Looking Forward at NSF

Jeannette M. Wing

President's Professor of Computer Science and Department Head
Computer Science Department, Carnegie Mellon University

and

Incoming CISE AD

CRA, February 26, 2007
NSF, February 27, 2007

Outline

- Looking Far Out
  - Vision: Making Computational Thinking Commonplace
  - Deep Questions for Computing
- Looking 3-10 years out
  - Big Idea #1: Super Data Center
  - Big Idea #2: The Brain
- Looking near term
  - CISE programmatic thoughts
  - Bread and butter thoughts
  - CDI, CCC, GENI

Warning: I still have a lot to learn about NSF and I come with expertise in some, but not all areas of computer science so my ignorance and biases will show through. I am an eager learner though!

Looking Far Out

My Grand Vision for the Field

- Computational thinking will be a fundamental skill used by everyone in the world by the middle of the 21st Century.
  - Just like reading, writing, and arithmetic.
  - Imagine every child knowing how to think like a computer scientist!
  - Incestuous: Computing and computers will enable the spread of computational thinking.
  - In research: scientists, engineers, …, historians, artists
  - In education: K-12 students and teachers, undergrads, …

Deep Questions for Computer Science

- P = NP ?
- What is computable?
  - What is the power of computing, by machine and human together?
- What is intelligence?
  - Understanding the brain
- How can we build complex systems simply?

Looking 3-10 Years Out

Super (Duper) Big Idea #1: ~~Google-sized~~ Data Center(s)

- 10K servers, $20M
- Open, shared research facility
- Intellectual justification
  - Theory
    - Algorithms (spectral graph analysis a la PageRank), programming languages (massively parallel and ultra reliable a la MapReduce)
    - A New Computing Platform: not a pc, not a supercomputer, not a distributed system, not a network (a la GENI)
  - Systems and Software
    - Self-*: self-configuring, self-managing, self-tuning, self-diagnosing, self-healing, self-repair
    - Power: Google and the Columbia River, consumption equal to a metropolitan area
    - *Ambitiously, two or more data centers networked*
      - speed of light constraint in Google response time
      - redundancy, backup, task repartitioning, networking (bandwidth, routing)
    - Software needed to program, operate, and manage apps, e.g., "cluster O/S"
  - Applications: "Data is Gold"
    - Beyond search and web-crawling!
    - CS: Natural language learning, data-driven graphics and animation, SAT solvers …
    - Beyond CS: astronomy (Sloan Digital Sky: 215 million unique objects and growing), biology (PDB's 41,687 protein structures and growing is only 1% of known), …
- Logistics, politics
  - Plan: Layout groundwork for eventual MREFC
    - Google and IBM: 500-1000 node cluster this summer (80-90% prob). Get Microsoft on board too.
    - MREFC: feasibility and justification laid out, timing right
    - OCI, MRI roles?
  - CCC 2nd pitch (helps disentangle GENI from CCC), need grassroots effort, community buy-in
  - Government-Industry-Academic Partnership (would look good to Congress)
    - Sponsored by NSF, Google, IBM, Microsoft, Yahoo!, HP, etc.
    - Google proof of concept: 1K servers for $1M. IBM helps with software virtualization, open source s/w

Big Idea #2: Understanding the Brain

- What is intelligence?
  - Computational neuroscience, theory of consciousness, fMRI data, cognitive science, social and behavioral sciences
- Suppose we were to build a brain?
- Cross-directorate, inter-agency program

Looking Near Term:

CISE Programmatics

Back to Basics

- NSF is about basic science and engineering.
- It's all about good ideas and good people.
- It's about "high risk" long term impact.
  - Impact may be far in the future.
  - Impact is long-lasting (that is real science).
  - Impact can create new economies and change societal behavior.

CISE Programmatic Thoughts: Small to Large

- CISE
  - Grant size. Mixed portfolio: small, medium, large
  - Review process. Goal: Improve quality. Revisit panels, individual, hybrid.
  - Clusters: Revisiting CCF and CNS. IIS has a new one.
- Cross-directorate programs, e.g., CDI
  - CISE + {MPS, Eng, Bio, Geo, SBE, OCI, Polar, …}
- Inter-agency programs
  - NSF/CISE + {NIH, DOE, …}
- Goal: Build communities.
  - NSF can spawn and foster.
    - We can play the role DARPA used to. Let's embrace it, not run from it!
  - CCC can sustain (we hope).

