ICCV2009

Chartered

SRAM Variability: Foundry perspectives

- Understanding SRAM Variability and Implications on Bitcell Performance -

Nam Sung Kim, John Sudijono, Hari Balan,Yong Meng Lee,
Young Way Teh, Jeffrey Chee and Liang Choo Hsia

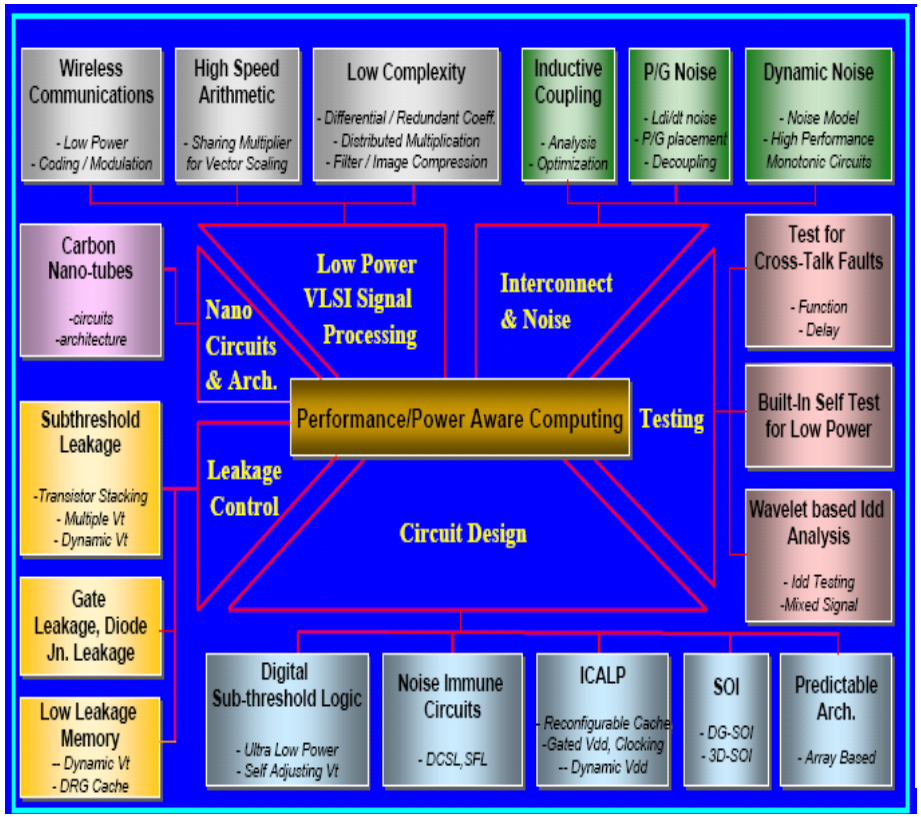
Technology Development
Chartered Semiconductor Manufacturing Ltd.

Outline

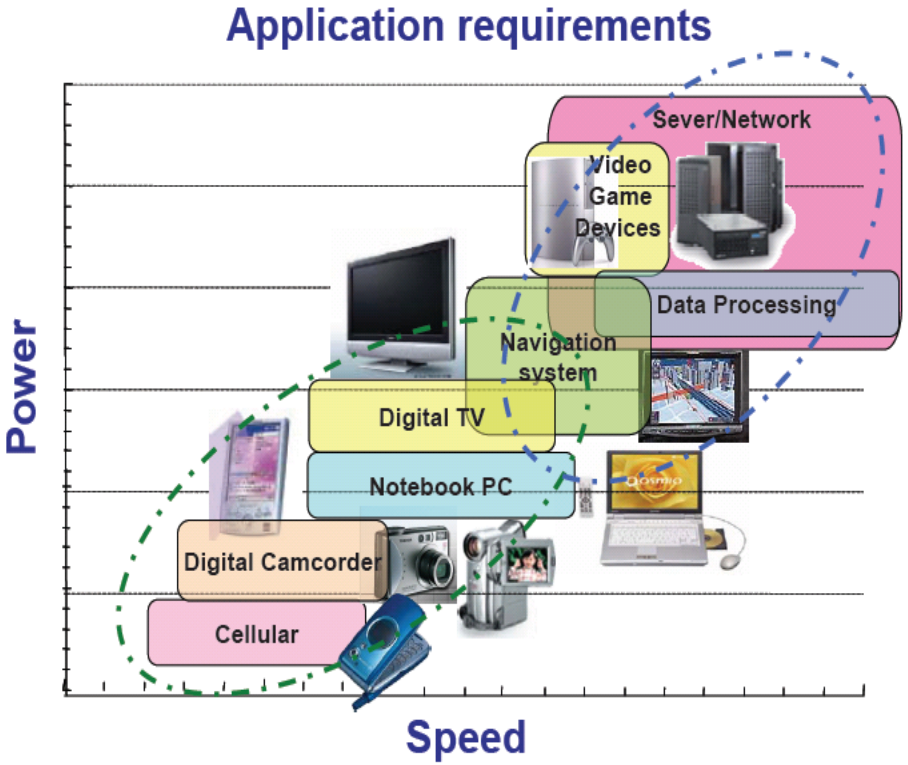
- Introduction
  - SRAM device variability is a critical challenge for further scaling technology
  - One of major roadblocks for advanced CMOS technology development path
- Device variability: SRAM variability is so critical!
  - Some historical review - critical sources of device variations
  - Predominant contributors to SRAM variability
- Discussion
  - Brief review of Sigma study for Vmin limited SRAM yield prediction
    - Vmin LY ~Sigma from min(ADM, WRM)
    - SRAM device variability impact on Vmin margin
  - Success examples of a few practical mitigation techniques in 45nm(poly/SiON) & 32nm(MG/HK) technologies
    - True benefits of MG/HK tech over poly/SiON tech
  - Some guidelines of SRAM variability managements with critical challenges for current & next generation technology development
- Summary

Chartered

LP and HP technology development is being focused on research floor for time-to-market



- Courtesy of K. Roy/IEDM2008

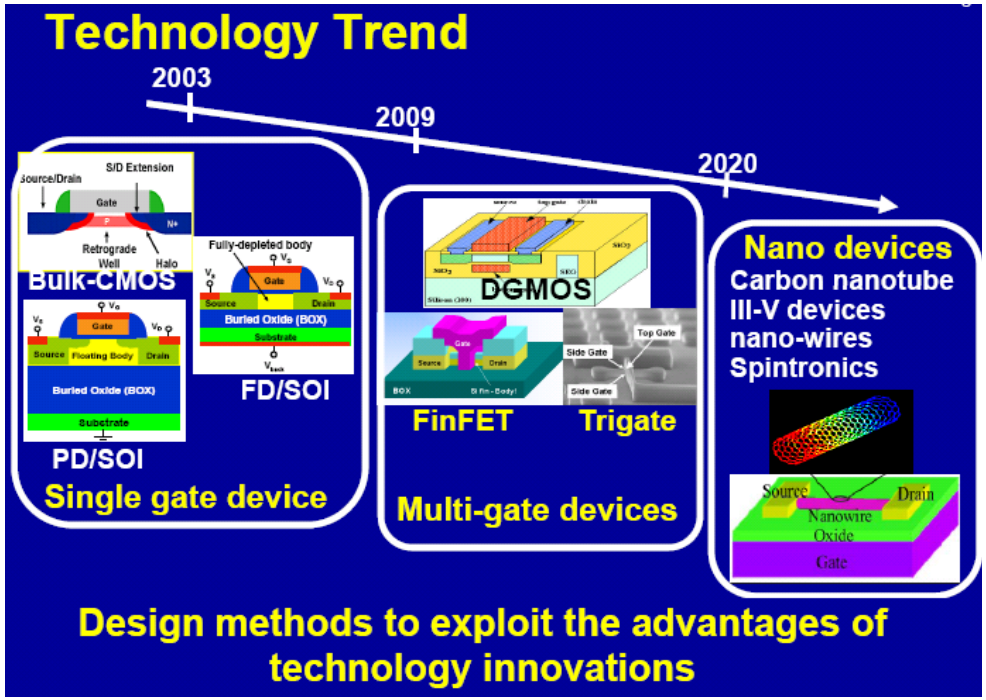


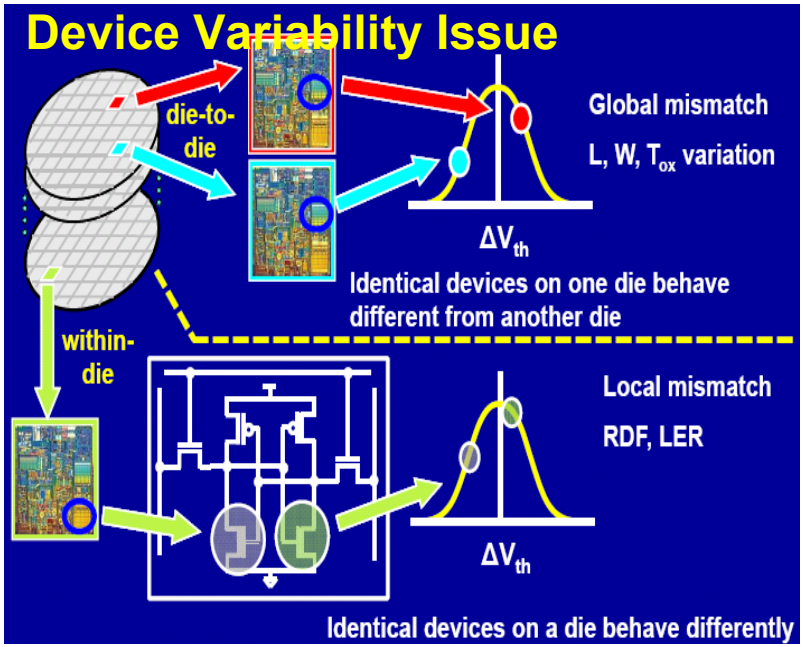
- Courtesy of Koji Miyamoto/IEDM2006

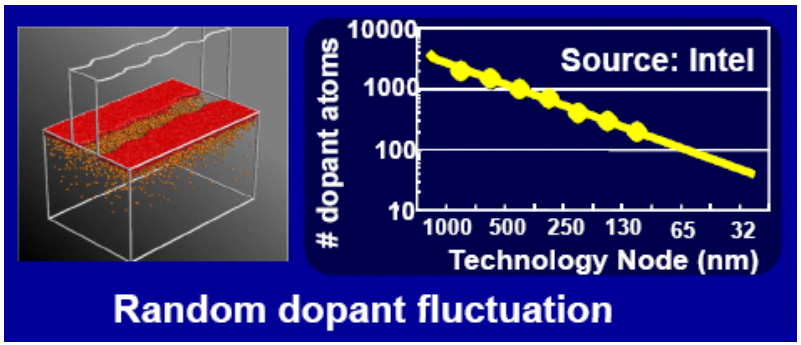
➔ *Today, aggressive technology scaling is being accelerated to meet various and dynamic market requirements*

Chartered

Variability – one of big challengers; roadblock for further technology scaling







- Very innovate technologies with feature size scaling are being developed in the current research floor
- SRAM device variability is a critical challenge for further scaling technology and one of major roadblocks for advanced CMOS technology development path

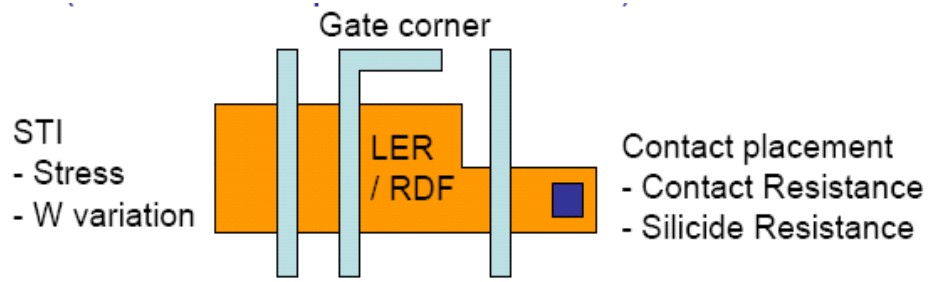
-Courtesy of K. Roy/IEDM2008 Short course

