2nd ENES HPC WS, Session 3
European Earth System Models with a focus on model performance

MPI-ESM

Reinhard Budich

MPI für Meteorologie

reinhard.budich@zmaw.de

With help from

Irina Fast (DKRZ), Marco Giorgetta, Helmuth Haak, Johann Jungclaus, Luis Kornblueh, Leonidas Linardakis

MPI-ESM

- MPI-M: An Intro
- Models at MPI-M
- **MPI-ESM1**
  - Improvements from COSMOS version
  - Configurations
  - Experiments
  - Performance
- Towards MPI-ESM2: **ICON**
- Summary and Conclusions

Max Planck Institute for Meteorology

http://mpimet.mpg.de

- Located in Hamburg, Germany
- Established in 1972
  - Former directors Hinzpeter, Hasselmann, Bengtsson, Brasseur, Grassl
- About 250 People
- Earth system research
  - Atmosphere in the Earth System, Björn Stevens, Mg. Dir.
  - Land in the Earth System, Martin Claussen
  - Ocean in the Earth System , Jochem Marotzke
  - International Max-Planck-Research School on Earth System Modelling
- DKRZ separate entity

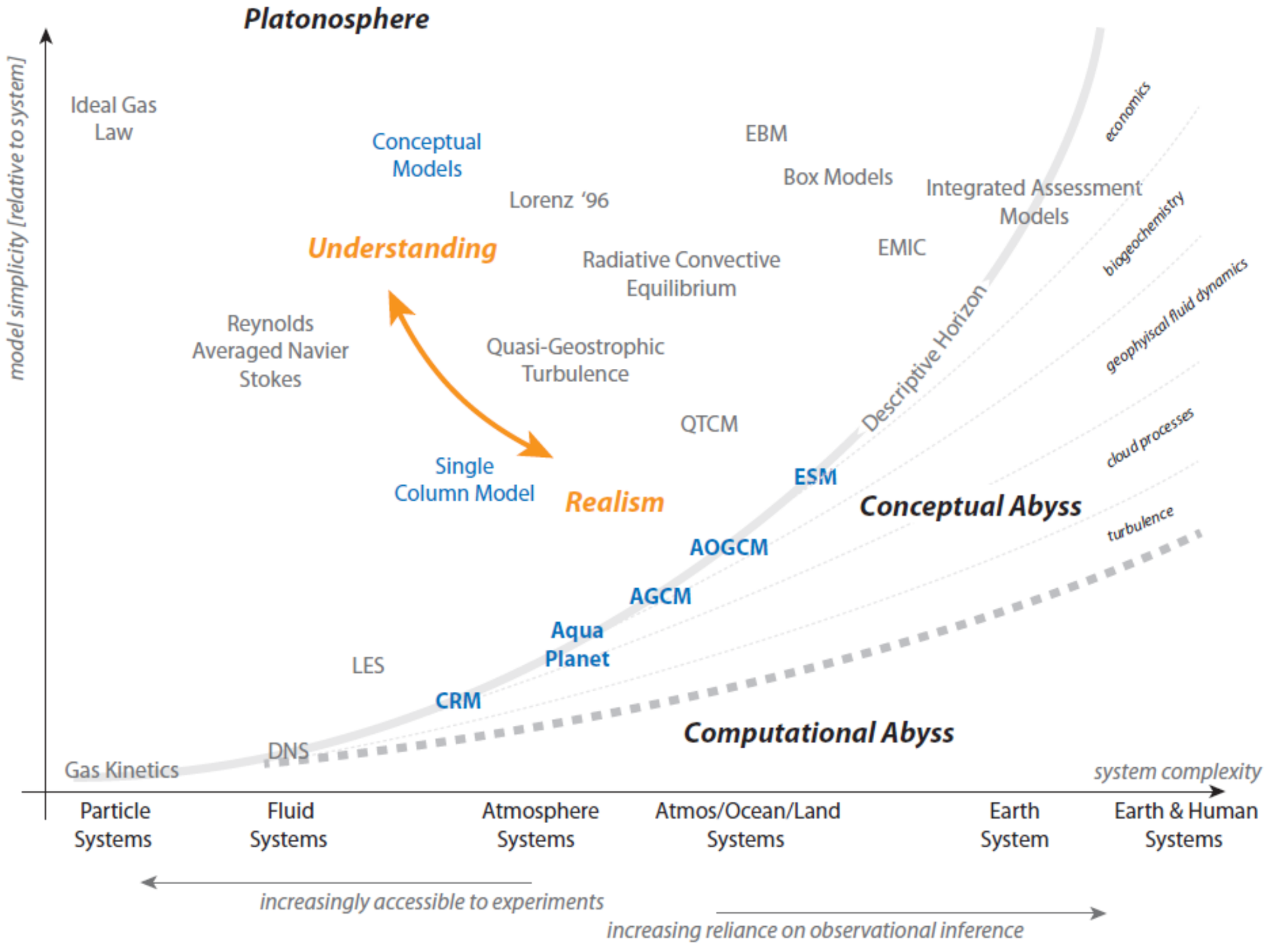
Max Planck Institute for Meteorology

http://mpimet.mpg.de

- Located in Hamburg, Germany
- Established in 1972
  - Former directors Hinzpeter, Hasselmann, Bengtsson, Brasseur, Grassl
- About 250 People
- Earth system research
  - Atmosphere in the Earth System, Björn Stevens, Mg. Dir.
  - Land in the Earth System, Martin Claussen
  - Ocean in the Earth System , Jochem Marotzke
  - International Max-Planck-Research School on Earth System Modelling
- DKRZ separate entity

