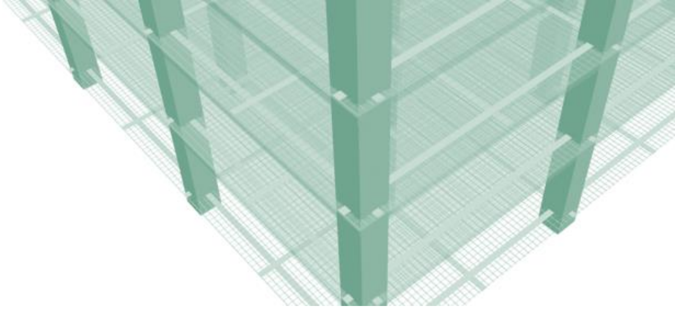
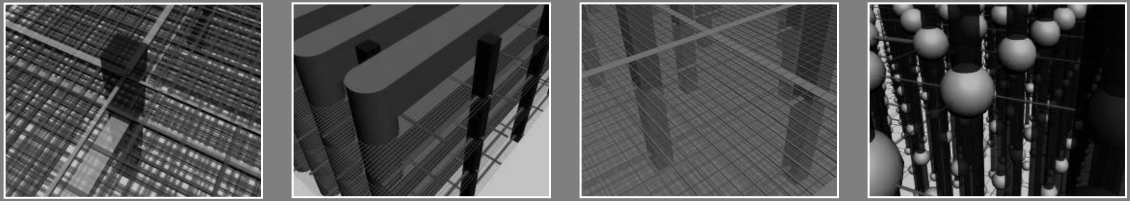


Design, Packaging, and Architectural Policy Co-Optimization for DC Power Integrity in 3D DRAM



Yarui Peng[1], Bon Woong Ku[1], Younsik Park[2],

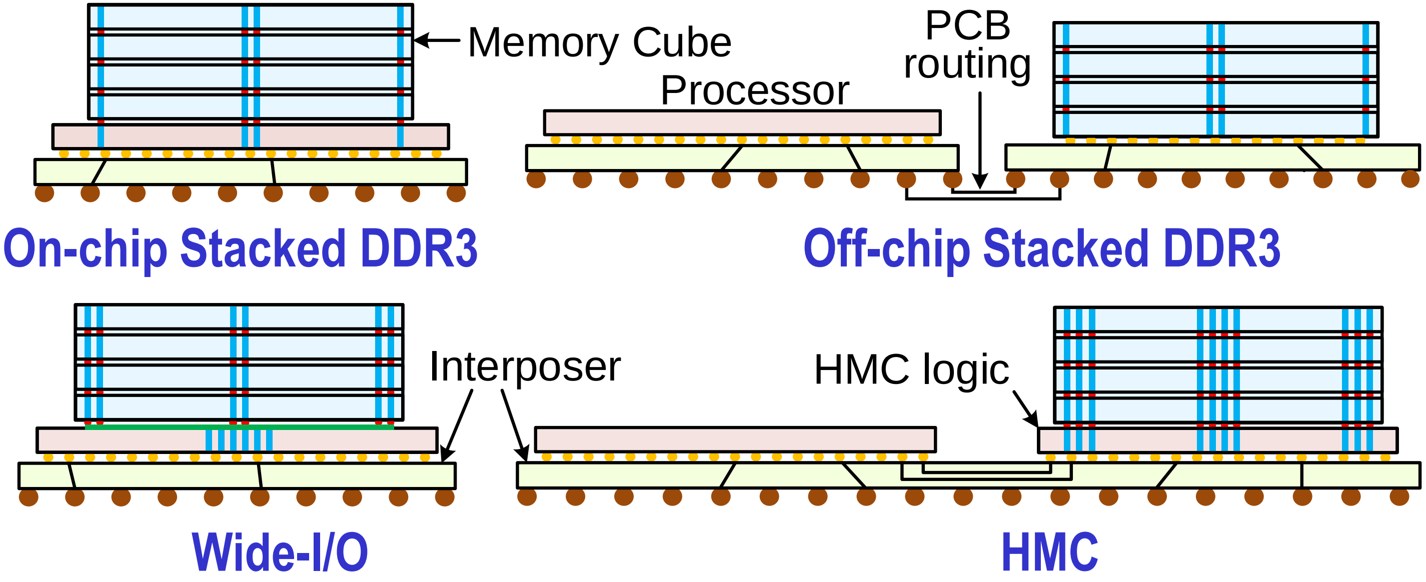
Kwang-Il Park[2], Seong-Jin Jan[2], Joo Sun Choi[2], and Sung Kyu Lim[1]

