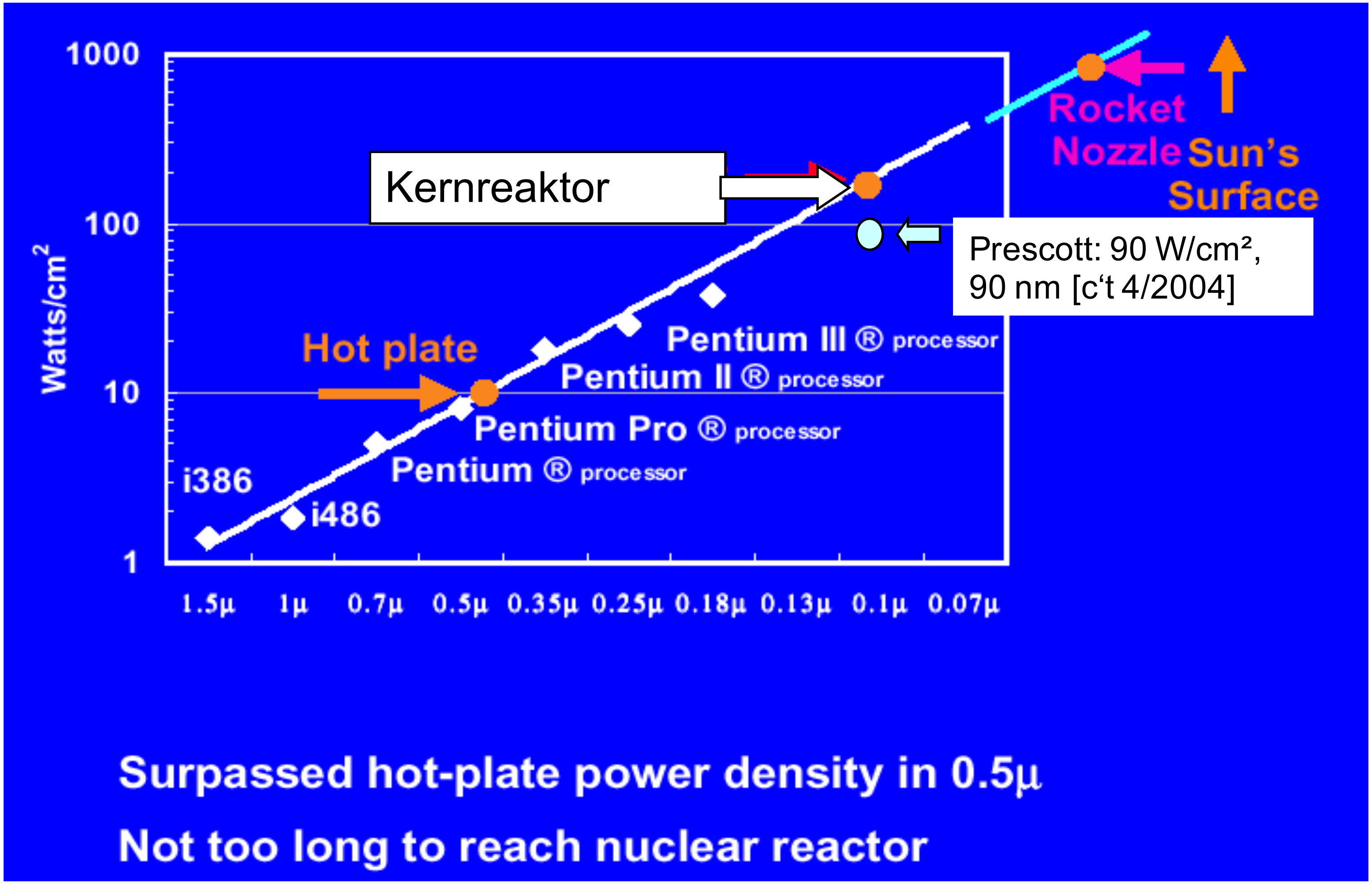
Temperaturprobleme der modernen Rechnerarchitekturen

Basis:

- Pagani et al., DATE 2015, CODES+ISSS 2014
- Babak Falsafi: Dark Silicon & Its Implications on Server Chip Design, Microsoft Research, Nov. 2010
  Siehe auch *publications* unter http://parsa.epfl.ch/~falsafi/
- Hadi Esmaeilzadeh: Dark Silicon and the End of Multicore Scaling, International Symposium on Computer Architecture (ISCA ’11)

PCs: Leistungsdichte (Power density) steigt!!!!!!



Prescott: 90 W/cm², 90 nm [c't 4/2004]

Surpassed hot-plate power density in 0.5µ

Not too long to reach nuclear reactor

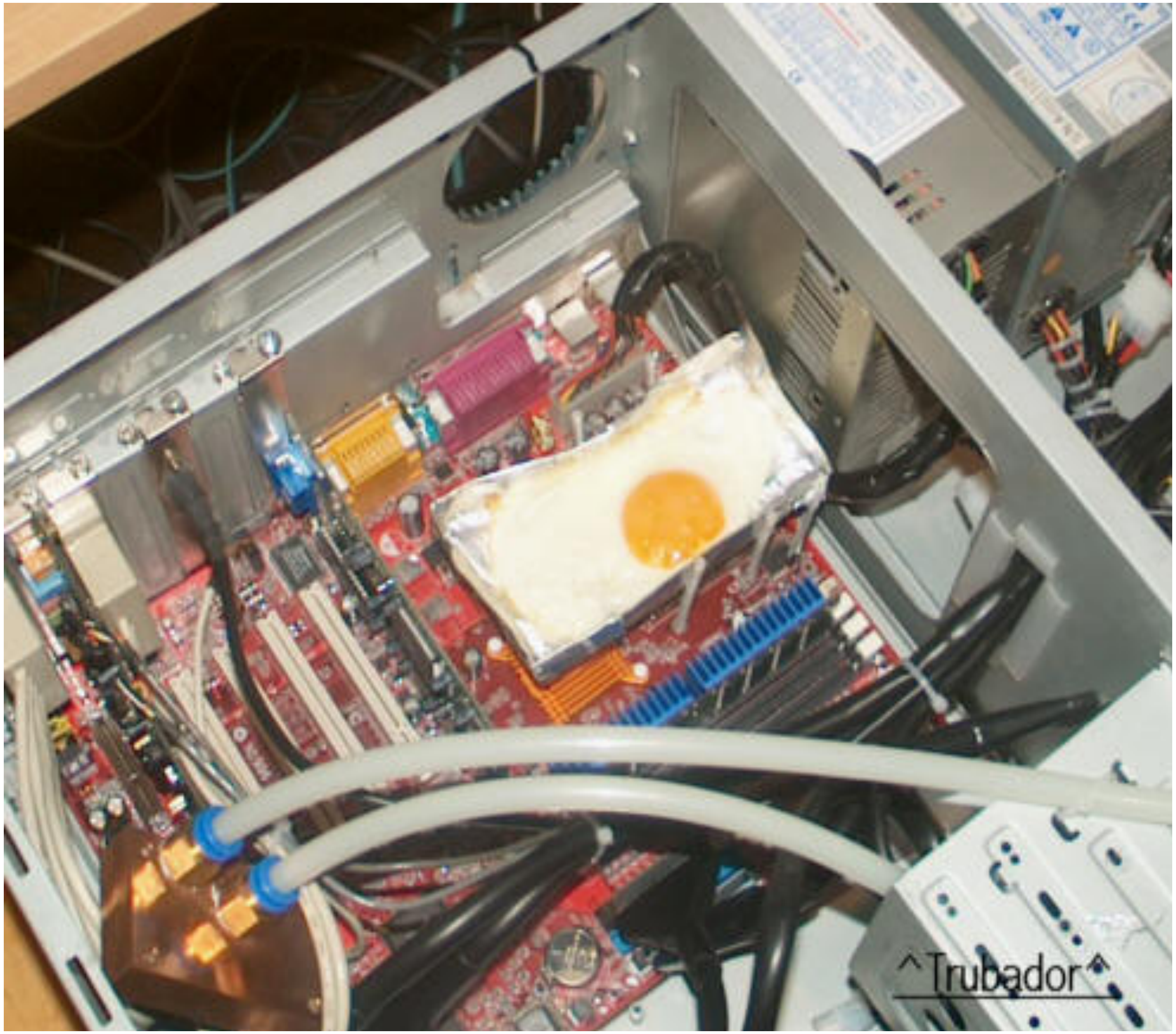
© Intel
M. Pollack,
Micro-32

PCs: Just adding transistors would have resulted in this:

Reuters: December 9, 2004: Men should keep their laptops off their laps because they could damage fertility, an expert said on Thursday. Laptops, which reach high internal operating temperatures, can heat up the scrotum which could affect the quality and quantity of men's sperm. "The increase in scrotal temperature is significant enough to cause changes in sperm parameters," said Dr Yefim Sheynkin, an associate professor of urology at the State University of New York at Stony Brook.

Wie kochen wir?

PCs: Besser als Herdplatte …? Warum nicht?

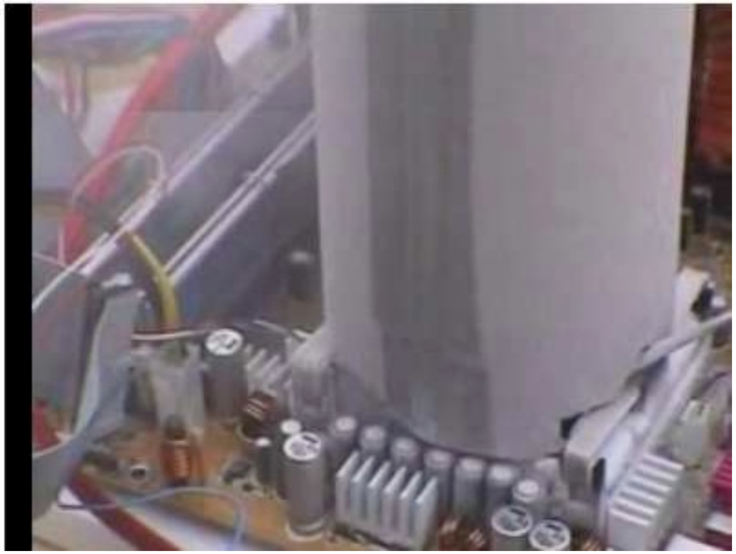


Strictly speaking, energy is not “consumed”, but converted from electrical energy into heat energy

http://www.phys.ncku.edu.tw/~htsu/humor/fry_egg.html

Verrückte Kühlungsmethode

- Thermoelektrische Kühlung
- Wasserkühlung
- Kältekühlung
- usw.



