Looking towards the future – NCAR Computing, Storage and our Models and Workflows

Anke Kamrath
Director, Computational and Information Systems Laboratory

Software Engineering Assembly
April 2, 2018

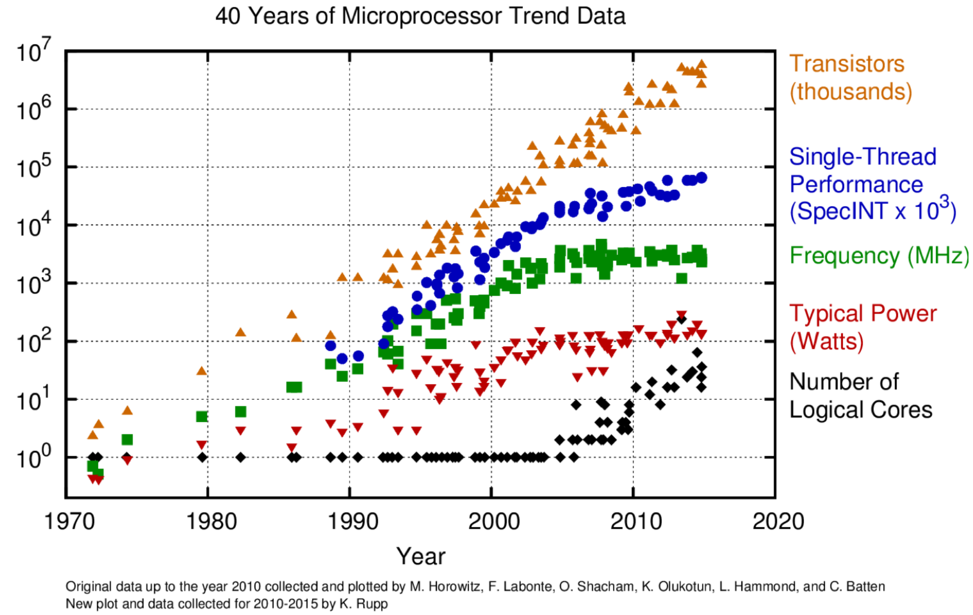
Challenging Times Ahead –
The end of modeling as we know it?

- **Atmospheric & rel. science relies on modeling complex systems with multi-scale, multi-physics**
- **Computers have big problems supporting our current approach (next slide)**
- **Is our future ability to advance atmospheric science in jeopardy?**

Computing Challenges

- **Performance:**
  - Individual cores are not getting faster - integration rate is stuck
- **Power:**
  - moving data on/off chip takes lots of energy
- **Complexity:**
  - increasing complexity both software, in human-crafted algorithms and in hardware architecture
- **Data:**
  - the volumes of data generated by climate and weather codes don’t play well with current storage technologies, and system software and analysis practices are not keeping up

