Variation in 45nm and Implications for 32nm and Beyond

Kelin J. Kuhn

Intel Fellow

Director of Advanced Device Technology

AGENDA

Technology scaling

I. Physical Variation Sources and Mitigation

II. Measurements, results and interpretation

II. Next generation challenges

Closing thoughts

Technology Scaling

Lithography Scaling Limitations
From Broers [1] IEDM Plenary Session 1980

In the limit, microscope objectives with 0.95 N.A. are available and, provided very small fields (200μ x 200μ) are adequate, linewidths < 0.4μ should be achievable under carefully controlled laboratory conditions, and in very thin resist layers.

Depth of field will be reduced to about $\pm 0.2\mu$. Deep U.V. (λ = 200nm - 260nm) lenses will be difficult to build because of the lack of materials that are transparent at these wavelengths and yet have relatively high refractive indices.

1980:
Optical Lithography Limit
~ 400nm

intel

Transistor Scaling Limitations
From Meindl [2] IEDM Plenary Session 1983

For conservative design margins, typical results suggest that IGFET channel lengths can be reduced to approximately 0.40 microns in E/D NMOS logic gates; 0.30 microns in CMOS transmission gates; and 0.20 microns in E/E CMOS logic gates. Smaller channel lengths can be projected for more aggressive designs. The dominant mechanism imposing these limits is subthreshold drain current due to short channel charge sharing and drain induced barrier lowering.

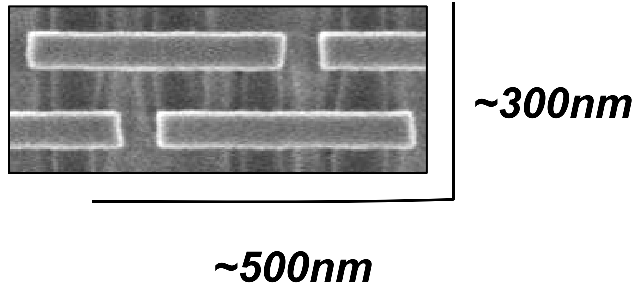
1983:
Transistor architecture limit
200-400nm (SCE)

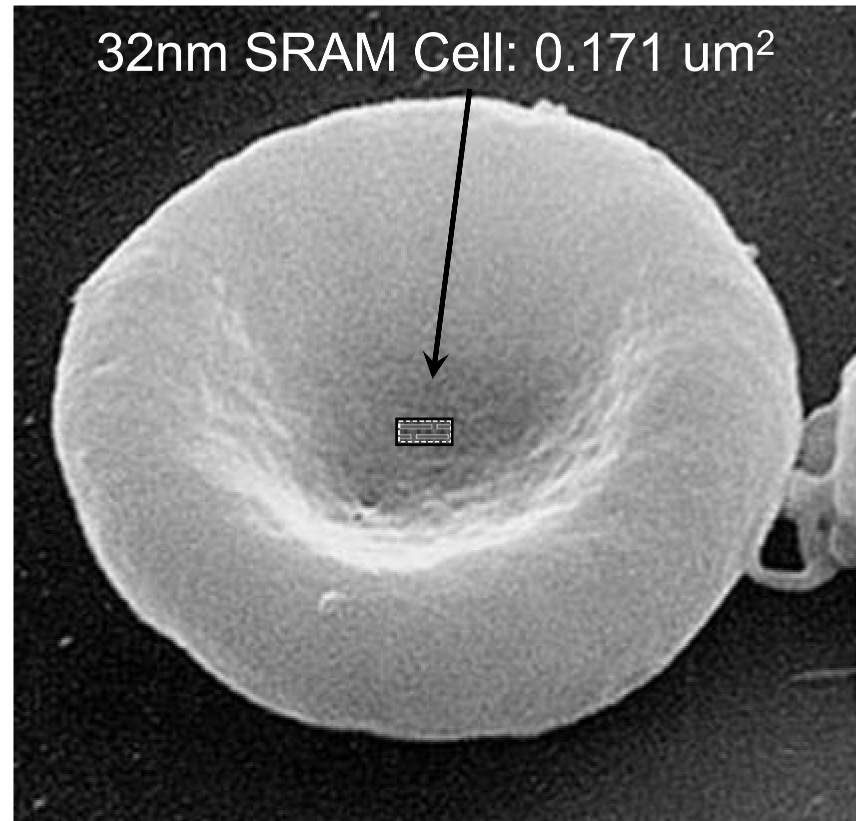
Transistor Scaling Limitations
From Heilmeier [4] IEDM Plenary Session 1984

other factors limiting the scaling of ICs come into play. Some of these factors are interconnect capacitance, channel capacitance, interconnect resistance, parasitic resistances, velocity saturation, ionizing radiation, drain breakdown, gate oxide breakdown, hot carrier injection, subthreshold current, punch through, and statistical control of oxide thickness and channel doping. It appears that minimum geometries for high-volume ICs will saturate in the range of 0.3 to 0.5 microns.

1984:
Transistor architecture limit
300-500nm (laundry list of reasons…)

How small is a 32nm memory cell?





Blood cell: Elec. Mic. Fac. (NCI-Frederick) 2007

1983-84 limits on gate size, are commensurate with the dimensions of 2008's entire 32nm SRAM cell!

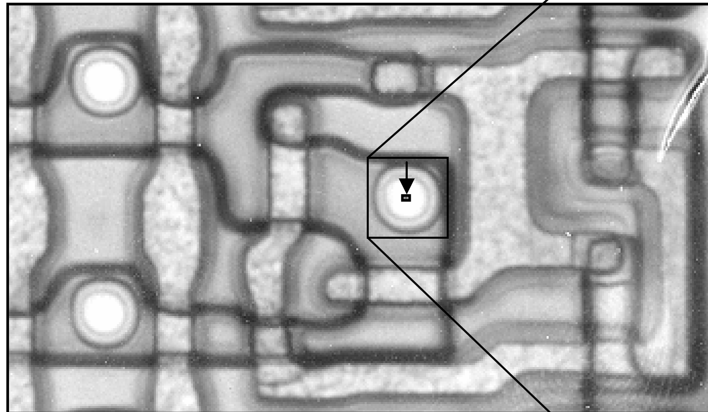
Small enough that a 2008 32nm SRAM cell is dwarfed by a human redblood cell

K. Kuhn 2007

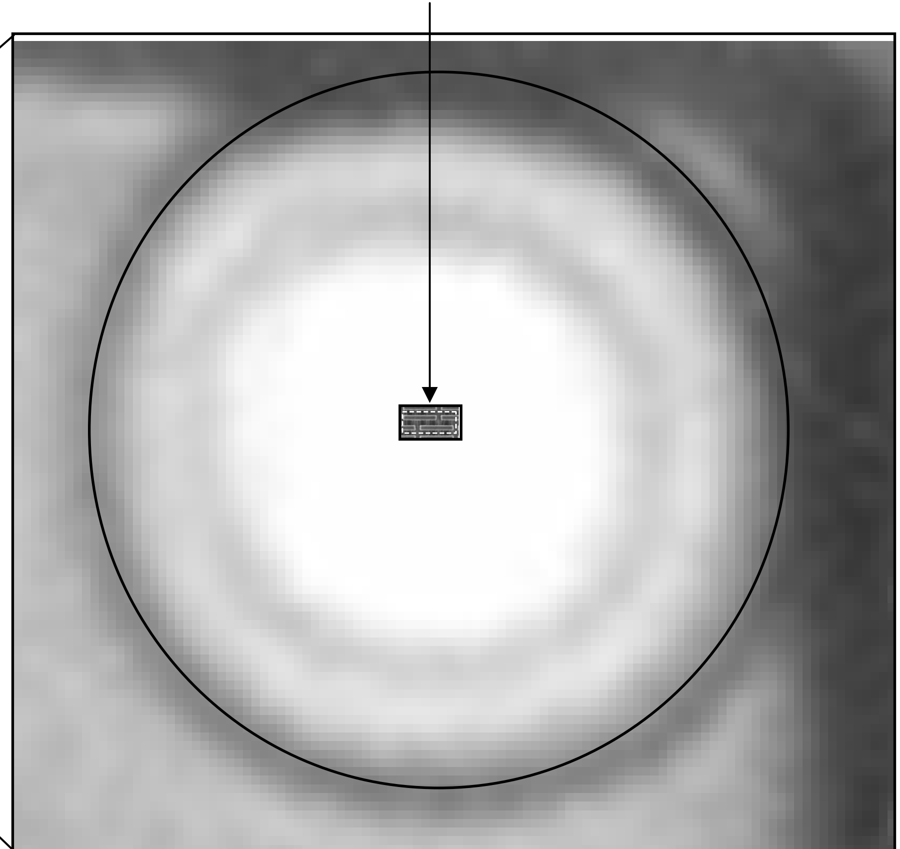
How small is a 32nm memory cell?

1980 SRAM Cell: 1700 um^2

32nm SRAM Cell: 0.171 um^2



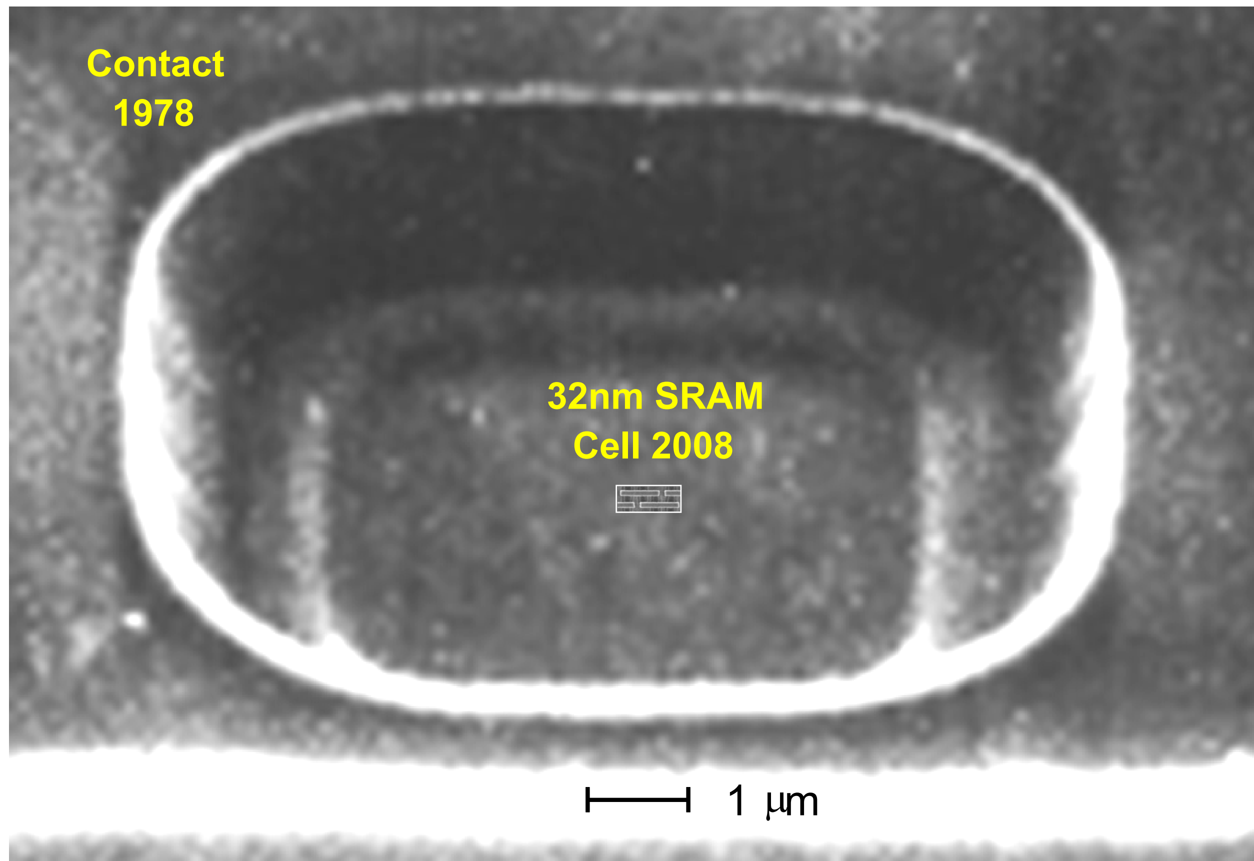
10000X



Small enough that a 2008 32nm SRAM cell is dwarfed by a 1980 SRAM cell CONTACT

M. Bohr 2007

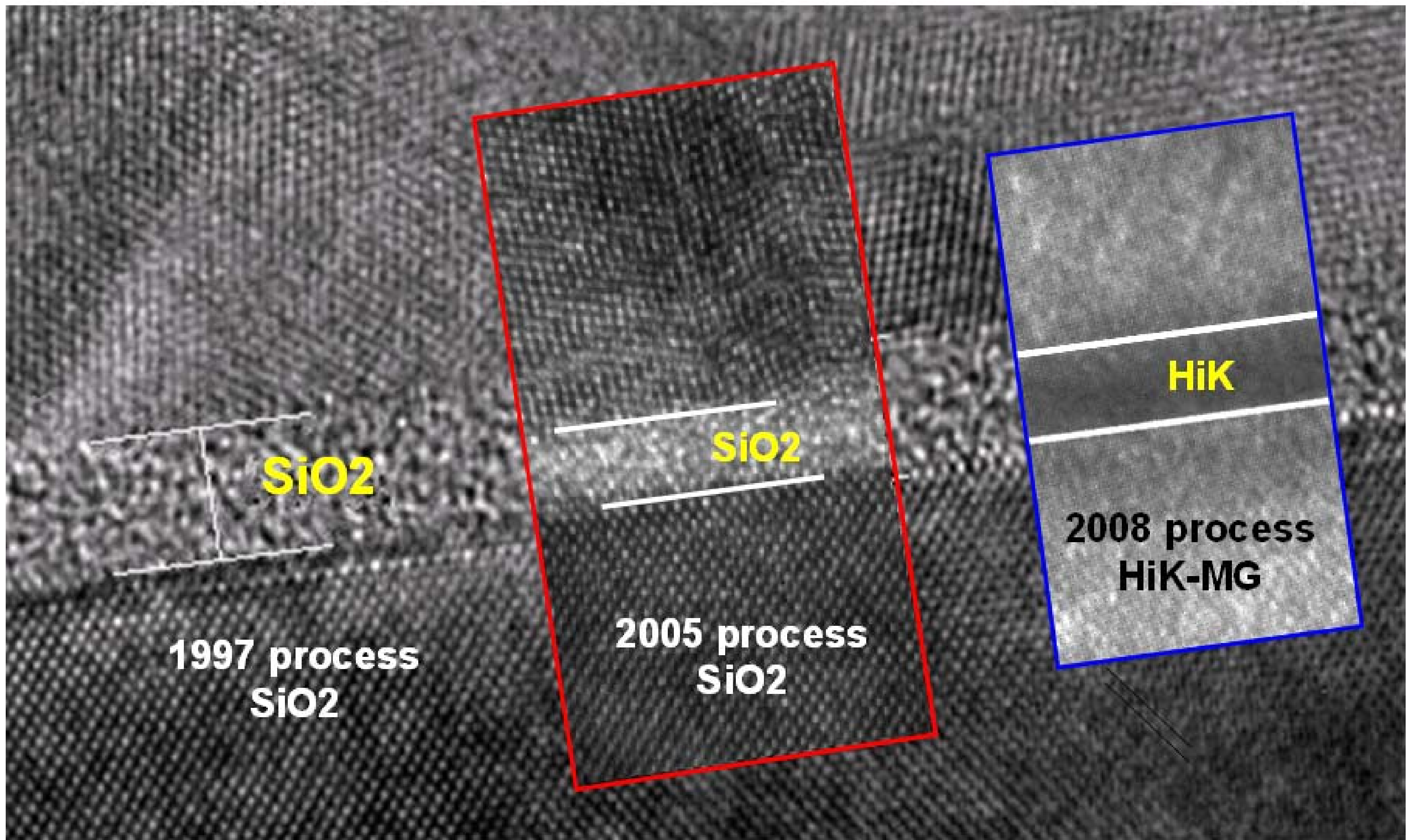
How small is a 32nm memory cell?



Small enough that a 2008 32nm SRAM cell is dwarfed by a 1980 SRAM cell CONTACT

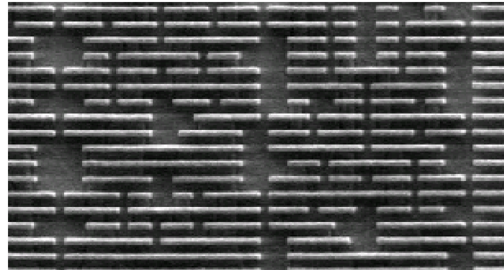
M. Bohr, ISCC, 2009

Atomic dimensions are now routine

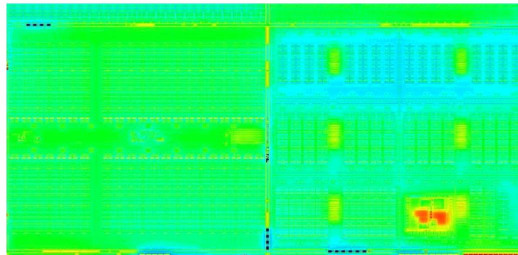


Part I: Physical Variation Sources and Mitigation

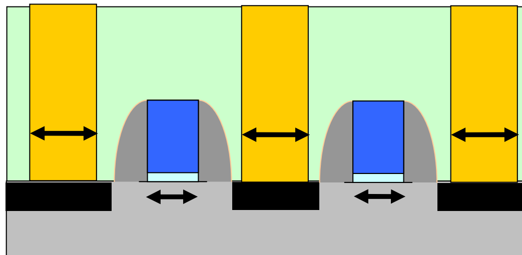
Part I – Physical Variation Sources and Mitigation



Patterning

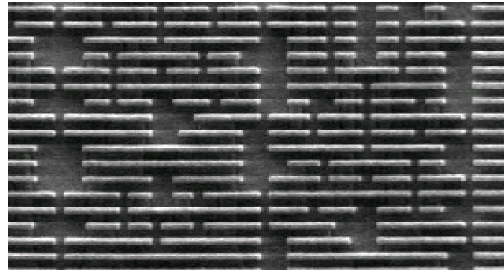


Polish

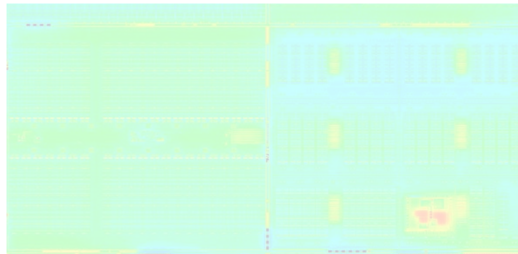


Strain

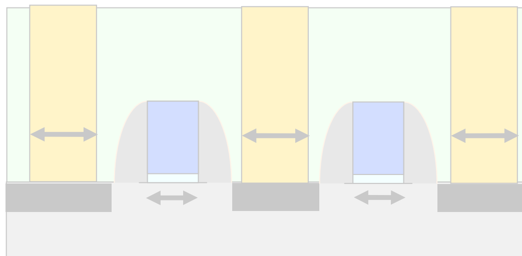
Part I – Physical Variation Sources and Mitigation



Patterning

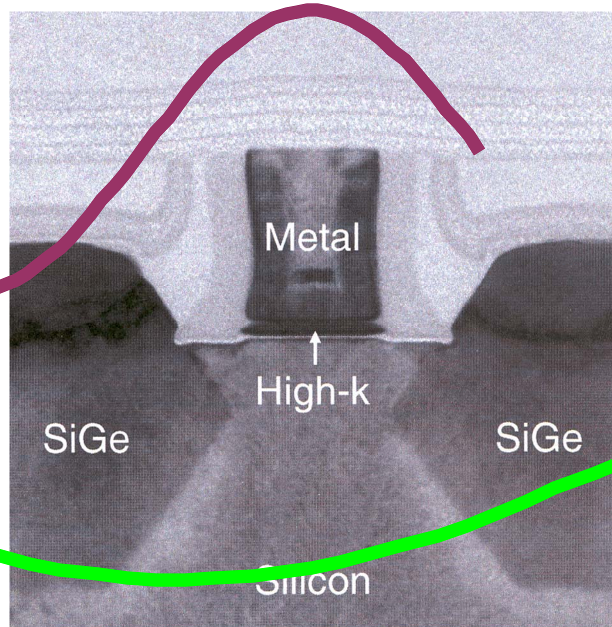


Polish



Strain

How small is a 45nm transistor?



- 5.5X smaller than the 193nm light that prints it
- ~15X smaller than visible green light

K. Kuhn 2007

Design

OPC/RET

Phase mask data

Trim mask data

Reticle manufacturing

Phase mask

Trim mask

Exposure

Etch

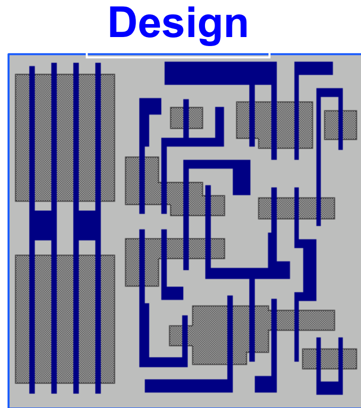
Putting it all together for the gate layer of a 65nm MPU

=

(magnified 25,000X)

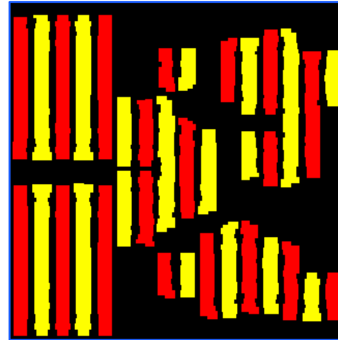
C. Kenyon TOK conf. Dec. 2008

Design



OPC/RET

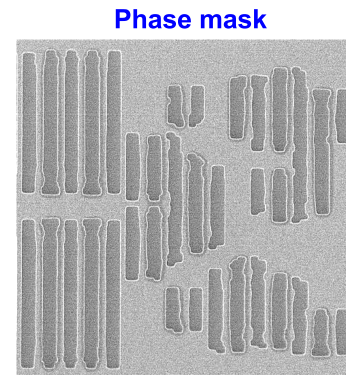
Phase mask data



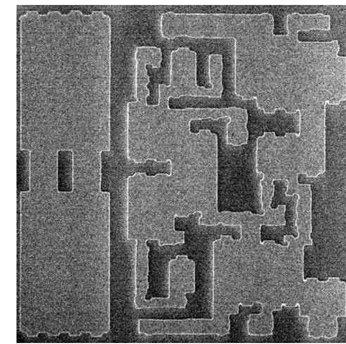
Trim mask data

Reticle manufacturing

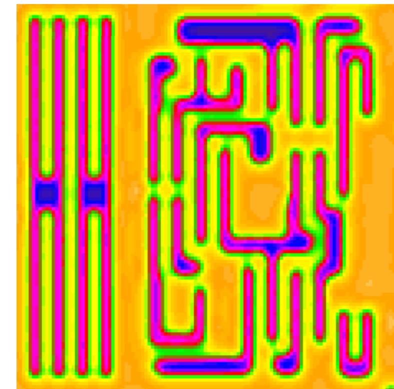
Phase mask



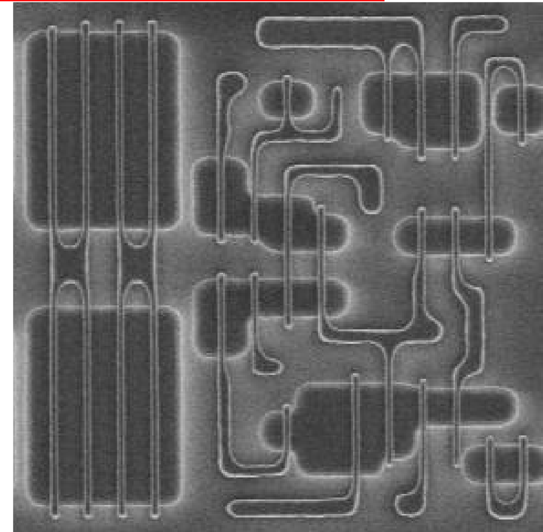
Trim mask



Exposure



Etch



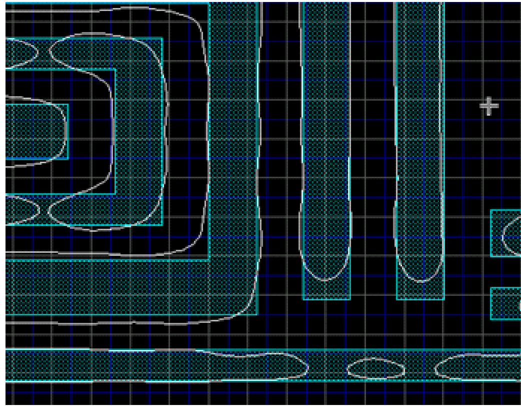
Putting it all together for the gate layer of a 65nm MPU

=

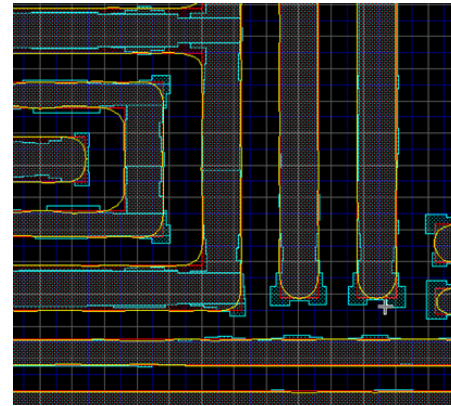
(magnified 25,000X)

C. Kenyon
TOK conf.
Dec. 2008

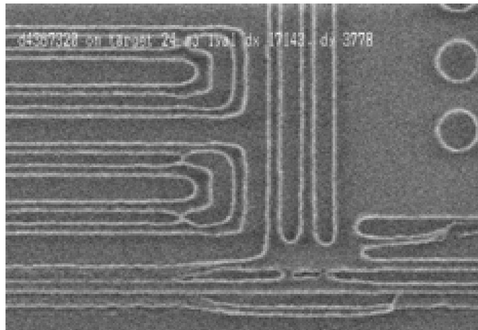
Optical Proximity Correction (OPC) As a Resolution Enhancement Technique



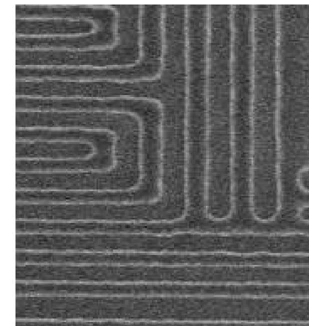
Contour prediction – no OPC



Contour prediction – with OPC



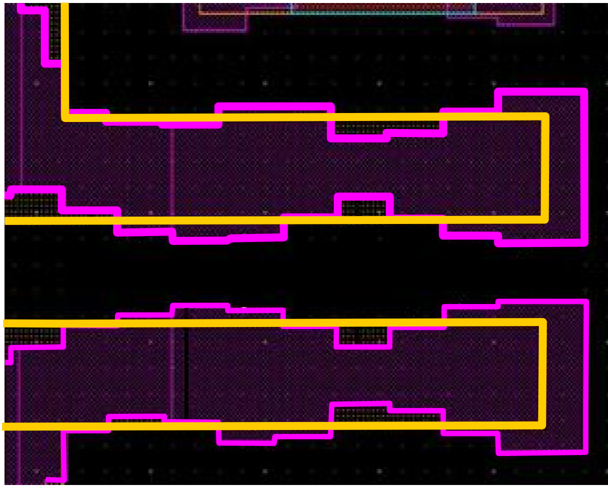
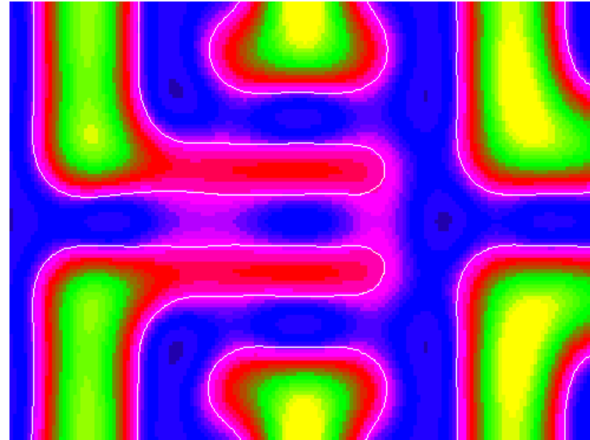
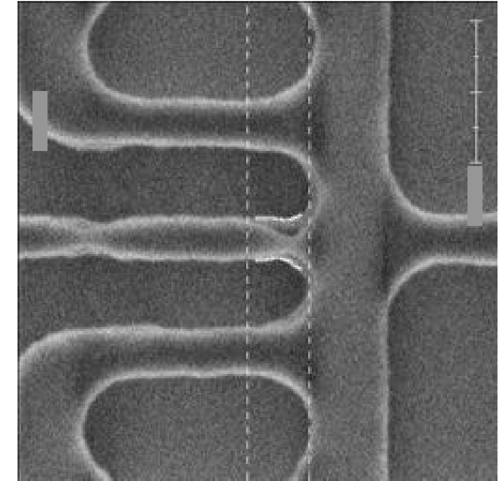
SEM Image – no OPC



