

Mentor[®]
A Siemens Business

Cross-Boundary Inductive Timing optimization for 2.5D Chiplet-Package Co-Design

MD Arafat Kabir¹, Dusan Petranovic², Yarui Peng¹
CSCE Dept., University of Arkansas, Fayetteville, AR¹
Mentor, a Siemens Business, Fremont, CA²

UNIVERSITY OF
ARKANSAS



Introduction



- ❑ **Package becomes increasingly critical in post-Moore's Law era**
 - Transistor scaling is saturated, and chips are reaching reticle limit.
 - 2.5D and 3D packages provide high bandwidth and compact size.
 - Novel design techniques like plug-and-play, Drop-in, Hardware security
 - Heterogeneous integration capabilities (AMD Milan-X, Intel Lakefield)
 - Supports large systems with tens of Known-Good-Dies
- ❑ **Need for a cross-boundary package-aware design strategy**
 - Interactions between the package and chiplets are significant
 - Package inductance is expected to play significant role on performance and signal integrity.
- ❑ **Objectives**
 - Study of RDL inductance impact on 2.5D system performance
 - A cross-boundary inductance aware timing optimization flow



Need for Inductance-Aware Flow



Need for cross-boundary inductance-aware timing optimization

- Package nets are the bottlenecks in a 2.5D system [1]
- Large I/O drivers are used for inter-chiplet communication [2]
- Bigger driver mean large power and area
- Custom drivers can save power, area, and improve performance
- Requires consideration of all circuit elements (RLC) and careful analysis and optimization

Limitation of existing flows

- Support for only RC elements in STA tools
- Driver optimization based on capacitive load only
- Timing impact of interconnect inductance is completely ignored

[1] M. A. Kabir, D. Petranovic and Y. Peng, "Coupling Extraction and Optimization for Heterogeneous 2.5D Chiplet-Package Co-Design," *IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, 2020, pp. 1-8.

[2] Minah Lee, Arvind Singh, Hakki M.Torun, Jinwoo Kim, Sung Kyu Lim, Madhavan Swaminathan, and Saibal Mukhopadhyay, "Automated I/O Library Generation for Interposer-based System-in-Package Integration of Multiple Heterogeneous Dies," *IEEE Transactions on Components, Packaging, and Manufacturing Technology*. Vol. 10, No. 1, pp. 111-122, 2020



Interconnect Delay Study



□ Interconnect model

- Inductance is modeled using the following partial inductance equation
- $k = l/r$, l is the length and r is the thickness of the wire
- At 2 GHz, skin depth of copper is 1.45 μm

RDL Parameter	Value
Width	10 μm
Spacing	10 μm
Thickness	1 μm
Resistance	0.05 $\Omega/\mu\text{m}$
Capacitance	0.068 fF/ μm
Partial Inductance	Equation (1)

$$L_{l,k} = \frac{\mu_0 l}{2\pi} \left[\ln(\sqrt{k^2 + 1} + k) - \sqrt{k^{-2} + 1} + \frac{0.9054}{k} + 0.25 \right] \quad (1) \quad [3]$$

[3] H. A. Aebischer and B. Aebischer, "Improved Formulae for the Inductance of Straight Wires," Advanced Electromagnetics, vol. 3, no. 1, pp. 31–43, 2014.

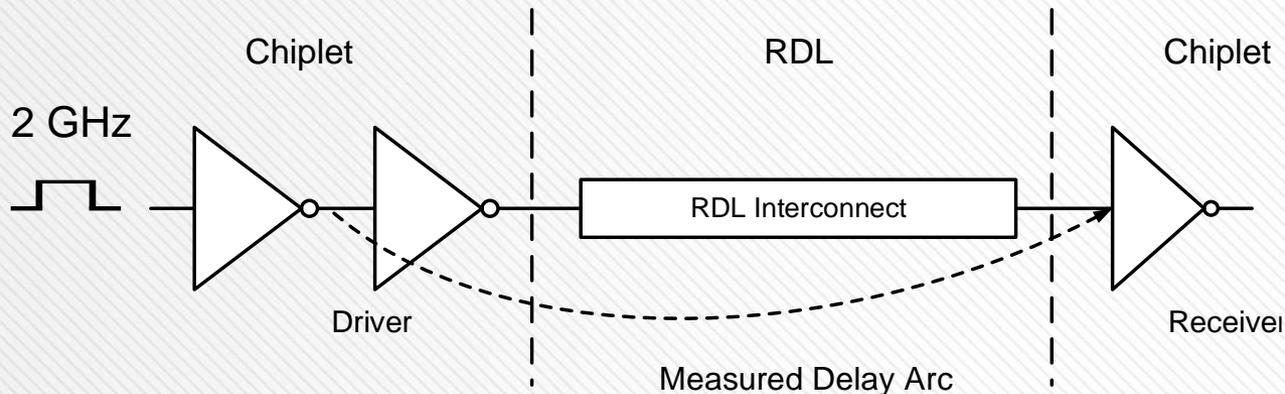


Interconnect Delay Study



Simulation setup

- Nangate45 cell library based on FreePDK45
- 2 GHz pulse source with 10 ps rise/fall time
- Ignoring the IO pads in this simulation
- Multiple simulation runs with different interconnect length
- Both RC and RLC models are simulated



