Preliminary Information Collection and Analysis for Earth System Model Framework

Shujia Zhou

Collaborators: Glenn Higgins, Bryan Talbot, Mark Clausen at

Litton/TASC

Program managers: James Fischer and Mile Seablom at

GSFC

Litton TASC

Outline

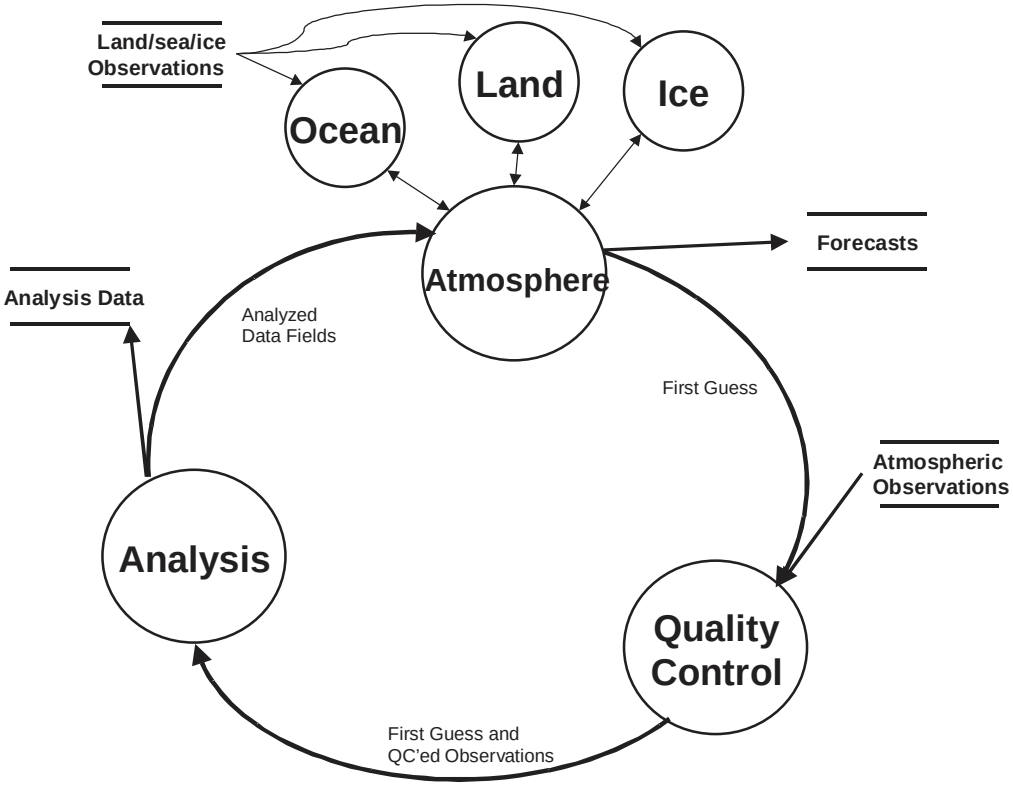
- Current climate models
- Existing frameworks
- Prototyping of strawman
- Unsolved questions
- Summary

Litton TASC

Review Current Climate Models

- Broad review of Earth Science numerical modeling and data analysis activities
  - Data Assimilation Office (DAO)
  - National Centers for Environmental Prediction (NCEP)
  - NASA Seasonal to Interannual Prediction Project (NSIPP)
  - Geophysical Fluid Dynamics Laboratory (GFDL)
  - National Center for Atmospheric Research (NCAR)
- Examine the most recent operational/production software versions of the global data circulation models and data analysis schemes and summarize:
  - Major modules and data flows
  - Major data structures and interface methods
  - Regions of parallelization
  - Programming languages and device dependencies
  - Functional overlap with other products
  - ESMF framework Issues

Climate Model and Data Assimilation Cycle



Litton TASC

Climate Model Summary

- In general *communication* is not good across five organizations. Their application models are considerably different , especially in the code structure. There is little or no documentation for codes and algorithms.
- Organizations are reusing application models through *reprogramming*.
- All codes are written in *FORTRAN*: A large amount of legacy FORTRAN 77 with subsequent development in FORTRAN 90
  - Current re-writes of DAO and NSIPP software heavily modified to take advantage of many FORTRAN 90 capabilities
- *MPI* and domain decomposition are used in most of the application models.

