Decoupling the Selection of Compression Algorithms from Required Precision with the Scientific Compression Library (SCIL)

Julian Kunkel, Eugen Betke,
Anastasiia Novikova, Armin Schaare

Scientific Computing
Department of Informatics
University of Hamburg

ISC HPC Project Poster Presentation
2017-06-20

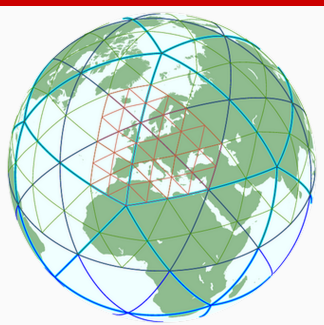
Outline

Goals of the AIMES Project

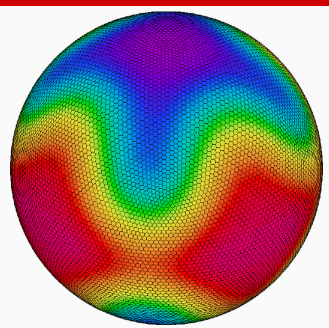
Address key issues of icosahedral earth-system models

- Enhance programmability and performance-portability
- Overcome storage limitations
- Shared benchmark for these models

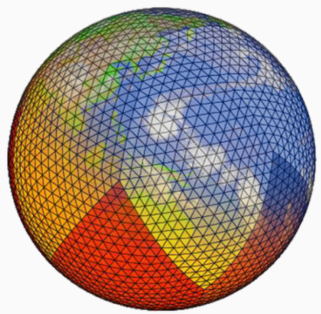
Covered models



ICON



DYNAMICO



NICAM

Goals for Compression

Goal: Design of domain-specific compression (ratio $> 10 : 1$)

Approach

1. Investigate metrics allowing to define accuracy per variable
2. Design user-interfaces for specifying accuracy
3. Implement compression schemes exploiting this knowledge
4. Develop a methodology for identifying the required accuracy

SCIL: Scientific Compression Interface Library

- Realizes user-interfaces
- Metacompressor providing many algorithms
- Ongoing work: Enable SCIL to select best-fitting algorithm
- `https://github.com/JulianKunkel/scil`

SCIL: Supported User-Space Quantities

Quantities defining the residual (error):

absolute tolerance: compressed can become true value ± absolute tolerance

relative tolerance: percentage the compressed value can deviate from true value

relative error finest tolerance: value defining the absolute tolerable error for relative compression for values around 0

significant digits: number of significant decimal digits

significant bits: number of significant decimals in bits

field conservation: limits the sum (mean) of field's change

Quantities defining the performance behavior:

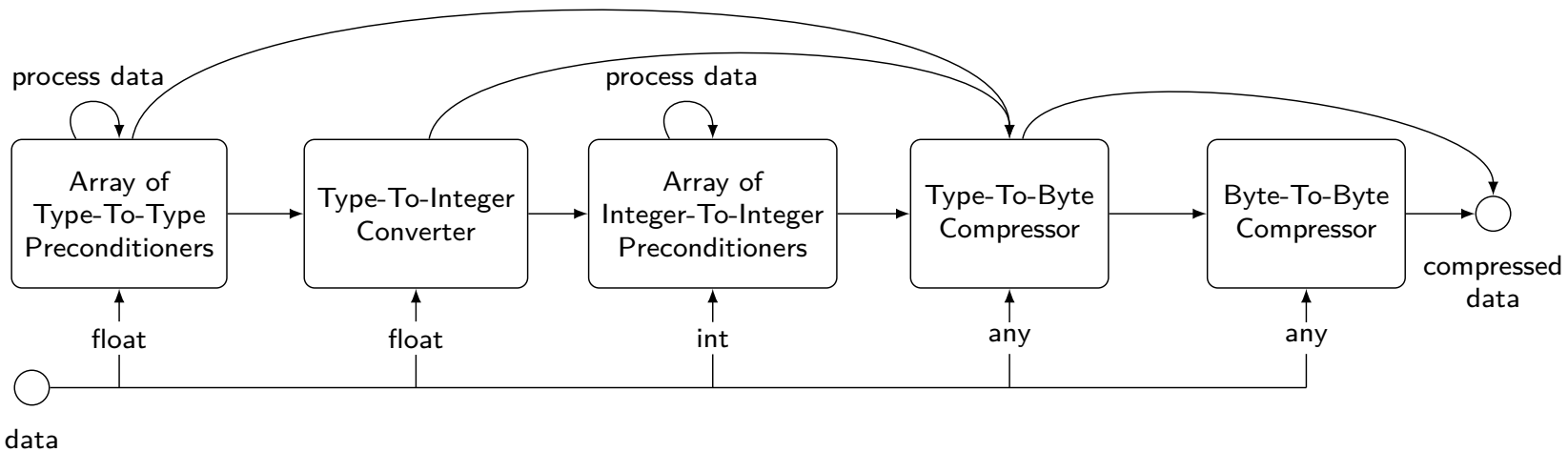
compression throughput

decompression throughput

- in MiB or GiB, or relative to network or storage speed

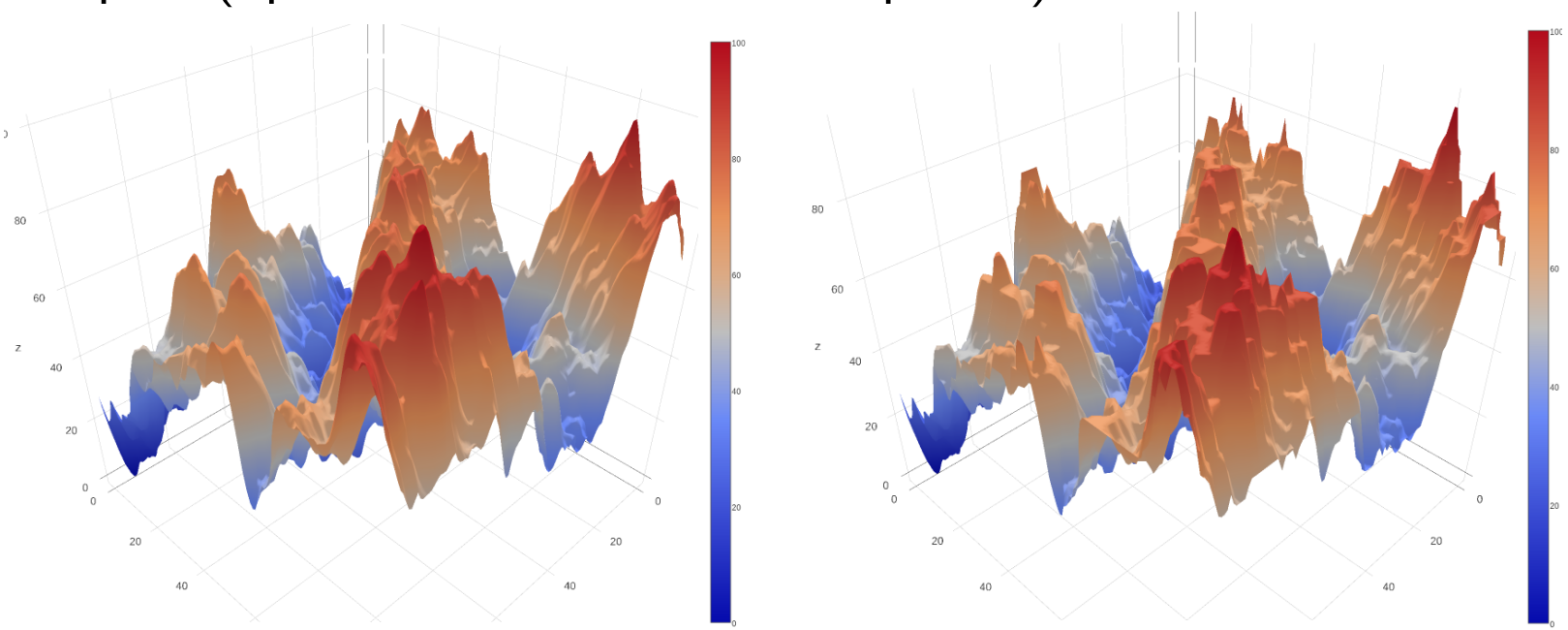
Architecture of SCIL

- Contains tools to
  - Create random patterns, compress/decompress, add noise, plot
- HDF5 and NetCDF4 integration
- Library offers
  - Automatic algorithm selection (under development)
  - Flexible compression chain:



Example Synthetic Data

Simplex (options 206, 2D: 100x100 points)



Right picture compressed with Sigbits 3bits (ratio 11.3:1)

Analyzing Performance of Lossy Compression using SCIL

An initial experiment is conducted with SCIL

Data

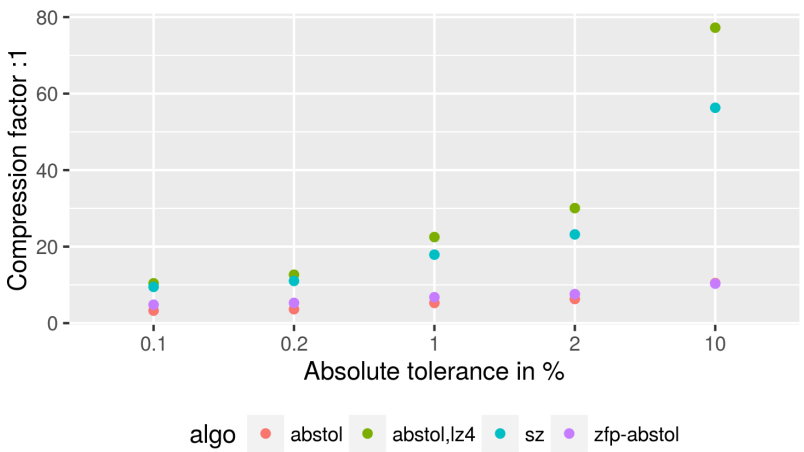
- Single precision (1+8+23 bits)
- Synthetic, generated by SCIL's pattern lib.
  - e.g., Random, Steps, Sinus, Simplex
- Data of the variables created by ECHAM
  - The climate model creates up to 123 vars

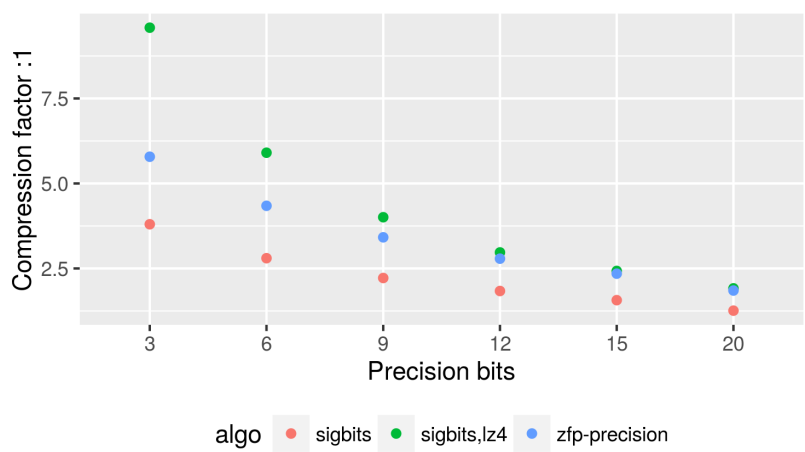
Experiments

- Single thread, 10 repeats
- Lossless (memcopy and lz4)
- Lossy compression with significant bits (zfp, sigbits, sigbits+lz4)
- Lossy compression with absolute tolerance (zfp, sz, abstol, abstol+lz4)
  - Tolerance: 10%, 2%, 1%, 0.2%, 0.1% of the data maximum value

Tolerance-Based Results

- Mean compression factor across all scientific files
- Factor 50:1 means space is reduced to 2% of the original size
- Note that ZFP does not always reach the set precision
  - Often the absolute and precision bit tolerance cannot be met





Comparing Algorithms for the Scientific Files

		Throughput [MiB/s]	
Algorithm	Ratio	Compression	Decompression
sigbits	0.448	462	615
sigbits,lz4	0.228	227	479
zfp-precision	0.299	155	252

