25–26 Nov. 2025

ANALYSIS AND OPTIMIZATION OF MEMORY REQUIREMENTS FOR STELLA (AND) CONTINUOUS PERFORMANCE MONITORING FOR GVEC

Joan Vinyals Ylla-Català joan.vinyals@bsc.es
Valentin Seitz valentin.seitz@bsc.es
Marta Garcia Gasulla marta.garcia@bsc.es

3nd Annual Meeting of EUROfusion HPC ACHs

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

25–26 Nov. 2025

STELLA

Valentin Seitz

3nd Annual Meeting of EUROfusion HPC ACHs

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Goal: Enabling multiscale simulations

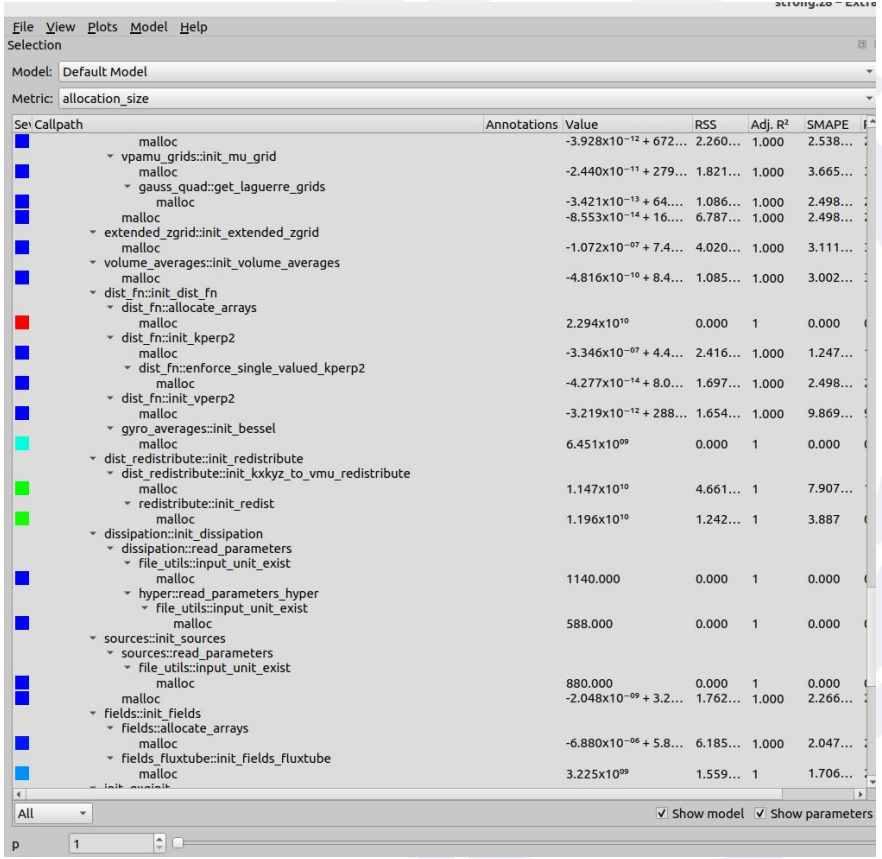
Stella is MPI only.
Scalability in terms of real-space resolution of Stella was majorly hindered by memory usage.

→ Detailed analysis of memory scaling of high resolution input case:
Run different configurations, measure memory allocations per function and build analytical model with Extra-P

- fields::allocate_arrays: 0.5GB*p
- init_g::ginit_noise: 2GB + 0.09GB*p
- dist_fn::init_kperp2: 0.04GB*p

→ linear dependence on p means memory replicated across processes

The models estimated the memory requirements for target run



Goal: Enabling multiscale simulations

Another Problem in Stella: Available Parallelism with MPI as it's often used in velocity space parallel mode

Distributed velocity space:

`nvpa * nmu * nspec =`

(2x48) x 12 x 2 = 2304 processes max -> 20 Nodes in MN5 when every process has exactly one velocity space point. (not really ideal?)

Distributed real space:

`naky * nakx * nzed * ntubes * nspec =`

1115 * 558 * 64 * 1? * 2 = 80e6 processes -> More than full MN5 of parallelism

Strategy: Port Parts to OpenMP

The memory models indicate that 2 processes per node would work on current platforms, so we chose to do add OpenMP parallelism.
Focus: timestepping

~60% of runtime for timestep OpenMP parallelized and partially verified before major refactor upstream.

