ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

System-Level Thermal Management of 3D MPSoCs with Active Cooling

Prof. David Atienza,
Embedded Systems Laboratory (ESL),
Ecole Polytechnique Fédérale de Lausanne (EPFL)

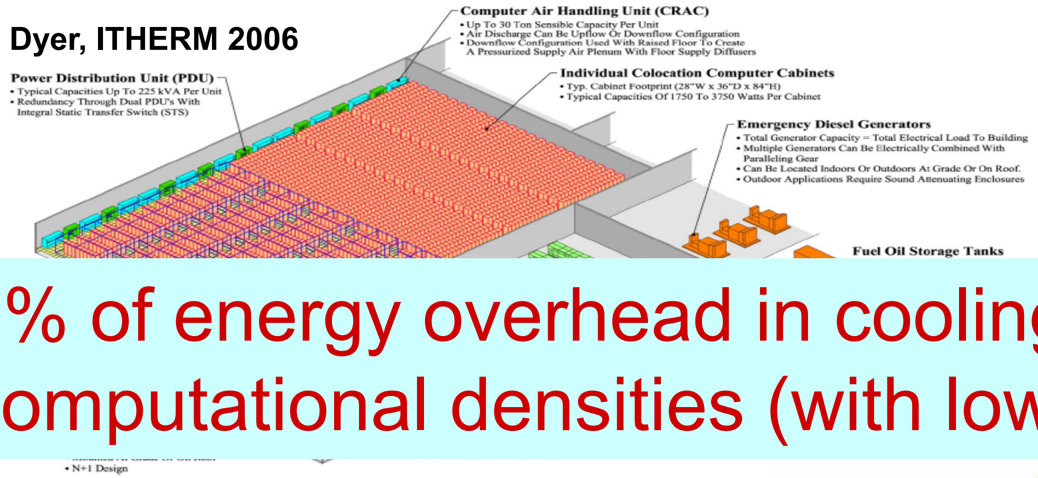
MPSoC'11

MPSoC'11, July 4th – 8th 2011 (Beaune, France)

Acknowledgements:
IBM Zürich, LSM and LTCM-EPFL, ECE-Boston University

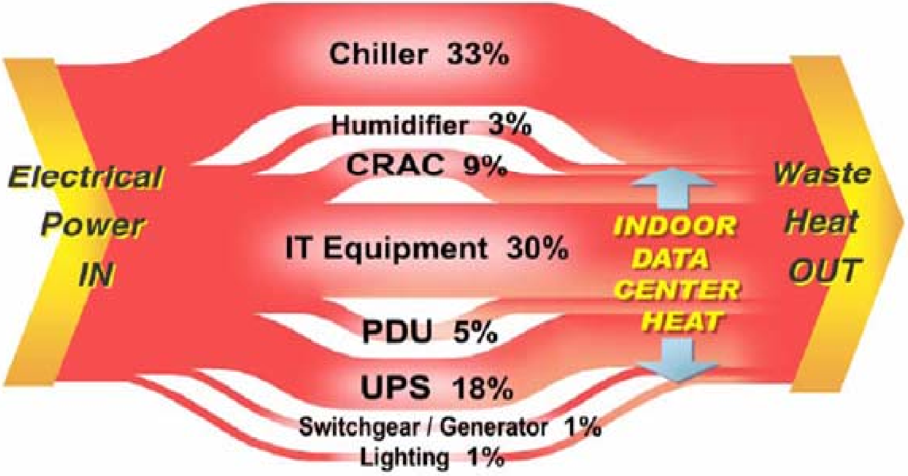
The Big Picture for Energy Efficiency: Large-Scale Computing in Datacenters

- Area is expensive, we try to get denser infrastructures
  - New containers: 2500 servers each, >10x density


Dyer, ITHERM 2006

45% of energy overhead in cooling, how to get higher computational densities (with lower cooling costs)?

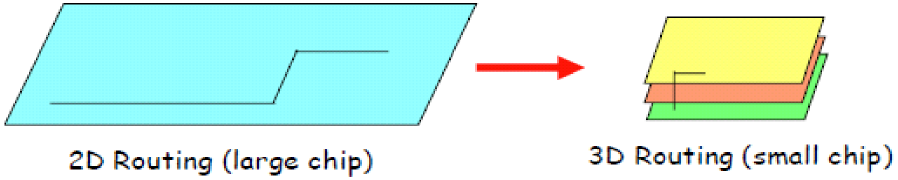
- **Air-cooled datacenters** are **very inefficient**
  - Cooling needs as much energy as IT… and thrown-away
- For a 10MW datacenter **~US$ 4M wasted** per year



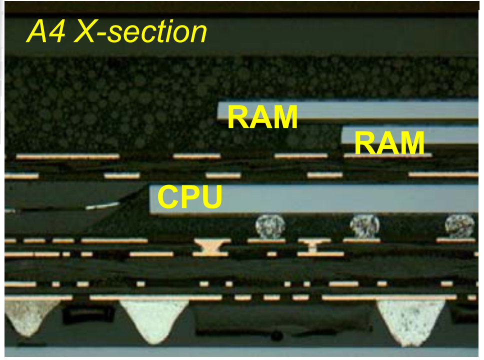
Datacenter energy overhead, ASHRAE

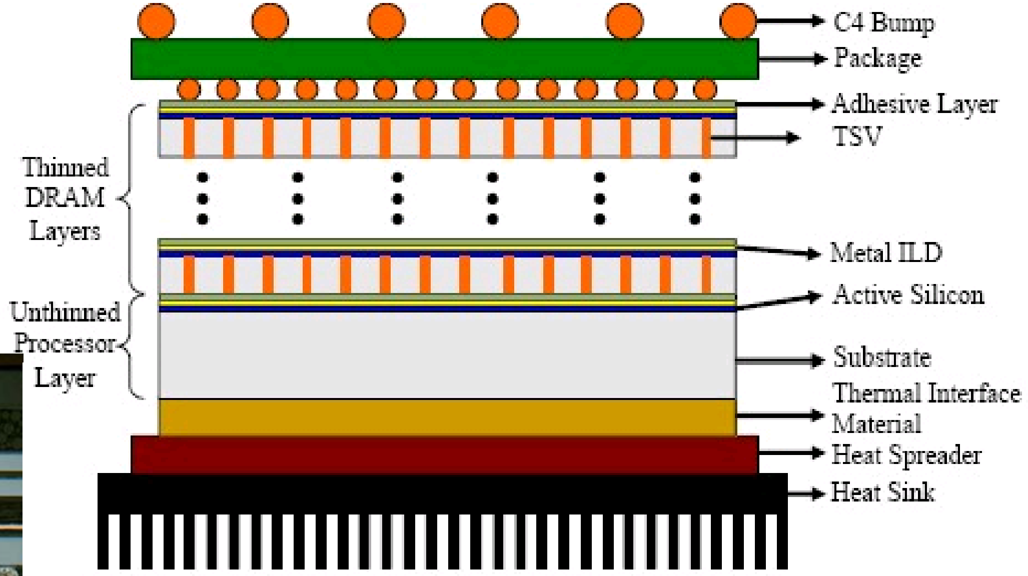
Advantages of 3D vs. 2D Multi-Processor System-on-Chip (MPSoCs) ICs

- Promises
  - Reduce average length of on-chip global wires
    - Increase number of devices reachable in given time budget



  - Greatly facilitate massive storage integration (i.e., logic-RAM stacks)

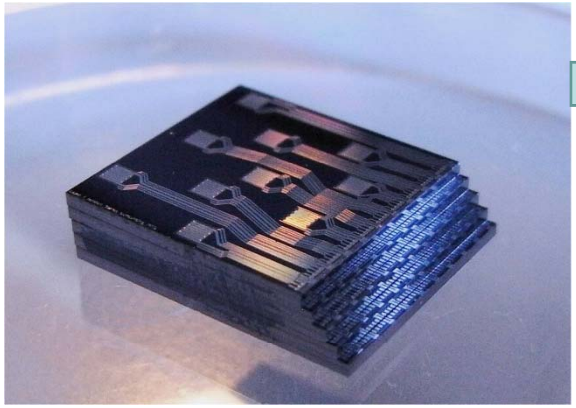




Three-Dimensional Multi-Processor System-on-Chip (MPSoC) IC Concept

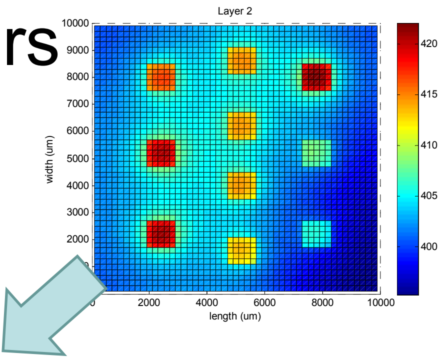
Run-Time Heat Spreading in 3D MPSoCs: More Complex Cooling Needs!

