Agenda: GPU Progress and Directions for CAE

- Introduction of GPUs in HPC
- Progress of CFD on GPUs
- Review of OpenFOAM on GPUs
- Discussion on WRF Developments

Note: This is not CAE, this is for atmospheric modeling
## Collaborations and Primary Application Focus

<table>
<thead>
<tr>
<th>Model</th>
<th>Domain</th>
<th>GPU Approach</th>
<th>Collaborators</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRF</td>
<td>NWP/Climate</td>
<td>OpenACC (OACC)</td>
<td>NCAR, Cray, NVIDIA (Carl)</td>
</tr>
<tr>
<td>COSMO</td>
<td>NWP/Climate</td>
<td>CUDA C, OACC</td>
<td>CSCS, SCS, NVIDIA (Peter)</td>
</tr>
<tr>
<td>CAM-SE/HOMME</td>
<td>Climate</td>
<td>CUDA Ftn, OACC (future)</td>
<td>ORNL, Cray, NVIDIA (Paulius, Jeff)</td>
</tr>
<tr>
<td>NIM/FIM</td>
<td>NWP/Climate</td>
<td>Dirs, OACC</td>
<td>NOAA, NVIDIA (Carl, Paulius)</td>
</tr>
<tr>
<td>GEOS-5</td>
<td>Climate</td>
<td>CUDA Ftn, PGIDirs</td>
<td>NASA, NVIDIA (Carl)</td>
</tr>
<tr>
<td>NEMO</td>
<td>Ocean Model</td>
<td>OACC</td>
<td>NVIDIA (Maxim)</td>
</tr>
</tbody>
</table>
GPU Considerations for CWO Modeling

Initial efforts are mostly dynamical core developments
- If dynamics ~50% of profile time – 2x overall speed-up is possible
- Implicit schemes – iterative sparse matrix linear algebra solvers
- Explicit schemes – no linear algebra, operations on i,j,k stencil

Increasing use of GPU-based libraries and directives
- Examples: SpMV for implicit; stencil libraries, OpenACC directives

Most models use a domain decomposition parallel method
- Fits GPU model very well and preserves costly MPI investment

Fortran programming on GPUs most critical for adoption
- NVIDIA investments in CAPS, PGI and Cray compilers; OpenACC
NVIDIA HPC Strategy for CWO Applications

- Industry leader in HPC development environment – emphasis on Fortran
  - Technical and business alliances with OpenACC members: Cray, PGI, CAPS
  - Demonstration of OpenACC success with WRF development (joint with Cray)

- Invest in technical collaborations toward 2013 proof of operational NWP
  - Current focus is NOAA with FIM/NIM and CSCS with COSMO/OPCODE

- Ensure Titan success with ORNL climate (CAM-SE) and NOAA applications
  - ORNL funded 14,000+ GPUs and NOAA funded 4000+ GPUs of the NVIDIA Kepler GPU’s in Titan and put 2 cabinets of NVIDIA Kepler GPU’s into their Gaea system ($8M total)

- Maintain presence in user community with industry event participation

- Expand GPU use to more applications – especially next-gen models
  - Next-gen Examples: UKMet GungHo, NCAR MPAS-A, LANL MPAS-O, NOAA MOMn
Important CWO Talks at NVIDIA GTC 2013

S3429 - Running the FIM and NIM Weather Models on GPUs
Mark Govett - Chief, Advanced Computing Section Global Systems Division, NOAA ESRL

S3417 - Towards GPU-accelerated Operational Weather Forecasting
Oliver Fuhrer - Senior Scientist, MeteoSwiss

S3326 - Hybrid Fortran 90: High Performance, Low Friction GPGPU for Weather Prediction
Michel Muller - Technical Staff, RIKEN

S3209 - Accelerating NEMO with OpenACC
Maxim Milakov - HPC Developer Technology Engineer, NVIDIA
NVIDIA Participation in Key Industry Events**

**APR:** Invited GPU tutorial at NCAR/UCAR Software Engineering Assembly (SEA) 2013: [http://sea.ucar.edu/conference/2013](http://sea.ucar.edu/conference/2013)

