Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-280
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 20 November 2018

Geoscientific
Model Development
Discussions
Open Access

Improving climate model coupling through a complete mesh representation: a case study with E3SM (v1) and MOAB (v5.x)

Vijay S. Mahadevan[1], Iulian Grindeanu[1], Robert Jacob[1], and Jason Sarich[1]

[1] Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL

Correspondence: V. S. Mahadevan (mahadevan@anl.gov)

Abstract.

One of the fundamental factors contributing to the spatiotemporal inaccuracy in climate modeling is the mapping of solution field data between different discretizations and numerical grids used in the coupled component models. The typical climate computational workflow involves evaluation and serialization of the remapping weights during the pre-processing step, which is then consumed by the coupled driver infrastructure during simulation to compute field projections. Tools like Earth System Modeling Framework (ESMF) Hill et al. (2004) and TempestRemap Ullrich et al. (2013) offer capability to generate conservative remapping weights, while the Model Coupling Toolkit (MCT) Larson et al. (2001) that is utilized in many production climate models exposes functionality to make use of the operators to solve the coupled problem. However, such multi-step processes present several hurdles in terms of the scientific workflow, and impedes research productivity. In order to overcome these limitations, we present a fully integrated infrastructure based on the Mesh Oriented datABase (MOAB) Tautges et al. (2004); Mahadevan et al. (2015) library, which allows for a complete description of the numerical grids, and solution data used in each submodel. Through a scalable advancing front intersection algorithm, the supermesh of the source and target grids are computed, which is then used to assemble the high-order, conservative and monotonicity preserving remapping weights between discretization specifications. The Fortran compatible interfaces in MOAB are utilized to directly link the submodels in the Energy Exascale Earth System Model (E3SM) to enable online remapping strategies in order to simplify the coupled workflow process. We demonstrate the superior computational efficiency of the remapping algorithms in comparison with other state-of-science tools and present strong scaling results on large-scale machines for computing remapping weights between the spectral-element atmosphere and finite-volume discretizations on the polygonal ocean grids.

1 Introduction

Understanding Earth's climate evolution through robust and accurate modeling of the intrinsically complex, coupled ocean-atmosphere-land-ice-biosphere models requires extreme-scale computational power Washington et al. (2008). In such coupled applications, the different component models may employ unstructured spatial meshes that are specifically generated to resolve problem-dependent solution variations, which introduces several challenges in performing a consistent solution coupling. It is known that operator decomposition and unresolved coupling errors in partitioned atmosphere and ocean model simulations Beljaars et al. (2017), or physics and dynamics components of an atmosphere, can lead to large approximation errors that cause severe numerical stability issues. In this context, one factor contributing to the spatiotemporal accuracy is the mapping between different discretizations of the sphere used in the components of a coupled climate model. Accurate remapping strategies in such multi-mesh problems are critical to preserve higher order resolution, but are in general computationally expensive given the disparate spatial scales across which conservative projections are calculated. Since the primal solution or auxiliary derived data defined on a source physics component mesh (donor model) needs to be transferred to its coupled dependent physics mesh (target model), robust numerical algorithms are necessary to preserve discretization accuracy during these operations Grandy (1999); de Boer et al. (2008), in addition to conservation and monotonicity properties in the field profile.

An important consideration is that in addition to maintaining the overall discretization accuracy of the solution during remapping, global conservation, and sometimes local element-wise conservation for quantities Jiao and Heath (2004) needs to be imposed during the workflow. Such stringent requirements on key flux fields that couple components along boundary interfaces is necessary in order to mitigate any numerical deviations in coupled climate simulations. Note that these physics meshes are usually never embedded or include trivial linear transformations, which render existence of exact projection or interpolation operators unfeasible, even if the same continuous geometric topology is discretized in the models. Additionally, the unique domain decomposition used for each of the component physics meshes complicates the communication pattern during intra-physics transfer, since aggregation of point location requests need to be handled efficiently in order to reduce overheads during the remapping workflow Plimpton et al. (2004); Tautges and Caceres (2009).

Adaptive block-structured cubed-sphere or unstructured refinement of icosahedral/polygonal meshes Slingo et al. (2009) are often used to resolve the complex fluid dynamics behavior in atmosphere and ocean models efficiently. In such models, conservative, local flux-preserving remapping schemes are critically important Berger (1987) to effectively reduce multimesh errors, especially during computation of tracer advection such as water vapor or CO_2 Lauritzen et al. (2010). This is also an issue in atmosphere models where physics and dynamics are computed on non-embedded grids Dennis et al. (2012), and the improper spatial coupling between these multi-scale models could introduce numerical artifacts. Hence, the availability of different consistent and accurate remapping schemes under one flexible climate simulation framework is vital to better understand the pros and cons of the adaptive multiresolution choices Reichler and Kim (2008).

The hub-and-spoke centralized model as shown in Fig. 1 (left) is used in the current Exascale Earth System Model (E3SM) driver, and relies on several tools and libraries that have been developed to simplify the regridding workflow within the climate community. Most of the current tools used in E3SM and the Community Earth System Model (CESM) Hurrell et al. (2013) are included in a single package called the Common Infrastructure for Modeling the Earth (CIME), which builds on previous couplers used in CESM Craig et al. (2005, 2012). These modeling tools approach the problem in a two-step computational process :

1. Compute the projection or remapping weights for a solution field from a source component physics to a target component physics as an offline process
2. During runtime, the CIME coupled solver loads the remapping weights from a file, and handles the partition-aware communication and weight matrix application to project coupled fields between components

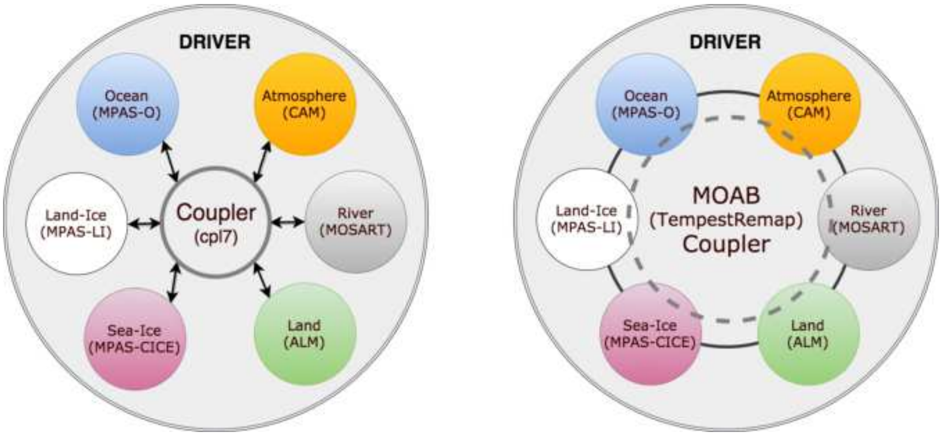


Figure 1. E3SM Coupled Climate Solver: (a) Current model (left), (b) Newer MOAB based coupler (right).

The first task in this workflow is currently accomplished through a variety of standard state-of-science tools such as the Earth Science Modeling Framework (ESMF) Hill et al. (2004), Spherical Coordinate Remapping and Interpolation Package (SCRIP) Jones (1999), TempestRemap Ullrich et al. (2013); Ullrich and Taylor (2015). The Model Coupling Toolkit (MCT) Larson et al. (2001); Jacob et al. (2005a) used in the CIME solver provides data structures for the second part of the workflow. Traditionally the first workflow phase is executed decoupled from the simulation driver during a pre-processing step, and hence any updates to the field discretization or the underlying mesh resolution immediately necessitates recomputation of the remapping weight generation workflow with updated inputs. This process flow also prohibits the component solvers from performing any runtime spatial adaptivity, since the remapping weights have to be re-computed dynamically after any changes in grid positions. To overcome such deficiencies, and to accelerate the current coupling workflow, recent efforts have been undertaken to implement a fully integrated remapping weight generation process within E3SM using a scalable infrastructure provided by the topology, decomposition and data-aware Mesh Oriented datABase (MOAB) Tautges et al. (2004); Mahadevan et al. (2015) and TempestRemap Ullrich et al. (2013) software libraries as shown in Fig. 1 (right).

The paper is organized as follows. In Section. (2), we present the necessary background and motivations to develop an online remapping workflow implementation in E3SM. Section. (3) covers details on the scalable, mesh and partition aware, conservative remapping algorithmic implementation to improve scientific productivity of the climate scientists, and to simplify the overall computational workflow for complex problem simulations. Then, the performance of these algorithms are first evaluated in serial for various grid combinations, and the parallel scalability of the workflow is demonstrated on large-scale machines in Section. (4).

2 Background

Conservative remapping of solution fields that nonlinearly couple multiple physics components is a critical task to ensure consistency and accuracy in climate and numerical weather prediction simulations Slingo et al. (2009). While there are various ways to compute a projection of a solution defined on a source grid Ω_S to a target grid Ω_T, the requirements related to global or local conservation in the remapped solution reduces the number of potential algorithms that can be employed for such problems.

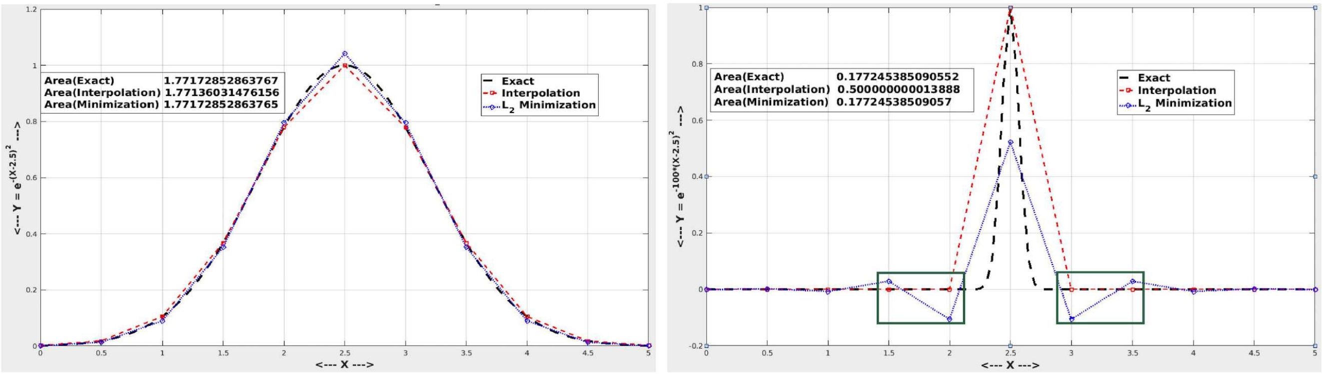
Depending on whether (global or local) conservation is important, and if higher-order, monotone interpolators are required, there are several consistent algorithmic options that can be used de Boer et al. (2008). All of these different remapping schemes usually have one of these characteristic traits: non-conservative (**NC**), globally-conservative (**GC**) and locally-conservative (**LC**).

1. **NC/GC**: Solution interpolation approximations
   - **NC**: (Approximate or exact) nearest neighbor interpolation
   - **NC/GC**: Radial Basis Function (RBF) Flyer and Wright (2007) interpolators and patch-based Least Squares reconstructions Fleishman et al. (2005)
   - **GC**: Consistent FEM interpolation and area re-normalization
2. **LC/GC**: L_2 or H_1 projection
   - **LC/GC**: Embedded FEM/FD/FV meshes in adaptive computations
   - **LC**: Intersection-based field integrators with consistent higher-order discretization Jones (1999)
   - **LC**: Constrained projections to ensure conservation Berger (1987); Aguerre et al. (2017) and monotonicity Rančić (1995)

Typically in climate applications, flux fields are interpolated using first-order (locally) conservative interpolation, while other scalar fields use non-conservative but higher-order interpolators (e.g. bilinear or biquadratic). For scalar solutions that do not need to be conserved, consistent FEM interpolation, patch-wise reconstruction schemes Fornberg and Piret (2008) or even nearest neighbor interpolation Blanco and Rai (2014) can be performed efficiently using Kd-tree based search and locate infrastructure. Vector fields like velocities or wind stresses are interpolated using these same routines by separately tackling each Cartesian-decomposed component of the field. However, conservative remapping of flux fields require computation of a supermesh Farrell and Maddison (2011), or a global intersection mesh that can be viewed as $\Omega_S \bigcup \Omega_T$, which is then used to compute projection weights that contain additional conservation and monotonicity constraints embedded in them.