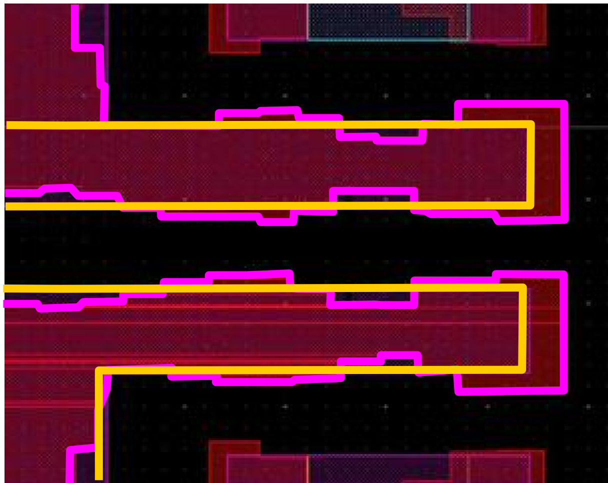
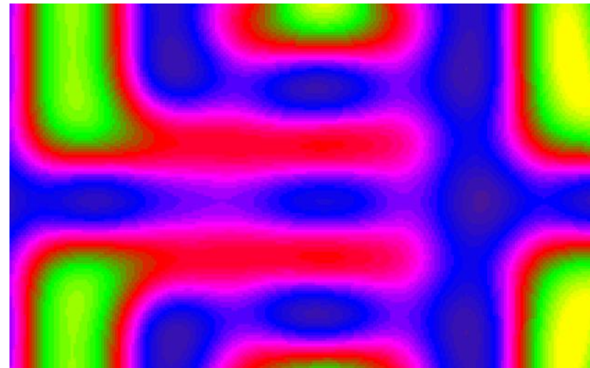
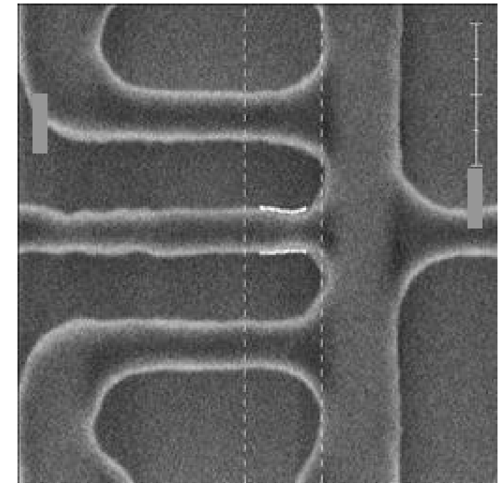
SEM Image – with OPC

K. Wells-Kilpatrick: 2007

45nm: OPC as a Variation Management Technique

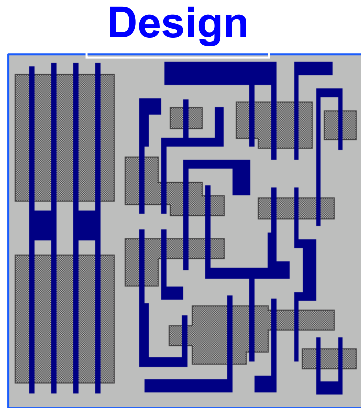




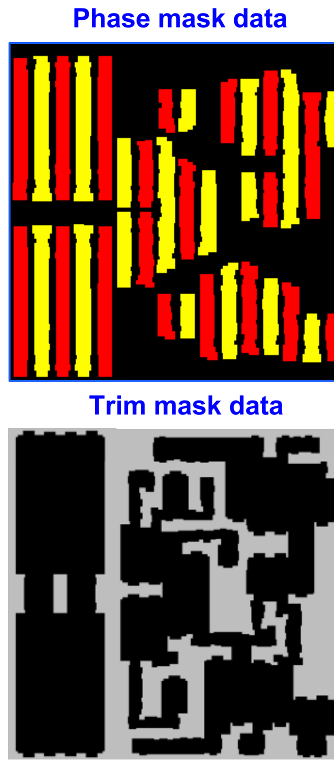
Top-down resist CD meets spec, but poor contrast leads to poor resist profile which gets transferred to metal pattern after etch, resulting in shorting marginality

Computational lithography solution

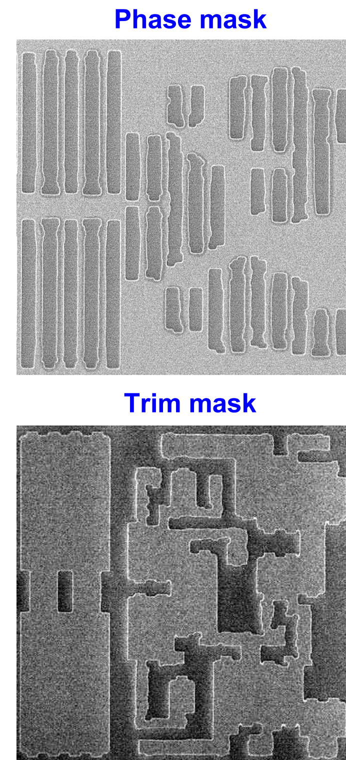
K. Kuhn, IEDM 2007

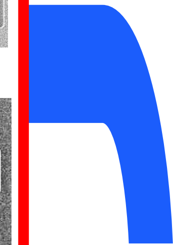




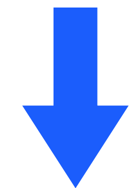
OPC/RET

Reticle manufacturing





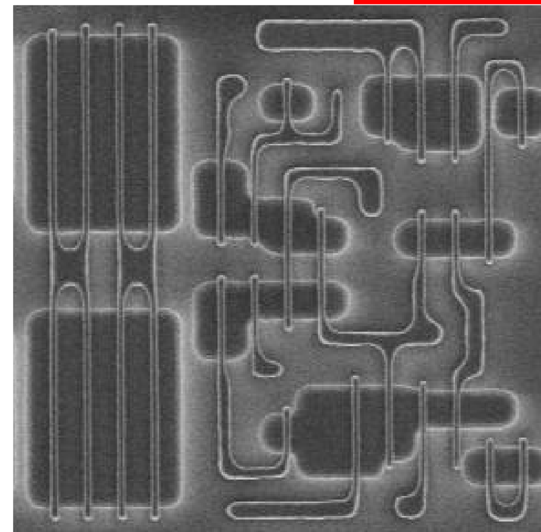
Exposure

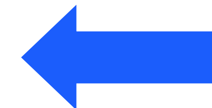


Putting it all together for the gate layer of a 65nm MPU

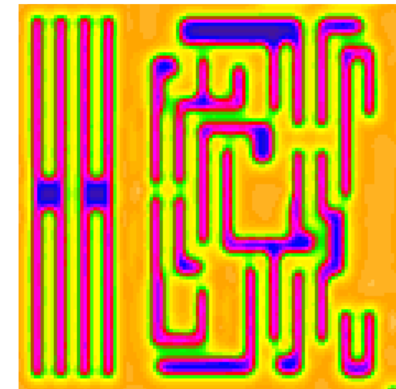
=

(magnified 25,000X)





Etch



C. Kenyon TOK conf. Dec. 2008

MEEF
Mask Error Enhancement Factor

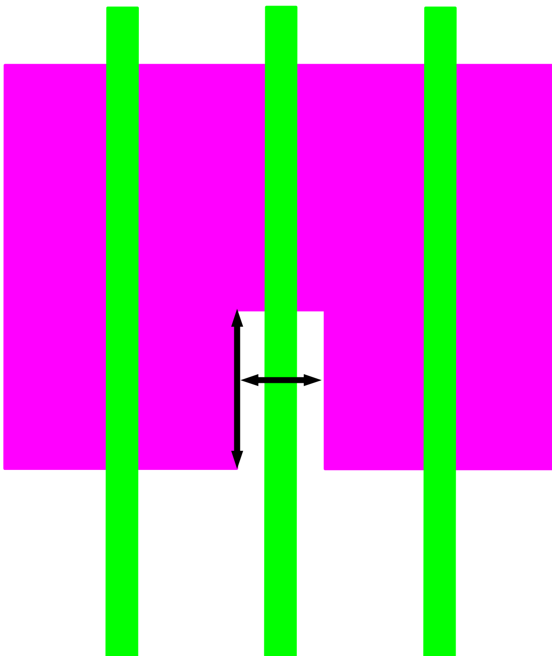
- MEEF is a scaling factor that causes certain layout geometries to exhibit a greater sensitivity to mask dimension tolerances.
- Any dimensional error in the mask is magnified on the wafer by the MEEF value.

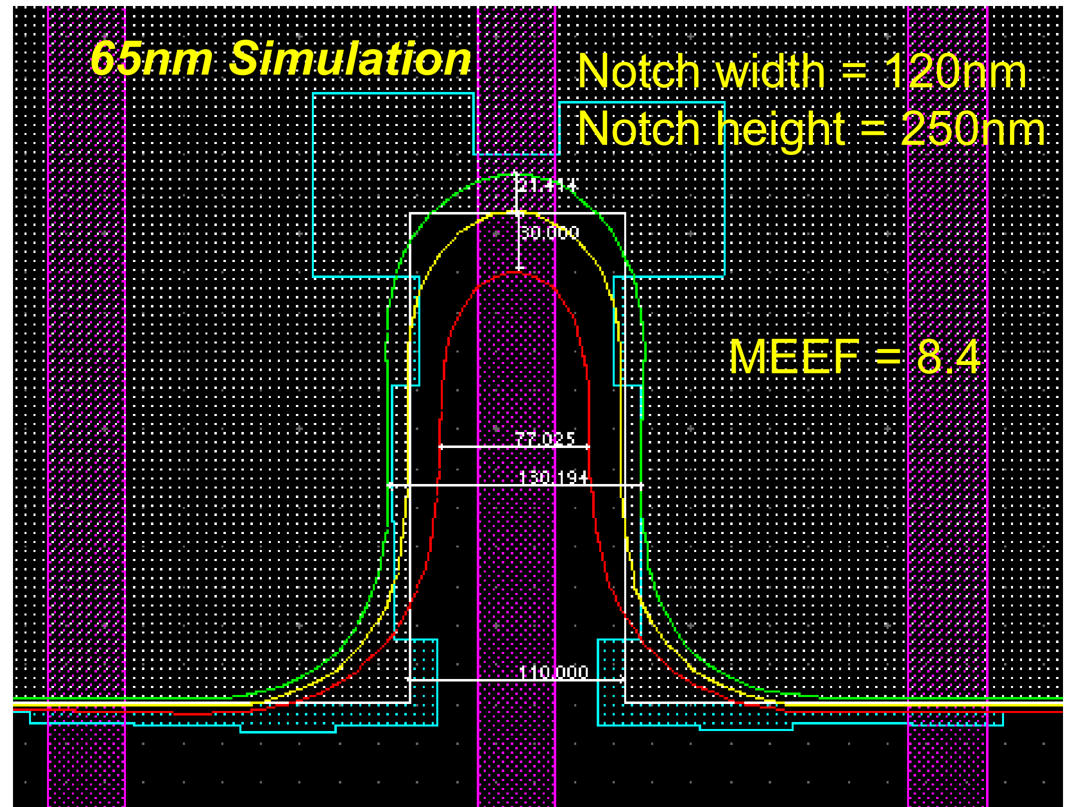
$$\Delta W_{wafer} = MEEF * \Delta W_{mask}$$

- Depending on the value of the mask error and the lithography exposure/focus conditions the final printed pattern can be either larger or smaller.

MEEF Impact on Ze Error

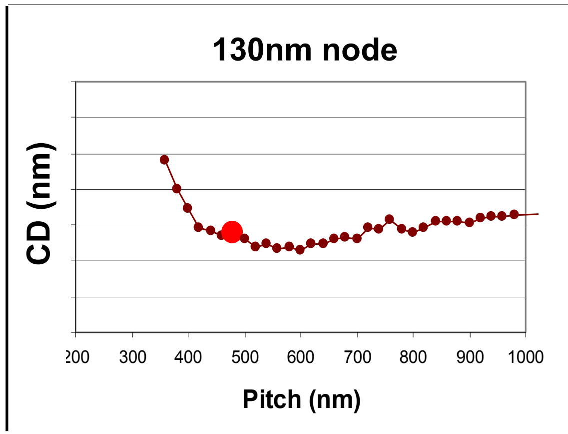
Ze error can be either positive or negative



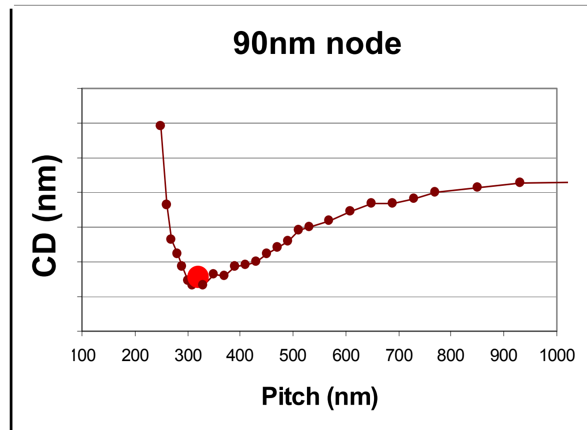


Yellow: DCCD contour after OPC
Green: with -3.375 nm mask making error
Red: with 3.375 nm mask making error

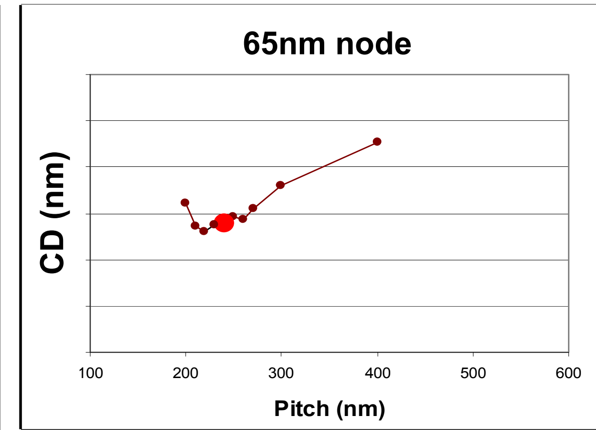
MEEF and Historical gate CD vs. pitch



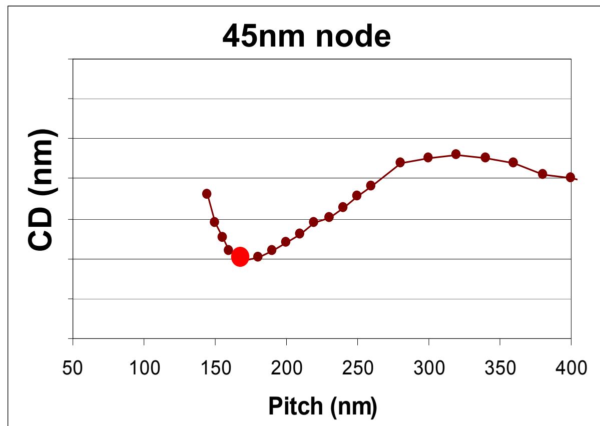
248nm; OAI; model-based OPC



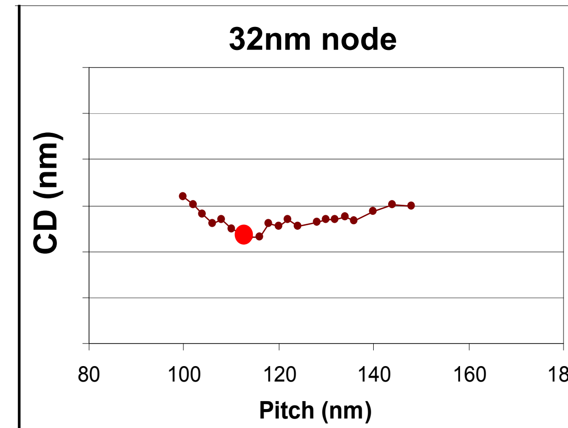
193nm; OAI; model-based OPC



193nm; APSM; model-based OPC



193nm; APSM; model- based OPC double patterning



193nm; immersion; APSM; model- based OPC; double patterning; polarization

● *Contacted gate pitch*

C. Kenyon
TOK conf.
Dec. 2008

Low MEEF requires targeting in the "flat" portion of CD vs. pitch
Process innovations continue this trend in the 32nm node

Design

OPC/RET

Phase mask data

Trim mask data

Reticle manufacturing

Phase mask

Trim mask

Exposure

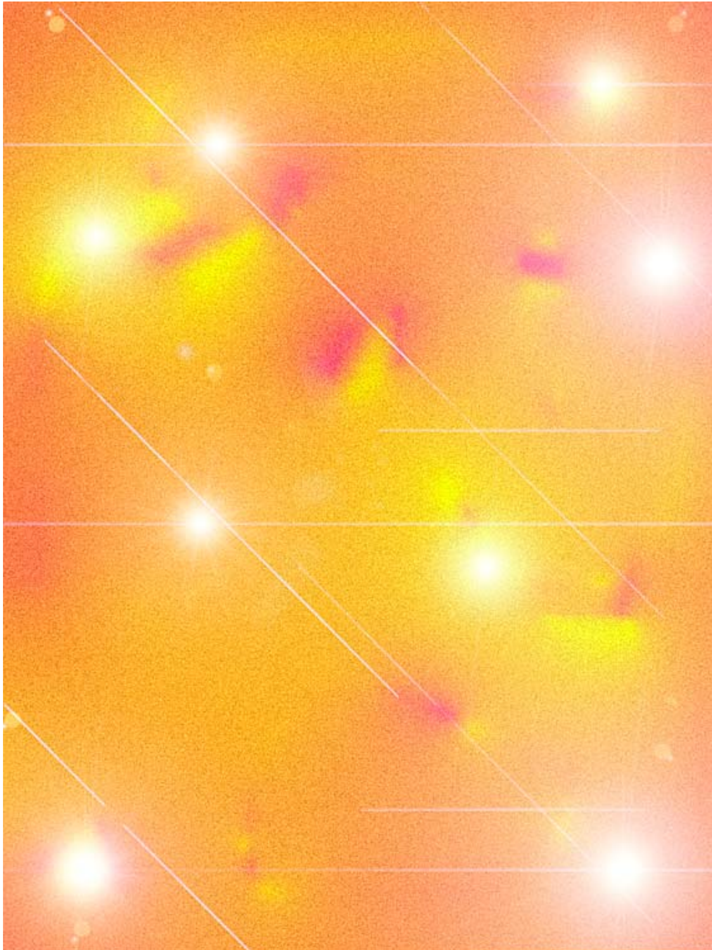
Etch

Putting it all together for the gate layer of a 65nm MPU

(magnified 25,000X)

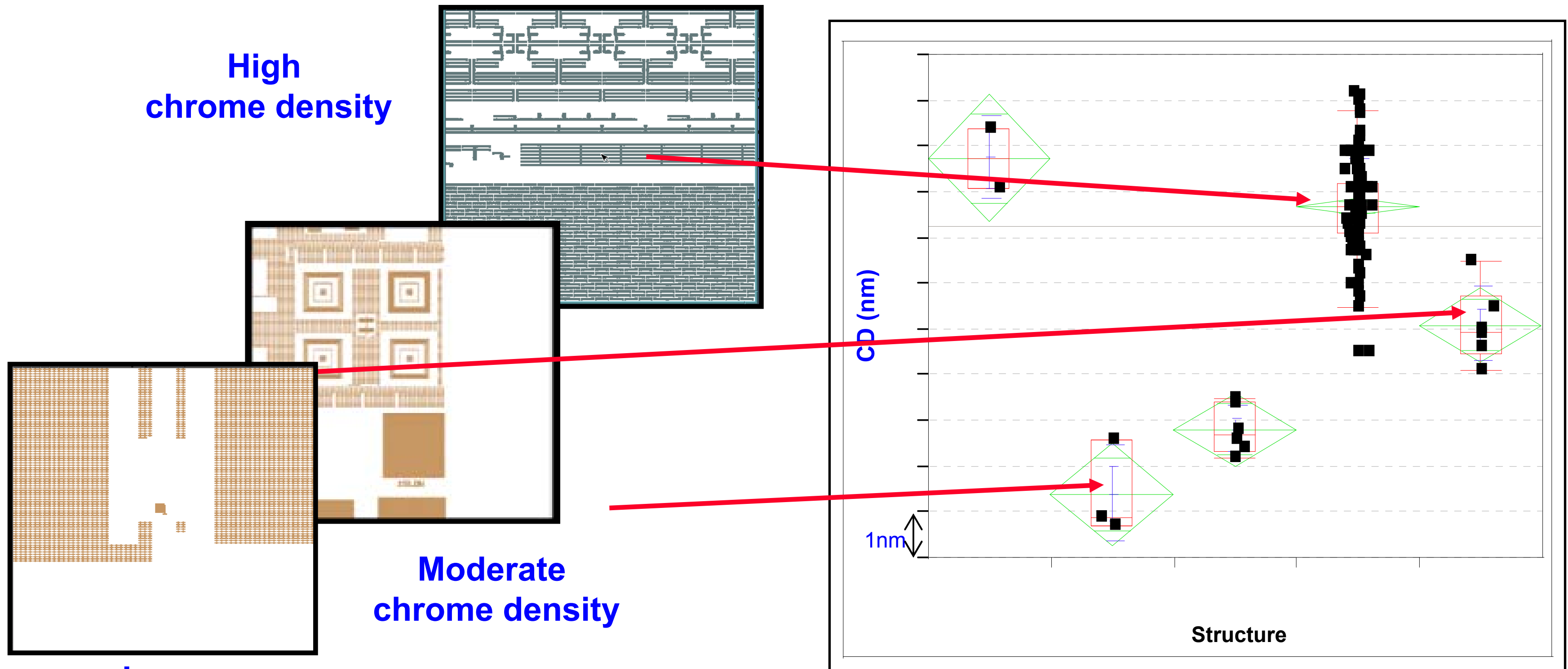
C. Kenyon TOK conf. Dec. 2008

FLARE



- Flare is unwanted scattered light arriving at the wafer
- Flare is caused by interactions that force the light to travel in a "non-ray trace" direction.
- Flare is both a function of local environment around a feature (short range flare) and the total amount of energy going through the lens (long range flare).

Impact of flare on gate CDs

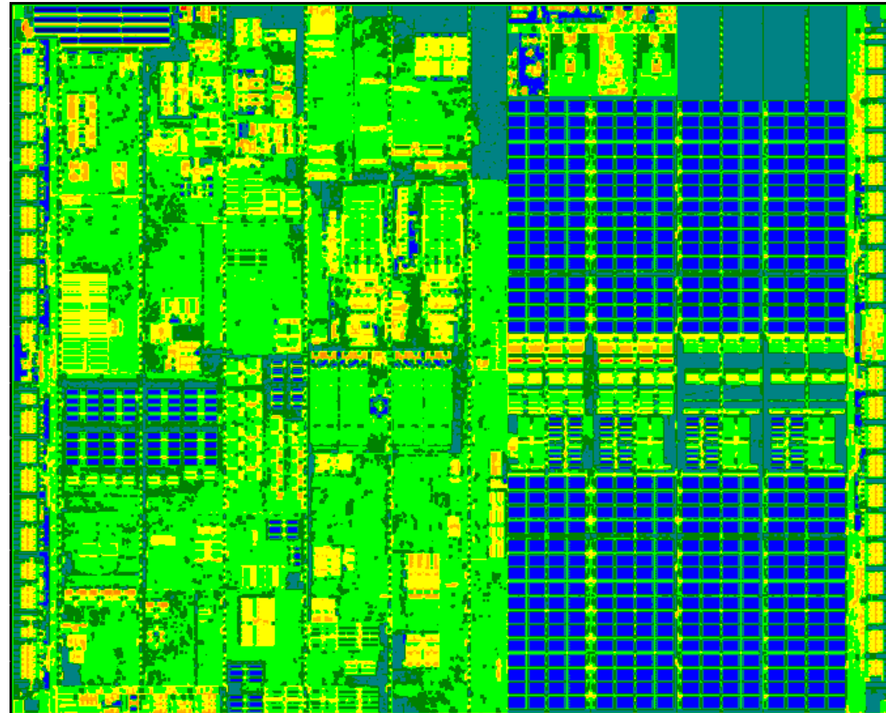


C. Kenyon
TOK conf.
Dec. 2008

- **During 65nm process development, large CD deviations were observed for structures having identical pitch and reticle CD due to flare**
- **Gates only 500μm away from one another could be >5nm different in CD**

Flare Variation Improvement with OPC

Color Code	Chrome Fraction
109:0	56.0 – 63.0
108:0	49.0 - 56.0
107:0	42.0 – 49.0
106:0	35.0 – 42.0
105:0	28.0 – 35.0
104:0	21.0 – 28.0
103:0	14.0 – 21.0
102:0	7.0 – 14.0
101:0	0.0 – 7.0



Development effort produced an algorithm capable of scanning designs and binning regions by local chrome fraction

Binning algorithm is combined with flare-calibrated OPC model

C. Kenyon, TOK conf., Dec. 2008

Design

OPC/RET

Phase mask data

Trim mask data

Reticle manufacturing

Phase mask

Trim mask

Exposure

Etch

Putting it all together for the gate layer of a 65nm MPU

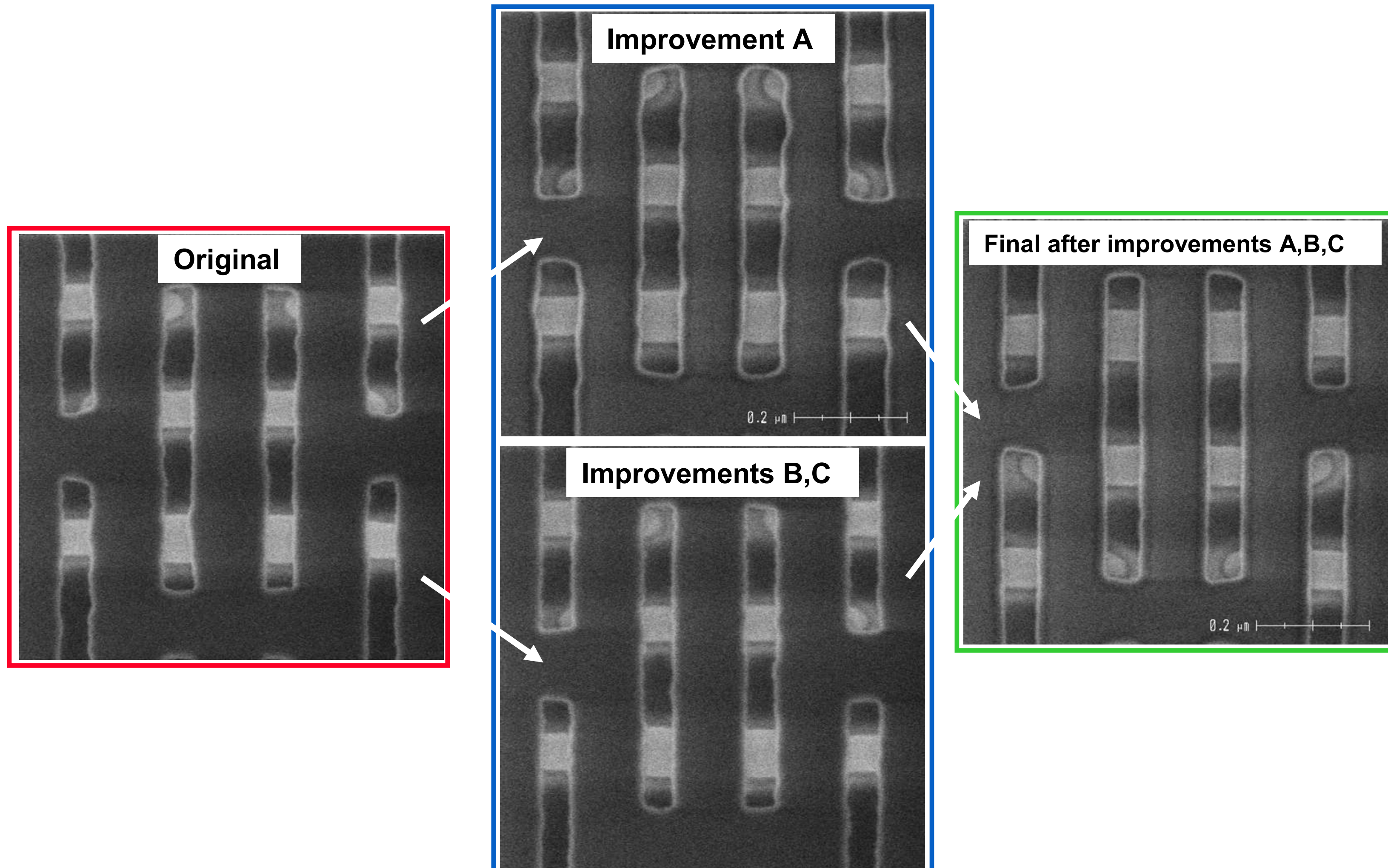
intel Centrino Duo

=

(magnified 25,000X)

