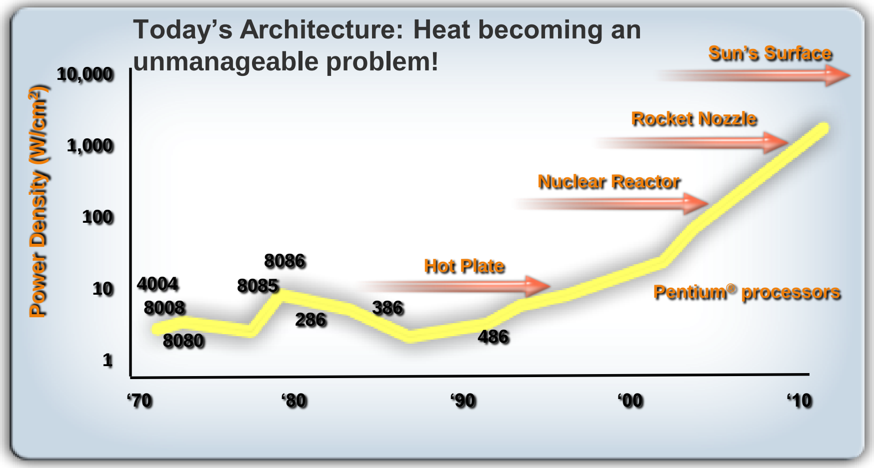
Supercomputing and Mass Market Desktops

John Manferdelli

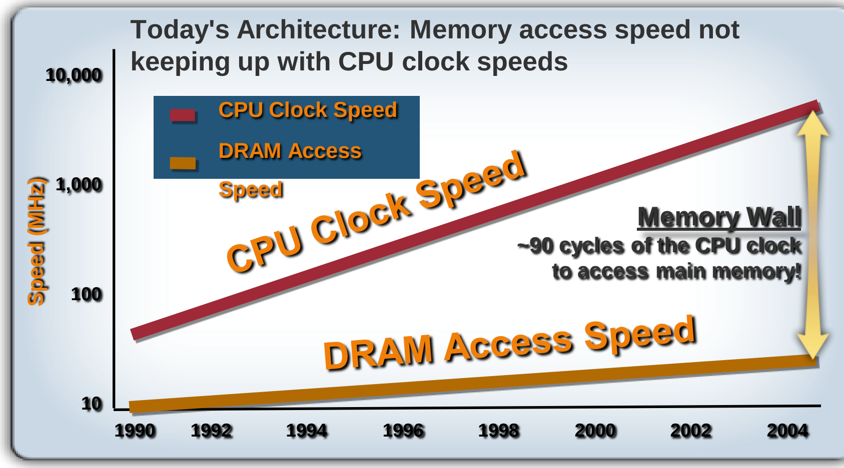
Microsoft Corporation

Capabilities on the rise for the end user

Hardware Paradigm Shift - 2004



Intel Developer Forum, Spring 2004 - Pat Gelsinger



Modern Microprocessors - Jason Patterson

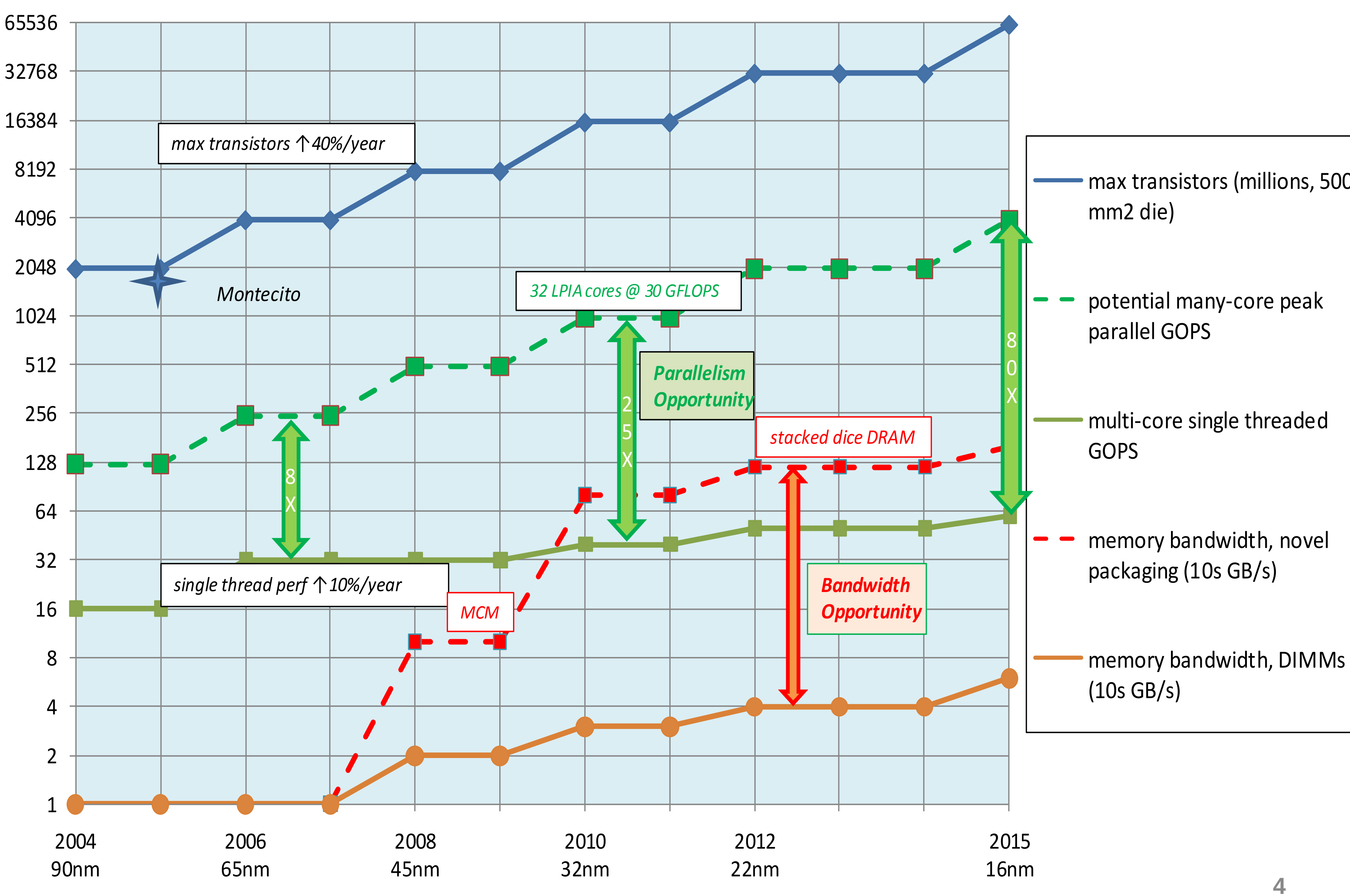
"… we see a very significant shift in what architectures will look like in the future …
fundamentally the way we've begun to look at doing that is to move from instruction level concurrency to … multiple cores per die. But we're going to continue to go beyond there. And that just won't be in our server lines in the future; this will permeate every arch[...]ture that we build. *All will have massively* multicore implementations."

Intel Developer Forum, Spring 2004
Pat Gelsinger
Chief Technology Officer, Senior Vice President
Intel Corporation
February, 19, 2004

Intel Cancels Top-Speed Pentium 4 Chip
Thu Oct 14, 6:50 PM ET Technology - Reuters
By Daniel Sorid

Intel …canceled plans to introduce its highest-speed desktop computer chip, ending for now a 25-year run that has seen the speeds of Intel's microprocessors increase by more than 750 times.