[1]Georgia Institute of Technology, Atlanta, GA, USA

[2]Samsung Electronics, Hwaseong-si, Gyeonggi-do, Korea

Introduction and Motivation

- One challenge in 3D DRAM is unreliable power delivery
  - More devices needs current while fewer bumps can fit into the footprint
- To solve this, we need to:
  - Assess special IR-drop issues in 3D IC system
  - Co-optimize PDNs in both memory cube and application processor (T2 chip)
  - Build the most efficient PDN design/package/architecture



On-chip Stacked DDR3

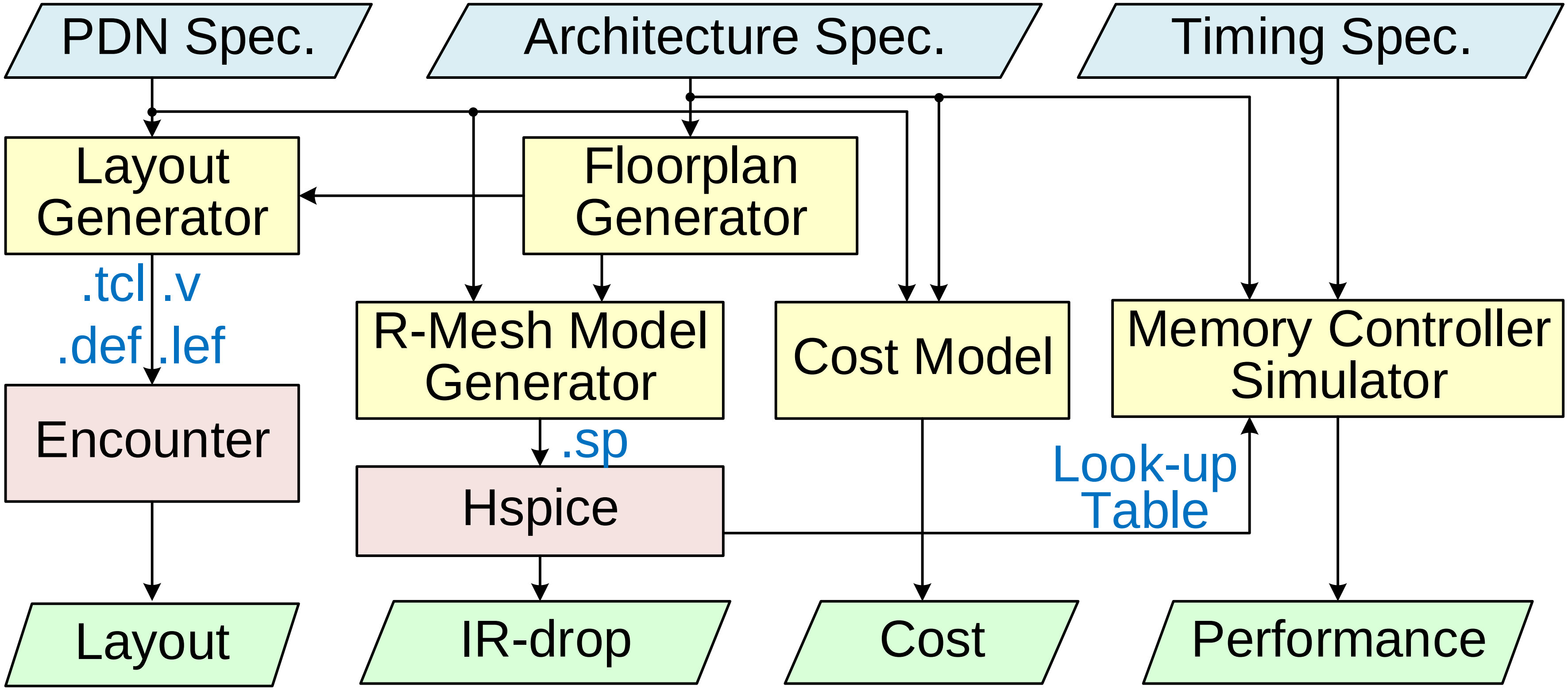
Off-chip Stacked DDR3

Wide-I/O

HMC

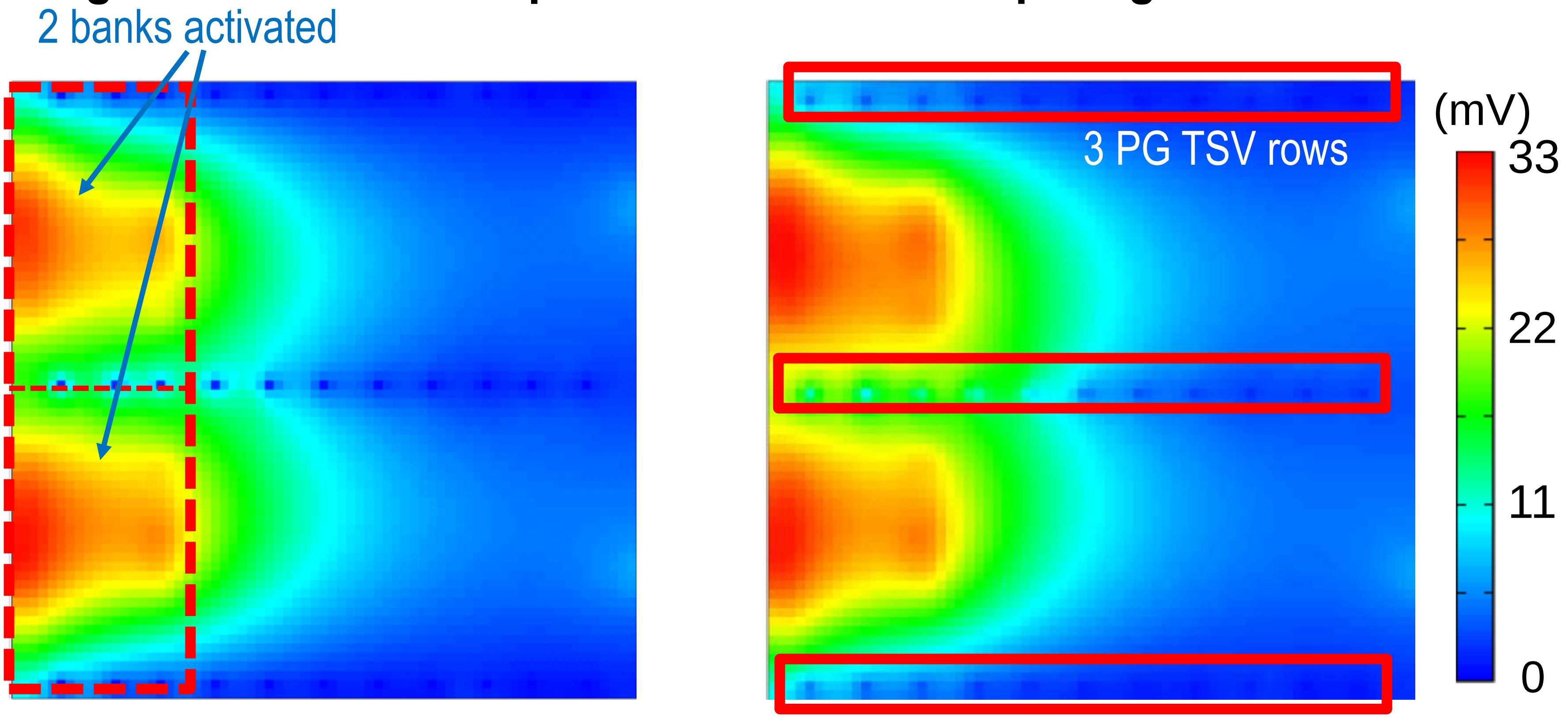
Our CAD Infrastructure

- Our study combines a floorplanner, a PDN generator, an R-Mesh model, a memory controller simulator, and a cost model altogether



Verification Results

- **Our R-Mesh model is fully verified against Cadence EPS with significant runtime improvement without requiring detail extraction**



