Locally implicit discontinuous Galerkin time-domain methods for electromagnetic wave propagation in biological tissues

Ludovic Moya

UNIVERSITÉ NICE SOPHIA ANTIPOLIS

ÉCOLE DOCTORALE SFA
SCIENCES FONDAMENTALES
ET APPLIQUÉES

T H È S E

pour obtenir le titre de

Docteur en Sciences

de l'Université de Nice Sophia Antipolis
Mention : MATHÉMATIQUES

Présentée et soutenue par

Ludovic MOYA

Méthodes Galerkine discontinues localement implicites en domaine temporel pour la propagation des ondes électromagnétiques dans les tissus biologiques

Thèse dirigée par Stéphane LANTERI
Co-dirigée par Stéphane DESCOMBES

préparée à l'Inria Sophia Antipolis

soutenue le 16 Décembre 2013

Jury :

Rapporteurs :	Marcus GROTE	-	Mathematisches Institut Universität Basel
	Willem HUNDSDORFER	-	Centrum voor Wiskunde en Informatica
Examinateurs :	Julien DIAZ	-	Inria (Magique-3D)
	Cédric GALUSINSKI	-	Université du Sud Toulon Var
	Robert STARAJ	-	LEAT
Directeur :	Stéphane LANTERI	-	Inria (Nachos)
Co-Directeur :	Stéphane DESCOMBES	-	Université Nice Sophia Antipolis

Contents

Introduction

L'électromagnétisme numérique

L'unification de l'électricité et du magnétisme a été achevée par le physicien et mathématicien James Clerk Maxwell dans les années 1860. Quand Maxwell établit sa théorie de l'électromagnétisme, il aboutit à un systéme d'Équations aux Dérivées Partielles (EDP), aujourd'hui définit comme les équations de Maxwell, qui lie les champs électriques et magnétiques entre eux et régit tous les phénomènes classiques de propagation d'ondes électromagnétiques. De nos jours, bien que les principes de l'électromagnétisme sont bien compris, leur application à des configurations pratiques est compliquée et bien au-delà du calcul manuel, excepté dans les cas les plus simples. Les premières analyses en électromagnétisme ont été menées pour des formes simples présentant un degré de symmétrie élevé telles que des sphères, des cylindres ou des plans ; la recherche de solution pour des géométries plus complexes a nécessité le développement de méthodes d'approximation. Avec l'avènement de l'informatique, les méthodes numériques ont été développées dans les années 1960 afin de permettre plus de souplesse et de précision dans l'approximation des solutions. De nos jours, les ingénieurs et les scientifiques utilisent des ordinateurs allant des simples machines de bureau aux supercalculateurs massivement parallèles pour obtenir des solutions aux équations de Maxwell.

Les phénomènes électromagnétiques sont si présents dans les sciences et technologies modernes que les champs d'application de l'électromagnétisme numérique sont extrêmement vastes. Dans le passé, la recherche de solutions précises aux équations de Maxwell a principalement été motivée par des enjeux dans le domaine de la défense militaire comme la technologie micro-ondes radar ou la vulnérabilité électromagnétique, puis l'électromagnétisme numérique s'est rapidement orienté vers des applications industrielles et sociales. L'une des premières applications de l'analyse numérique en électromagnétisme fut l'analyse et le design d'antennes. Les antennes jouent un rôle essentiel dans la communication sans fil, la télédétection, l'exploration de l'espace, la défense, la guerre électronique, et dans de nombreux autres systèmes électroniques. Avec la croissance explosive des appareils électroniques dans la vie quotidienne, des applications plus récentes ont émergé telles que l'analyse des interférences électromagnétiques et les problèmes de compatibilité électromagnétique. Tout appareil électronique émet un champ électromagnétique qui peut potentiellement interférer avec d'autres appareils, menaçant ainsi leur bon fonctionnement. Un composant dans un dispositif électronique peut également interférer avec d'autres composants du même dispositif et affecter sa performance globale. Dans de tels cas, il est essentiel d'identifier les sources d'interférence et d'éliminer ou réduire le niveau d'interférence. Depuis les années 1990, nous assistons à une très forte croissance des dispositifs de communication sans fil accessibles aux consommateurs tels que les téléphones portables dont les radiations interagissent avec les tissus humains. Une prérogative indispensable est alors de répondre à la norme de sécurité d'exposition aux micro-ondes. L'indice de Débit d'Absorption Spécifique (DAS) qui renseigne sur la quantité d'énergie véhiculée par les radiofréquences émises vers l'usager par un appareil radioélectrique doit être suffisamment faible pour ne pas représenter un danger pour l'utilisateur. La propagation des ondes électromagnétiques à travers les tissus humains est également au coeur de nombreuses applications biomédicales, telles que l'imagerie par micro-ondes des tumeurs cancéreuses, l'hyperthermie utilisée

comme une stratégie d'immunothérapie contre le cancer, ou encore la technique d'électroporation [Miklavčič *et al.* 2000, Šel *et al.* 2005, Serša 2005, Sukharev *et al.* 1992, Tsong 1991].

La simulation numérique a contribué à une meilleure compréhension des phénomènes physiques liés à la propagation des ondes électromagnétiques, et le large éventail d'applications a motivé des recherches approfondies pour obtenir des méthodes numériques bien adaptées restituant au mieux les caractéristiques mathématiques des équations de Maxwell. Les méthodes numériques sont fondamentales pour la simulation et la résolution de problèmes complexes qui n'admettent pas de solutions analytiques, cependant, plusieurs difficultés surviennent lorsqu'il s'agit de modéliser et de développer des méthodes numériques pour simuler la propagation des ondes électromagnétiques. D'une part, la plupart des problèmes de propagation d'ondes électromagnétiques sont des problèmes non-bornés. Tronquer un domaine non-borné par un domaine de calcul fini est alors l'un des défis auquel les méthodes numériques doivent répondre. D'autre part, les caractéristiques géométriques (formes irrégulières, singularités géométriques) et physiques (hétérogénéité, dispersion physique et dissipation) des milieux de propagation, ainsi que les caractéristiques des différentes sources (fils, etc) sont de plus en plus complexes. Résoudre les équations de Maxwell pour des milieux de propagation complexes prenant en compte la nature des différentes sources nécessite alors le développement d'algorithmes robustes. Enfin, les applications d'intérêt industriel conduisent à la résolution numérique de systèmes discrets dont la taille, évaluée en terme du nombre d'inconnues pour atteindre une précision donnée, est très grande. Par exemple, les longueurs caractéristiques sont généralement de l'ordre de quelques dizaines de longueur d'onde mais peuvent dépasser la centaine, conduisant à des maillages qui peuvent contenir jusqu'à dix millions de mailles pour des maillages volumiques. En trois dimension, avec un minimum de six champs, pour calculer chacune des trois composantes du champ électrique et du champ magnétique, la résolution numérique de problèmes de cette taille ne peut se faire qu'en exploitant pleinement les possibilités des calculateurs parallèles. L'algorithme généré par la méthode numérique doit alors avoir un haut degré de parallélisme.

Deux groupes de méthodes ont été développés pour résoudre les problèmes de propagation d'ondes électromagnétiques, selon que l'on souhaite considérer les équations de Maxwell en domaine temporel ou en domaine fréquentiel, les deux formulations étant liées l'une à l'autre par la transformée de Fourier. Le choix de la méthode est directement lié à l'application visée. Pour certains types de problèmes, une dépendance harmonique en temps peut être établie conduisant naturellement à la formulation des équations de Maxwell en domaine fréquentiel. Les méthodes numériques associées nécessitent alors la résolution d'un système linéaire pour chaque fréquence, et la méthode est donc naturellement implicite. Les méthodes fréquentielles sont bien adaptées pour des problèmes où de nombreuses excitations sont considérées comme dans une analyse de diffusion monostatique. En outre, étant donné que ces méthodes approximent les solutions des équations de Maxwell à chaque fréquence, elles permettent de traiter facilement les milieux dispersifs, dans lesquels les paramètres constitutifs dépendent de la fréquence. En effet, dans les équations de Maxwell le caractère dispersif du milieu de propagation est pris en compte à travers la permittivité électrique, la perméabilité magnétique et la conductivité électrique qui sont, en général, des tenseurs dont les entrées dépendent de l'espace et de la fréquence. Il existe plusieurs modèles mathématiques traduisant l'évolution fréquencielle de ces coefficients, tels que les modèles de Debye, de Lorentz, de Drude, etc. Dans cette thèse, nous allons considérer les équations de Maxwell en domaine temporel. Parmi les méthodes numériques destinées à la résolution des équations de Maxwell instationnaires la méthode DFDT (Différences Finies en Domaine Temporel) introduite par par K.S. Yee dans [Yee 1966]

reste la plus répendue et est souvent utilisée comme méthode de référence [Inan & Marshall 2011, Taflove & Hagness 2005]. Cette popularité est principalement due à la facilité d'implémentation et aux faibles temps de calcul. Dans la méthode DFDT, le domaine de calcul est discrétisé par un maillage structuré (Cartésien), simplifiant ainsi le processus de discrétisation mais représentant aussi la principale limitation de cette méthode lorsque des objets à géométrie complexe entrent en jeu. Cette incapacité à gérer efficacement des géométries complexes a conduit au développement de méthodes alternatives comme les méthodes EFDT (Éléments Finis en Domaine Temporel), VFDT (Volume Finis en Domaine Temporel) ou encore GDDT (Galerkine Discontinue en Domaine Temporel) [Cockburn *et al.* 2000, Hesthaven & Warburton 2008, Jin & Riley 2008, Rao 1999]. La grande majorité de ces méthodes repose sur un schéma d'intégration en temps explicite. Les méthodes numériques en domaine temporel sont bien adaptées pour des problèmes large bande où l'on cherche une solution sur une large bande de fréquences avec seulement quelques excitations. En outre, ces méthodes approximent les solutions des équations de Maxwell instationnaires étape par étape, par conséquent, elles peuvent traiter efficacement les problèmes non linéaires, où les propriétés électromagnétiques du domaine de calcul changent en fonction de l'intensité des champs. Dans le cas de milieux dispersifs, la modélisation requiert alors l'usage de techniques spécifiques, afin de transformer l'évolution fréquentielle des paramètres constitutifs des matériaux dans le domaine temporel.

Objectifs et plan de la thèse

Dans ce travail nous considérons les équations de Maxwell instationnaires du premier ordre. Le premier objectif est de proposer des méthodes d'éléments finis discontinues d'ordre élevé (interpolation polynomiale), reposant sur des triangulations (cas 2D) ou des tétraédrisations (cas 3D) du domaine de calcul et des schémas d'intégration en temps efficaces en présence de maillages localement raffinés. Nous allons examiner des méthodes GDDT s'appuyant sur une interpolation polynomial nodale d'ordre arbitrairement élevé (éléments finis de Lagrange) pour les composantes du champ électromagnétique sur chaque simplexe. La méthode Galerkine discontinue combine d'importantes propriétés des méthodes éléments finis et volumes finis permettant d'obtenir des schémas strictement locaux et une grande précision d'approximation. Ce choix est alors motivé par le fait que les méthodes GDDT sont particulièrement bien adaptées aux problèmes mentionnés ci-dessus, qui se posent lorsqu'il s'agit de simuler la propagation des ondes électromagnétiques. Les méthodes GDDT existantes pour la résolution des équations Maxwell instationnaires s'appuient le plus souvent sur des schémas d'intégration en temps explicites, les conditions de stabilité en présence de maillages localement raffinés peuvent alors conduire à des pas de temps particulièrement contraignants, notamment en 3D, entraînant des temps de calcul parfois rédhibitoires. Pour pallier à ce problème, nous allons étudier des schémas localement implicites (schémas hybrides implicites/explicites) ayant de bonnes propriétés de stabilité et de précision. Ces méthodes hybrides consistent à appliquer un schéma d'intégration en temps implicite dans les regions raffinées du maillage et un schéma explicite dans la partie complémentaire du domaine de calcul. Les techniques hybrides implicites/explicites sont classiques et bien connues en mécanique des fluides numérique, où par exemple le terme d'advection est traité implicitement et le terme de diffusion explicitement [Ascher *et al.* 1995, W. Hundsdorfer 2003, Hundsdorfer & Verwer 2003, Wesseling 2001]. En

électromagnétisme numérique de telles approches restent peu exploitées en comparaison des méthodes de pas de temps local explicites qui consitent à adapter le pas de temps aux cellules en les regroupant dans différentes classes d'intégration. Ces méthodes permettent l'utilisation de petits pas de temps lorsque cela est strictement nécessaire (sur les petites cellules) et l'utilisation d'un pas de temps moins contraignant sur les cellules de taille plus importante, économisant ainsi de nombreux calculs [Diaz & Grote 2009, Grote & Mitkova 2010, Monseny *et al.* 2008, Piperno 2006, Taube *et al.* 2009]. Contrairement aux méthodes de pas de temps local explicites, les méthodes localement implicites fournissent un unique pas de temps global assurant la stabilité, uniquement déterminé par la partie du maillage dédiée au traitement explicite (constituée des plus grosse cellules). Dans cette thèse, nous présentons une étude théorique complète (formulation, stabilité, convergence, analyse numérique du traitement implicite) et une comparaison des deux méthodes GDDT localement implicites pour la formulation du premier ordre des équations de Maxwell et plusieurs expériences numériques pour des problèmes de propagation d'ondes électromagnétiques en 2D et 3D, illustrant les résultats théoriques et mettant en évidence les propriétés attractives de ces approches. La première méthode localement implicite que nous considérons a initialement été introduite par Piperno dans [Piperno 2006]. La discrétisation en espace est issue d'une méthode Galerkine discontinue nodale et le schéma d'intégration en temps est basé sur le couplage du schéma explicite du second ordre saute-mouton (LF2) et du schéma implicite du second ordre Crank-Nicolson (CN2). Une analyse de stabilité et plusieurs experiences numériques en 3D ont été présentées dans [Dolean *et al.* 2010], cependant aucune preuve de convergence n'est donnée. Dans cette thèse, une analyse de convergence de la méthode est réalisée. En particulier, sur la base de cette preuve théorique et quelques expériences numériques, nous verrons que le découpage implicite/explicite peut entrainer une réduction de l'ordre de convergence. Cette possible réduction d'ordre a motivé l'étude d'un deuxième schéma d'intégration en temps localement implicite, initialement proposé par Verwer dans [Verwer 2010]. Cette méthode est également un couplage du schéma explicite saute mouton et du schéma implicite Crank-Nicolson qui est directement appliqué sur une formulation générale du système de Maxwell semi-discret. Verwer a prouvé qu'en adoptant le découpage implicite/explicite proposé dans [Verwer 2010] aucune réduction d'ordre n'apparaît, c'est à dire que la méthode conserve l'ordre deux. Enfin, en adoptant une discrétisation spatiale de type différences finies l'auteur donne des résultats numériques pour des problèmes de propagation d'ondes électromagnétiques en 1D et 2D. Dans cette thèse, nous étudions cette stratégie d'intégration en temps avec une discrétisation spatiale GD. Une analyse de la stabilité est menée et les grandes similitudes avec la méthode GDDT localement implicite de [Piperno 2006] sont mis en évidence. En outre, les résultats numériques obtenus pour des problèmes de propagation d'ondes électromagnétiques en 3D sur des maillages non-uniformes, localement raffinés, mettent en avant les propriétés attractives de ce schéma d'intégration en temps localement implicite couplé avec une discrétisation spatiale GD. Enfin, ces résultats ont conduit à une comparaison rigoureuse entre les deux méthodes GDDT localement implicites ainsi qu'à une comparaison avec les méthodes GDDT tout explicite et tout implicite.

La méthode Galerkine discontinue permet d'augmenter facilement l'ordre de convergence en espace, un autre objectif de ce travail est donc de proposer des stratégies permettant d'augmenter l'ordre de convergence des schémas d'intégration en temps localement implicites. En considérant le précédent schéma d'intégration en temps d'ordre deux comme schéma de base, des méthodes telles que les stratégies de compositions symétriques ou l'extrapolation de Richardson,

sont étudiées. De telles techniques pour augmenter l'ordre de convergence sont bien connues pour des problèmes faisant intervenir des Équations Différentielles Ordinaires (EDO) [Hairer *et al.* 1993, Hairer *et al.* 2010], mais restent peu exploitées pour des problèmes de type EDP. A partir d'un schéma de base, ces stratégies sont essentiellement une combinaison appropriée de ce dernier, appliqué avec différentes tailles de pas de temps, et donc leurs implémentations sont particulièrement aisées, ce qui rend ces techniques attractives.

Le traitement numérique de milieux de propagation complexes (c'est à dire de modèles physiques de dispersion) est également l'un des objectifs de la présente étude. L'application envisagée est l'interaction des ondes électromagnétiques avec les tissus biologiques. Comme mentionné précédemment, l'étude de l'interaction entre les ondes électromagnétiques et les tissus biologiques est au coeur de nombreuses applications dans le domaine biomédicales, telles que l'évaluation des effets potentiellement nocifs des champs électromagnétiques ou l'utilisation des ondes électromagnétiques à des fins thérapeutiques ou diagnostic. La modélisation de l'interaction des ondes électromagnétiques avec les tissus biologiques nécessite de résoudre le système d'équations de Maxwell couplé à des modèles appropriés de dispersion dans les tissus, comme le modèle de Debye. Ce dernier est le plus souvent utilisé pour les interactions des ondes électromagnétiques avec des substances à base d'eau, tels que les tissus biologiques. Dans cette thèse, nous formulons un schéma GDDT localement implicite pour ce modèle de dispersion. A notre connaissance, dans ce contexte, de telles méthodes n'ont jamais été exploitées, alors que les propriétés de ces dernières sont bien adaptées à la miniaturisation des dispositifs électroniques et des antennes ou encore à la petite taille des cellules cancéreuses, pour lesquelles l'utilisation de maillages non-uniformes, localement raffinés, est certainement une prérogative importante pour obtenir une solution numérique précise.

Le plan du manuscrit de thèse est le suivant. Le premier chapitre donne un bref aperçu des équations de Maxwell instationnaires et de leurs fondements, et introduit le modèle mathématique que nous considérons par la suite. Le Chapitre 2 est consacré à la méthode Galerkine discontinue. Après l'introduction des principes de base, un traitement complet de la méthode GD pour la discrétisation en espace des équations de Maxwell est présenté. À partir du système semi-discret obtenu, le Chapitre 3 introduit deux schémas d'intégration en temps localement implicites. Une analyse numérique complète des deux méthodes GDDT résultantes est alors présentée. Le Chapitre 4 est dédié aux expériences numériques pour des problèmes de propagation d'ondes électromagnétiques en 2D et 3D. Un soin particulier est apporté afin de confirmer les résultats théoriques obtenus dans le chapitre précédent, et une comparaison numérique entre les différentes stratégies d'intégration en temps est réalisée. Les schémas GDDT présentés dans le Chapitre 3 sont des méthodes d'ordre deux. Le Chapitre 5 introduit différentes stratégies pour augmenter l'ordre de convergence de ces méthodes. Comme mentionné précédemment, la méthode de Galerkine discontinue permet d'augmenter facilement l'ordre de convergence en espace, l'objectif est donc d'exploiter cette propriété afin d'évaluer l'approche localement implicite pour des schémas d'intégration en temps d'ordre plus élevé. Finalement, le Chapitre 6 est consacré à la propagation des ondes électromagnétiques dans des milieux dispersifs, en particulier dans les tissus biologiques. Dans ce dernier chapitre, nous introduisons la formulation du premier ordre des équations de Maxwell pour le modèle de dispersion de Debye. Le caractère dispersif est pris en compte par une équation différentielle auxiliaire et le traitement numérique du modèle mathématique sous-jacent est déduit de l'une des méthodes GDDT localement implicites présentées dans le Chapitre 3. En outre, la stabilité et la convergence de la méthode résultante sont étudiées, et des problèmes réalistes en 3D sont présentés.

Liste des publications et conférences

Articles

- S. Descombes C. Durochat. S. Lanteri, L. Moya, C. Scheid and J. Viquerat. *Recent advances on a DGTD method for time-domain electromagnetics*. To appear in Photonics and Nanostructures - Fundamentals and Applications (2013). http://dx.doi.org/10.1016/j.photonics.2013.06.005

- S. Descombes, S. Lanteri and L. Moya. *Locally implicit time integration strategies in a discontinuous Galerkin method for Maxwell's equations*. J. SCI. Comput., vol. 56, no. 1, pages 190-218, 2013.

- L. Moya. *Temporal convergence of a locally implicit discontinuous Galerkin method for Maxwell's equations*. ESAIM: M2AN, vol. 46, no. 5, pages 1225-1246, 2012.

Proceedings

- S. Descombes, S. Lanteri and L. Moya. *High-order locally implicit time integration strategies in a discontinuous Galerkin method for Maxwell's equations*. To appear in proceedings of the ICOSAHOM conference, June 25-29, 2012, Gammarth, Tunisia, Series: Lecture Notes in Computational Science and Engineering.

- L. Moya. *Locally implicit discontinuous Galerkin methods for time-domain Maxwell equations*. In Numerical Mathematics and Advanced Applications 2011: Proceedings of ENUMATH 2011, the 9th European Conference on Numerical Mathematics and Advanced Applications, Leicester, September 2011, pages 129-137, 2013.

Conferences

- NUMELEC 2012, 7ème Conférence Européenne sur les Méthodes Numériques en Electromagnétisme, Marseilles, France, 3-5 Juillet, 2012.

- ICOSAHOM 2012, International Conference on Spectral and High Order Methods, Gammarth, Tunisia, 25th-29th June, 2012.

- ENUMATH 2011 (European Numerical Mathematics and Advanced Applications), Leicester, United Kingdom, 5th-9th September 2011.

- SMAI 2011, 5ème Biennale Française des Mathématiques Appliquées et Industrielles, Guidel, France, 23-27 Mai, 2011.

Introduction

Computational Electromagnetics

The unification of electricity and magnetism was achieved by the physicist and mathematician James Clerk Maxwell in the 1860s. When Maxwell worked out his theory of electromagnetism, he ended up with a system of Partial Differential Equations (PDEs), now widely defined as Maxwell's equations, that relate electric and magnetic fields to each other and govern all classical electromagnetic wave phenomena. Nowadays, although the principles of electromagnetics are well understood, their application to practical configurations of current interest is significantly complicated and far beyond manual calculation in all but the simplest cases. The first electromagnetic analyses were carried out for simple shapes presenting a high degree of geometrical symmetry such as spheres, cylinders and planes; solutions to more complex geometries required the development of approximate methods. With the advent of computer technology, numerical methods were developed in the 1960s to allow more versatility and accuracy in the approximate solutions. Now, engineers and scientists use computers ranging from desktop machines to massively parallel arrays of processors to obtain solutions of Maxwell's equations.

The electromagnetic phenomena are so pervasive in modern sciences and technologies, that computational electromagnetics has an extremely wide range of applications. Solving Maxwell's equations accurately has been motivated in the past primarily by the requirements of military defense such as microwave radar technology or electromagnetic pulse vulnerability, and the numerical analysis in electromagnetics is shifting rapidly toward applications of societal and industrial relevance. An early application of numerical analysis in electromagnetics is antenna analysis and design. Antennas play a critical role in wireless communication, remote sensing, space exploration, defense, electronic warfare, and many other electronic systems. With the explosive growth of electronic devices in the daily life, a more recent application is the numerical analysis of electromagnetic interference and electromagnetic compatibility problems. Any electronic device emits electromagnetic fields that can potentially interfere with other devices, threatening their proper functioning. A component in an electronic device can also interfere undesirably with other components in the same device and affect its overall performance. In such cases, it is critical to determine the source of the interference and eliminate or reduce the interference level. Since the 1990s, there has been an explosive growth in the number of personal wireless communication devices available to the consumer such as cellphones whose radiations interact with the human head. A key requirement is then to meet the safety standard for microwave exposure to the user, namely, the peak Specific Absorption Rate (SAR). The propagation of electromagnetic waves through human tissues is also at the heart of many biomedical applications. Electromagnetic fields are involved in several biomedical technologies such as the microwave imaging of cancer tumours, the definition of microwave-based hyperthermia as an immunotherapy strategy for cancer, and the electroporation technique [Miklavčič *et al.* 2000, Šel *et al.* 2005, Serša 2005, Sukharev *et al.* 1992, Tsong 1991].

The numerical simulation has contributed to a better understanding of the physical phenomena associated with the propagation of electromagnetic waves, and the wide range of applications has motivated extensive researches to achieve well-suited numerical methods making the mathemati-

cal characteristics of Maxwell's equations. Numerical methods are fundamental for simulating and solving complex problems that do not admit analytical solutions, however, several difficulties arise when trying to model and develop numerical methods to simulate the propagation of electromagnetic waves. On one hand, most of electromagnetic waves propagation problems are unbounded problems that require an infinitely large solution domain. Truncating an infinitely large solution domain into a finite computational domain is then one of the major challenges for the numerical methods. On the other hand, the geometrical characteristics of the propagation medium (irregular shapes, geometrical singularities), the physical characteristics of the propagation medium (heterogeneity, physical dispersion and dissipation) and the characteristics of the sources (wires, etc.) are increasingly complex. Solving Maxwell's equations for such propagation media and sources needs the development of robust algorithms. Finally, the applications of industrial interest lead to the numerical solution of discrete systems whose size, in terms of the number of unknowns, is very large. A typical three-dimensional problem dedicated to industrial applications would require, at minimum, 10 million grid cells total. With a minimum of six fields to compute, three components each of the electric field and magnetic field, a high degree of parallelism is a key requirement for the algorithm generated by the numerical method.

Since Maxwell's equations can be solved in either the time- or the frequency-domain, two groups of methods have been developed to deal with electromagnetic problems, time- and frequency-domain methods, which are related to each other by the Fourier transform. The choice of a method is related to the intended application. For certain types of problems, a time-harmonic evolution can be assumed leading to the formulation of the frequency-domain Maxwell equations whose can be numerically solved using so-called frequency-domain methods. The numerical resolution requires the solution of a linear system of equations for each frequency, and the resulting numerical method is then naturally implicit. The frequency-domain methods are well suited for problems where many excitations are considered, such as in a monostatic scattering analysis. Furthermore, since the frequency-domain methods solve Maxwell's equations at each frequency, they can deal easily with dispersive media, where the constitutive parameters depend on the frequency. Indeed, the dispersive character of the propagation media is taken into account in the Maxwell equations, through the electrical permittivity, the magnetic permeability and the electric conductivity coefficients, which are in the general case tensors whose entries depend on space and frequency. There exist several mathematical models for the frequency evolution of these coefficients such as Debye model, Lorentz model, Drude model, etc. In this dissertation, we will focus on the time-domain Maxwell equations. The full system of unsteady Maxwell equations which is a first-order hyperbolic linear system of PDEs (if the underlying propagation media is assumed to be linear) can be numerically solved using so-called time-domain methods among which the Finite Difference Time Domain (FDTD) method introduced by K.S. Yee in [Yee 1966] is the most popular and often serves as a reference method [Inan & Marshall 2011, Taflove & Hagness 2005]. This popularity is mainly due to its simplicity and efficiency in discretising simple domain problems. However, its inability to effectively handle complex geometries has prompted to the development of alternatives methods, for instance, the Finite Element Time Domain (FETD) method, the Finite Volume Time Domain (FVTD) method or the Discontinuous Galerkin Time Domain (DGTD) method [Cockburn *et al.* 2000, Hesthaven & Warburton 2008, Jin & Riley 2008, Rao 1999]. In the vast majority of existing time-domain methods, time advancing relies on an explicit time integration scheme. The time-domain methods are well suited for broadband problems where one seeks for a

solution over a broad frequency band with only a few excitations, such as in a broadband antenna or device analysis. Furthermore, the time-domain methods solve Maxwell's equations step by step in time, therefore they can deal effectively with nonlinear problems, where the electromagnetic properties of the computational domain change with the field strengths. For dispersive propagation media, the time-domain numerical modeling of such materials requires specific techniques in order to switch from the frequency evolution of the electromagnetic coefficients to a time dependency.

Outline of the dissertation

This work deals with the time-domain formulation of Maxwell's equations. The first objective of the study is to propose arbitrary high-order finite element type methods on simplicial meshes for the discretization of Maxwell's equations and efficient time integration methods for dealing with grid induced stiffness when using non-uniform (locally refined) meshes. We will consider DGTD methods relying on an arbitrary high-order polynomial interpolation of the component of the electromagnetic field, and their computer implementations will make use of nodal (Lagrange) basis expansions on simplicial elements. The discontinuous Galerkin technique combines important features from both the finite element method and the finite volume method, producing a scheme that is strictly local, and which can achieve a high-order of accuracy. This choice is then motivated by the fact that DGTD methods are well suited to the above-mentioned issues which arise when trying to simulate the propagation of electromagnetic waves. Existing DGTD methods for the solution of the time-domain Maxwell equations often rely on explicit time integration schemes and are therefore constrained by a stability condition that can be very restrictive on highly refined meshes. To overcome this limitation, we will consider time integration methods that consist in applying an implicit time integration scheme locally i.e in the refined regions of the mesh, while preserving an explicit time scheme in the complementary part, resulting in locally implicit (or hybrid implicit-explicit) time integration schemes. Hybrid implicit-explicit approaches are classical and well known in computational fluid dynamics, where the non-stiff part of the model equations is treated explicitly, for instance an advection term, and the stiff part implicitly, for instance a diffusion term or a term modeling stiff reactions [Ascher *et al.* 1995, W. Hundsdorfer 2003, Hundsdorfer & Verwer 2003, Wesseling 2001]. In computational electrodynamics such approaches remain underexploited compared to explicit local time-stepping techniques, which consist to use smaller time steps, given by a local stability criterion, precisely where the smallest elements are located to overcome the bottleneck caused by local mesh refinement in explicit time integrators [Diaz & Grote 2009, Grote & Mitkova 2010, Monseny *et al.* 2008, Piperno 2006, Taube *et al.* 2009]. In contrast to explicit local time-stepping methods, the locally implicit methods provide a single time step for stability, independent of the fine grid. In this dissertation, we will present a full theoretical study (formulation, stability, convergence, numerical analysis of the implicit treatment) and a comparison of two locally implicit DGTD methods for the first-order formulation of Maxwell's equations and various numerical experiments for two- and three-dimensional electromagnetic wave propagation problems, illustrating the theoretical results and highlighting the attractive features of these approaches. The first locally implicit method, that we will consider, has been introduced by Piperno in [Piperno 2006]. This time integration scheme is a blend of the second-order explicit leap-frog scheme and the second-order implicit Crank-Nicolson scheme, based on a nodal DG method for the spatial discretization. A stability anal-

ysis and three-dimensional numerical experiments for this locally implicit DGTD method have been presented in [Dolean *et al.* 2010], however no proof of convergence is given. In this dissertation, a numerical convergence analysis of the fully discrete scheme is achieved. Particularly, based on this theoretical proof and some numerical experiments, we will see that the implicit-explicit component splitting can reduce by one the order of convergence of the method. Such possible reduction of the order of convergence has motivated the study of a second locally implicit time integration scheme, initially proposed by Verwer in [Verwer 2010]. This time integration method is also a blend of the leap-frog scheme and the Crank-Nicolson scheme, which is directly applied on a general form of the semi-discrete Maxwell equations. In [Verwer 2010], the author prove that no reduction of the order of convergence occurs with this implicit-explicit component splitting, and give numerical results for one- and two-dimensional problems, based on finite difference spatial discretizations. In this dissertation, we study this time integration strategy with a DG spatial discretization. An additional stability analysis is carried out and the the great similarities with the locally implicit DGTD method of [Dolean *et al.* 2010] are highlighted. Furthermore, numerical investigations for two- and three-dimensional electromagnetic wave propagation problems are achieved, allowing, on one hand, to assess the practical virtue of this component splitting scheme for a DG discretization on non-uniform meshes and, on the other hand, to compare the resulting method with the locally implicit DGTD method of [Dolean *et al.* 2010] and the fully explicit and fully implicit methods.

The discontinuous Galerkin method allows to easily increase the spatial convergence order; another objective of this work is therefore to assess the possibility of designing higher order locally implicit time integration schemes. In this dissertation, higher order time integration techniques based on the previous second-order locally implicit DGTD method, such as symmetric composition methods or Richardson extrapolation, are investigated. Such strategies to increase the order of convergence are well-known for numerical Ordinary Differential Equation (ODE) problems [Hairer *et al.* 1993, Hairer *et al.* 2010] but not widespread for numerical PDE problems. From a basic method, these strategies are essentially a suitable combination of this method applied with different step sizes and therefore their computer implementations are easy.

The numerical treatment of complex propagation media models (i.e. physical dispersion models) is also one of the objectives of the present study. The intended application is the interaction of electromagnetic waves with biological tissues. As previously mentioned, the study of the interaction between electromagnetic waves and living tissues is of interest to several applications of societal relevance such as the assessment of potential adverse effects of electromagnetic fields or the utilization of electromagnetic waves for therapeutic or diagnostic purposes. Numerical modeling of electromagnetic wave propagation in interaction with biological tissues requires to solve the system of Maxwell's equations coupled to appropriate models of physical dispersion in the tissues, such the Debye model. This model is the most often used for electromagnetic wave interactions with water-based substances, such as biological tissues. In this dissertation we will derive a locally implicit DGTD method for this dispersion model. In this context, to our knowledge, locally implicit DGTD methods were never considered, although the characteristics of these methods are well suited to the miniaturization of electronic devices and antennas or the small size of cancerous cells, for which the use of non-uniform (locally refined) meshes are certainly key for the efficient numerical solution of wave propagation problems in biological tissues.

The plan of the dissertation is the following. The first chapter provides a brief review of Maxwell's equations and their underlying foundations, and introduces the mathematical model that

we will consider in the remaining. Chapter 2 is dedicated to the discontinuous Galerkin method. After the introduction of the basic principles and their foundations, a full treatment of the DG method for the first-order formulation of the time-domain Maxwell equations is presented. Starting from the semi-discrete scheme, obtained in Chapter 2, Chapter 3 deals with two locally implicit time integration methods. A full numerical analysis of both schemes is achieved. Chapter 4 is dedicated to numerical experiments for two-dimensional and three-dimensional electromagnetic waves propagation problems. Particular care is taken to confirm the theoretical results obtained in the previous chapter, and a numerical comparison between the different time integration strategies is carried out. The DGTD methods presented in Chapter 3 are second-order temporal methods. Chapter 5 introduces an outlook to higher order time integration schemes. As previously mentioned, the discontinuous Galerkin method allows to easily increase the spatial convergence order, the objective is therefore to assess the practical virtue of locally implicit approach for higher order time integration schemes. Finally, Chapter 6 is dedicated to the propagation of electromagnetic wave in dispersive media, particularly in biological tissues. In this last chapter, we introduce the formulation of Maxwell equations for Debye dispersive media. The dispersive character is taken into account via an auxiliary differential equation and the numerical treatment of the underlying mathematical model is derived from one of the locally implicit discontinuous Galerkin methods presented in Chapter 3. Furthermore, stability and convergence analyses of the resulting scheme are achieved, and three-dimensional problems are presented.

List of publications and conferences

Articles

- S. Descombes C. Durochat. S. Lanteri, L. Moya, C. Scheid and J. Viquerat. *Recent advances on a DGTD method for time-domain electromagnetics*. To appear in Photonics and Nanostructures - Fundamentals and Applications (2013). http://dx.doi.org/10.1016/j.photonics.2013.06.005

- S. Descombes, S. Lanteri and L. Moya. *Locally implicit time integration strategies in a discontinuous Galerkin method for Maxwell's equations*. J. SCI. Comput., vol. 56, no. 1, pages 190-218, 2013.

- L. Moya. *Temporal convergence of a locally implicit discontinuous Galerkin method for Maxwell's equations*. ESAIM: M2AN, vol. 46, no. 5, pages 1225-1246, 2012.

Proceedings

- S. Descombes, S. Lanteri and L. Moya. *High-order locally implicit time integration strategies in a discontinuous Galerkin method for Maxwell's equations*. To appear in proceedings of the ICOSAHOM conference, June 25-29, 2012, Gammarth, Tunisia, Series: Lecture Notes in Computational Science and Engineering.

- L. Moya. *Locally implicit discontinuous Galerkin methods for time-domain Maxwell equations*. In Numerical Mathematics and Advanced Applications 2011: Proceedings of ENU-

MATH 2011, the 9th European Conference on Numerical Mathematics and Advanced Applications, Leicester, September 2011, pages 129-137, 2013.

Conferences

- NUMELEC 2012, 7ème Conférence Européenne sur les Méthodes Numériques en Electromagnétisme, Marseilles, France, 3-5 Juillet, 2012.
- ICOSAHOM 2012, International Conference on Spectral and High Order Methods, Gammarth, Tunisia, 25th-29th June, 2012.
- ENUMATH 2011 (European Numerical Mathematics and Advanced Applications), Leicester, United Kingdom, 5th-9th September 2011.
- SMAI 2011, 5ème Biennale Française des Mathématiques Appliquées et Industrielles, Guidel, France, 23-27 Mai, 2011.

CHAPTER 1

Maxwell's equations

Contents

The unification of electricity and magnetism was achieved by the Scottish physicist and mathematician James Clerk Maxwell in the 1860s. Maxwell made use of a powerful idea called a field, which had been invented by the British physicist Michael Faraday in the 1840s to explain how a force could be carried from one body to another across free space. Faraday established laws to describe how the electric field and magnetic field are influenced by neighboring charges and magnetic poles. In collaboration with colleagues, he has tested and found these laws which gave predictions that agreed with experiment. Among the discoveries of the time there were phenomena involving electric and magnetic effects. Maxwell realized that these discoveries pointed to the unification of electricity and magnetism; he had to modify the equations, only adding one new term, new equations allowed electric and magnetic fields to turn into each other. These transformations produce rotating configurations that travels through space as waves, in which there is an electric field, leaving a magnetic field. From his theory, Maxwell could calculate the velocity of these waves and found the same speed as light. New predictions immediately followed. Maxwell realized that there had electromagnetic waves at all frequencies, not just those that correspond to visible light. This has contributed to the discovery of the radio, infrared and ultraviolet light, etc. When Maxwell worked out his theory of electromagnetism, he ended up with not four but twenty equations that describe the behavior of electric and magnetic fields [Maxwell 1865]. It was Oliver Heaviside in Great Britain and Heinrich Hertz in Germany who combined and simplified Maxwell's equations into four equations in the two decades after Maxwell's death. Today we call these four equations Gauss's law for electric fields, Gauss's law for magnetic fields, Faraday's law, and the Ampere - Maxwell law. Since these four laws are now widely defined as Maxwell's equations. Whereas most of classical physics was fundamentally revised with Einstein's theory of special relativity, Maxwell's equations remained intact.

In Section 1.1 we present a brief review of the set of four coupled partial differential equations, that relates the electric and magnetic fields and defines Maxwell's equations. In Section 1.2 we state the initial and boundary value problem that we will consider in the remaining of the dissertation.

1.1 Physical statement

1.1.1 Gauss's law for electric fields

1.1.1.1 The integral form of Gauss's law for electric fields

The integral form of Gauss's law for electric fields relates the electric flux over a surface to the charge enclosed by that surface. The main idea is that an electric charge produces an electric field, and the flux of that field passing through any closed surface is proportional to the total charge contained within that surface. The integral form of Gauss's law can be written as

$$\oint_S \mathbf{E}\cdot\mathbf{n}\, ds = \frac{q_{enc}}{\varepsilon_0}, \tag{1.1}$$

where $\mathbf{E}$ denotes the electric field, S is a surface, $\mathbf{n}$ is the unit normal to the surface S, q_{enc} is the amount of enclosed charge (in coulombs) and ε_0 is the electric permittivity of free space (in coulombs per volt-meter). The permittivity of a material determines its response to an applied electric field, in nonconducting materials, charges do not move freely, but may be slightly displaced from their equilibrium positions. The relevant permittivity in Gauss's law for electric fields is the permittivity of free space ($\varepsilon_0 \simeq 8.85 \times 10^{-12}$ coulombs per volt-meter). The left hand side of (1.1) is a mathematical description of the electric flux, i.e. the number of electric field lines passing through a closed surface S; the right hand side is the total amount of charge contained within that surface divided by the electric constant.

1.1.1.2 The differential form of Gauss's law for electric fields

The transition from the integral form to the differential form of Gauss's law for electric fields is directly derived from the divergence theorem which may be state has follow

Theorem 1.1.1. *(The divergence theorem) The flux of a vector field through a closed surface S is equal to the integral of the divergence of that field over a volume V for which S is a boundary.*

We apply the divergence theorem to the left hand side of the integral form of Gauss's law (1.1)

$$\oint_S \mathbf{E}\cdot\mathbf{n}\, ds = \int_V \operatorname{div}(\mathbf{E})\, dV. \tag{1.2}$$

Denote ρ the charge density in coulombs per cubic meter. The enclosed charge q_{enc} is the volume integral of the charge density ρ, then from (1.2) and (1.1) we write

$$\int_V \operatorname{div}(\mathbf{E})\, dV = \int_V \frac{\rho}{\varepsilon_0} dV, \tag{1.3}$$

Since this equality must hold for all volumes, the integrands must be equal. Thus

$$\operatorname{div}(\mathbf{E}) = \frac{\rho}{\varepsilon_0}, \tag{1.4}$$

which is the differential form of Gauss's law for electric fields. The differential form of Gauss's law (1.4) is a mathematical description of the divergence of the electric field, i.e. the tendency of the field to flow away from a specified location. There is a fundamental difference between the differential and the integral form of Gausss law; the differential form deals with the divergence of the electric field and the charge density at individual points in space, whereas the integral form entails the integral of the normal component of the electric field over a surface. The only places at which the divergence of the electric field is not zero are those locations at which charge is present. If positive charge is present, the divergence is positive, meaning that the electric field diverge from positive charge. If negative charge is present, the divergence is negative, and the electric field converge negative charge.

1.1.2 Gauss's law for magnetic fields

1.1.2.1 The integral form of Gauss's law for magnetic fields

The main idea of the integral form of Gauss's law for magnetic fields is that the total magnetic flux passing through any closed surface is zero. The integral form of Gauss's law can be written as

$$\oint_S \mathbf{B}\cdot\mathbf{n}\, ds = 0, \tag{1.5}$$

where $\mathbf{B}$ denotes the magnetic field. The left hand side of (1.5) is a mathematical description of the magnetic flux passing through a closed surface S, while the right hand side is identically zero. This does not mean that zero magnetic field line penetrates the surface; it means that the inward (negative) magnetic flux, i.e. the magnetic field lines that enters the volume enclosed by the surface, must be exactly balanced by the outward (positive) magnetic flux.

1.1.2.2 The differential form of Gauss's law for magnetic fields

Similarly to Gauss's law for electric fields the theorem of divergence applied to the integral form (1.5) leads to the differential form of Gauss's law for magnetic fields

$$\operatorname{div}(\mathbf{B}) = 0. \tag{1.6}$$

Thus the divergence of the magnetic field at any point is zero, the amount of incoming field is exactly the same as the amount of outgoing field at every point.

1.1.3 Faraday's law

In a series of experiments in the 1830s, Michael Faraday demonstrated that an electric current may be induced in a circuit by changing the magnetic flux enclosed by the circuit.

1.1.3.1 The integral form of Faraday's law

The main idea of the integral form of Faraday's law is that a changing magnetic flux through a surface induces an electromotive force (emf) in any closed path of that surface, and a changing magnetic field induces a circulating electric field. It can be written as

$$\oint_C \mathbf{E}\cdot\mathbf{dl} = -\frac{d}{dt}\int_S \mathbf{B}\cdot\mathbf{n}\, ds, \tag{1.7}$$

where C denotes a closed path of any surface S, **dl** is an incremental segment of C. Note that **E** in Faraday's law (1.7) represents the induced electric field at each segment **dl** of the path C measured in the reference frame in which that segment is stationary. The left hand side of (1.7) represents the electric field circulation while the right hand side is the rate of change of magnetic flux.

1.1.3.2 The differential form of Faraday's law

The transition from the integral form to the differential form of Faraday's law can be derived from Stokes' theorem which may be state has follow

Theorem 1.1.2. *(Stokes' theorem) The circulation of a vector field over a closed path C is equal to the integral of the normal component of the curl of that field over a surface S for which C is a boundary.*

Applying Stokes' theorem to the left hand side of (1.7) we obtain

$$\int_S \text{curl}\,(\mathbf{E})\cdot\mathbf{n}\,ds = -\frac{d}{dt}\int_S \mathbf{B}\cdot\mathbf{n}\,ds. \tag{1.8}$$

For stationary geometries, the time derivative can be moved inside the integral

$$\int_S \text{curl}\,(\mathbf{E})\cdot\mathbf{n}\,ds = \int_S \left(-\frac{\partial \mathbf{B}}{\partial t}\cdot\mathbf{n}\right) ds. \tag{1.9}$$

Since this equality must hold for all surfaces, the integrands must be equal. Thus

$$\text{curl}\,(\mathbf{E}) = -\frac{\partial \mathbf{B}}{\partial t}, \tag{1.10}$$

which is the differential form of Faraday's law. Consequently a circulating electric field is produced by a magnetic field that changes with time.

1.1.4 The Ampere - Maxwell law

Ampere's law relating a steady electric current to a circulating magnetic field was well known when James Clerk Maxwell began his work on the unification of electricity and magnetism in the 1850s. Ampere's law was known to apply only to static situations involving steady currents. It was Maxwell's addition of another source term, a changing electric flux, that extended the applicability of Ampere's law to time-dependent conditions. The presence of this term in the equation now called the Ampere - Maxwell law ensured consistency of the unified theory, and allowed Maxwell to discern the electromagnetic nature of light and to develop a comprehensive theory of electromagnetism.

1.1.4.1 The integral form of the Ampere - Maxwell law

The main idea of the integral form of Ampere - Maxwell law is that an electric current or a changing electric flux through a surface produces a circulating magnetic field around any path that bounds that surface. It can be written as

$$\oint_C \mathbf{B}\cdot\mathbf{dl} = \mu_0\left(I_{enc} + \varepsilon_0\frac{d}{dt}\int_S \mathbf{E}\cdot\mathbf{n}\,ds\right), \tag{1.11}$$

where I_{enc} denotes the enclosed electric current (in amperes) and μ_0 is the magnetic permeability of free space (in volt-seconds per ampere-meter). Just as the electric permittivity characterizes the response of a dielectric to an applied electric field, the magnetic permeability determines a material's response to an applied magnetic field. The permeability in the Ampere - Maxwell law is that of free space ($\mu_0 = 4\pi \times 10^{-7}$ volt-seconds per ampere-meter). The left hand side of (1.11) represents the magnetic field circulation while the right hand side represents two sources for the magnetic field, a steady conduction current and a changing electric flux through any surface S bounded by path C.

1.1.4.2 The differential form of the Ampere - Maxwell law

Similarly to Faraday's law Stokes' theorem applied to the integral form (1.11) leads to the differential form of Ampere - Maxwell law

$$\operatorname{curl}(\mathbf{B}) = \mu_0 \left(\mathbf{J} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right), \tag{1.12}$$

where $\mathbf{J}$ is the electric current density

$$\int_S \mathbf{J} \cdot \mathbf{n} \, ds = I_{enc}. \tag{1.13}$$

The differential form of Ampere - Maxwell law (1.12) is a mathematical description of the circulation of a magnetic field produced by an electric current and by an electric field that changes with time.

1.1.5 Summary

Here is a summary of the differential forms of Maxwell's equations

$$\begin{cases} \operatorname{div}(\mathbf{E}) &= \dfrac{\rho}{\varepsilon_0} & \text{(Gauss's law for electric fields)}, \\ \operatorname{div}(\mathbf{B}) &= 0 & \text{(Gauss's law for magnetic field)}, \\ \operatorname{curl}(\mathbf{E}) &= -\dfrac{\partial \mathbf{B}}{\partial t} & \text{(Faraday's law)}, \\ \operatorname{curl}(\mathbf{B}) &= \mu_0 \left(\mathbf{J} + \varepsilon_0 \dfrac{\partial \mathbf{E}}{\partial t} \right) & \text{(the Ampere - Maxwell law)}. \end{cases} \tag{1.14}$$

These four partial differential equations are general, apply to electric and magnetic fields in matter as well as in free space. The system (1.14) constitute a complete description of the behavior of electric and magnetic fields. The first equation describes how electric fields are induced by electric charges. The second equation says that there is no such thing as a magnetic monopole. The third equation describes the induction of electric fields by changing magnetic fields, and the fourth equation describes the generation of magnetic fields by electric currents, and the induction of magnetic fields by changing electric fields. Note that, with the inclusion of the displacement current, these equations treat electric and magnetic fields on an equal footing, i.e., electric fields can induce magnetic fields, and vice versa. Maxwell's equations succinctly sum up the experimental results of Coulomb, Ampere, and Faraday. They are called Maxwell's equations because James Clerk Maxwell was the first to write them down (in component form). Maxwell also modified them so as to make them mathematically self-consistent.

Inside matter note that the charge density ρ in Gauss's law includes all charge bound as well as free. Similarly the current density $\mathbf{J}$ in the Ampere - Maxwell law includes all currents bound, polarization as well as free. Bound charge and current may be difficult to determine then versions of Maxwell's equations that depend only on the free charge and the free current are very useful.

1.1.6 Maxwell's equations in matter

1.1.6.1 Gauss's law for electric fields in matter

Within a dielectric material positive and negative charges may become slightly displaced when an electric field is applied. We introduce the electric polarization of the material, denoted by $\mathbf{P}$, which describes the dipole moment per unit volume induced by electric fields. If the polarization is uniform, bound charge appears only on the surface of the material. If the polarization varies from point to point within the dielectric, there are accumulations of charge within the material and the volume charge density is given by

$$\rho^{bound} = -\mathrm{div}\,(\mathbf{P})\,, \tag{1.15}$$

where ρ^{bound} is the volume density of bound charge i.e. charge displaced by the electric field but does not move freely through the material. Within matter the total charge density includes bound and free charges

$$\rho = \rho^{free} + \rho^{bound}. \tag{1.16}$$

Then using (1.16) and (1.15) in (1.4) we write Gauss's law for electric fields as

$$\mathrm{div}\,(\varepsilon_0\mathbf{E} + \mathbf{P}) = \rho^{free}. \tag{1.17}$$

Now we define the displacement (or the electric flux density) $\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}$, substituting in (1.17), Gauss's law in terms of $\mathbf{D}$ and the free charge density is

$$\mathrm{div}\,(\mathbf{D}) = \rho^{free}. \tag{1.18}$$

1.1.6.2 The Ampere - Maxwell law in matter

Within dielectric material applied magnetic fields induce magnetic dipole moment per unit volume, denoted by $\mathbf{M}$, within magnetic material. As bound electric charges act as the source of additional electric fields within the material, bound current $\mathbf{J}^{bound}$ may act as the source of additional magnetic fields

$$\mathbf{J}^{bound} = \mathrm{curl}\,(\mathbf{M})\,. \tag{1.19}$$

Moreover the time rate of change of the polarization may also become a contribution to the current density. We define the polarization current density by

$$\mathbf{J}^{pol} = \frac{\partial \mathbf{P}}{\partial t}. \tag{1.20}$$

The current density $\mathbf{J}$ in the Ampere - Maxwell (1.12) law includes all currents bound, polarization as well as free

$$\mathbf{J} = \mathbf{J}^{bound} + \mathbf{J}^{pol} + \mathbf{J}^{free}. \tag{1.21}$$

Inserting the expressions of the bounds and polarization current and using (1.21) and the definition of the displacement **D** into (1.12) we write the Ampere - Maxwell as

$$\operatorname{curl}\left(\frac{\mathbf{B}}{\mu_0}-\mathbf{M}\right)=\mathbf{J}^{free}+\frac{\partial \mathbf{D}}{\partial t}. \tag{1.22}$$

Now we define the magnetic field intensity $\mathbf{H}=\mathbf{B}/\mu_0-\mathbf{M}$, substituting into (1.22), the Ampere - Maxwell law in terms of **H**, **D** and the free current density is then

$$\operatorname{curl}(\mathbf{H})=\mathbf{J}^{free}+\frac{\partial \mathbf{D}}{\partial t}. \tag{1.23}$$

Thus, Maxwell's equations in matter can be written as (differential form)

$$\begin{cases} \operatorname{div}(\mathbf{D}) &= \rho^{free} & \text{(Gauss's law for electric fields)},\\ \operatorname{div}(\mathbf{B}) &= 0 & \text{(Gauss's law for magnetic field)},\\ \operatorname{curl}(\mathbf{E}) &= -\dfrac{\partial \mathbf{B}}{\partial t} & \text{(Faraday's law)},\\ \operatorname{curl}(\mathbf{H}) &= \mathbf{J}^{free}+\dfrac{\partial \mathbf{D}}{\partial t} & \text{(the Ampere - Maxwell law)}. \end{cases} \tag{1.24}$$

1.2 Problem statement

This dissertation deals with Maxwell's equation in non-dispersive and dispersive propagation media. First we will focus our analysis on the numerical modeling of the propagation of electromagnetic waves in non-dispersive media, while Maxwell's equations in dispersive media and the corresponding mathematical model will be presented in Chapter 6.

We assume that media are isotropic, linear, time-invariant and non-dispersive. This means that

- the response of the medium not depends on the orientation of the fields, the polarization **P** and the magnetization **M** are co-linear and proportional to the electric field **E** and the magnetic field intensity **H**, respectively (isotropic);
- the constitutive parameters do not depend on the magnitude of the applied fields (linear);
- the constitutive parameters do not vary with time (time invariant);
- the constitutive parameters do not depend on the frequency (non-dispersive).

In a large class of dielectric and magnetic materials, there exits an approximately linear relationship between **P** and **E** and **M** and **H**. If the material is isotropic then

$$\begin{cases} \mathbf{P}=\varepsilon_0\chi_e\mathbf{E},\\ \mathbf{M}=\mu_0\chi_m\mathbf{H}, \end{cases} \tag{1.25}$$

where χ_e and χ_h are called the electric and magnetic susceptibility, respectively.

Remark 1.2.1. *Equations (1.25) are strictly valid only for non-dispersive media. Effectively, for example, because of the dependence of the electric permittivity with frequency we generally have* $\mathbf{P}(\omega) = \varepsilon_0 \chi_e(\omega)\mathbf{E}(\omega)$. *Thus, according to the convolution theorem, for arbitrary time dependence the relationship between the electric flux density and the electric field becomes*

$$\mathbf{P}(\mathbf{x},t) = \varepsilon_0 \int_0^t \mathbf{E}(\mathbf{x},t-s)\,\chi_e(\mathbf{x},s)\; ds, \tag{1.26}$$

These expressions indicate that the response of the medium to an applied field is not instantaneous. As previously mentioned we will state Maxwell's equations in dispersive media in Chapter 6.

It follows from (1.25) that the constitutive relations relate the electric field intensity $\mathbf{E}$ to the electric flux density $\mathbf{D}$, and similarly the magnetic field intensity $\mathbf{B}$ to the magnetic flux density $\mathbf{H}$, are given by

$$\begin{cases} \mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E}, \\ \mathbf{B} = \mu_0 \mu_r \mathbf{H}, \end{cases} \tag{1.27}$$

where $\varepsilon_r = 1 + \chi_e$ and $\mu_r = 1 + \chi_m$ are the relative permittivity and the relative magnetic permeability of the material, respectively. Note that ε_r and μ_r are dimensionless. Now we divide the system of Maxwell's equations (1.24) into two systems

$$\begin{cases} \dfrac{\partial \mathbf{D}}{\partial t} &= \operatorname{curl}(\mathbf{H}) - \mathbf{J}^{free}, \\ \dfrac{\partial \mathbf{B}}{\partial t} &= -\operatorname{curl}(\mathbf{E}), \end{cases} \tag{1.28}$$

and

$$\begin{cases} \operatorname{div}(\mathbf{D}) &= \rho^{free}, \\ \operatorname{div}(\mathbf{B}) &= 0. \end{cases} \tag{1.29}$$

With the constitutive relation (1.27), equations (1.29) are just the consistency conditions for the system (1.28). Indeed take the divergence of (1.28), apply (1.29) and (1.27), the resultant equation represents nothing else but charge conservation law, i.e.

$$\frac{\partial \rho^{free}}{\partial t} + \operatorname{div}\left(\mathbf{J}^{free}\right) = 0. \tag{1.30}$$

Thus, as long as initial conditions satisfy (1.29) and the electromagnetic field evolves according to (1.28), the solution at any time will also satisfy (1.29). Consequently we can only consider the equations (1.28) in which the constitutive relations (1.27) are included i.e.

$$\begin{cases} \varepsilon_0 \varepsilon_r \dfrac{\partial \mathbf{E}}{\partial t} &= \operatorname{curl}(\mathbf{H}) - \mathbf{J}^{free}, \\ \mu_0 \mu_r \dfrac{\partial \mathbf{H}}{\partial t} &= -\operatorname{curl}(\mathbf{E}). \end{cases} \tag{1.31}$$

From now we will consider a normalized form of Maxwell's equations (1.31). We introduce the normalized space, time variables and physical fields through the relations

$$\widetilde{\mathbf{x}} = \mathbf{x}, \quad \widetilde{t} = c_0\, t, \quad \widetilde{\mathbf{E}} = \mathbf{E}, \quad \widetilde{\mathbf{H}} = Z_0\, \mathbf{H} \quad \text{and} \quad \widetilde{\mathbf{J}}^{free} = Z_0\, \mathbf{J}^{free}, \tag{1.32}$$

where $c_0 = 1/\sqrt{\varepsilon_0\mu_0}$ is the speed of light in vacuum ($c_0 \simeq 3\times 10^8 \mathrm{m\cdot s^{-1}}$) and $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ is the free space intrinsic impedance (Ohm, Ω =$\mathrm{V\cdot A^{-1}}$). The units of physical and normalized variables and fields are summarized in Table 1.1.

$\mathbf{x} \longrightarrow \widetilde{\mathbf{x}}$	$t \longrightarrow \widetilde{t}$	$\mathbf{E} \longrightarrow \widetilde{\mathbf{E}}$	$\mathbf{H} \longrightarrow \widetilde{\mathbf{H}}$	$\mathbf{J}^{free} \longrightarrow \widetilde{\mathbf{J}}^{free}$
$\mathrm{m} \longrightarrow \mathrm{m}$	$\mathrm{s} \longrightarrow \mathrm{m}$	$\mathrm{V\cdot m^{-1}} \longrightarrow \mathrm{V\cdot m^{-1}}$	$\mathrm{A\cdot m^{-1}} \longrightarrow \mathrm{V\cdot m^{-1}}$	$\mathrm{A\cdot m^{-2}} \longrightarrow \mathrm{V\cdot m^{-2}}$

Table 1.1: Units of physical and normalized variables and fields.

Substituting the normalized space, time variables and fields (1.32) into (1.31) we can write Maxwell's equations as form

$$\left\{\begin{array}{rcl} \varepsilon_r \dfrac{\partial \widetilde{\mathbf{E}}}{\partial \widetilde{t}} & = & \mathrm{curl}\left(\widetilde{\mathbf{H}}\right) - \widetilde{\mathbf{J}}^{free}, \\ \mu_r \dfrac{\partial \widetilde{\mathbf{H}}}{\partial \widetilde{t}} & = & -\mathrm{curl}\left(\widetilde{\mathbf{E}}\right), \end{array}\right. \tag{1.33}$$

For convenience of presentation we omit in the sequel the "$\sim$" notation and we denote the relative dielectric permittivity and magnetic permeability, ε_r and μ_r, by ε and μ into (1.33).

Finally note that the free current density $\mathbf{J}^{free}$ includes the conduction current density and the source current density denote $\mathbf{J}^c$ and $\mathbf{J}^s$, respectively. The relation between an electric field and the conduction current density that is generated at any point of the conducting material is given by Ohm's law

$$\mathbf{J}^c = \sigma\mathbf{E}, \tag{1.34}$$

where σ is the conductivity of the medium. The source current density $\mathbf{J}^s$ may be maintained by external sources or generators and is often called driven or impressed current. Thus the free current density can be written

$$\mathbf{J}^{free} = \mathbf{J}^c + \mathbf{J}^s = \sigma\mathbf{E} + \mathbf{J}^s. \tag{1.35}$$

Hence, the first equation of (1.33) reads

$$\varepsilon\frac{\partial \mathbf{E}}{\partial t} = \mathrm{curl}\,(\mathbf{H}) - \sigma\mathbf{E} - \mathbf{J}^s. \tag{1.36}$$

1.2.1 Summary

Here is a summary of the continuous problem considered in the following (except in Chapter 6 where we will state the initial and boundary value problem used to model the propagation of an electromagnetic wave in dispersive media). Let $\Omega \subset \mathbb{R}^3$ be a bounded, convex, polyhedral domain. We denote by $\mathbf{n}$ the normal outward to the domain boundary $\partial\Omega$ and we define the functional space

$$\mathscr{H} \equiv \left[H\left(\mathrm{curl},\Omega\right)\right]^3 \times \left[H\left(\mathrm{curl},\Omega\right)\right]^3, \tag{1.37}$$

where $H\left(\mathrm{curl},\Omega\right)$ is the classical subspace of $L^2\left(\Omega\right)$ fields with curl in $L^2\left(\Omega\right)$. We assume that media are isotropic, linear and time-invariant. Let $T > 0$, we state the time-domain Maxwell equations in

$[0,T]$ as follow

$$\begin{cases} \varepsilon \dfrac{\partial \mathbf{E}}{\partial t} &= \operatorname{curl}(\mathbf{H}) - \sigma \mathbf{E} - \mathbf{J}^s, \\ \mu \dfrac{\partial \mathbf{H}}{\partial t} &= -\operatorname{curl}(\mathbf{E}), \end{cases} \tag{1.38}$$

where $\mathbf{E}$ and $\mathbf{H}$ denote the electric and magnetic fields, ε, μ and σ are coefficients representing (relative) dielectric permittivity, magnetic permeability and conductivity, and $\mathbf{J}^s$ represents the source current density. Everywhere on $\partial\Omega$, a boundary condition is set which is either metallic, $\mathbf{n}\times\mathbf{E}=0$ or absorbing, $\mathbf{n}\times\mathbf{E}=\sqrt{\mu/\varepsilon}\,(\mathbf{n}\times\mathbf{H})\times\mathbf{n}$ (first-order Silver-Müller condition). Finally, initial conditions are given by $(\mathbf{E}_0,\mathbf{H}_0)\in\mathscr{H}(\Omega)$ such that $\mathbf{E}(\cdot,0)=\mathbf{E}_0$, $\mathbf{H}(\cdot,0)=\mathbf{H}_0$ and $(\mathbf{E}_0,\mathbf{H}_0)$ satisfy (1.29).

The following existence and uniqueness theorem has been proved in [18] for a regular domain, and can be generalized for Lipschitz-polyhedra [Fezoui *et al.* 2005].

Theorem 1.2.1. *For zero source current $\mathbf{J}^s$, the problem (1.38) admits a unique solution $(\mathbf{E},\mathbf{H})\in C^1\big([0,T],\ [L^2(\Omega)]^3\times[L^2(\Omega)]^3\big)\cap C^0\big([0,T],\ \mathscr{H}\big)$, for any initial data $(\mathbf{E}_0,\mathbf{H}_0)\in\mathscr{H}(\Omega)$ satisfying $\mathbf{n}\times\mathbf{E}=0$ on metallic boundary and $\mathbf{n}\times\mathbf{E}=\sqrt{\mu/\varepsilon}\,(\mathbf{n}\times\mathbf{H})\times\mathbf{n}$ on absorbing boundary.*

CHAPTER 2

The Discontinuous Galerkin method

Contents

Nowadays, many different types of methods exist for the numerical resolution of Maxwell's equations modeling electromagnetic wave propagation. The most prominent method among physicists and engineers is still the finite difference (FD) method based on Yee's scheme [Yee 1966]. This popularity is mainly due to its simplicity and efficiency in discretising simple domain problems. However, its inability to effectively handle complex geometries has prompted to search for alternatives methods. Also one of the main features of numerical methods based on finite element meshes like finite element (FE), finite volume (FV) or discontinuous Galerkin (DG) methods is the possibility of using locally refined and non-conformal space grids to easily deal with complex geometries. In the last decade there has been an increasing interest in the DG method, as illustrated in Figure (2.1). The latter is particularly well suited to the design of *hp*-adaptive strategies (i.e. where the characteristic mesh size h and the interpolation degree p change locally wherever it is needed). Thus the DG method can achieve a high-order of accuracy and is used in many applications [Cockburn *et al.* 2000, Hesthaven & Warburton 2008].

In Section 2.1 we introduce and discuss the Galerkin method, but also the FV method and the FE method. Our intention here is to introduce their basic principles, in order to show their relationship to the DG method, which uses features of both FV and FE methods. In Section 2.2 we present a full treatment of the DG method for the first-order formulation of the time-domain Maxwell equations (1.38), that we consider in the dissertation. Finally in Section 2.3 we discuss some algorithmic details related to the implementation of our DG method.

