GIRIH on A64FX

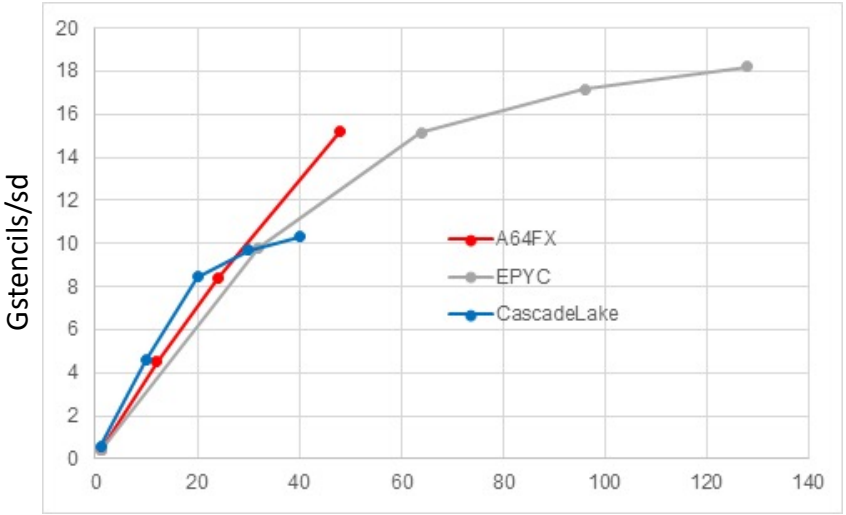
A64FX Oil and Gas webinar

Long Qu [long.qu@kaust.edu.sa]
Hatem Ltaief
David Keyes

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

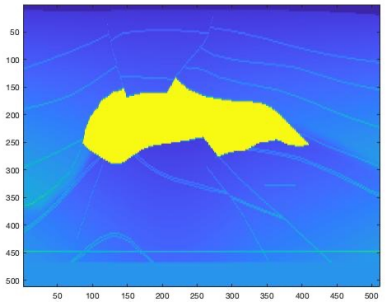
Preliminary results of GIRIH on A64FX

- MWD-TB stencil performance on A64FX.

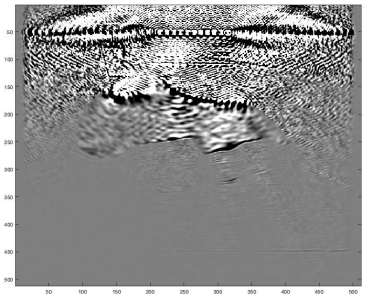


MWD-TB performance on various architectures using 1024x1024x512 grid

- MWD-TB-based RTM results on A64FX. (Salt3D dataset)



Velocity model



Migrated image (10 shots)

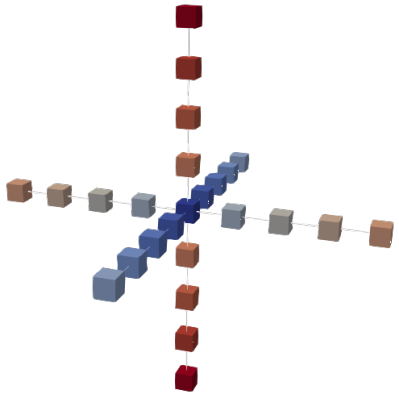
Outline

- Introduction
- Spatial blocking and MWD-Temporal blocking
- Experiments on A64FX
- Conclusion

Introduction

- GIRIH:
  - a tool to develop and analyze the performance of Multi-core Wavefront Diamond (MWD) tiling techniques for various stencil operators.

