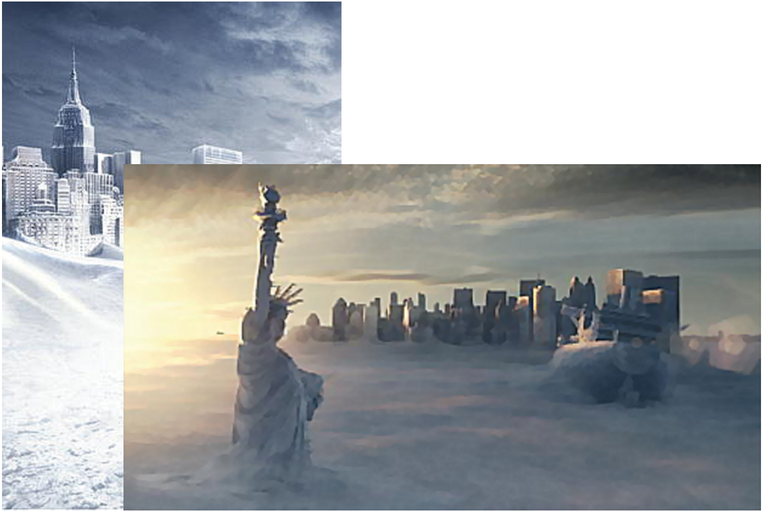
The performance of the COSMOS coupled model at DKRZ and ECMWF

Luis Kornblueh
Max Planck Institute for Meteorology

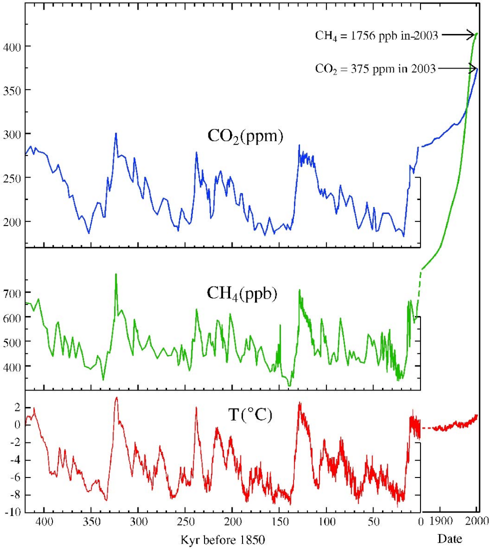
November 1, 2006

Outline

Climate - a fast developing desaster?



What is happening in reality



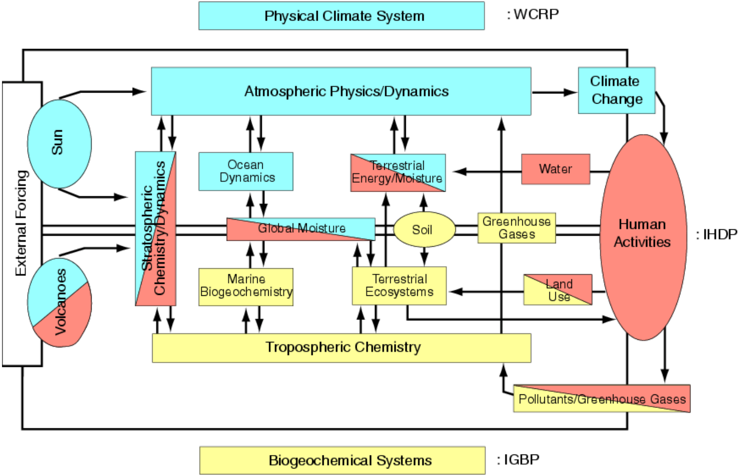
in the last 150 years we have introduced massive perturbations in some of the forcing variables.
How will the system respond???