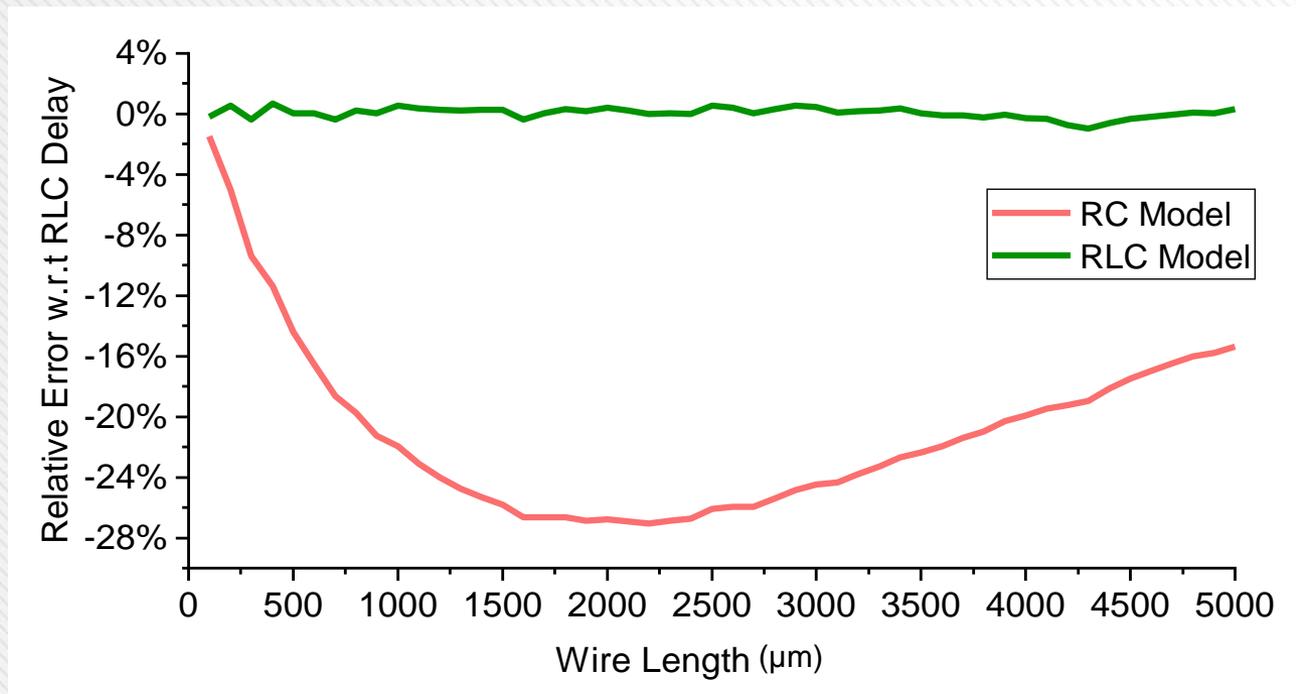


Simulation Result



❑ **Using only RC model can underestimate the propagation delay by approximately 30%**

- Following result is for INV_X16 as the driver
- Consistent with previous studies





Our RLC Delay Model



- **We are using a transmission line model to estimate the RLC delay**
 - The model equation is developed based on some previous studies [3]
 - RLC delay is approximated from the RC delay using a scaling factor
 - The scaling factor is later used in the parasitic scaling flow

Line Parameter	Definition
R_t	Total line resistance
C_t	Total line capacitance
L_t	Total line inductance
C_L	Total input capacitance of the receiver
ζ_{line}	Damping ratio of the line

$$scalingFactor = k + a\zeta_{line}^3 + b\zeta_{line}^2 + c\zeta_{line} + d\zeta_{line}^2 C_T$$

$$RLC\ Delay = scalingFactor \times RC\ Delay$$

$$C_T = \frac{C_L}{C_t}, \quad \zeta_{line} = \frac{R_t}{2} \sqrt{\frac{C_t}{L_t}}$$

[3] Y. I. Ismail and E. G. Friedman, "Effects of Inductance on the Propagation Delay and Repeater Insertion in VLSI Circuits," IEEE Transactions on Very Large Scale Integration Systems, vol. 8, no. 2, pp. 195–206, 2000.