In general, remapping implementations have three distinct steps to accomplish the solution field projection between grids. First, the target points of interest are identified and located in the source grid, such that, the target cells are a subset of the covering (source) mesh. Next, an intersection between this covering (source) mesh and the target mesh is performed, in order to calculate the individual weight contribution to each target cell, while consistently respecting the underlying discretization of the field data. Finally, application of the weight matrix yields the projection required to conservatively transfer the data onto the target grid.

To illustrate some key differences between some **NC** to **GC** or **LC** schemes, we show a 1-D Gaussian hill solution, projected onto a coarse grid through linear basis interpolation and L_2 minimization, as shown in Fig. 2. While the point-wise linear interpolator is computationally efficient, and second-order accurate (Fig. 2-(a)) for smooth profiles, it does not preserve the exact area under the curve. In contrast, the L_2 minimizer conserves the global integral area, but can exhibit spurious oscillatory modes as shown in Fig. 2-(b), when dealing with solutions with strong gradients (Gibbs phenomena Gottlieb and Shu (1997)). This demonstration confirms that even for the simple 1-D example, a conservative and monotonic projector is necessary to preserve both stability and accuracy for repeated remapping operator applications, in order to accurately transfer fields between grids with very different resolutions. These requirements are magnified manyfold when dealing with real-world climate simulation data.



(a) Gaussian hill $(x_0, \delta) = (2.5, 1.0)$ (b) Gaussian hill $(x_0, \delta) = (2.5, 0.1)$

Figure 2. An illustration: comparing point interpolation vs L_2 minimization; impact on conservation and monotonicity properties.

While there is a delicate balance in optimizing the computational efficiency of these operations without sacrificing the numerical accuracy or consistency of the procedure, several researchers have implemented algorithms that are useful for a variety of problem domains. In the recent years, the growing interest to rigorously tackle coupled multiphysics applications has led to research efforts focused on developing new regridding algorithms. The Data Transfer Kit (DTK) Slattery et al. (2013) from Oak Ridge National Labs was originally developed for Nuclear engineering applications, but has been extended for other problem domains through custom adaptors for meshes. DTK is more suited for non-conservative interpolation of scalar variables with either mesh-aware (using consistent discretization bases) or RBF-based meshless (point-cloud) representations Slattery (2016) that can be extended to model transport schemes on a sphere Flyer and Wright (2007). The Portage library Herring et al. (2017) from Los Alamos National Laboratory also provides several key capabilities that are useful for geology and geophysics modeling applications including porous flow and seismology systems. Using advanced clipping algorithms to compute the intersection of axis-aligned squares/cubes against faces of a triangle/tetrahedron in 2-d and 3-d respectively, general intersections of arbitrary convex polyhedral domains can be computed efficiently Powell and Abel (2015). Support for conservative solution transfer between grids and bound-preservation (to ensure monotonicity) Certik et al. (2017) has also been

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-280
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 20 November 2018

recently added. While Portage does support hybrid level parallelism (MPI + OpenMP), demonstrations on large-scale machines to compute remapping weights for climate science applications has not been pursued previously. It is also unclear whether DTK and Portage support remapping of vector fields with conservation constraints to be of direct use in climate workflows.

In earth science applications, the state-of-science regridding tool that is often used by many researchers is the ESMF library, and the set of utility tools that are distributed along with it Collins et al. (2005); Dunlap et al. (2013), to simplify the traditional offline-online computational workflow Section. (1). ESMF is implemented in a component architecture and provides capabilities to generate the remapping weights for different discretization combinations on the source and target grids in serial and parallel. ESMF provides a standalone tool, ESMF_REGRIDWEIGHTGEN, to generate *offline* weights that can be consumed by climate applications such as E3SM and OASIS3-MCT. ESMF also exposes interfaces that enable drivers to directly invoke the remapping algorithms in order to enable the fully-online workflow as well.

Currently, the E3SM components are integrated together in a hub-and-spoke model (Fig. 1 (left)), with the inter-model communication being handled by the Model Coupling Toolkit (MCT) Larson et al. (2001); Jacob et al. (2005a) in CIME. The MCT library consumes the offline weights generated with ESMF or similar tools, and provides the functionality to interface with models, decompose the field data, and apply the remapping weights loaded from a file during the setup phase. Hence, MCT serves to abstract the communication of data in the E3SM ecosystem. However, without the offline remapping weight generation phase for fixed grid resolutions and model combinations, the workflow in Fig. 1 (a) is incomplete.

Similar to the CIME-MCT driver used by E3SM, OASIS3-MCT Valcke (2013); Craig et al. (2017) is a coupler used by many European climate models, where the interpolation weights can be generated offline through SCRIP (included as part of OASIS3-MCT). An option to call SCRIP in an online mode is also available. The OASIS team have recently parallelized SCRIP to speed up its calculation time oas (2018). OASIS3-MCT also supports application of global conservation operations after interpolation, and does not require a strict hub-and-spoke coupler. Similar to the coupler in CIME, OASIS3-MCT utilizes MCT to perform both the communication of fields between components and for application of the pre-computed interpolation weights in parallel.

Even though ESMF and OASIS3-MCT have been used in online remapping studies, weight generation as part of a preprocessing step currently remains the preferred workflow for most climate models. While this decoupling provides flexibility in terms of choice of remapping tools, the data management of the mapping files for different discretization, field constraints and grids can render provenance, reproducibility and experimentation a difficult task. It also precludes the ability to handle moving or dynamically adaptive meshes in coupled simulations. Additionally, ESMF and SCRIP traditionally handle only cell-centered data that targets Finite Volume discretizations (FV to FV projections), with first-order conservation constraints. Hence, generating remapping weights for atmosphere-ocean grids with a Spectral Element (SE) source grid definition requires generation of an intermediate and spectrally equivalent, '*dual*' grid, which matches the areas of the polygons to the weight of each GLL node. Such procedures add more steps to the offline process and can degrade the accuracy in the remapped solution since the original spectral order is neglected (transformation from p-order to first order). These procedures may also introduce numerical uncertainty in the coupled solution that could produce high solution dispersion Ullrich et al. (2016).

Another production-ready remapping tool used in E3SM is the TempestRemap C++ library Ullrich et al. (2013). TempestRemap is a uni-process tool focused on the mathematically rigorous implementations of the remapping algorithms Ullrich and Taylor (2015); Ullrich et al. (2016) and provides higher order conservative and monotonicity preserving interpolators with different discretization basis such as (Finite Volume (FV), the spectrally equivalent continuous Galerkin Finite Element with Gauss-Lobatto quadrature (cGLL), and dis-continuous Galerkin Finite Element with Gauss-Lobatto quadrature (dGLL)). This library was developed as part of the effort to fill the gap in generating consistent remapping operators for non-FV discretizations without a need for intermediate dual meshes. Computation of conservative interpolators between any combination of these discretizations (FV, cGLL, dGLL) and grid definitions are supported by TempestRemap library. However, since this regridding tool can only be executed in serial, the usage of TempestRemap prior to the work presented here has been restricted primarily to generating the required mapping weights in the offline stage.

There are several challenges in scalably computing the regridding operators in parallel, since it is imperative to have both a mesh- and partition-aware datastructure to handle this part of the regridding workflow. In the E3SM workflow supported by CIME, the ESMF-regridder understands the component grid definitions, and generates the weight matrices (offline). The CIME driver loads these operators at runtime and places them in MCT datatypes which treats them as discrete operators to compute the interpolation or projection of data on the target grids. Additional changes in conservation requirements or monotonicity of the field data cannot be imposed as a runtime or post-processing step in such a workflow. In the current work, we present a new infrastructure with scalable algorithms implemented using the MOAB mesh library and TempestRemap package to replace the ESMF-E3SM-MCT remapper/coupler workflow. A detailed review of the algorithmic approach used in the MOAB-TempestRemap (MBTR) workflow, along with the software interfaces exposed to E3SM is presented next.

3 Algorithmic approach

Efficient, conservative and accurate multi-mesh solution transfer workflows Jacob et al. (2005b); Tautges and Caceres (2009) are a complex process. This is due to the fact that in order to ensure conservation of critical quantities in a given norm, exact cell intersections between the source and target grids have to be computed. This is complicated in a parallel setting since the domain decompositions between the source and target grids may not have any overlaps, making it a potentially all-to-all collective communication problem. Hence, efficient implementations of regridding operators need to be mesh, resolution, field and decomposition aware in order to provide optimal performance in emerging architectures.

Fully online remapping capability within a complex ecosystem such as E3SM requires a flexible infrastructure to generate the projection weights. In order to fullfill these needs, we utilize the MOAB mesh datastructure combined with the TempestRemap libraries in order to provide a in-memory remapping layer to dynamically compute the weight matrices during the setup phase of the simulations. The introduction of such a software stack allows higher order conservation of fields while being able to transfer and maintain field relations in parallel, within the context of the fully decomposed mesh view. This is an improvement to the E3SM workflow where MCT is oblivious to the underlying mesh datastructure in the component models. Having a fully mesh aware datastructure also provides opportunities to implement dynamic load-balancing algorithms to gain optimal performance on large-scale machines.

Let $N_{c,s}$ be the component processes for source mesh, $N_{c,t}$ be the component processes for target mesh and N_x be the coupler processes where the remapping operator is computed. More generally, the problem statement can be defined as: transfer a solution field U defined on the domain Ω_S and processes $N_{c,s}$, to the domain Ω_T and processes $N_{c,t}$, through a centralized coupler with domain information $\Omega_S \bigcup \Omega_T$ defined on N_x processes. Such a complex online remapping workflow for projecting the field data from a source to target mesh follows the algorithm shown in Alg. 1.

In the following sections, the new E3SM online remapping interface implemented with a combination of the MOAB and TempestRemap libraries is explained. Details regarding the algorithmic aspects to compute conservative, high-order remapping weights in parallel, without sacrificing discretization accuracy on next generation hardware are presented.

3.1 Interfacing to Component Models in E3SM

Within the E3SM simulation ecosystem, there are multiple component models (atmosphere-ocean-land-ice-runoff) that are coupled to each other. While the MCT infrastructure only allowed for a numbering of the grid points, the new MOAB-based coupler infrastructure provides the ability to natively interface to the underlying mesh, and understand the field DoF data layout associated with each model. MOAB can understand the difference between values on a cell center and values on a cell edge or corner. In the current work, the MOAB mesh database has been used to create the relevant integration abstraction for the HOMME atmosphere model (cubed-sphere SE grid) and the MPAS ocean model (polygonal meshes with holes). Since the details of the mesh are not available at the level of the coupler interface, additional MOAB calls are added to HOMME and MPAS to describe the details of the mesh to MOAB. The atmosphere-ocean coupling requires the largest computational effort in the coupler (since they cover about 70% of the coupled domain), and hence the bulk of the discussions in the current work will focus on remapping and coupling between these two models.

MOAB can handle the finite-element zoo of elements on a sphere (triangles, quadrangles, and polygons) making it an appropriate layer to store both the mesh layout (vertices, elements, connectivity, adjacencies) and the parallel decomposition for the component models along with information on shared and ghosted entities. While having a uniform partitioning methodology across components may be advantageous for improving the efficiency of coupled climate simulations, the parallel partition of the meshes are chosen according to the requirements in individual component solvers. Fig. 3 shows an example of a replicated SE and MPAS meshes, visualized through the native MOAB plugin for VisIt.

The coupled field data that is to be remapped from the source grid to the target grid also needs to be serialized as part of the MOAB mesh database in terms of a 'Tag'. For E3SM, we use element-based tags to store ${n_p}^2$ values per element, where for the atmosphere n_p is the order of SE discretization and for MPAS ocean, $n_p = 1$. With this complete description of the mesh and associated data for each component model, MOAB contains the necessary information to proceed with the remapping workflow.

