High Performance Computing and Modeling Infrastructure

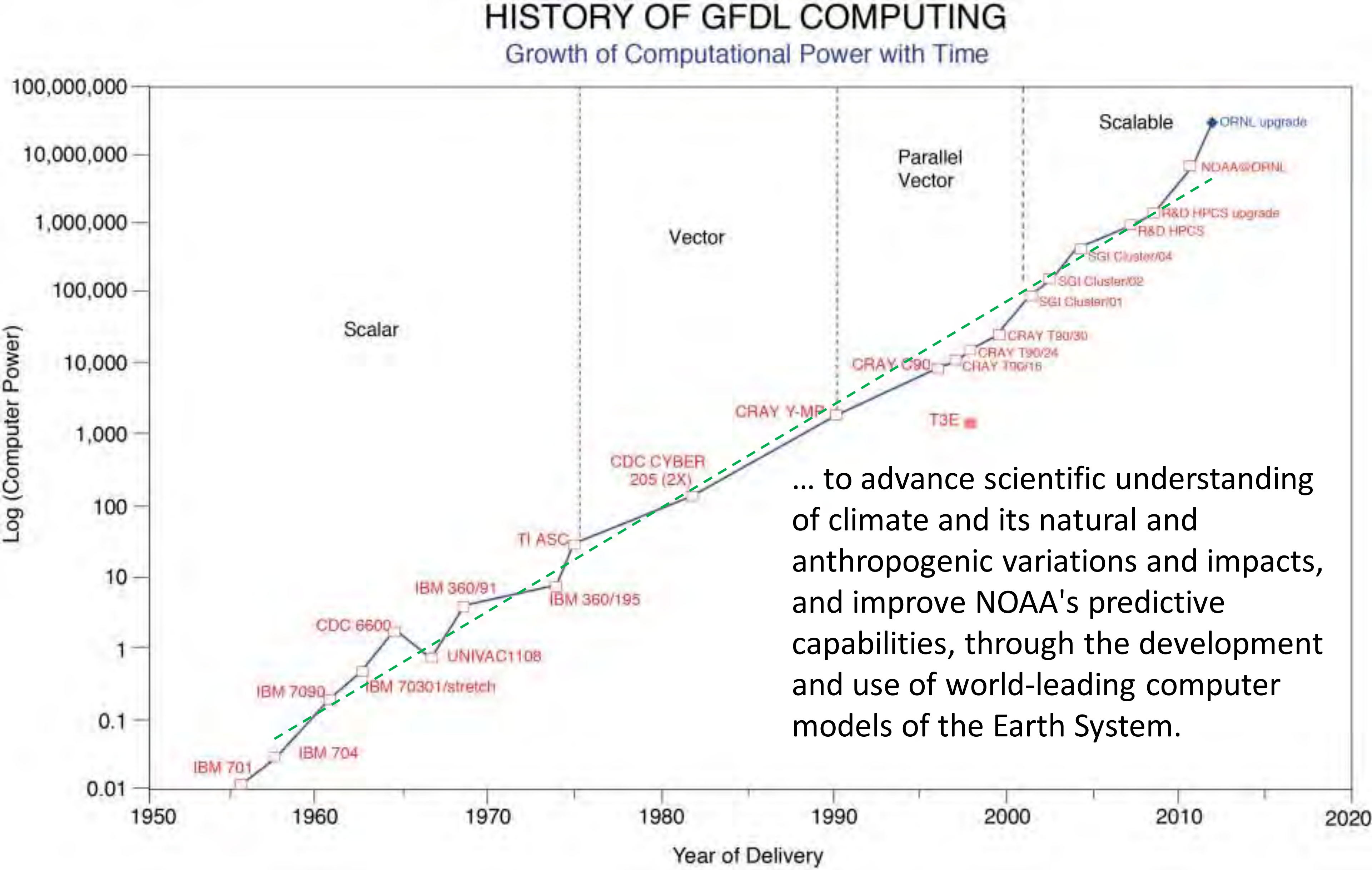
Presented by

Brian Gross (GFDL) and V. Balaji (Princeton University)