- 5-tier 3D stack: 10 heat sources and sensors

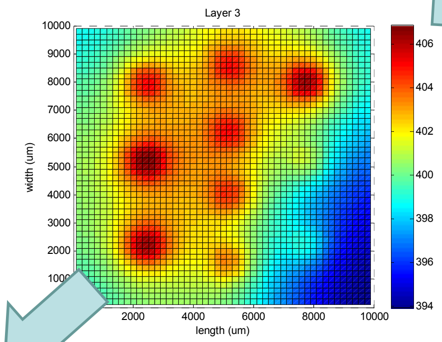


[Garcia and Atienza, *Microelectornics Journal 2010*]

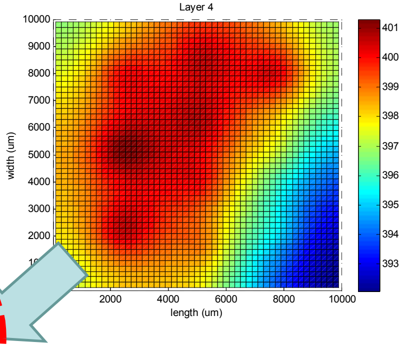
Inject between 4W – 1.5W



2nd Tier

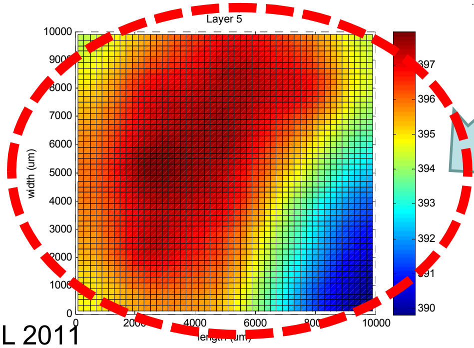


3rd Tier



4th Tier

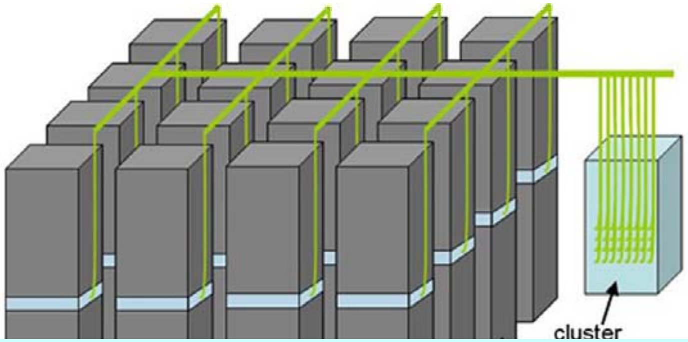
5th Tier



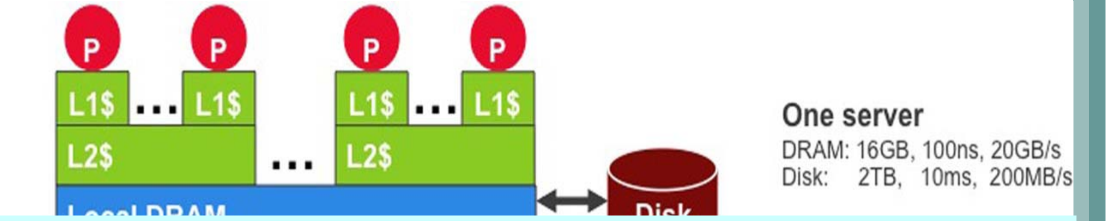
Large and non-uniform heat propagation!
(up to 130º C on top tier)

Energy-Centric Design for Datacenters: System-Level View

- Multi-scale energy management solution needed

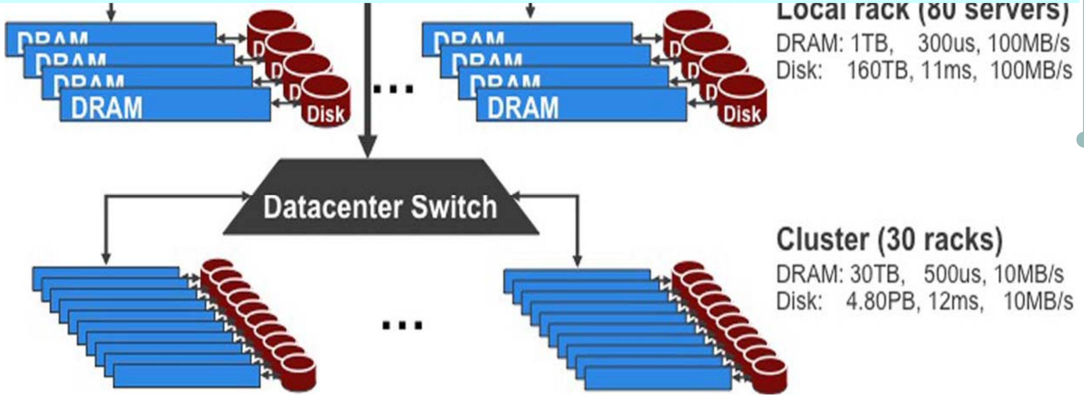


- Architecture SW: middleware, power management, network design, etc.



Energy-proportional datacenters: Integrated layers of software-hardware to maximize Mflops/Watt

- Technology HW: cooling, processing and storage systems, etc.

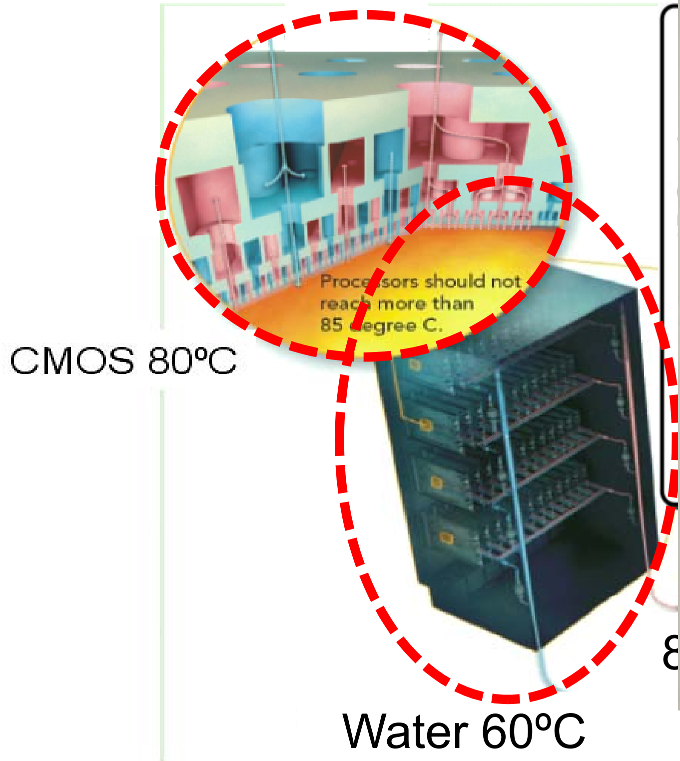


Barroso & Hölzle, 2009

Zero-Emission Datacenter: Liquid Cooling Technology and Predictive Energy Management

- Datacenters are “intelligent” heaters
  - 30-40% of carbon footprint in Europe using district heating networks
- Direct re-use of heat output
  - 3D MPSoC architectures