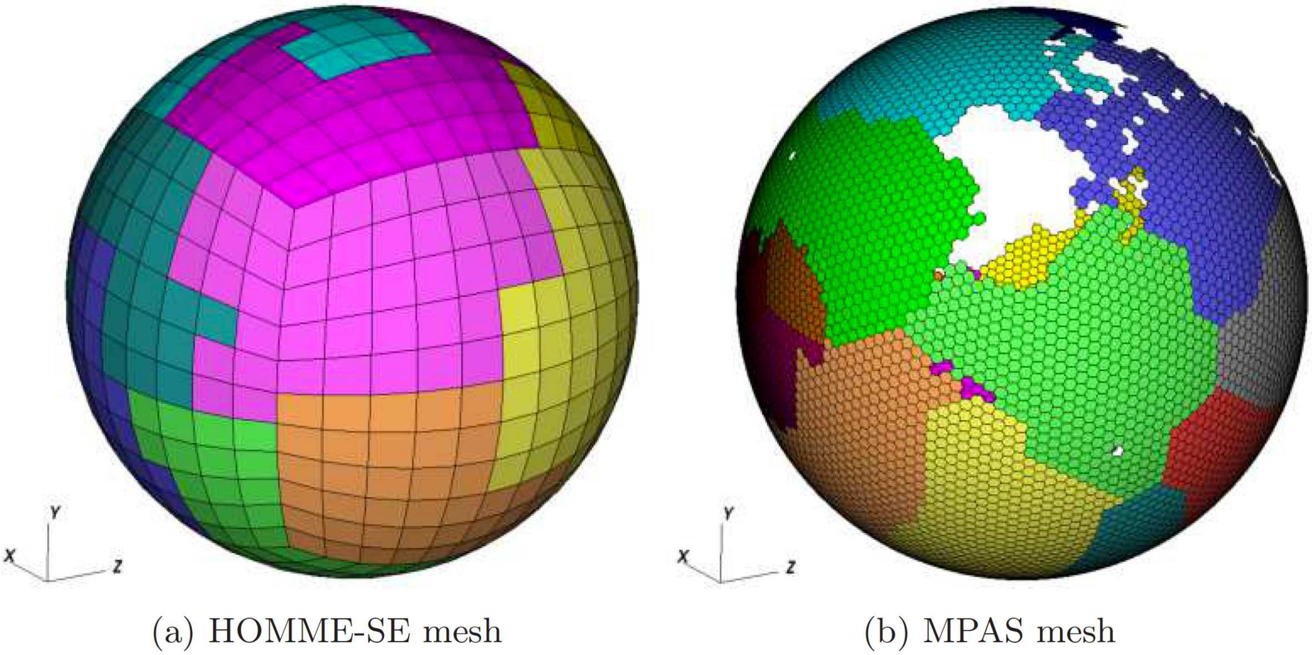
Algorithm 1 MOAB-TempestRemap parallel regridding workflow

```
Input: Partitioned and distributed native component meshes on $N_{c,l}$ processes
Result: Remapping weight matrix $W_{ij}$ computed for a source ($i$) and target ($j$) mesh pair on $N_x$ coupler processes

Scope: Coupler $N_x \leftarrow$ component mesh $N_{c,l}$
for each component $l \in [s,t]$ do
    – create in-memory copy of component mesh/data with MOAB
    – migrate MOAB component mesh to coupler; repartition from $N_{c,l} \rightarrow N_x$
end for

Scope: Compute pair-wise intersection mesh on coupler processes $N_x$
for each mesh pair to be regridded: $\Omega_S$ and $\Omega_T$ in $N_x$ do
    Ensure: {local source mesh fully covers target mesh}
    if $(\Omega_T - \Omega_T \cap \Omega_S) \neq 0$ then
        collectively gather coverage mesh $\Omega_{Sc}$ on $N_x \mid (\Omega_T - \Omega_T \cap \Omega_{Sc}) = 0$
    end if
    – store communication graph to send/receive between $N_{c,l}/N_x$
    – compute $\Omega_{ST} = \Omega_{Sc} \bigcup \Omega_T$ through an advancing-front algorithm Löhner and Parikh (1988); Gander and Japhet (2009)
    – evaluate source/target element mapping for $e_i \in \Omega_{ST}$
    – exchange ghost cell information for $\Omega_{ST}$
end for

Scope: Integrate over $\Omega_{ST}$ to compute remapping weights
for each intersection polygon element $e_i \in \Omega_{ST}$ do
    – Tessellate $e_i$ into triangular elements with reproducible ordering
    – Compute projection integral with consistent Triangular quadrature rules
    – Determine row/col DoF coupling through $e_i$ parent association to $\Omega_S/\Omega_T$
    – Assemble local matrix weights such that $W_{ij} = \sum_1^{N_x} w_{ij}$, where $w_{ij}$ represents the coupling between local target (row $i$) and source (col $j$) in projection operator
end for
```



(a) HOMME-SE mesh (b) MPAS mesh

Figure 3. MOAB representation of partitioned component meshes.

3.2 Migration of Component Mesh to Coupler

E3SM's driver supports multiple modes of partitioning the various components in the global processor space. This is usually fine tuned based on the estimated computational load in each physics, according to the problem case definition. A sample process layout for a E3SM run on 9000 processes with ATM on 5400 and OCN on 3600 tasks is shown in Fig. 4. In the case shown in the schematic, $N_{c,atm} = 5400$, $N_{c,ocn} = 3600$ and $N_x = 4800$. In such a processor execution layout, the atmosphere component mesh from HOMME, distributed on $N_{c,atm}$ (5400) tasks needs to be migrated and redistributed on N_x (4800 tasks). Similarly, from $N_{c,ocn}$ (3600) to N_x (4800) tasks for the MPAS ocean mesh. Since the remapping process is performed only in the coupler processing elements within the hub-and-spoke model Fig. 1, inference of a communication pattern becomes necessary to ensure scalable data transfers between the components and the coupler.

For illustration, let N_c be the number of component processing elements, and N_x be the number of coupler processing elements. In order to migrate the mesh and associated data from N_c to N_x, we first compute a trivial partition of elements that map directly in the partition space, the same partitioning as used in the CIME-MCT coupler. In MOAB, we have exposed parallel graph and geometric repartitioning schemes through interfaces to Zoltan or ParMetis, in order to evaluate optimized migration patterns to minimize the volume of data communicated between component and coupler processing units. We intend to analyze the impact of different migration schemes on the scalability of the remapping process in Section. (4). These optimizations have

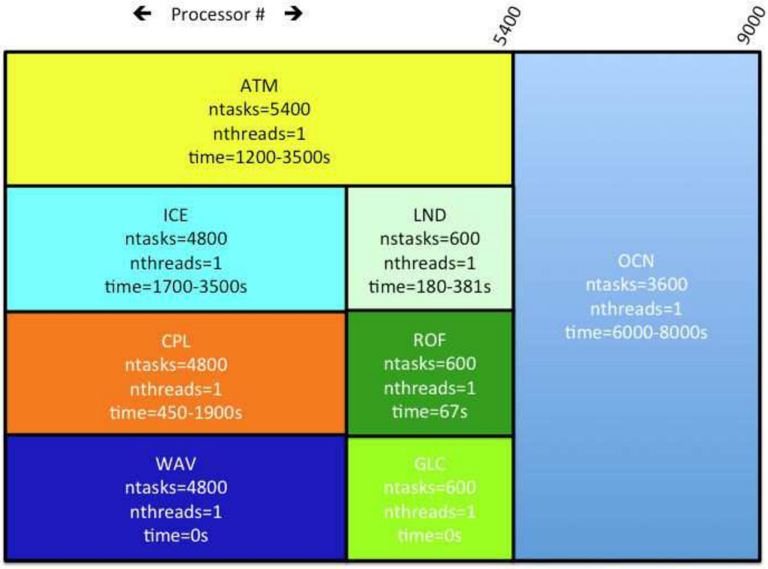
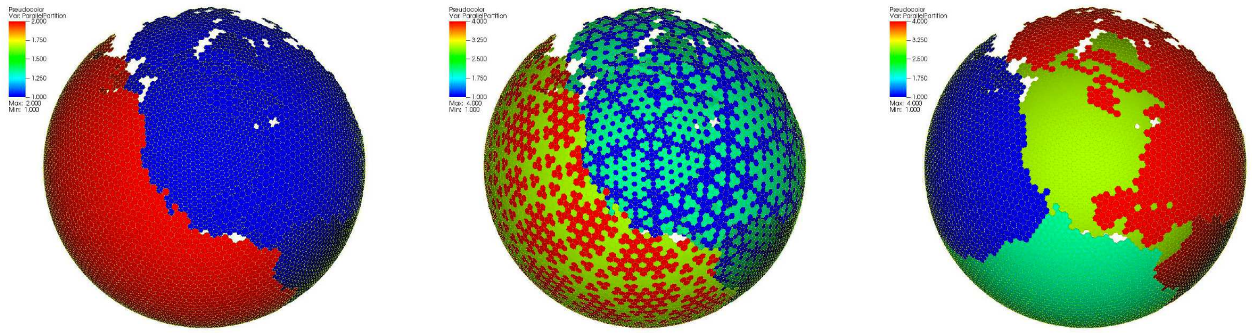


Figure 4. Example E3SM process execution layout for a problem case



(a) Component mesh on 2 tasks (b) Migrated mesh on 4 tasks (Trivial partitioner) (c) Migrated mesh on 4 tasks (Zoltan partitioner)

Figure 5. Migration strategies to repartition from $N_c \rightarrow N_x$

the potential to minimize data movement in the MOAB-based remapper, and to make it a competitive data broker to replace the current MCT Jacob et al. (2005a) coupler in E3SM.

We show an example of a decomposed ocean mesh (polygonal MPAS mesh) that is replicated in a E3SM problem case run on two processes in Fig. 5. Fig. 5-(a) is the original decomposed mesh on 2 tasks $\in N_c$, while Fig. 5-(b) and Fig. 5-(c) show the impact of migrating a mesh from 2 N_c tasks to 4 tasks $\in N_x$ with a trivial linear partitioner and a Zoltan based geometric online

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-280
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 20 November 2018

repartitioner. The decomposition in Fig. 5-(b) shows that the element ID based linear partitioner can produce bad data locality, which may require large number of nearest neighbor communications when computing a source coverage mesh. The resulting communication pattern also makes the migration and coverage computation process non-scalable on larger core counts. In contrast, in Fig. 5-(c), the Zoltan partitioners produce much better load balanced decompositions with Hypergraph (PHG), Recursive Coordinate Bisection (RCB) or Recursive Inertial Bisection (RIB) algorithms to reduce communication overheads in the remapping workflow.

3.3 Computing the Regridding Operator

Standard approaches to compute the intersection of two convex polygonal meshes involve the creation of a Kd-tree or BVH-tree datastructure to enable fast element location of relevant target points. Each target point of interest is located on the source mesh by querying the tree datastructure, and the corresponding (source) element is then marked as a contributor to the remapping weight computation of the target DoF. This process is repeated to form a list of unique source elements that interact directly according to the consistent discretization basis.

3.3.1 Advancing Front Intersection – A Linear Complexity Algorithm

The intersection algorithm used in this paper follows the ideas from Löhner and Parikh (1988); Gander and Japhet (2013), in which two meshes are covering the same domain. At the core is an advancing front method that aims to traverse through the source and target meshes to compute a union (super) mesh. First, two convex cells from the source coverage mesh and the target meshes that intersect are identified by using an adaptive Kd-tree search tree constructed during the setup phase. This also includes determination of the seed for the advancing front. Advancing in both meshes using face adjacency information, incrementally all possible intersections are computed Březina and Exner (2017) accurately to a user defined tolerance (default $= 1e-15$).

While the advancing front algorithm is not restricted to convex cells, the intersection computation is simpler if they are strictly convex. If concave polygons exist in the initial source or target meshes, they are recursively decomposed into simpler convex polygons, by splitting along interior diagonals. Note that the intersection between two convex polygons is results in a strictly convex polygon. Hence, the underlying intersection algorithm remains robust to resolve even arbitrary non-convex meshes covering the same domain space.

Fig. 6 shows how the algorithm advances; Each target cell is resolved by building a local queue of source cells that intersect the target cell. Source cells are added to the local queue incrementally, using adjacency information. At the same time, a global queue with seeds is formed, and it contains pairs of source/target cells that have the potential to intersect. When there are no more source cells in the local queue, the algorithm advances to the next seed from the global queue, and the algorithm repeats.

