

A GENERIC FINITE ELEMENT FRAMEWORK ON PARALLEL TREE-BASED ADAPTIVE MESHES

SANTIAGO BADIA^{1,2}, ALBERTO F. MARTÍN^{2,3}, ERIC NEIVA^{2,3}, AND FRANCESC VERDUGO³

ABSTRACT. We present highly scalable parallel distributed-memory algorithms and associated data structures for a *generic* finite element framework that supports h -adaptivity on computational domains represented as multiple connected adaptive trees—forest-of-trees—, thus providing multi-scale resolution on problems governed by partial differential equations. The framework is grounded on a rich representation of the adaptive mesh suitable for generic finite elements that is built on top of a low-level, light-weight forest-of-trees data structure handled by a specialized, highly parallel adaptive meshing engine. Along the way, we have identified the requirements that the forest-of-trees layer must fulfill to be coupled into our framework. Essentially, it must be able to describe neighboring relationships between cells in the adapted mesh (apart from hierarchical relationships) across the lower-dimensional objects at the boundary of the cells. Atop this two-layered mesh representation, we build the rest of data structures required for the numerical integration and assembly of the discrete system of linear equations. We consider algorithms that are suitable for both subassembled and fully-assembled distributed data layouts of linear system matrices. The proposed framework has been implemented within the FEMPAR scientific software library, using p4est as a practical forest-of-octrees demonstrator. A comprehensive strong scaling study of this implementation when applied to Poisson and Maxwell problems reveals remarkable scalability up to 32.2K CPU cores and 482.2M degrees of freedom. Besides, the implementation in FEMPAR of the proposed approach is up to 2.6 and 3.4 times faster than the state-of-the-art deal.II finite element software in the h -adaptive approximation of a Poisson problem with first- and second-order Lagrangian finite elements, respectively (excluding the linear solver step from the comparison).

Keywords: Partial Differential Equations, Finite Elements, Adaptive Mesh Refinement, Forest of Trees, Parallel algorithms, Scientific Software

1. INTRODUCTION

In this paper, we present the data structures and algorithms of a generic Finite Element (FE) simulation framework that is designed such that: (a) it supports generic FE spaces of arbitrary order, i.e., it is not restricted to a particular FE space (e.g., first order H^1 -conforming Lagrangian FE spaces), thus covering a wide range of applications governed by Partial Differential Equations (PDEs); (b) it supports Adaptive

¹SCHOOL OF MATHEMATICAL SCIENCES, MONASH UNIVERSITY, CLAYTON, VICTORIA, 3800, AUSTRALIA.

²CIMNE – CENTRE INTERNACIONAL DE MÈTODES NUMÈRICS EN ENGINYERIA, UPC, ESTEVE TERRADAS 5, 08860 CASTELLDEFELS, SPAIN.

³DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING, UNIVERSITAT POLITÈCNICA DE CATALUNYA, JORDI GIRONA 1-3, EDIFICI C1, 08034 BARCELONA, SPAIN.

E-mail addresses: sbadia@cimne.upc.edu, amartin@cimne.upc.edu, eneiva@cimne.upc.edu, fverdugo@cimne.upc.edu.

Date: July 9, 2019.

Financial support from the European Commission under the EMUSIC and ExaQUte projects within the Horizon 2020 Framework Programme is gratefully acknowledged (Grant Ids. 690725 and 800898, resp.). This work has been partially funded by the project MTM2014-60713-P from the “Ministerio de Economía, industria y Competitividad” of Spain. SB gratefully acknowledges the support received from the Catalan Government through the ICREA Acadèmia Research Program. EN gratefully acknowledges the support received from the Catalan Government through an FI fellowship (2019 FI-B2-00090; 2018 FI-B1-00095; 2017 FI-B-00219). FV gratefully acknowledges the support received from the *Secretaria d’Universitats i Recerca* of the Catalan Government in the framework of the Beatriu Pinós Program (Grant Id.: 2016 BP 00145). The authors thankfully acknowledge the computer resources at Marenostrum-IV and the technical support provided by the Barcelona Supercomputing Center (RES-ActivityID: FI-2018-2-0009 and FI-2018-3-0029).

Mesh Refinement and coarsening (AMR) (a.k.a. h -adaptivity) on computational domains that can be geometrically discretized as multiple connected adaptive trees—forest-of-trees—, thus being suitable for the efficient modeling of multi-scale problems; (c) it is highly scalable on current petascale distributed-memory supercomputers, thus suitable for large-scale simulations.

The framework follows a *two-layered meshing approach*, namely an inner, light-weight layer encoding the forest-of-trees, handled by a specialized meshing engine (to attain (b) and (c)), and an outer representation of the adaptive mesh suitable for the implementation of generic adaptive FE spaces (thus attaining (a)). The framework requires the forest-of-trees meshing engine to be able to provide local neighborhood information across the cell boundaries in the adapted mesh. The ideas underlying the meshing approach in our framework are general in that they apply for general cell topologies and an arbitrary number of dimensions. On top of this rich adaptive mesh representation, we also discuss the rest of algorithms and data structures in the framework towards the final assembly of the discrete system of linear equations. We pay close attention to algorithms that are well-suited for the construction of subassembled linear systems, although a fully-assembled data layout for the linear system can also be built. The subassembled data distribution is the most suitable for non-overlapping domain decomposition solvers, such as, e.g., the Balancing Domain Decomposition by Constraints (BDDC) preconditioner [1].

The inner light-weight engine in the framework relies on the significant progress made in recent years on scalable algorithms and software for fast AMR using adaptive trees endowed with Space-Filling Curves (SFCs) [2]. Most research work in this topic has been focused on quadrilateral (2D) or hexahedral (3D) adaptive meshes. The recursive application of the standard isotropic 1:4 (2D) and 1:8 (3D) refinement rule leads to adaptive trees that are referred as quadtrees [3] and octrees [4], resp. The data structure resulting from patching them together are called *forest-of-quadtrees* and *-octrees*, resp., although the latter term is typically employed in either case. We refer to [5, 6, 7] for works on single-octree handling, and to [8, 9] for their extension to a forest-of-octrees. All these works are based on the exploitation of the Morton SFC index for parallel data storage and scalable dynamic load balancing. The state-of-the-art in these techniques is available at the open source parallel forest-of-octrees meshing engine `p4est` [8]. However, tree-based AMR with SFCs can be generalized to other cell topologies as well. The authors of [10] propose an approach for simplicial adaptive meshes (i.e., composed of triangles or tetrahedra) that is grounded on Bey’s red-refinement rule, and the so-called Tetrahedral Morton (TM) SFC index [10]. Holke’s dissertation [2] goes even further by providing abstract, extensible tree-based AMR with SFCs manipulation algorithms, i.e., cell-topology- and dimension-independent. These ideas are being implemented in the on-going, open source software effort `t8code` [2].

The exploitation of single-octree or, more generally, forest-of-octrees in parallel adaptive FE solvers [11, 12, 13, 14] has been shown to be cornerstone in several large-scale application problems [14, 15, 16, 17, 18, 19, 20]. Forest-of-trees meshes provide multi-resolution capability by local adaptation. Indeed, this capability makes them perfectly suitable, among others, for the following three scenarios: for PDEs, in which the area that requires high-resolution is known *a priori*, they can be adapted before starting the actual numerical simulation (see, e.g., [15, 21]); for PDEs, in which solution features evolve spatially over time, and that are known only during the actual numerical simulation, they can be adapted dynamically, driven by a posteriori error estimates [22] during the actual simulation (see, e.g., [17]); for PDEs, posed on *complex* domains, and in combination with unfitted (a.k.a. immersed or embedded) FE discretization methods (as, e.g., the finite cell [23], cutFEM [24], or agFEM methods [25, 26]), they can be used to control geometry approximation errors by adaptation in regions of high geometric variability.

An implementation of the framework is available at FEMPAR [27], an open source Object-Oriented (OO) Fortran200X scientific software package for the High Performance Computing (HPC) simulation of complex multiphysics problems governed by PDEs at large scales. FEMPAR uses `p4est` as its specialized forest-of-octrees meshing engine, although any other engine able to fulfill a set of requirements described in the paper could be plugged in the framework as well. FEMPAR linear solver kernel relies on a highly scalable implementation of multilevel BDDC preconditioners [28, 29], thus requiring subassembled data

distribution of operators. In this sort of data distribution, Degrees Of Freedom (DOFs) laying on the interface of the partition into subdomains are actually replicated and become local to the subdomains sharing it. As a consequence, the global linear system is subassembled, i.e., the local portion of each subdomain holds partial contributions to the entries of the global matrix corresponding to couplings among interface DOFs. This is in contrast to the vast majority of widely available parallel distributed-memory FE and linear algebra software packages, such as, e.g., `deal.II` [30], FEniCS [31], PETSc [32], or TRILINOS [33], which are designed to work with globally-addressable, row-wise partitioned, distributed fully-assembled discrete linear systems. In order to deal with this data layout, one has to generate and handle a global labeling of DOFs across the whole domain. Also, it requires communication in order to set its sparsity pattern, and the entries themselves. These two steps have been proved to be time consuming and to affect parallel efficiency. *With subassembled operators, both requirements are avoided.* In any case, even if the algorithms are more efficient for the construction of subassembled linear systems, by means of an extra step, also discussed along the paper, one may as well build fully-assembled operators within the framework.

The excellent work in [34] discusses the particular approach followed by the state-of-the-art `deal.II` FE software library [30] in order to support generic adaptive FE spaces atop parallel forest-of-octrees meshes. The approach presented herein, and the one in [34], encode the distributed mesh in a two-layered approach. However, *there are two fundamental differences among these two approaches, which along this paper are shown experimentally to translate into performance and scalability improvements.* On the one hand, the outer layer mesh representation itself, and the approach followed to reconstruct the outer layer from the inner one, are fundamentally different. In particular, [34] follows the so-called match-tree-recursive approach, while the one here is based on local neighborhood information across the cell boundaries in the adapted mesh provided by the forest-of-trees meshing engine. On the other hand, the algorithms in [34] are suitable for the construction of fully-assembled linear systems, while the ones covered in this paper, for subassembled and fully-assembled ones. While differences among fully-assembled and subassembled linear systems have been already identified in [35], up to the authors' knowledge, there is no systematic experimental study in the literature, where performance and parallel scalability of generic adaptive FE pipelines suitable for either case are compared. The paper also tries to fill the aforementioned gap by experimentally comparing the algorithms, data structures, and software available within FEMPAR with those available at `deal.II` [34] on a massively parallel supercomputer.

This paper is structured as follows. In Sect. 2, we overview the so-called tree-based AMR with SFCs approach for the parallel scalable generation and partitioning of adaptive meshes. In Sect. 3 we present a rich representation of the distributed adaptive mesh that supports generic FE spaces built atop. This data structure is grounded on concepts which are cell-topology- and space-dimension-independent. In Sect. 4, we discuss the construction of conforming FE spaces on top of non-conforming meshes via the so-called hanging DOF constraints, while in Sect. 5 and Sect. 6 we focus on the high-level algorithms involved in the construction of parallel subassembled and fully-assembled linear systems, resp. In Sect. 7 we assess the parallel strong scalability of the proposed high-level algorithms and data structures in FEMPAR for the Poisson and Maxwell PDEs on the MN-IV petascale supercomputer, and compare them with their counterparts in `deal.II` for the former problems. Finally, we draw conclusions in Sect. 8.

2. OVERVIEW OF TREE-BASED AMR ENDOWED WITH SPACE-FILLING CURVES

We describe the so-called tree-based AMR with SFCs approach for scalable mesh generation and partitioning in a parallel distributed-memory framework. The idea was originally proposed in [8] for the particular case of octrees, and extended to general trees in [2]. Let $\Omega \subset \mathbb{R}^d$ be an open bounded domain in which our PDE problem is posed. AMR with SFCs can be seen as a two-level decomposition of Ω , referred to as macro and micro level, resp.¹ In the macro level, we assume that there is a partition C of Ω into cells $K \in C$ such that each of these cells can be expressed as a differentiable homeomorphism

¹We note that this two-level construction is for geometrical purposes only.

Φ_K over a set of admissible reference polytopes [27]. The mesh C is referred to as the coarse mesh and it is assumed to be a *conforming mesh* (see Sect. 2.1). In the micro level, each of the cells of C becomes the root of an adaptive tree. This two-tier adaptive structure is referred to as *forest-of-trees*. It represents a locally refined mesh \mathcal{T} . The process that generates it is essentially characterized by defining two complementary ingredients, namely a (recursive) refinement rule (see Sect. 2.2) and an SFC index (see Sect. 2.3).

2.1. Polytopes and coarse mesh. Reference polytopes for FE analysis are usually cubes or tetrahedra (and their extensions to arbitrary dimensions), but prisms and pyramids can also be used. The boundary of a polytope T is in turn composed of polytopes of lower dimension, denoted as n -faces, with $0 \leq n < d$ being their topological dimension; the d -face is the polytope itself. Furthermore, the n -faces on the boundary of an n -face of T are also n -faces of T . For example, with $d = 3$, the boundary is composed of Vertices, Edges and Faces (VEFs), representing the 0, 1, and 2-faces of the cell, resp. Hereafter, we may abuse notation by using the acronym VEF for any n -face with $n \in [0, d]$. The set of VEFs of T is represented with \mathcal{F}_T . The n -faces span a closed domain if $n = 0$, and open otherwise. We define an orientation for the polytope T (resp., a VEF $t \in \mathcal{F}_T$), e.g., via a coordinate system and lexicographical order, which makes the polytope (resp., a VEF) a totally ordered set represented with \vec{T} (resp., \vec{t}). A vertex is a degenerate case that does not require an orientation. An edge has two admissible orientations. In the case of squared faces, eight orientations are admissible, which are uniquely represented by its first two vertices for the corresponding order [27].

Cells in the physical domain $K \in \mathcal{T}$ are determined by Φ_K and the reference polytope T . The map Φ_K applied to every $t \in \mathcal{F}_T$ generates the set of *local* VEFs of K , denoted as \mathcal{F}_K . A local VEF can be understood as a tuple of the cell K it belongs to and the geometrical domain it represents, which we denote as $[f]$. Global VEFs only represent the geometrical domain. Thus, given a local VEF f , its corresponding global VEF is $[f]$. Formally, local VEFs are glued together into global VEFs as dictated by an equivalence relation \sim such that $f \sim g$ iff $[f] = [g]$. We represent with \mathcal{F} the set of *global* VEFs of \mathcal{T} , i.e., the quotient space of all equivalence classes (global VEFs) resulting from the application of \sim to the elements of $\bigcup_{K \in \mathcal{T}} \mathcal{F}_K$. $[\cdot]$ is the so-called local-to-global VEF map.

One can provide $K \in \mathcal{T}$ and $f \in \mathcal{F}_K$ with an orientation. The resulting ordered sets are denoted as \vec{K} and \vec{f} , resp. These are not arbitrary, but given by $\vec{K} \doteq \Phi_K(\vec{T})$ and $\vec{f} \doteq \Phi_K(\vec{t})$, resp., with $t \in \mathcal{F}_T$ such that $f = \Phi_K(t)$. Global VEFs $F \in \mathcal{F}$ are also given an orientation, represented with \vec{F} . The orientation of F may differ from the one of the local VEFs f such that $[f] = F$. One can determine the change of basis between \vec{f} and \vec{F} , and vice versa. We refer the reader to [27] for a detailed explanation of a change of basis approach that is extensible to arbitrary dimensions and cell topologies. A related, but different approach, tailored to hexahedral meshes, is presented in [8, Sect. 2.2], and it is extended to more general cases in [2, Sect. 6.1.3]. We note that tetrahedral meshes can easily be *oriented*, i.e., an orientation can be given for all global VEFs $F \in \mathcal{F}$ such that the change of basis is always the identity [36].

As mentioned at the beginning of Sect. 2, in AMR with SFCs, the coarse mesh C is assumed to be conforming. A formal definition of this property is as follows.

Definition 2.1 (Conforming mesh). *A mesh \mathcal{T} is conforming if, for any two cells $K, K' \in \mathcal{T}$ with $\bar{K} \cap \bar{K}' \neq \emptyset$, there exist $f \in \mathcal{F}_K$, $f' \in \mathcal{F}_{K'}$ such that $\bar{f} = \bar{f}' = \bar{K} \cap \bar{K}'$.*

2.2. Refinement rule. Let us consider a cell $K \in C$ in the coarse mesh of Ω . A *refinement rule* prescribes how one may subdivide the coarse cell K into finer cells that span the same region as K . K is referred to as the *parent* cell, and the latter ones, (its) *children* cells. The refinement rule is usually defined at the polytope and then mapped to the physical space using Φ_K . The reverse application of the refinement rule, i.e., the replacement of the children cells by its parent, is referred to as coarsening. Mesh generation in this context is a hierarchical process based on the recursive application of refinement and coarsening. At each level in the hierarchy, some cells are marked for refinement, and some other for coarsening. A cell marked for refinement is replaced by its children cells following the refinement rule. On the other hand,

if all children cells of a given parent are marked for coarsening, they are collapsed into the parent cell. As a result of this process for all cells in C , we obtain a *forest* of tree-like refinement structures rooted at every $K \in C$. We make the following assumption on the refinement rule.

Assumption 2.2 (Admissible refinement rule). *The cell refinement rule over a cell K provides a conforming mesh \mathcal{R}_K of K (into children cells) and an orientation to the n -faces of \mathcal{R}_K . The restriction of \mathcal{R}_K to any n -face $f \in \mathcal{F}_K$ with $n > 0$ is a non-trivial partition of f , i.e., these n -faces are subdivided. Besides, when recursively applying the refinement rule to a cell, the resulting mesh is non-degenerate: there is a constant $\rho > 0$ independent of the number of levels of refinement such that, for any cell K in the resulting mesh, there is a ball of diameter $\rho \operatorname{diam}(K)$ contained in K , where $\operatorname{diam}(K)$ denotes the diameter of K .*

Ass. 2.2 is not restrictive, since it holds, e.g., for uniform refinement rules of hexahedra, tetrahedra, prisms, and pyramids (and 2D counterparts) [2]. For hexahedra and prisms, the children cells $K_c \in \mathcal{R}_K$ can be expressed as a uniform scaling and translation of the parent cell K ; the resulting tree meshes are oriented. For simplicial meshes, the children cells $K_c \in \mathcal{R}_K$ can be expressed as a uniform scaling, translation and rotation of the parent cell K . The situation is more involved for pyramids, since some children cells must be tetrahedra.

We define the following relation between the VEFs of \mathcal{R}_K , represented with $\mathcal{F}_{\mathcal{R}_K} \doteq \bigcup_{K_c \in \mathcal{R}_K} \mathcal{F}_{K_c}$, and the n -faces in $\mathcal{F}_K \cup \{K\}$.

Definition 2.3 (Owner of refined VEF). *For every VEF $f \in \mathcal{F}_{\mathcal{R}_K}$, we define its owner n -face O_f as the unique n -face in $\mathcal{F}_K \cup \{K\}$ that contains it, i.e., $f \subset O_f$.²*

By Def 2.3, O_f has a topological dimension greater than or equal to the one of f .

A particular adaptive mesh \mathcal{T} of Ω to be used for FE discretization is defined as the union of all leaf cells (i.e., cells with no children) in the forest. We denote by $|\mathcal{T}|$ the number of cells in \mathcal{T} . For every cell $K \in \mathcal{T}$, we can define $\ell(K)$ as the *level of refinement* of K in the forest, with $\ell(K) = 0$ if $K \in C$ and $\ell(K) > 0$ otherwise. If, for any two cells $K, K' \in \mathcal{T}$, we have that $\ell(K) > \ell(K')$, then K is *finer* than K' , and K' is *coarser* than K . We say that a local (*refined*) n -face $f \in \mathcal{F} \doteq \bigcup_{K \in \mathcal{T}} \mathcal{F}_K$ is an *inter-tree* n -face if there is a local (*coarse*) n -face $f' \in \bigcup_{K \in C} \mathcal{F}_K$ such that $f \subseteq f'$, and *intra-tree* otherwise.

2.3. The SFC index. Let us assume that we bound the maximum level of refinement for any forest-of-trees that can be built by means of the hierarchical process described in Sect. 2.2; this is the case in practice since available memory is limited. Let us represent with \mathcal{S}_n the refinement tree that results from n recursive applications of the refinement rule to all leaves. \mathcal{S}_n includes the set of all cells with a maximum refinement level of n that can be potentially constructed from C by means of AMR. With this notation, we can readily introduce the concept of *SFC index*.³ The maximum level subscript is omitted in the definition.

Definition 2.4. *An SFC index \mathcal{I} on \mathcal{S} is formally defined as a map $\mathcal{I} : \mathcal{S} \rightarrow \mathbb{N}_0$ that fulfills the following three properties. First, the map $\mathcal{I} \times \ell : \mathcal{S} \rightarrow \mathbb{N}_0 \times \mathbb{N}_0$ is injective, and thus the pair composed by the SFC index and the refinement level of a cell uniquely identifies it among all cells in \mathcal{S} . Second, for any $K, K' \in \mathcal{S}$, where K' is a descendant of K , then $\mathcal{I}(K) \leq \mathcal{I}(K')$. Additionally, given a K'' that is not a descendant of K and $\mathcal{I}(K) < \mathcal{I}(K'')$, it holds that $\mathcal{I}(K') < \mathcal{I}(K'')$.*

A key requirement from the computational viewpoint is to choose the \mathcal{I} mapping such that functions that locally operate on a cell (or small set of cells) take *constant time*, independently of the level of refinement of the cell. Local cell functions include computing the SFC index of a cell, determining the SFC index of its parent, children, and neighbors with the same refinement level, or computing vertices

²Along the paper we will use \subset (resp., \supset) to denote subset (resp., superset) inclusion. The proper (a.k.a. strict) variant of this operator will be denoted as \subsetneq (resp., \supsetneq).

