LAYER-TO-LAYER INTERCONNECT VARIATION IS A SIGNIFICANT BUT UNMODELED SOURCE OF HOLD TIME OPTIMISM IN CONVENTIONAL BEOL CORNER MODELS

Christian Lütkemeyer and Ali Anvar

ACM International Workshop on Timing Issues in the Specification and Synthesis of Digital Systems.
March 10-11, 2016

Lütkemeyer+Anvar, Tau 2016, © 2016 Broadcom Corporation. All rights reserved.

MOTIVATION

BROADCOM.

- **Circuit timing depends on CMOS device properties <u>and</u> interconnect properties.**
  - Foundries provide a lot of detailed information about the devices (corners, random variation).
    - Device split lots are part of the device qualification process.
  - Much less detail is available for interconnect.
    - No interconnect corner split lots are available (not practical).
- **With continued scaling, the properties of interconnect become more important. Wire resistance increases, while capacitance per unit length stagnates.**
- **STA has gotten quite sophisticated in the modeling of device variation.**
  - AOCV, POCV, LVF
- **Interconnect modeling has not kept up.**
  - In 40 nm, fully statistical modeling of devices and interconnect was developed (Sensitivity SPEF) but this technology was put back on the shelf.
- **We have implemented test circuits to study interconnect layer variation. The results show strong layer-to-layer variation that cannot be modeled with the fully-correlated foundry interconnect corner models.**

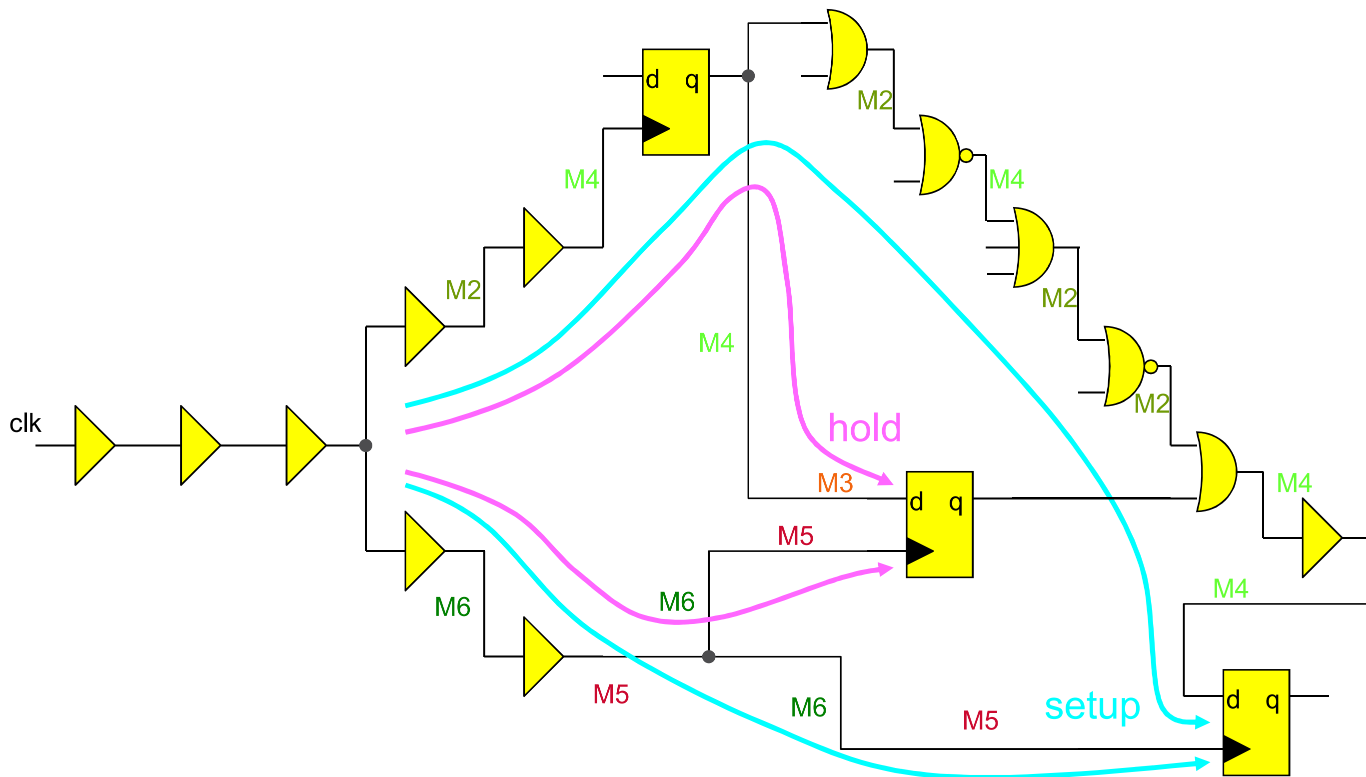
 © 2016 Broadcom Corporation. All rights reserved.

OUTLINE

- **The #1 concern with interconnect variation: Hold failures.**
- **Foundry interconnect layer models and modeling of interconnect in STA.**
- **Interconnect test structures.**
- **Review of the metrics that we use to measure interconnect variation.**
- **Review of interesting test data.**
- **Conclusions.**

Lütkemeyer+Anvar, Tau 2016, © 2016 Broadcom Corporation. All rights reserved.