Thermal Modeling: A Single Power Source

- Thermal conduction
  - Fourier's Law of Cooling: the temperate change is proportional to the different of the chip and the ambient temperature (or the heat sink temperature)
    - If the chip is hotter, the temperature change drops more
    - If the chip is cooler, the temperature change drops less
  - Heating generation is proportional to the power consumption
    - If the power consumption is larger, the temperature change increases more
    - If the power consumption is smaller, the temperature change increases less
  - *Therefore,* $T'(t) = uP(t) - v(T(t)-T_{amb})$
    - $T(t)$ is the temperature of the power source at time t
    - $P(t)$ is the power consumption of the power source at time t
    - T_{amb} is the ambient temperature. I will simple use it as 0. Why?
    - u and v are both hardware-dependent constants.

Solving Ordinary Differential Equation (ODE): *T‘(t) = uP(t) – vT(t)*

It is a standard linear ODE, where *u* and *v* are constants:

$$d\frac{T(t)e^{vt}}{dt} = e^{vt}d\frac{T(t)}{dt} + T(t)\cdot ve^{vt} = e^{vt}(uP(t) - vT(t)) + T(t)\cdot ve^{vt} = e^{vt}uP(t).$$

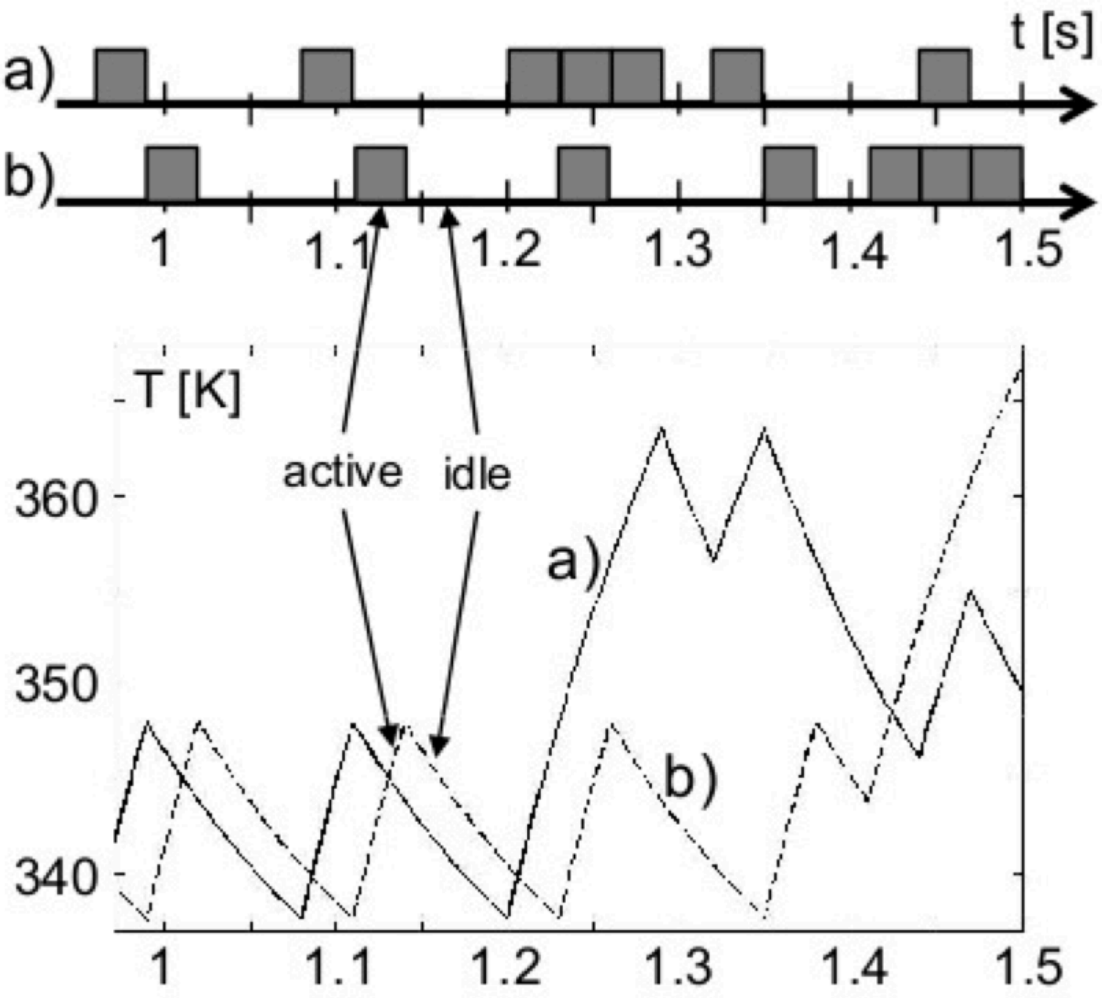
$$\int_{t_0}^{t} d\frac{T(t)e^{vt}}{dt} = \int_{t_0}^{t} e^{vt}uP(t) \Rightarrow T(t)e^{vt} - T(t_0)e^{vt_0)} = \int_{t_0}^{t} e^{vx}uP(x)dx$$

$$\Rightarrow T(t) - T(t_0)e^{-v(t-t_0)} = \int_{t_0}^{t} e^{v(x-t)}uP(x)dx$$

$$\Rightarrow T(t) = T(t_0)e^{-v(t-t_0)} + \int_{t_0}^{t} e^{v(x-t)}uP(x)dx$$

- The temperature effect at time t_0 decreases exponentially by $T(t_0)e^{-v(t-t_0)}$.
- The power consumption effect at time x decreases exponentially by $T(t_0)e^{v(x-t)}$, since $v > 0$ and $x - t \leq 0$ for $x \leq t$.

Different Traces versus Temperature

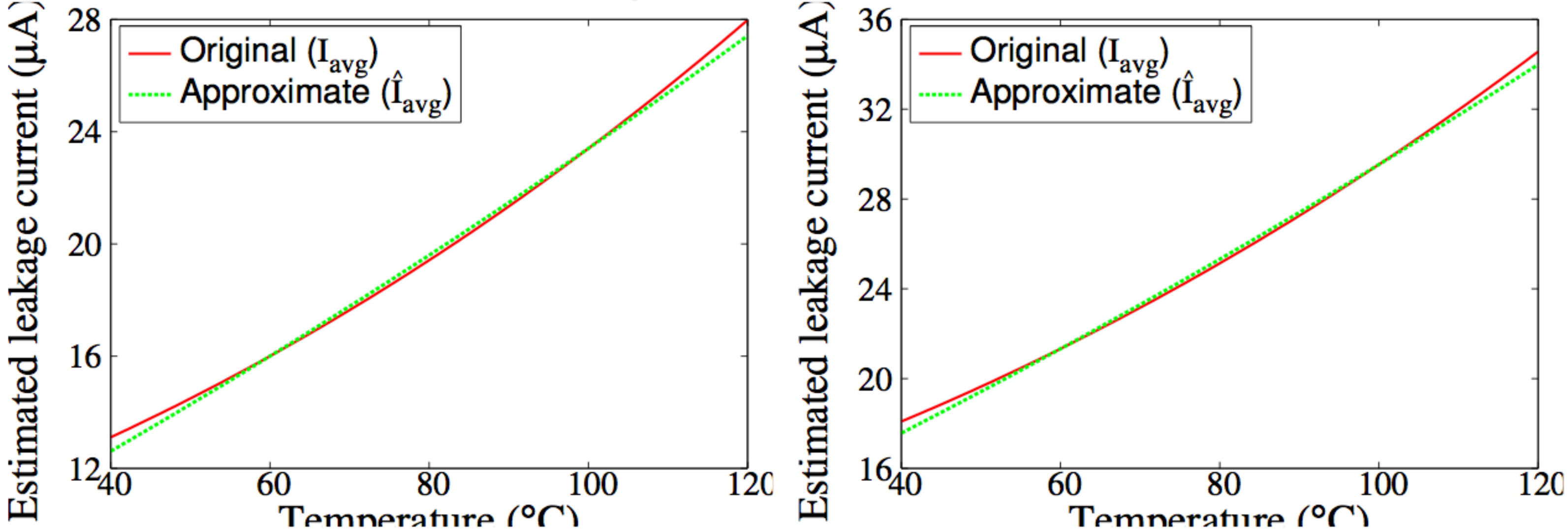


Thermal-Dependent Leakage Power Consumption

$$I_{avg}(T, V_{dd}) = I(T_0, V_0)\left(AT^2 e^{\left(\frac{q_1 \cdot V_{dd} + q_2}{T}\right)} + Be^{(\gamma \cdot V_{dd} + \delta)}\right),$$

However, the term $e^{(1/T)}$ dose not provide significant role in the accuracy. It is possible to use a simpler formula to formulate the leakage current.

