

2017年版 第4回 半導体技術の概要と動向

- 1)概要
- 2)高集積化技術動向
MPU、DRAM、NANDの高集積化推移
- 3)微細化技術動向
ITRSロードマップ、ITRSにおける寸法定義
スケーリング則
微細化に伴う高性能化技術
3D FinFET技術
- 4)NANDのチップ内3D技術
- 5)パッケージ内3次元実装技術
パッケージ内3次元実装技術概要
チップ積層ワイヤボンディング実装、PoP実装、TSVなど

2017.10.24

群馬大学非常勤講師

東京電機大学非常勤講師

中谷 隆之

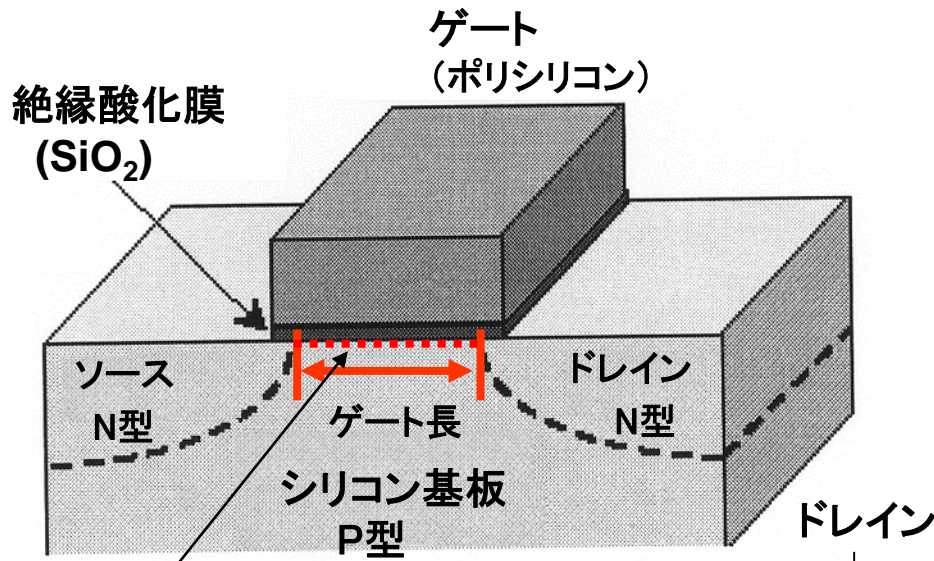
MOSTランジスタ(FET)の構造

MOSTランジスタの基本構造はバイポーラトランジスタに比べてシンプル。P型MOSとN型MOSがある。

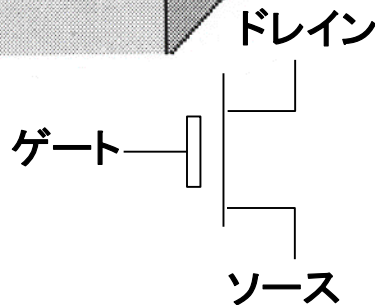
N型MOSTランジスタの場合

- ・P型基板に、N型のソース領域(ウェル)とドレイン領域(ウェル)および薄い酸化膜上にゲート電極を形成
- ・ゲート印加電圧ないときは、ソースとドレイン間に空乏層が形成されソースとドレイン間は導通せず
- ・ゲートにプラス電圧印加すると、ゲート電極下に反転層(N層)が現れ、ソースとドレイン間にチャネル形成され導通

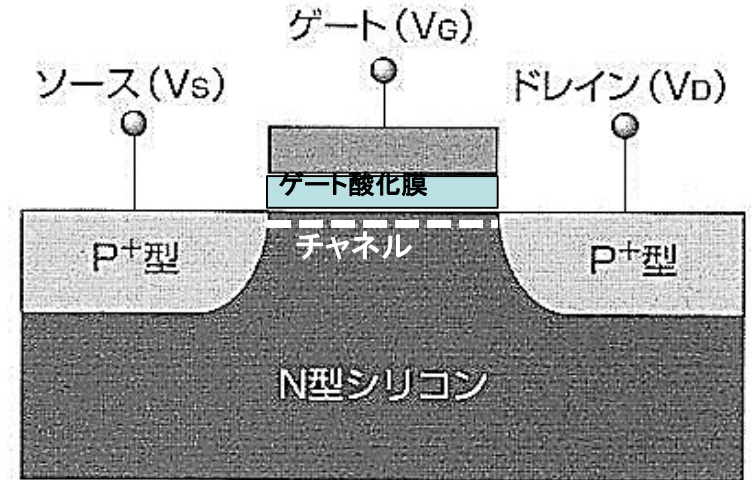
N型MOSトランジスタ



チャネル
ソース・ドレイン間電流は、ゲート直下の表層(反転層) 1-2nm程度の所を流れる。



P型MOSトランジスタ



不純物の例
N:リン(P)
P:ボロン(B) など

CMOS回路

CMOSとは
P型MOSトランジスタと
N型MOSトランジスタから構成

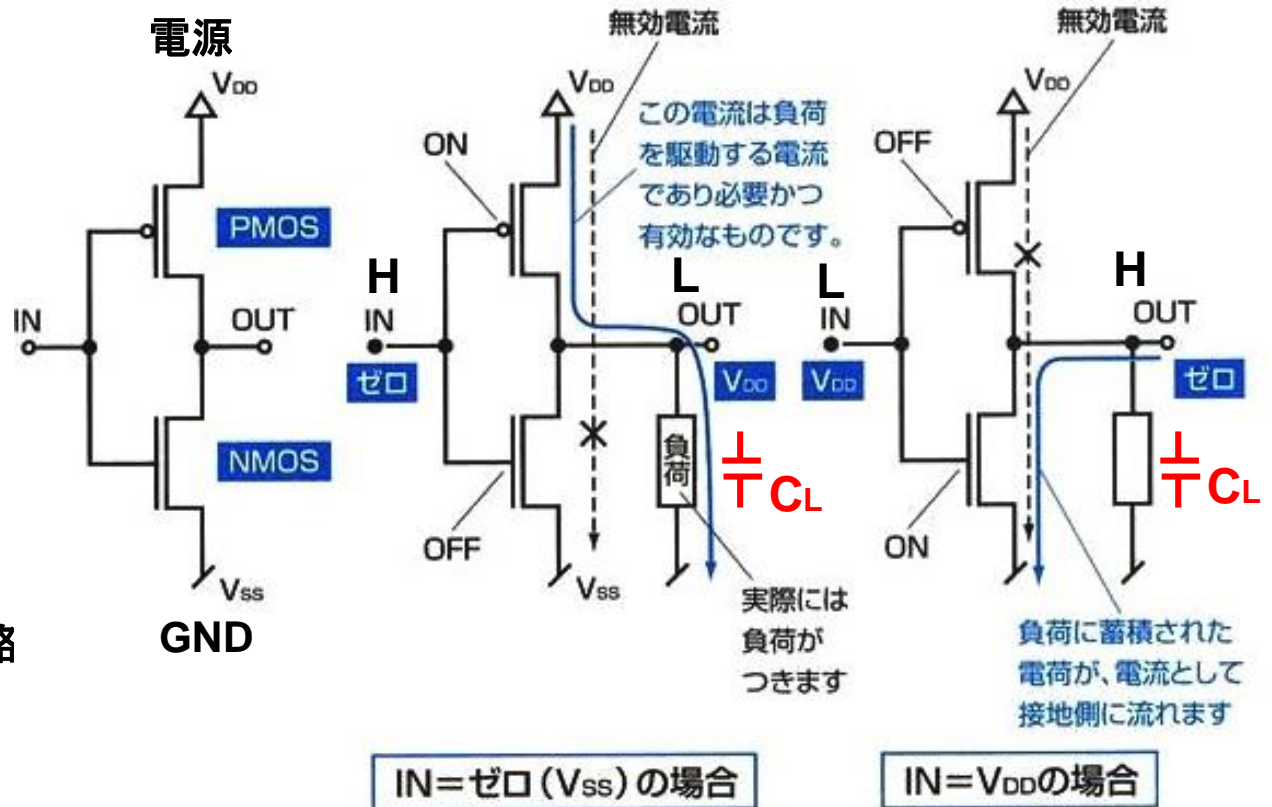
入力L(ゼロレベル)だと
P型MOSがOn
N型MOSがOff
すなわち
出力がH(VDD電源レベル)

入力がH(VDD)だと
P型MOSがOff
N型MOSがOn
すなわち
出力がL(ゼロ)

入力と出力の論理が反転
(インバータ)するデジタル回路

MOS: Metal Oxide Semiconductor
CMOS: Complementary (相補型) MOS

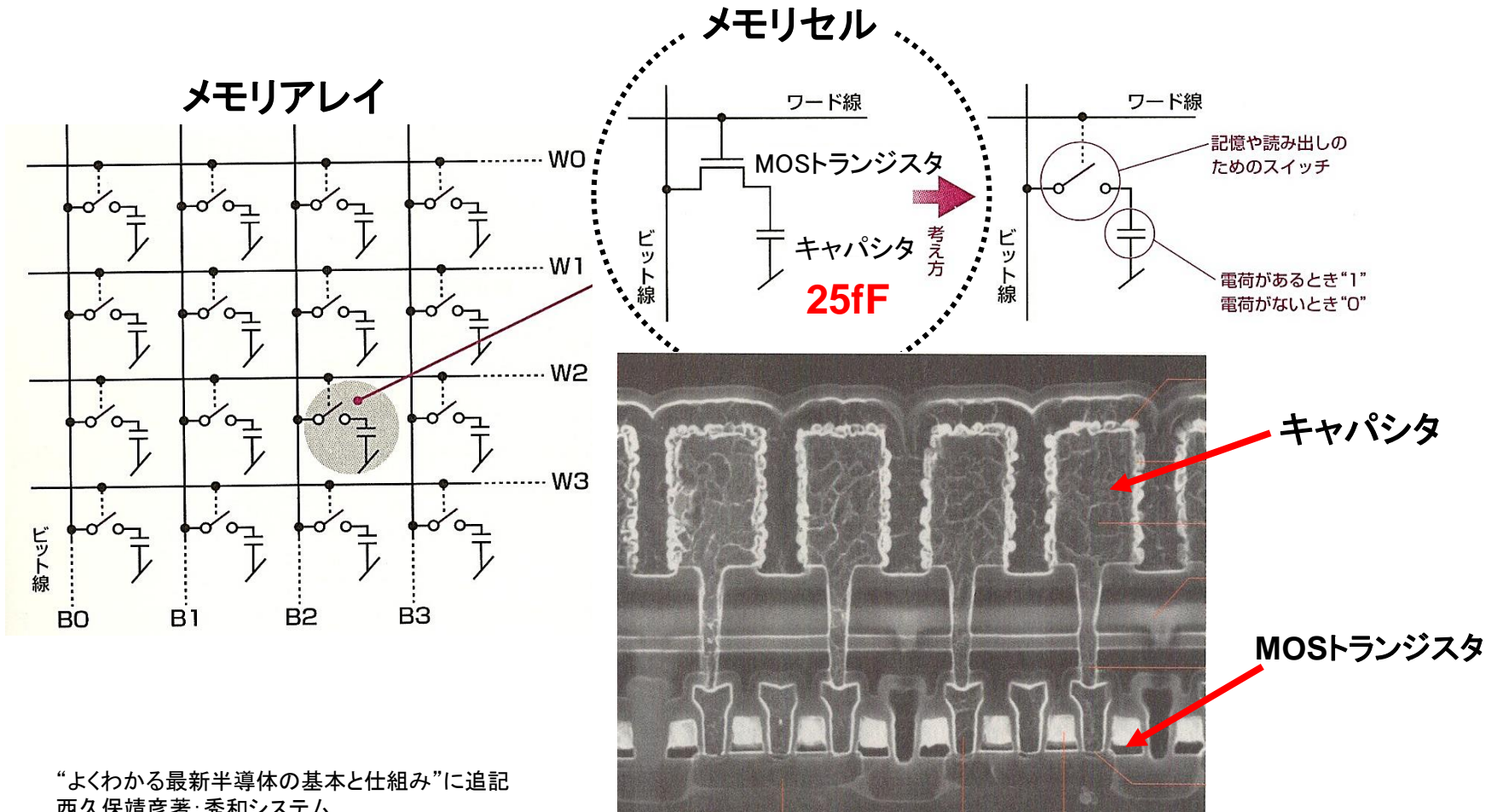
CMOSインバータ



DRAM

Dynamic Random Access Memory

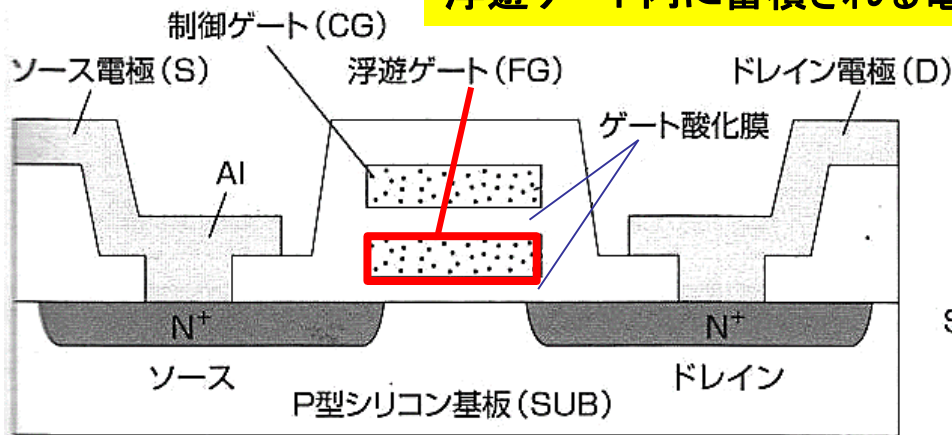
- ・DRAMのメモリセルは、MOSTランジスタ1個とキャパシタ1個から構成される揮発性メモリ
- ・キャパシタに電荷が蓄積された状態を“1”、ない状態を“0”として記憶
- ・MOSTランジスタはキャパシタ電荷の記憶や読み出しのスイッチとして働く。
- ・例えば、4GビットのDRAMでは、40億個のメモリセルが1チップ上に作られている。



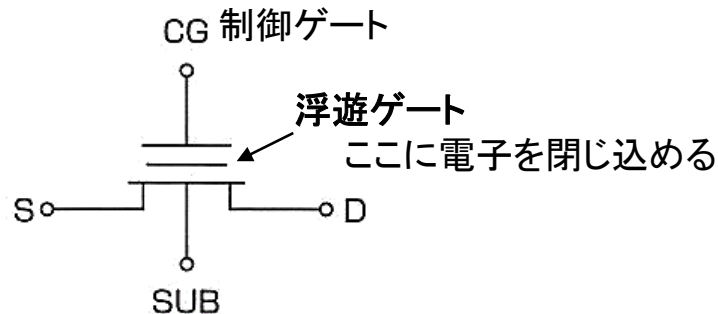
フラッシュメモリ

- ・フラッシュメモリは、酸化絶縁膜 (SiO_2) に囲まれた浮遊 (フローティング) ゲート電極内電子蓄積でデータを蓄え、電源を切ってもデータは消えない不揮発メモリ。
- ・ゲート電極に高電圧をかけ 浮遊ゲートへの電子の“**トンネル効果**”を用いて書込む

浮遊ゲート内に蓄積される電子数は約3,000個 (20nm世代)

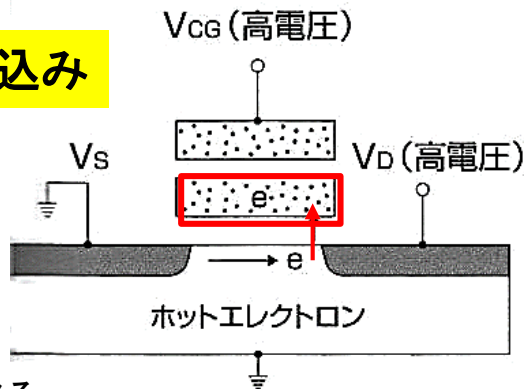


微細化すると電子数は減少

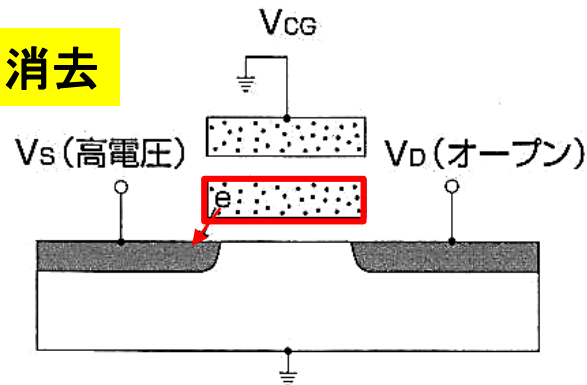


ゲートとドレイン電極に高電圧を印加すると、ソース内の電子がドレイン電極で加速されホットエレクトロンとなる。この内の一部電子が薄い絶縁膜をすり抜け(トンネル効果)浮遊ゲートに飛び移る。消去時はソースに高電圧印加すると浮遊ゲートから基板側に電子が飛び移る。