Fitted Cells



□ All driver cells are simulated and fitted to the delay model

- Fitted parameters of some of the Nangate45 library cells
- RC model is equivalent of $k=1.0$ and $a=b=c=d=0$
- Larger the driver, larger is the deviation from the RC model due to reduced driver resistance.

$$scalingFactor = k + a\zeta_{line}^3 + b\zeta_{line}^2 + c\zeta_{line} + d\zeta_{line}^2 C_T$$

Parameter	INV			BUF		
	X1	X4	X16	X1	X4	X16
a	-0.023	0.132	3.312	0.004	-0.036	1.931
b	0.047	-0.242	-5.783	0.007	0.000	-3.788
c	-0.013	0.156	2.804	-0.006	0.048	2.085
d	-0.008	-0.136	-1.009	-0.007	-0.076	-0.561
k	1.003	1.001	0.961	1.004	1.008	0.938

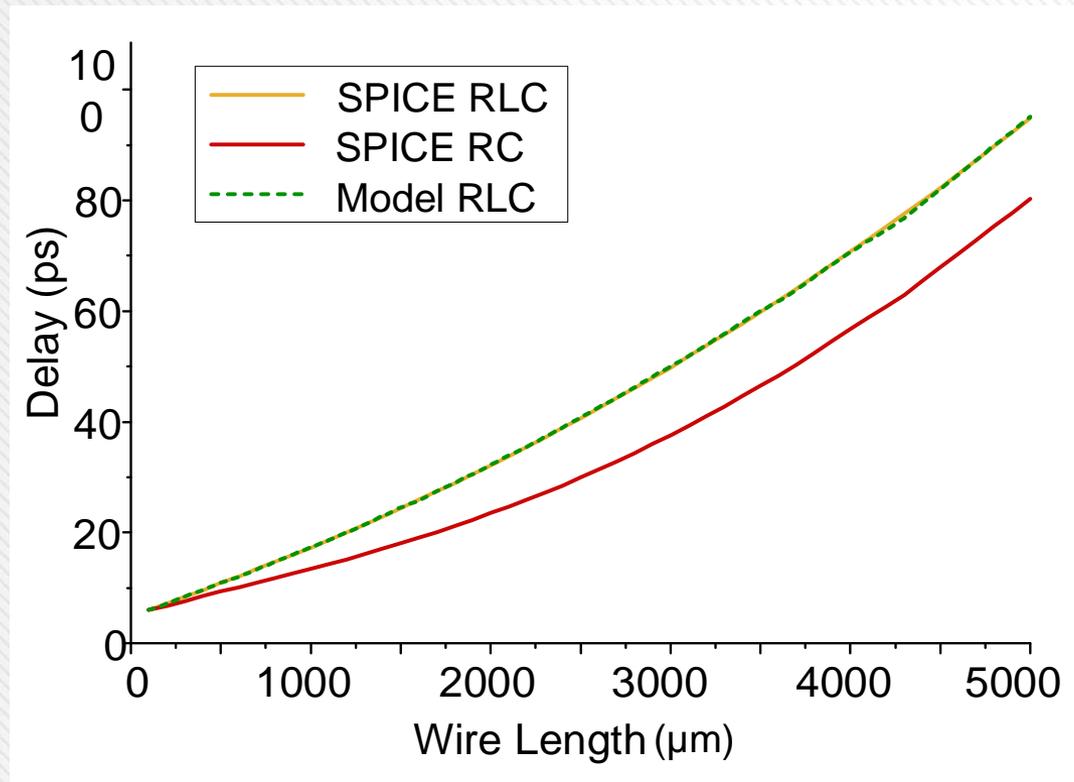


Our RLC Delay Model



□ **Our RLC delay model has only 1% error compared to SPICE simulation of the RLC interconnect model**

