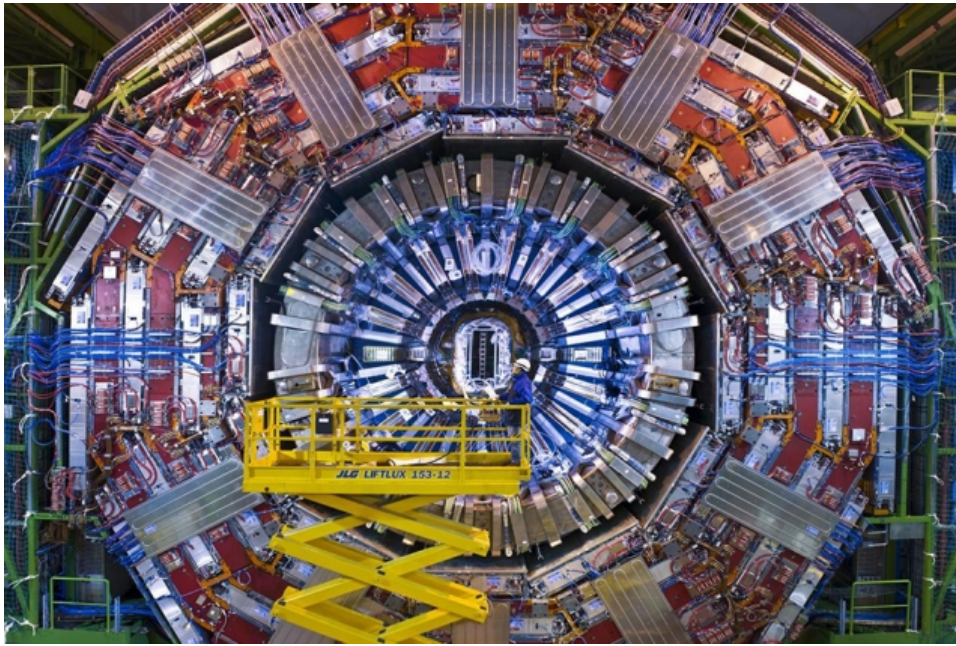


MIT News

ON CAMPUS AND AROUND THE WORLD



The Compact Muon Solenoid is a general-purpose detector at the Large Hadron Collider.

Image courtesy of CERN

First open-access data from large collider confirm subatomic particle patterns

CERN Open Data Portal results reveal predictable patterns from colliding high-energy protons.

Jennifer Chu | MIT News Office
September 29, 2017

In November of 2014, in a first, unexpected move for the field of particle physics, the Compact Muon Solenoid (CMS) experiment — one of the main detectors in the world's largest particle accelerator, the Large Hadron Collider — released to the public an immense amount of data, through a website called the CERN Open Data Portal.

The data, recorded and processed throughout the year 2010, amounted to about 29 terabytes of information, yielded from 300 million individual collisions of high-energy protons within the CMS detector. The sharing of these data marked the first time any major particle collider experiment had released such an information cache to the general public.

A new study by Jesse Thaler, an associate professor of physics at MIT and a long-time advocate for open access in particle physics, and his colleagues now demonstrates the scientific value of this move. In a paper published in *Physical Review Letters*, the researchers used the CMS data to reveal, for the first time, a universal feature within jets of subatomic particles, which are produced when high-energy protons collide. Their effort represents the first independent, published analysis of the CMS open data.

"In our field of particle physics, there isn't the tradition of making data public," says Thaler, who is a researcher in MIT's Laboratory for Nuclear Science. "To actually get data publicly with no other restrictions — that's unprecedented."

Part of the reason groups at the Large Hadron Collider and other particle accelerators have kept proprietary hold over their data is the concern that such data could be misinterpreted by

RELATED

Paper: "Exposing the QCD splitting function with CMS open data."

Jesse Thaler

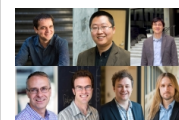
Center for Theoretical Physics

Laboratory for Nuclear Science

Department of Physics

School of Science

ARCHIVES



School of Science professors granted tenure

people who may not have a complete understanding of the physical detectors and how their various complex properties may influence the data produced.

“The worry was, if you made the data public, then you would have people claiming evidence for new physics when actually it was just a glitch in how the detector was operating,” Thaler says. “I think it was believed that no one could come from the outside and do those corrections properly, and that some rogue analyst could claim existence of something that wasn’t really there.”

“This is a resource that we now have, which is new in our field,” Thaler adds. “I think there was a reluctance to try to dig into it, because it was hard. But our work here shows that we can understand in general how to use this open data, that it has scientific value, and that this can be a stepping stone to future analysis of more exotic possibilities.”