Geophysical Fluid Dynamics Laboratory Review

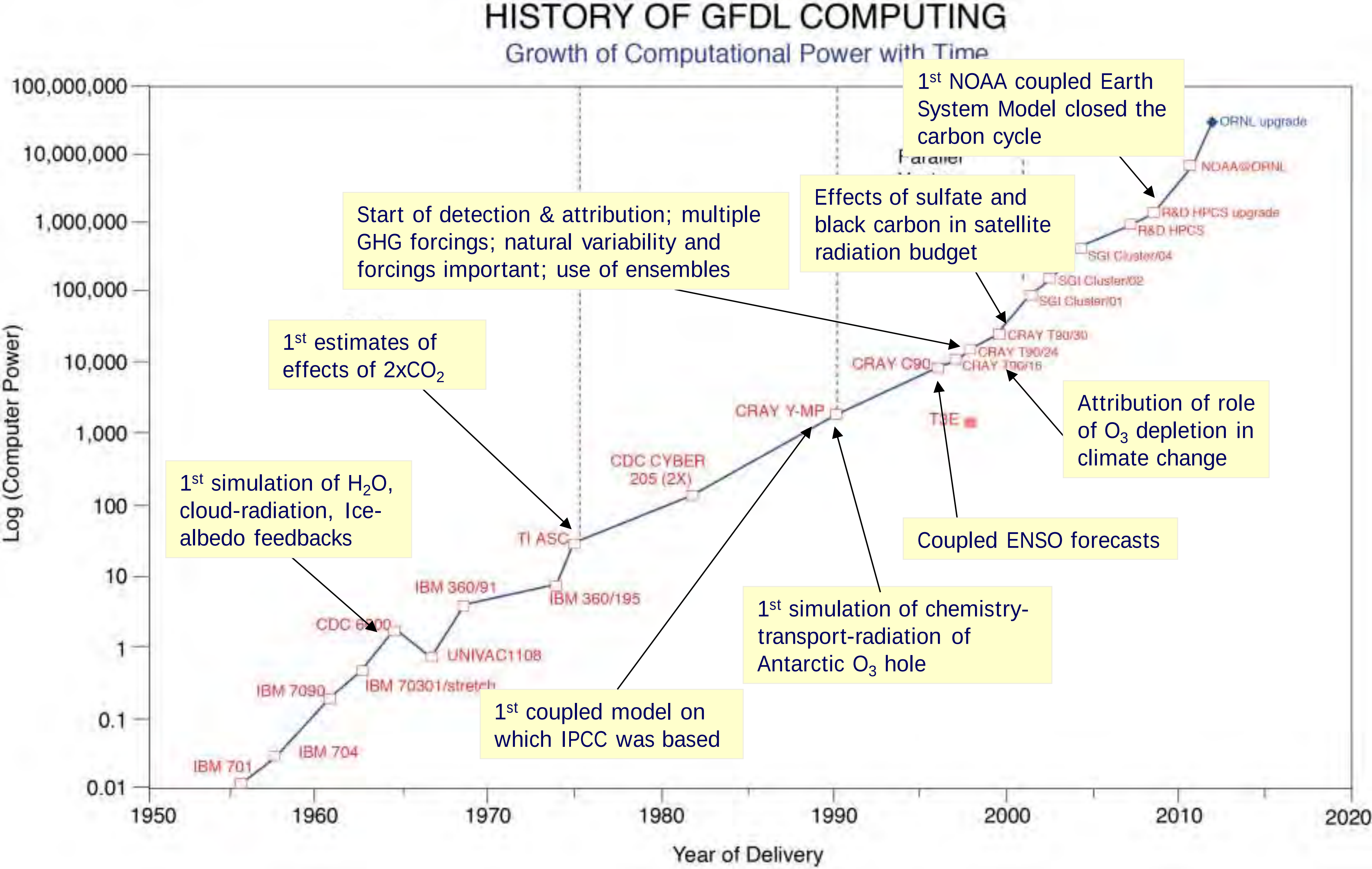
May 20 – May 22, 2014

GFDL's available computing doubles every ~2 years



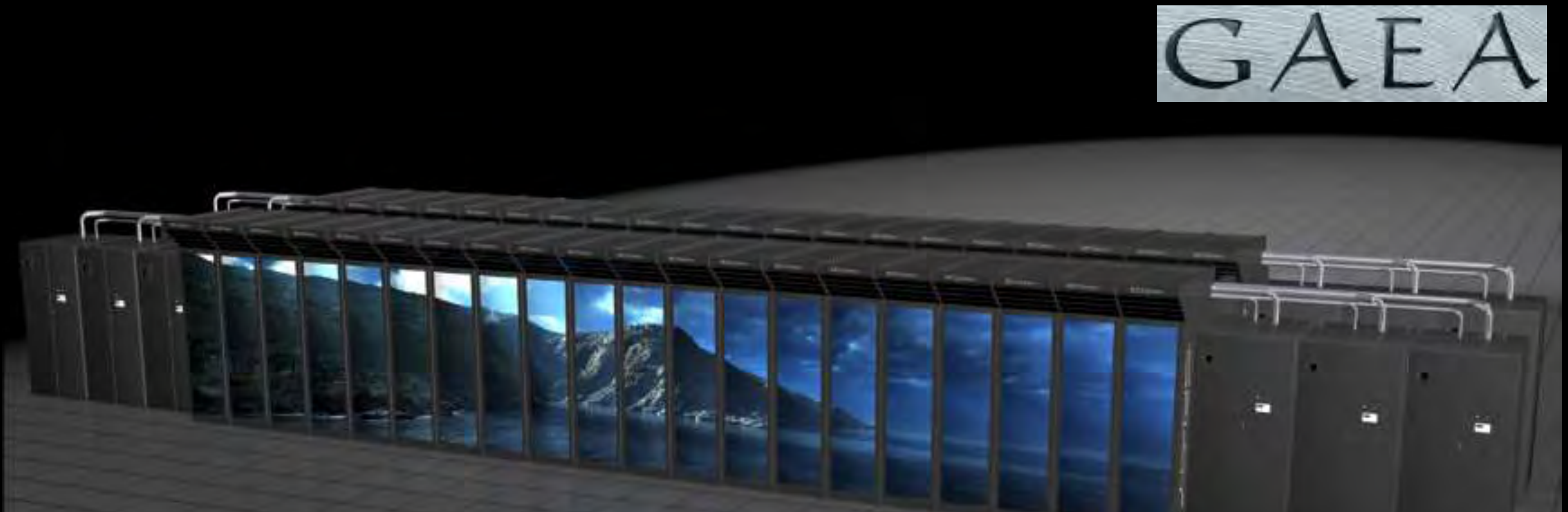
… to advance scientific understanding of climate and its natural and anthropogenic variations and impacts, and improve NOAA's predictive capabilities, through the development and use of world-leading computer models of the Earth System.

