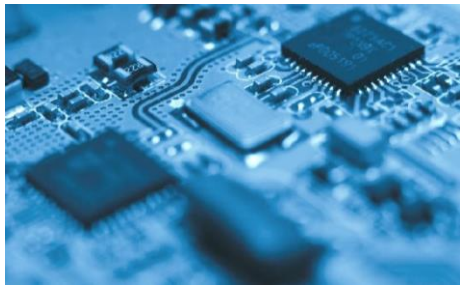
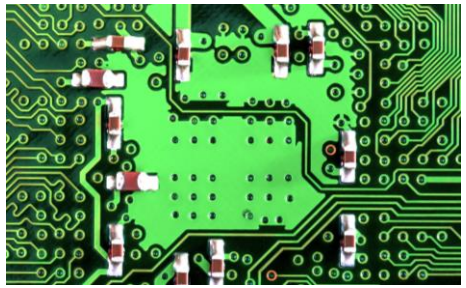
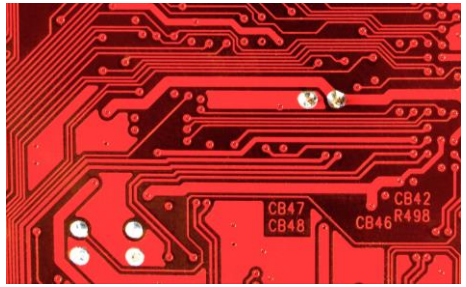
Funded by the European Union

Co-ordinated by ECMWF



ESCAPE 2: Energy-efficient Scalable Algorithms for Weather and Climate Prediction at Exascale

Andreas Mueller, Giovanni Tumolo, Willem Deconinck, Nils Wedi, Peter Bauer, et al.

ECMWF (European Centre for Medium-Range Weather Forecasts)

ESCAPE: ECMWF, MeteoSwiss, DMI, RMI, NVIDIA, METEO FRANCE Toujours un temps d'avance, ICHEC Irish Centre for High-End Computing, Deutscher Wetterdienst Wetter und Klima aus einer Hand, Optalysys, Bull atos technologies, PSNC, Loughborough University

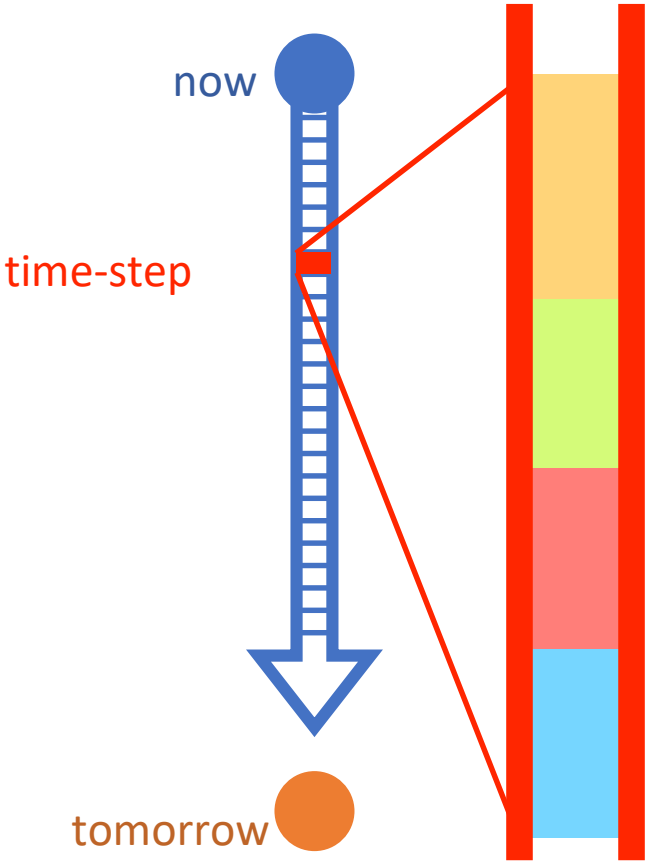
ESCAPE2: ECMWF, DKRZ Deutsches Klimarechenzentrum, Max-Planck-Institut für Meteorologie, MeteoSwiss, Barcelona Supercomputing Center Centro Nacional de Supercomputación, cea, Loughborough University, RMI, Politecnico Milano 1863, MOX, DMI, cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bull atos technologies

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 671627 (ESCAPE) and No 800897 (ESCAPE2)