Computing Wall

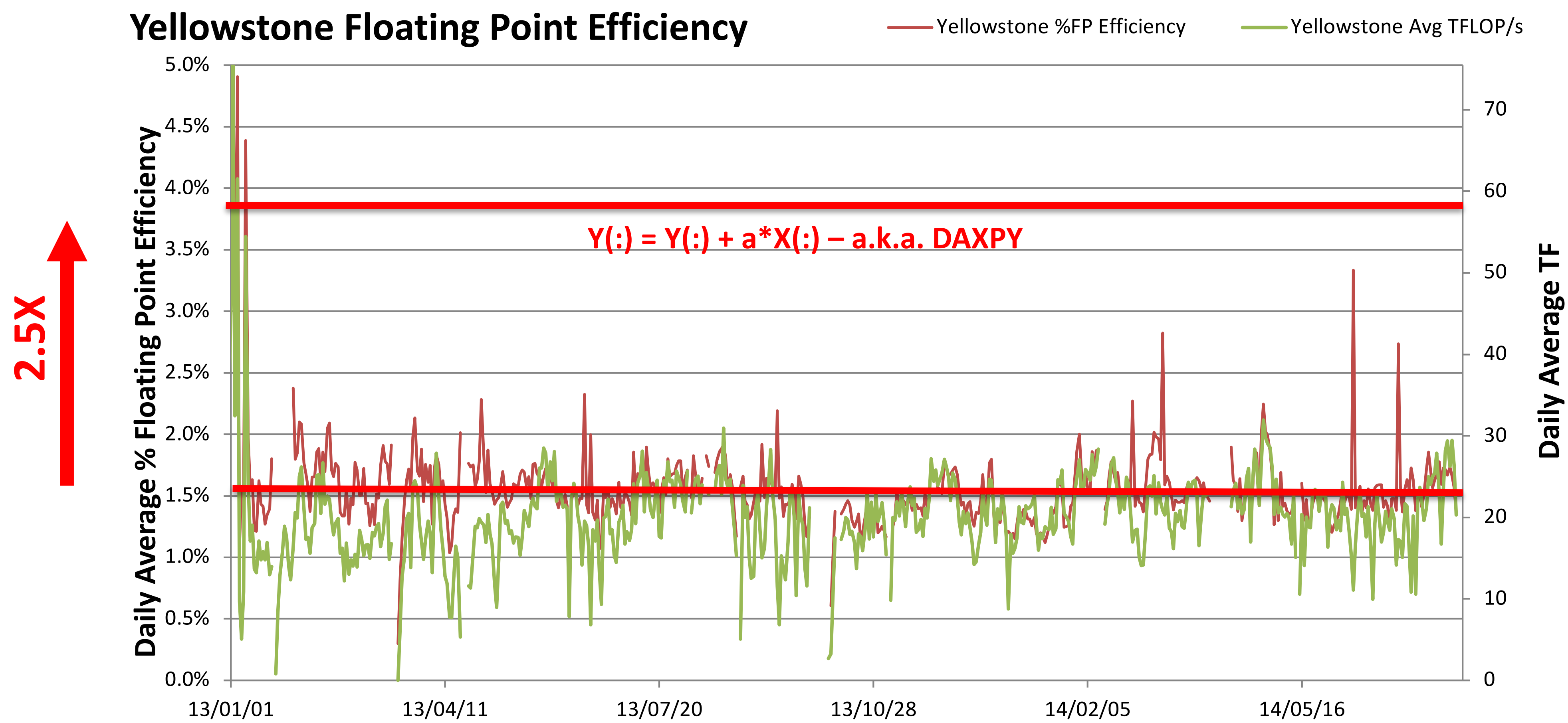


Source: Karl Rupp

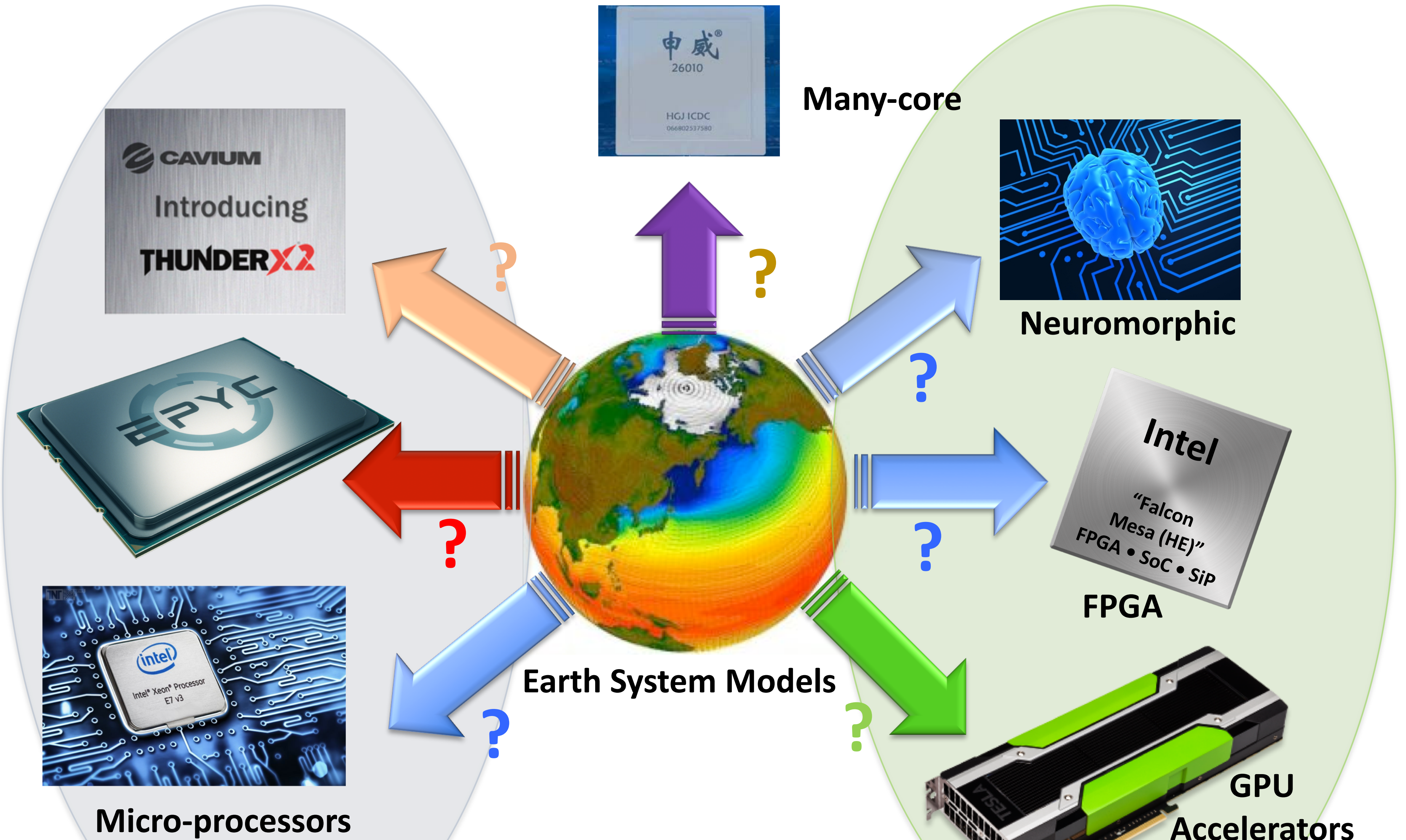
- **Processor trends**
  - More transistors
  - More cores (∝ transistors)
  - Flat clock speeds and power
  - Slowing thread performance
  - Increasing flops/byte of memory BW
    - SunWei processors ~25 flops/byte
    - KNL processors ~7 flops/byte
- **Climate modeling not well matched to trends**
  - Typically < 1 flop/byte
  - Climate applications are state heavy with low computational intensity.
  - Physics code is branchy, hard to vectorize, has divides and load imbalances

