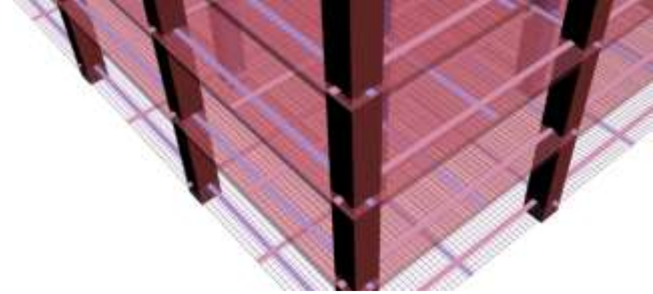
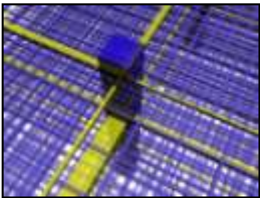
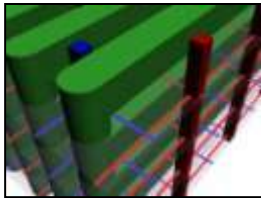
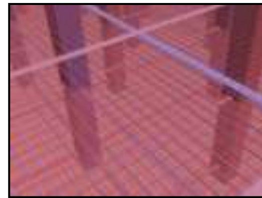
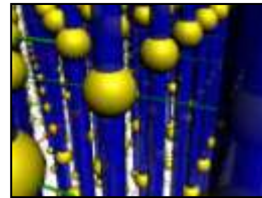


3D Circuit Design with Through-Silicon-Via: Challenges and Opportunities

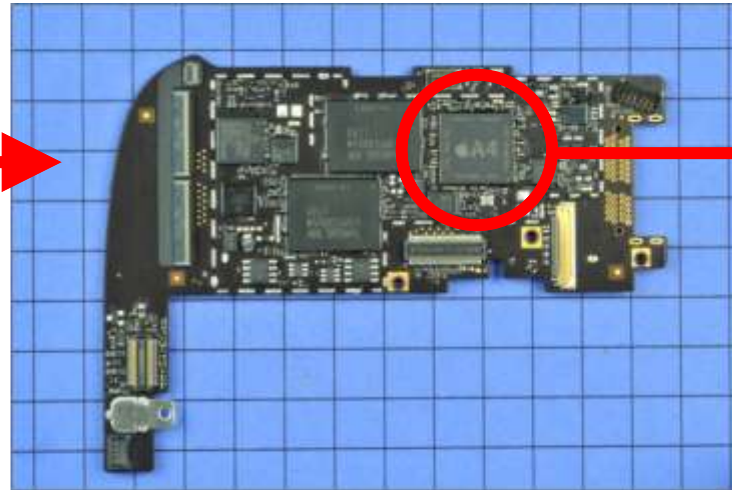





Sung Kyu Lim (limsk@ece.gatech.edu)

GTCAD Laboratory (www.gtcad.gatech.edu)

EDPS 2010 Talk, April 9, 2010

3D IC in the Mainstream (Almost)

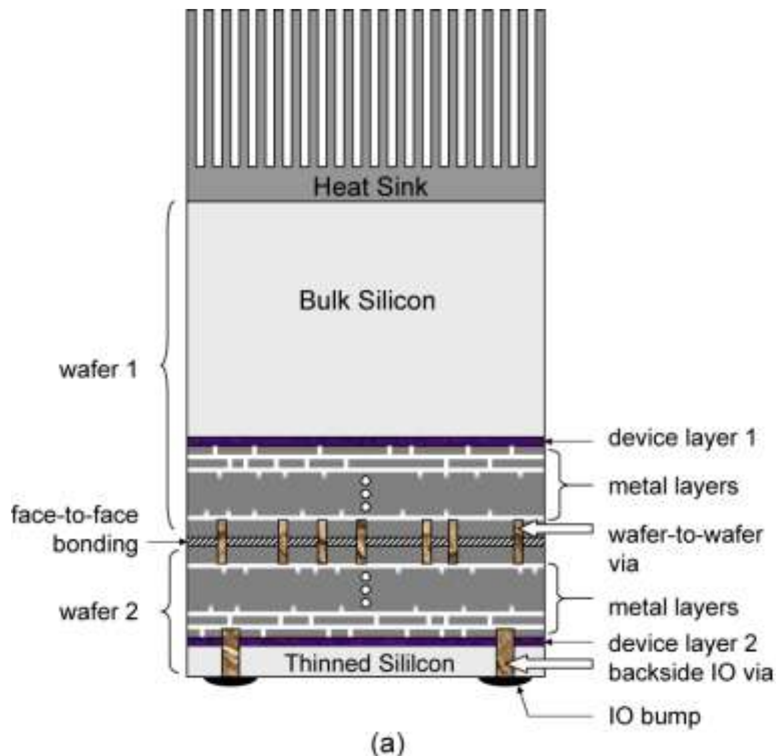


1 ARM die + 2 DRAM dies, stacked, wire-bonded (no TSVs… yet), and packaged

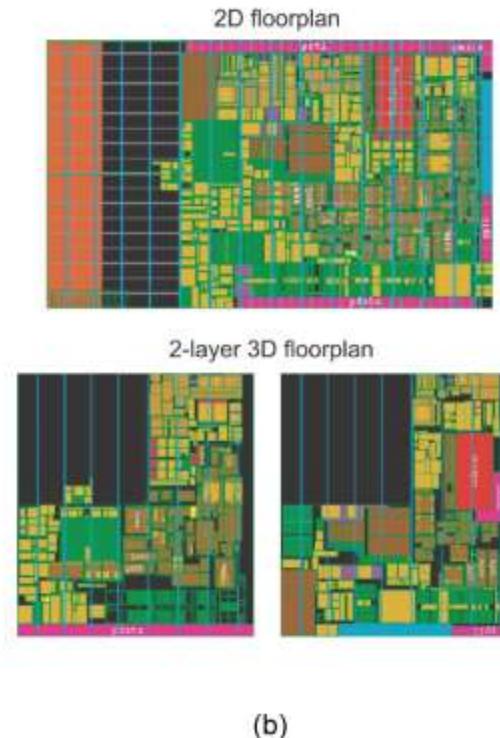
3D IC Design with TSVs

- Inter-tier connection with TSV
  - Pros: area/wirelength benefit (= power/delay benefit) … Wait!
  - Cons: manufacturing cost/yield, thermal/noise issues

actually
it depends



(a)



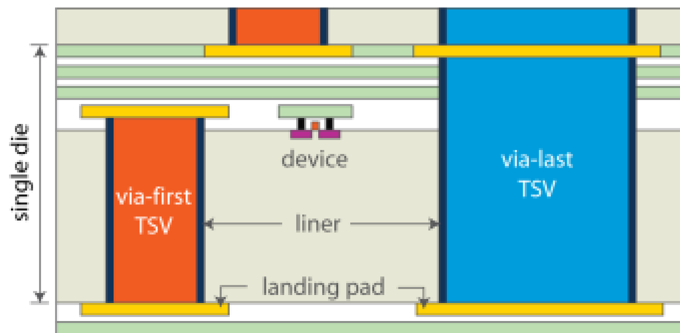
(b)

First, Our Observations

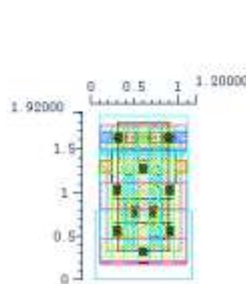
1. TSVs are huge
2. TSV count is crucial
3. TSV location is crucial
4. TSVs cause coupling
5. TSVs require design-for-manufacturability/reliability
6. TSVs require design-for-testing

Fact 1: TSVs are Huge

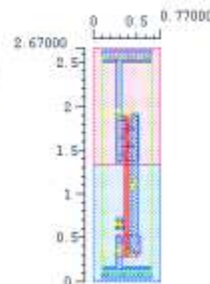
- **A few times larger** than gates & memory cells