Chartered

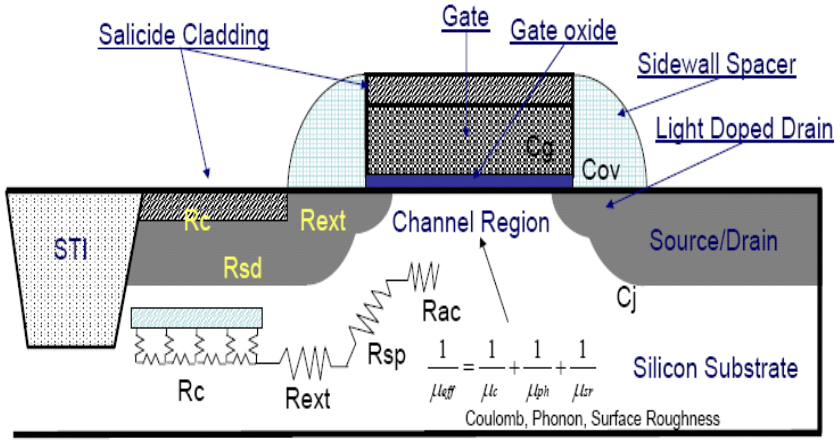
Historical overview: Critical sources of Variation

- **Systematic variations**
  - Process and design interaction
    - Lithography, RIE, RTA…with pattern layout and density
    - Dopant inter-diffusion btw PD & PU devices due to shared Poly line
  - New technology adds to variability, such as stress techniques, HK/MG complex gate stack
    - Increase in layout pattern dependent effects

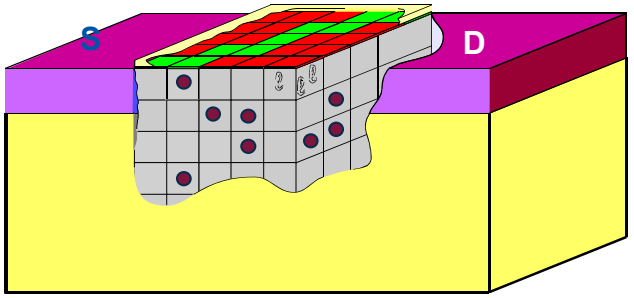
- **Random variations**
  - Fluctuation budget beyond process and device control
    - LER(Line edge roughness)
    - RDF(Random dopant fluctuation)
    - Poly grain size



Intra die, within wafer, wafer to wafer and lot to lot variations.



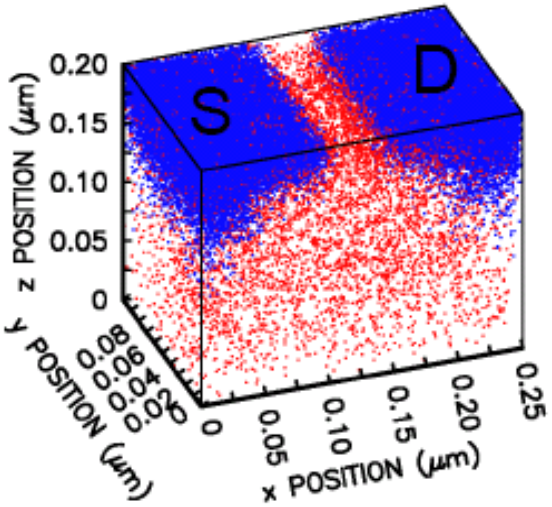
- Source: Koji Miyamoto/IEDM2006

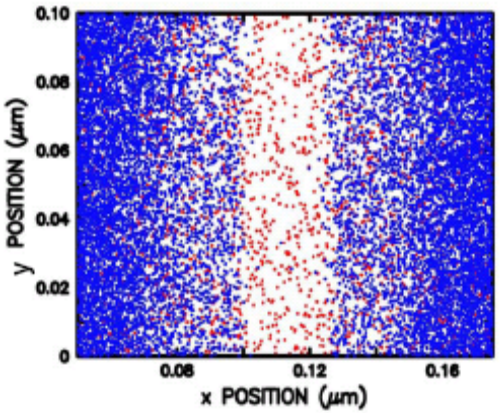


- Source: J. Meindl et al, ISSCC 1997

Chartered

Critical sources of Variation: RDF





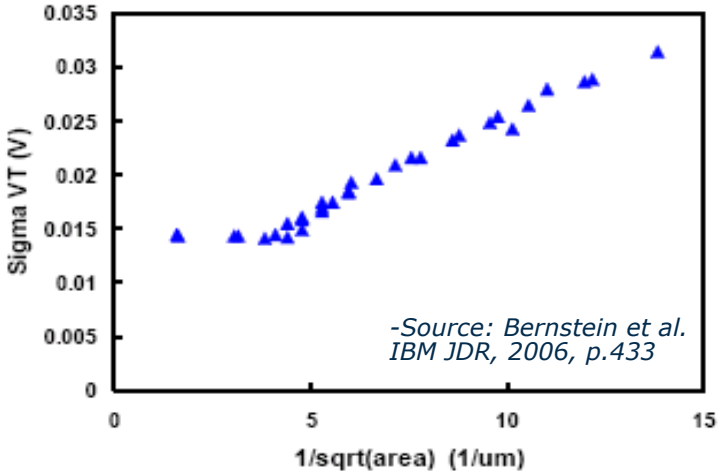
< Example: a 50 nm channel MOSFET with halos>

-Source: Bernstein et al. IBM JDR, 2006, p.433

Empirical fit to theoretical results accounting for both random dopant number and placement:

$$\sigma_{V_T}(RDF) = 3.2\times10^{-8}\left(\frac{t_{ox}N_{dop}^{0.4}}{\sqrt{L_{eff}W_{eff}}}\right)[V]$$

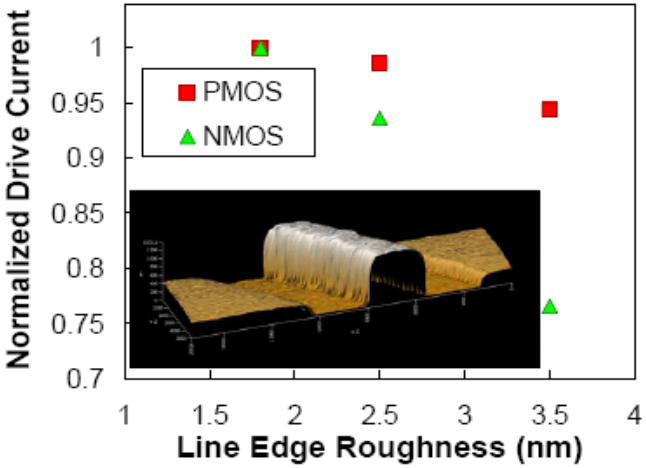
Source: Asenov et al. IEEE T-ED 2003, p.1837



- Only a few hundred atoms in channel set Vt in small devices
- Subject to Poisson statistics, number uncertainty ~ $N^{0.5}$
- Placement uncertainly also increases Vt fluctuations

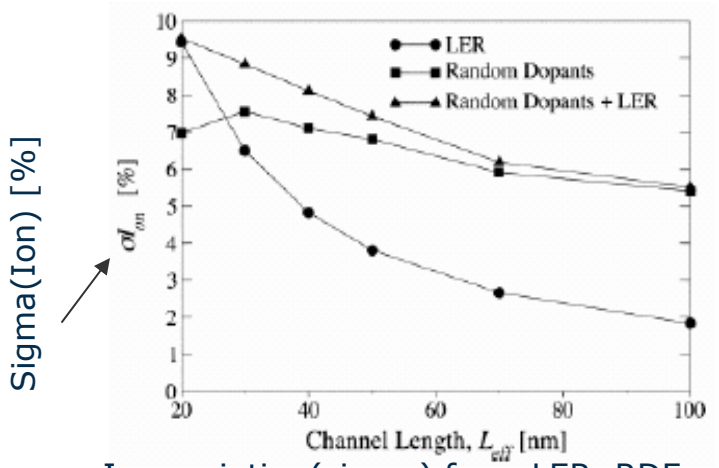
Chartered

Critical sources of Variation: LER & RTA temp



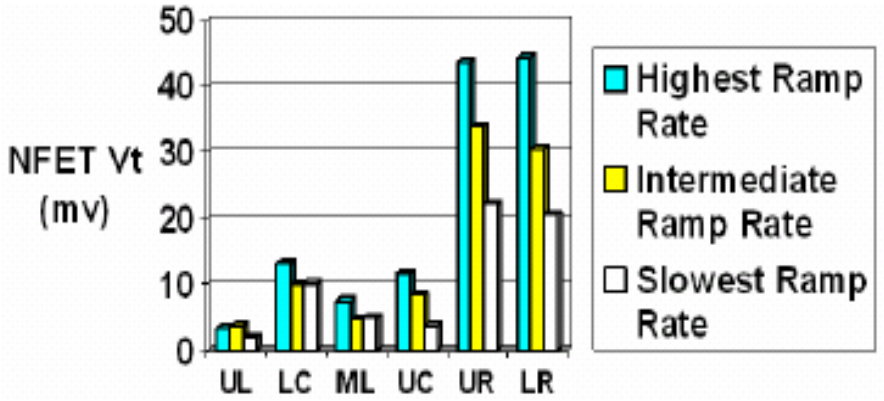
- Idrive current variation@fixed Ioff with different LER(Insert shows AFM profile)

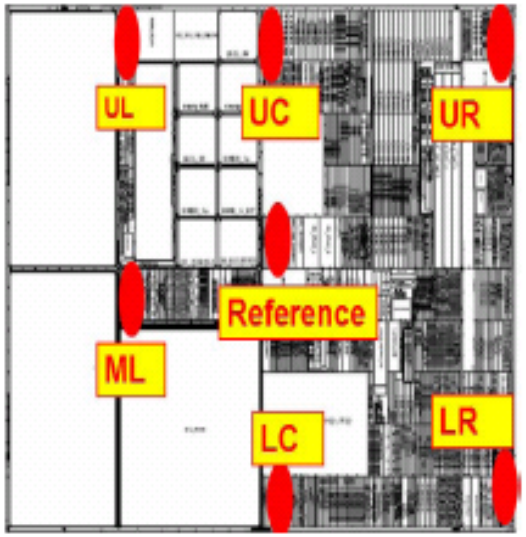
- Source: R.Khamankar et al., VLSI2004



- Ion variation(sigma) from LER, RDF, RDF+LER vs Leff

- Source: Asen Asenov et al.,IEEE TRANSACTIONS MAY 2003





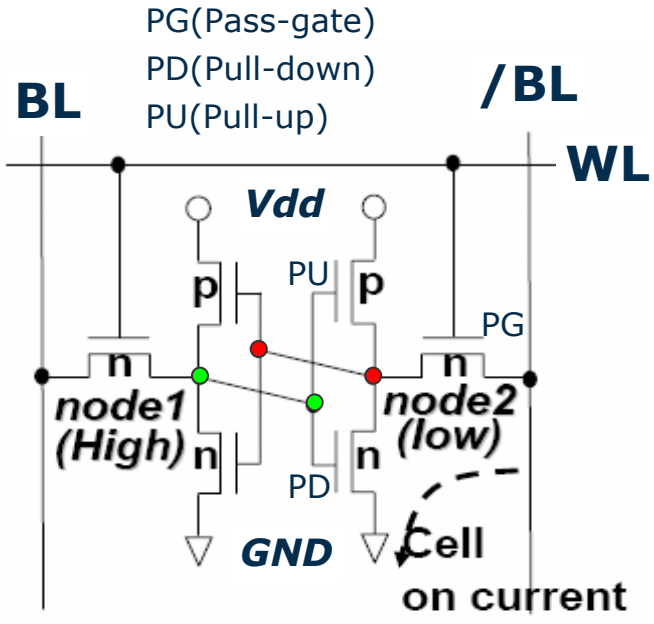
- RTA-Driven Intra-Die Variations

- Source: I. Ahsan et al., VLSI2006

Chartered

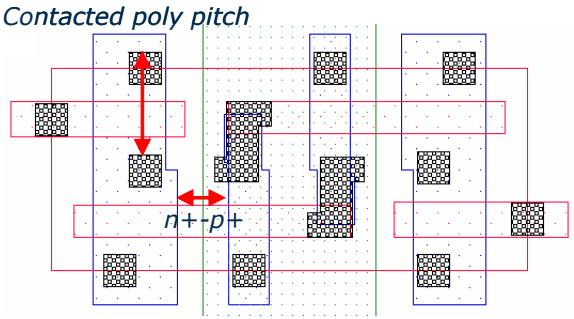
SRAM Devices – PG/PD/PU

- Basically the same as CMOS logic devices
- Aggressive ground rules are often considered for cell size reduction (e.g. tighter p+ to n+ spacing & contacted poly pitch size) as technology is scaled down
- Comparing to logic devices, in general, higher Vt may be applied to increase SNM (Static Noise Margin) as well as off-leakage control
- Basic SRAM cell operation is indicated in the table, which is well-known

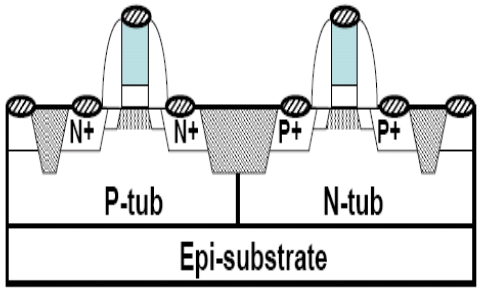


Data	node1	node2
0	low	high
1	high	low

Cell operation	BL	/BL	WL
Read	Vdd	Vdd	Vdd
Write0	GND	Vdd	Vdd
Write1	Vdd	GND	Vdd



<Conventional 6T cell layout>



<X-section view>

Chartered

SRAM Challenges with device scaling

- Design margin decrease with Vdd reduction
- Device variability increase
- Leakages increase

➔ SRAM suffers all the issues of CMOS scaling ahead of logic devices!

