

4 - 8 May 2020

Refactoring the memory access pattern to improve computational performance in NEMO

I.Epicoco, F. Mele, S. Mocavero, M. Chiarelli, A. D'Anca, G. Aloisio



OS4.8 - Numerical modelling of the ocean: new scientific advances in ocean models to foster exchanges within NEMO community and contribute to future developments



NEMO Ocean Model

Single core performance of the NEMO model is limited by memory access and poor exploitation of vector processing units on modern HPC architectures

The analysis of the memory access pattern shows that many repeated accesses occur for reading values not available in cache -> high rate of cache miss!!!



Goal of the work: enhance the exploitation of the cache memory of the modern parallel architectures through the loop fusion approach

Single core performance - Loop fusion

• Loop fusion technique aims at better exploiting the cache memory by fusing DO loops together

```
D0 j=1, n-1
    D0 i=1, n
        b (i,j) = in(i,j+1) - in(i,j)
    END D0
END D0
D0 j=2, n-1
    D0 i=1, n
        out (i,j) = b(i,j) - b(i,j-1)
    END D0
END D0
END D0
```

- advantages: reduction of cache misses and reduction of the memory footprint
- disadvantage: due to data dependencies redundant operations are needed

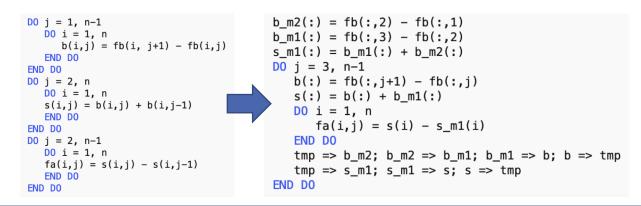
Single core performance - Loop fusion

- Loop fusion approach has been applied on the NEMO MUSCL advection kernel
- Three different levels of fusion have been implemented

□prototype1: has the maximum level of fusion with redundant operations

□prototype2: introduces the buffers rotation¹ in the outer loop

□prototype3: uses the buffers rotation in the outer and middle loop



¹buffers rotation technique avoid redundant operations by adopting pointers to arrays and implementing a rotation at each loop iteration as shown in the figure

First approach – prototype1

- The halo exchange is moved before all DO-loops (at the beginning of the routine)
 - The halo region needs to be extended up to two halo lines
- The advective trend is computed for each single (ji, jj, jk, jn) grid cell within a single big DO-loop
 - This approach implies also a duplication of the calculus up to a factor 3

