Large scale numerical simulations of the climate

Jean-Christophe Rioual, Climate Research IT

Met Office Hadley Centre
Met Office HQ locations
Exeter, Devon
## Basic facts

<table>
<thead>
<tr>
<th>Category</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover</td>
<td>~£208m (approx 16% Commercial)</td>
</tr>
<tr>
<td>People</td>
<td>~2000 Staff</td>
</tr>
<tr>
<td></td>
<td>~1400 at Exeter HQ</td>
</tr>
<tr>
<td>Locations</td>
<td>~50 manned locations</td>
</tr>
<tr>
<td></td>
<td>~Many more unmanned observing sites</td>
</tr>
<tr>
<td></td>
<td>~inc 5 permanent and 2 Mobile Met Unit overseas sites</td>
</tr>
<tr>
<td>Working areas</td>
<td>37% Forecasting &amp; Observations</td>
</tr>
<tr>
<td></td>
<td>28% Science &amp; Research</td>
</tr>
<tr>
<td></td>
<td>16% Technology (IT)</td>
</tr>
<tr>
<td></td>
<td>12% Commercial and Government Business</td>
</tr>
<tr>
<td></td>
<td>7% Corporate Services</td>
</tr>
</tbody>
</table>
Predictions across all timescales

Analysis of past weather observations to manage climate risks

Eg. Agriculture: informs crop choice, planting to yield optimisation and minimise crop failure risk.

Predicting routine and hazardous weather conditions.

Public, emergency response, international Disaster Risk Reduction

Monthly to decadal predictions - probability of drought, cold, hurricanes….

Contingency planners, national and international humanitarian response, government and private infrastructure investment

Global and regional climate predictions.

Informs mitigation policy and adaptation choices. Impacts on water resources, heat stress, crops, infrastructure.
Lewis Fry Richardson 1922

~64000 human ‘computers’
Computing - Yesterday

- 1959
- Ferranti Mercury
- 30000 FLOPS
Computers Used for Weather and Climate Prediction

Peak performance (FLOPS)

Year of First Use

1.0E+01 1.0E+02 1.0E+03 1.0E+04 1.0E+05 1.0E+06 1.0E+07 1.0E+08 1.0E+09 1.0E+10 1.0E+11 1.0E+12 1.0E+13 1.0E+14 1.0E+15 1.0E+16 1.0E+17

Cray XC40+ Cray XC40
IBM Power 7 IBM Power 6
Cray T3E NEC SX-8
NEC SX-6 Cray C-90
Cray Y-MP ETA-10
Cyber 205
IBM 360/195
KDF-9
Mercury
Leo

Best Fit
Moore’s Law
Computing - Today

Cray XC40

- Intel Broadwell processors
- 2 sockets, 36 cores per node
- 128GB RAM per node
- Proprietary interconnect Aries
Cray XC40

Operational Clusters
- NWP
- 2 x 2492 nodes (redundancy)
- 6 PB Lustre FS
- Met Office network

Research Cluster
- Research
- 6720 nodes
- 12 PB Lustre FS
- Shared Facility
- Largest research HPC for climate/weather (11 Top500)
Met Office Atmospheric Simulation Model

- Numerical Weather Prediction and Climate Modelling
- General Circulation Model
- ~1 million LOC
- 4 releases a year
- 100+ active developers per release
- Worldwide user community
Components

Dynamical Core - ENDgame
- Equations of motion on a sphere
- Finite Differences
- 3D Latitude-Longitude rectangular grid
- MPP code – 2D domain decomposition
- MPI + OpenMP
Components

Dynamical Core - ENDgame
- Equations of motion on a sphere
- Finite Differences
- 3D Latitude-Longitude rectangular grid
- MPP code – 2D domain decomposition
- MPI + OpenMP

Physical parametrisations
- Not fully resolved
- Convection, Clouds
- Column-based

Subsystems
- Land surface
- Chemistry
Coupled Models

- Particularly important on climatic timescales
- Met Office HadGEM3 model
Climate Models

- Complexity
- Feedback loops
Climate Models Numerics

- Validated until present day
- No reference solution past present day
- Societal demand for accuracy
Ensemble systems

- Perturbed initial conditions of the model
Ensemble systems

- Better: ensemble of ensembles
- Intercomparison of models
- Specific model biases compensated

Worldwide Coordinated Numerical Simulation Campaigns
CMIP6

- Organised by World Climate Research Program
- 5 years organisation
- Complex experimental apparatus
- Results analysed in time for IPCC AR6 in 2023
- Data hosted in ESGF database ~20+ Pbytes
- Billions of core hours
Workflows

