Computing needs for Earth System Modeling, Predictions and Projections

Presented by
V. Balaji, CIMES and GFDL

Geophysical Fluid Dynamics Laboratory Review

October 29-31, 2019

MSD: Vision and Mission

Modeling Systems Division aims at consistent and continuous innovation at the boundary of Earth system science and computational technology. We provide an infrastructure for robust, reliable, and reproducible science for the benefit of GFDL and the wider Earth system science community.

- Maintenance, optimization and development of Flexible Modeling System (FMS), FMS Runtime Environment (FRE) and the GFDL Data Portal
- CMIP6, NMME, and other production model workflows
- Liaison support to SB/RC-identified key lab projects
- Liaison activities to community development initiatives
- at NOAA and other (UFS, NGGPS, ESGF, ...)
- Novel approaches (SENA, Machine Learning, Cloud, …)
- Division Head sits on Science Board

Staff (~20) includes Federal, University, contract staff, and an active internship program

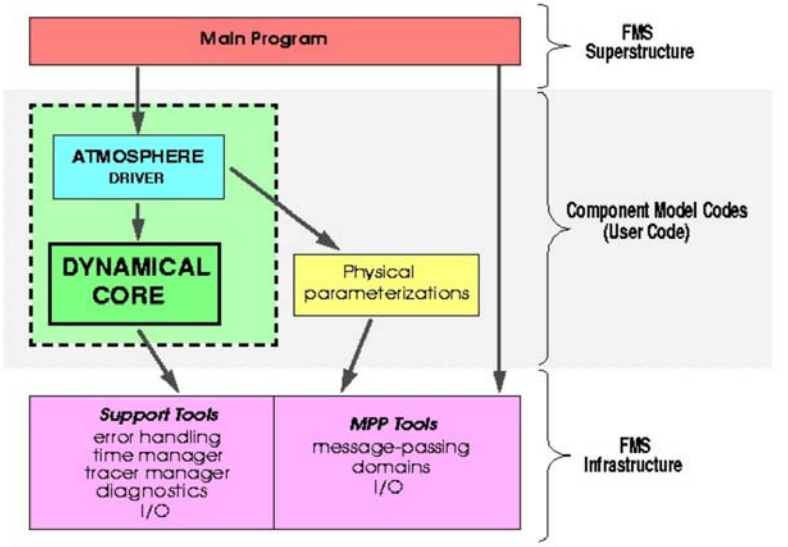
FMS provides the basis for 20 years of Unified Modeling

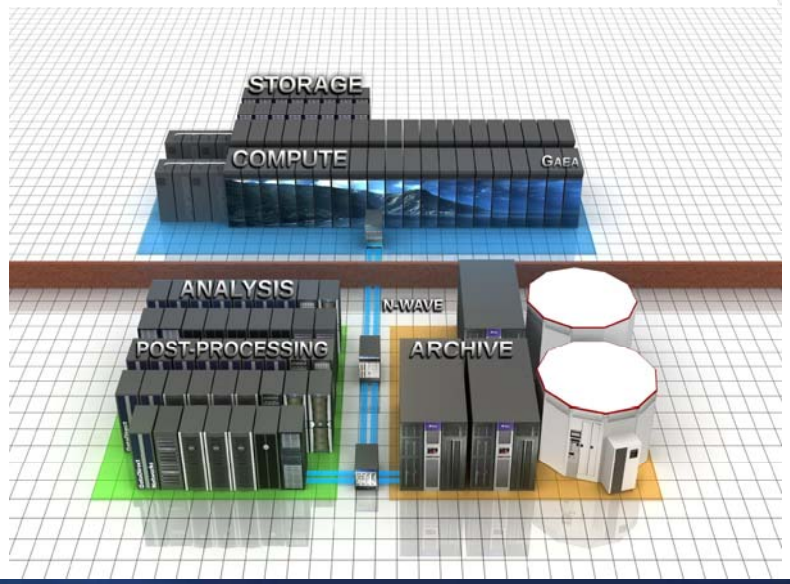
FMS: basis for all GFDL models since CM2.
Component based design: fast and scalable exchange of state vectors on independent grids, implicit and explicit coupling algorithms using exchange grid.
Since 1998:

- Median threadcount: 100X
- Top threadcount: 50000X.