ESCAPE 2

Funded by the European Union

Concept behind ESCAPE

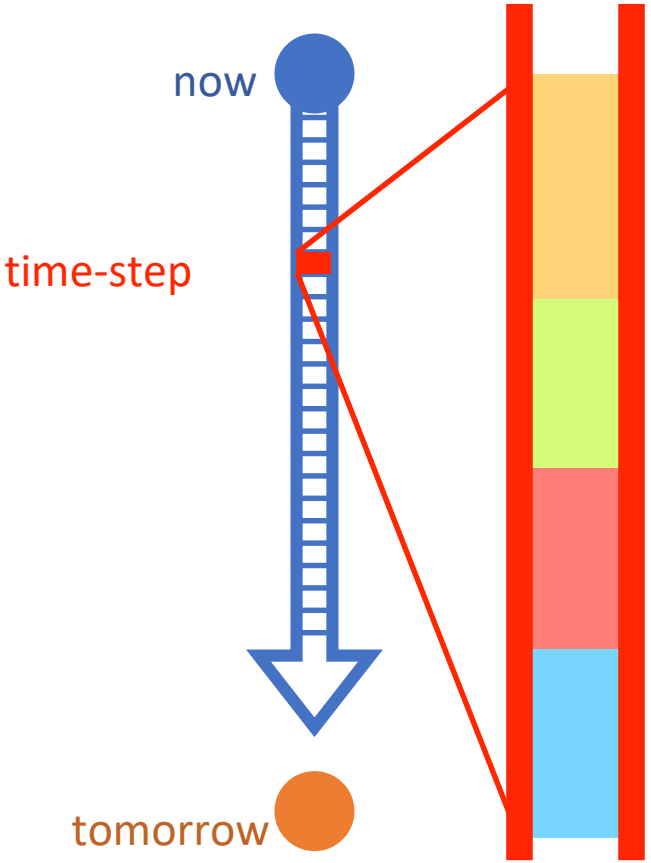


components:	challenges:
time-stepping	communication, memory
advection	halo-communication
gradient computation	expensive calculations
physics	expensive calculations

ESCAPE 2

Funded by the European Union

Dwarfs of ESCAPE 2



components:	options:	ocean		atmosphere		global		regional		D1.7		D1.8		HPCW	
discretisation	spectral transform*		✓	✓	✓							✓			
	finite volume	✓	✓	✓	✓							✓			
	discontinuous Galerkin	✓	✓	✓	✓					✓		✓			
time-stepping	multigrid elliptic solver	✓	✓	✓	✓			✓							
	fault tolerant elliptic solver	✓	✓	✓	✓					✓					
	horizontal explicit, vertical implicit	✓	✓	✓	✓							✓			
advection	semi-Lagrangian		✓	✓	✓			✓				✓			
	MPDATA*	✓	✓	✓	✓			✓				✓			
	MUSCL	✓	✓	✓	✓			✓				✓			
physics	CLOUDSC microphysics*		✓	✓	✓			✓				✓			
	ecRad radiation		✓	✓	✓							✓			
	ACRANEB2 radiation*		✓	✓	✓			✓				✓			
	machine learned radiation		✓	✓	✓					✓					

grey: work in progress, *from ESCAPE 1

work-steps for each dwarf: isolation into self-contained prototype, documentation, adaptation to different hardware, maintenance of repo

ESCAPE 2

Funded by the European Union

Dwarfs of ESCAPE 2

HPCW: Suite of weather and climate prediction benchmarks

components:	options:	ocean		atmosphere		global	regional	D1.7	D1.8	HPCW	
discretisation	spectral transform*		✓	✓	✓					✓	
	finite volume	✓	✓	✓	✓					✓	
	discontinuous Galerkin	✓	✓	✓	✓			✓		✓	
time-stepping	multigrid elliptic solver	✓	✓	✓	✓		✓				
	fault tolerant elliptic solver	✓	✓	✓	✓			✓			
	horizontal explicit, vertical implicit	✓	✓	✓	✓					✓	
advection	semi-Lagrangian		✓	✓	✓		✓			✓	
	MPDATA*	✓	✓	✓	✓		✓			✓	
	MUSCL	✓	✓	✓	✓		✓			✓	
physics	CLOUDSC microphysics*		✓	✓	✓		✓			✓	
	ecRad radiation		✓	✓	✓					✓	
	ACRANEB2 radiation*		✓	✓	✓		✓			✓	
	machine learned radiation		✓	✓	✓			✓			