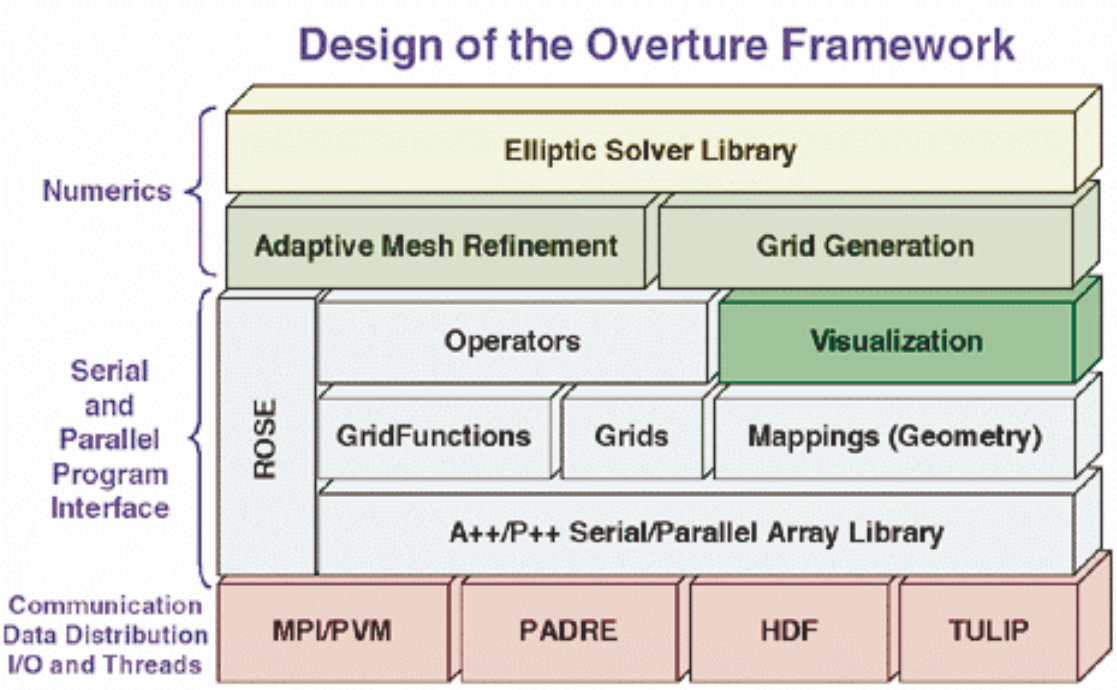
Litton TASC

Review Existing Frameworks

- **Detailed survey and analysis**
  - Overture
  - POOMA
  - GEMS
  - NCAR Flux Coupler
  - Cactus
  - UCLA Data Broker (DDB)
- **Limited Survey:** ALICE, PAWS, PETsc, ROOT, Grace, …
- **Very Limited Survey:** CAT,CCAT, DAHG, Globus, GlobalArrays, ISIS++, Kelp, PADRE, PAWS, PETsc, SAMRAI, Java, CC++, Charm++, CINT,HPC++, HPF, Java, Mentat, MPC++, Python, SILOON, Tcl/Tk, Titanium, ZPL, and others (links provided in web-site) …..

Litton TASC

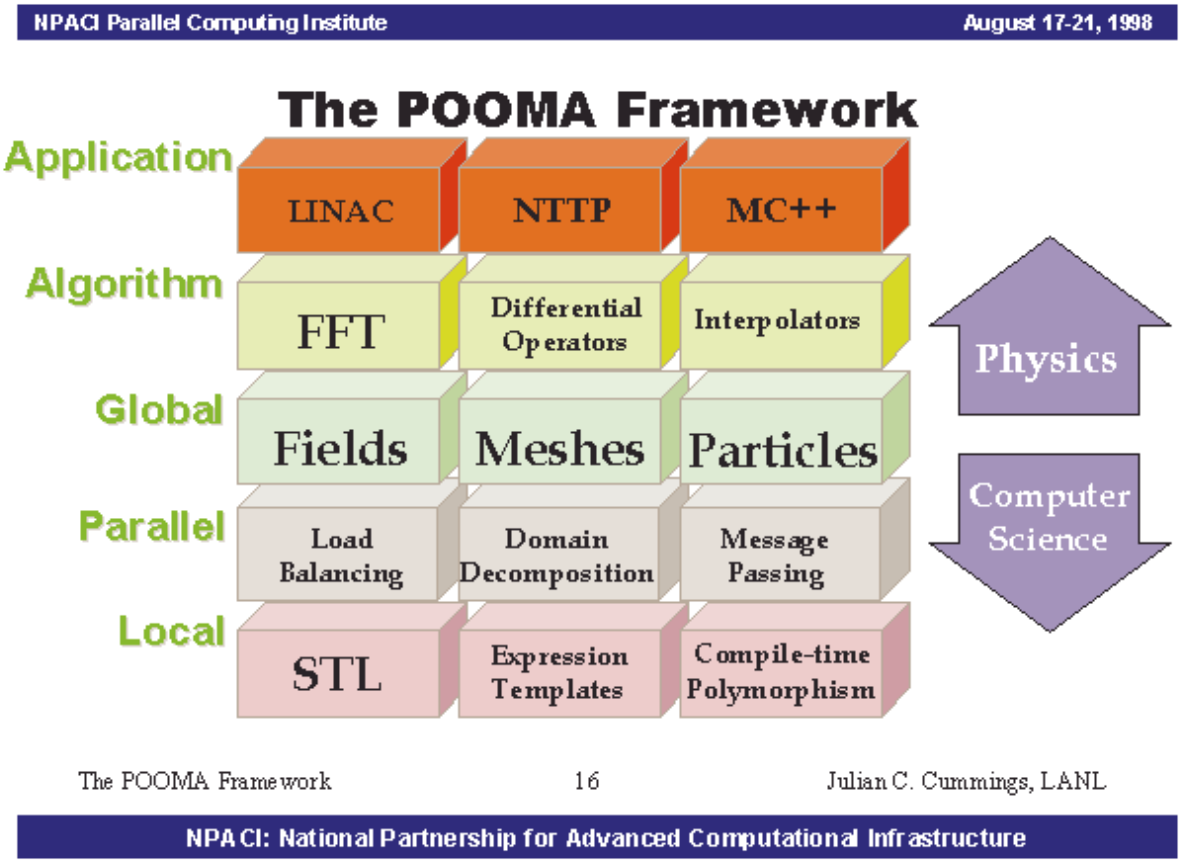
Overture (LLNL and LANL)



