Reducing Transistor Variability For High Performance Low Power Chips

HOT Chips 24

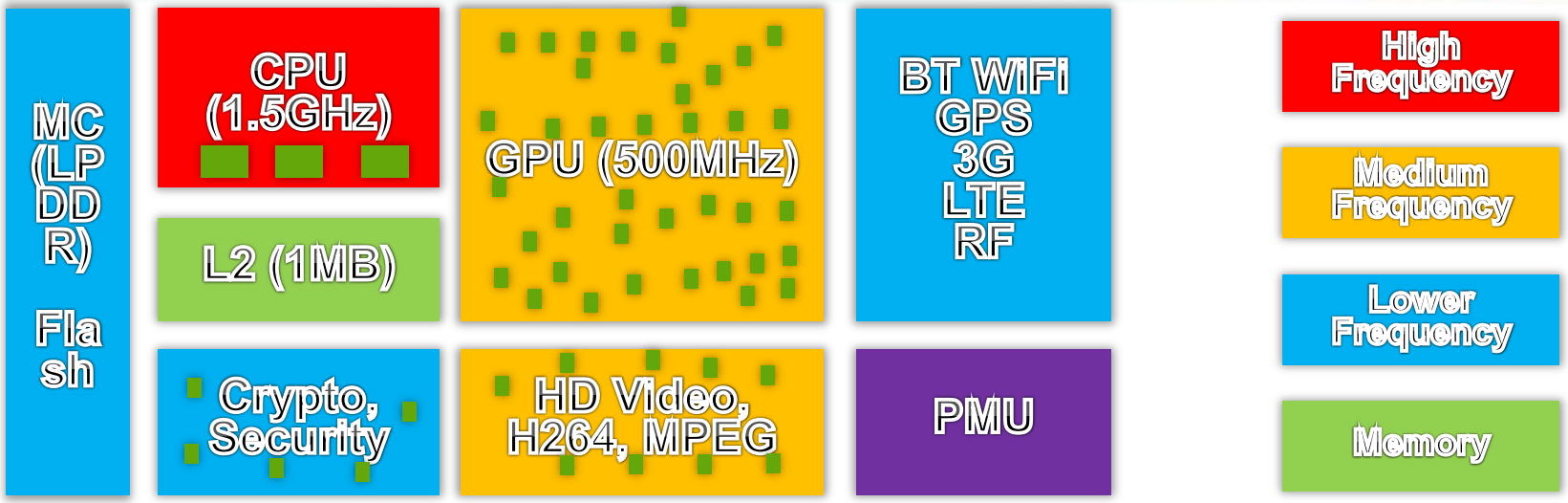
Dr Robert Rogenmoser

Senior Vice President Product Development & Engineering

Overview

- Transistor Variability Limits Chips
  - Impact on Mobile System on Chip (SOC)
  - Limited Low Power Design Techniques
  - Where does Variability come from?
- New Transistor Alternatives to Reduce Variability
  - Deeply Depleted Channel (DDC) technology
  - Silicon Impact
- Outlook
  - Taking advantage of Deeply Depleted Channel (DDC) in Mobile SOC

What is needed in Mobile System on Chip?



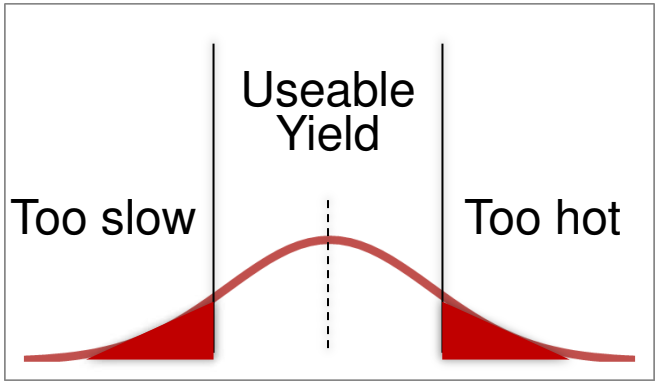
- Multiple blocks with different performance requirements
  - Integrated on the same die
  - Different power modes – would like to run at different supplies
  - Multiple V_T transistors used to control leakage
  - Single chip solution requires analog integration
- Need co-design of architecture, circuits and transistor technology for best solution

Variability Limits Design & Architecture

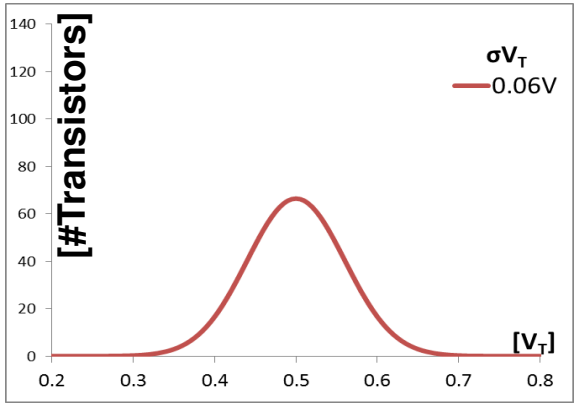
- Limited benefit using voltage scaling (DVFS)
  - Cannot overdrive much due to reliability and power restrictions
  - Dynamically lowering voltage limited to 100-200mV
  - Only lowering frequency leaves large leakage power
  - “Run to hold” beats DVFS despite overhead
- Finicky SRAM memories
  - High SRAM V_{MIN} leaves no room for memory voltage scaling
  - Many circuit tricks to improve V_{MIN} and noise margins
  - Design teams moved to dedicated power rail for SRAM
    - Works for CPU – difficult in GPU
    - Impacts power network integrity – more fluctuations
- Transistor variability limits chips

Transistor Variation Source of Chip Variation

- Global/Systematic/Manufacturing Variation
  - Shifts all the transistors similarly
    - Longer/shorter transistor lengths
    - More (or less) implant energy and dose
    - ➔ Will result in speed/power distribution



