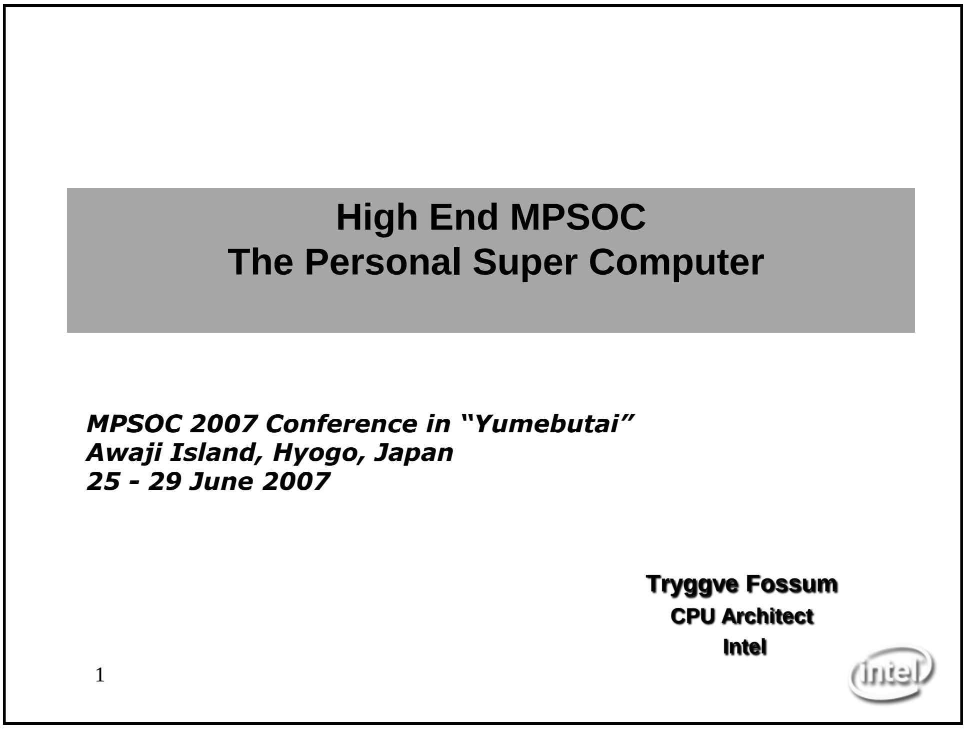
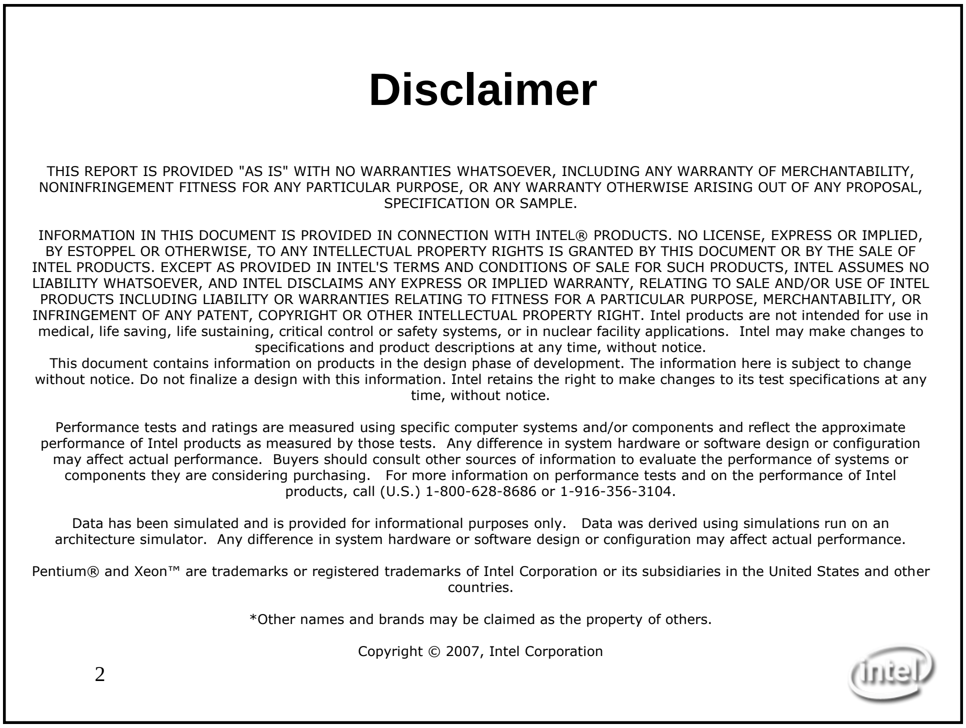


High End MPSOC
The Personal Super Computer

MPSOC 2007 Conference in "Yumebutai"
Awaji Island, Hyogo, Japan
25 - 29 June 2007

Tryggve Fossum
CPU Architect
Intel



Disclaimer

THIS REPORT IS PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, INCLUDING ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT FITNESS FOR ANY PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, SPECIFICATION OR SAMPLE.

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT OR BY THE SALE OF INTEL PRODUCTS. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. Intel products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications. Intel may make changes to specifications and product descriptions at any time, without notice.
This document contains information on products in the design phase of development. The information here is subject to change without notice. Do not finalize a design with this information. Intel retains the right to make changes to its test specifications at any time, without notice.

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, call (U.S.) 1-800-628-8686 or 1-916-356-3104.

Data has been simulated and is provided for informational purposes only. Data was derived using simulations run on an architecture simulator. Any difference in system hardware or software design or configuration may affect actual performance.