This flooding-like advancing front needs a stable and robust methodology of intersecting edges/segments in two cells that belong to different meshes. Any pair of segments that intersect can appear in four different pairs of cells. A list of intersection points is maintained on each target edge, so that the intersection points are unique. Also, a geometric tolerance is used to merge intersection points that are close to each other, or if they are proximal to the original vertices in both meshes. Decisions

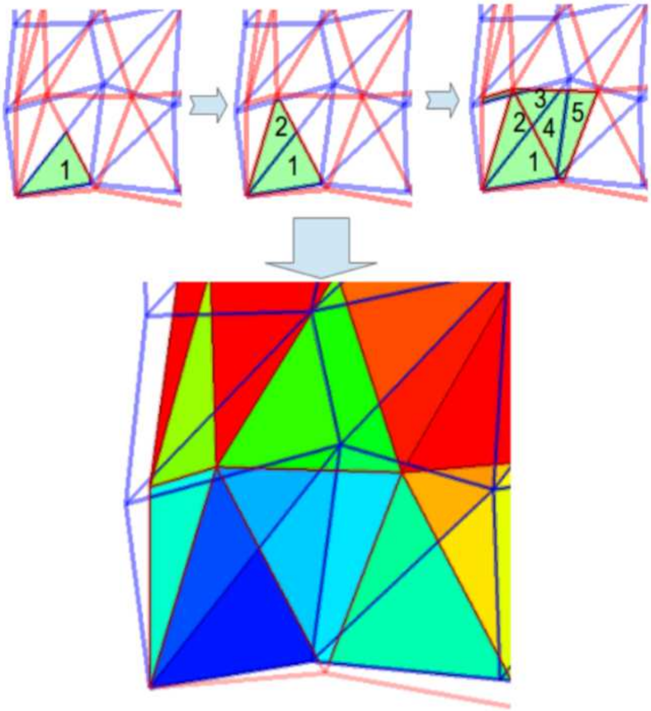


Figure 6. Illustration of the advancing front intersection algorithm.

regarding whether points are inside, outside or at the boundary of a convex enclosure are handled separately. If necessary, more robust techniques such as adaptive precision arithmetic procedures used in *Triangle* Shewchuk (1996), can be employed to resolve the fronts more accurately. Note that the advancing front strategy can be employed for meshes with topological holes (e.g. ocean meshes, in which the continents are excluded) without any further modifications by using a new seed for each disconnected region in the target mesh.

Note on Gnomonic Projection for Spherical Geometry

Meshes that appear in climate applications are often on a sphere. Cell edges are considered to be great circle arcs. A simple gnomonic projection is used to project the edges on one of the six planes parallel to the coordinate axis, and tangent to the sphere Ullrich et al. (2013). With this projection, all curvilinear cells on the sphere are transformed to linear polygons on a gnomonic plane, which simplifies the computation of intersection between multiple grids. Once the intersection points and cells are computed on the gnomonic plane, these are projected back on to the original spherical domain without approximations.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-280
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 20 November 2018

3.4 Parallel Implementation Considerations

Existing Infrastructure from MOAB Tautges et al. (2004) was used to extend the advancing front algorithm in parallel. The expensive intersection computation can be carried out independently, in parallel, once we redistribute the source mesh to envelope the target mesh areas fully, in a step we refer to as ‘source coverage mesh’ computation.

3.4.1 Computation of a Source Coverage Mesh

We select the target mesh as the driver for redistribution of the source mesh. On each task, we first compute the bounding box of the local target mesh. This information is then gathered and communicated to all tasks, and used for redistribution of the local source mesh. Cells that intersect the bounding boxes of other processors are sent to the corresponding owner task. This workflow guarantees that the target mesh on each processor is completely enveloped by the covering mesh repartitioned from its original source mesh decomposition, as shown in Fig. 7. In other words, the covering mesh is a superset of the target mesh in each task. It is important to note that some source coverage cells might be sent to multiple processors during this step, depending on the target mesh resolution and decomposition. The parallel infrastructure in MOAB is heavily leveraged Tautges et al. (2012) to utilize the scalable, crystal router algorithm in order to scalably communicate the covering cells to different processors.

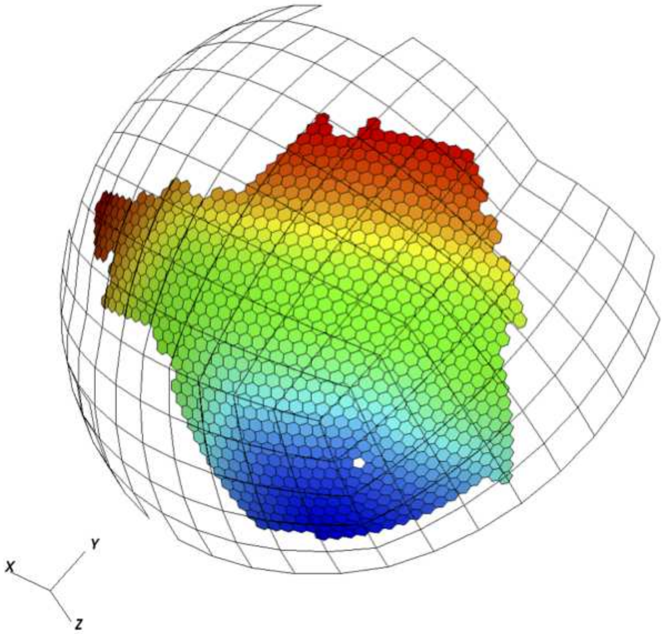


Figure 7. Source coverage mesh fully cover local target mesh; local intersection proceeds between atmosphere (Quadrangle) and ocean (Polygonal) grids.

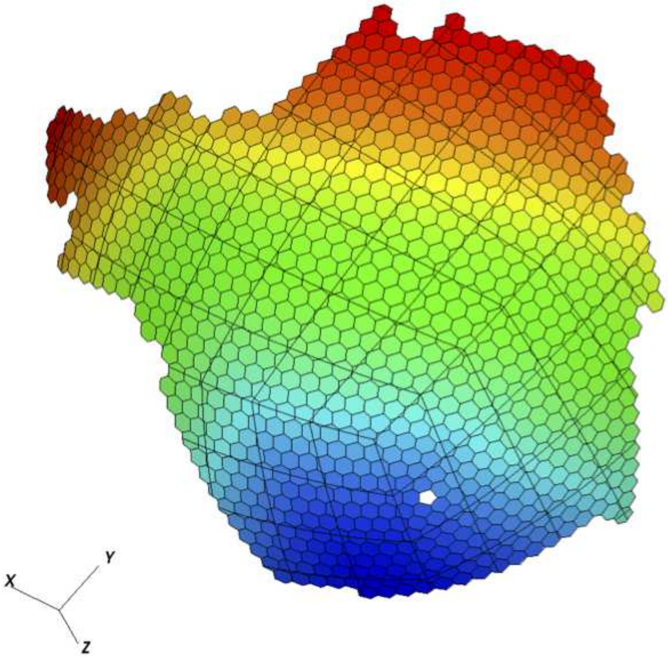


Figure 8. Intersection mesh computed with the coverage and target mesh in a single process.

Once the relevant covering mesh is accumulated locally on each process, the intersection computation can be carried out in parallel, completely independently, using the advancing front algorithm (Section. (3.3.1)), as shown in Fig. 8. Once each task computes its share of intersection polygons, the intersection vertices on the shared edges between processes needs to be communicated to avoid duplication. In order to ensure consistent local conservation constraints in the weight matrix in the parallel setting, there might be additional communication of ghost intersection elements to nearest neighbors. This extra communication step is only required for computing interpolators for flux variables and can be avoided when transferring scalar fields with non-conservative bilinear or higher-order interpolations in this workflow.

The parallel advancing front algorithm presented here to globally compute the intersection supermesh can be extended to expose finer grained parallelism with task-based execution models, where each task handles a unique front in the computation queue. With a local mesh decomposition with Metis or through coloring, each task can then proceed to compute the intersection elements until the front collides with another from a different thread, until all the overlap elements have been computed in each process. Such a hybrid MPI and threading algorithm has the potential to scale well even on heterogeneous architectures and provides options to improve the computational throughput of the regridding process Löhner (2014).

3.5 Computation of Remapping operator with TempestRemap

For illustration, consider a scalar field $\boldsymbol{U}$ discretized with standard Galerkin FEM on source Ω_1 and target Ω_2 meshes with different resolutions. The projection of the scalar field on the target grid is in general given as follows.

$$\boldsymbol{U}_2(\boldsymbol{\Omega_2}) = \boldsymbol{\Pi}_1^2 \boldsymbol{U}_1(\boldsymbol{\Omega_1}) \tag{1}$$

where, $\boldsymbol{\Pi}_1^2$ is the discrete solution interpolator of $\boldsymbol{U}$ defined on $\boldsymbol{\Omega_1}$ to $\boldsymbol{\Omega_2}$. This interpolator $\boldsymbol{\Pi}_1^2$ in Eq. (1) is often referred to as the remapping operator, which is pre-computed in the coupled climate workflows using ESMF and TempestRemap. For embedded meshes, the remapping operator can be calculated exactly as a restriction or prolongation from the source to target grid. However, for general unstructured meshes and in cases where the source and target meshes are topologically different, the numerical integration to assemble $\boldsymbol{\Pi}_1^2$ needs to be carried out on the supermesh Ullrich and Taylor (2015). Since a unique source and target parent element exists for every intersection element belonging to the supermesh $\boldsymbol{\Omega_1} \bigcup \boldsymbol{\Omega_2}$, $\boldsymbol{\Pi}_1^2$ is assembled as the sum of local mass matrix contributions on the intersection elements, by using the consistent discretization basis for the source and target field descriptions Ullrich et al. (2016). The intersection mesh typically contains arbitrary convex polygons and hence subsequent triangulation may be necessary before evaluating the integration. This global linear operator directly couples source and target DoFs based on the participating intersection element parents Ullrich et al. (2009).

MOAB supports point-wise FEM interpolation (bilinear and higher-order spectral) with local or global subset normalization Tautges and Caceres (2009), in addition to a conservative first-order remapping scheme. But higher order conservative monotone weight computations are currently unsupported natively. To fill this gap for climate applications, and to leverage existing developments in rigorous numerical algorithms to compute the conservative weights, interfaces to TempestRemap in MOAB were added to scalably compute the remap operator in parallel, without sacrificing field discretization accuracy. The MOAB interface to the E3SM component models provides access to the underlying type and order of field discretization, along with the global partitioning for the DoF numbering. Hence the projection or the weight matrix can be assembled in parallel by traversing through the intersection elements, and associating the appropriate source and target DoF parent to columns and rows respectively. The MOAB implementation uses a sparse matrix representation using the Eigen3 library to store the local weight matrix. Except for the particular case of projection onto a target grid with cGLL description, the matrix rows do not share any contributions from the same source DoFs. This implies that for FV and dGLL target field descriptions, the application of the weight matrix does not require global collective operations and sparse matrix vector applications scale ideally (still memory bandwidth limited). In the cGLL case, we perform a reduction of the parallel vector along the shared DoFs to accumulate contributions exactly. However, it is non-trivial to ensure full bit-for-bit reproducibility during such reductions.

It is also possible to use the transpose of the remapping operator computed between a particular source and target component combination, to project the solution back to the original source grid. Such an operation has the advantage of preserving the consistency and conservation metrics originally imposed in finding the remapping operator and reduces computation cost by avoiding recomputation of the weight matrix for the new directional pair. For example, when computing the remap operator between atmosphere and ocean models (with holes), it is advantageous to use the atmosphere model as the source grid, since the advancing front seed computation may require multiple iterations if the front begins within a hole. Additionally, such transpose vector applications can also make the global coupling symmetric, which may have favorable implications when pursuing implicit temporal integration schemes.

3.6 Note on MBTR Remapper Implementation

The remapping algorithms presented in the previous section are exposed through a combination of implementations in MOAB and TempestRemap libraries. Since both the libraries are written in C++, direct inheritance of key datastructures such as the GridElements (mesh) and OfflineMap (projection weights) are available to minimize data movement between the libraries. Additionally, Fortran codes such as E3SM can invoke computations of the intersection mesh and the remapping weights through specialized language-agnostic interfaces in MOAB: `iMOAB` Mahadevan et al. (2015). These interfaces offer the flexibility to query, manipulate and transfer the mesh between groups of processes that represent the component and coupler processing elements.

Using the iMOAB interfaces, the E3SM coupler can coordinate the online remapping workflow during the setup phase of the simulation, and compute the projection operators for component and scalar or vector coupled field combinations. For each pair of coupled components, the following sequence of steps are then executed to consistently compute the remapping operator and transfer the solution fields in parallel.

1. `iMOAB_SendMesh` and `iMOAB_ReceiveMesh`: Send the component mesh (defined on $N_{c,l}$ processes), and receive the complete unstructured mesh copy in the coupler processes (N_x). This mesh migration undergoes an online mesh repartition either through a trivial decomposition scheme or with advanced Zoltan algorithms (geometric or graph partitioners)

2. `iMOAB_ComputeMeshIntersectionOnSphere`: The advancing front intersection scheme is invoked to compute the overlap mesh in the coupler processes

3. `iMOAB_CoverageGraph`: Update the parallel communication graph based on the (source) coverage mesh association in each process

4. `iMOAB_ComputeScalarProjectionWeights`: The remapping weight operator is computed and assembled with discretization-specific (FV, SE) calls to TempestRemap, and stored in Eigen3 SparseMatrix object

Once the remapping operator is serialized in-memory for each coupled scalar and flux fields, this operator is then used at every timestep to compute the actual projection of the data.