³The definition in the sequel has been adapted from [2], where a novel approach to the theory of discrete SFCs suitable for tree-based AMR is presented.

coordinates. Besides, the exploitation of the SFC index for implementing such operations has to result in significant memory savings with respect to unstructured meshing, in which a list of neighbors has to be stored, as well as the coordinates of all vertices in the mesh, for each cell. The properties of the SFC index are such that it can be, thus, exploited for efficient mesh handling.

As a consequence of Def. 2.4, the refinement is a local operation in the sense that it preserves the ordering induced by the SFC index, i.e., the new cells generated by means of refinement have an SFC index inbetween the ones of those already existing in the forest. It also implies that the map $\mathcal{I} : \mathcal{T} \rightarrow \mathbb{N}_0$ is injective, and thus the SFC index uniquely identifies the leaves in \mathcal{T} . Therefore, the leaves of the forest at hand can be *uniquely arranged* in a linear array *in ascending order by their SFC index*, so that preserving this order when replacing a parent cell by its children or vice-versa becomes easy to implement (and efficient). Indeed, common practice in state-of-the-art codes is to store only the finest cells of the forest. This storage layout, referred to as linear forest-of-trees ordering in the literature, corresponds to an SFC embedded within the computational domain, with its shape depending on the particular definition of the SFC index. Apart from providing an efficient storage layout, SFCs also provide a linear runtime solution to the partitioning problem of the geometrical domain among processors. In particular, a cell-based partition of \mathcal{T} is simply generated by dividing the leaves in the linear ordering into as many equally-sized segments as processors involved in the computation. This approach in turn circumvents the high computational cost and lack of parallel scaling proneness typically associated with dynamic load balancing via unstructured graph partitioning algorithms [37].⁴ On the other hand, in contrast to these latter approaches, SFCs do not necessarily minimize the so-called *edge cut* [37] of the partition into subdomains, and may lead to disconnected subdomains [38].

We note that forest-of-trees can be exploited to consider more complex physical domains than single-tree approaches, i.e., conforming coarse meshes C with multiple cells. In any case, in order for forest-of-trees meshing to be efficient, scalable, and light-weight, the number of cells in the coarse unstructured mesh must be *small* (compared to the number of leaves).

2.4. Non-conformity and cell neighbors. Tree-based meshes provide multi-resolution capability by local adaptation, i.e., cells in $K \in \mathcal{T}$ might have different value for $\ell(K)$. However, these meshes are (potentially) *non-conforming*, i.e., they contain the so-called *hanging VEFs*. These occur at the interface of neighboring cells with different refinement levels.

In order to provide support to FE applications, tree-based AMR with SFCs engines also require to keep track of neighboring relationships between cells in \mathcal{T} (apart from hierarchical relationships). One possible way of describing these is by means of the notion of *cell neighbors across local n-faces*, which we introduce in the sequel. It turns out that this approach of representing cell connectivities best satisfies the requirements of the algorithms within the forest-of-trees meshing engine that require a description of the neighborhood of mesh cells (see, e.g., Balance or Ghost in Sect. 2.6).

Definition 2.5 (Neighbors of a cell across local n -faces). *The neighbors of $K \in \mathcal{T}$ across its n -face $f \in \mathcal{F}_K$ are classified into the sets $\mathcal{T}_{K,f}$, $\mathcal{T}_{K,f}^+$, and $\mathcal{T}_{K,f}^-$, referred to as *conformal*, *higher-level*, and *lower-level* neighbors, resp. They are composed of those cells $K' \in \mathcal{T} \setminus K$ that contain $f' \in \mathcal{F}_{K'}$ such that: (a) $f \odot f'$, with \odot being $=$, \supseteq , and \subsetneq for $\mathcal{T}_{K,f}$, $\mathcal{T}_{K,f}^+$, and $\mathcal{T}_{K,f}^-$, resp.; (b) $\bar{f} \cap \bar{f}' = \bar{K} \cap \bar{K}'$.*

In order to better grasp the $\mathcal{T}_{K,f}$, $\mathcal{T}_{K,f}^+$, and $\mathcal{T}_{K,f}^-$ sets, we enumerate several properties that these sets fulfill as an immediate consequence of their definition in Def. 2.5. First, if \mathcal{T} is conforming (see Def. 2.1), $\mathcal{T}_{K,f}^+$ and $\mathcal{T}_{K,f}^-$ are empty for all cells and local n -face combinations. Second, if f is a 0-face, the sets $\mathcal{T}_{K,f}^+$ and $\mathcal{T}_{K,f}^-$ are *always* void. Third, if f is a facet (i.e., $d - 1$ -face), the cardinality of $\mathcal{T}_{K,f}$ and $\mathcal{T}_{K,f}^-$ is at most one. Finally, facets f laying on $\partial\Omega$ can be distinguished as the ones for which all these sets

⁴Dynamic load balancing is the ability of an adaptive mesh to be re-distributed in the presence of an unacceptable amount of load imbalance, e.g., the one generated by means of AMR in a highly localized region of Ω .

are void. These boundary facets and the VEFs laying on its boundary compose the full set of VEFs of \mathcal{T} which lay on $\partial\Omega$.

Besides, the following holds for conformal and lower-level neighbors across vertices and edges in the particular case of *forest-of-octrees* [8]. On the one hand, if f is an intra-octree vertex, then $|\mathcal{T}_{K,f}| \leq 1$; else, $|\mathcal{T}_{K,f}|$ can be arbitrary since C might be an unstructured mesh. On the other hand, if f is an intra-octree edge, then $|\mathcal{T}_{K,f}| \leq 1$ and $|\mathcal{T}_{K,f}^-| \leq 1$; for inter-octree edges, the cardinality of these sets can be arbitrary since C might be an unstructured mesh.

With the cell neighbors across n -faces, one can compute the local-to-global VEF map $[\cdot]$; see Alg. 1 for more details. Given $f \in \mathcal{F}_K$ and $F = [f]$, we define $\mathcal{T}_{K,F} \doteq \mathcal{T}_{K,f}$ (analogously for $\mathcal{T}_{K,F}^+$ and $\mathcal{T}_{K,F}^-$). We denote by $[\mathcal{F}_K] \doteq \{[f] : f \in \mathcal{F}_K\}$ the set of global VEFs of K . We next introduce a pair of propositions on $\mathcal{T}_{K,F}$, $\mathcal{T}_{K,F}^+$, and $\mathcal{T}_{K,F}^-$. In order to prove these (and some of those in Sect. 3), we need the following assumption on the recursive refinement procedure.

Assumption 2.6 (Uniform refinement conformity). *The mesh composed of all leaves of the refinement tree \mathcal{S}_ℓ obtained after an arbitrary level ℓ of uniform refinements of C is also conforming.*

The uniform refinement conformity in Ass. 2.6 requires a refinement rule consistent among cells in the coarse mesh. We note that this assumption is already fulfilled by the standard uniform refinement rules mentioned after Ass. 2.2.

Proposition 2.7. *Given a cell $K \in \mathcal{T}$, and an n -face $F \in [\mathcal{F}_K]$ such that $n > 0$, then $\mathcal{T}_{K,F}$ can only be composed of neighbor cells at the same refinement level as K . On the other hand, $\mathcal{T}_{K,F}^-$ (resp., $\mathcal{T}_{K,F}^+$) can only be composed of neighbor cells with lower (resp., higher) refinement level than that of K .*

Proof. By Def. 2.5, for any $K' \in \mathcal{T}_{K,F}$, it holds $F \in [\mathcal{F}_{K'}]$ and $\bar{F} \subseteq \bar{K} \cap \bar{K}'$. If $\ell(K) > \ell(K')$ were true, then there would be an ancestor K'' of K at the same refinement level of K' . By Ass. 2.6, the mesh composed of all leaves of the refinement tree $\mathcal{S}_{\ell(K')}$ is a conforming mesh. Thus, by Def. 2.1, there exists an n -face $G \in [\mathcal{F}_{K'}] \cap [\mathcal{F}_{K''}]$ such that $\bar{G} = \bar{K}' \cap \bar{K}''$. Since $\bar{F} \subseteq \bar{K} \cap \bar{K}'$, $\bar{G} = \bar{K}' \cap \bar{K}''$, $K \subset K''$, and the refinement rule subdivides all cell n -faces with $n > 0$ (see Ass 2.2), then $F \subseteq G$. As a result, K' has two *different* VEFs $F, G \in [\mathcal{F}_{K'}]$ such that $F \cap G \neq \emptyset$. This is a contradiction, since the VEFs of a cell are disjoint by definition. We proceed analogously for $\ell(K) < \ell(K')$. It proves the result for $\mathcal{T}_{K,F}$. The result for $\mathcal{T}_{K,F}^-$ and $\mathcal{T}_{K,F}^+$ can be proved in a similar way. \square

However, for $n = 0$, $\mathcal{T}_{K,F}$ might be composed of neighbors at the same, higher, or lower-level of refinement than K . This can be readily observed, e.g., in Fig. 1a, for any corner that meets at the boundary of cells with different levels of refinement.

2.5. Balanced forest-of-trees. As mentioned in Sect. 2.4, forest-of-trees meshes are, in general, non-conforming. For FE applications, mesh non-conformity hardens the construction of conforming FE spaces, and the subsequent steps in the simulation; see Sect. 4. However, common practice in order to significantly alleviate this extra complexity consists on enforcing the so-called 2:1 balance constraint (a.k.a. balance or mesh regularity condition). A general definition of the 2:1 balance condition, parameterized by an integer $0 \leq k < d$, is as follows.

Definition 2.8. *A forest-of-trees mesh \mathcal{T} is 2:1 k -balanced if and only if, for any cell $K \in \mathcal{T}$, and n -face $F \in [\mathcal{F}_K]$, with $n \in [k, d]$, there is no neighbor K' of K across F (see Def 2.5) such that $|\ell(K') - \ell(K)| > 1$.*

In other words, geometrically neighboring cells may differ at most by a single level of refinement, with the notion of cell neighborhood depending on the value of k . Fig. 1a illustrates a forest-of-quadtrees, with two quadtrees (i.e., $|C| = 2$), which is 2:1 0-balanced, i.e., balance across corners and facets. We note that, as a consequence of Prop. 2.7, and Def 2.8, the $\mathcal{T}_{K,f}^+$, and $\mathcal{T}_{K,f}^-$ sets for local n -faces f with $n \geq \max(k, 1)$, are composed of cells K' such that $\ell(K') = \ell(K) + 1$ and $\ell(K') = \ell(K) - 1$, resp. On the

other hand, if $n = k = 0$, then $\mathcal{T}_{K,f}$ may be composed of neighbors at the same, one unit higher, or one unit lower level of refinement than K ; see Fig. 1a.

Along Sect. 4 and 5, it will become clear why this constraint extremely simplifies the construction of conforming FE spaces, especially in a distributed-memory context. However, enforcing the 2:1 k -balance constraint is a non-trivial operation, that may require to refine more cells than were initially marked for refinement. In particular, inherent to this operation is the so-called *ripple-effect*: a cell may cause a ripple of refinements that can propagate even for cells which are not immediate neighbors of it. As a consequence, this operation is a nonlocal and inherently iterative process. Fortunately, much progress has been made in recent years in this direction, and fast and scalable parallel algorithms are available in the literature [8, 7]. In FE applications, the ease in implementation complexity comes at the price of increasing computational complexity (i.e., increasing the size of the FE problem at hand). However, our experiments in Sect. 7, reveal that, in the worst case scenario evaluated, at most a moderate 10% increment in the number of cells is required to enforce the balance constraint.

2.6. Forest-of-trees handler operations. In practice, parallel tree-based AMR using SFCs is a specialized feature that numerical applications typically outsource to an external adaptive meshing engine. In the interface among these, a set of core application-level operations have been identified as cornerstone in order to fully realize this functionality in numerical applications. These are briefly outlined in the sequel. (a) **New**: creates a new uniformly refined forest, up to a user-provided level, from a data structure describing the connectivity of mesh cells in C ; (b) **Adapt**: refines and coarsens the current forest accordingly to a user-provided criterion; (c) **Partition**: redistributes the forest leaves among processors for dynamic load balancing; (d) **Balance**: ensures the 2:1 k -balance condition among neighboring cells by local refinement where required; (e) **Ghost**: creates a data structure that contains the so called s -ghost cell set, i.e., a layer of off-processor leaf cells that are neighbors of cells in the current processor (see Sect. 3); (f) **Iterate**: given a 2:1 k -balanced forest, provides a mechanism to iterate over its leaf cells, and a *subset* of the inter-cell interfaces (i.e., n -faces such that $n \geq k$), while letting applications get local neighborhood information of each interface visited, including both conforming and non-conforming cell interfaces. We refer to [8, 9, 2] and references therein for a comprehensive presentation of the algorithms behind these operations, implementation details, and cost and communication volume analysis, among others.

2.7. State-of-the-art software packages. There are essentially two software packages around that are grounded on the tree-based AMR with SFCs approach. As an evidence of the potential of this approach, both libraries have been shown to efficiently scale up to hundreds of thousands of computational cores. On the one hand, **p4est** [8, 9] provides parallel forest-of-octrees (thus restricted to either quadrilaterals or hexahedra) grounded on the standard 1:2^d isotropic refinement rule and the Morton SFC index [8]. On the other hand, **t8code** [2] is an ambitious, on-going software effort, which provides abstract tree-based AMR with SFCs algorithms for the operations outlined above. By means of implementing a set of low-level local-to-cell operations, these algorithms may be extended to work with general polytopes (e.g., lines, simplices, bricks, prisms, pyramids, etc.). For triangles and tetrahedra, it is grounded on Bey's red-refinement rule, and the recently proposed TM SFC index [10]. In its current status, up to the authors' knowledge, **t8code** only provides facet-oriented variants of the main operations above (e.g., **Balance** is restricted to $d - 1$ -balance), and thus cannot be readily used for the implementation of generic FEs. For this reason, in this paper we restrict to **p4est** as a practical demonstrator.

3. A GENERAL FE-SUITABLE DISTRIBUTED ADAPTIVE MESH REPRESENTATION

As mentioned in Sect. 1, our approach for mesh handling follows a two-layered meshing approach, namely an inner light-weight layer encoding the forest-of-trees, already overviewed in Sect. 2, and an outer rich FE-suitable mesh representation, that we cover in this section. In particular, we present a specification for a distributed adaptive mesh representation that supports generic FE spaces built atop. The sort of

data it needs to handle, and how it is internally laid out, has resulted from the authors' experience in accommodating the requirements of a wide range of state-of-the-art FE discretizations within a single framework [27]. The concepts underlying this specification are not tailored to a particular tree-based AMR with SFCs technology or cell topology.

We recall that \mathcal{T} is a partition of Ω defined as the union of the forest leaves, and \mathcal{F} is the set of global VEFs of \mathcal{T} . For the software implementation of boundary conditions [27, Sect. 10.4], among others, it becomes necessary to identify those global VEFs that lay at the interior of Ω , and those at its boundary, denoted as $\partial\Omega$. We represent the former and latter sets of VEFs with \mathcal{F}_Ω , and $\mathcal{F}_{\partial\Omega}$, resp. Clearly, $\{\mathcal{F}_\Omega, \mathcal{F}_{\partial\Omega}\}$ is a partition of \mathcal{F} .

3.1. Cells and global VEFs adjacencies. Our adaptive mesh representation considers three different kinds of adjacency relations among cells and global VEFs. First, given a cell $K \in \mathcal{T}$, $[\mathcal{F}_K]$ keeps track of the global VEFs (equivalence classes) corresponding to the local VEFs of K . The remaining two relations are neighborship relations. Given a global VEF $F \in \mathcal{F}$, the set of cells around the VEF is defined as $\mathcal{T}_F \doteq \{K \in \mathcal{T} : F \in [\mathcal{F}_K]\}$. For our purposes, the definition of the *coarser cells around a global VEF* is essential.

Definition 3.1 (Coarser cells around a global VEF). *The set of coarser cells around a VEF $F \in \mathcal{F}$ is represented with $\tilde{\mathcal{T}}_F$ and is composed by the cells $K \in \mathcal{T}$ such that : (a) $F \notin [\mathcal{F}_K]$; (b) $F \subset \bar{K}$.*

The following propositions hold for the \mathcal{T}_F and $\tilde{\mathcal{T}}_F$ sets.

Proposition 3.2. *For any forest-of-trees \mathcal{T} we have that, for any n -face $F \in \mathcal{F}$ with $n > 0$ and any two cells $K, K' \in \mathcal{T}_F$, then $\ell(K) = \ell(K')$. A vertex, however, may be shared by cells with different levels of refinement.*

Proof. The proof follows the same lines as the one for Prop. 2.7, which relies on Ass. 2.6 (the uniformly refined tree is conforming) and Ass. 2.2 (n -faces are subdivided for $n > 0$). \square

Proposition 3.3. *For any forest-of-trees \mathcal{T} , $F \in \mathcal{F}$, $K \in \mathcal{T}_F$, and $K' \in \tilde{\mathcal{T}}_F$, it holds $\ell(K) > \ell(K')$.*

Proof. Cell levels must be different due to the conformity in Ass. 2.6. Let us assume that $\ell(K) < \ell(K')$. We can consider the recursive refinement of K till reaching $\ell(K')$. In this process, by the definition of the refinement rule, any n -face of K with $n > 0$ is being partitioned into a set of VEFs at level $\ell(K')$. By the conformity of the uniformly refined tree at $\ell(K')$, any VEF in $\mathcal{F}_{K'}$ in touch with \bar{K} is a strict subset of a VEF in \mathcal{F}_K or a vertex in \mathcal{F}_K . Thus, the situation in the statement is not possible; $F \in \mathcal{F}_K$ can only be a vertex in $\mathcal{F}_{K'}$ or it a strict superset of a VEF in K' (i.e., $K' \notin \tilde{\mathcal{T}}_F$ in either case). Thus, $\ell(K) > \ell(K')$ must hold. \square

The \mathcal{T}_F and $\tilde{\mathcal{T}}_F$ sets can be readily obtained from the neighbors of cells across local n -faces, as stated in the following propositions. For conciseness in the formulation of these propositions, we define the sets

$$\mathcal{S}_{K,F} \doteq \bigcup_{G \in [\mathcal{F}_K]: F \subset \bar{G}} \mathcal{T}_{K,G}, \quad \mathcal{S}_{K,F}^- \doteq \bigcup_{G \in [\mathcal{F}_K]: F \subset \bar{G}} \mathcal{T}_{K,G}^-, \quad \mathcal{S}_{K,F}^+ \doteq \left\{ K' \in \left(\bigcup_{G \in [\mathcal{F}_K]: F \subset \bar{G}} \mathcal{T}_{K,G}^+ \right) : F \in [\mathcal{F}_{K'}] \right\}.$$

Proposition 3.4. *Given a cell $K \in \mathcal{T}$ and $F \in [\mathcal{F}_K]$, then we have that:*

$$\mathcal{T}_F = \{K\} \cup \mathcal{S}_{K,F} \cup \mathcal{S}_{K,F}^+ \cup \{K' \in \mathcal{S}_{K,F}^- : F \in [\mathcal{F}_{K'}]\}, \quad (1)$$

$$\tilde{\mathcal{T}}_F = \{K' \in \mathcal{S}_{K,F}^- : F \notin [\mathcal{F}_{K'}]\}. \quad (2)$$

Proof. Let us represent the right-hand side of (1) (resp., (2)) with \mathcal{S}_F (resp., $\tilde{\mathcal{S}}_F$). We want to prove that $\mathcal{T}_F = \mathcal{S}_F$ and $\tilde{\mathcal{T}}_F = \tilde{\mathcal{S}}_F$. First, we note that $\mathcal{T}_F \cup \tilde{\mathcal{T}}_F$ is the set of all cells $K' \in \mathcal{T}$ such that $F \subset \bar{K}'$. On the other hand, $\mathcal{S}_F \cup \tilde{\mathcal{S}}_F = \{K\} \cup \mathcal{S}_{K,F} \cup \mathcal{S}_{K,F}^+ \cup \mathcal{S}_{K,F}^-$ spans the same set of cells, which can be proved

using Def. 2.5. Thus, $\mathcal{T}_F \cup \tilde{\mathcal{T}}_F = \mathcal{S}_F \cup \tilde{\mathcal{S}}_F$. On the other hand, one can check that $\mathcal{S}_F \subseteq \mathcal{T}_F$, since: (a) $K \in \mathcal{T}_F$ by hypothesis, (b) $\mathcal{S}_{K,F} \subseteq \mathcal{T}_F$ by Def. 2.5, and (c) $\mathcal{S}_{K,F}^+$ and the fourth set in the definition of \mathcal{S}_F are subsets of \mathcal{T}_F by definition. One can also readily check that $\tilde{\mathcal{S}}_F \cap \mathcal{T}_F = \emptyset$ by its definition. Combining these results, we prove that $\mathcal{T}_F = \mathcal{S}_F$ and $\tilde{\mathcal{T}}_F = \tilde{\mathcal{S}}_F$. \square

Proposition 3.5. *Given an n -face $F \in \mathcal{F}$, with $n > 0$, and a cell $K \in \mathcal{T}$ such that $F \in [\mathcal{F}_K]$, then we have that: (a) $\mathcal{T}_F = \mathcal{S}_{K,F} \cup \{K\}$; and (b) $\tilde{\mathcal{T}}_F = \mathcal{S}_{K,F}^-$.*

Proof. Let us prove that for n -faces F with $n > 0$, it holds: (1) $\mathcal{S}_{K,F}^+ = \emptyset$ and (2) $\{K' \in \mathcal{S}_{K,F}^- : F \in [\mathcal{F}_{K'}]\} = \emptyset$ and $\{K' \in \mathcal{S}_{K,F}^- : F \notin [\mathcal{F}_{K'}]\} = \mathcal{S}_{K,F}^-$. We prove (1) using the same argument as in Prop. 3.3; a cell $K' \in \mathcal{S}_{K,F}^+$ belongs to $\mathcal{T}_{K,G}^+$ for $G \in [\mathcal{F}_K]$. It holds $\ell(K') > \ell(K)$ by Prop. 2.7, since G is an n -face with $n > 0$. On the other hand, $F \notin [\mathcal{F}_K] \cap [\mathcal{F}_{K'}]$ by using the fact that n -faces with $n > 0$ are partitioned with refinement. Thus, $\mathcal{S}_{K,F}^+$ is empty. The same arguments can be readily used to prove (2). Combining these results and Prop. 3.4, we end the proof. \square

Proposition 3.6. *Given a 2:1 k -balanced forest-of-trees \mathcal{T} and an n -face $F \in \mathcal{F}$ with $n \geq k$, $K \in \mathcal{T}_F$, and $K' \in \tilde{\mathcal{T}}_F$, it holds $\ell(K) = \ell(K') + 1$.*

Proof. It is a direct consequence of Def. 2.8 and Props. 3.3, 3.4, and 3.5. \square

The set $\tilde{\mathcal{T}}_F$ can be used to classify \mathcal{F} into *regular* and *hanging* VEFs.