Trends

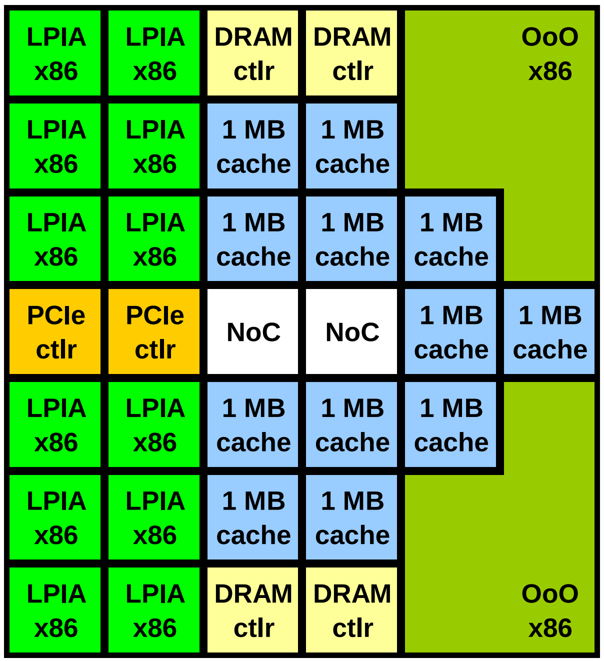


65536
32768
16384
8192
4096
2048
1024
512
256
128
64
32
16
8
4
2
1
max transistors ↑40%/year
Montecito
32 LPIA cores @ 30 GFLOPS
Parallelism Opportunity
8X
25X
80X
stacked dice DRAM
single thread perf ↑10%/year
MCM
Bandwidth Opportunity
2004 90nm
2006 65nm
2008 45nm
2010 32nm
2012 22nm
2015 16nm
max transistors (millions, 500 mm2 die)
potential many-core peak parallel GOPS
multi-core single threaded GOPS
memory bandwidth, novel packaging (10s GB/s)
memory bandwidth, DIMMs (10s GB/s)

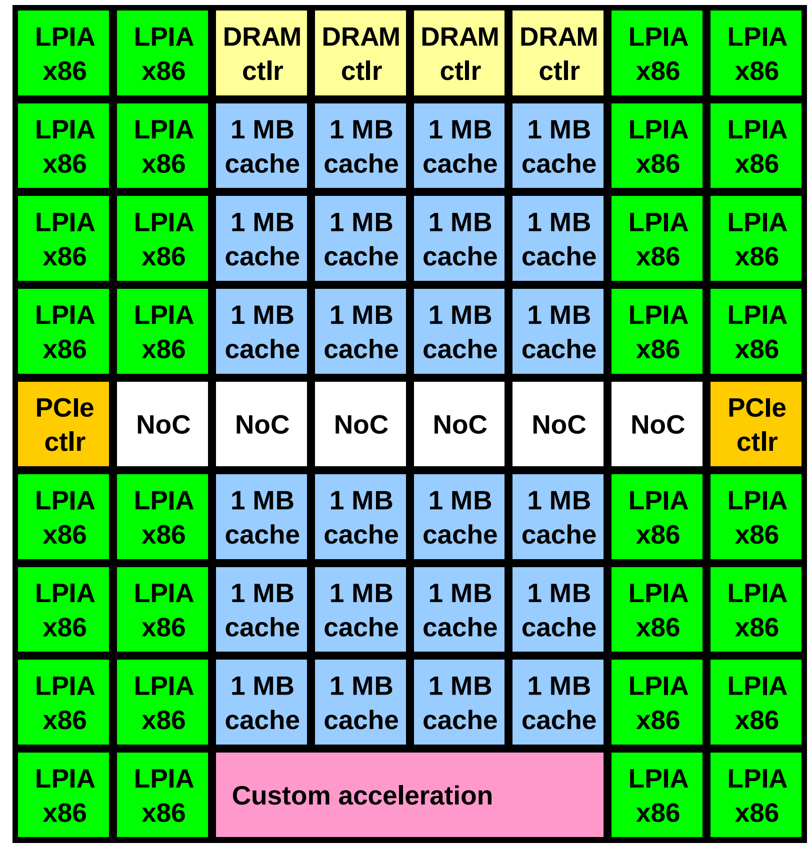
Lots more computing power

Los Alamos Computing Center? No a single CPU chip.

