Massive Data Computing
In a Connected World

2009 GRC ETAB Summer Study, June 29 2009

Pradeep K. Dubey
IEEE Fellow and Director of Throughput Computing
Intel Corporation
Predicting Future Apps is a tricky business

“**A reasonable man adapts himself to his environment. An unreasonable man persists in attempting to adapt his environment to suit himself** …

… *Therefore, all progress depends on the unreasonable man.*” -- **George Bernard Shaw**

Replace “man” with “application”, and you get one definition of a **killer app**, namely *that unreasonable application which succeeds in leaving its mark on the surrounding architecture.*

All architectural progress depends on such unreasonable apps!
Decomposing Emerging Applications

Mining
- Large dataset mining
- Semantic Web/Grid Mining
- Streaming Data Mining
- Distributed Data Mining
- Content-based Retrieval

Indexing
- Collaborative Filters
- Multidimensional Indexing
- Dimensionality Reduction
- Dynamic Ontologies
- Efficient access to large, unstructured, sparse datasets
- Stream Processing

Streaming
- Photo-real Synthesis
- Real-world animation
- Ray tracing
- Global Illumination
- Behavioral Synthesis
- Physical simulation
- Kinematics
- Emotion synthesis
- Audio synthesis
- Video/Image synthesis
- Document synthesis

Recognition (Modeling)
- Multimodal event/object Recognition
- Statistical Computing
- Machine Learning
- Cluster Analysis
- Multidimensional
- Bayesian Networks
- Markov Models
- Neural networks
- Probability networks
- LP/IP/QP/Stochastic Optimization

Graphics
- Real-world animation
- Photo-real Synthesis
- Real-world animation
- Ray tracing
- Global Illumination
- Behavioral Synthesis
- Physical simulation
- Kinematics
- Emotion synthesis
- Audio synthesis
- Video/Image synthesis
- Document synthesis
Interactive RMS Loop

Recognition
What is …?

Mining
Is it …?

Synthesis
What if …?

Model
Find an existing model instance
Create a new model instance

Most RMS apps are about enabling interactive (real-time) RMS Loop (iRMS)

Pradeep.Dubey@intel.com  6/29/2009
Visual Computing

Visual Computing (Graphics and Vision)

Procedural or Analytical

Model

Physical Simulation

Rendering

Data Acquisition and Mining Based

Non-Visual Computing (Search and Analytics)

Procedural or Analytical

Model

Real-time Indexing

What-if evaluation

Summary synthesis

Data Mining Based (Crawling)
iRMS Visual Computing Loop
Visual Computing -- Where are we headed

Rendering Simulation
- collision detection
- force solver
- global illumination
- ...

Machine learning
- Neural networks
- Probabilistic reasoning
- Fuzzy logic
- Belief networks
- Evolutionary computing
- Chaos theory
- ...

Physics
- Dynamics

Soft Computing
- Constraints

Soft Physics?
- Constraint Dynamics
Going Beyond Visual Senses …

- Haptic dynamics in haptic training apps
  - System fully usable in the operating room

- Real-time implicit simulation for 100K elements
  - For real-time training tools & very accurate prediction

- Interactive quasi-statics simulations of 100K elements
  - Good usability for planning and prediction

- Offline dynamic Simulations of 100K elements
  - Limited usability for prediction

- Offline quasi-statics simulations of 10K elements
  - Impractical for use in clinical environments

Force simulations for visual rendering: 10s of Hz
Force simulations for Haptic rendering: KHz or more
Analytics Computing

Visual Computing (Graphics and Vision)
- Model
  - Physical Simulation
  - Rendering
- Data Acquisition and Mining Based

Non-Visual Computing (Search and Analytics)
- Procedural or Analytical
  - Model
- Data Mining Based (Crawling)
  - Real-time Indexing
  - What-if evaluation
  - Summary synthesis

Pradeep.Dubey@intel.com
6/29/2009
Analytics Computing: Where are we headed

- **Web Today**
  - Tim Berners-Lee, et. al. “Most of the Web’s content today is designed for humans to read, not for computer programs to manipulate meaningfully”

- **Web Tomorrow**
  - Semantic Web according to Tim Berners-Lee: “*Adds logic to the Web*”

- **Implications**
  - Web focus shifts from “*data presentation to end-users*” to “*automatic data processing on behalf of end-users*”, also known as, “*analytics*”
  - Processing requirements growth shifts from ‘visual computing’ to ‘analytics’
    - Multiple inner loop iterations of “non-visual computing or analytics” for every outer loop of “visual computing”

http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

Pradeep.Dubey@intel.com  6/29/2009
What has been missing

- **Software Stack 😊:**
  - Tim Berners-Lee et. al.: “*For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning*”

- Almost there now …
  - XML – RDF – OWL etc.

http://www.w3.org/2003/Talks/0922-rsoc-tbl/
Growing Importance of Data-driven Models

