# Research Challenges for the Architecture Community in Temperature-Aware Design

Ronny Ronen
Senior Principal Engineer
Director of Architecture Research
Intel Labs - Haifa

Intel Corporation

*The views presented here are mine. They do not necessarily reflect Intel corp. position.*

TACS-1:
First Workshop on Temperature-Aware Computer Systems (w/ ISCA '2004)

# General trends

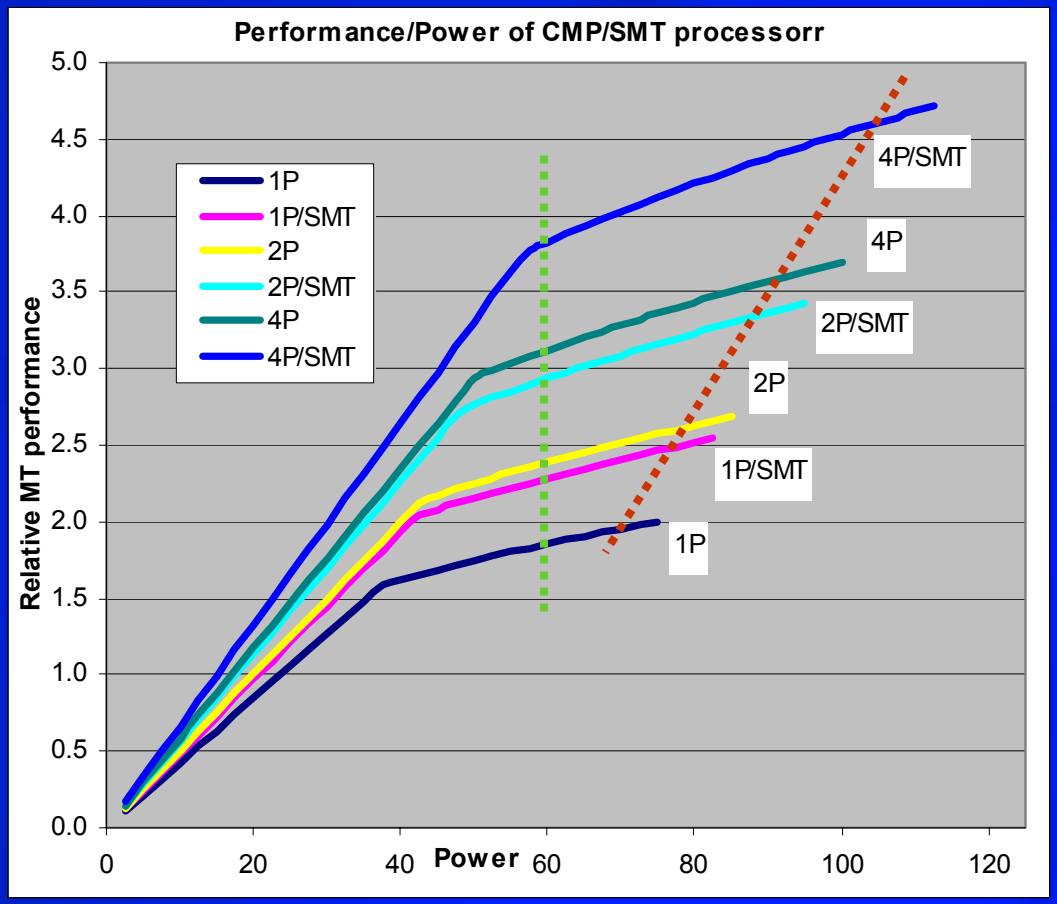
Process Technology:

- ☺ Smaller (2x), faster (1.4x) transistors, consuming less energy per switch (2x) – but:
- ☹ More transistors per $mm^2$, More switches per cycle ➔ More power density (~1.5X)

Micro-architecture:

- ☹ Traditional - More transistor switches per inst.
- 😐 New world - More efficiency ➔ More activity per area
- ➔ Higher power, higher power density
- ➔ Higher temperature

# CMP/SMT – performance/power trends
## With DVS




**Configurations**

- 1/2/4 cores
- 1/2 Way SMT
- Fixed area: 100 mm2
- Using Dynamic Voltage Scaling from 0 to Vmax

- Efficiency is up (performance per fixed power)
- Overall power increased
- Power density increased ➔ Temperature

Source – "Everything you always wanted to know about SCALABILITY and were afraid to ask". Ronny Ronen, Intel EMEA Academic Forum, Barcelona, 4/2004