$$\hat{I}_{avg}(T) = \hat{A}T^2 + \hat{B},$$



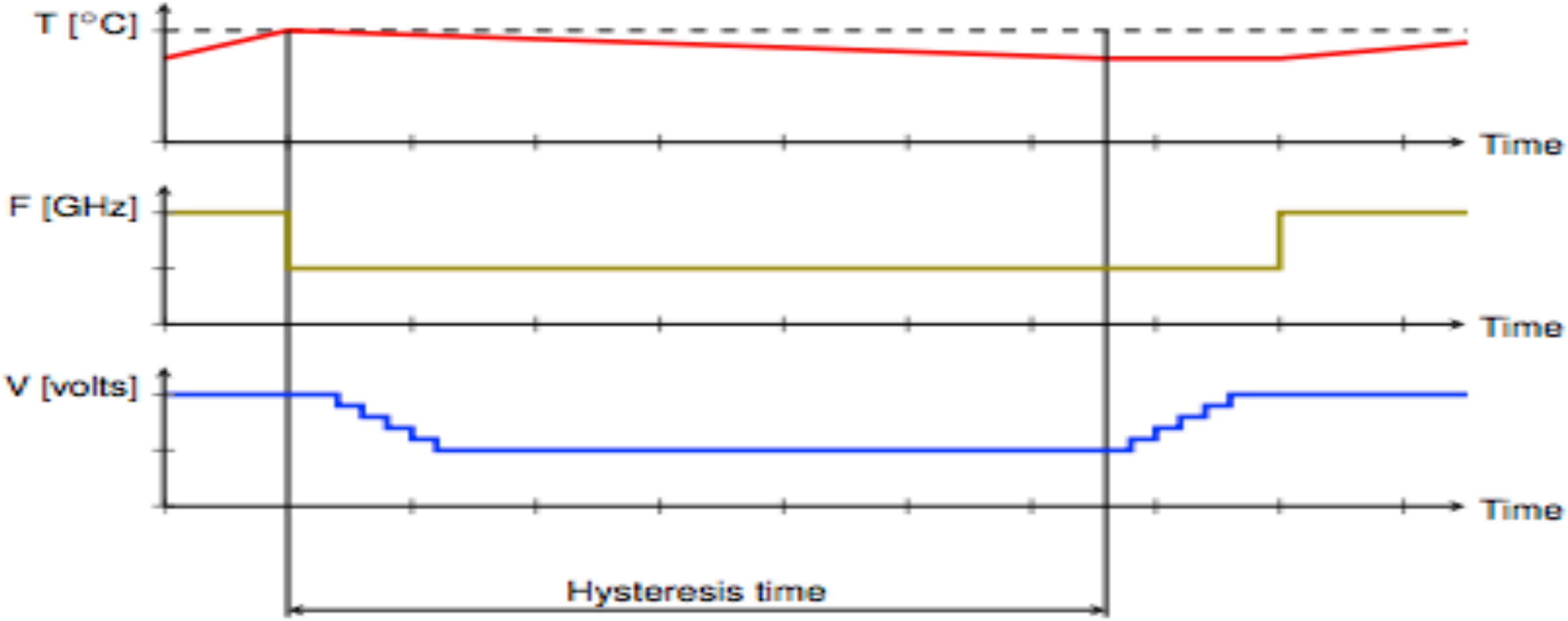
(a) $V_{dd} = 0.95$V

(b) $V_{dd} = 1.05$V

Chuan-Yue Yang, Jian-Jia Chen, Lothar Thiele, Tei-Wei Kuo: Energy-efficient real-time task scheduling with temperature-dependent leakage. DATE 2010: 9-14

Dynamic Thermal Management (DTM)

- Avoid possible over heating
  - DVFS
  - DPM



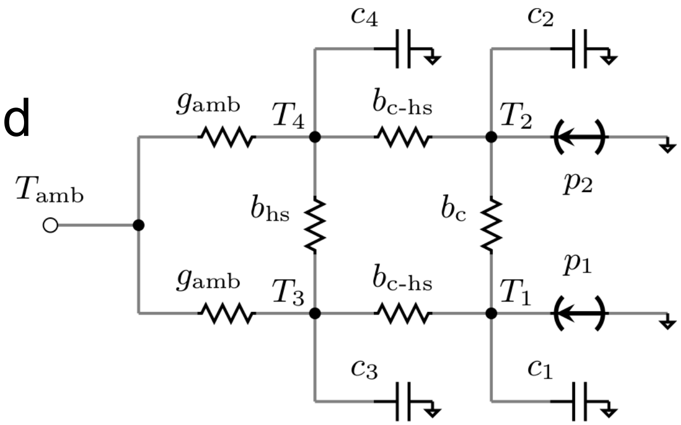
Thermal Networks – Multiple Heat Sources

- Thermal models of applications depend on neighbouring cores.
  - A resistance-capacitance (RC) thermal network is widely used
    - A set of first order differential equations
  - Steady states (the equilibrium temperatures if the power does not change)
    - Simple linear algebra
  - Transient states (temperature profile in time)
    - Approximate the solution by using fourth-order *Runge- Kutta* numerical method [HotSpot, Huang et al. 2009]
    - Exact solution by using matrix exponential (many approximations are available) methods [MatEx, Pagani et al. to be published in DATE 2015]

[P.-Y. Huang and Y.-M. Lee, "Full-chip thermal analysis for the early design stage via generalized integral transforms," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, no. 5, pp. 613–626, May 2009. ;]
[Santiago Pagani, Muhammad Shafique, Jian-Jia Chen and Jörg HenkelMatEx: Efficient Transient and Peak Temperature Computation for Compact Thermal Modelsin 18th Design, Automation & Test in Europe (DATE) 2015 ;]

Thermal Model

- Thermal model → System of first-order differential equations
  - Relates temperature with power values and T_{amb}
  - For example, RC thermal networks (like HotSpot)
- RC thermal network details



Temperature Vectors

$$\mathbf{A}\mathbf{T}' + \mathbf{B}\mathbf{T} = \mathbf{P} + T_{\mathrm{amb}}\mathbf{G}$$

Constant Vector

Constant Matrices

Power Vector

Ambient Temperature

$$\mathbf{T}_{\mathrm{steady}} = \mathbf{B}^{-1}\mathbf{P} + T_{\mathrm{amb}}\mathbf{B}^{-1}\mathbf{G}$$

Computing Transient Temperatures

RC thermal network

Express with matrix exponentials

Solve the matrix exponential

Analytical Solution: Tk(t)

Thermal equation with matrix exponentials

Steady-State Temperatures (where vector T converges)

$$\mathbf{T} = \mathbf{T}_{\text{steady}} + e^{\mathbf{C}t}(\mathbf{T}_{\text{init}} - \mathbf{T}_{\text{steady}})$$

Matrix Exponential

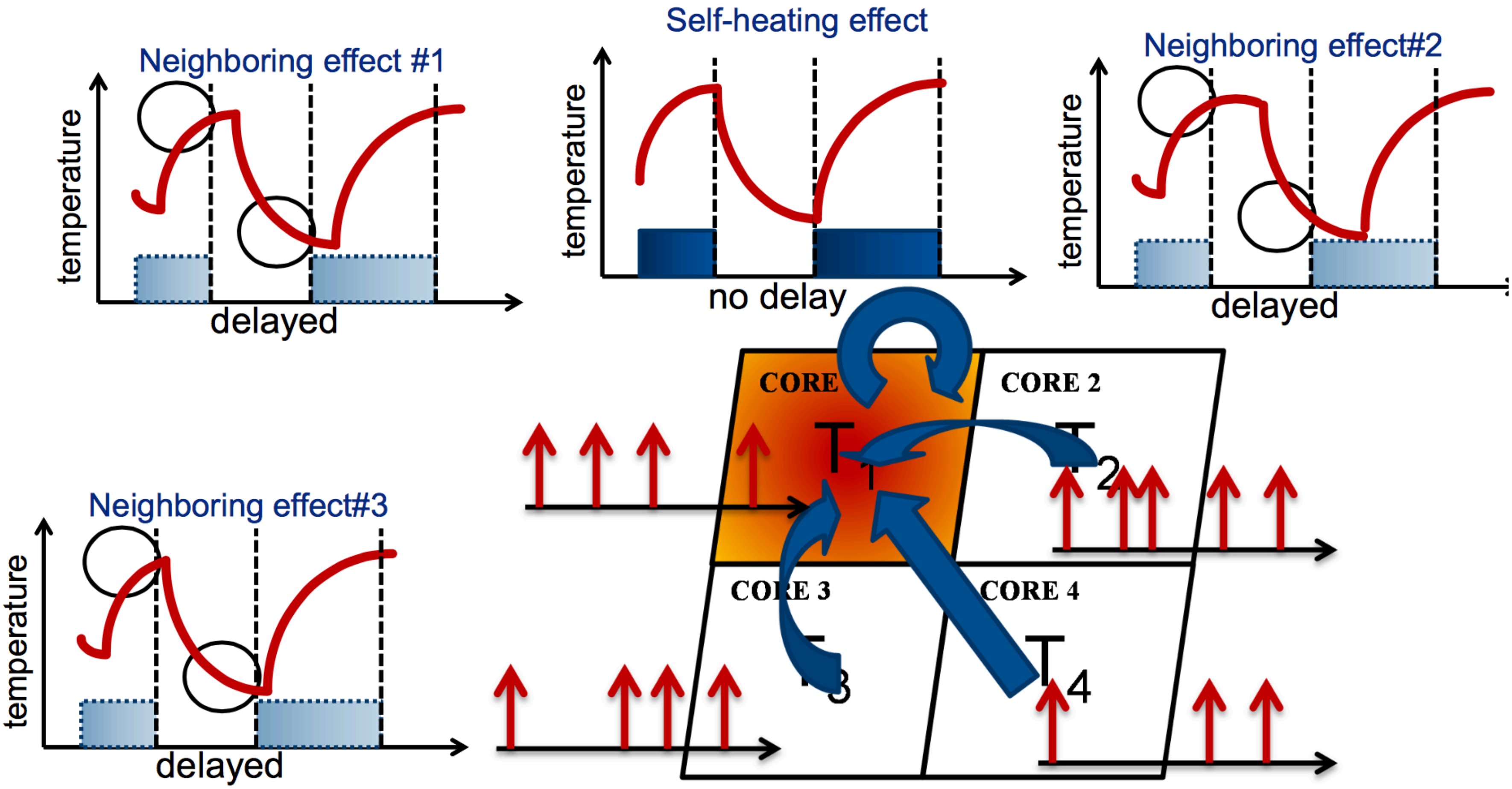
Initial Temperatures (at *t* = *0*)

$$\mathbf{C} = -\mathbf{A}^{-1}\mathbf{B}$$

$$\mathbf{T}_{\text{steady}} = \mathbf{B}^{-1}\mathbf{P} + T_{\text{amb}}\mathbf{B}^{-1}\mathbf{G}$$

