

The AI Singularity

Prospects for Digital Immortality

Rich Sutton

Reinforcement Learning & Artificial Intelligence Lab
Dept of Computing Science

The Technological (AI) Singularity

It seems plausible that with technology we can, in the fairly near future, create (or become) creatures who surpass humans in every intellectual and creative dimension

—Vernor Vinge, 1993, 2010

When the smartest thing on the planet is designed rather than evolved

Questions for today

- ✓ What is the AI singularity?
 - Will it occur?
 - When will it occur? In our lifetimes?
 - ~~Is it a good thing?~~
 - What are the implications for us?
 - Can people be made digital?

Will the AI singularity occur?

First, is it possible?

- If people are biological machines, then eventually we will reverse engineer them, and understand their workings
- Then, surely we can make improvements
 - with materials and technology not available to evolution
 - how could there not be something we can improve?
 - design can overcome local minima, make great strides, try things much faster than biology

Will the AI singularity occur?

Second, if it can be done, will it be done?

- Not if we destroy ourselves first of course
- If that doesn't happen, then there will be strong, multi-incremental economic incentives pushing inexorably towards an AI singularity
- It seems unlikely that they could be resisted
 - or successfully forbidden or controlled
 - there is too much value, too many independent actors

The Big Picture:

Technology as the 4th 'age' of the universe

The Age of Particles

0 The big bang
100s formation of elementary particles
1h formation of light (hydrogen and helium) nuclei
1My+ formation of light atoms
formation of gas galaxies

The Age of Stars

100My formation of first generation stars
formation of heavier elements within stars
stars explode and expel heavier elements
formation of planets from heavier elements
9By formation of our solar system

The Age of Replicators

13By origin of life on earth (formation of first replicators)
RNA and DNA
14By sexual reproduction
multi-cellular organisms
nervous systems
1Mya humans
culture

The Age of Design

100Kya language
10Kya agriculture, metal tools
5Kya written language
200ya industrial revolution
technology
70ya computers
artificial intelligence
singularity
transcendence

...

Questions for today

- ✓ What is the AI singularity?
- ✓ Will it occur?
 - When will it occur? In our lifetimes?
 - ~~Is it a good thing?~~
 - What are the implications for us?
 - Can people be made digital?

Evolution of Computer Power/Cost

MIPS per \$1000
Billion (1998 \$)

- Moravec's law:**
- Effective computation/\$ increases exponentially, doubling in 18-24 mos
 - This has held for 60+ years
 - And will continue for the foreseeable future