Aquasar datacenter server: 80% payback of electricity costs

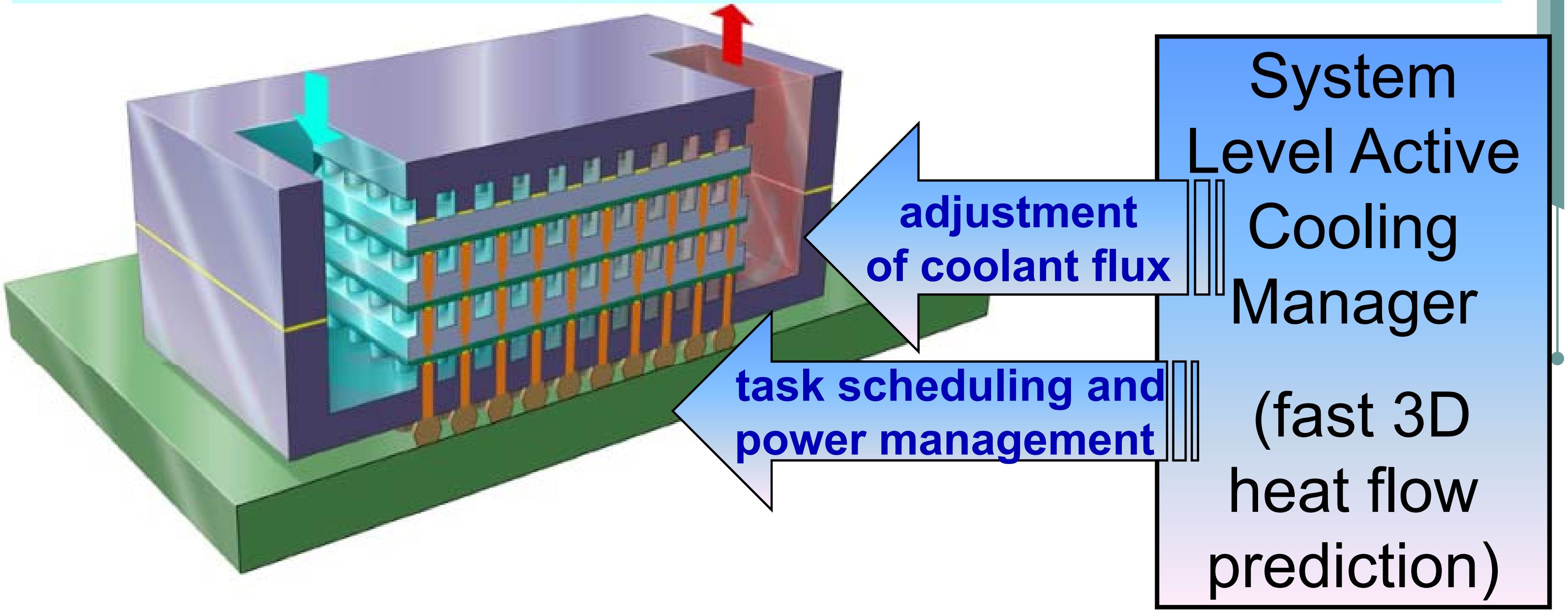


Courtesy: IBM Zürich

NanoTera CMOSAIC Project: Design of 3D MPSoCs with Advanced Cooling

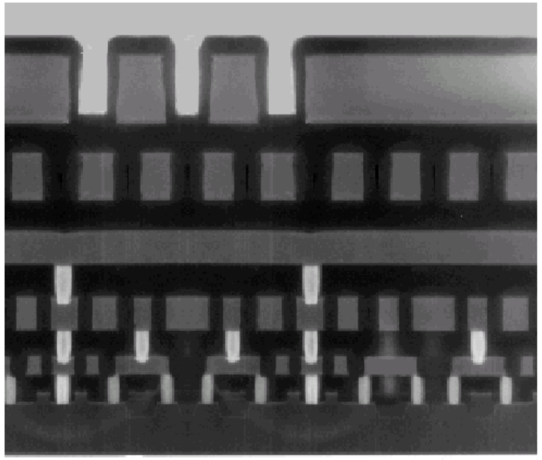
- 3D systems require novel electro-thermal co-design
  - Academic partners: EPFL and ETHZ
  - Industrial: IBM Zürich and T.J. Watson

3D MPSoC datacenter chip: microchannels etched on back side to circulate (controlled) liquid coolant



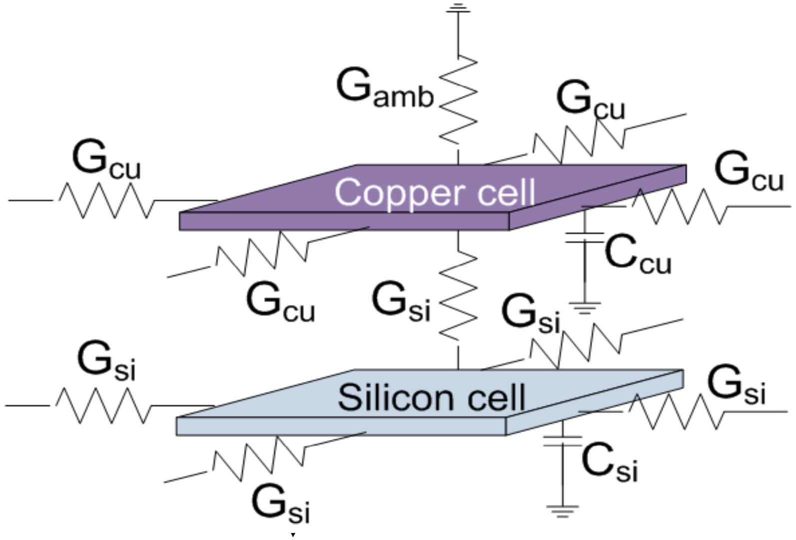
Creating a Fast Thermal Model: Compact RC-Based Stack Model with TSVs

- Chip-Level thermal model

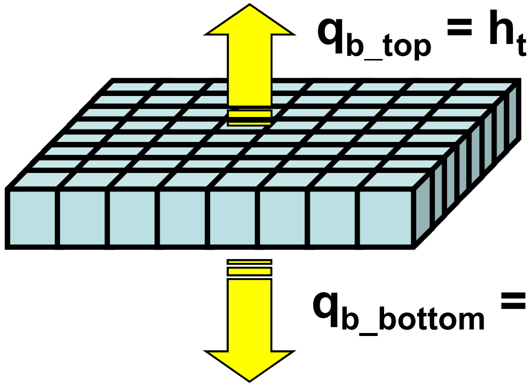


RC Network of

Si/metal layer cells



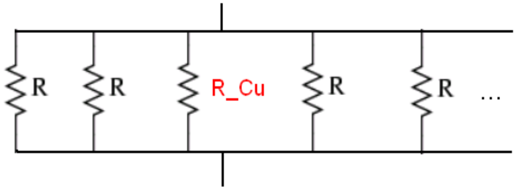
- Convective boundary conditions between layers **for each tier separately**

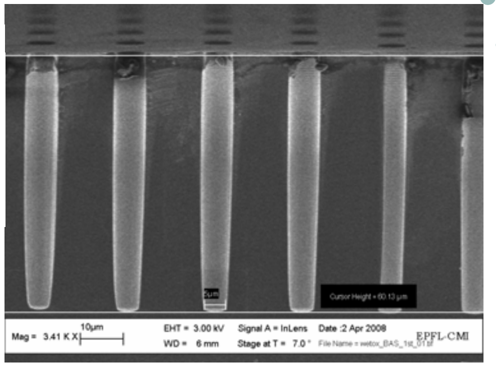


$$q_{b\_top} = h_{top}A(T_a-T_{top})$$

$$q_{b\_bottom} = h_{bottom}A(T_a-T_{bottom})$$

- TSVs change resistivity of interlayer material





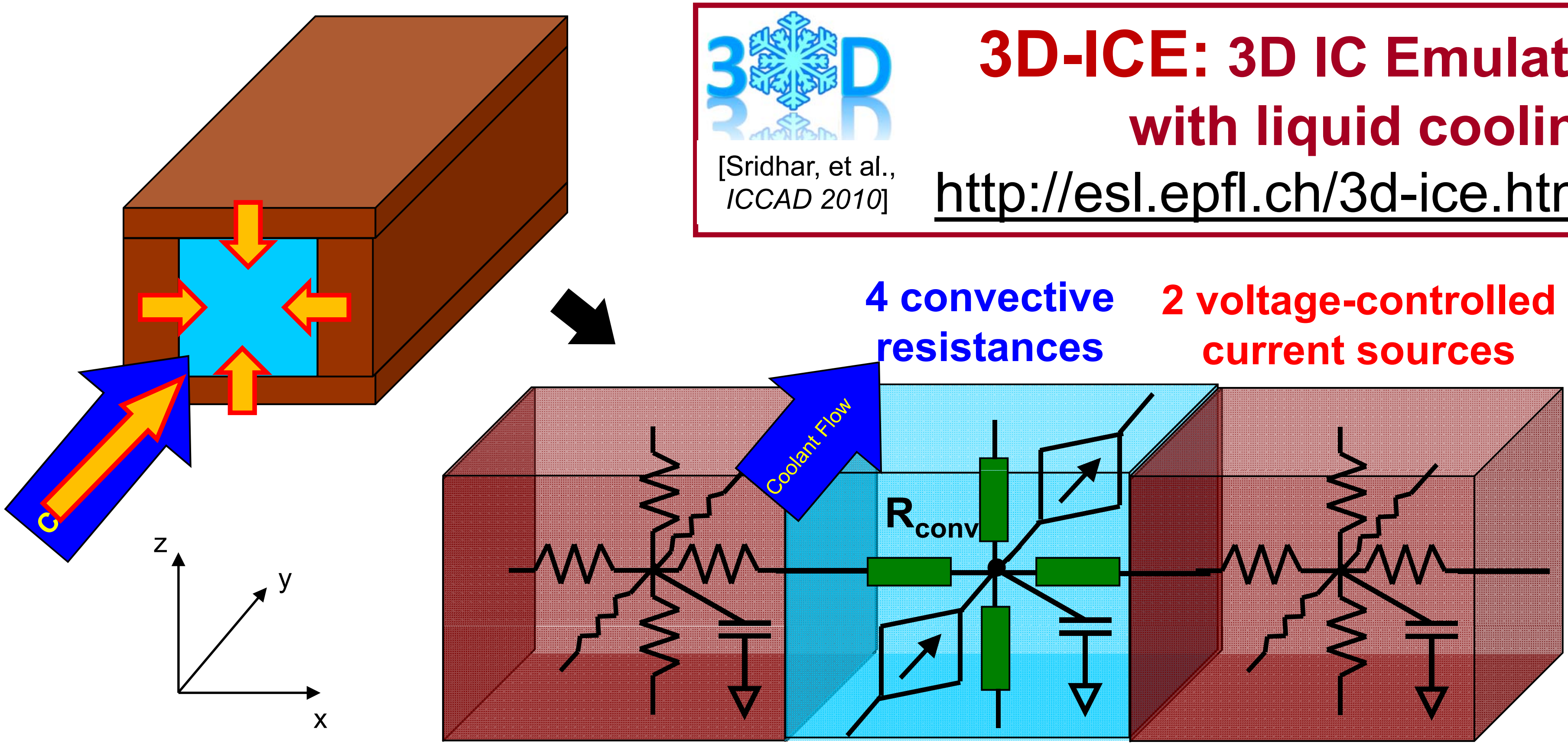
Modeling Liquid Cooling as RC-Network in 3D MPSoC stacks

