Innovations Enabling Semiconductor Roadmap

Chee Wee Liu (劉致為)

cliu@ntu.edu.tw

http://www.nanosioe.ee.ntu.edu.tw

High mobility/High K/ 3D x3D

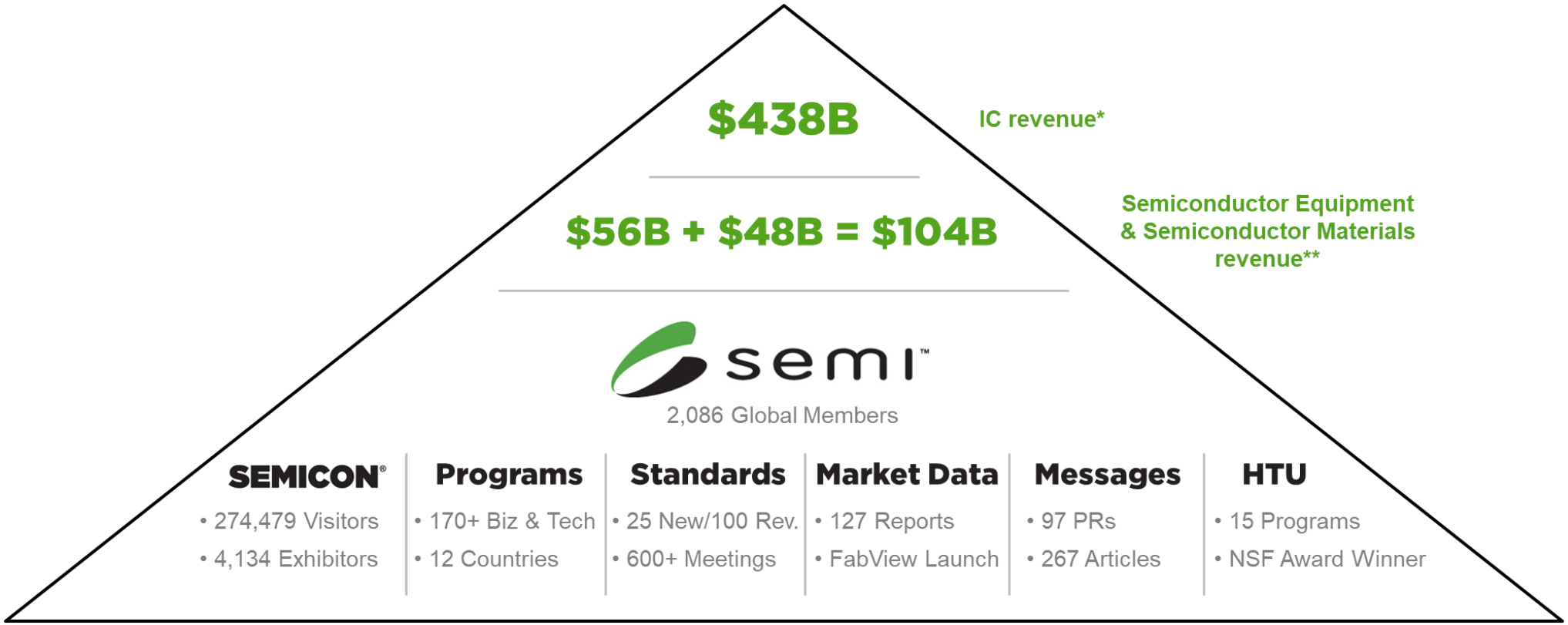
GDP similar to Philippine (1.2x) if semi dies

2017 年全球半導體廠商營收前十大排名

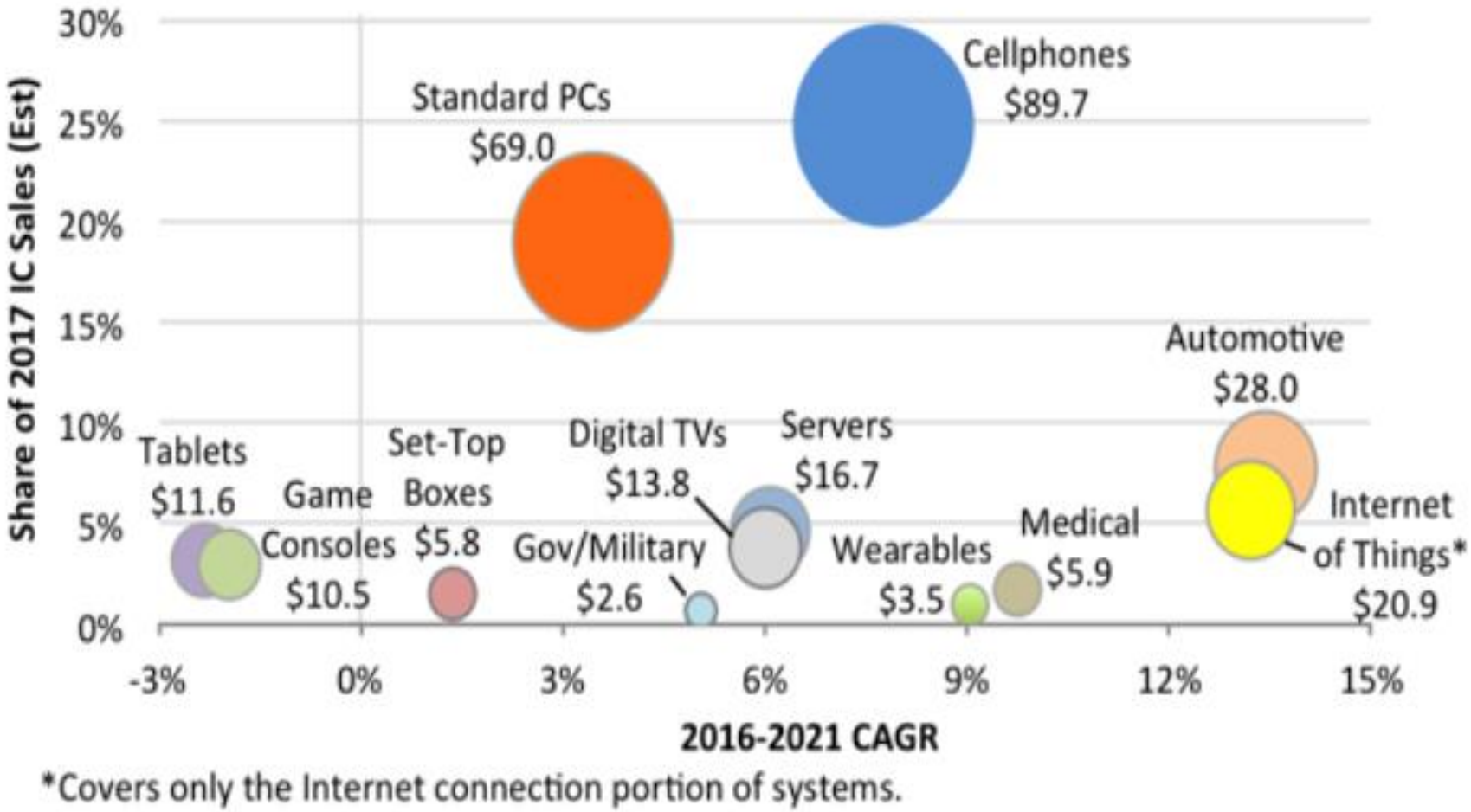
Source: Gartner (January 2018)

2017 Rank	2016 Rank	Vendor	2017 Revenue	2017 Market Share (%)	2016 Revenue	2016-2017 Growth (%)
1	2	Samsung Electronics	61,215	14.6	40,104	52.6
2	1	Intel	57,712	13.8	54,091	6.7
3	4	SK Hynix	26,309	6.3	14,700	79.0
4	6	Micron Technology	23,062	5.5	12,950	78.1
5	3	Qualcomm	17,063	4.1	15,415	10.7
6	5	Broadcom	15,490	3.7	13,223	17.1
7	7	Texas Instruments	13,806	3.3	11,901	16.0
8	8	Toshiba	12,813	3.1	9,918	29.2
9	17	Western Digital	9,181	2.2	4,170	120.2
10	9	NXP	8,651	2.1	9,306	-7.0
		Others	**174,418**	**41.6**	**157,736**	**10.6**
		Total Market	**419,720**	**100.0**	**343,514**	**22**

2017

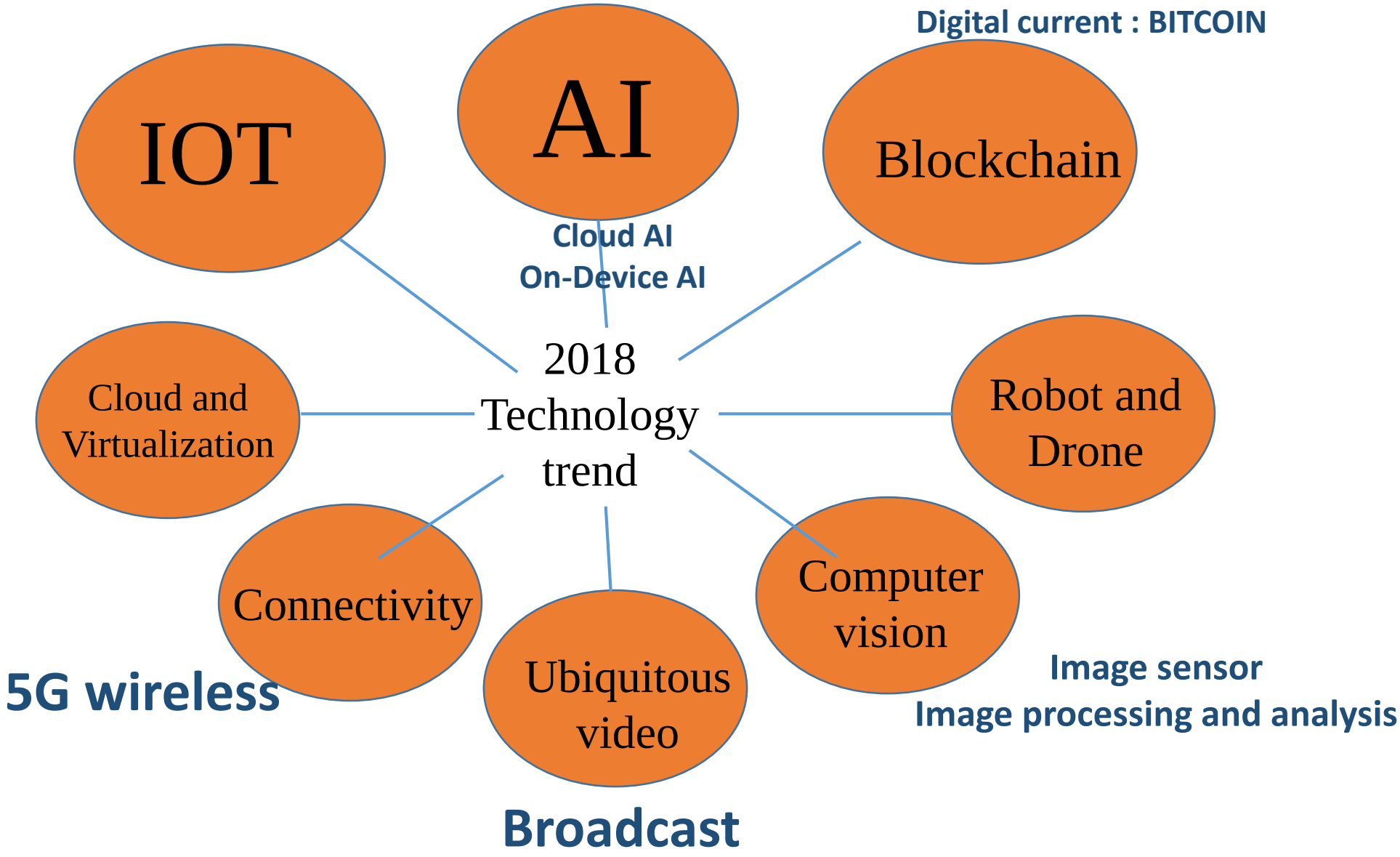


*IC Insights 10/17 forecast
**SEMI 12/17 forecast

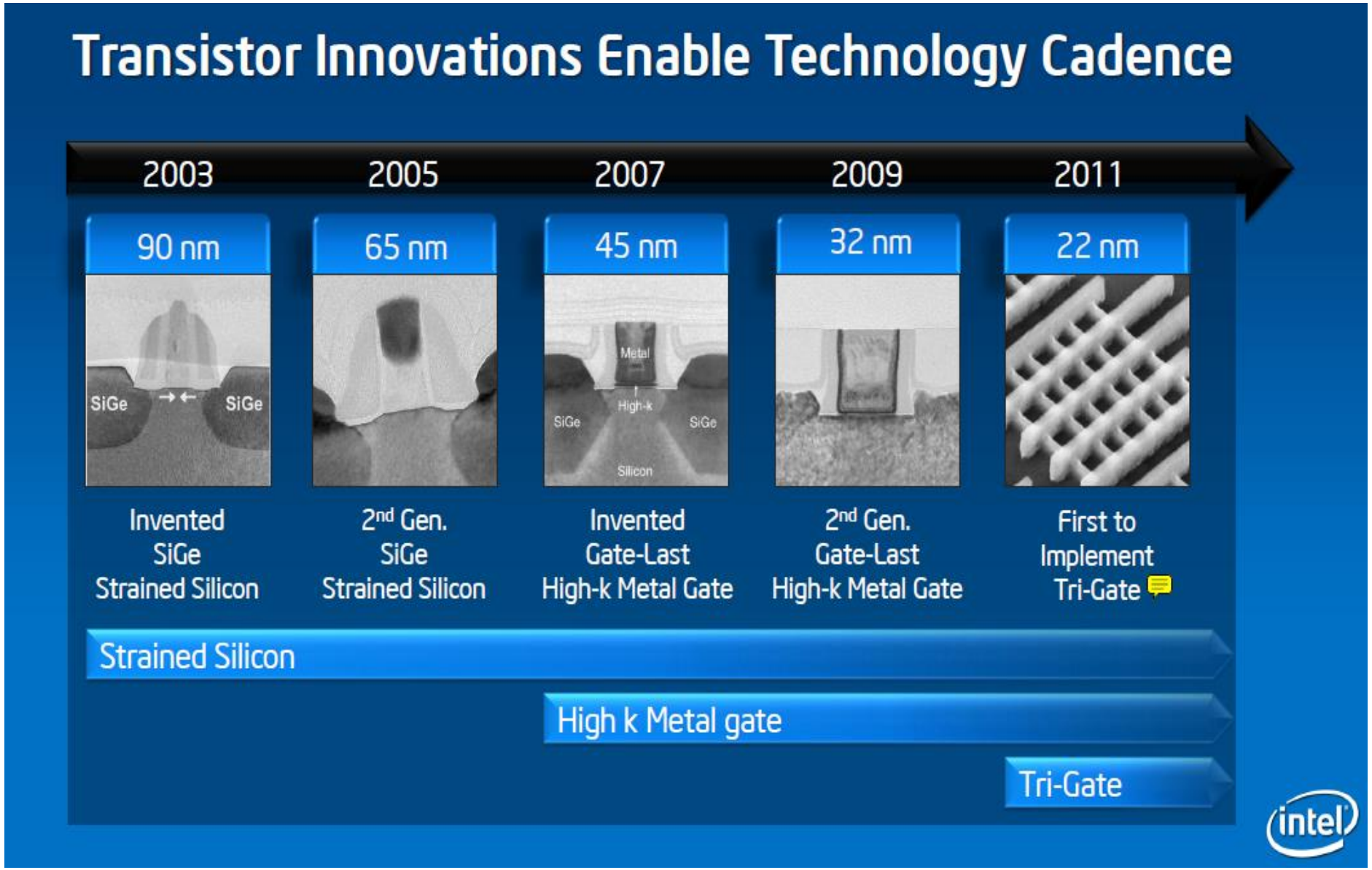


*Covers only the Internet connection portion of systems.

