Chapter 1

The Pursuit of Machoflops

The Rise and Fall of High-Performance Computing

Anne C. Fitzpatrick

Nearly twenty-eight years ago, I first encountered the writings of Carl Hempel, Rudolf Carnap, Sir Karl Popper, and other members of the Vienna Circle school of philosophy in Joseph C. Pitt's mandatory-for-science and technology studies (STS) philosophy graduate course at Virginia Tech. Reading Popper's *The Logic of Scientific Discovery* (1968) was hard going the first time.[1] The logical positivist syllabus that comprised Joe's class was the most daunting one in graduate school, with a paper due weekly; students largely analyzed the Vienna Circle school of philosophical thought and some other, more modern writings about science and technology. Carl Hempel's tome, *Aspects of Scientific Explanation* (1965), was a much-dreaded-by-students rite of passage to get through Joe's first-year course. But it challenged you and made you stronger intellectually and pushed your writing skills. If you could get through the logical empiricist philosophers, you could get through any course. Being more a student of history than philosophy, I subsequently read Popper's disquisition, *The Open Society and Its Enemies* (1994), on my own to try and understand what the Vienna Circle really stood for.

Today, nearly a century after they were written, the works of Popper and several of his other Viennese intellectual contemporaries are again becoming increasingly relevant. There is a renewed appreciation for their writings now in a time when our own democracy seems at risk, and science—especially climate change—is dismissed as political and unfounded, and the United States government is diminishing its general support for science. What I did not understand when initially exposed to the works of Popper, Quine, Carnap, and Hempel was that their writings were essentially a response against the rising tide of fascism and tyranny in Europe in the 1930s.

Many members of the Vienna Circle had to flee their beloved Vienna and emigrate abroad to find safety. Popper wrote *The Open Society* after Hitler invaded Austria. In *The Open Society*, Popper advocated that central direction is not the way to govern a society, but rather competition for ideas supported by critical thinking would lead to a better way of life. A lot of his beliefs were based on his earlier works that explored the scientific method where hypotheses are produced and advanced, and scientists try to falsify them, so any hypothesis that remains unfalsified must stand as some kind of credible knowledge. This notion carried forward into a concept of truth for Popper in *The Open Society*, where he argued that blind deference to great men and grand theories of history was dangerous to modern, civil, active participant society.

In this chapter, I will review and analyze a slice of the high tech government-industry partnership that builds America's high-performance computers (HPC, also called supercomputing), where I have worked for some time, and demonstrate how Joe's work in the philosophy of technology has deeply influenced my approach to my job. As part of a modern, civil, and open society and as a federal employee and leader, I need credible knowledge to make the most informed decisions I can when spending public money. Because of this, I find myself thinking a lot not only about Joe's work, but also about Popper and the logical empiricists' writings. I do this both in terms of their relevance to our American democracy today, and also in seeking credible scientific and technological knowledge in my daily work over the past two decades in the information technology and national security and policy realms, which when executed correctly is intended to support our democracy.

The logical empiricists' writings were not without their flaws, as postmodern and social constructivist scholars in the middle and latter twentieth century pointed out. In the STS coursework, we learned that women and minorities were usually left out of histories, and perhaps science and technology could be seen through relativistic lenses and were up to interpretation (Haraway, 1991; Rossiter, 1982). STS, as an outgrowth in part of Philosophy of Science, has come a long way as a discipline since academics began to question what science is, how it is conducted, and where science and technology fit into society and what their relationship to one another is. Now, more than ever, we need to figure these relationships out.

My STS education, the logical empiricists, and other philosophical points of view influenced my choice of career, while Joe's philosophical writings about technology, in particular, have shaped both my analytical work on technological futures and my efforts to bring closer together the "two cultures" of—for our purposes here—Science, Technology, Engineering, and Math (STEM) and non-STEM specialists (Snow, 1962). For some years I've supported efforts to help the United States government avoid technological surprise to ensure that we remain on the leading edge of HPC. My office colleagues and I have watched how over the past couple of decades China has been making huge strides in HPC, even surpassing the United States at times in sporting the fastest machines on the planet. Until November of 2018, China had the fastest supercomputer in the world: the TaihuLight, at the National Supercomputing Center in Wuxi. TaihuLight's microchips are all Chinese-designed—a landmark achievement for China, considering that prior to 2000 it had no computers on the biannual and increasingly questionable Top500 ranking of the fastest HPCs in the world.[2]

Today, China has 226 HPCs on this list. I'll return to the subject of HPC as a real-life STS case study later in this chapter, but first, I need to provide some context based on Joe's ideas about technology and the human condition. Advances in supercomputing matter a great deal both as they are an indicator of a nation's economic competitiveness and because of how pervasive information technology has become. Because of this, I find myself thinking about Joe's notion that as a society we are all dependent on a technological backbone, which is rapidly becoming more and more of a digital backbone that weaves through nearly all facets of modern life (Pitt, 2000). Joe would even say technology—or just "Tech" as it is increasingly called—is more than pervasive in modern life, where it defines the human condition today (Pitt, 2000).