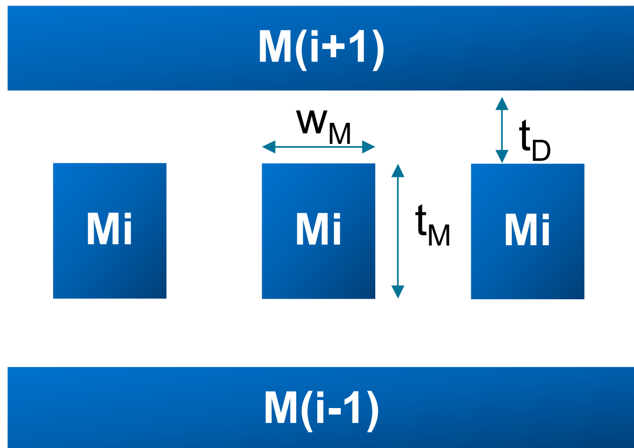
#1 CONCERN: POOR CLOCK CELL MATCHING AND INTERCONNECT VARIATION CAN LEAD TO HARD-TO-DETECT HOLD FAILURES



- Hold robustness has to be built into a design.
- Screening out marginal parts is not reliable as the screening process during a production test is limited (VDD, temperature, vectors).
- Aging during lifetime can push marginal parts over the edge.
- Chips that pass a hold test only marginally are likely to be unreliable in customer's systems.
- => Field returns, loss of trust.

Lütkemeyer+Anvar, Tau 2016, © 2016 Broadcom Corporation. All rights reserved.

FOUNDRY INTERCONNECT MODELS AND STA USE



Corner	Metal Width w_M	Metal Thickness t_M	Interlayer Dielectric Thickness t_D
Cworst	+	+	-
RCbest	+	+	+
typ	=	=	=
Cbest	-	-	+
RCworst	-	-	-

- **Interconnect layers have at least three variable parameters per layer: metal width, thickness, and interlayer dielectric thickness.**
  - N layers have 3N parameters. => 2^{3N} corners
- **In the foundry interconnect models, the variable parameters on all layers are pushed into identical corners, i.e. full correlation is assumed.**
  - => Five parasitic corners: Cworst, RCbest, typ, Cbest, RCworst.
- **In STA, foundries recommend to analyze and fix timing over all five foundry corners.**
  - <u>Pessimistic</u> in scenarios where gate delay is matched with interconnect dominated delay.
  - <u>Optimistic</u> in blended scenarios if significant layer-to-layer interconnect variation is present in silicon.

Lütkemeyer+Anvar, Tau 2016, © 2016 Broadcom Corporation. All rights reserved.

SIMPLIFIED SCHEMATIC OF THE PRECISION INTERCONNECT MEASUREMENT CIRCUIT

en[1]

$C_{viastack}+C_{load1}$

R_{TG1}

en[0]

$C_{viastack}$

R_{TG0}

start o_0 o_1 o_2 o_3 o_4 o_5 o_{n-1}

fb

Description:

- Weak inverter (o_2) creates a small current.
- Different test loads can be connected through transmission-gates to o_2.
- The added loads increase the RO period T proportional to the additional load divided by the unknown current from o_2.
- We can calculate capacitance ratios that are independent of the unknown current:

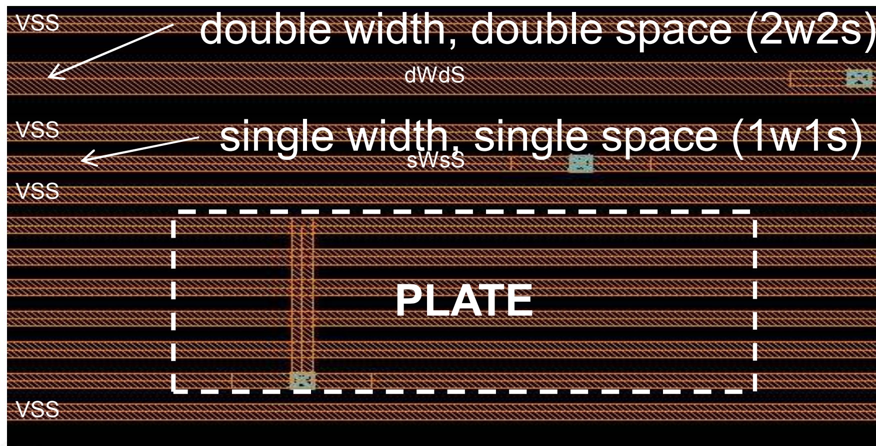
$$\frac{C_i}{C_j} = \frac{T(en=i)-T(en=0)}{T(en=j)-T(en=0)}$$

