

Where is my "Typical" Chip? Relating Silicon Back to the Timing Sign-Off Model

Christian Lütkemeyer, 3/21/2019

Objective of the Presentation

- Show real world measured data from a product chip to provide a practical perspective on the dynamic performance of complex CMOS SOCs
 - Dynamic performance of ring oscillators (ROs) for different VTs => estimate the silicon process corner
 - Estimation of intra die spatial performance gradients
 - Layer-to-layer interconnect capacitance ratios
 - Measured data from on-chip supply noise
- Contrast this data to the idealized assumptions that we make during the chip design phase.
- Provide recommendations on how we can collaborate with Foundries and EDA partners to provide a smooth path from design to Device Validation Testing (DVT) and to volume manufacturing.
 - Opportunities to grow the au Workshop

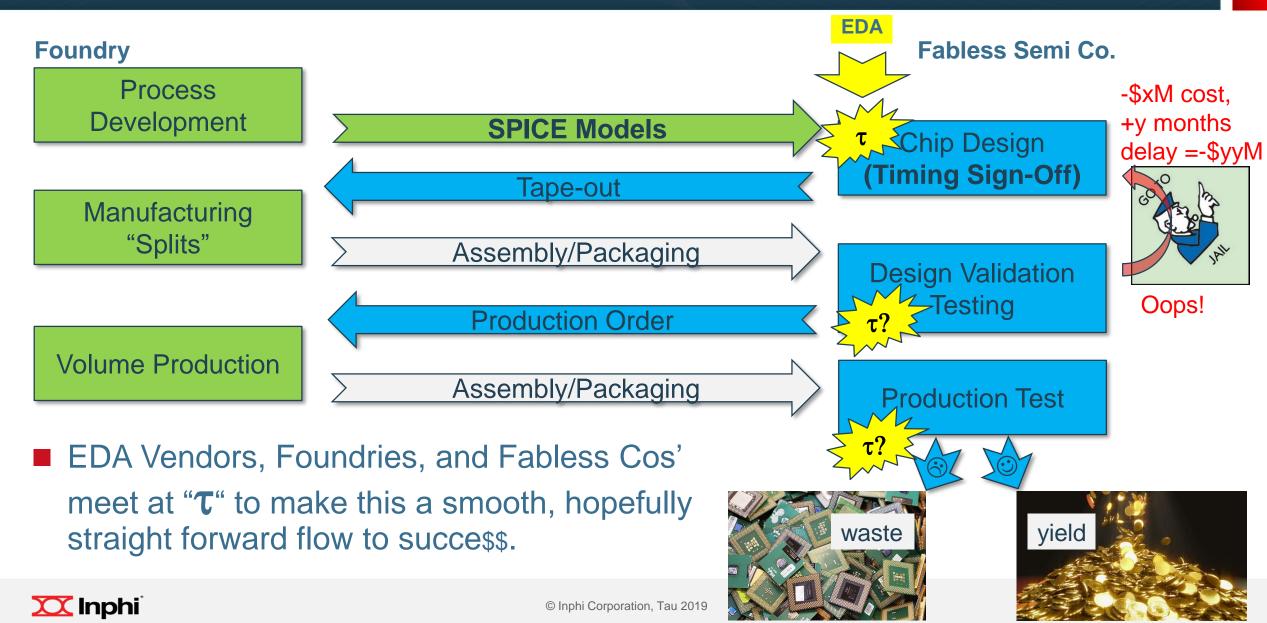


Outline

- Introduction
 - Business cooperation between Foundry and Fabless Semi Company
- A parametric model to map ring oscillator delay to a process estimate
- Example of estimated process data
 - Cumulative distributions in the design window
 - Intra-die gradients
- Layer-to-layer interconnect capacitance ratios
- On-chip supply variability
- Summary of the presented data, wishes and conclusions,
 - **T**-Workshop expansion opportunities

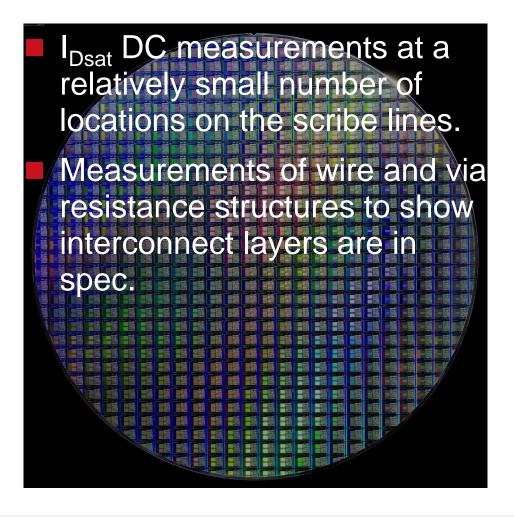


Foundry and Fabless Semiconductor Company Business Cooperation



Classifying a Wafer vs. Classifying Individual Chips

Foundry



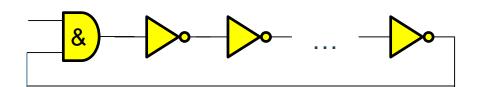
Inphi

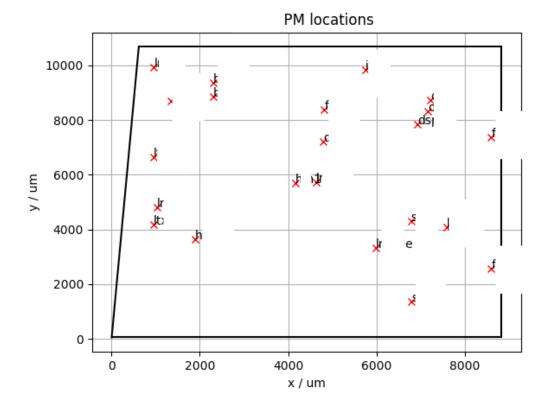
Each chip is classified individually in several locations with dynam oscillator measurements Measurement of layer-to-layer capacitance ratios to gauge the impact of interconnect variation on dynamic performance. Every die coun