Scientific Advances are Linked to Computer Power



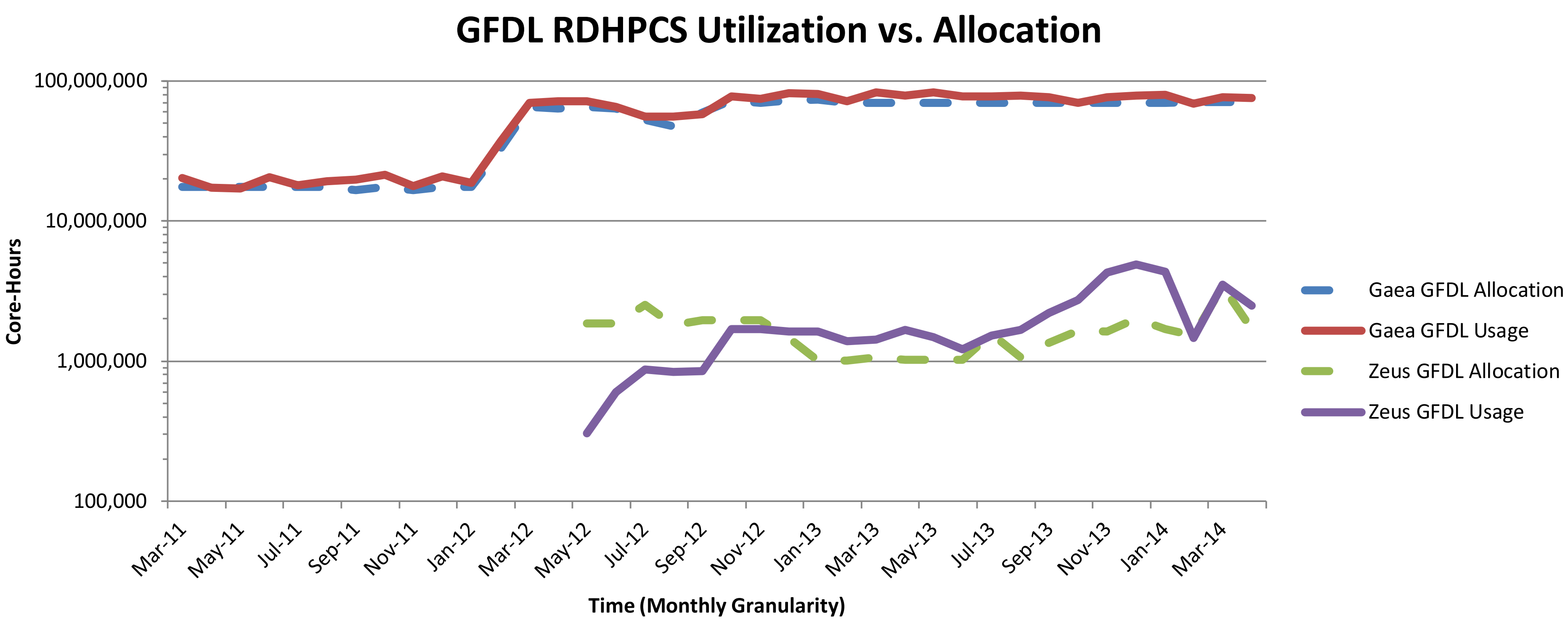
GFDL depends on sustained, dedicated production computing

An Infusion of Funds Accelerated NOAA's HPC Capacity



- ARRA provided $170M for climate modeling
  - $5M for climate data records
  - $165M for R&D HPC
- Acquired and implemented
  - Two HPC systems (Gaea at ORNL & Zeus at WV)
  - Storage and Analysis (remains a local activity)
  - Enhanced NOAA R&D HPC network
  - Associated operations and maintenance