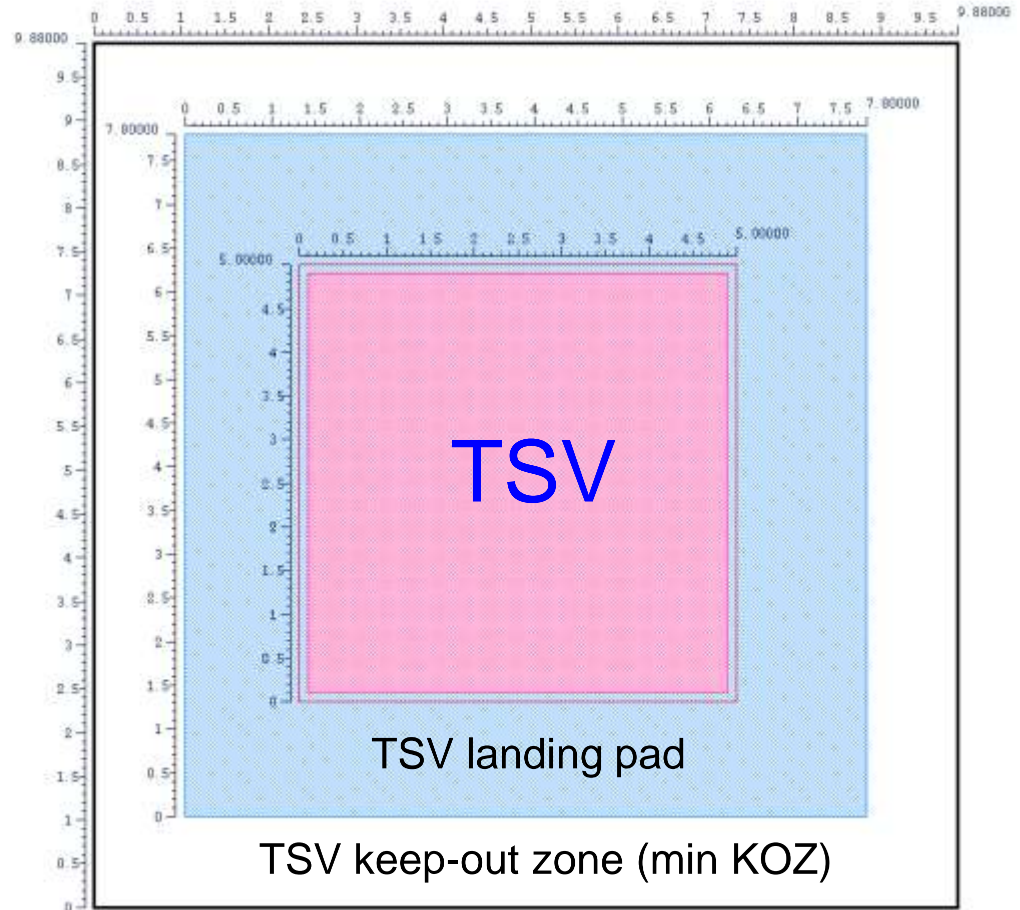


SRAM cell

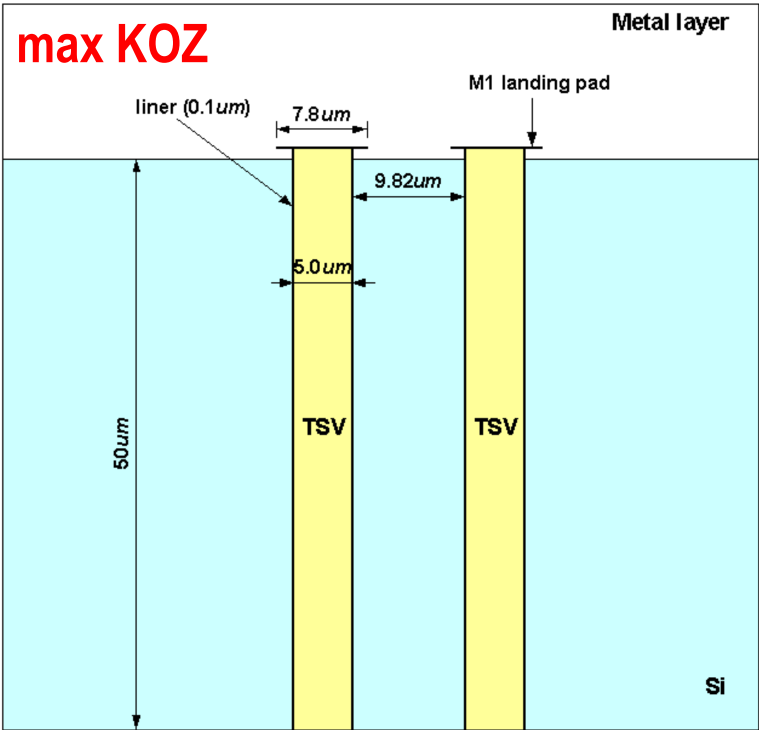


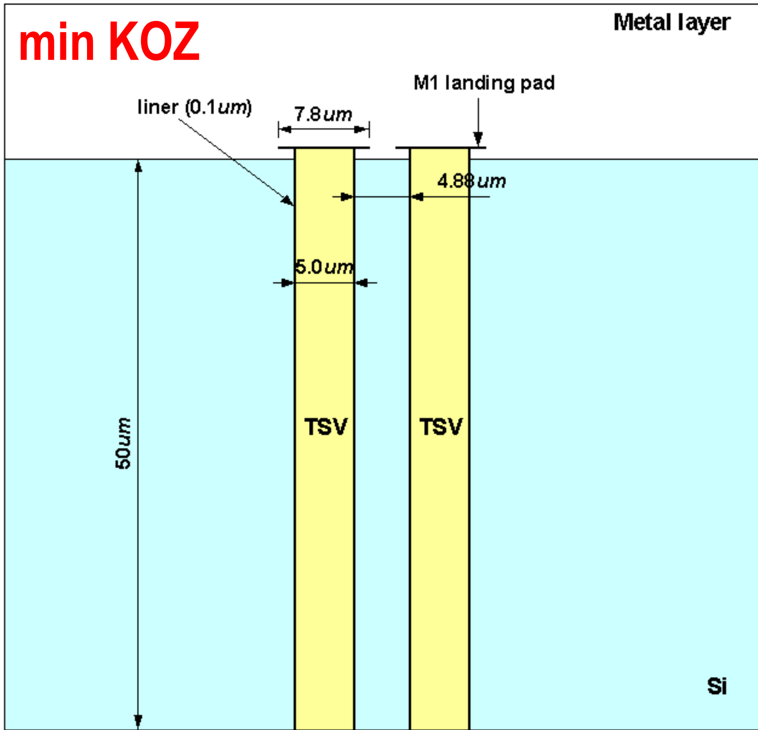
inverter



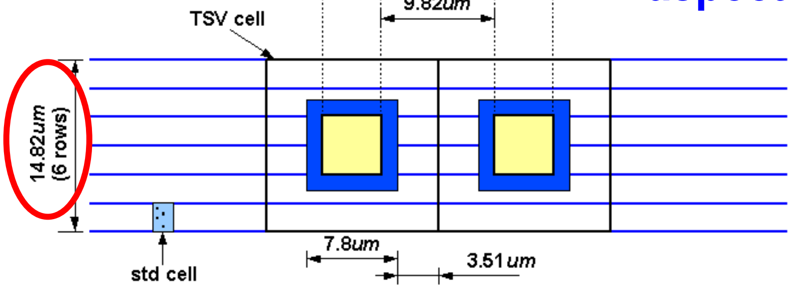


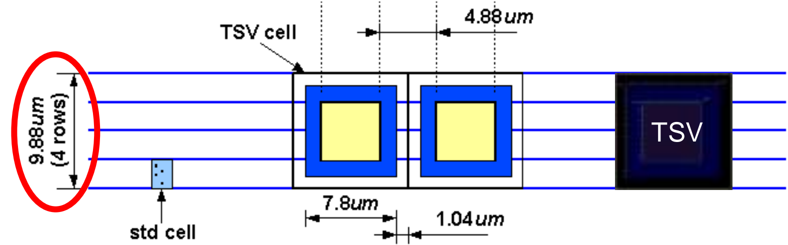
Popular TSV Dimensions





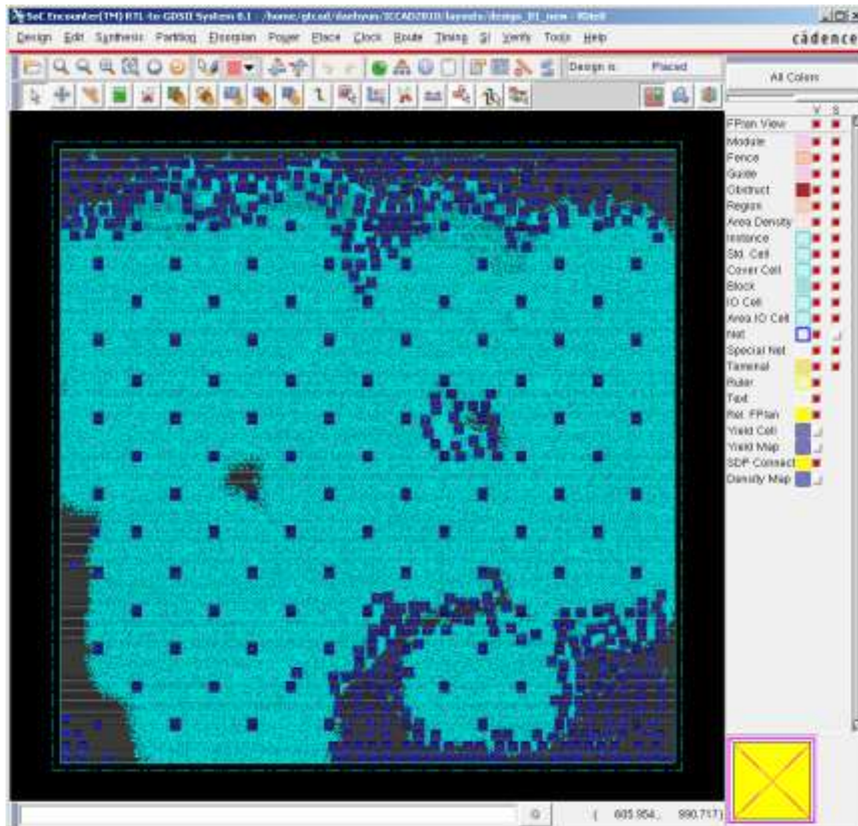
aspect ratio = 1:10

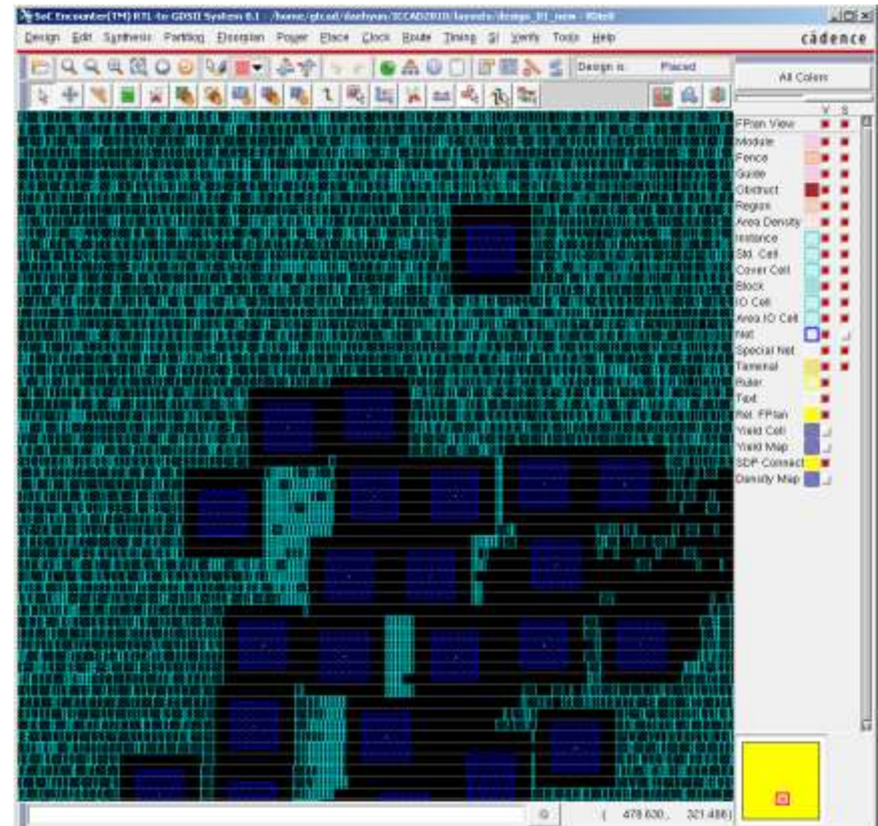




Layouts with TSVs

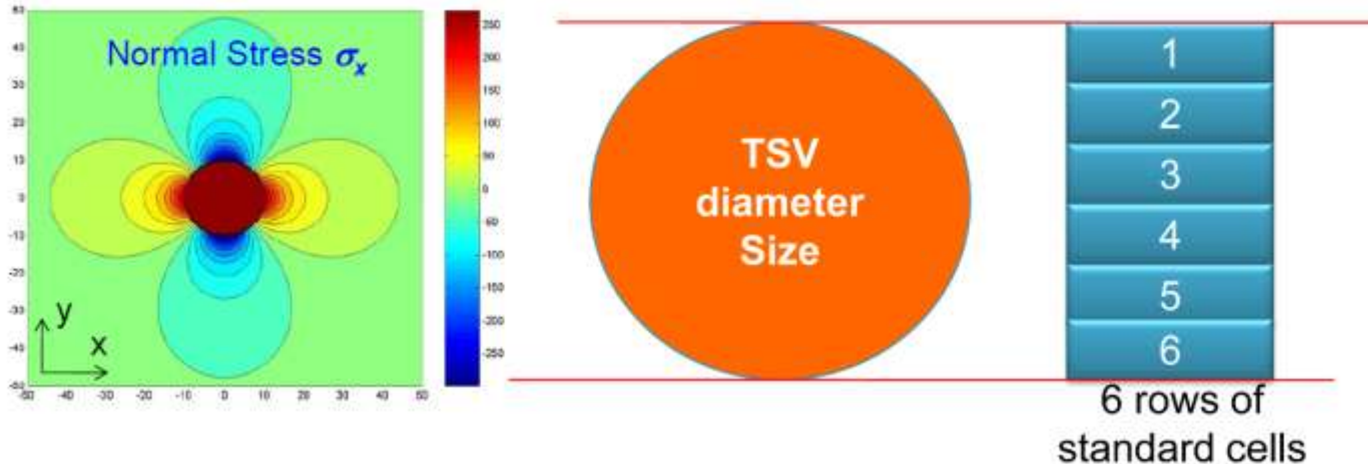
- Using our RTL-to-GDSII tool-flow for 3D ICs





According to Cadence

TSV size ➔ Impact on placement and routing



- TSV cut size is about 5-10X the height of standard cell in 32 nm technology.
  - TSV placement disturbs standard cell row placements
- TSV cut size is about 15-30X M1 min-width.
  - Special routing rules for M1: Use of max width wire
- TSV thermo-mechanical stress has impact on mobility of nearby devices
  - Best handled with keep out area from diffusion area
    - Small distance to digital cells and bigger distance near analog cells.

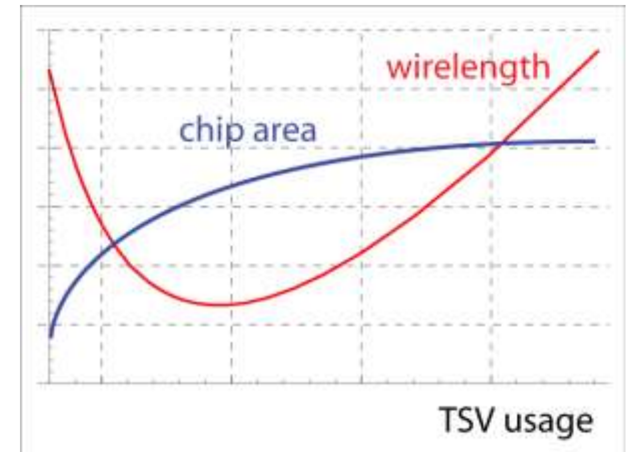
15 3/6/2010 Physical Design with 3D IC

cādence

David Noice & Vassilios Gerousis Cadence, ISPD 2010

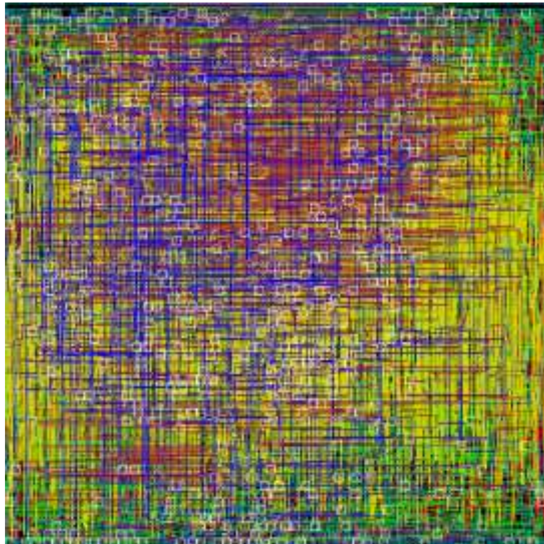
Fact 2: TSV Count is Crucial

- Stacking 2D dies
  - Small TSV count
  - Ex: 3D MAPS
- Placement of modules across 3D stack
  - Medium TSV count (hundreds to thousands)
  - Ex: some of our research prototypes
- Placement of gates across 3D stack
  - Large TSV count (thousands and up)
  - Ex: some of our research prototypes
- Placement of transistors across 3D stack
  - True 3D device: P-diff on top, N-diff on bottom
  - Extremely large TSV count

