Computer Science and the Singularity – The Prospect of Conscious Machines

Food for Thought II
August, 2014
David Gnabasik
Food for Thought II

• **White Chicken Chile**
  
  – 12-inch non-stick skillet.
  
  – Ingredients (makes 6x9-ounce servings) Cost: $4.50/serving
  
  – 1 tablespoon vegetable oil
  – 1 small onion, chopped (1/2 cup)
  – 2 15-ounce cans white beans, rinsed
  – 2 garlic cloves, finely chopped
  – 1 pound cubed chicken breasts
  – 2 can corn & peppers, undrained
  – 1 4-oz can diced green chilies, undrained
  – 3 1/2 cups premium chicken broth
  – 2 teaspoons ground cumin
  – 1/2 teaspoon cayenne pepper

  In skillet, heat oil over medium-high heat. Add chicken & onion; cook and stir until chicken is no longer pink (internal 165°F). Stir in remaining ingredients. Cover; cook 10 minutes, stirring once. Refrigerate any leftovers. Tip: top with fresh lime juice, cilantro or avocado.
Primary Research Resources For This Topic

- *The Coming Technological Singularity*, Vernor Vinge, 1993 (a)
  - [https://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html](https://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html)
- *Why the Future Doesn’t Need Us*, Bill Joy, 2000 (b)
- *The Singularity is Near: When Humans Transcend Biology*, Ray Kurzweil, 2005 (c)
- June, 2008 *IEEE Spectrum – The Rapture of the Geeks* (d)
  - [http://spectrum.ieee.org/static/singularity](http://spectrum.ieee.org/static/singularity)
- Various books and movies
Questions to Address

1) How is the Singularity defined, and what does it envision?
2) Do we need a new definition of consciousness (i.e., self-awareness, sentience) or intelligence?
3) What are the CS assumptions, components, dependencies required for its possible realization?
   What CS tools can we use to analyze this possibility? What’s testable?
4) How would we as computer scientists possibly engineer hyper-intelligent or conscious machines?
5) Should building a Singularity be a goal of computer science?
6) Does building the Singularity imply the end of computer science?
Artificial Intelligence Movie Sampler

- Forbidden Planet (1958)
- 2001: A Space Odyssey (1968)
- Star Trek, Tron, etc.
- Blade Runner (1982)
- WarGames (1983)
- The Matrix (1999)
- A.I. Artificial Intelligence (2001)
- The Singularity Is Near (2010)
- Transcendence (2014)
Singularity Books

- *The Singularity is Near: When Humans Transcend Biology*, Ray Kurzweil (inventor)
- *Superintelligence*, Nick Bostrom,
- *The Rapture of the Nerds: A tale of the singularity, posthumanity, and awkward social situations*, Cory Doctorow (SF author)
- *A Fire Upon the Deep*, Vernor Vinge (Computer Scientist & teacher)
- *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, Ray Kurzweil (inventor)
- *Singularity Rising: Surviving and Thriving in a Smarter, Richer, and More Dangerous World*, James D. Miller (economist)
How is the Singularity defined?

• Rise of artificial hyper-intelligence?
• The merging of man and technology?
• A point in history beyond which our predictions are meaningless?
• Networks of embedded μ-processors (digital Gaia)
• Extrapolate the idea of exponential technological growth and you end up with several versions of the singularity.
• Other (economic/technological) “singularities”
  – Agricultural revolution
  – Industrial revolution
• Vinge: “… the world will become intrinsically unintelligible to the likes of us.” → singularity
  • Anti- or unscientific?
What does the Singularity envision?

• Technotopia.
• Human super-intelligence.
• Machine/robotic hyper-intelligence.
• Extended life spans.
• Merger of ourselves and our machines (cochlear implants), gradually replacing ourselves with our robotic technology, achieving near immortality by uploading our consciousness -- no death.
• Fundamentally unpredictable.
Vernor Vinge’s 1993 essay “The Coming Technological Singularity” (5747 words)

• What is The Singularity?
• Can the Singularity be Avoided?
• Other Paths to the Singularity:
  – Intelligence Amplification
  – Human/computer team automation
  – Develop human/computer symbiosis in art
  – Allow human/computer teams at chess tournaments
  – Develop interfaces that allow computer and network access without requiring the human to be tied to one spot
  – Develop more symmetrical decision support systems
  – Use local area nets to make human teams that really work
  – Exploit the worldwide Internet as a combination human/machine tool
• Strong Super-humanity and the Best We Can Ask for
Quotes from Vinge’s 1993 Essay

• “Progress in computer hardware has followed an amazingly steady curve in the last few decades.” (one reference given)
• “…superhumanity is the essence of the Singularity.”
  – Comment: “ultraintelligent“ never defined (equivalence)
• “disturbing consequences“
  – Comment: humans are easily frightened. Good SF author.
• “intelligence is a function of brain power”
  – Comment: What does this mean?
• “…algorithms are of central importance to the existence of minds.”
• “…intellectual runaway (which) is the essence of the Singularity.”
Do we need a new definition of consciousness (i.e., self-awareness) or intelligence?