The Vectorization Gap

Yellowstone: Sustained fraction of FP peak is 1.57%

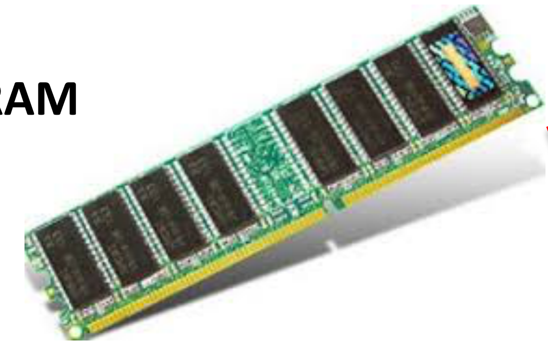


Increasing Computing Complexity



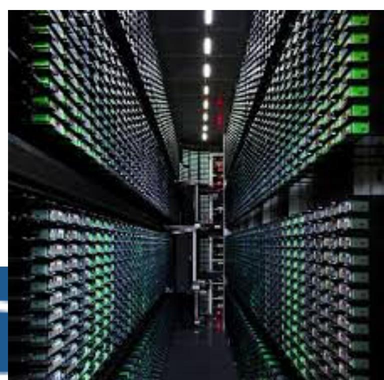
Increasing Storage Complexity

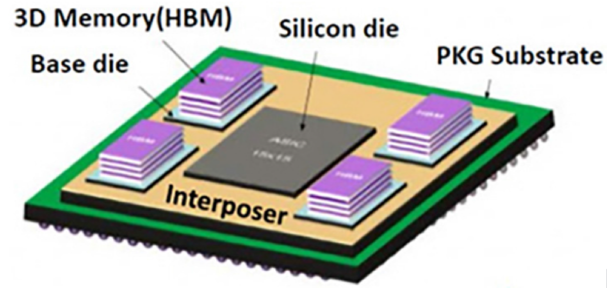
DRAM



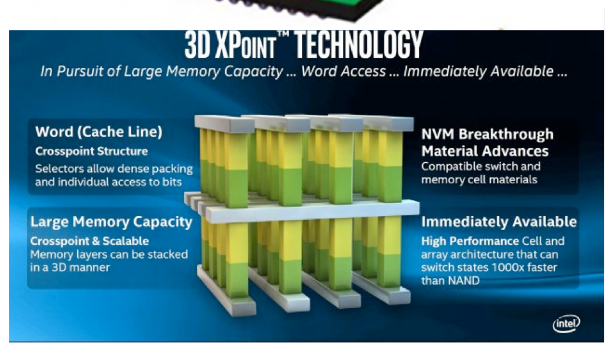
DISK

TAPE

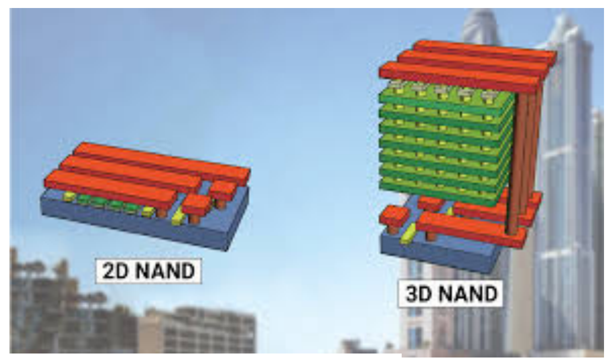




Stacked memory:
Fast, hot & small



Memory-class storage



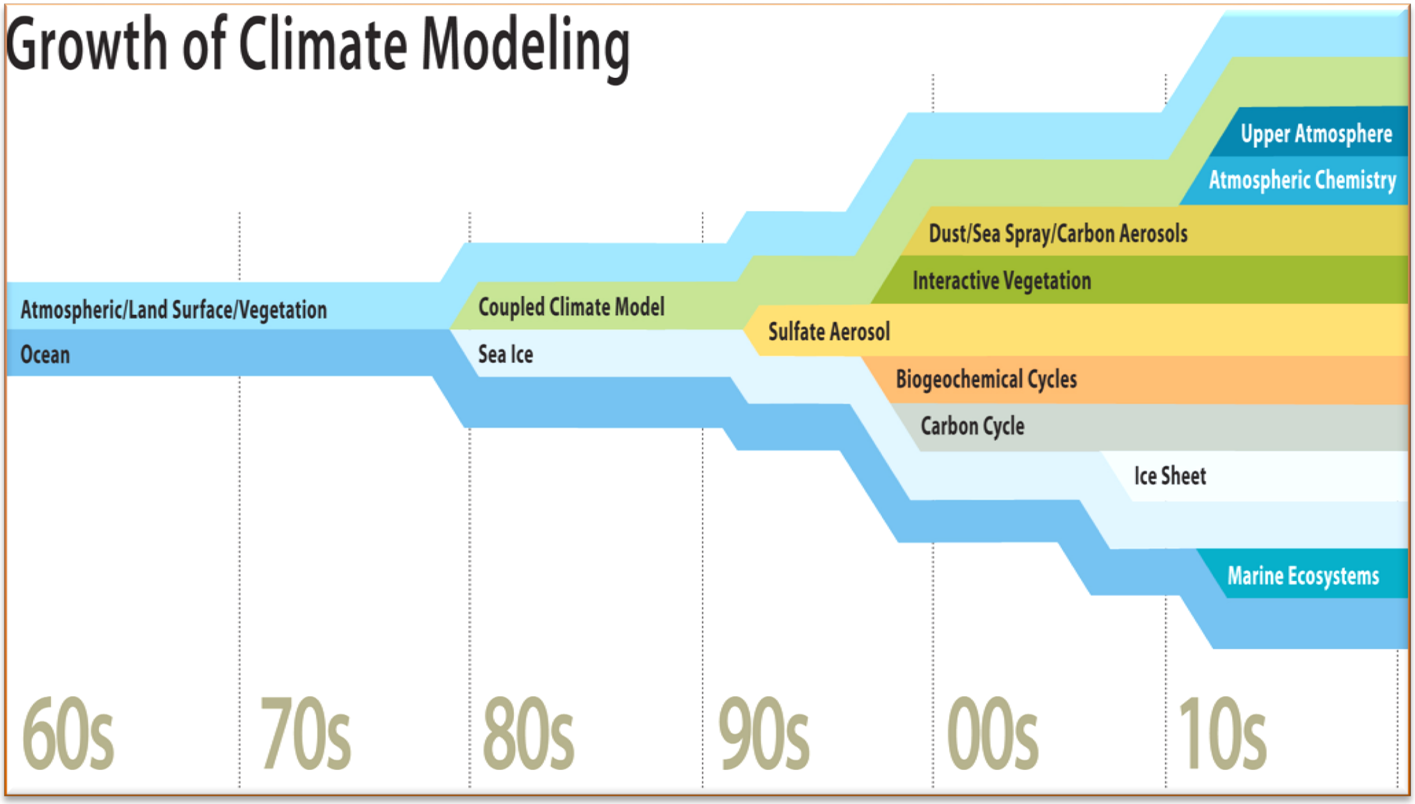
Storage-class memory

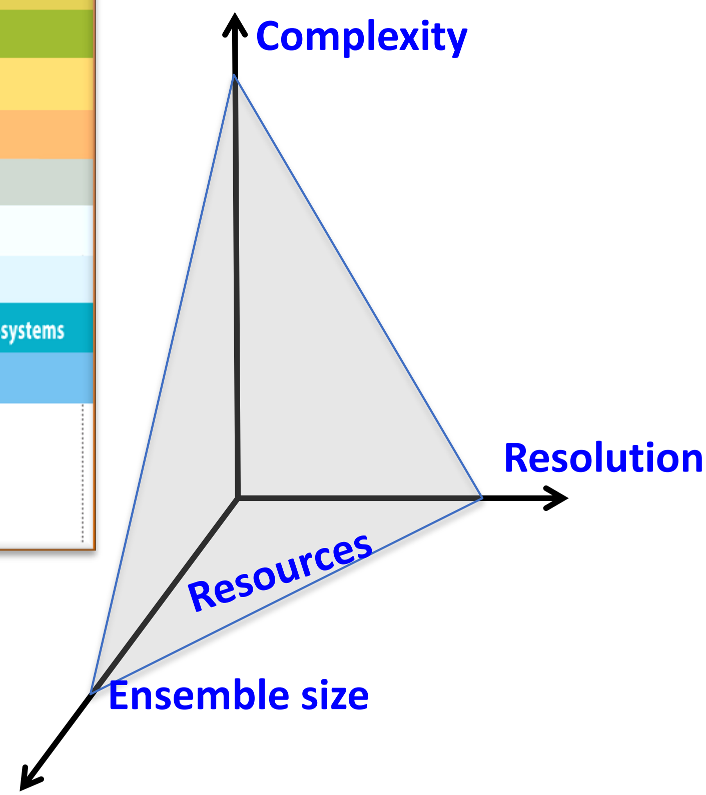
Cloud-base
object store
public or private

S3

ceph

Increasing Model Complexity





"Lines of Earth System model code increased 10x over the period 1993-2008", Easterbrook and Johns., 2009

What is the "limit" of model complexity?

- **Modularity, separation of concerns, etc. are strategies developed to manage software complexity. Is it enough?**
- **Possible sources of limits:**
  - Limits in human ability to engineer complex systems
  - Computational cost becomes too high /integration rate becomes too low.
  - Space of tunable parameters grows too large, or the region of stable parameter choices become infinitesimal.
  - Code becomes untestable/unverifiable.
  - Cost ($) of development/maintenance becomes too high.
- **Which is a real issue, and which an unfounded fear?**

"[software engineering] limits exist but are simply less obvious and more related to limitations in human abilities", Leveson, N.G.