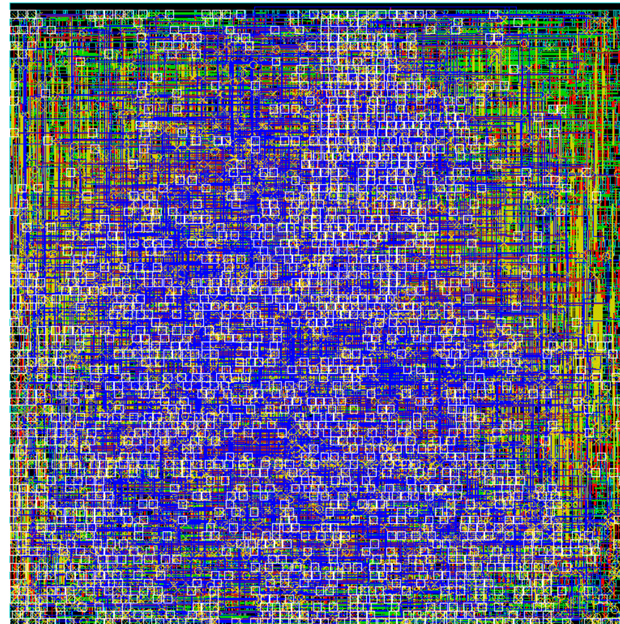


On the Bad Side of the Curve

- **Benchmark: 8086 (20K gates)**



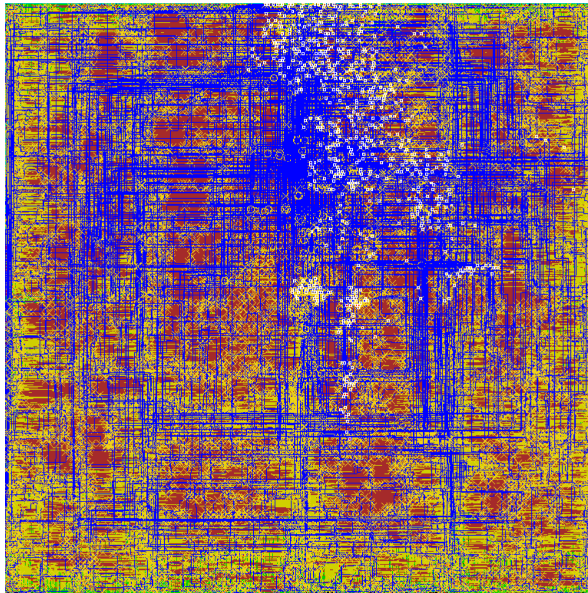
TSV = 467, area = 165x165um, WL = 165mm



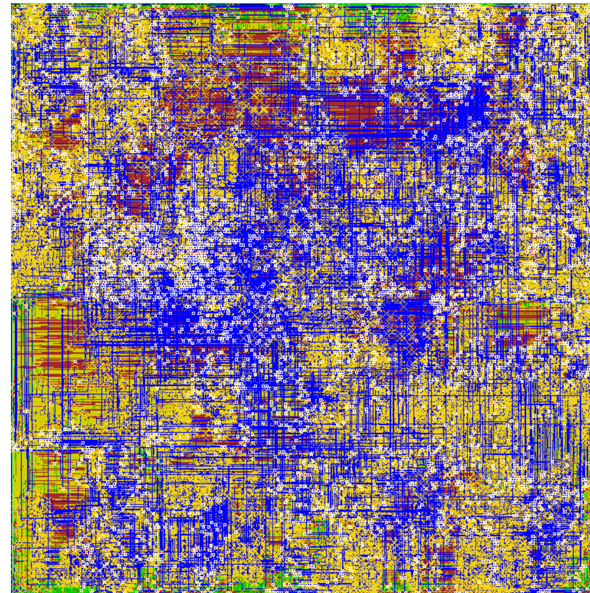
TSV = 1659, area = 195x195um, WL = 205mm

On the Good Side of the Curve

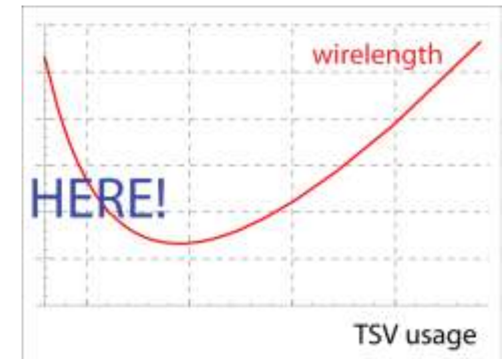
- **Benchmark: OpenSPARC (300K gates)**



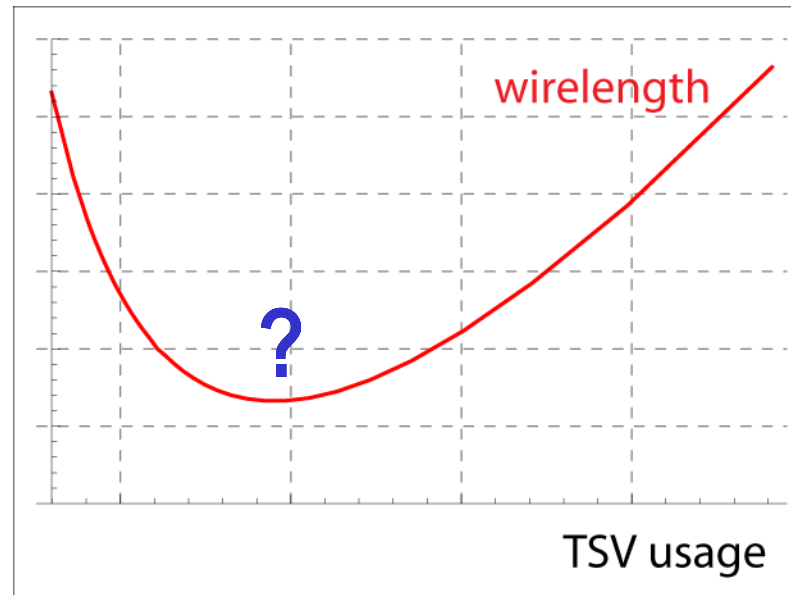
TSV = 629, area = 655x655um, WL = 3,015mm



TSV = 4482, area = 642x642um, WL = 2,606mm



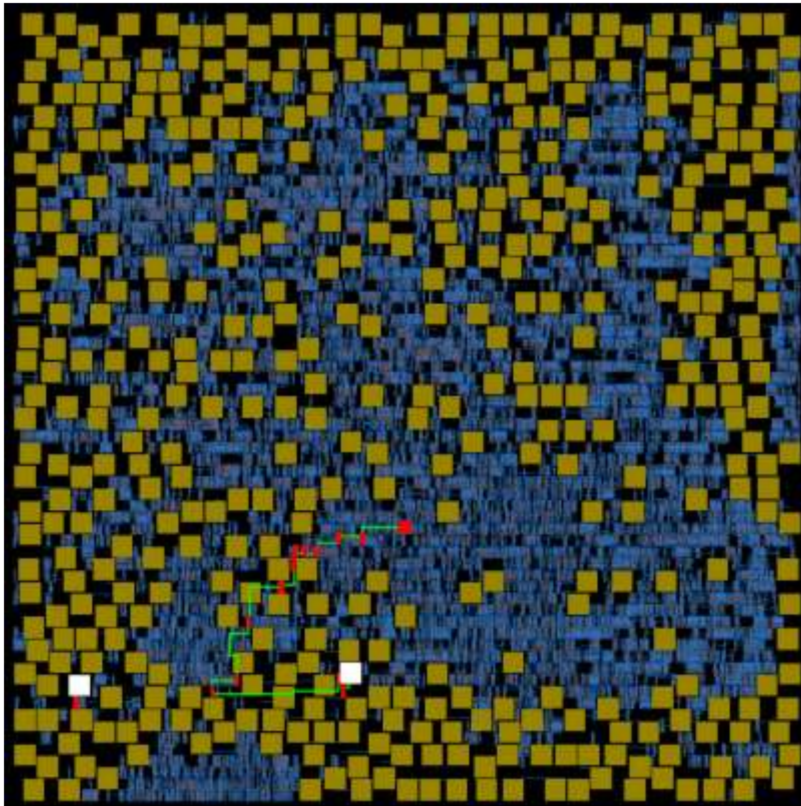
Then, the Question is…

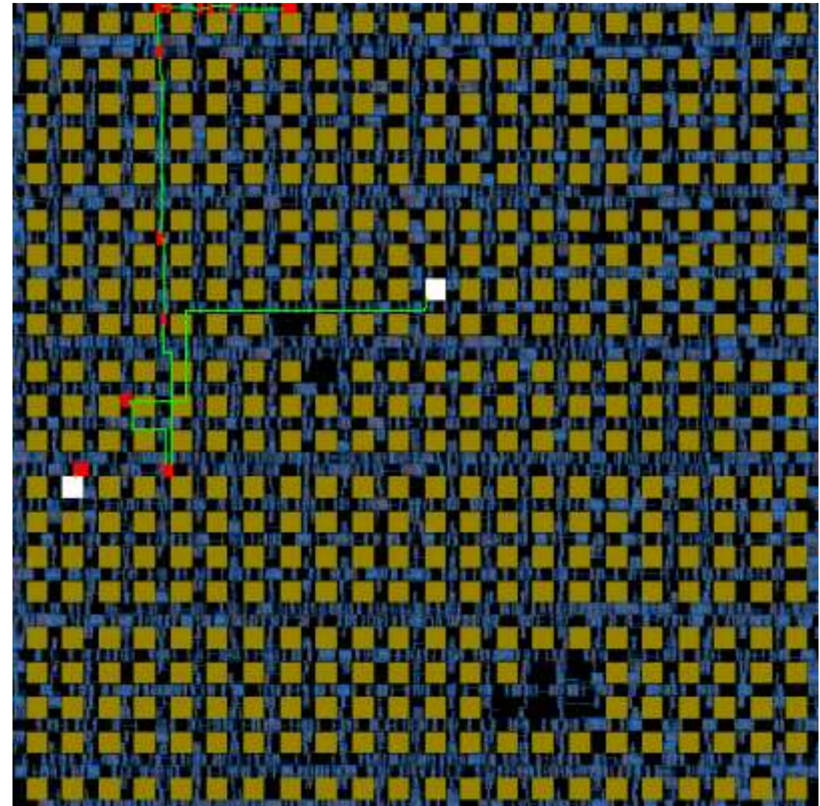


How do you find the sweet spot?
What about delay, power, manufacturability, cost?

Fact 3: TSV Location Is Crucial

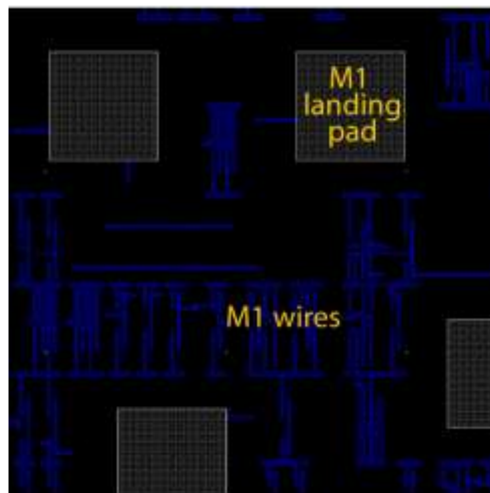
- Two extreme TSV placement style: **non-regular vs regular**
  - Wirelength/timing vs manufacturability tradeoffs



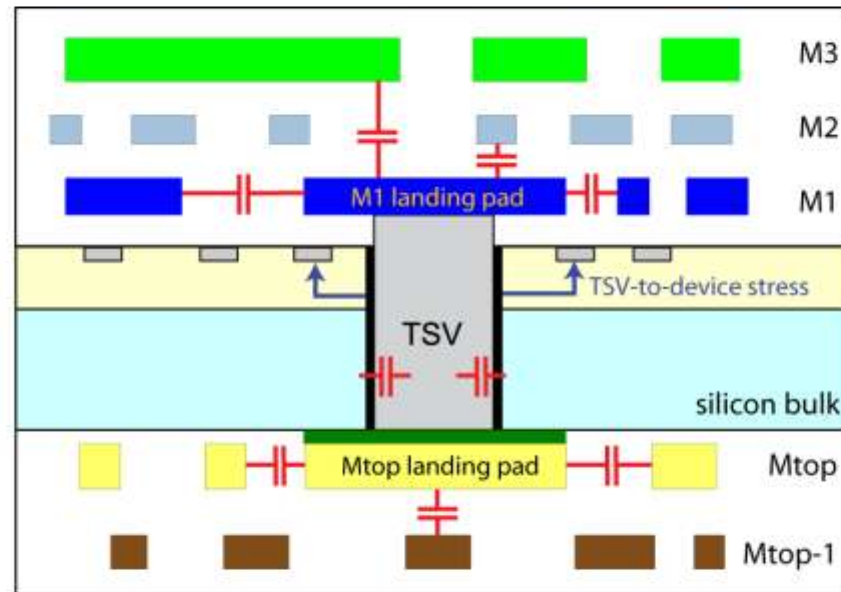


Fact 4: TSVs Cause Coupling

- TSVs (via-first) occupy M1, Mtop, device layer, substrate
  - Serious layout obstacles
- Coupling affects timing & power
  - What is the right KOZ (keep out zone) size then?