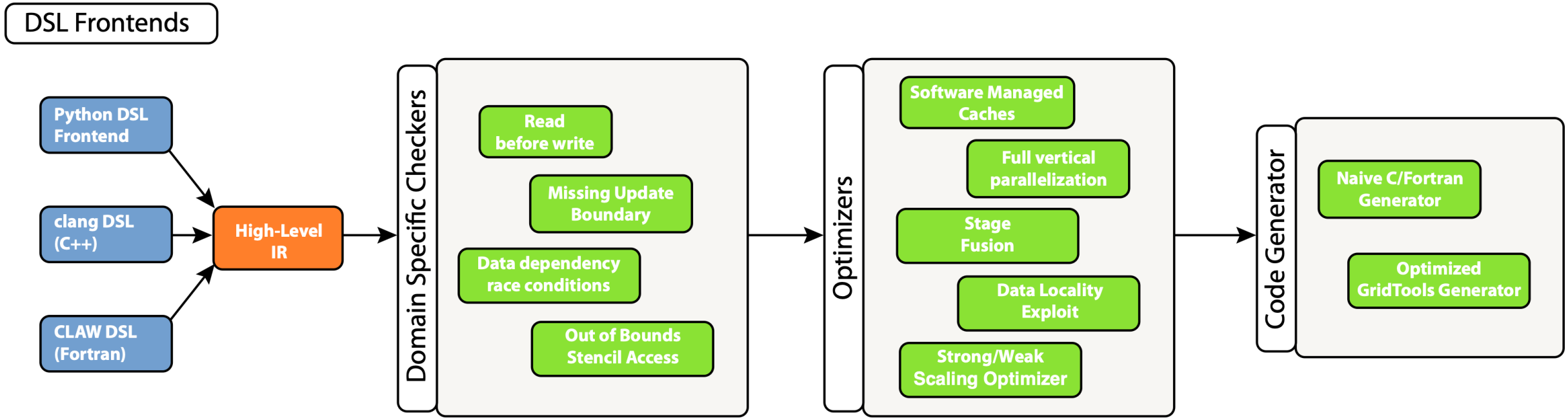
grey: work in progress, *from ESCAPE 1

work-steps for each dwarf: isolation into self-contained prototype, documentation, adaptation to different hardware, maintenance of repo

Funded by the European Union

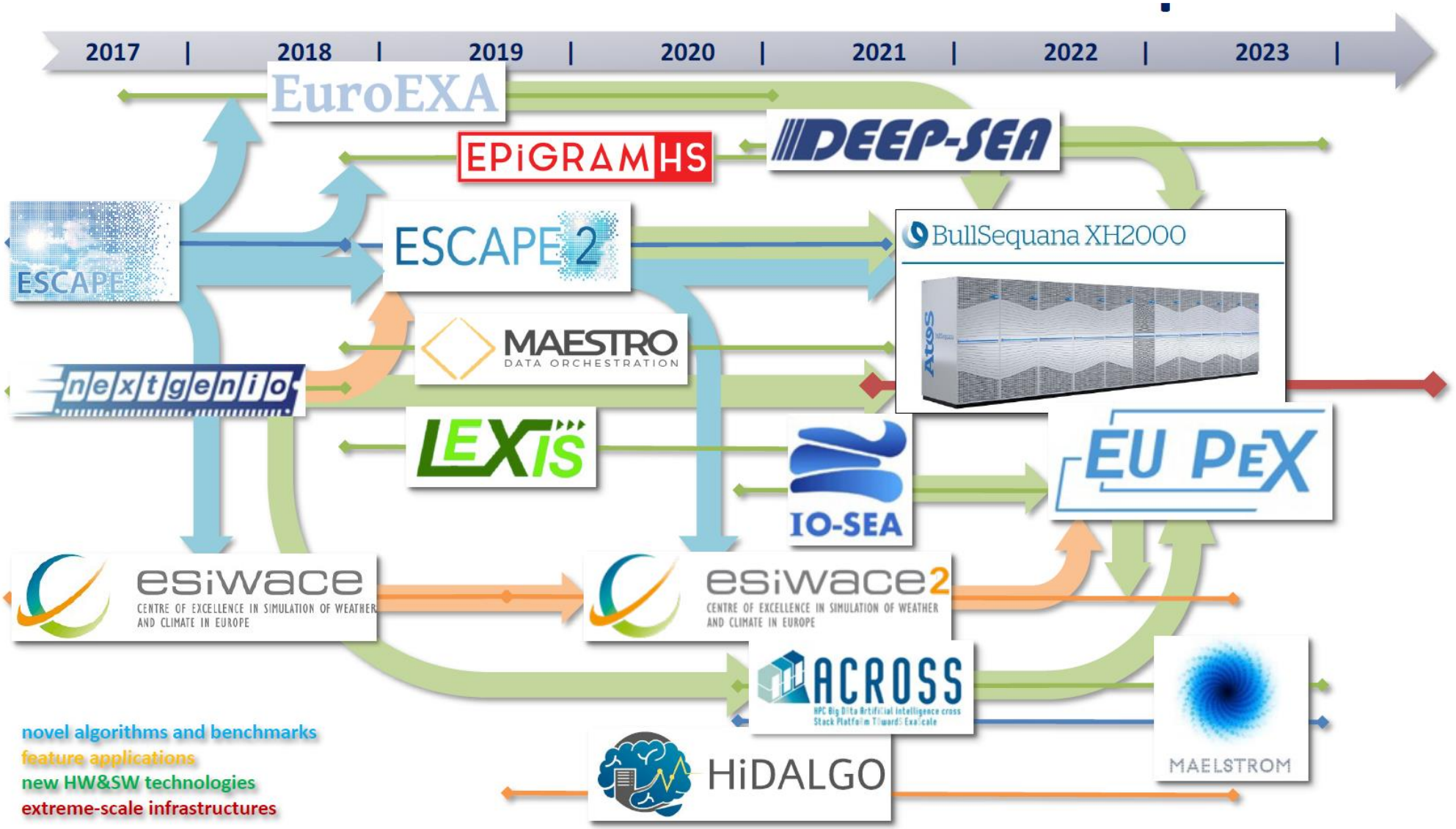
Research done in ESCAPE 2

- science: semi-Lagrangian Discontinuous Galerkin (see presentation by Stella and Giovanni on Wednesday), fault tolerant solvers (presentation by Tommaso on Thursday), machine learning
- domain specific language: developed high-level intermediate representation
- VVUQ: collaboration with CEA and their software package URANIE



