

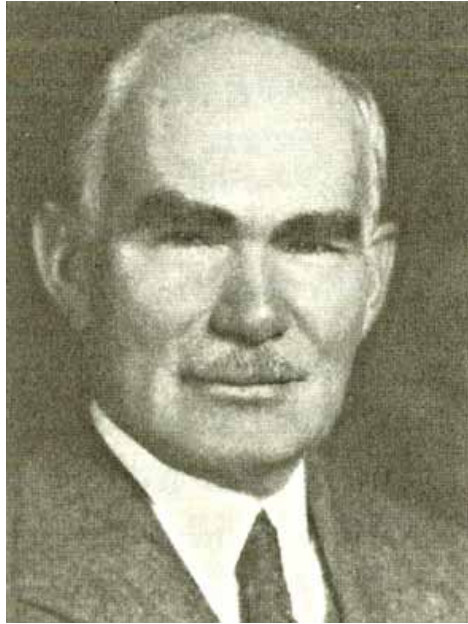
Downsizing of transistors towards its Limit

March 6, 2009

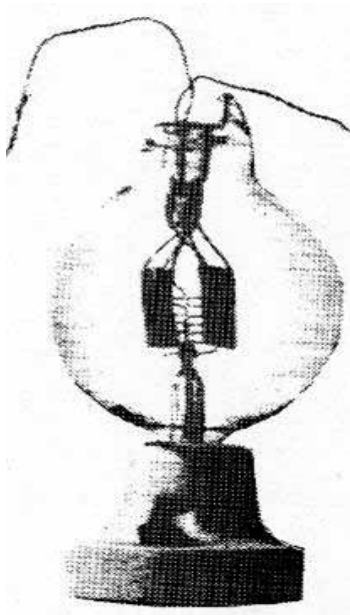
**@Bengal Institute of Technology
& Management**

**Hiroshi Iwai,
Tokyo Institute of Technology**

- There were many inventions in the 20th century:
Airplane, Nuclear Power generation, Computer,
Space aircraft, etc
- However, everything has to be controlled by
electronics
- Electronics
Most important invention in the 20th century
- What is Electronics: To use electrons,
Electronic Circuits

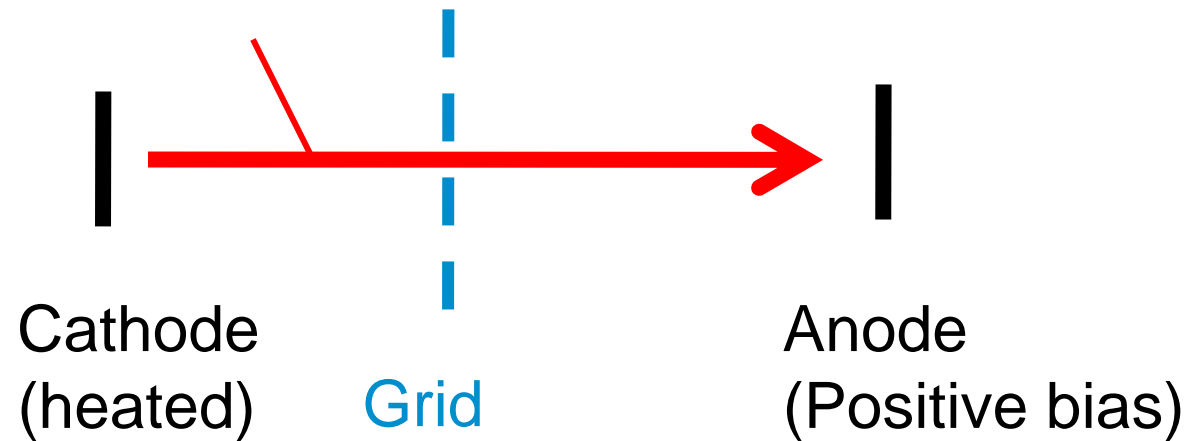


Lee De Forest



Electronic Circuits started by the invention of vacuum tube (Triode) in 1906

Thermal electrons from cathode controlled by grid bias



Same mechanism as that of transistor

4 wives of Lee De Forest

1906 Lucille Sheardown

1907 Nora Blatch

1912 Mary Mayo, singer

1930 Marie Mosquini, silent film actress



Mary

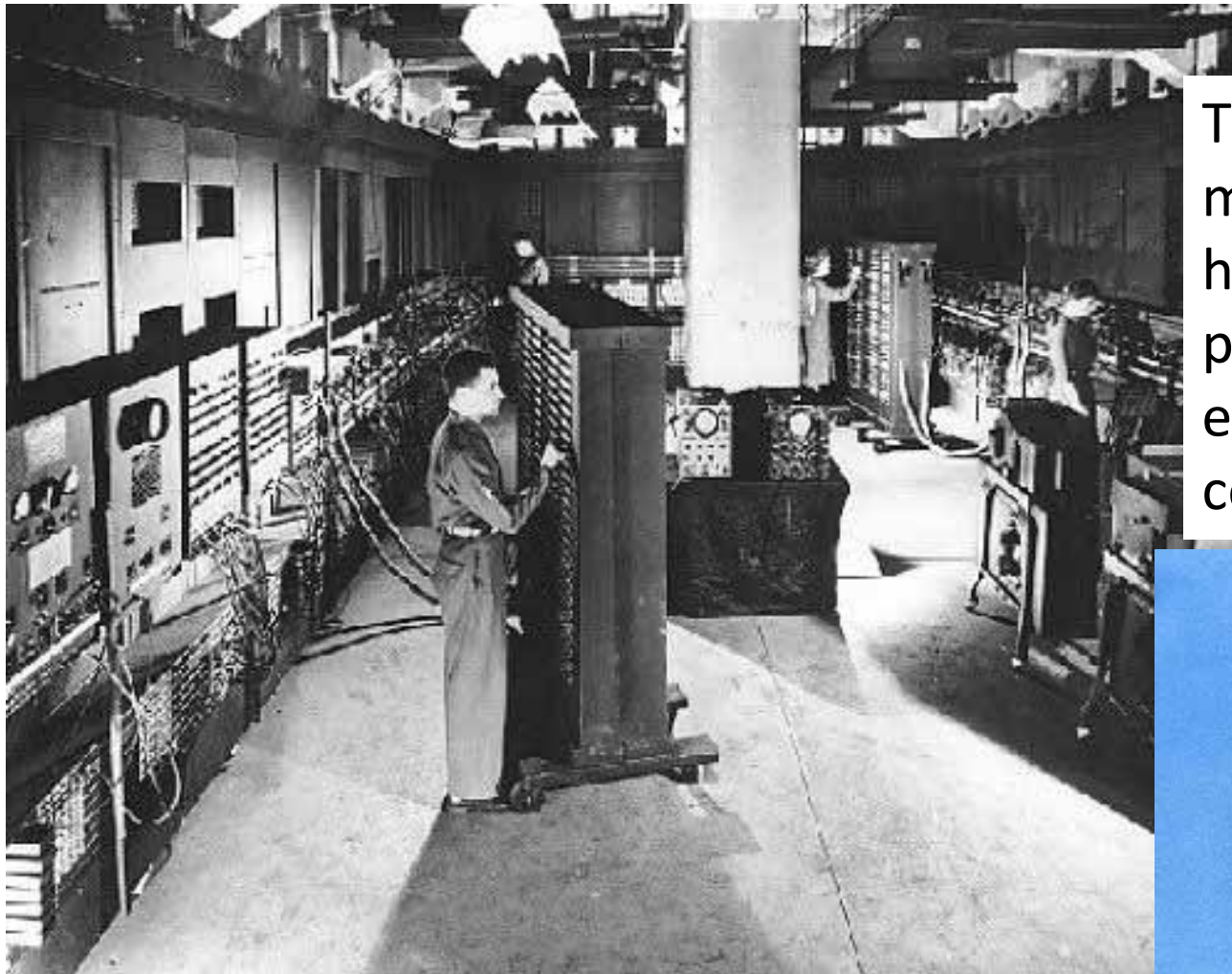


Marie



First Computer Eniac: made of huge number of vacuum tubes 1946
Big size, huge power, short life time filament

→ dreamed of replacing vacuum tube with solid-state device



Today's pocket PC
made of semiconductor
has much higher
performance with
extremely low power
consumption



History of Semiconductor devices

1947, 1st Point Contact Bipolar Transistor:

Ge Semiconductor, Bardeen, Brattin

→ Nobel Prize

1948, 1st Junction Bipolar Transistor,

Ge Semiconductor, Schokley

→ Nobel Prize

1958, 1st Integrated Circuits,

Ge Semiconductor, J.Kilby → Nobel Prize

1959, 1st Planar Integrated Circuits,

R.Noice

1960, 1st MOS Transistor, Kahng,

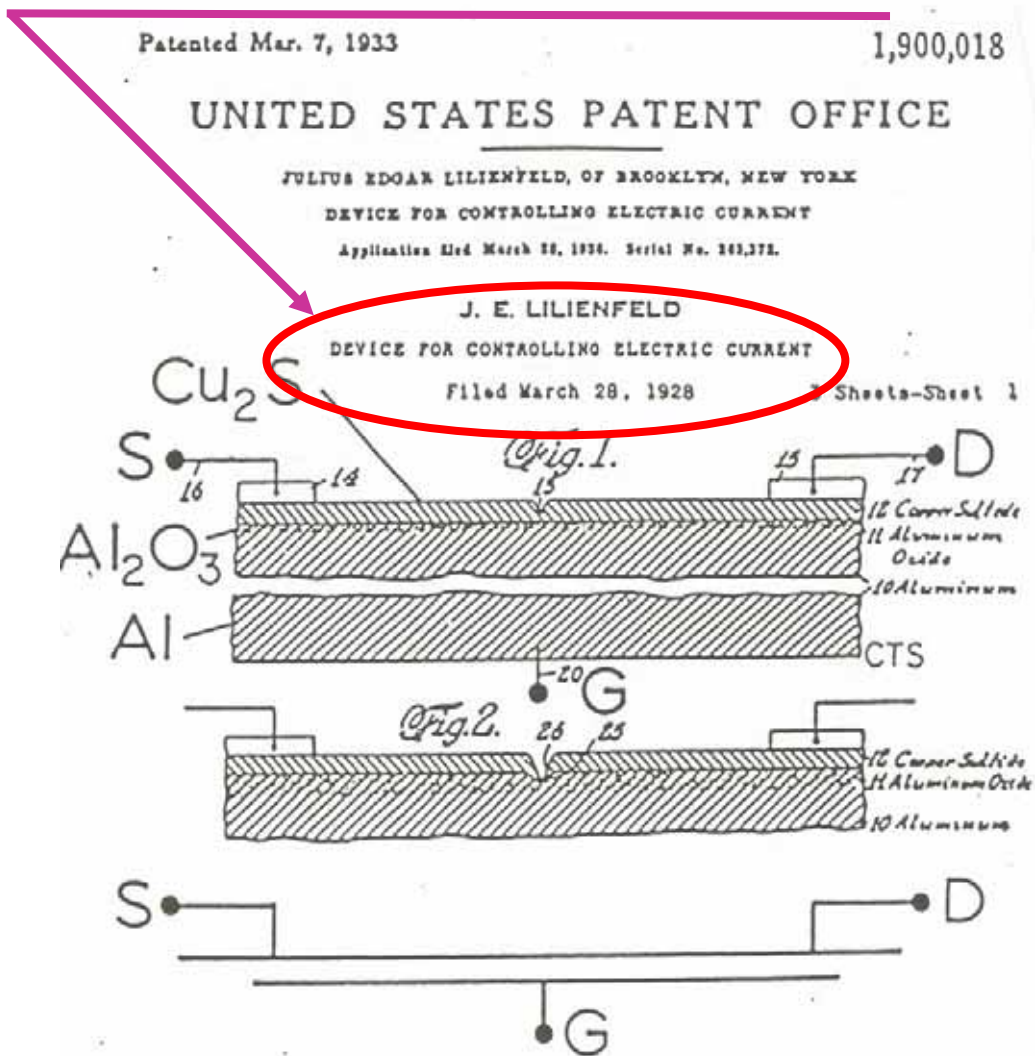
Si Semiconductor

1963, 1st CMOS Circuits, C.T. Sah and F. Wanlass

J. E. LILIENFELD

DEVICES FOR CONTROLLED ELECTRIC CURRENT

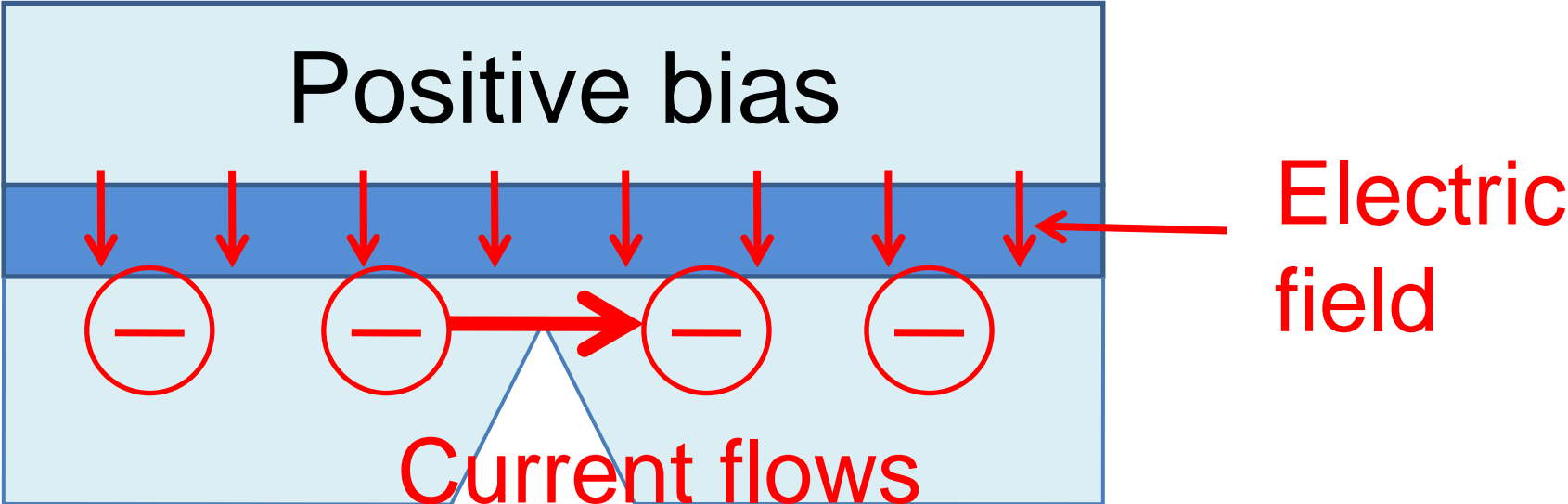
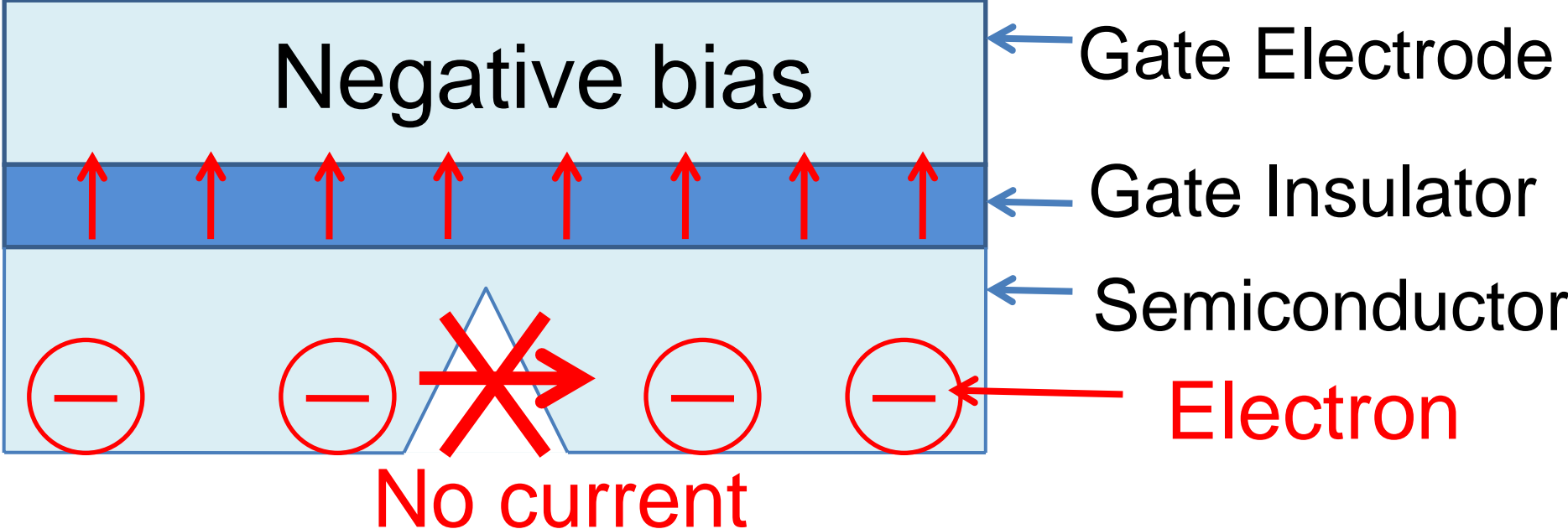
Filed March 28, 1928

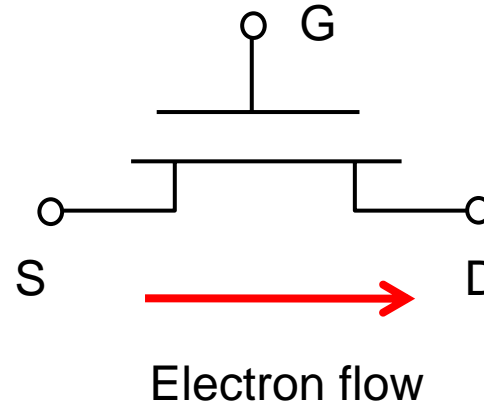
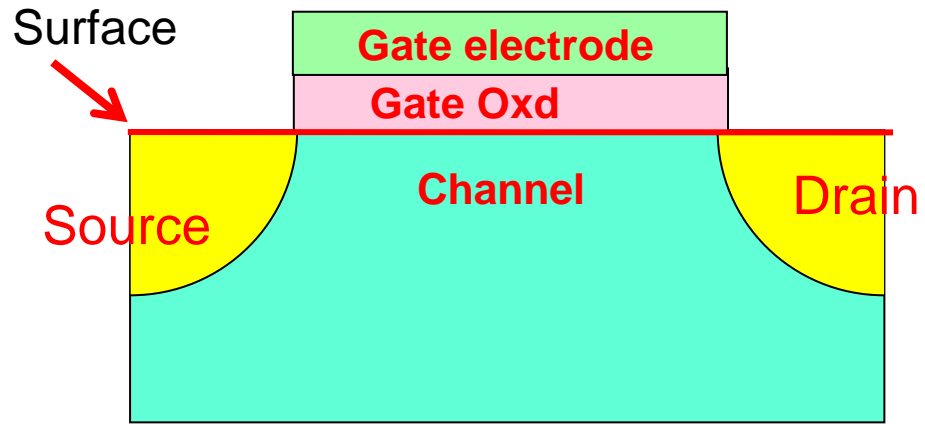


J.E.LILIENFELD



Capacitor structure with notch

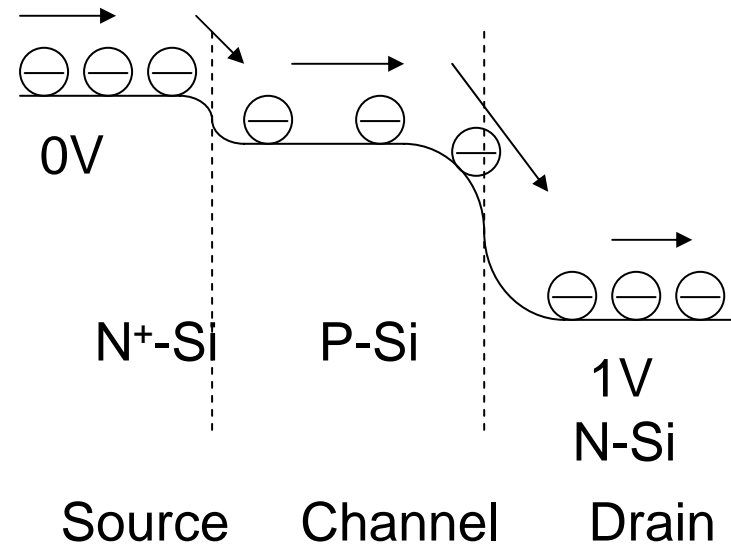
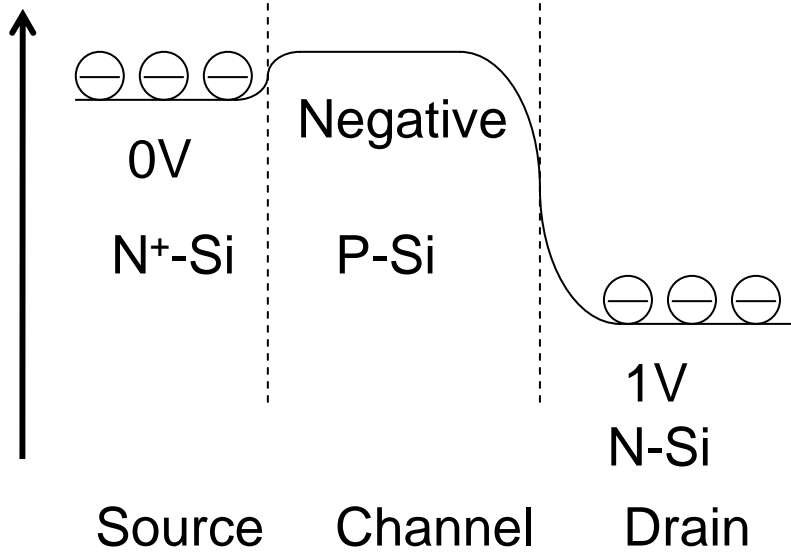




0 bias for gate

Positive bias for gate

Surface Potential (Negative direction)

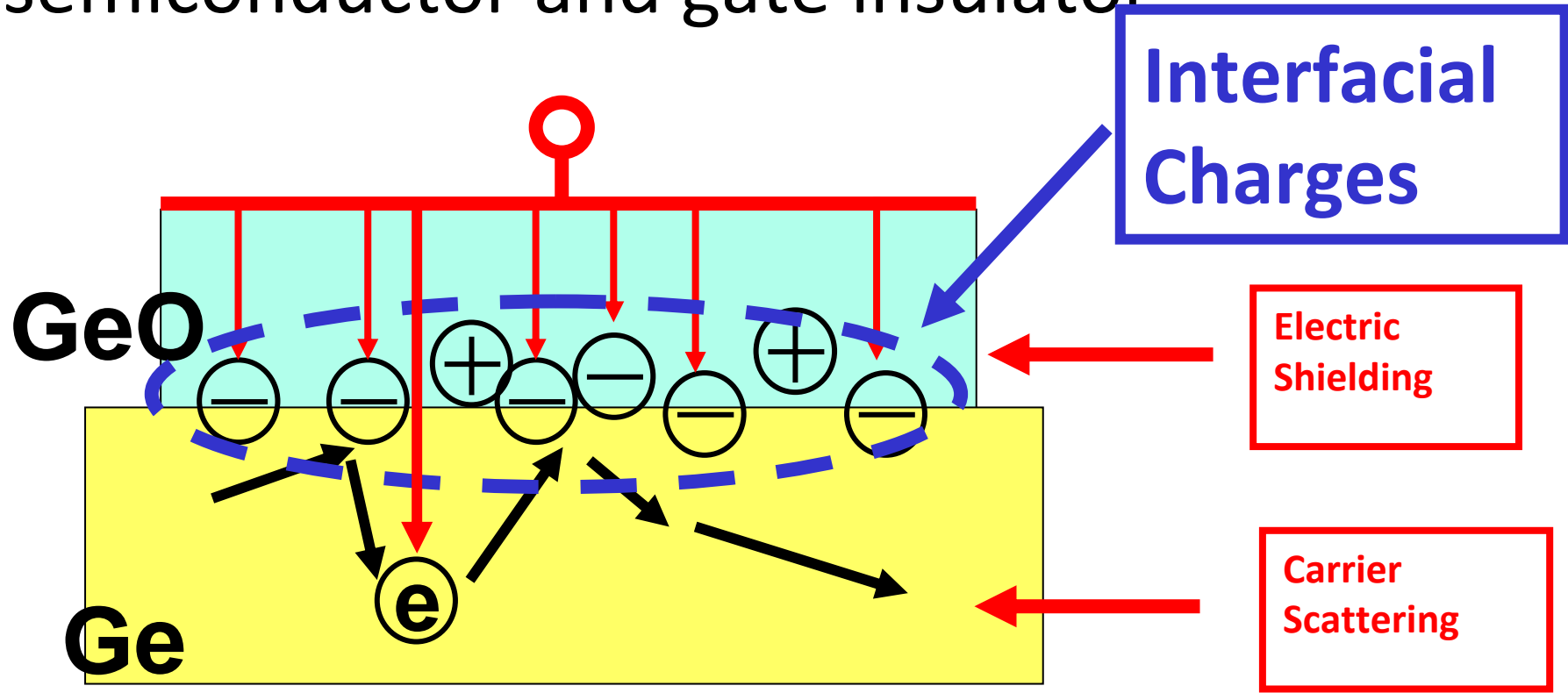


However, no one could realize MOSFET operation for more than 30 years.

Because of very bad interface property between the semiconductor and gate insulator

Even Shockley!

Very bad interface property between the semiconductor and gate insulator



Drain Current was several orders of magnitude smaller than expected

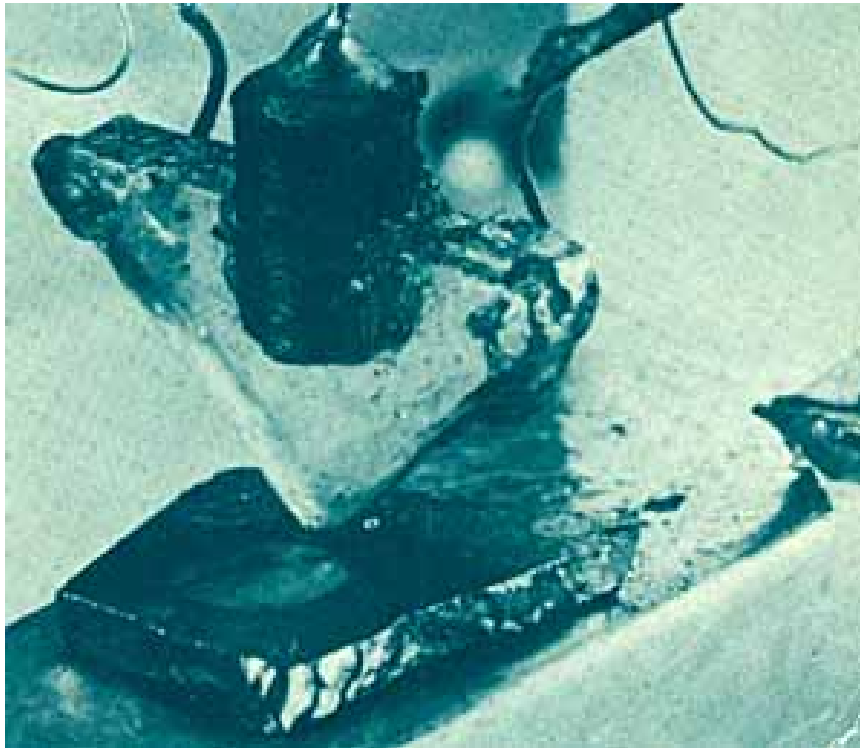
Even Shockley!