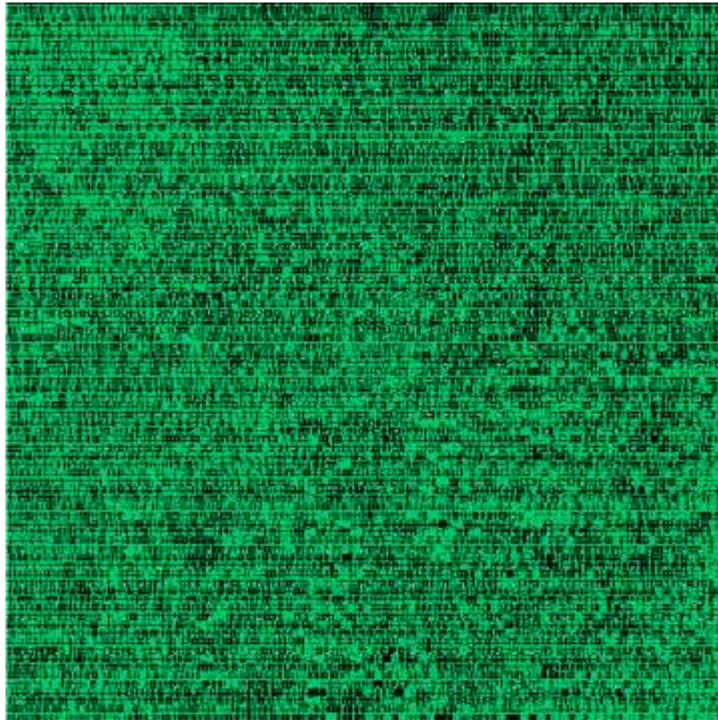
top-down view



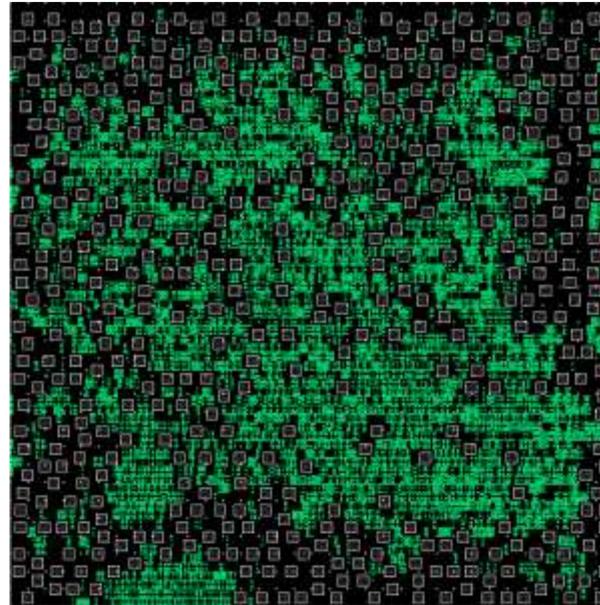
side view

Fact 5: TSVs Require DFM/DFR

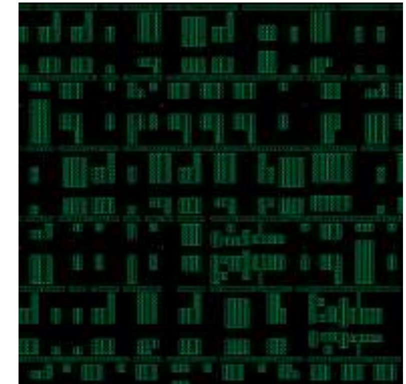
- TSV causes CMP and stress issues (no litho issue, fortunately)



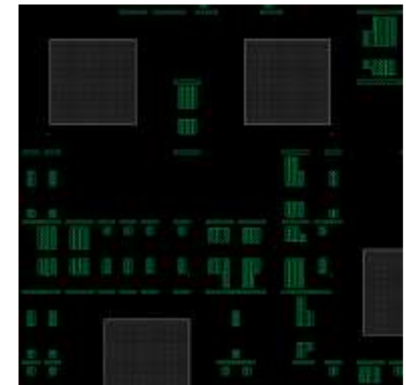
2D layout (no TSV)
Inverse discrete cosine transformation (IDCT) @ 45nm



3D layout (w/ TSV)
top-tier of 4-die IDCT
TSV diameter = 3um

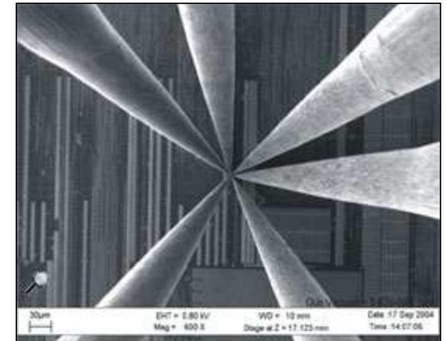


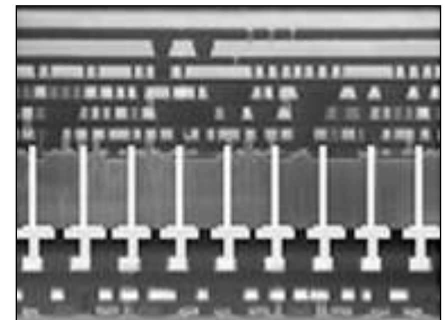
2D zoom-in



3D zoom-in

Fact 6: TSVs Require DFT



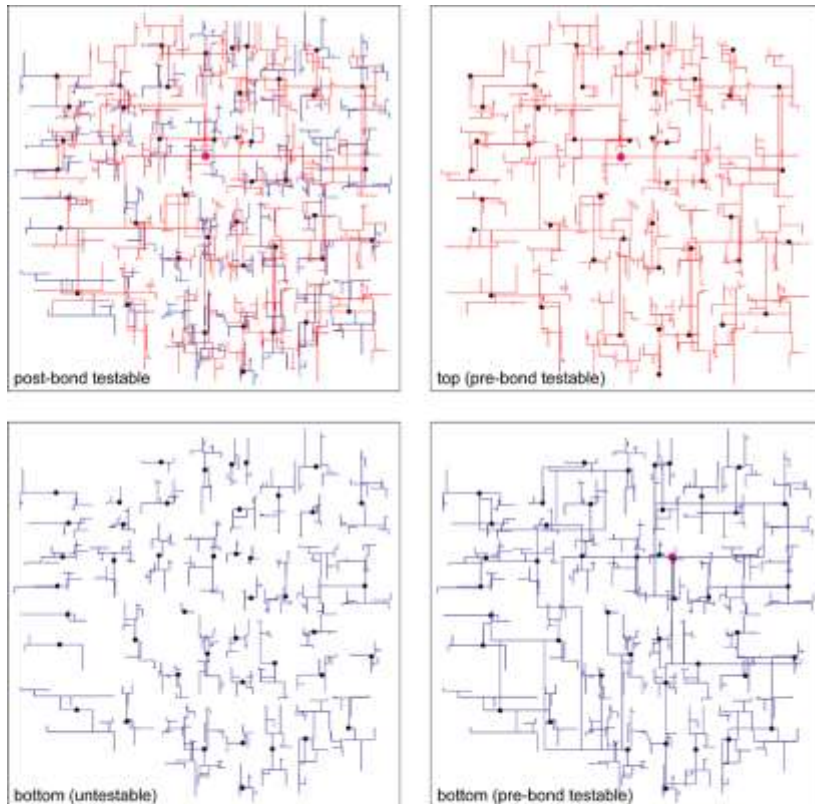


- **Pre-bond testing**
  - Important because we want to stack **known-good-dies**
  - Test probe pads are BIG: 50um < diameter
  - How to provide clock, power, signal to individual dies?
  - How do we test TSVs in each die?
  - How do we minimize/recycle pre-bond testing resources?
- **Post-bond testing**
  - Burn-in test: may permanently damage dies/TSVs
  - How do we test TSVs after bonding?

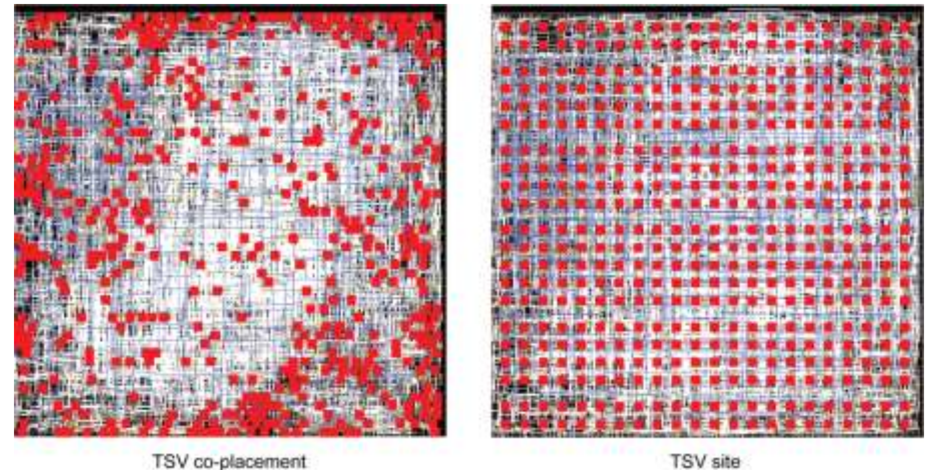
Ongoing GTCAD Projects

- Physical Design Automation for 3D Circuits (NSF)
- DFM/DFR for TSV-based 3D ICs (Intel)
- Design, Fabrication, and Testing of 3D-MAPS (DOD)
- Interconnect Limit Study with Many-Tier 3D System (FCRP/IFC)
- 3D Integration of Sub-Threshold Multi-core Co-processor (NSF)
- 3D IC Integration with Interposers (GT-PRC)
- Heterogeneous 3D Integration (SRC, IBM, Intel)
- DFT for TSV-based 3D ICs (SRC, IBM, Intel)

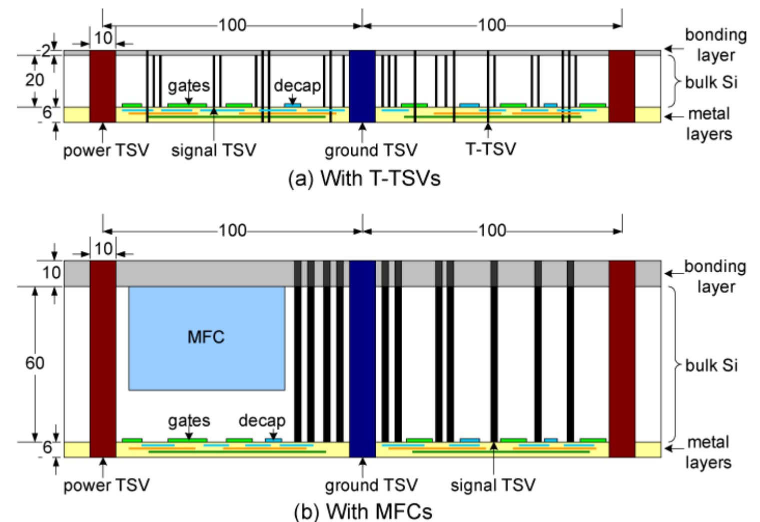
Our 3D Works at ICCAD 2009



1. Pre-bond testable 3D clock routing [ICCAD'09]
Best-paper award candidate



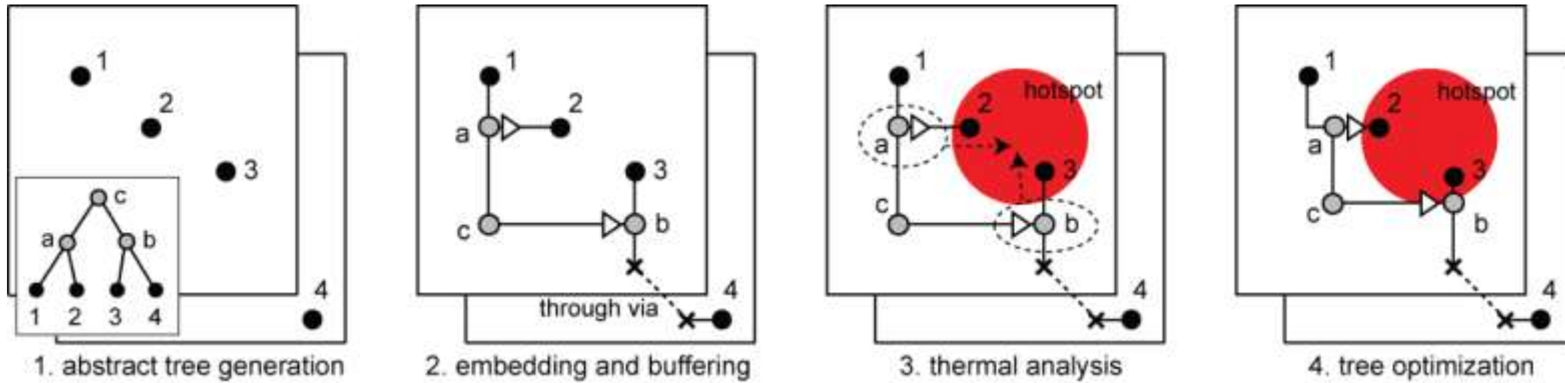
2. TSV placement and routing [ICCAD'09]

