Testimony

Before the Committee on Science, Space, and Technology, Subcommittee on Research and Technology and Subcommittee on Energy, of the U.S. House of Representatives on the hearing titled, "Big Data Challenges and Advanced Computing Solutions"

Submitted By

Anthony David Rollett

US Steel Professor of Materials Science and Engineering

Carnegie Mellon University

July 12th, 2018

Greetings, my name is Anthony Rollett and I am a Professor of Materials Science & Engineering in the College of Engineering at Carnegie Mellon University, in Pittsburgh, PA. At Carnegie Mellon I help lead the NextManufacturing Center, which is focused on advancing additive manufacturing and participate in the Manufacturing Futures Initiative—a campus wide effort focused on accelerating innovation and enhancing manufacturing in the Greater Pittsburgh region.

I thank Chairman Weber of the Energy Subcommittee and Chairwoman Comstock of the Research and Technology Subcommittee for inviting me to testify today. I also thank, Ranking Member Veasey, Ranking Member Lipinski and all of the Members of the Committee for your interest in Big Data and Computing, which are subjects of great interest and importance in my research.

In recent years, I have had the privilege of becoming closely involved in research on 3D printing, which is a key component of advanced manufacturing. It is clear to me that this is a seriously revolutionary technology because it forces us to think differently about how to make things. The design of a part is as intimately coupled to the printing process and the chosen material as a Stradivarius is to its wood and crafting. The difference is the importance of data as input and as output. Imagine that in a few years we will be able to, e.g., build a rocket that is tailored to the particular mission, instead of forcing the payload to match one of a limited set of vehicles. Or that "mass production" is transmuted into "mass individualization" such that Ford's proverbial "any color so long as black" becomes "any choice of color and size for dozens if not hundreds of parts of a car."

Let me begin by giving some context to the challenges and solutions by explaining that there are both practical applications and scientific advances to be gained from adapting to the availability of large amounts of data. I will outline three major challenges in advanced manufacturing, associated needs in STEM education, and a link between cybersecurity and manufacturing.

New and complex manufacturing technologies such as 3D printing are coming on stream that need further development, especially for qualification—the science of verifying that processes and/or materials will produce the characteristics required. Scientific research is generating ever larger streams of data that are challenging to handle and to interpret. There is an essential challenge in this that concerns the extraction of information from data. Current machine learning algorithms have been developed for big data, as we know, but the data concerned is information rich, e.g., faces or cats or cars. One could say that "big information" is as important as "big data."

Another important context is that education in all STEM subjects needs to include the use of advanced computing for data analysis. Our data acquisition instruments are essentially all run by computers and modeling our experiments essentially always involves large computer codes. These considerations and the examples that follow demonstrate the pervasive nature of big data and computing and the need to incorporate computer-aided learning into every aspect of how we function in science and engineering. I know that this Committee has focused on expanding computer science education. Those efforts are critical and appreciated.

Returning to the domain of advanced manufacturing, another exciting area of the application of big data, machine learning and advanced computing is to the manifold challenges of the materials used. In order to check the internal structure of materials and to understand their properties, we commonly cut, polish and photograph cross-sections.

An early lesson for those of us in materials science and engineering was to be told that although computer vision is well developed for finding cats, dogs, cars etc. in images or videos, the sort of cross-sectional images that we produce are for more complex and the features in each image much less obvious. An analogy might be a painting by Jackson Pollock with its seemingly analysis-defying random dribbles of paint. Yet there is an organization to the image that makes his paintings compelling art. I assert that many of the images of materials are equally complex and require domain-specific analysis. We have demonstrated success with examples of teaching the computer to recognize different kinds of metal powders used in additive manufacturing and recognizing different kinds of steel where the composition is constant but the processing history is varied. The bottom line is that advanced manufacturing is already a source of big data but there are many challenges in front of us to learn how to transmute the data into information and then discover the algorithms that allow us to learn from that data and optimize manufacturing.

These successes open up a wide range of potential impacts on improving materials, generating new materials, performing quality control on feedstocks, etc. We have also demonstrated that we can recognize failures in any individual powder spreading step that is essential to any powder bed 3D printing process; again, this points to an impact of improved machine control algorithms that exploit data, machine learning and high speed computing.

As another example of how data coupled with machine learning shows up almost everywhere, I ran across the "LettuceBot," which is a software that controls an agricultural machine towed by a tractor across lettuce plantations whose job is to decide which individual lettuce plants should stay in the ground to grow versus those that should be culled. Once again, cameras provide images that are then analyzed for a decision-making process followed by action (or not). The bigger picture is one where computer vision helps humans to make decisions about the manufacturing process that they are in charge of, i.e., advanced technology aiding workers .