Models at MPI-M

- From simple 0-D to comprehensive ESMs



From Bony et.al, 2011, Carbon Dioxide and Climate: Perspectives on a Scientific Assessment

Models at MPI-M

- From simple 0-D to comprehensive ESMs
- AES: DNS, LES, ECHAM, HAMOZZ
- LES: JSBACH, Climber
- OES: MPI-OM, HAMOCC
- MPI-M:
  - MPI-ESM1
  - ICON
    - Dynamics, together with DWD
    - "Physics" from departments

Models at MPI-M

- From simple 0-D to comprehensive ESMs
- AES: DNS, LES, ECHAM, HAMOZZ
- LES: JSBACH, Climber
- OES: MPI-OM, HAMOCC
- MPI-M:
  - MPI-ESM1
  - ICON
    - Dynamics, together with DWD
    - "Physics" from departments

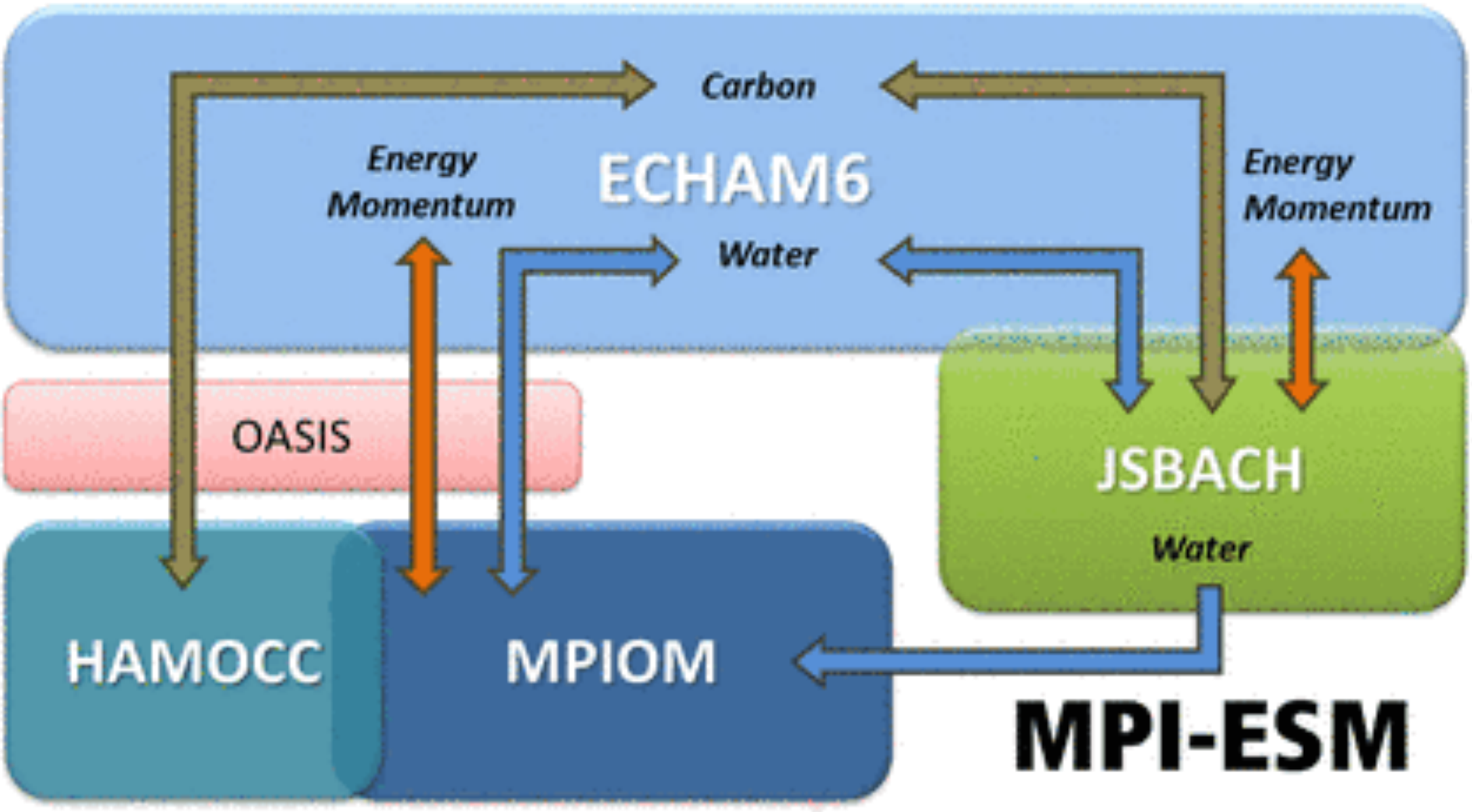
Performance

- Scientific
  - To assemble the right configuration for a given scientific question / project / experiment ....
  - ... that can be done in an acceptable time frame

Performance

- Scientific
  - To assemble the right configuration for a given scientific question / project / experiment ....
  - ... that can be done in an acceptable time frame
- Technical
  - can be done
    - on IS permanently
    - on those code pieces assembled for the scientific experiment only after the code is delivered
      - risk of failure!