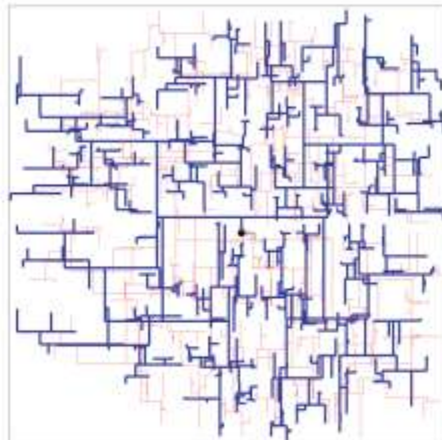


3. Liquid cooling with micro-fluidic channels [ICCAD'09]

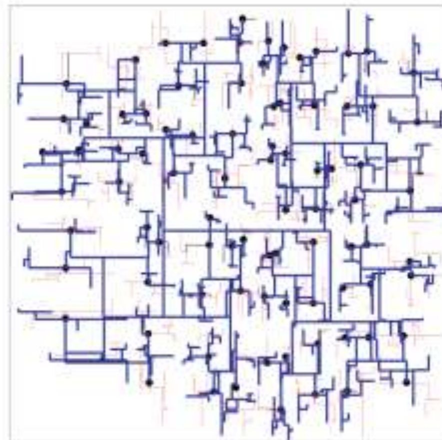
3D Clock Routing

- **TSVs reduce clock tree wirelength (= power)**

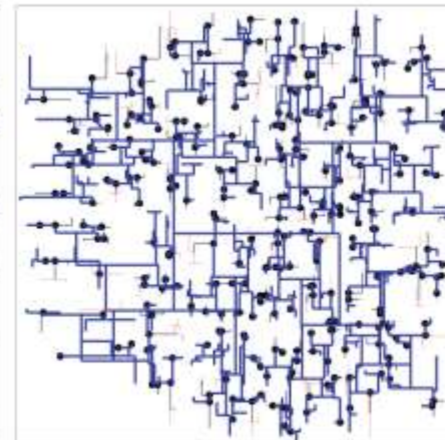




TSV = 1, WL=532mm



TSV = 87, WL=446mm



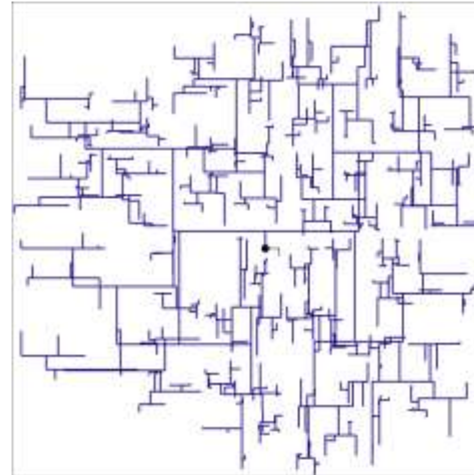
TSV = 306, WL=387mm

But, TSVs Cause Tree Fragments

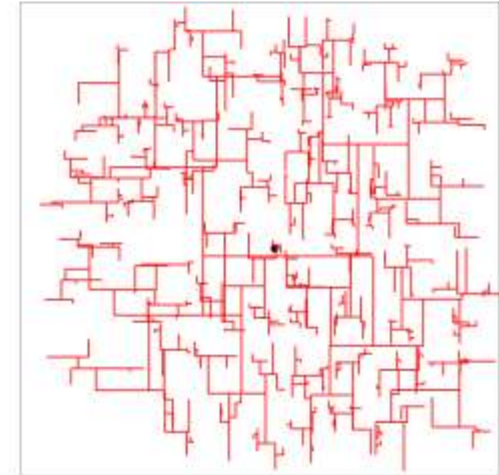
single TSV

pre-bond testable

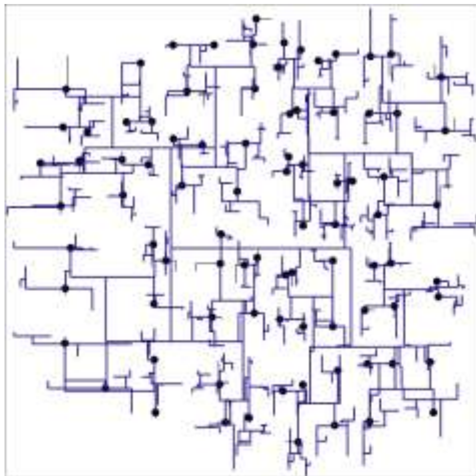
high power consumption



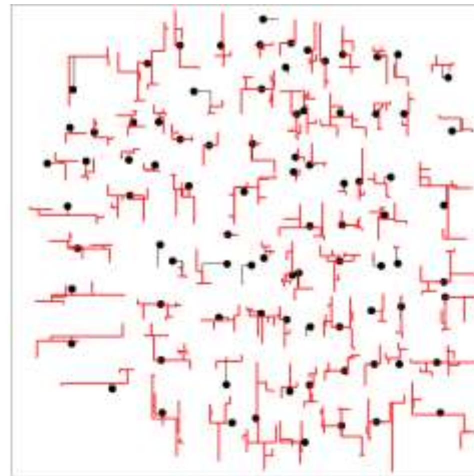
via = 1 (layer 1)



via = 1 (layer 2)



via = 87 (layer 1)



via = 87 (layer 1)

multiple TSVs

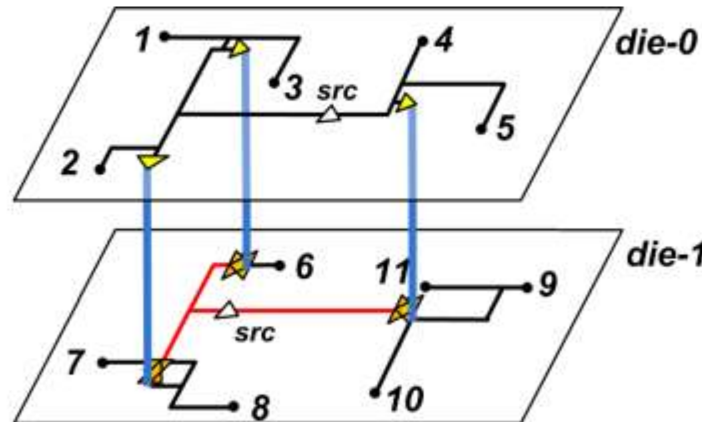
NOT pre-bond testable

low power consumption

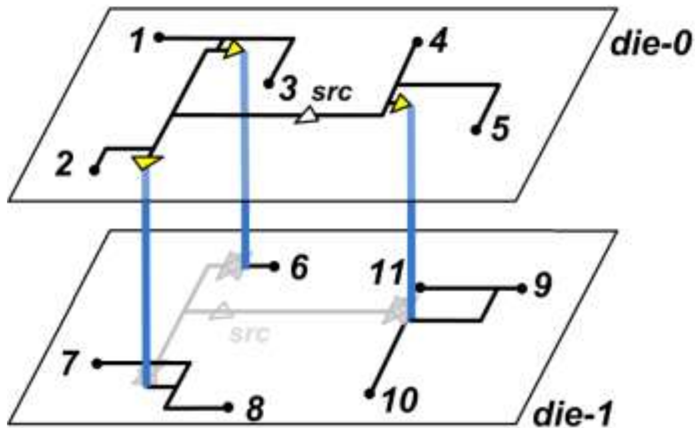
Adding Pre-bond Testability

Key ideas

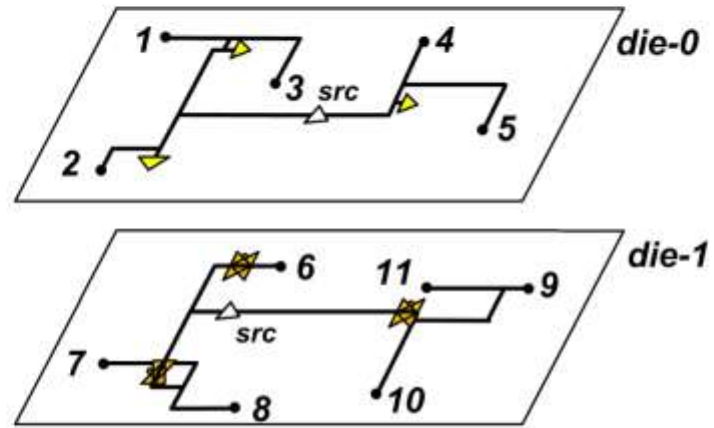
- Redundant tree
- TSV Buffer



Pre-bond testable 3D clock tree

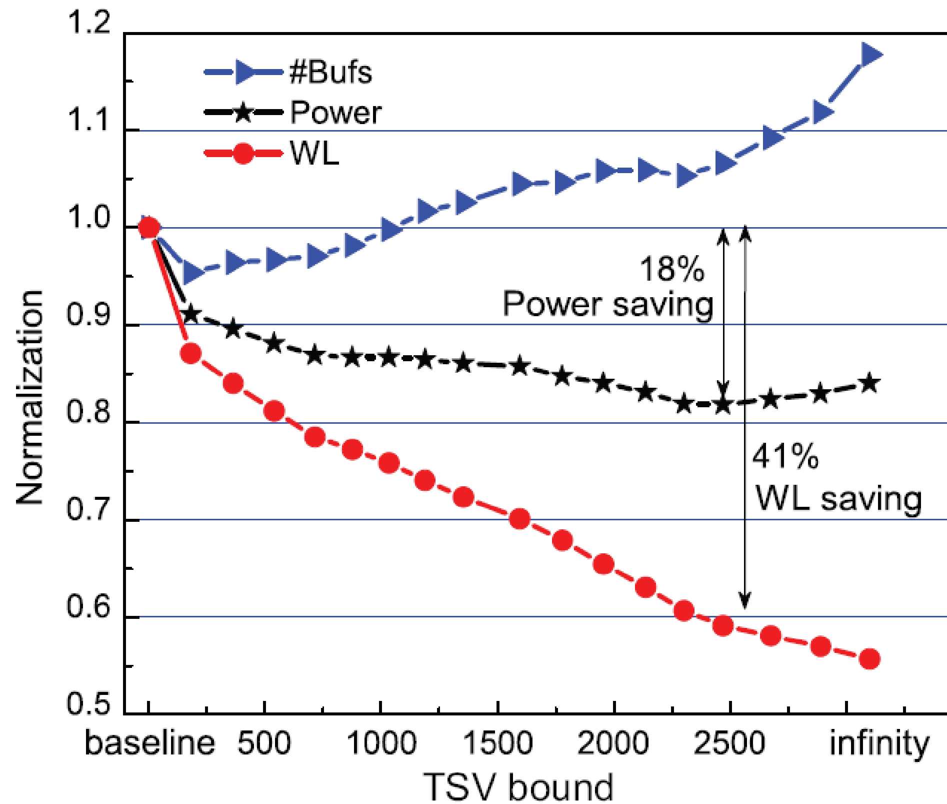


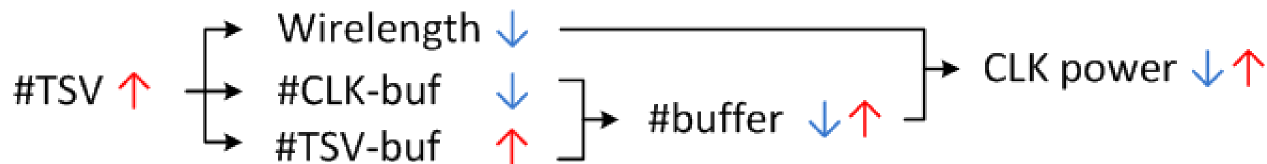
In post-bond operations



In pre-bond testings

Impact of TSV on Clock Tree





Obstacles during 3D Clock Routing

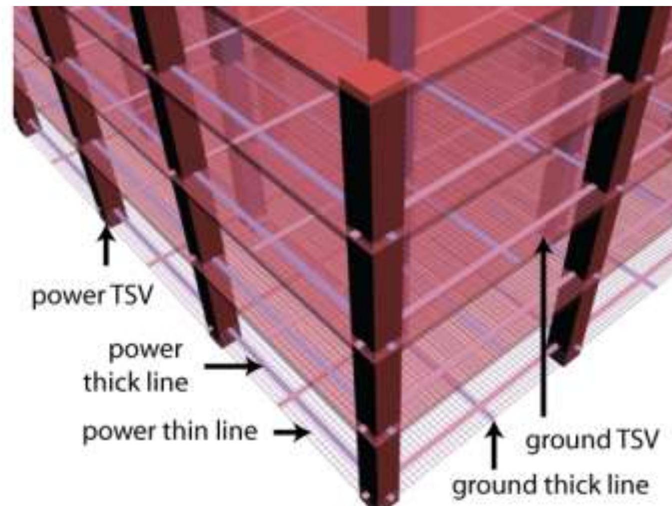
- TSVs come in your way
  - P/G TSVs: placement and routing obstacles
  - Signal TSVs: placement obstacles
  - Clock TSVs: same as signal TSVs, being added during routing
  - Clock buffers: same as clock TSVs
  - Sink nodes (= FFs): placement obstacles

P/G TSV
signal TSV
clock TSV
clock buffers

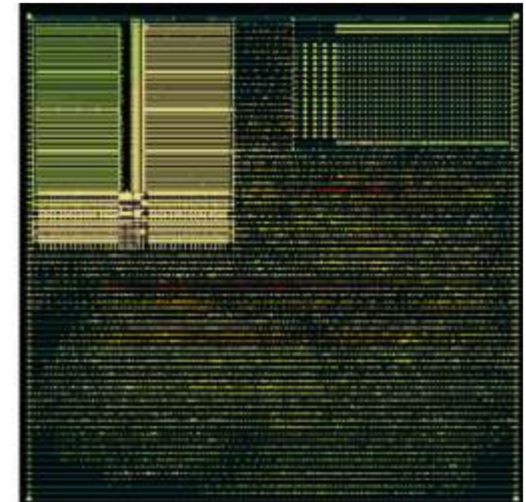
3D Power Delivery

- Big Challenge
  - Fewer power bumps available due to smaller footprint
  - Need many big P/G TSVs to deliver power vertically: causes congestion
  - On-chip solutions: dedicated decap tier, dedicated P/G TSVs
  - Off-chip solutions: on-package decaps, P/G TSVs