Ring Oscillators and Process Monitor (PM) Locations

Ring Oscillators for each VT class [SVT, LVT, UVT]

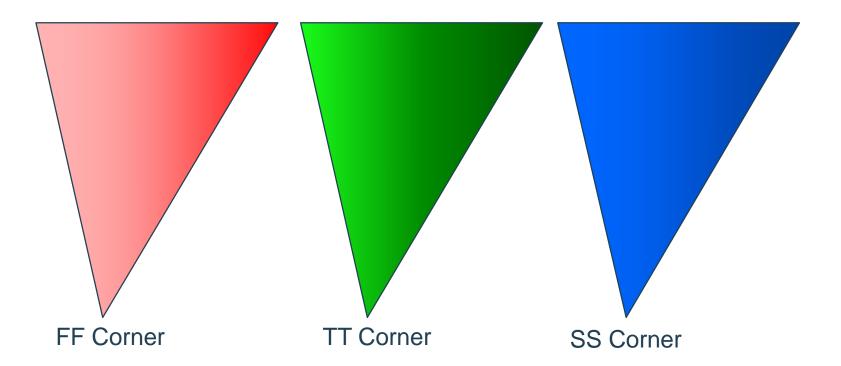






CMOS Delay Variation



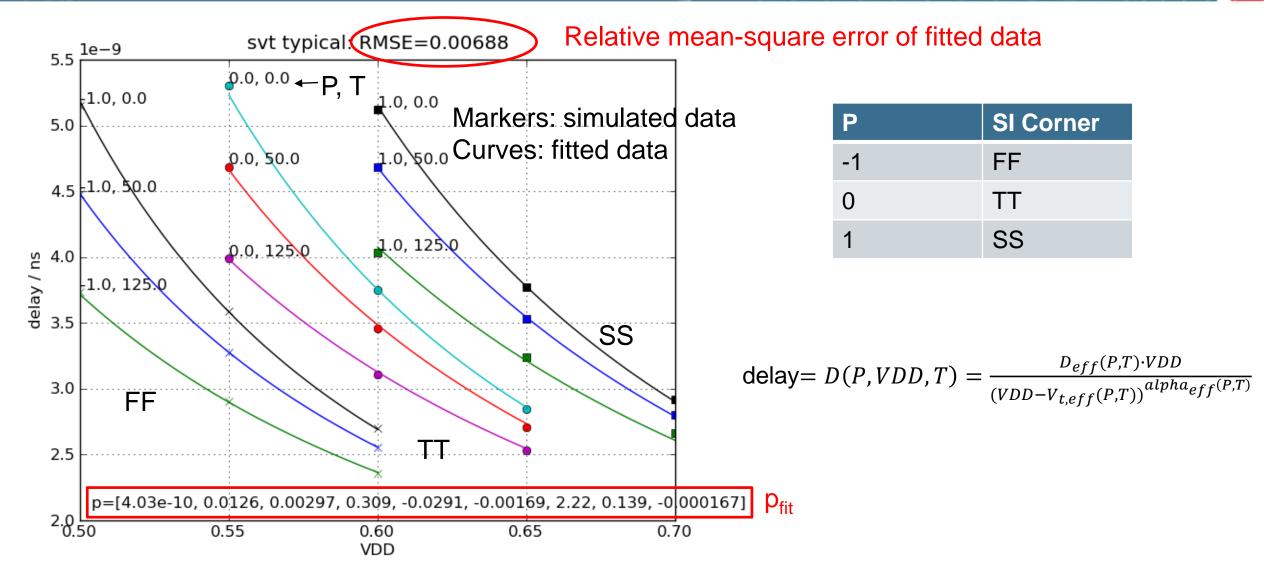


- Interconnect variation (Cmax, Cmin, RCmax, RCmin)
- Voltage variation
- Temperature variation
- local variation (LVF)
- modeling corners

How can we estimate the silicon corner that corresponds to a measured dynamic performance?



3D (PVT) Delay Fit: Process Monitor RO svt, typical Interconnect





Parameterized Delay Fit for Process Monitor ROs

- $D_{eff} = D_0(1 + k_{DP} \cdot P + k_{DT} \cdot T)$ "sensitivity of delay parameter to P and T"
- $V_{t,eff} = V_T(1 + k_{VtP} \cdot P + k_{VtT} \cdot T)$ "sensitivity of V_t parameter to P and T"
- $alpha_{eff} = alpha(1 + k_{alphaP} \cdot P + k_{alphaT} \cdot T)$ "sensitivity of alpha parameter to P and T"
- Fitting vector with 9 parameters:

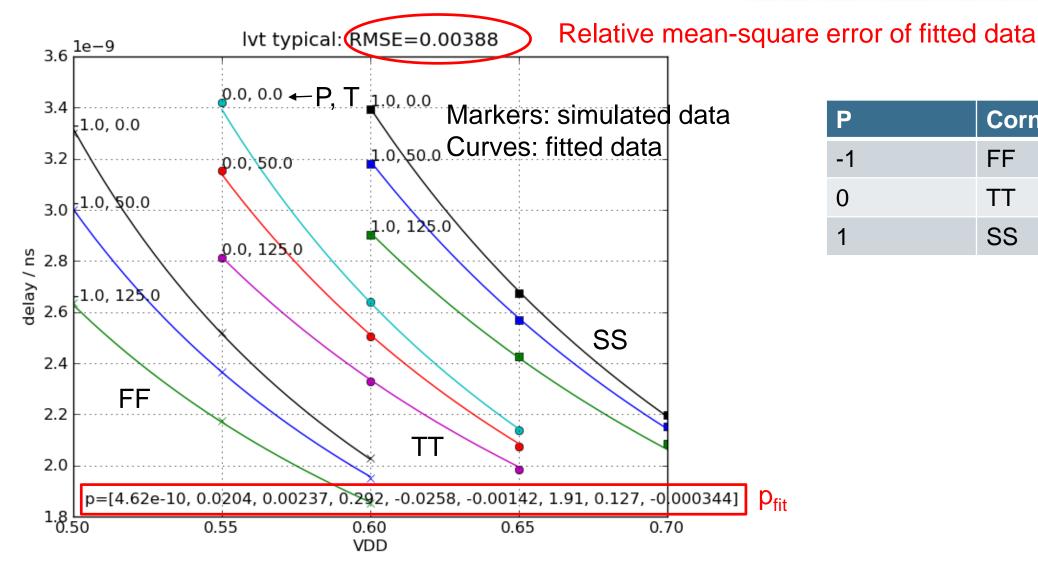
$$p = [D_0, k_{DP}, k_{DT}, V_T, k_{VtP}, k_{VtT}, alpha, k_{alphaP}, k_{alphaT}]$$

$$D = D(P, VDD, T) = \frac{D_{eff} \cdot VDD}{(VDD - V_{t,eff})^{alpha_{eff}}}$$

- The above fitting function can be used to estimate the process P if we measure D at a known VDD and T.
 - This is a dynamic estimate of the process corner. Interconnect variation of the connecting metal is translated into a process shift. => There will be small differences between a DC Idsat-based process corner classification and dynamic ring oscillator based classifications.



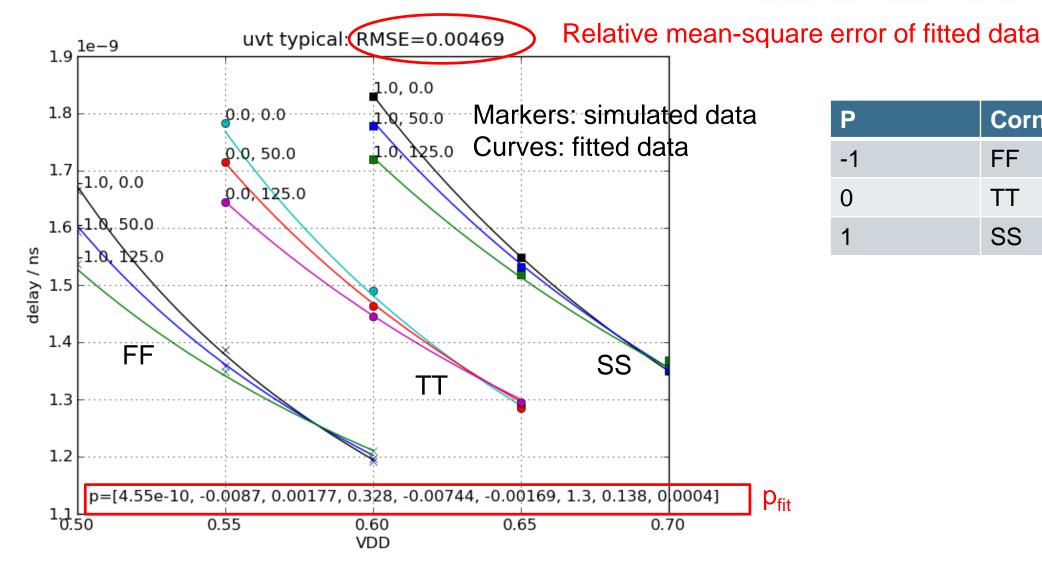
3D (PVT) Delay Fit: Process Monitor RO lvt, typical Interconnect



P	Corner
-1	FF
0	TT
1	SS



3D (PVT) Delay Fit: Process Monitor RO uvt, typical Interconnect

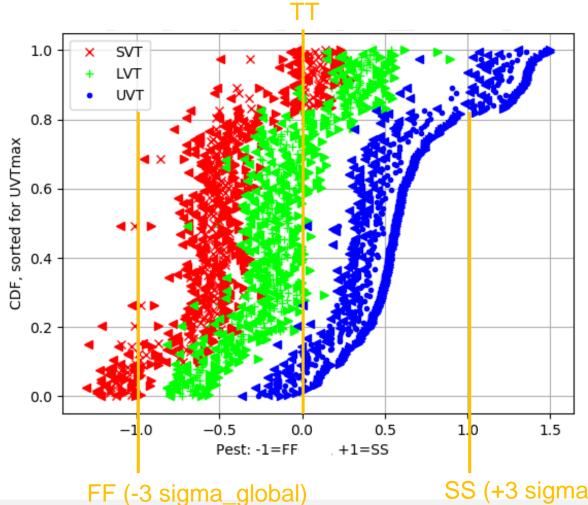


P	Corner
-1	FF
0	TT
1	SS



Cumulative Distribution Function of Initial Split Lot Parts

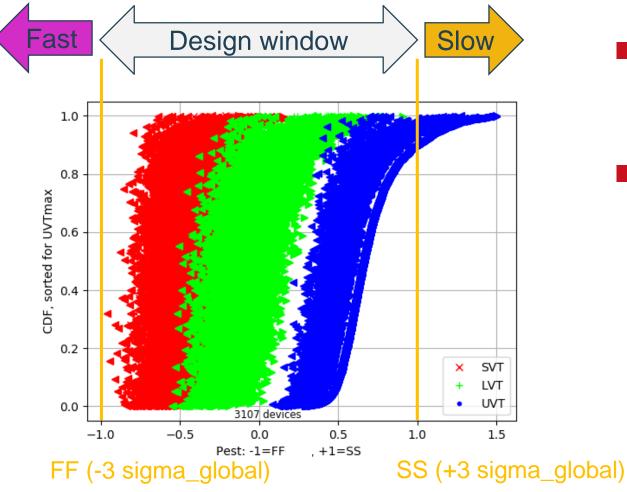
Performance Variation on Splits, ~200 Parts