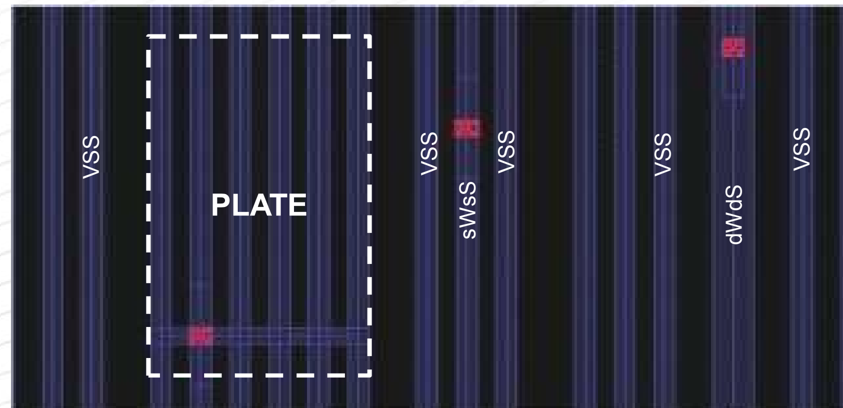
- The increase of the period has a small uncertainty due to local capacitance and resistance variation of the TGs.

Lütkemeyer+Anvar, Tau 2016, © 2016 Broadcom Corporation. All rights reserved.

EXAMPLE OF METAL TEST LOADS

- Ring oscillator will be active once “start” is “1”. Enable pins will engage a particular metal load with the ring as described below.
- In 6Mx version of the IP, en<18:21> does not add any load but the mux itself is present and added diffusion load will impact output frequency.

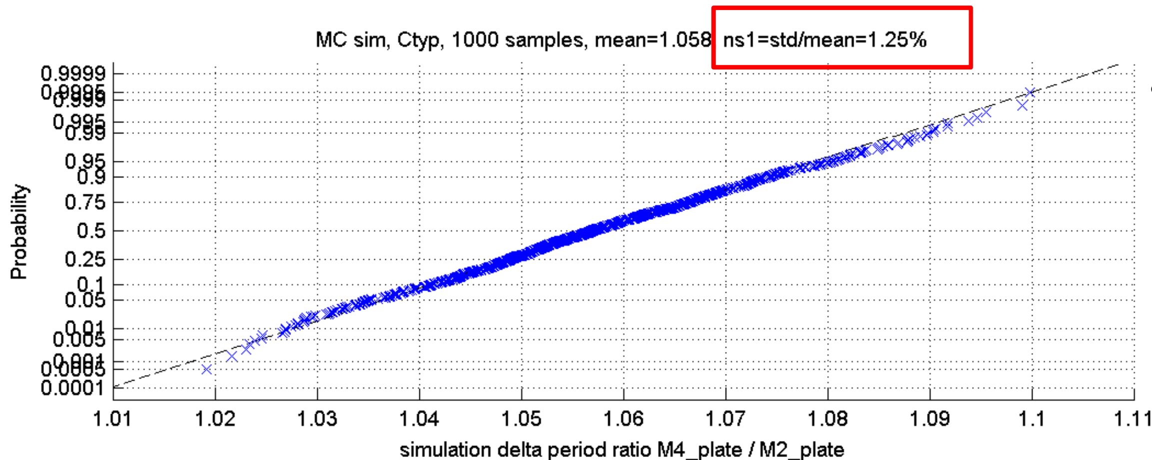


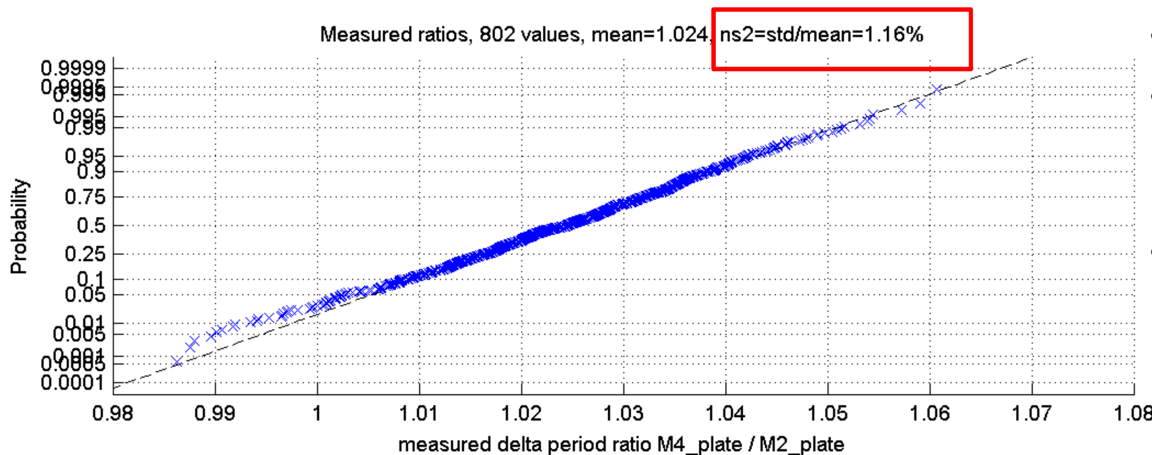


Enable pin	Metal layer	Cap type
en<0>	Metal 2	Plate
en<1>	Metal 2	Min width / Min spacing
en<2>	Metal 2	Double width / Double spacing
en<3>	Metal 3	Plate
en<4>	Metal 3	Min width / Min spacing
en<5>	Metal 3	Double width / Double spacing
en<6>	Metal 4	Plate
en<7>	Metal 4	Min width / Min spacing
en<8>	Metal 4	Double width / Double spacing
en<9>	Metal 5	Plate
en<10>	Metal 5	Min width / Min spacing
en<11>	Metal 5	Double width / Double spacing
en<12>	Metal 6	Plate
en<13>	Metal 6	Min width / Min spacing
en<14>	Metal 6	Double width / Double spacing
en<15>	Metal 7	Plate
en<16>	Metal 7	Min width / Min spacing
en<17>	Metal 7	Double width / Double spacing
en<18>	Metal 8	N.A.
en<19>	Metal 8	N.A.
en<20>	Metal 8	N.A.
en<21>	NA	Unloaded

Lütkemeyer+Anvar, Tau 2016, © 2016 Broadcom Corporation. All rights reserved.