# Observations

- Processors are becoming thermally limited –
  … and it will just get worse.
- Thermal solutions improve –
  But cost and form factor limit their benefit
- Designing for the worst case is not enough
  - Too risky
  - Too inefficient
- Temperature awareness is becoming crucial
  - This has already started – e.g., Thermal throttling

# Challenges

- Improve measurements
  - To know better where we are
- Reduce Temperature
  - To actively eliminate/relax heating
- Improve control
  - To get the maximum in a given environment

# Challenges – improve Measurements

## Simulation

- Better, more accurate, thermal estimate
  - More detailed and accurate  power modeling
  - Integrated active/leakage with thermal modeling
  - Accurate thermal solutions

## On chip

- Temperature measurement on die.
  - More accurate, more points on chip
  - Targeting hot spots

# Challenges – Reduce Temperature

Power Awareness

- *"Less is More"*
  Do the same "work" with less power
  - Less instruction per task
  - Less micro-operation per instruction
  - Less transistor switches per micro-operation
  - Less energy per transistor switch
- Better power/performance trade off
  - Optimal/adaptive structure sizes, …

Reduce power density

- Identify hot spots – distribute them as possible

Better cooling solutions

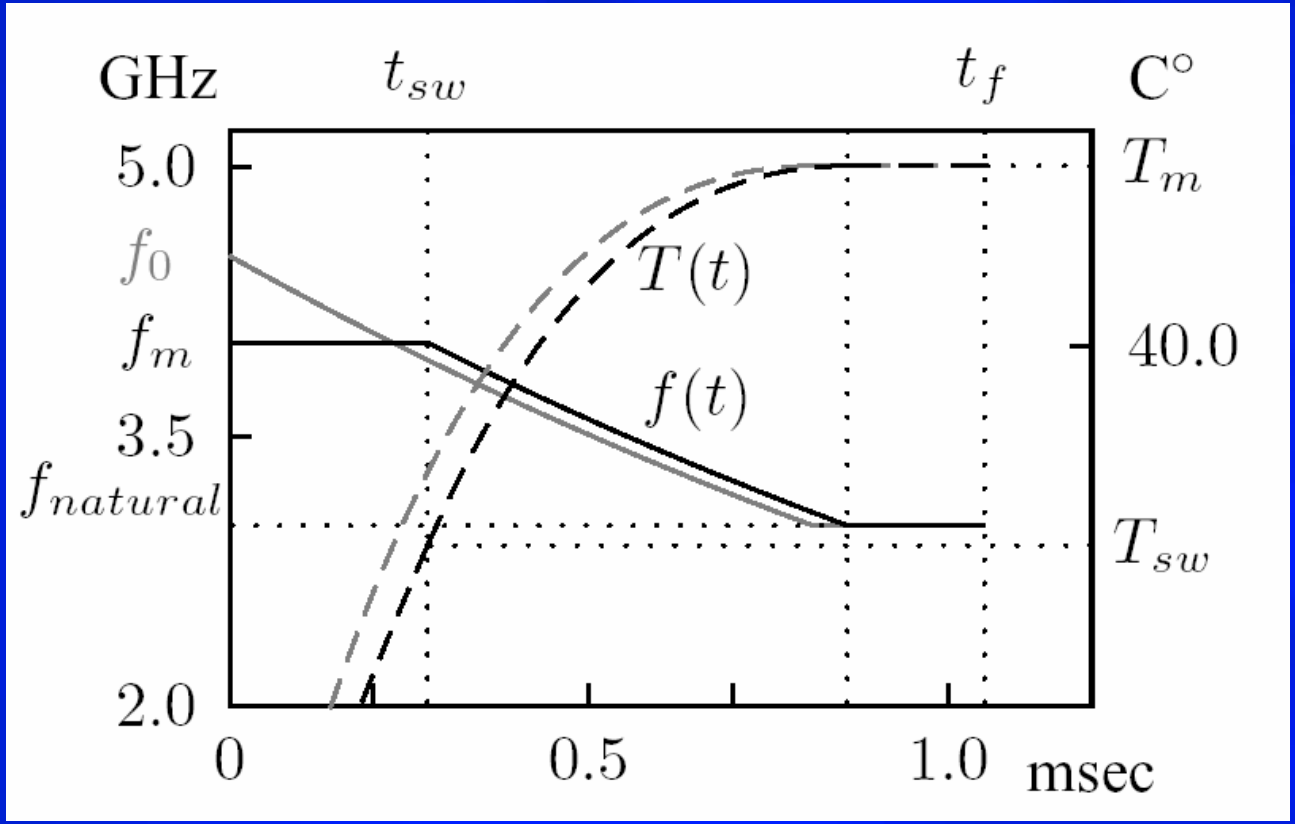
# Challenges – Improve Control

## Adaptive Thermal management

- Eliminate risk of thermal run-away
  - Already being done in today's processor
- Maximize performance in thermally limited environment
  - Technology in infancy
  - Lack of determinism is a concern for OEMs*
- This is an optimization problem!
  - How to control power ➔ thermal for best performance
  - Current Challenges: SMT/MP.