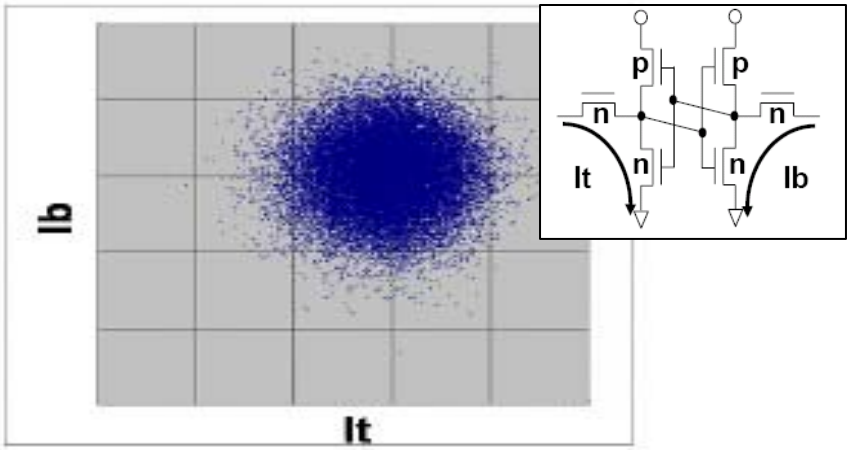
- With shrinking device geometries and increasing device variability, it is becoming increasingly important to consider the impact of variability on the SRAM characteristics during the design phase

- All key SRAM performance metrics: read fails, write fails, retention fails, static noise margins, etc., are sensitive to device variation

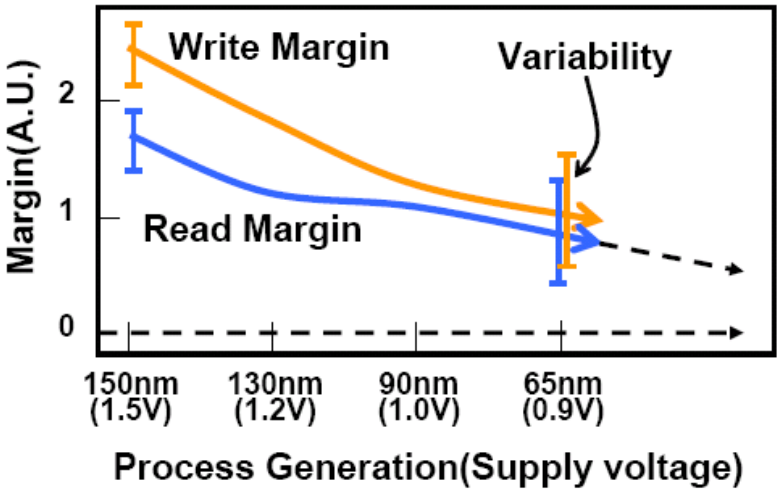
SRAM Variability is a big challenger for competitive SRAM Vmin requirement!

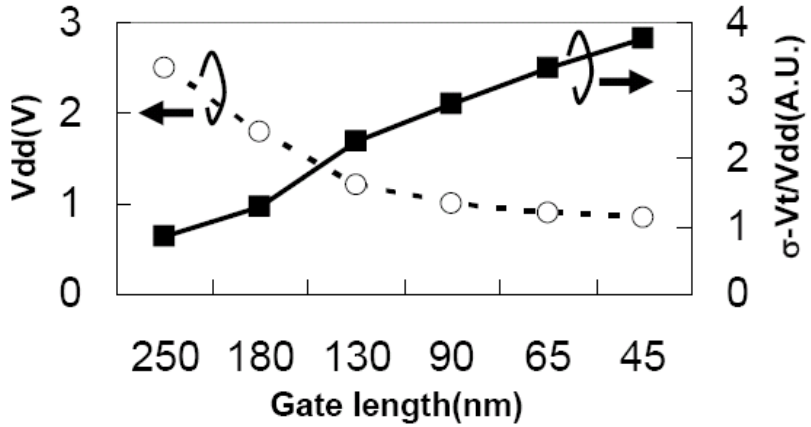
Chartered

SRAM Device Variability



No clear correlation between adjacent cell Trs.





As SRAM device scales down,

- Random variability occupies larger part of total variability
- No clear correlation between adjacent SRAM devices, resulting in cell functionality margin degradation

-Source: Yasushi Yamagata/IEDM'05

Chartered

Foundry Perspective on Vmin(Vdd_min) Limited SRAM Yield prediction

- There is a growing demand for low Vdd operation together with feature size scaling
  - Low Vdd operation for SRAM becomes more challenging as bitcell scales
  - Yield impact at low Vdd due to reduced read/write margin can be significant
- We need a practical method of predicting Vmin limited Yield
  - Vmin visibility needed to manage yield risk
  - Redundancy or larger bitcell

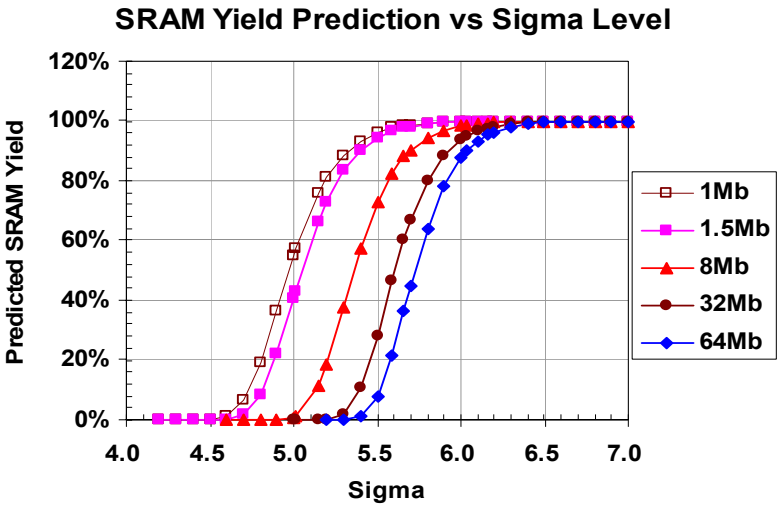
➔ Sigma study is being used by capturing the local variation including Vt-mismatch and global variation in order to estimate access-disturb-margin(ADM) & write-margin(WRM)

Components of process variation

- Local variation, within chip variation including mismatch
- Global variation, chip mean variation

- Within wafer variation
- Wafer to wafer variation
- Lot to lot variation

$$\sigma_{total}^{2} = \sigma_{global}^{2} + \sigma_{local}^{2}$$



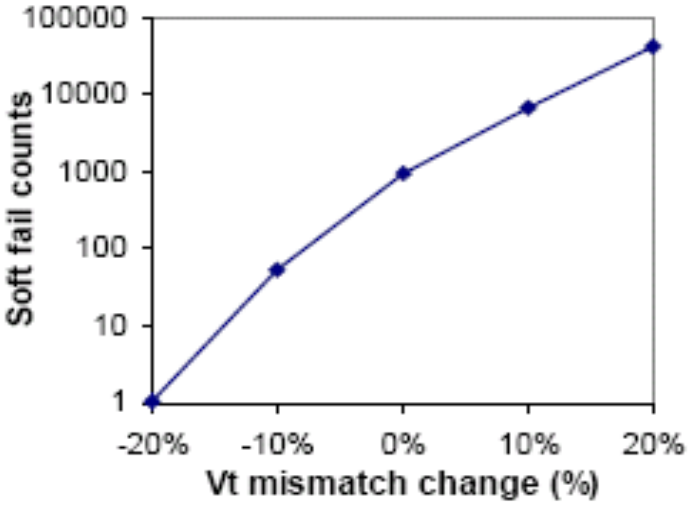
- Courtesy of SRAM team/IBM Alliance Partner

➔ *Vmin Limited Yield is correlated with Sigma ~min sigma(ADM, WRM)*

Chartered

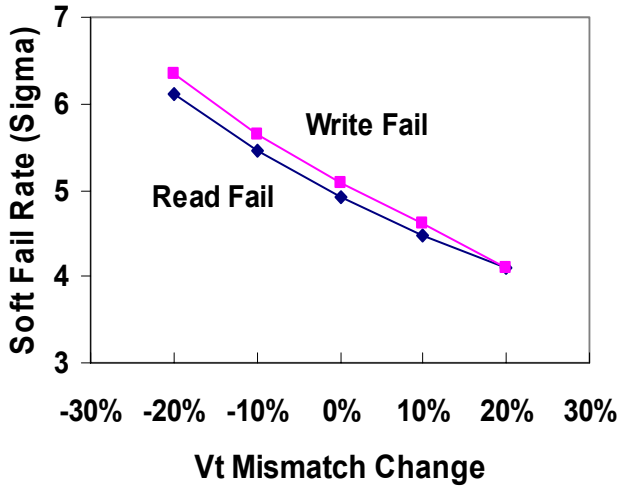
Sigma study for Vmin limited SRAM yield prediction - ADM/WRM dependency

- It illustrates SRAM soft fail rate as a function of the change of device Vt mismatch. Soft fail rate comprises read stability fail rate and write fail rate, which are shown in terms of sigma. For example, a 5.2σ corresponds to one fail among 5Mbit cells. This methodology is adopted as cell design criteria to statistically determine minimum SRAM operating voltage



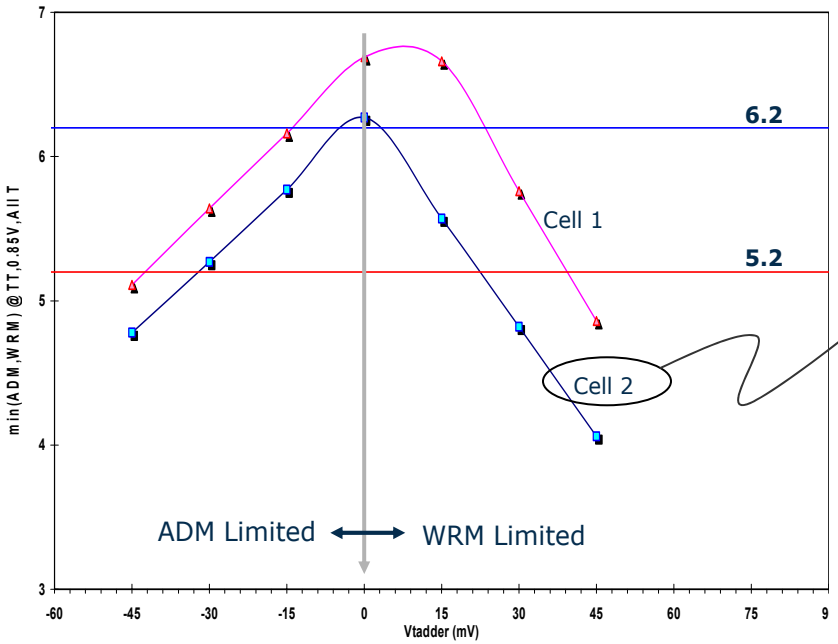
-Source: H. S. Yang, et al., IEDM2008

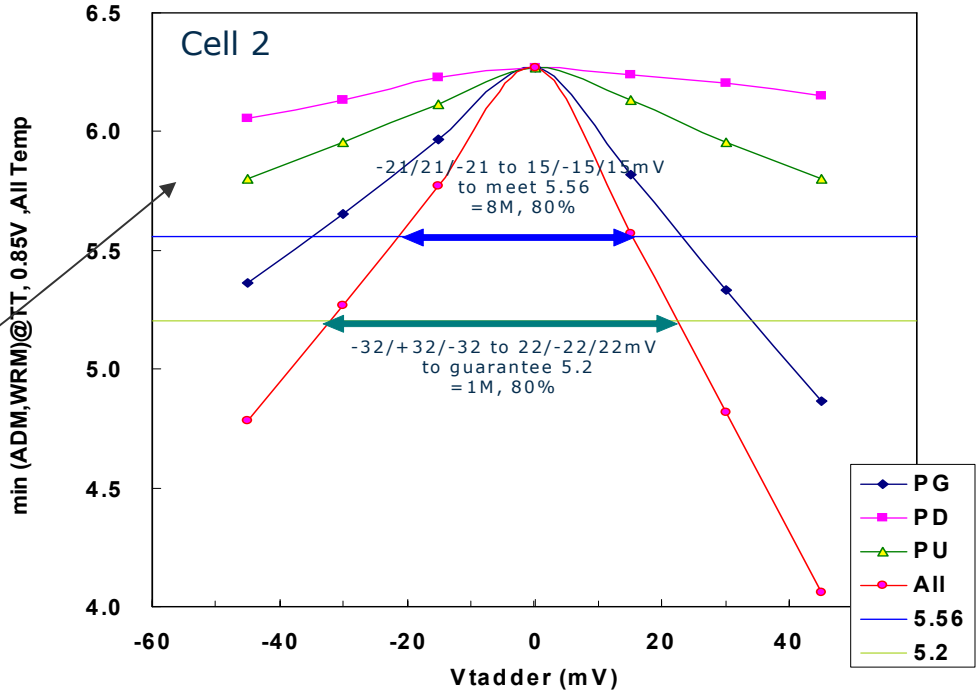
- Soft fails include access disturb fails and write fails
- Soft fails have been overtaking hard fail rates for the past few technology generations
- SRAM soft fails increase with degraded device mismatch