書込み



消去



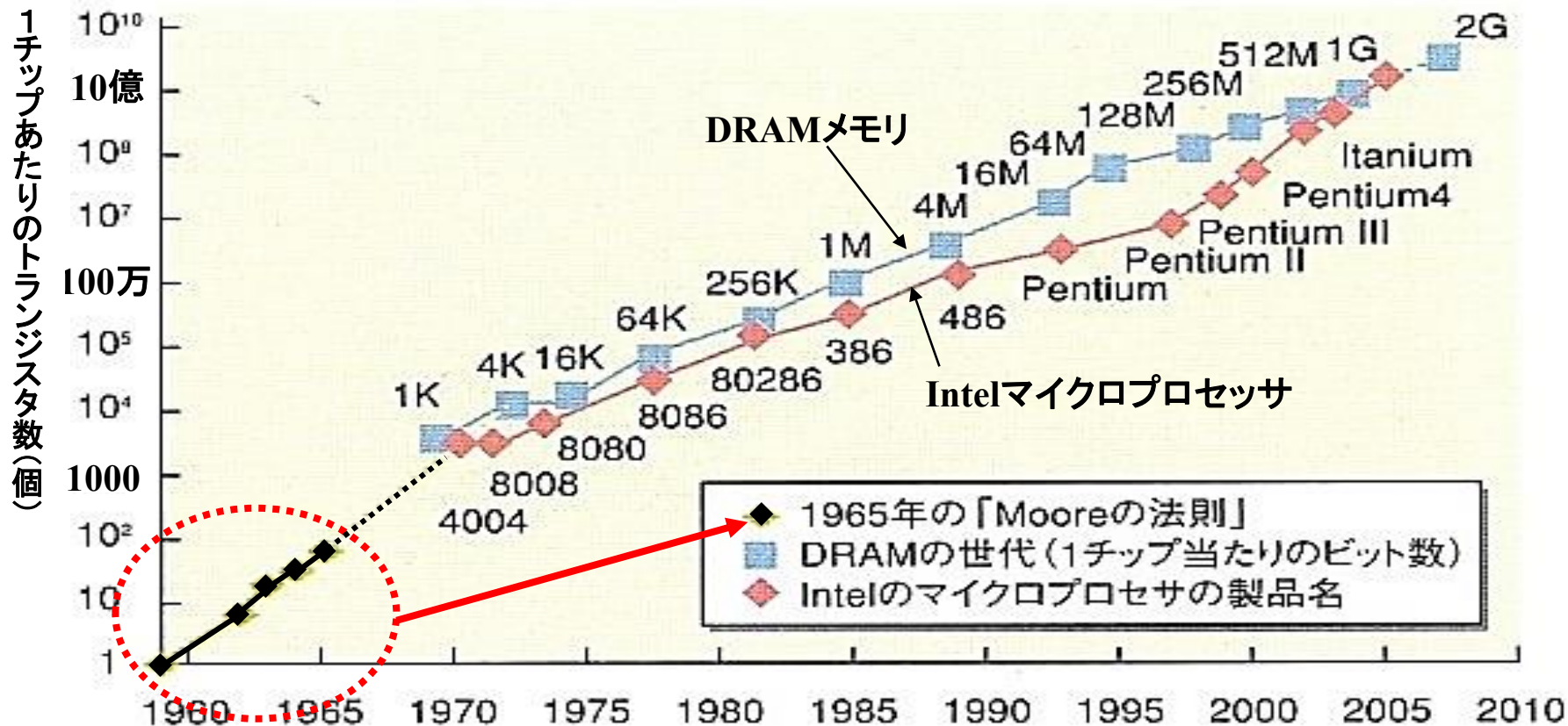
半導体(LSI)の高集積化推移

ムーアの法則: 1965年に集積度は18~24ヶ月(1世代)で2倍増加すると提唱。

1965年以降、現在にいたるまで半導体はムーアの法則により高集積化している。

現在、1チップLSI中に数十億個のトランジスタが集積されている。

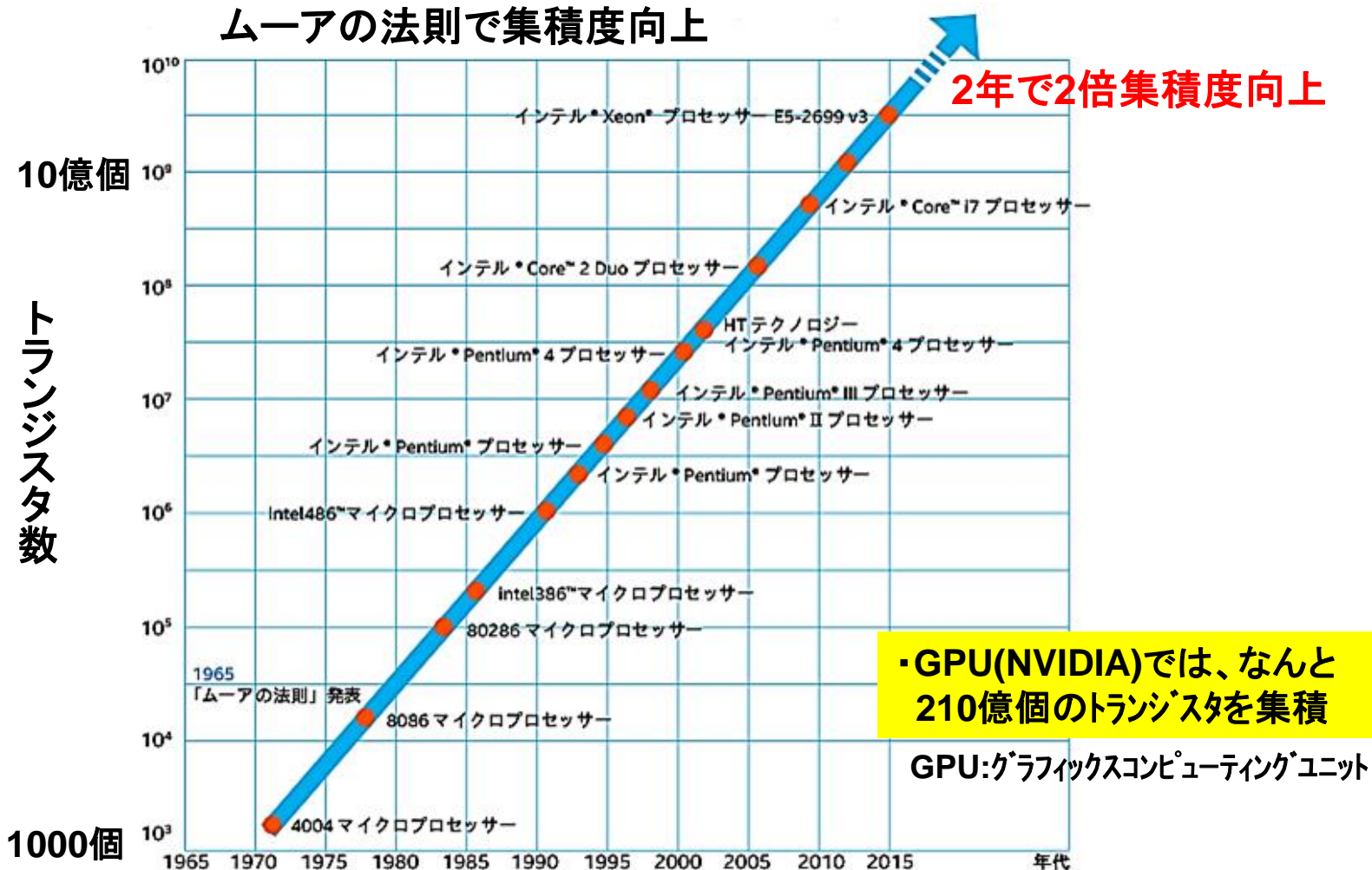
Gordon E. Moore : Intelの共同創始者



マイクロプロセッサに集積されるトランジスタ数

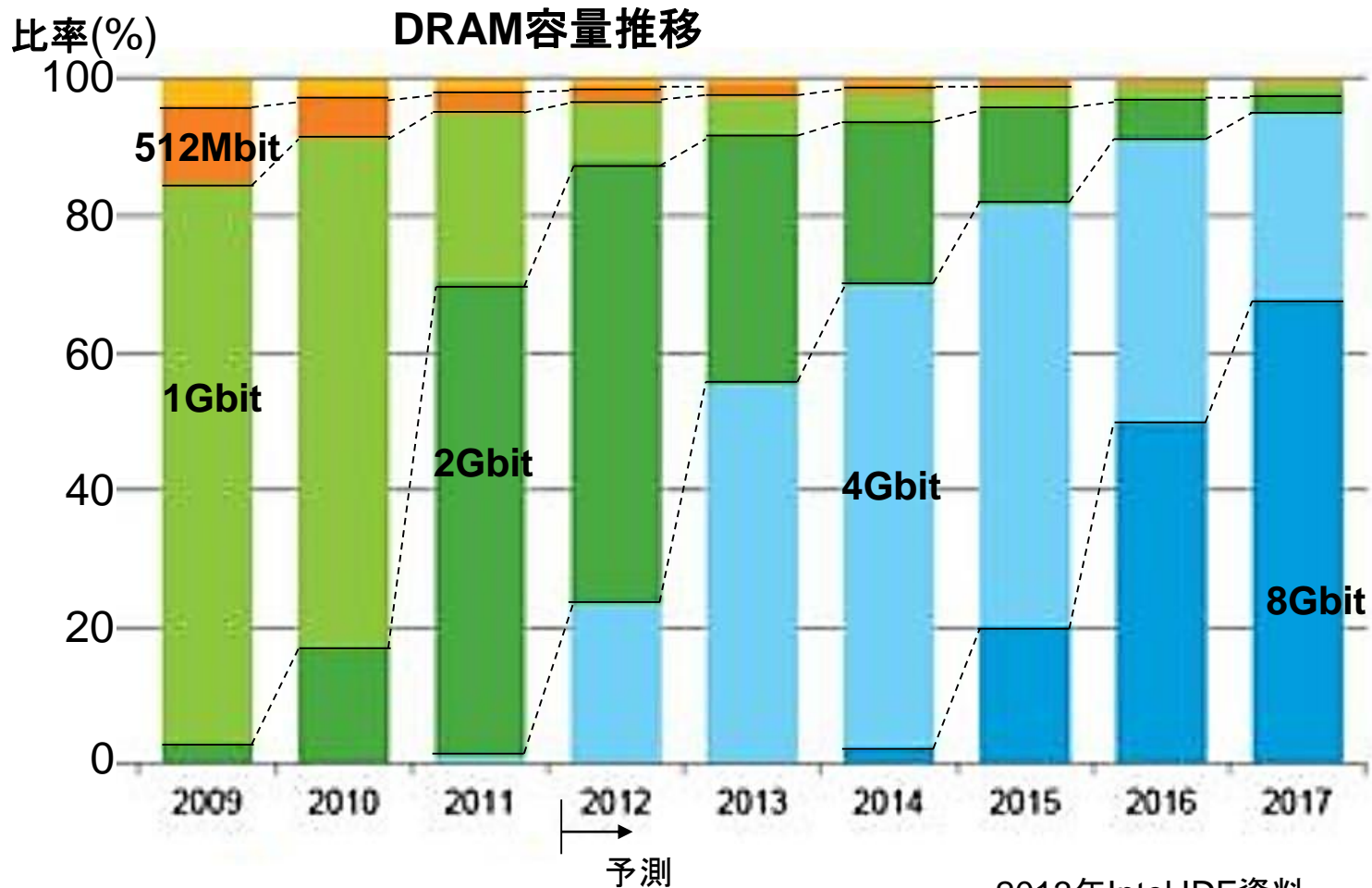
- ・1971年世界初のマイクロプロセッサは、**約2,300個**のトランジスタが集積されていた。
- ・2017年発売のiPhone8用A11プロセッサでは**約 43億個**のトランジスタを集積 (10nmFinFETプロセス)
- ・2017年発売のゲーム器(xBoxone x)プロセッサでは**約70億個**のトランジスタを集積(16nmFinFETプロセス)

ムーアの法則で集積度向上



DRAMの大容量化

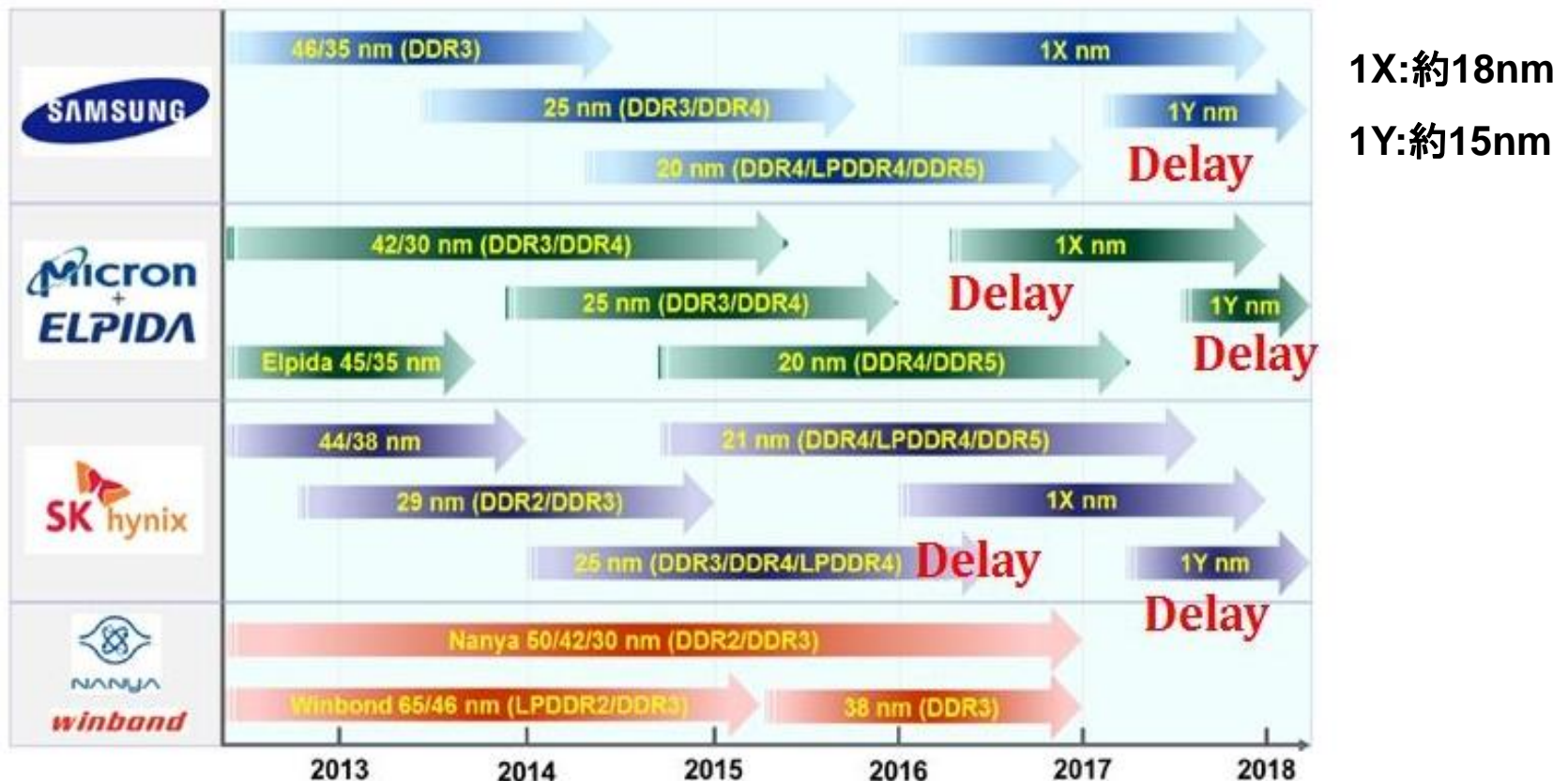
- ・DRAMは約2年半で世代交代し、容量が2倍へ高集積化
- ・2016年、4Gbitから8Gbit世代に移行
- ・DRAMの微細化による大容量化はほぼ限界。16Gbitは難しい



DRAM各社のプロセスロードマップ

DRAMの製造プロセスの微細化は16nm世代で限界に直面。結果大容量化も行きづまり
DRAM微細化ではSamsungが先行

■ DRAM Process Node Roadmap (Manufacturers)



NANDの大容量化

- ・プレーナ型による微細化は2016年、128Gbit(MLC)14nmプロセスで行き止まり
- ・チップ内3次元3D NANDは32層128Gb(TLC)から始まり、毎年2倍容量アップ
- ・2017年ISSCCでは64層512Gbit発表。2017年6月からサンプル出荷
- ・2017年6月にはQLC(4bit/cell)として**768Gbit/chip**発表
- ・768Gbit/chipを薄くして16枚パッケージ内積層し
1.5TByte(12.3Tbit)/パッケージ

多値化

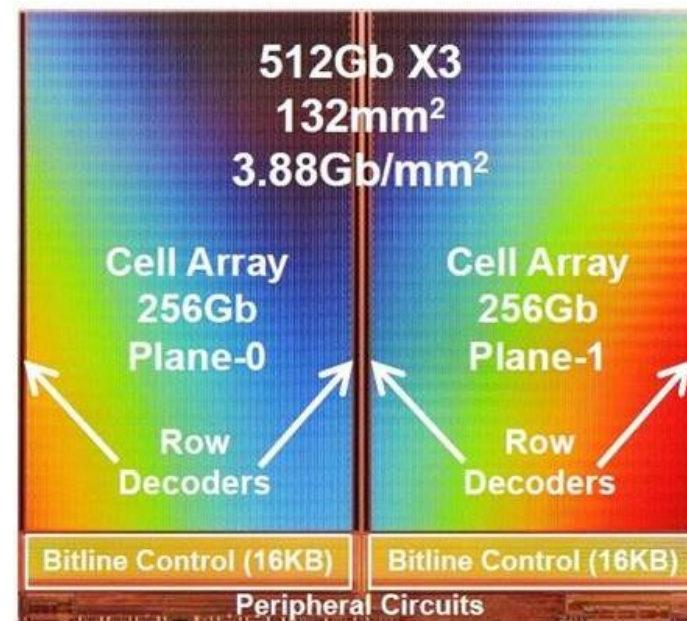
SLC:1bit/cell
MLC:2bit/cell
TLC:3bit/cell
QLC:4bit/cell

ISSCCに見るNAND大容量化推移

年	容量	多値レベル	プロセス
2010	32Gb	MLC	32nmプレーナ
2011	64Gb	MLC	24nmプレーナ
2012	64Gb	MLC	19nmプレーナ
2013	128Gb	TLC	20nmプレーナ
2014	128Gb	MLC	16nmプレーナ
2015	128Gb	TLC	32層3D NAND
2016	256Gb	TLC	48層3D NAND
2017	512Gb	TLC	64層3D NAND
2017/6	768Gb	QLC	64層3D NAND
2017/6	256Gb	TLC	96層3D NAND

768GbはVLSIシンポジウム、96層はプレスリリース(東芝)

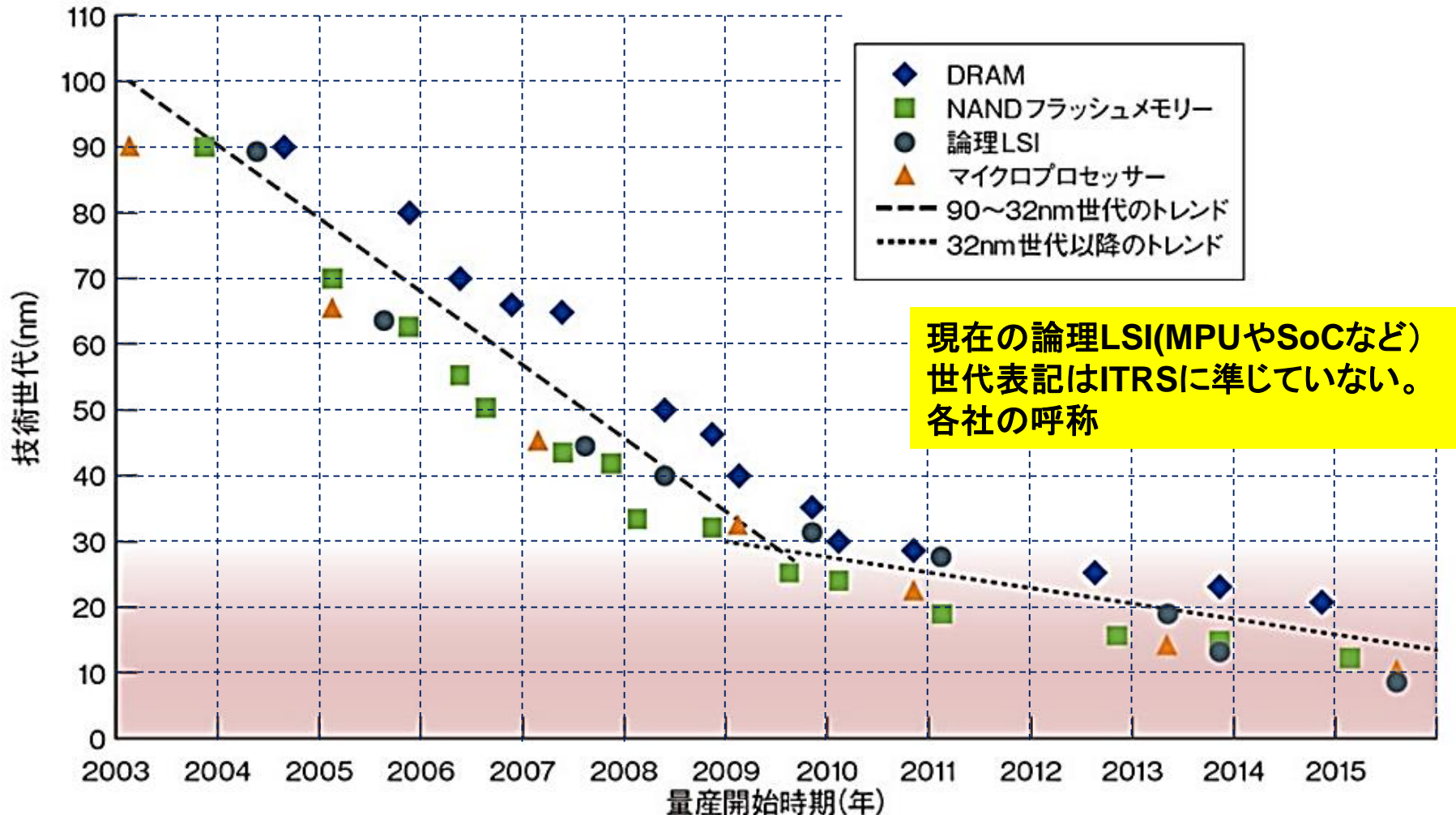
64層512Gb3D NANDチップ(東芝)



ISSCC2017

半導体の微細化トレンド

- ・微細化は2003～2009年頃までは、**約3年毎にx0.7倍**で微細化が進んできた
すなわち同じチップ面積とすると、約3年で集積度が2倍向上
- ・32nm以降、微細化ペースが大幅ダウン。2016年現在、量産最小加工寸法は15nmレベル(NAND)



微細化ロードマップ: 国際半導体技術ロードマップ (ITRS)

- ・ITRSは、ムーアの法則を継続し微細化を継続するために何が必要か、グローバル規模で議論され、重要な技術的課題を抽出し、それぞれの課題ごとに15年先まで見据えた定量的な表を多数作製し、毎年その表の更新を重ねてきた。**2016年2月でITRSが終焉**。原因あれこれ
 - ・継続的微細化の行きづまり。15年先のロードマップ意味なくなった
 - ・ロードマップ可能なのは連続性のある技術範囲。これまでも非連続な技術は予測と大きくずれた
 - ・半導体企業の淘汰 (ファブライツ、ファウンドリ化)。先端設備可能なのは数社

ITRS2013 Edition

1. Executive Summary
2. System Drivers
3. Design
4. Test & Test Equipment
5. Process Integration, Devices & Structures
6. RF and A/MS Technologies
7. Emerging Research Devices
8. Emerging Research Materials
9. Front End Processes
10. Lithography
11. Interconnect
12. Factory Integration
13. Assembly & Packaging
14. Environment, Safety & Health
15. Yield Enhancement
16. Metrology
17. Modeling & Simulation
18. MEMS

簡素化

ITRS2.0 2015 Edition

1. System Integration
2. Heterogeneous Integration
3. Heterogeneous Components
4. Outside System Connectivity
5. More Moore
6. Beyond CMOS
7. Factory Integration

終焉

2016年
2月

ポストITRS

半導体の微細化だけではなく、幅広い技術進歩へのアプローチを検討し、IoTに代表される新しいエレクトロニクス市場にマッチしたロードマップを提供。現在、様々な動きがある。

Memory Trends: ITRS2.0 2015 Edition

DRAM TECHNOLOGY	DRAM現状HP=20nmプロセス。以後の微細化は難しい						
YEAR OF PRODUCTION	2015	2017	2019	2021	2024	2027	2030
Half Pitch (Calculated Half pitch) (nm)	24	20	17	14	11	8.4	7.7
DRAM cell size (μm^2)	0.00346	0.00240	0.00116	0.00078	0.00048	0.00028	0.00024
DRAM cell FET structure	RCAT+Fin	RCAT+Fin	VCT	VCT	VCT	VCT	VCT
Cell Size Factor: a	6	6	4	4	4	4	4
Array Area Efficiency	0.55	0.55	0.5	0.5	0.5	0.5	0.5
V_{int} (support FET voltage) [V]	1.1	1.1	1.1	1.1	0.95	0.95	0.95
Support min. V_m (25C, $G_{m,max}$, $V_d=55mV$)	0.40	0.40	0.40	0.40	0.37	0.37	0.37
Minimum DRAM retention time (ms)	64	64	64	64	64	64	64
DRAM soft error rate (fits)	1000	1000	1000	1000	1000	1000	1000
<i>Gb/chip target</i>	8G	8G	16G	16G	32G	32G	32G

NAND Flash	プレーナ型NANDはHP=14nmプロセスまで以後は3D NAND						
Year of Production	2015	2016	2020	2022	2024	2028	2030
2D NAND Flash uncontacted poly 1/2 pitch - F (nm)	15	14	12	12	12	12	12
3D NAND minimum array 1/2 pitch -F (nm)	80nm	80nm	80nm	80nm	80nm	80nm	80nm
Number of word lines in one 3D NAND string	32	32-48	64-96	96-128	128-192	256-384	384-512
Dominant Cell type (FG, CT, 3D, etc.)	FG/CT/3D	FG/CT/3D	FG/CT/3D	FG/CT/3D	FG/CT/3D	FG/CT/3D	FG/CT/3D
<i>Product highest density (2D or 3D)</i>	256G	384G	768G	1T	1.5T	3T	4T
3D NAND number of memory layers	32	32-48	64-96	96-128	128-192	256-384	384-512
Maximum number of bits per cell for 2D NAND	3	3	3	3	3	3	3
Maximum number of bits per cell for 3D NAND	3	3	3	3	3	3	3