Types of Software Complexity

- **Source lines of code**
  - Can be tricky to measure
  - Definition: executable lines;
- **Number of subprograms**
  - Indication of modularity
  - CESM1 – 9832
  - CESM2 – 10,570
- **Cyclomatic complexity**
  - Number of independent paths through the code
  - 350 routines (CESM1) have >51 paths
  - Impacts application "testability"

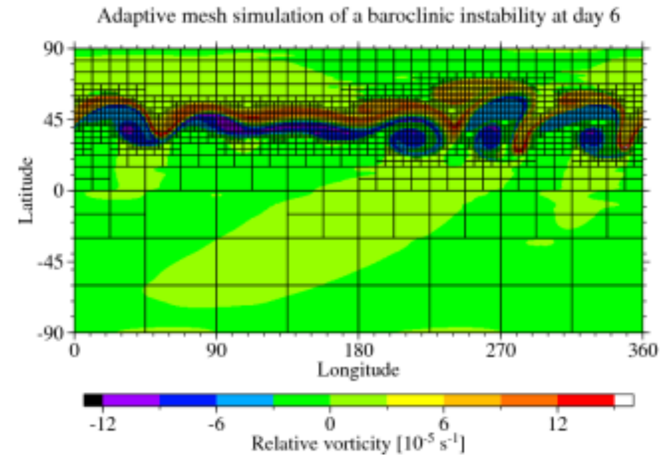
Table V. NUMBER OF SUBPROGRAMS IN EACH CYCLOMATIC COMPLEXITY RANGE

Model	0–10	11–20	21–50	>51
GISS	62	26	34	21
CSIRO-Mk3.6.0	116	63	65	55
INM-CM4	459	135	93	52
GFDL-CM2.1	1442	249	212	109
HadGEM2	1003	361	383	285
HadGEM3	1318	421	453	274
CCSM3	2053	335	289	63
CMCC-CESM	1806	438	381	197
GEOS-5	2301	427	308	135
IPSL	2573	375	284	129
MPI-ESM-LR	2191	531	454	234
CFSv2-2011	2397	511	603	263
BCC-CSM1.1	2705	527	422	130
ModelE	3026	459	312	147
CCSM4	4682	803	701	238
CESM1	7312	1223	947	350

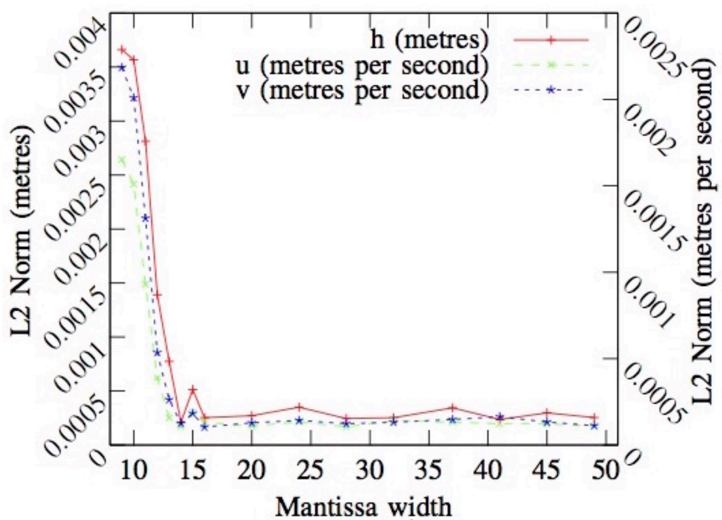
From: Mendez, et al. 2014

Algorithmic ways to deal with numerical complexity

- **Bigger timesteps**
  - Implicit integration
  - Parallel in time (PinT) methods
- **Fewer points**
  - More accurate numerics
  - Adaptive mesh refinement
- **Physics emulators**
  - Neural networks
  - Stochastic forcing
- **Reduced precision**
  - FPGAs



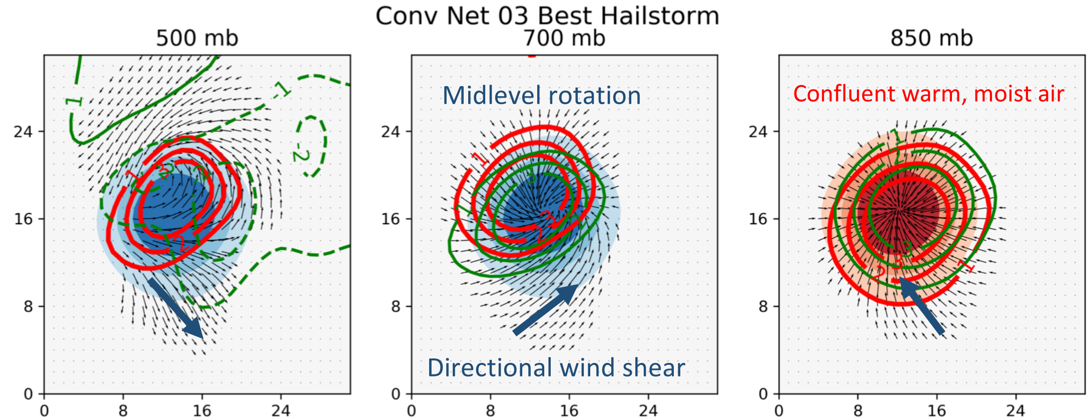
C/o: Paul Ulrich



Targett, et al. 2015

Tackling Model Complexity through Machine Learning (ML)

New ML Team - AnIMaL (Analytics and Integrative ML) in CISL/NCAR



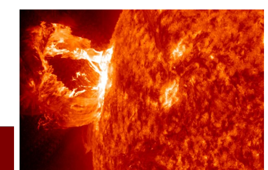
Red: Temperature
Green: Dewpoint
Filled: Height

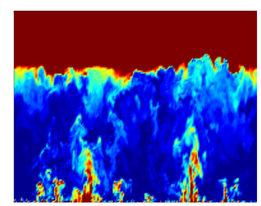
Deep learning models identify physically relevant features for hailstorm prediction from core weather fields. Running the network in reverse reveals these features.

What is a neural net's "dream" hailstorm?

Why machine-learned emulation? Replacing human-crafted parameterizations with machine learning algorithms may simplify, accelerate and improve models.

