Karlsruhe Institute of Technology

Locally Implicit Time Integration for Linear Maxwell's Equations

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der KIT-Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

M. Sc. ANDREAS STURM

Tag der mündlichen Prüfung: 26.04.2017

1. Referentin: Prof. Dr. Marlis Hochbruck
2. Referent: Prof. Dr. Tobias Jahnke
3. Referent: Prof. Dr. Marcus J. Grote

Acknowledgment

In German.

An erster Stelle möchte ich mich bei meiner Betreuerin Prof. Dr. Marlis Hochbruck bedanken. Seit ich vor 5 1/2 Jahren nach Karlsruhe gekommen bin, ist sie die Mentorin, die mich durch die Mathematik geleitet und die diese Arbeit ermöglicht hat. Zuerst als Hiwi und anschließend als Doktorand hatte sie immer Zeit und Rat für mich. Aus vielen Diskussionen ging ich (mit mehr Arbeit, aber) wesentlich schlauer hervor. Ihr Anspruch an die Klarheit und Struktur von mathematischen Arbeiten, an die Qualität von Vorträgen und an die Wichtigkeit der Lehre prägen mich und meine Arbeit in besonderer Weise. Auch in einer Zeit, in der es mir schlecht ging, war Marlis immer für mich da und hat dadurch in einer nicht zu überschätzenden Weise geholfen. Dafür, für die viele Freiheit auf meine Art und Weise zu arbeiten und für die hervorragende Betreuung möchte ich mich ganz herzlich bedanken.

Vielen Dank auch an meinen Zweitbetreuer Prof. Dr. Tobias Jahnke, dessen viele gute Korrekturvorschläge wesentlich zu dieser Arbeit beigetragen haben.

Mein Dank gebührt auch Prof. Dr. Marcus J. Grote für sein Gutachten und seine Expertise zu dieser Dissertation.

Sehr gerne möchte ich mich auch bei meiner Arbeitsgruppe bedanken. Für die viele Unterstützung und die immer gute Stimmung und Zusammenarbeit. Liebesten Dank dabei an Mathias und Christian für ihre Geduld mit mir bezüglich meiner Unwissenheit im Umgang mit Computern und allen technischen Geräten. Auch ganz herzlichen Dank an meine Mitdoktoranden Jonas, Simone, David, Lukas, Robin, Patrick, Julian und Lena, und natürlich auch an Michaela und Christian – abseits guter Zusammenarbeit hatten wir auch immer viel Spaß zusammen.

Nicht vergessen möchte ich mich bei meinen Freunden zu bedanken. Es war und ist mir ein großes Fest mit euch allen befreundet zu sein. Vielen Dank für unzählige tolle Erinnerungen an Steffen, Manu, Julian, Bianca, Uli, Simon, Adi, Cosi und an meine Freunde schon seit der Schulzeit Thomas, Lisa, Lukas, Jonas, Eddi, Daniel und Julian.

Zu allerletzt möchte ich mich bei meiner Familie bedanken. Bei Daniel und Matthias, den beiden besten Brüdern, die man sich wünschen kann. Und bei meinen Eltern für ihren Rat, ihre rückhaltlose Unterstützung und einfach alles, was sie für mich getan haben. Diese Arbeit ist für euch.

Contents

Notation

Throughout this thesis we use the following notation: We write the **scalar product** of two vectors a, $b \in \mathbb{R}^3$ as

$$a \cdot b = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \cdot \begin{pmatrix} b_x \\ b_y \\ b_y \end{pmatrix} = a_x b_x + a_y b_y + a_z b_z,$$

and denote their **cross product** by

$$a \times b = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \times \begin{pmatrix} b_x \\ b_y \\ b_y \end{pmatrix} = \begin{pmatrix} a_y b_z - a_z b_y \\ a_z b_x - a_x b_z \\ a_x b_y - a_y b_x \end{pmatrix}.$$

Let $\mathbb{R}_+ = (0, \infty)$ denote the positive real numbers. We often consider multivariate functions $u : \mathbb{R}_+ \times \mathbb{R}^3 \to \mathbb{R}$, where the first variable is the time variable t and the three remaining variables are the space variables x, y, z. We usually drop the space variables and just write $u(t) = u(t, x, y, z)$ and often also omit the time variable such that $u = u(t) = u(t, x, y, z)$.

We denote the **partial derivatives** of u by

$$\partial_t u = \frac{\partial}{\partial t} u, \qquad \partial_x u = \frac{\partial}{\partial x} u, \qquad \partial_y u = \frac{\partial}{\partial y} u, \qquad \partial_z u = \frac{\partial}{\partial z} u.$$

The spatial derivatives are collected in the **gradient** of u, which is given by

$$\operatorname{grad} u = \begin{pmatrix} \partial_x u \\ \partial_y u \\ \partial_z u \end{pmatrix}.$$

If a function $v : \mathbb{R}_+ \to \mathbb{R}$ only depends on the time, we write its **time derivative** by $\dot{v} = \frac{d}{dt} v$.

For a vector field $\mathbf{U} : \mathbb{R}_+ \times \mathbb{R}^3 \to \mathbb{R}^3$ we define the following differential operators acting on the spatial variables: The **divergence** of $\mathbf{U}$ is defined as

$$\operatorname{div} \mathbf{U} = \operatorname{div} \begin{pmatrix} \mathbf{U}_x \\ \mathbf{U}_y \\ \mathbf{U}_z \end{pmatrix} = \partial_x \mathbf{U}_x + \partial_y \mathbf{U}_y + \partial_z \mathbf{U}_z,$$

and its **curl** by

$$\operatorname{curl} \mathbf{U} = \operatorname{curl} \begin{pmatrix} \mathbf{U}_x \\ \mathbf{U}_y \\ \mathbf{U}_z \end{pmatrix} = \begin{pmatrix} \partial_y \mathbf{U}_z - \partial_z \mathbf{U}_y \\ \partial_z \mathbf{U}_x - \partial_x \mathbf{U}_z \\ \partial_x \mathbf{U}_y - \partial_y \mathbf{U}_x \end{pmatrix}.$$

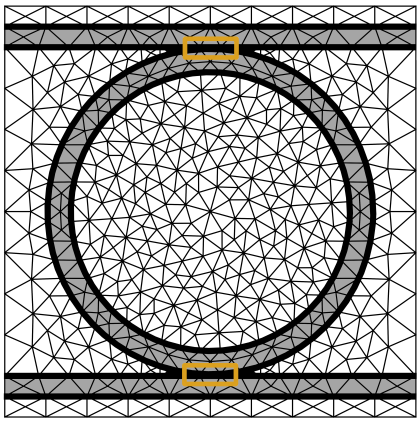
CHAPTER 0

Introduction

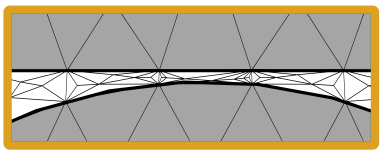
Motivation

Maxwell's equations are the fundamental laws in electromagnetism. They describe the interaction of time-dependent electromagnetic fields with each other, as well as their behavior in different materials and in the presence (or absence) of electrical currents and charges. Among many other applications they play a crucial role in the analysis and design of nanophotonic systems such as antennas, photonic cristals, waveguides and interferometers.

Despite the fact that Maxwell's equations have been a research objective for the last 150 years, they still pose significant challenges and analytic solutions can only be found for certain simplified systems. With the rise of computing power this shortcoming was cured by the techniques of numerical analysis providing (high order) approximations of Maxwell's equations. In many applications the numerical approximation is realized by a finite-difference space discretization in combination with an explicit time integrator. One of the oldest and most popular methods following this recipe to solve the time-dependent Maxwell's equations was proposed by Yee [1966]. This method comprises a finite-difference space discretization on a staggered spatial grid – the famous Yee grid – and the explicit Verlet (or leap frog) time integrator. However, there are two shortcomings in this popular approach. On the one hand, methods based on finite-differences are limited to domains with a regular geometry and their generalization to unstructured grids is difficult. Moreover, they do not allow for adaptivity and the numerical analysis requires high regularity of the exact solution of Maxwell's equations, which is not reasonable in realistic applications. As a remedy to this problem other space discretization techniques were proposed such as schemes based on Nédélec elements (Nédélec [1980]) or discontinuous Galerkin (dG) methods (Reed and Hill [1973]), see also the textbooks Monk [2003], Di Pietro and Ern [2012] and Hesthaven and Warburton [2008]. On the other hand, despite their wide spread application, explicit time integrators such as the Verlet method (Fahs [2009]), explicit two and three stage Runge-Kutta (RK) methods (Burman et al. [2010]) and low-storage RK schemes (Diehl et al. [2010]), suffer from severe stability issues when applied to stiff problems such as the spatially discretized Maxwell's equations. In fact, in order to guarantee stability the time-step size of these methods is subject to a strong limitation (CFL condition), which often renders the application of explicit time integrators inefficient.



(a) The ring resonator and the wave guides (grey areas) are separated by a small gap.



(b) Enlargement of the gap between the ring resonator and the wave guides.

Figure 1: Mesh of a ring resonator. The white and grey areas are made of different materials.

In particular, explicit time integration schemes perform poorly in the case of locally refined spatial grids, i.e. grids which consist mostly of coarse elements but also of a few (very) tiny elements (grid-induced stiffness). However, many applications require such a locally refined grid, e.g. to resolve tiny geometric details or to guarantee the optimal convergence order of the space discretization. A concrete example is that of a ring resonator where the different materials require a space discretization by a locally refined grid, see Figure 1. These problems demand for more adapted time integration methods and two classes of novel time integrators have been proposed in the literature. The first class are explicit local time stepping schemes. They were initially proposed in Diaz and Grote [2009] for the second order wave equation and extended to Maxwell's equations in Grote and Mitkova [2010]. In several succeeding papers these methods were extended and generalized, see Chapter 5 for a detailed discussion. The underlying idea of local time stepping methods is to treat the tiny elements in the spatial grid with a small time-step size, thus avoiding a restrictive CFL condition emanating from these fine elements, and treating the remaining coarse elements with a bigger time step. The second class consists of locally implicit time integrators and originates from Piperno [2006] and Verwer [2011]. Further insight into these methods was provided in Descombes et al. [2013]. The key ingredient in locally implicit time integration schemes is to treat the fine elements with an implicit time integrator while retaining an explicit time integration scheme for the remaining coarse elements.

Aims und results

In this thesis we provide a deeper understanding and a rigorous error analysis of the locally implicit time integrator proposed in Verwer [2011]. So far, the method was only constructed by considering the spatially discretized Maxwell's equations as a system of ODEs and the error analysis was limited to the non-stiff case since the error constants depended on the spatial mesh. Moreover, it was unclear which elements of the spatial grid exactly enter the CFL condition. We closed this gap by combining the idea of Hochbruck and Pažur [2015] to consider the spatially discretized Maxwell's equations in a variational setting with an adaption of the locally implicit scheme from Verwer [2011]. This allows us to control exactly which spatial elements are integrated implicitly and which explicitly and we can prove rigorously which of them enter the CFL condition. It turns out that in order to ensure a CFL condition, which only depends on the coarse elements of the spatial grid, not only all fine elements have to be integrated

implicitly but so do their (coarse) neighbors. Another result emanating from our new ansatz is an error analysis which is independent of the spatial grid and thus also valid in the relevant stiff regime. In fact, we can prove that the locally implicit method is of order two in the time step and of order k in the mesh parameter, when using a dG space discretization with polynomials of order k. We developed a novel technique for the stability and the convergence proof, which is –in our appreciation– simpler than an energy technique and which also provides a rigorous error analysis for the fully explicit Verlet method and the fully implicit Crank–Nicolson scheme. These results were published in Hochbruck and Sturm [2016].