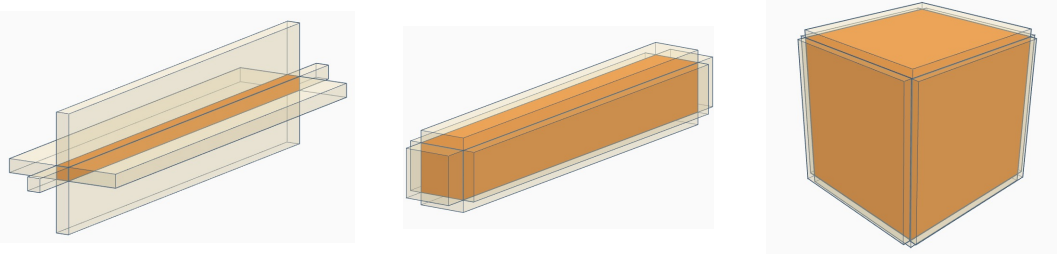
- We focus in this talk on:
  - 3D finite-difference time-domain stencil with 8^{th} order in space / 2^{nd} order in time and constant coefficients



8^{th} order in space / 2^{nd} order in time
and constant coefficients

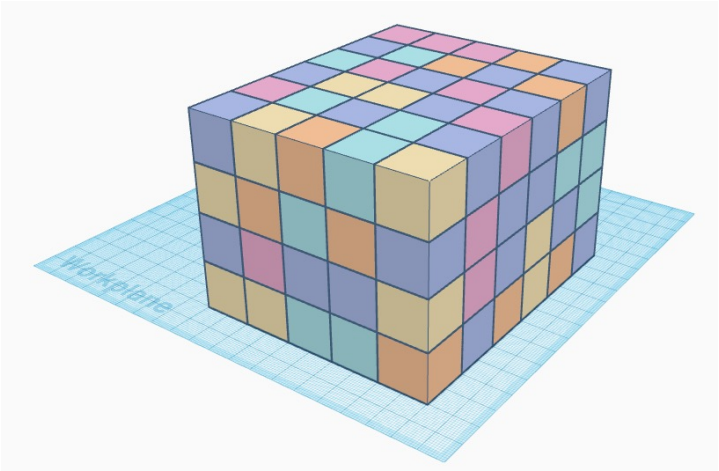
Spatial Blocking (SB)

- Enhancing cache reuse
- Most widely used technique
- Simplicity and Flexibility



1D 2D 3D

Illustration of Spatial Blocking applied to the stencil computation



3D domain tiled with 3D cube blocks and computed by 6 OpenMP threads (each thread uses a different color)

MWD-TB background

- **M**ulticore **W**avefront **D**iamond tiling **T**emporal **B**locking technique

- Advantage :
  - Reduction of memory bandwidth pressure
  - Enable natural overlapping of computation and communication

References:

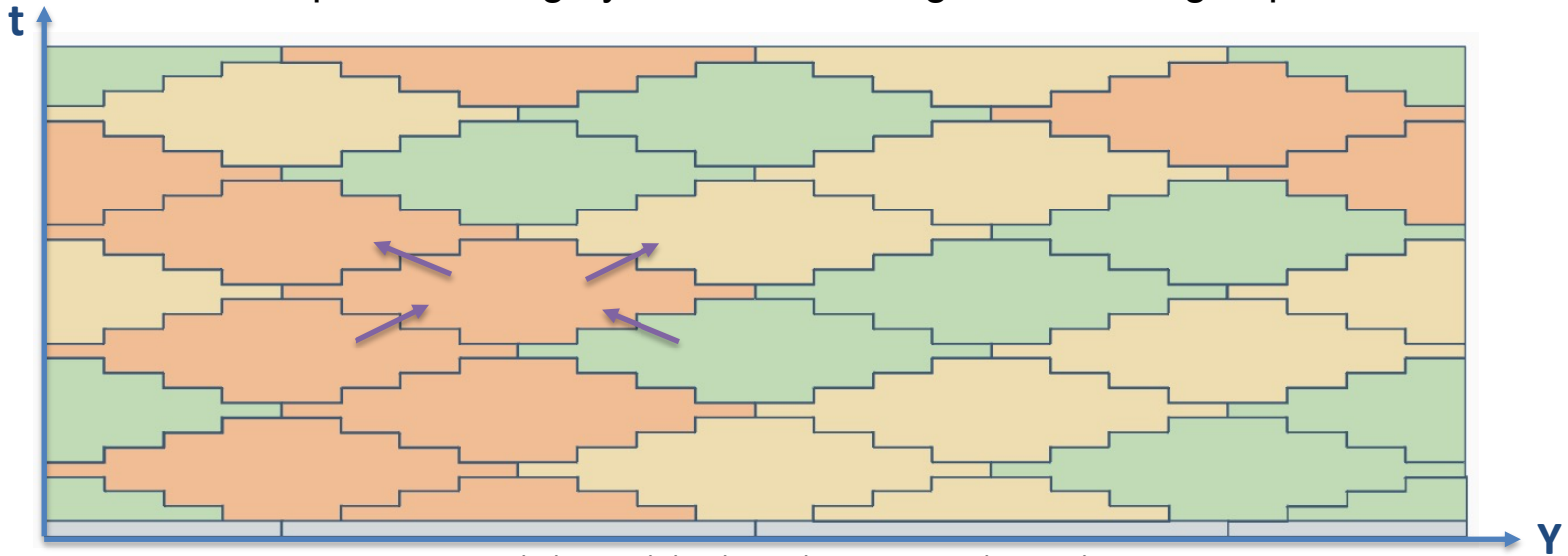
Tiling and Asynchronous Communication Optimizations for Stencil Computations, T. Malas, PhD thesis, KAUST, 2015

Multicore-Optimized Wavefront Diamond Blocking for Optimizing Stencil Updates, T. Malas, G. Hager, H. Ltaief, H. Stengel, G. Wellein & D. Keyes, SIAM J. Sci Comput. 37:C439-C464, 2015

Asynchronous computations for solving the acoustic wave propagation equation, K. Akbudak, H. Ltaief, V. Etienne, R. Abdelkhala,T. Tonellot & D. Keyes. IJHPCA. 34(4):377-393, 2020

MWD-TB background

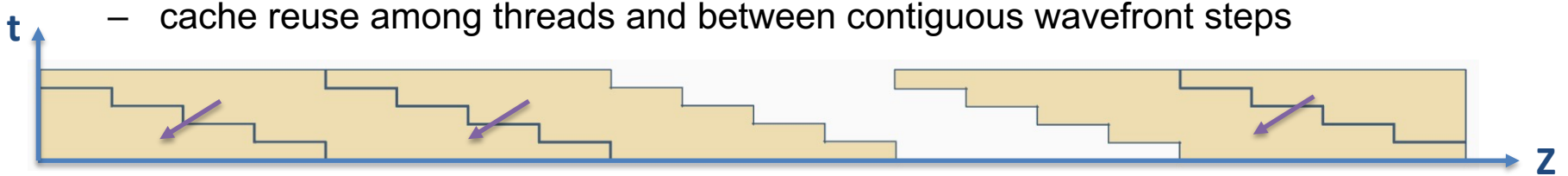
- Diamond tiling
  - outer level OpenMP, using dynamic scheduling on different groups of threads



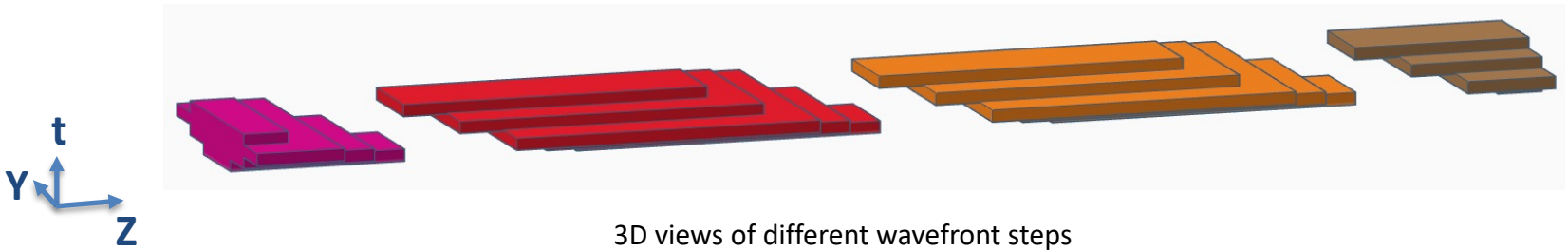
Diamond tiling and the dependency among diamonds
(color represents three diff. groups of threads)

