IEEE CS 2022 Report (DRAFT)

Hasan Alkhatib, Paolo Faraboschi, Eitan Frachtenberg,
Hironori Kasahara, Danny Lange, Phil Laplante,
Arif Merchant, Dejan Milojicic, and Karsten Schwan

with contributions by: Mohammed AlQuraishi, Angela Burgess,
Hiroyasu Iwata, Rick McGeer, and John Walz

Preface

In 2013-14, nine technical leaders wrote a report, entitled *IEEE CS 2022*, surveying 22 innovative technologies that could change the industry by the year 2022. The report covers 3D printing, big data and analytics, open intellectual property movement, massively online open courses, security cross-cutting issues, universal memory, 3D integrated circuits, photonics, cloud computing, computational biology and bioinformatics, device and nanotechnology, sustainability, high-performance computing, the Internet of Things, life sciences, machine learning and intelligent systems, natural user interfaces, networking and interconnectivity, quantum computing, software-defined networks, multicore, and robotics for medical care.

These technologies, tied into a scenario that we call seamless intelligence, present a view of the future. For each of the 22 technologies, there is a description of the state of the art, challenges, where we think the technology will go, and its disruption. To confirm the report's prediction, we surveyed IEEE members about technology drivers and disruptors. We also tried to predict what kind of society the world would require with these 22 technologies. Finally, we analyzed the IEEE digital library to better understand the degree to which these technologies are covered today and by which Societies, so that we can make better ties.

This document is intended for computer science professionals, students, and professors, as well as laymen interested in technology and technology use. While we tried to be complete and exhaustive, it is inevitable that some technologies have been omitted, such as Bitcoin, future transportation, and the general notion of what technology contributes to the mankind. Our position, as well as the premise that this document brings, is that technology is the enabler. What humanity takes out of it really depends on human society.

The *IEEE CS 2022* report was presented at the Computer Society of India Congress, at the Information Processing Society of Japan (IPSJ) Congress, at the IEEE CS Board of Governors, at the IEEE CS Industrial Advisory Board, and at Belgrade Chapter. We received positive feedback and excellent ideas for improvement. This is a living document, because the technology continuously changes. We intend to use this document for the IEEE CS strategic planning that takes place every three years. We hope that the IEEE CS will be able to come up with similar reports regularly in the future.

I thank Hasan Alkhatib, Paolo Faraboschi, Eitan Frachtenberg, Hironori Kasahara, Danny Lange, Phil Laplante, Arif Merchant, and Karsten Schwan for making this journey to 2022 together. Without their vision, technical knowledge, and creativity, this document would not be possible.

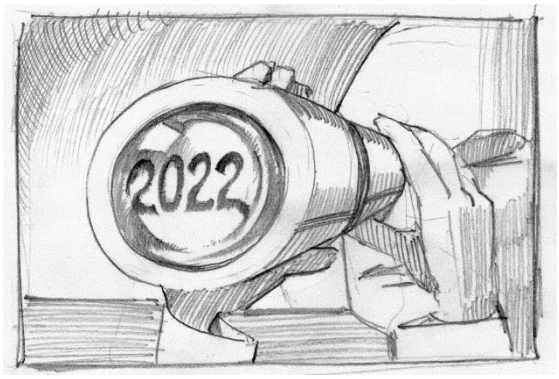
Dejan S Milojicic, IEEE Computer Society President 2014, February 2014

Table of Contents

Table of Figures

Table of Tables

1 Introduction



Predicting the future is hard and risky. Predicting the future in the computer industry is even harder and riskier due to dramatic changes in technology and limitless challenges to innovation. Only a small fraction of innovations truly disrupt the state of the art. Some are not practical or cost-effective, some are ahead of their time, and some simply do not have a market. There are numerous examples of superior technologies that were never adopted because others arrived on time or fared better in the market. Therefore this document is only an attempt to better understand where technologies are going. The book *Innovators Dilemma* and its sequels best describe the process of innovation and disruption.

Nine technical leaders of the IEEE Computer Society joined forces to write a technical report, entitled *IEEE CS 2022*, symbolically surveying 22 potential technologies that could change the landscape of computer science and industry by the year 2022. In particular, this report focuses on 3D printing, big data and analytics, open intellectual property movement, massively online open courses, security cross-cutting issues, universal memory, 3D integrated circuits, photonics, cloud computing, computational biology and bioinformatics, device and nanotechnology, sustainability, high-performance computing, the Internet of Things, life sciences, machine learning and intelligent systems, natural user interfaces, networking and interconnectivity, quantum computing, software-defined networks, multicore, and robotics for medical care.

1.1 Goals

As authors, we had the following goals in mind when we started writing the document:

- Predict the future technologies that will disrupt the state of the art.
- Help researchers understand the future impact of various technologies.
- Help laymen—a general audience—understand where technology is evolving and the implications for human society.
- Help the IEEE Computer Society understand how it should be organized for this future.

1.2 Target Audience

This document was intended for computer science professionals, students, and professors, as well as laymen interested in technology and technology use. It is equally targeted to the members of the Computer Society and similar Societies around the world, as we dare to predict what kind of future professional society will be best suited to take these technologies to the next level through its publications, conferences, communities, standards, courses, and artifacts in support of our profession and humanity.

While we tried to be complete and exhaustive, it is inevitable that some technologies and aspects have been omitted. Examples include electronic money, such as Bitcoin, and various forms of transportation, such as autonomous vehicles. Also missing is the general notion of what technology contributes to mankind, a question frequently asked by those who have seen this material to date. Our premise,

echoed in this document, is that technology is the enabler. What humanity takes out of it really depends on human society.

1.3 Process

The core team of nine technologists met twice by phone in preparation for a face-to-face meeting in Seattle, collocated with an IEEE Board of Governors gathering. We brainstormed about possible technologies and came up with a list that has since been trimmed. Each team member chose two to three technology areas to describe, and two members wrote the scenario.

We describe each of the 22 technologies by following a common approach—summary of the state of the art, challenges, where we think the technology will go, and its disruption—and tie them into a scenario that we call seamless intelligence. Together, they present a similar view of the future.

We held another face-to-face meeting in the IEEE Computer Society's Washington, DC, office to brainstorm the future of the IEEE Computer Society. Ultimately, we are attempting to predict what kind of future Society will be needed for our profession, for the professionals who will be learning, practicing, and putting into use the technologies we present here.

Most of our other interaction was by email. In a few cases, we reverted to technologists outside of our team who helped us write on the topics of life sciences, bioinformatics, robotics, and software-defined networks.

Independently, we surveyed a few thousand IEEE members on technology drivers and disruptors, and they confirmed some of our predictions and provided another perspective on the future of technology advancements.

Finally, we were helped by IEEE Computer Society staff for copyediting, pictures, and numerous other details.

1.4 Technologies Landscape

When we originally discussed the 22 technologies, we observed them all as equal. However, some of the feedback we received from the IEEE Computer Society Industrial Advisory Board was that our 22 technologies really fit into a larger landscape, comprising policies, human capital, technologies, and market categories. The image below is an attempt to classify the technology areas according to offered classification. We could have evolved this model further and tried to populate it with other elements, but we felt that our bottom-up approach was sufficient for this document's needs. We may revisit it in future attempts to categorize technology areas.

Figure 1. Landscape of 22 technologies.

(The numbers after the technology represent the subsection in Section 3*.)*

1.5 Document Organization

The rest of this document is organized as follows. Section 2 describes the seamless intelligence scenario that ties the 22 technology areas together and showcases their potential benefits. It also serves as a use case introduction for the individual technology areas presented in Section 3. Technology drivers and disruptors are presented in Section 4, based on the survey we conducted. Strengths, weaknesses, opportunities, and threats to IEEE are presented in Section 5. They serve the purpose of better understanding how the IEEE Computer Society should grow in the future, which is the theme of Section 6. Summary and next steps are provided in Section 7. The authors as well as other contributors are presented in Section 8.

2 Seamless Intelligence Scenario

2.1 Introduction

Since the inception of digital computing in the mid-1940s, society has witnessed a historic revolution in the acquisition, processing, and communication of information. This revolution has transformed every aspect of society through increased automation, ubiquitous access to information, and pervasive human networking.

2.2 State of the Art

We continue to witness an increase in the numbers, shapes, and sizes of computing devices, from micro-scale to mega-scale, as well as a combinatorial increase in connectivity, both local and global. As a result of this pervasive penetration of computing and communication capabilities, human knowledge, intelligence, and connectivity are increasingly enhanced and augmented by information technology. By 2022, we project that we will be well into a phase where intelligence becomes seamless and ubiquitous to those who can afford and use state-of-the-art information technology.

This new reality is the expected result of the confluence of multiple information and communication technologies. Computing devices—from the very small, such as wearable devices and chips embedded under the skin, to the computers inside our mobile devices, laptops, desktops, home servers, TV sets, and refrigerators, to the computing cloud that we reach via the Internet—are interconnected via different communication and networking technologies. Together, they form an intelligent mesh, a computing and communication ecosystem that augments reality with information and intelligence gathered from our fingertips, eyes, ears, and other senses, and even directly interfaced to our brain waves.

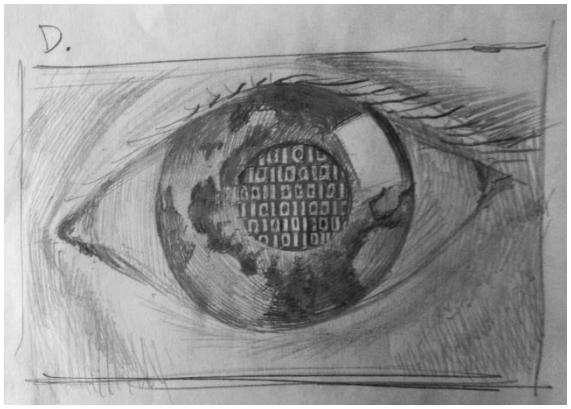
2.3 Challenges and Opportunities

At the heart of this revolution is seamless networking, where the transition from one network device to another is transparent and uninterrupted. Various wireless networking technologies—from Near-Field Communication (NFC), to Bluetooth, to Wi-Fi, 4G, and 5G—are integrated with high-speed wired networking and the Internet, allowing anywhere-to-anywhere access. But to achieve seamlessness and realize logical end-to-end connectivity, we will need communications to run independently on top of any form of physical networking, regardless of device or location. Through virtualized end-to-end connectivity, total integration of all the ecosystem devices that cater to our specific needs can be achieved. This new world will require sophisticated intelligent coordination software; voice, image, and motion recognition will transform human–computer interfaces into a seamless interaction between the user and all the computing devices in that person's life.

Another gap between today and 2022 is seamless reliance on federated identity and the use of more sophisticated identity technologies. Access will be authorized based on capabilities and access tokens rather than strictly on identity. Private applications will still require strict identity–for example, discovering from a specific social network that a specific friend happens to be at the same café as the

user will require notification of peers about their mutual presence. But to achieve interoperability, identity federation will require standards developed by and agreed upon among identity providers. In addition, meta-identity information will play a major role, capturing a person's profile and managing preferences while, for example, shopping, eating, and traveling (specifically, a hotel could detect a guest's preferred type of bed, floor level, or smoking status and automatically fulfill a reservation accordingly).

Cloud services that offer APIs to facilitate application mash-ups will lead to intelligent software that can integrate multiple services together and achieve results that are difficult to imagine today. We see the current power in mashing up location data with maps as an illustration of what future mash-ups might look like.



The combination of powerful voice and facial recognition, massive identity databases, and powerful tracking will likely result in a new norm that potentially translates into a significant loss of privacy compared to today. Technology will enable many benefits, but controlling its use and preventing misuse will require collective social action.

On the other hand, pervasive and massive identity recognition could also result in myriad benefits, such as cashless and contactless financial transactions, the ability to cross borders without stopping for inspection, and walking into a coffee shop in a foreign country and having the barista offer up your favorite coffee because your preferences appeared on her counter screen as you approached the shop.

The application of seamless and pervasive intelligence will penetrate many aspects of our lives, particularly healthcare. Imagine walking into a hospital and having your entire medical history be accessible to the attending medical professional from a centrally managed health vault: you won't need to state what medications you are currently taking or what immunization your child most recently received. Progress in 3D printing already lets your dentist automatically shape your crown molding while you wait. Physicians will also be able to use less invasive procedures, such as having a patient swallow a small camera to track the entire digestive track without needing to perform an incision; medication and medical devices could even be customized on the fly.

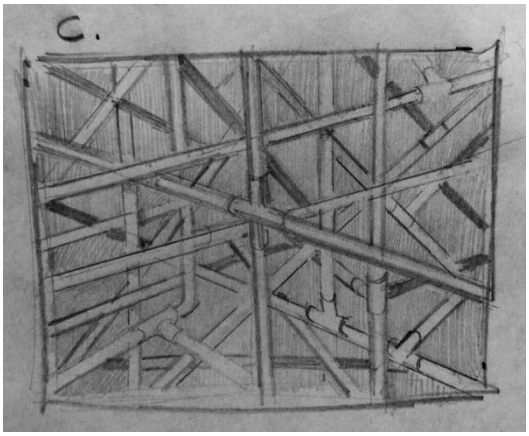
Seamless and pervasive intelligence is impacting education more disruptively. The traditional model of campus-based education is changing by virtue of the availability of better methods for both teaching and learning, augmented by automated and interactive learning outside the classroom as well as through distance participation. By 2022, we expect that the experiments with MOOCs will lead to a refined model in which they become complementary to ongoing instruction models. We also project that the classroom will involve less instruction and more dialogues with the expert professor, resulting from the ability to use technology to learn outside the classroom. Students will enjoy learning more, requiring less time and gaining deeper comprehension of their subject material. While MOOCs will become part of the education ecosystem, making them effective will be a challenge. The future holds even more for the integration of work and education via augmented reality. As someone is working, for

example, she will get customized information that progressively trains her. This will revolutionize several sectors, including customer care and the learning of services and products [Saracco].

Progress in robotics will likely transform the way mass transit is handled today to fully automated, autonomous vehicles. Imagine a driverless taxi, just large enough to accommodate you and your baggage, dispatched to your hotel to take you to the airport, automatically navigating the best route along the way. Naturally, it already knows your departure terminal from a prior seamless information exchange. Autonomous vehicles will transform the topology of urban areas, dynamically creating one-way streets and preferential lanes. The traffic layout will change continuously. This might also lead to a change in the concept of car ownership, transforming vehicles into utilities to use and drop [Saracco].

Continuity in computing—from basic sensory processing, to simple event and location tracking, to calendaring and collaboration support, to personal applications—will be augmented by powerful computing in the cloud and massively distributed systems. Big data analysis will take place in the background, providing continuous intelligence to executives who run major organizations, enabling both the tracking and coordination of major business activities and intelligent choices based on real-life data intelligence.

Developments in cloud computing will transition computing from a physical experience to a virtual one available to any user via a simple device operating on ubiquitous networks with seamless connectivity. The results of large computations running on massive cloud infrastructures will be available as affordable services that almost anyone can access and utilize. However, history has also taught us that more and more processing power becomes available at the edges and in the hands of customers/users. In that regard, the cloud can be seen as a processing fabric, part of the ambient environment, and a commodity. Cloud gets implemented as technicians and economists decide; to the end user, it is irrelevant [Saracco].



Seamless and ubiquitous intelligence aids in enforcing the strong security measures that can achieve unprecedented levels of safety in the service of peace. Smart sensors, surveillance cameras, and eavesdropping devices integrated with identity recognition systems will allow law enforcement to track and capture or quarantine individuals who might otherwise cause harm to others in society. Conversely, this access to such intrusive technology can violate individual rights and invade the privacy of innocent people. The onus is again on society to limit the use of seamless connectivity to acceptable norms.

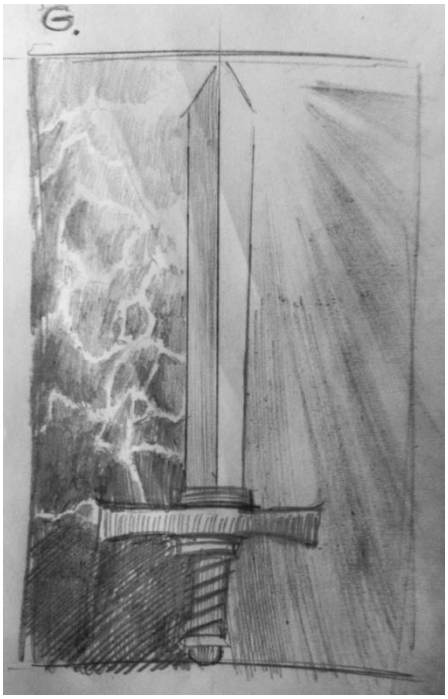
On the downside, the gap between developed and underdeveloped countries could continue to increase. The seamlessness enjoyed in developed countries will be missed when a traveler finds it hard to use a smart card at a merchant or a train station in an area that does not have that technology. The rapid evolution of increased automation and the spread of pervasive intelligence in traditional uses in everyday activities will accentuate the differences between the have and have-not nations. Nonetheless, underdeveloped countries will continue to enjoy access to advances in computing, particularly the use of inexpensive yet smart mobile devices. The trend seems to be toward facilitating further social

networking rather than real enhancement in productivity tools. Furthermore, ubiquitous computational and educational services will grow ever more accessible to any population that meets the basic connectivity requirements.

2.4 What Will Likely Happen

The future we want versus the future we do not want: information and communication technology is advancing at a pace that is surpassing our abilities as a society to direct. It is the scale and speed with which this progress is taking place that is creating this challenge. But there are choices that free nations can make through regulation and investment that can either lead to a better world or one that we do not desire.

Technology is a double-edged sword. It can be used for advancing healthcare, education, science, trade, financial services, social and political activism, security, and safety, or it can be used for militarization, to invade privacy, and to push the Big Brother phenomenon worldwide, even in countries that consider and pride themselves on being free. In general, any technology has its ups and downs. The man who invented the first ship also invented the shipwreck and the castaway! This is something that we need to understand. Even if technology is used in the best possible way, it will still bring along some downsides. It just changes the landscape, and along with it, the ups and downs [Saracco].



2.5 Potential Disruptions

The emergence of the mobile smart device sector in the past decade is likely to continue to disrupt the traditional model of desktops and laptops. Mobile applications are also expanding the common Web platform by enabling applications on mobile devices using their operating systems.

On the other end of the computing scale, the emergence of cloud computing primarily based on commodity server hardware is pushing and disrupting the traditional server sector, replacing it with computational power as a service over the network.

Another disruptive trend emerging as a result of the spread of social networking is resulting in countries and regions potentially creating their own regional Internets with imposed restrictions on access to global sites and universal services. This trend can have a negatively disruptive impact on the global Internet and freedom of individuals to access information and services regardless of geographic locations and political boundaries. There is also the impact of regulation, such as the different positions taken by the US and EU in the area of premium connectivity, which is allowed in the US but not in the EU.

Intellectual property wars among major players in the industry can present barriers to both the speed of progress and use of technology. Consumers of technology will ultimately be the victims of such wars.

2.6 Summary

By 2022, computing devices will range from nano-scale to mega-scale, with advanced networking enabling access to a world of integrated services. Virtual connectivity will enable the integration of relevant computing resources to provide users with seamless services. The resulting ecosystem will offer continuous, uninterrupted services that enhance automation, productivity, collaboration, and access to intelligence and knowledge that will be available not only at users' fingertips but accessible to all human senses, spontaneously, through emerging human–computer interfaces.

The benefit of technology is what we make of it. Societies will face further challenges in directing and investing in technologies that benefit humanity instead of destroying it or intruding on basic human rights of privacy and freedom of access to information. We should stop considering technology as something standalone. It is more than a piece of the quilt of life: it is reshaping it, and being reshaped itself by humanity. A holistic approach is needed.

2.6.1 References

[Saracco] Roberto Saracco, the author of COMSOC 2020 Report, Personal Communication.

3 22 Technologies in 2022

3.1 Security Cross-Cutting Issues

3.1.1 Introduction

Powerful forces are converging that are of great concern to individuals and private and public entities. These powerful forces will cause people, businesses, and groups to pause before releasing certain information to government, merchants, and even other citizens and to consider the consequences of every activity in which they engage.

The first of these forces is the exponential growth of large data repositories (see big data in Section 3.18) of personal and corporate information. The second is the enhanced capability to analyze this data for various patterns (see data analytics). The third force is the advancing technological ability to collect diverse data about citizens, private and public corporations, and profit and nonprofit entities alike through a variety of channels. This data includes financial transactions, personal and business correspondence, the movements of people and assets, and personal and business relationships. The fourth force is institution/municipalities and crowd-sourced information. This may be the first that will be exploited and will have an impact on society. The final force is the growing ability and determination of malevolent actors in acquiring information about people, business entities, and objects such as critical infrastructure. Malevolent actors can include adversarial government agents, criminals, malcontents, and personal or business enemies.

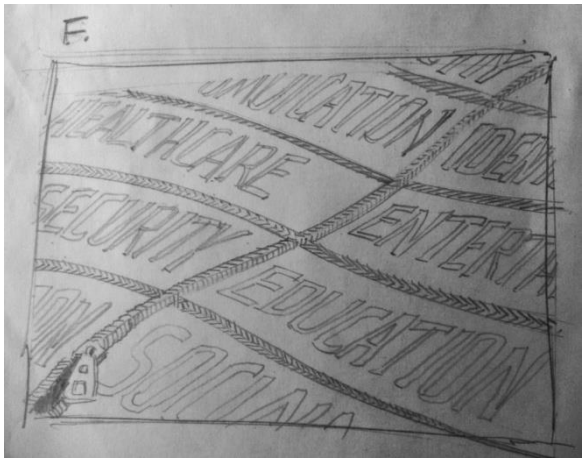
The convergence of these forces requires tradeoff decisions to be made about privacy versus security. In order to protect individuals and corporate entities from malevolent actors, governments must monitor personal and business transactions and examine associations of people with other people and with corporations and affinity groups. Governments must track movements of people and goods, monitor the utilization of private and public resources, and mine data repositories in order to investigate or predict crimes, all in the interest of protecting the public. In allowing governments to conduct these activities, however, individuals and corporations must surrender privacy. How much privacy should an individual, a corporation, or an affinity group surrender in order to ensure an acceptable level of security from threats? Should these limits be set legislatively? Is it feasible to protect one's privacy without legislative support, and what tools are available, or need to be available, to make it possible?

3.1.2 State of the Art

Social networking sites such as Facebook and Twitter can be monitored and predictive analytics used to investigate crimes or predict the potential for crimes. Machine-to-machine networks (see Internet of Things in Section 3.15) can be used to track individual products or subsystems of interests using RFID; whole systems and people can be tracked via GPS, terrestrial imaging, satellite imaging, black boxes, and other low-technology means. Powerful Internet search engines exist, and coupled with the ability to capture, store, and analyze large amounts of data from surface mail and parcels, email traffic, telephone conversations, financial transactions, consumer purchases, and Internet sites visited, government

agencies can mine this data and use predictive analytics to spot potential threats before they occur or to investigate crimes. Private entities and malevolent actors may also gain access to this information and conduct their own analytics for commercial or nefarious purposes. In some cases, such intrusions are limited by law, but the limitations vary by country, are hard to enforce, and offer little protection should the intrusion come from a government agency authorized to bypass the limits. Cryptographic methods to protect a user's privacy exist for some uses, such as OpenPGP for email, but are often hard to configure and use, and are sometimes blocked or not well supported.

3.1.3 Challenges



There is a balance between security and privacy. Citizens, corporations, and other groups accept a certain level of intrusion, provided a certain level of security is afforded. Every person, corporation, and group, however, has a different level of sensitivity to intrusion and a different notion of acceptable security risk.

There are political challenges of fostering public trust that transactions and movement are safe without being overly intrusive. Commercial organizations, such as Internet service providers, have little incentive to provide and support privacy-enhancement tools, and in some cases, are under pressure from regulators to avoid changes that will block law enforcement agencies from accessing private communications.

3.1.4 Where We Think It Will Go

Citizens and corporate entities and groups have always accepted a certain level of intrusion in order to ensure some level of security. Technological advances have simply focused more attention on this problem. If governments can show that real security is achieved through surrendering a certain level of privacy, then new technological advancements that can perform accurate predictive and forensic analytics will be embraced in exchange for a certain level of privacy being sacrificed. On the other hand, consumer demand for privacy-enhancing tools may lead to changes that make it easier for individuals to protect their privacy, perhaps at the cost of some effort and inconvenience.

3.1.5 Potential Disruptions

Access to vast quantities of personal information either in one repository (e.g. the Affordable Healthcare Database) or through aggregation of multiple databases creates an irresistible target for hackers. If infiltrated, no one can safely depend on their own identity being protected or can trust the identity of anyone else with whom they engage in a personal or business transaction. The public may rebel against any increases in intrusion if the benefit of increased security is not demonstrated.

3.1.6 Summary

The growth of large data repositories of personal information, where data from many sources may be aggregated, combined with data analytics that enable deduction of surprisingly detailed patterns of information regarding individuals and groups, has opened a Pandora 's Box of privacy issues. Privacy

intrusions can come from both authorized sources such as law enforcement and corporations that have been explicitly granted permission, as well as malevolent actors such as identity thieves. We face a tradeoff among privacy, security, and convenience. Changes in laws and improvements in privacy-enhancement tools and techniques may be needed to help users find a balance between the degree of intrusion they can tolerate and the security they desire.

3.1.7 References

W. Diffie and S.E. Landau, *Privacy on the Line: The Politics of Wiretapping and Encryption,* MIT Press, 2007.

S. Landau, "Politics, Love, and Death in a World of No Privacy," *IEEE Security & Privacy*, vol. 11, no. 3, 2013, pp. 11-13.

I. Goldberg, "Privacy-Enhancing Technologies for the Internet III: Ten Years Later," *Digital Privacy: Theory, Technologies, and Practices*, Auerbach, 2007, pp. 3–18.

3.2 The Open Intellectual Property Movement

3.2.1 Introduction

Open intellectual property (IP), such as that found in open source software, open standards, and open access publishing (along with crowd-sourcing as a means of producing information) is a significant positive byproduct of the ubiquity of the World Wide Web. It is rapidly expanding into areas where property was traditionally proprietary, such as hardware design. Continued growth of the open IP movement will continue to generate significant benefits to humankind.

But along with these benefits come significant challenges and risks, including security and trust, motivation for innovators, and diminishment of individuality.

3.2.2 State of the Art

Open IP is information contained in freely accessible repositories in which volunteers, often in very large numbers, produce and vet the content. Users of this information also provide feedback to the community, driving innovation, correcting errors, and acting as a police force to ensure that the content is not maliciously corrupted. The IP is generated for the beneficial use of humankind and is often covered by the well-known Creative Commons license.

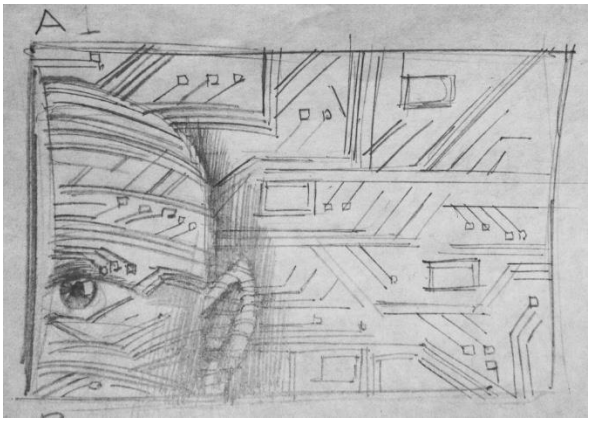
Open IP can be found in the form of information repositories (e.g., Wikipedia), open source software (e.g., Linux), media repositories (e.g., Flickr), open access publishing (e.g., Public Library of Science), open systems (e.g., World Wide Web), protocols (e.g., TCP/IP), programming languages (e.g., Ada), open hardware standards (e.g., USB), and even hardware designs (e.g., Open Compute Project) and 3D models for home printing (e.g., Blendswap).

Much of the intellectual underpinnings of the open IP movement can be found in *The Wisdom of Crowds* [Surowiecki]. Surowiecki argues that these crowd-driven movements are less subject to political forces and more dependent on expert knowledge, are necessarily more well-coordinated, and more trust is established than in plan-driven IP development by hierarchical teams. Interestingly, even the prerequisite initial funding for costly projects is now often obtained via crowd-funding. This further democratizes and expands the range of available open IP, which traditionally was limited on the higher end to a few organizations with appropriate resources and inclinations to share their results.

All crowd-sourcing applications, such as the aforementioned open IP ones, have four basic elements: a division of labor, computing and communications technology, a crowd of human workers, and a labor market [Grier]. Social media such as Facebook, Twitter, Digg, and others provide a ready platform for other kinds of open information applications. For example, social media information has been used to create contemporaneous trouble-spot maps and help relief agencies share information in response to disasters (Gao et al.). Large numbers of volunteer participants using community sites (e.g., PatientsLikeMe) have collaborated to share personal information in the creation of the medical data used for disease research and epidemiology (Reidl and Reidl). And crowd-sourcing has been used to

maintain contemporaneously accurate maps and to translate large quantities of text from one language to another. There are also various entertainment applications of crowd-source-like communities, such as in massively multiplayer games and in the creation of artistic and educational works. Similarly, there have been educational benefits from the proliferation of crowd-sourced classes online.

3.2.3 Challenges



Safety, truth, and accuracy: Is the information contained in open information repositories (e.g., Wikipedia) true? Is the open source software downloaded for use in a critical application safe to use, or does it contain a critical defect or a security flaw? Eric Raymond, one of the fathers of the open source software movement, contends that "with enough eyes, all bugs are shallow," but this observation isn't always correct. Crowds can be fooled, and collective intelligence can be wrong [Cox].

If open information creation displaces commercial information creation, what incentives are there for individuals to contribute? Not every human is altruistically motivated, and it is unlikely that someone can earn a living through the micropayments offered by some crowd-based initiatives. When information creators forego a copyright, there is a blurring of public-private relationships, and some measure of individuality is lost.

The distributed and often unchecked nature of the crowd-sourced worker can also lead to mistakes, cheating, and poor-quality work. While crowd-sourcing has built-in mechanisms for work-checking and fault-tolerance, these mechanisms are imperfect (Grier).

Ostensibly beneficial open intellectual movements could actually be ruses designed to trick human workers for some nefarious purposes. There are instances of such ruses being perpetrated already, for example, the notorious "Captcha Busting Trojan," in which a game was used to trick users into solving Captcha puzzles that were actually intended to thwart automated email account generation. Such an approach could be used to, say, recruit an army of volunteers who think they are working on some important mathematical problem into using brute-force techniques to crack passwords on a secure site.

Another large arena challenged by the open IP trend is legal. There is the obvious conflict with laws to protect IP from sharing and use, such as trademark and patent laws. Already, several notable examples exist where open IP was challenged in court (e.g., the *SCO versus IBM* lawsuit over Linux). As open data encompasses more and more fields, such as the ability to freely print 3D models at home, legal challenges will range from liability over manufactured parts (including weapons) to ownership of their design. Current copyright laws can limit the rights over a 3D model but are ill equipped to address the rights over the resulting physical output.

3.2.4 Where We Think It Will Go

Open IP generation will be very successful in certain niches, for example, encyclopedias, open standards, and open programming language. It will be only partially successful in certain niches (e.g., open access publishing, open source software). Open IP movements may fail in other domains.

3.2.5 Potential Disruptions

For certain market segments, it might be impossible for the free market to compete with open information counterparts, for example, in academic publishing, and we may see the end of the traditional paid-for scholarly journal.

Open IP could dramatically accelerate innovation, information dissemination, and quality of life improvements (particularly in disadvantaged nations). On the hardware side, open designs could accelerate technological developments and lower prices for devices from the hobbyist's toys to high-end servers.

The open movement could also greatly change the way society views IP ownership as it shifts from private to public.

A "scandal" involving flawed information in open IP (either through mistake or deliberate malfeasance) could cause a major disaster that calls into question the entire open information movement. The legal system has yet to adapt to the rapidly changing reality of open IP, and we may risk bottlenecking the promised progress in litigation and paperwork.

3.2.6 Summary

The open IP movement has moved beyond an experimental phase and will be a permanent fixture in society. How impactful this movement will be may largely depend on government actions or inactions regarding the treatment of this property; it might also be subject to the chaotic events of fate.

3.2.7 References

L.P. Cox, "Truth in Crowdsourcing," *IEEE Security & Privacy,* vol. 9, no. 5, 2011, pp. 74-76.

G. Huiji, G. Barbier, and R. Goolsby, "Harnessing the Crowdsourcing Power of Social Media for Disaster Relief," *IEEE Intelligent Systems,* vol. 26, no. 3, 2011, pp. 10-14

D.A. Grier, "Not for All Markets," *Computer*, vol. 44, no. 5, 2011, pp. 6-8.

J. Riedl and E. Riedl, "Crowdsourcing Medical Research," *Computer*, vol. 46, no. 1, 2013, pp. 89-92.

J. Surowiecki, *The Wisdom of Crowds*, Random House Digital, 2005.

3.3 Sustainability

3.3.1 Introduction

Sustainability in computer science is defined as a means of maintaining/preserving resources in IT service delivery to users. It is a confluence of supply and demand, where the IT ecosystem plays an important role (see Figure 2) [12].

Multiple Earth resources are the focus of sustainability. *Electricity* (gas, coal, etc.), for example, is critical in many datacenters, not only because it contributes to operational costs, but also because it impacts overall sustainability. The more power used from renewable energy sources, the more sustainable operations will be.

Water is used both to cool datacenters and to produce equipment. In certain areas, it is a scarce resource and must be handled with a lot of care—for example, in the Middle East and India. In some cases, water can be contaminated during the process and require treatment. *Carbon* is produced when burning fuel and needs to be removed by plants. *Materials* (steel aluminum, etc.) used during the production of various pieces of equipment must be recycled or otherwise contribute to ever-increasing garbage dumps. *Global warming* is a result of heating and cooling datacenters, in addition to other factors, and can have detrimental consequences to the Earth, especially as temperatures and water levels rise.

There are three aspects of sustainability in computer systems: *economical*, the financial impact of energy spent running CPUs, memory, networking, storage, etc.; *environmental*, the impact on the environment, such as how much CO_2 is spent or how much water is used in running datacenters; and *social*, summarizing the impact on the area where computer systems are executing, for example, the GDP of the region, stability of the region, any temporary influences, such as earthquakes, tsunamis, etc.

3.3.2 State of the Art

Today, sustainability-aware technologists prefer to consider cradle-to-cradle design, that is, resource consumption from a product's inception to its retirement. This includes all resources used to ship the product, its usage throughout its lifetime, and finally the recycling of it.

Standards have increasing importance in abiding by "green" energy usage guidelines, disposing of materials, and recycling equipment that is obsolete.

Technology can help in this regard by turning off infrastructure when it is not used or energy proportionate, optimizing the load by moving it around the datacenter to minimize energy consumption.

Sustainability can be achieved at all levels of the system, from savings in materials, such as NVM memories or photonics, which have much lower power consumption compared to DRAMs and electronic interconnects, to making tradeoffs in hardware architecture, including using dark silicon, which enables only parts of systems to be turned on, to using intelligent migration of virtual machines to enable consolidation and powering down parts of unused datacenters, all the way up to giving applications and services intelligent placement and design policies to enable optimal utilization.

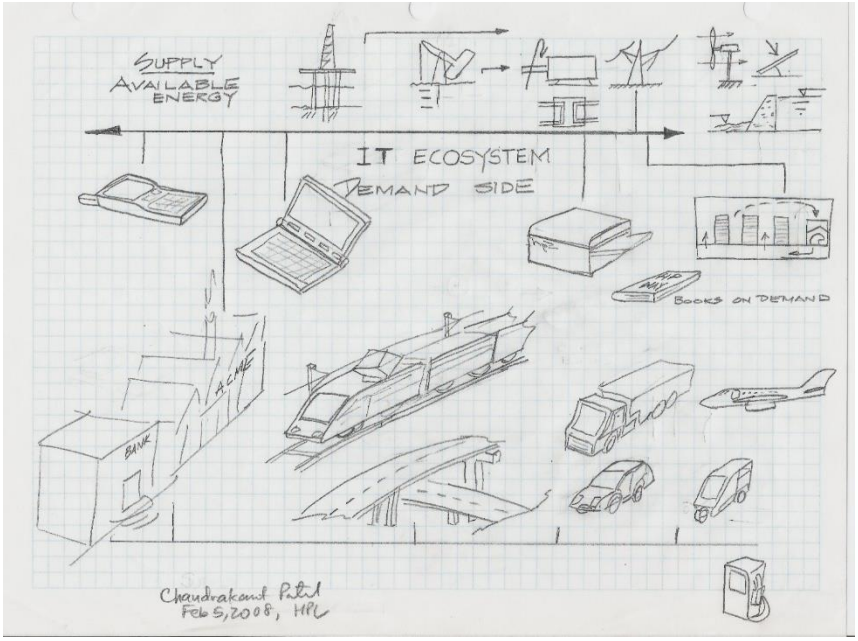


Figure 2. IT ecosystem from supply to demand.

3.3.3 Challenges

With the increasing population growth (3 billion more consumers predicted by 2030 [14]) and number of users connected to the Internet (expected to be near 3 billion by 2025 [15]), energy consumption will also rise. North America and Western Europe are saturated in terms of Internet users and mobile phones, but we can expect growth in continental China and India, South Asia, Africa, and South America. The amount of data produced is larger than the Moore's law equivalent in processing, and the Internet of Things will introduce additional data produced computer-to-computer and device-to-computer.

At the same time, scaling of technology, particularly for CPUs, has all but stopped, and new ways of using parallelism have been adopted. In the past, power was almost free and not on most people's minds. With increasing power consumption requirements, datacenters are now built near power plants or where ambient cooling reduces their cooling costs. Yet datacenter cooling still affects global warming. The increasing power consumption of hardware components has led to more power capping. Computer manufacturers have started to think upfront about recycling the materials used to produce IT devices. Computers are frequently assembled in areas where they will be sold or at hubs where the energy required to deliver equipment to customers is optimized.

However, the same production processes can be optimized in many other industries. Transporting computing products is the same as any other good. The opportunity to use technology in these areas are vast, such as deploying intelligent and sustainable data sensors, educating professionals, building and deploying sustainable resources into ecosystems to oversee processes, etc. [13].

3.3.4 Where We Think It Will Go

First and foremost, sustainability awareness is required at all levels. Technology can substantially help in many areas of productivity, making processes much more sustainable. Once sustainability can be measured, it can be controlled.

The second issue is the regulations and incentives governments can introduce to both prevent companies and individuals from malpractice in terms of sustainability and to encourage them to improve sustainability of their business.

Of most interest is power capping and power-driven management, the ability to think in terms of how much energy a program will consume, not just how long it will execute. For this to happen, more detailed instrumentation is required, to help software systems make policy decisions.

In addition, the increased configurability found in hardware, such as turning off parts of computers and using dark silicon, will enable better optimization in specific applications. For example, the existence of different CPUs (powerful or less powerful), GPGPUs, accelerators, FPGAs, etc., can be optimized for different applications, enabling better power utilization for different applications at different times and overall aggregate sustainability.

3.3.5 Potential Disruptions

There are several opportunities for disruptive improvement to sustainability. Frequently, it is the new or improved use of existing technologies that becomes disruptive. End-to-end resource management in manufacturing is one of the obvious examples. While it has been approached in computer equipment manufacturing, it has yet to be widely adopted in other areas of manufacturing and industries. There is a huge potential for conscious and sustainable approach to resource management.

One specific example of resource management is in sustainable or smart cities. The use of the Internet of Things further improves the benefits of smart cities, by enabling innovation at many levels. New environmental approaches to cooling with zero-energy datacenters combine solar energy with careful datacenter management, for example.

Electronic cars are another obvious disruptive technology whose benefits are in terms of reduced pollution and eliminating non-renewable energy sources. Remaining issues include the lack of acceleration (even though Tesla's line of cars addresses this), refueling time, battery capacity, and lifetime.

Light-emitting diodes (LEDs) are another example of an existing technology that can disrupt the future in terms of sustainability. Today's LEDs are used in automotive lighting (traffic lights, cars, planes, etc.) and are not widely deployed for general-purpose lightning. However, they have the advantage of being incandescent light sources in terms of lower energy consumption, small sizes, robustness, quick switching times, long lifetimes, etc. Once the cost is reduced and voltage/currency control improved, they could have a sustainable advantage of fluorescent lighting.

Consumer energy storage, new types of batteries (silicon anode; lithium iron phosphate), and renewable energies, including increased solar energy use and biofuels, can be disruptive technologies impacting many industries. Improved consumer and home energy management have many savings opportunities.

Appliance lifecycle assessment tools could predict when it is more sustainable to replace them; smart appliances can react and adjust to the grid disturbances and price changing to optimize consumption and cost; and a similar impact can be achieved by facilities energy management.

Another potential disruption is new generations of chips based on graphene and on metalferroids. In both cases, there might be a three-order magnitude reduction in power consumption [16].

3.3.6 Summary

There are many ways how technology can help improve sustainability. Big data analytics (see Section 3.18) and Internet of Things (see Section 3.15) will further enable and automate sustainable processes. Satellites sending images of air pollution could enable quick detection and early prevention. Governance, standards, and increased awareness will also help from the oversight and process perspective. Holistic approaches, such as cradle-to-cradle, will be increasing required. For example, it is estimated that there will be over 6 billion phones by 2017—only a holistic approach will be able to address this electronic waste. At the same time, cloud computing will help with reducing and optimizing power consumption through consolidating resources, and social media platforms, such as Twitter, can quickly increase public awareness in case of violations.

Sustainability has become an important factor in industry and public awareness. It has been substantially improved, but the growing needs are increasing a gap with the available reserves of water, energy, materials, and greenhouse gases. Therefore, humanity needs to continue and even increase sustainability to protect our future.

3.3.7 References

[1] X. Fan et al., “Power Provisioning for a Warehouse-sized Computer,” *ISCA*, ACM, 2007.
[2] P. Mahadevan et al., “A Power Benchmarking Framework for Networking devices," *IFIP Networking Conf.*, 2009.
[3] C. Belady et al., “Green Grid Data Center Power Efficiency Metrics: PUE AND DCIE,” White Paper 6, The Green Grid. 2008.
[4] R. Sharma et al., “Water Efficiency Management in Data Centers: Metrics and Methodology,” *ISSST*, 2009.
[5] R. Silliman, “Vendor Survey Analysis: Benchmarking Hardware Support Operations, N. Am., 2009,” ID Number G00165983.
[6] C.D. Patel et al., “Energy Flow in the Information Technology Stack,” *Proc. IMECE*, 2006.
[7] “The Ecoinvent v2 Database,” PRé Consultants; http://www.pre.nl/ecoinvent/default.htm
[8] T.J. Breen et al., “From Chip to Cooling Tower Data Center Modeling: Part I, Influence of Server Inlet Temperature and Temperature Rise across Cabinet,” *Proc. IEEE Intersociety Conf. Thermal and Thermomechanical Phenomena in Electronic Systems* (ITHERM), 2008.
[9] M. Manos and C. Belady, “What’s Your PUE Strategy?,” 2008; http://blogs.msdn.com/the_power_of_software/archive/2008/07/07/part-3-what-s-your-pue-strategy.aspx.
[10] World Bank World Development Indictors (WDI) database, http://data.worldbank.org/indicator.
[11] W. Adams, “The Future of Sustainability: Re-thinking Environment and Development in the Twenty-first Century,” *Report of the IUCN Renowned Thinkers Meeting*, 2006.
[12] C. Patel, “Sustainable Ecosystems: Enabled by Supply and Demand Management,” M.K. Aguilera et al., eds., *Proc. ICDCN 2011*, LNCS 6522, 2011, pp. 12-28.

[13] NSF Cyberinfrastructure for 21st Century Science and Engineering (CIF21), Advanced Computing Infrastructure, Vision and Strategic Plan. 2012.
[14] McKinsey Global Institute, McKinsey Sustainability & Resource Productivity Practice, *Resource Revolution: Meeting the World's Energy, Materials, Food, and Water Needs*, 2011.
[15] McKinsey Global Institute, "Disruptive Technologies: Advances that Will Transform Life, Business, and the Global Economy," 2013.
[16] Saracco, Personal Communication.

3.4 Massively Online Open Courses[1]

3.4.1 Introduction

Often drawing tens of thousands of students to a single section, massively open online courses (MOOCs) offer free, high-quality, university course content to anyone with Internet access. Requiring only a computer and Internet access to enroll, MOOCs can be used for continuing education courses and credit-bearing undergraduate courses, leading to degree programs and even graduation education.

The prospect of achieving huge economies of scale is alluring to deans and college presidents. World-renowned scholars can reach immense audiences. High-quality courses can be delivered to heretofore underserved and remote populations, particularly in disadvantaged countries, having enormous societal impact. These "universities without walls" have the potential to transform higher education. But there are significant unresolved issues relating to educational quality and financial sustainability.

3.4.2 State of the Art

A MOOC has two basic models. The first involves Web-based and emailed course content, with assessment achieved through automated exams. A notable example is Circuits & Electronics, one of the first MOOCs offered through EdX. The second "connective" learning model has less structure and content. The learning presumably occurs via crowd-sourced interactions through blogs, threaded discussion boards, and email. In either model, graduate assistants might moderate the interactions and answer questions, but instructor-initiated interaction is rare—if not nonexistent.

While online or remote delivery of college course content has been available for many decades, MOOCs differ in terms of scale and no-cost. Massive enrollments allow world-class faculty and curricula to be accessible to anyone. MOOCs can be taken anywhere that has Internet access, including sparsely populated areas, and those locations where it would be impractical to build a physical university. A MOOC will probably be completed by someone in Antarctica or on the International Space Station soon.

There are several major players in the MOOC space, including Coursera, a consortium of 33 colleges and universities; EdX, created by Harvard and MIT; Kahn Academy, backed by Google and Bill Gates; and Udacity. Currently, most MOOCs are taken as non-credit bearing, though several universities have recently begun awarding credit for completing certain MOOCs, passing additional tests, and providing certain authenticating artifacts.

MOOC courses can theoretically scale up without limit, from more than 100,000 students today to millions in a single course. To date, millions of course enrollments in MOOCs have been recorded, but it is unclear how many students have actually completed these courses and how many credit hours have been earned worldwide.

[1] Some of this article is adapted from Laplante 2013 with permission.

3.4.3 Challenges

Typical completion rates for MOOCs are less than 8 percent of enrolled students, which may include the curious as well as committed and ill-prepared students. These completion rates are an order of magnitude lower than in a traditional college course.

Assessment is another challenge. In order to allow for scale, MOOCs typically use multiple-choice, matching, simple fill-in-the-blank, and other forms of testing in which scoring can be automated. Some MOOCs require deliverables that must be assessed manually by instructors or teaching assistants, but these artifacts significantly limit course size.

Authentication of students is problematic, though this same problem exists for any online course. There are solutions available, such as using certified testing centers or biometric authentication. But these solutions can be expensive and logistically challenging and will limit the MOOC scale-up factor. Since most MOOCs use fully automated test grading, it is possible that an oracle will one day fool a MOOC test engine. We feel there is 50 percent chance someone will write a program that will pass enough MOOC courses to have obtained a degree by 2022, arguably passing the Turing test for artificial intelligence.

Critics of MOOCs highlight the lack of instructor-student and student-student interaction. While it is possible for some students to interact through group assignments, threaded discussion boards, and direct email, instructor-to-individual-student contact is limited to a select few students. In the United States, the Department of Education requires courses to have “significant instructor-initiated contact” in order for that course to be approved for financial aid credit.

Whether the MOOC is hosted by a not-for-profit entity or a for-profit business, the finances have to make sense. It takes significant investment to build and maintain the MOOC platform, fill course content and pay support staff, teaching assistants, and professors (if they are not working *pro bono*). A pure philanthropic model would see the financial burden met entirely through grants, donations, and earnings on some foundation. Some small financial successes have been reported, but no one has figured out how to make the finances work for MOOCs once they scale up and for the long run.

3.4.4 Where We Think It Will Go

The value proposition is so compelling that MOOCs will draw thousands of participating colleges and universities, thousands of investors, and millions of students from around the world, but in a limited way. Current MOOC offerings are targeted to the undergraduate market, but there will probably be a limited number of professional-, graduate-, and even doctoral-level MOOCs. Even today, however, there are signs of reluctance and disappointment on behalf of students, instructors, and universities.

We believe that most universities will either directly participate in MOOCs for a select few credit and non-credit courses or grant certain allowances to those who complete MOOCs, for example, by waiving a prerequisite if an appropriate MOOC has been successfully completed.

3.4.5 Potential Disruptions

With no tuition required, the convenience of online learning, and access to world-class faculty, MOOCs have the potential to draw vast numbers of students away from traditional bricks-and-mortar universities. A significant migration of students to MOOCs would threaten the viability of some

traditional colleges and universities, but we believe that there is a less than 10 percent likelihood that this disruption will occur.

MOOCs also threaten to change the role of faculty, student, and teaching assistants and the nature of the university. For example, one quality metric for traditional universities is the average number of students per class, with a lower ratio considered desirable. Automated course delivery and grading allows for immense upscaling of course enrollments. Does the growth of MOOCs mean we will need fewer professors but more teaching assistants? We believe that there may be pressures on traditional universities to scale course sizes by adopting partial MOOC attributes (e.g., more automated grading) but still preserving a high level of instructor-student interaction.