Chartered

SRAM Guaranteed Zone in terms of Vth Variation



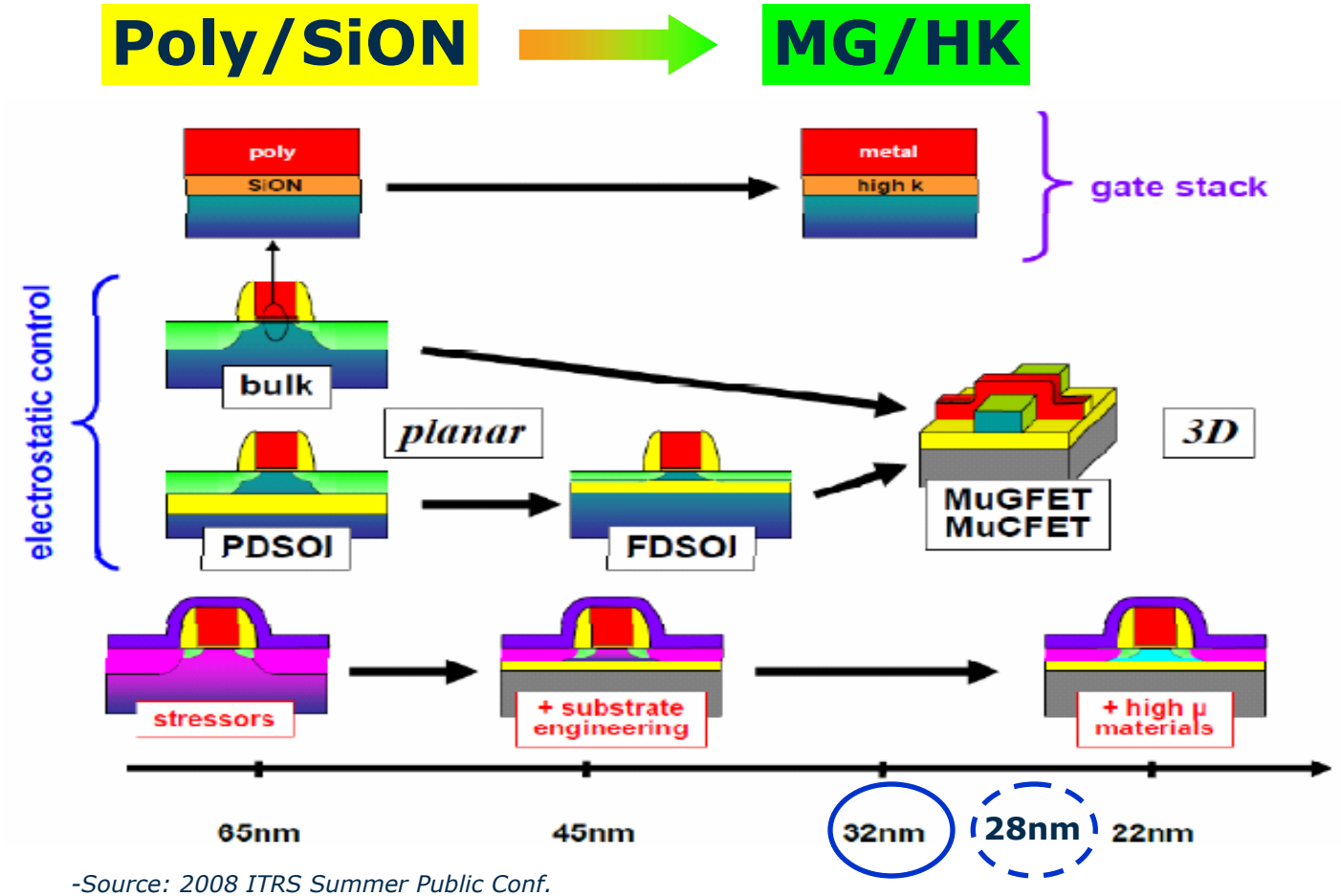


***Sigma ~min sigma(ADM, WRM)@TT, 0.85V, all temp ranges(-40C~125C)*

- Courtesy of K-W Kim/IBM Alliance Partner

- PG (Pass Gate) control is critical to get high Vmin yield
- To hit 5.2 sigma, it requires to control Idsat within ~±10% range

Structure and Technology Innovation Trend (ITRS 2007)

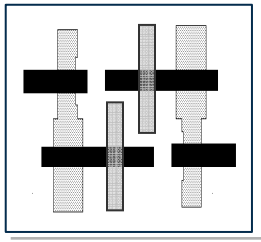


-Source: 2008 ITRS Summer Public Conf.

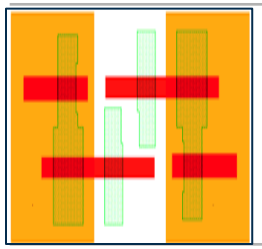
➔ *Currently, technology development focus is moving from poly/SiON to MG/HK technology*

Chartered

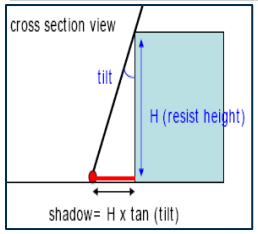
Success examples of a few practical mitigation techniques from 45nm(Poly/SiON) to 32nm(MG/HK)



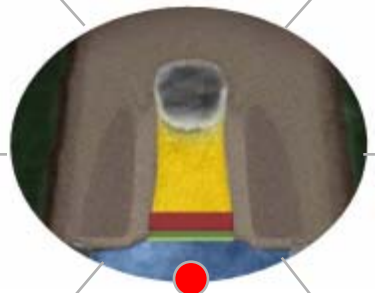
Active and Poly CD optimization of SRAM ✓

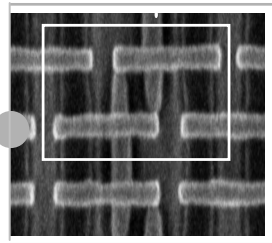


Poly pre-doping boundary control in SRAM ✓

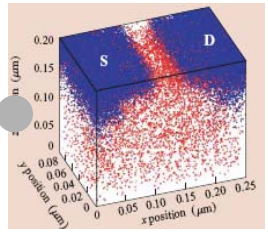


Implant shadowing-free technique ✓

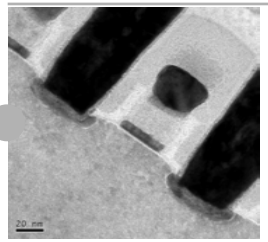




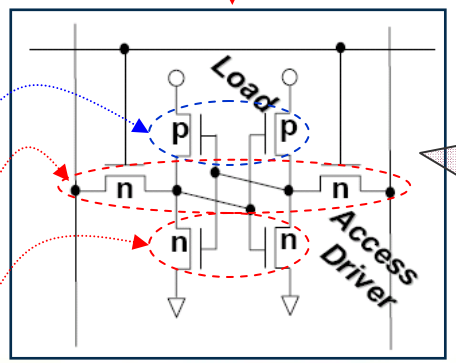
Double Exposure of Gate line ✓



Maximization of Co-ii effect ✓



MG/HK Technology Benefits ✓



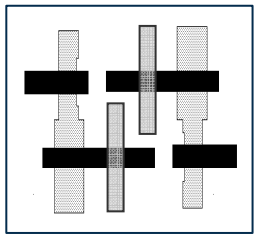
- *σVT (PU)*
- *σVT (PG)*
- *σVT (PD)*

SRAM Variability management to meet competitive Vmin requirement

Chartered

Active(W) & Poly(L) CD optimization in SRAM

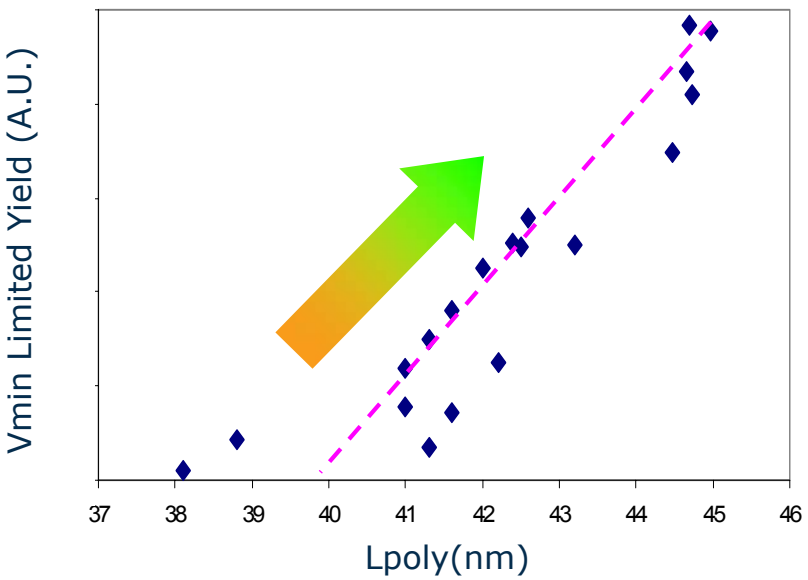
45nm(poly/SiON)



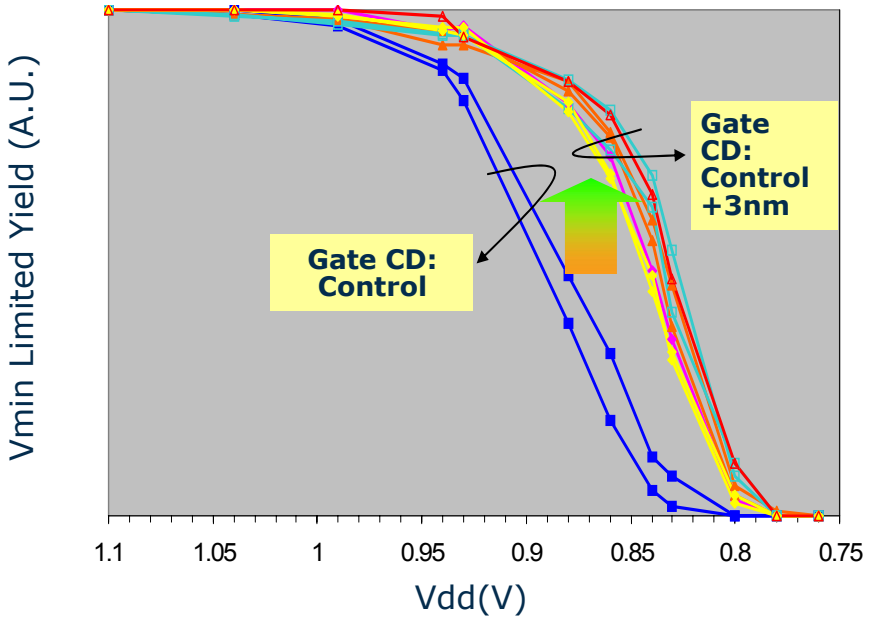
$$\sigma_{\Delta VT} = \frac{A_{VT}}{\sqrt{W.L}}$$

$\sigma_{\Delta VT}$: increases as device area(W*L) decreases. Hence, W & L CD optimization is critical at a given technology by taking into account – Iread/Istby/Gate-Contact process margin(~Vmax yield)