Overture provides an object-oriented framework for solving partial differential equations in serial and parallel computing environments.

Litton TASC

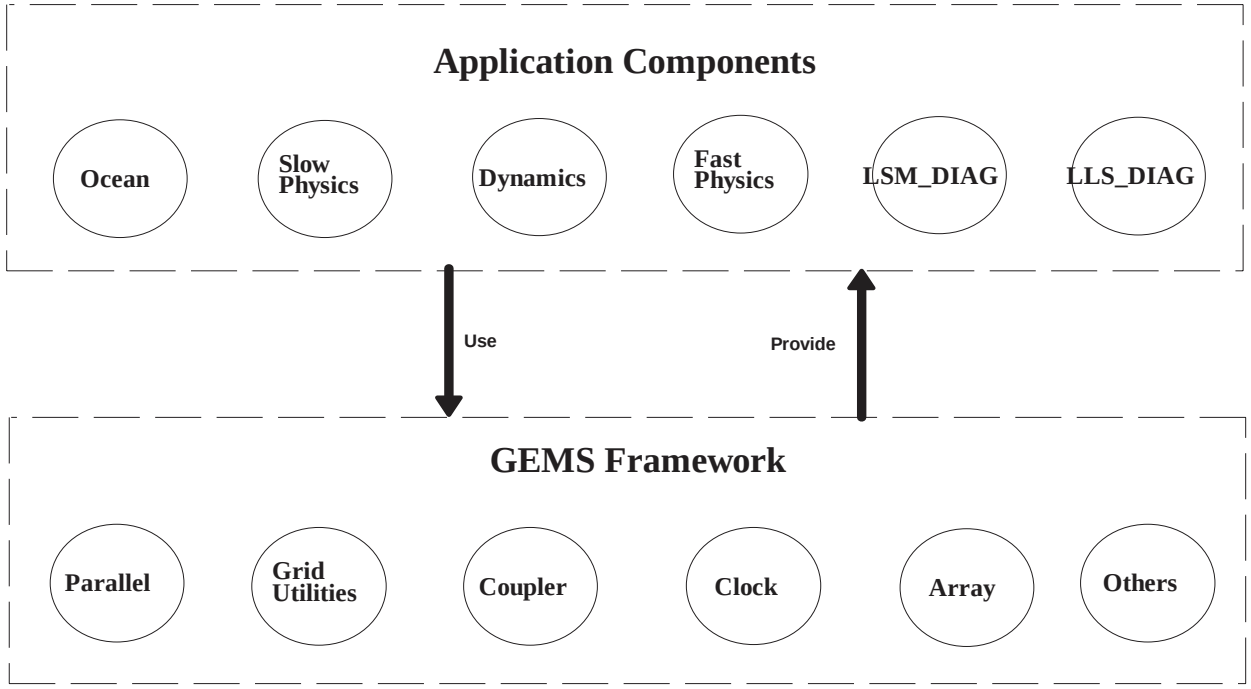
POOMA (LANL)



POOMA provides an object-oriented framework for applications in computational science requiring high-performance parallel computers

Shujia Zhou/TASC at GSFC

GEMS (NSIPP)