Max-Planck-Institute Earth-System-Model 1



Improvements from COSMOS version

- Advanced/improved treatment of **radiative transfer**
- Improved representation of **surface albedo**
- New and improved representation of the **aerosol**
- Much better representation of the **middle atmosphere**
- A capacity to simulate at a range of **different resolutions** depending on the question
- **Interactive vegetation dynamics**
- **Coupled carbon cycle**

Configurations of MPI-ESM1		
Name	Resolution	Remarks
Millenium	T31L19-GR30L40	COSMOS Millenium configuration, NEC
CR	T31L31-GR30L40	Summer school model Top at 10 hPa; IBM
LR	T63L47/GR15L40	up to 0.01 hPa (same as L31 to 10 hPa), Stratosphere resolved; IBM
MR	T63L95/TP04L40	ocean eddy permitting in many regions Strat. better resolved, QBO; IBM
P	T63L47/GR15L40	fixed orbit no dyn veg; IBM
HR	T127/L95/TP04L40	not published too expensive; IBM
XR	T255L199/TP6ML80	spin-up is a severe problem really expensive; IBM

Experiments

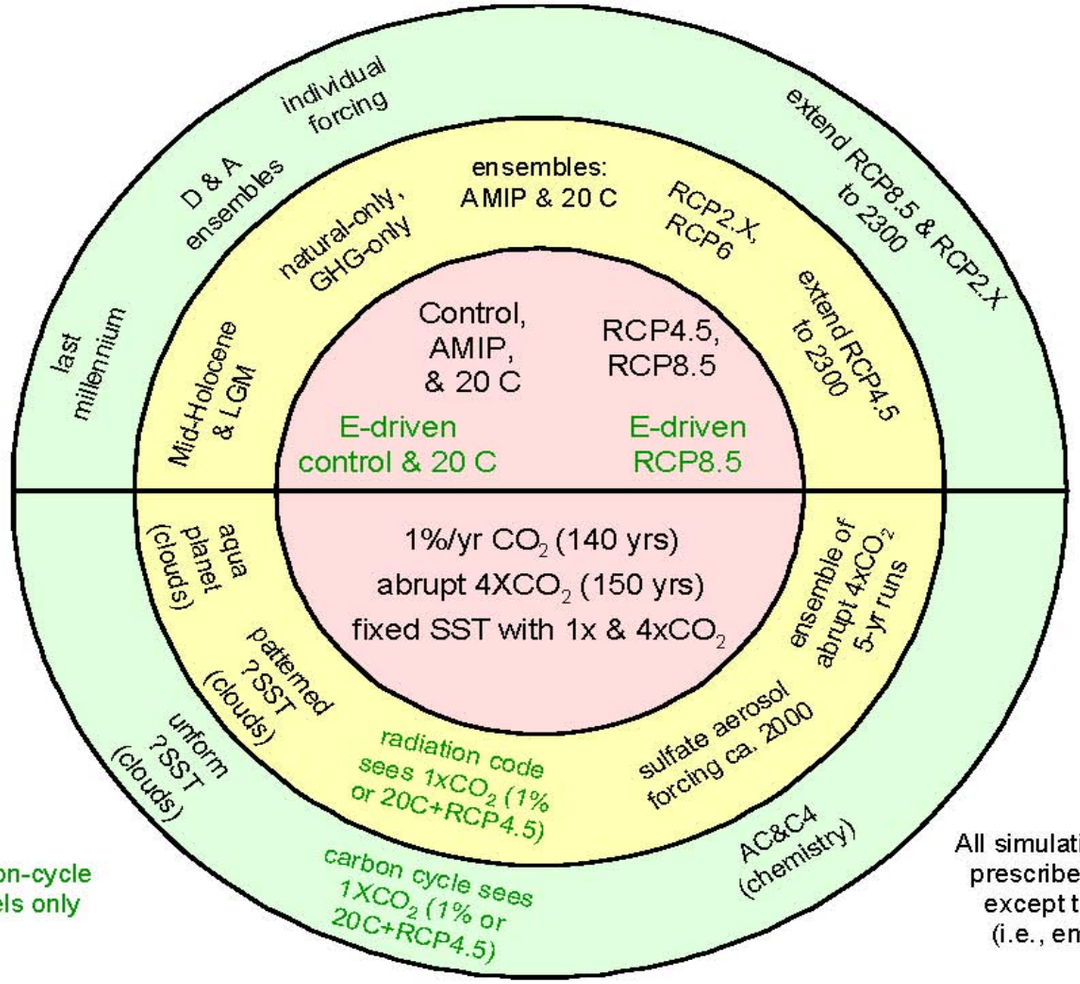
- Millenium
  - COSMOS → MPI-ESM Millenium Version, LR
- CMIP5
  - MPI-ESM1, LR, MR, HR