R-Mesh: 32.2mV Max IR-drop
Runtime: 5s

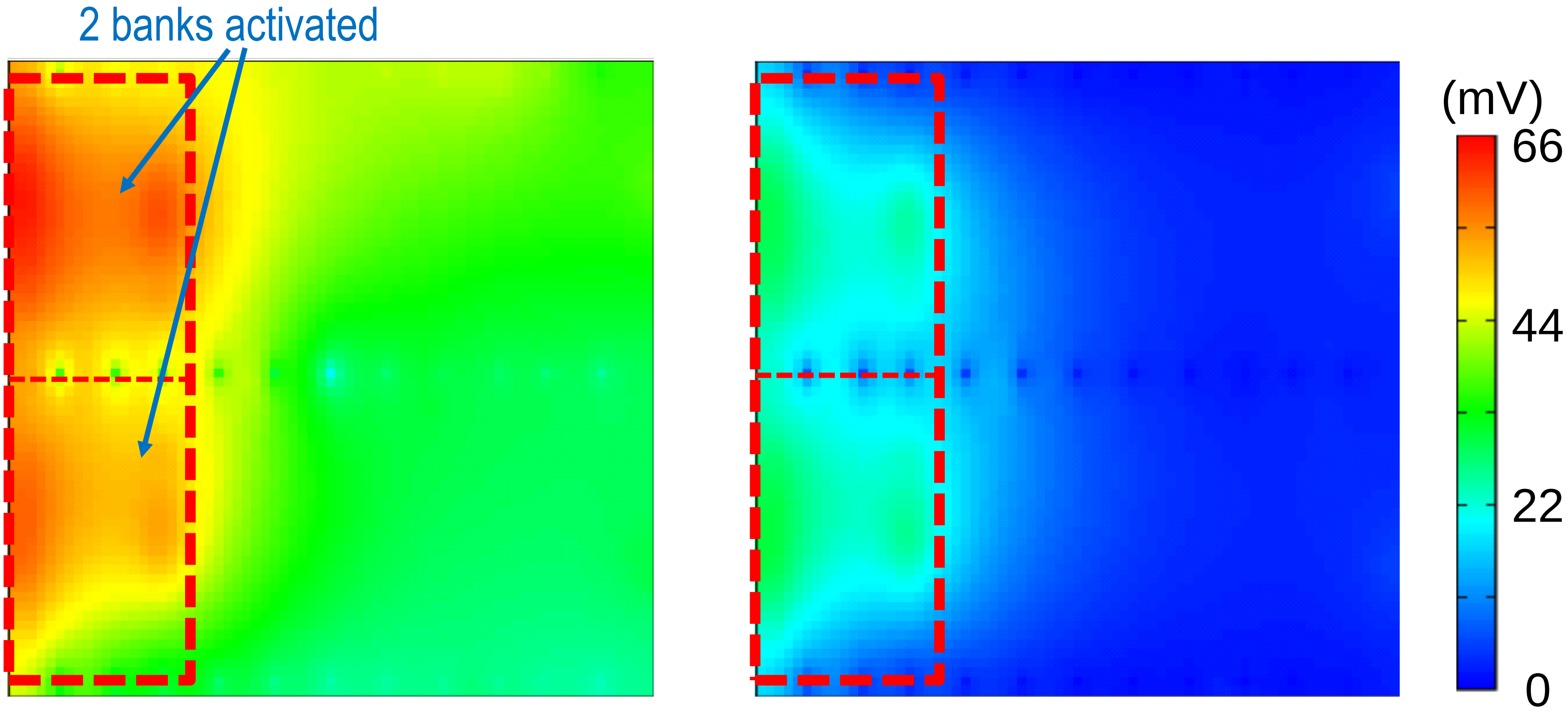
EPS: 32.6mV Max IR-drop
Runtime: 517s

Our Cross Domain Solutions

- Design domain
  - PDN wire usage
  - TSV count, location, and alignment with C4
  - RDL configuration: between memory and logic & between memories
- Packaging domain
  - Bonding style: F2F, F2B
  - Dedicated TSVs
  - Extra wire bonding
- Architectural domain
  - Read policy based on IR-drop look-up tables
  - Balancing read requests to multiple dies

Inter-die IR-drop Coupling Impact

- T2 has significant impact on stacked DDR3 with connected PDN
- Dedicated TSV helps IR-drop by decoupling the PDNs