* See http://www.vanshardware.com/articles/2004/05/040517_efficeonFreeze/040517_efficeonFreeze.htm

# Improved Control - example

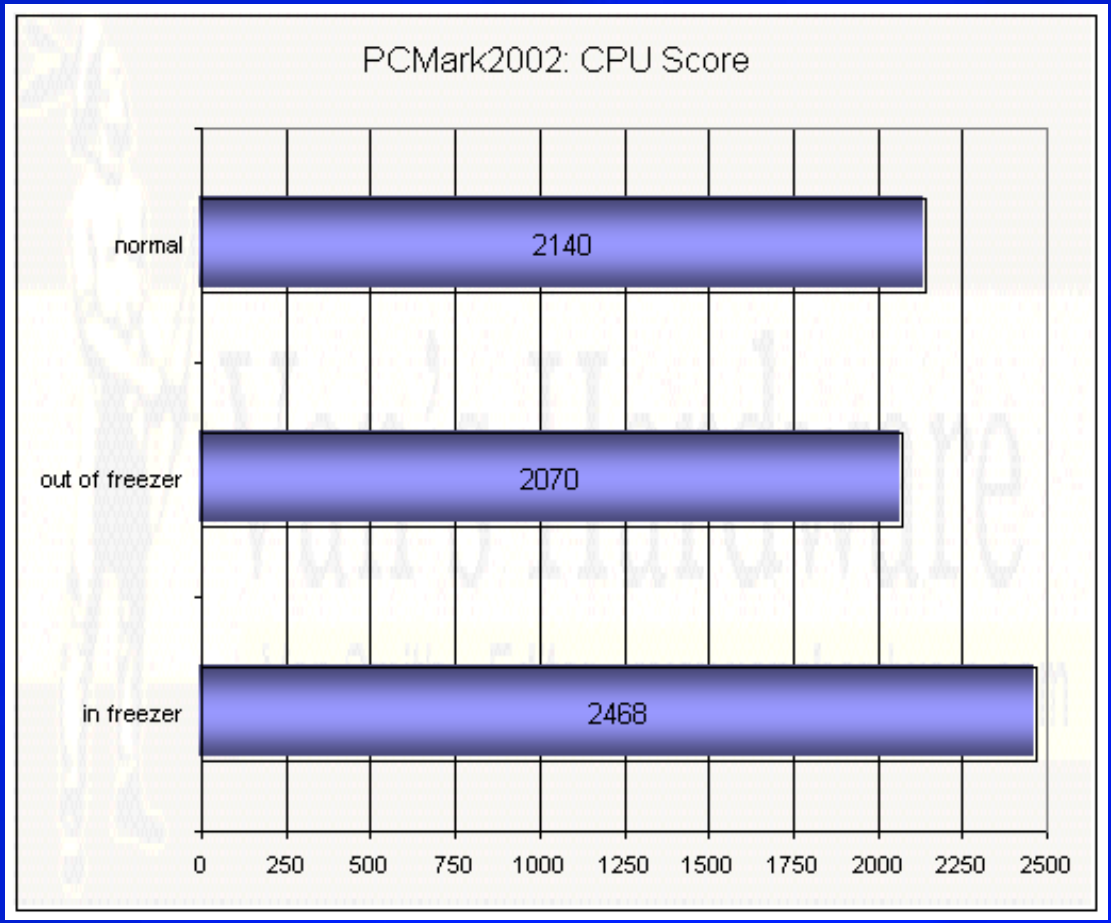



## Optimal strategy for a single-RC model[1]

[1]From Cohen et. al, "On Estimating Optimal Performance of CPU Dynamic Thermal Management." Computer Architecture Letters, Volume 2, Oct. 2003