- Local/Random Variation
  - Transistor next to each other vary widely
  - Small number of dopants in transistor channel
  - ➔ Random Dopant Fluctuation (RDF)
    - Apparent in threshold voltage mismatch (σV_T)
    - Impacts speed, leakage, SRAM & Analog

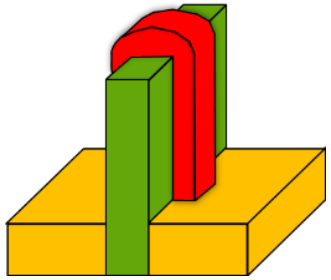


  - ➔ Industry solution: Remove RDF using **Undoped Channel**

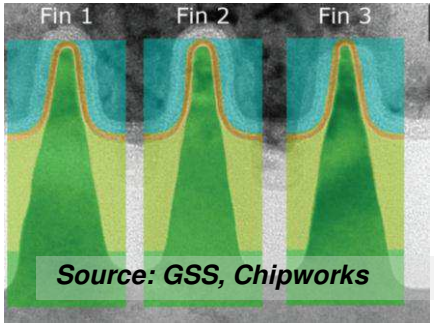
- What is the right silicon roadmap going forward?

Transistor Alternatives

- FinFET or TriGate
  - Promises high drive current
  - Manufacturing, cost, and IP challenge
  - Doped channel to enable multi V_T

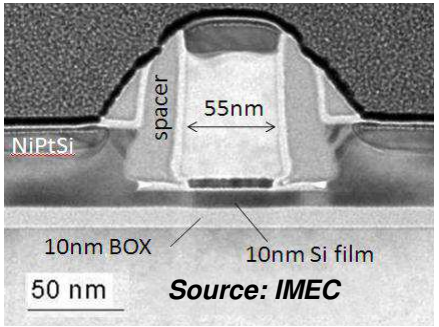


Textbook FinFET

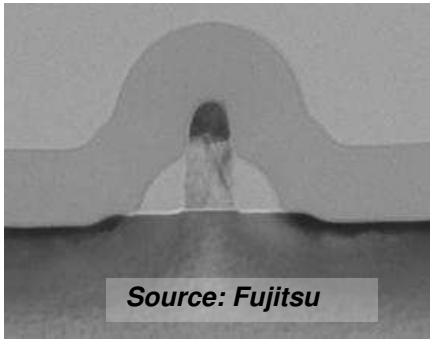


Intel TriGate

- FDSOI
  - Showing off undoped channel benefits
  - Good body effect, but lack of multi V_T capability
  - Restricted supply chain



- DDC – Deeply Depleted Channel transistor
  - Straight forward insertion into Bulk Planar CMOS
  - Undoped channel to reduce random variability
  - Good body effect and multi V_T transistors

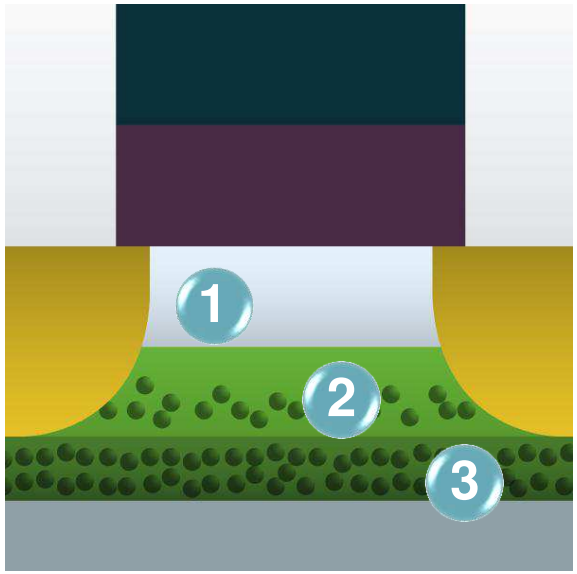


SuVolta®

Deeply Depleted Channel™ (DDC) Transistor

1 Undoped or very lightly doped region

- Significantly reduced transistor random variability σV_T
  - ➔ Lower leakage
  - ➔ Better SRAM (I_{READ}, lower V_{min} & V_{ret})
  - ➔ Tighter corners
  - ➔ Smaller area analog design
- Higher channel mobility (increased I_{eff}, lower DIBL)
  - ➔ Higher speed, improved voltage scaling



