Introduction to Reinforcement Learning

Rich Sutton
Reinforcement Learning and Artificial Intelligence Laboratory
Department of Computing Science
University of Alberta, Canada

Part 1: Why?

The coming of artificial intelligence

- When people finally come to understand the principles of intelligence—what it is and how it works—well enough to design and create beings as intelligent as ourselves

- A fundamental goal for science, engineering, the humanities, …for all mankind

- It will change the way we work and play, our sense of self, life, and death, the goals we set for ourselves and for our societies

- But it is also of significance beyond our species, beyond history

- It will lead to new beings and new ways of being, things inevitably *much <u>more powerful</u> than our current selves*

Milestones in the development of life on Earth

	year	Milestone	
	14Bya	Big bang	
	4.5Bya	formation of the earth and solar system	
The Age of Replicators	3.7Bya	origin of life on earth (formation of first replicators)	Self-replicated things most prominent
		DNA and RNA	
	1.1Bya	sexual reproduction	
		multi-cellular organisms	
		nervous systems	
	1Mya	humans	
		culture	
	100Kya	language	
	10Kya	agriculture, metal tools	
	5Kya	written language	
The Age of Design	200ya	industrial revolution	Designed things most prominent
		technology	
	70ya	computers	
		nanotechnology	
	?	artificial intelligence	
		super-intelligence	
		…	

~~When will super-human AI come?~~

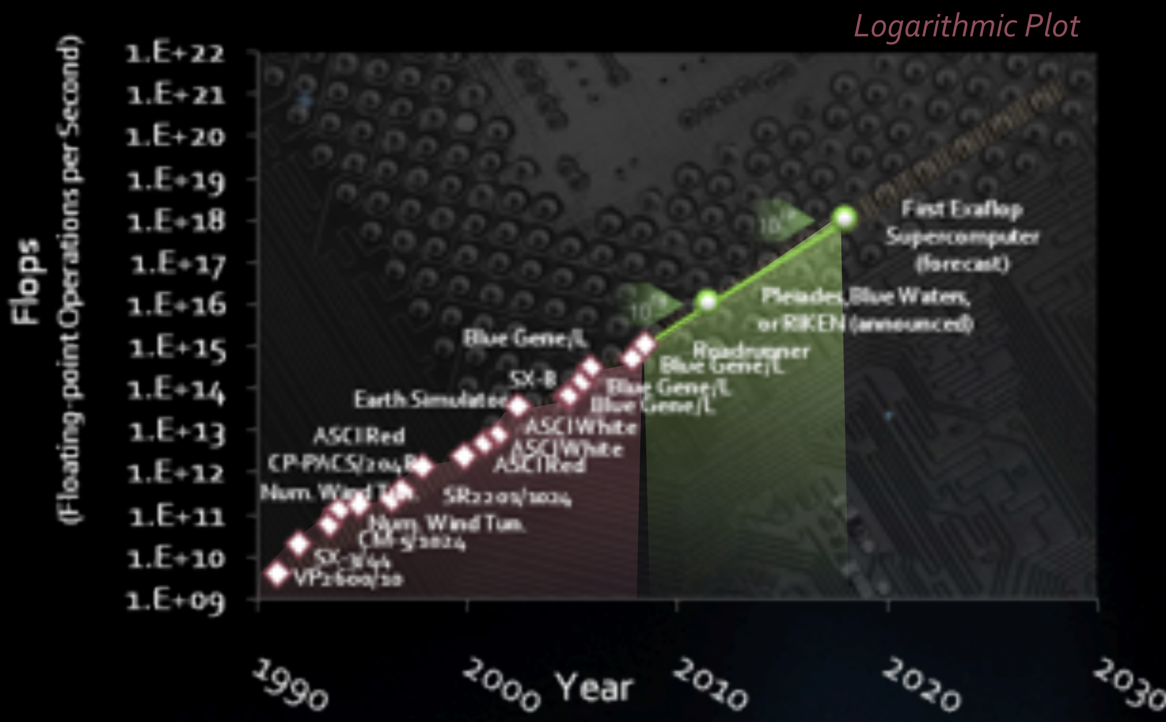
Might it come in our lifetimes?