- The simulation covers up to 5 mm RDL wirelength

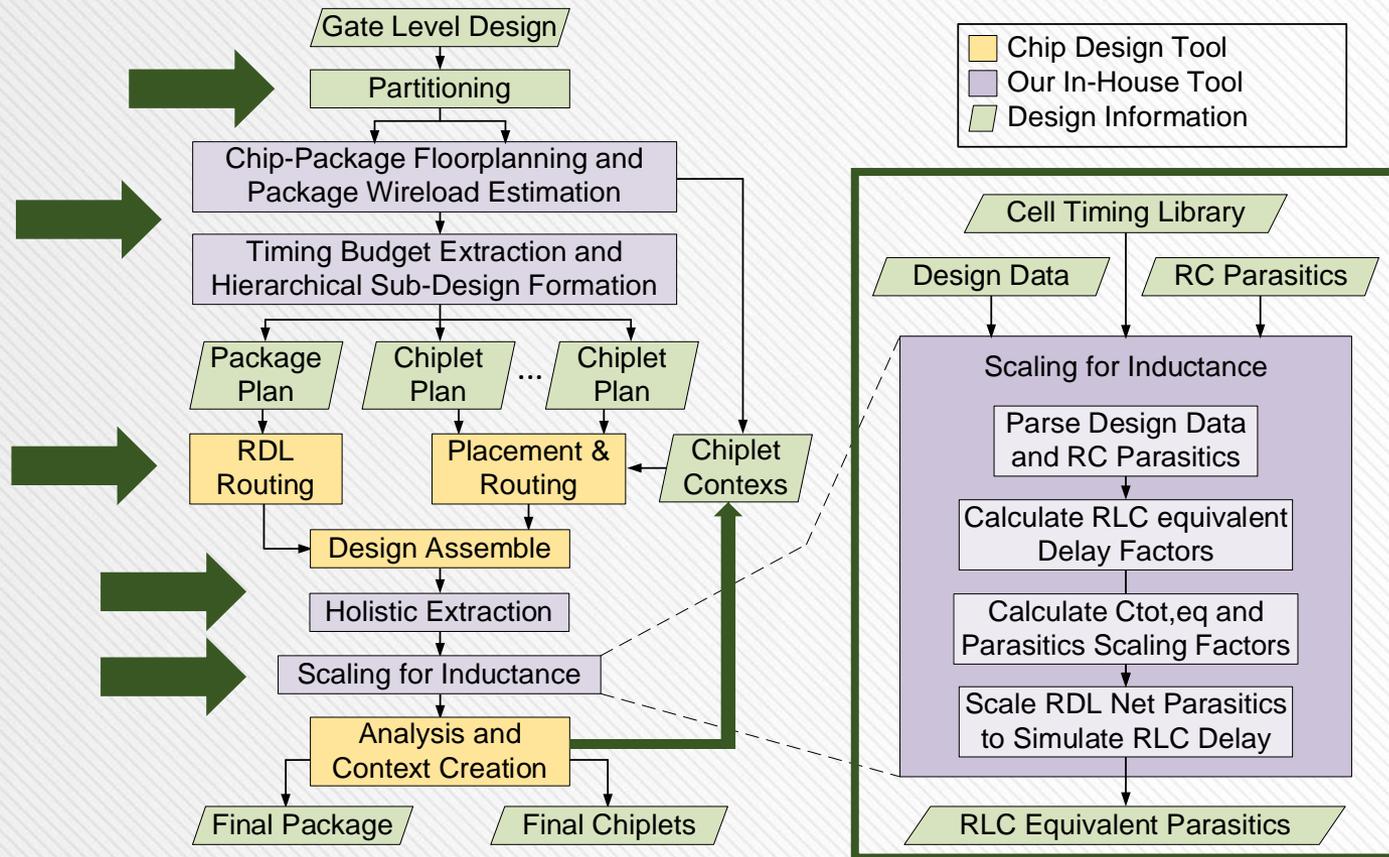




Our Holistic Flow



Exchange of cross-boundary design information in planning, design, analysis, and optimization steps



(a) Holistic Co-Optimization Flow

(b) Inductance Impact Modeling



Parasitics Scaling for Inductance



- ❑ **Industry standard tools does not directly support inductance in the STA and timing optimization steps**
 - Standard parasitics formats (like SPEF) support inductance.
 - STA tools (like PrimeTime) ignore the inductances in timing analysis.
 - PDK and timing optimization do not consider inductance either
 - Fundamental changes are required in existing tool flows for inductance

- ❑ **We come up with parasitic (RC) scaling as a direct solution**
 - No need to modify the existing timing analysis tool flow
 - Inherently compatible with the timing optimization flow

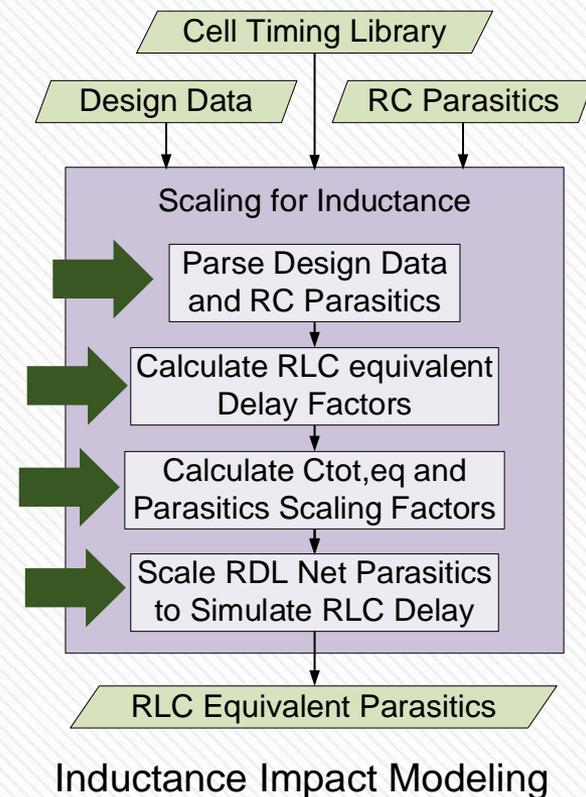


Parasitics Scaling Flow



□ RC parasitics are scaled to emulate RLC equivalent delay

- Our in-house tool performs the parasitic scaling
- Design data and RC parasitics are obtained from P&R tool and holistic extraction
- RLC delay for RDL wires is estimated using our RLC equivalent model
- The scaling factor to emulate RLC delay is determined
- Capacitance values of the RDL wires are scaled generate RLC equivalent parasitic