**Example implementation*

2 V_T setting offset region

- Enables multiple threshold voltages

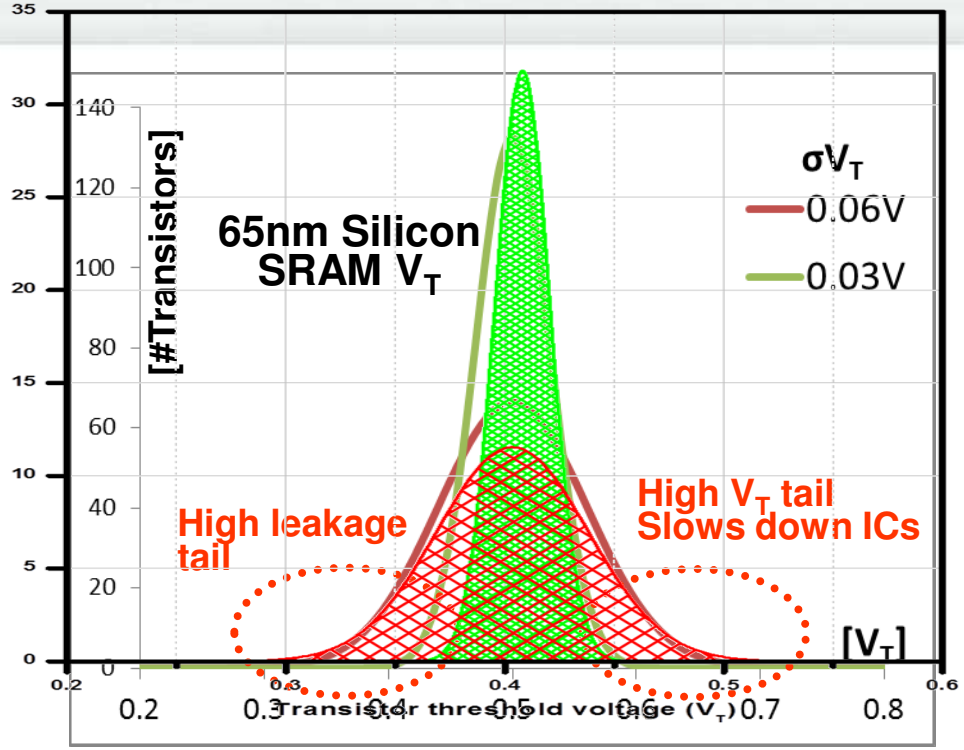
3 Screening region

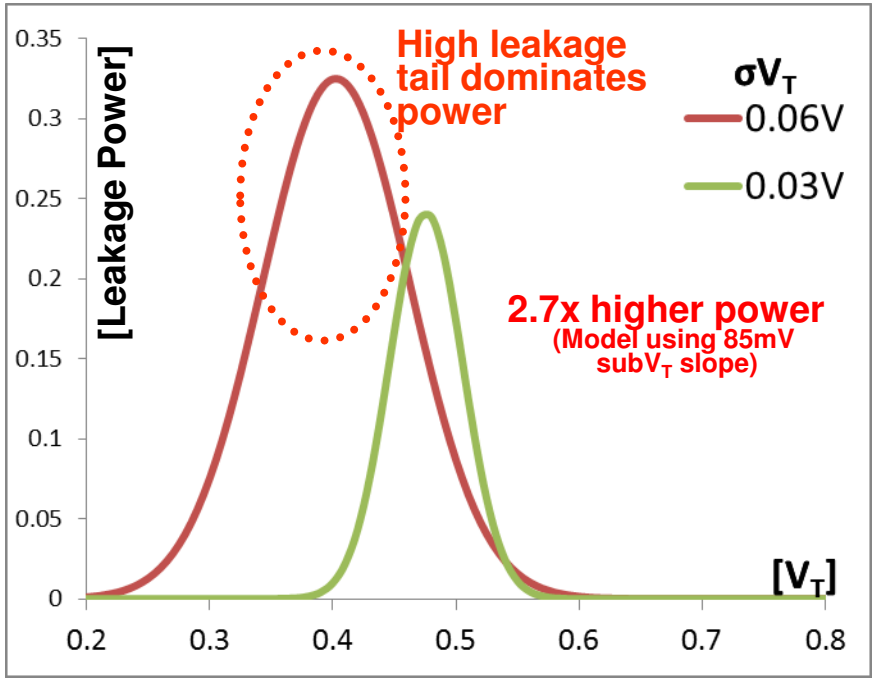
- Strong body coefficient
  - ➔ Bias bodies to tighten manufacturing distribution
  - ➔ Body biasing to compensate for temperature and aging

Benefits similar to FinFET in planar bulk CMOS

SuVolta®

Lower Transistor Variability Reduces Leakage

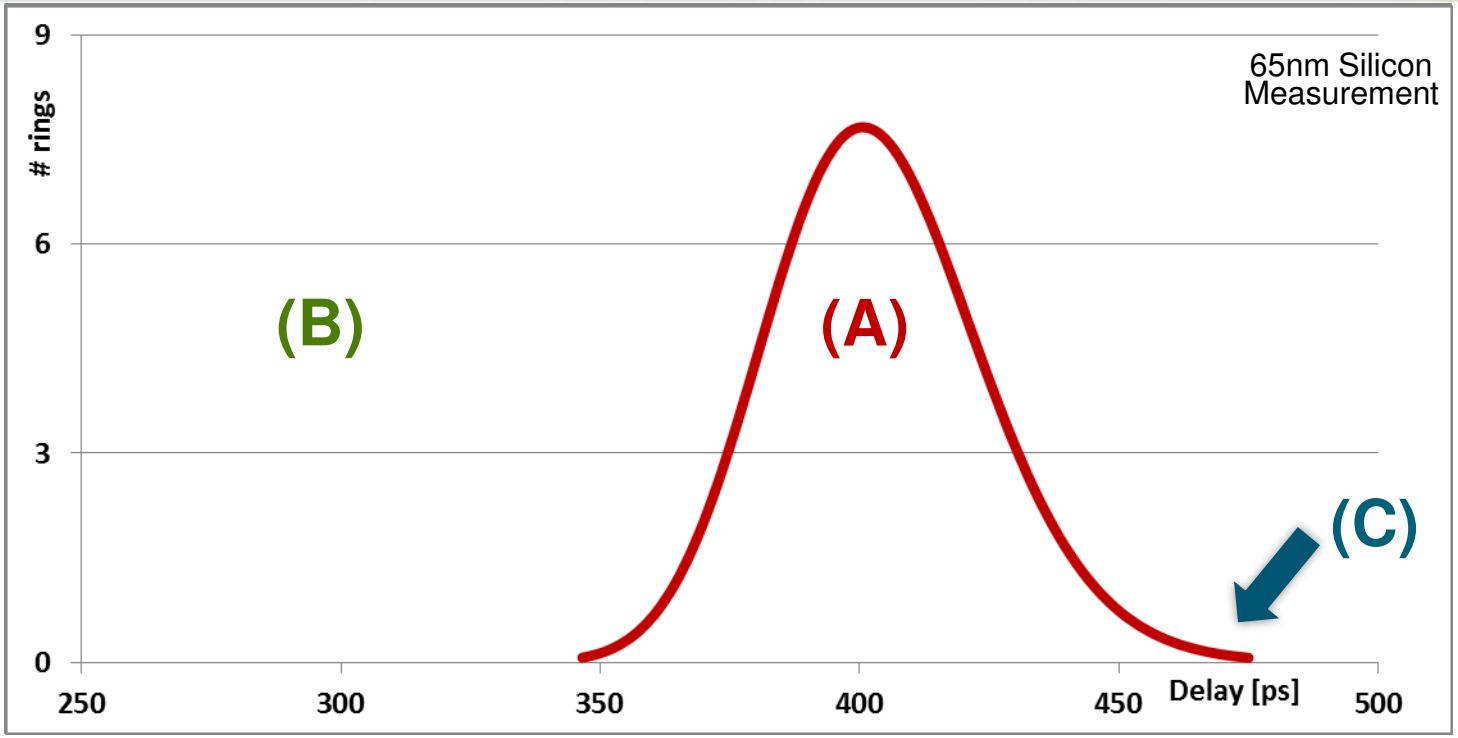




- Transistor variability is reflected in threshold voltage (V_T) distribution
- Leakage current is exponentially dependent on V_T
- Lower V_T variability (σV_T) reduces number of leaky low V_T devices
- Power dissipation is dominated by low V_T edge of distribution
- Smaller σV_T → Less leakage power for digital and memory/SRAM