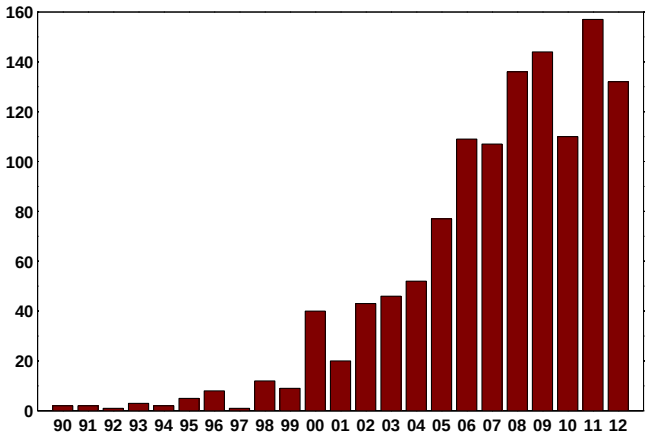


Figure 2.1: Yearly number of publication titles with the keyword *discontinuous Galerkin* in the MathSciNet database, since 1990.

2.1 Basic principles and characteristics of the DG method

2.1.1 A brief overview of the Galerkin method

The Galerkin method has been a precursor of the finite element method. Even though it does not have any numerical interest in general, it is very useful from a theoretical point of view. It appears in the framework of the internal variational approximation. The principle of the variational approach for the solution of PDEs is to replace the equation by an equivalent so-called variational formulation obtained by integrating the equation multiplied by an arbitrary function, called a test function.

Let V a Hilbert space, $b : V \times V \to \mathbb{R}$ a continuous and coercive bilinear form, and $L : V \to \mathbb{R}$ a continuous linear form, we consider the variational formulation

$$\text{find } u \in V \text{ such that } b(u,v) = L(v)\,,\ \forall v \in V. \tag{2.1}$$

The basic idea of the Galerkin method is to replace the Hilbert space V on which we pose the variational formulation by a subspace V_h of finite dimension. The approximate problem posed over V_h reduces to the simple solution of a linear system. In addition, we can choose the construction of V_h in such a way that the subspace V_h is a good approximation of V and that the solution u_h in V_h of the variational formulation is close to the exact solution u in V.

Assume that the Hilbert space V is separable and infinite dimensional, which implies that there exists a Hilbertian basis $(e_i)_{i\geq 1}$ of V. We then choose $\mathscr{V}$ as the subspace generated by this Hilbertian basis (generated by a finite linear combination) which is dense in V. By setting $h = 1/n$, we define V_h as the finite dimensional subspace generated by $(e_1, \ldots, e_n)$. The Galerkin method consists to solve the internal variational approximation associated to (2.1), i.e.

$$\text{find } u_h \in V_h \text{ such that } b(u_h, v_h) = L(v_h)\,,\ \forall v_h \in V_h. \tag{2.2}$$

The existence and uniqueness of $u_h \in V_h$, the solution of (2.1), follows from the Lax-Milgram theorem applied to V_h.

Theorem 2.1.1. *(Lax-Milgram) Let V be a real Hilbert space, $L(\cdot)$ a continuous linear form on V, $b(\cdot,\cdot)$ a continuous coercive bilinear form on V . Then the variational formulation (2.1) has a unique solution. Furthermore, this solution depends continuously on the linear form L.*

Then the following lemma, due to Jean Céa, shows that the distance between the exact solution u and the approximate solution u_h is bounded uniformly with respect to the subspace V_h by the distance between u and V_h.

Lemma 2.1.1. *(Céa's lemma) Let u be the solution of (2.1) and u_h the solution of (2.2) we have*

$$\|u-u_h\| \leq \frac{M}{\nu} \inf_{v_h \in V_h} \|u-v_h\|, \tag{2.3}$$

where $\nu > 0$ is the coercivity constant and $M > 0$ is the continuity constant of the bilinear form $b(u,v)$.

Proof. Because $V_h \subset V$, we deduce, by subtraction of the variational formulations (2.1) and (2.2), that $b(u-u_h, w_h) = 0,\ \forall w_h \in V_h$. By choosing $w_h = u_h - v_h$ we obtain

$$\nu\|u-u_h\|^2 \leq b(u-u_h, u-u_h) = b(u-u_h, u-v_h) \leq M\|u-u_h\|\|u-v_h\|, \tag{2.4}$$

from which we deduce (2.3). □

Finally we deduce that the approximate solution u_h converges to the exact solution u from the following lemma

Lemma 2.1.2. *If there exists a mapping r_h from $\mathscr{V}$ into V_h (called an interpolation operator) such that*

$$\lim_{h\to 0} \|v - r_h(v)\| = 0, \quad \forall v \in \mathscr{V}, \tag{2.5}$$

then the method of internal variational approximation converges, i.e.,

$$\lim_{h\to 0} \|u-u_h\| = 0. \tag{2.6}$$

Proof. Let $\varepsilon > 0$, by density of $\mathscr{V}$, there exists $v \in \mathscr{V}$ such that $\|u-v\| \leq \varepsilon$. From hypothesis (2.5) there exists $h_0(\varepsilon) > 0$ such that $\|v - r_h(v)\| = 0,\ \forall h \leq h_0(\varepsilon)$. From Lemma 2.1.1 we write

$$\|u-u_h\| \leq C\|u-r_h(v)\| \leq C(\|u-v\| + \|v-r_h(v)\|) \leq 2C\varepsilon, \tag{2.7}$$

from which we deduce the result. □

The interpolation operator r_h is then simply the orthogonal projection over V_h (which is here defined in all of V and not only in $\mathscr{V}$).

Recall that u_h is calculated by solving a linear system which is equivalent to the problem (2.2). We use the decomposition of the unknown u_h in the basis $(e_1,\ldots,e_n)$, i.e.,

$$u_h = \sum_{j=1}^{n} u_j e_j. \tag{2.8}$$

We set $U_h = (u_1, \ldots, u_n)$ the vector in $\mathbb{R}^n$ of coordinates of u_h. Problem (2.2) is equivalent to

$$\text{find } U_h \in \mathbb{R}^n \text{ such that } b\left(\sum_{j=1}^{n} u_j e_j, e_i\right) = L(e_i)\,,\ \forall 1 \leq i \leq n, \tag{2.9}$$

which can be written as the form of a linear system

$$K_h U_h = l_h, \tag{2.10}$$

with for $1 \leq i, j \leq n$, $(K_h)_{ij} = b(e_j, e_i)$ and $(l_h)_i = L(e_i)$.

Despite its usefulness in the theoretical framework, the Galerkin method is not helpful from a numerical point of view. The matrix K_h is generally full, that is, all the coefficients are nonzero in general, and ill-conditioned, that is, the numerical solution of the linear system will be unstable and so very sensitive to rounding errors on the computer.

2.1.2 A brief overview of Finite Element methods

Historically, the first premises of FE methods have been proposed by the mathematician Richard Courant in the 1940s, but it was mechanical engineers who have developed, popularized, and proved the efficiency of this method in the 1950s and 1960s (as well as giving it its actual name). After these first practical successes, mathematicians have then considerably developed the theoretical foundations of the method and proposed significant improvements. This is a good example of interdisciplinary cooperation where the joint efforts of engineers and applied mathematicians have made immense progress in numerical simulation (not neglecting the even more spectacular advances in the power of computers).

The principle of FE methods is to construct internal approximation spaces V_h from the usual functional spaces $H^1(\Omega)$, $H_0^1(\Omega)$, $H^2(\Omega)$, ... , whose definition is based on the geometrical concept of a mesh of the domain Ω. A mesh is a tessellation of the space by very simple elementary volumes: triangles, tetrahedra, parallelopipeds. In this context the parameter h of V_h corresponds to the maximum size of the mesh or the cells which comprise the mesh. Typically a basis of V_h will be composed of functions whose support is localized in one or few elements. This will have two important consequences: on the one hand, in the limit $h \to 0$, the space V_h will be more and more large and will approach little by little the entire space V, and on the other hand, the stiffness matrix K_h of the linear system (2.10) will be sparse, which will limit the cost of the numerical solution.

FE methods have seen wide use in numerical electromagnetics. The two primary reasons for using FE method for electromagnetic problems are its geometric flexibility and the possibility of using high-order interpolation of the field components. Geometric flexibility arises because the grid in FE can use arbitrary polygons or polyhedra, and these can be designed to match the shapes of objects in the simulation space or the expected field variations. Then, the fields are assumed to take a particular functional form (usually polynomial) over this element; accuracy can be improved by assuming higher order polynomial basis functions on each element. However, the method suffers from the non-locality of the scheme. Even if the basis functions are chosen to be local, enforcing continuity of the function or its derivatives at the boundary of each element means that a large (albeit sparse) matrix must be solved in order to solve the system of equations. For 2D or 3D problems, the overhead involved just in solving the system of equations can dominate the overall computational cost.

2.1.3 A brief overview of Finite Volume methods

FV methods were introduced in the field of computational fluid dynamics in the beginning of the seventies. As a matter of fact, they are particularly well suited to the numerical treatment of systems of conservation laws such as the Navier-Stokes equations. In general, conservation laws can be written in the form

$$\frac{\partial u}{\partial t} + \operatorname{div}\left(\mathscr{F}(u)\right) = \mathscr{S}(u), \tag{2.11}$$

where $\mathscr{F}$ is a flux tensor of some variable, and $\mathscr{S}$ is a source term. Now, application of the FV method involves an integration of this general equation over small volumetric domains (in 3D)

$$\int_{V_i} \frac{\partial u}{\partial t} + \int_{V_i} \operatorname{div}\left(\mathscr{F}(u)\right) = \int_{V_i} \mathscr{S}(u), \tag{2.12}$$

We next assume the field u is constant over the volume V_i and apply the divergence theorem (see Theorem 1.1.1)

$$\int_{V_i} \frac{\partial u}{\partial t} + \oint_{S_i} \overline{\mathscr{F}_i}(u) \cdot n \, ds = \int_{V_i} \mathscr{S}(u), \tag{2.13}$$

that is

$$\frac{du_i}{dt} + \frac{1}{V_i} \oint_{S_i} \overline{\mathscr{F}_i}(u) \cdot n \, ds = \mathscr{S}(u_i), \tag{2.14}$$

where n is the unit normal to the surface S_i which encloses V_i. The $\overline{\mathscr{F}_i}$ is the numerical flux, explicitly differentiated from the physical flux $\mathscr{F}$. This numerical flux is an estimate of the flux leaving or entering the volume V_i through its surface S_i; the estimation of this numerical flux is what defines the particular finite volume method.

The FV approach has been widely adopted by computational fluid dynamics scientists and has now nearly supplanted classical finite difference and finite element methods in solving problems of nonlinear convection. The success of the finite volume method is due to its ability to capture discontinuous solutions which may occur when solving nonlinear equations or more simply, when convecting discontinuous initial data in the linear case. The FV method easily leads to strictly local, explicit schemes for hyperbolic problems and is thus a natural choice for the solution of Maxwell's equations if they are cast in conservative form [Fumeaux *et al.* 2006, Remaki 2000, Shankar *et al.* 1990]. However, the most severe limitation of FV technique, is that achieving high-order accuracy destroys the locality of the scheme, requiring progressively larger and larger stencils for higher order accuracy.

2.1.4 Discontinuous Galerkin methods

The first DG method was introduced in 1973 by Reed and Hill for solving the steady-state neutron transport equation [Reed & Hill 1973], i.e. a time independent linear hyperbolic equation. The first analysis of this method was presented in [LeSaint & Raviart 1974]. Since then, this class of finite element methods has enjoyed an increasing interest and has been applied in many domains based on hyperbolic systems for acoustics, electromagnetics, elastodynamics, gas dynamics and plasma dynamics applications.

DG methods can be viewed as a clever combination of FE methods and FV methods. A space of basis and test functions is defined as in FE methods, while the equation is satisfied in a sense closer to FV methods. In DG approaches the approximate solution is allowed to be discontinuous across

element boundaries, thus a local variational formulation is established for each element, not on the whole space. The global solution is then the direct sum of local solutions and the global mass matrix is a sparse block diagonal matrix. The size of the blocks is then equal to the number of degrees of freedom inside the corresponding elements and the blocks can be inverted by hand or by using a symbolic manipulator once for all. Finally, the discontinuity across element boundaries impose to define an approximate trace on the boundaries leading to the notion of numerical flux as in FV methods.

Ideally, DG methods share almost all the advantages of FE methods (large spectrum of applications, complex geometries, etc.) and FV methods (ability to capture discontinuous solutions). The DG method has other nice properties which explain the renewed interest it gains in various domains in scientific computing as witnessed by books or special issues of journals dedicated to this method [Cockburn *et al.* 2000, Cockburn & Shu 2005, Dawson 2006, Hesthaven & Warburton 2008]:

- It is naturally adapted to a high-order approximation of the unknown field. Moreover, one may increase the degree of the approximation in the whole mesh as easily as for spectral methods but, with a DG method, this can also be done very locally. In most cases, the approximation relies on a polynomial interpolation method but the DG method also offers the flexibility of applying local approximation strategies that best fit to the intrinsic features of the modeled physical phenomena.

- When the discretization in space is coupled to an explicit time integration method, the DG method leads to a block diagonal mass matrix independently of the form of the local approximation (e.g the type of polynomial interpolation). This is a striking difference with classical, continuous finite element formulations. Moreover, the mass matrix is diagonal if an orthogonal basis is chosen.

- It is flexible with regards to the choice of the time stepping scheme. One may combine the DG spatial discretization with any global or local explicit time integration scheme, or even implicit, provided the resulting scheme is stable.

- It is naturally adapted to parallel computing. As long as an explicit time integration scheme is used, the DG method is easily parallelized. Moreover, the compact nature of DG discretization schemes is in favor of high computation to communication ratio especially when the interpolation order is increased.

2.2 Application of DG spatial discretization for Maxwell's equations

2.2.1 Problem statement and notations

We consider the time-domain Maxwell equations (1.38) on a bounded three-dimensional domain, denoted $\Omega \subset \mathbb{R}^3$

$$\left\{ \begin{array}{rcl} \varepsilon(\mathbf{x}) \dfrac{\partial \mathbf{E}(\mathbf{x},t)}{\partial t} & = & \operatorname{curl}(\mathbf{H}(\mathbf{x},t)) - \sigma(\mathbf{x})\mathbf{E}(\mathbf{x},t) - \mathbf{J}^s(\mathbf{x},t), \\ \mu(\mathbf{x}) \dfrac{\partial \mathbf{H}(\mathbf{x},t)}{\partial t} & = & -\operatorname{curl}(\mathbf{E}(\mathbf{x},t)), \end{array} \right. \tag{2.15}$$

where $\mathbf{E}(\mathbf{x},t) = (E^x, E^y, E^z)^T$ and $\mathbf{H}(\mathbf{x},t) = (H^x, H^y, H^z)^T$ denote the electric and magnetic fields. $\varepsilon(\mathbf{x})$, $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are coefficients representing dielectric permittivity, magnetic permeability and conductivity, respectively, and $\mathbf{J}^s(\mathbf{x},t)$ represents the source current density. We denote $\partial\Omega = \Gamma_m \cup \Gamma_a$ the boundary of Ω on which we impose the following boundary conditions

$$\begin{cases} \mathbf{n}\times\mathbf{E} = 0 \quad \text{on } \Gamma_m, \\ \mathbf{n}\times\mathbf{E} - \sqrt{\dfrac{\mu}{\varepsilon}}\mathbf{n}\times(\mathbf{H}\times\mathbf{n}) = \mathbf{n}\times\mathbf{E}^{inc} - \sqrt{\dfrac{\mu}{\varepsilon}}\mathbf{n}\times\left(\mathbf{H}^{inc}\times\mathbf{n}\right) \quad \text{on } \Gamma_a, \end{cases} \tag{2.16}$$

where $\mathbf{n}$ denotes the unit outward normal to $\partial\Omega$ and $\left(\mathbf{E}^{inc}, \mathbf{H}^{inc}\right)$ is a given incident field. The first boundary condition is called metallic (referring to a perfectly conductive surface) while the second one is called absorbing and takes the form of the Silver-Müller condition which is a first order approximation of the exact absorbing boundary condition. Finally the system of equations (2.15) is closed with the initial conditions

$$\mathbf{E}(\mathbf{x}, t=0) = \mathbf{E}_0(\mathbf{x}) \quad \text{and} \quad \mathbf{H}(\mathbf{x}, t=0) = \mathbf{H}_0(\mathbf{x}). \tag{2.17}$$

Let Ω_h be a partition of Ω into a set of N_h tetrahedra τ_i of size h_i with boundary $\partial\tau_i$

$$\Omega \simeq \Omega_h = \bigcup_{i=1}^{N_h} \tau_i. \tag{2.18}$$

By convention h denotes the maximum diameter of the (non-uniform) grid elements

$$h = \max_{\tau_i \in \Omega_h} h_i. \tag{2.19}$$

As in a finite element method for the given partition Ω_h we seek approximate electric and magnetic fields $(\mathbf{E}_h, \mathbf{H}_h)$ of $(\mathbf{E}, \mathbf{H})$ solution to Maxwell's equations (2.15) in a subspace V_h. We define the following finite dimensional subspace

$$V_h = \left\{ \mathbf{v} = (v_1, v_2, v_3)^T \in \left[L^2(\Omega)\right]^3 \ : \ v_{k|_{\tau_i}} \in \mathbb{P}_{p_i}(\tau_i), \quad \forall k, \quad \forall \tau_i \in \Omega_h \right\}, \tag{2.20}$$

where $\mathbb{P}_{p_i}(\tau_i)$ denotes the space of polynomial functions of degree at most p_i inside the element τ_i, that is, all $v \in \mathbb{P}_{p_i}(\tau_i)$ are written in the form

$$v(\mathbf{x}) = \sum_{\substack{i_1,i_2,i_3 \geq 0 \\ i_1+i_2+i_3 \leq p_i}} \alpha_{i_1,i_2,i_3}\, x^{i_1} y^{i_2} z^{i_3}. \tag{2.21}$$

Following the DG approach, inside each finite element τ_i, the local electric and magnetic fields $(\mathbf{E}_{h|_{\tau_i}}, \mathbf{H}_{h|_{\tau_i}}) = (\mathbf{E}_i, \mathbf{H}_i)$ are expressed as linear combination of linearly independent basis vector Φ_{il}, $1 \leq l \leq 3d_i$

$$\mathbf{E}_i(\mathbf{x},t) = \sum_{l=1}^{3d_i} E_{il}(t)\Phi_{il}(\mathbf{x}) \quad \text{and} \quad \mathbf{H}_i(\mathbf{x},t) = \sum_{l=1}^{3d_i} H_{il}(t)\Phi_{il}(\mathbf{x}), \tag{2.22}$$

where d_i denotes the local number of degrees of freedom associated to the interpolation degree p_i in τ_i, i.e.

$$d_i = \frac{(p_i+1)(p_i+2)(p_i+3)}{6}, \tag{2.23}$$

and E_{il}, H_{il} reflect nodal values of $\mathbf{E}_i$ and $\mathbf{H}_i$, respectively. The global solution of Maxwell's equations (2.15) is then given by

$$\mathbf{E}(\mathbf{x},t) \simeq \mathbf{E}_h(\mathbf{x},t) = \bigoplus_{i=1}^{N_h} \mathbf{E}_i(\mathbf{x},t) \quad \text{and} \quad \mathbf{H}(\mathbf{x},t) \simeq \mathbf{H}_h(\mathbf{x},t) = \bigoplus_{i=1}^{N_h} \mathbf{H}_i(\mathbf{x},t). \tag{2.24}$$

To avoid any ambiguity we specify the basis vector $(\Phi_{il})_{1\le l\le 3d_i}$ of the space $[\mathbb{P}_{p_i}(\tau_i)]^3$ and the sequences of nodal values $(E_{il})_{1\le l\le 3d_i}$ and $(H_{il})_{1\le l\le 3d_i}$. We denote by $(\varphi_{il})_{1\le l\le d_i}$ the basis of the polynomial functions space $\mathbb{P}_{p_i}(\tau_i)$, then we define

$$(\Phi_{il})_{1\le l\le 3d_i} = \left(\begin{pmatrix} \varphi_{i1} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \varphi_{i1} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \varphi_{i1} \end{pmatrix}, \ \dots\ , \begin{pmatrix} \varphi_{id_i} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \varphi_{id_i} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \varphi_{id_i} \end{pmatrix} \right). \tag{2.25}$$

Regarding $(E_{il})_{1\le l\le 3d_i}$ we order the sequence as follow

$$(E_{il})_{1\le l\le 3d_i} = \left(E^x_{i1},\ E^y_{i1},\ E^z_{i1},\ \dots\ ,\ E^x_{id_i},\ E^y_{id_i},\ E^z_{id_i} \right), \tag{2.26}$$

and similarly for the sequence $(H_{il})_{1\le l\le 3d_i}$.

For each τ_i, ε_i, μ_i and σ_i denote respectively the local electric permittivity, magnetic permeability of the medium and the conduction coefficient, which are assumed constant inside the element τ_i. For two distinct tetrahedra τ_i and τ_k in Ω_h, the intersection $\tau_i \cap \tau_k$ is a triangle a_{ik} which we call interface. The unitary normal vector of the interface a_{ik} is denoted $\mathbf{n}_{ik}$, oriented from τ_i to τ_k. For the boundary interface, the index k corresponds to a fictitious element outside the domain. We denote by ν_i the set of indices of the elements which have a common interface with τ_i. Finally we set

$$\nu_i = \nu_i^i \cup \nu_i^m \cup \nu_i^a, \tag{2.27}$$

where

- $\nu_i^i = \{k \in \nu_i \ :\ a_{ik} \text{ is an internal interface } \}$,
- $\nu_i^m = \{k \in \nu_i \ :\ a_{ik} \in \Gamma_m\}$,
- $\nu_i^a = \{k \in \nu_i \ :\ a_{ik} \in \Gamma_a\}$.

2.2.2 Local formulation of the DG method

We now derive the DG spatial discretization. To simplify the presentation we assume that the given source current density in (2.15) is equal to zero. Following the DG approach we establish local variational formulations on each element of the space grid. Dot-multiplying (2.15) by any given vector $\Phi \in Span(\Phi_{ij}, 1 \le j \le 3d_i)$, integrating over each element τ_i, integrating by part and finally replacing the exact fields $(\mathbf{E},\mathbf{H})$ by the approximate fields $(\mathbf{E}_h,\mathbf{H}_h)$, yields

$$\begin{cases} \displaystyle\int_{\tau_i} \Phi\cdot\varepsilon_i\partial_t\mathbf{E}_h\, d\mathbf{x} - \int_{\tau_i} \mathrm{curl}\Phi\cdot\mathbf{H}_h\, d\mathbf{x} + \int_{\partial\tau_i} \Phi\cdot(\mathbf{H}_h\times\mathbf{n})\, ds + \int_{\tau_i} \Phi\cdot\sigma_i\mathbf{E}_h\, d\mathbf{x} = 0, \\ \displaystyle\int_{\tau_i} \Phi\cdot\mu_i\partial_t\mathbf{H}_h\, d\mathbf{x} + \int_{\tau_i} \mathrm{curl}\Phi\cdot\mathbf{E}_h\, d\mathbf{x} - \int_{\partial\tau_i} \Phi\cdot(\mathbf{E}_h\times\mathbf{n})\, ds = 0. \end{cases} \tag{2.28}$$

One of the main features of the DG approach is that the approximate fields $(\mathbf{E}_h, \mathbf{H}_h)$ are allowed to be discontinuous across element boundaries. Then for such discontinuous fields we must define an approximate trace (i.e. a numerical trace) to evaluate the integrals over $\partial\tau_i$. In this study, we choose to use a centered approximation

$$\forall i,\ \forall k \in \nu_i, \quad \mathbf{E}_h|_{a_{ik}} = \frac{\mathbf{E}_i|_{a_{ik}} + \mathbf{E}_k|_{a_{ik}}}{2} \quad \text{and} \quad \mathbf{H}_h|_{a_{ik}} = \frac{\mathbf{H}_i|_{a_{ik}} + \mathbf{H}_k|_{a_{ik}}}{2}. \tag{2.29}$$

Now using $(\mathbf{E}_h|_{\tau_i}, \mathbf{H}_h|_{\tau_i}) = (\mathbf{E}_i, \mathbf{H}_i)$ for volume integrals and evaluating the surface integrals in (2.28) with the centered numerical flux (2.29) , yields

$$\left\{\begin{array}{rcl} \displaystyle\int_{\tau_i} \Phi\cdot\varepsilon_i\partial_t\mathbf{E}_i\ d\mathbf{x} & = & \displaystyle\int_{\tau_i} \mathrm{curl}\Phi\cdot\mathbf{H}_i\ d\mathbf{x} - \frac{1}{2}\sum_{k\in\nu_i}\int_{a_{ik}} \Phi\cdot\left((\mathbf{H}_i+\mathbf{H}_k)\times\mathbf{n}_{ik}\right)\ ds \\ & - & \displaystyle\int_{\tau_i} \Phi\cdot\sigma_i\mathbf{E}_i\ d\mathbf{x}, \\ \displaystyle\int_{\tau_i} \Phi\cdot\mu_i\partial_t\mathbf{H}_i\ d\mathbf{x} & = & \displaystyle\int_{\tau_i} \mathrm{curl}\Phi\cdot\mathbf{E}_i\ d\mathbf{x} + \frac{1}{2}\sum_{k\in\nu_i}\int_{a_{ik}} \Phi\cdot\left((\mathbf{E}_i+\mathbf{E}_k)\times\mathbf{n}_{ik}\right)\ ds. \end{array}\right. \tag{2.30}$$

Hence, by re-integration by parts

$$\left\{\begin{array}{rcl} \displaystyle\int_{\tau_i} \Phi\cdot\varepsilon_i\partial_t\mathbf{E}_i\ d\mathbf{x} & = & \displaystyle\frac{1}{2}\int_{\tau_i} \left(\mathrm{curl}\Phi\cdot\mathbf{H}_i + \mathrm{curl}\mathbf{H}_i\cdot\Phi\right)\ d\mathbf{x} - \frac{1}{2}\sum_{k\in\nu_i}\int_{a_{ik}} \Phi\cdot\left(\mathbf{H}_k\times\mathbf{n}_{ik}\right)\ ds \\ & - & \displaystyle\int_{\tau_i} \Phi\cdot\sigma_i\mathbf{E}_i\ d\mathbf{x}, \\ \displaystyle\int_{\tau_i} \Phi\cdot\mu_i\partial_t\mathbf{H}_i\ d\mathbf{x} & = & \displaystyle-\frac{1}{2}\int_{\tau_i} \left(\mathrm{curl}\Phi\cdot\mathbf{E}_i + \mathrm{curl}\mathbf{E}_i\cdot\Phi\right)\ d\mathbf{x} + \frac{1}{2}\sum_{k\in\nu_i}\int_{a_{ik}} \Phi\cdot\left(\mathbf{E}_k\times\mathbf{n}_{ik}\right)\ ds. \end{array}\right. \tag{2.31}$$

Then for each face on the boundary of Ω_h, $a_{ik} \in \Gamma_m \cup \Gamma_a$, the trace of a fictitious neighboring element is needed for the computation of the numerical flux. We treat the boundary conditions defined in (2.16) in a weak sense by defining appropriate values of the electric and magnetic fields in the fictitious element

$$\begin{array}{ll} \forall a_{ik} \in \Gamma_m & \left\{\begin{array}{l} \mathbf{E}_k|_{a_{ik}} = -\mathbf{E}_i|_{a_{ik}}, \\ \mathbf{H}_k|_{a_{ik}} = \mathbf{H}_i|_{a_{ik}}, \end{array}\right. \\ \forall a_{ik} \in \Gamma_a & \left\{\begin{array}{l} \mathbf{E}_k|_{a_{ik}} = \sqrt{\dfrac{\mu_i}{\varepsilon_i}}\left(\mathbf{H}_i|_{a_{ik}}\times\mathbf{n}_{ik}\right) + \mathbf{E}_i^{inc}|_{a_{ik}} - \sqrt{\dfrac{\mu_i}{\varepsilon_i}}\left(\mathbf{H}_i^{inc}|_{a_{ik}}\times\mathbf{n}_{ik}\right), \\ \mathbf{H}_k|_{a_{ik}} = -\sqrt{\dfrac{\varepsilon_i}{\mu_i}}\left(\mathbf{E}_i|_{a_{ik}}\times\mathbf{n}_{ik}\right) + \mathbf{H}_i^{inc}|_{a_{ik}} + \sqrt{\dfrac{\varepsilon_i}{\mu_i}}\left(\mathbf{E}_i^{inc}|_{a_{ik}}\times\mathbf{n}_{ik}\right). \end{array}\right. \end{array} \tag{2.32}$$

Remark 2.2.1. *The values of electric and magnetic fields (2.32) are used in the computation of boundary fluxes in (2.30). Then they both are cross-multiplied by the local normal $\mathbf{n}_{ik}$, which yields back the original form of the the Silver-Müller condition given in (2.16). Among many possible choices, the origin of these fluxes is not really obvious. Indeed, the absorbing boundary condition is exact for outgoing plane waves, with a wave vector collinear with $\mathbf{n}_{ik}$. It is a first-order approximation, asymptotically correct when the fictitious absorbing boundary is far enough and normal to the wave propagation. Finally, note that these values correspond to upwind fluxes at the absorbing boundary, based on the hyperbolic nature of the global six-component Maxwell system.*

The system of equations (2.31) can be written in terms of scalar unknowns. According to the decomposition (2.22) of the local electric and magnetic fields $(\mathbf{E}_i, \mathbf{H}_i)$, denoting the column vectors $(E_{il})_{1\le l\le 3d_i}$ and $(H_{il})_{1\le l\le 3d_i}$ by E_i and H_i, respectively, and replacing Φ by the basis vectors $(\Phi_{ij})_{1\le j\le 3d_i}$, we obtain the equivalent system

$$\begin{cases} M_i^{\varepsilon}\,\partial_t E_i = K_i H_i - \displaystyle\sum_{k\in\nu_i} S_{ik} H_k - D_i E_i, \\ M_i^{\mu}\,\partial_t H_i = -K_i E_i + \displaystyle\sum_{k\in\nu_i} S_{ik} E_k, \end{cases} \tag{2.33}$$

where

- M_i^{ς} (ς stands for ε or μ) are the symmetric, positive definite mass matrices

$$\left(M_i^{\varsigma}\right)_{jl} = \int_{\tau_i} \Phi_{ij}^T\, \varsigma_i\, \Phi_{il}\, d\mathbf{x} \quad (1\le j,l\le 3d_i), \tag{2.34}$$

- K_i is the symmetric stiffness matrix

$$(K_i)_{jl} = \frac{1}{2}\int_{\tau_i} \left(\Phi_{ij}^T\, \mathrm{curl}\Phi_{il} + \Phi_{il}^T\, \mathrm{curl}\Phi_{ij}\right)\, d\mathbf{x} \quad (1\le j,l\le 3d_i), \tag{2.35}$$

- S_{ik} are the rectangular interface matrices

$$(S_{ik})_{jl} = \frac{1}{2}\int_{a_{ik}} \Phi_{ij}^T (\Phi_{kl}\times\mathbf{n}_{ik})\, ds \quad (1\le j\le 3d_i,\ 1\le l\le 3d_k). \tag{2.36}$$

- D_i is the symmetric, positive semi-definite conduction matrix

$$(D_i)_{jl} = \int_{\tau_i} \Phi_{ij}^T\, \sigma_i\, \Phi_{il}\, d\mathbf{x} \quad (1\le j,l\le 3d_i), \tag{2.37}$$

Taking into account the trace of fictitious neighboring elements (2.32) we can rewrite the local system of ODEs (2.33) as

$$\begin{cases} M_i^{\varepsilon}\,\partial_t E_i = K_i H_i - \displaystyle\sum_{k\in\nu_i^i} S_{ik} H_k - \sum_{k\in\nu_i^m} S_{ik}^m H_i + \sum_{k\in\nu_i^a} \left(S_{ik}^E E_i - X_{ik}^E W_i^{inc}\right) - D_i E_i, \\ M_i^{\mu}\,\partial_t H_i = -K_i H_i + \displaystyle\sum_{k\in\nu_i^i} S_{ik} E_k - \sum_{k\in\nu_i^m} S_{ik}^m H_i + \sum_{k\in\nu_i^a} \left(S_{ik}^H H_i - X_{ik}^H W_i^{inc}\right), \end{cases} \tag{2.38}$$

where

- $(S_{ik}^m)_{jl} = \dfrac{1}{2}\displaystyle\int_{a_{ik}} \Phi_{ij}^T (\Phi_{il}\times\mathbf{n}_{ik})\, ds \quad (1\le j,l\le 3d_i),$
- $\left(S_{ik}^E\right)_{jl} = \dfrac{1}{2}\sqrt{\dfrac{\varepsilon_i}{\mu_i}}\displaystyle\int_{a_{ik}} \Phi_{ij}^T ((\Phi_{il}\times\mathbf{n}_{ik})\times\mathbf{n}_{ik})\, ds \quad (1\le j,l\le 3d_i),$
- $\left(S_{ik}^H\right)_{jl} = \dfrac{1}{2}\sqrt{\dfrac{\mu_i}{\varepsilon_i}}\displaystyle\int_{a_{ik}} \Phi_{ij}^T ((\Phi_{il}\times\mathbf{n}_{ik})\times\mathbf{n}_{ik})\, ds \quad (1\le j,l\le 3d_i),$

- W_i^{inc} is the column vector $\left(E_i^{inc}, H_i^{inc}\right)^T$ of length $6d_i$, and for $1 \leq j \leq 3d_i$

$$\left(X_{ik}^E\right)_{jl} = \begin{cases} -\frac{1}{2}\sqrt{\frac{\varepsilon_i}{\mu_i}} \int_{a_{ik}} \Phi_{ij}^T \left((\Phi_{il} \times \mathbf{n}_{ik}) \times \mathbf{n}_{ik}\right)\, ds \quad (1 \leq l \leq 3d_i), \\ \frac{1}{2} \int_{a_{ik}} \Phi_{ij}^T \left(\Phi_{i(l-3d_i)} \times \mathbf{n}_{ik}\right)\, ds \quad (3d_i+1 \leq l \leq 6d_i), \end{cases}$$

$$\left(X_{ik}^H\right)_{jl} = \begin{cases} -\frac{1}{2} \int_{a_{ik}} \Phi_{ij}^T \left(\Phi_{il} \times \mathbf{n}_{ik}\right)\, ds \quad (1 \leq l \leq 3d_i), \\ \frac{1}{2}\sqrt{\frac{\mu_i}{\varepsilon_i}} \int_{a_{ik}} \Phi_{ij}^T \left(\left(\Phi_{i(l-3d_i)} \times \mathbf{n}_{ik}\right) \times \mathbf{n}_{ik}\right)\, ds \quad (3d_i+1 \leq l \leq 6d_i), \end{cases}$$

2.2.3 From local to global formulation of the DG method

The set of local semi-discrete system (2.33) for each finite element τ_i can be transformed into a global system. Gathering all electric and magnetic unknowns in column vectors of size $d = \sum_{i=1}^{N_h} d_i$, denoted E and H, respectively, we write

$$\begin{cases} M^{\varepsilon} \partial_t E = KH - AH - BH + C^E E - X^E W^{inc} - DE, \\ M^{\mu} \partial_t H = -KE + AE - BE + C^H H - X^H W^{inc}, \end{cases} \tag{2.39}$$

where

- M^{ς} and K are $3d \times 3d$ block diagonal mass and stiffness matrices with diagonal blocks equal to M_i^{ς} and K_i, respectively. Then the matrices M^{ς} are symmetric, positive definite and K is symmetric.
- A is a $3d \times 3d$ block sparse matrix, whose nonzero blocks equal to S_{ik} when a_{ik} is an internal interface. From (2.36) and the equality $\mathbf{n}_{ki} = -\mathbf{n}_{ik}$ we can check that $S_{ki} = S_{ik}^T$ and then A is symmetric

$$(S_{ki})_{lj} = \frac{1}{2} \int_{a_{ik}} -\Phi_{kl}^T \left(\Phi_{ij} \times \mathbf{n}_{ik}\right)\, ds = \frac{1}{2} \int_{a_{ik}} \Phi_{ij}^T \left(\Phi_{kl} \times \mathbf{n}_{ik}\right)\, ds = (S_{ik})_{jl}\,.$$

- B is a $3d \times 3d$ block diagonal matrix, whose nonzero blocks equal to S_{ik}^m when a_{ik} is a metallic boundary face. In that case $(S_{ik}^m)_{jl} = -(S_{ik}^m)_{lj}$, then $S_{ik}^m = -(S_{ik}^m)^T$ and B is skew-symmetric.
- C^E and C^H are $3d \times 3d$ block diagonal matrices, whose nonzero blocks equal to S_{ik}^E and S_{ik}^H, respectively, when a_{ik} is an absorbing boundary face.
- W^{inc} denotes the column vectors $(E_{il}, H_{il})_{1 \leq l \leq 3d_i}$ for $1 \leq i \leq N_h$. X^E and X^H are $3d \times 6d$ block diagonal matrices, whose nonzero blocks equal to X_{ik}^E and X_{ik}^H, respectively, when a_{ik} is an absorbing boundary face.
- D is a $3d \times 3d$ positive semi-definite, block diagonal matrix with diagonal blocks equal to D_i.

Reintroducing the current density, and setting $S = K - A - B$ the system (2.39) reads

$$\begin{cases} M^{\varepsilon} \partial_t E = SH + C^E E - X^E W^{inc} - DE - j^s, \\ M^{\mu} \partial_t H = -S^T E + C^H H - X^H W^{inc}. \end{cases} \tag{2.40}$$

2.2.4 Stability and conservation properties

We assume that $\Gamma_a = \emptyset$ and $j^s = 0$. Let (E_h, H_h) the solution of the semi-discrete system (2.40), we define the semi-discrete energy

$$\mathscr{E}_h = \frac{1}{2}\left(\|E_h\|^2_{M^\varepsilon} + \|H_h\|^2_{M^\mu}\right), \tag{2.41}$$

where

$$\|E_h\|^2_{M^\varepsilon} = \langle\, M^\varepsilon E_h,\, E_h\rangle\,,\ \|H_h\|^2_{M^\mu} = \langle\, M^\mu H_h,\, H_h\rangle\,, \tag{2.42}$$

and $\langle\,\cdot\,,\,\cdot\,\rangle$ is the L^2 inner product. Taking inner products with E_h and H_h in (2.40) yields, respectively,

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}\langle\, M^\varepsilon E_h,\, E_h\rangle &= \langle\, SH_h,\, E_h\rangle - \langle\, DE_h,\, E_h\rangle\,, \\ \frac{1}{2}\frac{d}{dt}\langle\, M^\mu H_h,\, H_h\rangle &= -\left\langle\, S^T E_h,\, H_h\right\rangle. \end{aligned} \tag{2.43}$$

As D is symmetric semi-positive definite, it follows

$$\frac{d}{dt}\mathscr{E}_h = -\langle\, DE_h,\, E_h\rangle \le 0. \tag{2.44}$$

Hence $\mathscr{E}_h$ is a decreasing function in time so that $\mathscr{E}_h(t) \le \mathscr{E}_h(0)$, showing stability in the L^2 sense and energy conservation would D be zero.

2.3 Some algorithmic details

We call the N-simplex τ of $\mathbb{R}^N$ the convex envelope of $(N+1)$ points $(\mathbf{a}_j)_{1\le j\le N+1}$ of $\mathbb{R}^N$, called the vertices of τ (a 2-simplex is simply a triangle and a 3-simplex a tetrahedron). In the following we will assume that any N-simplex τ is non-degenerate, i.e. the vertices $(\mathbf{a}_j)_{1\le j\le N+1}$ do not belong to the same hyperplane of $\mathbb{R}^N$.

To simplify the implementation we can use an affine transformation to reduce every N-simplex τ of the mesh Ω_h to a reference N-simplex τ_0.

2.3.1 Two-dimensional problems ($N = 2$)

In practice we choose as reference triangle

$$\tau_0 = \left\{\mathbf{x} \in \mathbb{R}^2 \text{ such that } x+y \le 1 \text{ and } x, y \ge 0\right\}. \tag{2.45}$$

Every triangle τ with vertices $(\mathbf{a}_i)_{1\le i\le 3}$ is then the image of τ_0, by an affine transformation, F (Figure 2.2), defined by

$$\begin{array}{rccc} F: & \tau_0 & \longrightarrow & \tau \\ & \mathbf{x}_0 & \longrightarrow & \mathbf{x} = A\mathbf{x}_0 + \mathbf{b} \end{array} \tag{2.46}$$

where

$$A = \begin{pmatrix} a_2^x - a_1^x & a_3^x - a_1^x \\ a_2^y - a_1^y & a_3^y - a_1^y \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} a_1^x \\ a_1^y \end{pmatrix}. \tag{2.47}$$

Note that the Jacobian of the transformation, F, is equal to twice the algebraic area of the triangle, τ

$$det\,(A) = \begin{vmatrix} a_2^x - a_1^x & a_3^x - a_1^x \\ a_2^y - a_1^y & a_3^y - a_1^y \end{vmatrix} = 2 \times Area\,(\tau)\,, \tag{2.48}$$

because it is the ratio of the area of the triangle τ and the triangle of reference (which is $1/2$). With the non-degeneracy condition on τ we deduce that the affine transformation is invertible.

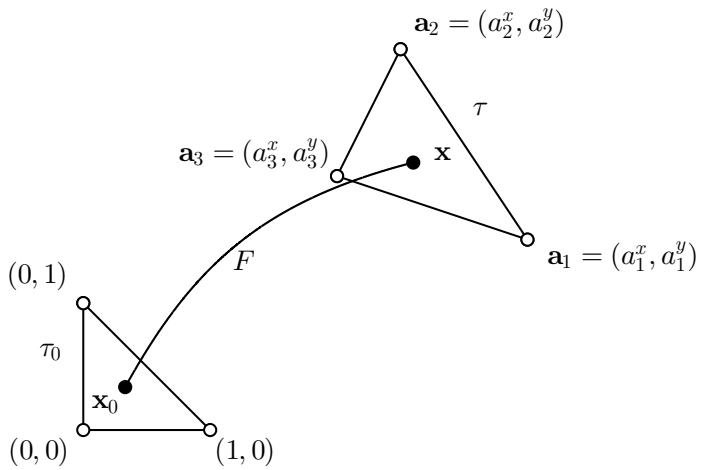


Figure 2.2: Affine transformation between the reference triangle, τ_0, and any triangle τ.

2.3.2 Three-dimensional problems ($N = 3$)

In practice we choose as reference tetrahedron

$$\tau_0 = \left\{ \mathbf{x} \in \mathbb{R}^3 \text{ such that } x + y + z \leq 1 \text{ and } x, y, z \geq 0 \right\}. \tag{2.49}$$

Every tetrahedron τ with vertices $(\mathbf{a}_i)_{1 \leq i \leq 4}$ is then the image of τ_0, by an affine transformation, F (Figure 2.3), defined by

$$\begin{array}{rccl} F: & \tau_0 & \longrightarrow & \tau \\ & \mathbf{x}_0 & \longrightarrow & \mathbf{x} = A\mathbf{x}_0 + \mathbf{b} \end{array} \tag{2.50}$$

where

$$A = \begin{pmatrix} a_2^x - a_1^x & a_3^x - a_1^x & a_4^x - a_1^x \\ a_2^y - a_1^y & a_3^y - a_1^y & a_4^y - a_1^y \\ a_2^z - a_1^z & a_3^z - a_1^z & a_4^z - a_1^z \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} a_1^x \\ a_1^y \\ a_1^z \end{pmatrix}. \tag{2.51}$$

Note that the Jacobian of the transformation, F, is equal to six times the algebraic volume of the tetrahedron, τ

$$det\,(A) = \begin{vmatrix} a_2^x - a_1^x & a_3^x - a_1^x & a_4^x - a_1^x \\ a_2^y - a_1^y & a_3^y - a_1^y & a_4^y - a_1^y \\ a_2^z - a_1^z & a_3^z - a_1^z & a_4^z - a_1^z \end{vmatrix} = 6 \times Volume\,(\tau)\,, \tag{2.52}$$

because it is the ratio of the volume of the tetrahedron τ and the tetrahedron of reference (which is $1/6$). With the non-degeneracy condition on τ we deduce that the affine transformation is invertible.

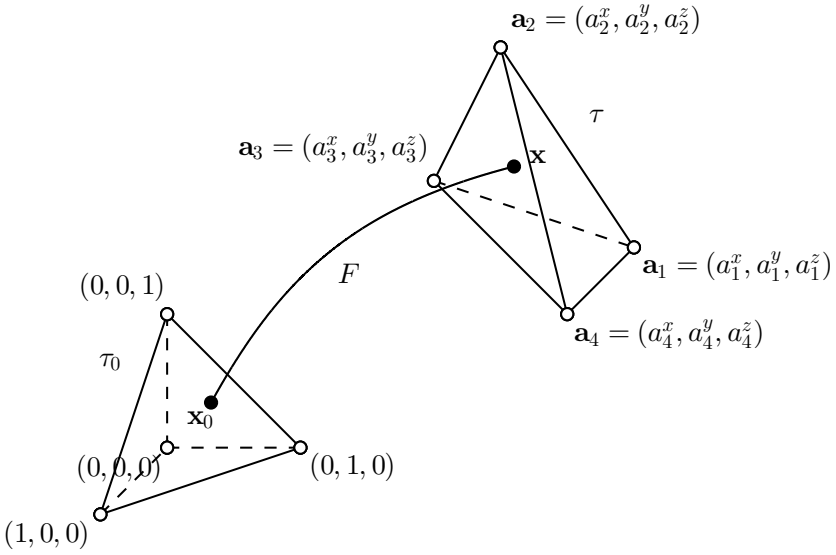


Figure 2.3: Affine transformation between the tetrahedron element, τ_0, and any tetrahedron, τ.

The analytical expressions of the basis functions of the space $\mathbb{P}_k(\tau_0)$ are given in the appendix A. From these expressions we can easily determine the coefficients of the elementary matrices involved in the DG method. Moreover, we have seen that every N-simplex τ is the image of the reference N-simplex τ_0 by an invertible affine transformation, hence by a simple change of variable all calculations can be reduced to calculation on τ_0. In practice the elementary matrices are stored once for all on τ_0; then the coefficients of matrices on any N-simplex are calculated by this change of variable.

CHAPTER 3

Locally implicit time integration methods

Contents

It is well known that, when combined with an explicit time integration method to numerically solve a partial differential equation, a high-order DG method can lead to a severe time step size restriction. An implicit time integration scheme is a natural way to obtain a time-domain method which is unconditionally stable. Starting from the explicit, non-dissipative, DGTD method introduced in [Fezoui *et al.* 2005] Catella et al. have proposed the use of Crank-Nicolson scheme in place of the explicit leap-frog scheme adopted in this method [Catella *et al.* 2010]. As a result, they obtain an unconditionally stable, non-dissipative, implicit DGTD method, but at the expense of the inversion of a global linear system at each time step, thus obliterating one of the attractive features of discontinuous Galerkin formulations. Starting from a generic semi-discrete system of Maxwell's equations, Verwer and Botchev have also studied the use of the implicit Crank-Nicolson scheme, and they compare the obtained results with the explicit leap-frog scheme, for a three-dimensional problem [Verwer & Botchev 2009]. The expense of the inversion of a global linear system at each time step is too large to consider a fully implicit approach as an attractive alternative to explicit methods, especially for three-dimensional problems, where the use of an effective preconditioner is then a mandatory requirement.

A more viable approach to overcome the most severe stability-based time step restrictions, generally induced by local mesh refinements in explicit methods, is to use smaller time steps, given by a local stability criterion, precisely where the smallest elements are located. The local character of DG methods is a very attractive feature for the development of explicit local time-stepping schemes. In [Piperno 2006], Piperno proposed an explicit second-order local time-stepping scheme

for solving Maxwell's equations in a non-conducting medium. The scheme is based on the symplectic Störmer-Verlet method combined with a DG discretization of the first-order formulation of Maxwell's equations. The resulting method conserves a discrete form of the electromagnetic energy. In addition, this scheme allows to treat the entire problem as a set of cell areas for which a local time step is given. Then the evolution of the fields in time can be easily done by a recursive process. However, the requirement of field values at time stations that are not computed implies the loss of the symplectic property of the scheme. In [Monseny *et al.* 2008], the authors combine a similar recursive integrator with a DG formulation on hexahedral elements. Although hexahedral elements are very efficient, the authors of [Buffa & Perugia 2006] have proven that DG methods on conformal hexahedral/quadrilateral meshes can produce spurious modes when the approximation spaces are made of elementwise polynomials of degree k in each variable. To reduce spurious modes, Montseny et al. use additional dissipative terms as penalization in the numerical scheme. In addition, automated grid generation only with hexahedral elements remains a non-trivial task. In [Taube *et al.* 2009], Taube et al. proposed an arbitrary high-order local time-stepping method based on the so-called arbitrary high-order derivatives (ADER) DG approach to Maxwell's equations. The solution is expanded in Taylor series in time and then the Cauchy-Kovalevskaya procedure is used to replace the time derivatives in the Taylor series by space derivatives. Starting from the standard leap-frog scheme, Diaz and Grote derived local time-stepping methods for the second-order wave equations, combined with a symmetric finite element discretization in space [Diaz & Grote 2009]. The resulting fully discrete scheme is explicit and conserves a discrete form of the energy. These methods can also be apply to more general second-order hyperbolic problems, as in electromagnetics, for which symmetric interior penality (IP) DG methods are available [Grote *et al.* 2007, Grote *et al.* 2008]. In [Grote & Mitkova 2010], Grote and Mitkova derived local time-stepping methods for Maxwell's equations from the standard leap-frog scheme. Two second-order accurate schemes in a non-conducting or conducting medium and a fourth-order scheme in a non-conducting medium are given. In a source-free and a non-conducting medium, the authors prove that a discrete form of the energy is conserved, which provide a rigorous criterion for the numerical stability. In addition, in this case, the method can be easily extended to an arbitrary high (even) order. For a conducting medium, the conduction term is treated implicitly, nevertheless since the conduction matrix is essentially diagonal the resulting scheme remains explicit. If there is no local mesh refinement the energy is then no longer conserved but decays with time, independently of the magnitude of the conduction term, under the same CFL condition as in the case without conduction.

Another possibility to overcome step size restrictions is to use implicit-explicit (IMEX) schemes. IMEX methods were originally developed to solve problems with natural splittings into two parts. Such approaches are frequently used in Computational Fluid Dynamics [Wesseling 2001]. The non-stiff part of the model equations is treated explicitly, for instance an advection or a convection term, and the stiff part implicitly, for instance a diffusion term or a term modeling stiff reactions. IMEX multistep methods were introduced by Crouzeix and Varah in [Crouzeix 1980, Varah 1980] for linear parabolic equations. A natural way to derive such methods is to start with an implicit scheme that is known to possess favorable stability properties, and then replace the implicit term by a linear combination of explicit terms using extrapolation. The stability region is then determined by the explicit method [Frank *et al.* 1997]. Theoretical analyses and examples of IMEX multistep methods for time-dependent advection-diffusion-reaction equations are given in [Hundsdorfer & Verwer 2003]. Examples of IMEX multistep schemes are the IMEX-

CNLF scheme using the leap-frog method for the explicit part and the Crank-Nicolson for the implicit part [Verwer 2009], the IMEX-BDF schemes derived from the implicit backward differentiation formula and its explicit counterpart [W. Hundsdorfer 2003], and the IMEX-Adams schemes based on explicit Adams-Bashford methods [Ascher *et al.* 1995]. IMEX RK schemes, based on explicit Runge-Kutta methods, have been also studied in [Ascher *et al.* 1997, Calvo *et al.* 2001, Kanvesky *et al.* 2007, Kennedy & Carpenter 2003]. In [Kanvesky *et al.* 2007], the authors are interested in alleviating the severe stability-based time step restrictions that affect explicit RK time integration schemes when applied to problems that exhibit high levels of geometry-induced stiffness. Starting from a multi-dimensional well-posed conservation law and nodal DG finite element spatial discretization, Kanvesky et al. defined the implicit-explicit splitting in IMEX-RK methods such as the explicit set contains the coarsest elements of the computational domain, while the implicit set contains the smallest elements. Then, the resulting locally implicit scheme allow to overcome the most severe stability-based time-step restrictions, imposed by the set treated explicitly.

In the present study we consider two locally implicit time integration methods for Maxwell's equations. Starting from the well-suited leap-frog scheme for the explicit treatment, similarly to the splitting defined in [Kanvesky *et al.* 2007], only solution variables associated to the smallest grid elements are treated implicitly. The downside of these approaches is the necessity to solve a linear system of equations at each time step. But if the region of refinement is of small size relatively to the computational domain, the overhead will also be small while the solution can be advanced in time with a step size only determined by the coarse elements. In contrast to explicit local time-stepping methods, the locally implicit methods provide a single time step for stability, independent of the fine grid.

The first locally implicit method, that we consider in Section 3.3 has been introduced by Piperno in [Piperno 2006]. This time integration scheme is a blend of the explicit second-order leap-frog scheme (LF2) and the implicit second-order Crank-Nicolson scheme (CN2), based on a nodal DG method for the spatial discretization. LF2 and CN2 schemes and their respective properties are summarized in Section 3.2. The second IMEX method which is treated in Section 3.4 has been proposed by Verwer in [Verwer 2010]. This method is also a combination of the LF2 and CN2 schemes, based on a clever component splitting which allows a significant gain in the sparsity of the matrix of the linear system enabling to solve this latter at lower cost. Although the latter method is more general since the locally implicit scheme is directly formulated on the semi-discrete Maxwell equations, we propose to study this approach for the DG spatial discretization presented in Chapter 2, since the local character of DG methods is particularly well suited to IMEX methods.

Much of this chapter is derived from the articles [Moya 2012] and [Descombes et al. 2013].

3.1 Problem statement

For convenience of presentation, we assume from now that $\Gamma_a = \emptyset$ and similarly to [Botchev & Verwer 2009, Moya 2012, Verwer 2010] we introduce the Cholesky factorization of the mass matrices

$$M^{\varepsilon} = L_{M^{\varepsilon}} L_{M^{\varepsilon}}^T \text{ and } M^{\mu} = L_{M^{\mu}} L_{M^{\mu}}^T, \tag{3.1}$$

where L_{M^ε} and L_{M^μ} are triangular matrices. Then by introducing the change of variables $\widetilde{E} = L_{M^\varepsilon}^T E$ and $\widetilde{H} = L_{M^\mu}^T H$ in (2.40), we write

$$\begin{cases} \partial_t \widetilde{E} &= \widetilde{S}\widetilde{H} - \widetilde{D}\widetilde{E} + \widetilde{j^s}, \\ \partial_t \widetilde{H} &= -\widetilde{S}^T \widetilde{E}, \end{cases} \tag{3.2}$$

where

$$\widetilde{S} = L_{M^\varepsilon}^{-1}\, S\, (L_{M^\mu}^{-1})^T,\ \widetilde{D} = L_{M^\varepsilon}^{-1}\, D\, (L_{M^\varepsilon}^{-1})^T \text{ and } \widetilde{j^s} = L_{M^\varepsilon}^{-1}\, j^s. \tag{3.3}$$

For convenience of notation and presentation we omit in the sequel the "$\sim$" notation in (3.2) i.e.

$$\begin{cases} \partial_t E &= SH - DE - j^s, \\ \partial_t H &= -S^T E. \end{cases} \tag{3.4}$$

The results obtained for (3.2) can always be carried over to (3.4) and vice versa. From now we will proceed with (3.4). Note that the matrix S within (3.4) satisfies [Moya 2012]

$$S \sim \frac{1}{h}, \quad \text{for } h \to 0, \tag{3.5}$$

and the conduction matrix D within (3.4) is diagonal with non-negative entries.

3.2 Explicit and implicit time integration methods

3.2.1 The second-order leap-frog scheme (LF2)

A popular time integration method for the semi-discrete Maxwell system (3.4) is the second order leap-frog scheme that we write in the three-stage form, emanating from Verlet's method, see [Piperno 2006]

$$\begin{cases} \dfrac{H^{n+1/2} - H^n}{\Delta t/2} &= -S^T E^n, \\ \dfrac{E^{n+1} - E^n}{\Delta t} &= SH^{n+1/2} - \dfrac{1}{2} D\left(E^{n+1} + E^n\right) + \dfrac{1}{2}\left(j^s\left(t_{n+1}\right) + j^s\left(t_n\right)\right), \\ \dfrac{H^{n+1} - H^{n+1/2}}{\Delta t/2} &= -S^T E^{n+1}, \end{cases} \tag{3.6}$$

where $\Delta t = t_{n+1} - t_n$ denotes the time step size and upper indices refer to time levels, as usual. This method has consistency two, is explicit in S, conditionally stable with a critical time step size proportional to h^{-1}, determined by the smallest grid element, [Botchev & Verwer 2009]

$$\Delta t \leq \frac{2}{\sqrt{\rho\left(SS^T\right)}}, \tag{3.7}$$

with strict inequality for zero conduction ($D = 0$) and where ρ denotes the spectral radius. Hence DG applied with its attractive feature of local grid refinement may lead to unduly step size restrictions.

3.2.2 The second-order Crank-Nicolson scheme (CN2)

An alternative to (3.6) is the second order, unconditionally stable Crank-Nicolson method that we write in the three-stage form

$$\begin{cases} \dfrac{H^{n+1/2}-H^n}{\Delta t/2} &= -S^T E^n, \\ \dfrac{E^{n+1}-E^n}{\Delta t} &= \dfrac{1}{2}S\left(H^{n+1}+H^n\right)-\dfrac{1}{2}D\left(E^{n+1}+E^n\right)+\dfrac{1}{2}\left(j^s\left(t_{n+1}\right)+j^s\left(t_n\right)\right), \\ \dfrac{H^{n+1}-H^{n+1/2}}{\Delta t/2} &= -S^T E^{n+1}, \end{cases} \tag{3.8}$$

which only differs in the middle stage in the time level for H. For consistency and stability we refer to [Verwer & Botchev 2009]. The expense for the implicit computation is too large to consider (3.8) as an attractive alternative to (3.6), especially in 3D (see e.g. [Verwer & Botchev 2009]).

If the region of refinement is of small size relatively to the computational domain, the unduly step size restriction of (3.6) and the overhead of (3.8) can be overcome by blending the two methods yielding locally implicit approaches where only variables associated to the smallest grid elements are implicitly treated.

3.3 The locally implicit method from [Piperno 2006]

The implicit-explicit method proposed in [Piperno 2006] is based on the following component splitting. The set of DG grid elements is divided into two subsets, one made of the smallest elements for implicit treatment and its complementary set for explicit treatment. Accordingly to this subdivision, the unknowns E and H are reordered as

$$E=\begin{pmatrix} E_e \\ E_i \end{pmatrix} \quad \text{and} \quad H=\begin{pmatrix} H_e \\ H_i \end{pmatrix}, \tag{3.9}$$

where the indices i and e are associated to the elements of the subsets treated implicitly and explicitly, respectively. In line with this splitting the semi-discrete curl operator S, the conduction matrix D and the source terms j^s are written as

$$S=\begin{pmatrix} S_e & -A_{ei} \\ -A_{ie} & S_i \end{pmatrix}, \quad D=\begin{pmatrix} D_e & 0 \\ 0 & D_i \end{pmatrix}, \quad j^s=\begin{pmatrix} j_e^s \\ j_i^s \end{pmatrix}. \tag{3.10}$$

Inserting this splitting into the semi-discrete DG Maxwell system (3.4) we obtain the system of ODEs

$$\begin{cases} \partial_t E_e &= S_e H_e - A_{ei} H_i - D_e E_e + j_e^s(t), \\ \partial_t E_i &= S_i H_i - A_{ie} H_e - D_i E_i + j_i^s(t), \\ \partial_t H_e &= -S_e^T E_e + A_{ie}^T E_i, \\ \partial_t H_i &= -S_i^T E_i + A_{ei}^T E_e. \end{cases} \tag{3.11}$$

The locally implicit time integration scheme from [Piperno 2006] is a blend of LF2 (3.6) and CN2 (3.8) applied to (3.11). It reads

$$\begin{cases} \dfrac{H_e^{n+1/2}-H_e^n}{\Delta t/2} &= -S_e^T E_e^n + A_{ie}^T E_i^n, \\ \dfrac{E_e^{n+1/2}-E_e^n}{\Delta t/2} &= S_e H_e^{n+1/2} - A_{ei} H_i^n - D_e E_e^n + j_e^s(t_n), \end{cases}$$

$$\begin{cases} \dfrac{E_i^{n+1}-E_i^n}{\Delta t} &= S_i\left(\dfrac{H_i^{n+1}+H_i^n}{2}\right) - A_{ie}H_e^{n+1/2} - D_i\left(\dfrac{E_i^{n+1}+E_i^n}{2}\right) \\ &+ \dfrac{j_i^s(t_{n+1})+j_i^s(t_n)}{2}, \\ \dfrac{H_i^{n+1}-H_i^n}{\Delta t} &= -S_i^T\left(\dfrac{E_i^{n+1}+E_i^n}{2}\right) + A_{ei}^T E_e^{n+1/2}, \end{cases} \tag{3.12}$$

$$\begin{cases} \dfrac{E_e^{n+1}-E_e^{n+1/2}}{\Delta t/2} &= S_e H_e^{n+1/2} - A_{ei} H_i^{n+1} - D_e E_e^{n+1} + j_e^s(t_{n+1}), \\ \dfrac{H_e^{n+1}-H_e^{n+1/2}}{\Delta t/2} &= -S_e^T E_e^{n+1} + A_{ie}^T E_i^{n+1}. \end{cases}$$

3.3.1 Computational work

Note that for $n \geq 1$ the derivative evaluations of the second explicit advance in the third block of (3.12) can be copied to the first explicit advance in the first block at the next time step. Furthermore from the second block of (3.12) we write

$$\begin{aligned} \left(I+\frac{\Delta t}{2}D_i\right)E_i^{n+1} &= \left(I-\frac{\Delta t}{2}D_i\right)E_i^n + \frac{\Delta t}{2}S_i H_i^{n+1} + \frac{\Delta t}{2}S_i H_i^n - \Delta t\, A_{ie}H_e^{n+1/2} \\ &+ \frac{\Delta t}{2}\left(j_i^s(t_n+1)+j_i^s(t_n)\right), \end{aligned} \tag{3.13}$$

and

$$H_i^{n+1} = H_i^n - \frac{\Delta t}{2}S_i^T\left(E_i^{n+1}+E_i^n\right) + \Delta t\, A_{ei}^T E_e^{n+1/2}. \tag{3.14}$$

Then by multiplying the equation (3.14) by S_i and inserting the obtained result in (3.13) we get

$$\mathscr{M}_1 E_i^{n+1} = b_i^{n+1}, \tag{3.15}$$

where

$$\begin{aligned} \mathscr{M}_1 &= I + \frac{\Delta t}{2}D_i + \frac{\Delta t^2}{4}S_i S_i^T, \\ b_i^{n+1} &= \left(I - \frac{\Delta t}{2}D_i - \frac{\Delta t^2}{4}S_i S_i^T\right)E_i^n + \Delta t\, S_i H_i^n + \frac{\Delta t^2}{2}S_i A_{ei}^T E_e^{n+1/2} - \Delta t\, A_{ie}H_e^{n+1/2} \\ &+ \frac{\Delta t}{2}\left(j_i^s(t_{n+1})+j_i^s(t_n)\right). \end{aligned} \tag{3.16}$$

Thus E_i^{n+1} can be obtained from the linear system (3.15) and successively H_i^{n+1} from (3.14). Note that $\mathscr{M}_1$ is a square symmetric positive definite matrix with dimension the length of E_i. Consequently if the region of local refinement is small relative to the computational domain, the workload induced by the implicit computation will be also small. Finally note that the evaluation of E_e^{n+1} in the third block of (3.12) is implicit in the conduction matrix D_e, but since D_e is diagonal this entails no additional overhead.

3.3.2 Stability and conservation properties

For the stability analysis of this method we refer to [Dolean *et al.* 2010]. The proof is based on the conservation of a quadratic form of the numerical unknowns E_e^n, E_i^n, H_e^n and H_i^n, previously established in [Piperno 2006]. The authors of [Dolean *et al.* 2010] show that under a condition on the time step size this quadratic form is positive definite and thus represents a discrete form of the electromagnetic energy. Consequently with the non-dissipative nature of the method they can conclude that this condition is sufficient for the stability of the locally implicit time integration scheme.

3.3.3 Convergence

In this section we are interested in the PDE convergence of method (3.12). More precisely, we will examine whether the method retains its second-order ODE convergence under stable simultaneous space-time grid refinement $\Delta t \backsim h$, $h \to 0$ towards the exact PDE solution. This is not a priori clear due to the component splitting which can introduce order reduction through error constants which grow with h^{-1}, for $h \to 0$.

This section is organized in four subsections. In Subsection 3.3.3.1 we are interested in the behavior of the matrices in (3.12) for $h \to 0$. This is an essential point for convergence analysis because some of these matrices can lie at the origin of order reduction. In Subsection 3.3.3.2 we will eliminate the intermediate values of (3.12) to get an equivalent one step formula from t_n to t_{n+1} that we will use for our convergence analysis. In Subsection 3.3.3.3 we will introduce the perturbed method obtained by substituting the exact PDE solution restricted to the assumed space grid into (3.11), and defects (space-time truncation errors) obtained by substituting this exact PDE solution into the equivalent one step formula of our method. In Subsection 3.3.3.4 we will define the common one-step recurrence relation for the global error. In Subsection 3.3.3.5 we will point out the order reduction mentioned above. Finally in Subsection 3.3.3.6 we will see that this order reduction, affecting the local error, may (partly) cancel in the transition from the local to the global error.

