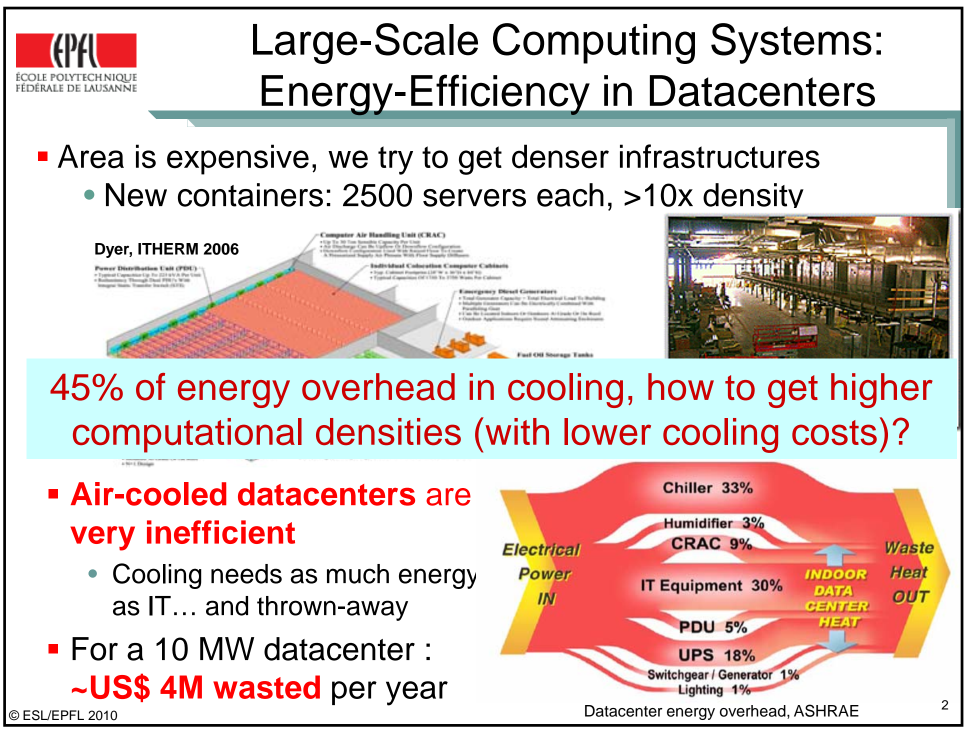
Green Considerations for System-Level Computing Architectures Design

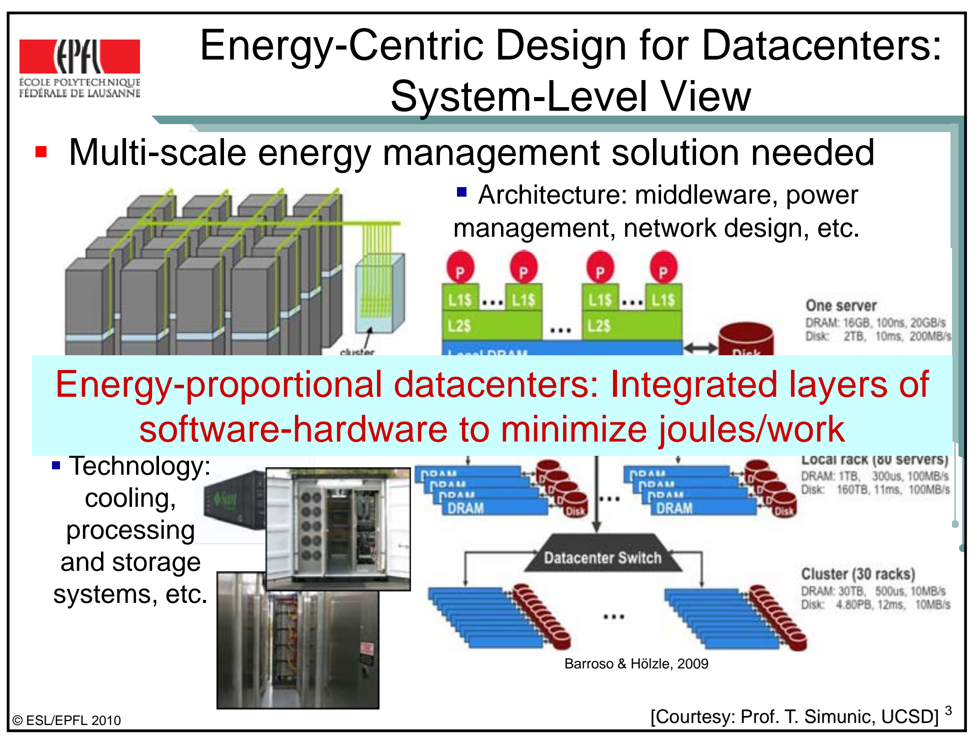
Prof. David Atienza,
Embedded Systems Laboratory (ESL),
Ecole Polytechnique Fédérale de Lausanne (EPFL)

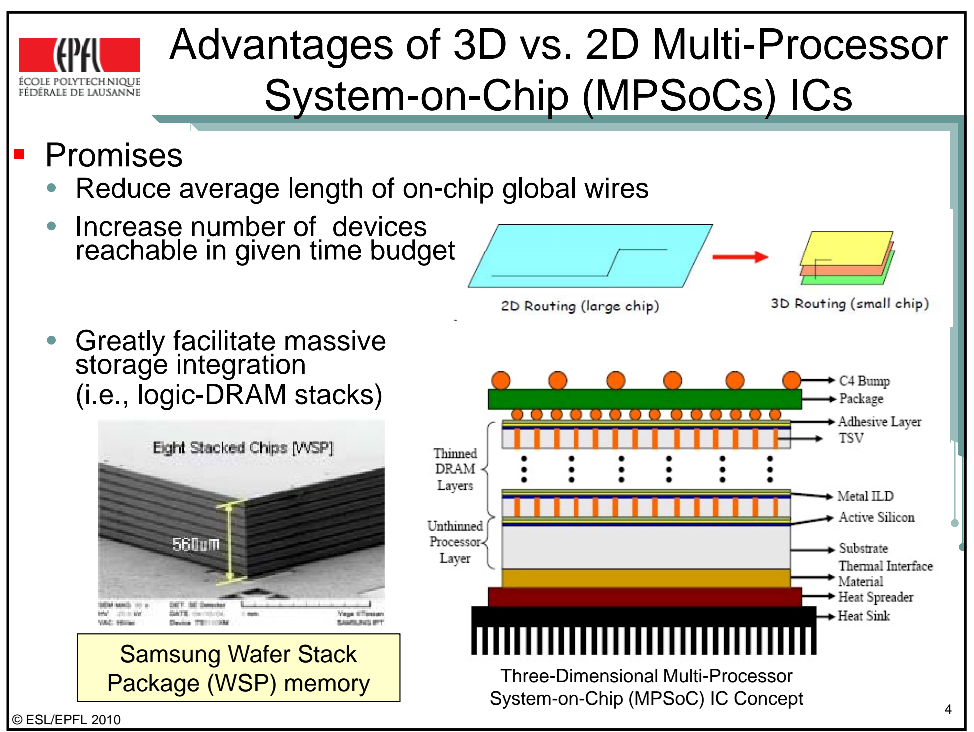
Acknowledgements:
LSM-EPFL, Advanced Packaging Group-IBM Zürich,
ECE-Boston University