MWD-TB background

- Wavefront parallelism
  - inner level OpenMP
  - cache reuse among threads and between contiguous wavefront steps



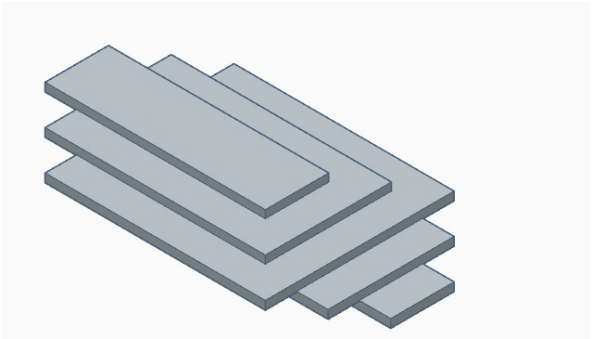
Wavefront and the dependency among different wavefront steps



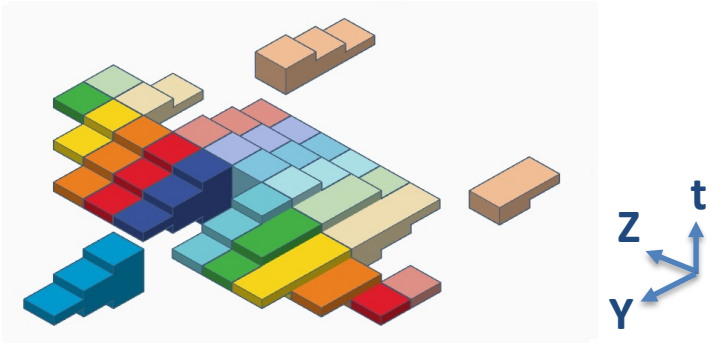
3D views of different wavefront steps

MWD-TB stencil computation

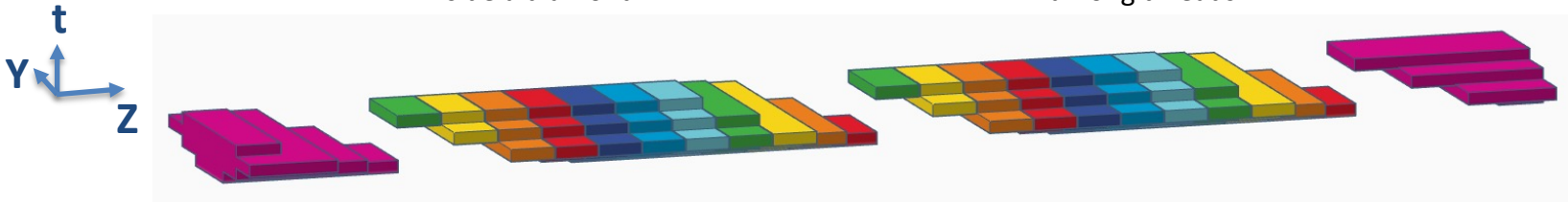
- Multi-core computation of one wavefront step
  - cache reuse inside each wavefront step
  - inner level OpenMP barrier between each time step