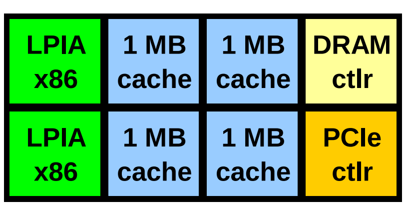
Desktop: 200 mm^2, 100 W, \$400



Server: 350 mm^2, 120 W, \$2000



Ultra-Mobile: 40 mm^2, 5 W, \$50



(2008 45 nm process)

Many Core in a Nutshell

1. We believe user experiences will benefit from 100 fold improvements in computational power.
2. Because of physics, you won't get this power from frequency scaling or programmer transparent hardware like ILP. The only way is parallelism.
3. Chips with many *heterogeneous* cores can be manufactured now. Only with chips like this can such performance scaling continue.
4. There are difficult but solvable issues related to memory bandwidth and I/O for this to work well over-all.
5. There are other benefits from such chips like good power and manufacturing characteristics .
6. Application of such hardware does not appear to be limited by any intrinsic lack of parallel algorithms.

Manycore Induced Hard Problems

Constructing parallel applications

- Encapsulating parallelism in reusable components
- Integrating concurrency & coordination into existing programming languages
- Raising the semantic level to eliminate sequencing
- Reducing the complexity of debugging, tuning and testing

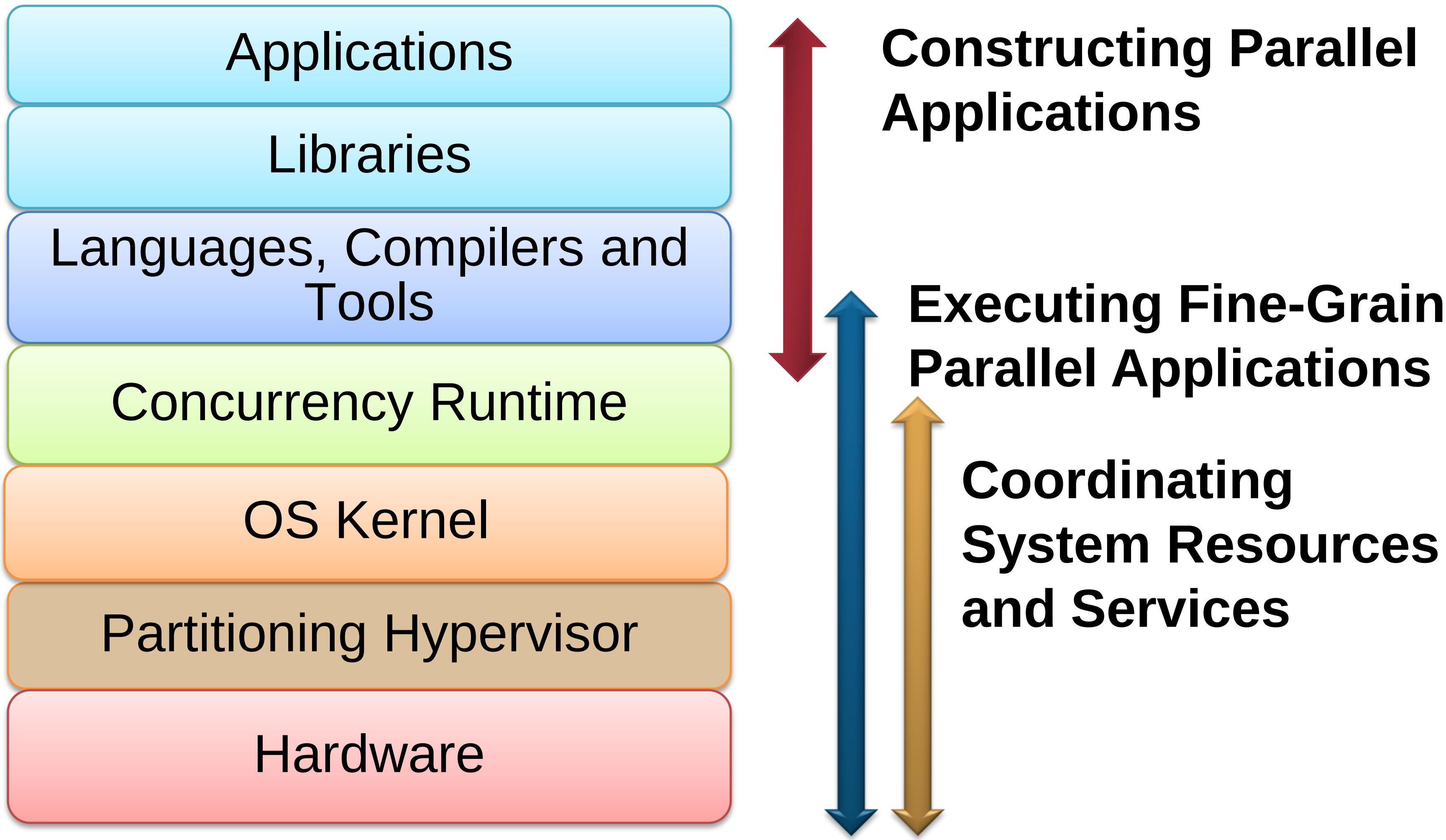
Executing fine-grain concurrent applications