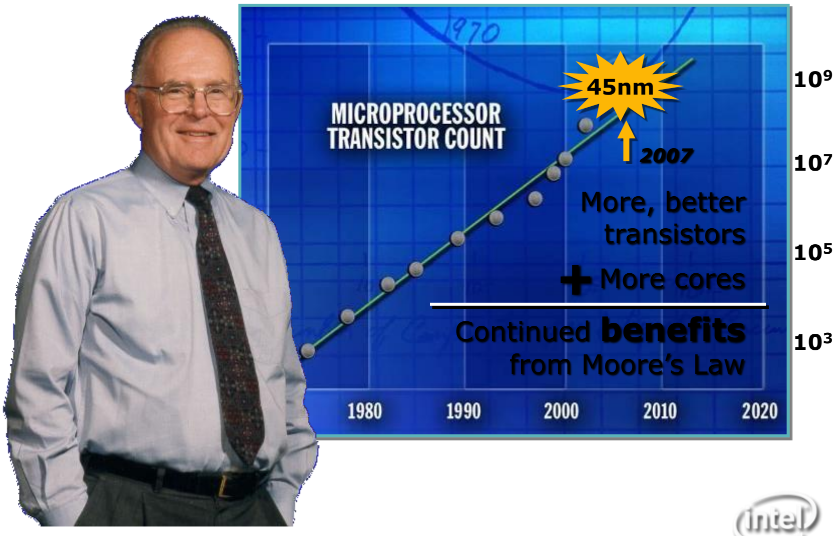
Pentium® and Xeon™ are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

*Other names and brands may be claimed as the property of others.

Agenda

- **Motivation for Chip Level Multiprocessing (CMP)**
- **Success: Moore's Law**
- **Challenges:**
  - Processor Core Design
  - Memory Access
  - Cache Behavior
    - Applications
  - Reliability
  - Power

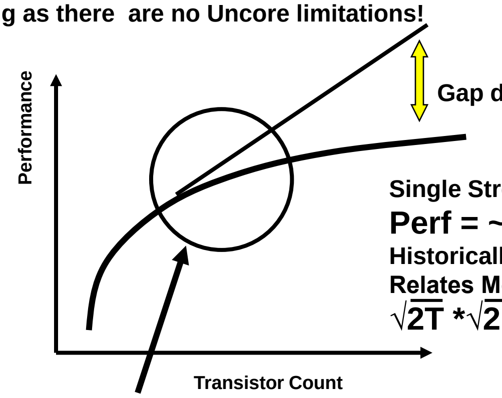
Moore's Law Motivates Multi-Core



Single Stream, Moore's Law, and CMP

CMP Performance: **Performance = ~ k x Transistor Count**

As long as there are no Uncore limitations!



Gap driving us to Multi-core

Single Stream Performance:

Perf = ~ $\sqrt{\text{Transistor Count}}$

Historically fairly accurate

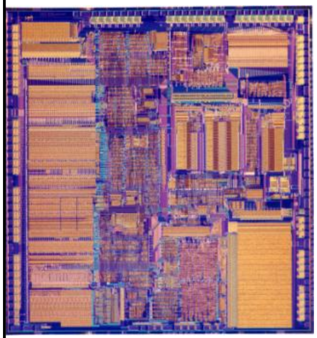
Relates Moore's Law to Performance:

$\sqrt{2T} * \sqrt{2} = 2\sqrt{T}$

Interesting Core Design Area:
Slopes are similar

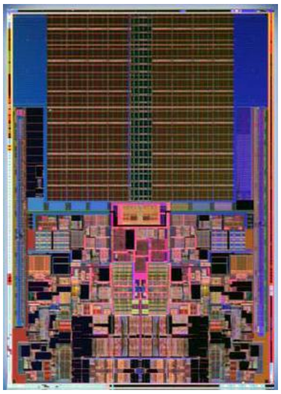
Transistor speedup due to Technology shrink: 0.7

386 Processor



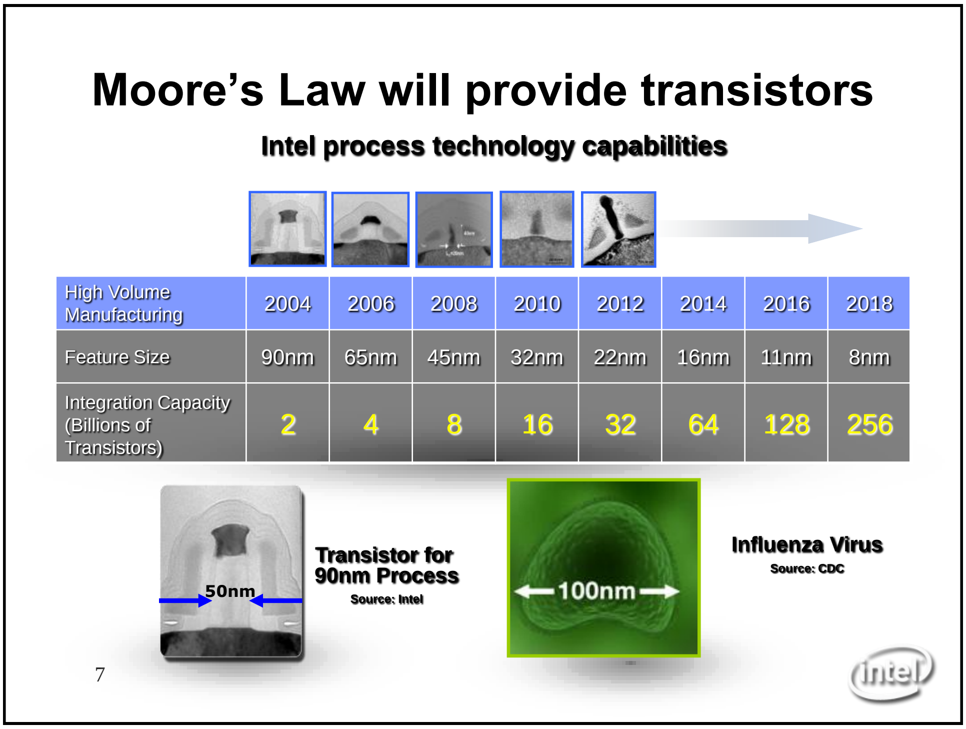
Pentium® 4 Processor

Penryn Processor



May 1986	17 Years	August 27, 2003
@16 MHz core	200x	@3.2 GHz core
275,000 1.5µ transistors	200x/11x	55 Million 0.13µ
~1.2 SPECint2000	1000x	1249 SPECint2000

Coming Soon!
Dual Core
45 nanometer
Large Cache
Energy Efficient



Moore's Law will provide transistors

Intel process technology capabilities

High Volume Manufacturing	2004	2006	2008	2010	2012	2014	2016	2018
Feature Size	90nm	65nm	45nm	32nm	22nm	16nm	11nm	8nm
Integration Capacity (Billions of Transistors)	2	4	8	16	32	64	128	256