However, they found amplification phenomenon when investigating Ge surface when putting needles.

This is the 1st Transistor:

**Not Field Effect Transistor,
But Bipolar Transistor (another mechanism)**

1947: 1st transistor



Bipolar using Ge

J. Bardeen

W. Bratten,

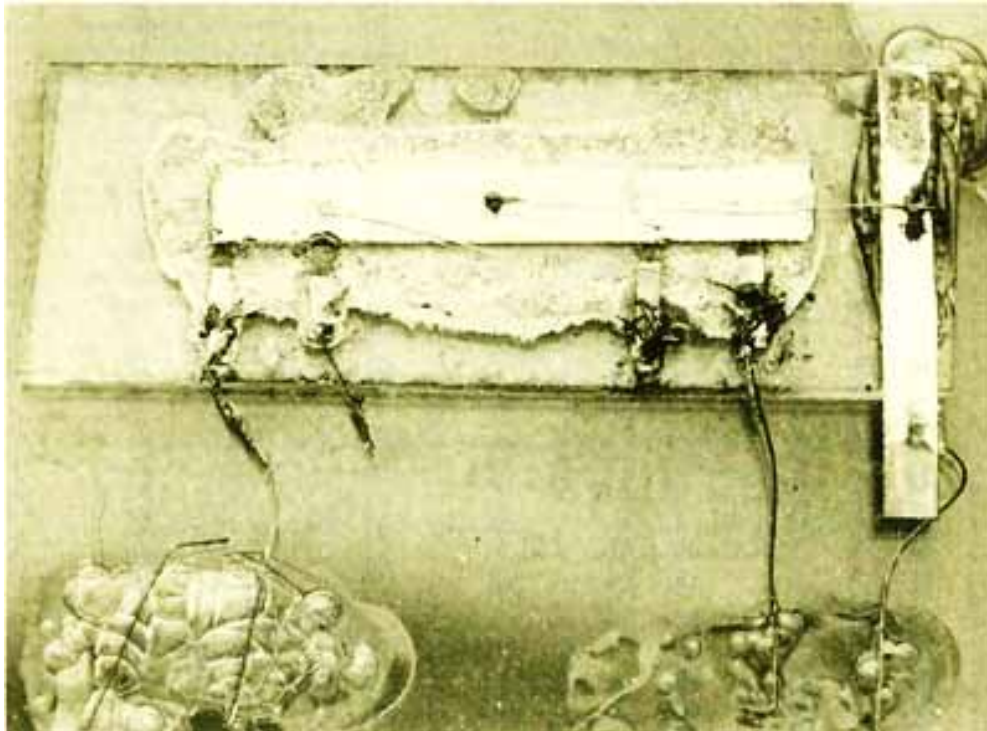


W. Shockley

1958: 1st Integrated Circuit

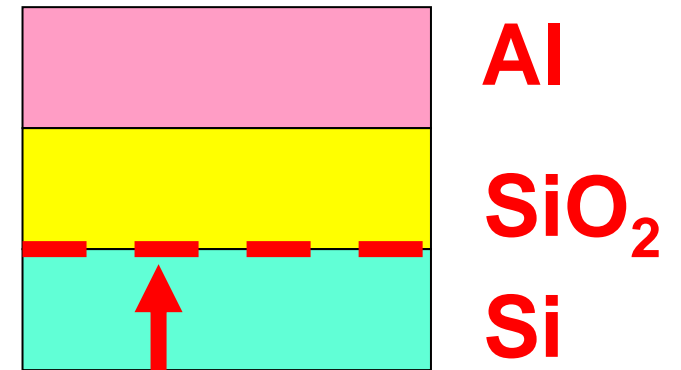
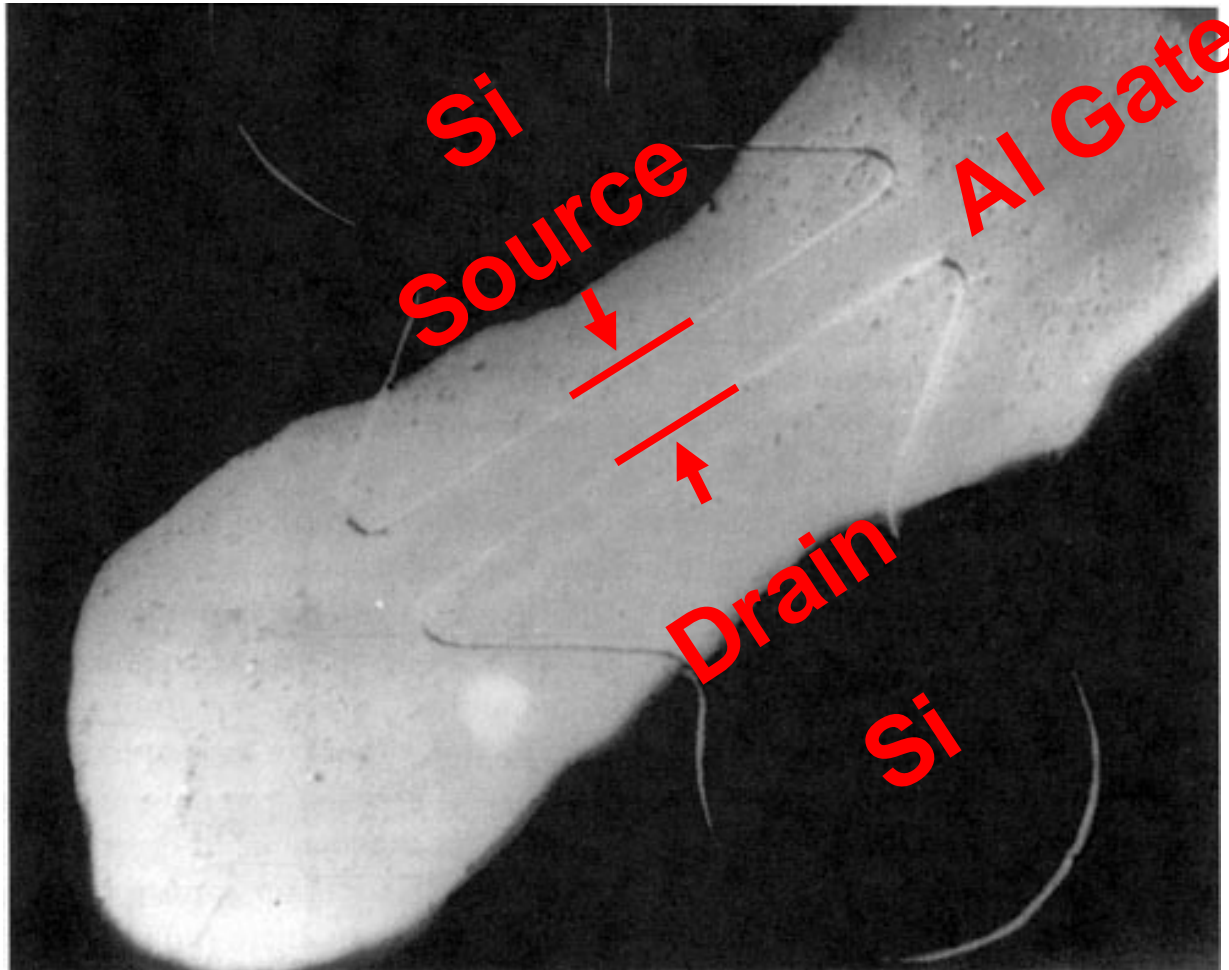
Jack S. Kilby

Connect 2 bipolar transistors in the Same substrate by bonding wire.



1960: First MOSFET
by D. Kahng and M. Atalla

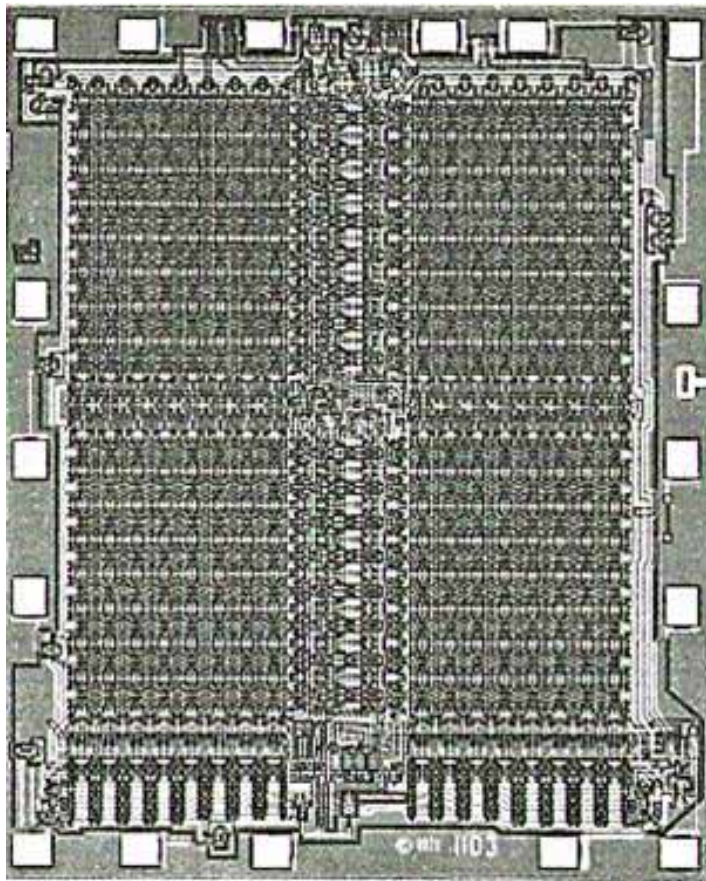
Top View



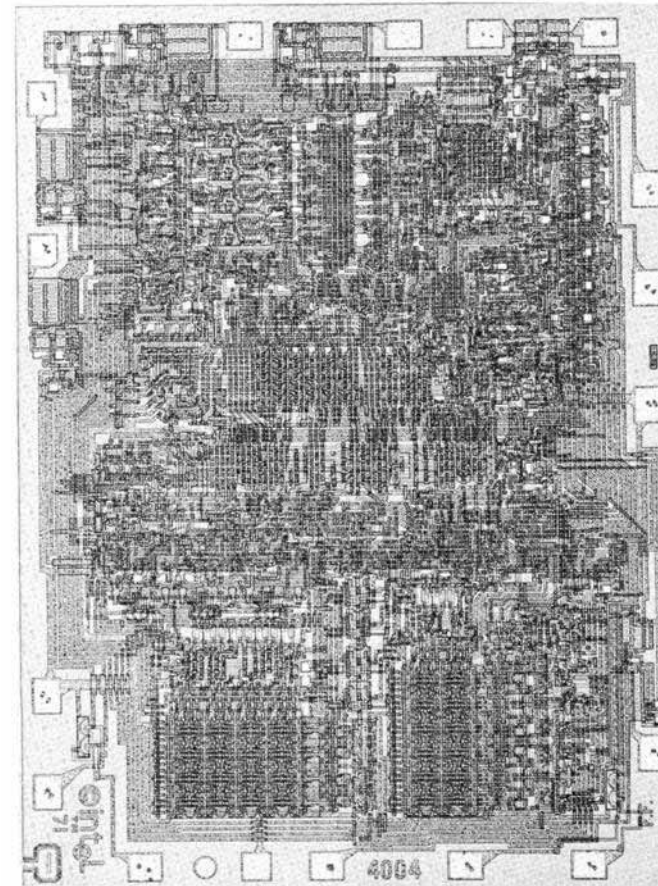
Si/SiO₂ Interface is
extraordinarily good

1970,71: 1st generation of LSIs

DRAM Intel 1103



MPU Intel 4004



MOS LSI experienced continuous progress for many years

	Name of Integrated Circuits	Number of Transistors
1960s	IC (Integrated Circuits)	~ 10
1970s	LSI (Large Scale Integrated Circuit)	~1,000
1980s	VLSI (Very Large Scale IC)	~10,000
1990s	ULSI (Ultra Large Scale IC)	~1,000,000
2000s	?LSI (? Large Scale IC)	~1000,000,000

Gate Electrode
Poly Si

Gate Insulator
SiO₂

Substrate
Si

MOSFET: Metal Oxide Semiconductor
Field Effect Transistor

Use Gate Field Effect for switching

Gate Electrode
Poly Si

Gate Insulator
SiO₂

Source

n-Si



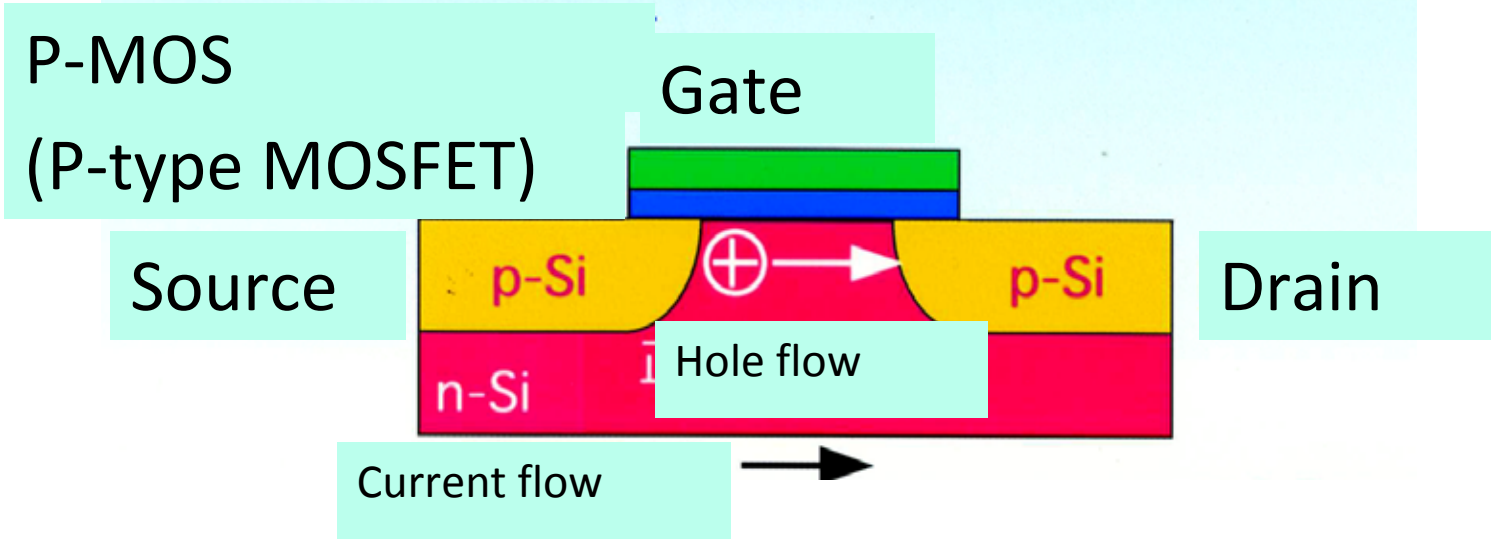
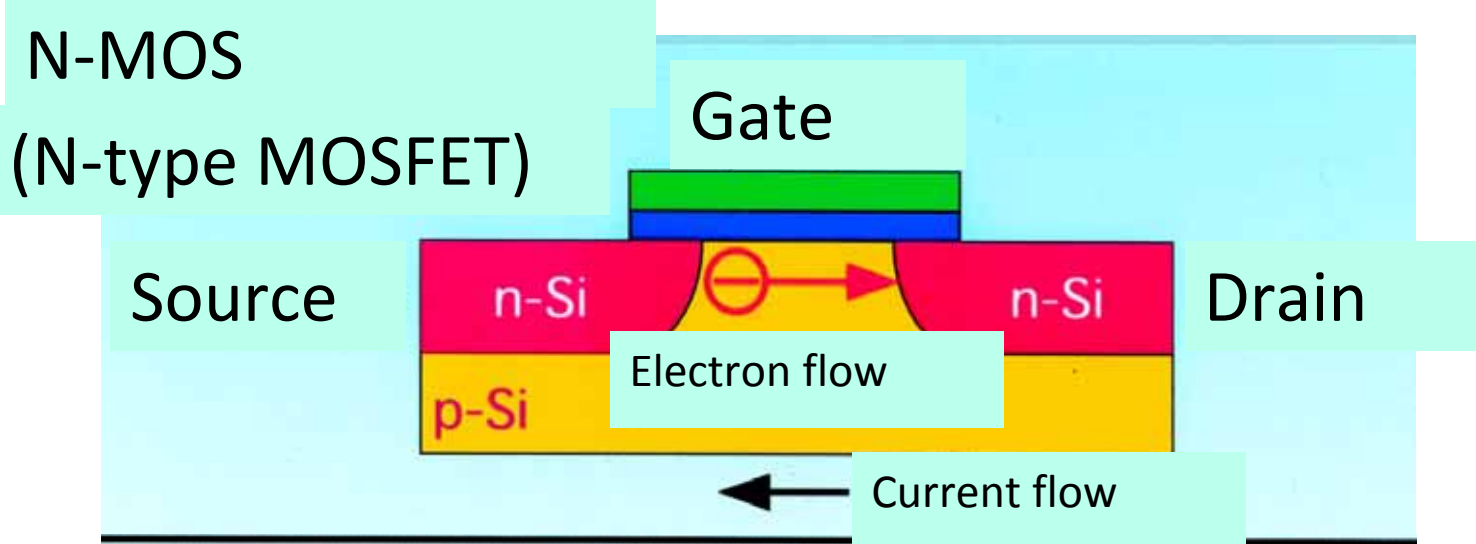
n-Si

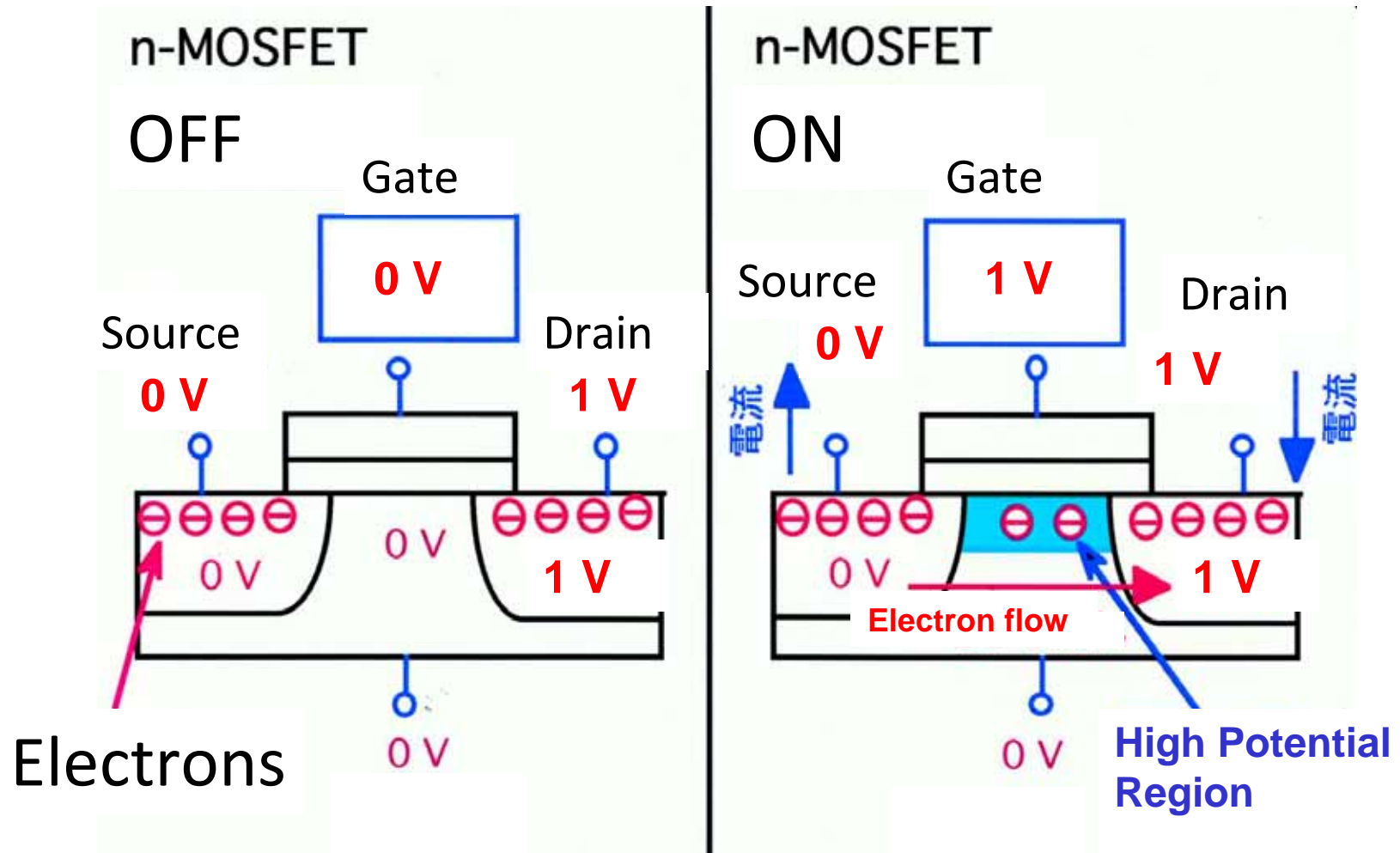
Drain

Channel

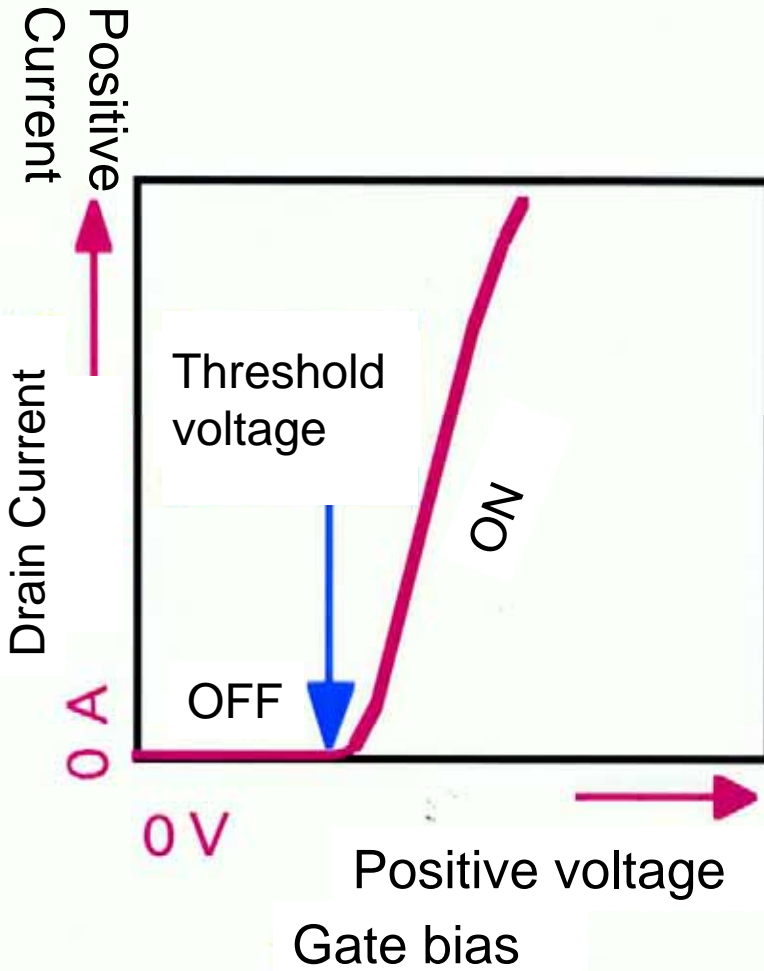
N-MOS (N-type MOSFET)

