An Experimental Study of Reduced-Voltage Operation in Modern FPGAs for Neural Network Acceleration

<u>Behzad Salami</u>

Fahrettin Koc

Osman Unsal

Baturay Onural

Oguz Ergin

Hamid Sarbazi-Azad

Ismail Yuksel

Adrian Cristal

Onur Mutlu



Barcelona Supercomputing Center Centro Nacional de Supercomputación









50th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 30th June, 2020

Executive Summary

<u>Goal</u>: Improve the **power-efficiency** of FPGA-based neural networks by:
✓ Undervolting (i.e., underscaling supply voltage) below nominal level

<u>Evaluation Setup</u>

- **5** Image classification workloads
- ✓ **3** Xilinx UltraScale+ ZCU102 platforms

• <u>Main Results</u>

- ✓ Large voltage guardband (i.e., 33%)
- >3X power-efficiency gain



• In the Main Talk

- ✓ Characterization of FPGA voltage **behavior** and its impact on the **power-reliability**
- ✓ Frequency underscaling
- ✓ Environmental temperature

An Experimental Study of Reduced-Voltage Operation in Modern FPGAs for Neural Network Acceleration

<u>Behzad Salami</u>

Fahrettin Koc

Osman Unsal

Baturay Onural

Oguz Ergin

Hamid Sarbazi-Azad

Ismail Yuksel

Adrian Cristal

Onur Mutlu



Barcelona Supercomputing Center Centro Nacional de Supercomputación









50th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 30th June, 2020