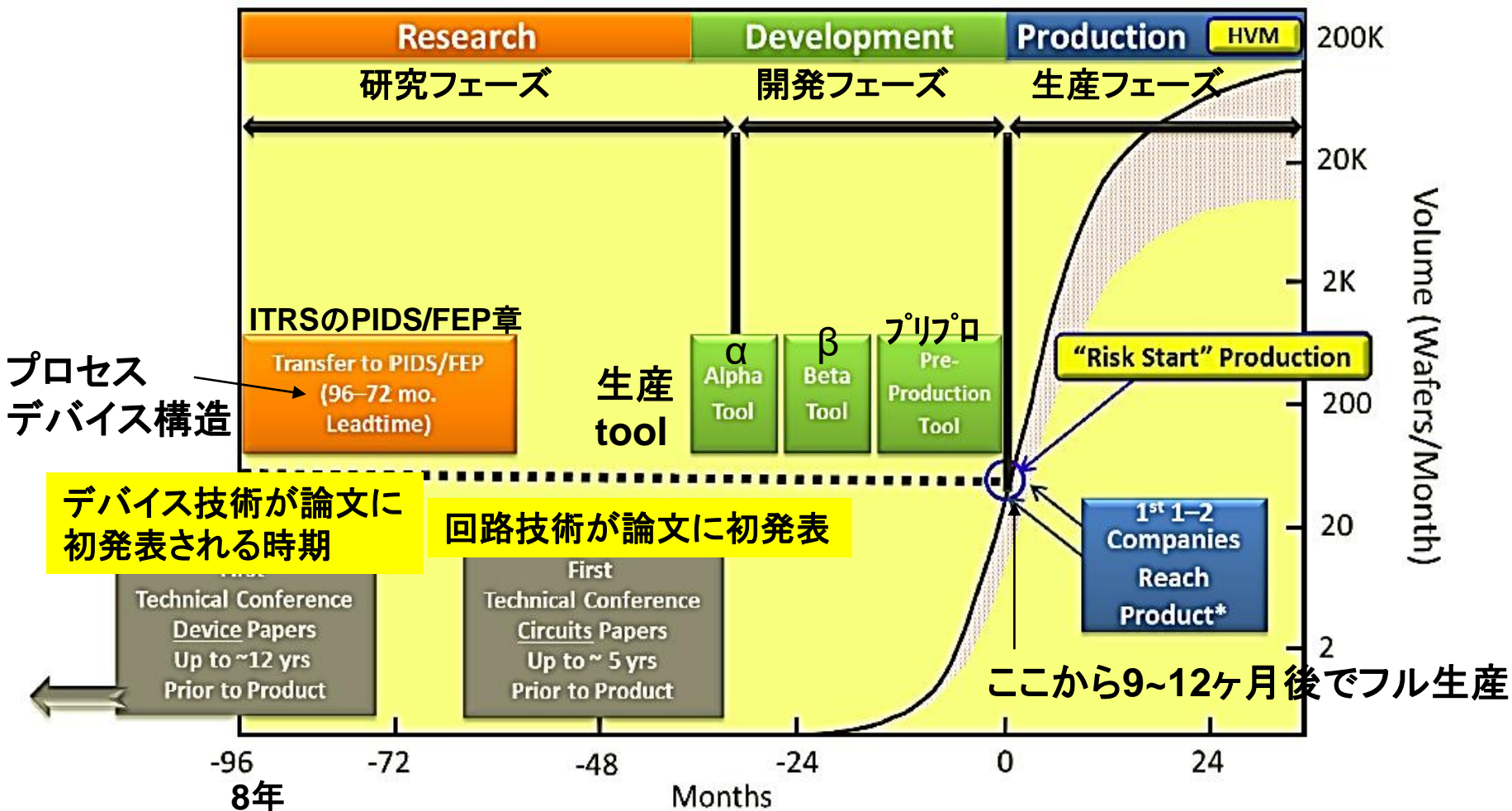
Logic Trends: ITRS2.0 2015 Edition

- ・2016年現在、Logicの最先端は**14/16nm FinFETプロセス**。ITRS定義ではHP(ハーフピッチ)=28nm
- 2017年、10nmFinFETプロセス。ITRS定義によるHP=18nm

YEAR OF PRODUCTION	2015	2017	2019	2021	2024	2027	2030
Logic device technology naming	P70M56	P48M36	P42M24	P32M20	P24M12G1	P24M12G2	P24M12G3
Logic industry "Node Range" Labeling (nm)	"16/14"	"11/10"	"8/7"	"6/5"	"4/3"	"3/2.5"	"2/1.5"
Logic device structure options	FinFET FDSOI	FinFET FDSOI	FinFET LGAA	FinFET LGAA VGAA	VGAA, M3D	VGAA, M3D	VGAA, M3D
LOGIC DEVICE GROUND RULES							
ITRS正式定義							
MPU/SoC Metaix ½ Pitch (nm)[1,2]	28.0	18.0	12.0	10.0	6.0	6.0	6.0
MPU/SoC Metal0/1 ½ Pitch (nm)	28.0	18.0	12.0	10.0	6.0	6.0	6.0
Contacted poly half pitch (nm)	35.0	24.0	21.0	16.0	12.0	12.0	12.0
L _g : Physical Gate Length for HP Logic (nm) [3]	24	18	14	10	10	10	10
L _g : Physical Gate Length for LP Logic (nm)	26	20	16	12	12	12	12
FinFET Fin Half-pitch (new) =0.75 or 1.0 M0/M1 (nm)	21.0	18.0	12.0				
FinFET Fin Width (nm)	8.0	6.0	6.0				
FinFET Fin Height (nm)	42.0	42.0	42.0		END OF	2D	DOMAIN
Footprint drive efficiency - FinFET	2.19	2.50	3.75				

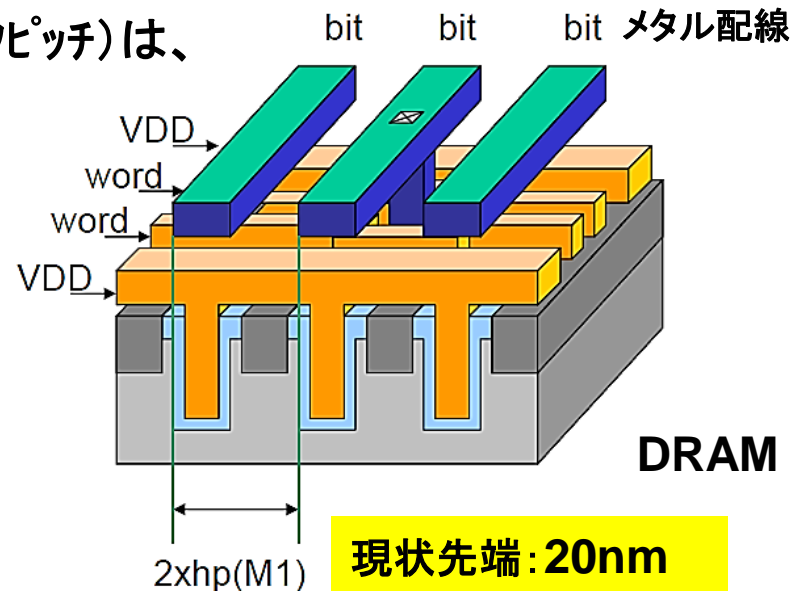
ITRS定義： Research, Development, Production

- ・Production(生産フェーズ)は、先行する1~2社が2万枚/月のウェハ生産(リスク生産)に至る時
- ・回路に関する新技術は、初論文から製品になるのに~5年かかる
- ・デバイスに関わる新技術は、初論文から製品化に至るのに~12年かかる
- ・近年、一社が2万枚/月に達してから2社目が2万枚/月に達するのに4年もかかる例あり(Intel先行)



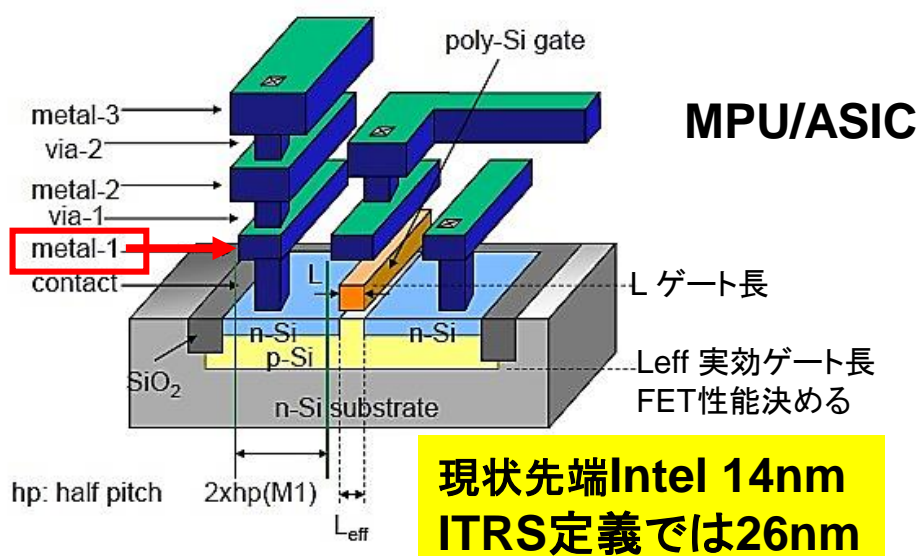
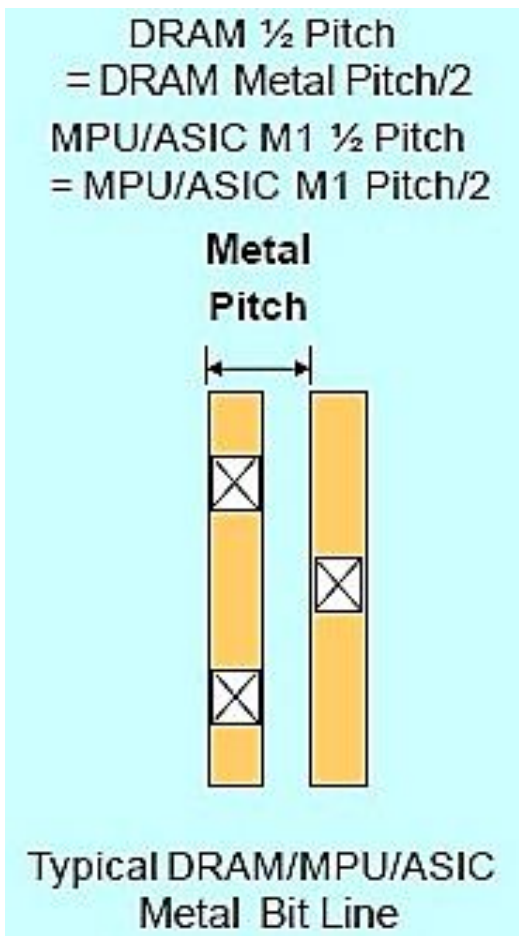
ITRSによるDRAMとMPU/ASICの寸法定義

DRAMおよびMPU/ASICでの1/2Pitch (hp: ハーフピッチ) は、
最下層メタル配線 (M1) 配線ピッチの1/2で定義



現状先端: 20nm

DRAM

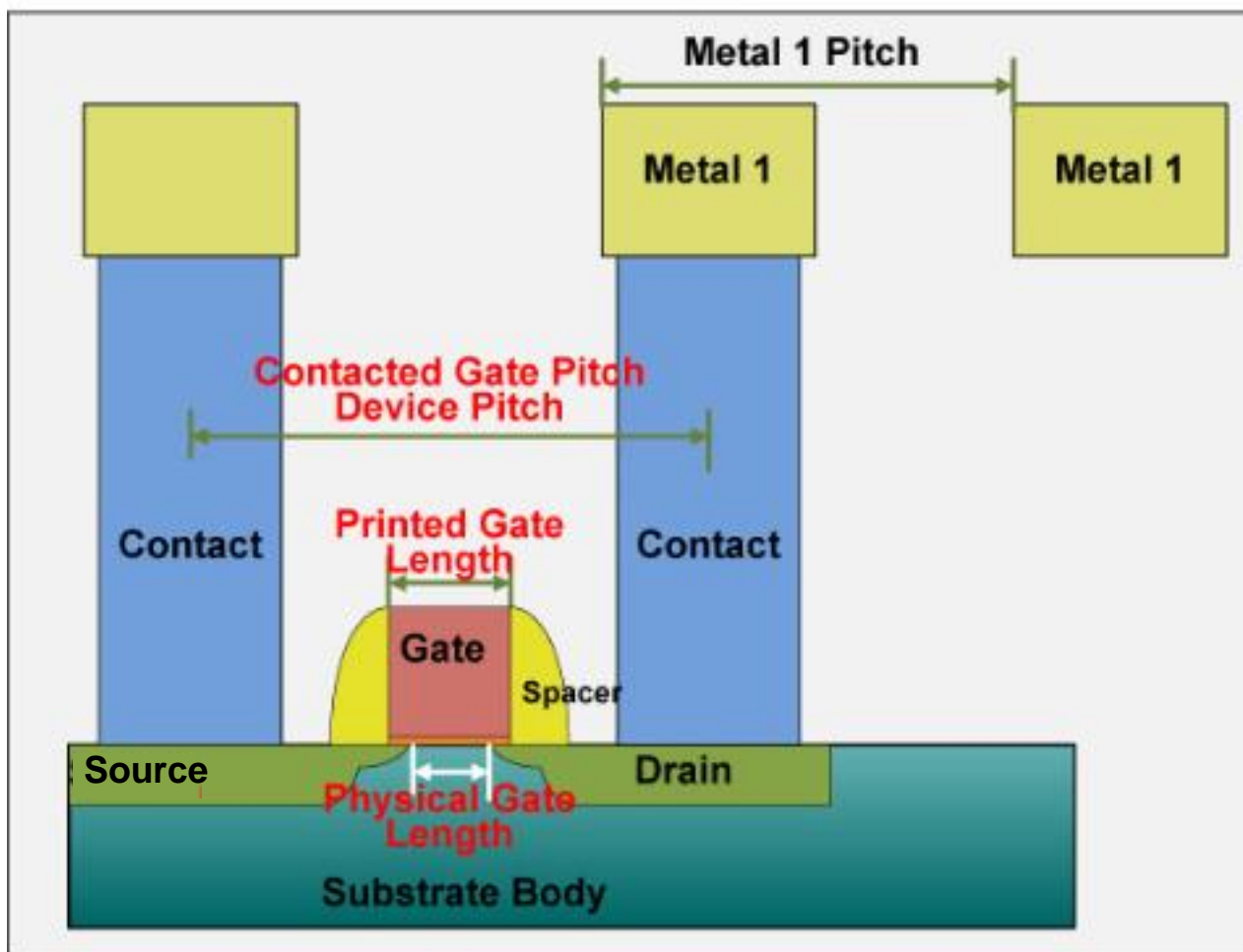


MPU/ASIC

現状先端Intel 14nm
ITRS定義では26nm

トランジスタの寸法定義詳細

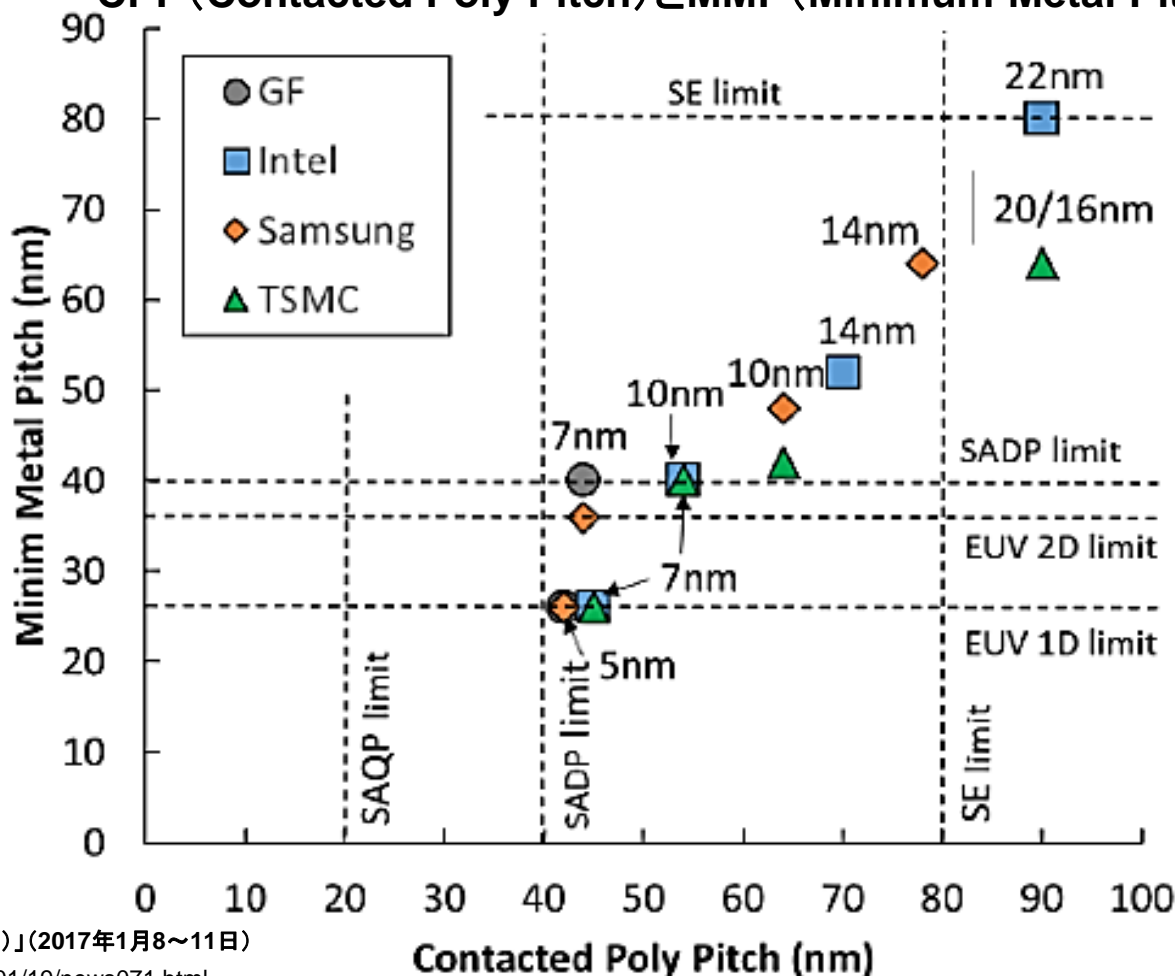
- ・半導体プロセスでは、配線層のことをメタルと呼ぶ。
- ・配線は、最も下のトランジスタに近い配線が「M1(Metal 1)」で、現在では「M10(Metal 10)」以上
- ・配線は、最下層のM1が最も配線幅やピッチが狭く、上層になるほど配線幅やピッチは広くなる
- ・**ITRSでは最下層のM1ピッチの1/2(HP)で寸法を定義**



ロジックプロセス各社呼称とITRS定義の関係

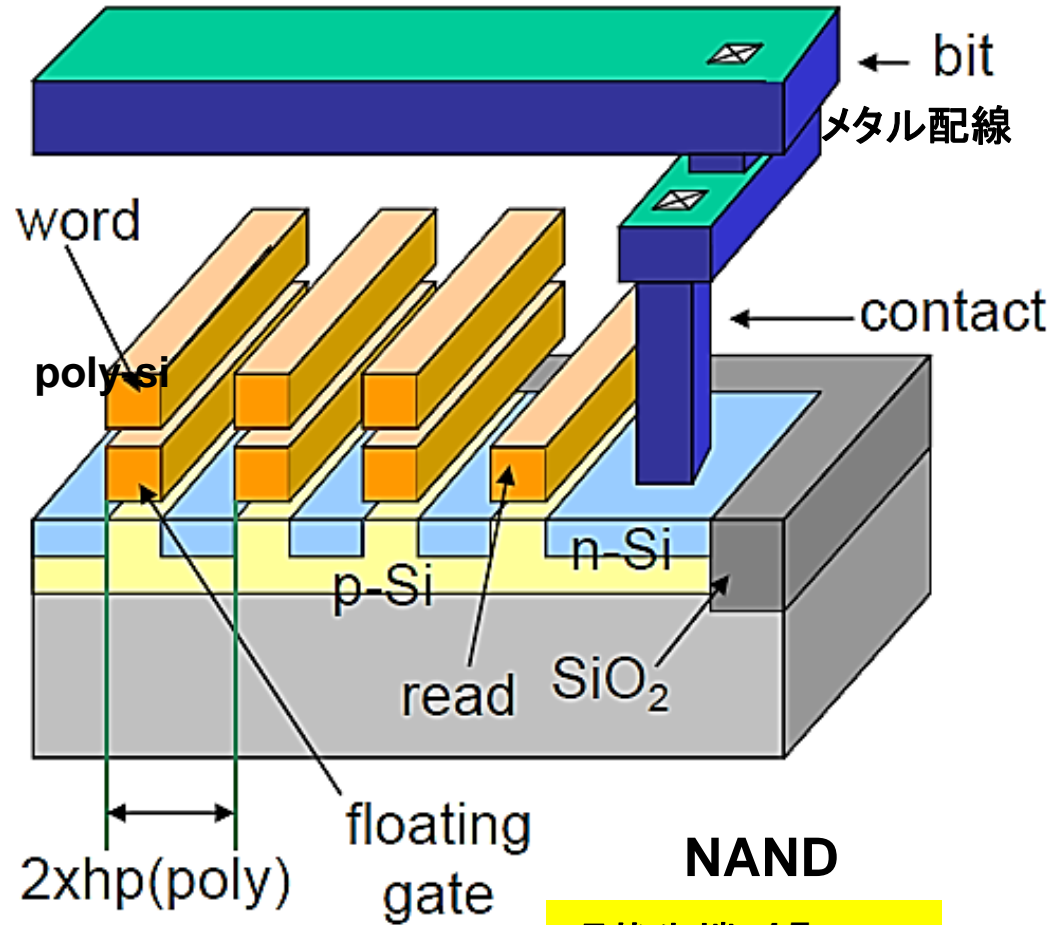
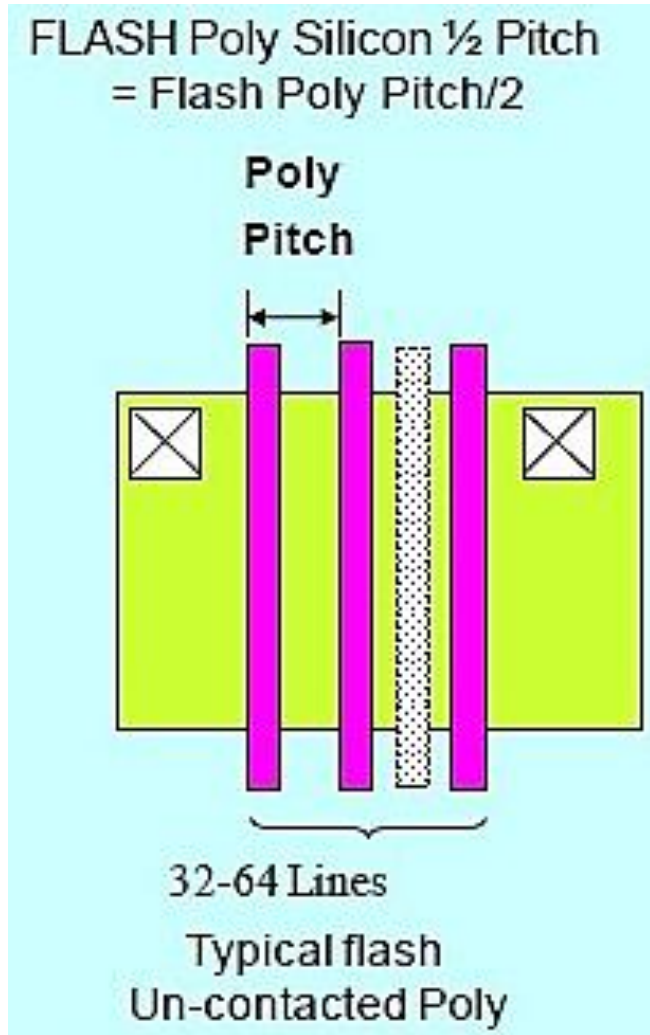
- ・Intelの10nm呼称は、ITRS定義の20nm相当
- ・Intelの10nm呼称は、他社(Samsung, TSMC, GF)の7nm相当
- ・Intelの7nm呼称は他社の5nm相当