3.3.3.1 Matrix behavior for $h \to 0$

Let us consider the general case of dimension d ($d = 1$, 2 or 3). First we investigate the behavior of the matrices in the formulation with the mass matrices. Thereafter we will be able to deduce the behavior of the matrices in (3.12).

We reintroduce the notation with a tilde for the elements involved in the formulation without mass matrix (see (3.2)) in order to avoid confusion. The specific meaning of the block-entries of the different matrices involved in the formulation with mass matrices can be found in [Dolean *et al.* 2010].

First we observe that the mass matrices are only composed of volumic terms, hence we have

$$M^{\varepsilon},\ M^{\mu} \backsim h^{d},\ \text{for } h \to 0. \tag{3.17}$$

Thus

$$L_{M^{\varepsilon}},\ L^{T}_{M^{\varepsilon}},\ L_{M^{\mu}} \text{ and } L^{T}_{M^{\mu}} \backsim h^{\frac{d}{2}},\ \text{for } h \to 0. \tag{3.18}$$

The matrices S_e and S_i are composed of volumic and surfacic terms, hence

$$S_e \text{ and } S_i \backsim h^{d-1},\ \text{for } h \to 0. \tag{3.19}$$

The matrices A_{ei} and A_{ie} represent surfacic terms (interface matrices). Hence

$$A_{ei} \text{ and } A_{ie} \backsim h^{d-1},\ \text{for } h \to 0 \tag{3.20}$$

and from the block form of the matrix S (see (3.10)) we deduce with (3.19) and (3.20) that

$$S \backsim h^{d-1},\ \text{for } h \to 0. \tag{3.21}$$

From (3.3) we get

$$\widetilde{S} = L^{-1}_{M^{\varepsilon}}\ S\ \left(L^{-1}_{M^{\mu}}\right)^{T}, \tag{3.22}$$

then with the behaviors above we deduce that for $h \to 0$

$$\widetilde{S}_e,\ \widetilde{S}_i,\ \widetilde{A}_{ei},\ \widetilde{A}_{ie} = \mathscr{O}\left(h^{-1}\right), \tag{3.23}$$

and we have the expected behavior (3.5) for $\widetilde{S}$.

3.3.3.2 Elimination of intermediates values

First we treat H_e. From the first and last equations of (3.12) we get

$$\begin{aligned} H_e^{n+\frac{1}{2}} &= H_e^{n} - \frac{\Delta t}{2} S_e^{T} E_e^{n} + \frac{\Delta t}{2} A_{ie}^{T} E_i^{n}, \\ H_e^{n+\frac{1}{2}} &= H_e^{n+1} + \frac{\Delta t}{2} S_e^{T} E_e^{n+1} - \frac{\Delta t}{2} A_{ie}^{T} E_i^{n+1}. \end{aligned} \tag{3.24}$$

Inserting the first equation of (3.24) into the last equation of (3.12) yields

$$H_e^{n+1} = H_e^{n} - \frac{\Delta t}{2} S_e^{T}\left(E_e^{n} + E_e^{n+1}\right) + \frac{\Delta t}{2} A_{ie}^{T}\left(E_i^{n} + E_i^{n+1}\right). \tag{3.25}$$

Next we treat E_e. From the second and fifth equations of (3.12) we get

$$\begin{aligned} E_e^{n+\frac{1}{2}} &= E_e^{n} + \frac{\Delta t}{2} S_e H_e^{n+\frac{1}{2}} - \frac{\Delta t}{2} A_{ei} H_i^{n} - \frac{\Delta t}{2} D_e E_e^{n} + \frac{\Delta t}{2} j_e^{s}(t_n), \\ E_e^{n+\frac{1}{2}} &= E_e^{n+1} - \frac{\Delta t}{2} S_e H_e^{n+\frac{1}{2}} + \frac{\Delta t}{2} A_{ei} H_i^{n+1} + \frac{\Delta t}{2} D_e E_e^{n+1} - \frac{\Delta t}{2} j_e^{s}(t_{n+1}). \end{aligned} \tag{3.26}$$

Inserting the first equation of (3.26) and half of each expression of (3.24) for $H_e^{n+\frac{1}{2}}$ into the fifth equation of (3.12) yields

$$\begin{aligned} E_e^{n+1} &= E_e^n + \frac{\Delta t}{2} S_e \left(H_e^n + H_e^{n+1}\right) - \frac{\Delta t}{2} A_{ei} \left(H_i^n + H_i^{n+1}\right) - \frac{\Delta t}{2} D_e \left(E_e^n + E_e^{n+1}\right) \\ &+ \frac{\Delta t}{2} \left(j_e^s(t_n) + j_e^s(t_{n+1})\right) + \frac{\Delta t^2}{4} S_e \left[\left(-S_e^T E_e^n + A_{ie}^T E_i^n\right) - \left(S_e^T E_e^{n+1} - A_{ie}^T E_i^{n+1}\right)\right]. \end{aligned} \tag{3.27}$$

Now we consider H_i. Inserting half of each expression of (3.26) for $E_e^{n+\frac{1}{2}}$ in the fourth equation of (3.12) gives

$$\begin{aligned} H_i^{n+1} &= H_i^n - \frac{\Delta t}{2} S_i^T \left(E_i^n + E_i^{n+1}\right) + \frac{\Delta t}{2} A_{ei}^T \left(E_e^n + E_e^{n+1}\right) + \frac{\Delta t^2}{4} A_{ei}^T A_{ei} \left(H_i^{n+1} - H_i^n\right) \\ &+ \frac{\Delta t^2}{4} A_{ei}^T \left[D_e \left(E_e^{n+1} - E_e^n\right) + \left(j_e^s(t_n) + j_e^s(t_{n+1})\right)\right]. \end{aligned} \tag{3.28}$$

Finally we treat E_i. Inserting half of each expression of (3.24) for $H_e^{n+\frac{1}{2}}$ in the third equation of (3.12) yields

$$\begin{aligned} E_i^{n+1} &= E_i^n + \frac{\Delta t}{2} S_i \left(H_i^n + H_i^{n+1}\right) - \frac{\Delta t}{2} A_{ie} \left(H_e^n + H_e^{n+1}\right) - \frac{\Delta t}{2} D_i \left(E_i^n + E_i^{n+1}\right) \\ &+ \frac{\Delta t}{2} \left(j_i^s(t_n) + j_i^s(t_{n+1})\right) - \frac{\Delta t^2}{4} A_{ie} \left[\left(-S_e^T E_e^n + A_{ie}^T E_i^n\right) - \left(S_e^T E_e^{n+1} - A_{ie}^T E_i^{n+1}\right)\right]. \end{aligned} \tag{3.29}$$

The equivalent method of (3.12) with its intermediate values eliminated thus reads

$$\begin{aligned} E_e^{n+1} &= E_e^n + \frac{\Delta t}{2} S_e \left(H_e^n + H_e^{n+1}\right) - \frac{\Delta t}{2} A_{ei} \left(H_i^n + H_i^{n+1}\right) - \frac{\Delta t}{2} D_e \left(E_e^n + E_e^{n+1}\right) \\ &+ \frac{\Delta t}{2} \left(j_e^s(t_n) + j_e^s(t_{n+1})\right) + \frac{\Delta t^2}{4} S_e \left[\left(-S_e^T E_e^n + A_{ie}^T E_i^n\right) - \left(S_e^T E_e^{n+1} - A_{ie}^T E_i^{n+1}\right)\right], \\ E_i^{n+1} &= E_i^n + \frac{\Delta t}{2} S_i \left(H_i^n + H_i^{n+1}\right) - \frac{\Delta t}{2} A_{ie} \left(H_e^n + H_e^{n+1}\right) - \frac{\Delta t}{2} D_i \left(E_i^n + E_i^{n+1}\right) \\ &+ \frac{\Delta t}{2} \left(j_i^s(t_n) + j_i^s(t_{n+1})\right) - \frac{\Delta t^2}{4} A_{ie} \left[\left(-S_e^T E_e^n + A_{ie}^T E_i^n\right) - \left(S_e^T E_e^{n+1} - A_{ie}^T E_i^{n+1}\right)\right], \\ H_e^{n+1} &= H_e^n - \frac{\Delta t}{2} S_e^T \left(E_e^n + E_e^{n+1}\right) + \frac{\Delta t}{2} A_{ie}^T \left(E_i^n + E_i^{n+1}\right), \\ H_i^{n+1} &= H_i^n - \frac{\Delta t}{2} S_i^T \left(E_i^n + E_i^{n+1}\right) + \frac{\Delta t}{2} A_{ei}^T \left(E_e^n + E_e^{n+1}\right) + \frac{\Delta t^2}{4} A_{ei}^T A_{ei} \left(H_i^{n+1} - H_i^n\right) \\ &+ \frac{\Delta t^2}{4} A_{ei}^T \left[D_e \left(E_e^{n+1} - E_e^n\right) + \left(j_e^s(t_n) + j_e^s(t_{n+1})\right)\right]. \end{aligned} \tag{3.30}$$

3.3.3.3 *The perturbed method and defects for the PDE solution*

Let $E_e^h(t)$ denote at time t the exact solution of the PDE problem restricted to the assumed space grid that we have approximated with the semi-discrete system (3.11). $E_e^h(t_n)$ thus represents the vector

that is approximated by E_e^n. Assume the same notation for E_i, H_e and H_i. Substituting $E_e^h(t)$, $E_i^h(t)$, $H_e^h(t)$ and $H_i^h(t)$ into (3.11) reveals the spatial truncation errors

$$\begin{cases} \frac{d}{dt}E_e^h(t) &= S_e H_e^h(t) - A_{ei} H_i^h(t) - D_e E_e^h(t) + j_e^s(t) + \zeta_e^E(t), \\ \frac{d}{dt}E_i^h(t) &= S_i H_i^h(t) - A_{ie} H_e^h(t) - D_i E_i^h(t) + j_i^s(t) + \zeta_i^E(t), \\ \frac{d}{dt}H_e^h(t) &= -S_e^T E_e^h(t) + A_{ie}^T E_i^h(t) + \zeta_e^H(t), \\ \frac{d}{dt}H_i^h(t) &= -S_i^T E_i^h + A_{ei}^T E_e^h(t) + \zeta_i^H(t), \end{cases} \tag{3.31}$$

where $\zeta_e^E(t)$, $\zeta_i^E(t)$, $\zeta_e^H(t)$ and $\zeta_i^H(t)$ denote the spatial truncation errors.

Substituting $E_e^h(t)$, $E_i^h(t)$, $H_e^h(t)$ and $H_i^h(t)$ into (3.30) reveals the defects for the PDE solution (space-time truncation errors) and gives what we call the perturbed method

$$\begin{aligned} E_e^h(t_{n+1}) &= E_e^h(t_n) + \frac{\Delta t}{2} S_e \left(H_e^h(t_n) + H_e^h(t_{n+1}) \right) - \frac{\Delta t}{2} A_{ei} \left(H_i^h(t_n) + H_i^h(t_{n+1}) \right) \\ &- \frac{\Delta t}{2} D_e \left(E_e^h(t_n) + E_e^h(t_{n+1}) \right) + \frac{\Delta t}{2} \left(j_e^s(t_n) + j_e^s(t_{n+1}) \right) \\ &+ \frac{\Delta t^2}{4} S_e \left[\left(-S_e^T E_e^h(t_n) + A_{ie}^T E_i^h(t_n) \right) - \left(S_e^T E_e^h(t_{n+1}) - A_{ie}^T E_i^h(t_{n+1}) \right) \right] \\ &+ \Delta t \ \delta_{e,n}^E, \\ E_i^h(t_{n+1}) &= E_i^h(t_n) + \frac{\Delta t}{2} S_i \left(H_i^h(t_n) + H_i^h(t_{n+1}) \right) - \frac{\Delta t}{2} A_{ie} \left(H_e^h(t_n) + H_e^h(t_{n+1}) \right) \\ &- \frac{\Delta t}{2} D_i \left(E_i^h(t_n) + E_i^h(t_{n+1}) \right) + \frac{\Delta t}{2} \left(j_i^s(t_n) + j_i^s(t_{n+1}) \right) \\ &- \frac{\Delta t^2}{4} A_{ie} \left[\left(-S_e^T E_e^h(t_n) + A_{ie}^T E_i^h(t_n) \right) - \left(S_e^T E_e^h(t_{n+1}) - A_{ie}^T E_i^h(t_{n+1}) \right) \right] \\ &+ \Delta t \ \delta_{i,n}^E, \\ H_e^h(t_{n+1}) &= H_e^h(t_n) - \frac{\Delta t}{2} S_e^T \left(E_e^h(t_n) + E_e^h(t_{n+1}) \right) + \frac{\Delta t}{2} A_{ie}^T \left(E_i^h(t_n) + E_i^h(t_{n+1}) \right) \\ &+ \Delta t \ \delta_{e,n}^H, \\ H_i^h(t_{n+1}) &= H_i^h(t_n) - \frac{\Delta t}{2} S_i^T \left(E_i^h(t_n) + E_i^h(t_{n+1}) \right) + \frac{\Delta t}{2} A_{ei}^T \left(E_e^h(t_n) + E_e^h(t_{n+1}) \right) \\ &+ \frac{\Delta t^2}{4} A_{ei}^T A_{ei} \left(H_i^h(t_{n+1}) - H_i^h(t_n) \right) \\ &+ \frac{\Delta t^2}{4} A_{ei}^T \left[D_e \left(E_e^h(t_{n+1}) - E_e^h(t_n) \right) + \left(j_e^s(t_n) + j_e^s(t_{n+1}) \right) \right] + \Delta t \ \delta_{i,n}^H, \end{aligned} \tag{3.32}$$

where $\delta_{e,n}^E$, $\delta_{i,n}^E$, $\delta_{e,n}^H$ and $\delta_{i,n}^H$ denote the defects for the PDE solution.

From (3.31) and (3.32) we write

$$\begin{aligned}
\delta^E_{e,n} &= \delta_{E^h_e} + \frac{\Delta t}{4} S_e \frac{d}{dt}\left(H^h_e(t_{n+1}) - H^h_e(t_n)\right) \\
&+ \frac{1}{2}\left(\zeta^E_e(t_n) + \zeta^E_e(t_{n+1})\right) + \frac{\Delta t}{4} S_e\left(\zeta^H_e(t_n) + \zeta^H_e(t_{n+1})\right), \\
\delta^E_{i,n} &= \delta_{E^h_i} - \frac{\Delta t}{4} A_{ie} \frac{d}{dt}\left(H^h_e(t_{n+1}) - H^h_e(t_n)\right) \\
&+ \frac{1}{2}\left(\zeta^E_i(t_n) + \zeta^E_i(t_{n+1})\right) - \frac{\Delta t}{4} A_{ie}\left(\zeta^H_e(t_n) + \zeta^H_e(t_{n+1})\right), \\
\delta^H_{e,n} &= \delta_{H^h_e} + \frac{1}{2}\left(\zeta^H_e(t_n) + \zeta^H_e(t_{n+1})\right) \\
\delta^H_{i,n} &= \delta_{H^h_i} - \frac{\Delta t}{4} A^T_{ei} S_e\left(H^h_e(t_{n+1}) - H^h_e(t_n)\right) + \frac{\Delta t}{4} A^T_{ei} \frac{d}{dt}\left(E^h_e(t_{n+1}) - E^h_e(t_n)\right) \\
&+ \frac{1}{2}\left(\zeta^H_i(t_n) + \zeta^H_i(t_{n+1})\right) + \frac{\Delta t}{4} A^T_{ei}\left(\zeta^E_e(t_n) + \zeta^E_e(t_{n+1})\right),
\end{aligned} \tag{3.33}$$

where $\delta_{E^h_e}$ denotes the implicit trapezoidal rule defect (see [Verwer 2010]) for variable E^h_e (similarly for E^h_i, H^h_e and H^h_i), i.e.

$$\delta_{E^h_e}(t) = \frac{E^h_e(t+\Delta t) - E^h_e(t)}{\Delta t} - \frac{1}{2}\frac{d}{dt}\left(E^h_e(t+\Delta t) + E^h_e(t)\right). \tag{3.34}$$

3.3.3.4 *The error scheme*

Let $\varepsilon^E_{e,n} = E^h_e(t_n) - E^n_e$ denote the global error (similarly we introduce $\varepsilon^E_{i,n}$, $\varepsilon^H_{e,n}$ and $\varepsilon^H_{i,n}$). Substracting (3.30) and (3.32) we obtain the error scheme

$$\begin{aligned}
\varepsilon^E_{e,n+1} &= \varepsilon^E_{e,n} + \frac{\Delta t}{2} S_e\left(\varepsilon^H_{e,n} + \varepsilon^H_{e,n+1}\right) - \frac{\Delta t}{2} A_{ei}\left(\varepsilon^H_{i,n} + \varepsilon^H_{i,n+1}\right) - \frac{\Delta t}{2} D_e\left(\varepsilon^E_{e,n} + \varepsilon^E_{e,n+1}\right) \\
&+ \frac{\Delta t^2}{4} S_e\left[\left(-S^T_e \varepsilon^E_{e,n} + A^T_{ie}\varepsilon^E_{i,n}\right) - \left(S^T_e \varepsilon^E_{e,n+1} - A^T_{ie}\varepsilon^E_{i,n+1}\right)\right] + \Delta t\ \delta^E_{e,n}, \\
\varepsilon^E_{i,n+1} &= \varepsilon^E_{i,n} + \frac{\Delta t}{2} S_i\left(\varepsilon^H_{i,n} + \varepsilon^H_{i,n+1}\right) - \frac{\Delta t}{2} A_{ie}\left(\varepsilon^H_{e,n} + \varepsilon^H_{e,n+1}\right) - \frac{\Delta t}{2} D_i\left(\varepsilon^E_{i,n} + \varepsilon^E_{i,n+1}\right) \\
&- \frac{\Delta t^2}{4} A_{ie}\left[\left(-S^T_e \varepsilon^E_{e,n} + A^T_{ie}\varepsilon^E_{i,n}\right) - \left(S^T_e \varepsilon^E_{e,n+1} - A^T_{ie}\varepsilon^E_{i,n+1}\right)\right] + \Delta t\ \delta^E_{i,n}, \\
\varepsilon^H_{e,n+1} &= \varepsilon^H_{e,n} - \frac{\Delta t}{2} S^T_e\left(\varepsilon^E_{e,n} + \varepsilon^E_{e,n+1}\right) + \frac{\Delta t}{2} A^T_{ie}\left(\varepsilon^E_{i,n} + \varepsilon^E_{i,n+1}\right) + \Delta t\ \delta^H_{e,n}, \\
\varepsilon^H_{i,n+1} &= \varepsilon^H_{i,n} - \frac{\Delta t}{2} S^T_i\left(\varepsilon^E_{i,n} + \varepsilon^E_{i,n+1}\right) + \frac{\Delta t}{2} A^T_{ei}\left(\varepsilon^E_{e,n} + \varepsilon^E_{e,n+1}\right) \\
&+ \frac{\Delta t^2}{4} A^T_{ei} A_{ei}\left(\varepsilon^H_{i,n+1} - \varepsilon^H_{i,n}\right) + \frac{\Delta t^2}{4} A^T_{ei} D_e\left(\varepsilon^E_{e,n+1} - \varepsilon^E_{e,n}\right) + \Delta t\ \delta^H_{i,n}.
\end{aligned} \tag{3.35}$$

Let

$$\varepsilon_n = \begin{pmatrix} \varepsilon^E_{e,n} \\ \varepsilon^E_{i,n} \\ \varepsilon^H_{e,n} \\ \varepsilon^H_{i,n} \end{pmatrix} \quad \text{and} \quad \delta_n = \begin{pmatrix} \delta^E_{e,n} \\ \delta^E_{i,n} \\ \delta^H_{e,n} \\ \delta^H_{i,n} \end{pmatrix}, \tag{3.36}$$

then from (3.35) we can write the global error in a more compact form (one-step recurrence relation)

$$\varepsilon_{n+1} = R\varepsilon_n + \Delta t \rho_n, \quad R = R_L^{-1} R_R, \quad \rho_n = R_L^{-1}\delta_n, \tag{3.37}$$

where

$$R_L = \begin{pmatrix} I + \frac{\Delta t}{2} D_e - \frac{\Delta t^2}{4} S_e S_e^T & \frac{\Delta t^2}{4} S_e A_{ie}^T & -\frac{\Delta t}{2} S_e & \frac{\Delta t}{2} A_{ei} \\ \frac{\Delta t^2}{4} A_{ie} S_e^T & I + \frac{\Delta t}{2} D_i - \frac{\Delta t^2}{4} A_{ie} A_{ie}^T & \frac{\Delta t}{2} A_{ie} & -\frac{\Delta t}{2} S_i \\ \frac{\Delta t}{2} S_e^T & -\frac{\Delta t}{2} A_{ie}^T & I & 0 \\ -\frac{\Delta t}{2} A_{ei}^T - \frac{\Delta t^2}{4} A_{ei}^T D_e & \frac{\Delta t}{2} S_i^T & 0 & I - \frac{\Delta t^2}{4} A_{ei}^T A_{ei} \end{pmatrix}, \tag{3.38}$$

$$R_R = \begin{pmatrix} I - \frac{\Delta t}{2} D_e - \frac{\Delta t^2}{4} S_e S_e^T & \frac{\Delta t^2}{4} S_e A_{ie}^T & \frac{\Delta t}{2} S_e & -\frac{\Delta t}{2} A_{ei} \\ \frac{\Delta t^2}{4} A_{ie} S_e^T & I - \frac{\Delta t}{2} D_i - \frac{\Delta t^2}{4} A_{ie} A_{ie}^T & -\frac{\Delta t}{2} A_{ie} & \frac{\Delta t}{2} S_i \\ -\frac{\Delta t}{2} S_e^T & \frac{\Delta t}{2} A_{ie}^T & I & 0 \\ \frac{\Delta t}{2} A_{ei}^T - \frac{\Delta t^2}{4} A_{ei}^T D_e & -\frac{\Delta t}{2} S_i^T & 0 & I - \frac{\Delta t^2}{4} A_{ei}^T A_{ei} \end{pmatrix}, \tag{3.39}$$

and ε_n, $\Delta t \rho_n$ and δ_n are respectively the (space-time) global, local and truncation errors.

Note that the recursion (3.37) has the standard form (see e.g. [Hundsdorfer & Verwer 2003]) for the convergence analysis of one-step integration methods. It transfers local errors to the global error, essentially by adding all local errors. Indeed, for a given time interval $[0, T]$ we make the usual stability hypothesis

$$\|R^n\| \leq K \quad \text{for } h \to 0 \text{ and } n \geq 0,\ n\Delta t \leq T. \tag{3.40}$$

On the other hand the elaboration of the error recursion (3.37) gives

$$\varepsilon_n = R^n \varepsilon_0 + R^{n-1} \Delta t \rho_0 + \ldots + R \Delta t \rho_{n-2} + \Delta t \rho_{n-1}, \tag{3.41}$$

which leads directly (with (3.40)) to

$$\|\varepsilon_n\| \leq K \|\varepsilon_0\| + K \Delta t \sum_{j=0}^{n-1} \|\rho_j\| \quad \text{for } n\Delta t \leq T. \tag{3.42}$$

Recall that $\rho_j = R_L^{-1}\delta_j$ and because we assume stability we may consider R_L inversely bounded for $\Delta t \sim h$, $h \to 0$. Consequently if $\delta_j = \mathcal{O}(\Delta t^k)$ we have $\|\rho_j\| \leq C\Delta t^k$. Assuming $\varepsilon_0 = 0$, we deduce from (3.42) that

$$\|\varepsilon_n\| \leq \widetilde{K}\Delta t^k \quad \text{for } n\Delta t \leq T, \tag{3.43}$$

with constant $\widetilde{K} = KTC$.

3.3.3.5 *Error analysis*

We assume that the exact PDE solutions E_e^h, E_i^h, H_e^h and H_i^h are sufficiently differentiable. Then, Taylor expansion of the trapezoidal rule defect (3.34) at the symmetry point $t_{n+1/2}$ gives

$$\delta_{E_e^h}(t_n) = \sum_{k=2'} \frac{-k}{2^k(k+1)!}(\Delta t)^k E_e^{h(k+1)}, \tag{3.44}$$

where $k = 2'$ means even values for k only and $E_e^{h(k)}$ denotes the k-th derivative of $E_e^h(t)$ at time $t = t_{n+1/2}$. We obtain similar expressions for $\delta_{E_i^h}(t_n)$, $\delta_{H_e^h}(t_n)$ and $\delta_{H_i^h}(t_n)$. Note that these defects start with Δt^2 and the third solution derivative of the exact PDE solution (in the first section we have assumed that the exact PDE solution is three times differentiable). We conclude that the trapezoidal rule defects are $\mathcal{O}(\Delta t^2)$ for $\Delta t \sim h$ and $h \to 0$. Then, from (3.33), we write the truncation error as follows

$$\delta_n = b_n + \mathcal{O}\left(\Delta t^2\right), \tag{3.45}$$

where

$$b_n = \begin{pmatrix} b_{e,n}^E \\ b_{i,n}^E \\ b_{e,n}^H \\ b_{i,n}^H \end{pmatrix} = \begin{pmatrix} \frac{\Delta t}{4} S_e \frac{d}{dt}\left(H_e^h(t_{n+1}) - H_e^h(t_n)\right) \\ -\frac{\Delta t}{4} A_{ie} \frac{d}{dt}\left(H_e^h(t_{n+1}) - H_e^h(t_n)\right) \\ 0 \\ -\frac{\Delta t}{4} A_{ei}^T S_e \left(H_e^h(t_{n+1}) - H_e^h(t_n)\right) + \frac{\Delta t}{4} A_{ei}^T \frac{d}{dt}\left(E_e^h(t_{n+1}) - E_e^h(t_n)\right) \end{pmatrix}, \tag{3.46}$$

and $\mathcal{O}(\Delta t^2)$ contains the trapezoidal rule defects. Note that we have voluntarily omitted the spatial error parts contained in the ζ_e^E, ζ_i^E, ζ_e^H, ζ_i^H contributions from (3.33) because our interest lies in temporal convergence order. Further carrying these spatial error contributions only complicates the formulas and will not lead to different conclusions for the temporal errors.

Next we Taylor expand the components of b_n at the symmetry point $t_{n+1/2}$

$$\begin{aligned} b_{e,n}^E &= \frac{\Delta t}{4} S_e \sum_{k=1'} \frac{1}{2^{k-1}k!}(\Delta t)^k H_e^{h(k+1)}, \\ b_{i,n}^E &= -\frac{\Delta t}{4} A_{ie} \sum_{k=1'} \frac{1}{2^{k-1}k!}(\Delta t)^k H_e^{h(k+1)}, \\ b_{i,n}^H &= -\frac{\Delta t}{4} A_{ei}^T S_e \sum_{k=1'} \frac{1}{2^{k-1}k!}(\Delta t)^k H_e^{h(k)} + \frac{\Delta t}{4} A_{ei}^T \sum_{k=1'} \frac{1}{2^{k-1}k!}(\Delta t)^k E_e^{h(k+1)}, \end{aligned} \tag{3.47}$$

where $k = 1'$ means odd values for k only. For a fixed spatial dimension we find the expected second-order ODE convergence, since S_e, A_{ie} and A_{ei}^T are bounded for fixed dimension. Indeed, with (3.47) we have $b_n = \mathcal{O}(\Delta t^2)$ and because we assume stability we may consider R_L inversely bounded, consequently $\rho_n = R_L^{-1}\delta_n = \mathcal{O}(\Delta t^2)$ and we conclude that we have the second-order convergence for a fixed dimension.

Now we observe, with (3.23), that under stable simultaneous space-time refinement, $\Delta t \sim h$ and $h \to 0$, we might lose one unit of Δt in $b_{e,n}^E$ and $b_{i,n}^E$ (due to S_e and A_{ie}, respectively) and two units of Δt in $b_{i,n}^H$ (due to the product $A_{ei}^T S_e$ in the first terms). Then $b_n = \mathcal{O}(1)$, for $\Delta t \sim h$ and $h \to 0$ and with (3.43) we should expect a severe order reduction. However, as mentioned in [Verwer 2010], this result is based on standard local error analysis and in the transition from local to global errors it can happen that the order reduction for local errors is (partly) canceled. Often this cancellation can be shown to exist through a transformation of the global error recurrence to one by which we may gain one unit of Δt in the transformed local error.

3.3.3.6 A transformed global error recursion

The transformation used in [Verwer 2010] emanates from [Hundsdorfer & Verwer 2003], Lemma II.2.3. We write the latter for our one-step global error recursion (3.37) and stability assumption (3.40)

Lemma 3.3.1. *Suppose the local error $\Delta\rho_n$ can be written as*

$$\Delta t\,\rho_n = (I - R)\,\xi_n + \eta_n,$$

with $\|\xi_n\| \le C\Delta t^k$, $\|\eta_n\| \le C\Delta t^{k+1}$ and $\|\xi_{n+1} - \xi_n\| \le C\Delta t^{k+1}$ for all n. Then there is a constant $C' > 0$, depending on C, K and T, such that $\|\varepsilon_n\| \le C'\Delta t^k$ for $n\Delta t \le T$.

The proof can be found in [Hundsdorfer & Verwer 2003].

First we assume for $\Delta t \sim h$, $h \to 0$ that

$$A_{ei}^T S_e H_e^{h(1)} = \mathcal{O}\left(\Delta t^{-1}\right). \tag{3.48}$$

Consequently we get $b_n = \mathcal{O}(\Delta t)$ (in this case $b_{i,n}^H = \mathcal{O}(\Delta t)$, see (3.47)). With the above-mentioned Lemma we can assume that if the local error $\Delta t\rho_n$ allows a decomposition

$$\Delta t\,\rho_n = (I - R)\,\xi_n + \eta_n \tag{3.49}$$

such that $\xi_n = \mathcal{O}(\Delta t^2)$, $\eta_n = \mathcal{O}(\Delta t^3)$ for $\Delta t \sim h$, $h \to 0$, then we have the desired second-order convergence for ε_n. So we need to verify (3.49), or equivalently,

$$\Delta t\,\delta_n = (R_L - R_R)\,\xi_n + R_L\eta_n, \tag{3.50}$$

such that $\xi_n = \mathcal{O}(\Delta t^2)$, $\eta_n = \mathcal{O}(\Delta t^3)$ for $\Delta t \sim h$, $h \to 0$.

Now we deal with the condition $\eta_n = \mathcal{O}(\Delta t^3)$. Recall that $\delta_n = b_n + \mathcal{O}(\Delta t^2)$, then (3.50) can be written as

$$\Delta t\left(b_n + \mathcal{O}\left(\Delta t^2\right)\right) = (R_L - R_R)\,\xi_n + R_L\eta_n. \tag{3.51}$$

Furthermore, R_L is inversely bounded, then by assigning the $\mathscr{O}(\Delta t^2)$ terms present in (3.51) multiplied by $\Delta t\ R_L^{-1}$ for η_n (i.e. $\eta_n = \Delta t\ R_L^{-1}\mathscr{O}(\Delta t^2)$), we have $\eta_n = \mathscr{O}(\Delta t^3)$. Consequently, we deduce from Lemma 3.3.1 that we have the desired second-order convergence if a vector ξ_n exists such that $\xi_n = \mathscr{O}(\Delta t^2)$ for $\Delta t \sim h$, $h \to 0$ and

$$\Delta t b_n = (R_L - R_R)\,\xi_n, \text{ i.e.} \tag{3.52}$$

$$\begin{pmatrix} D_e & 0 & -S_e & A_{ei} \\ 0 & D_i & A_{ie} & -S_i \\ S_e^T & -A_{ie}^T & 0 & 0 \\ -A_{ei}^T & S_i^T & 0 & 0 \end{pmatrix} \begin{pmatrix} \xi_{e,n}^E \\ \xi_{i,n}^E \\ \xi_{e,n}^H \\ \xi_{i,n}^H \end{pmatrix} = \begin{pmatrix} b_{e,n}^E \\ b_{i,n}^E \\ b_{e,n}^H \\ b_{i,n}^H \end{pmatrix}. \tag{3.53}$$

Equivalently, we have second-order convergence if a vector $\xi_n = [(\xi_n^E)^T, (\xi_n^H)^T]^T$ exists such that $\xi_n = \mathscr{O}(\Delta t^2)$ for $\Delta t \sim h$, $h \to 0$ and

$$\begin{aligned} D\xi_n^E - S\xi_n^H &= b_n^E, \\ S^T\xi_n^E &= b_n^H, \end{aligned} \tag{3.54}$$

where $\xi_n^E = \left[(\xi_{e,n}^E)^T, (\xi_{i,n}^E)^T\right]^T$, $\xi_n^H = \left[(\xi_{e,n}^H)^T, (\xi_{i,n}^H)^T\right]^T$, $b_n^E = \left[(b_{e,n}^E)^T, (b_{i,n}^E)^T\right]^T$, $b_n^H = \left[(b_{e,n}^H)^T, (b_{i,n}^H)^T\right]^T$.

Now we will check the existence of a such vector ξ_n. At this stage of the derivation we must be careful because the matrix S is not necessarily a square matrix (in 2D this is not the case) and consequently S may no be invertible. More precisely, if we denote n_{dof} the number of degrees of freedom, and if we consider the two-dimensional transversal magnetic (TM) model, then the size of the matrix S is $n_{dof} \times 2n_{dof}$. That is why we now use the notion of pseudo inverse.

Definition 3.3.1. *Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ and A^+ the Moore-Penrose pseudo inverse of A which is a generalization of the inverse and exists for any $m \times n$ matrix. If A has full rank, then*

$$\begin{aligned} A^+ &= A^{-1} && (m = n) \\ A^+ &= A^T\left(AA^T\right)^{-1} && (m < n), \\ A^+ &= \left(A^TA\right)^{-1}A^T && (m > n), \end{aligned} \tag{3.55}$$

and the solution of $Ax = b$ is $x = A^+b$.

Assume the size of the matrix S is $m \times n$ with $m \leq n$ (the case $m > n$ can be treated similarly). With the second equation of (3.54) and the above definition we derive

$$\xi_n^E = \left(S^T\right)^+ b_n^H. \tag{3.56}$$

With (3.5) we have

$$\left(S^T\right)^+ = \left(SS^T\right)^{-1} S \sim h, \quad \text{for } h \to 0, \tag{3.57}$$

and recall that $b_{e,n}^H = 0$ and $b_{i,n}^H = \mathscr{O}(\Delta t)$ (due to the initial assumption (3.48)) for $\Delta t \sim h$, $h \to 0$. Then we conclude with (3.56) and (3.57) that for $\Delta t \sim h$, $h \to 0$

$$\xi_n^E = \mathscr{O}\left(\Delta t^2\right). \tag{3.58}$$

From the first equation of (3.54) we get

$$\xi_n^H = -S^+\left(b_n^E - D\xi_n^E\right), \tag{3.59}$$

and with (3.5) we get

$$S^+ = S^T\left(SS^T\right)^{-1} \sim h, \quad \text{for } h \to 0. \tag{3.60}$$

Recalling that $b_{e,n}^E = \mathscr{O}(\Delta t)$, $b_{i,n}^E = \mathscr{O}(\Delta t)$ and $\xi_n^E = \mathscr{O}(\Delta t^2)$, we conclude from (3.59) and (3.60) that for $\Delta t \sim h$, $h \to 0$

$$\text{if } D = 0 \text{ (no conduction term } -\sigma E) \text{ or } D = \mathscr{O}\left(h^k\right) \text{ with } k \geq -1, \text{ then } \xi_n^H = \mathscr{O}\left(\Delta t^2\right). \tag{3.61}$$

With (3.58) and (3.61) we conclude that $\xi_n = \left[(\xi_n^E)^T, (\xi_n^H)^T\right]^T = \mathscr{O}(\Delta t^2)$ for $\Delta t \sim h$, $h \to 0$ and consequently through the Lemma 3.3.1 we have the second-order convergence uniformly in h under the assumption (3.48). Note that if we do not assume (3.48) (i.e. $b_n = \mathscr{O}(1)$) a similar proof, based on the same Lemma, only guarantees the first-order convergence. We can now state the following theorem

Theorem 3.3.1. *Let $j^s(t) \in C^2[0,T]$ and suppose a Lax-Richtmyer stable space-time grid refinement $\Delta t \sim h$, $h \to 0$. On $[0,T]$, the approximations H_e^n, H_i^n, E_e^n and E_i^n resulting from method (3.12) then converge to $H_e^h(t)$, $H_i^h(t)$, $E_e^h(t)$ and $E_i^h(t)$*

(i) at least at first order,

(ii) at least at second order, if in addition $A_{ei}^T S_e H_e^{h(1)}(t) = \mathscr{O}(\Delta t^{-1})$ for $h \to 0$.

To sum up, we can guarantee at least the first-order convergence of method (3.12). As might be feared, component splitting can be detrimental to the temporal convergence order (order reduction). We have also put forward a sufficient condition (3.48) on the the exact solution of the PDE problem for second-order convergence. However it would have been better if this sufficient condition could be controlled through the source term, because in general the exact solution is of course not (a priori) known.

3.4 The locally implicit method from [Verwer 2010]

This implicit-explicit time integration method is also a blend of (3.6) and (3.8) applied to the generic semi-discrete Maxwell system (3.4)

$$\begin{cases} \dfrac{H^{n+1/2} - H^n}{\Delta t/2} &= -S^T E^n, \\ \dfrac{E^{n+1} - E^n}{\Delta t} &= S_0 H^{n+1/2} + \dfrac{1}{2} S_1\left(H^{n+1} + H^n\right) \\ &- \dfrac{1}{2} D\left(E^{n+1} + E^n\right) + \dfrac{1}{2}\left(j^s(t_{n+1}) + j^s(t_n)\right), \\ \dfrac{H^{n+1} - H^{n+1/2}}{\Delta t/2} &= -S^T E^{n+1}, \end{cases} \tag{3.62}$$

where $S = S_0 + S_1$ is a matrix splitting. The method is implicit in S_1 and explicit in S_0. For $S_0 = 0$ we recover (3.8) and for $S_1 = 0$ the method (3.6).

3.4.1 Computational work

First note that for $n \geq 1$ the third stage derivative computation can be copied to the first stage at the next time step. Furthermore from the second and the third stage of (3.62) we write

$$\begin{aligned} \left(I + \frac{\Delta t}{2} D\right) E^{n+1} &= \left(I - \frac{\Delta t}{2} D\right) E^n + \frac{\Delta t}{2} S_1 H^{n+1} + \frac{\Delta t}{2} S_1 H^n - \Delta t\, S_0 H^{n+1/2} \\ &+ \frac{\Delta t}{2} \left(j^s(t_{n+1}) + j^s(t_n)\right), \end{aligned} \tag{3.63}$$

and

$$H^{n+1} = H^{n+1/2} - \frac{\Delta t}{2} S^T E^{n+1}. \tag{3.64}$$

Then by multiplying the equation (3.64) by S_1 and inserting the obtained result in (3.63) we get

$$\mathscr{M}_2 E^{n+1} = b_{n+1} \tag{3.65}$$

where

$$\begin{aligned} \mathscr{M}_2 &= I + \frac{\Delta t}{2} D + \frac{\Delta t^2}{4} S_1 S^T, \\ b_{n+1} &= E^n + \Delta t\, S_0 H^{n+1/2} + \frac{\Delta t}{2} S_1 \left(H^{n+1/2} + H^n\right) - \frac{\Delta t}{2} D E^n + \frac{\Delta t}{2} \left(j^s(t_{n+1}) + j^s(t_n)\right). \end{aligned} \tag{3.66}$$

Thus E^{n+1} can be obtained from the linear system (3.65) and successively H^{n+1} from (3.64). The initial splitting adopted in [Verwer 2010] is defined as follow

$$S_1 = S S_H, \tag{3.67}$$

where S_H is a diagonal matrix of dimension the length of H with

$$(S_H)_{jj} = \begin{cases} 0, \text{ component } H_j \text{ of } H \text{ to be treated explicitly,} \\ 1, \text{ component } H_j \text{ of } H \text{ to be treated implicitly.} \end{cases} \tag{3.68}$$

With this definition the author of [Verwer 2010] notes that $S_1 S^T$ is symmetric, $S_1 S^T = S S_H S^T = S S_H S_H S^T = S_1 S_1^T$, which facilitates the resolution of (3.65). The matrix $\mathscr{M}_2$ is then given by

$$\mathscr{M}_2 = I + \frac{\Delta t}{2} D + \frac{\Delta t^2}{4} S_1 S_1^T. \tag{3.69}$$

For $S_1 = S$ we recover the matrix of the linear system of the fully implicit method (3.8), [Verwer & Botchev 2009]

$$\mathscr{M} = I + \frac{\Delta t}{2} D + \frac{\Delta t^2}{4} S S^T. \tag{3.70}$$

With the adopted splitting the matrix $\mathscr{M}_2$ will be significantly more sparse than without splitting. This gain of sparsity for the matrix $\mathscr{M}_2$ enables to solve the linear system at lower costs.

Note that the locally implicit methods (3.12) and (3.62) are very similar. Indeed, giving an equivalent formulation of (3.62) using $E^{n+1/2}$

$$\left\{\begin{array}{lcl} \dfrac{H^{n+1/2}-H^n}{\Delta t/2} & = & -S^T E^n, \\ \dfrac{E^{n+1/2}-E^n}{\Delta t} & = & \dfrac{1}{2}S_0 H^{n+1/2}+\dfrac{1}{2}S_1 H^n-\dfrac{1}{2}DE^n+\dfrac{1}{2}j^s(t_n), \\ \dfrac{E^{n+1}-E^{n+1/2}}{\Delta t} & = & \dfrac{1}{2}S_0 H^{n+1/2}+\dfrac{1}{2}S_1 H^{n+1}-\dfrac{1}{2}DE^{n+1}+\dfrac{1}{2}j^s(t_{n+1}), \\ \dfrac{H^{n+1}-H^{n+1/2}}{\Delta t/2} & = & -S^T E^{n+1}, \end{array}\right. \tag{3.71}$$

Adopting the subdivisions (3.9) - (3.10) we adjust the splitting (3.67) of the DG matrix S

$$S_1 = SS_h = \begin{pmatrix} S_e & -A_{ei} \\ -A_{ie} & S_i \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} 0 & -A_{ei} \\ 0 & S_i \end{pmatrix}, \tag{3.72}$$

Hence,

$$S_0 = S - S_1 = \begin{pmatrix} S_e & 0 \\ -A_{ie} & 0 \end{pmatrix}. \tag{3.73}$$

Adopting the subdivisions (3.9) and considering the previous splitting into (3.71) we obtain the following locally implicit time integration method

$$\left\{\begin{array}{lcl} \dfrac{H_e^{n+1/2}-H_e^n}{\Delta t/2} & = & -S_e^T E_e^n + A_{ie}^T E_i^n, \\ \dfrac{E_e^{n+1/2}-E_e^n}{\Delta t/2} & = & S_e H_e^{n+1/2} - A_{ei} H_i^n - D_e E_e^n + j_e^s(t_n), \end{array}\right.$$

$$\left\{\begin{array}{lcl} \dfrac{E_i^{n+1}-E_i^n}{\Delta t} & = & S_i\left(\dfrac{H_i^{n+1}+H_i^n}{2}\right) - A_{ie}H_e^{n+1/2} - D_i\left(\dfrac{E_i^{n+1}+E_i^n}{2}\right) \\ & + & \dfrac{j_i^s(t_{n+1})+j_i^s(t_n)}{2}, \\ \dfrac{H_i^{n+1}-H_i^n}{\Delta t} & = & -S_i^T\left(\dfrac{E_i^{n+1}+E_i^n}{2}\right) + A_{ei}^T\left(\dfrac{E_e^{n+1}+E_e^n}{2}\right), \end{array}\right. \tag{3.74}$$

$$\left\{\begin{array}{lcl} \dfrac{E_e^{n+1}-E_e^{n+1/2}}{\Delta t/2} & = & S_e H_e^{n+1/2} - A_{ei} H_i^{n+1} - D_e E_e^{n+1} + j_e^s(t_{n+1}), \\ \dfrac{H_e^{n+1}-H_e^{n+1/2}}{\Delta t/2} & = & -S_e^T E_e^{n+1} + A_{ie}^T E_i^{n+1}. \end{array}\right.$$

Comparing this method with the locally implicit method (3.12), the only difference appears in the second block for the definition of the variable H_i, which reads

$$A_{ei}^T\left(\frac{E_e^{n+1}+E_e^n}{2}\right) \text{ instead of } A_{ei}^T E_e^{n+1/2}. \tag{3.75}$$

Albeit this difference concerning the treatment of the interface matrix A_{ei}^T seems minor, we will see that the consequences on the error behavior are significant, notably regarding convergence to the exact underlying PDE solution; while the method (3.12) suffers from order reduction [Moya 2012] this is not the case with the method (3.62) [Verwer 2010]. However we remark that the matrix $\mathscr{M}_2$ has dimension the length of E, while the matrix $\mathscr{M}_1$ has dimension the length of E_i. Hence the workload for solving the linear system of the method (3.12) is smaller. More precisely the linear system (3.65) of the method (3.62) is equivalent to

$$\begin{pmatrix} I+\dfrac{\Delta t}{2}D_e+\dfrac{\Delta t^2}{4}A_{ei}^TA_{ei} & -\dfrac{\Delta t^2}{4}A_{ei}S_i^T \\ -\dfrac{\Delta t^2}{4}S_iA_{ei}^T & I+\dfrac{\Delta t}{2}D_i+\dfrac{\Delta t^2}{4}S_i^TS_i \end{pmatrix} \begin{pmatrix} E_e^{n+1} \\ E_i^{n+1} \end{pmatrix} = \begin{pmatrix} b_e^{n+1} \\ b_i^{n+1} \end{pmatrix}, \tag{3.76}$$

where b_e^{n+1}, b_i^{n+1} collect available vectors. By omitting the identity matrix in the above (2,2)-block matrix and in the matrix $\mathscr{M}_1$ we expect that the overhead for solving the linear system of the method (3.62) will not be excessive compared to the method (3.12), because $A_{ei}^TA_{ei}$, $A_{ei}S_i^T$ and $S_iA_{ei}^T$ are very sparse block diagonal matrices.

3.4.2 Stability and conservation properties

In [Verwer 2010], the author shows that the locally implicit method (3.62) for time integration of the semi-discrete Maxwell system (3.4) exactly conserve the following quadratic form of the numerical unknowns E^n and H^n

$$\mathscr{Q}_n = \|H^n\|_2^2 + \|E^n\|_2^2 - \frac{\Delta t^2}{4}\left\langle S_0S^TE^n,\, E^n\right\rangle, \tag{3.77}$$

for zero damping matrix D, and where $\langle\,\cdot\,,\,\cdot\,\rangle$ is the L^2 inner product and $\|\,\cdot\,\|_2$ the corresponding norm. By definition of $S_0 = S(I-S_H)$ the matrix S_0S^T is symmetric, since $S_0S^T = S(I-S_H)S^T = S(I-S_H)(I-S_H)S^T = S_0S_0^T$. Then we have

$$\mathscr{Q}_n = \|H^n\|_2^2 + \|E^n\|_2^2 - \frac{\Delta t^2}{4}\left\langle S_0S_0^TE^n,\, E^n\right\rangle. \tag{3.78}$$

The minus sign in front of the third term does not allow to conclude directly on the stability of the locally implicit time integration method (3.62). In [Verwer 2010] the author notes that for zero diagonal matrix S_H one recover the conservation property of the explicit scheme (3.6) (see [Botchev & Verwer 2009]) and since the entries of S_H are either zero or one, the deviation from the exact energy is reduced for the method (3.62) compared to the method (3.6).

In this section we are interested in the stability of the fully discrete locally implicit scheme (3.62). The derivations in the remainder of this section follow an energy approach which provides a rigorous criterion for stability. In Section 3.4.2.1 we exhibit a discrete energy, which is a quadratic form of the numerical unknowns. In Section 3.4.2.2 we show that the energy is a positive definite quadratic form. Finally, in Section 3.4.2.3 we prove that the energy is decreasing, which achieves the stability analysis.

3.4.2.1 Discrete energy

We define the discrete electromagnetic energy, denoted $\mathscr{E}_n$, as

$$\mathscr{E}_n \;=\; \frac{1}{2}\mathscr{Q}_n \;=\; \frac{1}{2}\left(\|H^n\|_2^2+\|E^n\|_2^2-\frac{\Delta t^2}{4}\left\langle S_0S_0^TE^n,\,E^n\right\rangle\right). \tag{3.79}$$

3.4.2.2 Positivity of the energy

We state a condition on the time step Δt such that $\mathscr{E}_n$ is a positive definite quadratic form.

Lemma 3.4.1. *The quadratic form $\mathscr{E}_n$ given by (3.79) is a positive definite quadratic form of the numerical unknowns E^n and H^n if*

$$\Delta t < \frac{2}{\sqrt{\rho\left(S_0S_0^T\right)}}, \tag{3.80}$$

where ρ denotes the spectral radius.

Proof. From (3.79) we have

$$\begin{aligned}\mathscr{E}_n &= \frac{1}{2}\left(\|H^n\|_2^2+\|E^n\|_2^2-\frac{\Delta t^2}{4}\left\langle\, S_0S_0^TE^n,\,E^n\right\rangle\right)\\ &= \frac{1}{2}\left(\|H^n\|_2^2+\|E^n\|_2^2-\frac{\Delta t^2}{4}\|S_0^TE^n\|_2^2\right).\end{aligned} \tag{3.81}$$

Note that

$$\|S_0^TE^n\|_2 \le \|S_0^T\|_2\|E^n\|_2 = \sqrt{\rho\left(S_0S_0^T\right)}\|E^n\|_2, \tag{3.82}$$

hence

$$\mathscr{E}_n \ge \frac{1}{2}\left(\|H^n\|_2^2+\left(1-\frac{\Delta t^2}{4}\rho\left(S_0S_0^T\right)\right)\|E^n\|_2^2\right), \tag{3.83}$$

allows to obtain that under the condition (3.80), $\mathscr{E}_n$ is a positive definite quadratic form of the numerical unknowns E^n and H^n. □

3.4.2.3 Variation of the energy

We now prove the following result.

Lemma 3.4.2. *The discrete energy (3.79) is decreasing so that $\mathscr{E}_n \le \mathscr{E}_0$.*

Proof. From the first and fourth equation of (3.62) we have

$$\begin{aligned}H^{n+1/2} &= H^n-\frac{\Delta t}{2}S^TE^n,\\ H^{n+1/2} &= H^{n+1}+\frac{\Delta t}{2}S^TE^{n+1}.\end{aligned} \tag{3.84}$$

Substitute into (3.62) half of $H^{n+1/2}$ from the first stage and half of $H^{n+1/2}$ from the third one into the second stage and the first expression of (3.84) into the third stage of (3.62). Together with the final expression for H^{n+1}, this yields

$$\begin{aligned} E^{n+1}-E^n &= \frac{\Delta t}{2}S\left(H^{n+1}+H^n\right)-\frac{\Delta t}{2}D\left(E^{n+1}+E^n\right)+\frac{\Delta t^2}{4}S_0S^T\left(E^{n+1}-E^n\right), \\ H^{n+1}-H^n &= -\frac{\Delta t}{2}S^T\left(E^{n+1}+E^n\right). \end{aligned} \tag{3.85}$$

Taking inner product with $E^{n+1}+E^n$ and $H^{n+1}+H^n$ yields, respectively,

$$\begin{aligned} \|E^{n+1}\|_2^2-\|E^n\|_2^2 &= \frac{\Delta t}{2}\left\langle S\left(H^{n+1}+H^n\right),\, E^{n+1}+E^n\right\rangle-\frac{\Delta t}{2}\left\langle D\left(E^{n+1}+E^n\right),\, E^{n+1}+E^n\right\rangle \\ &+ \frac{\Delta t^2}{4}\left\langle S_0S^T\left(E^{n+1}-E^n\right),\, E^{n+1}+E^n\right\rangle, \\ \|H^{n+1}\|_2^2-\|H^n\|_2^2 &= -\frac{\Delta t}{2}\left\langle S^T\left(E^{n+1}+E^n\right),\, H^{n+1}+H^n\right\rangle. \end{aligned} \tag{3.86}$$

We recall that $S_0S^T=S_0S_0^T$, hence

$$\begin{aligned} \left\langle S_0S^T\left(E^{n+1}-E^n\right),\, E^{n+1}+E^n\right\rangle &= \left\langle S_0S_0^T\left(E^{n+1}-E^n\right),\, E^{n+1}+E^n\right\rangle, \\ &= \left\langle S_0S_0^TE^{n+1},\, E^{n+1}\right\rangle-\left\langle S_0S_0^TE^n,\, E^n\right\rangle, \\ &= \left\langle S_0S^TE^{n+1},\, E^{n+1}\right\rangle-\left\langle S_0S^TE^n,\, E^n\right\rangle. \end{aligned} \tag{3.87}$$

Substituting this expression into the first equation of (3.86) and adding the two equations (3.86) yields

$$\begin{aligned} \frac{\mathscr{E}_{n+1}-\mathscr{E}_n}{\Delta t} &= -\frac{1}{4}\left\langle D\left(E^{n+1}+E^n\right),\, E^{n+1}+E^n\right\rangle, \\ &\leq 0. \end{aligned} \tag{3.88}$$

□

Note that the condition (3.80) is similar to the stability condition of the LF2 scheme (3.7) with S_0 instead of S, allowing to let the definition Δt be restricted to the subset of the coarse grid elements. Thus in the presence of a local refinement, the purpose of the IMEX method is achieved since the most severe stability constraints on explicit time integration methods can be overcome.

3.4.3 Convergence

In [Verwer 2010] the author has proven that the subdivision into coarse and fine elements is not detrimental to the second-order ODE convergence of the method (3.62), under stable simultaneous space-time grid refinement towards the exact underlying PDE solution. Let $H^h(t)$ and $E^h(t)$ denote the exact solutions of the Maxwell problem under consideration, restricted to the space grid i.e. the exact solutions of the system of ODEs

$$\begin{cases} \partial_t E^h(t) &= SH^h(t)-DE^h(t)+j^s(t)+\zeta_E^h(t), \\ \partial_t H^h(t) &= -S^TE^h(t)+\zeta_H^h(t), \end{cases} \tag{3.89}$$

with $\zeta_H^h(t)$ and $\zeta_E^h(t)$ denoting the spatial truncation errors.

Theorem 3.4.1. *Let $j^s(t) \in C^2[0,T]$ and suppose a Lax-Richtmyer stable space-time grid refinement $\Delta t \sim h$, $h \to 0$. On the interval $[0,T]$ the approximations H^n and E^n of method (3.62) then converge with order two to $H^h(t)$ and $E^h(t)$.*

CHAPTER 4

Numerical results

Contents

In this chapter, we present some numerical results to illustrate the theoretical results obtained in Chapter 3 and highlight the attractive features of locally implicit methods versus the fully explicit method in the presence of local mesh refinements. The two- and three-dimensional Maxwell solvers are written in Fortran 77/90. We recall that we consider DGTD methods relying on an arbitrary high-order polynomial interpolation of the component of the electromagnetic field, and their computer implementations make use of nodal (Lagrange) basis expansions on simplicial elements, see Appendix A. Furthermore, every N-simplex, τ, is the image of the reference N-simplex, τ_0, by an invertible affine transformation, hence by a simple change of variable all calculations can be reduced to calculation on τ_0, see Section 2.3. In practice, the elementary matrices are stored once for all on τ_0, then the coefficients of matrices on any N-simplex are calculated by this change of variable. Finally, the linear systems associated to the locally implicit methods (3.12) and (3.62) (see (3.15) and (3.65), respectively) are solved using the MUMPS (MUltifrontal Massively Parallel sparse direct Solver) optimized sparse direct solver [Amestoy *et al.* 2000]. MUMPS is a package for solving systems of linear equations of the form $Ax = b$, where A is a square sparse matrix that can be either unsymmetric, symmetric positive definite, or general symmetric. MUMPS is direct method based on a multifrontal approach which performs a direct factorization $A = LU$ or $A = LDL^T$ depending on the symmetry of the matrix.

Section 4.1 deals with two-dimensional problems. The first test case that we present is the propagation of an eigenmode in a unitary perfectly electrically conducting (PEC) cavity. This problem is a useful numerical test to validate the two-dimensional Maxwell solvers and to study the numerical convergence of the two locally implicit methods, since an analytical solution is known. The second one, which is an artificial problem, allows us to highlight the possible reduction of the convergence rate when we use the locally implicit method (3.12) (see Theorem 3.3.1). In addition, we give a

first overview of the efficiency of the locally implicit time integration methods in the presence of local mesh refinements, even if for this specific problem the mesh refinement is not justified from a physical or a geometrical point of view. Finally, we consider two problems for which local mesh refinements are required. Then we compare the results obtained with both locally implicit methods, the fully implicit method (computational work for the implicit treatment), and the fully explicit method (error, final CPU times). Section 4.2 deals with three-dimensional problems. The first test case that we present is the propagation of a standing wave in a cubic PEC cavity, which is also a useful validation test for the three-dimensional Maxwell solvers. The second one is a more realistic problem, the exposure of head tissues to a localized source radiation.

4.1 Two-dimensional problems

We consider the three-dimensional (normalized) Maxwell equations

$$\left\{\begin{array}{rcl}
\mu(\mathbf{x})\dfrac{\partial H^x}{\partial t}(\mathbf{x},t) &=& \dfrac{\partial E^y}{\partial z}(\mathbf{x},t)-\dfrac{\partial E^z}{\partial y}(\mathbf{x},t),\\
\mu(\mathbf{x})\dfrac{\partial H^y}{\partial t}(\mathbf{x},t) &=& \dfrac{\partial E^z}{\partial x}(\mathbf{x},t)-\dfrac{\partial E^x}{\partial z}(\mathbf{x},t),\\
\mu(\mathbf{x})\dfrac{\partial H^z}{\partial t}(\mathbf{x},t) &=& \dfrac{\partial E^x}{\partial y}(\mathbf{x},t)-\dfrac{\partial E^y}{\partial z}(\mathbf{x},t),\\
\varepsilon(\mathbf{x})\dfrac{\partial E^x}{\partial t}(\mathbf{x},t) &=& \dfrac{\partial H^z}{\partial y}(\mathbf{x},t)-\dfrac{\partial H^y}{\partial z}(\mathbf{x},t)-\sigma(\mathbf{x})E^x(\mathbf{x},t)-J^{s,x}(\mathbf{x},t),\\
\varepsilon(\mathbf{x})\dfrac{\partial E^y}{\partial t}(\mathbf{x},t) &=& \dfrac{\partial H^x}{\partial z}(\mathbf{x},t)-\dfrac{\partial H^z}{\partial x}(\mathbf{x},t)-\sigma(\mathbf{x})E^y(\mathbf{x},t)-J^{s,y}(\mathbf{x},t),\\
\varepsilon(\mathbf{x})\dfrac{\partial E^z}{\partial t}(\mathbf{x},t) &=& \dfrac{\partial H^y}{\partial x}(\mathbf{x},t)-\dfrac{\partial H^x}{\partial y}(\mathbf{x},t)-\sigma(\mathbf{x})E^z(\mathbf{x},t)-J^{s,z}(\mathbf{x},t),
\end{array}\right. \tag{4.1}$$

where $\mathbf{E}=(E^x,E^y,E^z)$ and $\mathbf{H}=(H^x,H^y,H^z)$ denote the electric and magnetic fields, ε, μ and σ are coefficients representing (relative) dielectric permittivity, magnetic permeability and conductivity, and $\mathbf{J}^s=(J^{s,x},J^{s,y},J^{s,z})$ represents the source current density maintained by external sources. From this three-dimensional model we derive the two-dimensional Transverse Magnetic model (TM_z) for components H^x, H^y and E^z

$$\left\{\begin{array}{rcl}
\mu\dfrac{\partial H^x}{\partial t} &=& -\dfrac{\partial E^z}{\partial y},\\
\mu\dfrac{\partial H^y}{\partial t} &=& \dfrac{\partial E^z}{\partial x},\\
\varepsilon\dfrac{\partial E^z}{\partial t} &=& \dfrac{\partial H^y}{\partial x}-\dfrac{\partial H^x}{\partial y}-\sigma E^z-J^{s,z}.
\end{array}\right. \tag{4.2}$$

The equations (4.2) are space discretized using a DG method formulated on triangular meshes, see Chapter 2. In the preliminary implementation of this DG method, the approximation of the electromagnetic field components within a triangle τ_i relies on a nodal $\mathbb{P}_k$ interpolation method. The a priori convergence analysis for this DG method based on a centered numerical flux and formulated on simplicial meshes shows that the convergence rate is $\mathscr{O}(h^k)$ for a k-th interpolation order [Fezoui *et al.* 2005]. A triangle τ_i is characterized by its height h_i. Denote Ω_h the computational

domain and Ω_h^{exp} the set of triangles that belong to the region where the explicit method is used in the implicit-explicit splitting. The critical step size, denoted Δt_c, used in the numerical tests is given by

$$\Delta t_c = CFL \times \min_{\tau_i \in \Omega_h^{exp}} h_i. \tag{4.3}$$

The values of the CFL number, given in Table 4.1, corresponds to the numerical stability, i.e. the limit beyond which we observe a growth of the discrete energy.

Method	DGTD-$\mathbb{P}_1$	DGTD-$\mathbb{P}_2$	DGTD-$\mathbb{P}_3$	DGTD-$\mathbb{P}_4$
Numerical CFL	0.30	0.20	0.13	0.09

Table 4.1: Numerical value of the CFL number in (4.3).

The simulations discussed in this section have been performed on a workstation equipped with an Intel Xeon 2.67 GHz processor and 8 GB of RAM memory.

4.1.1 Propagation of an eigenmode in a unitary perfectly electrically conducting (PEC) cavity

We consider the propagation of an eigenmode in a unitary perfectly electrically conducting (PEC) cavity. In this problem there is no source term, i.e. $J^{s,z} = 0$ in (4.2) and the exact solution is given by

$$\begin{cases} H^x(x,y,t) &= -\dfrac{k\pi}{\omega}\sin(l\pi x)\cos(k\pi y)\sin(\omega t), \\ H^y(x,y,t) &= \dfrac{l\pi}{\omega}\cos(l\pi x)\sin(k\pi y)\sin(\omega t), \\ E^z(x,y,t) &= \sin(l\pi x)\sin(k\pi y)\cos(\omega t), \end{cases} \tag{4.4}$$

where the resonance frequency is given by

$$\omega = \pi\sqrt{k^2+l^2}. \tag{4.5}$$

For numerical tests we put $k = l = 1$, the frequency is $f = 212$ MHz and the wavelength is $\lambda = 1.41$ m. We initialize the electromagnetic field with the exact analytical solution at $t = 0$, i.e.

$$\begin{cases} H^x(x,y,t=0) &= 0, \\ H^y(x,y,t=0) &= 0, \\ E^z(x,y,t=0) &= \sin(\pi x)\sin(\pi y). \end{cases} \tag{4.6}$$

For the boundary conditions, we consider a PEC cavity such that the tangential component of the electric field vanishes on the boundaries (see (2.16), metallic boundary condition)

$$n \times E^z = 0 \text{ on } \partial\Omega, \tag{4.7}$$

where the domain $\Omega = [0,1]^2$ (the unitary PEC cavity) and n denotes the unit outward normal to $\partial\Omega$. Finally, the total simulation time, T, is set to $T = 7.5$ m (normalized unit) which corresponds to the physical time $T = 2.5 \times 10^{-8}$ s.

4.1.1.1 Numerical convergence

We investigate the space-time convergence order (i.e. for a stable simultaneous space-time grid refinement $\Delta t \backsim h$, $h \to 0$) of the locally implicit methods (3.12) and (3.62). We consider a sequence of four successively refined triangular meshes (see Figure 4.1 for an example of a mesh with the identification of the implicit region). The characteristics of the different meshes used in the numerical tests are summarized in Table 4.2. The size of the critical time step (see (4.3)) for the locally implicit methods (3.12) and (3.62), is determined by the smallest height in the region treated explicitly (for the structured meshes and the implicit regions used in numerical tests it is equal to h^{max}, since all refined triangles belong to the implicit region). Note that for the present problem the local refinement is not justified from a physical or a geometrical point of view; this is only a validation test which allows us to study the convergence of locally implicit methods. For this purpose, we measure the maximal L^2-norm of the error for different meshes of increased resolution and we plot this error as a function of the square root of the number of degrees of freedom (DOF), in logarithmic scale. The use of the logarithmic scale allows to visualize the convergence rates as the slopes of the curves. We use the DGTD-$\mathbb{P}_k$ methods, with $k = 1,2,3$ and 4. The a priori convergence analysis for the error in $C^0([0,T],L^2(\Omega))$ and the DG method presented in Chapter 2, formulated on simplicial meshes and based on a centered numerical flux for the approximation of the boundary integral term at the interface between neighboring elements, shows that the convergence rate is $\mathcal{O}(h^k)$ for a k-th interpolation order [Fezoui *et al.* 2005]. The convergence result is slightly weaker than available results for upwind fluxes [Cockburn *et al.* 2000, Sármány *et al.* 2007]. Consequently a suboptimal convergence rate is obtained for the semi-discrete system in the L^2-norm which is the result of the adopted formulation and numerical treatment of the problem, namely the use of a centered flux. Nevertheless this setting allows to obtain the conservation of a discrete form of the electromagnetic energy for the two locally implicit methods (3.12) - (3.62), see [Dolean *et al.* 2010] and Section 3.4.2, respectively.