GFDL uses as much computing as it can grab



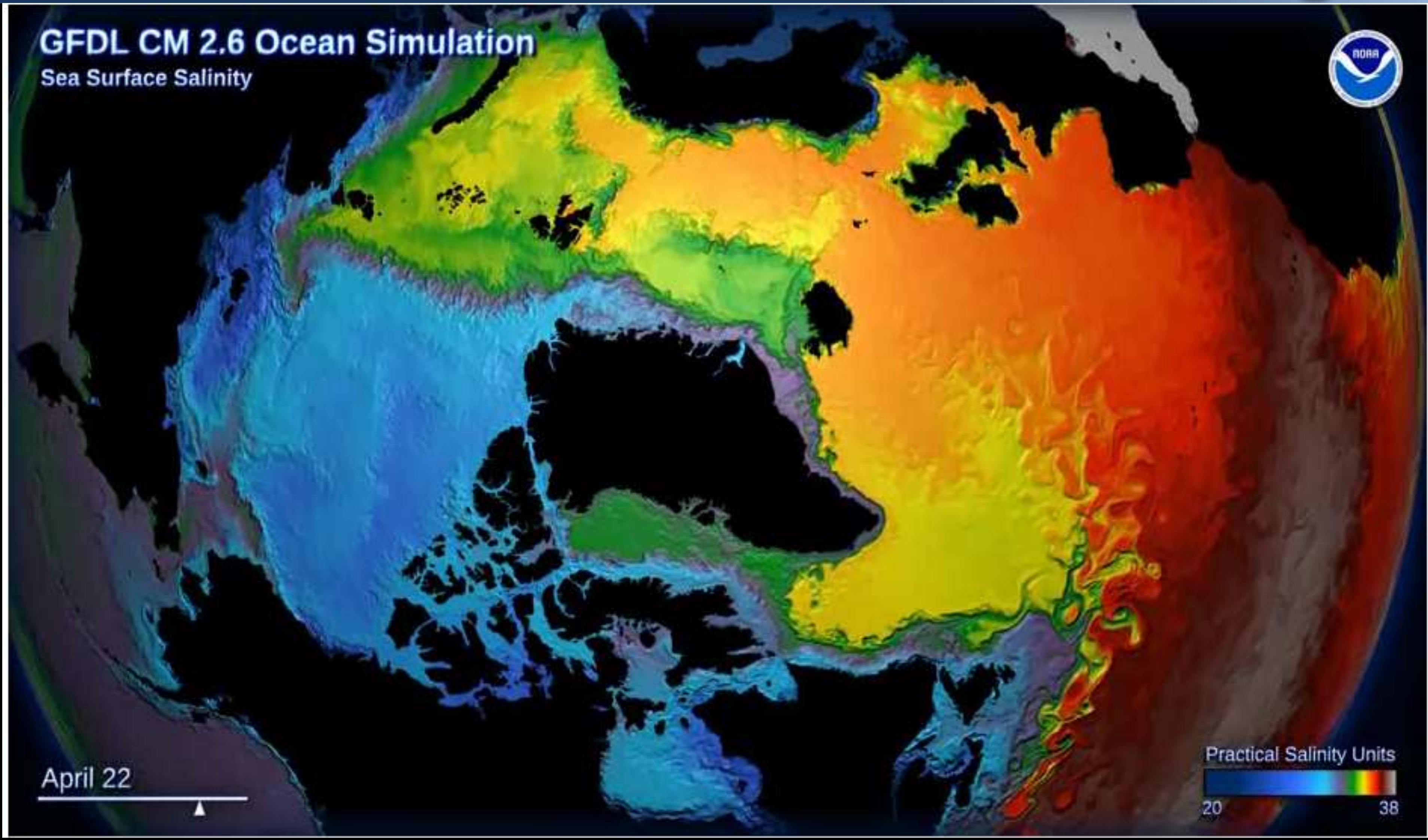
GFDL relies on external partners

- DOE/ORNL
  - GAEA system for production computing
  - TITAN to explore GPU architectures
  - Workflow research
- DOE/ANL (competitively awarded INCITE grant)
  - 150M core-hours on Mira (Blue Gene Q) to explore extreme scaling
- TACC (competitively awarded XSEDE grant)
  - To explore Intel Phi architecture on Stampede

On the horizon ….

- Disaster Recovery Act (Sandy Supplemental) supports
  - an upgrade to Zeus in FY15
  - A fine-grained parallel system in FY16
- Targeted for FY16
  - Upgrade Gaea
  - Additional support for software architecture re-engineering
- Continued partnerships to explore near-exascale performance