Si
Substrate

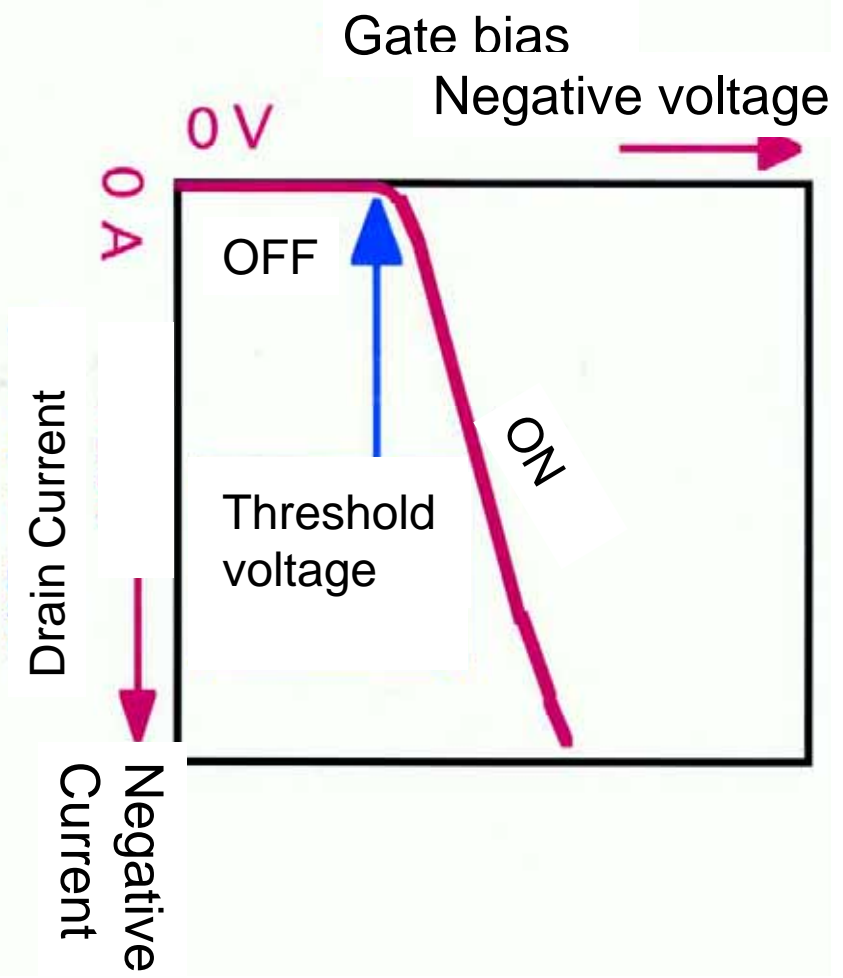




n-MOSFET

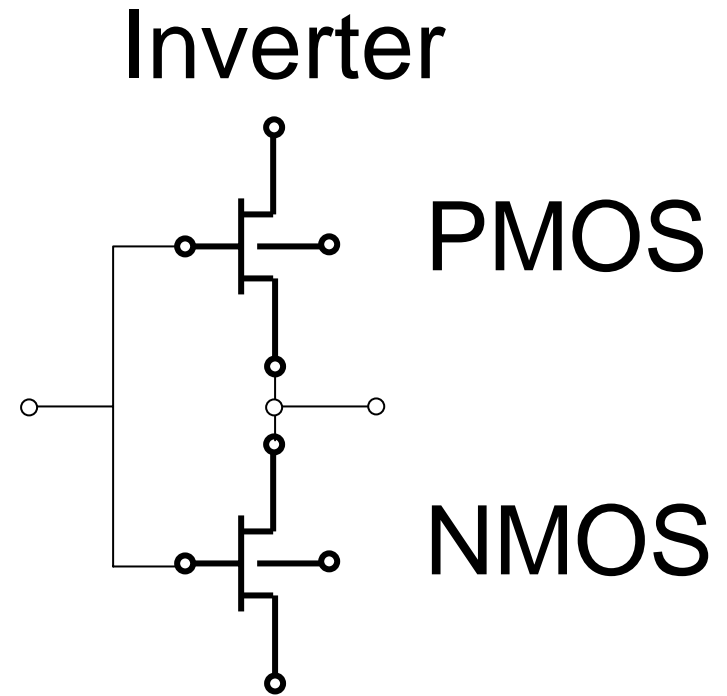


p-MOSFET



CMOS

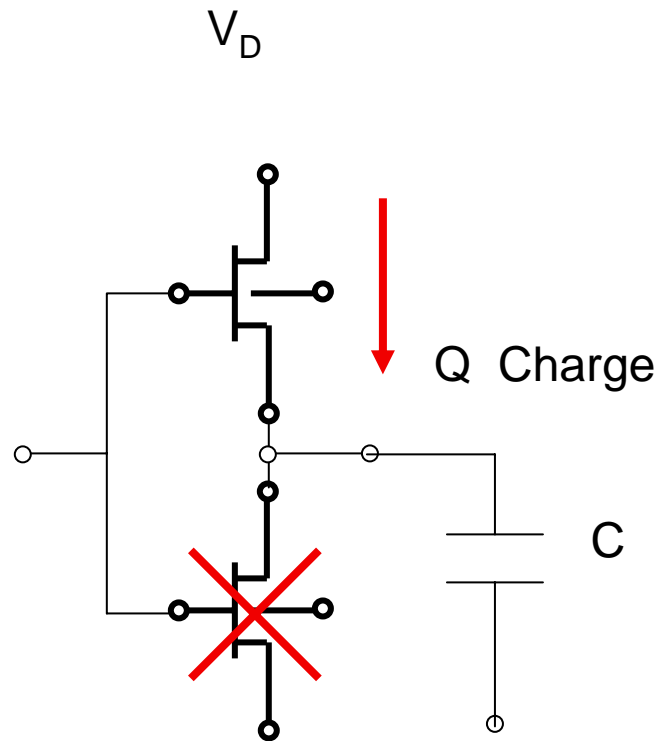
Complimentary MOS



When NMOS is ON, PMOS is OFF

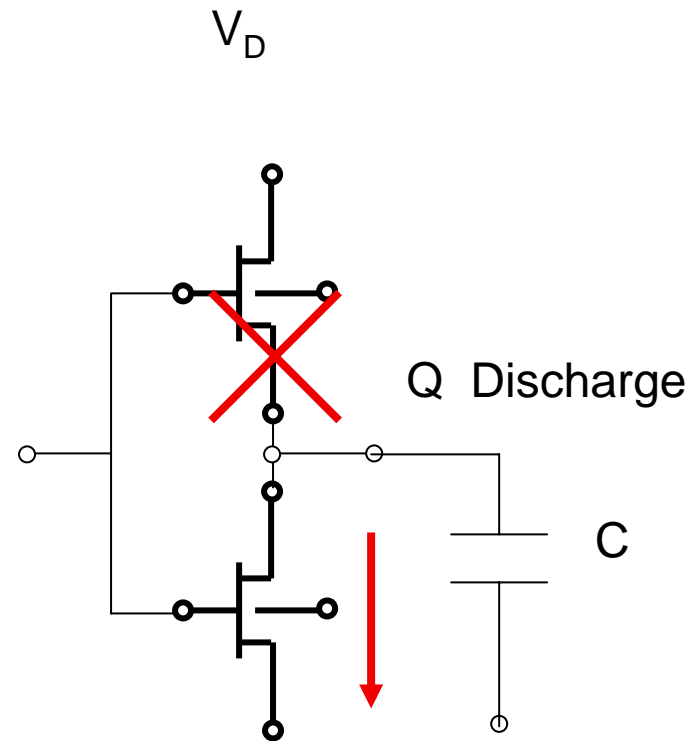
When PMOS is ON, NMOS is OFF

CMOS: Low Power: No DC current from Power supply to the ground



1 cycle

$$P = \frac{1}{2} CV_D^2$$



Clock frequency f

$$P = \frac{1}{2} fCV_D^2$$

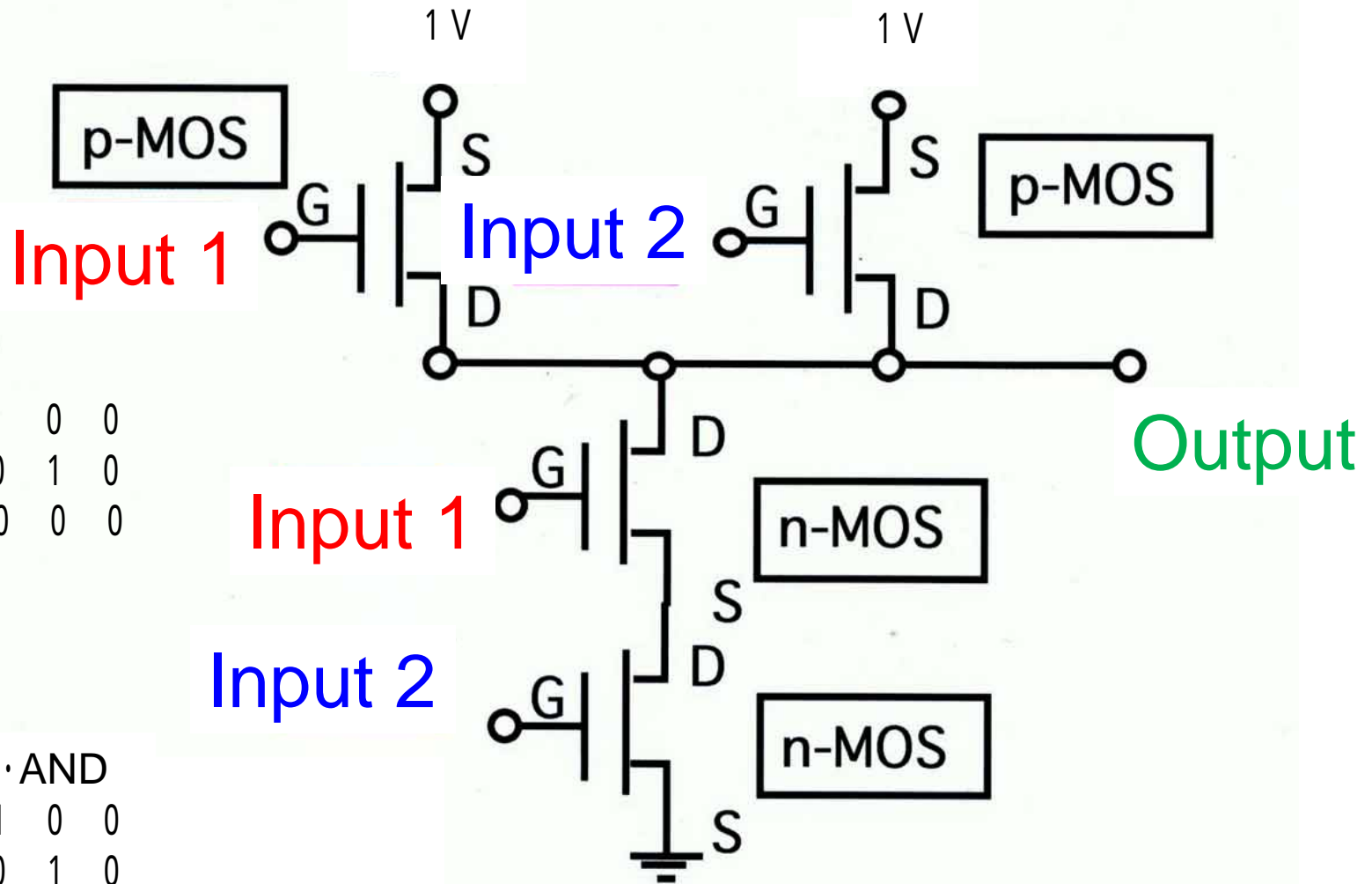
2 input NAND Circuit

AND

Input 1	1	1	0	0
Input 2	1	0	1	0
Output	1	0	0	0

NAND = NOT · AND

Input 1	1	1	0	0
Input 2	1	0	1	0
Output	0	1	1	1



Needless to say, but....

CMOS Technology:

Indispensable for our human society

All the human activities are controlled by CMOS

living, production, financing, telecommunication,
transportation, medical care, education,
entertainment, etc.

Without CMOS:

There is no computer in banks, and
world economical activities immediately stop.

Cellular phone does not exist

Downsizing of the components has been the driving force for circuit evolution



1900	1950	1960	1970	2000
Vacuum Tube	Transistor	IC	LSI	ULSI
10 cm	cm	mm	10 μ m	100 nm
10^{-1} m	10^{-2} m	10^{-3} m	10^{-5} m	10^{-7} m

In 100 years, the size reduced by one million times. There have been many devices from stone age. **We have never experienced such a tremendous reduction of devices in human history.**

Downsizing

1. Reduce Capacitance

→ Reduce switching time of MOSFETs

→ Increase clock frequency

→ Increase circuit operation speed

2. Increase number of Transistors

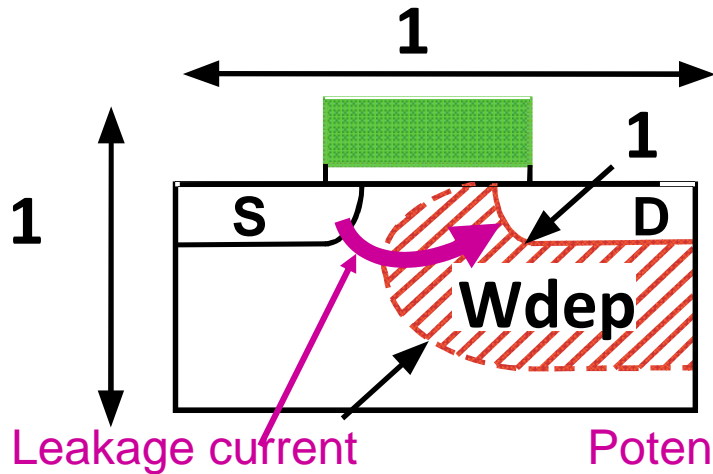
→ Parallel processing

→ Increase circuit operation speed

Downsizing contribute to the performance increase in double ways

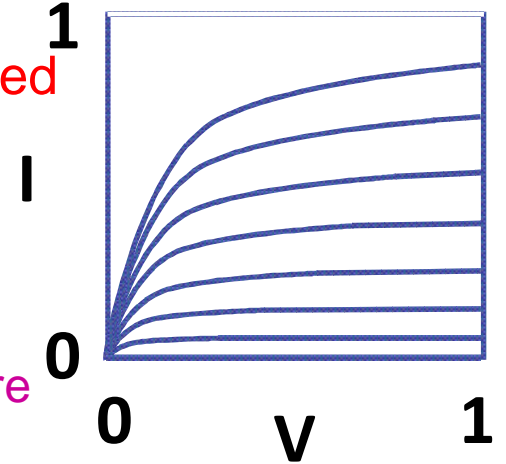
Thus, downsizing of Si devices is the most important and critical issue.²⁶

Scaling Method: by R. Dennard in 1974



Wdep: Space Charge Region (or Depletion Region) Width

Wdep has to be suppressed
Otherwise, large leakage
between S and D



Leakage current

Potential in space charge region is high, and thus, electrons in source are attracted to the space charge region.

**K=0.7
for
example**

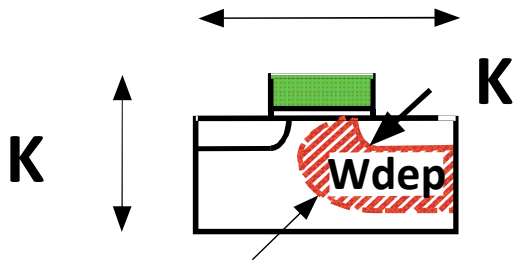


X , Y , Z : K, V : K, Na : 1/K

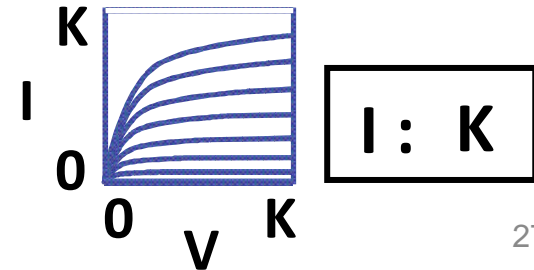
By the scaling, Wdep is suppressed in proportion,
and thus, leakage can be suppressed.

K

→ Good scaled I-V characteristics



**Wdep ∝ √(V/Na)
: K**



I : K

Downscaling merit: Beautiful!

Geometry & Supply voltage	L_g, W_g T_{ox}, V_{dd}	K	Scaling K : K=0.7 for example
Drive current in saturation	I_d	K	$I_d = v_{sat} W_g C_o (V_g - V_{th})$ C_o : gate C per unit area $\rightarrow W_g (t_{ox}^{-1})(V_g - V_{th}) = W_g t_{ox}^{-1}(V_g - V_{th}) = KK^{-1}K = K$
I_d per unit W_g	$I_d/\mu m$	1	I_d per unit $W_g = I_d / W_g = 1$
Gate capacitance	C_g	K	$C_g = \epsilon_o \epsilon_{ox} L_g W_g / t_{ox} \rightarrow KK/K = K$
Switching speed	τ	K	$\tau = C_g V_{dd} / I_d \rightarrow KK/K = K$
Clock frequency	f	1/K	$f = 1/\tau = 1/K$
Chip area	A_{chip}	α	α: Scaling factor \rightarrow In the past, $\alpha > 1$ for most cases
Integration (# of Tr)	N	α/K^2	$N \rightarrow \alpha/K^2 = 1/K^2$, when $\alpha=1$
Power per chip	P	α	$fNCV^2/2 \rightarrow K^{-1}(\alpha K^{-2})K(K^1)^2 = \alpha = 1$, when $\alpha=1$

$k = 0.7$ and $\alpha = 1$

Single MOFET

$V_{dd} \rightarrow 0.7$

$L_g \rightarrow 0.7$

$I_d \rightarrow 0.7$

$C_g \rightarrow 0.7$

P (Power)/Clock

$\rightarrow 0.7^3 = 0.34$

τ (Switching time) $\rightarrow 0.7$

$k = 0.7^2 = 0.5$ and $\alpha = 1$

$V_{dd} \rightarrow 0.5$

$L_g \rightarrow 0.5$

$I_d \rightarrow 0.5$

$C_g \rightarrow 0.5$

P (Power)/Clock

$\rightarrow 0.5^3 = 0.125$

τ (Switching time) $\rightarrow 0.5$

Chip

N (# of Tr) $\rightarrow 1/0.7^2 = 2$

f (Clock) $\rightarrow 1/0.7 = 1.4$

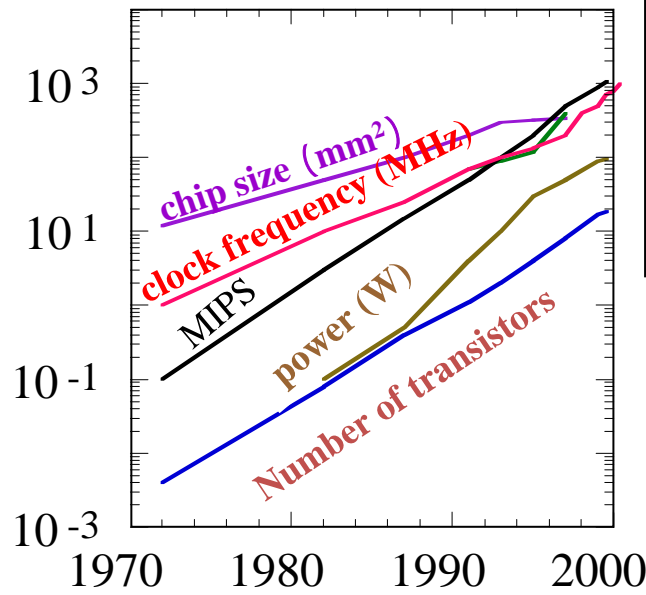
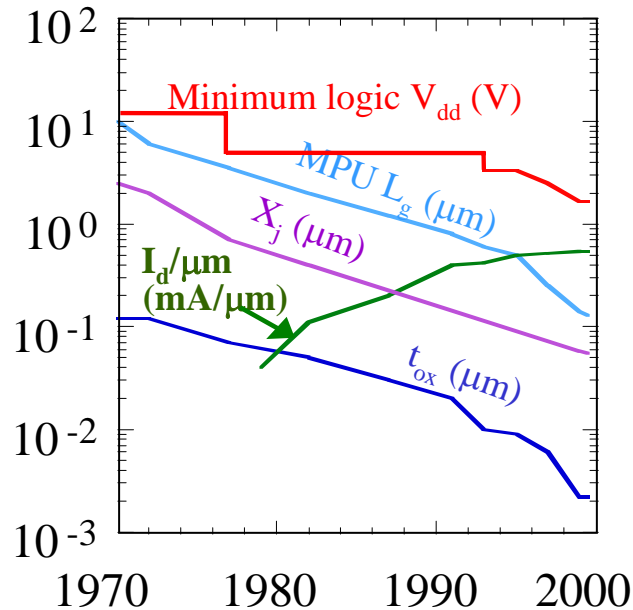
P (Power) $\rightarrow 1$

N (# of Tr) $\rightarrow 1/0.5^2 = 4$

f (Clock) $\rightarrow 1/0.5 = 2$

P (Power) $\rightarrow 1$

Actual past downscaling trend until year 2000



Past 30 years scaling
 Merit: N, f increase
 Demerit: P increase

V_{dd} scaling insufficient
 ↓
 Additional significant increase in I_d, f, P

Source: Iwai and S. Ohmi, Microelectronics Reliability 42 (2002), pp.1251-1268

Change in 30 years

	Ideal scaling	Real Change		Ideal scaling	Real Change		Ideal scaling	Real Change
L_g	K	10^{-2}	I_d	K (10^{-2})	10^{-1}	f	$1/K(10^2)$	10^3
t_{ox}	K(10^{-2})	10^{-2}	$I_d/\mu m$	1	10^1	P	$\alpha(10^1)$	10^5
V_{dd}	K(10^{-2})	10^{-1}	N	$\alpha/K^2(10^5)$	10^4	= $f\alpha NCV^2$		
A_{chip}	α	10^1						

V_d scaling insufficient, α increased → N, I_d , f, P increased significantly

Many people wanted to say about the limit. Past predictions were not correct!!

Period	Expected limit(size)	Cause
Late 1970's	1 μ m:	SCE
Early 1980's	0.5 μ m:	S/D resistance
Early 1980's	0.25 μ m:	Direct-tunneling of gate SiO ₂
Late 1980's	0.1 μ m:	'0.1 μ m brick wall'(various)
2000	50nm:	'Red brick wall' (various)
2000	10nm:	Fundamental?

Historically, many predictions of the limit of downsizing.

VLSI text book written 1979 predict that 0.25 micrometer would be the limit because of direct-tunneling current through the very thin-gate oxide.

INTRODUCTION TO **VLSI** SYSTEMS

CARVER MEAD • LYNN CONWAY





C. Mead

L. Conway

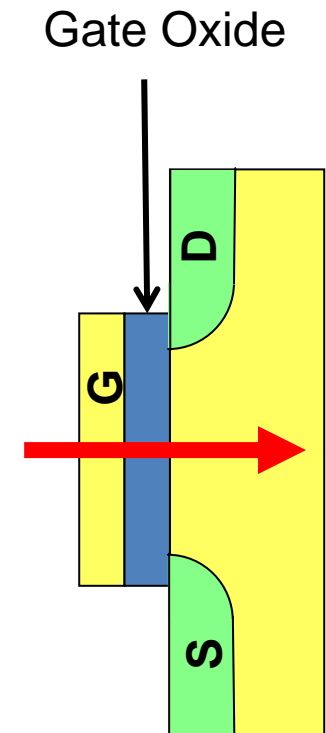
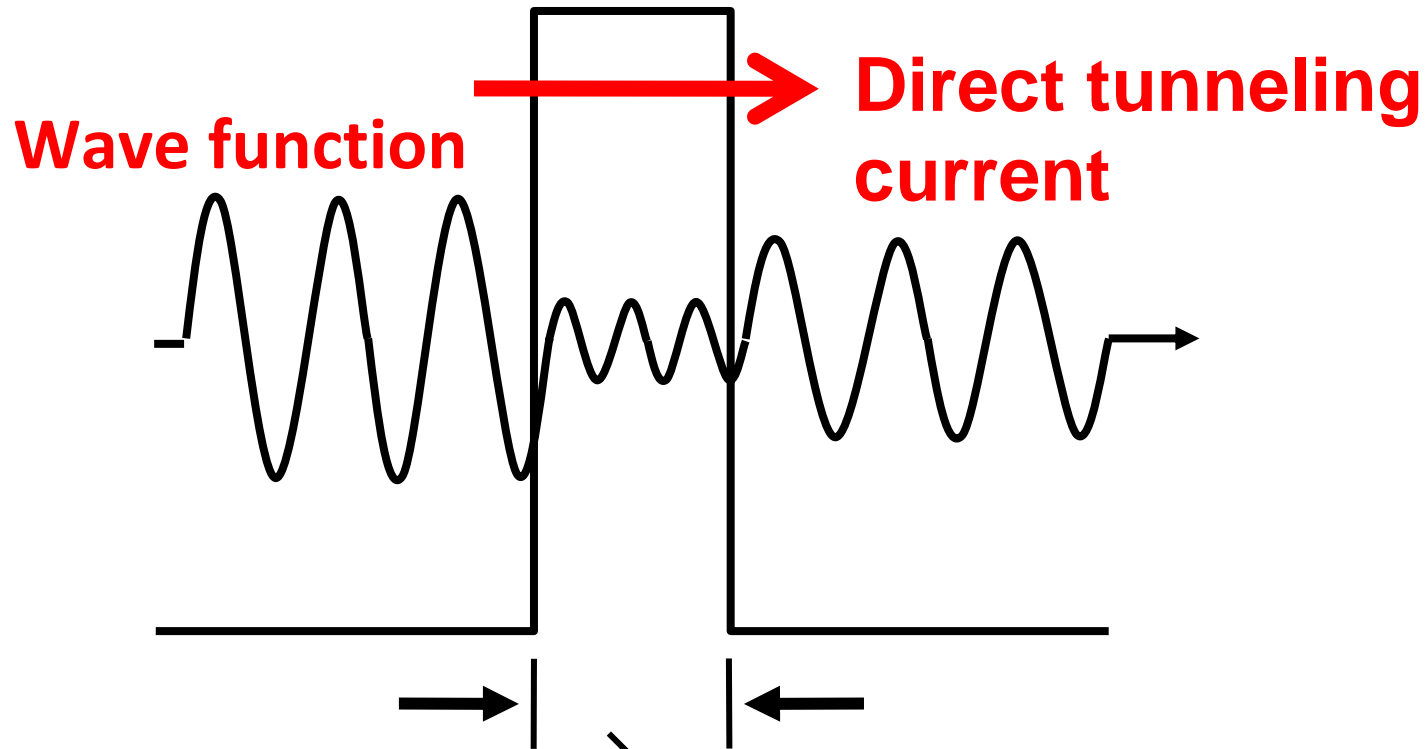
VLSI textbook

Finally, there appears to be a fundamental limit ¹⁰ of approximately quarter micron channel length, where certain physical effects such as the **tunneling through the gate oxide begin to make the devices of smaller dimension unworkable.**