Pagani et al., DATE 2015

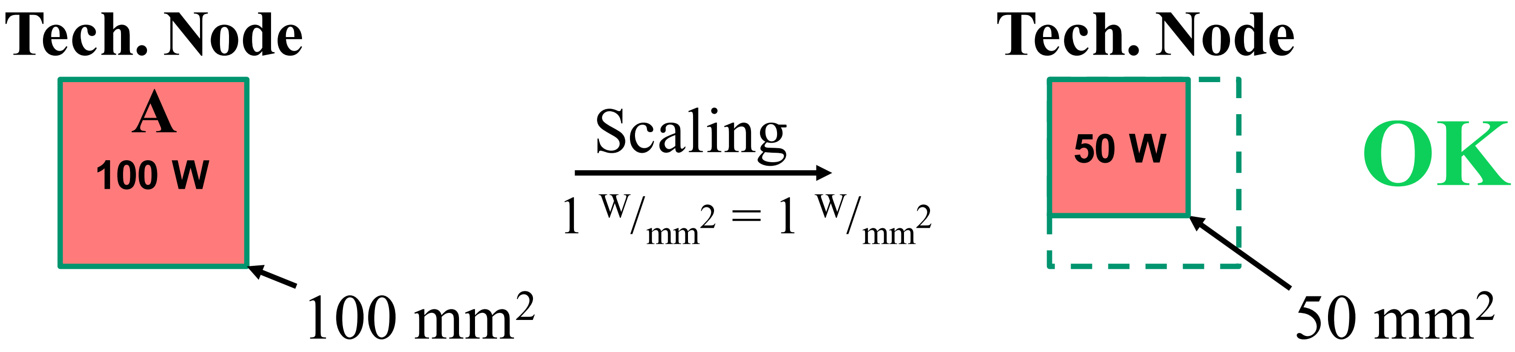
Heat Transfer: Impulse Response



Thiele @ ETHZ

The Dark Silicon Problem

So far: Constant power density



Tech. Node

A

100 W

$100\ mm^2$

Scaling

$1\ {}^{W}/_{mm^2} = 1\ {}^{W}/_{mm^2}$

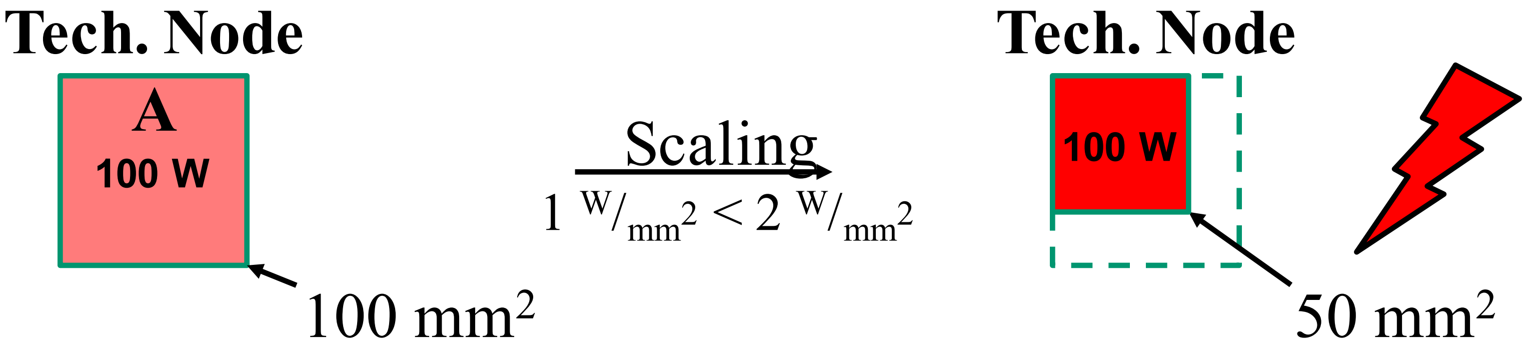
Tech. Node

50 W

$50\ mm^2$

OK

- Expected: Power density increases



Tech. Node

A

100 W

$100\ mm^2$

Scaling

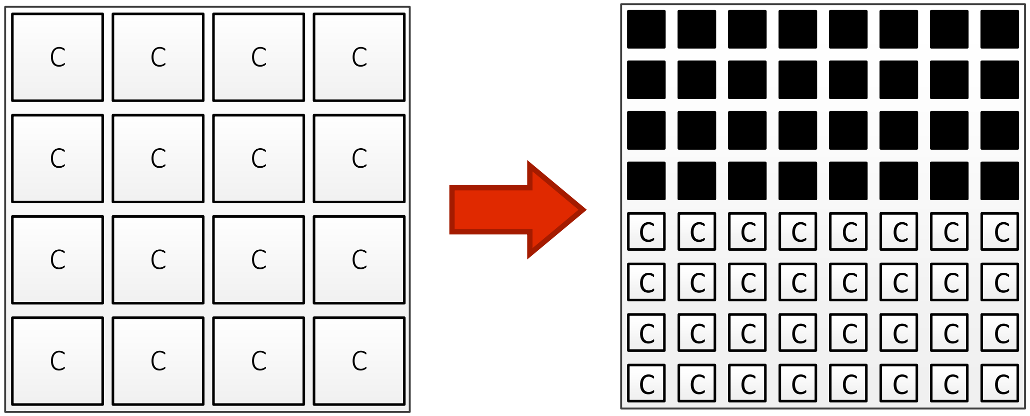
$1\ {}^{W}/_{mm^2} < 2\ {}^{W}/_{mm^2}$

Tech. Node

100 W

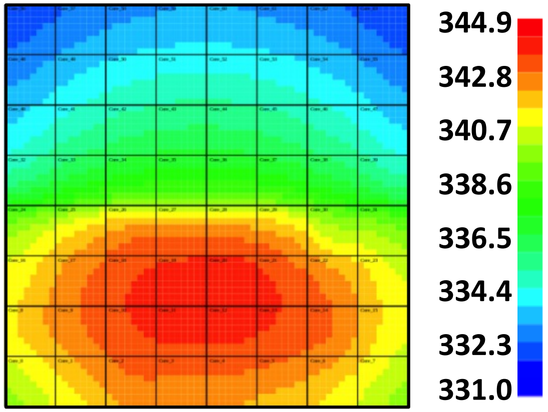
$50\ mm^2$

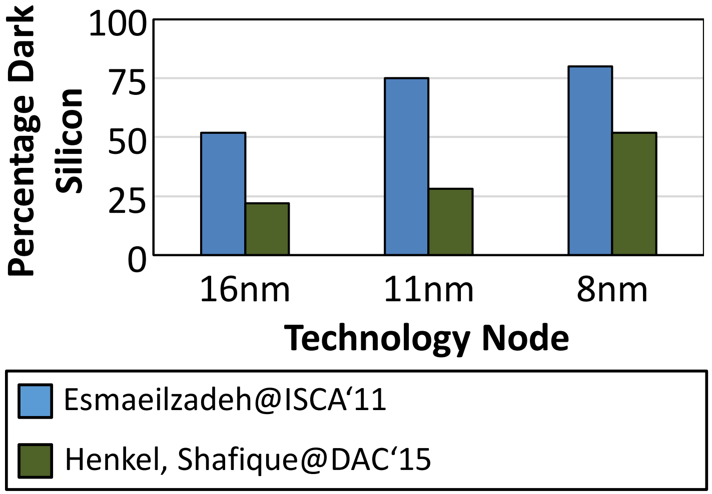
The Emerging Dark Silicon Problem



22nm

11nm and Beyond





Caveat: Simple Parallelizable Workloads

Workloads are assumed parallel

- Scaling server workloads is reasonable

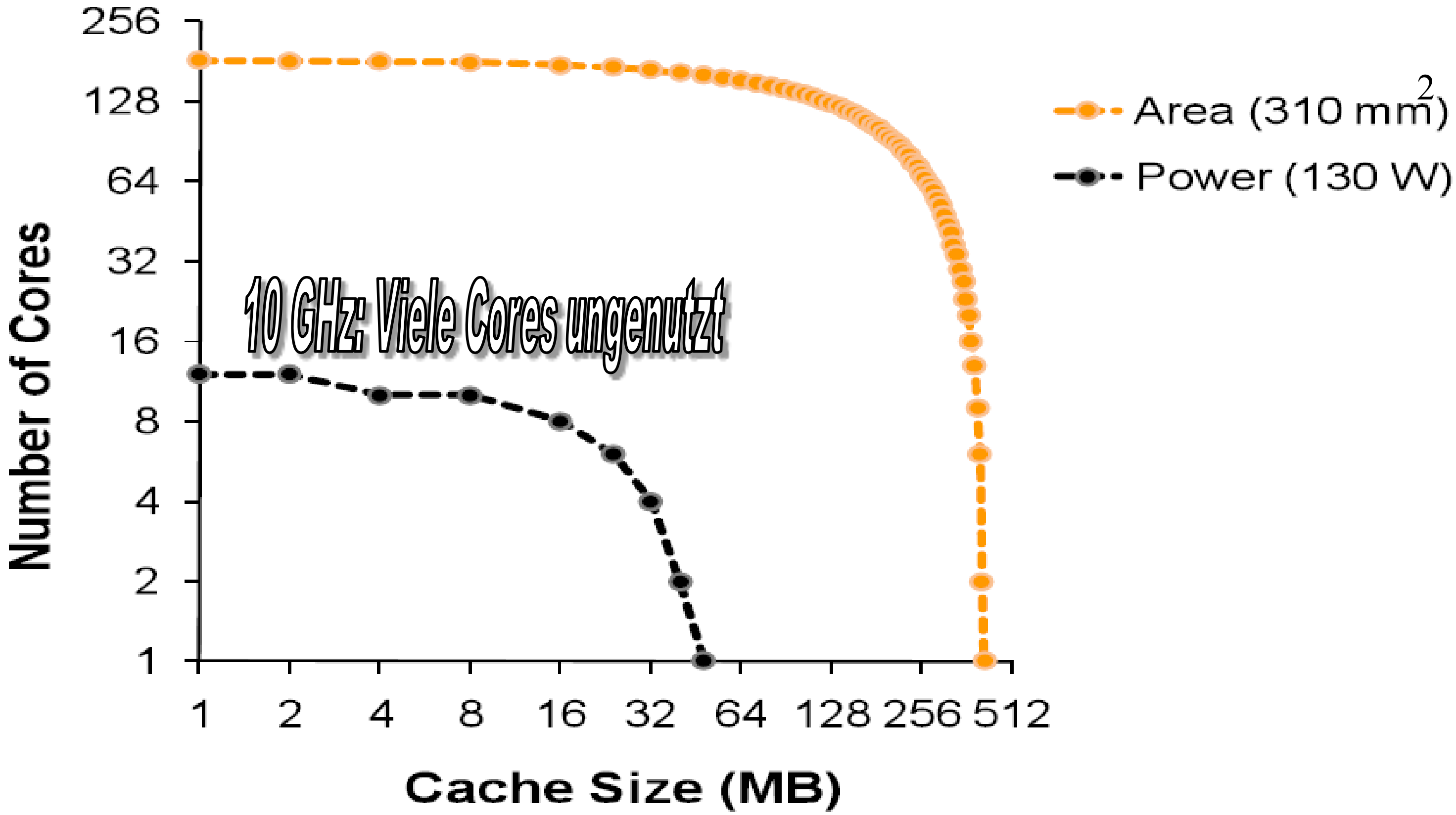
CPI model:

- Works well for workloads with low MLP
- OLTP, Web & DSS are mostly memory-latency dependent

Future servers will run a mix of workloads

DSS=decision support system

Area vs. Power Envelope (22nm)