Thaler’s co-authors are Andrew Larkoski of Reed College, Simone Marzani of the State University of New York at Buffalo, and Aashish Tripathy and Wei Xue of MIT’s Center for Theoretical Physics and Laboratory for Nuclear Science.

Seeing fractals in jets

When the CMS collaboration publicly released its data in 2014, Thaler sought to apply new theoretical ideas to analyze the information. His goal was to use novel methods to study jets produced from the high-energy collision of protons.

Protons are essentially accumulations of even smaller subatomic particles called quarks and gluons, which are bound together by interactions known in physics parlance as the strong force. One feature of the strong force that has been known to physicists since the 1970s describes the way in which quarks and gluons repeatedly split and divide in the aftermath of a high-energy collision.

This feature can be used to predict the energy imparted to each particle as it cleaves from a mother quark or gluon. In particular, physicists can use an equation, known as an evolution equation or splitting function, to predict the pattern of particles that spray out from an initial collision, and therefore the overall structure of the jet produced.

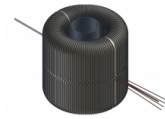
“It’s this fractal-like process that describes how jets are formed,” Thaler says. “But when you look at a jet in reality, it’s really messy. How do you go from this messy, chaotic jet you’re seeing to the fundamental governing rule or equation that generated that jet? It’s a universal feature, and yet it has never directly been seen in the jet that’s measured.”

Collider legacy

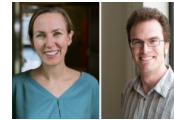
In 2014, the CMS released a preprocessed form of the detector’s 2010 raw data that contained an exhaustive listing of “particle flow candidates,” or the types of subatomic particles that are most likely to have been released, given the energies measured in the detector after a collision.

The following year, Thaler published a theoretical paper with Larkoski and Marzani, proposing a strategy to more fully understand a complicated jet in a way that revealed the fundamental evolution equation governing its structure.

“This idea had not existed before,” Thaler says. “That you could distill the messiness of the jet into a pattern, and that pattern would match beautifully onto that equation — this is what we found when we applied this method to the CMS data.”



Team simulates a magnetar to seek dark matter particle



Biobarrier explorer and dark matter theorist win MIT’s prestigious junior faculty award

To apply his theoretical idea, Thaler examined 750,000 individual jets that were produced from proton collisions within the CMS open data. He looked to see whether the pattern of particles in those jets matched with what the evolution equation predicted, given the energies released from their respective collisions.

Taking each collision one by one, his team looked at the most prominent jet produced and used previously developed algorithms to trace back and disentangle the energies emitted as particles cleaved again and again. The primary analysis work was carried out by Tripathee, as part of his MIT bachelor's thesis, and by Xue.

"We wanted to see how this jet came from smaller pieces," Thaler says. "The equation is telling you how energy is shared when things split, and we found when you look at a jet and measure how much energy is shared when they split, they're the same thing."

The team was able to reveal the splitting function, or evolution equation, by combining information from all 750,000 jets they studied, showing that the equation — a fundamental feature of the strong force — can indeed predict the overall structure of a jet and the energies of particles produced from the collision of two protons.

While this may not generally be a surprise to most physicists, the study represents the first time this equation has been seen so clearly in experimental data.

"No one doubts this equation, but we were able to expose it in a new way," Thaler says. "This is a clean verification that things behave the way you'd expect. And it gives us confidence that we can use this kind of open data for future analyses."

Thaler hopes his and others' analysis of the CMS open data will spur other large particle physics experiments to release similar information, in part to preserve their legacies.

"Colliders are big endeavors," Thaler says. "These are unique datasets, and we need to make sure there's a mechanism to archive that information in order to potentially make discoveries down the line using old data, because our theoretical understanding changes over time. Public access is a stepping stone to making sure this data is available for future use."

This research was supported, in part, by the MIT Charles E. Reed Faculty Initiatives Fund, the MIT Undergraduate Research Opportunities Program, the U.S. Department of Energy, and the National Science Foundation.

Topics: [Data](#) [Particles](#) [Open access](#) [Physics](#) [Research](#)

[School of Science](#) [National Science Foundation \(NSF\)](#)

[Laboratory for Nuclear Science](#) [Department of Energy \(DoE\)](#)

About This Website

This Website is maintained by the MIT News Office, part of the Office of Communications.

MIT News Office • Building 11-400
Massachusetts Institute of Technology • Cambridge, MA 02139-4307

