Formal model of the FPGA implemented neural networks

Martin Krčma, Zdeněk Kotásek

An artificial neural network is an abstract computer model of human brain. The human brain has an aproximately hundreds of trillion cells called neurons. These cells are interconnected with even more links called synapses. Similar to the human brain, the artificial neural networks (ANNs) are composed of units called the artificial neurons (or simply neurons) and interconnections between them. The human brain is the centre of perception, cognition, intelligence. It has a capability to learn from sense perception and experience. Similar to the brain, the ANNs are able to learn from patterns. This learning can be applied to various tasks, typically classification, prediction, clustering and control tasks. And if relearning is allowed, it makes the ANNs strongly adaptive. It is a known fact that the human brain loses neurons for various reasons during lifetime. But it is also capable of creating new connections (trough relearning) between remaining cells (and complete new neurons according to a new research). So the brain naturally adapts its own damages, and restores and maintains its functionality. If we could make the neural networks with this capability, we could use them to construct fault tolerant architectures. If we consider implementation, according to the massive parallelism of the neural networks structure, the field programmable gate arrays (FPGAs) are one of the suitable choices as they are parallel by design and capable of on-line reconfiguration.

Several alternatives exist how to implement neural networks in FPGAs. The field programmable neural arrays (FPNAs) are one of them. The FPNAs are generally a model of the ANNs composed of dedicated interconnected units. These units approximates neurons and synapses and structure of their interconnection is, unlike the ANNs, not predefined and completely optional. This leads to option of constructing the FPNAs structurally suitable for implementation in FPGAs. As the FPGA's interconnection matrix has a grid structure, we can design the FPNA as the grid as well thereby it is easy to be mapped into gate array's logic. In this kind of implementation some units might be more important than others. Some units have bigger influence on the rest of the architecture as the results of the computation performed by them are used in units which have more successors. This influence is also affected with the input data parameters and the parameters of units themselves. If we are going to construct fault tolerant architectures, we might need to determine which units are more important and need to be made up more fault tolerant than others.

In this work we present a formal model of the FPGAs implemented neural networks, especially the FPNAs, considering the fault tolerant design, and the methods for units and data paths importance determination leading to critical and non-critical architecture elements identification.