NORMALIZED STANDARD DEVIATIONS NS1 AND NS2



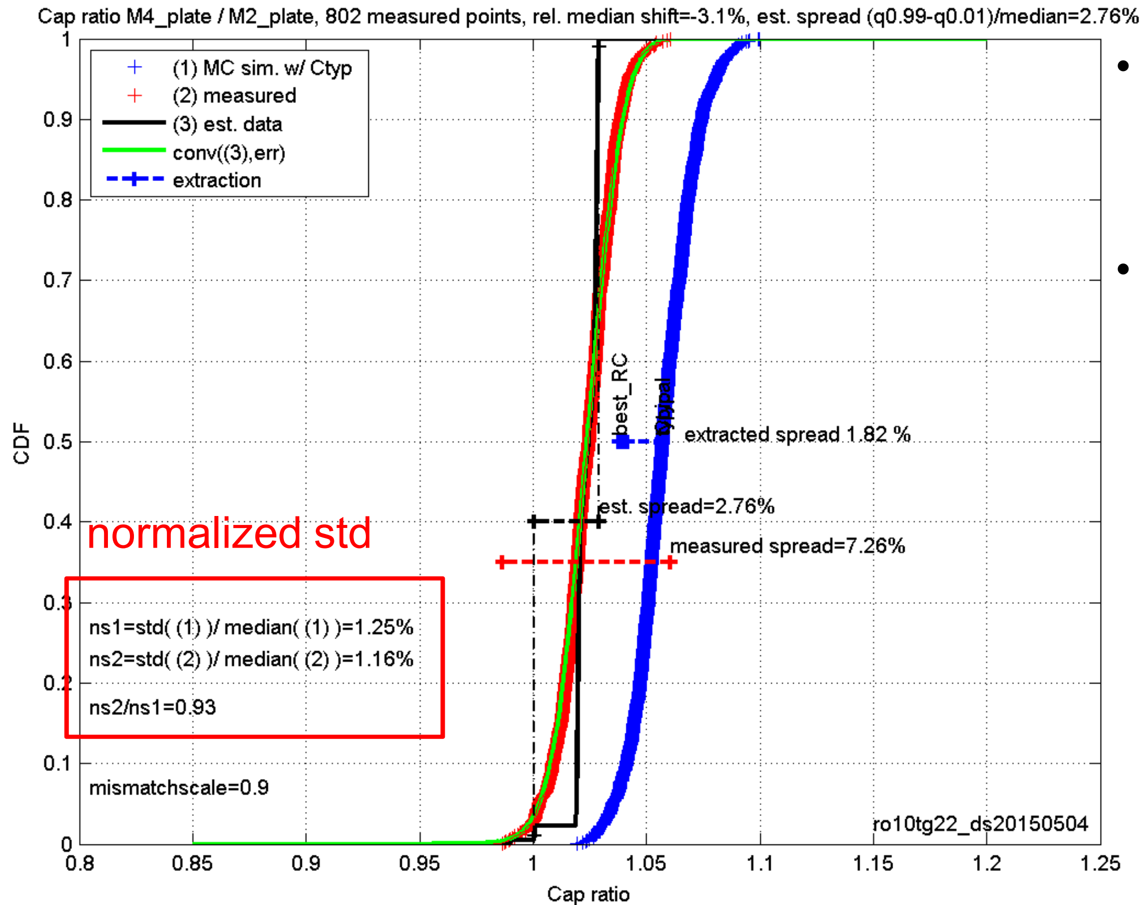


- Monte Carlo simulation with constant parasitic capacitances is used to estimate the measurement error in the capacitance ratio data due to local and global device variation.

- Example M4_plate /M3_plate:
- Mean of the data aligns well (simulated mean=1.058, measured mean=1.024).
- Normalized standard deviation of measured data is a bit smaller than simulated data (ns2=1.16%<ns1=1.25%)
=> process mismatch appears a bit better in silicon than in SPICE.

Lütkemeyer+Anvar, Tau 2016, © 2016 Broadcom Corporation. All rights reserved.