… but this is not enough



Computational constraints

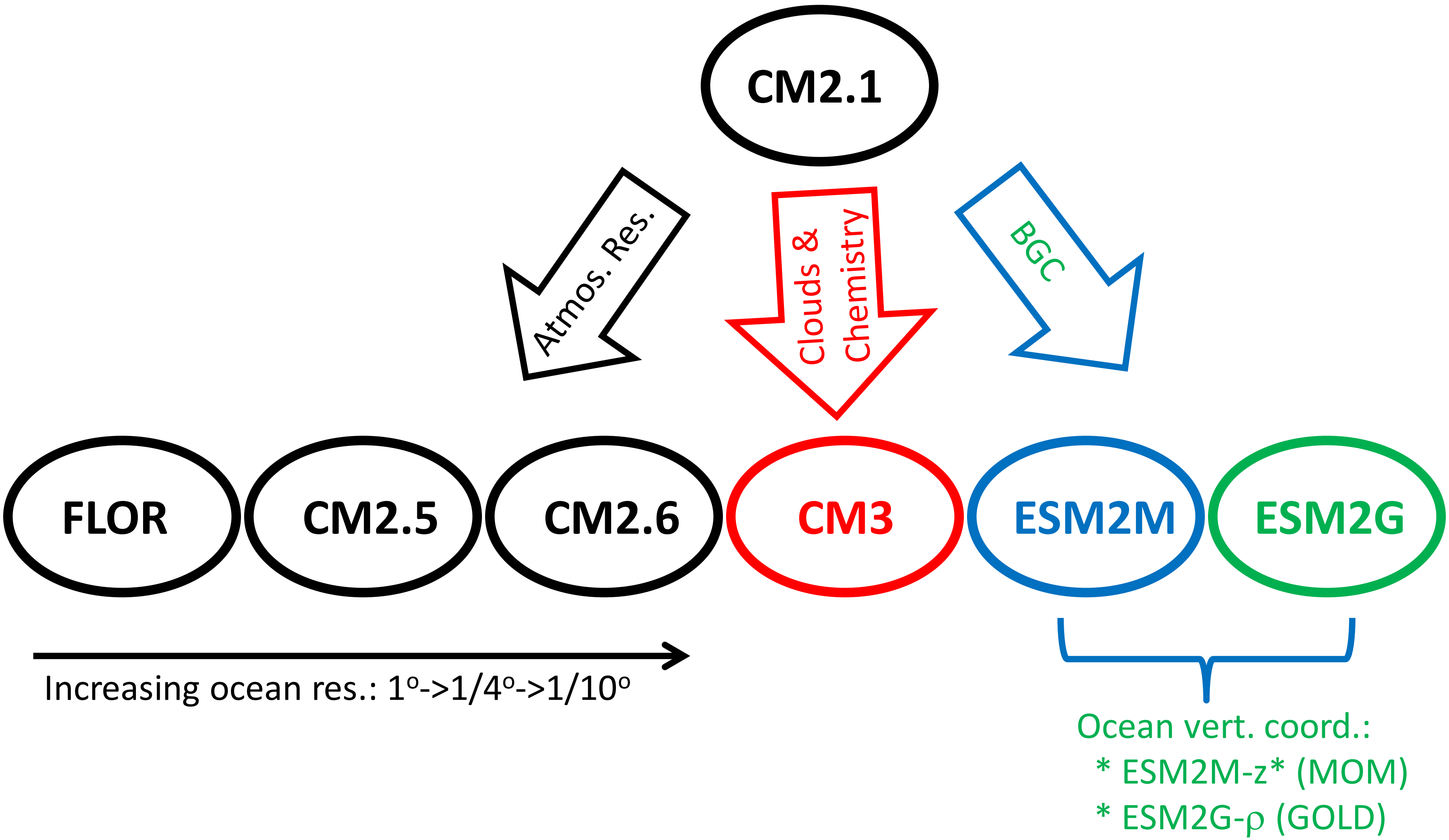
Capability: Maximum simulated years per day of a single model instance.
Capacity: Aggregate SYPD on available computing hardware.

Models choices (resolution, complexity, ensemble size) made based upon capability requirements (e.g 5-10 SYPD for dec-cen, 50-100 SYPD for carbon cycle) and available allocation.