Should we include its possibility
in our research and career plans?

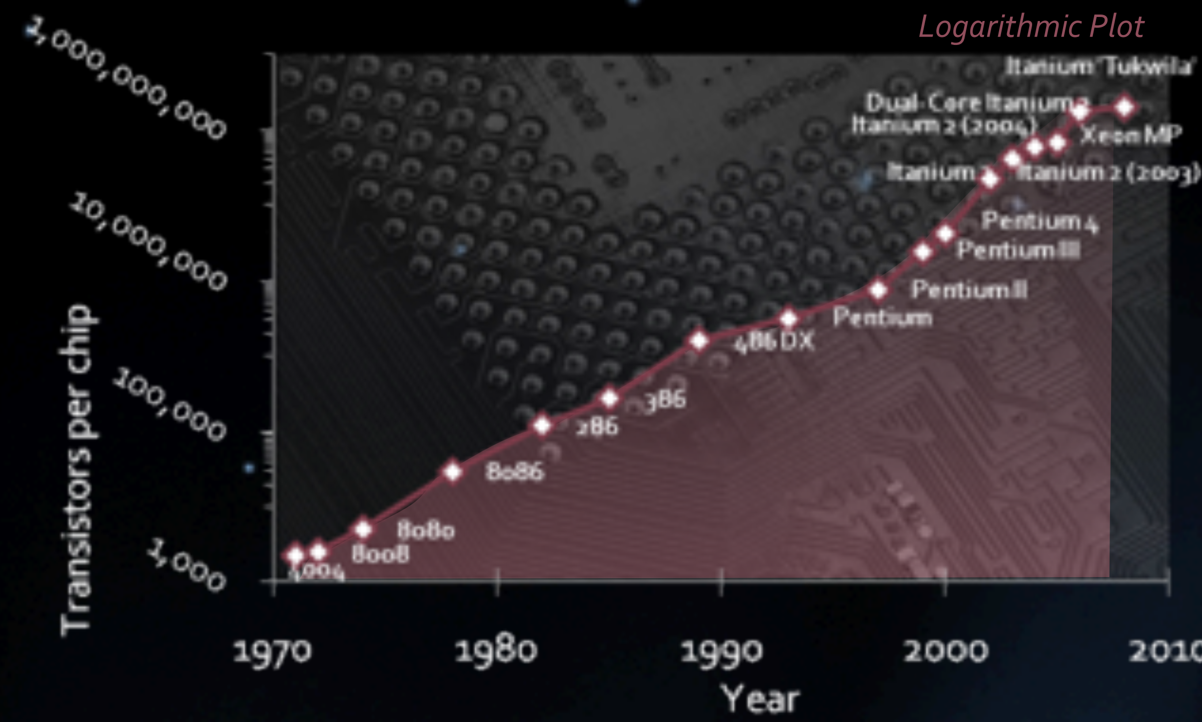
the computational mega-trend

- effective computation per $ increases exponentially, with a doubling time of 18-24 months
- this trend has held for the last sixty years
- and will continue for the foreseeable future