**JUN:** Plans for talk on WRF developments with OpenACC at WRF Users Workshop 2013: [http://www.mmm.ucar.edu/wrf/users/wrfda/workshop.html](http://www.mmm.ucar.edu/wrf/users/wrfda/workshop.html)

**SEP:** Accepted talk(s) at the NCAR/NOAA Heterogeneous Multicore Workshop 2013: [http://data1.gfdl.noaa.gov/multi-core/](http://data1.gfdl.noaa.gov/multi-core/)

**SEP:** Invited talk at NCAR CISL’s biennial CAS2K13 in Annecy France, 2013: [http://www2.cisl.ucar.edu/section/events/cas2k13?page=4](http://www2.cisl.ucar.edu/section/events/cas2k13?page=4)

**CWO talks on GPU progress are given by customers throughout 2013 at various conferences**
GPU Progress Reported at NCAR Workshop

Programming weather, climate, and earth-system models on heterogeneous multi-core platforms

September 7-8, 2011 at the National Center for Atmospheric Research in Boulder, Colorado

GPU related talks (11+) that cover application software such as:

NIM | WRF | GEOS-5 | HOMME | COSMO | CAM-SE | ICON

- Successes and Challenges using GPUs for Weather and Climate Models
  Mark Govett, NOAA
- Experience using FORTRAN GPU Compilers with the NIM
  Tom Henderson, NOAA
- GPU Acceleration of the RRTM in WRF using CUDA FORTRAN
  Greg Ruetsch, NVIDIA
- Lessons Learned adapting GEOS-5 GCM Physics to CUDA FORTRAN
  Matt Thompson, NASA
- Accelerated Cloud Resolving Model in Hybrid CPU-GPU Clusters
  Jose Garcia, NCAR
- Reworking Boundary Exchanges in HOMME for Many-Core Nodes
  Ilene Carpenter, NREL
- Performance optimizations for running an NWP model on GPUs
  Jacques Middlecoff, NOAA
- Rewrite of the COSMO Dynamical Core
  Mueller / Gysi, SCS/CSCS
- Experiences with the Finite-Volume Dynamical core and GEOS-5 on GPUs
  Bill Putman, NASA
- Progress in Accelerating CAM-SE
  Jeff Larkin, Cray/ORNL
- Porting the ICON Non-hydrostatic Dynamical Solver to GPUs
  Will Sawyer, CSCS

Source: http://data1.gfdl.noaa.gov/multi-core/
<table>
<thead>
<tr>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM-SE</td>
<td>CAM-SE</td>
</tr>
<tr>
<td>COSMO</td>
<td>COSMO</td>
</tr>
<tr>
<td>GEOS-5</td>
<td>GEOS-5</td>
</tr>
<tr>
<td>HOMME</td>
<td></td>
</tr>
<tr>
<td>ICON</td>
<td></td>
</tr>
<tr>
<td>NIM</td>
<td>NIM</td>
</tr>
<tr>
<td>gNEMO</td>
<td></td>
</tr>
<tr>
<td>WRF</td>
<td>WRF</td>
</tr>
<tr>
<td></td>
<td>RRTM</td>
</tr>
<tr>
<td></td>
<td>KernelGen</td>
</tr>
<tr>
<td></td>
<td>OpenACC</td>
</tr>
</tbody>
</table>

GPU Progress Reported at This Workshop

<table>
<thead>
<tr>
<th>2011</th>
<th>2012</th>
<th>Other GPU Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM-SE</td>
<td>CAM-SE</td>
<td>Model</td>
</tr>
<tr>
<td>COSMO</td>
<td>COSMO</td>
<td>Organization(s)</td>
</tr>
<tr>
<td>GEOS-5</td>
<td>GEOS-5</td>
<td>ASUCA</td>
</tr>
<tr>
<td>HOMME</td>
<td></td>
<td>JMA / TiTech</td>
</tr>
<tr>
<td>ICON</td>
<td>NIM</td>
<td>CFSv2/GFS</td>
</tr>
<tr>
<td>NIM</td>
<td>NIM</td>
<td>C-DAC (India)</td>
</tr>
<tr>
<td>gNEMO</td>
<td></td>
<td>GALES</td>
</tr>
<tr>
<td>WRF</td>
<td>WRF</td>
<td>KNMI</td>
</tr>
<tr>
<td>- RRTM</td>
<td>- KernelGen</td>
<td>GRAPES</td>
</tr>
<tr>
<td></td>
<td>- OpenACC</td>
<td>CMA (China)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HARMONIE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ICHEC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HIRLAM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KNMI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NEMO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NVIDIA</td>
</tr>
</tbody>
</table>