Scaling for Inductance



□ RLC equivalent parasitics is computed using equation (2)

- Cell delay depends on input transition and total output capacitance.
- Net delay is calculated using Elmore delay model.
- Total Elmore-delay is scaled if all capacitances are scaled keeping all resistances constant

$$\begin{aligned} RLC &= \text{cell delay} + \text{net delay} \\ &= LUT(C_{tot,eq}, t_r) + \text{scalePar} \times (RC \text{ net delay}) \end{aligned} \quad (2)$$

Where,

C_{tot} : Total Capacitance in the RC network,

t_r : Input transition time of the driver cell,

$C_{tot,eq}$: Equivalent total capacitance required to simulate RLC delay,

LUT : Cell timing library look-up table

$\text{scalePar} : C_{tot,eq}/C_{tot}$

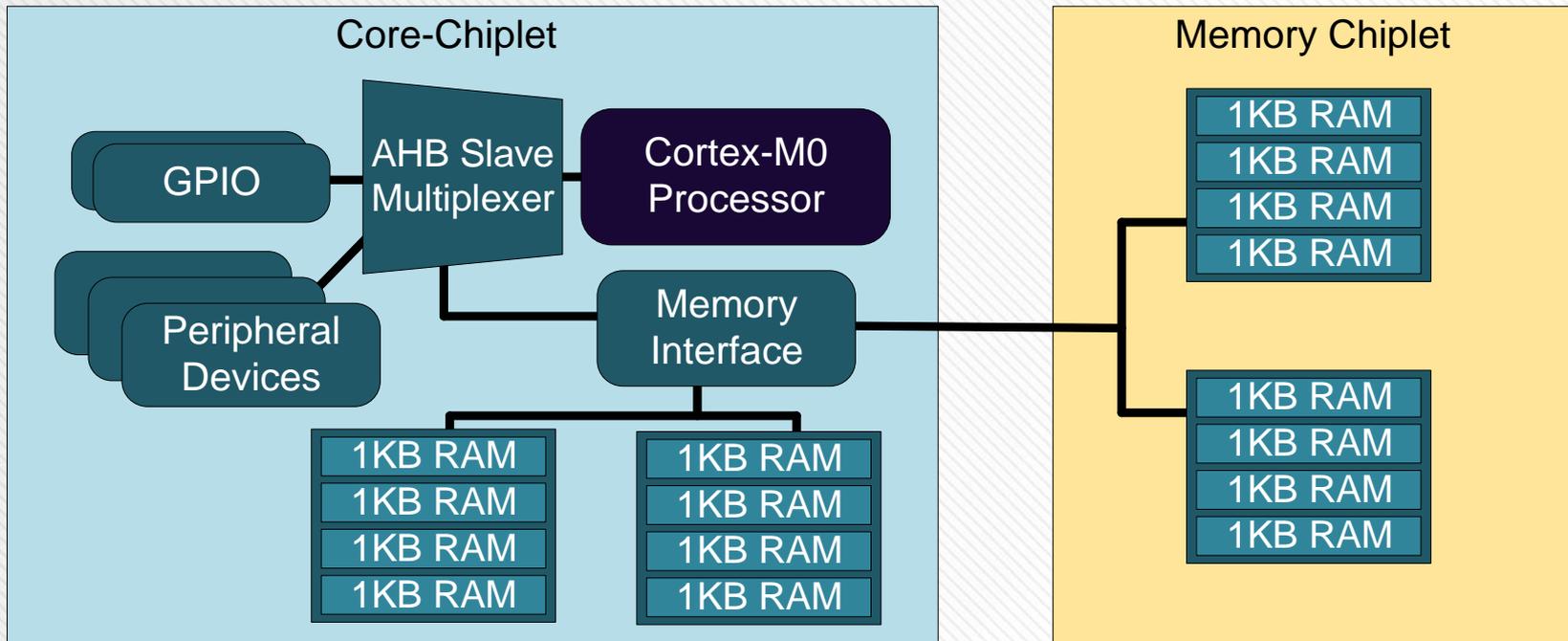


Experimental Study



ARM Cortex-M0 based micro-controller system

- Consists of an ARM Cortex-M0 core, 16 KB memory, and some common peripheral devices
- Two-chiplet system: Core and Memory
- The 16 KB memory is divided into two parts, 8 KB each.