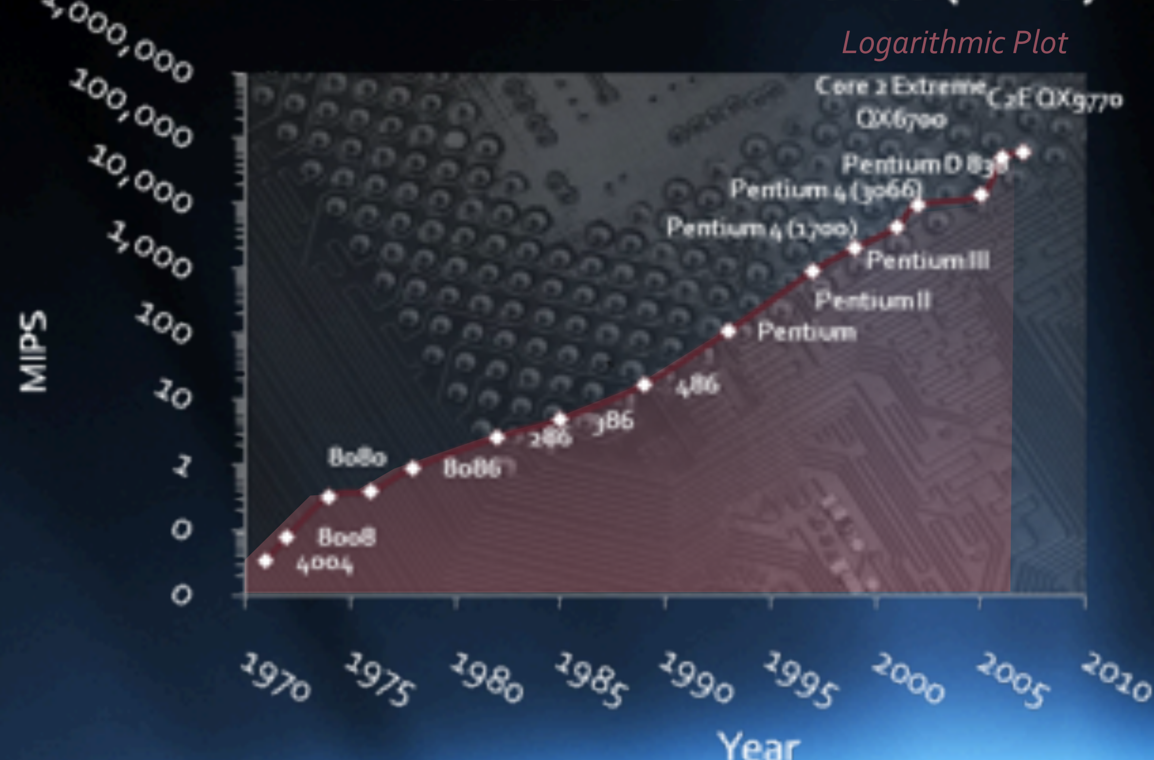
Growth in Supercomputer Power