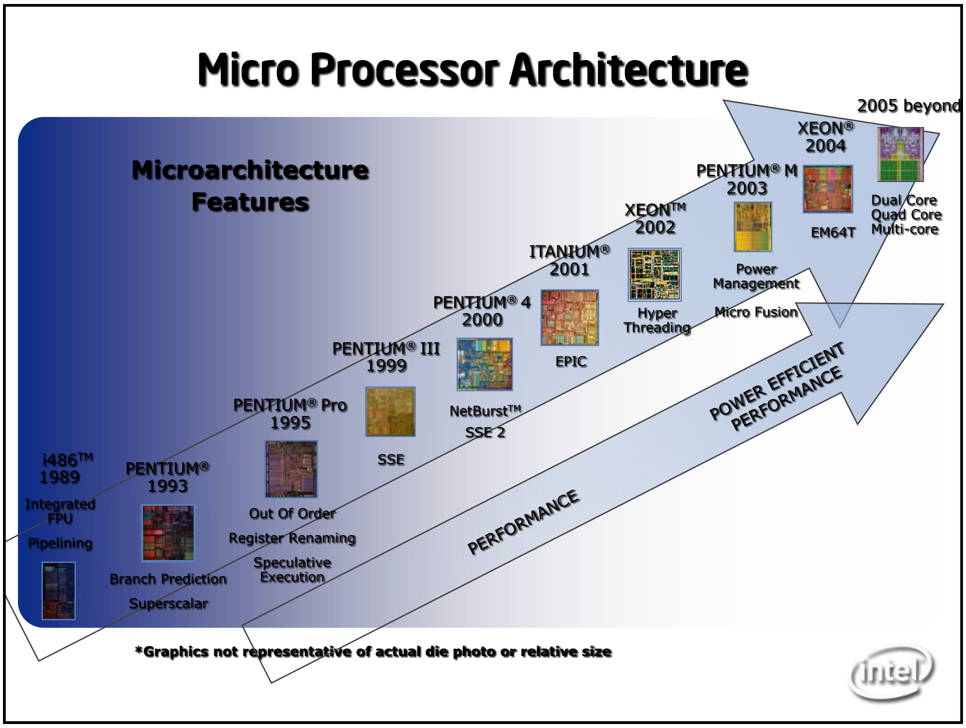
Transistor for 90nm Process

Source: Intel

Influenza Virus

Source: CDC

7



Micro Processor Architecture

Microarchitecture Features

i486™ 1989
Integrated FPU
Pipelining

PENTIUM® 1993
Branch Prediction
Superscalar

PENTIUM® Pro 1995
Out Of Order
Register Renaming
Speculative Execution

PENTIUM® III 1999
SSE

PENTIUM® 4 2000
NetBurst™
SSE 2

ITANIUM® 2001
EPIC

XEON™ 2002
Hyper Threading

PENTIUM® M 2003
Power Management
Micro Fusion

XEON® 2004
EM64T

2005 beyond
Dual Core
Quad Core
Multi-core

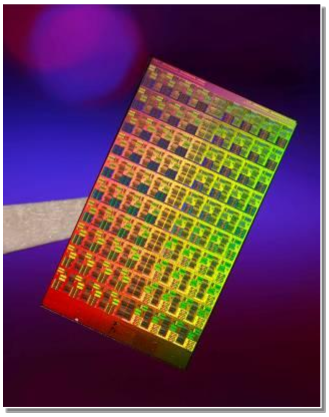
PERFORMANCE

POWER EFFICIENT PERFORMANCE

*Graphics not representative of actual die photo or relative size

Teraflops Research Chip

100 Million Transistors ● 80 Tiles ● 275mm^2



First tera-scale programmable silicon

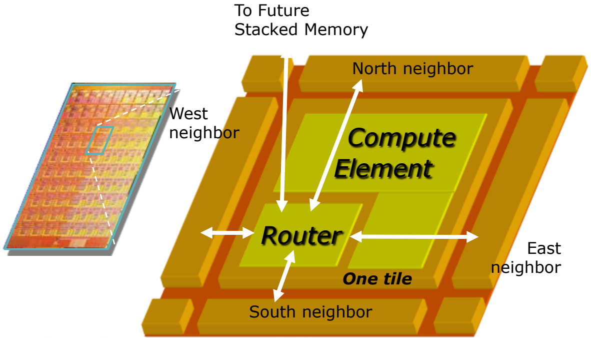
- Teraflops performance
- Tile design approach
- On-die mesh network
- Novel clocking
- Power-aware capability
- Supports 3D-memory

Not designed for IA or product

Tiled Design & Mesh Network

Repeated Tile Method:

- Compute + router
- Modular, scalable
- Small design teams
- Short design cycle



Mesh Interconnect:

- “Network-on-a-Chip”
  - Cores networked in a grid allows for super high bandwidth communications in and between cores
- 5-port, 80GB/s* routers
- Low latency (1.25ns*)
- Future: connect IA/or and special purpose cores

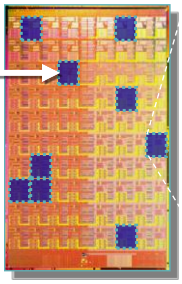
* When operating at a nominal speed of 4GHz

Fine Grain Power Management

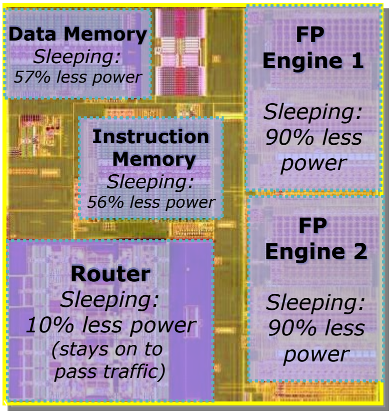
- Novel, modular clocking scheme saves power over global clock
- New instructions to make any core sleep or wake as apps demand
- Chip Voltage & freq. control (0.7-1.3V, 0-5.8GHz)

Dynamic sleep

STANDBY:
• Memory retains data
• **50%** less power/tile
FULL SLEEP:
•Memories fully off
•**80%** less power/tile