GLOBALFOUNDRIES (GF)、Intel、Samsung、TSMC各社のCPP (Contacted Poly Pitch) とMMP (Minimum Metal Pitch)



ITRSによるNANDの寸法定義

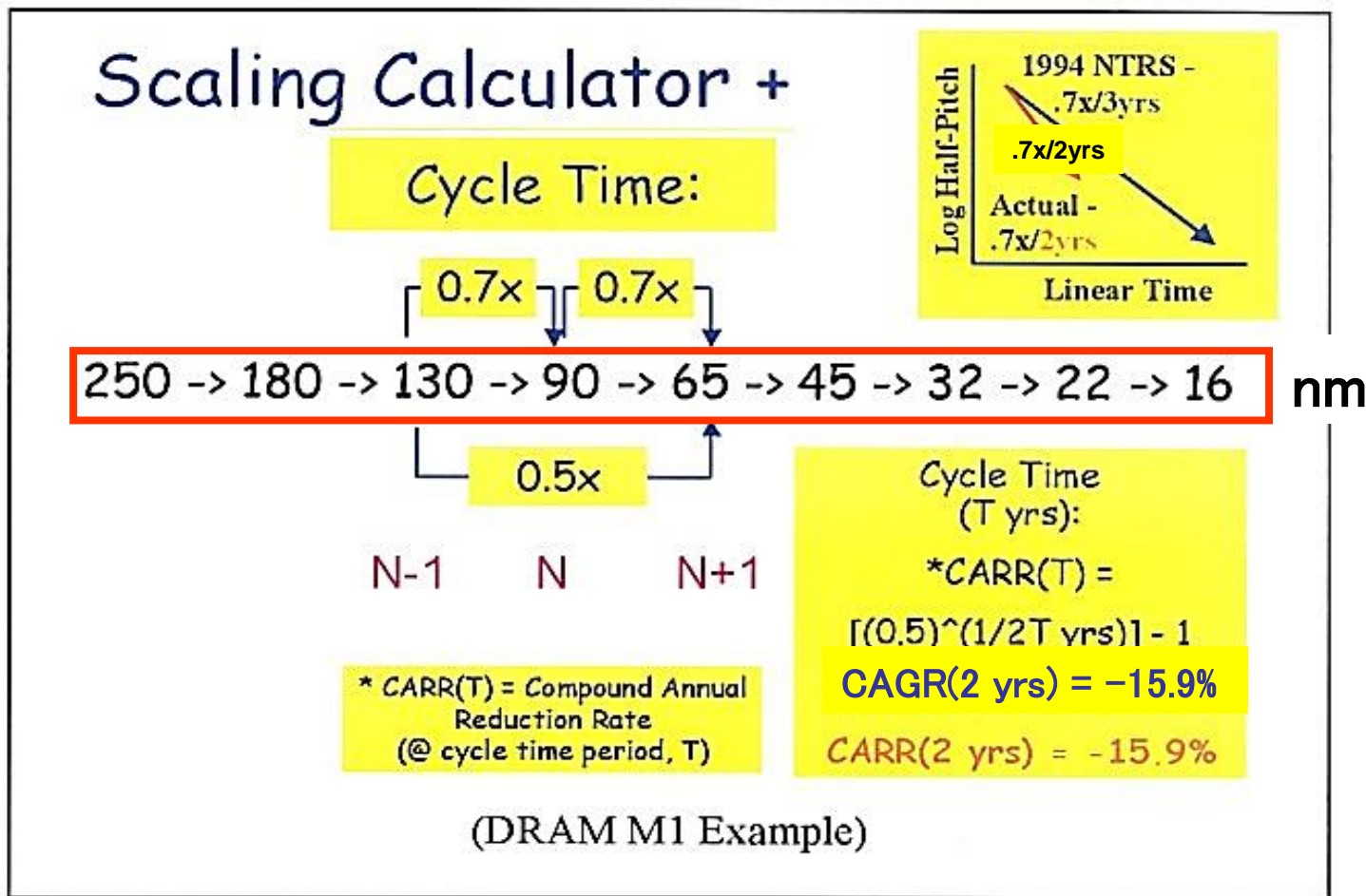
NANDフラッシュでの1/2Pitch (hp:ハーフピッチ)は、
ポリシリコン(フローティングゲート)ピッチの1/2で定義



現状先端: 15nm

微細化: スケーリング

- ・半導体の微細化は、最小加工寸法が1世代(約2年)でx0.7倍ずつ縮小(スケーリング)
- ・ITRS定義による微細化はNANDが最も進んでいる
- ・近年、ロジック(MPUやSOC)LSIで使用される22nm,16nm,14nm表記は、ITRS定義ではなく実効ゲート長。(3D FinFETではプレーナ相当ゲート長) M1ハーフピッチ

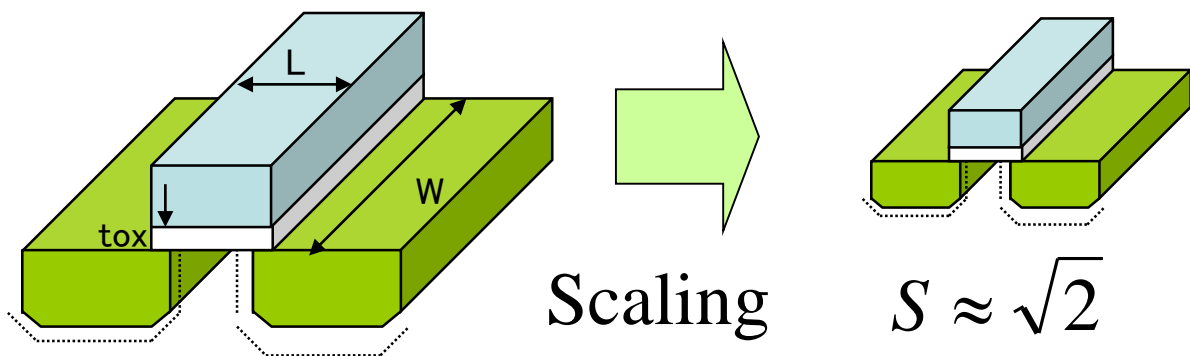


トランジスタの微細化：デナードのスケールング則

R. Dennard のスケールング理論(1974年)

- ・スケールングファクタをS
 - ・各ディメンジョンを1/Sに縮小
 - ・動作電圧を1/S
- として、**チャネル電界を一定**とする理論。

スケールング則はLSIの黄金則である



微細化で顕在化しつつある問題

- ・短チャネル効果による
リーク電流増大の顕在化
- ・スケールングによりゲート遅延は短縮されるが、**配線遅延増大**が顕在化

短チャネル効果：

ゲート長(L)が短くなり、ドレイン電界の影響が大きくなり、漏れ電流が増大する効果。

動作電圧も1/Sにする

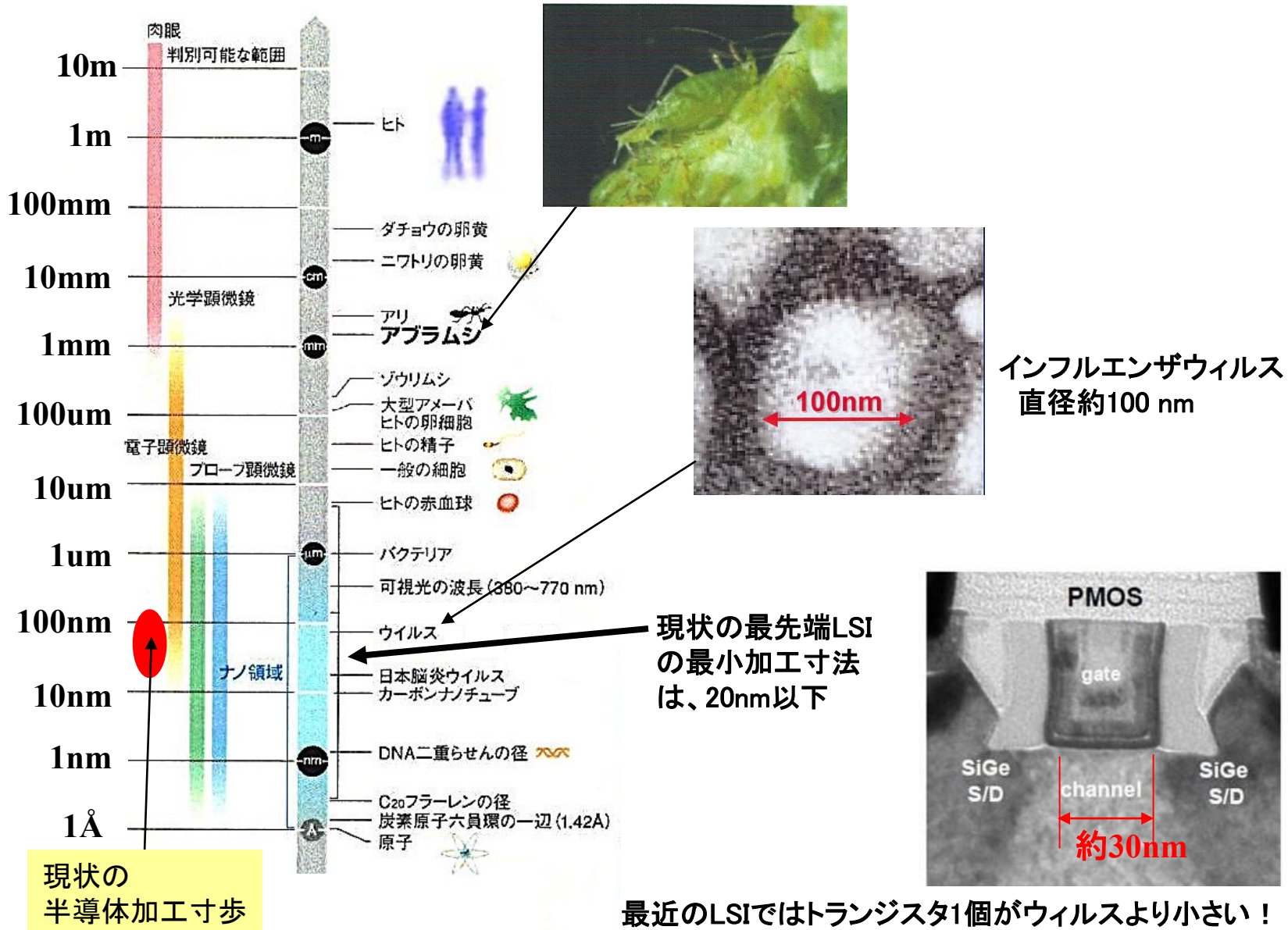
トランジスタのスケールングパラメータ

Device/Circuit parameter	Scaling Factor
Device dimensions L, W, Tox	1/S
Doping concentration	S ~ S ^{1.5}
Voltage	1/S
Field	1
Current	1/S
Gate Delay	1/S
Power dissipation/device	1/S ³ ~ 1/S ²

微細化が進み、スケールングパラメータが、右表から乖離してきた。

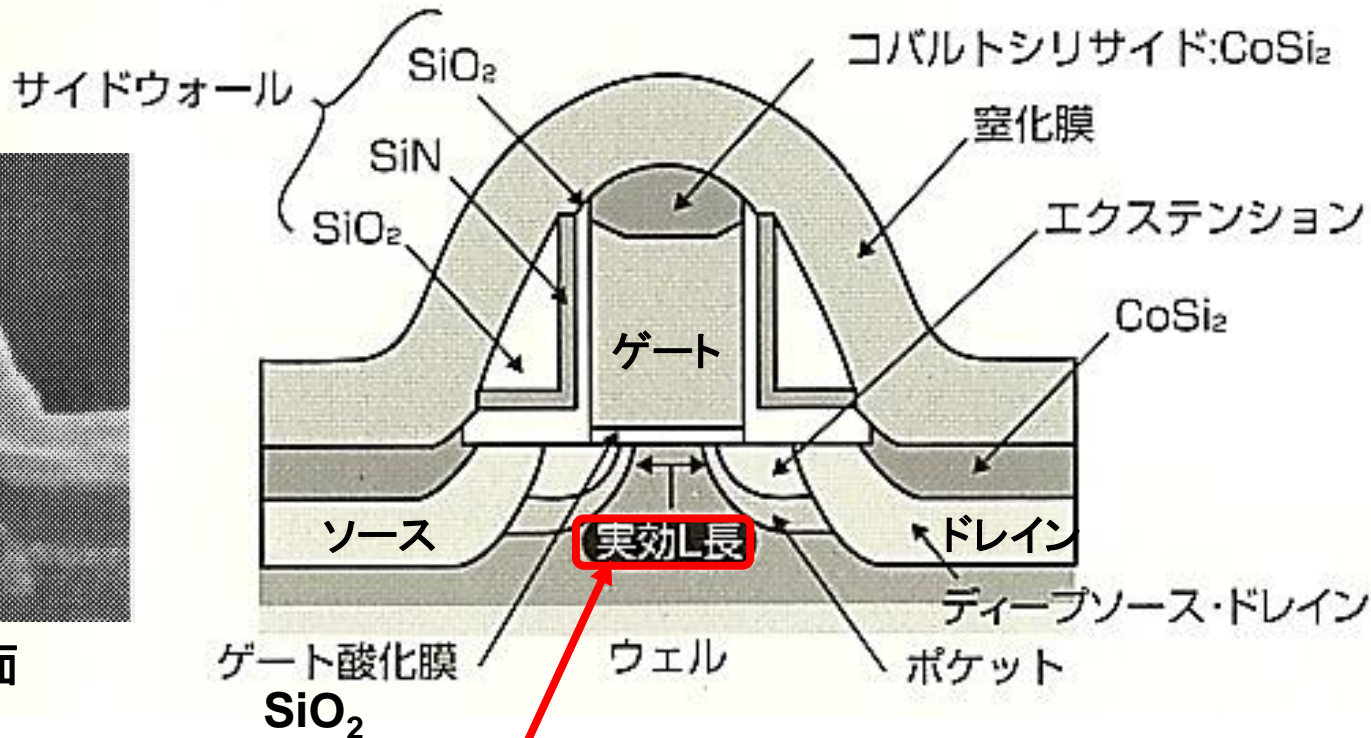
例えば電源が1/Sで下がらないなど。

様々な物質のサイズと半導体の微細化:



MOS FETは微細化に伴い構造は複雑化

- ・130nm以降の微細化では、様々な性能劣化対策のためMOS FET構造が複雑化
- ・**実効ゲート長(実効L長)**によりトランジスタ性能(動作速度)が決まる。
このためプロセス技術で実際のゲート長よりも短くする技術(エクステンション、ポケット構造など)
- ・微細化するとソース、ドレイン、ゲートが高抵抗化対策のためシリサイド CoSi_2 技術など



Chipworks社データ 2004.1

47.5nm

実際のMOS FET断面

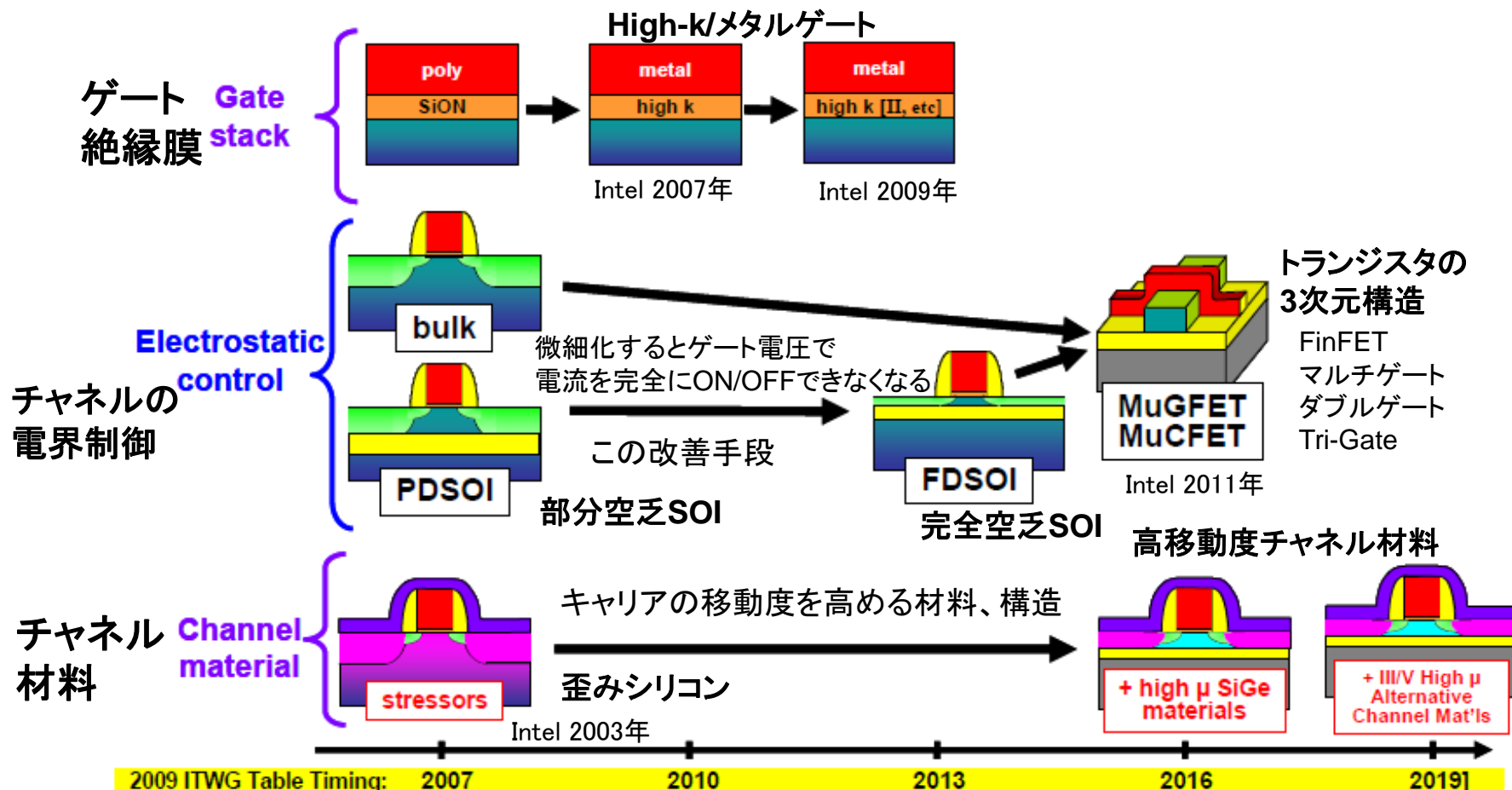
90nmプロセスで

実効ゲート長は47.5nm

MPUやSOCメーカーは、この寸法をプロセス世代と称してきた

微細化トランジスタにおける高性能化技術

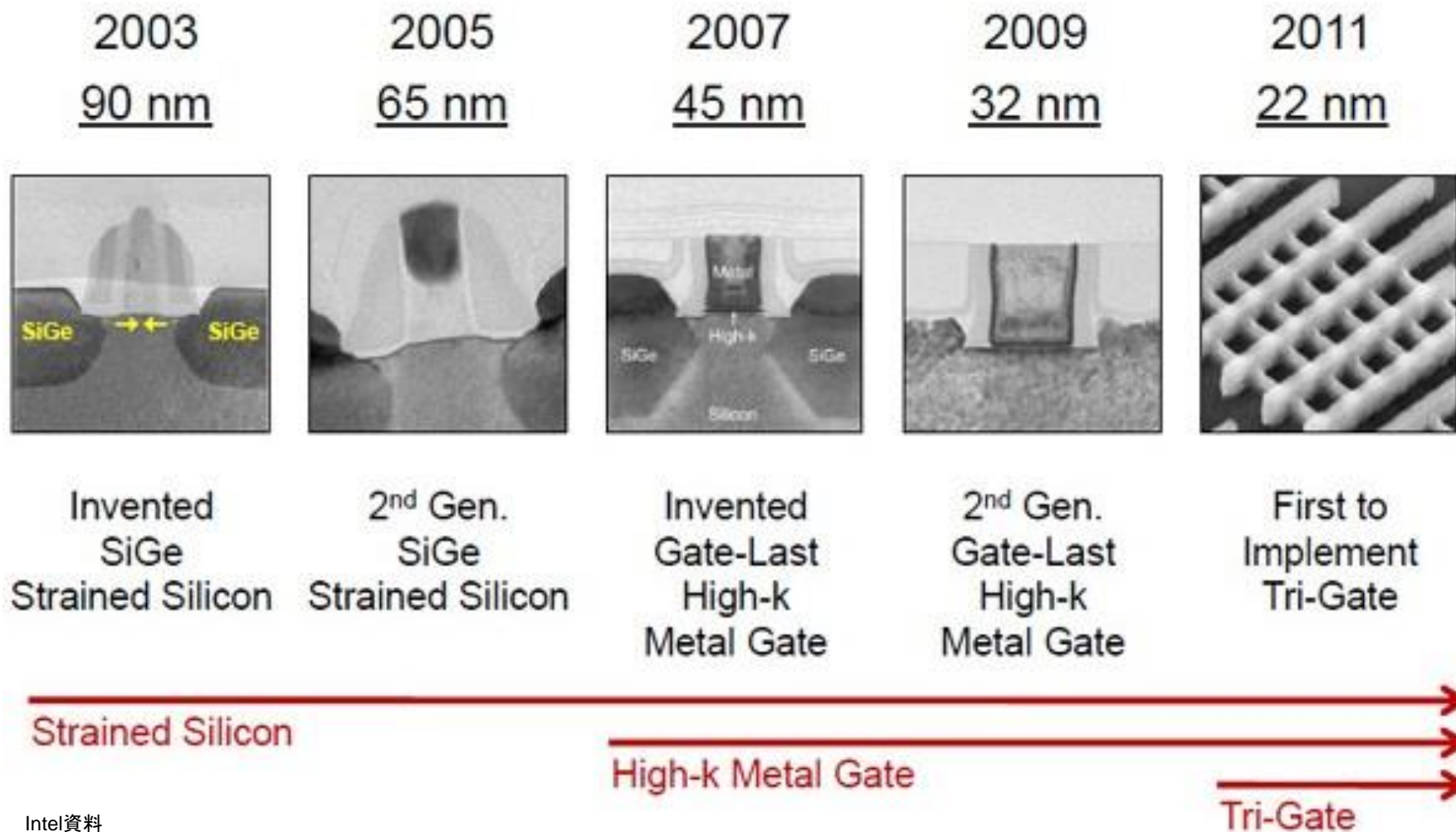
90nm以降の微細化では、特性劣化やリーク電流増加問題に対し新たな技術が採用
 90nm以降で歪シリコン技術が、45nm以降ではHigh-k/メタルゲート技術が、
 22nm以降では3Dトランジスタ構造であるマルチゲート(MuGFET)技術が量産使用。