45nm SRAM Vmin Limited Yied@32Mbits vs Lpoly



45nm SRAM Vmin Limited Yied@32Mbits vs Vdd

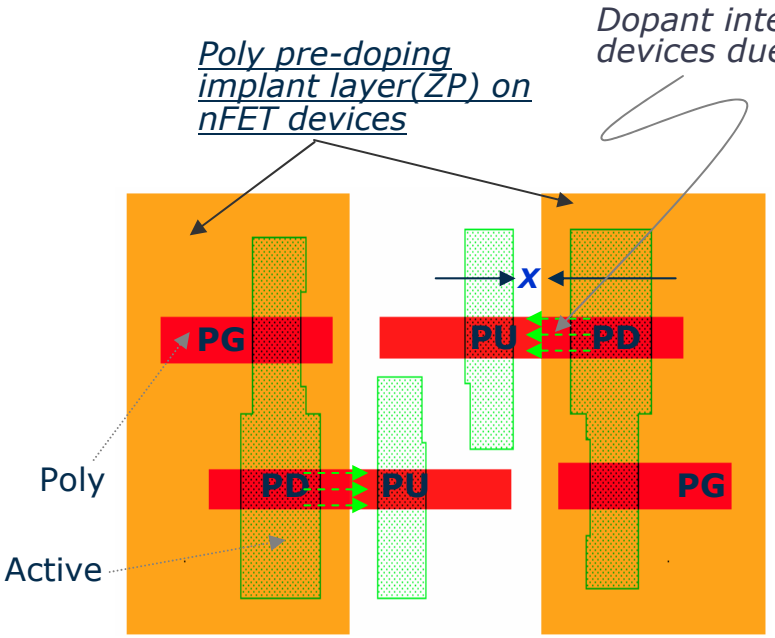


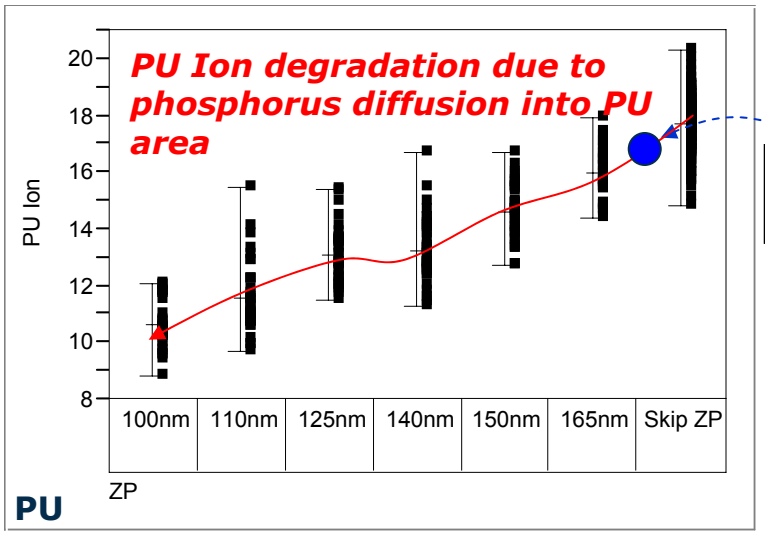
- Vmin limited yield as function of Lpoly

Chartered

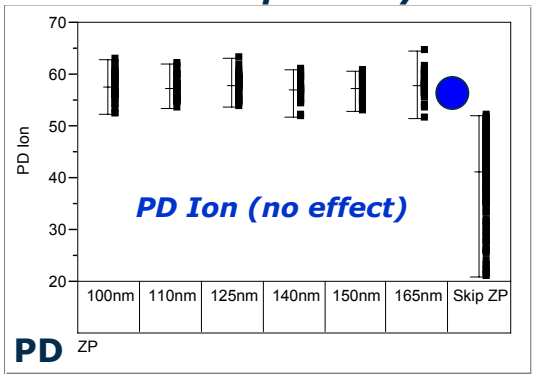
Poly pre-doping boundary control in SRAM(1/2)

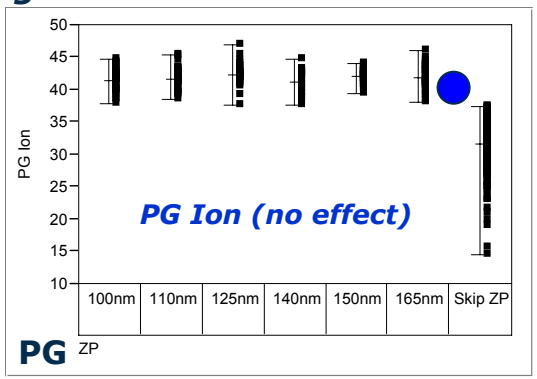
45nm(poly/SiON)





X = Implant layer Sizing





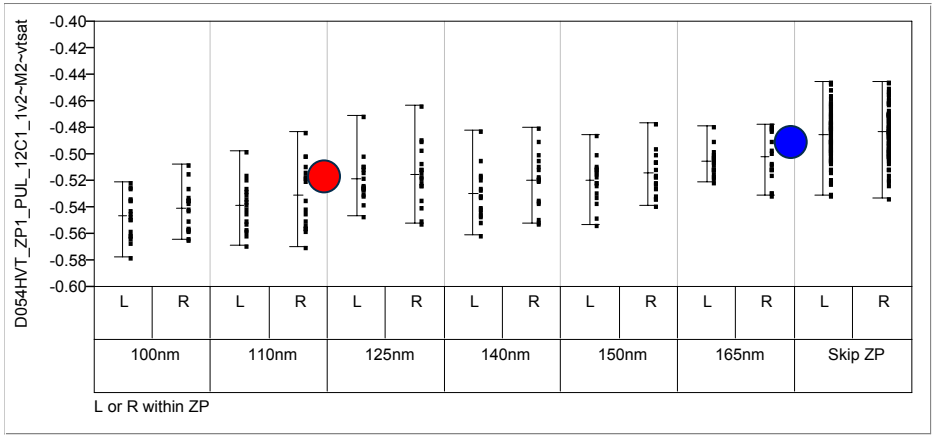
- PU Ion starts to degrade below x=165nm sizing
- nFET(PG,PD) pre-doping implant boundary optimizing is critical to eliminate PU Ion degradation with variability

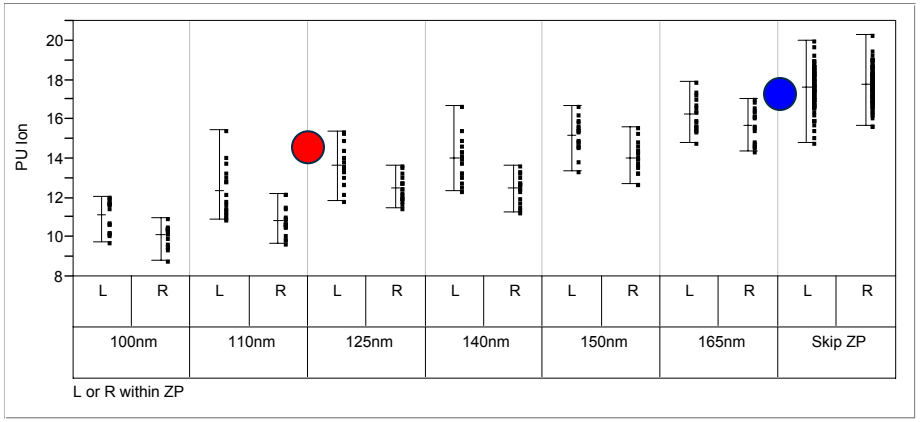
Chartered

Poly pre-doping boundary control in SRAM(2/2)

45nm(poly/SiON)

PU Vtsat & Ion vs pre-doping boundary sizing (X)

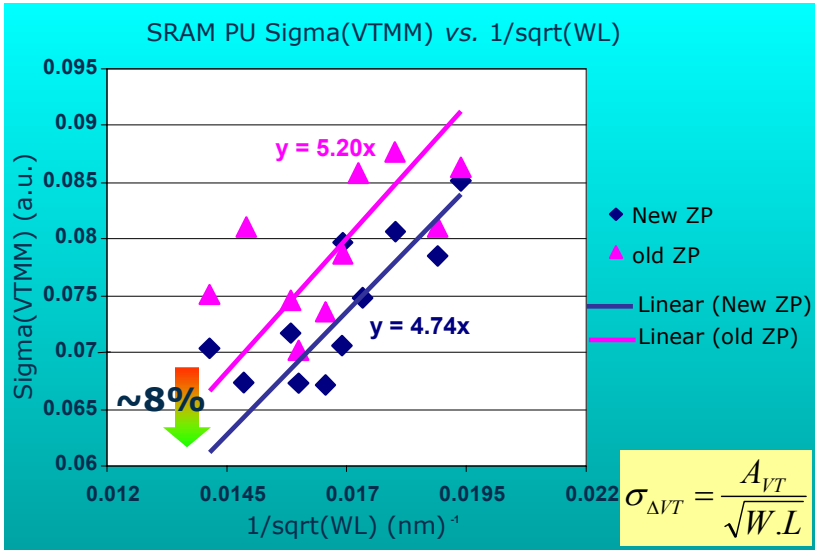




: Old ZP

: New ZP

PU Vt-mismatch(sigma): Old vs new ZP sizing (X)

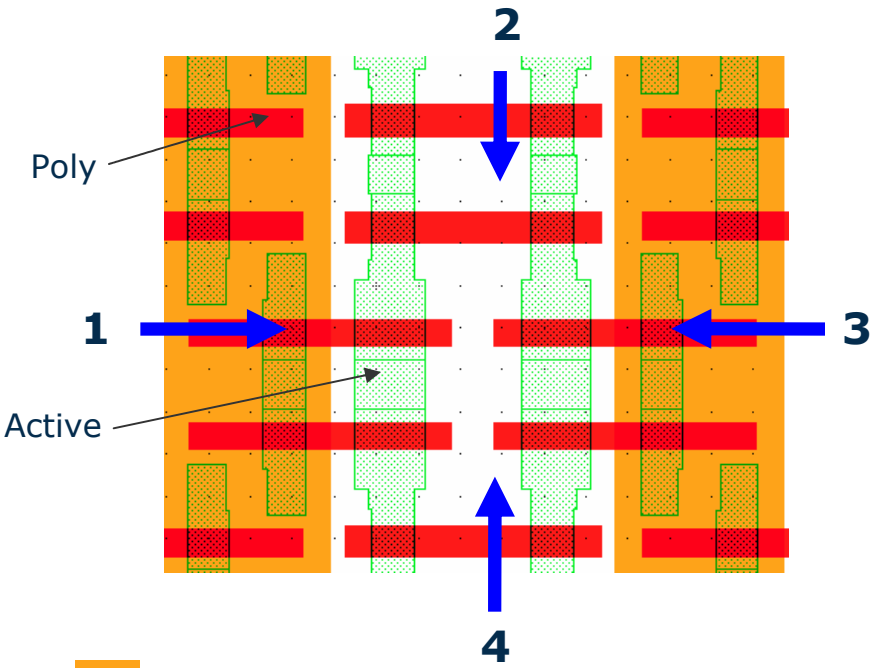


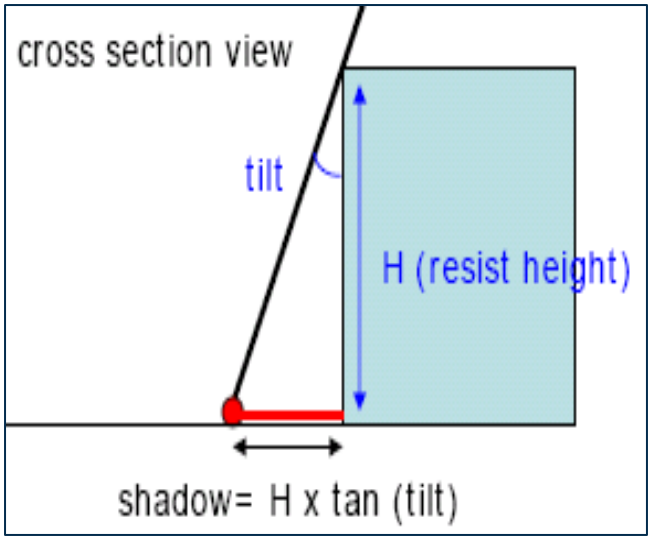
- Pelgrom plot illustrates the improvement:
  - Using ZP sizing optimization (old→new), PU Vt-mismatch improvement is obtained by ~8%(Avt: 5.2 → 4.7)

Chartered

Implant shadowing-free approach in SRAM

45nm(poly/SiON)