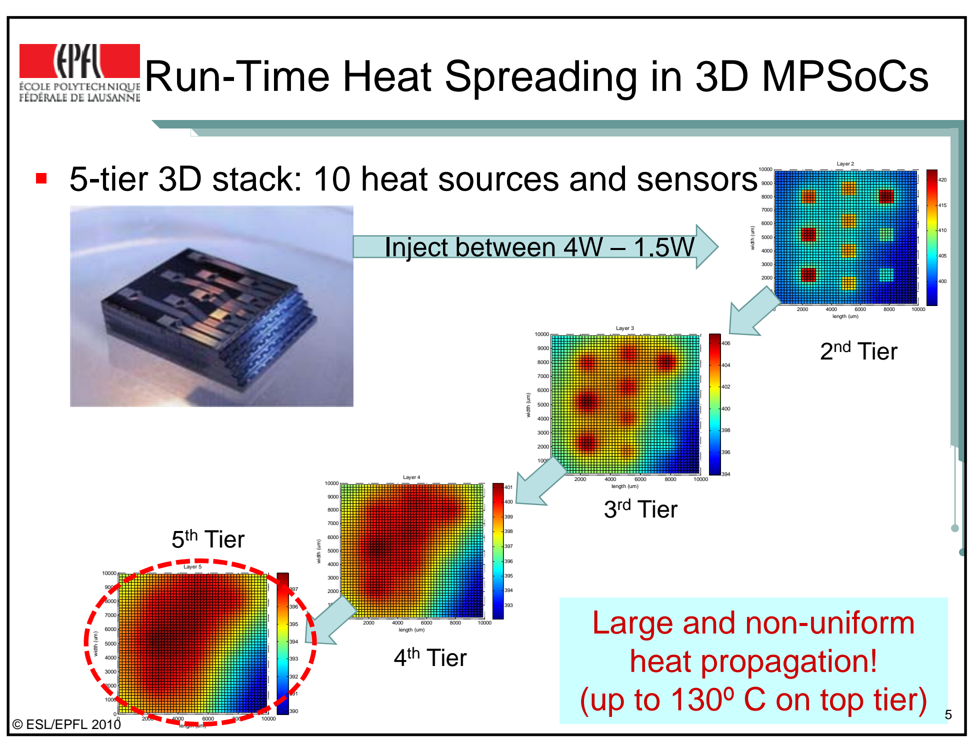
Smart Energy Day (EPFL, Lausanne, December 13th, 2010)