Millenium

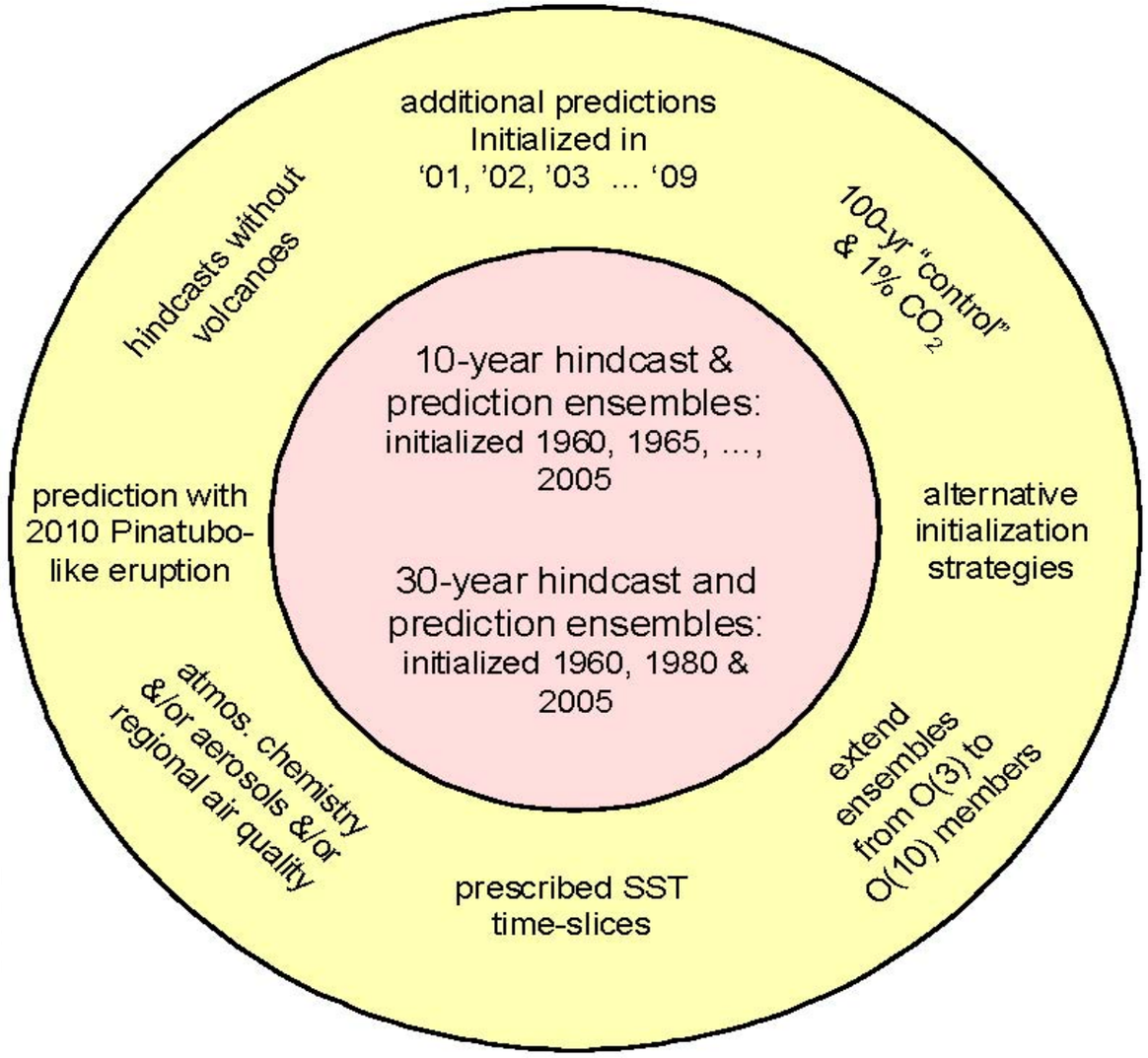
Millenium

- T31L19 ECHAM JSBACH / MPIOM3°L40 HAMOCC
- Full carbon cycle
- > 15000 yrs with COSMOS configuration
- many more to get a balanced carbon cycle
- 10s of person years
- substantial part of the DKRZ Hurrikan machine