So far, locally implicit schemes discussed in the literature were limited to an unstabilized spatial discretization, which is usually referred to as a central fluxes dG discretization. However, a stabilized (upwind fluxes) dG discretization provides many benefits such as a better stability behavior and a higher spatial convergence rate. We were able to adapt the locally implicit scheme to this space discretization ensuring that it also features a CFL condition which solely depends on the coarse elements in the spatial grid. Moreover, we can prove that it is convergent of order two in the time step and $k + 1/2$ on the coarse part of the grid and k in the fine part of the grid. It turns out that the construction of this method needs completely new ideas and that the error analysis has to be carried out with an energy technique. As byproduct of our analysis, we also give rigorous error bounds for a fully explicit Verlet-type time integrator for the upwind fluxes dG discretization of Maxwell's equations. A summary of the results can be found in Hochbruck and Sturm [2017].

Outline

This thesis is organized as follows. In Chapter 1 we introduce Maxwell's equations and discuss the particular case of linear, isotropic materials which lead to the linear Maxwell's equations we consider in this thesis. We provide the functional analytic framework in which Maxwell's equations are a well-posed problem. Chapters 2 and 3 are concerned with the spatial discretization of Maxwell's equations by means of a dG method. In Chapter 2 we introduce the discrete setting we need to formulate the dG method in Chapter 3. In this chapter we derive both the central fluxes dG discretization and the upwind fluxes dG discretization and discuss their differences. We end this chapter with an error analysis, which reveals the different techniques needed in the central fluxes case and in the upwind fluxes case. This distinction will also be employed in the fully discrete case. Chapter 4 is devoted to the time integration of the semidiscrete Maxwell's equations stemming from the dG space discretization of Chapters 2 and 3. In this chapter we study the explicit Verlet method and the implicit Crank–Nicolson method, which will be the underlying methods for our locally implicit scheme. We provide a stability analysis as well as an error analysis for both time integration methods in combination with a central fluxes dG scheme and with an upwind fluxes dG scheme. The presented techniques will be the basis for our analysis of the locally implicit scheme which we present in Chapter 5. We begin this chapter with a decomposition of the spatial mesh as preparation for the distinction of explicit and implicit time integration. Then, we derive the locally implicit scheme in combination with a central fluxes dG space discretization. Our main results for this scheme are its CFL condition (5.40) under which we can prove its stability and the convergence result given in Theorem 5.13. Next, we introduce the modifications needed to adapt the central fluxes locally implicit method to an upwind fluxes dG discretization. Our essential results for this method are the CFL condition (5.93) and the convergence result in Theorem 5.35. We conclude this thesis with Chapter 6 where we illustrate how the locally implicit methods can be implemented efficiently and where we provide numerical examples underlining the theoretical results.

CHAPTER 1

Maxwell's equations

In this chapter we present Maxwell's equations in their integral form and derive their differential form. Then, we focus on electromagnetic phenomena in isotropic, linear materials which are described by the linear Maxwell's equations and which are the underlying equations for this thesis. We shortly give an overview of the functional analytic framework in which we embed the linear Maxwell's equations and in which we can show their well-posedness. We end this chapter by discussing the energy conservation and the stability of solutions of Maxwell's equations. Our main references for this chapter are the books Monk [2003] and Kirsch and Hettlich [2015].

1.1 Maxwell's equations in integral and differential form

In the following, $\Omega \subset \mathbb{R}^3$ denotes a domain and $\mathbb{R}_+ = (0, \infty)$. The electromagnetic field is described by four vector fields called

electric field intensity	$\mathbf{E} : \mathbb{R}_+ \times \Omega \to \mathbb{R}^3$	$\left[\frac{\mathrm{V}}{\mathrm{m}}\right]$,
magnetic field intensity	$\mathbf{H} : \mathbb{R}_+ \times \Omega \to \mathbb{R}^3$	$\left[\frac{\mathrm{A}}{\mathrm{m}}\right]$,
electric displacement	$\mathbf{D} : \mathbb{R}_+ \times \Omega \to \mathbb{R}^3$	$\left[\frac{\mathrm{As}}{\mathrm{m}^2}\right]$,
magnetic induction	$\mathbf{B} : \mathbb{R}_+ \times \Omega \to \mathbb{R}^3$	$\left[\frac{\mathrm{Vs}}{\mathrm{m}^2}\right]$.

The interaction of these fields on each other as well as their dependence on the two sources

electric current density	$\mathbf{J} : \mathbb{R}_+ \times \Omega \to \mathbb{R}^3$	$\left[\frac{\mathrm{A}}{\mathrm{m}^2}\right]$,
electric charge density	$\varrho : \mathbb{R}_+ \times \Omega \to \mathbb{R}$	$\left[\frac{\mathrm{As}}{\mathrm{m}^3}\right]$,

is described by Maxwell's equations.
Remark: Frequently used units are also *Coulomb* $\mathrm{C} = \mathrm{As}$ and *Tesla* $\mathrm{T} = \mathrm{Vs/m^2}$.

1.1.1 Maxwell's equations in integral form

In order to state Maxwell's equations in integral form we consider the following setting: Let $S \subset \Omega$ be a connected, smooth surface with boundary ∂S. We denote by $n_S : S \to \mathbb{R}^3$ the continuous, unit normal vector which is always directed to the same side of S. We call this side the "positive side" of S. Moreover, we denote by $t_S : \partial S \to \mathbb{R}^3$ the unit tangent vector of ∂S that is directed counterclockwise when seen from the positive side of S. Last, let $V \subset \Omega$ be an open set with boundary ∂V and outer unit normal vector $n_V : \partial V \to \mathbb{R}^3$.

Maxwell's equations now consist of the following set of four equations which are often split into two equations containing time derivatives

$$\textbf{Faraday's law of induction} \qquad \int_{\partial S} \mathbf{E} \cdot t_S \, d\ell = -\frac{d}{dt} \int_S \mathbf{B} \cdot n_S \, ds, \tag{1.1a}$$

$$\textbf{Ampère's circuital law} \qquad \int_{\partial S} \mathbf{H} \cdot t_S \, d\ell = \frac{d}{dt} \int_S \mathbf{D} \cdot n_S \, ds + \int_S \mathbf{J} \cdot n_S \, ds, \tag{1.1b}$$

and two integral equations

$$\textbf{Gauss' magnetic law} \qquad \int_{\partial V} \mathbf{B} \cdot n_V \, ds = 0, \tag{1.2a}$$

$$\textbf{Gauss' electric law} \qquad \int_{\partial V} \mathbf{D} \cdot n_V \, ds = \int_V \varrho \, dx. \tag{1.2b}$$

The first equation (1.1a) means that a changing magnetic field induces an electric field. Equation (1.1b) states that a magnetic field can be generated by an (external) electrical current or by a changing electric field. Equation (1.2a) essentially states that there are no magnetic monopoles and that the magnetic field lines form closed loops. Finally, equation (1.2b) describes how electric charges generate an electric field.

1.1.2 Maxwell's equations in differential form

Now, we derive the differential form of Maxwell's equations from two famous theorems which hold for sufficiently smooth vector fields $\mathbf{F} : \Omega \to \mathbb{R}^3$:

$$\textbf{Stoke's theorem} \qquad \int_S \operatorname{curl} \mathbf{F} \cdot n_S \, ds = \int_{\partial S} \mathbf{F} \cdot t_S \, d\ell, \tag{1.3}$$

$$\textbf{Gauss' divergence theorem} \qquad \int_V \operatorname{div} \mathbf{F} \, dx = \int_{\partial V} \mathbf{F} \cdot n_V \, ds. \tag{1.4}$$

Applying (1.3) to (1.1) and (1.4) to (1.2), and furthermore using that S, V are arbitrary we obtain Maxwell's equations in differential form. They consist of two **curl-equations**

$$\partial_t \mathbf{B} = -\operatorname{curl} \mathbf{E}, \qquad (0,T) \times \Omega, \tag{1.5a}$$

$$\partial_t \mathbf{D} = \operatorname{curl} \mathbf{H} - \mathbf{J}, \qquad (0,T) \times \Omega, \tag{1.5b}$$

and two **div-equations**

$$\operatorname{div} \mathbf{B} = 0, \qquad (0,T) \times \Omega, \tag{1.6a}$$

$$\operatorname{div} \mathbf{D} = \varrho, \qquad (0,T) \times \Omega. \tag{1.6b}$$

These equations need to be supplemented with initial values and boundary conditions.

For sufficiently smooth $\mathbf{D}$ and $\mathbf{H}$ we can already gain a relation between the charge density ϱ and the current density $\mathbf{J}$ in the **continuity equation**

$$\partial_t \varrho + \operatorname{div} \mathbf{J} = 0. \tag{1.7}$$

This follows from (1.5b) and (1.6b) by

$$\partial_t \varrho = \operatorname{div}(\partial_t \mathbf{D}) = \operatorname{div}(\operatorname{curl} \mathbf{H}) - \operatorname{div} \mathbf{J} = -\operatorname{div} \mathbf{J},$$

where the last equation holds since $\operatorname{div}(\operatorname{curl} \cdot) = 0$.

On the other hand, if we assume the continuity equation (1.7), then the div-equations (1.6) become redundant in the sense that they only have to be ensured for $t = 0$ and then follow from the curl-equations (1.5) for all $t > 0$. We collect this in the following proposition.

Proposition 1.1. *Let* $\mathbf{B}$, $\mathbf{D}$, $\mathbf{H}$, $\mathbf{E}$ *be smooth solutions of* (1.5) *and let* ϱ *and* $\mathbf{J}$ *satisfy* (1.7). *Furthermore, assume that* (1.6) *is satisfied for* $t = 0$, *i.e.*

$$\operatorname{div} \mathbf{D}(0) = \varrho(0), \qquad \operatorname{div} \mathbf{B}(0) = 0. \tag{1.8}$$

Then, (1.6) *holds true for all* $t \in \mathbb{R}_+$.

Proof. By (1.8) we have that

$$\operatorname{div} \mathbf{D}(t) = \operatorname{div} \mathbf{D}(0) + \int_0^t \partial_t(\operatorname{div} \mathbf{D}(s))\, ds = \varrho(0) + \int_0^t \partial_t(\operatorname{div} \mathbf{D}(s))\, ds. \tag{1.9}$$

Furthermore, by (1.5b) and (1.7), we conclude

$$\partial_t(\operatorname{div} \mathbf{D}) = \operatorname{div}(\partial_t \mathbf{D}) = \operatorname{div}(\operatorname{curl} \mathbf{H}) - \operatorname{div} \mathbf{J} = \partial_t \varrho.$$

Inserting this into (1.9) shows that (1.6b) holds for all $t \in \mathbb{R}_+$. In order to prove (1.6a) we take the divergence of (1.5a) and obtain

$$\partial_t(\operatorname{div} \mathbf{B}) = \operatorname{div}(\partial_t \mathbf{B}) = -\operatorname{div}(\operatorname{curl} \mathbf{E}) = 0.$$

Together with (1.8) this yields

$$\operatorname{div} \mathbf{B}(t) \equiv \operatorname{div} \mathbf{B}(0) = 0,$$

which finishes the proof. □

Considering the set of equations (1.1)–(1.2) or (1.5)–(1.6), respectively, we see that we have 12 unknowns $\mathbf{B}$, $\mathbf{D}$, $\mathbf{H}$ and $\mathbf{E}$ but only eight independent equations (six if we assume (1.8)). Hence, we need additional conditions to ensure the well-posedness of Maxwell's equations.

