The use of inexact hardware to improve weather and climate predictions

Peter Düben, Tim Palmer

University of Oxford

Why should we use inexact hardware in weather and climate predictions?

- Double precision is used as standard in almost all global weather and climate models.
- Inexact hardware allows a reduction of power consumption and an increase in performance.
- This would allow simulations at higher resolution and possibly more accurate forecasts.

What is inexact hardware?

My definition: Inexact hardware is using less than double precision (64 bits).

What is inexact hardware?

My definition: Inexact hardware is using less than double precision (64 bits).

Easy: double $\rightarrow$ single $\rightarrow$ half (e.g. NVIDIA's Pascal GPUs).

What is inexact hardware?

My definition: Inexact hardware is using less than double precision (64 bits).

Easy: double → single → half (e.g. NVIDIA's Pascal GPUs).

Hard work: Field Programmable Gate Arrays (FPGAs).

What is inexact hardware?

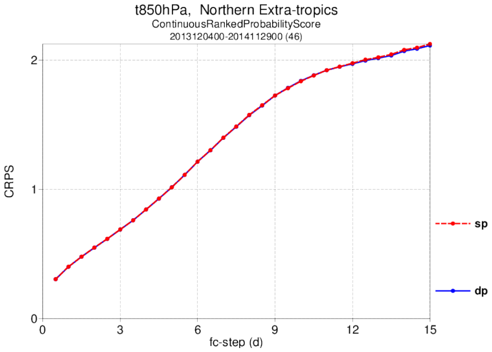
My definition: Inexact hardware is using less than double precision (64 bits).

Easy: double $\rightarrow$ single $\rightarrow$ half (e.g. NVIDIA's Pascal GPUs).

Hard work: Field Programmable Gate Arrays (FPGAs).

Future perspective: Pruned hardware or stochastic processors.

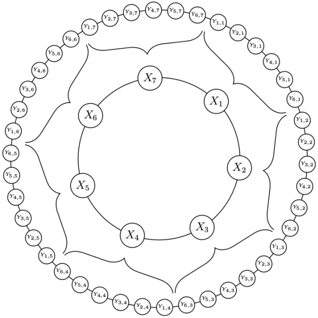
Example 1: IFS in single precision



- Ensemble forecasts and long-term simulations in double and single precision at T399 resolution are almost identical.
- ≈40% speed-up.
- Filip Vana is investigating single precision at ECMWF.

Düben et al. MWR 2015, Váňa et al. submitted.

Example 2: Lorenz ’96 on FPGAs



- We implemented the Lorenz ’96 model on FPGAs in cooperation with the group of Wayne Luk from Imperial College.
- We scale the size of Lorenz ’96 to the size of a high performance application with more than 100 million degrees-of-freedom.
- We compare results with reduced precision against results with perturbed parameters (by 1 %) or parametrised small scales.

Example 2: Lorenz ’96 on FPGAs

Model quality

Model setup	Hellinger distance
c and F times 1.01, single precision	0.005
Parametrised small scales, single precision	0.114
Reduced precision: 17 bits for X, 14 bits for Y	0.008

Speed and Power

Hardware	Speed	Energy efficiency
CPU, 12 cores, single precision	1.0	1.0
FPGA, single precision	2.8	10.4
FPGA, 17 bits for X, 14 bits for Y	6.9	23.9

Example 2: Lorenz ’96 on FPGAs

Model quality

Model setup	Hellinger distance
c and F times 1.01, single precision	0.005
Parametrised small scales, single precision	0.114
Reduced precision: 17 bits for X, 14 bits for Y	0.008

Speed and Power

Hardware	Speed	Energy efficiency
CPU, 12 cores, single precision	1.0	1.0
FPGA, single precision	2.8	10.4
FPGA, 17 bits for X, 14 bits for Y	6.9	23.9

Significant savings and no strong decrease in model quality.

Düben et al. JAMES 2015, Russel et al. FCCM 2015.