- For TT targeted wafers Pest for all VTs are inside the design window.
- However, when LVT devices are typical, SVT trends ~1.5 sigma fast and UVT trends ~1.5 sigma slow.
 - There are no FF or SS chips where all VTs are close to the corners.
 - There is no "typical" silicon. => Most chips will require a supply voltage above "typical" as the slower UVT devices need to be compensated.
 - Leakage power will be higher as faster LVT and SVT devices cause increased leakage.
- Significant intra-die performance variation.

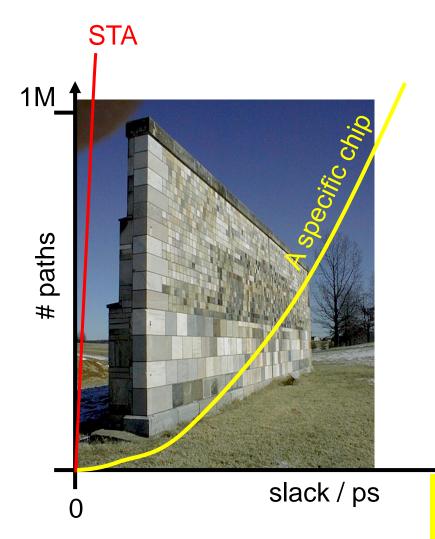
SS (+3 sigma global)

Estimated Process Data from an Early Production Lot



- About 10% of the parts have their worst UVT RO slower than the "SS" model point.
- The data clearly shows the different skew for the VT classes relative to the process window.
 - SVT trends fast around -0.5 (1.5 sigma_global fast)
 - LVT targets typical
 - UVT trends to +0.5 (1.5 sigma_global slow)

The Timing Closure "Slack Wall" – A Myth for Individual Chips!

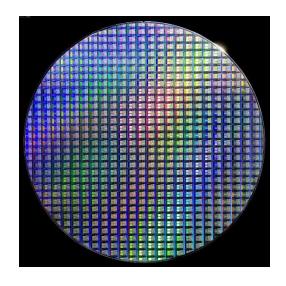


- Setup timing optimization with power recovery creates a "slack wall" in STA.
 - The timing reports suggest that there is a very large number of paths that are at the edge of being critical.
- However, on a particular die the number of close to critical paths is much smaller than the slack wall in STA suggests. STA assumes incorrectly:
 - All VTs are aligned at the edge of the manufacturing window. Skew between VTs will flatten the wall.
 - STA assumes all metal layers are at an extreme. In reality only a subset of the layers may be critical.
 - STA assumes SI has its worst impact whenever there is overlap to the early-late window. In many cases there may not be switching in the middle of the window.
 - STA assumes that all variations covered by derating margins push a path towards the wall. This will only happen for a few paths in a chip.
 - Most chips have performance gradients. Paths in the slow location fail first.
 - STA assumes worst case supply drops for all paths. In reality only a subset of paths is at the lowest supply.
- => The slack wall is a myth!
 - High sigma confidence analysis of setup slacks is overkill! (IMHO)

Analyzing Intra-Die Spatial Process Gradients



STA:
Flat Earth?
How do you
margin for spatial
gradients?



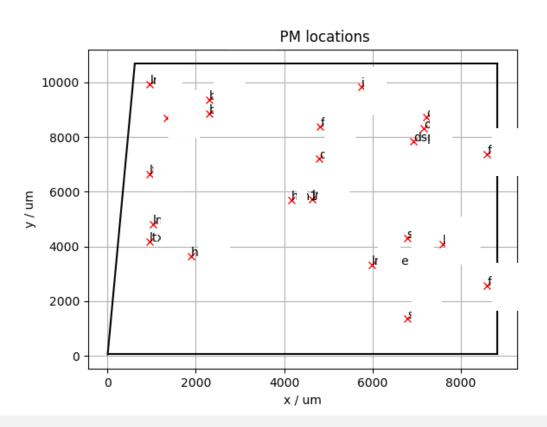
Wafers are physically very flat (CMP). Electrically not so much!



- We can use the Process Monitor ROs to estimate intra-die gradients of the process parameter P_{est}.
- Uncorrelated local variation in the individual RO instances creates estimation error of the gradients.
- To gauge the size of this estimation error we create an artificial RO dataset that assumes constant P and adds delay variation according to local variation from MonteCarlo simulations.