1.1.3 Constitutive equations

The **constitutive equations** provide a description of how the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$ give rise to the electric displacement $\mathbf{D}$ and the magnetic induction $\mathbf{B}$:

$$\mathbf{D} = \mathbf{D}(\mathbf{E}, \mathbf{H}), \qquad \mathbf{B} = \mathbf{B}(\mathbf{E}, \mathbf{H}).$$

In general, the relationships are complicated and strongly depend on the material (e.g. molecular character, density, temperature) in which the electromagnetic phenomena are examined.

For stationary media a typical representation is given by

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}, \qquad \mathbf{B} = \mu_0 \mathbf{H} + \mu_0 \mathbf{M},$$

where $\mathbf{P}$ and $\mathbf{M}$ denote the **polarization** and **magnetization**, respectively, and ε_0 and μ_0 are the **permittivity** and the **permeability of free space**. The values of the latter are given by

$$\varepsilon_0 = 8.854 \cdot 10^{-12} \frac{\mathrm{As}}{\mathrm{Vm}}, \qquad \mu_0 = 4\pi \cdot 10^{-7} \frac{\mathrm{Vs}}{\mathrm{Am}}.$$

These quantities are related to the **speed of light in vacuum** c_0 by

$$c_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} = 2.998 \cdot 10^8 \frac{\mathrm{m}}{\mathrm{s}}.$$

An example fitting in the upper framework is light propagating through optical materials in photonic crystals described by the **Kerr nonlinearity**

$$\mathbf{P}(\mathbf{E}) = \varepsilon_0(\boldsymbol{\varepsilon}_r - 1 + \chi|\mathbf{E}|^2)\mathbf{E} \quad (\chi \in \mathbb{R}), \qquad \mathbf{M} \equiv 0,$$

cf. [Busch et al., 2007, Section 3], [Dörfler et al., 2011, Chapter 1] and Pototschnig et al. [2009]. Here, $\boldsymbol{\varepsilon}_r : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ is the **relative permittivity**.

In non-ferroelectric and non-ferromagnetic media the electric displacement and the magnetic induction depend **linearly** on the electric field and the magnetic field, respectively, if the fields are relatively small. Then, we have

$$\mathbf{D} = \boldsymbol{\varepsilon}\mathbf{E}, \qquad \mathbf{B} = \boldsymbol{\mu}\mathbf{H},$$

with matrix-valued functions $\boldsymbol{\varepsilon} = \varepsilon_0 \boldsymbol{\varepsilon}_r : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$, the **dielectric tensor** with **relative permittivity** $\boldsymbol{\varepsilon}_r$, and $\boldsymbol{\mu} = \mu_0 \boldsymbol{\mu}_r : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$, the **permeability tensor** with **relative permeability** $\boldsymbol{\mu}_r$. We call such a material **linear** and **anisotropic**. Note that $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ need not to be continuous. If Ω is a **composite material**, i.e. made up of different materials, the coefficients $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ may jump at material interfaces.

In the special case where the polarization and the magnetization do not depend on the directions, the dielectricity and the permeability can be modeled as just real functions $\varepsilon_r, \mu_r : \mathbb{R}^3 \to \mathbb{R}$. We call such a material **isotropic**.

In the simplest case, ε_r and μ_r are constants and we call such a medium **homogeneous**. In such a medium light travels with speed

$$c = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{c_0}{n}, \qquad n = \sqrt{\varepsilon_r \mu_r},$$

where n is called the **refractive index** of the medium. An important example of a homogeneous medium is that of **vacuum**, where $\varepsilon_r = 1$, $\mu_r = 1$ and thus $n = 1$. For some other materials the refractive indices are given by

$$n_{\text{air}} = 1.000292, \qquad n_{\text{water}} = 1.33, \qquad n_{\text{glass}} \approx 1.46 \ldots 1.65, \qquad n_{\text{diamond}} = 2.42.$$

Last, we point out that both the current density $\mathbf{J}$ and the charge density ϱ can depend on the material and the fields. In **conducting** media the electric field $\mathbf{E}$ induces a current $\mathbf{J}$. In a linear approximation this is described by **Ohm's law**

$$\mathbf{J} = \sigma\mathbf{E} + \mathbf{J}_e,$$

where $\mathbf{J}_e$ is an **external current density**. For isotropic materials the function $\sigma : \mathbb{R}^3 \to \mathbb{R}$ is called the **conductivity**. In anistoropic media the function σ is matrix-valued and in vacuum we have $\sigma \equiv 0$.

In this thesis we focus on linear, isotropic materials. This results in the **linear Maxwell's equations**. Moreover, we assume that the material is nonconducting, i.e. $\sigma \equiv 0$.

1.2 Linear Maxwell's equations

We substitute the linear constitutive relations $\mathbf{D} = \varepsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$ into (1.5),

$$\mu\partial_t\mathbf{H} = -\operatorname{curl}\mathbf{E}, \qquad (0,T)\times\Omega, \tag{1.10a}$$
$$\varepsilon\partial_t\mathbf{E} = \operatorname{curl}\mathbf{H} - \mathbf{J}, \qquad (0,T)\times\Omega, \tag{1.10b}$$

and into (1.6),

$$\operatorname{div}(\mu\mathbf{H}) = 0, \qquad (0,T)\times\Omega, \tag{1.11a}$$
$$\operatorname{div}(\varepsilon\mathbf{E}) = \varrho, \qquad (0,T)\times\Omega. \tag{1.11b}$$

These equations are endowed with **initial values** $\mathbf{H}(0) = \mathbf{H}^0$ and $\mathbf{E}(0) = \mathbf{E}^0$ satisfying $\operatorname{div}(\mu\mathbf{H}^0) = 0$ and $\operatorname{div}(\varepsilon\mathbf{E}^0) = \varrho(0)$, respectively.

As mentioned above the coefficients ε and μ are allowed to have jumps. In this case we cannot use Maxwell's equations (1.10), (1.11) directly since the data is not smooth enough. Thus, we consider next interface conditions for $\mathbf{E}$ and $\mathbf{H}$ and also discuss appropriate boundary conditions.

1.2.1 Interface and boundary conditions

We consider the situation where Ω is made up of two different materials, say material 1 and 2, which share a common surface S. We denote by n_S the unit normal to S and by $\mathbf{E}_j$, $\mathbf{H}_j$, ε_j, μ_j the restriction of the respective functions to material $j \in \{1,2\}$.

From (1.1a) and (1.2a) one can obtain

$$n_S \times (\mathbf{E}_1 - \mathbf{E}_2) = 0 \qquad \text{on } S, \tag{1.12a}$$
$$n_S \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_S \qquad \text{on } S, \tag{1.12b}$$
$$n_S \cdot (\mu_1\mathbf{H}_1 - \mu_2\mathbf{H}_2) = 0 \qquad \text{on } S, \tag{1.12c}$$
$$n_S \cdot (\varepsilon_1\mathbf{E}_1 - \varepsilon_2\mathbf{E}_2) = \varrho_S \qquad \text{on } S, \tag{1.12d}$$

where ϱ_S is the surface charge density and $\mathbf{J}_S$ is the surface current density on S, cf. [Kirsch and Hettlich, 2015, Section 1.4] and [Monk, 2003, Section 1.2.2] for details. In many applications we can assume $\mathbf{J}_S \equiv 0$. Then (1.12b) becomes

$$n_S \times (\mathbf{H}_1 - \mathbf{H}_2) = 0 \qquad \text{on } S. \tag{1.12e}$$

The conditions (1.12a) and (1.12e) mean that both the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$ have **continuous tangential components** at interfaces. On the other hand (1.12c) and (1.12d) state that they exhibit **jumps in the normal components** if ε and μ are discontinuous, respectively. In the presence of material discontinuities, a numerical scheme has to take this behavior into account.

Since we are interested in solving Maxwell's equation in a bounded domain we need appropriate boundary conditions for $\mathbf{E}$ and $\mathbf{H}$ on $\partial\Omega$. In this thesis we consider the case of **perfectly conducting boundary conditions**, i.e. we assume that Ω is surrounded by an idealized perfect conductor. By letting $\sigma \to \infty$, Ohm's law shows that $\mathbf{E} \to 0$ if we demand that $\mathbf{J}$ stays finite. Thus, we conclude that inside a perfect conductor the electric field has to vanish, whence we deduce from (1.12a) the boundary condition

$$n \times \mathbf{E} = 0 \qquad \text{on } \partial\Omega. \tag{1.13}$$

Here and from now on we denote by n the **unit outward normal** to Ω. This condition implies

$$n \cdot (\mu\mathbf{H}) \equiv \text{const} \qquad \text{on } \partial\Omega, \tag{1.14}$$

since

$$\partial_t(n \cdot (\mu \mathbf{H})) = n \cdot (\mu \partial_t \mathbf{H}) = -n \cdot \operatorname{curl} \mathbf{E} = \operatorname{div}(n \times \mathbf{E}) - \mathbf{E} \cdot \operatorname{curl} n = 0.$$

Here, we used $\operatorname{div}(\mathbf{U} \times \mathbf{V}) = \mathbf{V} \cdot \operatorname{curl} \mathbf{U} - \mathbf{U} \cdot \operatorname{curl} \mathbf{V}$ for the third equation. The last equation holds because of (1.13) and since n can be written as gradient of a parametrization of $\partial\Omega$ and $\operatorname{curl}(\operatorname{grad} \cdot) = 0$. We conclude that

$$n \cdot (\mu \mathbf{H}(t)) = n \cdot (\mu \mathbf{H}^0), \qquad \text{for all } t \in \mathbb{R}_+.$$

Hence, it is sufficient to pose boundary conditions on the electric field $\mathbf{E}$ and on the initial value of $\mathbf{H}$ only. In the following we will assume that the normal components of $\mathbf{H}$ vanish on the boundary,

$$n \cdot (\mu \mathbf{H}) = 0 \qquad \text{on } \partial\Omega. \tag{1.15}$$

1.2.2 Reduction to two dimensions

If the underlying physical system is homogeneous in z-direction Maxwell's equations (1.10) decouple into two sets of three equations, cf. Niegemann [2009]. The first case is the **transverse-electric (TE) polarization** where the associated equations read

$$\begin{aligned}
\mu \partial_t \mathbf{H}_z &= -\partial_x \mathbf{E}_y + \partial_y \mathbf{E}_x, && (0,T) \times \Omega, \\
\varepsilon \partial_t \mathbf{E}_x &= \partial_y \mathbf{H}_z - \mathbf{J}_x, && (0,T) \times \Omega, \\
\varepsilon \partial_t \mathbf{E}_y &= -\partial_y \mathbf{H}_z - \mathbf{J}_y, && (0,T) \times \Omega, \\
n_x \mathbf{E}_y - n_y \mathbf{E}_x &= 0, && (0,T) \times \partial\Omega.
\end{aligned} \tag{1.16}$$

Here, the electric field vector lies in the (x, y)-plane and the magnetic field vector is directed in z-direction. In the second case, the **transverse-magnetic (TM) polarization**, it is the other way round. The associated equations read

$$\begin{aligned}
\mu \partial_t \mathbf{H}_x &= -\partial_y \mathbf{E}_z, && (0,T) \times \Omega, \\
\mu \partial_t \mathbf{H}_y &= \partial_x \mathbf{E}_z, && (0,T) \times \Omega, \\
\varepsilon \partial_t \mathbf{E}_z &= -\partial_y \mathbf{H}_x + \partial_x \mathbf{H}_y - \mathbf{J}_z, && (0,T) \times \Omega, \\
\mathbf{E}_z &= 0, && (0,T) \times \partial\Omega.
\end{aligned} \tag{1.17}$$

Our later numerical experiments will be carried out for the TM case.