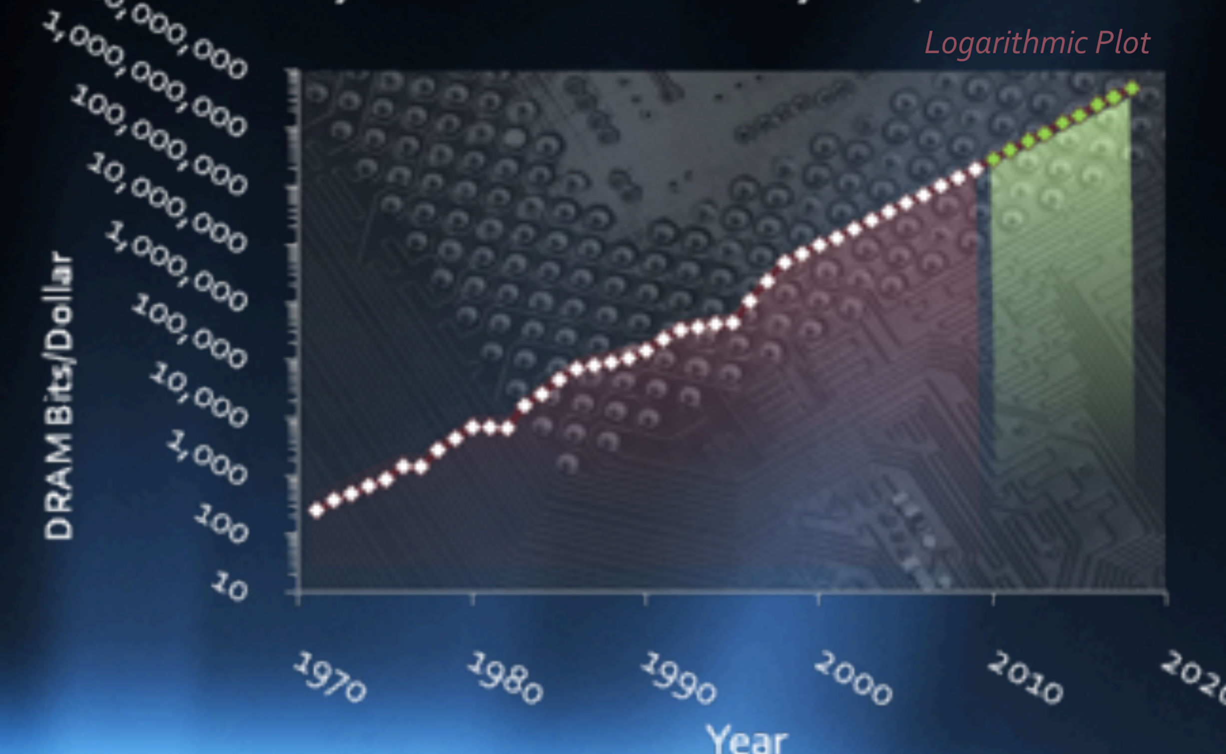
Transistors Per Chip (Intel processors)