The top eight transformative technologies for the global technology market in 2018, IHS Markt report **Ref: https://www.businesswire.com/news/home/20180104005040/en/IHS-Markit-Identifies-Top-Technology-Trends-2018**



Strained Si starting from 90 nm node

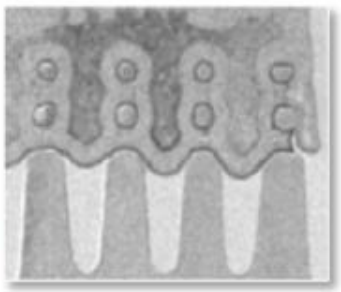


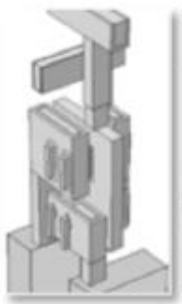
- Enabling technologies: high mobility, high-k/metal gate, 3D transistor

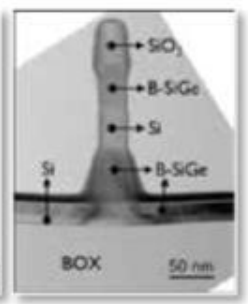
IMEC VIEW OF TRANSISTOR ROADMAP

6nm	5nm	4nm	3nm	2nm	
2020	2021	2022	2023	2024	
48nm 36nm 6.5T	42nm 32nm 6.5T	42nm 32nm 6.5T	36-42nm 21-24nm 5.5T-4.5T	36-42nm 21-24nm 4.5T-3T	Gate Pitch Metal Pitch # Tracks

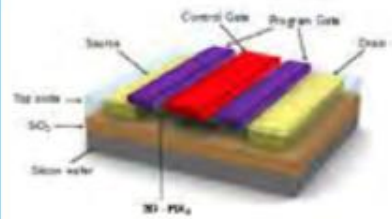

21nm finfet pitch

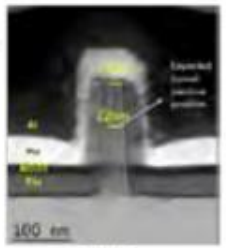

Nanowire FET

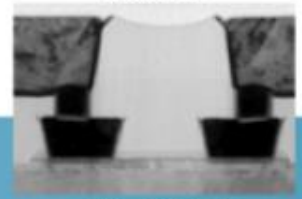

CFET

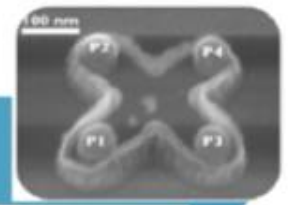


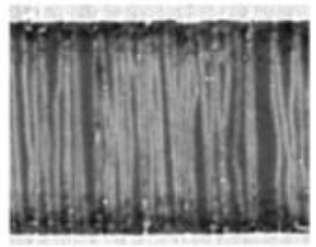
VFET

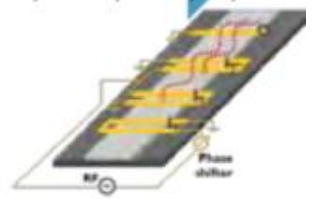

Ambipolar

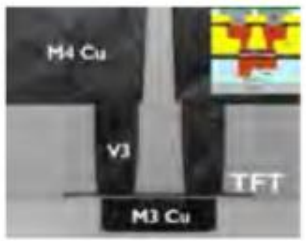

TFET


2D MX2 FET

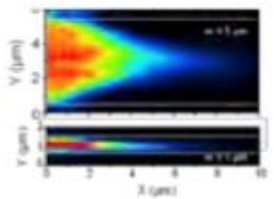

Spin-Torque Majority Gate


CNTFET





IGZO


Spin Wave Majority Gate

imec

TSMC's Roadmap

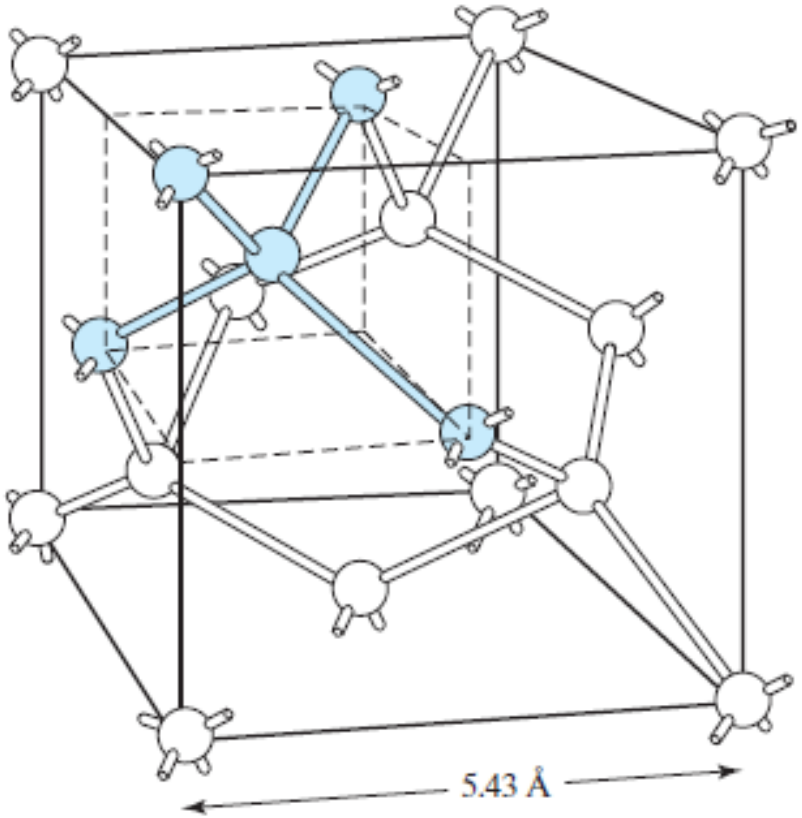
5/2/2018

https://www.eetimes.com/document.asp?doc_id=1333244&print=yes

- 7nm： 2018/5 in volume production
- 7nm+：early 2019, using EUV lithography ramping
- 5nm：
  - To start risk production of a 5-nm node in the first half of 2019.
  - To use EUV on multiple layers and 5-nm nodes.
  - To start 5-nm production in 2020.
- 2nm and beyond：
  - Stacked nanowires or nanosheets
  - Germanium channel with record-low contact resistance
  - 2D back-end materials including molybdenum disulfide
  - To enlarge copper grains to reduce resistance in interconnects.
  - selective dielectric-on-dielectric deposition process to enable self-aligning of copper vias.

Silicon Crystal Structure (Ge, SiGe, GeSn)

- ***Unit cell*** of silicon crystal is cubic.
- ***Each Si atom has 4 nearest neighbors*.**
- **Si sp3**



Large I_{on} and small V_{dd}

- 要馬兒好又要馬兒不吃草

$$I_{on} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{gs} - V_t\right)^2$$

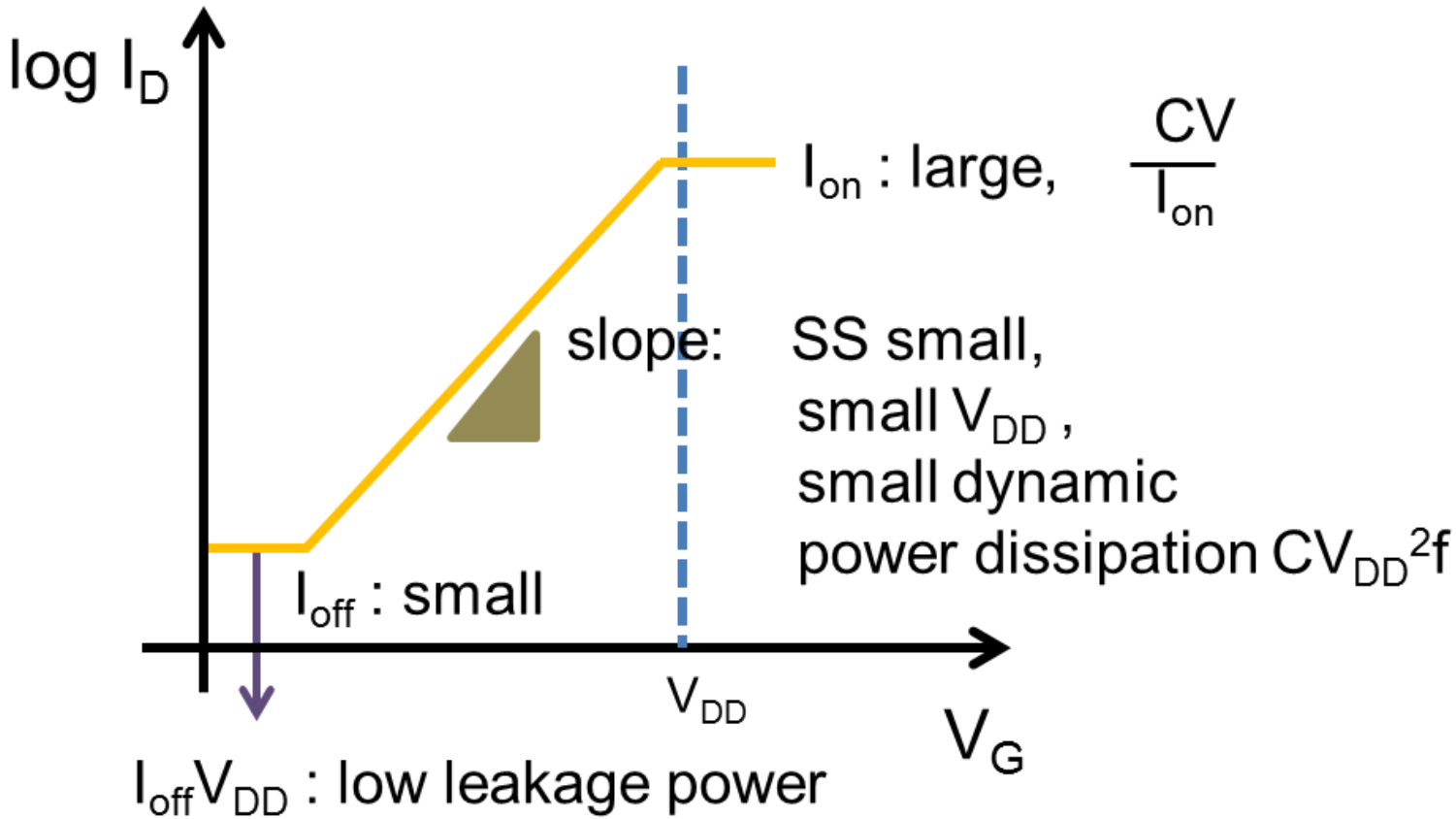
Large Ion → high speed CV/I

large μ_n , $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$ → large ϵ_{ox} small

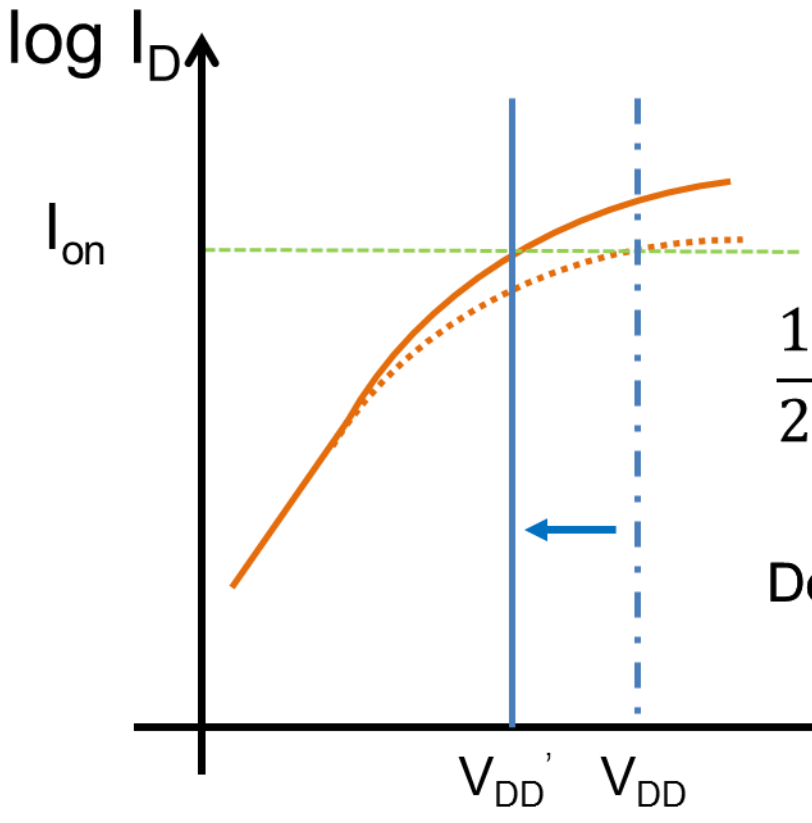
tox→ large tunneling current (big Ig)

$CV_{DD}{}^2 f$ → small V_{DD} → small SS → FinFET

I_{off} : small tunneling (many sources)


log I_D
I_{on} : large,
$\frac{CV}{I_{on}}$
slope:
SS small,
small V_{DD} ,
small dynamic
power dissipation CV_{DD}^2f
I_{off} : small
V_{DD}
V_G
$I_{off}V_{DD}$: low leakage power

Power Saving : To Increase Mobility



Mobility ↑

$$I_{on} = ne\mu$$

$$\frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{gs} - V_t\right)^2 \quad \uparrow$$

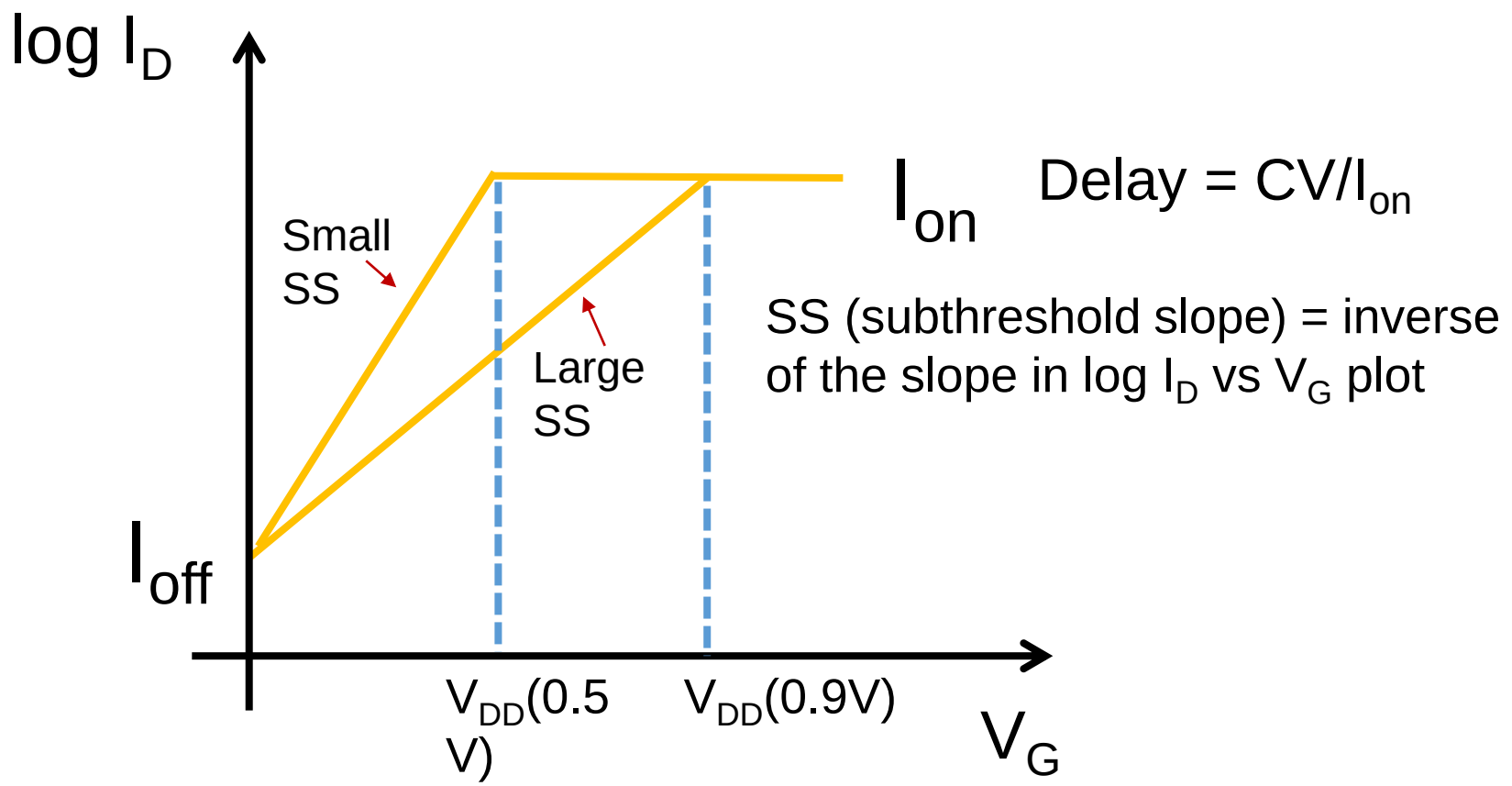
$$\text{Delay (Speed)} = \frac{CV}{I_{on}}$$

Power = $CV_{DD}{}^2 f$ ↓

Cox=ε/tox

- Channel: Strained Si => SiGe=> Ge ?

