# Computer Architecture in an Era of Multi-Core Chips

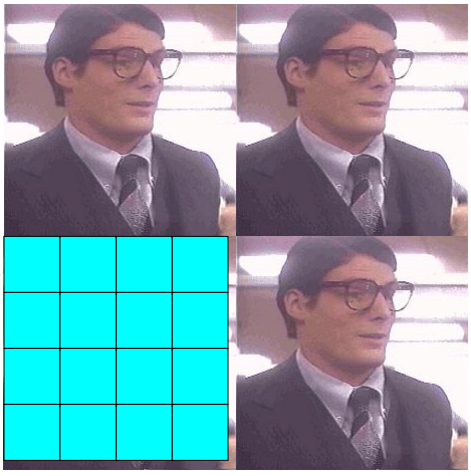
**Kevin Skadron**

**Univ. of Virginia**
**Dept. of Computer Science**
**LAVA Lab**



## A New Era of Multi-Core Architectures

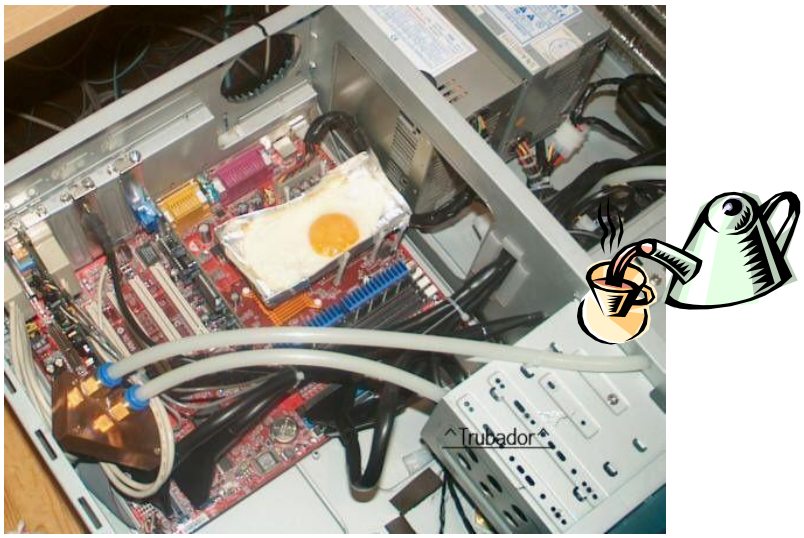
vs.



Source: Chrostopher Reeve Homepage, http://www.chrisreevehomepage.com/

Cores may also be heterogeneous, with a few powerful cores and very many small cores

# A New Era of Physical Constraints




- Will limit integration, system architecture



3

# Impact of Physical Constraints

- Thermal constraints shift optimum toward fewer and simpler cores ("Clark Kents")
  - Actually CPU-bound programs still want aggressive superscalar cores, minimal L2, despite throttling
  - Mem-bound programs want simpler cores, lots of L2
- Thermal constraints subsume power-delivery, maybe even pin-bandwidth constraints
- You can still have lots of cores
  - But they will be simpler
  - And they will be severely throttled (e.g., at 50-75% of max frequency)



4

## How many cores?

- Depends heavily on workload. Today it would be hard to fill up more than a few cores on a desktop except with games.
- Servers: easy – one thread per request, highly parallel
- *…It is a big open question*
  - Maybe rich multimedia tasks, e.g. videoconferencing
  - Computational photography?
  - Biomedical and other appliances?





## What kind of cores?

- A few beefy cores? (Pentiums)
- Lots of simple cores? (Maybe good for parallel tasks)
- Special-purpose cores (graphics, multimedia encode/decode, encryption/decryption, signal processing, etc.)
- A mix?
- What about single-thread latency vs. aggregate throughput?

## Putting it all together

- # cores
- Type of cores
- How much cache
- How to interconnect them
- How to meet power-delivery, pin-B/W, and thermal constraints
- How to deal with workloads that have different needs

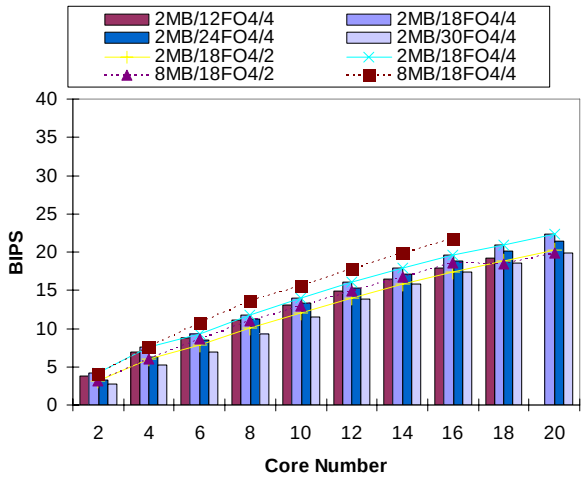




- **Questions?**

# CPU-bound, 400mm²

- Aggressive, expensive thermal solution



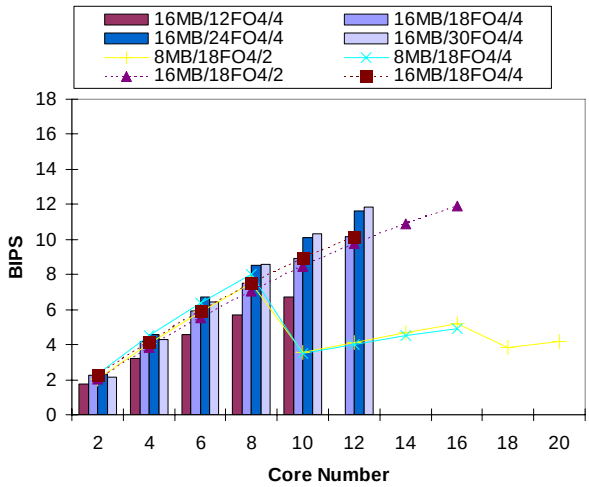

- Thermal limits still cost you 50% performance due to throttling



9

# Memory-bound, 400mm²

- Aggressive, expensive thermal solution



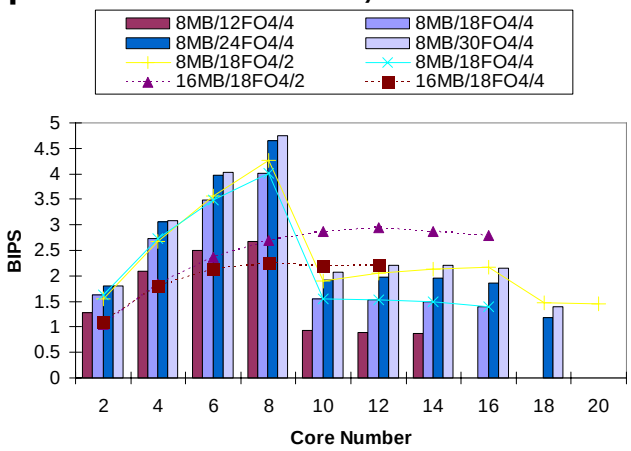

- Thermal limits still costs you 25% performance due to throttling



10

# Memory-bound, 400mm$^2$

- **Cheap, low-end thermal solution**
- **Drop from 16 to 8 cores, 8MB**




- **We are now at 30% of peak possible performance**





# Need for Physically-Aware Design

- **Can't optimize absent thermal constraints and then scale**