Definition 3.7 (Regular and hanging VEFs). *A VEF is regular if $\tilde{\mathcal{T}}_F = \emptyset$ and hanging otherwise.*

We denote as \mathcal{F}_R and \mathcal{F}_H the set of regular and hanging VEFs, resp. Clearly, $\{\mathcal{F}_R, \mathcal{F}_H\}$ is a partition of \mathcal{F} .

Proposition 3.8. *Let us consider a 2:1 k -balanced mesh \mathcal{T} and an n -face $F \in \mathcal{F}_H$ with $n \geq k$. For any $K \in \mathcal{T}$ with a local VEF $f \in \mathcal{F}_K$ satisfying $[f] = F$, there exists $K' \in \mathcal{T}$ and $g \in \mathcal{F}_{K'}$ such that $[g] = [O_f]$. Thus, one can define the owner VEF of F as $O_F \doteq [g]$ and $F \subsetneq O_F$. If F is a vertex, it also holds for a 2:1 1-balanced mesh.*

Proof. Due to Prop. 2.7 and the 2:1 k -balance, since the VEF F is hanging, it is in touch with a cell K' at level $\ell(K) - 1$. By definition, $O_f \in \mathcal{F}_{K_p}$ where K_p is the parent cell of K . By conformity of the uniformly refined tree at level $\ell(K) - 1$, there exists a $g \in \mathcal{F}_{K'}$ such that $[O_f] = [g]$. Thus, one can define O_F as $[g] \in \mathcal{F}$.

If $F \in \mathcal{F}_H$ is a vertex, there is a $K' \in \mathcal{T}$ such that $F \subset \overline{K'}$ but $F \notin [\mathcal{F}_{K'}]$. For any $K \in \mathcal{T}$, $f \in \mathcal{F}_K$ such that $[f] = F$, we consider the ancestor K_p of K at level $\ell(K')$. By conformity of the uniformly refined tree, we can pick $g \in \mathcal{F}_{K_p}$ and $g' \in \mathcal{F}_{K'}$ such that $\overline{K_p} \cap \overline{K'} = \bar{g} = \bar{g}'$. Clearly, $f \subsetneq g'$ (otherwise $f \in \mathcal{F}_{K'}$) and thus, g and g' are n -faces with $n > 0$. By the 2:1 1-balance, $\ell(K') = \ell(K_p) = \ell(K) - 1$. As a result, $O_f \in \mathcal{F}_{K_p}$ (i.e., K_p is the parent cell), O_f is a VEF of g (by the polytope definition), and there exists an $h \in \mathcal{F}_{K'}$ such that $[h] = [O_f]$. Therefore, $[O_f] \in \mathcal{F}$ and we can define $O_F \doteq [O_f]$. \square

Proposition 3.9. *Given a 2:1 k -balanced mesh \mathcal{T} and $F \in \mathcal{F}_H$ such that its owner VEF O_F is an n -face with $n \geq k$, then the sets $\tilde{\mathcal{T}}_F$ and \mathcal{T}_{O_F} are identical (or, equivalently, $O_F \in \mathcal{F}$).*

Proof. It is a direct consequence of the construction of O_F in Prop. 3.8. \square

The following *key property* holds for hanging VEFs.

Proposition 3.10. *Given a 2:1 k -balanced forest-of-trees mesh \mathcal{T} and $F \in \mathcal{F}_H$ such that O_F is an n -face with $n \geq k$, it holds that $O_F \in \mathcal{F}_R$. Besides, any n -face E of $\overline{O_F}$ with $n \geq k$ is also regular, i.e., $E \in \mathcal{F}_R$.*

Proof. If $F \in \mathcal{F}_H$, there exists a $K \in \mathcal{T}_F$ and a $K' \in \tilde{\mathcal{T}}_F$. If $O_F \in \mathcal{F}_H$, it would be in touch with a cell in level $\ell(K') - 1$ due to Prop. 3.3. It contradicts the 2:1 k -balanced assumption. Thus, $O_F \in \mathcal{F}_R$. Finally, an n -face E of O_F belongs to $\mathcal{F}_{K'}$; the n -faces of an n -face of K are n -faces of K by construction of a polytope. Thus, it is easy to check that E is in touch with K' and a sibling cell of K . As above, if E would be hanging, it would violate the 2:1 k -balance. \square

As a consequence of Prop. 3.10, one enjoys the following two cornerstone benefits: (a) *only* single-level (a.k.a. direct) hanging node constraints are required when constructing global conforming FE spaces; (b) in a distributed-memory context, all such constraints can be resolved locally with a single layer of ghost cells. This will be revisited in Sect. 4.2. A stronger result is proved below for 2:1 1-balanced forest-of-trees.

Proposition 3.11. *Given a 2:1 1-balanced forest-of-trees mesh \mathcal{T} and $F \in \mathcal{F}_H$, then $O_F \in \mathcal{F}_R$. Besides, any n-face E of $\overline{O_F}$ is also regular, i.e., $E \in \mathcal{F}_R$.*

Proof. To prove the first result, we must show that $O_V \in \mathcal{F}_R$ for any vertex $V \in \mathcal{F}_H$. The result for O_V being an n-face with $n > 0$ has been proved in Prop. 3.10. If O_V would be a vertex, $V = O_V$ and thus $O_V \in \mathcal{F}_H$. Thus, V would belong to a cell K and its parent cell in $\ell(K) - 1$ would be in touch with a cell K' in $\ell(K) - 2$ or lower in $\tilde{\mathcal{T}}_V$. Thus, there would be an n-face $G \in [\mathcal{F}_{K'}]$ such that $V \subsetneq G$, thus $n > 0$. Thus, G would violate the 2:1 1-balance property. It proves the first result and also shows that hanging vertices are always 2:1 balanced. Using this fact, we can prove the second result as in Prop. 3.10. \square

Prop. 3.11 has very important practical implications. In particular, for those global conforming FE spaces for which vertices carry out DOFs (e.g., those constructed from Lagrangian FEs), 2:1 0-balanced trees are not required to have (a) and (b) above, only 2:1 1-balance is required. Provided this is acceptable for the problem at hand from a numerical point of view, 2:1 1-balance may lead to computational savings, as enforcing it requires less balance refinement than 0-balance in general (see Sect. 2.5). This result seems to have gone unnoticed/neglected and not made clear in the literature [34, 8, 9]. For example, the deal.II software, which uses p4est as its inner layer forest-of-trees meshing engine, always instructs it to build 2:1 0-balanced meshes, i.e., k is not a parameter the user can play around with.⁵

3.2. Distributed-memory context. With the definitions associated to the global adaptive mesh \mathcal{T} so far, we can readily discuss *its representation in a parallel distributed-memory context*. All cells $K \in \mathcal{T}$ are assigned a *processor owner* $p = 1, \dots, P$, with P being the number of processors involved in the parallel computation. Given this cell-to-processor ownership mapping, the local portion of the global mesh \mathcal{T} that processor p stores locally, referred to as $\mathcal{T}^p \subset \mathcal{T}$, is defined as the union of two *disjoint* sets of cells, i.e., $\mathcal{T}^p \doteq \mathcal{T}_L^p \cup \mathcal{T}_G^p$, $\mathcal{T}_L^p \cap \mathcal{T}_G^p = \emptyset$, with \mathcal{T}_L^p being the *local* cells set, and \mathcal{T}_G^p the so-called *ghost* cells set. The former includes those cells that processor p owns. By construction, the set $\{\mathcal{T}_L^p\}$, for $p = 1, \dots, P$, is a partition of \mathcal{T} . The formal definition of the *ghost* cells set is also parameterized by a parameter s , with $0 \leq s < d$.

Definition 3.12. *The s -ghost cell set includes all cells $K \in \mathcal{T} \setminus \mathcal{T}_L^p$ (i.e., off-processor cells) that are neighbors of cells in \mathcal{T}_L^p across n-faces with $n \geq s$ (see Def 2.5).*

The sets \mathcal{T}_L^p and \mathcal{T}_G^p with $s = 0$, for $p = 1, 2$, are illustrated in Fig. 1b, 1c, resp., for the forest-of-quadtrees in Fig. 1a. We note that a given cell in \mathcal{T} belongs to only one \mathcal{T}_L^p set. It, however, belongs to either none, one, or more \mathcal{T}_G^p sets.

There are two highly relevant questions regarding the 2:1 k -balance and the s -ghost cell set that this paper answers along the way. First, what must a FE formulation fulfill such that it can be implemented in a distributed-memory framework given the constraint that each processor only has access to the off-processor cells in the s -ghost cell set? Second, which is the largest possible value of k and s that lets one still implement such FE formulation? The answer to both questions can be found in Sect. 4.

Since a given processor only stores a local portion of the global mesh, i.e., \mathcal{T}^p , it can only store a subset of the global VEFs set \mathcal{F} , denoted as $\mathcal{F}^p \subset \mathcal{F}$, and referred to as the proc-local VEFs set. \mathcal{F}^p is generated by gluing together the local VEFs that lay at the boundary of (the local and ghost) cells in \mathcal{T}^p . Likewise the proc-local restriction of \mathcal{F}_Ω and $\mathcal{F}_{\partial\Omega}$ to \mathcal{T}^p is denoted as $\mathcal{F}_\Omega^p \doteq \mathcal{F}_\Omega \cap \mathcal{F}^p$ and $\mathcal{F}_{\partial\Omega}^p \doteq \mathcal{F}_{\partial\Omega} \cap \mathcal{F}^p$,

⁵See line 2873 of https://www.dealii.org/9.0.1/doxygen/deal.II/distributed_2tria_8cc_source.html for more details.

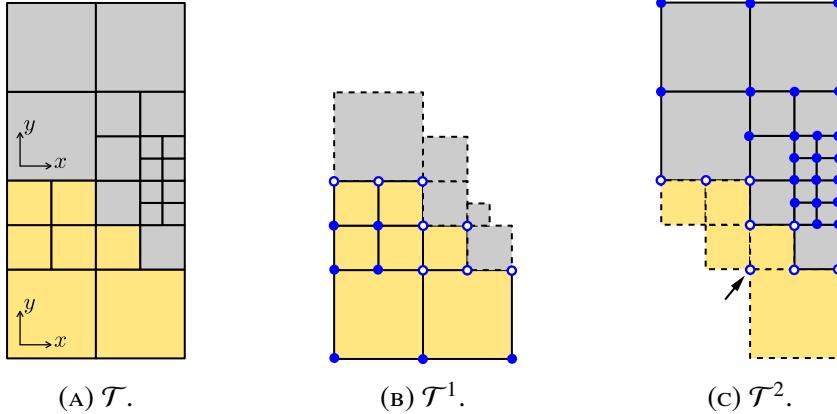


FIGURE 1. 2:1 0-balanced forest-of-quadtrees mesh with two quadtrees (i.e., $|C| = 2$) distributed (non-uniformly) among two processors (i.e., $P = 2$) with 0-ghost cells, 1:4 refinement and the Morton SFC index [8]. Cells in \mathcal{T}_L^p are depicted with continuous boundary lines, while those in the 0-ghost layer \mathcal{T}_G^p with dashed ones. All vertices in \mathcal{F}_L^p are depicted with circles. In particular, vertices in \mathcal{F}_I^p are depicted with blue circles, whereas vertices in \mathcal{F}_Γ^p are depicted with unfilled blue circles. The mesh vertex in \mathcal{T}^2 pointed by the arrow is in \mathcal{F}_L^p as it fulfills Def. 3.13 (d).

resp. The sets $\mathcal{T}_{K,f}$, $\mathcal{T}_{K,f}^-$, and $\mathcal{T}_{K,f}^+$ are also restricted to cells in \mathcal{T}^p . We denote their restriction as $\mathcal{T}_{K,f}^p \doteq \mathcal{T}_{K,f} \cap \mathcal{T}^p$, $\mathcal{T}_{K,f}^{p,+} \doteq \mathcal{T}^p \cap \mathcal{T}_{K,f}^+$, and $\mathcal{T}_{K,f}^{p,-} \doteq \mathcal{T}^p \cap \mathcal{T}_{K,f}^-$, resp., for all $K \in \mathcal{T}^p$. It turns out that, for all cells $K \in \mathcal{T}_L^p$, and local n -faces $f \in \mathcal{F}_K$ such that $n \geq s$, the processor-local and global sets are equivalent by Def. 3.12. Likewise, \mathcal{T}_F and $\tilde{\mathcal{T}}_F$ are restricted to the cells in \mathcal{T}^p . For a given proc-local VEF $F \in \mathcal{F}^p$, we denote their restriction as $\mathcal{T}_F^p \doteq \mathcal{T}_F \cap \mathcal{T}^p$ and $\tilde{\mathcal{T}}_F^p \doteq \tilde{\mathcal{T}}_F \cap \mathcal{T}^p$, resp. By the equivalences in Props. 3.5, 3.4, we have that $\mathcal{T}_F^p = \mathcal{T}_F$ and $\tilde{\mathcal{T}}_F^p = \tilde{\mathcal{T}}_F$ for those n -faces $F \in \mathcal{F}^p$ such that $n \geq s$, and there is at least one local cell in \mathcal{T}_F^p . As shown along the paper, this equivalence has key implications.

The set of proc-local VEFs \mathcal{F}^p , as in the case of cells, can be split into two disjoint sets of VEFs, i.e., $\mathcal{F}^p \doteq \mathcal{F}_L^p \cup \mathcal{F}_G^p$. This separation into the \mathcal{F}_L^p and \mathcal{F}_G^p sets is, however, of different nature compared to that of the cells.

Definition 3.13 (\mathcal{F}_L^p and \mathcal{F}_G^p sets). A VEF $F \in \mathcal{F}^p$ is in \mathcal{F}_L^p if it fulfills one of the four following conditions:
(a) at least one of the cells in \mathcal{T}_F^p is local, i.e., $\mathcal{T}_F^p \cap \mathcal{T}_L^p \neq \emptyset$; (b) all cells in \mathcal{T}_F^p are ghost but at least one cell in $\tilde{\mathcal{T}}_F^p$ is local, i.e., $\tilde{\mathcal{T}}_F^p \cap \mathcal{T}_L^p \neq \emptyset$; (c) all cells in \mathcal{T}_F^p are ghost and there exists $F' \in \mathcal{F}_L^p$ such that $F = O_{F'}$; (d) all cells in \mathcal{T}_F^p are ghost and F lays at the boundary of a VEF that fulfills (c). It is in \mathcal{F}_G^p otherwise.

Thus, a global VEF $F \in \mathcal{F}$ might be in \mathcal{F}_L^p at multiple processors, while a local cell cannot. We will refer to this kind of VEFs as *interface* VEFs, while we will use the term *interior* to refer to those VEFs which are in \mathcal{F}_L^p only at a single processor. We denote the former and latter sets as \mathcal{F}_Γ^p and \mathcal{F}_I^p , resp. Clearly, $\{\mathcal{F}_I^p, \mathcal{F}_\Gamma^p\}$ is a partition of \mathcal{F}_L^p . Those vertices in the sets \mathcal{F}_L^p , \mathcal{F}_I^p , and \mathcal{F}_Γ^p for the forest-of-quadtrees in Fig. 1a distributed among two processors, are illustrated in Fig. 1b, 1c for $p = 1, 2$, resp. It remains to define a classification of VEFs of \mathcal{T}^p into sets of regular and hanging VEFs suitable in a distributed context, i.e., that only requires processor-local information. With this aim, we define the concept of proc-regular and proc-hanging VEFs.

Definition 3.14 (Proc-regular and proc-hanging VEFs). A VEF $F \in \mathcal{F}^p$ is proc-regular if $\tilde{\mathcal{T}}_F^p = \emptyset$, and proc-hanging otherwise.

The set of proc-regular and proc-hanging VEFs are denoted as \mathcal{F}_R^p and \mathcal{F}_H^p , resp. We stress that, for VEFs in \mathcal{F}_G^p , this definition of proc-regular (resp., proc-hanging) VEFs is not equivalent to regular (resp., hanging) VEFs. In Fig. 2a, 2b, 2c, 2d, we illustrate the \mathcal{F}_R^p and \mathcal{F}_H^p sets, with $p = 1, 2$, resp., for the forest-of-quadtrees in Fig. 1a. The ghost vertex and ghost face pointed by an arrow in Fig. 2c and 2d are such that they belong to \mathcal{F}_H in \mathcal{T} but to \mathcal{F}_R^p in \mathcal{T}^p . Fortunately, the algorithms that run on top of the mesh *only* require these definitions to be equivalent for n -faces in \mathcal{F}_L^p , where the values of n are again determined by the FE space at hand. This invariant holds, as proved in the next proposition.

Proposition 3.15. *For a 2:1 k -balanced forest-of-trees mesh \mathcal{T} with k -ghost cells, then $\mathcal{F}_R^p \cap \mathcal{F}_L^p = \mathcal{F}_R \cap \mathcal{F}_L^p$ and $\mathcal{F}_H^p \cap \mathcal{F}_L^p = \mathcal{F}_H \cap \mathcal{F}_L^p$ for n -faces F with $n \geq k$. This statement also holds for 0-faces when $k = 1$ (i.e., 2:1 1-balance and 1-ghost cell set).*

Proof. Let us first consider the case $F \in \mathcal{F}_R^p \cap \mathcal{F}_L^p$. As $F \in \mathcal{F}_R^p$, then $\tilde{\mathcal{T}}_F^p = \emptyset$. Thus, $F \in \mathcal{F}_L^p$ because of (a), (c), or (d) in Def. 3.13 (i.e., it cannot be in \mathcal{F}_L^p due to (b)). Let us assume that $F \in \mathcal{F}_L^p$ because of (a) in Def. 3.13, i.e., at least one of the cells in \mathcal{T}_F^p is local. For the first proposition statement, as F fulfills (a), then $\mathcal{T}_F = \mathcal{T}_F^p$ and $\tilde{\mathcal{T}}_F = \tilde{\mathcal{T}}_F^p = \emptyset$ for those n -faces F for which $n \geq k$ (see discussion above where the \mathcal{T}_F^p and $\tilde{\mathcal{T}}_F^p$ sets are defined). For the second proposition statement, i.e., 0-faces F and $k = 1$, we have that $\tilde{\mathcal{T}}_F = \tilde{\mathcal{T}}_F^p = \emptyset$ due to Prop. 3.4. Thus, $F \in \mathcal{F}_R$ in both cases. If $F \in \mathcal{F}_R^p$ is in \mathcal{F}_L^p because it holds (c) or (d) in Def. 3.13, $F \in \mathcal{F}_R$ due to Prop. 3.10 in the case of the first proposition statement, and due to Prop. 3.11 in the case of the second. On the other hand, if $F \in \mathcal{F}_R \cap \mathcal{F}_L^p$ it is clearly in $\mathcal{F}_R^p \cap \mathcal{F}_L^p$, thus $\mathcal{F}_R^p \cap \mathcal{F}_L^p = \mathcal{F}_R \cap \mathcal{F}_L^p$. Using the fact that both $\{\mathcal{F}_R^p \cap \mathcal{F}_L^p, \mathcal{F}_H^p \cap \mathcal{F}_L^p\}$ and $\{\mathcal{F}_R \cap \mathcal{F}_L^p, \mathcal{F}_H \cap \mathcal{F}_L^p\}$ are partitions of \mathcal{F}_L^p , we readily have that $\mathcal{F}_H^p \cap \mathcal{F}_L^p = \mathcal{F}_H \cap \mathcal{F}_L^p$. \square

We note that, with the proc-local definitions for the cell and VEF sets presented along the section, we enjoy two benefits (see Sect. 3.4): (1) serial code bound to \mathcal{T} can be re-used verbatim in parallel computations at each \mathcal{T}^p separately; (2) \mathcal{F}_R^p and \mathcal{F}_H^p can be determined at each processor without any inter-processor communication. The only cells/VEFs for which this locally computed mesh information does not coincide with the global one are not required in practice for the generation of global FE spaces and their numerical integration (see Sect. 4). These benefits are precluded with the alternative definition of these sets based on the restriction of \mathcal{F}_R and \mathcal{F}_H to \mathcal{F}^p .

