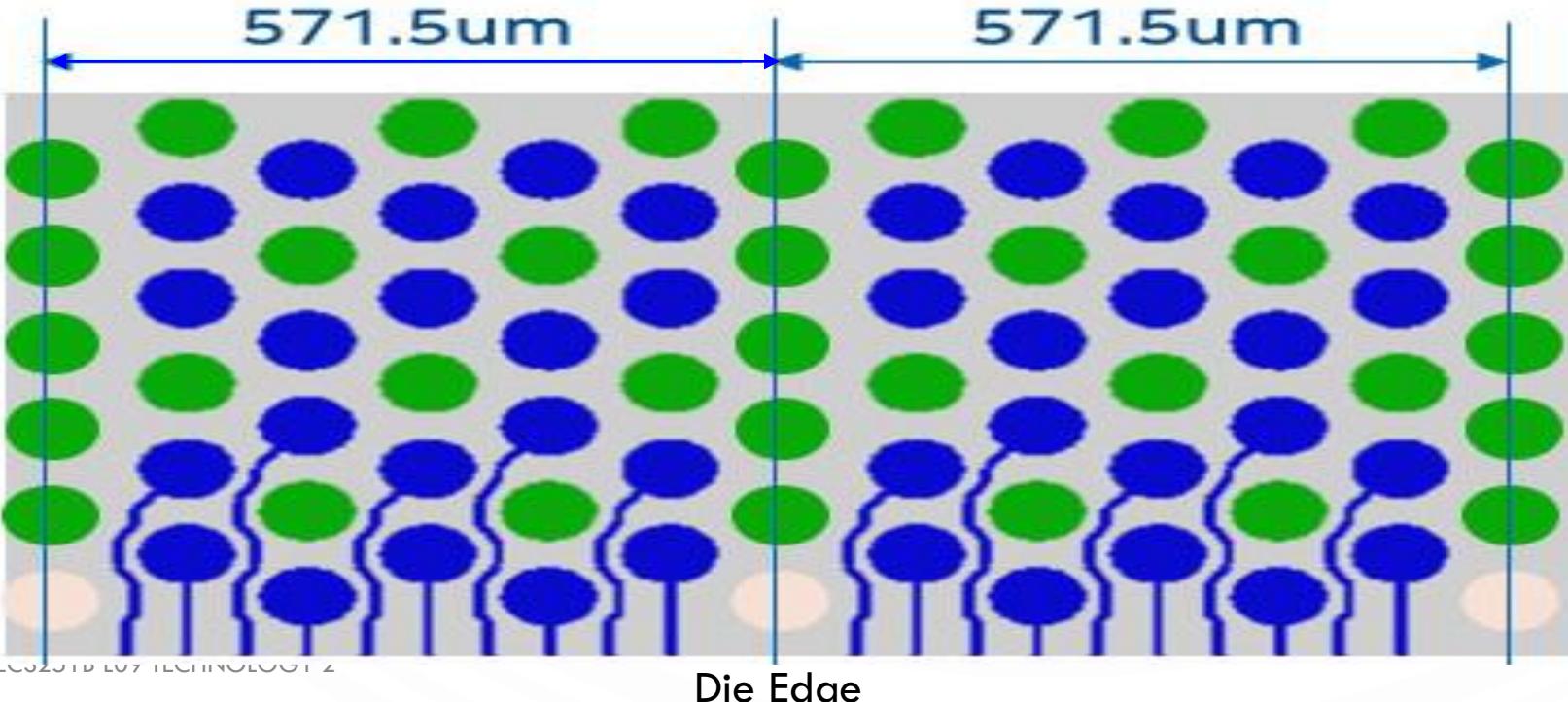


- Phase adjustment performed at Tx based on link training info from Rx

UCle: Standard Package Bump Map

Column 0	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11
	txdata5b		txcksb		vccao		vccao		rxcksb		rxdata5b
vccio		vccio		vccio		vccio		vccio		vccio	
	vss		vss								
vccio		txdata7		txdata9		vccio		rxdata8		rxdata6	
	txdata5		txckn		txdata11		rxdata10		rxckp		rxdata4
vss		vss		vss		vss		vss		vss	
	txdata4		txckp		txdata10		rxdata11		rxckn		rxdata5
vss		txdata6		txdata8		vss		rxdata9		rxdata7	
	vss		vss								
vccio		txdata3		txdata13		vccio		rxdata12		rxdata2	
	txdata1		txvld		txdata15		rxdata14		rxtrk		rxdata0
vccio		vss		vss		vccio		vss		vss	
	txdata0		txtrk		txdata14		rxdata15		rxvld		rxdata1
vss		txdata2		txdata12		vss		rxdata13		rxdata3	

Die Edge

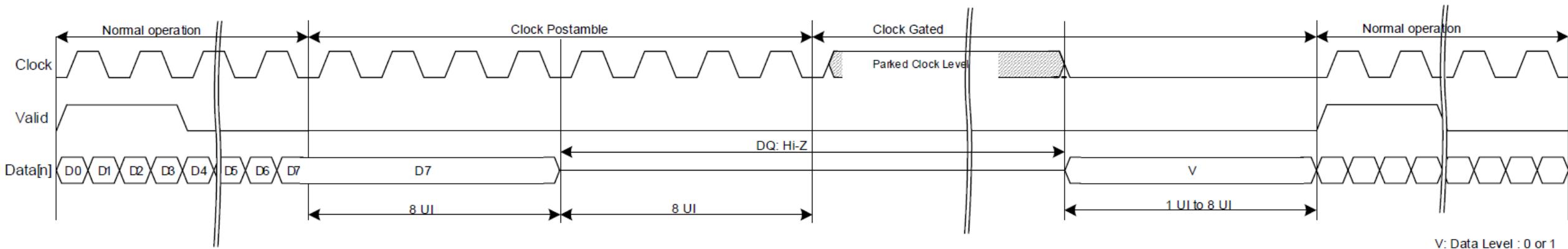


Reference design 1:
 $P=110\text{um}$,
 $P_x = 110\text{um}$,
 $P_y = 190.5\text{um}$

Reference design 2:
 $P=130\text{um}$,
 $P_x = 177\text{um}$,
 $P_y = 190.5\text{um}$

$$3*P_y = 571.5\text{um}$$

UCle: PHY layer – Clock gating



- **Entry**
 - TxData must send the last UI for at least 1UI and up to 8UIs and then Hi-z
 - Valid Lane must be held low
 - Clock idle state level must alternate between differential high and differential low during consecutive clock gating events
- **Exit**
 - TxData must precondition the Data Lanes to a 0 or 1 (1UI to 8UI) before normal transmission
 - Clock must drive a differential low (1UI to 8UI) before normal transmission

UCle: PHY Layer – Electrical Idle and Sideband signaling

- Some training states need electrical idle when Transmitters and Receivers are waiting for generate and receive patterns
 - Tx and Rx are enabled
 - Data, Valid and Track held low
 - Clock is at high and low
- Sideband Signaling
 - Sideband data 800MT/s
 - Sideband clock 800MHz