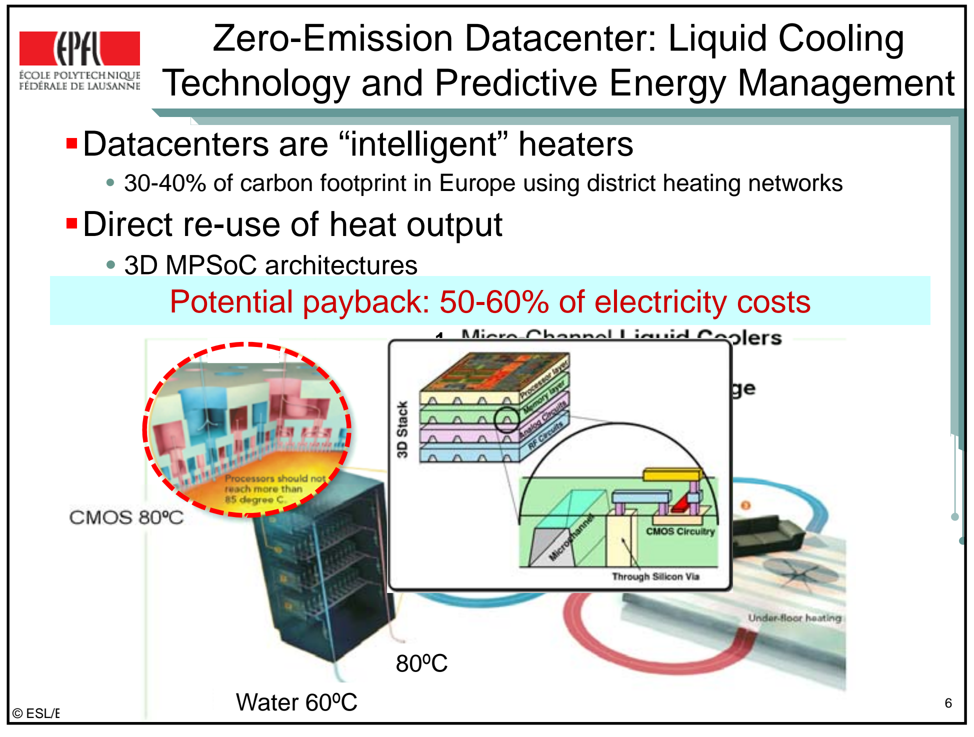
© ESL/EPFL 2010

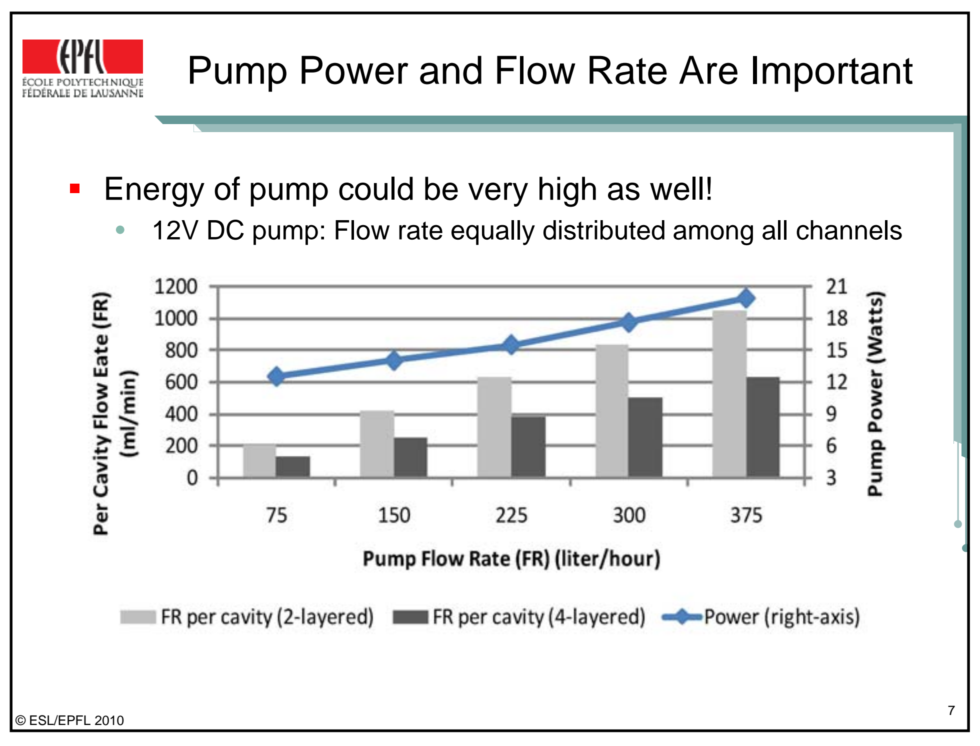


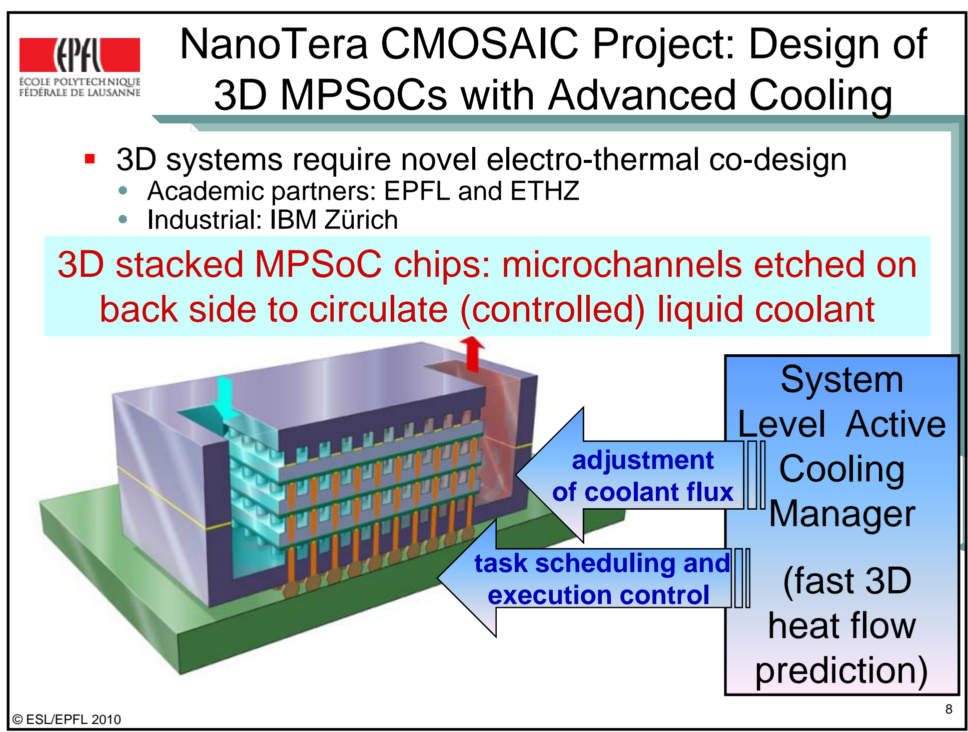

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Energy-Centric Design for Datacenters: System-Level View
Multi-scale energy management solution needed
Architecture: middleware, power management, network design, etc.
One server
DRAM: 16GB, 100ns, 20GB/s
Disk: 2TB, 10ms, 200MB/s
Energy-proportional datacenters: Integrated layers of software-hardware to minimize joules/work
Technology: cooling, processing and storage systems, etc.
Local rack (80 servers)
DRAM: 1TB, 300us, 100MB/s
Disk: 160TB, 11ms, 100MB/s
Datacenter Switch
Cluster (30 racks)
DRAM: 30TB, 500us, 10MB/s
Disk: 4.80PB, 12ms, 10MB/s
Barroso & Hölzle, 2009
© ESL/EPFL 2010
[Courtesy: Prof. T. Simunic, UCSD] 3