Funded by the European Union

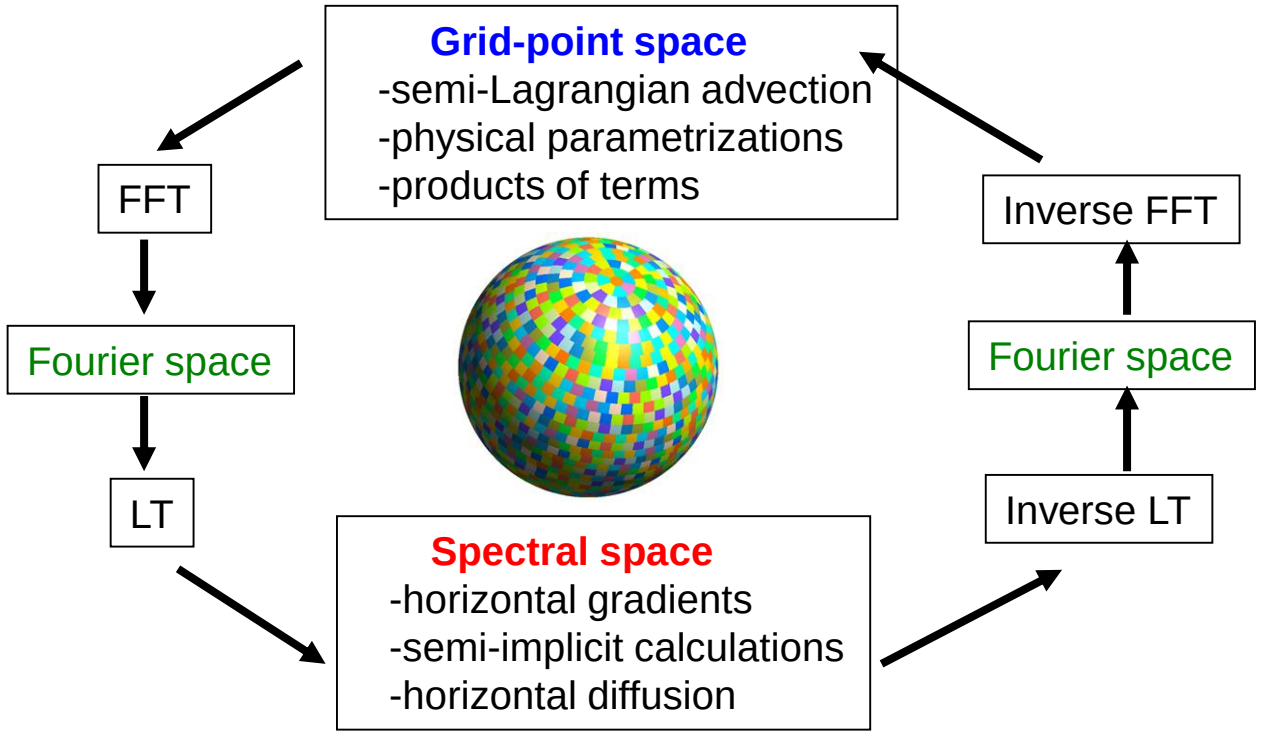
Weather & climate roadmap



ESCAPE 2

Funded by the European Union

Spectral transform cycle for each timestep of the IFS



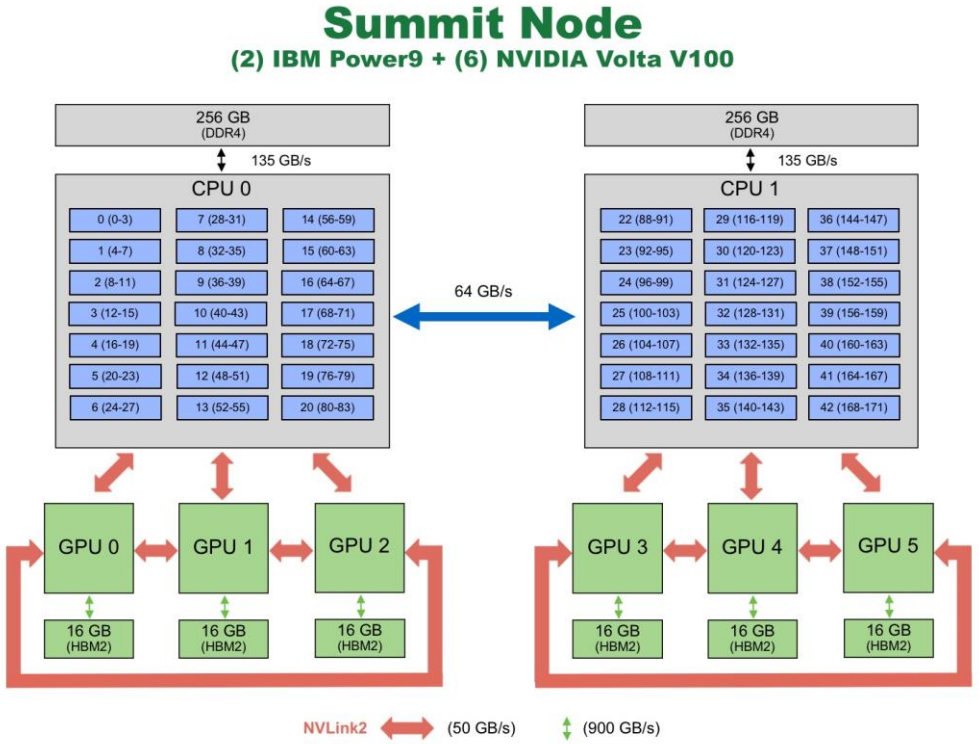
direct spectral transform

inverse spectral transform

FFT: Fast Fourier Transform, LT: Legendre Transform

GPU version of the spectral transform

- History:
  - 2014-2015: George Mozdzynski creates first GPU port of trans library and performs tests on Titan as part of the CRESTA project
  - 2017-2018: Alan Gray (NVIDIA) rewrites George's version as part of the ESCAPE project
  - since 2019: extending functionality of Alan's GPU version and making it work inside RAPS with lots of help from Nils Wedi, Alan Gray, Wayne Gaudin, Ioan Hadade, Sam Hatfield
- Current status:
  - focused so far on making it run inside RAPS (benchmarking version of IFS) at 1km resolution on Summit
  - mostly followed the strategies introduced by Alan Gray in the dwarf
  - work in progress, far from being finished