Example 3: Reduced precision in a spectral dycore

- We calculate weather forecasts with a spectral dynamical core (IGCM) in a “Held-Suarez world” and compare results against a high resolution truth.
- Floating point precision for the significand is reduced to 8 bits (instead of 52) using an emulator. Only 2% of the reduced precision simulation is calculated in double precision.
- We estimate savings for pruned hardware in cooperation with computer scientists (Krishna Palem Rice University, Christian Enz EPFL and John Augustine IITM).

Example 3: Reduced precision in a spectral dycore

- We calculate weather forecasts with a spectral dynamical core (IGCM) in a “Held-Suarez world” and compare results against a high resolution truth.
- Floating point precision for the significand is reduced to 8 bits (instead of 52) using an emulator. Only 2% of the reduced precision simulation is calculated in double precision.
- We estimate savings for pruned hardware in cooperation with computer scientists (Krishna Palem Rice University, Christian Enz EPFL and John Augustine IITM).

Resolution	Precision FP significand	Normalised Energy Demand	Mean error Z500 at day 2
235 km	52	1.0	2.3
315 km	52	0.47	4.5
235 km	**8**	0.29	2.5

Example 3: Reduced precision in a spectral dycore

- We calculate weather forecasts with a spectral dynamical core (IGCM) in a “Held-Suarez world” and compare results against a high resolution truth.
- Floating point precision for the significand is reduced to 8 bits (instead of 52) using an emulator. Only 2% of the reduced precision simulation is calculated in double precision.
- We estimate savings for pruned hardware in cooperation with computer scientists (Krishna Palem Rice University, Christian Enz EPFL and John Augustine IITM).

Resolution	Precision FP significand	Normalised Energy Demand	Mean error Z500 at day 2
235 km	52	1.0	2.3
315 km	52	0.47	4.5
235 km	**8**	0.29	2.5

To save power a reduction in precision is much more efficient when compared to a reduction in resolution.

Düben et al. MWR 2015, Düben et al. DATE 2015.

A precision analysis to improve models

- The influence of rounding errors on model dynamics can be described as random forcing that is added to the differential equations.
- The level of precision will influence the magnitude of this forcing.
- Let’s assume that we can identify the optimal level of precision for all parts of an atmosphere model.

A precision analysis to improve models

- The influence of rounding errors on model dynamics can be described as random forcing that is added to the differential equations.
- The level of precision will influence the magnitude of this forcing.
- Let's assume that we can identify the optimal level of precision for all parts of an atmosphere model.

Rounding errors can be used to identify model parts that do not have a strong influence on model dynamics and can be removed.

A precision analysis to improve models

- The influence of rounding errors on model dynamics can be described as random forcing that is added to the differential equations.
- The level of precision will influence the magnitude of this forcing.
- Let's assume that we can identify the optimal level of precision for all parts of an atmosphere model.

Rounding errors can be used to identify model parts that do not have a strong influence on model dynamics and can be removed.

A precision analysis can be used to secure double precision simulations against rounding errors.

A precision analysis to improve models

- The influence of rounding errors on model dynamics can be described as random forcing that is added to the differential equations.
- The level of precision will influence the magnitude of this forcing.
- Let's assume that we can identify the optimal level of precision for all parts of an atmosphere model.

Rounding errors can be used to identify model parts that do not have a strong influence on model dynamics and can be removed.

A precision analysis can be used to secure double precision simulations against rounding errors.

Numerical precision can be reduced with time in weather forecasts to allow long term high resolution simulations.

A precision analysis to improve models

- The influence of rounding errors on model dynamics can be described as random forcing that is added to the differential equations.
- The level of precision will influence the magnitude of this forcing.
- Let's assume that we can identify the optimal level of precision for all parts of an atmosphere model.

Rounding errors can be used to identify model parts that do not have a strong influence on model dynamics and can be removed.

A precision analysis can be used to secure double precision simulations against rounding errors.

Numerical precision can be reduced with time in weather forecasts to allow long term high resolution simulations.

Increased variability due to rounding errors can be beneficial for ensemble simulations and represent sub-grid-scale behaviour.

Example 1: Stochastic parametrisation schemes and rounding errors

- Stochastic parametrisation schemes use random noise to represent sub-grid-scale variability.
- We investigated a Burgers equation model that uses stochastic parametrisation schemes for turbulent closure (based on Dolaptchiev, Timofeyev and Achatz).