# Backup

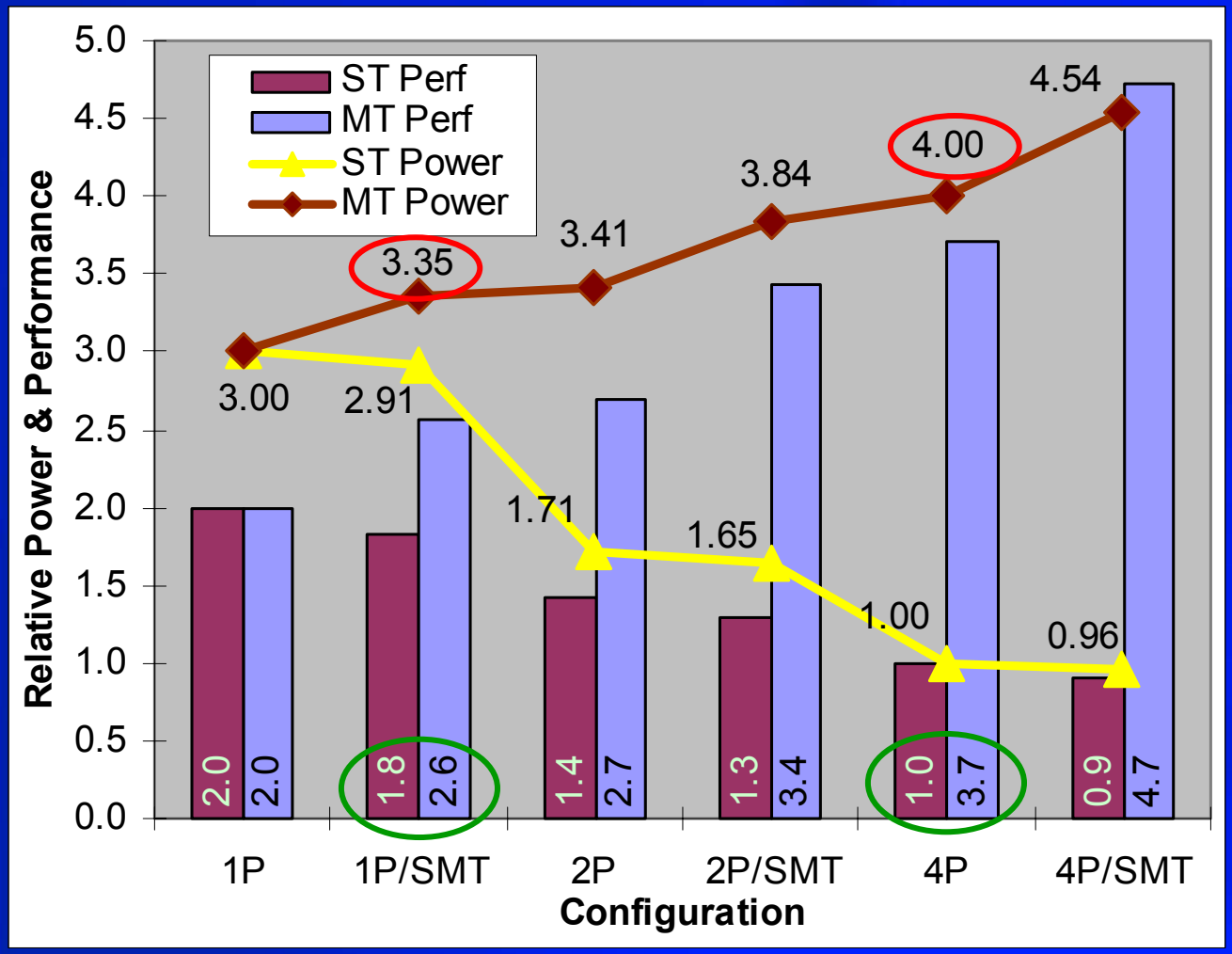
# Deepfreezing a Transmeta efficeon*




**Quote: 'Conclusion: Just Say "No" to Thermal Throttling!'**

* See http://www.vanshardware.com/articles/2004/05/040517_efficeonFreeze/040517_efficeonFreeze.htm

# CMP/SMT – performance/power trends




- Configurations
  - 1/2/4 cores
  - 1/2 Way SMT
  - Fixed area: 100 $mm^2$
  - Unlimited power
- With CMP
  - MT Perf goes up
  - ST Perf goes down
  - Power efficiency is up
  - Power goes up
    → Thermal goes up!
- With SMT
  - MT Perf goes up
  - ST perf stay same
  - Power efficiency is up
  - Power goes up
    → Thermal goes up!

Source - Everything you always wanted to know about SCALABILITY and were afraid to ask. Ronny Ronen, Intel EMEA Academic Forum, Barcelona, 4/2004