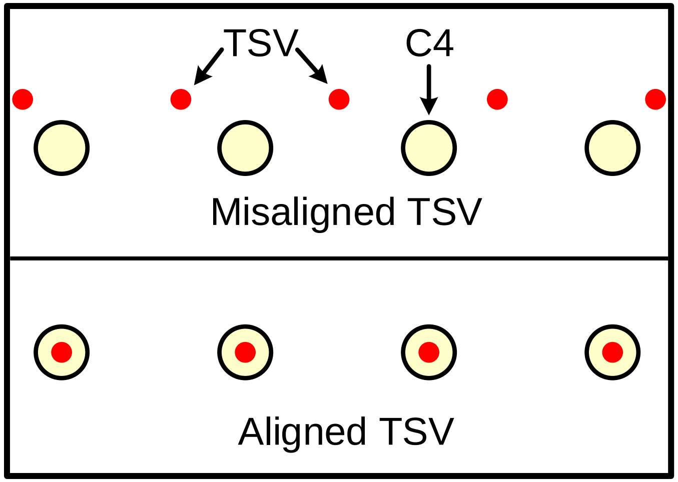
2 banks activated

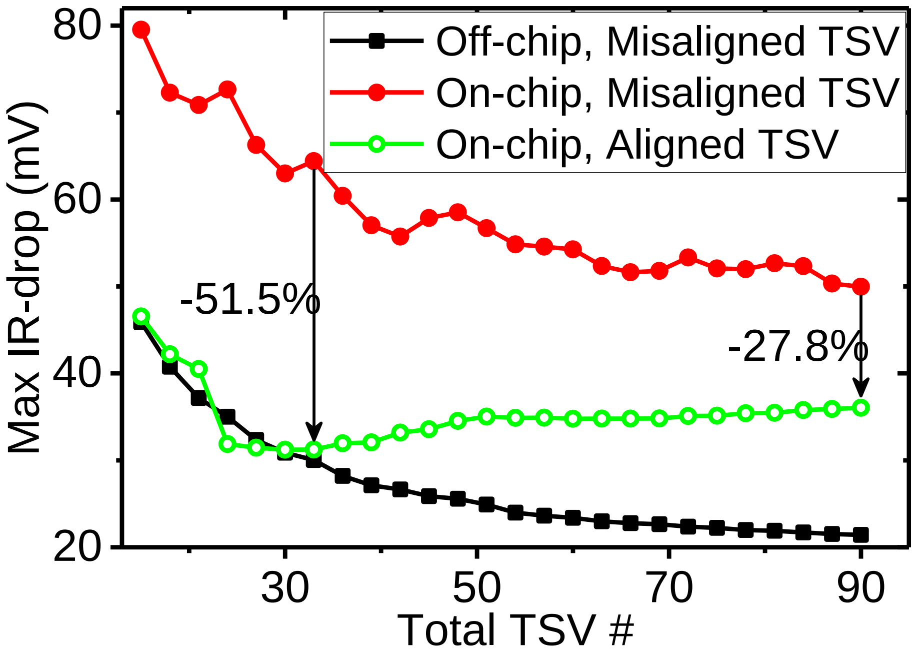
On-chip (64.41mV)

Off-chip (30.03mV)

TSV Count and Alignment Impact

- Good alignment between TSV and C4 reduces IR-drop up to 51.5%
  - Reduces horizontal IR supply path
- Increasing TSVs reduce IR-drop effectively, but the reduction saturates with large TSV count
  - Reduces vertical IR supply path

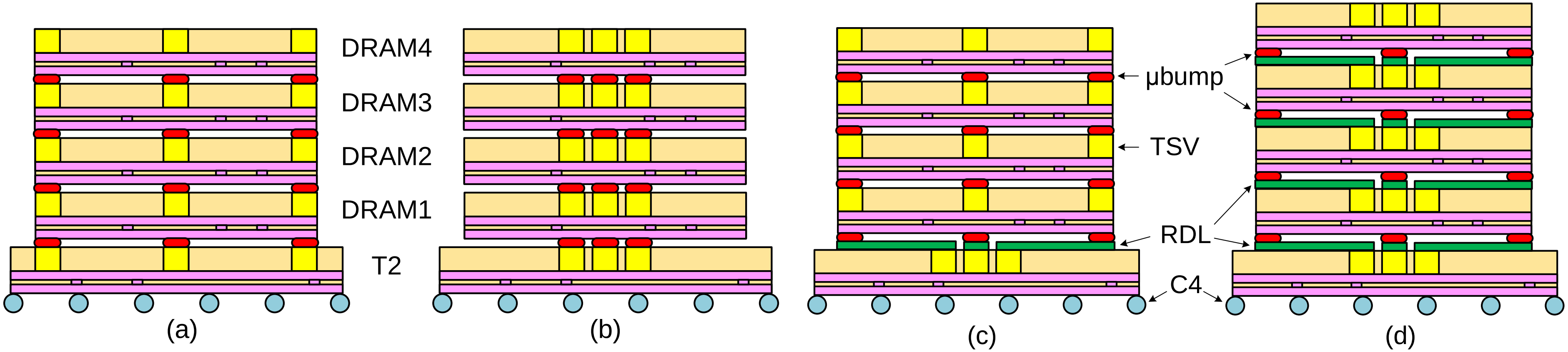




RDL Impact

- We studied four RDL configurations and their tradeoffs

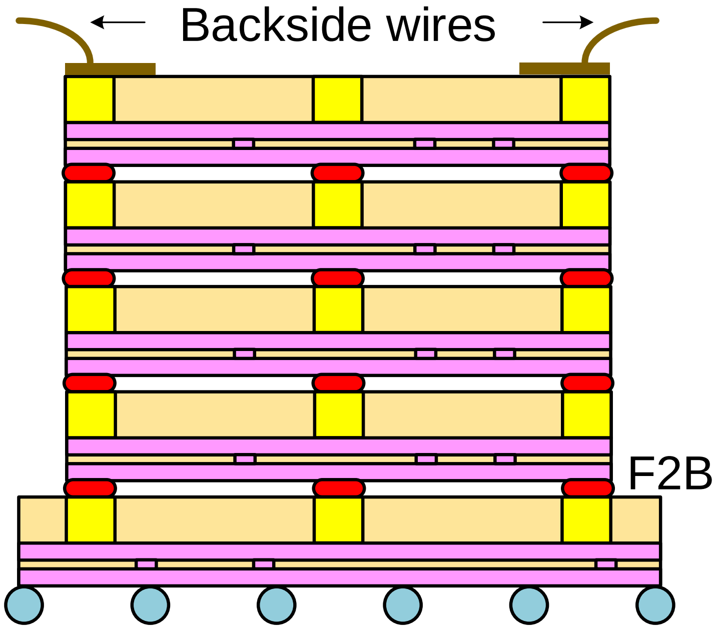
Design option	(a)	(b)	(c)	(d)
Logic TSV	Non-center	Center	Center	Center
DRAM TSV	Edge	Center	Edge	Center
Logic die cost	High	Low	Medium	Medium
DRAM die cost	High	Low	High	Medium
Overall cost	Highest	Lowest	High	Medium
IR drop (mV)	30.03	50.76	38.46	49.36