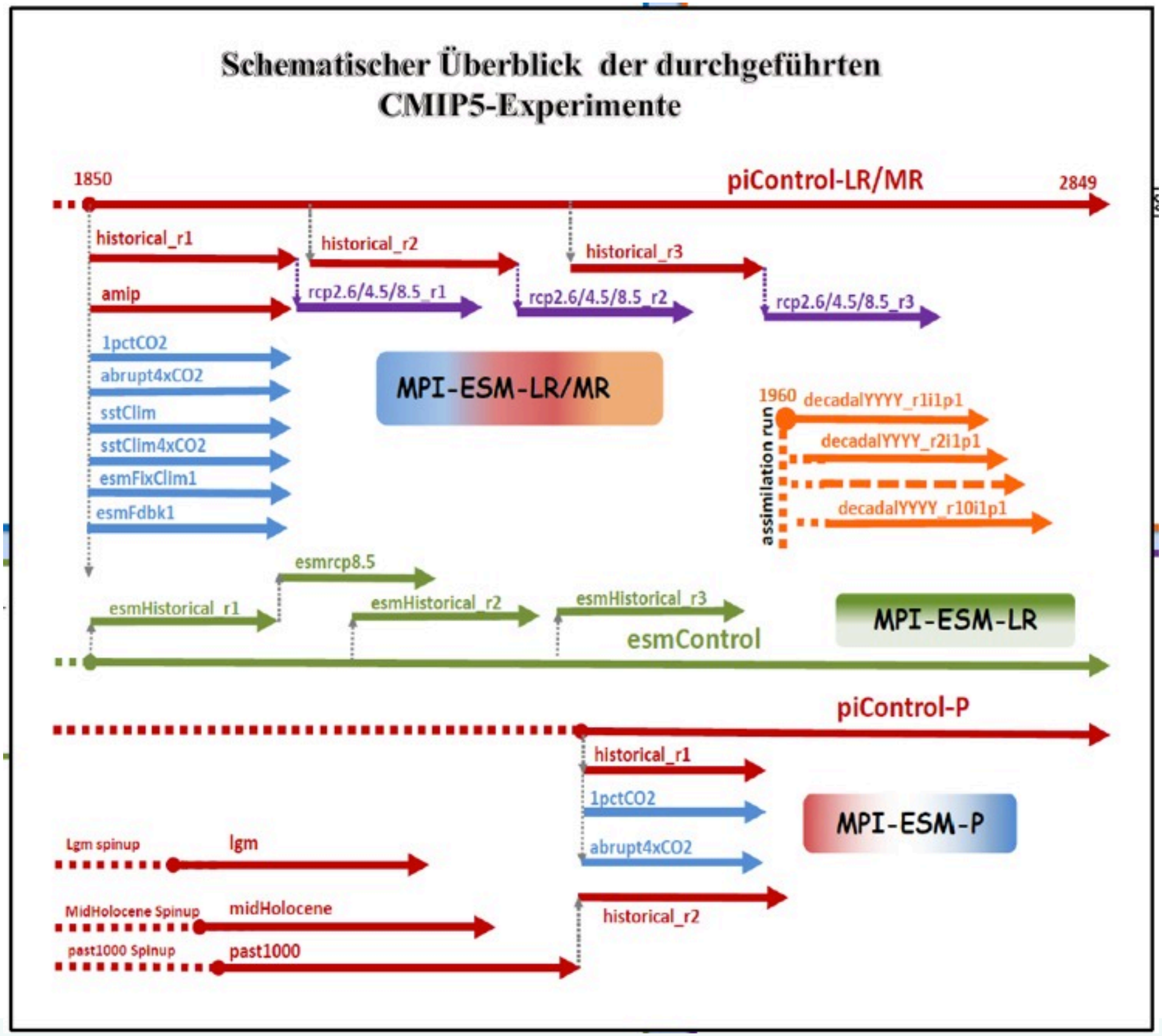
CMIP5



All simulations are 1
prescribed concen
except those "E-
(i.e., emission-dr



Schematischer Überblick der durchgeführten CMIP5-Experimente



CMIP5

CMIP5

- Modell develpoment, tuning, spinup
- Experiments at DKRZ:
- Post-Processing
- Quality checks
- “CMOR” formats and variables
- Input for Regionalisation
- Transfer of CMOR data into ESGF
- Publication with DOIs

CMIP5

- Modell develpoment, tuning, spinup
- Experiments at DKRZ:
- Post-Processing
- Quality checks
- “CMOR” formats and variables

MPI-ESM-LR	259 simulations	7751 yrs
MPI-ESM-P	10 simulations	3028 yrs
MPI-ESM-MR	72 simulations	2803 yrs

- Input for Regionalisation
- Transfer of CMOR data into ESGF
- Publication with DOIs

CMIP5

- Modell develpoment, tuning, spinup
- Experiments at DKRZ:
- Post-Processing
- Quality checks
- “CMOR” formats and variables

MPI-ESM-LR	259 simulations	7751 yrs
MPI-ESM-P	10 simulations	3028 yrs
MPI-ESM-MR	72 simulations	2803 yrs

- Input for Regionalisation
- Transfer of CMOR data into ESGF
- Publication with DOIs

~18 mth of 1/4 of DKRZ IBM P6

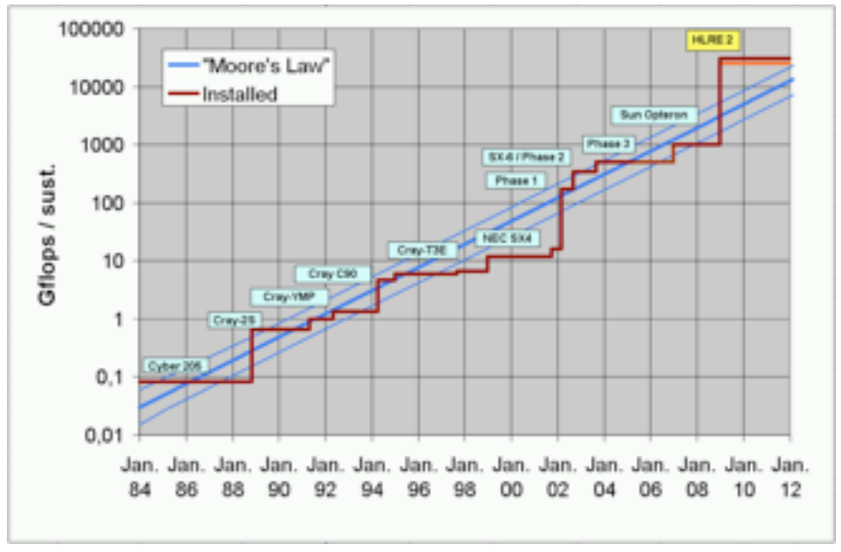
10s of person years

Performance Metrics

- Discussion in (IS)ENES (HPCTF)
- Forecast [yrs/mths/d] / day (on a given platform)
  - –Detailed specification of
    - model (aka system complexity)
    - machine/environment (compilers!)
  - –„CIM“?