3.4.6 Summary

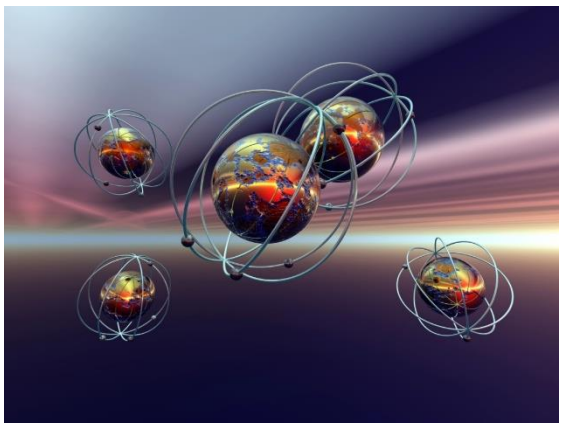
MOOCs have the potential to transform the higher educational landscape, but it is too soon to tell how significant this impact will be. MOOCs will likely play a future role predominately in continuing education, course prerequisites, and, on a limited basis, credit-bearing courses. It is unlikely, but possible, that complete credit-bearing courses from accredited universities will be available through MOOCs before 2022.

3.4.7 References

V.G. Cerf, "Running AMOOC," *IEEE Internet Computing,* vol. 17, no. 3, 2013, p. 88.

P.A. Laplante, "Courses for the Masses?," *IT Professional*, vol. 15, no. 2, 2013, pp. 57-59.

3.5 Quantum Computing



3.5.1 Introduction

Moore's law is still going strong and has been for several decades since Gordon Moore made his forecast in 1965. Continuing the pattern of Moore's law, we can expect that the limit for current lithographic manufacturing processes will be reached within the next few decades. While we will not speculate in the exact timing of its demise, it is a fact that current approaches to the fabrication of computer chips are starting to run up against the fundamental difficulties related to the extremely small scale of circuitry. As quantum effects are known to interfere in the proper functioning of electronic circuits as they decrease in size, we may reach the limit sooner. With time running out for Moore's law, it may be opportune to explore a paradigm shift from Newtonian or classic computing to alternative processing methods such as quantum computing (QC).

QC is based on the idea of using quantum mechanical phenomena to execute our computations instead of classical Newtonian physics. QC uses quantum properties to represent data and perform operations on data and offers—in theory—a decisive speed advantage over computers based on current technology. The promised speed advantage is so momentous that many researchers believe that no conceivable amount of progress in classical computer science will ever be able to bridge the gap between the power of QC and classical computation.

Shor's algorithm, [1] published in 1994, proved on a theoretical level that QC could efficiently factor natural numbers. The problem of finding efficient algorithms for factoring in classical computing remains an open challenge. In fact, the very lack of such an efficient factoring algorithm is the foundation for the security of public-key cryptosystems. Shor's discovery that QC can break the vast majority of cryptographic protocols in use today followed by a multitude of subsequent theoretical breakthroughs in QC research have generated significant public interest and kicked off a quest to build a practical QC device or a quantum computer.

3.5.2 State of the Art

QC is still in its infancy. Until now, experiments have been carried out in which QC has only been applied to a limited number of quantum bits, so-called qubits, the quantum equivalent to bits in classical computing. Numerous companies, academic institutions, and national governments support QC research to develop devices for both civilian and military purposes.

Practical and experimental QC is rapidly gaining momentum. From the beginning in 2001, when Shor's algorithm was first demonstrated by a group at IBM using a quantum computer with 7 qubits, to D-Wave Systems' 2007 announcement of the first fully functional QC device supporting 16 qubits, research efforts have started sprouting in many labs. More recently, D-Wave announced that a 512-qubits device would be installed at the new Quantum Artificial Intelligence Lab, a collaboration among NASA, Google, and USRA [2]. These organizations are investing in practical applications of QC because they believe it

may help solve some of their most challenging computer science problems, particularly in machine learning.

The fact that D-Wave's QC device is shrouded in a veil of commercial secrecy has raised questions about whether it has actually managed to build a viable QC device. While not all are convinced, a number of research papers exploring D-Wave's device are lending some credence to the claims made by the manufacturer.

D-Wave's QC device is optimized to find answers to problems that classical computers can only solve by exhaustively trying every possible solution, the so-called class of NP-hard problems. This QC device utilizes one of nature's own "algorithms," quantum annealing, which in a sense is hard-wired into the device's physical design. When datasets are transferred to the device, they are converted and represented as qubits. After that, the qubit configuration goes through a series of quantum mechanical transitions—quantum annealing—and a result emerges. The laws of nature dictate that systems want to sink to the lowest possible energy level with the most entropy. Fortunately, this particular property matches the core problem in most machine learning algorithms (minima detection in multidimensional space) and makes them ideal candidates for QC using D-Wave's device.

3.5.3 Challenges

What is the true potential of QC? Are there basic limits to our ability to control and manipulate quantum systems (qubits) that will prevent us moving QC from theory to practice and deploy real solutions on practical QC devices? A significant amount of research and development still needs to happen in this field to answer these crucial questions. D-Wave's QC device appears to be just one of many ways that practical QC can materialize. We need to understand what it takes to create general QC devices if at all possible.

QC is fundamentally changing our approach to computing and algorithm development. Deeply understanding the counterintuitive aspects of QC will be essential to fully exploit the potential possibilities it provides. With such a fundamentally different approach to both computation and computer architecture, few of today's computer scientists are well-equipped to take on this challenge.

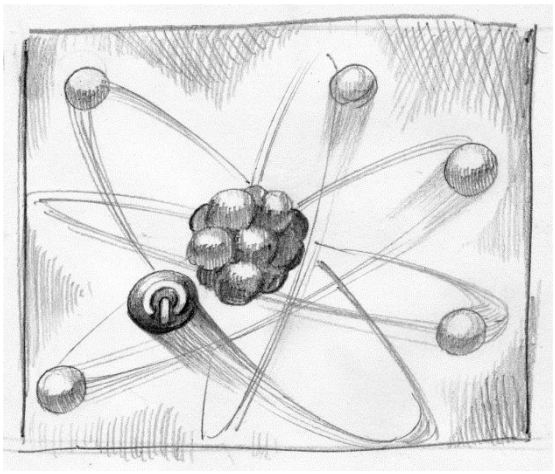
3.5.4 Where We Think It Will Go

Even the most immediate future of QC will be hard to predict, but we believe that early indications point in the direction of the integration of QC within large classic computing infrastructures where it will serve in specialized data-processing roles similar to what we saw in the early days of graphical processing unit (GPU) deployments dedicated to number crunching. Today, GPUs have become an integral part of datacenter servers and have taken over many tasks previously reserved for the CPU. Perhaps QC will take the same path.

There are other quantum effects worth considering, in particular, wave guide spin technology that promises dramatic increases in transistor density and a three-order decrease in power consumption. This can extend Moore's law into the next decade. In addition, graphene is a possible replacement for silicon [3].

3.5.5 Potential Disruptions

Practical quantum computers will be able to solve a class of problems much more efficiently and quickly than classical computer systems. Whether it is Shor's factorization algorithm or quantum search algorithms, they will execute much faster than any current algorithm can on a classical computing system.



The true impact of QC and the path it will take is not yet known. The potential is staggering since this computing approach at its most fundamental level is only constrained by the laws of physics. During the Industrial Revolution, technological progress was driven and constrained by our understanding of thermodynamics and Newtonian mechanics—fast forward to the 20th century, when our deeper understanding of physics shattered these constraints, bringing innovations such as lasers, transistors, chips, and computing devices to the mass market. Even these spectacular technologies seem too rudimentary to exploit the full potential of quantum mechanics. With the emergence of QC, it appears plausible that we are about to experience a new wave of innovations that will tear down many existing computational barriers.

Research and development in QC by nature is much broader in scope and further reaching than earlier technological innovations such as the transistor. Yet the transistor's amazing impact proved hard to predict. We believe that despite that, QC is at such an early application stage, it possesses a novelty and a potential that suggests the likelihood of an even greater impact than the transistor has had.

3.5.6 Summary

Our understanding of QC is currently undergoing radical changes as it moves from being an esoteric branch of physics and information theory and enters into the realm of practical applications. As commercial QC comes within reach, new breakthroughs are occurring at an accelerating pace. There is now evidence that QC can revolutionize crucial areas from chemistry that could have a dramatic impact on drug design to data processing with its ability to efficiently analyze vast amounts of data.

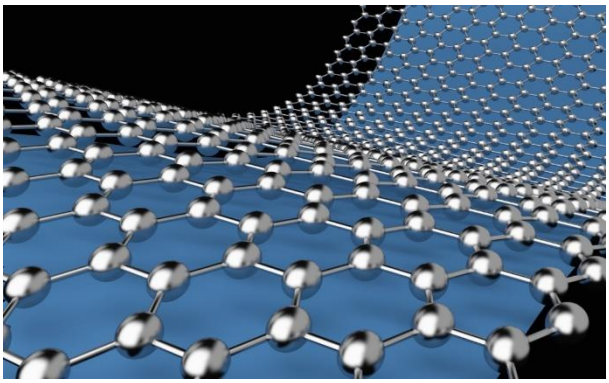
3.5.7 References

[1] M. Nielsen, and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2010.

[2] D-Wave, "D-Wave Two Quantum Computer Selected for New Quantum Artificial Intelligence Initiative, System to be Installed at NASA's Ames Research Center, and Operational in Q3," 2013; http://www.dwavesys.com/en/pressreleases.html#dwaveus_Google_NASA.

[3] Saracco, Personal Communication.

3.6 Device and Nanotechnology



3.6.1 State of the Art

MEMS and micromachines made from silicon are evolving into the nanotechnology field, where you might "imagine your life being saved by a custom-designed medical machine made from particles 50,000 times as small as a single strand of your hair" [GT-nanotech]. More generally, nanotechnology is about manipulating systems at the level of atoms, molecules, and larger structures. Popular depictions and current technology are showing the capability of rearranging atoms on a silicon substrate to spell a word or of moving them around to show a sketch or cartoon.

A wide range of science and engineering fields pursue nanotechnology, including biology and medicine, physics, chemistry, materials science, and other engineering disciplines. Nanotechnology is appearing in products like sunscreens and makeup, in automobile tires, and in vaccines. There are already cameras that can be swallowed (at least in the lab) and/or digested.

3.6.2 Challenges and Opportunities

Medical applications of nanotechnology appear to hold the most immediate promise for future computing environments: think of millions of extremely tiny sensors and actuators pervading some environment, like a human body under study, to understand and then fix it. Science fiction stories frequently raise the specter of self-healing bodies, where nanotech quickly heals a wound or rebuilds entire structures such as bones or organs.

The state of the art remains far from the active nanotech envisioned in sci-fi stories, but by 2020, we will likely see an increased use of nanotech-based devices in controlled settings. In medicine, swallowing little pills containing cameras may well be routine parts of office visits, with digestive processes removing them after some time, but will we have injections of nanotech into our bloodstream, to, say, better map the heart and the blood vessels connected to it or to trace blood vessels in the brain? Perhaps, but given the long lead times for safe medical technologies, we will not see nanotech that "cleans up" those blood vessels, removing debris, or that fights invasive organisms. On the other hand, it will be possible to manufacture such nanotech devices, creating a vision of millions of devices concerned with a single human body and their use in many natural and man-made settings.

There are many challenges in realizing the nanotech visions articulated above. There are ethical and privacy issues concerning "constant monitoring" by millions of tiny devices. There are dangers from long-lived nanoparticles, both in terms of potential unknown interactions with the human body and in terms of external influences able to use them for damaging rather than repairing human bodies. The popular press has already taken up this issue, worrying about nanoparticles entering the food chain or nanoparticles in sunscreen interacting with the human body. There are also entirely different issues, as when nanoparticles are used in advanced materials that raise entirely new challenges, such as the meaning of "metal fatigue" when metals are reinforced with nanoparticles. Can we still predict fatigue for such materials?

3.6.3 What Will Likely Happen

The use of nanodevices and nanoparticles has shown great promise (and profit, e.g., in sunscreen or makeup) in many fields. It remains unclear, however, to what extent and in what fields the dire visions painted in some of the sci-fi literature of "smart" nanomachines running amok will be realized. Certainly, current computing technologies still operate at length scales much larger than those of the nanoparticles used in current applications. Thus, it is really MEMS devices that have the capability of becoming increasingly smart and sophisticated in their actions. Studies have begun on nanoparticles' possible effects on humans and the ecosystem, but there will be a need for longitudinal studies, going beyond the short-term investigations already being carried out. There has not yet been widespread popular opposition, in contrast with the artificially induced changes in DNA that give rise to new plants banned in many countries. Whether there will be applications or usage models of nanoparticles that give rise to such opposition and ensuing legal or governmental actions remains unclear. Whether MEMS devices will find common application in medical and other areas by 2022 also remains unclear, in part because of the lengthy processes involved in launching new medical technologies.

Nanotech is frequently described in terms of "very small." But another important aspect is that nanotech is a different way of creating materials, from bottom up, as Mother Nature does, which is quite different from creating something top down. When something is manufactured in a top-down way, the original physical characteristics are not altered, but in manufacturing bottom up, one can design specific material characteristics [Saracco].

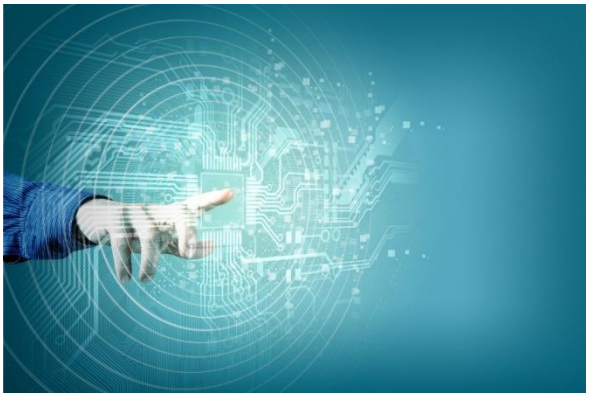
3.6.4 Summary

It is clear that MEMS devices, nanoparticles, and their use in a broad set of applications are here to stay, as the opportunities arising from their use are simply too numerous to ignore. It will be interesting, however, to watch their evolution.

3.6.5 References

[Saracco] Saracco, R., Personal Communication.

3.7 3D Integrated Circuits



3.7.1 Introduction

The desire to overcome the memory bottleneck caused by pin issues in planar circuits, along with the skyrocketing foundry costs of leading-edge process designs, have fueled the development of stacked 2.5D and 3D chips over the last few years [Kni12].

While the trend toward aggressive single-chip, SoC-level integration continues, some forces are also pulling in the opposite direction. A monolithic SoC is constrained to a single silicon process, cannot cope with mixed signal components, and has its volume economics reconciled with the nonlinear growth of NRE with complexity. The lack of cost-effective lithographic solutions is slowing down raw process scaling, and rapidly increasing volumes are required to absorb the design costs for each new process node, limiting the number of products that can be manufactured. Finally, several critical performance factors are shifting away from single-die CMOS scaling to system-level considerations, such as breaking the "memory wall" and providing more efficient paths to I/O.

Several factors are pushing stacking technologies to the mainstream, especially the desire to increase density (for volume-conscious products), to both decrease cost and power and increase performance [Ark12]. By using several smaller dies, stacking can enhance the single-die yield versus building a single large SoC. It can also avoid the capital cost recovery of the large NRE of a complex SoC design. And, of course, it reduces the bill-of-material (BOM) through the integration of multiple ICs in the same component. On the power dimension, the use of local low-power connections reduces the need of a large number of external power-hungry interconnects and PHYs, especially for memory. Using separate dies also enables adoption of separate silicon processes that can be power-optimized for a specific function. Finally, performance improves due to the increased interconnect speed related to the wire lengths (short-wide are faster than long-narrow wires), and the power recovery in interconnects helps offset the "dark silicon" problem.

3.7.2 State of the Art

In simple terms, a 3D-IC (also commonly called system-in-package, or SiP), can be seen as the modern incarnation of a multichip module (MCM). The two dominant flavors are PiP (package-in-package, mounted side by side) and PoP (package-on-package, mounted on top of one another). A variety of substrates are currently used in the industry [Woy13], ranging from laminates (similar to FR4 boards, supporting 5 to 25 layers) to ceramics (capable of hundreds of layers) to glass or metal covered with a layer of dielectric (around 5 layers) to semiconductors.

Relative to a single-die SoC, even a simple 2D SiP can provide several advantages, such as the possibility to mix signals, optimize the best technology process for each die, couple the IPs of different vendors, and offer greater flexibility with derivative designs using a different component mix.

The introduction of silicon interposers is what the industry refers to as 2.5D integration (Figure 3). It can support very fine tracks, enable active or passive configurations, and have mechanical properties (such

as the coefficient of thermal expansion) that match the individual silicon slices. The 2.5D packaging technologies provide tremendous increase in capacity and performance. While flip-chip bumps are around 100 µm, the micro bumps that connect 2.5D dies onto a Si interposer can be shrunk to about 10 µm. The interposer itself can be 200- to 700-µm thick and contain metal layer tracks (created using a standard Si process) and thru-silicon-vias (TSVs) that enable efficient connections between the upper layer and the package bumps. For example, the Xilinx Virtex-7 2000T device supports about 10,000 silicon-speed connections between adjacent slices.

The most complex form of SiP co-packaging is full 3D integration, which can also be combined with 2.5D integration to create very elaborate configurations (Figure 4). Unlike 2.5D, 3D integration directly connects multiple Si slices with TSVs etched in the dies themselves. This provides superior integration (no need for micro bumps) and higher interconnect density with TSVs (about 5 µm).

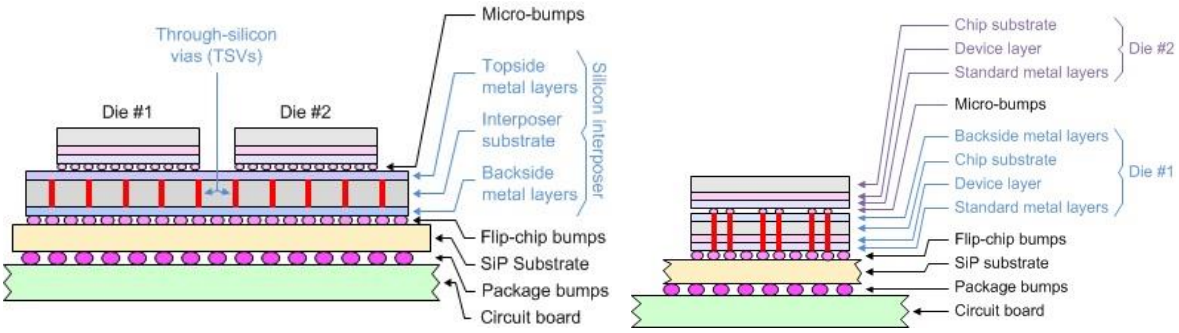


Figure 3. Two integration scenarios:
a 2.5D component using a silicon interposer (left) and a full 3D stack using TSVs (right).

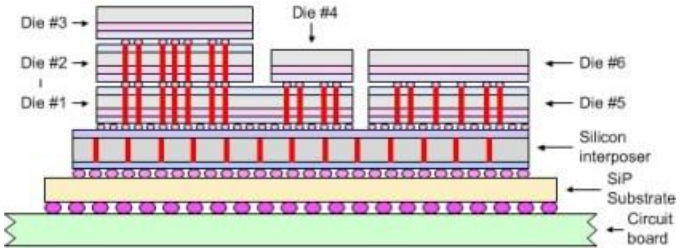


Figure 4. An integration scenario combining 2.5D integration of multiple 3D-stacked components.

3.7.3 Challenges

Major challenges remain in a 3D hybrid ecosystem, such as testing, dealing with the business aspect of compounding multi-die yields, and above all, management of the supply chain. Several players, such as Xilinx, Altera, Cisco, Huawei, and IBM, openly discuss their 2.5D and 3D roadmaps. However, it will take some time before 3D-ICs can reach mass production and at least until 2015 before full heterogeneous 3D (i.e., not just single-vendor memory chips) becomes mainstream. As of 2013, volume production of 2.5D and 3D SoCs has primarily used a "turnkey" approach, in which a single vertically integrated company provides both the front end (design of the individual ICs) and back end (testing and assembly) of the final part. The alternative "hybrid" approach, where the foundries deal with the front end and the packager with the back end, is constrained to niche specialty parts to date, mostly because of the ecosystem complexity of dealing with multivendor solutions.

For 3D-IC to evolve beyond vertical developments, all players in the ecosystem must find a way to work in a cost-effective manner. This implies providing fast turnaround times, defining a clear separation of responsibilities, and defining manufacturing and supply-chain roles.

Dissipating the heat building up within the 3D-IC is a major technical challenge, especially because these multiple high-speed components are placed in such a small physical proximity. New heat extraction technologies are required, especially to manage multilayer thermal hotspots and deal with the intermediate layers far from the package boundaries and heat sinks.

TSVs are large compared to other silicon structures (50 to 100 gates): placing them has significant impact on chip-floor planning. Manufacturability requirements for landing pads and keep-out zones result in placement obstacles. Because TSVs occupy the metal layers, they also result in additional routing obstacles.

Stacking could create effects that were never considered before, and signal integrity challenges emerge when dealing with die-to-die interconnects, shrinking wires and RC delays, unpredictable electro-migration, and so on. With the added complication of multipatterning, stress effects, and process variations, new design flows will become imperative to address some of these issues.

Separate testing of the independent layers is essential to keep yield issues under control. With full 3D structures, all but the first and last die are hidden, leaving no way to contact the stacked die for testing. Contact of test probes to thinner “naked” dies increases the probability of mechanical stress and fractures. For 2.5D integration, some of these issues are smaller, and constraining the TSVs to a silicon passive interposer eliminates the mechanical stress problem for active transistors.

Finally, the fundamental challenge of multivendor 3D-ICs is not technology or cost per transistor: it’s who takes responsibility when something goes wrong in a chip that costs several hundred million dollars to create but no longer works. At the foundation of the problem is the yield-compounding issue of multi-die packages, where a large stack can be “killed” by a single bad die. At 99 percent yield (of a simple die) and 5 layers, the compound yield goes down to 95 percent, which may still be acceptable. Stacking 8 larger dies, like CPU or memory, with a 95 percent yield results in a 66 percent yield, so 5 percent bad becomes 33 percent bad before any assembly loss. Even worse, if the individual layers have different value, a bad $1 die (e.g., a DRAM layer) can make a $1,000 stack (e.g., a complex CPU) into a keychain.

3.7.4 Where We Think It Will Go

Ultra-mobile and mobile products are already the first adopters of 3D-IC technology, but across the entire spectrum of IT products, we are increasingly observing a disruptive transition from printed circuit boards to 3D-ICs, and packaging will increasingly play a pivotal role in being able to provide value and differentiation.

Despite the technology and business challenges, we expect that over the next five years, 2.5D and 3D will become increasingly commonplace. Beyond mobile products, other cost- and energy-sensitive areas include the hyperscale server market, networking and storage products, and a variety of embedded applications, such as sensors.

Additional emerging technologies claim even better properties than TSV-based 3D. For example, the so-called "monolithic 3D" wafer-scale integration uses "patterned vias," about 50-nm wide, which translates into more than 10,000x higher vertical connections than TSV. It also uses a 100-nm thick silicon layer and yields a total reduction of 3x in Si area and 12x in chip footprint (a standard wafer with 8 to 9 metal layers could be 1-µm thick). We expect these (or other) technologies to mature by 2022.

3.7.5 Disruptions

As with any major technology shift, 3D-ICs will have a significant disruptive effect on the entire breadth of IT products, from mobile devices to enterprise servers. This technology will pose significant new threats to established players by fundamentally changing the supply-chain flows of important components, such as DRAM and CPUs. It will also create new business opportunities for the industry related to managing the co-packaging of IP blocks from the 3D ecosystem.

Because of the business challenges of multivendor 3D-ICs, we also expect a significant push toward vertically integrated products, where new or established players will act as catalysts to integrate complex 3D-ICs by leveraging a large portfolio of IP blocks (or dies) that will appear in the next few years.

3.7.6 Summary

The expectation in the semiconductor industry is that multi-die co-packaging will be a steady and rapidly growing trend to address these concerns. Combining SoC integration and co-packaging will help the continued scaling of power-efficient system performance, while enabling each die to be made in an optimized process node and enabling the design re-use of individual dies across multiple products. Co-packaging logic and memory dies can break the memory wall by using short-length, low-capacitance, and wide interconnects. Because of the nonlinear relationship of complexity and NRE, the cost of a very complex chip could also be reduced by co-packaging two smaller dies with half the functionality (for example, building a 16-core CPU out of a pair of 8-core dies). Different technologies are at play here, progressing from SoC (same die) to SiP (multiple dies on interposer, or 2.5D) to full die stacking (multiple dies with TSVs, or 3D).

3.7.7 References

[Ark12] S. Arkalgud, "2.5D/3D Scaling Walls (presentation)," *IMAPS Device Packaging Conf*., 2013; http://www.invensas.com/Company/Documents/Invensas_IMAPSDevicePkg2013_Keynote25D3DScalingWalls.pdf.

[Wy13] C. Woychik, "Emerging 3D and TSV Packaging Technology (presentation)," *SMTA Int'l*, 2013; http://www.invensas.com/Company/Documents/Invensas_SMTA2013_Emerging3DandTSVPackaging.pdf.

[Kni12] J.U. Knickerbocker et al., "2.5D and 3D Technology Challenges and Test Vehicle Demonstrations," *62nd IEEE Electronic Components and Technology Conf*. (ECTC), 2012, pp. 1068-1076.

3.8 Universal Memory

3.8.1 Introduction

The next five to seven years will cause very significant shifts to the IT infrastructure, and we believe that memory and processor architectures are two areas that will change profoundly. We focus here on memory.

Because of the charge retention issues and manufacturability challenges dictated by the laws of physics (Figure 5), and despite manufacturers' heroic efforts to continue scaling, DRAM's end is in sight [Mut13]. DRAM has had a remarkable lifespan of over 40 years, starting in the late 1960s when it was invented and then manufactured in 1970 by Intel. It has scaled in capacity by a factor of over 8 million, from 1 kbits on a die in 1970 to 8 Gbits today. From this perspective, DRAM's capacity has been one of the more consistent incarnations of Moore's law and has become one of the foundational commodities of the entire IT industry. Notwithstanding DRAM's incredible success, the number of memory manufacturers has been steadily decreasing: 20 in 1985, 11 in 1995, 8 in 2007, and only 3 today.

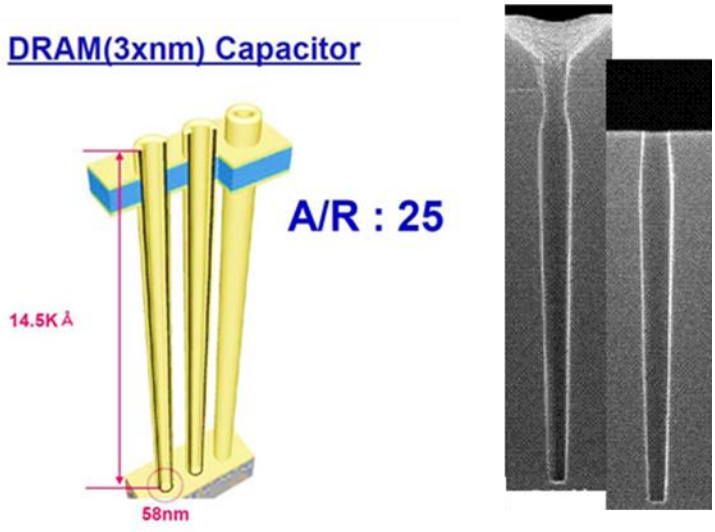


Figure 5. Illustration of the severity of the DRAM capacitor "trench."
On the left is a schematic representation of the aspect ratio of a DRAM cell in 3x nm process, showing a "trench" aspect ratio (depth/aperture width) of over 25x. On the right is the actual silicon cross section of two DRAM cells.

Between now and 2022, we expect that a new form of nonvolatile "universal" memory (NVM) will replace DRAM [Xie10]. While it is difficult to predict exactly when and how, we believe that this transition is inevitable, and there are already signs of it happening today. This new "universal memory" will combine the fast random access characteristics of DRAM and the nonvolatility properties of Flash. As such, it will have the potential to replace a large fraction of the memory and storage hierarchy, and

consequently cause a tectonic shift in architectures and the corresponding software to take advantage of it [Ran11].

3.8.2 State of the Art

The two most visible, expected, and desired metrics for memory are capacity (bits per device) and cost per bit. To keep advancing these metrics, DRAM manufacturers are resorting to 2.5D/3D packaging and stacking techniques such as the Hybrid Memory Cube (HMC) from Micron [Paw11]. The additional manufacturing steps and yield loss for this type of memory device increases the price per bit, but may be able to keep the capacity scaling growth for a few more years.

In parallel, the memory industry has been actively developing possible replacement technologies for DRAM, all of which are flavors of NVM. A survey of literature, patents, and manufacturers' disclosures indicates three technologies as the leading contenders: STT-RAM (spin-transfer torque RAM)[Mos05], PCM (phase-change memory) [Rao08,Lee09], and Memristor [Stru08]. They all have significant investments by the major memory manufacturers, show a different balance of advantages and disadvantages (which we are not covering here), and have in common that the state is defined by differences in the cell resistance (unlike DRAM, which stores charge).

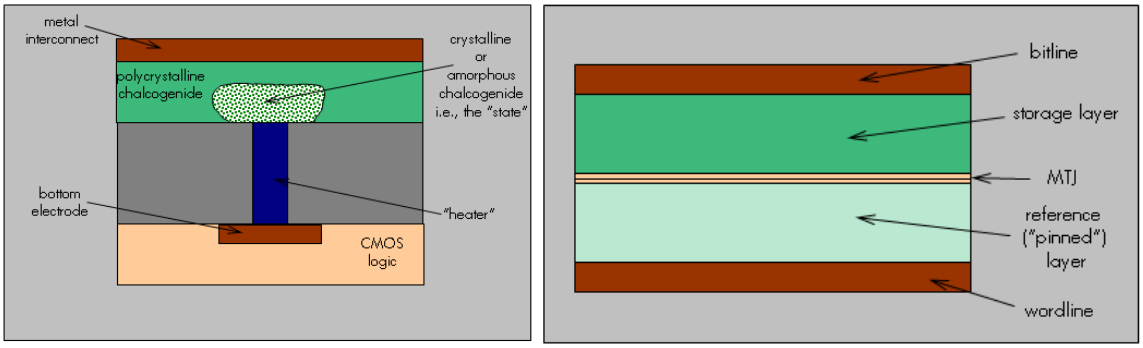


Figure 6. Simplified phase-change memory (PCM) cell (left) and spin-transfer torque (STT) cell (right).

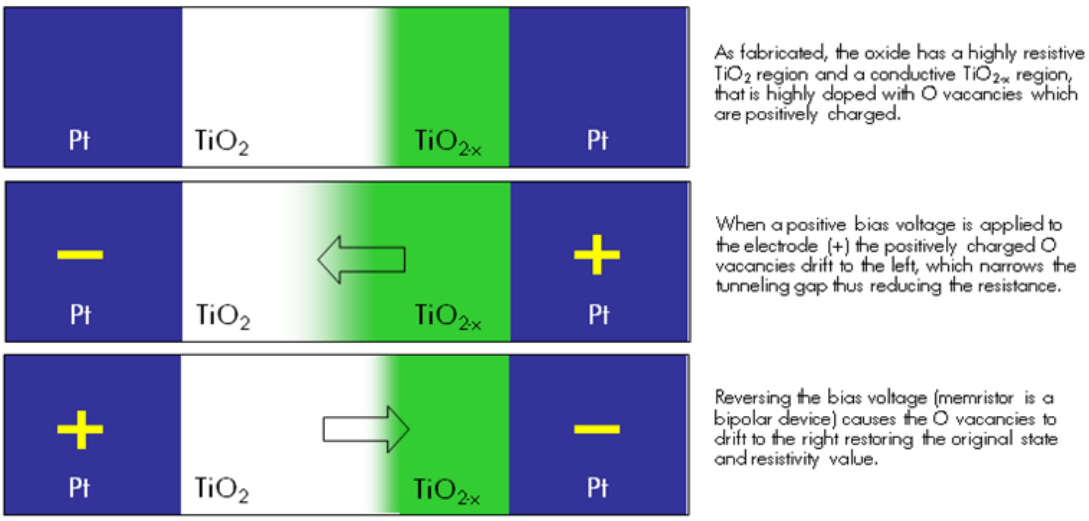


Figure 7. Simplified memristor (ReRAM) cell.

Of the NVM technologies, at least one (memristor) appears to offer a substantial greater bit density than DRAM [Rib12]. This comes from two factors: they are "crossbar" memories that do not need an isolation device per cell (leading to greater planar density), and they can be layered (on the same piece of silicon) as multiple planes for increased "effective" planar density. This is in addition to the stacking of dies within a 3D package.

3.8.3 Challenges

DRAM holds state through electric charge, but the shrinking of silicon structures has reduced the physical size of the cell capacitors to a point that makes it challenging to retain the charge. This is made even worse by the increasingly thinner insulation layer of deep submicron semiconductor processes. The industry is already starting to see scaling-related issues with DRAM, which is causing quality issues; there appears to be no remedy.

A second big DRAM issue is the ability to manufacture cells at increasingly smaller semiconductor nodes. The aspect ratio (depth versus opening) of the "trench" used to construct the DRAM cell already has an aspect ratio of about 25x. As the semiconductor node gets finer, the aperture (surface opening) reduces in size, making cell manufacturing increasingly difficult. Because the trench volume cannot be made arbitrarily small (it determines capacitance), DRAM manufacturability will be harder and harder.

The top NVM challenge is not related to technology, but business. Whereas DRAM has had an industry-wide consistency and commonality over time, we expect a much wider set of generational and manufacturer differences in NVM. As a consequence, no individual NVM technology will likely have the longevity of DRAM. This is especially true as the technologies climb the maturity curve, which means that there may be a two-stage market for the DRAM replacement: a given media may get to market first but lack all the features relative to a later arrival, and the important market players could switch to the better media.

3.8.4 Where We Think It Will Go

We expect at least one NVM technology to reach maturity and volume manufacturing capabilities within the next three to five years. The first place where NVM will materialize is most likely going to be today's primary storage, top-tier layer, where NVM's superior properties will gradually replace NAND and NOR Flash. This will happen across the board, from mobile client devices to high-end enterprise storage products. The SNIA NVM Programming Technical Working Group is already actively working on developing a new NVM Programming Model so that hardware and software vendors can align their effort behind a well-defined standard that presents a simple and consistent method of exposing persistent memory to applications.

In parallel, we also expect NVM to appear in the main memory space, initially as a memory extension, similar to the efforts already appearing today that combine Flash and DRAM [Diablo Memory Channel]. This is when the real disruption will occur, since the presence of NVM in the memory space has the potential to fundamentally change the way in which we persist information in the storage layer: we no longer need to think in terms of serialization and deserialization.

As NVM universal memory appears, there will be consequences in other components of the compute infrastructure, the first being the memory controller. Since about 2005, memory controllers have been integrated with the CPU chip, so that the memory itself (DRAM and DIMMs) is no more than a passive slave to the microprocessor. We expect that this will change with NVM: because of the technology and architecture variability, the most logical evolution will involve breaking NVM access functionality into a high-level asynchronous protocol controller (which remains with the CPU) and a low-level media controller (integrated as part of the memory system itself). This way, the high-level interface can only specify the "intent" (e.g., read, write, copy requests), and the memory will be free to optimize and re-order low-level operations to better match media properties. As these interfaces standardize over time, they will also allow for some level of computation to move into the memory system itself.

Finally, it is important to observe that any incremental approach to remedying the current memory ecosystem and accommodating NVM will be suboptimal and only delay the inevitable. The best thing the industry can, and should, do is to fully re-architect the memory ecosystem, but that involves fighting the inertia in existing legacy constraints. While a complete replacement would be the best technical solution, market forces and inertia will resist and possibly delay adoption beyond our prediction.

3.8.5 Disruptions

The availability of a much larger byte-addressable and persistent physical memory will cause a major re-thinking and re-architecting of end user applications and algorithms, as well as the operating system components (such as file systems and object stores) that are more closely related to storage.

The nonvolatility of a large physical memory offers several benefits to application writers. At a certain point in size, NVM can be viewed and used as the union of "memory" and "storage." Today's hard disks have very high random-access latency, relatively low bandwidth, access via I/O calls that require a long code path and OS context switches, and block-based semantics. With sufficient persistent physical address space, a large fraction of what today sits on a hard disk can be moved to NVM. The advantages are compelling: a short load/store path and simplified random-access semantics to access file systems or object stores, and possibly minimal or no OS involvement. Even more importantly, data structures can

be natively persisted without the need for serializing them to a disk-friendly block-based format—all with a bandwidth comparable to memory and a latency several orders of magnitude lower than today's storage.

For transactional applications that rely on high IOPS rates (I/O operations per second), we expect NVM solutions to gradually start replacing today's caching and I/O acceleration solutions. While these are quite successful today, they require more power, yield lower performance, add maintenance complexity, and have additional points of failure relative to an in-memory file system or object store.

Finally, the packaging of this large amount of memory requires further consideration. Due to the much lower power of NVM, we anticipate that the trend of stacking a large number of die slices within the same component will continue and accelerate. Coupled with the increased silicon-level density advantage of NVM media, we can expect line-of-sight of 20 to 40x greater density per part, most likely very different from today's DIMMs and probably more similar to an evolution of the recently proposed HMC 3D structure.

3.8.6 Summary

As DRAM approaches its end of life, we are witnessing the emergence of new NVM technologies that have the potential to address DRAM's scaling and capacity issues. We expect a gradual replacement will occur between now and 2022. These new NVM technologies have a set of characteristics that will make them amenable to becoming a "universal" memory that takes over the entire hierarchy from main memory to storage (or at least the top tier of storage). This will cause disk and Flash technology to move to lower-level tiers, a transition of similar disruptive magnitude to what happened when tape was ubiquitously replaced by spinning hard disks.

Because NVM technologies combine the fast access patterns of DRAM and the persistence and capacity of disks, they will cause a collapsing of the memory-storage hierarchy that will permeate all the way into the software we write across the board, from operating systems to middleware to applications. This will be a deep-reaching fundamental, powerful, and beneficial change.

3.8.7 References

[Hos05] M. Hosomi et al., "A Novel Nonvolatile Memory with Spin Torque Transfer Magnetization Switching: Spin-Ram," *IEDM IEEE Int'l Technical Digest*, vol. 5, no. 5, 2005, pp. 459-462.

[Lee09] B.C. Lee et al., "Architecting Phase Change Memory As a Scalable DRAM Alternative," *Proc. 36th Int'l Symp. Computer Architecture* (ISCA '09), 2009, pp. 2-13.

[Mut13] O. Mutlu, "Memory Scaling: A Systems Architecture Perspective," *Proc. 5th Int'l Memory Workshop* (IMW), 2013; http://users.ece.cmu.edu/~omutlu/pub/mutlu_memory-scaling_memcon13_talk.pdf.

[Paw11] J.T. Pawlowski, ''Micron Hybrid Memory Cube (HMC),'' *HotChips 23*, 2011.

[Ran11] P. Ranganathan et al., "From Microprocessors to Nanostores: Rethinking Data-Centric Systems," *Computer*, vol. 44, no. 1, 2011, pp. 39-48.

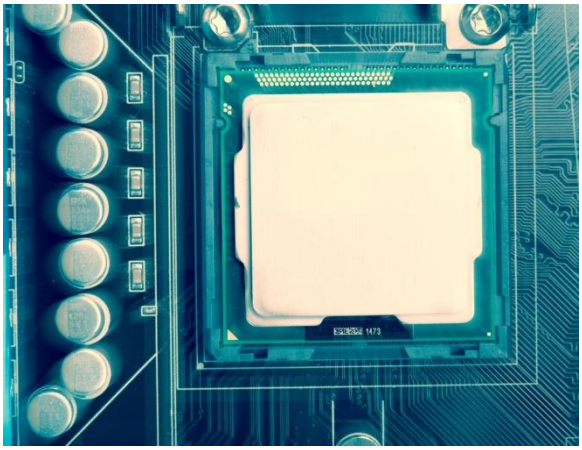
[Rao08] S. Raoux et al., “Phase-Change Random Access Memory: A Scalable Technology,” *IBM J. Research and Development*, vol. 52, no. 4/5, 2008.

[Rib12] G.M. Ribeiro et al., "Designing Memristors: Physics, Materials Science and Engineering," *IEEE Int’l Symp. Circuits and Systems* (ISCAS), 2012, pp.2513-2516.

[Stru8] D.B. Strukov et al., "The Missing Memristor Found," *Nature*, vol. 453, no. 7191, 2008, pp. 80–83.

[Xie10] Y. Xie, “Modeling, Architecture, and Applications for Emerging Memory Technologies,” *IEEE Design and Test of Computers*, Special Issues on Memory Technologies, 2010.

3.9 Multicore



3.9.1 Introduction

Multicore has attracted wide attention from the embedded systems community in such areas as automobiles, smartphones, cameras, tablets, PCs, and medical systems to high-performance computing systems such as cloud servers and supercomputers. It is known widely that the consumed dynamic power is proportional to the clock frequency cube. So, if we lower the frequency to 1/4, the dynamic power will go down to 1/64, and if we increase processor cores 4 times to compensate for performance degradation, the power will be 1/16. Also, we should consider static power caused by leakage current. To reduce static power, power gating on non-active parts is effective. In multicore, power gating can be applied to each small processor core and its local memories.

However, to obtain good multicore performance, software is key for decomposing an original sequential program into parallel program parts and assigning them to processor cores, to minimize the execution time (including the data transfer and synchronization overheads among processor cores). So far, such parallelization has been performed by application programmers, but it is very difficult, takes a long time, and has a high cost. Therefore, to use multicore in a wide variety of applications, automatic parallelization tools such as compilers will be very important [1].

3.9.2 State of the Art

There are many options for the low-power embedded multicore processors on smartphones and tablets, such as homogeneous multicores with 2, 4, and 8 cores, heterogeneous multicores combining super low-power 4 cores and ordinary 4 cores [2], heterogeneous multicores with low-power general-purpose processor cores and accelerator cores like GPU cores [3]. The accelerators are very important for realizing low-power computation since accelerators give us high performance with low clock frequency. However, programming for GPUs is often difficult and time-consuming, and the communication overhead among general-purpose processor cores and GPU cores is sometimes very large. Coping with these problems will be crucial for the next generation of heterogeneous multicores.

For routers, servers, and supercomputers relatively high-performance multicores and many-cores are becoming available. For example, 8- to 16-cores homogeneous multicores [4] or more than 50-core co-processors [5] are available for servers, and more than 100-core homogenous many-core processors have been planned for network processors [6]. As heterogeneous multiprocessors, the 8- to 16-core general-purpose multicores are connected with high-performance GPGPUs, including more than tens of Pflops supercomputers. In these high-performance systems, the most difficult problem is how to efficiently program many processor and accelerator cores.

Another important problem is how to realize low-power hardware and software combinations. The most advanced low-power technology is compiler control [1][7][8]. So far, low-power software has been realized in the operating system by power gating idle processors through virtualization among different

application programs. In the latest technology, each application program accomplishes low power through a parallelizing compiler—for example, a program is parallelized by the compiler, which inserts DVFS (dynamic voltage and frequency scaling), clock gating, or power gating APIs into programs to slowly operate or completely stop light load or busy-waiting processors for synchronization. Especially in real-time computation, program parts on the critical path are slowly executed by DVFS to satisfy a given deadline. With this control, program parts not on the critical path have more chances to be slowly executed or stopped. In other words, speedup by parallel processing gives us low-power execution for real-time computation, such as moving picture applications.

3.9.3 Challenges

The top challenges for multicore are as follows:

- low-power scalable homogeneous and heterogeneous architectures and their programming;
- hard real-time architectures with local memory and their programming;
- automatic parallelization and low power control;
- debugging and tuning tools;
- reliable architectures and software; and
- solar-powered multicores for everything from embedded to high-performance computation.

3.9.4 Where We Think It Will Go

In 2022, multicore will be everywhere, from wearable IT systems, smartphones, cameras, games, automobiles, medical systems such as drinkable inner-cameras for health diagnosis, cancer treatment systems that use carbon ions or protons, and solar-powered cloud servers to exascale supercomputers for super-low-power high-performance computation. Multicores and many-cores will allow us to recharge our smartphones just once a month or even enable solar-power recharging.

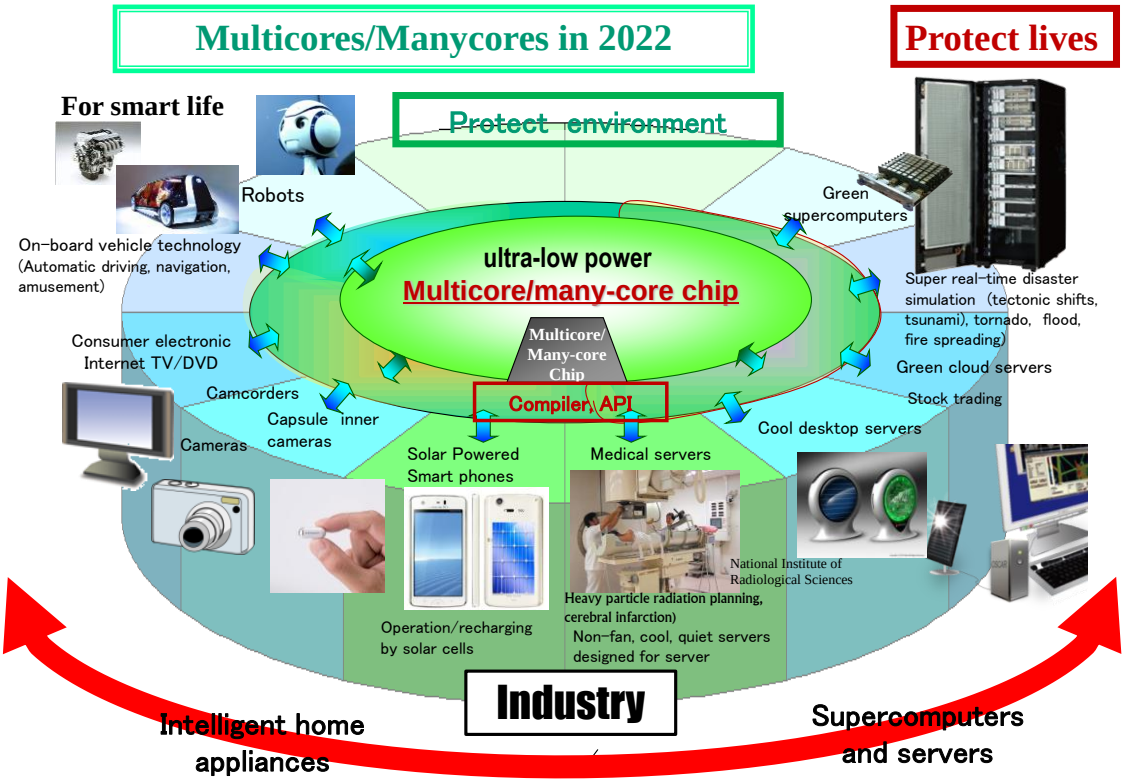


Figure 8. Multicore. many-cores landscape.

3.9.5 Potential Disruptions

A few technology innovations could disrupt multicore/many-core systems:

- *Automatic multigrain parallelizing and low-power compilers*. Multicore or multiprocessor systems have been researched or used for several decades with varying degrees of difficulty. Before 2022, automatic multigrain parallelizing compilers that use coarse-grain task parallelization, traditional loop parallelization, and fine- or near-fine-grain parallelization will be available for most multicore or many-core processors. The compilers will automatically insert DVFS and clock-gating APIs for dynamic power reduction and power-gating APIs for static power reduction with the efficient use of nonvolatile memory. The compiler will let application developers parallelize in a few minutes compared to several months of careful hand-written programing. Furthermore, the automatic power reduction with clock and power gating will reduce or entirely eliminate deadlocks caused by manual power tuning.
- *Many-cores for super low-power execution.* Many-cores will be used not only for high-power computation but also for super-low-power computation. For example, a small capsule could contain a camera that a person can easily swallow and continuously take pictures for 7 or 8 hours; the processor cores could perform pattern matching for cancer detection using a very small battery inside the capsule. This kind of application requires a 1/100 to 1/1,000 processor power reduction, which would require low clock frequency and low voltage via speedup through many-core parallel processing. More processor cores are important not only for high performance but also for low power.

- *Low-power multicores will be everywhere.* As killer micros took over almost all computer fields, lower-power multicore processors including low-power embedded multicores will be used by almost all IT systems in everyday life.

3.9.6 Summary

Multicores and many-cores will be everywhere from wearable devices, cameras, smartphones, automobiles, medical systems, cloud servers to exa-scale supercomputers in 2022. Those multicore architectures will be designed with the automatic parallelizing and power lowering compilers and multiplatform API to make parallel programming easier and power consumption lower. Such low-power multicores will open a road to a solar powered electronics society.

3.9.7 References

[1] Jun Shirako, Nato Oshiyama, Yasutaka Wada, Hiroaki Shikano, Keiji Kimura, Hironori Kasahara, "Compiler Control Power Saving Scheme for Multi Core Processors," *Proc. 18th International Workshop on Languages and Compilers for Parallel Computing* (LCPC2005), Oct. 2005.

[2] http://www.arm.com/ja/products/processors/cortex-a/cortex-a17-processor.php

[3] http://www.nvidia.com/object/nvidia-kepler.html

[4] http://www.amd.com/en-us/products/server/6000/6300#

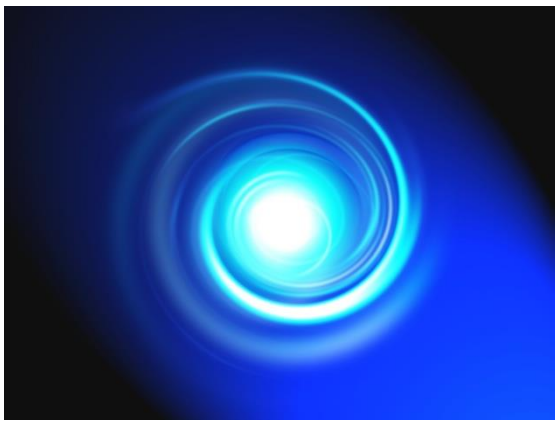
[5] http://www.intel.com/content/www/us/en/processors/xeon/xeon-phi-detail.html

[6] http://www.tilera.com/products/processors

[7] Keiji Kimura, Cecilia Gonzales-Alvarez, Akihiro Hayashi, Hiroki Mikami, Mamoru Shimaoka, Jun Shirako, Hironori Kasahara, "OS CAR API v2.1: Extensions for an Advanced Accelerator Control Scheme to a Low-Power Multicore API," *17th Workshop on Compilers for Parallel Computing* (CPC2013), Lyon, France, Jul . 2013.