- Cloud Microphysics emulation (Dr. Andrew Gettelman, CGD)
    - improved weather and climate modeling
- Interplanetary Coronal Mass Ejection (CME) (Dr. Sarah Gibson, HAO)
    - space weather prediction
- Sub-grid-scale Turbulence (Dr. Sue Haupt, RAL)
    - Application: improved meteorological models

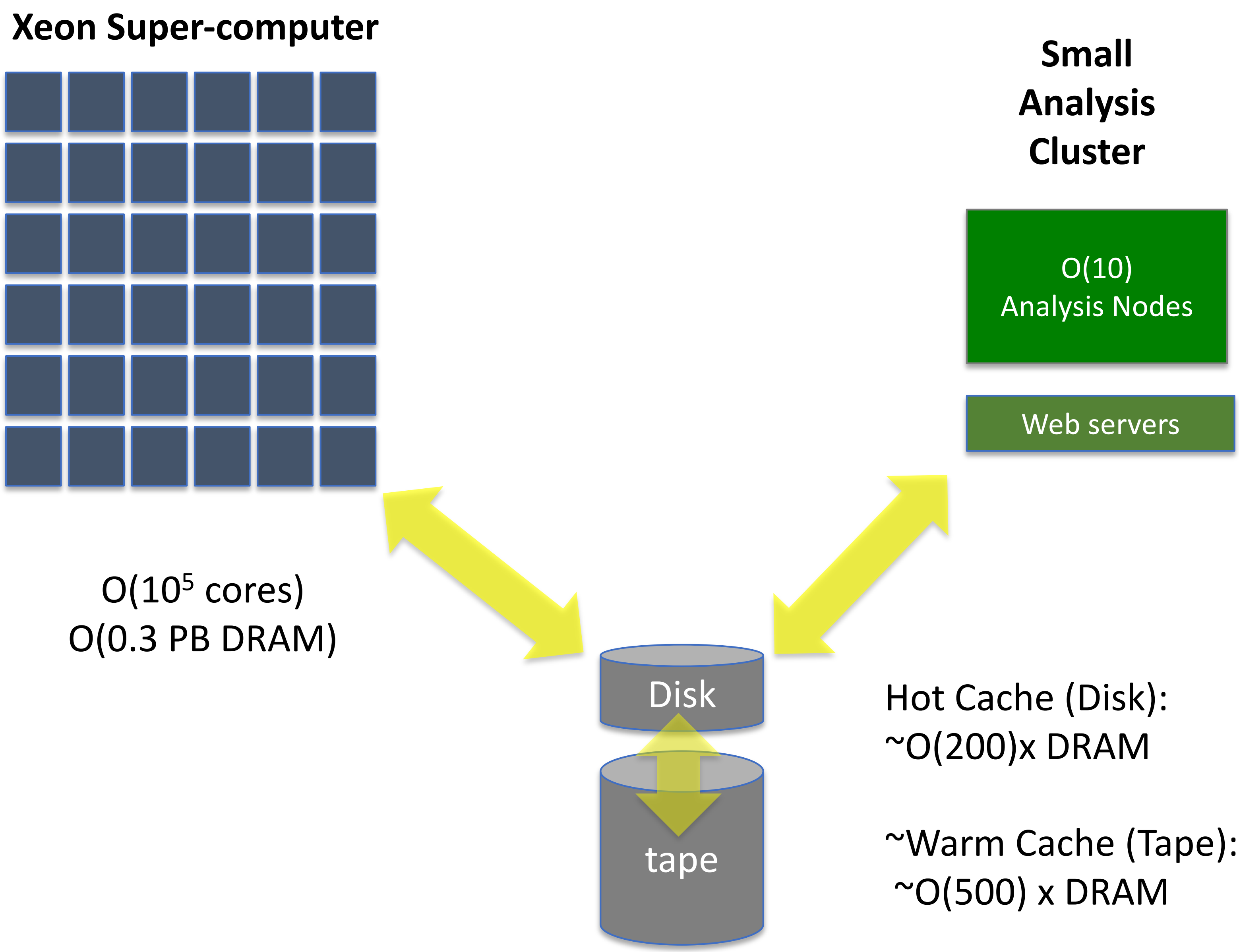




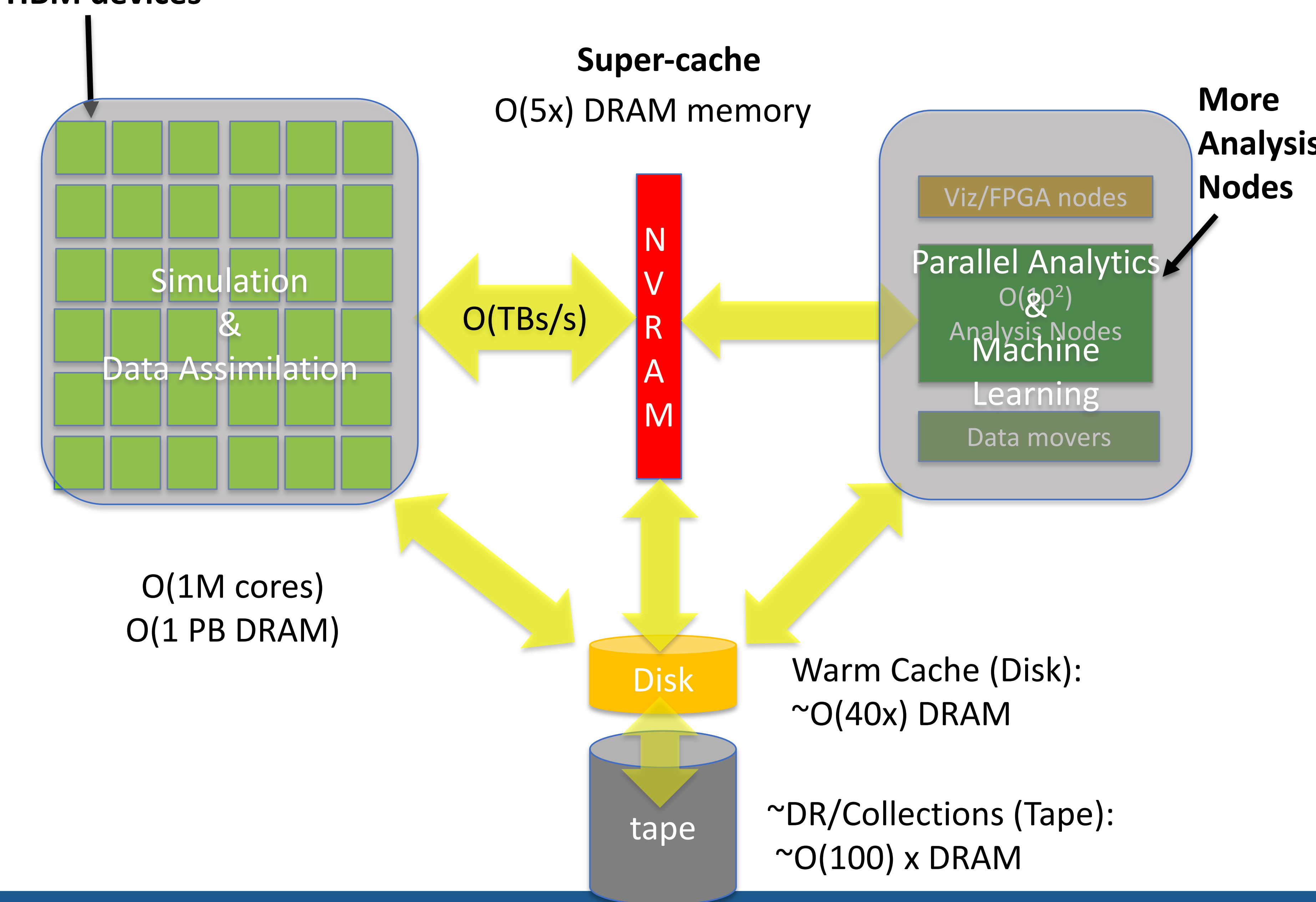
NCAR’s Next Super ”NWSC-3” (2021)