1.3 Well-posedness of linear Maxwell's equations

From now on we consider the system

$$\begin{aligned}
\mu \partial_t \mathbf{H} &= -\operatorname{curl} \mathbf{E}, && (0,T) \times \Omega, \\
\varepsilon \partial_t \mathbf{E} &= \operatorname{curl} \mathbf{H} - \mathbf{J}, && (0,T) \times \Omega, \\
\mathbf{H}(0) &= \mathbf{H}^0, \quad \mathbf{E}(0) = \mathbf{E}^0, && \Omega, \\
n \times \mathbf{E} &= 0, && (0,T) \times \partial\Omega.
\end{aligned} \tag{1.18}$$

We assume the continuity equation (1.7) and for the initial values we demand

$$\begin{aligned}
&\operatorname{div}(\mu \mathbf{H}^0) = 0, \quad \operatorname{div}(\varepsilon \mathbf{E}^0) = \varrho(0), && \Omega, \\
&n \cdot (\mu \mathbf{H}^0) = 0, && \partial\Omega.
\end{aligned} \tag{1.19}$$

Furthermore, we assume that the coefficients ε, μ are **bounded** and **uniformly positive definite**, i.e.

$$\varepsilon, \mu \in L^\infty(\Omega), \qquad \varepsilon, \mu \geq \delta, \tag{1.20}$$

for a constant $\delta > 0$. We can write Maxwell's equations (1.18) as the **abstract Cauchy problem**

$$\partial_t \mathbf{u}(t) = \mathcal{C}\mathbf{u}(t) + \mathbf{j}(t), \qquad \mathbf{u}(0) = \mathbf{u}^0, \tag{1.21}$$

where we collected the electric field and the magnetic field in $\mathbf{u} = (\mathbf{H}, \mathbf{E})$ and the current density in $\mathbf{j} = (0, -\varepsilon^{-1}\mathbf{J})$, and where $\mathcal{C}$ is the **Maxwell operator**

$$\mathcal{C} = \begin{pmatrix} 0 & -\mathcal{C}_{\mathbf{E}} \\ \mathcal{C}_{\mathbf{H}} & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\mu^{-1}\operatorname{curl} \\ \varepsilon^{-1}\operatorname{curl} & 0 \end{pmatrix}. \tag{1.22}$$

We will specify the exact mathematical setting in which $\mathcal{C}$ is a well-defined operator in Section 1.3.2. We already indicate that this setting has to incorporate the boundary condition on the electric field $\mathbf{E}$ otherwise the Cauchy problem (1.21) is not equivalent to Maxwell's equations (1.18).

In the next section we give a short overview on the well-posedness of more general abstract evolution equations.

1.3.1 Abstract evolution equations and semigroups

The material in this section is taken from Engel and Nagel [2000], [Jacob and Zwart, 2012, Chapters 5 and 6] and Pazy [1983]. We also considered the lecture notes Schnaubelt [2010–2011], Schnaubelt [2012–2013] and Schnaubelt [2015].

Let $(X, (\cdot,\cdot)_X)$ be a Hilbert space with corresponding norm $\|\cdot\|_X^2 = (\cdot,\cdot)_X$. By $\mathcal{L}(X)$ we denote the space of all bounded linear operators from X into X with operator norm

$$\|\mathcal{A}\|_{X\leftarrow X} = \sup_{\substack{x\in X\\ x\neq 0}} \frac{\|\mathcal{A}x\|_X}{\|x\|_X}.$$

Definition 1.2. *A one-parameter family $(T(t))_{t\geq 0}$ of bounded linear operators from X to X is called a **semigroup of bounded linear operators on** X if*

(a) $T(0) = I$ *and*

(b) $T(t+s) = T(t)T(s)$ *for all* $t, s \geq 0$.

*A semigroup $(T(t))_{t\geq 0}$ is called a **strongly continuous semigroup** or **C_0-semigroup*** if for all $x \in X$,*

$$\lim_{t\to 0+} \|T(t)x - x\|_X = 0;$$

*i.e. $t \mapsto T(t)$ is **strongly continuous** at 0.*

We call X the **state space**. If we replace in Defintion 1.2 "$t, s \geq 0$" by "$t, s \in \mathbb{R}$" and "$t \to 0+$" by "$t \to 0$" we obtain the concept of a (**strongly continuous**) **group**.

Lemma 1.3. *A strongly continuous semigroup $(T(t))_{t\geq 0}$ has the following properties:*

(a) *There exist constants $M \geq 1$ and $\omega \geq 0$ such that*

$$\|T(t)\|_{X\leftarrow X} \leq Me^{\omega t}, \qquad \text{for all } t \geq 0. \tag{1.23}$$

*C_0 abbreviates "Cesàsro summable of order 0"

(b) *The mapping $t \mapsto T(t)$ is strongly continuous on $[0,\infty)$, i.e.*

$$\lim_{s\to 0} \|T(t+s)x - T(t)x\|_X = 0, \qquad \text{for all } t \geq 0.$$

If $M = 1$ and $\omega = 0$ the semigroup is called a **contraction semigroup**.

Example 1.4. We illustrate the connection between semigroups and Cauchy problems with the simple example of $X = \mathbb{C}^n$. Let $A \in \mathbb{C}^{n\times n}$ and $u^0 \in \mathbb{C}^n$ be given and consider the following system of ordinary differential equations:

$$\dot{u}(t) = Au(t), \qquad u(0) = u^0. \tag{1.24}$$

It is well-known that its solution $u : [0,\infty) \to \mathbb{C}^n$ can be written as

$$u(t) = e^{tA}u^0,$$

where e^{tA} is the **exponential of the matrix** tA. This exponential itself is again a $n \times n$ matrix, or in other words a linear operator from $\mathbb{C}^n$ to $\mathbb{C}^n$. Even more, it is easy to see that $(e^{tA})_{t\geq 0}$ is a strongly continuous semigroup. The semigroup and A are directly linked via

$$A = \left(\frac{d}{dt}e^{tA}\right)\Big|_{t=0}. \tag{1.25}$$

One can easily prove that for an arbitrary matrix (1.23) holds with $M = 1$ and $\omega = \|A\|$. For our purposes, we are mostly interested in matrices with a **field of values**

$$\mathcal{F}(A) = \left\{ \frac{x^*Ax}{x^*x} \;\middle|\; x \in \mathbb{C}^n \setminus \{0\} \right\} \tag{1.26}$$

contained in $\mathbb{C}^- = \{z \in \mathbb{C} \mid \operatorname{Re} z \leq 0\}$. Then e^{tA} is a contraction semigroup. For matrices with $\mathcal{F}(A) \subset i\mathbb{R}$, e.g. skew-hermitian matrices, the matrix exponential e^{tA} is unitary and thus satisfies $\|e^{tA}\| = 1$. The latter two properties can be shown by considering the ODE (1.24). ⋄

We generalize (1.25) by associating an operator $\mathcal{A}$ to a generic C_0-semigroup $(T(t))_{t\geq 0}$.

Definition 1.5. *Let $(T(t))_{t\geq 0}$ be a C_0-semigroup. We define the linear operator $\mathcal{A} : D(\mathcal{A}) \to X$ by*

$$\mathcal{A}x = \lim_{t\to 0+} \frac{T(t)x - x}{t}, \tag{1.27}$$

*where the **domain** $D(\mathcal{A})$ consists of all $x \in X$ for which the limit in (1.27) exists.*

*We call $\mathcal{A}$ the **infinitesimal generator** of the strongly continuous semigroup $(T(t))_{t\geq 0}$.*

The next lemma shows that for every $x \in D(\mathcal{A})$ the function $t \mapsto T(t)x$ is differentiable.

Lemma 1.6. *Let $(T(t))_{t\geq 0}$ be a C_0-semigroup with infinitesimal generator $\mathcal{A}$. Then, the following results hold:*

(a) *For $x \in D(\mathcal{A})$ and $t \geq 0$ we have $T(t)x \in D(\mathcal{A})$.*

(b) *For all $x \in D(\mathcal{A})$ and all $t \geq 0$ we have the relation*

$$\frac{d}{dt}(T(t)x) = \mathcal{A}T(t)x = T(t)\mathcal{A}x. \tag{1.28}$$

(c) *The domain of $\mathcal{A}$ is dense in X and $\mathcal{A}$ is a closed operator.*

Definition 1.5 implies that every C_0-semigroup has a **unique generator**. The following corollary of Lemma 1.6 shows the converse, namely that every generator belongs to a **unique semigroup**.

Corollary 1.7. *Let $(T_1(t))_{t\geq 0}$ and $(T_2(t))_{t\geq 0}$ be two C_0-semigroups with generators $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively. If $\mathcal{A}_1 = \mathcal{A}_2$, then $T_1(t) = T_2(t)$ for all $t \geq 0$.*

Moreover, Lemma 1.6 enables us to link a strongly continuous semigroup to the **abstract Cauchy problem**

$$\partial_t u(t) = \mathcal{A}u(t), \qquad u(0) = u^0. \tag{1.29}$$

Theorem 1.8. *Let $\mathcal{A}$ be the infinitesimal generator of the strongly continuous semigroup $(T(t))_{t\geq 0}$. Then, for every $u^0 \in D(\mathcal{A})$ the abstract Cauchy problem (1.29) has the unique solution $u(t) = T(t)u^0 \in C^1(\mathbb{R}_+; X) \cap C(\mathbb{R}_+; D(\mathcal{A}))$.*

In Example 1.4 we saw that the semigroup belonging to a matrix A can be written in the form of a matrix exponential. We adopt this notation also for an unbounded operator $\mathcal{A}$ by writing $e^{t\mathcal{A}}x$ instead of $T(t)x$ if $\mathcal{A}$ generates the C_0-semigroup $T(t)$.

Having established the correspondence between ODEs and the abstract Cauchy problem (1.29) we can carry over many concepts from the ODE case to general Cauchy problems. For instance, the variation of constants formula is also valid in the more general situation. More precisely, for the inhomogeneous abstract Cauchy problem

$$\partial_t u(t) = \mathcal{A}u(t) + f(t), \qquad u(0) = u^0, \tag{1.30}$$

the following result holds true.

Theorem 1.9. *Let $\mathcal{A}$ be the infinitesimal generator of the strongly continuous semigroup $\left(e^{t\mathcal{A}}\right)_{t\geq 0}$ and $u^0 \in D(\mathcal{A})$. Moreover, assume that either $f \in C^1(0,T;X)$ or that $f \in C(0,T;D(\mathcal{A}))$. Then, there exists a unique solution $u \in C^1(0,T;X) \cap C(0,T;D(\mathcal{A}))$ of (1.30) given by*

$$u(t) = e^{t\mathcal{A}}u^0 + \int_0^t e^{(t-s)\mathcal{A}} f(s)\, ds.$$

Next, we give two sufficient conditions for an operator $\mathcal{A}$ to generate a C_0-semigroup (Theorem 1.12) or a C_0-group (Theorem 1.17), respectively.