Processor Performance (MIPS)



Dynamic RAM Memory Bits per Dollar



The computational mega-trend

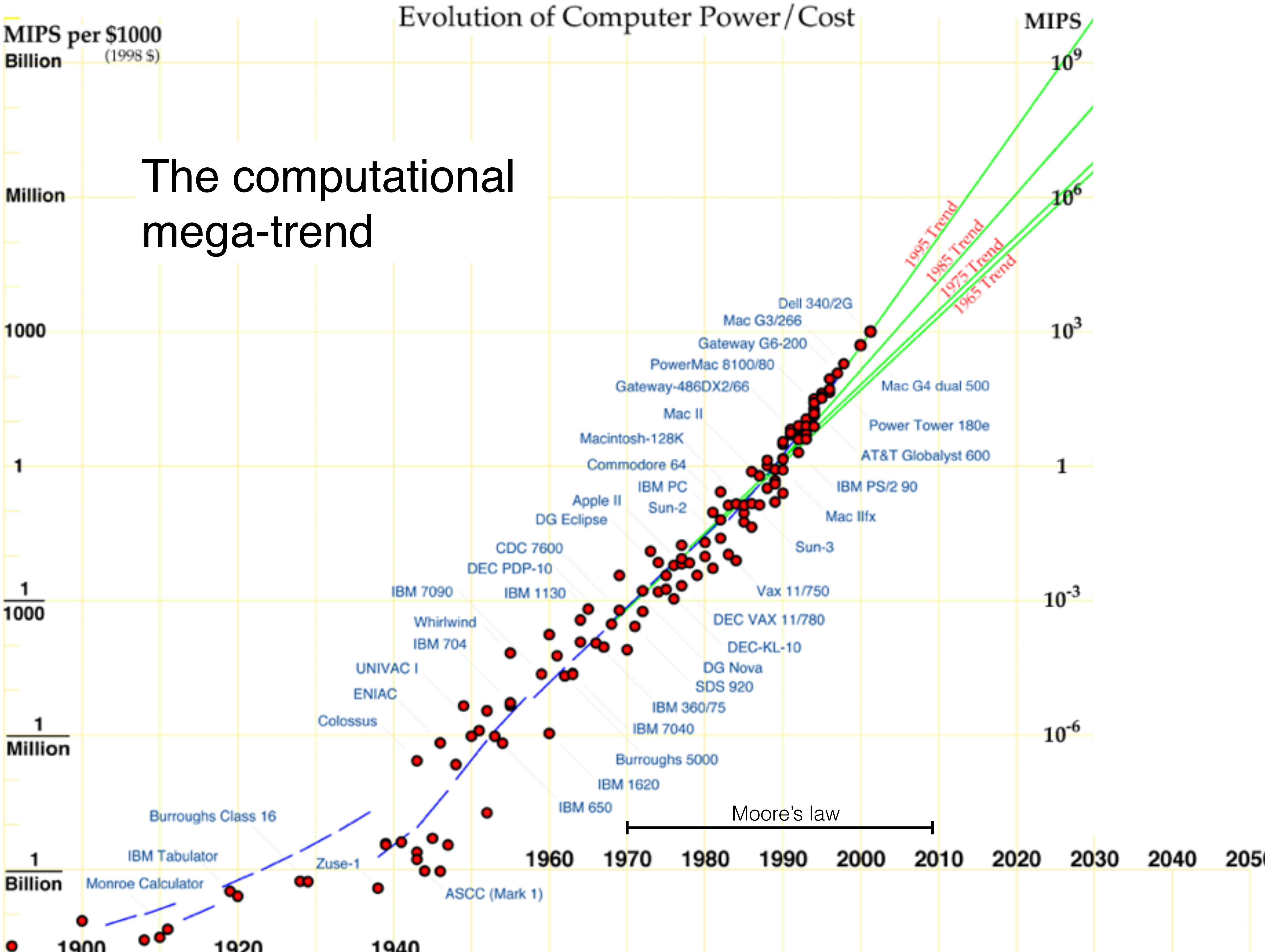


figure by Hans Moravec, 1998

The possibility of AI 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 (recent president of AAAI)
- Should occur in ≈2030 for ≈$1000
- We don’t yet have the needed AI “software” (designs, ideas)
- But the hardware will be a tremendous economic spur to development of the ideas...perhaps at nearly the same time

What’s an AI researcher to do?

- The prize—understanding mind—is great
    - And may be within reach
- How then can we do something relevant to acquiring the prize?
    - What strategy should we follow?
    - What subproblem should we address?
    - which issues should we consider, and which should we ignore?

Implications for AI

- Diminishes the importance of special-case solutions
- Amplifies the importance of general “smart force” solutions
- Data and computation becomes more important than leveraging human understanding
  - witness the demise of classical AI
    and the rise of machine learning, search, and statistics
- Scalability of an algorithm—the ability to leverage more computation—becomes crucial

Do computational costs matter?

- As computation becomes vastly cheaper (but not free), do computation costs become less important?
  - No. Quite the opposite
- Computational costs become more, not less, important
  - because demand is essentially infinite
  - performance will depend on how efficiently computation is used; other influences will decline in relative import

We have seen this story before

- In chess

  we thought human ideas were key, but it turned out (deep Blue 1997) that big, efficient, heuristic search was key

- In computer Go

  we thought human ideas were key, but it turned out (MCTS 2006–) that big, sample-based search was key