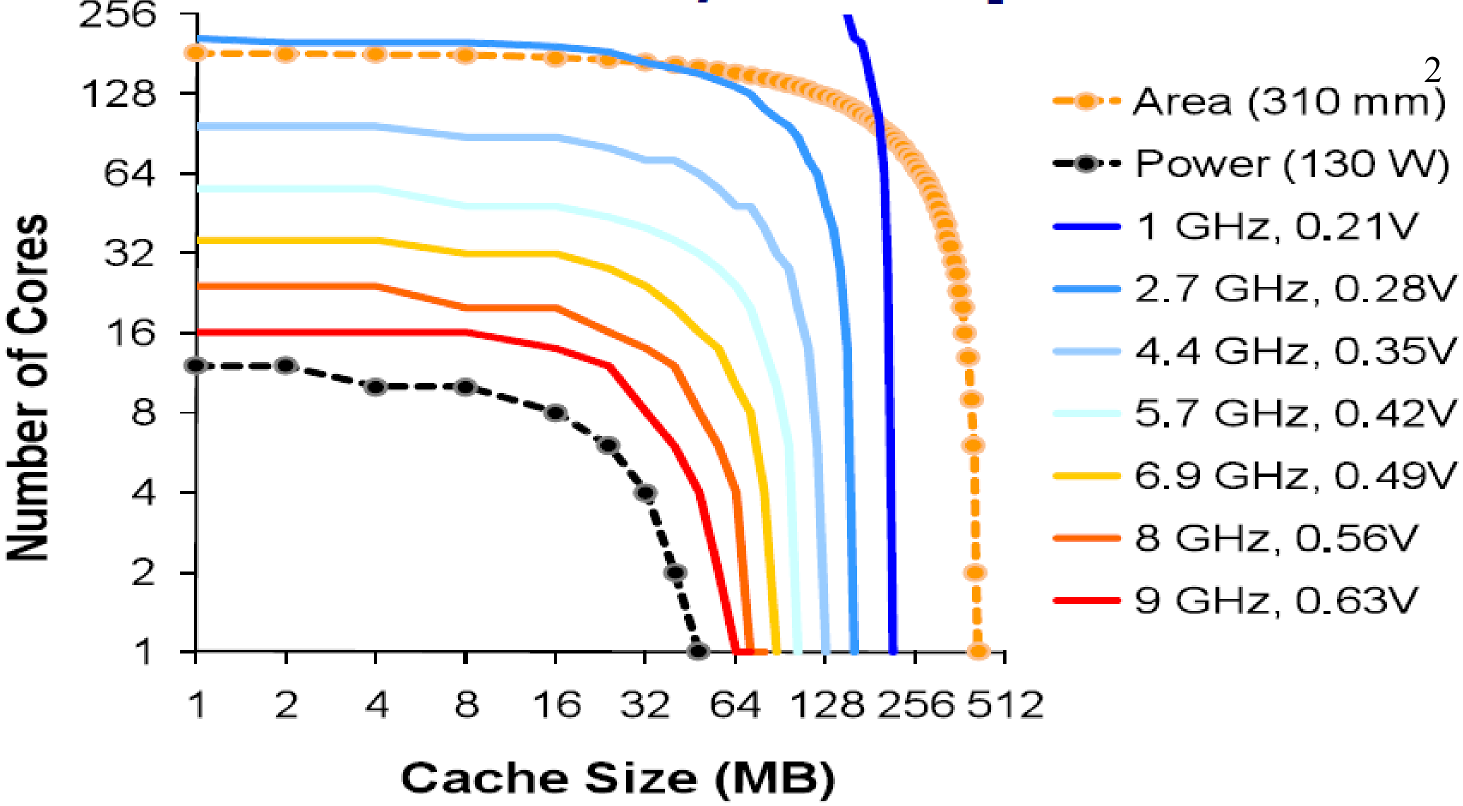


10 GHz: Viele Cores ungenutzt

✓ Good news: can fit hundreds of cores

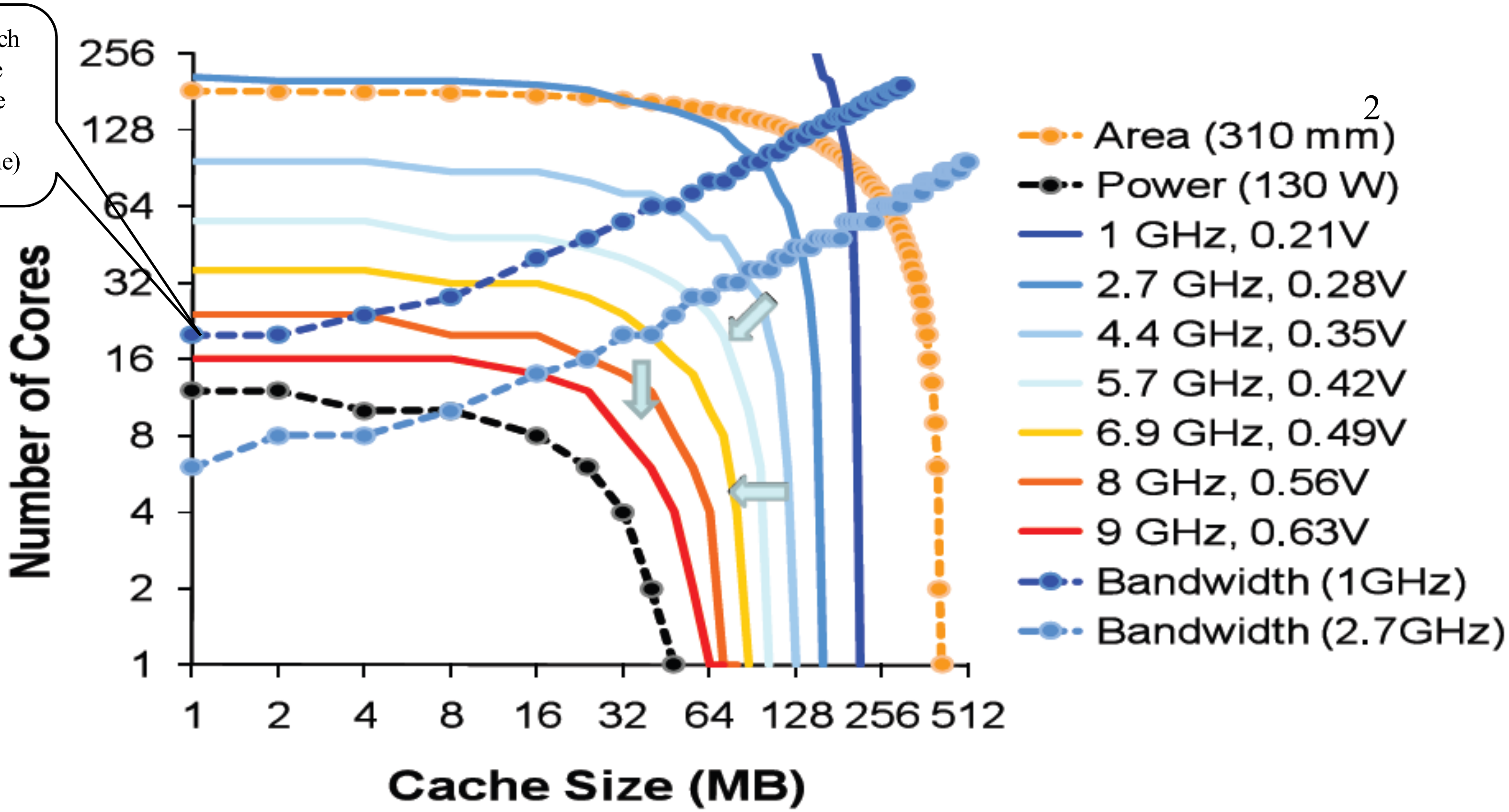
× Can not use them all at highest speed

Of course one could pack more slower cores, cheaper cache



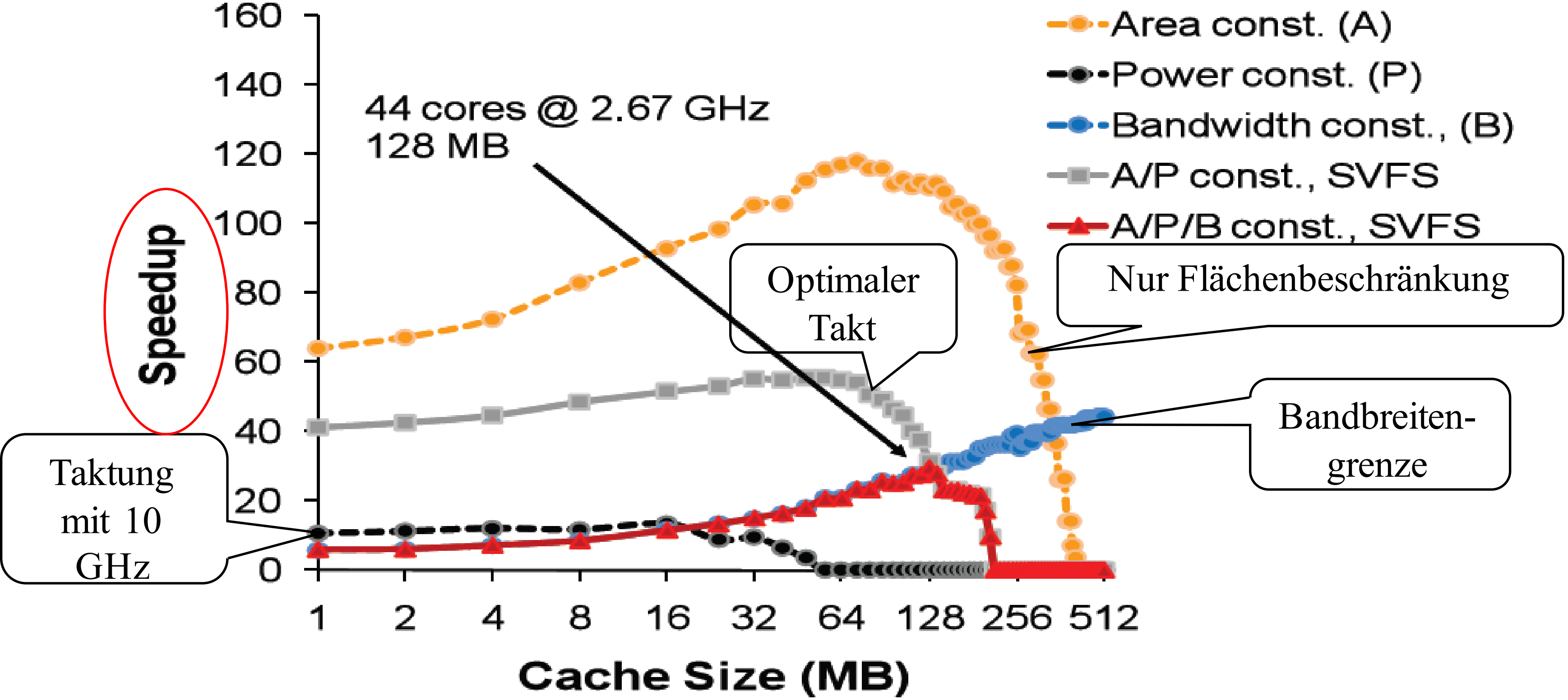
- Result: a performance/power trade-off
- Assuming bandwidth is unlimited

But, limited pin b/w favors fewer cores + more cache