TSVs, gates, and P/G strips



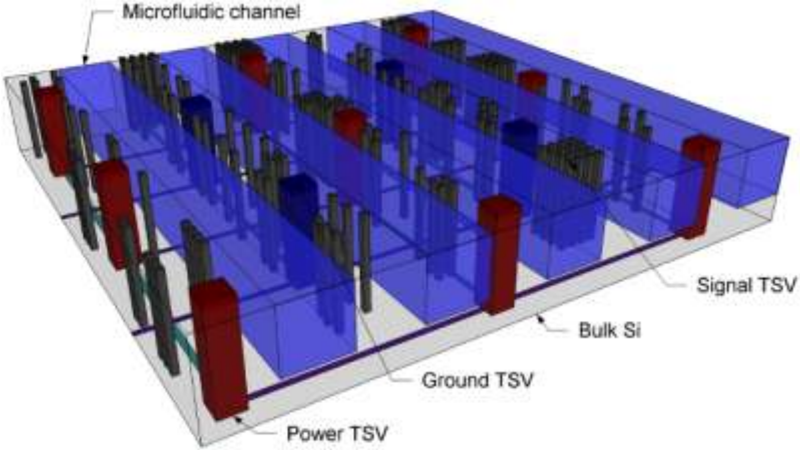
P/G network for 4-tier 3D IC

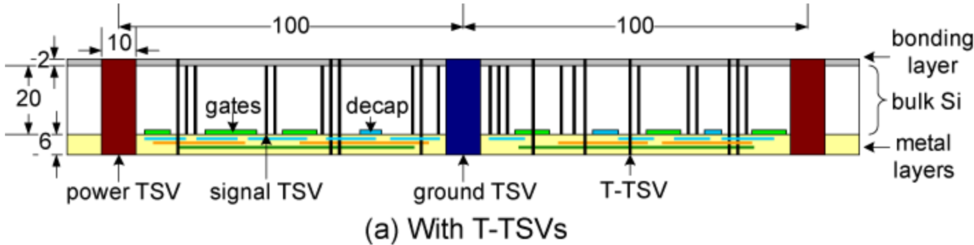


IR-drop analysis

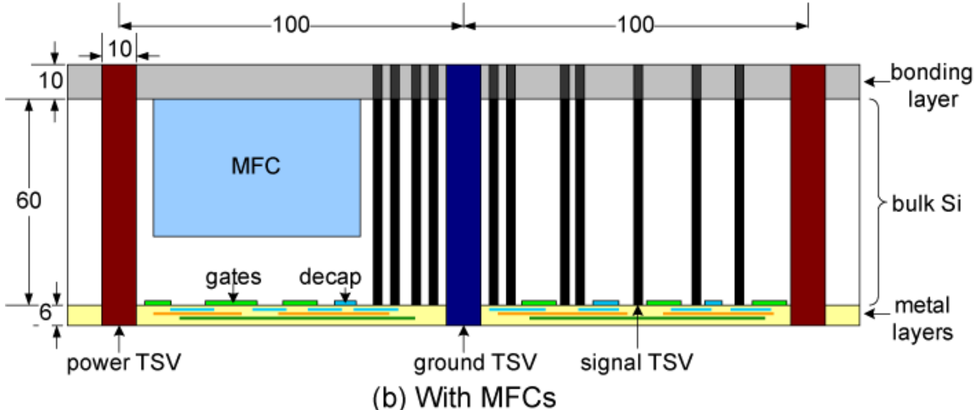
MFC-based Liquid Cooling

- MFC vs thermal TSVs [ICCAD'09]
  - Cooling capacity vs layout resource overhead tradeoff exist

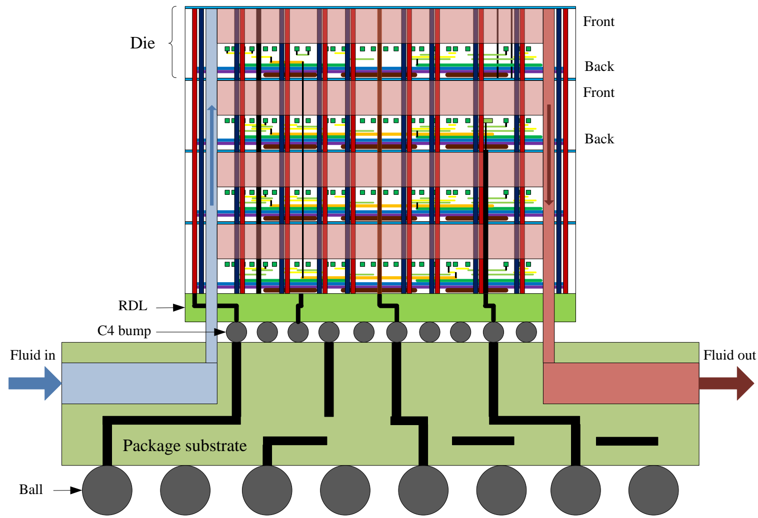




(a) With T-TSVs



(b) With MFCs



Sizing Interconnects: DOE

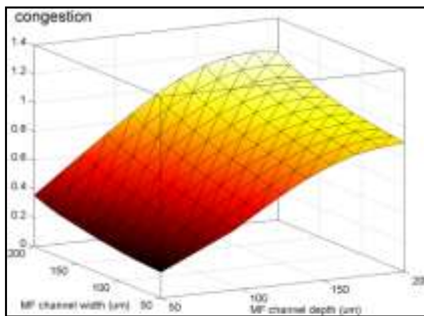
- 7 Tuning Knobs and Ranges

Microfluidic channel depth	50-200 um
Microfluidic channel width	50-200 um
Microfluidic channel pitch	200/400/600/800 um
Pressure drop of working fluid	100/120/140/160/180 kPa
PG TSV diameter	20/40/60/80 um
PG TSV pitch	400/800 um
PG grid thin wire ratio	0.2-0.8

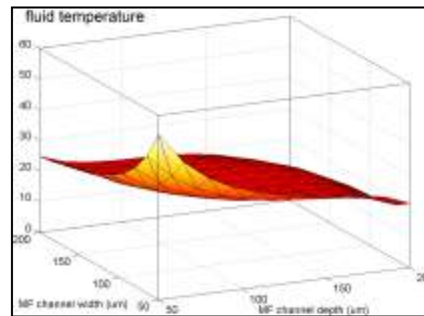
- 7 Assessing Metrics
  - Total signal net wirelength, total number of signal TSVs, congestion,
  - Max Silicon wall temperature, max working fluid temperature, pump power
  - Max power noise level

Response Surface Method

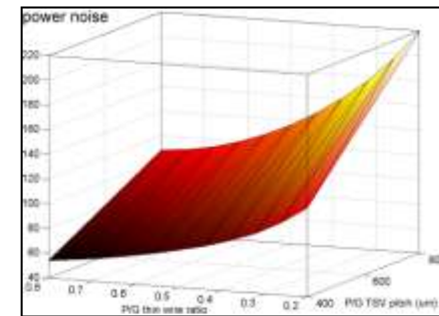
- **Response Surface Method**
  - identifies important knobs for each metric and tunes them



congestion (MFC width vs depth)



fluid temp (MFC width vs depth)



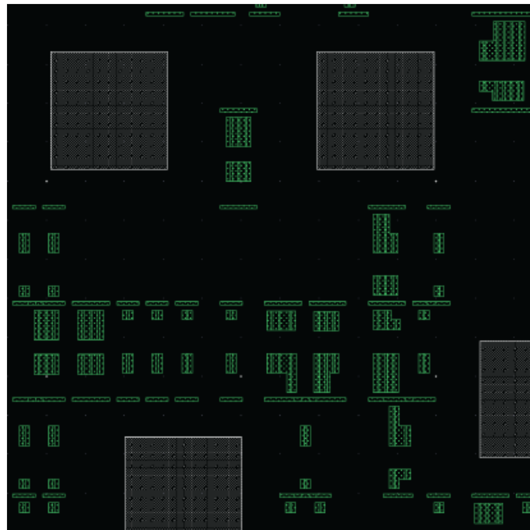
noise (P/G wire density vs TSV pitch)

knob	optimized	initial
MF channel depth	57um	100um
MF channel width	86um	100um
MF channel pitch	800um	200um
Pressure drop	140kPa	140kPa
P/G TSV diameter	60um	40um
P/G TSV pitch	400um	400um
P/G thin wire ratio	0.8	0.5

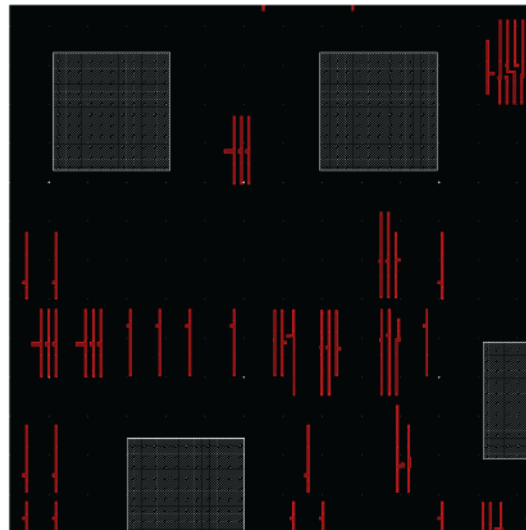
metric	predicted	actual
Total wirelength	3.77e8	3.77e8
Total # signal TSVs	43516	43304
Max total TSV utilization	0.228	0.220
Max fluid temp	43.4	40.0
Max Si wall temp	84.9	80.8
Pump power	0.119	0.097
Power noise	54.1	55.2

TSVs Are Large: DFM Issues

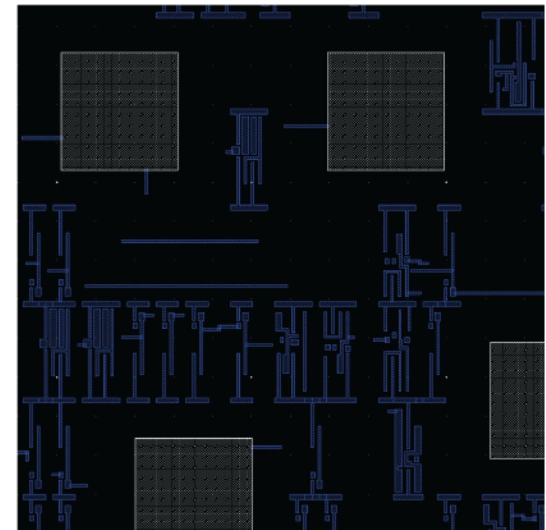
- What kind of DFM issues do TSV cause?
  - Density, CMP (serious)
  - Lithography (turned out not too much)
  - Stress and reliability (serious)



active layer (zoom in)



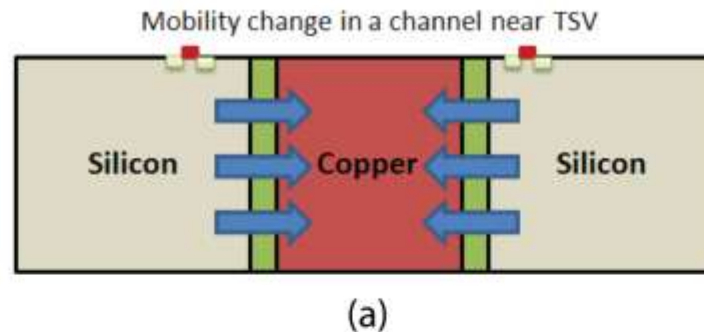
poly layer (zoom in)



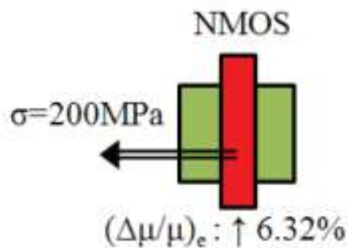
M1 layer (zoom in)

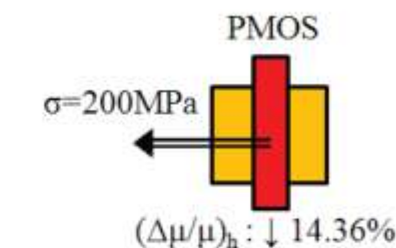
TSV Impact On Carrier Mobility

- Caused by CTE mismatch during fabrication and operation
  - Vertical tension is good for both PMOS and NMOS
  - Horizontal: TSV pulls the substrate, and only NMOS benefits

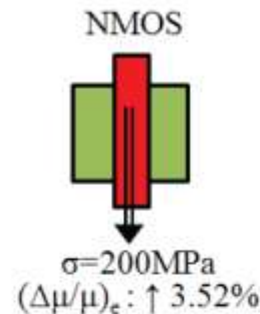


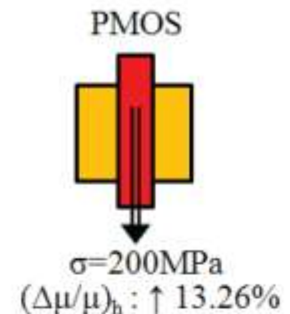
(a)





(b)