- Local junction temperature modeled as 4-resistor based compact transient thermal model (4RM-based CTTM)
  - Rtot = Rcond + Rconv + Rheat

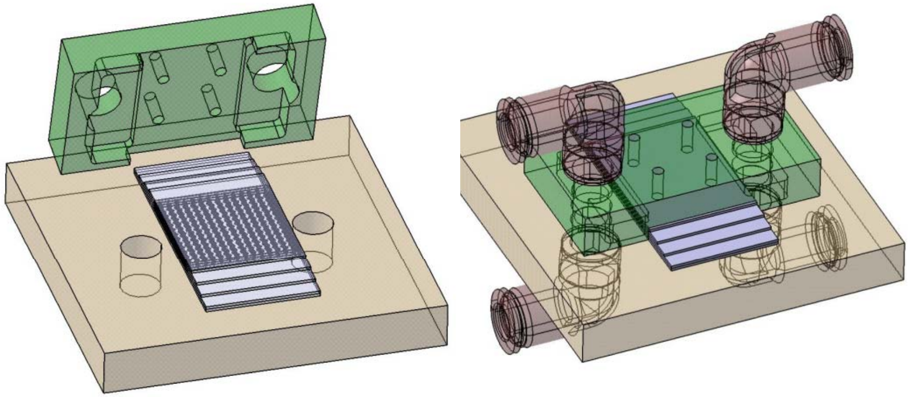
3D-ICE: 3D IC Emulator with liquid cooling

http://esl.epfl.ch/3d-ice.html

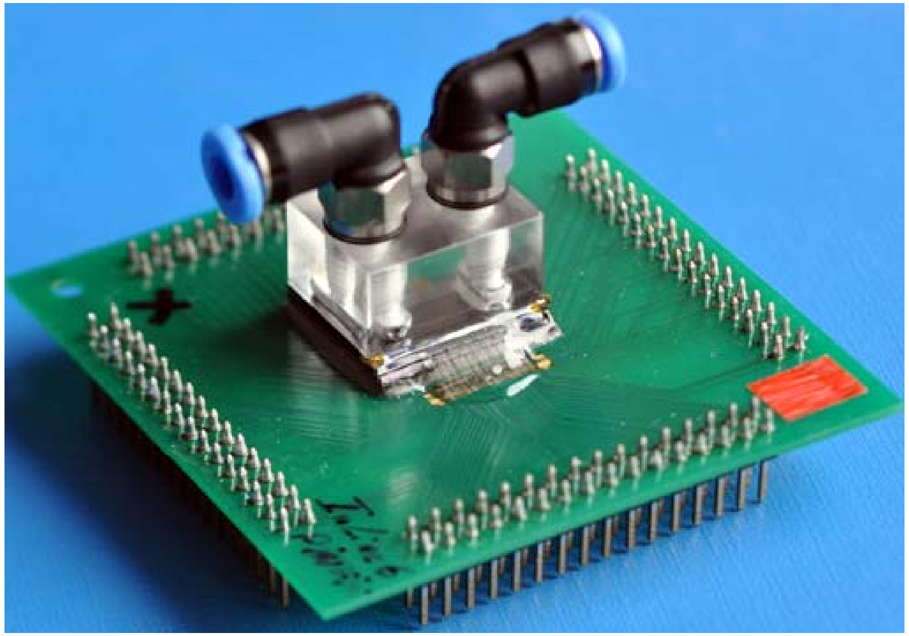
[Sridhar, et al., *ICCAD 2010*]

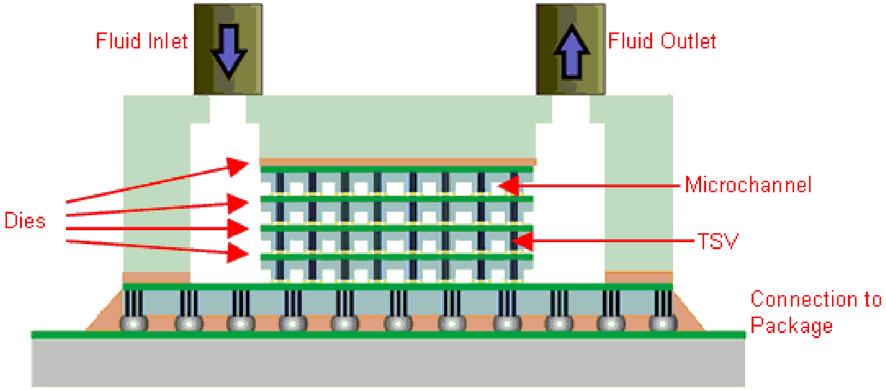


Manufacturing of 5-Tier 3D Chips with Liquid Channels in Multiple Tiers

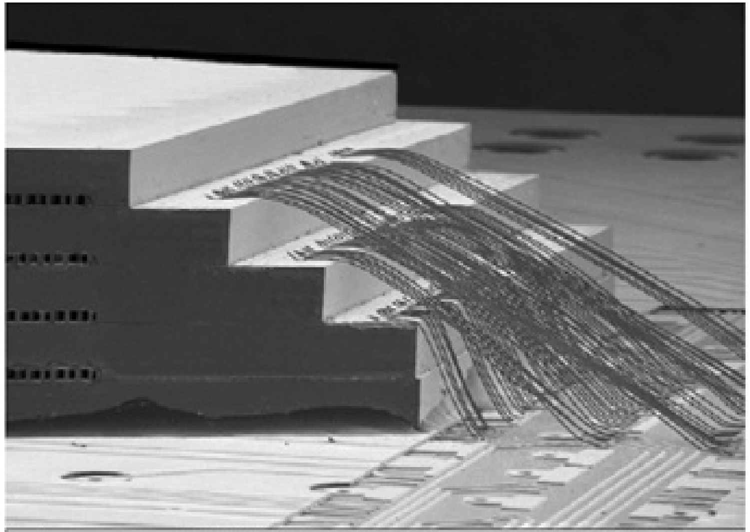


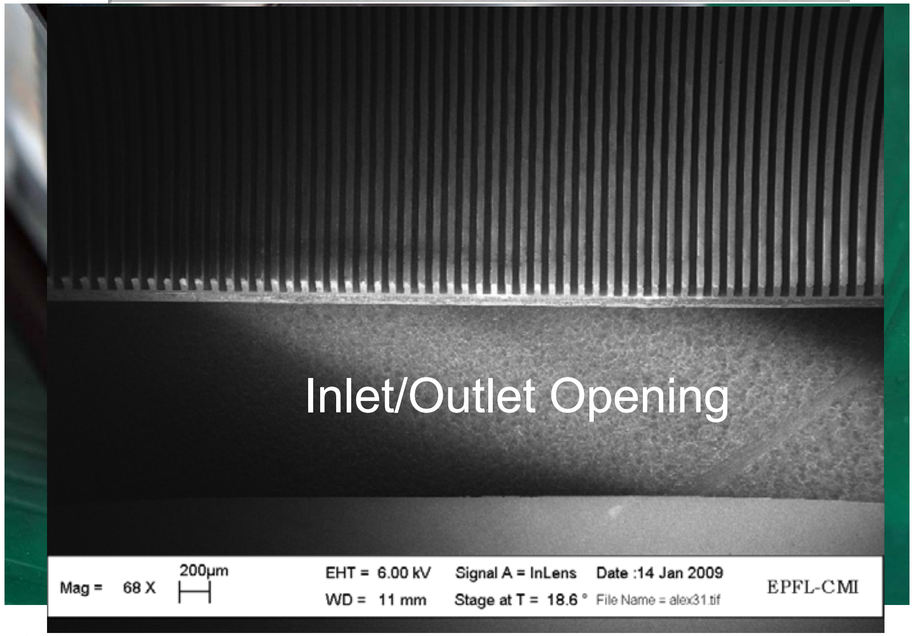
Adding multi-tier liquid cooling in-/out-lets