Definition 1.10. *A linear operator $\mathcal{A}$ on a Hilbert space $\left(X, (\cdot,\cdot)_X\right)$ is called **dissipative** if for every $x \in D(\mathcal{A})$ we have that*

$$\operatorname{Re}(\mathcal{A}x, x)_X \leq 0.$$

Example 1.11. A matrix $A \in \mathbb{C}^{n\times n}$ whose field of values is contained in the left complex half-plane, $\mathcal{F}(A) \subset \mathbb{C}^-$, is dissipative. In fact, every skew-hermitian matrix is dissipative. ⋄

We note that the concept of dissipative operators, like most of the considerations above, can be carried out also in Banach spaces, see [Pazy, 1983, Section 1.4], [Engel and Nagel, 2000, Chapter IIb.]. Moreover, the famous **Lumer–Phillips Theorem** [Engel and Nagel, 2000, Theorem II.3.15] holds true in this setting. We give its statement for the simpler case of Hilbert spaces, see [Jacob and Zwart, 2012, Theorem 6.1.7] and also [Engel and Nagel, 2000, Corollary II.3.20].

Theorem 1.12. *Let $\mathcal{A}$ be a linear operator with domain $D(\mathcal{A})$ on a Hilbert space X. Then, the following statements are equivalent:*

(a) *$\mathcal{A}$ is densely defined and generates a contraction semigroup.*

(b) *$\mathcal{A}$ is dissipative and* $\operatorname{ran}(\lambda - \mathcal{A}) = X$ *for some* $\lambda > 0$.

For the condition that $\mathcal{A}$ generates a C_0-group we first have to introduce the notion of the adjoint operator.

Definition 1.13. *Let $\mathcal{A} : D(\mathcal{A}) \to X$ be a linear operator with dense domain $\overline{D(\mathcal{A})} = X$. The **adjoint operator** $\mathcal{A}^*$ of $\mathcal{A}$ is defined as follows. The domain $D(\mathcal{A}^*)$ consists of all $y \in X$ such that there exists a $z \in X$ satisfying*

$$\big(\mathcal{A}x, y\big)_X = \big(x, z\big)_X \qquad \text{for all } x \in D(\mathcal{A}).$$

For $y \in D(\mathcal{A}^*)$, the adjoint is defined as $\mathcal{A}^* y = z$.

Note that for a bounded operator $\mathcal{A} \in \mathcal{L}(X)$ the definition of the adjoint simplifies significantly, since in this case $D(\mathcal{A}) = D(\mathcal{A}^*) = X$. Then, the adjoint is given by $\mathcal{A}^* : X \to X$,

$$\big(\mathcal{A}x, y\big)_X = \big(x, \mathcal{A}^* y\big)_X \qquad \text{for all } x, y \in X.$$

Definition 1.14. *Let $\mathcal{A} : D(\mathcal{A}) \to X$ be densely defined. The operator $\mathcal{A}$ is called*

(a) ***symmetric** if $\mathcal{A}x = \mathcal{A}^* x$ for all $x \in D(\mathcal{A}) \subset D(\mathcal{A}^*)$,*

(b) ***skew-symmetric** if $\mathcal{A}x = -\mathcal{A}^* x$ for all $x \in D(\mathcal{A}) \subset D(A^*)$,*

(c) ***self-adjoint** if $\mathcal{A} = \mathcal{A}^*$, i.e. if $\mathcal{A}$ is symmetric and $D(\mathcal{A}) = D(\mathcal{A}^*)$,*

(d) ***skew-adjoint** if $\mathcal{A}^* = -\mathcal{A}$, i.e. if $\mathcal{A}$ is skew-symmetric and $D(\mathcal{A}) = D(\mathcal{A}^*)$.*

Remark. Note that by the previous definition, a (skew-) hermitian matrix $A \in \mathbb{C}^{n\times n}$ represents a (skew-) symmetric linear operator $\mathcal{A} : \mathbb{C}^n \to \mathbb{C}^n$ and vice versa.

The following lemma provides a useful criterion to decide whether a skew-symmetric operator is also skew-adjoint.

Lemma 1.15. *Let $\mathcal{A} : D(\mathcal{A}) \to X$ be skew-symmetric. Then, $\mathcal{A}$ is skew-adjoint if $\mathcal{I} \pm \mathcal{A}$ has dense range, i.e. if*

$$\overline{\operatorname{ran}(\mathcal{I} \pm \mathcal{A})} = X.$$

Definition 1.16. *A C_0-group $(T(t))_{t\in\mathbb{R}}$ is called a **unitary group** if*

$$\|T(t)x\|_X = \|x\|_X \qquad \text{for all } x \in X, t \in \mathbb{R}.$$

Eventually, we can state the announced condition for C_0-groups. This theorem can be found in [Engel and Nagel, 2000, Theoremm II.3.24].

Theorem 1.17 (Stone's Theorem). *Let $\mathcal{A} : D(\mathcal{A}) \to X$ be a linear operator with dense domain $\overline{D(\mathcal{A})} = X$. Then, the following statements are equivalent:*

(a) *$\mathcal{A}$ generates a unitary C_0-group $(T(t))_{t\in\mathbb{R}}$ on X.*

(b) *$\mathcal{A}$ is skew-adjoint.*

1.3.2 Application to Maxwell's equations

In this section we apply the previously obtained results to Maxwell's equations. For this purpose, we first provide an appropriate framework in which Maxwell's equations fit in and in which the previous results are applicable. We start by introducing abbreviations for inner products that we will use throughout the thesis.

Functional analytic setting

For a set $K \subset \Omega$ and vector fields $\mathbf{U}, \widehat{\mathbf{U}}, \mathbf{V}, \widehat{\mathbf{V}} : K \to \mathbb{R}^3$ we denote the $L^2(K)$-inner product by

$$\left(\mathbf{U}, \widehat{\mathbf{U}}\right)_K = \int_K \mathbf{U} \cdot \widehat{\mathbf{U}} \, dx, \tag{1.31}$$

and for $F \subset \partial K$ we write

$$\left(\mathbf{U}, \widehat{\mathbf{U}}\right)_F = \int_F \mathbf{U}|_F \cdot \widehat{\mathbf{U}}|_F \, d\sigma. \tag{1.32}$$

Let $\mathbf{u} = (\mathbf{U}, \mathbf{V})$ and $\widehat{\mathbf{u}} = (\widehat{\mathbf{U}}, \widehat{\mathbf{V}})$. Given uniformly positive weight functions $\omega_1, \omega_2 : \Omega \to \mathbb{R}_{>0}$ we write the weighted inner products as

$$\left(\mathbf{U}, \widehat{\mathbf{U}}\right)_{\omega_1, K} = \left(\omega_1 \mathbf{U}, \widehat{\mathbf{U}}\right)_K, \qquad \left(\mathbf{u}, \widehat{\mathbf{u}}\right)_{\omega_1 \times \omega_2, K} = \left(\mathbf{U}, \widehat{\mathbf{U}}\right)_{\omega_1, K} + \left(\mathbf{V}, \widehat{\mathbf{V}}\right)_{\omega_2, K}. \tag{1.33}$$

By $\|\cdot\|_{\omega_1}$ and $\|\cdot\|_{\omega_1 \times \omega_2}$ we denote the corresponding norms. We abbreviate $\left(\cdot, \cdot\right) = \left(\cdot, \cdot\right)_\Omega$ and $\|\cdot\| = \|\cdot\|_\Omega$ and analogously for the weighted inner products and norms.

We want to analyze Maxwell's equations (1.18) in the state space $L^2(\Omega)^6$. This requires to clarify what we mean by writing $\operatorname{curl} \mathbf{U}$, since in general functions $\mathbf{U} \in L^2(\Omega)^3$ are not differentiable (and thus do not possess a "classical curl"). In the following we denote by $C^k(\Omega)$ the space of k times differentiable functions in Ω and by $C^k(\overline{\Omega})$ the space of k times differentiable functions in $\Omega \cup \partial\Omega$. Furthermore, we write

$$C_0^\infty(\Omega) = \{v \in C^\infty(\Omega) \mid \operatorname{supp}(v) \subset \Omega \text{ is compact}\}.$$

Note that the space $C_0^\infty(\Omega)$ (and also $C(\overline{\Omega})$) is dense in $L^2(\Omega)$ if the boundary is smooth enough, e.g. if it satisfies the segment condition, see [Adams and Fournier, 2008, Chapter 3, page 68]. For our purpose we do not need differentiability of $\mathbf{U}$ but it is sufficient that we have $\operatorname{curl} \mathbf{U} \in L^2(\Omega)^3$. This statement means that the functional

$$\ell_{\mathbf{U}} : C_0^\infty(\Omega)^3 \to \mathbb{R}, \qquad \ell_{\mathbf{U}}(\varphi) = \int_\Omega \mathbf{U} \cdot \operatorname{curl} \varphi \, dx,$$

is bounded in $L^2(\Omega)^3$, i.e., there is a constant $C_{\mathbf{U}}$ such that

$$|\ell_{\mathbf{U}}(\varphi)| \leq C_{\mathbf{U}} \|\varphi\|, \qquad \text{for all } \varphi \in C_0^\infty(\Omega)^3.$$

Then, by the Riesz representation theorem there is a unique $\mathbf{V} \in L^2(\Omega)^3$ such that

$$\ell_{\mathbf{U}}(\varphi) = \int_\Omega \mathbf{V} \cdot \varphi \, dx, \qquad \text{for all } \varphi \in C_0^\infty(\Omega)^3.$$

This $\mathbf{V}$ is called the variational curl of $\mathbf{U}$ and we denote it (for the moment) by $\widehat{\operatorname{curl}}\, \mathbf{U}$. In Definition 1.19 we fix this concept. Before, we show that for smooth functions the classical curl operator equals the variational curl.

Example 1.18. Consider $\mathbf{U} \in C^1(\Omega)^3$ and $\varphi \in C_0^\infty(\Omega)^3$. Applying integration by parts we obtain

$$\ell_{\mathbf{U}}(\varphi) = \int_\Omega \operatorname{curl} \mathbf{U} \cdot \varphi \, dx,$$

where the boundary term vanishes due to $\varphi|_{\partial\Omega} = 0$. By the Cauchy–Schwarz inequality we obtain $|\ell_{\mathbf{U}}(\varphi)| \leq C_{\mathbf{U}} \|\varphi\|$ with $C_{\mathbf{U}} = \|\operatorname{curl} \mathbf{U}\|_{L^2(\Omega)^3}$. As above, this means that there is a unique $\mathbf{V} \in L^2(\Omega)^3$ such that $\widehat{\operatorname{curl}}\, \mathbf{U} = \mathbf{V}$ and

$$\int_\Omega \widehat{\operatorname{curl}}\, \mathbf{U} \cdot \varphi \, dx = \int_\Omega \operatorname{curl} \mathbf{U} \cdot \varphi \, dx, \qquad \text{for all } \varphi \in C_0^\infty(\Omega)^3.$$

Since $C_0^\infty(\Omega)^3$ is dense in $L^2(\Omega)^3$ we can conclude that for $\mathbf{U} \in C^1(\Omega)^3$ we have that $\widehat{\operatorname{curl}}\, \mathbf{U} = \operatorname{curl} \mathbf{U}$ (in $L^2(\Omega)^3$). This motivates to use the notation curl also for the variational curl in the following definition.

The same holds true for $\mathbf{U} \in H^1(\Omega)^3$, if the partial derivatives in the definition of the curl are replaced by weak derivatives. Here, we have $C_{\mathbf{U}} = 2\,|\mathbf{U}|_{H^1(\Omega)^3}$.

Definition 1.19. *A function $\mathbf{U} \in L^2(\Omega)^3$ possesses a **variational curl** if there exists $\mathbf{V} \in L^2(\Omega)^3$ such that*

$$\int_\Omega \mathbf{U} \cdot \operatorname{curl} \varphi \, dx = \int_\Omega \mathbf{V} \cdot \varphi \, dx \qquad \text{for all } \varphi \in C_0^\infty(\Omega)^3. \tag{1.34}$$

In this case we write $\operatorname{curl} \mathbf{U} = \mathbf{V}$.