1. `iMOAB_SendElementTag` and `iMOAB_ReceiveElementTag`: Using the coverage graph computed previously, direct one-to-one communication of the field data is enabled between $N_{c,l}$ and N_x, before and after application of the weight operator

2. `iMOAB_ApplyScalarProjectionWeights`: In order to compute the field interpolation or projection from the source component to the target component, a matvec product of the weight matrix and the field vector defined on the source grid is performed. The source field vector is received from source processes $N_{c,s}$ and after weight application, the target field vector is sent to target processes $N_{c,l}$

Additionally, to facilitate offline generation of projection weights, a MOAB based parallel tool `mbtempest` has been written in C++, similar to ESMF and TempestRemap (serial) standalone tools. `mptempst` can load the source and target meshes from files, in parallel, and compute the intersection and remapping weights through TempestRemap. The weights can then be written back to a SCRIP-compatible file format, for any of the supported field discretization combinations in source and destination components. Added capability to apply the weight matrix onto the source solution field vectors, and native visualization plugins in VisIt for MOAB, simplify the verification of conservation and monotonicity for complex remapping workflows.

4 Results

Evaluating the performance of the in-memory, MOAB-TempestRemap (MBTR) remapping infrastructure requires recursive profiling and optimization to ensure scalability for large-scale simulations. In order to showcase the advantage of using the mesh-aware MOAB datastructure as the MCT coupler replacement, we need to understand the per task performance of the regridder in addition to the parallel point locator scalability, and overall time for remapping weight computation. Note that except for the weight application for each solution field from a source grid to a target grid, the in-memory copy of the component meshes, migration to coupler PEs, computation of intersection elements and remapping weights are done only once during the setup phase in E3SM, per coupled component model pair.

4.1 Serial Performance

We compare the total cost for computing the supermesh and the remapping weights for several source and target grid combinations through three different methods to determine the serial computational complexity.

1. ESMF: Kd-tree based regridder and weight generation for first/second order FV→FV conservative remapping
2. TempestRemap: Kd-tree based supermesh generation and conservative, monotonic, high-order remap operator for FV→FV, SE→FV, SE→SE projection
3. MBTempest: Advancing front intersection with MOAB and conservative weight generation with TempestRemap interfaces

Fig. 9 shows the serial performance of the remappers for computing the conservative interpolator from Cubed-Sphere grids to polygonal MPAS grids of different resolutions for a FV→FV field transfer. This total time includes the computation of intersection mesh or supermesh, in addition to the remapping weights with field conservation specifications. These serial runs were executed on a machine with 8x Intel Xeon(R) CPU E7-4820 @ 2.00GHz (total of 64 cores) and 1.47 TB of RAM. As the

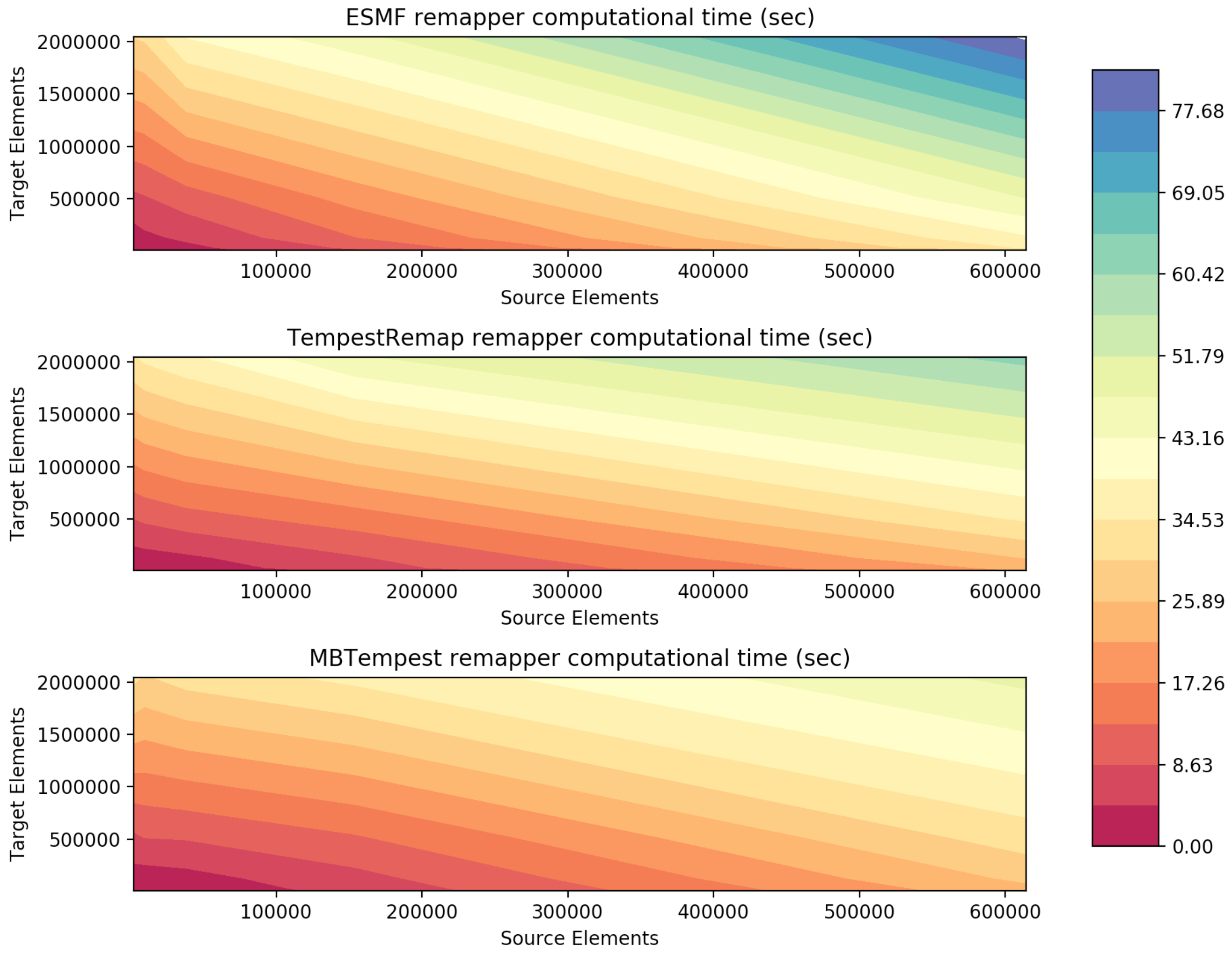


Figure 9. Comparison of serial regridding computation (supermesh and projection weight generation) between ESMF, TempestRemap, and MBTempest

source grid resolution increases, the advancing front intersection with linear complexity outperforms the Kd-tree intersection algorithms used by TempestRemap and ESMF. The time spent in the remapping task, including the overlap mesh generation, provides an overall metric on the single task performance when memory bandwidth or communication concerns do not dominate in a parallel run. In this comparison with three remapping software libraries, the total computational time in the fine resolution limit as $\frac{nele(source)}{nele(target)} \approx 1$ consistently increases (going diagonally from left to right in Fig. 9). We note that the serial version of TempestRemap is comparable to ESMF and can even provide better timings on the highly refined cases, while the MBTempest remapper consistently outperforms both the tools, with a 2x speedup on average. The relatively better performance in MBTempest is accomplished through the linear complexity advancing front algorithm, which further offers avenues to incorporate task level parallelism to accelerate the on-node performance on multicore and GPGPU hybrid architectures.

4.2 Scalability of the MOAB Kd-tree Point Locator

In addition to being able to compute the supermesh between Ω_S and Ω_T, MOAB also offers datastructures to query source elements containing points that correspond to the target DoFs locations. This operation is critical in evaluating bilinear and biquadratic interpolator approximations for scalar variables when conservative projection is not required by the underlying coupled model. The solution interpolation for the multi-mesh case involves two distinct phases.

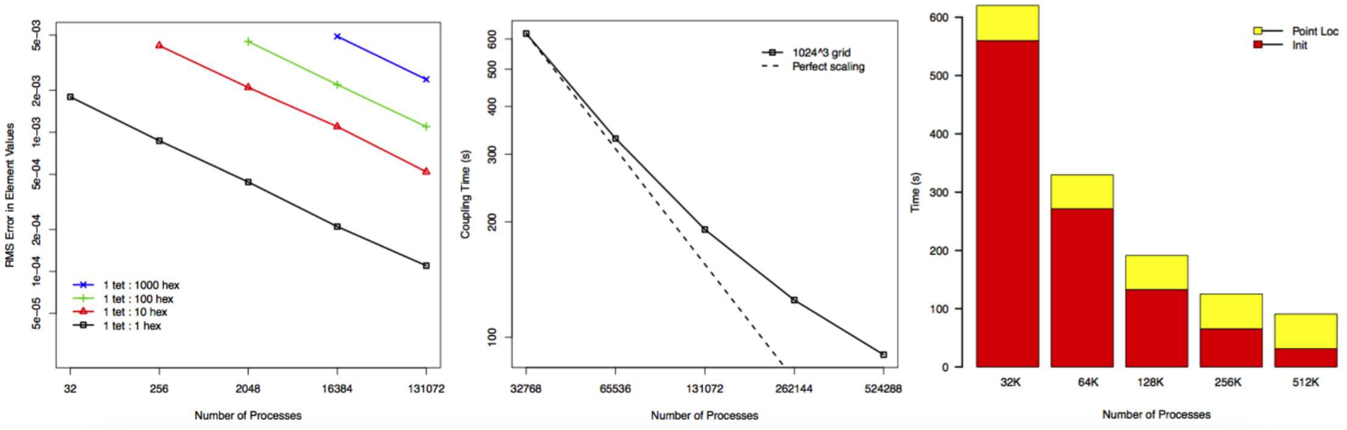
1. Setup phase: Use Kd-tree to build the search datastructure to locate points corresponding to vertices in the target mesh on the source mesh
2. Run phase: Use the elements containing the located points to compute consistent interpolation onto target mesh vertices

Studies were performed on the BlueGene-Q machine (Mira) at ANL to evaluate the strong and weak scalability of the parallel Kd-tree point search implementation in MOAB. The scalability results were generated with the CIAN2 coupling mini-app Morozov and Peterka (2016), which links to MOAB to handle traversal of the unstructured grids and transfer of solution fields between the grids. For this case, a series of hexahedral and tetrahedral meshes were used to interpolate an analytical solution. By changing the basis interpolation order, and mesh resolutions, the convergence of the interpolator was verified to provide theoretical accuracy orders of convergence in the asymptotic fine limit.

The performance tests were executed on the IBM BlueGene/Q Mira at 16 MPI ranks per node, with 2GB RAM per MPI rank, at up to 500K MPI processes. The strong scaling results and error convergence were computed with a grid size of 1024^3. The solution interpolation on varying mesh resolutions were performed by projecting an analytical solution from a Tetrahedral→Hexahedral→Tetrahedral grid, with total number of points/rank varied between [2K, 32K] in the study.

Fig. 10 shows a strong scaling efficiency of around 50% is achieved on a maximum of 512K cores (66% of Mira). We note that the computational complexity of the Kd-tree data structure scales as $O(nlog(n))$ asymptotically, and the point location phase during initial search setup dominates the total cost on higher core counts. This is evident in the timing breakdown for each phase shown in Fig. 10-(c). Since the point location is performed only once during simulation startup, while the interpolation is performed multiple times per timestep during the run, we expect the total cost of the projection for scalar variables to be amortized over transient climate simulations with fixed grids. Further investigations with optimal BVH-tree Larsen et al. (1999) or R-tree implementations for these interpolation cases could help reduce the overall cost.

Currently, only the NC bilinear or biquadratic interpolation of scalar fields with subset normalization Tautges and Caceres (2009) is supported directly in MOAB (via Kd-tree point location and interpolation), and advancing front intersection algorithm does not make use of these data-structures. In contrast, TempestRemap and ESMF use a Kd-tree search to not only compute the location of points, but also to evaluate the supermesh $\Omega_S \bigcup \Omega_T$, and hence the computational complexity for the intersection mesh determination scales as $O(nlog(n))$, in contrast to the linear complexity ($O(n)$) of the advancing front intersection algorithm implemented in MOAB.