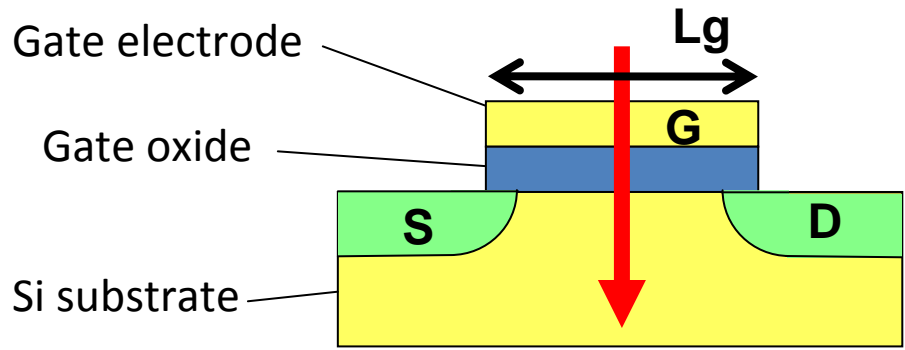
Direct-tunneling effect

Gate Electrode Gate Oxide Si Substrate

Potential Barrier

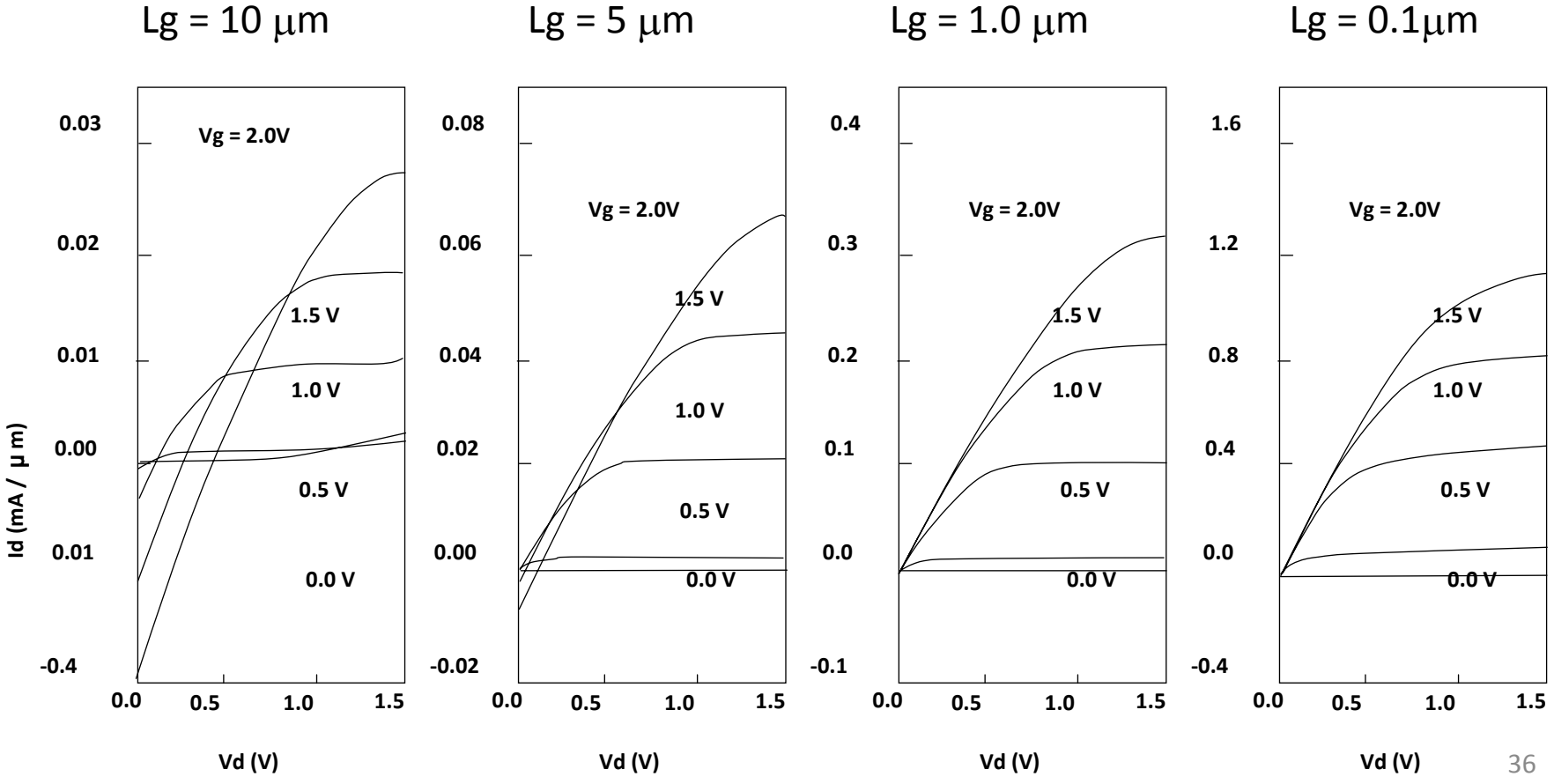


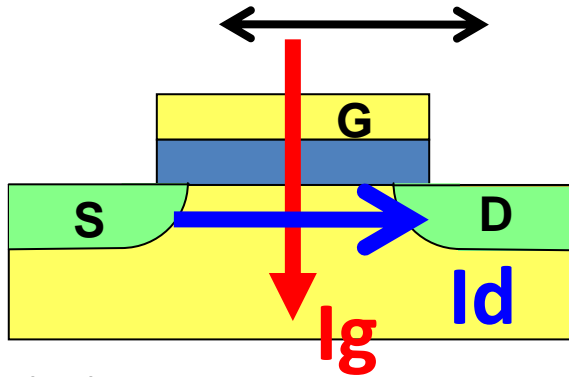
Direct tunneling leakage current start to flow when the thickness is 3 nm.



Direct tunneling leakage was found to be OK! In 1994!

MOSFETs with 1.5 nm gate oxide





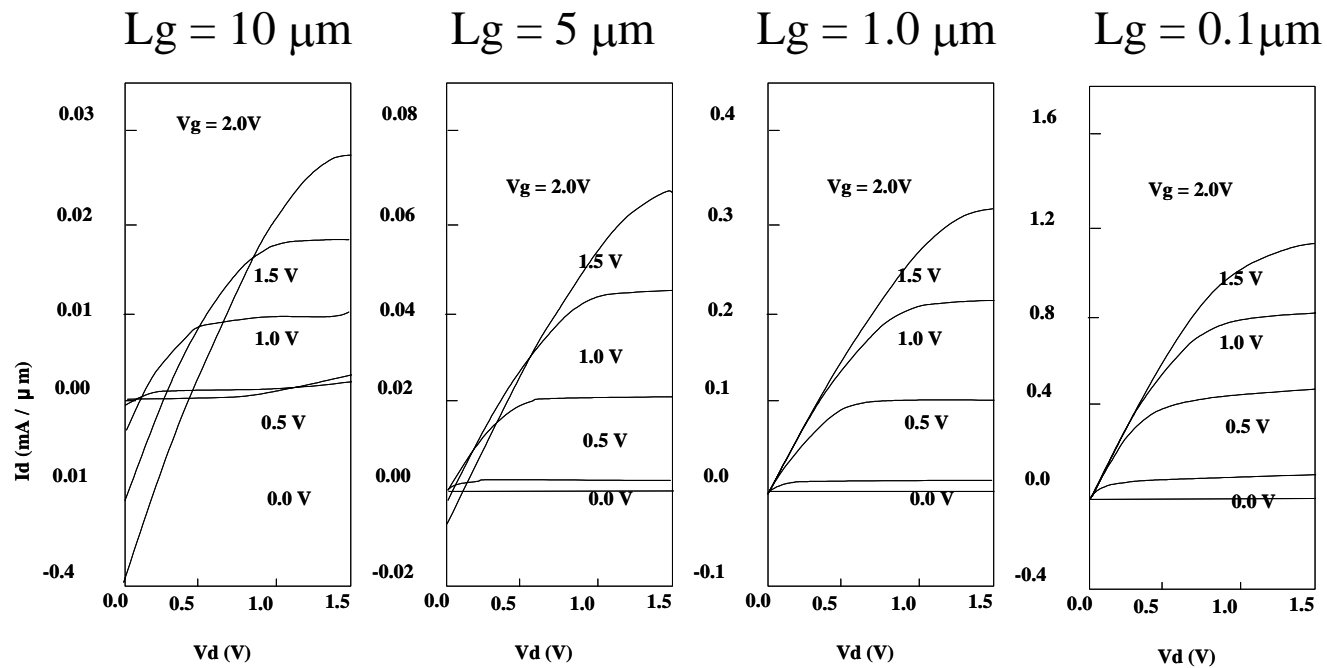
Gate leakage: $I_g \propto \text{Gate Area} \propto \text{Gate length (L}_g\text{)}$

Drain current: $I_d \propto 1/\text{Gate length (L}_g\text{)}$

$L_g \rightarrow \text{small,}$

Then, $I_g \rightarrow \text{small, } I_d \rightarrow \text{large, Thus, } I_g/I_d \rightarrow \text{very small}$

I_d
→



Do not believe a text book statement, blindly!

Never Give Up!

No one knows future!

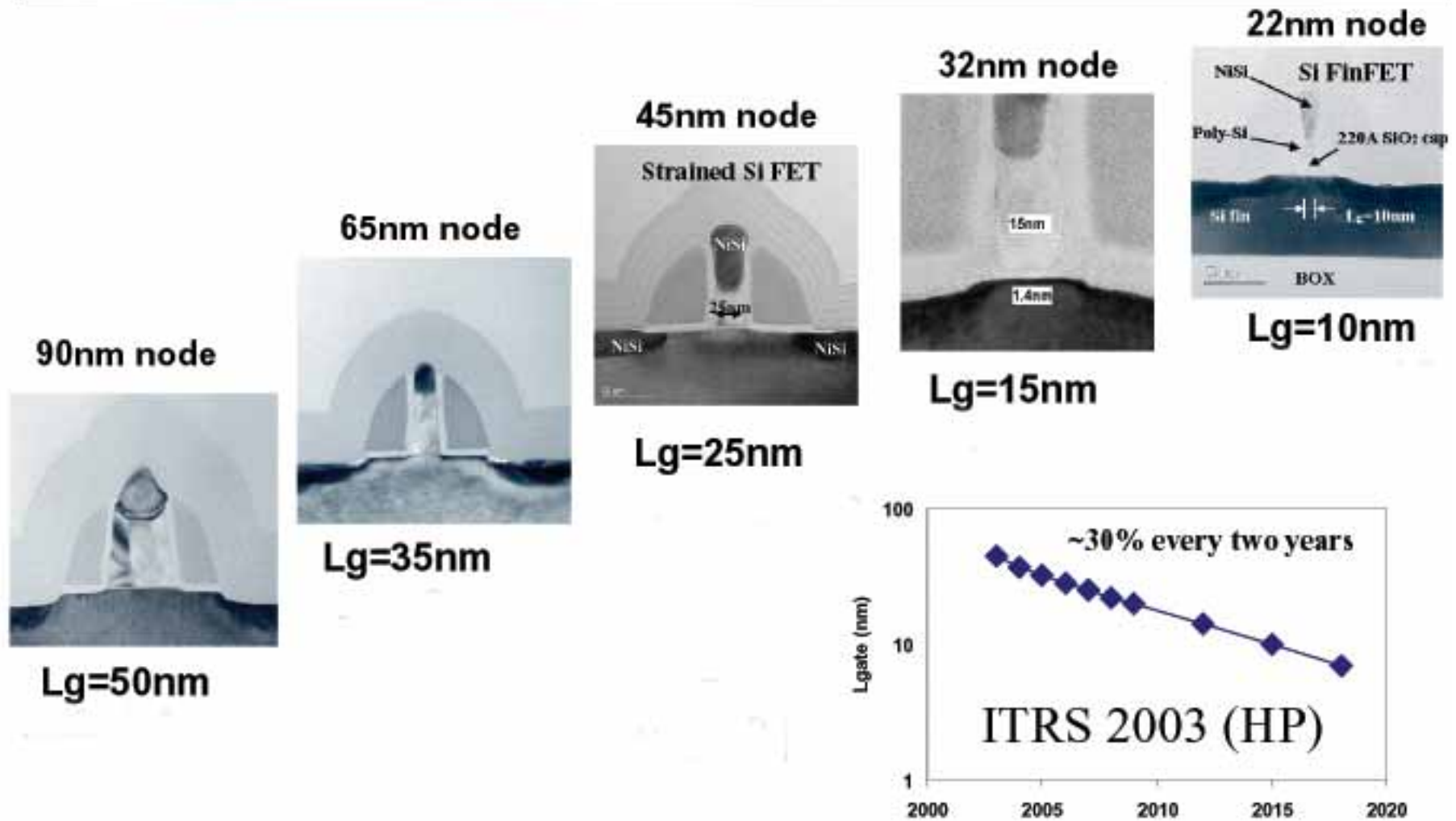
There would be a solution!

Think, Think, and Think!

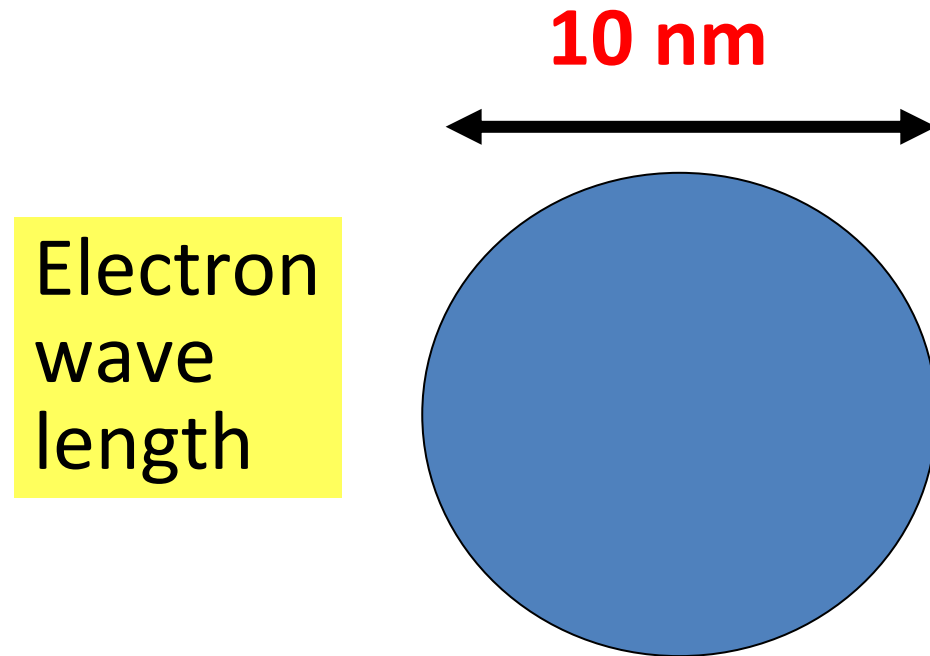
Or, Wait the time!

Some one will think for you

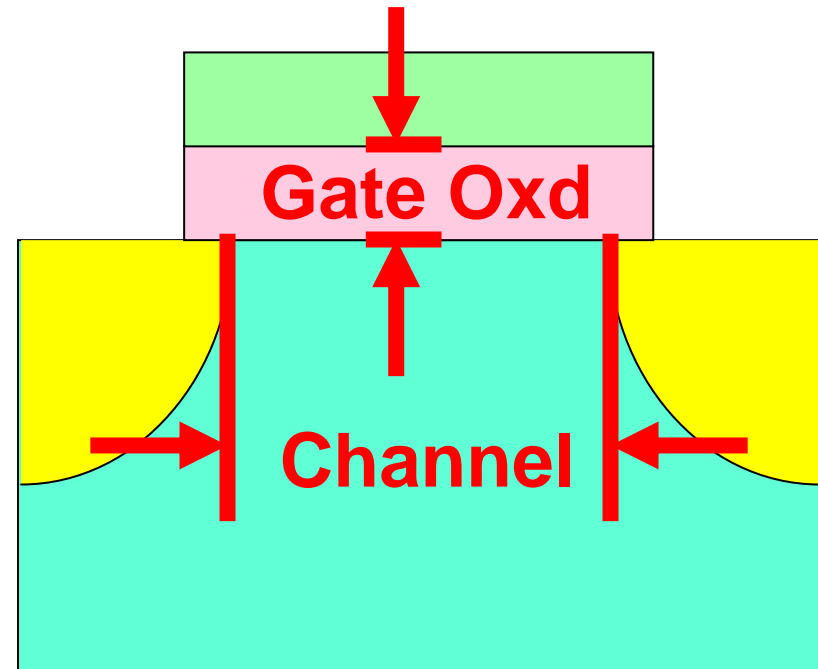
Transistor Scaling Continues



Downsizing limit?

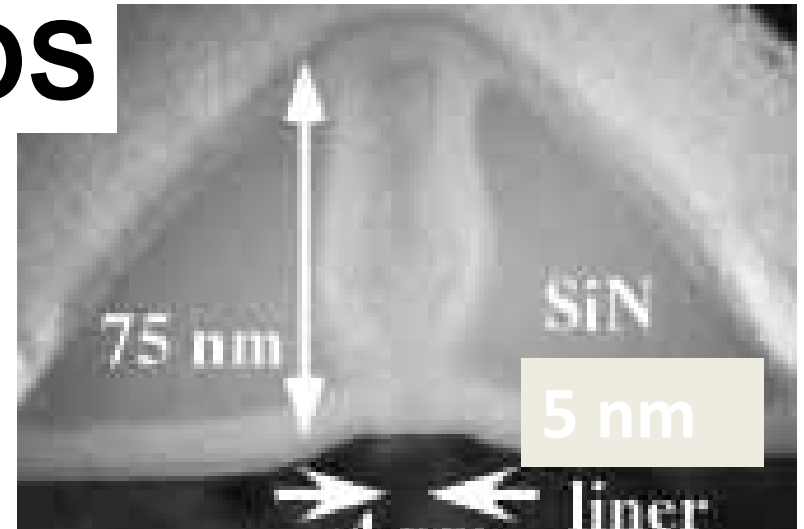


Channel length?

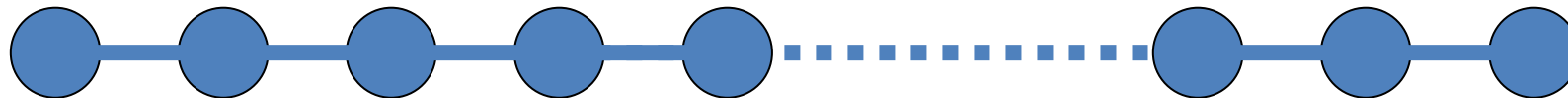


5 nm gate length CMOS

Is a Real Nano Device!!

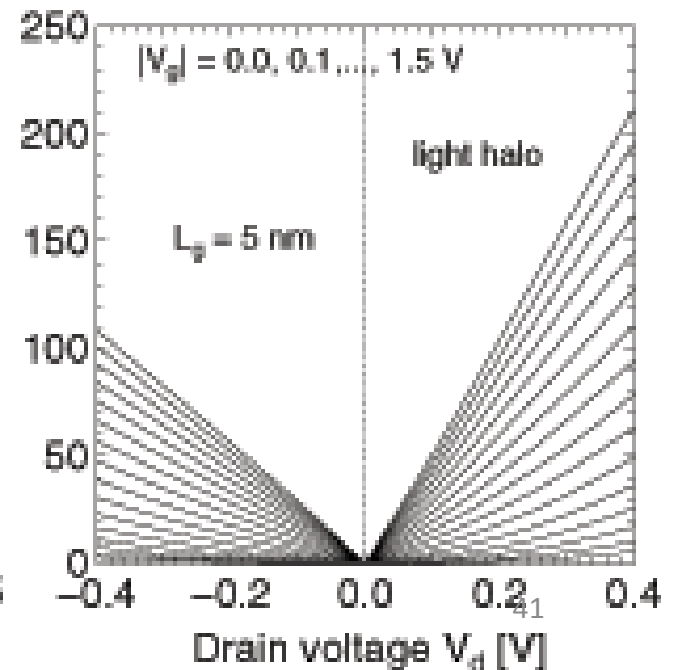
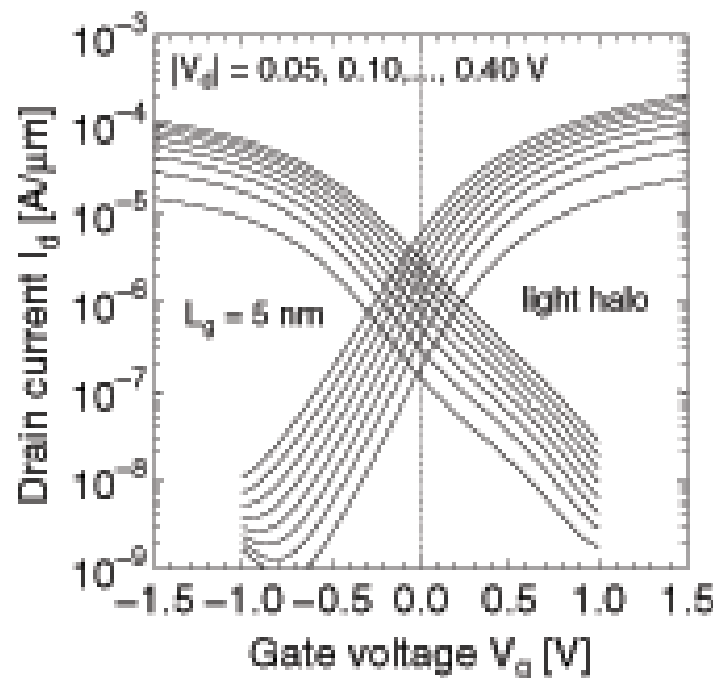


Length of 18 Si atoms



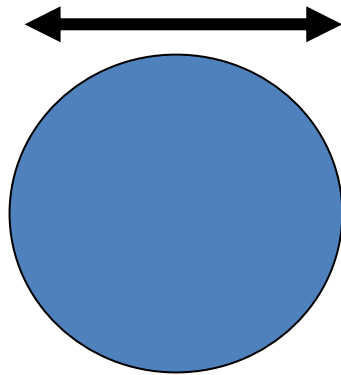
H. Wakabayashi
et.al, NEC

IEDM, 2003



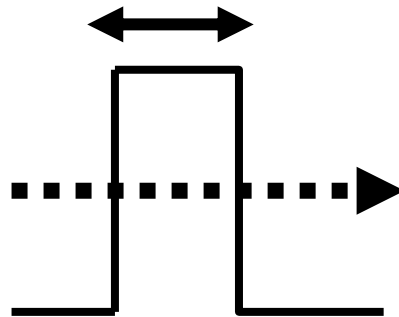
Electron
wave
length

10 nm



Tunneling
distance

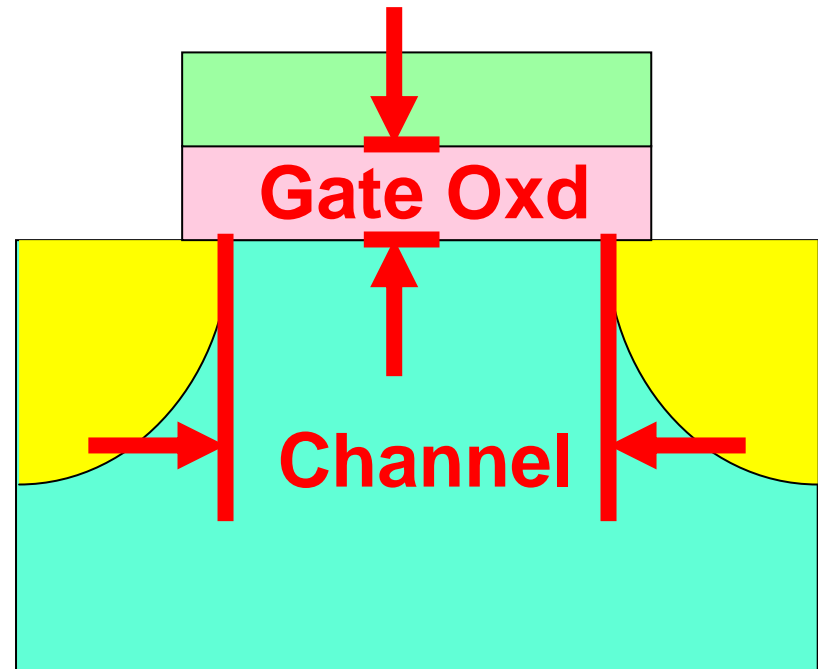
3 nm



Downsizing limit!

Channel length

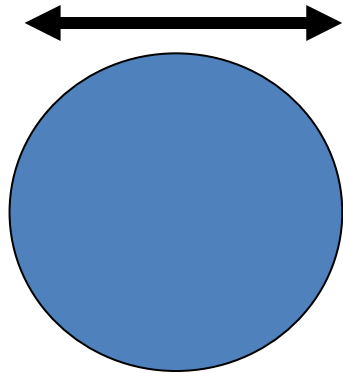
Gate oxide thickness



Prediction now!

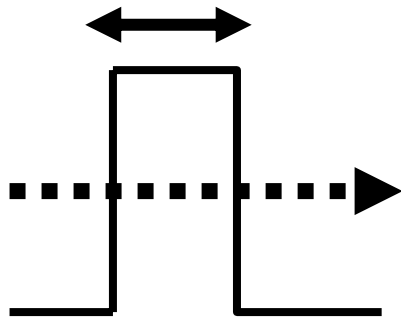
Electron wave length

10 nm



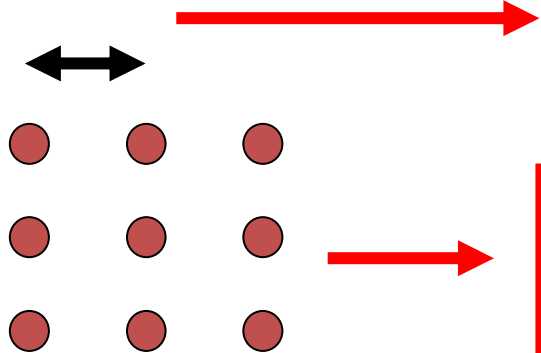
Tunneling distance

3 nm



Atom distance

0.3 nm



MOSFET operation

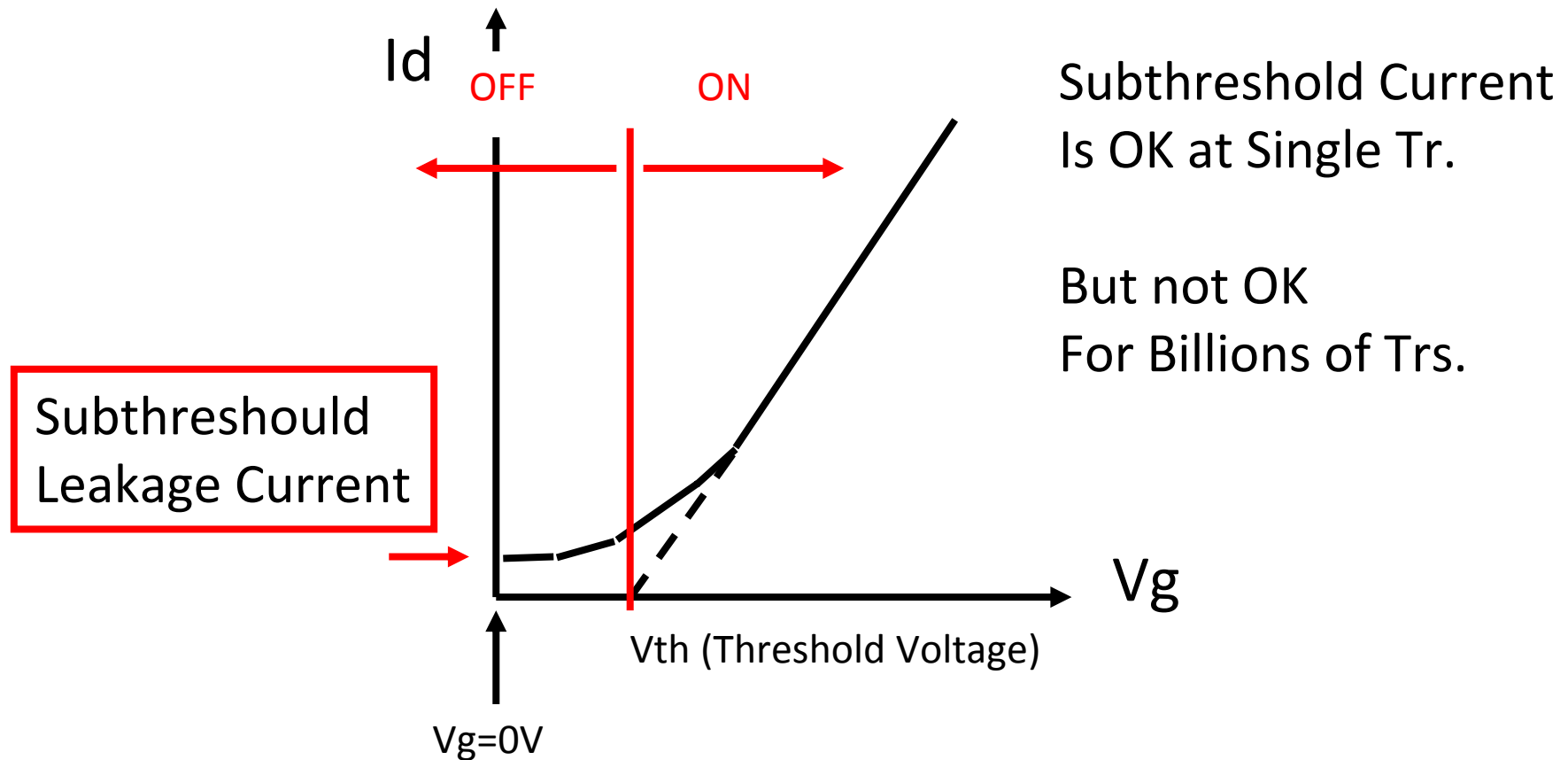
$L_g = 2 \sim 1.5 \text{ nm?}$

**Below this,
no one knows future!**

Maybe, practical limit around 5 nm.