Funded by the European Union

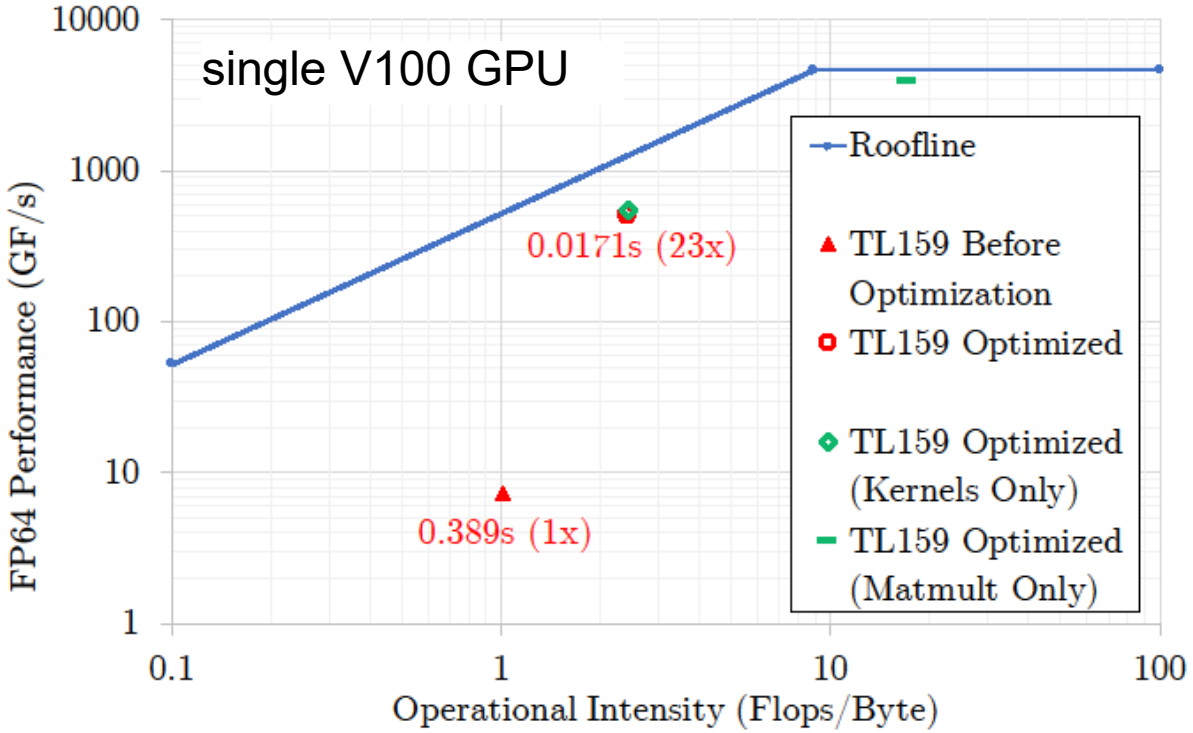
Spectral transform with GPUs in IFS

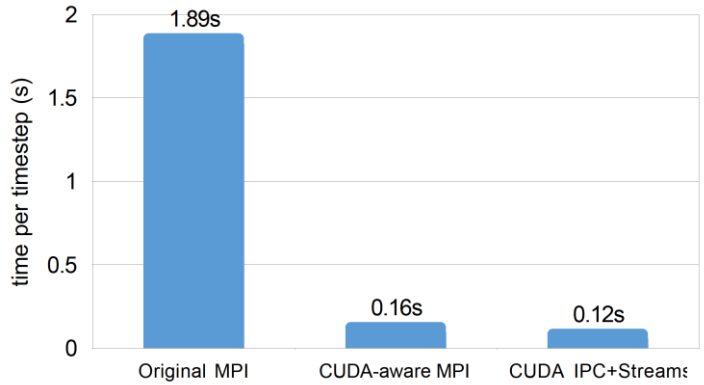
	RAPS on CPU	ESCAPE1 dwarf on GPU	RAPS on GPU
single precision	yes	no	yes
compiler	IBM	NVIDIA	NVIDIA
processor	everything on CPU	entire transform on GPU (CPU only printing norms)	SGEMM and FFT on GPU, reordering of data on CPU
programming standards	MPI + OpenMP	MPI + OpenACC + CUDA (GEMM+FFT)	MPI + OpenMP + OpenACC + CUDA (GEMM+FFT)
parallelisation	OpenMP over zonal wavenumbers (entire Legendre transform)	loop over zonal wavenumbers in innermost functions to parallelize them with OpenACC => changed data layout	
CUDA-aware MPI		yes	no
can run 1km on Summit	yes	no	yes (multiple trans calls)
levels distributed	yes	not used => max 1333 nodes (7998 GPUs) at 1km	

