

TER project - 2D tiling

Atte Torri - atte.torri@universite-paris-saclay.fr

Matrix multiplication with 2D tiling

Doing a matrix multiplication taking into account possible transpositions, with A, B, and C tiled in 2D.

Note: full operation is $C = \alpha \times A^{N/T} B^{N/T} + \beta \times C$

B _{1,1}	B _{1,2}
B _{2,1}	B _{2,2}

A _{1,1}	A _{1,2}
A _{2,1}	A _{2,2}

C _{1,1}	C _{1,2}
C _{2,1}	C _{2,2}

Matrix multiplication with 2D tiling

Important note: Matrix multiplication should still work for last weeks 1D case! In the presentation/report I expect comparisons between them.

Tasks

- StarPU permits to disable sequential consistency and manually specify dependencies
- Multiple approaches to wait for tasks to complete or data to be available
- Possibility to have multiple implementations of a kernel and choose according to capabilities of the machine or the problem requirements (codelet field `can_execute`)
- Tasks can be attributed to parallel workers using codelet field type and `max_parallelism` (this might be in part 2 of the TER)

Data management

- Resizing data handle buffers in tasks dynamically
- Data allocation using `starpu_malloc`
- Controlling StarPUs data caching
- Access data from data handle with `starpu_data_acquire` and release it to StarPU with `starpu_data_release`
- Prefetching with `starpu_data_idle_prefetch_on_node` (useful to exclude data transfer time from benchmarks)
- Possible to advise StarPU on data usage needs
- Custom data interfaces
- Multi-format interfaces
- Specifying the node in which data should be for a codelet

Next

Continue working on the TER project.

```
ssh qdcster_XX@chome.metz.supelec.fr  
salloc --partition gpu_tp_resa --reservation M1QDCS_TERSTARPU13  
--exclusive --time 4:00:00
```

For next week read at least chapters 24, 25, 38 of the StarPU documentation

<https://github.com/TER-StarPU/ter-starpup-gemm/blob/main/docs/starpup-documentation.pdf>