Reduce to Power (= $CV_{DD}^2 f$) in transistor level → small V_{DD}



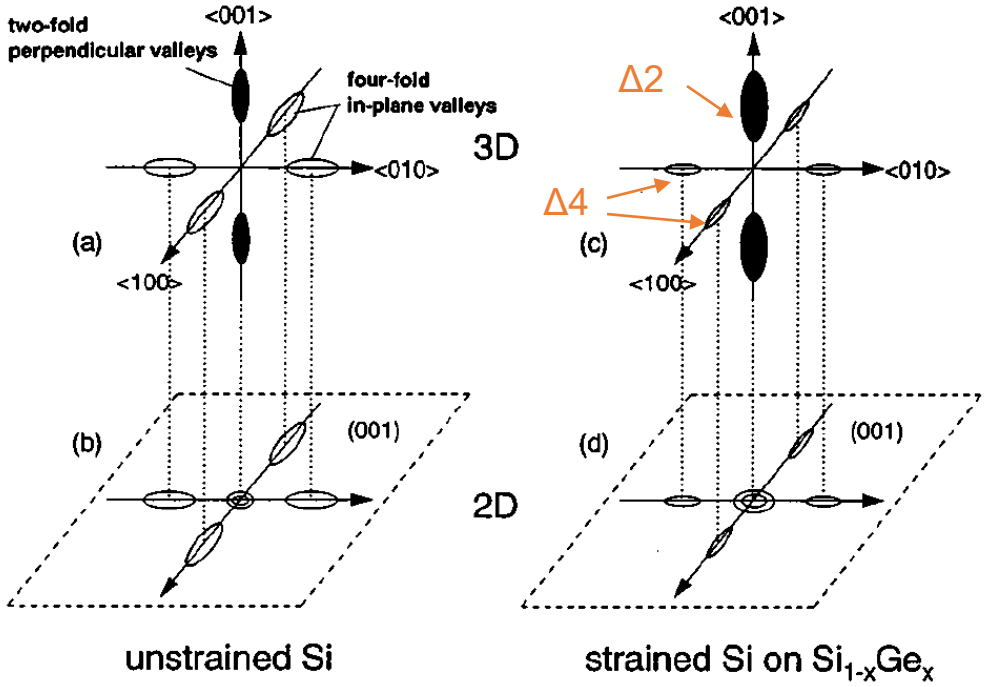
- SS ≥ 60 mV/decade for thermionic emission
- SS ≤ 60 mV/decade → NCFET , TFET

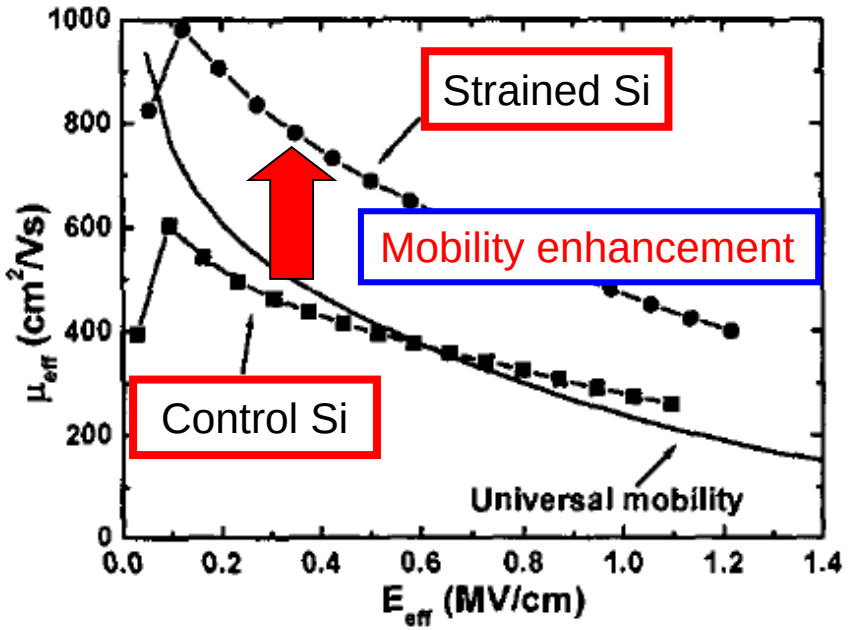
Moore's Law (scaled FETs)

- Intel co-founder Gorden Moore noticed in 1964
- Transistor density on a chip doubles per generation 2x/generation
- Amazingly still correct, likely to keep until 20xx

MORE Than MOORE: **Sensor/MEMS, Analog, RF, …**

Strained Si at room temperature (90 nm)

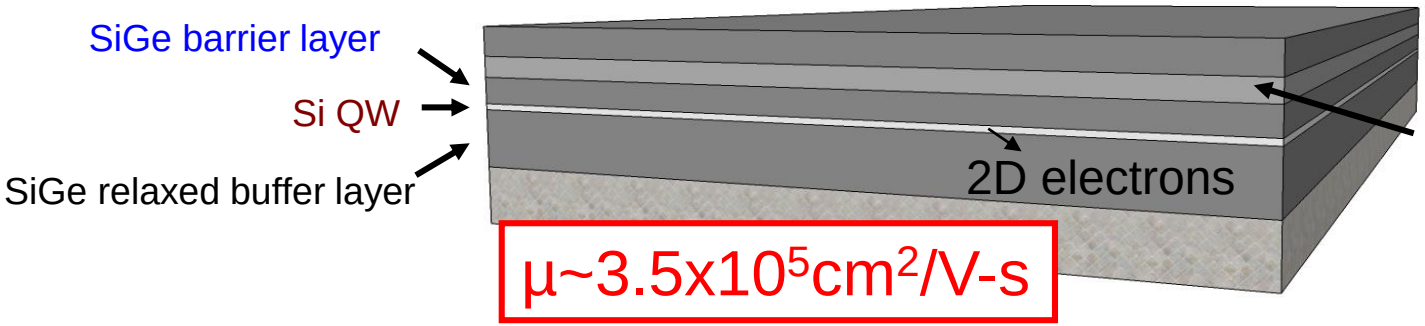




M. H. Lee et al., IEDM, 69 (2003).

- The tensile strain lowers Δ2 valleys with low in-plane effective mass.
- More electron in high mobility Δ2 valleys
- ~2x enhancement

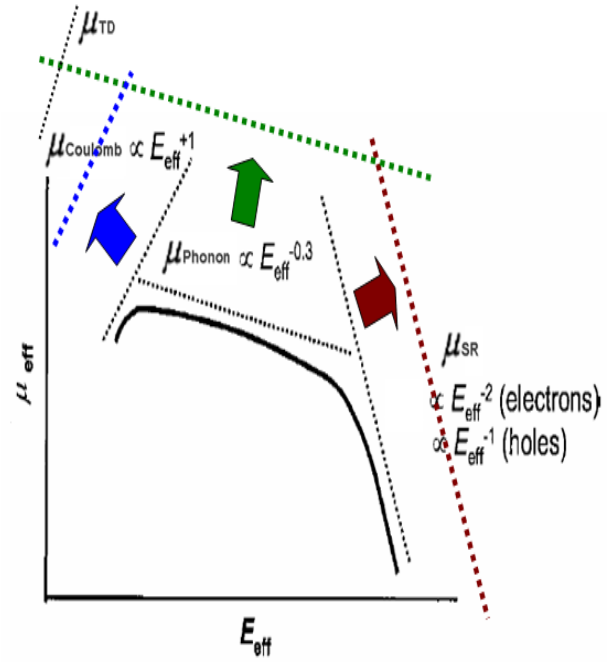
High mobility strained Si



modulation doped layer

Ref: F. Schaffler et al., Semicond. Sci. Tech., 12, 1515 (1997).

Our approach



- *To reduce Coulomb scattering:*

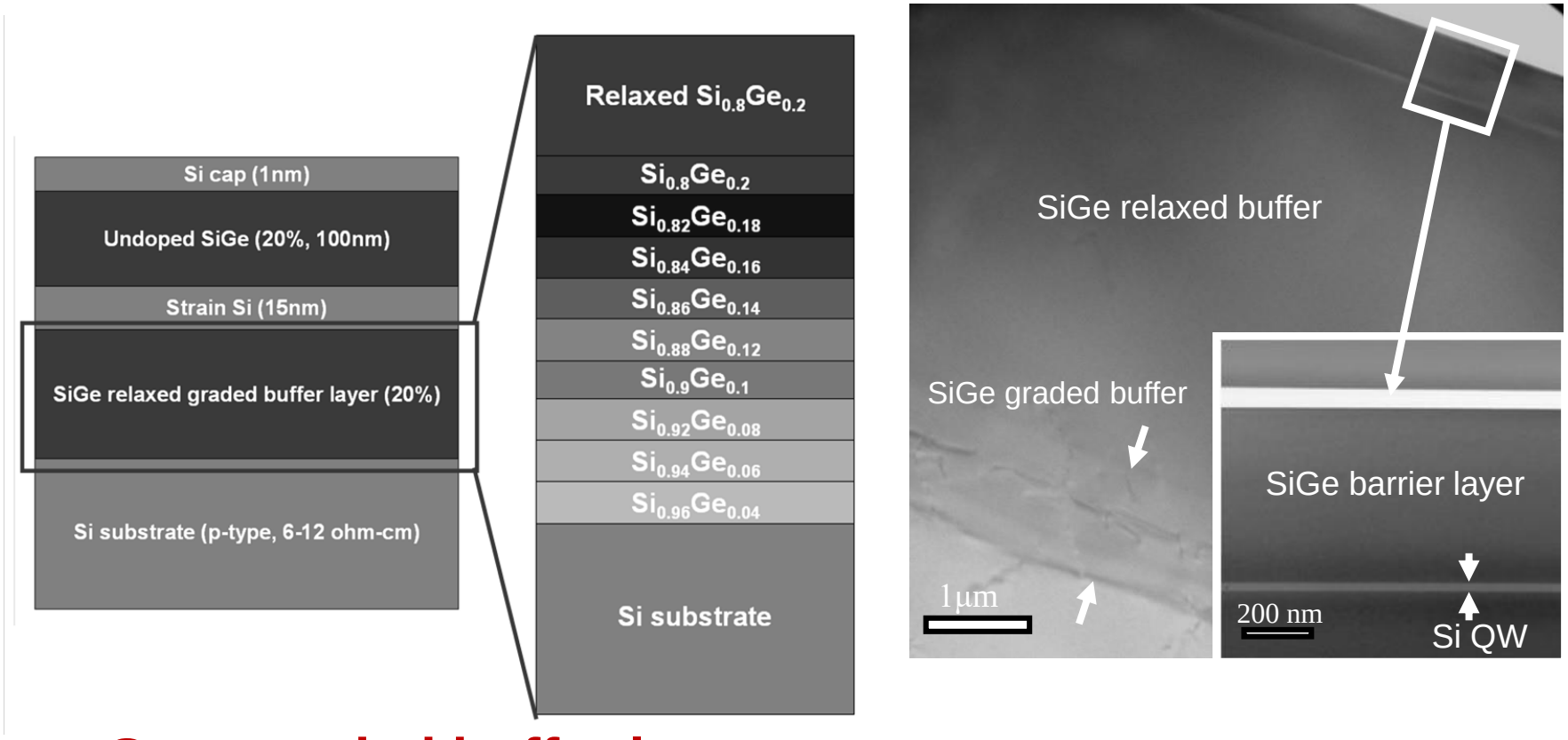
1. No doping in the SiGe/Si/SiGe heterostructure.

2. Thick SiGe barrier layer. To prevent remote charged impurity scattering from gate stack

- *To reduce phonon scattering:*

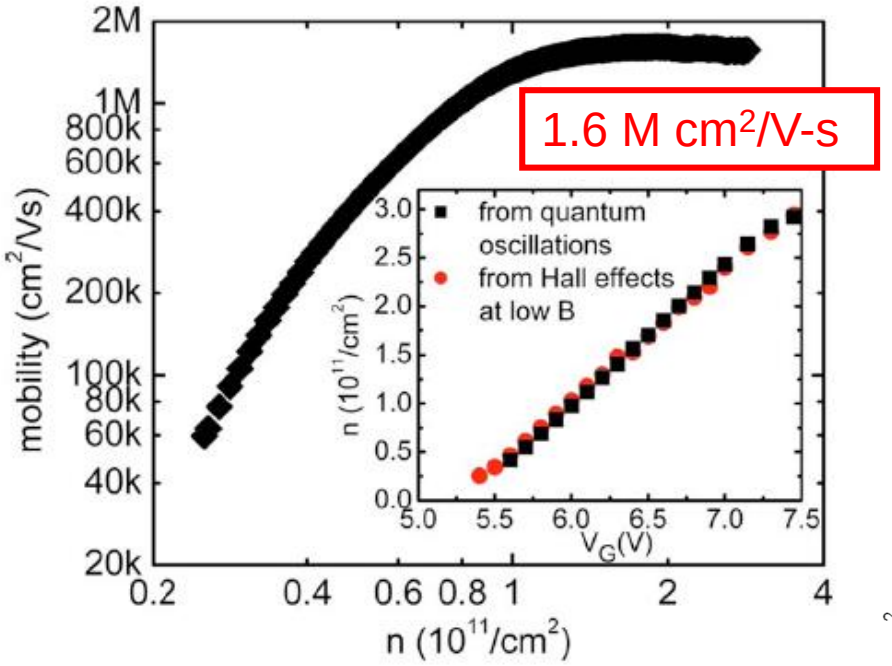
2DEG mobility is measured at 0.3K.

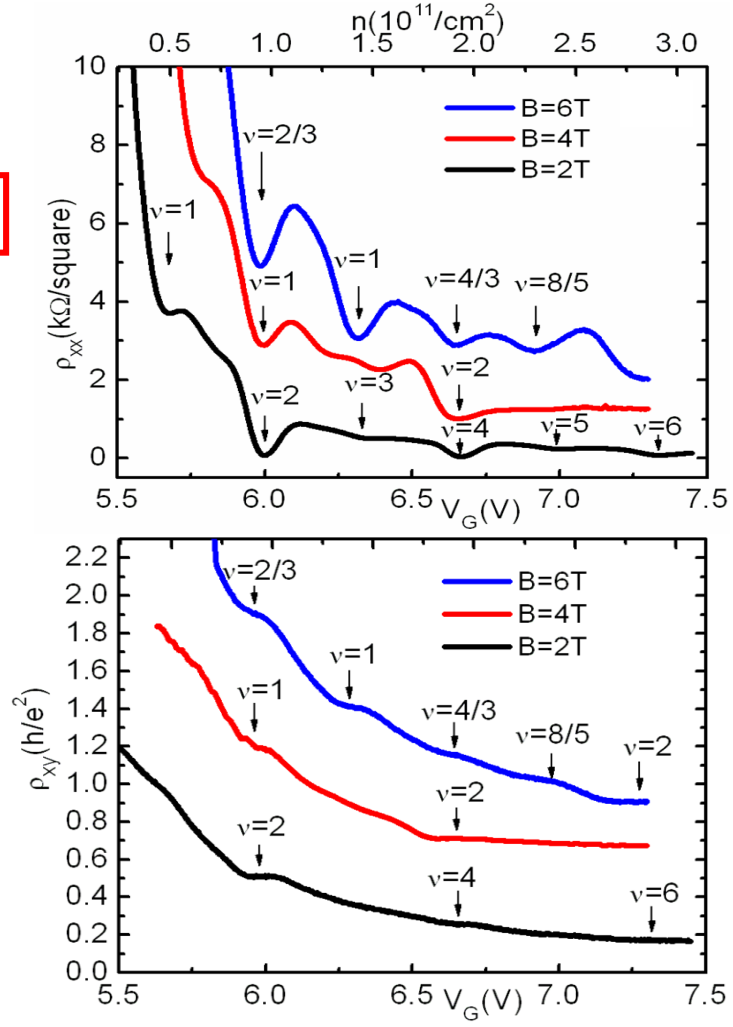
SiGe Graded Buffer Layers



- **Step graded buffer layer**
- **Dislocations confined in the compositionally graded buffer layer.**

Quantum Hall Effect

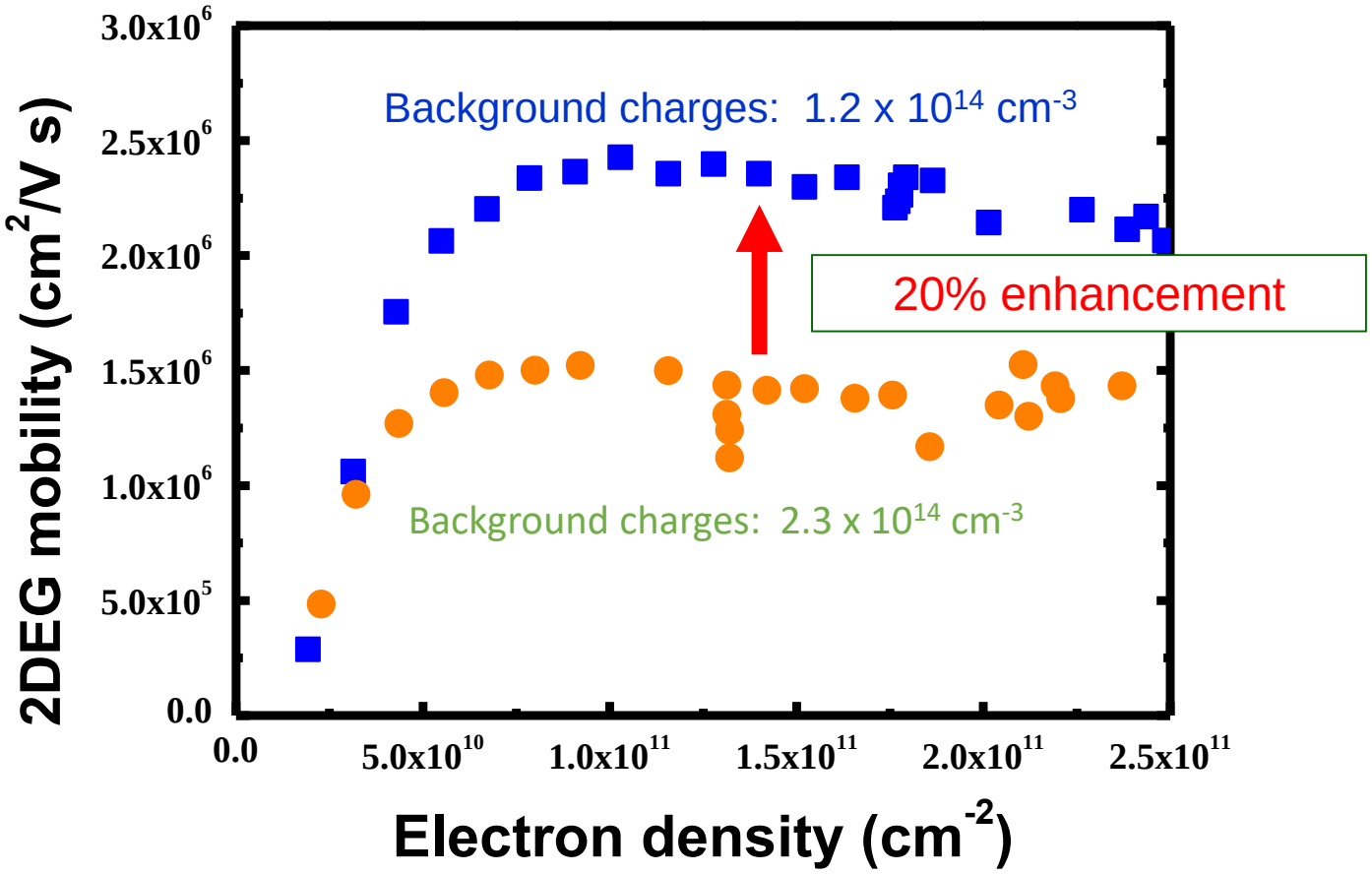




T. M. Lu et al., APL, 94, 182102 (2009).
Collaborate with Prof. D. C. Tsui, Princeton Univ.

- Integral and fraction quantum Hall Effect can is observed in the 2DEG structure.