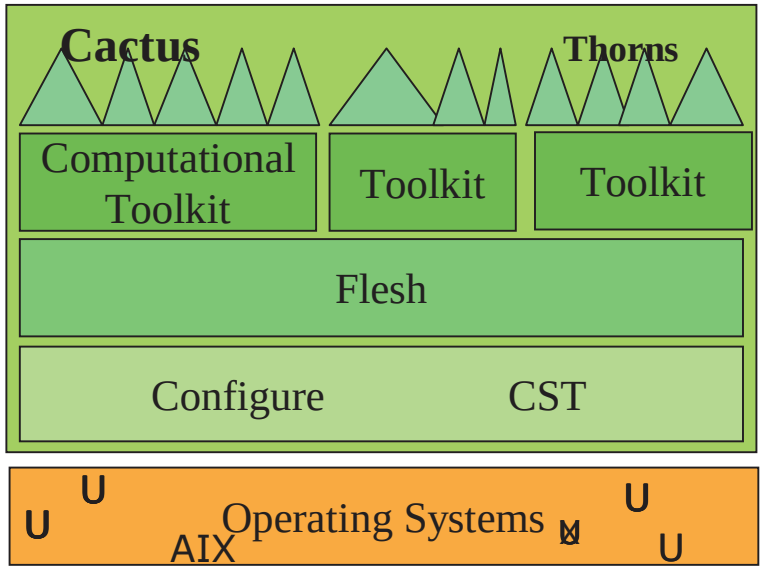


GEMS provides tools for climate modeling components, which are modularized with F90

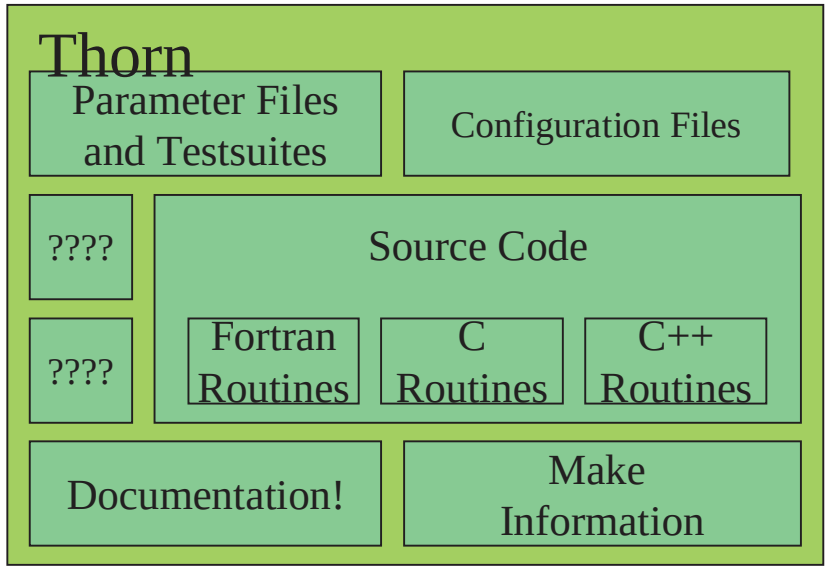
Shujia Zhou/TASC at GSFC

Cactus: (Max Plank Institute, NCSA)