[8] http://www.youtube.com/watch?v=M63W2RAjXfc

3.10 Photonics



3.10.1 Introduction

The technology roadmap for data communication faces three challenges: achieving *energy efficiency*, scaling *bandwidth* to track processor roadmaps, and delivering *low latency across systems*, with exponentially increasing numbers of cores and processors [Moo11].

The energy to move data today exceeds the energy to actually compute on the data itself [Dal10], and this trend is expected to continue. For today's high-end systems, the fraction of power and cost for communications is comparable to processors or memory. Hence, data communication efficiencies must be sought at essentially every scale from execution of instructions within the processor to the machine room floor.

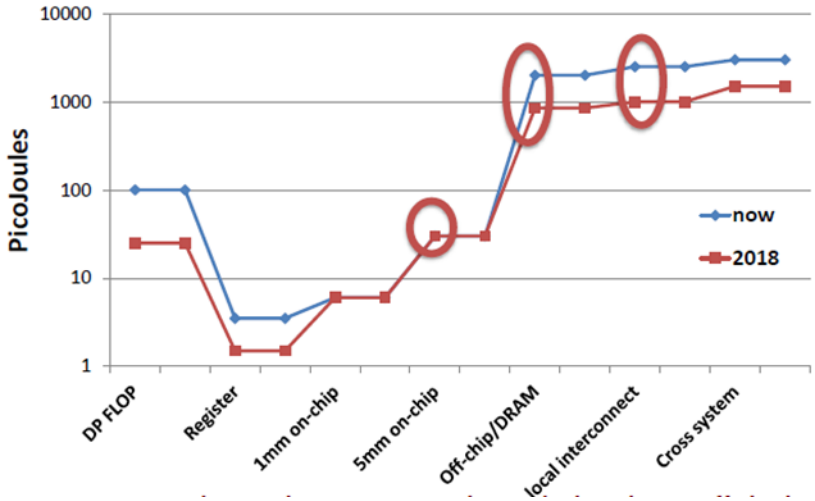


Figure 9. Data movement cost: the unbalance of computing vs. moving data energy efficiency [Sha13].

The fact that data communication is energy inefficient relative to computation and storage (Figure 9) is only part of the problem. Systems increasingly require wider bandwidths, and the communication energy growing in a nonlinear way, with bandwidth, makes the problem even worse.

At the processor-memory level, tighter integration of memory and processor using 3D-ICs will address several of the communication challenges. Looking at past trends shows that evolutionary electronic solutions will neither reduce the data communications energy nor substantially increase the bandwidths, and certainly not both at the same time.

For communication beyond the individual socket, photonic interconnects offer the best path to low-bit transfer energies and the bandwidth scaling needed to track increases in CPU performance [Ast09]. Emerging silicon photonic technologies for interconnect fabrics have radically different performance characteristics when compared to existing CMOS electronics solutions. Silicon photonics offers lower

power and higher bandwidth density, and eliminates the link-length restrictions associated with electronic interconnects.

There are compelling arguments showing that silicon photonics is a foundational technology for high-end systems [Bea11]. In the 2022 timeframe, high-end computing is expected to be in the exascale range. If an exa-operation application requires a communication ratio of only 0.04 bytes/operation, and each message goes on average through three hops (a very aggressive estimate), the total communication rate adds up to 40 TB/s, which at 4 pJ/bit per hop is about 4 MW, or 20 percent of the expected 20-MW system energy budget. Aiming at 4 pJ/bit per hop implies that the link energy has to be within 1 pJ/bit. Pervasive silicon photonics, all the way down to the compute elements, is the only technology that can reach this objective (Figure 10).

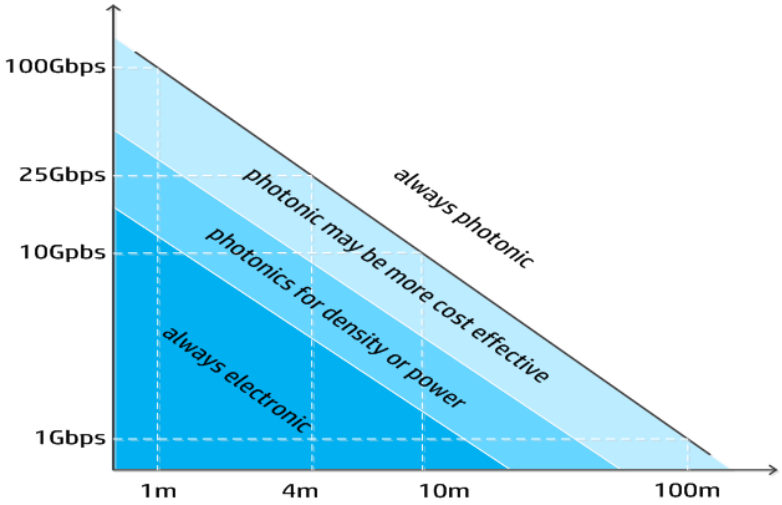


Figure 10. Rule-of-thumb of using photonics vs. electronics based on distance and required bandwidth. What emerges from the graph is that roughly above 100 Gbps per meter, photonics is clearly a win. Below 10 Gbps per meter, electronics is clearly a win (and a gray area in between). As we move to exascale and massively scale-out systems, the pressure for more bandwidth increases, and so does the appeal to use photonics for shorter-distance communication.

3.10.2 State of the Art

In high-end systems today, most optical interconnects use VCSEL-based transmitters, large area detectors, and multimode fibers [Bow13]. Several players are improving the cost effectiveness of these interconnects through simplifications in packaging and by increasing their bandwidth with improvements in VSCEL technology.

Silicon photonic transmitters with limited wavelength division multiplexing (WDM) capabilities have been demonstrated by several companies (such as Luxtera, Intel, and IBM) by using Mach-Zender-based modulators. Since these devices exploit a weak effect to modulate the light, they have to be relatively large in silicon real estate, which leads to high power requirements and limited scope for integration. Alternate approaches using resonant structures such as micro-rings as modulators have also been demonstrated by several companies (such as IBM, Sun, and HP). These resonators are more compact

and enable dense wavelength division multiplexing (DWDM), but tuning the resonators and matching them to an appropriate laser source remain unsolved technical challenges.

Finally, hybrid silicon ring lasers [Lia11] use rings of silicon waveguide as resonators and as a laser cavity stimulated by a layer of III/V material bonded to the silicon. As the laser's wavelength is determined by cavity geometry, several highly compact lasers in a range of wavelengths can be formed on the same substrate, simply by varying the diameter of the resonant cavity. Samples of these devices have been shown to be capable of direct modulation at 10 Gbps. The directly modulated ring laser has several advantages: no requirement for an external laser source and optical power distribution, a greatly simplified tuning, and power proportionality (the devices can be powered off when not in use).

3.10.3 Challenges

While limited WDM is possible with VCSELs, the cost of such links is still proportional to the bandwidth, as each additional wavelength requires additional components. Silicon photonics has the potential for much higher bandwidth density through WDM and lower power through the use of low-loss, single-mode fiber and waveguide detectors. With silicon photonics, bandwidth can be scaled by adding very compact transmitters and detectors to an integrated photonic die, at a minimal increase in the overall cost.

Although photonic interconnects are in principle more power efficient than electronic interconnects for rack-to-rack distances and beyond, the use of active optical cables (AOCs) has negated much of this advantage. The full benefit of optical interconnect can only be realized when the entire physical link path is designed for photonics.

A complete integrated photonic link requires detectors, optical drop filters, and a range of waveguide and coupling technologies to suit different applications, which in turn involves ecosystem and supply-chain issues that are being addressed. An additional problem lies in coupling light on and off the integrated photonic die: while several approaches have been shown (including tapers and grating couplers), significant challenges remain.

Finally, all photonics device technologies require innovative packaging that allows large numbers of single-mode waveguide connections to be made between devices and subassemblies. The development of appropriate packaging and connector and waveguide technologies is an obvious area where additional development is necessary.

3.10.4 Where We Think It Will Go

The design of packet switches for processor networks is constrained by the bandwidth density available at the chip perimeter. As the connections to switches span distances ranging 10 cm to 20 m, the link-length independence of photonics is particularly attractive. For these reasons, we believe that switches will be the first to exploit the benefit of close integration between high-density CMOS logic and silicon photonic communication. Figure 11 shows the time progression of photonics technologies, from active cables (single wavelength) to component-based photonics (CWDM) and on-chip interconnects (DWDM).

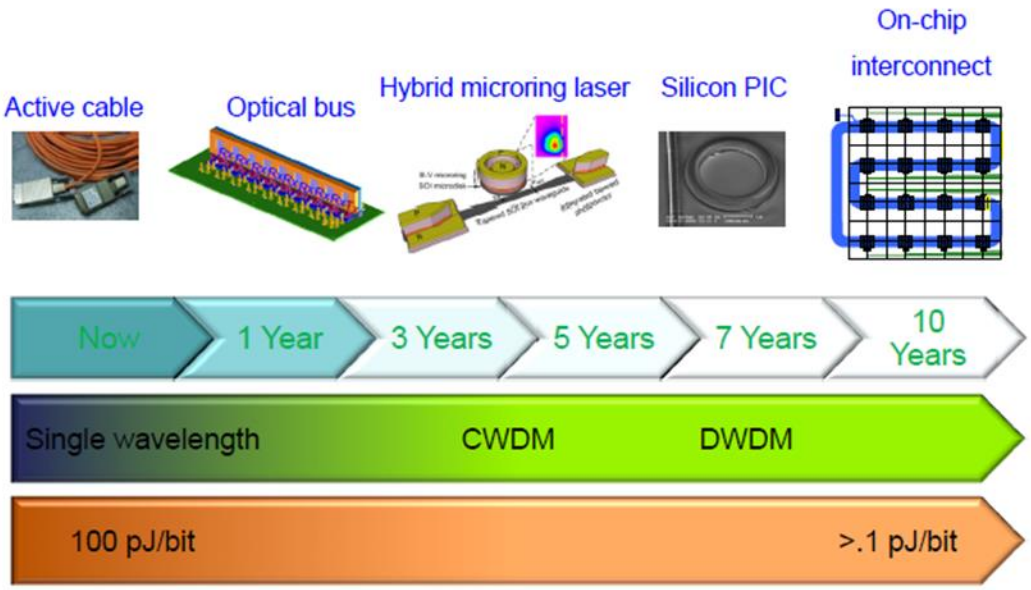


Figure 11. Roadmap of industrial photonics technologies. (source: HP)

DWDM silicon photonics will enable the development of very high-radix switch components while continuing to scale switch port bandwidths to track improvements in processor performance. Increasing the radix of switches allows low-diameter networks to be deployed with consequent advantages in latency, energy efficiency, and reliability. By exploiting silicon photonics, switch components with bandwidth up to 50 Tbit/s will become a reality. An attraction of high-radix network topologies is that the switches are distributed with the processors in a regular way, avoiding the wiring complexity of centralized switches.

By incorporating additional logic in the switch fabric, the network will become more intelligent, and operations such as collective broadcast and reduction, whose importance increases with larger node counts, will become a reality through intelligent support in the network itself.

3.10.5 Disruptions

A critical application of photonics will be to build highly energy-efficient router components that are interconnected optically, use CMOS electronics internally for packet processing and buffering, and connect to high-performance computing engines more efficiently than what can be achieved by co-packaging. Micro-solder bumps and face-to-face copper bonds allow much smaller connections between devices, allowing arrays of closely packed transceivers to be bonded to the CMOS switch device. This increases the CMOS device's effective chip-edge bandwidth, a performance bottleneck in today's systems, and enables the development of higher port count switches without reducing port bandwidths. Higher port count switches further contribute to lower communication energy by enabling networks with a lower diameter to be constructed that require fewer retransmissions

The availability of these new switches will enable whole new classes of network topologies that combine the ease of deployment of grids and meshes with the high levels of path diversity of logarithmic networks such as fat trees.

Network topologies such as HyperX [Ahn09] and flattened butterfly are ideally suited to high-radix switch components: with a high degree of path diversity, they have the potential to provide a highly resilient interconnect fabric scaling to millions of nodes.

3.10.6 Summary

Silicon photonics will be a fundamental technology to address the bandwidth, latency, and energy challenges in the fabric of high-end systems.

This area opens up the opportunity to build high-radix switches, with integrated support for important collective operations such as multicast barriers, reductions, scatter, and gather. In the 2022 timeframe, such a photonics-enabled high-radix switch could reach 64 to 128 ports, 640 Gbps per port, and 1 pJ/bit of link energy, which will enable connecting 1 million ports.

Bringing photonics inside chips has another effect: it gets rid of distance constraints, which in turns leads to flatter networks. A full photonics-based network is nothing but a giant supercomputer, where processing units are distributed geographically. This is going to change the software architecture of the switches in a telecommunications network and will eventually collapse telecom networks onto the computer's inner network, the one connecting chips in a supercomputer. This will create disruption for telco manufacturers [Sar14].

3.10.7 References

[Ahn09] D. Vantrease et al., "Corona: System Implications of Emerging Nanophotonic Technology," *35th Int'l Symp. Computer Architecture* (ISCA '08), 2008, pp. 153–164.

[Ast09] G. Astfalk, "Why Optics and Why Now?," *Applied Physics A*, vol. 95, 2009, pp. 933-940.

[Beau11] R.G. Beausoleil, "Large-Scale Integrated Photonics for High-Performance Interconnects," *ACM J. Emerg. Technol. Comput. Syst*., vol. 7, 2011.

[Bin11] N. Binkert et al., "The Role of Optics in Future High Radix Wwitch Design," *Proc. 38th Annual Int'l Symp. Computer Architecture* (ISCA '11), ACM, 2011, pp. 437–448.

[Bow13] J. Bowers, "Trends, Possibilities and Limitations of Silicon Photonic Integrated Circuits and Devices," *A Tutorial at the IEEE Custom Integrated Circuits Conf*., 2013.

[Dal10] B. Dally, "To ExaScale and Beyond (presentation)," *Supercomputing 2010 keynote*, 2010; http://www.nvidia.com/content/PDF/sc_2010/theater/Dally_SC10.pdf.

[Lia11] D. Liang et al., "Low Threshold Electrically-Pumped Hybrid Silicon Microring Lasers," *IEEE J. Sel. Topics Quantum Electron*., vol. 17, no. 6, 2011, pp. 1528–1533.

[Moo11] C. Moore, "Data Processing in ExaScale-Class Computer Systems (presentation)," *Salishan Conf.*, 2011; http://www.lanl.gov/orgs/hpc/salishan/salishan2011/3moore.pdf.

[Sha13] J. Shalf, "Active Power Management Technology Challenges and Implications for Programming Models (presentation)," *Teratec Forum*, 2014;

http://www.teratec.eu/library/pdf/forum/2013/Pr%C3%A9sentations/A4_01_John_Shalf_LBNL_FT2013.pdf.

[Sar14] Saracco, R., Personal Communication, 2014.

3.11 Networking and Interconnectivity

3.11.1 Introduction

Computer networks have become the basis for a broad set of critical applications, including sensor networks and their extension to the Internet of Things; Wi-Fi networks in homes, offices, and stores; high-end networks for large-scale parallel machines and on-chip versions of those linking the many cores and memories of many-core platforms; and, of course, the Internet itself, which is composed of a federation of networks connecting sites across the Earth and even reaching into the solar system.

In other words, computer systems cannot operate without connectivity between their components, whether on chip or in larger form factors like the networks connecting datacenter machines and, of course, the Internet.

3.11.2 State of the Art

Wherever power is plentiful, there have been great strides in communication technology. Datacenter networks, once ruled by 1-Gbyte Ethernet links, are fast evolving to 40-Gbyte+ interconnects, perhaps by 2020, even integrated into the same chips where data is stored and processed. It remains unclear, though, whether those on-chip interconnect technologies will be proprietary or open like Ethernet. Rapid improvements are also seen for home and urban networking, where Wi-Fi technology is improving rapidly, with many sites in cities now well-connected, even subways and buses. Yet there still remains a steep difference in connectivity availability between more versus less developed countries and within countries, between urban and rural areas. Nonetheless, between 2006 and 2011, for instance, the number of countries with commercially available fixed broadband grew from 166 to 206 (www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx), and the number of national broadband plans and policies in the world has more than doubled since 2009. Yet the ITU also estimates that while roughly 2.5 billion people used the Internet in 2012, this included only a quarter of people in the developing world, and in the US, there are still 19 million Americans who cannot buy fixed broadband Internet service. Wireless is more widely available, but even in October 2012, there were still 1.9 million Americans without access, and many rural access speeds remain low. As a result, while technology permits us to connect with high bandwidths, access to technology remains a limiting factor [bustamante13].

All this said, however, the world now operates with cellphones and, more and more, with smartphones, where in less developed countries, this former luxury is sometimes critically important to the daily lives of their citizens. Cell towers, in fact, are easier to construct than landlines, and the sudden connectivity they have created is replacing existing industries with new ones, one example being financial transactions via phones versus physical banks. In 2020, even more of our world will be smartphone-connected, with higher bandwidth connections in developed countries, and much more connectivity everywhere else, particularly in urban settings. In fact, for the first time, in 2013, sales of smartphones were no longer dominated in total volume by the US and European markets.

3.11.3 Challenges, Opportunities, and What Will Likely Happen

Many questions and issues remain concerning modern communication systems and infrastructures. For using them in extremely small devices, there are power issues, which make it unclear whether communications can occur continuously or whether there will only be intermittent connectivity, perhaps when devices can acquire additional energy and/or dock with other systems. The question arises because the network technologies needed for such communications are inherently limited by their energy consumption. Even on a single chip, we need vastly more energy/bit to move data between CPU and cache versus CPU and DRAM memory, and this ratio is getting worse as technology progresses (although on-chip optical interconnects offer some hope for increased bandwidth). Power issues are even worse when operating in cyber-physical environments, an example being nanotech devices in your body having to communicate with the outside world: this likely cannot be done without the infusion of outside energy, e.g., via radio beams directed at your body. Or imagine sensors in rivers or in the forest: water movement or wind/solar energy may permit them to communicate, but such energy must be harvested effectively.

At larger scales, communications require infrastructure, an example being the aforementioned cell towers, but there are interesting evolutions in this infrastructure. Specifically, regarding the ability of end devices to interact with the cloud, there is ongoing evolution from the current model—end devices either directly interact with each other, via peering, or with the Internet, via nearby communication endpoints like routers or cell towers—to a new model offering an additional intermediate layer, such as micro-cells, smarter home gateways, or public access points in, say, coffee shops. Those flexible infrastructure components will not only offer the communication support already present in current systems, but they may also provide useful computational or storage services offloaded from but still interacting with today's giant datacenters. An obvious example is data caching, but there are other, more interesting opportunities, such as orientation services warning cars away from streets currently under construction or experiencing a traffic jam. Thus, we are likely to evolve to a world in which communications become more tightly bound with services than those in the strictly layered systems now in place. This same trend, in fact, is evident in datacenter systems, where there is a rapid ongoing evolution from traditional to software-defined networks: the idea is to make networks more programmable to meet existing or future application needs. We do not further comment on those rich developments, as they are already reaching into standards bodies, but note that the broader topic of software-defined systems or datacenters is now an active field of study.

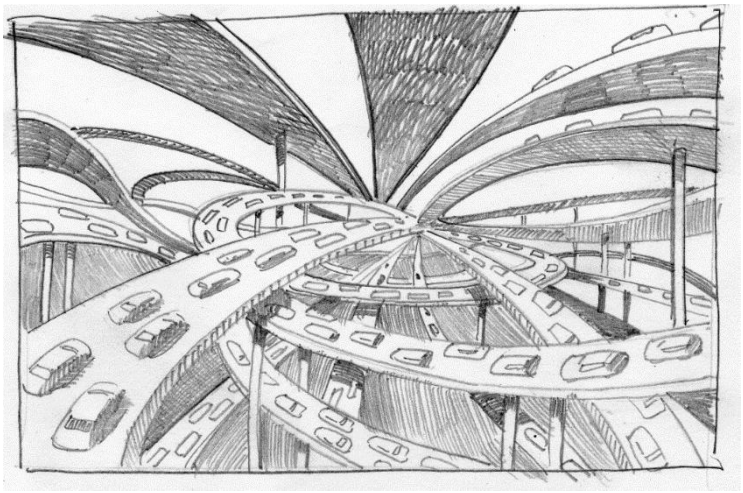
3.11.4 Potential Disruptions

Interesting developments underway today may lead to significant leaps forward in networking and communications. There is the promise of silicon photonics, able to move data at energy costs substantially less than with current technologies and to assist both with on-chip and rack-level communications. With these technologies, we can almost envision a rack of machines able to act like a single many-core chip, leading to levels of diversity in the processing and/or storage components available to applications much beyond that offered by current mostly homogeneous many-core chips. There are additional promises derived from more closely integrating services with communications. A promising beginning is in software-defined networks and the many network appliances in use and being developed for datacenter systems. Their services go beyond just assisting with communications to deep

packet inspection, data cleaning, threat detection and mitigation, and more. Will future services actually mine incoming data, continuously, to extract useful information and/or to discard data otherwise imposing undue load on back-end systems? An example of the latter is spam: Why should we first store it, paying those costs, to only then recognize its lack of value and discard it?

In the near future, within this decade, halo nets will become a significant player in the creation of the communication fabric. Indeed, one can say that the evolution is from telecommunications infrastructures (designed top down and owned by a few operators, requiring great CAPEX) to communication fabric (aggregated bottom up and owned by a variety of players with a variety of business and sustainability models). Terminals such as smartphones will generate these halo nets, creating a continuously evolving network at the edges of the big backbones. [Saracco14].

3.11.5 Summary

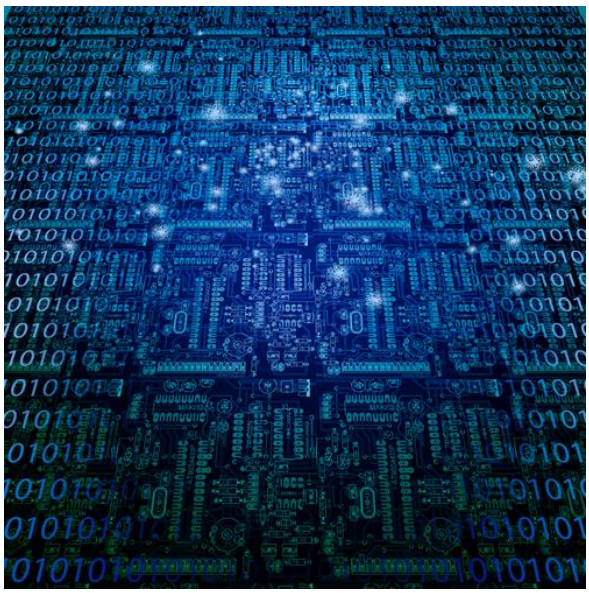


3.11.6 Communications and interconnects are seeing new opportunities, open issues, and potential disruptions from new technologies (silicon photonics), new use cases (online data mining), new challenges (the increasingly high energy costs of moving data), and infrastructure investments (like those in developing countries). Developments at all levels of the network stack, from interconnects on single chips to the worldwide networks connecting datacenter systems, will continue to drive the research and the Internet economy.References

[bustamante13] F.E. Bustamante, “Broadly Available Broadband,” *IEEE Internet Computing*, vol. 17, no. 5, 2013, pp. 3-5.

[Saracco14] Saracco, R., Personal Communication, 2014.

3.12 Software-Defined Networks



3.12.1 Introduction

Scott Shenker, University of California, Berkeley, was one of the core contributors to the OpenFlow protocol and has a one-sentence description of software-defined networking: an SDN is a set of abstractions for the control plane in networking.

This is very much a computer scientist's description, for abstraction is the key tool in the system computer scientist's toolkit. It means separating the function of a device from its implementation. This permits independent development of the implementation, and applications, of a device—programming language, operating system, architecture, or protocol.

Of course, this is ubiquitous in computer science and information technology, and examples are easy to enumerate: the x86 architecture (or any standardized ISA), the Linux kernel, the OpenStack API in cloud services, any programming language, and any programming language's standard library. In fact, computer systems generally are hierarchies of abstract architectures, with each one built on top of the stack below. Changes in implementation are essentially hidden from upper layers. For example, changes in the implementation of the Linux kernel are invisible to application programs, programs port easily between one x86 chip and another, and so on. Abstraction is so ubiquitous that we just assume it.

Given that, it is remarkable that, to this date, the network control plane has had no similar set of abstractions. The network data plane, in contrast, has an extremely successful layer of abstractions: the familiar standard OSI protocol stack, with its physical, media access, Internet, transmission control, and application layers. This set of abstractions has been standardized for a generation and has been stunningly successful: a Web server is (largely) independent of the Transmission Control Protocol implementation upon which it relies, and the physical layer is completely invisible to application programs.

However, for all of the success of abstraction in the network data plane, there is no set of accompanying abstractions for the network control plane. And, to quote Shenker, "this is crazy." With every new control protocol, network engineers have to re-specify and re-implement the general methods common to all control protocols: propagation of distributed state, failover, recovery from error, and so on. If this were done for (for example) storage systems, the filesystem API wouldn't exist, and each application writer would have to re-specify the layout of blocks on disk, error-correcting codes, a two-phase commit protocol to the device, and so on. We do this for network control protocols, however, and we do it all the time—so often that we barely notice we're doing it. For example, RFC 2328 specifies the Open Shortest Path First Protocol, the basic routing algorithm of the Internet. RFC 2328 runs to 250 pages, of which 13 are devoted to the method used to calculate the appropriate paths; the remaining 237 pages specify details of how local information is propagated to neighbors in the network graph, maintenance of distributed state, security and authentication provisions, and so on. Given a fully developed set of

abstractions for the control plane, OSPF—and virtually every other Layer 2 and Layer 3 control protocol—will be shortened by a significant degree.

We have only begun the journey toward standardizing and specifying a set of abstractions for the control plane. Shenker specified three general layers: the switch protocol, the network operating system, and the specification layer—a.k.a., the programming language. The initial switch protocol, OpenFlow, is now fairly mature; the network operating system is becoming so, with the emergence of a number of open source controllers. The final step, the specification language, is fairly nascent, with the emergence of the FreNetic language from Princeton and Cornell.

3.12.2 State of the Art

The state of the art in OpenFlow networks is the implementation of a standard protocol, in which a switch is reduced to a simple forwarding table. The table matches incoming packets based on specifications of the packet header—values of bits in specific fields, with wildcards. On match, one of four actions can be taken: send the packet out on a port, ask the off-board controller for help, drop the packet, or send the packet through the switch's normal processing pipeline (so-called "hybrid mode"). The most recent implemented version of the protocol offers the ability to match a packet multiple times, offering the prospect of multiple actions on a single packet.

The preceding description of OpenFlow suggests two things: one, this is within the capabilities of almost any switch on the market today, and two, there is less need for on-switch software in a "software-defined network" than in today's networks. Both these observations are correct: OpenFlow was deliberately designed to be implementable on the current generation of switches, and there is less software on a pure OpenFlow switch than in a standard switch. For the first, the original authors of the OpenFlow protocol observe, "While each vendor's flow-table is different, we've identified an interesting common set of functions that run in many switches and routers. OpenFlow exploits this common set of functions."

For the second, there is no more software in an SDN than in a classic network. An OpenFlow-based network routes and forwards packets via on-switch hardware, and provides no more services than any other classic L2/L3 network. The software in an SDN, defining the rules and policies concerning packet forwarding and transmission, has been moved off the individual switches and routers, and centralized and opened up to the network administrator. And thus the network as a whole is more transparent and more controllable by the network administrator. The switch is a simple forwarding table.

Since forwarding rules and policies, not the physical topology of a network, essentially defines what we mean by a "network," this factoring of the control plane offers the possibility of virtual networks. A virtual network is an application- or purpose-specific network with its own forwarding rules, segregated address and rulespace, quality-of-service guarantees, and admission control that can be set up and torn down on a dynamic basis. This is dream in classic networks: an easy reality in an OpenFlow network. You simply identify the virtual network by some combination of address space, VLAN tag, ethertype, protocol, and port and write the forwarding rules for this virtual network in a specification to the controller; the controller then forwards these rules to the individual switches.

This capability of OpenFlow of isolated virtual networks suggests per-virtual-network controllers. This in turn leads to the concept of a network hypervisor, which plays the same role for multiple network controllers that a hypervisor does for separate virtual machines on a single common substrate. The prototype network hypervisor is the FlowVisor, which partitions the matching header spaces among component controllers and permits each controller to write rules over its own header space (the "flowspace"). Rules from the controllers are sent to the FlowVisor, which checks to ensure that each controller is writing rules over its own flowspace and then transmits those rules to the switches.

Further, since the network itself is reduced to a collection of forwarding tables in an SDN, verification of network rules is much simpler. Verification of networks is difficult today because the control plane is embedded in the network and consists of switches running a distributed, Turing-complete computation; verification of this is undecidable. However, packets are always forwarded by the data plane; checking the packet-forwarding rules for compliance with desired properties verifies the network. Mathematically, verifying a forwarding ruleset is identical to verifying a loop-free logic circuit. This problem is far easier than verifying Turing machines (it is NP-complete instead of undecidable) and has been successfully attacked over a 25-year period in the VLSI industry.

3.12.3 Challenges

The primary challenge in implementing OpenFlow is that current-generation switch hardware was built for a very different networking use case, and existing switch ASIC pipelines fall far short of that required to implement a pure OpenFlow protocol. In order to explore alternatives to traditional forwarding with off-the-shelf hardware, we need to explain how that hardware is built to exploit the realities of existing forwarding rules and whether it is suitable for new forwarding requirements.

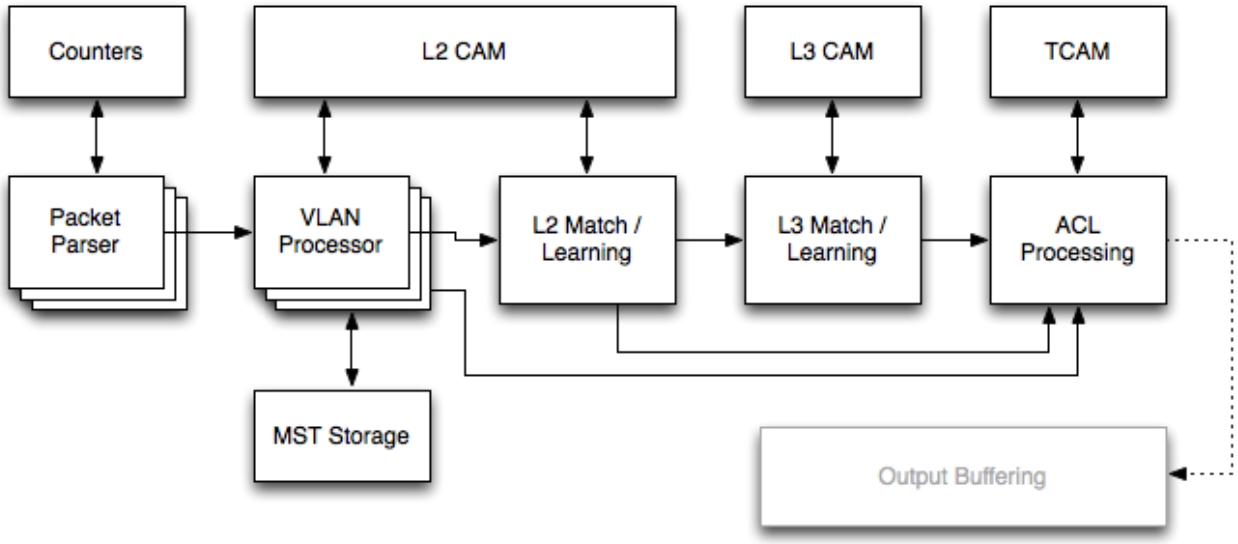


Figure 12. A representative switch ASIC pipeline.

In a traditional (pre-SDN "legacy") L2 forwarding environment, basically all switches are the same—they have slightly different pipelines depending on how versatile a single piece of silicon is (e.g., whether it is built to only be a switch, or whether it can also be sold for load balancers, firewalls, etc.), but they have

effectively baked IEEE and IETF standards into the silicon. To the extent that you want to do something that is nontraditional, you have limited flexibility.

Pre-SDN standard L2 forwarding is fairly simple and based on MAC learning and L2 destination-based forwarding. The network device learns what MAC addresses are connected to each port in the L2 MAC learning table (per-VLAN in a high-end device) and then builds a table for all forwarding decisions based on this information. This results in a fairly simple high-capacity forwarding table, which is optimized for a single field (typically, destination MAC address) and no wildcards.

ASIC pipelines have reasonably strict rules about how you get into various tables in the packet flow; they are somewhat configurable, but we must keep in mind that the vendor-imagined packet flow has a strong fidelity to existing standards. Thus, leveraging L3 matching tables often involves a previous stage in the pipeline indicating that a packet needs to be routed (often by marking a packet for a destination MAC address modification). This means that if you want to use the L3 table, you must mark a packet in this way, and if you mark a packet in this way, you should be aware that your packet is now going to traverse the L3 table, which may have adverse effects on your real intent. It is also important to keep in mind that regardless of the rewrite you may actually want to execute, there are various L3 capabilities that must be preserved, specifically in the L3 multicast case, for normal host and switch function to work properly.

As such, it is relevant to discuss common SDN (typically OpenFlow) firmware, which attempt to avoid the majority of these issues the easy way—by leveraging the table of last resort, the ACL table. This table is a small TCAM, usually a few megabits in size, and thus has a very limited match capability (and as discussed previously, the rewrite capability of the device may also be significantly limited). A TCAM is capable of wildcard matches (unlike a CAM, which must make exact matches), so is very useful for L3 and L4 matching. The mechanical implications of adding a “don’t care” bit to every possible part of a match include the very real problem that the die space required for TCAM is significantly larger than that required for CAM, limiting the amount available (and the expense of table space in general). The effective table size is further limited if L3 matching allows for IPv6, as this now significantly increases the number of bits required for each match line.

Given that existing devices are built for existing L2/L3 protocols, this table is further limited by the expectation of the vendor. A TCAM is not designed for standard protocol-based L3 forwarding—it is exclusively intended for use by the user or firmware developer to “correct” a limited number of standard forwarding behaviors. Because this table most directly maps to the 12-tuple OpenFlow match structure, it is very commonly the only table exposed by the firmware developer to the OpenFlow control channel. However, for many reasons, including those listed above, this table is very small, making it an ineffective place to make any kind of nuanced forwarding decisions at scale.

To some extent, this limitation can be overcome in a controller by using sophisticated algorithms to map OpenFlow rules to existing switch hardware while preserving semantics, maximizing the use of cheap, large memories, and minimizing the use of TCAMs. However, to date, these algorithms are not implemented in any controller, though their use has been explored in the literature.

Two other challenges are secure communication between the controller and the switch and making the controller robust against network failures and outages. Secure communication between the switch and

controller is in the OpenFlow spec, based on a Transport Layer Security implementation on both switch and controller. However, few commercial switches today implement TLS.

A robust, high-availability, distributed controller is a sine qua non for real OpenFlow deployments: no network can take a single point of failure. Fortunately, the design of high-availability, robust, distributed (HARD) software systems is now well understood. These lessons must be applied to the controller space.

3.12.4 Where We Think It Will Go

Many of the shortcomings that we discussed in the previous section are due to the relative immaturity of OpenFlow and the time constants inherent in new ASIC designs. OpenFlow today runs on a generation of switches designed for a far different use case; a new generation of switches, with flexible pipelines and larger TCAMs, will run OpenFlow more efficiently and far more effectively. Already, a number of existing vendors and startups are working on switches optimized for OpenFlow deployment.

The new generation of switches will be aided by improvements in controller technology, designed to optimize the ruleset. As we mentioned above, optimization technologies based on the algorithms used to minimize digital logic circuits are known but not yet incorporated in existing controllers. Preliminary investigations have indicated that substantial savings in rulespace are achievable using these techniques.

We expect significant improvement in the controller space, implementing features in the literature and addressing commonly recognized shortcomings. Controller technology has progressed rapidly, from Nox, which was a thin overlay on the simple OpenFlow API, to Floodlight, a scalable implementation, to OpenDaylight, which incorporates a number of optimizations for specific switch families. We expect future controllers to have the HARD properties and to implement the algorithms for both network verification and safe update that have appeared in the literature.

We also expect that the hypervisor-like capabilities of FlowVisor and FOAM will become an integral part of controller and network operating system design. In this picture, the developer of a distributed system will develop a virtual network as an integral part of his system, where the network control is simply a part of his application. The network operating system will then check to ensure that his generated rules work only over his virtual network and will mediate communication between the physical network and application.

If this picture seems exotic, note that it already exists for all other system devices. In today's world, a system developer creates and manipulates a virtual filesystem for his application, consisting of the files and directories he reads and writes over the course of the application. This capability will simply bring the network up to the programmability of other application resources.

We expect that the controller API will become standardized, much as the operating system API has become standardized through POSIX, and a number of different implementations with a common API will emerge.

Finally, we expect that the next few years will see the integration of the controller API with a distributed cloud controller, which will site virtual machines across a wide-area computing fabric. This is similar to the existing GENI mesoscale deployment in the United States, and we believe that this will become

standardized and ubiquitous. The fabric of the future will be a network of virtual machines, sited close to data sources and users, interconnected by SDN-enabled virtual networks.

3.12.5 Potential Disruptions

From the foregoing, it's clear that SDNs will be the most disruptive force in networking since the standardization of the OSI stack in the late 1970s. SDNs are the perfect complement to the OSI stack in the sense that they introduce a set of abstractions for the control plane to match the OSI stack's set of abstractions for the control plane. However, the impact goes far beyond this evident symmetry. OpenFlow and SDN herald a disruption of the networking market similar to the disruption in the server market caused by the introduction of the Linux OS on the x86 platform. Within a decade, the Linux-on-x86 had largely displaced the proprietary server stacks that had dominated the IT world: Solaris-on-SPARC, HPUX-on-PA-RISC, AIX-on-POWER, etc. In a similar sense, the router and switch market is today dominated by complex, expensive devices that are highly feature-rich. In the future, that functionality will move to the controller, and switches and routers will become commodity devices. Virtually every control plane protocol can be realized in a logically centralized software controller running on a standardized platform. This is a far more developer- and administrator-friendly environment, offering far greater visibility, controllability, and verifiability over the network.

SDN is likely to happen first at the edges of the network rather than in the telecom operators' big networks (although a few are experimenting with SDN). Furthermore, SDN goes hand in hand with NFV, or network function virtualization. They are not the same but are likely to leverage one another.

3.12.6 Summary

OpenFlow and SDN are the greatest advances in networking in a generation, and will change the fundamental activity from configuring the network to programming it. This will make the network far more secure, transparent, flexible, verifiable, and functional. Fully achieving this promise will take some years; a new generation of switches must emerge, and the promise of SDNs must be incorporated into the controllers. This will not be automatic or quick, but we know how to do it, and over time, these changes will take place and SDNs will exceed even the high expectations of today.

3.12.7 References

[1] N. McKeown et al., "OpenFlow: Enabling Innovation in Campus Networks," *SIGCOMM Computer Communications Rev*., vol. 38, no. 2, 2008, pp. 69-74; DOI=10.1145/1355734.1355746 http://doi.acm.org/10.1145/1355734.1355746.

[2] S. Zhang, S. Malik, and R. McGeer, "Verification of Computer Switching Networks: An Overview," *Proc. 10th Int'l Conf. Automated Technology for Verification and Analysis* (ATVA'12), S. Chakraborty and M. Mukund, eds., Springer-Verlag, 2012, pp. 1-16; DOI=10.1007/978-3-642-33386-6_1 http://dx.doi.org/10.1007/978-3-642-33386-6_1.

[3] A. Khurshid et al., "VeriFlow: Verifying Network-Wide Invariants in Real Time," *Proc. 10th USENIX Conf. Networked Systems Design and Implementation* (nsdi'13), N. Feamster and J. Mogul, eds., USENIX Association, 2013, pp. 15-28.

[4] R. McGeer and P. Yalagandula, "Minimizing Rulesets for TCAM Implementation," *INFOCOM 2009*, IEEE, 2009.

[5] S. Shenker, "A Gentle Introduction to Software-Defined Networking," *Lecture at Technion*, 2012; http://www.youtube.com/watch?v=eXsCQdshMr4.

[6] R. Sherwood et al., "FlowVisor: A Network Virtualization Layer," OpenFlow Switch Consortium, Tech. Rep., 2009.

[8] N. Foster et al., "Frenetic: A Network Programming Language," *Proc. 16th ACM SIGPLAN Int'l Conf. Functional Programming* (ICFP '11), ACM, 2011, pp. 279-291; DOI=10.1145/2034773.2034812 http://doi.acm.org/10.1145/2034773.2034812.

[9] P. Kazemian, G. Varghese, and N. McKeown, "Header Space Analysis: Static Checking for Networks," *NSDI*, 2012.

[10] M. Reitblatt et al., "Abstractions for Network Update," *Proc. ACM SIGCOMM 2012 Conf. Applications, Technologies, Architectures, and Protocols for Computer Communication*, ACM, 2012.

[11] S. Gutz et al., "Splendid Isolation: A Slice Abstraction for Software-Defined Networks," *Proc. 1st Workshop on Hot Topics in Software Defined Networks*, ACM, 2012.

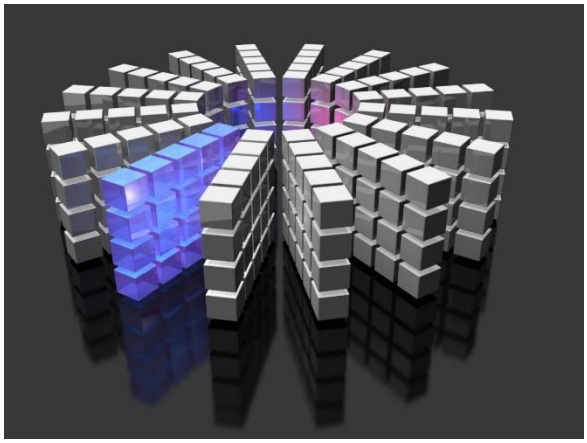
[12] R. McGeer, "A Safe, Efficient Update Protocol for OpenFlow Networks," *Proc. 1st Workshop on Hot Topics in Software Defined Networks*, ACM, 2012.

[13] R. McGeer, "Verification of Switching Network Properties Using Satisfiability," IEEE Int'l Conf. Communications (ICC), IEEE, 2012.

[14] The GENI Project. http://www.geni.net

[15] Saracco, R., Personal Communication, 2014.

3.13 High-Performance Computing (HPC)



3.13.1 Introduction

High-performance computing is a sector that entails hardware, systems software and tools, and applications/services. It is strategically important to many areas in industry, such as biotechnology, chemical, life sciences, pharmaceutical, national security and homeland defense, automotive, gas and oil, financial, weather forecasting, computer-aided engineering, and many others. Multiple vendors develop and sell HPC equipment, such as Bull, Cray, DDN, Fujitsu, HP, IBM, Intel, Mellanox, NEC, NVIDIA, and SGI, to name a few. The market is still divided based on server cost (from over $500,000 to less than $100,000) into supercomputers, division, department, and workgroup servers.

At the hardware layer, HPC systems are typically designed from compute-intensive processor farms with memories as large as possible (depending on application footprint) and specially designed interconnects to enable low latency and high bandwidth. Storage is optimized to receive large amounts of data, which is stored possibly in hierarchies and across sites. Because of the large processing power, these machines can be liquid-cooled instead of the traditional air-cooling.

At the systems software layer, operating systems are optimized to reduce any noise, to enable parallelism. The presence of noise (various daemons, TLB and cache flushes, etc.) accumulates and aggregates into delay, preventing applications to scale. HPC operating systems are usually stripped down or work with lean microkernels and runtimes.

3.13.2 State of the Art

State of the art at the very high end of the HPC field leap-frogs between the vendors exchanging the lead in the 10-Pflop range of supercomputers. The most recent entrants on the list are computers from China, including the current top computer 500 computer Tianhe-2, from the National University of Defense Technology, with almost 55 Pflops peak and over 3 million cores. (Reference top 500).

An even more interesting race becomes in the power efficiency of these computers. Similar to the large datacenters that host Internet providers (Google, Yahoo, etc.), the operating costs start to outweigh the capital costs. The primary contributor to operating costs is power consumption. Peak usage becomes a bottleneck as it becomes hard to bring sufficient power to supercomputers in certain geographical areas.

At the lower end of HPC, general-purpose graphics processing units (GPGPUs) are turning workstations into supercomputers. By including GPGPUs in personal servers and even laptops, it is possible to have the power of a supercomputer by using a high level of parallelism.

Current programming models still rely on established libraries, such as MPI (message-passing interface), although the degrees of parallelism will require better suited and finer granularity sharing. CUDA

programming relies on explicit memory sharing and requires programming wizards to extract parallelism at scale.

File systems have been optimized to sustain high-bandwidth throughput—for example, most of the Lustre file system's design is to work around the bottleneck of disk performance and its mechanical parts (moving head). This will likely change as solid-state disks and other new NVM technologies get introduced.

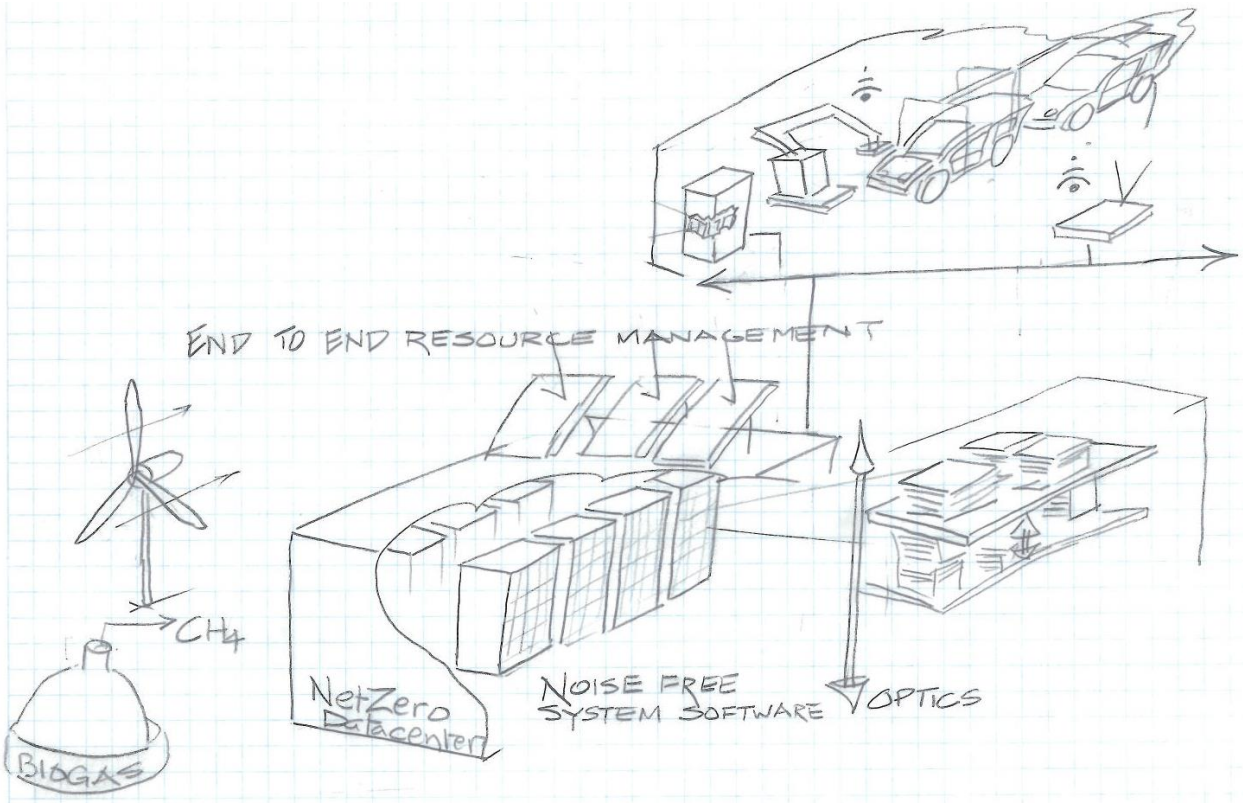


Figure 13. High-performance computing.

3.13.3 Challenges

The top challenges for HPC are as follows.

Scaling within power limits is the first and largest challenge, followed closely by reaching the next levels of performance, such as exascale. This can only be accomplished by careful optimization at all levels. Reaching the power budget of 20 MW for exascale would require 50- to 60-fold power savings compared to today's systems power consumption.

Interconnect bottlenecks are the second major challenge. With increased scale, there is a greater need to communicate across many systems, which puts stress on interconnect latency and bandwidth. Photonic interconnects offer potential here (see Section 3.10).

Low-noise system software that enables applications to increase parallelism has a similar importance to interconnects. At the scale of 100,000 nodes as predicted for exascale, the frequency of failures will be

much higher, while requirement to retain similar reliability to today's systems will remain. This will require new approaches to reliability at all levels of the system.

Ease of use of programming languages and tools that can enable applications to be more intuitively written in a parallel fashion are required to enrich today's MPI versus shared-memory (multi-threads) models.

New applications and algorithms that can leverage parallelism, computational power, and large amounts of available memory are needed to evolve existing applications and algorithms designed many years ago. Uncertainty quantification, combustion, and many new fields will be enabled, but scientist need to learn how to reason about them and how to program in these new fields, which were unfeasible to even consider with previous generations of IT.