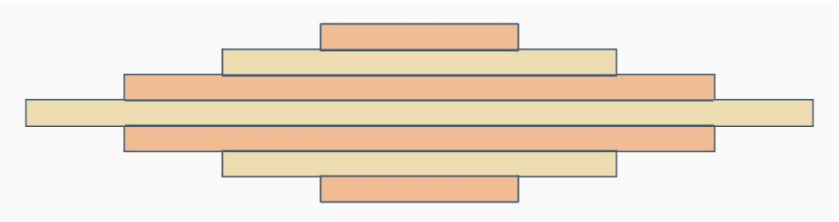
3D view of one wavefront step inside a diamond



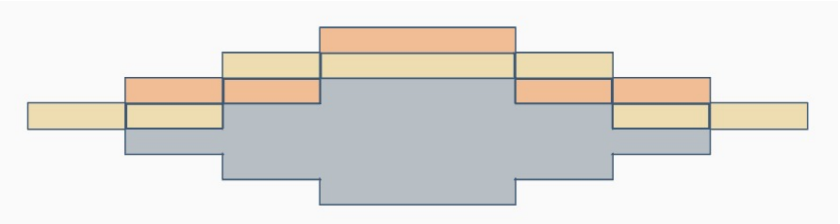
Well-balanced workload among threads



Memory during the computation



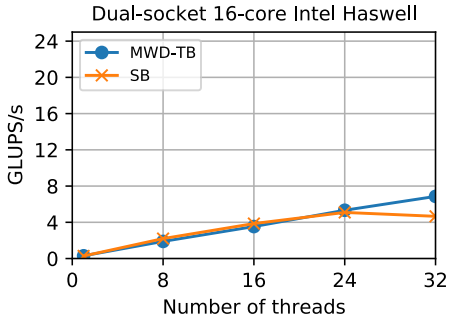
Alternated memory usage
between each time step

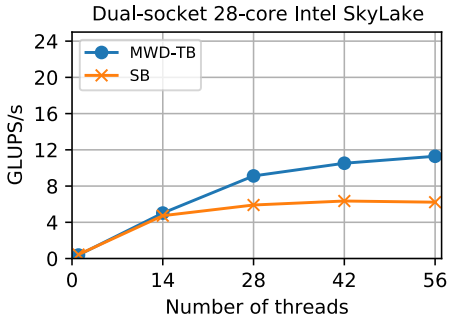


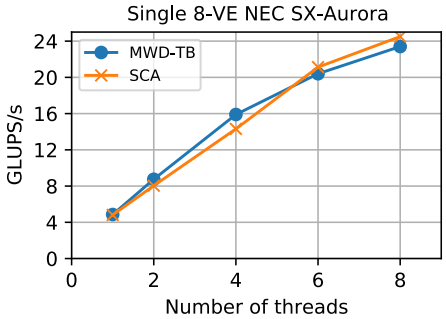
In-memory status
after the computation of the diamond

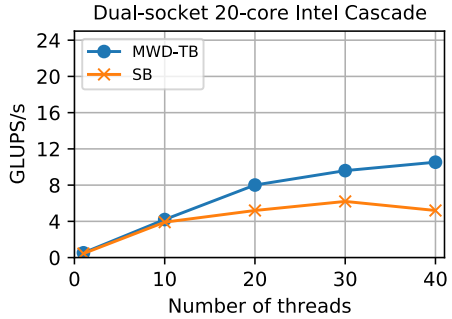
MWD-TB stencils performance

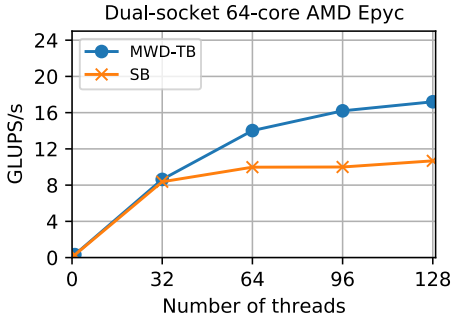
- Results of 1024 x 1024 x 512 domain size with 1000 time steps