- **Focus on a design that:**
  - Enhances the end-to-end rate of science throughput
  - Reduces costs and/or enhance reliability
- **Harness emerging technologies:**
  - Accelerators (GPUs)
  - New memory technologies (stacked, NV memory)
  - Machine learning techniques (ML/DL)
- **Prepare application/workflow codes:**
  - scalability and performance
  - Performance-portability

Existing System Architecture

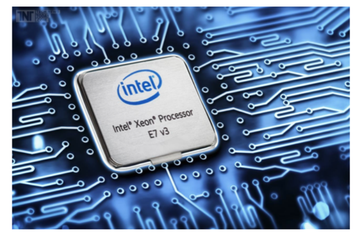


Next-Gen (NWSC-3) Architecture



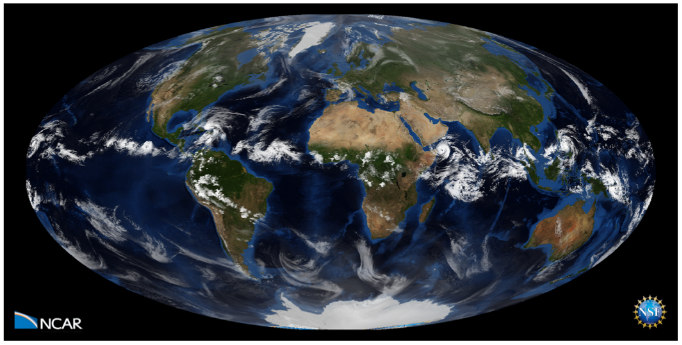
CPU-GPU Performance Portability

- **Strategy - Single Source Programming Model**
  - CPU: OpenMP directives
  - GPU: OpenACC directives
- **Why?** *GPUs have substantially better energy efficiency and performance per device than CPUs*
- **Growing GPU-enabled Portfolio:**
  - MPAS Dynamics
  - MPAS Physics underway
  - MOM6 about to start
  - MURAM kicking off



- **Annual Multicore workshop** provides forum to better understand application of new HPC technologies for the next generation of weather, climate, and earth-system models
- **Active University partnerships (Wyoming, Delaware) to address workforce challenges**

Weather and Climate Alliance (WACA): Refactoring MPAS for CPU-GPU portability



High resolution global atmospheric models: too slow for long-range simulations

- **Goal:**
  - Single source performance portability across CPUs and GPUs using the OpenACC & OpenMP directive-based system.
- **Participants:**
  - NCAR (CISL and MMM)
  - NVIDIA Corporation and IBM Corporation/The Weather Company
  - University of Wyoming, CE&EE Department
  - Korean Institute of Science and Technology Information (KISTI)
- **MPAS 64-bit dynamical core results (40,962 grid-points)**
  - Comparison Intel Broadwell node (CPU) vs NVIDIA GP100 (GPU)
  - 8-15% better performance on CPU after porting to GPUs
  - **1 GPU = 2.8x CPU nodes**

GPU accelerated code

```
!$ACC KERNEL
!$ACC LOOP
do i=1,imax
    a(i)-a(i)+z
enddo
!$ACC END KERNELS
```

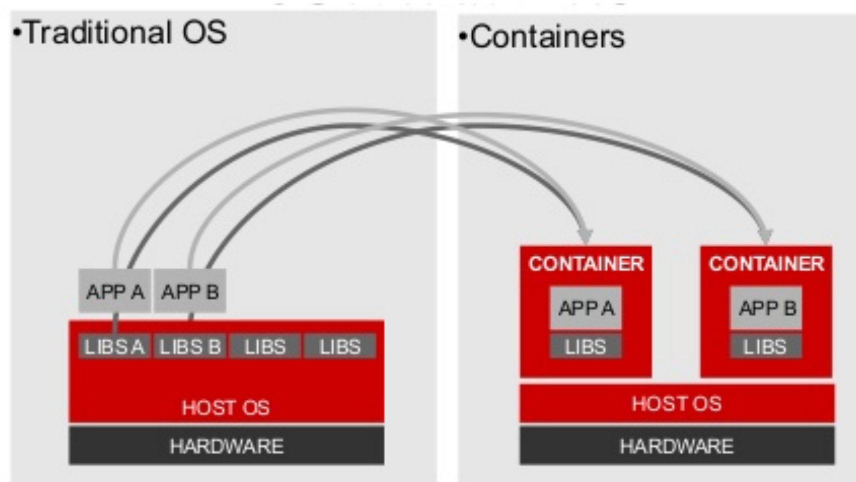
OpenACC directives avoid a total rewrite in CUDA.

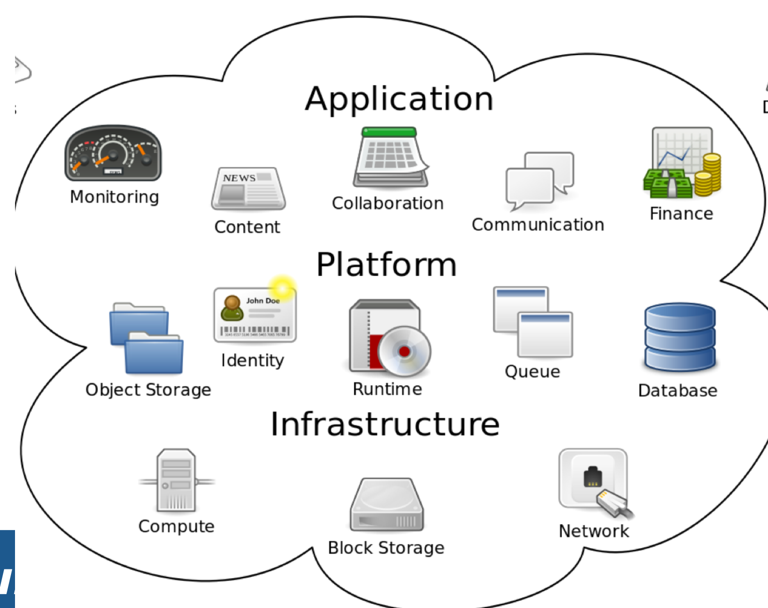
NCAR | ***Computing, Storage, and our Models and Workflows***

air • planet • *people*

Enabling and Supporting Cloud-based Solutions

- **HPC Cloud-based solutions emerging across NCAR**
  - AMPs Forecast on Penguin Cloud for High Availability Needs
  - CESM2 1º running on AWS EC2 at 10 SYPD with I/O
- **Environment "around apps" most valuable (i.e., Containerization)**
  - Working with SW Development teams for containerization (capturing complex environment needs)
- **Data Analysis and Storage**