(See also PIDS, FEP, ERD, and ERM chapters' text and tables for additional detail)

Intelは微細化や高性能化でリードしてきた

- Intelは微細化、高性能化技術を他社に先駆け導入してきた。
 - 歪シリコン、High-k MGや3DTri-Gate(FinFET)技術は約3年半先行
- 2014年頃から、TSMCやSamsungなどが追い上げ、時間差が縮小へ



高性能化技術：歪シリコン技術

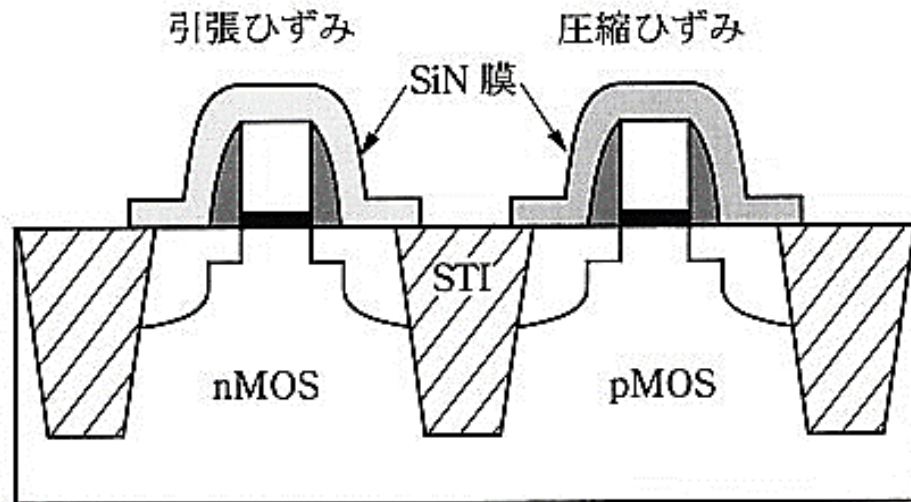
MOSTランジスタのチャンネル部に物理的ひずみを印加してキャリア移動度を高める。

局所ひずみ印加方法

一軸性ひずみの場合は、NMOSとPMOSで印加すべき歪方向が異なる。

SiN膜の成膜条件を工夫し、2種類の膜を堆積。NMOSが引っ張りひずみ、PMOSが圧縮ひずみを印加すると移動度が高まる。

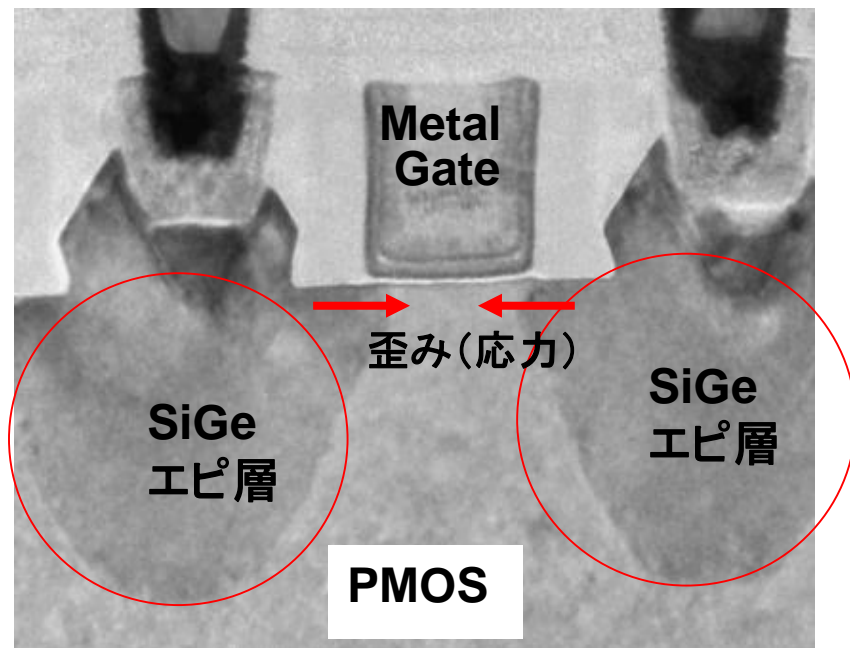
PMOSでは55%も移動度が高まる。



Intel第4世代歪みシリコン技術(32nm)

SiGeエピ層のGeがチャンネルに歪み与える

GeはSiに比べて約4%原子サイズが大きい。このためSiに物理的歪が加わり、電子やホールがSi原子の間を通りやすくなり移動度が高まる。

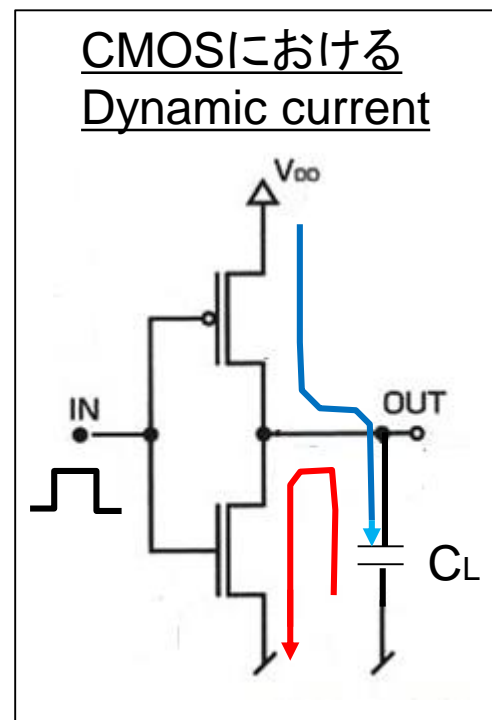
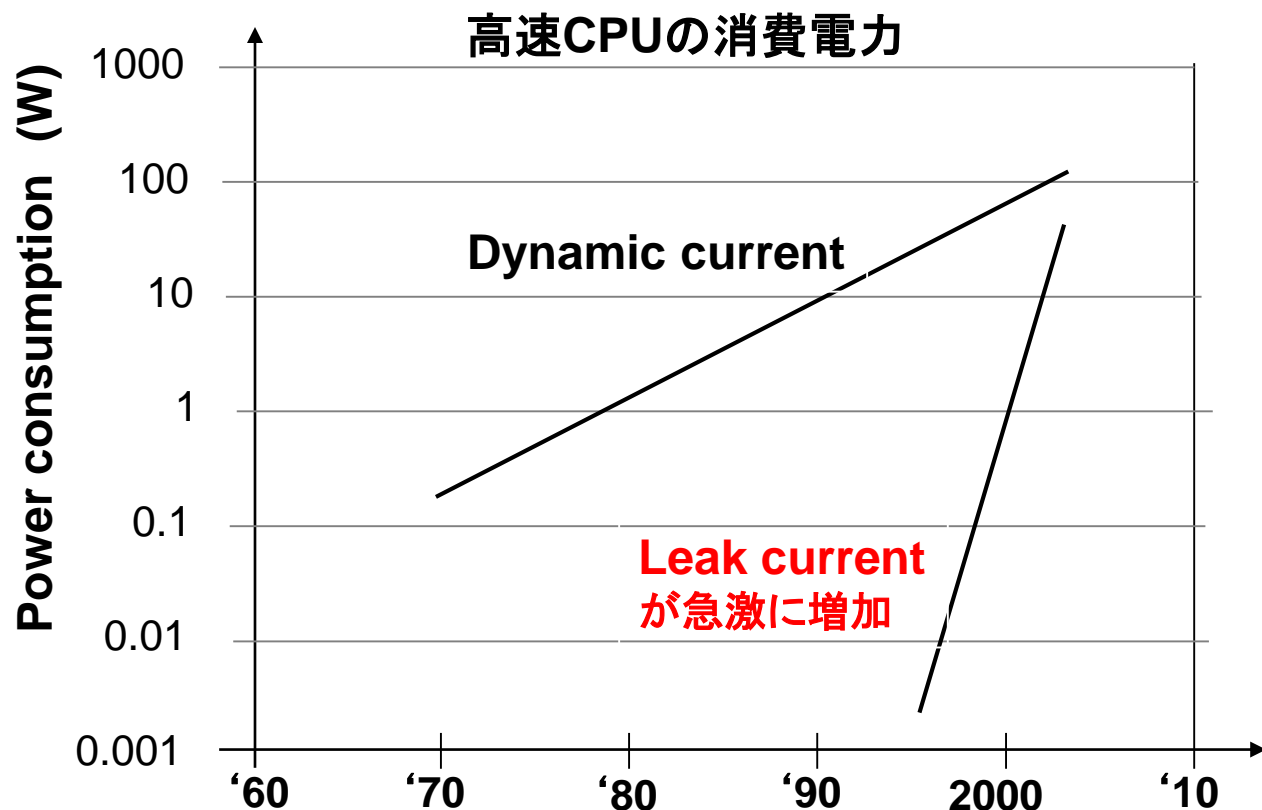


トランジスタの微細化問題:リーク電流の増加

微細化でリーク電流による消費電力増大が大きな問題に(特にスマホなどの携帯機器)

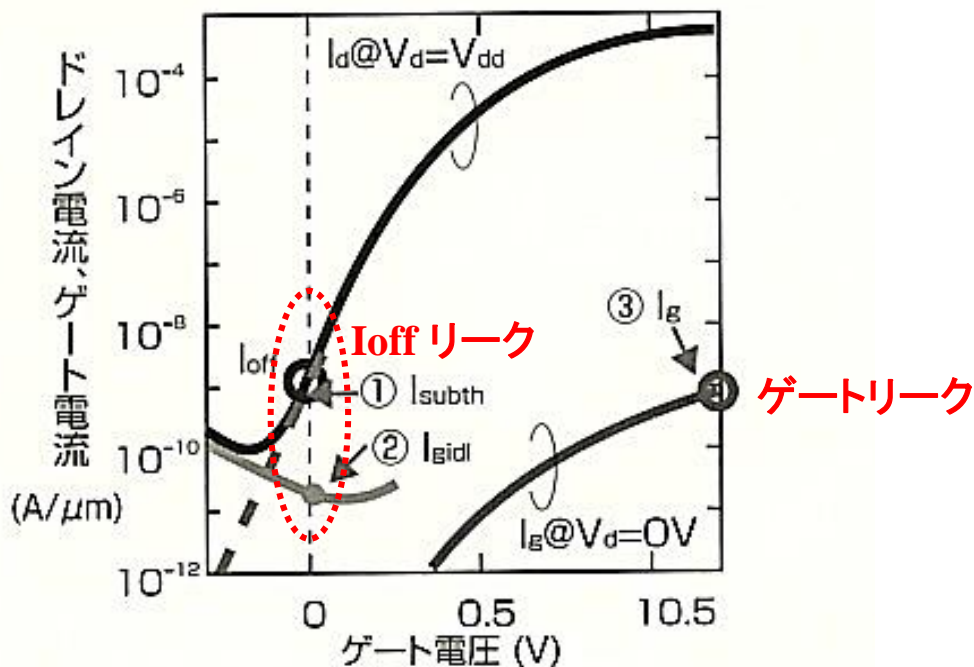
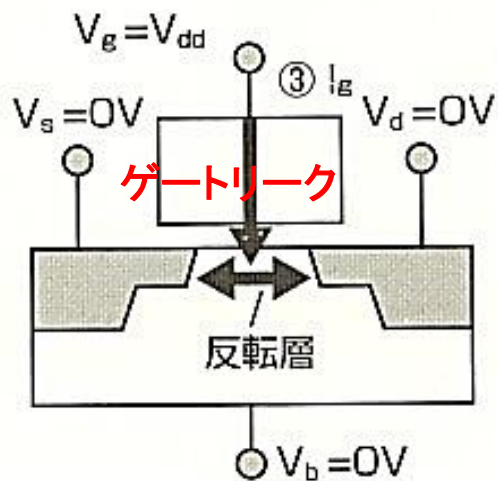
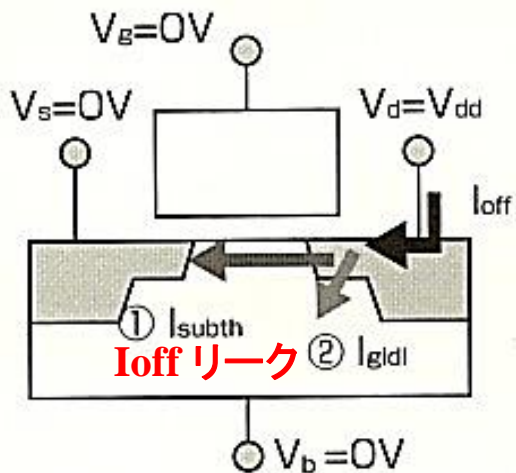
消費電力 $P_d \approx \underbrace{fC_L V_{dd}^2}_{\text{容量のスイッチング電力}} + \underbrace{I_{leak} V_{dd}}_{\text{微細化により影響が顕著に}}$

f: 動作周波数、 C_L : 容量、 V_{dd} : 電源電圧、 I_{leak} : 漏れ電流



微細化における問題点：リーク電流の増大

微細化によるスタンバイリーク電流の増大



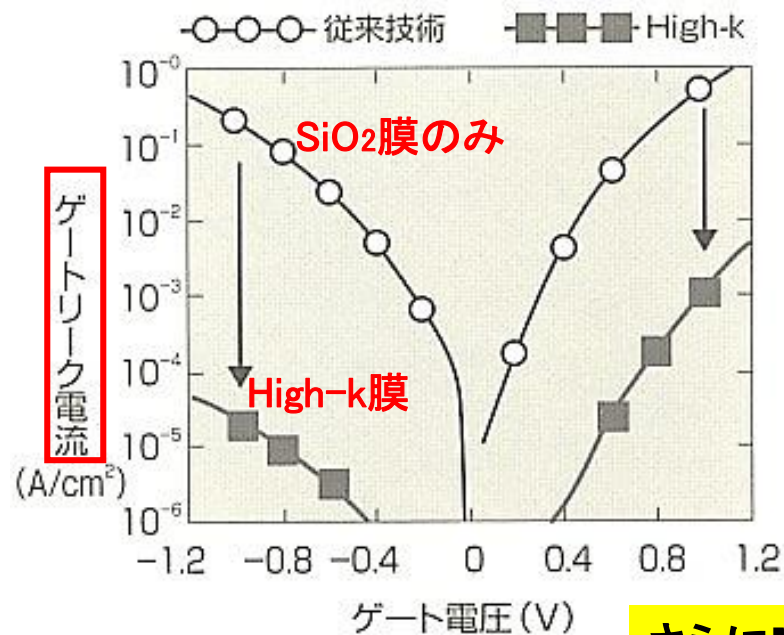
スタンバイリークを構成する3つのリーク成分

- ① I_{subth} : サブスレシヨルドリーク
 - ② I_{gidl} : Gate Induced Drain Leakage
 - ③ I_g : ゲートリーク
- } I_{off} オフリーク

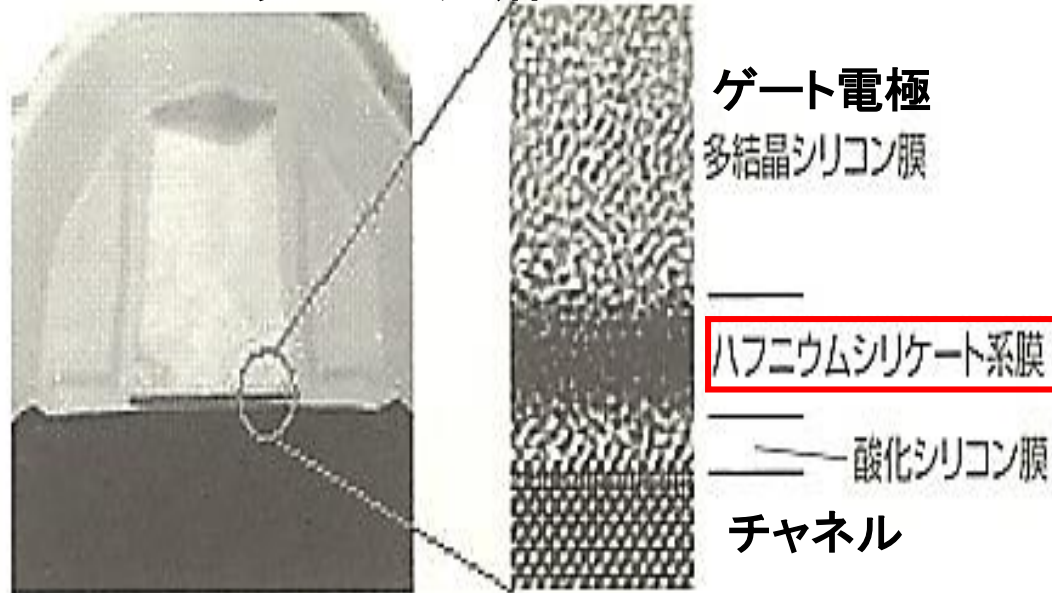
ゲートリーク電流対策: HKMG(High-k/Metal Gate)

ゲートリークの低減 (High-k材料)

ゲート酸化膜が薄くなり(1nm)、トンネル電流(リーク)が増大する。
そこで、**High-Kゲート絶縁膜**使用によりゲート酸化膜の厚みを確保して、大幅なゲートリークを低減。
(SiO₂酸化膜のみより2桁以上の低減)



MOSTランジスタ断面



代表的なゲート絶縁膜の誘電率

- SiO₂(二酸化シリコン)の比誘電率=3.9
- Si₃N₄(窒化シリコン)=7
- HfSiO(ハフニウムシリケート)=12~20**
- HfO(ハフニウム酸化物)=20~24

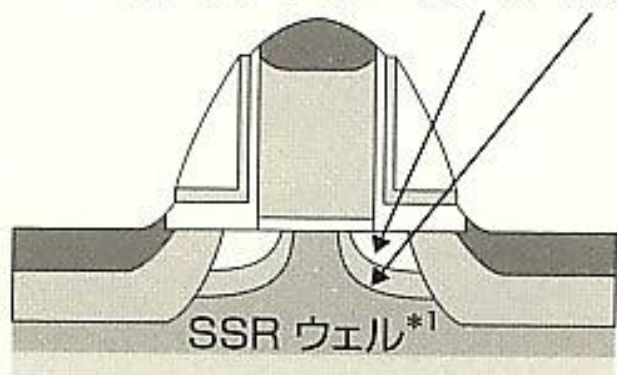
さらにMG(Metal Gate):ゲート材料をポリシリコンからメタル材料にかえてランジスタ性能を向上

loff リーク電流対策

ボディバイアス:

サブスレッショルドリークの低減。
ウェルにバイアス印加し閾値を上昇。
(NMOSでは負電圧印加)

最適化した
エクステンション・ポケット構造

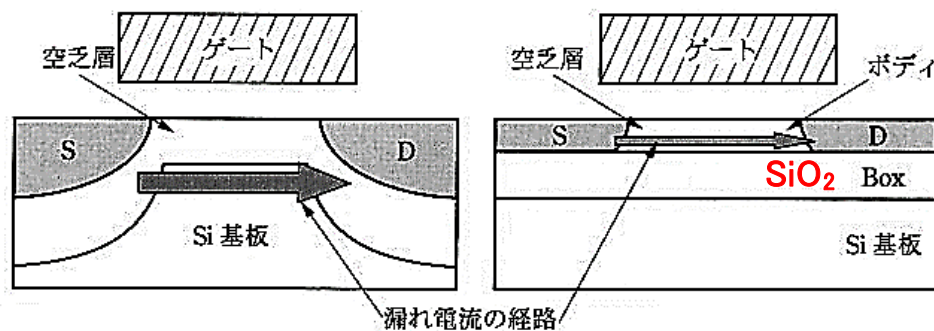


90nm以降でのボディバイアス技術

1. エクステンション/ポケット構造の最適化によりショートチャンネル効果およびGIDLを抑制
2. SSRウェル*1採用 (表面は低濃度、深部で高濃度の逆傾斜構造)

SOI (Silicon on Insulator):

SOI基板では、埋め込み酸化膜 (BOX) 上の薄いSi層にチャネルが形成される。
これがドレイン-ソース間の電流通路となるが、この通路が極めて薄いため、ゲート電界で完全制御され漏れ電流が抑えられる。
右図のFD SOI (完全空乏型SOI)