- Christof Koch & Giulio Tononi: Can Machines Be Conscious? (d)
- Much brain activity has nothing to do with consciousness.
- Being conscious (subjective experience) does not require:
  - emotions - dreaming, frontal lobe damage
  - either explicit or working memory
  - attention & self-reflection, but coupled to selective attention.
    - neocortex-damaged patients have nearly intact perceptual abilities.
  - language - animals, infants.
- Essential properties of consciousness:
  - related to the amount of integrated information generated.
  - information is defined as the reduction of uncertainty that occurs when one among many possible outcomes is chosen.
  - A photodiode has just 2 states (1 bit each), but we have an uncountable number of things that are not seen.
Do we need a new definition of consciousness (i.e., self-awareness) or intelligence?

• To be conscious you need to be a single integrated entity with a large repertoire of available states.
  – Conscious experience implies differentiating among many states.
  – Photodiode states are completely independent, but our states are very interdependent $\rightarrow$ unity.
  – The repertoire of available states cannot be subdivided.

• Theoretical framework: Integrated Information Theory of Consciousness [*]

• IITC suggests a test of assessing machine consciousness.
  – Describe a scene that efficiently differentiates the scene's key features from the immense range of other possible scenes.

• Why not simply say “provides intelligent responses”??
What action or statement best demonstrates that the HAL 9000 computer is conscious or intelligent?

- “It can only be attributable to human error.”
- Dave: “Do you know what happened?” HAL: “I’m sorry, Dave. I don’t have enough information.”
- “Without your space helmet, Dave? You're going to find that rather difficult.”
- “That's a very nice rendering, Dave. I think you've improved a great deal. Can you hold it a bit closer? That's Dr. Hunter, isn't it?”
- “I'm sorry, Frank, I think you missed it. Queen to Bishop 3, Bishop takes Queen, Knight takes Bishop. Mate.”
- “I'm afraid. I'm afraid, Dave. Dave, my mind is going. I can feel it. I can feel it. My mind is going. There is no question about it. I can feel it.”
What are the CS assumptions, components, dependencies required for its possible realization?

• Artificial Intelligence (AI), genetic algorithms, neural networks, machine learning in general.

• Analyzing the brain as a circuit diagram:
  – Ultrahigh-resolution brain scans ➔ computer simulation

• Machines that reproduce: nano-technology, robotics, genetic engineering of synthetic organisms.
  – Now driven by commercial and not military interests.

• Define machine: an apparatus using or applying mechanical power and having several parts, each with a definite function and together performing a particular task. Now it seems to mean algorithm, a set of deterministic & well-defined rules.
Development of Necessary AI Capabilities

• Rodney Brooks (MIT's Humanoid Robotics Group) (d)
  – “…the brain is a machine, whether this machine is a computer is another question.”

• Possible development program:
  1. Object-recognition capabilities of a 2-year old.
  2. Language capabilities of a 4-year old.
  4. Social understanding of an 8-year old.

• Most significant factor is continued HW development.
  – Completely predicated upon Moore’s Law!
Gordon E. Moore's law: the number of transistors in a dense integrated circuit doubles approximately every two years. (1965)
Dennard Scaling = *Constant* Electric Field Scaling

- Dennard scaling theory deals with switching speeds and physical characteristics of transistors → heat dissipation, max clock speeds.
- MOSFETs continue to function as voltage-controlled switches while all key figures of merit such as layout density, operating speed, and energy efficiency improve – provided geometric dimensions, voltages, and doping concentrations are consistently scaled to maintain a *constant* electric field.
- Dennard scaling *made* Moore’s law work and it stopped working in 2004!
- Leakage, etc. *Power* becomes the limiting problem. Scaling → *variability*!!