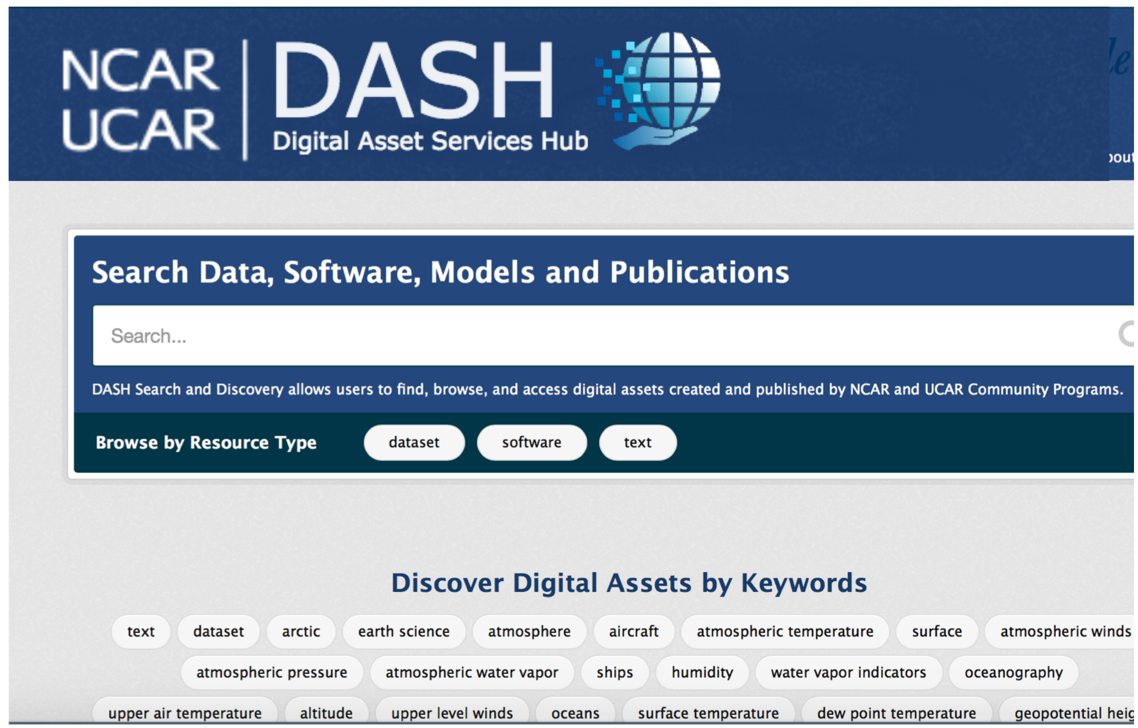
Cloud computing

Confronting the Data Challenges

- **Positioning NCAR for increasing Data Focus**
  - **DSET (Data Stewardship Engineering Team) –** Cross NCAR team to tackle NCAR-wide data challenges
  - **DASH -** Digital Asset Services Hub deployed
  - **CISL -** creating an "Information Systems Division" (Leader is being actively recruited)
- **Prioritization and Collaboration around Data Portfolio**
  - Creation of "DASH" – for better data discovery
  - CMIP Analysis Platform
  - Capstone
  - Pangeo: Open Source Big Data Climate Science Platform
  - NCAR's Research Data Archive (RDA)
  - Globus and Globus+
- **Architecting and integrating our data discovery, serving and analysis portfolio**

DSET and DASH

- Integrated front door to data discovery digital scientific assets (datasets and supporting metadata, publications, software applications, and model code) across NCAR
- Improve coordination, shared expertise, and data management standards across organization
- DSET identified 102 distinct NCAR digital assets http://data.ucar.edu
- Expanding DASH to support Public Access Mandate and NCAR PI Data Plans
- Hosting: "*Geoscience Digital Data Resource and Repository Service (GeoDaRRS) workshop*" (Aug 7-9, 2018)



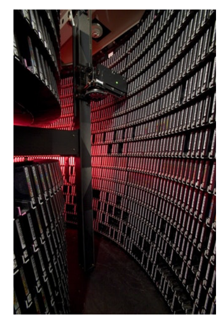
CMIP Analysis Platform

- **New Service - Lending Library for CMIP with DAV access**
  - Still early days for CMIP6. Good test experience with CMIP5 data
  - Up to 10 PB will be allocated
- **Working to meet community needs**
  - Reaching out to NSF Funded CMIP researchers
  - Providing a platform for developing analytics tools
  - Future - look to provide all the data on “the Cloud”?
- **More than 70 allocation requests for CMIP AP received to date**
  - More than 40 show some amount of DAV cluster use
  - And 45 user requests for data to be added to the lending library

Pangeo: Open Source Big Data Climate Science Platform

Scalable analytics solutions are required to work with large datasets

- **Goal:** create an open-source toolkit for the analysis of climate datasets, built on the Python language ecosystem, Xarray multi-dimensional array tools, and Dask parallel analytics system
- **Funding: $1.2M NSF-funded EarthCube project**
- **Pangeo Participants:**
  - Lamont-Doherty Earth Observatory
  - Columbia University's Data Science Institute
  - NCAR (Kevin Paul)
  - Anaconda