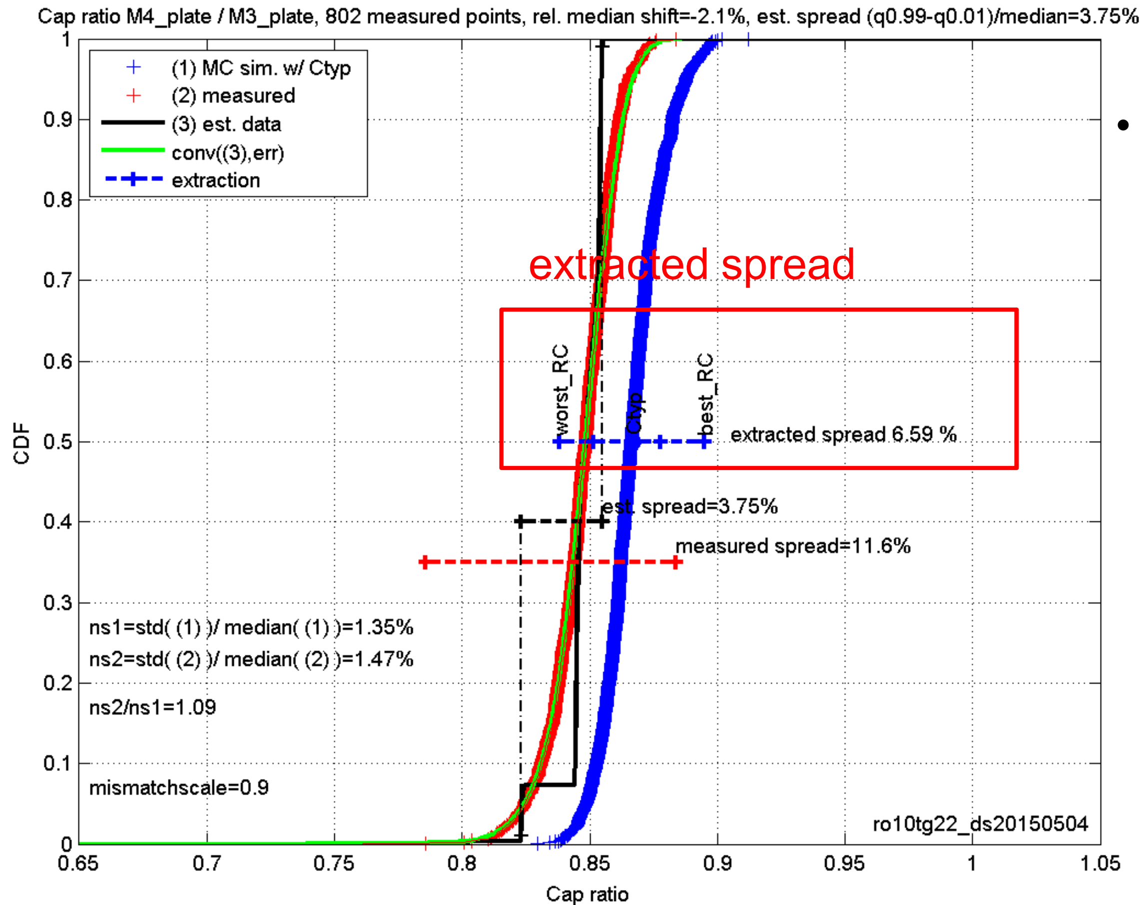
NORMALIZED STANDARD DEVIATIONS, M4 PLATE / M2 PLATE



- The normalized standard deviation ns1 characterizes the measurement error due to device mismatch.
- ns2 is the normalized standard deviation of the measured data. It includes device mismatch as well as interconnect variation.

 © 2016 Broadcom Corporation. All rights reserved.