GPU Status of WRF Developments

Several non-trunk efforts at various stages:
- Dynamics and some physics by Thomas Nipen at UBC – source at NVIDIA
- KernelGen project: [www.kernelgen.org](http://www.kernelgen.org) update at NCAR 2012 workshop
- Cray and OpenACC (Pete Johnsen) with results at 2012 NCAR workshop
- C-DAC and HPC-FTE group working with NVIDIA India (Priyanka)
- Shortwave radiation model by NV software group and PGI (G. Ruetsch)
- NIM to include WRF physics using PGI and/or HMPP, OpenACC
- Several physics schemes by Space Science and Engineering Center, WI, USA

Trunk efforts at various stages:
- WSM5 physics model (15% - 25%) in release 3.2 from 2009
- WRF 3.5 with OpenACC – NVIDIA and NCAR (MMM – Dave Gill) collaboration
# CWO Models and GPU Progress

<table>
<thead>
<tr>
<th>Climate</th>
<th>Weather</th>
<th>Ocean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEOS-5 (US)</td>
<td>WRF (US)</td>
<td>MITgcm (US)</td>
</tr>
<tr>
<td>CESM (US)</td>
<td>FIM/NIM (US)</td>
<td>HYCOM (US)</td>
</tr>
<tr>
<td>- CAM-SE (HOMME)</td>
<td>COAMPS (US)</td>
<td>ROMS (US)</td>
</tr>
<tr>
<td>- POP (US)</td>
<td>COSMO (EU)</td>
<td>MOM4 (US)</td>
</tr>
<tr>
<td>CFSv2 (US)</td>
<td>ICON (DE)</td>
<td>GOLD (US)</td>
</tr>
<tr>
<td>- GFS</td>
<td>AROME (FR)</td>
<td>POP (US)</td>
</tr>
<tr>
<td>- MOM4</td>
<td>HARMONIE (EU)</td>
<td>NEMO (EU)</td>
</tr>
<tr>
<td>ASUCA (JP)</td>
<td>- Hirlam + Aladin</td>
<td></td>
</tr>
<tr>
<td>GRAPES (CN)</td>
<td></td>
<td>NVIDIA Accelerated</td>
</tr>
<tr>
<td>OLAM (BR)</td>
<td></td>
<td>Tesla in Process</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tesla Not Started</td>
</tr>
</tbody>
</table>
Next-Gen CWO Models and GPU Status

**Climate**
- GEOS-5 (US) → GEOS-6 (US)
- CESM (US)
  - CAM-SE (HOMME)
  - POP (US) → MPAS-O
- CFSv2 (US)
  - GFS
  - MOM4 (US) → MOMn

**Weather**
- WRF (US) → MPAS-A or NIM
- ICON (DE)
  - UM (UK) → GungHo
- IFS (UK) → OOPS

**Ocean**
- MOM4 (US) → MOMn
- POP (US) → MPAS-O

**GPU Status**
- NVIDIA Accelerated
- Port to Tesla in Process
- Port not yet started
UBC Developments of WRF

- Dynamics and some physics by Thomas Nipen at UBC, with John Michalakes ~2010

Subset of WRF model:
- Various dynamics components
- Microphysics (Kessler)
- Shortwave radiation (Dudhia)

Simulation on:
- 189 x 150 x 27 domain
- CPU: 2.4GHz Opteron*
- GPU: GeForce 9800 GX2
  *using one core
OpenACC Developments for WRF 3.4/3.5

WRF Experiments on GPU Accelerators using OpenACC
- Pete Johnsen, Cray, Inc.

WRF routine `advance_w`

- Dynamics routine to advance vertical velocity
- Standard Fortran use with OpenACC directives
- 2.1x speedup for 16 cores

Thank You and Questions?

Stan Posey
NVIDIA, Santa Clara, CA, USA; sPosey@nvidia.com