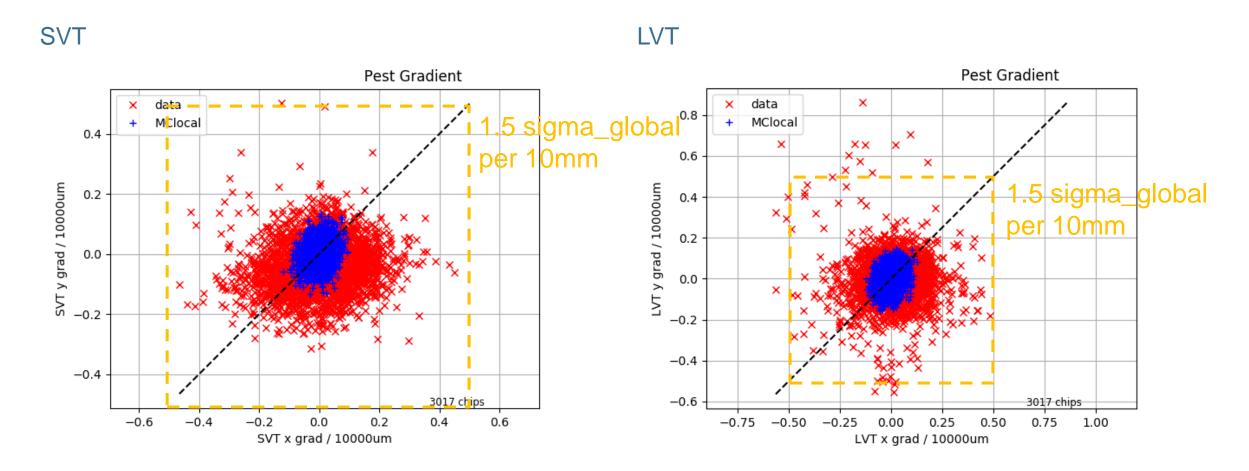
Locations of Process Monitors; Linear Gradient Fit of Process P



- Fit of performance planes for each VT class [SVT, LVT, UVT]
- $P_{fit}(x,y) = P_0 + P_{grad,x} \cdot x + P_{grad,y} \cdot y$



Gradient Correlation x-grad, y-grad for SVT and LVT Devices

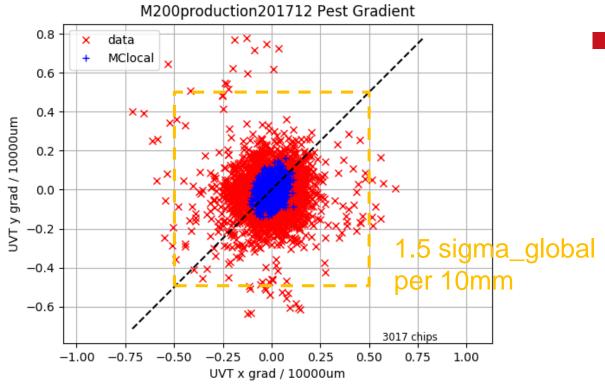


Observed gradients exceed significantly what would be observed if no gradients and only local variation existed as in the synthetic MClocal dataset.



Gradient Correlation x-grad, y-grad for UVT

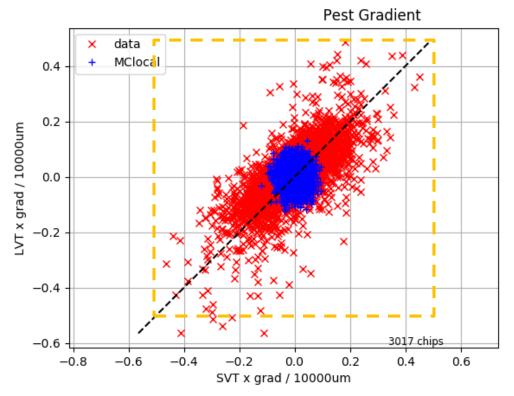
UVT



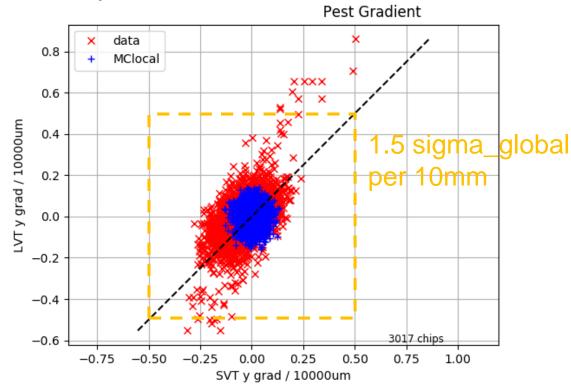
Observed gradients exceed significantly what would be observed if no gradients and only local variation existed as in the synthetic MClocal dataset.

Gradient Correlation SVT, LVT

Gradient x



Gradient y

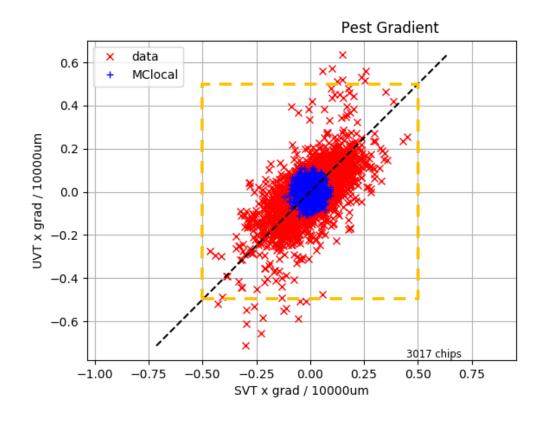


Observed gradients exceed significantly what would be observed if no gradients and only local variation existed. Significant correlation between SVT and LVT gradients

