ABSTRAK

Abstrak—RISC-V adalah Arsitektur Set Instruksi (*Instruction Set Architecture* atau ISA) berbasis *open-source* yang dapat diimplementasikan dan dikembangkan tanpa adanya royalti. Meskipun ISA *base* RISC-V termasuk ke dalam tipe prosesor *general-purpose*, set instruksi standarnya sering kali kurang efisien untuk menjalankan algoritma yang komputasional intensif, seperti *neural network*. Algoritma *neural network* dapat dipercepat secara signifikan menggunakan *Floating-Point Multiply-Accumulator* (MAC) untuk menangani proses perkalian matriks-vektor yang berulang. Namun, *Floating-Point Unit* sering kali memerlukan konsumsi energi dan penggunaan area yang besar. Oleh karena itu, pengembangan *floating-point multiplier* yang efisien sangat penting untuk meningkatkan performa tanpa penggunaan sumber daya yang berlebihan.

Penelitian ini mengembangkan modul *floating-point adder* dan *multiplier* yang kemudian digabungkan untuk membuat modul *Floating-Point* MAC. Terdapat dua varian *multiplier* yang dikembangkan, yaitu *exact multiplier* dan *approximate multiplier*. Modul - modul ini dimodelkan menggunakan bahasa Verilog *Register Transfer Level* (RTL) lalu diintegrasikan ke dalam prosesor berbasis RISC-V, yaitu Nios V sebagai instruksi khusus. Performa dan akurasi dari instruksi tersebut dievaluasi dengan menjalankan proses *inference* pada algoritma *neural network*.

Hasil penelitian menunjukkan bahwa penggunaan *Floating-Point* MAC yang dikembangkan mampu mempercepat proses *neural network inference* hingga 2.73x. Selain itu, *approximate multiplier* meningkatkan efisiensi *Area-Delay Product* (ADP) sebesar 3.36% dibandingkan dengan desain *exact multiplier*. Dalam hal akurasi model *neural network*, penggunaan *approximate multiplier* tidak secara langsung menurunkan akurasi. Sebaliknya, akurasi bergantung pada nilai *weights* dan bias yang diperoleh selama proses *training* model, sehingga bisa menghasilkan akurasi yang lebih tinggi, lebih rendah, atau setara.

Kata kunci—RISC-V, *Floating-Point Multiply-Accumulator*, *neural network*, *custom instruction*, *approximate multiplier*.

ABSTRACT

Abstract—RISC-V is an open-source Instruction Set Architecture (ISA) that enables royalty-free implementation and development. While the RISC-V base ISA serves as a versatile general-purpose processor, its standard instruction set often lacks the efficiency required for computationally intensive tasks, such as neural network inference. These algorithms can be significantly accelerated using a Floating-Point Multiply-Accumulator (MAC) to handle repetitive matrix-vector multiplications. However, conventional floating-point units often incur high energy consumption and significant silicon area overhead. Therefore, developing efficient floating-point multipliers is essential to enhance performance without excessive resource costs.

This research focuses on the development of custom floating-point adder and multiplier modules, which are combined to create a dedicated Floating-Point MAC unit. Two multiplier variants were developed: an exact multiplier and an approximate multiplier. These custom instructions were modeled using Verilog Register Transfer Level (RTL) and integrated into the Nios V, a RISC-V-based processor. The performance and accuracy of these instructions were evaluated by executing neural network inference tasks.

The experimental results demonstrate that the custom Floating-Point MAC accelerates neural network inference by up to 2.73x. Furthermore, the approximate multiplier improved the Area-Delay Product (ADP) efficiency by 3.36% compared to the exact multiplier design. In terms of model performance, the use of the approximate multiplier does not inherently reduce inference accuracy. The accuracy is dependent on the specific weights and biases obtained during training, resulting in outcomes that may be higher, lower, or equivalent to the exact multiplier baseline.

Keywords—RISC-V, Floating-Point Multiply-Accumulator, neural network, custom instruction, approximate multiplier.