61908697 佐々木良輔

cpu kernel と gpu kernel についてそれぞれ 5 回時間を計測した．ただし gpu kernel については Reduction をホスト側で行う場合とデバイス側で行う場合とでそれぞれ 5 回時間を計測した. 単位は全て ms である．デバイスで Reduction を行うときは `__shfl_down_sync` を用いて同一 Warp 内の部分和を集計しそれをホスト側で再度集計している．

表 1 から cpu kernel に比べて gpu kernel は 1/700 程度の時間で計算が実行できている．また Reduction をデバイス側で行った場合はそうでない場合に比べて 1 割程度高速化している．

表 1

No.	gpu kernel (Reduction in host)	gpu kernel (Reduction in device)	cpu kernel
1	2.589760	2.370848	16721.085938
2	2.568512	2.399328	16708.503906
3	2.569280	2.369952	16679.330078
4	2.680256	2.368896	16724.927734
5	2.568864	2.384512	16635.873047
平均	2.595	2.379	1.670×10^4