Climate - or what are we looking for

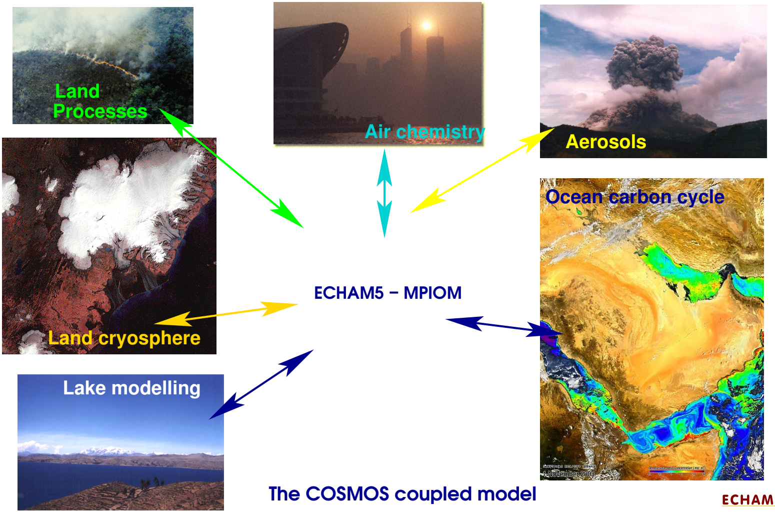
Earth System Research - the general objective

- To understand how physical, chemical, and biological processes, as well as human behavior contribute to the dynamics of the Earth system, and specifically how they relate to global and regional climate changes.
- To observe, monitor, analyze, understand, and predict in order to better manage the Earth system.

The system to look for



Standard model extensions ...



The COSMOS coupled model

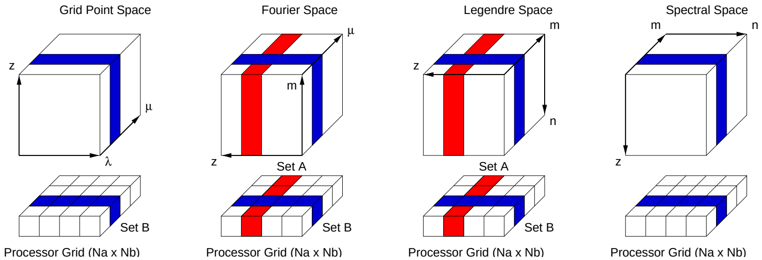
The climate model ECHAM5

- Spectral dynamical core
- Semi-implicit leapfrog time differencing
- Flux-form semi-Lagrangian transport of passive tracers
- Shortwave radiation based on δ-Eddington approximation
- Longwave radiation based on the correlated-k method
- Stratiform clouds based on micro-physical prognostic equations for water vapour, water, and ice
- Convection solved by a mass-flux scheme based on Tiedtke
- Subgrid-scale induced gravity wave drag
- Vertical diffusion (subgrid-scale turbulence closure by TKE)
- Solving the surface energy balance equation

Model states

- spectral space
  - Legendre space
  - Fourier space
- grid point space
- transport flux-form space
- grid point space
  - Fourier space
  - Legendre space
- spectral space

Double transposition strategy



Redistribution for tracer transport

- Resort in latitudinal direction (Na)
  - higher wind speed (transporting) in east-west direction, hardly predictable Courant number
  - low wind speed in north-south direction, Courant number < 1
- and vertical levels and tracers (Nb)

Transposition back into grid point space before calculation of vertical transport.

Loop-level parallelization

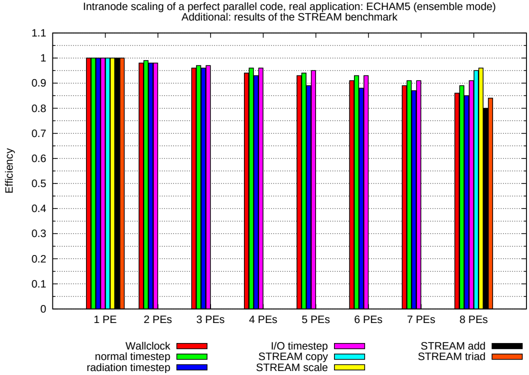
- Physics block - OpenMP orphaning
- Remaining model - OpenMP on loop level
- Performance problem: OS jitter