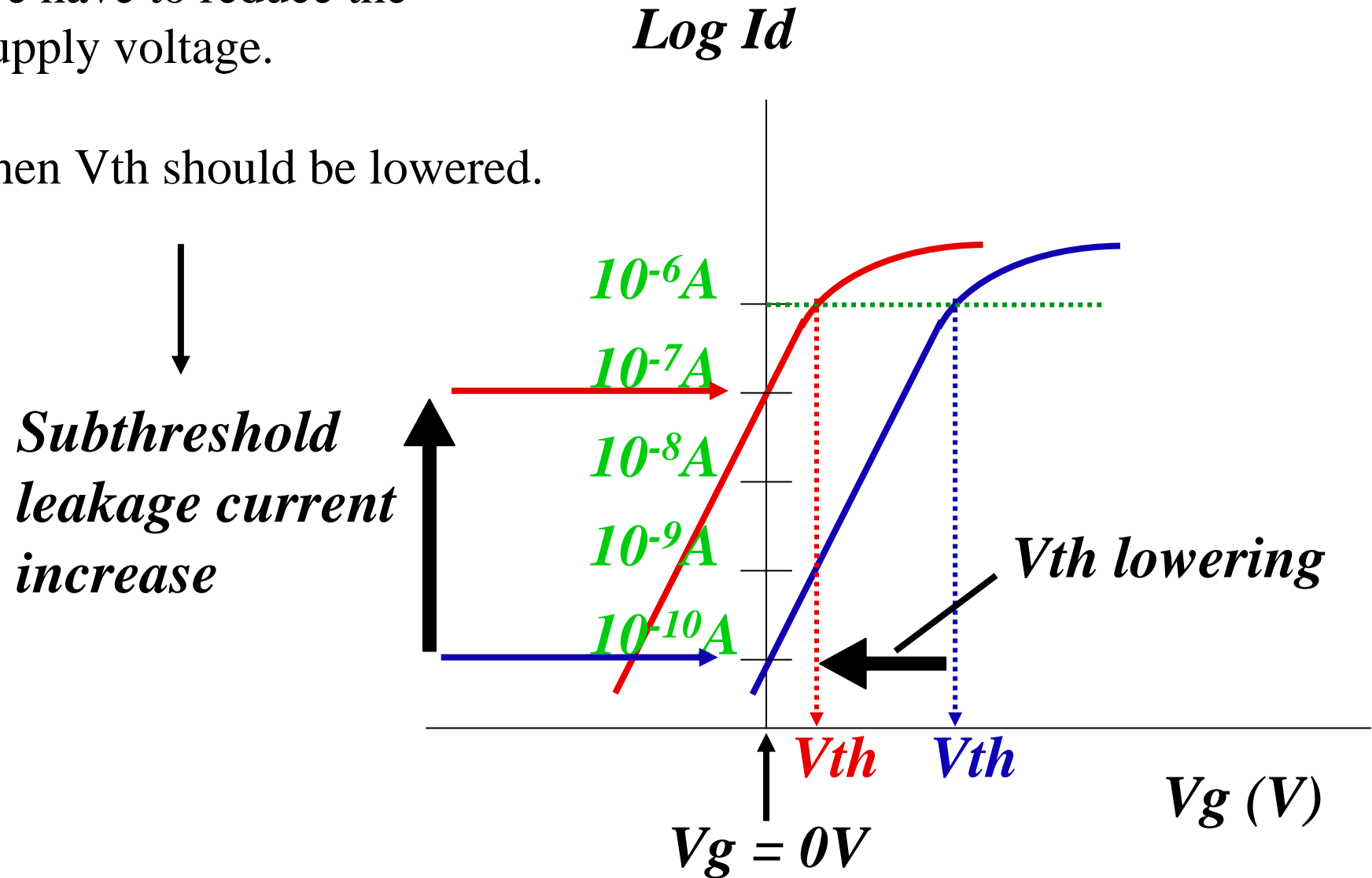
When Gate length Smaller,

→ Subthreshold Leakage Current Larger



We have to reduce the
Supply voltage.

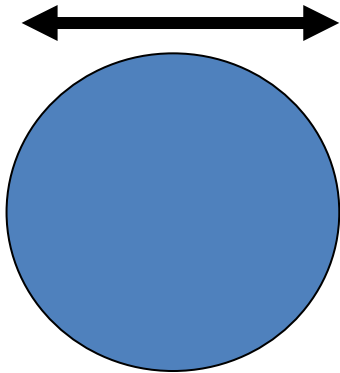
Then V_{th} should be lowered.



Prediction now!

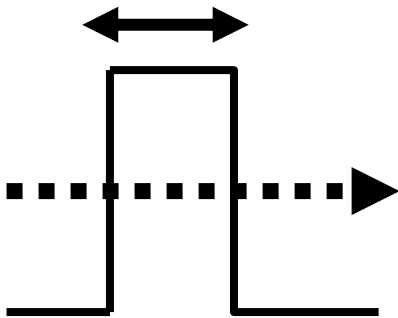
Electron wave length

10 nm



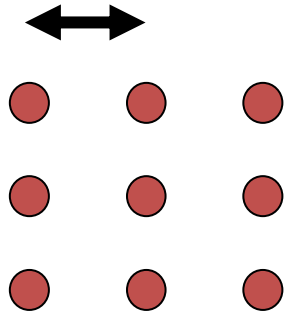
Tunneling distance

3 nm



Atom distance

0.3 nm



Practical limit for integration

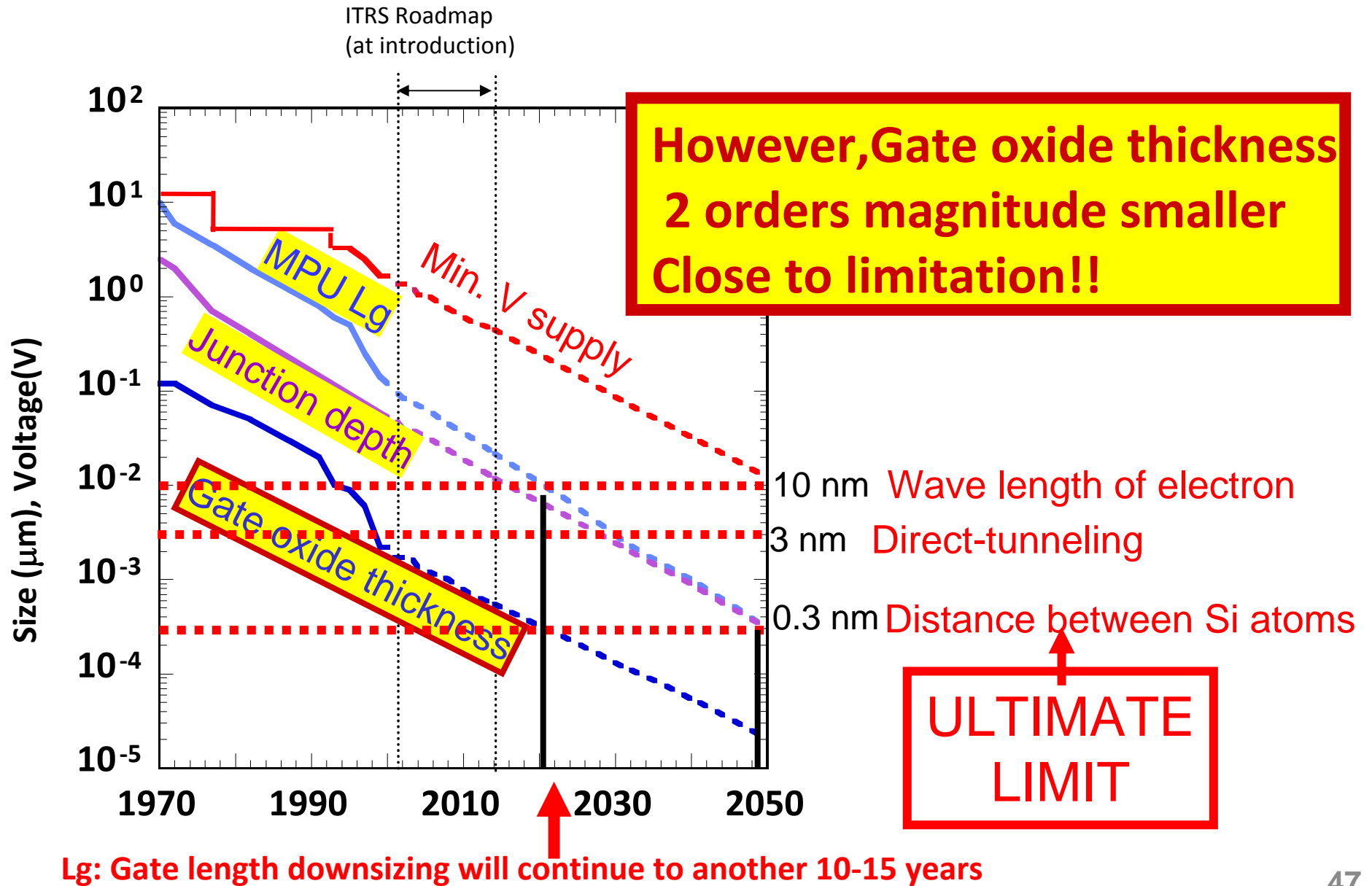
$L_g = 5 \text{ nm?}$

MOSFET operation

$L_g = 2 \sim 1.5 \text{ nm?}$

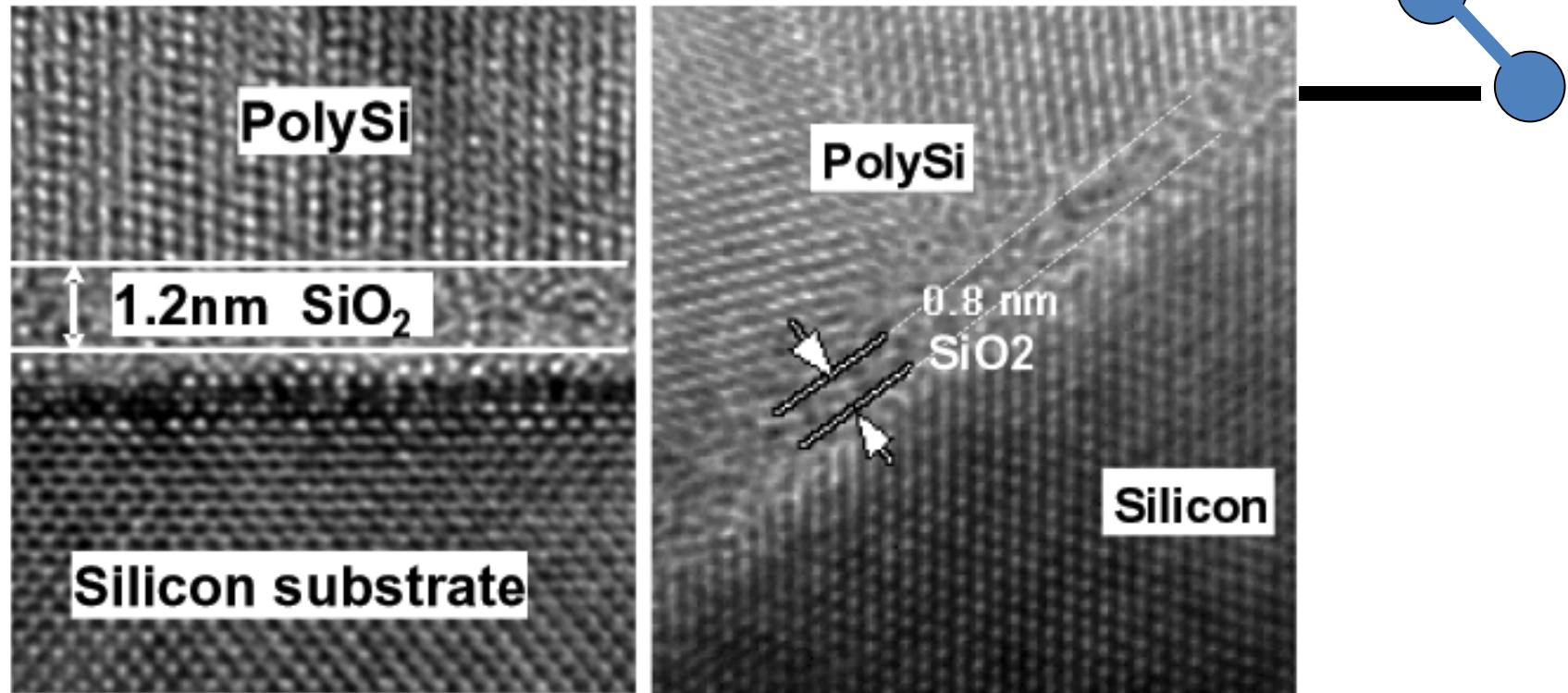
Below this, no one knows future!

Ultimate limitation



0.8 nm Gate Oxide Thickness MOSFETs operates!!

0.8 nm: Distance of 3 Si atoms!!



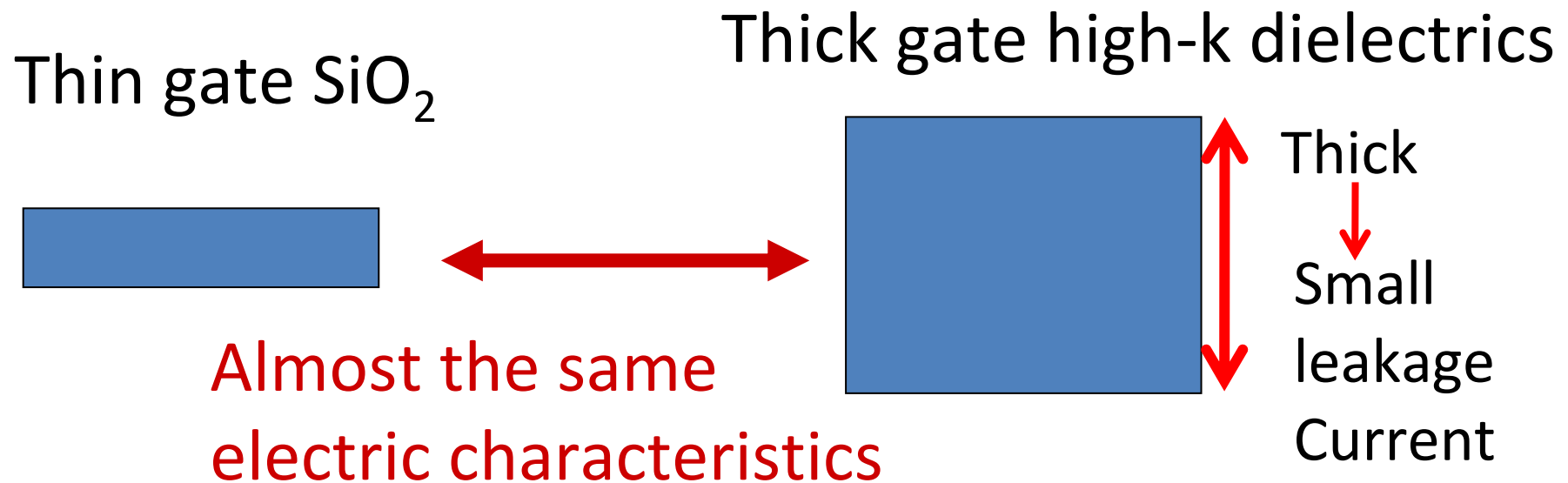
- 1.2nm physical SiO₂ in production (90nm logic node)
- 0.8nm physical SiO₂ in research transistors

By Robert Chau, IWGI 2003

So, we are now in the limitation
of downsizing?

Do you believe this or do not?

There is a solution! **K: Dielectric Constant**
To use high-k dielectrics



However, very difficult and big challenge!

Remember MOSFET had not been realized
without Si/SiO₂!

Choice of High-k elements for oxide

Candidates 														Gas or liquid at 1000 K											
Unstable at Si interface														Radio active											
H																				He					
Li	Be																			B	C	N	O	F	Ne
Na	Mg																			Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr								
Rh	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rb	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe								
Cs	Ba		Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn								
Fr	Ra	Rf	Ha	Sg	Ns	Hs	Mt																		
		La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu									
Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr											

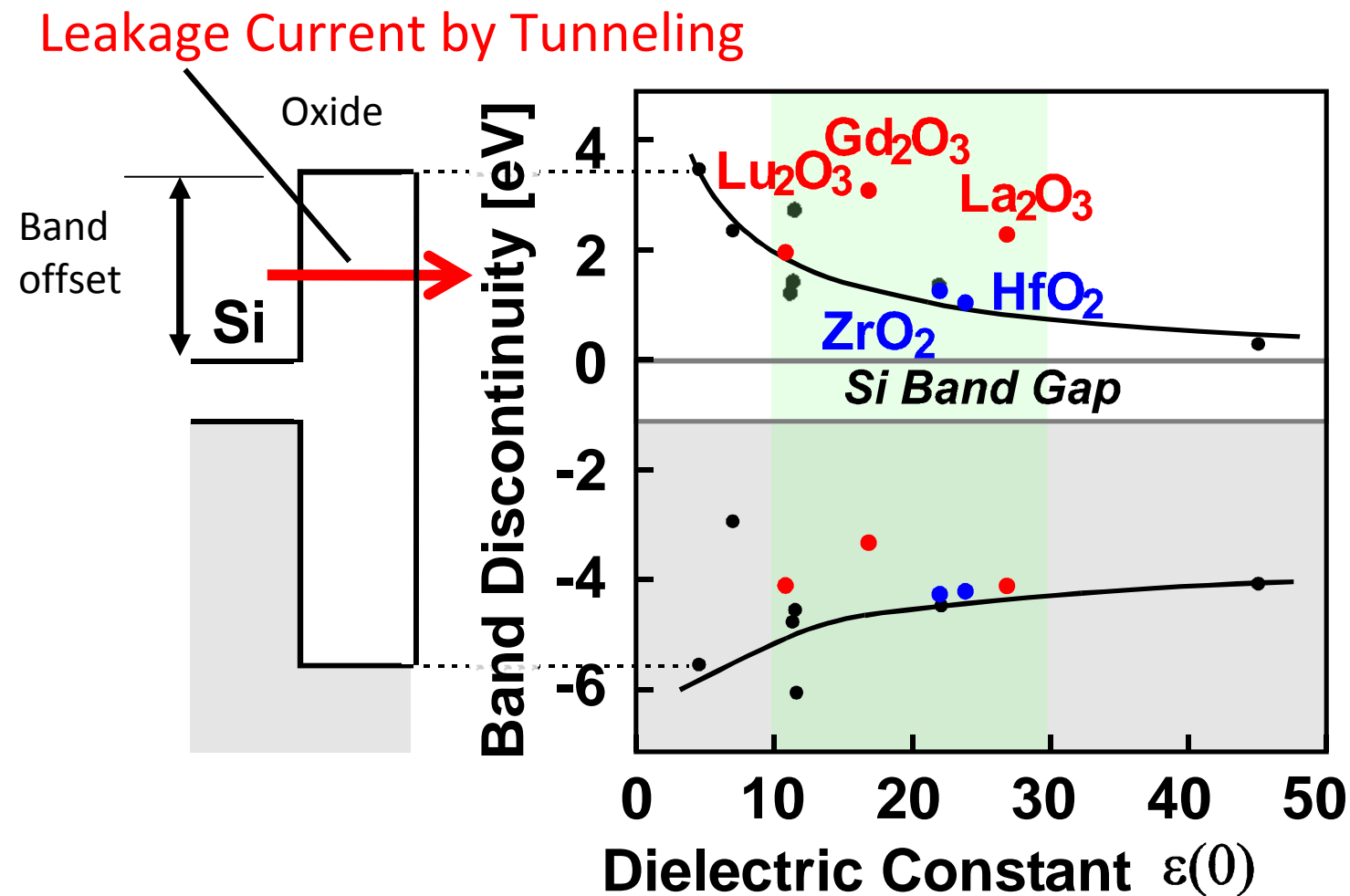
HfO₂ based dielectrics are selected as the first generation materials, because of their merit in

- 1) band-offset,
- 2) dielectric constant
- 3) thermal stability

La₂O₃ based dielectrics are thought to be the next generation materials, which may not need a thicker interfacial layer

R. Hauser, IEDM Short Course, 1999
Hubbard and Schlom, J Mater Res 11 2757 (1996)

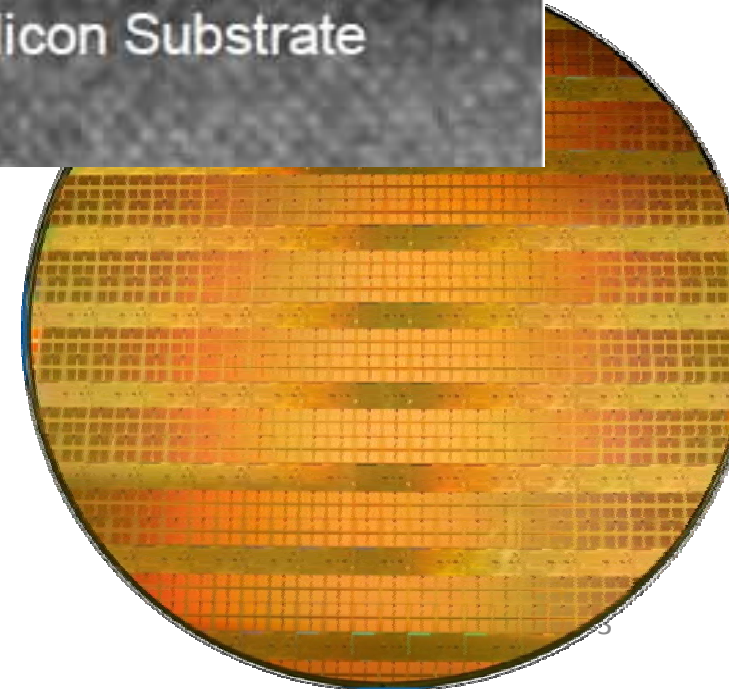
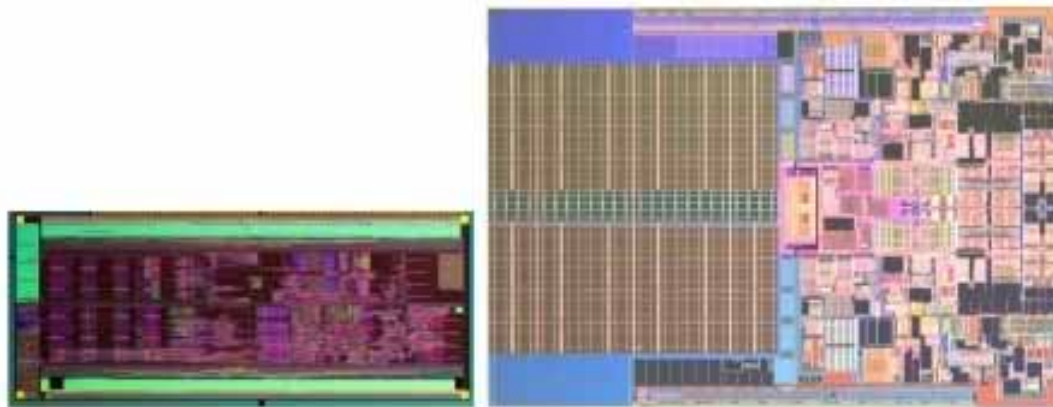
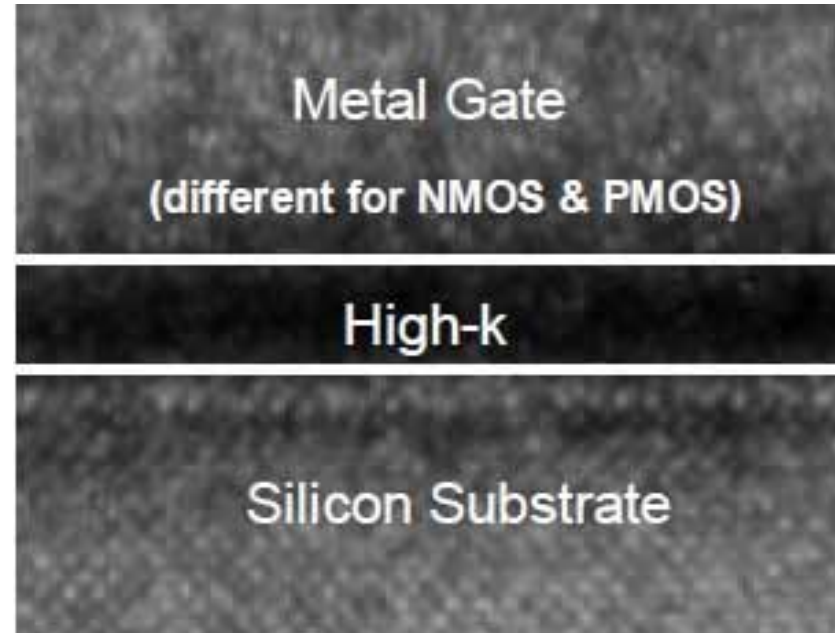
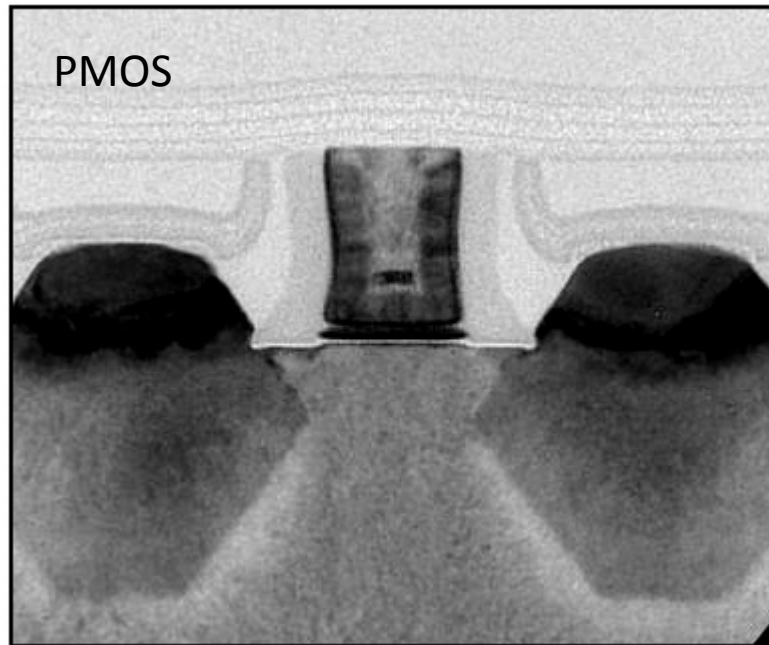
Conduction band offset vs. Dielectric Constant