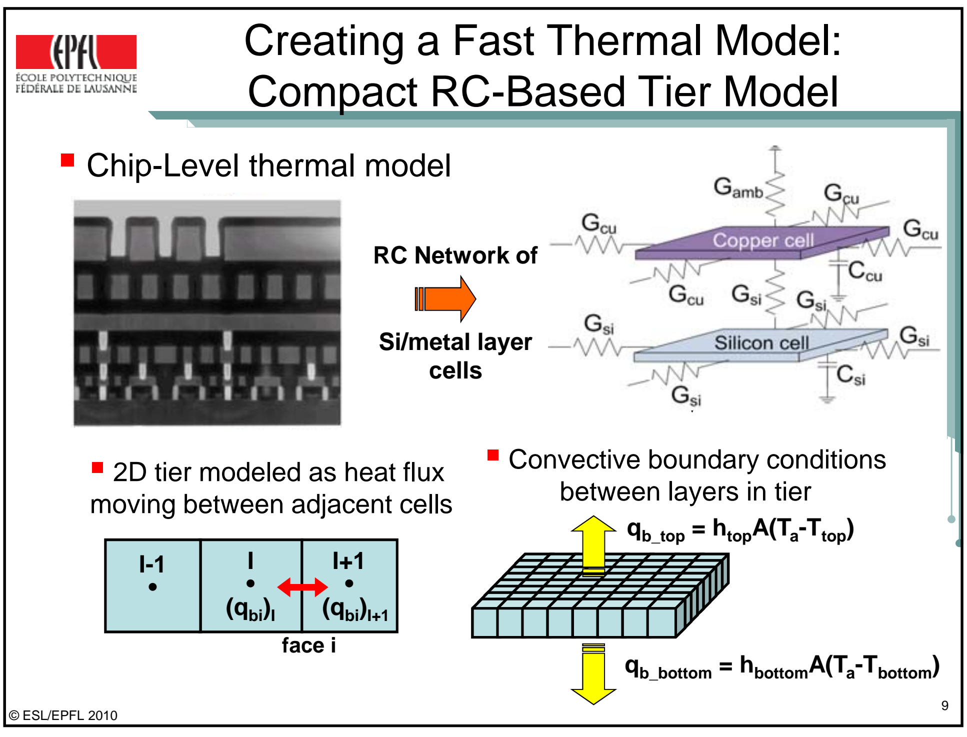

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Advantages of 3D vs. 2D Multi-Processor System-on-Chip (MPSoCs) ICs
Promises
Reduce average length of on-chip global wires
Increase number of devices reachable in given time budget
2D Routing (large chip)
3D Routing (small chip)
Greatly facilitate massive storage integration (i.e., logic-DRAM stacks)
Eight Stacked Chips [WSP]
560um
Samsung Wafer Stack Package (WSP) memory
C4 Bump
Package
Adhesive Layer
TSV
Thinned DRAM Layers
Metal ILD
Active Silicon
Unthinned Processor Layer
Substrate
Thermal Interface Material
Heat Spreader
Heat Sink
Three-Dimensional Multi-Processor System-on-Chip (MPSoC) IC Concept
© ESL/EPFL 2010
4

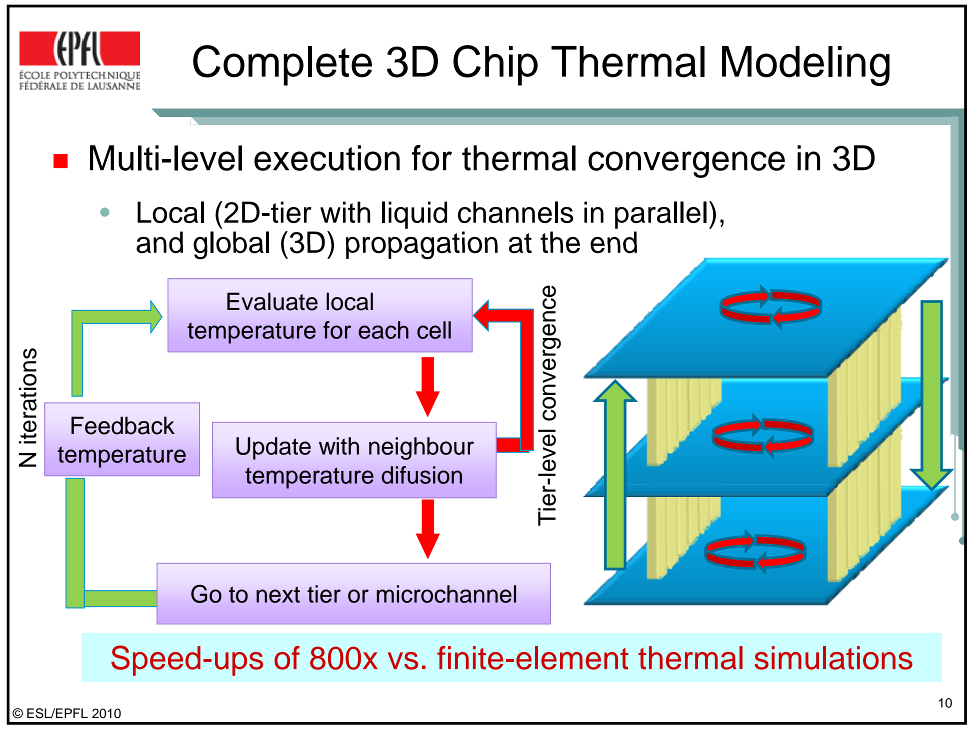

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Run-Time Heat Spreading in 3D MPSoCs
5-tier 3D stack: 10 heat sources and sensors
Inject between 4W – 1.5W
2nd Tier
3rd Tier
4th Tier
5th Tier
Large and non-uniform heat propagation!
(up to 130º C on top tier)
© ESL/EPFL 2010
5