EXTRACTED SPREAD, M4 PLATE / M3 PLATE

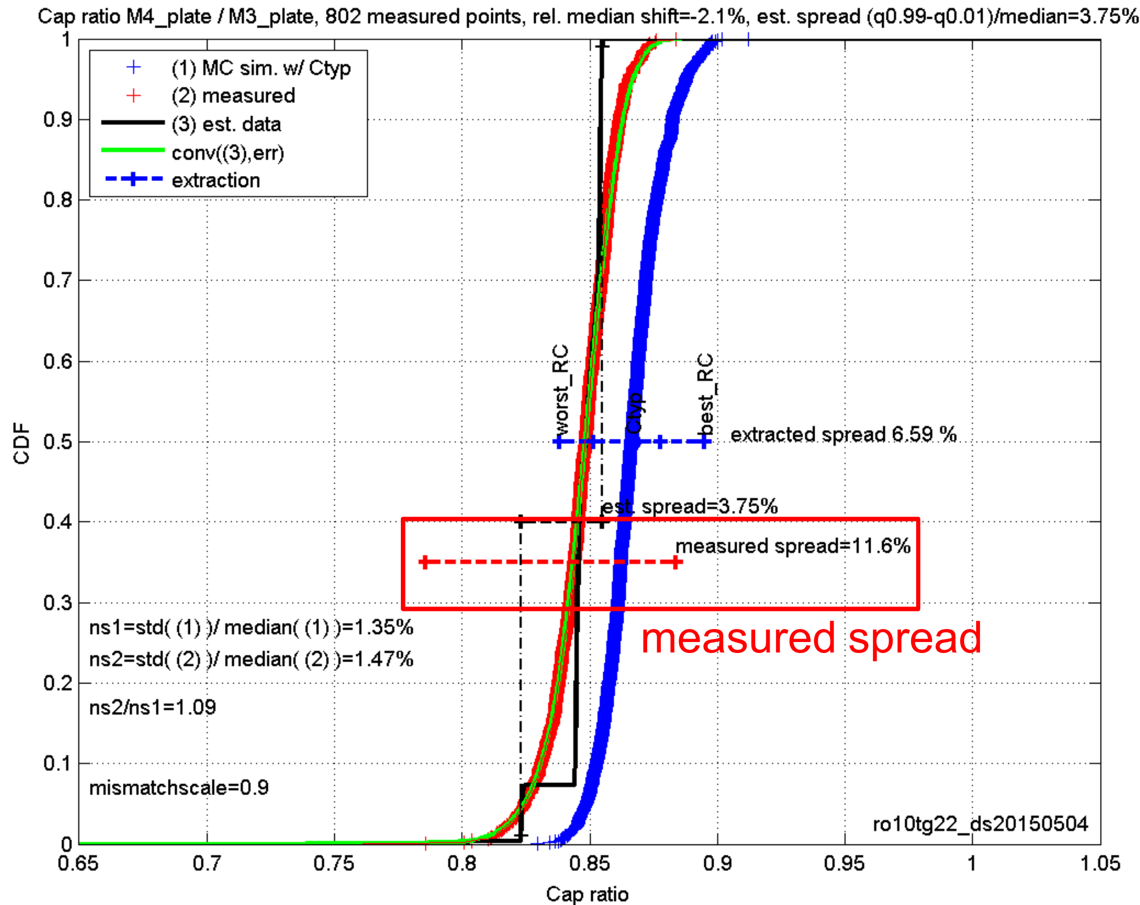


- The extracted spread is determined from corner simulations of the RO22 with typical silicon and the extreme foundry interconnect corners (worst_RC, worst_C, best_RC, best_C, typ_C). In each interconnect corner we calculate the Capacitance ratio.

$$extracted_spread = \frac{capratio_{\max} - capratio_{\min}}{(1/2)\cdot(capratio_{\max} + capratio_{\min})}$$

 © 2016 Broadcom Corporation. All rights reserved.

MEASURED SPREAD, M4 PLATE / M3 PLATE

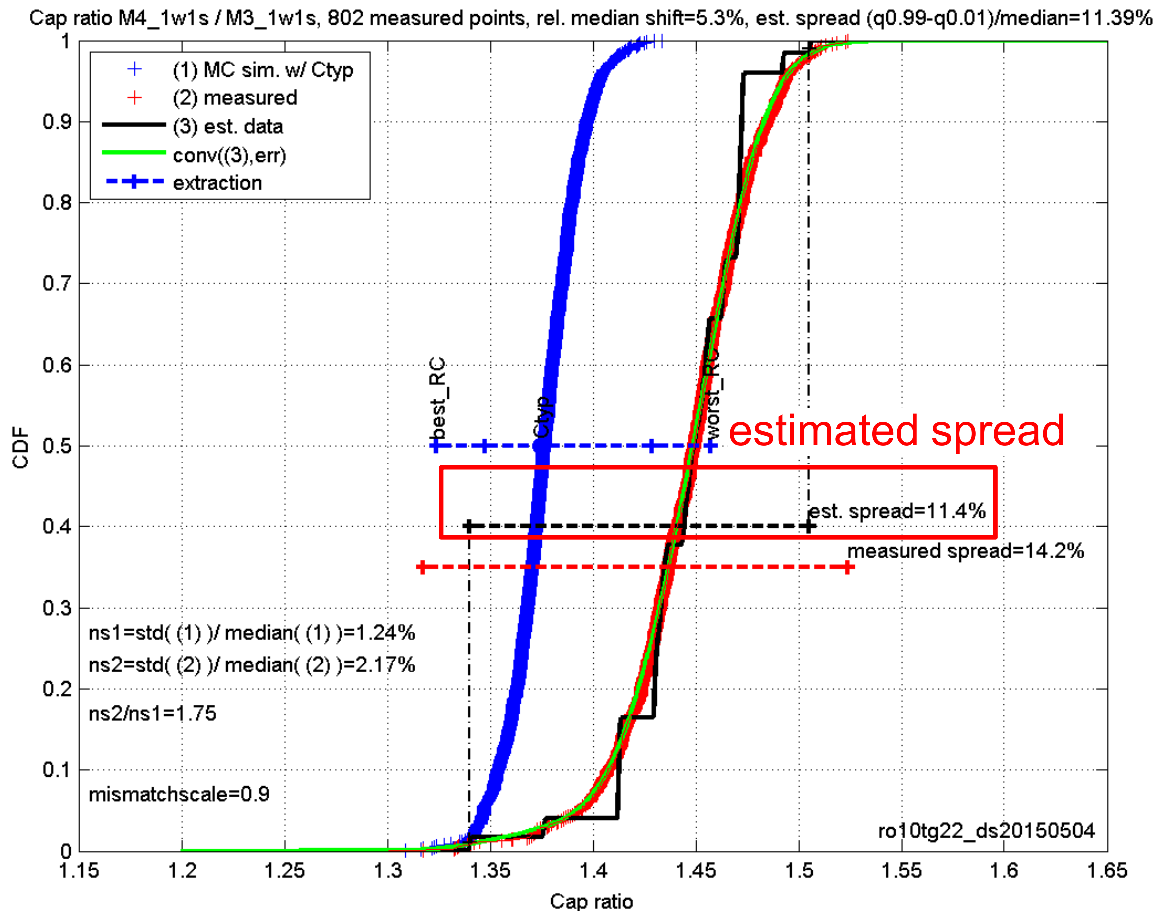


- Measured spread:

$$measured_spread = \frac{measured_capratio_{max} - measured_capratio_{min}}{mean(measured_capratios)}$$

 © 2016 Broadcom Corporation. All rights reserved.