In the following, $\operatorname{curl} \mathbf{U}$ always denotes the variational curl of $\mathbf{U}$. We consider the subspace of $L^2(\Omega)^3$ functions which possess a variational curl.

Definition 1.20. *The **graph space of the curl operator** is given by*

$$H(\operatorname{curl}, \Omega) = \big\{ \mathbf{U} \in L^2(\Omega)^3 \;\big|\; \operatorname{curl} \mathbf{U} \in L^2(\Omega)^3 \big\}. \tag{1.35a}$$

We endow this space with the inner product

$$\big(\mathbf{U}, \mathbf{V}\big)_{H(\operatorname{curl},\Omega)} = \big(\mathbf{U}, \mathbf{V}\big) + \big(\operatorname{curl} \mathbf{U}, \operatorname{curl} \mathbf{V}\big), \qquad \text{for all } \mathbf{U}, \mathbf{V} \in H(\operatorname{curl}, \Omega), \tag{1.35b}$$

and the associated norm given by $\|\mathbf{U}\|^2_{H(\operatorname{curl},\Omega)} = \big(\mathbf{U}, \mathbf{U}\big)_{H(\operatorname{curl},\Omega)}$.

Let us compare $H(\operatorname{curl}, \Omega)$ with the standard Sobolev space $H^1(\Omega)$. While the former space is vector valued, the latter consists of scalar valued functions. Nevertheless, these spaces share some similarities. Either space consists of L^2-functions such that the associated functionals remain bounded in L^2. In fact, a function $u \in L^2(\Omega)$ possess a variational gradient if $\ell_u(\varphi) = -\int_\Omega u \operatorname{grad} \varphi \, dx$ can be bounded in $L^2(\Omega)^3$ for all $\varphi \in C_0^\infty(\Omega)$, see [Kirsch and Hettlich, 2015, Definition 4.1].

The space $H^1(\Omega)$ has among others the following three important properties. It is a Hilbert space, it can be defined as the closure of $C^\infty(\overline{\Omega})$ (or $C^1(\overline{\Omega})$) with respect to its graph norm, i.e. w.r.t. the $H^1(\Omega)$-norm, and there is an integration by parts formula. Analog properties also hold for the space $H(\operatorname{curl}, \Omega)$.

Theorem 1.21. *Let $\Omega \subset \mathbb{R}^3$ be a bounded, simply connected Lipschitz domain.*

(a) *The space $H(\operatorname{curl}, \Omega)$ is a Hilbert space.*

(b) *The space $H(\operatorname{curl}, \Omega)$ is the closure of $C^\infty(\overline{\Omega})^3$ with respect to $\|\cdot\|_{H(\operatorname{curl},\Omega)}$.*

(c) *For* $\mathbf{U} \in H^1(\Omega)^3 \subset H(\operatorname{curl}, \Omega)$ *we have*

$$\big(\operatorname{curl} \mathbf{U}, \varphi\big) = \big(\mathbf{U}, \operatorname{curl} \varphi\big) + \big(n \times \mathbf{U}, \varphi\big)_{\partial\Omega}, \qquad \text{for all } \varphi \in C^\infty(\overline{\Omega})^3. \tag{1.36}$$

Proof. For part (a) we refer to [Kirsch and Hettlich, 2015, Section 4.1.2], parts (b) and (c) are shown in [Monk, 2003, Theorem 3.26] and [Monk, 2003, Cororally 3.20], respectively. □

Remark 1.22. In general, functions in $H(\operatorname{curl}, \Omega)$ do not admit a trace in $L^2(\partial\Omega)^3$ but only in $H^{-1/2}(\partial\Omega)^3$. Part (c) of Theorem 1.21 can be extended to the case $\mathbf{U} \in H(\operatorname{curl}, \Omega)$, see [Monk, 2003, Theorem 3.29], but the integration by parts formula (1.36) is sufficient for this thesis and we omit these details.

For the boundary condition we recall once more standard Sobolev spaces, where the space $H_0^1(\Omega)$ is defined as the closure of $C_0^\infty(\Omega)$ with respect to the $H^1(\Omega)$-norm. This motivates the following definition.

Definition 1.23. *The space* $H_0(\operatorname{curl}, \Omega)$ *is defined as the closure of* $C_0^\infty(\Omega)^3$ *with respect to the norm* $\|\cdot\|_{H(\operatorname{curl},\Omega)}$.

We illustrate the meaning of Definition 1.23 by considering the space $H_0(\operatorname{curl}, \Omega) \cap H^1(\Omega)^3$ which admits traces in $L^2(\partial\Omega)^3$. However, we point out that the following results hold also true without the assumption that $\mathbf{U} \in H^1(\Omega)^3$, cf. [Monk, 2003, Section 3.5.3].

Owing to Definition 1.23, for every $\mathbf{U} \in H_0(\operatorname{curl}, \Omega) \cap H^1(\Omega)^3$ there is a sequence $(\mathbf{U}_k)_k \subset C_0^\infty(\Omega)^3$ such that $\mathbf{U}_k \to \mathbf{U}$ w.r.t. $\|\cdot\|_{H(\operatorname{curl},\Omega)}$ as $k \to \infty$. Hence, $\mathbf{U}_k \to \mathbf{U}$ and $\operatorname{curl} \mathbf{U}_k \to \operatorname{curl} \mathbf{U}$ w.r.t. $\|\cdot\|$. Applying integration by parts we infer

$$\big(\operatorname{curl} \mathbf{U}_k, \varphi\big) = \big(\mathbf{U}_k, \operatorname{curl} \varphi\big), \qquad \text{for all } \varphi \in C^\infty(\overline{\Omega})^3,$$

where the boundary term vanishes due to $\mathbf{U}_k|_{\partial\Omega} = 0$. Taking the limit $k \to \infty$ we obtain

$$\big(\operatorname{curl} \mathbf{U}, \varphi\big) = \big(\mathbf{U}, \operatorname{curl} \varphi\big), \qquad \text{for all } \varphi \in C^\infty(\overline{\Omega})^3. \tag{1.37}$$

Since $H_0(\operatorname{curl}, \Omega)$ is a subspace of $H(\operatorname{curl}, \Omega)$, Theorem 1.21 is applicable and we deduce by comparing (1.36) with (1.37) that

$$\big(n \times \mathbf{U}, \varphi\big)_{\partial\Omega} = 0 \qquad \text{for all } \varphi \in C^\infty(\overline{\Omega})^3.$$

This means that the space $H_0(\operatorname{curl}, \Omega) \cap H^1(\Omega)^3$ only contains functions $\mathbf{U}$ with vanishing tangential components on the boundary,

$$(n \times \mathbf{U})|_{\partial\Omega} = 0, \qquad \text{for all } \mathbf{U} \in H_0(\operatorname{curl}, \Omega) \cap H^1(\Omega)^3. \tag{1.38}$$

The converse is true as well, i.e. if a function $\mathbf{U} \in H(\operatorname{curl}, \Omega) \cap H^1(\Omega)^3$ satisfies (1.37), then we have that $\mathbf{U} \in H_0(\operatorname{curl}, \Omega)$, see [Monk, 2003, Lemma 3.27, Theorem 3.33]. The following lemma can be concluded from this.

Lemma 1.24. *If* $\mathbf{H} \in H(\operatorname{curl}, \Omega)$ *and* $\mathbf{E} \in H_0(\operatorname{curl}, \Omega)$. *Then, we have*

$$\big(\operatorname{curl} \mathbf{H}, \mathbf{E}\big) = \big(\mathbf{H}, \operatorname{curl} \mathbf{E}\big). \tag{1.39}$$

Proof of the well-posedness of linear Maxwell's equations

We can now compose the different parts together to show the well-posedness of linear Maxwell's equations (1.18) with perfectly conducting electric boundary conditions. In order to apply the semigroup theory of Section 1.3.1, we consider the Cauchy problem formulation (1.21) for $X = L^2(\Omega)^3 \times L^2(\Omega)^3 = L^2(\Omega)^6$ with the weighted inner product $\big(\cdot,\cdot\big)_{\mu\times\varepsilon}$.

Theorem 1.25. *The Maxwell operator $\mathcal{C}$ defined in (1.22) with domain*

$$D(\mathcal{C}) = D(\mathcal{C}_{\mathbf{H}}) \times D(\mathcal{C}_{\mathbf{E}}) = H(\operatorname{curl},\Omega) \times H_0(\operatorname{curl},\Omega) \tag{1.40}$$

generates a unitary C_0-group $e^{t\mathcal{C}}$.

Proof. The concept of this proof is taken from [Hochbruck et al., 2015a, Proposition 3.1]. We prove the assertion via Stone's theorem (Theorem 1.17), i.e. we show that $\mathcal{C}$ is skew-adjoint. We begin by observing that due to Lemma 1.24 the Maxwell operator $\mathcal{C}$ is skew-symmetric w.r.t. the weighted inner-product $\big(\cdot,\cdot\big)_{\mu\times\varepsilon}$, i.e.,

$$\big(\mathcal{C}\mathbf{u},\widehat{\mathbf{u}}\big)_{\mu\times\varepsilon} = -\big(\mathbf{u},\mathcal{C}\widehat{\mathbf{u}}\big)_{\mu\times\varepsilon}\,, \qquad \text{for all } \mathbf{u},\widehat{\mathbf{u}} \in D(\mathcal{C}). \tag{1.41}$$

In order to prove that $\mathcal{C}$ is skew-adjoint we apply Lemma 1.15. Hence, we have to show that

$$\overline{\operatorname{ran}(\mathcal{I} \pm \mathcal{C})} = L^2(\Omega)^6. \tag{1.42}$$

Because $C_0^\infty(\Omega)^6$ is dense in $L^2(\Omega)^6$ we infer that (1.42) is equivalent to show that for every $\mathbf{f} = (\mathbf{F},\mathbf{G}) \in C_0^\infty(\Omega)^6$ there is a $\mathbf{u} = (\mathbf{H},\mathbf{E}) \in D(\mathcal{C})$ such that

$$(\mathcal{I} \pm \mathcal{C})\mathbf{u} = \mathbf{f}, \tag{1.43a}$$

or, equivalently,

$$\mathbf{H} \mp \mu^{-1}\operatorname{curl}\mathbf{E} = \mathbf{F}, \tag{1.43b}$$
$$\mathbf{E} \pm \varepsilon^{-1}\operatorname{curl}\mathbf{H} = \mathbf{G}. \tag{1.43c}$$

Formally inserting $\mathbf{H}$ from (1.43b) into (1.43c) yields

$$\varepsilon\mathbf{E} + \operatorname{curl}(\mu^{-1}\operatorname{curl}\mathbf{E}) = \varepsilon\mathbf{G} \mp \operatorname{curl}\mathbf{F} := \widehat{\mathbf{G}} \in L^2(\Omega)^3. \tag{1.44}$$

In order to solve this problem we consider the bilinear form

$$a(\mathbf{E},\varphi) = \int_\Omega \varepsilon\mathbf{E}\cdot\varphi + \mu^{-1}\operatorname{curl}\mathbf{E}\cdot\operatorname{curl}\varphi\;dx, \qquad \mathbf{E},\varphi \in H_0(\operatorname{curl},\Omega).$$