Vectorisation/Cache blocking

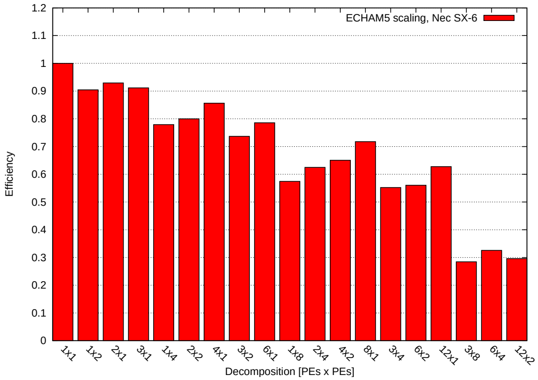
- Physics block - high level strip mining (VL/average optimal cache blocking)
- Remaining model - blocking and *fast and dirty* code
- Performance problem 1: tracer transport
- performance problem 2: *non-mathematical* formulated part
- However reaches 2 GFlops on SX-6 on a small resolution (T63L31) run on a single CPU

Intranode performance degradation

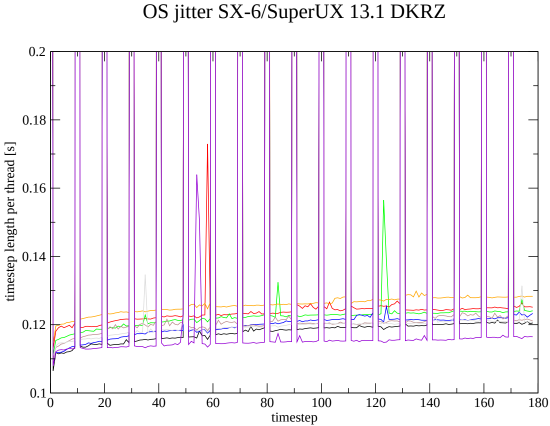
By using a perfectly parallel code the degradation by technical issues of a single SMP node can be determined



Efficiency of parallel model



OS jitter effect

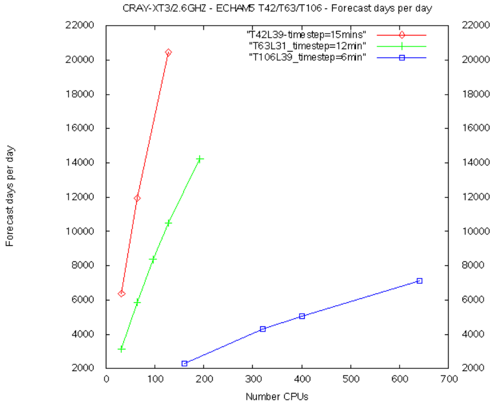


OS jitter continued

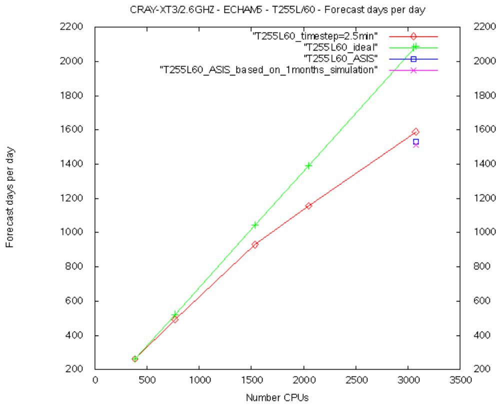
Solution development underway inside NEC:

- reducing number of daemons
- synchronisation of processes other than application over the whole cluster nodes
- frequency adaption for affected applications

