

# Improving Locality of Unstructured Mesh Algorithms on GPUs

András Attila Sulyok<sup>1</sup>, Gábor Dániel Balogh<sup>1</sup>, István Zoltán Reguly<sup>1</sup> and Gihan R. Mudalige<sup>2</sup>

<sup>1</sup>Pázmány Péter Catholic University, Information Technology and Bionics

<sup>2</sup>The University of Warwick, Department of Computer Science

To most efficiently utilize modern parallel architectures, the memory access patterns of algorithms must make heavy use of the cache architecture: successively accessed data must be close in memory (spatial locality) and one piece of data must be reused as many times as possible (temporal locality).

Unstructured mesh algorithms are notoriously difficult in this sense, especially due to computations that indirectly modify data, leading to race conditions. In this work we address this problem through a number of optimisations on GPUs, specifically the use of the shared memory and a two-layered colouring strategy to cache the data. We also look at different block layouts to analyse the trade-off between data reuse and the amount of synchronisation.

We developed a standalone library that can transparently reorder the operations done and data accessed by a kernel, without modifications to the algorithm by the user. Using this, we performed measurements on relevant scientific kernels from different applications, such as Airfoil, Volna, Bookleaf, Lulesh and miniAero; using Nvidia Pascal and Volta GPUs. We observed significant speedups (1.2--2.5x) compared to the original codes.