

NC STATE UNIVERSITY

# 3D Technologies For Low Power Integrated Circuits

---

---

**Paul Franzon**

North Carolina State University

Raleigh, NC

[paulf@ncsu.edu](mailto:paulf@ncsu.edu)

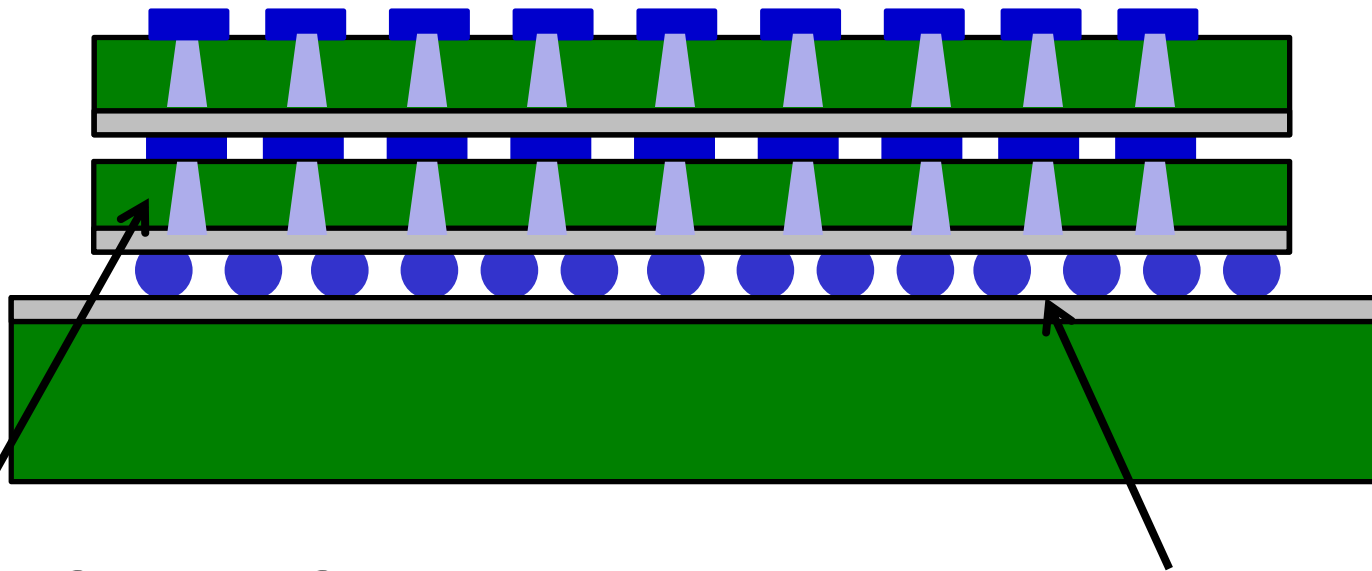
919.515.7351

# Outline

---

- 3DIC Technology Set
- Approaches to 3D Specific Power Minimization
  - ⊙ Comparative energy/operations
  - ⊙ Leverage Rent's Rule?
  - ⊙ The low-hanging fruit : Energy-optimized memory interfaces
  - ⊙ 3D-specific architectures
  - ⊙ Next: Heterogenous Integration
- CAD approaches to 3D optimization

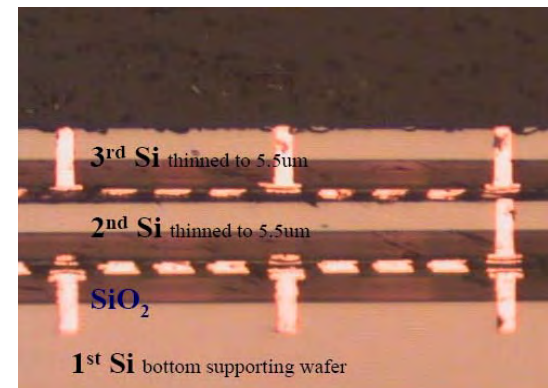
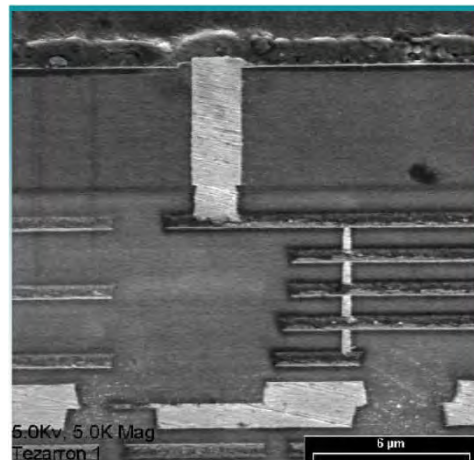
# 3DIC Technology Set



**Bulk Silicon TSVs and bumps  
(25 - 40  $\mu\text{m}$  pitch)**

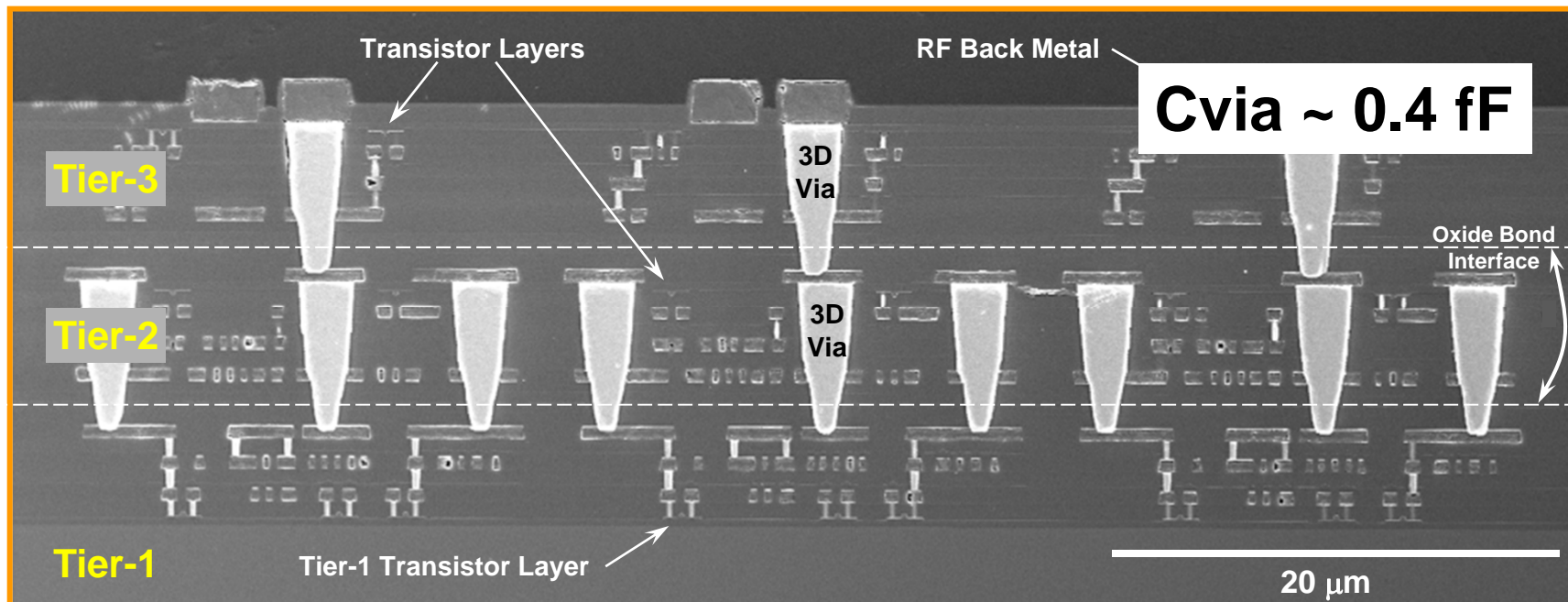
**Face to face microbumps  
(2 - 30  $\mu\text{m}$  pitch)**

