Design and Implementation of a Bitonic Sorter Based DCNN Using Adiabatic Superconducting Logic

O. Chen¹, T. Tanaka¹, R. Cai², Y. Wang², N. Yoshikawa¹

¹Yokohama National University, Japan
²Northeastern University, USA

Abstract— The Adiabatic Quantum-Flux-Parametron (AQFP) superconducting technology has been recently developed, which achieves the highest energy efficiency among superconducting logic families, potentially 104 -105 gain compared with state-of-the-art CMOS. Besides ultrahigh energy efficiency, AQFP exhibits two unique characteristics: the deep pipelining nature since each AQFP logic gate is connected with an AC clock signal, which increases the chance to adopt time-independent computing scheme; the second is the unique opportunity of true random number generation (RNG) using a single AQFP buffer, far more efficient than RNG in CMOS. We point out that these two characteristics make AQFP especially compatible with the stochastic computing (SC) technique, which uses a time-independent bit sequence for value representation and is compatible with the deep pipelining nature. Further, the application of SC has been investigated in DNNs in prior work, and the suitability has been illustrated as SC is more compatible with approximate computations. This work is to design an SC-based DNN acceleration components using AQFP technology. Being different to prior design scheme using approximate parallel counter (APC) based architecture, bitonic sorter-based implementation are utilized as an alternative to decrease the design complexity since the latter requires smaller hardware footprint when comparing to the previous design. We further implemented a 32-bit bitonic sorter using AIST 10kA/cm² high-speed standard process (HSTP), which consists of 7557 Josephson junctions and occupies an area of 2.3mm×6.6mm. The energy dissipation per clock cycle (at 5GHz) of this design is 4.99 aJ, achieving up to 31902 times higher energy efficiency compared to its CMOS implemented counterpart.

Keywords (Index Terms) — DNN, AQFP, Stochastic computing, deep learning acceleration.