

# The past, present and future of computing in high-energy physics

26 Nov 2019 [Tushna Commissariat](#)

This article first appeared in the 2019 *Physics World Focus on Computing* under the headline "Open to new ideas"

**Federico Carminati** is the chief innovation officer at CERN openlab, a public–private partnership that facilitates collaborations between CERN and computing firms. He spoke to Tushna Commissariat about how particle physics has shaped (and been shaped by) computing trends such as big data and machine learning



**Emerging challenges** Federico Carminati addresses an audience at a quantum computing workshop hosted by CERN openlab. (Courtesy: CERN/Andrew Purcell)

**One of the main computing tasks at CERN is to sort through huge numbers of particle collisions and identify the ones that are interesting. What impact has this “big data” had on high-energy physics?**

I'd like to invert the question, because rather than talking about the impact of big data on high-energy physics, I think it's more interesting to talk about the impact of high-energy physics on big data.

High-energy physicists started working with very large datasets in the 1990s. Out of all the scientific disciplines, our datasets were among the largest, and we had to develop our own solutions for handling them – firstly because there was nothing else, and secondly because CERN operates within a specific social, economic and political framework that encourages us to spread work around different member countries. This is only natural: we're getting a lot of money for computing from national funding agencies, and they naturally privilege local investments.

So, high-energy physicists were doing big data before big data was a thing. But somehow, we failed to capitalize on it, because we didn't communicate this well at the time. A similar thing happened with open-source software. We have a philosophy of openness and sharing at CERN. We could have invented open-source and popularized it. But we didn't. Instead, when open-source started to become widespread, we said, "Oh, yeah, that's interesting. This is what we have been doing for 20 years. Nice."

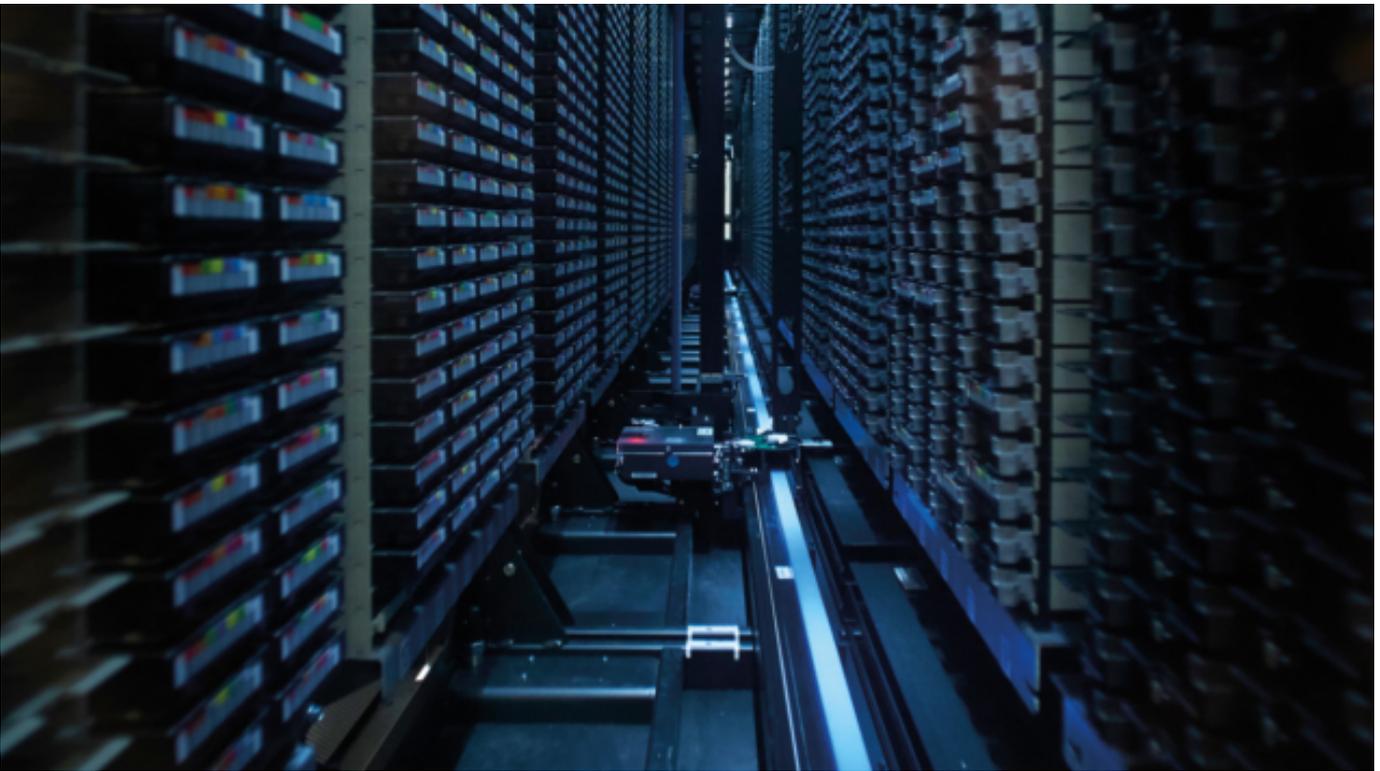
The problem is that we have very little room to capitalize on our ideas beyond fundamental physics. Our people are working day and night on experimental physics. Everything else is just something we do because we want to do physics. This means that once we've done something, we don't have time to develop it further. Sure, Tim Berners-Lee came up with the concept of the World Wide Web when he was at CERN, but the Web was taken up and developed by the rest of the world, and the same thing happened with big data. Compared to the amount of data Google and Facebook now have, we have very little indeed.