- Managing large amounts potential concurrency
- Supporting lightweight transactions
- Interoperating with legacy thread model and interfaces
- Evolving hardware to effectively support parallel programs

Coordination of system resources and services

- Assigning resources securely
- Hosting concurrent operating environments
- Managing resources cooperatively
- Providing concurrent system services
- Managing heterogeneous resources

Manycore Stack



Developer Impact

Programming Abstractions

- Asynchronous Agents
- Concurrent Collections
- Transactions

Existing Languages

- Data Parallel Extensions
- Transactions
- API for messaging

New Languages

- *Safety & Portability*
- Streaming Data Parallel
- Functional Parallel

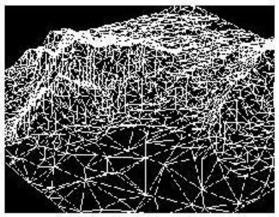
Infrastructure

- *Transaction & Parallel Safe*
- *Numerical Libraries*
- *Declarative Languages*

Developer Tools

- Visual Designers
  Agents, Dataflow
- Diagnostics
  Debugging & Performance Tuning
- Testing & Verification

Simulating The Physical World

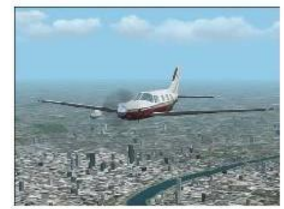


DEM
Space Shuttle
GTOPO 30
NED (US)
Earthsat (Europe)
GeoScience (Australia)

Landclass
Tiling Textures
Satellite Imagery

Vectors
Roads
Rivers
Lakes
Coastlines
Pipelines
Power Lines
Etc.

World
Weather
Seasons
Celestial Catalogue
Moon Phases
Time Zones
Magnetic Variation

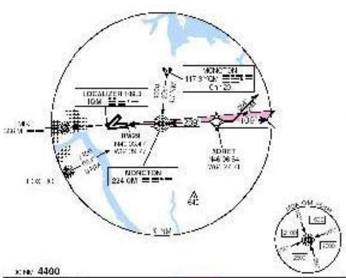


Autogen
Buildings
Houses
Trees
Fences
Etc.