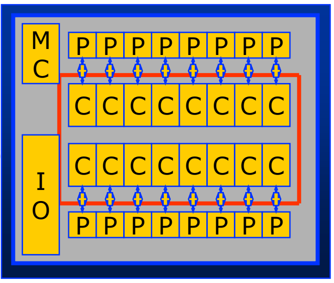
21 sleep regions per tile *(not all shown)*



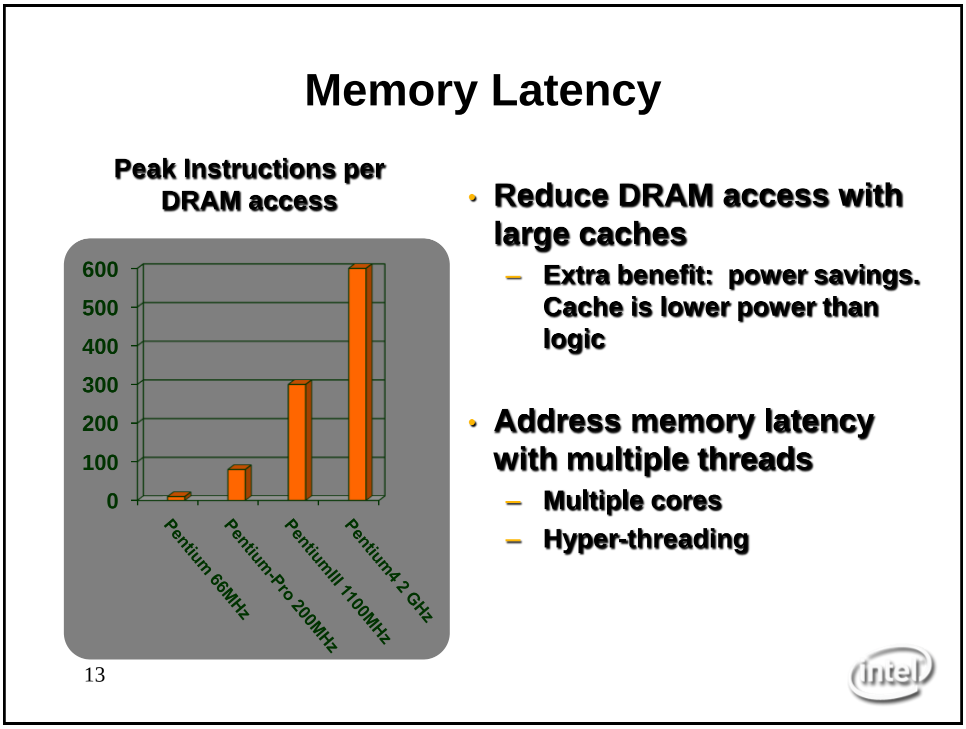
Industry leading energy-efficiency of 16 Gigaflops/Watt

Multi Core System Benefits

- **Performance scaling:**
  - On die interconnect:
    - Higher Bandwidth --- TB/sec vs GB's/sec
    - Shorter Latency --- ns's vs. 100 ns
    - Fast Communication with Shared cache
    - Better cache Hit rate
    - Fast Synchronization --- Locks and Barriers
    - Reduced false sharing
  - Memory
    - Simplifies System Design
    - Reduces NUMA effects
    - Simplifies performance tuning
    - Simplifies application development
    - Enables Fine Grained Parallelism



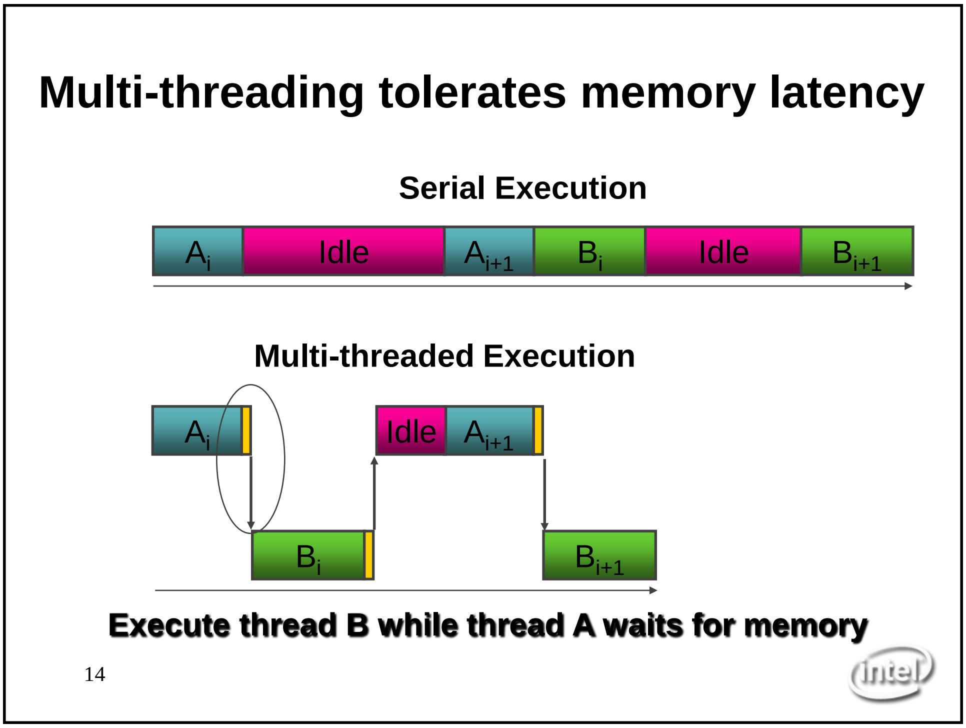
On-die performance can grow almost linearly with core count!