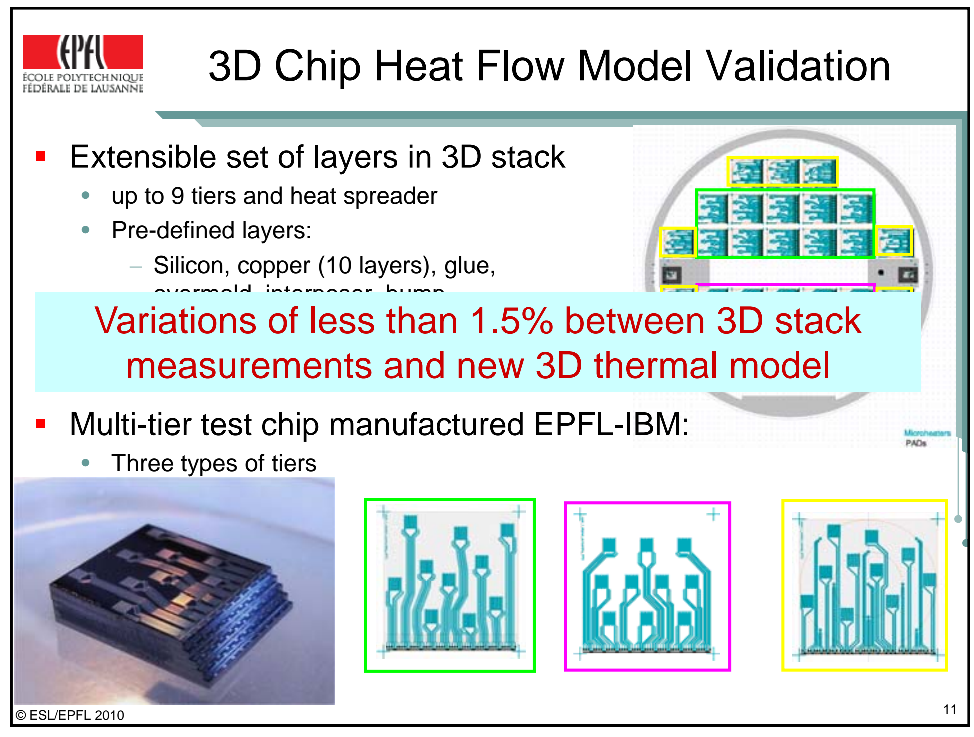

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Zero-Emission Datacenter: Liquid Cooling Technology and Predictive Energy Management
Datacenters are "intelligent" heaters
30-40% of carbon footprint in Europe using district heating networks
Direct re-use of heat output
3D MPSoC architectures
Potential payback: 50-60% of electricity costs
CMOS 80ºC
80ºC
Water 60ºC
3D Stack
CMOS Circuitry
Through Silicon Via
Under-floor heating
© ESL/E
6


Pump Power and Flow Rate Are Important
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Energy of pump could be very high as well!
12V DC pump: Flow rate equally distributed among all channels
Per Cavity Flow Eate (FR) (ml/min)
Pump Power (Watts)
Pump Flow Rate (FR) (liter/hour)
75
150
225
300
375
FR per cavity (2-layered)
FR per cavity (4-layered)
Power (right-axis)
© ESL/EPFL 2010
7

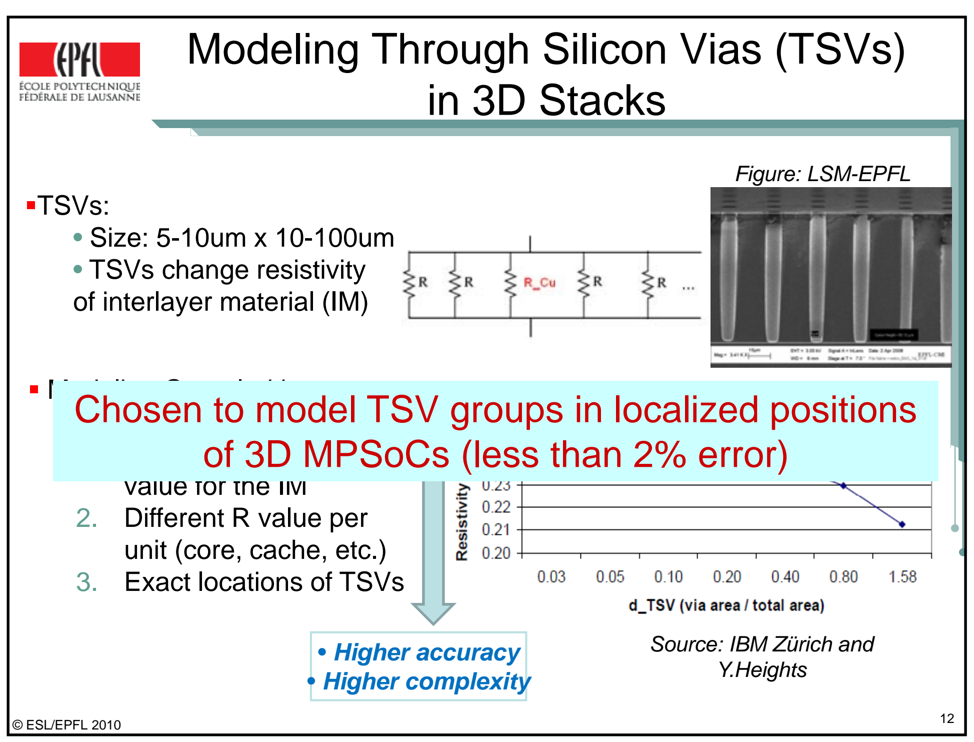

NanoTera CMOSAIC Project: Design of 3D MPSoCs with Advanced Cooling
3D systems require novel electro-thermal co-design
Academic partners: EPFL and ETHZ
Industrial: IBM Zürich
3D stacked MPSoC chips: microchannels etched on back side to circulate (controlled) liquid coolant
adjustment of coolant flux
task scheduling and execution control
System Level Active Cooling Manager
(fast 3D heat flow prediction)
© ESL/EPFL 2010
8


Creating a Fast Thermal Model: Compact RC-Based Tier Model
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Chip-Level thermal model
RC Network of
Si/metal layer cells
Copper cell
Silicon cell
2D tier modeled as heat flux moving between adjacent cells
Convective boundary conditions between layers in tier
face i
© ESL/EPFL 2010
9