- For clarity, only showing two bandwidth lines
- Where would the best performance be?

Peak Performing with Conventional Memory



44 cores @ 2.67 GHz
128 MB

Optimaler Takt

Nur Flächenbeschränkung

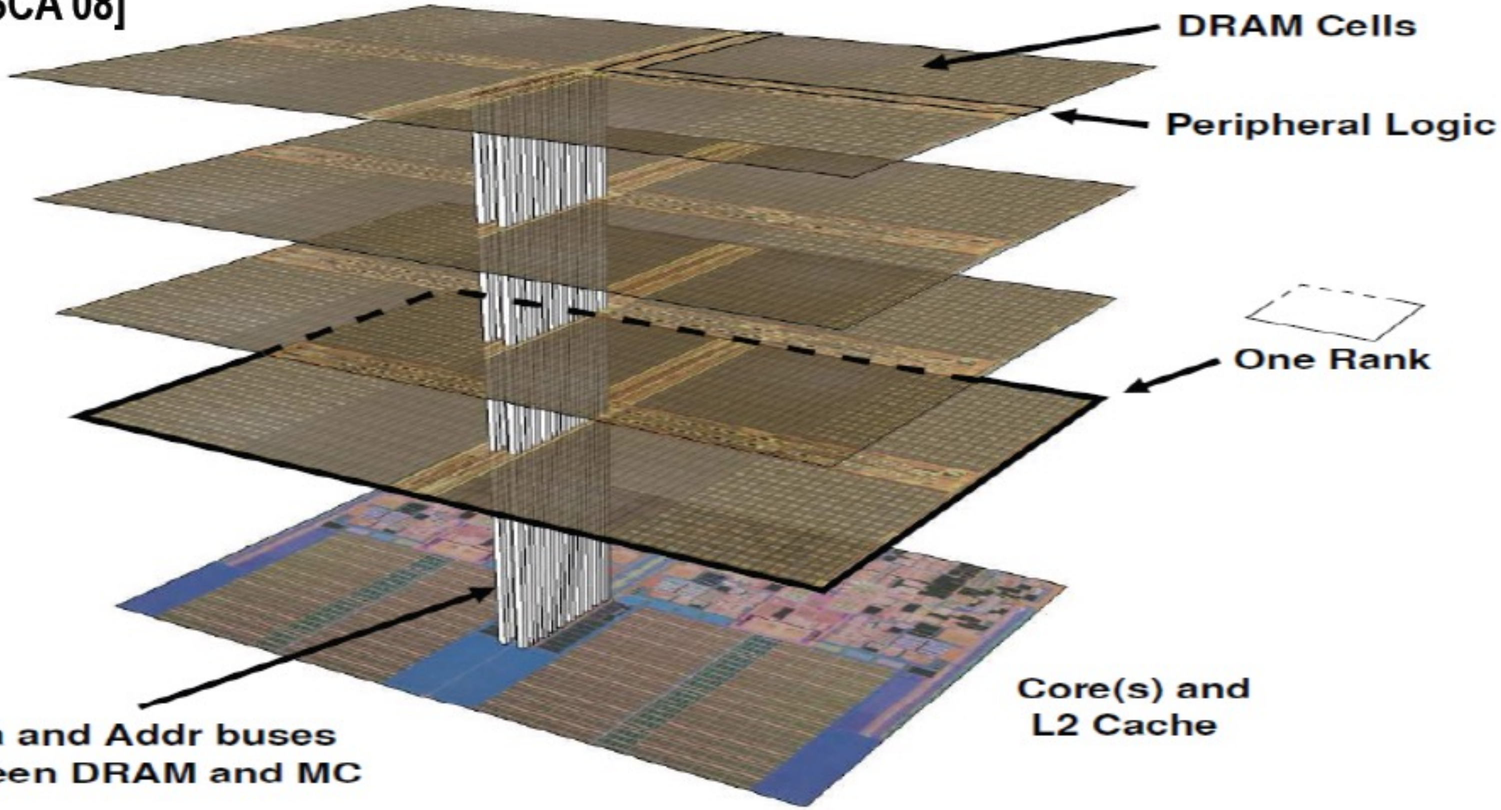
Bandbreiten-grenze

Taktung mit 10 GHz

- B/W constrained, then power constrained
- Fewer slower cores, lots of cache

Mitigating B/W Limitations: 3D-stacked Memory

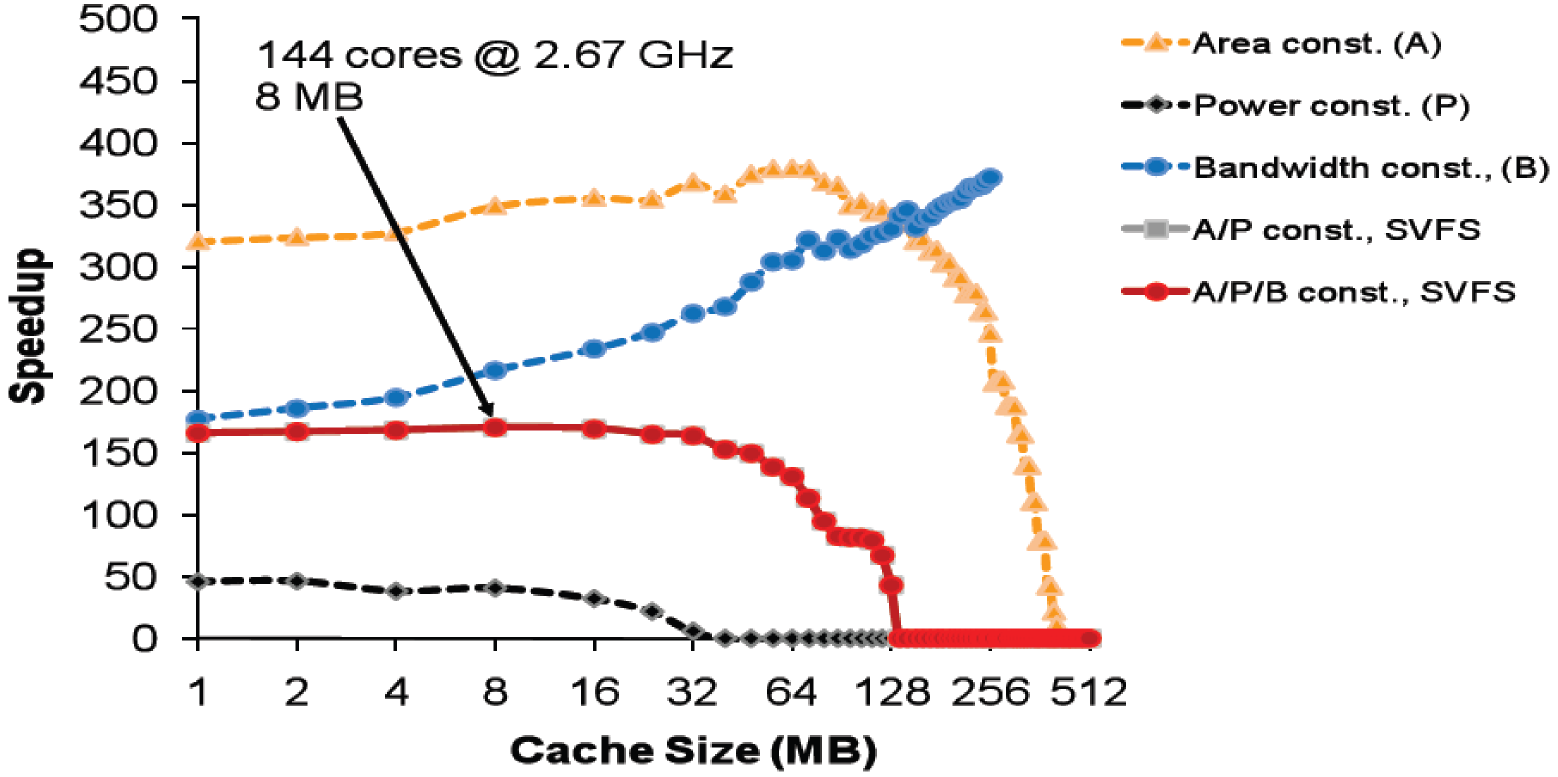
[Loh, ISCA'08]