Joe was right. It's easy to see the pervasiveness of tech where the speed of change and innovation has become so blindingly fast. Joe foresaw what is becoming obvious to all now—that tech is a way of life today, especially in more affluent communities and parts of the world. From smart fridges in our homes to driverless cars in cities and our phones essentially extensions of our bodies and personalities, technology defines the human condition in the developed world: we're often expected to have work-issued phones and laptops, and to be available during nonstandard hours; we come home at the end of the day and without a thought check work email and order pizza via Amazon Echo without touching a button or talking with a real person. For many people, there is little "turning off" from all this unless one deliberately and mindfully chooses to, and in today's world that is not always easy given the dependence we have on these systems.

The last twenty years or so have witnessed a critical transition from the twentieth-century US-led military-industrial complex, which is by no means dead, to the globalized Big Tech economy, dominated by Silicon Valley and its Chinese competitors and suppliers. Facebook, Amazon, Apple, Netflix, and Google (acronymized as the FAANG stocks by investors) are collectively valued at over $3 trillion as of 2018. At the time of writing this chapter, these and other competitors are sitting on unprecedented amounts of cash. These companies' influence on our daily lives is unmistakable, both providing convenience and creating serious risks.

Big Tech is no stranger to all kinds of ethical issues, such as issues of privacy and surveillance including hackers spying inside homes via unsecured baby monitors, and theme parks requiring every guest to register themselves and wear a tracking bracelet during their visit so their every stop, purchase, and move can be recorded and processed in a cloud-based computer system, creating a data-driven personal profile (Marsden, 2019). Such phenomena demonstrate that Tech is not neutral; if humans made it, that tech has a human imprint on it in some way. For example, artificial intelligence (AI) programmers are discovering that their machines can carry the racial and gender biases of those who created them, even though the programmers did not likely intentionally program their AI in that way. As Joe has pointed out repeatedly throughout his career, technology is not inherently evil or destructive, but technological change is a complex set of events that can have far-reaching economic, physical, and psychological effects on people, and it's expected that there is a fear of the unknown consequences that will arise from new and evolving Tech (Pitt, 2000).

At this point in time, no one truly understands how deep neural networks (DNN) as part of an AI system "learn" to recognize images, words, and other data points. It is important to point out that AI itself is not intelligence; it is a collection of computational techniques used to help us close the performance gap that is stressed by high-speed computing. AI techniques always trade search algorithms for knowledge. Using words like "learning" and "deep" doesn't make it so. Even if algorithms are tuned to a great data set for better precision and speed, they may be confused by a changing environment, averaging multimodal data, sampling errors, and assumptions about what is happening. And presently, in a commercial sense, AI is a buzzword widely used by vendors to sell often useless products to the government.

Joe's philosophical view of AI is more grounded. He has characterized the goal of AI as figuring out the mysteries of human cognition and allowing for new paths leading from human sensory input to coherent thinking and knowledge, resulting in radically new outputs (Pitt, 2017). AI right now is a fascinating journey, not an endpoint.

In most cases, technology is inevitably entangled in politics and often driven by personal and corporate interests. A classic example—one that Joe always had a keen interest in and wrote a great deal about—is the famous Italian astronomer Galileo Galilei, who is probably best known in today's popular culture as the revolutionary scientist who invented the telescope and was branded a heretic by the Catholic Church—essentially a powerful corporation—after proclaiming that the Copernican theory of the universe, where the planets revolve around the sun, was correct. Most people—and authorities—at that time believed in an earth-centric universe. Galileo eventually wound up dying under house arrest in 1642 for teaching and defending the

Copernican system. Nevertheless, Galileo's innovative telescope lived on, was improved upon, played an important role in changing the way humans perceived their place in the universe, and had long-reaching impacts on society and science (Pitt, 1992).

Out of all his writings, I've found Joe's *Thinking about Technology: Foundations of the Philosophy of Technology* (2000) the most useful when helping me to think through and solve problems at the office because it describes many practical applications toward real-life technology quandaries. In *Thinking about Technology*, Joe defines technology broadly as humanity at work, with work being "the deliberate design and manufacture of the means to manipulate the environment to meet humanity's changing needs and goals" (Pitt, 1992). Importantly, Joe establishes that technology is a form of knowledge that does not necessarily have to be tied to or dependent on science and should not be considered as inferior to the latter.

Tech's not taking a second place to science is nowhere more evident than in the twentieth and twenty-first centuries. During this time, I would argue that there has been no technology—not spacecraft, jet engines, or nuclear weapons—that has had as large an overall impact on humanity as the computer. *Computing is the technological backbone undergirding modern life today.* As part of the Big Tech economy, computers are an inextricable part of our society, homes, relationships, and culture; they underpin commerce, medicine, finance, education, research, and most everything else we engage in. How we even define what a computer is remains up for grabs: their capabilities have been advancing incrementally for decades and they have taken on many forms, from game consoles and handheld smartphones, to larger than football field-sized data centers. Computing is arguably the backbone of our global economy today, and semiconductor products are the third-largest class of U.S. exports. Recent research has shown that one-third of productivity increases in the United States since the early 1970s came from the broad field of computing (Byrne, Oliner, and Sichel, 2013).