C. Kenyon TOK conf. Dec. 2008

45nm highlights role of lithography/etch in resolving LER/LWR



K. Kuhn, ITJ, 2008

Technology Trend
Systematic Gate CD Lithography Variation

LOG (Variation normalized to 130nm)

10

1

0.1

Gate CD variation improvements
with technology scaling

0.7X

WID-total

TOTAL

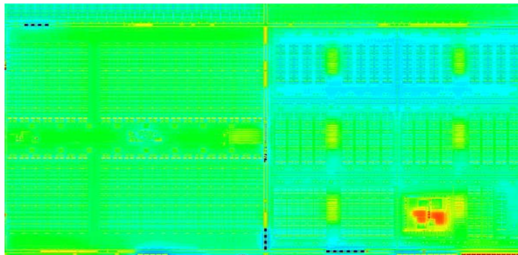
WIW-total

130nm	90nm	65nm	45nm	32nm	GENERATION
336	260	220	160	112.5	GATE PITCH

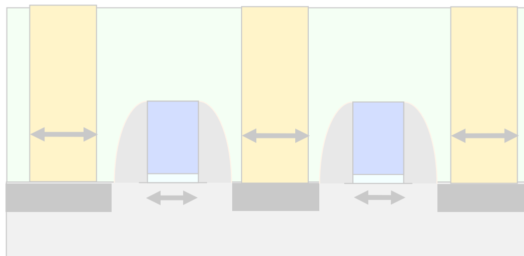
Critical to management of variation is the ability to deliver a 0.7X gate CD variation improvement in each generation enabled by continuous process technology improvements

Part I – Physical Variation Sources and Mitigation

Patterning

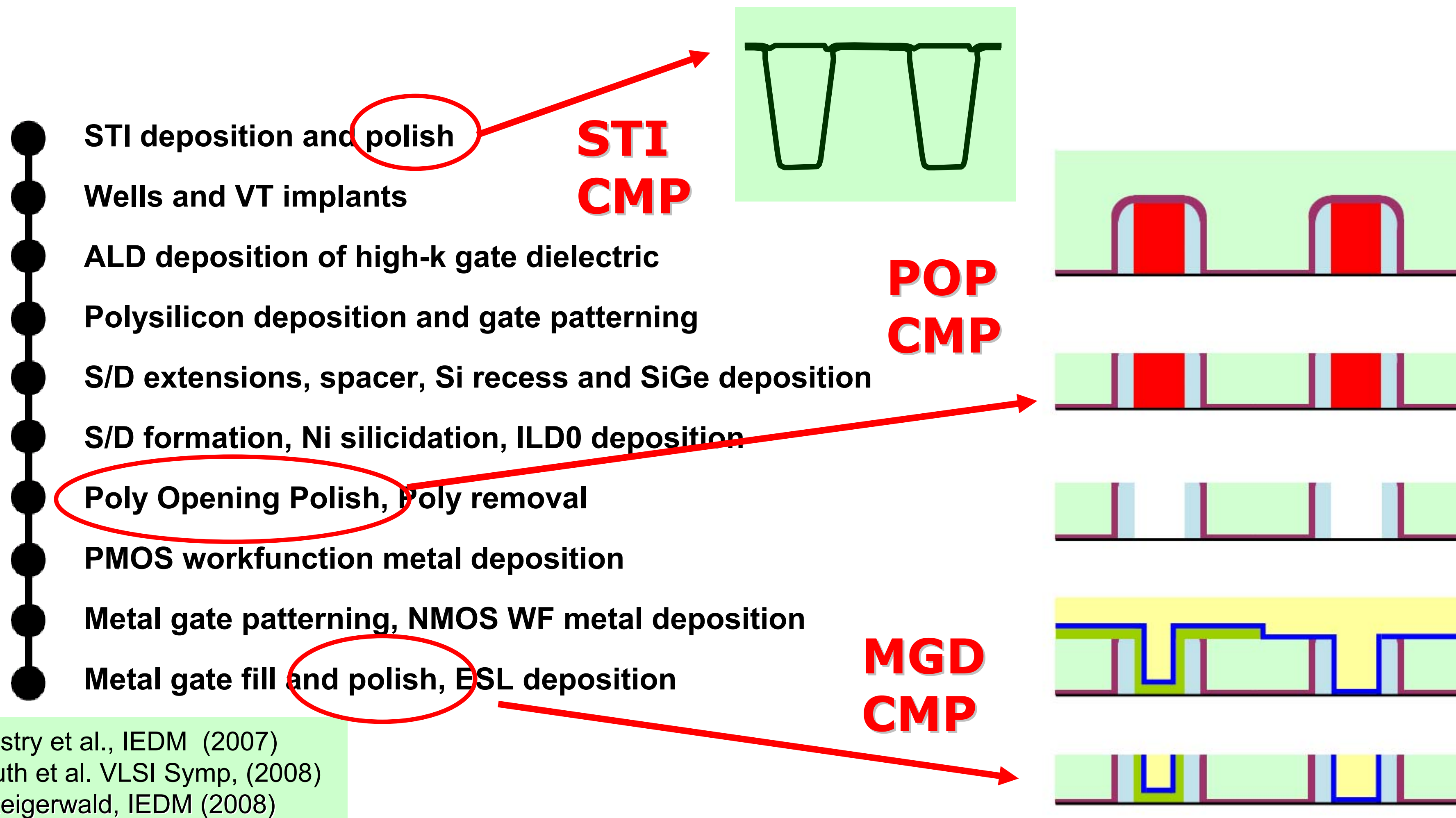


Polish



Strain

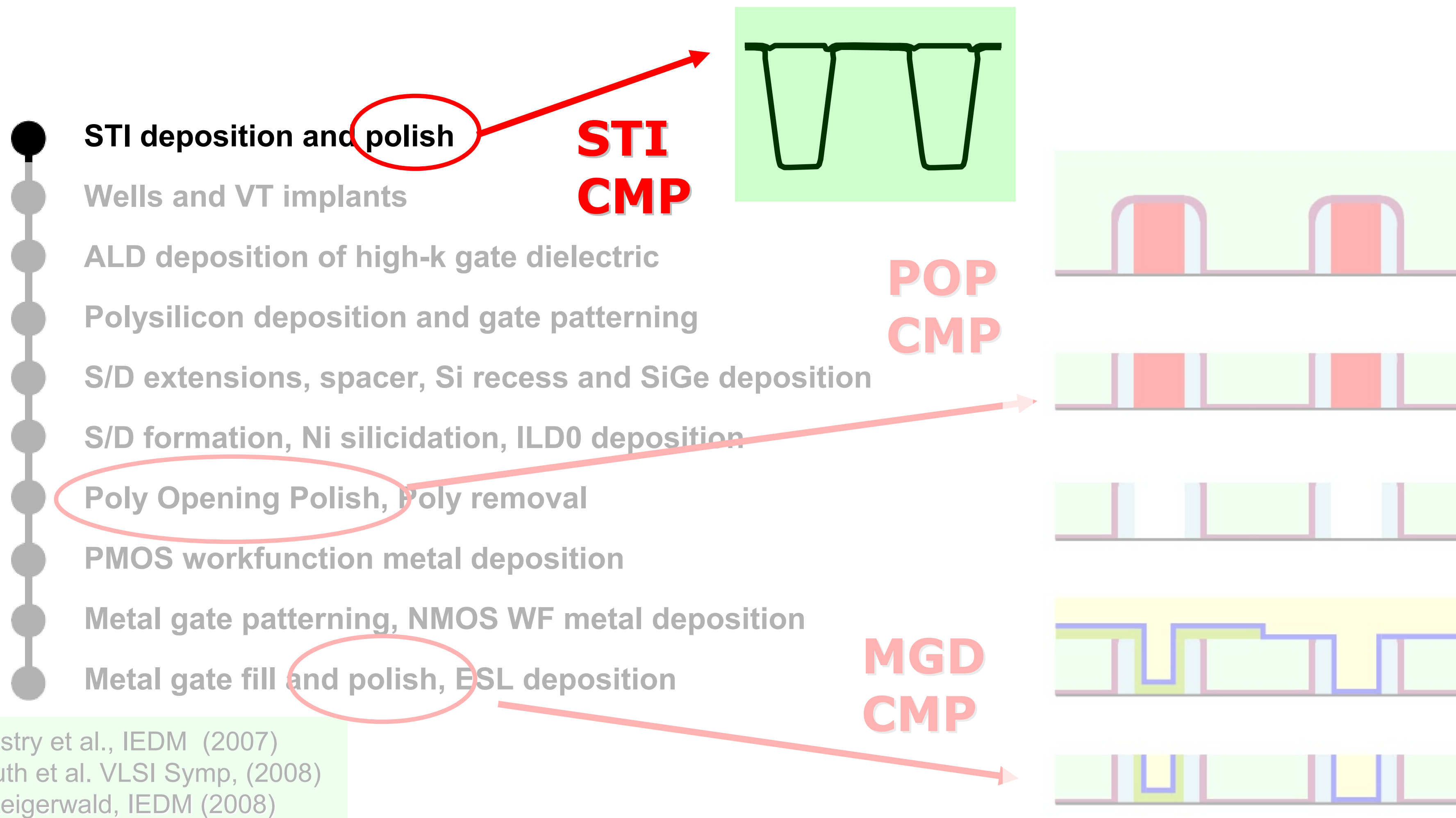
CMP Integration at 45 nm – HiK Metal Gate



K.Mistry et al., IEDM (2007)
C.Auth et al. VLSI Symp, (2008)
J. Steigerwald, IEDM (2008)

First Generation HiK – Replacement Metal Gate
Three critical CMP operations in the FE

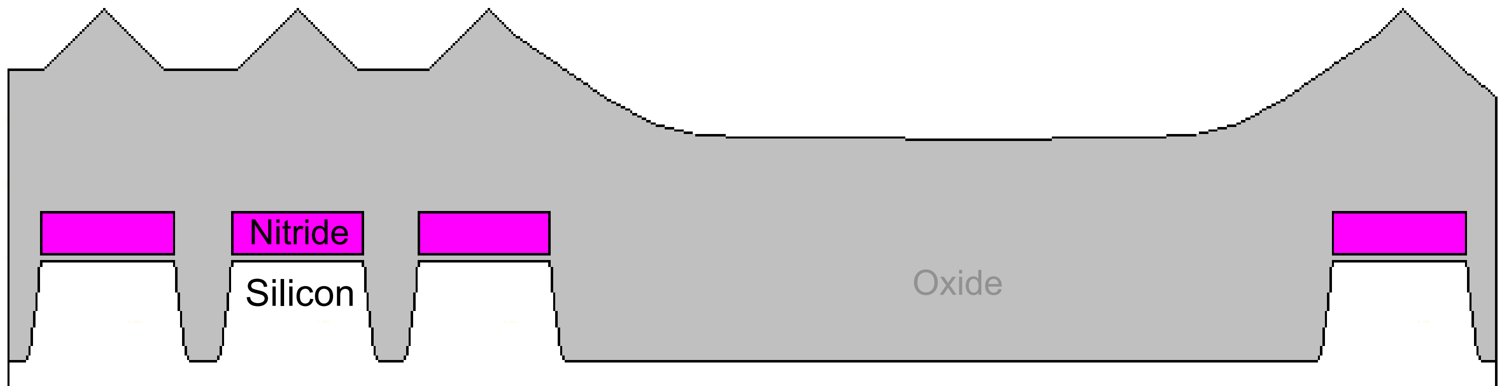
CMP Integration at 45 nm – HiK Metal Gate

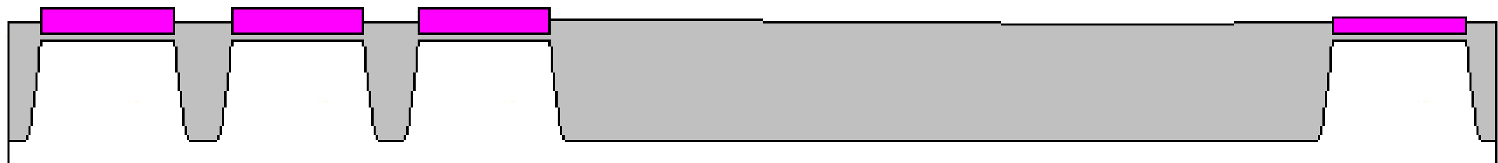


K.Mistry et al., IEDM (2007)
C.Auth et al. VLSI Symp, (2008)
J. Steigerwald, IEDM (2008)

First Generation HiK – Replacement Metal Gate
Three critical CMP operations in the FE

STR Pattern Density Variation Impact

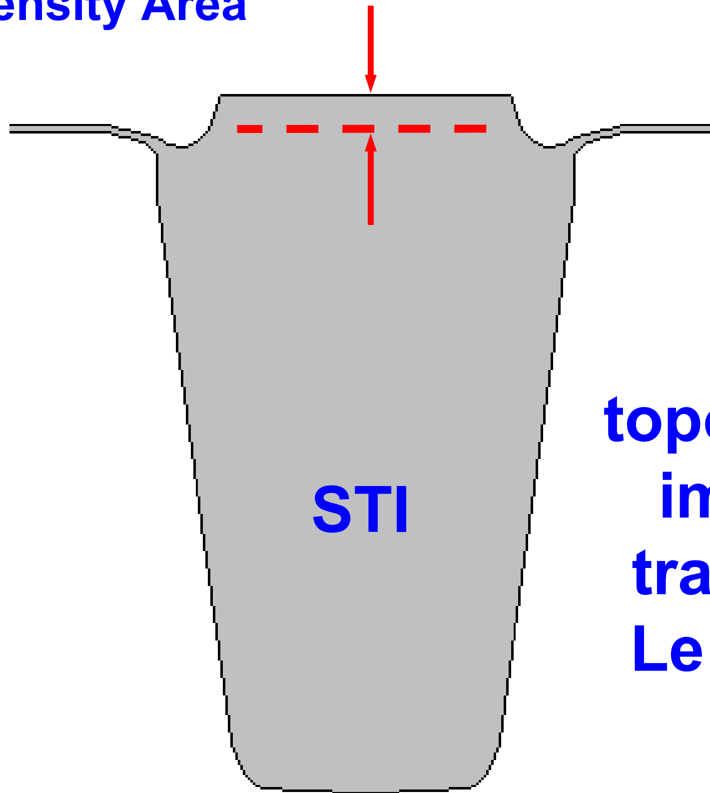




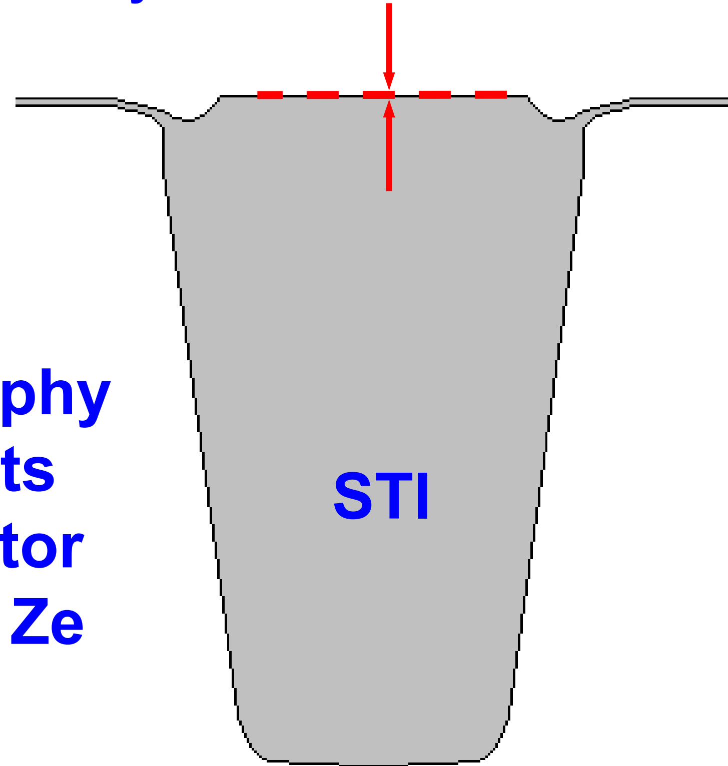
STI Step Height Variation

High Pattern Density Area

Low Pattern Density Area



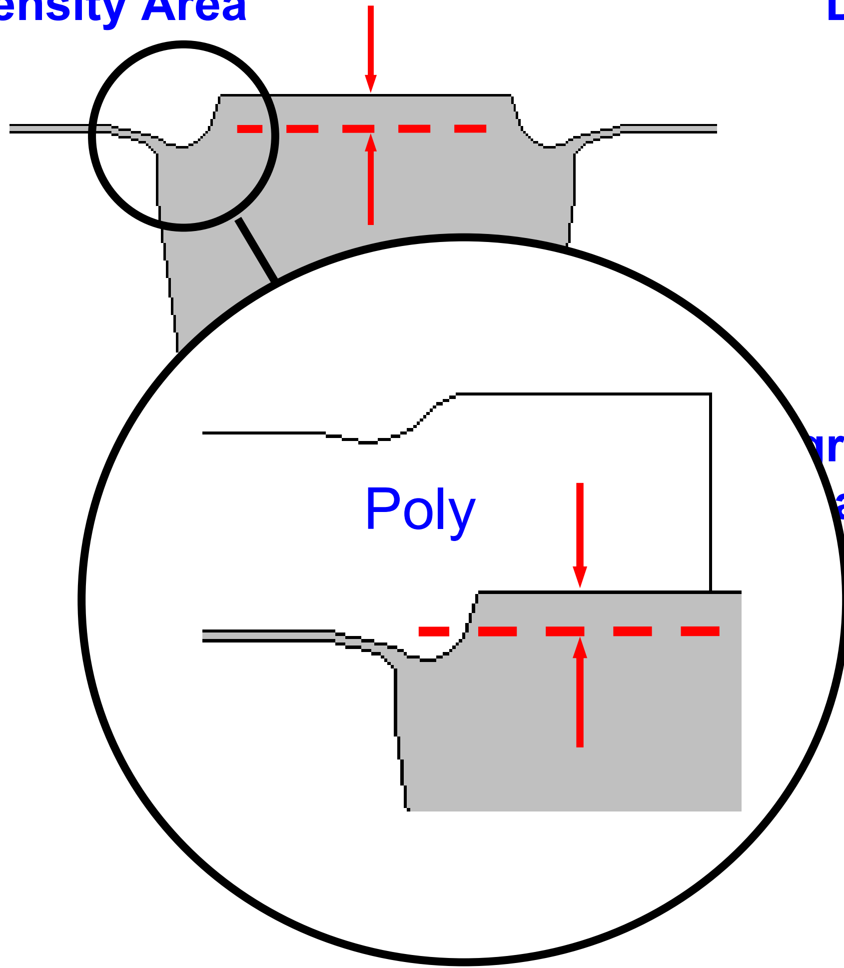
STI topography impacts transistor Le and Ze

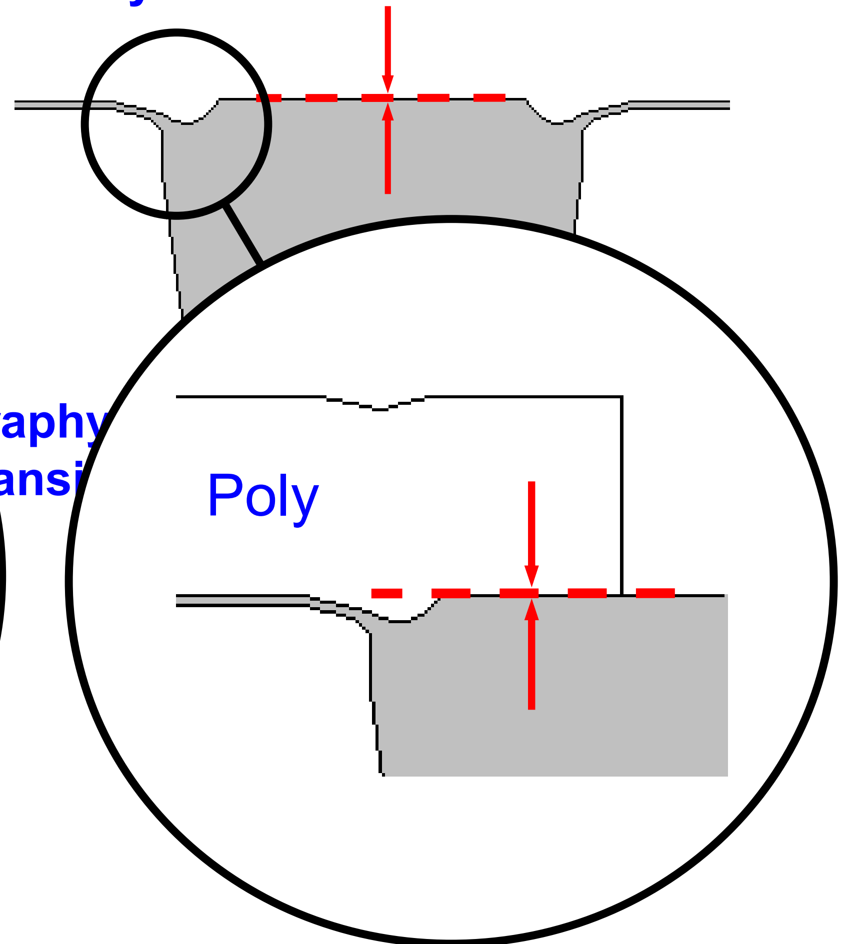


Positive Step Height

Zero Step Height

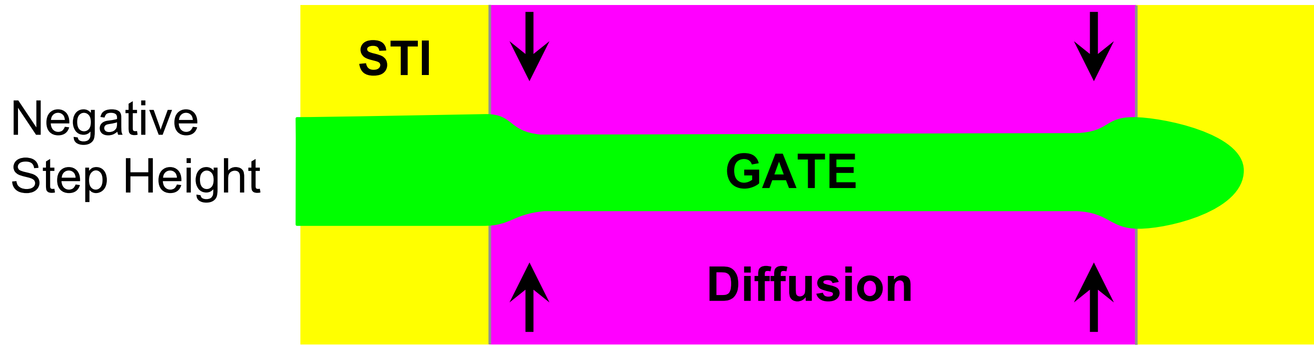
STI Step Height Variation





STI Step Height Impact on Gate CD

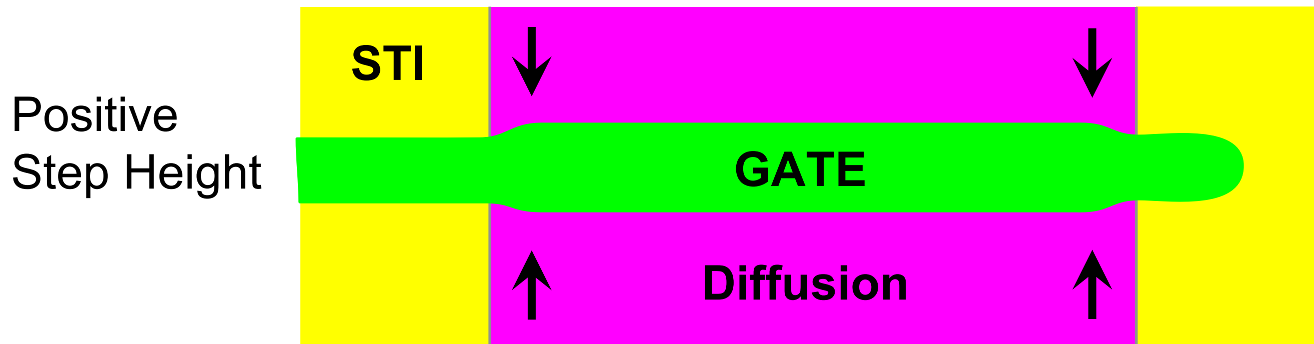
Negative Step Height



"Dogbone"

Lg is longer at the diffusion boundary

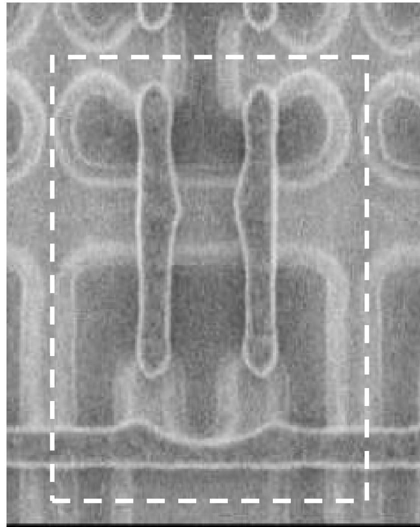
Positive Step Height



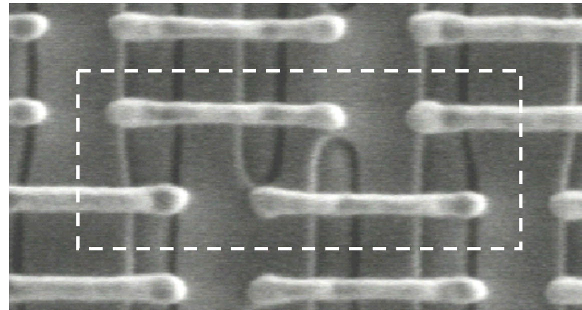
"Icicle"