- Delivers TB/sec of bandwidth

© 2010 Babak Falsafi

Peak Performing w/ 3D-stacked Memory



- Only power-constrained
- Virtually eliminates on-chip cache

© 2010 Babak Falsafi

Long-term:
Where to go from here?

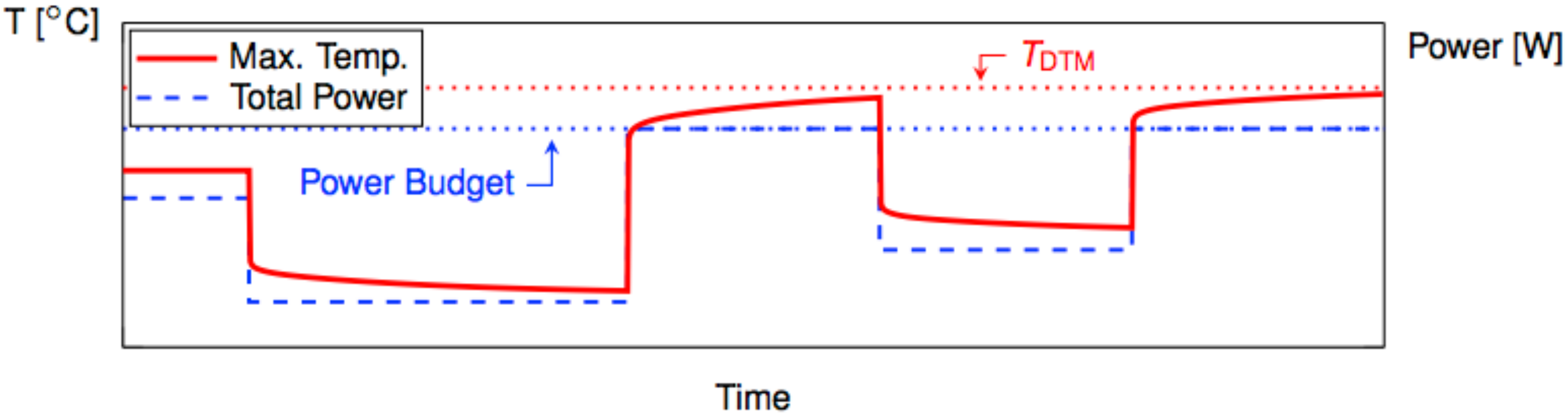
1. Redo SW stack
   - ❑ Minimize joules/work (algo. down to HW)
   - ❑ Program for locality + heterogeneity

2. Pray for technology
   - ❑ Energy-scalable silicon devices
   - ❑ Emerging nanoscale technologies?

3. Infrastructure technology
   - ❑ Renewable/carbon-neutral energy
   - ❑ Scalable cooling + power delivery

Power Budget / Power Constraint

- Abstraction: Not deal directly with temperature.
- Generally, a power budget (for thermal safety) is a single value:
  - For each core (per-core).
  - For the entire chip (per-chip).

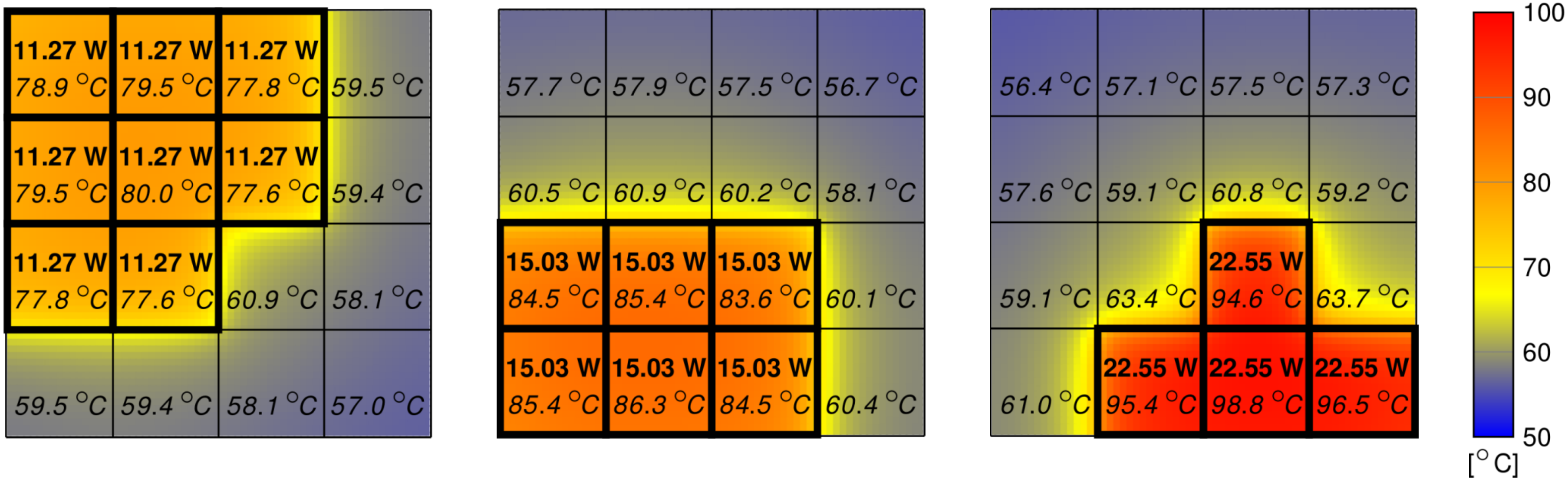


Per-Chip / Per-Core Power Budgets

16 cores with area 5.3 mm^2

Threshold temperature: 80°C

Power budget: 90 W



11.27 W 78.9 °C	**11.27 W** 79.5 °C	**11.27 W** 77.8 °C	59.5 °C
11.27 W 79.5 °C	**11.27 W** 80.0 °C	**11.27 W** 77.6 °C	59.4 °C
11.27 W 77.8 °C	**11.27 W** 77.6 °C	60.9 °C	58.1 °C
59.5 °C	59.4 °C	58.1 °C	57.0 °C

Highest Temperature: 80.0° C

8 active cores

57.7 °C	57.9 °C	57.5 °C	56.7 °C
60.5 °C	60.9 °C	60.2 °C	58.1 °C
15.03 W 84.5 °C	**15.03 W** 85.4 °C	**15.03 W** 83.6 °C	60.1 °C
15.03 W 85.4 °C	**15.03 W** 86.3 °C	**15.03 W** 84.5 °C	60.4 °C

Highest Temperature: 86.3° C

6 active cores

56.4 °C	57.1 °C	57.5 °C	57.3 °C
57.6 °C	59.1 °C	60.8 °C	59.2 °C
59.1 °C	63.4 °C	**22.55 W** 94.6 °C	63.7 °C
61.0 °C	**22.55 W** 95.4 °C	**22.55 W** 98.8 °C	**22.55 W** 96.5 °C