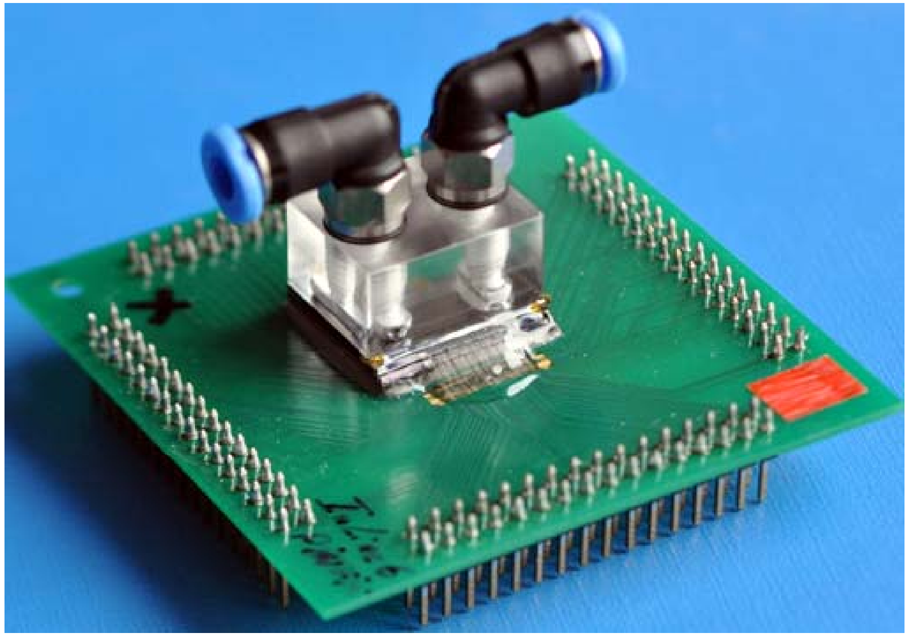


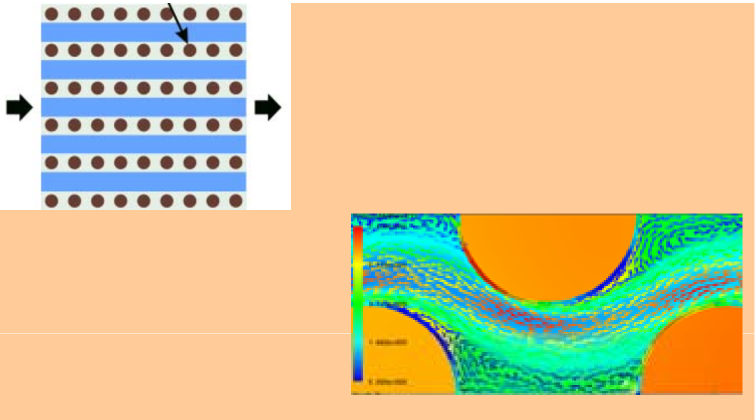
Manufacturing of 5-Tier 3D Chips with Liquid Channels in Multiple Tiers





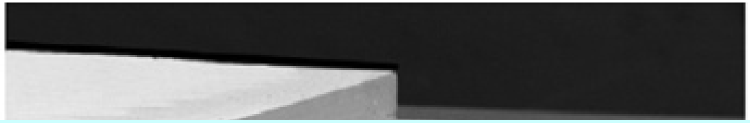
Inlet/Outlet Opening

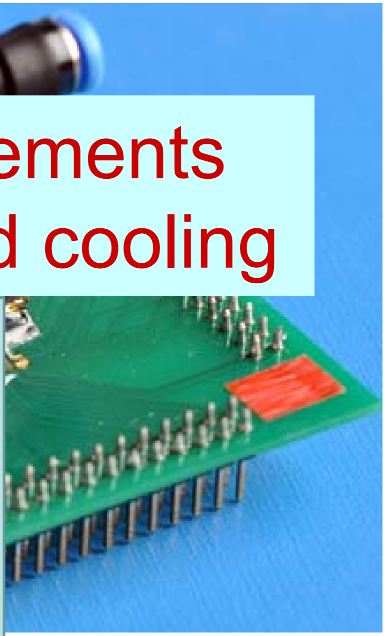




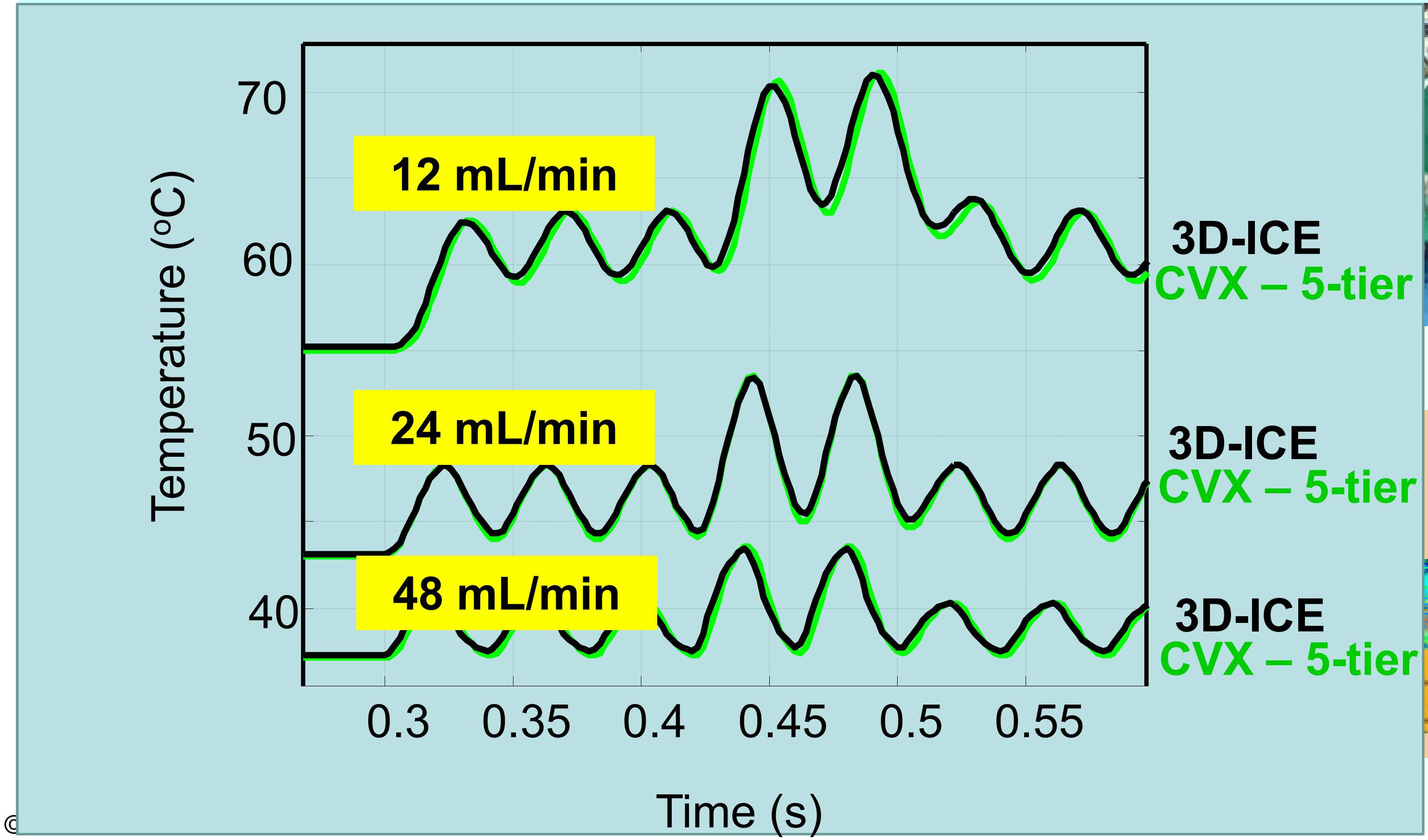
© ESL/EPFL 2011

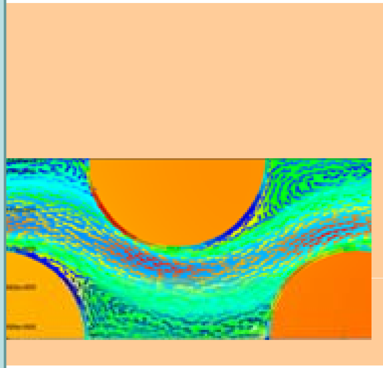
Manufacturing of 5-Tier 3D Chips with Liquid Channels in Multiple Tiers