red underlined: temporary limitation which we are currently working on

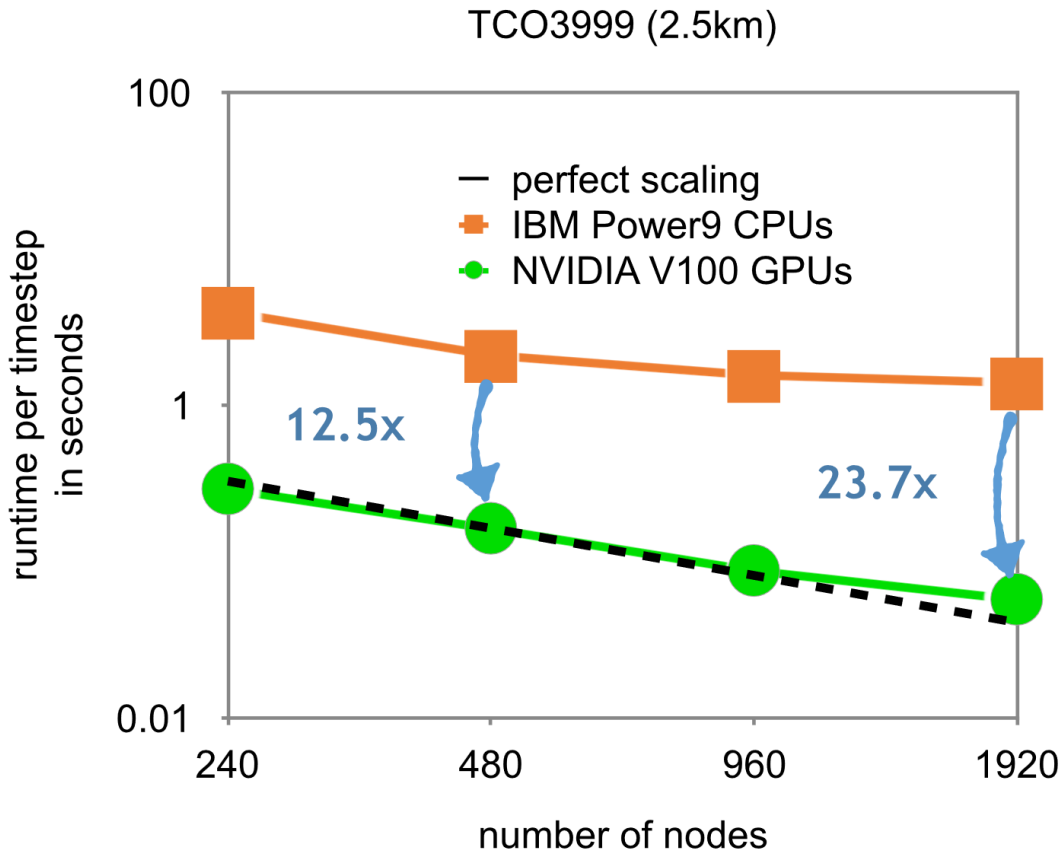
Funded by the European Union

Speedup for scalar transforms in ESCAPE1 dwarf





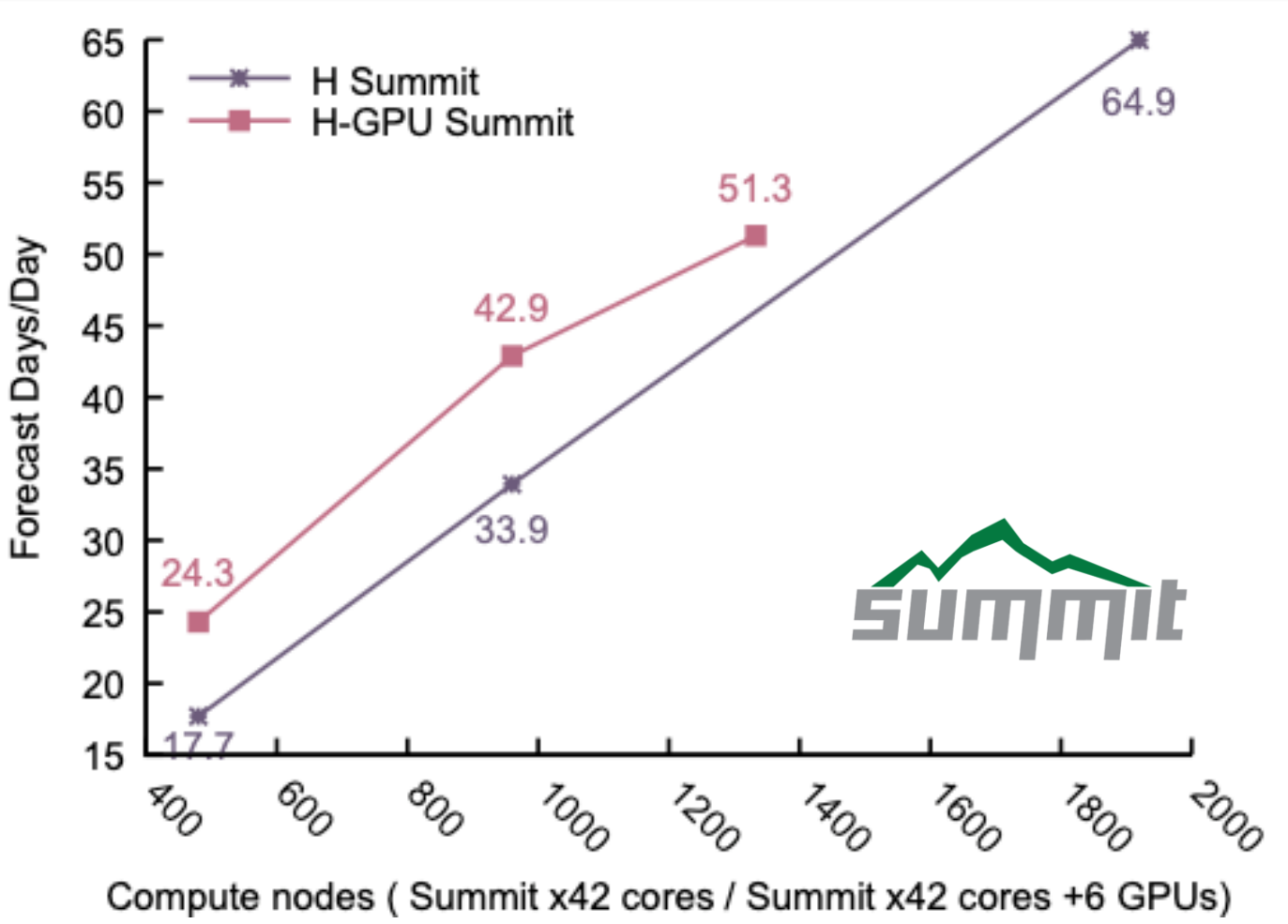