Modern computing has its roots in World War II and has evolved extensively since then, growing steadily but slowly in terms of memory and performance. Very large, general-purpose, and powerful supercomputers were once almost exclusively the realm of secret government defense laboratories. Specifically, after World War II, the Los Alamos and Lawrence Livermore National Laboratories as part of the U.S. Department of Energy (previously known as the U.S. Atomic Energy Commission) became key sponsors of and customers for specialized scientific HPC systems. They quickly established the speed of a machine's floating-point arithmetic operations as the performance criterion defining supercomputing (MacKenzie, 1991). Incredibly, over seventy years later, this is still primarily how HPC performance is gauged: specifically, speeds are measured in Floating Point Operations Per Second (FLOPS).

While the first "supercomputers" such as the Electronic Numeric Integrator and Calculator (ENIAC) boasted speeds of around 500 FLOPS, today we have reached the petascale or petaflop (PF) era, with machines that can run one quadrillion FLOPS.

HPCs are still found in government settings, but today they are also in major research universities and in certain commercial industry laboratories. In these settings, they solve all sorts of specialized scientific and engineering problems ranging from oil exploration to climate change, and on Wall Street automatically execute millions of daily stock market trades often without many humans in the loop. Yet, HPCs have largely remained unadopted by middle- and lower-end users, which is unfortunate because of HPC's potential benefits to them; I will return to this subject later and explain it more thoroughly.

As with Galileo's telescope, scientific instruments can make fascinating case studies of human faith in and their relationships with technology. HPC is an interesting case study in technological advancement because it's a very specialized, even rarified, segment of the Big Tech economy; yet, HPC is embroiled with political wrangling, big egos, personal ambitions, and curious behavior where many of its practitioners compete to be able to boast of owning the fastest HPC on the planet. Japan currently holds this title: as of June 2020, the world's fastest HPC, Fugaku, located in Kobe, demonstrated a speed of 415.5 PF. At that time, the US held the number 2 position with Summit at 148 PF, while China's TaihuLight placed number 4 on the Top500 list at 93 PF Beyond this, HPC builders are now jockeying to be first to reach exascale computing, machines capable of at least one exaFLOPS, or a quintillion (or 10^{18}) calculations per second. Yet, HPC builders are incorporating very little fundamentally new technology into the design of these devices, instead scaling out from what technology they have already and with the limited choice of products that the supercomputer vendors will provide.

While the Top500 list is neither the only nor the single most important measure of HPC's advancement, these numbers do signal two important trends: first, that the United States is no longer significantly far ahead of foreign competition and, second, that HPC has become an inextricable and commoditized part of a global industrial ecosystem that is rapidly changing. China's ascent in HPC is an important bellwether for significant shifts in the global economy and international security.[3]

Times have changed but HPC practitioners have not. The Top500 project was started when it became clear in the early 1990s that a new definition of supercomputer was needed to produce meaningful statistics, intending to provide a reliable basis for tracking and detecting trends in high-performance computing and bases rankings on a portable implementation of the Fortran-based LINPACK benchmark for distributed-memory computers.[4] Every machine that is measured for the Top500 list is given the same LINPACK benchmark to run.[5] Lately, detractors have vocally criticized the LINPACK benchmark, saying that it has succeeded only because of its scalability, and for showing unrealistic performance levels that would generally be unobtainable by all but a very few programmers who optimize their code for their specific machine, because it only tests the resolution of dense linear systems, which are not representative of all the complex operations usually performed in scientific computing. Most HPC builders today recognize the faults with the Top500 list, but continue to adhere to it; changing ingrained culture and established ways is always challenging.

The reality is that there has been too much emphasis among supercomputer builders and their government sponsors on merely attaining the highest score on the list, with little regard for the actual quality or true performance of the machine when it tackles real-world problems. Within the HPC community this superfluous chase to be number one on the Top500 list is today known unofficially as the quest for "machoflops," where HPC builders and their sponsors boast in nerd-speak that "my machine has more FLOPS than yours," or "my HPC is bigger than theirs." According to HPC practitioners, the term "machoflops" has been in colloquial use since the 1980s, where allegedly vendors such as the (now defunct) Thinking Machines Corporation, IBM, and others—in order to undercut each other—would accuse competitors of overhyping that their products were the fastest available for purchase. Machoflops never went away, and neither did the Top500 list.[6] Alternatives to the LINPACK benchmark do exist, such as the Green500 list, which measures HPCs in terms of energy efficiency. Such alternative rankings, however, are not that widely accepted by the HPC community. Why not? The answer is a complex mix of human decisions and values, technology, politics, and economics.

HPC technology is based on long-established, complementary metal-oxide semiconductors (CMOS), sometimes referred to as "classical computing," as opposed to over-the-horizon concepts like quantum or superconducting computing, which are still in the basic research stages and may not be realized for some time. A looming problem for CMOS is the likely death of semiconductors as we know them: since coined by Intel Corporation founder Gordon Moore in 1965, Moore's Law states that the number of transistors in a dense integrated circuit doubles about every two years. Moore's Law has held true for a long time, but if experts in the semiconductor industry are correct and we continue to cram ever more transistors on each chip, we're going to run out of room on the silicon slices. To be clear, Moore's Law is not a law in any legal sense and is more akin to an observation and projection of technological change. It is viewed by some as a human construct. Philosopher Cyrus Mody has even suggested that Moore's be viewed as a *regula*, which

> conveys multiple senses that reflect the heterogeneity of Moore's Law and phenomena like it: it is a rule to be obeyed; it is something made with a ruler, i.e., a human construct that straightens out complexity; it is a regularity observed in the world; and it has a regulatory function. (Mody, 2017, 242)

However we characterize Moore's, along with the unraveling of it and closely related, Dennard Scaling is also breaking down. Named after a 1974 paper coauthored by Robert Dennard, this "law" states, roughly, that as transistors decrease in size, their power density stays constant, so that the power use stays in proportion with area (Dennard et al., 1974). But since around 2005 the reduction in the size of transistors associated with a decrease in power has ended, leading to an inability to increase clock frequencies up as the transistor size decreases. The inability to operate within the same power envelope led the HPC industry to transition to multicore architectures, creating significant challenges for memory technology where with each new generation of Central Processing Unit (CPU) the number of memory controllers per core decreased and the memory system burden increased.

