Co-Designing a Productive and Energy-Efficient Weather Simulation System

O. Fuhrer, X. Lapillonne, G. Peretti-Pezzi, C. Osuna, Thomas C. Schulthess
1000-fold performance improvement per decade

www.top500.org
“Only” 100-fold improvement for climate codes

Source: Peter Bauer, ECMWF
Has efficiency of climate codes dropped 10-fold every decade decade?
Peak performance is algorithm dependent

Peak performance varies with arithmetic density of algorithm / code / benchmark

source: lorena a. barba group (lorenabarba.com)
Generic performance metrics in HPC

Energy & Time
Optimising Time and Energy to Solution

Time to solution (TTS):
- do we have to minimise time to solution?
- no, it just needs to be good enough to meet operational constraints

Energy to solution (ETS):
- energy is directly proportional to cost (energy = power x time)
- given all operational constraints, energy should be minimised
Today’s (2015) production suite of Meteo Swiss

COSMO-7
3x per day 72h forecast
6.6 km lateral grid, 60 layers

ECMWF
2x per day
16 km lateral grid, 91 layers

COSMO-2
8x per day 24h forecast
2.2 km lateral grid, 60 layers

Some of the products generate from these simulations:
- Daily weather forecast on TV / radio
- Forecasting for air traffic control (Sky Guide)
- Safety management in event of nuclear incidents
“Albis” & “Lema”, CSCS production systems for Meteo Swiss

Cray XE6 procured in spring 2012 based on 12-core AMD Opteron multi-core processors
Cloud resolving simulations

Institute for Atmospheric and Climate Science Study at ETH Zürich (Prof. Schär) demonstrates cloud resolving models converge at 1-2km resolution (at least for convective clouds over the alpine region)

Cloud ice

Cloud liquid water

Rain

Accumulated surface precipitation

COSMO model setup: $\Delta x=550$ m, $\Delta t=4$ sec

Orographic convection – simulation: 11-18 local time, 11 July 2006 ($\Delta t_{\text{plot}}=4$ min)

Plots generated using INSIGHT

Source: Wolfgang Langhans and Christoph Schär, Institute for Atmospheric and Climate Science, ETH Zurich
Higher resolution is necessary for quantitative agreement with experiment

18 days for July 9-27, 2006

**Observation**  Average wind speed (——) and direction (◊)

---

source: Oliver Fuhrer, MeteoSwiss
Improve resolution of Meteo Swiss model from 2 to 1 km

Doubling the resolution requires ~10x performance increase

Run on 4x the number of processors
Prognostic uncertainty

The weather system is chaotic
→ rapid growth of small perturbations (butterfly effect)

**Ensemble method:** compute distribution over many simulations
Benefit of ensemble (heavy thunderstorms July 24, 2015)

Adelboden
Improving simulation quality requires higher performance – what exactly and by how much?

Resource determining factors for Meteo Swiss’ simulations

Current model running through mid 2016

**COSMO-2**: 24h forecast running in 30 min. 8x per day

**COSMO-2E**: 21-member ensemble, 120h forecast in 150 min., 2x per day (~26x COSMO-2)

**KENDA**: 40-member ensemble, 1h forecast in 15 min., 24x per day (~5x COSMO-2)

New model starting operation on in Jan. 2016

**COSMO-1**: 24h forecast running in 30 min. 8x per day (~10x COSMO-2)

New production system must deliver ~40x the simulations performance of “Albis” and “Lema”
State of the art implementation of new system for Meteo Swiss

- New system needs to be installed Q2-3/2015
- Assuming 2x improvement in per-socket performance: 
  ~20x more X86 sockets would require 30 Cray XC cabinets

Albis & Lema: 3 cabinets Cray XE6 installed Q2/2012

New system for Meteo Swiss if we build it like the German Weather Service (DWD) did theirs, or UK Met Office, or ECMWF … (30 racks XC)

Current Cray XC30/XC40 platform  
(space for 40 racks XC)

Thinking inside the box is not a good option!
Co-Design our way out?

