

ANALYSIS OF PERFORMANCE/ACCURACY TRADEOFFS FOR
FLOATING POINT APPLICATIONS ON GPUS

Huyen Nguyen

Dr. Michela Becchi, Thesis Supervisor

ABSTRACT

Floating-point computations produce approximate results, which can lead to inaccuracy problems. Existing work addresses two issues: first, the design of high precision floating-point representations; second, the study of methods to trade-off accuracy and performance of serial CPU applications. However, a comprehensive study of the trade-off between accuracy and performance for multithreaded applications is missing. In my thesis, I study this trade-off on GPU. In particular, my study covers the use of different floating-point precisions (i.e., single and double floating-point precision in IEEE 754 standard, GNU Multiple Precision, and composite floating-point precision) on a variety of real-world and synthetic benchmark applications. I explore how the use of multithreading and instruction-level parallelism on GPU can allow the use of higher precision arithmetic to improve accuracy without paying in terms of execution time. As a result of my analysis, I provide insights to guide users to the selection of the arithmetic precision leading to a good performance/accuracy tradeoff depending on the arithmetic operations used in their program (addition, multiplication, division), the degree of multithreading of their program, and its arithmetic intensity.

Keywords – floating-point arithmetic; parallel computing; multithreading.