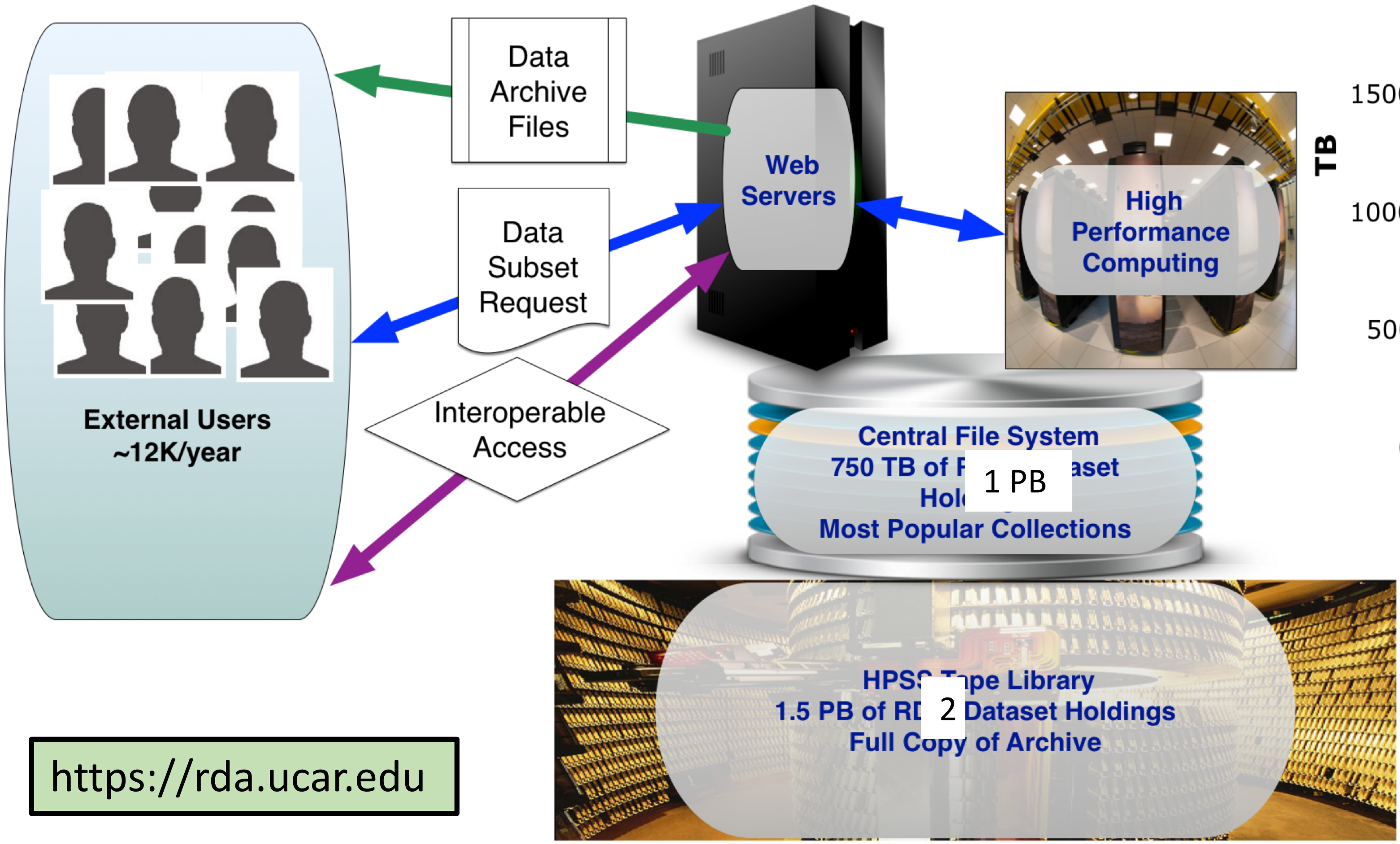


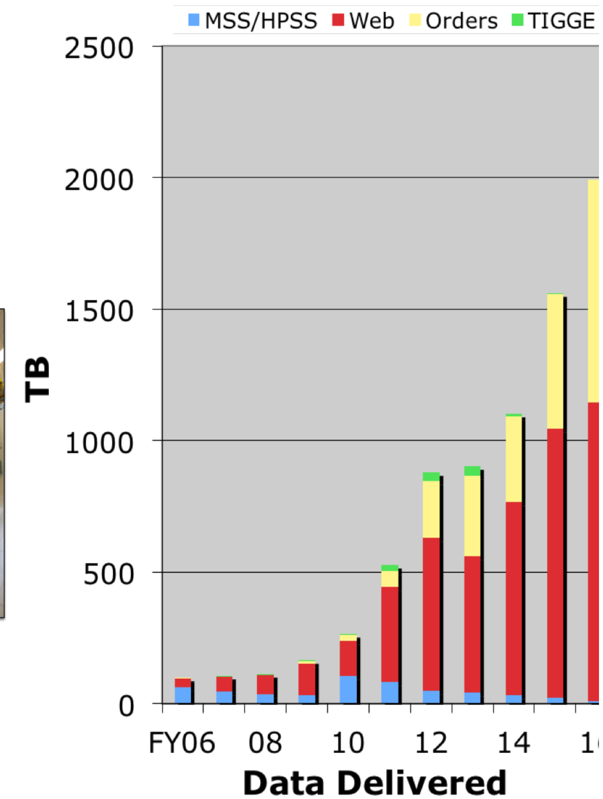
Parallelism is key: single device performance is falling behind!

RDA Data Processing

12,000+ Unique Web users in FY 2016

~2 PB Data Delivered in FY 2016





https://rda.ucar.edu

Data Movers, Sharing and Globus+

- **Data Transfer Protocols**
  - Globus
  - GridFTP
  - SCP/SFTP
- **Science Gateway Support**
  - RDA/ESG/CDP
- **Globus Features**
  - Reliable, Secure, High-performance file transfer
  - "Fire and Forget"
  - Automatic fault recovery
  - Powerful GUI, APIs and CLI
  - Integrated with RDA file lists
  - Integration with authentication systems

Developing an integrated Data Architecture for Discovery, Management and Analysis

g plus

CMIP AP

Pangeo

DASH Digital Asset Services Hub Repository

DASH Digital Asset Services Hub Discovery

NCL NCAR Command Language

Capstone

Research Data Archive

Data Sharing Environments

Tools

Analysis Environments

Data Serving & Management Environments

Data Discovery

- **Goals:**
  - Data portfolio needs more coordination/prioritization (do fewer things better ... headroom to tackle new initiatives)
  - Managing Legacy Needs
  - Move to Open Development
  - Cloud-enabled
  - Reusable components - Reduce/remove duplication

Thank you

Questions?

Complexity Metrics

Table V. NUMBER OF SUBPROGRAMS IN EACH CYCLOMATIC COMPLEXITY RANGE

Model	0–10	11–20	21–50	>51
GISS	62	26	34	21
CSIRO-Mk3.6.0	116	63	65	55
INM-CM4	459	135	93	52
GFDL-CM2.1	1442	249	212	109
HadGEM2	1003	361	383	285
HadGEM3	1318	421	453	274
CCSM3	2053	335	289	63
CMCC-CESM	1806	438	381	197
GEOS-5	2301	427	308	135
IPSL	2573	375	284	129
MPI-ESM-LR	2191	531	454	234
CFSv2-2011	2397	511	603	263
BCC-CSM1.1	2705	527	422	130
ModelE	3026	459	312	147
CCSM4	4682	803	701	238
CESM1	7312	1223	947	350

Table III. NUMBER OF SUBPROGRAMS

Model	Subpr.	Func.	In Mod.	Subr.	In Mod.
GISS	143	18	0	125	0
CSIRO-Mk3.6.0	299	3	0	296	0
INM-CM4	739	42	0	697	45
GFDL-CM2.1	2012	331	329	1681	1670
HadGEM2	2032	89	14	1943	319
HadGEM3	2566	222	184	2344	1219
CCSM3	2740	327	320	2414	1950
CMCC-CESM	2822	524	451	2298	1360
GEOS-5	3171	721	487	2450	1645
IPSL	3361	441	413	2920	2222
MPI-ESM-LR	3410	453	393	2957	2198
CFSv2-2011	3774	1113	818	2661	1248
BCC-CSM1.1	3784	802	781	2982	2090
ModelE	3944	619	474	3325	1697
CCSM4	6424	1150	1117	5274	4649
CESM1	9832	1852	1809	7980	7516
total	51053	8707	7590	42534	29828

From: Mendez, et al. 2014