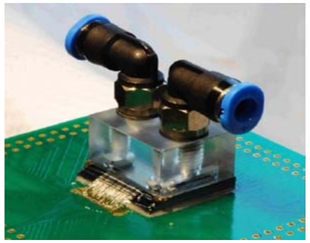
Maximum error of 3.5% between measurements and RC-based 3D thermal model with liquid cooling





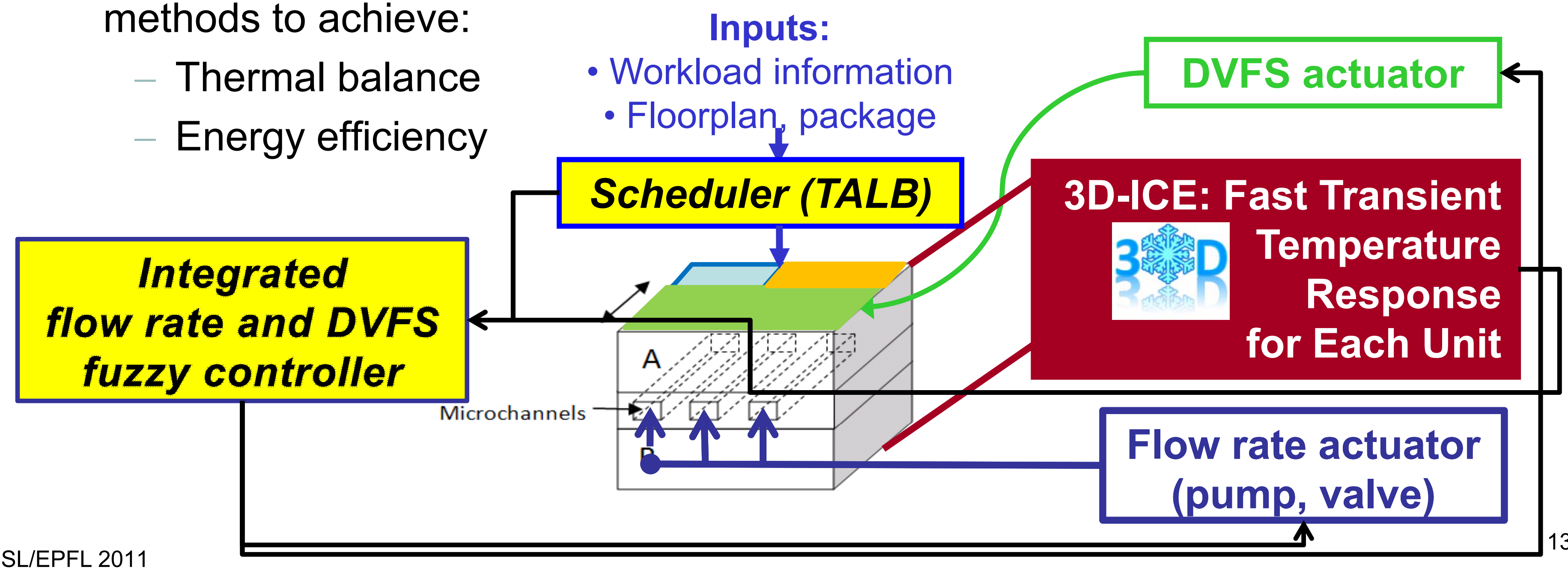
Active-Adapt3D: Active cooling management for 3D MPSoCs

- 3D MPSoC temperature control at system-level:
  - **Electrical based**: task scheduling, and DVFS (µsec or few ms)
  - **Mechanical based**: run-time varying flow rate (hundreds of ms)



- *Fuzzy logic-based controller and thermal-aware scheduler*
  1. **Design-time analysis**: extraction of set of thermal management rules
  2. **Run-time thermal management**: utilization of rules in scheduler and subsequently fuzzy logic controller using both mechanical and electrical methods to achieve:
     - Thermal balance
     - Energy efficiency

[Coskun , Atienza, et al., *MICRO 2011*]



Inputs:
- Workload information
- Floorplan, package

Scheduler (TALB)

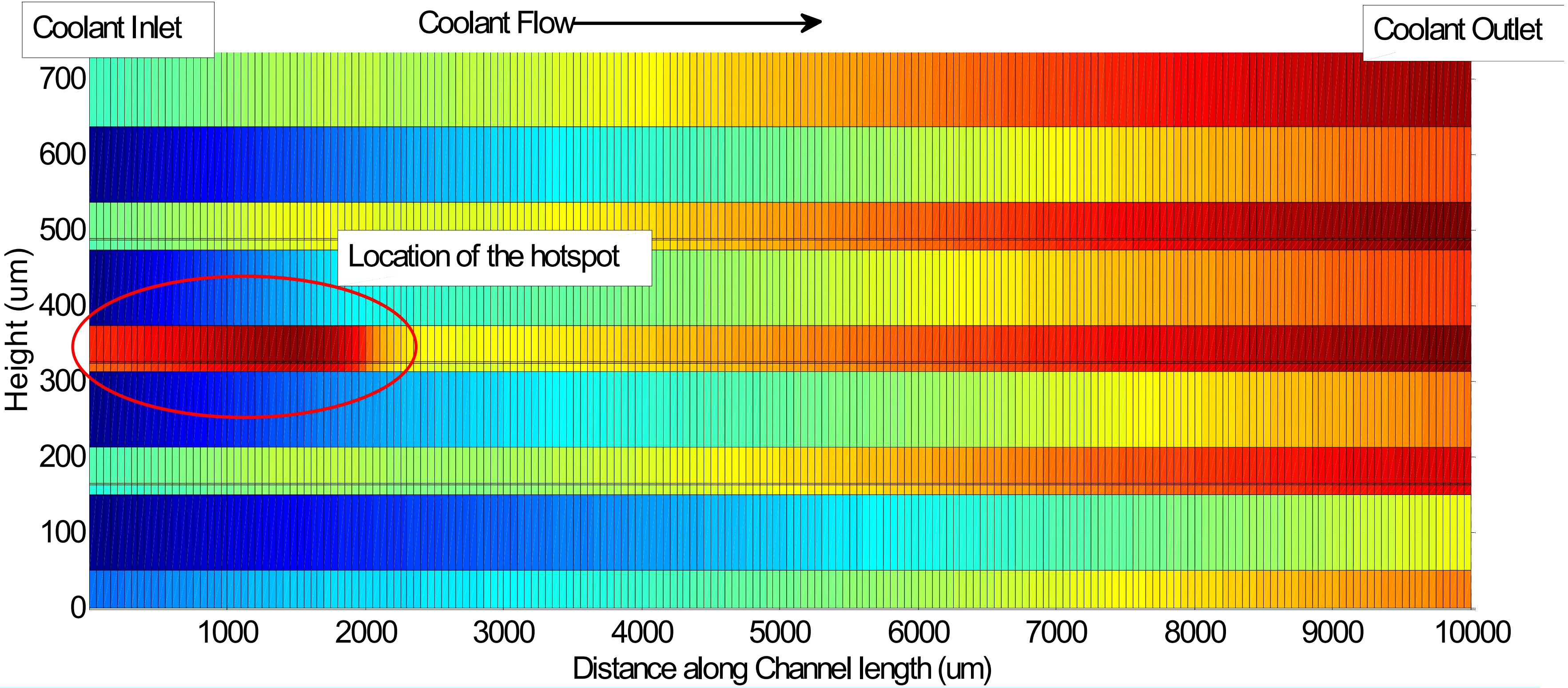
Integrated flow rate and DVFS fuzzy controller

DVFS actuator

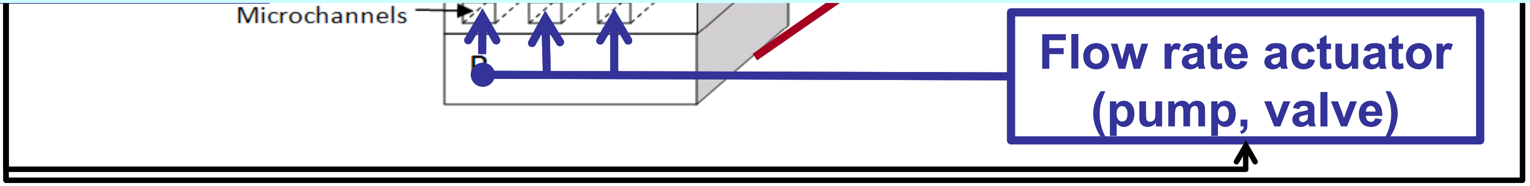
3D-ICE: Fast Transient Temperature Response for Each Unit

Flow rate actuator (pump, valve)