Memory Latency

Peak Instructions per DRAM access

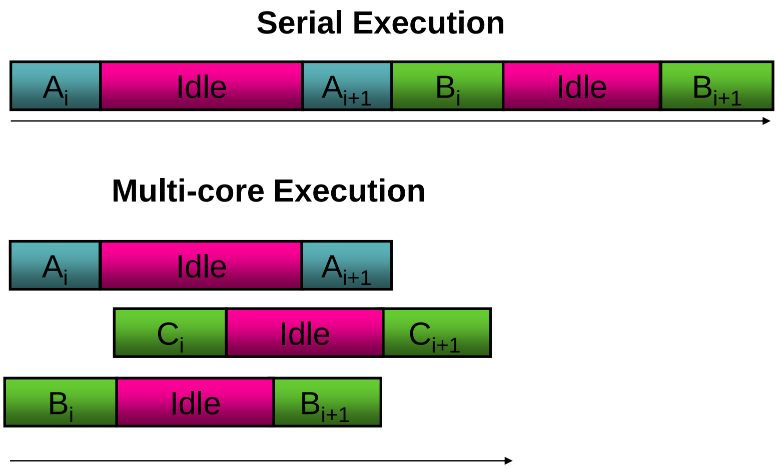
- **Reduce DRAM access with large caches**
  - **Extra benefit: power savings. Cache is lower power than logic**
- **Address memory latency with multiple threads**
  - **Multiple cores**
  - **Hyper-threading**



Multi-threading tolerates memory latency

Execute thread B while thread A waits for memory

Multi-core parallelizes memory access



Serial Execution

A_i | Idle | A_{i+1} | B_i | Idle | B_{i+1}

Multi-core Execution

A_i | Idle | A_{i+1}

C_i | Idle | C_{i+1}

B_i | Idle | B_{i+1}

15 **Execute thread A, B and C simultaneously**

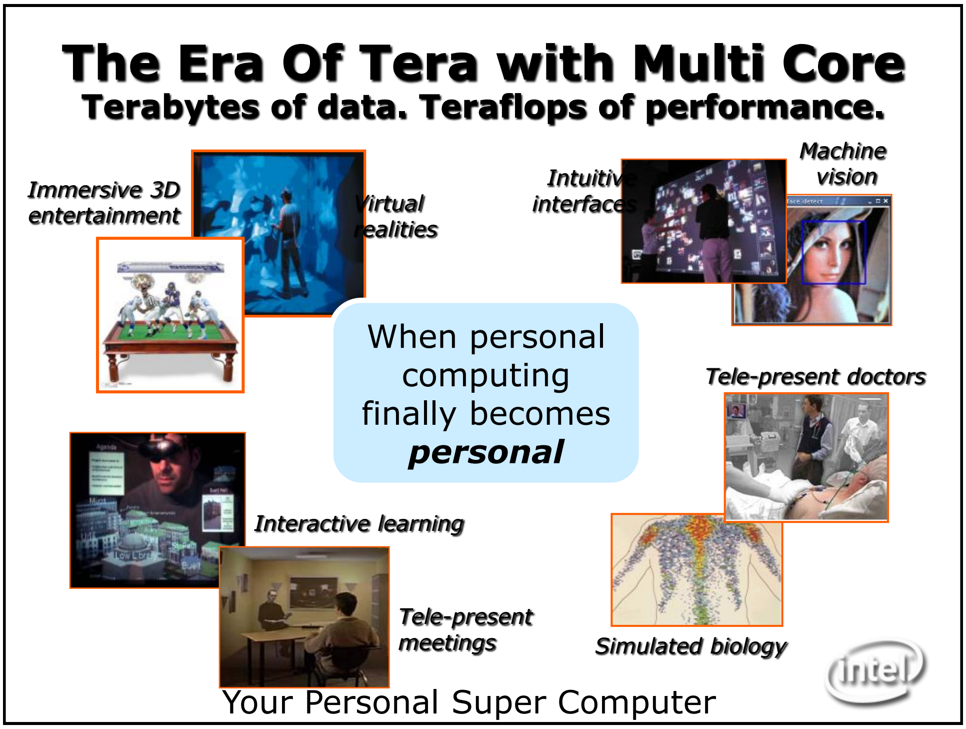
Cache Design Issues

Cache Design for Parallel Programming

- **Transactional Memory most visible feature**
  - **Simplifies parallel programming**
- **Thread Communication and Synchronization – take advantage of on-die latencies and bandwidth**
- **Support for fine-grained parallelism**
- **Debug Hooks, Performance Monitoring, Visualization**

Optimize Cache Hierarchy Opportunities for CMP

- **Technology is just now emerging**
  from 1 to 4 levels of cache
- **Keep Bandwidth up and Latency down:**
  Multiple Levels, Sharing, Inclusion, Hybrids, Prefetching, Fairness, Quality of Service, Interconnects, Single Socket
- **Fertile Ground for Research**



The Era Of Tera with Multi Core

Terabytes of data. Teraflops of performance.

Immersive 3D entertainment

Virtual realities

Intuitive interfaces

Machine vision

When personal computing finally becomes ***personal***

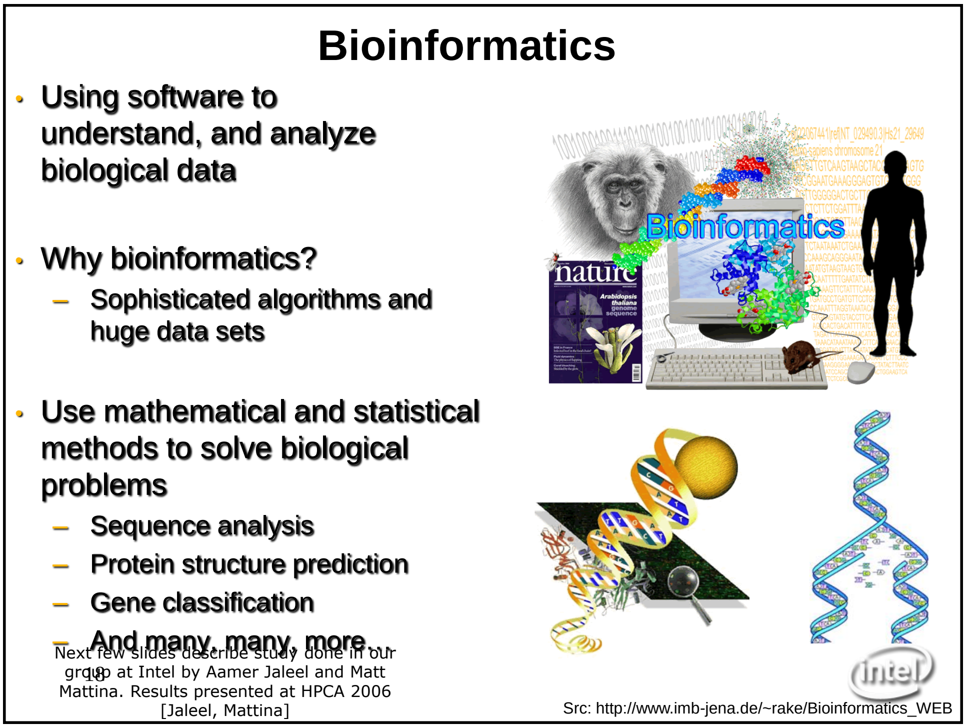
Tele-present doctors

Interactive learning

Tele-present meetings

Simulated biology

Your Personal Super Computer



Bioinformatics

- Using software to understand, and analyze biological data
- Why bioinformatics?
  - Sophisticated algorithms and huge data sets
- Use mathematical and statistical methods to solve biological problems
  - Sequence analysis
  - Protein structure prediction
  - Gene classification
  - And many, many, more…