Technology Settings



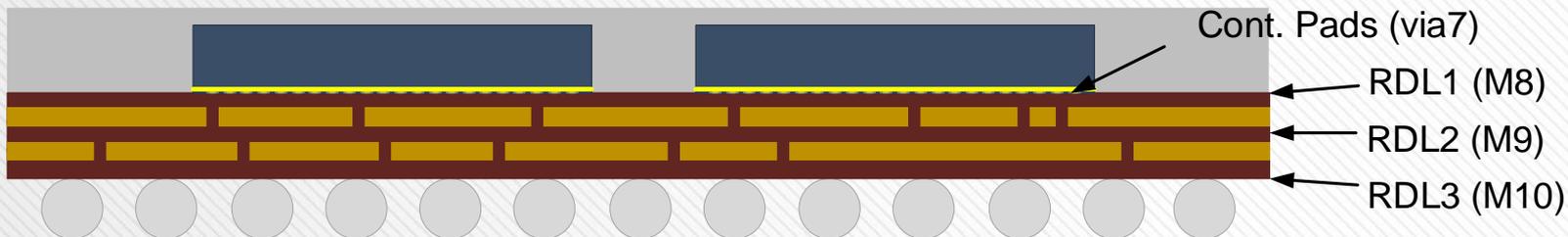
We use Nangate45nm as our PDK

- M1-M7 used for chiplet routing

We modify the top three layers to include 2.5D package RDLs

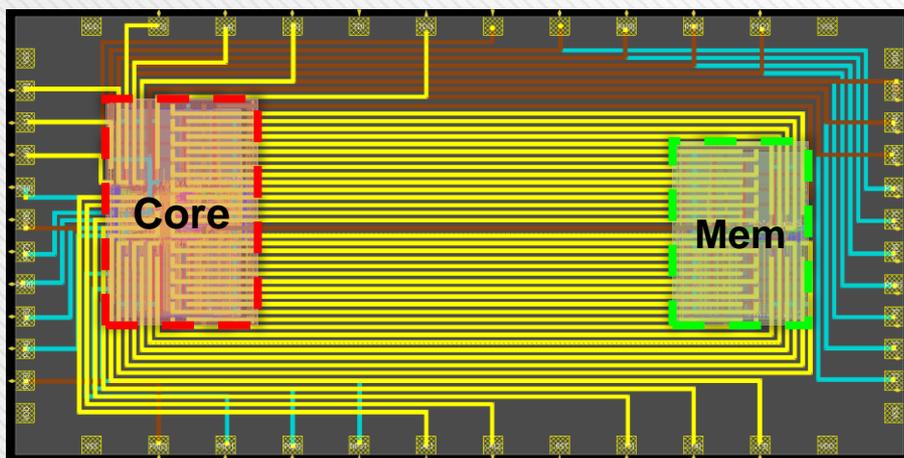
- Dimensions are similar to the TSMC 2.5D InFO technology

	M6	via6	M7	via7	RDL1	viar1	RDL2	viar2	RDL3
Height	2.28	3.08	3.9	7.5	12.5	17.5	22.5	27.5	32.5
Thickness	0.8	0.82	3.6	5	5	5	5	5	5
Width	0.4	0.4	2	5	10	10	10	10	10
Spacing	0.4	0.44	2	10	10	20	10	20	10

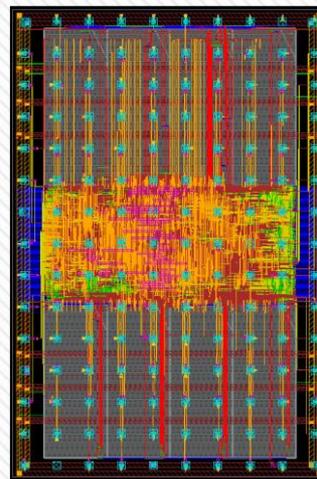


□ The system is implemented keeping the chiplets 1 mm apart

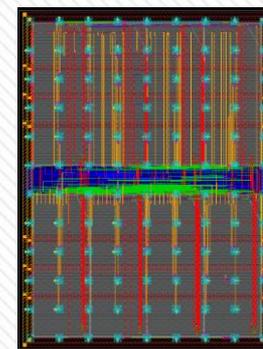
- A small system is easy to control for experimental study
- RDL wirelength varies between 1-2.50 mm
- Final system frequency of 300 MHz
- Two different designs are prepared,
 - Using holistic RC analysis and optimization flow: **RC-Design**
 - Using our proposed flow with parasitic scaling: **RLC-Design**