System Architecture

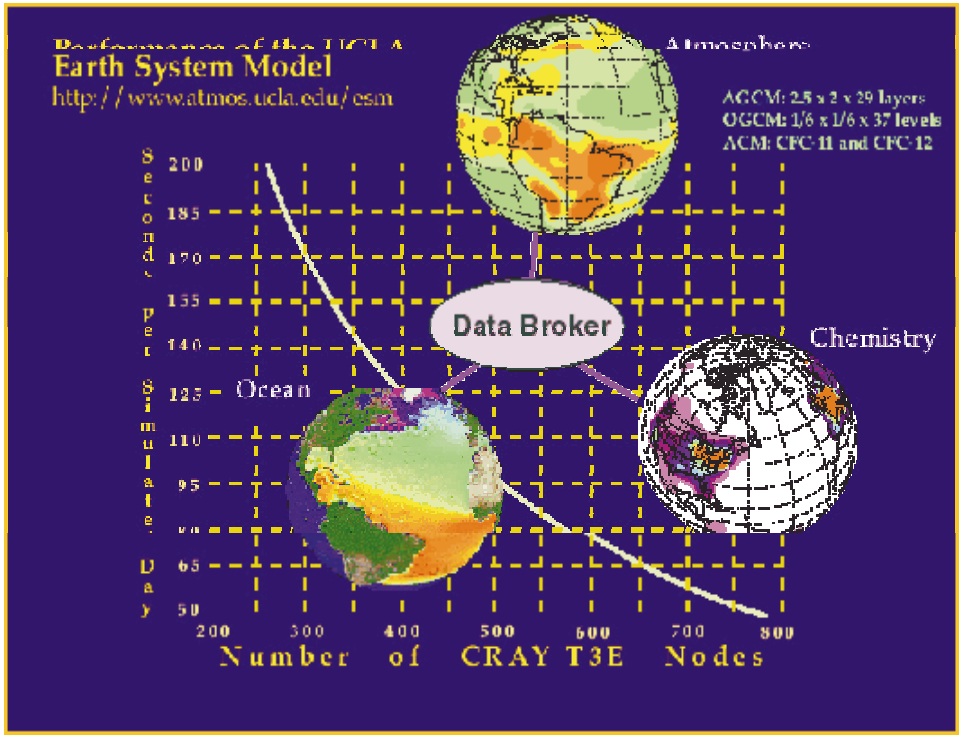


Thorn Architecture



Cactus is a component-level framework, written in C, C++, and Perl. It can support the thorn written in Fortran, C, and C++. Its parallel communication toolkit enables the code portable from single-processor to multi-processor computers.

DDB – Distributed Data Broker (UCLA)



DDB provides data interchange services between multiple climate-related models.

Litton TASC

Motivation for ESMF (from the CAN)

- Reduce the *time* required to develop and modify application codes for research and production
- Structure systems for better *management* of evolving and increasingly complex codes
- Enable software exchange and *interoperation* between major research centers
- Foster *reusability* among software components and *portability* among high-end computing architectures

Litton TASC

Earth Science Community Needs

- Develop *data-exchange* libraries to couple existing ESM components (e.g. ocean, land, ice, atmosphere)
  - Similar to UCLA/Distributed Data Broker and NCAR/Flux Coupler but more flexible interface and extensible capabilities
- Develop *tools/utilities* to facilitate developing new ESM components and sub-components (e.g. radiation, cloud)
  - Similar to GEMS, Overture, POOMA
    - Parallel communication
    - I/O
    - Grid generation, transformation utilities
    - Calendar and timer
    - Performance optimization

Litton TASC

Constraints for ESMF

- ESMF needs to serve the missions of five Federal agencies (NASA, DOE, NSF, NOAA, DOD)
- Poor communication among various research groups
- Considerable legacy *FORTRAN* code
- Ability to continue to apply expertise to write new FORTRAN code
- Object oriented concepts and languages are not familiar to researchers

Correlation of Frameworks with Community Needs: Each framework has both desirable and undesirable characteristics.

CACTUS

Desirable: Next best example of comprehensive community framework, more easily adapted to climate modeling in FORTRAN than ROOT.

Undesirable: Coupling function may need adaptation, awkward scheduling**.**

Overture

Desirable: Powerful framework shows that the combination of PDE solver and visualization is possible with OOD.

Undesirable: Lacks utilities for earth system modeling, team support is limited.

POOMA

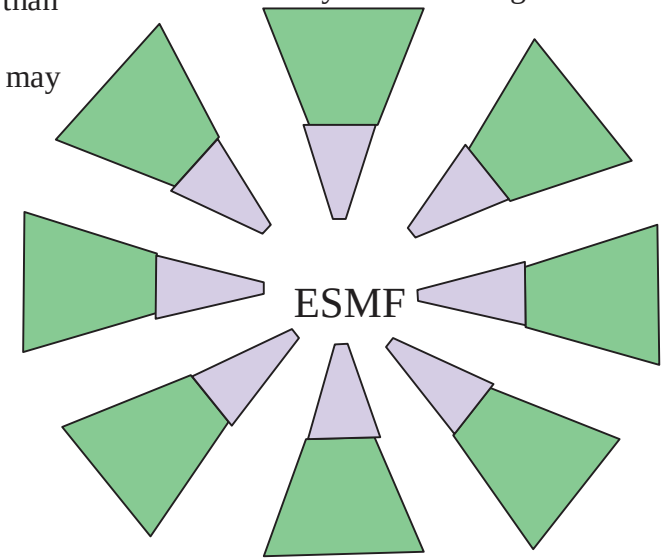
Desirable: Portable across parallel, non-parallel platforms with OOD.

Undesirable: Too complex, team support uncertain.

ROOT

Desirable: Best overall example of community-level framework

Undesirable: C++, not specifically adapted to earth system modeling.



DDB

Desirable: Excellent coupler architecture

Undesirable: Maturity, documentation, number of users, organizational support

FMS

Desirable: Basic utilities for climate modeling

Undesirable: Maturity, documentation, number of users, organizational support

GEMS

Desirable: Basic utilities for climate modeling

Undesirable: Maturity, documentation, number of users, organizational support

Flux Coupler

Desirable: Basic coupling utilities, documentation, organizational support, number of users

Undesirable: Poor coupler architecture

Framework Summary: Observations

- No single existing framework can be used without change to construct the ESMF
  - All incorporate *some* desirable features
  - All lack *some* desirable features
- Most of frameworks intend to encapsulate parallel programming and isolate machine-dependent parts.
- Majority of Climate community code written in FORTRAN 77/90
  - Most researchers do not want to migrate to other languages (e.g., C++)
  - C++ performance relative to FORTRAN questioned
  - FORTRAN is not extensively used outside the research community