Greatest missing region is and advance_implicit(20%)

Whats next?

Merge OpenMP porting upstream.

With OpenMP mostly in the timestep, we find that the computation of the response terms and the corresponding matrices is a major bottleneck. While this is initially only done in the initialization, it's also required once the time step changes (which can happen quite often).

The next step is to find solution strategies to initialization of response terms (varying matrix sizes).

25–26 Nov. 2025

GVEC

Joan Vinyals Ylla-Català

3nd Annual Meeting of EUROfusion HPC ACHs

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Background

- **Code**: GVEC (Galerkin Variational Equilibrium Code) is an open-source software for the generation of three-dimensional ideal magnetohydrodynamic (MHD) equilibria.
- **Programming**: Fortran, OpenMP, MPI
- **Inputs**: hmap_RZ, hmap_axisNB
- **Analysis**: Scalability of OpenMP threads, with two inputs.
- **Analysis platform:** MareNostrum 5 (GPP partition, 112 CPUs/node, more info)

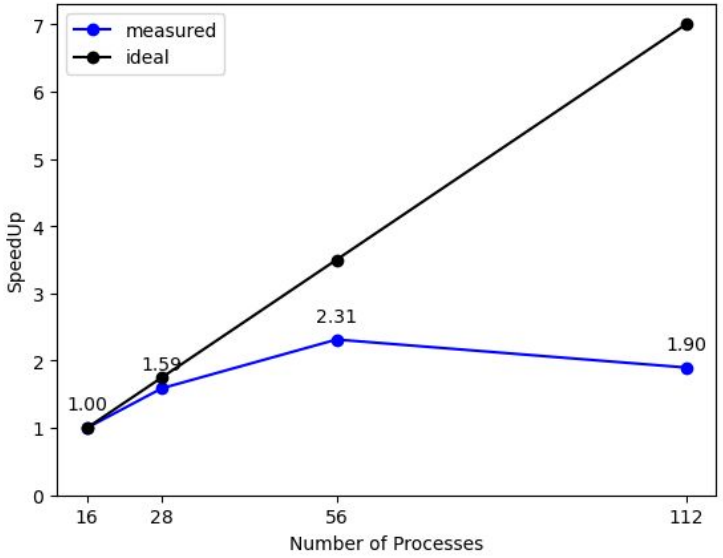
Structure & Focus of analysis - Input: hmap_RZ

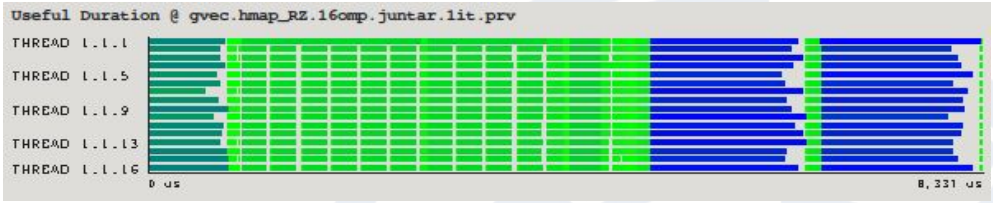
We observe an iterative pattern

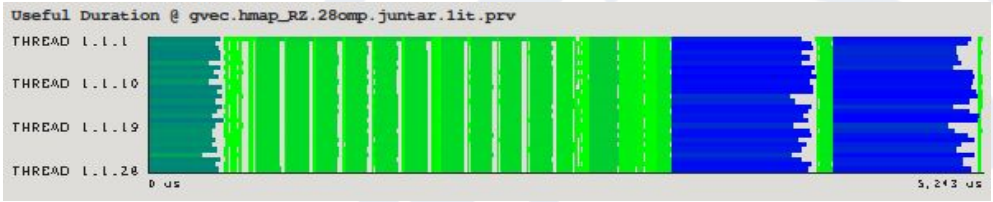
Within this iterative pattern we observe a sub-structure, which correspond to different OpenMP Parallel regions.