huge speedup with CUDA-aware MPI on 4 DGX−1V

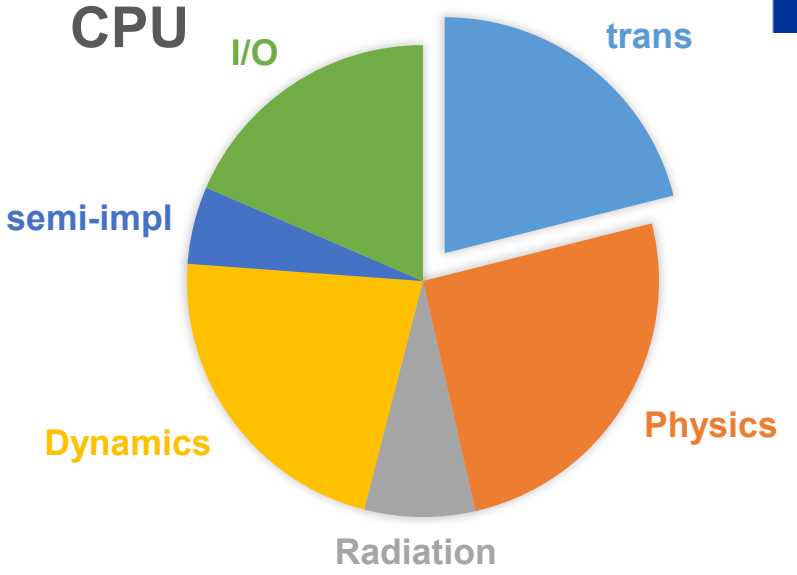


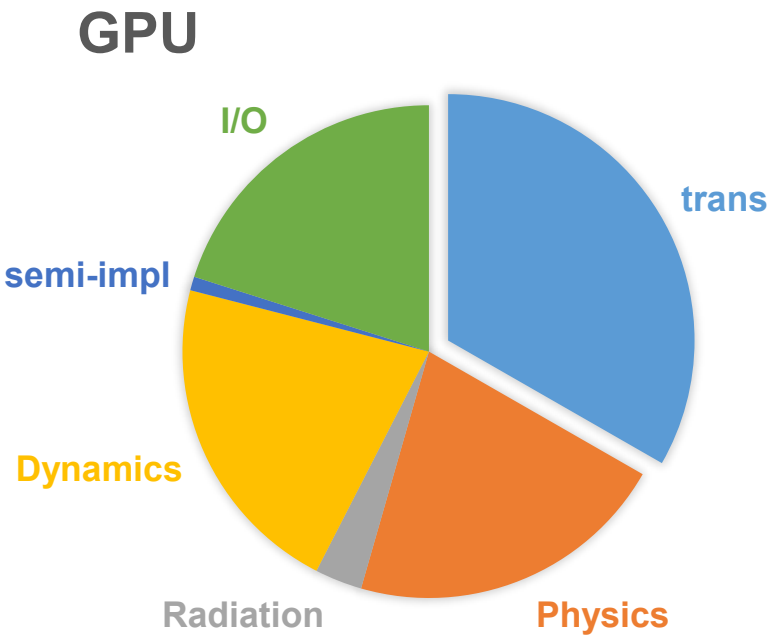
This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE office of Science User Facility supported under contract DE-AC05-00OR22725.