Wire Bonding Impact

- **Backside wire bonding provides additional power supply to DRAM cube and helps reducing IR-drop significantly**
  - **Allow power supply from both sides of the DRAM cube**
  - **Provides direct supply to DRAM cube similarly as dedicated TSVs**

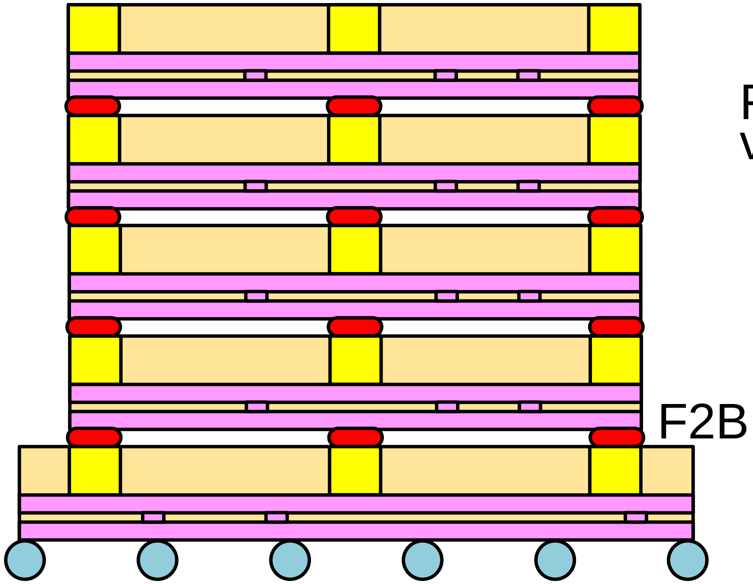
Design	Dedicated TSV?	IR-drop (mV)		
		Baseline	Wire bonded	Δ%
On-chip	no	64.41	30.04	-53.4%
	yes	31.18	27.18	-12.8%
Off-chip	yes	30.03	27.10	-9.8%



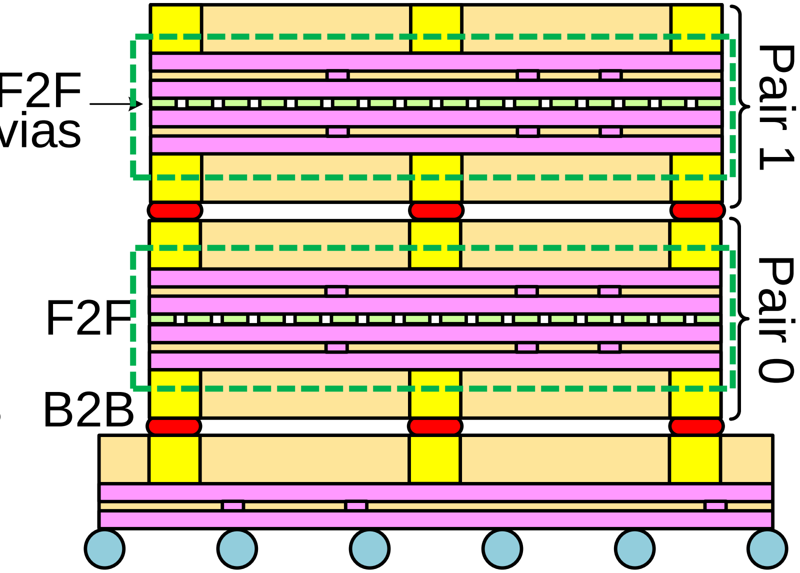
With wire bonding

Bonding Style Impact

- **Due to PDN sharing, swapping die orientation and using F2F+B2B reduces IR-drop significantly when there is no intra-pair overlap**
  - **A pair of dies is able to share their PDNs together with identical PDN routing**
  - **Provides additional power supply path for active banks**