- Degrees of freedom need to be added
  - –gird points, timespteps, number of variables

Performance



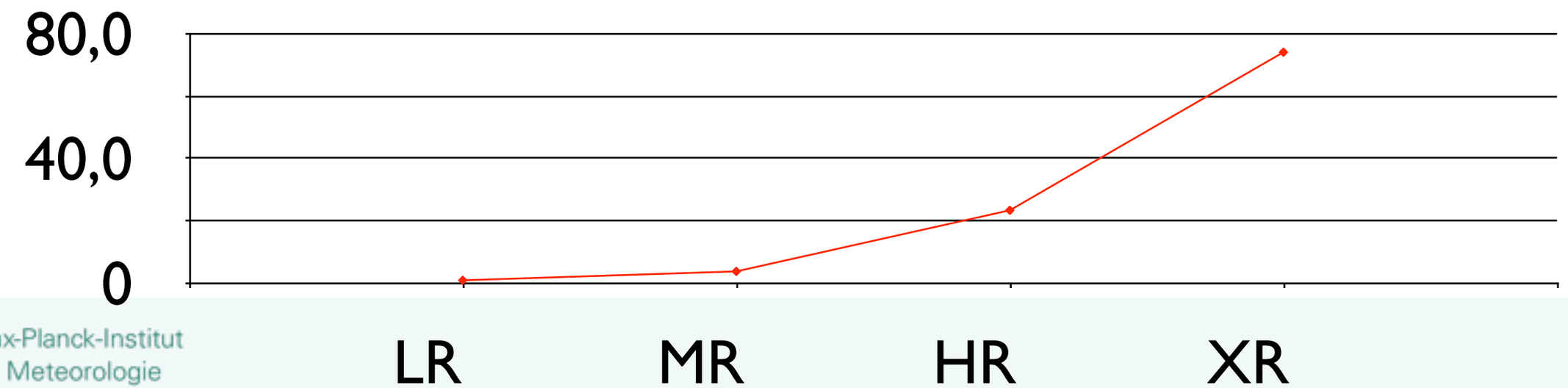
- Blizzard@DKRZ
  - –IBM p575 Power 6 cluster
  - –Peak performance: 158 TeraFlop/s
  - –264 IBM Power6 nodes
  - –16 dual core 4.7 GHz CPUs per node (8448 cores)
  - –20 TB of main memory and 7 PB of disk space
  - –Infiniband network: 7.6 TB/s (aggregated)

Performance [FCyrs/day]

NEC SX9 | IBM P6

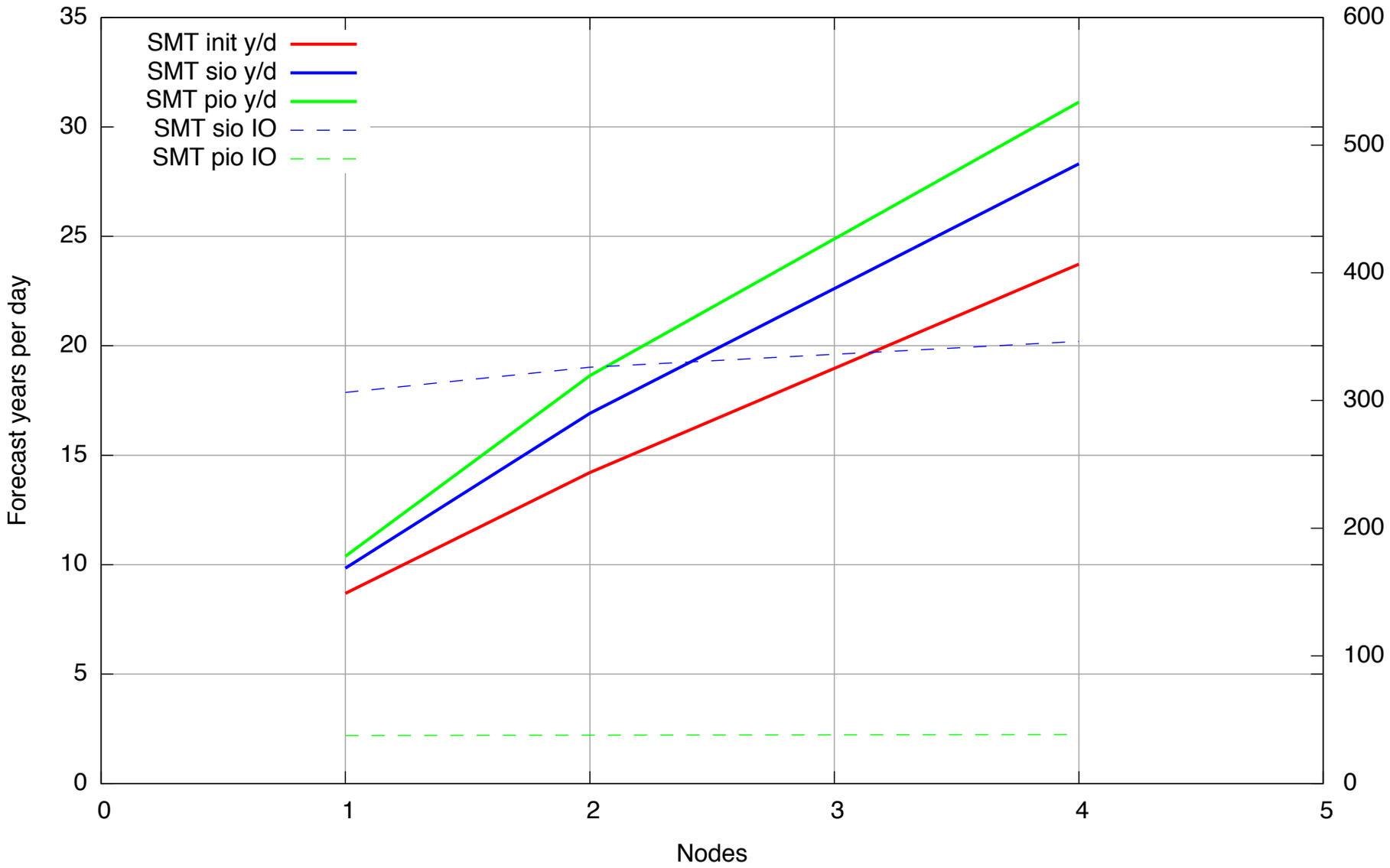
Mill.	~30			on NEC, T31L19/OM3°L40
Model	***FCyrs / d***	***# nodes DKRZ blizzard***	***„normed efficiency“***	***Remark***
LR	18,4 *	4	1	4 nodes, 4*64=256 MPI-tasks (16x12 (echam6) + 9x7 (mpiom) + 1(oasis3)) => 4700 s/yr => 18.4 FCyrs/d; MCT I/O
MR	10,7	10	3,51	10 nodes, 10*64=640 MPI-tasks (16x16 (echam6) + 23*16 (mpiom) + 16(oasis3)) => 8100 s/yr => 10.7 FCyrs/d
HR	~ 5	30	23,44	Way too expensive!
XR	0,85	63	74,11	63 nodes, 2001 MPI-tasks (with decomposition 32x24 (echam6) + 38*32 (mpiom) +17(oasis3)) => 102000 s/yr => 0.85 FCyrs/d