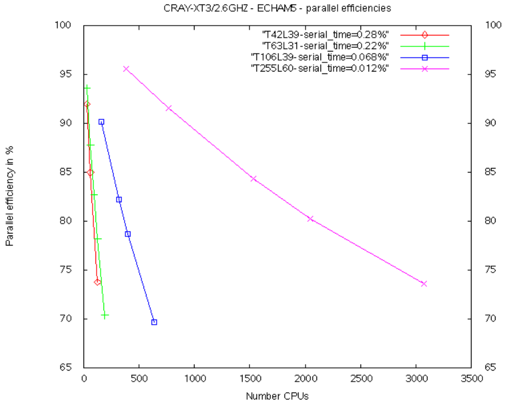
Low resolution examples



High resolution example



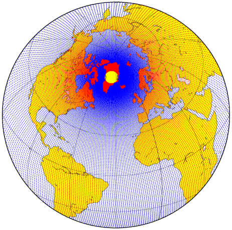
Parallel efficiency



Peak: 1.4 TFlops

MPIOM

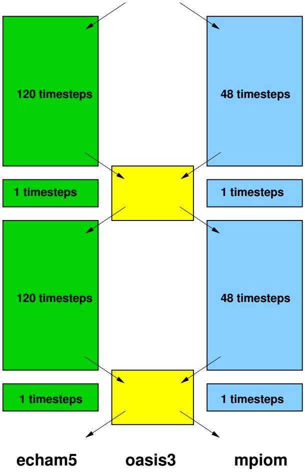
- Primitive Equations
- C-Grid, z-level, partial cells
- mixed upwind/centered difference advection scheme (other schemes possible)
- Hibler sea ice model incl. snow and fractional ice cover
- Sub-gridscale parameterisations
- Isopycnal diffusion
- Edddy induced tracer transports
- Slope convection
- Conformal mapping



The ocean model development level

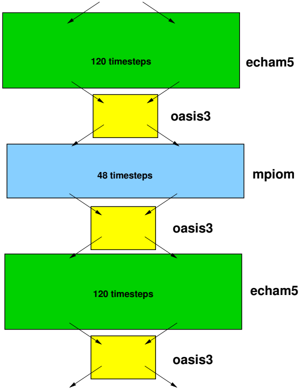
- hybrid MPI - OpenMP
- good vectorization level
- very poor global solver (2d linear equation system, SOR or direct LU)
- 3.75 GFlops on a singel SX-6 CPU

The coupling itself



Inherent load inbalance causes half of a total of 30 % loss in performance by coupling.

Improved performance for coupling



3 MPI tasks - models overlay using OpenMP (MPI suspend mode - no aggresive spinning).

Final performance

DKRZ	ECMWF
NEC SX-6	IBM POWER 4+
5.5 years/day	5.5 years/day
1 node/8 CPUs	1 node/32 CPUs

Note 1: expected result from single model comparison on both platforms
Note 2: scaling on more CPUs at EMWF *easily* possible, but very difficult at DKRZ.

I/O improvements

Current planed experiments:

- climate of the last 1000 years
- several ensemble members at different sites
- *raw* GRIB1 output: 25 TB per member

I/O improvements continued

Improvement by additional packing of the BDS data section!

Properties:

Resolution Source	GRIBZip2d Frauenhofer	grib-szip MPI-M	grib-szip MPI-M	gzip (external) GNU
E5 T42 L19	2.32	2.13	2.13	2.06
E5 T63 L31	1.85	1.78	1.78	1.35
E5 T106 L60	5.17	4.75	4.75	3.81
E5 T213 L31	3.09	3.06	2.71	2.41
mean	3.03	2.93	2.73	2.15

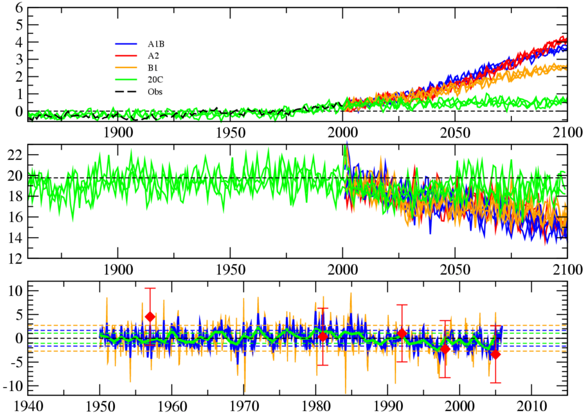
I/O improvements continued

- Frauenhofer solution patented - expensive (80000 Euro/year)
- Nasa szip solution patented - no cost for non-commercial organisations
- GRIBZip2d : 13 MB/s throughput on a standard PC
- grib -szip: 84 MB/s throughput on a standard PC

