

Talk 4: Radiation Experiments on AI accelerators: Current and Future Challenges and Opportunities

Wednesday, 12 June 2024 15:05 (20 minutes)

Abstract:

Complex AI accelerators, such as Graphics Processing Units (GPUs) or dedicated accelerators implemented in Field Programmable Gate Arrays (FPGAs) or in Application Specific Integrated Circuits (ASICs), such as the Google's Tensor Processing Unit (TPU) are rapidly making their way in the chip market. Embedding AI is extremely interesting for automotive, production lines, and aerospace. To be implemented, a self-driving system needs to analyze a huge amount of images and signals in real time. Nonetheless, guaranteeing sufficient reliability is challenging since both the hardware architecture and the running software are highly complex. It is then hard to characterize the radiation reliability of the framework and the experimental data risks to be biased to the specific configuration chosen for the experiment.

In the talk we will investigate the challenges related to the reliability evaluation of GPUs, FPGAs, and TPUs executing neural networks. The evaluation, to be accurate and precise, is based on the combination of beam experiments and fault injection at different levels of abstractions (RTL, microarchitectural, and software). This combination allows us to have a realistic evaluation of the error rate, distinguish between tolerable errors and critical errors, and to design efficient and effective hardening solutions for neural networks.

CV:

Paolo Rech received his master and Ph.D. degrees from Padova University, Padova, Italy, in 2006 and 2009, respectively. Since 2022 Paolo is an associate professor at Università di Trento, in Italy and since 2012 he is an associate professor at UFRGS in Brazil. He is the 2019 Rosen Scholar Fellow at the Los Alamos National Laboratory, he received the 2020 impact in society award from the Rutherford Appleton Laboratory, UK. In 2020 Paolo was awarded the Marie Curie Fellowship at Politecnico di Torino, in Italy. His main research interests include the evaluation and mitigation of radiation-induced effects in autonomous vehicles for automotive applications and space exploration, in large-scale HPC centers, and quantum computers.

Presenter: RECH (UNIVERSITÀ DI TRENTO), Paolo

Session Classification: Session 2: Atmospheric Industrial Session