Gate CD is shorter at the diffusion boundary

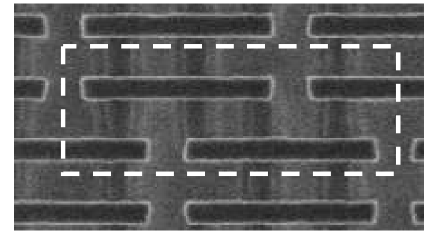
SRAM Density Scaling



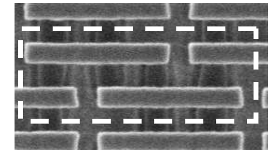
90nm – TALL
1.0 μm²



65nm – WIDE - 0.57 μm²



45nm – WIDE
0.346 μm²

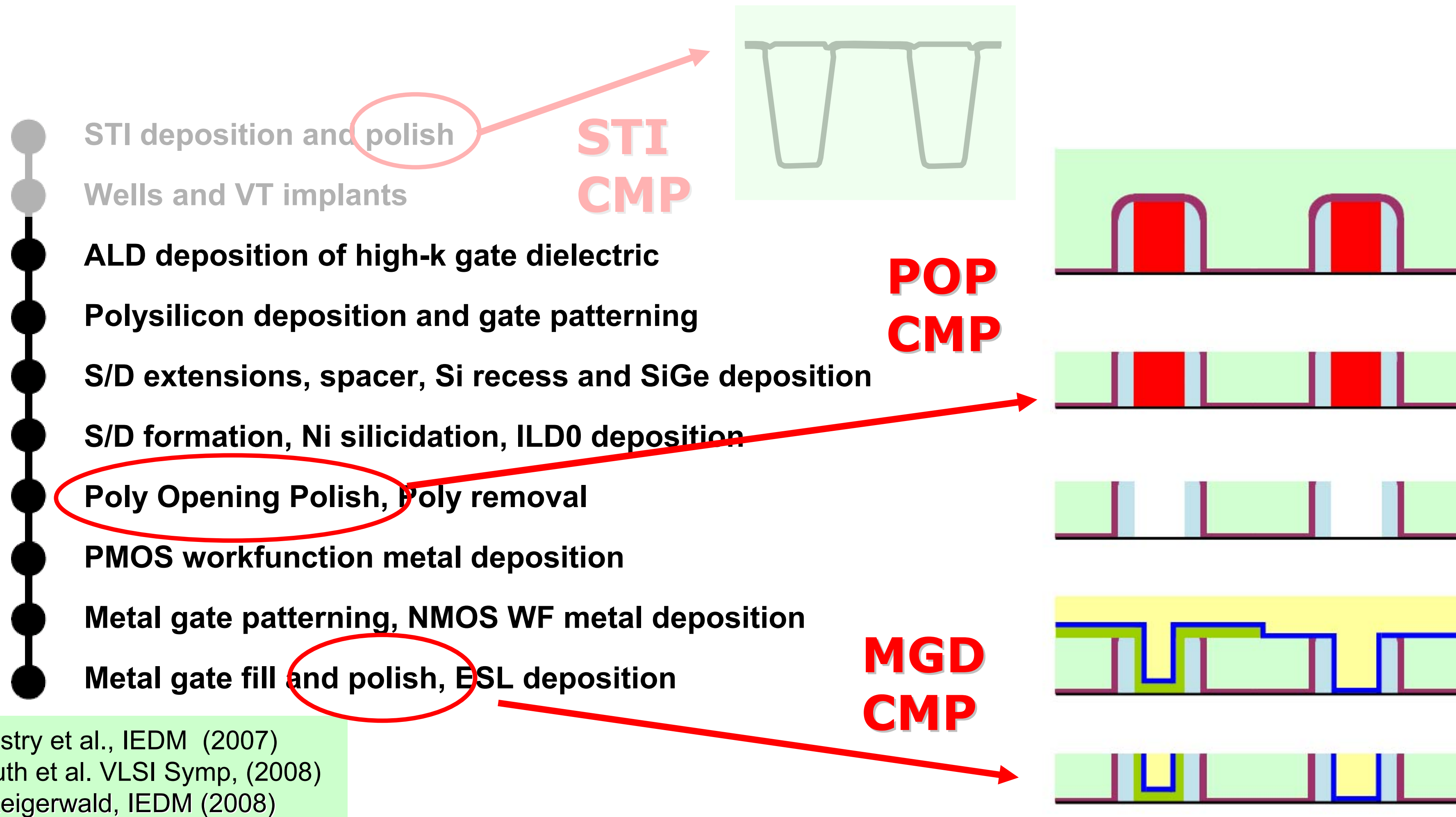


32nm – WIDE
0.171 μm²

65nm to 32nm: Patterning and polish enhancements

- Improved CD uniformity across STI boundaries
- Square corners (eliminate "dogbone" and "icicle" corners)

CMP Integration at 45 nm – HiK Metal Gate

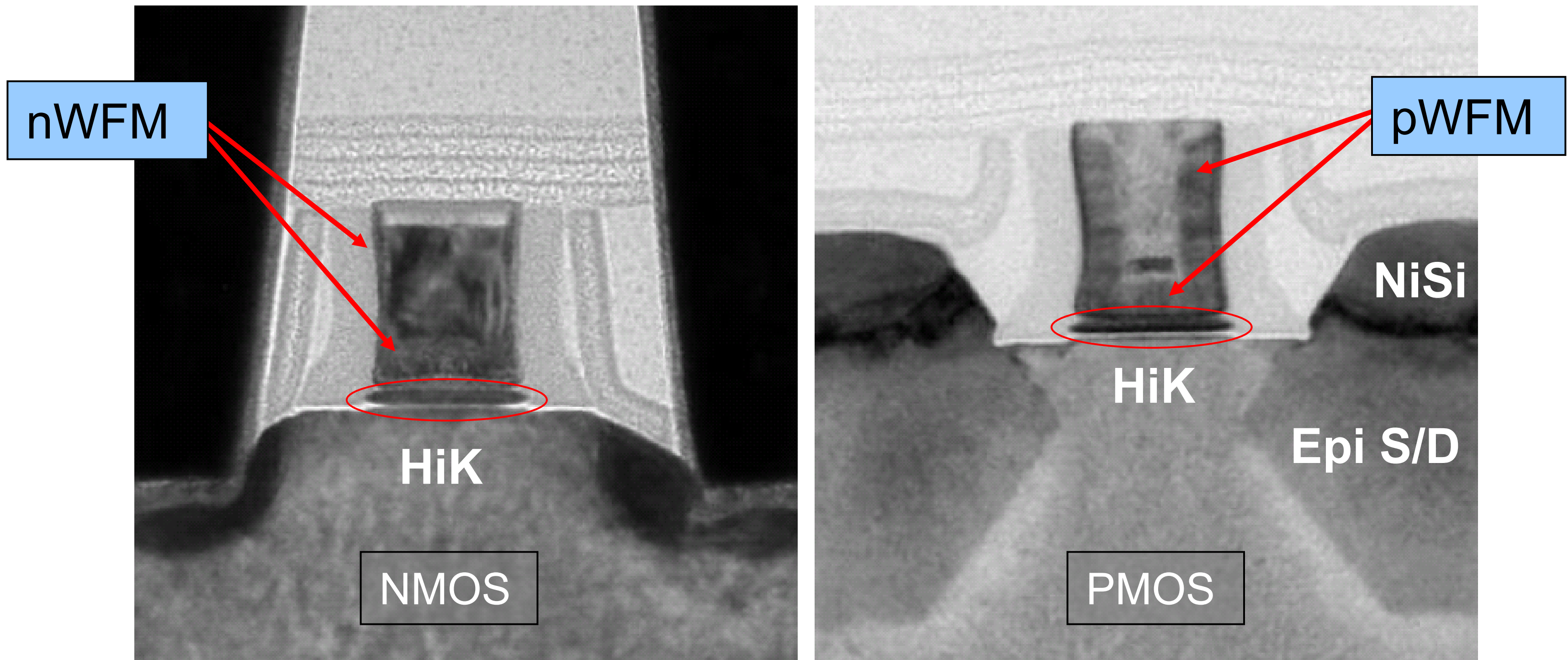


K.Mistry et al., IEDM (2007)
C.Auth et al. VLSI Symp, (2008)
J. Steigerwald, IEDM (2008)

First Generation HiK – Replacement Metal Gate
Three critical CMP operations in the FE

Variation Challenges of RMG CMP Steps

- Gate height control critical to reducing variation
- PMOS/NMOS differences complicate CMP



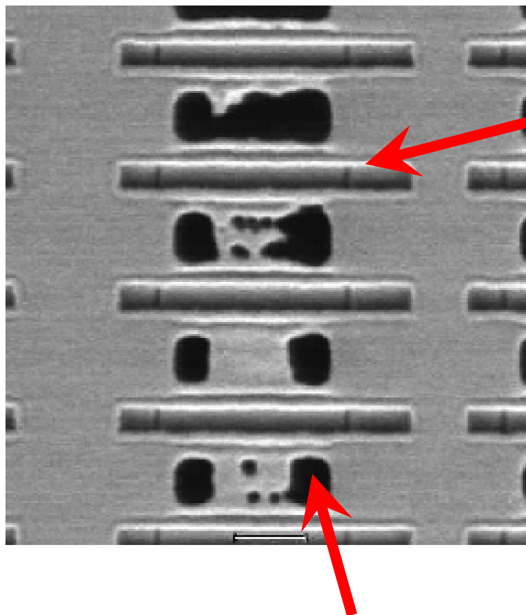
C.Auth et al. VLSI Symp, (2008)

J. Steigerwald, IEDM 2008

Variation Challenges of RMG CMP Steps

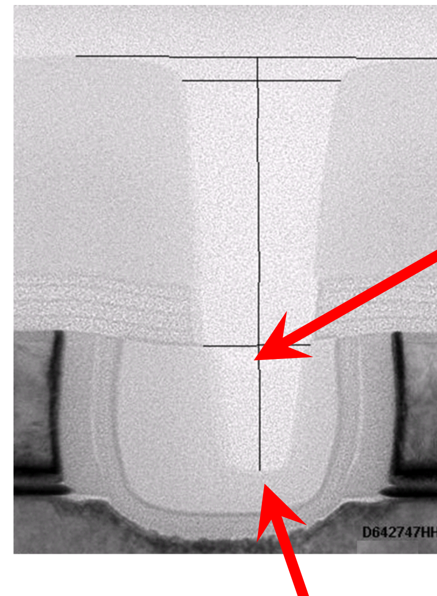
OVERPOLISH
Exposes raised S/D
Rext/mobility impact

UNDERPOLISH
Underetched contact
Rext impact



Gate region

S/D region – attacked during poly etch



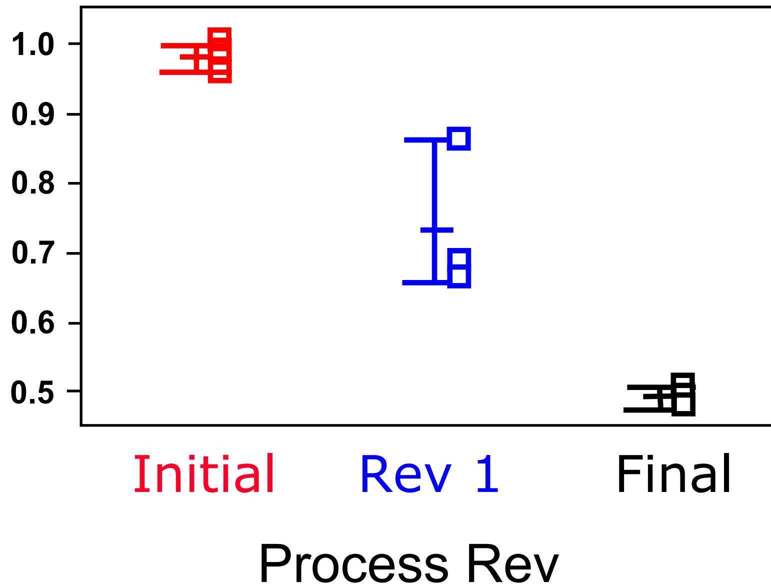
NMOS S/D region contact

S/D region – marginal contact

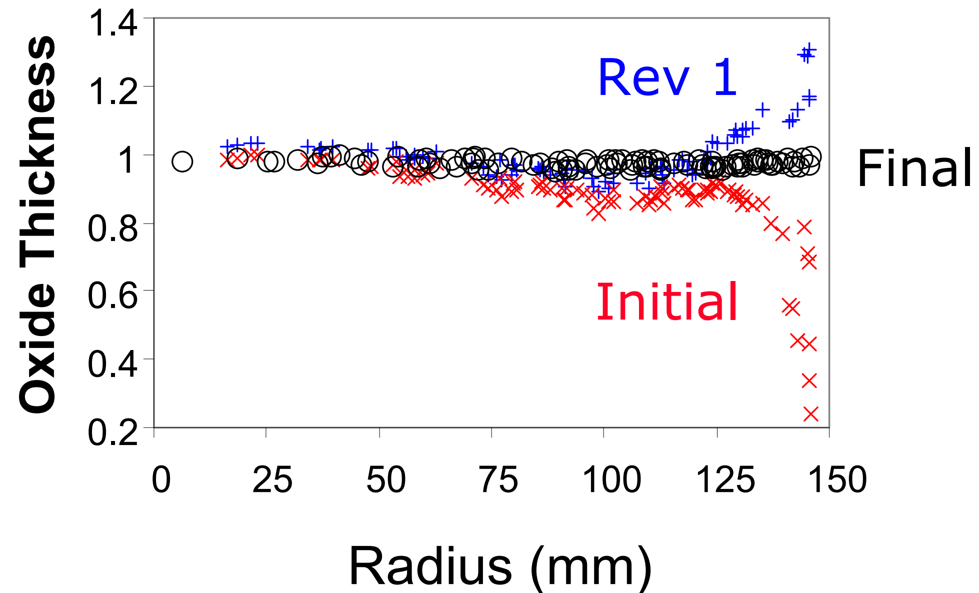
J. Steigerwald, IEDM 2008

Poly Opening Polish (POP)
Thickness Control

POP: WID by process Rev.



Patterned Wafers: WIW Profiles



45nm: with-in die (WID) and with-in wafer (WIW) improvement
High selectivity between films is required.
Key aspect is control of polish rate at edge of wafer.

J. Steigerwald, IEDM 2008

45 nm: POP CMP Improvement
Overscaling Topography Improvement

CMP Topography

1

0.1

0.01

350 250 180 130 90 65 45

Technology node (nm)

0.7X improvement

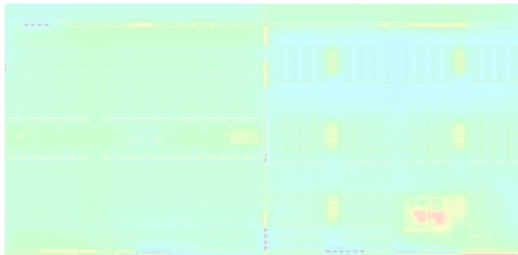
45nm: 2X greater than standard technology scale

J. Steigerwald, IEDM 2008

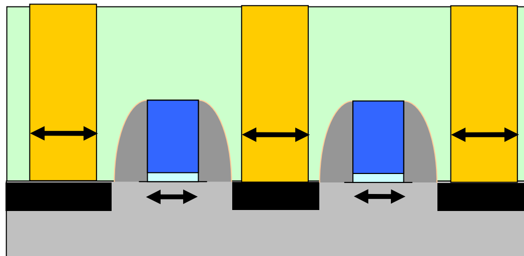
Improvements in polish enabled dramatic improvements in topography variation

Part I – Physical Variation Sources and Mitigation

Patterning

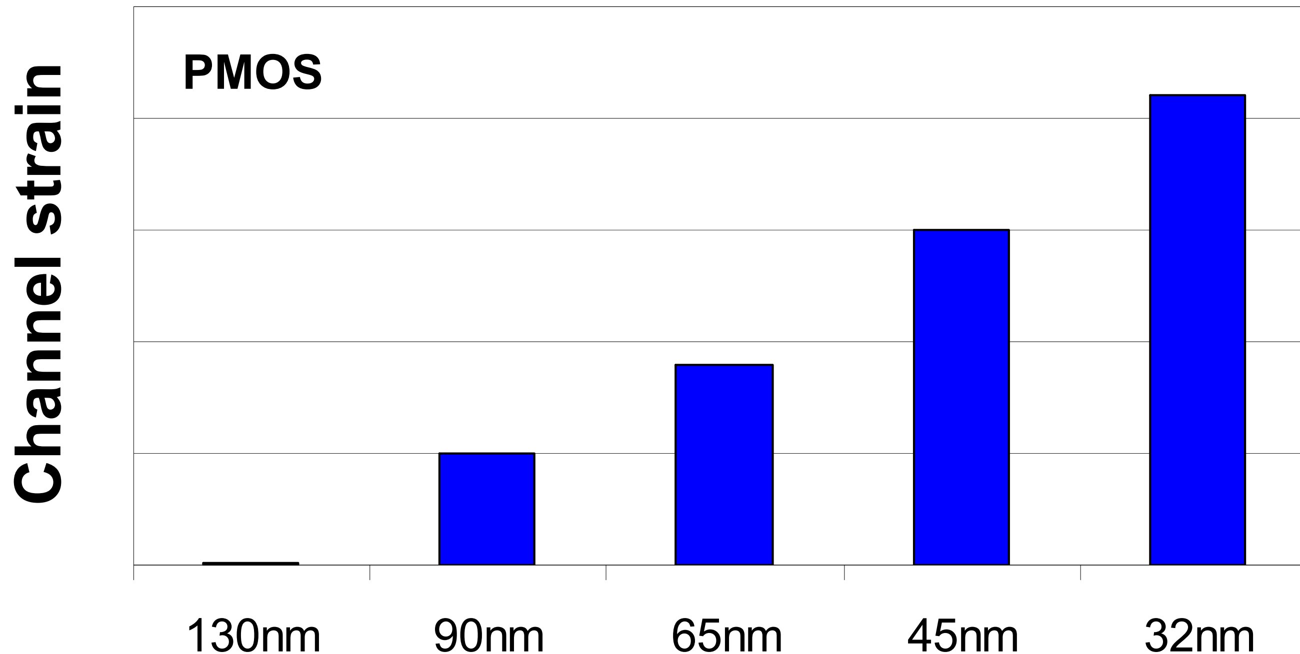


Polish



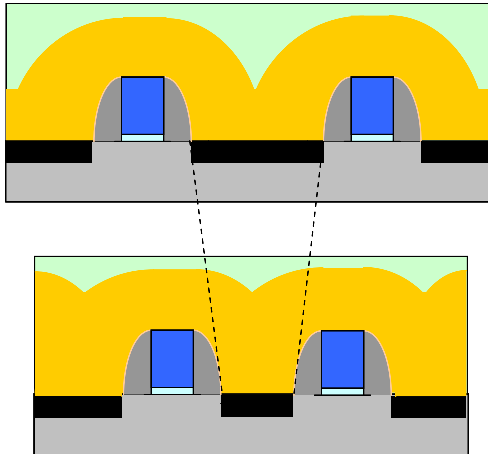
Strain

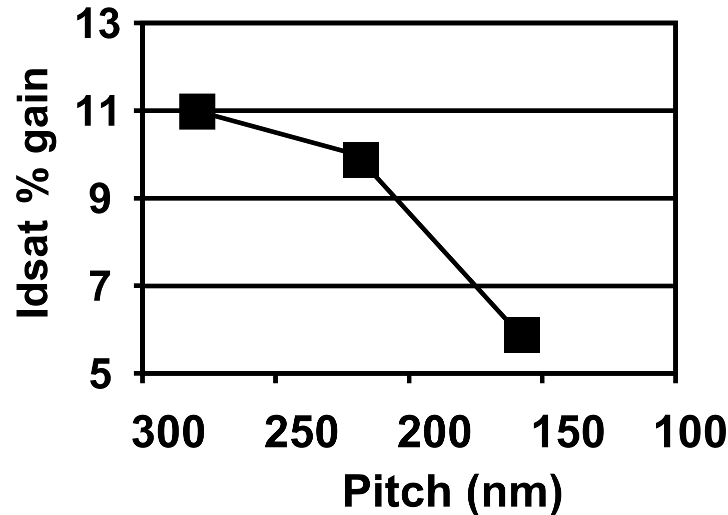
Strain: Importance in scaling



Strain (first introduced at 90nm) is a critical ingredient in modern transistor scaling

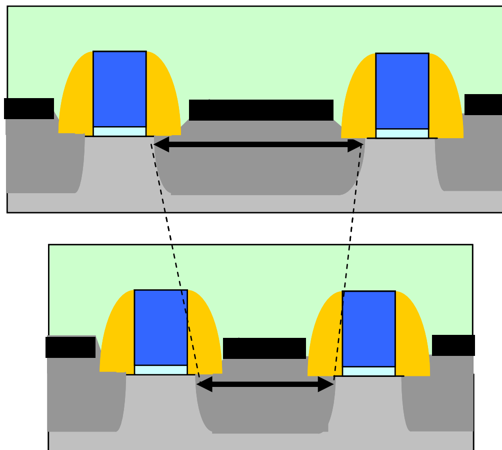
Strain: Pitch dependence

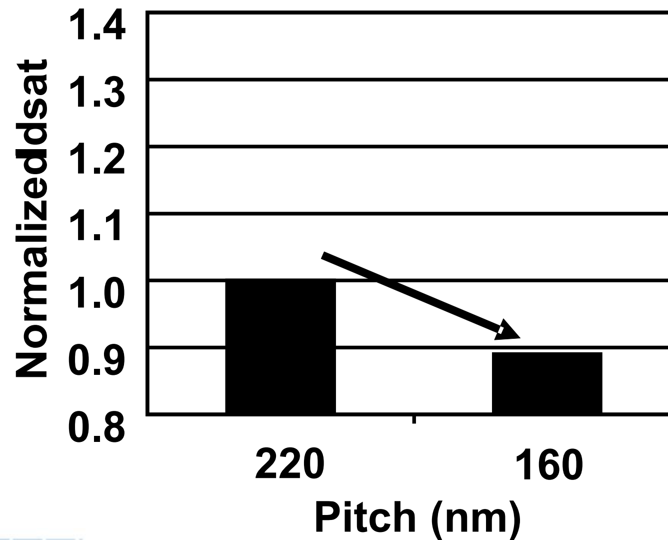




NMOS

Pitch degradation increases with film pinchoff, requires higher stress, thinner films