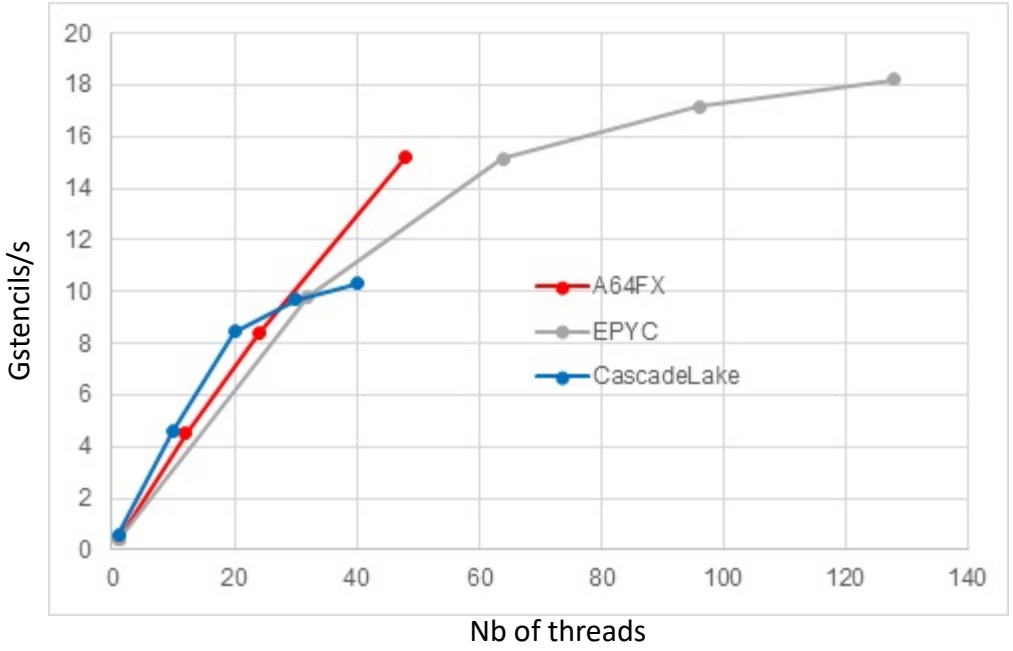


SB: spatial blocking

MWD-TB: multicore wavefront diamond temporal blocking

MWD-TB stencils performance

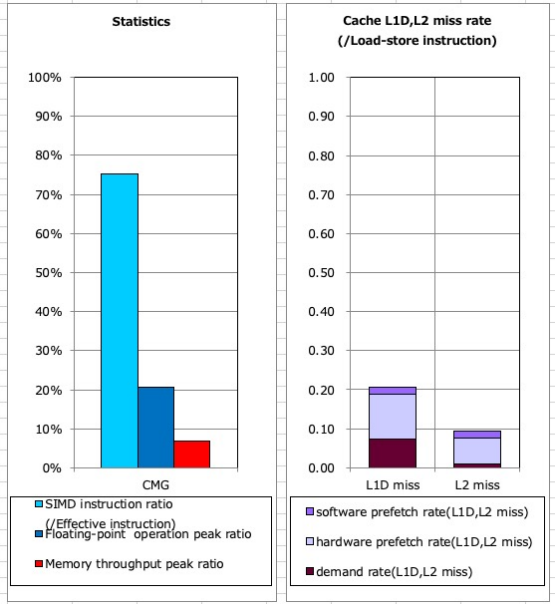
- AMD Epyc Rome shows good performance with 128 cores.
- A64FX shows similar performance than single-socket Epyc Rome. (lower cores)
- Intel Cascade lake shows better performance in 20 cores but saturates in larger cores.