*Was ~ 15 for CMIP5



A Word on I/O

- See presentation by Luis Kornblueh during workshop



Performance improvement by about 15 %.
Done in cdi-pio, also helps analysis

Towards MPI-ESM2: ICON

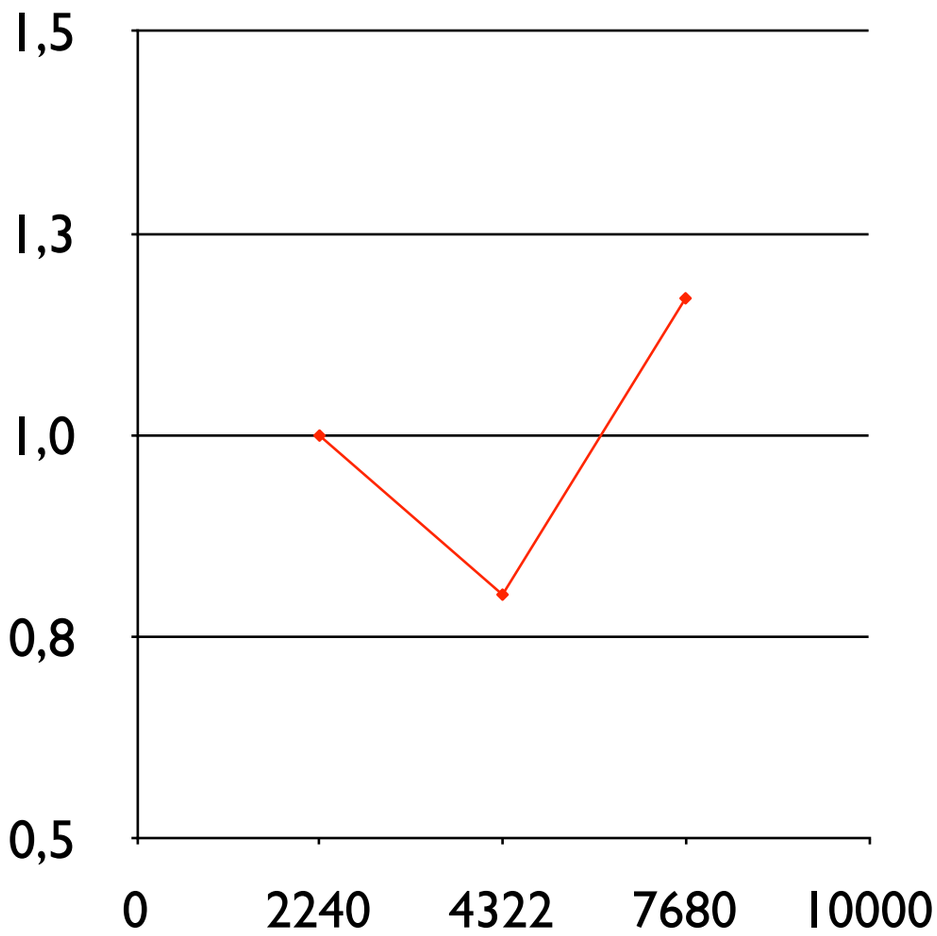
- See presentation ICON by Günther Zängl, DWD
- Trick for us the implementation of "our" physics
  - Technically
    - Don't spoil the code
      - readability
      - reliability
      - performance
      - ease of use
  - Scientifically
    - Reproduce results for the same cases