(a) Tet→Hex RMS Error

(b) Strong Scaling

(c) Strong Scaling by Component

Figure 10. MOAB 3-d Kd-tree Point Location: Strong scaling on Mira (BG/Q)

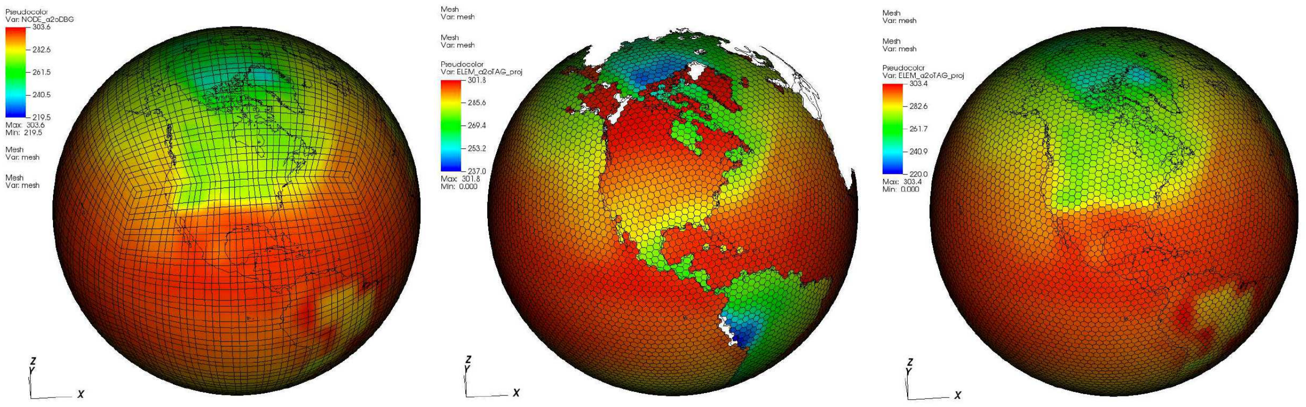
4.3 The Parallel MBTR Remapping Algorithm

The MBTR online weight generation workflow within E3SM was employed to verify and test the projection of real simulation data generated during the coupled atmosphere-ocean model runs. A choice was made to use the model-computed temperature on the lowest level of the atmosphere, since the heat fluxes that nonlinearly couples the atmosphere and ocean models are directly proportional to this interface temperature field. By convention, the fluxes are computed on the ocean mesh, and hence the atmosphere temperature must be interpolated onto MPAS polygonal mesh. We use this scenario as a test case for demonstrating the strong scalability results in this section.

The NE11 (with approximately 4 degree grid size) atmosphere run in E3SM, and the projection of its lowest level temperature onto two different MPAS meshes (with approximate grid size of 240km) are shown in Fig. 11. The conservative projection from SE→FV on a mesh with holes (Fig. 11-(b)) and without holes (Fig. 11-(c)) corresponding to land regions, is presented here to show the difference in the remapped solutions.

4.3.1 Scaling Comparison of Conservative Remappers (FV→FV)

The strong scaling studies for computation of remapping weights to project a FV solution field between CS grids of varying resolutions was performed on the Blues large-scale cluster (with 16 Sandy Bridge Xeon E5-2670 2.6GHz cores and 32 GB RAM per node) at ANL, and the Cori supercomputer at NERSC (with 64 Haswell Xeon E5-2698v3 2.3GHz cores and 128 GB RAM per node). Fig. 12 shows that the MBTR workflow consistently outperforms ESMF on both the machines as the number of processes used by the coupler is increased. The timings shown here represent the total remapping time i.e., cumulative computational time for generating the super mesh and the (conservative) remapping weights.



(a) Original SE field on CS atmosphere mesh (b) Remapped field on MPAS mesh with holes (c) Remapped field on MPAS mesh without holes

Figure 11. Projection of the NE11 SE bottom atmospheric temperature field onto the MPAS ocean grid

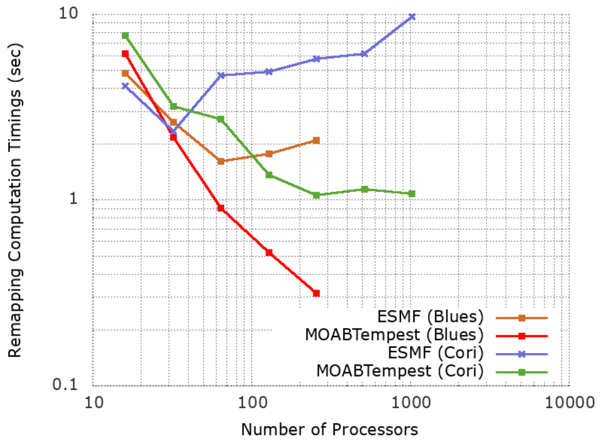


Figure 12. CS (E=614400 quads) → CS (E=153600 quads) remapping (-m conserve) on LCRC/ALCF and NERSC machines

The relatively better scaling for MOAB on the Blues cluster is due to faster hardware and memory bandwidth compared to the Cori machine. The strong scaling efficiency approaches a plateau on Cori Haswell nodes as communication costs for the coverage mesh computation start dominating the overall remapping processes, especially in the limit of $\frac{nele}{process} \rightarrow 1$ at large node counts.

4.3.2 Strong Scalability of Spectral Projection (SE→FV)

To further evaluate the characteristics of in-memory remapping computation, along with cost of application of the weights during a transient simulation, a series of further studies were executed to determine the spectral projection of a real dataset between atmosphere and ocean components in E3SM. The source mesh contains 4th order spectral element temperature data defined on Gauss-Lobatto quadrature nodes (cGLL discretization) of the CS mesh, and the projection is performed on a MPAS polygonal mesh with holes (FV discretization). A direct comparison to ESMF was unfeasible in this study since the traditional workflow requires the computation of a dual mesh transformation of the spectral grid. Hence, only timings for MBTR workflow is shown here.

Two specific cases were considered for this SE→FV strong scaling study with conservation and monotonicity constraints.

1. **Case A (NE30):** 1-degree CS SE mesh (nele=5400 quads) with $p = 4$ to MPAS mesh (nele=235160 polygons)
2. **Case B (NE120):** 0.25-degree CS SE mesh (nele=86400 quads) with $p = 4$ to MPAS mesh (nele=3693225 polygons)

The performance tests for each of these cases were launched with three different process execution layouts for the atmosphere, ocean components and the coupler.

(a) Fully colocated PE layout: $N_{atm} = N_x$ and $N_{ocn} = N_x$

(b) Disjoint-ATM model PE layout: $N_{atm} = N_x/2$ and $N_{ocn} = N_x$

(c) Disjoint-OCN model PE layout: $N_{atm} = N_x$ and $N_{ocn} = N_x/2$

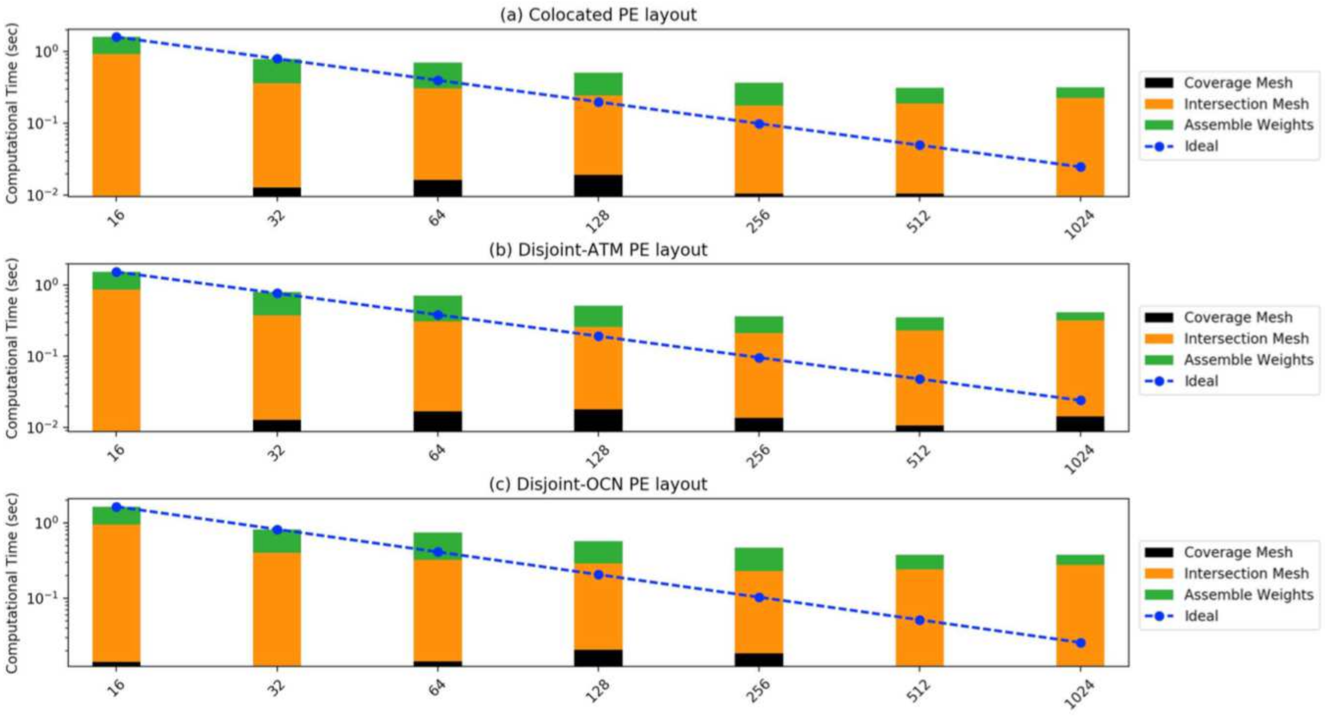
Table 1. Strong scaling on Cori for SE→FV projection with two different resolutions

Number of processors	**Case A**		**Case B**	
	Intersection (sec)	**Compute Weights (sec)**	**Intersection (sec)**	**Compute Weights (sec)**
16	0.936846	0.64983	145.623	9.732
32	0.449022	0.429028	53.1244	5.78093
64	0.377767	0.373476	22.7167	4.92151
128	0.255154	0.270574	6.70485	2.79397
256	0.180136	0.18272	2.26435	1.71835
512	0.162388	0.104737	1.25471	0.928622
1024	0.203354	0.0932475	0.680122	0.618943

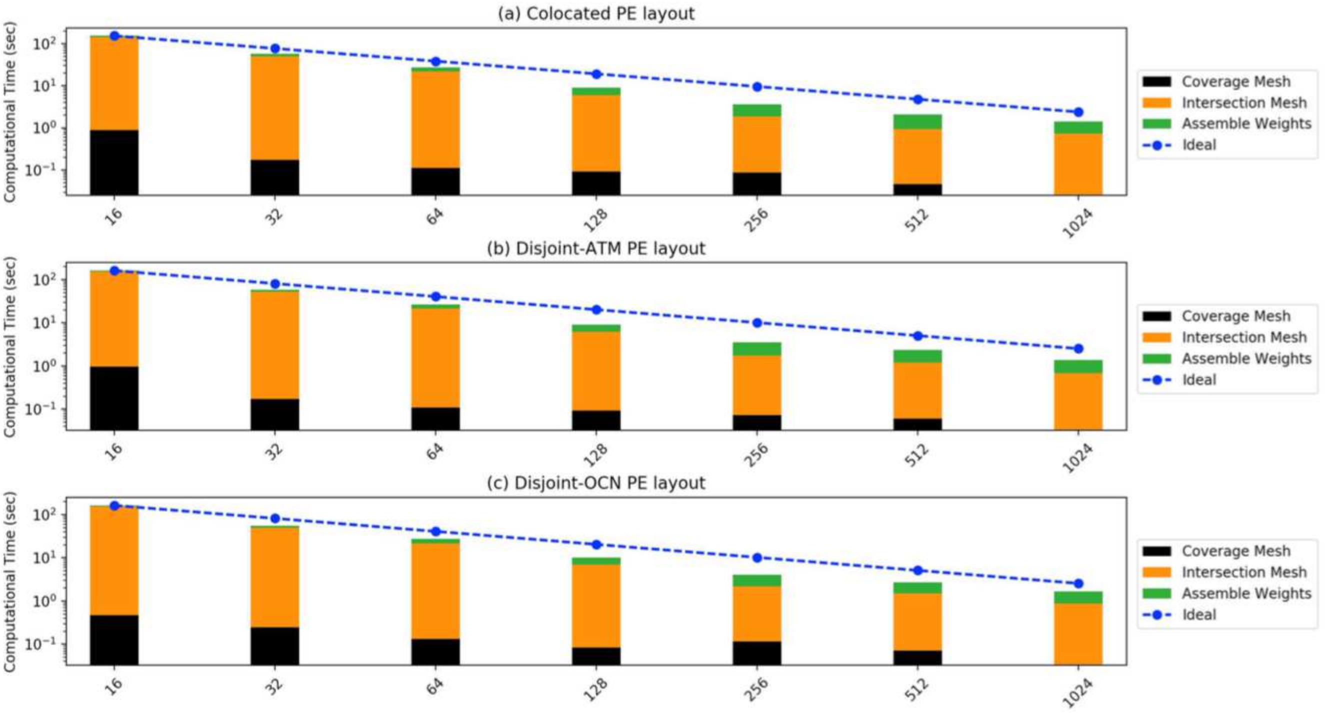
A breakdown of computational time for key tasks on Cori with up to 1024 processes for both the cases is tabulated in Table 1 on a fully colocated decomposition i.e., $N_{ocn} = N_{atm} = N_x$. It is clear that the computation of parallel intersection mesh strong scales well for these production cases, especially for larger mesh resolutions (Case B). For the smaller source and target mesh resolution (Case A), we notice that the intersection time hits a lower bound that is dominated by the computation of the coverage mesh to enclose the target mesh in each task. It is important to stress that this one time setup call to compute remap operator, per component pair, is relatively much cheaper compared to individual component and solver initializations and get amortized over longer transient simulations.