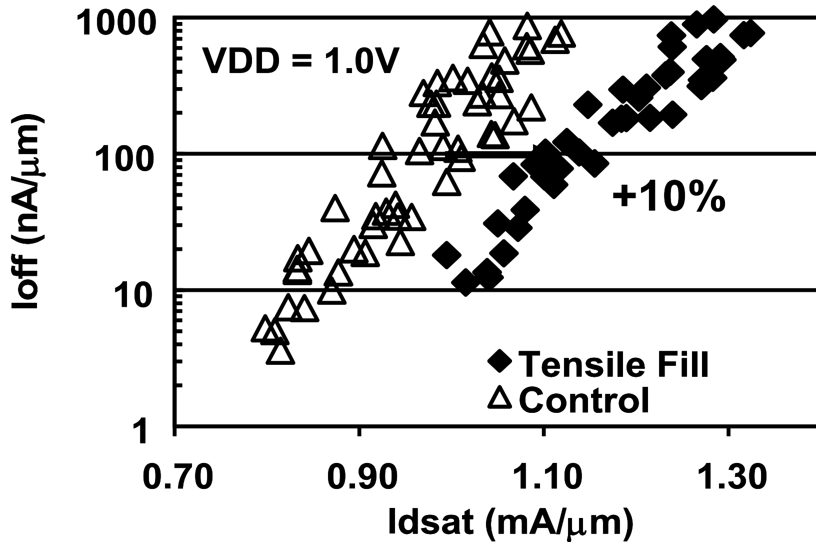


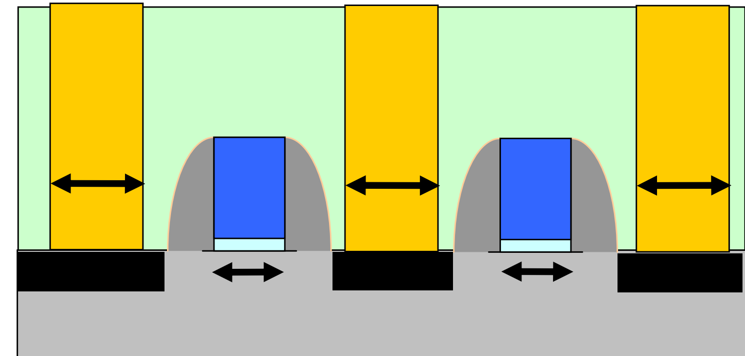
PMOS

eSiGe S/D mobility strongly dependent on pitch

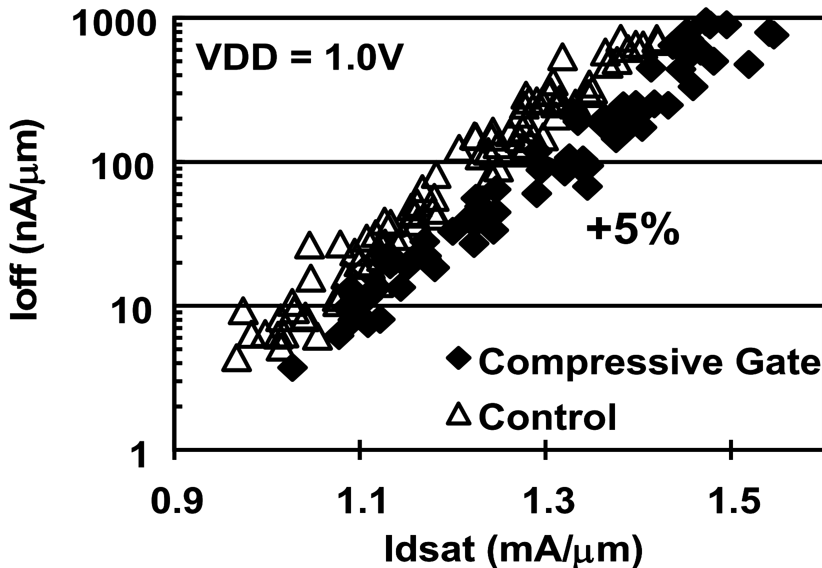
C. Auth, VLSI 2008

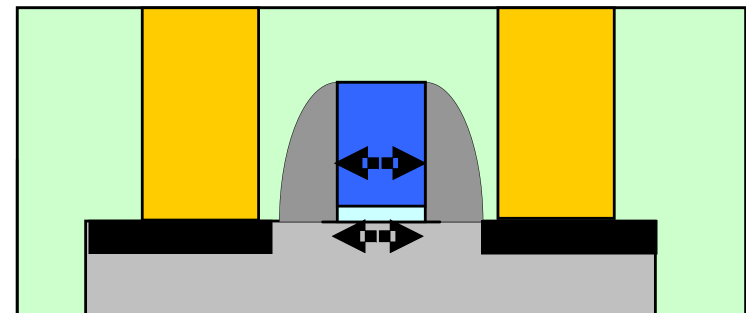
NMOS strain: Scaling with pitch





Tensile trench contacts

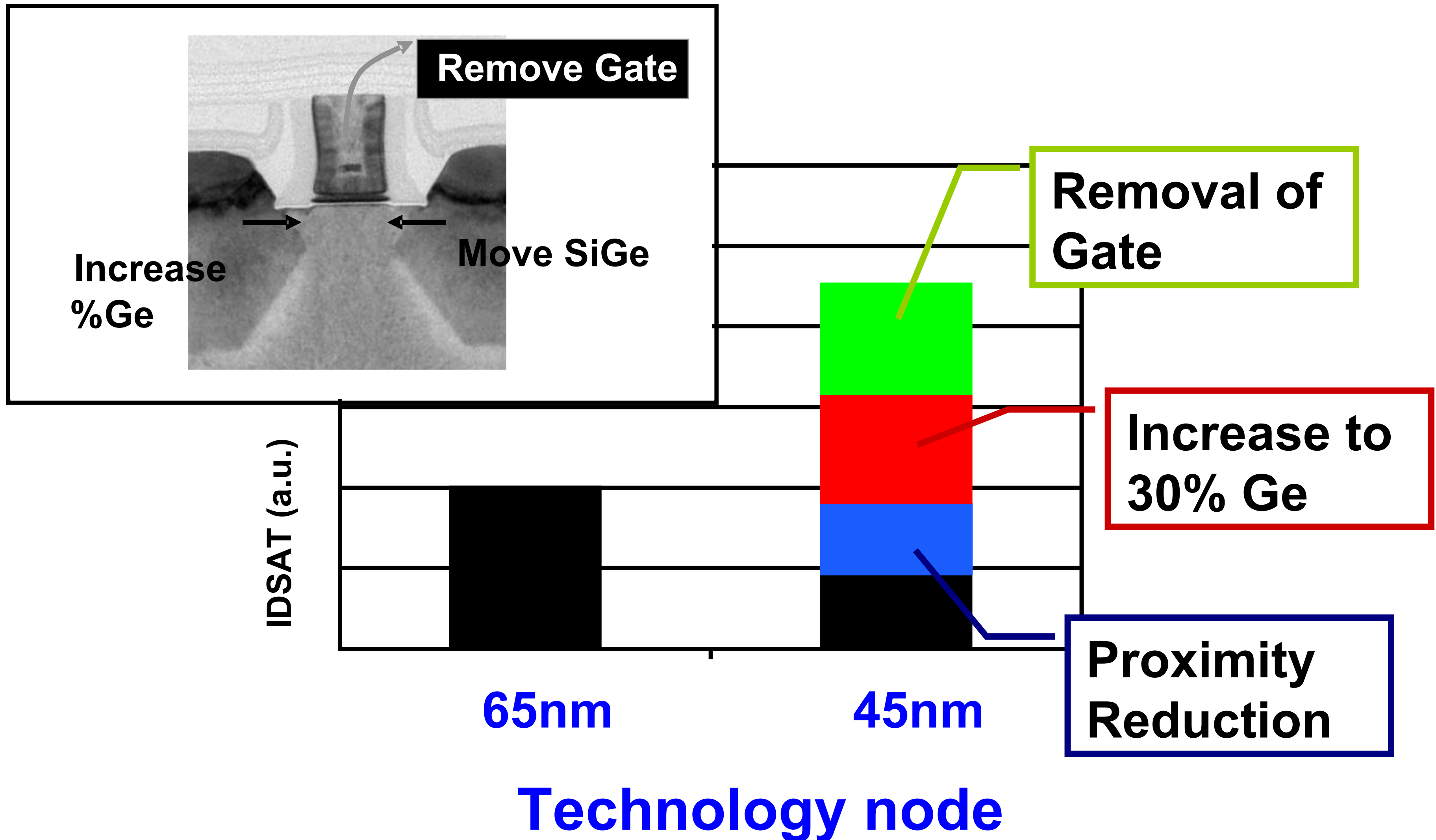




Compressive gate stress

C. Auth, VLSI 2008

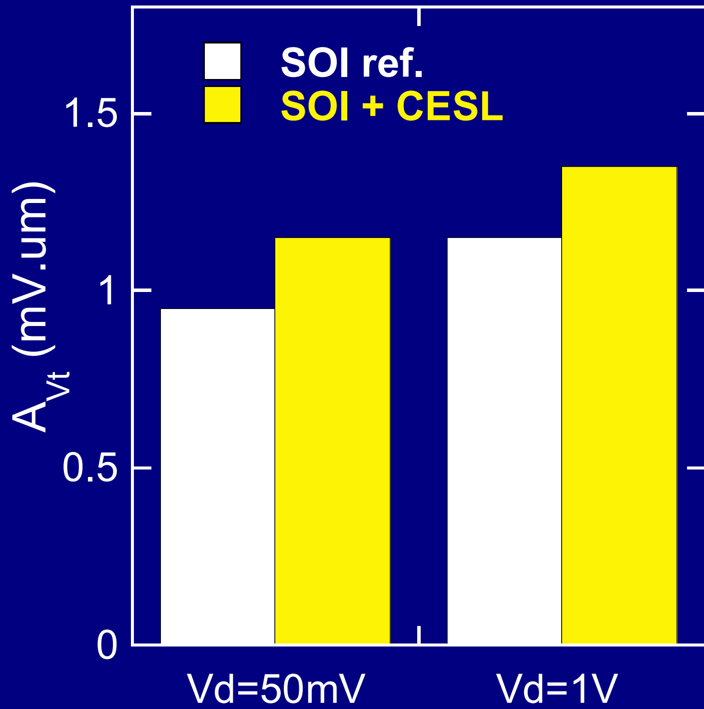
PMOS strain: Scaling with pitch

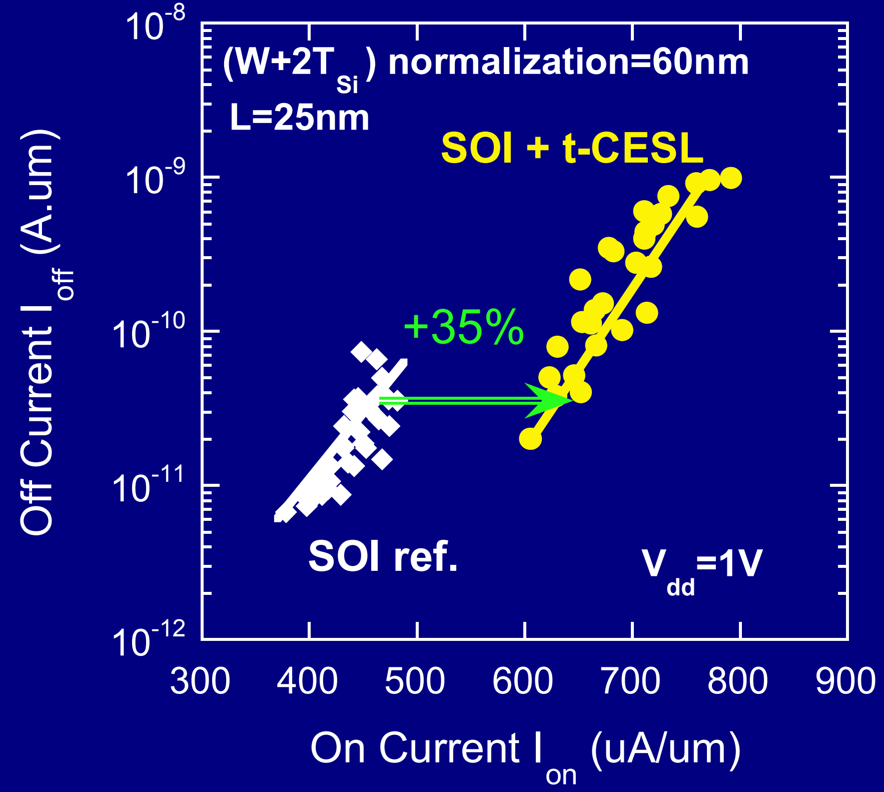


C. Auth, VLSI 2008

Random V_T variability and strain

Weber et al.
IEDM 2008
pp. 245-248





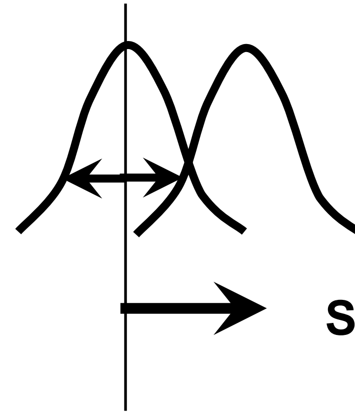
Similar V_T matching with CESL while 35% I_{ON} enhancement is achieved

650μA/μm – 30pA/μm at V_{dd}=1V and L_g=25nm

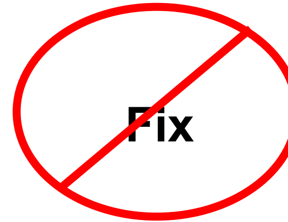
Part II: Measurements, results and interpretation

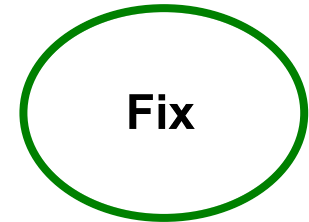
Systematic and Random

- **Statistician's viewpoint:**



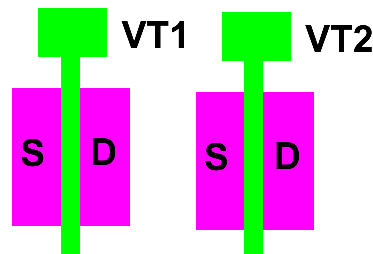
- **Process engineer's viewpoint:**

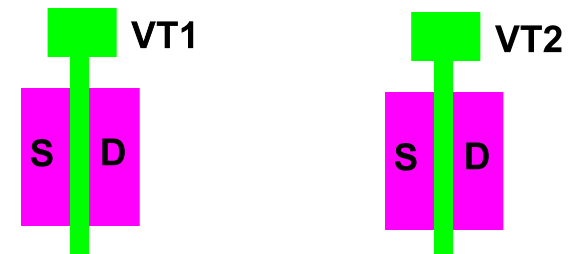




Random Systematic

- **Device engineer's viewpoint:**



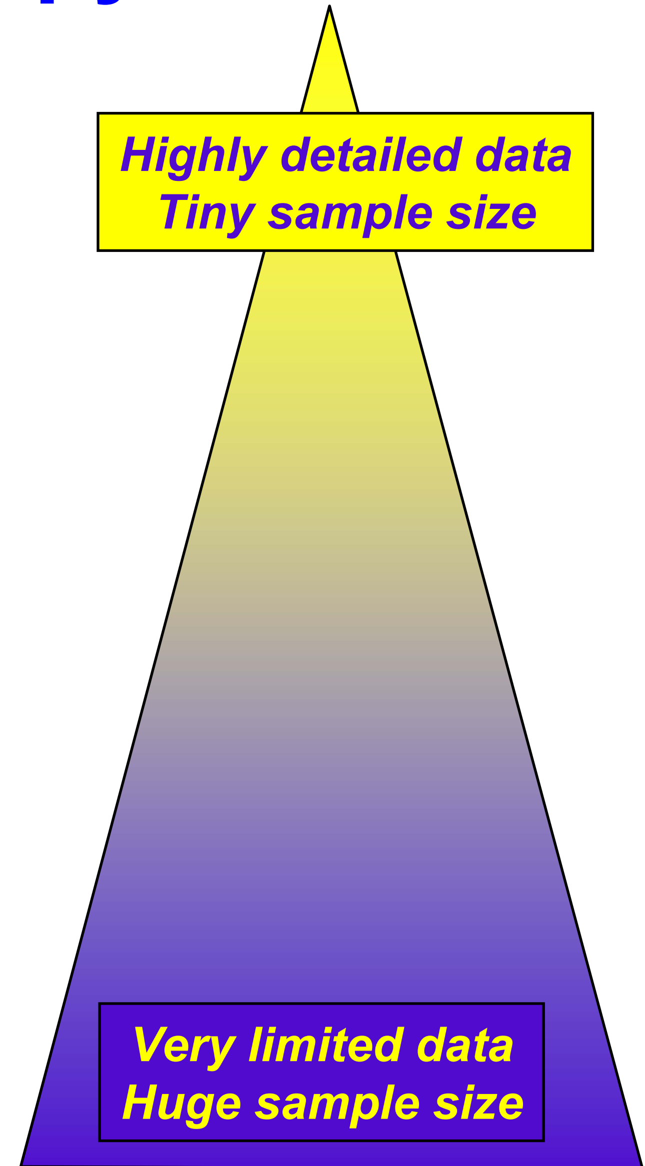


Random Systematic

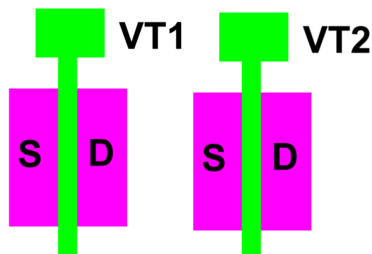
Measurement "food pyramid"

INCREASING DATA QUANTITY
DECREASING ABILITY TO SEGMENT ORIGIN

- In-line or off-line physical measurements of test wafers (TEM, SIMs, Auger, etc.)
- Device parametric measurements on test material (Ion/Ioff, IG/VG etc.)
- In-line physical measurements of selected sites in product (CD, thickness, etc.)
- Device parametric measurements on product (Idsat/lin, VT)
- Device parametric measurements on simple circuits (fmax, fmin, etc)
- Device sort on completed product (Vccmin and performance)



Measurement of Random and Systematic VT Variation at the Device Level



Traditional method:

1. Measure two identical adjacent devices and extract the difference $\sigma(VT_A\text{-}VT_B)$
2. Measure the entire population of all devices and extract $\sigma(VT_{pop})$

Random Variation for a matched pair

$$Random_{mp} = StdDev(VT_A - VT_B) = \sigma(DVT)$$

Random Variation for a single device

$$Random_{one-device} = \frac{StdDev(VT_A - VT_B)}{\sqrt{2}} = \frac{\sigma(DVT)}{\sqrt{2}}$$

Systematic Variation for a single device

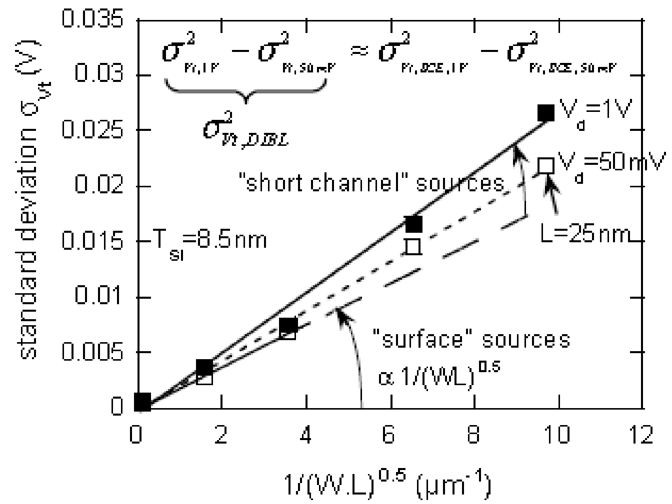
$$Systematic = \sqrt{(\sigma VT_{pop})^2 - \left(\frac{\sigma(DVT)}{\sqrt{2}}\right)^2}$$

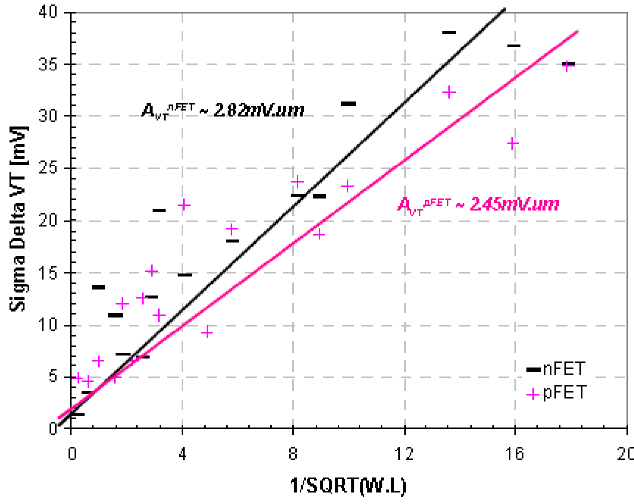
Pelgrom Plots: What is AVT anyway?

Two choices are widely used in the literature

IEDM 2008: Weber

IEDM 2008: Arnaud





$\sqrt{2}$

Choice A ⟷ **Choice B**

Slope of σVT vs $1/\sqrt{LW}$ < Slope of $\sigma\Delta$VT vs $1/\sqrt{LW}$

What did Pelgrom say?

Pelgrom "Matching properties of MOS transistors"
(IEEE Journal of Solid-State Circuits, Vol. 24, No. 5, Oct. 1989)

- Eq. 5 defines a generic A_P for a parameter ΔP; implying A_{VT} would then be the parameter for ΔVT

$$\sigma^2(\Delta P) = \frac{A_P^2}{WL} + S_P^2 D_x^2.$$

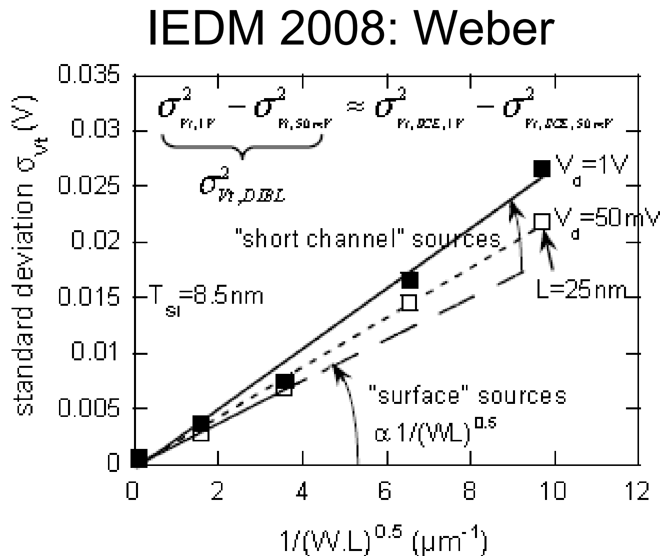
- However, one page further in the paper, he explicitly defines A_{VT} in terms of V_T only in equation 8:

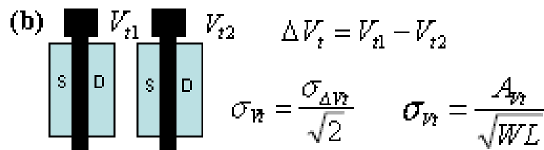
$$\sigma^2(V_{T0}) = \frac{A_{VT0}^2}{WL} + S_{VT0}^2 D^2$$

- So – which is did he mean? Well, I asked him.

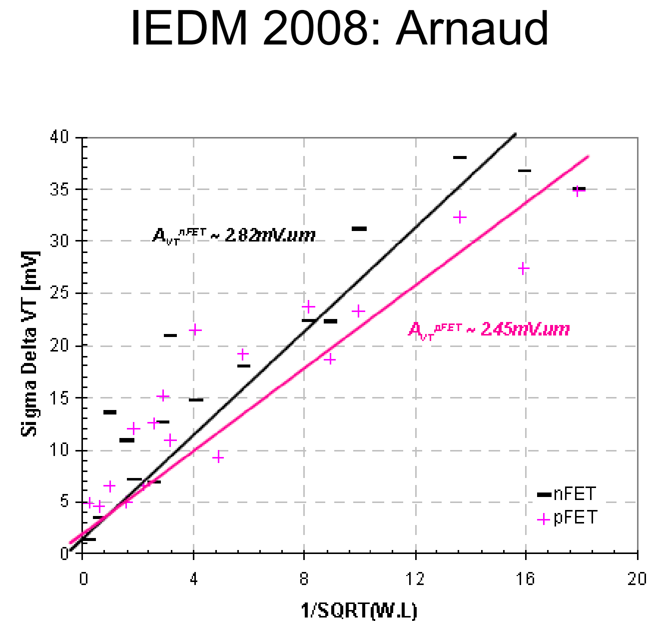
What is AVT anyway?

IEDM 2008: Weber





IEDM 2008: Arnaud



This is AVT

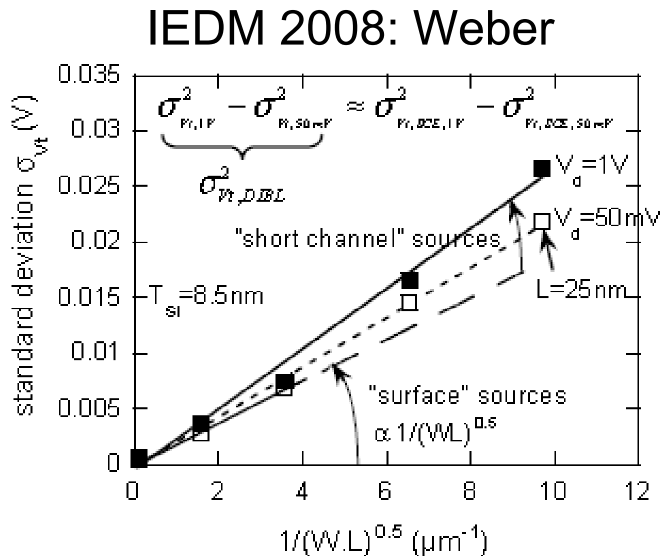
$\sqrt{2}$

Choice A ⟷ Choice B

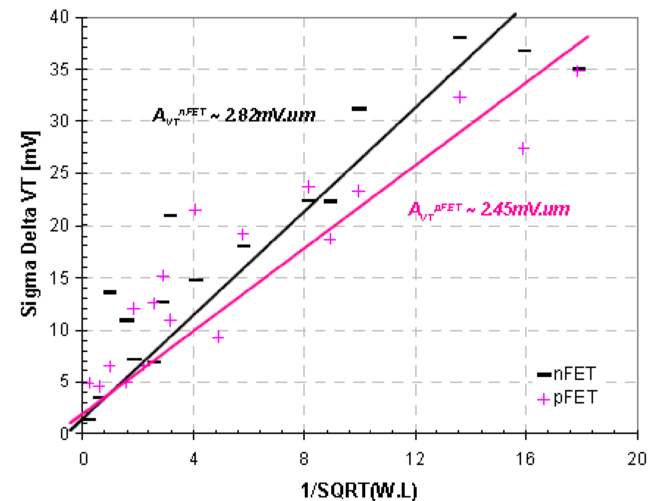
Slope of σVT vs 1/√LW **<** Slope of σ**Δ**VT vs 1/√LW

What is A_{VT} anyway?

IEDM 2008: Weber



IEDM 2008: Arnaud



I will call this C_{VT} (or C_2)

$C_{VT} = A_{VT} / \sqrt{2}$

This is A_{VT}

Choice A

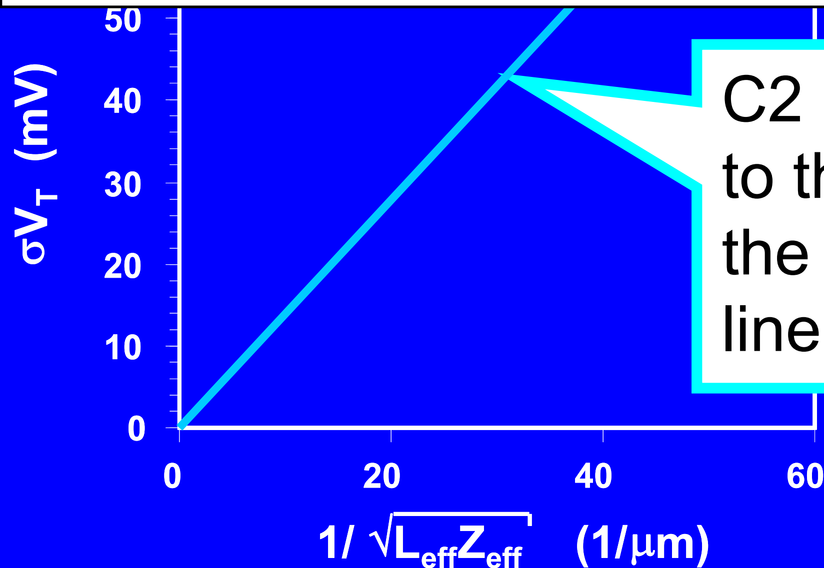
Slope of σVT vs $1/\sqrt{LW}$

<

Choice B

Slope of $\sigma\Delta$VT vs $1/\sqrt{LW}$

Additional propagation of confusion (By me, it turns out …)



C2 is proportional to the slope of the 1 / $\sqrt{Leff \cdot Zeff}$ line

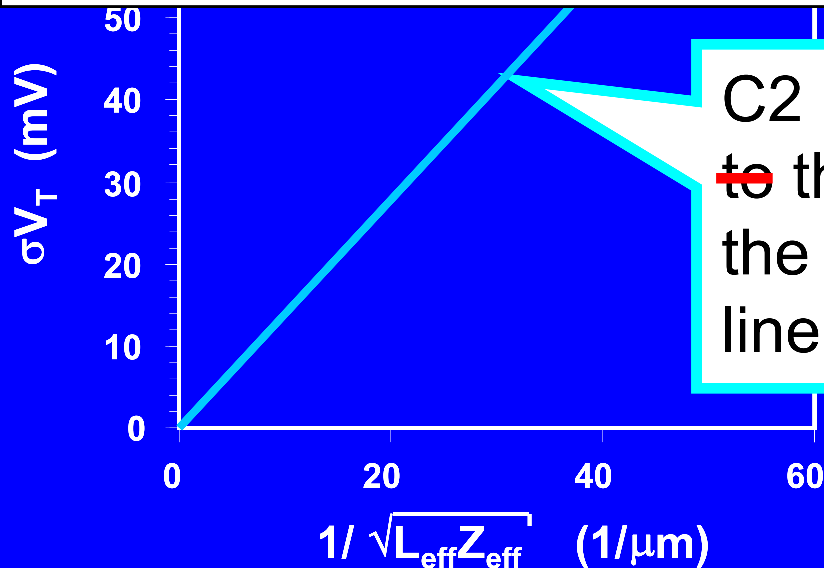
K. Kuhn, IEDM 2007

RDF is frequently described by (Stolk):

$$\sigma V_{Tran} = \left(\frac{\sqrt[4]{4q^3 \varepsilon_{si} \phi_B}}{2} \right) \cdot \frac{T_{ox}}{\varepsilon_{ox}} \cdot \left(\frac{\sqrt[4]{N}}{\sqrt{Leff \cdot Zeff}} \right) = \frac{1}{\sqrt{2}} \left(\frac{c_2}{\sqrt{Leff \cdot Zeff}} \right) \quad (1)$$

P. Stolk, F. Widdershoven, and D/ Klaassen, “Modeling statistical dopant fluctuations in MOS transistors” IEEE Trans. on Elec. Dev., 45:9, pp 1960-1971, Sept. 1998

Additional propagation of confusion (By me, it turns out …)



C2 is ~~proportional to~~ the slope of the 1 / $\sqrt{Leff \cdot Zeff}$ line **vs σVT**

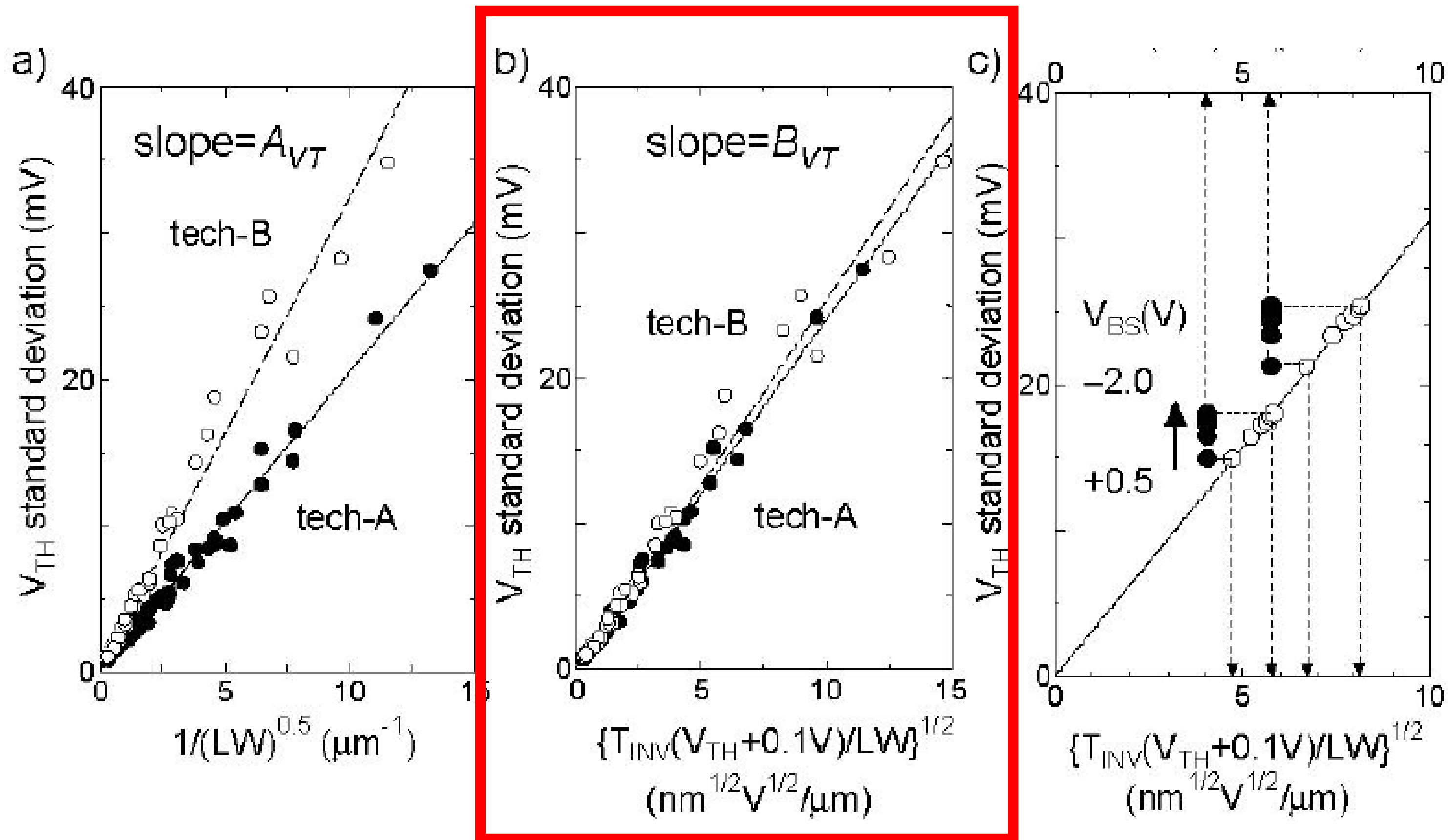
K. Kuhn, IEDM 2007

RDF is frequently described by (Stolk):

$$\sigma V_{Tran} = \left(\frac{\sqrt[4]{4q^3 \varepsilon_{si} \phi_B}}{2} \right) \cdot \frac{T_{ox}}{\varepsilon_{ox}} \cdot \left(\frac{\sqrt[4]{N}}{\sqrt{Leff \cdot Zeff}} \right) = \frac{1}{\sqrt{2}} \left(\frac{c_2}{\sqrt{Leff \cdot Zeff}} \right) \quad (1)$$

P. Stolk, F. Widdershoven, and D/ Klaassen, "Modeling statistical dopant fluctuations in MOS transistors" IEEE Trans. on Elec. Dev., 45:9, pp 1960-1971, Sept. 1998

What is B_{VT} then?