Cities

Airports

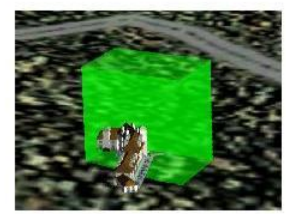
Objects
Landmarks
Bridges
Dams



Facilities Data
Jeppesen Naviads
NOAA Hazards
DAFIF
FAA

Vehicle Simulation
Airplanes
Boats
Ground Vehicles
-Cars & Trucks
-Trains
Ambient Traffic

AI Paths and Waypoints
Jeppesen
Navtech



Triggers
Text
Audio
Cameras
Animations
Cut-Scenes

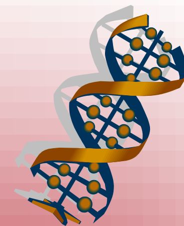
After Action Review
Goals
Timing
Scoring

Gameplay Simulation

Technical Computing Is Far Reaching

Earth Sciences

Life Sciences



Social Sciences

Multidisciplinary Research

Technical Computing

New Materials, Technologies and Processes

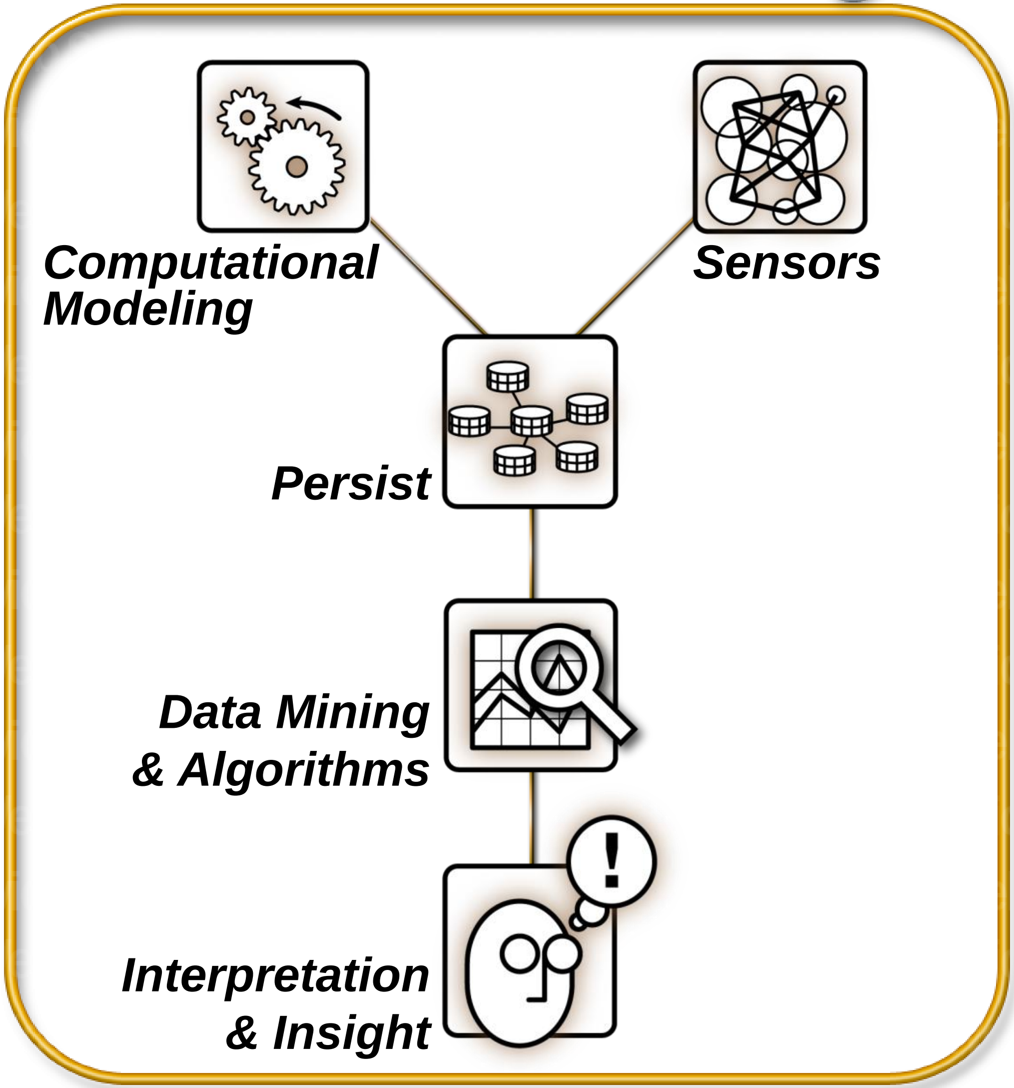
Computer and Information Sciences

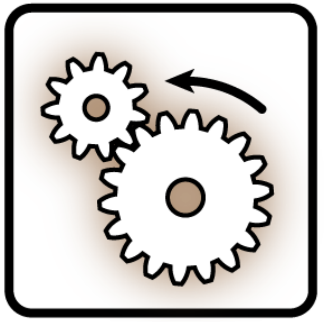
Math and Physical Science

$$E=MC^2$$

Technical Computing

Reduced Time To Insight





Computational Modeling

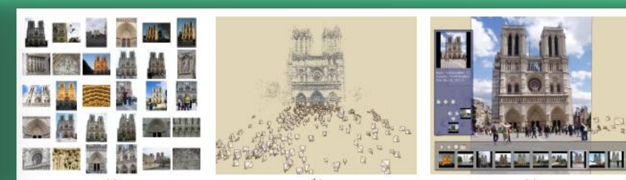
A transformative example

Modeling of the instrumented world

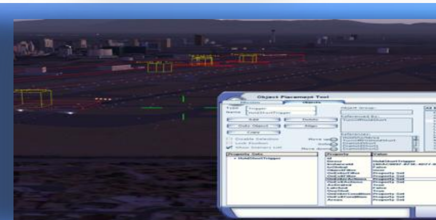
First Responder Scenario (empowering users to make good decisions)

First responders need to quickly and safely experience unfamiliar areas under different disaster scenarios before they deploy into it. For example, earthquakes in San Francisco.