Heterogeneity of the infrastructure with newly introduced accelerators and how to maintain compatibility with the systems software supporting them will be critical for enabling and leveraging dark silicon.

Finally, *managing the complexity at a large scale*, as well as heterogeneity of CPUs, storage, and software components will be a challenge to keep these large systems operational.

3.13.4 Where We Think It Will Go

Exascale is a goal set by many governments in the US, Europe, and Asia. Many teams are trying to achieve this level of computing at a predefined power envelope—for example, the US Department of Energy has issued a target of 20 MW.

At the other end of the spectrum, a lot of HPC scientists are pushing to move HPC to the cloud. This makes sense for embarrassingly parallel applications or for development at a smaller scale, but at the larger scale and, in particular, for finer granularity sharing, current virtualization technologies and multitenancy introduce too much noise, preventing HPC applications from scaling.

The figure below classifies different types of HPC systems and applications, indicating that the very high-end HPC applications will likely remain executing on dedicated supercomputers, just like some of the top banks still run on mainframes, while the lower-end applications will likely move to the cloud. The remaining "in between" applications have the potential to move to the cloud, assuming that it is adapted for HPC applications in terms of improved interconnects and virtualization (multiple references).

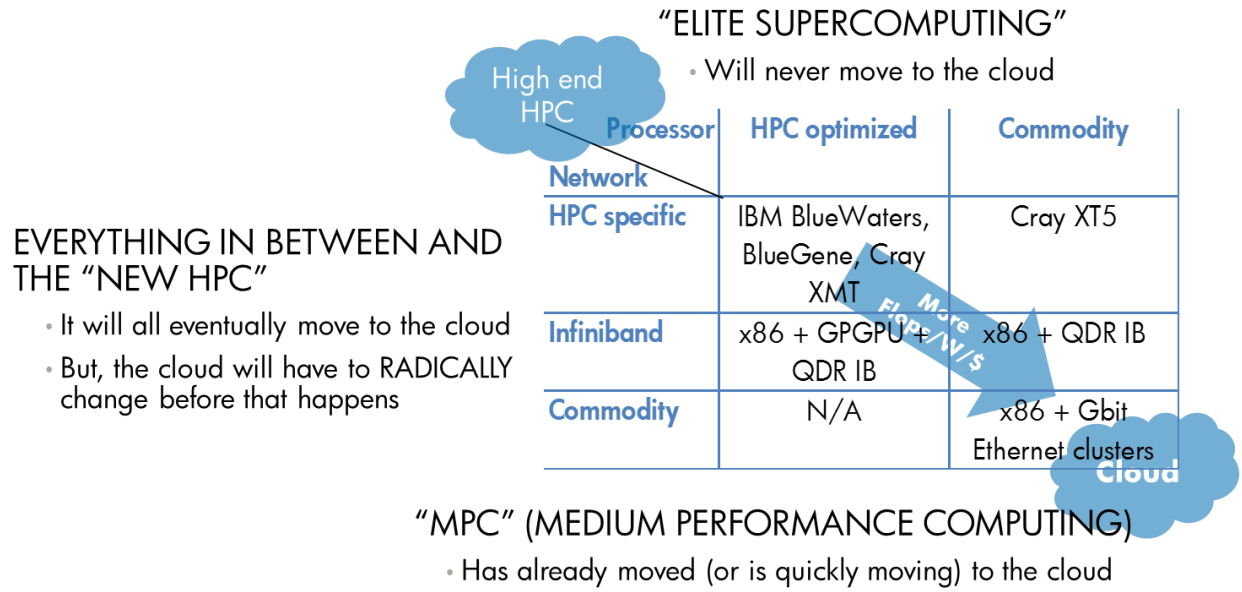


Figure 14. Comparing classes of HPC and their feasibility to deliver in the cloud.

3.13.5 Potential Disruptions

Some technology innovations could disrupt the rise HPC in particular, but also general-purpose computing.

Low-power components, such as ARM processors, are increasingly used to build high-end computers. They represent a perfect match for the Innovator's Dilemma model, where low-cost products enter the market with lower characteristics (performance, robustness, efficiency, etc.) and start gaining market share based on cost. As they improve in quality, they push the existing market leaders up the food chain. **Coprocessors and accelerators** also fall in this category and can likewise disrupt the state of the art of technology. They are important for power savings and for optimizing underlying hardware for specific applications and workloads.

Nonvolatile memory (NVM) is truly disruptive to computing in general but especially to HPC (see Section 3.8). NVM-based benefits for HPC are in terms of checkpoint-restart, file systems, and memory size. The long-term execution of HPC applications and the inevitable failures that increase with the infrastructure's increasing complexity and scale require checkpoint applications that can then restart from a previous checkpoint in case of failures. Because of the nonvolatility, memory checkpointing is not needed anymore, only the nonvolatile state, such as flushing caches and processor state. This will substantially reduce both checkpoint and restore times. The files produced as a result of HPC will be much more quickly written, removing the bottleneck introduced by disks. Finally, the larger memories enabled by new NVM technologies will enable new classes of algorithms, replacing previous algorithms that had to write the data to disks.

Photonic interconnects are critical to HPC because they will enable lower latency and higher bandwidth, decreasing the delays due to communication across parallel components in HPC systems. In addition, photonics will increase the scale and reduce power consumption.

Big data analytics continues to be a disruptor for many technologies, and HPC is no exception. Traditionally, HPC was both a producer and consumer of big data. However, new techniques and algorithms for big data analytics may be considered for use in HPC. One variant of this is the **ability to perform analytics (or any application processing) in real time.** As hardware capabilities increase, it will be increasingly possible to execute many more algorithms in real time or near-real time. There are many examples, such as simulations, financial what-if analyses, fraud detection, etc., that can dramatically change the way in which business is conducted in many market segments and verticals. In many cases, real time is back-end processing whose results are leveraged in real time, similarly to how Google prepares serialization of all pages on the Internet and then searches the serialized structure rather than parsing the Internet. In this way, a lot of data can be precomputed, such as insurance quotes, medical results, and others, which will further increase the demand for technology improvements and enable more functionality to be executed in real time.

Third-world countries entering the race for HPC will set new types of requirements for HPC, such as stringent power consumption guidelines, new cooling technologies in extremely hot countries, different kinds of support and reliability, etc.

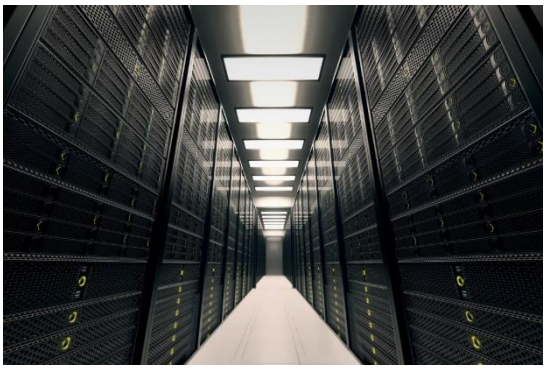
3.13.6 Summary

High-performance computing is still leading the advances in computing, but it is also being commoditized. Power bottlenecks are becoming the biggest challenge for advancing the state of the art. But new advances in NVM, photonics, and integrated circuits (see Sections 3.8, 3.10, and 3.7 respectively) are promising for overcoming new barriers, such as exascale. Yet, the next set of challenges will remain in programming models at the new levels of scale.

3.13.7 References

[1] http://www.top500.org/
[2] http://www.green500.org/
[3] Summary Report of the Advanced Scientific Computing Advisory Committee (ASCAC), "The Opportunities and Challenges of Exascale Computing," Fall 2010.
[4] "Magellan Final Report," U.S. Department of Energy (DOE), Tech. Rep., 2011.
[5] P. Mehrotra et al., "Performance Evaluation of Amazon EC2 for NASA HPC Applications," *Proc. 3rd Workshop on Scientific Cloud Computing*, ACM, 2012, pp. 41–50.
[6] K.R. Jackson et al., "Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud," *CloudCom'10*, 2010.
[7] "High Performance Computing (HPC) on AWS," http://aws.amazon.com/hpc-applications.
[8] T. Hoefler, T. Schneider, and A. Lumsdaine, "Characterizing the Influence of System Noise on Large-Scale Applications by Simulation," *Supercomputing 10*, 2010.
[9] Gupta et al., "The Who, What, Why and How of High Performance Computing Applications in the Cloud," Proc. CloudCom 2013, Best Paper Award; Also available at http://www.hpl.hp.com/techreports/2013/HPL-2013-49.html

3.14 Cloud Computing



3.14.1 Introduction

Computing has transitioned from centralized mainframes to clients networked locally to servers to the current generation of Web services and mobile applications using service-oriented architectures (SOA). A new generation of computing technology is emerging rapidly, in which computing is made available as a virtualized resource, accessible via a network. This emerging technology is called cloud computing. It is a response to a need for simplifying the administration and management of physical computing resources in order to focus on business logic and utility of the computing resources.

One may argue justifiably that cloud computing has existed since the early 1990s in the form of SOAs and network accessed services. There is, however, a new innovation that comes with today's cloud computing—it combines the virtualization of computing resources at all levels through automation, making these resources available for assembly and use remotely via networks. This definition of cloud computing was first invented at Amazon. Retail business is seasonal, and datacenters are designed to accommodate peak demand at any time of day or season. The result for Amazon was that many machines in its datacenters were idle most of the time and only activated during hours of peak demand. The opportunity to leverage these expensive resources to utilize them for computing services during off peak hours became obvious. The result was a revolution in computing, where computing resources became available as a utility service over the network.

3.14.2 State of the Art

The literature often implies that cloud computing is no more than virtualization, or that it is simply the next generation of automated hosting services. But cloud computing is actually the outcome of many advances in computing and communication technologies, particularly through virtualization that unleashes new opportunities for automation of what traditionally used to be manual tasks. The result is the ability to build IT applications without having to endure the long cycle of ordering and deploying equipment, setting up physical space for it locally or remotely, and having to babysit it 24x7 to keep it running, while still dealing with mundane issues such as power, cooling, and depreciation. In addition, with cloud computing services, IT leaders gain a new option for investing in IT equipment, which has become simplified to a pay-as-you-go model, meaning you only pay for the resources you use when you use them, enjoying the ability to scale your resources to support increased or decreased demand programmatically and almost instantaneously.

The term *cloud computing* has become a marketing catchall, but it is, in fact, a new technology. We offer a definition here that we hope will further clarify it, but we are certain that it will increase the debate as to what it is.

Definition

Cloud computing is based on a datacenter-scale virtualization of computing resources, in which through the collective automation of these virtualized resources, a virtualized subset of compute, storage,

connectivity, and application/middleware services are carved out to serve as a virtualized computing substrate accessed via a network.

There are several forms of manifestation of cloud computing, including infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). They all share the above definition in that each offers access to a virtualized computing substrate at a different level of abstraction.

IaaS

IaaS offers a virtualized substrate where the individual compute, storage, and connectivity resources are made more visible to the workload developer. As such, the toolset a developer uses allows for the instantiation of specific virtual machines with specific versions of an operating system, size of memory and type of mass storage, network attachment, and network subnets and VLANs all automated to create a mini-datacenter for a computing workload. Amazon's AWS is a good example of this approach.

PaaS

PaaS offers a virtualized substrate that appears as if it were an integrated single system; the compute, storage, and network configuration details are further hidden from the developer, and the developer is only concerned with working on that cloud computing platform as if it were a server platform. The toolset gives the developer access for specifying the layout and relationship between the components that make up a workload such as the relationship between the front-end, the application, and the back-end layers. Programming becomes specific to the respective cloud computing platform, and its portability is limited. Windows Azure from Microsoft is a good example of this approach.

SaaS

SaaS offers virtualized finished application/middleware services. The developer is offered a running, virtualized version of the workload required and provisions it to make it accessible to his/her end users. If it were just middleware, then the developer can use the middleware API to receive services that can be integrated with some other workload he/she is developing. Google App Engine and Force.com from Salesforce.com are good examples of this approach.

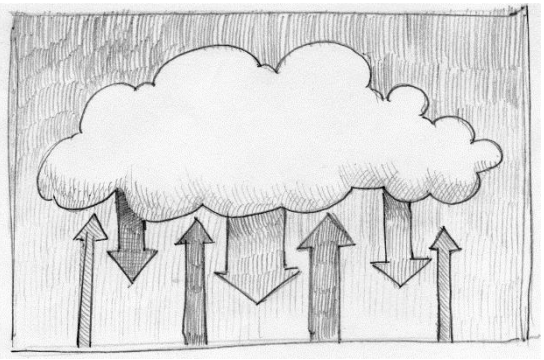
Cloud computing is also about employing automation to raise the level of resource utilization in a datacenter to significantly higher levels than what traditionally is possible, from as little as 15 percent without resource virtualization to as much as 50 percent with it. Furthermore, utilization can be increased to upwards of 85 percent with collective cloud computing automation of virtualized resources (DatacenterUtilization).

Public clouds are publicly available cloud computing services, and private clouds are datacenter implementations that dedicate datacenter resources to a private party. Public clouds are multitenant, whereas private clouds are intended for a single tenant. A secure subset of a public cloud can be used as a dedicated private cloud, such as a VPC (virtual private cloud) in the AWS offering.

Vendors today, including VMware, Cisco, and EMC, offer modular building blocks to construct datacenters that employ cloud computing technology for building private and public clouds. Public cloud services are available worldwide through mega-datacenters built by companies like Amazon, Google,

and Microsoft. Since access to cloud services is made via a network, accelerated network services, such as content delivery networks (CDNs) and edge compute networks (ECNs) tend to facilitate and augment cloud computing services.

3.14.3 Challenges and Opportunities



Building today's computing clouds is fairly complicated. Standards are scattered and lacking, and do not often follow a methodical approach. An effort to develop a "divide-and-conquer" approach to standardizing layers of building block services can go a long way in accelerating the penetration of cloud computing into datacenter infrastructures. Today's solutions are proprietary and rarely interoperable. IEEE and the IEEE CS are working toward standards in this space, specifically on P2301 cloud portability and commonality and P2302 cloud-to-cloud interoperability.

Layered Approach to Cloud Computing Infrastructure

Cloud computing is about managing a datacenter's virtualized resources collectively to present a unified system that can be allocated on demand and managed automatically, to deal with availability and resilience and to engage the pay-as-you-go model. We envision and propose a community effort to develop a layered model for constructing cloud computing infrastructures that will allow the industry to cooperate on creating interoperable modules that complement each other in a manner similar to the ISO-OSI layered network model developed in the early 1980s to address the complexity of creating interoperable networking solutions. Standardizing the building block layers of cloud computing infrastructure has the opportunity to achieve the same results seen in networking over the past three decades.

Virtual Connectivity for a Virtual Cloud Computing Substrate to a Workload

The purpose of clouds is to create and offer to developers a virtualized computing substrate on top of which to run a workload. Unfortunately, today's networking and connectivity technology to integrate resources from different clouds as well as resources from private enterprise datacenters falls short in enabling the flexible creation of the connectivity that could integrate such virtualized distributed resources into an integrated substrate. Furthermore, automated enforcement of security measures is lacking, as is federated identity management that could enable seamless interoperability among security zones from multiple independent identity providers. Software-defined networking solutions such as Nicera's (NiceraSDN) are steps in the right direction, but they do not address the full requirements in integrating resources across multiple clouds and enterprise datacenters. Project Sydney (Sydney) at Microsoft attempted to achieve this objective, but it failed to come to fruition for nontechnical reasons.

Developing for and Migrating to the Cloud

Provisioning a cloud environment is itself a challenge that requires resources with new IT skills that are also familiar with the requirements and capabilities of specific IaaS clouds such as AWS. The challenge is not in developing applications, but rather in provisioning the environment to host an application along

with the resources to support, operate, and monitor it. For each public cloud, provisioning the environment, deploying the application, and sometimes adapting it to run is proprietary and can take a lot of dedicated expertise. Nonetheless, the resulting benefits in time to market, total cost of ownership, and increased flexibility, scalability, and resilience outweigh the cost of the learning curve.

The case for PaaS is challenging, the equivalent of introducing a new server platform for development. The battle between Linux and Windows is a good illustration of the barriers involved, because this is an effort only giants like Microsoft can tackle. Although there are benefits to developing new scalable applications that can take advantage of cloud computing's benefits, the reality is that it is very hard to develop to a new platform without taking advantage of existing software. In the absence of a secure and virtualized connectivity solution that integrates the new development with existing software or enterprise data, such a platform is nice to have but not very useful unless the enterprise dedicates its entire development to the target platform and accepts the risks of not being able to migrate the application to other environments in the future.

The case for SaaS is more compelling, both for the service operator and the consumer. The model of finished services offered via a cloud to enterprises and users at large is strong, as service margins can be higher for the providing operators and the benefits of paying as you go to the consumer—with all the other cloud computing benefits of scalability, time to market, and not having to manage infrastructure—are obvious. There is an opportunity to offer middleware SaaS that requires developing standard interoperability interfaces that combine the same APIs between shrink-wrapped software and software offered as a service.

The Potential of Mash-ups

SaaS offers an interesting opportunity dynamically compose new services from existing ones. Publishing such services' APIs enables developers to build more sophisticated software applications by utilizing the APIs for finished services through mash-ups. The resulting solutions can be developed and brought to market in a much shorter timeframe and at a much smaller cost. Dependency on SaaS service continuity and how to assure continuity are a risk here. Standards may offer an answer. If standards are developed for such APIs, then the presence of multiple sources can alleviate this risk.

Big Data and Analytics

The emergence of clouds enables enterprises to utilize commodity computing to run applications that in the past were prohibitively expensive. Technologies like Hadoop and MapReduce let researchers tackle problems that were previously impossible. But IT expertise in using such tools is still rare and presents opportunities to introduce educational programs that can meet this need.

3.14.4 Where We Think It Will Go

We expect cloud computing to continue to improve, depending on how quickly the related challenges and opportunities are addressed and resolved. By 2022, it is likely that most new installations of datacenters will be based on cloud computing technologies. Computing as a utility will not likely become a reality by then. But short time to market in introducing application services will become the norm. The use of hosted application services in the public cloud will also likely increase dramatically, as the

economic incentives are too high to ignore, which will drive the market to seek expert IT resources that can work with clouds.

3.14.5 Potential Disruptions

Cloud computing is a disruptive technology. It changes the economic dynamics of how datacenters are built and operated, which impacts their total cost of ownership and drives server, storage, and network vendors to transition from the form factors and technologies they produce today to technologies that are more suited to the cloud. The challenge/opportunity we presented above regarding the layering of cloud infrastructure will drive new standards and present opportunities for new technology vendors to introduce new disruptive solutions to today's existing vendor solutions.

Disruption can also be felt on shrink-wrapped software products, as they will be challenged by offerings that can be hosted in the cloud as a service. Traditional applications offered as cloud-hosted services have not yet matched in quality and versatility the traditional shrink-wrapped versions, such as Microsoft Office. Cloud-hosted services also face the challenging requirement that network connectivity is robust and high in performance, which is still not the case everywhere around the world. There is an opportunity for such vendors to introduce cloud-enhanced shrink-wrapped applications, where when connecting to the cloud, the user gets significant feature enhancements that can leverage the cloud's power without compromising the power and sophistication of native applications on PCs and tablets.

Traditional Hosting

Traditional hosting services face major disruption if they do not transition to the cloud computing model of operation, as they will not be able to compete economically against the benefits of total cost of ownership resulting from using cloud computing technology in the datacenter. Furthermore, traditional datacenter architectures are changing, and with the traditional server, storage and network form factors and solutions are no longer suitable for datacenters that are built based on cloud computing technologies and managed through the cloud.

Time to Market

Cloud computing will invigorate the ability of entrepreneurs to create and deliver to the market at a much lower cost and in a much shorter time the solutions that today's vendors took years and huge investment to develop. This new phenomenon will shake up the market and create a new dynamic for competition.

3.14.6 Summary

By 2022, cloud computing will likely become more entrenched, and a significantly larger segment of computing workloads will be run on cloud computing infrastructures, whether public or private. This promising market faces many challenges and opportunities. Standardization and inter-vendor cooperation on breaking up the puzzle of building and managing cloud infrastructures is a major challenge, and successes here can drive this market toward expansion much more rapidly.

The real promise of cloud computing is the way that it changes the game for software development. Once IT administrators and developers have the ability to create true virtual datacenter infrastructure substrates, where resources are connected virtually across clouds and premises, and developers are

able to tap into APIs of services to mash up applications and middleware from different providers, there is the potential for experiencing a Cambrian-like explosion in the next generation of software. The sophistication of newly developed offerings that leverage already developed SaaS can potentially exceed our wildest imaginations.

3.15 The Internet of Things

3.15.1 State of the Art

Technology drivers for the Internet of Things (IoT) include sensor and actuator evolution and ubiquity, along with increased interconnectivity for such devices with each other and with compute and memory capacities. To understand the IoT, therefore, we must begin by understanding devices and device connectivity.

Electronic devices and sensors are becoming both increasingly cheap and common, and device miniaturization is ongoing relentlessly. An early driver for these technologies has been the military's need for cheap, ubiquitous sensing. "Smart dust" technology was funded by DARPA and the US military in the early 1990s, often based on MEMS devices, and smart fabrics or e-textiles date back to a similar period of time and sometimes called "wearable computers." In fact, these technologies have become sufficiently mature to result in annual meetings dedicated to discussing smart fabrics and their roadmaps, for entirely commercial applications, whether for fashion, sports, or medical purposes.

A second driver for ubiquitous sensing and computing has been industrial applications, for example, to track fleets of trucks on the road, perform detailed mapping of environments (recall Google's 2013 efforts to map sites like the Grand Canyon for Google Earth), engage in environmental sensing and monitoring that utilizes either special devices or the many sensors now integrated into and/or available for smartphones or small form-factor tablets, and locating and tracking products in warehouses, transit, and stores. In sports, there have been attempts to instrument balls and/or players to help improve goal-kicking accuracy, and numerous studies of swimmers' abilities use sensors built into sports devices or clothes. Recent research has explored ways to reduce re-stocking costs, using vision processing by wandering robots. Related work in datacenter systems attempts to track temperature profiles for improved cooling efficiency. More recent work aims to find less intrusive ways of monitoring or reacting to inputs, such as commands from gesture recognition. Games like Microsoft's Kinect and Sony's Wii have seen substantial market acceptance; underlining the upcoming importance of these technologies for broad sets of consumer electronics, Intel bought Omek Interactive, a gesture recognition company, in 2013.

Finally, we have all heard about consumer applications such as smart homes, which can monitor electricity consumption and adjust to current pricing, give owners remote access for monitoring their properties, etc. These are already deployed in European countries, along with smart grids, smart city facilities for security and monitoring, and many other such facilities. In fact, self-monitored oil and gas pipelines predated many of these technologies, already deployed in the late 1980s to help watch for pipeline failures.

3.15.2 Where We Think It Will Go

But where are sensors/actuators and devices going? And what is the future of the IoT? Certainly, by 2020, many heretofore manual business processes will have been automated, whether via active

devices built into products and supply chains or via external sensors such as cameras. It will be possible to dress in clothes that completely and thoroughly monitor the wearer's activities, which is evidently useful in sports and sports training, in the arts (e.g., instrumented dancers whose moves are amplified and displayed), and in medical settings.

3.15.3 Challenges and Opportunities

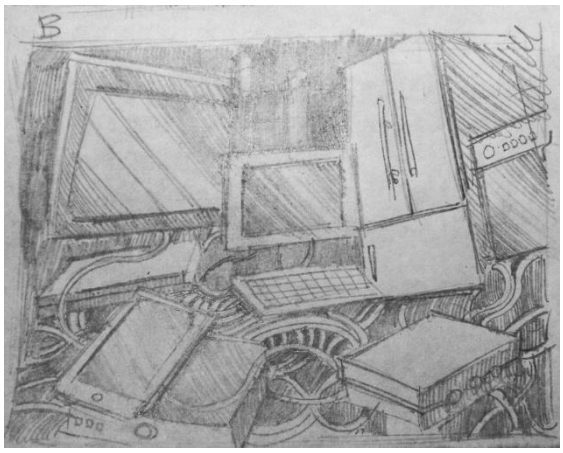
What about our everyday actions, however, and the privacy concerns raised by the IoT? Yes, it is convenient to walk into a coffee shop, have your wearable glasses recognize customers' faces and tell you that the person sitting by the window is "Dave Smith," someone you last met at some business meeting. This way, you won't be embarrassed by having forgotten his name and affiliation. However, do you really want the "cloud" to know where you are right now, with whom you are talking, and who else is around? Numerous commercial opportunities are raised by such monitoring, making it of interest to many businesses, but such monitoring is also rife for abuse, as with stalkers hounding someone or profilers using the data to update credit reports or inform potential loan agents. Yes, you can measure your gait and activity level, continuously, and such information can be invaluable to your doctor, but what if your health insurance rates rise because you did not move around enough last month? In other words, concerns with privacy and security will affect the IoT's growth and acceptance, particularly in lieu of differences in government actions and in the legal environments of the US versus Europe versus Asia.

Setting aside privacy concerns, amazing possibilities arise not just for future wearable fabrics but also for other materials, such as weight-bearing supports for bridges monitoring their own status and reporting it, and smart airplane wings adjusting their surface structure to current air flow and flight requirements. Further, the IoT can improve the lives of citizens and tourists alike, with cities guiding visitors to important sites, and the sites self-narrating their histories and importance; rather than going by guidebooks and what everyone "has to" see, tourists can re-live what their friends found exciting or re-trace the paths of an ancestor who used to live in that city. City dwellers can bypass crowds or avoid traffic jams, guided on the current best routes to their destinations, or, when using public transportation or walking, guided to meet friends along the way. In suburbia, car sharing may be automated, with self-directed cars arranging for meeting places, and for overland trips, tedious long hours of driving along country roads replaced by a self-driving car narrating when beacons are passed, without the need for constant hands-on driving. Even better, city traffic jams become more bearable (and more efficient) when cars move automatically when the light turns green and when cars move at some common, safe speed.

In nature settings, hikers encountering wildlife may result in, say, bears being tagged, virtually, to track their movements, via image and face recognition technologies. Long-term research on bear behavior and preservation of their natural habitat can use such data as well. Anglers may be directed to where the fish are, not only by their own sonar fish finders, but because all such finders are linked, via the IoT, with back-end processing computing likely fish concentrations and possibly enforcing fishing quotas. In fact, it is the global scope of the IoT—its knowledge about the many sensors deployed in many different settings—that presents entirely new opportunities for enriching or facilitating our daily lives, saving the planet, and helping research. In one demo, a phone company, simply by tracking its cellphones' current locations, was able to draw a precise map of London, including boats in the river, demonstrating progress in the mapping technologies enabled by the IoT. Or, in science, by understanding water

temperature and pollutants, along with fish concentrations, and by tracking the types of fish being caught, we can understand for, say, the entire state of Minnesota, how and where fish populations thrive or suffer. Such data can be used to identify, e.g., industrial, urban, or home pollution sources, not only in extent but also in their effects on the environment, informing enforcement or lawmakers. It can also help attract and maintain tourism.

3.15.4 Potential Disruptions



It should be apparent from the descriptions above that the IoT is more than just the plethora of distributed sensors and actuators embedded into home, urban, or natural settings, but that it is critically enabled by and dependent on the tremendous data collections and compute capacities in the back-end machines and datacenters that use such data to understand our world and improve it. Its growth and continued evolution, therefore, depends on our ability to deploy the computational and storage capacities needed to support the IoT. That, however, depends on the potential profits and gains obtained from building and paying for such infrastructure. It appears today that these profits and gains exist, but there are many potential disruptions. First, what entities will obtain those gains and will, therefore, be willing to make continued investments? Currently, large companies like Google, etc., are capitalizing on end user services, but are there other models? Certainly, yes, as evident from developing countries' use of "microservices" as in banking via cellphones versus using local and unreliable banks. But if such developments become more ubiquitous, the will to continue infrastructure investment and the large entities needed for doing so may be reduced. Second, as already evident from past experiences with peer-to-peer technology, there are legal and governmental issues, as well, which if not resolved, can substantially slow progress. A notable current case in point in the US is the legal case against someone wearing Google Glass while driving a car receiving a traffic fine in lieu of laws passed against texting while driving.

3.15.5 Summary

In summary, the IoT is here to stay, driven, among others, by device technology advances, the opportunities created by the billions of smartphones with their rich built-in sensors, Internet connectivity to fixed facilities, increased mobile connectivity, the new functionalities it enables, and business reasons, such as the desire to reduce cost through automation, reduced loss/wastage, and shorter durations for supply chains.

3.16 Natural User Interfaces

3.16.1 Introduction

Since the beginning of the computing age, the public has dreamed of computers that could interact in a natural way using speech, gestures, and intelligence, and interface with humans in the same natural ways as we communicate with one another. But what was once in the realm of popular science fiction culture is rapidly becoming a part of everyday life. Using techniques from touch and gesture to speech recognition lets users increasingly interact with computing devices just as naturally as they interact with each other. Such natural ways to interact with new technologies intend to make it easier to operate them and speed up their adoption.

Since the first appearance of the graphical user interface (GUI) almost 40 years ago, engineers have envisioned increasingly natural ways to interact with the systems they design and develop. But what has become known as the natural user interface (NUI) was in reality until recently only little more than an enhanced GUI. The capture of a limited range of human interactions such as speech, handwriting with pen, and simple touch was used as an alternative way to click a button or hit a key on a keyboard. Speech recognition took the form of command and control, and pen-based input was essentially about character recognition. With relatively poor dependability, these fragmented elements of a NUI did not experience wide adoption.

However, today's NUI is in the midst of a big transition. New display technologies turn any surface into an interactive screen. Megapixel cameras and microphones are embedded in every device, enabling seamless understanding of speech and gestures. The incumbent keyboard and mouse are giving way to gesture, touch, and the spoken language. The traditional desktop and laptop computing devices are being supplemented, if not directly displaced, by an aggregate of powerful connected devices giving a sense of ambient intelligence. [1]

3.16.2 State of the Art

Microsoft's Kinect [2] is revolutionizing more than interactive gaming. One of the fastest selling consumer devices of all times is being adapted for a wide range of applications outside the living room. At the heart of the Kinect experience is its ability to analyze and process images, gestures, and voice. These inputs can be used to create a NUI that can change the way users interact with computers. Example applications range from American Sign Language readers to shopping applications that analyze your body shape and select a pair of jeans that both fits and flatters.

Smartphones, tablets, and a new generation of laptops with multitouch screens are transforming how we interact with our surroundings. With recent developments in connectivity and cloud services, devices today have continuous access to unimaginable computing resources and mind-boggling amounts of data. The developer ecosystems evolving around these devices are empowered to flight their innovations with millions of real users while monitoring detailed usage in real time. This feedback loop creates a cycle of increasingly more refined NUI interactions. Smartphones and other mobile devices

have led to the creation of the largest workbench in human history, and it is churning out numerous applications that are exploring new ground in NUIs, whether that is real-time machine translation services that break down language barriers between people or applications that let your eyes control your phone without actually touching it.

There is another dynamic driving NUI development that is coming from a transformation of the limited interface of individual computing devices to the opportunities arising from turning our living rooms, workspaces, and vehicles into computing spaces. While many devices become increasingly smaller, these computing spaces are inherently physically larger than the user. Whether it is a self-driving car, a video conferencing room, or immersive video gaming, touch, gesture, and speech have been integrated to provide a more natural and efficient way of interacting with these complex, multisensory systems.

3.16.3 Challenges

But for all the progress and developments, the vast majority of our interactions with technology remain little changed. The software productivity tools at the core of our working lives take only limited advantage of a NUI. Most of our home appliances are still firmly rooted in the dark ages of VCR programming. Modern technologies are often crudely integrated and appear too often more unintelligent than we would expect in this day and age.

To successfully leverage NUI trends and opportunities, applications must perform one particular function exceptionally well: in the context of a specific task, they must enable the user to interact with the application as if the user were interacting with a capable human assistant. Accomplishing this goal involves a complete rethinking of the interaction between human and computer. A truly natural interface goes beyond the interface between the user and the system and focuses on what the interface actually enables and the processes required for this to happen. Applications that employ a NUI have to address these challenging “intelligent” interaction paradigms to be successful:

- **Predictive, anticipatory, and adaptive.** Use past and current user actions to assist with task completion or perhaps even automation. Predict user behavior and act in synergy with the user.
- **Contextual awareness.** Understand the user’s context similarly to what would be expected of a human assistant. Capture intent and emotion to better aid with task completion.
- **Multisensory input.** Ability to capture multimodal input including but not limited to speech, touch, and gestures; ability to respond to the user in the most appropriate way. Can utilize information about a user’s current and past environments, such as proximity to other physical devices and resources, geo-location, and movement, to assist with task completion.
- **Language and inference.** Understand natural language with ability to correctly infers users’ intentions and goals, and engage in a dialog to resolve ambiguity and simplify collaboration.
- **Augmented reality.** Create the most natural extension to the reality in which the user operates. The ability to capture the surrounding natural environment and create an augmented environment tailored to complete the task at hand.

By designing computing systems around these interaction paradigms, we redefine the relationship between users and their computing devices. We will no longer force our users into unintuitive and arcane interactions typically required by existing computers. We do not envision NUIs to be static; they will dynamically adapt to the user and fine-tune themselves as usage evolves. This NUI vision has the

potential of enabling applications to play a greater role in tasks such as driving, walking, and reading, where traditional interfaces would either be a distraction or even intolerable.

With a NUI, we envision a world where no device is an island, and ambient intelligence is achieved through natural interactions and integrated software architecture. Devices will work together and perhaps more importantly, they will both feed into and take advantage of cloud services to exhibit the human-like intelligence required for a truly natural interface. This mesh of devices and computing services will constantly collect a wide range of sensory information and combine this with contextual data via cloud services. In turn, cloud services will interpret aggregated datasets against patterns to anticipate tasks, activities, and events, and in this way, adapt its behavior to the particular devices used in a given situation.

The promise of a brave NUI world is the emergence of applications that will make possible a new model of interaction between human and computer. Our “intelligent” interaction patterns and continued new hardware innovations are a challenge, but they will surely enable a multitude of computing devices that know so much about us that they can increasingly work on our behalf.

3.16.4 Summary

After years of being the Next Big Thing on the technology horizon, NUIs are rapidly becoming mainstream. Interactions between human and machine become more natural and intuitive when people can use touch, gesture, and speech to interact with their computing devices.

Hardware prices are falling rapidly and capabilities rising at an even faster pace. These developments are making it easier to embed sensors, extreme processing power, and connectivity into devices and surroundings. The software that runs these technologies is the result of years of research into computer vision, machine learning, big data, user interfaces, and speech recognition and natural language processing.

The NUI is starting to make its way into the mainstream, but the work to make it real has been going on for years. We should expect innovation to continue, with the emergence of entirely new kinds of computing form factors combined with a wide range of significant hardware and software technological breakthroughs leading to far more radical types of NUI. This an amazing opportunity for both researchers in academia and for the technology industry to create even more exciting products with the NUI at their core.

3.16.5 References

[1] E.H.L. Aarts and J.L. Encarnação, *True Visions: The Emergence of Ambient Intelligence*, Springer, 2006.

[2] G. Goth, “Brave NUI World,” *Comm. ACM*, vol. 54, no. 12, 2011.

3.17 3D Printing

3.17.1 Introduction

3D printing promises a revolution in fabrication. In today's manufacturing, products are usually assembled from components that are separately created using specialized machinery. The scale required to make such specialized manufacturing cost effective often requires a network of far-flung suppliers and complex supply chains. With 3D printers capable of handling multiple materials, it may become possible to fabricate many such items entirely in one place, close to the consumer. 3D printers can already create many shapes using combinations of materials that would be very difficult to create with conventional machining methods. Moreover, they can handle products from a few inches to many feet in size, and materials ranging from plastic to metal to edible foodstuffs to stem cells for creating living tissue. The possibilities of what can be made with 3D printers are endless.

3.17.2 State of the Art

Also known as additive manufacturing, the basic 3D printing technology was invented and commercialized several decades ago. Carl Deckard and Joseph Beaman invented a selective laser sintering printer at the University of Texas, Austin, in 1986. That same year, Charles Hull received a patent on stereolithography, a method for building up a solid object by depositing successive thin layers of a liquid polymer that could be cured (solidified) by exposure to ultraviolet light. These basic techniques have been refined and form the basis of many commercial-grade 3D printers.

3D printing is currently used in a wide variety of small-scale and custom fabrication jobs. For example, in the movie *The Hobbit: An Unexpected Journey*, most of the animatronics for goblin eyeballs, facial muscles, lips, and tongues were 3D printed. Such props were traditionally made by hand, requiring weeks of work by a skilled artisan, whereas they can now be prototyped, refined, and produced in days. Hobbyists and artists fabricate small items in personal printers or send them to 3D printing services for items requiring printers that can handle multiple materials or larger scale. Dental labs produce custom dental crowns, bridges, and orthodontic appliances in hours using digital oral scanning, specialized CAD/CAM software, and 3D printers. Large manufacturers, such as aircraft and automobile companies, use 3D printing for rapidly prototyping parts and for producing specialized production parts such as jigs for aircraft assembly.

Inexpensive 3D printers typically use plastic as the building material, but there are experimental modifications to print with food pastes such as cookie dough and frosting to produce elaborate confections. Industrial 3D printers can handle a variety of materials, including ceramics and metals such as bronze, steel, tungsten, and titanium. Technologies include deposition-based methods, where material is deposited in paper-thin layers to build up the desired shape and methods where lasers melt and fuse powdered metals or polymers. These printers can produce shapes and forms that are difficult or impossible to create using conventional techniques. For example, it is possible to create a mesh

consisting of seamless interlocking rings, which is difficult using conventional processes. Precisely shaped internal voids can be created by filling the space with material that can later be removed, such as a gel or unfused powder. Multi-material objects can be formed by filling voids in a strong structure with a different material later, yielding the possibility of strong yet light composite structures.

3.17.3 Challenges

A major challenge in using 3D printing to its full potential is that the available software and algorithms for driving the hardware are still nascent and limited. Input to 3D printers is typically in the form of an STL (Standard Tesselation Language) file generated by a CAD package. While several CAD packages are designed for 3D printing, generating a correct, printable object description from a concept in a user's mind is a fairly complex process. Solid modeling software often uses metaphors adapted from machine shop operations, such as drilling and milling; however, 3D printers can generate many forms that are difficult to produce through machining. Other techniques such as surface modeling borrow from video game and animation design, but these methods are primarily intended for modeling the surface of the solid, not the interior. Even surface modeling requires consideration of surface properties such as color, texture, reflectivity, and hardness. Solid forms have many aspects that are difficult to specify and optimize using current software, including strength requirements, rigidity, weight, and center of gravity. In addition, for industrial use, the design must consider other aspects such as cost, ease of manufacturing, and workflow if multiple steps are required. Overall, there are significant challenges in designing software that will make it easy for users to visualize, refine, and realize forms from their imagination into an input for a 3D printer.

Besides the technological issues, 3D printing brings with it several social, legal, and ecological challenges. Given appropriate design files, it is possible to print weapons and contraband items such as counterfeit goods and paraphernalia for manufacturing illegal drugs, making it difficult to regulate such items. With improvements in methods for deriving designs from 3D scanning, it will become easier to reproduce proprietary designs, thus evading intellectual property regulations. Parts made from uncertified designs, such as replacement parts for automobiles, may be dangerous in use. Taxation of product sales will become harder if users can simply purchase designs that they can print themselves. Inexpensive and easily printed products made from non-recyclable and non-biodegradable materials may lead to more pollution and other ecological issues. New legal regulations may be required to address these issues.

3.17.4 Where We Think It Will Go

As 3D printer hardware and software improve and become cheaper and more widely available, we expect that more products will be customized to specific use cases. Custom prosthetics and even replacements for body parts may be created with 3D printers—initial research already shows that cartilage can be formed using 3D printed molds. Combining different materials will allow the creation of composite materials with new properties, such as the ability to heal after failure. The ability to print batteries and sensors directly onto objects will enable more "smart" mechanisms that can sense changes in surrounding temperature and light levels, as well as impending failure. By adding articulated joints and electrical connections, it will become possible to print complete, fully functioning devices, both electronic and mechanical, rather than assembling them from parts produced separately.

Manufacture of many products will move from large, centralized factories to local workshops and even the user's home.

3.17.5 Disruptions

3D printing may be highly disruptive because it could make many jobs obsolete. For example, if entire mechanisms can be created directly by 3D printing, then this may eliminate many assembly jobs. With local manufacturing of goods, there may be less freight to transport. If consumers can purchase or otherwise obtain designs for items over the Internet and print them either in their homes or in local print shops, then there may be less need for retail personnel. On the other hand, it may also create design jobs and a need for teachers and equipment to help train designers.

3.17.6 Summary

3D printing, also known as additive manufacturing, is a technology with enormous potential that will allow us to produce objects with designs that in the past would have been prohibitively expensive or impossible to manufacture. As the printing hardware and design software improve, we expect that a wide variety of products will be manufactured mostly or even entirely using 3D printers in a manufacturing plant, at local printing services, or in the consumer's home. These changes may be quite disruptive because the increased automation may reduce jobs in manufacturing, assembly, freight transportation, and retailing. Changes will be required in education to train a new generation of designers, as well as in laws to manage new issues in intellectual property, taxation, and certification of product safety and effectiveness.

3.17.7 References

1. H. Lipson and M. Kurman, *Fabricated: The New World of 3D Printing*, Wiley, 2013.
2. S. Bradshaw, A. Bowyer, and P. Haufe, "The Intellectual Property Implications of Low-Cost 3D Printing," *SCRIPTed*, vol. 7, no. 1, Apr. 2010.
3. Saracco, R., Personal Communication, 2014.

3.18 Big Data and Analytics

3.18.1 Executive Summary

More data is collected, shared, and analyzed every day. The growing availability of data and demand for its insights hold great potential to improve many data-driven decisions, from the mundane to the strategic. But this growth also poses significant challenges, both technological and societal. To harness the deluge of data to beneficial use, we will need to address rapid changes in data acquisition, storage, and processing technologies; education, both of the analytics workforce and everyday users; and complex privacy issues.

3.18.2 Introduction

Imagine, if you will, the following scenario. You sit down at a restaurant for lunch, wondering what to order. You take a picture of the menu with your phone, starting a whirlwind of activity on your behalf. The software combs through the data it has collected during the day about you, such as your breakfast, exercise and calorie expenditure, blood pressure and blood sugar levels, etc. It combines this data with long-term information such as previous user reviews of the different dishes in this restaurant, your weight loss goals, food preference, and sensitiveness, and perhaps even your individual genetic properties. In milliseconds, it makes its top three suggestions, from which you choose your meal. In the meantime, data is collected anonymously in the aggregate about your choices and decisions, as well as other patrons'. The restaurant manager can use it to adjust the menu. Researchers can use it to better understand the relationship between nutrition, fitness, and health. Your friends can use it to obtain personalized food recommendations, and you can use it to track your progress toward your health goals.

This scenario, already more science than fiction, exemplifies how so-called "big data" can be used to seamlessly affect decisions from the prosaic (your choice of lunch) to the strategic (the FDA's nutrition guidelines). It represents but one of many opportunities envisioned for the large-scale analytics of diverse data. Big data is finally transitioning from the computer science and machine learning classrooms into numerous real-world scenarios in business, government and military, science, politics, medicine, climate, and personal analytics—a trend that we expect to grow rapidly through 2022.

3.18.3 State of the Art

Big data is exploding, with no signs of slowing down. The growth is manifest on two separate axes: more data is collected, and more data is shared. Growth along both axes is exponential, and the combined growth results in a very rapid increase in total available data indeed. IDC estimates that the amount of data created and shared on the Internet will reach around 8 zettabytes by 2015[2]. Let's look at a few examples:

[2] IDC report "Extracting Value from Chaos," June 2011

- Photos and videos taken and shared are growing at an exponential rate,[3] a result of three multiplicative trends. One, people are taking more photos, because more cameras are ubiquitously available through the proliferation of smart and feature phones. Two, these photos increasingly contain more data (pixels) through the rapid growth in sensor technology. And three, more people share their photos and videos on Facebook, YouTube, Twitter, Snapchat, and other fast-growth social networks.
- Crowd-sourced and individuals' data from day-to-day life is proliferating through a variety of mobile applications. Such information includes service reviews, traffic and geo-tagged check-ins, health and exercise metrics from wearable devices, etc. (see Section 3.2).
- More studies, more instrumentation, more simulations, and more sharing facilitated by the Internet (e.g., CERN's Grid) is translating into more scientific data.
- The increase in resolution of freely available elevation data has led to a lot more mapping.

3.18.4 Where We Think It Will Go

We see tremendous opportunities in big data. In the sciences, for example, the growth in experimental data and in simulations—the fourth paradigm of science—has already advanced our understanding of the universe and of life. The growth in ubiquity of mobile computing devices, as well as in the applications that collect data, means that much more data is available about a lot more people (and possibly *to* many more people). This data is often used in quotidian decisions such as picking a driving route. In business, much more data is collected on every aspect of operation, increasing efficiency, customer marketing, and pivoting to new markets[4]. In medicine, big data combined with rapid advances medical science can bring us to a point where all major health decisions are tailored to an individual's situation[5].

If big data fulfills its promise, we think it will have tremendous impact in reducing uncertainty around large domains of decisions, both before they're made and, retrospectively, afterward, too.

3.18.5 Technological Challenges

The collection, organization, validation, interpretation, and management of large datasets present multiple technical and technological challenges. As the amount of data grows rapidly, additional computational resources are required to process the data in a timely manner. This time and resource pressure is increased for ongoing analysis by a recurring deadline (such as daily business metrics) and even more so for interactive and exploratory data exploration.

Accordingly, computational resources dedicated to big data are growing explosively. The demand for data storage and processing can, however, grow faster than the underlying technologies. Take the recent growth of informatics-based scientific disciplines, for example. In genomics, advances in gene sequencer technology have brought down the cost and delay of sequencing to the point where many labs around the world can produce copious genetic data. Worldwide, we can now produce around 15

---

3 See Kleiner Perkin's "Internet Trends 2013" report at http://www.slideshare.net/kleinerperkins/kpcb-internet-trends-2013

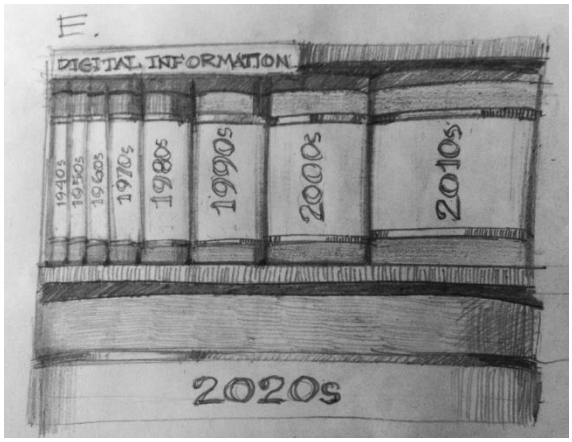
4 http://www.wired.com/insights/2013/07/putting-a-dollar-value-on-big-data-insights/

5 http://spectrum.ieee.org/tech-talk/computing/software/predictive-analytics-and-deciding-who-should-receive-organ-transplants

Pbytes of compressed genetic data per year, which is growing at a rate of 3x to 5x a year[6]. In high-energy physics, the Large Hadron Collider and other instruments at CERN alone produce a similar amount of data annually[7].

Pervasive big data tools such as Hadoop are already prevalent in the analysis of these massive genomic databases. But despite the carefully designed scalability of the software tools, they are still limited by hardware constraints, such as power, acquisition, and operation costs; capacity growth in processors, hard disks, and networks; and increased complexity in management and cooling. For example, CERN is planning for its Wigner datacenter to double its processing and storage capability in the next three years. Although impressive, this rate is a far cry from the growth rate of the data to be processed, creating an increasing gap between the amount of data to process and the hardware to process it. If cost weren't an issue in scaling the hardware, power still remains a stubborn constraint, limiting practical datacenter size to several MW. And even if that constraint were to be removed by advances in power efficiency, the speed of light effectively limits the scale of a datacenter to the tolerable limits of latency in data fetching before computation grinds to a halt.

3.18.6 Potential Technological Disruptions



To fully exploit the opportunity promised by big data, we must find ways to bridge the gap between data growth and processing capability.

On the hardware side, it is a simple matter to extrapolate current growth trends to predict increased storage density; continued processor growth along Moore's law; and better power efficiency and networks. But the trends set in the past two decades, even if exponential in growth, are not disruptive enough to close this gap. It would take radical technological shifts to match resource growth with data growth, or we could experience a significant decline in the current rate of data growth and, with it, the predictive capabilities of its analysis.

Barring an unpredictable disruptive technology, a more feasible path to closing the gap is innovation in software. Big data software can be considered still in its infancy, with plenty of opportunities for growth. Areas of possible improvement include

- reducing the amount data to be processed: better compression; early detection of irrelevant data; and more effective sampling techniques;
- algorithmic effort/processing reduction: more efficient machine learning algorithms to produce predictions in less time, etc.;
- improving systems effort: better utilization and sharing of available hardware resources; and

[6] IEEE Spectrum 07-2013 "The DNA Data Deluge" http://spectrum.ieee.org/biomedical/devices/the-dna-data-deluge

[7] http://home.web.cern.ch/about/computing

- generating insightful analysis: something that produces much higher-level analytics and answers than is standard today.

Such innovations will not only reduce the requirements of labor and expertise from analysts but may in fact drive efficiency through more parsimonious representations of data.

3.18.7 Societal Challenges

Big data hardware and software are no panacea. In fact, they are currently useless without specialized human skills. These skills range from the selection, preparation, and cleaning of data; the exploration of different analyses on the data; the application of error checking and strong statistical reasoning to reduce bias and type I/II errors; and finally, the interpretation, visualization (for the higher-bandwidth visual sensory), and domain application of the results. Although we believe that many important parts of these processes can and will undergo increased automation, we do not foresee an elimination of the skilled human element. If anything, big data analytics falls in line with the workforce migration we observed in the past century from labor- to knowledge-intensive industries.