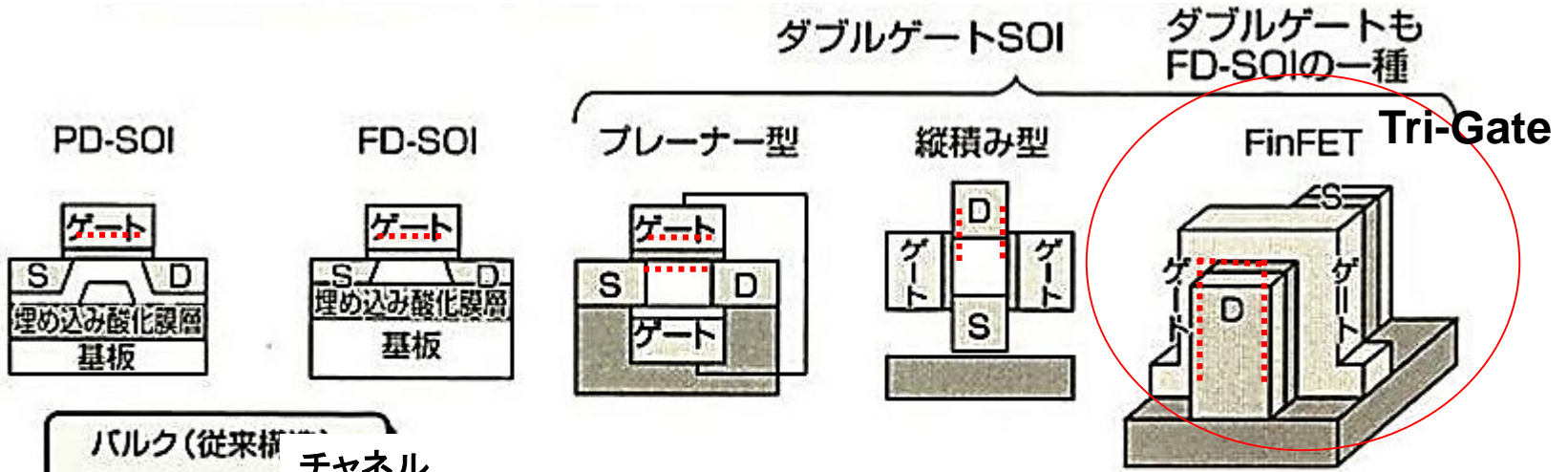
(a) バルク MOSトランジスタ

(b) 完全空乏型 SOI MOSトランジスタ

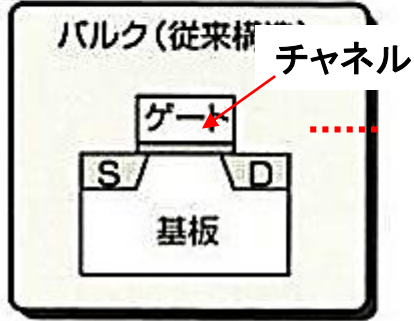
SOI技術はIBM、AMDやSTMなどが積極的に利用。
IntelはSOI技術は使用していない。

3D Tri-Gate/FinFETトランジスタ

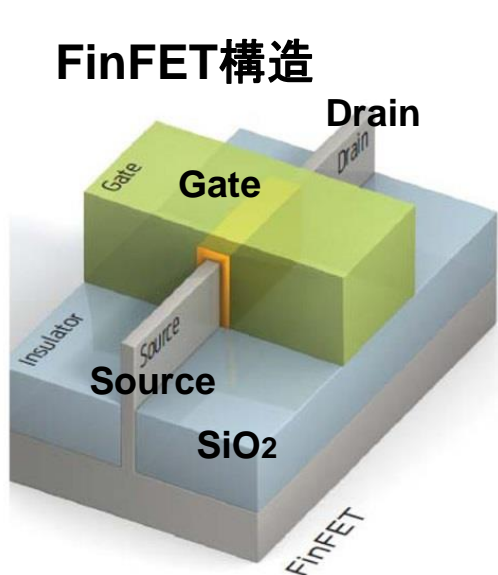
従来の2次元構造FETではゲート電圧でS-D間の電流を完全にON/OFF制御できなくなる。
このためチャネル周囲をゲートで囲うTrigate/FinFETなどの3次元構造が必要となる



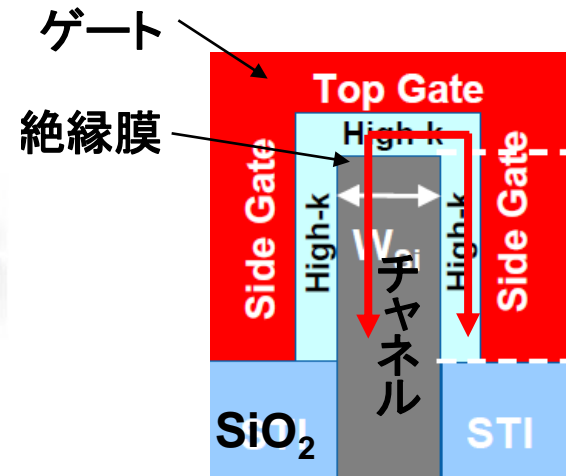
現状の
プレーナ
FET構造



- ・3D Tri-Gate/Fin FETトランジスタは、チャネルを3方向に囲われたゲートで制御する3次元構造
- ・動作は完全空乏型
- ・高性能化と低消費電力（リーク電流低減）を同時に満たす



Trigate/FinFET断面



Intel 第一世代22nmFinFETと第2世代14nmFinFET

2011年後半～

22nm FinFET

2013年後半～ 14nm FinFET

60nm Finピッチ

Finの本数で
駆動能力調整

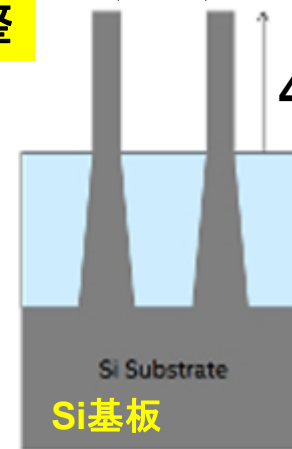
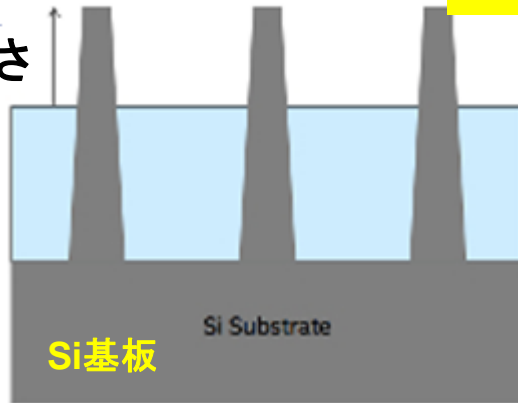
42nm Fin ピッチ

34nm Fin高さ

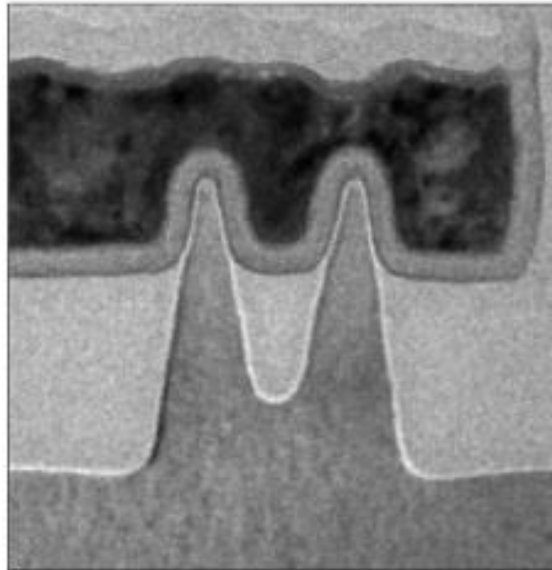
42nmFin 高さ

Gateピッチ: 90nm
配線ピッチ: 80nm

Gateピッチ: 70nm
配線ピッチ: 52nm



SEM
断面写真



第2世代は
狭ピッチ化
高Fin化

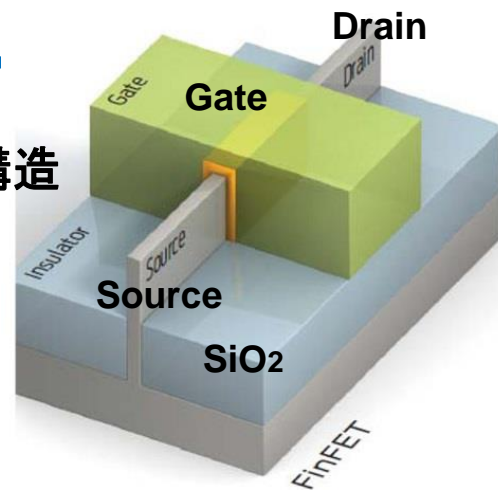
IntelのFinFETプロセス

IntelはIEDM2006で22nm Tri-Gate技術を発表
2011年に22nmで第1世代、2013年に14nmで第2世代、
2017年に10nmで第3世代量産開始

Intelの10nmは、フィンピッチが34nm、フィンの高さが53nm、
メタルピッチは最小で36nm、ゲートピッチは54nm

Intel 10nmは他社のFinFET 7nmプロセス相当

FinFET構造

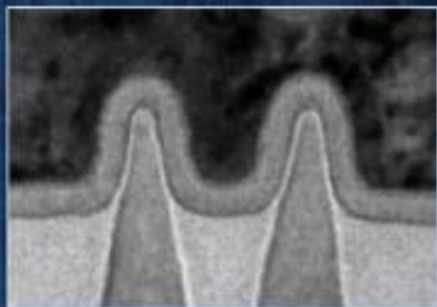


第1世代(2011年～)

第2世代(2013年～)

第3世代(2017年～)

22 NM



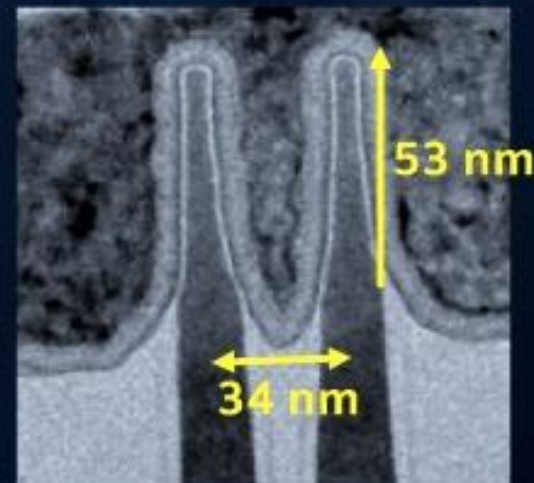
Finの高さ: 34nm
Finピッチ: 60nm
Fin幅: 8nm

14 NM



Finの高さ: 42nm
Finピッチ: 42nm

10 NM



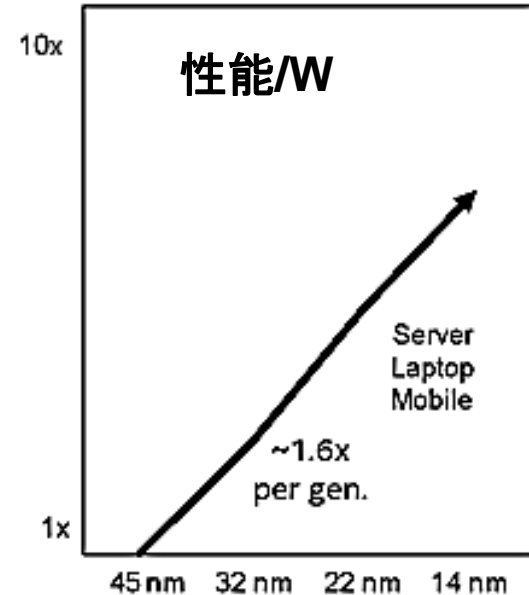
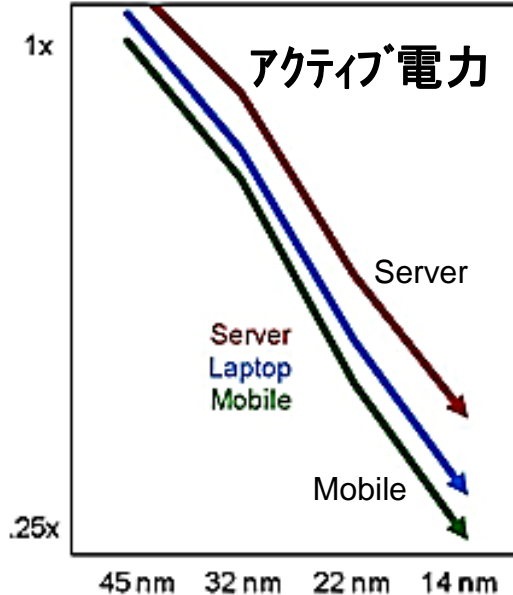
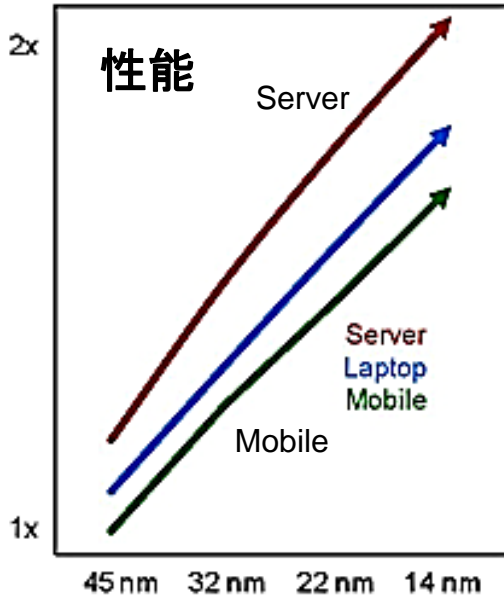
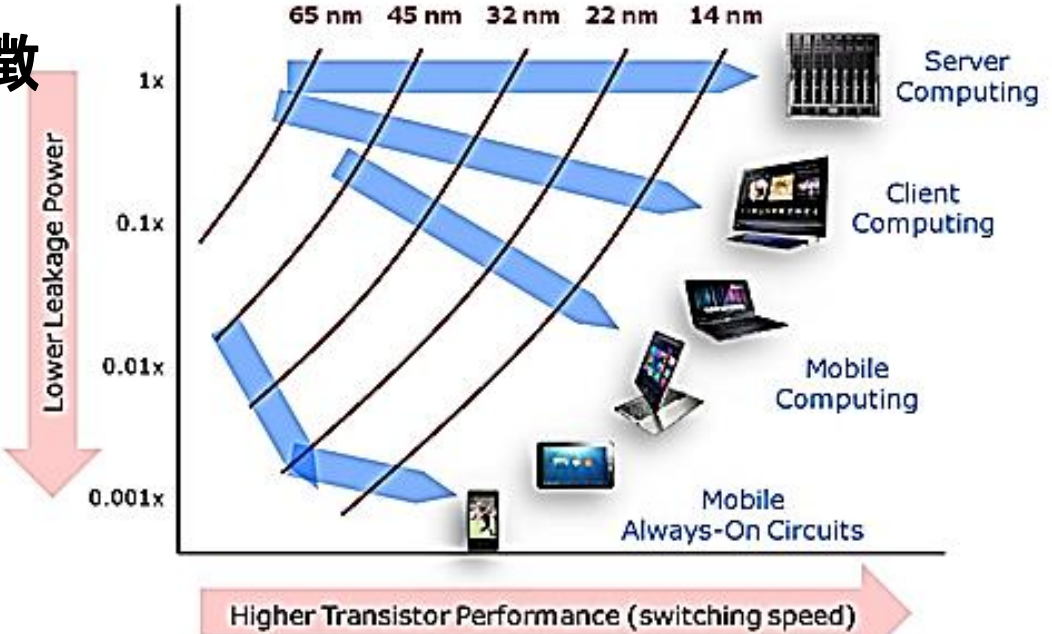
Finの高さ: 53nm
Finピッチ: 34nm

Intel 22nm/14nm

3D FinFETプロセス特徴

- Intelは2014年に第2世代FinFET量産
22nmから14nmに微細化
- FinFETは、高性能に活かす事も
低リーク低消費電力に活かす事も可能

性能 vs リーク電流



プレーナ型NANDの微細化限界

- プレーナ型NANDは、ほぼ16nm世代で微細化の限界。以後はチップ内3D化が必須とのシナリオ
東芝は2014年9月に、“プレーナNANDは15nmで終了し以後は3D化へ移行”と発表
- 2009年当時、2011～2012年に微細化限界を迎えるとしたが、NANDの微細化のスローダウンで、3D化は3～4年程度先延ばし

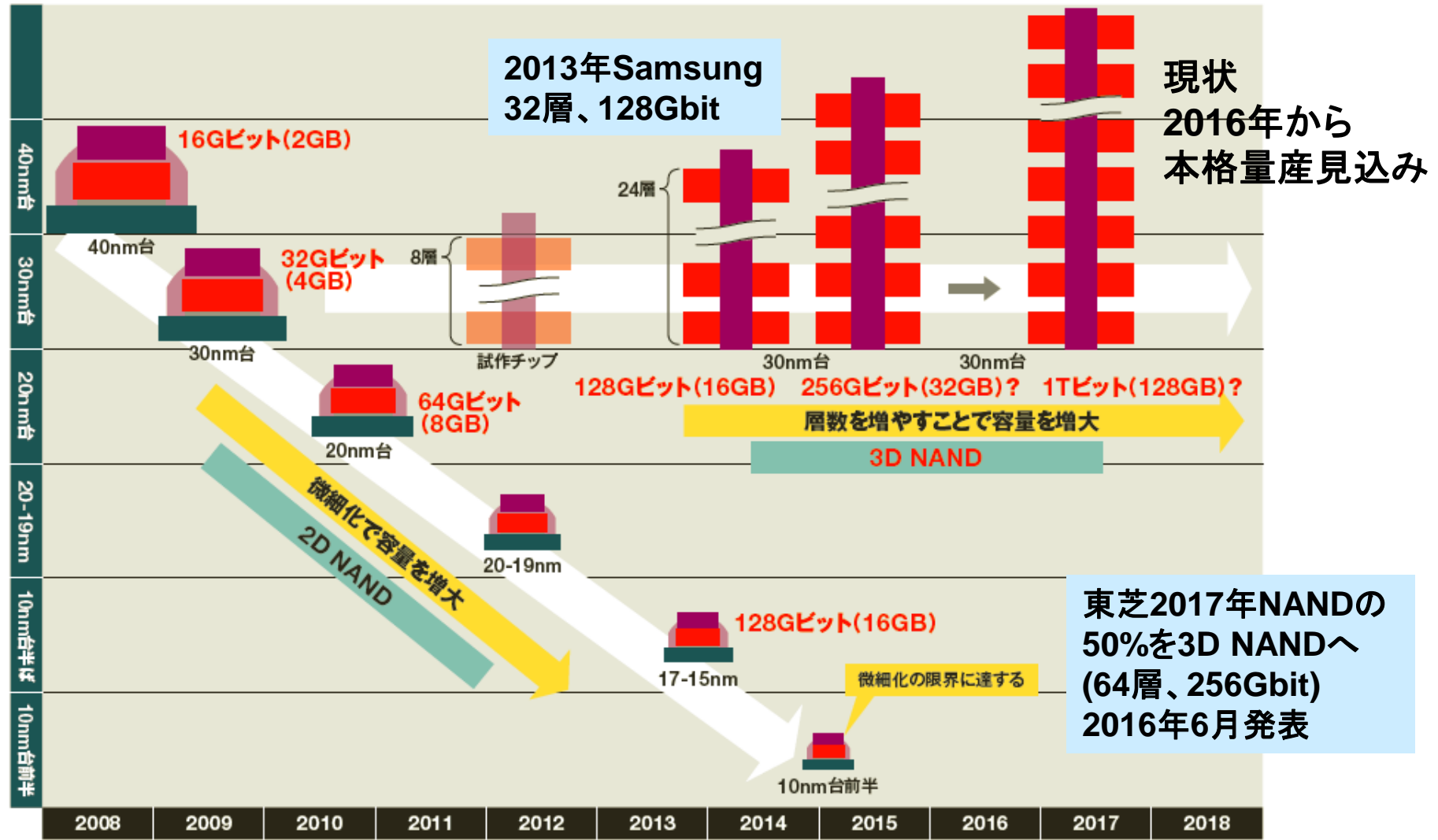
NAND FLASH ROADMAP SUMMARY



NANDの微細化は東芝が先行

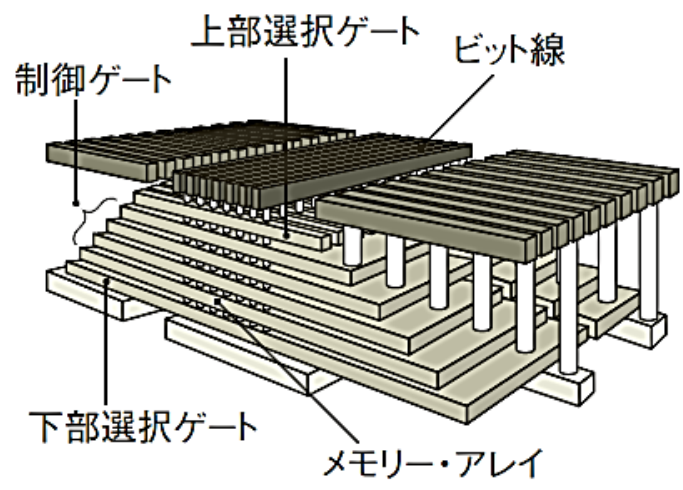
2013年に示されたNANDフラッシュのシナリオ

- ・2014年に10nm代前半プロセスでプレーナ型NAND限界に達する
- ・2015年に3D化に本格移行して256GbitNAND量産とのシナリオ
- ・2014年現在、3D NANDのコスト問題は解決していない(同じ容量で微細化プレーナより高コスト)



3D NAND技術は東芝が2007年に発表: BiCS Bit Cost Scalable

2007年に東芝が発表した3次元NANDフラッシュメモリ技術: BiCS
 BiCSは多層化しても製造コストが上がらない特徴(微細露光回数増えない)



(a) 「BiCS」のアレイ構造

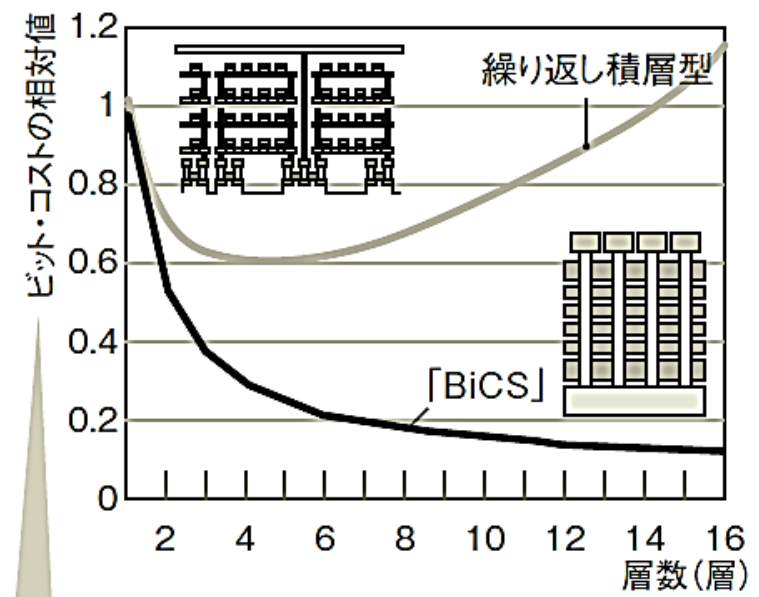


図4 ● 積層数を増やすほど低コストに
 東芝の「BiCS」は、従来のセル積層技術とは異なり、積層数を増やすほどビット・コストを低減できる。1回のリソグラフィ工程で、複数の層を形成できるためである。同社のデータ。

$$\text{ビット・コストの相対値} = \frac{1}{n} (C_f + nC_v) \left(\frac{1+A}{1-Y} \right)^{n-1}$$

