

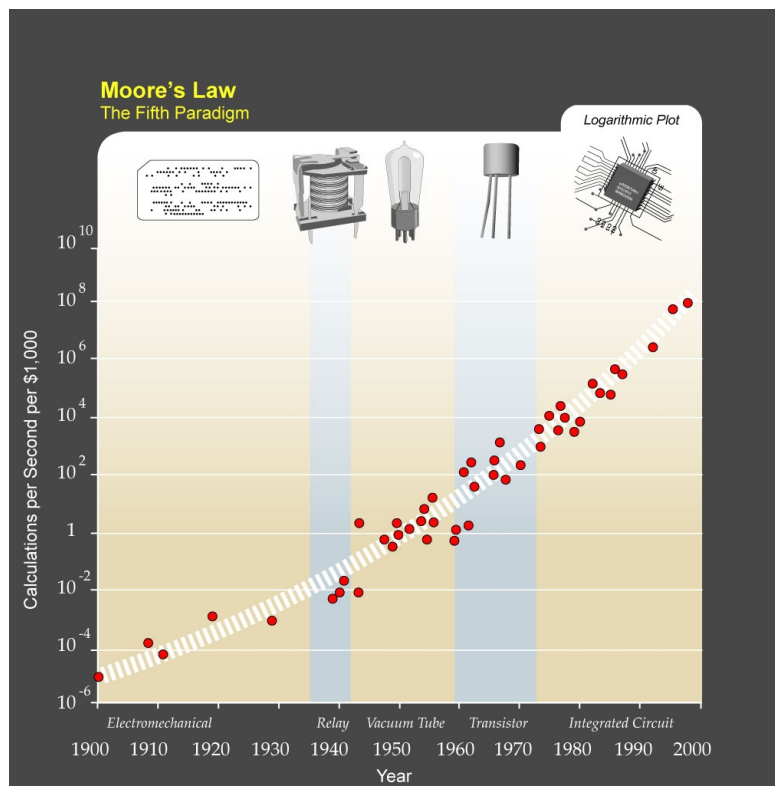
The Singularity

Michael Maynard
CMSC 828D
Fall 2012
Report 4

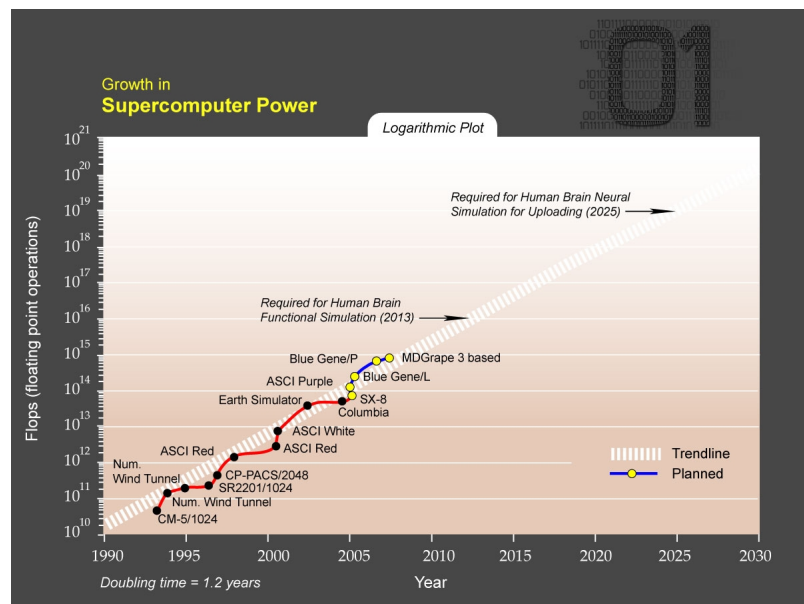
The technological singularity¹, or just the singularity, is a hypothetical point in the future of technological development beyond which we, humans in 2012, cannot make accurate predictions. More specifically, after the singularity, the nature of the human experience may have changed in ways which are today incomprehensible. The name comes from an analogy with the event horizon of black holes – black holes are singularities and nothing beyond the event horizon can be perceived. The idea is that technology evolves exponentially, and exponential growth appears mild towards its beginnings, but at some point appears to “explode”. This “explosion” is the beginning of the singularity - note that the “singularity” is not in actuality a singularity, there is no asymptote. The driving force behind the singularity is postulated to be an ever upward trend in intelligence(biological and/or AI).

Two people in particular have promoted the idea of an impending singularity; Mathematician and science fiction writer Vernor Vinge, and inventor and futurist Ray Kurzweil, though the term originated in the 1950's with John von Neumann. Vinge postulates that a singularity could arise out of any of: AI, human biological enhancement, or brain-computer interfaces. Kurzweil has plotted the trends of various technologies, observed that they follow a pattern of exponential growth, and argues that this trend will continue to and past the point of singularity.