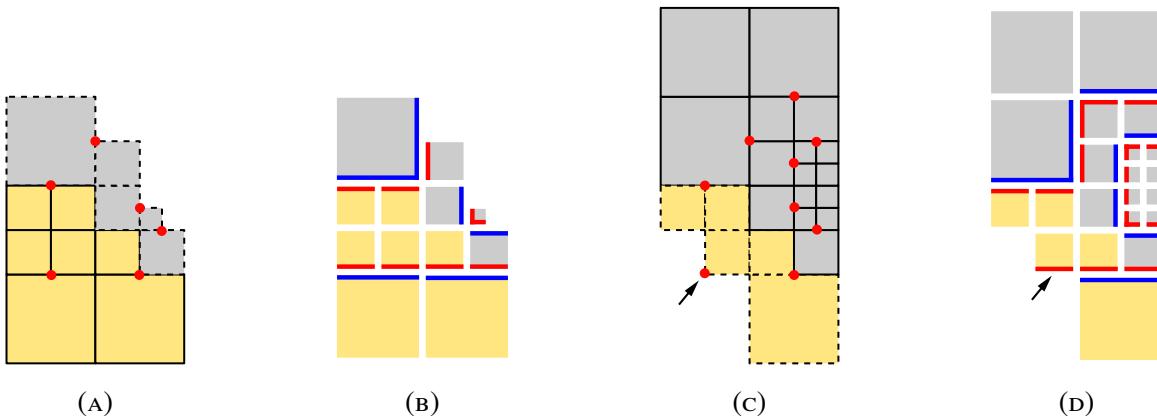


FIGURE 2. Hanging vertices (red circles) and faces (red lines) in \mathcal{T}^1 (a and b, resp.) and \mathcal{T}^2 (c and d, resp.). Those regular faces that are the owner VEF of a hanging VEF are depicted in blue. The hanging VEFs pointed out by the arrows in \mathcal{T}^2 are two examples of ghost VEFs that, while being hanging in \mathcal{T} , cannot be identified as such by processor $p = 2$, i.e., they are proc-regular.

3.3. Computer representation. Local VEFs f in \mathcal{F}_K are represented in the computer with a cell-local VEF identifier within the range $1, \dots, |\mathcal{F}_K|$. If \mathcal{T} is *not* distributed, then global VEFs F in \mathcal{F} are labelled with a global VEF identifier within $1, \dots, |\mathcal{F}|$. The equivalence classes corresponding to local VEFs, i.e., $[\cdot]$, are represented with a cell-wise array that stores the local-to-global index map. If \mathcal{T} is distributed, it is not needed in practice to define a global indexing for global VEFs. Instead, $[\cdot]$ provides a processor-wise VEF index within $1, \dots, |\mathcal{F}^p|$. For VEFs on subdomain interfaces, this processor-wise VEF index must be combined with the overlapped mesh partition information to implement the equivalence class of VEFs (see Remark 5.3). For the sake of simplicity, we do not make explicit this implementation detail in the following exposition. In any case, this observation allows us to reduce parallel overhead by eliminating the communications among processors required to generate a global VEF indexing. Furthermore, using proc-local VEF identifiers, to determine whether a given proc-local VEF $F \in \mathcal{F}^p$ is in a given set, say \mathcal{F}_H^p , can be implemented *in constant time*, whereas with global VEF identifiers, one has to pay the cost of a search in a global index set.

3.4. Construction of FE-suitable distributed adaptive mesh data structure. Alg. 1 shows a simplified pseudocode of the process in charge of building the distributed adaptive mesh representation covered so far. The algorithm assumes the global forest-of-trees mesh \mathcal{T} to be 2:1 k -balanced, with k being a user-level parameter to be set up accordingly to Prop. 4.3; see Sect. 4.2. The input of Alg. 1 are the proc-local variants of the cell neighbors across local n -faces sets presented in Sect 2.4. From this input, Alg. 1 glues together the local VEFs laying at the boundary of the mesh cells in \mathcal{T}^p that satisfy the condition in Line 4, thus generating: (a) $[\mathcal{F}_K]$, for $K \in \mathcal{T}^p$; (b) the proc-local \mathcal{F}^p , \mathcal{F}_H^p and \mathcal{F}_R^p VEFs sets; (c) the proc-local \mathcal{T}_F^p and $\tilde{\mathcal{T}}_F^p$ sets; (d) O_F for all $F \in \mathcal{F}_H^p$. Although omitted from Alg. 1 in order to keep the presentation short, the actual algorithm also builds the \mathcal{F}_Ω^p , $\mathcal{F}_{\partial\Omega}^p$, \mathcal{F}_L^p , \mathcal{F}_G^p , \mathcal{F}_I^p , and \mathcal{F}_Γ^p sets (see Sect. 3.2). The correctness of Alg. 1 is mathematically supported by the propositions in Sect. 3.1 and 3.2. For example, Line 31 of Alg. 1, triggers the generation of the \mathcal{T}_F^p and $\tilde{\mathcal{T}}_F^p$ sets for the owner VEF of a hanging VEF (if the former has not been visited yet). By Prop. 3.9, we know that the owner VEF exists in \mathcal{F}^p , and due to Prop. 3.10, that the depth of the recursive call in Line 31 is always one, as the owner VEF of a hanging VEF is always a regular VEF (i.e., $\tilde{\mathcal{T}}_{O_F}^p = \emptyset$). Besides, due to Prop. 3.9, Line 32 is correct. In a distributed-memory context, each processor p executes its own instance of Alg. 1, *with no communication at all involved among processors*. Line 6 in Alg. 1 essentially amounts to generating a new proc-local VEF identifier from the previously generated one (starting from one locally in each processor), resulting in a local-to-proc-local index map $[\cdot]$; see Sect.3.3.

Alg. 1, at several points, has to determine the local VEF \hat{f} of a neighbor cell \hat{K} of K across g that either satisfies $\hat{f} \sim f$ or $\hat{f} \sim O_f$, with $g \in \mathcal{F}_K$ such that $f \subset \bar{g}$. This becomes a requirement that the forest-of-trees inner layer meshing engine has to be able to fulfill. p4est, as a forest-of-octrees engine to be plugged into our generic FE framework, satisfies this requirement as follows. When $g = f$, the data structures encoding $\mathcal{T}_{K,f}^p$ and $\mathcal{T}_{K,f}^{p,-}$ in p4est provide \hat{f} directly. However, when f lays at the boundary of g , and \hat{K} is a conformal cell neighbour of K across g , to determine \hat{f} is more involved, as it requires a change of basis among the coordinate systems \vec{g} and $\hat{\vec{g}}$ (using the algorithm in [8, Sect. 2.2]), with $\hat{g} \in \mathcal{F}_{\hat{K}}$ such that $\hat{g} \sim g$.⁶ Such change of basis is in particular required when g (and thus f) is an inter-octree n -face, as the coordinate systems of cells in the coarse mesh might not be aligned; see Sect. 2.1.

Remark 3.16. *Alg. 1 follows a fundamentally different approach from its counterpart in deal.II [34, Fig. 1], and imposes a different set of requirements to the tree-based AMR mesh engine. This latter algorithm matches recursively the forest-of-octrees within the deal.II mesh data structure and the one in p4est. In order to do so, it starts from the root octants of the forest, and by means of a set of queries to p4est, reconstructs the local part of the mesh by means of a full hierarchical AMR process in which the mesh is incrementally transformed by means of successive refinement and coarsening steps, until the*

⁶We note that, when f lays at the boundary of g , $f \neq g$, and \hat{K} is a lower level neighbour of K across g , then p4est provides $\hat{f} \sim O_f$ as well, as $O_f = O_g$; see Line 30 of Alg.1.

Algorithm 1: Construction of FE-suitable adaptive mesh from cell neighbors across n -faces.

```

1  $\mathcal{F}_R^P \leftarrow \emptyset; \mathcal{F}_H^P \leftarrow \emptyset$ 
2 for  $K \in \mathcal{T}^P$  do                                /* loop over all cells in the local portion of processor  $p$  */
3   for  $f \in \mathcal{F}_K$  do                      /* loop over current cell local VEFs */
4     if ( $f$  is a 0-face and  $k \leq 1$ ) or ( $f$  is an  $n$ -face with  $n \geq k$ ) then
5       if  $[f]$  not created yet then                /* current VEF first visited from any cell */
6          $[f] \leftarrow F \leftarrow \text{new\_proc\_local\_vef}();$           /* create new VEF equivalence class */
7          $\mathcal{T}_F^P \leftarrow \{K\}; \tilde{\mathcal{T}}_F^P \leftarrow \emptyset$            /* initialize the  $\mathcal{T}_F^P$  and  $\tilde{\mathcal{T}}_F^P$  sets */
8         for  $g \in \mathcal{F}_K : f \subset g$  do           /* Generate  $\mathcal{T}_F^P$  (Lines 8-28) */
9           if  $\mathcal{T}_{K,g}^P \neq \emptyset$  then
10            for  $\hat{K} \in \mathcal{T}_{K,g}^P$  do             /* loop over conformal neighbors of  $K$  across  $g$  */
11               $[\hat{f}] \leftarrow F$ , with  $\hat{f} \in \mathcal{F}_{\hat{K}} : \hat{f} \sim f$ ;  $\mathcal{T}_F^P \leftarrow \mathcal{T}_F^P \cup \{\hat{K}\}$ 
12            end
13          end
14          if  $\mathcal{T}_{K,g}^{P,-} \neq \emptyset$  then
15            if  $f$  is a 0-face then
16              for  $\hat{K} \in \mathcal{T}_{K,g}^{P,-} : \exists \hat{f} \in \mathcal{F}_{\hat{K}}$  satisfying  $\hat{f} \sim f$  do
17                 $[\hat{f}] \leftarrow F; \mathcal{T}_F^P \leftarrow \mathcal{T}_F^P \cup \{\hat{K}\}$ 
18              end
19            end
20          end
21          if  $\mathcal{T}_{K,g}^{P,+} \neq \emptyset$  then
22            if  $f$  is a 0-face then
23              for  $\hat{K} \in \mathcal{T}_{K,g}^{P,+} : \exists \hat{f} \in \mathcal{F}_{\hat{K}}$  satisfying  $\hat{f} \sim f$  do
24                 $[\hat{f}] \leftarrow F; \mathcal{T}_F^P \leftarrow \mathcal{T}_F^P \cup \{\hat{K}\}$ 
25              end
26            end
27          end
28        end
29        if  $\{K' \in \mathcal{S}_{K,f}^- : \nexists f' \in \mathcal{F}_{K'}, f' \sim f\} \neq \emptyset$  then      /* is  $f$  hanging? Generate  $\tilde{\mathcal{T}}_F^P$  (Lines 29-34) */
30          Let  $\hat{K}$  be an arbitrary neighbor cell in  $\{K' \in \mathcal{S}_{K,f}^- : \nexists f' \in \mathcal{F}_{K'}, f' \sim f\}$  and  $\hat{f} \in \mathcal{F}_{\hat{K}}$  such that  $\hat{f} \sim O_f$ 
31          Execute lines 5-36 replacing  $f \equiv \hat{f}$  and  $K \equiv \hat{K}$ 
32           $\tilde{\mathcal{T}}_F^P \leftarrow \mathcal{T}_{O_f}^P$ 
33           $O_F \leftarrow [O_f]$ 
34        end
35         $(\tilde{\mathcal{T}}_F^P = \emptyset) ? \mathcal{F}_R^P \leftarrow \mathcal{F}_R^P \cup \{F\} : \mathcal{F}_H^P \leftarrow \mathcal{F}_H^P \cup \{F\}$ 
36      end
37    end
38  end
39 end
40  $\mathcal{F}^P \leftarrow \mathcal{F}_R^P \cup \mathcal{F}_H^P$ 

```

leaves in the former match those in the latter. At each step, the aforementioned set of queries lets `deal.II` to determine which cells in the current mesh have to be refined and coarsened in the path towards the final match. While this approach takes into account the prior state of the forest-of-octrees within the `deal.II` mesh data structure, and thus does not need to actually perform changes into the data structure if no changes have been performed in its `p4est` counterpart, it still has to go over a full hierarchical AMR process to determine whether there is a match or not. Besides, when significant changes have been performed in `p4est`, it turns to be a quite complex, harder-to-update data structure (i.e., cell hierarchy, n -faces hierarchy, etc.), than the one introduced in this paper. As shown in Sect. 7, the implementation of

the latter data structure in FEMPAR turns out to be significantly faster than the one in deal.II, even if Alg. 1 (as now conceived) does not try to exploit the previous state of the mesh data structure.

Remark 3.17 (Implementation remark). *Alg. 1 can be implemented for forest-of-octrees using p4est as the inner layer mesh engine. In particular, this library offers the so-called pXest_mesh_t data structure, with X=4,8 for d = 2, 3, resp. This data structure provides neighbors across vertices, edges, and faces laid out into separate global arrays. pXest_mesh_t is constructed by means of the so-called universal mesh topology iterators, introduced in [9]; see also Iterate in Sect. 2.6. As a consequence, pXest_mesh_t only provides neighbors for cells in \mathcal{T}_L^P . However, Alg. 1 requires these for cells in \mathcal{T}_G^P as well. To this end, we developed a nearest neighbors exchange communication stage to complete pXest_mesh_t prior to the actual execution of Alg. 1. Apart from this, it turns out that, in p4est v2.0 (latest stable release at the date of writing), we have the following for the p8est_mesh_t data structure: (a) it lacks for the cell neighbors across edges sets; (b) vertices hanging on a coarse edge across which there are vertex neighbors satisfying Def. 2.5 are missing. Fortunately, we could obtain them using edge topology iterators [9].*

3.5. Mesh handling primitives. Users of our FE framework are provided with a distributed mesh data structure that hides the high complexity underlying the specialized tree-based AMR engine. To this end, it is equipped with three high-level mesh handling primitives:

- **Create.** Creates a new, yet unadapted forest-of-trees mesh provided a coarse mesh C . The coarse mesh can be either described manually by the user for simple domains, or generated by an unstructured mesh generator and imported into the framework.
- **Refine_and_coarsen.** Adapts the forest-of-trees mesh based on cell flags set by the user, while transferring data that the user might have attached to the mesh objects (i.e., cells and VEFs) to the new mesh objects generated after mesh adaptation. The mesh data structure provides a set of mechanisms that, prior to the execution of this primitive, lets them walk through over the mesh cells and associate a per-cell flag that tells the framework whether a cell is to be refined, coarsened, or kept as it is.
- **Redistribute.** Dynamically balances the computational load by redistributing the forest-of-trees mesh among processors involved in the parallel simulation. The default criteria is to balance the number of cells in each processor. Alternatively, the user might associate to each cell a partition weight. In this case, the primitive balances the sums of the cell partition weights among processors. The data that the user might have attached to the mesh objects (i.e., cells and VEFs) is also migrated among processors. This latter operation only involves non-blocking point-to-point message exchanges among processors (see discussion below).

The implementation of these mesh handling primitives is grounded on the following tree-based AMR operations; see Sect. 2.6. Create invokes New and Ghost, Refine_and_coarsen invokes Adapt, Balance, and Ghost, and Redistribute invokes Adapt, Partition, and Ghost. We note that these three mesh handling primitives also have to: (a) set up the data structures providing cell neighbors across n-faces; (b) call Alg. 1 (Sect. 3.4). The process underlying (a) is implemented by means of invocations to Iterate (Sect. 2.6).

As mentioned above, the Redistribute mesh handling primitive is in charge of migrating the data that the user might have attached to the mesh objects (i.e., cells and VEFs). The underlying tree-based AMR engine may have a built-in mechanism that lets users to attach, cell-wise, application-related data to the forest leaves, and migrate them *in one-shot* during the call to partition. This is indeed the approach followed by [8] atop p4est in order to migrate, not only mesh-related data, but that within any other data structure built on top of the mesh (such as, e.g., the nodal values of FE functions involved in the simulation). However, in our framework, we preferred to implement the communication pattern among processors required to carry out the data migration. This full control over data migration provides greater flexibility as we do not require the data to be migrated (e.g., FE function nodal values, or application-related cell-wise data arrays) to be fully available whenever the user calls to Redistribute.

The execution of the communication pattern required for data migration is easily implemented in terms of non-blocking point-to-point MPI communication primitives. The determination of such pattern is more involved. To this end, `Redistribute` performs a (temporary) copy of the forest-of-trees data structure right before calling `partition`. This copy consumes only a (temporary) small amount of extra memory due to the low memory footprint of tree-based AMR with SFCs. We stress that neither the full forest-of-trees mesh connectivity, nor the data structure storing ghost cells has to be copied. Then, `Redistribute` invokes `partition` on one of these two copies, so that it ends up with the same forest-of-trees, but (potentially) distributed differently among the processors involved. The algorithm that builds the communication pattern in turn relies on an algorithm that takes as input two instances of the forest-of-trees data structure. Let us refer to these as the “first” and “second” forest-of-trees, although we stress that they correspond to the same forest, but distributed differently among processors. On each processor p , the algorithm performs two passes over the leaf cells locally owned by p in the first forest. In a first pass, the algorithm determines, for each of these leaves, whether the leaf is still locally owned by p in the second forest, or by a remote processor q . This can be determined, for each leaf, in $O(\log_2(P))$ operations [8]. A second pass lets processor p know the number and list of remote processors q that own, in their local portion of the second forest, leaves that are owned by p , in its local portion of the first forest. The number and list of leaves transferred from p to q is determined in the same pass. It can be easily seen that the send side of the communication pattern is determined if we call the algorithm with the first and second forest being the forest before and after calling `partition`, resp., while the receive side, the other way round (i.e., first forest = after `partition`, second forest = before `partition`).

4. HANDLING CONFORMING FE SPATIAL DISCRETIZATIONS ON NON-COMFORMING MESHES

4.1. Problem statement and its FE spatial discretization. We aim at solving a PDE problem in Ω , subject to boundary conditions on $\partial\Omega$. We assume homogeneous Dirichlet conditions on $\partial\Omega_D \subset \partial\Omega$. For non-homogeneous Dirichlet conditions on $\partial\Omega_D \subset \partial\Omega$, we can consider an equivalent problem with homogeneous conditions and a modified right-hand side (see [27, Sect. 3.1]). Without loss of generality, we can state the problem in an abstract setting as follows: find $u \in \mathcal{V}$ such that

$$a(u, v) = \ell(v), \quad \forall v \in \mathcal{W}, \quad (3)$$

where \mathcal{V} and \mathcal{W} are infinite-dimensional Sobolev spaces of functions with null trace on $\partial\Omega_D$, $a : \mathcal{V} \times \mathcal{W} \rightarrow \mathbb{R}$ is a continuous bilinear form, and $\ell : \mathcal{W} \rightarrow \mathbb{R}$ a continuous linear form. For the general case $\mathcal{V} \neq \mathcal{W}$, problem (3) is referred to as a *Petrov-Galerkin*-type problem, and *Galerkin*-type otherwise. For ease in the exposition, hereafter we consider Galerkin-type problems, i.e., $\mathcal{V} = \mathcal{W}$. The extension to Petrov-Galerkin cases is straightforward.

In order to end up with a computable version of problem (3), we must introduce finite-dimensional spaces of functions \mathcal{V}_h with some approximability properties. In particular, if \mathcal{V}_h is a subspace of its infinite-dimensional counterpart, i.e., $\mathcal{V}_h \subset \mathcal{V}$, then \mathcal{V}_h is said to be a *conforming FE space*. In such a case, the discrete version of (3) reads: find $u_h \in \mathcal{V}_h$ such that:

$$a(u_h, v_h) = \ell(v_h), \quad \text{for any } v_h \in \mathcal{V}_h. \quad (4)$$

On the other hand, if $\mathcal{V}_h \not\subset \mathcal{V}$, then \mathcal{V}_h is said to be non-conforming. Non-conforming FE schemes, such as Discontinuous Galerkin (DG) schemes, require judiciously numerical perturbations of the continuous bilinear and linear forms in (3) in order to *weakly* enforce conformity.

4.2. Construction of conforming FE spaces. A conforming FE space \mathcal{V}_h can be built as follows. First, we define the local space of functions $Q(K)$ (defined as the composition of Φ_K^{-1} and a polynomial space $Q(T)$ on a reference polytope T) unisolvant with respect to a set of local DOFs (functionals) \mathcal{M}_K for every cell $K \in \mathcal{T}$; such construction leads to a canonical basis of local shape functions $\{\phi_K^\alpha\}_{\alpha \in \mathcal{M}_K}$ for $Q(K)$.⁷ We use the notation $v^\alpha \doteq \alpha(v)$ for any $v \in Q(K)$, $\alpha \in \mathcal{M}_K$. Second, with these cell-wise spaces, the

⁷Here, we assume the same polynomial space in all cells.

global FE space \mathcal{V}_h is determined by an equivalence class to glue together local DOFs and create the set of global DOFs \mathcal{N} ; we use the notation $\alpha \sim \beta$ to say that two local DOFs are equivalent. The equivalent class (global DOF) of a local DOF α is represented with $[\alpha]$; $[\cdot]$ is the so-called local-to-global DOF map. The equivalence relation depends on the particular FE at hand (see discussion below). For those cases in which Dirichlet boundary conditions are strongly enforced, global DOFs on the Dirichlet boundary are not actually free DOFs, and thus are not considered to be in \mathcal{N} in the following exposition.

For non-conforming DG methods the rest of this section does not apply, since there is no conformity to be enforced. In order to define conforming FE spaces, we need to introduce additional notation. Let us consider a map O_α that, for every local DOF α in \mathcal{M}_K , provides its *owner* n-face, with $O_\alpha \in \mathcal{F}_K \cup K$. For local VEFs $f \in \mathcal{F}_K$, we define its inverse map \mathcal{M}_K^f as the set of local DOFs owned by f , i.e., $\mathcal{M}_K^f \doteq \{\alpha \in \mathcal{M}_K : O_\alpha = f\}$. We also define $\mathcal{M}_K^{\bar{f}} \doteq \{\cup_g \mathcal{M}_K^g : g \in \mathcal{F}_K, g \subseteq \bar{f}\}$. We denote by $\mathcal{N}_F \subset \mathcal{N}$ (resp., $\mathcal{N}_{\bar{F}} \subset \mathcal{N}$) the set of all equivalence classes (i.e., global DOFs) resulting from the application of the equivalence relation \sim to all cell-local DOFs $\alpha \in \mathcal{M}_K^f$ (resp, $\alpha \in \mathcal{M}_K^{\bar{f}}$), for all $K \in \mathcal{T}_F$ such that $F = [f]$.

