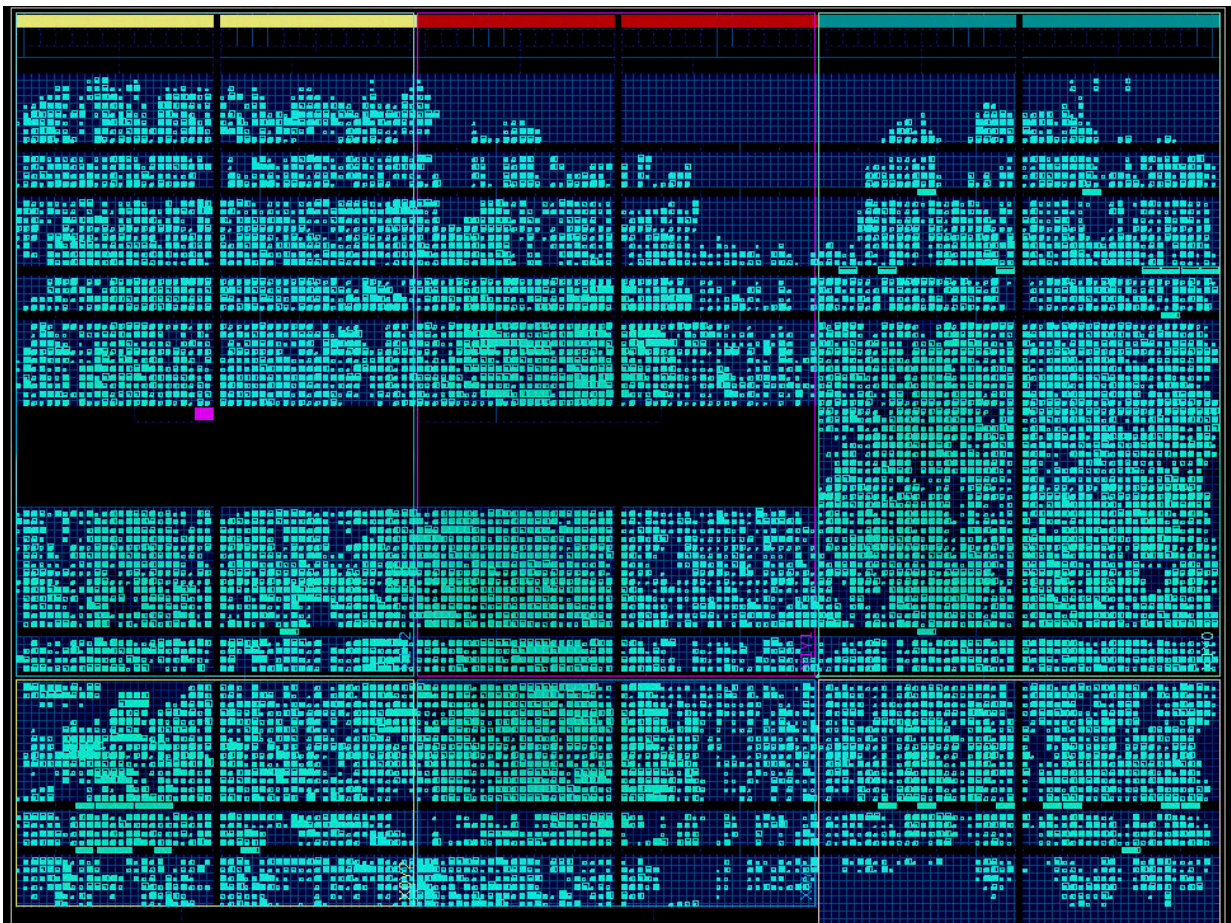


Speeding up machine learning for particle physics

June 22 2021, by Ana Lopes



An ultra-compressed deep neural network on a field-programmable gate array.
Credit: Sioni P. Summers

Machine learning is everywhere. For example, it's how Spotify gives you suggestions of what to listen to next or how Siri answers your questions. And it's used in particle physics too, from theoretical calculations to data analysis. Now a team including researchers from CERN and Google has come up with a new method to speed up deep neural networks—a form of machine-learning algorithms—for selecting proton–proton collisions at the Large Hadron Collider (LHC) for further analysis. The technique, described in a paper just published in *Nature Machine Intelligence*, could also be used beyond particle physics.

The particle detectors around the LHC ring use an electronic hardware "trigger" system to select potentially interesting particle collisions for further analysis. With the current rate of proton–[proton collisions](#) at the LHC, up to 1 billion collisions per second, the software currently in use on the detectors' trigger systems chooses whether or not to select a collision in the required time, which is a mere microsecond. But with the collision rate set to increase by a factor of five to seven with the future upgraded LHC, the HL-LHC, researchers are exploring alternative software, including [machine-learning](#) algorithms, that could make this choice faster.

Enter the new study by CERN researchers and co-workers, which builds on previous work that introduced a software tool to deploy [deep neural networks](#) on a type of hardware, called field-programmable gate arrays (FPGAs), that can be programmed to perform different tasks, including selecting particle collisions of interest. The CERN researchers and their colleagues developed a technique that reduces the size of a deep neural [network](#) by a factor of 50 and achieves a network processing time of tens of nanoseconds—well below the time available to choose whether to save or discard a [collision](#).

"The technique boils down to compressing the deep neural network by reducing the numerical precision of the parameters that describe it," says

co-author of the study and CERN researcher Vladimir Loncar. "This is done during the training, or learning, of the network, allowing the network to adapt to the change. In this way, you can reduce the network size and processing time, without a loss in network performance."

In addition, the technique can find which numerical precision is best to use given certain hardware constraints, such as the amount of available hardware resources.

If that wasn't enough, the technique has the advantage that it is easy to use for non-experts, and it can be used on FPGAs in particle detectors and in other devices that require networks with fast processing times and small sizes.

Looking forward, the researchers want to use their technique to design a new kind of trigger system for spotting collisions that would normally be discarded by a conventional trigger system but that could hide new phenomena. "The ultimate goal is to be able to capture collisions that could point to new physics beyond the Standard Model of [particle physics](#)," says another co-author of the study and CERN researcher Thea Aarrestad.

More information: Claudionor N. Coelho et al, Automatic heterogeneous quantization of deep neural networks for low-latency inference on the edge for particle detectors, *Nature Machine Intelligence* (2021). [DOI: 10.1038/s42256-021-00356-5](https://doi.org/10.1038/s42256-021-00356-5)

Provided by CERN

Citation: Speeding up machine learning for particle physics (2021, June 22) retrieved 25 April 2024 from <https://techxplore.com/news/2021-06-machine-particle-physics.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.