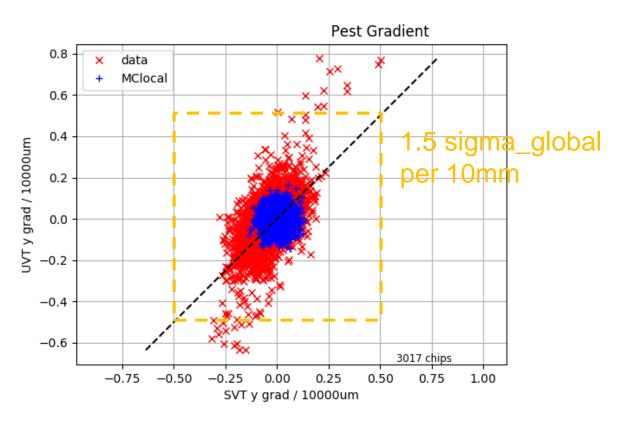


Gradient Correlation SVT, UVT

Gradient x



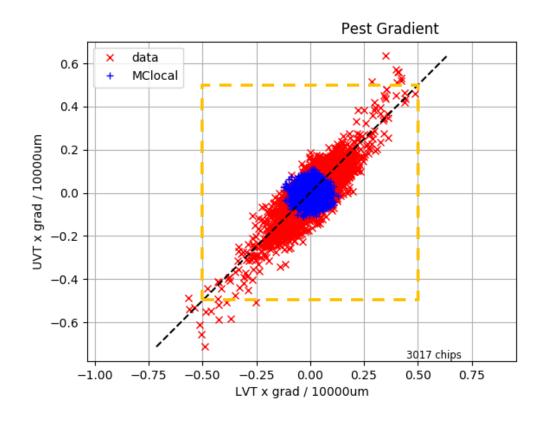
Gradient y



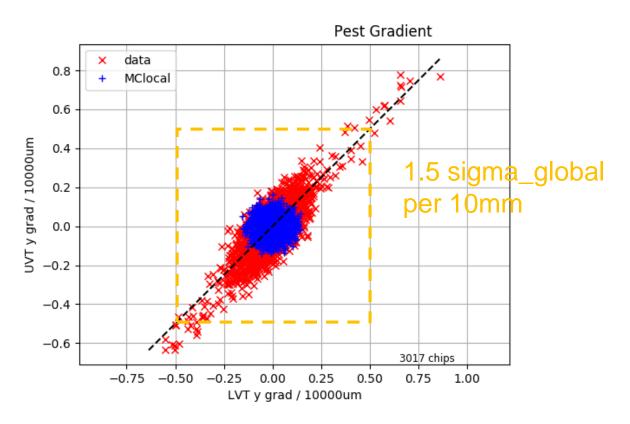
Significant correlation between SVT and UVT gradients. There is a significant number of chips where the UVT gradient is significantly larger than the SVT gradient.

Gradient Correlation LVT, UVT

Gradient x



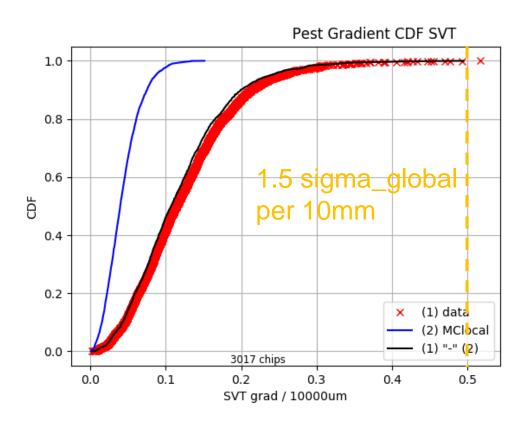
Gradient y



Strong correlation between LVT and UVT gradients. UVT gradient up to 2.4 sigma_global / 10mm.



CDF of Gradient= $\sqrt{gradx^2 + grady^2}$ (SVT)



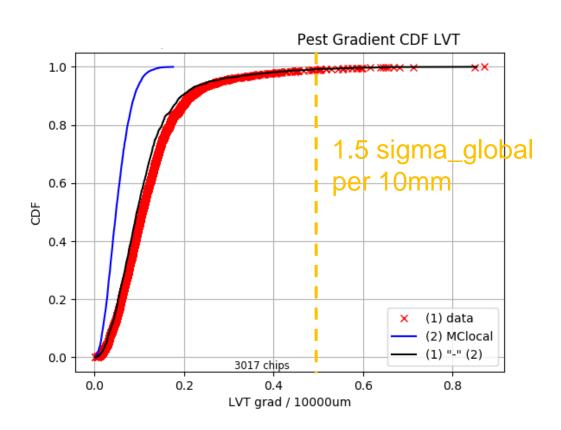
- Over 10mm distance, the worst case gradient change is 0.5, i.e. 1.5 sigma_global.
- (1)"-" $(2) = \sqrt{data^2 MClocal^2}$ is a subtraction of the estimation uncertainty for statistical uncorrelated random variation in the devices.
 - => The estimation error due to local variation does not significantly reduce the magnitude of the observed maximum gradient.

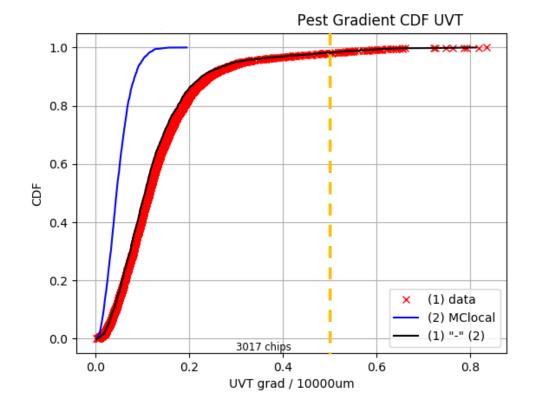


CDF of Gradient= $\sqrt{gradx^2 + grady^2}$ (LVT and UVT)