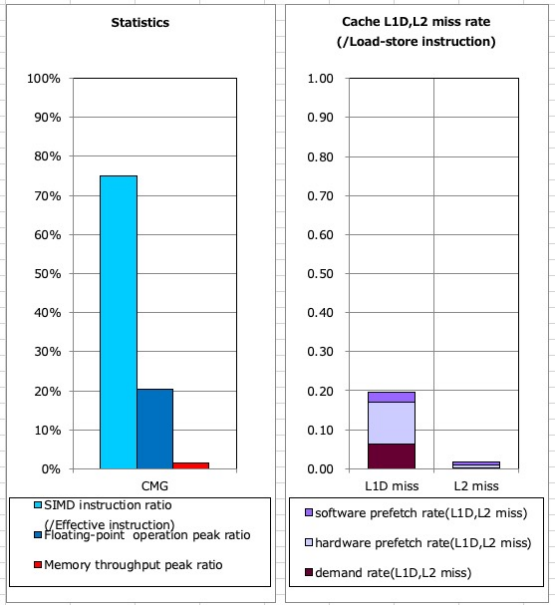
MWD-TB performance on various architectures using 1024x1024x512 grid

Comparaison SB and MWD-TB (single-thread)

- On A64FX, MWD-TB has lower L2 miss, and lower memory throughtput peak ratio.
- But SB (0.59 Gstencils/s) is faster than MWD-TB (0.36 Gstencils/s).



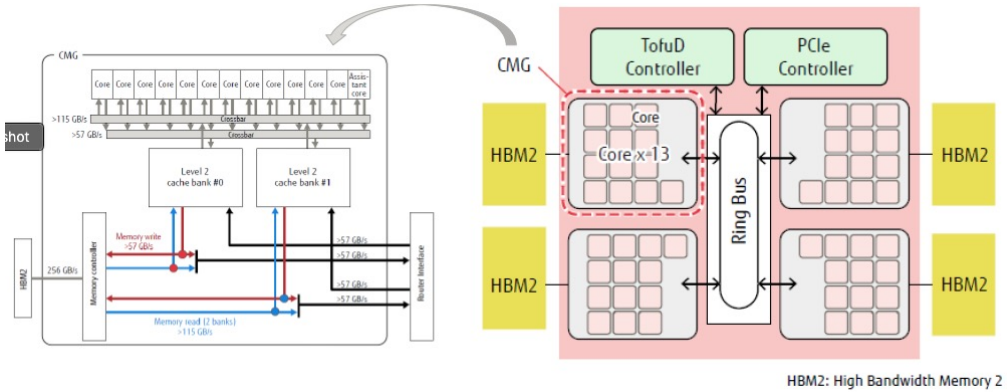
SB



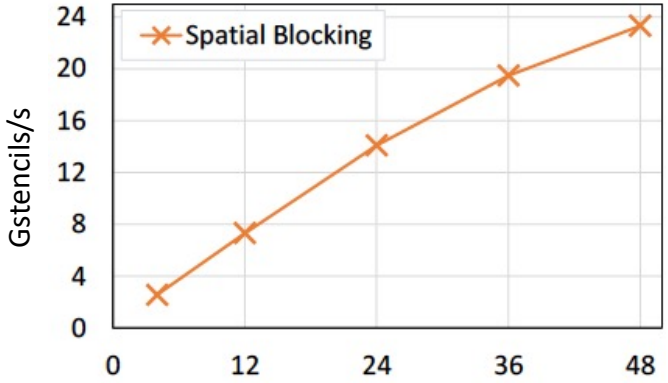
MWD-TB

Comparaison SB and MWD-TB on A64FX

- On a single CMG (12 threads), TB (4.8Gstencils/s) is still slower than SB (5.6Gstencils/s). The height of the diamond is **3** due to the limited size of L2 cache.
- On all 48 cores using 4 MPI threads, SB reachs 22.4Gstencils/s.



A64FX architecture (Fujitsu copyright)



Conclusion

- A64FX : 1MPI process on each CMG for NUMA-aware.
- SB performance is impressive on A64FX compared to other architectures.
  - the SB performance is cache-bound.
- Good scalability thanks to HBM2.
- MWD-TB does not improve on A64FX because
  - 8MBx4 L2 cache is small to make MWD-TB competitive.
  - Multi-MPI MWD-TB needs to be improved.