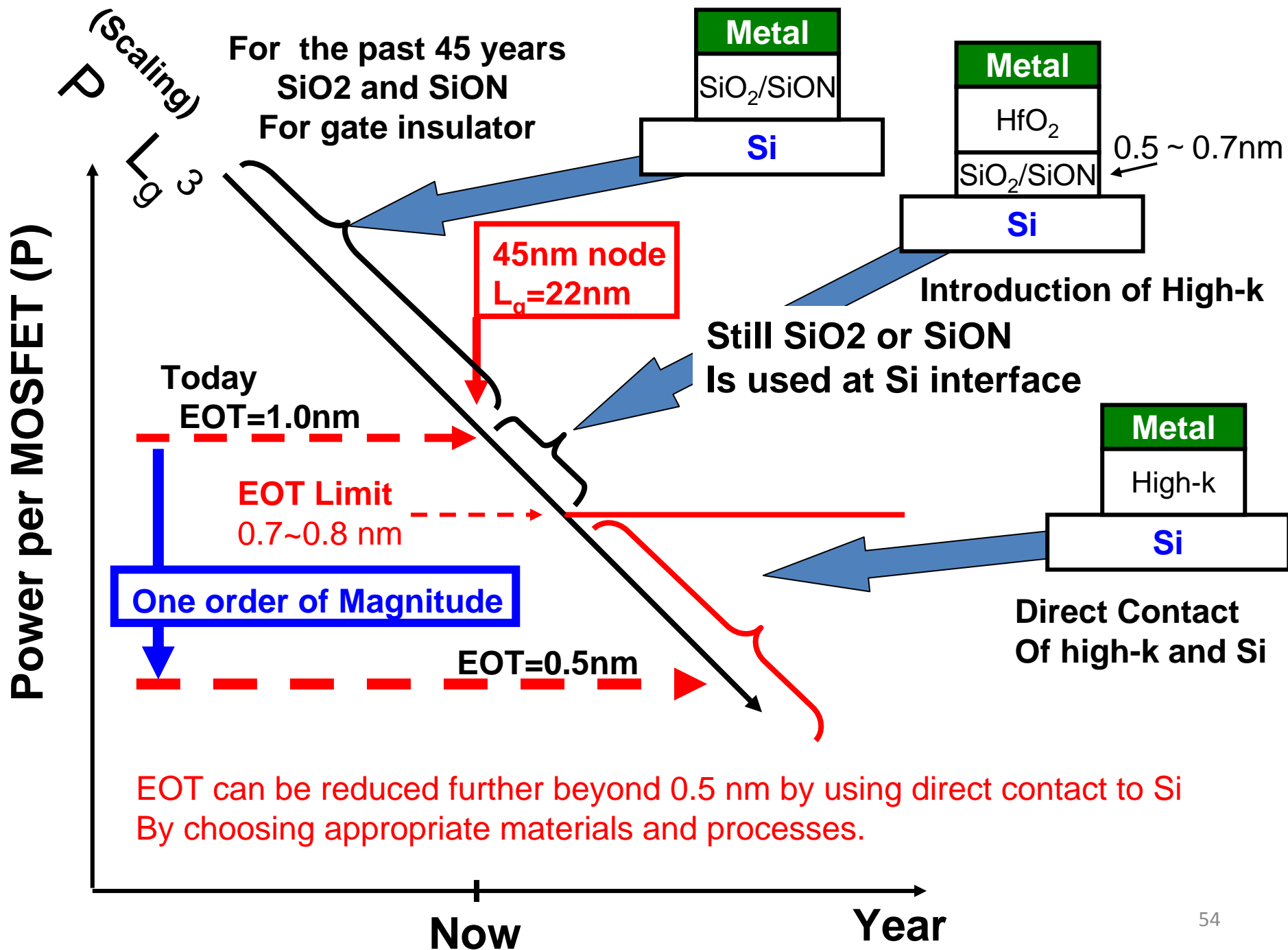


XPS measurement by Prof. T. Hattori, INFOS 2003

High-k gate insulator MOSFETs for Intel: EOT=1nm

EOT: Equivalent Oxide Thickness





Choice of High-k elements for oxide

Candidates 														Gas or liquid at 1000 K											
Unstable at Si interface														Radio active											
H																				He					
Li	Be																			B	C	N	O	F	Ne
Na	Mg																			Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fc	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr								
Rh	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rb	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe								
Cs	Ba		Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn								
Fr	Ra			Rf	Ha	Sg	Ns	Hs	Mt																
		La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu									
Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr											

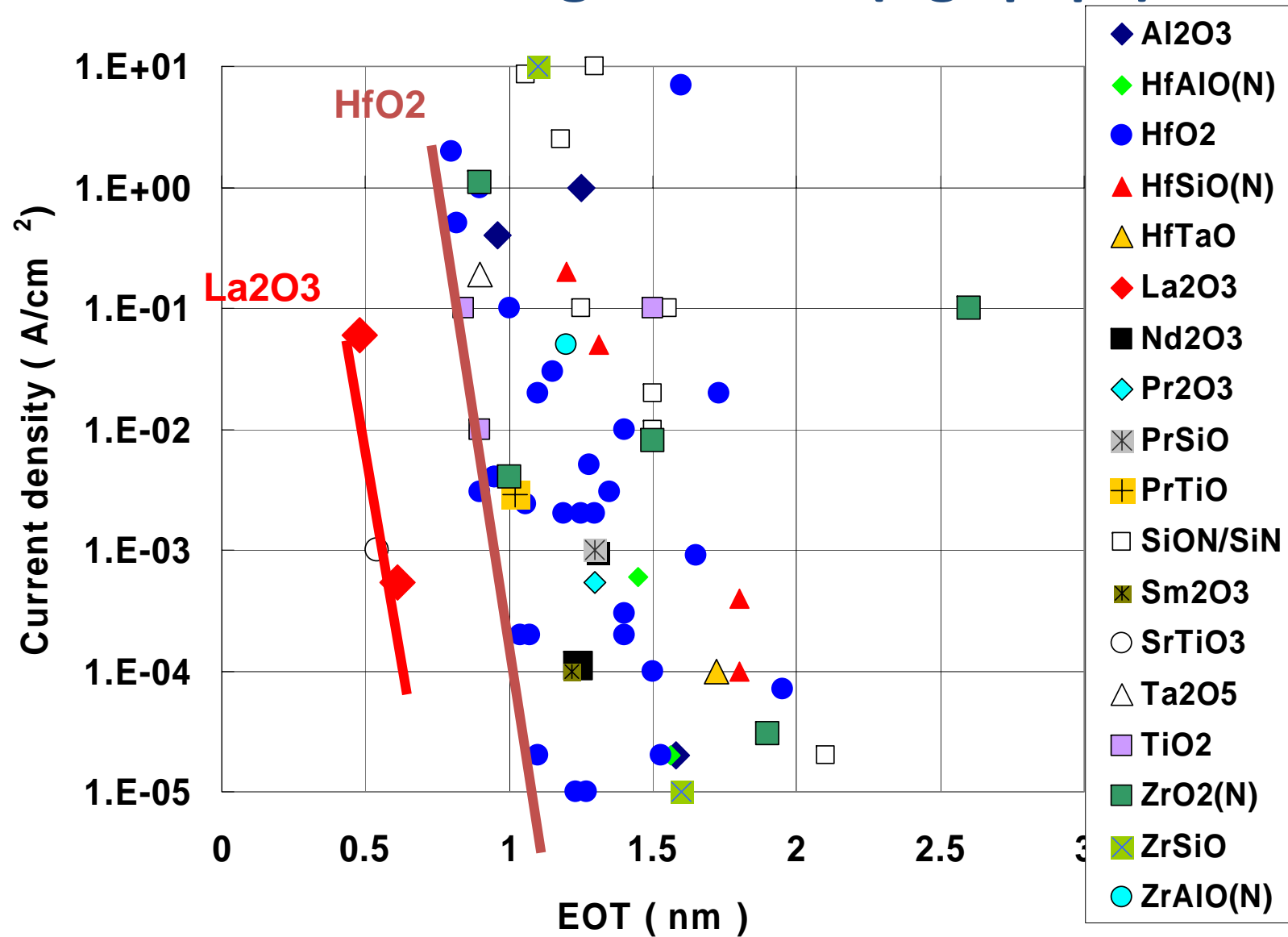
HfO₂ based dielectrics are selected as the first generation materials, because of their merit in

- 1) band-offset,
- 2) dielectric constant
- 3) thermal stability

La₂O₃ based dielectrics are thought to be the next generation materials, which may not need a thicker interfacial layer

R. Hauser, IEDM Short Course, 1999
Hubbard and Schlom, J Mater Res 11 2757 (1996)

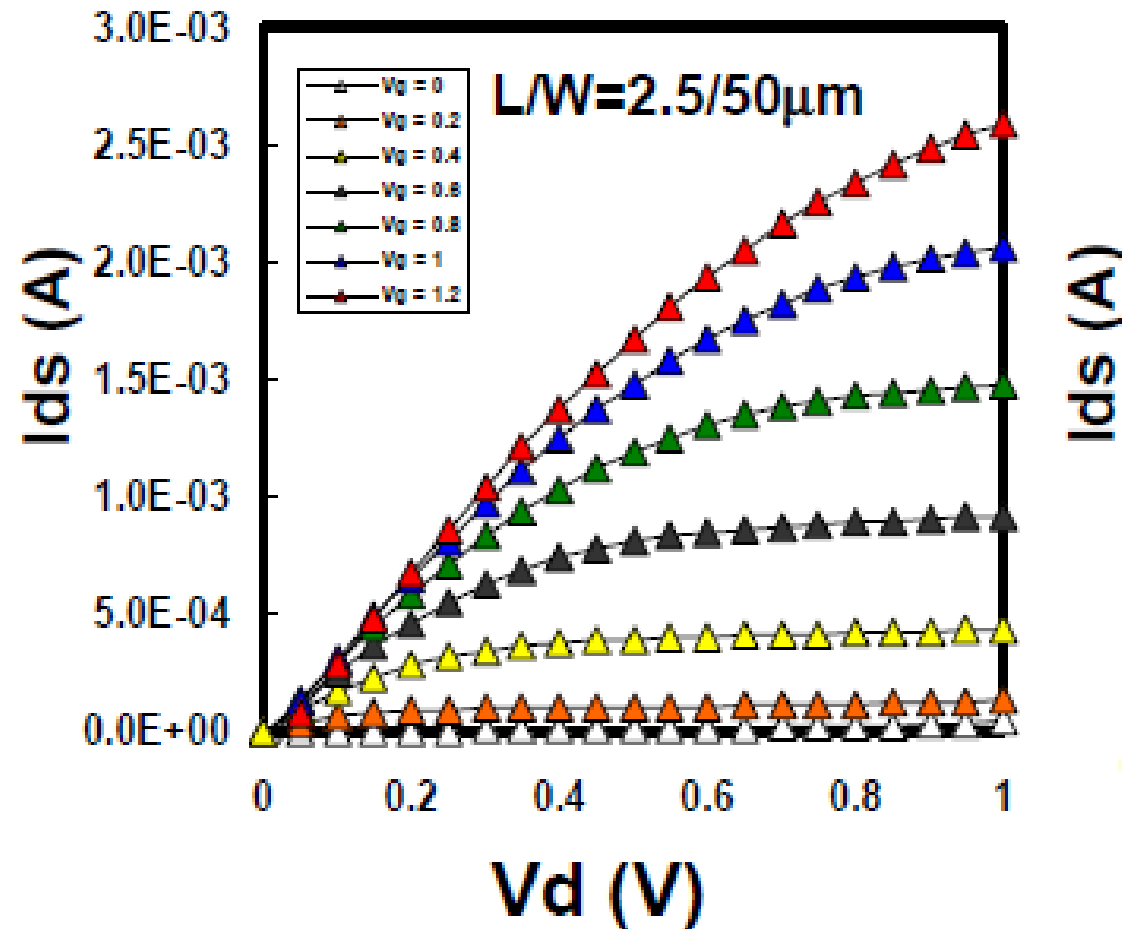
Gate Leakage vs EOT, ($V_g = |1|V$)



EOT = 0.48 nm

Our results

Transistor with La₂O₃ gate insulator



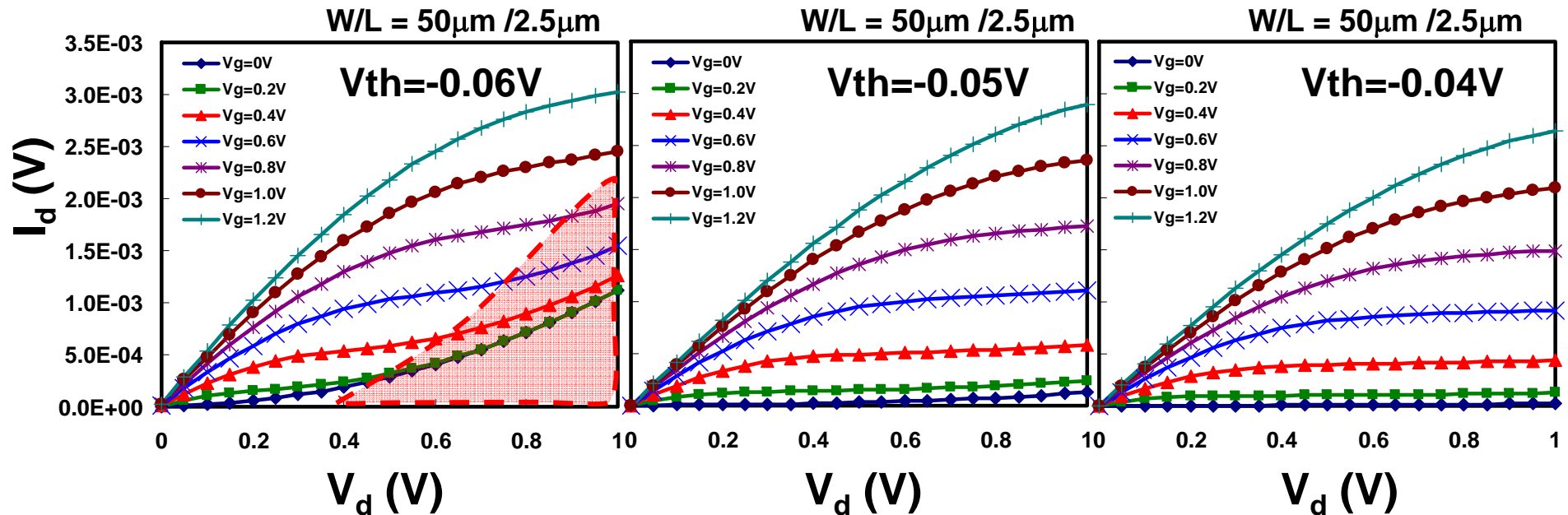
EOT=0.37nm

La2O3

EOT=0.37nm

EOT=0.40nm

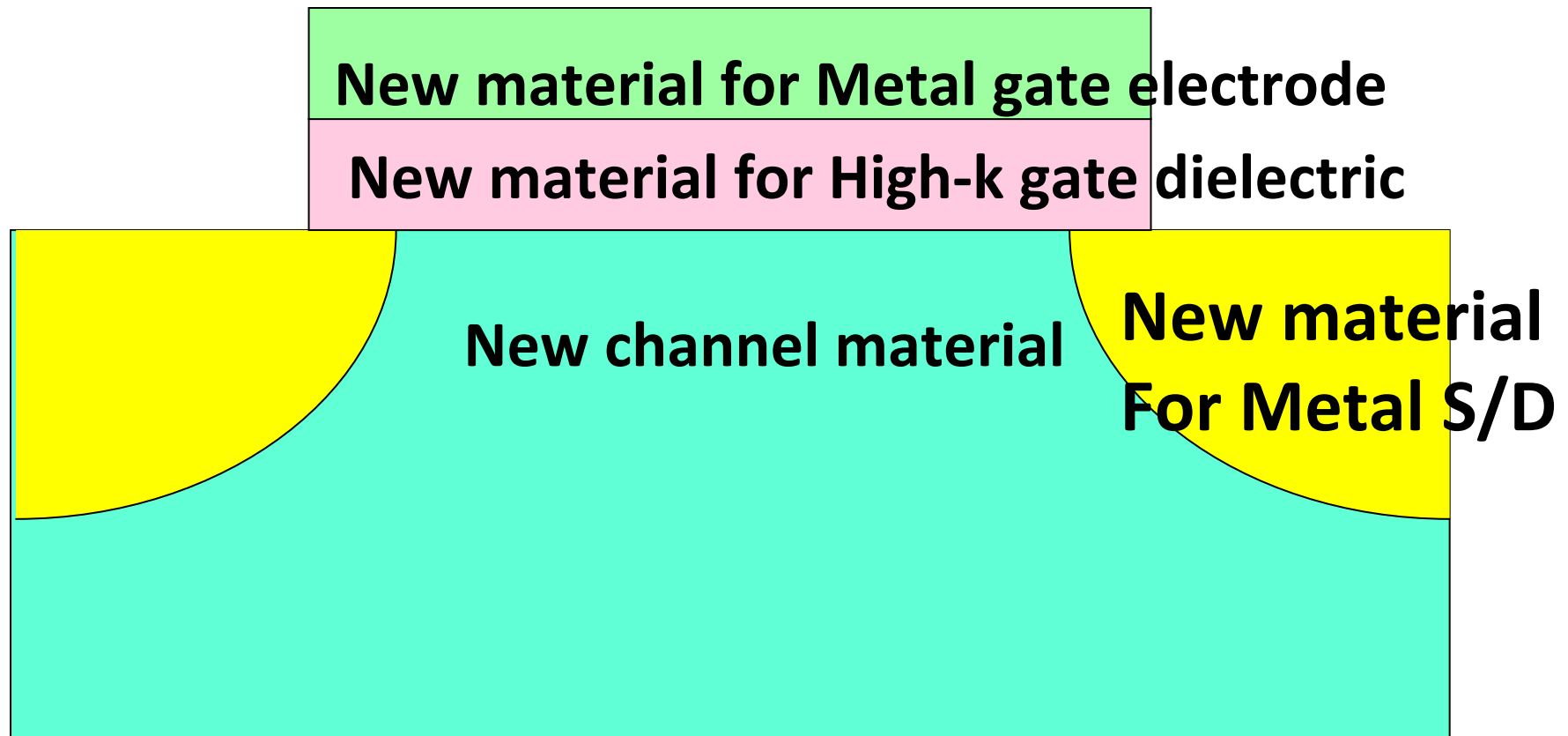
EOT=0.48nm



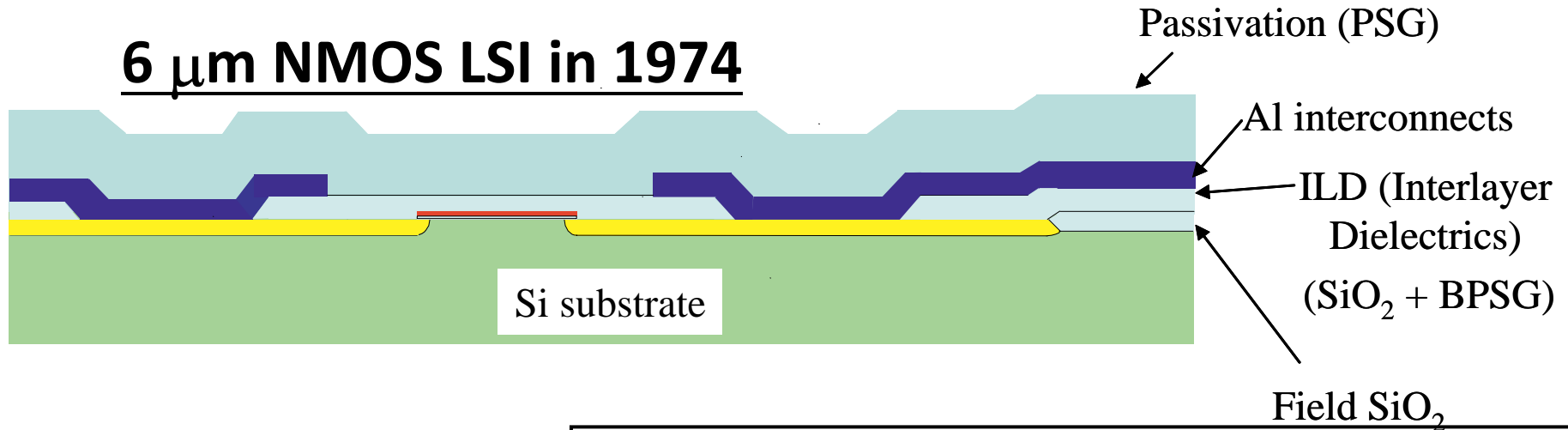
← 0.48 → 0.37nm Increase of I_d at 30%

New material research will give us many future possibilities and the most important for Nano-CMOS!

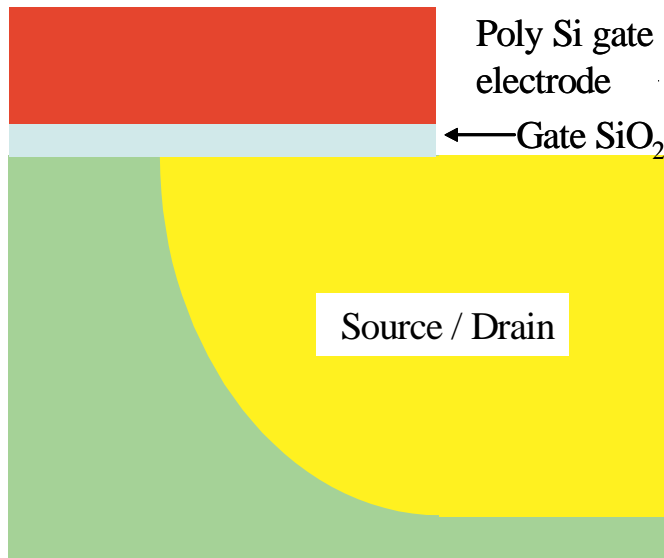
Not only for high-k!