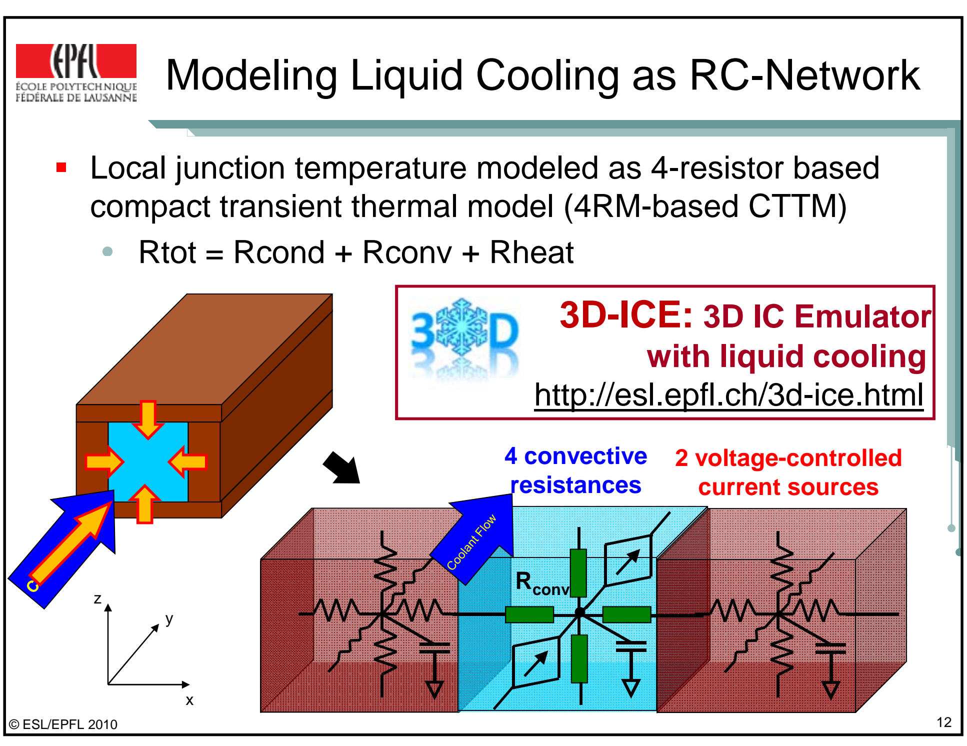

Complete 3D Chip Thermal Modeling
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Multi-level execution for thermal convergence in 3D
Local (2D-tier with liquid channels in parallel), and global (3D) propagation at the end
Evaluate local temperature for each cell
Update with neighbour temperature difusion
Feedback temperature
Go to next tier or microchannel
N iterations
Tier-level convergence
Speed-ups of 800x vs. finite-element thermal simulations
© ESL/EPFL 2010
10



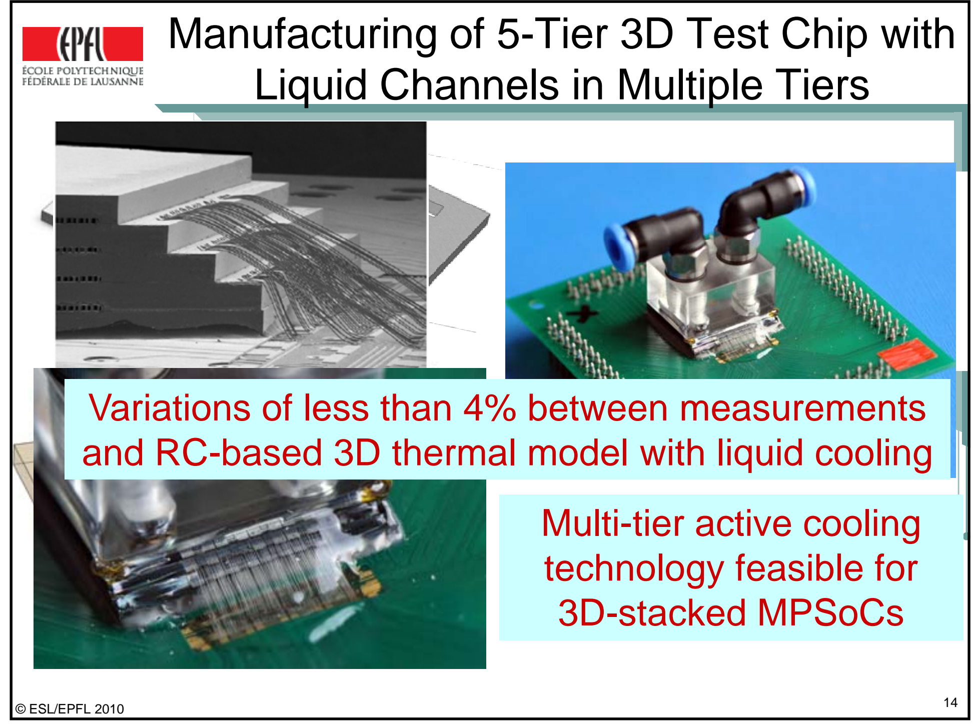
3D Chip Heat Flow Model Validation
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Extensible set of layers in 3D stack
up to 9 tiers and heat spreader
Pre-defined layers:
Silicon, copper (10 layers), glue,
Variations of less than 1.5% between 3D stack measurements and new 3D thermal model
Multi-tier test chip manufactured EPFL-IBM:
Three types of tiers
© ESL/EPFL 2010
11



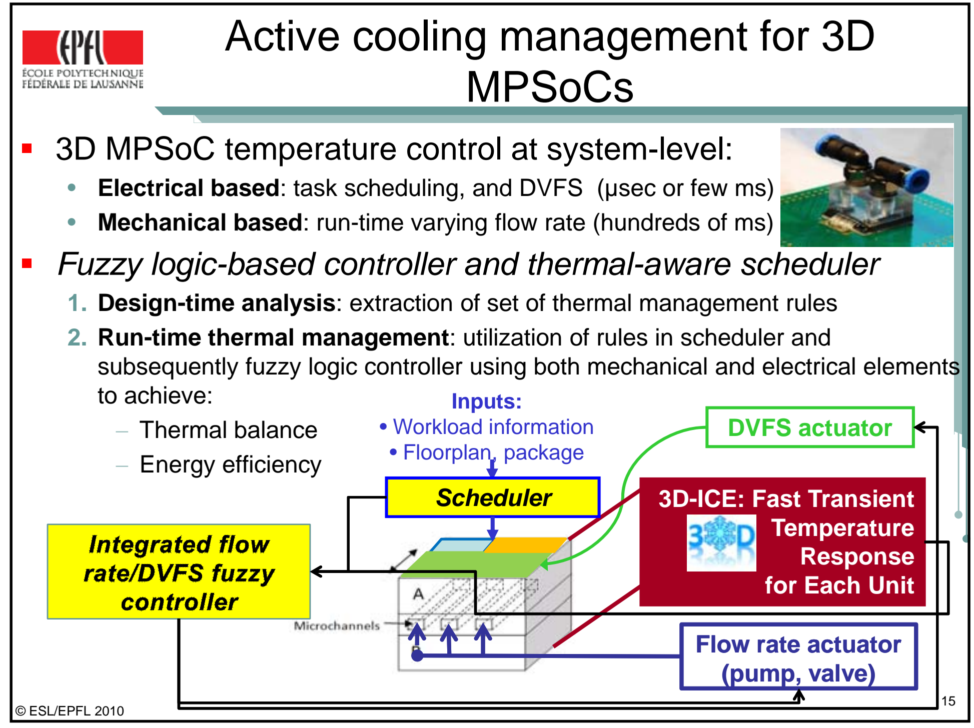
Modeling Through Silicon Vias (TSVs) in 3D Stacks
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Figure: LSM-EPFL
TSVs:
Size: 5-10um x 10-100um
TSVs change resistivity of interlayer material (IM)
R
R_Cu
Chosen to model TSV groups in localized positions of 3D MPSoCs (less than 2% error)
2. Different R value per unit (core, cache, etc.)
3. Exact locations of TSVs
Resistivity
0.22
0.21
0.20
0.03
0.05
0.10
0.20
0.40
0.80
1.58
d_TSV (via area / total area)
Higher accuracy
Higher complexity
Source: IBM Zürich and Y.Heights
© ESL/EPFL 2010
12