I/O improvements continued

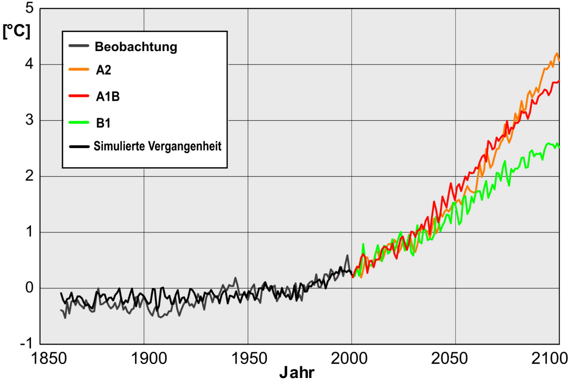
- we go with the szip solution
- all post-processing tools are adapted
- reduces primary output from 25 to 10 TB
- setting up a project for putting compression on clearspeed card

Some result ...



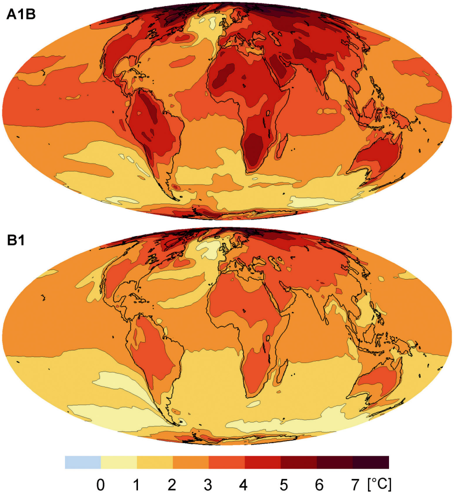
Time Evolution of Atlantic Meridional Overturning Circulation under different IPCC Scenarios

... more results ...



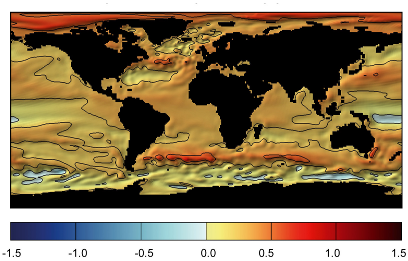
Global mean temperature development [K]
(relative to the mean of the period 1961-1990.

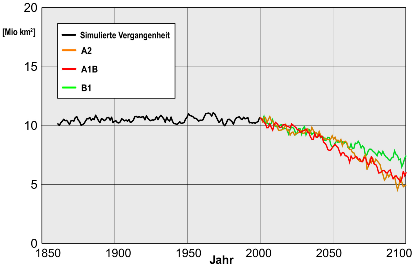
... more results ...



Global temperature anomaly difference between the mean of 2071-2100 to the period 1961-1990.

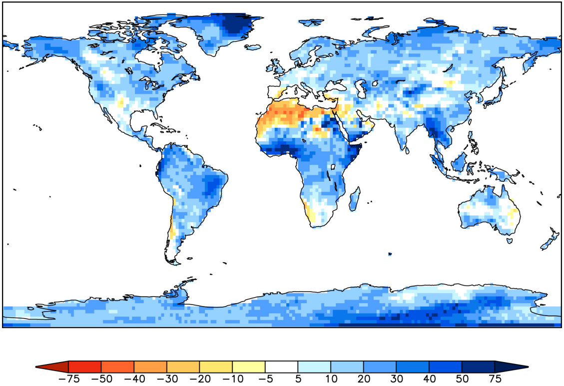
… more results …





Right: global difference of sea level height [m] (2100-2000 SRES A1B)
Left: global sea ice coverage development

... extreme weather



Extreme precipitation change [%]
(difference between the maximum precipitation amount in a 5-day period between 2071-2100 and 1961-1990.

Conclusion

We have to work more on achieving high performance on the coupled model . . .