Kurzweil has proposed what he calls the “law of accelerating returns”. This is essentially Moore's law applied to technologies beyond semiconductor circuits. The bellow plot illustrates²:



With respect to supercomputer performance we see exponential growth as well²:



The assertions “Required for Functional Brain Simulation” and “Required for Human Brain Neural Simulation for Uploading” are not without contention. However, even if Kurzweil's predictions of the computational capacity necessary for functional brain or neural simulation are off by several orders of magnitude, assuming sustained exponential growth in computational capacity, this will only have the effect of shifting the time at which such capacity is reached back a few decades. Many or most of those who argue that a coming singularity is probable place it at some point in the 21st century.

Given a sustained exponential growth of technological capacity, it seems inevitable that something analogous to a singularity will occur at some point in the future. With respect to an intelligence explosion specifically, there are various scenarios through which a singularity could arise. One is the development of intelligent software, either AI or uploaded minds, of intelligence comparable to human intelligence. This software intelligence would then design hardware of a greater speed than that on which it currently runs, and then transfer itself to this new hardware. The subjective experience of time by the intelligent software would then be compressed, allowing it more subjective time per unit objective time with which to work on any given problem. The effect would be that to an outside observer the software would then appear more intelligent. The intelligent software would be able to engage in a cycle of recursive self-improvement, creating ever faster hardware on which to run, and ever compressing its subjective time.

A second scenario is intelligent software which, rather than constructing better hardware, constructs better intelligent software. Similarly to the first scenario, this intelligent software would be capable of entering a recursive loop of self-improvement where ever more intelligent agents were designed, not by virtue of compressed subjective time, but by virtue of better intelligence architectures.

Before a self-improving loop can begin in the above scenarios, we must construct a software agent with intelligence comparable to that of a human. Such a software agent could be constructed as a result of research into AI – that is, today's research could eventually lead to an understanding of 'intelligence' sufficient to replicate it in a software agent. Such an agent could also be constructed simply by mimicking the internal workings of currently existing intelligent agents, humans. There are currently projects working towards an understanding of the architecture of the human brain sufficient in detail for replication on a machine; specifically see the Blue Brain and connectome projects. If a brain can be simulated in sufficient detail, or a human mind uploaded onto a simulated neural architecture, intelligence need not be fully understood in order to replicate and improve upon it.

The intelligent software scenario seems the most common in contemporary discourse, though there are other scenarios. Through genetic engineering for instance, it may be possible to construct organisms, humans perhaps, of greater intelligence than current humans. These organisms could then enter a loop of recursive self-improvement similar to the above scenarios. Another scenario is an increase in intelligence emerging from the merger of biology and technology. Neural implants could in principle be constructed which either give greater cognitive capacity directly, or serve as links to computers to which sub-tasks and queries could be siphoned off by their biological users. Such enhanced humans could then enter a loop of recursive self-improvement similar to the above scenarios.

How precisely a technological singularity would play out is not at present apparent, however some kind of singularity would seem to be the inevitable result of sustained exponential technological growth. Thus, many of those who wish to argue against the possibility or probability of something as apparently outlandish as a technological singularity which has the potential to change the human condition in radical ways, argue against the likelihood that the exponential growth of our technological capacity can be sustained.

One valid argument against the singularity is that extrapolating current trends so far into the future, to the point of singularity, is invalid reasoning, we have no way of projecting that far and no reason to think that exponential growth will be sustained. The logic used in support of the singularity is vulnerable to a *reductio ad absurdum* counter-argument by applying that logic to other domains. “The Economist” for instance pointed out¹ that extrapolating the increase of number of blades on a razor leads to razor blades consisting of perhaps thousands of blades in the not so distant future. Proponents of the singularity acknowledge that exponential growth cannot be sustained indefinitely, but argue that the point at which growth ceases to appear exponential lies well beyond the singularity.

The possibility of a singularity presents potentially existential risks. Humans dominate their biological peers because of their intelligence, and if entities of superhuman intelligence were constructed, then they may come to be the dominant entities on earth. If such entities' desires do not align with the interests of humanity, humanity could face extinction. There are ways of attempting to mitigate such risks, including restraining the capabilities of super-humanly intelligent entities, or increasing the intelligence of humans, perhaps through cognitive implants, concurrently with that of AI's. If at some point in the future an intelligence explosion appears imminent, the situation will need to be dealt with with delicacy.

In summary, our technological capacity has for some time been experiencing exponential growth. There are those who argue that this trend will continue for some time, and that probably some time in the 21st century we will hit the “knee of the curve” of this exponential growth. If this occurs, the result will be something analogous to a singularity in that today we are unable to see beyond its “event horizon”. Such a singularity presents great potential and great risk, at one extreme everyone could live forever as greatly enhanced humans, at the other extreme the human race could go extinct.

Sources:

1. http://en.wikipedia.org/wiki/Technological_singularity

2. http://en.wikipedia.org/wiki/Accelerating_change#Kurzweil.27s_The_Law_of_Accelerating_Returns