SuVolta®

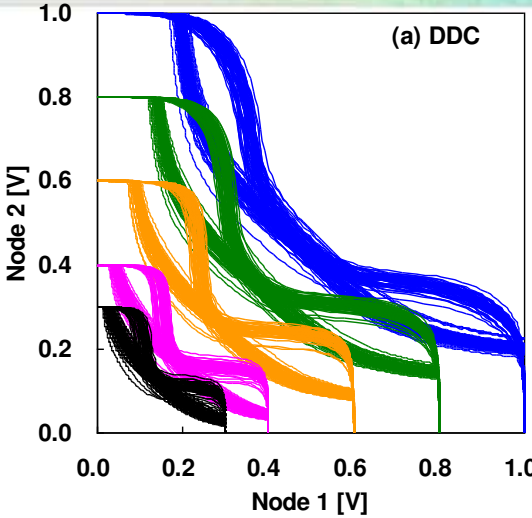
Lower Transistor Variability Improves Speed



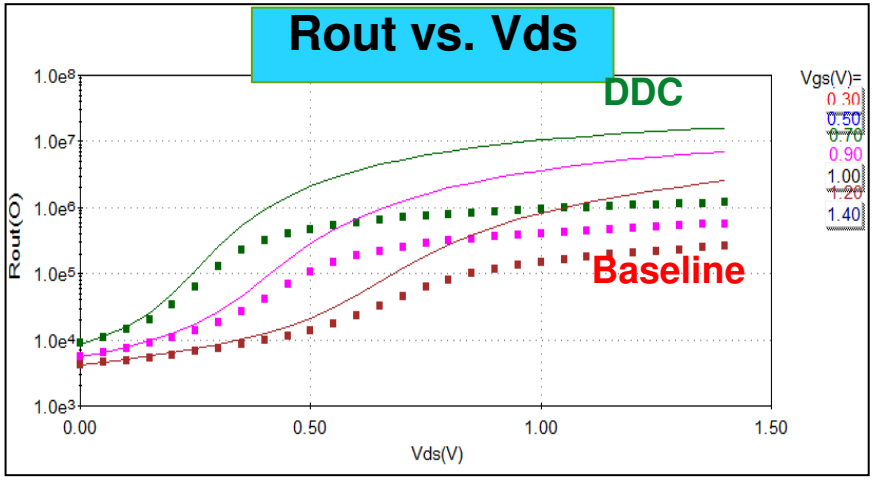
- Nominal (TT) ring oscillator speed expected to be 400ps **(A)**
  - Equivalent to having many similar critical paths in a chip
  - V_T variation will randomly affect paths within the same die limiting speed to 470ps
- Undoped channel reduces variability and increases mobility **(B)**
  - 25% faster mean, 30% faster tail due to tighter distribution
- To match performance lower V_{DD} until tails have same speed **(C)**
  - Large impact on power due square dependence $P=CV^2f + IV$

Lower Variability Improves Transistor Matching

- SRAM memories built using 6-T SRAM cell
  - Smallest transistors on every chip, worst V_T mismatch
  - Higher V_{DD} is required to avoid failures
  - Demonstrated SRAM to V_{min} of 0.425V
- In analog circuits, matching is key
  - Large transistors used to improve relative variability in current mirrors, differential pairs, etc.
  - Better transistor matching allows for
    - Area savings
    - Higher performance
    - Lower power
  - Undoped channel improves R_{OUT} ➔ higher gain