(a) Assembled 2.5D system with chiplets and the package together



(b) Core-Chiplet



(c) Memory Chiplet

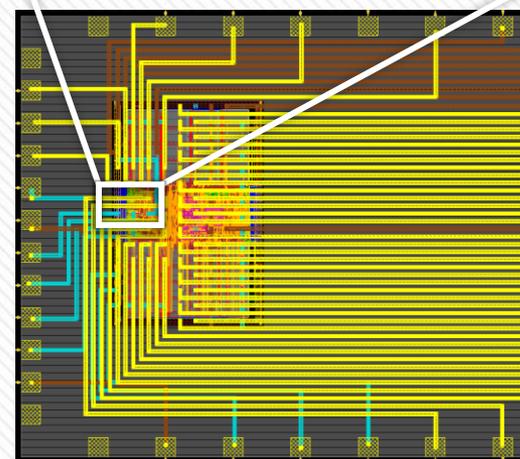
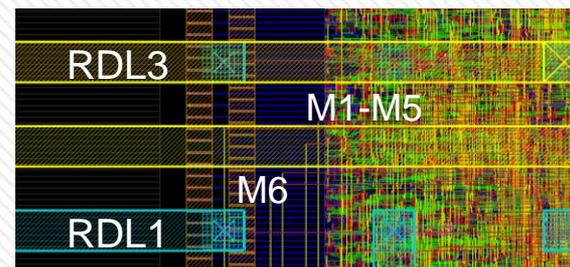


Holistic RC Extraction



❑ Package and chiplet designs are assembled for holistic extraction

- The extraction environment has everything together
- The extraction tool can capture the cross-boundary RC parasitics between the package and chiplets



Coupling Capacitance (CCAP)						
Metal Layer	M1-M5	M6	M7	RDL1	RDL2	RDL3
M1-M5	6116	413.1	38.45	57.60	10.13	7.316
M6	413.1	494.4	92.67	109.2	12.09	10.17
M7	38.45	92.67	41.11	18.32	2.097	2.354
RDL1	57.60	109.2	18.32	721.7	2646	45.63
RDL2	10.13	12.09	2.097	2646	750.2	2623
RDL3	7.315	10.17	2.353	45.62	2623	1135
Ground Capacitance (GCAP)						
Metal Layer	M1-M5	M6	M7	RDL1	RDL2	RDL3
Capacitance	21640	2142	288.9	2118	365.3	681.7

Assembled System SITY OF NSAS

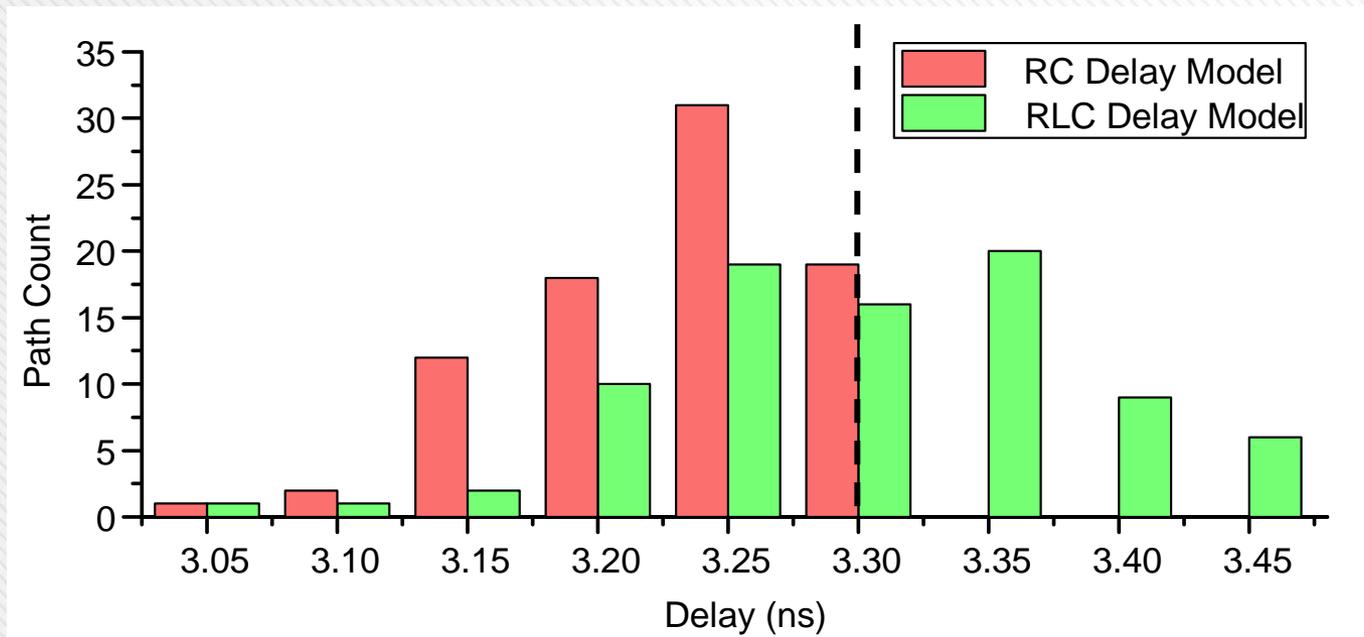


RC vs RLC Model Comparison



❑ RC only analysis and optimization keeps 35% of the paths in timing violation

- These violations remain undetected in RC analysis.
- The worst violation is by 0.15 ns
- Finished system will fail to run at nominal speed