**$C_{\text{TSV}} \sim 30 \text{ fF}$**



# ... 3DIC Technology Set

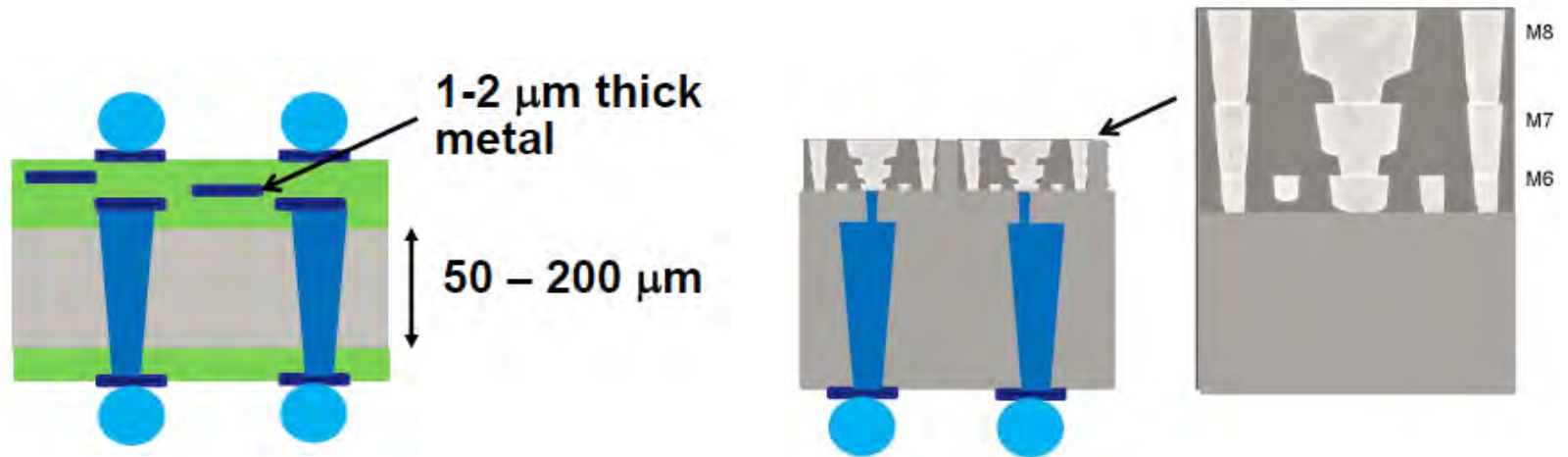
TSVs in an SOI process



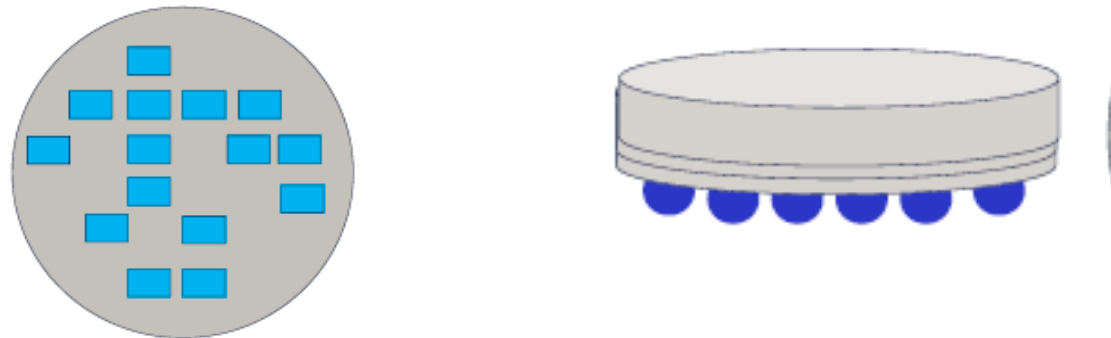
MIT Lincoln Labs

# 3DIC Technology Set

- Interposers: Thin film or 65/90 nm BEOL



- Assembly : Chip to Wafer or Wafer to wafer

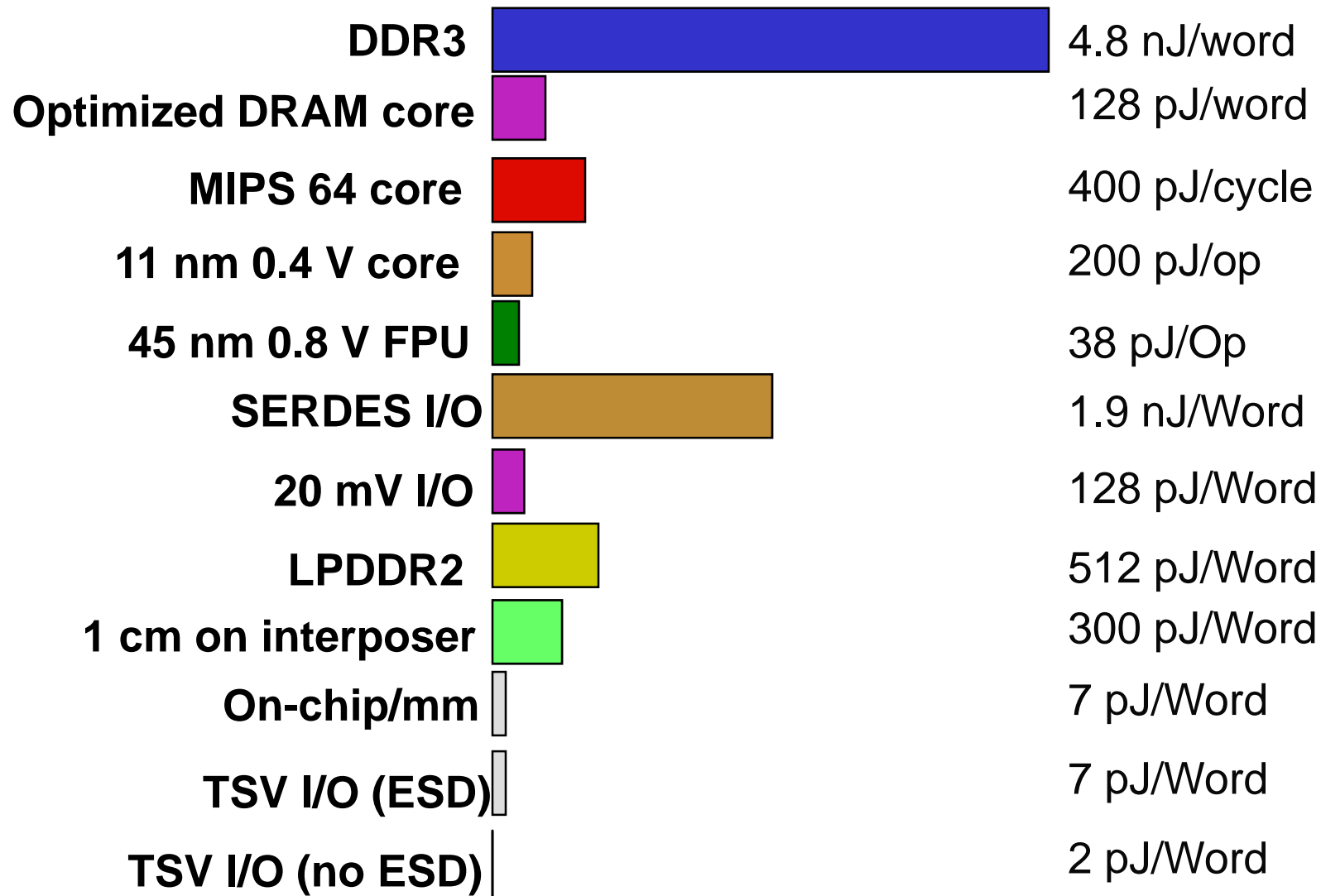


# Outline

---

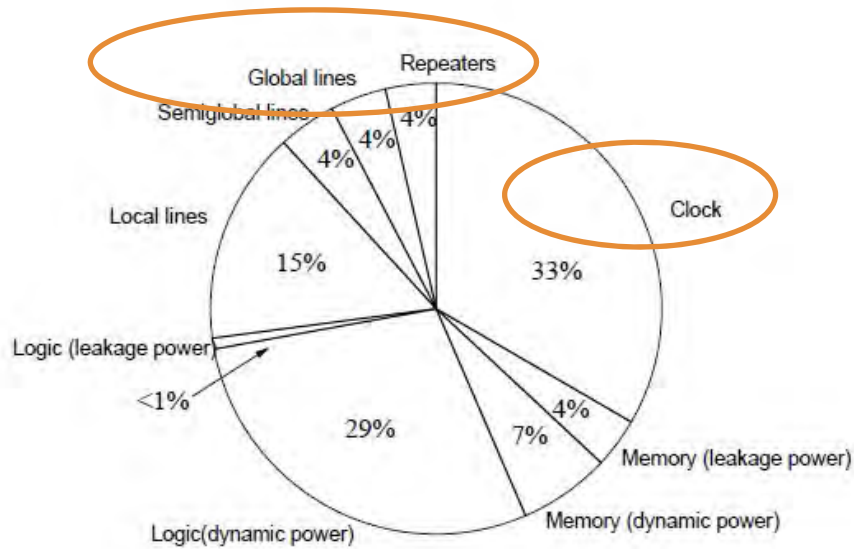
- 3DIC Technology Set
- Approaches to 3D Specific Power Minimization
  - Comparative energy/operations
  - Is leveraging Rent's Rule enough?
  - Energy-optimized memory interfaces
  - 3D-specific architectures
  - Heterogeneous Integration