Litton TASC

Framework Summary: Observations (Cont'd)

- The most powerful frameworks identified (ROOT and Overture) are written in C++
  - No equivalently powerful frameworks written in FORTRAN
  - There are, however, many useful FORTRAN libraries
- FORTRAN is missing two key OO features available in C++ and Java: inheritance and polymorphism
  - FORTRAN 2000 is scheduled to have these features in a few years
- C++ can be combined with FORTRAN in a limited way
  - No standard way of having a C++ program read dynamically allocated FORTRAN 90 data structures
  - Problem will still exist with FORTRAN 2000

Litton TASC

Framework Summary: Beliefs and Opinions

- OO design and algorithmic design are **complements** to one another, not competitors
  - Some system components are better expressed with algorithms
  - Some system components are better expressed with objects
- Most complex systems incorporate complementary components:
  - Microsoft EXCEL: Data visible (spreadsheet) and algorithms visible (Visual Basic Macros)
  - Overture framework: OO high-level structure. Some of the linear solvers for PDEs are written in FORTRAN and are legacy codes.

Litton TASC

Frameworks Summary: Beliefs and Opinions (Cont'd)

- Physics focuses on algorithmic expressions, which may be best represented by a functional language such as FORTRAN.
- Frameworks focus on relationships between participating objects (models), which are best represented by an object-oriented language such as C++, Java, or Smalltalk.
- It may be that in the climate community, the issue of language, instead of evolving properly into a question regarding the proper blend of complements (OO and algorithmic expression, Java and FORTRAN, etc.) has evolved into a competition between complements (FORTRAN or C++, etc.)
- The issue of language is absolutely central to the design of the ESMF.

Litton TASC

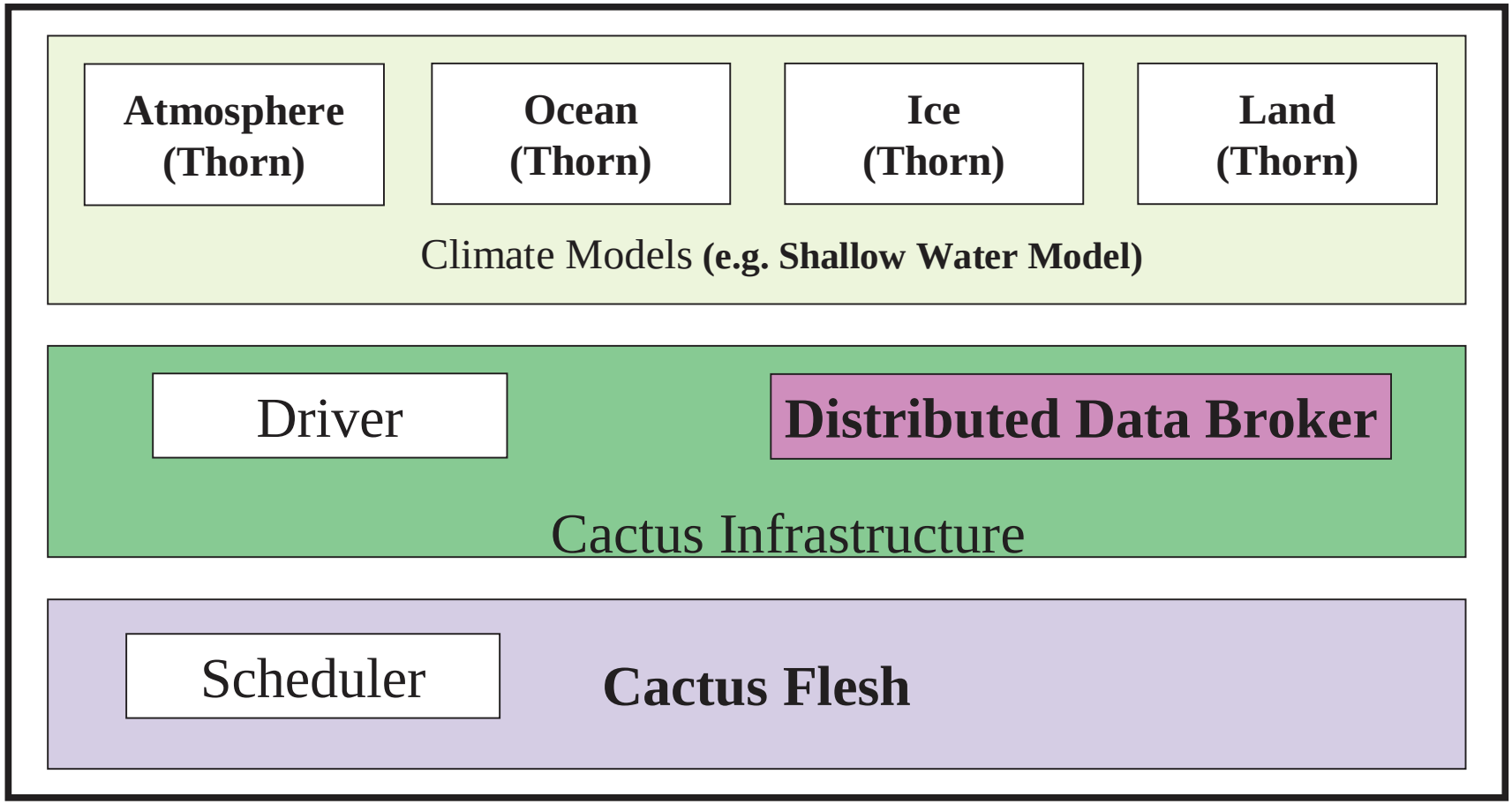
Framework Summary: Beliefs and Opinions (Cont'd)