Active-Adapt3D: Active cooling management for 3D MPSoCs

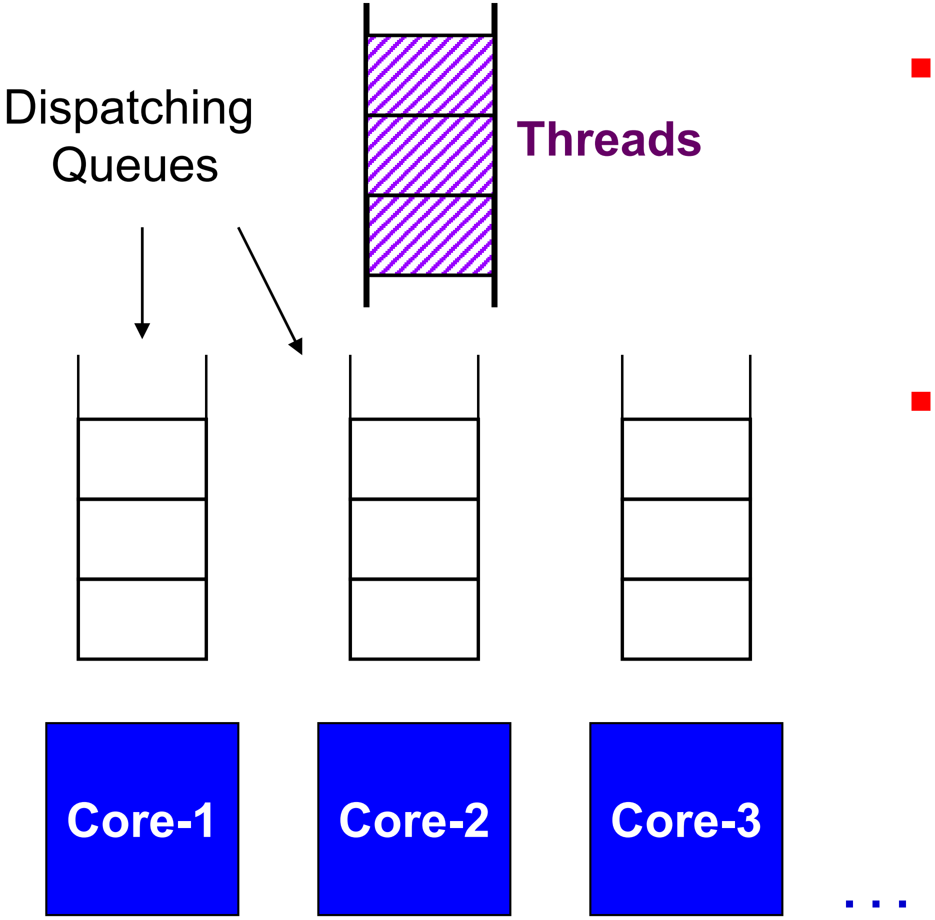


New insights about suitable thermal-aware scheduling and task assignment for 3D MPSoCs!



Flow rate actuator (pump, valve)

Temperature-Aware Load Balancing (TALB) Scheduler for 3D MPSoCs

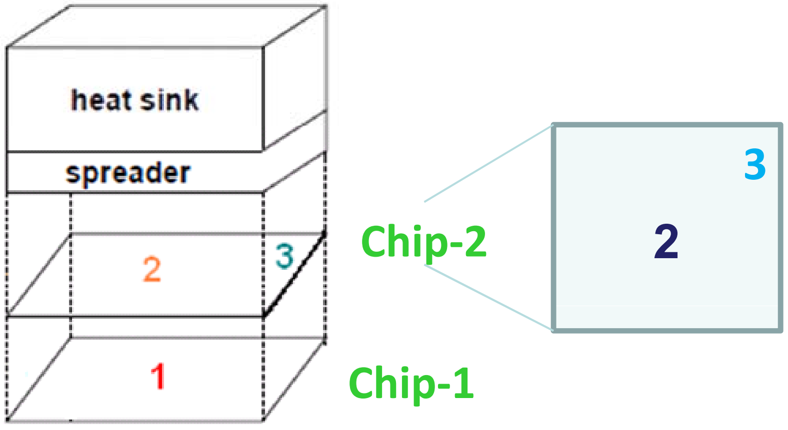


For cores at locations 1, 2 and 3:

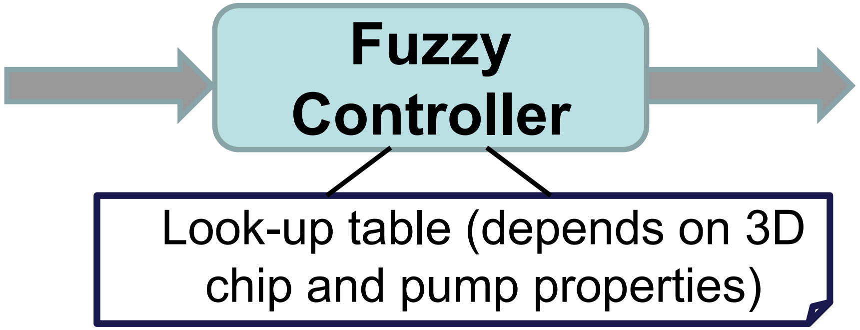
$$\alpha^1 > \alpha^2 > \alpha^3$$

- Load balancing (Default):
  - Balances queue-length, but does not consider core locations
- TALB: [Coskun , et al., *DATE 2010*]
  - Thermal index (α) per core

$$l^i_{\text{weighted}} = l^i_{\text{queue}} \cdot \alpha^i_{\text{thermal}}(T(k))$$



Integrated Flow Rate and DVFS Fuzzy Controller for 3D MPSoCs



- 3D-ICE: predicted max. core temperat.
- TALB: Workload
- Pump power

Fuzzy Controller

Adjusted Flow Rate and DVFS

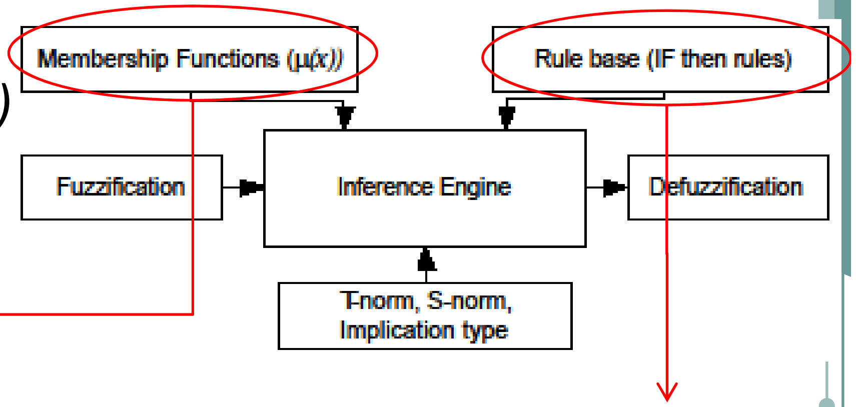
Look-up table (depends on 3D chip and pump properties)

[Sabry, et al., *ICCAD 2010*]

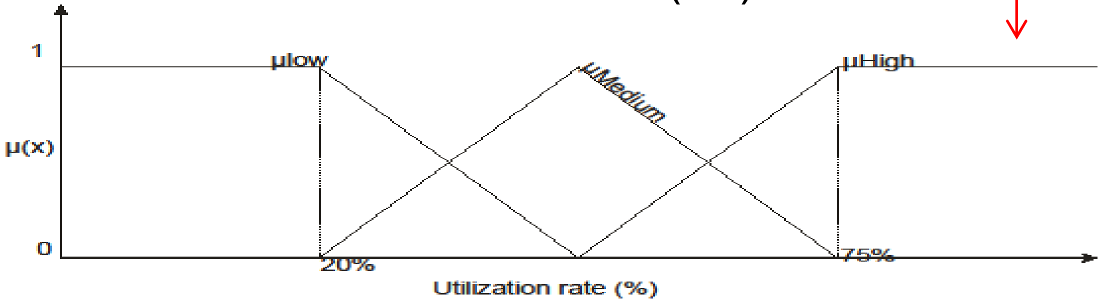
- Takagi-Sugeno fuzzy controller

IF $x_1 = a_1$ **AND** $x_2 = a_2$ … **THEN** $y_1=f_1(x_1,x_2,..)$

  - Extract cooling rules at design time
  - Stable (circle criteria, BIBO stability,…)



- Rules for multi-outputs at run-time
  - Electrical: DVFS (VF) per core
  - Mechanical: Flow rate (FL)