The concept of n -face owner of a DOF depends on the FE at hand. For example, for scalar Lagrangian FEs, there is a bijective map between DOFs and the corresponding set of nodes (points). The owner n -face of a DOF is defined as the one among $\mathcal{F}_K \cup K$ that contains the node. Two local DOFs are equivalent if they are positioned at the same node (point) in the physical space. Using this equivalence, it is easy to check that Ass. 4.1 below holds. Similar constructions are possible for edge and face FEs. See, e.g., [27, Sect. 3] for a detailed exposition of these concepts.

We assume that the following property holds for the FE at hand. It states that the continuity of global DOFs is equivalent to the trace continuity across cells required for conformity.

Assumption 4.1. *Given a set of global DOF values $\{v_h^g\}_{g \in \mathcal{N}}$, a function $v_h \in \mathcal{V}_h$, defined as $v_h \doteq \sum_{K \in \mathcal{T}} \sum_{\alpha \in \mathcal{M}_K} v_h^{[\alpha]} \phi_K^\alpha$, belongs to \mathcal{V} . Furthermore, given a function $v \in \mathcal{V}$ and local DOFs α, β with $[\alpha] = [\beta]$, $\alpha(v) = \beta(v)$.*

Any \mathcal{V}_h built from a FE that satisfies this assumption is conforming (on conforming meshes). For example, edge and face FEs are designed such that they satisfy this assumption with respect to tangent and normal component trace continuity, respectively.

Unfortunately, for non-conforming meshes, this is not enough. Under Ass. 4.1, given a tree-based adapted mesh \mathcal{T} with the properties in Sect. 2, it is easy to check that conformity on regular VEFs is readily obtained gluing together equivalent DOFs. However, on hanging VEFs this is not true anymore, and conformity on these VEFs must be enforced explicitly through additional constraints. We have the following result.

Proposition 4.2. *Any piecewise function $v_h = \sum_{K \in \mathcal{T}} \sum_{\alpha \in \mathcal{M}_K} v_h^{[\alpha]} \phi_K^\alpha$ such that*

$$v_h^{[\alpha]} = \sum_{\beta \in \mathcal{M}_{K_p}^{\bar{O}_f}} C_{[\alpha][\beta]} v_h^{[\beta]} \quad \text{with } C_{[\alpha][\beta]} \doteq \alpha(\phi_{K_p}^\beta), \quad \text{for any } \alpha \in \mathcal{M}_K^{\bar{f}}, f \in \mathcal{F}_H, \quad (5)$$

where K_p denotes the parent cell of K , belongs to \mathcal{V} .

Proof. First, let us note that the constraints over global DOFs in (5) can be enforced, i.e., they do not depend on the cell K being used to compute them. Let us consider a cell $K \in \mathcal{T}$ with a hanging VEF f , the parent cell K_p of K , and a neighbor cell $K' \in \mathcal{T}$ of K that contains $O_f \in \mathcal{F}_{K'}$. Using the following facts: (1) the space of polynomials in $Q(K_p)$ restricted to K is identical to $Q(K)$; (2) given $v \in Q(K_p)$, $v|_K = \sum_{\alpha \in \mathcal{M}_K} \alpha(v) \phi_K^\alpha$; and (3) Ass. 4.1, the required trace continuity for conformity readily holds for f . Using this result for all hanging VEFs and cells, it proves the proposition. \square

The constraints in (5) lead to a sparse constraint matrix C_{gh} ; $g \in \mathcal{N}_F$ is only constrained by $h \in M(g) \doteq \{g' \in \mathcal{N} : g' \in \mathcal{N}_{O_f}\}$. We note that the entries of C can be computed once at the reference polytope T for all possible pairs $t \in \mathcal{F}_{\mathcal{R}_T}$, $O_t \in T$ (such that $t \neq O_t$).

Let us denote with $D(\mathcal{V}_h)$ the dimension of the n -face of lowest dimension that has a non-empty set of owned DOFs. For instance, it is 0 for Lagrangian FEs (vertices own DOFs), 1 for edge FEs (edges own DOFs, vertices do not), 2 for face FEs (faces own DOFs, vertices/edges do not), and d (the space dimension) for DG (only cells own DOFs since no inter-cell continuity must be enforced). We have the following result.

Proposition 4.3. *Let us consider the FE problem (4) and a distributed 2:1 k -balanced forest-of-trees mesh \mathcal{T} with k -ghost cells. If $\max(1, D(\mathcal{V}_h)) \geq k$, the constraints in (5) are direct, i.e., they only depend on regular DOFs. Furthermore, all constraints of DOFs owned by n -faces in $\mathcal{F}_H \cap \mathcal{F}_L^p$ only depend on DOFs owned by $\mathcal{F}_R \cap \mathcal{F}_L^p$, and thus can be solved in parallel with processor-local information.*

Proof. The fact that the constraints are direct is an immediate consequence of (5), Prop. 3.10, and (for $D(\mathcal{V}_h) = 0$) Prop. 3.11. On the other hand, for DOFs owned by VEFs $F \in \mathcal{F}_H \cap \mathcal{F}_L^p$, then O_F , and any of its boundary VEFs are in \mathcal{F}_L^p , due to Def. 3.13, i.e., such DOFs are only constrained by DOFs owned by $\mathcal{F}_R \cap \mathcal{F}_L^p$. Since k -ghost cells are locally accessible, then O_F , and any of its boundary VEFs are locally accessible, and thus the constraints on DOFs owned by such VEFs F can be computed locally. \square

As shown in Sect. 5.1, being able to compute all such constraints locally enables each processor to build locally a (subassembled) local portion of the global linear system.

4.3. Generation of proc-local DOFs. In a parallel distributed-memory context, each processor restricts itself to the generation of those equivalence classes (i.e., global DOFs) of \mathcal{N} corresponding to n -faces (i.e., VEFs and cells) in \mathcal{T}^p . Given $F \in \mathcal{F}^p$, we denote by \mathcal{N}_F^p the set of all equivalence classes on F at processor p . We define $\mathcal{N}^p \doteq \bigcup_{F \in \{\mathcal{F}^p \cup \mathcal{T}^p\}} \mathcal{N}_F^p$, with $\mathcal{N}^p \subset \mathcal{N}$, and refer to it as the proc-local DOFs set. We also define $\mathcal{N}_R^p \doteq \bigcup_{F \in \{\mathcal{F}_R^p \cup \mathcal{T}^p\}} \mathcal{N}_F^p$, $\mathcal{N}_H^p \doteq \bigcup_{F \in \mathcal{F}_H^p} \mathcal{N}_F^p$, $\mathcal{N}_L^p \doteq \bigcup_{F \in \{\mathcal{F}_L^p \cup \mathcal{T}_L^p\}} \mathcal{N}_F^p$, $\mathcal{N}_G^p \doteq \bigcup_{F \in \{\mathcal{F}_G^p \cup \mathcal{T}_G^p\}} \mathcal{N}_F^p$, $\mathcal{N}_I^p \doteq \bigcup_{F \in \{\mathcal{F}_I^p \cup \mathcal{T}_L^p\}} \mathcal{N}_F^p$, $\mathcal{N}_\Gamma^p \doteq \bigcup_{F \in \{\mathcal{F}_\Gamma^p\}} \mathcal{N}_F^p$. Assuming that one only requires to build a distributed subassembled linear system (see Sect. 5), the need for a local-to-global index map $[\cdot]$ can be by-passed by combining a local-to-proc-local index map, denoted hereafter as $[\cdot]_p$, and Remark 5.3. In other words, local DOFs α in \mathcal{M}_K , for $K \in \mathcal{T}^p$, are labeled with a proc-local identifier $[\alpha]_p$ in the range $\{1, \dots, |\mathcal{N}^p|\}$. We refer to Sect. 3.3 for the computational benefits resulting from this decision.

Alg. 2 shows a simplified pseudocode of the process in charge of building the \mathcal{N}^p , \mathcal{N}_R^p , \mathcal{N}_H^p , \mathcal{N}_L^p , and \mathcal{N}_G^p proc-local DOFs sets. The construction of \mathcal{N}_I^p and \mathcal{N}_Γ^p is omitted from Alg. 2 in order to keep the presentation short, although it follows the same lines as with the other sets. Each processor p executes its own instance of Alg. 2, *with no communication at all involved among processors*. The algorithm is grounded on the `generate_nface_dofs(F)` primitive, with $F \in \mathcal{F}^p$ or $F \in \mathcal{T}^p$. When $F \in \mathcal{F}^p$, this primitive applies to equivalence relation \sim to all cell-local DOFs $\alpha \in \mathcal{M}_K^f$, for $K \in \mathcal{T}_F^p$ such that $[f] = F$, in order to generate the equivalence classes in \mathcal{N}_F^p .

Remark 4.4. *In practice, proc-local DOF identifiers $[\cdot]_p$ are required, among others, in order to address in memory the DOF values of the proc-local portion of global FE functions $v_h \in \mathcal{V}_h$. Alg. 2 generates \mathcal{N}_F^p (and thus, $[\cdot]_p$), for all $F \in \mathcal{F}^p \cup \mathcal{T}^p$, and not only for $F \in \{\mathcal{F}_L^p \cap \mathcal{F}_R^p\} \cup \mathcal{T}_L^p$, i.e., those which are strictly required for FE (sub)assembly (see Sect. 5). Assuming DOF-wise (versus cell-wise) storage of FE functions, this results in significant memory savings. On the other hand, in a parallel context, it turns out that each processor needs these values for all DOFs on ghost cells, thus justifying why the \mathcal{N}_G^p set is generated in Alg. 2 as well. This is required, among others, for the evaluation of function gradient jumps across faces in the interface among processors, as typically required by a-posteriori error estimators [22]. The values of DOFs in \mathcal{N}_G^p are fetched from remote processors by means of a cell-wise nearest-neighbor communication.*

5. SETTING UP PARALLEL DISTRIBUTED SUBASSEMBLED LINEAR SYSTEMS

Algorithm 2: Generation of proc-local DOFs on \mathcal{T}^p .

```

1  $\mathcal{N}_R^p \leftarrow \emptyset; \mathcal{N}_H^p \leftarrow \emptyset;$ 
2 for  $F \in \{\mathcal{F}_L^p \cap \mathcal{F}_R^p\} \cup \mathcal{T}_L^p$  do                                /* loop over VEFs in  $\mathcal{F}_L^p \cap \mathcal{F}_R^p$  and local cells in  $\mathcal{T}_L^p$  */
3    $\mathcal{N}_F^p \leftarrow \text{generate\_nface\_dofs}(F)$       /* generate proc-local DOFs (equivalence classes) owned by  $F$  */
4    $\mathcal{N}_R^p \leftarrow \mathcal{N}_R^p \cup \mathcal{N}_F^p$ 
5 end
6 for  $F \in \mathcal{F}_L^p \cap \mathcal{F}_H^p$  do                                /* loop over VEFs in  $\mathcal{F}_L^p \cap \mathcal{F}_H^p$  */
7    $\mathcal{N}_F^p \leftarrow \text{generate\_nface\_dofs}(F)$       /* generate proc-local DOFs (equivalence classes) owned by  $F$  */
8    $\mathcal{N}_H^p \leftarrow \mathcal{N}_H^p \cup \mathcal{N}_F^p$ 
9 end
10  $\mathcal{N}_L^p \leftarrow \mathcal{N}_R^p \cup \mathcal{N}_H^p$ 
11 Execute Lines 2-5 for  $F \in \{\mathcal{F}_G^p \cap \mathcal{F}_R^p\} \cup \mathcal{T}_G^p$           /* generate  $\mathcal{N}_R^p \cap \mathcal{N}_G^p$  */
12 Execute Lines 6-9 for  $F \in \mathcal{F}_G^p \cap \mathcal{F}_H^p$                           /* generate  $\mathcal{N}_H^p \cap \mathcal{N}_G^p$  */
13  $\mathcal{N}_G^p \leftarrow \{\mathcal{N}_R^p \cup \mathcal{N}_H^p\} \setminus \mathcal{N}_L^p$ 
14  $\mathcal{N}^p \leftarrow \mathcal{N}_L^p \cup \mathcal{N}_G^p$ 

```

5.1. Finite element (sub)assembly. Assuming that the bilinear and linear forms in Problem (4) can be split into cell contributions as $a(u, v) = \sum_{K \in \mathcal{T}} a_K(u, v)$ and $\ell(v) = \sum_{K \in \mathcal{T}} \ell_K(v)$, resp., then, this problem, with \mathcal{V}_h as defined in Prop. 4.2, can be assembled into a global linear system $\mathbf{Au} = \mathbf{b}$ as a sum of cell matrices $(\mathbf{A}_K)_{\alpha\beta} \doteq a_K(\phi_K^\alpha, \phi_K^\beta)$ and vectors $(\mathbf{b}_K)_\alpha \doteq \ell_K(\phi_K^\alpha)$, for $K \in \mathcal{T}$, $\alpha \in \mathcal{M}_K$, $\beta \in \mathcal{M}_K$, using the assembly operations collected in Tab. 1. Using this approach, hanging DOF constraints are first resolved locally at each cell, and the entries are then transferred (assembled) into the global linear system. Besides, hanging DOFs are not associated to an equation of the global linear system, i.e., $\mathbf{Au} = \mathbf{b}$ has $|\mathcal{N}_R|$ equations/unknowns. Therefore, after solving the linear system, one has to use (5) in order to compute the DOF values of u_h in (4) associated to hanging DOFs.

Case	Assembly operation
Right hand side vector	
$\alpha \in \mathcal{M}_K^f : [f] \in \mathcal{F}_K^R$	$\mathbf{b}_{[\alpha]} += (\mathbf{b}_K)_\alpha$ (standard case)
$\alpha \in \mathcal{M}_K^f : [f] \in \mathcal{F}_K^H$	$\mathbf{b}_g += C_{[\alpha]g}(\mathbf{b}_K)_\alpha$, $g \in M([\alpha])$
System matrix	
$\alpha \in \mathcal{M}_K^f : [f] \in \mathcal{F}_K^R$ and $\beta \in \mathcal{M}_K^{f'} : [f'] \in \mathcal{F}_K^R$	$\mathbf{A}_{[\alpha][\beta]} += (\mathbf{A}_K)_{\beta\alpha}$ (standard case)
$\alpha \in \mathcal{M}_K^f : [f] \in \mathcal{F}_K^H$ and $\beta \in \mathcal{M}_K^{f'} : [f'] \in \mathcal{F}_K^R$	$\mathbf{A}_{g[\beta]} += C_{[\alpha]g}(\mathbf{A}_K)_{\beta\alpha}$, $g \in M([\alpha])$
$\alpha \in \mathcal{M}_K^f : [f] \in \mathcal{F}_K^R$ and $\beta \in \mathcal{M}_K^{f'} : [f'] \in \mathcal{F}_K^H$	$\mathbf{A}_{[\alpha]h} += C_{[\beta]h}(\mathbf{A}_K)_{\beta\alpha}$, $h \in M([\beta])$
$\alpha \in \mathcal{M}_K^f : [f] \in \mathcal{F}_K^H$ and $\beta \in \mathcal{M}_K^{f'} : [f'] \in \mathcal{F}_K^H$	$\mathbf{A}_{gh} += C_{[\alpha]g}C_{[\beta]h}(\mathbf{A}_K)_{\beta\alpha}$, $g \in M([\alpha])$, $h \in M([\beta])$

TABLE 1. Assembly operations with hanging DOF constraints. We use the notation $\mathcal{F}_K^R \doteq [\mathcal{F}_K] \cap \mathcal{F}_R$ and $\mathcal{F}_K^H \doteq [\mathcal{F}_K] \cap \mathcal{F}_H$. Local DOFs $\alpha \in \mathcal{M}_K$ such that $O_\alpha = K$ (i.e., cell-interior DOFs) are omitted from the table to keep it concise. These local DOFs are treated in the same way as local DOFs owned by VEFs $f \in \mathcal{F}_K^R$ in the table.

In a parallel distributed-memory setting, each processor p executes the operations in Tab. 1 *only for cells* $K \in \mathcal{T}_L^p$, using $[\cdot]_p$ instead of $[\cdot]$; see Sect. 4.3. This results in a proc-local matrix \mathbf{A}^p and a proc-local vector \mathbf{b}^p (with as many rows/cols and entries, resp., as $|\mathcal{N}_L^p \cap \mathcal{N}_R^p|$) that are fully assembled for interior DOFs, but only *partially assembled* (or *subassembled*) for interface DOFs. More precisely, if $(\mathbf{A}^p)_{ij}$ is an

entry of \mathbf{A}^p such that i and j belong to \mathcal{N}_Γ^p , then $(\mathbf{A}^p)_{ij}$ holds a partial contribution to the corresponding entry in the global matrix \mathbf{A} (and a fully summed entry otherwise). The same applies to $(\mathbf{b}^p)_i$ if i belongs to \mathcal{N}_Γ^p .

In order to be able to build \mathbf{A}^p and \mathbf{b}^p , processor p *only needs to handle hanging DOF constraints* (i.e., to determine the sets $M([\alpha]_p)$, and to compute the coefficients $C_{[\alpha]_p [\beta]_p}$ defined in Sect. 4.2) *for those hanging DOFs touched by local cells in \mathcal{T}_L^p* . If the requirements of Prop. 4.3 are satisfied, we know that such constraints can be resolved in parallel with processor-local information, i.e., it is guaranteed that hanging DOF constraints dependencies required to build \mathbf{A}^p and \mathbf{b}^p cannot expand beyond a single layer of ghost cells.

5.2. Subdomain-wise assembly of proc-regular interface DOF values. In a distributed-memory context, non-overlapping domain decomposition solvers need, in the iterative solution process of $\mathbf{Au} = \mathbf{b}$, to assemble, subdomain-wise, the (subassembled) values corresponding to proc-regular interface DOFs (i.e., DOFs in the $\mathcal{N}_R^p \cap \mathcal{N}_\Gamma^p$ set).⁸ This operation can be stated as follows. Given a distributed *subassembled* (i.e., partially-summed) vector \mathbf{x} , i.e., a vector such that entries of \mathbf{x}^p corresponding to proc-regular interface DOFs hold partial contributions to the corresponding entries in \mathbf{x} , *transform \mathbf{x} such that it becomes fully-assembled*, i.e., entries of \mathbf{x}^p corresponding to proc-regular interface DOFs contain the (i.e., fully summed) value of the corresponding entries in \mathbf{x} .

We define the set $\mathcal{S}^{\text{proc}}(g) \doteq \{q : g \in \mathcal{N}_R^q \cap \mathcal{N}_\Gamma^q\}$ for any $g \in \mathcal{N}_R^p \cap \mathcal{N}_\Gamma^p$. For each of these DOFs, we assign a *processor owner* among the set of processors in $\mathcal{S}^{\text{proc}}(g)$ (using an algorithm presented later in this section). The owner processor is denoted by $O^{\text{proc}}(g) \in \mathcal{S}^{\text{proc}}(g)$. The rest of processors in $\mathcal{S}^{\text{proc}}(g)$ become non-owner processors. As required for the exposition, we also define equivalently $\mathcal{S}^{\text{proc}}(F)$ and $O^{\text{proc}}(F)$ for any VEF $F \in \mathcal{F}_\Gamma^p \cap \mathcal{F}_R^p$. Thus, we can write $\mathcal{S}^{\text{proc}}(g) \doteq \mathcal{S}^{\text{proc}}(F)$ (resp. $O^{\text{proc}}(g) \doteq O^{\text{proc}}(F)$), for any $g \in \mathcal{N}_F^p$. In any case, the algorithms in this section do not compute $\mathcal{S}^{\text{proc}}(F)$ and $O^{\text{proc}}(F)$, but $O^{\text{proc}}(g)$ and $\mathcal{S}^{\text{proc}}(g)$ directly. This may lead to computational savings, as not all VEFs necessarily own DOFs. (This obviously depends on the FE at hand.) The subdomain-wise assembly operation is split into two communication stages. In the first stage, referred to as **S1**, processor owners receive the corresponding partially-summed contributions from non-owner processors, and reduce-sum the partial sums received into fully-assembled values (at processor owners). In the second stage, referred to as **S2**, processor owners send fully-summed values resulting from the first stage to non-owner processors, so that all processors end up with fully-assembled proc-regular interface DOF values. In the sequel, we sketch the process that sets up the communication patterns underlying **S1** and **S2**.

Each processor needs to figure out the following information in order to set up the **S1** communication pattern. *On the receive side*, p has to determine the set of processor identifiers from which it receives data. Let us denote this set by $\mathcal{S}_{\text{rcv}}^p$. Besides, p needs to figure out, for each processor $q \in \mathcal{S}_{\text{rcv}}^p$, the DOFs whose values have to be accumulated to the (partially-summed) values received from q . Let us denote this set by $\mathcal{N}_{\text{rcv}}^{p \leftarrow q}$, for all $q \in \mathcal{S}_{\text{rcv}}^p$. Thus, $|\mathcal{N}_{\text{rcv}}^{p \leftarrow q}|$ is the amount of data items that p receives from q . *On the send side*, each processor p has to determine $\mathcal{S}_{\text{snd}}^p$ and $\mathcal{N}_{\text{snd}}^{p \rightarrow q}$, for each $q \in \mathcal{S}_{\text{snd}}^p$, i.e., the set of processor identifiers to which p sends data, and the DOFs whose values have to be sent to each processor $q \in \mathcal{S}_{\text{snd}}^p$, resp. Likewise, $|\mathcal{N}_{\text{snd}}^{p \rightarrow q}|$ is the amount of vector entries that p sends to q . On the other hand, the communication pattern for **S2** can be very easily determined from the one of **S1** by swapping send-side sets and receive-side sets.