In comparison to the computation of the intersection mesh, the time to assemble the remapping weight operator in parallel is generally smaller. Even though both of these operations are performed only once during the setup phase of the E3SM simulation, the weight operator computation involves several validation checks that utilize collective MPI operations, which do destroy the embarrassingly parallel nature of the calculation, once appropriate coverage mesh is determined in each task.

The component-wise breakdown for the advancing front intersection mesh, the parallel communication graph for sending and receiving data between component and coupler, and finally, the remapping weight generation for the SE→FV setup for NE30 and NE120 cases are shown in Fig. 13. The cumulative time for this remapping process is shown to scale linearly for NE120 case, even if the parallel efficiency decreases significantly in the NE30 case, as expected based on the results in Table 1. Note that the MBTR workflow provides a unique capability to consistently and accurately compute SE→FV projection weights in parallel, without any need for an external pre-processing step to compute the dual mesh (as required by ESMF) or running the entire remapping process in serial (TempestRemap).



(a) NE30 component-wise strong scaling

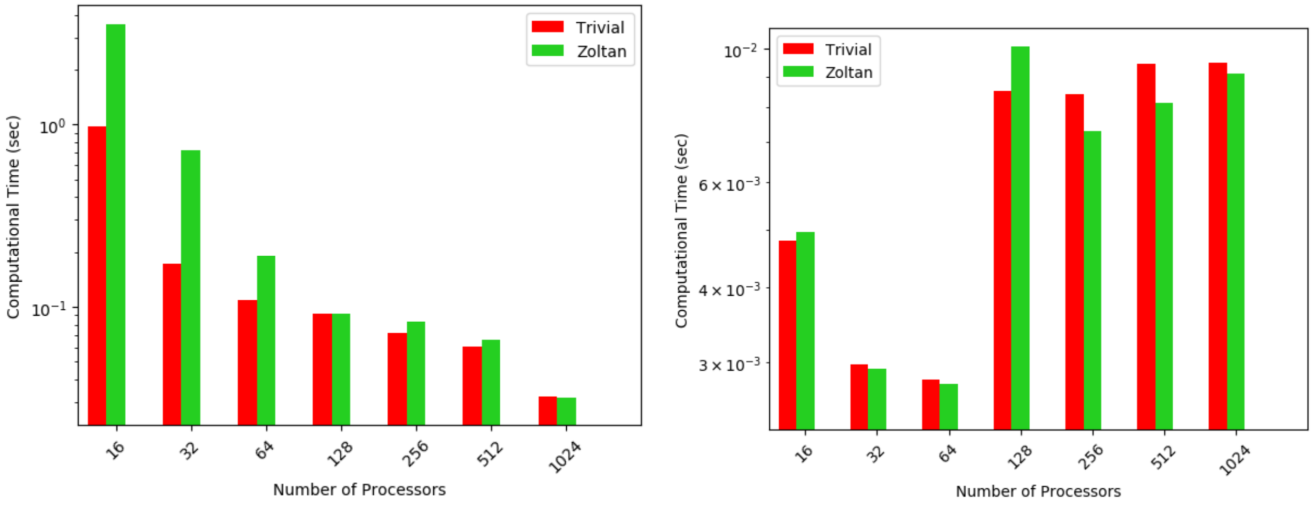


(b) NE120 component-wise strong scaling

Figure 13. Strong scaling study for the NE30 and NE120 cases for spectral projection with Zoltan repartitioner

4.4 Effect of partitioning strategy

In order to determine the effect of partitioning strategies described in Fig. 5, the NE120 case with the trivial decomposition and Zoltan geometric partitioner (RCB) were tested in parallel. Fig. 14 compares the two strategies for optimizing the mesh migration from the component to coupler. These strategies play a critical role in task mapping and data locality for the source coverage mesh computation, in addition to determining the communication graph complexity between the components and the coupler. This comparison highlights that the coverage mesh cost reduces uniformly at scale, while the trivial partitioning scheme behaves better on lower core counts. The communication of field data between the atmosphere component and the coupler resulting from the partitioning strategy is a critical operation during the transient simulation, and generally stays within network latency limits in Cori, as the message size reduces. Eventhough the communication kernel does not show good scaling on increasing node counts, the relative cost of the operation is insignificant in comparison to total time spent in individual component solvers. Note that production climate model solvers require multiple data fields to be remapped at every rendezvous timestep, and hence the size of the packed messages may be larger for such simulations (volume should remain similar to Fig. 14-(b)).



(a) Source coverage mesh computation

(b) Send/Receive field data between N_x and N_{atm}

Figure 14. Scaling of the communication kernels driven with the parallel graph computed with a trivial redistribution and the Zoltan geometric (RCB) repartitioner for the NE120 case with $N_{ocn} = N_x$ and $N_{atm} = N_x/2$

4.5 Note on Application of Weights

Generally, operations involving Sparse Matrix-Vector (SpMV) products are memory bandwidth limited Bell and Garland (2009), and occur during the application of remapping weights operator on to the source solution field vector, in order to compute the field projection onto the target grid. In addition to the communication of field data shown in Fig. 14-(b), the cost of remapping weight application in parallel (presented in Fig. 15) determines the total cost of the remapping operation during runtime. Except for the case of cGLL target discretizations, the parallel SpMV operation during the weight application do not involve any global collective reductions. In the current E3SM and OASIS3-MCT workflow, these operations are handled by the MCT library. In high resolution simulations of E3SM, the total time for the remapping operation in MCT is primarily dominated by the communication costs based on the communication graph, similar to the MBTR workflow. However, a direct comparison between these two workflows is not yet possible, but we expect the aggregated communication strategies in the crystal router algorithm Fox et al. (1989) in MOAB, to provide relatively better performance at scale.

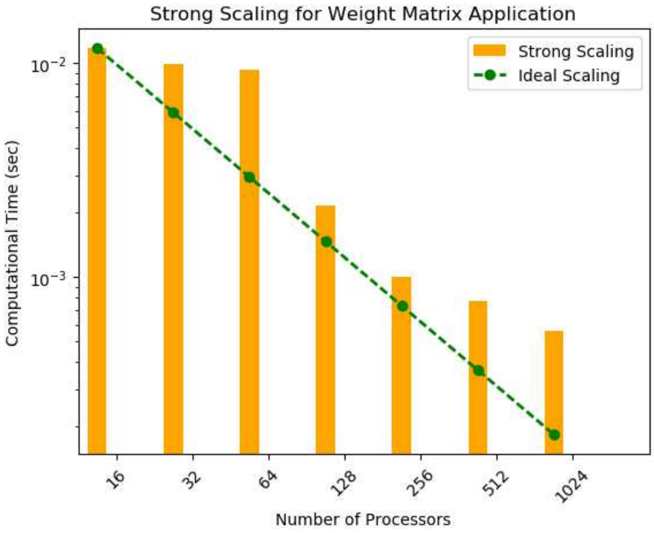


Figure 15. SE→FV remapping weight operator application

5 Conclusion

Understanding and controlling primary sources of errors in a coupled system dynamically, will be key to achieving predictable and verifiable climate simulations on emerging architectures. Traditionally, the computational workflow for coupled climate simulations has involved two distinct steps, with an offline pre-processing phase using remapping tools to generate solution field projection weights (ESMF, TempestRemap, SCRIP), which is then consumed by the coupler to transfer field data between the component grids.

The offline steps include generating grid description files and running the offline tools with the problem-specific options. Additionally many of state-of-science tools such as ESMF and SCRIP require additional steps to specially handle interpolators from SE grids. Such workflows create bottlenecks that do not scale, and can inhibit scientific research productivity. When experimenting with refined grids, a goal for E3SM, this tool chain has to excercised repeatedly. Additionally, when component meshes are dynamically modified, either through mesh adaptivity or dynamical mesh movement to track moving boundaries, the underlying remapping weights must be recomputed on the fly.

To overcome some of these limitations, we have presented scalable algorithms and software interfaces to create a direct component coupling with online regridding and weight generation tools. The remapping algorithms utilize the numerics exposed by TempestRemap, and leverage the parallel mesh handling infrastructure in MOAB to create a scalable in-memory remapping infrastructure that can be integrated with existing coupled climate solvers. Such a methodology invalidates the need for dual grids, preserves higher-order spectral accuracy, and locally conserves the field data, in addition to monotonicity constraints, when transferring solutions between grids with non-matching resolutions.

The serial and parallel performance of the MOAB advancing front intersection algorithm with linear complexity ($O(n)$) was demonstrated for a variety of source and target mesh resolution combinations, and compared with the current state-of-science regridding tools such as ESMF (serial/parallel) and TempestRemap (serial) that have a $O(nlog(n))$ complexity using the Kd-tree datastructure. The MOAB-TempestRemap (MBTR) software infrastructure yields a balance of both the scalable performance on emerging architectures without sacrificing discretization accuracy for component field interpolators. There are also several optimizations in the MBTR algorithms that can be implemented to improve finer-grained parallelism on hybrid architectures and to minimize data movement with better partitioning strategies. Such a software infrastructure provides a foundation to build a new coupler to replace the current offline-online, hub-and-spoke MCT-based coupler in E3SM, and offer extensions to enable a fully distributed coupling paradigm (without the need for a centralized coupler) to minimize computational bottlenecks in a task-based workflow.

Code availability. Information on the availability of source code for the algorithmic infrastructure and models featured in this paper is tabulated below.

Short name	Code availability
E3SM	E3SM Project (2018) is under active development funded by the US Department of Energy. E3SM version 1.1 has been publicly released under an open-source 3-clause BSD license in August 2018, and available at GitHub.
MOAB	MOAB Tautges et al. (2004) is an open-source library under the umbrella of the SIGMA toolkit (2014) Mahadevan et al. (2015), and is publicly available under the Lesser GNU Public License (v3) on BitBucket.
TempestRemap	The TempestRemap Ullrich and Taylor (2015); Ullrich et al. (2016) source code is available under a BSD open-source license and hosted in GitHub.

Video supplement. The video supplements for the serial and parallel advancing front mesh intersection algorithm to compute the supermesh ($\mathbf{\Omega_S \bigcup \Omega_T}$) of a source ($\mathbf{\Omega_S}$)and target ($\mathbf{\Omega_T}$) grid is demonstrated.

Short name	Video description and availability
Serial advancing front mesh intersection	Intersection between CS and MPAS grids on a single task is illustrated. DOI:10.6084/m9.figshare.7294901
Parallel advancing front mesh intersection	Simultaneous parallel Intersection between CS and MPAS grids on two different tasks are illustrated side by side. DOI:10.6084/m9.figshare.7294919

Author contributions. VM and RJ wrote the paper (with comments from IG and JS). VM and IG designed and implemented the MOAB integration with TempestRemap library, along with exposing the necessary infrastructure for online remapping through iMOAB interfaces. IG and JS configured the MOAB-TempestRemap remapper within E3SM, and verified weight generation to transfer solution fields between atmosphere and ocean component models. VM conducted numerical verification studies and executed both the serial and parallel scalability studies on Blues and Cori LCF machines to quantify performance characteristics of the remapping algorithms. The broader project idea was conceived by Andy Salinger (SNL), RJ, VM, and IG.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Support for this work was provided by the Climate Model Development and Validation - Software Modernization project, and partially by the SciDAC Coupling Approaches for Next Generation Architectures (CANGA) project, which are funded by the US Department of Energy (DOE), Office of Science Biological and Environmental Research Program. CANGA is also funded by the DOE Office of Advanced Scientific Computing Research. This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357, and resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We gratefully acknowledge the computing resources provided on Blues, a high-performance computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory. We would also like to thank Dr. Paul Ullrich at University of California, Davis for several helpful discussions regarding remapping schemes and, in particular, implementations in TempestRemap.