To illustrate the challenges in my own research, I often use the light sources, i.e., x-rays from synchrotrons, most of which are curated by the Department of Energy. I use several modes of experiment such computed tomography, high energy diffraction microscopy and dynamic x-ray radiography. Computed tomography (CT) is very similar to getting an MRI scan except that the high energy x-rays from the synchrotron allows one to see inside the sort of dense materials from which we build aircraft, engines, rockets etc. at the micrometer scale (about 1/100th of a human hair). The results have been invaluable in understanding the feedstocks used in, e.g., 3D metals printing. High energy diffraction microscopy (HEDM) functions as a microscope, again at the micrometer scale, that provides, in full 3-dimensional form, a map with crystal information of the material in millimeter-sized samples. We can then heat and deform the sample and measure how it responds under load, which is proving invaluable for understanding what controls the durability of components, for example. Once again, a central challenge is how to transmute this ever expanding data stream into information and to discover the algorithms that allow us to learn from that data.

Dynamic x-ray radiography (DXR) provides movies of the melting of powder layers just as occurs in 3D printing with a laser, again at the micrometer scale. The dynamic nature of the process means that one must capture the process at the same rate as the more familiar case of a bullet penetrating armor. Over the last couple of years this technique has provided many deep insights into how additive manufacturing really works at the appropriate scale of length and time. From the perspective of data and computing, each experiment lasts for about a millisecond, i.e., 500 times faster than you can blink, and provides gigabytes of images. It is not difficult , therefore, to appreciate that a few days' worth of what we call "beam time" provide big data, so much so that we typically take it home on our own hard disk drives. Storing such large amounts of data is a challenge for this extremely important public resource. Transmitting such large amounts of data, e.g., to one's own university, is challenging. The mechanisms exist but they are not quick as they ought to be and accessing high speed transfers is definitely something for experts, even if the institutions at either end have the appropriate speed of access. The light sources themselves are well aware of the forthcoming challenge that is posed by the rapidly accelerating rate at which data is generated in the aggregate as detectors become ever larger and more sensitive.

Acquiring the data may only occupy a few days, but analyzing it often consumes months of time on the part of a graduate student. Perhaps a better way to say this is that we have the problem of converting raw data to useful information, by contrast with the data available from, e.g., social media, which are intrinsically information-rich. My judgment is that advanced computing algorithms to aid researchers in the conversion of data to information are nascent at best. Although there is a plethora of data analytics and machine learning techniques available, applying such techniques in any given domain requires time and effort.

Giving more serious attention to such challenges requires funding agencies to adopt the right vision in terms of recognizing the need for a fusion of research activities. We are in essence building the infrastructure for digital engineering and manufacturing.

A closely related issue is the timescale on which new methods are developed. The canonical 3 or 4-year research program rarely allows one to take a technique development to a reasonable point of maturity or technical readiness level in the modern argot. The high energy diffraction microscopy mentioned above is a case in point where an agency sustained the effort over roughly a decade, which enabled it to mature to the point where the research community was able to start using it more generally.

Additive manufacturing provides an excellent example of an application domain for big data and computing. Consider 3D printing of metals as a particular facet that has grown with dramatic speed from a small specialized activity that most believed (as did we) would only provide business cases in aerospace and only in rare instances, to a technology that essentially all OEMs consider that they must pay attention to. It is also provoking a reaction in education, where universities are acting at something faster than the proverbial glacial pace and instituting new programs across the scale, e.g., MS programs in additive manufacturing.

To print a part with a powder bed machine requires thousands-fold repetition of spreading a hairsbreadth layer of powder, writing the desired shape in that layer, shifting the part by the hairsbreadth, and repeating. Divide a part dimension by a hairsbreadth, multiply by yards of laser melt track, and one readily estimates that each part contains miles upon miles of melt tracks. There is a great deal of physics and chemistry detail required at the melt track scale.

Thus, the data stream is commensurately enormous ("big"), but the impact has to be such that useful information about the integrity of the part is obtained. Please do not be intimidated by the scale because the machines are highly functional and produce good results. Nevertheless, if we are to be able to qualify the machines to produce reliable parts that can be used in, e.g., commercial aviation, there is work to do.

Moreover if we as a country are to maintain our competitiveness in this area, we need the full range of tools that, crucially, include the application of big data and advanced computing. As a brief illustration, consider taking high speed movies of the melt pool using visible light (as opposed to the highly specialized x-ray approach). This generally has to be done at an angle to the laser beam and the images are confused by particle spatter and metal vapor plumes. This means that substantial processing must be done on the videos to render them useful to the researcher. We are only at the very beginning of being to use this type of data, let alone knowing how to incorporate the lessons learned as improved control algorithms. Permit me to underscore the importance of the research community and the publication of results so that companies involved in advanced manufacturing can adopt the results without necessarily revealing where they obtained the knowledge.