**Visual Computing (Graphics and Vision)**
- Procedural or Analytical
- Model
- Physical Simulation
- Rendering

**Data Acquisition and Mining Based**

**Non-Visual Computing (Search and Analytics)**
- Procedural or Analytical
- Model
- Real-time Indexing
- What-if evaluation Summary synthesis

**Data Mining Based (Crawling)**
Massive Data & Ubiquitous Connectivity

- Data-driven models are now tractable and usable
  - We are not limited to analytical models any more
  - No need to rely on heuristics alone for unknown models
  - Massive data offers new algorithmic opportunities
    - Many traditional compute problems worth revisiting

- Web connectivity significantly speeds up model-training

- Real-time connectivity enables continuous model refinement
  - Poor model is an acceptable starting point
  - Classification accuracy improves over time
**Image Completion**

- **Scene completion using millions of photographs** James Hays, Alexei A. Efros, Siggraph 2007, Also, October 2008 Communications of the ACM, Volume 51 Issue 10

- “Dramatic improvement when moving from ten thousand to one million images”

- “Brute force many currently unsolvable vision and graphics problems!”
  - For example: “Feasibility of sampling from the entire space of scenes as a way of exhaustively modeling our visual world”
Rule-based system exceeds human performance in a structured, deterministic domain

Newcomer Google beats decades of rule-based translation research with a language-unaware statistical approach to MT

- Statistical inference (not rules)
- 100s of TB of training data
- Racks of computation

Semantic Search

- Google Rolls Out Semantic Search
  - http://www.pcworld.com/businesscenter/article/161869/google_rolls_out_semantic_search_capabilities.html

- "What we're seeing actually is that with a lot of data, you ultimately see things that seem intelligent even though they're done through brute force," she said. "Because we're processing so much data, we have a lot of context around things like acronyms. Suddenly, the search engine seems smart, like it achieved that semantic understanding, but it hasn't really. It has to do with brute force. That said, I think the best algorithm for search is a mix of both brute-force computation and sheer comprehensiveness and also the qualitative human component."

  -- Marissa Mayer, VP of Search Products, Google
Mind Reading

How Technology May Soon "Read" Your Mind

- [Link](http://www.cbsnews.com/stories/2008/12/31/60minutes/main4694713.shtml)

- “… Tom Mitchell at Carnegie Mellon University have done is combine fMRI's ability to look at the brain in action with computer science's new power to sort through massive amounts of data. The goal: to see if they could identify exactly what happens in the brain when people think specific thoughts.”
Nested RMS

Recognition
What is …?

Mining
Is it …?

Synthesis
What if …?

Graphics Rendering + Physical Simulation

Semantic Web Mining

Learning & Filling Ontologies
Structured Data + Unstructured Blogs
Synthesized Structures

Learning & Modeling
Visualization Streams
Synthesized Visuals

Computer Vision
Structured Augmentation
Reality Augmentation

Pradeep.Dubey@intel.com 6/29/2009
Nested RMS Instance: Virtual World

Outer loop:
- iRMS Visual Loop:
  - One Per Real User

Inner Loop:
- iRMS Analytics Loop
  - One per bot per user

Performance Needs:
- Can far exceed typical i/o limits of human perception

Performance Needs:
- Limited by input/output limits of human perception

Chat Bots
Gambler Bots
Player Bots
Reporter Bots
Trade Bots
Shop Bots

Pradeep.Dubey@intel.com
6/29/2009
Putting it all together

Sensory Immersion

Behavioral Immersion

Super Immersion

Rendering
Simulation

Machine learning
Neural networks
Probabilistic reasoning

Fuzzy logic

Performance Needs:
Can far exceed typical I/O limits of human perception

Immersive Computing can bridge Norman’s Gulf
Computational requirements are huge!
Where is my computer 😊

Visual Computing
(Clients)

Private data, sensory inputs, streams/feeds
immersive 3D graphics output, interactive visualization

Massive Data Analytics
(Cloud of Servers)

Intersection of massive data with massive compute
real-time analytics, massive data mining-learning

Architectural Implications Are Even More Radical!
Our Contribution

- **Throughput Computing: Research to Realization**
  - Application-driven Architecture Research
  - Larrabee/manycore Opportunities

- **Workload focus:**
  - Nested iRMS: *Server-side analytics loop nested inside Client-side visual computing loop*
  - Such as: Massive data computing, Multimodal real-time physical simulation, Behavioral simulation, Interventional medical imaging, Large-scale optimization (FSI), Video Karaoke 😊

- **Research Focus** ➔ Server-Client decomposition
Our Contribution ...

Architectural Research

- Cores
- Cache memory
- On-die fabric
- I/O, network, storage
- Platform firmware/Ucode
- Execution environments
- Programming environments
- Workloads

Workloads used to validate designs

Workload requirements drive design decisions