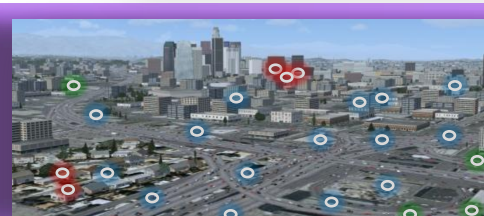
Setup Environment



Explore Possibilities



Monitor and Take Action



14

Scenario Features → Tech Domains

Modeling the instrumented world

Scenario: First responders need to quickly and safely experience unfamiliar areas under different disaster scenarios before they deploy into it.

Technology Domains: To realize this scenario, we need to compose these technologies and accelerate their performance at least 10-fold on the manycore client.

	Machine Learning	Math	Physics	Statistics	Image Processing	3d Modeling	Rendering	Database	Query engine	Index engine	Data compression	Motion Input
Stitching scenes together (2D, 3D)		✓			✓	✓	✓	✓				
Object recognition in scene	✓	✓				✓		✓				
Interaction (navigation) in scene			✓		✓	✓	✓	✓	✓			✓
Linking scene objects to live data	✓							✓	✓	✓	✓	
Filling in missing information	✓	✓			✓		✓	✓	✓		✓	
Image and video correction	✓	✓		✓	✓	✓	✓					
Long running scene statistics		✓		✓				✓	✓	✓		
Personalization of content	✓							✓		✓		
Simulating physical events		✓	✓		✓	✓	✓					
Real-time scene update		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