LVT > 2.4 sigma_global gradient / 10mm

UVT > 2.4 sigma_global gradient / 10mm







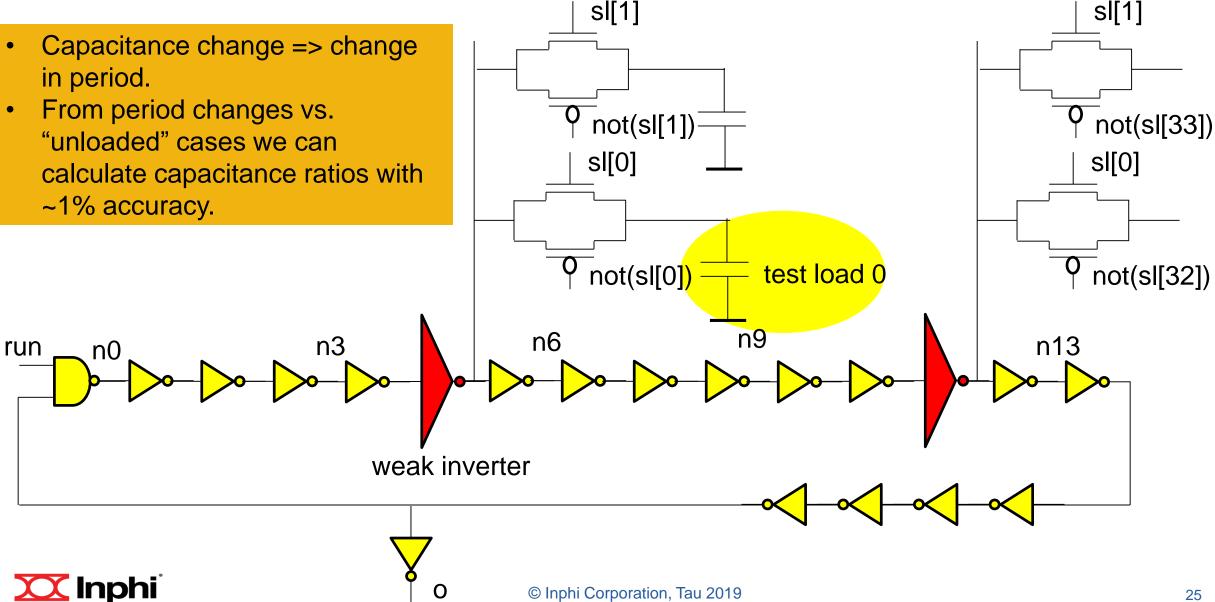
Layer-to-Layer Interconnect Variation

- Previous work from **T** 2016: "Layer-to-layer interconnect variation is a significant but unmodeled source of hold time optimism in conventional BEOL corner models"

 Christian Lutkemeyer, Ali Anvar, Broadcom
- Interconnect variation continues to be a growing concern and with the increased resistance it is unfortunate that modeling has not been significantly improved after 3 years!



Christian Lutkemeyer, Ali Anvar, Broadcom Interconnect Variation: ITCRO Top Level Schematic



Test Loads

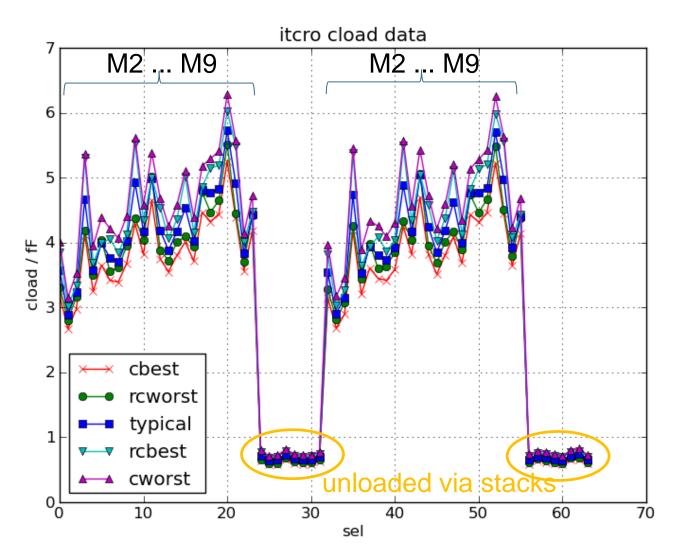
■ For each layer M2-M9 we implement 3 test structures (Example for a vertical layer)



test load, load capacitance approximately 4fF total (~20um wire) additional routing to connect to the test load has to match, including the via stack



Extracted Test Load Capacitance vs. Interconnect Corners

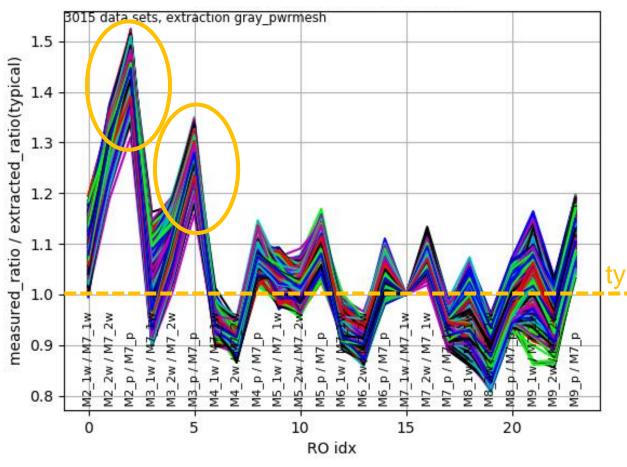


- Test structures are repeated once.
 This allows us to reduce random variation in the measurements from mismatch in the transmission gate VT.
- There is significant capacitance variation over the various interconnect corners (-13% to +15%).
- Interconnect variation can have a significant impact on dynamic performance.