ESTIMATED SPREAD OF THE TRUE ON-CHIP DATA

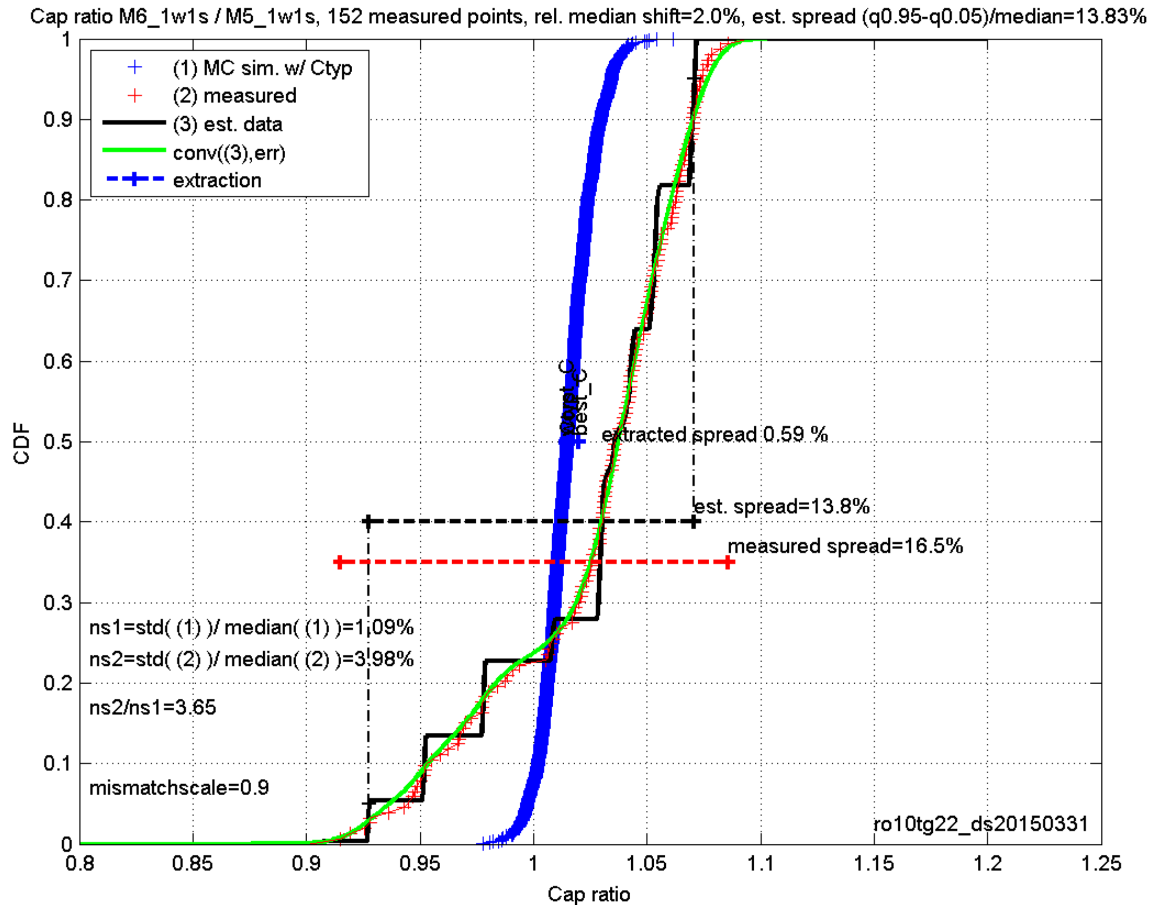


- Estimated spread:
  The measurement circuit adds observation uncertainty. The distribution of the actual interconnect capacitance ratios in silicon is tighter than the measured spread, if sufficient data is available to obtain samples on the tails of the distribution. We estimate the distribution of the actual data by deconvolving an estimated observation error (err) from the CDF of the measured data.

$$estimated_spread = \frac{est_data_{q99\%} - est_data_{q01\%}}{median(measured_capratios)}$$

Lütkemeyer+Anvar, Tau 2016, © 2016 Broadcom Corporation. All rights reserved.