Tech Domains → Platform

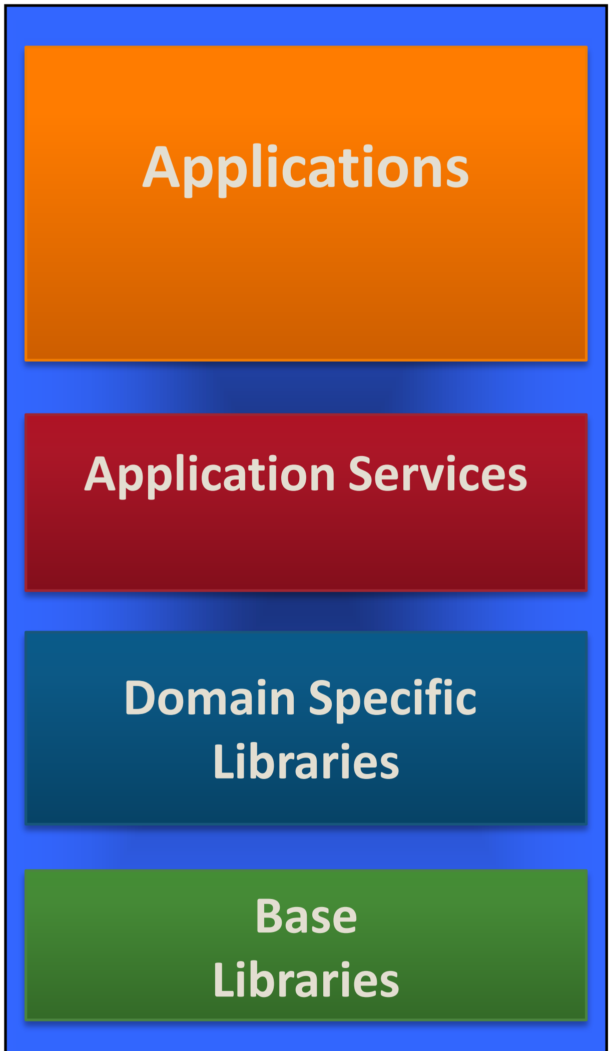
Technology Domains

To realize this scenario, we need to compose these technologies and accelerate their performance at least 10-fold on the manycore client..

Platform : To compose technology domains and accelerate their performance 10-fold, we need to build the following platform components

	Integrating concurrency into existing languages	Encapsulating parallel-ism in reusable modules	Tools for debugging and profiling parallel code	Support for fine-grain data parallelism	Support for lightweight transactions	Optimization of available bandwidth	System-wide dynamic resource management	Connectivity of services	Managing manycore hardware platform
Machine Learning	✓	✓	✓	✓	✓	✓	✓		✓
Math	✓		✓				✓		✓
Physics	✓	✓	✓	✓		✓	✓		✓
Statistics	✓	✓	✓				✓		✓
Image Processing	✓	✓	✓	✓	✓	✓	✓	✓	✓
3D Modeling	✓	✓	✓	✓	✓	✓	✓		✓
Rendering	✓	✓	✓	✓	✓	✓	✓	✓	✓
Database	✓	✓	✓	✓	✓	✓	✓	✓	✓
Query engine	✓	✓	✓	✓	✓	✓	✓	✓	✓
Data compression	✓		✓				✓		✓
Gesture/Motion Input		✓		✓	✓	✓	✓	✓	✓

Application/Libraries Architecture



Personal Assistant
Natural UI
Program Development
Technical Computing
System Design

Information Management
Games & Entertainment
Biology/Health care
Business Intelligence
Robotics

Speech Engine, Gaming (Physics/AI), Unified Communications, **Database**, Vision, **Machine Learning, Semantic Processing**

Single domain physics, search algorithms, audio/video characterization and extraction, biology simulations, optimization and constraint resolution

Common data structures and algorithms: trees, graphs, tables, sorting, traversal.

System Architecture