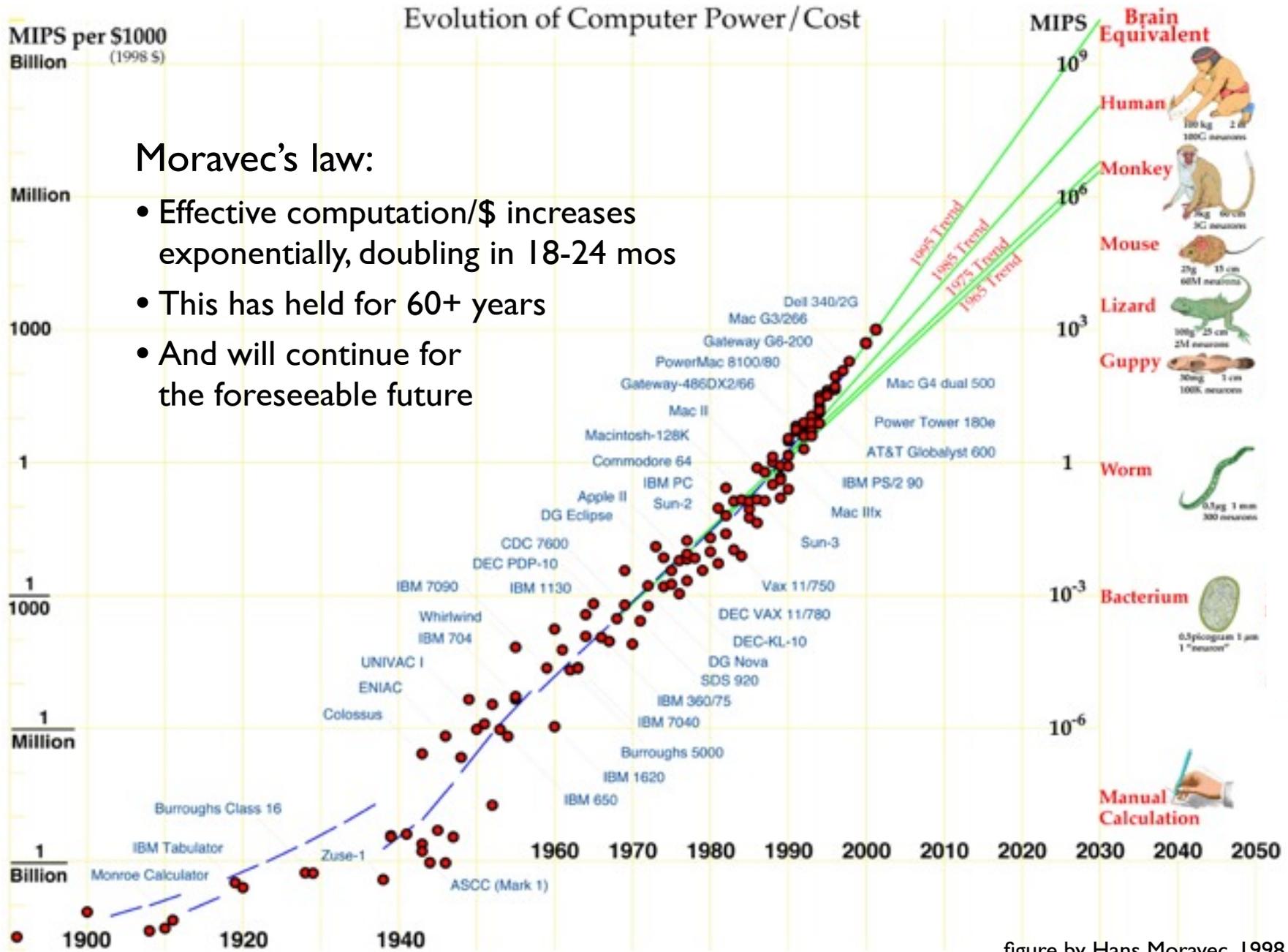


figure by Hans Moravec, 1998

The AI Singularity is near

— [“We are nearing an important milestone in the history of life on earth, the point at which we can construct machines with the potential for exhibiting an intelligence comparable to ours.” – David Waltz, 1988

— [Should occur in ≈ 2030 for $\approx \$1000$

— [We don't currently have the needed AI software (design)

— [But the hardware is and will be a tremendous economic spur to its development...perhaps at nearly the same time

Questions for today

- ✓ What is the AI singularity?
- ✓ Will it occur?
- ✓ When will it occur? In our lifetimes?
- ~~Is it a good thing?~~
- What are the implications for us?
- Can people be made digital?

Implications for AI?

— [Computation will continue to get vastly cheaper

— [In the long run, computation and data are more powerful than trying to leverage imperfect human understanding

— witness the demise of classical AI, and the rise of machine learning, search, and statistics

— [Methods that use massive computation *efficiently* will become more and more important

— [The future will be interesting for computer scientists

Implications for us?

— [We can help make the AIs

— [We can profit from them

— [We can become them

— [We probably can not

— prevent them

— or control them

— [We will have to ride the storm

Prospects for digital immortality

- Can people be made digital?
 - New people? Als certainly.
 - Existent people?
- If so, then it seems that they can live indefinitely
 - digitization is key to all kinds of longevity
 - enables backups

Mental digitization

- The transfer of the essence of a biological person's mind to a new physical substrate, a brain that is digital in that all of its state can be recorded and reproduced at a later time without significant loss
- That would be nice, but...
 - Is it at all possible? Or is it nonsense?

Is digitization a medical issue?

- Absolutely, it is the ultimate medical issue
 - How to preserve life and function despite illness, health failures, and design errors
 - No doctor should deny it to a patient in need
- Can we offer the digital cure?
Would we want it?
- I think it is obvious that many would want it

Is the essence of a mind bound to its physical substrate?

- No
- All the atoms in a person's body are replaced every few years
 - Many of cells are replaced as well, but others remain
 - The digital part -- the DNA -- is retained even as the atoms are changed
- We are not our atoms... we are the pattern in our atoms

Must the pattern be reproduced exactly?

- No
- Our pattern is not the same as it was 10 years ago, or when we were a child
- These changes are essential to living, and do not diminish our sense of identity
- Identity is like horseshoes: Close should be good enough

Digital life need not be linear

- We can make backups... and restore after death
- We can make copies... and merge
 - body may become more precious than mind
 - many more different senses of identity are possible
- We can pause... and thus travel forward in time
- We can experiment, change ourselves then restore

The dangers of self design and augmentation

- What would we change?
 - sensory and motor abilities of course
 - but also mental abilities, concentration, focus, memory, internet access, and more
- What if we change our goals?
 - to avoid addictions, obsessions, bad memories
- Will we lose our sense of self?
- What if we split and don't want to remerge?

So, the prospects for digital immortality are...good

- No clear reason why digitization of existing people should be nonsense
- Yet it would be very desirable, medically
- The most crazy/confusing part is what happens afterwards, with backups and copies, splitting and merging
- Maybe we won't even care about immortality; maybe identity won't have the clear, compelling meaning that it has for us now

Questions for today

- ✓ What is the AI singularity?
- ✓ Will it occur?
- ✓ When will it occur? In our lifetimes?
- ~~Is it a good thing?~~
- ✓ What are the implications for us?
- ✓ Can people be made digital?