:nFET(PG & PD) Halo implant block layer

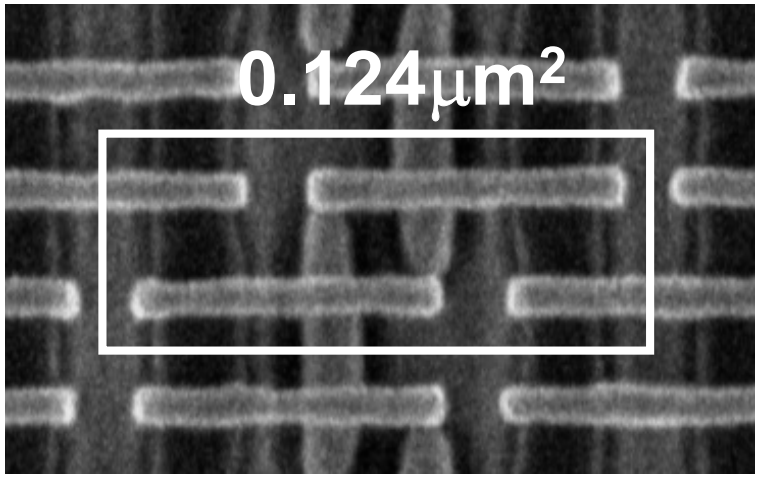
- From the four directions of the Quad implants, the implants from the direction 2 & 4 are the most important for SRAM device and not shadowed even at relatively higher degree tilt → Need to optimize angle and implant dose
- The direction 1 & 3, however, may generate shadowing, which is one of SRAM device variability sources
- In 45nm technology and below, dual-halo implant schemes (direction 2 & 4 only) are widely adopted

Chartered

Double Exposure of Poly Gate line in SRAM

32nm(MG/HK)

Double-Patterned Gate Features



-Source: H. S. Yang, et al., IEDM2008

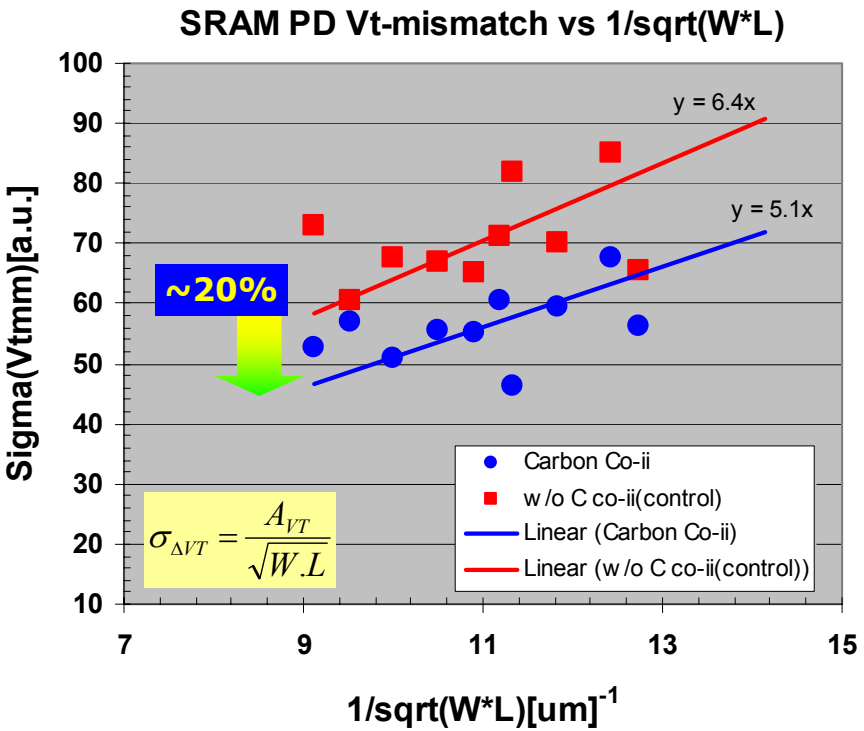
Square endcaps are defined using double-patterning tech

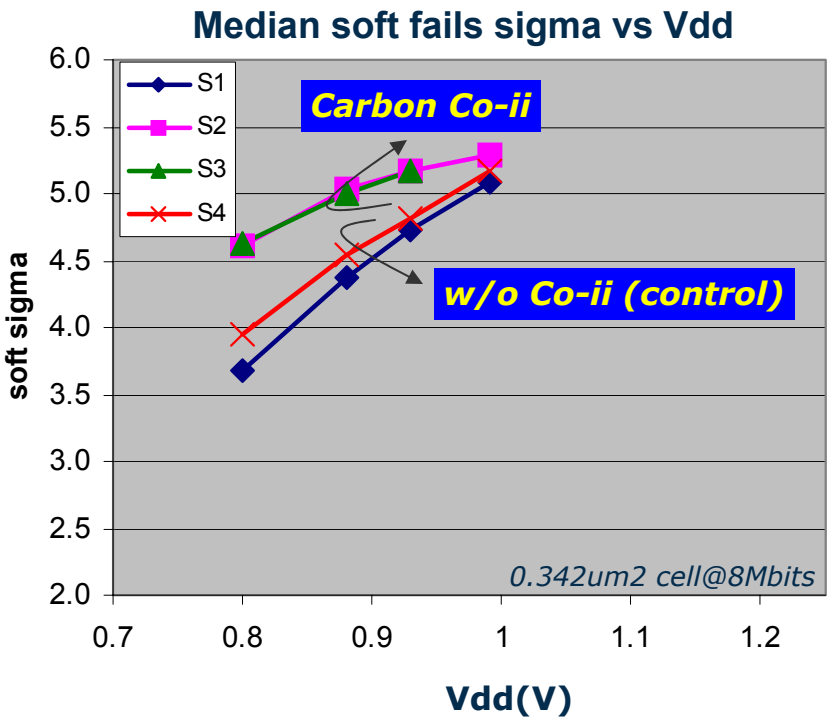
- Double-patterning with a "cut-mask" approach has been demonstrated to achieve good gate tip-to-tip spacing (50nm) in order to eliminate the systematic variation associated with "dogbone" and "icicle" endcaps
- Patterning fidelity demonstrated for a ultra-dense cell with 0.124mm2 cell size

Chartered

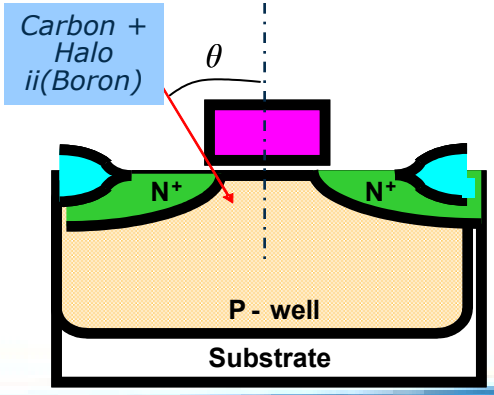
Co-imp effect – Carbon co-ii at Halo area w/ optimum halo profile control

45nm(poly/SiON)





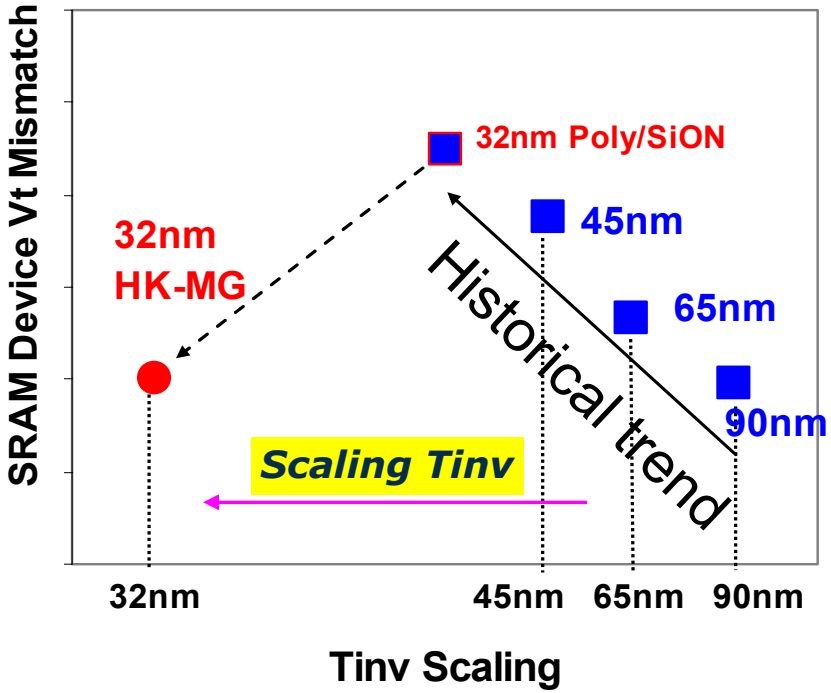
- Carbon co-ii followed by Halo(pocket) w/ angle implanted to get SC control
- Carbon forms carbon-interstitial clusters → reduce boron diffusivity



- Carbon co-ii effect leads to significant SRAM nFET(PD,PG) Vt-mismatch reduction(~20%) due to boron retardation at Halo region
- It contributes to sigma improvement from 4.5 to 5.0

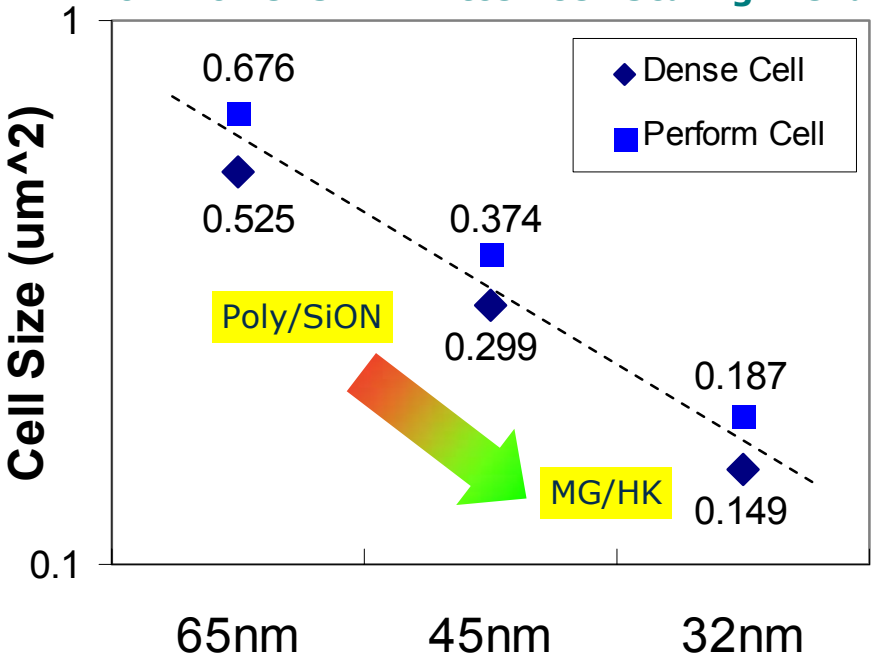
Device-mismatch trend & Scaling Trends in SRAM

Device Mismatch & Scaling Trends w/ High-k



- Random dopant fluctuation and gate leakage limits mismatch reduction
- SRAM scaling with Poly-SiON is unsustainable
- MG/HK dielectric overcomes the Tox-Gate leakage limitations of Poly/SiON, reversing the mismatch trend, allowing the continual bitcell scaling

Low-Power SRAM Bitcell Cell Scaling Trend



- The continual 50% bit cell area scaling trend for both the dense & performance cells from 45nm has been demonstrated with the combination of High-k gate stack & 32nm processes
- Dense cell scales from 0.299 mm2 to 0.149 mm2 from 45nm to 32nm

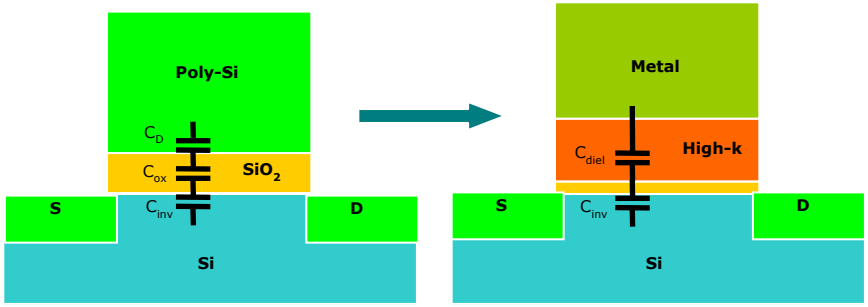
-Source: H. S. Yang, et al., IEDM2008

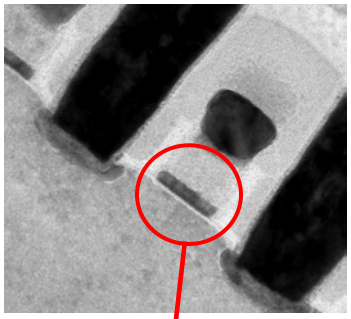
Chartered

Aggressive 32nm(28nm) SRAM Area Scaling enabled by MG/HK

- Demonstration of world's smallest SRAM cell with "gate first" MG/HK in 32nm
- Aggressive SRAM scaling with improved Vmin behavior & leakage
- Extendibility to 22nm