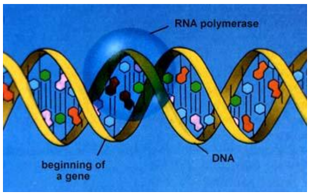
18

Next few slides describe study done in our group at Intel by Aamer Jaleel and Matt Mattina. Results presented at HPCA 2006 [Jaleel, Mattina]

Src: http://www.imb-jena.de/~rake/Bioinformatics_WEB

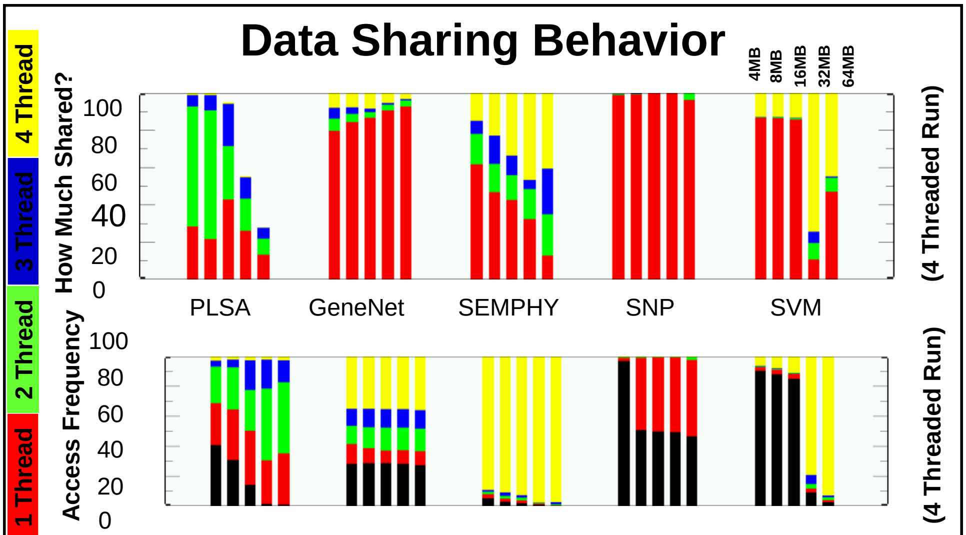
Parallel Bioinformatics Workloads

- Structure Learning:
  - GeneNet – Hill Climbing, Bayesian network learning
  - SNP – Hill Climbing, Bayesian network learning
  - SEMPHY – Structural Expectation Maximization algorithm
- Optimization:
  - PLSA – Dynamic Programming
- Recognition:
  - SVM-RFE – Feature Selection
- OpenMP workloads developed by Intel Corporation
  - Now part of Northwestern University, NU-MineBench Suite
  - http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html
  - Also made available at: http://www.ece.umd.edu/biobench/



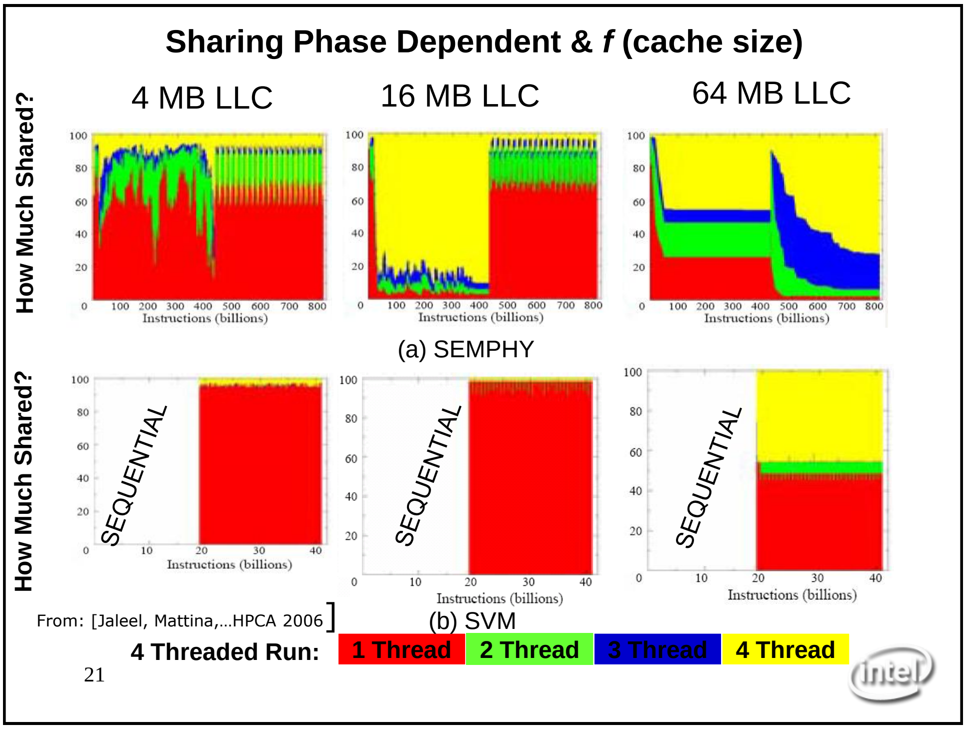
From: [Jaleel, Mattina,…HPCA 2006]

Data Sharing Behavior



- Sharing is dependent on algorithm and varies with cache size
- Workloads fully utilize a 64MB LLC
- Reducing cache misses improves data sharing
- Despite size of shared footprint, shared data frequently referenced