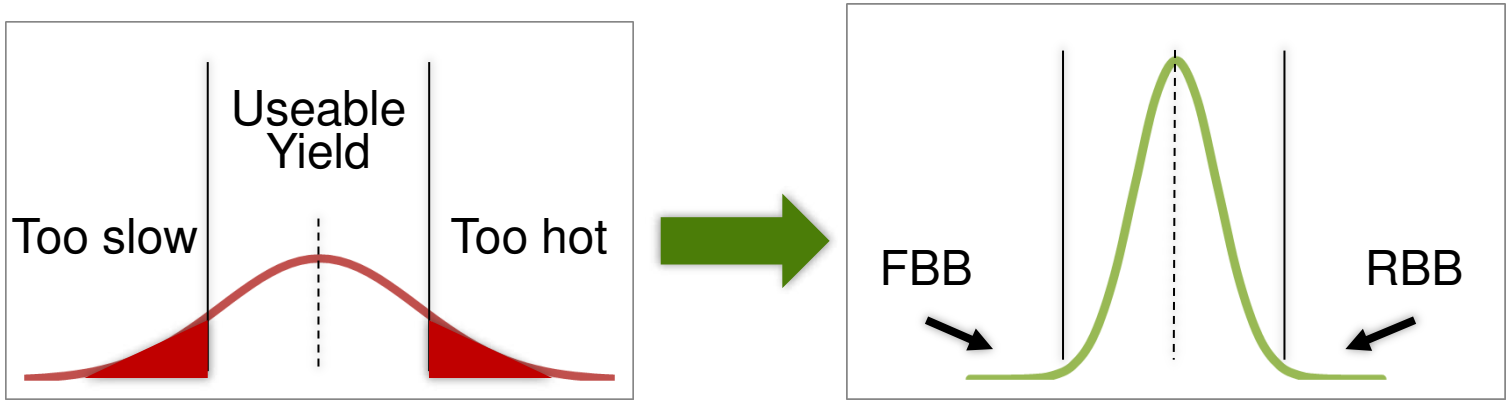


65nm Silicon Measurement



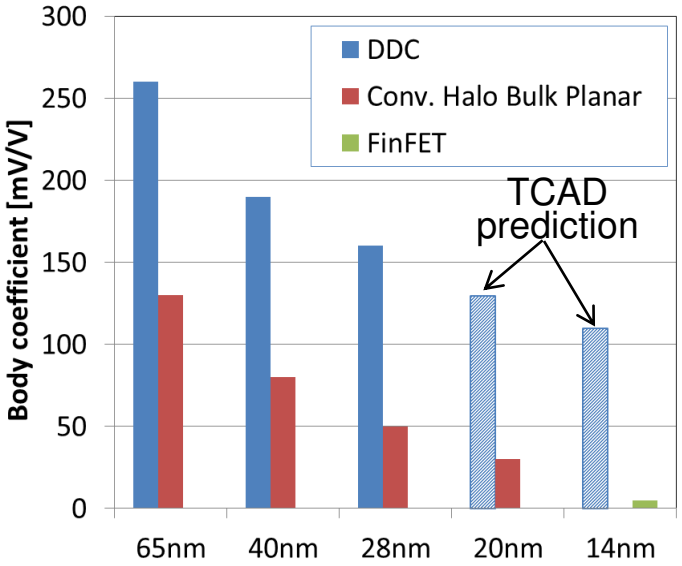
SuVolta®

Better Chips with Body Biasing



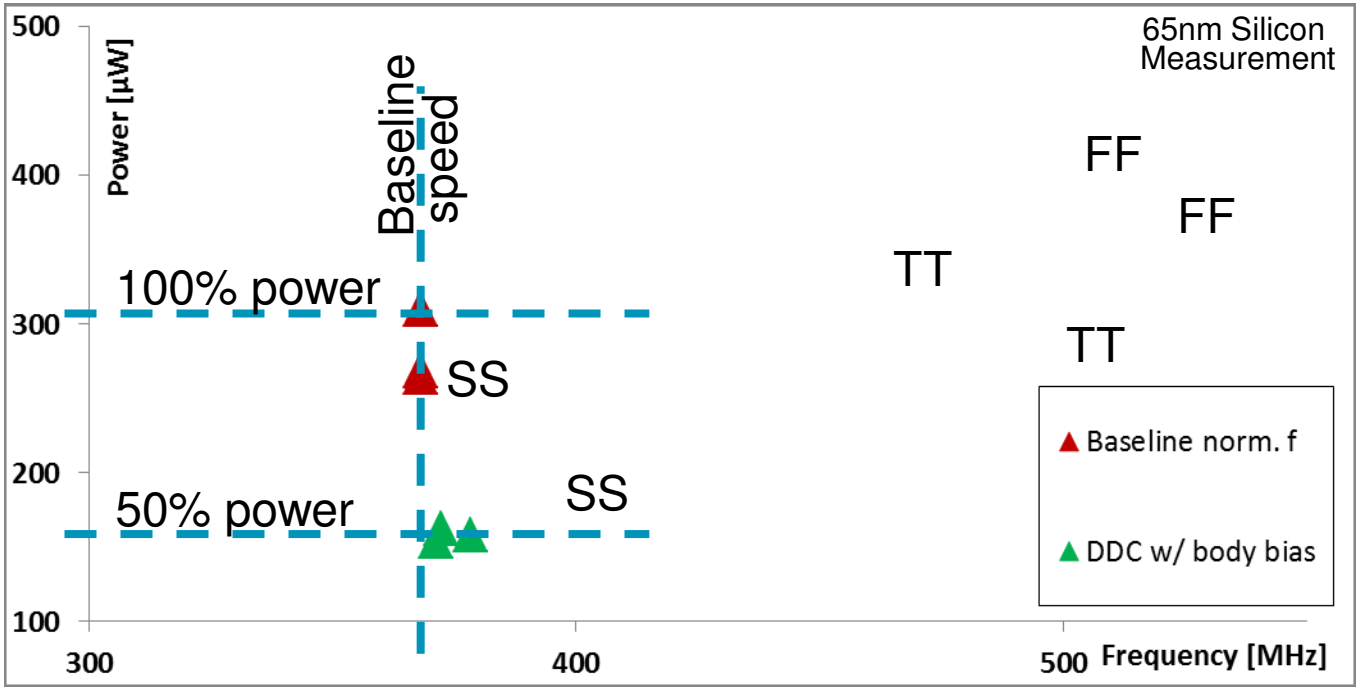
- Body Bias to fix systematic variation
  - Speed-up (forward bias - FBB) slow parts
  - Cool down (reverse bias - RBB) hot parts

  ➔ Increase manufacturing yield
- Body bias enables multiple modes of operation
  - Active → minimize power at every performance
  - Standby → leakage reduction, power gating
- DDC provides 2-4x larger body factor