Record High 2DEG Mobility (2.4 M $cm^2/V\,s$)



M. Yu. Melnikov et al., APL, 106, 092102 (2015).

Collaborate with Prof. S. V. Kravchenko, Northeastern Univ.

- 2.4×10^6 $cm^2/V\,s$ by the reduction of background charges from 2.3 x 10^{14} cm^{-3} to 1.2 x 10^{14} cm^{-3}.

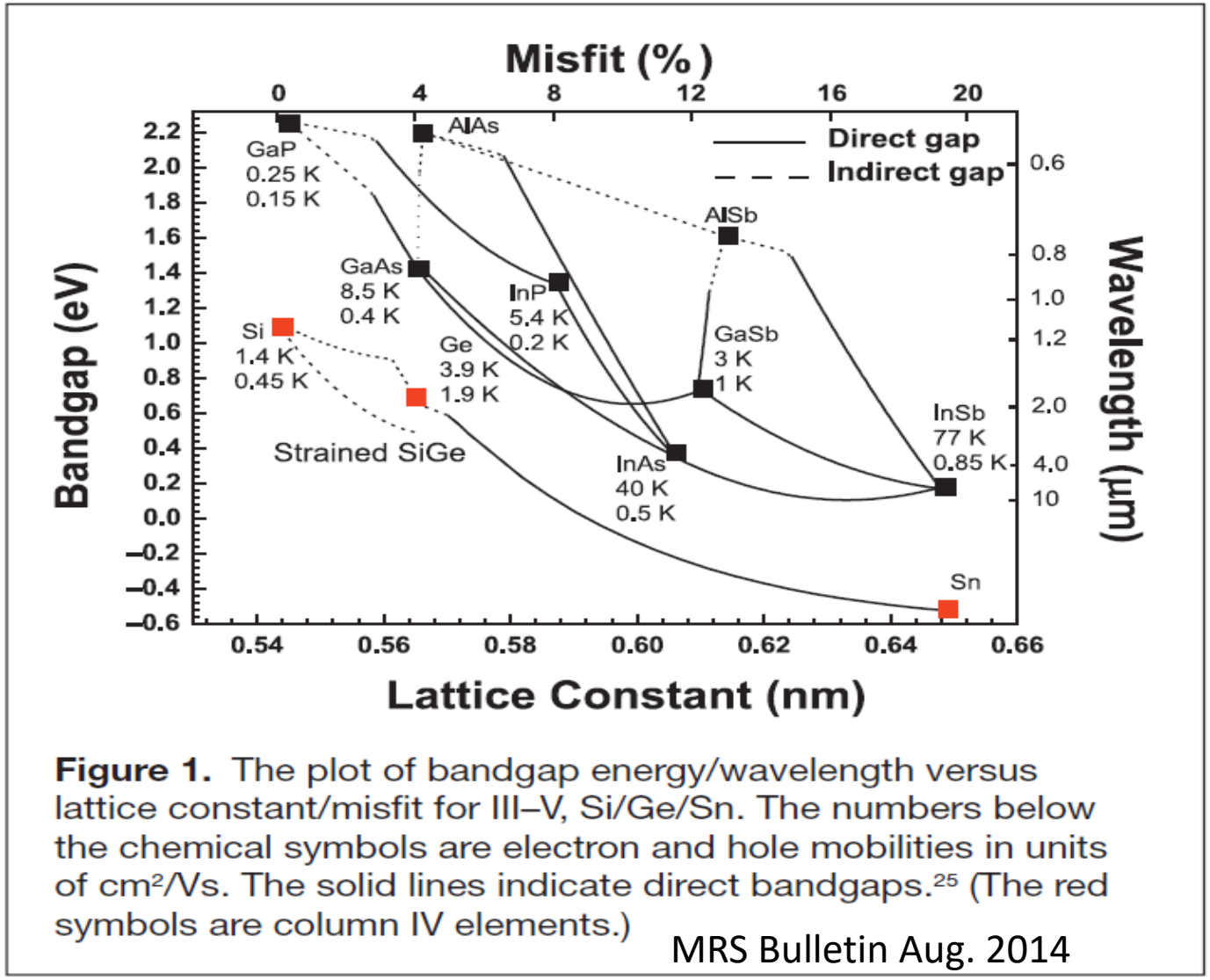
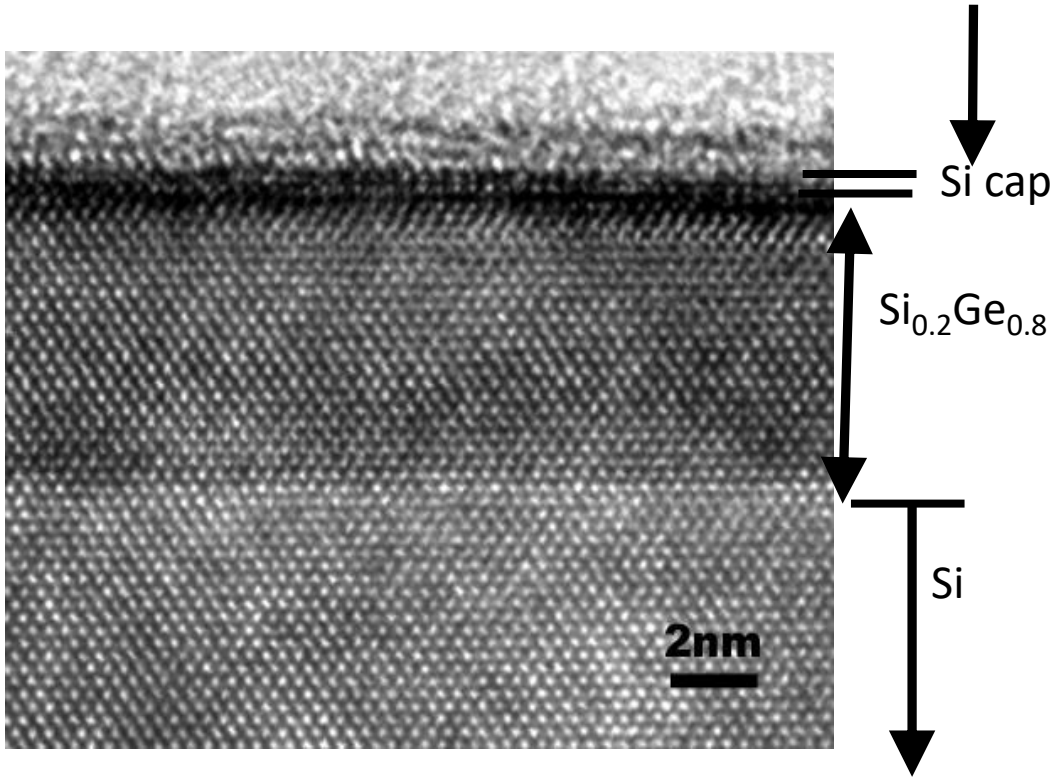


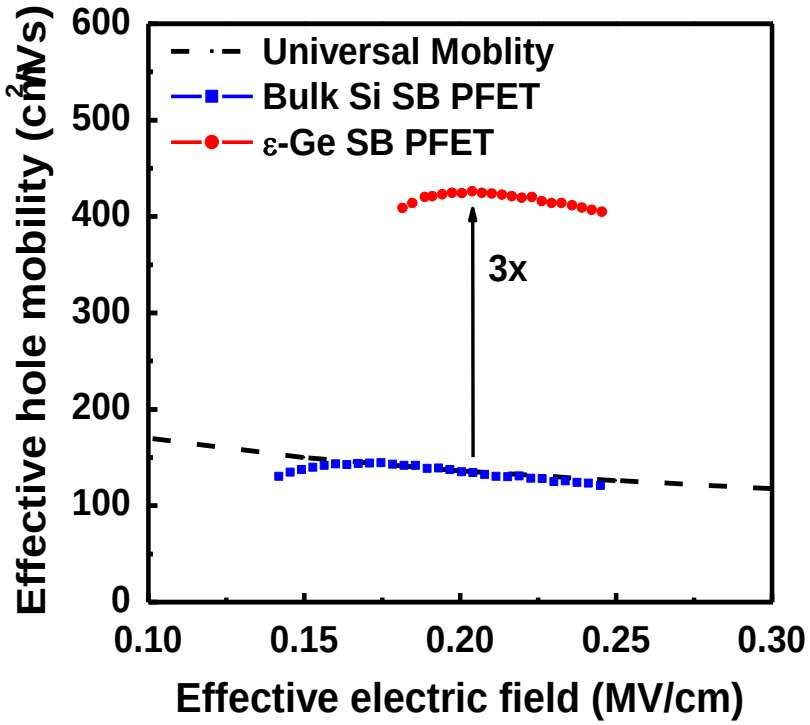
Figure 1. The plot of bandgap energy/wavelength versus lattice constant/misfit for III–V, Si/Ge/Sn. The numbers below the chemical symbols are electron and hole mobilities in units of cm^2/Vs. The solid lines indicate direct bandgaps.[25] (The red symbols are column IV elements.)

MRS Bulletin Aug. 2014

- **SiGe/Ge has higher mobility than Si**
- **GeSn mobility is not known!**

SiGe channel PFET using Si cap

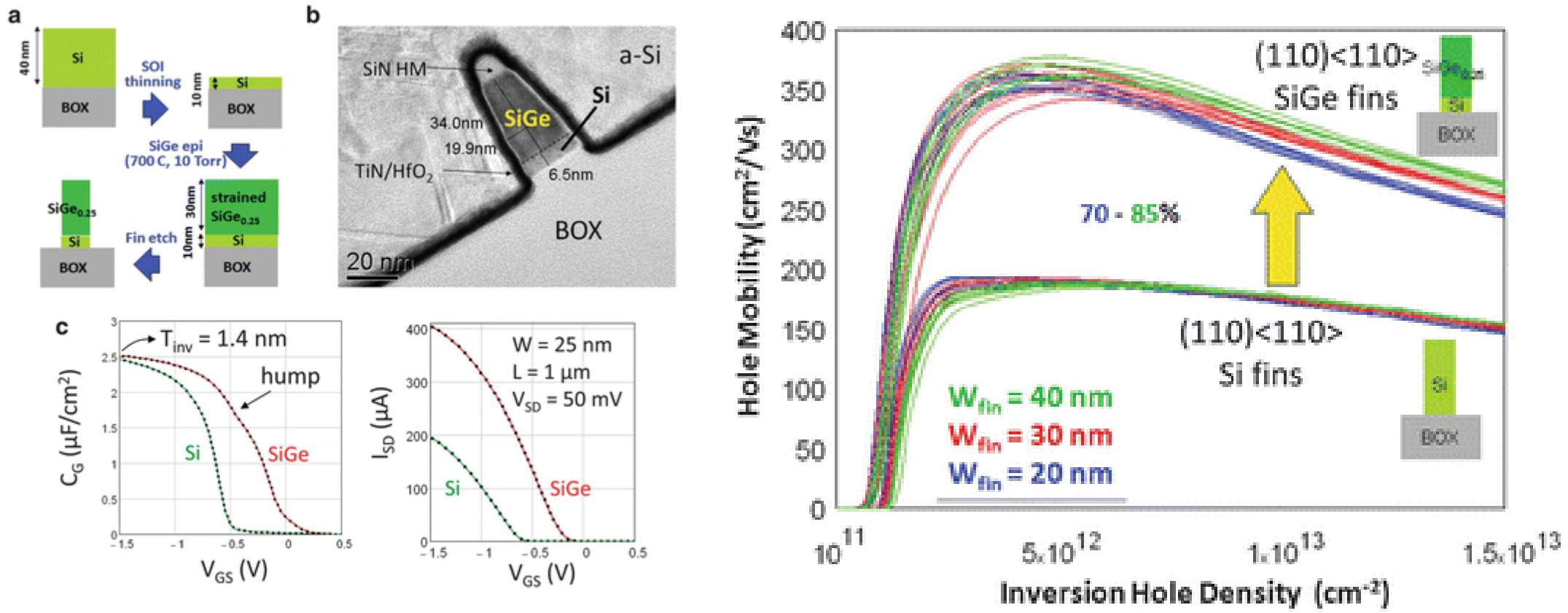




- Si cap for high K/metal gate
- 3x mobility enhancement

C.Y. Peng APL 2007, ROC patent 2006

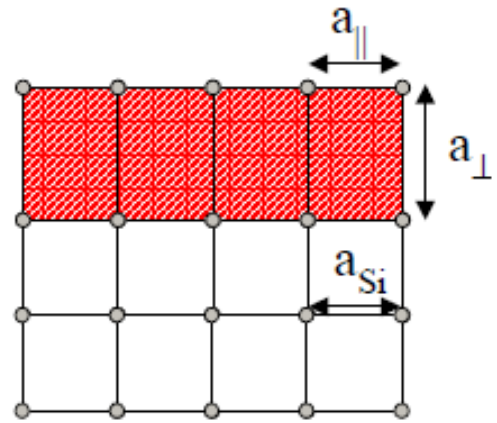
Strained $Si_{0.75}Ge_{0.25}$ channel pFET

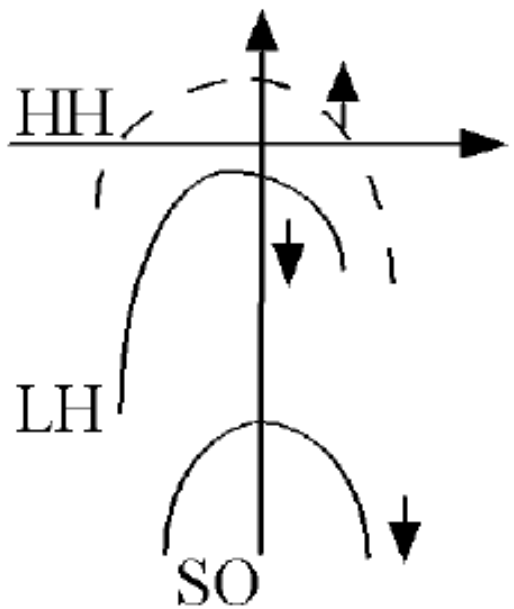


- A substantial improvement of 70–85 % with SiGe fins is noticed at high carrier density of 10^{13} cm^{-2}.

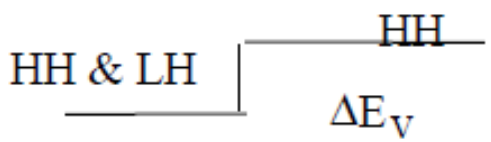
Ref: Toward Quantum FinFET pp 55-79

Electronic Properties of Strained $Si_{1-x}Ge_x$



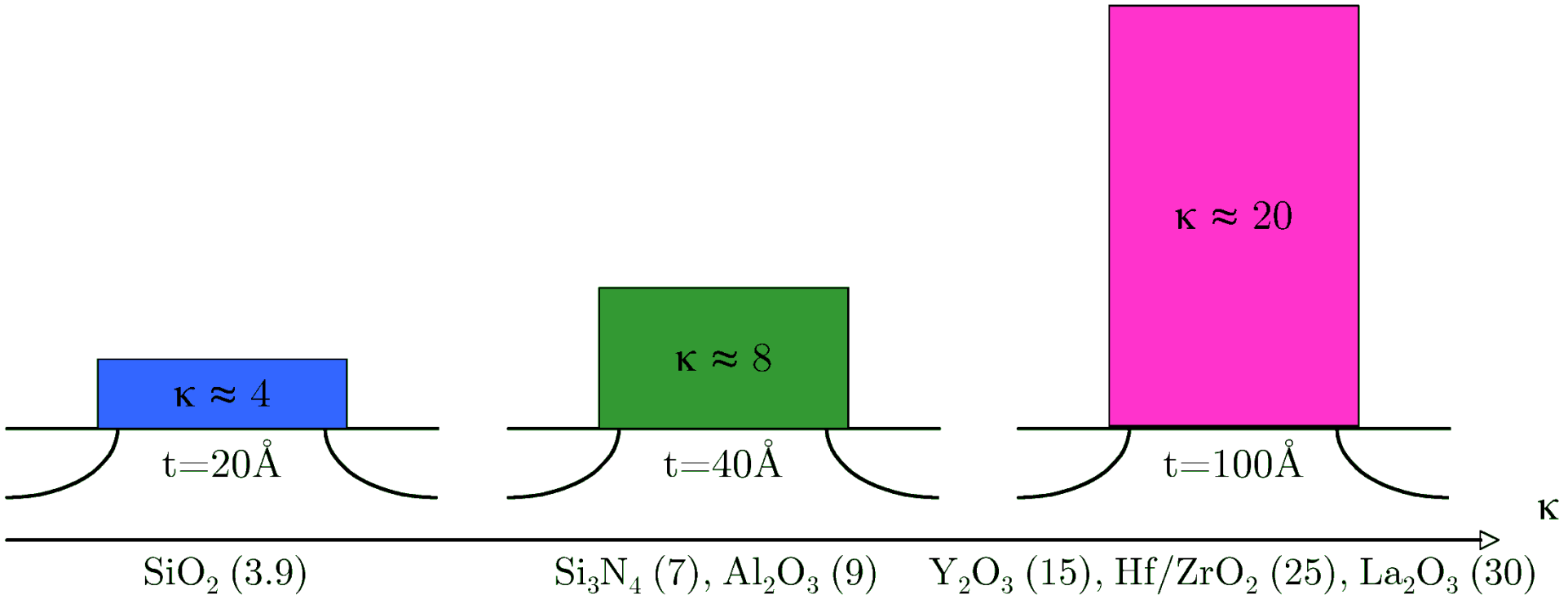


Relaxed Si | Strained $Si_{1-x}Ge_x$



Type-I (Compressive)