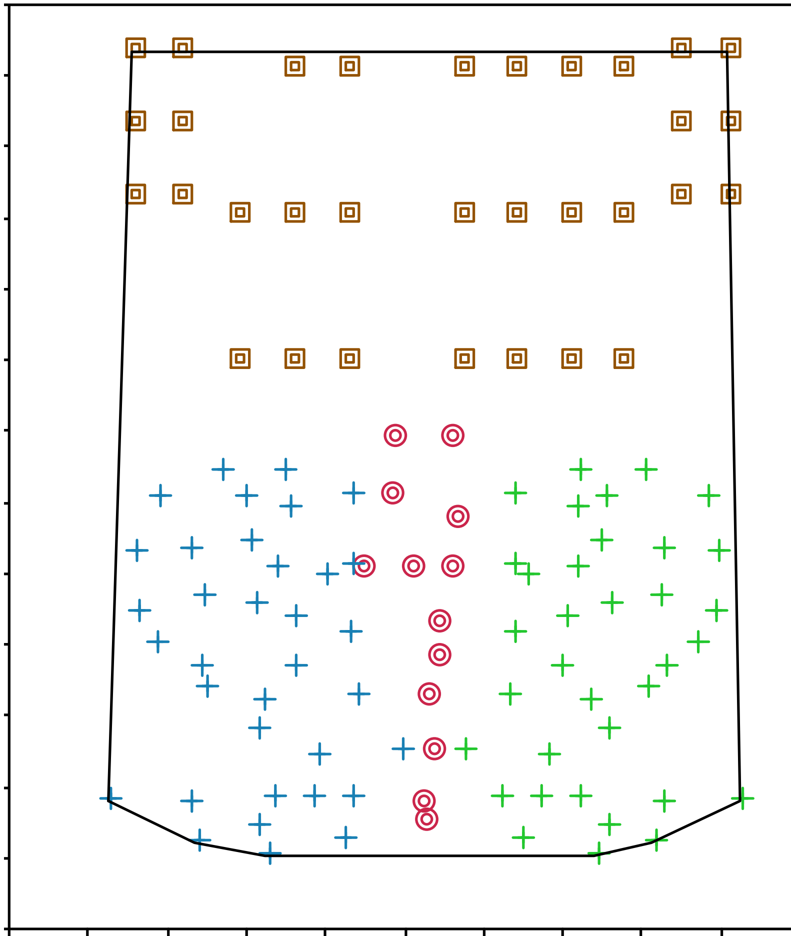


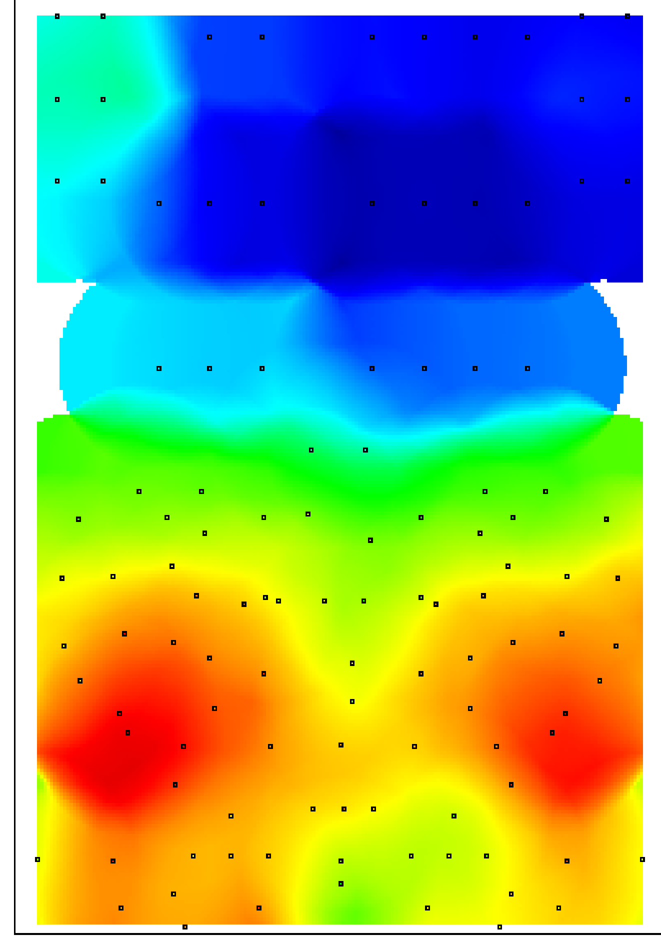
B_{VT}

Slope of σVT vs $\sqrt{(T_{inv}(V_T+0.1)/LW)}$

Fig. 3: Takeuchi, IEDM 2007

But what about simple circuits?





One powerful tool for assessment of variation is locating ring-oscillators (ROs) routinely in all product designs

Random and Systematic Variation for Matched Ring Oscillators

Random:

- Calculate Delta

$$Delta = \frac{FreqA - FreqB}{FreqA + FreqB} * \frac{200}{\sqrt{2}}$$

- **Random Variation**

$$Rand = StdDev(Delta)$$

per data unit

Systematic:

- Total Sigma

$$\sigma = StdDev(FreqA)$$

per data unit

- Grand Mean

$$\mu = \frac{Mean(FreqA) + Mean(FreqB)}{2}$$

- **Systematic Variation**

$$Syst = \sqrt{\left(\frac{\sigma}{\mu} * 100\right)^2 - Rand^2}$$

per data unit

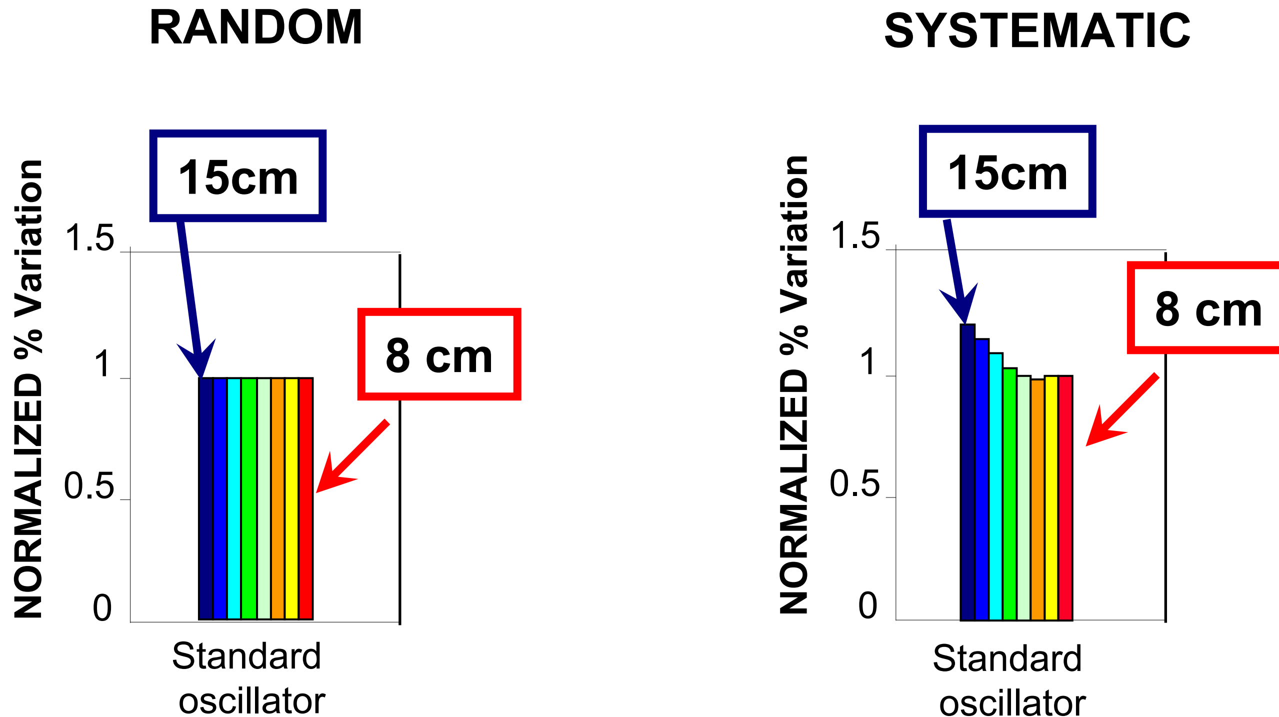
Total Variation:

$$Total = \frac{StdDev(FreqA)}{Mean(FreqA)} * 100$$

per data unit

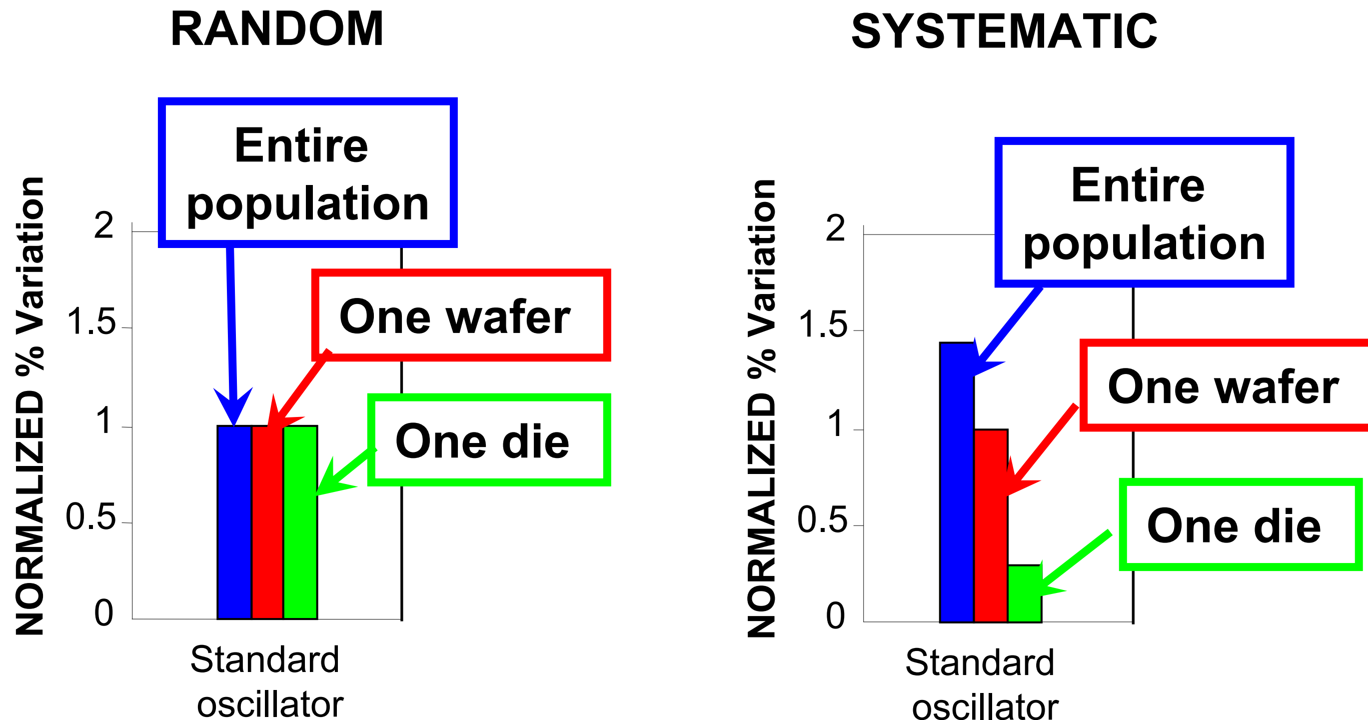
intel

45nm: Within Wafer Variation



For random variation: Uniform across wafer
For systematic variation: More variation at the wafer edge

45nm: Within Die (WID), Within Wafer (WIW) and Wafer to Wafer (WTW)



For random variation: Uniform with population choice
For systematic variation: Variation increases significantly going from within-die (WID) to within-wafer (WIW)

45nm Product wafer: Random variation

Delta 4-5

K. Kuhn, ITJ 2008

NORMALIZED % Variation

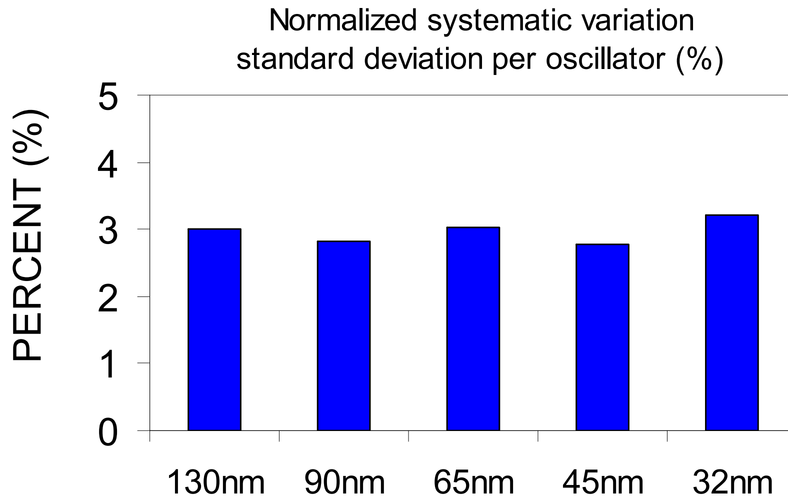
15cm

8 cm

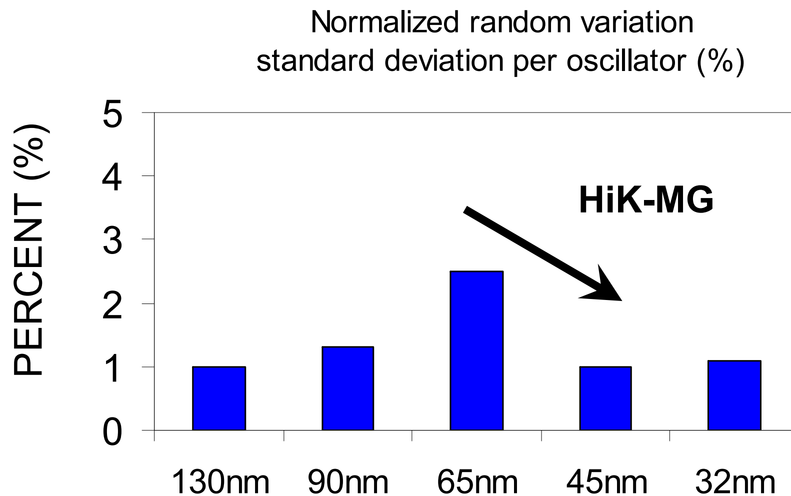
1.5
1
0.5
0

Standard oscillator

Random and Systematic Variation Trends



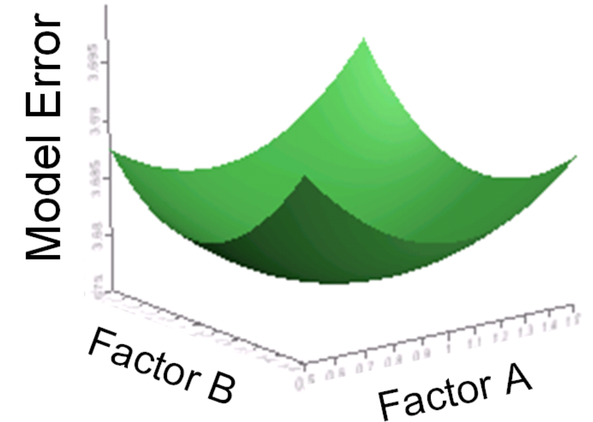
Systematic WIW variation is comparable from one generation to the next

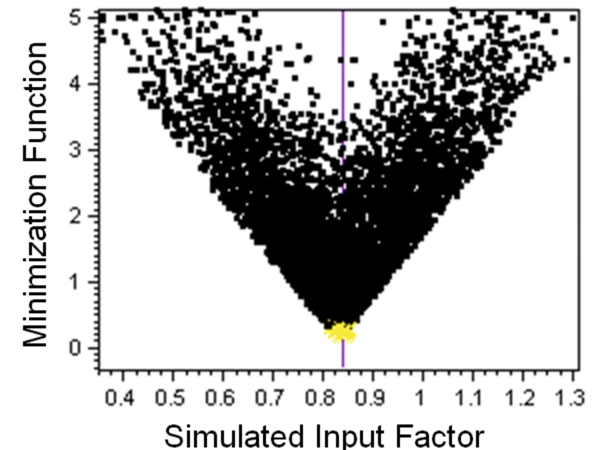


Random WIW variation in 32nm is comparable to 45nm and significantly improved over 65nm and 90nm due to HiK-MG

What about more complex circuits?
RSM Methodology for Variation Model Parameters

- Identify the set of input parameters in variation modeling files that can be allowed to vary
- Create DOE to vary all parameters within selected limits
- Create a series of variation modeling files, using the matrix of parameters from the DOE
- Simulate an appropriate set of circuits and devices to obtain responses to the set of variation modeling files
- Enter simulation results back into DOE to determine sensitivity to model parameters
- Optimize variation modeling file parameters to get best match to measured data





Example Matrix of Inputs and Associated Responses

Responses

1

2

3

4

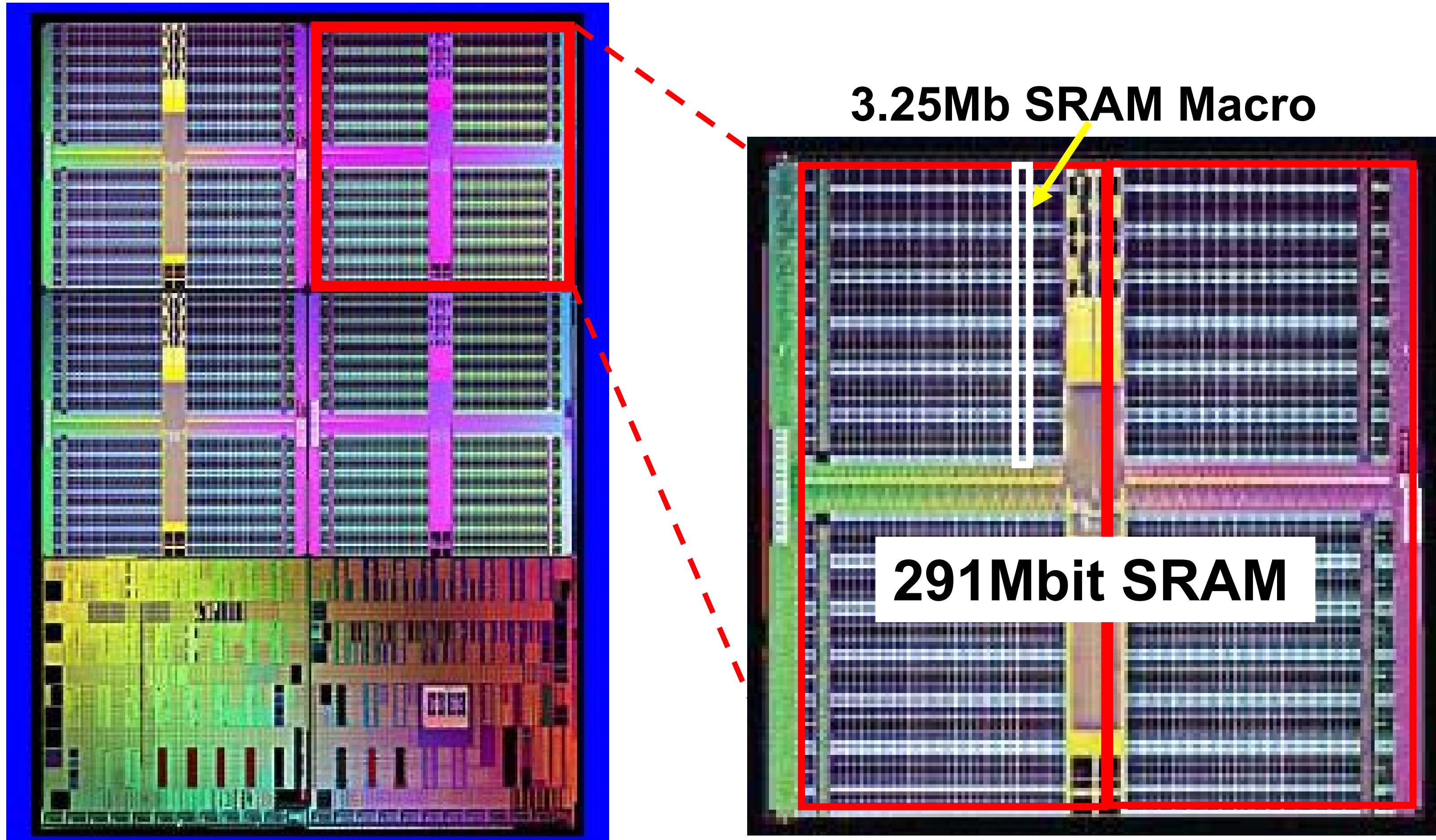
Sensitivity of response “2” to input “E”

A B C D E F G H I J

Inputs

Not all responses are sensitive to all inputs– key is to determine which responses are appropriate for setting each input parameter

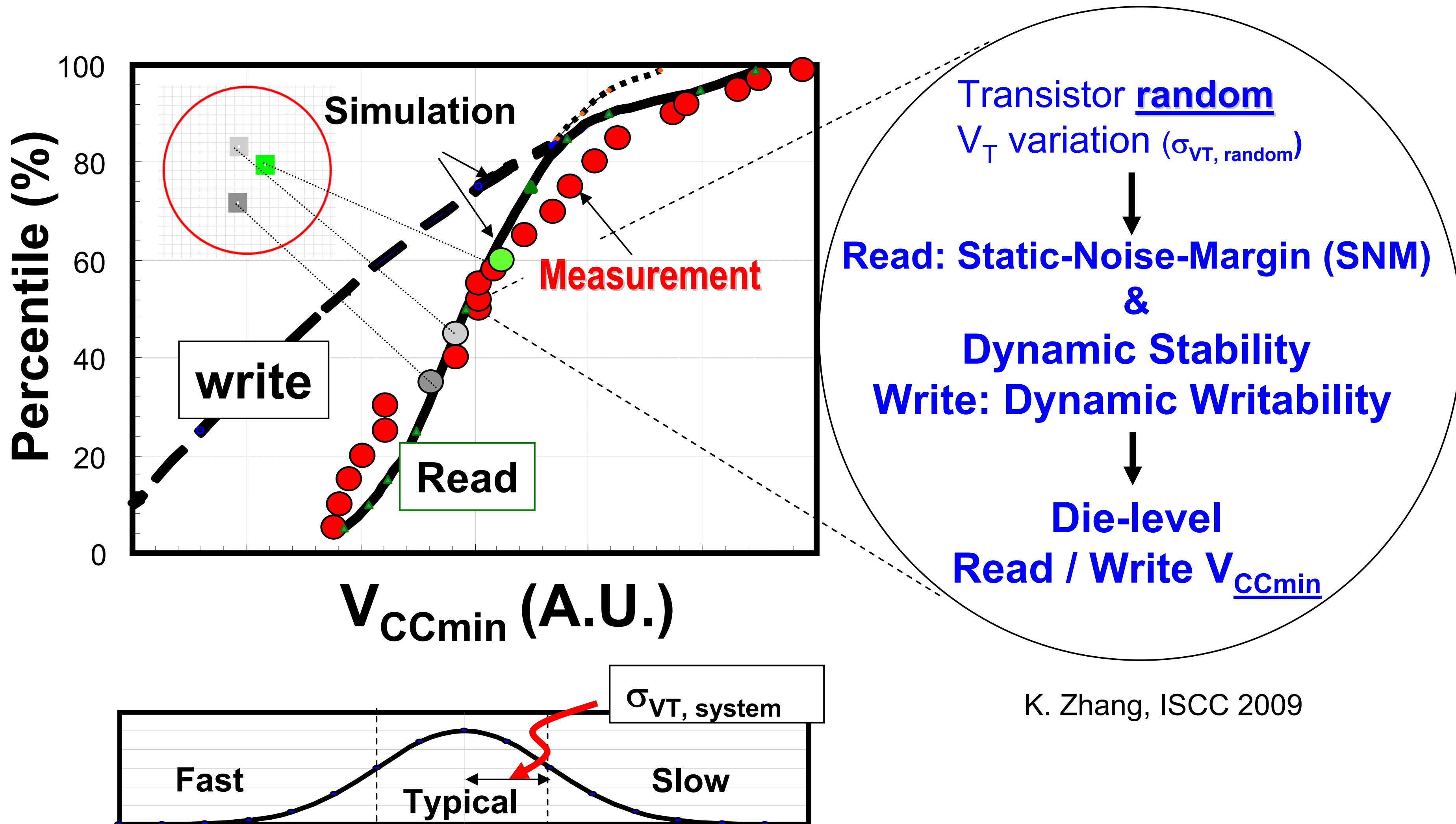
32nm SRAM Test Chip



SRAM test chip with advanced test features (PBIST, eFUSE, ECC, etc.) to support development of 32nm high-volume manufacturing process