SRAM yield is driven by soft fails at low V_{dd} operation





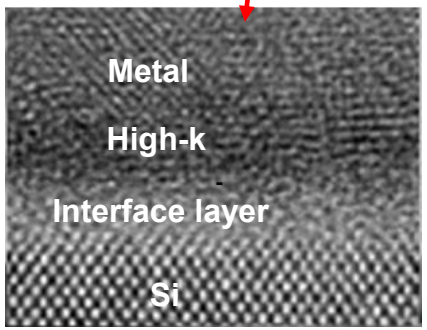
Reduce Vt mismatch to increase Read/Write margins $\sigma^{VT} \sim T_{ox}/(W*L)^{1/2}$

Reduce Tox to improve σ^{VT} ➔ MG/HK is a key advantage

MG/HK approach will allow Ultra-High Density cell for 32nm and beyond

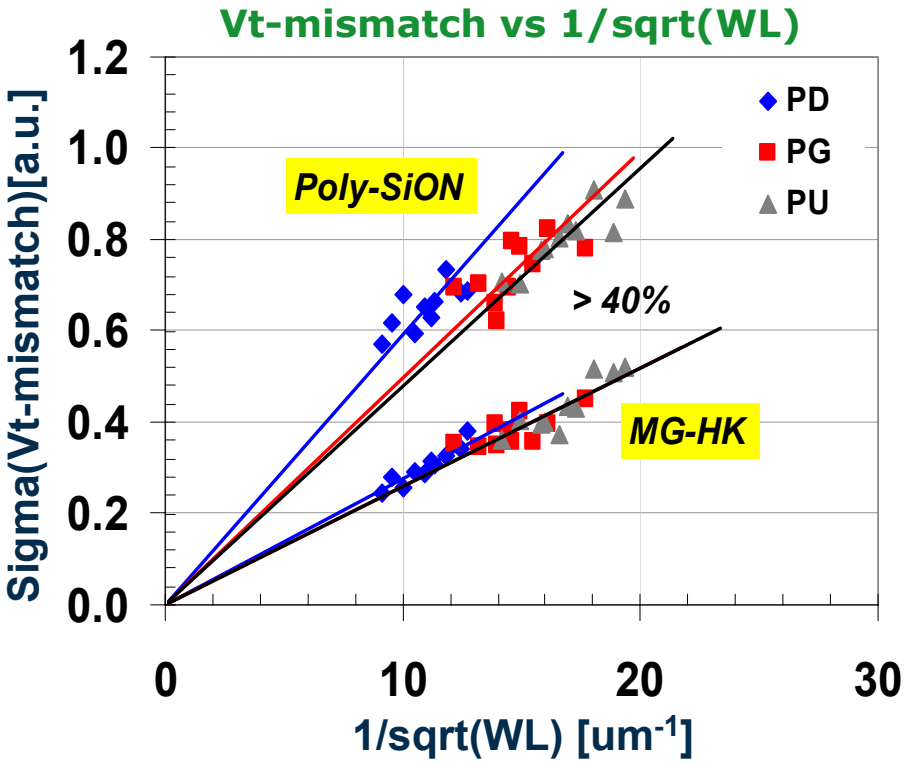
Key elements and function

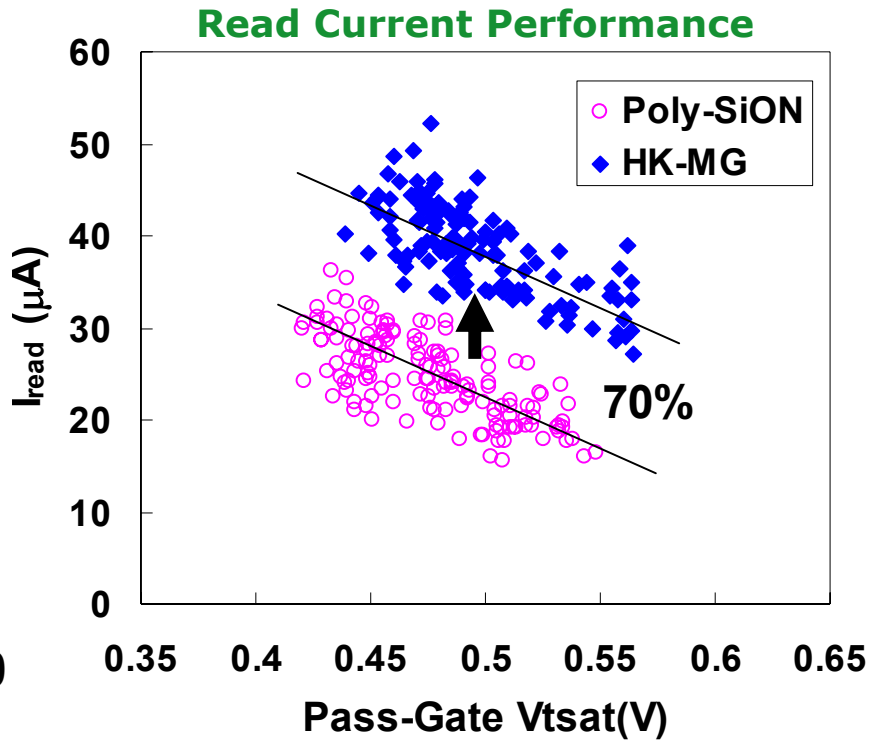
1. SiO2 interlayer: Stability and performance
2. High-k: Gate leakage reduction (Power) & improved EOT scaling (L scaling)
3. Metal: Vt control & eliminate poly-Si depletion



Chartered

SRAM Device Vt-mismatch & Iread performance: Poly/SiON vs MG/HK in 0.299um2 cell





- SRAM device Vt-mismatch improvement by > 40% from MG-HK is observed over Poly-SiON on 0.299um2 cell

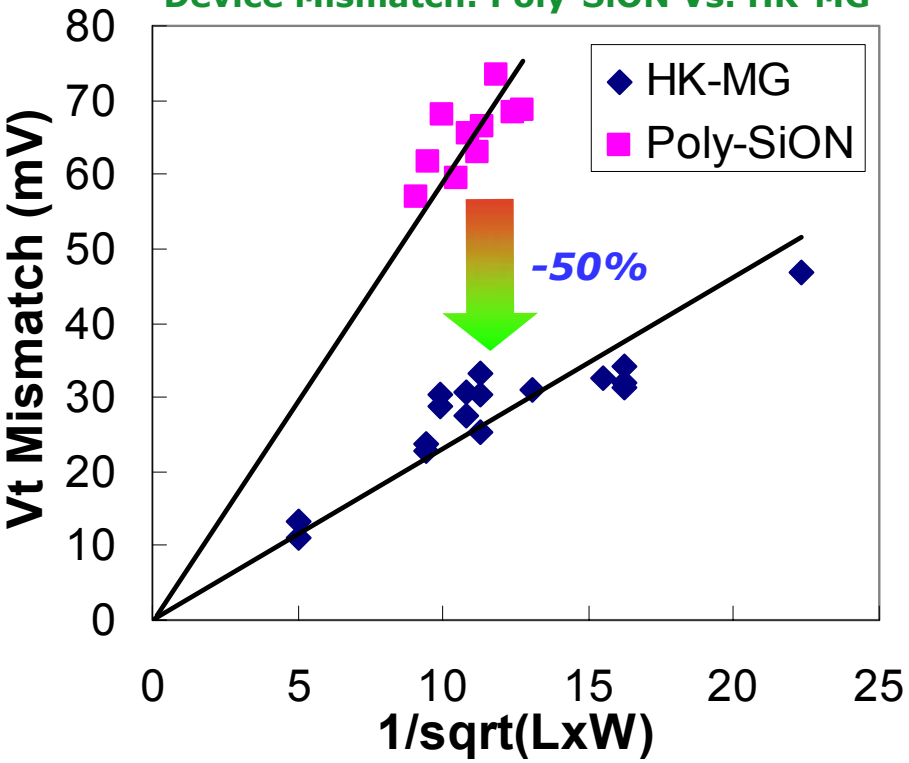
- SRAM read current increases by 70% from HK-MG over Poly-SiON, as demonstrated on 0.299um2 cell

Chartered

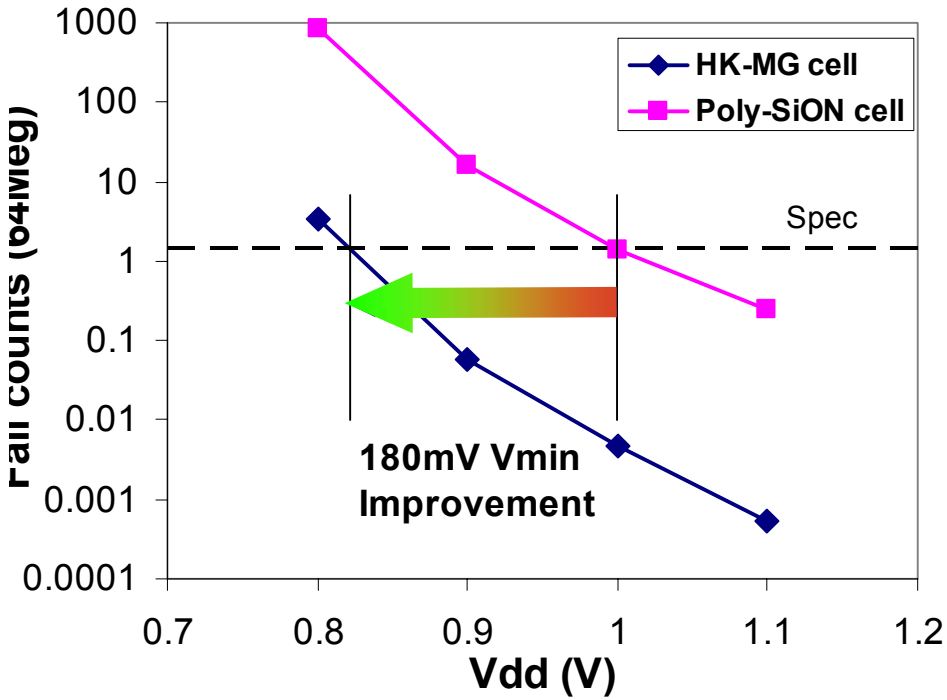
Benefits of MG/HK tech on SRAM Vt-mismatch

32nm(MG/HK)

Device Mismatch: Poly-SiON Vs. HK-MG



Vmin Improvement with 32nm Bit Cell



-Source: H. S. Yang, et al., IEDM2008

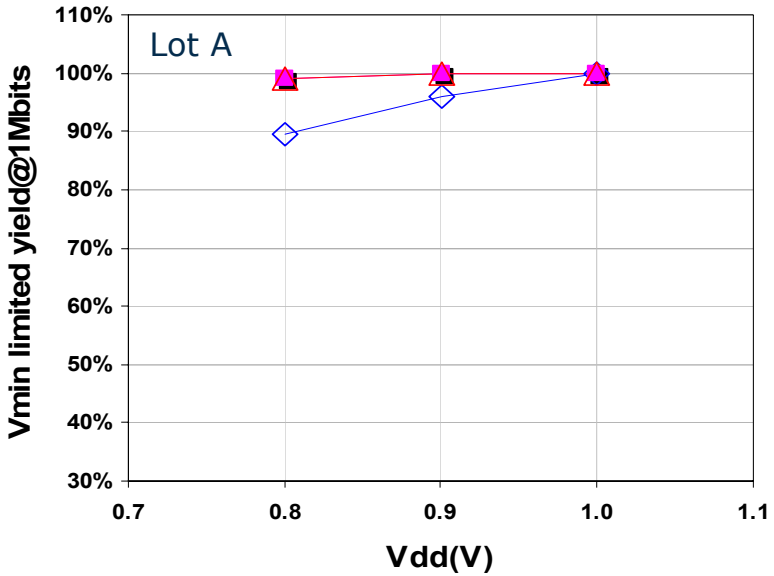
- After optimizing for leakage & performance, 50% Vt mismatch reduction can be realized by using HK-MG over Poly-SiON