In a recent study, 88 percent of companies surveyed have already reported a talent shortage to successfully execute on big data initiatives[8]. Because big data is growing at such a rapid rate, along with the demand for data scientists and analysts, and because the skills required encompass a wide range of advanced computing, statistics, communication, and domain expertise, we may potentially face a critical shortfall in this workforce[9]. This challenge needs to be met by a correspondingly large challenge in workforce education, both in academia and industry.

Other aspects of the big data shift will require societal response. Perhaps the biggest one is the concern about eroding privacy and data leaks, with a potential for very significant personal, business, or military damage. The concentration of big data in the hands of governments also evokes concerns about the risk to democracy and civil rights. The challenge is then to find ways to collect, share, and benefit from big data technologies while still preserving the privacy, trust, and rights of the individuals whose data is collected.

3.18.8 Potential Disruptions

Academia is already mobilizing to develop new programs around big data and to train thousands of new data scientists and analysts[10]. Perhaps a more radical approach in workforce education is required to meet the rapid demand. Interestingly, one potential disruption in the training of the workforce in general, and big data in particular, also comes from a new scale-out field: MOOCs (refer to Section 3.4). Distributed online education, with its various levels of certification and cost, is already training many thousands of individuals in big data-related fields and showing a strong growth trend[11].

One of the characteristics of MOOCs is that successful training no longer requires physical school attendance and is therefore independent of geography. This is just one aspect that may require

[8] http://thehiringsite.careerbuilder.com/2013/07/16/careerbuilder-big-data-study/

[9] http://spectrum.ieee.org/podcast/at-work/tech-careers/is-data-science-your-next-career

[10] http://www.wired.com/insights/2013/07/the-growing-need-for-big-data-workers-meeting-the-challenge-with-training/

[11] http://gigaom.com/2012/10/14/why-becoming-a-data-scientist-might-be-easier-than-you-think/

employers too to radically adjust to the changing landscape of big data professionals, even if universities and MOOC train an adequate number of them. We may therefore experience a disruption from the traditional employment model of accredited employees sharing an office space. Instead, we may see the increasing demand for these professionals met by companies who successfully adapt to a distributed workforce of varying formal education.

Finally, we may find that the explosive success of big data may hinge on a significant disruption in the field of data security and privacy. There is certainly a technological challenge and opportunity here, to come up with provable standards of privacy and security. But there is also one that may require legal, normative, and educational changes to place acceptable limits on the use of big data.

3.18.9 References

G. Brumfiel, "High-Energy Physics: Down the Petabyte Highway", *Nature* vol. 469, 19 Jan. 2011, pp. 282–83.

P. Webster, "Supercomputing the Climate: NASA's Big Data Mission". *CSC World*, Computer Sciences Corporation, 2012.

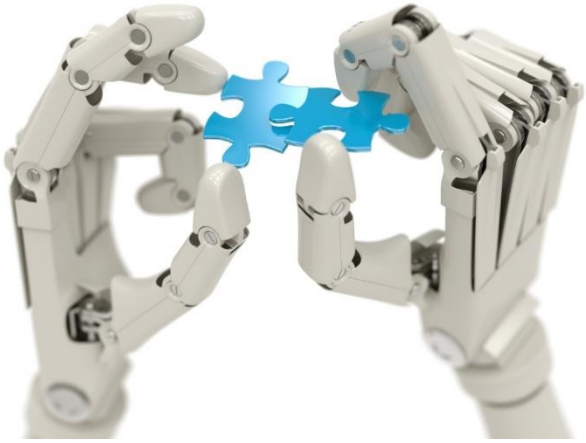
S. Shah, A. Horne, and J. Capella, "Good Data Won't Guarantee Good Decisions," *Harvard Business Review*, Apr. 2012.

"Data, Data Everywhere", *The Economist*, 25 Feb. 2010.

O.J. Reichman, M.B. Jones, and M.P. Schildhauer, "Challenges and Opportunities of Open Data in Ecology," *Science*, vol. 331, no. 6018, 2011, pp. 703–705; doi:10.1126/science.1197962.

S. Cherry, "Is Data Science Your Next Career?", *IEEE Spectrum*. 28 May 2013.

3.19 Machine Learning and Intelligent Systems



3.19.1 Introduction

In the past decade, we have witnessed a dramatic increase in the use of machine learning (ML). The application of ML plays an increasingly important role in our daily lives, whether it is ranking search results for Google and Bing or recommending products on Amazon.com and movies on Netflix. Innovative ML techniques have led to emerging businesses that are able to identify influential users on Twitter and capture product sentiments for marketers.

ML is the discipline of artificial intelligence aimed at creating computing systems that can learn from data. For example, a classical ML system can be trained on email messages to learn to distinguish between spam and ham. Optical character recognition is another example in which printed characters are recognized automatically based on previous examples.

The simplified ML process consists of (i) training, (ii) test, and (iii) prediction. A dataset is partitioned into the training set and the test set, where the training set trains the system, and the test set is kept secret from it. The output of the training process is a learned model that will be used for prediction; we use the test set to test the learned model's accuracy. Because of a wide range of choice of training parameters as well as training algorithms, the ML process often becomes an iterative one, with the final model selection resulting from an empirical process.

There are different ways of training in ML:

- *Supervised learning*. In this kind of learning, the algorithm is given training data that consist of examples with both the input data and the desired output, also known as labels. The learner should be able to generalize from the presented previously unseen data.
- *Unsupervised learning*. Here, the algorithm is presented with examples from the input data only and will fit a model to these observations without prior human knowledge.
- *Reinforcement learning*. This algorithm learns how to respond given an observation of the environment. Every action has an impact on the environment, which provides feedback in the form of positive or negative rewards.

3.19.2 State of the Art

With the dramatic increase in processing power and storage capacity, the field of ML has changed dramatically in recent years. This change can be attributed to a deeper understanding of ML algorithms, inventions such as multicore processors, distributed computing, and new storage technology, as well as to an explosion of available data from an increasingly connected world—the so-called big data phenomenon. The aggregate of these changes has resulted in copious new developments in ML that are fundamentally characterized by large scale.

The ML community is vibrant and diverse. While numerous ML techniques have made significant strides recently—too many to be covered here—one stands out as a representative of the power of both looking to understanding how the human brain works and utilizing technological advancements, namely,

deep learning (DL) [1]. DL algorithms and systems have enjoyed remarkable success in the speech, language, and image-processing fields over the past few years. These new algorithms swiftly beat current approaches to image analysis, acoustic modeling, and natural language understanding. Several factors have contributed to these achievements, starting with building artificial neural networks that mimic the behavior of the human brain. Much like the brain, these multilayered networks can capture information and respond to it. Arguably, these network architectures build up an understanding of what image objects look, or phonemes sound, like. Across a wide range of application domains, the abilities of DL to learn from unlabeled data have been broadly useful and have led to significant advances. In acoustic modeling, the ability of DL architectures to separate multiple factors of variation in the input, such as speaker-dependent effects on speech acoustics, has led to extraordinary improvements in speech recognition, as popularized by Apple's Siri service.

3.19.3 Challenges

Most essential algorithms around us operate in near-linear or better time. We often think of these algorithms as meeting the high bar of unlimited scalability. Unfortunately, the ML field often works with super-linear algorithms in both time and space. Many ML algorithms display quadratic or worse behavior and are inherently tailored to operate in a single-address space. While learning algorithms that operate in linear time that allow us to train on very large datasets would be preferable, in practice, the growing volume of data often precludes the application of standard single-machine training algorithms. However, DL has successfully been explored in scaled-up environments involving clusters of GPU and very large amounts of RAM as well as scaled-out environments of thousands of networked commodity servers. Initial results have been promising, and further progress should be expected in this area.

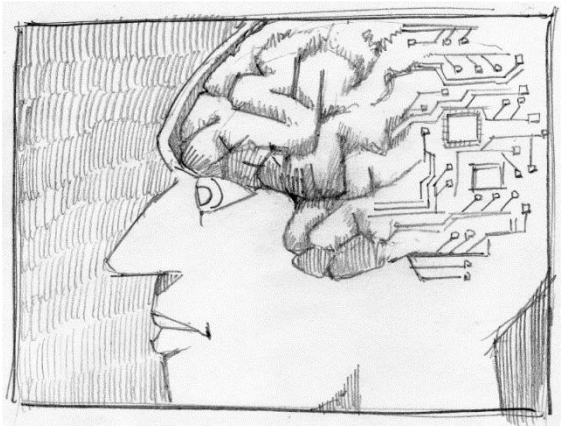
Most ML algorithms come with levers and knobs for tuning the learning process to the data at hand. Questions often arise about which configuration to use for a particular dataset and learning task. There are also challenges with collecting, cleaning, and preparing large datasets for this process. New tools are needed to help people specify what they want to learn and determine how to measure the accuracy of the predictions made by the learned models.

Finally, there is the societal challenge of how to guide actions and public policies in a world increasingly based on large-scale predictions made by computing systems that no single human being can fathom the scope of. Will people trust predictions from these systems?

3.19.4 Where We Think It Will Go

For ML, the best is yet to come. Improvements in ML, including new DL architectures and optimization strategies, are being explored broadly in the ML field with applications normally exclusively reserved for humans including facial recognition, image object recognition (tagging), language analysis, conversation, and translation. With computers' ability to process and store vast amounts of information at extremely high speeds, we must expect that ML-based computing systems in some cases soon will exceed human capabilities. Combining ML-based systems in order to create ensembles that exhibit human-like intelligent behavior in the aforementioned domains has the potential to enact an even greater change to human society than was already experienced by the computing revolution of the last 50 years. We are facing a unique opportunity to build systems that really become empowering agents that fundamentally understand our intent and continue to work on our behalf to complement us in our daily life.

3.19.5 Potential Disruptions



The NIH Brain Research through Advancing Innovative Neurotechnologies (BRAIN) [2] Initiative is part of a new Presidential focus aimed at revolutionizing our understanding of the human brain. The objective of this initiative is to produce a new dynamic picture of the brain that shows how individual cells and complex neural aggregates interact in both time and space. The results from BRAIN will fill major gaps in our current knowledge and create opportunities for exploring exactly how the brain enables the human body to capture, process, analyze, store, and retrieve vast quantities of information. These insights may lead to new hardware and software learning architectures that have the potential to revolutionize the ML field. The Human Brain Flagship project of the European Commission has similar goals [3].

We already mentioned that most ML algorithms require tradeoffs to execute in reasonable time and space. We suspect that computers as we know them today are not truly optimized for this class of problems. This is where quantum computing comes into play. By mixing quantum computing—which is extremely well suited at finding global minima in multidimensional spaces—with traditional computing systems, we may experience a true revolution in the capabilities of ML. Quantum ML may enable the most creative problem-solving process allowed by the laws of nature.

3.19.6 Summary

ML is about building better models of the environment in order to make predictions that are more accurate. If we want to cure diseases, we need better models of how they evolve. If we want to combat climate change, we need better models of what is happening to our global climate.

The directions and goals of ML fields are bold. They span explorations of the basic science of ML to understanding how to best solve practical problems and perform specific predictions. The development of more efficient and powerful tools to support the engineering practices of ML are strongly needed. Tools and methods that let nonexperts do a great job with their own predictive modeling are needed to truly empower users with machines that learn.

3.19.7 References

[1] A. Ng and J. Dean, “Building High-level Features Using Large Scale Unsupervised Learning,” 2012

[2] The NIH BRAIN Initiative, http://www.nih.gov/science/brain, 2013.

[3] The human Brain Project, https://www.humanbrainproject.eu/, 2014.

3.20 Life Sciences

3.20.1 Introduction

The life sciences (LS) industry uses modern biological techniques and supporting technologies with a goal to improve human and animal health; address threats to the environment; improve crop production; contain emerging and existing diseases; and improve currently used manufacturing technologies. LS industry sectors include pharmaceuticals, biotechnology, chemicals, medical devices, medical products and technology, and healthcare services. LS industry employment has significant size and growth worldwide. AAAS reports job ads today are roughly spread evenly across Europe, Asia, and US. In the US there were 1.61 million jobs in 2010, spanning over 70,000 individual companies [Battelle/BiO]. US healthcare spending grew 3.9 percent in 2011, reaching $2.7 trillion or $8,680 per person [CMS].

Currently, LS is experiencing five mega-trends: increasing expectations and cost of healthcare, quality of life of aging populations, major challenges in biology/medicine, and compliance challenges. These trends have increased LS rate of growth, which exceeds most major sectors of the economy in developed countries. This growth has resulted in a fierce economic competition to motivate industry (and government) to look for competitive advantages derived from a world-class academic research establishmentS. These stakeholders are increasing their demand for and expectation of “intelligence” in devices and systems, e.g., ubiquitous computing, communications, sensing, etc. [Khargonekar]

LS disciplines include bioengineering, biomedical engineering, healthcare technology, communications, and computational technical domains such as big data analytics, dependable and secure computing, high-performance computing, information technology, knowledge and data engineering, machine learning, multimedia, parallel and distributed processing systems, pattern analysis, security and privacy, software engineering, and visualization and computer graphics.

3.20.2 State of the Art

There is a shift occurring in the character of LS. The pharmaceutical, medical device, and biotechnology sectors have led the way, while a persistent focus on discovery, delivery, and continuous innovation remains a driver for growth. The subsectors of engineering have started to bring efficiency, effectiveness, and modern processes to advance the theories and research that bring LS to important and practical applications. [Thakor]

For example, the grand challenge of affordable and effective healthcare has spurred Singapore to help physicians provide more complete diagnoses; conduct regular treatments; follow evidence-based practice; and conform to workflow treatments by using enhanced, accessible big patent/healthcare data and analytics. Singapore has also started a national medical records database that has one record per citizen for that person’s lifespan. This big data approach will expand to physicians and their healthcare centers, and this new environment will help shape how professional cultures can work better together. [Ying-l,]

Convergence as the Fourth Revolution

The convergence of the life sciences, physical sciences, and engineering in advancing healthcare has four phases or revolutions:

- 1st golden age of biochemistry, 1900-1950 [Radda]
- 2nd revolution: molecular and cellular biology with Watson and Crick, 1950-2000
- 3rd revolution: genomics, 2000-present
- 4th revolution: integration of LS at the molecular level with engineering, physical sciences, and mathematics/computational science. This will increase understanding of how components collaborate to create complex biological systems and promote the flow of results into practice. [Sharp]

The converging, synergistic power of the biochemical and digital revolutions enables us to read every letter of life's code, create precisely targeted drugs, and tailor their use to individual patients. Cancer, diabetes, Alzheimer's, and countless other killers could be vanquished—if we make full use of the tools of modern drug design and allow doctors the use of modern data gathering and analytical tools when prescribing drugs to their patients. [Huber]

LS is not just the interface between the disciplines of engineering and biomedical sciences but also the convergent overlaps among bio-, nano-, and info-technologies. These interfaces are very exciting and fertile zones for highly original ideas, experiments, and discoveries. LS requires quicker translation—to bring discoveries into useful applications that help our patients, restore or assist human function, and address major needs in our society. [Chuan]

The public is another participant in this revolution as people are downloading and running fold-it-to-solve puzzles for LS research.

Framework standards are important. Classifying and modeling LS information is daunting, and using a multiscale modeling approach could support the goal of building a complete virtual physiological human. Similar to the computer stack, the LS stack starts at the nanometer scale and builds up to one-meter scale: quantum mechanics ➔ molecular ➔ network proteins ➔ cell types ➔ organ tissue ➔ organ ➔ organ system, e.g., torso ➔ organism. [Hunter]

Teamwork

There is now a widespread recognition of the critical importance of multidisciplinary team research in government, industry, and academy. Real-world problems do not come in disciplinary-shaped boxes [Jeffrey]. Large-space opportunities, such as LS, requiring interdisciplinary work have more risks than specific projects. [Jeffrey]

Barriers to communication between disciplines as they have naturally grown into a "stovepipe" reinforce the communication problem [Dewulf]. Cross-disciplinary project issues [Khargonekar] include differences in terminology and methods, setting priorities, effort needed to gain real understanding of the key technical and nontechnical issues, promotion and tenure, professional recognition, publications in discipline-based journals, intellectual property negotiations, dealing with government regulations, and potential loss of proprietary information.

An interdisciplinary shift in demand for talent within the biotech industry is moving away from hiring narrowly focused specialists to individuals with interdisciplinary academic training, highlighting the latest LS workforce trends. Hiring managers and industry leaders are starting to profile their workforce-related capability needs, which include soft skills and the ability to work effectively across disciplines. There is a clear shift in the industry's demand for talent away from senior scientist positions that tend to be more highly specialized and narrowly focused to a talent pool consisting of individuals who have interdisciplinary academic training and the ability to work broadly across multiple areas and in project teams where not everyone has to be an expert in everything. [Nugent]

Ultimately, scientists and engineers must learn how to work in teams. An outstanding teamwork example is the 6,500 technologists at CERN working on one problem (Higgs), a physics grand challenge. [Radda]

To encourage graduates with multidisciplinary experience, universities need to 1) promote cross-disciplinary interactions among their students, e.g., educational, sports, arts, and social facilities and dormitories; 2) develop programs specifically focused on the interfaces of key disciplines; and 3) encourage them to collaborate together in international research "collaboratories" working on interdisciplinary research projects. [Chuan]

The US National Science Foundation is funding new models for graduate education and training in an environment fostering collaborative research that transcends the traditional disciplinary boundaries and facilitates diversity in student participation and preparation. One of these NSF programs is Integrative Graduate Education and Research Traineeship (IGERT) [Ramasubramanian]

3.20.3 Challenges

In 2008, the NSF identified 14 engineering grand challenges, and in 2013, the US National Academy of Engineering revisited them. In both studies, four LS-related grand challenges remain:

- advance healthcare informatics,
- engineer better medicines,
- manage the nitrogen cycle, and
- reverse engineer the brain.

Some proposals transforming health and wellness include genomics-enabled personalized medicine, which would replace the creation of generic proprietary medicines. [Kun] In a related case study, patients with one type of leukemia received a one-time experimental therapy several years ago and some remain cancer-free today. At least six research groups have treated more than 120 patients with many types of blood and bone marrow cancers, with stunning results. [Marchione]

For quick transfer of medical device development to the patient, proposals have been drafted to use modeling (virtual prototyping) as a tool for regulatory approval. [Schiestl] Cardiovascular diseases are the major cause of death, and the cardiovascular health informatics used in wearable medical devices technologies and unobtrusive measurements connect through a body sensor network. Requirements

attributes include advances in miniaturization, and intelligent, network, digital, and standardization. [Zhang]

Healthcare once consisted solely of killing germs, but tomorrow's regimens will be guided and adjusted using relevant biomarkers specific to individual patients. 21st-century medicine is hampered by a regulatory regime built for the science of the 20th century. The search for cancer's silver bullet, something that meets the FDA gauntlet, is still going on, but there has been limited success in reducing per capita deaths from cancer since 1950 [Bashir]. Furthermore, new medical devices are going to Europe for regulatory approval because it takes half the time of obtaining FDA approval. [PWC] The FDA still operates according to the requirements of the age of mass drugs and must be reformed. [Huber]

3.20.4 Where We Think It Will Go

As the pace of intertwined discovery and invention increases, we are on parallel paths of evolution and inspiration, where our computer scientists can both learn and provide insights. Automation in all walks of life will be the most disruptive technology in coming decades. For healthcare, this means

- instant, expert diagnostic advice;
- personal preventative health advice;
- enhanced bedside care; and
- big data analysis of clinical trials and unstructured research data.

Big data in health and medicine will pull together databases with patients' outcomes, leading to a translation of research results directly into medical practice without delay. [Wah]

The role of scholarly Societies is to provide guidance not only on technical feasibility but on social and psychological impact. Our challenge is to optimize deployment of willingly tolerated, naturally intelligent computers for healthcare and clinical research. [Finkel]

3.20.5 Potential Disruptions

A *New Biology for the 21st Century* report from the National Academy of Sciences (NAS), National Academy of Engineering (NAE), Institute of Medicine (IOM), and National Research Council (NRC) announces biology is at an inflection point, poised to help solve major societal problems related to food, environment, energy, and health using a cross-discipline integration of LS research by physical, computational, Earth scientists, and engineers. [Sharp et al]

Despite the potential of recent advances, there is still much to be done to move from identifying parts to defining complex biological systems. Furthermore, the systems design, manipulation, and prediction needed for practical applications such as ecosystem repair or individualized medicine are still well beyond current capabilities. The "new biology" will provide a framework to connect biological research with advances in other branches of science and engineering. [Kamm].

3.20.6 Summary

LS industry is experiencing a large growth in the 21st century, surpassing most other sectors. Most of the growth is in addressing new needs with new solutions. These solutions were created with the help of

new computational technologies and the technologists who are comfortable and effective in cross-disciplinary teams. There future team members will need cross-discipline education and training.

3.20.7 References

[Bashir] Bashir, Rashid, "Cell-based Systems, Bio-fabrication, and Cellular Machines," IEEE Life Sciences Grand Challenges Conference (LSGCC 2012), 2012.

[Battelle/BiO] Battelle/BiO, Battelle and Biotechnology industry Organization, State Bioscience Industry Development, 2012.

[Chuan] Chuan, Tan Chorh, "Life Sciences at the National University of Singapore," IEEE Life Sciences Grand Challenges Conference (LSGCC 2013), 2013.

[CMS] CMS, "National Health Expenditures Data, Historical," 2012.

[Dewulf] Dewulf, A., "A framing approach to cross-disciplinary research collaboration: experiences from a large-scale research project on adaptive water management," *Ecology and Society*, 2007.

[Finkel] Finkel, Alan, "Panel on International Cooperation in Life Sciences," IEEE Life Sciences Grand Challenges Conference (LSGCC 2013), 2013.

[Huber] Huber, Peter W., *The Cure in the Code: How 20th Century Law Is Undermining 21st Century Medicine*, 2013.

[Hunter] Hunter, Peter, "Frontiers of Computational Medicine," IEEE Life Sciences Grand Challenges Conference (LSGCC 2013), 2013.

[Jeffrey] Jeffrey, P., "Smoothing the Waters: Observations on the Process of Cross-Disciplinary Research Collaboration," *Social Studies of Science*, 33 (4):539–562, 2003.

[Kamm] Kamm, Roger D, "Engineering Microvascular Networks for Therapeutic and in vitro Applications," IEEE Life Sciences Grand Challenges Conference (LSGCC 2012), 2012.

[Khargonekar] Khargonekar, Pramod P., "Issues and Perspectives on Cross-Disciplinary Research and the Role of Industry," IEEE CDC 2003.

[Kun] Kun, Luis G., National Defense University, personal communication.

[Marchione] Marchione, Marilynn, "Gene therapy scores big wins against blood cancers," The Associated Press, 2013.

[Nugent] Nugent, Kathy L. & Kulkarni, Avi, *Nature Biotechnology*, September 2013.

[PWC] PWC 2011, "Medical Technology Innovation Scorecard: US Falling Behind in the Race for Global Leadership," 2011.

[Radda] Radda, George, "Biomedical Knowledge in the Service of Man: Social Responsibility of the Scientist," IEEE Life Sciences Grand Challenges Conference (LSGCC 2013), 2013.

[Ramasubramanian] Ramasubramanian, Melur K., "Next Generation Convergence- Driven Research and Interdisciplinary Workforce and Development Efforts," IEEE Life Sciences Grand Challenges Conference (LSGCC 2012), 2012.

[Schiestl], Schiestl, Randy, "Translation - From Bench to Bedside, Medical Device Development," IEEE Life Sciences Grand Challenges Conference (LSGCC 2012), 2012.

[Sharp] Sharp, Phillip A. et al., "New Biology for the 21st Century," IEEE Life Sciences Grand Challenges Conference (LSGCC 2012), 2012.

[Thakor] Thakor, Nitish, "Frontiers of Human Brain," IEEE Life Sciences Grand Challenges Conference (LSGCC 2012), 2012.

[Wah] Wah, Benjamin, "Chinese University of Hong Kong, Life Sciences Big Data," IEEE Life Sciences Grand Challenges Conference (LSGCC 2013), 2013.

[Ying-I] Ying-I, Yong, "Affordability and Effective Healthcare in Singapore," IEEE Life Sciences Grand Challenges Conference (LSGCC 2013), 2013.

[Zhang] Zhang, Y.T., "Cardiovascular Health Informatics: Wearable Technologies and Unobtrusive Measurements," IEEE Life Sciences Grand Challenges Conference (LSGCC 2013), 2013.

3.21 Computational Biology and Bioinformatics

3.21.1 Background

The importance of computation in the acquisition, analysis, and modeling of biological systems has been steadily increasing for the past several decades. Contemporary bioinformatics and computational biology, twin fields divided roughly along the lines of data acquisition and analysis for the former and phenomenological modeling for the latter, comprise a strikingly wide range of topics and disciplines. Owing to the intrinsic breadth of biological phenomena, which ranges from the molecular to the cellular to the organismal and ecological, computational biologists and bioinformaticists must grapple with a diverse set of problems and devise an equally diverse set of tools to solve them. As a result, a number of distinct subdisciplines have come to define the field, coarsely demarcated by the scale and type of phenomena they address.

Genomic Bioinformatics

The acquisition and analysis of genomic data comprise the field of genomic bioinformatics. This includes the initial acquisition of raw sequencing data, the interpretation and assembly of such data into partial and complete genomes, the analysis of sequenced genomes for statistical correlations indicative of diseases and other traits, and the mining of genomes for overrepresented motifs and other sequence features [1].

Structural Bioinformatics

The analysis, modeling, and simulation of biological macromolecules—namely, proteins, DNA (DeoxyriboNucleic Acid), and RNA (RiboNucleic Acid)—comprise structural bioinformatics. The holy grail of the field has been, for several decades, the prediction of the three-dimensional structure of proteins from their amino acid sequence [2]. A similar challenge remains open for RNA molecules [3]. Beyond structure prediction, structural bioinformatics is concerned with the analysis and simulation of biomolecules to predict their interactions with other biomolecules and to infer useful physico-chemical properties [4].

Systems Modeling

The modeling and simulation of a set of biological parts is the domain of systems biology. What constitutes an appropriate set for study can range from a small subsystem of a biological organism, such as a single signaling pathway [5], to an entire biological cell with its complete metabolic and transcriptional networks [6], [7]. A plethora of modeling and simulation techniques are typically employed, depending on the complexity of the underlying phenomena and the availability of experimental data.

Phylogenetics and Evolutionary Modeling

The phenomena of the three aforementioned fields, the submolecular, molecular, and supramolecular, can all be studied in light of evolution. Evolutionary genomics concerns the use and comparison of multiple genomes to infer functional regions that are more likely to be conserved over evolutionary timescales. The use of evolutionary analysis of structures similarly helps identify functional hotspots on biomolecules and informs the prediction of protein structure [8]. Finally, the analysis of biological pathway evolution elucidates how the rewiring of cellular circuitry leads to new behaviors.

3.21.2 Current State of the Field

While bioinformatics and computational biology constitute a broad field, genomic bioinformatics currently occupies an oversized role within the field. This has been driven by significant changes in both supply and demand over the past few years. On the supply side, progress in sequencing technology resulted in explosive growth in the availability of genomic sequences, with the rate of increase outpacing Moore's law for over a decade now [9], [10]. In 2000, the first human genome draft was completed at a cost of $3 billion after a 10-year effort. Today, an entire human genome can be sequenced in less than a week and for less than $10,000 [11]. This abundance of sequence information, while a great scientific opportunity, has also created an unprecedented demand for new computing tools and infrastructure capable of analyzing enormous amounts of data. The trajectory of genomics is a classic example of a disruptive technology, particularly on the computational side. To underscore the point, the cost of computation in the overall sequencing pipeline has historically been fractional and inconsequential. As of 2010, the costliest aspect of the sequencing pipeline is the computational analysis required to turn raw data into completed genomes [12]. This presents a tremendous challenge to bioinformaticists and computer scientists to develop new algorithms and computational infrastructures capable of keeping up with the unrelenting growth in genomic data predicted for the next several years.

3.21.3 Challenges

The explosive growth in the availability of genomic data, in particular when compared to other bioinformatics fields that have not benefited from similar data growth, has resulted in a high fraction of the bioinformatics effort being focused on solving sequencing problems. The overarching focus of this effort has been the acquisition and assembly of genomic data, but not necessarily its interpretation, as captured by the classic *Cell* article titled, "Sequence First, Ask Questions Later" [13].

While this approach was appropriate during the initial stages of the genomic revolution, our ability to analyze genomic data now lags our ability to acquire it. One area where this is clear is genome-wide association studies, or GWAS. In such studies, a large number of patient genomes are sequenced, and individual genomic loci are tested for statistical correlations with diseases. Despite the initial high expectations for such studies, the current consensus is that most GWAS studies have been unsuccessful, because the typical strength of most disease correlations found has been very weak [14]. So serious is the problem that it has acquired its own name, "missing heritability," which refers to the many diseases that are known to be heritable but whose precise genetic causes have escaped elucidation [15].

The causes of the so-called missing heritability are myriad, including lack of sufficient data to provide the statistical power necessary to find very weak correlations. But equally important are the statistical and computational techniques used to mine genomic data, which were conceived in an era when hundreds, instead of trillions, of data points were the norm. Furthermore, such methods typically assume a simple,

even linear, mapping between inputs (genomes) and outputs (phenotypes), when in reality the functions mapping human genomes to disease phenotypes are likely to be extremely complex.

3.21.4 Where We Think It Will Go

The coming decade will see a shift in focus from genome acquisition to genome interpretation. This will likely be precipitated by three important developments.

Qualitative increase in data quantity. Advances in sequencing technology continue to be made, and if the exponential trajectory is maintained, a 100- to 300-fold increase in the number of sequenced genomes by the end of the decade is possible. Such increases will provide a qualitative improvement in available statistical power.

Improved statistical methodologies. Statistical inference methods designed specifically to tackle genomic bioinformatics will become increasingly more common and will exploit the unique structure of genomic data to infer subtle correlations, particularly ones in which a disease is dependent on the state of multiple mutations.

Convergence of genomic, structural, and systems approaches. Perhaps most importantly, the currently separate fields of genomic, structural, and systems bioinformatics will converge. The underlying driving force behind this shift is the complex mapping function between genotypes and phenotypes. Even with improvements in statistical methodologies and increases in data sizes, if every human genome were sequenced, the scientific community would obtain around 10^{10} genomes. In contrast, the mutational landscape of the human genome is around $4^{3,000,000,000}$ in size. Brute-force statistics and data acquisition will be insufficient to decode the human genotype-phenotype function. Instead, the interpretation of genomic data will need to proceed in a stepwise fashion, with the initial focus on understanding the molecular consequences of genomic changes. Doing so will require an understanding of how sequence determines structure, elevating structural bioinformatics to a central role in a disruptive manner. The types of analyses done within structural bioinformatics will be different from today's, as the emphasis shifts from coarse-grained prediction of de novo structures to the precise prediction of mutational effects on structure. The end result of this shift will be the convergence of genomic and structural bioinformatics.

As the ability to interpret genomic data molecularly improves, the next step will be to interpret genomic data in terms of systems-level phenotypes, at least on the pathway and cellular level. To do so will require that genotypes are first mapped onto structural phenotypes, which are then mapped onto systems phenotype, in a bottom-up approach. In a similar vein to the first shift, understanding the effects of structural changes on system behavior will necessitate a move away from the study of individual biomolecules to the study of complexes of molecules and their interactions. This is currently the domain of systems biology, but it is done in a top-down fashion in which high-level experimental data is used to fit observed systems-level phenomena, instead of a bottom-up approach in which known molecular interactions are simulated to obtain, in an emergent manner, the observed systems-level behavior. Achieving this will result in the convergence of structural and systems bioinformatics, where systems-scale structural simulations play a central role. Such a shift is already underway, although on a limited scale [16], [17].

3.21.5 Potential Disruptions

All the above shifts will prove disruptive. To a first order, they will push the computational and data storage requirements far beyond today's limits, potentially by several orders of magnitude, to the point where computation, instead of experiment, could become the major bottleneck. More importantly, these shifts will also require new types of computation, which in the long term may prove to be a more substantial disruption.

On the data analysis side, machine learning methods, including deep architectures that have recently seen a resurgence [18], will play an increasingly important role. Such methods have shown great potential for scalability when run on GPUs [19], which will likely further cement the role of GPUs in bioinformatics. On the structural simulation side, long time scale molecular dynamics simulations will likely play an increasingly important role, and specialized hardware, such as the Anton computer [20], have shown exceptional effectiveness at tackling such problems [21].

The broader impact of these changes will first be felt in the basic life sciences, where the convergence of disparate bioinformatics fields will help elucidate the mechanistic basis of biological pathways. In the longer term, this newfound understanding will be translated into new treatment strategies and therapeutic targets for human diseases. Two examples help illustrate the potential impact of these shifts.

Cancer Modeling

Many types of cancers are caused by somatic mutations, i.e., mutations acquired during the lifetime of an individual, which disrupt important signaling pathways in human cells. Currently, many large-scale projects are underway to identify the specific mutations responsible for different types of cancers [22], [23]. These projects rely on acquiring a large number of tumor genomes and searching for overrepresented mutations that may be indicative of a causal role. Unfortunately, as described earlier, finding such causal links is difficult, as many cancers are affected through a large number of mutations acting in concert. Furthermore, the disruptions caused by these mutations often affect multiple proteins in a signaling pathway, such that the integrative effect cannot be ascertained without a systems-level model of how the signaling pathway functions. The coming advances in structural and systems bioinformatics will make it possible to translate genomic data into molecular and systems phenotypes, and to establish a causal link between genotype and disease that may ultimately be disrupted therapeutically.

Polypharmacology

The development of therapeutic drugs is currently centered on finding a "target," typically a protein believed to play a causal role in a disease and whose activity is to be suppressed or enhanced. Much of the effort in medicinal chemistry is in finding drugs with a "clean" profile, i.e., ones that only affect their intended target while leaving all other proteins unperturbed. In the current era of one molecule one disease, this approach makes sense. However, as our understanding of the complex interactions underlying disease states improves, therapeutic approaches will take on an increasingly polypharmacological bent, meaning they will by design target multiple molecules because the disease state is induced by multiple molecules. Furthermore, even when a single molecule is targeted,

understanding the polypharmacology of a drug is important, as some lack of specificity may be more problematic than another. The integration of structural and systems approaches will play a crucial role in making designed polypharmacology a reality. By enabling the analysis and simulation of a drug's molecular interaction with all proteins in a given pathway, its systems-level behavior can be predicted, and possibly designed. In addition, the information gained from a more sophisticated understanding of the basic science of disease will provide additional targets for drugs to act on.

3.21.6 Summary

The past decade has been an exciting time in bioinformatics and the life sciences broadly, as fundamental breakthroughs in technology have made it possible to amass unparalleled amounts of data. The core challenges of this and upcoming decades will be the translation of such data into actionable knowledge, one that can improve human health and shed light on the principal mysteries of life. Much as mathematics, particularly group theory and topology, played a critical role in the development of 20th century physics, computation and machine learning are playing an analogous role in the development of 21st century biology. And much as physics proved to be a constant source of disruptive developments in the past century, it is likely that the intersection of computation and biology will play a similarly disruptive role in this and upcoming decades.

3.21.7 References

[1] D.W. Mount, *Bioinformatics: Sequence and Genome Analysis*, Cold Spring Harbor Laboratory Press, 2004.

[2] K.A. Dill and J.L. MacCallum, "The Protein-Folding Problem, 50 Years On," *Science*, vol. 338, no. 6110, 2012, pp. 1042–1046.

[3] M.G. Seetin and D.H. Mathews, "RNA Structure Prediction: An Overview of Methods," *Methods Mol. Biol. Clifton Nj*, vol. 905, 2012, pp. 99–122.

[4] J. Gu and P.E. Bourne, *Structural Bioinformatics*, Wiley-Blackwell, 2009.

[5] K. Sachs et al., "Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data," *Science*, vol. 308, no. 5721, 2005, pp. 523–529.

[6] J.R. Karr et al., "A Whole-Cell Computational Model Predicts Phenotype from Genotype," *Cell*, vol. 150, no. 2, 2012, pp. 389–401.

[7] J. Gunawardena, "Silicon Dreams of Cells into Symbols," *Nat. Biotechnol.*, vol. 30, no. 9, 2012, pp. 838–840.

[8] D.S. Marks et al., "Protein 3D Structure Computed from Evolutionary Sequence Variation," *Plos One*, vol. 6, no. 12, 2011, p. e28766.

[9] N.D. DeWitt, M.P. Yaffe, and A. Trounson, "Building Stem-Cell Genomics in California and Beyond," *Nat. Biotechnol.*, vol. 30, no. 1, 2012, pp. 20–25.

[10] E.R. Mardis, "A Decade's Perspective on DNA Sequencing Technology," *Nature*, vol. 470, no. 7333, 2011, pp. 198–203.

[11] K. Wetterstrand, "DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)," *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*; www.genome.gov/sequencingcosts.

[12] L.D. Stein, "The Case for Cloud Computing in Genome Informatics," *Genome Biol.*, vol. 11, no. 5, 2010, p. 207.

[13] "Sequence First. Ask Questions Later," *Cell*, vol. 111, no. 1, 2002, pp. 13–16.
[14] D.B. Goldstein, "Common Genetic Variation and Human Traits," *N. Engl. J. Med.*, vol. 360, no. 17, 2009, pp. 1696–1698.
[15] T.A. Manolio et al., "Finding the Missing Heritability of Complex Diseases," *Nature*, vol. 461, no. 7265, 2009, pp. 747–753.
[16] R.L. Chang et al., "Structural Systems Biology Evaluation of Metabolic Thermotolerance in Escherichia coli," *Science*, vol. 340, no. 6137, 2013, pp. 1220–1223.
[17] M. Duran-Frigola, R. Mosca, and P. Aloy, "Structural Systems Pharmacology: The Role of 3D Structures in Next-Generation Drug Development," *Chem. Biol.*, vol. 20, no. 5, 2013, pp. 674–684.
[18] S. Bengio et al., "Guest Editors' Introduction: Special Section on Learning Deep Architectures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, 2013, pp. 1795–1797.
[19] A. Coates et al., "Deep Learning with COTS HPC Systems," Proc. 30th Int'l Conf. Machine Learning, 2013, pp. 1337–1345.
[20] D.E. Shaw et al., "Anton, a Special-purpose Machine for Molecular Dynamics Simulation," *Comm. ACM*, vol. 51, no. 7, 2008, pp. 91–97.
[21] R.O. Dror et al., "Structural Basis for Modulation of a G-Protein-Coupled Receptor by Allosteric Drugs," *Nature*, 2013.
[22] Cancer Genome Atlas Research Network, "Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways," *Nature*, vol. 455, no. 7216, 2008, pp. 1061–1068.
[23] T.J. Hudson et al., "International Network of Cancer Genome Projects," *Nature*, vol. 464, no. 7291, 2010, pp. 993–998.

3.22 Robotics Challenges for Emergency Medical Care

3.22.1 Introduction

In the field of medicine, technology advancements (e.g., CT [computed tomography] scans, MRI [magnetic resonance imaging], and heavy ion radiotherapy) have allowed a remarkable evolution. In particular, the advancement of computer science has played an important role in overcoming the technical difficulties in invasive measurement and image processing for screening undiscovered diseases. Computer-aided surgery (CAS) is expected to become standard soon, due to its low invasiveness, lower hand vibration, and highly accurate positioning, all of which leads to remarkably reduced physical burdens on the patient and a general enhancement to quality of life. The da Vinci Surgical System, one of the most well-known CAS systems, has already been commercialized, induced a surgical revolution, and installed in more than 2,700 operating environments around the world. However, its distribution and surgical application are still limited. About 70 percent of total robotic procedures are applied in the pelvic region, such as hysterectomy and prostatectomy. Other types of surgical robotic systems have been reported in the literatures, but most are not yet commercially available. The practical development of medical robots in association with ICT (information and communication technology) and RT (robot technology) in the field of emergency medical care has been launched, but very few RT systems have been implemented, due to technical difficulties and severe conditions such as time and space limitations.

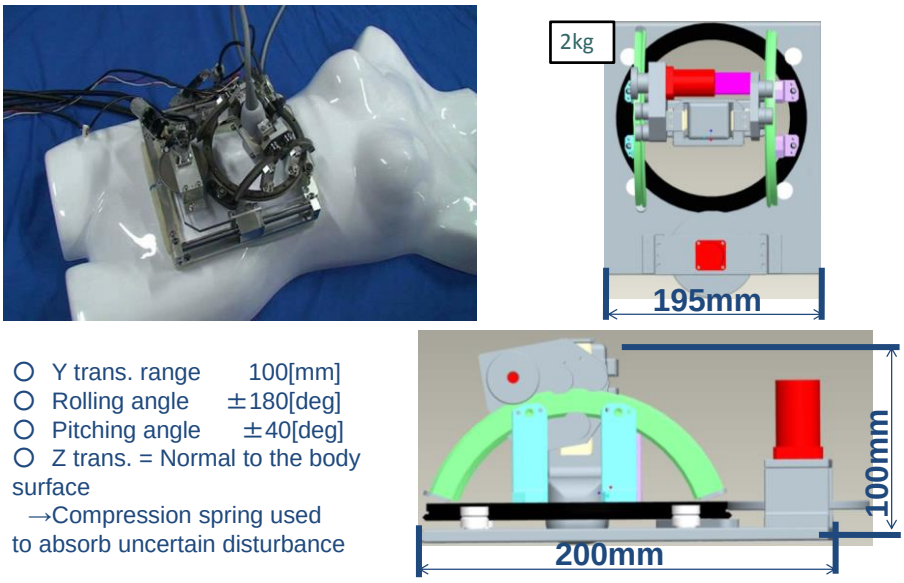


Figure 15. Portable tele-echography robot: FASTele.

3.22.2 State of the Art

Primary emergency care requires quick diagnosis and treatment especially for patients who might by bleeding internally. Focused Assessment with Sonography for Trauma (FAST) is widely used as a

first life-saving step for patients suffering from internal bleeding. However, transport to the hospital usually prevents trauma patients from an immediate FAST diagnosis. To resolve this issue, a portable tele-echography robot that a paramedic can attach to the patient to help doctors remotely and noninvasively search for internal bleeding with an ultrasound (US) image while the patient is in transit has been developed. Three mounted motors and the software running them allow intuitive control over position and orientation of an echo-probe through a smartphone or touchscreen. Experiments have indicated that the robot can be used to complete FAST under an MD's control within 9 minutes and that the extracted US images were clear enough for analysis. These results indicate that the robot is worth using, suitable for FAST, and effective in emergency medical care.

Table 1. Sensitivity of the FAST (ER, Creighton University).

	FAST Positive	FAST Negative
Internal Bleeding	**88 (42.7%)**	118
Not Found	5	1894

3.22.3 Challenges

A report from Creighton University pointed out that FAST screening unfortunately has a low sensitivity, approximately 42.7 percent (Table 1), and that delays in life-saving treatments because of internal bleeding being missed have become a serious problem in emergency medical care. A US image-processing method (See Figure 17) that helps emergency physicians detect internal bleeding during FAST improves the sensitivity and will be installable to the echography robot, FASTele, as well. The US method involves 1) extraction of low-brightness areas with several filters for edge detection and brightness gradients (See Figure 16), 2) extraction of organ boundary and identification of each organ with a ratio of low-brightness to high-brightness areas after brightness gradient filtering (See Figure 18); and 3) identification of bleeding on low-brightness areas around organ boundaries (See Figure 19). The

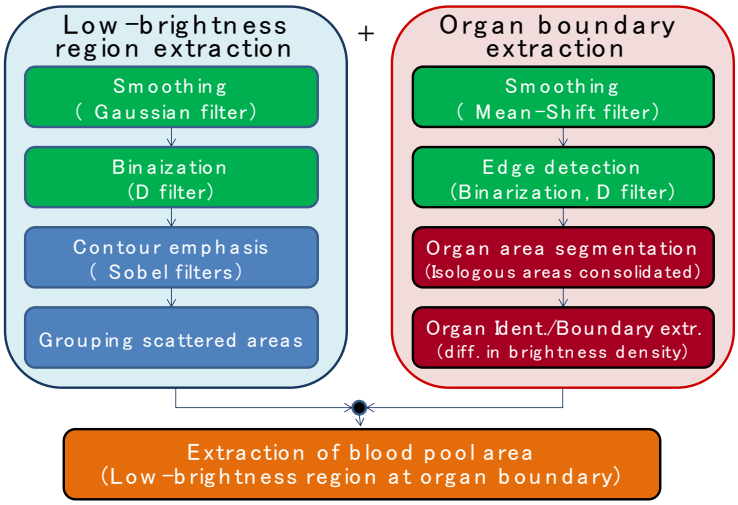


Figure 17. Internal bleeding extracting algorithm.

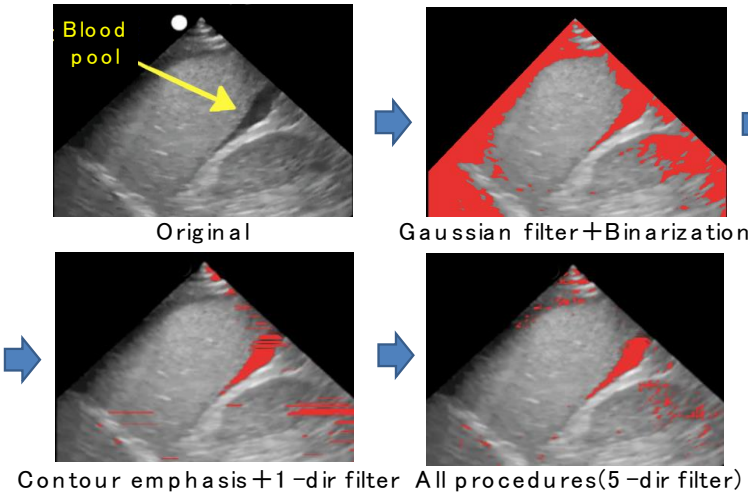


Figure 16. Extraction of low-brightness area.

proposed algorithm detects internal bleeding from clinical images with a much higher sensitivity, achieving 77.8 percent of accurate detection. The average time taken for bleeding detection was 4 seconds with an orginal PC system, which is sufficient for clinical application. However, automatic identification of internal bleeding requires processing at least 30 images per second—much more computer power is required to get 4 seconds processing time over 30 images and retain clinical value.

3.22.4 Potential Disruptions

Africa's economic growth and trans-Pacific strategic economic partnership agreement will bring about a new wave of healthcare globalization by means of ICT and RT coordinated techonology. Robotics in medical care has great potential to allow elderly/pregnant patients to undertake medical services such as telecheckup/telediagnosis/teletherapy beyond hospital/region/country/continent (See Figure 20). The early stage of deployment will require ICT advancements in higher-speed, secure, stable connectivity so as to allow medical doctors to share image data (US, CT, MRI) anytime and anywhere. At the latter stage, RT technology such as FASTele, bleeding detectors, and other RT systems will have a chance to provide new clinical values to the ICT infrastructure. For example, the FASTele tele-echography robot can work for patients requiring critical care on cruiseships or aircraft, and could reduce travel burdens related to prenatal telecheckups.

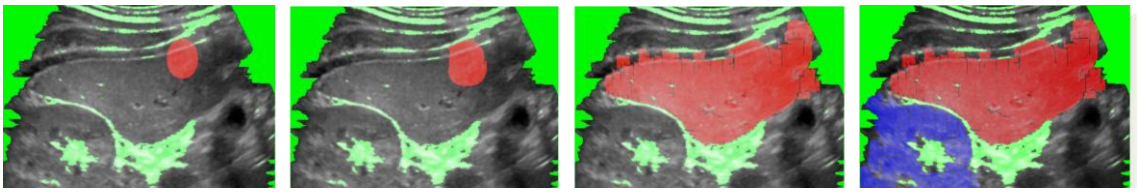


Figure 18. Organ area segmentation.

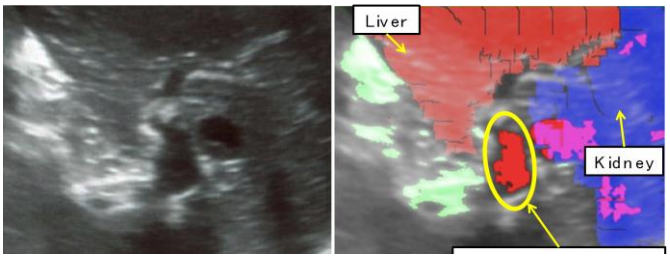


Figure 19. Internal bleeding detection by extracting low-brightness areas around organ boundary.

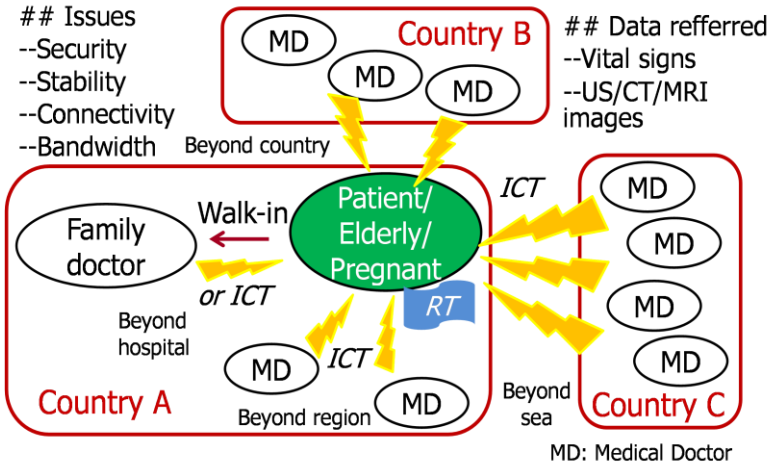


Figure 20. Medical services beyond country with ICT and RT.