Table: Preserving 9 precision bits (instead of 23 from float) ≤ 0.56

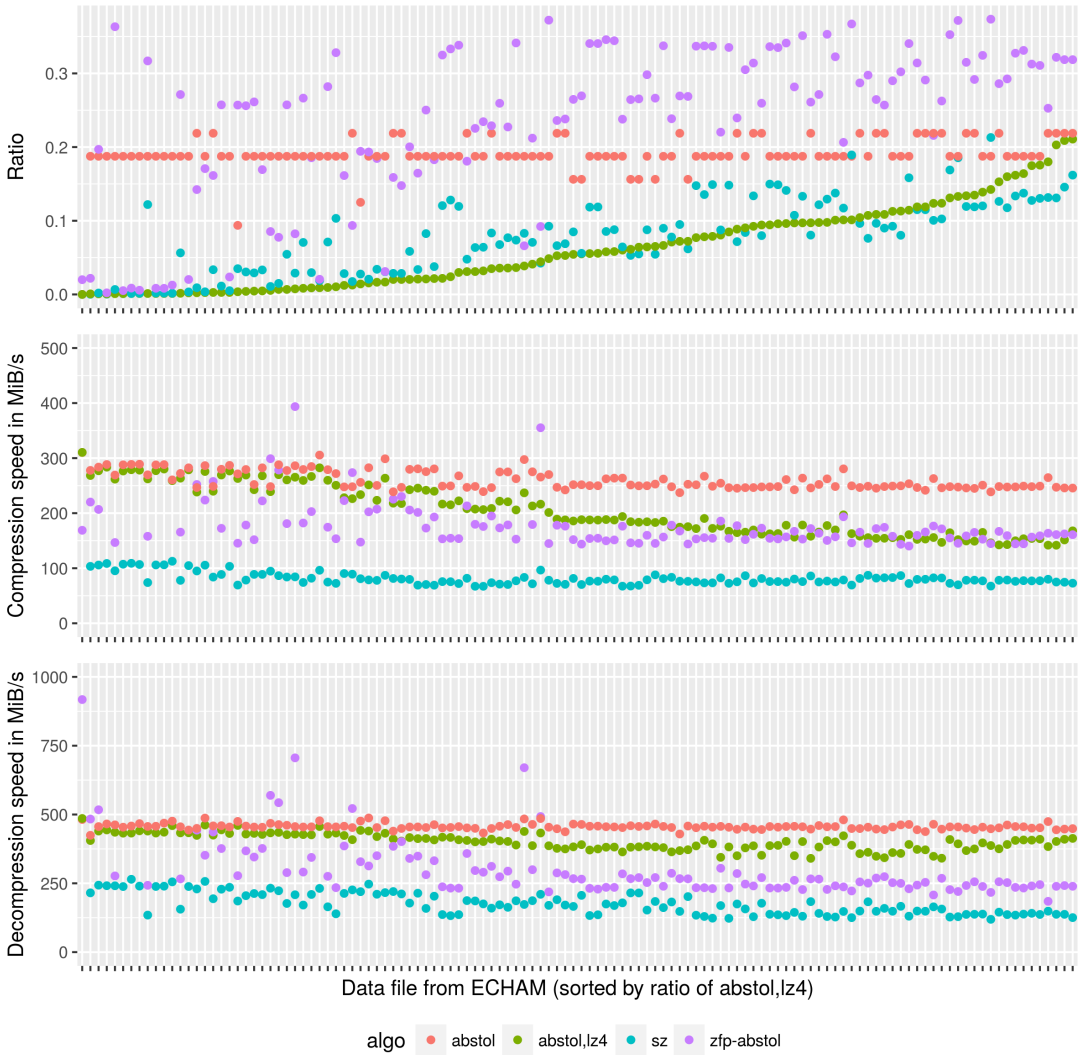
		Throughput [MiB/s]	
Algorithm	Ratio	Compression	Decompression
abstol	0.19	260	456
abstol,lz4	0.062	196	400
sz	0.078	81	169
zfp-abstol	0.239	185	301

Table: For absolute tolerance with 1% of max value < 0.22

The harmonic mean has been used

Results for Absolute Tolerance

Comparing algorithms using an absolute tolerance of 1% of the maximum value



Acknowledgement

This work was supported in part by the German Research Foundation (DFG) through the Priority Programme 1648 “Software for Exascale Computing” (SPPEXA) (GZ: LU 1353/11-1).

We thank Luis Kornblüh from the Max Planck Institute for Meteorology for providing the ECHAM6 data.