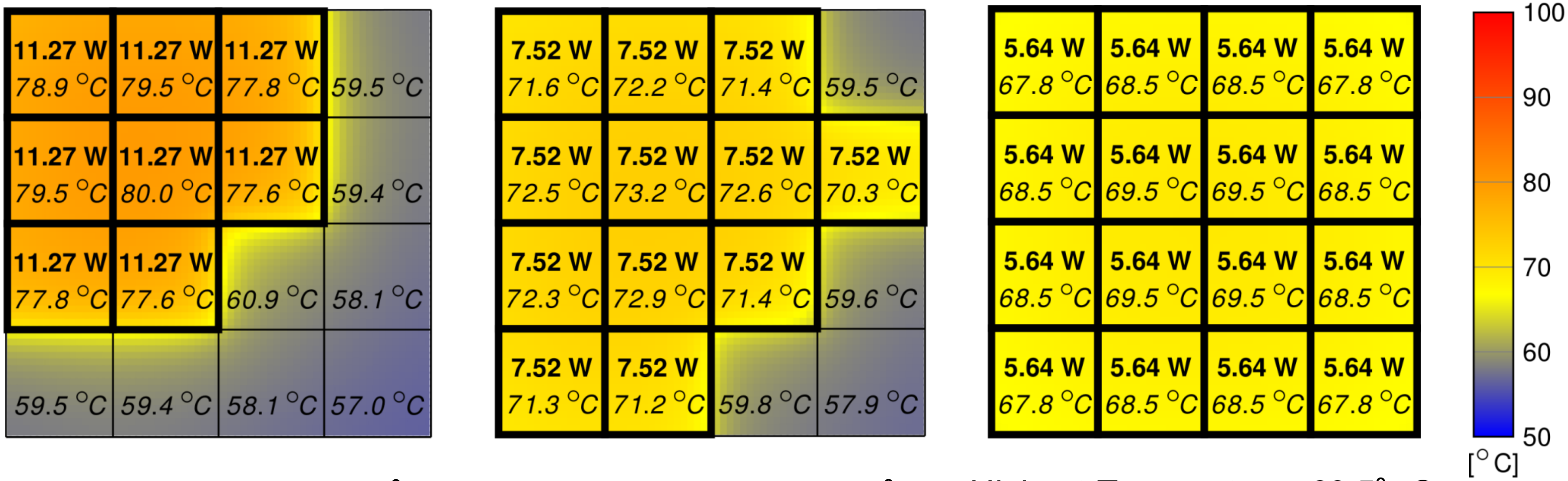
Highest Temperature: 98.8C

4 active cores

Pagani et al., CODES+ISSS 2014

Per-Chip / Per-Core Power Budgets

16 cores with area 5.3 mm^2
Threshold temperature: 80°C
Power budget: 90 W



11.27 W 78.9 °C	11.27 W 79.5 °C	11.27 W 77.8 °C	59.5 °C
11.27 W 79.5 °C	11.27 W 80.0 °C	11.27 W 77.6 °C	59.4 °C
11.27 W 77.8 °C	11.27 W 77.6 °C	60.9 °C	58.1 °C
59.5 °C	59.4 °C	58.1 °C	57.0 °C

Highest Temperature: 80.0° C

8 active cores

7.52 W 71.6 °C	7.52 W 72.2 °C	7.52 W 71.4 °C	59.5 °C
7.52 W 72.5 °C	7.52 W 73.2 °C	7.52 W 72.6 °C	7.52 W 70.3 °C
7.52 W 72.3 °C	7.52 W 72.9 °C	7.52 W 71.4 °C	59.6 °C
7.52 W 71.3 °C	7.52 W 71.2 °C	59.8 °C	57.9 °C

Highest Temperature: 73.2° C

12 active cores

5.64 W 67.8 °C	5.64 W 68.5 °C	5.64 W 68.5 °C	5.64 W 67.8 °C
5.64 W 68.5 °C	5.64 W 69.5 °C	5.64 W 69.5 °C	5.64 W 68.5 °C
5.64 W 68.5 °C	5.64 W 69.5 °C	5.64 W 69.5 °C	5.64 W 68.5 °C
5.64 W 67.8 °C	5.64 W 68.5 °C	5.64 W 68.5 °C	5.64 W 67.8 °C

Highest Temperature: 69.5° C

16 active cores

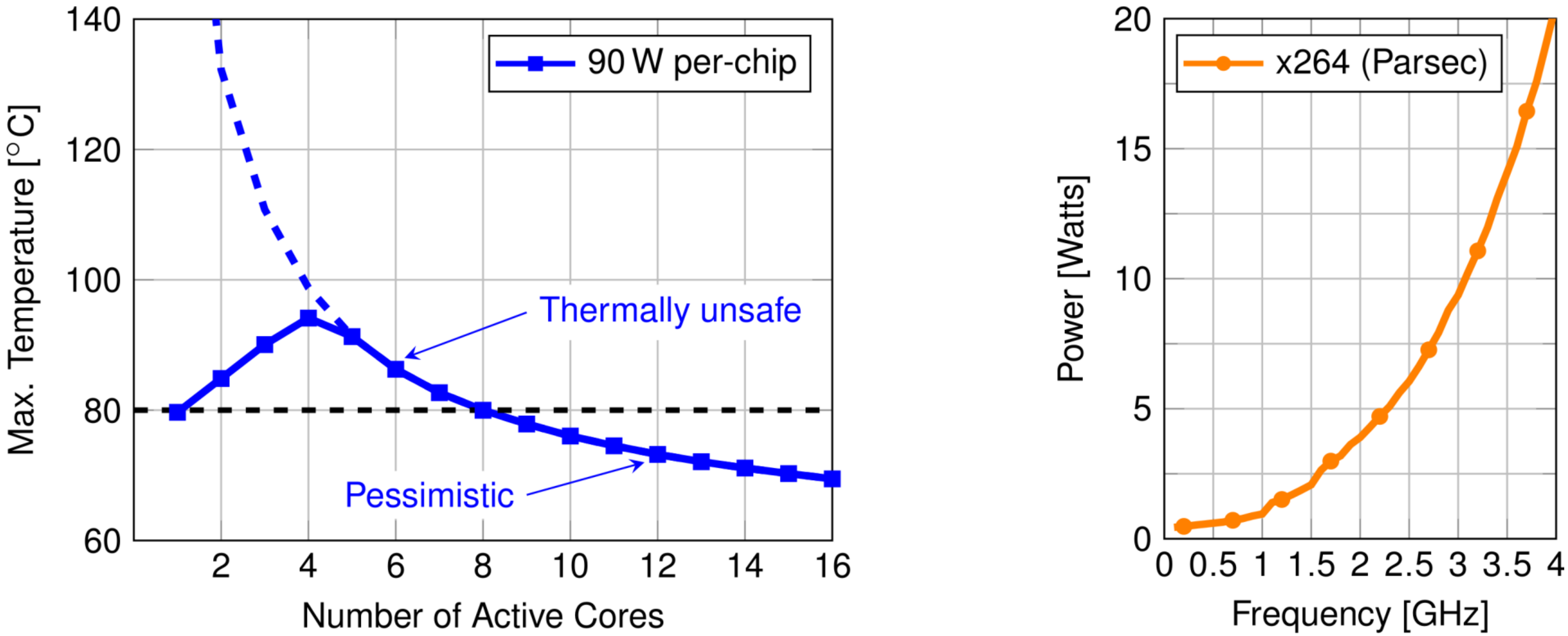
Pagani et al., CODES+ISSS 2014

Problem with Per-Chip / Per-Core Power Budgets

16 cores with area 5.3 mm^2

Threshold temperature for DTM: 80°C

Power budget: 90 W



Pagani et al.,
CODES+ISSS 2014

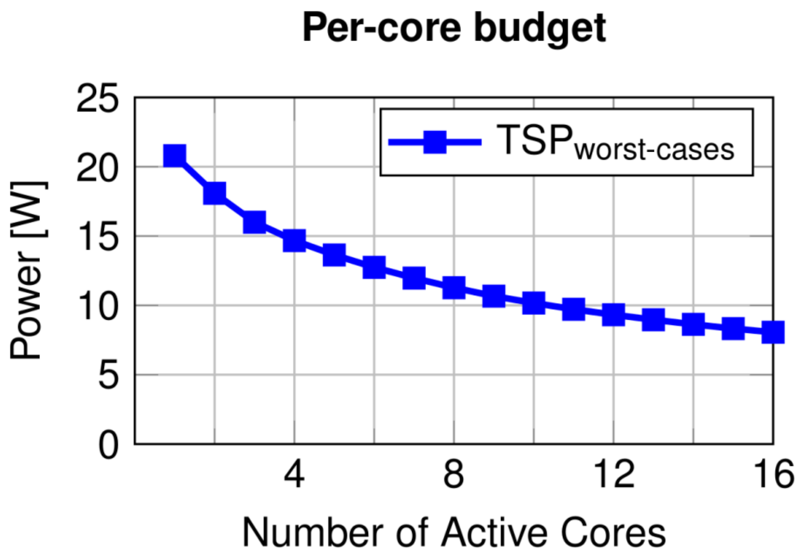
Thermal Safe Power (TSP): Power Budget depending on # of activated cores

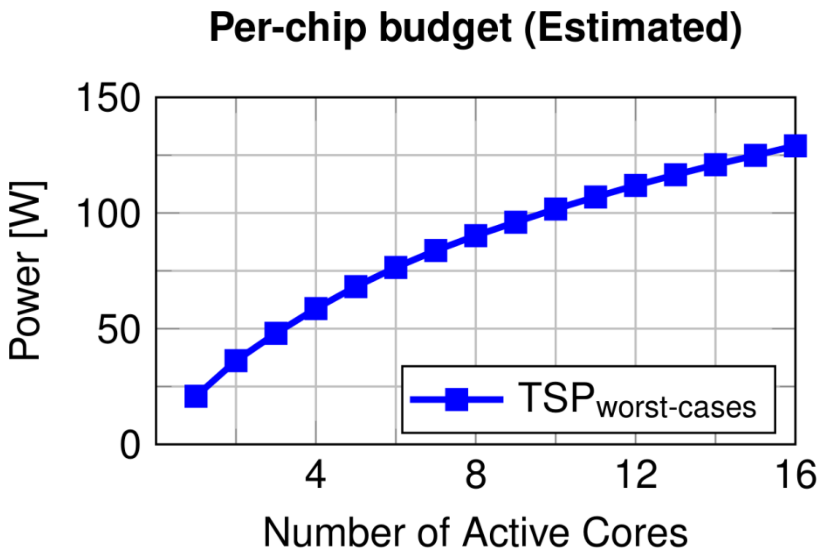
Power budget depends on the number of active cores

Safe for **any** ‘*m*’ active cores => Abstract mapping decisions

TSP table:

Active Cores	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
TSP per-core [W]	20.79	18.08	16.00	14.67	13.64	12.74	11.97	11.27	10.67	10.17	9.72	9.33	8.96	8.63	8.33	8.06





Pagani et al., CODES+ISSS 2014

Cooling or Darkening Matters

- Thermoelektrische Kühlung
- Wasserkühlung
- Kältekühlung
- usw.