In the meantime, HPC component technology became a broad set of globally produced and desired commodities, while worldwide competition has stiffened to outcompute other nations, and almost anyone with the resources can build an HPC, which means most developed nations. U.S. government and private investments in HPC have been inconsistent in the past decade, whereas investment in data analytic-oriented or data computing—driven by consumer demand for AI-enabled smartphones, gaming, and scores of other personal consumer products—has skyrocketed. There is far more money to be made in data computing; therefore, the industry is chasing that more than government HPC projects. The supercomputer industry and the government both failed to foresee the rise of data-oriented computing and its seemingly limitless ever hungry for cool, new devices customer base.

For well over two decades HPC architectures have strayed further and further away from an optimal balance between processor speed, memory access, and input/output (I/O) speed. Since 1991, successive generations of HPC systems have upped peak processor performance without corresponding advances in per-core memory capacity and speed, and thus the systems have become increasingly compute-centric and the well-known "memory wall" in these devices has gotten worse. This means that every time the HPC industry comes out with a new flagship machine it is in practice slower than the previous generation: what this means is that when a real-world complicated code representing a hard science problem is run on the machine, users get single-digit efficiencies and much slower speeds than when tested by LINPACK. HPC designers have made a few noteworthy technological advances in the

---

past two dec
cantly much
Many cor
HPC paralle
tasks across
and other o
PCs, and th
computing.
distributed
grammers a
The adva
as modeling
some years
tions and s
ism. The n
reliability,
ures. Incre
processors
cycle: whil
often incre
Besides
political re
steadfastly
puting tec
already pr
not make
too focuse
ments. An
sors woul
them to th
usable sup
Witness
back to h
sense of c
my STS h
want "the
competiti
described.
political s
ers who a
The nat
also facto

Law and phe-
th a ruler, i.e.,
ty observed in

ling of it and
Named after a
s, roughly, that
nstant, so that
al., 1974). But
ociated with a
ease clock fre-
operate within
n to multicore
nology where
) the number
system burden

ad set of glob-
mpetition has
the resources
S. government
e past decade,
ng—driven by
cores of other
ore money to
that more than
e government
its seemingly

d further and
eed, memory
generations of
corresponding
systems have
mory wall" in
HPC industry
han the previ-
plicated code
rs get single-
y LINPACK.
vances in the

past two decades, such as message passing interface (MPI), but not significantly much else.

Many complex scientific problems are best solved by being parallelized; HPC parallel computing is accomplished by splitting up large and complex tasks across multiple processors. The 1990s saw the introduction of Linux and other open-source software that could run on inexpensive, off-the-shelf PCs, and this work opened the door for low-cost, high-performance cluster computing. New standards and tools were also developed that allowed for distributed memory parallel computing systems that made it easier for programmers and a variety of scalable and portable parallel applications.

The advance of parallelism has been a boon to scientific problem sets, such as modeling nuclear weapons and understanding supernovas. But twenty and some years ago, HPC practitioners' fundamental assumptions about applications and system software did not anticipate exponential growth in parallelism. The number of system components increased faster than the component reliability, while undetected error rates continued to increase alongside failures. Increasing power requirements are driving HPC builders to use more processors instead of developing newer and faster ones. This is a vicious cycle: while US vendors are struggling to reduce chip power, their solutions often increase fault rates and involve compromises in capabilities.

Besides the technical obstacles and lack of foresight, there are social and political reasons for the near halt in HPC's evolution: the government has steadfastly refused to invest in much new and alternative-to-CMOS computing technologies, instead insisting on mere scale-out of what vendors already provide. HPC industry vendors such as Intel, Cray, and IBM will not make the investment in either new technology or software as they are too focused on quarterly profit results instead of long-term research investments. And, as we saw earlier, HPC specialists and some government sponsors would often remain too focused on vanity projects that would propel them to the top of the Top500 ranking, rather than build the most efficiently usable supercomputers.

Witnessing all this unfold over a number of years, I find myself looking back to how Joe taught me that thinking philosophically can help us make sense of complex problems such as the state HPC has found itself in. Putting my STS hat on, I ask the question: why do we even want or perceive we want "the best" when it comes to technology? Is it the American, workaholic, competitive ethos? Perhaps somewhat, but it is all these factors I've just described, coupled with the reality that the United States does not have the political support to do this because HPC remains invisible to most policymakers who are in positions to have an impact on it.

