

MadGraph5 GPU Development

Kernel Profiling

A. Thete & C. Vuosalo

University of Wisconsin-Madison

30.07.2024

ALOHA Subroutine Variations

Routine	Wall Time [ms]	Memory Throughput [%]	Warp Stall Statistics	Comments
NoFact	67.45 ± 0.01 (+0.02%)	62.32 ± 0.00 (+0.05%)	13.21% (-0.92 pp)	45.11% LS, 19.31% W, 11.12% NI
FactByWFact	67.41 ± 0.02 (-0.04%)	62.31 ± 0.00 (+0.03%)	17.04% (+2.91 pp)	47.00% LS, 17.55% W, 12.39% NI
VectorFirst	67.34 ± 0.04 (-0.14%)	62.28 ± 0.00 (-0.02%)	14.06% (-0.06 pp)	43.42% LS, 19.06% W, 13.42% NI
VectorByHand	67.41 ± 0.02 (-0.03%)	62.34 ± 0.00 (+0.08%)	16.94% (+2.81 pp)	45.52% LS, 18.44% W, 11.55% NI
VectorByHand2	67.33 ± 0.00 (-0.15%)	62.30 ± 0.00 (+0.02%)	12.8% (-1.33 pp)	39.88% LS, 19.43% W, 14.10% NI
Baseline	67.44 ± 0.01	62.29 ± 0.00	14.13%	41.79% LS, 19.77% W, 12.69% NI

Long Scoreboard (LS): Warp stalled waiting for L1TEX data

Wait (W): Warp stalled waiting on a fixed latency execution; typically shows up as top contributor in highly-optimized kernels

No Instruction (NI): Warp stalled waiting to be selected to fetch an instruction or waiting on an instruction cache miss.

ALOHA Subroutine Variations: FFV1_0

Routine	Wall Time [us]
NoFact	2.38 ± 0.01
FactByWFact	2.39 ± 0.01
VectorFirst	2.37 ± 0.01
VectorByHand	2.37 ± 0.01
VectorByHand2	2.37 ± 0.01
Baseline	2.39 ± 0.01