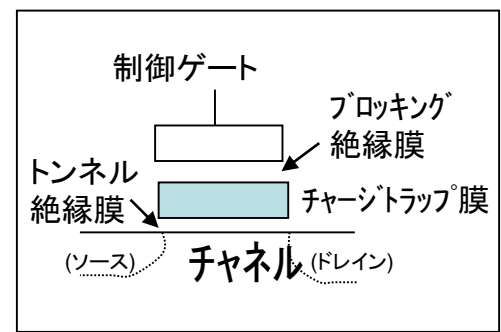
n = 層数
 C_f = 各層が共有する領域のコスト
 C_v = 1層当たりのコスト
 A = 1層当たりの面積ペナルティ率
 Y = 1層当たりの歩留まり損

(b) 「BiCS」のビット・コスト

チップ内3D NANDフラッシュBiCSの構造

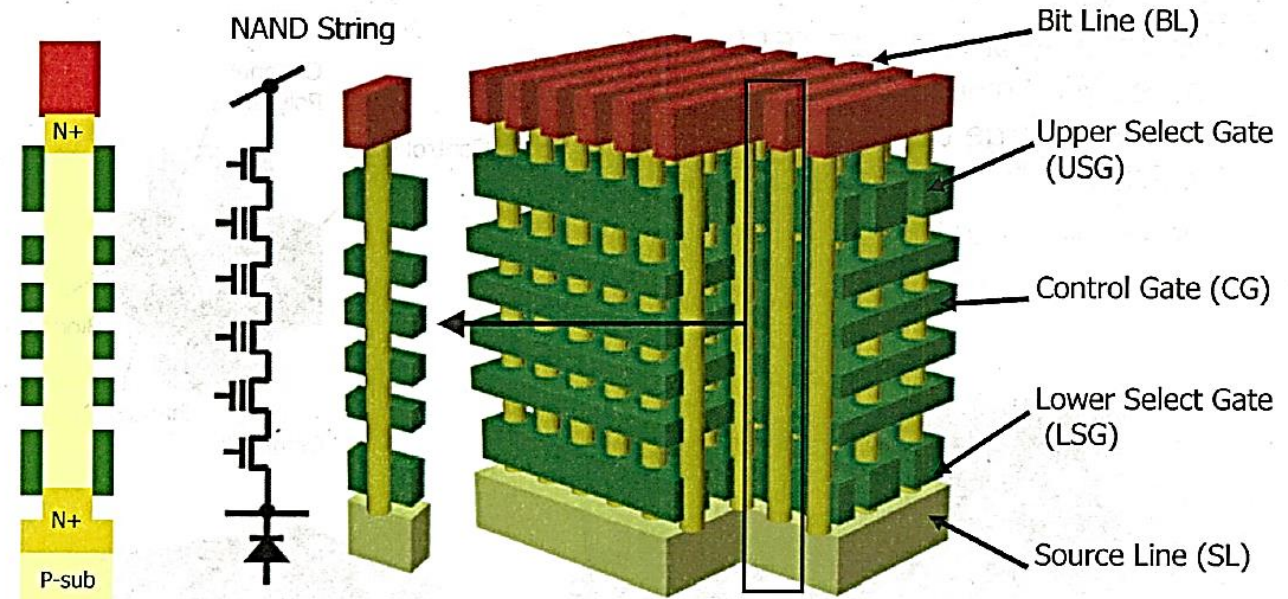
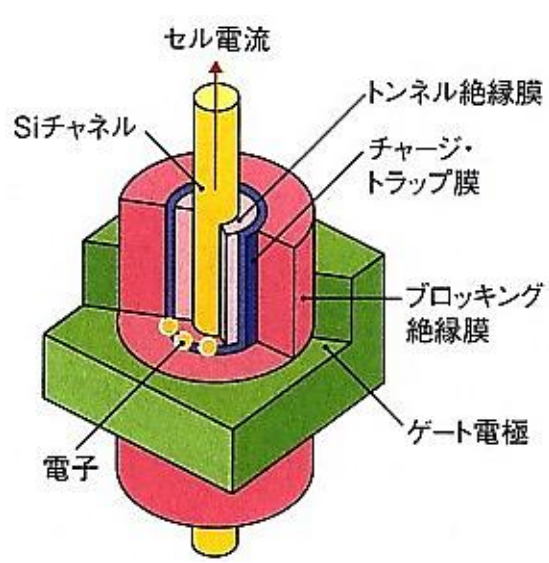
BiCS概要:

電極膜と絶縁膜を交互に積層し、この積層構造を貫通する孔をRIE(反応性イオンエッチング)を用いて一度にあける。孔の側面にメモリ膜(トンネル絶縁膜、チャージトラップ膜、ブロッキング膜)を形成し、孔をポリシリコンで埋める。積層電極膜、孔の側面に形成したメモリ膜、孔を埋めたポリシリコンがチャンネルとなり、セルが多層に一度に形成される。



チャージトラップ型 NANDフラッシュ

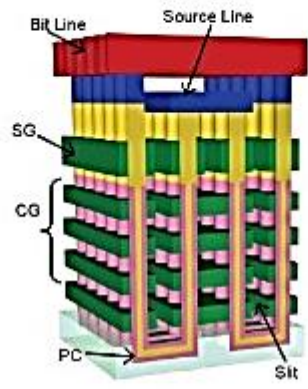
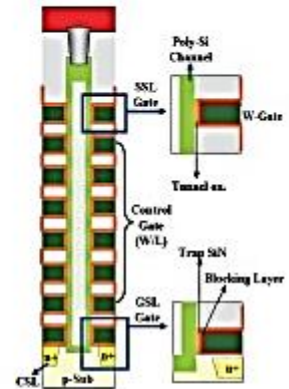
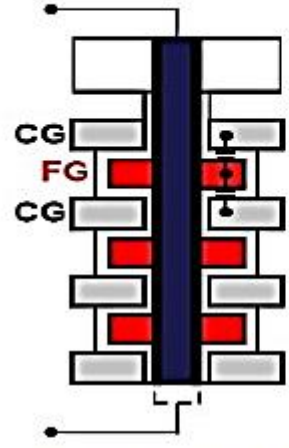
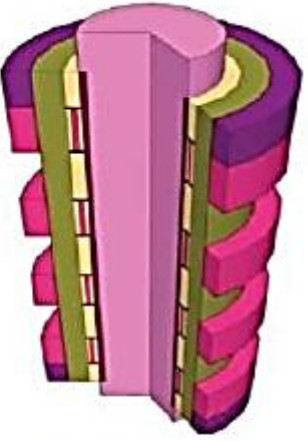
BiCS構造



No S/D for Cell Transistor
Asymmetric S/D for SG

各社Vertical Channel 3D NAND構造

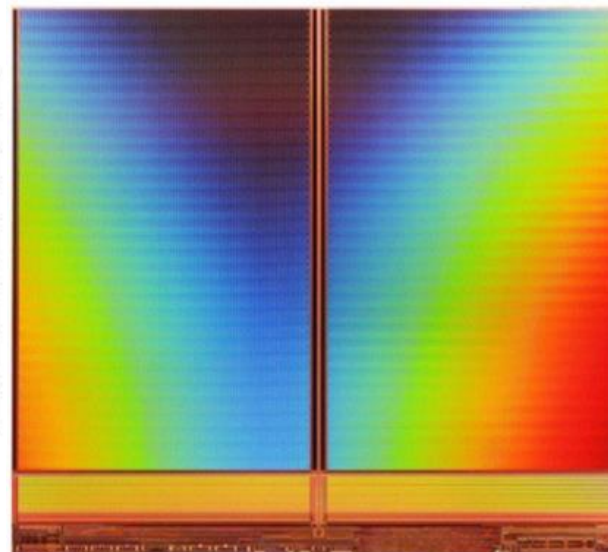
- ・3次元NANDは2007年に東芝がBiCSとして発表
- ・Samsung,Hynix,MicronがBiCS構造をベースに開発
- ・Samsungは2009年にBiCSの問題を改善したTCAT発表。2013年8月V-NANDとして量産開始

	p-BiCS (Toshiba)	TCAT (Samsung)	3D FG (Hynix)	Micron
Structure	 <p>Tanaka, H, VLSIT 2007</p>	 <p>J. Jang, VLSIT 2009</p>	 <p>S. Whang, IEDM 2010</p>	 <p>G. Hawk, FMS 2011</p>
Key Features	- P+ SONOS Cell	- TANOS Cell	- Floating Gate	?
Key Issue	- Large Cell Size - Reliability	- Large Cell Size - SL Resistance	- Process of bit separation - Disturbance	?

V-NANDとして
2013年商品化

東芝3D NAND BiCS 512Gbit 64層

	[1]	[2]	[4]	[5]	[6]	This work
ISSCC paper	2012	2013	2015	2016	2016	2017
Technology	2D 19nm	2D 20nm	3D 32WL layers	3D 48 WL layers	3D Floating Gate (No. of layers: n/a)	3D 64 WL layers
Bits per cell	3	3	3	3	3	3
Capacity	128Gb	128Gb	128Gb	256Gb	768Gb	512Gb
Die size [mm ²]	170.6	146.5	68.9	97.6	179.2	132
Bit density [Gbit/mm ²]	0.75	0.87	1.86	2.62	4.29	3.88
Read operation time	n/a	90us	45us (4 KB page)	45us (4 KB page)	n/a	64us (8 KB page) 80us (16 KB page)
Program throughput	18MB/s	n/a	n/a	53MB/s	44MB/s	46MB/s



Technology:
64-WL-layer 3D Flash Memory

Density and Die Size:
512Gb (3.88Gb/mm²)
132mm²

Organization:
3 Bits/Cell,
(16KB + ECC) / Page,
768 Pages / Block,
(2732 + EXT) Blocks / Plane,
2 Planes

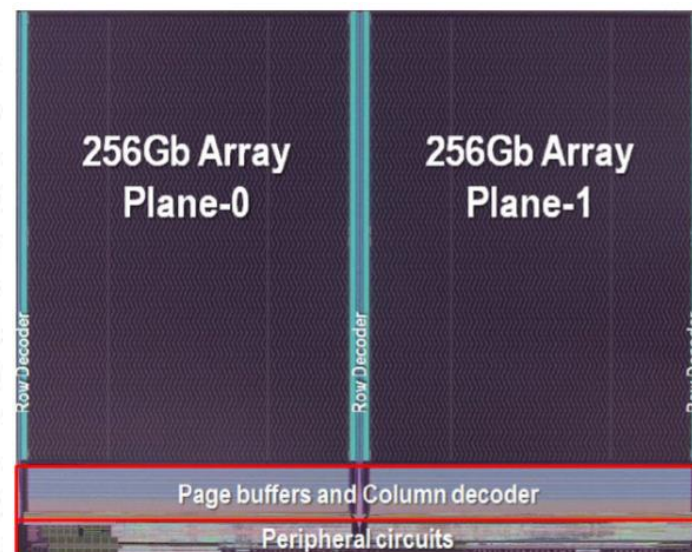
Throughput:
Read: 64us (SBL), 80us (ABL)
Prog: 46MB/s

I/O:
533Mbps double data rate, X8

Power Supply:
Vcc: 2.35V to 3.6V
Vccq: 1.8V

Samsung 3D NAND BiCS 512Gbit 64層

	Previous work [3]	This work
Bits per cell	3	←
Density	256Gb	512Gb
Chip size	97.6mm ² (2.62Gb/mm ²)	128.5mm ² (3.98Gb/mm ²)
Technology	3D NAND with 48 stacked WL layer	3D NAND with 64 stacked WL layer
Organization	16KB/Page, 576Pages/Block, 3776Blocks/Die	16KB/Page, 768Pages/Block, 57488Blocks/Die
I/O Bandwidth	Max. 1Gb/s	←
VCCQ	1.8V	1.8V(Legacy), 1.2V(Low power)
tBERS	3.5ms (Typ.)	←
tPROG	660us	700us
tR (4KB)	45us	60us



3D NAND動向

3D NANDは当初予定より遅延。プロセスが予想以上に難しかった(エッチング工程や膜形成など)装置性能向上や、歩留まり向上にAI技術活用し、2017年やっと本格量産に至ってきた

- ・NAND3社(Samsung, 東芝/WD、Micron-Intel)は64層256G~512Gbit(TLC)を量産開始
- ・東芝は64層で768Gbit(QLC)サンプル提供(2017.6)
- ・東芝は96層512Gbit(TLC)を開発

Samsung Electronicsの3D NAND技術世代

技術世代名称	ワード線の積層数	多値化方式(nbit/セル)	シリコンダイの記憶容量	シリコンダイ面積	発表年月
V1	24層	MLC(2bit/セル)	128Gbit	133平方mm (ISSCC2014)	2013年8月(FMS) 2014年2月(ISSCC)
V2	32層	MLC版とTLC(3bit/セル)版がある	128Gbit	68.9平方mm(TLC版、ISSCC2015)	2014年8月(FMS) 2015年2月(ISSCC)
V3	48層	TLC(3bit/セル)	256Gbit	97.6平方mm (ISSCC2016)	2015年8月(FMS) 2016年2月(ISSCC)
V4	64層	TLC(3bit/セル)	512Gbit	128.5平方mm (ISSCC2017)	2016年8月(FMS) 2017年2月(ISSCC)
V5	未公表(96層?)	QLC(4bit/セル)	1Tbit	不明	2017年8月(FMS)

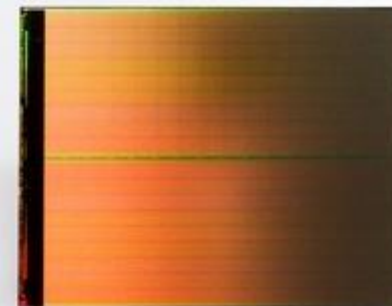
東芝-Western Digital(WD)連合の3D NAND技術世代

技術世代名称	ワード線の積層数	多値化方式(nbit/セル)	シリコンダイの記憶容量	シリコンダイ面積	発表年月
BiCS1	24層	MLC(2bit/セル)	不明	不明	2012年2月 (SanDiskのアナリスト向け説明会)
BiCS2	48層	MLC(2bit/セル)とTLC(3bit/セル)	128Gbit(MLC)と256Gbit(TLC)	不明	2015年8月(FMS)
BiCS3	64層	TLC(3bit/セル)	256Gbit(TLC)と512Gbit(TLC)	132平方mm(512Gbit, ISSCC2017)	2016年8月(FMS) 2017年2月(ISSCC)
BiCS4	96層	TLC(3bit/セル)	256Gbit(TLC)と512Gbit(TLC)	不明	2017年6月(リリース) 2017年8月(FMS)

Intel+Micro 128Gbit 3Dクロスポイントメモリ

- ・X,Y方向配線のクロス部分にメモリセルを形成。
- ・20nmプロセス、128Gビット容量で640億個の2層構造メモリセル
- ・メモリセルは抵抗変化する材料(ReRAMか相変化メモリと推測。詳細発表なし)
- ・NAND型フラッシュメモリよりも1000倍高速で、DRAMよりも8~10倍、記憶密度が高い
- ・NANDよりも1000倍の耐久性

20nmプロセス128Gbit 3D XPoint メモリ



**3D XPoint™ Technology:
An Innovative, High-Density Design**

Cross Point Structure
Perpendicular wires connect submicroscopic columns. An individual memory cell can be addressed by selecting its top and bottom wire.

Stackable
These thin layers of memory can be stacked to further boost density.

Non-Volatile
3D XPoint™ Technology is non-volatile—which means your data doesn't go away when your power goes away—making it a great choice for storage.

High Endurance
Unlike other storage memory technologies, 3D XPoint™ Technology is not significantly impacted by the number of write cycles it can endure, making it more durable.

Selector
Whereas DRAM requires a transistor at each memory cell—making it big and expensive—the amount of voltage sent to each 3D XPoint™ Technology selector enables its memory cell to be written to or read without requiring a transistor.

Memory Cell
Each memory cell can store a single bit of data.

Transforming the Memory Hierarchy
For the first time, there is a fast, inexpensive and non-volatile memory technology that can serve as system memory and storage.

~8x to 10x Greater Density than DRAM!
3D XPoint™ Technology's simple, stackable, transistor-less design packs more memory into less space, which is critical to reducing cost.

3D XPoint™ Technology Processor





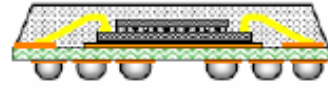
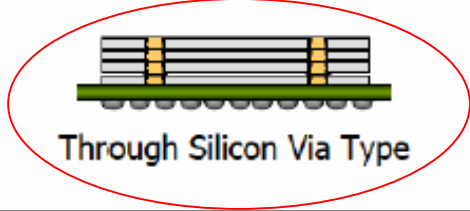
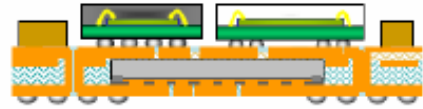
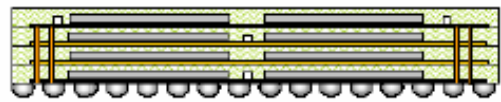

1GB DRAM

3D XPoint™ Technology

Intel+Micronも
3DXPointメモリ量産にたいへん
苦勞している様子。
新たな材料が数十以上も必要。

3次元(3D)実装技術のいろいろ

現在の量産レベル3次元実装は、ワイヤボンディングやパッケージオンパッケージ(PoP)積層レベル

3次元構造		平面構造	 Wire Bonding Type	 Flip Chip Type	WL : Wafer Level
		チップ間 ビア経由接続	 Wire bonding Die-stacked	 Wire Bonding + Flip Chip (WLCSP) Type	
		チップ間 直接接続	 Wire Bonding + Flip Chip (WLCSP) Type		TSV
		内蔵構造	 Chip Embedded + Package on Surface Type	 3D Chip Embedded Type	
			 Package in Package		

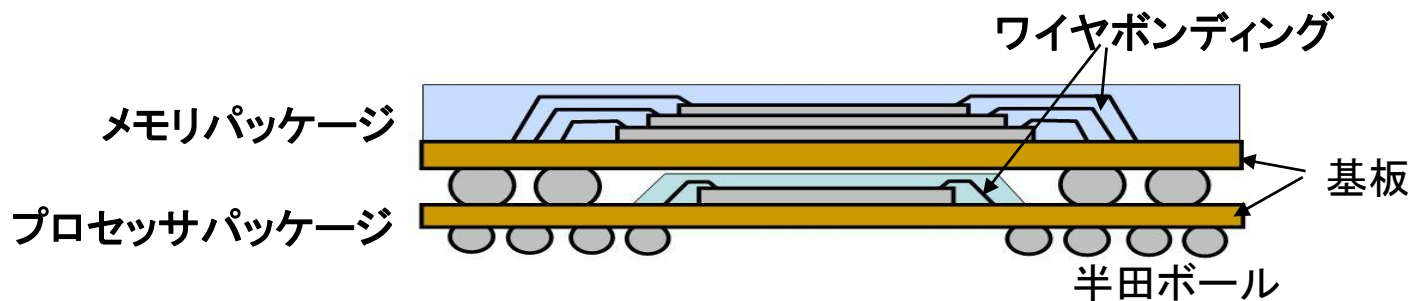
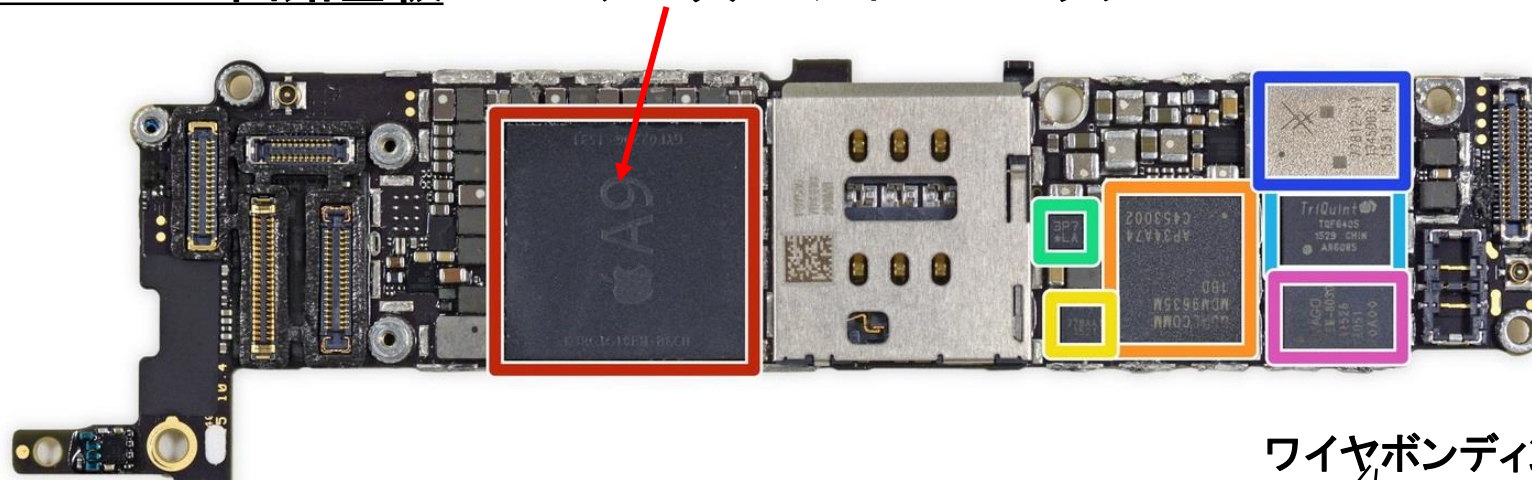
3次元実装技術: PoP(Package on Package)

現在のスマホ用プロセッサとDRAMでは、ほとんどのこのPoP実装技術が使用されている

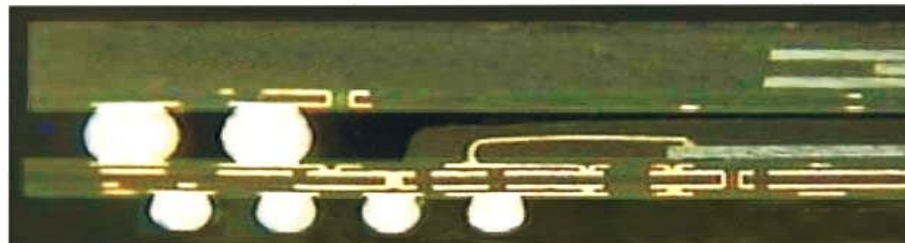
iPhone 6S 回路基板

PoP (Package on Package)