Ratio of Ratios vs. M7 (Measured Ratio / Extracted Typical Ratio)

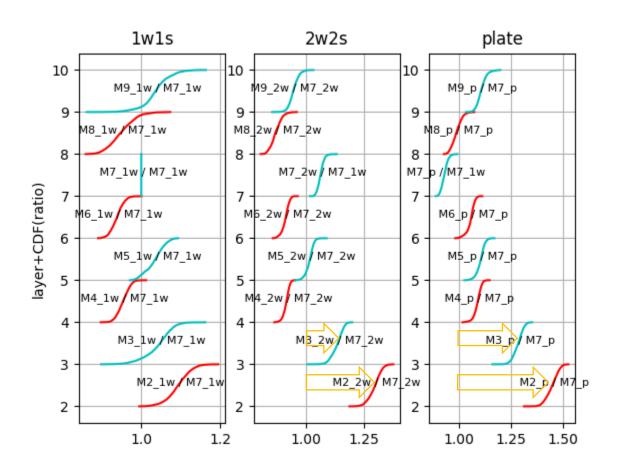
Ratio of Ratios vs. M7



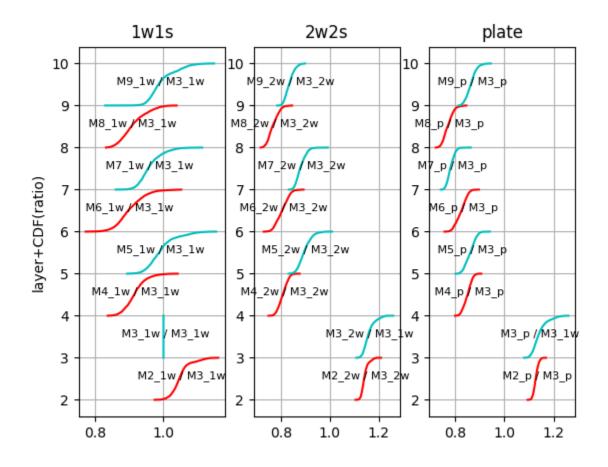
- We see a significant increase for the M2_2w / M7_2w, M2_p / M7_p, M3_2w / M7_2w, M3_p / M7_p ratios.
 - We suspect that the extracted capacitance values are missing capacitances to the bottom (inside the standard cells). The "gray" box extraction may not work properly?
 This is consistent with increased impact on M2 vs. M3 as M2 is closer to the cells.

Cumulative Distribution Functions of Interconnect Ratios vs. Typical Ratios

Normalized to M7 as Reference Layer



Normalized to M3 as Reference Layer

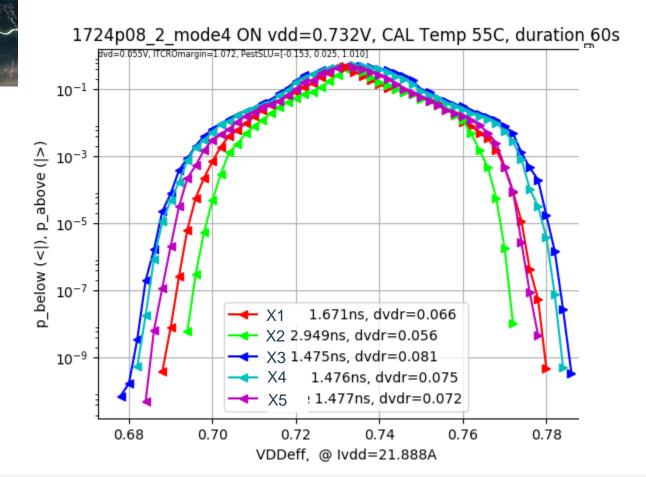




On-Chip Supply Variability

- τ 2017
 - Christian Lutkemeyer:
 Rogue Waves, On-Chip Power Integrity,
 and Static Timing Analysis
- VDD is the variable with the biggest impact on digital performance.
- Quantifying it accurately with high confidence in simulations is impossible.
- It can vary significantly
 - for different clock frequencies.
 - between ATE, socketed board, soldered board, or customer boards.

Measured data





Summary of the Presented Data

- We presented a fitting equation for Process Monitor RO delays that can be used to map dynamic RO speeds to a process estimate P_{est}.
 - By analyzing early split lot and production data we showed that our process
 - appears to have significant skew between different VT classes (+/- 1.5 sigma_global).
 - has intra-die performance gradients of up to 2.4 sigma_global / 10mm
- Measurements of different interconnect patterns show significant layer-to-layer variation which is completely ignored in fully correlated interconnect models in STA.
- Supply noise on a complex SOC is a significant factor that drives dynamic worst-case performance.



Wishes and Conclusions

Wishes

- It would be nice if foundries could expand WAT test structures to measure dynamic digital performance in line with customer use models and share this data early.
- Minimize skew between VTs in the process window so the selection of test candidate chips for DVT provides higher confidence for production ramp-up.

Conclusions

- We still have a significant number of variables in timing sign-off that are highly idealized and correlate poorly with manufactured dies
 - Supply voltage, interconnect, SI coupling (aggressor windows), blanket margins, spatial variation.
- There is no slack wall on individual dies.
 - High sigma confidence slack modeling for setup paths seems overkill in light of so many other rough parameters.



T Workshop Expansion Opportunities

- Develop embedded test circuits to measure wire capacitance, wire resistance, and via resistance.
- Collaborate to agree on meaningful dynamic test structures to classify silicon.
 - Standardized dynamic silicon test IP.
- Embedded supply voltage measurement structures.
- Production Testing:
 - Accurate understanding of how much supply noise is created during test, and how that noise is different between the ATE and the product soldered on the board.
 - How good are critical path test vectors in rejecting parts with marginal dynamic performance?
 - Limited test times reduce the probability that critical paths are evaluated with worst-case coupling and at a worst-case supply voltage.



Know Your Chips!





Thanks!