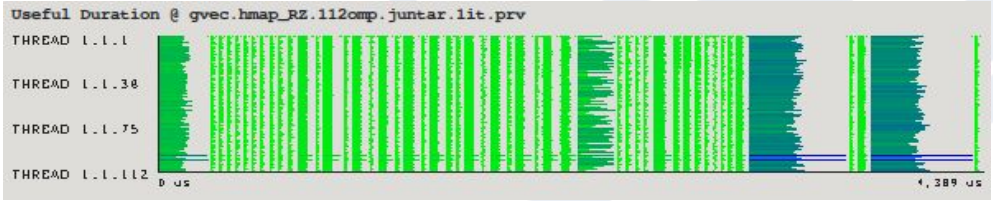
We observe that the repetitive pattern doesn't change across iterations, therefore, our FoA will be one iteration.

Scalability - Input: hmap_RZ

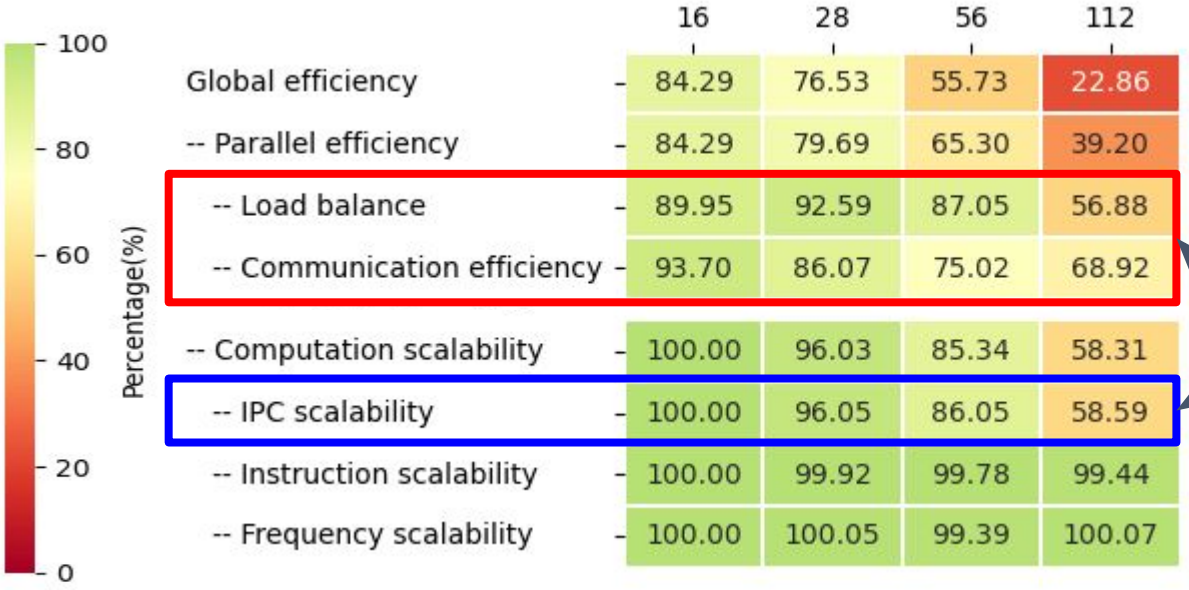








Efficiency Metrics - Input: hmap_RZ

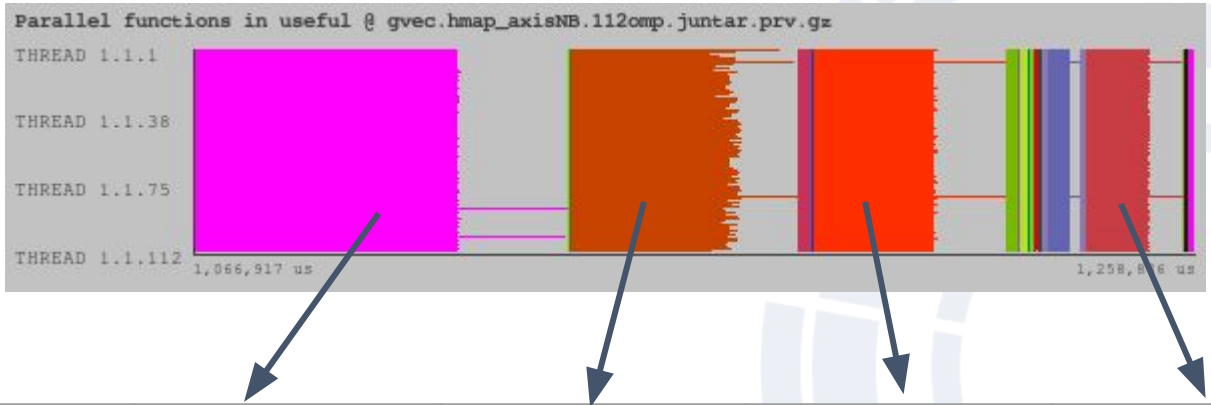


Further analysis: showed that the load imbalance was caused by **differences in IPC** across processes.