ICON Performance

- Experiment Setup
  - non-hydrostatic APE
- Grid resolution: 35km
- Vertical levels: 96
- Dynamics step: 60 secs
- Physics step: 300 secs
- Radiation step: 1800 secs
- No I/O except basic diagnostics

- Machine:
  - IBM P6,
  - 1 mpi process per core,
  - 2 threads (SMT mode)

P6 Cores	FCyrs/day	„normed efficiency“
2240	2,24	1
4322	5,38	0,803345725
7680	6,52	1,17791411



Performance

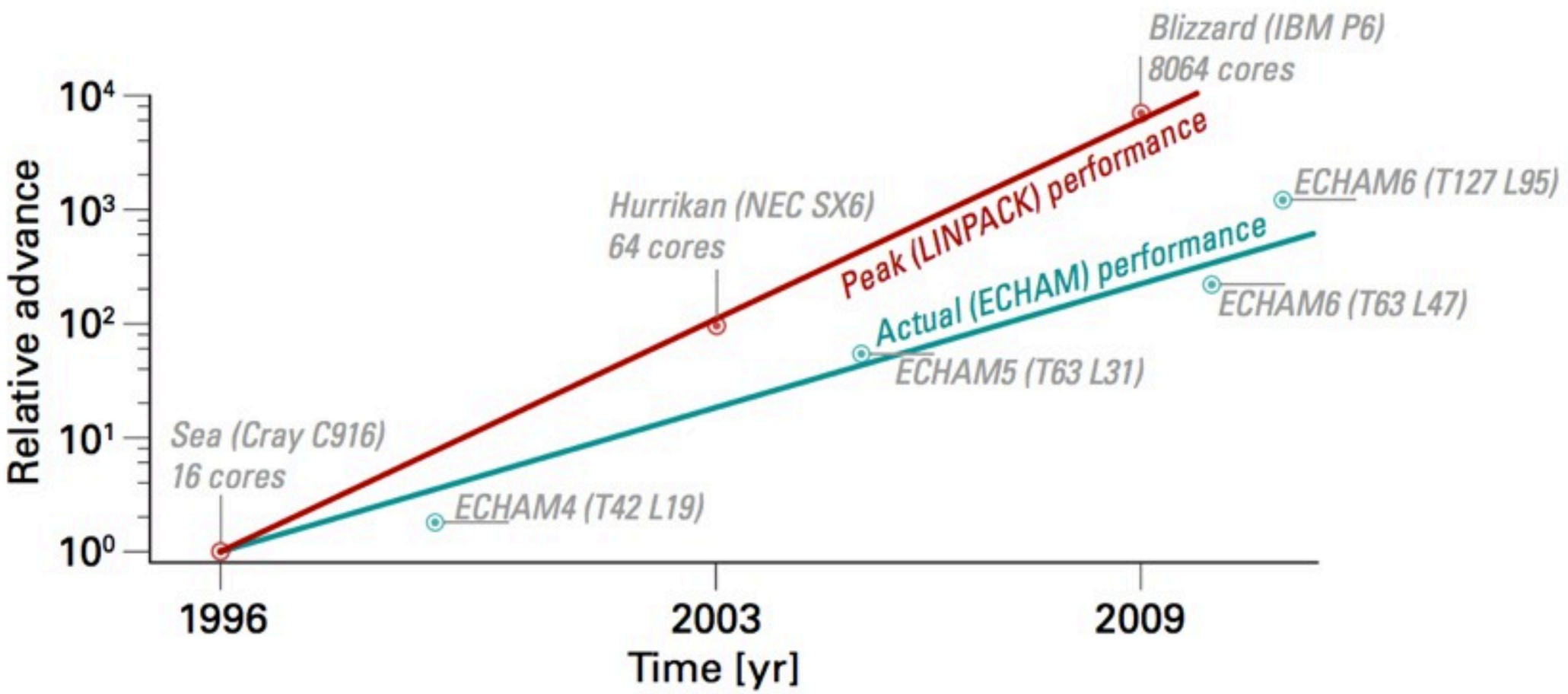
- Overall performance of coupled models looks flat:
  - 10s of FCyrs/day as a constant?
- Models improve in scientific performance
- Optimization in terms of technical performance is often eaten up by better scientific performance
- You can only optimize against single routines, and their interplay
  - Stable code helps
  - Isolate critical portions
  - If profile of modules is flat, rethink!

Summary and Conclusions

- Improving performance is a very expensive business
  - Needs to involve many layers of expertise
    - Physics/Components
    - Numerics
    - SysAdmin

- Special problem MPI-M: We need to tame two models at the same time
  - MPIESM1 still major scientific workhorse
  - Major development work needs to go into MPIESM2

Be aware:


Computing struggles to keep up with computers
Relative advance
10^4
10^3
10^2
10^1
10^0
Blizzard (IBM P6)
8064 cores
Hurrikan (NEC SX6)
64 cores
Peak (LINPACK) performance
ECHAM6 (T127 L95)
Actual (ECHAM) performance
ECHAM6 (T63 L47)
ECHAM5 (T63 L31)
Sea (Cray C916)
16 cores
ECHAM4 (T42 L19)
1996
2003
2009
Time [yr]