From: [Jaleel, Mattina,…HPCA 2006]



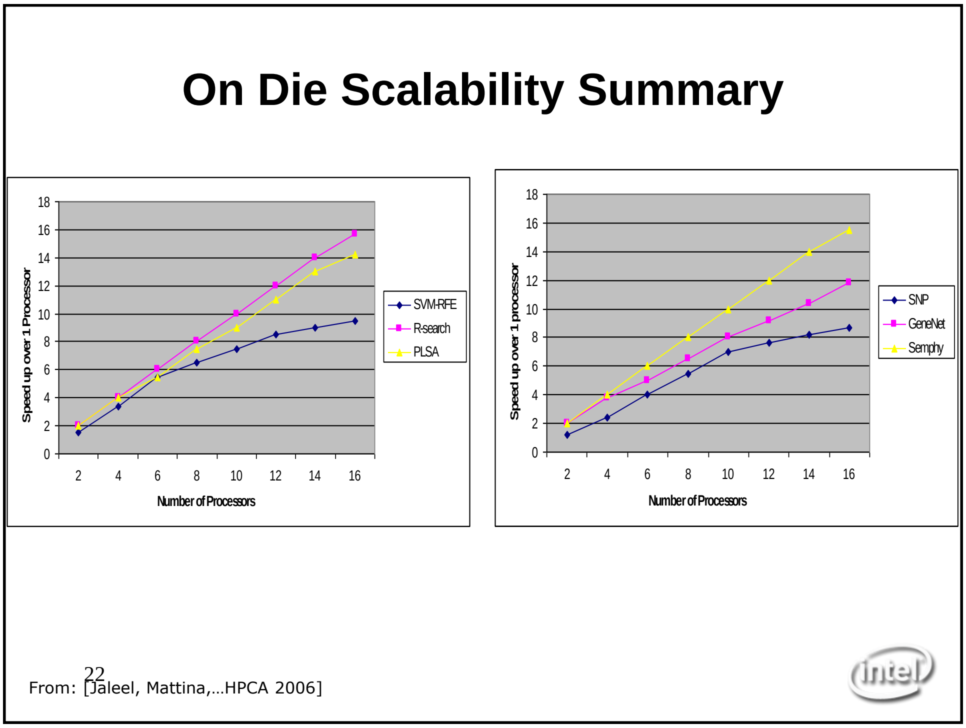
Sharing Phase Dependent & *f* (cache size)

(a) SEMPHY

(b) SVM

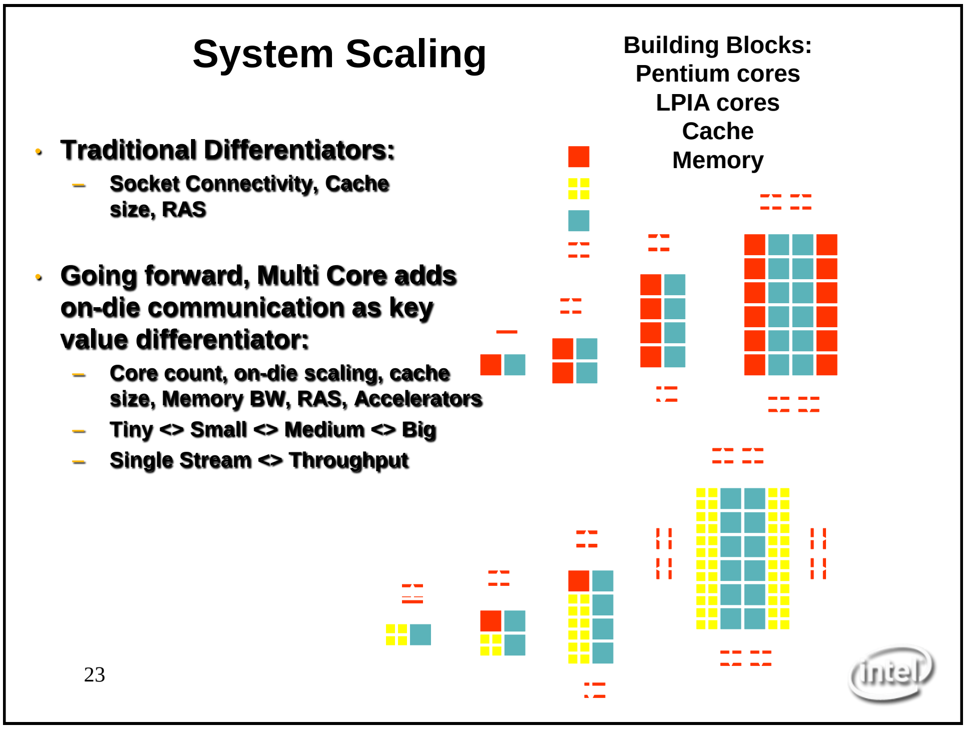
From: [Jaleel, Mattina,…HPCA 2006]