Efficiency metrics - hmap_RZ vs hmap_axisNB

Input	hmap_RZ				hmap_axisNB			
NUM THREADS	16	28	56	112	16	28	56	112
Global efficiency	84.29	76.53	55.73	22.86	0.89	0.88	0.88	0.58
-- Parallel efficiency	84.29	79.69	65.30	39.20	99.11	99.05	97.77	66.13
-- Load balance	89.95	92.59	87.05	56.88	99.73	99.38	99.11	83.66
-- Communication efficiency	93.70	86.07	75.02	68.92	99.37	99.67	98.65	79.04
-- Computation scalability	100.00	96.03	85.34	58.31	0.90	0.89	0.90	0.87
-- IPC scalability	100.00	96.05	86.05	58.59	83.52	83.42	83.37	82.05
-- Instruction scalability	100.00	99.92	99.78	99.44	1.06	1.06	1.06	1.06
-- Frequency scalability	100.00	100.05	99.39	100.07	101.63	101.01	102.02	100.53

Load Balance - Input: hmap_axisNB

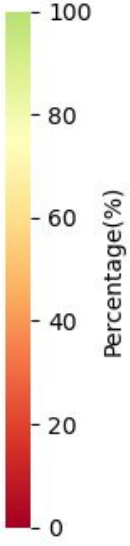


Parallel Eff.	70.41%	71.12%	63.05%	64.45%
- Communication Eff.	99.25%	99.59%	99.61%	99.51%
- Load Balance Eff.	70.94%	71.41%	63.30%	64.77%

Further analysis: that all unbalanced regions were caused by variations in IPC across processes.

Background

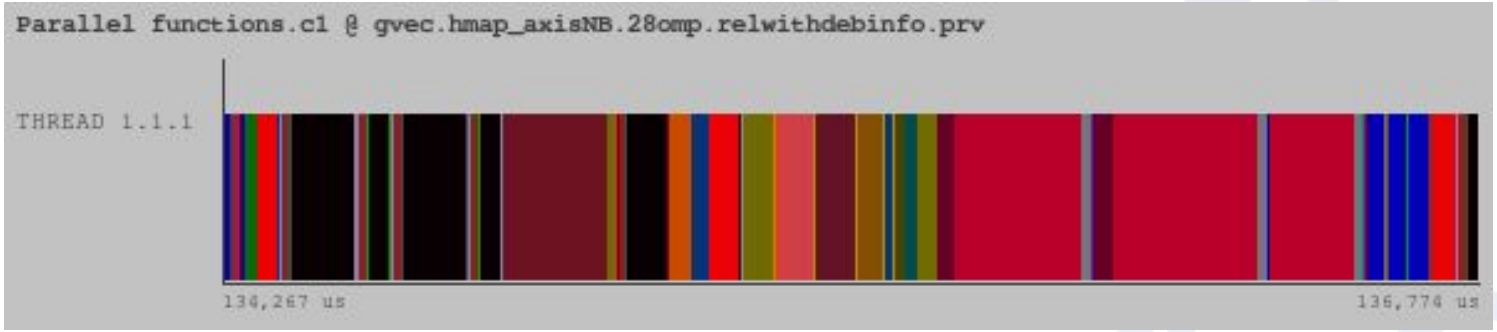
- At this point, based on the analysis comparison of both inputs, the code developer implemented some changes, to improve its performance.



Input	baseline				improved			
NUM THREADS	16	28	56	112	16	28	56	112
Global efficiency	98.78	98.77	94.38	96.18	9822.29	8814.66	6805.19	2923.91
-- Parallel efficiency	98.78	99.11	96.03	96.91	83.99	78.57	68.73	44.10
-- Load balance	99.23	99.54	98.80	99.15	90.68	91.07	93.10	69.01
-- Communication efficiency	99.55	99.57	97.19	97.74	92.63	86.28	73.82	63.91
-- Computation scalability	100.00	99.66	98.28	99.25	11694.65	11218.67	9901.93	6629.57
-- IPC scalability	100.00	99.95	99.77	99.04	119.78	114.80	101.59	68.11
-- Instruction scalability	100.00	100.00	100.00	99.99	9957.53	9952.56	9936.81	9902.55
-- Frequency scalability	100.00	99.71	98.51	100.22	98.05	98.19	98.09	98.29

Continuous performance monitoring - Manual Instrumentation

- We used the code annotations they already had implemented in the code to plug in TALP DLB.
- To provide the user with more insight on regions performance.