Moore's Law: capability increases 2X every 18 months.
We are in the post-Moore era: increased concurrency, but arithmetic does not get faster (quite likely slower!)
Harder to program, understand behavior and performance, possible risks to reproducibility

Requires: judicious balanced investment between hardware and software.

GFDL ESM Genealogy



FMS and FRE

All of the models shown above were built with codes in the Flexible Modeling System (FMS): model components sharing a common codebase, common infrastructure (e.g parallelism and I/O) and superstructure (coupling interface)

The FMS Runtime Environment (FRE) provides a fault-tolerant, reproducible environment for configuring, testing, running and analyzing FMS-based models.

The FRE workflow includes publication of datasets to an external server (ESGF).

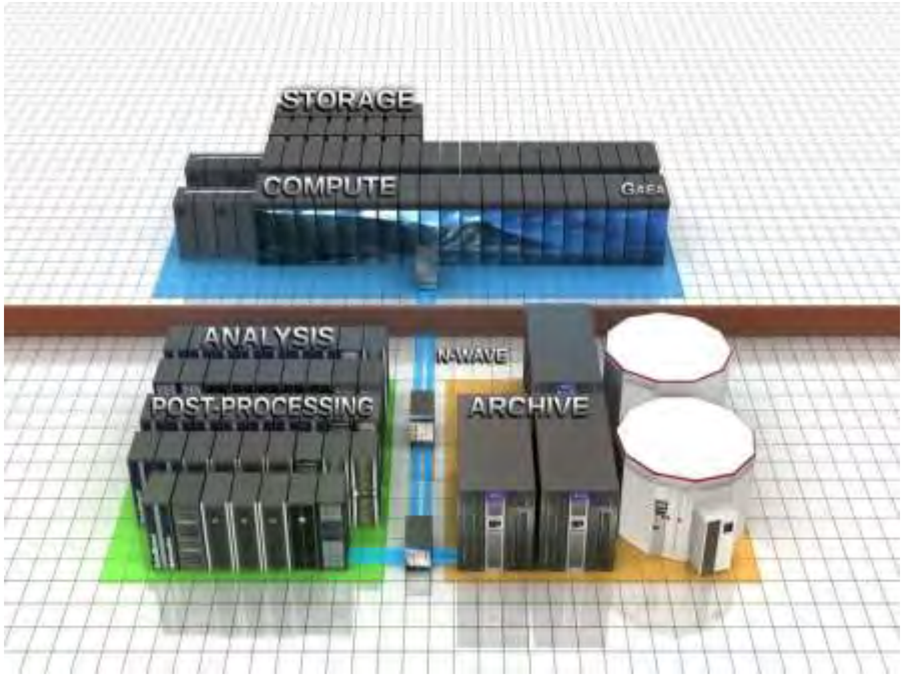
The hardware jungle and the software zoo

Processor clock speeds have stalled: it all hinges now on increased concurrency

- Hosted systems (e.g CPU+GPU)
- Many-core systems (e.g MICs)
- Equally many programming techniques! MPI, OMP, OACC, PGAS… harder to program and achieve performance…

GFDL's conservative approach: standards-based programming model (messages, threads, vectors), offloaded I/O. Extensive prototyping on experimental hardware.

Fault-resilient workflow



- Reproducible, fault-tolerant workflow across remote compute and local archive, including publication to ESGF node
- Large scale automation and testing
- Also the basis for disaster recovery plan

Summary

- GFDL Strategic Plan: process studies, development of comprehensive models, climate extremes, experimental prediction, downstream science.
- Continued development of atmospheric dynamical core, unification of ocean modeling capabilities.
- Convergence of multiple model branches into trunk model CM4.
- Forecast workflow.

Challenges

- Right way to program the next generation of parallel machines
- Component and process concurrency
- Reproducibility: what if models became more like experimental biological systems (where an individual cell “culture” is not reproducible, only the ensemble is)?
- How to understand and analyze performance on a “sea of functional units”?