- CAD approaches to 3D optimization

# Energy per Operation

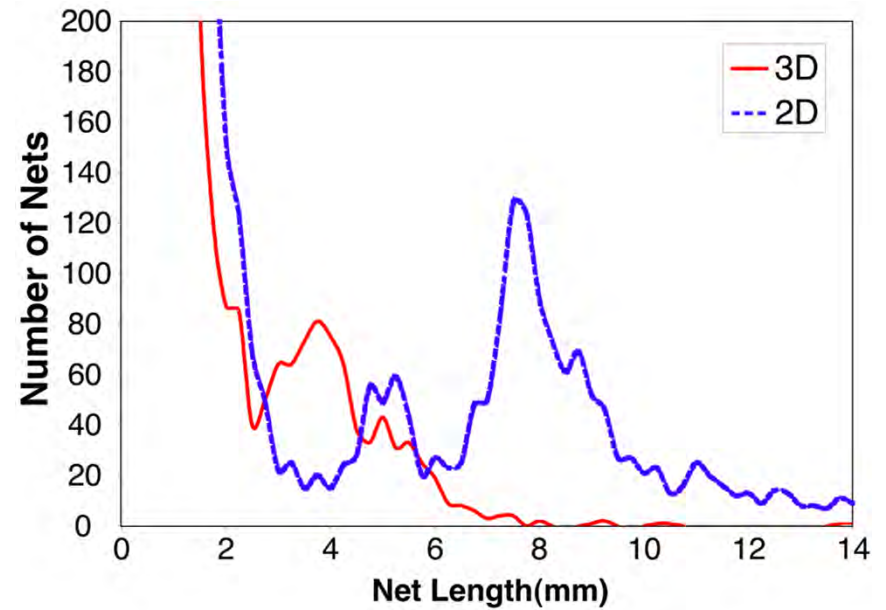


**(64 bit words)**

# Wire Power Reduction



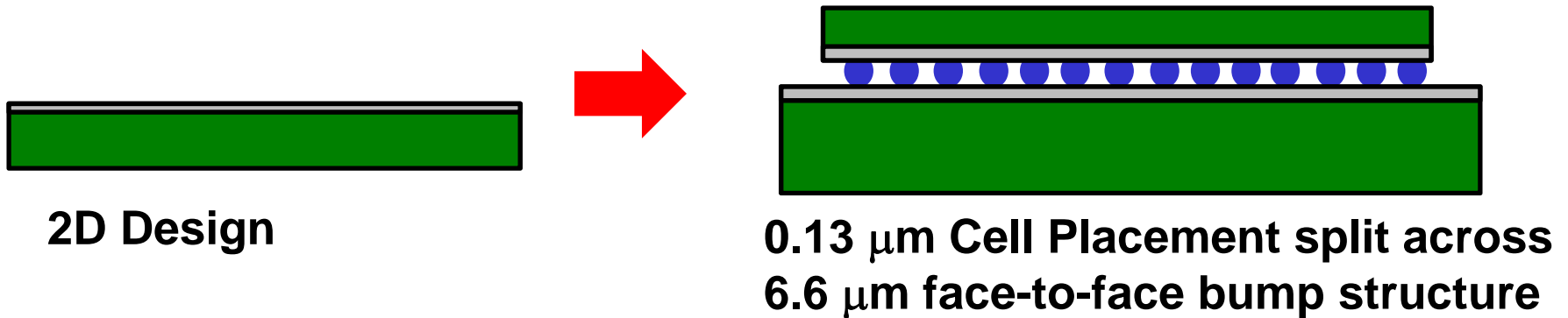
**Exemplar power distribution**



**FFT Processor**

# Shorter wires → Modest Returns

- Relying on wire-length reduction alone is not enough



	Total Wire Length (% Change)	Max Frequency (% Change)	Parasitic Power (% Change)	Total Power ) (% Change)
PE 3D Seq.	-17.1%	+7.1%	-15.5%	-4.7%
PE 3D Sim.	-17.7%	+16.2%	-27.9%	-7.7%
PE 3D True	-21.0%	+22.6%	-45.2%	-12.9%
AES 3D Seq.	-8.0%	+15.3%	-19.6%	-2.6%
MIMO 3D Seq.	+216.1%	+17.1%	-34.9%	-5.1%

