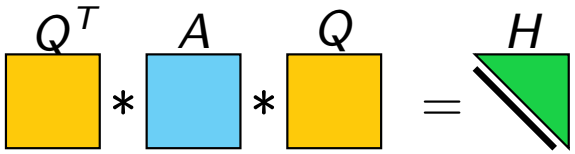
NUMA-Aware Blocked Hessenberg Reduction Using Parallel Cache Assignment

Hessenberg reduction is a similarity transformation:



where Q is an orthogonal matrix.

Challenge

- The computation is **memory-bound**.

Solution requirements

- NUMA-aware algorithm.
- High utilization of low-level cache memory.

Solution

- Using a recently published technique by Castaldo and Whaley called **parallel cache assignment** (PCA).

Key results

- Up to 8.4 times faster than LAPACK.
- Up to 2.4 times faster than ScaLAPACK.