Centro Operativo per la Meteorologia C.O.MET.

Italian Air Force Meteorological Centre

WG 6, - 10 September 2019

Performance and experience porting Bechtold convection scheme on GPU

In collaboration with:

Federal Office of Meteorology and Climatology ***MeteoSwiss***

Lt. Scatamacchia Riccardo
NWP Modelling Branch

Aeronautica Militare

Outline

- *NWP system at COMET*
- *ITAF HPC facility*
- *GPU porting with OpenACC*
- *Future plans*

Operational Numerical Weather Prediction System

Ensemble Data Assimilation:

Operational since June 2011

7 km
49 v.l.

COMET-LETKF
(40+1 members)

Operational since nov 2018

2.2 km
65 v.l.

LETKF
(40+1 members)

Observations

RAOB (also 4D), PILOT, SYNOP, ISPRA rep., SHIP, BUOY, Wind Profilers, AMDAR-AIREP, Mode-S, MSG3-MET8 AMV, Metop scatt. winds, NOAA/Metop AMSUA/MHS and NPP/NOAA 20 ATMS radiances

ECMWF Boundary Conditions
(HRES, ENS also from BC Progr.)

Ensemble Analysis

Deterministic Analysis

Local Area Modeling:

5 km
45 v.l.

parameterized convection

2.2 km
65 v.l.

explicit convection

Ensemble Prediction System:

7 km
45 v.l.
40 m.

COSMO-ME EPS
00-12 UTC up to 72h

2.2 km
65 v.l.
20 m.

COSMO-IT EPS
00-12 UTC up to 48h

NETTUNO-ME EPS 3'
30 freq., 36 dir.
00 UTC up to 48h

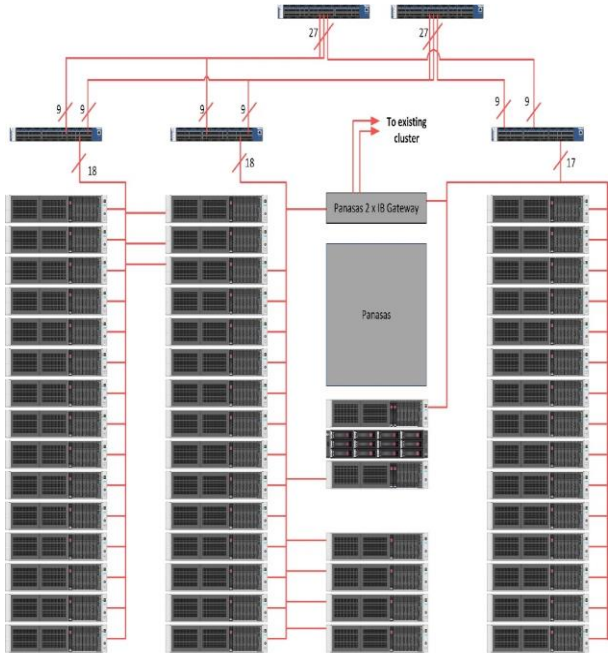
COSMO-ME
00-06-12-18 UTC up to 72h

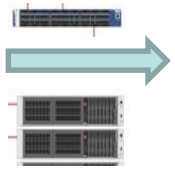
COSMO-IT
00-06-12-18 UTC up to 30/48h

NETTUNO-ME 3'
30 freq., 36 dir.
00-12 UTC up to

NETTUNO-IT 1'
30 freq., 36 dir.
00-12 UTC up to
48h

ITAF HPC Facility





Hybrid CPU/GPU (CPU Intel, GPU INVIDIA)

- 66 x DL380 G9 Computing Nodes (2 x 12 Haswell cores - 256 GB)
- 10 x DL380 G9 Computing Nodes (2 x 14 Haswell cores - 256 GB)
- 2 x DL380 G9 Management Nodes
- 4 x PANASAS AS12 (1 Director Blade each)
- 6 x Infiniband 36p FDR switches
- 152 x Kepler K80 GPUs

- K80 Features:

> Up to 2.91 Teraflops of double precision performance with NVIDIA GPU Boost™
> Up to 8.74 Terfalops of single precision performance with NVIDIA GPU Boost
> 24 GB of GDDR5 memory (12 GB per GPU)

- 19.45 TB RAM
- 240 TFLOPS peak (related to HPL Benchmark)

Next upgrade:

- + 7 computing nodes (+14 GPU v100)
  + 2 IB EDR 36p Switches

Operational Models at COMET

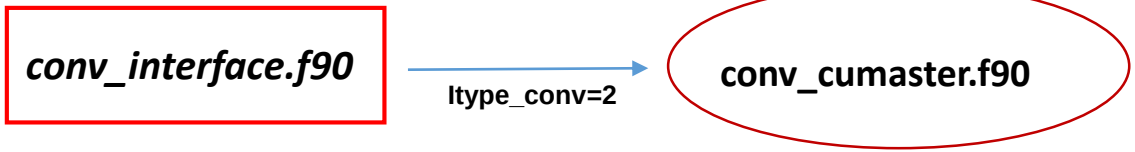
- COSMO ME CPU (version 5,06) 24 nodes, 576 cores in total. ***Tiedke*** convection scheme. Simulation time about 1.5 h (+72h)

- COSMO IT CPU (version 5,06) 24 nodes, 576 cores in total. ***Bechtold*** shallow convection scheme 24 nodes, 576 cores in total. Simulation time about 2.3 h (+48h)

- COSMO ME EPS (pompa) hybrid CPU-GPU 40 members one node per each member. ***Tiedke*** convection scheme. Simulation time about 1,3 h (+72h)

- COSMO IT EPS (pompa) hybrid CPU-GPU 20 members two nodes per each member. ***Tiedke*** shallow convection scheme. Simulation time about 1.9 h.

Bechtold Convection Scheme in COSMO



Main module: *it computes the physical tendecies of the prognostic variables T, Q, U, V and tracers due to convective processes*

conv_cuinit.f90
Inizialization of physical fields

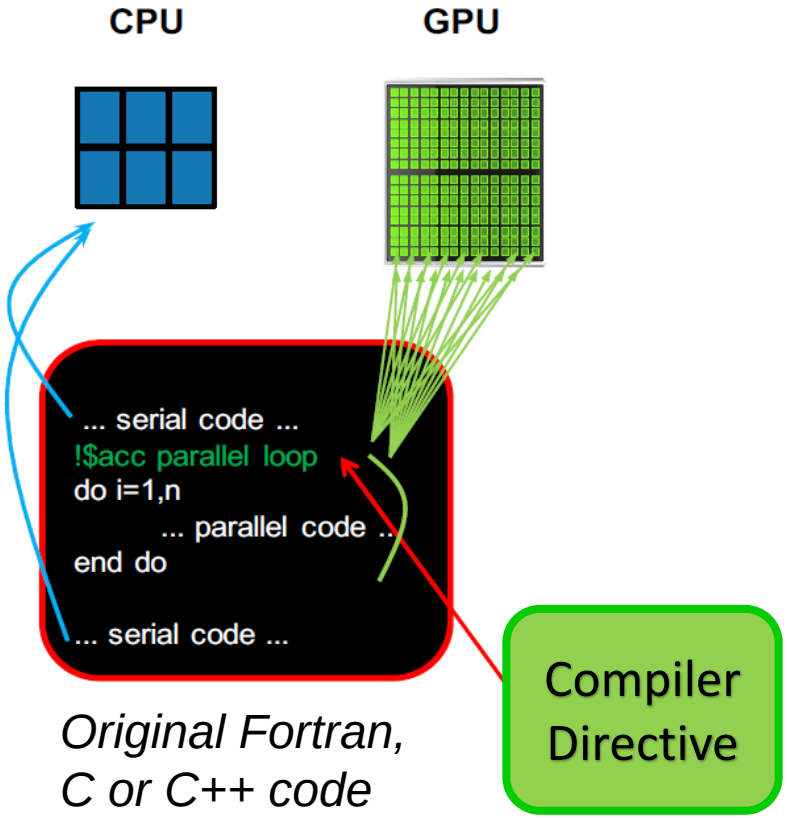
conv_cuascn.f90
Calculations for cloud ascents for cumulus

conv_cudescn.f90
Calculates cumulus downdraft descent

conv_cufluxtends.f90
Calculates convective fluxes in cloud and in subcloud layers. Update of T, Q, U, V and computation of transport of chemical tracers

12 subroutines ported on GPU – about 10000 code lines

OpenACC approach



- Compiler directives are comments or pragmas which can be inserted into existing code
- When a compiler directive is encountered the compiler in runtime will:

1. Generate parallel code for GPU
2. Allocate GPU memory and copy input data
3. Execute parallel code on GPU
4. Copy output data to CPU and deallocate GPU memory

- Directives are ignored without GPU

Validation

In general CPU and GPU results are not bitwise identical.
Note: CPU codes compiled with two compilers are in general not bitwise identical

Validations require to define acceptable thresholds, e.g. by perturbing CPU reference results to the last bit

- A reference test with Bechtold scheme has been implemented in MeteoSwiss Testsuite on KESCH supercomputer. The GPU code has been validated step by step during GPU porting with both cray and pgi compilers.