The obtained results are presented on Figure 4.2. For both locally implicit methods the numerical results are in accordance with the prediction of the theoretical analysis. Indeed, since the theoretical convergence rate is $\mathcal{O}(\Delta t^2 + h^k)$ for $\mathbb{P}_k$ interpolation method, with $k \leq 2$ the global convergence rate should be dominated by the spatial approximation order. This is what we observe in the present simulation (slightly higher for $k = 1$, 1.2 instead of 1.0). Considering $k > 3$ will lead to a rate of convergence bounded by two since this latter is in this case dominated by the temporal approximation. This is also what we observe for both locally implicit methods. Note that the condition (3.48) is not satisfied, hence we could observe a reduction by one of the convergence rate for the locally implicit method (3.12), see Theorem 3.3.1. This does not occur for the present test case (the condition is sufficient but not necessary), the component splitting is not detrimental for the convergence order of the locally implicit method (3.12). Finally, we can observe on Figure 4.2 that the locally implicit method (3.62) is more accurate than the locally implicit method (3.12), however we must also take into account the computational work for the implicit treatment of both methods.

4.1.1.2 Computational work for the implicit treatment

Now we focus on the cost of solving the linear systems of each locally implicit method. In particular, we are interested in the sparsity of the matrices to be inverted and the cost of the factorization step. Indeed, the matrices are factored only once before the main time stepping loop. Then, each linear

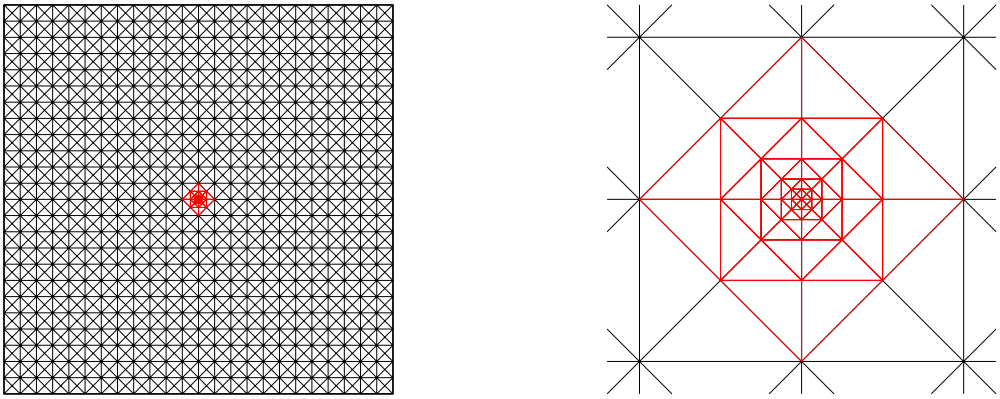


Figure 4.1: Propagation of an eigenmode in a unitary PEC cavity: example of a mesh used in numerical tests (1233 vertices, 2368 triangles, implicit treatment: 72 triangles, *red region*).

	# elements	h^{min}	h^{max}
Mesh (a)	848	0.00316	0.07143
Mesh (b)	2368	0.00184	0.04167
Mesh (c)	4688	0.00130	0.02941
Mesh (d)	7808	0.00100	0.02273

Table 4.2: Propagation of an eigenmode in a unitary PEC cavity: data of the four successively triangular meshes.

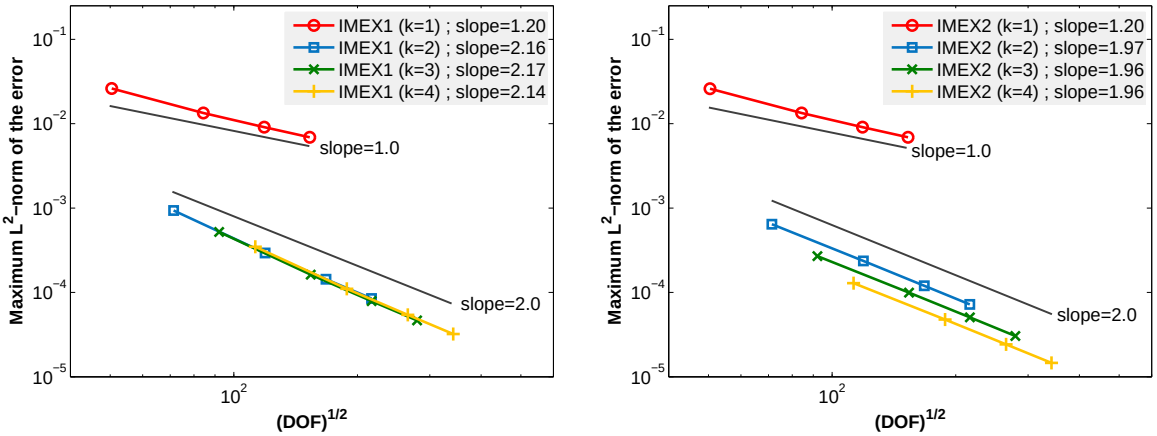
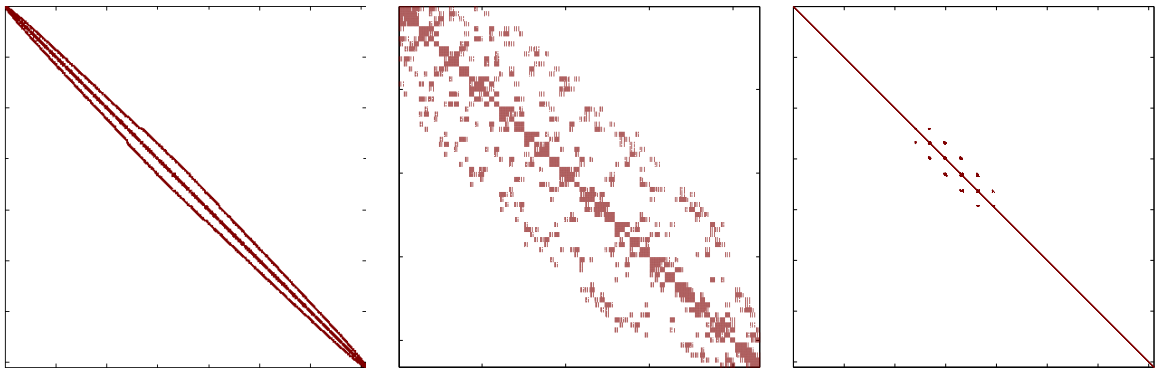


Figure 4.2: Propagation of an eigenmode in a unitary PEC cavity: convergence of the locally implicit methods (3.12) - (3.62) (left - right, respectively).

system inversion amounts to a forward and a backward solve using the L and U factors. All the numerical tests presented in this section have been conducted on the regular triangular mesh of the Figure 4.1. In Table 4.3, for each matrix of the linear system to be solved, we indicate the number of nonzero elements and the memory requirements for storing the L and U factors. In order to appreciate the properties of the locally implicit methods we also reported the results of the fully implicit method (3.8). To further illustrate the gain in sparsity, Figure 4.3 shows the sparsity pattern for the matrices $\mathscr{M}$, $\mathscr{M}_1$ and $\mathscr{M}_2$ defined in (3.70), (3.16) and (3.69), respectively, based on a DGTD-$\mathbb{P}_2$ method. As anticipated in Section 3.4.1 we remark that if we omit the identity matrix in the matrices $\mathscr{M}_1$ and $\mathscr{M}_2$ we recover a very close number of nonzero elements. The overhead for solving the linear system of the method (3.62) compared to the method (3.12), due to the matrix orders and the number of nonzero, leads to a higher final simulation time, as illustrated in Table 4.4. The locally implicit method (3.62) requires about 1.5 times more CPU time. However, as previously mentioned, when we compare the L^2-norm of the error, we observe that for an interpolation degree $k \geq 2$ the locally implicit method (3.62) is more accurate, especially for $k > 2$ (see Table (4.4)).



$\mathscr{M}$ **Matrix order = 14208 nz = 577410**

$\mathscr{M}_1$ **Matrix order = 432 nz= 16476**

$\mathscr{M}_2$ **Matrix order = 14208 nz =31092**

Figure 4.3: Propagation of an eigenmode in a unitary PEC cavity: sparsity pattern of the matrix of the linear system to solve for the fully implicit method (3.8) and the locally implicit methods (3.12) - (3.62) (from left to right), based on DGTD-$\mathbb{P}_2$ method. N.B.: the middle matrix (for IMEX method (3.12)) has a smaller dimension than the other two.

	Fully implicit method (3.8) : $\mathscr{M} = I + \frac{\Delta t^2}{4} SS^T$		
DGTD	Matrix order	# nonzero (nz)	RAM size (MB)
$\mathbb{P}_1$	7104	166939	8.914
$\mathbb{P}_2$	14208	577410	33.644
$\mathbb{P}_3$	23680	1481079	91.243
$\mathbb{P}_4$	35520	3133552	188.690

	Locally implicit method (3.12) : $\mathscr{M}_1 = I + \frac{\Delta t^2}{4} S_i S_i^T$		
DGTD	Matrix order	# nonzero (nz)	RAM size (MB)
$\mathbb{P}_1$	216	4728	0.139
$\mathbb{P}_2$	432	16476	0.556
$\mathbb{P}_3$	720	42489	1.399
$\mathbb{P}_4$	1080	90361	2.858

	Locally implicit method (3.62) : $\mathscr{M}_2 = I + \frac{\Delta t^2}{4} S_1 S_1^T$		
DGTD	Matrix order	# nonzero (nz)	RAM size (MB)
$\mathbb{P}_1$	7104	11880	0.235
$\mathbb{P}_2$	14208	31092	0.758
$\mathbb{P}_3$	23680	67557	1.828
$\mathbb{P}_4$	35520	128980	3.652

Table 4.3: Propagation of an eigenmode in a unitary PEC cavity: data and factorization of the matrix of the linear system to be solved for the fully implicit method (3.8) and the locally implicit methods (3.12) - (3.62), based on DGTD-$\mathbb{P}_k$ methods ($k = 1,2,3,4$).

	Locally implicit method (3.12)		**Locally implicit method** (3.62)	
$\mathbb{P}_k$	CPU time (s)	Error (L^2-norm)	CPU time (s)	Error (L^2-norm)
$\mathbb{P}_1$	9	1.3407e-2	14	1.3402e-2
$\mathbb{P}_2$	33	2.9383e-4	50	2.3632e-4
$\mathbb{P}_3$	113	1.6231e-4	175	9.9544e-5
$\mathbb{P}_4$	373	1.1048e-4	483	4.7695e-5

Table 4.4: Propagation of an eigenmode in a unitary PEC cavity: final simulation time and final L^2-norm of the error of the locally implicit methods (3.12) - (3.62), based on DGTD-$\mathbb{P}_k$ methods ($k = 1,2,3,4$).

4.1.2 Model test problem with an exact solution and a volume source term

We set $\varepsilon = \mu = 1$ and $\sigma = 0$ and we solve (4.2) in an unitary cavity, $\Omega = [0,1]^2$, with an artificial source term $J^{s,z}$ defined by

$$J^{s,z}(x,y,t) = -\mathrm{e}^t \left[x(1-x)y(1-y) + 2x(1-x) + 2y(1-y)\right]. \tag{4.8}$$

The exact solution of this problem is given by

$$\begin{cases} H^x(x,y,t) &= -\mathrm{e}^t x(1-x)(1-2y), \\ H^y(x,y,t) &= \mathrm{e}^t y(1-y)(1-2x), \\ E^z(x,y,t) &= \mathrm{e}^t x(1-x)y(1-y). \end{cases} \tag{4.9}$$

We impose a metallic boundary condition on the boundary $\partial\Omega$ such that the tangential component of the electric field vanishes on the boundaries, see (4.7). The total simulation time, T, is set to $T = 5.0$ m (normalized unit) which corresponds to the physical time $T = 1.67 \times 10^{-8}$ s.

First we investigate the space-time convergence order of the locally implicit methods (3.12) and (3.62) with a $\mathbb{P}_k$ interpolation method ($k = 1,2,3$ and 4). We consider the same sequence of four successively refined triangular meshes, see Figure 4.1 for an example of a mesh with the identification of the implicit region and Table 4.2 for the characteristics of the different meshes used in the numerical tests. We report the results on Figure 4.4. We recall that the theoretical convergence rate is $\mathscr{O}(\Delta t^2 + h^k)$ for a $\mathbb{P}_k$ interpolation method. For both locally implicit methods we obtain the expected convergence rate with $k = 1$, i.e., the first-order since in this case the convergence rate corresponds to the spatial convergence rate. For $k = 2$ or 3 we obtain the third-order for the locally implicit method (3.62) and the second-order for $k = 4$. For $k \geq 2$ and the locally implicit method (3.12) we obtain the second-order convergence rate. As in the previous problem, the condition (3.48) is not satisfied, but unlike the previous problem we observe a reduction by one of the convergence rate for the locally implicit method (3.12), with $k = 2$ or 3, compared to the locally implicit method (3.62). Consequently the possible reduction of the convergence rate, due to the component splitting, seems to occur for this problem (Theorem 3.3.1). The possible loss in accuracy could originate from the spatial discretization itself, regardless of component splitting. To eliminate this latter possibility we also repeat the numerical tests using the fully explicit and the fully implicit methods (3.6) and (3.8), respectively. We also report the results on Figure 4.4. Note that for the fully explicit method (3.6) the size of the critical time step (see (4.15)) is determined by the smallest height of the mesh, h^{min}, while for the locally implicit methods (3.12) and (3.62), it is determined by the smallest height in the region treated explicitly (for the structured meshes and the implicit regions used in numerical tests it is equal to h^{max}, since all refined triangles belong to the implicit region). For the unconditional stable implicit method (3.8) we chose to use the same critical time step that for the locally implicit methods. We observe the same behavior for the fully explicit method, the fully implicit method and the locally implicit method (3.62), i.e. the first-order convergence rate for $k = 1$, the third-order convergence rate for $k = 2$ or 3 and the second-order for $k = 4$. These results clearly confirm the reduction order by one for the locally implicit method (3.12) (see Theorem 3.3.1) while the subdivision into coarse and fine elements is not detrimental for the convergence order of the locally implicit method (3.62) (see Theorem 3.4.1) since we obtain the same behavior for the fully explicit and the fully implicit methods (3.6) - (3.8).

Regarding the computational work for the implicit treatment of the locally implicit methods and the fully implicit method, the comments of section 4.1.1.2 for the previous problem remain valid, since we consider the same meshes and the same implicit zones. Hence, Table 4.3, which indicates the number of nonzero entries and the memory requirements for storing the L and U factors, and Figure 4.3 which depicts the sparsity pattern for the different matrices of linear systems to solve, are valid for the present problem.

Even if the refinement is not justified for the present problem, we show in Table 4.5 and Figure 4.5 some numerical results for the locally implicit methods (3.12) and (3.62) and the fully explicit method (3.6), obtained on the regular triangular mesh of the Figure 4.1, in order to provide an idea of gains that we can expect by using an implicit-explicit approach. We observe in Table 4.5 that the locally implicit methods allow to overcome the step size limitations caused by the local refinement. With implicit-explicit approaches the sizes of the time step are about 16 times larger which yields significant gains of final CPU time, about 8 times lower than the fully explicit case. Regarding the error (see Figure 4.5) we observe that the locally implicit scheme (3.62) and the fully explicit scheme (3.6) based on DGTD-$\mathbb{P}_{1,2,3}$ methods give similar results. For a DGTD-$\mathbb{P}_4$ method the difference is more pronounced; we can assume that in the latter case the error in time is predominant on the error in space. As expected, with the previous numerical convergence study, Figure 4.5 also confirms that the locally implicit method (3.12) is less accurate, for $k \geq 2$. Because of the reduction by one of the order of convergence, a high-order spatial discretization is less advantageous for method (3.12) than for method (3.62) which retain its second-order PDE convergence.

The numerical results presented in this section show that, for a workload per time step about equal, the IMEX method (3.62) is more accurate than the IMEX method (3.12) and the results are very close that obtained with the fully explicit method. The reduction by one of the convergence rate for the locally implicit method (3.12) raises the question of the efficiency of this latter when high-order approximation polynomials are used within the DG method.

	Fully exp. method (3.6)		**Loc. imp. method** (3.12)			**Loc. imp. method** (3.62)		
$\mathbb{P}_k$	Δt (m)	CPU (s)	Δt	CPU	$\frac{\text{CPU}_{(3.6)}}{\text{CPU}_{(3.12)}}$	Δt	CPU	$\frac{\text{CPU}_{(3.6)}}{\text{CPU}_{(3.62)}}$
$\mathbb{P}_1$	3.91e-4	48.4	6.25e-3	10.4	4.7	6.25e-3	6.8	7.2
$\mathbb{P}_2$	2.60e-4	202.3	4.17e-3	27.5	7.4	4.17e-3	25.2	8.1
$\mathbb{P}_3$	1.69e-4	731.7	2.71e-3	84.2	8.7	2.71e-3	87.4	8.4
$\mathbb{P}_4$	1.17e-4	2328.4	1.87e-3	241.5	9.7	1.87e-3	257.4	9.1

Table 4.5: Model test problem with an exact solution and a volume source term: critical time step size and CPU time for the fully explicit method (3.6) and the locally implicit methods (3.12) - (3.62).

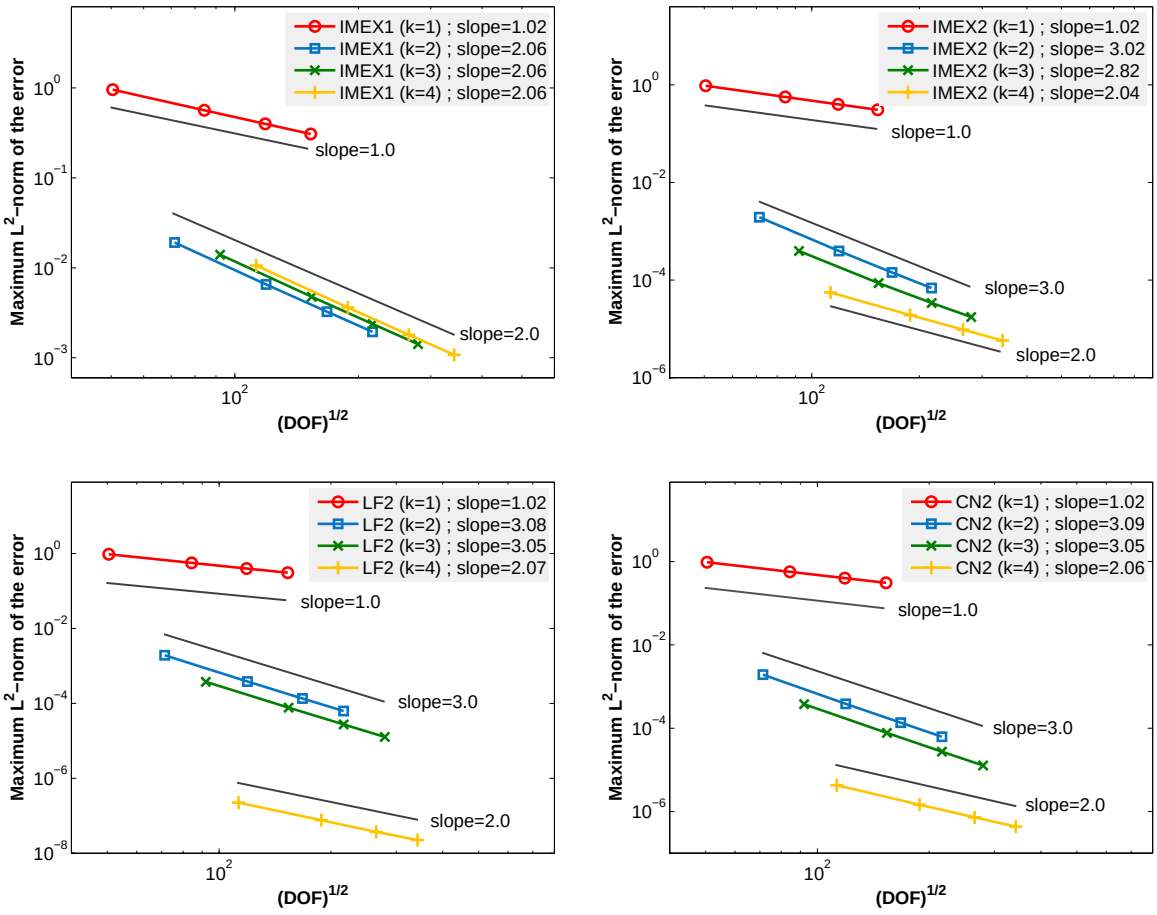


Figure 4.4: Model test problem with an exact solution and a volume source term: convergence of the locally implicit methods (3.12) - (3.62) (top left - top right, respectively), the fully explicit and the fully implicit method (3.6) - (3.8) (bottom left - bottom right, respectively).

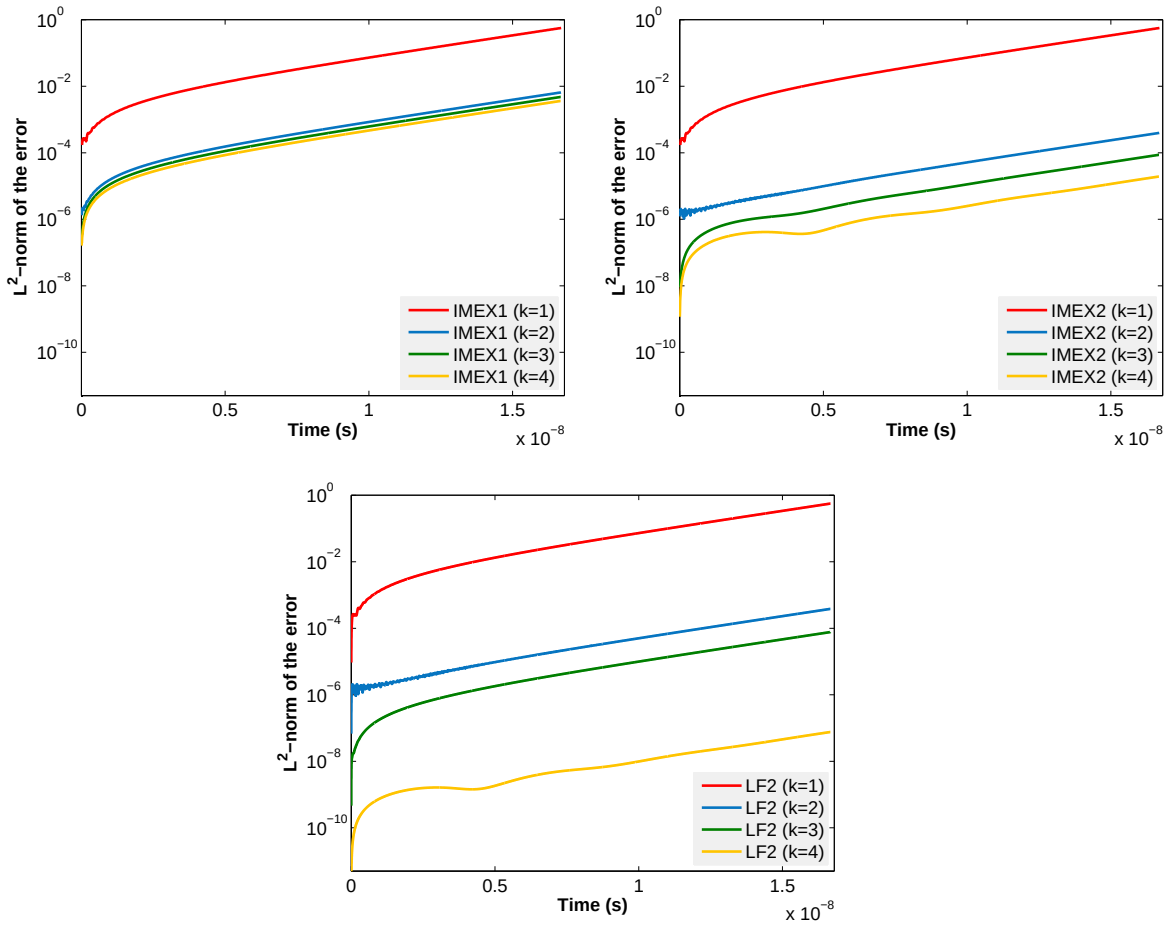


Figure 4.5: Model test problem with an exact solution and a volume source term: time evolution of the L^2-norm of the error of the locally implicit methods (3.12) - (3.62) (top left - top right, respectively), and the fully explicit method (3.6) - (3.8) (bottom), based on DGTD-$\mathbb{P}_k$ method.

4.1.3 Scattering of a modulated Gaussian by an airfoil profile

We consider an electromagnetic wave propagating into a rectangular domain $\Omega = [-1,2] \times [-1,1]$, with an airfoil profile, see Figure 4.6. We set $\varepsilon = \mu = 1$ and $\sigma = 0$ in (4.2), and we impose an absorbing boundary condition on the boundary of the domain (with a zero incident field) and a metallic boundary condition on the boundary of the airfoil profile, see (2.16). The electromagnetic field is excited by a local source term

$$J^{s,z}(x,y,t) = f(t)g(x,y), \tag{4.10}$$

where f is a time dependent function (Figure 4.7), g is a two-dimensional Gaussian function with $(x_0,y_0) = (-0.3,0.0)$ the center of the Gaussian spatial support

$$g(x,y) = A\mathrm{e}^{-\alpha((x-x_0)^2+(y-y_0)^2)}. \tag{4.11}$$

For the numerical experiments, $A = 1/\|g\|$ and the parameter α has been chosen such that the source term $J^{s,z}$ is strongly localized, which justifies the local refinement around the center of the Gaussian (the support of the Gaussian g is strictly included in the implicit zone, see Figures 4.6 and 4.7). The total simulation time, T, is set to $T = 10$ m (normalized unit) which corresponds to the physical time $T = 3 \times 10^{-8}$ s. Figure 4.8 shows the intensity of the electric field at different times.

First we focus on the cost of solving the linear systems of both locally implicit methods based on DGTD-$\mathbb{P}_k$ methods ($k = 1,2,3,4$). In Table 4.6, for each matrix of linear system to solve, we indicate the number of nonzero entries and the memory requirements for storing the L and U factors. We observe significant gains with the locally implicit methods compared to the fully implicit method. Similarly to the previous problems we remark that if we omit the identity matrix we recover a very close number of nonzero elements for the matrices of linear system to solve, for locally implicit methods (3.12) and (3.62). Consequently, we can expect that the additional cost to solve the linear system of the method (3.62) is weak compared to the method (3.12).

We also report some numerical results for the IMEX methods and the fully explicit method. In Table 4.7 we observe that the locally implicit methods allow to overcome the step size limitations caused by the local refinement. With implicit-explicit approaches the time step sizes are about 25.5 times larger which yields significant gains of final CPU time, about 12.5 times lower than the fully explicit case. Note that the obtained results for the final simulation time of two locally implicit methods are very close, which shows, as expected, that the additional cost to solve the linear system of the method (3.62) is weak compared to the method (3.12). For this problem we do not have an exact solution, then we compare the approximate solutions obtained with the locally implicit methods to that obtained with the fully explicit method based on DGTD-$\mathbb{P}_4$ method, that we consider as the reference method. In Table 4.7 we indicate a relative error for the L^2-norm of the discrete Fourier transform of the electromagnetic field given by

$$\text{relative error} = \frac{|\,\|DFT\,(W_{(3.12)\text{ or }(3.62)})\,\|_2 - \|DFT\,(W_{(3.6)})\,\|_2|}{\|DFT\,(W_{(3.6)})\,\|_2} \times 100, \tag{4.12}$$

where $\|DFT\,(W_{(3.12)\text{ or }(3.62)})\,\|_2$ is the L^2-norm of the discrete Fourier transform of the approximate electromagnetic field obtained with the locally implicit methods (3.12) or (3.62) and a $\mathbb{P}_k$ interpolation method ($k = 1,2,3$ or 4); $\|DFT\,(W_{(3.6)})\,\|_2$ is the L^2-norm of the discrete Fourier transform of

the approximate electromagnetic field obtained with the fully explicit method (3.6) and a $\mathbb{P}_4$ interpolation method. For $k \leq 2$ the results are very close while for $k > 2$ we can observe once again that the locally implicit method (3.62) is more accurate. Finally, these results are confirmed in Figures 4.9 and 4.10 where we plot the time evolution of the electric field at a selected point of the mesh (near the source center) for both locally implicit methods based on DGTD-$\mathbb{P}_k$ method ($k = 1, 2, 3$ and 4) and the fully explicit method based on DGTD-$\mathbb{P}_4$ method (reference method).

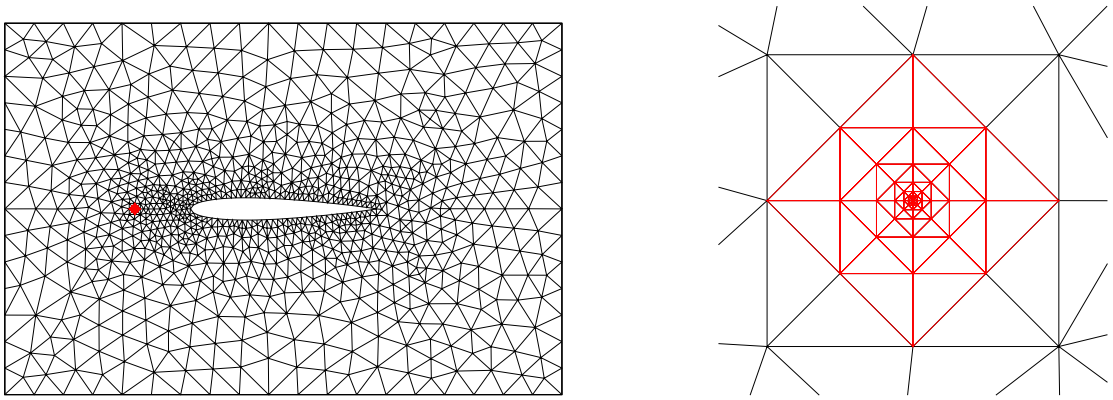


Figure 4.6: Scattering of a modulated Gaussian by an airfoil profile: mesh used in numerical tests (980 vertices, 1817 triangles, implicit treatment: 88 triangles, *red region*).

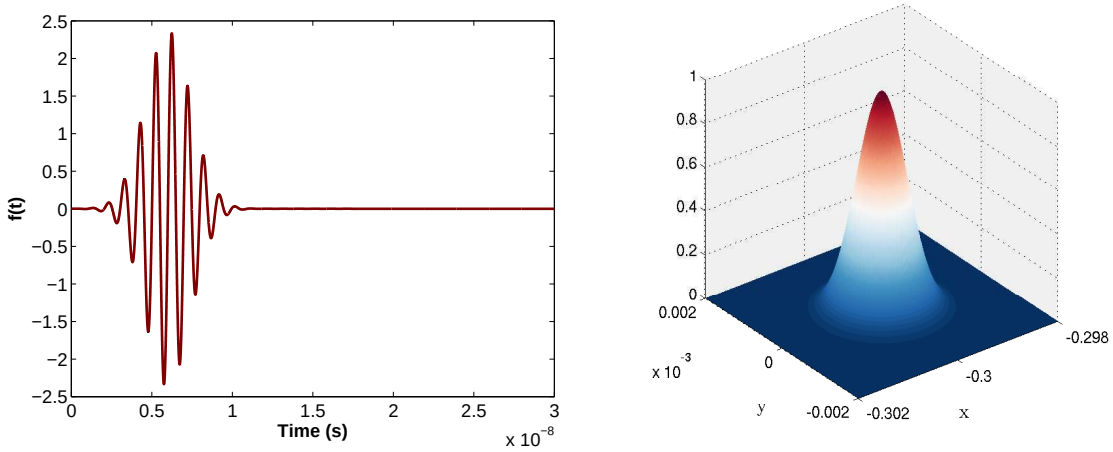


Figure 4.7: Scattering of a modulated Gaussian by an airfoil profile: temporal signal $f(t)$ and Gaussian function $g(x,y)/A$ on $[-0.302, -0.298] \times [-0.002, 0.002]$ (left and right, respectively).

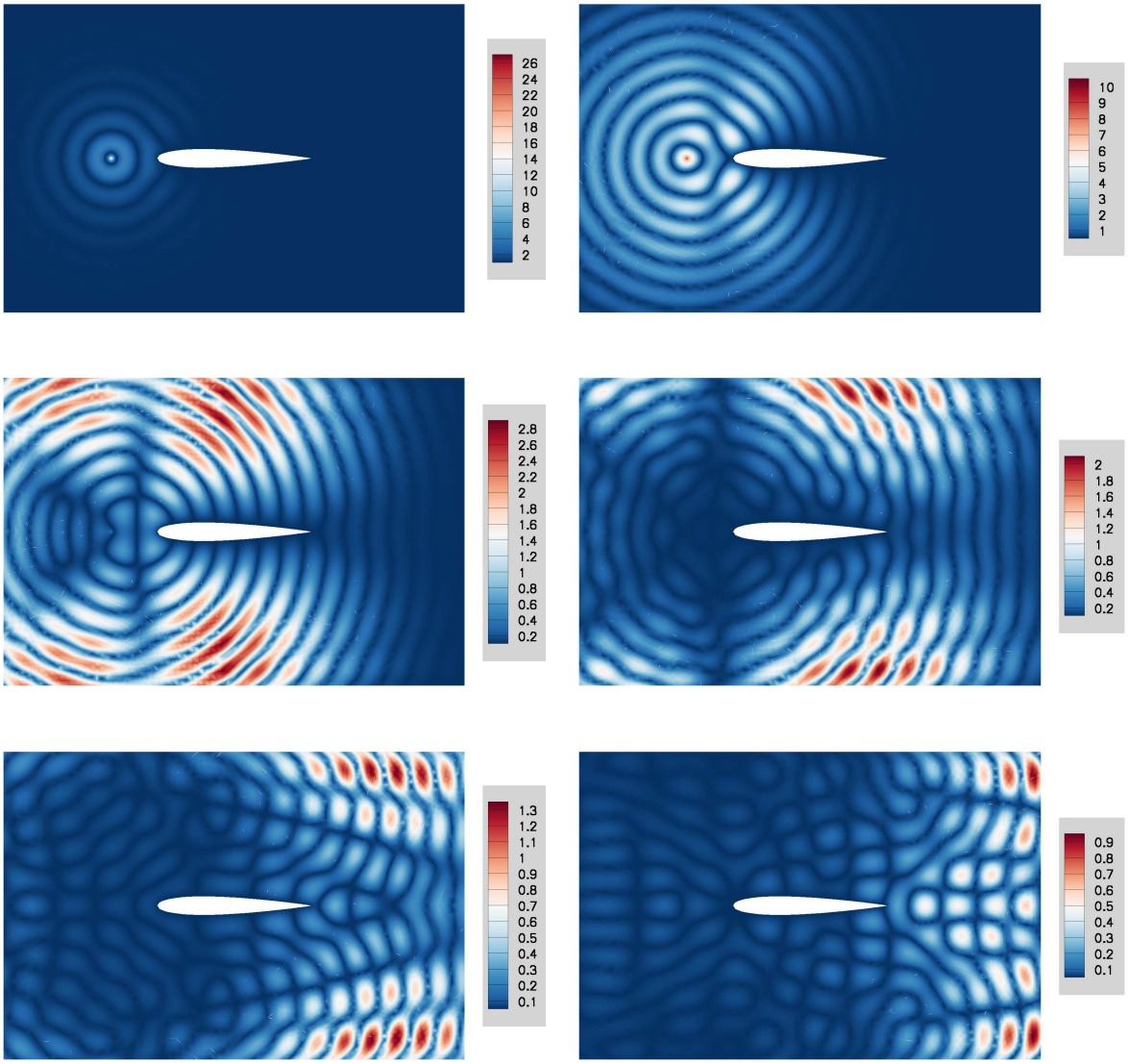


Figure 4.8: Scattering of a modulated Gaussian by an airfoil profile: time evolution of the intensity of the electric field, $|E^z|$, at time $t = 1.5$, 2.3, 3.0, 3.5, 4.0 and 4.5 m (normalized unit).

Fully implicit method (3.8) : $\mathscr{M} = I - \frac{\Delta t}{2} C^E + \frac{\Delta t^2}{4} SS^T$			
DGTD	Matrix order	# nonzero (nz)	RAM size (MB)
$\mathbb{P}_1$	5451	125625	7.323
$\mathbb{P}_2$	10902	436518	28.355
$\mathbb{P}_3$	18170	1119372	72.258
$\mathbb{P}_4$	27255	2367548	145.024

Locally implicit method (3.12) : $\mathscr{M}_1 = I - \frac{\Delta t}{2} C_i^E + \frac{\Delta t^2}{4} S_i S_i^T$			
DGTD	Matrix order	# nonzero (nz)	RAM size (MB)
$\mathbb{P}_1$	264	5592	0.180
$\mathbb{P}_2$	528	19533	0.724
$\mathbb{P}_3$	880	50782	1.821
$\mathbb{P}_4$	1320	108323	3.774

Locally implicit method (3.62) : $\mathscr{M}_2 = I - \frac{\Delta t}{2} C_1^E + \frac{\Delta t^2}{4} S_1 S_1^T$			
DGTD	Matrix order	# nonzero (nz)	RAM size (MB)
$\mathbb{P}_1$	5451	11043	0.250
$\mathbb{P}_2$	10902	30736	0.942
$\mathbb{P}_3$	18170	70190	2.284
$\mathbb{P}_4$	27255	138425	4.619

Table 4.6: Scattering of a modulated Gaussian by an airfoil profile: data and factorization of the matrix of the linear system to be solved for the fully implicit method (3.8) and the locally implicit methods (3.12) - (3.62), with DGTD-$\mathbb{P}_k$ methods ($k = 1,2,3,4$).

DGTD	$\frac{\Delta t_{(3.12)}}{\Delta t_{(3.6)}}$	$\frac{\text{CPU}_{(3.6)}}{\text{CPU}_{(3.12)}}$	$\frac{\lvert \lVert DFT\,(W_{(3.12)})\rVert_2 - \lVert DFT\,(W_{(3.6)})\rVert_2 \rvert}{\lVert DFT\,(W_{(3.6)})\rVert_2} \times 100$
	Locally implicit method (3.12)		
$\mathbb{P}_1$	25.8	9.5	6.9652 %
$\mathbb{P}_2$	25.5	10.5	0.4157 %
$\mathbb{P}_3$	25.6	14.2	0.1454 %
$\mathbb{P}_4$	25.7	17.1	0.0626 %
	Locally implicit method (3.62)		
DGTD	$\frac{\Delta t_{(3.62)}}{\Delta t_{(3.6)}}$	$\frac{\text{CPU}_{(3.6)}}{\text{CPU}_{(3.62)}}$	$\frac{\lvert \lVert DFT\,(W_{(3.62)})\rVert_2 - \lVert DFT\,(W_{(3.6)})\rVert_2 \rvert}{\lVert DFT\,(W_{(3.6)})\rVert_2} \times 100$
$\mathbb{P}_1$	25.8	10.2	7.0368 %
$\mathbb{P}_2$	25.5	12.0	0.4738 %
$\mathbb{P}_3$	25.6	12.9	0.0429 %
$\mathbb{P}_4$	25.7	14.3	0.0048 %

Table 4.7: Scattering of a modulated Gaussian by an airfoil profile: critical time step size, CPU time and relative error of the L^2-norm of the DFT of the electromagnetic field, for the locally implicit methods (3.12) - (3.62) compared to the fully explicit method (3.6).

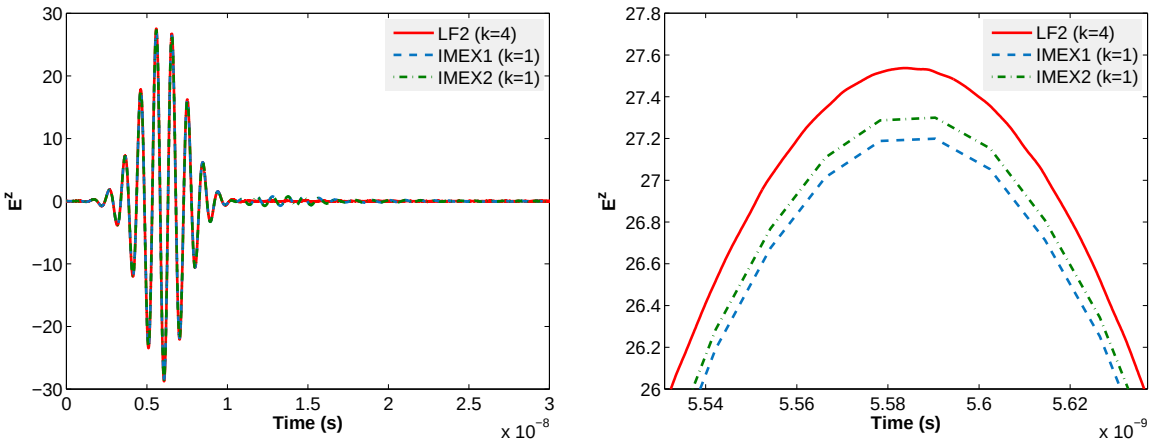


Figure 4.9: Scattering of a modulated Gaussian by an airfoil profile: time evolution of the electric field, E^z, at point $(x,y) = (-0.298, 0.0)$, for the locally implicit methods (3.12) and (3.62) based on the DGTD-$\mathbb{P}_1$ method, compared to the fully explicit method (3.6), based on the DGTD-$\mathbb{P}_4$ method.

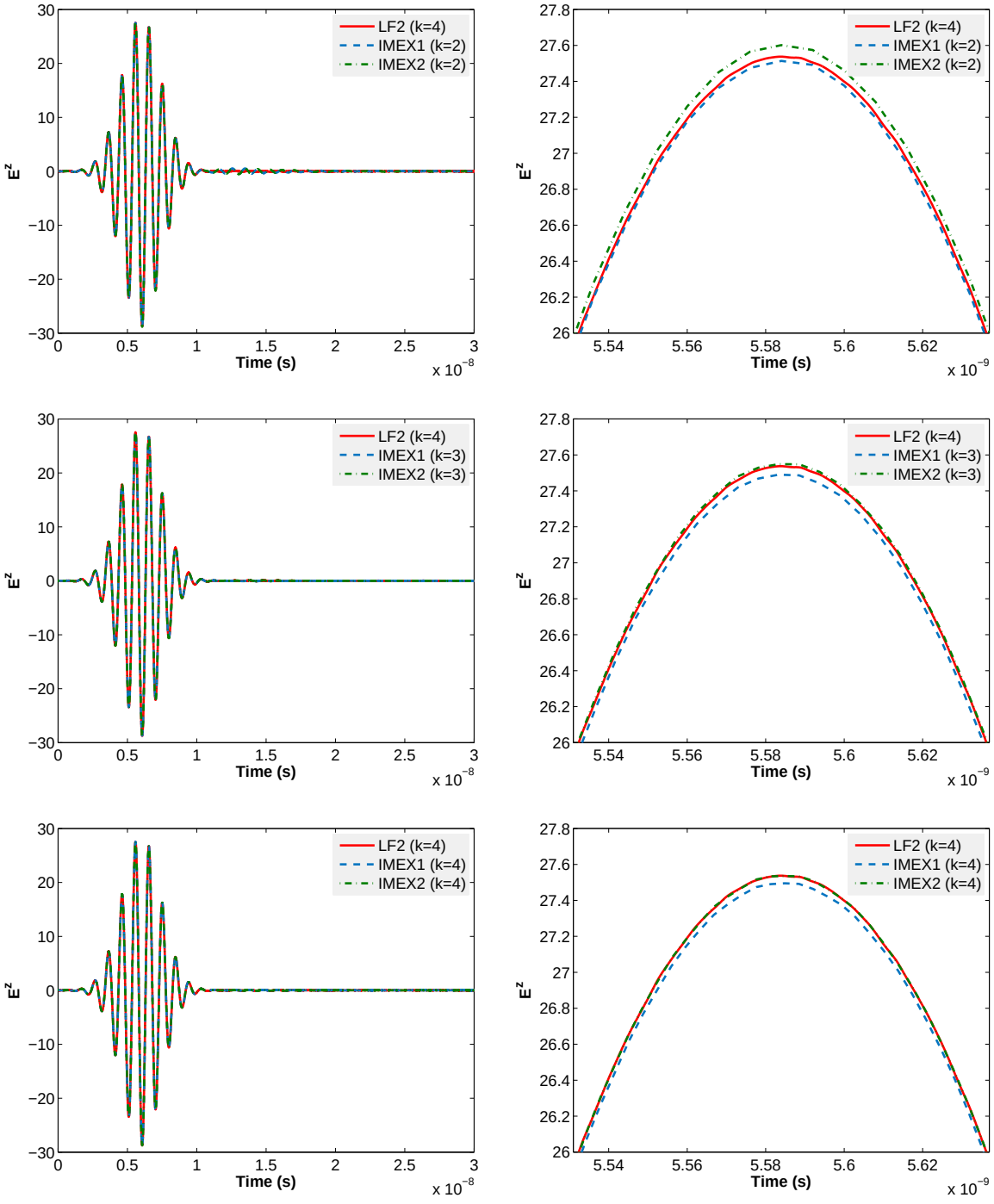


Figure 4.10: Scattering of a modulated Gaussian by an airfoil profile: time evolution of the electric field, E^z, at point $(x,y) = (-0.298, 0.0)$, for the locally implicit methods (3.12) and (3.62), based on DGTD-$\mathbb{P}_k$ ($k = 2, 3$ and 4) method, compared to the fully explicit method (3.6), based on DGTD-$\mathbb{P}_4$ method.

4.1.4 Scattering of a plane wave by a dielectric cylinder

We consider the scattering of a plane wave by a dielectric cylinder in free space. The cylinder is assumed to have a radius of r_2. The geometry of the problem is shown on Figure 4.11.

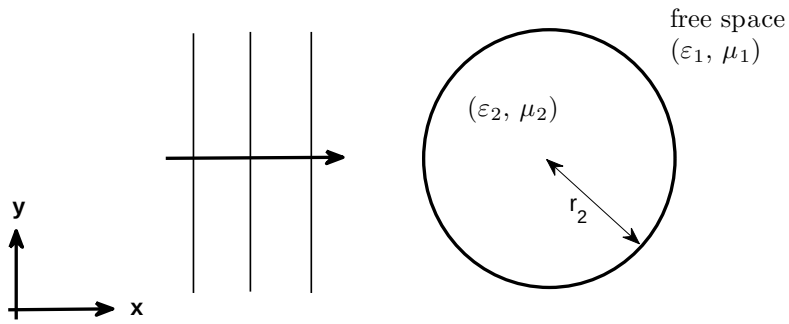


Figure 4.11: Scattering of a plane wave by a dielectric cylinder.

We assume that the cylinder is illuminated by a time-harmonic plane wave of the form

$$E^z_{inc} = \mathrm{e}^{-i(k_1 x - \omega t)}, \quad H^y_{inc} = -\mathrm{e}^{-i(k_1 x - \omega t)}, \tag{4.13}$$

where $k_1 = \omega\sqrt{\varepsilon_1\mu_1}$ is the propagation constant for homogeneous, isotropic free space medium and $\omega = 2\pi f$ is the angular frequency (or angular speed). The analytical solution of this problem is given by (see for example [Cai & Deng 2003, Fahs 2009])

$$E^z(x,y,t) = E^z(r,\theta,t) = \mathrm{e}^{i\omega t}\begin{cases} \displaystyle\sum_{n=-\infty}^{\infty} C_n^{\mathrm{tot}} J_n(k_2 r)\mathrm{e}^{in\theta}, & r \le r_2, \\ \displaystyle\sum_{n=-\infty}^{\infty} (i^{-n}J_n(k_1 r) + C_n^{\mathrm{scat}} H_n^{(2)}(k_1 r))\mathrm{e}^{in\theta}, & r > r_2, \end{cases}$$

$$H^\theta(r,\theta,t) = -\mathrm{e}^{i\omega t}\begin{cases} \dfrac{-ik_2}{\omega\mu_2}\displaystyle\sum_{n=-\infty}^{\infty} C_n^{\mathrm{tot}} J_n'(k_2 r)\mathrm{e}^{in\theta}, & r \le r_2, \\ \dfrac{-ik_1}{\omega\mu_1}\displaystyle\sum_{n=-\infty}^{\infty} (i^{-n}J_n'(k_1 r) + C_n^{\mathrm{scat}} H_n^{(2)'}(k_1 r))\mathrm{e}^{in\theta}, & r > r_2, \end{cases}$$

$$H^r(r,\theta,t) = -\mathrm{e}^{i\omega t}\begin{cases} \dfrac{i}{\omega\mu_2 r}\displaystyle\sum_{n=-\infty}^{\infty} in\, C_n^{\mathrm{tot}} J_n(k_2 r)\mathrm{e}^{in\theta}, & r \le r_2, \\ \dfrac{i}{\omega\mu_1 r}\displaystyle\sum_{n=-\infty}^{\infty} in(i^{-n}J_n(k_1 r) + C_n^{\mathrm{scat}} H_n^{(2)}(k_1 r))\mathrm{e}^{in\theta}, & r > r_2, \end{cases}$$

where $(r,\theta) = (\sqrt{x^2+y^2}, \arctan(y/x))$ are the usual polar coordinates, H^θ and H^r are the angular and radial components of the total magnetic field. J_n and $H_n^{(2)}$ denote the n-th order Bessel function of the first kind and the Hankel function of the second kind, and $k_2 = \omega\sqrt{\varepsilon_2\mu_2}$ is the propagation constant for homogeneous dielectric medium. Finally the expansion coefficients for the total field interior to the cylinder and the scattered field are given by

$$C_n^{\mathrm{tot}} = i^{-n}\frac{(k_1/\mu_1)J_n'(k_1 r_2)H_n^{(2)}(k_1 r_2) - (k_1/\mu_1)H_n^{(2)'}(k_1 r_2)J_n(k_1 r_2)}{(k_2/\mu_2)J_n'(k_2 r_2)H_n^{(2)}(k_1 r_2) - (k_1/\mu_1)H_n^{(2)'}(k_1 r_2)J_n'(k_2 r_2)},$$

and

$$C_n^{\text{scat}} = i^{-n} \frac{(k_1/\mu_1)J_n'(k_1r_2)J_n(k_2r_2) - (k_2/\mu_2)J_n'(k_2r_2)J_n(k_1r_2)}{(k_2/\mu_2)J_n'(k_2r_2)H_n^{(2)}(k_1r_2) - (k_1/\mu_1)H_n^{(2)'}(k_1r_2)J_n'(k_2r_2)}.$$

For the numerical tests we consider the following situation. The computational domain is artificially bounded by a cylinder with radius $r_1 = 0.045$ m on which the Silver-Müller absorbing condition is applied. The radius of the inner cylinder is set to $r_2 = 0.002$ m and bounds a material of relative permittivity $\varepsilon_2 = 7$ and relative permeability $\mu_2 = 1$ (non-magnetic material). The medium exterior to the dielectric cylinder is assumed to be vacuum, i.e. $\varepsilon_1 = \mu_1 = 1$. The frequency is set to $f = 30$ GHz, then the angular frequency is $\omega = 2\pi \times 30 \times 10^9$ rad·s^{-1} and the wavelength is $\lambda = 0.01$ m. The total simulation time is set to $T = 6.67 \times 10^{-10}$ s which corresponds to a propagation of the initial wave over 20 wavelengths. In this case all components of the electromagnetic field are continuous across the interface. The first derivative of E^z is also continuous while the first derivatives of H^x and H^y are discontinuous. We make use of a non-uniform triangular mesh which consists of 4860 vertices and 9342 triangles, see Figure 4.12. The local refinements are strongly localized on the absorbing boundary (circular geometry), inside the cylinder where the relative permittivity is high, and more specifically at the interface between the two materials where the solution is of low regularity. For the locally implicit method the implicit treatment is applied to the elements inside these regions (*red region* on Figure 4.12). Precisely the implicit region denoted by Ω_h^{imp} is defined by

$$\Omega_h^{imp} = \left\{ \tau \in \Omega_h \ / \ \alpha \times area(\tau) \leq \max_{\tau \in \Omega_h} area(\tau) \right\}, \tag{4.14}$$

where the coefficient α has been chosen such that the number of elements treated implicitly represent less than 15% of the total number of elements, and the critical time step size for the locally implicit methods is at least 10 times higher than the critical time step size for the fully explicit method.

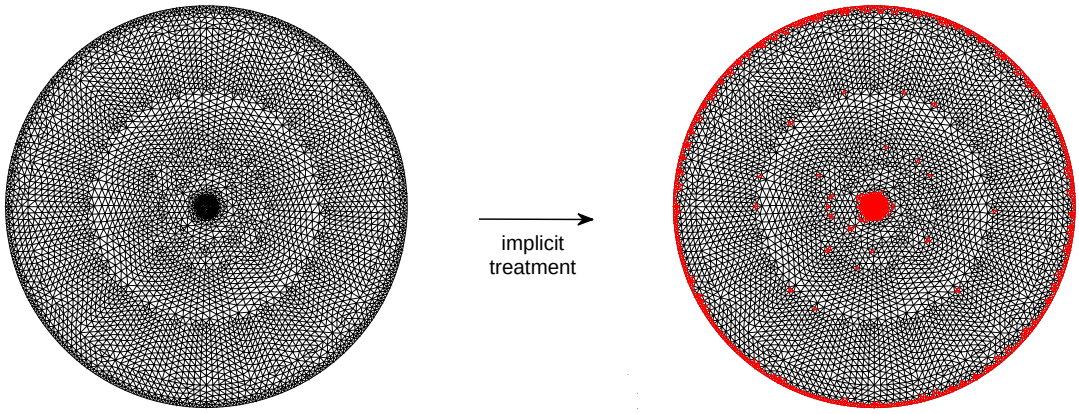


Figure 4.12: Scattering of a plane wave by a dielectric cylinder: triangular mesh used in numerical tests (4860 vertices, 9342 triangles, implicit treatment: 1365 triangles, *red region*).

We present results for DGTD-$\mathbb{P}_k$ methods with $k = 1, 2, 3$. The increasing accuracy with increasing k is here limited by the approximation error on the circular geometry of the interface between the two materials as well as the use of the first order Silver-Müller absorbing condition. Furthermore the low regularity of the solution also limit the convergence rate.

To measure the cost of the linear system to solve we indicate in Table 4.8 the number of nonzero elements and the memory requirement for the factorization of the matrix. We observe significant gains with the locally implicit methods compared to the fully implicit method. The figures reported in this Table also illustrate the overhead for storing the L and U factors for the IMEX method (3.62) compared to the IMEX method (3.12) (about 1.6 times more memory requirement).

In Table 4.9, we summarize the numerical results for the IMEX methods and the fully explicit method. We observe that the locally implicit methods allow to overcome the step size limitations caused by the local refinements. With implicit-explicit approaches the time step sizes are about 10 times larger which yields gains of final CPU time, about 6 times lower than the fully explicit case. Regarding the L^2-norm errors, the obtained results are very close for the IMEX method (3.62) and the explicit method (3.6), regardless of the interpolation degree used in the DGTD method, while for $k \geq 2$ the locally implicit method (3.12) is less accurate. The use of an interpolation degree $k \geq 2$ does not translate in a further reduction of the L^2-norm error for the IMEX method (3.12). Note that the L^2-norm error is calculated on the last period of simulation, using a discrete Fourier transform (DFT) of the approximate electromagnetic field. Similarly to the previous test case these numerical results show that the IMEX method (3.62) is more accurate than the IMEX method (3.12) for a similar CPU time. Figure 4.13 shows the contour lines of the DFT of E^z and H^y over the last period of simulation, and Figures 4.14 - 4.15 the obtained results by using the locally implicit method (3.62). Inside the dielectric cylinder we can see that for $k = 1$ a finer mesh would be more suitable, while for $k \geq 2$ the considered mesh is fine enough. Finally Figure 4.16 depicts the 1D distributions of E^z and H^y over the last period of simulation along $y = 0.0$.

	Fully implicit method (3.8) : $\mathscr{M} = I - \frac{\Delta t}{2} C^E + \frac{\Delta t^2}{4} SS^T$		
DGTD	Matrix order	# nonzero (nz)	RAM size (MB)
$\mathbb{P}_1$	28026	660216	61.622
$\mathbb{P}_2$	56052	2291999	243.202
$\mathbb{P}_3$	93420	5865412	605.482

	Locally implicit method (3.12) : $\mathscr{M}_1 = I - \frac{\Delta t}{2} C_i^E + \frac{\Delta t^2}{4} S_i S_i^T$		
DGTD	Matrix order	# nonzero (nz)	RAM size (MB)
$\mathbb{P}_1$	4095	67077	2.054
$\mathbb{P}_2$	8190	238830	8.095
$\mathbb{P}_3$	13650	620828	20.510

	Locally implicit method (3.62) : $\mathscr{M}_2 = I - \frac{\Delta t}{2} C_1^E + \frac{\Delta t^2}{4} S_1 S_1^T$		
DGTD	Matrix order	# nonzero (nz)	RAM size (MB)
$\mathbb{P}_1$	28026	110721	3.334
$\mathbb{P}_2$	56052	351225	12.726
$\mathbb{P}_3$	93420	859676	33.014

Table 4.8: Scattering of a plane wave by a dielectric cylinder: data and factorization of the matrix of the linear system to solve for the fully implicit method and the locally implicit methods.

	Fully exp. method (3.6)		**Loc. imp. method** (3.12)			**Loc. imp. method** (3.62)		
$\mathbb{P}_k$	Δt (m)	Error	Δt	Error	$\frac{\text{CPU}_{(3.6)}}{\text{CPU}_{(3.12)}}$	Δt	Error	$\frac{\text{CPU}_{(3.6)}}{\text{CPU}_{(3.62)}}$
$\mathbb{P}_1$	1.46e-5	2.92e-2	1.52e-4	2.93e-2	6.1	1.52e-4	2.93e-2	6.1
$\mathbb{P}_2$	9.70e-6	1.32e-3	1.01e-4	2.40e-2	6.2	1.01e-4	1.85e-3	6.0
$\mathbb{P}_3$	6.36e-6	8.29e-4	6.58e-5	1.42e-2	6.0	6.58e-5	8.78e-4	5.7

Table 4.9: Scattering of a plane wave by a dielectric cylinder: critical time step size, error (L^2-norm) and final CPU time for the fully explicit method and the locally implicit methods.

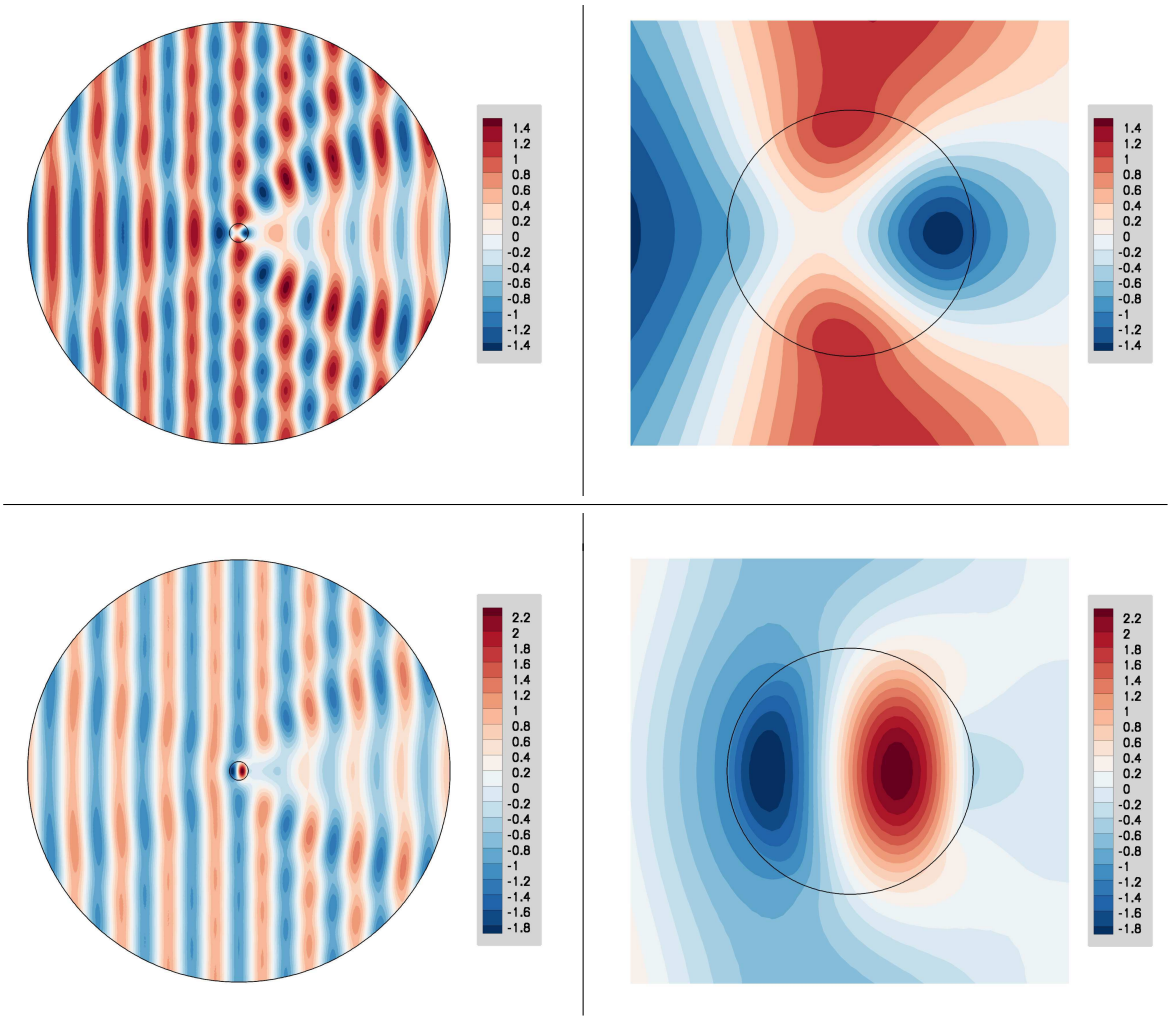


Figure 4.13: Scattering of a plane wave by a dielectric cylinder: contour lines of the real part of the DFT of E^z and H^y (top and bottom, respectively) over the last period of simulation (exact solution).

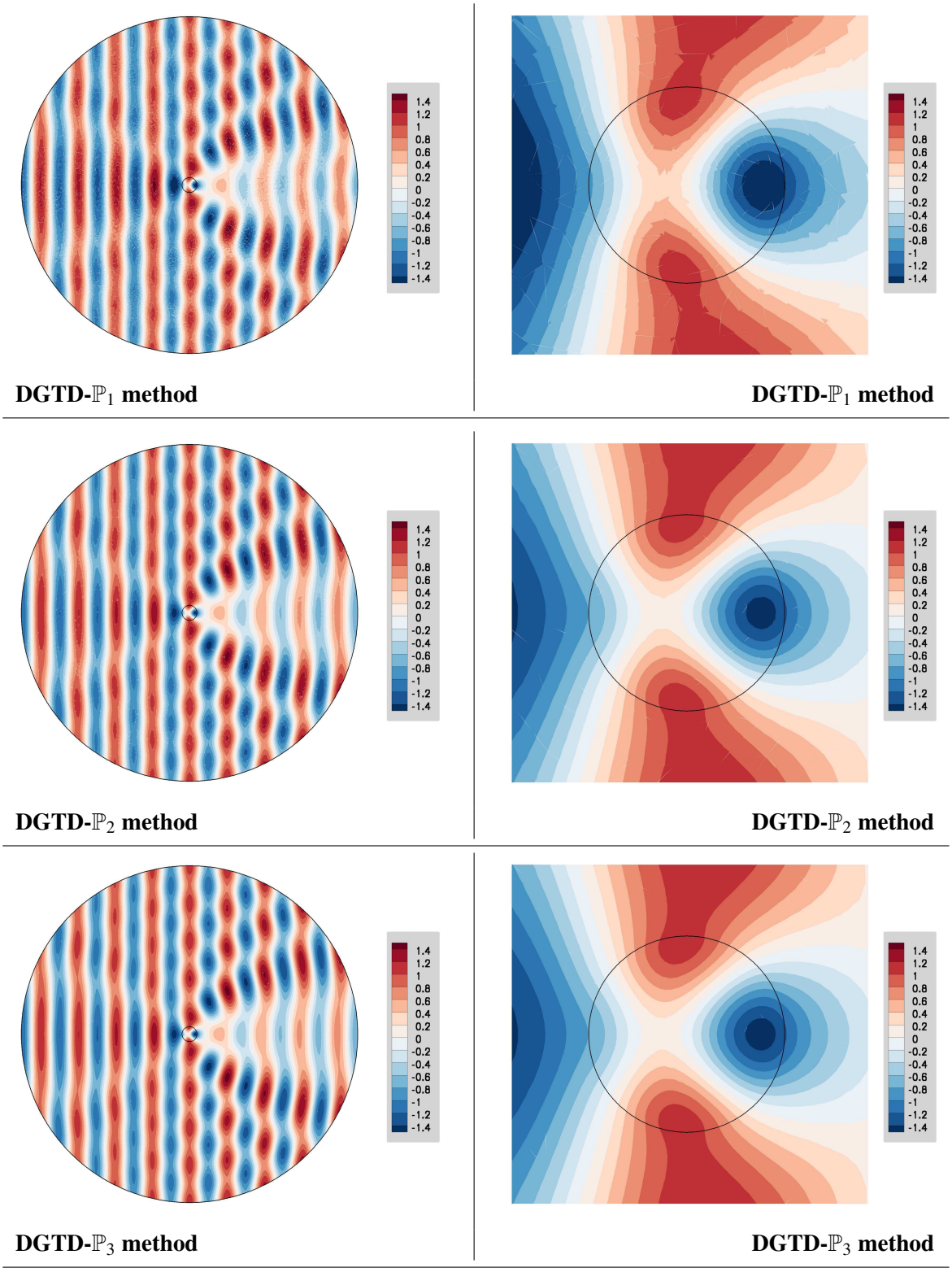
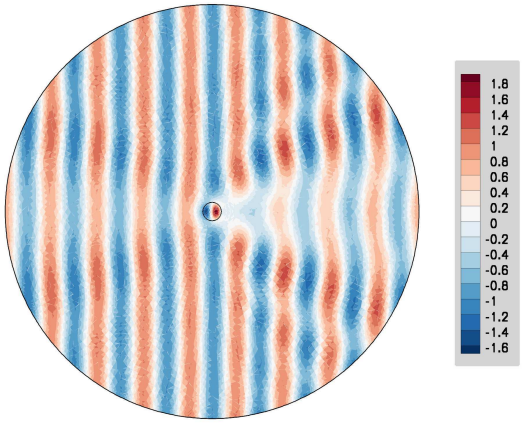
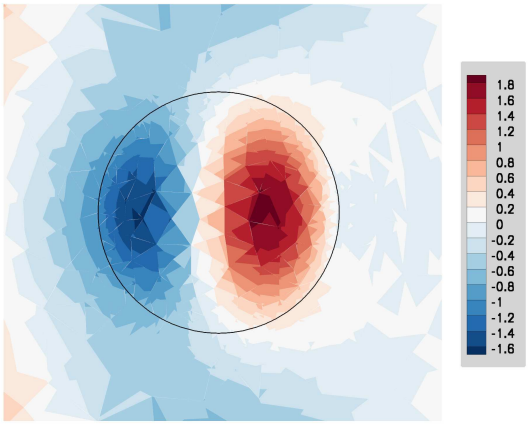


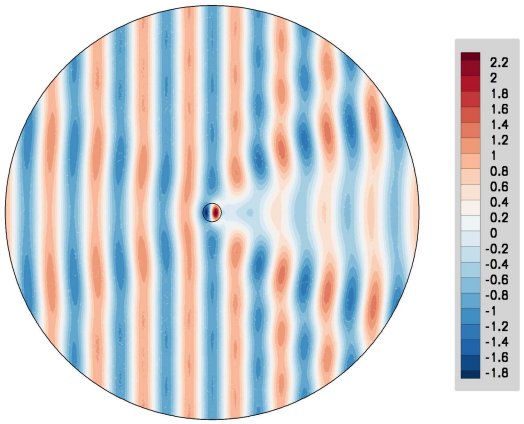
Figure 4.14: Scattering of a plane wave by a dielectric cylinder: contour lines of the DFT of E^z (real part) over the last period of simulation, for the locally implicit method (3.62).