Potential for co-design

- Time-to-solution driven (specs are clear)
- Exclusive usage
- Only one performance critical application
- Stable configuration (code & system)
- Current code can be improved
- Novel hardware has yet to be exploited

Challenges for making it work

- Community code
  - Large user base
  - Performance portability
  - Knowhow transfer
- Complex workflow
- High reliability required
- Rapidly evolving technology (hardware and software)
“Piz Daint,” a productive supercomputer with CPU-GPU nodes

- Cray XC30 with 5272 compute nodes, each with one 8-core Xeon CPU and one K20X GPU
- Fully populated dragonfly: global bandwidth per node matches injection bandwidth
- Developed with application performance in mind: CP2K, COSMO, SPECFEM, GROMACS, Q.E.
- Co-designed with CP2K and COSMO-OPCODE
- Final upgrade 10/2013; accepted 12/2013; early science 01-03/2014; full operation since 04/2014
Co-design approach

• Co-design software / workflow / hardware paying attention to
  • Portability to other users and hardware architectures
  • Achieve specified time-to-solution
  • Optimise hardware footprint and energy
• Several collaboration pre-existed
  • Software development since 2010: MeteoSwiss / C2SM@ETH Zurich / CSCS
  • CSCS with Cray and NVIDIA for development of “Piz Daint” in 2013
  • Domain scientists and computer scientists
• Substantial software investments from HPCN Strategy: HP2C and PASC
• Extreme programming team
  • Oliver Fuhrer (the perfect product owner)
  • Tobias Gysi, Carlos Osuna, Xavier Lapillonne, Mauro Bianco, Andrea Arteaga (not all at the same time)
  • CSCS experts: Ben Cumming, Gilles Fourestey, Guilherme Peretti-Pezzi
  • NVIDIA experts: Peter Messmer, Christoph Angerer
COSMO: current and new (refactored) code

- Main (current / Fortran)
  - Physics (Fortran)
  - Dynamics (Fortran)
  - MPI system

- Main (new / Fortran)
  - Dynamics (C++)
  - Physics (Fortran) with OpenMP / OpenACC
  - Stencil library
  - Boundary conditions & halo exchg.
  - Shared Infrastructure
  - Generic Comm. Library
  - MPI or whatever system
A factor 40 improvement with the same footprint

Current production system: Albis & Lema

New system: Kesch & Escha
Piz Kesch / Piz Escha: appliance for meteorology

- Water cooled rack (48U)
- 12 compute nodes with
  - 2 Intel Xeon E5-2690v3 12 cores @ 2.6 GHz
  - 256 GB 2133 MHz DDR4 memory
- 8 NVIDIA Tesla K80 GPU
- 3 login nodes
- 5 post-processing nodes
- Mellanox FDR InfiniBand
- Cray CLFS Luster Storage
- Cray Programming Environment
Origin of factor 40 performance improvement

Performance of COSMO running on new “Piz Kesch” compared to current production systems

- Current production system installed in 2012
- New Piz Kesch/Escha installed in 2015
  - Processor performance \(2.8x\)
  - Improved system utilisation \(2.8x\)
  - General software performance \(1.7x\)
  - Port to GPU architecture \(2.3x\)
  - Increase in number of processors \(1.3x\)
  - Total performance improvement \(~40x\)
- Bonus: simulation running on GPU is \(3x\) more energy efficient compared to conventional state of the art CPU

Moore’s Law 2012-2015

Software refactoring

Moore’s Law
References and Collaborators

- Peter Messmer and his team at the NVIDIA co-design lab at ETH Zurich
- Teams at CSCS and Meteo Suisse, group of Christoph Schaer @ ETH Zurich
Join us @ the PASC16 Conference

PASC16 provides an opportunity for scientists and practitioners to discuss key issues in the use of High Performance Computing (HPC) in branches of science that require computer modelling and simulations. The scientific program will offer invited lectures, minisymposia, contributed talks and poster presentations. The active participation of graduate students and postdocs is strongly encouraged.

Contributions

Researchers from the academic and from the corporate world are invited to participate and present their research area in the form of minisymposia, contributed talks and/or poster presentations. PASC16 welcomes submissions in the following scientific fields:

- CLIMATE & WEATHER
- SOLID EARTH
- LIFE SCIENCE
- CHEMISTRY & MATERIALS
- PHYSICS
- COMPUTER SCIENCE & MATHEMATICS
- ENGINEERING
- EMERGING DOMAINS

Abstracts should describe original, interesting, and solid scientific content that is relevant to computational sciences and HPC. Cross-disciplinary approaches are highly encouraged.

Organization Committee

- Maria Grazia Giuffreda, ETH Zurich / CSCS
- Jan Hesthaven, EPFL
- Torsten Hoefler, ETH Zurich
- David Keyes, KAUST
- Nicola Marzari, EPFL
- Olaf Schenk, USI
- Thomas Schulthess, ETH Zurich / CSCS

Venue

EPF Lausanne
Swiss Tech Convention Center
Lausanne
Switzerland