Effects of High-k



$$I_{on} = \frac{\mu_{eff}\, C_{ox,\,inv}}{2}\frac{W}{L}(V_{gs} - V_T)^2, \text{ where } C_{ox,\,inv} = \frac{\kappa\varepsilon_0 A}{T_{inv}} \quad (1)$$

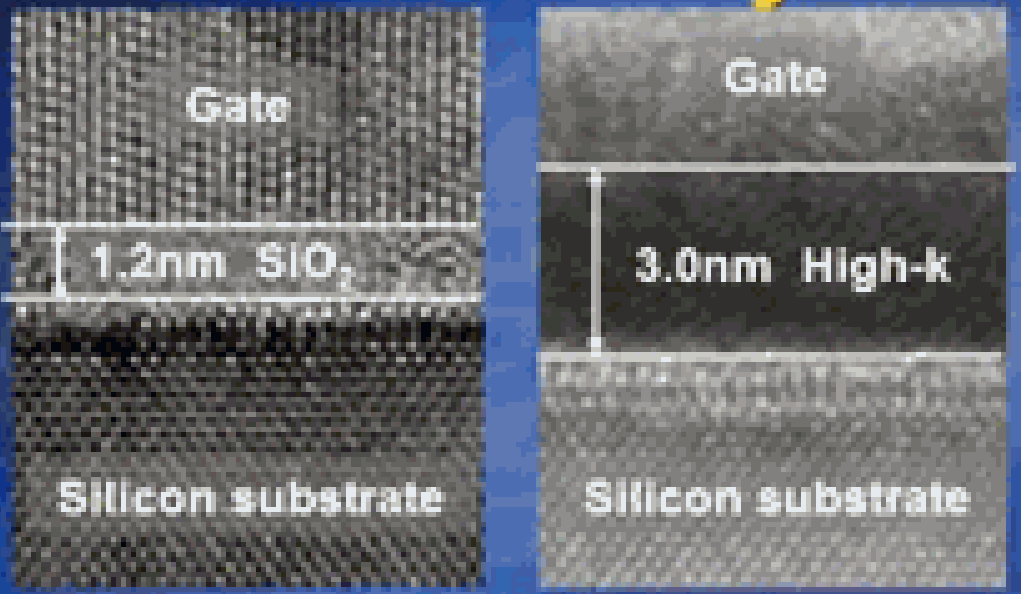
High-k

Metal Gate

$$EOT = t_{\text{high-k}}\left(\frac{k_{SiO_2}}{k_{\text{high-k}}}\right)$$

- k ↑, $C_{ox,\,inv}$ ↑, I_{on} ↑
- k ↑, $t_{high\text{-}k}$ ↑, leakage ↓

High-k Dielectric reduces leakage substantially



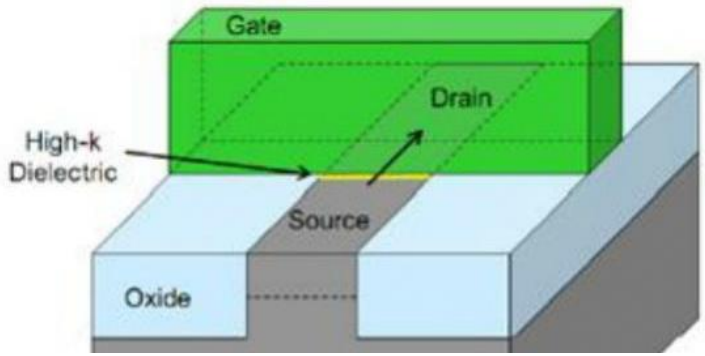
Benefits compared to current process technologies

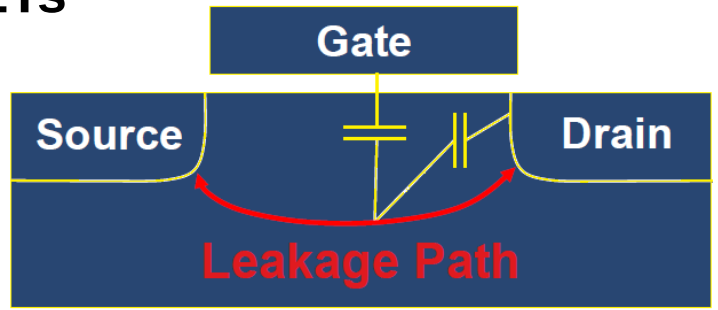
	High-k vs. SiO_2	Benefit
Capacitance	60% greater	*Much faster transistors*
Gate dielectric leakage	> 100x reduction	*Far cooler*

intel

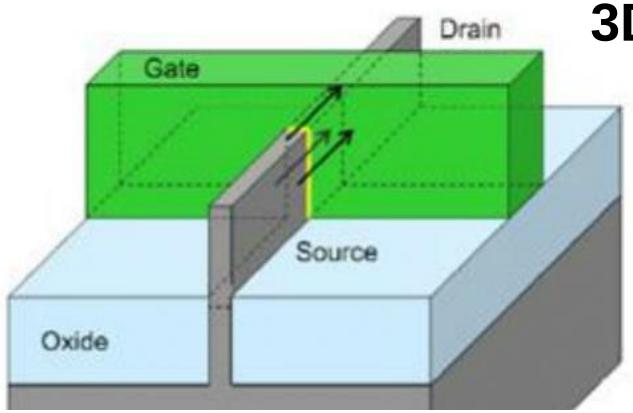
3D Transistors to Reduce SS

Planar FETs

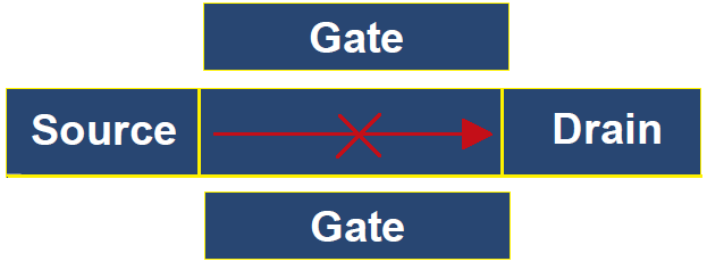




3D transistors



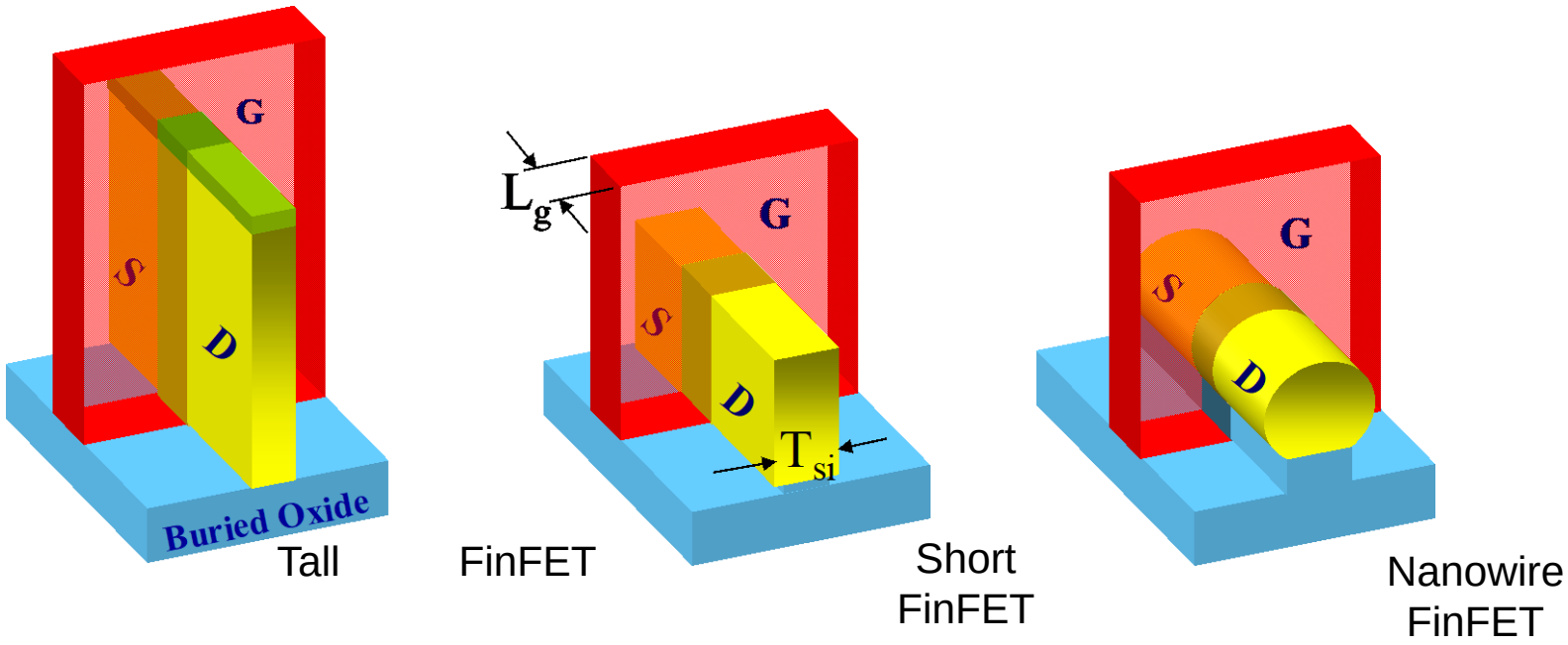
Source: Intel



- **Current can NOT go through at Vgs=0V.**
- **SS approaches 60 mV/dec (70 mV in real devices)**

ref: C.Hu, “Modern Semicon. Devices for ICs” 2010, Pearson

Variations of FinFET



- **Tall FinFET** has the advantage of providing a large W and therefore large I_{on} while occupying a small footprint.
- **Short FinFET** has the advantage of less challenging lithography and etching.
- **Nanowire FinFET** gives the gate even more control over the silicon wire by surrounding it.

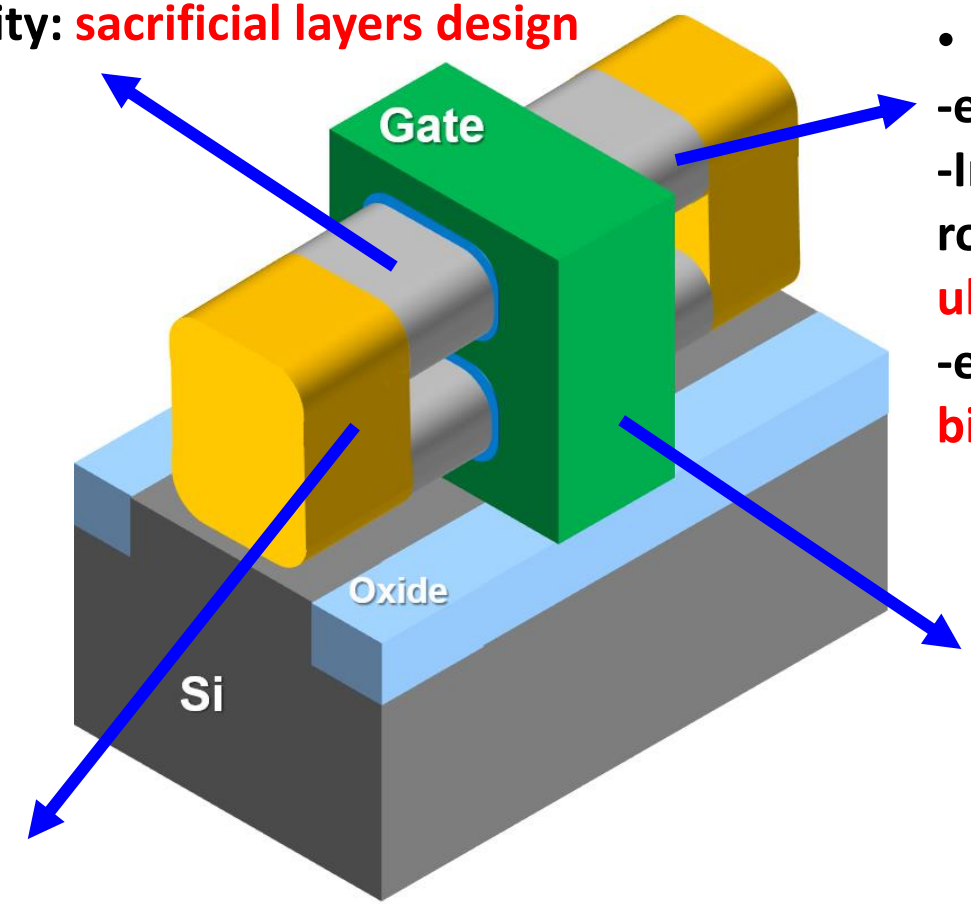
Beyond FinFET: stacked GAA

- Epi

-strain engineering: fully compressive strain for stacked GeSn channels

-defect and dislocation: defect confinement at Ge/SOI interface

-inter-channel uniformity: sacrificial layers design



- Channel release

-etching selectivity: H_2O_2 etching

-Inter-channel uniformity & line edge roughness: sacrificial layers etching with ultrasonic assist technique

-effective mass & strain after channel release: biaxial to uniaxial strain & microbridge effect

- S/D:

-high doping

-low parasitic resistance

➔ Optimization for *in-situ* doped GeSn S/D and PtGeSn/PtGe formation to decrease parasitic resistance are achieved.

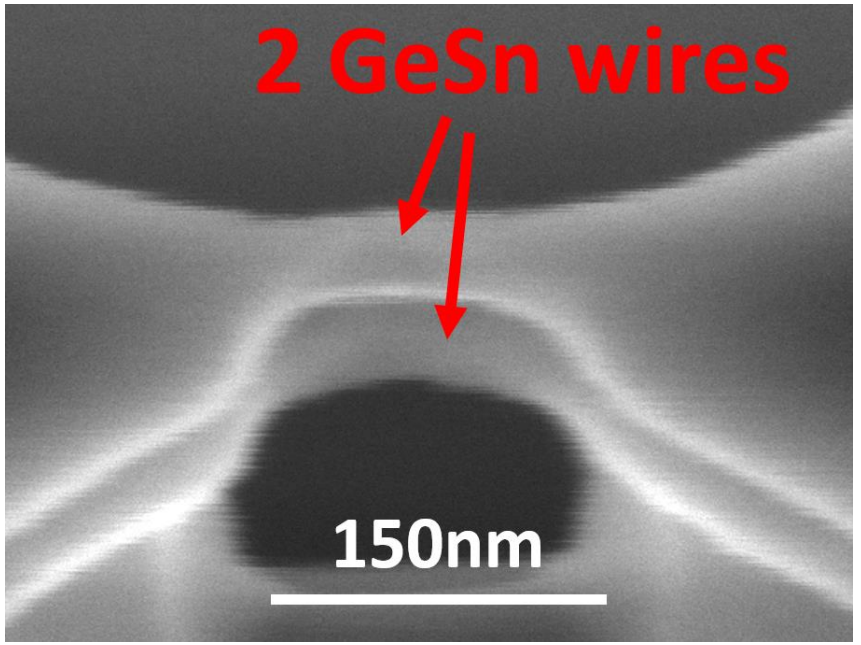
- Low thermal budget gate stacks

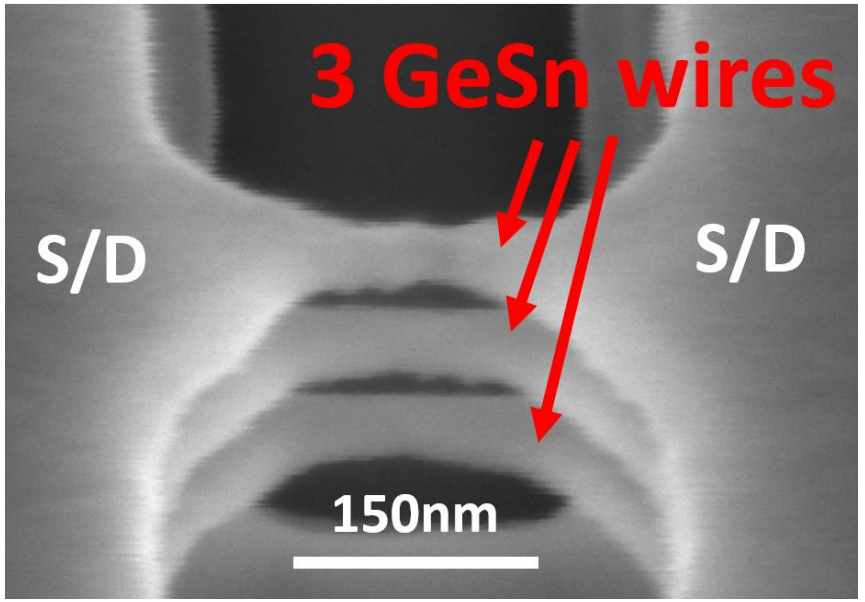
-low D_{it} of IL, high-k material, and low dispersion: TiN/ZrO_2/Al_2O_3+RTO with 400°C thermal budget