**Results get less compelling with technology scaling, as the microbumps don't scale**

Logic-on-Logic 3D Integration and Placement

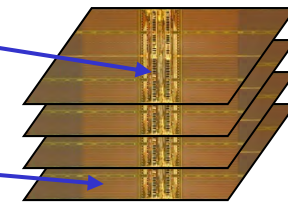
Thorlindur Thorolfsson<sup>□</sup>, Guojie Luo<sup>†</sup>, Jason Cong<sup>†</sup> and Paul D. Franzon<sup>□</sup>  
<sup>□</sup> Department of Electrical & Computer Engineering, North Carolina State University, Raleigh, NC 27695  
<sup>†</sup> Computer Science Department, University of California, Los Angeles, CA 90095  
 Email: [thor@ece.ucla.edu](mailto:thor@ece.ucla.edu) and [cong@ece.ucla.edu](mailto:cong@ece.ucla.edu)

# Memory on Logic

## Conventional



## TSV Enabled



Less Overhead

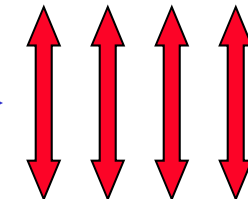
Flexible bank access

x32  
to  
x128

Less interface power

3.2 GHz @ >10 pJ/bit

→ 1 GHz @ 0.3 pJ/bit

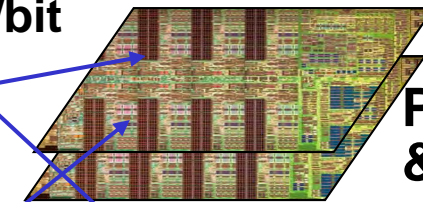


N x 128  
“wide I/O”

Flexible architecture

Short wires

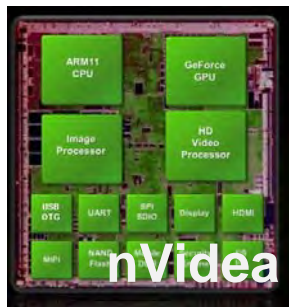
Exploit dense face-to-face



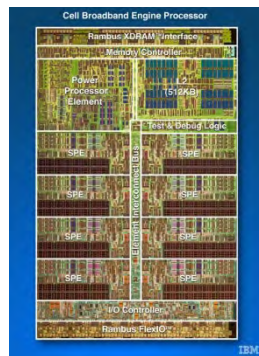
Processor  
& SRAM

or

Mobile



or

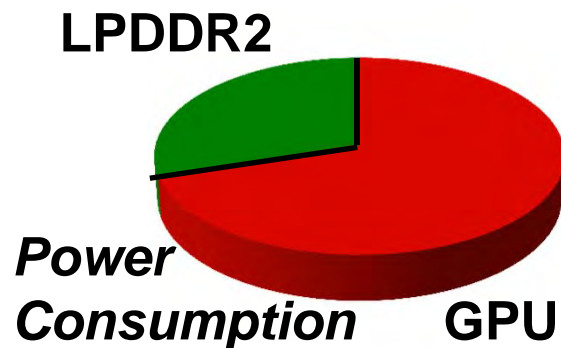


# Mobile Graphics



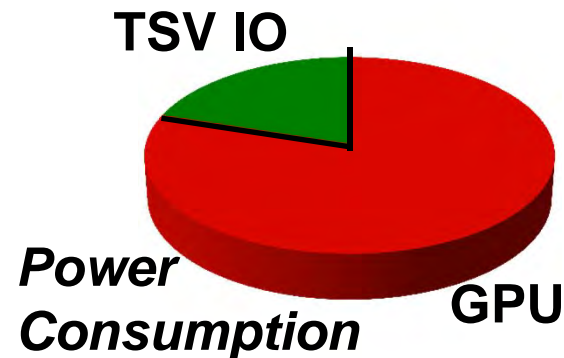
- **Problem:** Want more graphics capacity but total power is constrained
- **Solution:** Trade power in memory interface with power to spend on computation