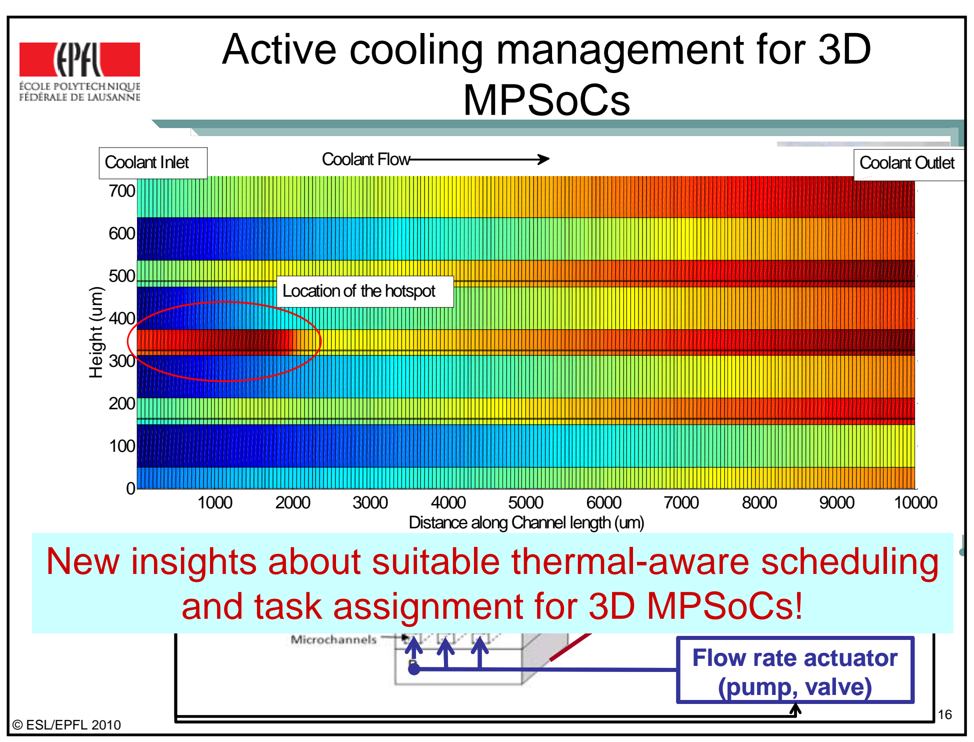


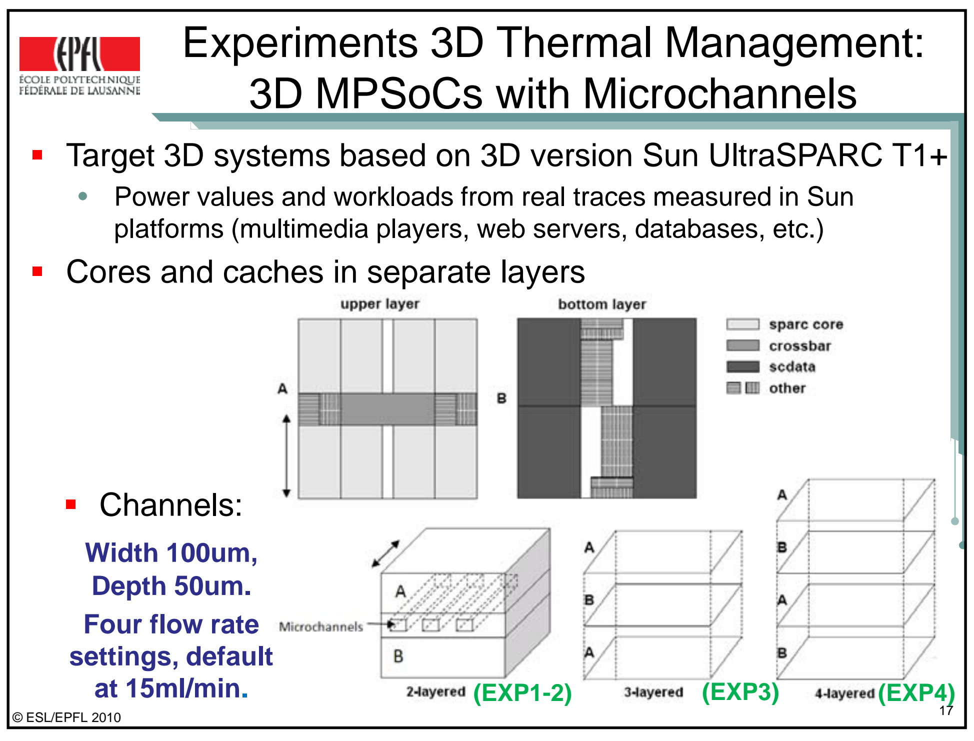
Modeling Liquid Cooling as RC-Network
Local junction temperature modeled as 4-resistor based compact transient thermal model (4RM-based CTTM)
Rtot = Rcond + Rconv + Rheat
3D-ICE: 3D IC Emulator with liquid cooling
http://esl.epfl.ch/3d-ice.html
4 convective resistances
2 voltage-controlled current sources
Coolant Flow
Rconv
z
y
x
© ESL/EPFL 2010
12