F2B bonding

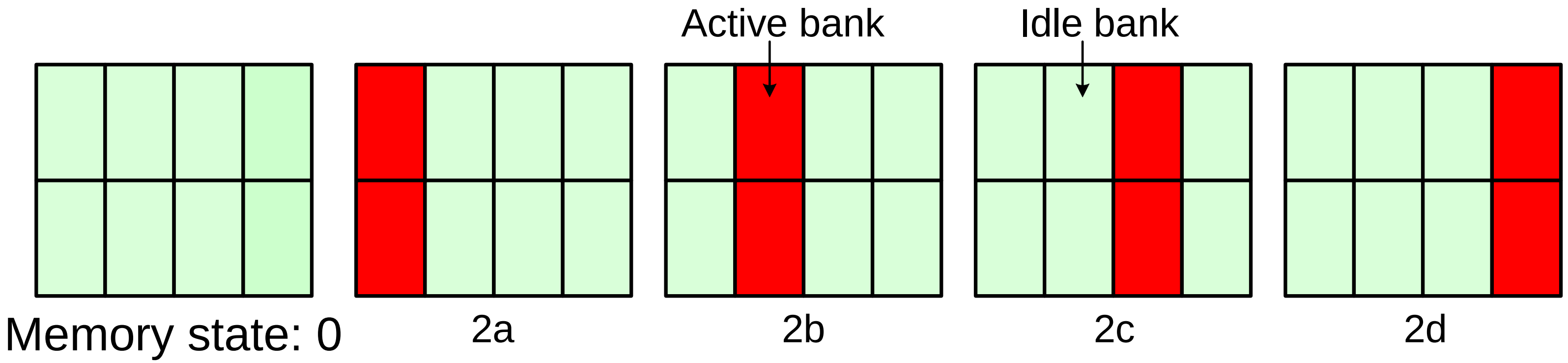


F2F+B2B bonding

Intra-die Overlap Impact

- **Without intra-pair overlapping, F2F benefits maximize**

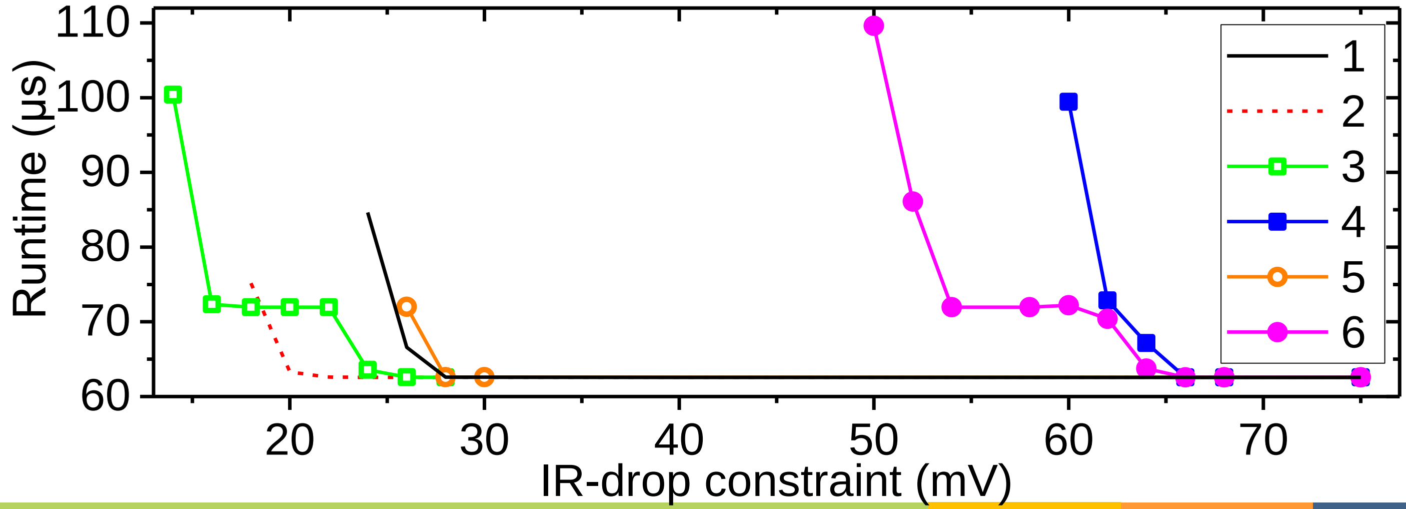
Memory state	Intra-pair overlapping	Max IR drop (mV)		
		F2B	F2F+B2B	Δ%
0-0-2a-2a	yes	28.14	27.21	-3.3%
0-0-2b-2b		18.06	17.42	-3.5%
0-2a-0-2a	no	27.32	15.24	-44.2%
2a-0-0-2a		26.51	15.24	-42.5%
0-0-2b-2a	no	27.38	17.98	-34.3%
0-0-2c-2a		27.04	17.1	-36.8%
0-0-2d-2a		26.86	15.27	-43.1%