A9アプリケーションプロセッサ+2GB DRAM



一般的なPoP構造の例

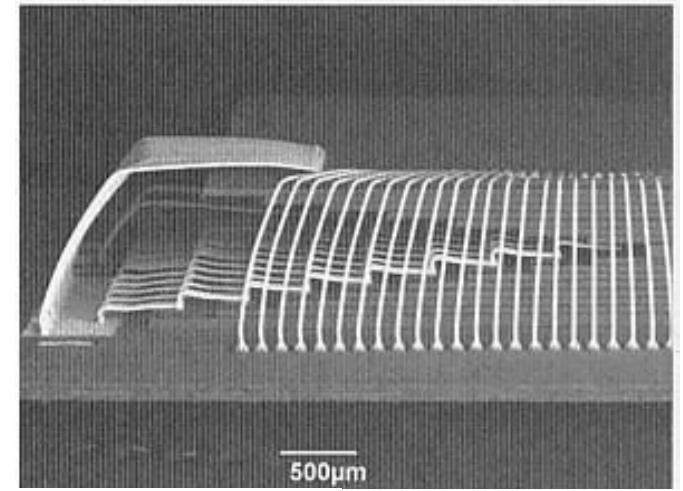


NAND積層技術 ワイヤボンディング：東芝

2009年資料

18 μm極薄のNANDを33段ワイヤボンディングで積層

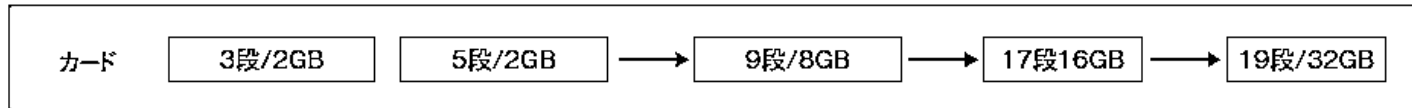
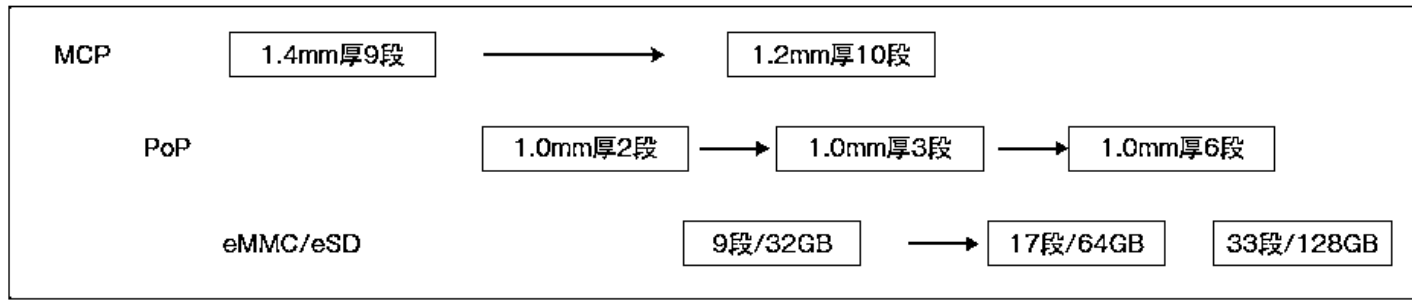
- ・東芝はダイシングしてから裏面を削るDBG技術 (Dicing before grinding)によりウエハを18 μmまで極薄化。
- ・15 μmまでは動作確認済みだが、量産は18 μm限界か。
- ・TSV技術はコスト競争激しいNANDでは当面採用無理
TSV技術は高速DRAMの積層化など高速化でメリット大きい。



要求容量	2005年	2006年	2007年	2008年	2009年	2010年
	4GB	8GB	16GB	32GB	64GB	



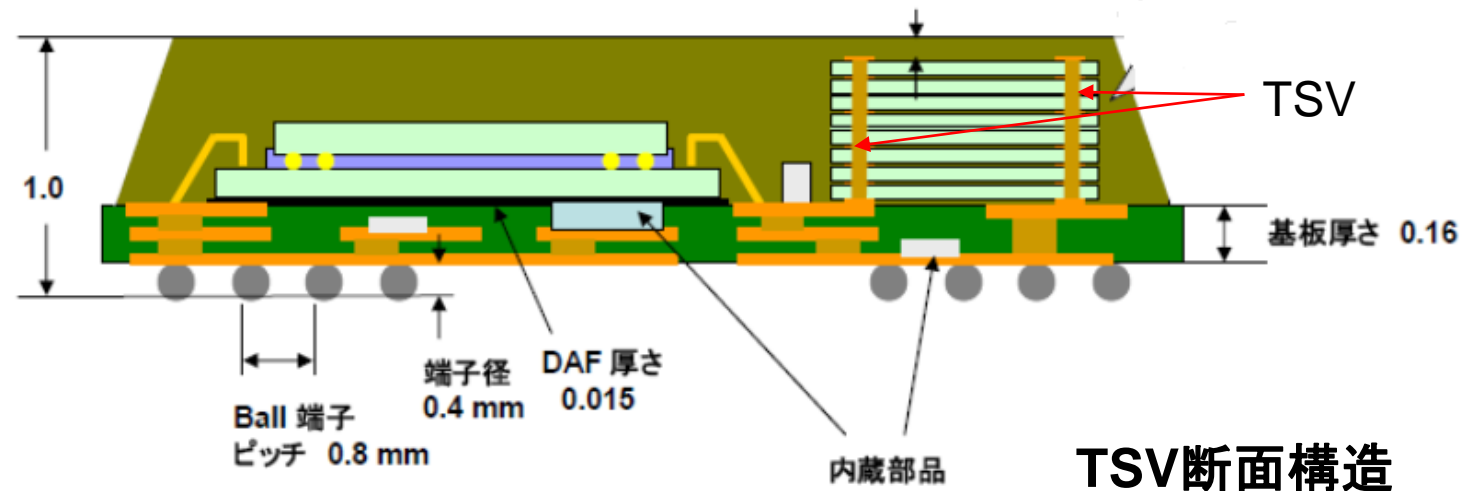
現状の最大容量
SDXCカード512GB
128Gbit Chipだと33枚
256Gbit Chipで17枚
のチップが必要



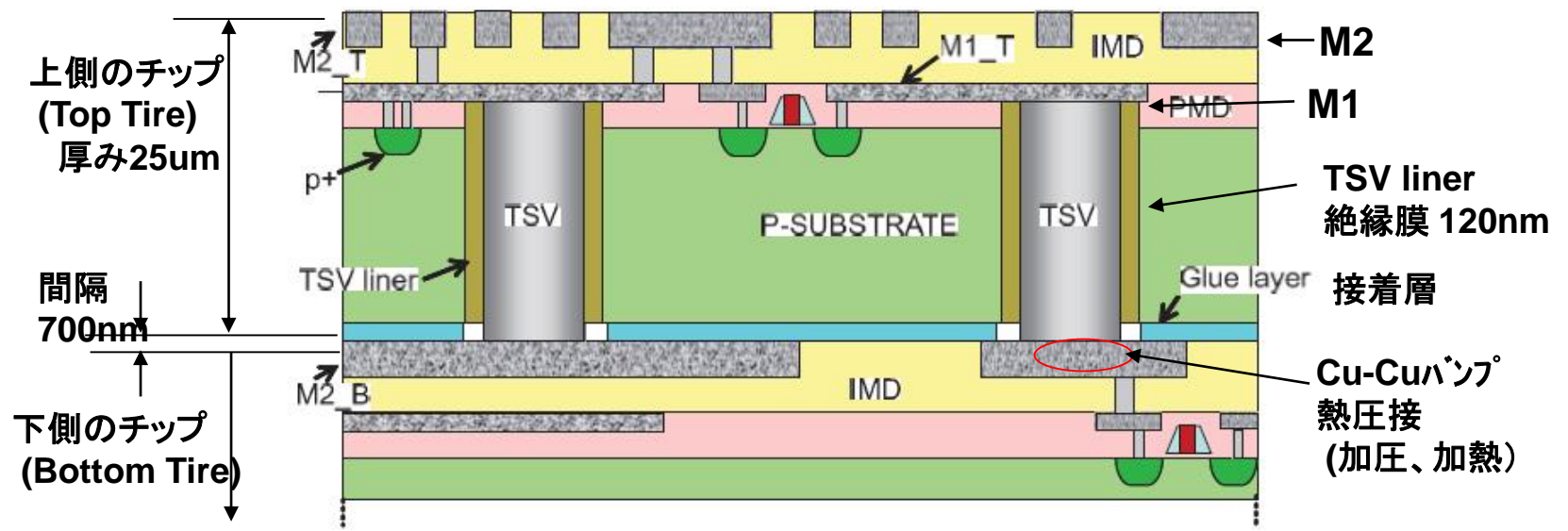
チップ厚	70μm	50μm	25μm	20μm	18μm
------	------	------	------	------	------

3次元実装技術: TSV (Through silicon Via: シリコン貫通ビア)

TSVはSi基板を薄化して、基板上下を貫通する穴を開け、Si基板を積層して接続する技術

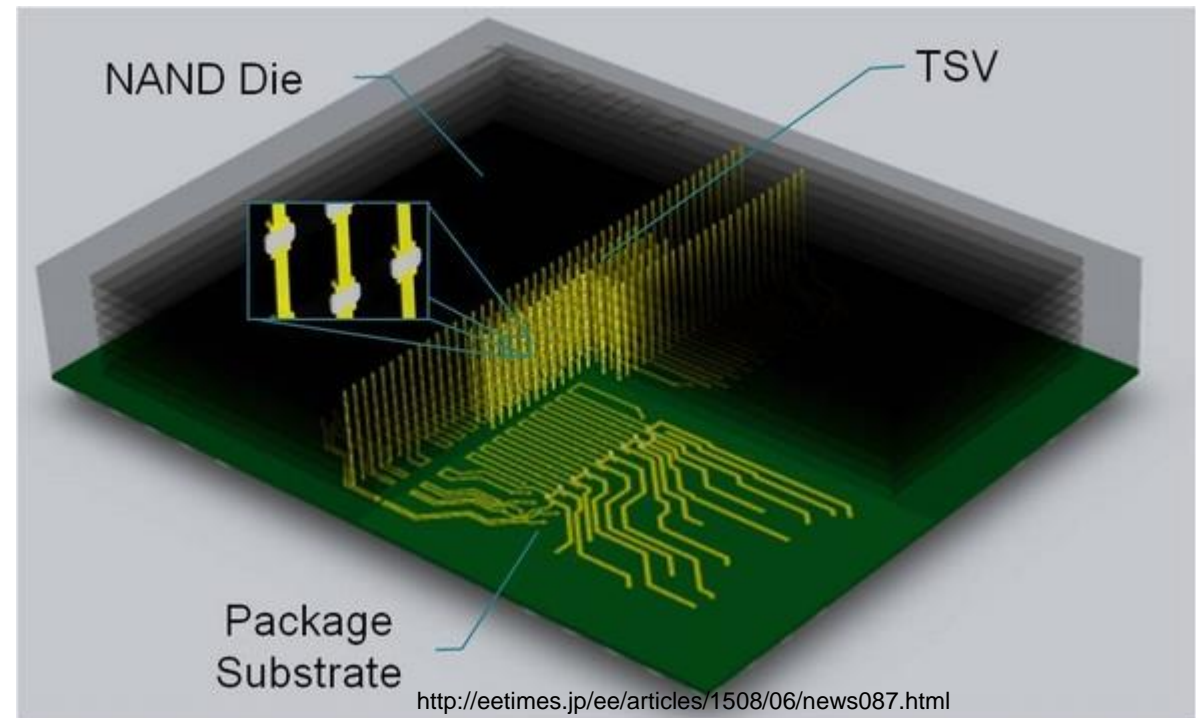
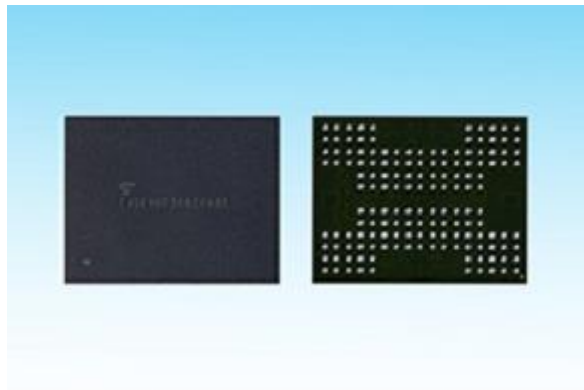


TSV断面構造



IMD (Inter-Metal Dielectric) 層間絶縁膜
PMD (Pre-Metal Dielectric) 膜

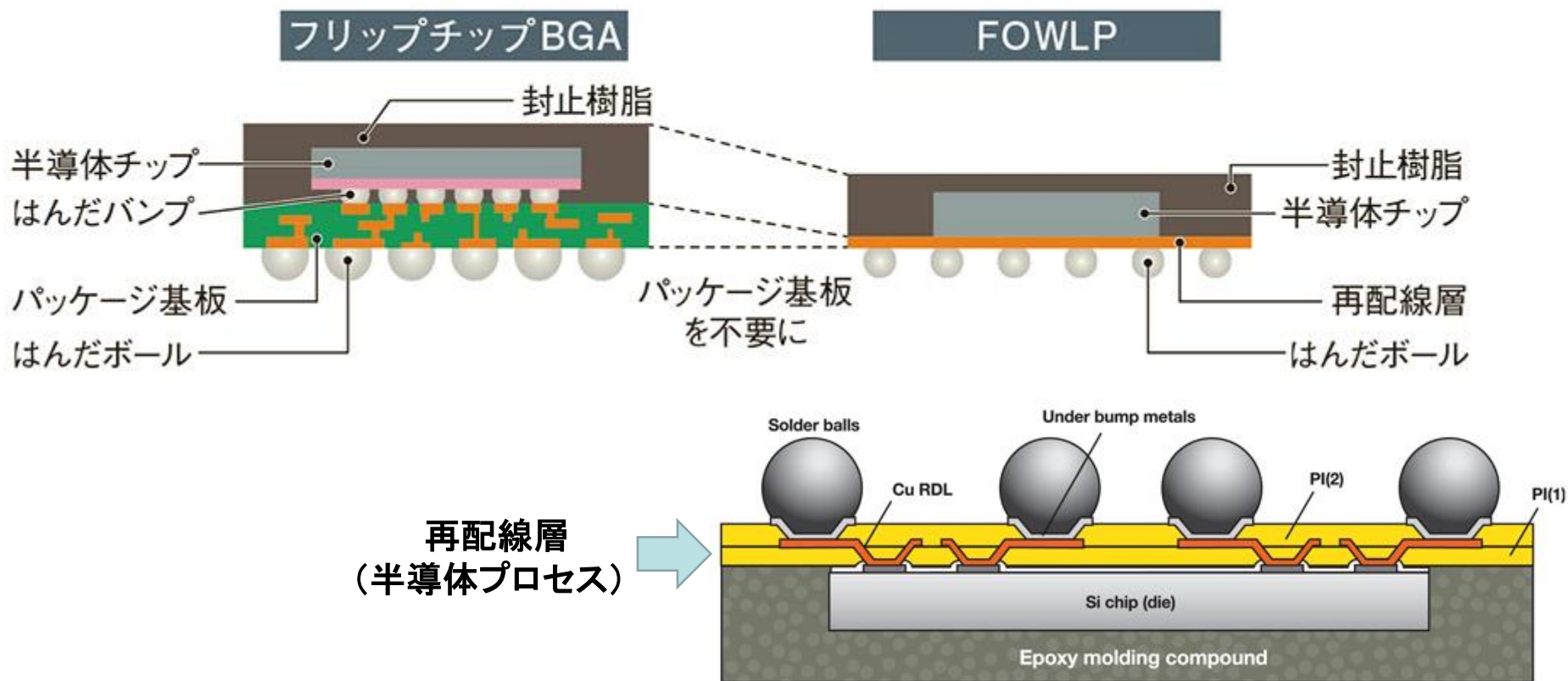
- ・48層3D NAND にTSVを適用し、書き込みバンド幅と低消費電力性能を向上
ワイヤボンディングを用いた製品に比べて電力効率を約2倍向上
- ・512G bit(TLC)チップを単一パッケージ内に16段積層し、総容量1Tバイトの大容量化
- ・BiCS FLASHはパッケージがNAND Dual x8 BGA-152で、インタフェース規格がToggle DDR
- ・容量512GB(8層)の製品と容量1TB(16層)の製品があり、
サイズは512GBが14×18×1.35mm、後者が14×18×1.85mm
- ・サンプル出荷は2017年中に開始予定



FOWLP (Fan-out Wafer Level Package)

AppleがA10で採用 (iPhone7)

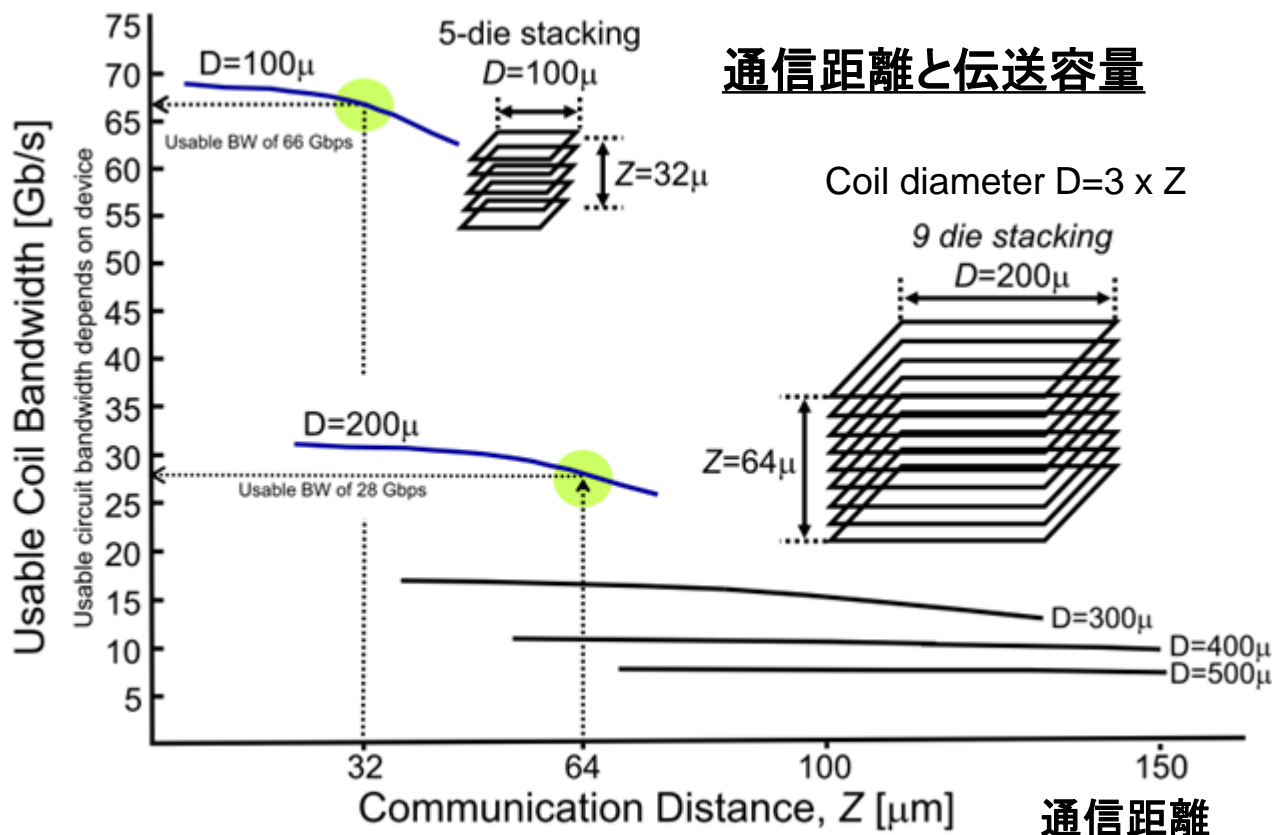
従来のフリップチップBGAに置き換わる実装方式として、「FOWLP」が注目を集めている。大きな違いはパッケージ基板がないこと。代わりにチップの端子から配線を引き出す再配線層を半導体工程で作成し、外部端子につなげる。パッケージ基板がないため、「パッケージが薄い」、「配線長が短く伝送が速い」、「パッケージ基板の分、コストが安い」といった特性を実現できる



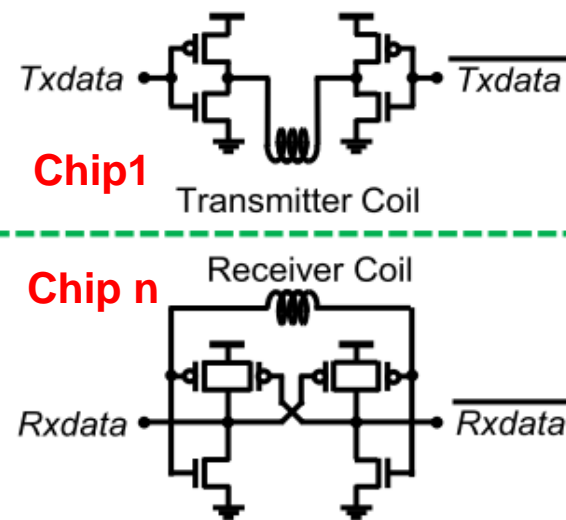
チップ間、磁界結合 : Thru Chip Interface (TCI)

- ・慶応大学黒田先生チームが研究してきたチップ間、磁界結合技術が実用化に向かいだした
- ・従来コイル寸法が大きすぎたが、チップの薄化技術(DRAMを4 μm 厚に研磨しても正常動作)で、8枚積層して60 μm サイズのコイルで通信可能になった。
- ・チップの薄化以外に追加プロセス不要。
またTSVのようにESD対策ダイオードも不要

PEZY(国内ベンチャ)は次期スパコンchipに、このTCI技術を採用予定



通信回路は極めてシンプル



送信、受信用コイルの巻き数は4 turn

More MooreとMore than Moore:

- ・More Mooreによる微細化がスローダウン。等価スケールリングにてMooreの法則維持
 今後はチップ内3次元化と高移動度チャネル材料がMore Mooreのキー
- ・近年More than Mooreの多様化に関する論文発表が多い。異種チップの融合など