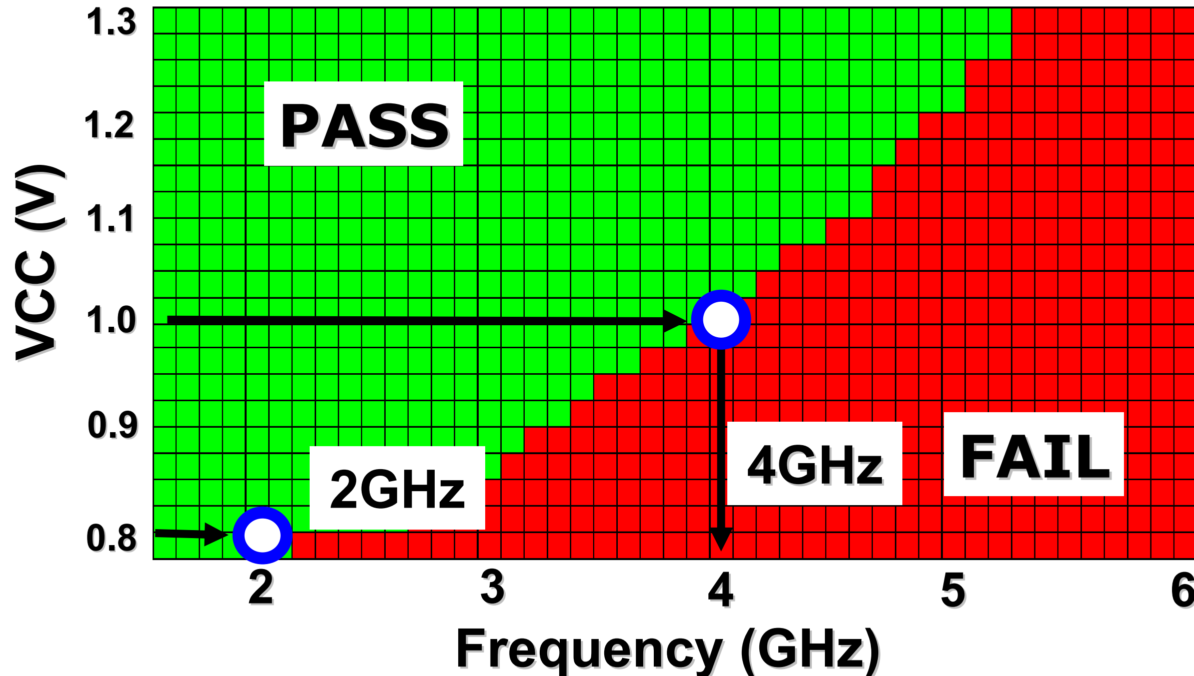
K. Zhang, ISCC 2009

SRAM V_{CCmin} – Silicon to Simulation



Wafer-level SRAM P/NMOS transistor **systematic** V_T variation

32nm Voltage-Frequency Shmoo

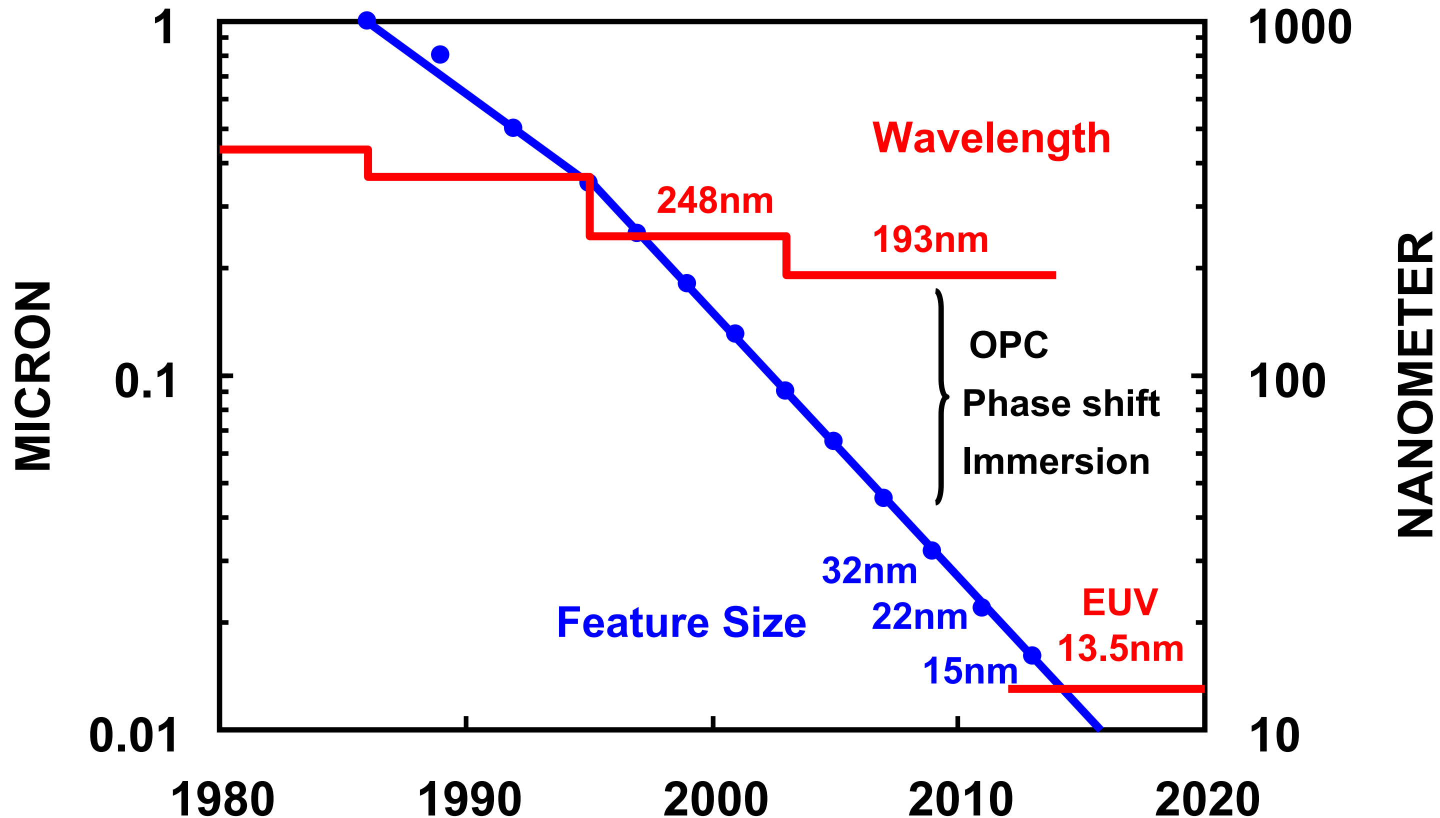


- 32nm SRAM operates over a broad range of supply voltages, enabling dynamic voltage scaling for low-power application
- 32nm SRAM achieves operating frequency of 4GHz at 1.0V, 15% better than 45nm design

K. Zhang, ISCC 2009

Part III: Next generation challenges

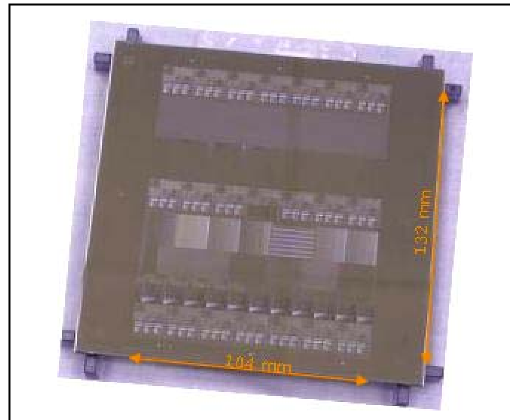
Lithography Pipeline



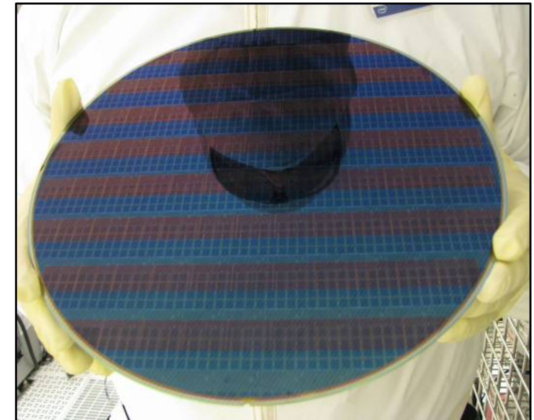
Extend 193nm Optical Lithography as far as possible
Deploy EUV Lithography when available/affordable

Extreme Ultraviolet Lithography

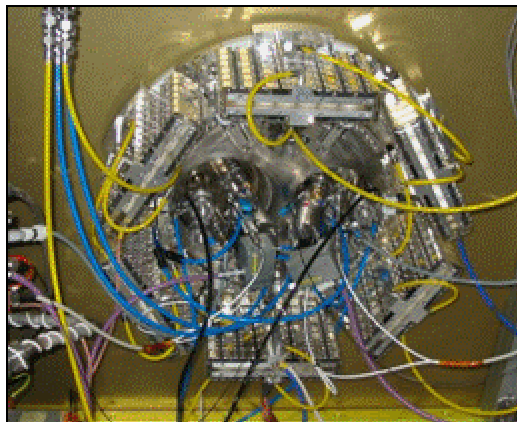
Cymer beta source



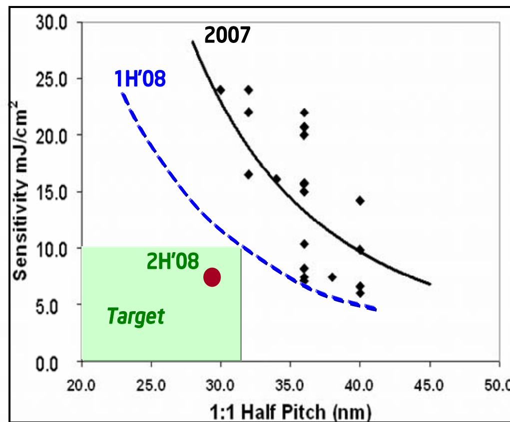
Intel EUV Mask



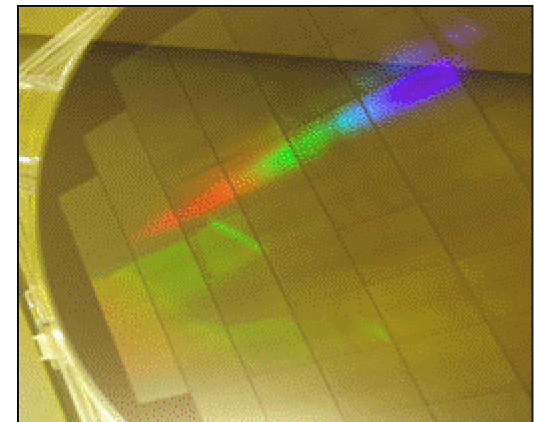
ASML ADT printed wafer



Philips beta source



Photoresist Development



Nikon EUV1 printed wafer

Continued progress towards EUV implementation

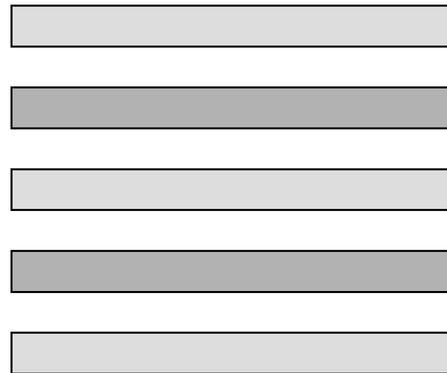
M. Bohr, ISCC, 2009

Non-EUV Lithography Beyond 32 nm

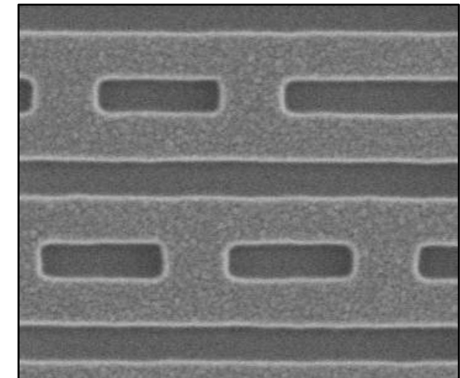
Double Patterning

- Pitch doubling
- Improved 2-D features

Pitch Doubling

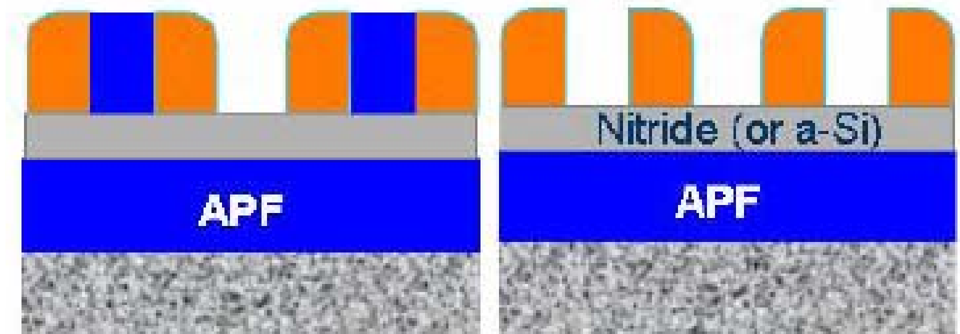


2-D Features



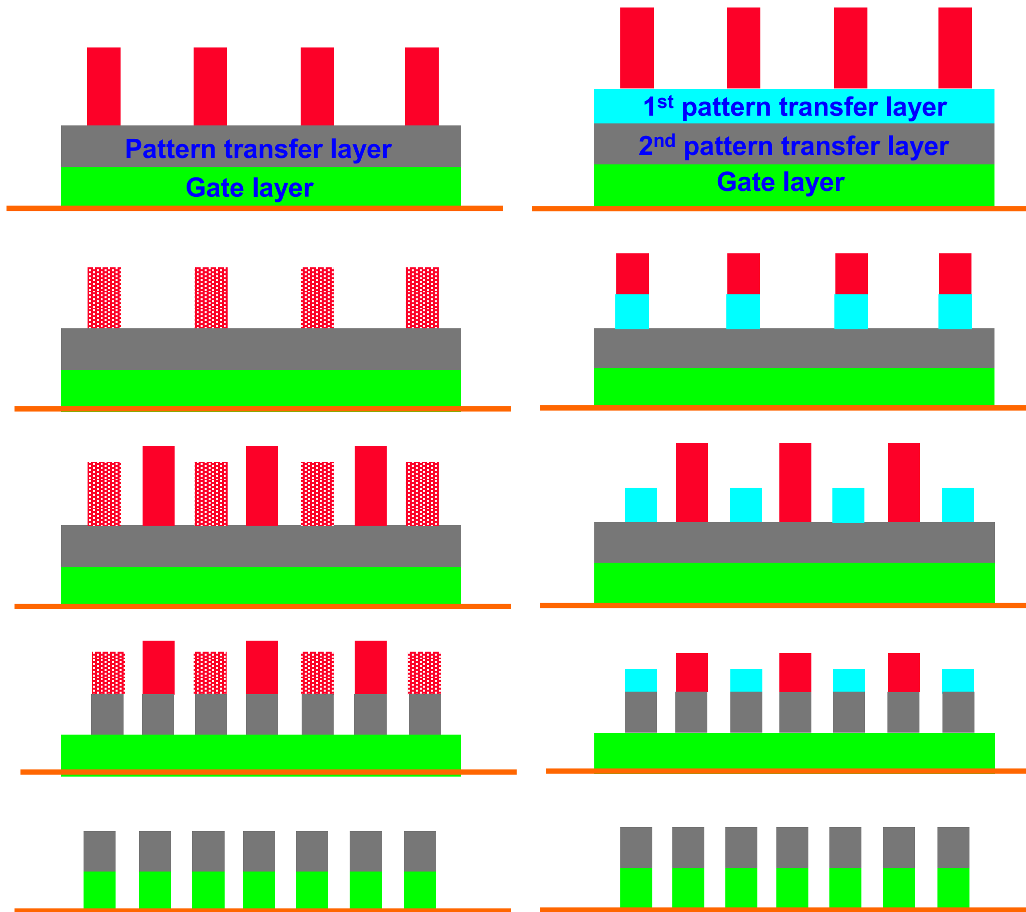
Spacer Gate Patterning

- Pitch doubling
- Improved variation



M. Bohr, ISCC, 2009
Bencher et al, Proc. of SPIE Vol. 6924 69244E-7

Pitch doubling and gate CD control



Resist freeze

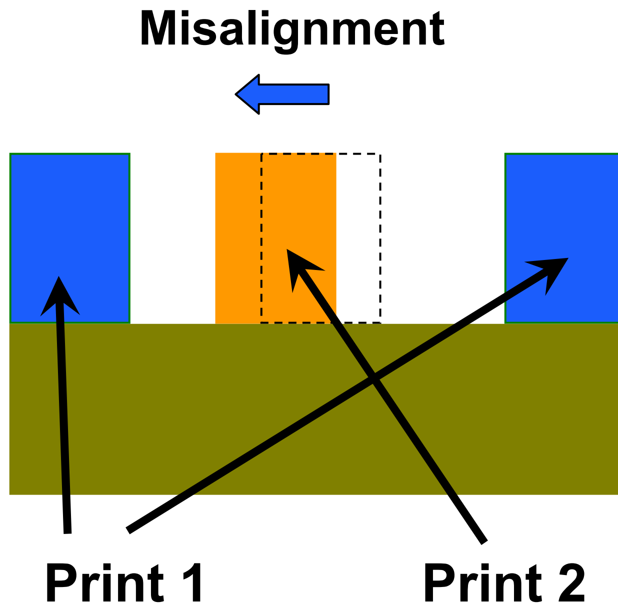
Double Pattern Transfer

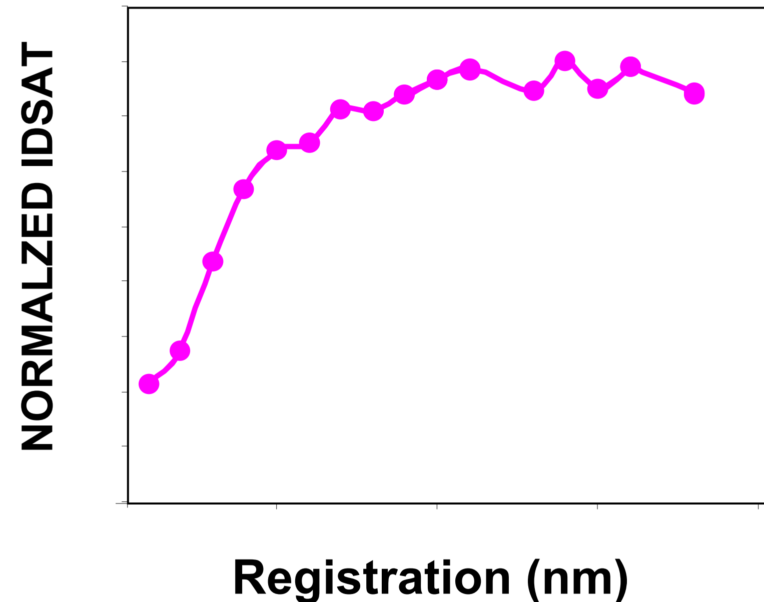
Neither Resist Freeze nor Double Pattern Transfer achieve full benefit of patterning at ½ pitch

Both techniques still require resolution of a very small space (MEEF, LWR etc.)

C. Kenyon, TOK conf., Dec. 2008

Disadvantages of Double-patterning

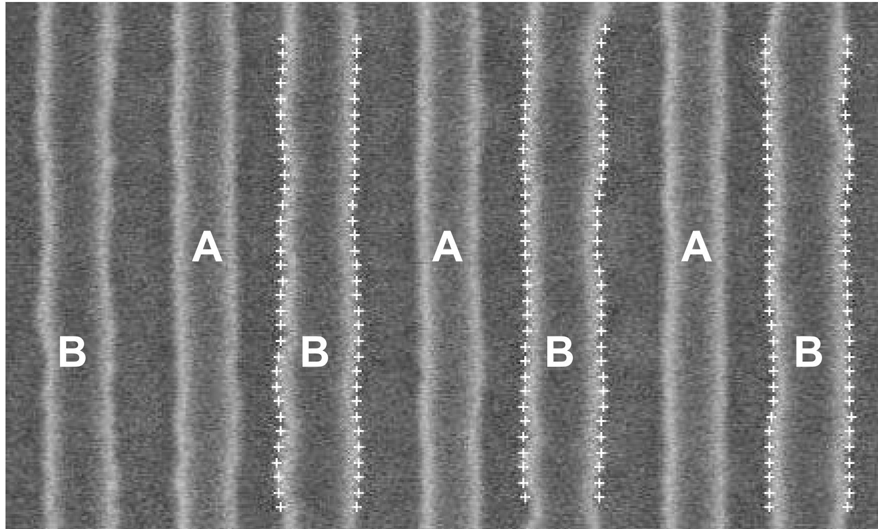


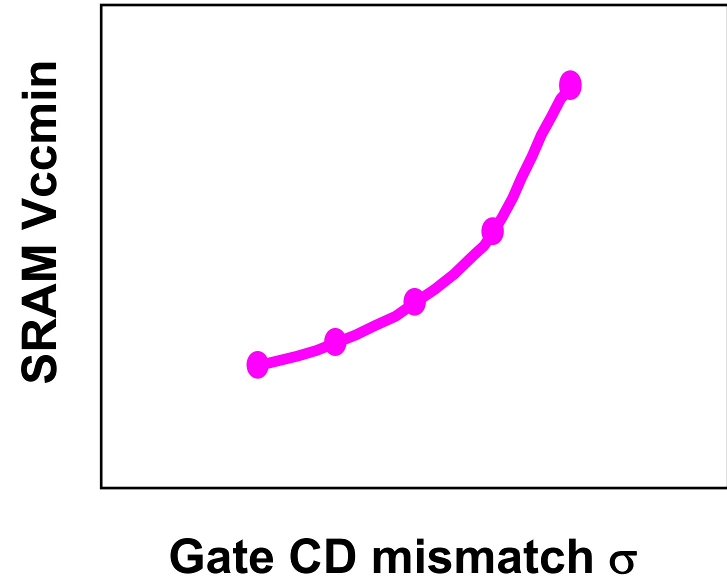


Misalignment between the 2 exposures is a crucial liability for this technique and can limit its usability

Transistor parameters can be affected by asymmetry between the source and drain regions

Pitch doubling and gate CD matching



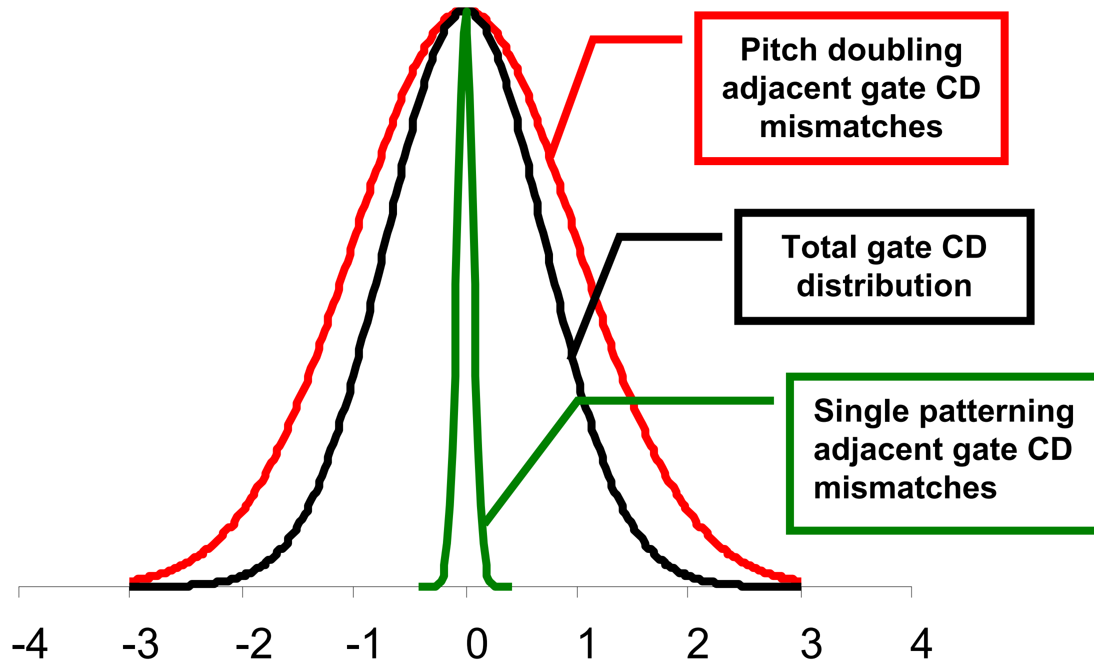


Pitch doubling eliminates the close correlation which currently exists between the CDs of adjacent gates

This has implications for memory cells and other circuits which depend upon this CD matching

C. Kenyon, TOK conf., Dec. 2008

Pitch doubling and gate CD matching



Single patterning: the distribution of CD mismatches between adjacent gates is a very small fraction of total gate CD variation

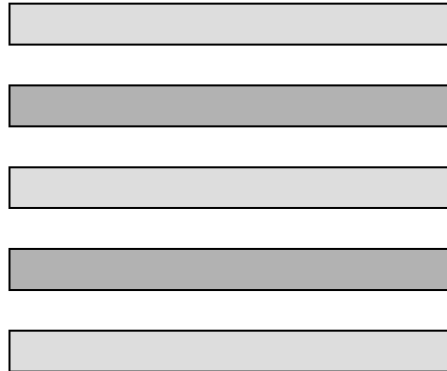
Pitch doubling: the distribution of CD mismatches is GREATER than the total gate CD variation

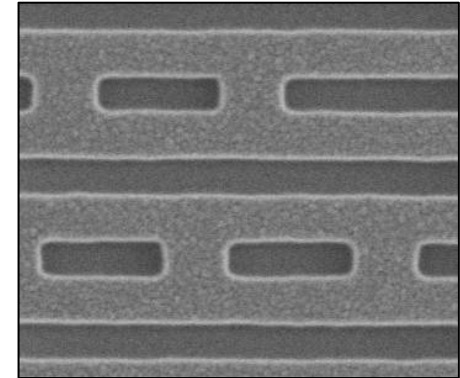
C. Kenyon, TOK conf., Dec. 2008

Non-EUV Lithography Beyond 32 nm

Pitch Doubling

2-D Features



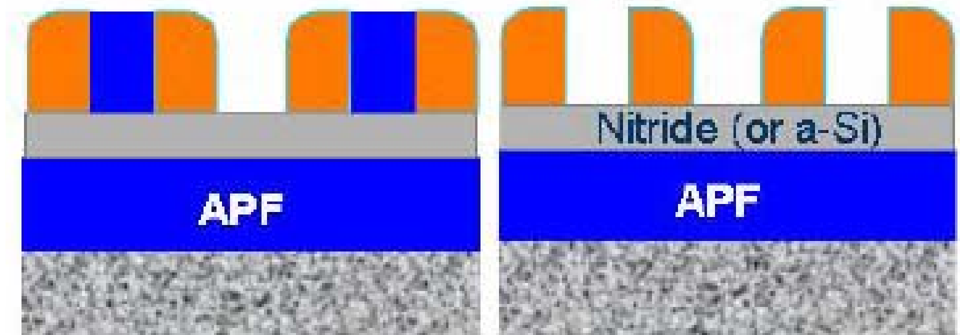


Double Patterning

- Pitch doubling
- Improved 2-D features

Spacer Gate Patterning

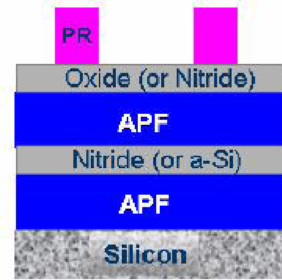
- Pitch doubling
- Improved variation



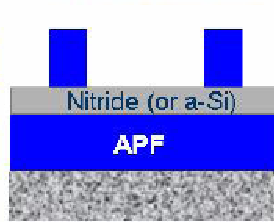
M. Bohr, ISCC, 2009
Bencher et al, Proc. of SPIE Vol. 6924 69244E-7