Example 1: Stochastic parametrisation schemes and rounding errors

- Stochastic parametrisation schemes use random noise to represent sub-grid-scale variability.
- We investigated a Burgers equation model that uses stochastic parametrisation schemes for turbulent closure (based on Dolaptchiev, Timofeyev and Achatz).

We replaced the random noise term of the stochastic parametrisation schemes by engineered rounding errors.
$\rightarrow$ Model quality is comparable.

Example 1: Stochastic parametrisation schemes and rounding errors

- Stochastic parametrisation schemes use random noise to represent sub-grid-scale variability.
- We investigated a Burgers equation model that uses stochastic parametrisation schemes for turbulent closure (based on Dolaptchiev, Timofeyev and Achatz).

We replaced the random noise term of the stochastic parametrisation schemes by engineered rounding errors.
$\rightarrow$ Model quality is comparable.

Rounding errors can represent sub-grid-scale variability.

Example 1: Stochastic parametrisation schemes and rounding errors

- Stochastic parametrisation schemes use random noise to represent sub-grid-scale variability.
- We investigated a Burgers equation model that uses stochastic parametrisation schemes for turbulent closure (based on Dolaptchiev, Timofeyev and Achatz).

We replaced the random noise term of the stochastic parametrisation schemes by engineered rounding errors.
$\rightarrow$ Model quality is comparable.

Rounding errors can represent sub-grid-scale variability.

We could base ensemble simulations on rounding error forcings.

Düben and Dolaptchiev TCFD 2015

Example 2: Parameter uncertainty in superparametrisation

Parameter	Precision	64 bits	Reduced prec.	Error
specific heat	7	1004.0	1004.0	0%
grav. acceleration	5	9.81	9.75	0.6%
latent heat of condensation	4	2.5104e+06	2.490368e+06	0.8%
latent heat of fusion	0	3.336e+05	0.262144e+05	21%
latent heat of sublimation	3	2.8440e+06	2.883584e+06	1.4%
gas constant	0	461.0	512.0	11%
diffusivity water vapour	1	2.21e-05	2.2888184e-05	3.6%
thermal conductivity	1	2.40e-02	2.34375e-02	2.3%
dynamic viscosity	0	1.717e-05	1.5258789E-5	11%
...	...	...	...	...

Example 2: Parameter uncertainty in superparametrisation

Parameter	Precision	64 bits	Reduced prec.	Error
specific heat	7	1004.0	1004.0	0%
grav. acceleration	5	9.81	9.75	0.6%
latent heat of condensation	4	2.5104e+06	2.490368e+06	0.8%
latent heat of fusion	0	3.336e+05	0.262144e+05	21%
latent heat of sublimation	3	2.8440e+06	2.883584e+06	1.4%
gas constant	0	461.0	512.0	11%
diffusivity water vapour	1	2.21e-05	2.2888184e-05	3.6%
thermal conductivity	1	2.40e-02	2.34375e-02	2.3%
dynamic viscosity	0	1.717e-05	1.5258789E-5	11%
...	...	...	...	...

The minimal level of precision provides plenty of information on model uncertainty.

Conclusions

- Double precision as default is overcautious.
- A reduction in precision will allow significant savings.
- Savings can be reinvested to allow higher resolution or more ensemble members to improve forecasts.
- A precision analysis can help to improve models.

References

PD Düben, J Joven, A Lingamneni, H McNamara, G De Micheli, KV Palem, TN Palmer, Phil. Trans. A, 2014

PD Düben, TN Palmer, H McNamara, JCP, 2014

TN Palmer, PD Düben, H McNamara, Phil. Trans. A, 2014

PD Düben, TN Palmer, Mon. Weath. Rev., 2014

PD Düben, J Schlachter, Parishkrati, S Yenugula, J Augustine, C Enz, K Palem and TN Palmer, DATE, 2015

F Russell, PD Düben, X Niu, W Luk, TN Palmer, FCCM, 2015

PD Düben, S Jeffress, TN Palmer, EMIT, 2015

PD Düben, SI Dolaptchiev, TCFD, 2015

PD Düben, F Russel, X Niu, W Luk, TN Palmer, JAMES, 2015