Looking Near Term:

Bread and Butter

Foundations: "flour and water"

- P = NP?
- Algorithms, data structures, and complexity
- Models, logics, and semantics

Celera (shotgun sequencing), RSA (crypto), Google (PageRank, MapReduce), MapQuest (shortest path), Akamai (consistent hashing), Intel/IBM/…/Microsoft (temporal logic, model checking), ….

Foundations: New Math Models (Examples)

- Game Theory
  - Computational microeconomics
    - e.g., ad placement, on-line auctions, organ exchange
  - Networking
    - e.g., congestion control, adaptive wireless networks
  - Security
    - e.g., two-party games (adversary and administrator)
- Topology
  - Distributed systems, sensor nets
  - Robotics, e.g., self-configurable, robot arms, motion planning
  - Protein structure, e.g., knot theory
- Spectral Graph Analysis (algebraic graph theory meets linear algebra)
  - Image segmentation, e.g., medical, face recognition
  - Data clustering, e.g., data mining, market research, social networks
  - Scientific computing, e.g., energy and angular momentum of electrons; absorption spectrum of chemicals

Foundations: New and Future Models of Computation

- Multi-core, many-core
- Human and machine
- Networks of…
- …
- Nano
- Bio
- Quantum

Systems and Software

How can we build complex systems simply?

- Nature of tomorrow's systems
  - Dynamic, ever-changing, 24/7 reliability
  - Self-* (aware, diagnosing, healing, repairing, managing)
- Two important classes converging
  - Embedded
    - Networked architecture, e.g., sensor nets (see below)
    - Safety-critical apps, e.g., medical, automotive, aero&astro
    - Challenge: Reasoning about uncertainty, e.g., Human, Mother Nature, the Adversary
  - Pervasive and mobile
    - Focus on sensors and actuators, not just the devices and communication links
    - Prevalence of cell phones, iPods, RFIDs, …
    - Implications for HCI, embedded systems, sensor nets (see above)

Evaluation of System/Software Quality

- Different notions of "correctness" (reliability)
  - Context of use (reality) vs. intended/expected use (ideal)
  - Promising trend: Convergence of program analysis, type theory, formal methods (model checking, theorem proving, decision procedures)

- Metrics
  - Not just performance (space and time), but power (energy)
  - S/W: beyond LOCs and bug counts
  - Manageability, especially in the presence of continuous change
  - Trustworthiness: Security, Privacy, Usability, Reliability

Trustworthy Computing

- Security
  - New models for networks: security without borders
  - Threats of the future
  - Not just code, but component level and above
- Privacy
  - Models and logics for reasoning about privacy
  - System mechanisms, protocols, and tools for enforcing privacy policies
- Usability
  - Human is often the weakest link. Social engineering
  - Today's mechanisms are obscure and hard to use.
- Reliability
  - See other slides.

Intelligent Systems: Don't Forget Humans

- Human and Machine
  - What is computable, by human and computer together?
    - Human Computation, e.g., ESP
- Pervasive sensing and actuation
  - Speech, natural language processing
  - Vision, graphics, animation
- Robotics
  - HRI: Humanoid robots, social robots, quality-of-life robots
  - Medical robotics
- Virtual communities
  - Social networks, networks of brains
- Computing and the Arts and Humanities

All this requires/drives advances in theory and systems!

Looking Near Term:

CDI, CCC, GENI

Looking Near Term

- Cyber-enabled Discovery and Innovation
  - $52M cross-directorate, $20M to CISE

- CCC
  - Chair will soon be announced, council to be formed.
  - I intend to work with CCC as closely as possible.
  - Super Data Center could be another CCC thrust.

- GENI
  - I am still talking to various parties.

Major Themes Looking Ahead

- Human (as a First-Class Computer) + Machine
  - What is computable?
  - What is intelligence?
  - Understanding complexity.

- Emerging Models of Computation
  - Now: Multi-core, ubiquitous perception and actuation
  - Tomorrow: GENI, Super Data Center, The Brain
  - Day after Tomorrow: Nano, Bio, Quantum

- Computational Thinking for All
  - Research: science, engineering, …, humanities, arts
  - Education: K-12, …

Priorities

Guiding Principles

- Looking Far Out
  - Vision: Making Computational Thinking Commonplace
  - Deep Questions for Computing
- Looking 3-10 years out
  - Big Idea #1: Super Data Center
  - Big Idea #2: The Brain
- Looking near term
  - CISE programmatic thoughts
  - Bread and butter thoughts
  - GENI, CCC

I Am

to being

Looking Forward at NSF!