### **How does CERN openlab help to tackle the computing challenges faced at CERN?**

We identify computing challenges that are of common interest and then set up joint research-and-development projects with leading ICT companies to tackle these.

In the past, CERN has taken a sound engineering-style approach to computing. We bought the computing we needed at the most convenient conditions, and off we went. Nowadays, though, computing is evolving so quickly that we need to know what is coming down the pipeline. Evaluating new technologies after they are on the market is not good enough, so it is important to work closely with leading companies to understand how technologies are evolving and to help shape this process.

Another aspect of CERN openlab's work involves working with other research communities to share technologies and techniques that may be of mutual benefit. For example, we are working with Unosat, the UN technology platform that deals with satellite imagery and which is hosted at CERN, to help estimate the population of refugee camps. This is a big challenge, and part of the problem is that it's difficult – sometimes even dangerous – to count the actual number of people living in these camps. So, we are developing machine-learning algorithms that will count the number of tents in satellite photos of the camps, which refugee agencies can then use to estimate the number of people.



**Future proofing** Carminati expects machine learning to play an increasingly important role in data analysis at CERN. (Courtesy: CERN/Maximilien Brice and Julien Marius Ordan)

### **What are some of the technologies you see becoming more important in the future?**

We're already using machine learning and artificial intelligence (AI) across the board, for data classification, data analysis and simulations. With AI, you can make very subtle classifications of your data, which is of course a large part of what we do to find new particles and elaborate on new physics.

The high-energy physics community actually started looking at machine learning in the 1990s, but every time we started to do something, we saw the tremendous possibilities, and then we had to stop for lack of computing power. Computers were just not fast enough to do what we wanted to do. Now that computers are faster, we can explore deep learning using deep networks. But these networks are very slow to train, and this is something we are working on now.

### **What about quantum computers? What impact will they have on high-energy physics?**

This is very hard to say, because it's crystal-ball thinking. But I can tell you that it's important to explore quantum computing, because in 10 years we will have a massive shortage of computing power.

High-energy physics is in a very funny situation. We have two theories: general relativity and quantum mechanics. General relativity explains the behaviour of stars and planets and the evolution of the universe, and it is the epitome of elegance. It's a beautiful theory, and it works: you use it every day when you use GPS on your smartphone. Quantum mechanics, in contrast, is a very complex theory, and although it's pretty successful – the fact that we found the Higgs boson is a testament to its success – there are a lot of unsolved questions that it doesn't answer. The other problem is that quantum mechanics does not work with general relativity. When we try to unify

them, the outcomes are really weird-looking, and the few predictions that we make are not borne out by reality. So we have to find something else.

But how do you find that “something else”? Usually, you find something that is not explained by present theories of physics, and that unexplained thing gives theorists hints about how to proceed. So we have to find something that contradicts our current view of the Standard Model. However, we know that this cannot be a big thing, because every observation we make is more or less confirmed by the model. Instead, we are looking for something subtle. The old game was to find a needle in the haystack of our data. The new game is to look in a stack of needles for a needle that is slightly different. This new game will involve an incredible amount of data, and incredible precision in processing it, so we are increasing the amount of data we take and increasing the quality of our detectors. But we will also need much more computing power to analyse these data, and we cannot expect our computing budget to increase by a factor of 100.

That means we have to find new sources of very fast computing. We don't know what they will be, but quantum computing may be one of them. We are looking into it as a candidate to provide us with computing power in the future, and also because there is one really exciting thing that you can do with quantum computing that you cannot do nearly as well with normal computing, and that is to directly simulate a quantum system.

### **What is the particle-physics community doing to meet its more immediate computing challenges?**

We are moving forward along several axes. One of them is to exploit current technology as well as possible. It used to be that when computers got faster, it was because they had a faster clock rate. That kind of increase in power was readily useable, because you could just park your old program on a new machine and it would run faster. More recently, though, computing power has gone up because of increases in the number of transistors on a chip, meaning that you can make more operations in parallel. But to take advantage of this power you need to rewrite your programs to exploit this parallelism, and that is not easy.

We are also exploring different computing architectures, such as graphical processing units GPUs, to see how well they can fit into our computing environment. And of course, there is the work we are doing on novel technologies such as machine learning, which could really improve the speed of certain operations.

The biggest prize, though, would be quantum computing. Quantum computers would be useful across our entire workload. But we will have to develop new ways of thinking to exploit them, and this is why it is so important that CERN openlab is working with different companies in this area. Can we imagine software that is independent of the type of computing we are using? Will we have to write a different program for each type of quantum computer? Or will we develop algorithms that can be ported onto different quantum computers? For the moment, nobody knows. An extension of C++ or Python that could be ported to different quantum computers is, for the moment, science

fiction. But we have to think creatively in this direction if we want to assess the capabilities and opportunities.