6 μm NMOS LSI in 1974



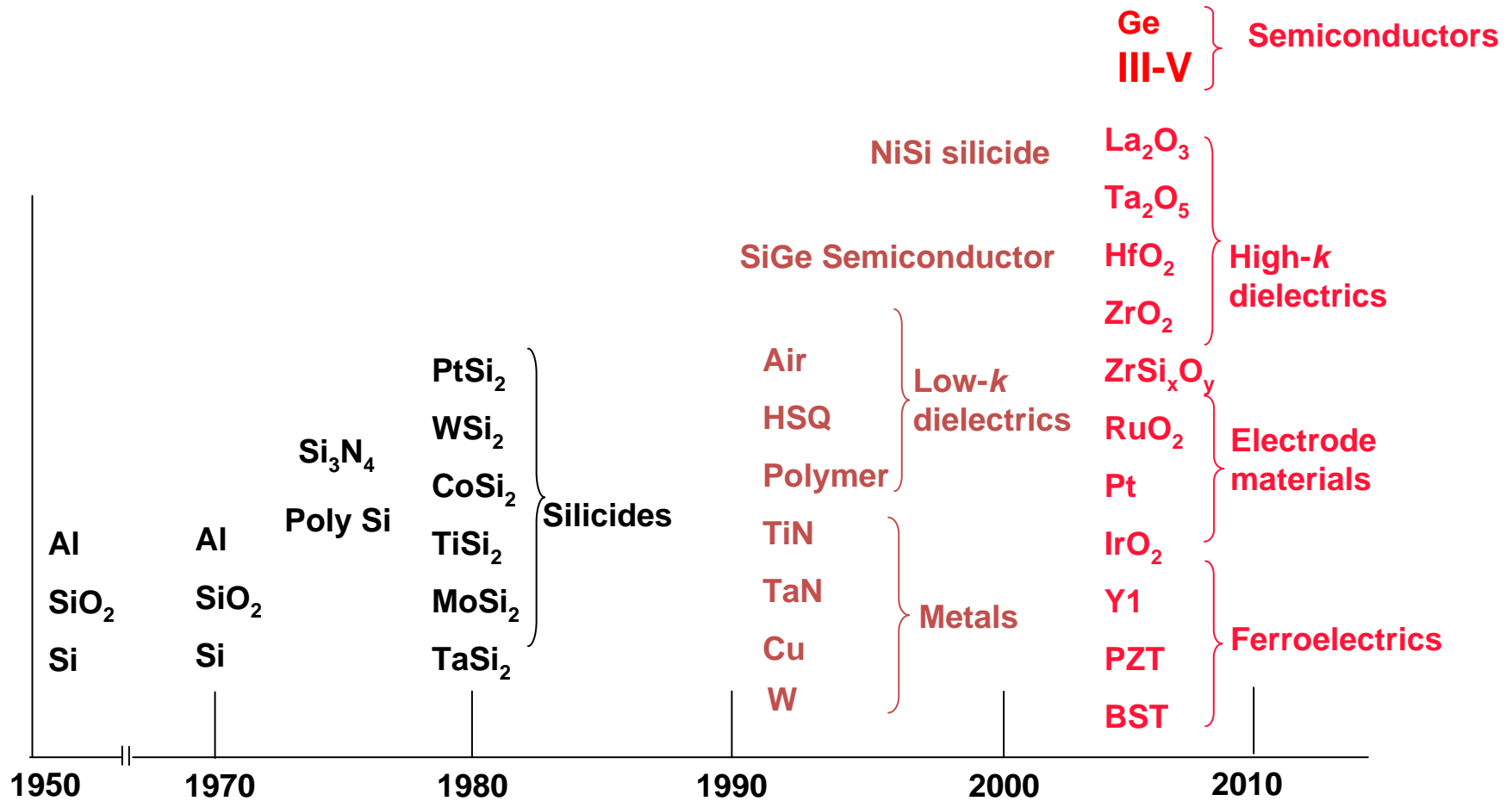
magnification
↓



<u>Layers</u>	<u>Materials</u>	<u>Atoms</u>
1. Si substrate	1. Si	1. Si
2. Field oxide	2. SiO ₂	2. O
3. Gate oxide	3. BPSG	3. P
4. Poly Si	4. Al	4. B
5. S/D	5. PSG	5. Al
6. Interlayer		(H, N, Cl)
7. Aluminum		60
8. Passivation		

New materials

Just examples!
Many other candidates

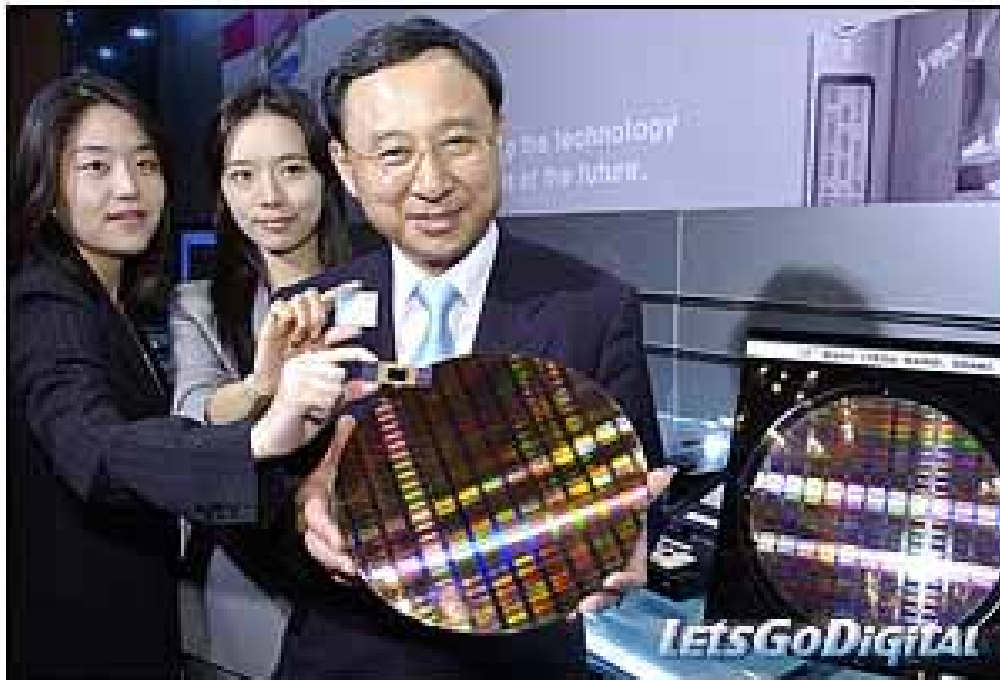


Y. Nishi, Si Nano Workshop, 2006,

(S. Sze, Based on invited talk at Stanford Univ., Aug. 1999)

Now: After 45 Years from the 1st single MOSFETs

*32 Gb and 16Gb NAND,
SAMSUNG*



Samsung's NAND flash trend

Capacity	Node	1 st Fabrication	Production
512Mbit	120nm	2000	2001
1Gbit	100nm	2001	2002
2Gbit	90nm	2002	2003
4Gbit	70nm	2003	2004
8Gbit	60nm	2004	2005
16Gbit	50nm	2005	2006
32Gbit	40nm		
256Gbit	20nm		

Even Tbit would be possible in future!

Already 32 Gbit:

larger than that of world population
comparable for the numbers of neurons
in human brain

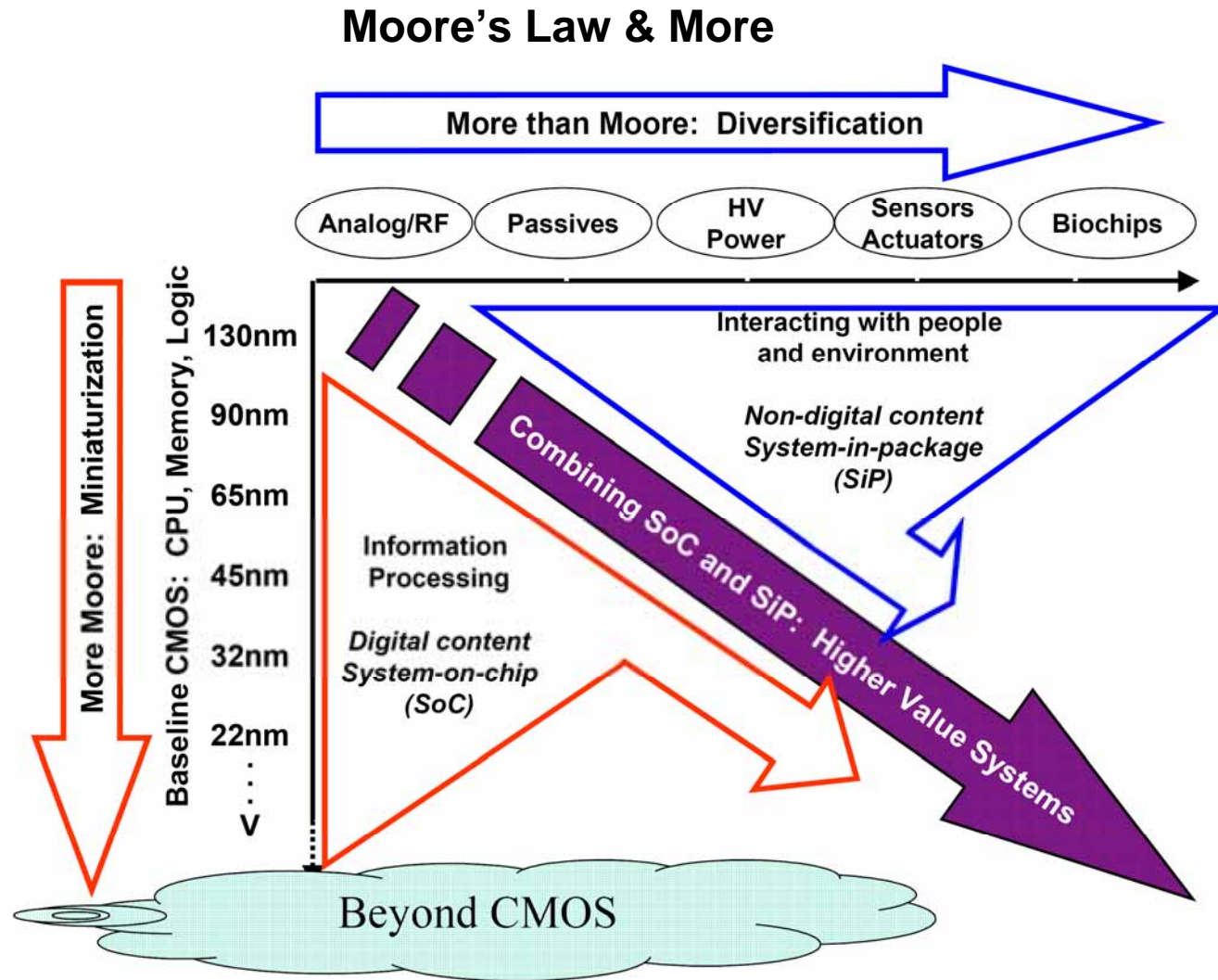
Samsung announced 256 Gbit will be produced in 2010.

Only 4 years from now.

256Gbit: larger than those of # of stars in galaxies



More Moore and More than Moore

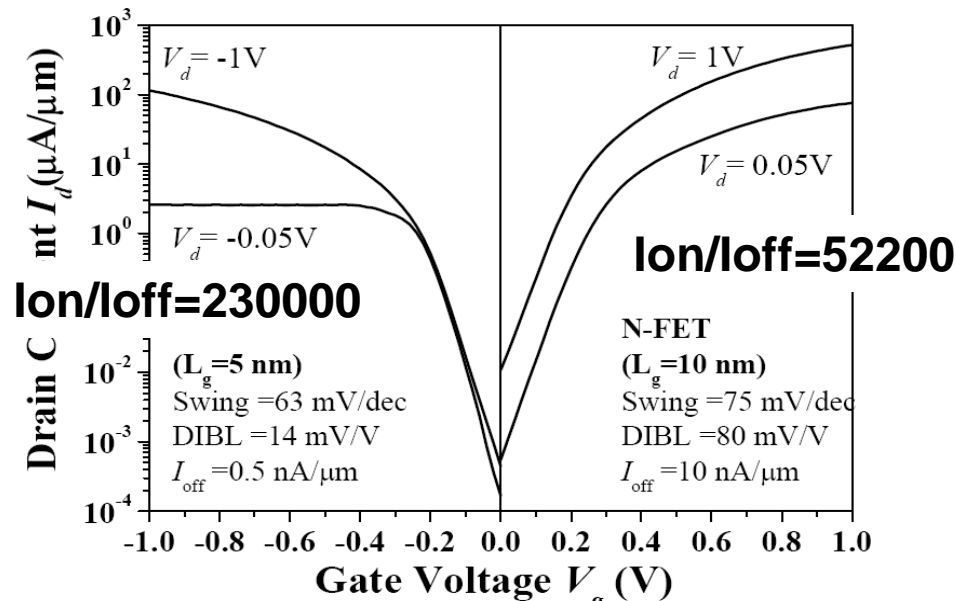
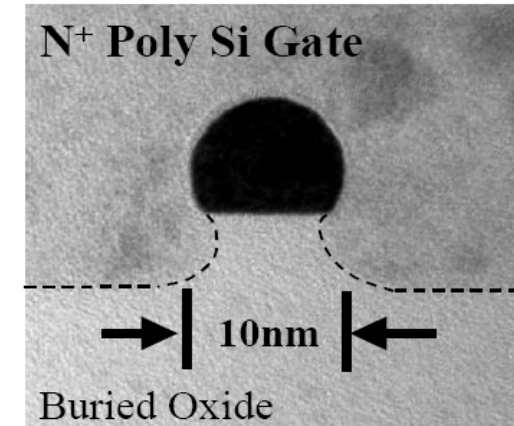
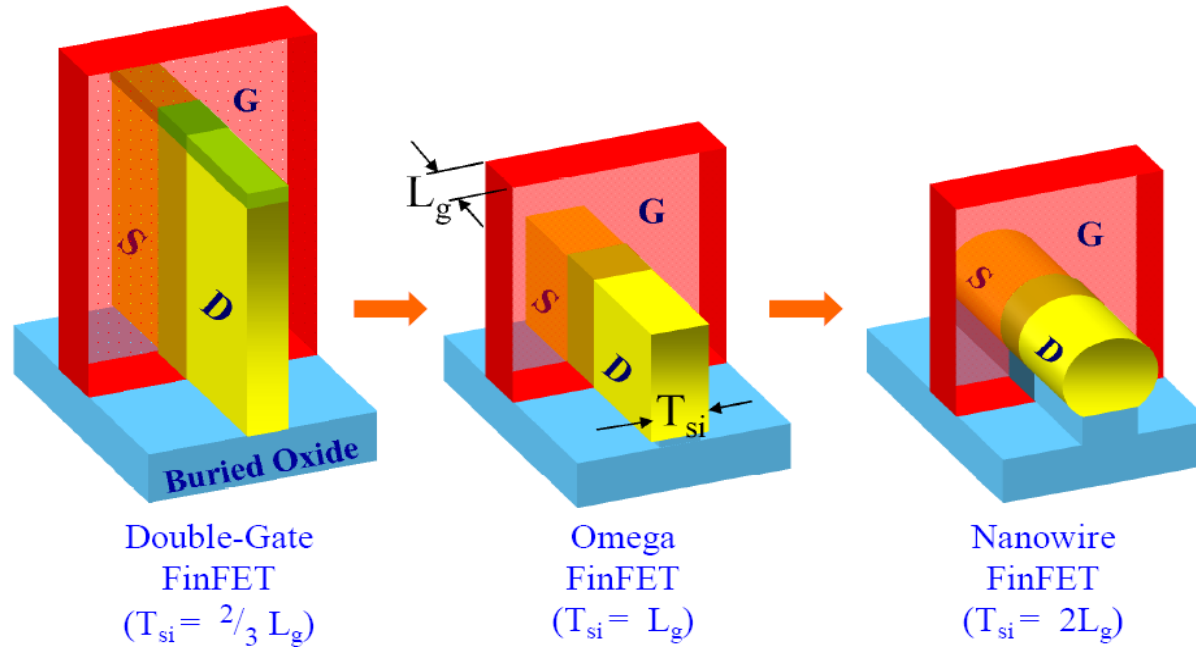


Question what is the other side of the cloud?

ITRS 2005 Edition

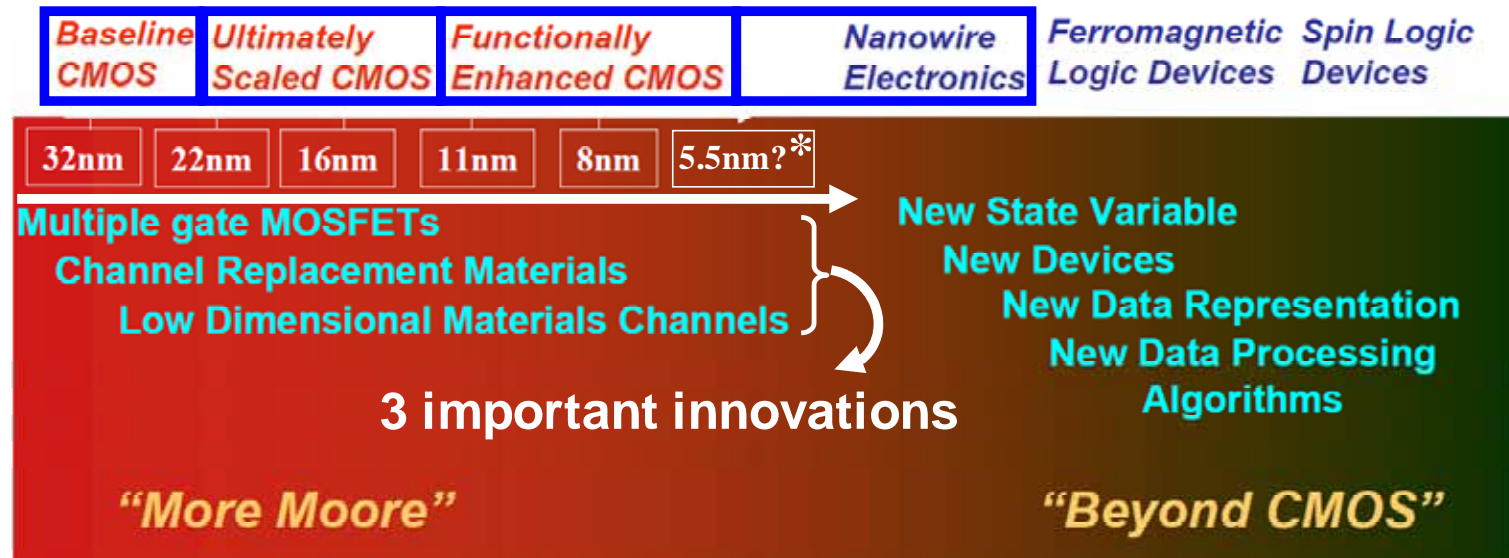
http://strj-jeita.elisasp.net/pdf_ws_2005nendo/9A_WS2005IRC_Ishiuchi.pdf

FinFET to Nanowire



Channel conductance is well controlled by Gate even at $L=5\text{nm}$

- There will be still 4~6 cycles (or technology generations) left until we reach 11 ~ 5.5 nm technologies, at which we will reach down-scaling limit, in some year between 2020-30 (H. Iwai, IWJT2008).
- Even After reaching the down-scaling limit, we could still continue R & D, seeking sufficiently higher I_{d-sat} under low V_{dd} .
- Two candidates have emerged for R & D
 1. Nanowire/tube MOSFETs
 2. Alternative channel MOSFETs (III-V, Ge)
- Other Beyond CMOS devices are still in the cloud.



ITRS figure
edited by Iwai

Si nanowire FET with Semi-1D Ballistic Transport

Merit of Si-nanowire

Source: Y. Lee., T. Nagata., K. Kakushima., K. Shiraishi, and H. Iwai, IWDTF 2008, Tokyo, November, 2008

Trade off

Carrier scattering probability

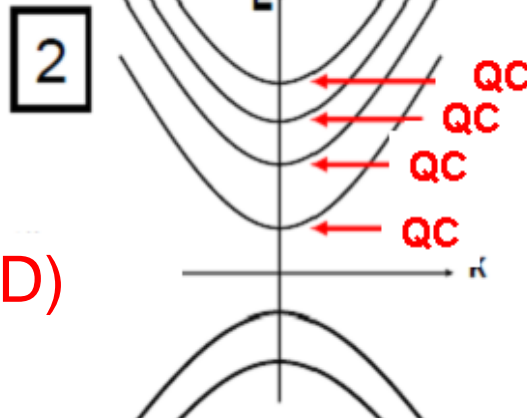
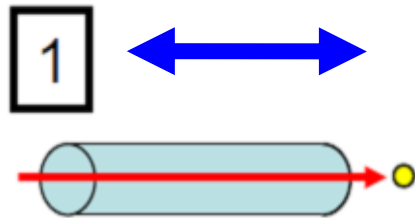
Small

Large

of quantum channel

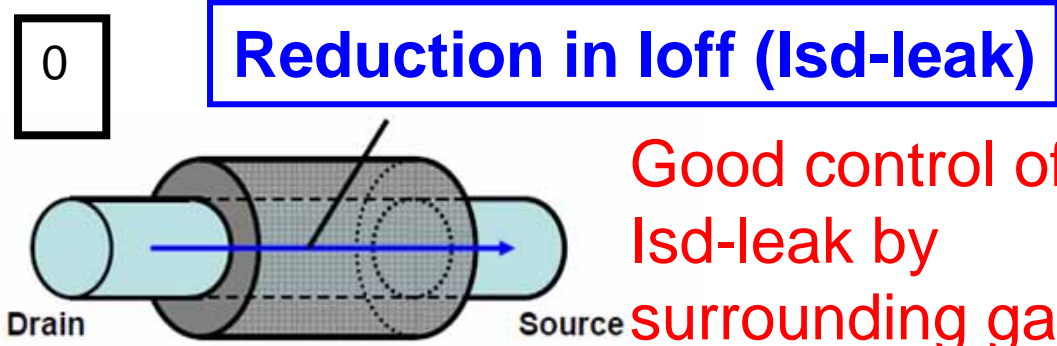
Small

Large



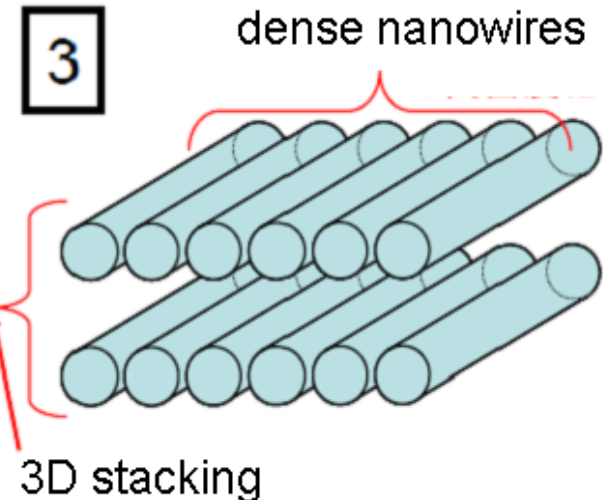
High Conduction (1D)
 $G_0 = 77.8 \mu S / \text{wire}$

Multiple quantum channel (QC) used for conduction



Good control of Isd-leak by surrounding gate

Increase in Ion (Id-sat)



High-density lateral and vertical integration

Selection of MOSFET structure for high conduction:
Nano-wire or Nano-tube FETs is promising

3 methods to realize High-conduction at Low voltage

M1 . Use 1D ballistic conduction

M2 . Increase number of quantum channel

M3 . Increase the number of wire or tube per area
3D integration of wire and tubes

For suppression of I_{off} , the Nanowire/tube is also good.

1D conduction per one quantum channel:

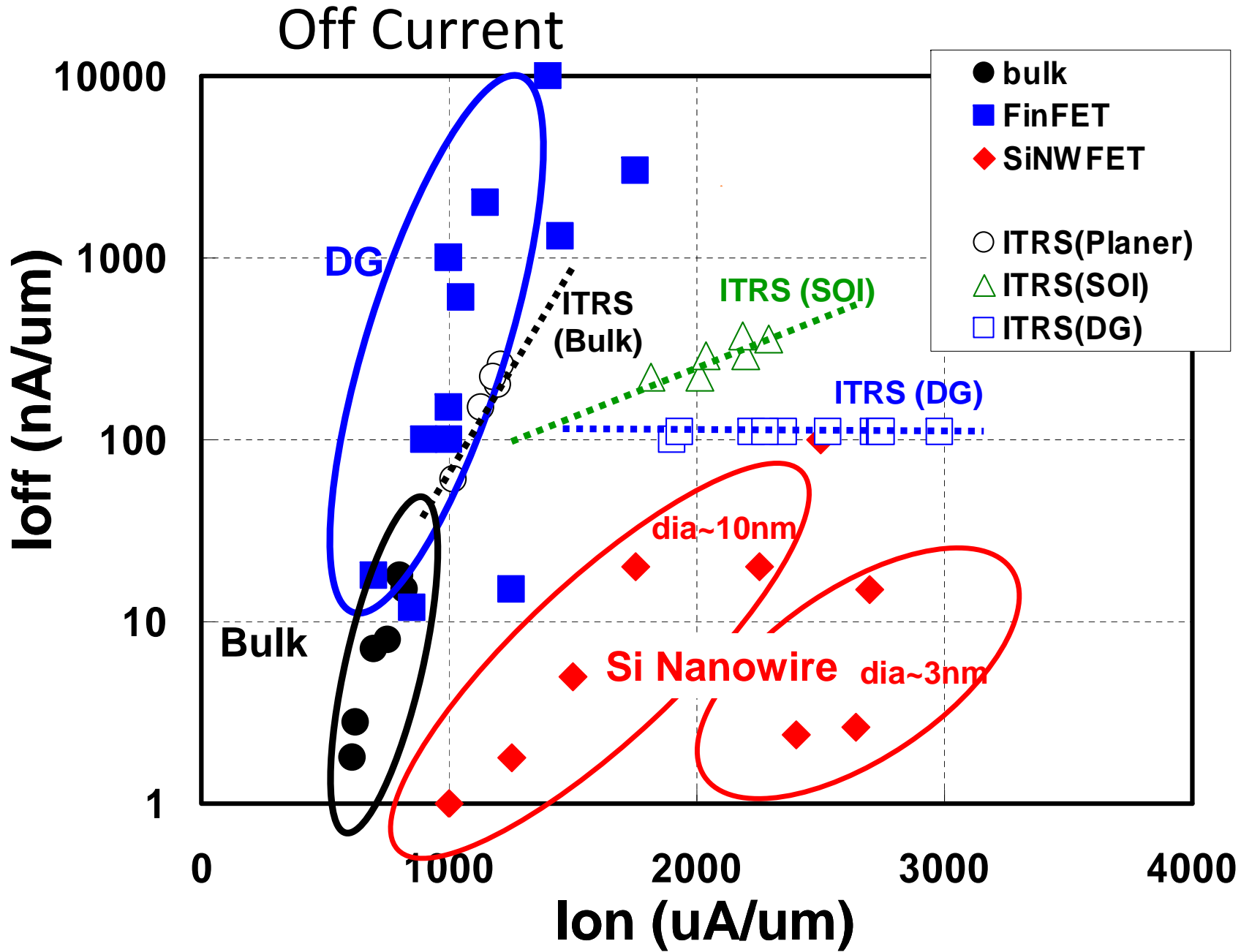
$$G = 2e^2/h = 77.8 \mu\text{S/wire or tube}$$

regardless of gate length and channel material

That is $77.8 \mu\text{A/wire}$ at 1V supply

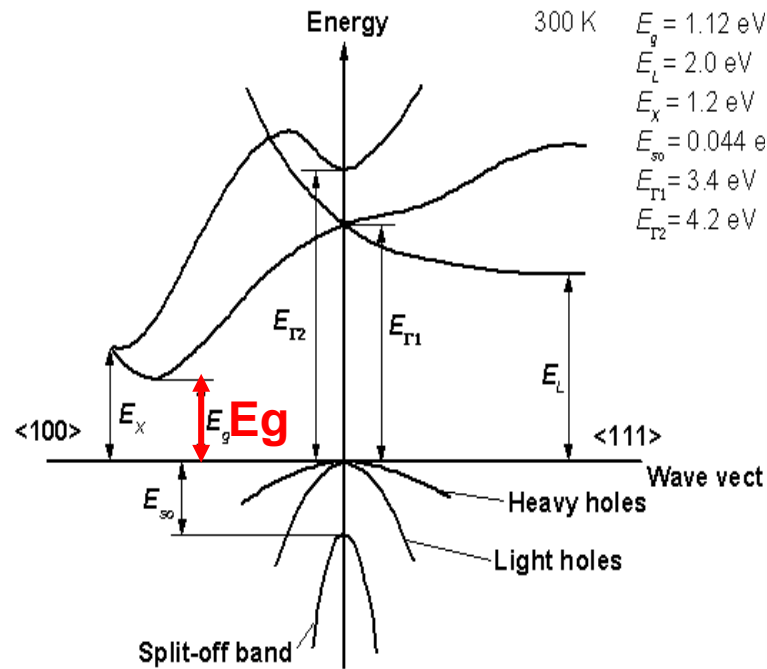
This an extremely high value

However, already 20mA/wire was obtained experimentally
by Samsung

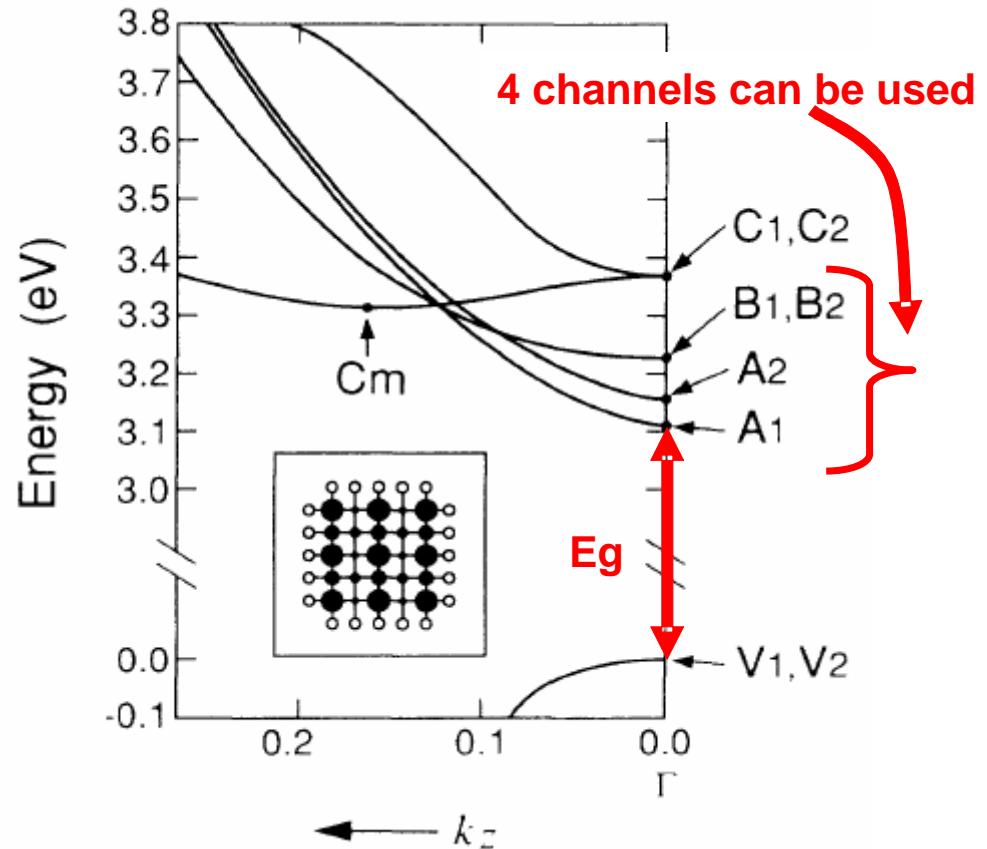


Increase the Number of quantum channels

By Prof. Shiraishi of Tsukuba univ.