Finally, cybersecurity is widely understood to be an important problem, with almost weekly stories about data leaks and hacking efforts. What is less well understood is how manufacturing and cybersecurity must interface to each other. At the consumer level, concern has already been expressed about the ability of bad actors to gain access to IoT-enabled gadgets in one's home and control them or acquire data from them. With companies touting their ability to provide customer solutions that are based on networked machines, the importance of cybersecurity in manufacturing takes on a new significance and urgency. The caution in this instance is to not underestimate the importance of the domain-specific knowledge for determining which existing cybersecurity solutions will work and, more importantly, adapting the methods to suit a given domain. This is analogous to the way in which computer vision is applicable to materials science but has to be adapted to the particularities of the field.

Recommendations

As others have testified, the various agencies that provide federal funding for R&D have done an excellent job over the years of identifying worthwhile areas for development of new ideas. Please continue to support them.

Specialized facilities are tremendously important to the scientific and engineering community. The DOE has done an exceptional job in this regard and my own research is all the richer for it. In addition, DOE is investing in building machine learning capabilities. The manufacturing institutes—such as America Makes—have also been critical to advancing more applied research.

I suggest that there are three areas of opportunity. First, federal agencies should continue to support the application of machine learning to advanced manufacturing particularly for the qualification of new technologies and materials. Currently, no additive manufacturing processes or materials are qualified for mission critical defense or aerospace parts (non-mission critical additive parts are in use). As noted above, this requires advances in scientific research and strong collaboration with industry and among research and application and regulatory agencies. Winning the innovation race in the science of qualification is essential for future competitiveness and job creation in these technologies. In the future, research initiatives can also seize the potential for "moonshot" efforts on objectives such as integrating artificial intelligence capabilities directly into advanced manufacturing machines and advancing synergy between technologies such as additive manufacturing and robotics.

Second, we need to continue to energize and revitalize STEM education at all levels to reflect the importance of data, learning and computing, with a focus on manufacturing. Data analytics will play a vital role across the entire manufacturing enterprise—from the lab, to product design, production and product service functions. It will not be necessary for all workers to have a computer science degree. But varying degrees of comfort and capability with statistics and data analytics will be vital. As a step in this direction, the NextManufacturing Center at Carnegie Mellon has begun engaging teachers and students with the most advanced additive manufacturing machines. Investments that creatively stimulate a co-development of manufacturing with cybersecurity innovation will be essential.

Third, based on the evidence that machine learning is being successfully applied in many areas, we should encourage agencies to seek programs in areas where it is not so obvious how to apply the new tools and to to instantiate programs in communities where data, machine learning and advanced computing are not yet prevalent. Not only is domain-specific knowledge essential but, in manufacturing and research, the process of transmuting data into knowledge is a challenge in itself. In fact, one could say that "big information" is the twin of "big data."

Having traveled abroad extensively, I can report that the competition in science and technology is serious. Countries that we used to dismiss out of hand are publishing more than we are and securing more patents than we do. Time and again, national investment in new ideas and technology, coupled with an expectation that industry will strive to pick these up in their innovation process, has kept this country in the lead.

A best practice in my experience is where a funding agency has a well-established mechanism through which the scientific community can articulate needs and directions. Although a variety of mechanisms is appropriate, some work better than others. It is important that the community recognizes and is comfortable with whatever mechanism an agency uses.

Program Managers should have some discretion in what they fund so that they are able to respond quickly when an interesting new idea arises. High risk with high impact is often touted but less often encouraged.

Although some effort has been made to facilitate the transport of big data around the country, arranging to ship data at the terabyte scale requires substantial effort for ordinary researchers. This is, of course, is linked to the availability of data storage systems ("servers") that have the capacity and delivery speed to support such transfers. It would be helpful if one of the agencies were to be empowered to support such capabilities.

Again, thank you very much for the opportunity to share my views on this vital subject. I would be glad to answer any follow up questions you may have.

Rollett's research focuses on microstructural evolution and microstructure-property relationships in 3D, using both experiments and simulations. Interests include 3D printing of metals, materials for energy conversion systems, strength of materials, constitutive relations, microstructure, texture, anisotropy, grain growth, recrystallization, formability, extreme value statistics and stereology. Relevant techniques highlight spectral methods in micro-mechanics, Dynamic X-ray Radiography and High Energy Diffraction Microscopy. Important recent results include definition of process windows in 3D printing through characterization of porosity, 3D comparisons of experiment and simulation for plastic deformation in metals, the appearance of new grains during grain growth, and grain size stabilization. He has been a Professor of Materials Science & Engineering at Carnegie Mellon University since 1995 and before that was with the Los Alamos National Laboratory. His most recent honor was the award of US Steel Professor of Metallurgical Engineering & Materials Science in 2017. He is the co-Director of CMU's NextManufacturing Center that is dedicated to advancing manufacturing especially through 3D printing. He has over 200 peer-reviewed publications