## POP with LPDDR2



532 M triangles/s

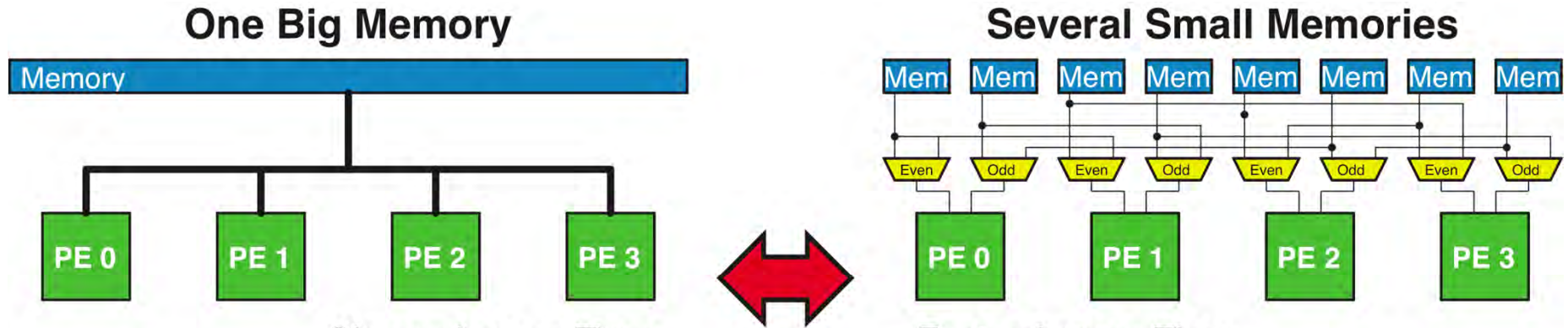
## TSV Enabled



695 M triangles/s

# Synthetic Aperture Radar Processor

- Built FFT in Lincoln Labs 3D Process

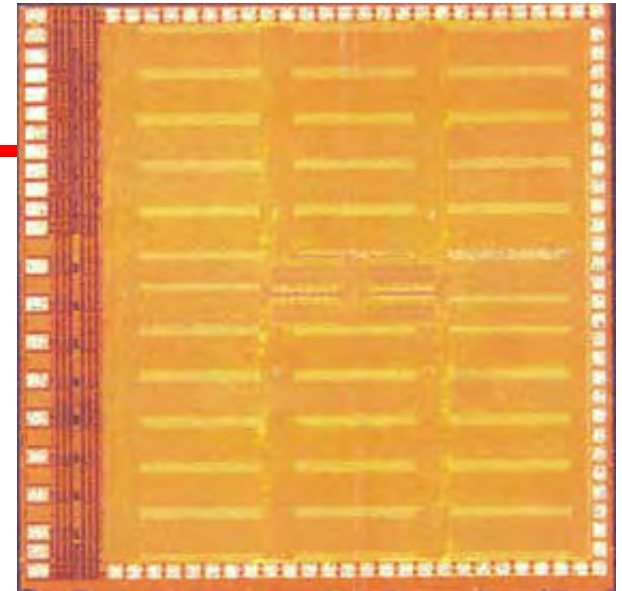


Metric	Undivided	Divided	%
Bandwidth (GBps)	13.4	128.4	+854.9
Energy Per Write(pJ)	14.48	6.142	-57.6
Energy Per Read (pJ)	68.205	26.718	-60.8
Memory Pins (#)	150	2272	+1414.7
Total Area (mm <sup>2</sup> )	23.4	26.7	+16.8%

Thor Thorolfsson

# 3D FFT Floorplan

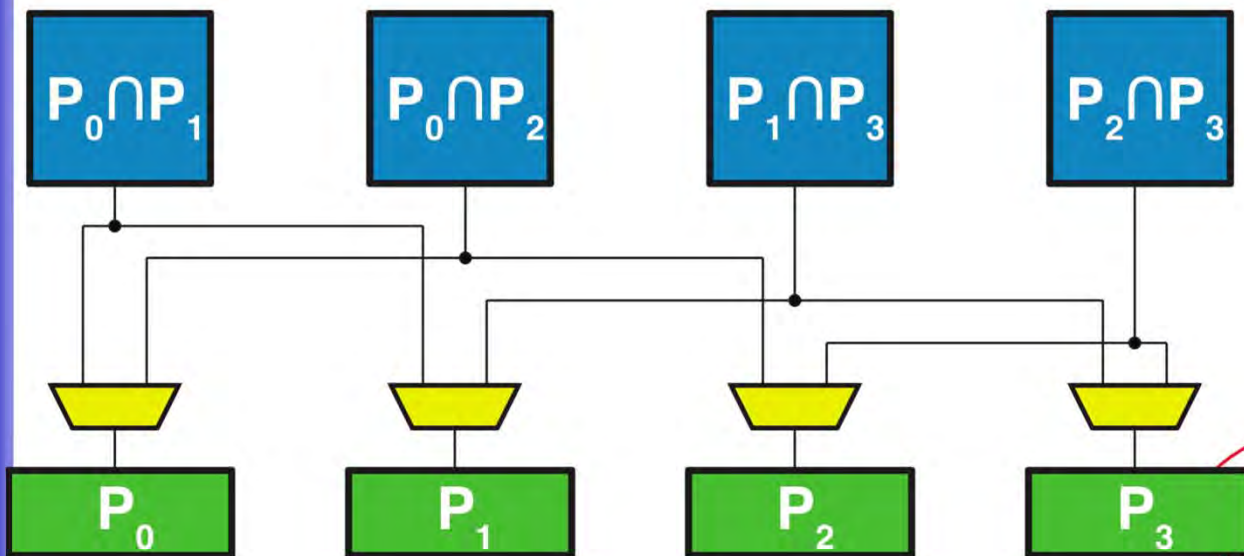
- All communications is vertical
- Support multiple small memories WITHOUT an interconnect penalty
  - AND Gives 60% memory power savings