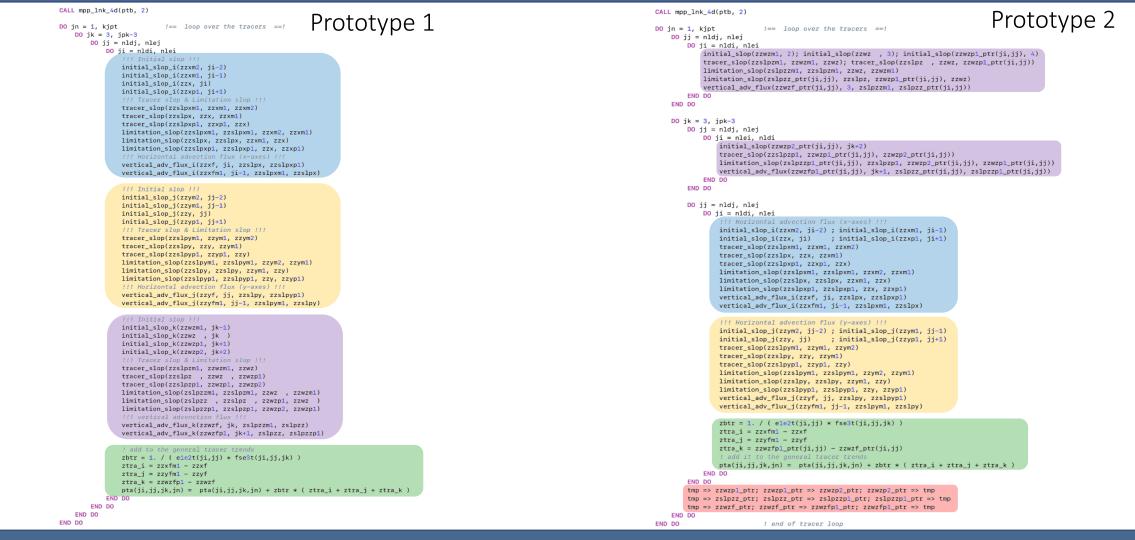
First approach – prototype1



Second approach – prototype2

- Along the vertical direction:
 - The advective flux at level jk=1 is computed before the loop over jk (we call this flux F_{jk-1})
 - Inside the jk loop, the flux at level jk (which we call F_{ik}) is computed
 - We use F_{jk-1} and F_{jk} to calculate the advective trend at level jk and to update the RHS variable at level jk
 - Before incrementing the jk level, we update the flux at level jk-1: $F_{jk} \rightarrow F_{jk-1}$
- This approach reduces the number of redundant operations, but it introduces a data dependencies in the jk loop, hence the jk loop can not be vectorized, neither executed in parallel.

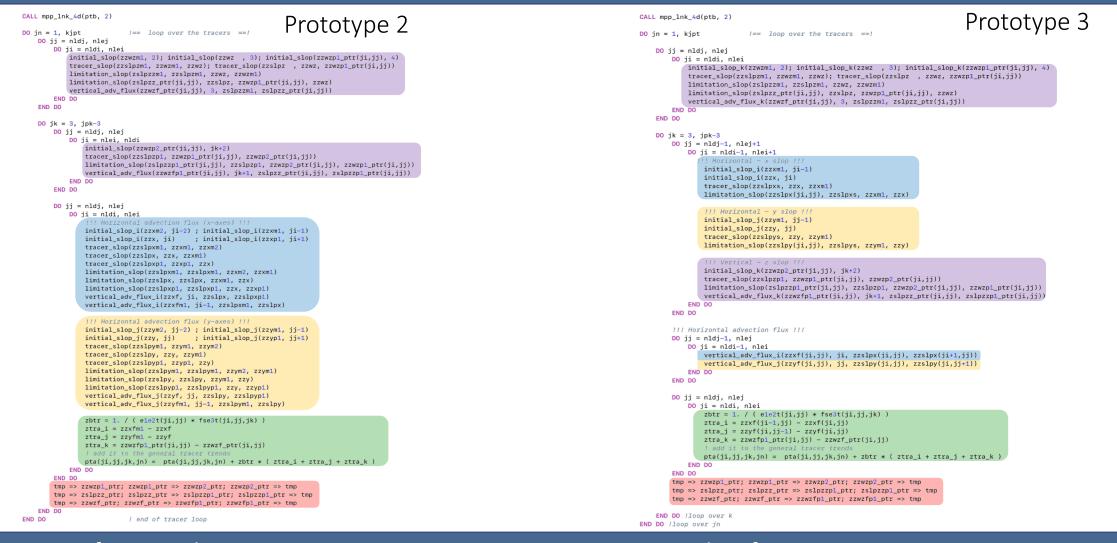
Second approach – prototype2



Third approach – prototype3

- Along the horizontal direction:
 - In the jk loop, before updating the RHS variable with the advective trend, the fluxes for the whole horizontal domain are computed
 - The fluxes are hence used to compute the advective trend and to update the RHS variable
- This approach further reduces the number of redundant operations.