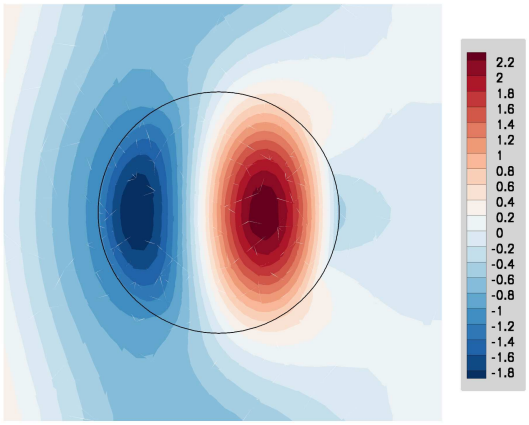
DGTD-$\mathbb{P}_1$ method



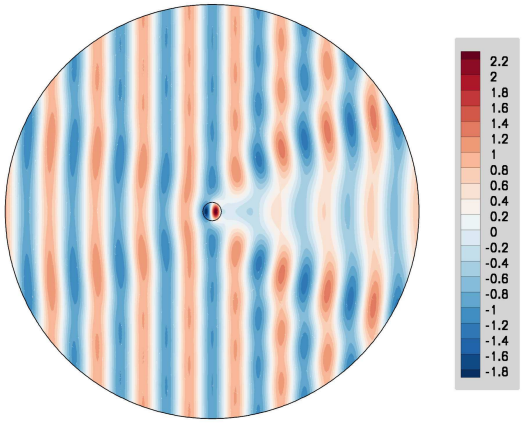
DGTD-$\mathbb{P}_1$ method



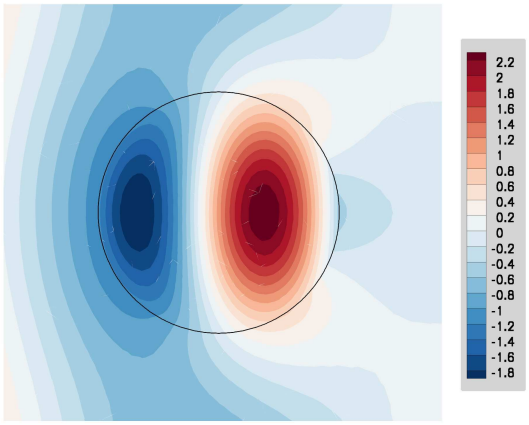
DGTD-$\mathbb{P}_2$ method



DGTD-$\mathbb{P}_2$ method



DGTD-$\mathbb{P}_3$ method



DGTD-$\mathbb{P}_3$ method

Figure 4.15: Scattering of a plane wave by a dielectric cylinder: contour lines of the DFT of H^y (real part) over the last period of simulation, for the locally implicit method (3.62).

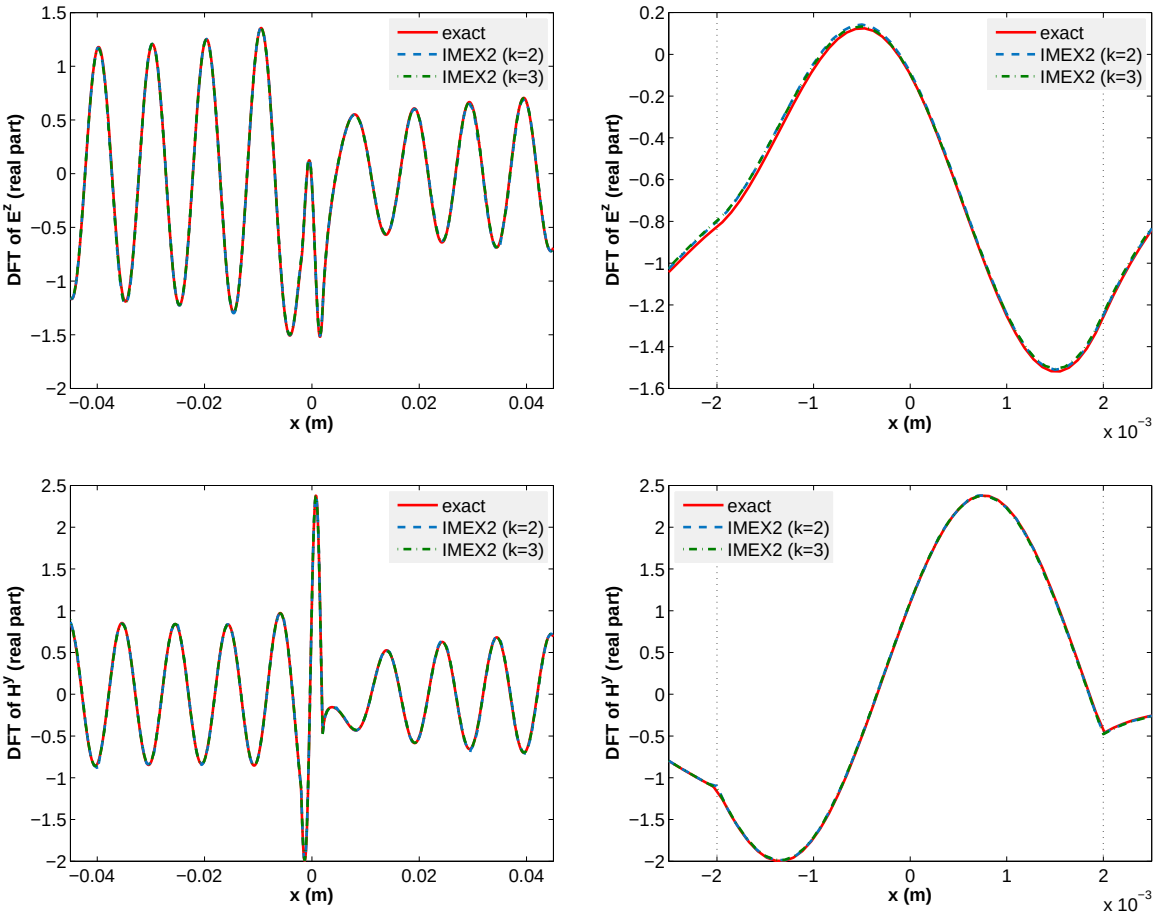


Figure 4.16: Scattering of a plane wave by a dielectric cylinder: 1D distribution of the DFT of E^z and H^y over the last period of simulation, along $y = 0$, for the locally implicit method (3.62), based on DGTD-$\mathbb{P}_k$ method ($k = 2$ and 3).

4.2 Three-dimensional problems

We observed through the previous two-dimensional problems when the reduction by one of the convergence rate occurs in the locally implicit method (3.12) that the use of an interpolation degree $k \geq 2$ does not translate in a further reduction of the error and raises the question of the efficiency of this method when high-order approximation polynomials are used within the DG method. Hence, from now, we will focus our analysis on the locally implicit method (3.62) which is the most accurate implicit-explicit approach and does not suffer from a reduction of convergence order.

We consider the three-dimensional (normalized) Maxwell equations (4.1). These equations are space discretized using a DG method formulated on a tetrahedral mesh. In the preliminary implementation of this DG method, the approximation of the electromagnetic field components within a tetrahedron τ_i relies on a nodal $\mathbb{P}_k$ interpolation method. We denote by Ω_h the computational domain and Ω_h^{exp} the set of tetrahedra that belong to the region where the explicit method is used into the implicit-explicit splitting. The critical step size, denoted Δt_c, used in the numerical tests is then given by

$$\Delta t_c = CFL \times \min_{\tau_i \in \Omega_h^{exp}} \left(\frac{4 \times V_i}{A_i} \right), \tag{4.15}$$

where V_i and A_i are the volume and the surface area of tetrahedron τ_i, respectively. The values of the CFL number corresponds to the numerical stability, i.e. the limit beyond which we observe a growth of the discrete energy.

The simulations discussed in this section have been performed on a workstation equipped with an Intel Xeon 2.40 GHz processor and 16 GB of RAM memory.

4.2.1 Propagation of a standing wave in a cubic PEC cavity

We compute the $(1,1,1)$ mode which is a standing wave of frequency $f = 260$ MHz and wavelength $\lambda = 1.15$ m in unitary PEC cavity, $\Omega = [0,1]^3$, where relative permittivity and permeability set to $\varepsilon = \varepsilon_0 = 1$ and $\mu = \mu_0 = 1$ are the constant vacuum values, $\sigma = 0$ (no conductivity) and $\mathbf{J}^s = 0$. Similarly to the validation test of the propagation of an eigenmode in a PEC cavity for the two-dimensional Maxwell solvers, the $(1,1,1)$ mode is also useful to validate the three-dimensional Maxwell solvers and to study the numerical convergence since the exact time-domain solution is given by

$$\begin{cases} H^x(\mathbf{x},t) &= -\dfrac{\pi}{\omega} \sin(\pi x)\cos(\pi y)\cos(\pi z)\sin(\omega t), \\ H^y(\mathbf{x},t) &= \dfrac{2\pi}{\omega} \cos(\pi x)\sin(\pi y)\cos(\pi z)\sin(\omega t), \\ H^z(\mathbf{x},t) &= -\dfrac{\pi}{\omega} \cos(\pi x)\cos(\pi y)\sin(\pi z)\sin(\omega t), \\ E^x(\mathbf{x},t) &= -\cos(\pi x)\sin(\pi y)\sin(\pi z)\cos(\omega t), \\ E^y(\mathbf{x},t) &= 0, \\ E^z(\mathbf{x},t) &= \sin(\pi x)\sin(\pi y)\cos(\pi z)\cos(\omega t), \end{cases} \tag{4.16}$$

where the angular frequency (or pulsation) is given by $\omega = 2\pi f$ (rad$\cdot$s^{-1}). We impose a metallic boundary condition on the boundary of the cube such that the tangential component of the electric

field vanishes on the boundaries

$$\mathbf{n} \times \mathbf{E} = 0 \text{ on } \partial\Omega, \tag{4.17}$$

where $\mathbf{n}$ denotes the unit outward normal to $\partial\Omega$. Throughout this section the total simulation time is set to $T = 1.67 \times 10^{-8}$s which corresponds to a normalized time $T = 5$m and we initialize the electromagnetic field with the exact analytic solution (4.16) at $t = 0$, i.e. $H^x = H^y = H^z = E^y = 0$, $E^x = -\cos(\pi x)\sin(\pi y)\sin(\pi z)$ and $E^z = \sin(\pi x)\sin(\pi y)\cos(\pi z)$.

We investigate the space-time convergence order (i.e. for a stable simultaneous space-time grid refinement $\Delta t \backsim h$, $h \to 0$) of the fully explicit method and the locally implicit methods. We measure the maximal L^2-norm of the error for a sequence of four successively locally refined tetrahedral meshes, see Figure 4.17 for examples of meshes with the identification of the region for implicit treatment. We plot this error as a function of $1/h$, in logarithmic scale, the use of the logarithmic scale allows to visualize the convergence rates as the slopes of the curves. We use the DGTD-$\mathbb{P}_k$ method, with $k = 1$ or 2. The obtained results are summarized in Figure 4.18 (left plots); for both methods, the results are similar and slightly better than the theoretical behaviors. Furthermore these results clearly confirm that the subdivision into coarse and fine elements is not detrimental for the convergence order of the locally implicit method. We also plot in Figure 4.18 (right plots) the error as a function of the CPU time. For a given error we can observe the gain of final CPU time with the locally implicit approach compared to the fully explicit method.

Now we focus on the the linear systems to be solved and the efficiency of the locally implicit method. We consider the locally refined tetrahedral mesh composed of 40616 tetrahedra and 7756 vertices (Figure 4.17, on right). First, we are interested in the sparsity of the matrix to be inverted and the cost of the factorization step. The matrix is factorized only once before the main time stepping loop, then, each linear system inversion amounts to a forward and a backward solve using the L and U factors. In Table 4.10 for each matrix of linear system to solve, we indicate the number of nonzero elements and the fill-in ratio percentage. We also indicated the total size (memory requirement) of all internal data used during numerical factorization and the total CPU time for analysis and factorization. Note that for the fully implicit scheme based on DGTD-$\mathbb{P}_{1,2}$ methods the number of nonzero are 75081406 and 408726916, respectively, and will require too much memory to be a reasonable alternative to the fully explicit method. Finally we present some numerical results for the locally implicit method and the fully explicit method. In Table 4.11 we observe that the locally implicit method allows to overcome the step size limitations caused by the local mesh refinement. With implicit-explicit approaches the sizes of the time step are about 32 times larger which yields significant gains of final CPU time, which is about 7 times lower than the fully explicit case. Regarding the time evolution of the error, Figure 4.19, we observe that the IMEX scheme and the fully explicit scheme based on DGTD-$\mathbb{P}_{1,2}$ methods give similar results.

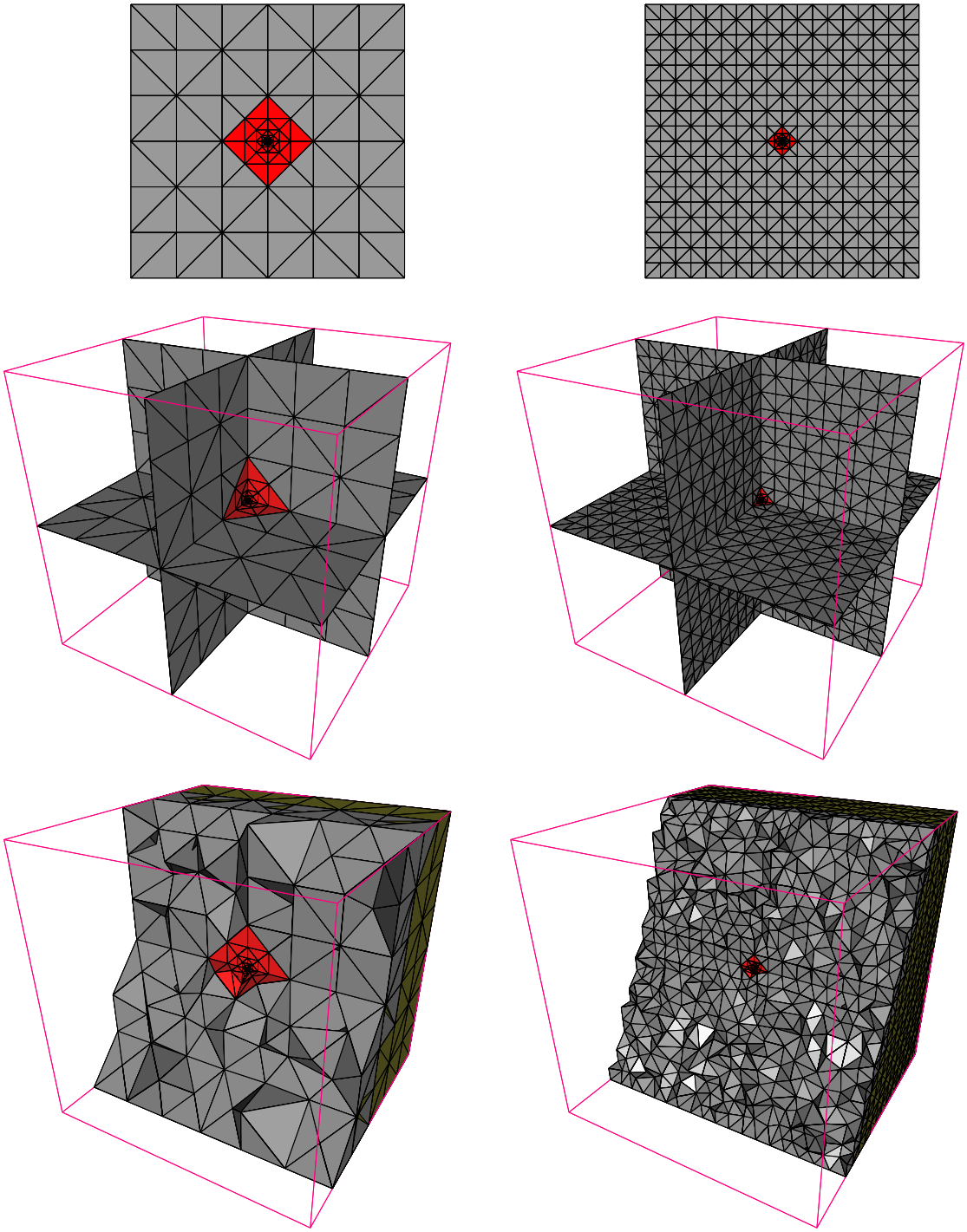


Figure 4.17: Propagation of a standing wave in a cubic PEC cavity: examples of two locally refined meshes used for the numerical convergence study (implicit treatment: *red regions*). From top to bottom: base 2D meshes; interior surfaces of 3D meshes; cross sections of the 3D meshes resulting (on left: 2968 tetraedra, 635 vertices; on right: 40616 tetraedra, 7759 vertices).

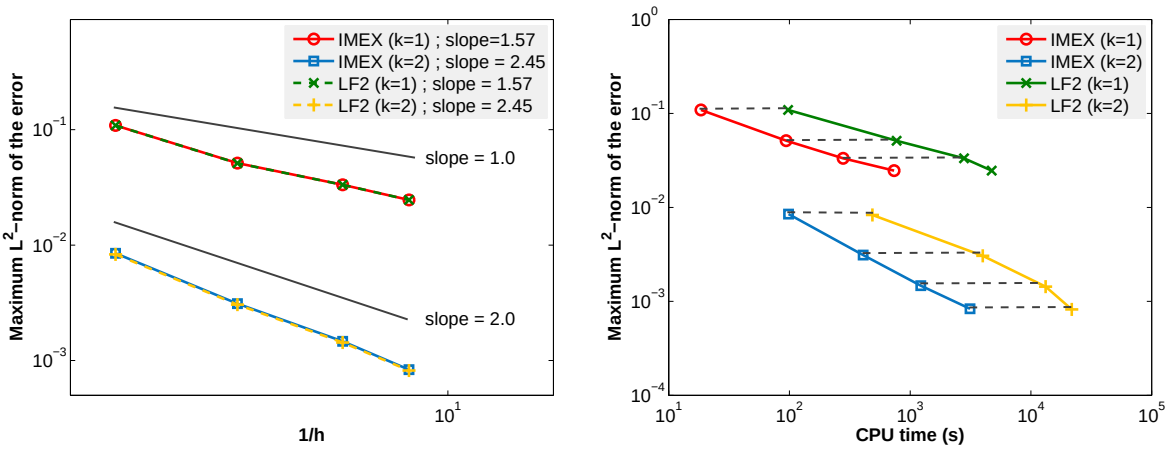


Figure 4.18: Propagation of a standing wave in a cubic PEC cavity: numerical convergence and maximum error (L^2-norm) in function of final CPU time for the locally implicit and fully explicit DGTD-$\mathbb{P}_k$ methods (left - right, respectively).

DGTD-$\mathbb{P}_k$	Matrix order	# nonzero (nz)	Fill-in ratio (percentage)	RAM size (MB)	CPU time (s)
$\mathbb{P}_1$	487392	1247266	0.000525	210	8
$\mathbb{P}_2$	1218480	5437770	0.000366	937	72

Table 4.10: Propagation of a standing wave in a cubic PEC cavity: data and factorization of the matrix of the linear system to be solved for the locally implicit methods, based on DGTD-$\mathbb{P}_k$ methods ($k = 1,2$).

	Fully exp. method (LF2)		**Loc. imp. method (IMEX)**		
$\mathbb{P}_k$	Δt (m)	CPU time (s)	Δt (m)	CPU time (s)	$\frac{\text{CPU}_{(\text{LF2})}}{\text{CPU}_{(\text{IMEX})}}$
$\mathbb{P}_1$	1.4920e-4	4736	4.8600e-3	701	6.76
$\mathbb{P}_2$	9.0496e-5	21810	2.9448e-3	3144	6.94

Table 4.11: Propagation of a standing wave in a cubic PEC cavity: critical time step size and CPU time for the fully explicit method and the locally implicit method, based on DGTD-$\mathbb{P}_k$ method ($k = 1,2$).

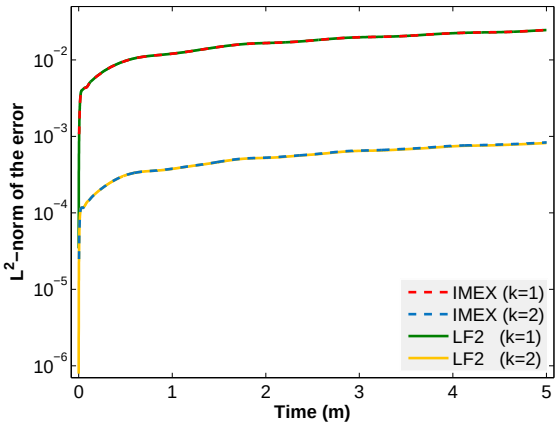


Figure 4.19: Propagation of a standing wave in a cubic PEC cavity: time evolution of the error (L^2-norm) for the locally implicit and fully explicit DGTD-$\mathbb{P}_k$ methods.

4.2.2 Exposure of head tissues to a localized source radiation

We now consider a more realistic problem which is concerned with the simulation of the exposure of a geometrical model of head tissues to an electromagnetic wave emitted by a localized source. Head tissues are segmented and the interfaces of a selected number of tissues (namely, the skin, the skull, the CSF - Cerebro Spinal Fluid and the brain) are triangulated, see Figure 4.20. Then, these surface meshes are used as inputs for the generation of volume meshes. Note that the exterior of the head must also be meshed, up to a certain distance from the skin. The computational domain is here artificially bounded by a sphere on which the Silver-Müller condition is imposed. In the present case, the constructed geometrical model involves four tissues (skin, skull, CSF, brain) and the global tetrahedral mesh consists of 61358 vertices and 366208 tetrahedra, see Figure 4.21. The minimum and maximum lengths of the mesh edges, are respectively, equal to 0.405 mm and 40.372 mm (in the vacuum zone). The characteristics of the tissues are summarized in Table 4.12 where the values of the relative electrical permittivity correspond to a frequency $f = 1800$ MHz and have been obtained from a special purpose online database. A dipolar type source is localized near the right ear of the head yielding a current of the form

$$J^{s,z}(\mathbf{x},t) = Z_0 \delta(\mathbf{x}-\mathbf{x}^s) f(t), \tag{4.18}$$

where Z_0 is the free space intrinsic impedance, δ is the Dirac delta function, $f(t)$ is a sinusoidal varying temporal signal and $\mathbf{x}^s$ is the localization point of the source. The total physical simulation time is set to $T = 2.78 \times 10^{-8}$ ns which corresponds to five periods of the temporal signal.

For the locally implicit method, the implicit region has been chosen such that the number of elements treated implicitly is 5092, which represents 1.4% of the total number of elements, and the critical time step size for the stability of the locally implicit method is about 15.5 times larger than the critical time step size for the stability of the the fully explicit method. Note that the elements for the implicit treatment are essentially localized near the source, i.e. near the right ear of the head (in the free space), see Figure 4.21 and Table 4.13. In Table 4.14 we indicate the number of nonzero

elements and the fill-in ratio percentage for the matrix of the linear system to be solved. We also indicate the total size (memory requirement) of all internal data used during numerical factorization and the total CPU time for analysis and factorization. Note that for the fully implicit scheme based on the DGTD-$\mathbb{P}_1$ method the number of nonzero will be 735944179, which requires too much memory to be a reasonable alternative to the fully explicit method.

A discrete Fourier transform of the components of the electric field is computed during the last period of the simulation. Contour lines of the module of the real part of the discrete Fourier transform of the electric field on the skin, skull, CFS and brain surfaces for the approximate solutions resulting from the IMEX DGTD-$\mathbb{P}_1$ are shown on Figure 4.22. Time evolution of the magnetic and electric components, H^y and E^z, at three selected points in the free space near the right ear, where the source is localized, in the brain and in the free space near the left ear are compared on Figure 4.23, using the fully explicit DGTD-$\mathbb{P}_1$ method and the locally implicit DGTD-$\mathbb{P}_1$ method. One can note on this figure that the approximate solutions resulting from the fully explicit and the locally implicit methods are almost indistinguishable. The total computing time for the fully explicit method is 11 h 40 min, while for the locally implicit method it is 3 h 05 min. Hence, for this problem, the locally implicit DGTD-$\mathbb{P}_1$ method allows a reduction of the computing time by a factor of 3.8.

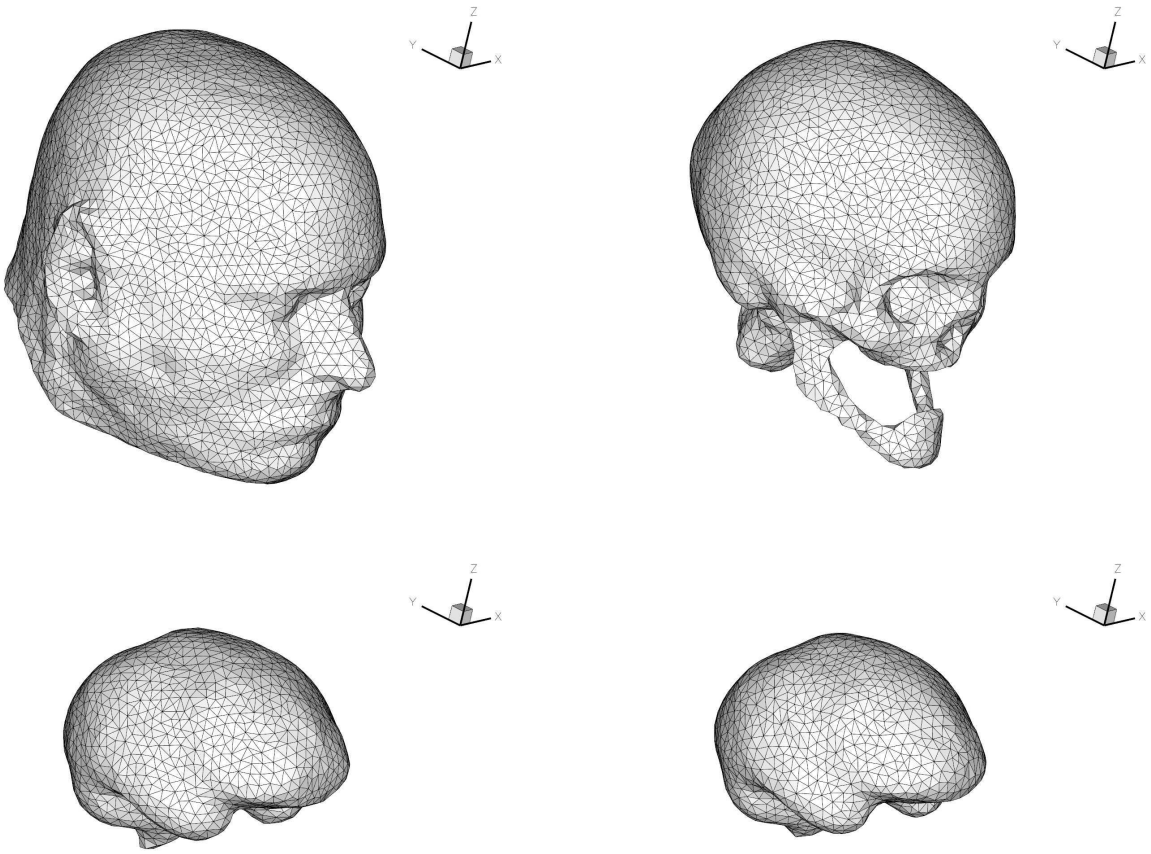


Figure 4.20: Exposure of head tissues to a localized source radiation: surface meshes of the skin, the skull, the CSF and the brain.

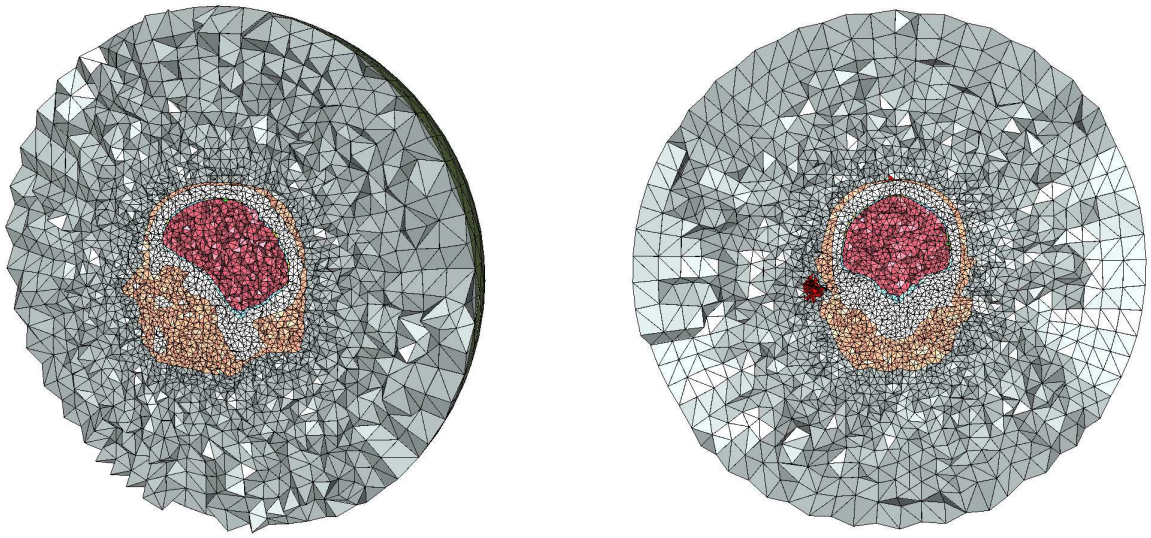


Figure 4.21: Exposure of head tissues to a localized source radiation: cross sections of the 3D mesh (61358 vertices, 366208 tetrahedra, implicit treatment: 5092 tetrahedra, *red region*).

	Vacuum	Skin	Skull	CSF	Brain
ε	1.00	43.85	15.56	67.20	43.55
σ (S·m^{-1})	0.00	1.23	0.43	2.92	1.15
λ (mm)	166.66	27.73	42.25	20.33	25.26

Table 4.12: Exposure of head tissues to a localized source radiation: electromagnetic characteristics of tissues.

Vacuum	Skin	Skull	CSF	Brain	Total
4209	102	21	720	40	5092

Table 4.13: Exposure of head tissues to a localized source radiation: distribution of elements in the implicit region.

Method	Matrix order	# nonzero (nz)	Fill-in ratio (percentage)	RAM size (MB)	CPU time (s)
DGTD-$\mathbb{P}_1$	4394496	15048090	0.000078	1298	95

Table 4.14: Exposure of head tissues to a localized source radiation: data and factorization of the matrix of the linear system to be solved for the locally implicit methods, based on DGTD-$\mathbb{P}_1$ method.

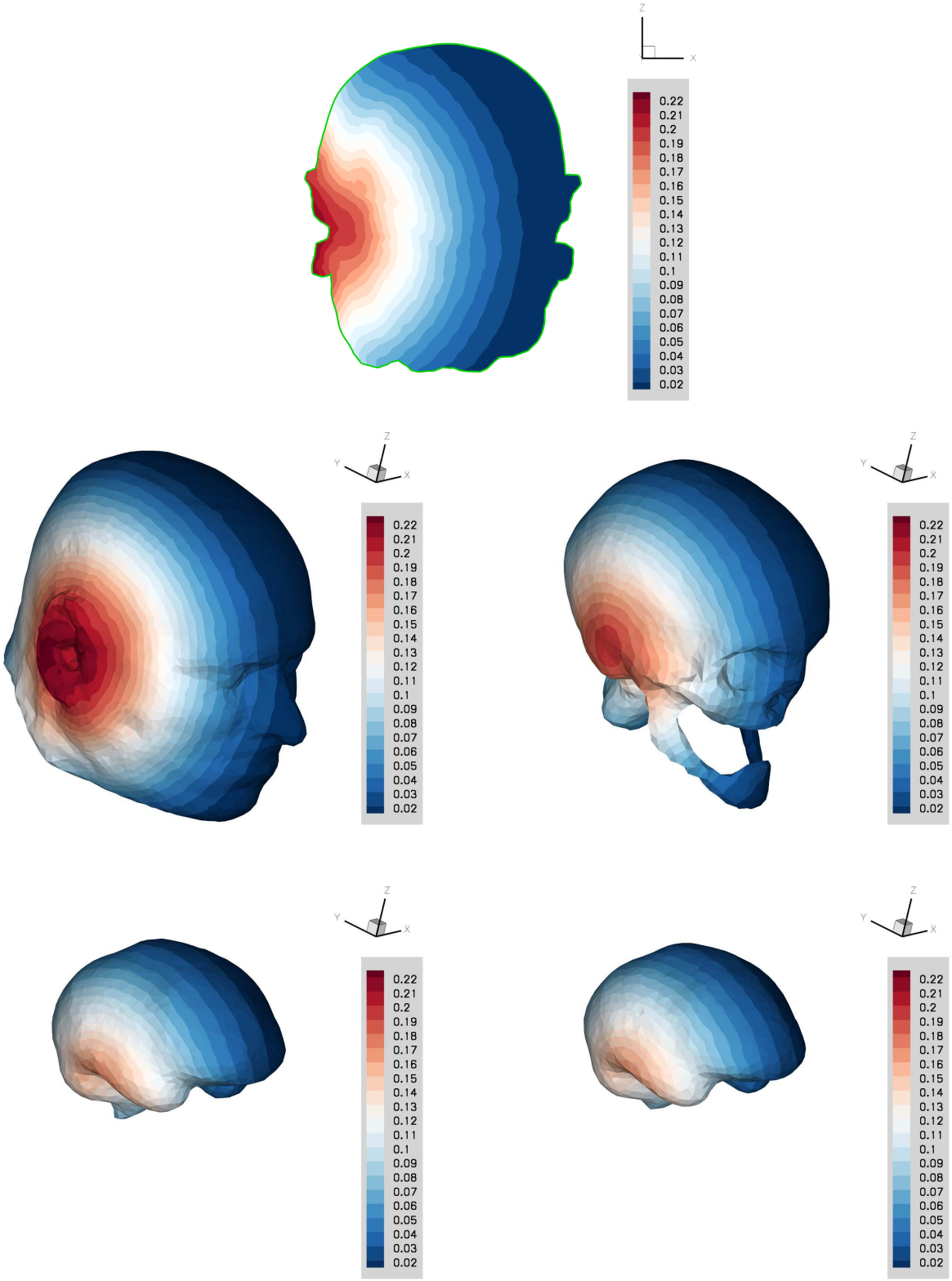


Figure 4.22: Exposure of head tissues to a localized source radiation: contour lines of the module of the real part of the DFT of the electric field, i.e. $\sqrt{(\mathbf{E}^x_{four})^2 + (\mathbf{E}^y_{four})^2 + (\mathbf{E}^z_{four})^2}$, on the skin, skull, CSF and brain surfaces for the approximate solutions resulting from the locally implicit DGTD-$\mathbb{P}_1$ method, over the last period of simulation.

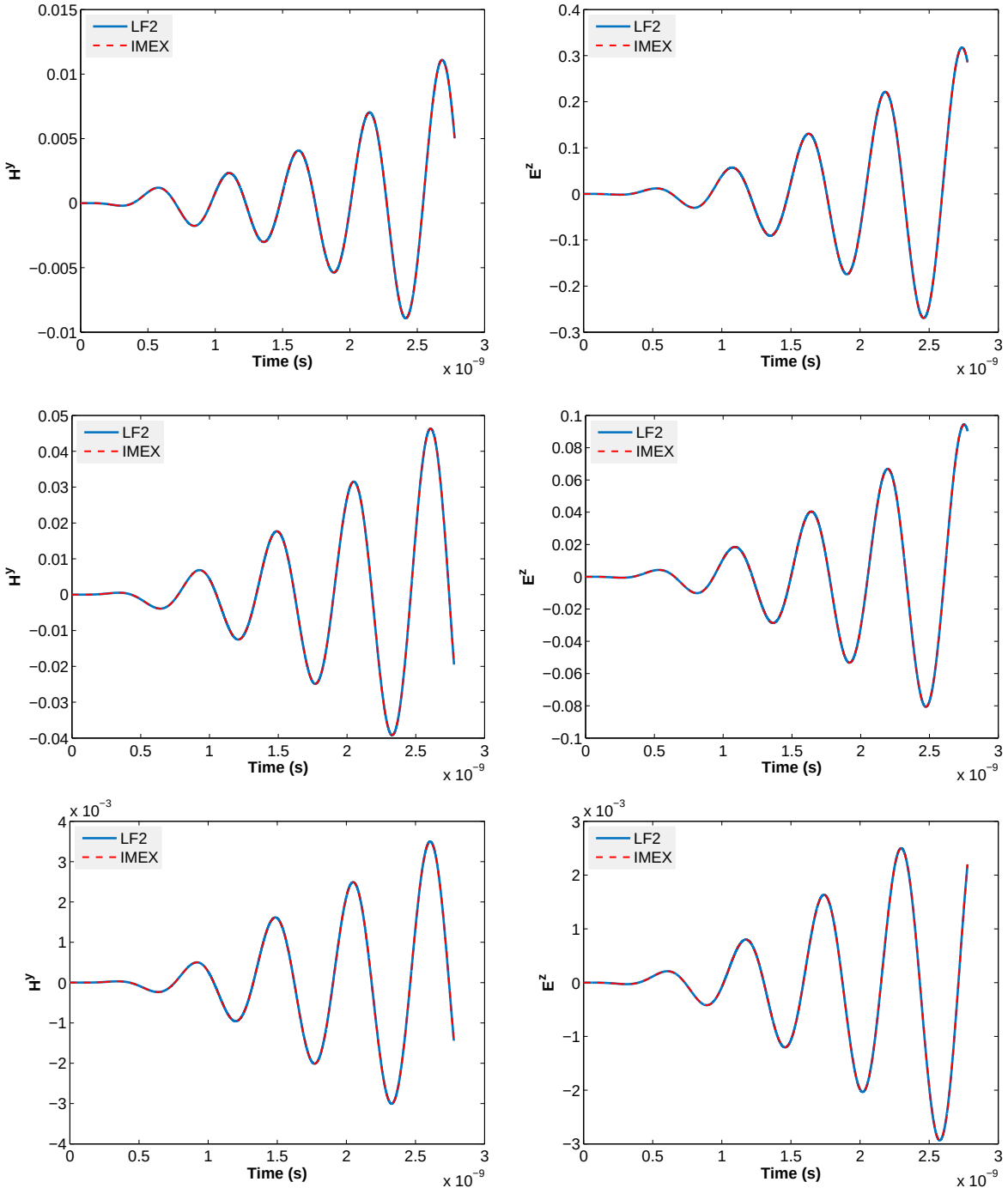


Figure 4.23: Exposure of head tissues to a localized source radiation: time evolution of the H^y and E^z components at three selected points. From top to bottom $\mathbf{x} = (-0.125, 0.025, -0.015)$, $\mathbf{x} = (0.0, 0.025, -0.015)$ and $\mathbf{x} = (0.125, 0.025, -0.015)$.

CHAPTER 5

High-order locally implicit time integration strategies

Contents

The topic of this chapter is to propose higher order time integration techniques based on the second-order locally implicit method (3.62) to fully exploit the attractive features of the implicit-explicit approach combined with a DG discretisation which allows to easily increase the spatial convergence order. Starting from the locally implicit method (3.62), we focus our analyses on strategies which are essentially suitable combinations of this basic method applied with different time step sizes such that their computer implementation are particularly easy.

One possibility is to exploit composition methods, known to be accurate with well-designed composition coefficients to minimize truncation errors, see e.g. [Hairer & Wanner 1996, Hairer *et al.* 2010, McLachlan 1995, Suzuki 1990, Yoshida 1990]. The composition methods with orders beyond two for solving Maxwell's equations are restricted to problems with small (non-stiff) dissipative terms [Botchev & Verwer 2009]. Furthermore in the presence of source functions, the convergence order may be lower than the chosen composition order [Verwer 2012]. In the latter reference, Verwer proposed an analysis of fourth-order composition methods. Starting from the second-order explicit scheme (3.6) as the basic method in a composition and a source function perturbation to solve the semi-discrete Maxwell equations, he obtains at least a third-order scheme and a fourth-order one with additional sufficient conditions.

Another possibility is to rely on local or global Richardson extrapolations which are easy to implement and straightforward to parallelize, see e.g. [Faragó *et al.* 2010, Hairer *et al.* 1993]. The idea of Richardson, announced in his classical paper [Richardson 1910] which treats partial differential equations, and explained in full detail in [Richardson 1927], is to use the known behavior of the error as a function of step size. Then, applying the same basic method for solving an initial value problem by using different step sizes and combining appropriately the obtained numerical solutions at each time step, the convergence order of the method can be increased. These high-order extensions are well-known in the numerical solution of ODEs and have been already considered with the fully explicit method (3.6) as the basic method to solve

damped Maxwell equations [Botchev & Verwer 2009]. With Dirichlet boundary conditions the authors of [Botchev & Verwer 2009] observe that the global approach, which preserves the CFL condition of the basic method to assume the stability of the resulting scheme, does not suffer from order reduction; this is not the case for composition methods and local extrapolations. However, when the reduction order is unavoidable, they advocate the local approach that allows to eliminate error terms instantaneously. In this chapter we propose to apply these techniques with the second-order locally implicit method from [Verwer 2010] for time-dependent Maxwell equations.

In Section 5.1, we introduce the concept of symmetric methods. Theoretical analyses of this concept and many examples can be found in [Hairer *et al.* 2010]. Next, we present composition methods and Richardson extrapolations. Particularly, we observe that the symmetry of the locally implicit method (3.62) is an attractive feature for applying such strategies to increase the temporal convergence order of this basic second-order method. Finally, Section 5.2 deals with two-dimensional problems to study the numerical convergence of the resulting locally implicit methods.

5.1 High-order time integration methods

5.1.1 Symmetric methods

Symmetry is an essential property of numerical methods with regards to order of accuracy and geometric properties of the solution. Before giving the definition of a symmetric method we need to introduce several concepts as exact and approximate flow maps and adjoint methods.

5.1.1.1 The group of flow maps

Let us take an autonomous ODE system

$$\dot{y} = f(y), \quad y \in \mathbb{R}^d, f : \mathbb{R}^d \to \mathbb{R}^d. \tag{5.1}$$

Given a fixed time τ, we can consider any point of phase space y_0, as a starting point of a trajectory $y(t, y_0)$ which is continued up to time τ, assuming the solution exists on the entire interval. Solving the differential equation defined a map from the starting points of trajectories to their end points. We then define the flow map associated to the differential equation (5.1)

$$\varphi_\tau(y_0) = y(\tau), \tag{5.2}$$

where $y(t)$ is the solution of the initial value problem

$$\dot{y}(t) = f(y), \quad y(0) = y_0, \quad t \in [0, \tau]. \tag{5.3}$$

The set of flow maps $\{\varphi_t,\ t \in \mathbb{R}\}$ is a one-parameter group with the (commutative) group operation being composition of maps and in particular φ_t satisfies $\varphi_0 = id$ and

$$\varphi_{-t} = \varphi_t^{-1}. \tag{5.4}$$

We could regularly sample the exact solution on a succession of time intervals of length Δt according to the rule

$$y(t_{n+1}) = \varphi_{\Delta t}(y(t_n)). \tag{5.5}$$

If $y(t, y_0)$ represents the solution of the initial value problem (5.3) we can write

$$y(\Delta t, y_0) = \varphi_{\Delta t}(y_0),\ y(2\Delta t, y_0) = \varphi_{\Delta t}(\varphi_{\Delta t}(y_0)) = \varphi_{\Delta t} \circ \varphi_{\Delta t}(y_0),\ \text{etc.} \tag{5.6}$$

This means that we can view the iteration of $\varphi_{\Delta t}$ as snapshots of the solution at equally spaced points in time. The concept of flow map is very helpful in exploring the qualitative behavior of numerical methods, since it is often possible to think of numerical methods as approximations of the flow map.

5.1.1.2 Approximate flow maps

Numerical methods for (5.1) implement numerical flow maps $\Phi_{\Delta t} : y_n \to y_{n+1}$ which, for small enough Δt, approximate $\varphi_{\Delta t}$, and where y_n represents a numerical approximation of $y(t_n)$ as usual. The approximation would satisfy

$$\Phi_0 = Id \text{ and } y_1 = \Phi_{\Delta t}(y_0),\ y_2 = \Phi_{\Delta t}(\Phi_{\Delta t}(y_0)) = \Phi_{\Delta t} \circ \Phi_{\Delta t}(y_0),\ \text{etc.} \tag{5.7}$$

A fundamental difference between the exact flow map and its numerical approximation is that the mappings $\Phi_{\Delta t}$ do not form a one-parameter group and the property (5.4) is not, in general, shared by the one-step map $\Phi_{\Delta t}$ of a numerical method.

5.1.1.3 Adjoint and symmetric methods

The key for understanding the concept of symmetric methods is the concept of adjoint method. A numerical flow map is usually also defined for negative time step $-\Delta t$, and by the inverse function theorem it is invertible for sufficiently small Δt.

Definition 5.1.1. *The adjoint method $\Phi^*_{\Delta t}$ of a method $\Phi_{\Delta t}$ is the inverse of the original method with reversed time step $-\Delta t$, i.e.*

$$\Phi^*_{\Delta t} = \Phi^{-1}_{-\Delta t}. \tag{5.8}$$

A numerical method $\Phi_{\Delta t}$ is then called symmetric if $\Phi^*_{\Delta t} = \Phi_{\Delta t}$.

For the computation of the adjoint method we observe that $y_1 = \Phi^*_{\Delta t}(y_0)$ is implicitly defined by $\Phi_{-\Delta t}(y_1) = y_0$, i.e. y_1 is the value which yields y_0 when the method $\Phi_{\Delta t}$ is applied with negative time step $-\Delta t$.

Example 5.1.1. *Let $\Phi_{\Delta t}$ the mapping of the explicit Euler method for the system (5.1), i.e.,*

$$\Phi_{\Delta t} : y_n \to y_{n+1} = y_n + \Delta t f(y_n). \tag{5.9}$$

Exchanging $y_n \leftrightarrow y_{n+1}$ and $\Delta t \leftrightarrow -\Delta t$ yields $y_n = y_{n+1} - \Delta t f(y_{n+1})$. It follows that the adjoint method of $\Phi_{\Delta t}$ is

$$\Phi^*_{\Delta t} : y_n \to y_{n+1} = y_n + \Delta t f(y_{n+1}), \tag{5.10}$$

which is the implicit Euler method. Therefore the adjoint method of explicit Euler is implicit Euler and the method is not symmetric.

Example 5.1.2. *Let $\Phi_{\Delta t}$ the mapping of the implicit midpoint rule method for the system (5.1), i.e.,*

$$\Phi_{\Delta t} : y_n \to y_{n+1} = y_n + \Delta t f\left(\frac{y_{n+1}+y_n}{2}\right). \tag{5.11}$$

We derive the adjoint method by exchanging $y_n \leftrightarrow y_{n+1}$ and $\Delta t \leftrightarrow -\Delta t$. The formula is unaltered after this exchange, thus $\Phi^*_{\Delta t} = \Phi_{\Delta t}$, i.e., the implicit midpoint rule is a symmetric method.

Example 5.1.3. *The fully explicit method (3.6), the fully implicit method (3.8) and the locally implicit method (3.62) are symmetric methods. Indeed, the three numerical methods remain unaltered after exchanging $E^n \leftrightarrow E^{n+1}$, $H^n \leftrightarrow H^{n+1}$ and $\Delta t \leftrightarrow -\Delta t$. Note that unlike the leap-frog method written in the three-stage form (3.6), the leap-frog method written in the two-stage form, i.e.,*

$$\Phi_{\Delta t} : \left\{ \begin{array}{rcl} \dfrac{E^{n+1}-E^n}{\Delta t} & = & SH^{n+1/2} - \frac{1}{2}D\left(E^{n+1}+E^n\right) + j^s\left(t_{n+1/2}\right), \\ \dfrac{H^{n+3/2}-H^{n+1/2}}{\Delta t} & = & -S^T E^{n+1}, \end{array} \right. \tag{5.12}$$

is not symmetric, by exchanging $E^n \leftrightarrow E^{n+1}$, $H^{n+1/2} \leftrightarrow H^{n+3/2}$ and $\Delta t \leftrightarrow -\Delta t$ into (5.12) it follows that $\Phi^*_{\Delta t} \neq \Phi_{\Delta t}$.

The adjoint method satisfies the usual properties such as $\left(\Phi^*_{\Delta t}\right)^* = \Phi_{\Delta t}$ and $\left(\Phi_{\Delta t} \circ \Psi_{\Delta t}\right)^* = \Phi^*_{\Delta t} \circ \Psi^*_{\Delta t}$ for any two one step methods. The adjoint method has also the same order as the original method.

Theorem 5.1.1. *Let $\varphi_{\Delta t}$ be the exact flow of (5.1) and $\Phi_{\Delta t}$ a one-step method of order p satisfying*

$$\Phi_{\Delta t}(y_0) = \varphi_{\Delta t}(y_0) + C(y_0)\Delta t^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right). \tag{5.13}$$

Then the adjoint method $\Phi^*_{\Delta t}$ has the same order p and

$$\Phi^*_{\Delta t}(y_0) = \varphi_{\Delta t}(y_0) + (-1)^p C(y_0)\Delta t^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right). \tag{5.14}$$

Proof. The main idea of the proof is depicted in Figure 5.1. From a given initial value y_0 we compute $\varphi_{\Delta t}(y_0)$ and $y_1 = \Phi^*_{\Delta t}(y_0)$ whose difference e^* is the local error of $\Phi^*_{\Delta t}$. Indeed by definition of the adjoint method and the flow property (5.4) we write

$$e^* = y_1 - \varphi_{\Delta t}(y_0) = \Phi^*_{\Delta t}(y_0) - \varphi_{\Delta t}(y_0) = \Phi^{-1}_{-\Delta t}(y_0) - \varphi^{-1}_{-\Delta t}(y_0). \tag{5.15}$$

This error is then projected by $\Phi_{-\Delta t}$ to become e i.e.

$$e = \Phi_{-\Delta t}(y_1) - \Phi_{-\Delta t}\left(\varphi_{\Delta t}(y_0)\right) = y_0 - \Phi_{-\Delta t}\left(\varphi_{\Delta t}(y_0)\right). \tag{5.16}$$

With the flow property (5.4) we write

$$e = \varphi_{-\Delta t}\left(\varphi_{\Delta t}(y_0)\right) - \Phi_{-\Delta t}\left(\varphi_{\Delta t}(y_0)\right), \tag{5.17}$$

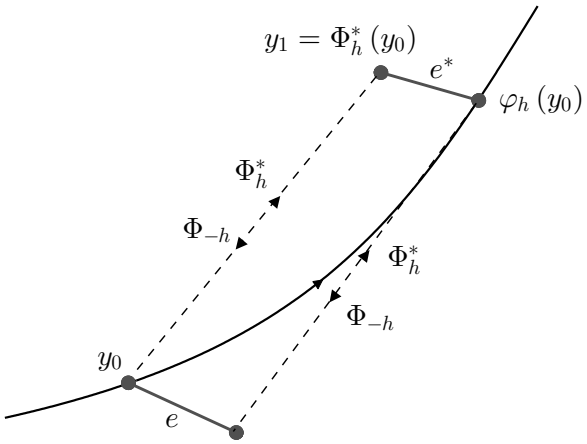


Figure 5.1: Schematic diagram of the proof of Theorem 5.1.1.

and we see that $-e$ is the local error of $\Phi_{-\Delta t}$, i.e. by hypothesis (5.13)

$$e = (-1)^p C(\varphi_{\Delta t}(y_0)) \Delta t^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right). \tag{5.18}$$

Since $\varphi_{\Delta t}(y_0) = y_0 + \mathscr{O}(\Delta t)$ and $e = (I + \mathscr{O}(\Delta t)) e^*$, it follows that

$$e^* = (-1)^p C(y_0) \Delta t^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right), \tag{5.19}$$

which proves (5.14). □

Theorem 5.1.1 implies that a symmetric method is necessarily of even order $p = 2q$, since $\Phi_{\Delta t}(y_0) = \Phi_{\Delta t}^*(y_0)$ means that $C(y_0) = (-1)^p C(y_0)$, and therefore $C(y_0)$ can be different from zero only for even p. This property of symmetric methods plays a key role in the construction of composition methods and it also explains why symmetric methods are used in conjunction with Richardson extrapolation techniques.

5.1.2 Symmetric composition of symmetric methods

High-order composition methods have been extensively studied for geometric composition, see e.g. [Hairer & Wanner 1996, Hairer *et al.* 2010, McLachlan 1995, Suzuki 1990, Yoshida 1990].

Definition 5.1.2. *Let $\Phi_{\Delta t}$ be a basic method and $\alpha_1, \cdots, \alpha_s$ real numbers. We call its composition with step sizes $\alpha_1 \Delta t, \cdots, \alpha_s \Delta t$, the corresponding composition method, i.e.*

$$\Psi_{\Delta t} = \Phi_{\alpha_s \Delta t} \circ \cdots \circ \Phi_{\alpha_1 \Delta t}. \tag{5.20}$$

Theorem 5.1.2. *Let $\Phi_{\Delta t}$ be a one-step method of order p. If*

$$\begin{aligned} \alpha_1 + \cdots + \alpha_s &= 1, \\ \alpha_1^{p+1} + \cdots + \alpha_s^{p+1} &= 0, \end{aligned} \tag{5.21}$$

then the composition method (5.20) is at least of order $p+1$.

Proof. The main idea of the proof is depicted in Figure 5.2. Let $\Phi_{\Delta t}$ be a one-step method of order p. From a given initial value y_0 we compute $y_1 = \Phi_{\alpha_1 \Delta t}(y_0)$, $y_2 = \Phi_{\alpha_2 \Delta t}(y_1)$ and $y_3 = \Phi_{\alpha_3 \Delta t}(y_2)$., by hypothesis (5.13),

$$\begin{aligned}
\Phi_{\alpha_1 \Delta t}(y_0) &= \varphi_{\alpha_1 \Delta t}(y_0) + C(y_0)(\alpha_1 \Delta t)^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right), \\
\Phi_{\alpha_2 \Delta t}(y_1) &= \varphi_{\alpha_2 \Delta t}(y_1) + C(y_1)(\alpha_2 \Delta t)^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right), \\
\Phi_{\alpha_3 \Delta t}(y_2) &= \varphi_{\alpha_3 \Delta t}(y_2) + C(y_2)(\alpha_3 \Delta t)^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right).
\end{aligned} \tag{5.22}$$

For all i we denote by e_i the local error of $\Phi_{\alpha_i \Delta t}$, i.e.,

$$e_i = \Phi_{\alpha_i \Delta t}(y_{i-1}) - \varphi_{\alpha_i \Delta t}(y_{i-1}). \tag{5.23}$$

Since $y_i = y_0 + \mathscr{O}(\Delta t)$ and $E_i = (I + \mathscr{O}(\Delta t))\, e_i$, it follows from $\sum \alpha_i = 1$ and (5.22) that

$$\Psi_{\Delta t}(y_0) - \varphi_{\Delta t}(y_0) = E_1 + E_2 + E_3 = C(y_0)\left(\alpha_1^{p+1} + \alpha_2^{p+1} + \alpha_3^{p+1}\right)\Delta t^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right). \tag{5.24}$$

Consequently with condition (5.21) the $\mathscr{O}\left(\Delta t^{p+1}\right)$-term vanishes, i.e. that the composition method is at least of order $p+1$. □

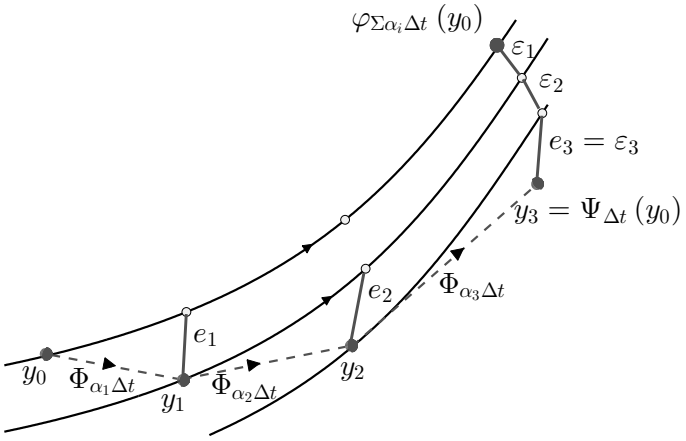


Figure 5.2: Schematic diagram of the proof of Theorem 5.1.2.

Theorem 5.1.2 is the key to understand the Triple Jump compositions, as proposed in [Forest 1989, Creutz & Gocksch 1989, Suzuki 1990, Yoshida 1990]. Starting from a symmetric method $\Phi_{\Delta t}$ of (even) order $2q$, the composition method obtained for the set of coefficients

$$\alpha_1 = \alpha_3 = \frac{1}{2 - 2^{1/(2q+1)}}, \quad \alpha_2 = -\frac{2^{1/(2q+1)}}{2 - 2^{1/(2q+1)}}, \tag{5.25}$$

is symmetric since

$$\Psi^*_{\Delta t} = \Phi^*_{\alpha_1 \Delta t} \circ \Phi^*_{\alpha_2 \Delta t} \circ \Phi^*_{\alpha_3 \Delta t} = \Phi_{\alpha_3 \Delta t} \circ \Phi_{\alpha_2 \Delta t} \circ \Phi_{\alpha_1 \Delta t} = \Psi_{\Delta t}, \tag{5.26}$$

and with Theorem 5.1.2 of order $2q+1$. Since the order of symmetric method is even, $\Psi_{\Delta t}$ is in fact of order $p = 2q+2$. This procedure can be repeated recursively to construct arbitrarily high-order symmetric methods of orders $2q+2$, $2q+4$, $2q+6$, etc., with respectively 3, 9, 27-stage, etc., symmetric composition method. However, the construction is known for not being the most efficient, for the combined coefficients become large, some of which being negative. A partial remedy is to envisage compositions with $s = 5$. We hereby give the set of coefficients obtained by Suzuki [Suzuki 1990]

$$\alpha_1 = \alpha_2 = \alpha_4 = \alpha_5 = \frac{1}{4 - 4^{1/(2q+1)}}, \quad \alpha_3 = -\frac{4^{1/(2q+1)}}{4 - 4^{1/(2q+1)}}, \tag{5.27}$$

which give rise to very efficient methods for $q = 1$ and $q = 2$.

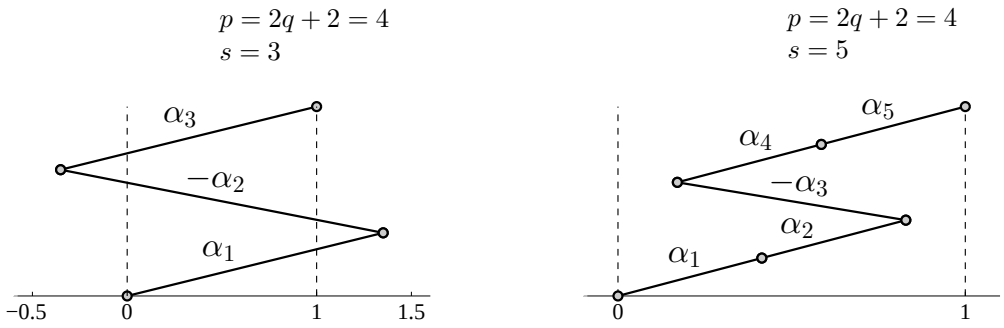


Figure 5.3: The Triple Jump (on left) and Suzuki's fractal (on right) of order four.

In [Botchev & Verwer 2009, Verwer 2012], the authors remark that negative coefficients cannot be avoided for composition methods with orders beyond two [Suzuki 1990], and then conduction terms limit the critical step size which guarantees the stability of composition methods. This property thus restricts such methods to problems with small (non-stiff) dissipative terms. Furthermore, for $s = 5$ the stability region is larger due to smaller coefficients α_k. Taking into account the workload, five substages compared to three, the advantage of a larger stability region still exists. The most efficient high-order composition methods are nevertheless those obtained by solving the full system of order conditions, without going through the intermediate steps described above. This requires much more effort though, first derive the order conditions, and then to solve the resulting polynomial system. We refer to [Murua1 & SanzSerna 1999] on the use of $\mathbf{B}_\infty$-series and to [Hairer *et al.* 2010] for numerical comparisons.