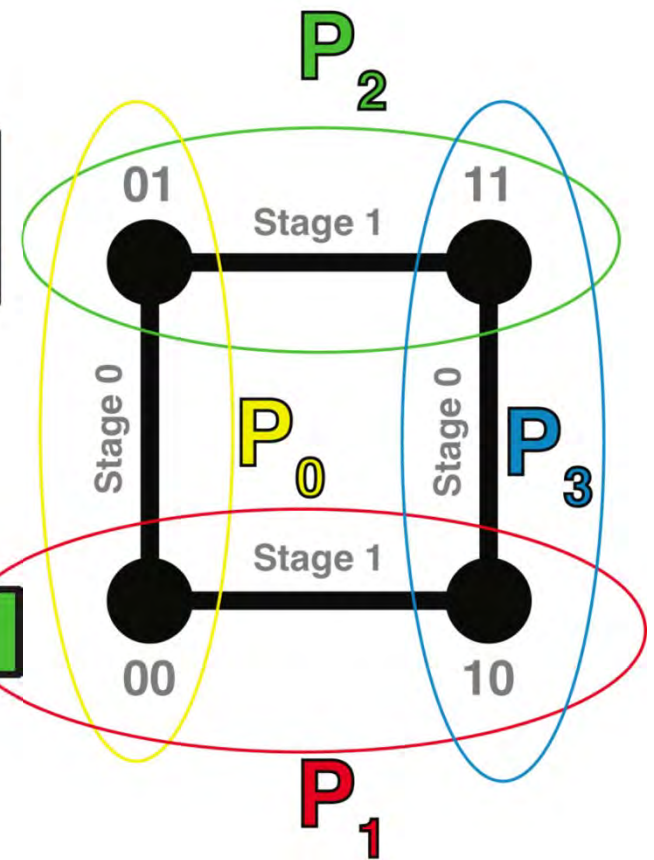
# RePartition FFT to Exploit Locality

- Every partition is a PE
- Every unique intersection is a memory

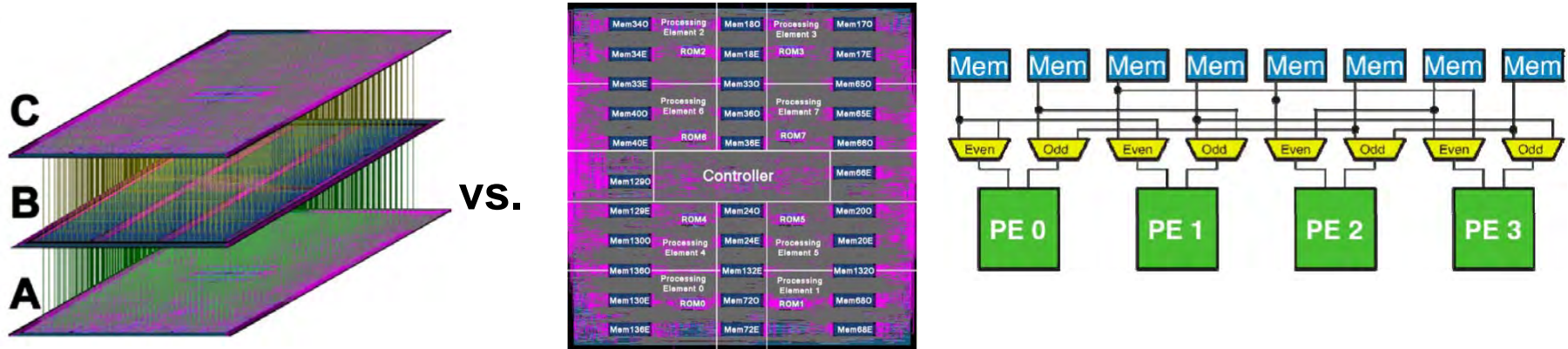
## Memories



## Processing Elements



# 2DIC vs. 3DIC Implementation



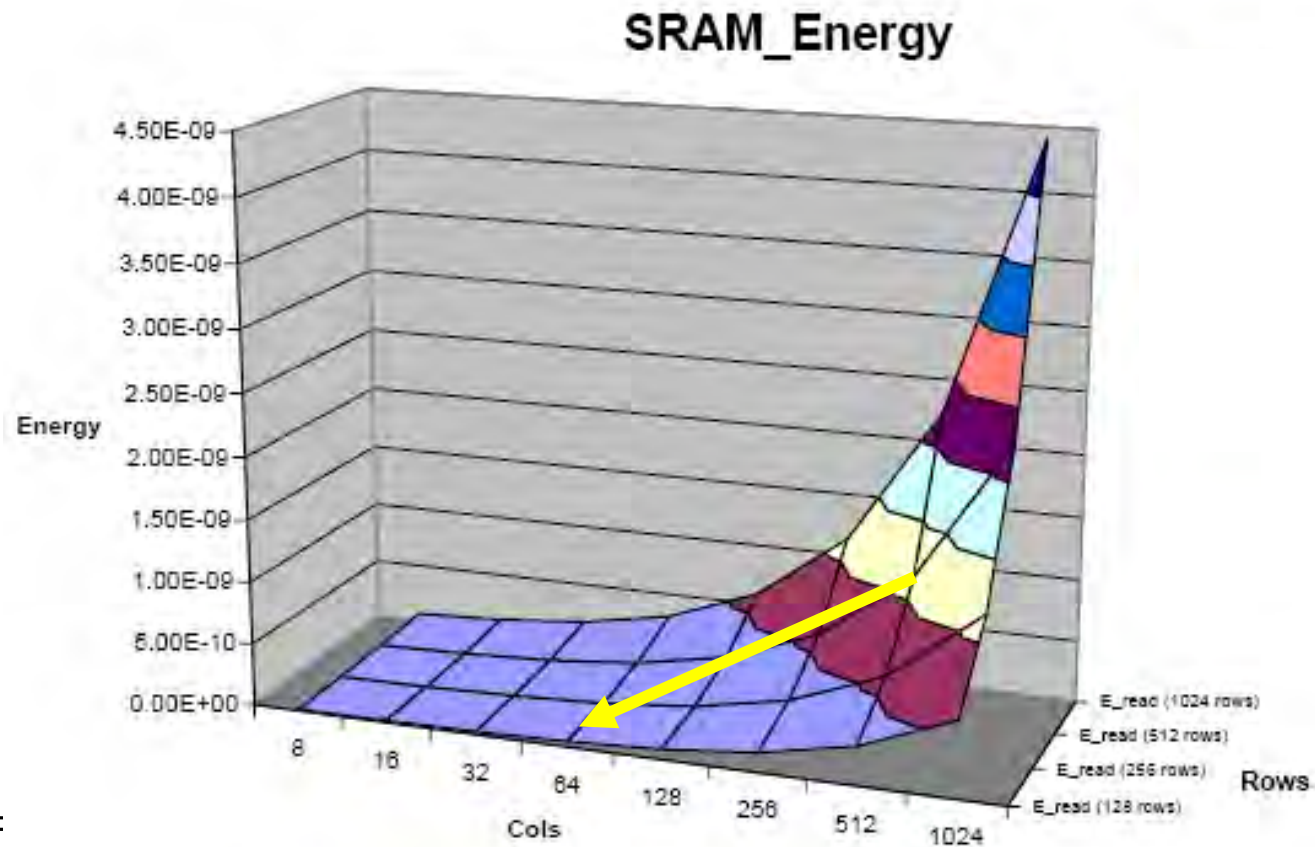
Metric	2D	3D	Change
Total Area (mm <sup>2</sup> )	31.36	23.4	-25.3%
Total Wire Length (m)	19.107	8.238	-56.9%
Max Speed (Mhz)	63.7	79.4	+24.6%
Power @ 63.7MHz (mW)	340.0	324.9	-4.4%
FFT Logic Energy (μJ)	3.552	3.366	-5.2%

# Memory bank size tradeoffs

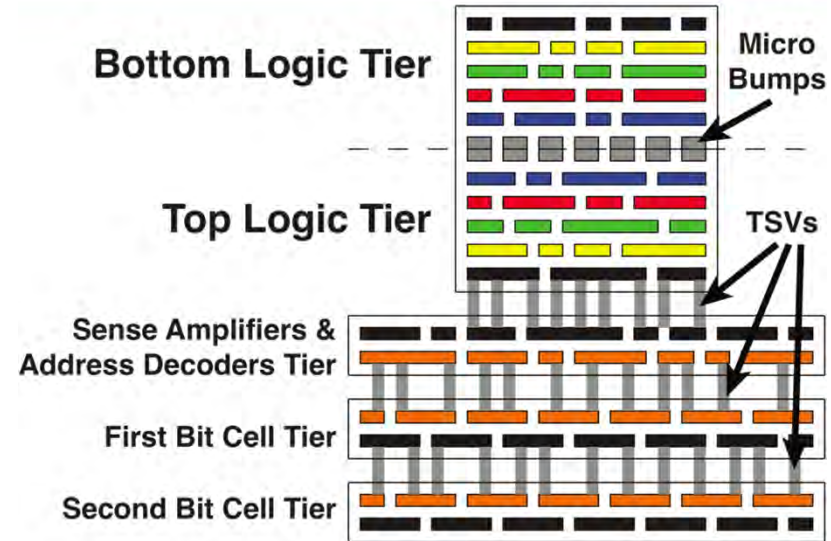
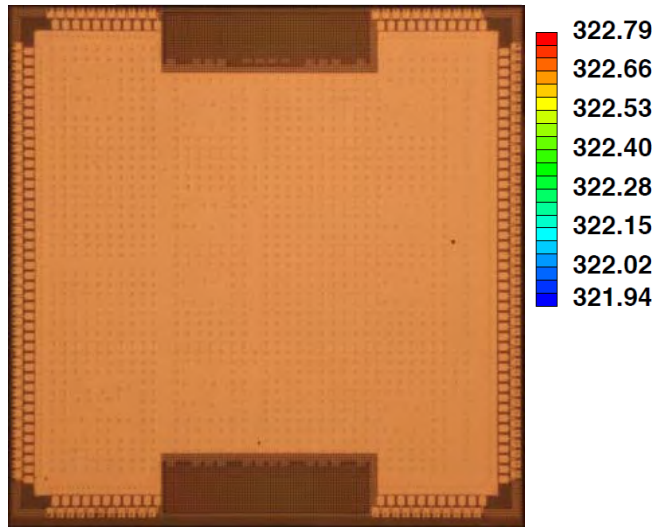
---