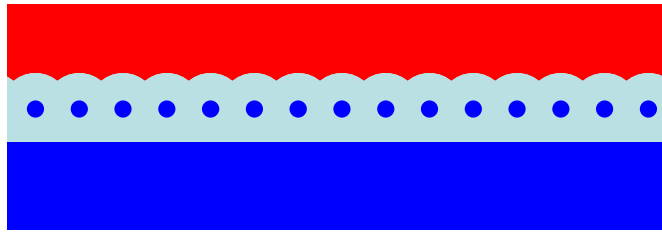
Energy band of Bulk Si



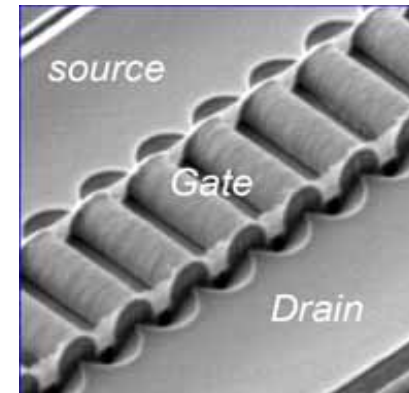
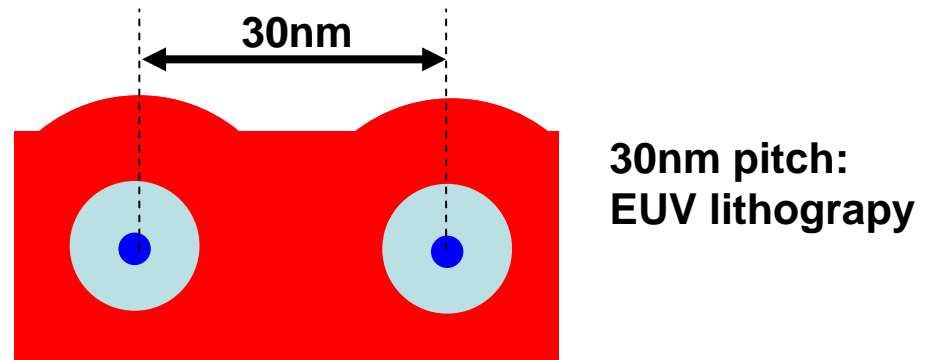
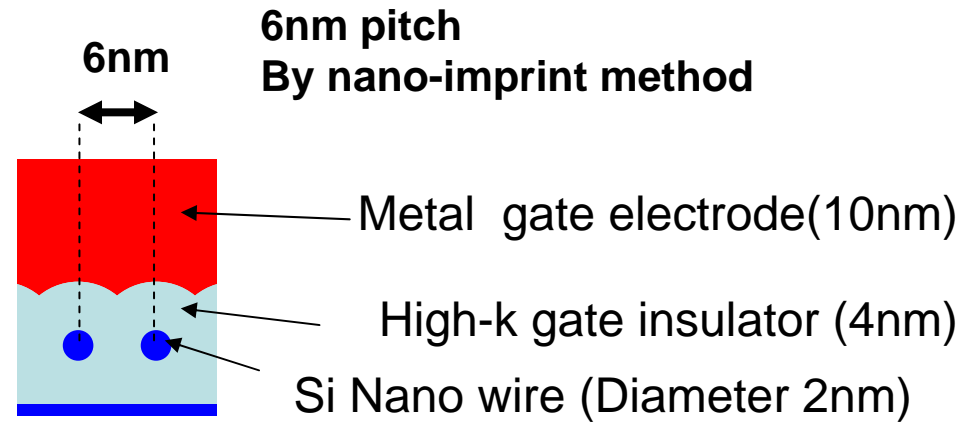
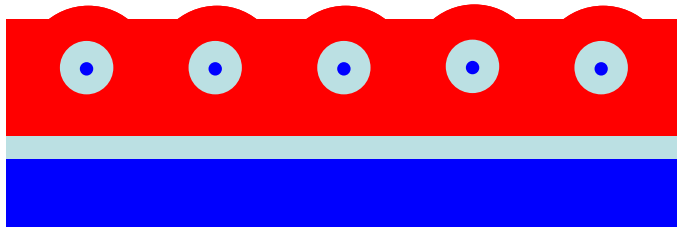
Energy band of 3 x 3 Si wire

Maximum number of wires per 1 μm

Front gate type MOS 165 wires / μm

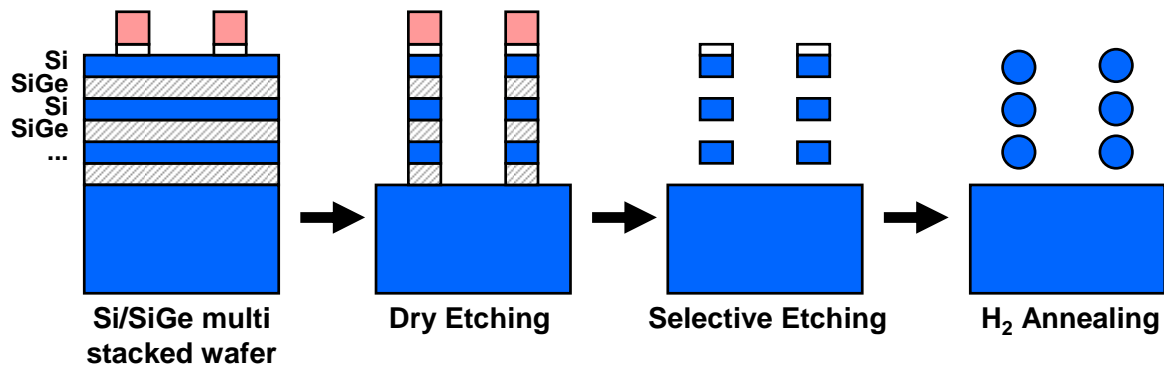
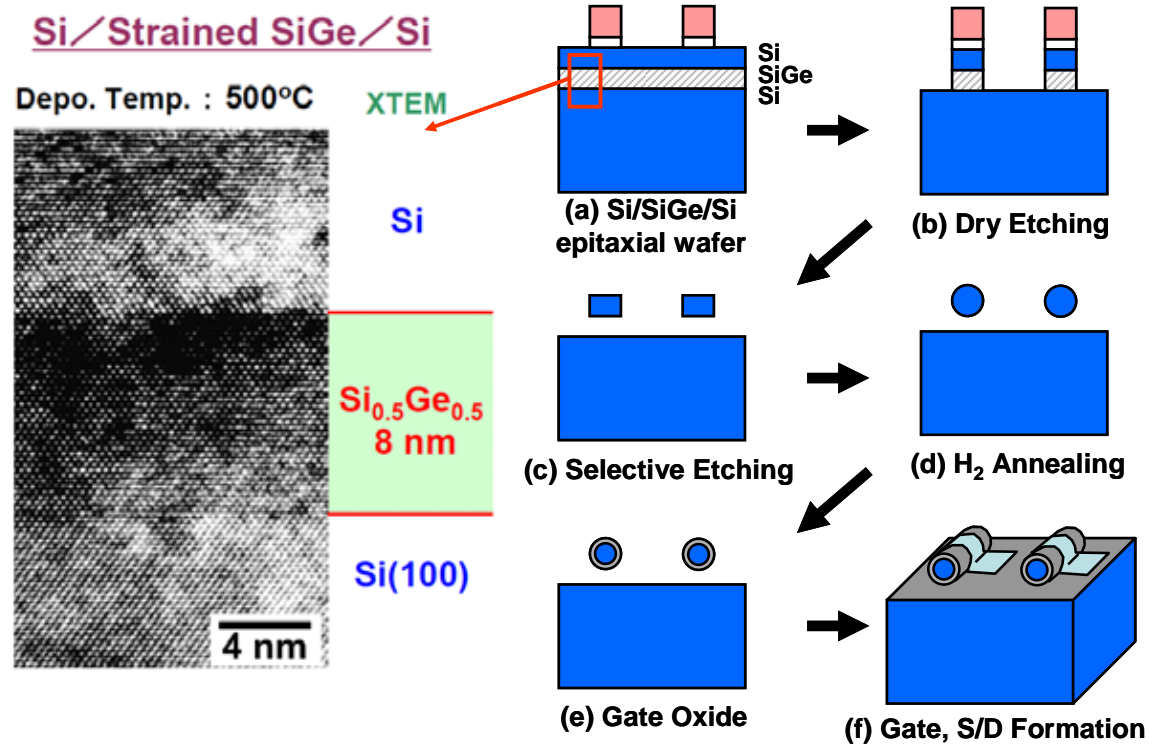


Surrounded gate type MOS 33 wires / μm

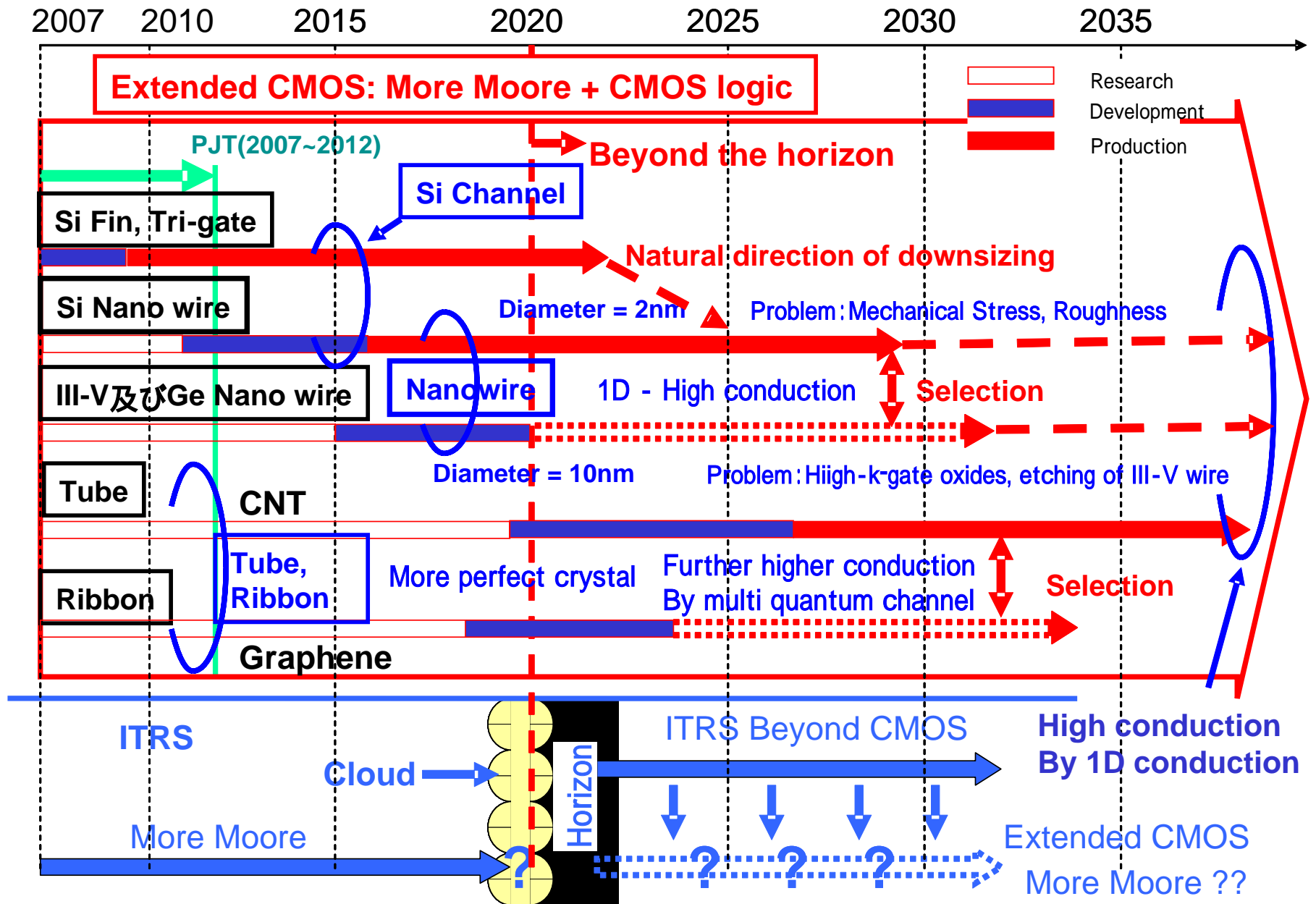


Surrounded gate MOS

Increase the number of wires towards vertical dimension



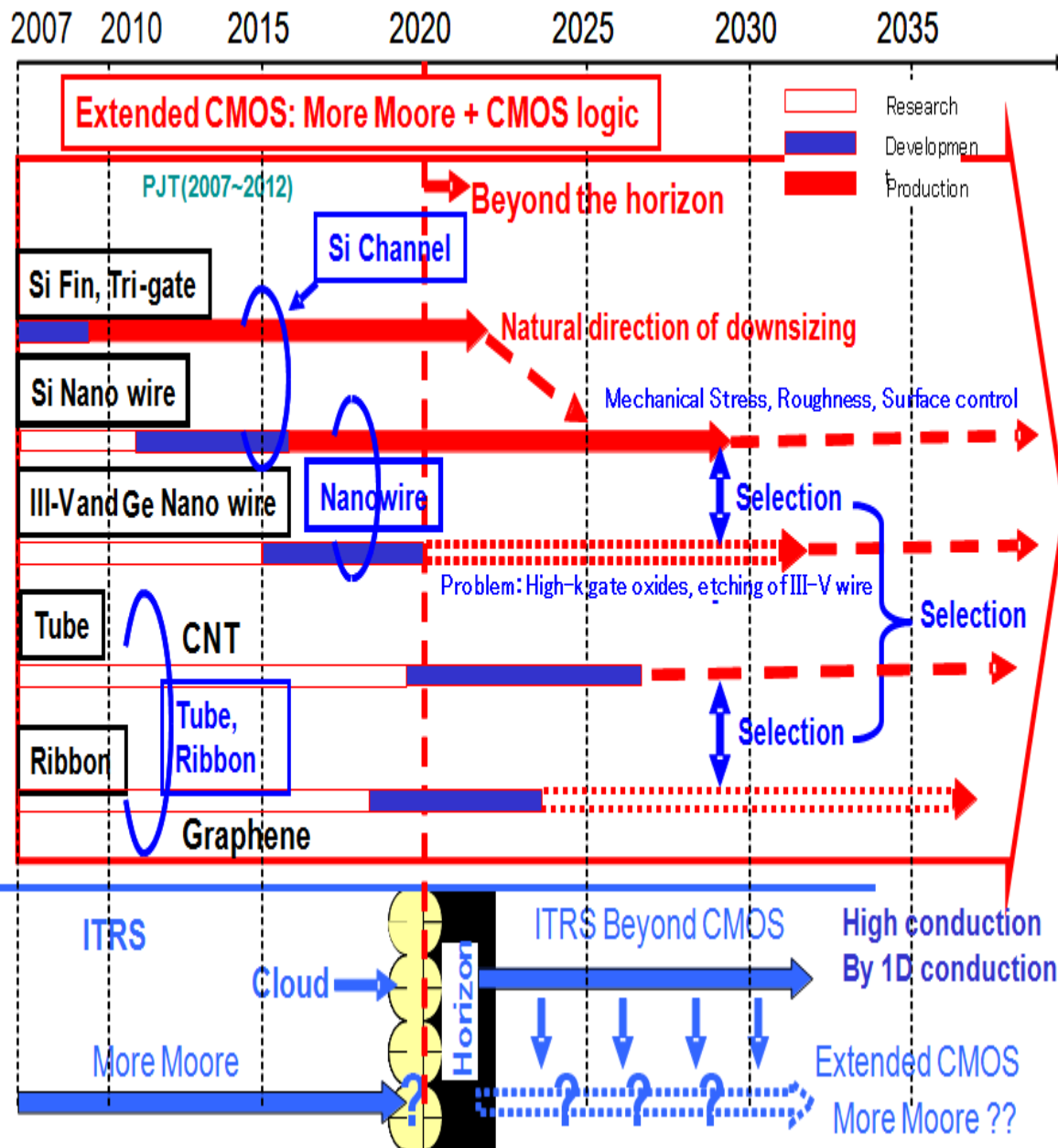
Our new roadmap



Our roadmap for R & D

Source: H. Iwai, IWJT 2008

Current Issues



Si Nanowire

- Control of wire surface property
- Source Drain contact
- Optimization of wire diameter
- Compact I-V model

III-V & Ge Nanowire

- High-k gate insulator
- Wire formation technique

CNT:

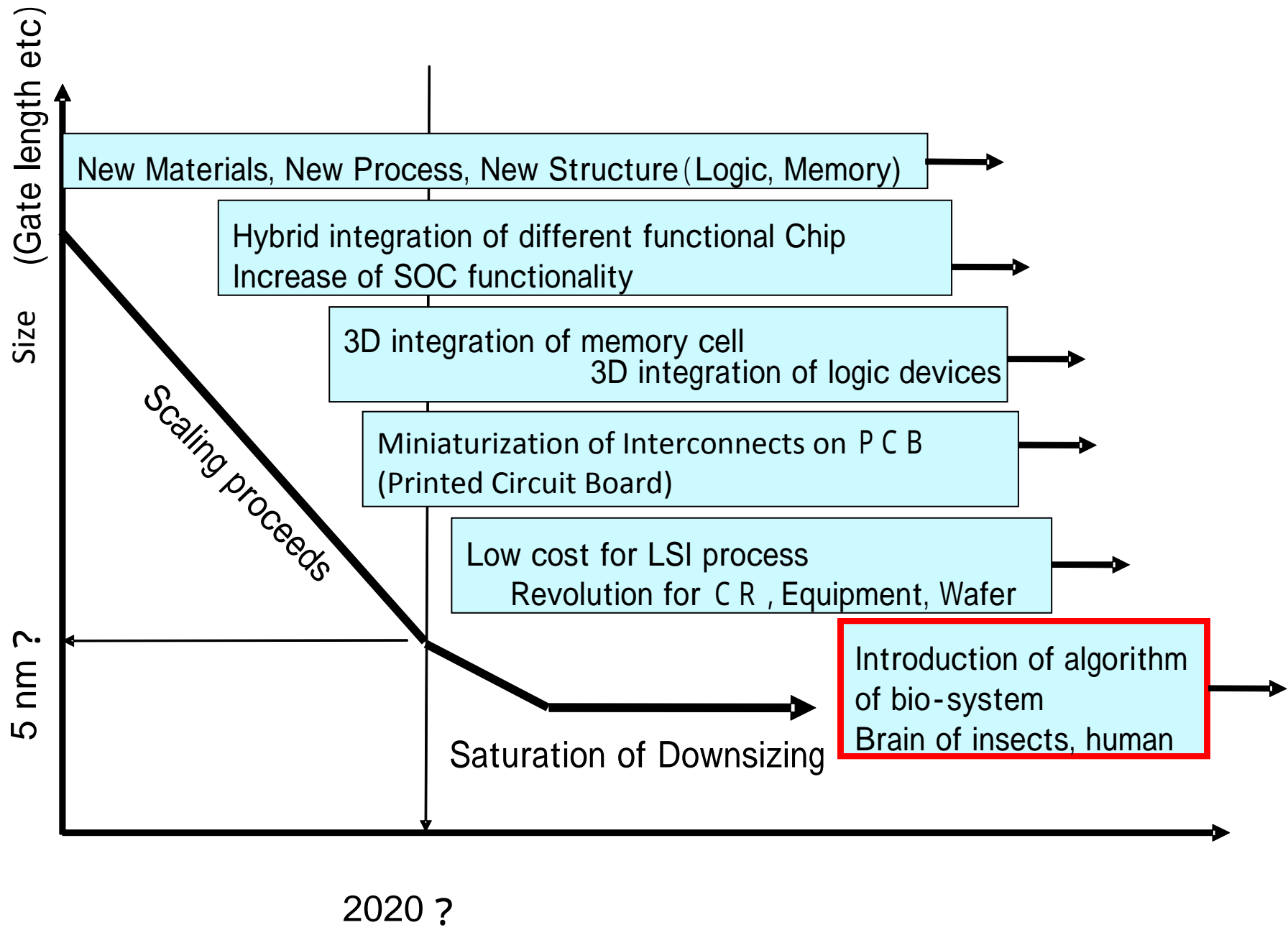
- Growth and integration of CNT
- Width and Chirality control
- Chirality determines conduction types: metal or semiconductor

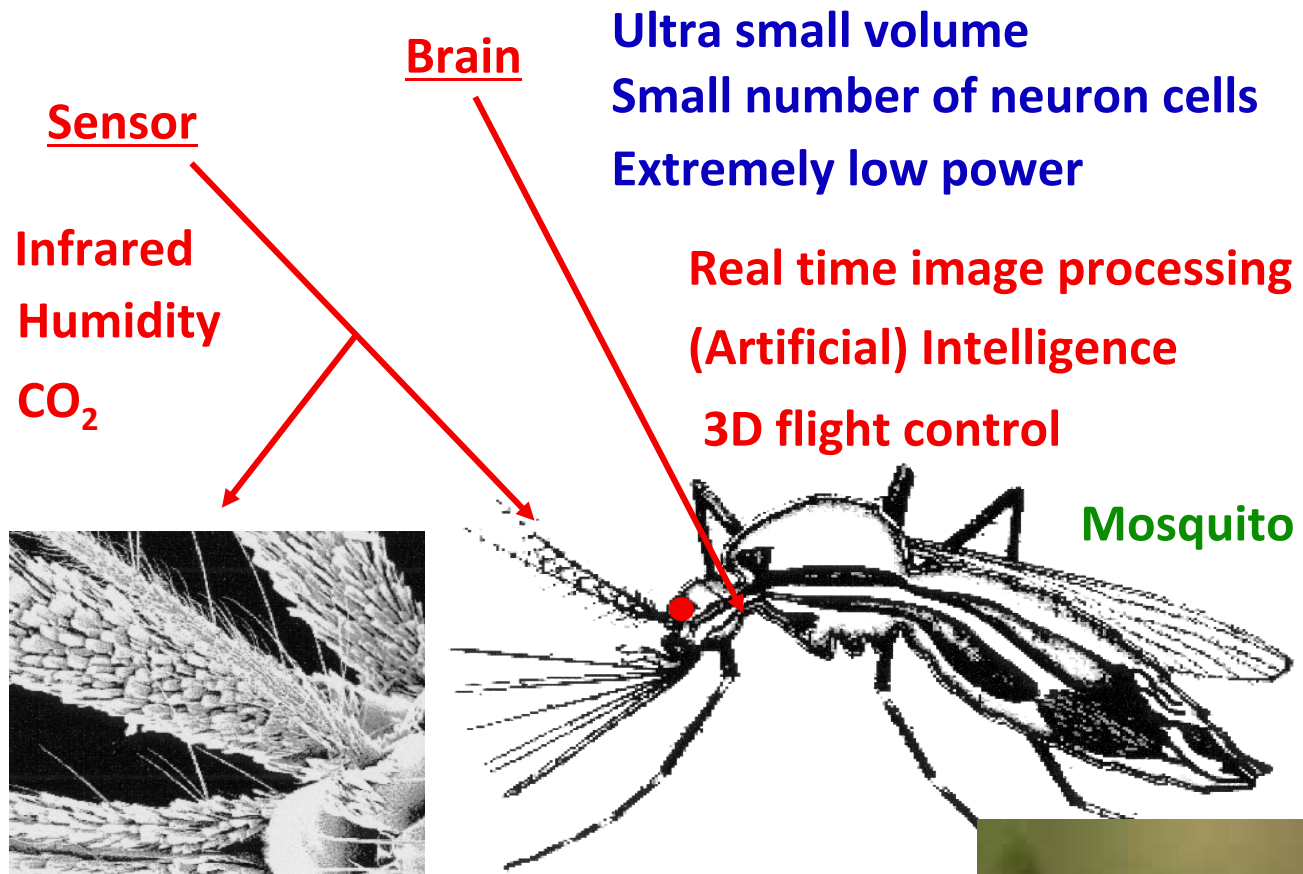
Graphene:

- Graphene formation technique
- Suppression of off-current

Very small bandgap or no bandgap (semi-metal)

- Control of ribbon edge structure which affects bandgap





System and Algorithm becomes more important!

But do not know how?

Dragonfly is further high performance



Thank you
for your attention!