- **Workstation**
  - Build Model

- **Cray HPC**
  - Jan 1978
  - Feb 1978
  - Dec 2099

- **Post-Processing Platforms, Archives**
  - Process Data
  - Process Data
  - Process Data

- **Fully automated** (Rose)
- **Version controlled**
Numerical Reproducibility

Cray HPC

Jan 1978 → Feb 1978 → Dec 2099

Archives

Jan 1999 → Cray HPC → Mar 2030
Numerical Reproducibility

Cray HPC

Jan 1978 ➔ Feb 1978 ➔ Dec 2099

Archives

Bit level Reproducibility

Jan 1999 ➔ Mar 2030
Numerical Reproducibility

Cray HPC

Jan 1978 → Feb 1978 → Dec 2099

Archives

Jan 1978 → Feb 1978 ← Dec 2099

Bit level Reproducibility

Cray HPC

Jan 1999 ← Feb 1999 ← Dec 2030

Archives

Jan 1999 → Feb 1999 → Dec 2030
Numerical reproducibility

Good Coding

```fortran
use timestep, only : current_timestep
use submodel, only : initialise, integrate

if (current_timestep().eq.0) then
   call initialise()
else
   call integrate(current_timestep())
end if
```
Numerical reproducibility

Good Coding

```fortran
use timestep, only : current_timestep
use submodel, only : initialise, integrate

if (current_timestep().eq.0) then
    call initialise()
else
    call integrate(current_timestep())
end if
```

Bad/Lazy Coding

```fortran
use my_favorite_module
logical first_call = .true.

if (first_call) then
    call interpolate(a,a_grid)
    first_call=.false.
end if

call do_stuff(a)
```
Quality Assurance

- *is important*
- Coding Standards
- Code reviews
- Unit tests
- Version Control

- Investment
- But can save a lot of time (and computing resources)
Numerical Reproducibility

Cray HPC

Jan 1978 ➔ Feb 1978 ➔ Dec 2099

m nodes

Archives

Cray HPC

Jan 1999 ➔ Mar 2030

n nodes
Bit Reproducibility

Across domain decompositions

\[ m \times n \text{ vs } n \times m \text{ vs } k \times l \]

All numerical computations are local except
• Halos Exchange
• Global Sums (Iterative Linear Solvers)
Bit reproducible global sums

Original implementation
Bit reproducible global sums

Original implementation
Bit reproducible global sums

Original implementation
Bit reproducible global sums

Original implementation
Bit reproducible global sums

New implementation

\[ \text{Local sum (+] \quad (+)} \]
Bit reproducible global sums

New implementation

- Local sum (+)
- Double-Double (complex) precision
- Knuth, D.H Bailey
Bit reproducible global sums

New implementation

- Local sum (+)
- (+)
- **ALL_REDUCE**
- Double-Double (complex) precision
- Knuth, D.H Bailey

- **MPI_ALLREDUCE**
- Double-Double Operator
Bit reproducible global sums

New implementation

- Local sum (+)
- (+)
- ALL_REDUCE

- Double-Double (complex) precision
- Knuth, D.H Bailey

- MPI_ALLREDUCE
- Double-Double Operator

<table>
<thead>
<tr>
<th>Cores</th>
<th>Double</th>
<th>Double-Double</th>
</tr>
</thead>
<tbody>
<tr>
<td>896</td>
<td>2264 s</td>
<td>2186 s</td>
</tr>
</tbody>
</table>
Bit reproducible global sums

- Do we really need this?
- *It is a proxy for good quality code (QA)*
- Coarse grained *Unit Test*
Using single precision

Always use double precision? … No.

Understanding the error
Accuracy of Krylov subspace solver – BiCGStab
Iterative solver, it improves the answer each iteration

In our model $\varepsilon < 10^{-3}$
answer is good enough
What precision is needed to satisfy this?
Single precision is good enough

$$\left\| \mathbf{r}_k \right\| = \left\| \mathbf{b} - A\mathbf{u}_k \right\| < \varepsilon$$

Modern FPU single prec operation is not significantly faster than double
Single prec words are half the size of double prec words
Compute values of $A$ in double precision
Store them in single precision
Doubles the effectiveness of cache
Max total time for TRI_SOR

diagonal projection

time (secs)

2.0
1.0
0.5
0.25
0.125

24x32
32x32
48x48
64x64
96x96

total cores (EWxNS)

