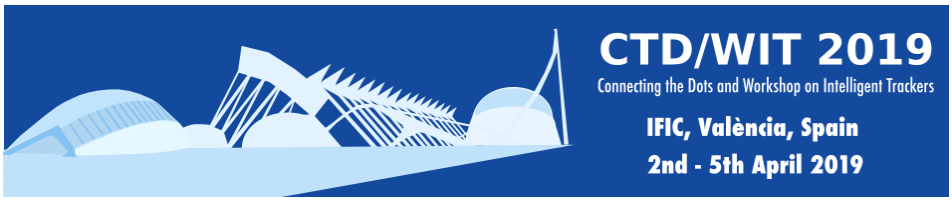

Contribution ID: **102** Type: **Talk**

Low Latency Neural Networks using Heterogenous Resources on FPGA for the Belle II Trigger

Tuesday, April 2, 2019 12:00 PM (25 minutes)

One of the major components of the Belle II trigger system is the neural network trigger. Its task is to estimate the z-Vertex particle tracks observed in the experiments drift chamber. The trigger is implemented on FPGAs to ensure flexibility during operation and leverage their IO capabilities. Meanwhile the implementation has to estimate the vertex in a few hundred nanoseconds to fulfill the requirements of the experiment. A first version of that trigger was operational during the first collisions. While it was able to estimate the vertex, it had some drawbacks regarding the possible throughput and timing closure. These are the focus of this work, which modifies the original design to allow two networks running in parallel and less routing congestion. We conducted a rescheduling of multiply and accumulate which are the basic operations in such networks. While the original design tried to parallelize as much as possible, the rescheduling tries to reduce the number of parallel data transmission by reusing processing modules. This way resource consumption was reduced by 40% for DSPs. To further increase the throughput by operating an additional network in parallel, we investigated the balanced use of SRAM-LUTs and DSPs for multiply and accumulate operations. With the found balancing ratio the trigger is able to operate two neural networks in parallel on the targeted FPGA within the required latency.

Primary authors: BAEHR, Steffen (Karlsruhe Institute of Technology); Mr POEHLER, Julian (Karlsruhe Institute of Technology)

Co-authors: Mr UNGER, Kai (Karlsruhe Institute of Technology); Mr HOCHSTUHL, Adam (Karlsruhe Institute of Technology); Mr BECKER, Juergen (Karlsruhe Institute of Technology); KIESLING, Christian (Werner-Heisenberg-Institut)

Presenter: BAEHR, Steffen (Karlsruhe Institute of Technology)

Track Classification: 6: Architectures and techniques for fast track reconstruction