Memory Performance with IR-drop Constraint

- **Max number of banks is limited by the IR-drop. Under an low IR-drop constraint, designs with lower IR-drop perform better.**

Mounting style	Off-chip			On-chip		
Case #	1	2	3	4	5	6
Bonding style	F2B	F2B	F2F	F2B	F2B	F2F
Metal usage	1x	1.5x	1x	1x	1x	1x
Wire bonding	no	no	no	no	yes	no
IR-drop (mV)	**30.03**	**22.15**	**17.18**	**64.41**	**30.04**	**65.43**



Architectural Policy Impact

- **Standard JEDEC policy uses tRRD (Activate to Activate delay) and tFAW (Four Activation Window) to control max IR-drop. But it lowers the performance**
- **Our IR-drop aware policy solves this with a detailed IR-drop look-up table to control max IR-drop**
- **A distributed read policy further improves performance by balancing the load across multiple DRAM dies**

IR-drop policy	Standard	Our IR-drop aware policy	
Scheduling policy*	FCFS	FCFS	DistR
IR-drop constraint	none	24mV	24mV
Runtime (us)	109.3	84.68 (-22.6%)	75.85 (-30.6%)
Bandwidth (read/clk)	0.114	0.148 (+29.2%)	0.165 (+44.2%)
Max IR-drop (mV)	30.03	23.98 (-20.2%)	23.98 (-20.2%)

*FCFS: first come first serve, DistR: distributed read

Cost Model for DRAM

- We build a cost model and use Matlab regression analysis to estimate IR-drop based on sampled R-Mesh simulation
- An IR-Cost term is used to calculate best options:

$$\mathrm{IR{-}Cost} = \mathrm{IR{-}Drop}^{\alpha} \times \ \mathrm{Cost}^{1-\alpha}$$

Solution	Abbreviation	Type	Cost Range
M2 metal usage	M2	Continuous integer	0.025-0.05
M3 metal usage	M3		0.025-0.10
Power TSV count	TC		0.078-0.44
Dedicated TSV	TD	Yes(Y)/No(N)	0.06/0
Bonding style	BD	F2B/F2F	0.045/0.06
RDL routing	RL	Yes(Y)/No(N)	0.05/0
Wire bonding	WB	Yes(Y)/No(N)	0.03/0
TSV location	TL	Center only(C)	0
		Edge and center(E)	0.5xTC
		Distributed(D)	TC

Put it Altogether: Best Options

Design	α	M2	M3	TC	TL	TD	BD	RL	WB	IR-drop(mV)		Cost
										Matlab	R-Mesh	
Off-chip DDR3	0	10	10	15	C	Y	F2B	N	N	88.73	88.73	0.23
	0.3	20	22	24	E		F2F	N	N	22.75	23.01	0.37
	1	20	40	360	E		F2F	N	Y	9.733	9.54	0.87
	Baseline	10	20	33	E		F2B	N	N	30.03	30.03	0.35
On-chip DDR3	0	10	10	15	C	N	F2B	N	N	117.6	117.6	0.17
	0.3	20	22	21	E	N	F2B	N	Y	25.51	27.09	0.32
	1	20	40	420	E	Y	F2F	N	Y	9.864	9.843	0.92
	Baseline	10	20	33	E	Y	F2F	N	N	31.18	31.18	0.35
Wide-I/O	0	10	10	160	C	N	F2B	N	N	110.1	110.2	0.35
	0.3	20	40		E	Y	F2F	Y	Y	4.864	4.841	0.73
	1	20	40		E	Y	F2F	Y	Y	4.864	4.841	0.73
	Baseline	10	20		E	Y	F2B	Y	N	13.56	13.62	0.62
HMC	0	10	10	160	C	N	F2B	N	N	459.7	459.7	0.35
	0.3	20	25	160	D	Y	F2B	N	Y	18.63	18.65	0.76
	1	20	40	480	D	Y	F2B	N	Y	13.76	13.84	1.17
	Baseline	10	20	384	E	Y	F2B	N	N	47.9	47.9	0.77

Conclusions

- We investigated impact of various design, packaging, and architectural policy options on 3D DRAM DC power integrity.
- Inter-die coupling, the TSV count, location, and alignment strongly affected the IR drop.
- Backside wire bonding and F2F bonding reduced the IR drop significantly with low cost overhead.
- Our IR-drop-aware policies and distributing activity optimized performance under a tight IR-drop constraint.
- We proposed best co-optimization solutions for the stacked DDR3, Wide I/O, and HMC designs based on Matlab regression analyses.