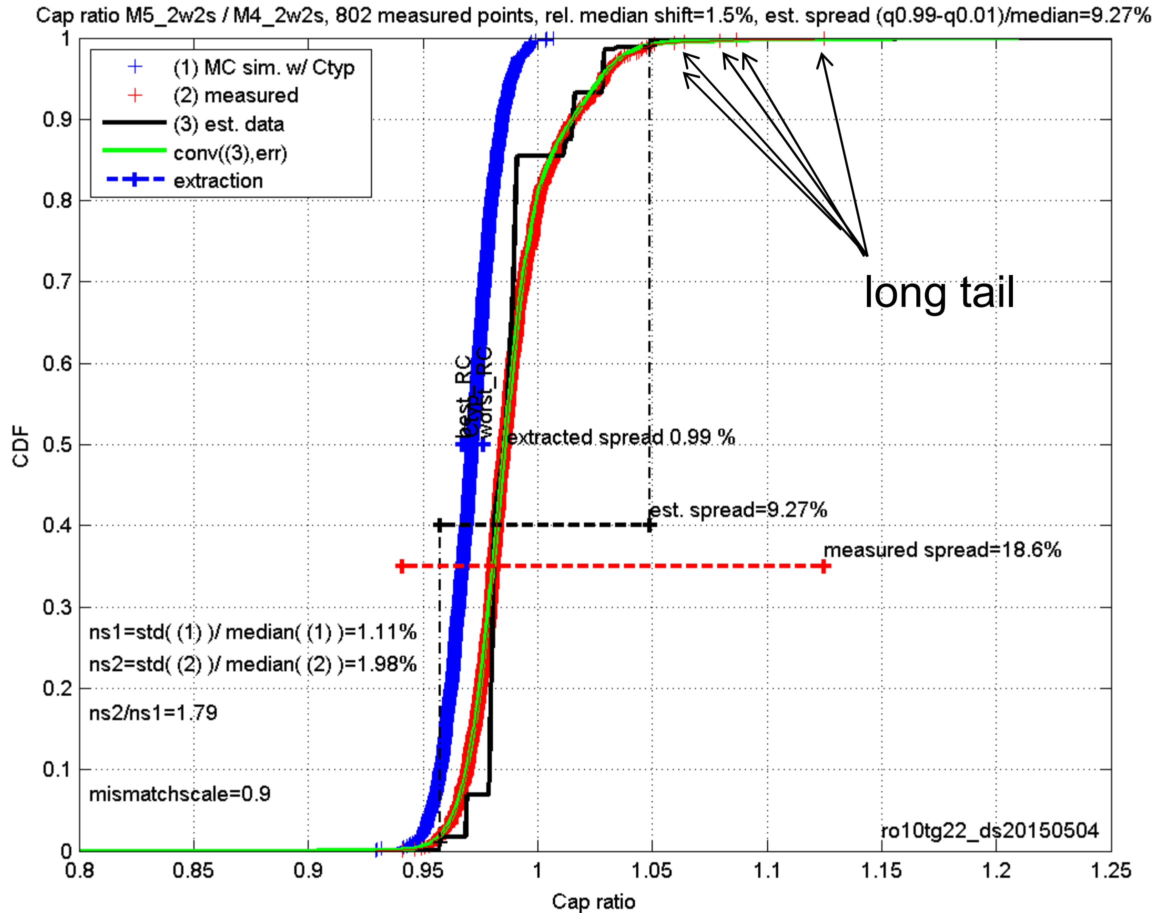
EXAMPLE: M6 1W1S / M5 1W1S



- Estimated spread of the cap ratio is 13.8%.
- The extracted spread with the fully correlated foundry corners is only 0.6%!
- => Closing timing in the foundry corners does not cover the layer-to-layer variation that is observed.

Lütkemeyer+Anvar, Tau 2016, © 2016 Broadcom Corporation. All rights reserved.

M5 2W2S / M4 2W2S



- Estimated spread for M5/M4 is significantly larger than for M3/M2 (9.3% vs 5%).
- There is a long tail that could indicate the spread is worse than estimated.
- The extracted spread is very tight (1%).
- Foundry corners do not cover the layer-to-layer variation that is observed.

Lütkemeyer+Anvar, Tau 2016, © 2016 Broadcom Corporation. All rights reserved.

SUMMARY

- **We have shown significant (>10%) layer-to-layer capacitance ratio variation in a mature 28 nm CMOS process on just a few wafers.**
- **The impact of such variation is not covered by the traditional, fully correlated, foundry interconnect corners.**
- **The timing impact of such large variations depends on the ratio of gate loads vs. wire loads on individual nets.**
  - Gate load dominated nets show small variations.
  - Interconnect dominated nets have larger variations.
- **Today's STA flows use RC delay derating to create margin against some interconnect variation.**
  - Effective for long nets with significant RC delay.
  - Ineffective on shorter nets with significant capacitance but small RC delay.
- **Other margins have to fill in to cover for the interconnect related variation of nets w/o significant RC delay.**
  - Gate delay derates that are covering for supply variation and aging.

 © 2016 Broadcom Corporation. All rights reserved.

SUMMARY II

- **To maintain robust hold margins in the latest technologies it would be highly desirable to be able to model the impact of layer-to-layer interconnect variation more specifically.**
  - This would allow us to separate the interconnect variation margin from other margins. We expect a benefit as the impact of interconnect variation can vary significantly for different connection scenarios (wire vs. gate load dominated).
- **We would like to get the SSPEF (IEEE Standard 1481-2009) resuscitated and supported in the latest STA tools to be able to better account for interconnect layer mismatch.**
- **More data from foundries on actual interconnect variation would be helpful.**

 © 2016 Broadcom Corporation. All rights reserved.

Thank You!

Broadcom Proprietary and Confidential. © 2015 Broadcom Corporation. All rights reserved.