4 Drivers and Disruptors

To verify the premises and conclusions that the *IEEE CS 2022* team made, we surveyed a few thousand IEEE members. Questionnaires went out after we selected the technologies and wrote initial drafts.

We posted two classes of questions, asking those who responded to rank driver and disruptor technologies. We offered the following items to be ranked:

Drivers

- Increases in average life expectancy
- Increasing ratio of retirees to workers
- Public concern over control over access/amount of personal information
- Desire for sustainable energy sources
- Reduction in availability of grants and philanthropic resources
- Widening economic inequality worldwide
- Reduced job security in a global market economy
- Climate change
- Global terrorism
- Use of big data and analytics
- Reduction in cost of data collection and retention (for use in analytics)
- Quickening pace of knowledge transfer (e.g., instantaneous global communication)
- Long-term availability of certain energy sources
- Alternative distribution chains (such as manufacturers selling directly to consumers)
- Use of technology for medical procedures
- Wireless/broadband connectivity

Disruptors

- Crowd-sourcing/open-sourcing of hardware development
- Changes in educational structure/design (e.g., MOOCs)
- Virtual/alternative currencies (such as Bitcoin)
- Smartphone use as a device for payment
- Cloud computing
- Use of robots as a source of labor
- Nonvolatile memory influencing big data accessibility and portability
- Quantum/nondeterministic computing
- Use of 3D printing
- Green computing
- New user interfaces (e.g., Siri, Kinect instead of traditional keyboards)

We received the following answers, represented in two figures.

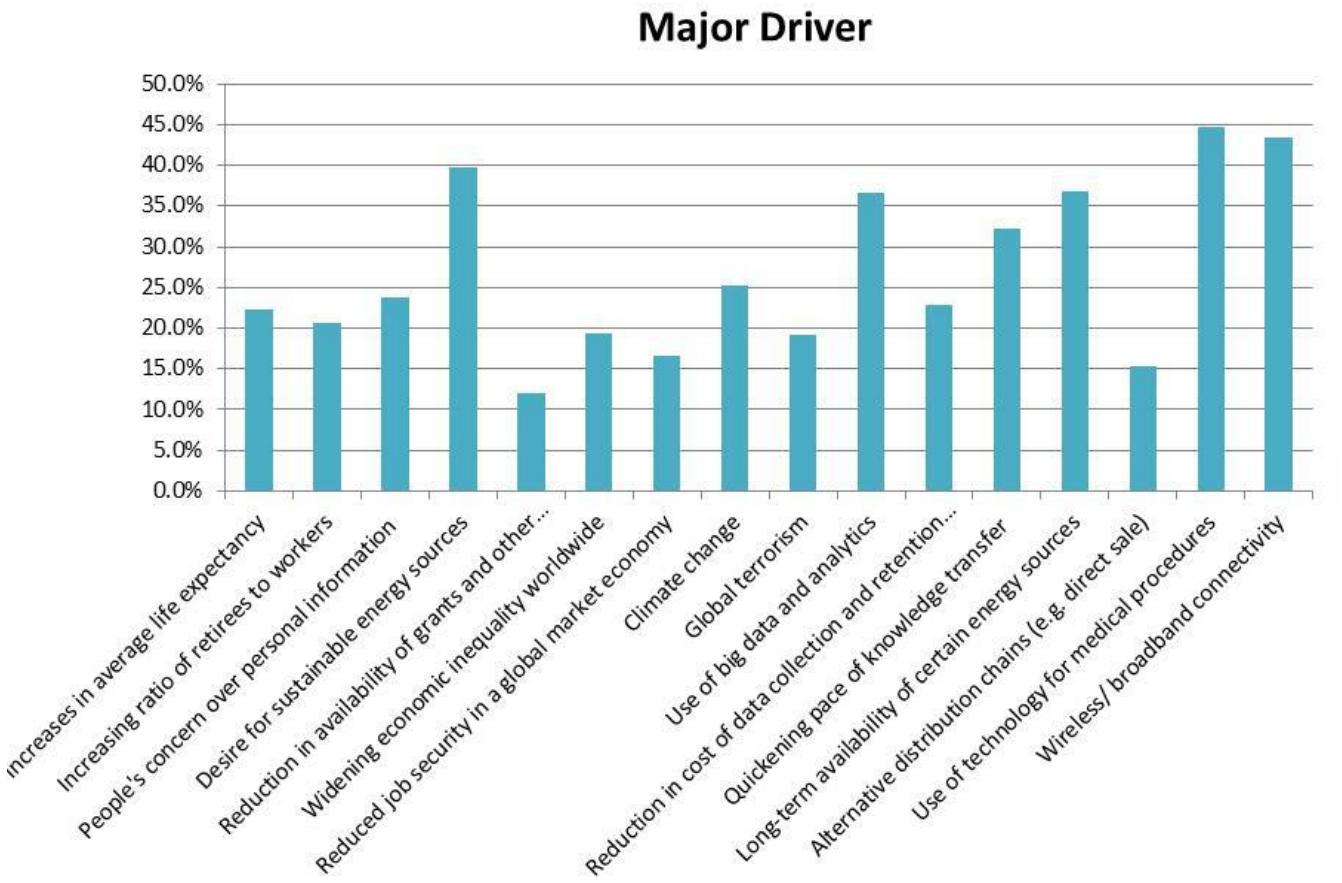


Figure 21. Comparison of major drivers.

Results aligned well with independent report findings. For example, the highest ranked drivers were the use of technology for medical procedures, followed by wireless/broadband connectivity and desire for sustainable energy sources. Also highly ranked were the use of big data and analytics; long-term availability of energy resources; and quickening pace of knowledge transfer. All of these drivers are discussed in the report.

Similarly, for major disruptors, use of robots as labor and 3D printing led the votes, followed by cloud computing, MOOCs, and new user interfaces.

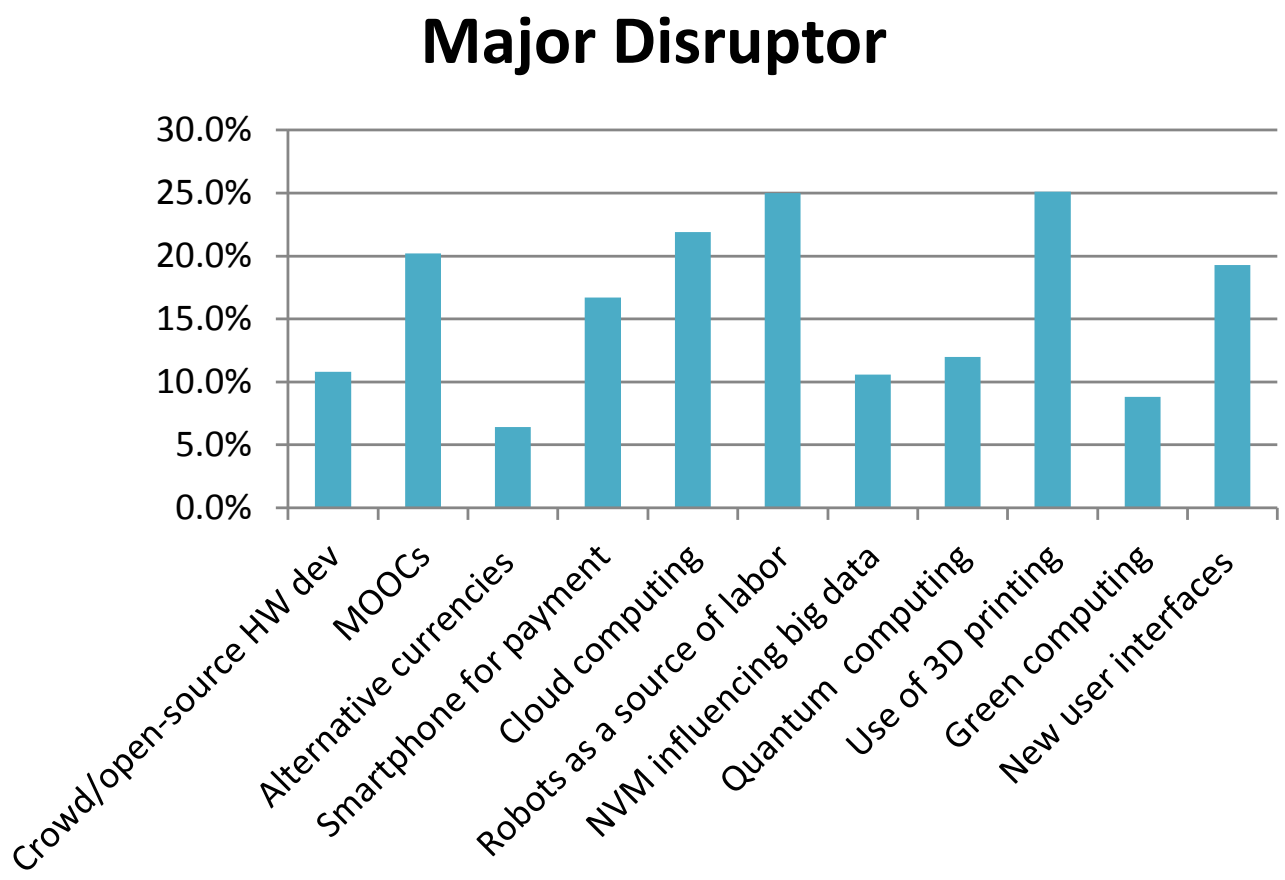


Figure 22. Comparison of major disruptors.

5 Technology Coverage in IEEE Xplore and by IEEE Societies

5.1 Introduction

These 22 technologies were analyzed for how IEEE Computer Society volunteer leaders can recruit existing experts and organizers and then invest in these technologies for growth. As technology convergence is inevitable, the Society will also partner with other major technical Societies to grow these areas. As IEEE encourages thought leadership in key technologies, funding might be available for these investments.

IEEE periodicals, conferences, and standards are key knowledge creation centers. IEEE-published articles have high quality and show the current state of the art. These articles are widely available to support additional research, attract funding, and attract other authors to build out the technical roadmap to drive the technology evolution. These knowledge centers are franchised by Societies, so Societies need to work together on emerging technology trends. Working together includes sharing investments, branding, publicity, leadership, and stewardship for future investments. The IEEE Computer Society continues to attracted many authors to our periodicals, conferences, and standards, thus the Society will use this base as an important stake to grow these 22 future technologies. As we get organized around these technologies to invest in them, these authors will be attracted to publish again.

The motivation for this analysis included the need to identify potential Society partners to best collaborate and to better understand overlaps for future consolidation. This was reinforced by the Communications Society's 2012 ComSoc 2020 report, which stated that its current technology trends were going to require partnering with other IEEE Societies, such as the Computer Society.

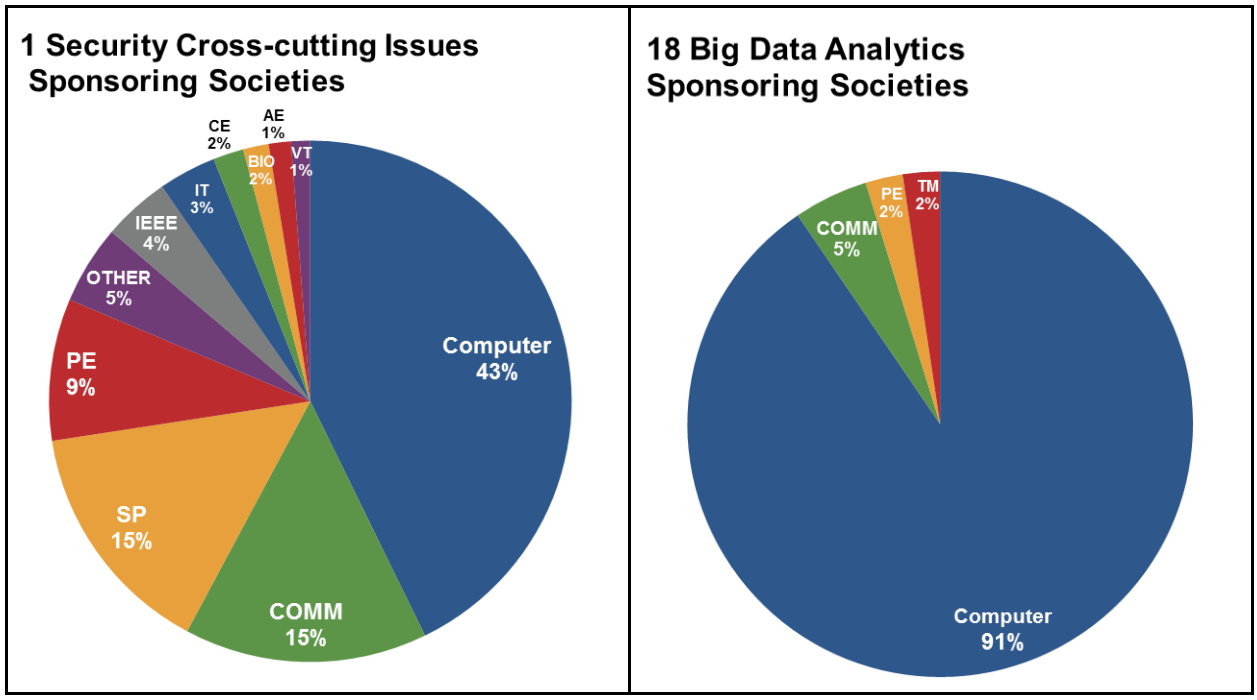
Within our society, Special Technical Communities (STCs) are formed to address emerging technologies such as these 22. This analysis helps STC leaders reach out to other Societies' technical communities for collaboration on conference tracks, special issues in our periodicals, or draft standards to focus on an emerging technology.

5.2 Comparison

Each of these 22 technologies was analyzed for its coverage of IEEE periodical articles published from 2000 to 2013. These periodicals are technically and financially sponsored by Societies, whose stewardship is for growth in quality reputation, author prestige, author submissions, relevance, and subscribers and readers. Mapping the numbers of periodical articles to the sponsoring Societies reveals each Society stake or coverage in our 22 technologies. This coverage is shown in each technology pie chart, where portions show the major IEEE technical Societies contribution according to their share of related periodical articles. The final chart shows the all the IEEE technical Societies for all 22 technologies along with a companion word cloud. [Note that neither conferences nor standards were included in this analysis as the true sponsoring influence is difficult to assign to a Society.]

The objective measure of a Society's "coverage" in a technical area was measured by the number of IEEE periodical articles published. As every Society (co-)sponsors periodicals, the Society was given credit by the number of articles discovered when searching with technical area's keywords. These keywords frame the technical area and were provided by our contributing technology area experts; we tested them in Xplore's advanced search for IEEE periodicals articles' metadata between 2000 and 2013 to avoid popular terms that included articles outside the technical areas. These keywords were arranged into a Boolean search expression to select the most relevant articles within the technology area. A large

sample (up to 2,000) of articles' metadata was downloaded to assign the Society sponsoring the periodical. We tabularized the Society's articles into a pie chart to highlight the top 10 Societies plus "other" collected Societies with lower article counts. Below are the four pie charts that our survey identified as "drivers" (See Footnote[12] for the description of acronyms for societies):



[12] AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

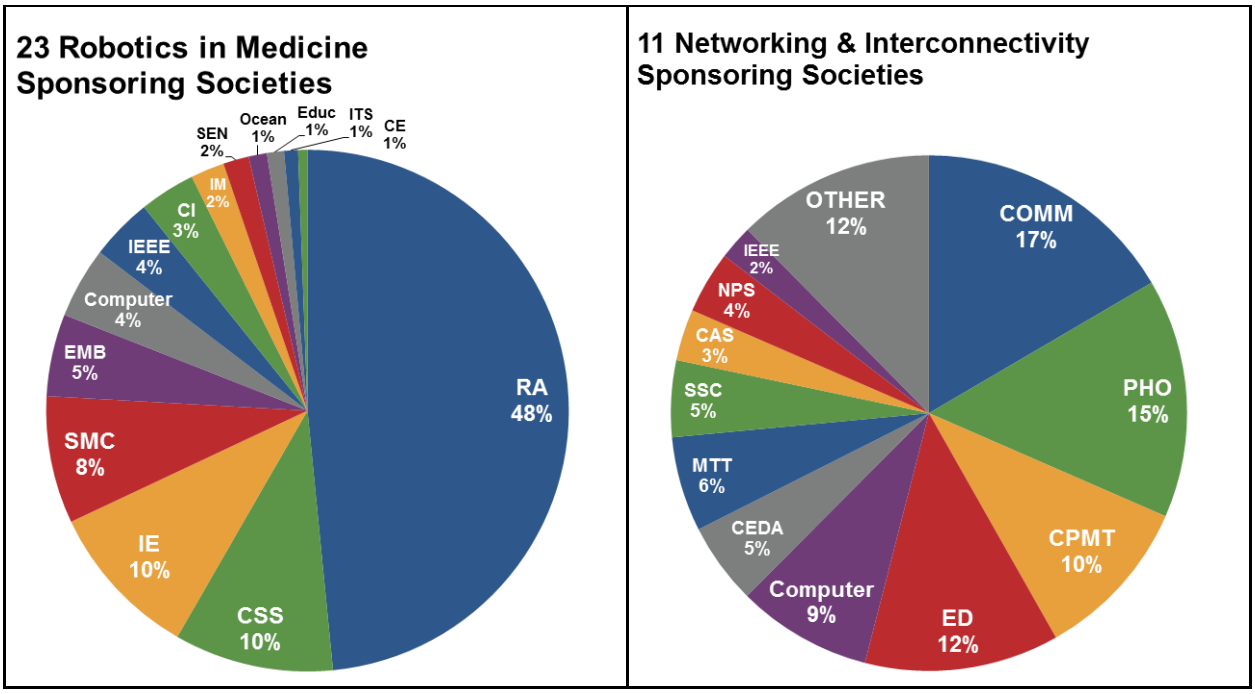


Figure 23. Coverage of some of the top drivers in IEEE Libraries by individual societies.

The appendix provides additional and enlarged chart views. Note that two of the four show the Computer Society with the largest coverage: the first indicates that many potential partners could be organized, whereas the second hints at going alone. In the other two, the Computer Society has a minority of the coverage and desire to achieve higher growth the first graph showcases the large leadership coverage and could organize the many contributing Societies to work together, the second shows many equal players that may lack leadership to accomplish together larger goals.

The survey also identified four major disruptors whose pie charts are as follows:

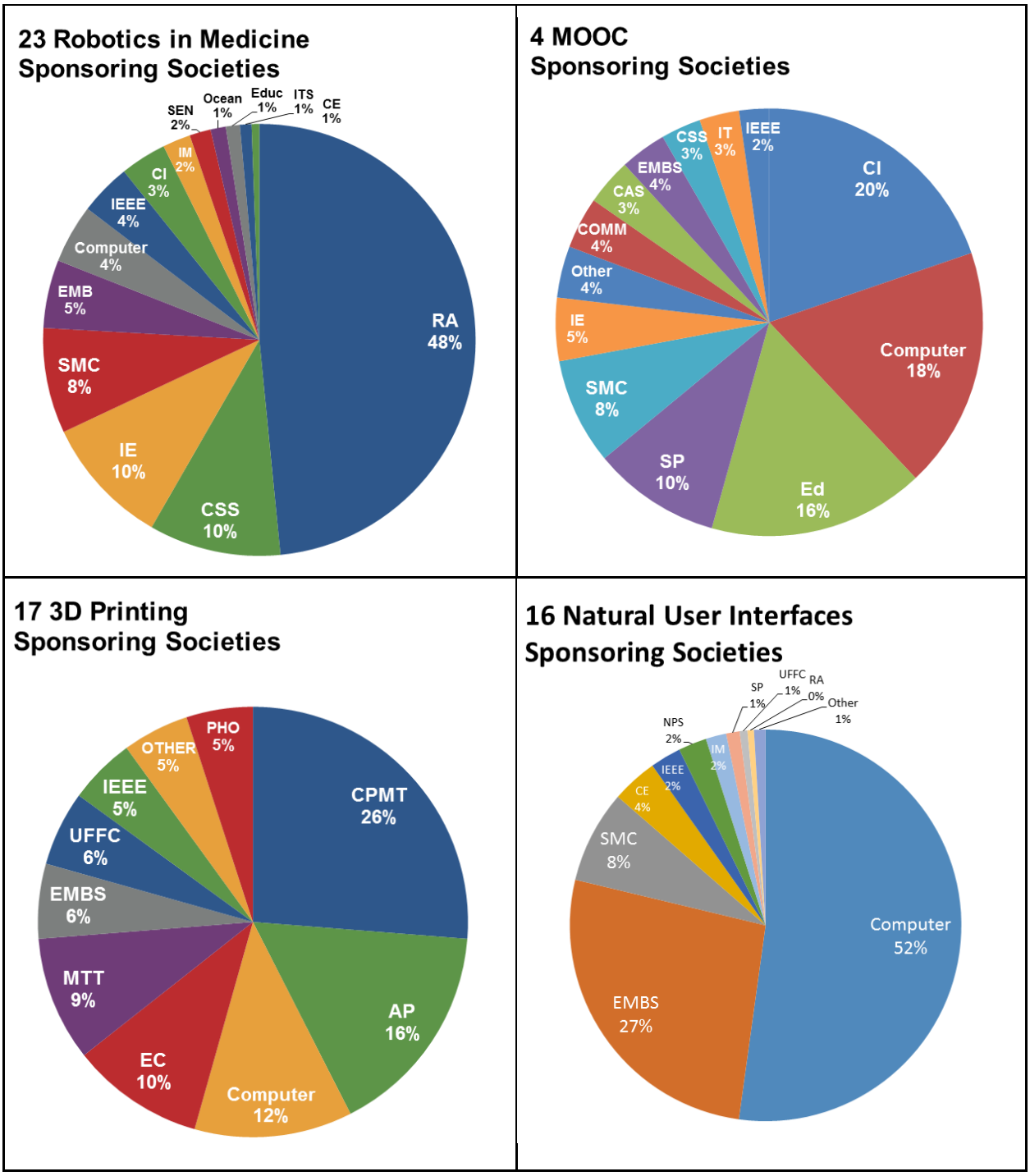


Figure 24. Coverage of some of the top disruptors in IEEE Libraries by individual societies

Again, we see two Societies with the most coverage that are also able to provide the leadership to organize this technology domain. The other two show wide coverage with no obvious organizing leader.

Because this quantitative analysis looks at the Societies' current technical assets in Xplore as the foundation for future technologies, there is the risk of disruptive forces that could kill some current technology trends, revive dead-end "solutions," or create new marriages/mergers of technologies.

This document and its analysis is the Computer Society's perspective, but it could compliment the Communication Society's perspective in its ComSoc 2020 report. At the higher IEEE Technical Activities level, the Future Technologies Committee has its own perspective. The Standards Association enlisted several Societies in its future view of smart grid, for example, the IEEE Smart Grid Vision for Computing: 2030 and Beyond. At the highest level, the IEEE Board of Governors is also defining our future world of technologies. All IEEE operating units need to share and plan for the future

5.3 Summary of Quantitative Analysis Findings

As mentioned above, our future technologies require partnering with other Societies with more assets, authors, experts, and organizers.

While our Society is known for its software engineering heritage, there are now many hardware-focused Societies using our technologies, and they are ripe for future collaboration.

IEEE has encouraged Societies in the stewardship of their own technical field of interest "silo" for excellence and growth, but our analysis shows the Societies need to partner on many emerging/future technologies. Of our chosen 22 technologies, the Computer Society has the top share in only 11, or 50 Percent; our rivals include EMB (Engineering in Medicine and Biology), RA (Robotics and Automation), NANO (Nanotechnology), PHO (Photonics), COMM (Communications), CI (Computational Intelligence), ED (Electronics Devices), CEDA (Council on Electronic Design Automation), and CPMT (Components, Packaging, and Manufacturing Technology). It is easy to partner when you are in the top position, but it is more difficult to reach out to those Societies with more "wealth" and "talent" as a contributing partner. Thus, IEEE Societies need incentives to reach out of their field to work with other Societies in moving IEEE to thought leadership. For example, while the Computer Society trails in a distant second place to the EMB Society in life sciences, we will have to work hard on this partnership to jointly reap the benefits of future growth.

6 IEEE Computer Society in 2022

The previous sections of this document have focused on the "what" and the "why." What technologies will be important? Why will they be important (drivers and disruptors)?

Left to be addressed are the questions of "who" and "how."

- Who will take up the work of tomorrow? This requires a focus on the development of the profession and professionals. The IEEE Computer Society must foster a highly skilled workforce.
- How can the CS fulfill its mission to benefit humanity? This requires a focus on the impact of technology on society. The CS must take responsibility for the *beneficial* implementation of technology.

The CS will support "seamless intelligence" for our members. Our members will all be global—*truly global*—and truly connected. Truly global means almost anyone on the globe who has interest can instantly become a member and participate in the Society's special technical communities, get access to all of its products and services, and be connected with other members virtually or at face-to-face meetings.

Early use of technology by next-generation professionals will drive the average age of our members down 10 years or more. The invent-to-publish cycle will be much shorter, with almost instant access to materials, ability to collaborate, and physical/virtual meetings. Crowd-sourced peer review will be the norm, but more important are new standards that will drive high quality: communities will tend to code that "lives," and community members will enforce professional codes of practice and collaborate to develop "building codes" for secure infrastructure (as the CS is pursuing through its cybersecurity initiative).

The value of these new products will be known immediately: users will rate their benefits. While technology will always have its "cool factor," traditional engineering principles and rigor will not be compromised. What will be different is how knowledge is transferred. You will learn *by doing* and learn *from doers*.

This new culture will give rise to more interactive events like hackathons, gaming conventions, and meetups. Traditional academic meetings will be joined by more practitioner conferences; the two will complement each other.

Recent developments in Internet security and privacy have eroded the universal view that technology advancement is always for the good. "Seamless intelligence" may not sound so positive to everyone. The CS must be societally mindful of preserving privacy as well as overall use of technology. We must make sure that societal changes are at pace with technology, but not slower.

More generally, all Societies should realize that technology is just one tile in a complex puzzle, and we can understand technology implications not only by looking at technology itself but also by looking at the whole puzzle. In a way, technology is about to be commoditized, and we need to appreciate that its value is in the economic, social, and cultural domain. These aspects are so intertwined with technology that we can no longer claim that technology is neutral. We need to consider in our publications and

conferences voices from other sectors and progress the whole puzzle (Saracco, personal communication).

We must also make sure our Society changes at pace with technology, but no slower. While this document does not go into detail about the internal workings of IEEE or the CS, some things become apparent from a simple strengths, weaknesses, opportunities, and threats (SWOT) analysis.

Table 2. IEEE SWOT.

Strengths	**Weaknesses**
Brand denotes quality	**Not nimble: IEEE has a hierarchical structure, whereas flat structures are known to be more nimble and less bureaucratic**
Scope encompasses all technologies (multidisciplinary)	**No presence in workforce development, which will be widely needed**
Global reach puts it in a position to develop technologies that cross borders	**IEEE's delivery channels are not keeping pace**
Neutral organization, which is an advantage in technical fields that are highly proprietary	**Very little professionally written content in CS publications, making it difficult to quickly reflect technology trends**
Opportunities	Threats
Need for highly skilled technical workforce is rising	**Organizations in emerging regions may be able to offer similar or better content, communications, services, communities, and chapter models that reflect local languages and norms; the same organizations can offer more universal products and services via the Internet at lower prices**
CS can lead IEEE in how to do open access	**Open access will slow IEEE's financial engine and disrupt its business model: intellectual property will not be granted to the IEEE at little to no cost**
IEEE can expand its influence in technology applications, such as healthcare	**Self-credentialing makes membership in a Society obsolete (stack overflow); peer review bodies are also self-organizing (research gate)**
Exciting time to reframe what it means to be a technology professional, at all levels	**Google Scholar-like services undermine IEEE collections**
The need for technical education will grow; IEEE can develop professional standards of practice, such as software engineering; IEEE can offer certifications that require membership	**Distrust of technology could grow, putting IEEE in an unfavorable light**
Asia is a huge opportunity at the moment: IEEE can uncover what's next	**Other professional associations and information services companies are moving faster than IEEE right now**

A SWOT analysis can be used for scenario planning. The CS used scenario planning in 2004 when it was developing Strategic Plan 5,[1] to imagine what the year 2020 might look like. Much of the 2004 scenario includes elements that are still current. This table summarizes the SP5 predictions as to who the CS would be serving in 2020, and how.

Table 3. Breakdown of who and how will be benefiting from IEEE CS.

Who	How
Multidisciplinary professionals	Deliver highest quality content, but in small units
Global citizens	Digest, synthesize, summarize, and repackage content
Constantly become "instant experts"	Offer skills development and training
Volunteers who can devote small bits of time intermittently	Focus on "high touch" but small face-to-face meetings on the latest topics

If these 2020 future states are still relevant, how is the CS doing? SWOT shows that we have developed some strength in these areas, but most of these challenges remain.

The second part of the 2022 report, CS Strategic Plan 8, will address how the Computer Society must be organized to have an impact in the year 2022, in light of the technologies outlined in this report.

The opportunities for IEEE Computer Society in 2022 are endless, and the future is exciting.

7 Summary and Next Steps

In this report, several technologists evaluated 22 technology areas that have the potential to disrupt the world we perceive today. Some of them are already known and being adopted today, such as multicore, high-performance computing, cloud computing, and software-defined networks. Others are only being explored at this time, such as 3D printing, nonvolatile memories, and quantum computing. These 22 technology areas cover a spectrum of policies (open intellectual property movement, massively online open courses), technologies (15 areas), market categories (computational biology and bioinformatics, life sciences, and robotics in medical care), and some vertically applied areas (sustainability and security cross-cutting issues).

These 22 areas have resulted out of brainstorming by technologists. They have been confirmed through subsequent questionnaires and also compared through digital libraries exploration. All areas have been tied into a single scenario that we call "seamless intelligence." While similar to past pervasive and ubiquitous computing scenarios, the seamless intelligence scenario has deep roots in technology advancement that did not exist in the near past. In particular, by 2022, computing devices will vary from nano- to mega-scale, and wireless/wired networks will enable access to integrated services. Virtual connectivity will enable integration of relevant computing resources to provide users with integrated and seamless services. The resulting ecosystem will offer seamless, continuous, uninterrupted services that enhance automation, productivity, collaboration, and access to intelligence and knowledge through emerging HCI.

However, the benefit of technology is what we make of it. Societies will face challenges in realizing technologies that benefit humanity instead of destroying and intruding on the human rights of privacy and freedom of access to information. How will these advancements will help humanity will depend on the pace of the policies and regulations that accompany the technologies' evolution. Like many times in the past, technology is an enabler. It is up to the human race to leverage it in the best possible way to advance human society.

This report is made freely available, but it was gradually distributed within the IEEE Computer Society to get the best feedback from our readership. Over the course of the following year, it will be used in preparation of the IEEE Computer Society's strategic plan. While *IEEE CS 2022* is more technology focused, the strategic plan will be more IEEE Computer Society focused.

Ultimately, this was a rewarding exercise. It was very interesting to lead and participate in technology discussions about the future and encouraging to see how many of the technologists converged on a single scenario as well as how many similar concerns about privacy and security transpired through all other discussions. We all have learned a lot from this process, and we hope that the readers of this document will learn something, too.

Core Authors:

Arif Merchant, Danny Lange, Dejan Milojicic, Eitan Frachtenburg, Hasan Alkhatib, Hironori Kasahara, Karsten Schwan, Paolo Faraboschi, and Phil Laplante.

8 Authors

This document was a team effort, spearheaded by a core team of authors who formulated the overall text and process. This team, organized by Dejan Milojicic, met twice in face-to-face meetings and had a few phone conferences. In addition, other people contributed to various parts of the document; the rest of this section lists all contributors.

8.1 The Core Team of Authors

The core team of authors included Hasan Alkhatib, Paolo Faraboschi, Eitan Frachtenburg, Hironori Kasahara, Danny Lange, Phil Laplante, Arif Merchant, Dejan Milojicic, and Karsten Schwan.

Hasan Alkhatib, (Entrepreneur and President of SSN Services, LLC)

Alkhatib is a Silicon Valley entrepreneur and veteran. He was a Computer Engineering Professor at Santa Clara University from 1981-1998, specializing in networking and distributed computing. In 1998, he founded IP Dynamics, a venture-backed start-up, where he was CEO until 2007. IP Dynamics developed the industry's first solution for policy-based software-defined connectivity from anywhere to anywhere, regardless of location and underlying physical networks. In 2007, Alkhatib joined Microsoft as General Manager of Enterprise Networking, and then became Chief Architect of Networking and Network Security for Microsoft's cloud computing platform, Windows Azure, in 2008. He's been President of SSN Services, a consulting firm specializing in network virtualization, cloud computing, and innovations in higher education, since 2011. Alkhatib has published over 50 papers and holds 26 patents and 37 other pending patent applications on networking, virtualization, security, and cloud computing. He has chaired five IEEE/CS conferences and was keynote speaker at five others. He has served as guest editor for *IEEE Micro* and chaired TCMM in 1991-1992. Alkhatib holds a PhD in Electrical and Computer Engineering from UC Santa Barbara.

Hasan Alkhatib wrote the Seamless Intelligence Scenario and Cloud Computing sections.

Paolo Faraboschi, HP Labs, Spain

Paolo Faraboschi is a Distinguished Technologist at HP Labs, working on energy-efficient servers. From 2004 to 2009, he led a group on system-level simulation. From 1995 to 2003, at HPL Cambridge (Mass.), he was the Principal Architect of the Lx/ST200 family of VLIW embedded cores. Faraboschi is an active member of the computer architecture community: he has served as guest editor of *IEEE Micro*'s TopPicks 2012, and Program Chair for CF2012, HiPEAC2010, MICRO2008 and MICRO2001. He has authored 23 patents, the book *Embedded Computing: A VLIW Approach*, and over 50 papers. Faraboschi holds a PhD in EECS (1993) from University of Genoa, Italy.

Paolo Faraboschi contributed the sections on Universal Memory, 3D Integrated Circuits, and Photonics.

Eitan Frachtenberg, Facebook

Eitan Frachtenberg is a Research Scientist at Facebook, analyzing social behavior on large-scale datasets. His research interests include data mining, performance evaluation and optimization, Web technologies, parallel algorithms, and computer architecture. Prior to Facebook, Frachtenberg was an Applied Researcher at Microsoft/Powerset (working on Semantic Web search), and before that, a Postdoctoral Fellow at Los Alamos National Laboratory (working on supercomputer operating systems). He obtained his PhD in Computer Science from Hebrew University.

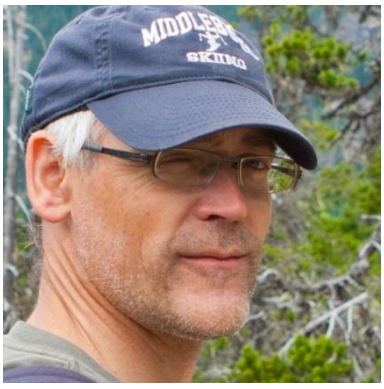
Frachtenberg wrote the sections on Big Data and Analytics, and Open Intellectual Property (together with Phil Laplante and encompassing crowd-sourcing).

Hironori Kasahara, Waseda University, Japan

Hironori Kasahara has been an IEEE CS BoG member since 2009 and a Chair of the Multicore STC since 2012. In 1985, he received a PhD in EE from Waseda University, Tokyo, where he has been a professor of computer science since 1997, and a Director of the Advanced Multicore Research Institute. He was a visiting scholar at UC Berkeley and the University of Illinois at Urbana-Champaign's Center for Supercomputing R&D. Kasahara received the IFAC World Congress Young Author Prize, and IPSJ Sakai Memorial Special Research Award. He has led Japanese national projects on parallelizing compilers, multicore, and green computing systems.
http://www.kasahara.cs.waseda.ac.jp/kasahara.html.en

Hironori Kasahara wrote the Multicore section of the report.

Danny Lange, Microsoft



Danny B. Lange is Manager of Elastic Machine Learning at Amazon.com. Prior to Amazon, he was Principal Development Manager at Microsoft, where he was leading the product team for large-scale machine learning. Previously, he was the Bing Software Architect responsible for mobile search. Lange was the co-founder of Vocomo Software, a speech technology company, and as CTO of General Magic, built the architecture for its OnStar voice response service. Prior to joining General Magic, he was Computer Scientist at IBM Tokyo Research. Lange has made significant contributions in the areas of distributed computing, big data analytics, cloud computing, mobile agents, speech recognition, program visualization, and hypertext. He holds an MS and a PhD in Computer Science from the Technical University of Denmark.

Lange has numerous patents to his credit, has presented his work at leading conferences, and published articles in many journals.

Danny Lange contributed the Machine Learning and Intelligent Systems, Natural User Interfaces, and Quantum Computing sections.

Phil Laplante, Pennsylvania State University

Phil Laplante is Professor of Software Engineering at The Pennsylvania State University. He received his BS, M.Eng., and PhD from Stevens Institute of Technology and an MBA from the University of Colorado. Laplante is a Fellow of IEEE and SPIE and has won several international awards for his teaching, research, and service. He has worked in avionics, CAD, and software testing systems and has published 27 books and more than 200 scholarly papers. Laplante's research interests are in software testing, requirements engineering, and software quality and management.

Phil Laplante wrote the section on Open Intellectual Property (together with Eitan Frachtenbery) and the section on MOOCs.

Arif Merchant, Google

Arif Merchant is a Research Scientist with the Storage Analytics group at Google, where he studies interactions between components of the storage stack. Prior to this, he was with HP Labs, where he worked on storage QoS, distributed storage systems, and stochastic models of storage. Merchant holds a B.Tech. from IIT Bombay and a PhD in Computer Science from Stanford University.

Arif Merchant contributed the 3D Printing section.

Dejan Milojicic, HP Labs, Palo Alto

Dejan Milojicic is a senior researcher at HP Labs and the IEEE Computer Society 2014 President. He was a founding editor in chief of IEEE ComputingNow and has been on many conference program committees and journal editorial boards. Milojicic worked at the OSF Research Institute, Cambridge, MA [1994-1998] and Institute "Mihajlo Pupin," Belgrade, Serbia [1983-1991]. He received his PhD from University of Kaiserslautern, Germany (1993) and MSc/BSc from Belgrade University, Serbia (1983/86). Milojicic is an IEEE Fellow, ACM Distinguished Engineer, and USENIX member. He has published over 130 papers and 2 books; he has 12 patents and 25 patent applications. His areas of expertise include systems software, distributed computing, mobile computing, and services.

Dejan Milojicic wrote the sections on High-Performance Computing and Sustainability. He also initiated and organized this effort and contributed to the remaining general sections.

Karsten Schwan, GaTech

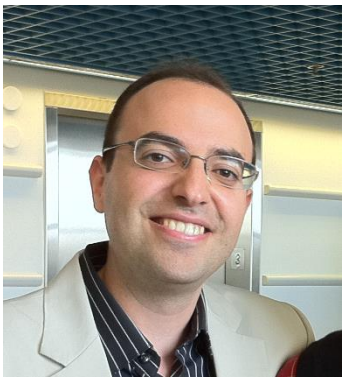
Karsten Schwan is a Regents' Professor in the College of Computing at the Georgia Institute of Technology, where he is also a Director of the Center for Experimental Research in Computer Systems (CERCS), with co-directors from both GT's College of Computing and School of Electrical and Computer Engineering. His MS and PhD are from Carnegie-Mellon University; his PhD concerned high-performance computing, addressing operating and programming systems support for the Cm* multiprocessor, after which he conducted extensive research in real-time and distributed systems. His current work ranges from topics in operating systems to middleware to parallel and distributed systems, focusing on information-intensive distributed applications in the enterprise domain and in the high-performance domain. www.cc.gatech.edu/~schwan

Karsten Schwan contributed the sections on Device and Nanotechnology, Internet of Things, and Networking and Interconnectivity.

8.2 Major Contributors of Individual Sections

In addition to the core team, a few individual contributed to substantial parts of the document. These valuable contributors include Mohammed AlQaraishi, Angela Burgess, Hiroyasu Iwata, Rick McGeer, and John Walz.

Mohammed AlQuraishi, Harvard Medical School



Mohammed AlQuraishi is a Systems Biology Fellow at Harvard Medical School. Prior to joining Harvard, he completed his PhD in Genetics from Stanford University under the supervision of Harley McAdams and Lucy Shapiro. His research interests lie at the intersection of systems and structural biology. AlQuraishi aims to obtain a systems-level understanding of biological processes through a molecular-level understanding of biological structures and their interactions, and to that end, he is developing computational methods for predicting the binding partners and quantitative binding affinities of biological molecules from their atomic structure. His work combines recent advances in machine learning and information theory with concepts from statistical mechanics and biophysics.

Mohammed AlQuraishi wrote the section on Bioinformatics.

Angela Burgess, IEEE Computer Society

Angela R. Burgess is the Executive Director of the IEEE Computer Society, the world's leading membership organization for computing professionals. The IEEE Computer Society is the largest technical organization within the IEEE, which has more than 400,000 members worldwide. As head of staff, Burgess oversees the Computer Society's Digital Library, 17 journals, 12 technical magazines, and 300 conference proceedings annually, along webinars, podcasts, courseware, bodies of knowledge, and certifications. She has more than 25 years' experience with the IEEE Computer Society and has been the Executive Director since 2007. Burgess received a BS in Journalism and International Studies from Iowa State University and an Executive MBA from the Peter F. Drucker School of Management, Claremont Graduate University.

Angela Burgess wrote the section on SWOT analysis and also contributed to the Section on 2022 Technologies Coverage in IEEE.

Hiroyasu Iwata, Waseda University

Hiroyasu Iwata received a BS, MS, and PhD in mechanical engineering from Waseda University, Tokyo, Japan. He was a Research Associate and an Assistant Professor at Waseda University from 2001 to 2004, and 2005, has been an Assistant Professor at the Institute for Biomedical Engineering, Consolidated Research Institute for Advanced Science and Medical Care, Waseda University. He is also a member of the Humanoid Robotics Institute and the WABOT HOUSE Laboratory of Waseda University.

Hiroyasu Iwata wrote the section on Robotics in Medical Care.

Rick McGeer, Communications and Design Group, SAP America

Rick McGeer received his PhD in Computer Science from UC Berkeley. He was an Assistant Professor in the Computer Science Department at the University of British Columbia, before returning to UC Berkeley as a Research Engineer in 1991. In 1993, he co-founded Cadence Berkeley Laboratories, the research arm of Cadence Design Systems, and five years later, he co-founded Softface, Inc., where he remained as Chief Scientist until 2003, when he joined Hewlett-Packard Laboratories, leaving in 2014 as a Distinguished Technologist. McGeer co-founded the PlanetLab consortium in 2003, and currently serves on the Steering Committee. In 2013, he joined US Ignite on a part-time, volunteer basis as Chief Scientist. He is currently a Principal Investigator with the Communications and Design Group, a research arm of SAP America. McGeer is the author of over 100 papers and one book in the fields of CaD, circuit theory, programming languages, distributed systems, networking, and information system design. His research interests include logic synthesis, timing analysis, formal verification, circuit simulation, programming languages, networking, wide-area distributed systems, and cloud systems. He has acted as a Principal Investigator on three DARPA and three GENI projects over the past two decades. He is also an Adjunct Professor of Computer Science at the University of Victoria, Victoria, BC, Canada.

Rick McGeer wrote the section on Software-Defined Networks.

John Walz, Retired from Lucent/AT&T

Former IEEE Computer Society President John W. Walz has been elected 2014 IEEE Division VIII Delegate-Elect/Director-Elect. Walz retired from Lucent Technologies/AT&T with more than 20 years of management/coaching experience, covering positions in hardware and software engineering, quality planning and auditing, standards implementation, and strategic planning. He has coauthored three books covering the use of IEEE software engineering standards to support CMMI, ISO 9001, and Lean Six Sigma. Walz has held leadership positions in national and international industry and professional organizations, including US Technical Advisory Group on Quality Management ISO 9001 and Risk Management ISO 31000; American Society for Quality (ASQ) Electronics and Communications Division and its Sarbanes-Oxley Forum; the Quality Excellence for Suppliers of Telecommunications Forum; and the Information Integrity Coalition.

John Walz wrote the section on Life Sciences and lead writing the section on 2022 Technologies Coverage in IEEE.

8.3 Acknowledgements

Greg Astfalk, HP Labs, for contributions to the Universal Memory section.

Evan Butterfield, IEEE Computer Society, for securing a number of images for final production of the document

Elena Gerstman, for facilitating the first 2022 Report core team meeting.

Moray McLaren, HP Labs, for contributions to the Photonics section.

John Reimer, IEEE Computer Society, for conducting searches and producing pie charts for appendix.

Jenny Stout, IEEE Computer Society, for copyediting this document.

Michael Werhman, for conducting the survey on drivers and disruptors.

Chandrakant Patel for contributing Figures 2 and 13.

Robert Stack contributing black and white drawings on pages 5, 9, 10, 11, 13, 16, 28, 52, 67, 73, 82, and 87.

APPENDIX I. 22 Technologies Coverage in IEEE Publications

Many IEEE Societies and councils are publishing content about each of the 22 technologies. These potential partners are identified on the subsequent pages of this Appendix.

The percentages represent one measurement of the degree of involvement each Society/council (S/C) has in each technology. It is calculated as the share of relevant content among the top-publishing S/Cs of that material in Xplore (the lowest-publishing S/Cs for any one technology are aggregated into the "Other" category for clarity).

Also included are summary charts of the amount of this content published in each technology area. For an example of how to utilize this data, life sciences accounts for 27 percent of the content assessed for this exercise, and the Computer Society accounts for 12 percent of that content. Or, taken as a whole, the Computer Society accounts for 25 percent of the 22 technologies' content identified.

The keywords chosen to identify the relevant content were developed by the 2022 technologies' subject matter experts, and are as follows:

Table 4. Search keywords summary.

Technology & indexing terms	Boolean search query	Total # Xplore articles
1. Security Cross-Cutting Issues	((Privacy OR Security OR Intrusion) OR Intrusion OR "Security legislation") OR (((cyber) OR cybersecurity) OR cyber-security) OR "cyber security"	12,389
2. Open Intellectual Property Movement	(((“Crowd sourcing”) OR “Open IP”) OR Open AND “Intellectual Property”) OR “Open standards”	1,416
3. Sustainability	((((“Energy usage”) AND computing)) OR (“Sustainability”) OR (“Green computing”) OR (“Carbon footprint”) OR (“Earth friendly”)) OR (Green ICT)) OR (Sustainable Computing)	882
4. Massively Online Open Courses	((((“Open Courses”) OR (“Massively Online”) NOT "Games") OR “Massively” AND “Courses”) OR "Online learning") OR "Automated grading"	458
5. Quantum Computing	“Quantum Computing”) OR (“Quantum” AND “mechanical phenomena”) OR (“Quantum properties”) OR (“Quantum annealing”) OR (“factorization algorithm” OR “Shor”) OR (“Qubit”	2,823
6. Device and Nano-technology	(((“Microelectromechanical systems”) OR “Nano-technology” OR “Nanotechnology” OR “Nano technology”) OR “Microelectromechanical systems”) OR “Micro machine” OR “Micro machines” OR “Micromachines” OR “Micromachine” OR “Micro-machine” OR “Micro-machines”	7,546

Technology & indexing terms	Boolean search query	Total # Xplore articles
7. 3D Integrated Circuits	((((((“2.5D chip" OR "2.5-D chip" OR "2.5D chips" OR "2.5-D chips"))) OR ("3D chip" OR "3-D chip" OR "3D chips" OR "3-D chips")) OR "System on a Chip") OR "System in a Package")	1,759
8. Universal Memory	((((“Non-volatile memory”) OR Memristor) OR “Spin Transfer Torque” RAM) OR "Phase Change Memory") OR “Universal Memory”	460
9. Multicore	(((((“Multicore") OR "Multiprocessor") OR GPU) OR ("Accelerators") AND "processor")) OR GPGPU) OR "Manycore"	2,276
10. Photonics	((“Photonics interconnect”) AND “Silicon photonics”) OR VCSEL OR “Vertical Cavity Surface Emitting Laser”)	1,313
11. Networking and Inter-connectivity	((("Interconnects") OR (((“Inter-connectivity") OR “Interconnectivity”) OR “Inter connectivity”) AND Networking) OR (“Ethernet”) AND "internet") OR "Ethernet" AND Networking	2,939
12. Software Defined Networks	(((((("Software Defined Networks") OR "Software defined networking") OR "Index Terms":SDN) OR OpenFlow) OR "Software radio") OR "Active networking") OR "Virtual Local Area Networks") OR VLAN	496
13. High Performance Computing	((((((“High Performance Computing" OR HPC)) OR Supercomputers) OR "Message Passing Interface") OR GPGPU) OR "Compute-intensive") OR Petascale) OR Exascale	1,068
14. Cloud Computing	(((((((Cloud Computing) OR “Grid computing”) OR “Cluster computing”) OR Virtualization) OR “-as-a-Service”) OR IaaS) OR PaaS) OR SaaS) OR “Pay as you go”	4,252
15. Internet of Things	(((((“Internet of Things") OR "Smart homes") OR Ubiquity) OR Pervasiveness) OR Interconnectivity) OR "Smart dust"	442
16. Natural User Interfaces	(“Natural User Interfaces” OR “NUI”) OR ((“gesture recognition”) OR (“Speech and gesture recognition”)) OR (“Graphical user interface” OR “NUI”) OR (“Human Computer Interface” OR “HCI”) OR (“Multi sensor input” OR “Multiple sensor input”) OR (“Augmented reality”)	3,581
17. 3D Printing	((((3-D) OR 3D) AND Printing) OR "Additive manufacturing") OR "Selective laser sintering"	215
18. Big Data and Analytics	(("Big data") OR “Massive Data”) AND Analytics	42
19. Machine Learning and Intelligent Systems	(((((“Artificial intelligence") OR "Machine Intelligence") OR "Intelligent systems") OR "Machine Learning") OR "Supervised learning") OR "Reinforcement learning"	13,199
20. Life Sciences	(((((((Bioinformatic) OR Biology) OR Biomedical) OR Biometrics) OR (Health) OR "Health care") OR Healthcare) OR “Life Sciences”) OR Medical) OR Medicine	28,510
21. Computational Biology and Bioinformatics	((((“Computational Biology") OR Bioinformatics) OR "Structural bioinformatics") OR Phylogenetics and evolutionary modeling) OR Phylogenetics	2,145
22. Robotics	(Robotics) OR Robot	8,817

Xplore Boolean searches were performed for each technology's keyword set using the following parameters:

Search for: metadata only

Publisher: IEEE only [inclusive of its S/Cs]

Content type: Journals and Magazines

Publication years: 2000 through 2013 [for currency as well as replication ability by excluding the newest 2014 content which is added daily]

Various combinations of the search terms were assessed to determine the keyword set best identifying the subject matter of interest. The search process for each technology is replicable and documentation is available upon request detailing: the search terms of interest, the number of search hits for the various combinations of keywords tried, the final Boolean aggregate search term chosen (see above table), and the Xplore URL for the search results.