SuVolta®

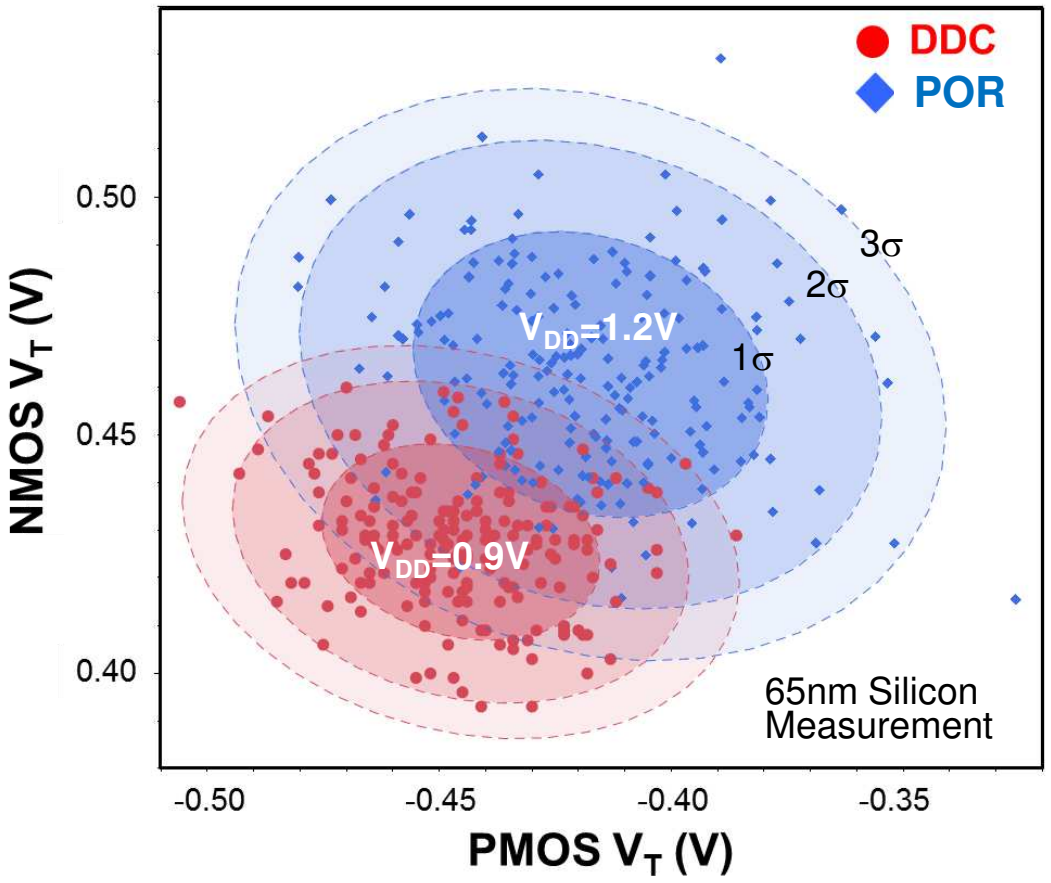
Half the Power at Matched Performance



- Inverter ring-oscillators (RO) fabricated at process corners
  - Baseline @ 1.2V V_{DD} and DDC @ 0.9V V_{DD}
- For each corner, DDC RO is faster and lower power
- Using strong body coefficient to pull in corners
  - Half the power (50% less power) while matching speed

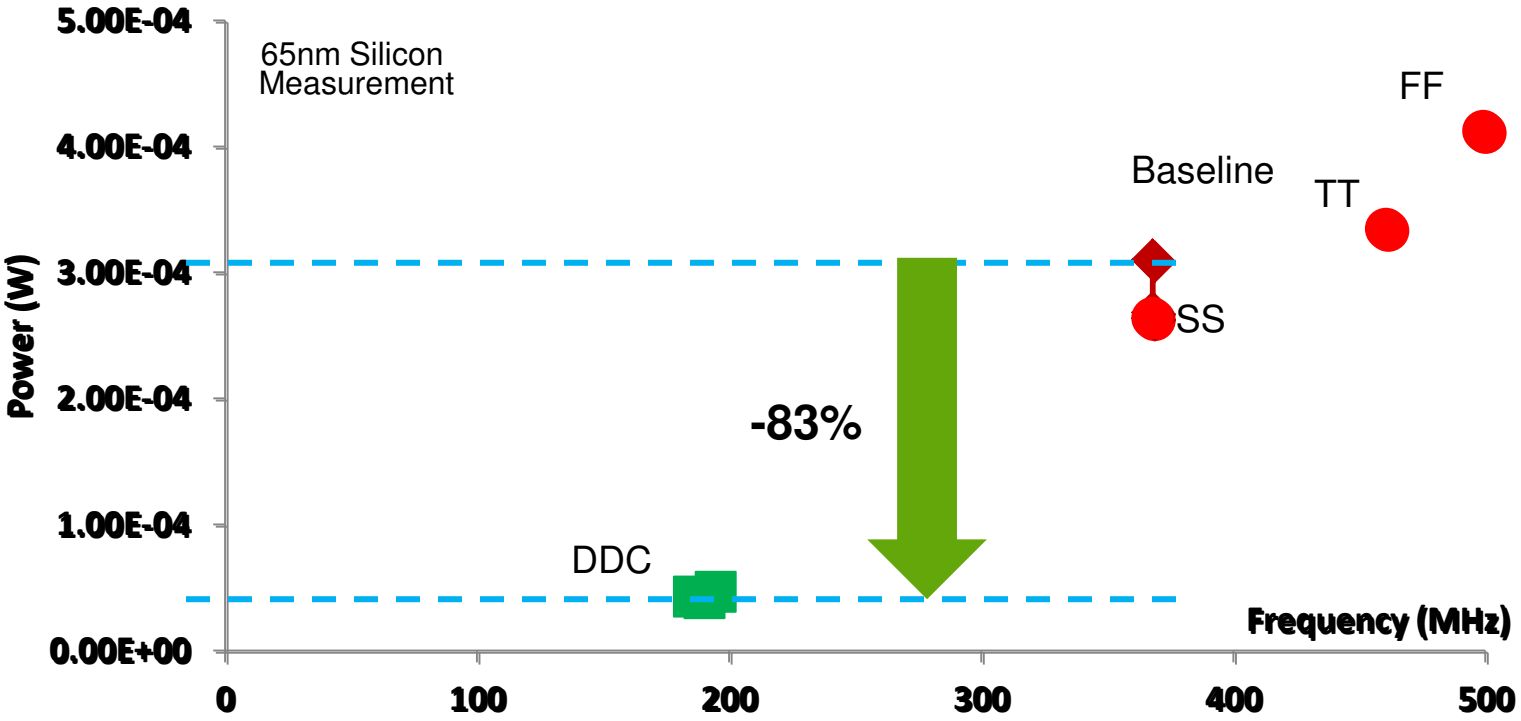
Tighter Manufacturing Corners w/ DDC

- Better process control leads to tighter corners
  - Manufacturing flow further reduces layout effects
- 1 sigma tighter wafer to wafer and within wafer variation for DDC
- Less overdesign as max paths and min (hold) paths are closer
- Faster design closure
  ➔ earlier tapeout
  ➔ shorter TTM