- OO design is critical for building a flexible and extensible ESMF meeting CAN requirements
  - Inheritance and polymorphism are the important OO features that are building blocks of a comprehensive framework
  - Inheritance is a way of distributing capabilities to models that minimizes the programming burden on the researchers.
  - Polymorphism provides a simple mechanism for researchers to modify models for various situations.
  - Together, inheritance and polymorphism reduce cost and error. They provide clear hierarchical structure for a complex problem

Litton TASC

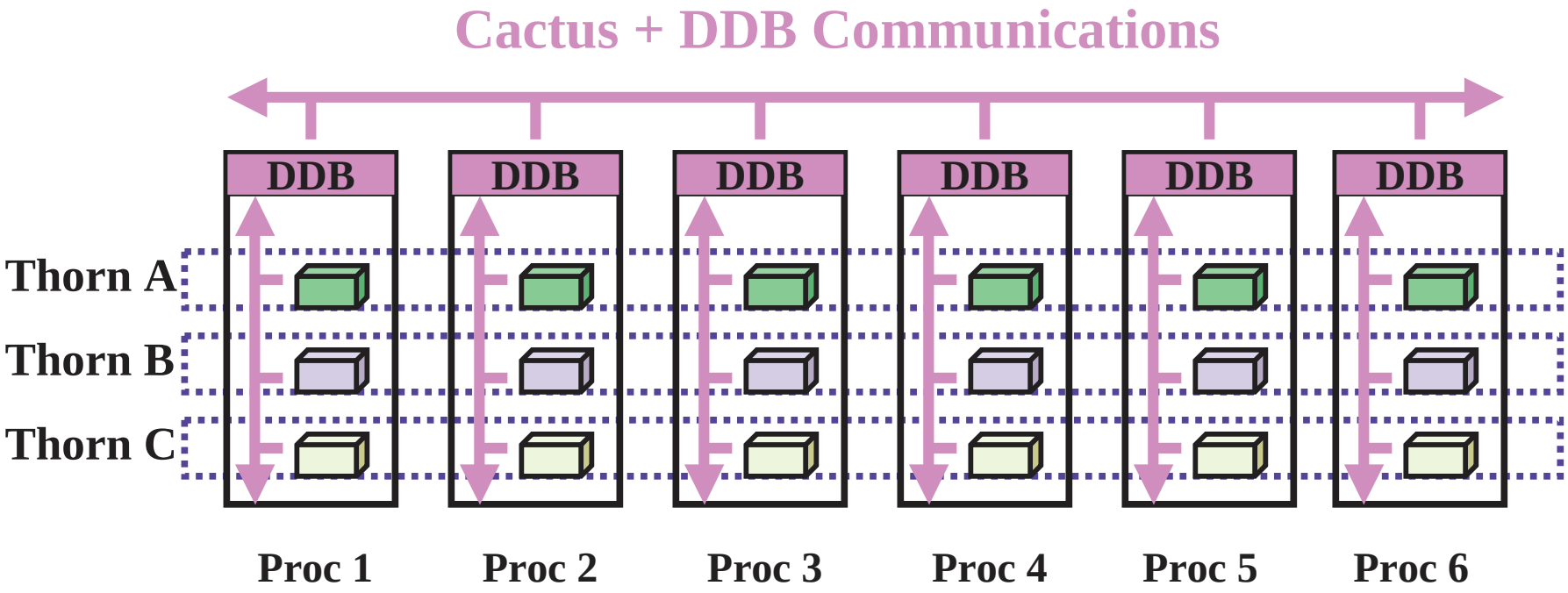
Strawman Design For ESMF

Language, Infrastructure, and Reuse

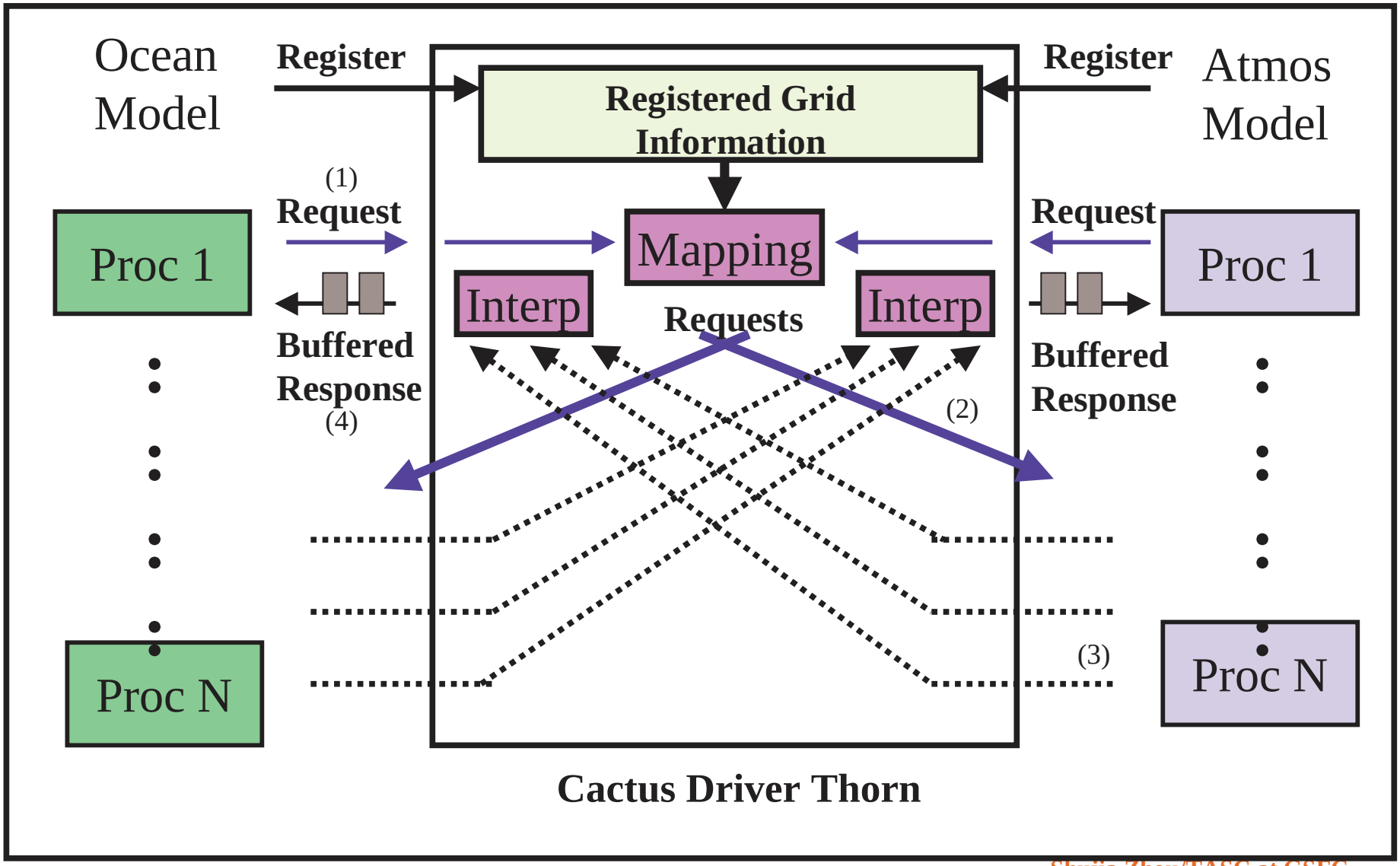


Litton TASC

Integrate DDB with Cactus to Support Data Exchange



Integrate DDB with Cactus to Support Data Exchange (Cont'd)



Litton TASC

Unsolved Questions

- How to develop a flexible interface for the community framework to support various current and future applications/models?
- How to deal with legacy Fortran codes while introducing new languages and paradigms such as object-oriented programming?
- ESMF is a framework for scientific research, which is constantly evolving. Industrial products are based on relatively mature technology. How much "software engineering principle" shall be applied?

Litton TASC

Summary

- Developing ESMF will be very interesting and challenging (the sociology more than the technology)
- Welcome ideas and suggestions
- More information on our survey and analyses can be found at http://esdcd.gsfc.nasa.gov/ESS/esmf_tasc/index.html

Litton TASC

List of Potential Collaborations

- Expert support for Investigator Team Design Reviews
- Propose to ESS “Plug-in” and “Open Source” solicitations
- Some ESS Round-3 Investigators should participate in the next Frameworks Workshop