- New Software enviroment has been installed and tested on HAL system at COMET in order to run validation tests with the last version of COSMO-GPU

  PGI 18.4 - MVAPICH 2.2rc1 - CUDA 8.0 - GCC 4.8.5 - NETCDF 4 - ECCODE 2.8.2

Performance

	CPU (Tiedke)	CPU (Bechtold)	GPU (Tiedke)	GPU (Bechtold)
Total Time	9679 sec	10968 sec	1079 sec	1179 sec
Convection Time	23 sec	121 sec	1,88 sec	58 sec

Convection Speed up with respect to CPU code:

Bechtold x 2
Tiedke x 12

Results for +6h COSMO v5.06 with lgsp_first=true using 8 GPU sockets (4 K80 NVIDIA cards) or 8 CPU sockets (8 Intel Hasswell CPUs with 12 cores each) – Measured on HAL System at COMET

Performance

	COSMO-POMPA DP (Tiedke)	COSMO v5.06 DP (Bechtold)	COSMO v5.06 SP (Bechtold)
Total Time	3050 sec	4267 sec	3803 sec
Convection Time	8,48 sec	253 sec	252 sec

Bechtold scheme is too slow compared to Tiedke scheme.
Further optimizations are needed

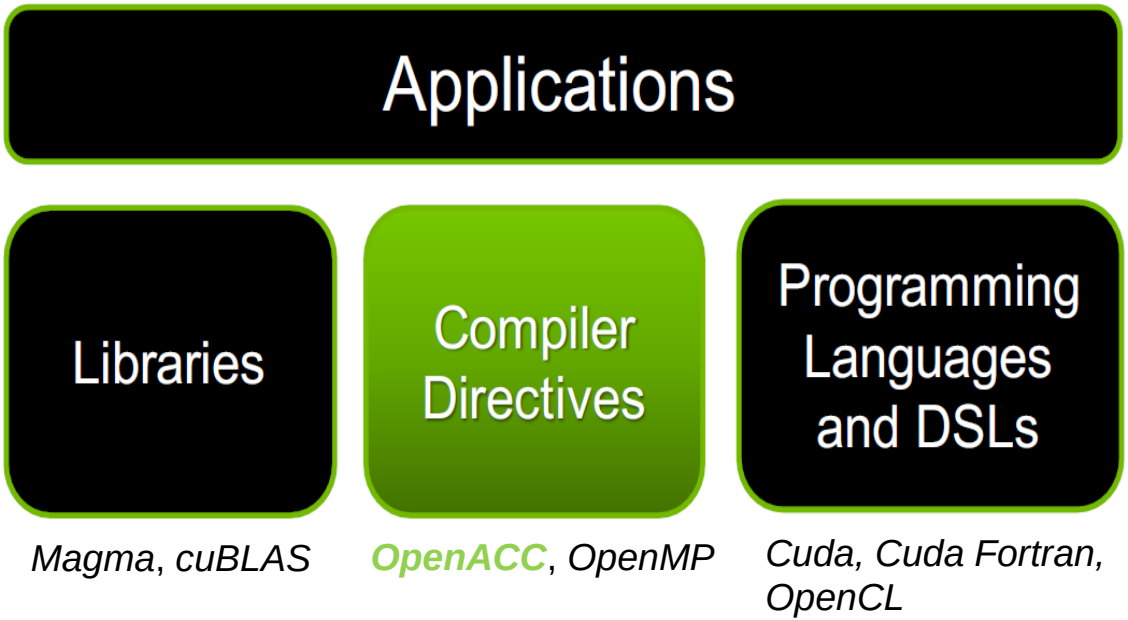
Results for +22h COSMO_POMPA and COSMO v5.06 with lgsp_first=true using 8 GPU sockets (4 K80 NVIDIA cards) – Measured on HAL System at COMET

Future plans

- Implement further optimizations of GPU Bechtold scheme

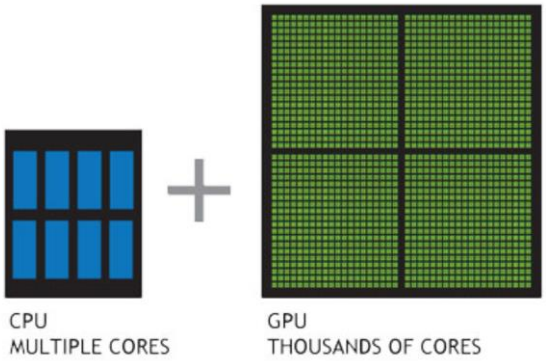
- Make COSMO IT v5.06 operational with Becholtd convection scheme