### Device or Circuit Parameter | Scaling Factor
--- | ---
Device dimension $t_{ox}, L, W$ | $1/\kappa$
Doping concentration $N_a$ | $\kappa$
Voltage $V$ | $1/\kappa$
Current $I$ | $1/\kappa$
Capacitance $\varepsilon A/t$ | $1/\kappa$
Delay time per circuit $V C/I$ | $1/\kappa$
Power dissipation per circuit $VI$ | $1/\kappa^2$
Power density $VI/A$ | $1$

What Really Stopped Growing: Power/Chip

[Watts]

[Courtesy Mark Horowitz, Stanford]
Exponential Growth of Computing
Twentieth through twenty-first century

Logarithmic Plot

Calculations per Second per $1,000$

Year

1900 1920 1940 1960 1980 2000 2020 2040 2060 2080 2100

- All Human Brains
- One Human Brain
- One Mouse Brain
- One Insect Brain
Kurzweil believes that the exponential growth of Moore's law will continue beyond the use of integrated circuits into technologies that will lead to the technological singularity.
Application of Moore's Law

• In 2003, Intel predicted the end would come between 2013 and 2018 with 16 nanometer manufacturing and 5 nanometer gates, due to quantum tunneling.

• Joel Hruska, Intel’s former chief architect: Moore’s law will be dead within a decade, Aug. 30, 2013 [1]:
  – “It is a common (but mistaken) belief that Moore's law makes predictions regarding all forms of technology, when it was originally intended to apply only to semiconductor circuits. Kurzweil has hypothesized that Moore's law will apply – at least by inference – to any problem that can be attacked by digital computers as is in its essence also a digital problem. But nothing else we’ve ever discovered has scaled like semiconductor design. A 2011 study in the journal Science showed that the peak of the rate of change of the world's capacity to compute information was in the year 1998, when the world's technological capacity to compute information on general-purpose computers grew at 88% per year.[2] Since then, technological change has clearly slowed.”

• This is extrapolating an extrapolation (Moore's law) to another field.
Extending Moore’s Law - Continued Scaling

- Molecular integrated circuits; nanotechnology; quantum computing.
- Use spin state of electron spintronics, tunnel junctions, advanced confinement of channel materials via nano-wire geometry.
- Microarchitecture techniques exploiting the growth of available transistor count.
  - Out-of-order execution and on-chip caching and prefetching reduce the memory latency bottleneck at the expense of using more transistors and increasing processor complexity.
  - Pollack's Rule: performance increases due to microarchitecture techniques are square root of the number of transistors or the area of a processor.

**IBM SyNAPSE Chip** (Aug, 2014) [^1]
- A neurosynaptic computer chip with 1 million programmable neurons, 256 million programmable synapses and 46 billion synaptic operations per second per watt, operating at a minuscule 70mW.
- On August 7 2014, IBM announced their second generation SyNAPTIC chip, which contains the most transistors in a Neurosynaptic chip to date at 5.4 billion.[3]

**Stanford’s Neurocore chip: Neurogrid** [4,5] (April, 2014)
- Stanford bioengineers developed a circuit board modeled on the human brain:
  - 16 custom designed Neurocore chips simulate 1 million neurons and billions of synaptic connections.
  - Neurogrid is claimed to be 9,000 times faster and more energy efficient than a PC.
How would we as Computer Scientists engineer hyper-intelligent or conscious machines?

• Caution! None of the reviewed articles ever spoke of the consequences of BAD data $\rightarrow$ ultimate-GIGO!
  – Conclusions/decisions/computations are as good as the quality of the input data.

• Vernor Vinge’s options: (d)
  – Create super-human AI in computers
  – Human intelligence amplification through novel UI.
  – Biomedical: enhance neurological operation of our brains.
  – The Internet becomes self-aware.

• What is the best way to build a conscious machine?
  – Copy the mammalian brain (harder)
  – Evolve a machine (easier).
Should building the Singularity be a goal of computer science?

• This is the best time to be in computer science!
  – Innovations needed at all levels of abstraction.
  – Lots of exciting research across full HW / SW stack.

• Brooks’ development program.

• Integrated Information Theory of Consciousness.
  – http://www.biomedcentral.com/1471-2202/5/42/

• **Yes**, guided by scientific measurement, replicating fail-safe mechanisms, *managing* complexity & reliability.