Alternative: Spacer patterning

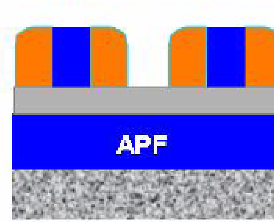
(1) Print and Resist Trim



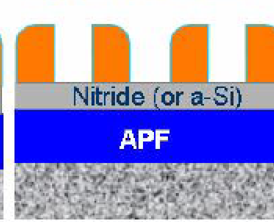
(2) Etch Template



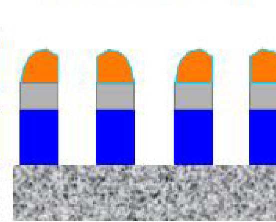
(3) Form Spacers



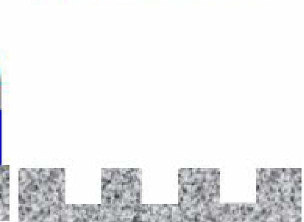
(4) Strip Template

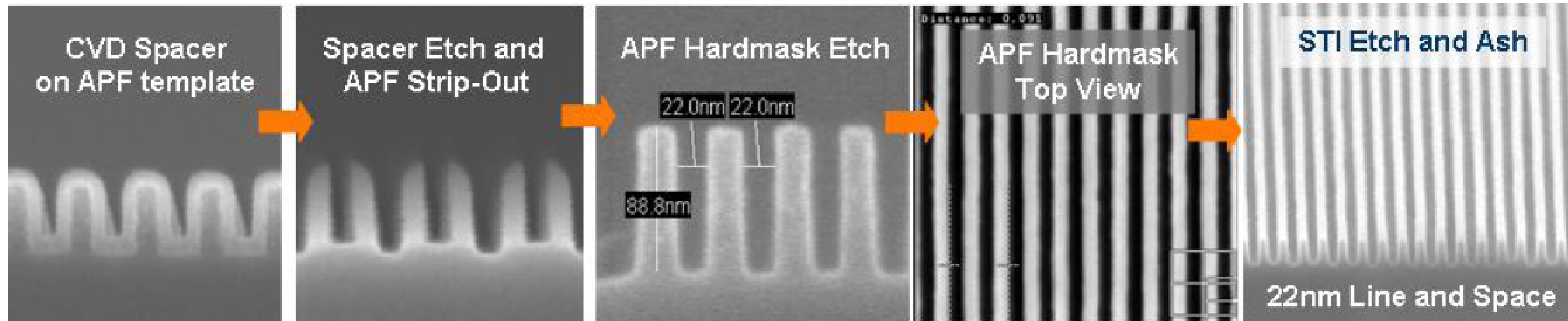


(5) Transfer Etch



(6) STI Etch and ash



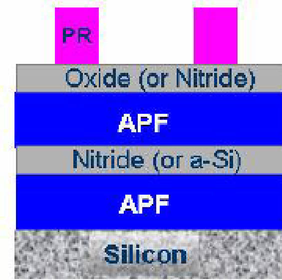


Spacer patterning retains correlation between doubled features

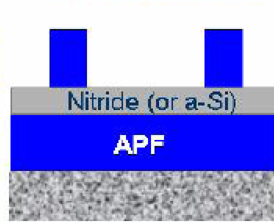
Bencher et al, Patterning by CVD Spacer Self Alignment DoublePatterning (SADP), Proc. of SPIE Vol. 6924 69244E-7

Alternative: Spacer patterning

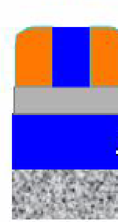
(1) Print and Resist Trim



(2) Etch Template



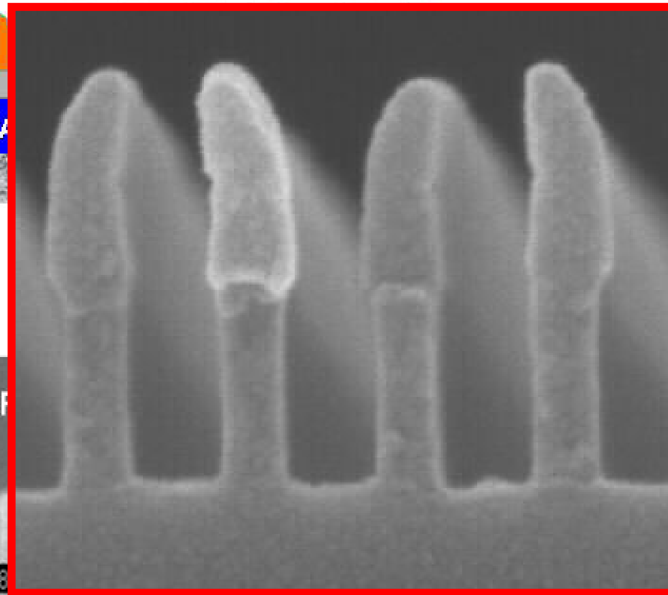
(3) Form Spacers

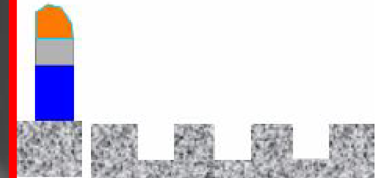


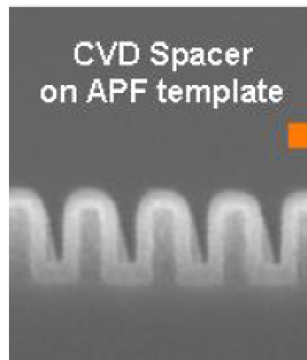
(4) Strip Template

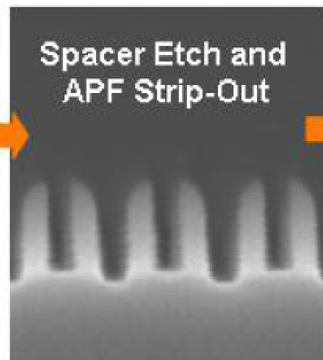
(5) Transfer Etch

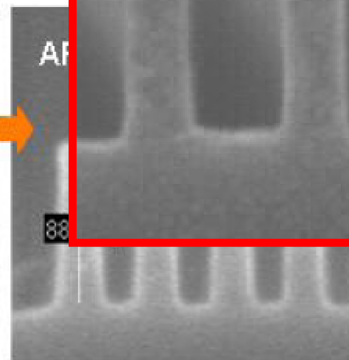
(6) STI Etch and ash

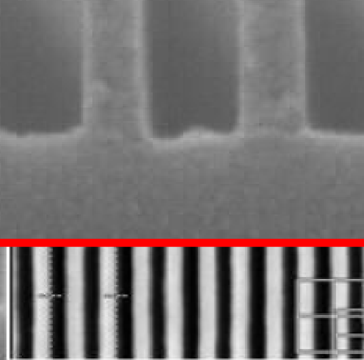


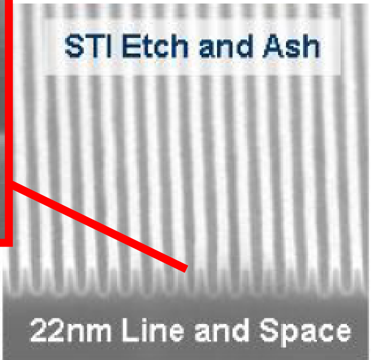








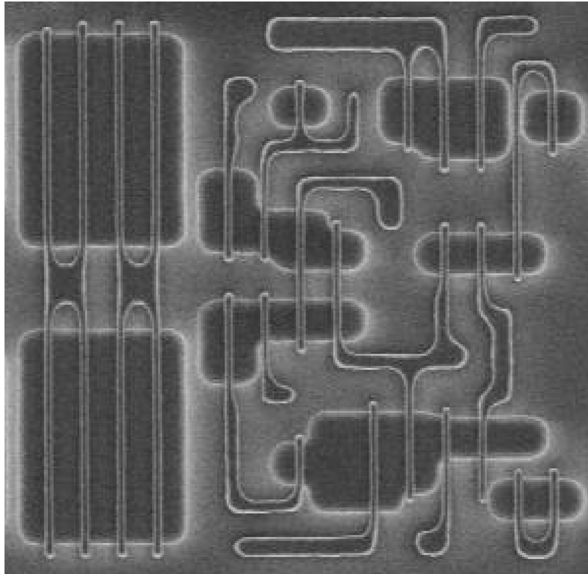




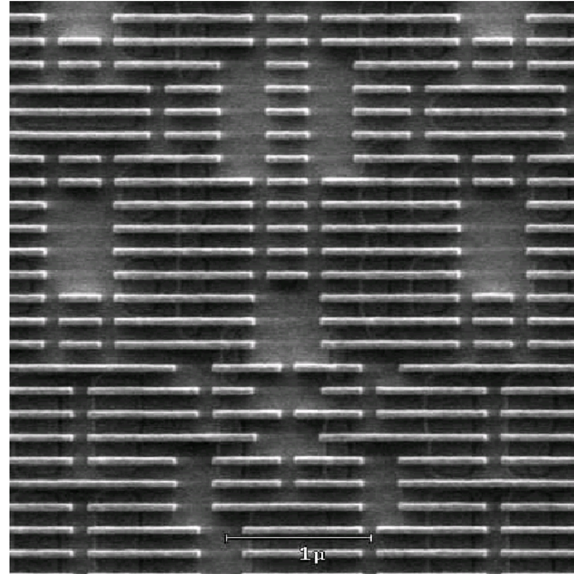
Spacer inhomogenities not transferred to patterned features

Bencher et al, Patterning by CVD Spacer Self Alignment DoublePatterning (SADP), Proc. of SPIE Vol. 6924 69244E-7

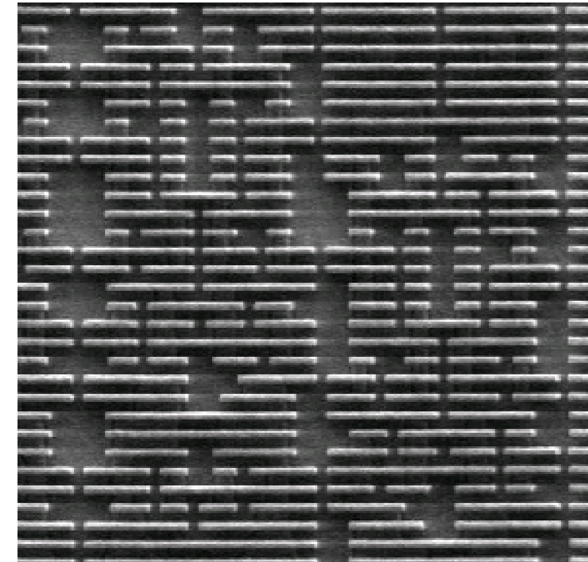
Uniformity matters: Logic images vs. technology node



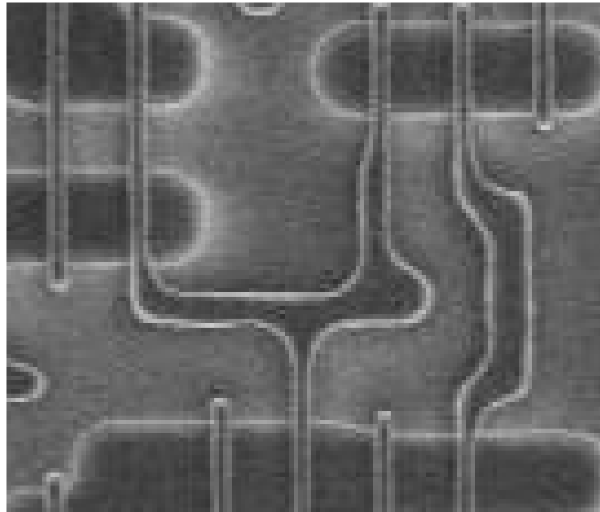
65nm node

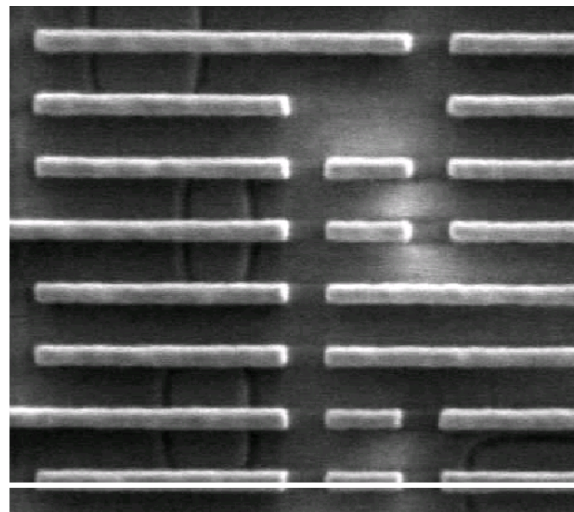


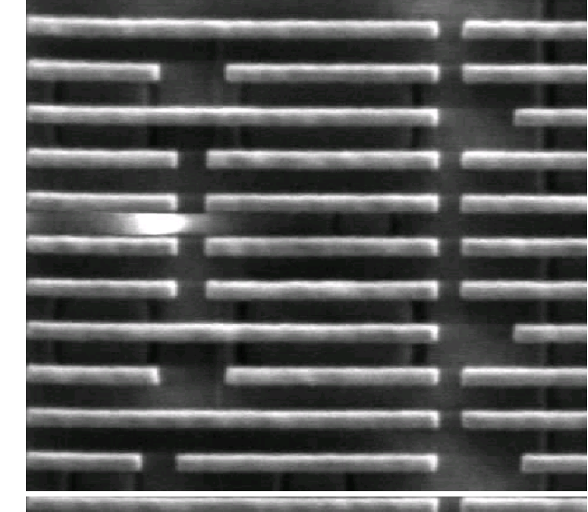
45nm node



32nm node

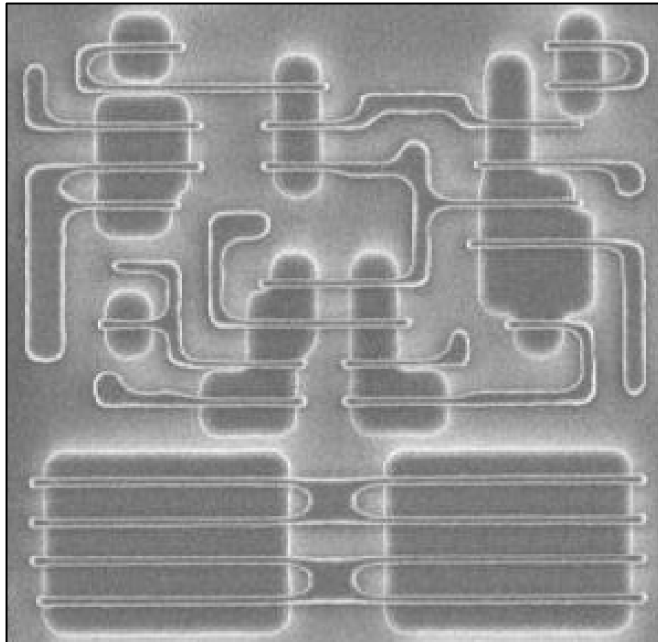






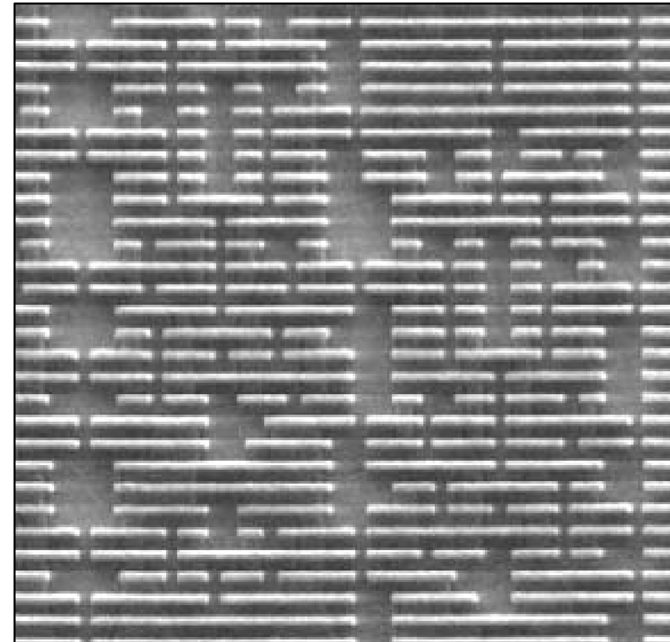
Layout Restrictions 65nm to 32nm

65 nm Layout Style



- Bi-directional features
- Varied gate dimensions
- Varied pitches

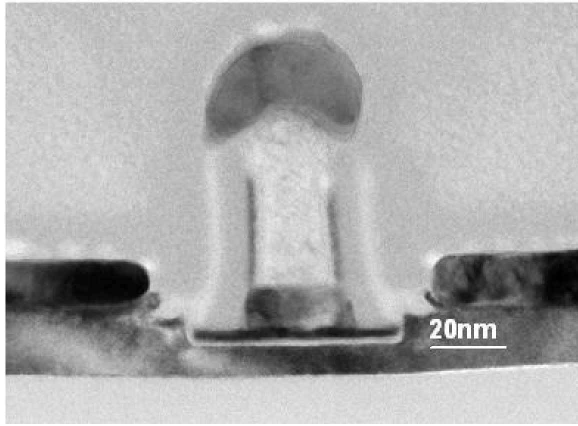
32 nm Layout Style



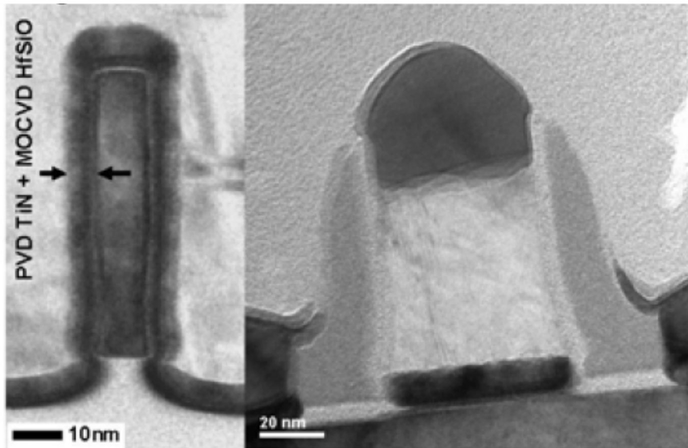
- Uni-directional features
- Uniform gate dimension
- Gridded layout

M. Bohr, ISCC, 2009

Transistor Architecture Enhancements



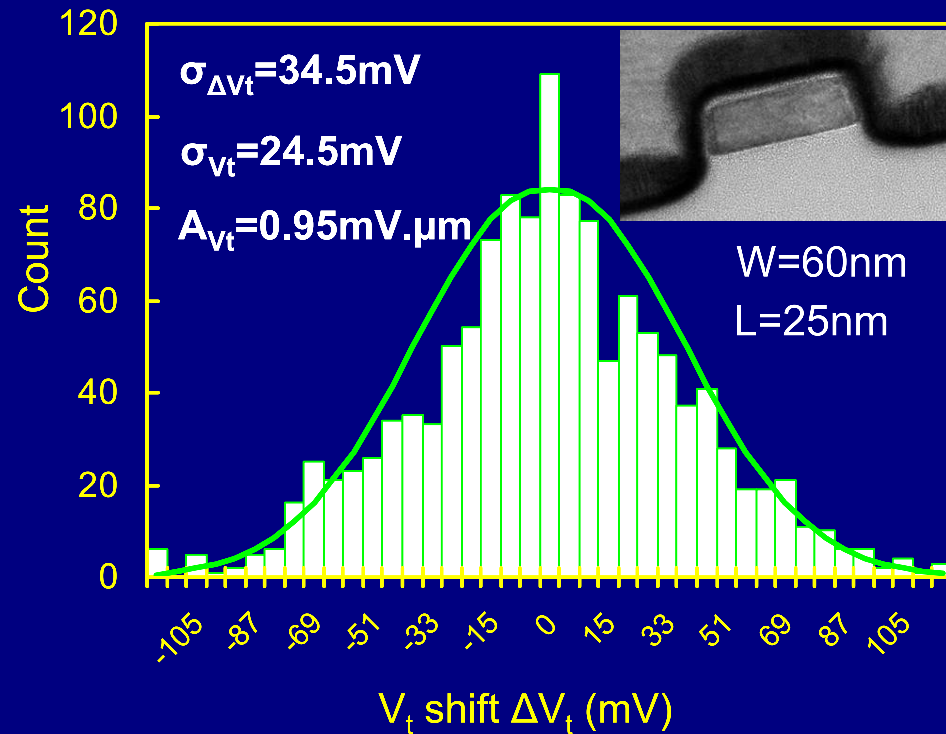
Weber et al. IEDM 2008 pp. 245-248

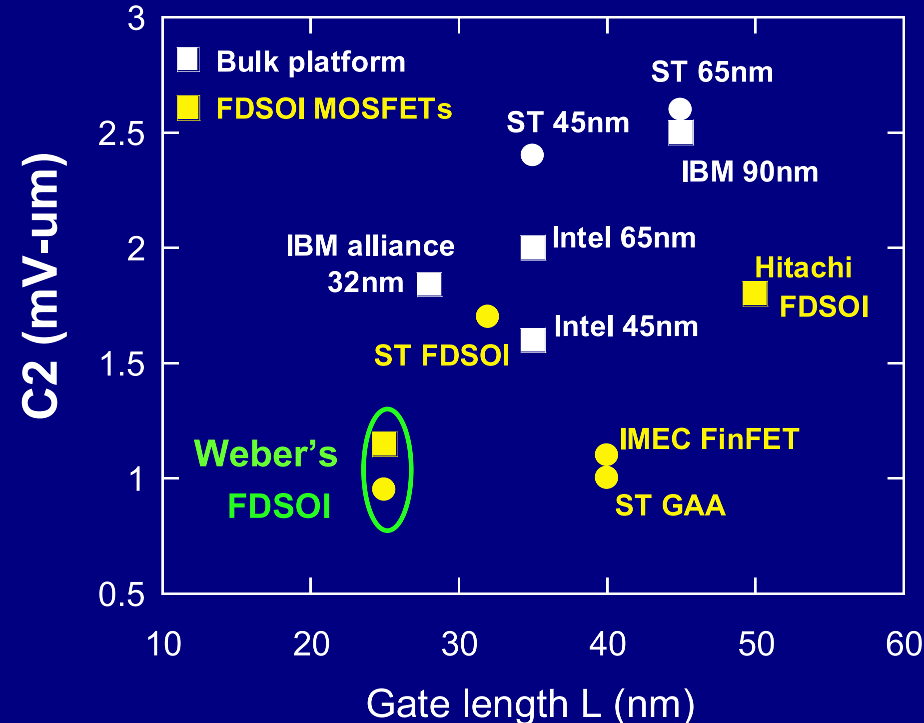


Vellianitis et al. IEDM 2008 pp. 681-683

Fully depleted devices (such as UTB or FinFET) are examples of innovations which permit significant improvement in RDF due to the ability to maintain channel control at lower channel doping.

V_T matching performance





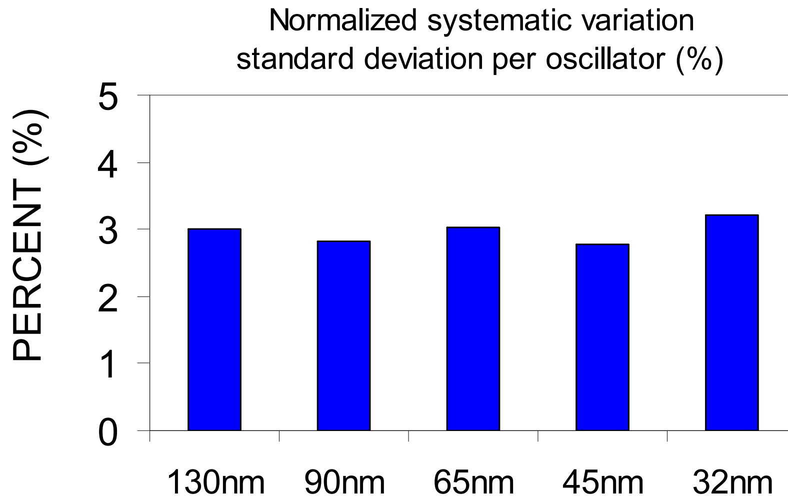
($\sigma_{Vt}=\sigma_{\Delta Vt}/\sqrt{2}$ to compare measurements on pairs and on arrays of transistors in the literature)

Fully depleted devices (such as UTB or FinFET) are examples of innovations which permit significant improvement in RDF due to the ability to maintain channel control at lower channel doping.

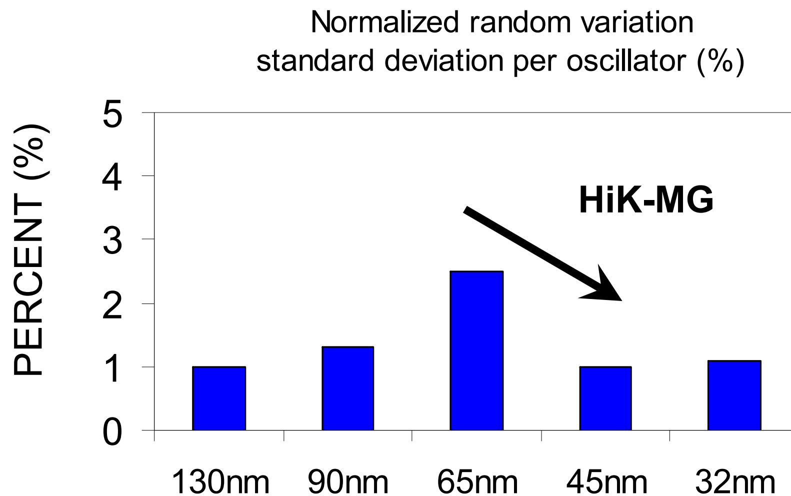
**Weber et al.
IEDM 2008
pp. 245-248**

Closing Thoughts

Random and Systematic Variation Trends



Systematic WIW variation is comparable from one generation to the next



Random WIW variation in 32nm is comparable to 45nm and significantly improved over 65nm and 90nm due to HiK-MG

45 nm: POP CMP Improvement
Overscaling Topography Improvement

CMP Topography

1

0.1

0.01

0.7X improvement

45nm: 2X greater than standard technology scale

J. Steigerwald, IEDM 2008

350 250 180 130 90 65 45

Technology node (nm)

Improvements in polish enabled dramatic improvements in topography variation

Technology Trend
Systematic Gate CD Lithography Variation

LOG (Variation normalized to 130nm)

10

1

0.1

Gate CD variation improvements
with technology scaling

0.7X

WID-total

TOTAL

WIW-total

130nm	90nm	65nm	45nm	32nm	**GENERATION**
336	260	220	160	112.5	**GATE PITCH**

Critical to management of variation is the ability to deliver a 0.7X gate CD variation improvement in each generation enabled by continuous process technology improvements

SRAM Density Scaling

Bitcell Area (μm²)

10.00

1.00

0.10

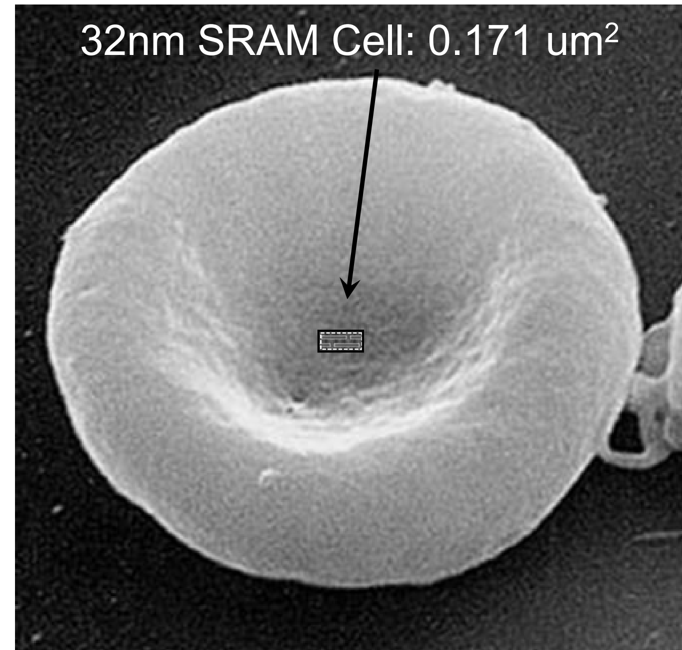
2X bitcell area scaling

90nm 65nm 45nm 32nm

Improved fidelity / uniformity on 32nm vs 90nm

K. Zhang, ISCC, 2009

Q&A



Blood cell: Elec. Mic. Fac. (NCI-Frederick) 2007

For further information on Intel's silicon technology, please visit our Technology & Research page at

www.intel.com/technology