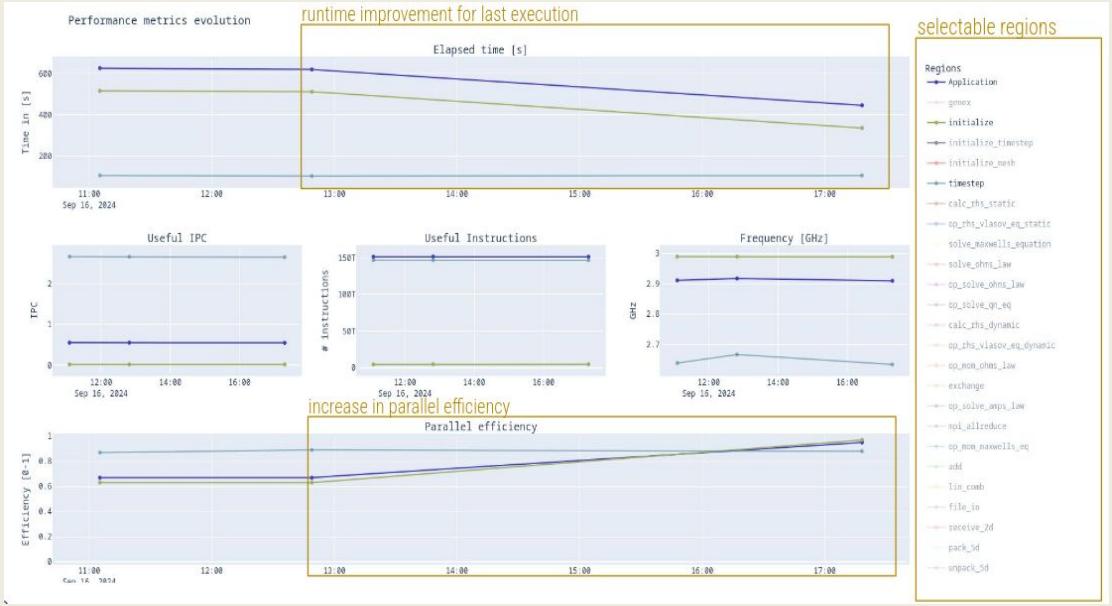
Continuous performance monitoring - TALP Pages

Metrics	8xOpenMP	36xOpenMP	72xOpenMP
Global Effiency	0.95	0.65	0.43
- Parallel efficiency	0.95	0.68	0.48
-- OpenMP Parallel efficiency	0.95	0.68	0.48
--- OpenMP Scheduling efficiency	1	1	1
--- OpenMP Load balance	0.95	0.68	0.48
--- OpenMP Serialization efficiency	1	1	0.99
- Computation Scalability	1	0.96	0.89
-- Instructions scaling	1	1	1
-- IPC scaling	1	1.03	1.03
-- Frequency scaling	1	0.93	0.87
Useful IPC	1.98	2.04	2.03
Frequency [GHz]	2.25	2.1	1.95
Elapsed time [s]	44.02	14.2	10.85

- With TALP Pages we get performance scaling efficiency tables for every commit like the one above.

Continuous performance monitoring - Next steps

Example from another code:



More examples:
https://tu-dresden.de/zih/das-department/ressourcen/dateien/toolsworkshop/2024_09_19-ValentinSeitz.pdf?lang=en

Integrate TALP Pages to the GitLab CI/CD infrastructure of GVEC, to provide region specific efficiency and execution time evolutions across commits.

25–26 Nov. 2024

ANALYSIS AND OPTIMIZATION OF MEMORY REQUIREMENTS FOR STELLA (AND) CONTINUOUS PERFORMANCE MONITORING FOR GVEC

Joan Vinyals Ylla-Català joan.vinyals@bsc.es
Valentin Seitz valentin.seitz@bsc.es
Marta Garcia Gasulla marta.garcia@bsc.es

3nd Annual Meeting of EUROfusion HPC ACHs

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.