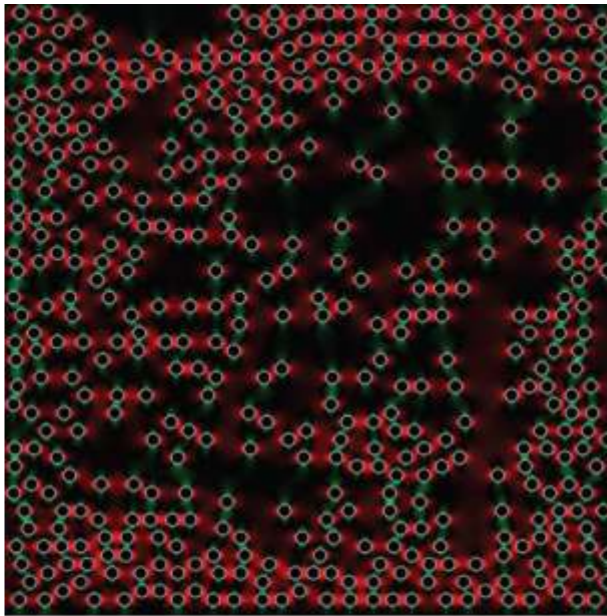




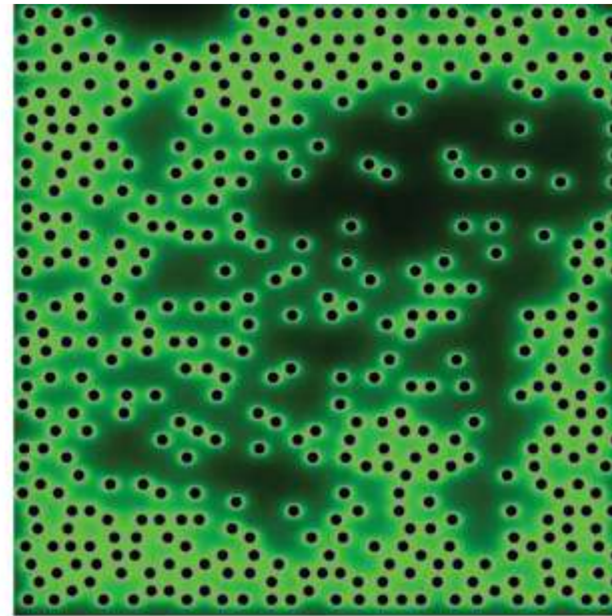
(c)

TSV Stress Map and STA

- TSV stress causes timing variations [DAC 2010]
  - Degrades/improves timing
  - Useful to guide placement



hole mobility change



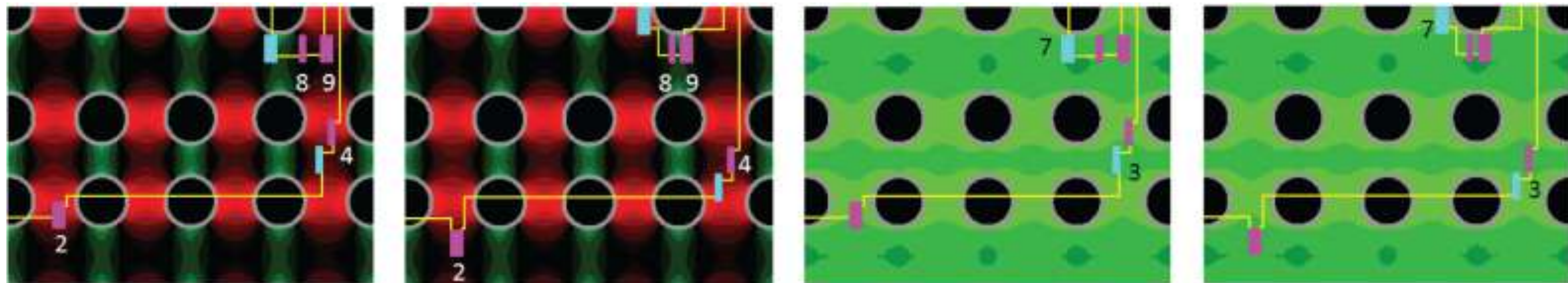
electron mobility change

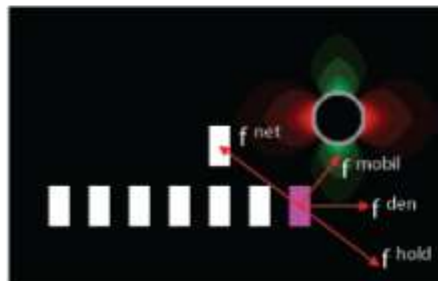
Collaborator: Prof. David Pan (UT)

TSV Stress-aware Optimization

- **Placement optimization [DAC'10]**
  - Manual optimization: bring timing critical gates closer to brighter green
  - Automatic placement: add forces to balance area, stress, density

hole mobility map　　　　electron mobility map





(a)

(b)

(c)

3D MAPS Processor V1

Architecture and Memory Model

- number and type of cores: 64, 5-stage, in-order, 2-way VLIW
- memory capacity: 256KB SRAM
- 3D stacking: 2 tiers face-to-face bonded (= core + memory)
- memory model: dedicated 4KB SRAM tile per core
- memory latency: 1 clock cycle, 1 read per every instruction
- memory bandwidth achieved: 61.3GB/sec peak achievable

Technology, Performance, and Power

- technology: Chartered Semiconductor 130nm
- footprint area: 5mm x 5mm
- clock frequency: 277MHz
- operating voltage: 1.5V
- maximum power consumption: up to 6

Reliability

- maximum IR-drop: up to 78mV
- maximum coupling noise: 574 mV
- clock skew/slew: skew = 82ps, slew = 117ps
- maximum temperature: coming up

TSVs, Face-to-face (F2F) Vias, and IOs

- TSV diameter and pitch: 1.2um, 5um (Tezzaron)
- F2F via diameter and pitch: 3.4um, 5um (Tezzaron)
- total TSV count: 2240 (= 35x64) dummy, 27940 (= 204x235) IO
- total F2F via count: 7424 (= 116x64) signal, 43776 (= 684x64) P/G
- total IO count: 14 signal, 205 P/G (1.5V), 16 P/G (2.5V)

arguably the FIRST many-core 3D processor from academia

- designed to demonstrate memory BW/power benefit of 3D processor

core tier (64 cores)

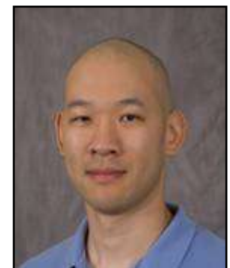
memory tier (SRAM)

The Team

- Faculty
  - Profs. Hsien-Hsin S. Lee, Sung Kyu Lim, Gabriel H. Loh
- Students
  - Core team (4): Mohammad Hossain, Dean Lewis, Tzu-Wei Lin, Dong Hyuk Woo
  - Memory team (1): Guanhao Shen
  - CAD team (11): Krit Athikulwongse, Rohan Goel, Michael Healy, Moongon Jung, Dae Hyun Kim, Young-Joon Lee, Chang Liu, Brian Ouellette, Mohit Pathak, Hemant Sane, Xin Zhao
- Collaborators
  - Package/board design: Dr. Daehyun Chung (GT), Prof. Joungho Kim (KAIST), Prof. Madhavan Swaminathan (GT)

LEE, Core

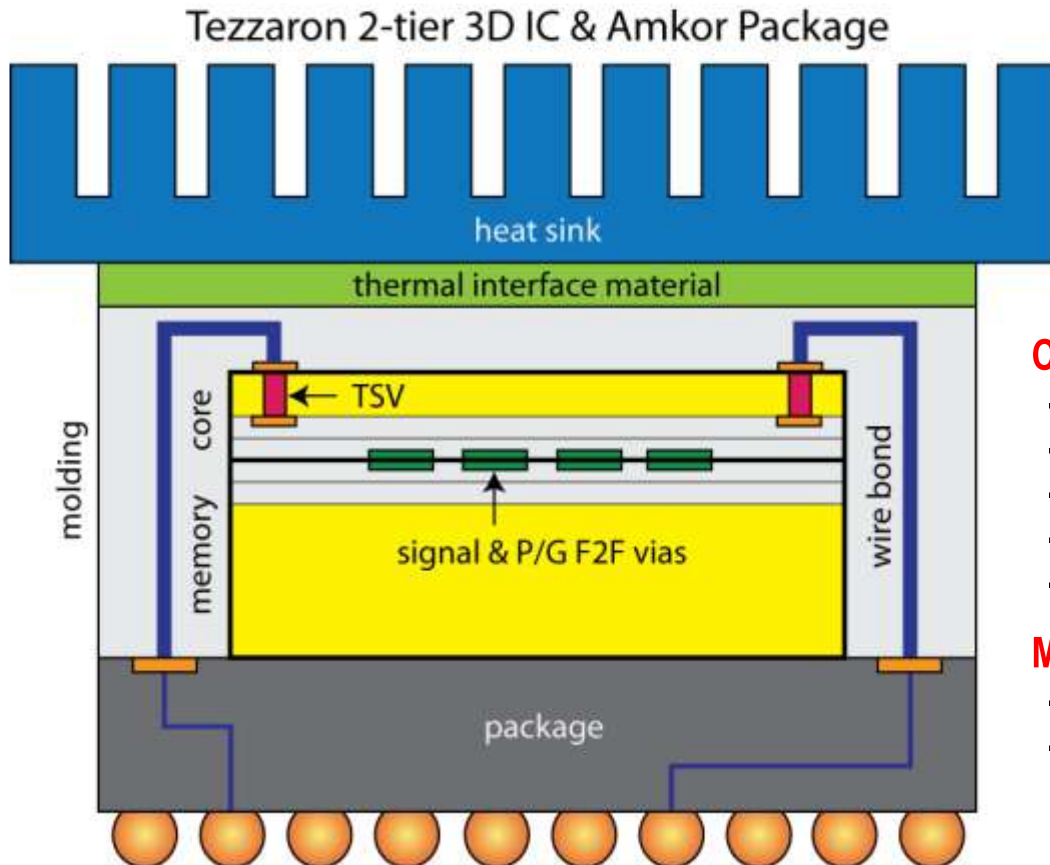
LIM, CAD Tool



LOH, Memory

Tezzaron 3D Process

- **DARPA MPW Run**
  - Chartered's 130nm technology + Artisan library/IP

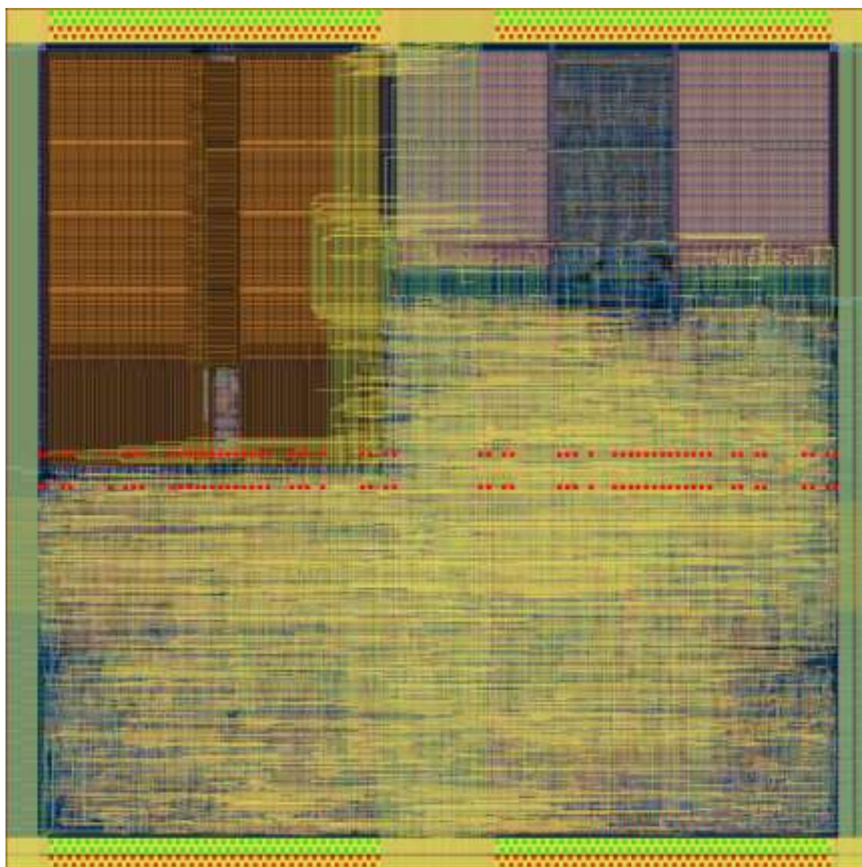


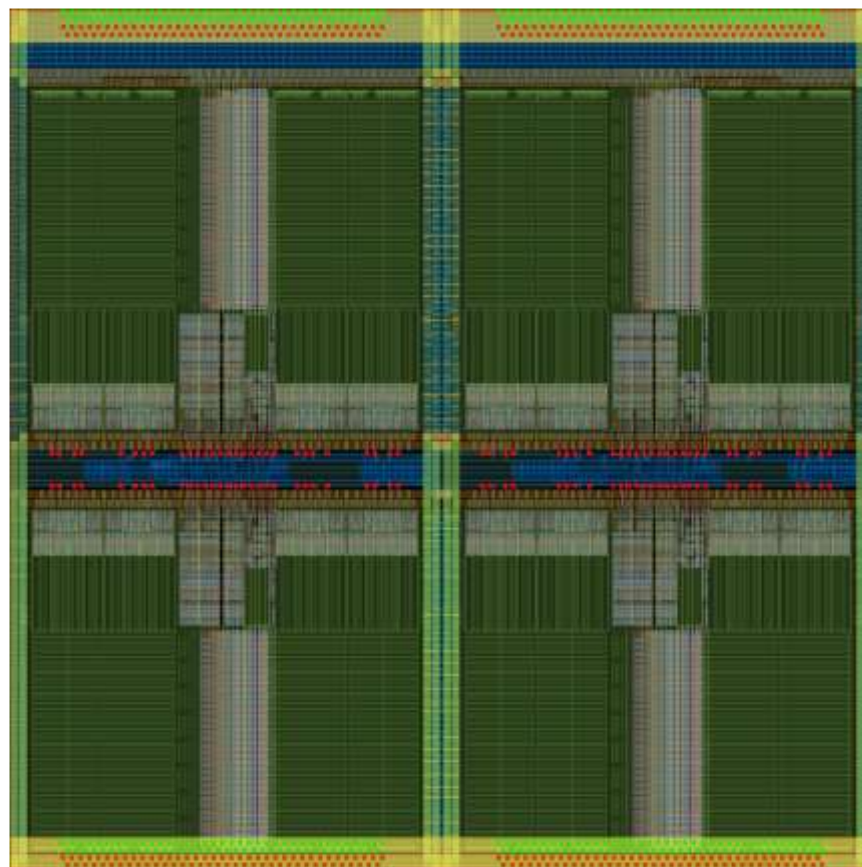
Core-tier
- thinned to 12um
- TSV height becomes 6um
- closer to heat sink
- talk to package via wire-bond
- requires dummy TSVs for density