Device fabrication

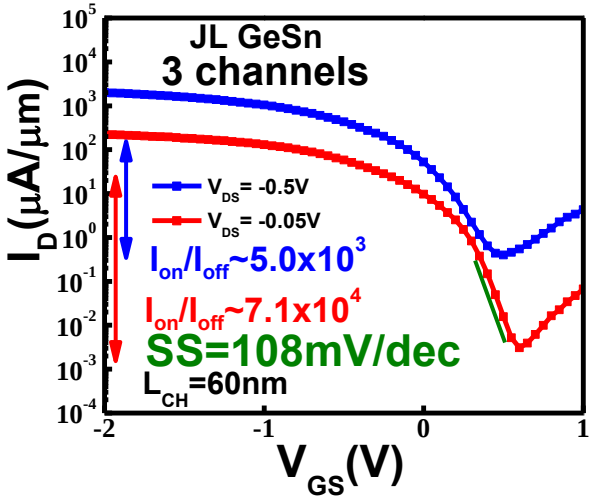
Process flow

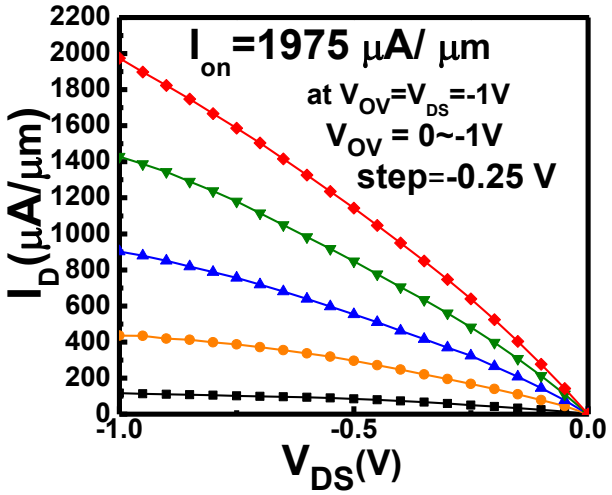
- CVD epitaxy on 200mm SOI
- Fin formation by E-beam patterning and anisotropic etching (pure Cl_2 or HBr)
- Channel release (H_2O_2(wet)) along <110>
- Gate dielectric and in-situ electrode formation (ALD @ 250°C)
  Al_2O_3+RTO 400°C +ZrO_2 + *in-situ* TiN
- Gate stack annealing (FGA 400°C 10min)
- Gate metal pad (TiN), S/D contact formation by sputtering (Pt), and PMA at 400°C

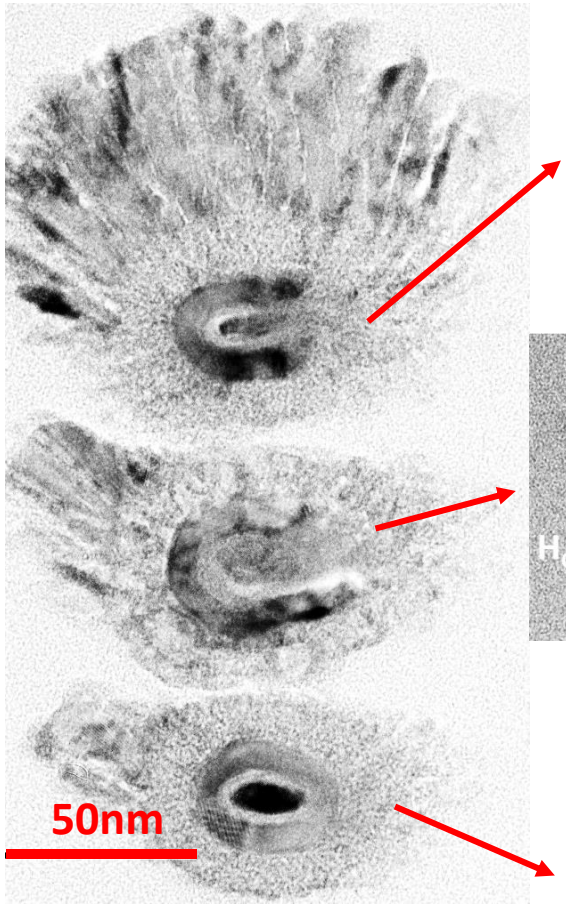


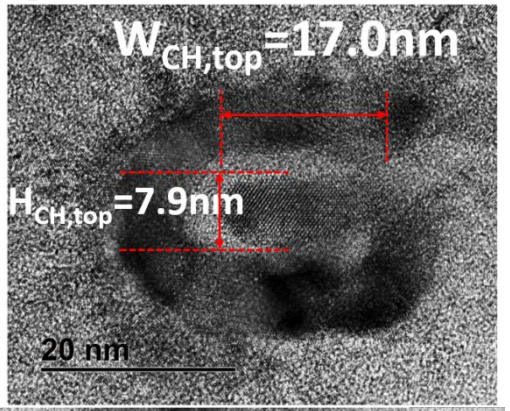


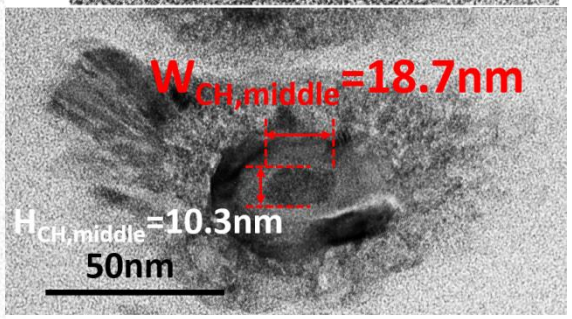
Stacked 3 channels GeSn JL pGAAFETs with optimum S/D

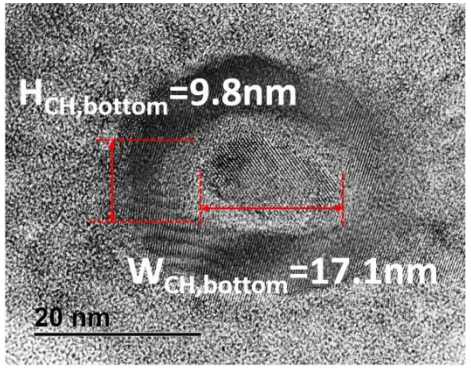


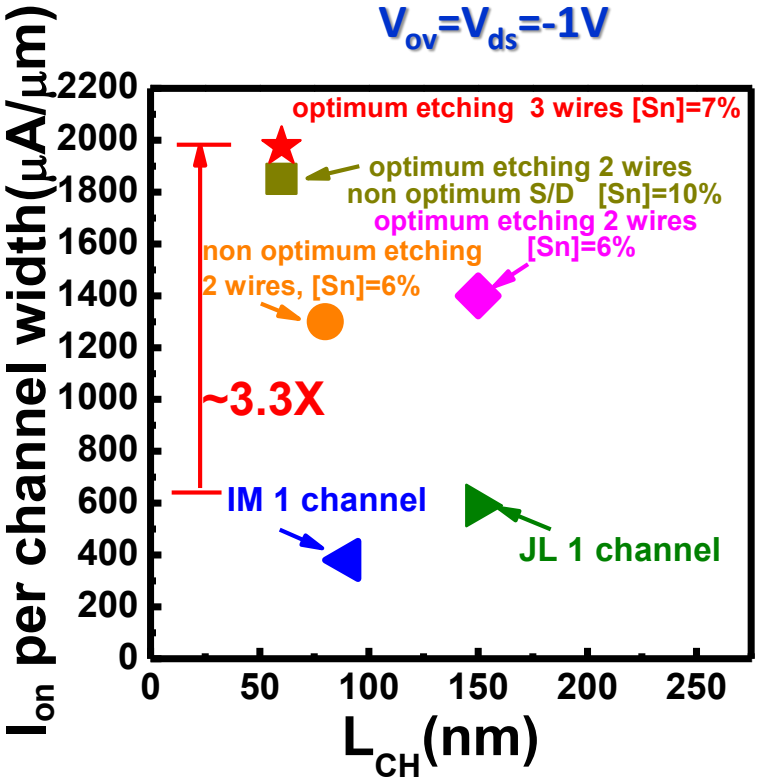












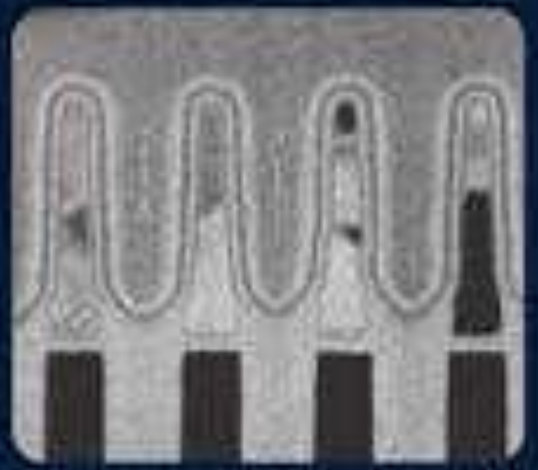
- Parasitic resistance is reduced by optimum S/D.

➔ I_{on}=1975μA/μm at V_{OV}=V_{DS}=-1V and I_{on}=554μA/μm at V_{OV}=V_{DS}=-0.5V with 3 channels.

Industry Transitioning from 2D to 3D

2D Logic → 3D FinFET

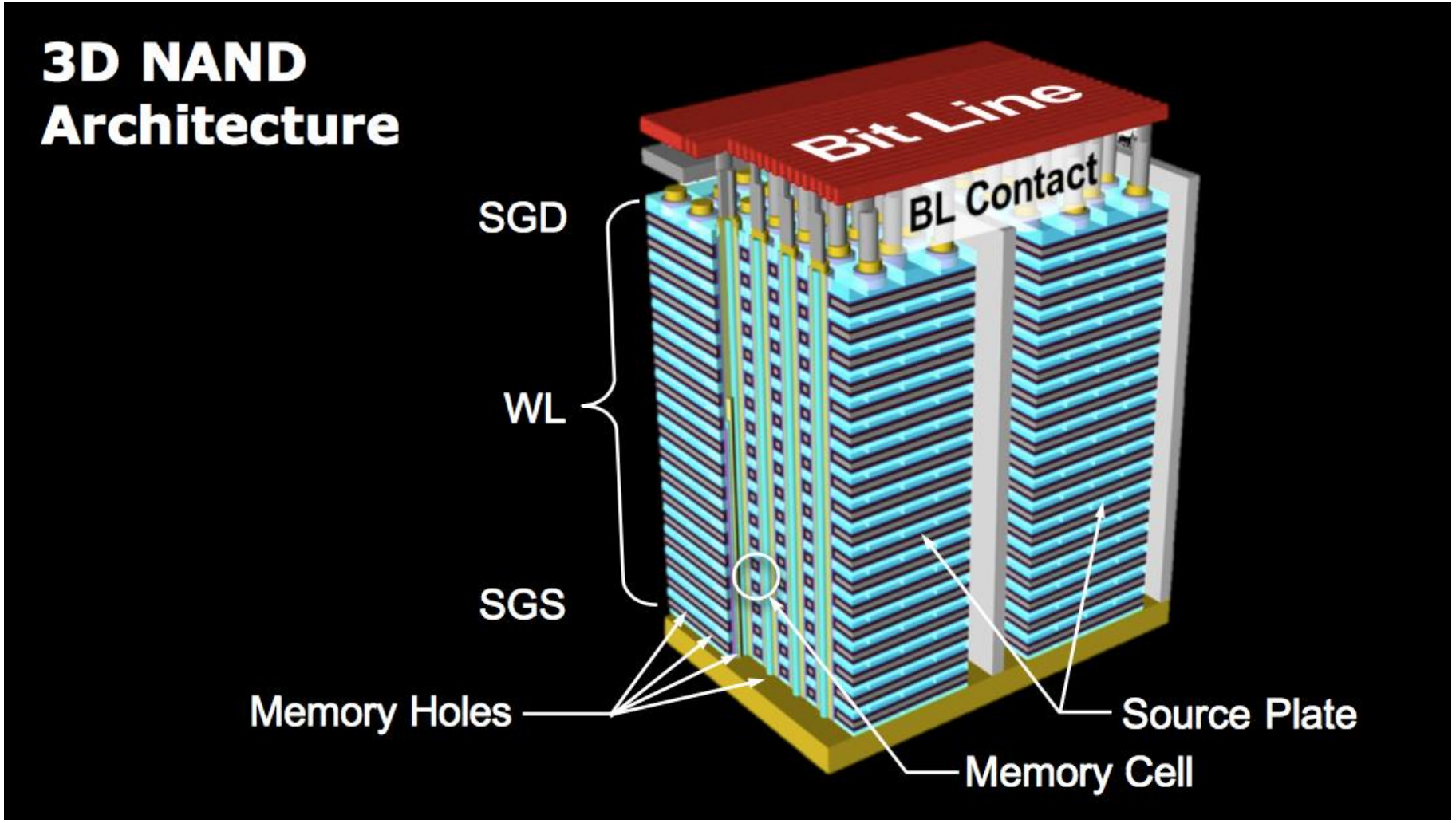
2D NAND → 3D NAND



- Fin and gate size variation can degrade device performance
- 3D NAND reliability depends on device dimension control along deep channels

Source: Applied Materials

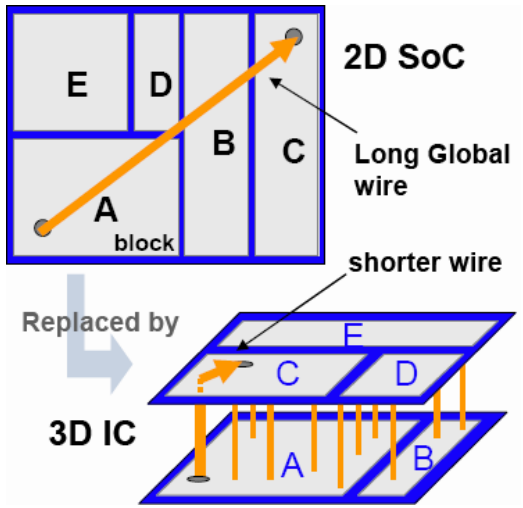
3 D NAND Flash



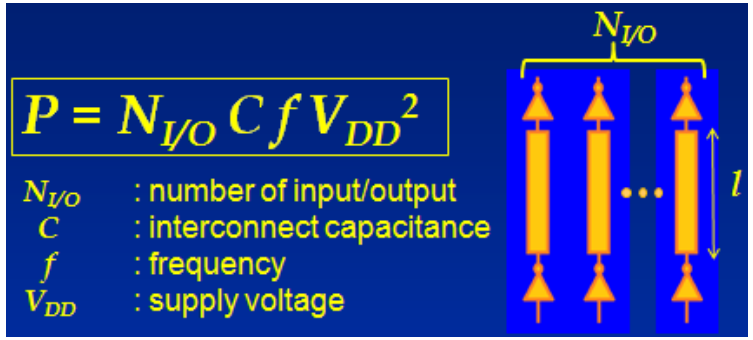
SGD: select gate at the drain end
SGS: select gate at the source end

Source: Western Digital, DevelopEX 2017

Why 3-D IC

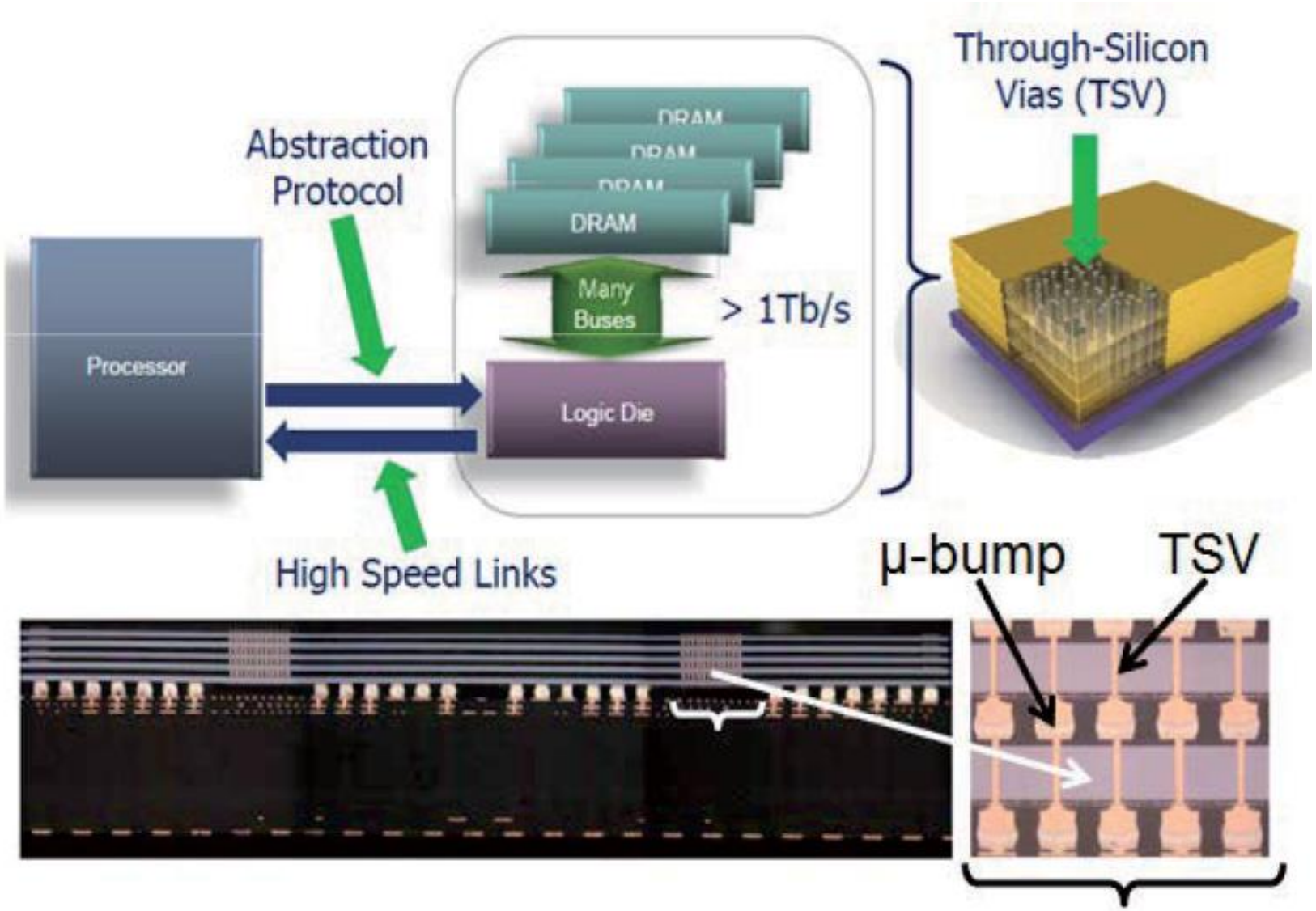


- $P = N_{I/O}\uparrow C\downarrow f\downarrow V_{DD}^2$
  - → **Less power consumption**
- **Hetero-integration**
  - → **More than Moore**
  - → **Cost effective**