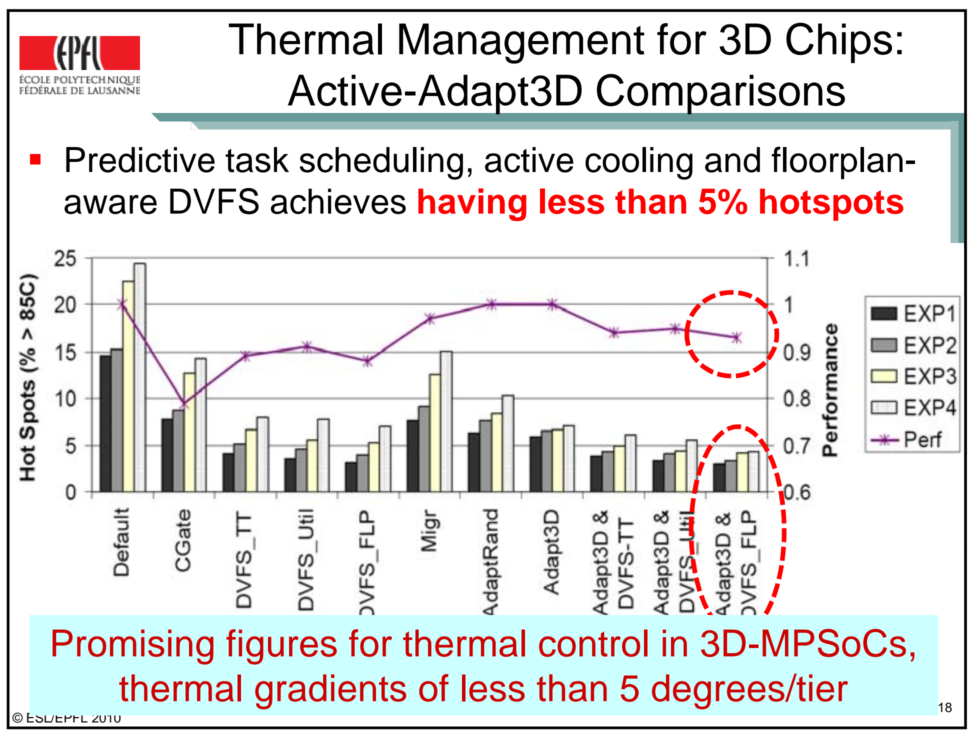


Manufacturing of 5-Tier 3D Test Chip with Liquid Channels in Multiple Tiers
Variations of less than 4% between measurements and RC-based 3D thermal model with liquid cooling
Multi-tier active cooling technology feasible for 3D-stacked MPSoCs
© ESL/EPFL 2010
14


Active cooling management for 3D MPSoCs
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
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 elements to achieve:
– Thermal balance
– Energy efficiency
Inputs:
• Workload information
• Floorplan, package
DVFS actuator
Scheduler
3D-ICE: Fast Transient Temperature Response for Each Unit
Integrated flow rate/DVFS fuzzy controller
Microchannels
Flow rate actuator (pump, valve)
© ESL/EPFL 2010
15


Active cooling management for 3D MPSoCs
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Coolant Inlet
Coolant Flow
Coolant Outlet
Location of the hotspot
Height (um)
700
600
500
400
300
200
100
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Distance along Channel length (um)
New insights about suitable thermal-aware scheduling and task assignment for 3D MPSoCs!
Microchannels
Flow rate actuator (pump, valve)
© ESL/EPFL 2010
16


Experiments 3D Thermal Management: 3D MPSoCs with Microchannels
Target 3D systems based on 3D version Sun UltraSPARC T1+
Power values and workloads from real traces measured in Sun platforms (multimedia players, web servers, databases, etc.)
Cores and caches in separate layers
upper layer
bottom layer
sparc core
crossbar
scdata
other
Channels:
Width 100um, Depth 50um.
Four flow rate settings, default at 15ml/min.
Microchannels
2-layered (EXP1-2)
3-layered (EXP3)
4-layered (EXP4)
© ESL/EPFL 2010
17


Thermal Management for 3D Chips: Active-Adapt3D Comparisons
Predictive task scheduling, active cooling and floorplan-aware DVFS achieves having less than 5% hotspots
Hot Spots (% > 85C)
Performance
EXP1
EXP2
EXP3
EXP4
Perf
Default
CGate
DVFS_TT
DVFS_Util
DVFS_FLP
Migr
AdaptRand
Adapt3D
Adapt3D & DVFS-TT
Adapt3D & DVFS_Util
Adapt3D & DVFS_FLP
Promising figures for thermal control in 3D-MPSoCs, thermal gradients of less than 5 degrees/tier
© ESL/EPFL 2010
18

Conclusions

- Green datacenters require novel system-level thermal-aware design methods
  - Application of new 3D chips MPSoC architectures and novel liquid cooling methods for energy-efficient solutions
  - Handling complexity in cooling and control implies an inter-disciplinary research work
    - Fast RC thermal models applicable even for variable liquid fluxes in 3D stacks (**errors of less than 4%)**
    - Multi-scale thermal management opportunities (**mechanical and electrical methods**) need to be managed at system-level
- New generation of thermal managers: proactive controllers adjusts flow rate considering task assignment and DVFS
  - Proactive control improves the **hot spot reduction to 95%** for systems with variable flow rates, and **reduces thermal variations**
  - Holistic control reduces significantly the energy cost for the whole system (**67% power savings!**)

© ESL/EPFL 2010

Key References and Bibliography

- Thermal modeling and FPGA-based emulation
  - "**HW-SW Emulation Framework for Temperature-Aware Design in MPSoCs**", David Atienza, et al. *ACM Trans. on Design Automation for Embedded Systems (TODAES)*, Vol. 12, Nr. 3, pp. 1–26, August 2007.
  - "**Emulation-Based Transient Thermal Modeling of 2D/3D Systems-on-Chip with Active Cooling**", David Atienza, *Proc. of the 15th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC '09)*, IEEE Press, Belgium, October, 2009.
- Thermal management for 3D MPSoCs
  - "**Dynamic Thermal Management in 3D Multicore Architectures**", Ayse K. Coskun, et al., *Proc. of Design, Automation and Test in Europe (DATE '09)*, France, April 2009.
  - "**Modeling and Dynamic Management of 3D Multicore Systems with Liquid Cooling**", Ayse K. Coskun, et al., *Proc. of 17th Annual IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC '09)*, Brazil, October 2009. **(Best Paper Award)**
  - "**Through Silicon Via-Based Grid for Thermal Control in 3D Chips**", Jose L. Ayala, et al., *Proc. of the 4th International ICST Conference on Nano-Networks (Nano-Net '09)*, Switzerland, October 2009.

© ESL/EPFL 2010

EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Thank You
nano-tera.ch
Nano-Tera.ch Swiss Engineering Programme
European Commission
FNSNF
Swiss National Science Foundation
QUESTIONS ?
© ESL/EPFL 2010