Alg. 3 sketches the process in charge of generating the **S1** communication pattern. The inputs of the algorithm are $O^{\text{proc}}(g)$, for $g \in \mathcal{N}_R^p \cap \mathcal{N}_\Gamma^p$, and $\mathcal{S}^{\text{proc}}(g)$ for $g \in \mathcal{N}_R^p \cap \mathcal{N}_\Gamma^p$ such that $O^{\text{proc}}(g) = p$. (Note that the algorithm does not actually require $\mathcal{S}^{\text{proc}}(g)$ if $O^{\text{proc}}(g) \neq p$.) From these inputs, the algorithm generates the $\mathcal{S}_{\text{rcv}}^p$, $\mathcal{N}_{\text{rcv}}^{p \leftarrow q}$, $\mathcal{S}_{\text{snd}}^p$, and $\mathcal{N}_{\text{snd}}^{p \rightarrow q}$ sets.

⁸Recall from Sect. 5.1 that hanging DOF constraints are eliminated from the system during FE assembly, i.e., we do not allocate an equation/unknown for DOFs in the $\mathcal{N}_H^p \cap \mathcal{N}_L^p$ set.

Algorithm 3: Determine S_{recv}^p , $N_{\text{recv}}^{p \leftarrow q}$, S_{snd}^p , and $N_{\text{snd}}^{p \rightarrow q}$.

```

1  for  $F \in \mathcal{F}_\Gamma^p$  do                                /* loop over interface VEFs */
2    if  $F \in \mathcal{F}_R^p$  then                      /* current VEF is proc-regular */
3      for  $g \in N_F^p$  do
4         $q \leftarrow O^{\text{proc}}(g)$ 
5        if  $p = q$  then                         /* i am the owner of g */
6          for  $q \in S^{\text{proc}}(g) \setminus \{p\}$  do
7            |  $S_{\text{recv}}^p \leftarrow S_{\text{recv}}^p \cup \{q\}$ ;  $N_{\text{recv}}^{p \leftarrow q} \leftarrow N_{\text{recv}}^{p \leftarrow q} \cup \{g\}$ 
8          end
9        else                                     /* a remote processor q is the owner of g */
10         |  $S_{\text{snd}}^p \leftarrow S_{\text{snd}}^p \cup \{q\}$ ;  $N_{\text{snd}}^{p \rightarrow q} \leftarrow N_{\text{snd}}^{p \rightarrow q} \cup \{g\}$ 
11       end
12     end
13   end
14 end

```

It remains to discuss how $O^{\text{proc}}(g)$ and $S^{\text{proc}}(g)$ are generated. Alg. 4 presents an sketch of the process in charge of determining $O^{\text{proc}}(g)$, for $g \in N_R^p \cap N_\Gamma^p$, and $S^{\text{proc}}(g)$ for $g \in N_R^p \cap N_\Gamma^p$ such that $O^{\text{proc}}(g) = p$ (i.e., the input of Alg. 3). Alg. 4 is split into two stages. On the one hand, a local stage spanning Lines 1-13, in which each processor p builds as much as it can of $O^{\text{proc}}(\cdot)$ and $S^{\text{proc}}(\cdot)$ only using local information. The local stage of Alg. 4 also walks through proc-hanging interface VEFs; see Line 1 and Lines 8-12. This is justified by the fact that, due to the hanging DOFs constraints, all $g \in N_{\overline{O_F}}$ become local in all processors that own cells in \mathcal{T}_F^p . On the other hand, Alg. 3 features a communication stage spanning Lines 14-30, in which the processors complete the partially constructed version of $O^{\text{proc}}(\cdot)$ and $S^{\text{proc}}(\cdot)$ during the first stage. The communication stage is composed by pack (Lines 15-20), nearest-neighbor exchange (Line 21), and unpack (Lines 22-30) steps. In the rest of the section we mathematically prove why this algorithm is correct.

Let us define the owner processor of a DOF $g \in N_F^p$ for $F \in \mathcal{F}_\Gamma^p \cap \mathcal{F}_R^p$ as $O^{\text{proc}}(g) \doteq \max_{K \in \mathcal{T}_F} O^{\text{proc}}(K)$, with $O^{\text{proc}}(K)$ denoting the processor owner of $K \in \mathcal{T}$; see Sect. 3.2. The following result holds.

Proposition 5.1. *Let us consider a distributed forest-of-trees mesh \mathcal{T} with s -ghost cells, where $D(\mathcal{V}_h) \geq s$. For any processor p , the owner processor of $g \in N_F^p$, for $F \in \mathcal{F}_\Gamma^p \cap \mathcal{F}_R^p$, can be locally computed as $O^{\text{proc}}(g) = \max_{K \in \mathcal{T}_F^p} O^{\text{proc}}(K)$ if $\mathcal{T}_F^p \cap \mathcal{T}_L^p \neq \emptyset$ (see Line 6). Otherwise (i.e., $\mathcal{T}_F^p \cap \mathcal{T}_L^p = \emptyset$), any neighbor processor that owns a ghost cell in \mathcal{T}_F^p can compute $O^{\text{proc}}(g)$ (and, thus, p can fetch this information from any of these neighbors).*

Proof. If $D(\mathcal{V}_h) \geq s$, then for $g \in N_R^p \cap N_\Gamma^p$ such that $g \in N_F^p$ with $\mathcal{T}_F^p \cap \mathcal{T}_L^p \neq \emptyset$, then $\mathcal{T}_F^p = \mathcal{T}_F$ (see Sect. 3.2). Thus, all processors in $S^{\text{proc}}(g)$ such that g is surrounded by at least one local cell can compute $O^{\text{proc}}(g)$ locally. The last part of the proposition is straightforward. \square

As a result of this proposition, the communication stage in Alg. 4 (Lines 14-30) is guaranteed to obtain $O^{\text{proc}}(g)$ on processors such that $\mathcal{T}_F^p \cap \mathcal{T}_L^p = \emptyset$ as well.

On the other hand, with regard to $S^{\text{proc}}(g)$, we note that a processor that is owner of a proc-regular interface DOF g , i.e., $O^{\text{proc}}(g) = p$, may not able to determine the full $S^{\text{proc}}(g)$ set solely using local information (i.e., local cells plus a layer of s -ghost cells). This scenario is illustrated in Fig. 3c for the DOF at processor $p = 5$ pointed by an arrow. Using only local information, $S_5(g) = \{3, 4, 5\}$. However, processor 2 is an element of $S_5(g)$ as well. The cell owned by processor 2 that would let processor 5 complete the $S_5(g)$ set is not in the ghost cells set of processor 5. However, processor 5 retrieves such missing local information fetching data from processor 3 in the processor 5 ghost cells which are owned by processor 3, in particular in Line 25 of Alg. 4. The approach followed by Alg. 4 in order to build $S^{\text{proc}}(g)$ is correct as proved in the following proposition.

Algorithm 4: Determine $O^{\text{proc}}(\cdot)$ and $S^{\text{proc}}(\cdot)$ for proc-regular interface DOFs.

```

1  for  $F \in \mathcal{F}_\Gamma^P$  do                                /* loop over interface VEFs */
2     $\mathcal{W} \leftarrow \bigcup_{K \in \mathcal{T}_F^P} O^{\text{proc}}(K)$ 
3    if  $F \in \mathcal{F}_R^P$  then                         /* current VEF is proc-regular */
4      for  $g \in \mathcal{N}_F^P$  do
5         $S^{\text{proc}}(g) \leftarrow S^{\text{proc}}(g) \cup \mathcal{W}$ 
6        if  $\mathcal{T}_F^P \cap \mathcal{T}_L^P \neq \emptyset$  then  $O^{\text{proc}}(g) \leftarrow \max_{K \in \mathcal{T}_F^P} O^{\text{proc}}(K)$  /* at least one cell surrounding  $F$  is local */
7      end
8    else                                         /* current VEF is proc-hanging */
9      if  $\mathcal{N}_F^P \neq \emptyset$  then
10        for  $g \in \mathcal{N}_{\overline{O_F}}$  do  $S^{\text{proc}}(g) \leftarrow S^{\text{proc}}(g) \cup \mathcal{W}$            /* loop over DOFs owned by  $\overline{O_F}$  */
11      end
12    end
13  end
14 Allocate communication buffers  $O_{\text{buf}}(\cdot, \cdot)$  and  $S_{\text{buf}}(\cdot, \cdot)$  to size  $|\mathcal{T}^P| \times |\mathcal{M}_K|$ 
15 for  $F \in \mathcal{F}_R^P \cap \mathcal{F}_\Gamma^P$  do          /* pack local data into communication buffers (Lines 15-20) */
16   for  $K \in \mathcal{T}_F^P$  such that  $K \in \mathcal{T}_L^P$  do          /* loop over local cells around  $F$  */
17     Find  $f \in \mathcal{F}_K$  such that  $[f] = F$ 
18     for  $\alpha \in \mathcal{M}_K^f$  do  $X_{\text{buf}}(K, \alpha) \leftarrow X^{\text{proc}}([\alpha]_p)$ , with  $X = O, S$ 
19   end
20 end
21 Fetch ghost cell data of  $O_{\text{buf}}(\cdot, \cdot)$  and  $S_{\text{buf}}(\cdot, \cdot)$  from remote processors via nearest-neighbors exchange
22 for  $K \in \mathcal{T}_G^P$  do                      /* unpack data from communication buffers (Lines 22-30) */
23   for  $f \in \mathcal{F}_K$  such that  $[f] \in \mathcal{F}_L^P \cap \mathcal{F}_R^P$  do
24     if  $\mathcal{T}_F^P \cap \mathcal{T}_L^P \neq \emptyset$  then
25       for  $\alpha \in \mathcal{M}_K^f$  do  $S^{\text{proc}}([\alpha]_p) \leftarrow S^{\text{proc}}([\alpha]_p) \cup S_{\text{buf}}^P(K, \alpha)$       /* augment  $S^{\text{proc}}(\cdot)$  with neighbors data */
26     else
27       for  $\alpha \in \mathcal{M}_K^f$  do  $O^{\text{proc}}([\alpha]_p) \leftarrow O_{\text{buf}}^P(K, \alpha)$                   /* remote neighbors determine  $O^{\text{proc}}(\cdot)$  */
28     end
29   end
30 end

```

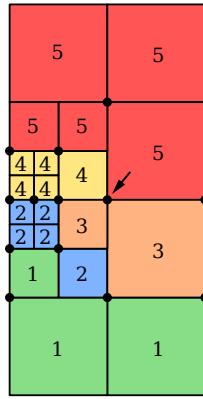
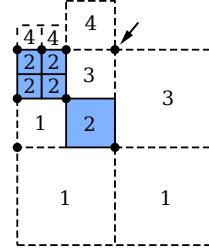
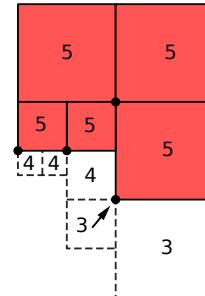
(a) $\mathcal{N} \cap \bigcup_p \mathcal{N}_R^p \cap \mathcal{N}_L^p$.(b) $\mathcal{N}_R^2 \cap \mathcal{N}_\Gamma^2$.(c) $\mathcal{N}_R^5 \cap \mathcal{N}_\Gamma^5$.

FIGURE 3. Proc-regular interface DOFs (black circles) of a global conforming FE space grounded on bilinear Lagrangian FEs on top of a 2:1 0-balanced forest-of-quadtrees mesh with two quadtrees distributed (non-uniformly) among $P = 5$ processors with 0-ghost cells. The DOF pointed out by an arrow in Fig. 3c is such that processor $p = 5$, even being its owner, is not able to determine the full $S^{\text{proc}}(g)$ set solely using local information.

Proposition 5.2. *Let us consider a conforming FE space \mathcal{V}_h and a distributed 2:1 k -balanced forest-of-trees mesh \mathcal{T} with k -ghost cells, where $\max(1, D(\mathcal{V}_h)) \geq k$. Then, for any $g \in \mathcal{N}_R^p \cap \mathcal{N}_\Gamma^p$ such that $O^{\text{proc}}(g) = p$, the full $\mathcal{S}^{\text{proc}}(g)$ set can be computed combining local information and a single nearest-neighbor communication.*

Proof. Given a processor $q \in \mathcal{S}^{\text{proc}}(g)$, by definition, the VEF F that owns g must be in \mathcal{F}_L^q because of (a), (c), or (d) in Def. 3.13. Using the definition of k -ghost cells in Def. 3.12, when F satisfies (a) or (c) at processor q , the fact that q is in $\mathcal{S}^{\text{proc}}(g)$ can be locally determined at $O^{\text{proc}}(g)$. When (d) holds for some VEF J such that $F \subseteq \bar{J}$ (where J is regular by Prop. 3.11), it is easy to check that $O^{\text{proc}}(J)$ (which cannot be q , as J is only surrounded by ghost cells in q) is neighbor of $O^{\text{proc}}(g)$ and can determine that $q \in \mathcal{S}^{\text{proc}}(g)$. Thus, q can be fetched at $O^{\text{proc}}(g)$ with a single nearest neighbor communication. \square

Remark 5.3 (Implementation remark). *In order to keep the presentation short, we have omitted from Alg. 3 and 4 that DOFs in $\mathcal{N}_{\text{snd}}^{p \rightarrow q}$ and $\mathcal{N}_{\text{rcv}}^{q \leftarrow p}$, for any pair of neighbors p and q , have to be glued together at both sides of the interface, as a final step to complete the equivalence classes corresponding to DOFs at the subdomain interfaces. We achieve this is in practice as follows. Alg. 4 generates, *in situ*, a global DOF identifier for all DOFs $g \in \mathcal{N}_R^p \cap \mathcal{N}_\Gamma^p$. The algorithm that generates such identifiers follows the same pattern to the one that determines $O^{\text{proc}}(g)$, and thus is mathematically supported by Prop. 5.1 as well. In particular, if a proc-regular interface DOF is surrounded by at least one local cell, a global DOF identifier can be generated solely using processor-local information, while if not, it can be fetched on ghost cells via a nearest-neighbor exchange. Then, using such global DOF identifiers, Alg. 3 arranges the DOFs in $\mathcal{N}_{\text{snd}}^{p \rightarrow q}$ and $\mathcal{N}_{\text{rcv}}^{q \leftarrow p}$ in increasing global DOF numbering within each set, such that they are consistently paired at both sides of the interface among p and q . The global DOF identifiers are then discarded, and not used elsewhere.*

Remark 5.4 (Implementation remark). *Let us consider the subset of $\mathcal{N}_R^p \cap \mathcal{N}_\Gamma^p$ composed by all DOFs g such that: (1) they are present in p and in one (and only one) remote processor q , i.e., $|\mathcal{S}^{\text{proc}}(g)| = 2$; (2) both in processor p and q , there is at least one local cell around g , so that either processor can actually become the owner of g . While Alg. 4 always assigns ownership to the processor with the maximum identifier, its actual implementation in FEMPAR does not. The latter, however, assigns evenly processor ownership to the DOFs in the aforementioned subset, i.e., p becomes owner of a half of its DOFs, and q of the other half. This approach balances the message sizes to be exchanged among any two processors, thus resulting in increased parallelism.*

6. SETTING UP PARALLEL DISTRIBUTED FULLY-ASSEMBLED LINEAR SYSTEMS

The vast majority of parallel linear algebra packages (e.g., PETSc [32], or TRILINOS [33]) use a data distribution model in which the global linear system is partitioned by rows, i.e., each processor *owns* a consecutive, non-overlapping range of *fully-assembled* global rows of \mathbf{A} and \mathbf{b} , resp. Besides, the data structures are *globally addressable*, i.e., a given processor may contribute to a global matrix row (or a global vector entry) which is not necessarily owned by it. Leveraging such packages essentially amounts to design a distributed algorithm able to generate the equivalence classes in \mathcal{N}_R with the added constraint that the local-to-global index map $[\cdot]$ has to be built such that the DOFs owned by the first processor are the ones to be numbered first, immediately followed by those of the second, and so on.

The distributed algorithm presented in [34, Sect. 3.1] (available in deal.II) follows a cell-wise approach that exploits the overlapped mesh partition to directly generate such index map. Here, we instead follow a DOF-wise approach which exploits the locally generated local-to-proc-local index map $[\cdot]_p$ (see Sect. 4.3) and Remark 5.3 to assign a global DOF identifier to DOFs in $\mathcal{N}_L^p \cap \mathcal{N}_R^p$, for all p . We define $\mathcal{N}_O^p \doteq \{g \in \mathcal{N}_L^p \cap \mathcal{N}_R^p : O^{\text{proc}}(g) = p\}$, and refer to it as the set of DOFs owned by processor p . This algorithm encompasses four main steps: (a) each processor locally computes $|\mathcal{N}_O^p|$; (b) a root processor gathers $|\mathcal{N}_O^p|$, for all p , then computes an starting offset global DOF identifier for each p from the data gathered, and finally scatters these to the corresponding processors; (c) each processor locally assigns a

global, consecutive DOF identifier to owned DOFs starting from the global offset identifier computed in (b); (d) a DOF-wise nearest neighbor communication with $S_{\text{rcv}}^p, N_{\text{rcv}}^{p \leftarrow q}$ as send side, and $S_{\text{snd}}^p, N_{\text{snd}}^{p \rightarrow q}$ as receive side (i.e., owners send, non-owners receive) lets each processor to fetch from remote neighbors the global DOF identifiers for those DOFs in $N_L^p \cap N_R^p$ which are not owned by it.

For the actual set up (i.e., assembly) of the fully-assembled linear system, we re-use the subassembled matrix and right-hand-side vector computed by means of the process described in Sect. 5.1. In a post-process stage, we go over the subassembled non-zero matrix entries $(\mathbf{A}^p)_{ij}$ and vector entries $(\mathbf{b}^p)_i$, and assemble them into the corresponding global entries of \mathbf{A} and \mathbf{b} using the global identifiers computed above and the services provided by the external package linear algebra data structures. A final inter-processor assembly stage involves the transfer of matrix and vector (partial) contributions to local matrix rows and local vector entries but actually stored in (i.e., owned by) remote processors. This communication phase is resolved by the linear algebra package and, thus, we do not cover it here for simplicity.

Remark 6.1. *The algorithm available in deal.II [34, Sect. 3.1] assigns ownership to the processor with the smallest identifier; see [34, Remark 2] for the underlying rationale. We, however, follow Remark 5.4. Our experiments in Sect. 7 reveal a significantly larger parallel efficiency drop of the assembly stage in deal.II compared to its counterpart in FEMPAR, which we believe to be caused by this difference.*

7. NUMERICAL EXPERIMENTS

7.1. Experimental environment. The numerical experiments are run at the Marenostrum-IV (MN-IV) supercomputer, hosted by BSC. This petascale machine is equipped with 3,456 compute nodes interconnected together with the Intel OPA HPC network. Each node has 2x Intel Xeon Platinum 8160 multi-core CPUs, with 24 cores each (i.e. 48 cores per node) and 96 GBytes of RAM.

With respect to the software, we used FEMPAR v1.0.0 [27], linked against p4est v2.0 [8, 9] as its forest-of-octrees manipulation engine. Besides, we used deal.II v9.0.0 [30], linked against p4est v2.0, and PETSc v3.9.0 [32] for distributed-memory linear algebra data structures and solvers. These software were compiled with Intel v18.0.1 compilers using system recommended optimization flags and linked against the Intel Message Passing Interface (MPI) Library (v2018.1.163) for message-passing and the BLAS/LAPACK and PARDISO available on the Intel MKL library for optimized dense linear algebra kernels and sparse direct solvers, resp. All floating-point operations were performed in IEEE double precision.

7.2. Poisson problem. In this section we evaluate the performance and strong scalability of the different stages in our framework as implemented in FEMPAR, and compare them against those in deal.II for the numerical solution of the Poisson problem. We consider a 3D cube domain $\Omega = [0, 1]^3$ with homogeneous Dirichlet boundary conditions (strongly) imposed over the entire domain boundary. The problem is discretized using a global Lagrangian FE space $\mathcal{V}_h \subset H_0^1(\Omega)$. As $D(\mathcal{V}_h) = 0$, we may use $k = 1$ and $s = 0$; see Prop. 4.3, and 5.1. However, we use $k = 0$ and $s = 0$ to have a fair comparison against deal.II; see Sect. 3.1. The weak formulation of this problem reads: find $u_h \in \mathcal{V}_h$ such that

$$(\nabla u_h, \nabla v_h) = (f, v_h), \quad \forall v_h \in \mathcal{V}_h. \quad (6)$$

The right-hand-size $f(x, y, z)$ is a piece-wise function defined as:

$$f(x, y, z) = \begin{cases} 1 & z > \frac{1}{2} + \frac{1}{4} \sin(4\pi x) \sin(4\pi y) \\ -1 & z \leq \frac{1}{2} + \frac{1}{4} \sin(4\pi x) \sin(4\pi y) \end{cases}.$$

The discontinuity in f leads to a solution that is non-smooth along a sinusoidal surface through the domain, so that very localized AMR is required in order to reduce the error in that area, while keeping the computational requirements reasonably low.

The experiment is designed as follows. We start with a coarse mesh that is obtained after uniformly refining the root octree up to four times. This mesh has 16 cells (hexahedra) per coordinate direction, and 4096 cells in total. Then, the Poisson problem is solved on a hierarchy of

meshes where the mesh at a given level is obtained from the one at the previous one by means of the `Refine_and_coarsen` mesh handling primitive presented in Sect. 3.5. With load balancing in mind, the mesh is dynamically redistributed at each level by means of the `Partition` primitive right after each call to `Refine_and_coarsen`. For a given (fixed) number of AMR levels, we measure the elapsed time (i.e., wall clock time) spent in each of the stages in the simulation of process, *aggregated across all levels*, and evaluate at which rate they are reduced with increasing number of CPU cores (i.e., strong scalability test). The number of AMR levels is adjusted such that a “sufficiently large” problem size at the last level is obtained for the CPU core range on which we run the strong scaling test.