Clearly, a is symmetric. Moreover, by using the Cauchy–Schwarz inequality (A.5) we infer

$$\begin{aligned}|a(\mathbf{E},\varphi)| &\le \left(\int_\Omega \varepsilon|\mathbf{E}|^2 + \mu^{-1}|\operatorname{curl}\mathbf{E}|^2\;dx\right)^{1/2}\left(\int_\Omega \varepsilon|\varphi|^2 + \mu^{-1}|\operatorname{curl}\varphi|^2\;dx\right)^{1/2}\\ &\le \max\left(\|\varepsilon\|_{L^\infty(\Omega)},\delta^{-1}\right)\|\mathbf{E}\|_{H(\operatorname{curl},\Omega)}\|\varphi\|_{H(\operatorname{curl},\Omega)}.\end{aligned}$$

Hence, a is bounded. It is also coercive, since

$$\begin{aligned}a(\mathbf{E},\mathbf{E}) = \int_\Omega \varepsilon|\mathbf{E}|^2 + \mu^{-1}|\operatorname{curl}\mathbf{E}|^2\;dx &\ge \delta\|\mathbf{E}\|^2 + \|\mu\|^{-1}_{L^\infty(\Omega)}\|\operatorname{curl}\mathbf{E}\|^2\\ &\ge \min\left(\delta,\|\mu\|^{-1}_{L^\infty(\Omega)}\right)\|\mathbf{E}\|^2_{H(\operatorname{curl},\Omega)}.\end{aligned}$$

As a consequence, the Lax–Milgram theorem, see e.g. [Di Pietro and Ern, 2012, Lemma 1.4], shows that there is a unique $\mathbf{E} \in H_0(\operatorname{curl}, \Omega)$ which satisfies

$$a(\mathbf{E}, \varphi) = \big(\widehat{\mathbf{G}}, \varphi\big), \qquad \text{for all } \varphi \in H_0(\operatorname{curl}, \Omega).$$

Furthermore, we have that

$$\big(\mu^{-1} \operatorname{curl} \mathbf{E}, \operatorname{curl} \varphi\big) = \big(\widehat{\mathbf{G}} - \varepsilon \mathbf{E}, \varphi\big), \qquad \text{for all } \varphi \in H_0(\operatorname{curl}, \Omega).$$

By Definition 1.19 we deduce that $\mu^{-1} \operatorname{curl} \mathbf{E} \in H(\operatorname{curl}, \Omega)$ and thus $\mathbf{E}$ satisfies (1.44). If we now define $\mathbf{H} \in H(\operatorname{curl}, \Omega)$ by (1.43b) we obtain $\mathbf{u} = (\mathbf{H}, \mathbf{E}) \in D(\mathcal{C})$ which solves (1.43a) as asserted. □

Remark 1.26. The skew-adjointness of the Maxwell operator $\mathcal{C}$ can also be proven by showing that it is skew-symmetric and furthermore that $D(\mathcal{C}) = D(\mathcal{C}^*)$ holds.

As a direct consequence of Theorem 1.25 we obtain the well-posedness of Maxwell's equations.

Corollary 1.27. *Let* $\mathbf{u}^0 = (\mathbf{H}^0, \mathbf{E}^0) \in D(\mathcal{C})$ *and let* $\mathbf{j} = (0, -\varepsilon^{-1}\mathbf{J}) \in C^1(0, T; X)$ *or* $\mathbf{j} \in C(0, T; D(\mathcal{C}))$. *Then, the linear Maxwell's equations* (1.21) *have a unique solution* $\mathbf{u}(t) = (\mathbf{H}(t), \mathbf{E}(t))$ *in* $C^1(0, T; X) \cap C(0, T; D(\mathcal{C}))$ *given by*

$$\mathbf{u}(t) = e^{t\mathcal{C}}\mathbf{u}^0 + \int_0^t e^{(t-s)\mathcal{C}}\mathbf{j}(s)\, ds. \tag{1.45}$$

Proof. The statement follows from Theorem 1.9. □

Remark 1.28. It is possible to incorporate the divergence conditions and the boundary condition on the magnetic field (1.19) into the domain of the Maxwell operator $\mathcal{C}$. This enables to prove a well-posedness result (such as Corollary 1.27) for the whole Maxwell system (1.10)–(1.11), see [Hochbruck et al., 2015a, Prop. 3.5] and [Pažur, 2013, Theorems 3.4, 3.6].

1.3.3 Energy conservation and stability

The **electromagnetic energy** $\mathcal{E}$ is given by

$$\mathcal{E}(\mathbf{H}, \mathbf{E}) = \frac{1}{2}\big(\|\mathbf{H}\|_\mu^2 + \|\mathbf{E}\|_\varepsilon^2\big).$$

In the absence of sources, the solution of Maxwell's equations conserves the electromagnetic energy.

Corollary 1.29. *Let* $\mathbf{u}(t) = (\mathbf{H}(t), \mathbf{E}(t))$ *be the solution of Maxwell's equations* (1.21) *with* $\mathbf{j} \equiv 0$. *Then, for all* $t \geq 0$ *we have that*

$$\mathcal{E}(\mathbf{H}(t), \mathbf{E}(t)) = \mathcal{E}(\mathbf{H}^0, \mathbf{E}^0). \tag{1.46}$$

Proof. This result follows directly from Theorem 1.25, since $e^{t\mathcal{C}}$ is a unitary group. □

We conclude this chapter by giving two stability results for the solution of Maxwell's equations.

Corollary 1.30. *For the solution* $\mathbf{u}(t)$ *of* (1.21) *we have the following bounds:*

$$\|\mathbf{u}(t)\|_{\mu\times\varepsilon} \leq \|\mathbf{u}^0\|_{\mu\times\varepsilon} + \frac{1}{\sqrt{\delta}} \int_0^t \|\mathbf{J}(s)\|\, ds, \tag{1.47a}$$

$$\|\mathbf{u}(t)\|_{\mu\times\varepsilon}^2 \leq \mathrm{e}^1 \|\mathbf{u}^0\|_{\mu\times\varepsilon}^2 + \mathrm{e}^1 \frac{T+1}{\delta} \int_0^t \|\mathbf{J}(s)\|^2\, ds. \tag{1.47b}$$

Proof. Taking the norm of (1.45) and using the triangle inequality (A.4) we get

$$\|\mathbf{u}(t)\|_{\mu\times\varepsilon} \leq \|e^{t\mathcal{C}}\mathbf{u}^0\|_{\mu\times\varepsilon} + \int_0^t \|e^{(t-s)\mathcal{C}}\mathbf{j}(s)\|_{\mu\times\varepsilon}\, ds.$$

Since $e^{t\mathcal{C}}$ is unitary, cf. Theorem 1.25, we have $\|e^{t\mathcal{C}}\mathbf{u}\|_{\mu\times\varepsilon} = \|\mathbf{u}\|_{\mu\times\varepsilon}$ for all $\mathbf{u} \in L^2(\Omega)^6$. The bound (1.47a) follows from

$$\|\mathbf{j}\|_{\mu\times\varepsilon} = \|-\varepsilon^{-1}\mathbf{J}\|_{\varepsilon} = \|\varepsilon^{-1/2}\mathbf{J}\| \leq \delta^{-1/2}\|\mathbf{J}\|. \tag{1.48}$$

In order to prove (1.47b) we consider

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|\mathbf{u}(t)\|_{\mu\times\varepsilon}^2 &= \big(\mathbf{u}(t), \partial_t\mathbf{u}(t)\big)_{\mu\times\varepsilon} \\ &= \big(\mathbf{u}(t), \mathcal{C}\mathbf{u}(t)\big)_{\mu\times\varepsilon} + \big(\mathbf{u}(t), \mathbf{j}(t)\big)_{\mu\times\varepsilon} \\ &= \big(\mathbf{u}(t), \mathbf{j}(t)\big)_{\mu\times\varepsilon},\end{aligned}$$

where the second equality follows by (1.21) and the last equality holds since $\mathcal{C}$ is skew-symmetric, see (1.41). Applying the Cauchy–Schwarz inequality (A.5) and Young's inequality (A.2) we get

$$\frac{1}{2}\frac{d}{dt}\|\mathbf{u}(t)\|_{\mu\times\varepsilon}^2 \leq \|\mathbf{u}(t)\|_{\mu\times\varepsilon}\|\mathbf{j}(t)\|_{\mu\times\varepsilon} \leq \frac{T+1}{2}\|\mathbf{j}(t)\|_{\mu\times\varepsilon}^2 + \frac{1}{2(T+1)}\|\mathbf{u}(t)\|_{\mu\times\varepsilon}^2.$$

Integrating from 0 to t shows

$$\|\mathbf{u}(t)\|_{\mu\times\varepsilon}^2 \leq \|\mathbf{u}^0\|_{\mu\times\varepsilon}^2 + (T+1)\int_0^t \|\mathbf{j}(s)\|_{\mu\times\varepsilon}^2\, ds + \frac{1}{T+1}\int_0^t \|\mathbf{u}(s)\|_{\mu\times\varepsilon}^2\, ds.$$

Gronwall's lemma (Lemma A.1) yields

$$\|\mathbf{u}(t)\|_{\mu\times\varepsilon}^2 \leq \mathrm{e}^{\frac{t}{T+1}}\|\mathbf{u}^0\|_{\mu\times\varepsilon}^2 + \mathrm{e}^{\frac{t}{T+1}}(T+1)\int_0^t \|\mathbf{j}(s)\|_{\mu\times\varepsilon}^2\, ds.$$

The stated bound (1.47b) now follows from (1.48) and $t \leq T$. □

CHAPTER 2

Spatial discretization: discrete setting

The following two chapters are devoted to the spatial discretization of Maxwell's equations (1.18) by means of a discontinuous Galerkin (dG) method. In the current chapter we introduce the necessary discrete setting and provide essential tools which we will use frequently in this thesis. This chapter closely follows the concepts presented in the book of Di Pietro and Ern [2012].

First of all, let us note that the domain Ω can by approximated by a polyhedron. Because this can be done of arbitrary accuracy we neglect the error of this approximation in this thesis and henceforth assume the following simplification.

Assumption 2.1. *We assume that the domain Ω is a polyhedron in $\mathbb{R}^d$.*

This assumption enables us to cover the domain with a mesh consisting of polyhedral elements.

2.1 Meshes

Our first step is to discretize Ω using a mesh. The simplest choice is a simplicial mesh.

Definition 2.2. *Let $\{x_0, \dots, x_d\}$ be a set of $d+1$ points in $\mathbb{R}^d$ such that the vectors $x_1 - x_0, \dots, x_d - x_0$ are linearly independent. We call the interior of the convex hull of $\{x_0, \dots, x_d\}$ a **non-degenerate simplex** in $\mathbb{R}^d$.*

For $d = 1$ a non-degenerate simplex is an interval, for $d = 2$ a triangle and for $d = 3$ a tetrahedron.

Definition 2.3. *A finite set $\mathcal{T} = \{K\}$ is called a **simplicial mesh** of the domain Ω if it satisfies:*

(a) *Every $K \in \mathcal{T}$ is a non-degenerate simplex.*

(b) *$\mathcal{T}$ forms a partition of Ω, i.e. $\overline{\Omega} = \bigcup_{K \in \mathcal{T}} \overline{K}$ and $K \cap \widehat{K} = \emptyset$ for all $K, \widehat{K} \in \mathcal{T}, K \neq \widehat{K}$.*

*Each $K \in \mathcal{T}$ is called a **mesh element**.*

Note that a simplicial mesh is allowed to have hanging nodes. For (continuous) finite elements simplicial meshes without hangin nodes are a quite convenient choice. An advantage of dG methods is that they allow more easily to work with more general meshes.