Memory-tier
- thickness is 765um
- requires dummy TSVs for density

F2F Via Connections

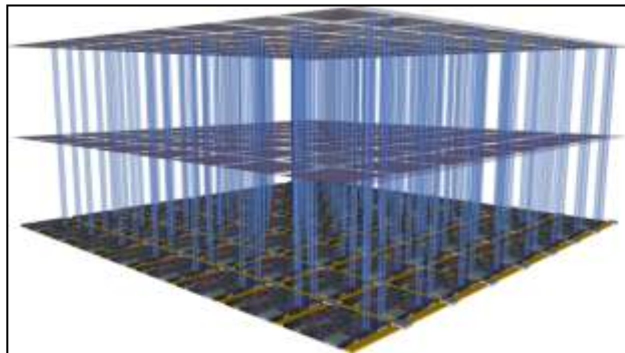
- **116 signal F2F-vias, 684 P/G F2F-vias per core**



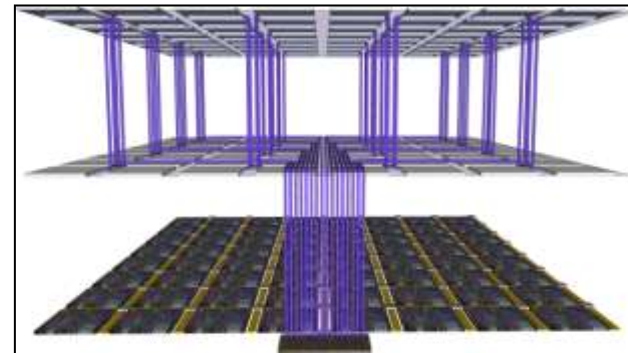


3D MAPS Test Configurations

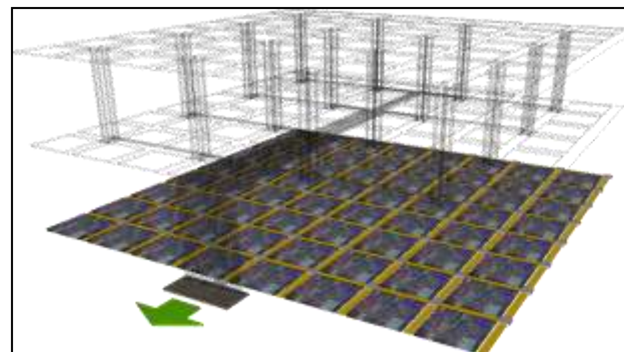
- Our design will support three configurations:
  - 3D local memories, 3D centralized memory, and 2D/off-chip
  - Directly measure performance and power benefits



this for now



next version

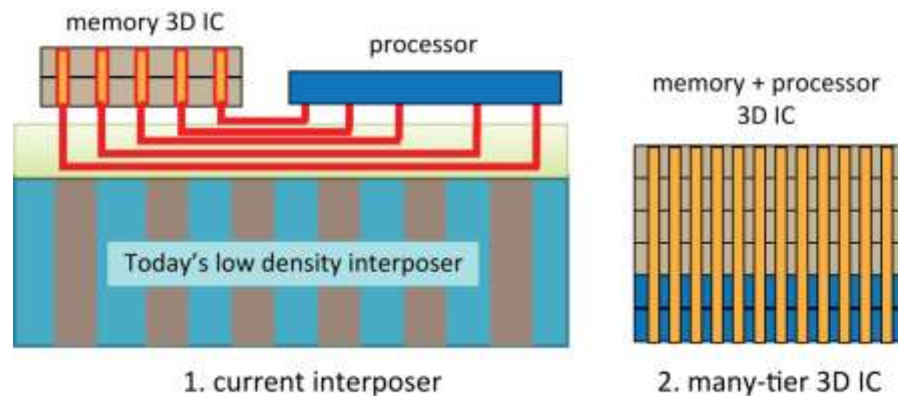


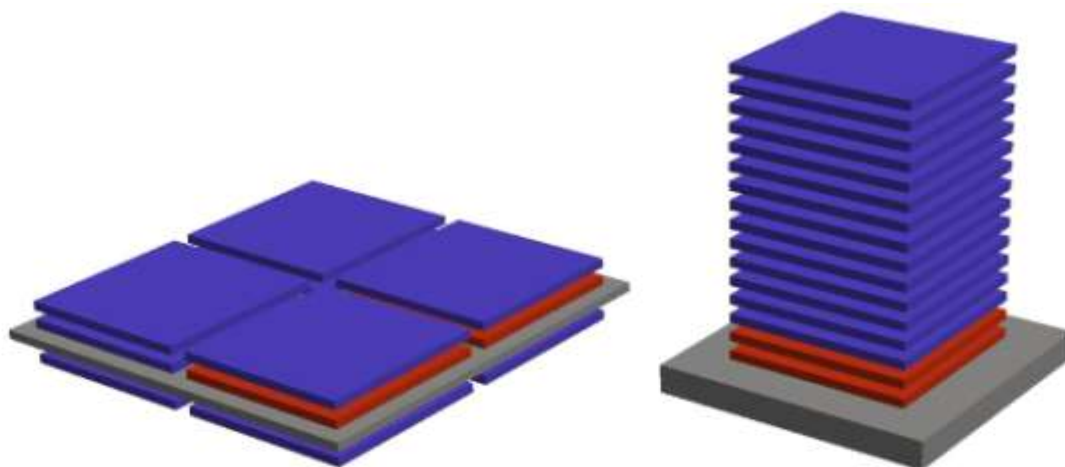
next version

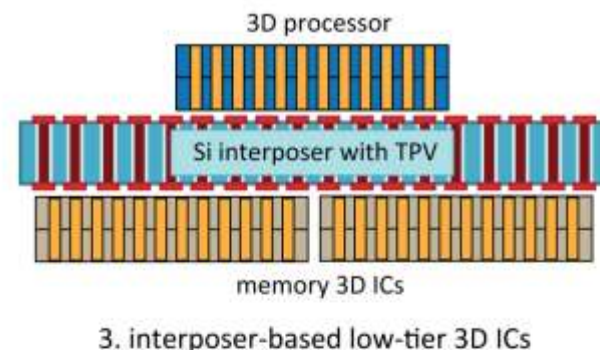
Many-Tier vs Interposer

With GT-PRC

- Interposers are improving
  - TSVs are also used, allowing double-sided integration
  - Could be better than many-tier 3D IC for 1000-core + memory integration





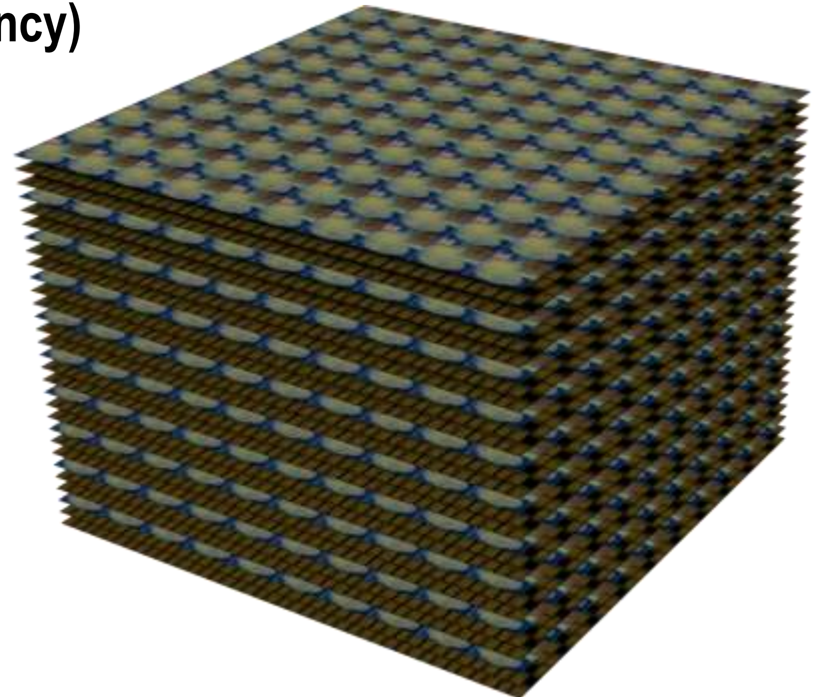


Interposer-based vs many-tier 3D integration of core+memory

How Many Tiers Can We Stack?

With GT-IFC

- Show how future 3D IC will look like
  - Will be MANY-TIERS of cores and memory
- Show what future 3D IC can do (compared w/ 2D and low-tier 3D)
  - Few orders of magnitude larger memory BW
  - Higher performance (IPC, clock frequency)
  - Low energy (joule/bit)
- Show what needs to be paid
  - Cooling, power, clock delivery
  - High fabrication cost and low yield
  - testing



Heterogeneous 3D ICs

SRC, IBM, Intel

- **Show how to integrate heterogeneous tiers into a single 3D IC**

HETEROGENEOUS 3D IC

CPU
GPU
memory
sub-Vth
package
die-to-heatsink coupling
die-to-die coupling
die-to-die coupling
die-to-die coupling
die-to-package coupling

MODELING

TSV stress and capacitance model

3D heat couple model

DESIGN

die-to-die communication

runtime power management

CASE STUDIES

sub-Vth+super-Vth 3D IC

CPU+GPU+memory 3D IC

Worked on ITRS 2009 Update

First Year of IC Production	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Automated handshake logic/circuit tools																
Synthesis and timing accounting for variability																
Circuit/layout enhancement accounting for variability																
Macro/chip leakage analysis																
Power management analysis and logic insertion SOI SoC tools																
Analog synthesis (circuit/layout)																
Non-static logic implementation																
Cost-driven implementation flow																
3D system design space exploration tools																
Native 3D power/thermal analyses, optimizations																

This legend indicates the time during which research, development, and qualification/pre-production should be taking place for the solution.

Research Required
Development Underway
Qualification / Pre-Production
Continuous Improvement

New in ITRS 2009

Figure DESN6 Logical/Circuit/Physical Design Potential Solutions

ITRS predicts that 3D/TSV will be important, and so will be 3D design technology.

ITRS 2009 edition includes two new items in the Requirements Table for the logic/circuit/physical (L/C/P) design technologies.

1. 3D system DSE tools: is 3D is better than 2D for a given design? If so, what level of granularity: core, block, or gate-level?

2. Native 3D analysis and optimization tools: power, thermal, performance, signal integrity, cost, manufacturability, etc

New metric: % of native 3D design technologies in the entire design flow

Cross-Cut 3D Research

- SRC runs 6 "focus centers"
  - All 6 centers believe 3D is important (and are working on it)
  - Need to collaborate: cross-center activities are important
  - First workshop in early 2010

Paul Franzon
NCSU

Sung Kyu Lim
GT, lead

Andrew Kahng
UCSD

Tomas Palacios
MIT

Xin Li
CMU

Phillip Wong
Stanford

Conclusions

- TSVs as layout objects
  - TSVs are large, intrusive
  - TSV count and location are important
  - TSVs cause manufacturability, reliability, testing issues
- 3D Research at GTCAD Lab
  - Physical Design Automation for 3D Circuits
  - DFM/DFR for TSV-based 3D ICs
  - Design, Fabrication, and Testing of 3D-MAPS
  - Interconnect Limit Study with Many-Tier 3D System
  - 3D Integration of Sub-Threshold Multi-core Co-processor
  - 3D IC Integration with Interposers
  - Heterogeneous 3D Integration
  - DFT for TSV-based 3D ICs