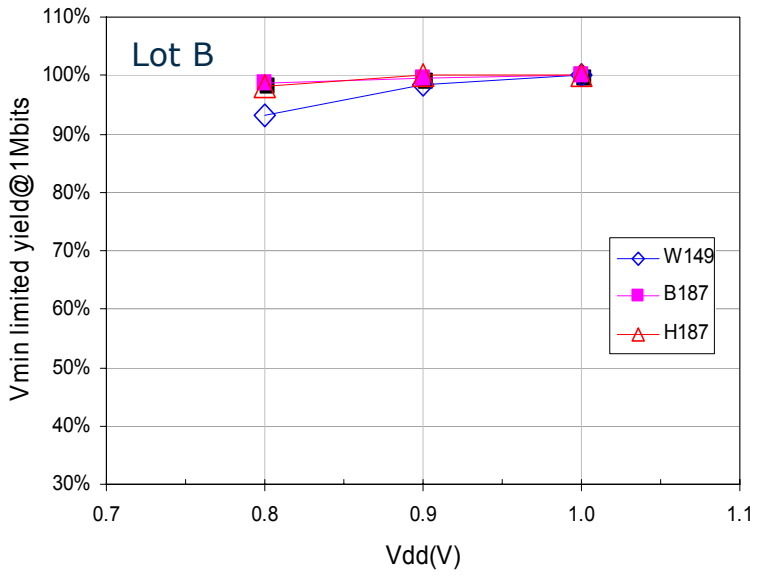
- Simulations show mismatch improvement translates to 180mV Vmin improvement for the 32nm low-power dense cell (0.149um2) with HK-MG
- The Poly-SiON test case is optimized for similar leakage

Chartered

Promising SRAM Vmin limited yield achieved from 32nm MG/HK technology

32nm(MG/HK)





- Vmin limited yield at Vdd=0.8V
  - 89.5% for 0.149um2 cell(W149)
  - 99% for 0.187um2 cell(B187)
  - 99% for 0.187um2 cell(H187, high perf cell@ Regular Vt)

- Vmin limited yield at Vdd=0.8V
  - 93.2% for 0.149um2 cell(W149)
  - 99.8% for 0.187um2 cell(B187)
  - 99.2% for 0.187um2 cell(H187)

➔ Excellent Vmin Performance obtained thanks to MG/HK benefit!

Chartered

Some guidelines for SRAM variability management for current and future technology development

Requirement: Improvement in voltage and temperature tolerance in SRAM characteristics

Critical challenges are ahead!

- Most difficult part of SRAM down-scaling is Vdd scaling down
- Density of on-chip cache SRAM memory is high and thus, Vth cannot be down-scaled too much due to large Isd-leak
- Also, under low Vdd, read- and write margin degrades → Vmin yield drops
- Thus, Vdd down-scaling is more severe in SRAM than logic part of the circuits

- ***From various mitigation techniques, manufacturing-friendly solutions are still required and attractive***

  - Process cost: mask counts, process complexity and process compatibility

Cell & process optimization approach to mitigate device variability

- Improved bit cell optimization
  - nFET(PG & PD) /pFET(PU) centering and Beta/Gamma balancing
  - Maximize the utilization of co-implant(Carbon/Nitrogen/ Germanium /Flourine,..) to minimize TED or dopant diffusivity from the subsequent thermal budgets
- Minimize device fluctuation by limiting device-geometry scaling in SRAM
  - Lg is often designed to be larger than Min. logic Lg → Different Contacted poly pitch btw logic and SRAM
  - Lpoly, Weff, LER optimization – DFM/OPC robustness
- Radom variability controls
  - Delta of Interface traps and local depletion of gate electrode
  - Delta of junction depth
  - Strain and interface roughness
  - Discrete dopant distributions
- Enjoy the true benefits of MG/HK in terms of Vt-mismatch reduction (~Tinv scaling), but still requires careful process optimization to maximize the realistic advantage of MG/HK technology

$$\sigma_{V_T} = \sqrt{\sigma_{V_T}(LER)^2 + \sigma_{V_T}(RDF)^2 + \sigma_{V_T}(gate)^2}$$

Chartered

Summary

- The predominant contributors to SRAM variability in current CMOS devices are briefly reviewed by looking through the historical overview of process variations - critical sources of variation
- Some of key learning for SRAM variability controls was discussed along with detailed characterized results obtained from the application of several mitigation techniques from 45nm(Poly/SiON) to 32nm(MG/HK) technologies
- In particular, the true benefit of MG/HK technology over the conventional poly/SiON technology is highlighted in terms of dramatic Vt-mismatch reduction(~50%) with superior Vmin response in dense SRAM cell
- Profitable utilization of MG/HK technology enables ultra dense SRAM cell process window to be expanded with optimized process controls in 32nm and beyond
- Some guidelines for SRAM variability management for current and future technology development are proposed by underlining the preference of manufacturing-friendly process solutions
- In addition, the accurate and efficient characterization and modeling of SRAM device variability are indispensable for the continuous technology scaling to estimate competitive SRAM Vmin requirement

Chartered

Acknowledgements

- *This work was performed at the Chartered Semiconductor Manufacturing Ltd. in Singapore, as well as a joint development with the ISDA 32nm Bulk alliance partners in IBM East-Fishkill, New York*

- *The authors would also like to thank ISDA 32nm Bulk SRAM & Device team for the support with valuable discussion*

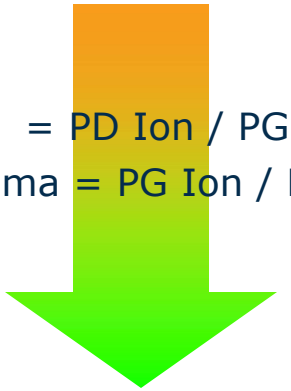
Chartered

Visit us at
www.charteredsemi.com

Backup Foils

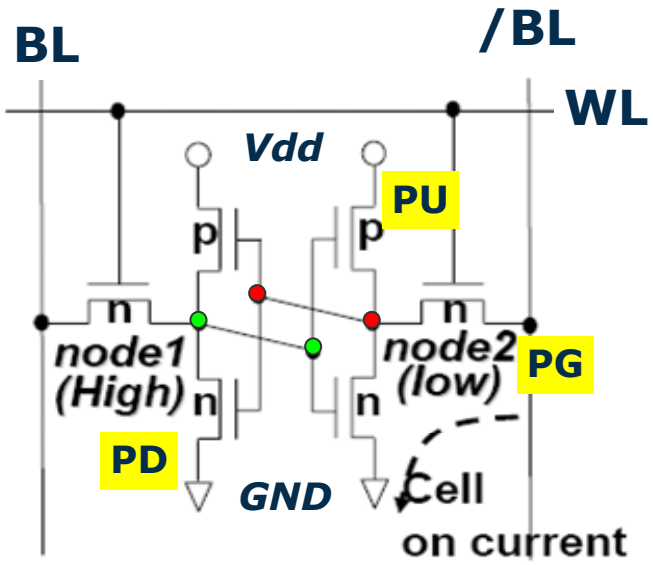
SRAM - Read margin vs Write margin

- Read and Write margin improvement direction is opposite!



- Beta = PD Ion / PG Ion
- Gamma = PG Ion / PU Ion

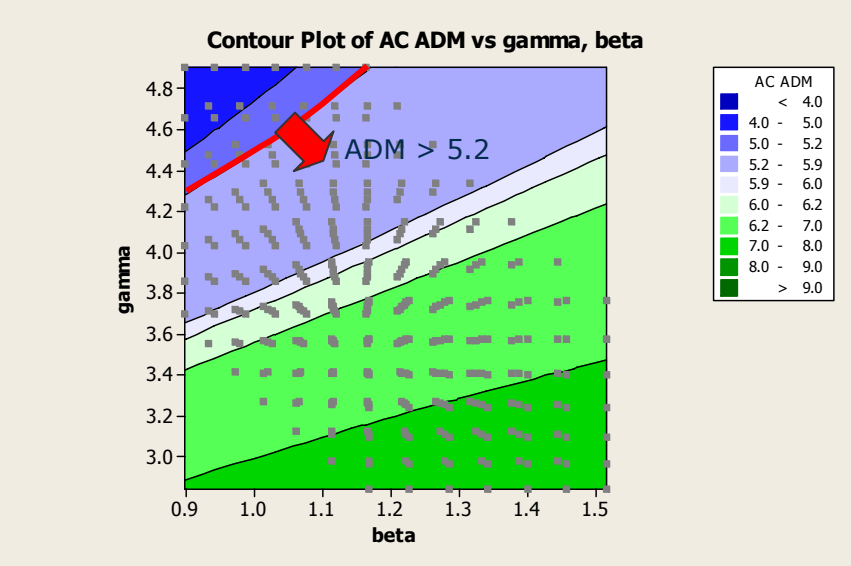
➔ *Beta and Gamma optimum-balancing is critical to maximize cell margin at given technology*

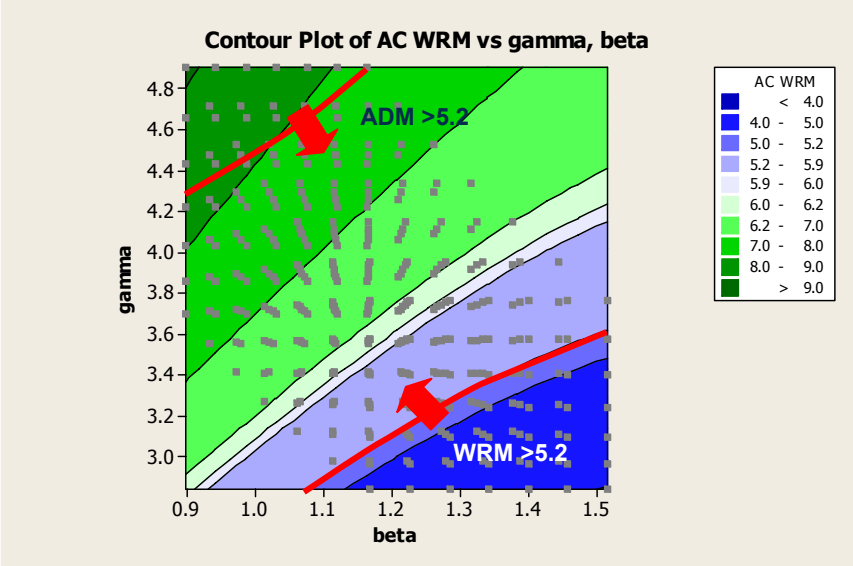


- Read Margin Improvement
  - PU: Vt lower
  - PG: Vt higher
  - PD: Vt higher
- Write Margin Improvement
  - PU: Vt higher
  - PG: Vt lower
  - PD: ---

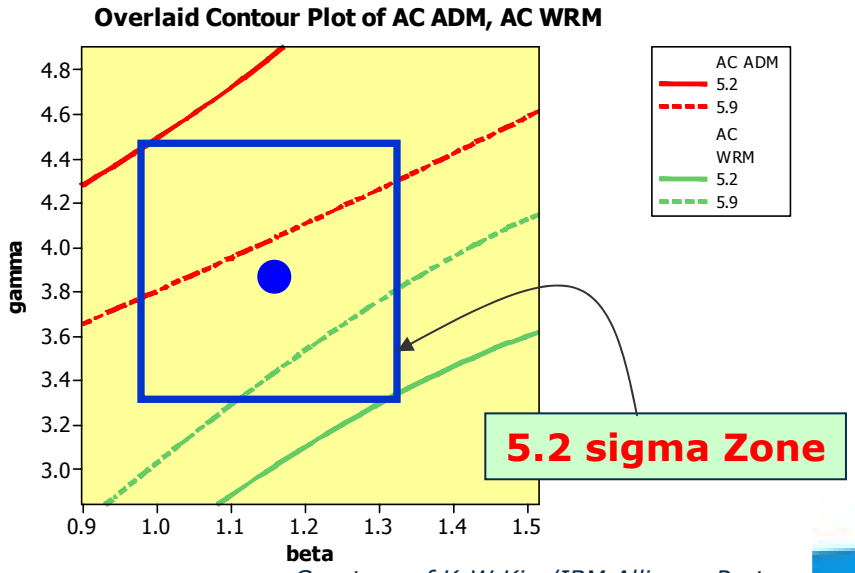
Chartered

Guaranteed Zone for beta & gamma balancing





- Beta & Gamma balancing is critical during SRAM device optimization by taking into account Iread vs Istby current
- The simulation data indicates the guaranteed zone of beta & gamma window to hit 5.2 sigma based on ADM and WRM dependency



- Courtesy of K-W Kim/IBM Alliance Partner

Chartered