Popular Science Summary

Aleko Lilius

June 2024

Machine learning is a well-known concept that you probably have heard a lot about. Have you ever wondered how scientists are harnessing the power of machine learning to revolutionize data compression? This thesis explores integrating machine learning-based compression with Field-Programmable Gate Arrays (FPGAs) to enhance bandwidth compression, which is crucial for scientific research where vast amounts of data are generated rapidly in real-time.

Imagine your average streaming service buffering your videos 17 times faster because the data transferred is reduced to 1/17 of its original size before being sent. This is precisely the potential that FPGA-accelerated machine learning offers for scientific data processing. In the future, there is a possibility that data collected in real-time will be compressed in real-time as well.

But it's not just about compressing more data in less time. The research has to be thoroughly evaluated. By assessing the performance of advanced compression algorithms like Baler on FPGAs, key factors influencing optimal performance, such as model size, precision levels, and clock period, have been identified. These findings not only advance our understanding of FPGA-accelerated machine learning but also hold significant potential for advancing computational methods and enabling more efficient, adaptable solutions in various fields of science and technology.

So, where are the findings from this thesis most applicable, and when is bandwidth compression most needed? Many research centers worldwide, such as CERN in Switzerland and MAX IV in Sweden, are experiencing an abundance of produced data. The amount of data produced is currently getting out of hand, and to save time, resources, and data storage, they require some sort of data reduction solution. By compressing data in real-time as it is produced, the amount of data that can be saved for storage will increase. This potentially enables further research at CERN, MAX IV, and other research centers that would benefit from this.

This thesis lays a foundation for further developing the hardware implementation of the Baler algorithm and highlights the exciting intersection of machine learning and FPGA technology. Continuing to explore the possibilities of this innovative approach brings us one step closer to unlocking the full potential of data-driven discovery in the digital age.