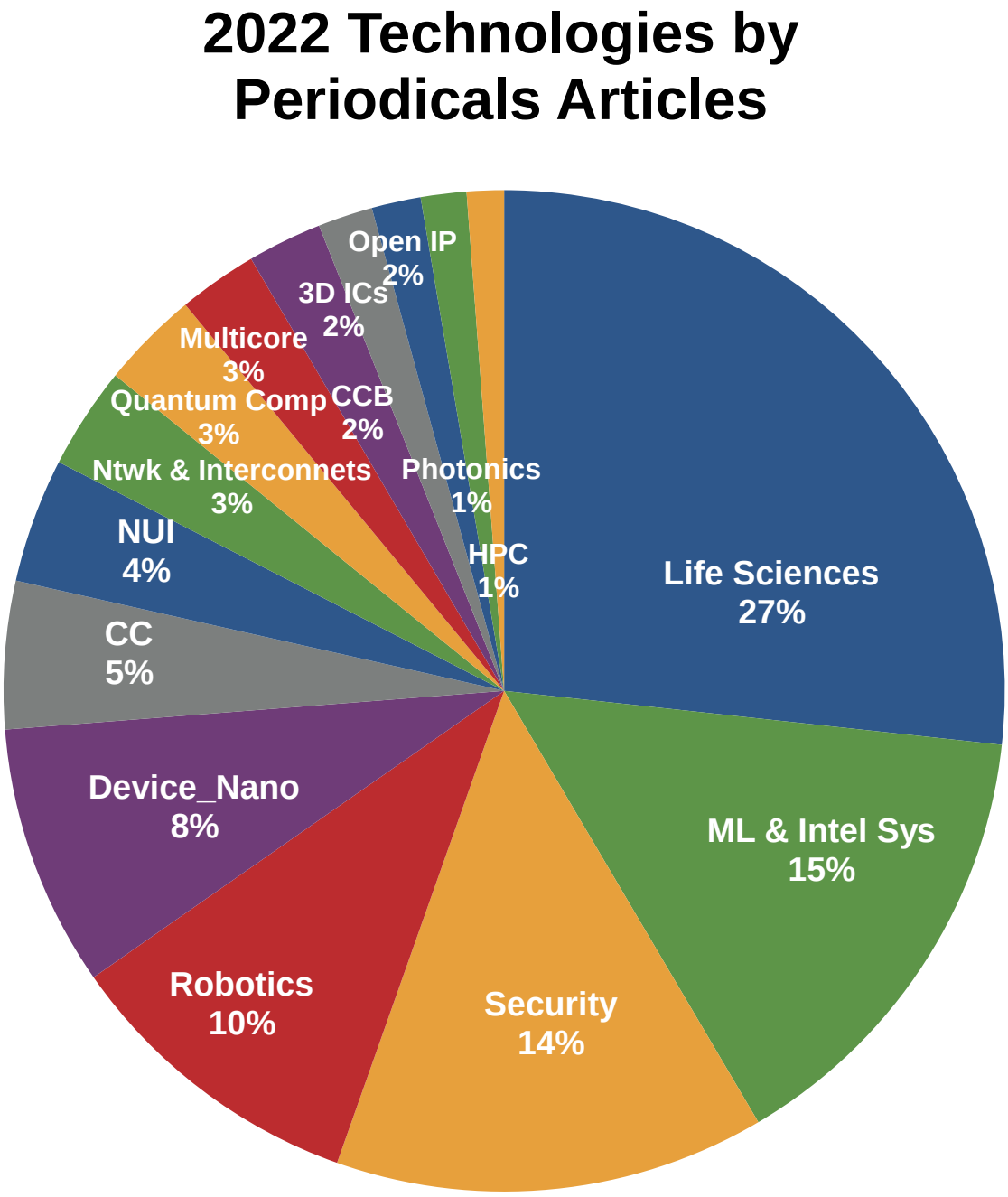
The Xplore search results were exported* for further analysis. The sponsoring S/Cs for each article's periodical were identified, except in the case of S/Cs with only one or two articles published (they were also assigned to the "Other" category). The number of articles published by each sponsoring S/C, across all of the titles with relevant content, was then tabulated. Finally, if an S/C had a total number of articles too low to be shown on a technology pie chart, its total was also added to the "Other" category.

In the case of co-sponsoring partner S/Cs, each S/C received an equal pro rata share of the aggregate article count for its articles. For example, if a co-sponsoring partnership of three S/Cs had published 24 articles of content relevant to the technology, each of the three S/Cs would be assigned a share of eight articles. An exception was made when the count was less than five articles for that co-sponsoring group of S/C, to avoid negligible per-S/C counts.

[*Note: Xplore is currently limited to 2,000 records for its export ability. A supplemental process was used for technologies' searches resulting in more than 2,000 articles. The periodicals expected to have a significant number of articles published in the technology were identified until all periodicals with significant involvement were ranked. The article count estimates were then assigned to the sponsoring S/Cs of those titles.]

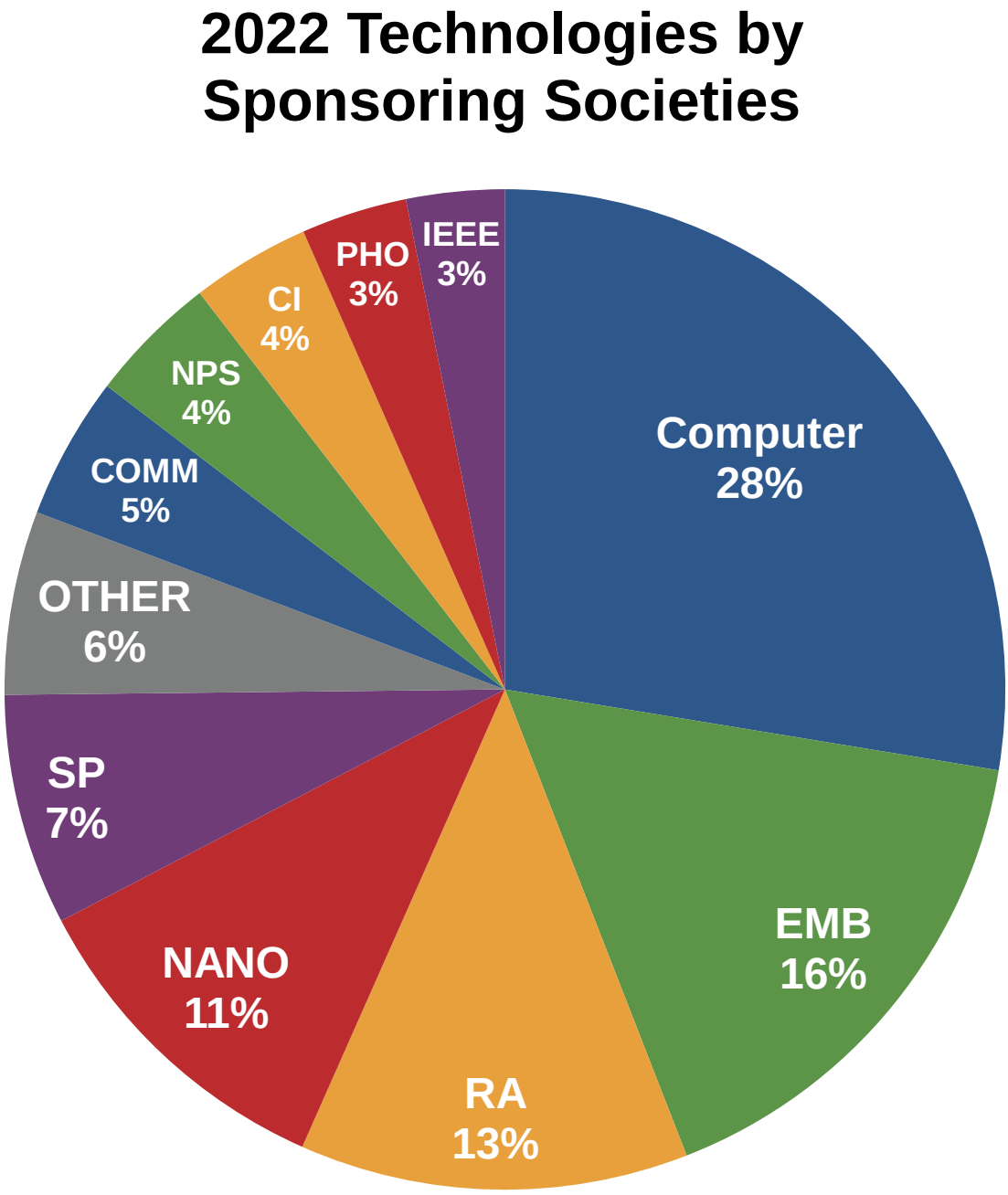
The following pages consist of:

- Summary by technology
- Summary by S/C
- Sponsoring S/C for each technology, in detail
- Tabulation of article counts by technology and S/C



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

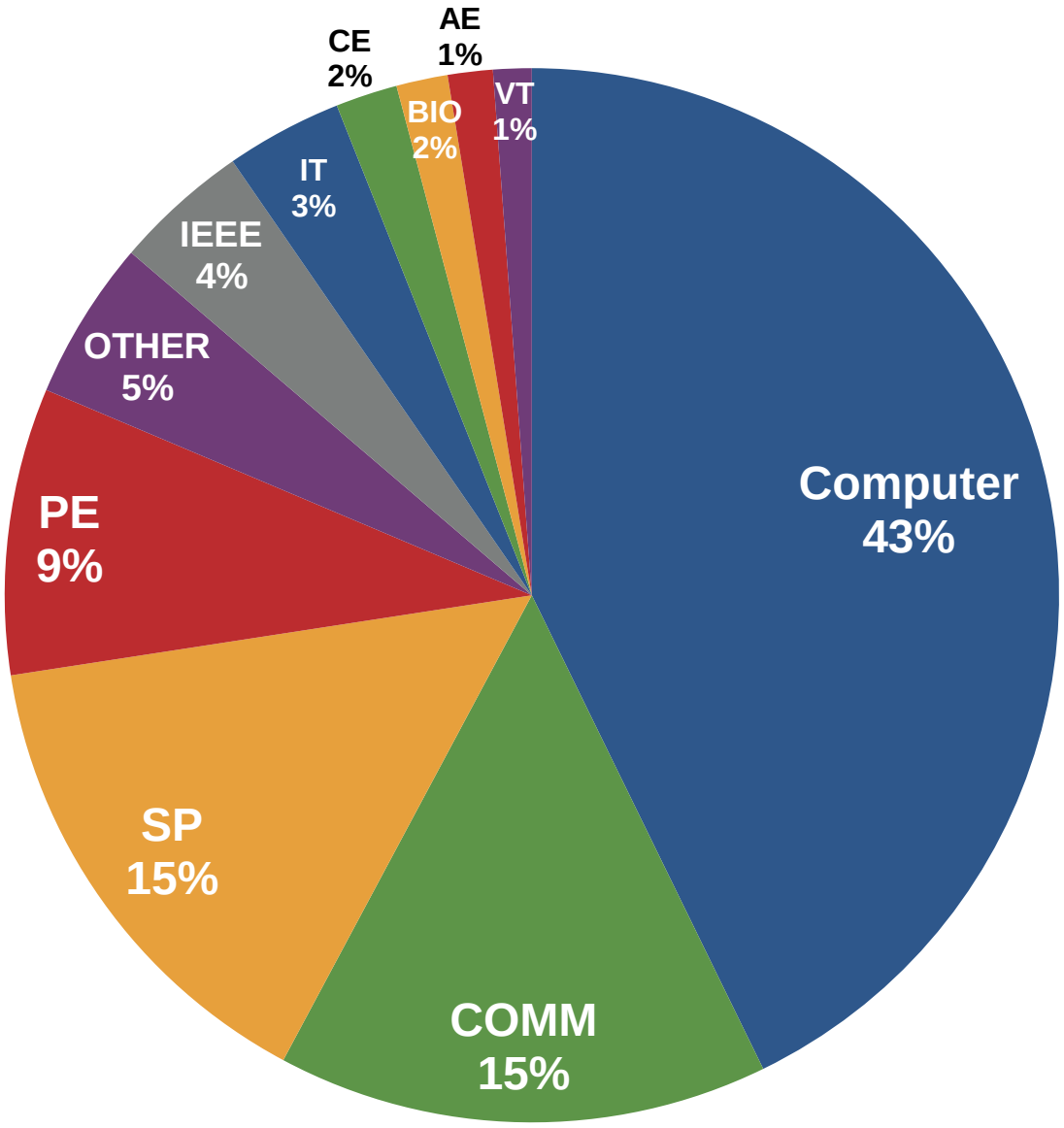
Figure 25. The breakdown of 22 technologies by periodical articles.



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 26. The breakdown of 22 technologies by sponsoring societies.

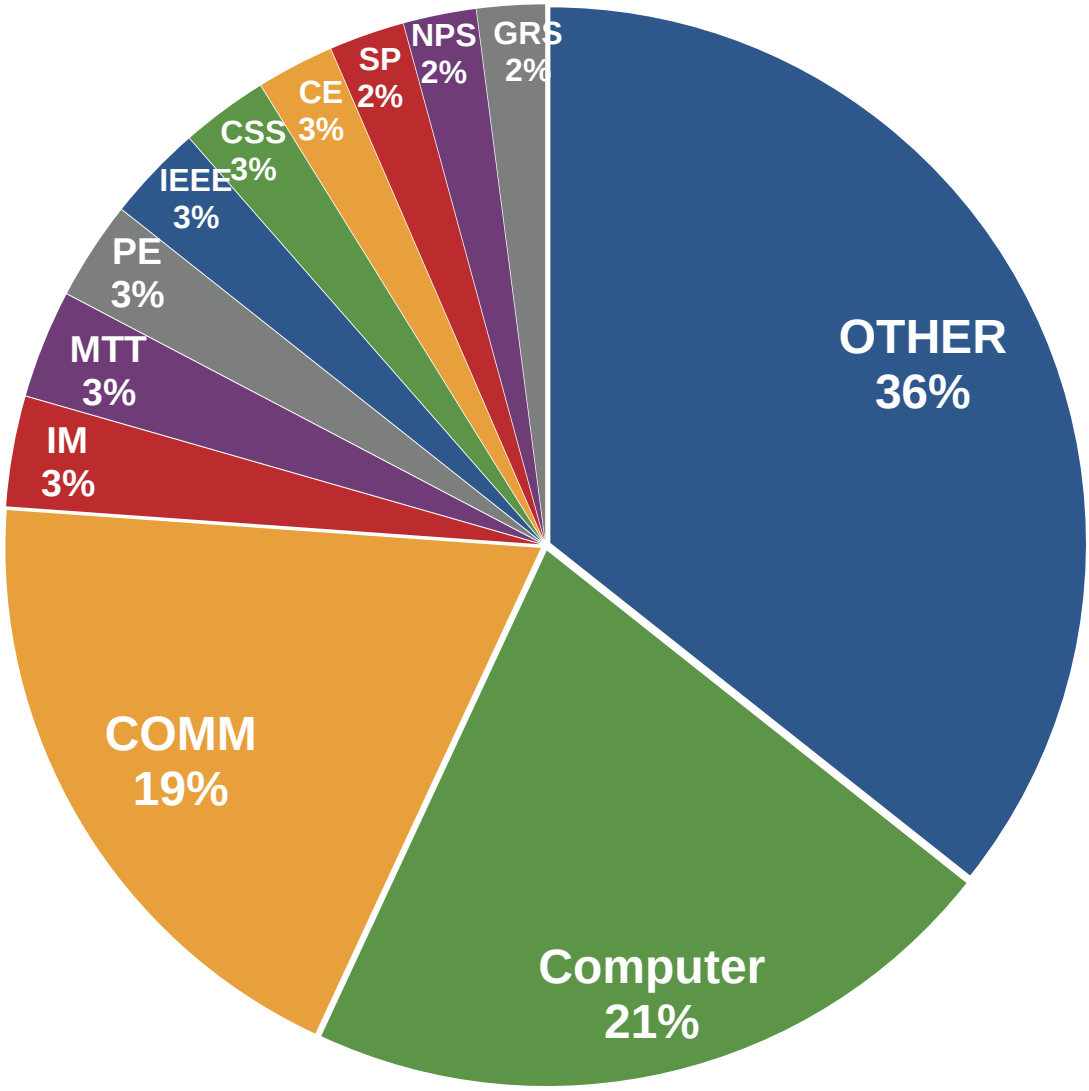
1 Security Cross-cutting Issues Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 27. The breakdown of security cross-cutting issues by sponsoring societies.

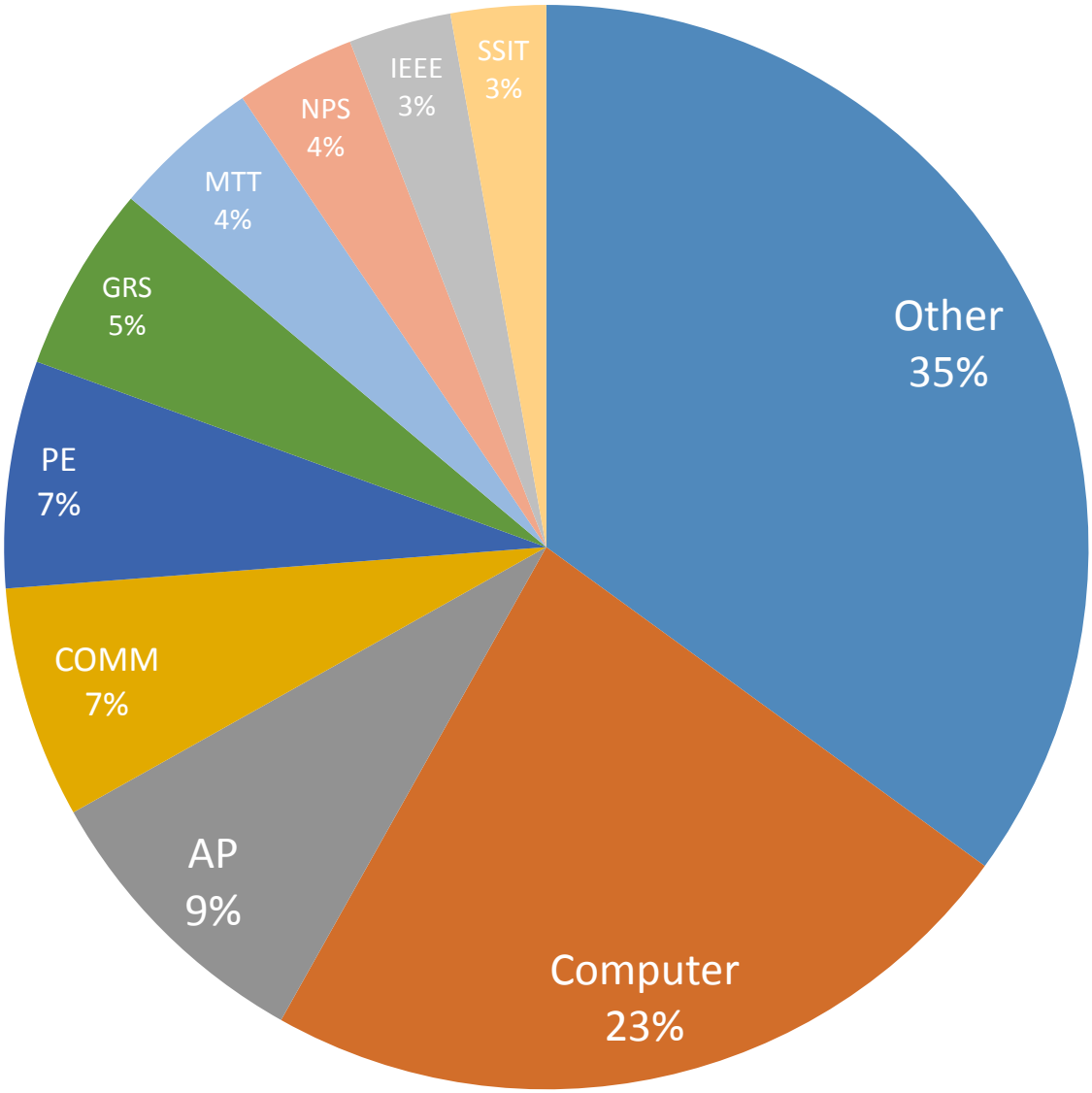
2 Open Intellectual Property Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 28. The breakdown of open intellectual property by sponsoring societies.

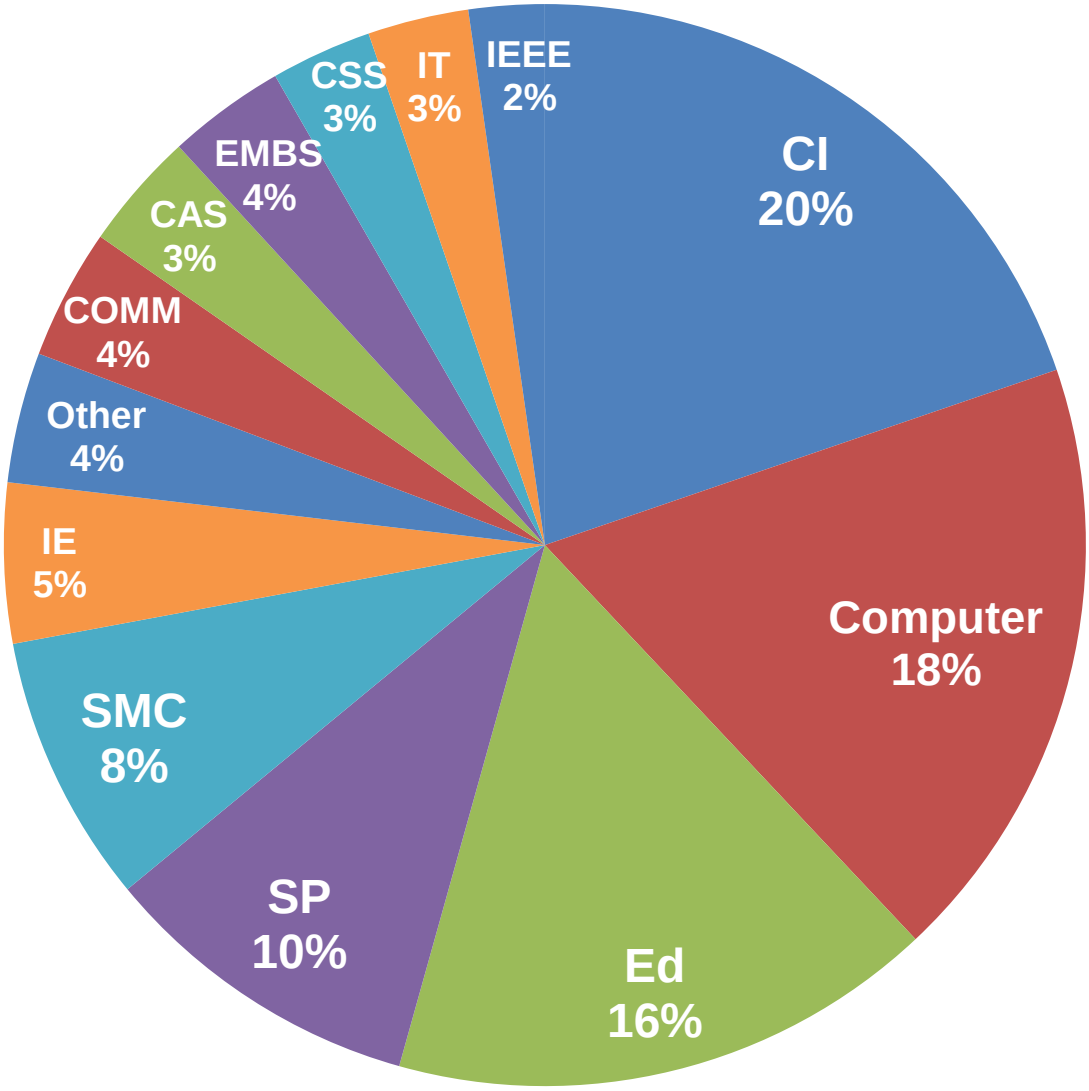
3 Sustainability Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

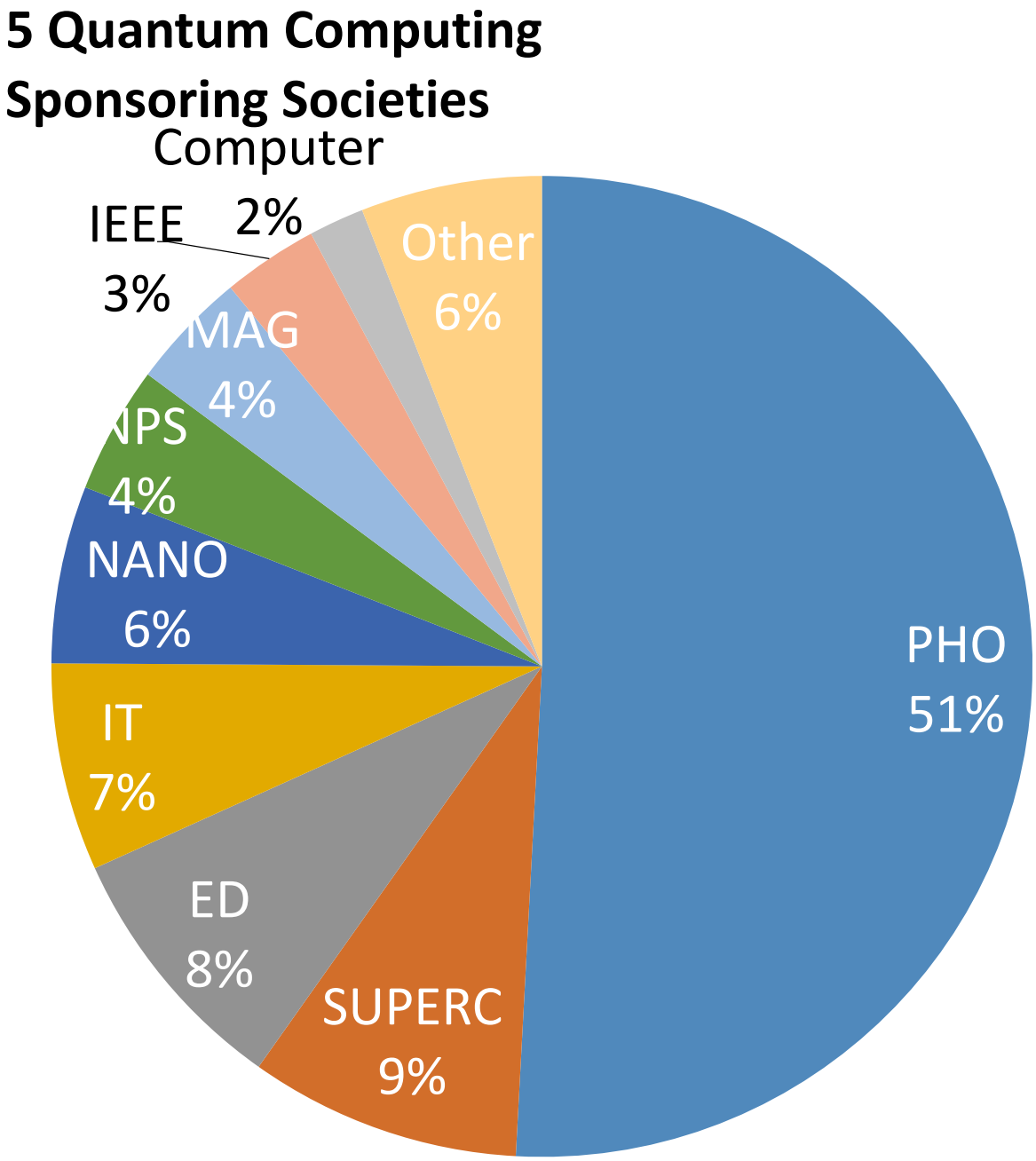
Figure 29. The breakdown of sustainability by sponsoring societies.

4 MOOC Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

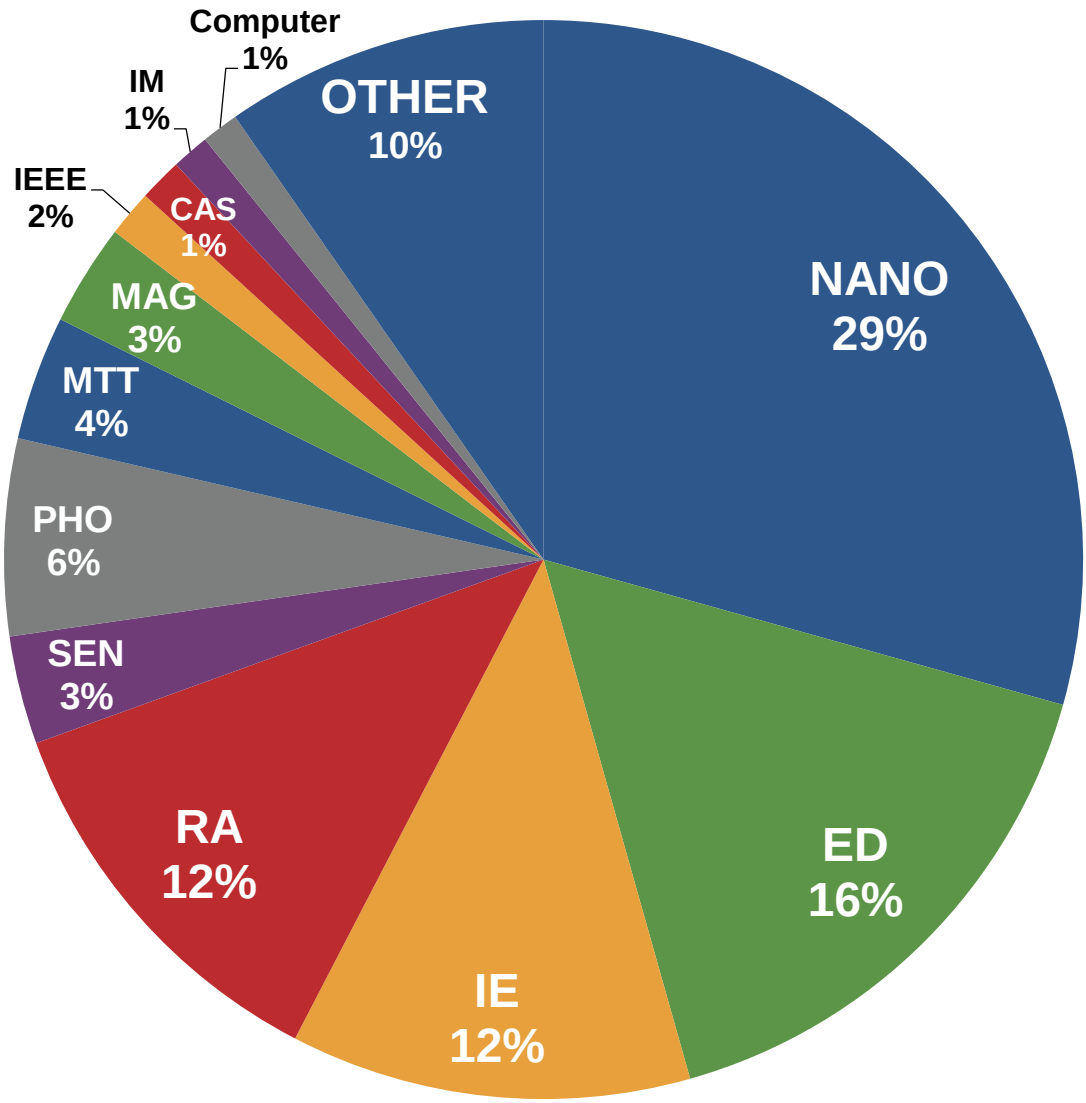
Figure 30. The breakdown of MOOC by sponsoring societies.



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 31. The breakdown of quantum computing by sponsoring societies.

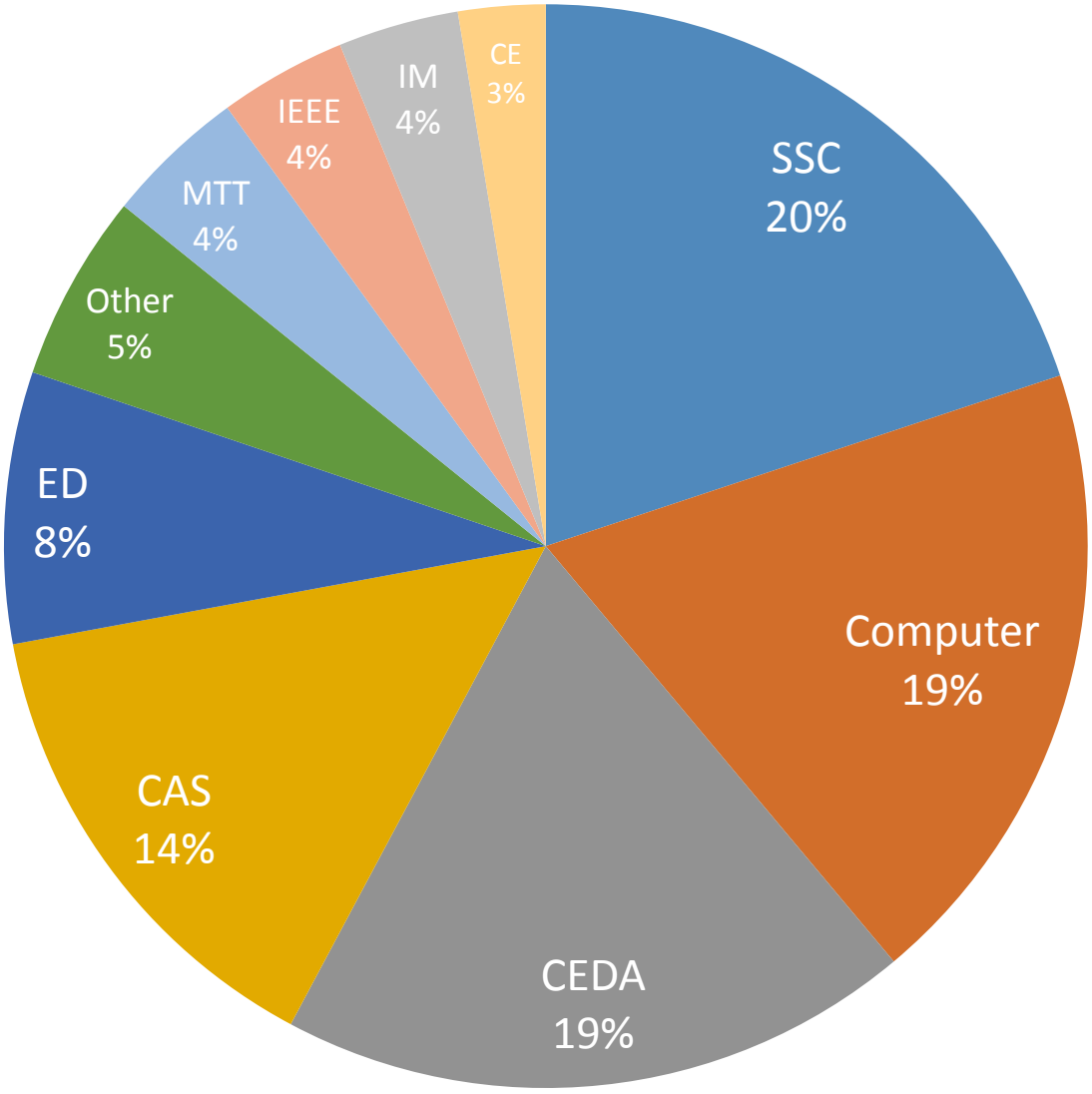
6 Device and Nano-tech. Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 32. The breakdown of device and nano-technology by sponsoring societies.

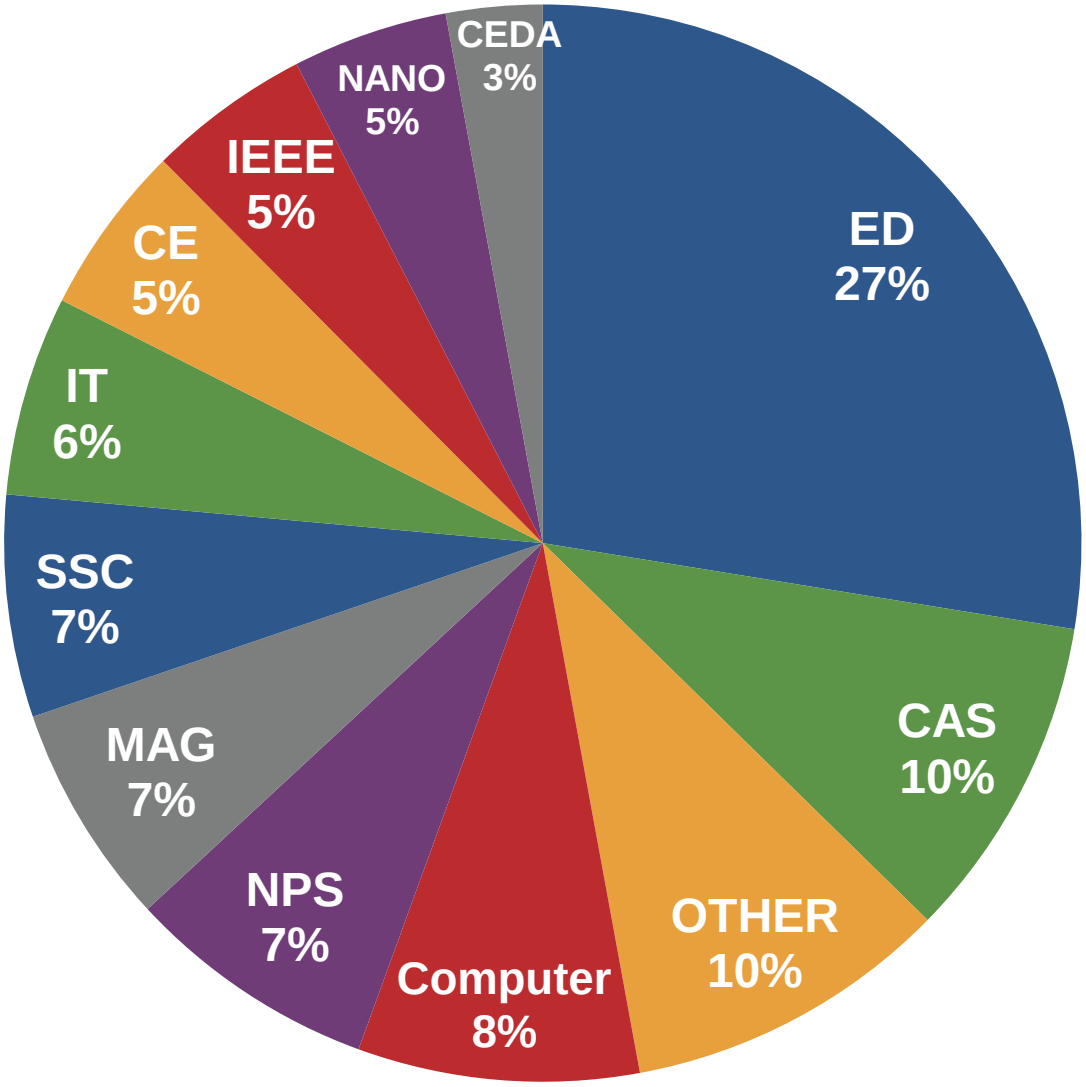
7 3D Integrated Circuits Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

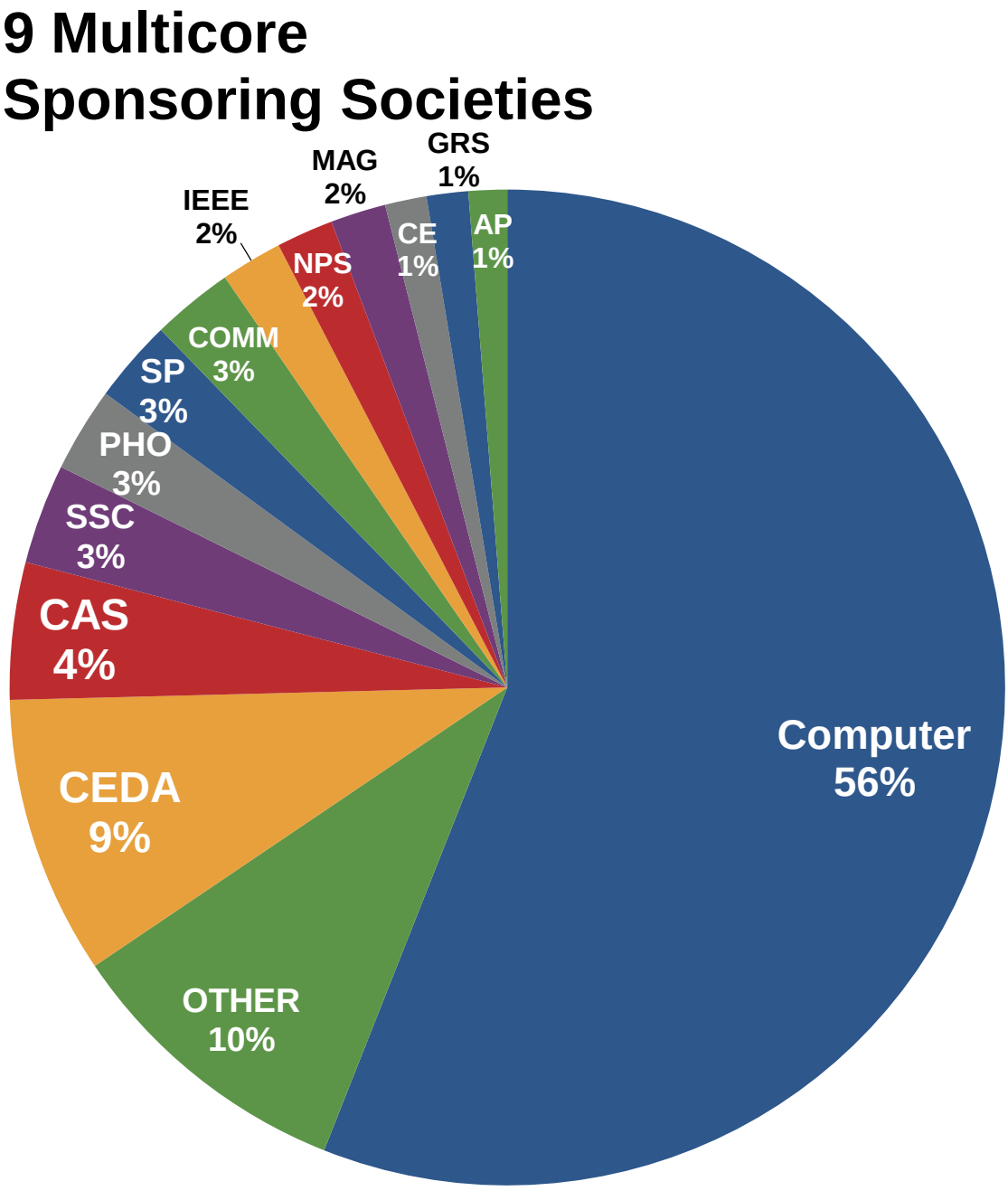
Figure 33. The breakdown of 3D integrated circuits by sponsoring societies.

8 Universal Memory Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

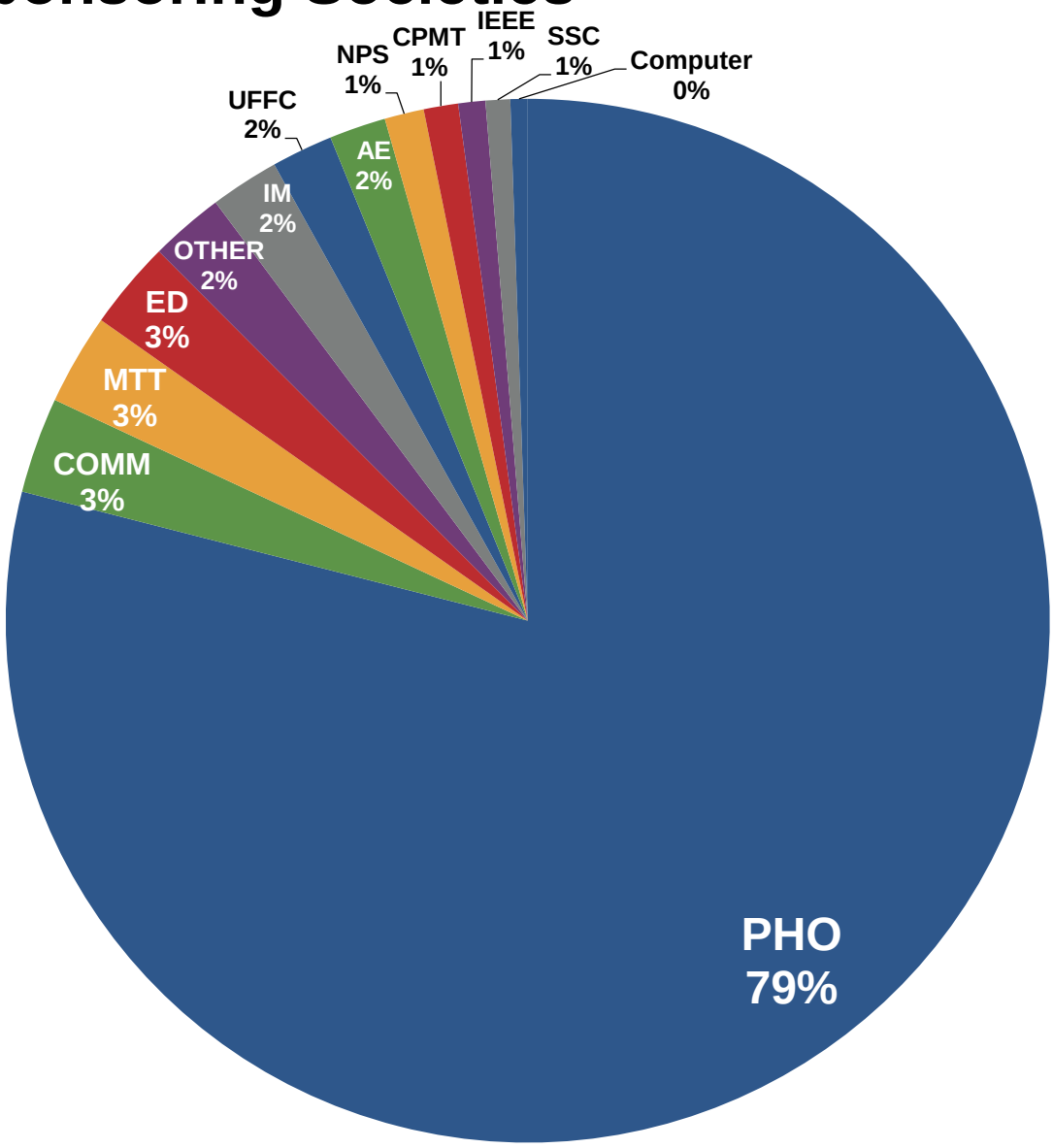
Figure 34. The breakdown of universal memory by sponsoring societies.



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 35. The breakdown of muticore by sponsoring societies.

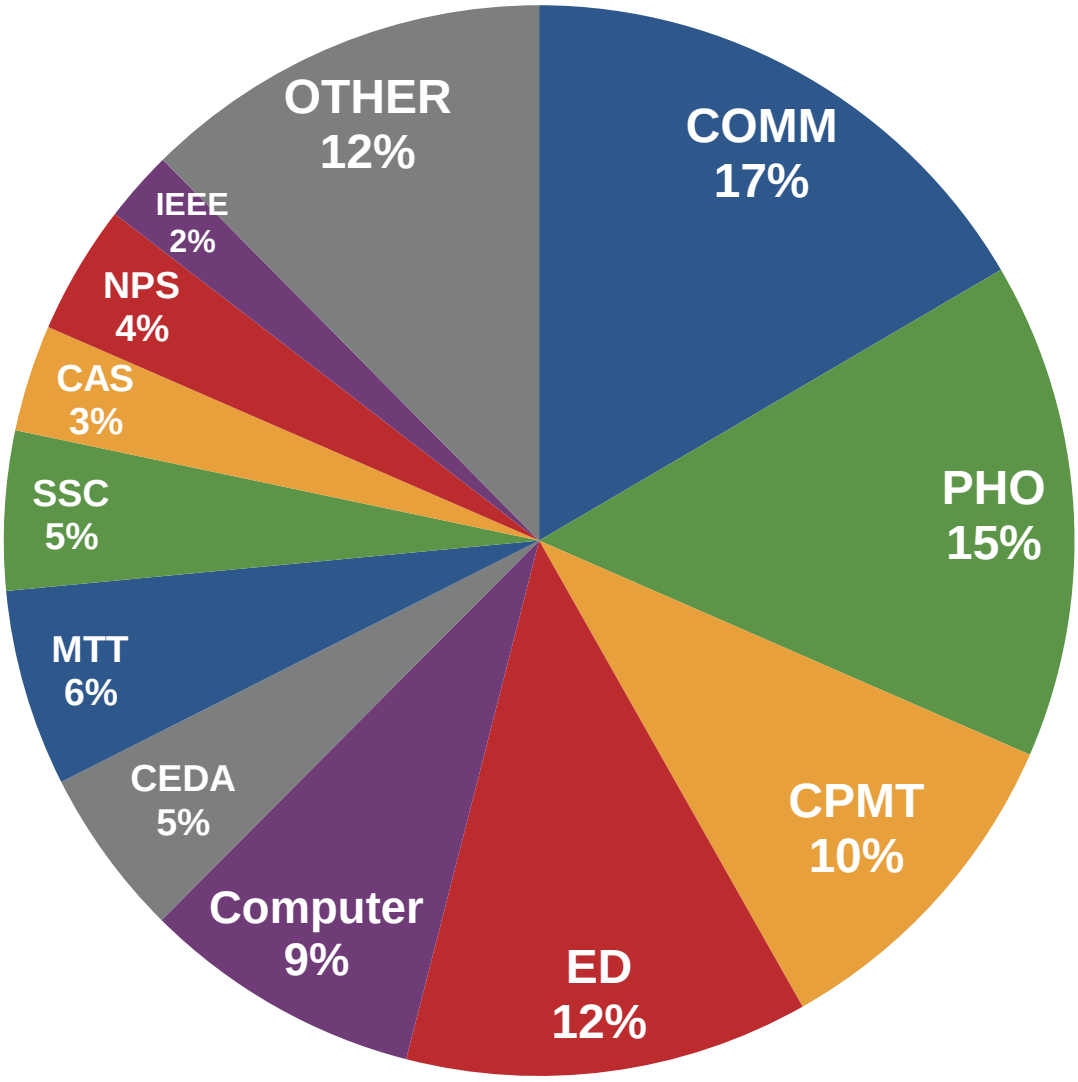
10 Photonics Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 36. The breakdown of photonics by sponsoring societies.