E.g. 32 x 2 kbit SRAM **10x** less energy/bit than 1 x 64 kbit SRAM

- With 17% increase in area (partially recoverable by in 3D)



# Tezzaron SAR Processor



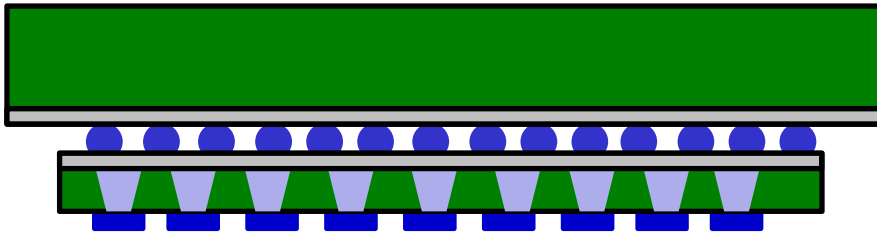
Metric	2D	3D		mPI 3D	
Total Wire length (mm)	588	487.3	-1.17%	464.8	-21%
Max. Frequency (MHz)	31.6	33.84	+7.1%	38.74	+22.6%
Max Performance (MFlops)	316.1	338.4	+7.1%	387.4	+22.6%
Parasitic Power (mW)	1.51	1.79	-15.5%	0.984	-45.2%
Logic Power (mW)	5.975	5.692	-4.8%	5.21	-12.9%
Memory Power (W)	10.3	3.1	-71%		

Logic-on-Logic 3D Integration and Placement

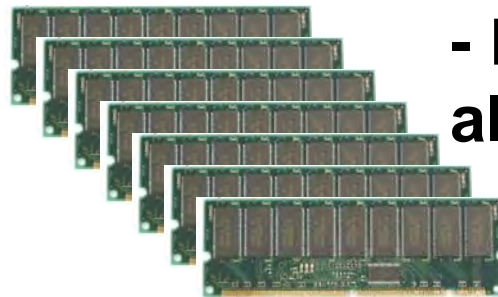
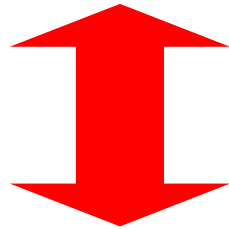
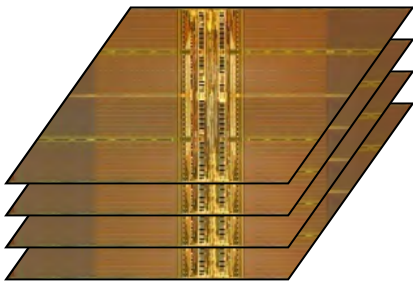
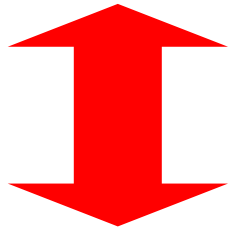
# Next level of Exploitation

---

- Key: Architecture Optimized for 3D Exploitation



**Heterogeneous Integration**  
(different voltages, processes)  
→ Power optimization without compromise

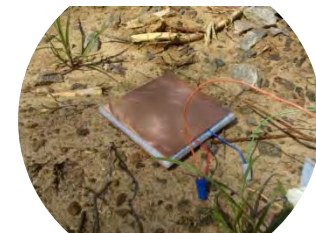
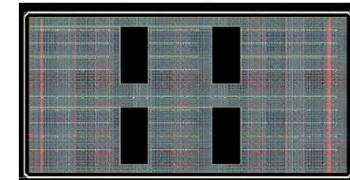
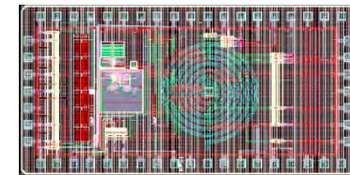
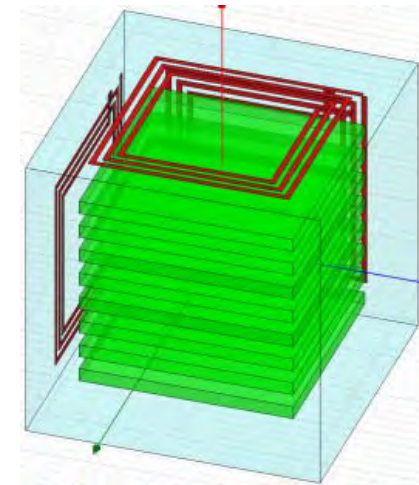


**Heterogeneous Memory**  
- Energy optimized  
algorithmic approach

# 3D Miniaturization

## Miniature Sensors

- ⊙ mm<sup>3</sup> scale - Human Implantable (with Jan Rabaey, UC(B))
- ⊙ cm<sup>3</sup> scale - Food Safety & Agriculture (with KP Sandeep, NCSU)
- Focii
  - ⊙ System Design
  - ⊙ Right sizing power harvesting for application
- Problems:
  - ⊙ Maximizing RF power harvesting potential
  - ⊙ Technology integration – MEMS, Energy storage, passives in “true 3D”



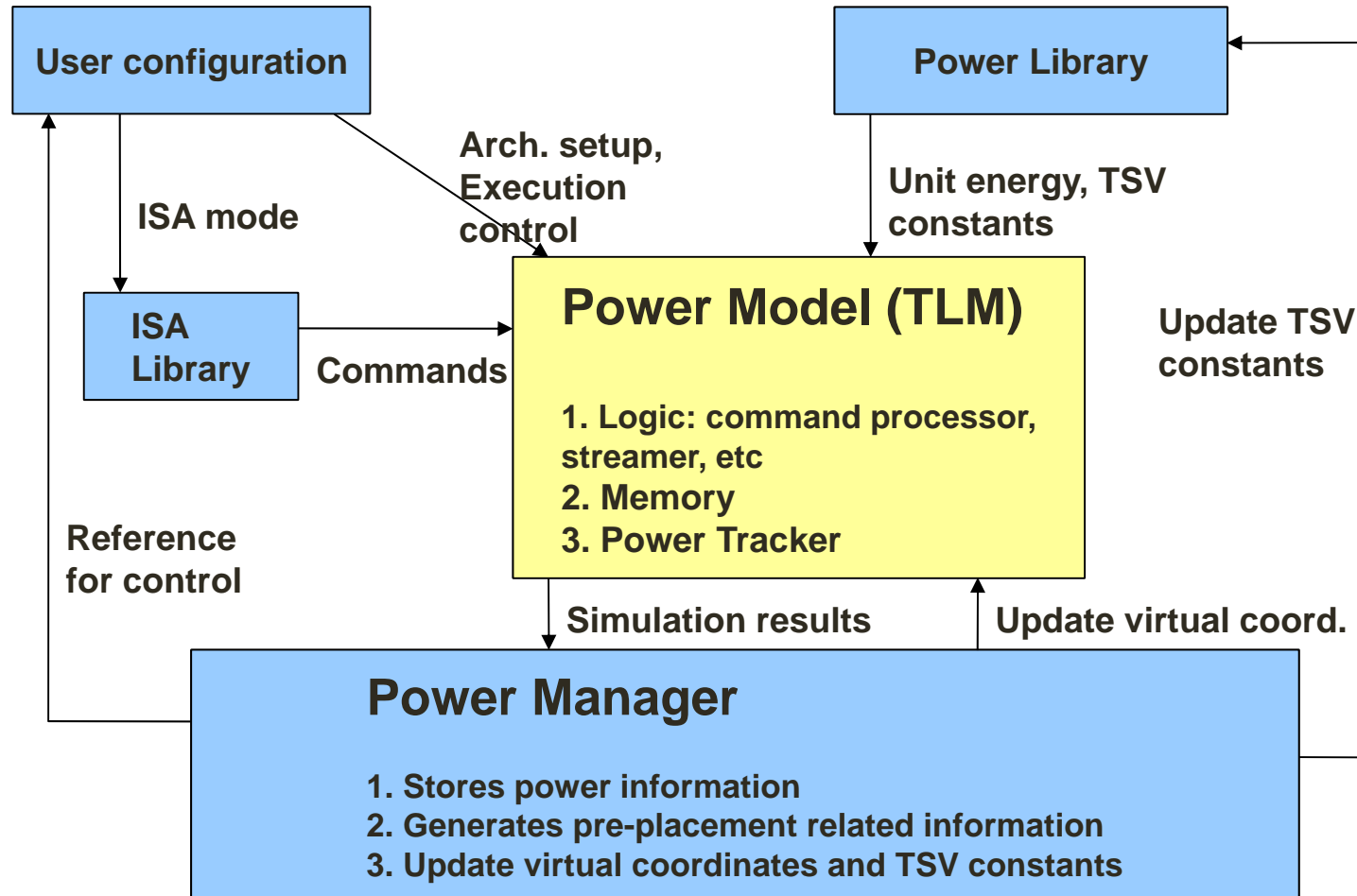
Peter Gadfort, Akalu Lentiro, Steve Lipa