The nature and culture of the HPC research and development community also factor into where this industry stands today. This is not to assign blame on individuals, but taking an applied STS view of how this failed, some description of the HPC designer culture and community is necessary to fully understand the big picture. HPC is a nerdy, niche, overwhelmingly over-age-fifty male segment of the tech landscape—the realm of an elite set of weapons and rocket scientists who live and work in a vacuum and have no reason to consider how HPC could benefit the larger world.[7] When attending an HPC conference, one notices that the field does not attract many young (meaning under age forty) professionals, who are today drawn to the excitement of data computing based on newer languages such as Python, SQL, and others which are easier for new developers to learn and master than programming HPCs.

In *The Soul of a New Machine* (1981), author Tracy Kidder chronicled the team of engineers in the then startup Data General Corporation. The team worked under enormous pressure and at a furious pace to complete the Eclipse MV/8000 computer in 1980. In this book, Kidder observed that computer engineers often harbor strong feelings toward their new designs (Kidder, 1981). Likewise, specialist HPC designers are usually passionate about the technology they are building, and territorial about their space in the industry. Moreover, those who are funded by the government have to propose a project, defend their budget, and as in the private sector fight off rivals who are designing competing machines and fighting for the same sources of money. This process becomes part of their norms and reinforces a culture of inward focus and exclusivity.

Although there are exceptions, users of HPC and the people who program them in the Department of Defense (DOD) and government laboratories are nearly all advanced specialists in their fields with many years of experience using HPC to solve difficult problems. In many cases the users and authors of codes are the same group of researchers, forming—because of the rarified nature of the work—something of an exclusive club.

Because of this, HPC systems and their applications have very steep learning curves for novice users who have not spent years programming them. HPC applications often solve a broad range of problems ranging from modeling the universe to designing new aircraft, and users are required to specify hundreds of input variables related to the problem to be solved; mis-specifying even one of these parameters can result in an aborted run or, worse, a completed run with a nonsensical answer. Because HPC system vendors and large federal HPC programs have had the luxury of a large base of experienced users, the systems themselves remain quite difficult to use for outsiders and in general do not reflect many of the lessons in human-computer interaction that make today's cellphones, tablets, and laptop computers relevant and accessible to a wide range of users in our society.

All these factors have limited the adoption of advanced technical computing in emerging areas by nonspecialists, or lower- to middle-end users. A focus on creating more usable HPC systems that utilize modern interfaces and provide more cognitive support to non-HPC specialists would lower the barrier to entry for new application areas, such as HPC-enabled analysis of complex social and behavioral structures and allow for new forms of non-CMOS computing to be experimented with. An emphasis on user productivity with mission-critical application codes could simplify correct specification of problems and execution of advanced simulation applications by nonspecialists, increasing the applicability of tools in existing areas of use to users from new communities.

The United States, as a nation, is at a major inflection point in developing next-generation, high-end, extreme-scale systems. We are moving from computing based on processors that are programmed to follow a predesigned sequence of instructions, to the cognitive computing era based on massive amounts of data and systems evolving into systems that can "learn," such as Amazon's Echo platform that will figure out that you like veggie pizza and mystery novels. Cognitive systems can modify and optimize projections or weigh the value of information based on experience and results. This new approach to computing requires entirely new strategies and skills to maintain U.S. leadership in technology. If we solve the challenges in exascale that include power, energy, memory, storage, concurrency and locality, and resiliency, not only will we advance technology as a discipline, but if humanity chooses to put this technology to good use we will also bolster economic competitiveness and bring unprecedented understanding of the natural world and the cosmos.

The new imperative must be to design for data, not for processor performance. Industry's focus has largely turned to Graphical Processing Unit (GPU) handhelds and consumer products.[8] The rest of the world is going in a different direction than the U.S. HPC community: everyone's iPhone is, in principle, a supercomputer. Consumer demand has given rise to a massive data intensive-based computing industry of generalist users. Notably, Japan's Fugaku is the first system powered by ARM processors—a significant architectural milestone of which a detailed analysis of is beyond the scope of this chapter. Unfortunately, many leaders in the HPC community see the rise of data computing as an either-or question: either we invest in exascale or in data computing. We need to push on both. Thinking back to Joe's class, Thomas Kuhn's landmark *The Structure of Scientific Revolutions* (1962) comes to mind where he introduced the concept of a paradigm shift in scientific research and belief (Kuhn, 1962). A paradigm shift is now desperately needed in supercomputing, because, as things stand today, *meaningful, usable, and reliable exascale computing is still unattainable.*

It's hard to see the future, but economics will no doubt drive change in the HPC community and industry. And it may well be a small startup company working on non-CMOS computing that overthrows the current HPC industry. This scenario would be truly disruptive, where the future may be the forced production of very different types of microprocessors based on volume economics. Indeed, as Neil Thompson and Svenja Spanuth have argued in a recent working paper, the combination of the end of Moore's Law causing notable rising fixed costs for the semiconductor industry and general-purpose universal processors becoming a mature technology, more and more advanced computing users will switch to specialized processors for their needs, leaving HPC as we know it behind (Thompson and Spanuth, 2019).

Adding to the economic aspect, individual government agencies do not have uniform HPC needs for their respective missions. Historically, while the Department of Energy (DOE) and the Department of Defense (DOD) have driven the very high end of HPC—often at hefty price tags of hundreds of millions of dollars per each new machine acquisition—areas such as weather and climate modeling, and basic military vehicle design and ordnance computing needs generally can be carried out using lower end computing such as grids, clouds, or white box clusters. Being its biggest customers, the DOE and DOD have been essentially keeping the HPC industry alive for the past several decades. This worked through the 1980s and perhaps 1990s. But this business model is no longer sustainable.