The decision of which cells to be refined, coarsened, or untouched (previous step required for the `Refine_and_coarsen` primitive) is performed as follows. Given the solution of the linear system, each processor computes independently of each other an error indicator for cells $K \in \mathcal{T}_L^P$ using the a-posteriori error estimator in [39]. Then, given user-defined refinement and coarsening fractions, denoted by α_r and α_c , resp., we find the thresholds θ_r and θ_c such that the *total* number of cells with error indicator larger (resp., smaller) than θ_r (resp., θ_c) is (approximately) a fraction α_r (resp., α_c) of the *total* number of cells. The computation of such thresholds is not trivial in a distributed-memory context, and requires interprocessor communication. In particular, we implemented the iterative algorithm proposed in [34, Fig. 5]. In all experiments, the iterative process underlying this algorithm was stopped (at most) after 25 iterations. Besides, we set $\alpha_r = 0.15$ and $\alpha_c = 0.03$, so that, the number of cells is at least doubled at each AMR level (assuming that no cells can be coarsened). The actual number of cells at each mesh in the hierarchy, however, also depends on the algorithm within `p4est` that 2:1 balances the forest-of-octrees, as it may need to apply more refinement in order to keep this constraint; see Sect. 2.5.

In the sequel, we will focus on evaluating the following stages, which are labeled as:

- **MESH:** Includes the calls to the `Refine_and_coarsen` and `Partition` primitives; see Sect. 3.5.
- **FE SPACE SUB-ASSEMBLY:** Includes Algs. 2-4, and the computation of the C_{gh} sparse matrix and the $M(g)$ sets for all proc-local DOFs $g \in \mathcal{N}_H^P \cap \mathcal{N}_L^P$.
- **FE SPACE FULL-ASSEMBLY:** Includes **FE SPACE SUB-ASSEMBLY** and the algorithm outlined in Sect. 6.
- **ASSEMBLY SUB-ASSEMBLY:** Includes the computation of \mathbf{A}_K and \mathbf{b}_K for all local cells, the treatment of hanging DOF constraints accordingly to Tab. 1, and the incremental assembly and final compression of the local-to-subdomain coefficient matrix data structure; see [27, Sect. 11.1] for more details on this last step.
- **ASSEMBLY FULL-ASSEMBLY:** Includes **ASSEMBLY SUB-ASSEMBLY** plus the setup of the fully-assembled linear system (using the data structures in PETSc). This latter step is split into three substeps: (A) the sparsity pattern of locally owned rows is completed by means of a nearest neighbor exchange. We note that, by means of local subassembly, each processor can only *partially* determine the location of non-zero entries in locally owned rows; (B) the full sparsity pattern of locally owned rows computed in (A) is used to set up the fully-assembled matrix data structure. (C) the subassembled non-zero matrix entries $(\mathbf{A}^P)_{ij}$ and vector entries $(\mathbf{b}^P)_{ij}$ are injected, on a row-by-row fashion, into the global, fully-assembled data structures, including the final communication stage described in Sect. 6.
- **ERROR ESTIMATOR:** Includes the computation of the error estimators for all local cells, the computation of θ_r and θ_c thresholds, and to flag the cells for the `Refine_and_coarsen` primitive.
- **TOTAL SUB-ASSEMBLY:** The aggregation of **MESH**, **FE SPACE SUB-ASSEMBLY**, **ASSEMBLY SUB-ASSEMBLY**, and **ERROR ESTIMATOR**.
- **TOTAL FULL-ASSEMBLY:** The aggregation of **MESH**, **FE SPACE FULL-ASSEMBLY**, **ASSEMBLY FULL-ASSEMBLY**, and **ERROR ESTIMATOR**.

In this section, we also compare the performance and scalability of the algorithms and data structures within **FEMPAR** with those in **deal.II**. In particular, we adapted the **deal.II step-40** tutorial program⁹

⁹Documentation available at https://www.dealii.org/9.0.0/doxygen/deal.II/step_40.html.

in order to solve problem (6), and grouped the different steps in this program into stages accordingly to the classification above. There are, though, still several significant worth-noting differences among the following stages:

- (1) The **MESH** stage in **deal.II** does not actually call Alg. 1, but a fundamentally different algorithm; see Remark 3.16.
- (2) The **FE SPACE FULL-ASSEMBLY** stage in **deal.II** does not call Alg. 2, Alg. 4, nor the algorithm in Sect. 6, but the algorithm in [34, Sect. 3.1]. Besides, it also sets up the so-called *locally owned* and *locally relevant* index sets of global identifiers. These data structures are required to manipulate globally addressable, fully-assembled linear algebra matrices and vectors in **deal.II**. We refer to [34, Sect. 3] for additional details.
- (3) The **ASSEMBLY FULL-ASSEMBLY** stage in **deal.II**, as its counterpart in **FEMPAR**, comprises three different steps, with step (B) in **deal.II** being equivalent to its **FEMPAR** counterpart. Step (A) first determines, *by means of symbolic assembly*, the location of non-zero elements in the local subassembled matrix and then, a communication stage, equivalent to the one in **FEMPAR** above, allows all processors to complete the non-zero pattern of locally owned rows; in step (C) the contribution of local cells are computed and *directly* assembled into the global linear system data structures, including the transfer of entries locally computed but stored in remote processors. We refer to [34, Sect. 4] for additional details.

We do not report the computational time spent in the linear system solve stage, but focus on the algorithms presented in this paper instead. For subassembled linear systems, **FEMPAR** provides non-overlapping domain decomposition preconditioners, while, for fully-assembled ones, a bunch of algebraic preconditioners provided by third party packages through its PETSc interface.¹⁰ On the other hand, **deal.II** provides the latter set of preconditioners. We used in particular the PETSc interfaces (available at both packages) to Hypre-BoomerAMG package with default parameter values. A fair comparison among the performance and scalability of both solver types would require a much more systematic/comprehensive study to the one that we performed here, including algorithmic components selection, parameter tuning, etc., which is beyond the scope of this paper.

Fig. 4 shows the strong scalability of the algorithms and data structures in **FEMPAR** (left) for problem (6) discretized with trilinear ($Q_1(K)$) Lagrangian FEs. On the other hand, Fig. 4 (right) shows the ratio, R , among the computational times spent in the stages in **deal.II** and their counterparts in **FEMPAR**, except for the **FE SPACE SUB-ASSEMBLY** and **ASSEMBLY SUB-ASSEMBLY** stages, which are not actually implemented in **deal.II**. For these latter curves, Fig. 4 (right) reports the ratio among **FE SPACE FULL-ASSEMBLY**, **ASSEMBLY FULL-ASSEMBLY**, as implemented in **deal.II**, and **FE SPACE SUB-ASSEMBLY**, **ASSEMBLY SUB-ASSEMBLY**, resp. The reader can use these curves to observe the extra overhead which is incurred by **deal.II** in order to handle the construction of globally-addressed, fully-assembled linear systems of equations. If R is larger than one, then the corresponding stage in **FEMPAR** is faster than the one in **deal.II** by a factor R . For readability purposes, Fig. 4 (left) also provides the ideal strong scaling slope (solid black line). The more parallel a given strong scaling curve is to the ideal slope the more strongly scalable the corresponding stage is. Two different global problem sizes suitable for the [48, 12228] and [384, 30672] CPU cores range were tested. In particular, the results in Fig. 4a correspond to a problem in which the AMR hierarchy has 13 levels, resulting into a 49.8 MDOFs problem size at the last level, while in Fig. 4b, we report the ones for a 16 AMR levels hierarchy, and a 415.5 MDOFs problem size.¹¹ *These DOFs counts, and the ones in the rest of the paper, include both regular and hanging DOFs.* The

¹⁰We refer the reader to [28, 29] for the expected performance and scalability of the domain decomposition solver within **FEMPAR**.

¹¹We stress that this is by no means a parallel scaling limit of the algorithms proposed, but the largest number of cores we can exploit on MN-IV provided the access constraints that we have to this supercomputer. We expect the algorithms proposed to be able to scale up to hundreds of thousands of cores in the solution of hundred of billions DOFs problem sizes.

average load per core at the last level ranges, in Fig. 4a, from 1.04M DOFs/core to 4.06K DOFs/core, and, in Fig. 4b, from 1.08M DOFs/core to 13.53K DOFs/core.

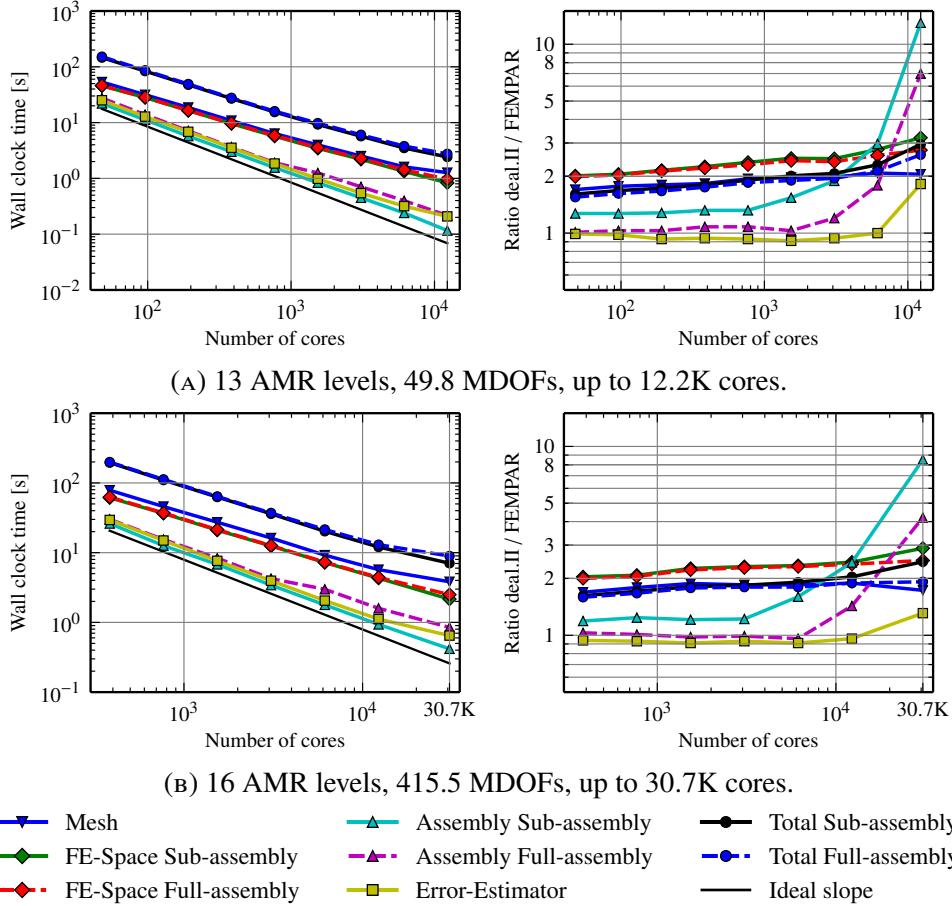


FIGURE 4. Strong scaling test results on MN-IV for FEMPAR (left) and ratio deal.II versus FEMPAR (right). Poisson problem with $Q_1(K)$ Lagrangian FEs.

Fig. 4a and 4b (left) reveal high strong scalability for the ASSEMBLY SUB-ASSEMBLY and ERROR ESTIMATOR stages in the full CPU cores range tested. For example, the parallel efficiency for the ASSEMBLY SUB-ASSEMBLY stage only decays to 74% and 78% with the largest number of cores tested in Fig. 4a and Fig. 4b, resp. An intermediate degree of strong scalability is observed for the ASSEMBLY FULL-ASSEMBLY stage, that decays to 50% and 44% in Fig. 4a and Fig. 4b, resp., due to the communication overhead involved in stages (A) and (C) outlined above. Also worth noting is the extra overhead of ASSEMBLY FULL-ASSEMBLY compared to ASSEMBLY SUB-ASSEMBLY, which ranges from 25% for 48 cores to 85% for 12.2K cores in Fig. 4a, and from 15% for 384 to 100% for 30.7K cores in Fig. 4a. This is caused by the extra stages (A)-(C) involved in the setup of fully-assembled linear systems. While less strong scalability is observed for the MESH, FE SPACE SUB-ASSEMBLY, and FE SPACE FULL-ASSEMBLY stages, the computational time reduction with the number of cores is still very significant for these stages, with the parallel efficiency decaying to 16%, 21%, and 19%, in Fig. 4a, and 26%, 36%, and 23%, resp., in Fig. 4b. This is not surprising as these three stages, and specially MESH, involves significantly more communication volume (relative to local work) than, e.g., ASSEMBLY SUB-ASSEMBLY. We also note that the computation times and strong scaling of FE SPACE FULL-ASSEMBLY compared to FE SPACE SUB-ASSEMBLY are slightly higher and lower, resp., due to the extra computation and communication involved in the execution of Alg. 4.

In Fig. 4a and 4b (right), we can observe that the algorithms and data structures in FEMPAR are either competitive (ERROR ESTIMATOR and ASSEMBLY FULL-ASSEMBLY stages, the latter provided the number

of cores is not large enough) or *significantly faster* than the ones in deal.II (**MESH**, **FE SPACE FULL-ASSEMBLY**, **ASSEMBLY FULL-ASSEMBLY**, the latter provided that the number of cores is large). The ratio in the **MESH** stage is, e.g., as large as 2.08x for the largest number of cores in Fig. 4a, and increases at a moderate pace from 1.70x to 2.08x from 48 to 12.2K cores. These results confirm that the approach selected for reconstructing our outer layer mesh representation in Sect. 2 results in a more efficient coupling of the data structures in FEMPAR and those of p4est. Even larger ratios for the **FE SPACE FULL-ASSEMBLY** stage are observed in Fig. 4a and 4b, as large as 2.75x, and 2.49x, resp., for the largest number of cores evaluated. The extra communication volume and overhead associated to handling a global numbering of DOFs suitable for fully-assembled operators in deal.II (e.g., searches on index sets of global identifiers) is the most contributing factor to this trend. If we focus on the ratios for the **ASSEMBLY FULL-ASSEMBLY** in Fig. 4a and 4b, we can clearly observe the following trend. For a moderate number of cores, the performance of both codes is almost equivalent. At these scales, computational times are dominated by local work (e.g., numerical calculations involved in the evaluation of the cell contributions to the coefficient matrix and right hand side). However, beyond a certain number of cores, the parallel efficiency of the **ASSEMBLY FULL-ASSEMBLY** stage in deal.II starts dropping significantly compared to the one in FEMPAR. Indeed, the **ASSEMBLY FULL-ASSEMBLY** stage is up to 7x and 4.2x times faster in FEMPAR compared to its counterpart in deal.II; see Fig. 4a and 4b, resp. We confirmed that the source of this difference is the large communication volume involved in steps (A) and (C) of **ASSEMBLY FULL-ASSEMBLY** in deal.II. We guess that this is in turn caused by the differences among both packages in the selection of the owner processors of interface DOFs shared among two processors; see Remark 6.1. This sudden parallel scaling drop was actually not observed in [34] (at least as evidently). We believe that this might be caused by the fact that experiments in [34] considered strong scaling results up to larger loads per core to those considered here.

Overall, *the balance achieved among all these stages ends up with **TOTAL SUB-ASSEMBLY** and **TOTAL FULL-ASSEMBLY** in FEMPAR being up to 2.95x and 2.60x faster in Fig. 4a, resp., and 2.45x and 1.92x faster in Fig 4b, resp., for the largest number of cores tested, compared to **TOTAL FULL-ASSEMBLY** in deal.II.*

In order to showcase the suitability of the proposed h -adaptive pipeline for higher than linear-order FEs, Fig. 5 shows its strong scalability for problem (6) discretized with triquadratic ($Q_2(K)$) Lagrangian FEs. The average load per core at the last level ranges, in Fig. 5a, from 0.92M DOFs/core to 3.57K DOFs/core, and, in Fig. 5b, from 1.25M DOFs/core to 15.72K DOFs/core. These loads per core are similar to those in Fig. 4. However, for a given number of cores, the load per core, in terms of number of mesh cells at the last level, is smaller in Fig. 5, compared to the one in Fig. 4. This is justified by the fact that $Q_2(K)$ FEs, in 3D, have 3.375 times as much DOFs/cell than $Q_1(K)$ FEs (i.e., 27 versus 8), and thus, a higher coupling in the sparsity pattern of the coefficient matrix (i.e., larger number of non-zero entries per row). This is reflected in our experiment, were the results in Fig. 5a correspond to a problem in which the AMR hierarchy has 9 levels (4 refinement levels less than with trilinear FEs), resulting into a 43.9 M DOFs problem size at the last level, while in Fig. 5b, we have 12 AMR levels hierarchy (3 refinement levels less), and a 482.2 MDOFs problem size.

All stages in Fig. 5 are less strongly scalable than their counterparts in Fig. 4, except for the **ASSEMBLY SUB-ASSEMBLY** stage, which is as scalable as in Fig. 4, and, by far, the most strongly scalable stage among those in Fig. 5. In any case, the computational time reduction with the number of cores is still significant for the rest of stages, except for the **MESH** and **ERROR ESTIMATOR** stages and the two largest core counts, for which any benefit of parallelism is totally absorbed by parallel overheads. Also worth noting in Fig. 5 is that **ASSEMBLY SUB-ASSEMBLY** and **ASSEMBLY FULL-ASSEMBLY** concentrate a higher percentage of **TOTAL SUB-ASSEMBLY** and **TOTAL FULL-ASSEMBLY**, resp. (compared to Fig. 4). This hides the higher drop of parallel efficiency that is observed for the **MESH** and **ERROR ESTIMATOR** stages.

As mentioned in the previous paragraph, the parallel efficiency of the **MESH** stage starts decaying sharply earlier in the number of cores for $Q_2(K)$ FEs (e.g., beyond 6.1K cores in Fig. 5b) until an extent that, for the largest two number of cores, any benefit of parallelism is totally absorbed by parallel overheads. With

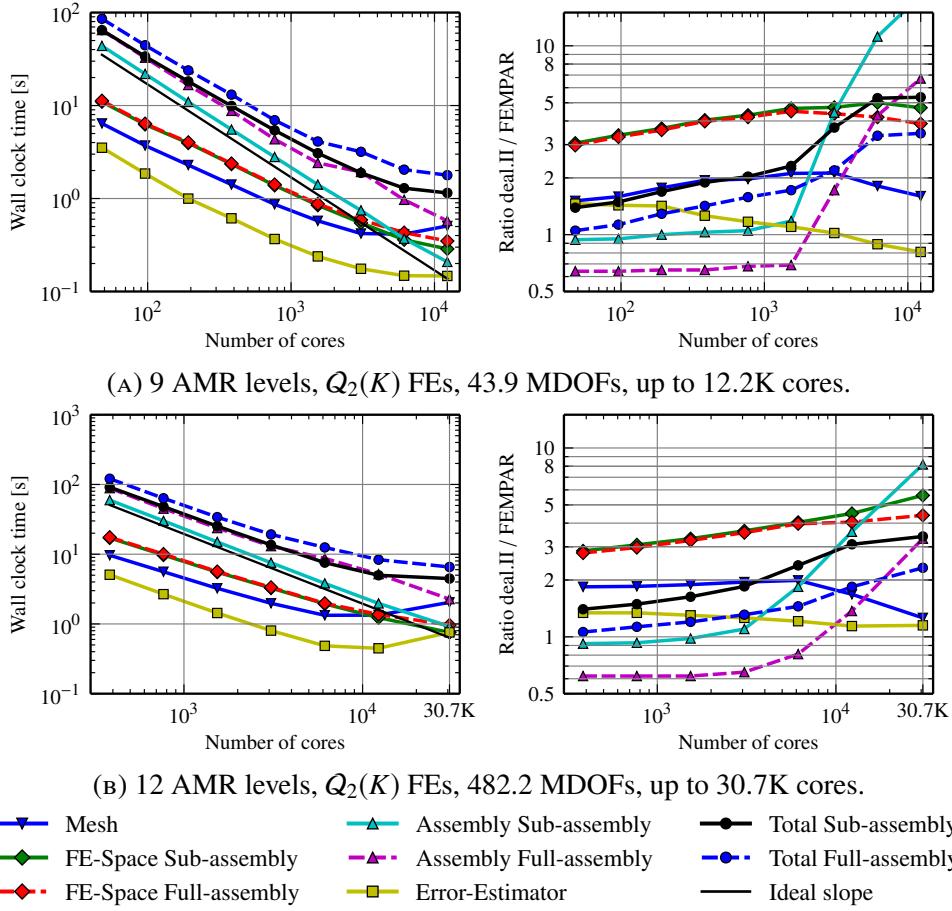


FIGURE 5. Strong scaling test results on MN-IV for FEMPAR (left) and ratio `deal.II` versus FEMPAR (right). Poisson problem with $Q_2(K)$ Lagrangian FEs.

$Q_2(K)$ FEs, the error estimate on those cells that are not concentrated around the surface, on which the force term is discontinuous, becomes rapidly tiny with the number of AMR levels, resulting in a much higher localized AMR around the sinusoidal surface (and quite different distribution of the mesh resolution across the domain). Besides, the ratio of the number of mesh cells among every two consecutive levels in the AMR hierarchy becomes consistently larger than a factor of 2x because of the extra refinement required in order to keep the forest-of-octrees 2:1 balanced. For example, for the hierarchy corresponding to Fig.4a, the average ratio among the number of cells for every two consecutive meshes across all levels, is 2.2x, while for the one in Fig.5a, 2.05x. These localized refinement patterns cause, to some extent, load and communication imbalance, and extra communication volume in the mesh handling primitives. Indeed, `p4est_balance` was observed to concentrate a significantly larger part of the computational time of the MESH stage, where the larger the number of processors, the larger the concentration. Besides, another contributing factor is the lower load per core in terms of number of cells for $Q_2(K)$ FEs.