# Outline

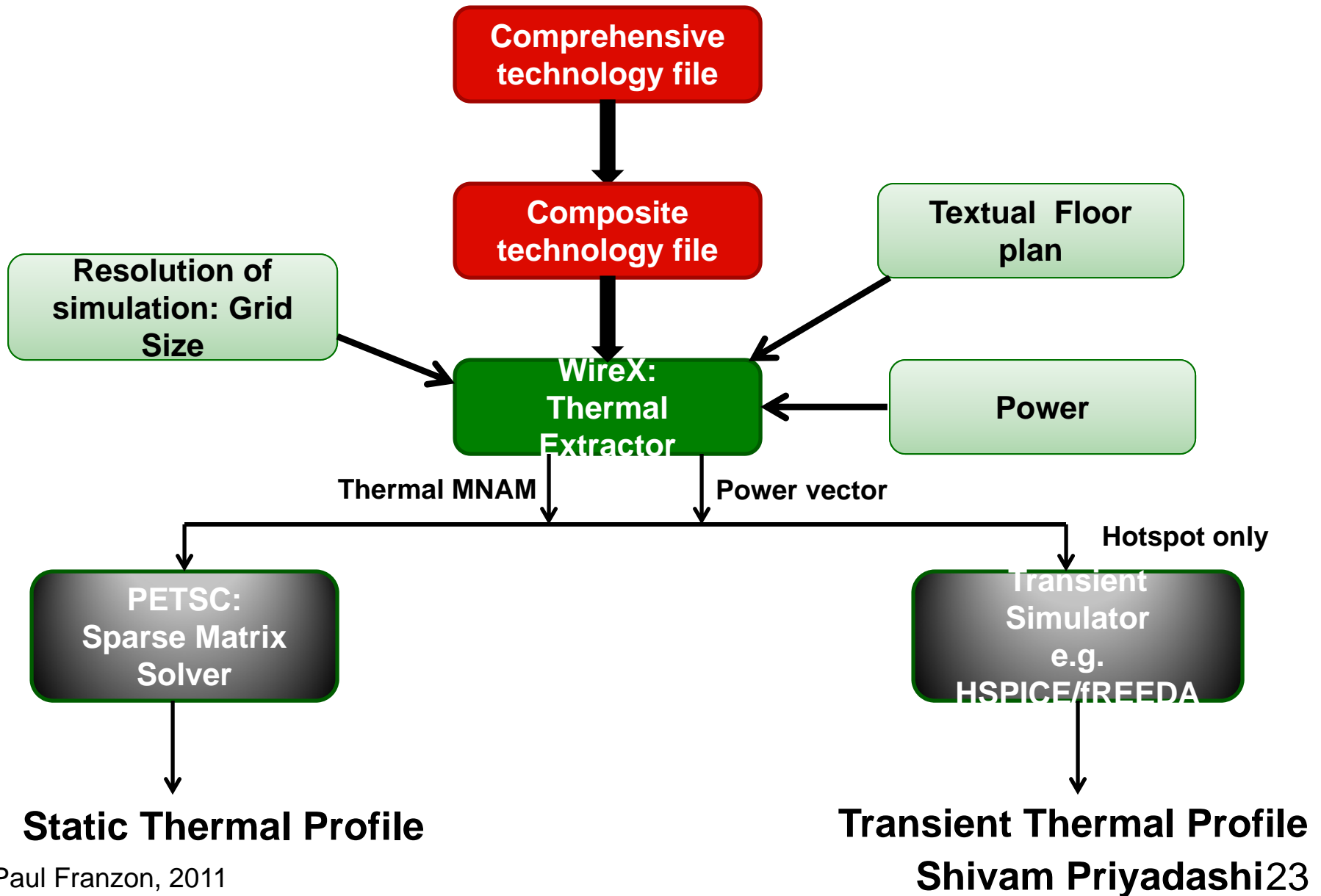
---

- 3DIC Technology Set
- Approaches to 3D Specific Power Minimization
  - Comparative energy/operations
  - Leverage Rent's Rule?
  - The low-hanging fruit : Energy-optimized memory interfaces
  - 3D-specific architectures
  - Next: Heterogenous Integration
- CAD approaches to 3D optimization

# SystemC Methodology for Pathfinding



# Thermal and Physical Flow:



# Conclusions

---

Getting Power Advantages from 3DIC:

Method	Advantages
Logic-on-logic shorter wires	2% - 13% power improvement
Exploiting DRAM on logic as a low power interface	30% improvement if application (mostly) fits in stacked DRAM
Architecting System for Vertical Spatial Locality	Varies. 5% to 60% (?)
Trade area for power	Up to 20% at system level, when done from memories
Exploiting Heterogeneity	15%+ (??)

# Acknowledgements



Faculty: William Rhett Davis, Michael B. Steer,  
Professionals: Steven Lipa, Neil DiSpigna,  
Students: Hua Hao, Samson Melamed, Peter Gadfort, Akalu Lentiro,  
Shivam Priyadarshi, Christopher Mineo, Julie Oh, Won Ha Choi,  
Zhou Yang, Ambirish Sule, Gary Charles, Thor Thorolfsson,  
Department of Electrical and Computer Engineering  
NC State University