At the Harvard Business School, the late Clay Christensen wrote extensively about innovation and technology and his work has had almost as much influence on my thinking as Joe's. Christensen came up with the disruptive innovation concept, which describes a process by which a product or service takes root initially in simple applications at the bottom of a market and then relentlessly moves upmarket, eventually displacing established competitors. In Christensen's *The Innovator's' Dilemma* (1997), he examines several cases of disruptive innovation occurring in the disk drive, steel, and mechanical excavator industries, where the large established corporations in each of these respective areas did not anticipate the small disruptor upending their markets and found themselves losing to the new competitors. HPC today is a cutthroat, rapidly consolidating business, and not a high margin one to boot.[9] Competition from upstarts may bury the HPC business.

Christensen wrote more recently about how modern companies' pursuit of profits and myopic focus on return on net assets (RONA), internal rate of return (IRR), and return on capital employed (ROCE) is killing America's ability to generate empowering innovations (Christensen, 2012). The HPC industry is as guilty of this behavior as any other big industry, where, as Christensen has described:

> In the semiconductor industry, for instance, there are almost no companies left in America that fabricate their own products besides Intel [INTC] [*sic*]. Most of them have become "fab-less" semiconductor companies. These companies are even proud of being "fab-less" because their profit as a percent of assets is much

> higher than at Intel. So they outsource the fabrication of the semi-conductors to Taiwan and China. (Christensen, 2012)

The U.S. government is no longer driving innovation in tech. Consumer capitalism is. Silicon Valley's ascent and strong innovation culture was early on dependent on the U.S. government. China's current rapid ascent in technology is strongly supported by its government, which focuses on the long-term view of investment, and not on RONA, IRR, or ROCE. Furthermore, Chinese industry is not bifurcated from its government the way it is here at home. The United States may be facing a sputnik moment if China reaches meaningful exascale computing first; China has publicly announced plans to achieve exaFLOP speeds by 2021.

The United States, for the first time in decades, faces technological uncertainty from a formidable economic and growing military competitor. The logical positivists shouted a warning call to the Western world when they struck out against fascism and called for ground truth on scientific inquiry. Although we already live in a near completely globalized world in terms of industrial supply chain and manufacturing, China's stunning ascent in HPC should serve as another kind of warning to Western nations: that we are entering a new era of global balance of power that is very different than that which characterized the Cold War.

The U.S. government has had a great deal of influence on both scientific and technological discoveries and developments for the past seven decades. Yet, the government has gradually been shifting away from funding big technological projects over the last thirty years. Consider the Human Genome Project that was completed only when J. Craig Venter stepped in with private money, or the current privatization of space travel, or the quest to end malaria that has largely been funded by the Gates Foundation. Wealthy patrons are taking up the mantle of sponsoring some of these big projects where the government has lagged. US HPC could see the same fate, which might allow for US supercomputing to get back on track toward convergence of classical and data computing and solve the challenges to exascale.

In writing this chapter, I do not, nor does anyone I know or work with, claim to offer a single correct solution to getting US HPC on track because there isn't one all-encompassing answer to this multifaceted problem. What is clear is that HPC, as a technological case study, demonstrates how hard it is for human beings to *foresee* and *prepare* for major technology changes. And as human beings in the thick of the HPC business, it's easy to miss all the technological and economic changes that are going on right in front of us. Compared to Galileo's telescope, technology today is very complex in terms of components and their global supply chain, politics, personalities, culture, corporate and national interests, and innovation drivers. Market forces are as important in shaping the direction of technology as are all these other factors covered in this chapter.

Given that HPC as a case study has been affected by all these forces, technology does stand on its own apart from science as Joe has shown us in his work, and not merely technology, but I would argue that increasingly, given its pervasiveness, it is computing that is defining the human condition today. If technology is humanity at work, then computing is humanity running a marathon given the rate of change in this field. Computing today looks very different and substantively is very different than what it was forty years ago. And it will likely be unrecognizable by early twenty-first-century humans if we could see what computing will look like forty years from now.

One of the valuable lessons about studying with a philosopher like Joe is that sometimes it pays to look at a problem from a completely different view. So from another perspective, perhaps "saving" HPC as an industry or enterprise is not the right answer. Usable exascale may simply not be reachable. We know that machoflops don't matter. General-purpose computing may be too mature of a technology to advance rapidly anymore, as the slowing down in their improvements in performance-per-dollar seems to be showing, and specialized processors have become the norm (Thompson and Spanuth, 2019). Computer platforms and systems are changing, as we've seen, and people will ultimately adapt to new technology offerings.

Long ago, Joe spurred an effort to see and analyze technology as something that people do and how they go about doing it (Pitt, 2000). Joe's work has influenced me to approach the problem of how to move advanced computing forward in a holistic way that includes considering humans and technology both as part of the solution, and gives me hope that I am making the most informed decisions when I am contributing to public sector decision-making—or "making the sausage" in policy speak—that involves nine-figure budgets. No one can ultimately see the future of technology, but Joe's writings on the philosophy of technology have helped the technological horizon seem a little less murky, and for that knowledge, I am grateful to him.