4 Threaded Run: 🟥 **1 Thread** 🟩 **2 Thread** 🟦 **3 Thread** 🟨 **4 Thread**



On Die Scalability Summary

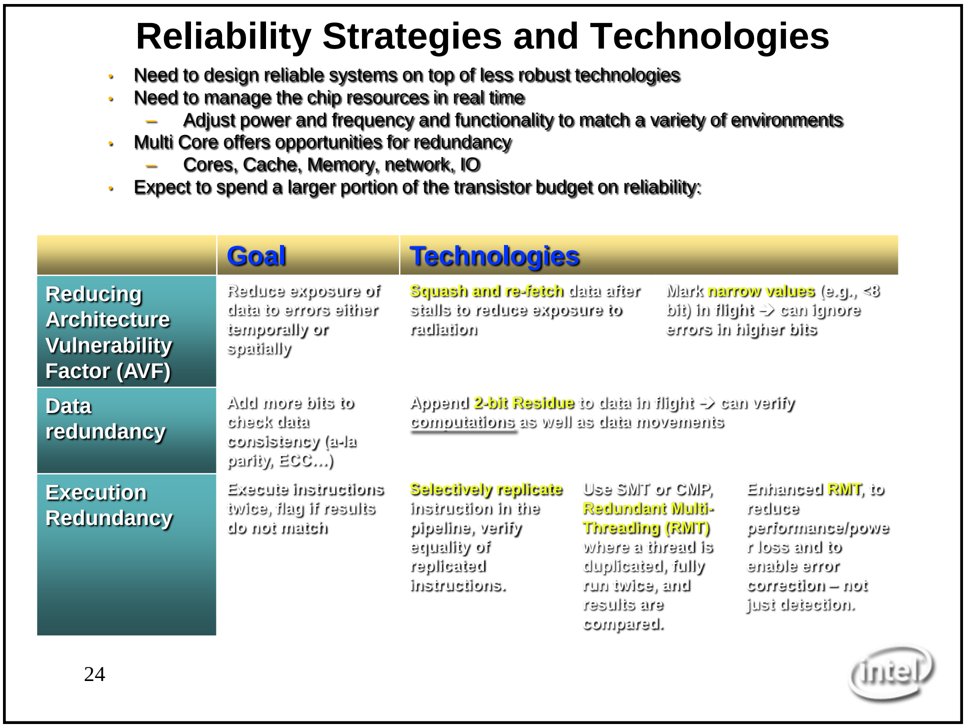
From: [Jaleel, Mattina,…HPCA 2006]



System Scaling

- **Traditional Differentiators:**
  - **Socket Connectivity, Cache size, RAS**
- **Going forward, Multi Core adds on-die communication as key value differentiator:**
  - **Core count, on-die scaling, cache size, Memory BW, RAS, Accelerators**
  - **Tiny <> Small <> Medium <> Big**
  - **Single Stream <> Throughput**

Building Blocks:
Pentium cores
LPIA cores
Cache
Memory



Reliability Strategies and Technologies

- Need to design reliable systems on top of less robust technologies
- Need to manage the chip resources in real time
  - Adjust power and frequency and functionality to match a variety of environments
- Multi Core offers opportunities for redundancy
  - Cores, Cache, Memory, network, IO
- Expect to spend a larger portion of the transistor budget on reliability:

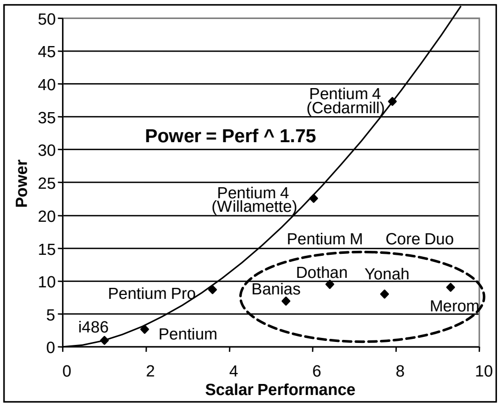
	Goal	Technologies		
Reducing Architecture Vulnerability Factor (AVF)	Reduce exposure of data to errors either temporally or spatially	Squash and re-fetch data after stalls to reduce exposure to radiation		Mark narrow values (e.g., <8 bit) in flight → can ignore errors in higher bits
Data redundancy	Add more bits to check data consistency (a-la parity, ECC…)	Append 2-bit Residue to data in flight → can verify computations as well as data movements		
Execution Redundancy	Execute instructions twice, flag if results do not match	Selectively replicate instruction in the pipeline, verify equality of replicated instructions.	Use SMT or CMP, Redundant Multi-Threading (RMT) where a thread is duplicated, fully run twice, and results are compared.	Enhanced RMT, to reduce performance/power loss and to enable error correction – not just detection.

Control CMP Activity for Power

- **Intel can now pack more transistors on a die than reasonably power and cool at max voltage & frequency**
  - Recall: Dynamic Power = VDD²x Cap x freq
  - Traditional Methods: Voltage scaling, Clock Gating
- **Wide variance between worst case and typical demands on power supply and cooling system**
  - Max current flow
  - di/dt swings at several frequencies
  - Total power dissipation
- **Goal: Maximize performance, accounting for physical constraints**
  - Controlling activity limits di/dt, max current draw and temp.
  - Take advantage of the activity constraints to run at a higher freq.

Energy Per Instruction (EPI)

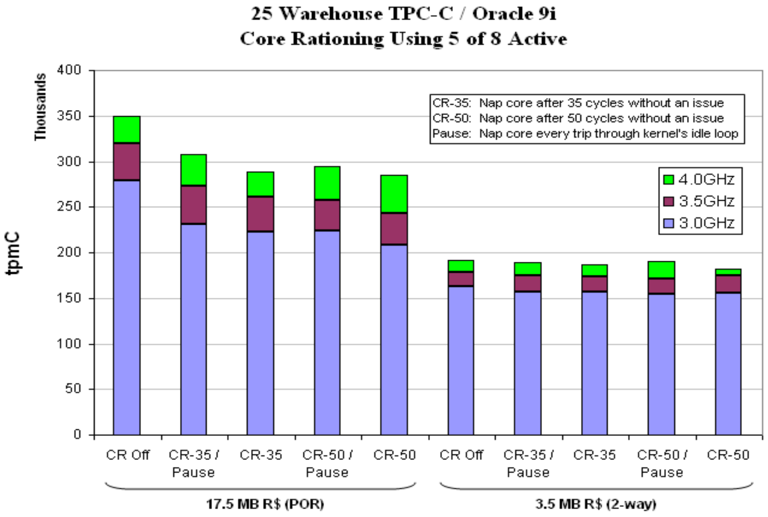
EPI looking better for Intel Architecture (IA) Just in time!



(From study by Ed Grochowski, Intel, MTL. IDF Spring 2006 White Paper *ftp://download.intel.com/technology/EEP/epi-trends.pdf)*

Energy Efficiency is key to CMP scaling

Power Management with CMP



Results from Michael Adler and others at Intel, DAP using ASIM / SoftSDV

Summary

- **Silicon integration continues to be a driving force**
- **Multi Core is an exciting opportunity to increase performance and simplify system design**
  - Great on-die scaling, high bandwidth, short latency, power and area efficiency
- **Will stress chip resources**
  - Don't over-subscribe available power and bandwidth
- **Intel is working to:**
  - Design balanced Multi Core Processor Chips
    - Power, Die area, Bandwidth, Caches
  - Analyze core sizes and functionality for different markets
  - Help solve the software scaling problems