- $N_{I/O}\uparrow$ + **Vertical Interconnect**
  - → **Higher bandwidth**
  - → **Size reduction**
  - → **Lower RC delay**

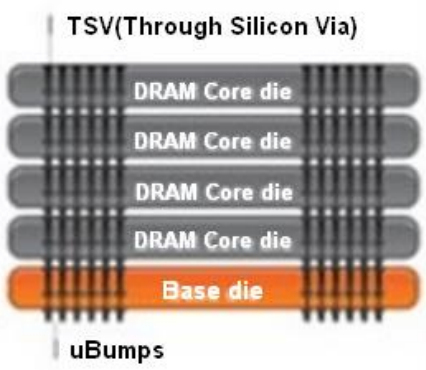
Hybrid Memory Cube

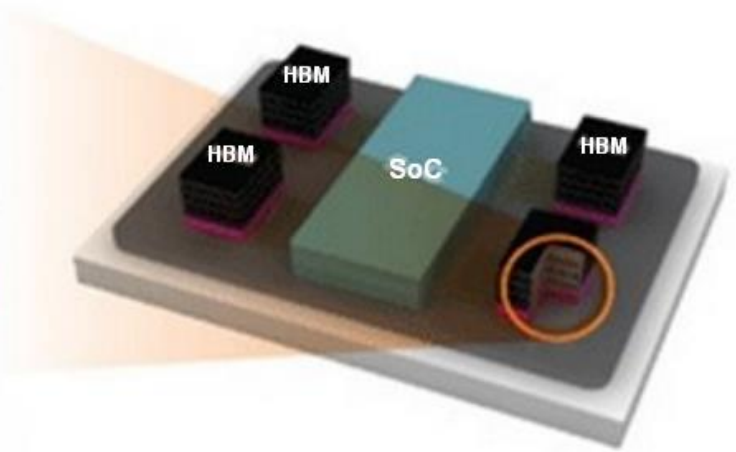


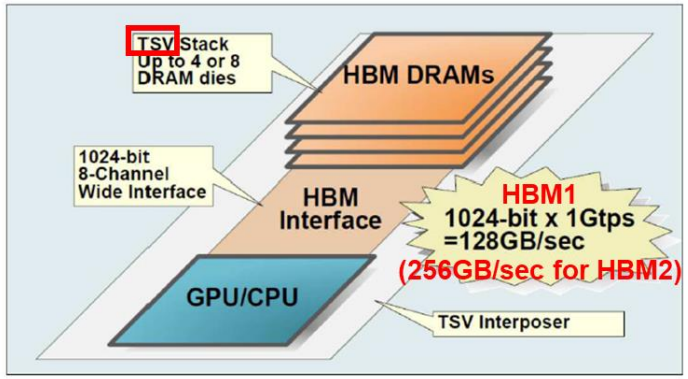
Ref:2017 VLSI SC

High Bandwidth Memory

Source: SK Hynix

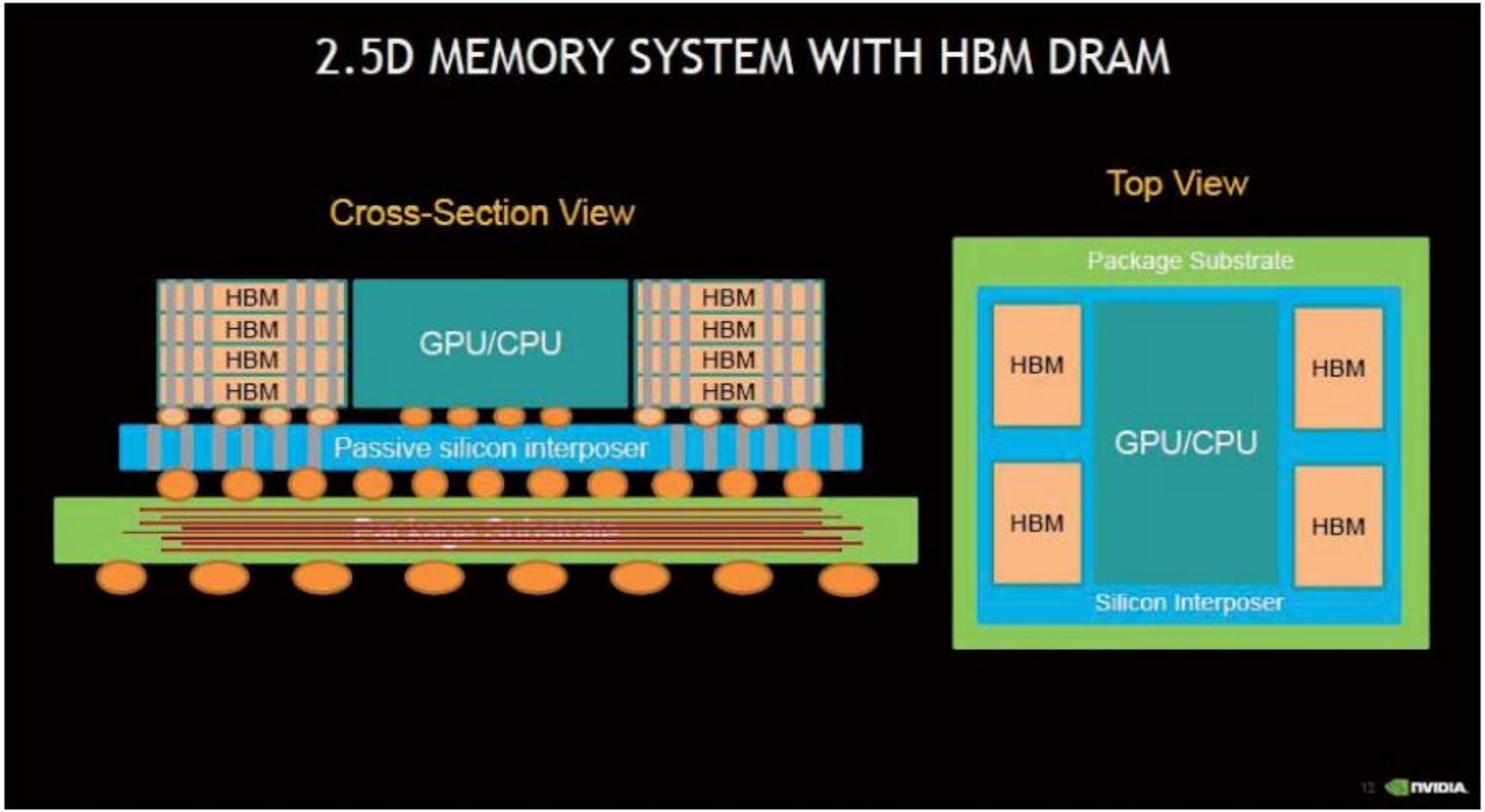






- HBM → 2.5D, memory dies packaged TOGETHER with processor and connected through silicon interposer.
- Stacked DRAM dies, connected by TSVs
- Size reduction vs GDDR5

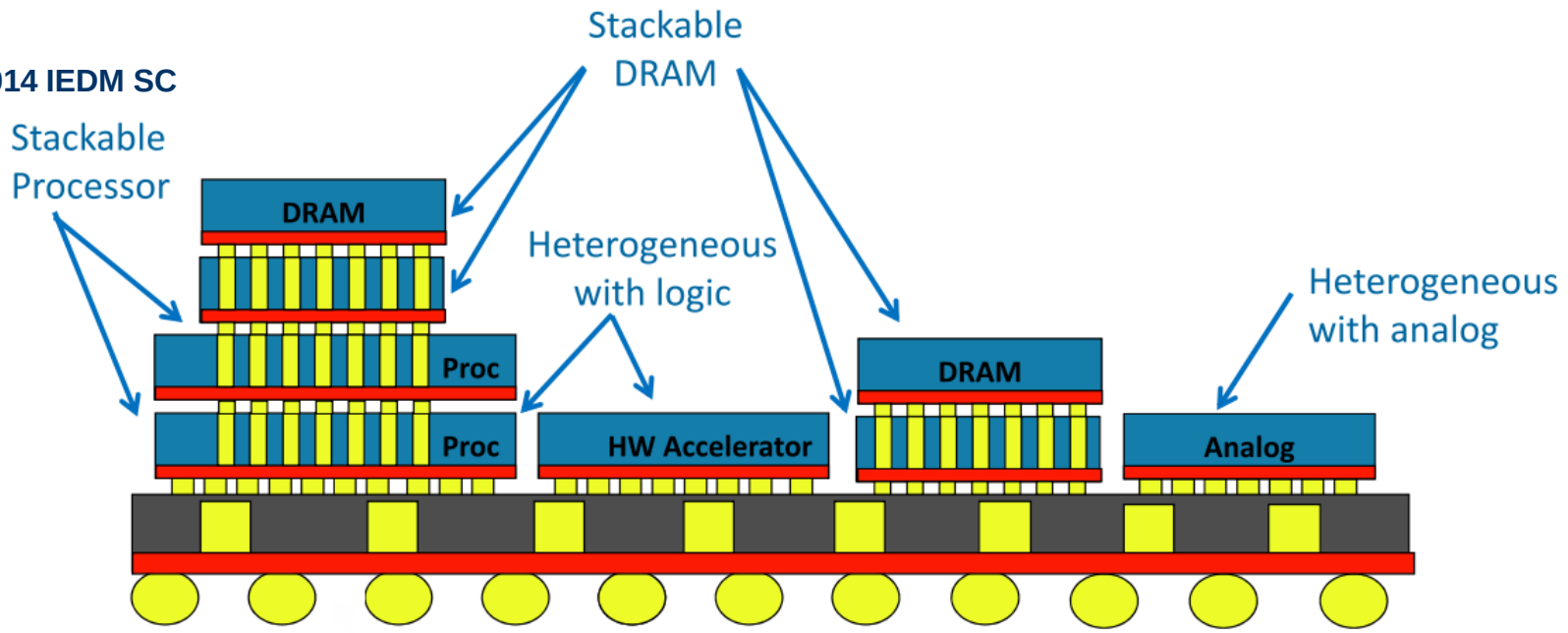
High Density Memory System with HBM on Si Interposer

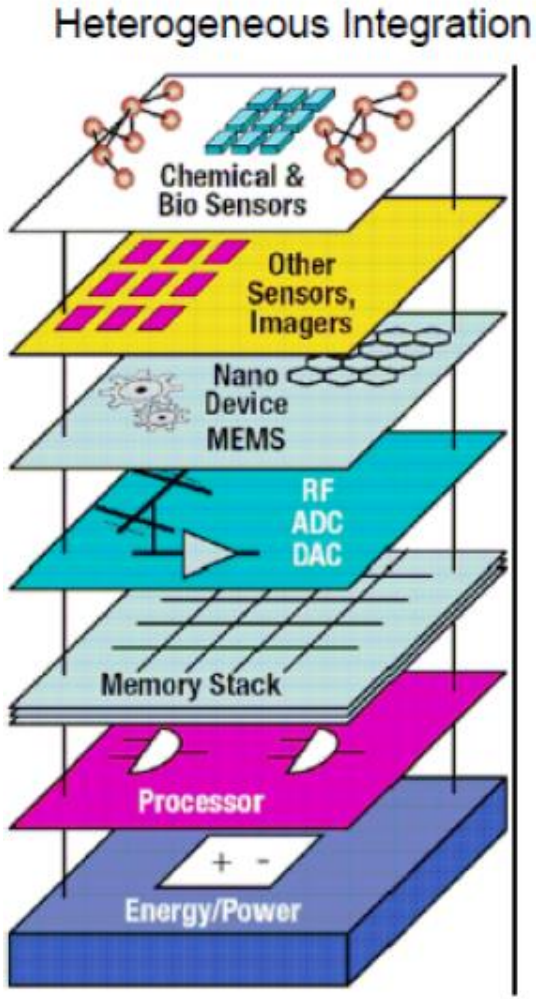


Ref:2017 VLSI SC

Hetero-integration

Ref:2014 IEDM SC

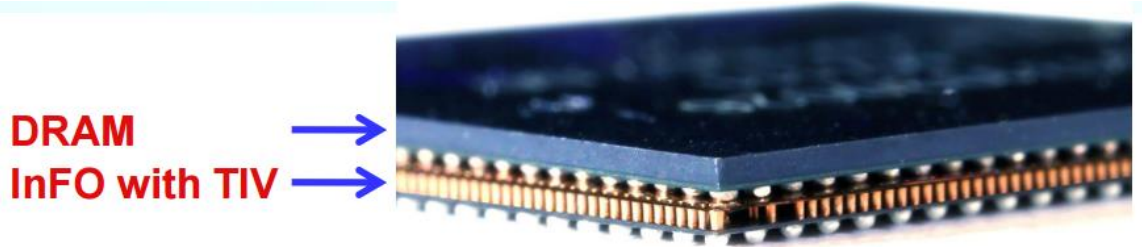


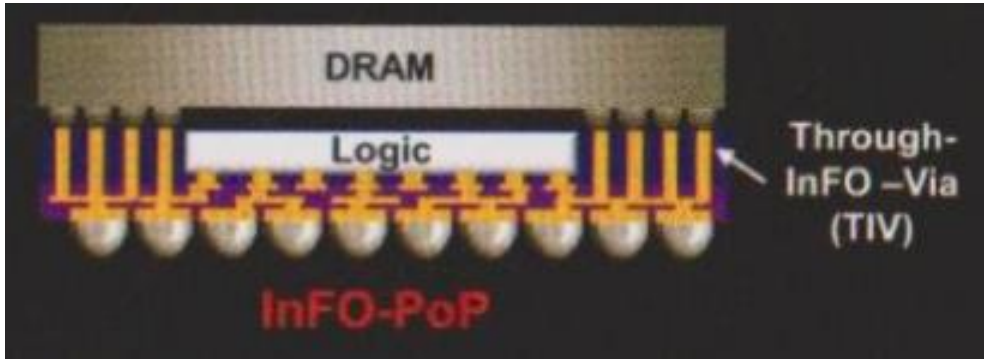


- **3D-IC is the promising technology for heterogeneous integration such as RF, photonics, memories, and logics**

DRAM on Application Processor

Source: TSMC





- Vertical interconnects in molding compound (TIV)
- Stacked DRAM dies on application processor (AP)
- No TSV, no additional substrate → Low cost

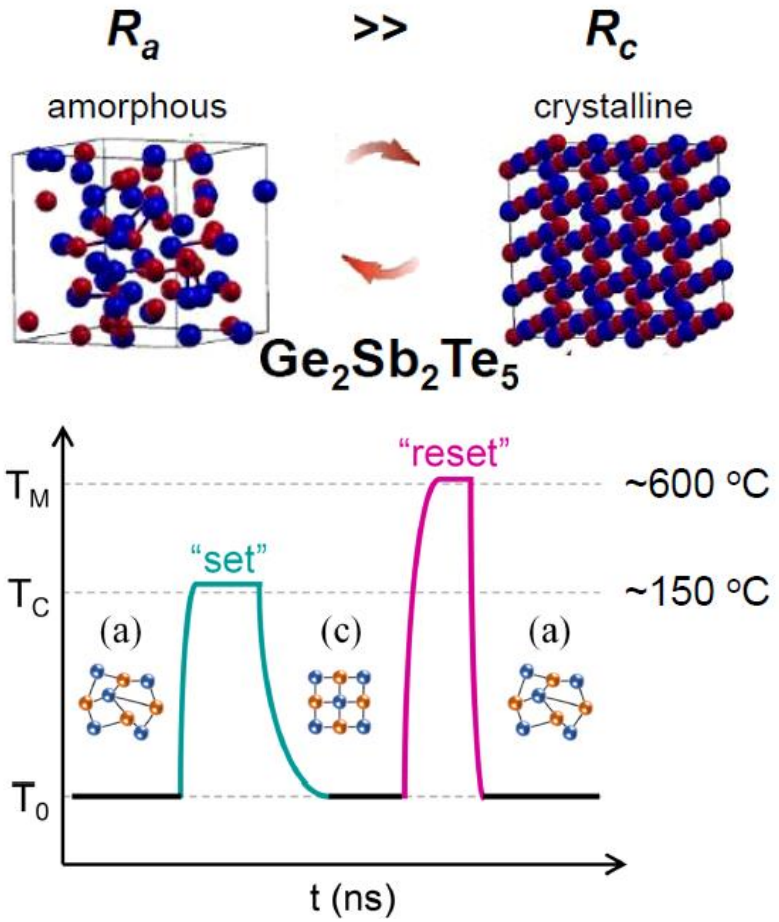
Emerging memory (source: IEEE Solid State circuit Mag. Col.8 no.43, 2016))

TABLE 1. DEVICE CHARACTERISTICS OF MAINSTREAM AND EMERGING MEMORY TECHNOLOGIES.