SuVolta®

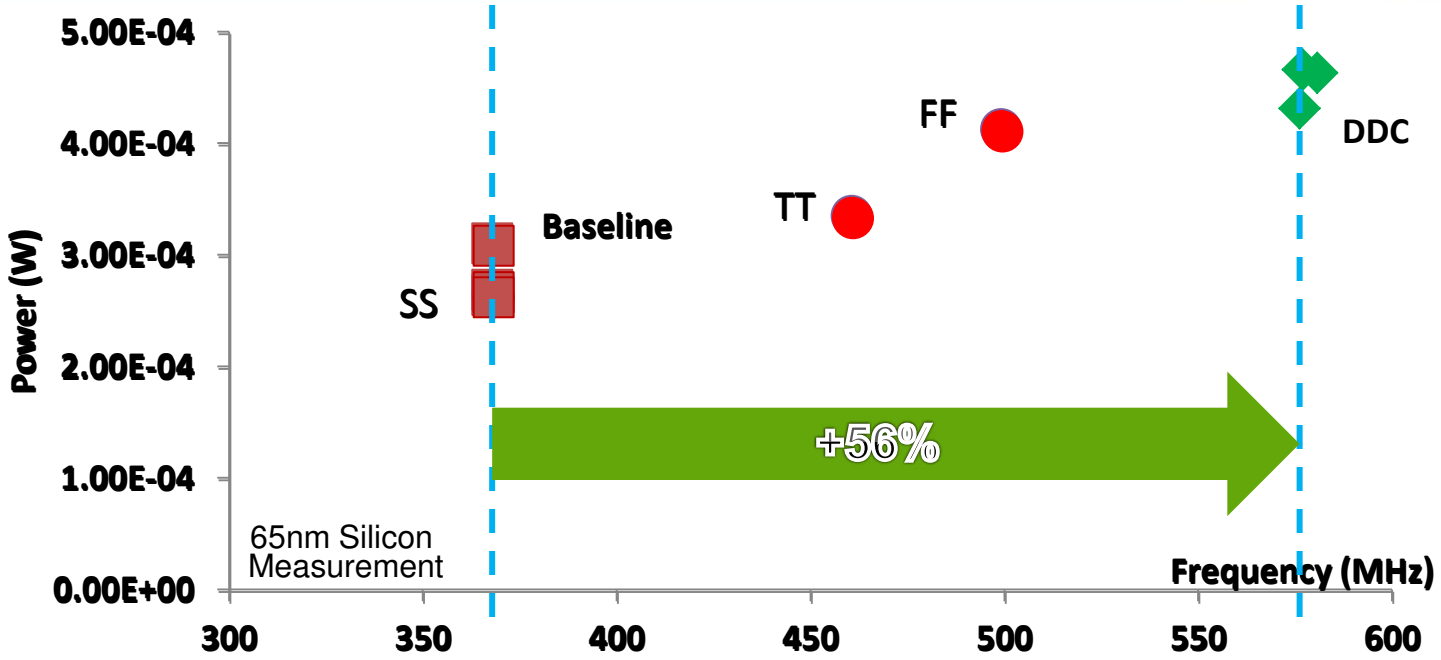
Voltage Scaling to 0.6V V_{DD}



- Achieve half the speed at 1/6 the power @0.6V V_{DD}
- Use body bias to compensate for temperature and aging
  - Critical for low V_{DD} operation
  - Enable workable design window – avoid overdesign

SuVolta®

This is HotChips – Go Faster!