Third approach – prototype3

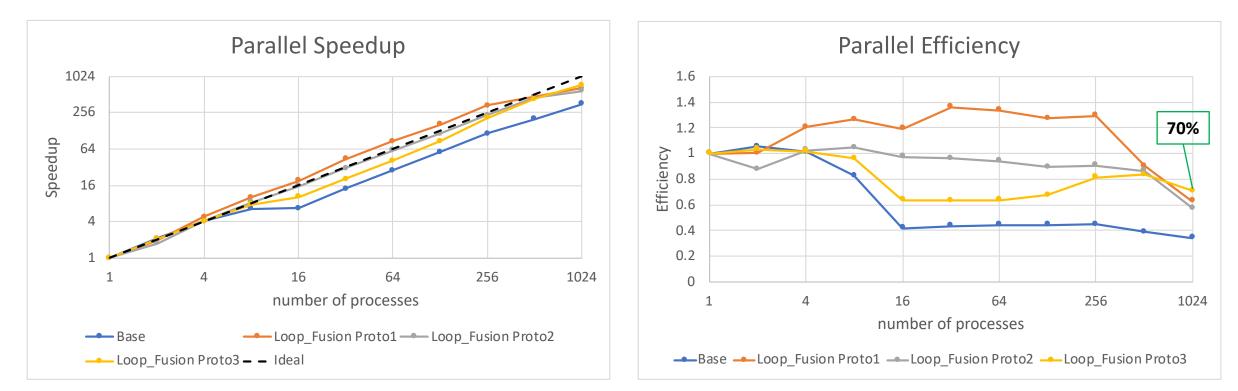


• Smallest sub-domain (with 1024 cores): 74 x 51 x 31 points

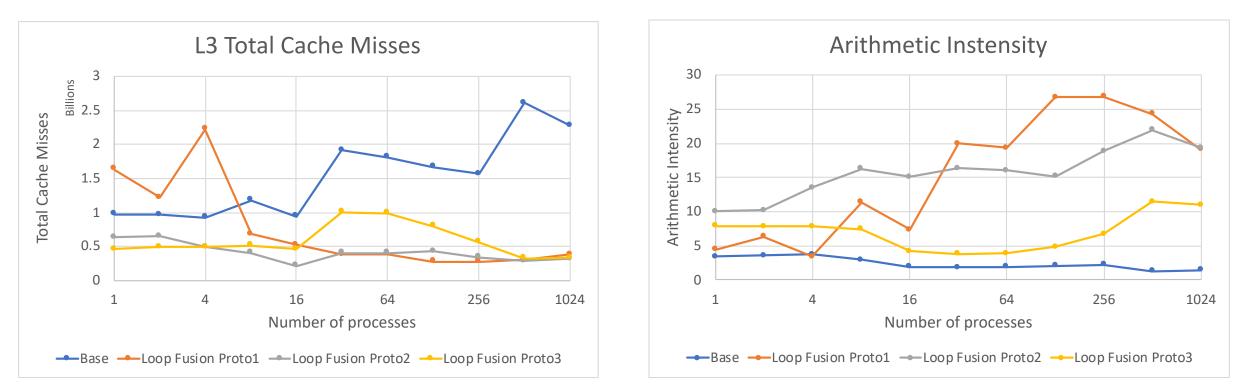
• Global domain: 2240 x 1500 x 31 points

- *Test executed on Intel Xeon based Architecture 16 cores per node*
- Prototypes Speedup **Elapsed Time** 2.5 100 2 Elapsed time (sec) 10 d 1.5 Speedup 1 0.1 0.5 0.01 64 16 32 128 256 1 Δ 16 256 1024 2 Δ 8 64 512 1024 number of processes number of processes ---Base --- Loop Fusion Proto1 --- Loop Fusion Proto2 --- Loop Fusion Proto3 ■ Loop Fusion Proto1 Loop Fusion Proto2 ■ Loop Fusion Proto3