NOTES

1. The views expressed are the author's own and do not represent those of the US Government.

2. See Top500.org. Since 1993 the Top500 project ranks and describes the 500 most powerful HPC systems in the world. The list is updated twice per year.

3. HPC is often viewed as a measure of a country's economic competitiveness. To emphasize this, the US Council on Competitiveness coined the phrase, "To

outcompete is to outcompute" several years ago. See https://www.compete.org/programs/compete-innovation.

4. LINPACK is a software library for performing numerical linear algebra on digital computers by running a program that solves a system of linear equations. It was written in the 1970s.

5. Countries voluntarily participate in this ranking: there exist many fast supercomputers in classified military settings around the world that do not appear on the Top500 list.

6. I have often wondered what the HPC industry would look like and if machoflops would even exist as a term if the supercomputing business were dominated by women instead of men. Would the Top500 list exist at all? Or would computing architecture look very different than it does now?

7. HPC culture has some similar characteristics to that of high energy physicists as described by Sharon Traweek in *Beamtimes and Lifetimes: The World of High Energy Physics*, Cambridge: Harvard University Press, 1992.

8. In contrast to a Central Processing Unit (CPU), which Intel makes, for example. CPUs and GPUs are physically not that dissimilar. They are both composed of hundreds of millions of transistors, and can process thousands of operations per second. CPUs are often colloquially referred to as a computer's brain, handling an astonishingly wide variety of types of problems and forming the basis of general-purpose computing. A GPU is a specialized type of microprocessor optimized to display graphics and does very specific computational tasks. It runs at a lower clock speed than a CPU but has many times the number of processing cores. GPUs are popular among gamers and are used increasingly to perform data computing.

9. In contrast, public cloud building such as that done by Amazon Web Services is a high margin business.

REFERENCES

Byrne, D. M., S. D. Oliner, and D. E. Sichel. 2013. "Is the Information Technology Revolution Over?" *International Productivity Monitor*. 25: 20–36.

Christensen, Clayton M. 2012. "A Capitalist's Dilemma, Whoever Wins on Tuesday." *The New York Times*, November 3 [Accessed February 15, 2019].

Dennard, Robert H., Fritz Gaensslen, Hwa-Nien Yu, Leo Rideout, Ernest Bassous, and Andre LeBlanc. 1974. "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions." *IEEE Journal of Solid-State Circuits*. 9(4): 256–68.

Haraway, Donna J. 1991. *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.

Hempel, Carl G. 1965. *Aspects of Scientific Explanation: And Other Essays in the Philosophy of Science*. New York: Free Press.

Kidder, Tracy. 1981. *The Soul of a New Machine*. New York: Avon Books.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.

MacKenzie, Donald. 1991. "The Influence of the Los Alamos and Livermore National Laboratories on the Development of Supercomputing." *IEEE Annals of the History of Computing*. 13(2): 179–201.

Marsden, Rhodri. 2013. "Mickey Is Watching You: Does Disney's New 'Magic Band' Infringe on Consumer Freedoms?" *The Independent*, February 20, 2013 [Accessed February 9, 2019].

Mody, Cyrus C. M. 2017. "Moore's Regula." In *Spaces for the Future: A Companion to the Philosophy of Technology*, edited by Joseph C. Pitt and Ashley Shew, 238–47. New York: Routledge.

Pitt, Joseph C. 1992. *Galileo, Human Knowledge, and the Book of Nature: Method Replaces Metaphysics*. Dordrecht: Springer Science + Business Media.

Pitt, Joseph C. 2000. *Thinking about Technology: Foundations of the Philosophy of Technology*. New York: Seven Bridges Press.

Pitt, Joseph C. 2017. "Transcendence in Space." In *Spaces for the Future: A Companion to the Philosophy of Technology*, edited by Joseph C. Pitt and Ashley Shew, 340–45. New York: Routledge.

Popper, Karl. 1968. *The Logic of Scientific Discovery*. New York: Harper & Row.