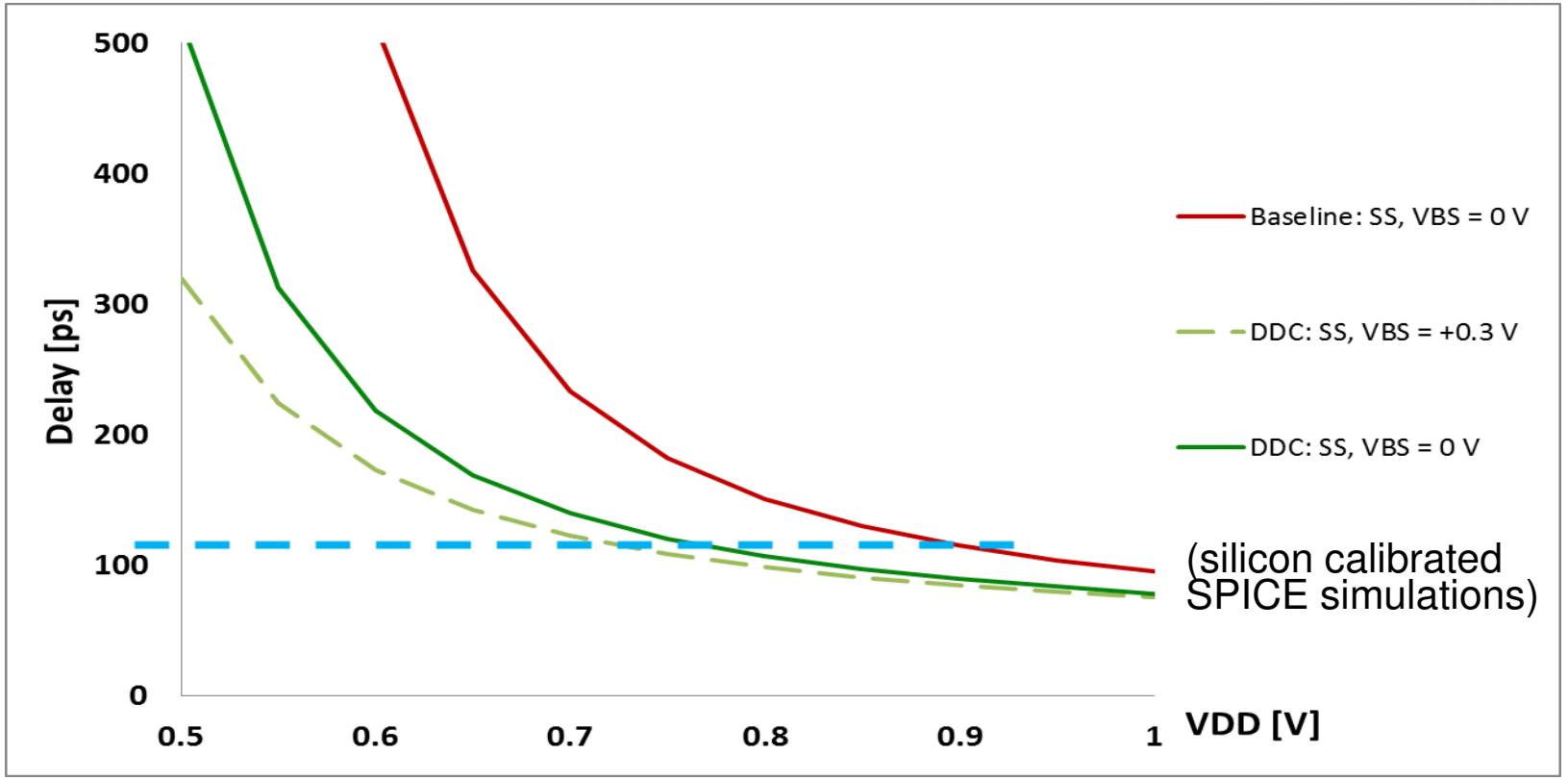


- Turbo Mode: DDC achieves over 50% speedup @ 1.2V V_{DD}
  - All corners for DDC run at 580MHz vs 370MHz for baseline

DVFS	Baseline	DDC			
V_{DD}	1.2V	0.6V	0.9V	1.05V	1.2V
Speed	**1**	0.5	**1**	1.28	1.56
Power	**1**	0.17	0.52	**1**	1.51

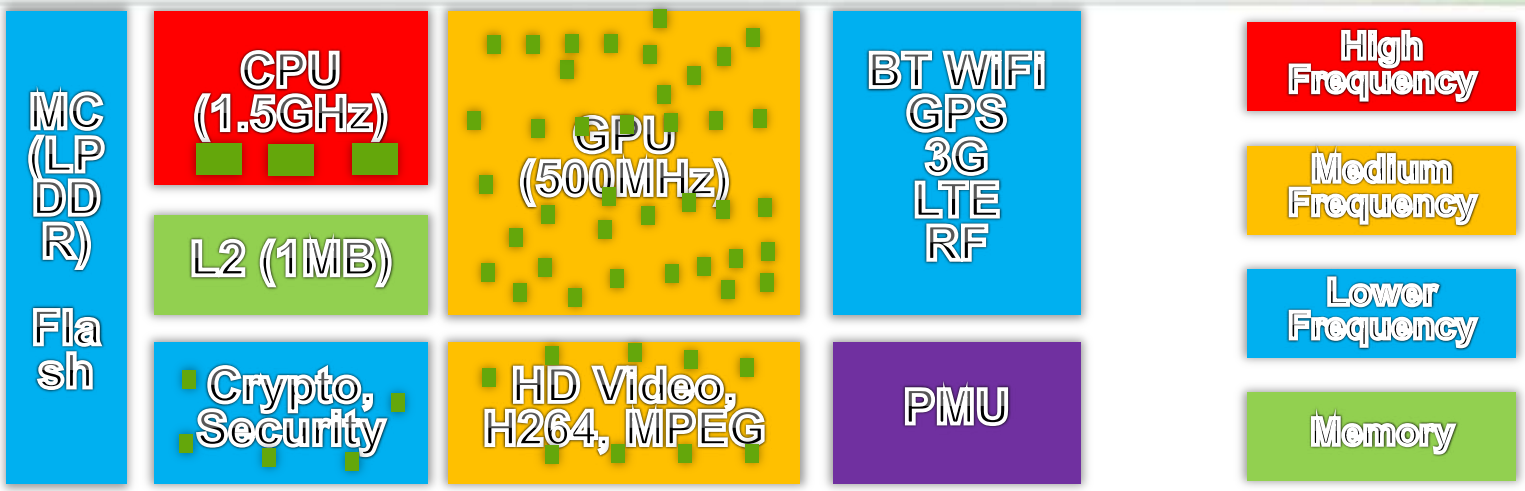
SuVolta®

28nm and Beyond



- Same performance at 0.75V V_{DD} as baseline at 0.9V V_{DD}
  - 30% lower power
  - Alternatively 25% faster at same voltage
- Even better when using body bias to pull in corners

Applying DDC to Lower Variability in Mobile SOC



- CPU: Single thread performance critical
  - Push frequency by temporarily raising voltage in turbo mode
  - DVFS with body biasing becomes DVBFS
- GPU: High number of cores using small transistors
  - Less overdesign due to lower delay variability
  - Increase parallelism, lower voltage, body bias dynamically for more pixels/Watt
- Lower frequency blocks
  - In addition to high V_T transistors also run at lower voltage and optimal body bias
- Whole chip: Use body bias to adjust for manufacturing variation
  - Take advantage of improved memory and analog performance
  - Lowering variability while compatible with existing bulk planar silicon IP

Conclusions

- Variability limits chips
  - DDC reduces random variability through its undoped channel
  - DDC's strong body factor can be used to fix systematic variation and compensate for temperature variation
- DDC provides performance kicker from 90nm to 20nm
  - Straight forward integration into existing nodes
  - Compatible with existing bulk planar CMOS silicon IP
  - Use existing CAD flow
- DDC brings back low power tools
  - Large range DVFS
  - Body biasing
  - Low voltage operation
- Taking advantage of reduced variability DDC in design and architecture will lead to next level in mobile SOC