OPTIMAL THREAD CONFIGURATIONS FOR SOLVING RECURRENCE EQUATION WITH NON-CONSTANT PARAMETERS

AKIYOSHI WAKATANI (KONAN UNIVERSITY, KOBE, JAPAN)

Summary: The details of the GPGPU implementation of the rearrangement of array configurations for recurrence equation solver, called P-scheme, are described and its effectiveness is considered. Our method alleviates branch divergences caused by the data accesses to the adjacent data and can improve the performance. We also illustrate the effectiveness of the optimal thread configuration for the recurrence equation with constant and non-constant parameters. Our experiments with GTX 590 GPU show that the implementation of the rearrangement using the shared memory improves the performance by 200% to 300% and the validity of the policy of the thread configuration is confirmed for the constant parameter case as well as the non-constant parameter case.

Effectiveness of Optimal Thread Configurations

Effectiveness of Rearrangement of Array Configurations

A simple SOA (Structure Of Array) transformation results in many branch divergence because some calculations of this parallel scheme needs adjacent data. But, by duplicating the elements of the last index, the branch divergences can be alleviated.

In order to rearrange the array located in the global memory, the array should be fetched and stored, but one of the memory accesses is non coalesced communication. But, using the shared memory in the intermediate form and rearranging the array within the shared memory, both memory accesses can be done in a coalesced communication way.

By using the shared memory, the performance of the matrix transposition can be improved by 200% to 300%.