• **No**, specific objections follow.
Nanotechnology

• Richard Jones: Rupturing the Nanotech Rapture (d)
  – Molecular nanotechnology reduces all material things to the status of software.
  – Molecular machines can do parallel processing (duroquinone).
  – Complex nano-assembly is certainly possible, but the watery nanoscale environment of cell biology is hostile to engineering!
  – Biology: lack of rigidity, excessive stickiness/friction, constant random motion, strong surface forces: propels the self-assembly of sophisticated structures.
  – The range of environments in which rigid nano-machines could operate is quite limited.
  – DNA itself can be used as a construction (self-assembly) machine to make programmed structures.
  – Brain-interface systems have already been built (256 neurons)
  – New polymer, nano-particle based photovoltaics \( \rightarrow \) solar cells.
• Critique of Kurzweil’s 2005 book *The Singularity is Near: When Humans Transcend Biology*.

• Main factors necessary for the achievement of AI:
  – AI is the main tool for the paradigm shift required by the singularity.
  – Computational power is *the* requirement for the development of AI.

• Preconditions for the continuity of exponential trends:
  – continuity of economic growth
  – limited energy usage
  – availability of enabling knowledge

• Demand side: trend in energy consumption on computer CPU’s has been decreasing in the last 14 years.

• Supply side: Kurzweil’s reliance on photovoltaics as a major future source of energy seems questionable due to the slow rate of improvement this technology has had during recent years.
Bill Joy: *Why the Future Doesn’t Need Us (b)*

- Risk of human extinction quite real.
- Limit development of machines that *reproduce*: nano-technology, robotics, synthetic organisms.
- *Unintended consequences*, a well-known problem with the design and use of technology, because the systems involved are *complex*, involving interaction among and feedback between many parts.
- We’ve been building these types of systems for at least 40 years!
A Dystopian Scenario: The Age of Spiritual Machines

THE NEW LUDDITE CHALLENGE

"First let us postulate that the computer scientists succeed in developing intelligent machines that can do all things better than human beings can do them. In that case presumably all work will be done by vast, highly organized systems of machines and no human effort will be necessary. Either of two cases might occur. The machines might be permitted to make all of their own decisions without human oversight, or else human control over the machines might be retained.

If the machines are permitted to make all their own decisions, we can't make any conjectures as to the results, because it is impossible to guess how such machines might behave. We only point out that the fate of the human race would be at the mercy of the machines. It might be argued that the human race would never be foolish enough to hand over all the power to the machines. But we are suggesting neither that the human race would voluntarily turn power over to the machines nor that the machines would willfully seize power. What we do suggest is that the human race might easily permit itself to drift into a position of such dependence on the machines that it would have no practical choice but to accept all of the machines' decisions. As society and the problems that face it become more and more complex and machines become more and more intelligent, people will let machines make more of their decisions for them, simply because machine-made decisions will bring better results than man-made ones. Eventually a stage may be reached at which the decisions necessary to keep the system running will be so complex that human beings will be incapable of making them intelligently. At that stage the machines will be in effective control. People won't be able to just turn the machines off, because they will be so dependent on them that turning them off would amount to suicide." <more>

Who wrote this? Theodore Kaczynski in his Unabomber manifesto.
Does building the Singularity imply the end of computer science as we know it?

• Theoretically, yes.
  – Machines directly programming other machines
  – No source code, compilation, linking.
  – No human intervention.

• Practically, no.
  – Huge amount of remaining work
Conceptual Summary

• Moore’s law. Dennard’s scaling law.
• Consciousness/sentience. Intelligence.
• Exponential growth rates. Extrapolation
• Reproducing machines.
• Complexity, reliability.
• Our need to preserve our mental identities -- myself only better.
• New tradeoffs to consider.
Slide References

- [http://en.wikipedia.org/wiki/Moore%27s_law](http://en.wikipedia.org/wiki/Moore%27s_law)
- Machine Intelligence Research Institute (MIRI) [http://intelligence.org/](http://intelligence.org/)
- Location of this pdf: [http://cse.ucdenver.edu/~dgnabasik/Food%20for%20Thought%20II%20-%20Singularity.pdf](http://cse.ucdenver.edu/~dgnabasik/Food%20for%20Thought%20II%20-%20Singularity.pdf)
- 12 - [*] http://www.biomedcentral.com/1471-2202/5/42/
- 16 - Moore's second law, also called Rock's law, which is that the capital cost of a semiconductor fab also increases exponentially over time.
- 17 - As of 2014, the highest transistor count in a commercially available CPU is over 4.3 billion transistors, in Intel's 15-core Xeon IvyBridge-EX. [http://en.wikipedia.org/wiki/Transistor_count](http://en.wikipedia.org/wiki/Transistor_count)
- 23 - [*] The chip started out as a software model running in one of the "Blue-Gene" series of super computers. It was then moved into FPGAs a few years back, and now into its own ASIC implemented using a 28nm process.