GPU Computing: Ways to use GPUs



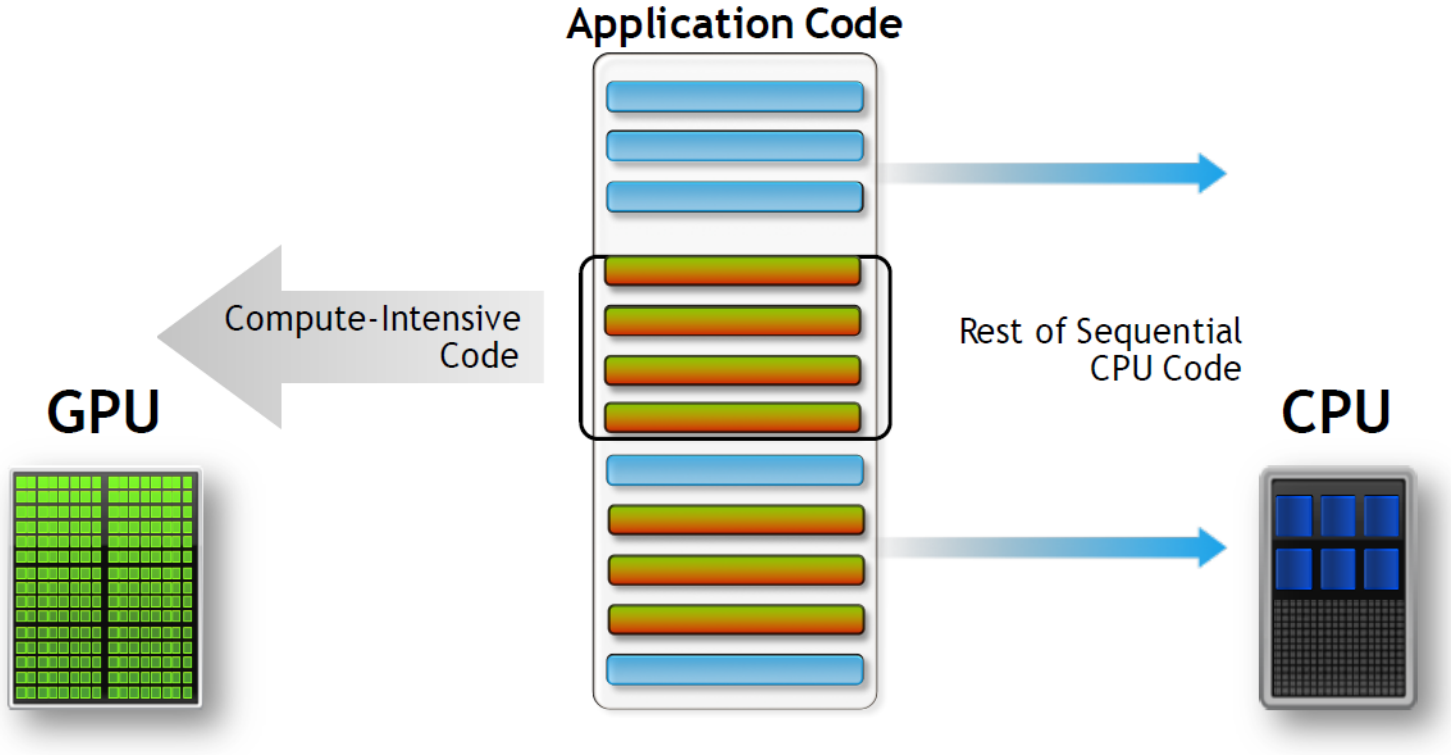
GPU Computing: key aspects

- GPUs have thousand of computing cores which allow to express a fine-grain parallelism



- GPUs and CPUs have separate physical memory
  - *require specific data management .*
  - *data transfer may be a performance issue due to slow transfer via PCI bus.*

GPU Computing: Hybrid approach



Bechtold VS Tiedke

Short motivation about using bechtold in cosmo it

Bechtold Convection Scheme in COSMO

(about 10000 code lines)

Cumastrn, Satur, Conv_cufunctions, Cuinin, Cubasen, Cuascn, Cudlfsn, Cuddrafn, Cuflxn, Cudtdqn, Cududv, Cutracer

12 subroutines ported on GPU