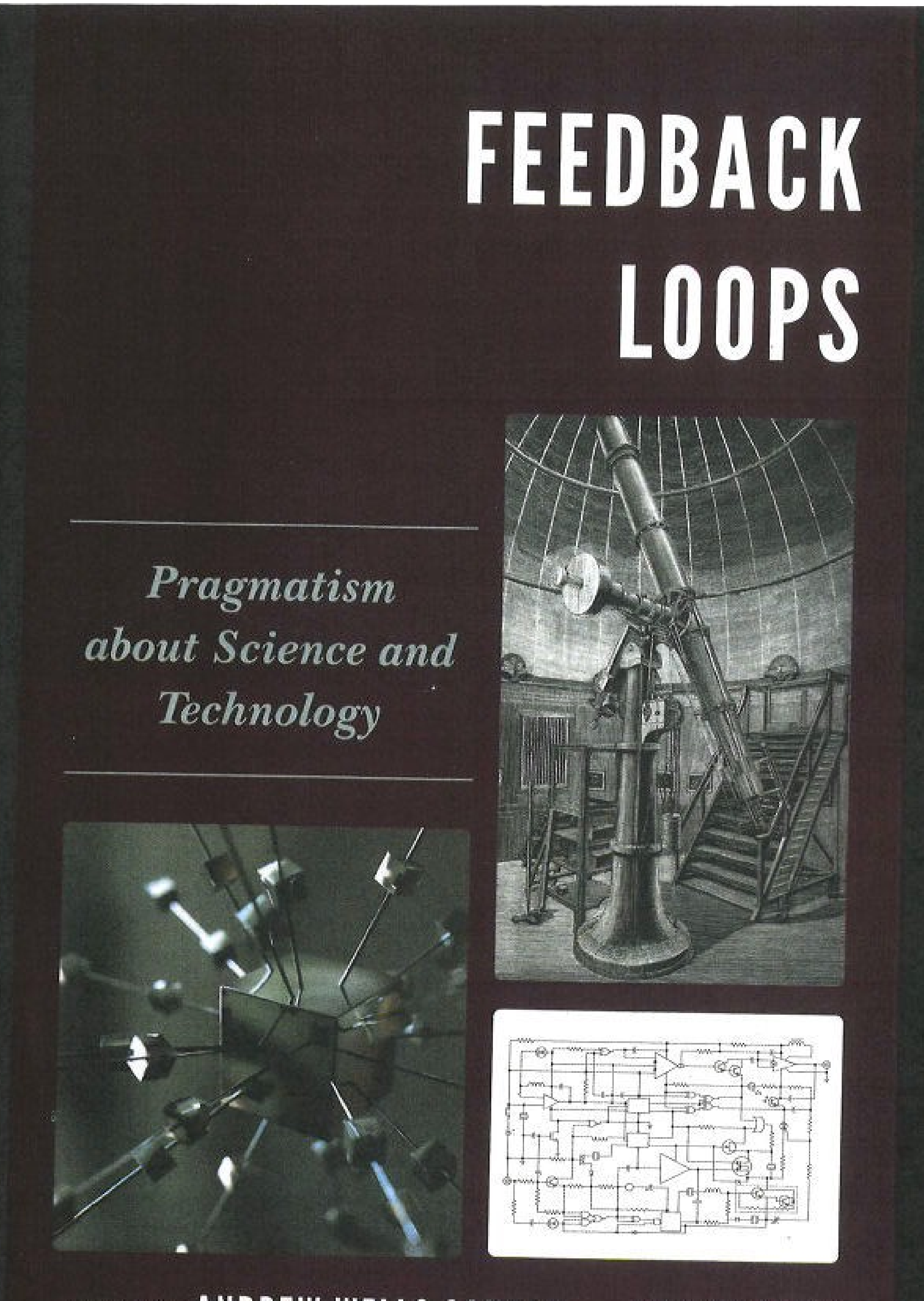
Popper, Karl. 1994. *The Open Society and Its Enemies*. Princeton: Princeton University Press.

Rossiter, Margaret. 1982. *Women Scientists in America: Struggles and Strategies to 1940*. Baltimore, MD: Johns Hopkins University Press.

Snow, C. P. 1962. *The Two Cultures and the Scientific Revolution*. New York: Cambridge University Press.

Thompson, Neil C. and Svenja Spanuth. 2019. The Decline of Computers as a General Purpose Technology: Why Deep Learning and the End of Moore's Law Are Fragmenting Computing. Working Paper at https://ssrn.com/abstract=3287769 [Accessed April 2, 2019].

Traweek, Sharon. 1992. *Beamtimes and Lifetimes: The World of High Energy Physics*. Cambridge: Harvard University Press.



FEEDBACK LOOPS

Pragmatism about Science and Technology

Edited by **ANDREW WELLS GARNAR** and **ASHLEY SHEW**

Afterword by Joseph C. Pitt

Published by Lexington Books
An imprint of The Rowman & Littlefield Publishing Group, Inc.
4501 Forbes Boulevard, Suite 200, Lanham, Maryland 20706
www.rowman.com

6 Tinworth Street, London SE11 5AL, United Kingdom

Copyright © 2020 The Rowman & Littlefield Publishing Group, Inc.

All rights reserved. No part of this book may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without written permission from the publisher, except by a reviewer who may quote passages in a review.

British Library Cataloguing in Publication Information Available

Library of Congress Cataloging-in-Publication Data

Names: Garnar, Andrew Wells, 1974- editor. | Shew, Ashley, 1983- editor.
Title: Feedback loops : pragmatism about science and technology / edited by Andrew Wells Garnar and Ashley Shew ; afterword by Joseph C. Pitt.
Description: Lanham : Lexington Books, [2020] | Series: Postphenomenology and the philosophy of technology | Includes bibliographical references and index. | Summary: "This volume explores the arrangement of science, technology, society, and education. Using the concept of 'feedback loop', this book processes subjects dear to the work of Joseph C. Pitt: technology as humanity at work, pragmatism, Sicilian realism, pragmatist pedagogy, instrumentation in science, and more"— Provided by publisher.
Identifiers: LCCN 2020034182 (print) | LCCN 2020034183 (ebook) | ISBN 9781498597623 (cloth ; permanent paper) | ISBN 9781498597630 (epub)
Subjects: LCSH: Technology—Philosophy. | Pitt, Joseph C.
Classification: LCC T14 .F39 2020 (print) | LCC T14 (ebook) | DDC 601—dc23
LC record available at https://lccn.loc.gov/2020034182
LC ebook record available at https://lccn.loc.gov/2020034183

Contents