In this study we are only interested in composition methods at most of order four with set of coefficients (5.25) and (5.27) for $s = 3$ and $s = 5$, respectively. Let $\Phi_{\Delta t}^{IMEX2}$ be the second-order locally implicit method (3.62) and $\Phi_{\Delta t}^{CO4}$ be a fourth-order composition method, following the construction presented in [Verwer 2012] we write

$$\Phi_{\Delta t}^{CO4} = \Phi_{\alpha_s \Delta t}^{IMEX2} \circ \cdots \circ \Phi_{\alpha_1 \Delta t}^{IMEX2}, \tag{5.28}$$

where $\alpha_1 + \cdots + \alpha_s = 1$ and $\alpha_1^3 + \cdots + \alpha_s^3 = 0$. Denote $(E^{\alpha_0}, H^{\alpha_0}) = (E^n, H^n)$ and $(E^{\alpha_s}, H^{\alpha_s}) =$

(E^{n+1}, H^{n+1}), the composition scheme can be written as

$$\text{for } k=1:s \begin{cases} \dfrac{H^{\beta_k}-H^{\alpha_{k-1}}}{\alpha_k\Delta t} &= -\dfrac{1}{2}S^T E^{\alpha_{k-1}}, \\ \dfrac{E^{\alpha_k}-E^{\alpha_{k-1}}}{\alpha_k\Delta t} &= S_0 H^{\beta_k} + \dfrac{1}{2}S_1(H^{\alpha_{k-1}}+H^{\alpha_k}) \\ &- \dfrac{1}{2}D(E^{\alpha_{k-1}}+E^{\alpha_k}) + \dfrac{1}{2}(j^s(t_{\alpha_{k-1}})+j^s(t_{\alpha_k})), \\ \dfrac{H^{\alpha_k}-H^{\beta_k}}{\alpha_k\Delta t} &= -\dfrac{1}{2}S^T E^{\alpha_k}, \end{cases} \tag{5.29}$$

with time levels $t_{\alpha_0}=t_n$, $t_{\beta_k}=t_{\alpha_{k-1}}+\alpha_k\Delta t/2$, $t_{\alpha_k}=t_{\alpha_{k-1}}+\alpha_k\Delta t$, spanning the interval $[t_n,t_{n+1}]$, for $k=1,\cdots,s$. The implementations of the composition method on the interval $[t_n,t_{n+1}]$ is depicted in Figure 5.4, for $s=3$. The composition method now reads

$$\text{for } k=1:s \begin{cases} H^{\beta_k} &= H^{\alpha_{k-1}} - \dfrac{\alpha_k\Delta t}{2}S^T E^{\alpha_{k-1}}, \\ \mathscr{M}_k E^{\alpha_k} &= b_k, \\ H^{\alpha_k} &= H^{\beta_k} - \dfrac{\alpha_k\Delta t}{2}S^T E^{\alpha_k}, \end{cases} \tag{5.30}$$

where

$$\begin{aligned} \mathscr{M}_k &= I + \frac{\alpha_k\Delta t}{2}D + \frac{(\alpha_k\Delta t)^2}{4}S_1 S^T, \\ b_k &= E^{\alpha_{k-1}} + \alpha_k\Delta t\, S_0 H^{\beta_k} + \frac{\alpha_k\Delta t}{2}S_1\left(H^{\alpha_{k-1}}+H^{\beta_k}\right) - \frac{\alpha_k\Delta t}{2}DE^{\alpha_{k-1}} \\ &+ \frac{\alpha_k\Delta t}{2}\left(j^s(t_{\alpha_{k-1}})+j^s(t_{\alpha_k})\right). \end{aligned} \tag{5.31}$$

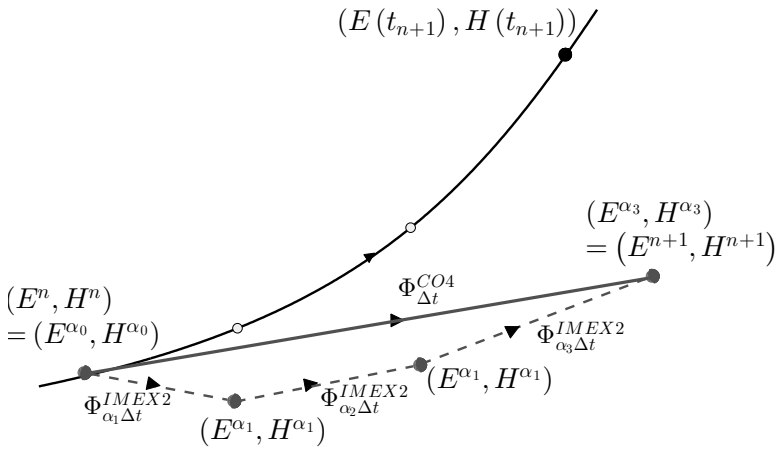


Figure 5.4: Schematic diagram of the composition method (5.28) based on the second-order locally implicit method (3.62), on the interval $[t_n,t_{n+1}]$ and for $s=3$.

5.1.3 Richardson extrapolation

It is known from the literature that applying the same basic method for solving an initial value problem as (5.3) by using two different step sizes and combining appropriately the obtained numerical solutions at each time step, we can increase the convergence order of the method (e.g. [Faragó *et al.* 2010, Hairer *et al.* 1993]). In this part we are going to apply this procedure, widely known as Richardson extrapolation. The idea of Richardson, announced in his classical paper [Richardson 1910] which treats mainly partial differential equations, and explained in full detail in [Richardson 1927], is to use more carefully the known behavior of the error as a function of step size. Typically, for low precision, extrapolation methods have not been competitive with Runge-Kutta methods. For high precision, however, the arbitrary order means that they can be arbitrarily faster than fixed order methods for very precise tolerances.

Let us take an autonomous ODE system (5.1) and the exact flow map (5.2) associated to the initial value problem (5.3). From the given initial value y_0 and stepsize $\Delta t/2$ we compute two steps, using a fixed one-step method of order p, and obtain the numerical results $\Phi_{\Delta t/2}(y_0)$ and $\Phi_{\Delta t/2} \circ \Phi_{\Delta t/2}(y_0)$. We then compute, starting from y_0, one step with stepsize Δt to obtain the solution $\Phi_{\Delta t}(y_0)$. These two computations are depicted in Figure 5.5. From Theorem 5.13 we write the error of $\Phi_{\Delta t/2}(y_0)$ as

$$e_1 = \Phi_{\Delta t/2}(y_0) - \varphi_{\Delta t/2}(y_0) = C(y_0)\left(\frac{\Delta t}{2}\right)^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right). \tag{5.32}$$

The error of $\Phi_{\Delta t/2} \circ \Phi_{\Delta t/2}(y_0)$ is composed of two parts the transported error of the first step $\varepsilon_1 = (I + \mathscr{O}(\Delta t)\, e_1$ and the local error of the second step e_2, which is the same as (5.32) evaluated at $\Phi_{\Delta t/2}(y_0) = y_0 + \mathscr{O}(\Delta t)$. Thus we obtain

$$\Phi_{\Delta t/2} \circ \Phi_{\Delta t/2}(y_0) - \varphi_{\Delta t}(y_0) = C(y_0)\left(\frac{1}{2^{p+1}} + \frac{1}{2^{p+1}}\right)\Delta t^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right). \tag{5.33}$$

Similarly to (5.32) the error of $\Phi_{\Delta t}$ is

$$e = \Phi_{\Delta t}(y_0) - \varphi_{\Delta t}(y_0) = C(y_0)\,\Delta t^{p+1} + \mathscr{O}\left(\Delta t^{p+2}\right). \tag{5.34}$$

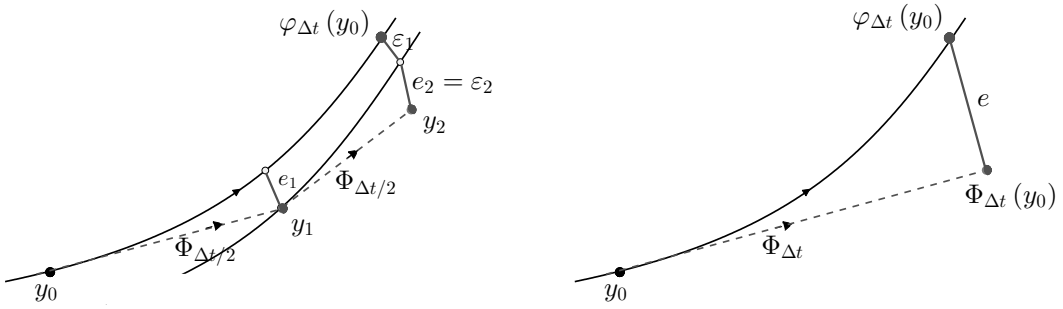


Figure 5.5: Schematic diagrams of $\Phi_{\Delta t/2} \circ \Phi_{\Delta t/2}$ (on left) and $\Phi_{\Delta t}$ (on right).

Theorem 5.1.3. *Let $\Psi_{\Delta t}$ be the Richardson extrapolation method of a one-step method $\Phi_{\Delta t}$ of order p. The method $\Psi_{\Delta t}$ defined by*

$$\Psi_{\Delta t}(y_0) = \Phi_{\Delta t/2} \circ \Phi_{\Delta t/2}(y_0) + \frac{\Phi_{\Delta t/2} \circ \Phi_{\Delta t/2}(y_0) - \Phi_{\Delta t}(y_0)}{2^p - 1} \tag{5.35}$$

is at least of order $p+1$.

Proof. With definition (5.35) from (5.33) and (5.34) we write

$$\begin{aligned}
\Psi_{\Delta t}(y_0) &= \varphi_{\Delta t}(y_0) + C(y_0)\left(\frac{1}{2^{p+1}} + \frac{1}{2^{p+1}}\right)\Delta t^{p+1} \\
&+ \frac{\varphi_{\Delta t}(y_0) + C(y_0)\left(\frac{1}{2^{p+1}} + \frac{1}{2^{p+1}}\right)\Delta t^{p+1} - \varphi_{\Delta t}(y_0) - C(y_0)\Delta t^{p+1}}{2^p - 1} \\
&+ \mathcal{O}\left(\Delta t^{p+2}\right), \\
&= \varphi_{\Delta t}(y_0) + C(y_0)\left(\frac{1}{2^p}\right)\Delta t^{p+1} + \frac{C(y_0)\left(\frac{1-2^p}{2^p}\right)\Delta t^{p+1}}{2^p - 1} + \mathcal{O}\left(\Delta t^{p+2}\right).
\end{aligned} \tag{5.36}$$

Consequently the $\mathcal{O}\left(\Delta t^{p+1}\right)$-term vanishes, i.e., the Richardson extrapolation method is at least of order $p+1$. □

Starting from a symmetric method $\Phi_{\Delta t}$ of (even) order $2q$, the Richardson extrapolation method $\Psi_{\Delta t}$ is also symmetric since

$$\begin{aligned}
\Psi_{\Delta t}^* &= \left(\Phi_{\Delta t/2} \circ \Phi_{\Delta t/2}\right)^* + \frac{\left(\Phi_{\Delta t/2} \circ \Phi_{\Delta t/2}\right)^* - \Phi_{\Delta t}^*}{2^p - 1}, \\
&= \Phi_{\Delta t/2}^* \circ \Phi_{\Delta t/2}^* + \frac{\Phi_{\Delta t/2}^* \circ \Phi_{\Delta t/2}^* - \Phi_{\Delta t}^*}{2^p - 1}, \\
&= \Phi_{\Delta t/2} \circ \Phi_{\Delta t/2} + \frac{\Phi_{\Delta t/2} \circ \Phi_{\Delta t/2} - \Phi_{\Delta t}}{2^p - 1}, \\
&= \Psi_{\Delta t},
\end{aligned} \tag{5.37}$$

and with Theorem 5.1.3 of order $2q+1$. Since the order of symmetric method is even, $\Psi_{\Delta t}$ is in fact of order $p = 2q+2$.

We examine the extension of the second-order locally implicit method $\Phi_{\Delta t}^{IMEX2}$ to fourth-order through the Richardson extrapolation technique for symmetric methods. Let $\Phi_{\Delta t}^{REX4}$ be the Richardson extrapolation of the IMEX method (3.62), we write

$$\Phi_{\Delta t}^{REX4} = \frac{4}{3}\Phi_{\Delta t/2}^{IMEX2} \circ \Phi_{\Delta t/2}^{IMEX2} - \frac{1}{3}\Phi_{\Delta t}^{IMEX2}. \tag{5.38}$$

Richardson extrapolation can be implemented in two different ways: local (or active) or global (or passive). We denote the approximate electromagnetic field at time t_n by $(E_{\Delta t}^n, H_{\Delta t}^n)$ when we apply the IMEX method (3.62) with time step Δt, by $(E_{\Delta t/2}^n, H_{\Delta t/2}^n)$ when we apply the composition of the IMEX method (3.62) with time step $\Delta t/2$, and by (E^n, H^n) when we apply the Richardson

extrapolation. The two implementations of the Richardson extrapolation are depicted in Figures 5.6 and 5.7. We observe that in the local form the value of (E^n, H^n) is used to calculate $(E^{n+1}_{\Delta t}, H^{n+1}_{\Delta t})$ and $(E^{n+1}_{\Delta t/2}, H^{n+1}_{\Delta t/2})$ while in the global form the value of the approximation (E^n, H^n) is never used in the further computations. The global Richardson extrapolation has the same stability properties as the second-order method while the local Richardson extrapolation leads to a new time integration method which might not share the good stability properties of the base method, it may cause instability of the computational process. For example the computation of the fully implicit method (3.8) together with the local Richardson extrapolation will in general be unstable, see e.g. [Faragó *et al.* 2010]. This also happens for the IMEX method (3.62) which is a blend of the methods (3.6) - (3.8). Thus, to extend the IMEX method to fourth-order we will only consider the global Richardson extrapolation which has the same stability properties as the second-order method and requires only three times more computation.

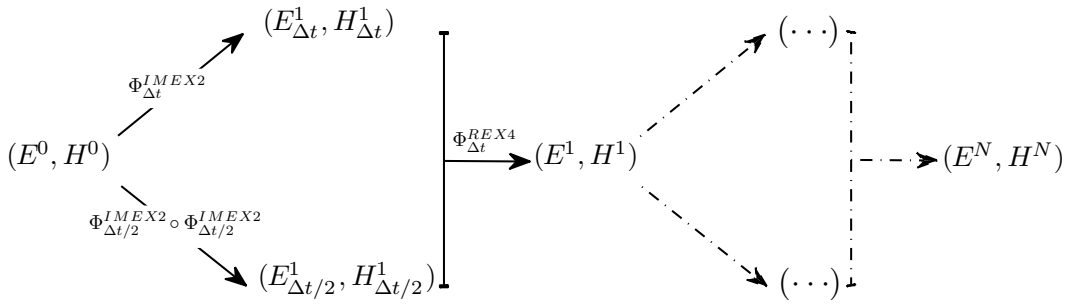


Figure 5.6: Schematic diagrams of the local Richardson extrapolation of the second-order locally implicit method (3.62) (final time $T = N\Delta t$).

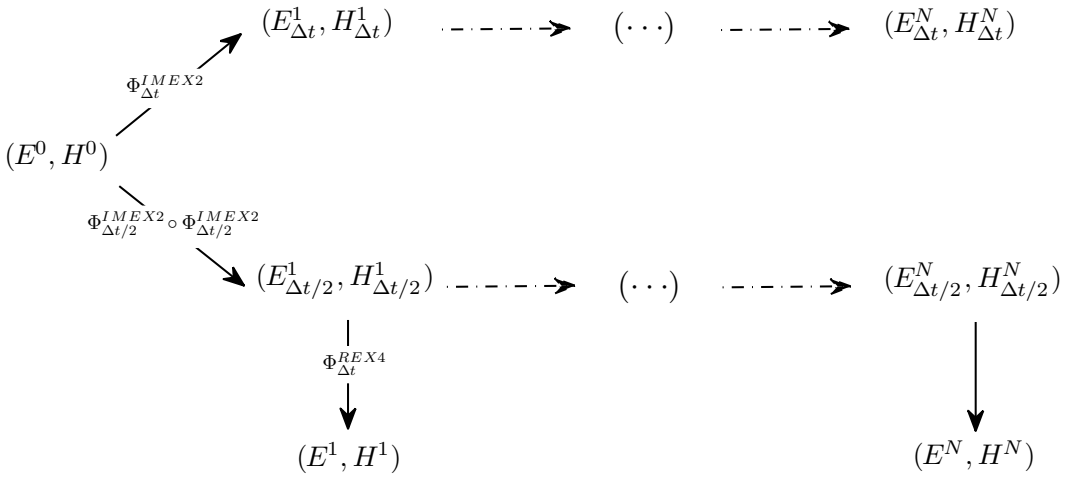


Figure 5.7: Schematic diagrams of the global Richardson extrapolation of the second-order locally implicit method (3.62) (final time $T = N\Delta t$).

5.2 Numerical results

In this section we consider two-dimensional electromagnetic wave propagation problems for which an exact solution is known: the propagation of an eigenmode in a unitary PEC cavity and a model test problem with a volume source term. For a full description of these problems we refer to Section 4.1.1 and Section 4.1.2, respectively.

The total simulation time is set to $T = 5$ m (renormalized unit). We investigate the space-time convergence order of the composition method (5.28) for $s = 3$ and 5, and the global Richardson extrapolation (5.38) based on the second-order IMEX method (3.62). We consider a sequence of five successively refined triangular meshes, see Figure 4.1 and Table 5.1 for an example and the characteristics of the different meshes. The critical time step is determined by the smallest height in the region treated explicitly; for the structured meshes and the implicit regions used in numerical tests it is equal to h^{max}, since all refined triangles belong to the implicit region. To estimate the order of convergence we measure the maximal L^2-norm of the error and we plot this error as a function of the square root of the number of degrees of freedom (DOF), in logarithmic scale. We use DGTD-$\mathbb{P}_4$ methods so that the spatial error is not detrimental to the temporal convergence orders since the theoretical convergence rate is $\mathcal{O}(\Delta t^4 + h^k)$ for a $\mathbb{P}_k$ polynomial interpolation and our DG discretization method.

For the propagation of an eigenmode in a unitary PEC cavity and the locally implicit DGTD-$\mathbb{P}_4$ method (3.62) Figure 5.8 shows the order two, which is in accordance with theory since in this case the convergence rate is dominated by the temporal approximation order, while for composition methods and the Richardson extrapolation the orders are slightly higher than the theoretical convergence rate, about 4.5 for composition methods and about 5.5 for the Richardson extrapolation. We also plot the error as a function of the final CPU time, see the right plot on Figure 5.8. For a given error or a given CPU time we observe the high efficiency of the fourth-order time integration methods compared to the second-order method. Finally for this problem Richardson extrapolation is the most accurate method.

For the model test problem with a volume source term, Figure 5.9 shows orders of convergence in accordance with the theoretical behaviours, i.e., the second-order for the locally implicit DGTD-$\mathbb{P}_4$ method (3.62) and the fourth-order for composition methods and the Richardson extrapolation. We also plot the error as a function of the final CPU time, see the right plot on Figure 5.9. As for the propagation of an eigenmode in a unitary PEC cavity, for a given error or a given CPU time we can see the high efficiency of the fourth-order time integration methods compared to the second-order method and the Richardson extrapolation is also the most accurate method.

These promising results should not prevent us to be cautious; we can not conclude that the fourth-order will be preserved whichever the problem considered. Indeed, the source term and the presence of a damping term which models conduction may be the cause of a reduction order, see for e.g. [Botchev & Verwer 2009, Verwer 2012]. Nevertheless, even if a reduction order is an unavoidable issue in some cases, the accuracy of the high-order locally implicit DG methods proposed in this chapter remains certainly a very interesting feature.

# elements	# DOF	h^{min}	h^{max}
208	3120	0.00736	0.16667
464	6960	0.00442	0.10000
848	12720	0.00316	0.07143
2368	35520	0.00184	0.04167
4688	70320	0.00130	0.02941

Table 5.1: Data of five successively refined triangular meshes (the total number of DOF is indicated for a DGTD-$\mathbb{P}_4$ method)

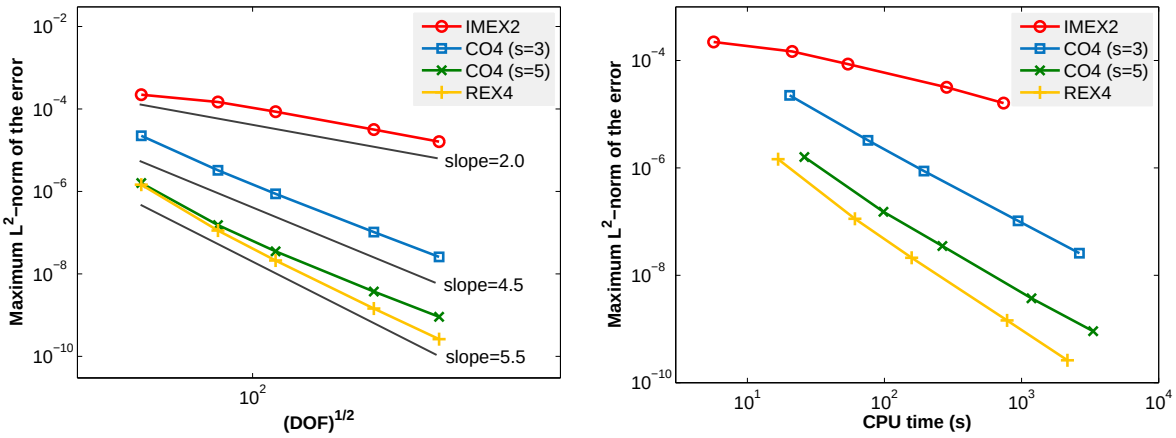


Figure 5.8: Propagation of an eigenmode in a unitary PEC cavity: numerical convergence and maximum error (L^2-norm) in function of final CPU time for the locally implicit DGTD-$\mathbb{P}_4$ methods (left plot - right plot, respectively).

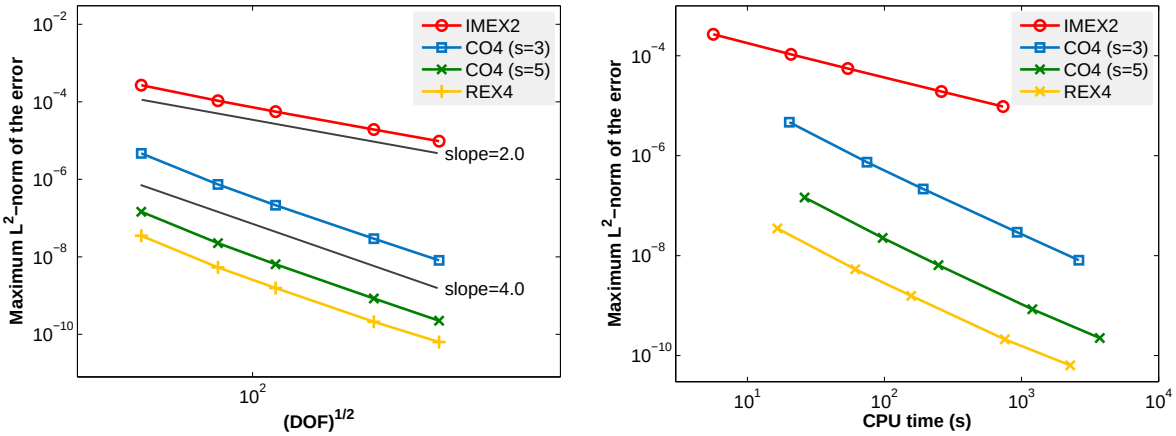


Figure 5.9: Model test problem with an exact solution and a volume source term: numerical convergence and maximum error (L^2-norm) in function of final CPU time for the locally implicit DGTD-$\mathbb{P}_4$ methods (left plot - right plot, respectively).

CHAPTER 6

Locally implicit DGTD method for Maxwell's equations in dispersive media

Contents

The present chapter is concerned with the propagation of electromagnetic waves in dispersive media. These are materials in which either or both of the electromagnetic material parameters ε and μ are functions of frequency. Note that the conductivity σ may also be a function of frequency, but its effect can be rolled into the complex permittivity. In reality, all materials have frequency-dependent ε and μ, but many materials can be approximated as frequency-independent over a frequency band of interest, simplifying their analysis and simulation. We will focus on the much more common case of frequency-dependent permittivity. A lot of practical problems involve such propagation media, such as modeling the interaction of an electromagnetic wave with biological tissues. The numerical modeling of the propagation of electromagnetic waves through human tissues is at the heart of many biomedical applications such as the microwave imaging of cancer tumours. For example, microwave imaging for breast cancer detection is expected to be safe for the patient and has the potential to detect very small cancerous tumors in the breast [Fear *et al.* 2003, Klemm *et al.* 2010, Schea *et al.* 2010]. The definition of microwave-based hyperthermia as an immunotherapy strategy for cancer can also be cited [Converse *et al.* 2004, Karanasiou *et al.* 2008]. The electroporation technique can also be

an application, which consists of applying nanopulses to the tissues, enabling only intracellular membranes to be affected, and then envisage treatments such as electrochemotherapy or gene transfer [Miklavčič *et al.* 2000, Šel *et al.* 2005, Serša 2005, Sukharev *et al.* 1992, Tsong 1991]. All these applications need accurate and efficient numerical modeling techniques, able to deal with the complex issues characterizing the associated propagation problems.

Numerical simulation of wave propagation in dispersive media started in early 1990's in the framework of FDTD methods, for details and references see e.g. [Inan & Marshall 2011] or [Taflove & Hagness 2005]. FETD methods were not explored until 2001 [Jiao & Jin 2001] and DGTD methods for solving Maxwell's equations in dispersive media have been considered more recently. In [Lu *et al.* 2004, Lu *et al.* 2005], a DGTD method, which uses piecewise high-order polynomials for spatial discretization and Runge-Kutta method for time integration, is derived for linear dispersive media of Debye type, the treatment of the dispersive character relies on an Auxiliary Differential Equation (ADE). Numerical results for two-dimensional problems are given and no proof of convergence is provided. In [Huang & Li 2009, Huang *et al.* 2011], a priori error estimates are proved for the second-order formulation of Maxwell's equations coupled to dispersive models discretized by an interior penalty DG formulation. Some two-dimensional numerical tests are included for supporting their analysis. In [Wang *et al.* 2010], different dispersive media are treated, considering a locally divergence-free DG method. The scheme is written and studied in its semi-discretized version, while the fully discrete scheme is described but not analyzed. Finally, in [Lanteri & Scheid 2013], which deals with the Debye model, the DG method is the same as presented in Chapter 2 with a second-order leap-frog scheme for time integration. Stability estimates are derived through energy conservation and convergence is proved for both the semi-discrete and the fully discrete scheme. A two-dimensional artificial numerical problem is presented to validate the theoretical findings.

In Section 6.1 we present the formulation of Maxwell's equations for Debye dispersive media. The Debye model is most often used to model electromagnetic wave interactions with water-based substances, such as biological materials. In particular, biological tissue is well represented by multipole Debye models. The dispersive character will be taken into account via an auxiliary differential equation, which relates the electric polarization to the electric field. In Section 6.2 Maxwell's equations in dispersive media are discretized according to the DG formulation discussed in Chapter 2. We adapt the locally implicit time integration method from [Verwer 2010] with the additional ADE and its stability is analyzed, via an energy approach which provides a rigorous stability criterion. We also derive a convergence analysis to prove that the locally implicit DGTD method for Maxwell's equations in dispersive media retains its second-order convergence. Finally, in Section 6.3, we present some numerical results for three-dimensional problems.

6.1 Maxwell's equations in dispersive media

First we recall Maxwell's equations which govern the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$ in matter (Section 1.1.6). Let $\Omega \subset \mathbb{R}^3$, from time 0 to T, Maxwell's equation are given as

$$\begin{cases} \dfrac{\partial \mathbf{D}}{\partial t} = \operatorname{curl}(\mathbf{H}) - \mathbf{J}^c - \mathbf{J}^s, \\ \dfrac{\partial \mathbf{B}}{\partial t} = \operatorname{curl}(\mathbf{E}), \\ \operatorname{div}(\mathbf{D}) = 0, \\ \operatorname{div}(\mathbf{B}) = 0, \end{cases} \tag{6.1}$$

where $\mathbf{D}$ and $\mathbf{B}$ are the electric and magnetic flux densities, $\mathbf{J}^c$ represents the conduction current density and $\mathbf{J}^s$ a given source current density. We will assume that Ohm's law governs the electric conductivity, $\mathbf{J}^c = \sigma\mathbf{E}$, and that $\mathbf{J}^s = 0$. Finally we will consider dielectric media for which magnetic effects are negligible, that is $\mathbf{M} = 0$ in the constitutive relation which relates the magnetic flux $\mathbf{B}$ to the magnetic field $\mathbf{H}$, i.e. $\mathbf{B} = \mu_0\mathbf{H}$. For linear dispersive materials, the constitutive relationship between the electric flux density and electric field phasors, valid at each frequency and at each point in space, is

$$\widehat{\mathbf{D}}(\mathbf{x},\omega) = \varepsilon_0\varepsilon_r(\mathbf{x},\omega)\widehat{\mathbf{E}}(\mathbf{x},\omega), \tag{6.2}$$

where ε_r is the complex relative permittivity. Equation (6.2) is often written as

$$\begin{aligned} \widehat{\mathbf{D}}(\mathbf{x},\omega) &= \varepsilon_0\varepsilon_\infty(\mathbf{x},\omega)\widehat{\mathbf{E}}(\mathbf{x},\omega) + \widehat{\mathbf{P}}(\mathbf{x},\omega), \\ &= \varepsilon_0\varepsilon_\infty(\mathbf{x},\omega)\widehat{\mathbf{E}}(\mathbf{x},\omega) + \varepsilon_0\widehat{\chi}_e(\mathbf{x},\omega)\widehat{\mathbf{E}}(\mathbf{x},\omega), \end{aligned} \tag{6.3}$$

where $\widehat{\mathbf{P}}$ is the electric polarization phasor, $\widehat{\chi}_e$ is the frequency-domain electric susceptibility of the material and ε_∞ is the infinite frequency relative permittivity. In general, complex permittivity results from the fact that the electric flux density $\mathbf{D}$ in a dielectric, resulting from an applied alternating electric field $\mathbf{E}$ has, in general, a different phase with respect to $\mathbf{E}$. This behavior is due to the inertia of the polarization $\mathbf{P}$ which, when frequency becomes high enough, cannot follow the rapid variations of the field, giving rise to a relaxation (i.e., a momentary delay or lag) of the permittivity. Due to the product of two frequency dependent quantities in (6.3), the time-domain equivalent is given by the following convolution integral

$$\begin{aligned} \mathbf{D}(\mathbf{x},t) &= \varepsilon_0\varepsilon_\infty\mathbf{E}(\mathbf{x},t) + \mathbf{P}(\mathbf{x},t), \\ &= \varepsilon_0\varepsilon_\infty\mathbf{E}(\mathbf{x},t) + \varepsilon_0\int_0^t \mathbf{E}(\mathbf{x},t-s)\,\chi_e(\mathbf{x},s)\,ds, \end{aligned} \tag{6.4}$$

where $\chi_e(\mathbf{x},t)$ is the electric susceptibility, i.e. the inverse Fourier transform of $\widehat{\chi}_e(\mathbf{x},\omega)$. Note that the electric polarization $\mathbf{P}(\mathbf{x},t)$, called the relaxation polarization, depends on the values of $\mathbf{E}$ at $\mathbf{x}$ not only at time t, but also at time prior to t. At first glance, it may appear as if all we need to do is to incorporate the evaluation of the convolution integral in (6.4) into our DGTD method. In practice, there are two problems which prohibit direct implementation of (6.4). First, the convolution integral must be evaluated for each electric field component, at each time step and spatial location within the frequency-dependent medium. Second, evaluation of the convolution requires the storage of the entire time history of the electric fields. Recursive Convolution and ADE approaches have been developed for implementing (6.4), see e.g. [Luebbers *et al.* 1990, Luebbers & Hunsberger 1992, Kelley & Luebbers 1996] and [Kashiwa & Fukai 1990, Joseph *et al.* 1991], respectively. In this chapter we concentrate our analysis on the ADE technique for dispersive dielectric media characterized by a single pole Debye model.

6.1.1 Debye media

Debye media are characterized by a complex value, frequency-domain susceptibility function $\widehat{\chi}_e(\omega)$ that has one or more real poles at separate frequencies. For a single-pole Debye medium, we have

$$\widehat{\chi}_e(\omega) = \frac{\varepsilon_s - \varepsilon_\infty}{1 + j\omega\tau} = \frac{\Delta\varepsilon}{1 + j\omega\tau}, \tag{6.5}$$

where ε_s, called the static relative permittivity, is the permittivity at zero frequency ($\varepsilon_s > \varepsilon_\infty$) and τ is the Debye relaxation time constant, characteristic of the material. Hence the relative permittivity is given by

$$\varepsilon_r(\omega) = \varepsilon_\infty + \frac{\Delta\varepsilon}{1 + j\omega\tau}. \tag{6.6}$$

The time-domain susceptibility function can be obtained by inverse Fourier transformation of (6.5), yielding the decaying exponential function

$$\chi_e(t) = \frac{\Delta\varepsilon}{\tau}\, \mathrm{e}^{-t/\tau} U(t), \tag{6.7}$$

where $U(t)$ is the Heaviside step function (or the unit step function).

6.1.2 Auxiliary Differential Equation method

The goal of the ADE approach is to express the relationship between electric polarization $\mathbf{P}$ and electric field $\mathbf{E}$ with a differential equation rather than a convolution integral. This section discusses the formulation of the ADE technique for dispersive dielectric media characterized by a single pole Debye model. First we express the relationship between $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{E}}$ from (6.3)

$$\widehat{\mathbf{P}}(\mathbf{x}, \omega) = \varepsilon_0 \widehat{\chi}_e(\mathbf{x}, \omega)\, \widehat{\mathbf{E}}(\mathbf{x}, \omega). \tag{6.8}$$

With the particular functional form of the susceptibility $\widehat{\chi}_e$ for a single pole Debye medium (6.5) we obtain

$$\widehat{\mathbf{P}}(\mathbf{x}, \omega) = \varepsilon_0 \frac{\Delta\varepsilon}{1 + j\omega\tau} \widehat{\mathbf{E}}(\mathbf{x}, \omega). \tag{6.9}$$

We write (6.9) as

$$(1 + j\omega\tau)\, \widehat{\mathbf{P}}(\mathbf{x}, \omega) = \varepsilon_0 \Delta\varepsilon\, \widehat{\mathbf{E}}(\mathbf{x}, \omega). \tag{6.10}$$

Finally by inverse Fourier transformation and by recognizing multiplication with $j\omega$ as a first-order time derivative, we obtain the ADE in terms of $\mathbf{P}$

$$\mathbf{P} + \tau \frac{\partial \mathbf{P}}{\partial t} = \varepsilon_0 (\varepsilon_s - \varepsilon_\infty)\, \mathbf{E}. \tag{6.11}$$

Note that with the constitutive relation (6.4) and the ADE (6.11) the Ampere - Maxwell law becomes

$$\begin{aligned} \varepsilon_0 \varepsilon_\infty \frac{\partial \mathbf{E}}{\partial t} &= \operatorname{curl}(\mathbf{H}) - \sigma\mathbf{E} - \frac{\partial \mathbf{P}}{\partial t}, \\ &= \operatorname{curl}(\mathbf{H}) - \frac{\varepsilon_0 (\varepsilon_s - \varepsilon_\infty)}{\tau}\mathbf{E} - \sigma\mathbf{E} + \frac{1}{\tau}\mathbf{P}. \end{aligned} \tag{6.12}$$

We can now state Maxwell's equations in a Debye dispersive medium.

6.1.3 The continuous problem formulation

Let $\Omega \subset \mathbb{R}^3$ be a bounded, convex, polyhedral domain. We denote by $\mathbf{n}$ the normal outward to $\partial\Omega$ and we define the functional spaces

$$\mathscr{H} \equiv \left[H(\mathrm{curl},\Omega)\right]^3 \times \left[H(\mathrm{curl},\Omega)\right]^3 \times \left[L^2(\Omega)\right]^3, \tag{6.13}$$

where $H(\mathrm{curl},\Omega)$ is the classical subspace of $L^2(\Omega)$ fields with curl in $L^2(\Omega)$;

$$\mathscr{H}_0 \equiv \left[H(\mathrm{curl},\Omega)\right]^3 \times \left[H_0(\mathrm{curl},\Omega)\right]^3 \times \left[L^2(\Omega)\right]^3, \tag{6.14}$$

where $H_0(\mathrm{curl},\Omega)$ is the classical subspace of $H(\mathrm{curl},\Omega)$ fields with zero tangential trace. Let $T>0$, the magnetic field $\mathbf{H}$, the electric field $\mathbf{E}$ and the polarization $\mathbf{P}$ satisfy the following system of equations in $[0,T]$

$$\left\{\begin{array}{l} \mu_0 \dfrac{\partial \mathbf{H}}{\partial t} = -\mathrm{curl}(\mathbf{E}), \\ \varepsilon_0\varepsilon_\infty \dfrac{\partial \mathbf{E}}{\partial t} = \mathrm{curl}(\mathbf{H}) - \dfrac{\varepsilon_0(\varepsilon_s-\varepsilon_\infty)}{\tau}\mathbf{E} - \sigma\mathbf{E} + \dfrac{1}{\tau}\mathbf{P}, \\ \dfrac{\partial \mathbf{P}}{\partial t} = \dfrac{\varepsilon_0(\varepsilon_s-\varepsilon_\infty)}{\tau}\mathbf{E} - \dfrac{1}{\tau}\mathbf{P}, \end{array}\right. \tag{6.15}$$

For convenience of presentation we concentrate our analysis by imposing metallic boundary conditions, i.e. $\mathbf{E}\times\mathbf{n}=0$, but the same analysis can be carried out in the presence of an artificial truncation of the computational domain modelled by absorbing boundary conditions (this requires some additional terms in the DGTD method presented in the following). Finally the system of equations (6.15) is closed with initial conditions, $(\mathbf{H}_0,\mathbf{E}_0,\mathbf{P}_0) \in \mathscr{H}(\Omega)$ such that $\mathbf{H}(\cdot,0)=\mathbf{H}_0$, $\mathbf{E}(\cdot,0)=\mathbf{E}_0$ and $\mathbf{P}(\cdot,0)=\mathbf{P}_0$.

Remark 6.1.1. *The existence of a solution of the weak formulation of system* (6.15) *is proved in [Lanteri & Scheid 2013]; i.e. there is a solution* $(\mathbf{H},\mathbf{E},\mathbf{P})$ *in* $C([0,T],\mathscr{H}_0(\Omega))$ *such that for all* (ϕ,ψ,φ) *in* $\mathscr{H}_0(\Omega)$

$$\left\{\begin{array}{l} \mu_0 \displaystyle\int_\Omega \frac{\partial \mathbf{H}}{\partial t}\cdot\phi = -\int_\Omega \mathrm{curl}(\mathbf{E})\cdot\phi, \\ \varepsilon_0\varepsilon_\infty \displaystyle\int_\Omega \frac{\partial \mathbf{E}}{\partial t}\cdot\psi = \int_\Omega \mathrm{curl}(\mathbf{H})\cdot\psi - \frac{1}{\tau}\left[\varepsilon_0(\varepsilon_s-\varepsilon_\infty)+\sigma\right]\int_\Omega \mathbf{E}\cdot\psi + \frac{1}{\tau}\int_\Omega \mathbf{P}\cdot\psi, \\ \displaystyle\int_\Omega \frac{\partial \mathbf{P}}{\partial t}\cdot\varphi = \frac{1}{\tau}\left[\varepsilon_0(\varepsilon_s-\varepsilon_\infty)\right]\int_\Omega \mathbf{E}\cdot\varphi - \frac{1}{\tau}\int_\Omega \mathbf{P}\cdot\varphi. \end{array}\right. \tag{6.16}$$

Remark 6.1.2. *One can prove that if the fields are initially divergence free, then the fields are divergence free at any time, see a proof in [Li 2011]. That's why we do not need to consider the third and fourth equations of* (6.1) *in the system* (6.15)

From now we consider a normalized form of Maxwell's equations (6.15). We introduce the normalized space, time variables and physical fields through the relations

$$\begin{array}{l} \widetilde{\mathbf{x}} = \mathbf{x}, \quad \widetilde{t} = c_0\, t, \quad \widetilde{\tau} = c_0\, \tau, \quad \widetilde{\sigma} = Z_0\, \sigma, \\ \widetilde{\mathbf{E}} = \mathbf{E}, \quad \widetilde{\mathbf{H}} = Z_0\, \mathbf{H} \quad \text{and} \quad \widetilde{\mathbf{P}} = \mathbf{P}/\varepsilon_0, \end{array} \tag{6.17}$$

$\mathbf{x} \longrightarrow \widetilde{\mathbf{x}}$	$t \longrightarrow \widetilde{t}$	$\tau \longrightarrow \widetilde{\tau}$	$\sigma \longrightarrow \widetilde{\sigma}$
$\mathrm{m} \longrightarrow \mathrm{m}$	$\mathrm{s} \longrightarrow \mathrm{m}$	$\mathrm{s} \longrightarrow \mathrm{m}$	$\mathrm{S{\cdot}m^{-1}} \longrightarrow \mathrm{m^{-1}}$

$\mathbf{E} \longrightarrow \widetilde{\mathbf{E}}$	$\mathbf{H} \longrightarrow \widetilde{\mathbf{H}}$	$\mathbf{P} \longrightarrow \widetilde{\mathbf{P}}$
$\mathrm{V{\cdot}m^{-1}} \longrightarrow \mathrm{V{\cdot}m^{-1}}$	$\mathrm{A{\cdot}m^{-1}} \longrightarrow \mathrm{V{\cdot}m^{-1}}$	$\mathrm{A{\cdot}m^{-2}} \longrightarrow \mathrm{V{\cdot}m^{-2}}$

Table 6.1: Units of physical and normalized variables and fields.

where $c_0 = 1/\sqrt{\varepsilon_0\mu_0}$ is the speed of light in vacuum ($c_0 \simeq 3\times 10^8 \mathrm{m{\cdot}s^{-1}}$) and $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ is the free space intrinsic impedance (Ohm, $\Omega = \mathrm{V{\cdot}A^{-1}}$). The units of physical and normalized variables and fields are summarized in Table 6.1.

Substituting the normalized space, time variables and fields (6.17) into (6.15) we can write Maxwell's equations as

$$\begin{cases} \mu_r \dfrac{\partial \widetilde{\mathbf{H}}}{\partial \widetilde{t}} = -\mathrm{curl}\big(\widetilde{\mathbf{E}}\big), \\ \varepsilon_\infty \dfrac{\partial \widetilde{\mathbf{E}}}{\partial \widetilde{t}} = \mathrm{curl}\big(\widetilde{\mathbf{H}}\big) - \dfrac{(\varepsilon_s-\varepsilon_\infty)}{\widetilde{\tau}}\widetilde{\mathbf{E}} - \widetilde{\sigma}\widetilde{\mathbf{E}} + \dfrac{1}{\widetilde{\tau}}\widetilde{\mathbf{P}}, \\ \dfrac{\partial \widetilde{\mathbf{P}}}{\partial \widetilde{t}} = \dfrac{(\varepsilon_s-\varepsilon_\infty)}{\widetilde{\tau}}\widetilde{\mathbf{E}} - \dfrac{1}{\widetilde{\tau}}\widetilde{\mathbf{P}}, \end{cases} \tag{6.18}$$

For convenience of presentation we omit in the sequel the "$\sim$" notation and we denote the relative magnetic permeability, μ_r, by μ into (6.18).

6.2 The locally implicit DGTD method for Maxwell's equations in dispersive dielectric media

6.2.1 Semi-discretization by the DG method

We write the semi-discrete system where the electric polarization $\mathbf{P}$ is discretized according to the discontinuous Galerkin formulation discussed in Chapter 2

$$\begin{cases} \mu M \dfrac{\partial H}{\partial t} = -S^T E, \\ \varepsilon_\infty M \dfrac{\partial E}{\partial t} = SH - \dfrac{(\varepsilon_s-\varepsilon_\infty)}{\tau} ME - \sigma ME + \dfrac{1}{\tau} MP, \\ M \dfrac{\partial P}{\partial t} = \dfrac{(\varepsilon_s-\varepsilon_\infty)}{\tau} ME - \dfrac{1}{\tau} MP, \end{cases} \tag{6.19}$$

where the column vector H, E, P and the block entries of matrices M and S are defined as in Chapter 2. Now we use the Cholesky factorization of the mass matrix $M = L_M L_M^T$, where L_M is a triangular

matrix, and we introduce the change of variables $\widetilde{E} = L_M^T E$, $\widetilde{H} = L_M^T H$ and $\widetilde{P} = L_M^T P$ in (6.19), hence

$$\begin{cases} \mu \dfrac{\partial \widetilde{H}}{\partial t} = -\widetilde{S}^T \widetilde{E}, \\ \varepsilon_\infty \dfrac{\partial \widetilde{E}}{\partial t} = \widetilde{S}\widetilde{H} - \dfrac{(\varepsilon_s - \varepsilon_\infty)}{\tau}\widetilde{E} - \sigma\widetilde{E} + \dfrac{1}{\tau}\widetilde{P}, \\ \dfrac{\partial \widetilde{P}}{\partial t} = \dfrac{(\varepsilon_s - \varepsilon_\infty)}{\tau}\widetilde{E} - \dfrac{1}{\tau}\widetilde{P}, \end{cases} \tag{6.20}$$

where $\widetilde{S} = L_M^{-1}\ S\ (L_M^{-1})^T$. For convenience of presentation we omit from now the notation tilde in (6.20).

6.2.2 Stability of the semi-discrete scheme

Let (H_h, E_h, P_h) the solution of the semi-discrete system (6.20), we define the semi-discrete energy by

$$\mathscr{E}_h \ = \ \frac{1}{2}\left(\mu\|H_h\|_2^2 + \varepsilon_\infty\|E_h\|_2^2 + \frac{1}{(\varepsilon_s - \varepsilon_\infty)}\|P_h\|_2^2\right), \tag{6.21}$$

where

$$\|E_h\|_2^2 = \langle\, E_h,\, E_h\rangle\,,\ \|H_h\|_2^2 = \langle\, H_h,\, H_h\rangle\,,\ \text{and}\ \|P_h\|_2^2 = \langle\, P_h,\, P_h\rangle\,, \tag{6.22}$$

and $\langle\,\cdot\,,\,\cdot\,\rangle$ is the L^2 inner product. Taking inner products with H_h, E_h and P_h in (6.20) yields, respectively,

$$\begin{aligned} \frac{\mu}{2}\frac{d}{dt}\langle\, H_h,\, H_h\rangle &= -\big\langle\, S^T E_h,\, H_h\big\rangle\,, \\ \frac{\varepsilon_\infty}{2}\frac{d}{dt}\langle\, E_h,\, E_h\rangle &= -\langle\, SH_h,\, E_h\rangle - \frac{(\varepsilon_s - \varepsilon_\infty)}{\tau}\langle\, E_h,\, E_h\rangle - \sigma\langle\, E_h,\, E_h\rangle \\ &+ \frac{1}{\tau}\langle\, P_h,\, E_h\rangle\,, \\ \frac{1}{2(\varepsilon_s - \varepsilon_\infty)}\frac{d}{dt}\langle\, P_h,\, P_h\rangle &= \frac{1}{\tau}\langle\, E_h,\, P_h\rangle - \frac{1}{\tau(\varepsilon_s - \varepsilon_\infty)}\langle\, P_h,\, P_h\rangle\,. \end{aligned} \tag{6.23}$$

Adding the three equations of (6.23), it follows

$$\begin{aligned} \frac{d}{dt}\mathscr{E}_h &= -\frac{1}{\tau(\varepsilon_s - \varepsilon_\infty)}\|P_h\|_2^2 + \frac{1}{2\tau}\langle\, P_h,\, E_h\rangle - \frac{(\varepsilon_s - \varepsilon_\infty)}{\tau}\|E_h\|_2^2 - \sigma\|E_h\|_2^2, \\ &= -\frac{1}{\tau(\varepsilon_s - \varepsilon_\infty)}\|P_h - (\varepsilon_s - \varepsilon_\infty)E_h\|_2^2 - \sigma\|E_h\|_2^2, \\ &\leq 0. \end{aligned} \tag{6.24}$$

Hence $\mathscr{E}_h$ is a decreasing function in time so that $\mathscr{E}_h(t) \leq \mathscr{E}_h(0)$, showing stability in the L^2 sense.

6.2.3 Time integration methods

First we apply the second order leap-frog scheme to the semi-discrete system (6.20) that we write in the three-stage form, emanating from Verlet's method,

$$\left\{\begin{array}{rcl} \mu\dfrac{H^{n+1/2}-H^n}{\Delta t/2} &=& -S^T E^n, \\ \varepsilon_\infty\dfrac{E^{n+1}-E^n}{\Delta t} &=& SH^{n+1/2}-\dfrac{(\varepsilon_s-\varepsilon_\infty)}{2\tau}\left(E^{n+1}+E^n\right) \\ &-& \dfrac{1}{2}\sigma\left(E^{n+1}+E^n\right)+\dfrac{1}{2\tau}\left(P^{n+1}+P^n\right), \\ \dfrac{P^{n+1}-P^n}{\Delta t} &=& \dfrac{(\varepsilon_s-\varepsilon_\infty)}{2\tau}\left(E^{n+1}+E^n\right)-\dfrac{1}{2\tau}\left(P^{n+1}+P^n\right), \\ \mu\dfrac{H^{n+1}-H^{n+1/2}}{\Delta t/2} &=& -S^T E^{n+1}. \end{array}\right. \tag{6.25}$$

Now we apply the second order Crank-Nicolson scheme to the semi-discrete system (6.20) that we write in the three-stage form

$$\left\{\begin{array}{rcl} \mu\dfrac{H^{n+1/2}-H^n}{\Delta t/2} &=& -S^T E^n, \\ \varepsilon_\infty\dfrac{E^{n+1}-E^n}{\Delta t} &=& \dfrac{1}{2}S\left(H^{n+1}+H^n\right)-\dfrac{(\varepsilon_s-\varepsilon_\infty)}{2\tau}\left(E^{n+1}+E^n\right) \\ &-& \dfrac{1}{2}\sigma\left(E^{n+1}+E^n\right)+\dfrac{1}{2\tau}\left(P^{n+1}+P^n\right), \\ \dfrac{P^{n+1}-P^n}{\Delta t} &=& \dfrac{(\varepsilon_s-\varepsilon_\infty)}{2\tau}\left(E^{n+1}+E^n\right)-\dfrac{1}{2\tau}\left(P^{n+1}+P^n\right), \\ \mu\dfrac{H^{n+1}-H^{n+1/2}}{\Delta t/2} &=& -S^T E^{n+1}. \end{array}\right. \tag{6.26}$$

Finally we blend (6.25) and (6.26) to obtain the implicit-explicit time integration scheme for the semi-discrete system (6.20)

$$\left\{\begin{array}{rcl} \mu\dfrac{H^{n+1/2}-H^n}{\Delta t/2} &=& -S^T E^n, \\ \varepsilon_\infty\dfrac{E^{n+1}-E^n}{\Delta t} &=& S_0H^{n+1/2}+\dfrac{1}{2}S_1\left(H^{n+1}+H^n\right)-\dfrac{(\varepsilon_s-\varepsilon_\infty)}{2\tau}\left(E^{n+1}+E^n\right) \\ &-& \dfrac{1}{2}\sigma\left(E^{n+1}+E^n\right)+\dfrac{1}{2\tau}\left(P^{n+1}+P^n\right), \\ \dfrac{P^{n+1}-P^n}{\Delta t} &=& \dfrac{(\varepsilon_s-\varepsilon_\infty)}{2\tau}\left(E^{n+1}+E^n\right)-\dfrac{1}{2\tau}\left(P^{n+1}+P^n\right), \\ \mu\dfrac{H^{n+1}-H^{n+1/2}}{\Delta t/2} &=& -S^T E^{n+1}, \end{array}\right. \tag{6.27}$$

where $S=S_0+S_1$ is a matrix splitting. The method is implicit in S_1 and explicit in S_0. For $S_0=0$ we recover (6.26) and for $S_1=0$ the method (6.25). As in Chapter 3 we adopt the splitting defined

in [Verwer 2010], i.e. $S_1 = SS_H$, where S_H is a diagonal matrix of dimension the length of H, see (3.68). From the third and the fourth equations of (6.27) we express the electric polarization P^{n+1} and the magnetic field H^{n+1} as

$$P^{n+1} = \left(\frac{2\tau - \Delta t}{2\tau + \Delta t}\right) P^n + \left(\frac{\Delta t}{2\tau + \Delta t}\right) (\varepsilon_s - \varepsilon_\infty) \left(E^{n+1} + E^n\right), \tag{6.28}$$

and

$$H^{n+1} = H^{n+1/2} - \frac{1}{\mu}\frac{\Delta t}{2} S^T E^{n+1}. \tag{6.29}$$

Now we substitute these expressions into the second stage of (6.27)

$$\begin{aligned} \varepsilon_\infty \frac{E^{n+1} - E^n}{\Delta t} &= S_0 H^{n+1/2} + \frac{1}{2} S_1 H^{n+1/2} \\ &- \frac{1}{2\tau}(\varepsilon_s - \varepsilon_\infty) E^n - \frac{1}{2}\sigma E^n + \frac{1}{2\tau}\left(\frac{\Delta t}{2\tau + \Delta t}\right)(\varepsilon_s - \varepsilon_\infty) E^n \\ &- \frac{1}{2\tau}(\varepsilon_s - \varepsilon_\infty) E^{n+1} - \frac{1}{2}\sigma E^{n+1} + \frac{1}{2\tau}\left(\frac{\Delta t}{2\tau + \Delta t}\right)(\varepsilon_s - \varepsilon_\infty) E^{n+1} \\ &- \frac{1}{\mu}\frac{\Delta t}{4} S_1 S^T E^{n+1} + \frac{1}{2\tau} P^n + \frac{1}{2\tau}\left(\frac{2\tau - \Delta t}{2\tau + \Delta t}\right) P^n. \end{aligned} \tag{6.30}$$

Note that $S_1 S^T = S_1 S_1^T$, the second stage (6.30) is then equivalent to the following linear system

$$\mathscr{M} E^{n+1} = b_n, \tag{6.31}$$

where

$$\mathscr{M} = \left[1 - \frac{1}{\varepsilon_\infty}\frac{\Delta t}{2\tau}\left(\frac{\Delta t}{2\tau + \Delta t}\right)(\varepsilon_s - \varepsilon_\infty) + \frac{1}{\varepsilon_\infty}\frac{\Delta t}{2\tau}(\varepsilon_s - \varepsilon_\infty) + \frac{1}{\varepsilon_\infty}\frac{\Delta t}{2}\sigma\right] I + \frac{1}{\varepsilon_\infty \mu}\frac{\Delta t^2}{4} S_1 S_1^T, \tag{6.32}$$

and

$$\begin{aligned} b_n &= \left[1 + \frac{1}{\varepsilon_\infty}\frac{\Delta t}{2\tau}\left(\frac{\Delta t}{2\tau + \Delta t}\right)(\varepsilon_s - \varepsilon_\infty) - \frac{1}{\varepsilon_\infty}\frac{\Delta t}{2\tau}(\varepsilon_s - \varepsilon_\infty) - \frac{1}{\varepsilon_\infty}\frac{\Delta t}{2}\sigma\right] E^n \\ &+ \frac{1}{\varepsilon_\infty}\Delta t S_0 H^{n+1/2} + \frac{1}{\varepsilon_\infty}\frac{\Delta t}{2} S_1 \left(H^{n+1/2} + H^n\right) \\ &+ \frac{1}{\varepsilon_\infty}\frac{\Delta t}{2\tau}\left[1 + \left(\frac{2\tau - \Delta t}{2\tau + \Delta t}\right)\right] P^n. \end{aligned} \tag{6.33}$$

Hence we can write the locally implicit scheme (6.27) as

$$\begin{cases} H^{n+1/2} &= H^n - \dfrac{1}{\mu}\dfrac{\Delta t}{2} S^T E^n, \\ \mathscr{M} E^{n+1} &= b_n, \\ P^{n+1} &= \left(\dfrac{2\tau - \Delta t}{2\tau + \Delta t}\right) P^n + \left(\dfrac{\Delta t}{2\tau + \Delta t}\right)(\varepsilon_s - \varepsilon_\infty)\left(E^{n+1} + E^n\right), \\ H^{n+1} &= H^{n+1/2} - \dfrac{1}{\mu}\dfrac{\Delta t}{2} S^T E^{n+1}, \end{cases} \tag{6.34}$$

where the matrix $\mathscr{M}$ and the right hand side b_n are defined by (6.32) and (6.33), respectively.

6.2.4 Stability of the fully discrete scheme

In this section we are interested in the stability of the fully discrete locally implicit scheme (6.27). The derivations in the remainder of this section follow an energy approach which provides a rigorous criterion for stability. In Section 6.2.4.1 we exhibit a discrete energy, which is a quadratic form of the numerical unknowns. In Section 6.2.4.2 we show that the energy is a positive definite quadratic form. Finally, in Section 6.2.4.3 we prove that the energy is decreasing, which achieves the stability analysis.

6.2.4.1 Discrete energy

We define the discrete electromagnetic energy, denote $\mathscr{E}_n$, as

$$\mathscr{E}_n = \frac{1}{2}\left(\mu\|H^n\|_2^2+\varepsilon_\infty\|E^n\|_2^2+\frac{1}{(\varepsilon_s-\varepsilon_\infty)}\|P^n\|_2^2-\frac{\Delta t^2}{4\mu}\left\langle S_0S^TE^n,\,E^n\right\rangle\right), \tag{6.35}$$

where $\langle\,\cdot\,,\,\cdot\,\rangle$ is the L^2 inner product and $\|\cdot\|_2$ the corresponding norm.

6.2.4.2 Positivity of the energy

We state a condition on the time step Δt such that $\mathscr{E}_n$ is a positive definite quadratic form.

Lemma 6.2.1. *The quadratic form $\mathscr{E}_n$ given by (6.35) is a positive definite quadratic form of the numerical unknowns H^n, E^n and P^n if*

$$\Delta t < \frac{2\sqrt{\varepsilon_\infty\mu}}{\sqrt{\rho\left(S_0S_0^T\right)}}, \tag{6.36}$$

where ρ denotes the spectral radius.

Proof. By definition of $S_0 = S(I-S_H)$ the matrix S_0S^T is symmetric, since $S_0S^T = S(I-S_H)S^T = S(I-S_H)(I-S_H)S^T = S_0S_0^T$. Then, from (6.35) we write the electromagnetic energy as

$$\begin{aligned}\mathscr{E}_n &= \frac{1}{2}\left(\mu\|H^n\|_2^2+\varepsilon_\infty\|E^n\|_2^2+\frac{1}{(\varepsilon_s-\varepsilon_\infty)}\|P^n\|_2^2-\frac{\Delta t^2}{4\mu}\left\langle S_0S_0^TE^n,\,E^n\right\rangle\right),\\ &= \frac{1}{2}\left(\mu\|H^n\|_2^2+\varepsilon_\infty\|E^n\|_2^2+\frac{1}{(\varepsilon_s-\varepsilon_\infty)}\|P^n\|_2^2-\frac{\Delta t^2}{4\mu}\|S_0E^n\|_2^2\right).\end{aligned} \tag{6.37}$$

Furthermore, we have

$$\|S_0^TE^n\|_2 \le \|S_0^T\|_2\|E^n\|_2 = \sqrt{\rho\left(S_0S_0^T\right)}\|E^n\|_2, \tag{6.38}$$

hence

$$\mathscr{E}_n \ge \frac{1}{2}\left(\mu\|H^n\|_2^2+\left(\varepsilon_\infty-\frac{\Delta t^2}{4\mu}\rho\left(S_0S_0^T\right)\right)\|E^n\|_2^2+\frac{1}{(\varepsilon_s-\varepsilon_\infty)}\|P^n\|_2^2\right), \tag{6.39}$$

from which it follows that under the condition (6.36), the discrete energy $\mathscr{E}_n$ is a positive definite quadratic form of the numerical unknowns H^n, E^n and P^n. □

6.2.4.3 *Variation of the energy*

We now prove the following result.

Lemma 6.2.2. *The discrete energy* (6.35) *is decreasing so that* $\mathscr{E}_n \leq \mathscr{E}_0$.

Proof. From the first and fourth equation of (6.27) we have

$$\begin{aligned} H^{n+1/2} &= H^n - \frac{\Delta t}{2\mu} S^T E^n, \\ H^{n+1/2} &= H^{n+1} + \frac{\Delta t}{2\mu} S^T E^{n+1}. \end{aligned} \tag{6.40}$$

Substitute half of each expressions of (6.40) for $H^{n+1/2}$ into the second stage of (6.27) and the first expression of (6.40) into the fourth stage of (6.27). Together with the final expression of P^{n+1} this yields

$$\begin{aligned} \varepsilon_\infty \left(E^{n+1} - E^n\right) &= \frac{\Delta t}{2} S \left(H^{n+1} + H^n\right) + \frac{\Delta t}{2\tau} \left(P^{n+1} + P^n\right) \\ &- \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\tau} \left(E^{n+1} + E^n\right) - \frac{\Delta t}{2} \sigma \left(E^{n+1} + E^n\right) \\ &+ \frac{\Delta t^2}{4\mu} S_0 S^T \left(E^{n+1} - E^n\right), \\ \frac{1}{(\varepsilon_s - \varepsilon_\infty)} \left(P^{n+1} - P^n\right) &= \frac{\Delta t}{2\tau} \left(E^{n+1} + E^n\right) - \frac{\Delta t}{2\tau(\varepsilon_s - \varepsilon_\infty)} \left(P^{n+1} + P^n\right), \\ \mu \left(H^{n+1} - H^n\right) &= -\frac{\Delta t}{2} S^T \left(E^{n+1} + E^n\right). \end{aligned} \tag{6.41}$$

Taking inner product with $E^{n+1} + E^n$, $P^{n+1} + P^n$ and $H^{n+1} + H^n$ yields, respectively,

$$\begin{aligned} \varepsilon_\infty \left(\|E^{n+1}\|_2^2 - \|E^n\|_2^2\right) &= \frac{\Delta t}{2} \left\langle S \left(H^{n+1} + H^n\right), E^{n+1} + E^n \right\rangle \\ &+ \frac{\Delta t}{2\tau} \left\langle P^{n+1} + P^n, E^{n+1} + E^n \right\rangle \\ &- \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\tau} \|E^{n+1} + E^n\|_2^2 - \frac{\Delta t}{2} \sigma \|E^{n+1} + E^n\|_2^2 \\ &+ \frac{\Delta t^2}{4\mu} \left\langle S_0 S^T \left(E^{n+1} - E^n\right), E^{n+1} + E^n \right\rangle, \\ \frac{1}{(\varepsilon_s - \varepsilon_\infty)} \left(\|P^{n+1}\|_2^2 - \|P^n\|_2^2\right) &= \frac{\Delta t}{2\tau} \left\langle E^{n+1} + E^n, P^{n+1} + P^n \right\rangle \\ &- \frac{\Delta t}{2\tau(\varepsilon_s - \varepsilon_\infty)} \|P^{n+1} + P^n\|_2^2, \\ \mu \left(\|H^{n+1}\|_2^2 - \|H^n\|_2^2\right) &= -\frac{\Delta t}{2} \left\langle S^T \left(E^{n+1} + E^n\right), H^{n+1} + H^n \right\rangle. \end{aligned} \tag{6.42}$$

We recall that $S_0 S^T = S_0 S_0^T$, hence

$$\begin{aligned} \left\langle S_0 S^T \left(E^{n+1} - E^n\right), E^{n+1} + E^n \right\rangle &= \left\langle S_0 S_0^T \left(E^{n+1} - E^n\right), E^{n+1} + E^n \right\rangle, \\ &= \left\langle S_0 S_0^T E^{n+1}, E^{n+1} \right\rangle - \left\langle S_0 S_0^T E^n, E^n \right\rangle, \\ &= \left\langle S_0 S^T E^{n+1}, E^{n+1} \right\rangle - \left\langle S_0 S^T E^n, E^n \right\rangle. \end{aligned} \tag{6.43}$$

Substituting this expression into the first equation of (6.42) and adding the three equations of (6.42) yields

$$\begin{aligned}
\frac{\mathscr{E}_{n+1}-\mathscr{E}_n}{\Delta t} &= -\frac{1}{4\tau\,(\varepsilon_s-\varepsilon_\infty)}\|P^{n+1}-P^n\|_2^2+\frac{1}{2\tau}\left\langle P^{n+1}+P^n,\,E^{n+1}+E^n\right\rangle \\
&- \frac{(\varepsilon_s-\varepsilon_\infty)}{4\tau}\|E^{n+1}-E^n\|_2^2-\frac{1}{4}\sigma\|E^{n+1}-E^n\|_2^2, \\
&= -\frac{1}{4\tau\,(\varepsilon_s-\varepsilon_\infty)}\|\left(P^{n+1}+P^n\right)-(\varepsilon_s-\varepsilon_\infty)\left(E^{n+1}+E^n\right)\|_2^2 \\
&- \frac{1}{4}\sigma\|E^{n+1}-E^n\|_2^2, \\
&\leq 0.
\end{aligned} \tag{6.44}$$

□

6.2.5 Convergence

In this section we are interested in the PDE convergence of the locally implicit method (6.27). We will examine whether the method retains its second-order ODE convergence under stable simultaneous space-time grid refinement $\Delta t \backsim h$, $h \to 0$ towards the exact PDE solution. As we have seen previously in Chapter 3 this is not a priori clear due to the component splitting which can introduce order reduction through error constants which grow with h^{-1}, for $h \to 0$.

The derivations in the remainder of this section follow a method of lines analysis related to that of [Verwer 2010] for the locally implicit method (3.62), which deals with Maxwell's equations in non-dispersive media. The proof of second temporal order in the PDE sense presented here is organized in three subsections. In Section 6.2.5.1 we will introduce the so-called perturbed scheme obtained by substituting the true PDE solution restricted to the assumed space grid into the locally implicit scheme (6.27). Herewith we introduce defects (truncation errors) composed of a temporal and a spatial error part. Our focus lies on temporal order, so for simplicity of derivation we will omit the spatial error part after this subsection. For our purpose the spatial error part can be omitted without loss of generality. In Section 6.2.5.2 we derive the common temporal recurrence for the full global error which is the difference of the PDE solution restricted to the space grid and the numerical solution on this grid generated by scheme (6.27). Here we point out that this global error scheme needs to be transformed to overcome a spatial inconsistency in the local error emanating from component splitting. The crucial observation hereby is that this spatial inconsistency enters the temporal error by the negative power h^{-1} which kills one power of Δt as we assume $\Delta t \backsim h$, $h \to 0$ (order reduction). Fortunately, this order reduction is present in the local error only and cancels in the transition from local to global errors. That this cancellation occurs can be proved by transforming the global error scheme, which is shown in the third Section 6.2.5.3.