If we compare the performance and scalability of FEMPAR and `deal.II` stages in Fig. 5a and 5b (right), rather different conclusions can be extracted depending on the stage. The MESH stage in FEMPAR is faster than the one in `deal.II` for the whole CPU core range evaluated. However, while it scales better within a first range of CPU cores (e.g., in the [48, 6.1K] range in Fig. 5a), a drop of the ratio `deal.II` versus FEMPAR is observed for this stage in Fig. 5a and Fig. 5b, with the drop in the latter being more significant than the one of the former. We believe that this drop is caused by the extra synchronization and communication volume incurred in the communication stage discussed in Remark 3.17. Although this extra overhead can be attenuated by overlapping computation and communication, we finally decided not to do it for software readability reasons. On the contrary, ASSEMBLY FULL-ASSEMBLY stage in `deal.II` is 1.6x faster

than its counterpart in FEMPAR within a first range of CPU cores in which **ASSEMBLY FULL-ASSEMBLY** is dominated by local work; this reveals that more effort should be invested in the optimization of stages (A)-(C) for $Q_2(K)$ FEs, for which, at present, FEMPAR only provides a naive, reference implementation. However, provided the number of CPU cores is sufficiently large, the parallel efficiency of the **ASSEMBLY FULL-ASSEMBLY** stage in **deal.II** starts dropping significantly compared to the one in FEMPAR, up to an extent that large **deal.II** versus FEMPAR ratios are observed, for the same reason pointed out above for $Q_1(K)$ FEs.

The FE SPACE **FULL-ASSEMBLY** stage in FEMPAR is faster and scales better than the one **deal.II** for the whole CPU core range evaluated. Indeed, larger ratios for the FE SPACE **FULL-ASSEMBLY** stage (and higher ratio increase with the number processors) are observed in Fig. 5 compared to those in Fig. 4. For example, the ratio in the FE SPACE **FULL-ASSEMBLY** stage is as large as 4.4x for the largest number of cores in Fig. 5b, and increases from 2.78x to 4.4x from 384 to 30.7K cores due to the extra communication volume and overhead associated to handling a global numbering of DOFs suitable for fully-assembled operators in **deal.II**. The **ERROR ESTIMATOR** stage is faster in FEMPAR within the whole CPU cores range in Fig. 5b, and the same applies to Fig. 5a except for the two largest number of cores. However, the drop in parallel efficiency observed in the left part of Fig. 4b and Fig. 5b is accompanied with a drop of the FEMPAR to **deal.II** ratio, moderate in the right part of Fig. 5b (in particular, a 14% ratio drop), but much more apparent in the right part of Fig. 4b (a 42% ratio drop). After thorough inspection with the help of performance analysis tools, we could confirm so far that the drop is not caused by the algorithm which determines the θ_r and θ_c thresholds, but in the actual numerical computations performed locally within each processor during this stage. Although we struggled trying to find the cause of this behaviour, and failed so far (due to the high intricacy of the software subject of study), we believe that the actual source(s) of this last behaviour might be related to technical differences in the HPC implementation and set up of this stage among both codes, such as programming language, data structure design, compiler optimizations, algorithmic optimizations, parameters set up, etc., and the ability of the resulting codes to efficiently exploit the underlying hardware architecture, i.e., cache locality exploitation, vectorization, etc.

Overall, the balance achieved among all these stages ends up with **TOTAL SUB-ASSEMBLY** and **TOTAL FULL-ASSEMBLY** in FEMPAR being up to 5.34x and 3.4x faster in Fig. 4a, resp., and 3.4x and 2.3x faster in Fig 4b, resp., for the largest number of cores tested, compared to **TOTAL FULL-ASSEMBLY** in **deal.II**.

7.3. Maxwell problem. In this section we showcase the application of our framework to the numerical solution of the Maxwell equations. We consider a 3D L-shaped domain $\Omega = [-1, 1]^3 \setminus [-1, 0]^3$ with non-homogeneous Dirichlet boundary conditions (strongly) imposed over the entire domain boundary. The problem is discretized using a global space \mathcal{V}_h that conforms to $H_0(\mathbf{curl}, \Omega) \doteq \{\mathbf{u} \in H(\mathbf{curl}, \Omega) : \mathbf{n} \times \mathbf{u} = \mathbf{0} \text{ on } \partial\Omega_D\}$. Such \mathcal{V}_h is built from local Nédélec (a.k.a. edge) FEs of first kind. (We refer to [36] for a comprehensive coverage on the general software implementation of this sort of FEs within FEMPAR.) As $D(\mathcal{V}_h) = 1$, we may use $k = 1$ and $s = 1$; see Prop. 4.3, and 5.1. However, we still used $k = 0$ and $s = 0$ for the experiments in this section. The problem reads: find $\mathbf{u}_h^0 \in \mathcal{V}_h$ such that

$$(\nabla \times \mathbf{u}_h^0, \nabla \times \mathbf{v}_h) + (\mathbf{u}_h^0, \mathbf{v}_h) = (\mathbf{f}, \mathbf{v}_h) - (\nabla \times \mathbf{E}_{\mathbf{u}_D}, \nabla \times \mathbf{v}_h) + (\mathbf{E}_{\mathbf{u}_D}, \mathbf{v}_h), \quad \forall \mathbf{v}_h \in \mathcal{V}_h, \quad (7)$$

where \mathbf{u}_D is the Dirichlet data to be imposed, and $\mathbf{E}_{\mathbf{u}_D}$ an arbitrary extension, i.e., $\mathbf{E}_{\mathbf{u}_D} = \mathbf{u}_D$ on $\partial\Omega_D$. The magnetic solution field is given by $\mathbf{u}_h \doteq \mathbf{u}_h^0 + \mathbf{E}_{\mathbf{u}_D}$. The input data \mathbf{u}_D and \mathbf{f} are set up such that the (analytical) solution of this problem becomes $\mathbf{u} = \nabla \left(r^{\frac{2}{3}} \sin \left(\frac{2t}{3} \right) \right)$, with $t = \arccos \left(\frac{xyz}{r} \right)$, (x, y, z) the Cartesian coordinates of any point within Ω , and r is the corresponding radius in 3D polar coordinates. This analytical solution has a singular behaviour near the origin and $\mathbf{u} \notin H^1(\Omega)$. We note that, for this problem, a forest-of-octrees with three adaptive octrees is sufficient in order to geometrically discretize Ω . Besides, as it has known analytical solution, we do not actually use a-posteriori error estimators, but the true error among the analytical and FE solution measured in the L_2 -norm instead. For this sort of

problem, up to the authors' knowledge, there is no deal.II step tutorial available at the public domain, so that we could not drive the comparison of FEMPAR against deal.II. In any case, deal.II supports edge h -adaptive FEs as well.

Fig. 4 shows the strong scalability of FEMPAR for problem (7) discretized with first-order Edge FEs. A single global problem size suitable for the [48, 12228] CPU cores range was tested. In 3D, these sort of FEs have 12 DOFs/cell, with a single DOF owned per each cell edge. Thus, among others, there is a higher coupling in the sparsity pattern of the coefficient matrix (i.e., larger number of non-zero entries per row) compared to first-order Lagrangian FEs. This is reflected in the problem corresponding to Fig. 6, where an AMR hierarchy with 11 levels already leads to a problem with 115.6 MDOFs at the last level. The averaged load per core at the last level ranges from 2.41M DOFs/core (48 processors) to 9.40K DOFs/core (12.2K processors).

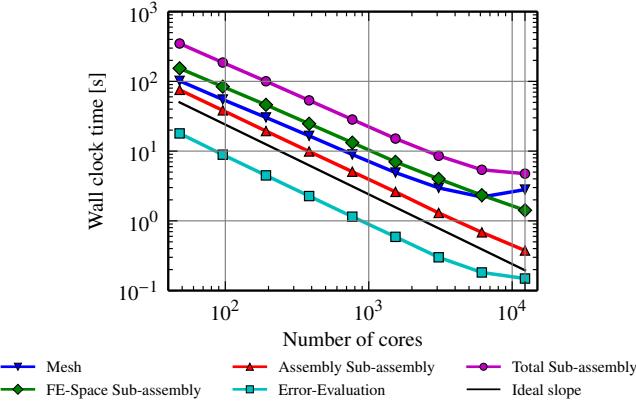


FIGURE 6. 11 AMR levels, $\mathcal{ND}_1(K)$ FEs, 115.6 MDOFs, up to 12.2K cores.

FIGURE 7. FEMPAR strong scaling test results on MN-IV. Maxwell problem with first order ($\mathcal{ND}_1(K)$) Edge FEs.

Fig. 6 confirms, as with the Poisson problem discretized with Lagrangian FEs, remarkable strong scalability for all stages, and significant computational time reductions for TOTAL SUB-ASSEMBLY in the whole range of CPU cores analyzed, despite the more general structure and complexity of implementation underlying H -curl conforming FE spaces.

8. CONCLUSIONS

In this paper we have presented the building blocks, i.e., data structures and algorithms, of a highly scalable parallel *generic FE* simulation framework that supports AMR via forest-of-trees endowed with SFCs, thus being suitable for the efficient modelling of physical problems with multiple scales on a wide range of large scale application problems governed by PDEs. We, in particular, discussed mathematically-proven algorithms for three key stages in the simulation pipeline, namely scalable mesh handling, construction of generic global conforming FE spaces on top of non-conforming meshes using hanging DOF constraints, and the parallel subdomain-wise sub-assembly and full-assembly of the discrete linear system of algebraic equations. Central to the framework is a generic FE-suitable adaptive mesh representation that is grounded on concepts that are not tailored to a particular cell topology, SFC, or number of space dimensions. All that the framework requires from the forest-of-trees layer meshing engine to be able to build such mesh data structure is a description of the local neighborhood across the cell boundaries in the adapted mesh. Along the way, we have (a) mathematically justified which are the crucial benefits (i.e., ease of implementation and parallelization, high scalability) that one enjoys from enforcing the 2:1 k -balance constraint, (b) identified what a conforming FE formulation must fulfill such that it can be implemented in parallel given the constraint that each processor only has access to

the off-processor cells in the s -ghost cell set, and (c) and determined the largest possible value of k and s that lets one still implement such FE formulation. The software implementation of these algorithms is available at FEMPAR [27], an open source OO Fortran200X scientific software package for the HPC simulation of complex multiphysics problems governed by PDEs at large scales.

Besides, we conducted a comprehensive strong scaling study of FEMPAR on a petascale supercomputer when applied to both Poisson and Maxwell PDE problems, for which the usage of AMR is highly beneficial for computational efficiency (if not essential). The study reveals remarkable scalability of the framework up to several tens of thousands of CPU cores in the solution of problems with several hundreds of millions of degrees of freedom. Besides, we also conducted a performance and strong scalability comparison of FEMPAR against an excellent, state-of-the-art scientific software, i.e., the `deal.II` FE library. This comparison reveals that the algorithms proposed here are at least competitive with the ones of `deal.II`, and, at most, *significantly faster*.

The usage of parallel scalable AMR is still rather limited by most scientific computing practitioners and researchers, mostly because of the underlying complexity behind developing such tool. We expect this paper to be successful in convincing the reader that the design and development of the algorithms and data structures behind this tool is affordable and of reasonable complexity. With this in mind, while still using a high-level approach when presenting the algorithms, we introduced sufficient discussion and associated details to be helpful in this regard. Besides, wherever applicable, and in contrast to the other related work in the literature, the algorithms are supplemented with mathematical propositions and proofs that support their correctness.

REFERENCES

- [1] C. R. Dohrmann. A Preconditioner for Substructuring Based on Constrained Energy Minimization. *SIAM Journal on Scientific Computing*, 25(1):246–258, 2003. doi:10.1137/S1064827502412887.
- [2] J. Holke. Scalable Algorithms for Parallel Tree-based Adaptive Mesh Refinement with General Element Types. 2018.
- [3] R. A. Finkel and J. L. Bentley. Quad trees a data structure for retrieval on composite keys. *Acta Informatica*, 4(1):1–9, 1974. doi:10.1007/BF00288933.
- [4] D. Meagher. Geometric modeling using octree encoding. *Computer Graphics and Image Processing*, 19(2):129–147, 1982. doi:10.1016/0146-664X(82)90104-6.
- [5] H. Sundar, R. S. Sampath, S. S. Adavani, C. Davatzikos, and G. Biros. Low-constant parallel algorithms for finite element simulations using linear octrees. In *Proceedings of the 2007 ACM/IEEE conference on Supercomputing - SC '07*, page 1, New York, New York, USA, 2007. ACM Press. doi:10.1145/1362622.1362656.
- [6] T. Tiankai Tu, D. O'Hallaron, and O. Ghattas. Scalable Parallel Octree Meshing for TeraScale Applications. In *ACM/IEEE SC 2005 Conference (SC'05)*, pages 4–4. IEEE, 2005. doi:10.1109/SC.2005.61.
- [7] H. Sundar, R. S. Sampath, and G. Biros. Bottom-Up Construction and 2:1 Balance Refinement of Linear Octrees in Parallel. *SIAM Journal on Scientific Computing*, 30(5):2675–2708, 2008. doi:10.1137/070681727.
- [8] C. Burstedde, L. C. Wilcox, and O. Ghattas. p4est: Scalable Algorithms for Parallel Adaptive Mesh Refinement on Forests of Octrees. *SIAM Journal on Scientific Computing*, 33(3):1103–1133, 2011. doi:10.1137/100791634.
- [9] T. Isaac, C. Burstedde, L. C. Wilcox, and O. Ghattas. Recursive Algorithms for Distributed Forests of Octrees. *SIAM Journal on Scientific Computing*, 37(5):C497–C531, 2015. doi:10.1137/140970963.
- [10] C. Burstedde and J. Holke. A Tetrahedral Space-Filling Curve for Nonconforming Adaptive Meshes. *SIAM Journal on Scientific Computing*, 38(5):C471–C503, 2016. doi:10.1137/15M1040049.
- [11] R. S. Sampath, S. S. Adavani, H. Sundar, I. Lashuk, and G. Biros. Dendro: Parallel algorithms for multigrid and AMR methods on 2:1 balanced octrees. In *2008 SC - International Conference*

- for High Performance Computing, Networking, Storage and Analysis*, pages 1–12. IEEE, 2008. doi:10.1109/SC.2008.5218558.
- [12] C. Burstedde, M. Burtscher, O. Ghattas, G. Stadler, T. Tu, and L. C. Wilcox. ALPS: A framework for parallel adaptive PDE solution. *Journal of Physics: Conference Series*, 180(1):012009, 2009. doi:10.1088/1742-6596/180/1/012009.
- [13] R. S. Sampath and G. Biros. A Parallel Geometric Multigrid Method for Finite Elements on Octree Meshes. *SIAM Journal on Scientific Computing*, 32(3):1361–1392, 2010. doi:10.1137/090747774.
- [14] L. C. Wilcox, G. Stadler, C. Burstedde, and O. Ghattas. A high-order discontinuous Galerkin method for wave propagation through coupled elastic–acoustic media. *Journal of Computational Physics*, 229(24):9373–9396, 2010. doi:10.1016/J.JCP.2010.09.008.
- [15] M. Olm, S. Badia, and A. F. Martín. Simulation of High Temperature Superconductors and experimental validation. *Computer Physics Communications*, 237:154–167, 2018. doi:10.1016/J.CPC.2018.11.021.
- [16] G. Stadler, M. Gurnis, C. Burstedde, L. C. Wilcox, L. Alisic, and O. Ghattas. The dynamics of plate tectonics and mantle flow: from local to global scales. *Science (New York, N.Y.)*, 329(5995):1033–8, 2010. doi:10.1126/science.1191223.
- [17] J. Rudi, O. Ghattas, A. C. I. Malossi, T. Isaac, G. Stadler, M. Gurnis, P. W. J. Staar, Y. Ineichen, C. Bekas, and A. Curioni. An extreme-scale implicit solver for complex PDEs: highly heterogeneous flow in earth’s mantle. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis on - SC ’15*, pages 1–12, New York, New York, USA, 2015. ACM Press. doi:10.1145/2807591.2807675.
- [18] C. Burstedde, O. Ghattas, G. Stadler, T. Tu, and L. C. Wilcox. Parallel scalable adjoint-based adaptive solution of variable-viscosity Stokes flow problems. *Computer Methods in Applied Mechanics and Engineering*, 198(21-26):1691–1700, 2009. doi:10.1016/J.CMA.2008.12.015.
- [19] V. Akcelik, T. Tu, J. Urbanic, J. Bielak, G. Biros, I. Epanomeritakis, A. Fernandez, O. Ghattas, E. J. Kim, J. Lopez, and D. O’Hallaron. High Resolution Forward And Inverse Earthquake Modeling on Terascale Computers. In *Proceedings of the 2003 ACM/IEEE conference on Supercomputing - SC ’03*, page 52, New York, New York, USA, 2003. ACM Press. doi:10.1145/1048935.1050202.
- [20] C. Burstedde, O. Ghattas, M. Gurnis, G. Stadler, Eh Tan, T. Tu, L. C. Wilcox, and S. Zhong. Scalable adaptive mantle convection simulation on petascale supercomputers. In *2008 SC - International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–15. IEEE, 2008. doi:10.1109/SC.2008.5214248.
- [21] E. Neiva, S. Badia, A. F. Martín, and M. Chiumenti. A scalable parallel finite element framework for growing geometries. Application to metal additive manufacturing. *International Journal for Numerical Methods in Engineering*, In press, 2019. doi:10.1002/nme.6085.
- [22] M. Ainsworth, J. T. J. T. Oden, and Wiley InterScience (Online service). *A posteriori error estimation in finite element analysis*. Wiley, 2000.
- [23] D. Schillinger and M. Ruess. The Finite Cell Method: A Review in the Context of Higher-Order Structural Analysis of CAD and Image-Based Geometric Models. *Archives of Computational Methods in Engineering*, 22(3):391–455, 2015. doi:10.1007/s11831-014-9115-y.
- [24] E. Burman, S. Claus, P. Hansbo, M. G. Larson, and A. Massing. CutFEM: Discretizing Geometry and Partial Differential Equations. *International Journal for Numerical Methods in Engineering*, 104(7):472–501, 2015. doi:10.1002/nme.4823.
- [25] S. Badia, A. F. Martín, and F. Verdugo. Mixed aggregated finite element methods for the unfitted discretization of the Stokes problem. *SIAM Journal on Scientific Computing*, 40(6):B1541–B1576, 2018. doi:10.1137/18M1185624.
- [26] S. Badia, F. Verdugo, and A. F. Martín. The aggregated unfitted finite element method for elliptic problems. *Computer Methods in Applied Mechanics and Engineering*, 336:533–553, 2018. doi:10.1016/j.cma.2018.03.022.

- [27] S. Badia, A. F. Martín, and J. Principe. FEMPAR: An Object-Oriented Parallel Finite Element Framework. *Archives of Computational Methods in Engineering*, 25(2):195–271, 2018. doi:10.1007/s11831-017-9244-1.
- [28] S. Badia, A. Martín, and J. Principe. A Highly Scalable Parallel Implementation of Balancing Domain Decomposition by Constraints. *SIAM Journal on Scientific Computing*, 36(2):C190—C218, 2014. doi:10.1137/130931989.
- [29] S. Badia, A. F. Martín, and J. Principe. Multilevel Balancing Domain Decomposition at Extreme Scales. *SIAM Journal on Scientific Computing*, 38(1):C22—C52, 2016. doi:10.1137/15M1013511.
- [30] W. Bangerth, R. Hartmann, and G. Kanschat. deal.II—A general-purpose object-oriented finite element library. *ACM Transactions on Mathematical Software*, 33(4), 2007. doi:10.1145/1268776.1268779.
- [31] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells. The FEniCS Project Version 1.5. *Archive of Numerical Software*, 3(100), 2015. doi:10.11588/ans.2015.100.20553.
- [32] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, K. Rupp, B. F. Smith, S. Zampini, H. Zhang, and H. Zhang. PETSc Users Manual. Technical Report ANL-95/11 - Revision 3.7, Argonne National Laboratory, 2016.
- [33] M. A. Heroux, R. A. Bartlett, V. E. Howle, R. J. Hoekstra, J. J. Hu, T. G. Kolda, R. B. Lehoucq, K. R. Long, R. P. Pawlowski, E. T. Phipps, A. G. Salinger, H. K. Thornquist, R. S. Tuminaro, J. M. Willenbring, A. Williams, and K. S. Stanley. An Overview of the Trilinos Project. *ACM Trans. Math. Softw.*, 31(3):397–423, 2005. doi:10.1145/1089014.1089021.
- [34] W. Bangerth, C. Burstedde, T. Heister, and M. Kronbichler. Algorithms and data structures for massively parallel generic adaptive finite element codes. *ACM Trans. Math. Softw.*, 38(2):14:1–14:28, 2012. doi:10.1145/2049673.2049678.
- [35] P. Kus and J. Sistek. Coupling parallel adaptive mesh refinement with a nonoverlapping domain decomposition solver. *Advances in Engineering Software*, 110:34–54, 2017. doi:10.1016/j.advengsoft.2017.03.012.
- [36] M. Olm, S. Badia, and A. F. Martín. On a general implementation of h- and p-adaptive curl-conforming finite elements. *Advances in Engineering Software*, 132:74–91, 2019. doi:10.1016/J.ADVENGSOFT.2019.03.006.
- [37] G. Karypis and V. Kumar. Parallel Multilevel series k-Way Partitioning Scheme for Irregular Graphs. *SIAM Review*, 41(2):278–300, 1999. doi:10.1137/S0036144598334138.
- [38] C. Burstedde, J. Holke, and T. Isaac. On the Number of Face-Connected Components of Morton-Type Space-Filling Curves. *Foundations of Computational Mathematics*, pages 1–26, 2018. doi:10.1007/s10208-018-9400-5.
- [39] D. W. Kelly, J. P. De S. R. Gago, O. C. Zienkiewicz, and I. Babuska. A posteriori error analysis and adaptive processes in the finite element method: Part I—error analysis. *International Journal for Numerical Methods in Engineering*, 19(11):1593–1619, 1983. doi:10.1002/nme.1620191103.