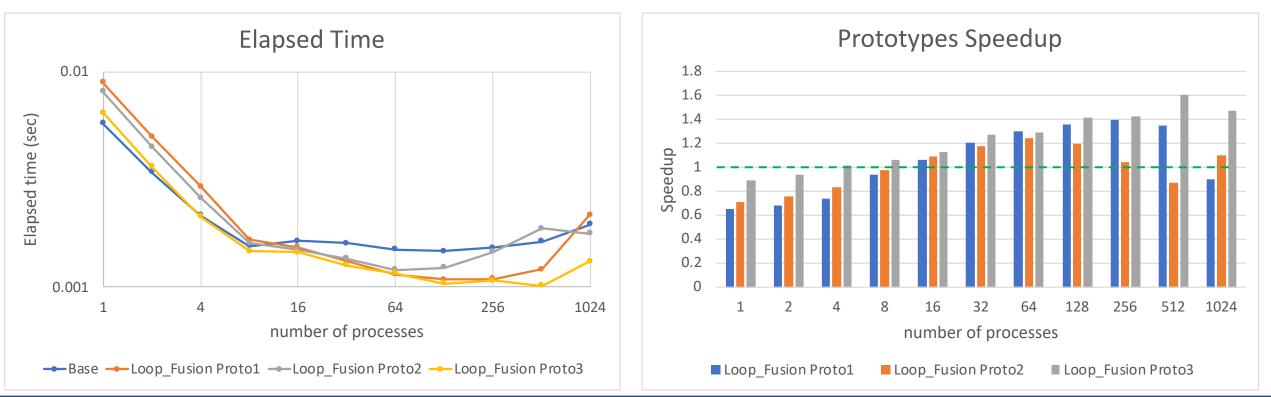
- Global domain: 2240 x 1500 x 31 points
- Smallest sub-domain (with 1024 cores): 74 x 51 x 31 points



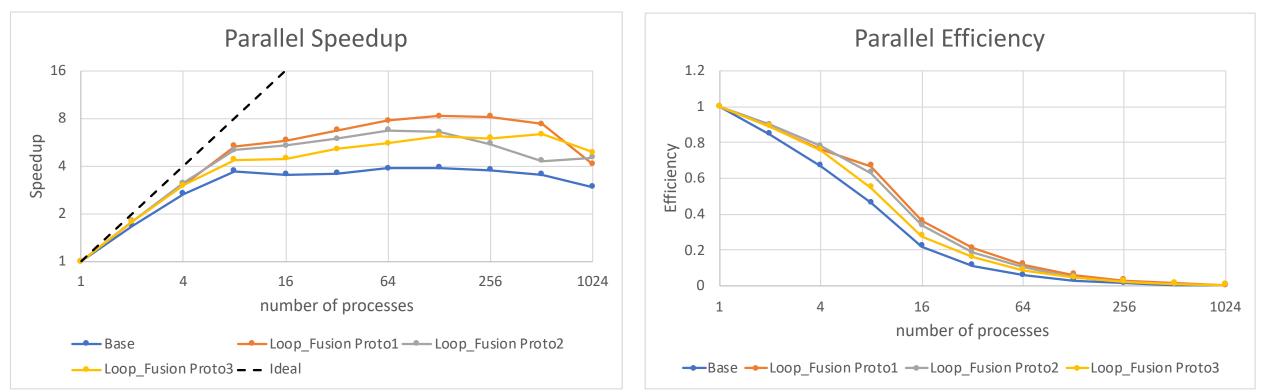
- Global domain: 2240 x 1500 x 31 points
- Smallest sub-domain (with 1024 cores): 74 x 51 x 31 points



- Global domain: 70 x 46 x 19 points
- Sub-domain with 64 cores: 13 x 10 x 19 points



- Global domain: 70 x 46 x 19 points
- Sub-domain with 64 cores: 13 x 10 x 19 points



- Prototypes 1 and 2 provide a good improvement up to 256 cores then the redundant operations lead to a loss of performance
- Prototypes 3 improves parallel efficiency by ~30% on 1024 cores
- This approach enhanced the vectorization level and the cache reuse, reducing L3 Total Cache Misses by ~80% on 1024 cores
- Loop-fusion is strictly linked to the computing architecture —> A fully portable performance improvement can be ensured by the adoption of a DSL.