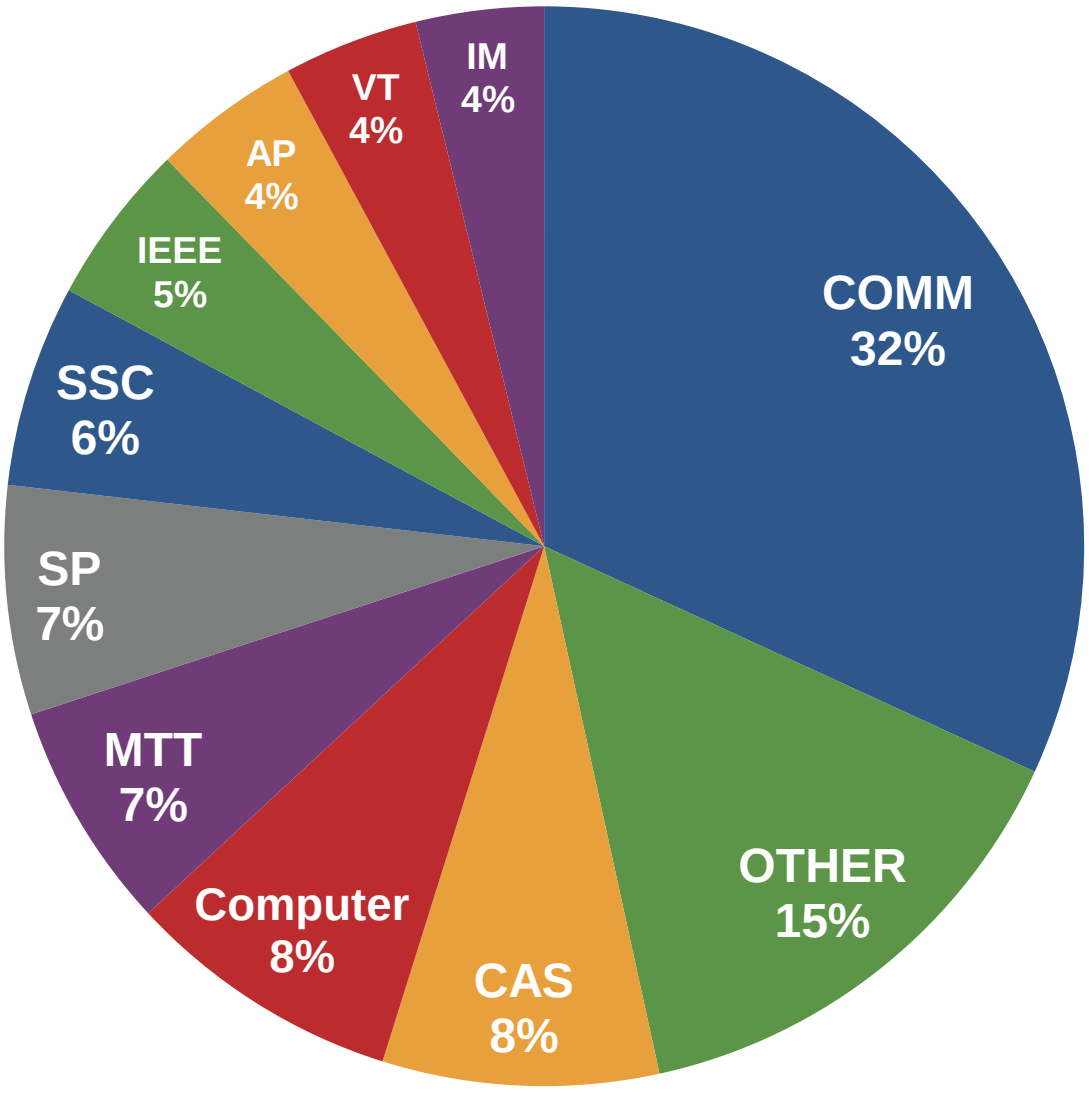
11 Networking & Interconnectivity Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

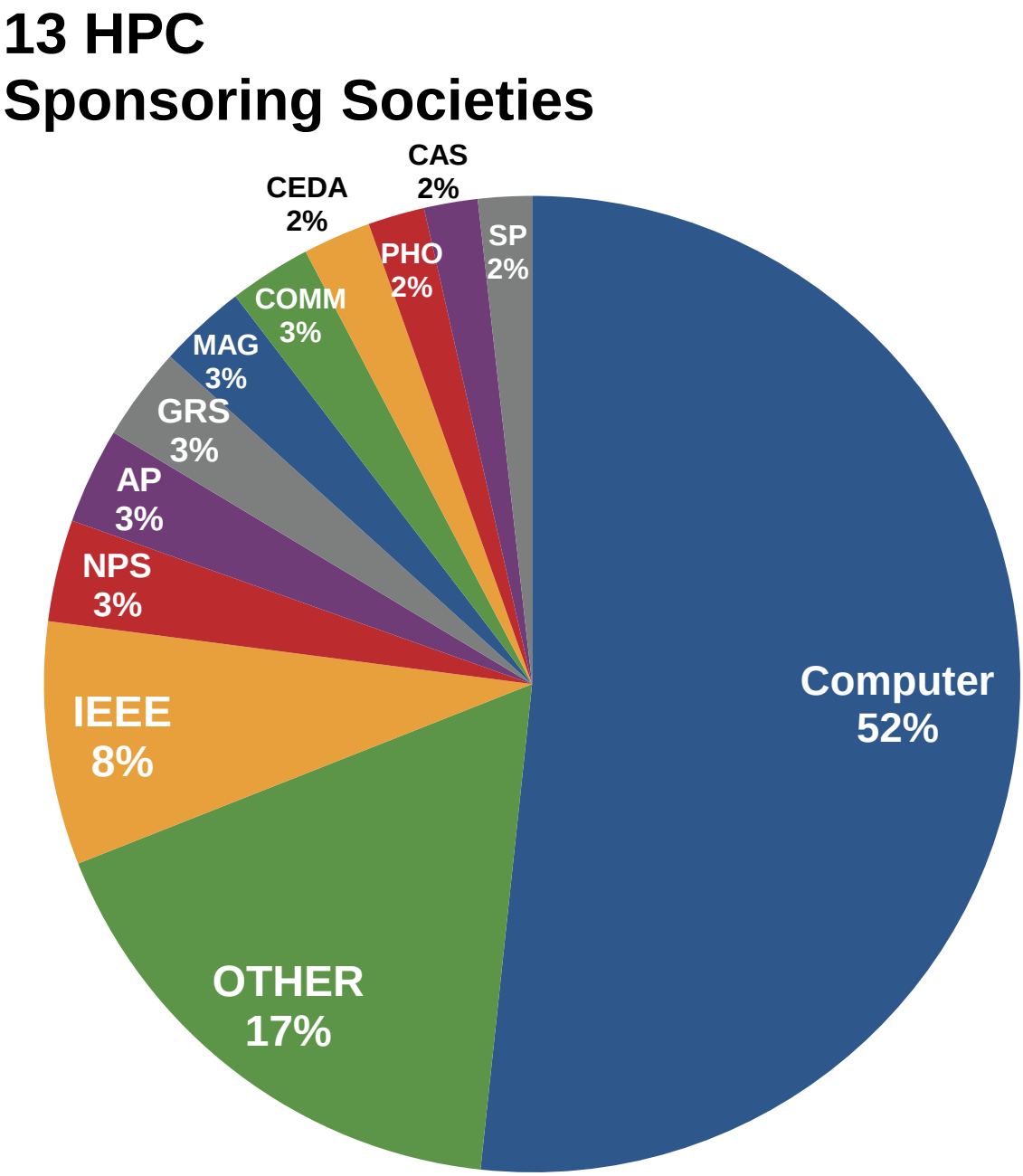
Figure 37. The breakdown of networking and interconnectivity by sponsoring societies.

12 Software Defined Networks Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

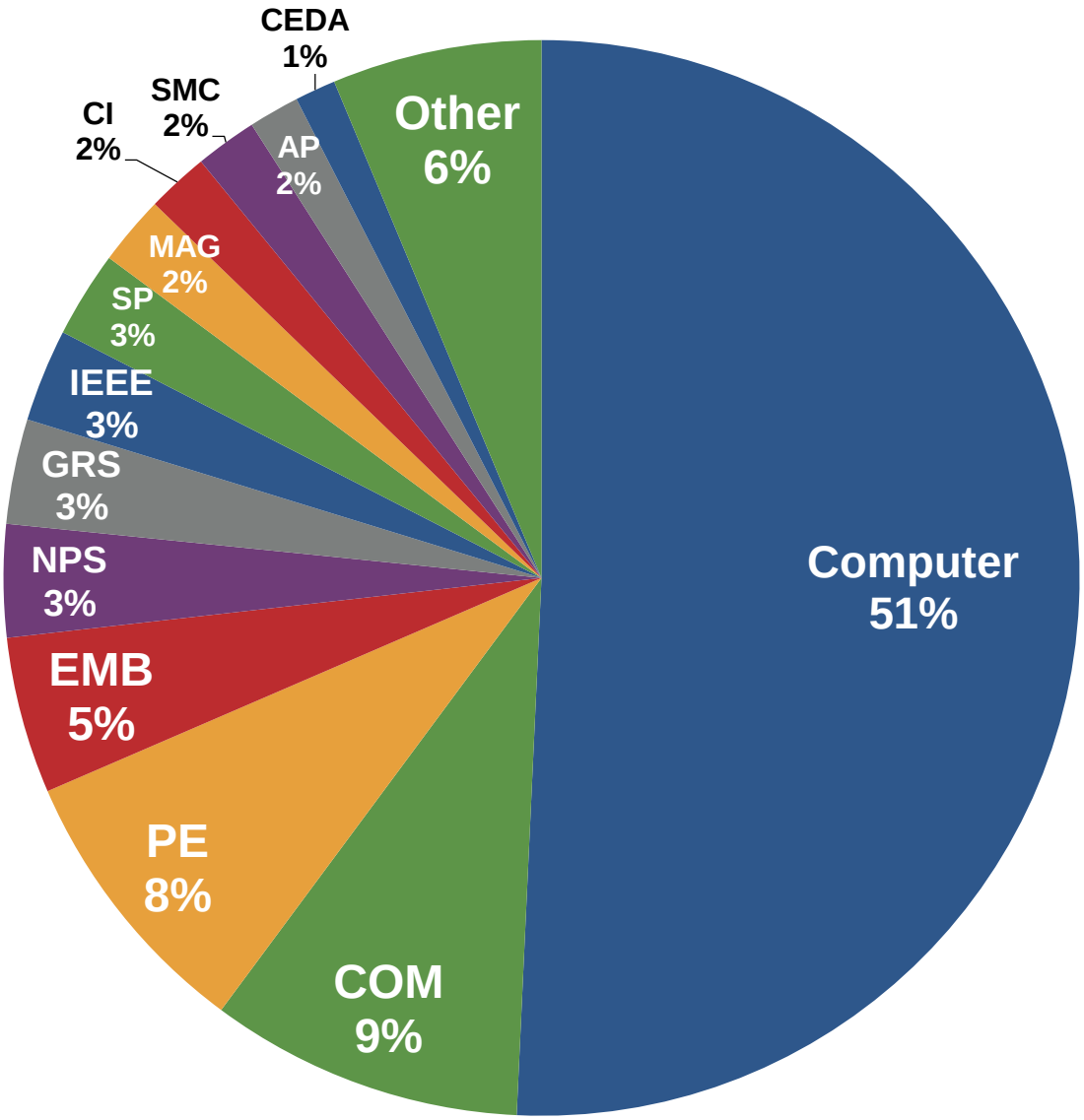
Figure 38. The breakdown of software-defined networks by sponsoring societies.



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

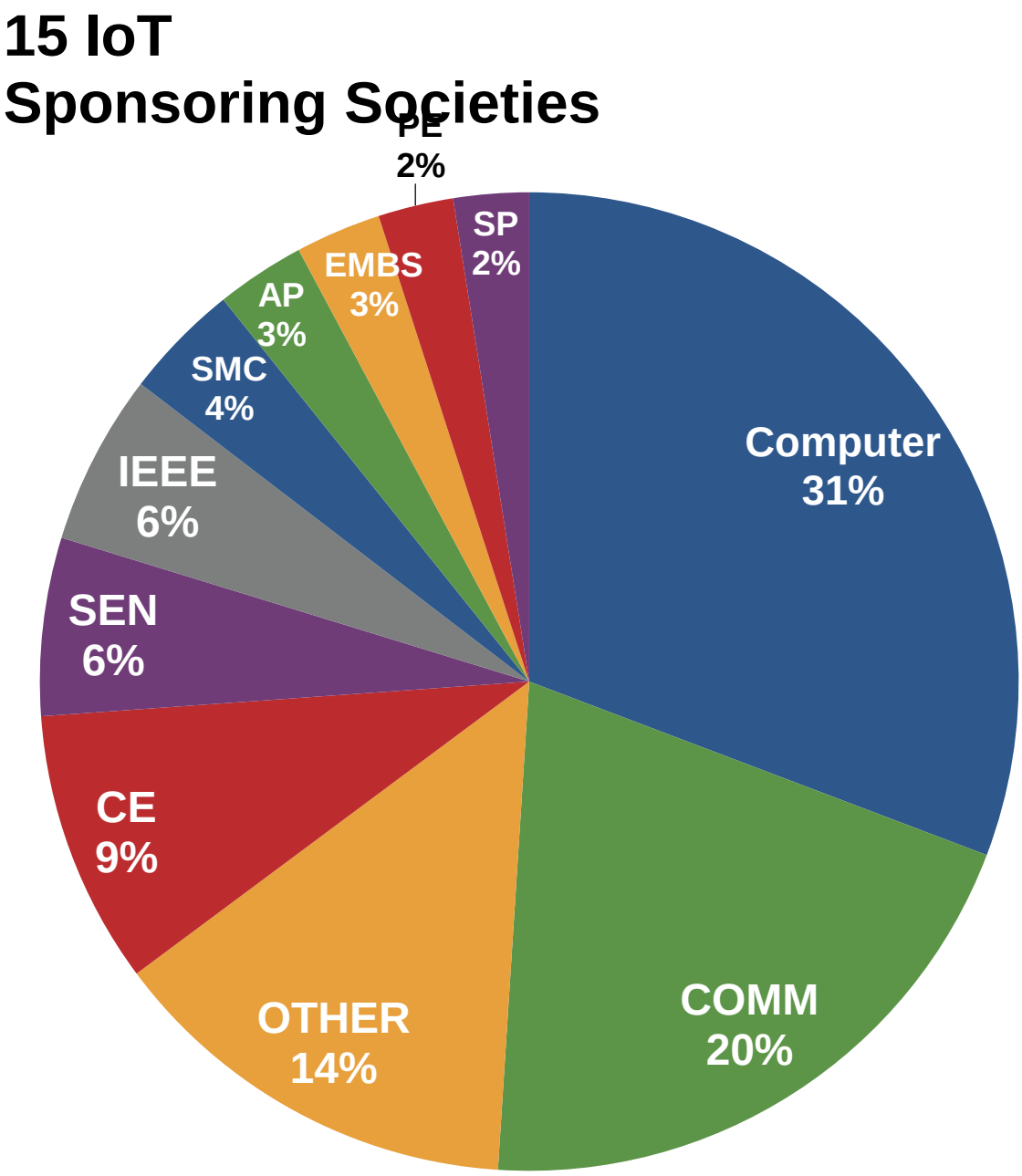
Figure 39. The breakdown of HPC by sponsoring societies.

14 Cloud Computing Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

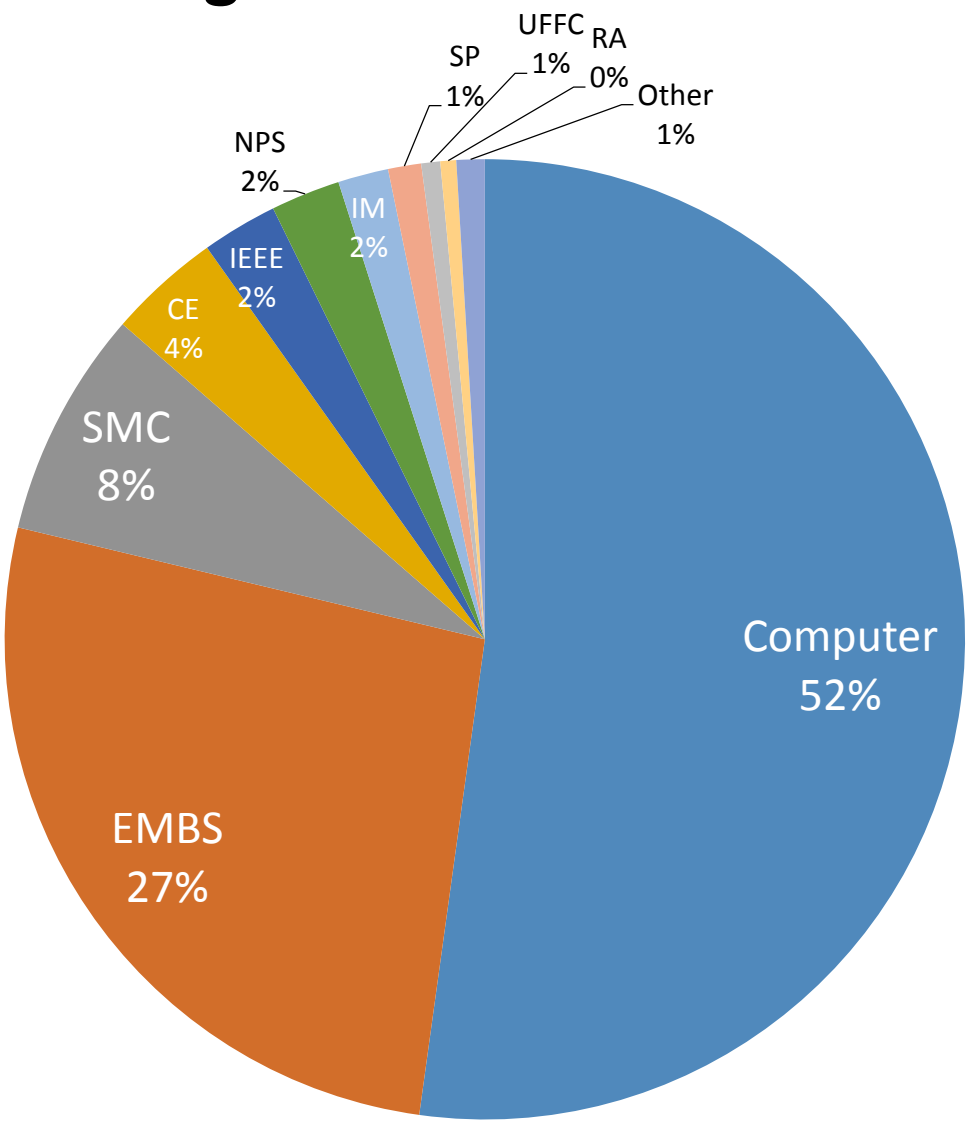
Figure 40. The breakdown of cloud computing by sponsoring societies.



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 41. The breakdown of IoT by sponsoring societies.

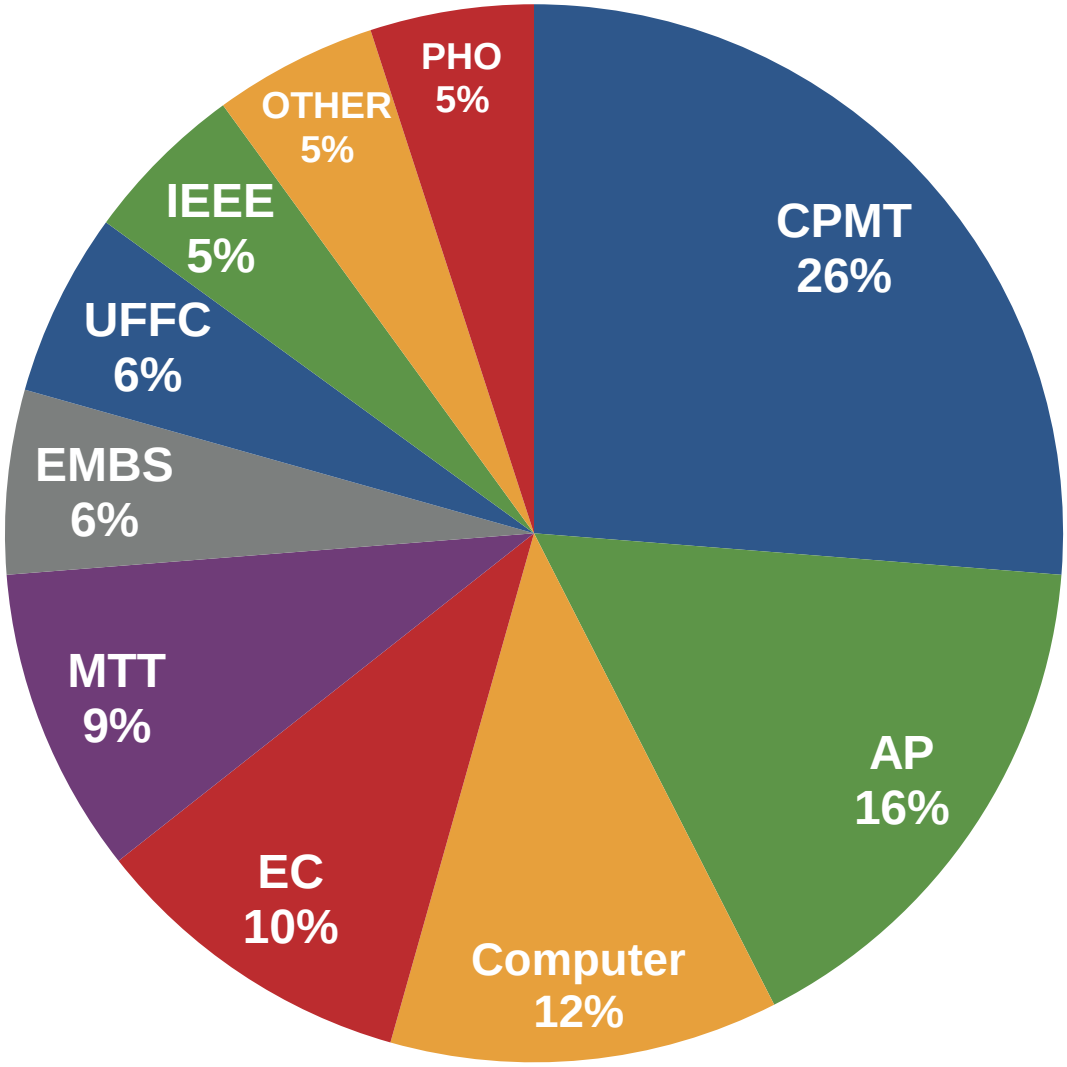
16 Natural User Interfaces Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 42. The breakdown of natural user interfaces by sponsoring societies.

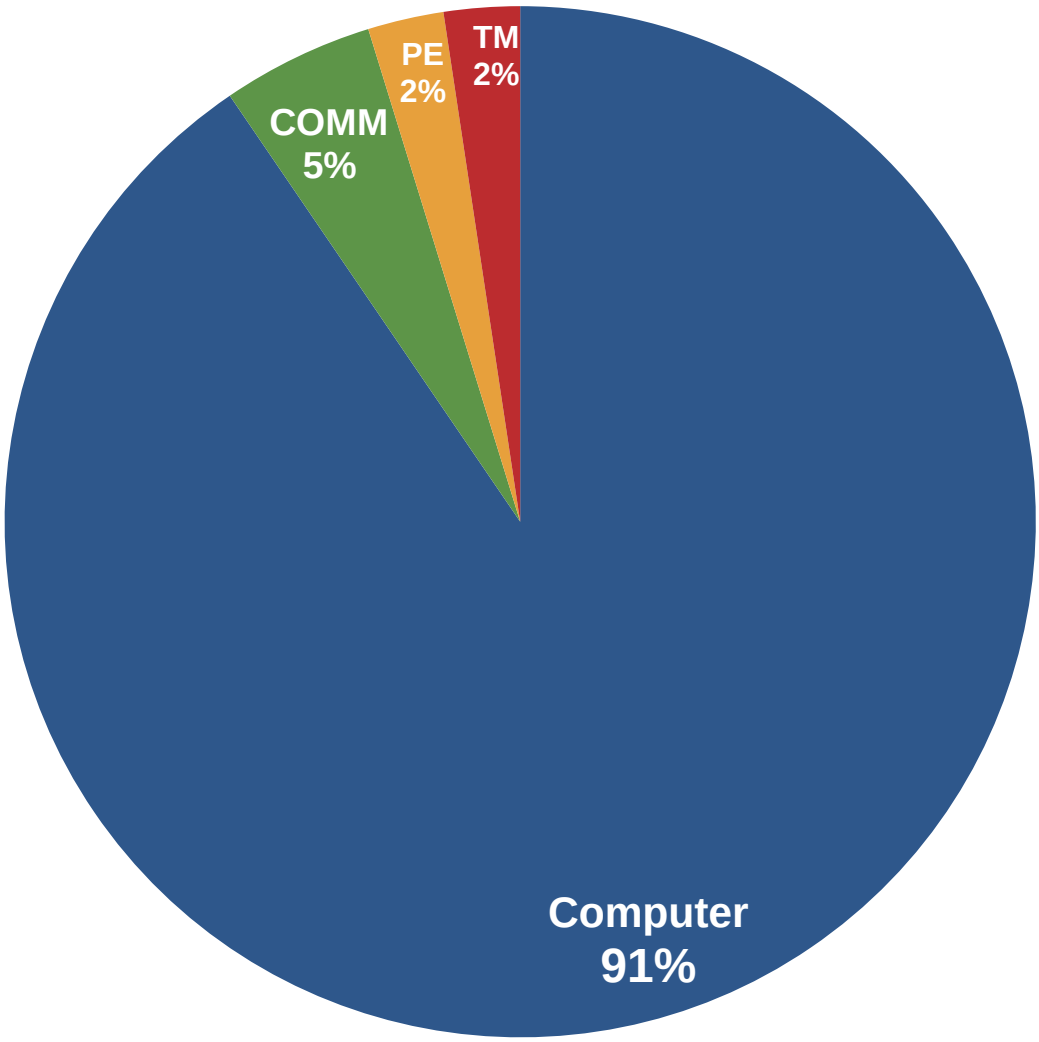
17 3D Printing Sponsoring Societies



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

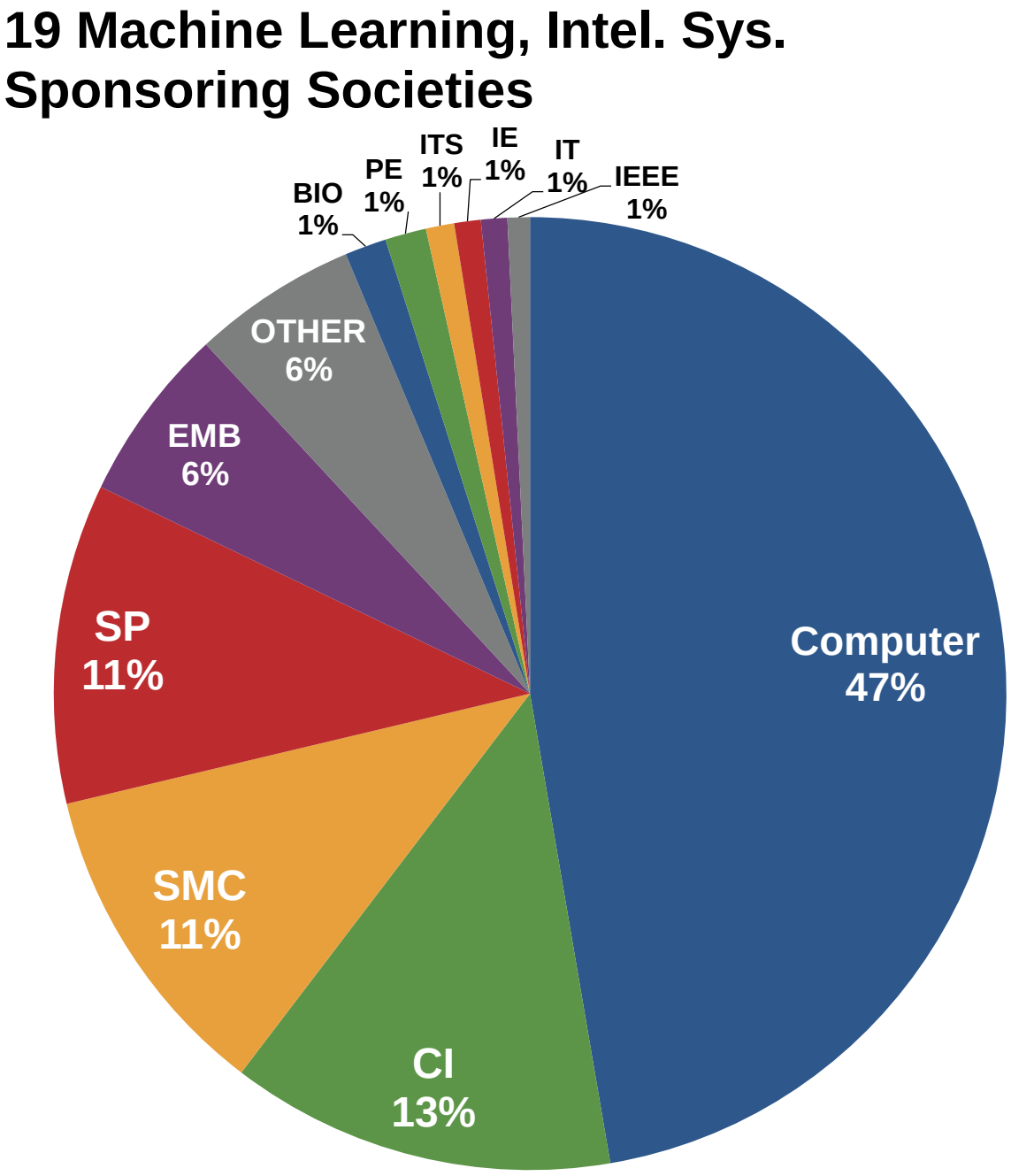
Figure 43. The breakdown of 3D printing by sponsoring societies.

18 Big Data Analytics Sponsoring Societies



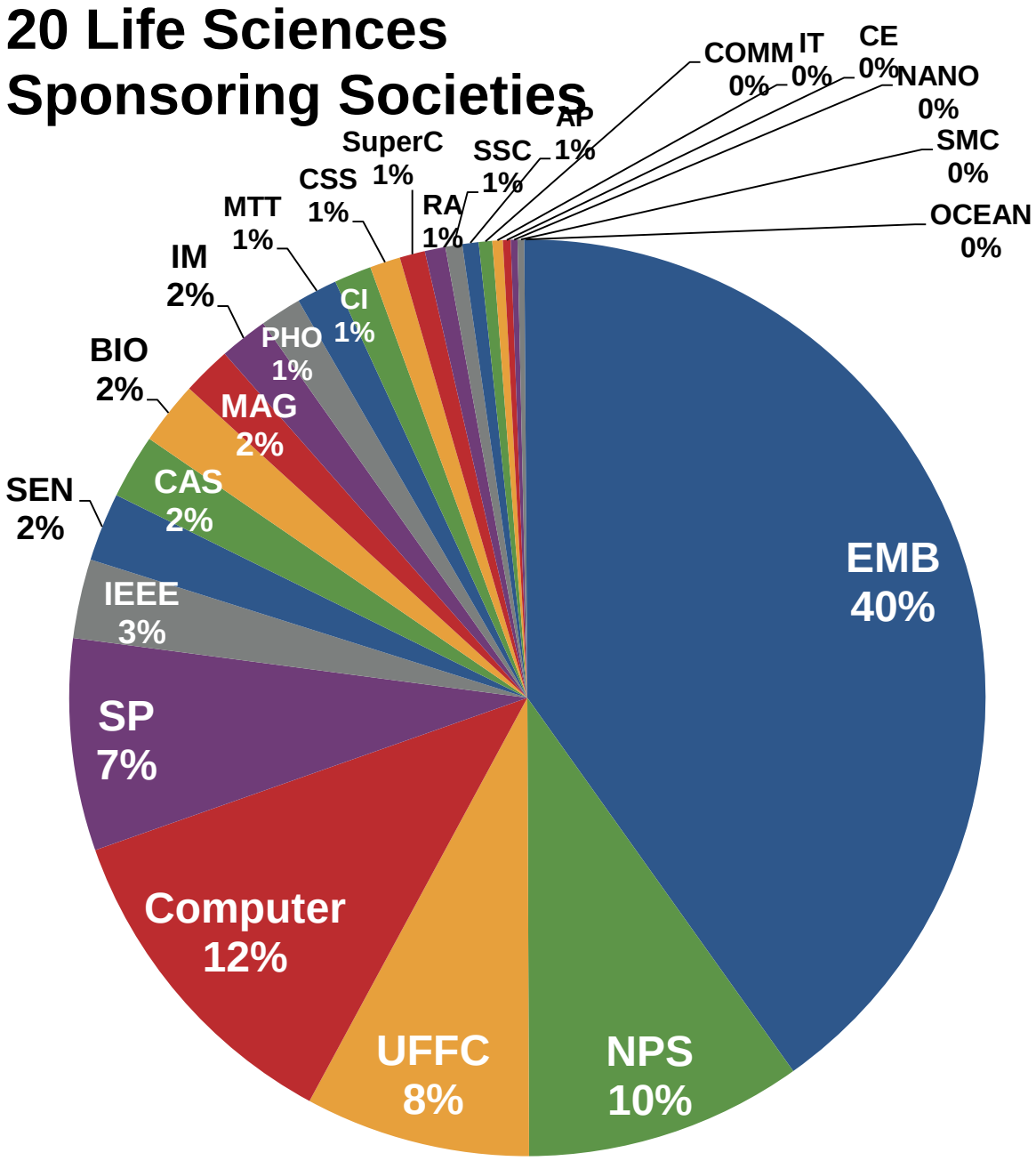
AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 44. The breakdown of big data analytics by sponsoring societies.



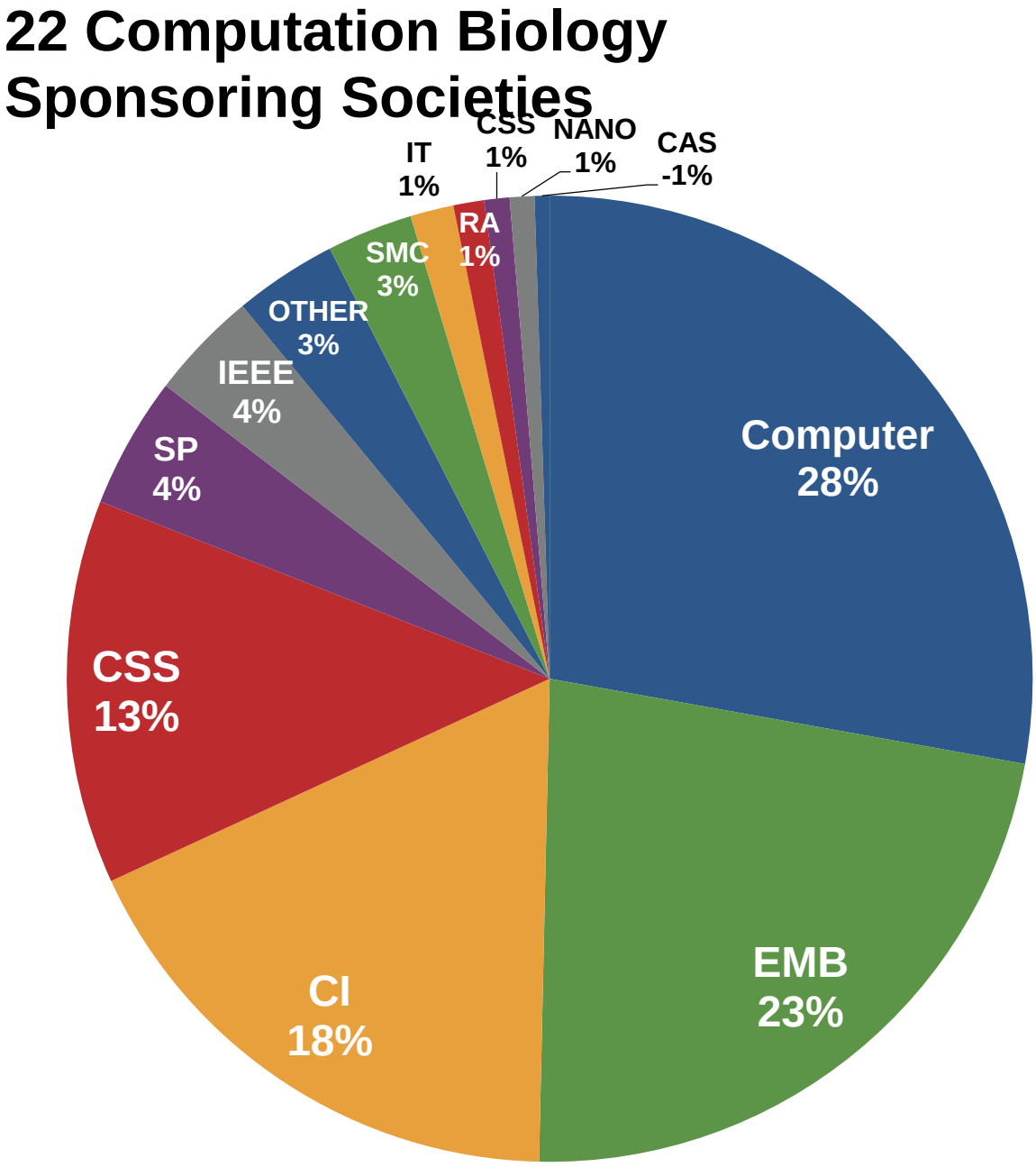
AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 45. The breakdown of machine learning and intelligent systems by sponsoring societies.



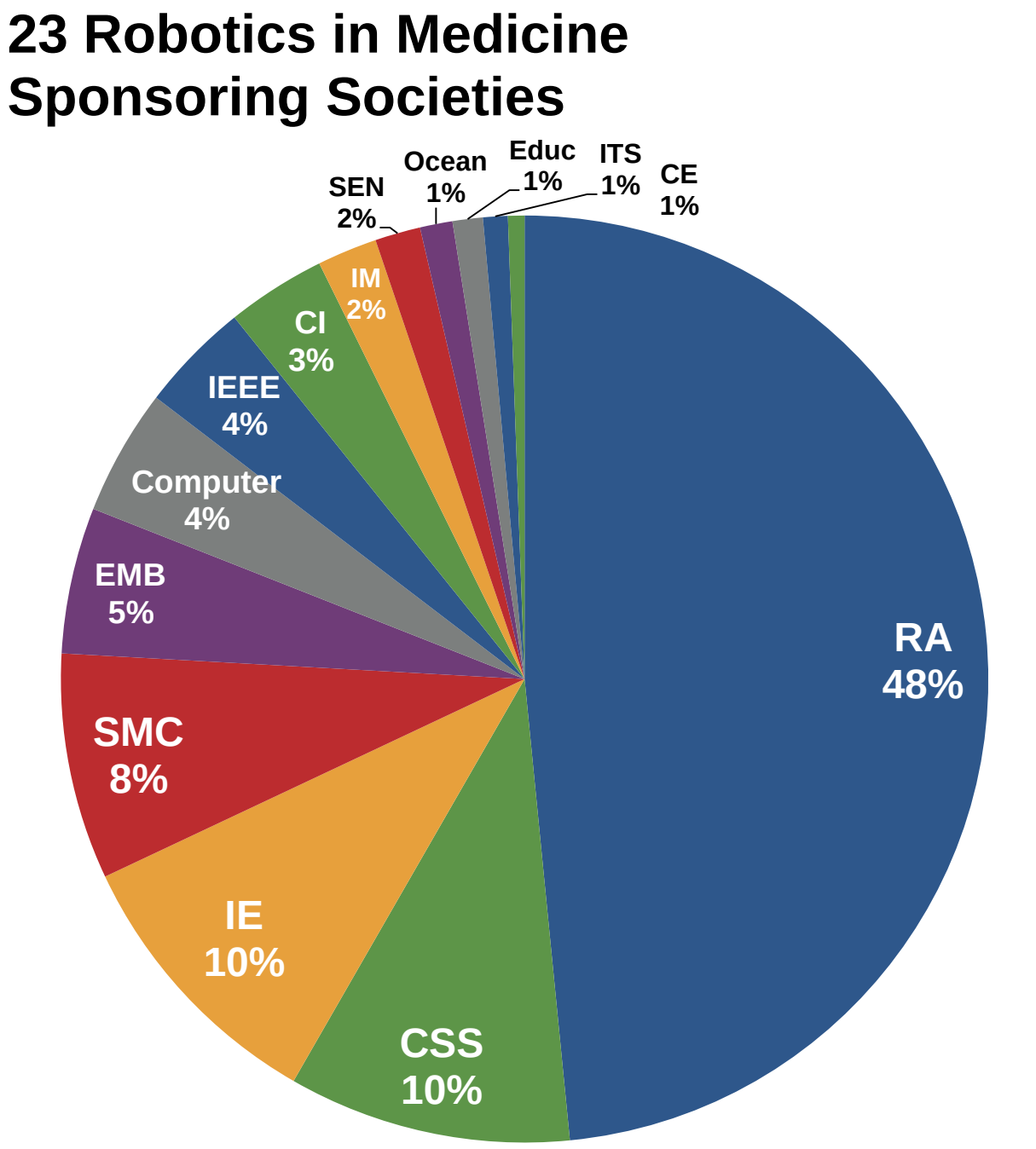
AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 46. The breakdown of life sciences by sponsoring societies.



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 47. The breakdown of computational biology by sponsoring societies.



AE=Aerospace & Electronic Systems AP=Antennas & Propagation BIO=Biometrics Council CAS=Circuits & Systems CE=Consumer Electronics CEDA=Council on Electronic Design Automation CI=Computational Intelligence COMM=Communications CPMT=Components, Packaging, & Manufacturing Technology CSS=Control Systems EC=Electromagnetic Compatibility Ed=Education ED=Electron Devices EMB=Engineering in Medicine & Biology GRS=Geoscience & Remote Sensing IE=Industrial Electronics IM=Instrumentation & Measurement IT=Information Theory ITS=Intelligent Transportation Systems MAG=Magnetics MTT=Microwave Theory & Techniques NANO=Nanotechnology Council NPS=Nuclear & Plasma Sciences OCEAN=Oceanic Engineering PE=Power & Energy PHO=Photonics RA=Robotics & Automation SEN=Sensors Council SMC=Systems, Man, & Cybernetics SP=Signal Processing SSC=Solid-State Circuits SSIT=Social Implications of Technology SUPERC=Council on Superconductivity TM=Technology Management Council UFFC=Ultrasonics, Ferroelectrics, and Frequency Control VT=Vehicular Technology

Figure 48. The breakdown of robotics in medicine by sponsoring societies.

We also did the search on Google Scholar and plan to do similar effort on MS Academic Research. We present the former in the table below.

Table 5. Google and IEEE Xplore search results combined.

Technology	Boolean search query	# Xplore articles	Google Scholar (K)	Google /IEEE ratio
1. Security Cross-Cutting Issues	((Privacy OR Security OR Intrusion) OR Intrusion OR "Security legislation") OR (((cyber) OR cybersecurity) OR cyber-security) OR "cyber security"	12,389	523.0	9.2
2. Open Intellectual Property Movement	(((“Crowd sourcing”) OR “Open IP”) OR Open AND “Intellectual Property”) OR “Open standards”	1,416	16.2	13.3
3. Sustainability	((((“Energy usage”) AND computing)) OR (“Sustainability”) OR (“Green computing”) OR (“Carbon footprint”) OR (“Earth friendly”)) OR (Green ICT)) OR (Sustainable Computing)	882	11.7	10.7
4. Massively Online Open Courses	((((“Open Courses”) OR (“Massively Online”) NOT "Games") OR “Massively” AND “Courses”) OR "Online learning") OR "Automated grading"	458	0.4	0.9
5. Quantum Computing	“Quantum Computing”) OR (“Quantum” AND “mechanical phenomena”) OR (“Quantum properties”) OR (“Quantum annealing”) OR (“factorization algorithm” OR “Shor”) OR (“Qubit”	2,823	26.1	1.3
6. Device and Nano-technology	(((“Microelectromechanical systems”) OR “Nano-technology” OR “Nanotechnology” OR “Nano technology”) OR “Microelectromechanical systems”) OR “Micro machine” OR “Micro machines” OR “Micromachines” OR “Micromachine” OR “Micro-machine” OR “Micro-machines”	7,546	335.0	5.6
7. 3D Integrated Circuits	((((("2.5D chip" OR "2.5-D chip" OR "2.5D chips" OR "2.5-D chips"))) OR ("3D chip" OR "3-D chip" OR "3D chips" OR "3-D chips")) OR "System on a Chip") OR "System in a Package")	1,579	22.8	14.4
8. Universal Memory	((((“Non-volatile memory”) OR Memristor) OR “Spin Transfer Torque” RAM) OR "Phase Change Memory") OR “Universal Memory”	460	10.4	44.4

9. Multicore	(((((("Multicore") OR "Multiprocessor") OR GPU) OR ("Accelerators") AND "processor")) OR GPGPU) OR "Manycore"	2,276	72.6	15.8
10. Photonics	((“Photonics interconnect”) AND “Silicon photonics”) OR VCSEL OR “Vertical Cavity Surface Emitting Laser”)	1,313	20.0	22.6
11. Networking and Inter-connectivity	((("Interconnects") OR (((“Inter-connectivity") OR “Interconnectivity”) OR “Inter connectivity”) AND Networking) OR (“Ethernet”) AND "internet") OR "Ethernet" AND Networking	2,939	16.6	221.4
12. Software Defined Networks	(((((("Software Defined Networks") OR "Software defined networking") OR "Index Terms":SDN) OR OpenFlow) OR "Software radio") OR "Active networking") OR "Virtual Local Area Networks") OR VLAN	496	5.3	9.1
13. High Performance Computing	((((("High Performance Computing" OR HPC)) OR Supercomputers) OR "Message Passing Interface") OR GPGPU) OR "Compute-intensive") OR Petascale) OR Exascale	1,068	51.1	47.8
14. Cloud Computing	((((((((Cloud Computing) OR “Grid computing”) OR “Cluster computing”) OR Virtualization) OR “-as-a-Service”) OR IaaS) OR PaaS) OR SaaS) OR “Pay as you go”	4,252	521.0	31.9
15. Internet of Things	(((((("Internet of Things") OR "Smart homes") OR Ubiquity) OR Pervasiveness) OR Interconnectivity) OR "Smart dust"	442	262.0	122.5
16. Natural User Interfaces	(“Natural User Interfaces” OR “NUI”) OR ((“gesture recognition”) OR (“Speech and gesture recognition”)) OR (“Graphical user interface” OR “NUI”) OR (“Human Computer Interface” OR “HCI”) OR (“Multi sensor input” OR “Multiple sensor input”) OR (“Augmented reality”)	3,581	13.4	178.6
17. 3D Printing	((((3-D) OR 3D) AND Printing) OR "Additive manufacturing") OR "Selective laser sintering"	215	38.4	42.2
18. Big Data Analytics	(("Big data") OR “Massive Data”) AND Analytics	42	9.3	11.4
19. Machine Learning and Intelligent Systems	(((((("Artificial intelligence") OR "Machine Intelligence") OR "Intelligent systems") OR "Machine Learning") OR "Supervised learning") OR "Reinforcement learning"	13,199	17.6	3.7

20. Life Sciences	(((((((Bioinformatic) OR Biology) OR Biomedical) OR Biometrics) OR (Health) OR "Health care") OR Healthcare) OR “Life Sciences”) OR Medical) OR Medicine	28,510	450.0	592.8
21. Computational Biology and Bioinformatics	((((("Computational Biology") OR Bioinformatics) OR "Structural bioinformatics") OR Phylogenetics and evolutionary modeling) OR Phylogenetics	2,145	19.5	15.2
22. Robotics	(Robotics) OR Robot	8,817	658.0	74.6