Funded by the European Union

Full model forecast at 1 km





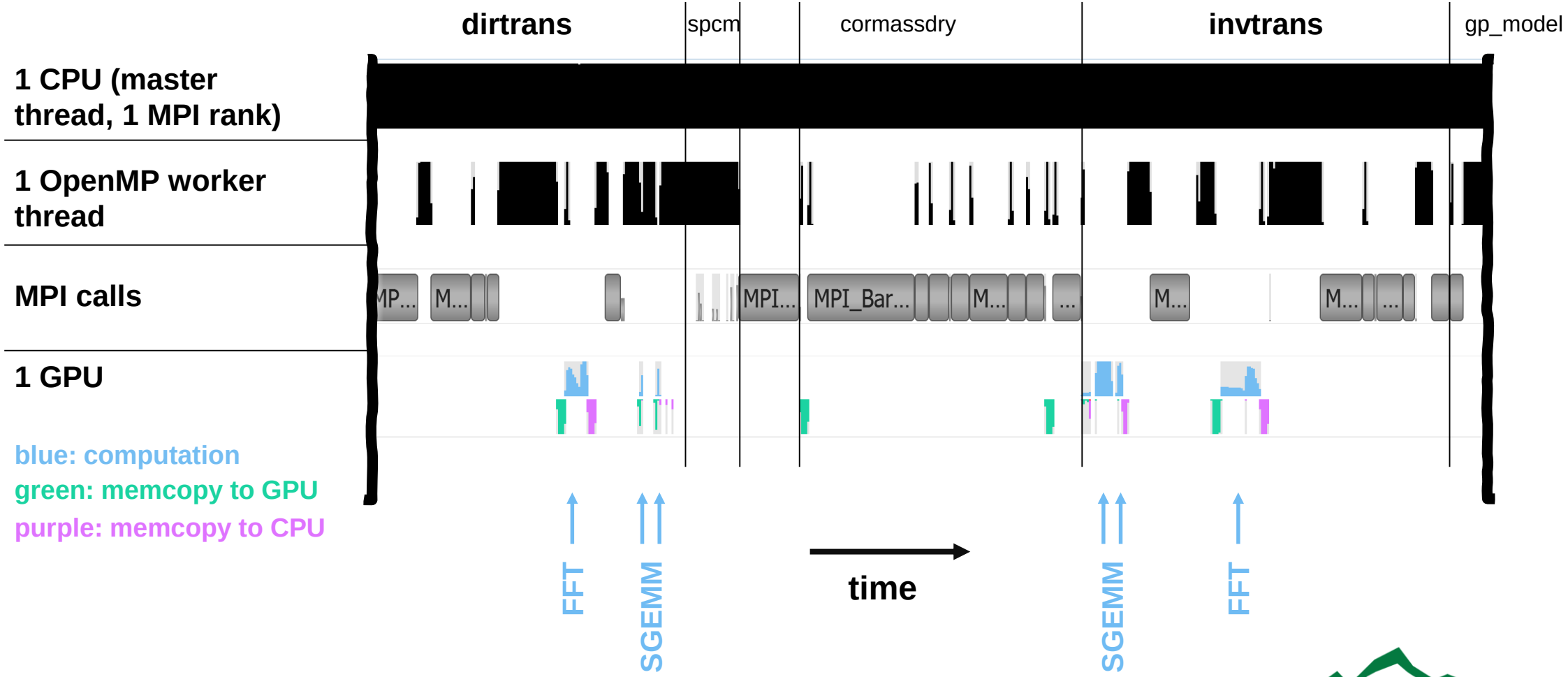


ESCAPE 2

Funded by the European Union

Profiling with NVIDIA nsight:

spectral transform computations of one time-step (9km resolution)



20 nodes, 120 V100 GPUs, 1 MPI rank per GPU, 14 OpenMP threads per MPI rank

summit

Funded by the European Union

Outlook for spectral transform on GPUs

Ongoing tasks

- optimisation of GPU version in the full model
- exploring Fast Legendre Transform on GPU to reduce memory footprint

Next tasks

- try MAGMA library (provides batched GEMM without padding matrices and provides internal computations in half precision)
- bring tangent linear and adjoint model parts of the transform to GPU

Open questions

- How much GPU memory will other parts of IFS need?
- What level of performance can the GPU version achieve on CPUs?