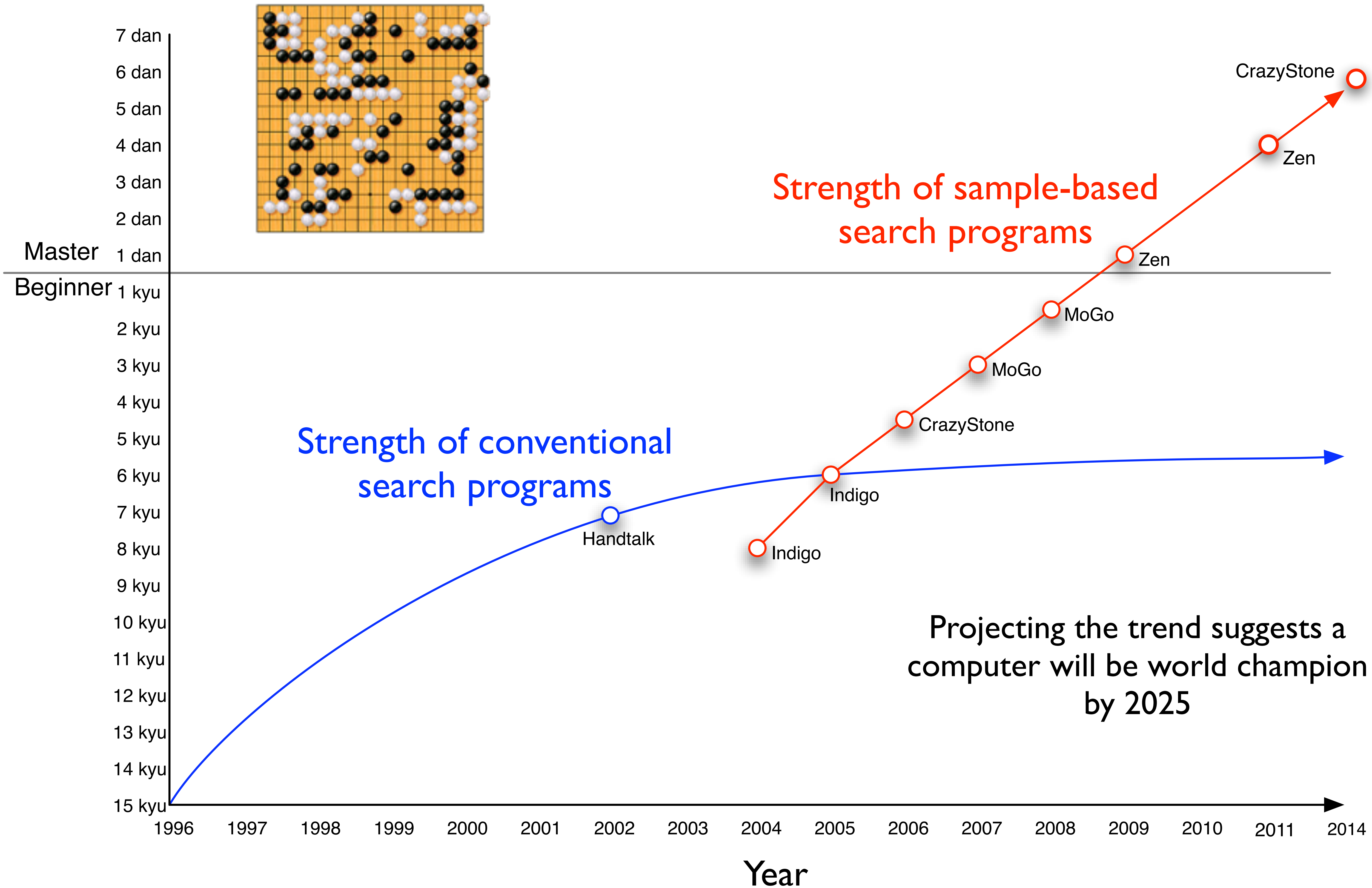
- In natural language processing

  we thought that human-written rules were key, but it turned out (~1988) that statistical machine learning and big data were key

- In visual object recognition

  we thought human ideas were key, but it turned out (deep learning 2012–) that big data sets, many parameters, and long training was key

Steady, exponential improvement (since MCTS, 2005) in the strength of the best computer Go programs



Two diverging approaches to AI

Weak methods (general)

- Strong automation (computation & data rule)
  - The AI gradually constructs an understanding of the world from data; this understanding may be opaque to the researcher
  - Success is defined in a self-verifiable way, for example, as predicting and controlling the data

Strong methods (specific)

- Engineering/design (human understanding rules)
  - The researcher gradually understands the possible worlds, and makes the AI behave appropriately in each
  - Success may be defined in terms unavailable to the AI, for example, as labelling objects in the same way a person would