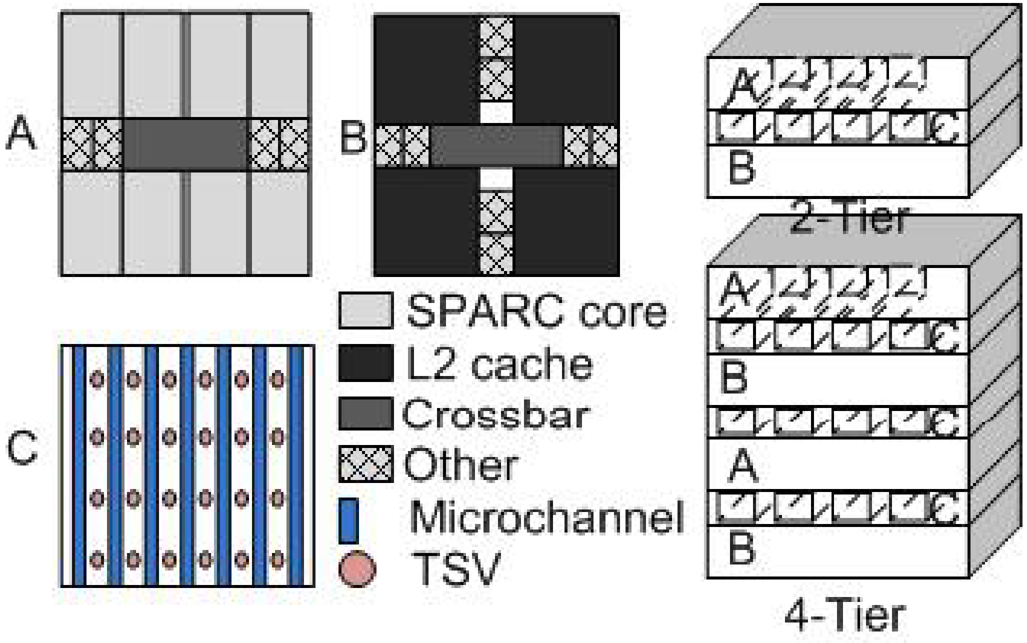
- Cores next to the inlet: no thermal reduction needed

IF D is L **THEN** VF is H **AND** FL is L

4-tier chip: ~150 rules

Experiments Active Thermal Management 3D MPSoCs with Microchannels

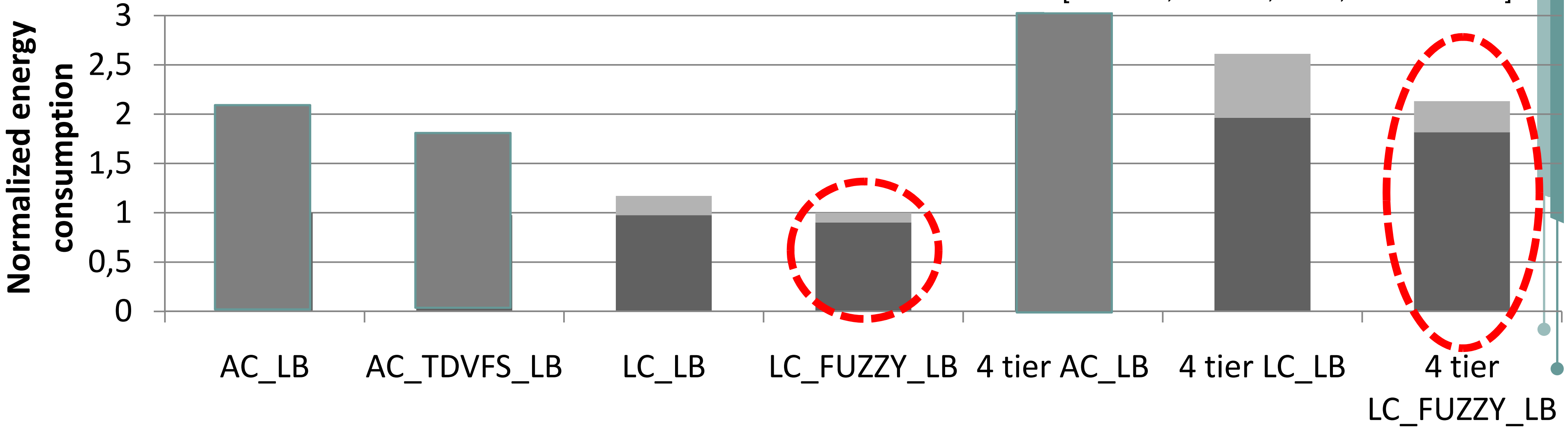
- Target 3D systems based on 3D ICs with Sparc-Power cores
  - Power values and workloads from real traces measured in Sun platforms (database queries, web services, etc.)
- Cores and caches in separate layers
  - 3D crossbar as interconnect
- Channels:
  - Width 100μm and height 50μm
  - Three flow rate settings, default at 32ml/min



Run-time thermal Management for 3D Chips: thermal evaluation

- For hot spot threshold 85°C, thermal violations: **0%**
- Energy reduction:
  - **70%** average coolant energy (**max. savings: 77%**)
  - **52%** average total system energy (**max. savings: 85%**)

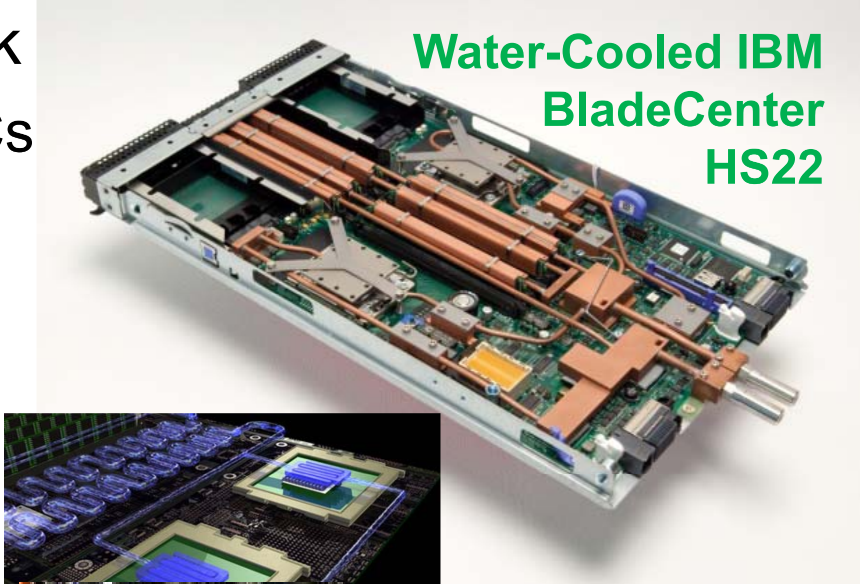
[Coskun , Atienza, et al., *MICRO 2011*]



Promising figures for thermal control in 3D MPSoCs, thermal gradients of less than five degrees/tier

Conclusions: Aquasar 2010
First Chip-Level Liquid Cooled Server

- 3D MPSoCs: Interdisciplinary work
  - Fast RC thermal models for 3D ICs with inter-tier variable liquid fluxes (**error of less than 4%**)
  - Layout combining **electrical and mechanical** constraints
- Next generation of thermal-aware proactive controllers (task control, flow rate and DVFS)
  - Holistic control reduces significantly the energy cost for the whole system (**80% power savings**)
  - "Green" datacenters: energy efficient
    - Roadrunner: 445 Mflops/Watt
    - **Aquasar: 2250 MFlops/Watt**


Water-Cooled IBM BladeCenter HS22


Back side


Water conditioning

Courtesy: IBM Zürich

Key References and Bibliography

- 3D Thermal modeling and FPGA-based emulation
    - "**3D-ICE: Compact transient thermal model for 3D ICs with liquid cooling via enhanced heat transfer cavity geometries**", A. Sridhar, et al. *Proc. of ICCAD 2010*, USA, November 2010 (http://esl.epfl.ch/3d-ice.html).
    - **"Emulation-based transient thermal modeling of 2D/3D systems-on-chip with active cooling"**, Pablo G. Del Valle, David Atienza, *Elsevier Microelectronics Journal*, December 2010.
- Thermal management for 3D MPSoCs
    - **"3D Stacked Systems with Active Cooling: The Road to Single-Chip High-Performance Computing**", Ayse K. Coskun, David Atienza, Mohamed M. Sabry, J. Meng, *IEEE MIC*RO, June/July 2011.
    - **"Fuzzy Control for Enforcing Energy Efficiency in High-Performance 3D Systems"**, M. Sabry, Ayse K. Coskun, David Atienza, *Proc. of ICCAD 2010*, USA, November 2010.
    - **"Energy-Efficient Variable-Flow Liquid Cooling in 3D Stacked Architectures"**, Ayse K. Coskun, David Atienza, et al., *Proc. of DATE 2010*, Germany, March 2010.
    - "**Modeling and Dynamic Management of 3D Multicore Systems with Liquid Cooling**", Ayse K. Coskun, et al., *Proc. of VLSI-SoC 2009*, Brazil, October 2009. **(Best Paper Award)**

Thank You

QUESTIONS ?

Nano-Tera.ch Swiss Engineering Programme

European Commission

Swiss National Science Foundation