6.2.5.1 The perturbed scheme

Let $E_h(t)$ denote at time t the exact solution of the PDE problem restricted to the assumed space grid that we have approximated with the semi-discrete system (6.20). $E_h(t_n)$ thus represents the vector that is approximated by E^n. Assume the same notation for H and P. Substituting $E_h(t)$, $H_h(t)$ and

$P_h(t)$ into (6.20) reveals the spatial truncation errors which we denote by ζ_h^E, ζ_h^H and ζ_h^P

$$\begin{cases} \mu \dfrac{d}{dt} H_h(t) = -S^T E_h(t) + \zeta_h^H(t), \\ \varepsilon_\infty \dfrac{d}{dt} E_h(t) = S H_h(t) - \dfrac{(\varepsilon_s - \varepsilon_\infty)}{\tau} E_h(t) - \sigma E_h(t) + \dfrac{1}{\tau} P_h(t) + \zeta_h^E(t), \\ \dfrac{d}{dt} P_h(t) = \dfrac{(\varepsilon_s - \varepsilon_\infty)}{\tau} E_h(t) - \dfrac{1}{\tau} P_h(t) + \zeta_h^P(t). \end{cases} \tag{6.45}$$

Next, substituting the exact solutions $E_h(t)$, $H_h(t)$ and $P_h(t)$ into the locally implicit scheme (6.27) gives the perturbed scheme containing defects (truncation errors) composed of a temporal and a spatial error part. Let δ_k denote the defects for the stages $k = 1, 2, 3$ and 4, we then have the following perturbed scheme

$$\begin{cases} \mu \dfrac{H_h(t_{n+1/2}) - H_h(t_n)}{\Delta t} &= -\dfrac{1}{2} S^T E_h(t_n) + \delta_1, \\ \varepsilon_\infty \dfrac{E_h(t_{n+1}) - E_h(t_n)}{\Delta t} &= S_0 H_h(t_{n+1/2}) + \dfrac{1}{2} S_1 (H_h(t_{n+1}) + H_h(t_n)) \\ &- \dfrac{(\varepsilon_s - \varepsilon_\infty)}{2\tau} (E_h(t_{n+1}) + E_h(t_n)) - \dfrac{1}{2} \sigma (E_h(t_{n+1}) + E_h(t_n)) \\ &+ \dfrac{1}{2\tau} (P_h(t_{n+1}) + P_h(t_n)) + \delta_2, \\ \dfrac{P_h(t_{n+1}) - P_h(t_n)}{\Delta t} &= \dfrac{(\varepsilon_s - \varepsilon_\infty)}{2\tau} (E_h(t_{n+1}) + E_h(t_n)) - \dfrac{1}{2\tau} (P_h(t_{n+1}) + P_h(t_n)) \\ &+ \delta_3, \\ \mu \dfrac{H_h(t_{n+1}) - H_h(t_{n+1/2})}{\Delta t} &= -\dfrac{1}{2} S^T E_h(t_{n+1}) + \delta_4. \end{cases} \tag{6.46}$$

From the first equation of (6.45) we write

$$-\frac{1}{2} S^T E_h(t_n) = \frac{\mu}{2} H_h'(t_n) - \frac{1}{2} \zeta_h^H(t_n), \tag{6.47}$$

and

$$-\frac{1}{2} S^T E_h(t_{n+1}) = \frac{\mu}{2} H_h'(t_{n+1}) - \frac{1}{2} \zeta_h^H(t_{n+1}). \tag{6.48}$$

From the second equation of (6.45) we write

$$\begin{aligned} &-\frac{(\varepsilon_s - \varepsilon_\infty)}{2\tau} (E_h(t_{n+1}) + E_h(t_n)) - \frac{1}{2} \sigma (E_h(t_{n+1}) + E_h(t_n)) + \frac{1}{2\tau} (P_h(t_{n+1}) + P_h(t_n)) \\ &= \frac{1}{2} \varepsilon_\infty \left(E_h'(t_{n+1}) + E_h'(t_n) \right) - \frac{1}{2} S (H_h(t_{n+1}) + H_h(t_n)) - \frac{1}{2} \left(\zeta_h^E(t_{n+1}) + \zeta_h^E(t_n) \right). \end{aligned} \tag{6.49}$$

And from the third equation of (6.45) we write

$$\begin{aligned} \frac{(\varepsilon_s - \varepsilon_\infty)}{2\tau} (E_h(t_{n+1}) + E_h(t_n)) - \frac{1}{2\tau} (P_h(t_{n+1}) + P_h(t_n)) &= \frac{1}{2} \left(P_h'(t_{n+1}) + P_h'(t_n) \right) \\ &- \frac{1}{2} \left(\zeta_h^P(t_{n+1}) + \zeta_h^P(t_n) \right). \end{aligned} \tag{6.50}$$

Inserting the expressions (6.47), (6.48), (6.49) and (6.50) into the perturbed scheme (6.46) yields the defect expressions

$$\begin{aligned}
\delta_1 &= \mu\frac{H_h\left(t_{n+1/2}\right)-H_h\left(t_n\right)}{\Delta t}-\frac{\mu}{2}H_h^{'}\left(t_n\right)+\frac{1}{2}\zeta_h^H\left(t_n\right),\\
\delta_2 &= \varepsilon_\infty\frac{E_h\left(t_{n+1}\right)-E_h\left(t_n\right)}{\Delta t}-\frac{\varepsilon_\infty}{2}\left(E_h^{'}\left(t_{n+1}\right)+E_h^{'}\left(t_n\right)\right)\\
&- S_0\left[H_h\left(t_{n+1/2}\right)-\frac{1}{2}\left(H_h\left(t_{n+1}\right)+H_h\left(t_n\right)\right)\right]+\frac{1}{2}\left(\zeta_h^E\left(t_{n+1}\right)+\zeta_h^E\left(t_n\right)\right),\\
\delta_3 &= \frac{P_h\left(t_{n+1}\right)-P_h\left(t_n\right)}{\Delta t}-\frac{1}{2}\left(P_h^{'}\left(t_{n+1}\right)+P_h^{'}\left(t_n\right)\right)+\frac{1}{2}\left(\zeta_h^P\left(t_{n+1}\right)+\zeta_h^P\left(t_n\right)\right),\\
\delta_4 &= \mu\frac{H_h\left(t_{n+1}\right)-H_h\left(t_{n+1/2}\right)}{\Delta t}-\frac{\mu}{2}H_h^{'}\left(t_{n+1}\right)+\frac{1}{2}\zeta_h^H\left(t_{n+1}\right).
\end{aligned} \tag{6.51}$$

Herein we can distinguish the temporal error parts and the spatial error parts contained in the ζ_h^E, ζ_h^H and ζ_h^P contributions. Our interest lies in the temporal errors. We therefore simplify our derivations by omitting these spatial contributions. This is not essential. Carrying the spatial contributions along in the derivations just complicates the formulas and will not lead to different conclusions for the temporal errors. Finally, the formal Taylor expansion at $t_{n+1/2}$ delivers the temporal defect expressions

$$\begin{aligned}
\delta_1 &= \mu\sum_{k=2}\left(\frac{1}{(k-1)!}-\frac{1}{k!}\right)\frac{(-1)^k}{2^k}\left(\Delta t\right)^{k-1}H_h^{(k)},\\
\delta_2 &= \varepsilon_\infty\sum_{k=2'}\frac{-k}{2^k\left(k+1\right)!}\left(\Delta t\right)^k E_h^{(k+1)}+S_0\sum_{k=2'}\frac{1}{2^k k!}\left(\Delta t\right)^k H_h^{(k)} \;\equiv\; \delta_5+S_0\delta_6,\\
\delta_3 &= \sum_{k=2'}\frac{-k}{2^k\left(k+1\right)!}\left(\Delta t\right)^k P_h^{(k+1)},\\
\delta_4 &= \mu\sum_{k=2}\left(\frac{1}{k!}-\frac{1}{(k-1)!}\right)\frac{1}{2^k}\left(\Delta t\right)^{k-1}H_h^{(k)},
\end{aligned} \tag{6.52}$$

where $k=2'$ means even values for k only, and $E_h^{(k)}$, $H_h^{(k)}$ and $P_h^{(k)}$ the k-th derivatives of $E_h\left(t\right)$, $H_h\left(t\right)$ and $P_h\left(t\right)$ at $t=t_{n+1/2}$. Note that δ_1 and δ_4 start with Δt and δ_3, δ_5 and δ_6 with Δt^2.

6.2.5.2 The global error recursion

We introduce the global errors $\mathscr{E}_n^E=E_h\left(t_n\right)-E^n$, $\mathscr{E}_n^H=H_h\left(t_n\right)-H^n$ and $\mathscr{E}_n^P=P_h\left(t_n\right)-P^n$ and the intermediate global error $\mathscr{E}_{n+1/2}^H=H_h\left(t_{n+1/2}\right)-H^{n+1/2}$. Subtracting (6.46) from (6.27) gives the

global errors

$$\begin{aligned}
\mathscr{E}^H_{n+1/2} &= \mathscr{E}^H_n - \frac{\Delta t}{2\mu} S^T \mathscr{E}^E_n + \frac{\Delta t}{\mu}\delta_1, \\
\mathscr{E}^E_{n+1} &= \mathscr{E}^E_n + \frac{\Delta t}{\varepsilon_\infty} S_0 \mathscr{E}^H_{n+1/2} + \frac{\Delta t}{2\varepsilon_\infty} S_1 \left(\mathscr{E}^H_{n+1} + \mathscr{E}^H_n\right) - \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\varepsilon_\infty \tau}\left(\mathscr{E}^E_{n+1} + \mathscr{E}^E_n\right) \\
&- \frac{\Delta t}{2\varepsilon_\infty}\sigma\left(\mathscr{E}^E_{n+1} + \mathscr{E}^E_n\right) + \frac{\Delta t}{2\varepsilon_\infty \tau}\left(\mathscr{E}^P_{n+1} + \mathscr{E}^P_n\right) + \frac{\Delta t}{\varepsilon_\infty}\delta_2, \\
\mathscr{E}^P_{n+1} &= \mathscr{E}^P_n + \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\tau}\left(\mathscr{E}^E_{n+1} + \mathscr{E}^E_n\right) - \frac{\Delta t}{2\tau}\left(\mathscr{E}^P_{n+1} + \mathscr{E}^P_n\right) + \Delta t\,\delta_3, \\
\mathscr{E}^H_{n+1} &= \mathscr{E}^H_{n+1/2} - \frac{\Delta t}{2\mu} S^T \mathscr{E}^E_{n+1} + \frac{\Delta t}{\mu}\delta_4.
\end{aligned} \tag{6.53}$$

From the first and fourth equations of (6.53) we get

$$\begin{aligned}
\mathscr{E}^H_{n+1/2} &= \mathscr{E}^H_n - \frac{\Delta t}{2\mu} S^T \mathscr{E}^E_n + \frac{\Delta t}{\mu}\delta_1, \\
\mathscr{E}^H_{n+1/2} &= \mathscr{E}^H_{n+1} + \frac{\Delta t}{2\mu} S^T \mathscr{E}^E_{n+1} - \frac{\Delta t}{\mu}\delta_4.
\end{aligned} \tag{6.54}$$

Eliminating the intermediate error in the second equation of the global error scheme by inserting half of each expression of (6.54) yields

$$\begin{aligned}
\mathscr{E}^E_{n+1} &= \mathscr{E}^E_n + \frac{\Delta t}{\varepsilon_\infty} S_0 \left[\frac{1}{2}\left(\mathscr{E}^H_{n+1} + \mathscr{E}^H_n\right) + \frac{\Delta t}{4\mu} S^T \left(\mathscr{E}^E_{n+1} - \mathscr{E}^E_n\right) + \frac{\Delta t}{2\mu}\left(\delta_1 - \delta_4\right)\right] \\
&+ \frac{\Delta t}{2\varepsilon_\infty} S_1 \left(\mathscr{E}^H_{n+1} + \mathscr{E}^H_n\right) - \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\varepsilon_\infty \tau}\left(\mathscr{E}^E_{n+1} + \mathscr{E}^E_n\right) - \frac{\Delta t}{2\varepsilon_\infty}\sigma\left(\mathscr{E}^E_{n+1} + \mathscr{E}^E_n\right) \\
&+ \frac{\Delta t}{2\varepsilon_\infty \tau}\left(\mathscr{E}^P_{n+1} + \mathscr{E}^P_n\right) + \frac{\Delta t}{\varepsilon_\infty}\delta_2, \\
&= \mathscr{E}^E_n - \left(\frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\varepsilon_\infty \tau} + \frac{\Delta t}{2\varepsilon_\infty}\sigma\right)\left(\mathscr{E}^E_{n+1} + \mathscr{E}^E_n\right) + \frac{\Delta t^2}{4\varepsilon_\infty \mu} S_0 S^T \left(\mathscr{E}^E_{n+1} - \mathscr{E}^E_n\right) \\
&+ \frac{\Delta t}{2\varepsilon_\infty} S\left(\mathscr{E}^H_{n+1} + \mathscr{E}^H_n\right) + \frac{\Delta t}{2\varepsilon_\infty \tau}\left(\mathscr{E}^P_{n+1} + \mathscr{E}^P_n\right) + \frac{\Delta t}{\varepsilon_\infty}\left(\delta_2 + \frac{\Delta t}{2\mu} S_0 \left(\delta_1 - \delta_4\right)\right).
\end{aligned} \tag{6.55}$$

We also eliminate the intermediate error in the fourth equation of (6.53) by using the first expression of (6.54) to obtain the following global errors scheme

$$\begin{aligned}
\mathscr{E}^E_{n+1} &= \mathscr{E}^E_n - \left(\frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\varepsilon_\infty \tau} + \frac{\Delta t}{2\varepsilon_\infty}\sigma\right)\left(\mathscr{E}^E_{n+1} + \mathscr{E}^E_n\right) + \frac{\Delta t^2}{4\varepsilon_\infty \mu} S_0 S^T \left(\mathscr{E}^E_{n+1} - \mathscr{E}^E_n\right) \\
&+ \frac{\Delta t}{2\varepsilon_\infty} S\left(\mathscr{E}^H_{n+1} + \mathscr{E}^H_n\right) + \frac{\Delta t}{2\varepsilon_\infty \tau}\left(\mathscr{E}^P_{n+1} + \mathscr{E}^P_n\right) + \frac{\Delta t}{\varepsilon_\infty} + \Delta t\,\delta^E_n, \\
\mathscr{E}^P_{n+1} &= \mathscr{E}^P_n - \frac{\Delta t}{2\tau}\left(\mathscr{E}^P_{n+1} + \mathscr{E}^P_n\right) + \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\tau}\left(\mathscr{E}^E_{n+1} + \mathscr{E}^E_n\right) + \Delta t\,\delta^P_n, \\
\mathscr{E}^H_{n+1} &= \mathscr{E}^H_n - \frac{\Delta t}{2\mu} S^T \left(\mathscr{E}^E_{n+1} + \mathscr{E}^E_n\right) + \Delta t\,\delta^H_n,
\end{aligned} \tag{6.56}$$

where

$$\begin{aligned}\delta_n^E &= \frac{1}{\varepsilon_\infty}\left(\delta_2+\frac{\Delta t}{2\mu}S_0\left(\delta_1-\delta_4\right)\right) = \frac{1}{\varepsilon_\infty}\left(\delta_5+S_0\left(\frac{\Delta t}{2\mu}\left(\delta_1-\delta_4\right)+\delta_6\right)\right),\\ \delta_n^P &= \delta_3,\\ \delta_n^H &= \frac{1}{\mu}\left(\delta_1+\delta_4\right).\end{aligned} \tag{6.57}$$

From the expressions of δ_k with $k=1,2,3,4,5$ in (6.52) we observe that these three new defects contain only even terms in Δt and start with Δt^2. At this stage we assume that $E_h(t)$, $H_h(t)$ and $P_h(t)\in C^3[0,T]$. It follows from the remainder in Taylor's theorem that for $\Delta t \backsim h$, $h\to 0$

$$\delta_5=\mathscr{O}\left(\Delta t^2\right),\ \frac{\Delta t}{2\mu}\left(\delta_1-\delta_4\right)+\delta_6=\mathscr{O}\left(\Delta t^2\right),\ \delta_n^P=\mathscr{O}\left(\Delta t^2\right) \text{ and } \delta_n^H=\mathscr{O}\left(\Delta t^2\right). \tag{6.58}$$

Let

$$\mathscr{E}_n=\begin{pmatrix}\mathscr{E}_n^H\\ \mathscr{E}_n^P\\ \mathscr{E}_n^E\end{pmatrix} \quad \text{and} \quad \delta_n=\begin{pmatrix}\delta_n^H\\ \delta_n^P\\ \delta_n^E\end{pmatrix}, \tag{6.59}$$

then from (6.56) we can write the global error in a more compact form (one-step recurrence relation)

$$\mathscr{E}_{n+1}=R\mathscr{E}_n+\Delta t\rho_n,\quad R=R_L^{-1}R_R,\quad \rho_n=R_L^{-1}\delta_n, \tag{6.60}$$

where

$$R_L=\begin{pmatrix} I & 0 & \frac{\Delta t}{2\mu}S^T\\ 0 & \left(1+\frac{\Delta t}{2\tau}\right)I & -\frac{(\varepsilon_s-\varepsilon_\infty)\Delta t}{2\tau}I\\ -\frac{\Delta t}{2\varepsilon_\infty}S & -\frac{\Delta t}{2\varepsilon_\infty\tau}I & \left(1+\frac{(\varepsilon_s-\varepsilon_\infty)\Delta t}{2\varepsilon_\infty\tau}+\frac{\Delta t}{2\varepsilon_\infty}\sigma\right)I-\frac{\Delta t^2}{4\varepsilon_\infty\mu}S_0S^T\end{pmatrix}, \tag{6.61}$$

$$R_R=\begin{pmatrix} I & 0 & -\frac{\Delta t}{2\mu}S^T\\ 0 & \left(1-\frac{\Delta t}{2\tau}\right)I & \frac{(\varepsilon_s-\varepsilon_\infty)\Delta t}{2\tau}I\\ \frac{\Delta t}{2\varepsilon_\infty}S & \frac{\Delta t}{2\varepsilon_\infty\tau}I & \left(1-\frac{(\varepsilon_s-\varepsilon_\infty)\Delta t}{2\varepsilon_\infty\tau}-\frac{\Delta t}{2\varepsilon_\infty}\sigma\right)I-\frac{\Delta t^2}{4\varepsilon_\infty\mu}S_0S^T\end{pmatrix}, \tag{6.62}$$

and $\mathscr{E}_n$, $\Delta t\rho_n$ and δ_n are respectively the (space-time) global, local and truncation errors.

The recursion (6.60) has the standard form featured in the convergence analysis of one-step integration methods, see e.g. [Hundsdorfer & Verwer 2003]. Assuming Lax-Richtmyer stability, whereby we include R_L inversely bounded for $\Delta t \backsim h$, $h\to 0$, it transfers local errors to the global error by essentially adding all local errors. It reveals second-order ODE convergence for a fixed

spatial dimension since then S_0 within the defect δ_n^E is bounded and hence $\rho_n = \mathscr{O}\left(\Delta t^2\right)$ for $\Delta t \to 0$, because (6.57) and (6.58) yield δ_n^E, δ_n^P and δ_n^H are $\mathscr{O}\left(\Delta t^2\right)$. However, for a simultaneous space-time grid refinement, the local error component δ_n^E must have components which will grow with h^{-1}. This growth is unavoidable, since by definition of $S_1 = SS_H$ we have

$$S_0 = S - S_1 = S\left(I - S_H\right), \tag{6.63}$$

and $S = \mathscr{O}\left(h^{-1}\right)$. Therefore we have $\delta_n^E = \mathscr{O}\left(\Delta t\right)$ for $\Delta t \backsim h$, $h \to 0$. The local errors δ_n^P and δ_n^H causes no problem as it contain only solution derivatives, thus δ_n^P and δ_n^H are $\mathscr{O}\left(\Delta t^2\right)$, for $\Delta t \backsim h$, $h \to 0$.

Fortunately, this order reduction by one unit of Δt manifests itself only in the local error and cancel in the transition from local to global error. This cancellation can be proven by transforming the global error scheme (6.60) into one where local errors remain second-order for $\Delta t \backsim h$, $h \to 0$.

6.2.5.3 A transformed global error recursion

The transformation emanates from [Hundsdorfer & Verwer 2003], Lemma II.2.3., reveals that the second-order will be maintained for any stable space-time grid refinement $\Delta t \backsim h$, $h \to 0$. Note that we have already used this transformation in Chapter 3, see Section 3.3.3.6, Lemma 3.3.1. From this lemma we can assume that if the local error $\Delta t \rho_n$ allows a decomposition

$$\Delta t \rho_n = \left(I - R\right)\xi_n + \eta_n, \text{ such that } \xi_n = \mathscr{O}\left(\Delta t^2\right) \text{ and } \eta_n = \mathscr{O}\left(\Delta t^3\right), \text{ for } \Delta t \backsim h, h \to 0, \tag{6.64}$$

then we have the desired second-order convergence for $\mathscr{E}_n$. Therefore, there remains to check (6.64), which amounts to examining

$$\Delta t \rho_n = \Delta t R_L^{-1} \delta_n = \left(I - R\right)\xi_n + \eta_n = \left(I - R_L^{-1} R_R\right)\xi_n + \eta_n, \tag{6.65}$$

or equivalently,

$$\Delta t R_L^{-1} \begin{pmatrix} \delta_n^H \\ \delta_n^P \\ \delta_n^E \end{pmatrix} = R_L^{-1}\left(R_L - R_R\right) \begin{pmatrix} \xi_n^H \\ \xi_n^P \\ \xi_n^E \end{pmatrix} + R_L^{-1} R_L \begin{pmatrix} \eta_n^H \\ \eta_n^P \\ \eta_n^E \end{pmatrix}, \tag{6.66}$$

or

$$\begin{pmatrix} \Delta t \delta_n^H \\ \Delta t \delta_n^P \\ \Delta t \delta_n^E \end{pmatrix} = \left(R_L - R_R\right) \begin{pmatrix} \xi_n^H \\ \xi_n^P \\ \xi_n^E \end{pmatrix} + R_L \begin{pmatrix} \eta_n^H \\ \eta_n^P \\ \eta_n^E \end{pmatrix}, \tag{6.67}$$

or

$$\begin{aligned}
\Delta t \delta_n^H &= \frac{\Delta t}{\mu} S^T \xi_n^E + \eta_n^H + \frac{\Delta t}{2\mu} S^T \eta_n^E, \\
\Delta t \delta_n^P &= \frac{\Delta t}{\tau} \xi_n^P - \frac{\left(\varepsilon_s - \varepsilon_\infty\right)\Delta t}{\tau} \xi_n^E + \left(1 + \frac{\Delta t}{2\tau}\right) \eta_n^P - \frac{\left(\varepsilon_s - \varepsilon_\infty\right)\Delta t}{2\tau} \eta_n^E, \\
\Delta t \delta_n^E &= -\frac{\Delta t}{\varepsilon_\infty} S \xi_n^H - \frac{\Delta t}{\varepsilon_\infty \tau} \xi_n^P + \frac{\left(\varepsilon_s - \varepsilon_\infty\right)\Delta t}{\varepsilon_\infty \tau} \xi_n^E + \frac{\Delta t}{\varepsilon_\infty} \sigma \xi_n^E \\
&\quad - \frac{\Delta t}{2\varepsilon_\infty} S \eta_n^H - \frac{\Delta t}{2\varepsilon_\infty \tau} \eta_n^P + \eta_n^E + \frac{\left(\varepsilon_s - \varepsilon_\infty\right)\Delta t}{2\varepsilon_\infty \tau} \eta_n^E + \frac{\Delta t}{2\varepsilon_\infty} \sigma \eta_n^E - \frac{\Delta t^2}{4\varepsilon_\infty \mu} S_0 S^T \eta_n^E.
\end{aligned} \tag{6.68}$$

Thus, our task is now to identify error vectors ξ_n^E, ξ_n^H, ξ_n^P and η_n^E, η_n^H, η_n^P in accordance with (6.64) such that (6.68) are satisfied. Let us first define

$$\eta_n^H = \frac{\Delta t}{\mu}(\delta_1 + \delta_4), \eta_n^P = \Delta t \delta_3 \text{ and } \eta_n^E = \frac{\Delta t}{\varepsilon_\infty}\delta_5. \tag{6.69}$$

From (6.58), we observe that η_n^H, η_n^P and η_n^E are $\mathscr{O}\left(\Delta t^3\right)$. Next from (6.68) we identify error vectors ξ_n^E, ξ_n^H, ξ_n^P. From the first equation of (6.68) and the definition of δ_n^H (see (6.57)) we write

$$\frac{\Delta t}{\mu}(\delta_1 + \delta_4) = \frac{\Delta t}{\mu} S^T \xi_n^E + \eta_n^H + \frac{\Delta t}{2\mu} S^T \eta_n^E. \tag{6.70}$$

Substituting $(\Delta t/\mu)(\delta_1 + \delta_4) = \eta_n^H$ into (6.70) reveals the error vector ξ_n^E

$$\xi_n^E = -\frac{1}{2}\eta_n^E \quad \left(= -\frac{\Delta t}{2\varepsilon_\infty}\delta_5\right), \tag{6.71}$$

hence $\xi_n^E = \mathscr{O}\left(\Delta t^3\right)$. Now from the second equation of (6.68) and the definition of δ_n^P (see (6.57)) we write

$$\Delta t \delta_3 = \frac{\Delta t}{\tau}\xi_n^P - \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{\tau}\xi_n^E + \left(1 + \frac{\Delta t}{2\tau}\right)\eta_n^P - \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\tau}\eta_n^E. \tag{6.72}$$

Substituting $\Delta t \delta_3 = \eta_n^P$ into (6.72) reveals the error vector ξ_n^P

$$\xi_n^P = -\frac{1}{2}\eta_n^P \quad \left(= -\frac{\Delta t}{2}\delta_3\right), \tag{6.73}$$

hence $\xi_n^P = \mathscr{O}\left(\Delta t^3\right)$. Thus, it remains to identify ξ_n^H and to check if $\xi_n^H = \mathscr{O}\left(\Delta t^2\right)$. From the third equation of (6.68) and the definition of δ_n^E (see (6.57)) we have

$$\begin{aligned}
\frac{\Delta t}{\varepsilon_\infty}\left(\delta_5 + S_0\left(\frac{\Delta t}{2\mu}(\delta_1 - \delta_4) + \delta_6\right)\right) &= -\frac{\Delta t}{\varepsilon_\infty} S \xi_n^H - \frac{\Delta t}{\varepsilon_\infty \tau}\xi_n^P + \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{\varepsilon_\infty \tau}\xi_n^E \\
&+ \frac{\Delta t}{\varepsilon_\infty}\sigma \xi_n^E - \frac{\Delta t}{2\varepsilon_\infty} S \eta_n^H - \frac{\Delta t}{2\varepsilon_\infty \tau}\eta_n^P + \eta_n^E \\
&+ \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\varepsilon_\infty \tau}\eta_n^E + \frac{\Delta t}{2\varepsilon_\infty}\sigma \eta_n^E - \frac{\Delta t^2}{4\varepsilon_\infty \mu} S_0 S^T \eta_n^E.
\end{aligned} \tag{6.74}$$

Substituting $\frac{\Delta t}{\varepsilon_\infty}\delta_5 = \eta_n^E$ and the expressions of ξ_n^E and ξ_n^P into (6.74) yields

$$\begin{aligned}
\eta_n^E + \frac{\Delta t}{\varepsilon_\infty} S_0\left(\frac{\Delta t}{2\mu}(\delta_1 - \delta_4) + \delta_6\right) &= -\frac{\Delta t}{\varepsilon_\infty} S \xi_n^H + \frac{\Delta t}{2\varepsilon_\infty \tau}\eta_n^P - \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\varepsilon_\infty \tau}\eta_n^E \\
&- \frac{\Delta t}{2\varepsilon_\infty}\sigma \eta_n^E - \frac{\Delta t}{2\varepsilon_\infty} S \eta_n^H - \frac{\Delta t}{2\varepsilon_\infty \tau}\eta_n^P + \eta_n^E \\
&+ \frac{(\varepsilon_s - \varepsilon_\infty)\Delta t}{2\varepsilon_\infty \tau}\eta_n^E + \frac{\Delta t}{2\varepsilon_\infty}\sigma \eta_n^E - \frac{\Delta t^2}{4\varepsilon_\infty \mu} S_0 S^T \eta_n^E,
\end{aligned} \tag{6.75}$$

hence,

$$S\xi_n^H = -\frac{1}{2} S \eta_n^H - \frac{\Delta t}{4\mu} S_0 S^T \eta_n^E - S_0\left(\frac{\Delta t}{2\mu}(\delta_1 - \delta_4) + \delta_6\right). \tag{6.76}$$

Then we are done if we can choose ξ_n^H to satisfy (6.76) such that $\xi_n^H = \mathscr{O}\left(\Delta t^2\right)$. Inserting $S_0 = S\left(I - S_H\right)$ yields

$$\xi_n^H = -\frac{1}{2}\eta_n^H - \frac{\Delta t}{4\mu}\left(I - S_H\right)S^T\eta_n^E - \left(I - S_H\right)\left(\frac{\Delta t}{2\mu}\left(\delta_1 - \delta_4\right) + \delta_6\right). \tag{6.77}$$

The first term is $\mathscr{O}\left(\Delta t^3\right)$, due to $\Delta t S^T = \mathscr{O}(1)$ for $\Delta t \backsim h$, $h \to 0$, the second term is also $\mathscr{O}\left(\Delta t^3\right)$. Finally, from (6.58) the third term is $\mathscr{O}\left(\Delta t^2\right)$, consequently $\xi_n^H = \mathscr{O}\left(\Delta t^2\right)$, which completes the error analysis.

Consequently, we have proven that the subdivision into coarse and fine elements is not detrimental to the second-order ODE convergence of the method (6.27), under stable simultaneous space-time grid refinement towards the exact underlying PDE solution. We summarize this convergence result with the following theorem.

Theorem 6.2.1. *Let $H_h(t)$, $E_h(t)$ and $P_h(t)$ denote the exact solutions of the Maxwell problem in dispersive media under consideration, restricted to the space grid, i.e. the exact solutions of the system of ODEs*

$$\begin{cases} \mu\dfrac{d}{dt}H_h(t) = -S^T E_h(t) + \zeta_h^H(t), \\ \varepsilon_\infty\dfrac{d}{dt}E_h(t) = SH_h(t) - \dfrac{(\varepsilon_s - \varepsilon_\infty)}{\tau}E_h(t) - \sigma E_h(t) + \dfrac{1}{\tau}P_h(t) + \zeta_h^E(t), \\ \dfrac{d}{dt}P_h(t) = \dfrac{(\varepsilon_s - \varepsilon_\infty)}{\tau}E_h(t) - \dfrac{1}{\tau}P_h(t) + \zeta_h^P(t). \end{cases} \tag{6.78}$$

where ζ_h^H, ζ_h^E and ζ_h^P denote the spatial truncation errors. Assume a Lax-Richtmyer stable space-time grid refinement $\Delta t \sim h$, $h \to 0$. On the interval $[0,T]$ the approximations H^n, E^n and P^n of method (6.27) then converge with order two to $H_h(t)$, $E_h(t)$ and $P_h(t)$.

6.3 Numerical results

The simulations discussed in this section have been performed on a workstation equipped with an Intel Xeon 2.40 GHz processor and 16 GB of RAM memory. As in Chapter 4 the linear system associated to the locally implicit method (6.27) (see (6.31) and (3.65)) is solved using the MUMPS (MUltifrontal Massively Parallel sparse direct Solver) optimized sparse direct solver [Amestoy *et al.* 2000].

We consider the three-dimensional (normalized) Maxwell's equations in Debye dispersive media

$$
\begin{cases}
\mu \dfrac{\partial H^x}{\partial t}(\mathbf{x},t) = \dfrac{\partial E^y}{\partial z}(\mathbf{x},t) - \dfrac{\partial E^z}{\partial y}(\mathbf{x},t), \\
\mu \dfrac{\partial H^y}{\partial t}(\mathbf{x},t) = \dfrac{\partial E^z}{\partial x}(\mathbf{x},t) - \dfrac{\partial E^x}{\partial z}(\mathbf{x},t), \\
\mu \dfrac{\partial H^z}{\partial t}(\mathbf{x},t) = \dfrac{\partial E^x}{\partial y}(\mathbf{x},t) - \dfrac{\partial E^y}{\partial z}(\mathbf{x},t), \\
\varepsilon_\infty \dfrac{\partial E^x}{\partial t}(\mathbf{x},t) = \dfrac{\partial H^z}{\partial y}(\mathbf{x},t) - \dfrac{\partial H^y}{\partial z}(\mathbf{x},t) - \left(\dfrac{\varepsilon_s - \varepsilon_\infty}{\tau} + \sigma\right) E^x(\mathbf{x},t) + \dfrac{1}{\tau} P^x(\mathbf{x},t) - J^{s,x}(\mathbf{x},t), \\
\varepsilon_\infty \dfrac{\partial E^y}{\partial t}(\mathbf{x},t) = \dfrac{\partial H^x}{\partial z}(\mathbf{x},t) - \dfrac{\partial H^z}{\partial x}(\mathbf{x},t) - \left(\dfrac{\varepsilon_s - \varepsilon_\infty}{\tau} + \sigma\right) E^y(\mathbf{x},t) + \dfrac{1}{\tau} P^y(\mathbf{x},t) - J^{s,y}(\mathbf{x},t), \\
\varepsilon_\infty \dfrac{\partial E^z}{\partial t}(\mathbf{x},t) = \dfrac{\partial H^y}{\partial x}(\mathbf{x},t) - \dfrac{\partial H^x}{\partial y}(\mathbf{x},t) - \left(\dfrac{\varepsilon_s - \varepsilon_\infty}{\tau} + \sigma\right) E^z(\mathbf{x},t) + \dfrac{1}{\tau} P^z(\mathbf{x},t) - J^{s,z}(\mathbf{x},t), \\
\varepsilon_\infty \dfrac{\partial P^x}{\partial t}(\mathbf{x},t) = \dfrac{(\varepsilon_s - \varepsilon_\infty)}{\tau} E^x(\mathbf{x},t) + \dfrac{1}{\tau} P^x(\mathbf{x},t), \\
\varepsilon_\infty \dfrac{\partial P^y}{\partial t}(\mathbf{x},t) = \dfrac{(\varepsilon_s - \varepsilon_\infty)}{\tau} E^y(\mathbf{x},t) + \dfrac{1}{\tau} P^y(\mathbf{x},t), \\
\varepsilon_\infty \dfrac{\partial P^z}{\partial t}(\mathbf{x},t) = \dfrac{(\varepsilon_s - \varepsilon_\infty)}{\tau} E^z(\mathbf{x},t) + \dfrac{1}{\tau} P^z(\mathbf{x},t),
\end{cases}
\tag{6.79}
$$

where $\mathbf{E} = (E^x, E^y, E^z)^T$, $\mathbf{H} = (H^x, H^y, H^z)^T$ and $\mathbf{P} = (P^x, P^y, P^z)^T$ denote the electric field, the magnetic field and the electric polarization, respectively. ε_∞ and ε_s are the infinite and static relative permittivity and τ is the Debye relaxation time constant, characteristic of the material. Finally, μ, σ are coefficients representing relative magnetic permeability and conductivity, and $\mathbf{J}^s\,(J^{s,x}, J^{s,y}, J^{s,z})^T$ represents the source current density maintained by external sources. In the following, the equations (6.79) are space discretized using the DG method formulated on tetrahedral meshes, introduced in Chapter 2. In the preliminary implementation of this DG method, the approximation of the electromagnetic field components within a tetrahedron τ_i relies on a nodal $\mathbb{P}_k$ interpolation method. The a priori convergence analysis for this DG method based on a centered numerical flux and formulated on simplicial meshes shows that the convergence rate is $\mathcal{O}(h^k)$ for a k-th interpolation order [Lanteri & Scheid 2013]. We denote by Ω_h the computational domain and Ω_h^{exp} the set of tetrahedra that belong to the region where the explicit method is used into the implicit-explicit splitting. The critical step size, denoted Δt_c, used in the numerical tests is then given by (4.15).

6.3.1 An artificial validation test

We establish an artificial three-dimensional problem to validate the code for solving the time domain Maxwell's equations in dispersive media with the Debye model. To that purpose we substitute the solution of the propagation of a standing wave in a cubic PEC cavity (see (4.16)) into the third equation of (6.18) to obtain the analytic expression of the electric polarization. Then we substitute the electric field and the polarization into the second equation of (6.18) to obtain the current density

to impose. This yields the exact solution

$$\begin{cases} H^x(\mathbf{x},t) = -\dfrac{\pi}{\omega}\sin(\pi x)\cos(\pi y)\cos(\pi z)\sin(\omega t), \\ H^y(\mathbf{x},t) = \dfrac{2\pi}{\omega}\cos(\pi x)\sin(\pi y)\cos(\pi z)\sin(\omega t), \\ H^z(\mathbf{x},t) = -\dfrac{\pi}{\omega}\cos(\pi x)\cos(\pi y)\sin(\pi z)\sin(\omega t), \\ E^x(\mathbf{x},t) = -\cos(\pi x)\sin(\pi y)\sin(\pi z)\cos(\omega t), \\ E^y(\mathbf{x},t) = 0, \\ E^z(\mathbf{x},t) = \sin(\pi x)\sin(\pi y)\cos(\pi z)\cos(\omega t), \\ P^x(\mathbf{x},t) = -\dfrac{(\varepsilon_s-\varepsilon_\infty)\tau}{1+\omega^2\tau^2}\cos(\pi x)\sin(\pi y)\sin(\pi z)\left[\dfrac{1}{\tau}\cos(\omega t)+\omega\sin(\omega t)\right], \\ P^y(\mathbf{x},t) = 0, \\ P^z(\mathbf{x},t) = -\dfrac{(\varepsilon_s-\varepsilon_\infty)\tau}{1+\omega^2\tau^2}\sin(\pi x)\sin(\pi y)\cos(\pi z)\left[\dfrac{1}{\tau}\cos(\omega t)+\omega\sin(\omega t)\right], \end{cases} \tag{6.80}$$

if we impose an artificial source current density defined by

$$\begin{cases} J^{s,x}(\mathbf{x},t) &=& \cos(\pi x)\sin(\pi y)\sin(\pi z) \\ &\times& \left[\dfrac{\varepsilon_s-\varepsilon_\infty}{1+\omega^2\tau^2}\left(\dfrac{1}{\tau}\cos(\omega t)+\omega\sin(\omega t)\right)-\left(\dfrac{\varepsilon_s-\varepsilon_\infty}{\tau}+\sigma\right)\cos(\omega t)\right], \\ J^{s,y}(\mathbf{x},t) &=& 0, \\ J^{s,z}(\mathbf{x},t) &=& \dfrac{\varepsilon_s-\varepsilon_\infty}{\tau}\sin(\pi x)\sin(\pi y)\sin(\pi z)\cos(\omega t) \\ &\times& \left[\dfrac{\varepsilon_s-\varepsilon_\infty}{1+\omega^2\tau^2}\left(\dfrac{1}{\tau}\cos(\omega t)+\omega\sin(\omega t)\right)-\left(\dfrac{\varepsilon_s-\varepsilon_\infty}{\tau}+\sigma\right)\cos(\omega t)\right], \end{cases} \tag{6.81}$$

where the angular frequency is given by $\omega = 2\pi f$ (rad·s^{-1}) with $f = 260$ MHz and wavelength $\lambda = 1.15$ m. The computational domain is the cube $\Omega = [0,1]^3$ where we impose a metallic boundary condition $\mathbf{n}\times\mathbf{E} = 0$ on $\partial\Omega$. In our simulations we choose $\mu = \varepsilon_\infty = 1$, $\varepsilon_s = 5$, $\sigma = 0$ and $\tau = 9.4$ picoseconds (ps). The total simulation time is set to $T = 1.67\times10^{-8}$ s which corresponds to $T = 5$ m (normalized unit).

We investigate the space-time convergence order (i.e. for a stable simultaneous space-time grid refinement $\Delta t \backsim h$, $h \to 0$) of the fully explicit method and the locally implicit methods. We measure the maximal L^2-norm of the error for a sequence of four successively locally refined tetrahedral meshes, see Figure 4.17 for examples of meshes with the identification of the region for implicit treatment. We plot this error as a function of $1/h$, in logarithmic scale, the use of the logarithmic scale allows to visualize the convergence rates as the slopes of the curves. We use the DGTD-$\mathbb{P}_k$ method, with $k = 1$ or 2. The obtained results, for both methods, are summarized in Figure 6.1 (on left). The results are similar for both time integration methods, which confirm that the subdivision into coarse and fine elements is not detrimental for the convergence order of the locally implicit

method. The order of convergence is slightly stronger than the theoretical behavior for $k = 1$ (about 1.6 instead of 2.0), while for $k = 2$ we obtain the expected order, i.e. the order two. We also plot in Figure 6.1 (on right) the error as a function of the CPU time. For a given error we can observe the gain of final CPU time with the locally implicit approach compared to the fully explicit method.

Now we focus on the the linear systems to be solved and the efficiency of the locally implicit method. We consider the locally refined tetrahedral mesh composed of 40616 tetrahedra and 7756 vertices (Figure 4.17, on right). First, we are interested in the sparsity of the matrix to be inverted and the cost of the factorization step. In Table 6.2 for each matrix of linear system to solve, we indicate the number of nonzero elements and the fill-in ratio percentage. We also indicated the total size (memory requirement) of all internal data used during numerical factorization and the total CPU time for analysis and factorization. Note that for the fully implicit scheme based on DGTD-$\mathbb{P}_{1,2}$ methods the number of nonzero will be 75081406 and 408726916, respectively, and will require too much memory to be a reasonable alternative to the fully explicit method. Finally we present some numerical results for the locally implicit method and the fully explicit method. In Table 6.3 we observe that the locally implicit method allows to overcome the step size limitations caused by the local refinement. With implicit-explicit approaches the sizes of the time step are about 34 times larger which yields significant gains of final CPU time, about 8 times lower than the fully explicit case. Regarding the time evolution of the error, Figure 6.2, we observe that the IMEX scheme and the fully explicit scheme based on DGTD-$\mathbb{P}_{1,2}$ methods give similar results.

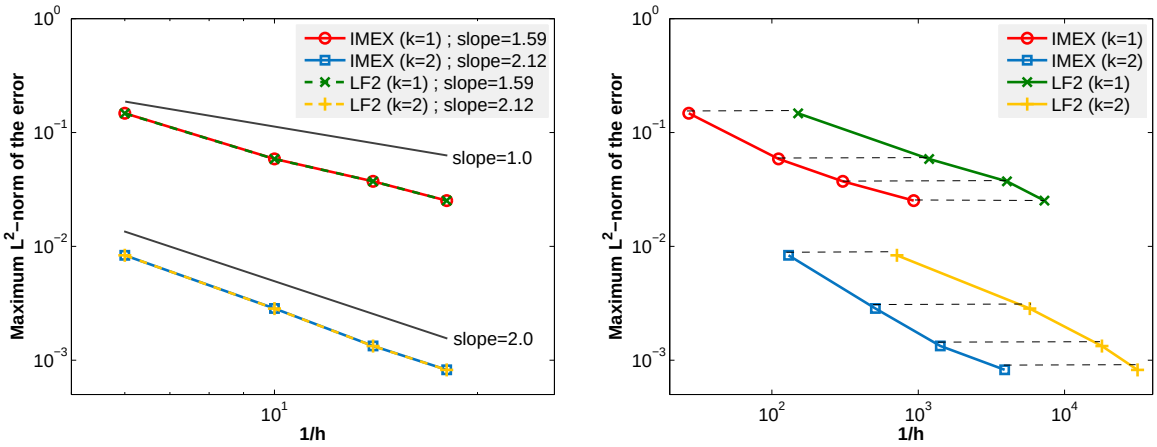


Figure 6.1: Numerical convergence and maximum error (L^2-norm) in function of final CPU time for the locally implicit and fully explicit DGTD-$\mathbb{P}_k$ methods (left - right, respectively).

DGTD-$\mathbb{P}_k$	Matrix order	# nonzero (nz)	Fill-in ratio (percentage)	RAM size (MB)	CPU time (s)
$\mathbb{P}_1$	487392	1247266	0.000525	210	8
$\mathbb{P}_2$	1218480	5437770	0.000366	937	72

Table 6.2: Data and factorization of the matrix of the linear system to be solved for the locally implicit methods, with DGTD-$\mathbb{P}_k$ methods ($k = 1,2$).

	Fully exp. method (LF2)		**Loc. imp. method (IMEX)**		
$\mathbb{P}_k$	Δt (m)	CPU time (s)	Δt (m)	CPU time (s)	$\frac{\text{CPU}_{(\text{LF2})}}{\text{CPU}_{(\text{IMEX})}}$
$\mathbb{P}_1$	1.1247e-4	7285	2.6040e-3	922	7.9
$\mathbb{P}_2$	7.8267e-5	31568	4.1502e-3	3786	8.3

Table 6.3: Critical time step size and CPU time for the fully explicit method and the locally implicit method, with DGTD-$\mathbb{P}_k$ method ($k = 1,2$).

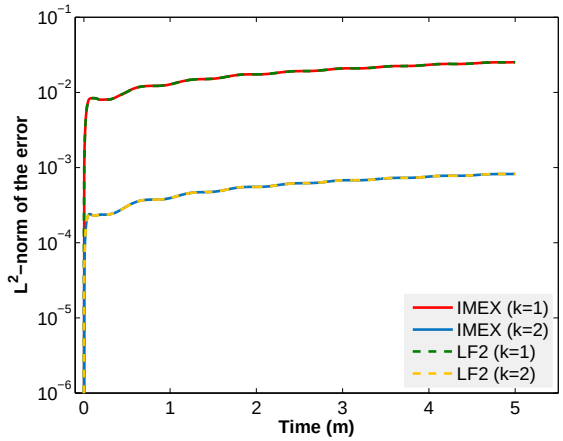


Figure 6.2: Time evolution of the error (L^2-norm) for the locally implicit and fully explicit DGTD-$\mathbb{P}_k$ methods.

6.3.2 Microwave propagation in head tissues

We now present numerical results for a more realistic problem demonstrating the application of the proposed locally implicit DGTD method (6.27) to microwave propagation in biological tissues. For that purpose, we consider a heterogeneous geometrical model of the head tissues based on an unstructured tetrahedral mesh consists of 61358 vertices and 366208 tetrahedra, see Figure 4.21. This geometrical model of the head consists of four tissues namely, the skin, the skull, the CFS and the brain, see Figure 4.20 for the surface meshes of these tissues. The Debye model parameters that we have used for the tissues are given in Table 6.4. The computational domain is here artificially bounded by a sphere on which the Silver-Müller condition is imposed

$$\mathbf{n}\times\mathbf{E}-\mathbf{n}\times(\mathbf{H}\times\mathbf{n})=\mathbf{n}\times\mathbf{E}^{inc}-\mathbf{n}\times\left(\mathbf{H}^{inc}\times\mathbf{n}\right)\quad\text{on }\partial\Omega, \tag{6.82}$$

where $\partial\Omega$ denotes the boundary of the sphere, $\mathbf{n}$ the unit outward normal to $\partial\Omega$ and $\left(\mathbf{E}^{inc},\mathbf{H}^{inc}\right)$ is a given incident field. In the present case, the incident field is a plane wave propagating in the z direction, with a temporal evolution given by a modulated gaussian pulse,

$$s(t)=\mathrm{e}^{-\left(\frac{t-4t_p}{t_p}\right)^2}\sin\left(2\pi f_c\left(t-t_p\right)\right), \tag{6.83}$$

where $t_p=400$ ps and $f_c=1.8$ GHz. The total physical simulation time is set to $T=15\times T_c$ where $T_c=1/f_c$, and a discrete Fourier transform of the electromagnetic field at the frequency f_c is computed on the fly. For the simulations we only consider DGTD-$\mathbb{P}_1$ methods and we choose the same implicit region as in Section 4.2.2 that consists of 5092 tetrahedra, which represents 1.4% of the total number of elements. We recall that the critical time step size for the stability is then about 15.5 times larger than the critical time step size for the stability of the fully explicit method. The localization and the distribution of the elements for the implicit treatment are given in Figure 4.21 and Table 4.13, respectively. Finally, Table 4.14 gives the number of nonzero elements, the fill-in ratio percentage for the matrix of the linear system to be solved, the total size (memory requirement) of all internal data used during numerical factorization and the total CPU time for analysis and factorization. Note that for the above-mentioned incident field the local refinement is not really justified but our aim is here to validate the locally implicit DGTD method (6.27) for a realistic electromagnetic wave propagation problem. For that purpose, we compare the results of two simulations. A first simulation is performed with the locally implicit DGTD-$\mathbb{P}_1$ method (6.27) using the above-mentioned incident field. Then, a second simulation is realized using the original locally implicit DGTD-$\mathbb{P}_1$ method (3.62) for non-dispersive media, using a monochromatic plane wave incident field with frequency f_c and adopting the appropriate values of the electric permittivity and the electric conductivity for each tissue, which are deduced from the Debye model and that we give in Table 6.5.

The total simulation times are equal to 9 h 57 min (dispersive calculation) and 9 h 28 min (non-dispersive calculation). The difference between these two simulation times is due to the additional ADE, but also for a great part due to the fact that the dispersive calculation includes the on the fly calculation of the discrete Fourier transform of the field components during the 15 periods of simulation while, in the non-dispersive case, this operation is only realized during the last period. Time evolution of the electric component E^z is shown on Figure 6.3 for the dispersive and the non-dispersive calculations. On Figures 6.5 - 6.6 we show the contour lines of the local SAR (Specific Absorption Rate) normalized by the maximal local SAR, in logarithmic scale, for the calculations

with the locally implicit DGTD-$\mathbb{P}_1$ method (6.27) (dispersive calculation) and the locally implicit DGTD-$\mathbb{P}_1$ method (3.62) (non-dispersive calculation), respectively. On Figure 6.4 we also show the contour lines of the local SAR normalized by the total emitted power for each simulation. We recall that the SAR is a measure of the rate at which electric energy is absorbed by the tissues when exposed to a radio-frequency electromagnetic field. It represents the power absorbed per mass of tissue and has units of watts per kilogram ($\mathrm{W{\cdot}kg^{-1}}$). The SAR is then defined as $\sigma|\mathbf{E}_{four}|^2/\rho$, where $\mathbf{E}_{four}$ denotes the electric field in the frequency-domain, resulting from the discrete Fourier transform of the temporal field, and ρ is the density which depends on the tissues. We do not observe noticeable differences in the different tissues for the SAR patterns, the obtained results for the locally implicit method (6.27) and the locally implicit method (3.62) are globally similar.

Tissue	Skin	Skull	CSF	Brain
ε_∞	1.0	1.0	2.0	1.0
$\varepsilon_s - \varepsilon_\infty$	38.0	10.5	66.0	43.0
τ (ps)	10.0	20.0	10.0	10.0
σ ($\mathrm{S{\cdot}m^{-1}}$)	0.7	0.1	2.0	0.7

Table 6.4: Microwave propagation in head tissues: Debye model parameters for dispersive calculation using a pulse in time plane wave as the incident field.

Tissue	Skin	Skull	CSF	Brain
ε	38.66	11.60	68.25	43.80
σ ($\mathrm{S{\cdot}m^{-1}}$)	1.18	0.27	2.28	0.97

Table 6.5: Microwave propagation in head tissues: electromagnetic characteristics of tissues for non-dispersive calculation using a monochromatic plane wave as the incident field, at frequency $f_c = 1.8$ GHz.

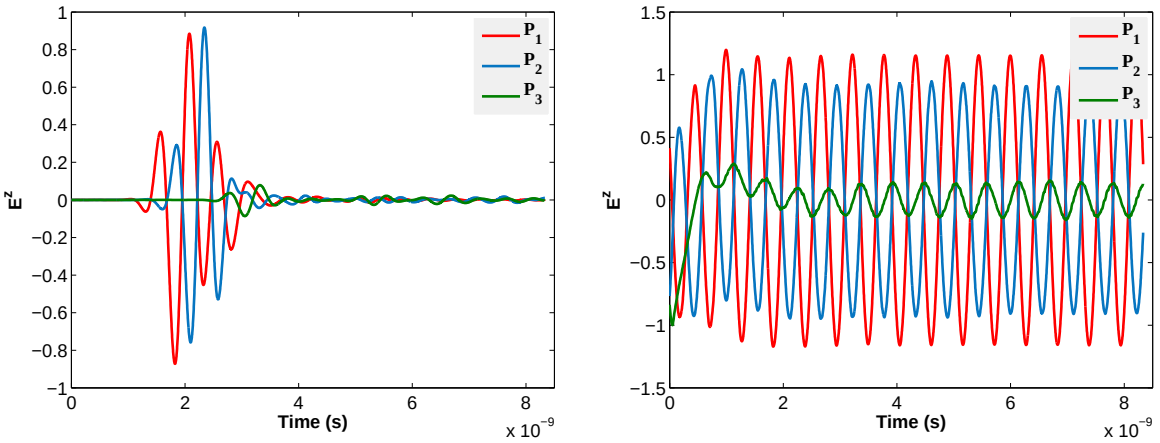


Figure 6.3: Microwave propagation in head tissues: time evolution of the E^z component of the electric field at selected spatial locations, $\mathbf{P}_1 = (-0.1962, -0.0027, -0.0032)$, $\mathbf{P}_2 = (-0.1013, -0.0009, 0.0000)$ and $\mathbf{P}_3 = (0.0985, -0.0019, -0.0004)$ (left plot: dispersive calculation, right plot: non-dispersive calculation).

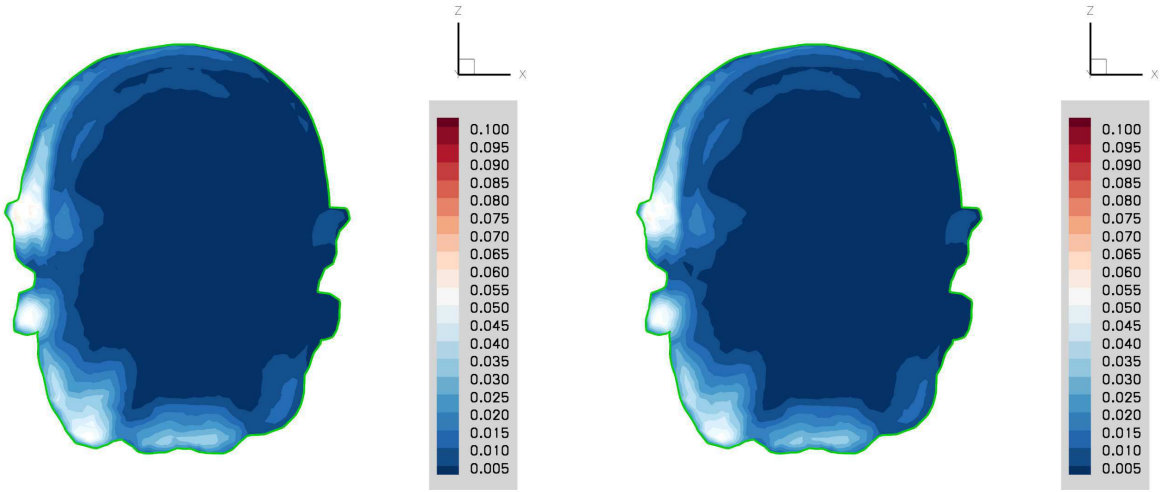


Figure 6.4: Microwave propagation in head tissues: dispersive calculation using a pulse in time plane wave as the incident field (left plot) and non-dispersive calculation using a monochromatic plane wave as the incident field (right plot). Contour lines of the local SAR normalized by the total emitted power.

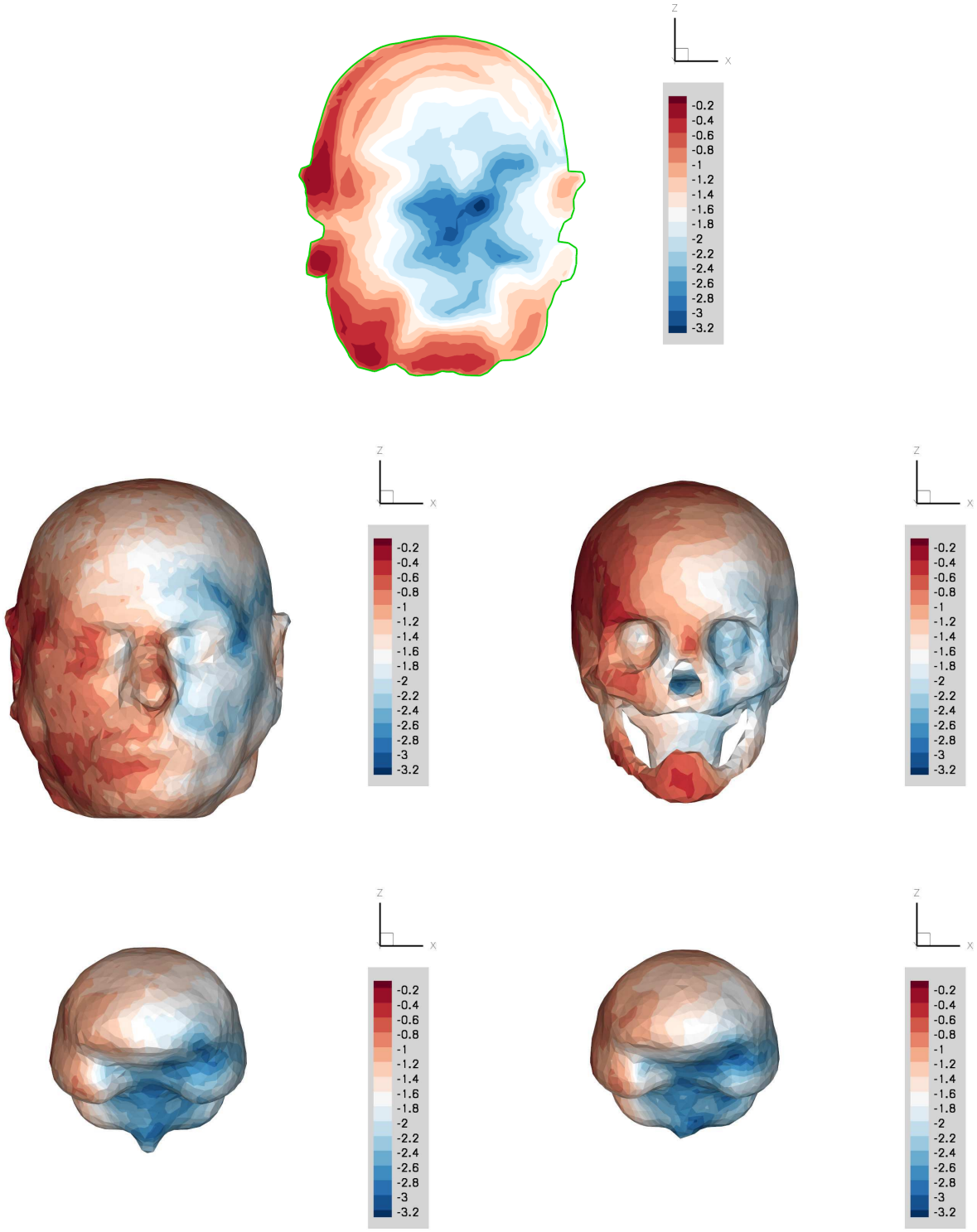


Figure 6.5: Microwave propagation in head tissues: dispersive calculation using a pulse in time plane wave as the incident field. Contour lines of the local SAR normalized by the maximal local SAR (logarithmic scale).

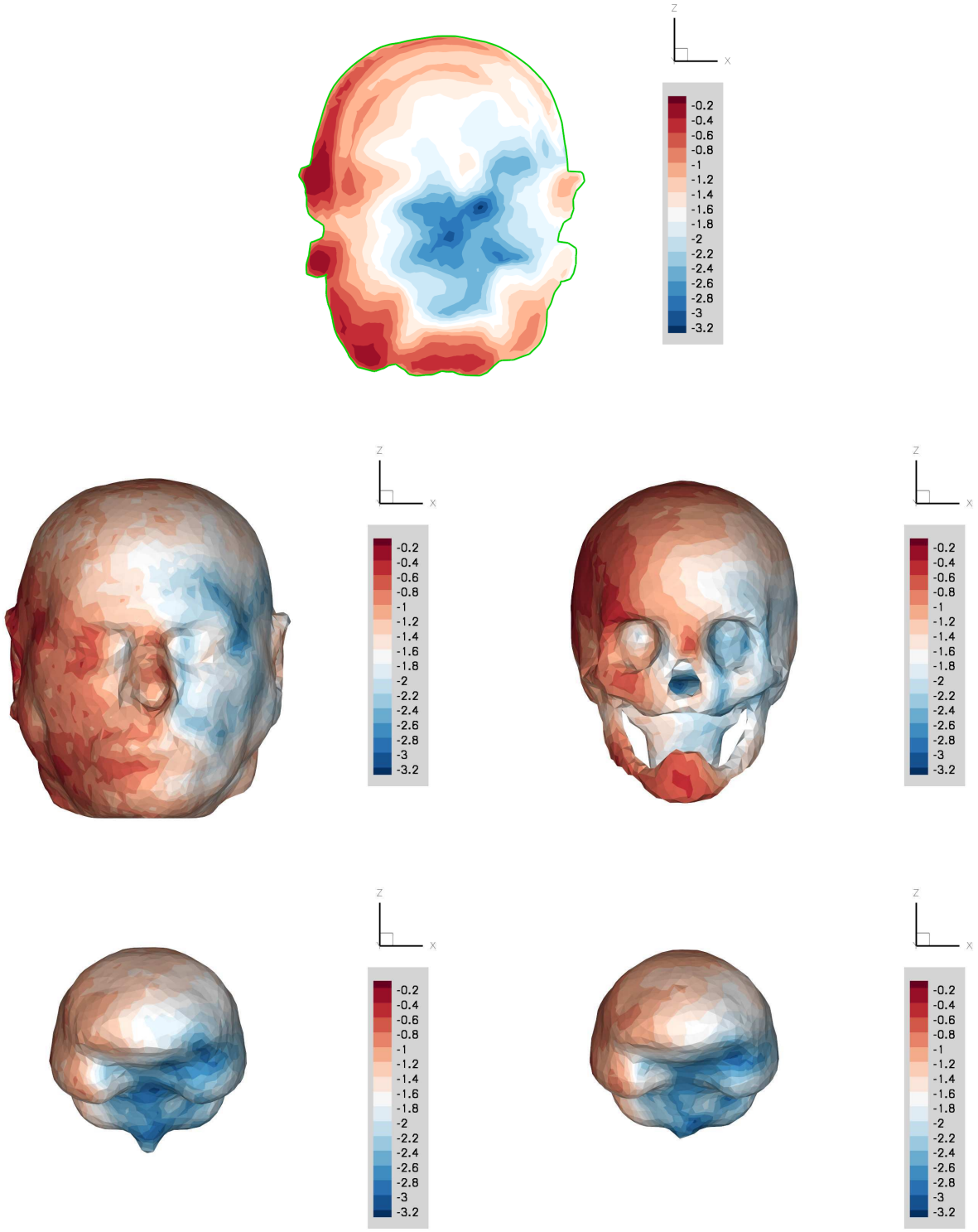


Figure 6.6: Microwave propagation in head tissues: non-dispersive calculation using a monochromatic plane wave as the incident field. Contour lines of the local SAR normalized by the maximal local SAR (logarithmic scale).

6.3.3 Exposure of head tissues to a localized source radiation

We now present numerical results for the exposure of head tissues to a localized source radiation demonstrating the usefulness of the proposed locally implicit DGTD method (6.27). For that purpose, we consider the same heterogeneous geometrical model of the head tissues as the previous test case based on an unstructured tetrahedral mesh consists of 61358 vertices and 366208 tetrahedra, see Figure 4.21. We recall that this geometrical model of the head consists of four tissues namely, the skin, the skull, the CFS and the brain, see Figure 4.20 for the surface meshes of these tissues. The Debye model parameters that we have used for the tissues are given in Table 6.4. The computational domain is artificially bounded by a sphere on which the Silver-Müller condition is imposed

$$\mathbf{n}\times\mathbf{E}-\mathbf{n}\times(\mathbf{H}\times\mathbf{n})=0 \quad \text{on } \partial\Omega, \tag{6.84}$$

where $\partial\Omega$ denotes the boundary of the sphere, $\mathbf{n}$ the unit outward normal to $\partial\Omega$. In the present case, there is no incident field and the source term is given by

$$J^{s,z}(\mathbf{x},t)=f_0\,s(t)\,\frac{g(\mathbf{x})}{\|g(\mathbf{x})\|}, \tag{6.85}$$

where f_0 is the amplitude of the signal, $s(t)$ is the temporal evolution that is a modulated gaussian pulse of the form (6.83) where $t_p = 400$ ps and $f_c = 1.8$ GHz, and $g(\mathbf{x})$ is a three-dimensional gaussian function with $(x_0,y_0,z_0)=(-0.100,0.025,-0.015)$ the center of the gaussian spatial support

$$g(\mathbf{x})=\mathrm{e}^{-\alpha\left((x-x_0)^2+(y-y_0)^2+(z-z_0)^2\right)}. \tag{6.86}$$

The total physical simulation time is set to $T = 10\times T_c$ where $T_c = 1/f_c$, and a discrete Fourier transform of the electromagnetic field at the frequency f_c is computed on the fly. For the simulations we only consider DGTD-$\mathbb{P}_1$ methods and we choose the same implicit region as in Section 4.2.2 that consists of 5092 tetrahedra, which represents 1.4% of the total number of elements. The parameter α in (6.86) was chosen such that the source term $J^{s,z}$ is strongly localized, which justifies the local refinement around the center of the gaussian, the support of the gaussian g is then strictly included in the implicit region. The localization and the distribution of the elements for the implicit treatment are given in Figure 4.21 and Table 4.13, respectively. Finally, Table 4.14 gives the number of nonzero elements, the fill-in ratio percentage for the matrix of the linear system to be solved, the total size (memory requirement) of all internal data used during numerical factorization and the total CPU time for analysis and factorization.

