

# Dynamic Load Balancing for BLonD-MPI (II)

Konstantinos Iliakis

Supervisors:

PhD Candidate  
CERN, CH - NTUA, GR  
*konstantinos.iliakis@cern.ch*

Dr. Helga Timko, CERN  
Dr. Sotirios Xydis, NTUA  
Dr. Dimitrios Soudris, NTUA

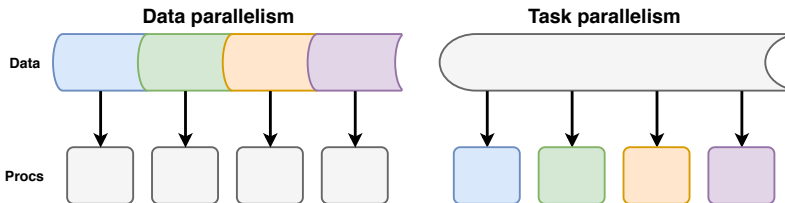


June 14, 2019

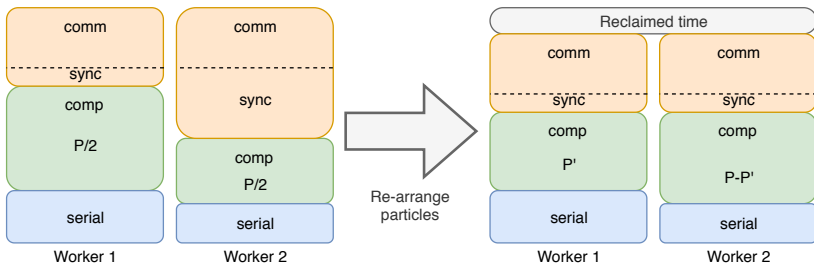
# Problem Description

## Challenge

- Node allocations more favourable than others.
- Workers faster than others.
- Load imbalance due to Task-Parallelism.
- Result: Huge spread in run-time across workers.
- Problem: Final run-time is the run-time of slowest worker.



# Model Assumptions/ Conclusions



- ①  $T_{comm}^i(p) = \text{constant}$ ,  $T_{serial}^i(p) = \text{constant}$ .
- ②  $T_{comp}^i(p) = a^i \times p^i + b^i$ , where  $p$ : #particles,  $i$ : worker id.
- ③ Perfect LB  $\Leftrightarrow T_{serial}^i + T_{comp}^i = t \Leftrightarrow T_{sync}^i \rightarrow \min$ .
- ④ A worker exhibits the same behavior for long periods.
- ⑤  $T_{serial}^i$ ,  $T_{comp}^i$ ,  $a^i$ ,  $b^i$  can be calculated.

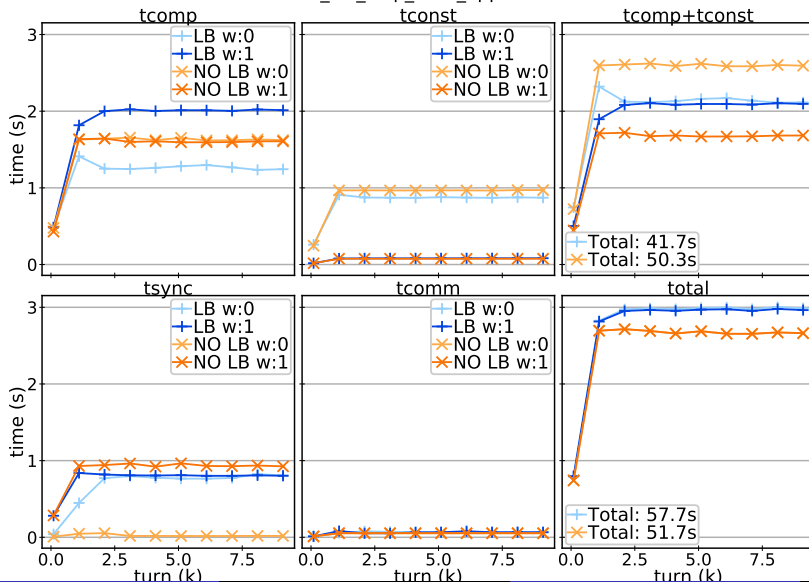
# Evaluation

## Set-up

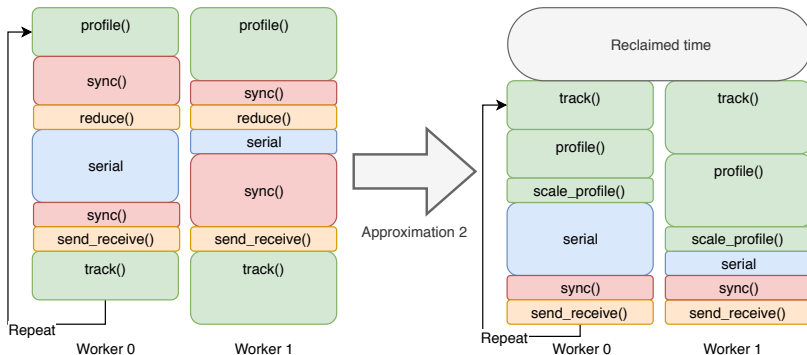
- LHC test-case.
- 2 MPPB, 2 bunches, 10K turns.
- 2 Workers, 4 threads each.
- With Task-Parallelism, no approximation.
- Run in the Haswell platform, not in the cluster.
- Legend: Blue with LB, Orange No LB.

# First attempt

LHC\_2w\_4Mp\_10Kt\_approx0



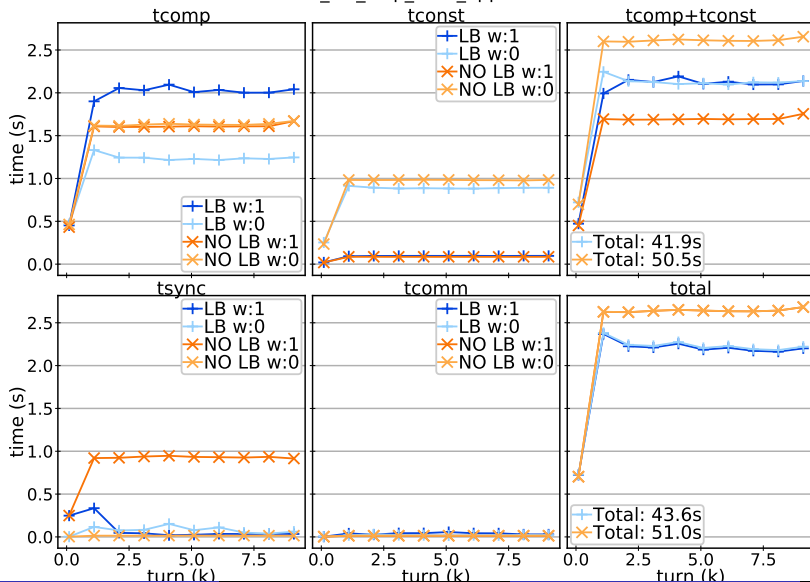
# What went wrong?



- LHC test-case with Task-Parallelism.
- Approximation 2 to avoid to sync() before reduce().

# And finally..

LHC\_2w\_4Mp\_10Kt\_approx2



# New Blondmath functions

- Added `bm.where()`, similar to `np.where()` ([link to PR](#)).
- Added `bm.rfft()`, `bm.irfft()`, `bm.rfftfreq()` ([link to PR](#)).
- Readme with instructions on how to enable the FFTW ffts ([link](#)).



# Questions