- EW=NS
- EW<NS
- EW>NS
- ideal
Max total time for TRI_SOR
diagonal projection

- EW=NS
- EW<NS
- EW>NS
- ideal
- Hlm single-prec

Graph showing the relationship between total cores (EWxNS) and time (secs) for different configurations.
Accuracy of answers

After 5 time-steps, level 10 biggest differences

\[ f = (\xi_{64} - \xi_{32}) \times 10^{-4} \]

\[ \xi = \text{exner} = \left( \frac{p}{p_0} \right)^k \]

24x32 EWxNS Proc
1024x769x70 grid

Plot by T. Allen
Accuracy of answers

After 5 time-steps, level 10 biggest differences

\[ f = (\xi_{64} - \xi_{32}) \times 10^{-4} \]

\[ \xi = \text{exner} = \left( \frac{p}{p_0} \right)^k \]

24x32 EWxNS Proc 1024x769x70 grid

<table>
<thead>
<tr>
<th>Method</th>
<th>64 bit</th>
<th>32 bit</th>
<th>Speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>N512 6 time-steps (s)</td>
<td>3.884</td>
<td>2.836</td>
<td>1.4</td>
</tr>
<tr>
<td>EG_SL_HELMHOLZ</td>
<td>2.876</td>
<td>1.809</td>
<td>1.6</td>
</tr>
<tr>
<td>EG_BICGSTAB</td>
<td>2.075</td>
<td>1.124</td>
<td>1.8</td>
</tr>
<tr>
<td>TRI_SOR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Accuracy of answers

After 5 time-steps, level 10 biggest differences

\[ f = (\xi_{64} - \xi_{32}) \times 10^{-4} \]

\[ \xi = \text{exner} = \left( \frac{p}{p_0} \right)^k \]

24x32 EWxNS Proc 1024x769x70 grid

<table>
<thead>
<tr>
<th>Method</th>
<th>64 bit</th>
<th>32 bit</th>
<th>Speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>N512 6 time-steps (s)</td>
<td>3.884</td>
<td>2.836</td>
<td>1.4</td>
</tr>
<tr>
<td>EG_SL_HELMHOLZ</td>
<td>2.876</td>
<td>1.809</td>
<td>1.6</td>
</tr>
<tr>
<td>EG_BICGSTAB</td>
<td>2.075</td>
<td>1.124</td>
<td>1.8</td>
</tr>
<tr>
<td>TRI_SOR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Single precision physics

• Precision required by subscale phenomena?
• Less obvious answer than for Krylov solver
• Ongoing research work
• But careful assessment required
32-bit LS-Precip: Early Results

- Build-time selection of scheme precision
- Minimal effect on model evolution
- Retune segment size to aid cache blocking
- Modest speedup. More understanding required.
32-bit GW-Drag: Early Results

Same implementation approach as LS-Precip

• Very different experience

Technically much easier, but…

…minimal performance gain at 32-bit

…big change in model evolution at 32-bit

• Cause still under investigation
Future Architectures

- We are not tied to a particular architecture
- Have changed in the past (vector to MPP)
- Engagement with new technologies
Future Architectures

- We are not tied to a particular architecture
- Have changed in the past (vector to MPP)
- Engagement with new technologies

- HPC Tier 2 system *Isambard*
- Cray inc
- Met Office
- GW4 Universities
  - Cardiff, Bristol, Bath, Exeter
Future Architectures

Phase 1
- 8x1 Intel Xeon Phi nodes
- 4x2 NVIDIA Pascal GPUs
- InfiniBand connection

HPC Tier 2 system Isambard
- Cray inc
- Met Office
- GW4 Universities
  - Cardiff, Bristol, Bath, Exeter
Future Architectures

**Phase 1**
- 8x1 Intel Xeon Phi nodes
- 4x2 NVIDIA Pascal GPUs
- Infiniband connection

**Phase 2**
- + "10000+" ARMv8 cores

- HPC Tier 2 system *Isambard*
- Cray inc
- Met Office
- GW4 Universities
  - Cardiff, Bristol, Bath, Exeter
Performance Portability

- Unified Model
- Fortran90 + MPI + OpenMP
- 25 years development
- CUDA, Open ACC difficult to implement

- New LFRic model
- Complete rewrite
- Finite elements

- Opportunity
- Modern Programming Paradigms
- Python generating F2003 code (PsyKAI) / STFC
- Architecture Agnostic
- Will be used for 20+ years
Summary

- Insatiable computing demands
- Numerical reproducibility important
- (Some) flexibility on precision
- Involvement with new architectures
- and new programming paradigms

Thank You!

Questions?

jean-christophe.rioual@metoffice.gov.uk