A first simulation is performed with the locally implicit DGTD-$\mathbb{P}_1$ method (6.27) and a second simulation is realized using the original locally implicit DGTD-$\mathbb{P}_1$ method (3.62) for non-dispersive media with the same localized source term and adopting the appropriate values of the electric permittivity and the electric conductivity for each tissue, that are given in Table 6.5. Time evolution of the electric component E^z at two selected points is shown on Figure 6.7 for the dispersive and the non-dispersive calculations. We also plot the time evolution of the polarization component P^z for the second point localized in the head. As expected, the curves obtained for the first point, localized at the center of the gaussian in vacuum, i.e. at $\mathbf{P}_0=(-0.100,0.025,-0.015)$ are similar. For the second point $\mathbf{P}_1=(0.000,0.025,-0.015)$, localized in the head, we can observed the dispersion and a slight attenuation of the signal for the dispersive calculation compared to the non-dispersive calculation.

A third simulation is performed with the fully explicit DGTD-$\mathbb{P}_1$ method (6.25). We recall that the critical time step size for the stability of the locally implicit method is then about 15.5 times larger than the critical time step size for the stability of the fully explicit method. On Figures 6.8 we show the contour lines of the local SAR normalized by the maximal local SAR, in logarithmic scale, for the calculations with the locally implicit DGTD-$\mathbb{P}_1$ method (6.27). For convenience of presentation we do not show the results obtained with the fully explicit method because they are indistinguishable from that shown in Figure 6.8. Finally, the total simulation times are equal to 6 h 38 min for the locally implicit DGTD-$\mathbb{P}_1$ method and 24 h 56 min for the fully explicit DGTD-$\mathbb{P}_1$ method. Hence, the locally implicit DGTD-$\mathbb{P}_1$ method allows a reduction of the computing time by a factor of 3.8.

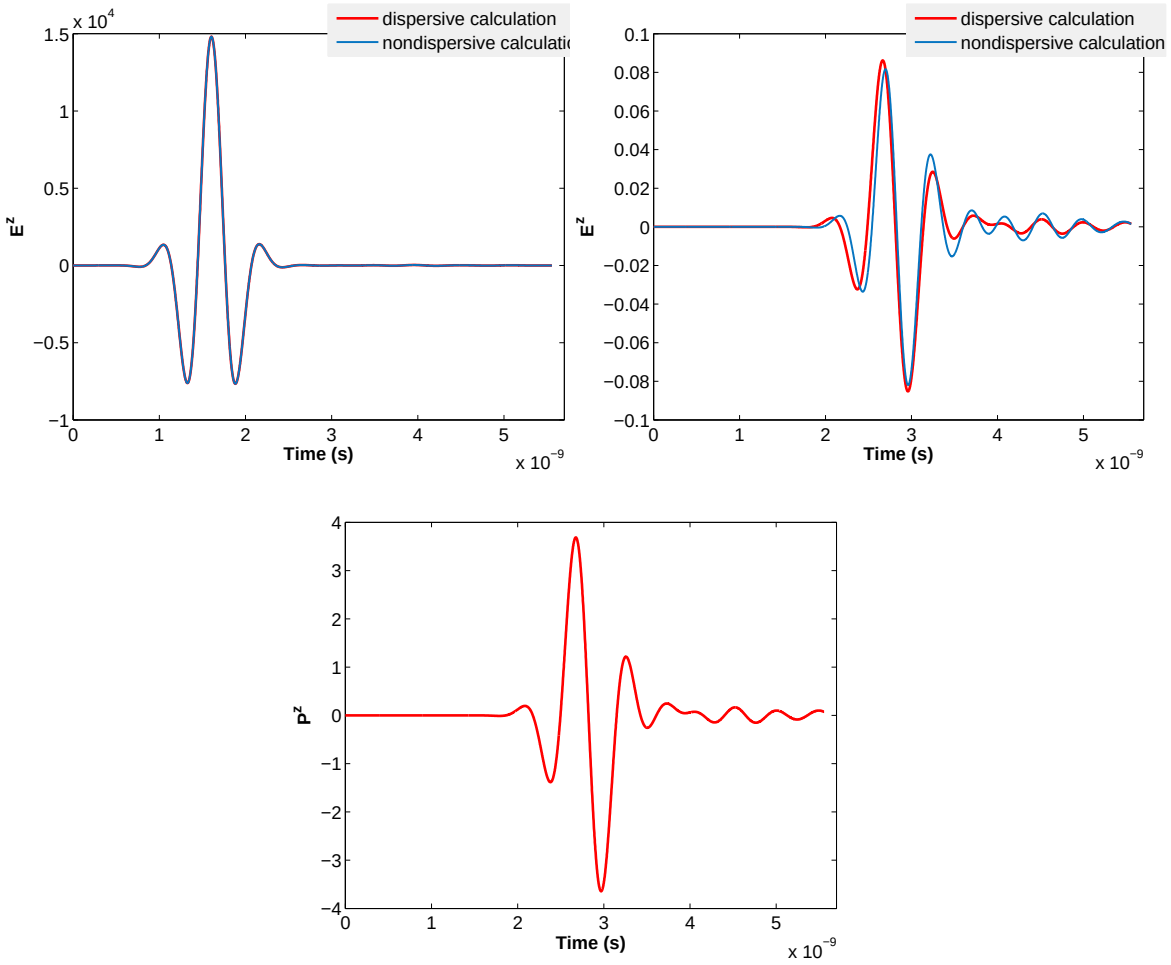


Figure 6.7: Exposure of head tissues to a localized source radiation: time evolution of the E^z component of the electric field at selected spatial locations, $\mathbf{P}_0 = (-0.1962, -0.0027, -0.0032)$ and $\mathbf{P}_1 = (-0.1013, -0.0009, 0.0000)$ (left plot and right plot, respectively) ; time evolution of the P^z component of the polarization field at $\mathbf{P}_1$ (bottom).

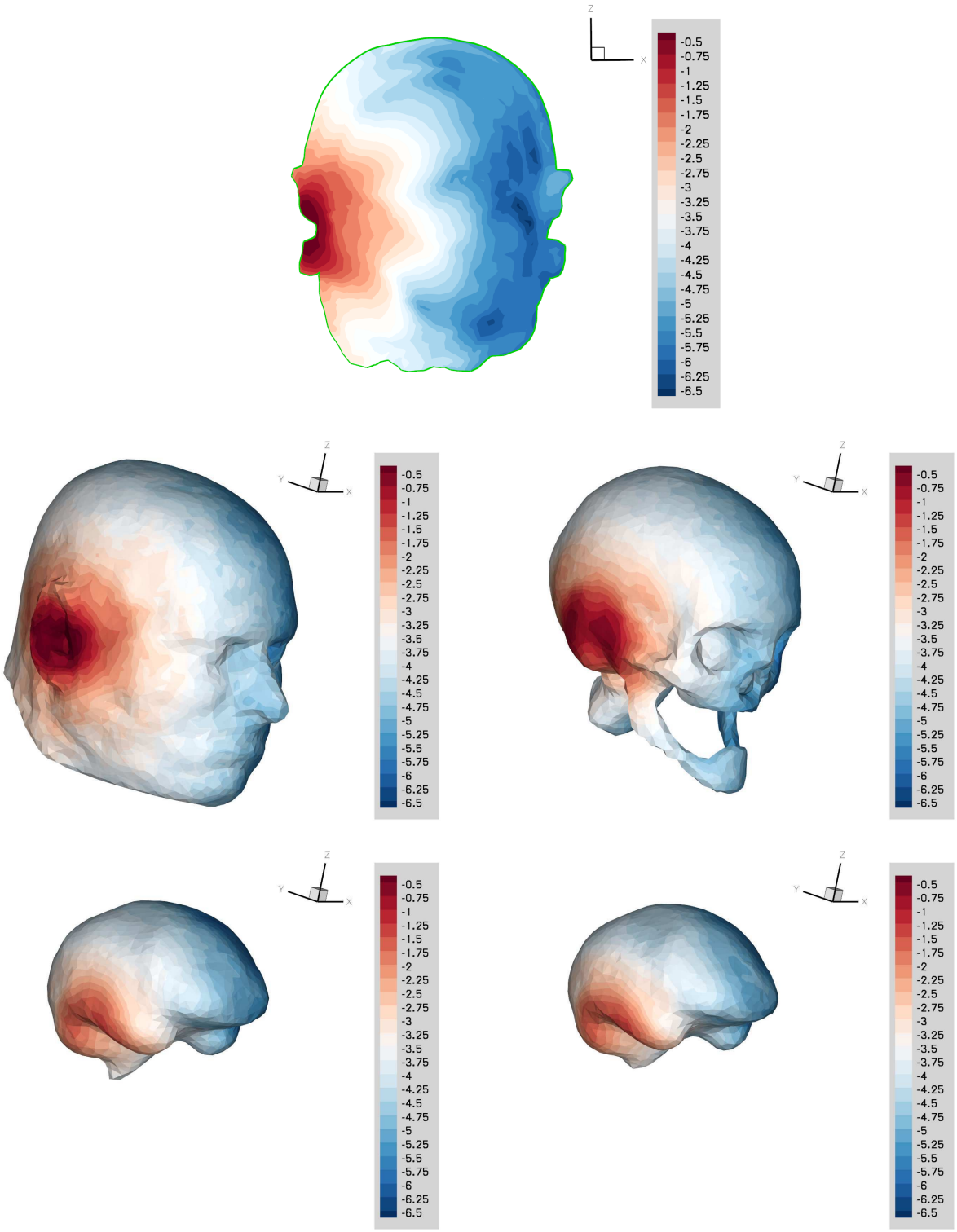


Figure 6.8: Exposure of head tissues to a localized source radiation: contour lines of the local SAR normalized by the maximal local SAR (logarithmic scale).

Conclusion

Rappelons les objectifs de cette thèse qui étaient, d'une part, de proposer et étudier des méthodes d'éléments finis discontinues d'ordre élevé (interpolation polynomiale), reposant sur des triangulations (cas 2D) ou des tétraédrisations (cas 3D) du domaine de calcul et des schémas d'intégration en temps efficaces en présence de maillages localement raffinés, pour la résolution des équations de Maxwell en domaine temporel. D'autre part, d'étendre et adapter ces méthodes à des milieux de propagation complexes, les applications visées étant l'étude des interactions des ondes électromagnétiques avec les tissus biologiques.

Résumé des résultats et contributions

Dans ce but, nous avons tout d'abord considéré deux schémas d'intégration en temps localement implicites, initialement introduits dans [Piperno 2006] et [Verwer 2010], pour la résolution des équations de Maxwell en domaine temporel dans des milieux non-dispersifs. La discrétisation spatiale est issue d'une méthode Galerkine discontinue conduisant à la formulation de deux méthodes GDDT localement implicites. Les résultats connus ont été complétés afin de présenter une étude théorique complète des deux schémas (formulation, stabilité, convergence, coût du traitement implicite). Ainsi, nous avons mené une analyse de convergence de la méthode proposée par Piperno dans laquelle nous avons démontré que le découpage implicite/explicite peut être à l'origine d'une réduction de l'ordre de convergence, lorsque nous considérons un raffinement simultané et stable en espace et en temps (convergence au sens des EDP). Nous avons notamment exhibé une condition suffisante sur la solution exacte du système d'EDP de Maxwell permettant de retrouver un ordre de convergence égal à deux. En conséquence, dans le cas général, nous ne pouvons garantir qu'une convergence du premier ordre pour la méthode introduite dans [Piperno 2006]. Nous avons également conduit une analyse de stabilité du schéma localement implicite proposé dans [Verwer 2010]. Cette étude est basée sur une approche énergétique permettant d'obtenir un critère de stabilité rigoureux et montrant que la taille du pas de temps assurant la stabilité de la méthode est uniquement déterminé par la région du maillage destinée au traitement explicite. Par conséquent, en présence de raffinements locaux, le but des méthodes GDDT proposées est atteint. En effet, les pas de temps les plus contraignants imposés par les petites cellules pour assurer la stabilité des méthodes explicites sont alors remplacés par des pas de temps plus larges dans les méthodes localement implicites, économisant ainsi de nombreux calculs. Par ailleurs, les temps de simulation obtenus pour des problèmes de propagation d'ondes électromagnétiques en 2D sur des maillages non-uniformes, localement raffinés, nous ont permis de montrer que les méthodes localement implicites permettaient d'obtenir des niveaux d'erreurs similaires et étaient plus efficaces que les approches tout explicite ou tout implicite. En ajustant le découpage implicite/explicite nous avons exhibé les grandes similtudes entre les deux approches localement implicites dans le cadre d'une discrétisation en espace par une méthode GD. Bien que les deux méthodes soient très proches, la méthode GDDT localement implicite issue de [Verwer 2010] est plus précise car elle ne souffre pas d'une réduction éventuelle de l'ordre de convergence, en contrepartie elle est aussi plus implicite (la taille du système linéaire à résoudre est plus grande). Nous avons montré que le nombre d'éléments non nuls des matrices

des différents systèmes linéaire associés aux deux méthodes est proche. En comparant les deux approches pour des problèmes 2D, nous avons observé que le coût de résolution du système linéaire par une méthode directe ne se traduit pas par un temps de simulation excessif du schéma localement implicite introduit dans [Verwer 2010]. Par conséquent, pour des problèmes de propagation d'ondes électromagnétiques de complexité comparable, nous recommandons l'utilisation de cette méthode. Dans la suite, nous avons concentré notre analyse sur cette méthode GDDT localement implicite et des résultats prometteurs ont été obtenus pour des problèmes de propagation en 3D sur des maillages non-uniformes, localement raffinés.

Nous avons également considéré des stratégies d'intégration en temps d'ordre plus élevé telles que des compositions symétriques ou l'extrapolation de Richardson, en choisissant pour méthode de base la méthode GDDT localement implicite du second ordre. Le but était de pouvoir exploiter pleinement l'une des propriétés attractives de la méthodes GD, à savoir la possibilité d'augmenter facilement l'ordre de convergence en espace. Nous avons concentré notre analyse sur ces stratégies car elles consistent essentiellement en une combinaison appropriée d'une méthode de base appliquée avec différentes tailles de pas de temps, l'implémentation est ainsi particulièrement facile. Les ordres de convergence prévus par la théorie ont été obtenus pour des problèmes de propagation en 2D. Cette partie de la thèse ne représente qu'une étude préliminaire. La présence d'un terme source ainsi que d'un terme de conduction peut être à l'origine d'une réduction de l'ordre de convergence pour de telles méthodes. Néanmoins, dans le cas où cette réduction est inévitable, la grande précision des méthodes GDDT d'ordre élevé proposées dans cette étude reste une propriété intéressante.

Enfin, nous avons considéré des problèmes de propagation d'ondes électromagnétiques dans des milieux dispersifs, en vue d'applications dans les tissus biologiques. A notre connaissance, dans ce contexte, les méthodes GDDT localement implicites n'ont jamais été exploitées, alors que leurs propriétés sont bien adaptées à la miniaturisation des dispositifs électroniques et des antennes ou encore à la petite taille des cellules cancéreuses, pour lesquelles l'utilisation de maillages localement raffinés est certainement une prérogative importante pour obtenir une solution numérique précise. La modélisation de l'interaction des ondes électromagnétiques avec les tissus biologiques nécessite de résoudre le système d'équations de Maxwell couplé à un modèle approprié de dispersion. Nous avons choisi d'utiliser le modèle de Debye qui est le plus souvent utilisé pour modéliser les interactions des ondes électromagnétiques avec des substances à base d'eau. A l'aide d'une équation différentielle auxiliaire, qui traduit la relation entre la polarisation et le champ électrique, nous avons alors adapté la méthode GDDT localement implicite. Une étude théorique complète de la méthode résultante a été conduite. Une analyse de stabilité basée sur une approche énergétique a permis d'obtenir une condition de stabilité rigoureuse et une étude de convergence a prouvé que la méthode GDDT localement implicite conservait un ordre de convergence égal à deux (au sens des EDP). Ce dernier résultat a été confirmé à l'aide d'un cas test artificiel en 3D, puis nous avons complété cette étude par deux problèmes réalistes de propagation d'ondes électromagnétiques dans un modèle de tête humaine.

Travaux en cours et perspectives

Plusieurs extensions de la méthode GDDT localement implicite proposée dans cette étude peuvent être envisagées afin de définitivement montrer et prouver les propriétés attractives du découpage

implicite/explicite dans le cadre de problèmes réalistes de propagation d'ondes électromagnétiques.

Parmi ses développements, le plus indispensable à court terme est sans doute la parallélisation du solveur Maxwell localement implicite 3D, afin de répondre aux exigences liées à la résolution du système linéaire (en terme de mémoire), et ainsi pouvoir considérer des domaines de calculs plus larges et des ordres d'interpolation plus élevés. Il s'agit alors de distribuer intelligemment les différentes tâches entre les processeurs, d'une part pour le stockage de la matrice, d'autre part pour la résolution elle-même. Ce travail est actuellement en cours en utilisant le solveur parallèle MUMPS (MUltifrontal Massively Parallel sparse direct Solver) [Amestoy *et al.* 2000].

Dans cette thèse nous avons considéré une EDA (Équation Différentielle Auxiliaire) pour modéliser l'évolution temporelle de la polarisation électrique dans un milieu dispersif de type Debye. D'autres modèles de dispersion comme ceux de Lorentz ou de Drude pourraient être considérés. Le modèle de Lorentz est approprié aux matériaux dont le comportement résulte d'une polarisation électronique ou ionique, alors que le modèle de Drude s'applique bien à la plupart des métaux, dont les électrons libres répondent à l'application d'un champs électromagnétique. De nombreuses applications dans des domaines tels que l'optique ou l'informatique sont concernées par ces modèles de dispersion. Comme pour les milieux de Debye, les matériaux de Lorentz ou de Drude se prêtent bien à l'utilisation d'une méthode EDA, grâce à la forme de la susceptibilité électrique en domaine fréquentiel. Similairement à la méthode EDA-GDDT explicite proposée dans [Viquerat *et al.* 2013] pour le modèle de dispersion de Drude, nous pourrions adapter notre méthode GDDT localement implicite à ces différents modèles. En outre, l'utilisation de maillages localement raffinés est adaptée aux matériaux de Drude, puisque le comportement macroscopique des métamatériaux résulte de leur structure nanoscopique.

Rappelons que tout au long de cette thèse nous avons uniquement considéré des conditions de Silver-Müller du premier ordre pour l'approximation des conditions absorbantes. L'introduction de conditions absorbantes d'ordre plus élevé permettrait d'améliorer la précision de la méthode GDDT localement implicite, ainsi que l'ordre de convergence des schémas d'ordre élevé que nous avons considéré au Chapitre 5.

À long terme, l'efficacité globale de la méthode GDDT localement implicite pourrait être également améliorée en adaptant la méthode pour des maillages localement raffinés, non-conformes, à l'image de la méthode explicite présentée dans [Fahs 2009, Fahs & Lanteri 2010]. Dans ce contexte l'utilisation de maillages non-conformes, multi-éléments, qui combine une discrétisation structurée des zones régulières par des héxaèdres orthogonaux avec une discrétisation non structurée par des tétraèdres (cas 3D) des objets de formes irrégulières [Durochat 2013], serait sans doute d'un grand intérêt.

Enfin, il serait intéressant de réaliser une étude comparative entre notre méthode GDDT localement implicite et des méthodes de pas temps locale explicites, dont le but est identique. Les méthodes proposées dans [Grote & Mitkova 2010, Piperno 2006, Taube *et al.* 2009], également formulées pour les équations de Maxwell en domaine temporel et s'appuyant sur une discrétisation en espace GD, représentent de possibles candidates pour une telle étude comparative.

Conclusion

We recall that the main objectives of this study were, on the one hand, to propose arbitrary high-order finite element type methods on simplicial meshes for the discretization of Maxwell's equations and efficient time integration methods for dealing with grid induced stiffness when using non-uniform (locally refined) meshes and, on the other hand, to adapt these methods for complex propagation media models, the intended application being the interaction of electromagnetic waves with biological tissues.

Summary of results and contributions

For this purpose, we first considered two locally implicit time integration approaches from [Piperno 2006] and [Verwer 2010], respectively, for the time-domain Maxwell equations in non-dispersive media, with a DG spatial discretization. To complete the already known theoretical results on both methods, we derived a convergence analysis for the locally implicit method from [Piperno 2006] that showed that the component splitting implicit - explicit can be detrimental to the temporal convergence uniformly in h. We presented a sufficient condition on the exact PDE solution to recover the second-order convergence. Thus, in the general case the convergence order can be reduced by one and we can only guarantee the first-order convergence. We also analyzed the stability of the locally implicit method from [Verwer 2010], via an energy approach which provided a rigorous stability criterion and showed that the stability-based time step is determined by the subset of the mesh elements treated explicitly. Therefore, in the presence of a local refinement, the potential of the two DGTD locally implicit methods is achieved, i.e., the severe stability constraint characterizing globally explicit time integration schemes is overcome. Numerical results for two-dimensional electromagnetic waves propagation problems in the presence of local refinements showed that the two locally implicit schemes are computationally more efficient than either the fully explicit or the fully implicit approach. By adjusting the component splitting, we highlighted the similarities between the two locally implicit time integration approaches for the time-dependent Maxwell equations with our DG spatial discretization. We observed that the two schemes are closely related but while the DGTD method adapted from [Verwer 2010] is more accurate because, unlike the method from [Piperno 2006], it does not suffer from order reduction, it is also more implicit, regarding the dimension and the sparsity of the matrix of the linear system to be solved. However, we showed that the number of nonzero elements of the matrices for the two methods is close. By comparing the two approaches for two-dimensional problems, we observed that the overhead for solving the linear system by a direct method using the LU decomposition does not translate into an excessive final simulation time for the locally implicit method from [Verwer 2010] compared to the DGTD method from [Piperno 2006]. Consequently, for electromagnetic waves propagation problems with similar computational complexity, we can advise the use of the method from [Verwer 2010]. In the sequel, we focused our analysis on this locally implicit DGTD method from [Verwer 2010]. Promising results were obtained for three-dimensional electromagnetic wave propagation problems on locally refined meshes.

Next we considered higher order time integration techniques based on the second-order locally implicit DGTD method, such as symmetric composition methods or Richardson extrapolation, to fully exploit an attractive feature of DG methods which is the possibility to easily increase the spatial convergence order. Such strategies to increase the order of convergence in time are well-known for numerical ODE problems but not widespread for numerical PDE problems. We focused on such techniques because these strategies are essentially a suitable combination of the basic symmetric second-order method applied with different step sizes and therefore their computer implementations are particularly easy. We obtained the expected convergence orders for two-dimensional electromagnetic waves propagation problems, however this part remains a preliminary investigation. The source term and the presence of a damping term which models conduction may be the cause of a reduction order, nevertheless, even if a reduction order is an issue, the accuracy of the high-order locally implicit DGTD methods proposed in this study with well-designed coefficients remains an attractive feature.

Finally, we considered the propagation of electromagnetic waves in dispersive media. As previously mentioned the intended applications are the interaction of electromagnetic waves with biological tissues, that are applications of societal relevance such as the assessment of potential adverse effects of electromagnetic fields or the utilization of electromagnetic waves for therapeutic or diagnostic purposes. To our knowledge, locally implicit DGTD methods were never considered, although the characteristics of these methods are well suited to the miniaturization of electronic devices and antennas or the small size of cancerous cells, for which the use of locally refined meshes are certainly key for the efficient numerical solution of electromagnetic wave propagation problems in biological tissues. In this context, we considered the Debye model which is most often used to model electromagnetic wave interactions with water-based substances, such as biological materials. The dispersive character was taken into account via an auxiliary differential equation (ADE), which relates the electric polarization to the electric field, and we adapted the locally implicit DGTD method to the resulting mathematical model. We presented a full theoretical study of the resulting locally implicit DGTD scheme. A stability analysis via an energy approach was realized, providing a rigorous stability criterion. We also derived a convergence analysis to prove that the locally implicit DGTD method for Maxwell's equations in dispersive media retains its second-order convergence. This latter result was confirmed with an artificial three-dimensional problem in a dispersive medium. Finally, we completed this study by applying the locally implicit DGTD method to a realistic wave propagation problem: microwave propagation in head tissues.

Work in progress and prospects

Several extension paths of the proposed locally implicit DGTD method for solving the time-domain Maxwell equations in non-dispersive and dispersive media can be considered to definitely assess the practical features of component splitting based approaches for realistic electromagnetic wave propagation problems.

Among all the developments that can be considered, a mandatory short-term topic is probably the parallelization of the three-dimensional locally implicit Maxwell solver. This is a mandatory issue for complex three-dimensional problems, imposed by the memory requirement for storing the matrix of the linear system to be solved, in order to consider computational domains with a great

number of elements and to use higher interpolation degrees in the DG method. A key requirement is then an appropriate distribution of the work tasks among the processors, on the one hand for the storage of the matrix, on the other hand for the resolution itself. This work is currently in progress using the MUMPS (MUltifrontal Massively Parallel sparse direct Solver) optimized sparse direct solver [Amestoy *et al.* 2000].

In this dissertation we considered an approach based on an ADE modeling the time evolution of the electric polarization for a dispersive medium of Debye type. Other dispersive media such as the so-called Lorentz and Drude materials can be considered. Lorentz materials are those whose behavior arises from electronic or ionic polarization, where the electrons or ions are displaced from the molecules to which they are bounded. The Drude model of dispersion applies very well to most metals, whose free electrons respond to applied electromagnetic fields. Many applications in optics, informatics, etc., are concerned by these models. As Debye materials, Lorentz and Drude materials lend themselves well to the use of the ADE method, thanks to the form of the frequency-domain susceptibility. As in [Viquerat *et al.* 2013], where an explicit ADE-DGTD method is designed for Drude materials, we will adapt our locally implicit DGTD method for Lorentz and Drude materials subsequently. Furthermore, the use of locally refined meshes are certainly well-suited to Drude materials, since the macroscopic behavior of metamaterials results from their nanoscopic structure, which induces particular interactions with propagating waves.

Recall that throughout this dissertation we considered the first-order Silver-Müller condition to approximate the absorbing boundary condition. The introduction of higher order absorbing boundary conditions would probably allow to improve the accuracy and the order of convergence of the locally implicit DGTD method, particularly for the higher order locally implicit time integration strategies.

In the longer-term, the overall efficiency of the locally implicit method could be improved by adapting it for locally refined non-conforming meshes, as in [Fahs 2009, Fahs & Lanteri 2010]. In this context the use of a non-conforming multi-element mesh (hexahedral/tetrahedral three-dimensional mesh) which combines a structured (orthogonal) discretization of the regular zones of the computational domain with an unstructured discretization of the irregularly shaped objects, as in [Durochat 2013], will be no doubt of interest.

Finally, it would be of great interest to conduct a numerical comparative study between our locally implicit DGTD method and explicit local time-stepping methods, which also allow to overcome step size limitation induced by local mesh refinements. For instance, the methods from [Grote & Mitkova 2010, Piperno 2006, Taube *et al.* 2009] which are also designed for the time-domain Maxwell equations with a spatial DG discretization, are possible candidates.

APPENDIX A

Nodes of the degrees of freedom and basis functions

Contents

A.1 Basis functions for the reference triangle

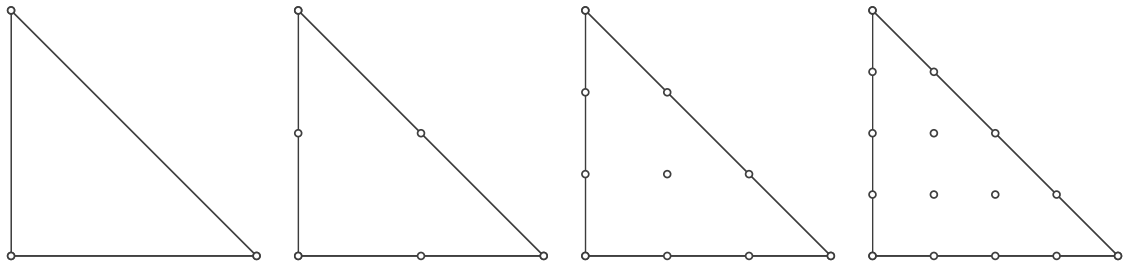


Figure A.1: Lattice of order $1, 2, 3$ and 4 for the reference triangle (the circles represent the points of the lattice), $\tau_0 = \{\mathbf{x} \in \mathbb{R}^2 \text{ such that } x+y \leq 1 \text{ and } x, y \geq 0\}$.

A.1.1 $\mathbb{P}_1$ interpolation

For a $\mathbb{P}_1$ interpolation the coordinates of nodes of the degrees of freedom are $(0,0)$, $(1,0)$ and $(0,1)$, and the basis functions for the reference triangle, τ_0, are given by

$$\begin{cases} \varphi_{01}(\mathbf{x}) &= 1-x-y, \\ \varphi_{02}(\mathbf{x}) &= x, \\ \varphi_{03}(\mathbf{x}) &= y. \end{cases} \tag{A.1}$$

A.1.2 $\mathbb{P}_2$ interpolation

For a $\mathbb{P}_2$ interpolation the coordinates of nodes of the degrees of freedom are $(0,0)$, $(1,0)$, $(0,1)$, $(1/2,0)$, $(1/2,1/2)$ and $(0,1/2)$ and the basis functions for the reference triangle, τ_0, are given by

$$\begin{cases} \varphi_{01}(\mathbf{x}) &= (1-x-y)(1-2x-2y), \\ \varphi_{02}(\mathbf{x}) &= x(2x-1), \\ \varphi_{03}(\mathbf{x}) &= y(2y-1), \\ \varphi_{04}(\mathbf{x}) &= 4x(1-x-y), \\ \varphi_{05}(\mathbf{x}) &= 4xy, \\ \varphi_{06}(\mathbf{x}) &= 4y(1-x-y). \end{cases} \tag{A.2}$$

A.1.3 $\mathbb{P}_3$ interpolation

For a $\mathbb{P}_3$ interpolation the coordinates of nodes of the degrees of freedom are $(0,0)$, $(1,0)$, $(0,1)$, $(1/3,0)$, $(2/3,0)$, $(2/3,1/3)$, $(1/3,2/3)$, $(0,2/3)$, $(0,1/3)$ and $(1/3,1/3)$ and the basis functions for the reference triangle, τ_0, are given by

$$\begin{cases} \varphi_{01}(\mathbf{x}) &= \frac{1}{2}(1-x-y)(2-3x-3y)(1-3x-3y), \\ \varphi_{02}(\mathbf{x}) &= \frac{1}{2}x(3x-1)(3x-2), \\ \varphi_{03}(\mathbf{x}) &= \frac{1}{2}y(3y-1)(3y-2), \\ \varphi_{04}(\mathbf{x}) &= \frac{9}{2}x(1-x-y)(2-3x-3y), \\ \varphi_{05}(\mathbf{x}) &= \frac{9}{2}x(1-x-y)(3x-1), \\ \varphi_{06}(\mathbf{x}) &= \frac{9}{2}xy(3x-1), \\ \varphi_{07}(\mathbf{x}) &= \frac{9}{2}xy(3y-1), \\ \varphi_{08}(\mathbf{x}) &= \frac{9}{2}y(1-x-y)(3y-1), \\ \varphi_{09}(\mathbf{x}) &= \frac{9}{2}y(1-x-y)(2-3x-3y), \\ \varphi_{010}(\mathbf{x}) &= 27xy(1-x-y). \end{cases} \tag{A.3}$$

A.1.4 $\mathbb{P}_4$ interpolation

For a $\mathbb{P}_4$ interpolation the coordinates of nodes of the degrees of freedom are $(0,0)$, $(1,0)$, $(0,1)$, $(1/4,0)$, $(1/2,0)$, $(3/4,0)$, $(3/4,1/4)$, $(1/2,1/2)$, $(1/4,3/4)$, $(0,3/4)$, $(0,1/2)$, $(0,1/4)$, $(1/4,1/4)$, $(1/2,1/4)$ and $(1/4,1/2)$ and the basis functions for the reference triangle, τ_0, are given by

$$
\begin{cases}
\varphi_{01}(\mathbf{x}) &= \frac{1}{3}(x+y-1)(2x+2y-1)(4x+4y-1)(4x+4y-3), \\
\varphi_{02}(\mathbf{x}) &= \frac{1}{3}x(2x-1)(4x-1)(4x-3), \\
\varphi_{03}(\mathbf{x}) &= \frac{1}{3}y(2y-1)(4y-1)(4y-3), \\
\varphi_{04}(\mathbf{x}) &= \frac{16}{3}x(1-x-y)(2x+2y-1)(4x+4y-1), \\
\varphi_{05}(\mathbf{x}) &= 16x(x+y-1)(4x-1)(4x+4y-3), \\
\varphi_{06}(\mathbf{x}) &= \frac{16}{3}x(1-x-y)(2x-1)(4x-1), \\
\varphi_{07}(\mathbf{x}) &= \frac{16}{3}xy(2x-1)(4x-1), \\
\varphi_{08}(\mathbf{x}) &= 4xy(4x-1)(4y-1), \\
\varphi_{09}(\mathbf{x}) &= \frac{16}{3}xy(2y-1)(4y-1), \\
\varphi_{010}(\mathbf{x}) &= \frac{16}{3}y(1-x-y)(2y-1)(4y-1), \\
\varphi_{011}(\mathbf{x}) &= 4y(x+y-1)(4y-1)(4x+4y-3), \\
\varphi_{012}(\mathbf{x}) &= \frac{16}{3}y(1-x-y)(2x+2y-1)(4x+4y-1), \\
\varphi_{013}(\mathbf{x}) &= 32xy(x+y-1)(4x+4y-3), \\
\varphi_{014}(\mathbf{x}) &= 32xy(1-x-y)(4x-1), \\
\varphi_{015}(\mathbf{x}) &= 32xy(1-x-y)(4y-1).
\end{cases}
\tag{A.4}
$$

A.2 Basis functions for the reference tetrahedron

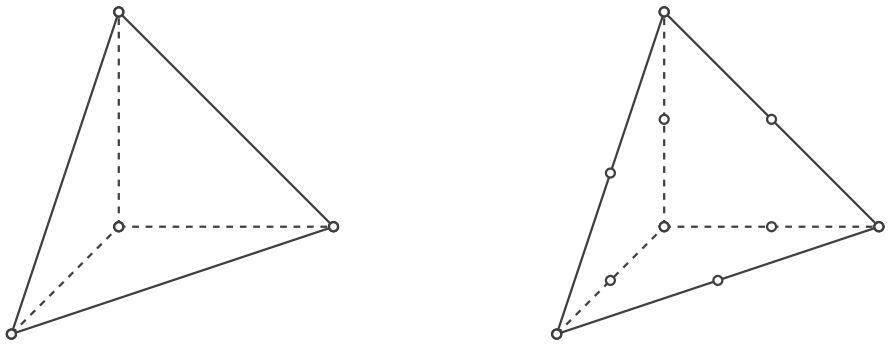


Figure A.2: Lattice of order 1 and 2 for the reference tetrahedron (the circles represent the points of the lattice), $\tau_0 = \left\{\mathbf{x} \in \mathbb{R}^3 \text{ such that } x+y+z \leq 1 \text{ and } x,y,z \geq 0\right\}$.

A.2.1 $\mathbb{P}_1$ interpolation

For a $\mathbb{P}_1$ interpolation the coordinates of nodes of the degrees of freedom are $(0,0,0)$, $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$, and the basis functions for the reference triangle, τ_0, are given by

$$\begin{cases} \varphi_{01}(\mathbf{x}) &= 1-x-y-z, \\ \varphi_{02}(\mathbf{x}) &= x, \\ \varphi_{03}(\mathbf{x}) &= y, \\ \varphi_{04}(\mathbf{x}) &= z. \end{cases} \tag{A.5}$$

A.2.2 $\mathbb{P}_2$ interpolation

For a $\mathbb{P}_2$ interpolation the coordinates of nodes of the degrees of freedom are $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, $(1/2,0,0)$, $(1/2,1/2,0)$, $(0,1/2,0)$, $(1/2,0,1/2)$, $(0,0,1/2)$ and $(0,1/2,1/2)$, and the basis functions for the reference triangle, τ_0, are given by

$$\begin{cases} \varphi_{01}(\mathbf{x}) &= (x+y+z-1)(2x+2y+2z-1), \\ \varphi_{02}(\mathbf{x}) &= x(2x-1), \\ \varphi_{03}(\mathbf{x}) &= y(2y-1), \\ \varphi_{04}(\mathbf{x}) &= z(2z-1), \\ \varphi_{05}(\mathbf{x}) &= -4x(x+y+z-1), \\ \varphi_{06}(\mathbf{x}) &= 4xy, \\ \varphi_{07}(\mathbf{x}) &= -4y(x+y+z-1), \\ \varphi_{08}(\mathbf{x}) &= 4xz, \\ \varphi_{09}(\mathbf{x}) &= -4z(x+y+z-1), \\ \varphi_{010}(\mathbf{x}) &= 4yz. \end{cases} \tag{A.6}$$

List of Tables

List of Figures

Bibliography

[Amestoy *et al.* 2000] P.R. Amestoy, I.S. Duff and J.-Y. L'Excellent. *Multifrontal parallel distributed symmetric and unsymmetric solvers*. Comput. Meth. Appl. Mech. Eng., vol. 184, no. 2-4, pages 501–520, 2000. (Cited on pages 59, 127, 143 and 147.)

[Ascher *et al.* 1995] U.M. Ascher, S.J. Ruuth and B. Wetton. *Implicit-Explicit Methods for Time-Dependent Partial Differential Equations*. SIAM J. Numer. Anal., vol. 32, no. 3, pages 797–823, 1995. (Cited on pages 3, 9 and 39.)

[Ascher *et al.* 1997] U.M. Ascher, S.J. Ruuth and R.J. Spiteri. *Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations*. Appl. Numer. Math., vol. 25, no. 2, pages 151–167, 1997. (Cited on page 39.)

[Botchev & Verwer 2009] M.A. Botchev and J.G. Verwer. *Numerical Integration of Damped Maxwell Equations*. SIAM J. SCI. Comput., vol. 31, no. 2, pages 1322–1346, 2009. (Cited on pages 39, 40, 55, 95, 96, 101 and 106.)

[Buffa & Perugia 2006] A. Buffa and I. Perugia. *Discontinuous Galerkin approximation of the Maxwell eigenproblem*. SIAM J. Numer. Anal., vol. 44, no. 5, pages 2198–2226, 2006. (Cited on page 38.)

[Cai & Deng 2003] W. Cai and S. Deng. *An upwinding embedded boundary method for Maxwell's equations in media with material interfaces: 2D case*. J. Comput. Phys., vol. 190, no. 1, pages 159–183, 2003. (Cited on page 76.)

[Calvo *et al.* 2001] M.P. Calvo, J. de Frutos and J. Novo. *Linearly implicit Runge-Kutta methods for advectionreactiondiffusion equations*. Appl. Numer. Math., vol. 37, no. 4, pages 535–549, 2001. (Cited on page 39.)

[Catella *et al.* 2010] A. Catella, V. Dolean and S. Lanteri. *An implicit discontinuous Galerkin time-domain method for two-dimensional electromagnetic wave propagation*. COMPEL, vol. 29, no. 3, pages 602–625, 2010. (Cited on page 37.)

[Cockburn & Shu 2005] B. Cockburn and C.W. Shu. *Special issue on discontinuous Galerkin methods*. J. Sci. Comput., Springer, vol. 22-23, 2005. (Cited on page 28.)

[Cockburn *et al.* 2000] B. Cockburn, G.E. Karniadakis and C.W. Shu. Discontinuous Galerkin Methods. Theory, Computation and Applications, volume 11 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2000. (Cited on pages 3, 8, 23, 28 and 62.)

[Converse *et al.* 2004] M. Converse, E. Bond, S. Hagness and B.V. Veen. *Ultrawide-band microwave space-time beamforming for hyperthermia treatment of breast cancer: a computational feasibility study*. IEEE Trans. Microwave Theory Tech., vol. 52, no. 8, pages 1876–1889, 2004. (Cited on page 109.)

[Creutz & Gocksch 1989] M. Creutz and A. Gocksch. *Higher-order hybrid Monte Carlo algorithms*. Phys. Rev. Lett., vol. 63, no. 1, pages 9–12, 1989. (Cited on page 100.)

[Crouzeix 1980] M. Crouzeix. *Une méthode multipas implicite-explicite pour l'approximation des équations d'évolution paraboliques*. Numer. Math., vol. 35, no. 3, pages 257–276, 1980. (Cited on page 38.)

[Dawson 2006] C. Dawson. *Special issue on discontinuous Galerkin methods*. Comput. Meth. App. Mech. Engng., Elsevier, vol. 195, 2006. (Cited on page 28.)

[Descombes *et al.* 2013] S. Descombes, S. Lanteri and L. Moya. *Locally implicit time integration strategies in a discontinuous Galerkin method for Maxwell's equations*. J. SCI. Comput., vol. 56, no. 1, pages 190–218, 2013. (Cited on page 39.)

[Diaz & Grote 2009] J. Diaz and M.J. Grote. *Energy conserving explicit local time-stepping for second-order wave equations*. SIAM J. SCI. Comput., vol. 31, no. 3, pages 1985–2014, 2009. (Cited on pages 4, 9 and 38.)

[Dolean *et al.* 2010] V. Dolean, H. Fahs, L. Fezoui and S. Lanteri. *Locally implicit discontinuous Galerkin method for time domain electromagnetics*. J. Comput. Phys., vol. 229, no. 2, pages 512–526, 2010. (Cited on pages 4, 10, 43 and 62.)

[Durochat 2013] C. Durochat. *Méthode de type Galerkin discontinu en maillages multi-éléments (et non-conformes) pour la résolution numérique des équations de Maxwell instationnaires*. PhD thesis, Université de Nice - Sophia Antipolis, 2013. (Cited on pages 143 and 147.)

[Fahs & Lanteri 2010] H. Fahs and S. Lanteri. *A high-order non-conforming discontinuous Galerkin method for time-domain electromagnetics*. J. Comput. Appl. Math., vol. 234, no. 4, pages 1088–1096, 2010. (Cited on pages 143 and 147.)

[Fahs 2009] H. Fahs. *Development of a hp-like discontinuous Galerkin time-domain method on non-conforming simplicial meshes for electromagnetic wave propagation*. Int. J. Numer. Anal. Mod., vol. 6, no. 2, pages 193–216, 2009. (Cited on pages 76, 143 and 147.)

[Faragó *et al.* 2010] I. Faragó, Á. Havasi and Z. Zlatev. *Efficient implementation of stable Richardson extrapolation algorithms*. Comput. Math. Appl., vol. 60, no. 8, pages 2309–2325, 2010. (Cited on pages 95, 103 and 105.)

[Fear *et al.* 2003] E.C. Fear, P.M. Meaney and M. A. Stuchly. *Microwaves for breast cancer detection*. IEEE Potentials, vol. 22, no. 1, pages 12–18, 2003. (Cited on page 109.)

[Fezoui *et al.* 2005] L. Fezoui, S. Lanteri, S. Lohrengel and S. Piperno. *Convergence and stability of a discontinuous Galerkin time-domain method for the 3D heterogeneous Maxwell equations on unstructured meshes*. ESAIM: M2AN, vol. 39, no. 6, pages 1149–1176, 2005. (Cited on pages 22, 37, 60 and 62.)

[Forest 1989] E. Forest. *Canonical integrators as tracking codes*. In AIP Conference Proceedings, volume 184, pages 1106–1136, 1989. (Cited on page 100.)

[Frank *et al.* 1997] J. Frank, W. Hundsdorfer and J.G. Verwer. *On the stability of implicit-explicit linear multistep methods*. Appl. Numer. Math., vol. 25, no. 2, pages 193–205, 1997. (Cited on page 38.)

[Fumeaux *et al.* 2006] C. Fumeaux, D. Baumann, P. Bonnet and R. Vahldieck. *Developments of finite-volume techniques for electromagnetic modeling in unstructured meshes*. In Dig. 17th Int. Symp. Electromagnetic Compatibility (EMC 2006), pages 5–8, Zurich, Switzerland, 2006. (Cited on page 27.)

[Grote & Mitkova 2010] M.J. Grote and T. Mitkova. *Explicit local time stepping methods for Maxwell's equations*. J. Comput. Appl. Math., vol. 234, no. 12, pages 3283–3302, 2010. (Cited on pages 4, 9, 38, 143 and 147.)

[Grote *et al.* 2007] M.J. Grote, A. Schneebeli and D. Schotzau. *Interior penalty discontinuous Galerkin method for Maxwell's equations: Energy norm error estimates*. J. Comput. Appl. Math., vol. 204, no. 2, pages 375–384, 2007. (Cited on page 38.)

[Grote *et al.* 2008] M.J. Grote, A. Schneebeli and D. Schotzau. *Interior penalty discontinuous Galerkin method for Maxwell's equations: Optimal L^2-norm error estimates*. IMA J. Numer. Anal., vol. 28, no. 2, pages 440–468, 2008. (Cited on page 38.)

[Hairer & Wanner 1996] E. Hairer and G. Wanner. Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second édition, 1996. (Cited on pages 95 and 99.)

[Hairer *et al.* 1993] E. Hairer, S.P. Nørsett and G. Wanner. Solving Ordinary Differential Equations I. Nonstiff Problems, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second édition, 1993. (Cited on pages 5, 10, 95 and 103.)

[Hairer *et al.* 2010] E. Hairer, C. Lubich and G. Wanner. Geometrical Numerical Integration : Structure-Preserving Algorithms for Ordinary Differential Equations, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, New York, second édition, 2010. (Cited on pages 5, 10, 95, 96, 99 and 101.)

[Hesthaven & Warburton 2008] J.S. Hesthaven and T. Warburton. Nodal discontinuous galerkin methods: Algorithms, analysis, and applications, volume 54 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 2008. (Cited on pages 3, 8, 23 and 28.)

[Huang & Li 2009] Y. Huang and J. Li. *Interior penalty discontinuous Galerkin method for Maxwell's equations in cold plasma*. J. Sci. Comput., vol. 41, no. 3, pages 321–340, 2009. (Cited on page 110.)

[Huang *et al.* 2011] Y. Huang, J. Li and W. Yang. *Interior penalty discontinuous Galerkin methods for Maxwell's equations in dispersive media*. J. Comput. Phys., vol. 230, no. 12, pages 4559–4570, 2011. (Cited on page 110.)

[Hundsdorfer & Verwer 2003] W. Hundsdorfer and J.G. Verwer. Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations, volume 33 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2003. (Cited on pages 3, 9, 38, 48, 50, 124 and 125.)

[Inan & Marshall 2011] U.S. Inan and R.A. Marshall. Numerical Electromagnetics: The FDTD Method. Cambridge University Press, California, 2011. (Cited on pages 3, 8 and 110.)

[Jiao & Jin 2001] D. Jiao and J.-M. Jin. *Time-domain finite-element modeling of dispersive media*. IEEE Microw. Wirel. Compon. Lett., vol. 11, pages 220–223, 2001. (Cited on page 110.)

[Jin & Riley 2008] J.-M Jin and D.J. Riley. Finite Element Analysis of Antennas and Arrays. Wiley, Hoboken, NJ, USA, 2008. (Cited on pages 3 and 8.)

[Joseph *et al.* 1991] R.M. Joseph, S.C. Hagness and A. Taflove. *Direct time integration of Maxwell's equations in linear dispersive media wi th absorption for scattering and propagation of femtosecond electromagnetic pulses*. Optics Lett., vol. 16, no. 18, pages 1412–1414, 1991. (Cited on page 111.)

[Kanvesky *et al.* 2007] A. Kanvesky, M.H. Carpenter, D. Gottlieb and J.S. Hesthaven. *Application of implicit-explicit high order Runge-Kutta methods to discontinuous-Galerkin schemes*. J. Comput. Phys., vol. 225, no. 2, pages 1753–1781, 2007. (Cited on page 39.)

[Karanasiou *et al.* 2008] I.S. Karanasiou, K.T. Karathanasis, A. Garetos and N.K. Uzunoglu. *Development and laboratory testing of a noninvasive intracranial focused hyperthermia system*. IEEE Trans. Microwave Theory Tech., vol. 56, no. 9, pages 2160–2171, 2008. (Cited on page 109.)

[Kashiwa & Fukai 1990] T. Kashiwa and I. Fukai. *A treatment by FDTD method of dispersive characteristics associated with electronic polarization*. Microwave Optics Tech. Lett., vol. 3, no. 6, pages 203–205, 1990. (Cited on page 111.)

[Kelley & Luebbers 1996] D.F. Kelley and R.J. Luebbers. *Piecewise linear recursive convolution for dispersive media using FDTD*. IEEE Trans. Antennas Propagat., vol. 44, no. 6, pages 792–797, 1996. (Cited on page 111.)

[Kennedy & Carpenter 2003] C.A. Kennedy and M.H. Carpenter. *Additive Runge-Kutta schemes for convection-diffusion-reaction equations*. Appl. Numer. Math., vol. 44, no. 1-2, pages 139–181, 2003. (Cited on page 39.)

[Klemm *et al.* 2010] M. Klemm, J.A. Leendertz, D.R. Gibbins, I.J. Craddock, A.W. Preece and R. Benjamin. *Towards contrast enhanced breast imaging using ultra-wideband microwave radar system*. In IEEE Radio and Wireless Symposium. Special session on biomedical applications, pages 516–519, New Orleans, USA, 2010. (Cited on page 109.)

[Lanteri & Scheid 2013] S. Lanteri and C. Scheid. *Convergence of a Discontinuous Galerkin scheme for the mixed time domain Maxwell's equations in dispersive media*. IMA J. Numer. Anal., vol. 33, no. 2, pages 432–459, 2013. (Cited on pages 110, 113 and 128.)

[LeSaint & Raviart 1974] P. LeSaint and P.A. Raviart. *On a Finite Element Method for Solving the Neutron Transport Equation*. In Mathematical Aspects of Finite Elements in Partial Differential Equations, pages 89–145. Academic Press, New York, 1974. (Cited on page 27.)

[Li 2011] J. Li. *Unified analysis of leap-frog methods for solving time-domain Maxwells equations in dispersive media*. J. Sci. Comput., vol. 47, no. 1, pages 1–26, 2011. (Cited on page 113.)

[Lu *et al.* 2004] T. Lu, P. Zhang and W. Cai. *Discontinuous Galerkin methods for dispersive and lossy Maxwell's equations and PML boundary conditions*. IEEE Trans. Geosc. Rem. Sens., vol. 200, no. 2, pages 549–580, 2004. (Cited on page 110.)

[Lu *et al.* 2005] T. Lu, W. Cai and P. Zhang. *Discontinuous Galerkin time domain methods for GPR simulation in dispersive media*. IEEE Trans. Geosc. Rem. Sens., vol. 43, no. 1, pages 72–80, 2005. (Cited on page 110.)

[Luebbers & Hunsberger 1992] R.J. Luebbers and F. Hunsberger. *FDTD for Nth-order dispersive media*. IEEE Trans. Antennas Propagat., vol. 40, no. 11, pages 1297–1301, 1992. (Cited on page 111.)

[Luebbers *et al.* 1990] R.J. Luebbers, F. Hunsberger, K.S. Kunz, R.B. Standler and M. Schneider. *A frequency-dependent finite-difference time-domain formulation formulation dispersive materials*. IEEE Trans. on Electr. Comp., vol. 32, no. 3, pages 222–227, 1990. (Cited on page 111.)

[Maxwell 1865] J.C. Maxwell. *A dynamical theory of the electromagnetic field*. Philosophical Transactions of the Royal Society of London, vol. 155, pages 459–513, 1865. (Cited on page 13.)

[McLachlan 1995] R.I. McLachlan. *On the numerical integration of ordinary differential equations by symmetric composition methods*. SIAM J. Sci. Comput., vol. 16, no. 1, pages 151–168, 1995. (Cited on pages 95 and 99.)

[Miklavčič *et al.* 2000] D. Miklavčič, D. Šermov, H. Mekid and L.M. Mir. *A validated model of in vivo electric field distribution in tissues for electrochemotherapy and for DNA electrotransfer for gene therapy*. Biochimica et Biophysica Acta, vol. 1523, pages 73–83, 2000. (Cited on pages 2, 7 and 110.)

[Monseny *et al.* 2008] E. Monseny, S. Pernet, X. Ferrières and G. Cohen. *Dissipative terms and local time-stepping improvements in a spatial high order Discontinuous Galerkin scheme for the time-domain Maxwell's equations*. J. Comput. Phys., vol. 227, no. 14, pages 6795–6820, 2008. (Cited on pages 4, 9 and 38.)

[Moya 2012] L. Moya. *Temporal convergence of a locally implicit discontinuous Galerkin method for Maxwell's equations*. ESAIM: M2AN, vol. 46, no. 5, pages 1225–1246, 2012. (Cited on pages 39, 40 and 55.)

[Murua1 & SanzSerna 1999] A. Murua1 and J.M. SanzSerna. *Order conditions for numerical integrators obtained by composing simpler integrators*. Phil. Trans. R. Soc. Lond. A, vol. 357, no. 1754, pages 1079–1100, 1999. (Cited on page 101.)

[Piperno 2006] S. Piperno. *Symplectic local time-stepping in non-dissipative DGTD methods applied to wave propagation problem*. ESAIM: M2AN, vol. 40, no. 5, pages 815–841, 2006. (Cited on pages i, 4, 9, 37, 39, 40, 41, 42, 43, 45, 47, 49, 51, 141, 143, 145 and 147.)

[Rao 1999] S.M. Rao. Time-Domain Electromagnetics. Academic Press Series in Engineering. Academic Press, first édition, 1999. (Cited on pages 3 and 8.)

[Reed & Hill 1973] W.H. Reed and T.R. Hill. *Triangular mesh methods for the neutron transport equation*. Los Alamos Scientific Laboratory Report LAUR-73-479, 1973. (Cited on page 27.)

[Remaki 2000] M. Remaki. *A new finite volume scheme for solving Maxwell's system*. COMPEL, vol. 19, no. 3, pages 913–931, 2000. (Cited on page 27.)

[Richardson 1910] L.F. Richardson. *The Approximate Arithmetical Solution by Finite Differences of Physical Problems Involving Differential Equations, with an Application to the Stresses in a Masonry Dam*. Phil. Trans. R. Soc. Lond. A, vol. 210, pages 307–357, 1910. (Cited on pages 95 and 103.)

[Richardson 1927] L.F. Richardson. *The deferred approach to the limit*. Phil. Trans. R. Soc. Lond. A, vol. 226, no. 1754, pages 299–349, 1927. (Cited on pages 95 and 103.)

[Sármány *et al.* 2007] D. Sármány, M.A. Botchev and J.J.W. Van der Vegt. *Dispersion and dissipation error in high-order Runge-Kutta discontinuous Galerkin discretisations of the Maxwell equations*. J. Sci. Comput., vol. 33, no. 1, pages 47–74, 2007. (Cited on page 62.)

[Schea *et al.* 2010] J.D. Schea, P. Kosmas, B.D. Van Veen and S.C. Hagness. *Contrast-enhanced microwave imaging of breast tumors: a computational study using 3D realistic numerical phantoms*. Inverse Problems, vol. 26, no. 7, 2010. (Cited on page 109.)

[Serša 2005] G. Serša. *Application of electroporation in electrochemotherapy of tumors*. In Electroporation based Technologies and Treatment. International scientific workshop and postgraduate course, pages 42–45, Ljubljana, Slovenia, 2005. (Cited on pages 2, 7 and 110.)

[Shankar *et al.* 1990] V. Shankar, A.H. Mohammadian and W.F. Hall. *A time-domain, finite-volume treatment for the Maxwell equations*. Electromagnetics, vol. 10, no. 1, pages 127–145, 1990. (Cited on page 27.)

[Sukharev *et al.* 1992] S.I. Sukharev, V.A. Klenchin, S.M. Serov, L.V. Chernomordik and Y.A. Chizmadzhev. *Electroporation and electrophoretic DNA transfer into cells: the effect of DNA interaction with electropores*. Biophys J., vol. 63, no. 5, pages 1320–1327, 1992. (Cited on pages 2, 7 and 110.)

[Suzuki 1990] M. Suzuki. *Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations*. Phys. Lett. A, vol. 146, no. 6, pages 319–323, 1990. (Cited on pages 95, 99, 100 and 101.)

[Taflove & Hagness 2005] A. Taflove and C. Hagness. Computational Electrodynamics: The Finite-Difference Time-Domain Method. MA: Artech House, Norwood, third édition, 2005. (Cited on pages 3, 8 and 110.)

[Taube *et al.* 2009] A. Taube, M. Dumbser, C.D. Munz and R. Schneider. *A high order discontinuous Galerkin method with local time stepping for the Maxwell equations*. Int. J. Numer. Model.: Elec. Netw. Dev. and Fields, vol. 22, no. 1, pages 77–103, 2009. (Cited on pages 4, 9, 38, 143 and 147.)

[Tsong 1991] T.Y. Tsong. *Electroporation of cell membranes*. Biophys J., vol. 60, no. 2, pages 297–306, 1991. (Cited on pages 2, 7 and 110.)

[Varah 1980] J.M. Varah. *Stability restrictions on second, three-level finite-difference schemes for parabolic equations*. SIAM J. Numer. Anal., vol. 17, no. 2, pages 300–309, 1980. (Cited on page 38.)

[Verwer & Botchev 2009] J.G. Verwer and M.A. Botchev. *Unconditionaly stable integration of Maxwell's equations*. Linear Algebra and its Applications, vol. 431, no. 3-4, pages 300–317, 2009. (Cited on pages 37, 41 and 53.)

[Verwer 2009] J.G. Verwer. *Convergence and component splitting for the Crank-Nicolson-Leap-Frog integration method*. CWI Technical Report MAS-E0902, February 2009. (Cited on page 39.)

[Verwer 2010] J.G. Verwer. *Component splitting for semi-discrete Maxwell equations*. BIT Numer. Math., vol. 51, no. 2, pages 427–445, 2010. (Cited on pages i, 4, 10, 37, 39, 47, 50, 52, 53, 55, 57, 96, 110, 117, 120, 141, 142 and 145.)

[Verwer 2012] J.G. Verwer. *Composition methods, Maxwell's equations and source term*. SIAM J. Numer. Anal., vol. 50, no. 2, pages 439–457, 2012. (Cited on pages 95, 101 and 106.)

[Viquerat *et al.* 2013] J. Viquerat, M. Klemm, S. Lantri and C. Scheid. *Theoretical and numerical analysis of local dispersion models coupled to a discontinuous Galerkin time-domain method for Maxwell's equations*. Inria Reasearch Report RR-8298, May 2013. (Cited on pages 143 and 147.)

[Šel *et al.* 2005] D. Šel, D. Cukjati, D. Batiuskaite, T. Slivnik, L.M. Mir and D. Miklavčič. *Sequential finite element model of tissue electropermeabilization*. IEEE Trans. Biomed. Engrg., vol. 52, no. 5, pages 816–827, 2005. (Cited on pages 2, 7 and 110.)

[W. Hundsdorfer 2003] J. Jaffr W. Hundsdorfer. *Implicit-Explicit Time Stepping with Spatial Discontinuous Finite Elements*. Appl. Num. Math., vol. 45, pages 231–254, 2003. (Cited on pages 3, 9 and 39.)

[Wang *et al.* 2010] B. Wang, Z. Xie and Z. Zhang. *Error analysis of a discontinuous Galerkin method for Maxwell's equations in dispersive media*. J. Comput. Phys., vol. 229, no. 22, pages 8552–8563, 2010. (Cited on page 110.)

[Wesseling 2001] P. Wesseling. Principles of Computational Fluid Dynamics, volume 29 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2001. (Cited on pages 3, 9 and 38.)

[Yee 1966] K.S. Yee. *Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media*. IEEE Trans. Antenn. Propag., vol. 14, no. 3, pages 302–307, 1966. (Cited on pages 2, 8 and 23.)

[Yoshida 1990] H. Yoshida. *Construction of higher order symplectic integrators*. Phys. Lett. A, vol. 150, no. 5,6,7, pages 262–268, 1990. (Cited on pages 95, 99 and 100.)

Méthodes Galerkine discontinues localement implicites en domaine temporel pour la propagation des ondes électromagnétiques dans les tissus biologiques

Résumé : Dans cette étude nous considérons les équations de Maxwell en domaine temporel pour des milieux non-dispersifs et dispersifs. Le principal objectif est de proposer des méthodes de type éléments finis d'ordre arbitrairement élevé pour les équations de Maxwell sur des maillages simplexes et des schémas d'intégration en temps efficaces sur des maillages non-uniformes, localement raffinés.

Nous considérons des méthodes GDDT (Galerkine Discontinues en Domaine Temporel) nodales s'appuyant sur une interpolation polynomiale d'ordre arbitrairement élevé des composantes du champ électromagnétique. Les méthodes GDDT pour les équations de Maxwell s'appuient le plus souvent sur des schémas d'intégration en temps explicites dont la condition de stabilité peut łtre très restrictive pour des maillages localement raffinés. Pour surmonter cette limitation, nous considérons des schémas d'intégration en temps localement implicites (ou hybrides implicites-explicites) qui consistent à appliquer un schéma implicite localement, dans les région raffinées, tout en préservant un schéma explicite sur le reste du maillage. Nous présentons une étude théorique complète (formulation, stabilité, convergence, coût du traitement implicite) et une comparaison de deux méthodes GDDT localement implicites. Des expériences numériques en 2D et 3D valident la théorie et illustrent l'utilité des schémas proposés.

Le traitement numérique de milieux de propagation complexes (c'est à dire de modèles de dispersion) est également l'un des objectifs. Nous considérons l'interaction des ondes électromagnétiques avec les tissus biologiques qui est au coeur d'applications telles que l'évaluation des effets nocifs de l'exposition à des ondes électromagnétiques ou encore l'utilisation de ces dernières à des fins thérapeutiques ou diagnostic. La modélisation numérique de l'interaction des ondes électromagnétiques avec les tissus biologiques nécessite alors de résoudre le système de Maxwell avec des modèles appropriés de dispersion. Nous proposons une méthode GDDT localement implicite pour le modèle de Debye. Dans ce contexte, les caractéristiques de cette méthode sont bien adaptées à la miniaturisation des dispositifs électroniques ou encore à la petite taille des cellules cancéreuses, pour lesquelles l'utilisation de maillages non-uniformes, localement raffinés est certainement une prérogative importante pour obtenir une solution numérique précise.

Mots clés : équations de Maxwell, maillages localement raffinés, méthode Galerkine discontinue en domaine temporel, schémas d'intégration en temps localement implicites, milieux de propagation complexes, tissus biologiques, modèle de Debye

Locally implicit Discontinuous Galerkin time-domain methods for electromagnetic wave propagation in biological tissues

Abstract: This work deals with the time-domain formulation of Maxwell's equations in non-dispersive and dispersive media. The main objective of the study is to propose arbitrary high-order finite element type methods on simplicial meshes for the discretization of Maxwell's equations and efficient time integration methods for dealing with grid induced stiffness when using non-uniform (locally refined) meshes.

We consider Discontinuous Galerkin Time-Domain (DGTD) methods relying on an arbitrary high-order polynomial interpolation of the component of the electromagnetic field, and their computer implementations make use of nodal (Lagrange) basis expansions on simplicial elements. Existing DGTD methods for the solution of the time-domain Maxwell equations often rely on explicit time integration schemes and are therefore constrained by a stability condition that can be very restrictive on highly refined meshes. To overcome this limitation, we consider time integration methods that consist in applying an implicit time integration scheme locally i.e. in the refined regions of the mesh, while preserving an explicit time scheme in the complementary part, resulting in locally implicit (or hybrid implicit-explicit) time integration schemes. In this dissertation, we present a full theoretical study (formulation, stability, convergence, numerical analysis of the implicit treatment) and a comparison of two locally implicit DGTD methods for the first-order formulation of Maxwell's equations. Numerical experiments for two- and three-dimensional electromagnetic waves propagation problems validate the theory and illustrate the usefulness of the proposed time integration schemes.

The numerical treatment of complex propagation media models (i.e. physical dispersion models) is also one of the objectives of the present study. We consider the interaction of electromagnetic waves with biological tissues that is of interest to several applications of societal relevance such as the assessment of potential adverse effects of electromagnetic fields or the utilization of electromagnetic waves for therapeutic or diagnostic purposes. Numerical modeling of electromagnetic wave propagation in interaction with biological tissues requires to solve the system of Maxwell's equations coupled to appropriate models of physical dispersion in the tissues, such the Debye model. In this dissertation we derive a locally implicit DGTD method for this dispersion model. In this context, the characteristics of locally implicit methods are well suited to the miniaturization of electronic devices or the small size of cancerous cells, for which the use of non-uniform (locally refined) meshes are certainly key for the efficient numerical solution of wave propagation problems in biological tissues.

Keywords: Maxwell's equations, locally refined meshes, Discontinuous Galerkin Time-Domain method, locally implicit time integration schemes, complex propagation media models, biological tissues, Debye model