References

OASIS3-MCT_4.0 official release, https://portal.enes.org/oasis/news/oasis3-mct_4-0-official-release, [Online; accessed 10-October-2018], 2018.

Aguerre, H. J., Damián, S. M., Gimenez, J. M., and Nigro, N. M.: Conservative handling of arbitrary non-conformal interfaces using an efficient supermesh, Journal of Computational Physics, 335, 21–49, 2017.

Beljaars, A., Dutra, E., Balsamo, G., and Lemarié, F.: On the numerical stability of surface-atmosphere coupling in weather and climate models, Geoscientific Model Development Discussions, 10, 977–989, 2017.

Bell, N. and Garland, M.: Implementing sparse matrix-vector multiplication on throughput-oriented processors, in: Proceedings of the conference on high performance computing networking, storage and analysis, p. 18, ACM, 2009.

Berger, M. J.: On conservation at grid interfaces, SIAM journal on numerical analysis, 24, 967–984, 1987.

Blanco, J. L. and Rai, P. K.: nanoflann: a C++ header-only fork of FLANN, a library for Nearest Neighbor (NN) wih KD-trees, https://github.com/jlblancoc/nanoflann, 2014.

Březina, J. and Exner, P.: Fast algorithms for intersection of non-matching grids using Plücker coordinates, Computers & Mathematics with Applications, 74, 174 – 187, https://doi.org/https://doi.org/10.1016/j.camwa.2017.01.028, http://www.sciencedirect.com/science/article/pii/S0898122117300792, 5th European Seminar on Computing ESCO 2016, 2017.

Certik, O., Ferenbaugh, C., Garimella, R., Herring, A., Jean, B., Malone, C., and Sewell, C.: A Flexible Conservative Remapping Framework for Exascale Computing, https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-17-21749, sIAM Conference on Computational Science & Engineering Feb 2017, 2017.

Collins, N., Theurich, G., Deluca, C., Suarez, M., Trayanov, A., Balaji, V., Li, P., Yang, W., Hill, C., and Da Silva, A.: Design and implementation of components in the Earth System Modeling Framework, The International Journal of High Performance Computing Applications, 19, 341–350, 2005.

Craig, A., Valcke, S., and Coquart, L.: Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0, Geoscientific Model Development, 10, 3297–3308, https://doi.org/10.5194/gmd-10-3297-2017, https://www.geosci-model-dev.net/10/3297/2017/, 2017.

Craig, A. P., Jacob, R., Kauffman, B., Bettge, T., Larson, J., Ong, E., Ding, C., and He, Y.: CPL6: The New Extensible, High Performance Parallel Coupler for the Community Climate System Model, The International Journal of High Performance Computing Applications, 19, 309–327, https://doi.org/10.1177/1094342005056117, https://doi.org/10.1177/1094342005056117, 2005.

Craig, A. P., Vertenstein, M., and Jacob, R.: A new flexible coupler for earth system modeling developed for CCSM4 and CESM1, The International Journal of High Performance Computing Applications, 26, 31–42, https://doi.org/10.1177/1094342011428141, https://doi.org/10.1177/1094342011428141, 2012.

de Boer, A., van Zuijlen, A., and Bijl, H.: Comparison of conservative and consistent approaches for the coupling of non-matching meshes, Computer Methods in Applied Mechanics and Engineering, 197, 4284–4297, https://doi.org/10.1016/j.cma.2008.05.001, http://dx.doi.org/10.1016/j.cma.2008.05.001, 2008.

Dennis, J. M., Edwards, J., Evans, K. J., Guba, O., Lauritzen, P. H., Mirin, A. A., St-Cyr, A., Taylor, M. A., and Worley, P. H.: CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model, The International Journal of High Performance Computing Applications, 26, 74–89, 2012.

Dunlap, R., Rugaber, S., and Mark, L.: A feature model of coupling technologies for Earth System Models, Computers & geosciences, 53, 13–20, 2013.

E3SM Project: Energy Exascale Earth System Model (E3SM), [Computer Software] https://dx.doi.org/10.11578/E3SM/dc.20180418.36, https://doi.org/10.11578/E3SM/dc.20180418.36, https://dx.doi.org/10.11578/E3SM/dc.20180418.36, 2018.

Farrell, P. and Maddison, J.: Conservative interpolation between volume meshes by local Galerkin projection, Computer Methods in Applied Mechanics and Engineering, 200, 89 – 100, https://doi.org/http://dx.doi.org/10.1016/j.cma.2010.07.015, http://www.sciencedirect.com/science/article/pii/S0045782510002276, 2011.

Fleishman, S., Cohen-Or, D., and Silva, C. T.: Robust moving least-squares fitting with sharp features, ACM transactions on graphics (TOG), 24, 544–552, 2005.

Flyer, N. and Wright, G. B.: Transport schemes on a sphere using radial basis functions, Journal of Computational Physics, 226, 1059–1084, 2007.

Fornberg, B. and Piret, C.: On choosing a radial basis function and a shape parameter when solving a convective PDE on a sphere, Journal of Computational Physics, 227, 2758–2780, 2008.

Fox, G., Johnson, M., Lyzenga, G., Otto, S., Salmon, J., Walker, D., and White, R. L.: Solving Problems On Concurrent Processors Vol. 1: General Techniques and Regular Problems, Computers in Physics, 3, 83–84, 1989.

Gander, M. J. and Japhet, C.: An algorithm for non-matching grid projections with linear complexity, in: Domain Decomposition Methods in Science and Engineering XVIII, pp. 185–192, Springer, 2009.

Gander, M. J. and Japhet, C.: Algorithm 932: PANG: software for nonmatching grid projections in 2D and 3D with linear complexity, ACM Transactions on Mathematical Software (TOMS), 40, 6, 2013.

Gottlieb, D. and Shu, C.-W.: On the Gibbs phenomenon and its resolution, SIAM review, 39, 644–668, 1997.

Grandy, J.: Conservative remapping and region overlays by intersecting arbitrary polyhedra, J. Comput. Phys., 148, 433–466, 1999.

Herring, A. M., Certik, O., Ferenbaugh, C. R., Garimella, R. V., Jean, B. A., Malone, C. M., and Sewell, C. M.: (U) Introduction to Portage, Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States), https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-17-20831, 2017.

Hill, C., DeLuca, C., Balaji, Suarez, M., and Silva, A. d.: The architecture of the earth system modeling framework, Computing in Science & Engineering, 6, 18–28, 2004.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., et al.: The community earth system model: a framework for collaborative research, Bulletin of the American Meteorological Society, 94, 1339–1360, 2013.

Jacob, R., Larson, J., and Ong, E.: M× N communication and parallel interpolation in Community Climate System Model Version 3 using the model coupling toolkit, The International Journal of High Performance Computing Applications, 19, 293–307, 2005a.

Jacob, R., Larson, J., and Ong, E.: M× N communication and parallel interpolation in Community Climate System Model Version 3 using the model coupling toolkit, The International Journal of High Performance Computing Applications, 19, 293–307, 2005b.

Jiao, X. and Heath, M. T.: Common-refinement-based data transfer between non-matching meshes in multiphysics simulations, International Journal for Numerical Methods in Engineering, 61, 2402–2427, 2004.

Jones, P. W.: First-and second-order conservative remapping schemes for grids in spherical coordinates, Monthly Weather Review, 127, 2204–2210, 1999.

Larsen, E., Gottschalk, S., Lin, M. C., and Manocha, D.: Fast proximity queries with swept sphere volumes, Tech. rep., Technical Report TR99-018, Department of Computer Science, University of North Carolina, 1999.

Larson, J. W., Jacob, R. L., Foster, I., and Guo, J.: The model coupling toolkit, in: International Conference on Computational Science, pp. 185–194, Springer, 2001.

Lauritzen, P. H., Nair, R. D., and Ullrich, P. A.: A conservative semi-Lagrangian multi-tracer transport scheme (CSLAM) on the cubed-sphere grid, Journal of Computational Physics, 229, 1401 – 1424, https://doi.org/http://dx.doi.org/10.1016/j.jcp.2009.10.036, http://www.sciencedirect.com/science/article/pii/S002199910900597X, 2010.

Löhner, R.: Recent advances in parallel advancing front grid generation, Archives of Computational Methods in Engineering, 21, 127–140, 2014.

Löhner, R. and Parikh, P.: Generation of three-dimensional unstructured grids by the advancing-front method, International Journal for Numerical Methods in Fluids, 8, 1135–1149, 1988.

Mahadevan, V., Grindeanu, I. R., Ray, N., Jain, R., and Wu, D.: SIGMA Release v1.2 - Capabilities, Enhancements and Fixes, https://doi.org/10.2172/1224985, 2015.

Morozov, D. and Peterka, T.: Block-Parallel Data Analysis with DIY2, 2016.

Plimpton, S. J., Hendrickson, B., and Stewart, J. R.: A parallel rendezvous algorithm for interpolation between multiple grids, Journal of Parallel and Distributed Computing, 64, 266–276, 2004.

Powell, D. and Abel, T.: An exact general remeshing scheme applied to physically conservative voxelization, Journal of Computational Physics, 297, 340 – 356, https://doi.org/https://doi.org/10.1016/j.jcp.2015.05.022, http://www.sciencedirect.com/science/article/pii/S0021999115003563, 2015.

Rančić, M.: An efficient, conservative, monotonic remapping for semi-Lagrangian transport algorithms, Monthly Weather Review, 123, 1213–1217, 1995.

Reichler, T. and Kim, J.: How well do coupled models simulate today's climate?, Bulletin of the American Meteorological Society, 89, 303–312, 2008.

Shewchuk, J. R.: Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator, in: Applied computational geometry towards geometric engineering, pp. 203–222, Springer, 1996.

SIGMA toolkit: Scalable Interfaces for Geometry and Mesh based Applications (SIGMA) toolkit, http://sigma.mcs.anl.gov, http://sigma.mcs.anl.gov, 2014.

Slattery, S., Wilson, P., and Pawlowski, R.: The data transfer kit: a geometric rendezvous-based tool for multiphysics data transfer, in: International conference on mathematics & computational methods applied to nuclear science & engineering (M&C 2013), pp. 5–9, 2013.

Slattery, S. R.: Mesh-free data transfer algorithms for partitioned multiphysics problems: Conservation, accuracy, and parallelism, Journal of Computational Physics, 307, 164–188, 2016.

Slingo, J., Bates, K., Nikiforakis, N., Piggott, M., Roberts, M., Shaffrey, L., Stevens, I., Vidale, P. L., and Weller, H.: Developing the next-generation climate system models: challenges and achievements, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 367, 815–831, 2009.

Tautges, T. J. and Caceres, A.: Scalable parallel solution coupling for multiphysics reactor simulation, Journal of Physics: Conference Series, 180, 2009.

Tautges, T. J., Meyers, R., Merkley, K., Stimpson, C., and Ernst, C.: MOAB: A Mesh-Oriented datABase, SAND2004-1592, Sandia National Laboratories, 2004.

Tautges, T. J., Kraftcheck, J., Bertram, N., Sachdeva, V., and Magerlein, J.: Mesh Interface Resolution and Ghost Exchange in a Parallel Mesh Representation, IEEE, Shanghai, China, 2012.

Ullrich, P. A. and Taylor, M. A.: Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part I, Monthly Weather Review, 143, 2419–2440, 2015.

Ullrich, P. A., Lauritzen, P. H., and Jablonowski, C.: Geometrically Exact Conservative Remapping (GECoRe): Regular latitude–longitude and cubed-sphere grids, Monthly Weather Review, 137, 1721–1741, 2009.

Ullrich, P. A., Lauritzen, P. H., and Jablonowski, C.: Some considerations for high-order ‘incremental remap’-based transport schemes: edges, reconstructions, and area integration, International Journal for Numerical Methods in Fluids, 71, 1131–1151, 2013.

Ullrich, P. A., Devendran, D., and Johansen, H.: Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part II, Monthly Weather Review, 144, 1529–1549, 2016.

Valcke, S.: The OASIS3 coupler: a European climate modelling community software, Geoscientific Model Development, 6, 373–388, 2013.

Washington, W., Bader, D., and Collins, B, e. a.: Challenges in climate change science and the role of computing at the extreme scale, in: Proc. of the Workshop on Climate Science, 2008.

Government License

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan. http://energy.gov/downloads/doe-public-access-plan.