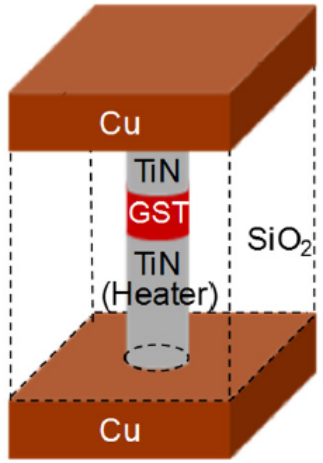
	MAINSTREAM MEMORIES				EMERGING MEMORIES		
			FLASH				
	SRAM	DRAM	NOR	NAND	STT-MRAM	PCRAM	RRAM
Cell area	>100 F^2	6 F^2	10 F^2	<4F^2 (3D)	6~50F^2	4~30F^2	4~12F^2
Multibit	1	1	2	3	1	2	2
Voltage	<1 V	<1 V	>10 V	>10 V	<1.5 V	<3 V	<3 V
Read time	~1 ns	~10 ns	~50 ns	~10 µs	<10 ns	<10 ns	<10 ns
Write time	~1 ns	~10 ns	10 µs–1 ms	100 µs–1 ms	<10 ns	~50 ns	<10 ns
Retention	N/A	~64 ms	>10 y	>10 y	>10 y	>10 y	>10 y
Endurance	>1E16	>1E16	>1E5	>1E4	>1E15	>1E9	>1E6~1E12
Write energy (J/bit)	~fJ	~10fJ	~100pJ	~10fJ	~0.1pJ	~10pJ	~0.1 pJ

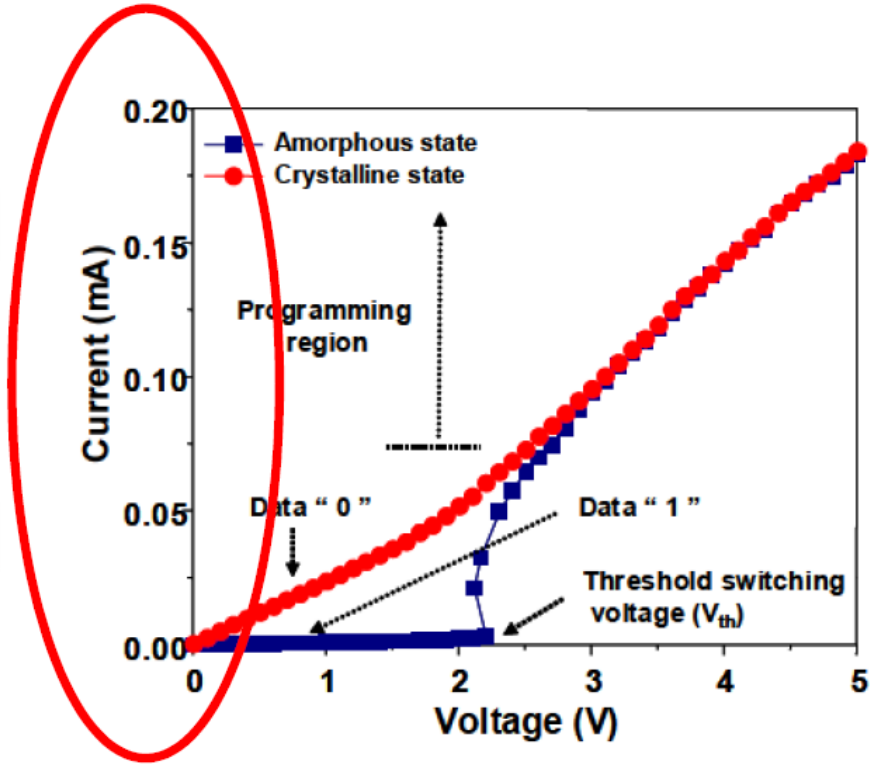
Notes: F: feature size of the lithography. The energy estimation is on the cell-level (not on the array-level). PCRAM and RRAM can achieve less than 4F^2 through 3D integration. The numbers of this table are representative (not the best or the worst cases).

Phase change memory (PCM)



Phase change induced by Joule heat

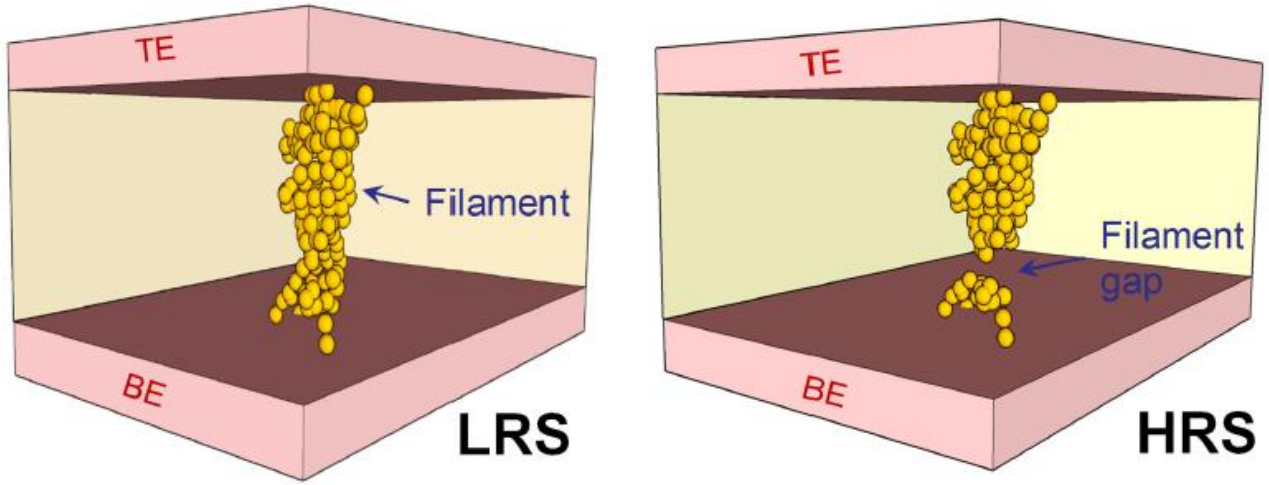


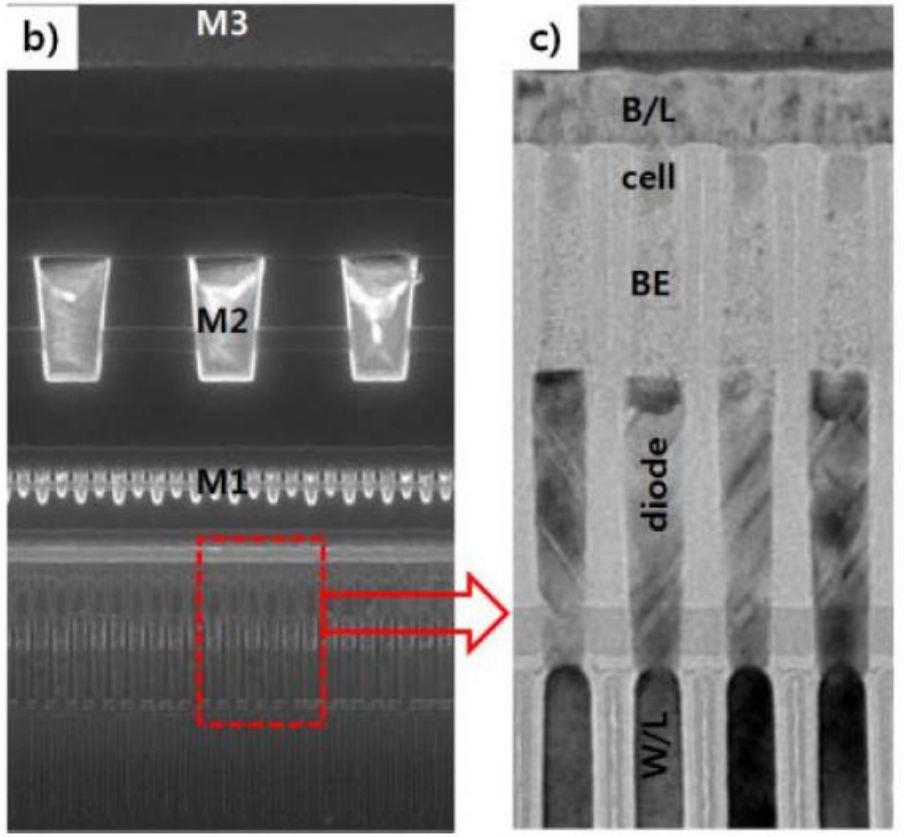


- **Large change in resistance > 100X**
- **But high programming current (~0.1 mA)**

IEDM Tutorial, 2017
Kolobov *et al*., Nature Materials, 2004
Lee *et al*., Nature Nano, 2007
Chen and Pop, TED, 2009

Resistive RAM (RRAM)



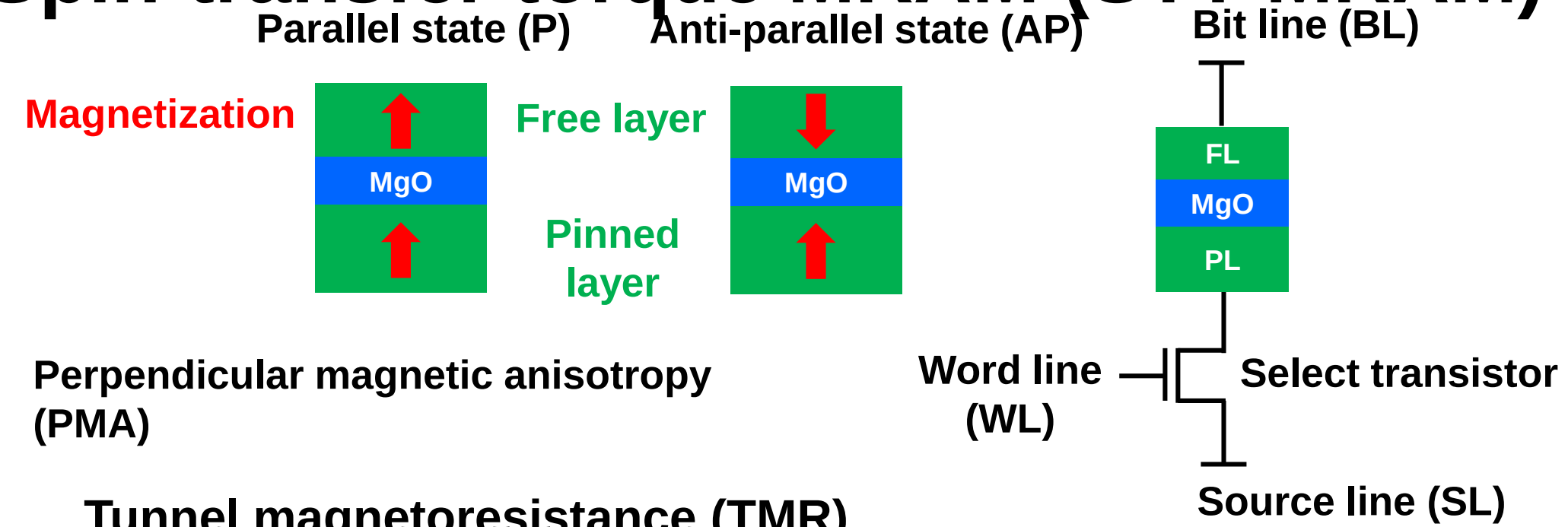


Samsung 8 Gb PCM 20 nm w/ integrated diode (IEDM, 2011)

- **Metal-insulator-metal (MIM) geometry, "I" = e.g. HfO2**
- **The conducting filament is formed by oxygen vacancies or metal precipitates**

H.-S.P. Wong et al., *Proc. IEEE* 100, 1951 (2012)

Spin-transfer torque MRAM (STT-MRAM)

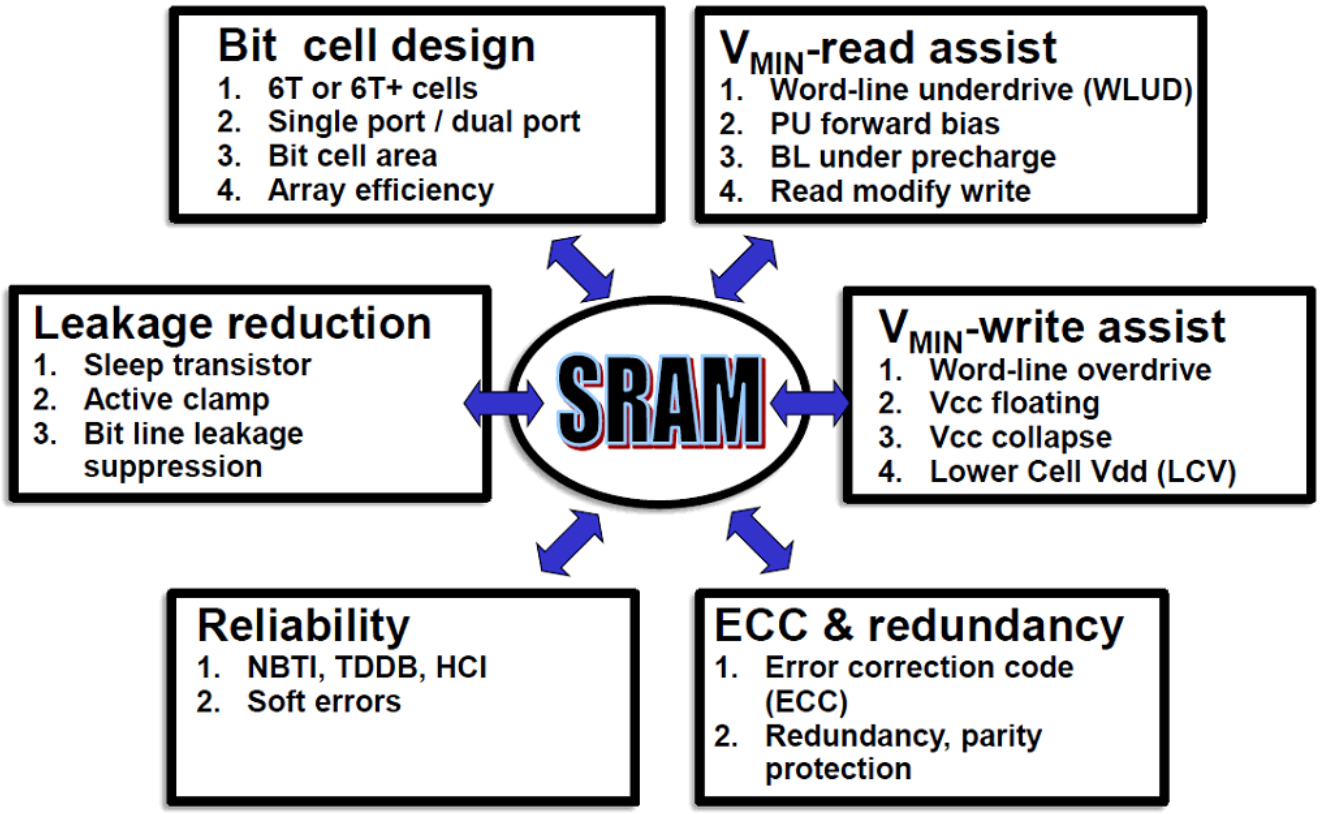


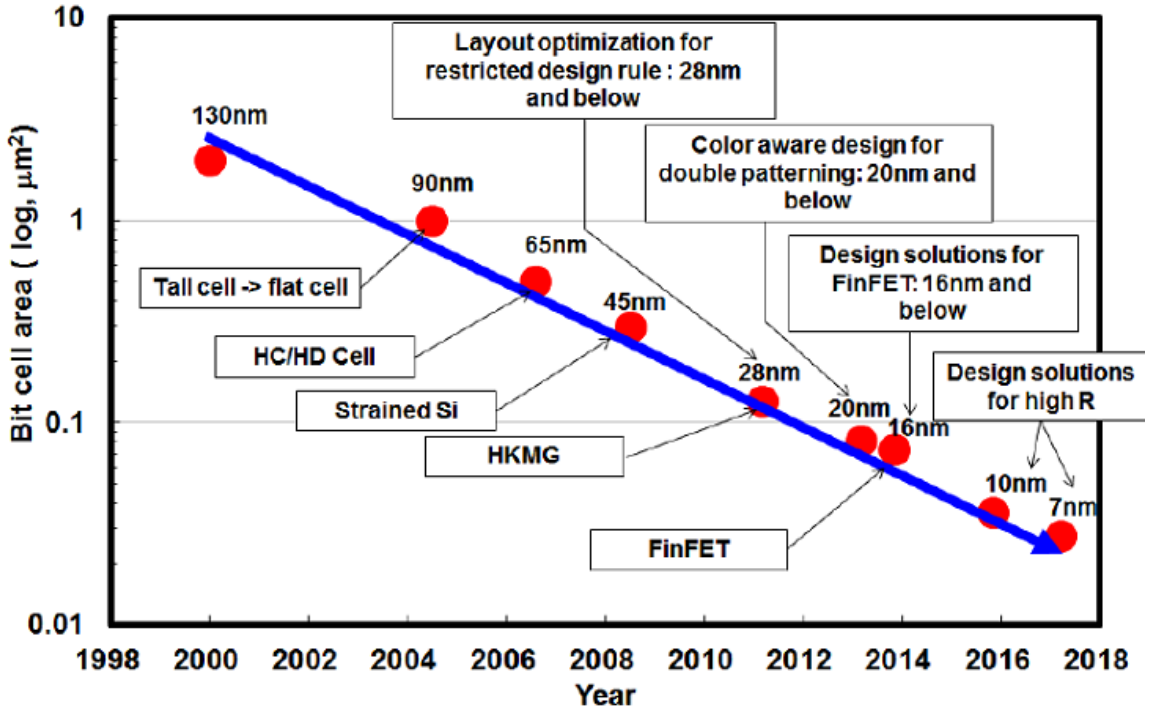
Perpendicular magnetic anisotropy (PMA)

Tunnel magnetoresistance (TMR)
$= (R_{AP}-R_P)/R_P \times 100\% = 100\%\sim200\%$

- Magnetic tunnel junctions (MTJ): Two ferromagnetic layers (CoFeB) sandwich a tunnel barrier (MgO).
- The spin polarized current can switch the magnetization of free layers by spin-transfer torque.

SRAM





IEDM Short course, 2017
J. Chang et al., VLSI Symp. T2-2, 2017

- **SRAM bit cell area scaling is enabled by**
  - ➢ **Innovations in process technology**
  - ➢ **Design solutions of read and write assist techniques**