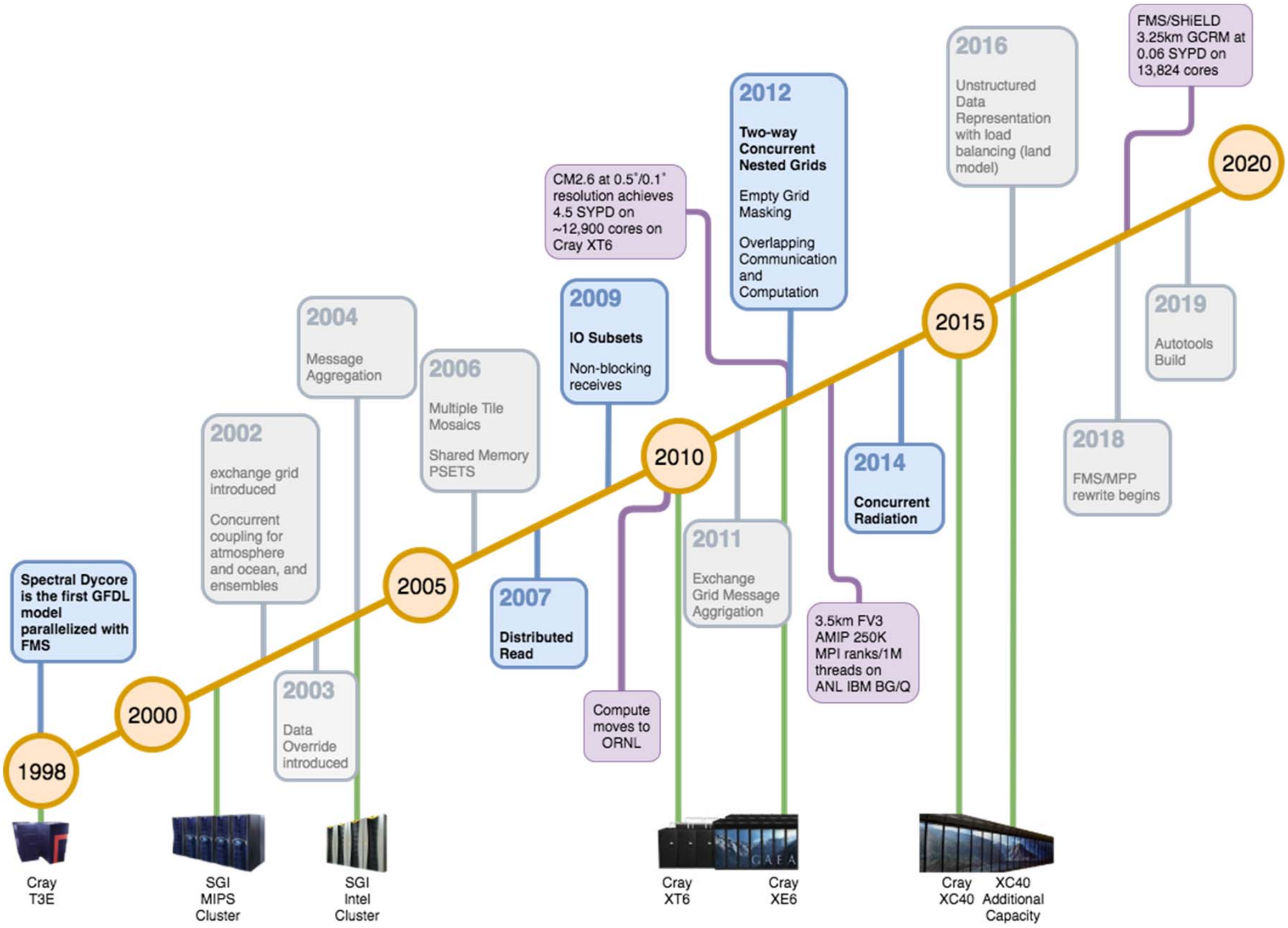
FRE: manages job graphs across distributed computing resources.
Since 2003:

- Seamless across multiple compute sites
- Data rates: 20X





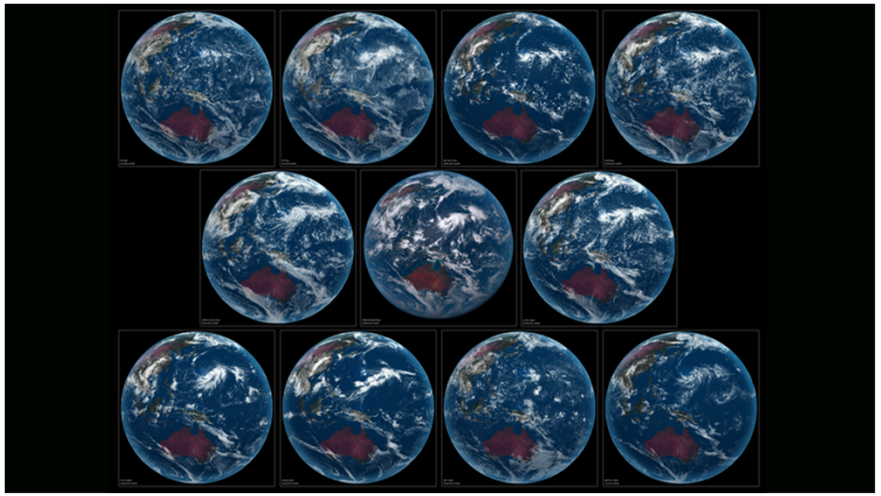
FMS provides the basis for 20 years of Unified Modeling



What can we expect at an exaflop?

Extrapolating from DYAMOND, on pre-exascale technology, current models will run at 0.06 SYPD at 1-km resolution, 16X improvement (strong scaling) needed for 1 SYPD

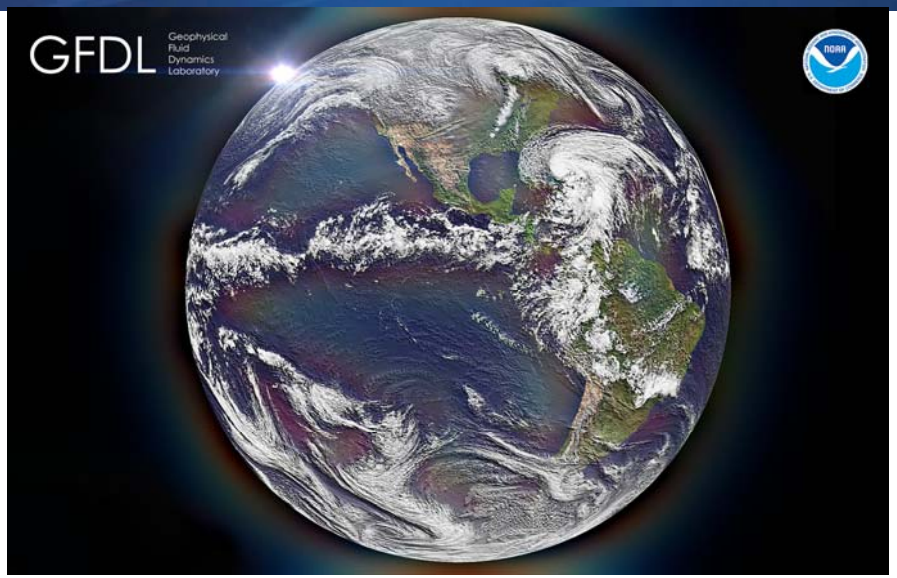
- This will be on roughly 2xGaea
- DECK: 1000 SY
- Full suite of hindcasts for seasonal prediction research: 10,000 SY.
- Weather requires 0.5 SYPD (NGGPS, ECMWF requirements)

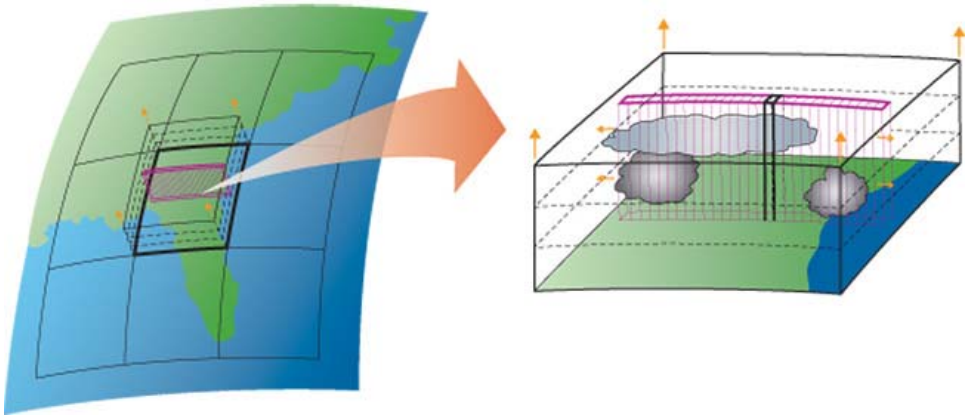


Courtesy Neumann et al, Phil Trans A (2018).

Industry is turning toward deep learning, low precision arithmetic.. should we do the same?