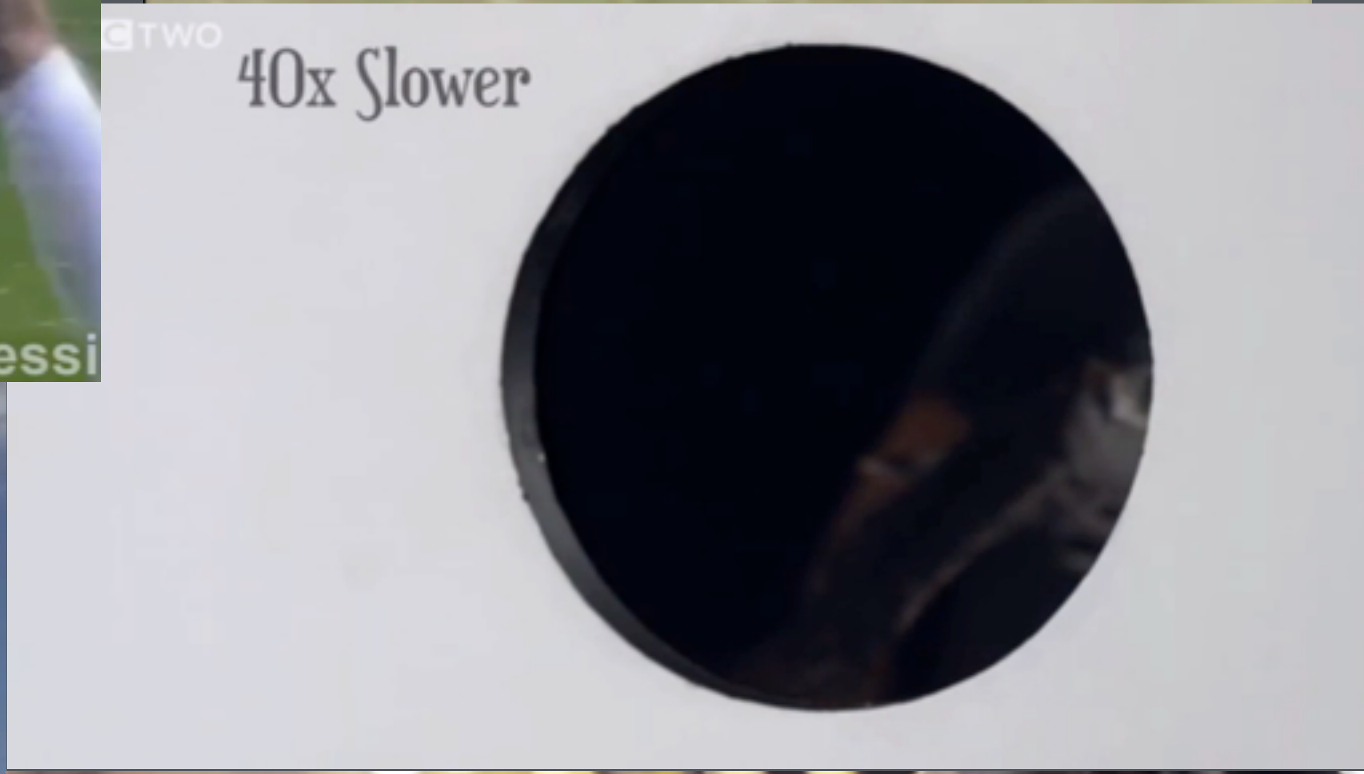
A new view of the AI problem

and its implications for (and against) solution methods

- Minds are real-time information processors interacting with a firehose of data from a complex and arbitrary world
    - we must find *scalable* and *general* methods, to *learn* arbitrary stuff (no domain knowledge, no taking advantage of structure)
- We have immense computational resources, but it's never enough; the complexity of the world is <u>always vastly greater</u>
    - we seek *computationally frugal* methods for finding *approximate* solutions (optimality is a distraction; relying on it is untenable)
- We have immense data, but not labeled examples
    - we must be able to learn from *unsupervised* interaction with the world, a.k.a. *self-labelling* (no human labels, not even from the web)

the mind's first responsibility is
real-time sensorimotor information processing

- Perception, action, & anticipation
  - as fast and reactive as possible



AI slowly, tectonically, shifting toward scalability

Increasing desire for:

- Generality
- Approximation
- Massive, efficient computation
- Learning without labels

The longest trend in AI

- There have always been two general approaches/directions
  - *Design*. Use our intuition about how our intelligence works to engineer AIs that work similarly; leverage our human design abilities
  - *Meta-design.* Design only general principles and general algorithms; leverage computation and data to determine the rest
- My reading of AI history is that the former has always been more appealing in the short run, but the latter more successful in the long run

Reinforcement learning (RL) and temporal-difference learning (TDL) are consilient with the new view

- RL is learning to *control* data
- TDL is learning to *predict* data
- Both are weak (general) methods
- Both proceed without human input or understanding
- Both are computationally cheap and thus potentially computationally massive