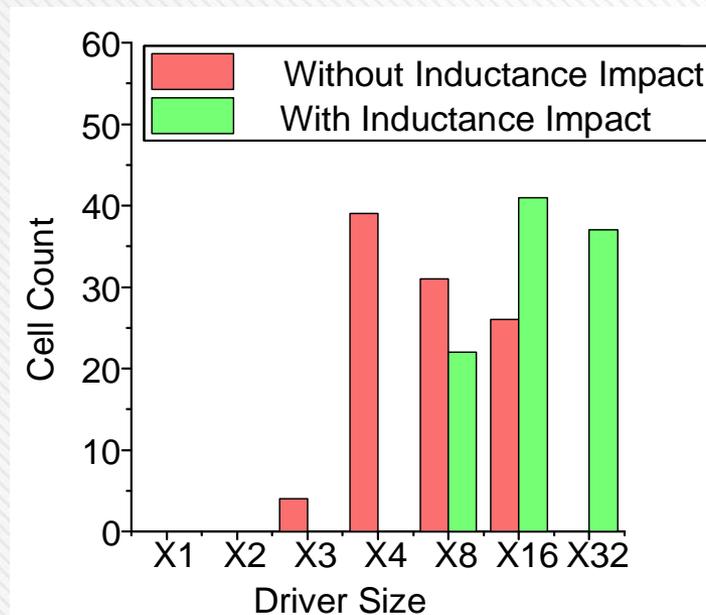




RDL Driver Optimization



- ❑ **Our optimization flow automatically adjusts drivers for inductance delay overhead**
 - Smaller drivers are used in the RC-Design.
 - Drivers are upsized in the to compensate for the inductance impact
 - This shift in driver size distribution is **ONLY** to compensate for the inductance overhead.



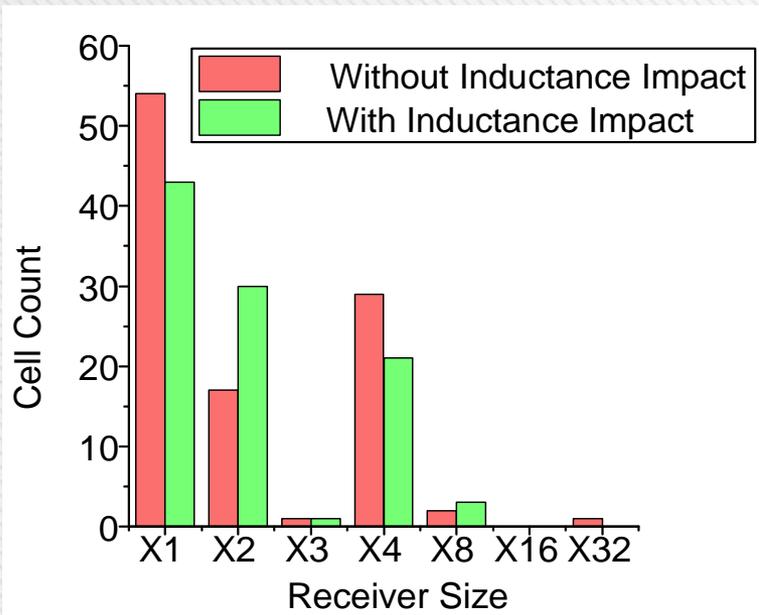


Receiver Optimization



Receiver cells are adjusted to reduce total load capacitance

- Shift in receiver distribution to reduce TOTAL input capacitance
 - Larger receiver cells are downsized
 - Many small receivers (X1) replaced with single a bit larger receiver (X2)
- Reduction in total path delay
 - The logic cell itself is downsized instead of inserting a smaller buffer.



Design	Path-1	Path-2	Path-3
RC	BUF_X4	AOI21_X1	AOI22_X4
	BUF_X1	NAND4_X1	
	BUF_X8	XOR2_X1	
	BUF_X2	BUF_X1	
RLC	BUF_X2	BUF_X1	AOI22_X2



Conclusions and Future Work



□ Conclusions

- Chiplet-package interactions need to be considered during analysis and optimization of 2.5D systems.
- RDL wire delays are significantly underestimated with RC-only model.
- Our RLC delay model can accurately capture the inductance impact on timing delay through RDL wires.
- Parasitic scaling for inductance is compatible with the existing tools.
- Our parasitic scaling flow performs cross-boundary optimization to reduce RDL overhead.

□ Future Work

- Model 2.5D interconnects with native RLC models and CAD tools
- Extend our model to multi-point connections
- Signal and Power Integrity Study with RCLM elements



Thank You