ML: learning from high-res models



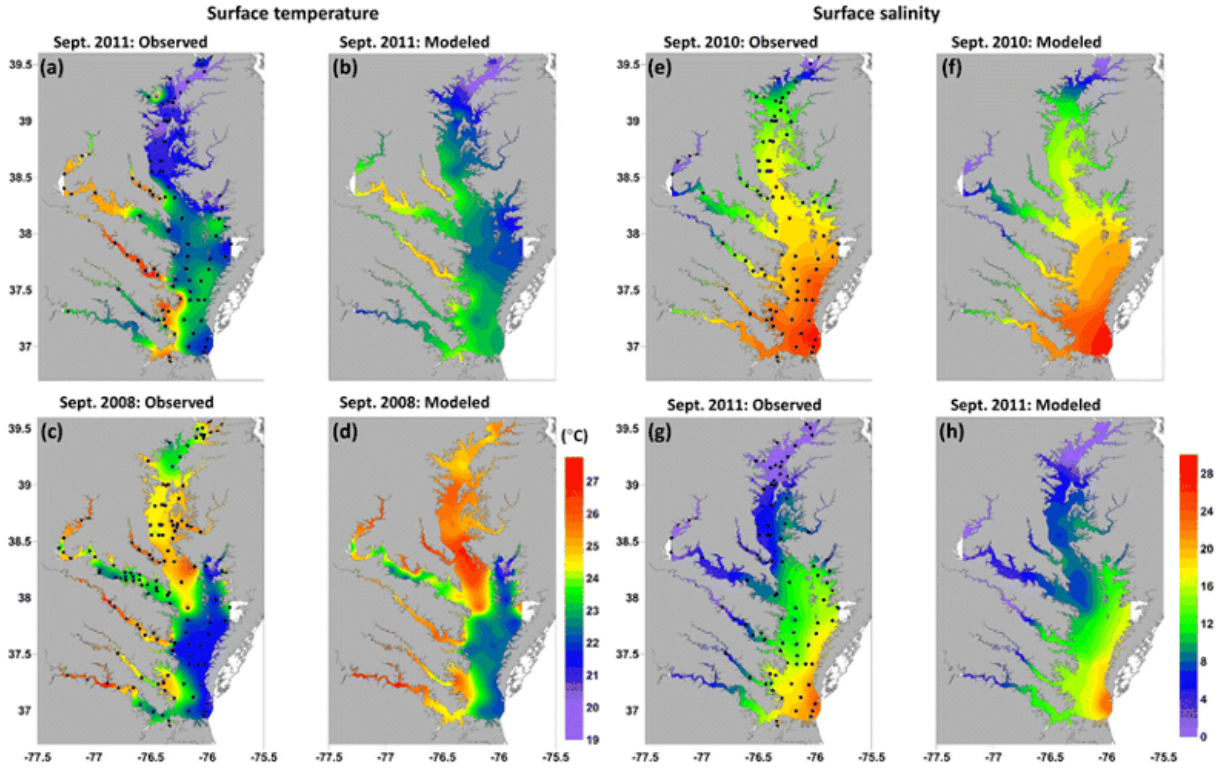


Can we learn the statistical aggregate of small scale motions? Emerging literature, e.g Schneider et al (2017), Gentine et al (2018), Bolton and Zanna (2019), Brenowitz and Bretherton (2018)

Emerging efforts at GFDL in atmosphere and ocean groups

- Atmosphere: in collaboration with Vulcan Inc.
- Ocean: funding from NOAA/OWAQ.

ML: adding high-resolution detail



ESD methods based on large scale predictors (air temperature and river flow) enhanced by ML trained on high-resolution estuarine flow and salinity anomalies
Resulting ensemble projections
used to assess climate change
impacts on Vibrio pathogens

From Muhling et al (2018). Other examples from GFDL (see talks by Shevliakova and Stock): Muhling et al (2017), Ross and Stock (2019), Chaney et al (2018)

Summary

- Several technology transitions in computing and data may stall the steady progression in model resolution, complexity, and ensemble size.
  - Resources are being devoted to the transition (see Benson talk).
- Close partnership between MSD / Science Divisions on unified modeling has helped us weather many transitions.
  - But this will be the most challenging yet!
- Turn in hardware technology toward ML offers promising avenues of research.
  - Considerable effort before they are in any way operational for science.
- NOAA making significant investments (HPC, SENA, BDP, cloud computing, AI).
  - Coordination across science teams.
- Cooperative Institute offers promising research pathways in collaboration with PICSciE.