Definition 2.4. *We call a finite set $\mathcal{T} = \{K\}$ of polyhedra K a **general mesh** of the domain Ω if it satisfies (b) of Definition 2.3. Each $K \in \mathcal{T}$ is called a **mesh element**.*

Clearly, a simplicial mesh is just a particular case of a general mesh.

Assumption 2.5. *We suppose that the coefficients μ and ε are **piecewise constant** and that the mesh $\mathcal{T}$ is **matched** to them such that $\mu|_K \equiv \mu_K$ and $\varepsilon|_K \equiv \varepsilon_K$ are constant for each $K \in \mathcal{T}$.*

Definition 2.6. *Let $\mathcal{T}$ be a mesh of Ω. For all $K \in \mathcal{T}$ we denote the **diameter** of K by h_K and the **radius of the largest ball** inscribed in K by r_K. Furthermore, we define the **meshsize** as*

$$h = \max_{K \in \mathcal{T}} h_K,$$

and use the notation $\mathcal{T}_h$ for a mesh with meshsize h.

Definition 2.7. *Let $\mathcal{T}_h$ be a mesh of Ω. We say that a closed subset F of $\overline{\Omega}$ is a **mesh face** if F has positive $(d-1)$-dimensional Hausdorff measure and if either one of the following two conditions is satisfied:*

(a) *There are distinct mesh elements $K, \widehat{K} \in \mathcal{T}_h$ such that $F = \partial K \cap \partial \widehat{K}$; in this case, we call F an **interface**.*

(b) *There is a mesh element $K \in \mathcal{T}_h$ such that $F = \partial K \cap \partial\Omega$; in this case, we call F a **boundary face**.*

The set of interfaces is denoted by $\mathcal{F}_h^{\text{int}}$ and the set of boundary faces by $\mathcal{F}_h^{\text{bnd}}$. With $\mathcal{F}_h = \mathcal{F}_h^{\text{int}} \cup \mathcal{F}_h^{\text{bnd}}$ we denote the set of all faces and

$$N_\partial = \max_{K \in \mathcal{T}_h} \operatorname{card}\{F \in \mathcal{F}_h \mid F \subset \partial K\}$$

*denotes the **maximum number of mesh faces** composing the boundary of a mesh element.*

For simplicial meshes we have $N_\partial = 2$ for $d = 1$, $N_\partial = 3$ for $d = 2$, and $N_\partial = 4$ for $d = 3$.

Definition 2.8. *Let $\mathcal{T}_h$ be a mesh of Ω. For all $K \in \mathcal{T}_h$ we define n_K a.e. on ∂K as the **unit outward normal** to K.*

*For every interface $F \in \mathcal{F}_h^{\text{int}}$ we choose arbitrarily one of the outer unit normals of the two mesh elements composing the face F. We fix this **face normal** and denote it with n_F. We use the notation K and K_F for two neighboring elements $\partial K \cap \partial K_F = F \in \mathcal{F}_h^{\text{int}}$, whereby the face normal n_F points from K to K_F. For a boundary face the orientation of n_F is always outwards.*

Figure 2.1 shows the face normal n_F and the associated elements K and K_F.

In dG methods we will consider functions $v : \Omega \to \mathbb{R}$ which are only piecewise smooth, i.e. smooth on every mesh element K but not on Ω (e.g. $v \in H^1(K)$ for all $K \in \mathcal{T}_h$ but $v \notin H^1(\Omega)$). The restriction of such a function to an element $v|_K$ admits a well-defined trace on ∂K. However, for all $F \in \mathcal{F}_h^{\text{int}}$, v has a (possibly) two-valued trace. Thus, the following concepts of the average and the jump of a function at an interface are essential for dG methods.

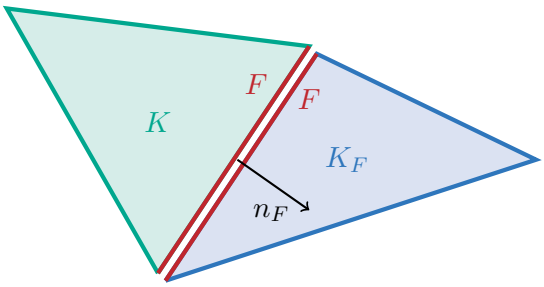


Figure 2.1: Convention for K, K_F.

Definition 2.9. *Let $v : \Omega \to \mathbb{R}$ be a function such that for every mesh element $K \in \mathcal{T}_h$ its restriction $v|_K$ admits a trace a.e. on the boundary ∂K. Then, for all interfaces $F \in \mathcal{F}_h^{\text{int}}$ we define the **jump** of v on F as*

$$[\![v]\!]_F = (v|_{K_F})|_F - (v|_K)|_F.$$

*Let $\omega : \Omega \to \mathbb{R}_+$ be a piecewise constant weight function, i.e. $\omega|_K \equiv \omega_K$ for all $K \in \mathcal{T}_h$. Then, we define the **weighted average** of v on F as*

$$\{\!\{v\}\!\}_F^\omega = \frac{\omega_K (v|_K)|_F + \omega_{K_F}(v|_{K_F})|_F}{\omega_K + \omega_{K_F}}.$$

For vector fields these operations act componentwise.

We abbreviate the average with $\omega \equiv 1$ by $\{\!\{v\}\!\}_F$. For later purpose we already state an important identity, which constitutes an essential trick in dG methods that we frequently will use.

Lemma 2.10. *Assume that the weight functions ω and $\overline{\omega}$ satisfy*

$$0 \neq \omega\overline{\omega} \equiv \text{const.} \tag{2.1a}$$

Then, for vector valued functions $\mathbf{U}, \mathbf{V} : \Omega \to \mathbb{R}^3$ we have that

$$[\![\mathbf{U} \cdot \mathbf{V}]\!]_F = \{\!\{\mathbf{U}\}\!\}_F^\omega \cdot [\![\mathbf{V}]\!]_F + [\![\mathbf{U}]\!]_F \cdot \{\!\{\mathbf{V}\}\!\}_F^{\overline{\omega}}. \tag{2.1b}$$

Proof. By (2.1a) we have

$$\begin{aligned} \omega_K\overline{\omega}_K = \omega_{K_F}\overline{\omega}_{K_F} \quad &\Longleftrightarrow \quad \frac{\omega_{K_F}}{\omega_K} = \frac{\overline{\omega}_K}{\overline{\omega}_{K_F}} \quad \Longleftrightarrow \quad \frac{1}{1 + \frac{\omega_{K_F}}{\omega_K}} = \frac{1}{\frac{\overline{\omega}_K}{\overline{\omega}_{K_F}} + 1} \\ &\Longleftrightarrow \quad \frac{\omega_K}{\omega_K + \omega_{K_F}} = \frac{\overline{\omega}_{K_F}}{\overline{\omega}_K + \overline{\omega}_{K_F}}. \end{aligned}$$

Using this, we obtain

$$\{\!\{\mathbf{U}\}\!\}_F^\omega \cdot [\![\mathbf{V}]\!]_F + [\![\mathbf{U}]\!]_F \cdot \{\!\{\mathbf{V}\}\!\}_F^{\overline{\omega}} = \mathbf{U}_{K_F} \cdot \mathbf{V}_{K_F} - \mathbf{U}_K \cdot \mathbf{V}_K = [\![\mathbf{U} \cdot \mathbf{V}]\!]_F,$$

which is the stated identity. □

We do not want to consider only a single approximation associated with a fixed grid $\mathcal{T}_h$, say $\mathbf{u}_h(t)$, to the exact solution $\mathbf{u}(t)$ of Maxwell's equations (1.21). Instead, we want to analyze how the quality (i.e. the error) of a sequence $(\mathbf{u}_h(t))_h$ improves when the associated meshes $(\mathcal{T}_h)_h$ consist of finer and finer elements. In other words, we want to analyze the convergence $\mathbf{u}_h(t) \to \mathbf{u}(t)$ when $h \searrow 0$. This requires that our meshes have a certain quality.

We consider a mesh sequence

$$\mathcal{T}_\mathcal{H} = (\mathcal{T}_h)_{h \in \mathcal{H}}$$

where $\mathcal{H}$ is a countable subset of $\mathbb{R}_+$ having 0 as only accumulation point.

Definition 2.11. *We call $\mathcal{T}_h$ a **matching simplicial mesh** if it is a simplicial mesh and if for every $K \in \mathcal{T}_h$ with vertices $\{x_0, \dots, x_d\}$, the set $\partial K \cap \partial \widehat{K}$, $\widehat{K} \in \mathcal{T}_h$, is the convex hull of a (possibly empty) subset of $\{x_0, \dots, x_d\}$.*

In $\mathbb{R}^2$ the set $\partial K \cap \partial \widehat{K}$ for two distinct elements of a matching simplicial mesh is either empty, or a common vertex, or a common edge of the two elements.

Definition 2.12. *Let $\mathcal{T}_h$ be a general mesh. We call $\mathcal{T}_h'$ a **matching simplicial submesh** if:*

(a) *$\mathcal{T}_h'$ is a matching simplicial mesh.*

(b) *For all $K' \in \mathcal{T}_h'$ there is only one $K \in \mathcal{T}_h$ such that $K' \subset K$.*

(c) *For all $F' \in \mathcal{F}_h'$, the set collecting the mesh faces on $\mathcal{T}_h'$, there is at most one $F \in \mathcal{F}_h$ such that $F' \subset F$.*

Definition 2.13. *Let $\mathcal{T}_\mathcal{H}$ be a mesh sequence which admits a matching simplicial submesh $\mathcal{T}_h'$ for all $h \in \mathcal{H}$.*

(a) *$\mathcal{T}_\mathcal{H}$ is **shape-regular** if there is $\rho_1 > 0$, independent of h, such that for all $K' \in \mathcal{T}_h'$ we have that*

$$h_{K'} \leq \rho_1 r_{K'}. \tag{2.2}$$

(b) *$\mathcal{T}_\mathcal{H}$ is **contact-regular** if there is $\rho_2 > 0$ such that for all $K \in \mathcal{T}_h$ and all $K' \in \mathcal{T}_h', K' \subset K$, we have that*

$$h_K \leq \rho_2 h_{K'}. \tag{2.3}$$

We denote the product of the mesh parameters ρ_1 and ρ_2 by

$$\rho = \rho_1 \rho_2.$$

If $\mathcal{T}_h$ is itself simplicial and matching, then $\mathcal{T}_h' = \mathcal{T}_h$, and thus $\rho_2 = 1$. So, in this case, one only has to require shape-regularity (2.2).

An important observation is that the number of faces of a shape- and contact-regular mesh sequence is bounded independently of the mesh parameter h.

Lemma 2.14. *Let $\mathcal{T}_\mathcal{H}$ be a shape- and contact-regular mesh sequence. Then, for all $h \in \mathcal{H}$, N_∂ is bounded uniformly in h. In fact, we have*

$$N_\partial \leq (d+1)\, |B_d|_d^{-1}\, \rho^d,$$

where $|\cdot|_d$ denotes the d-dimensional Hausdorff measure and B_d is the unit ball in $\mathbb{R}^d$.

Proof. We follow the proof in [Di Pietro and Ern, 2012, Lemmas 1.40, 1.41]. Let, for all $K \in \mathcal{T}_h$, the set $\mathcal{S}_K'$ collect the subelements $K' \in \mathcal{T}_h'$ composing the element K, i.e.

$$\mathcal{S}_K' = \{K' \in \mathcal{T}_h' \mid K' \subset K\}.$$

Then, we have

$$\begin{aligned} h_K^d \geq |K|_d = \sum_{K' \in \mathcal{S}_K'} |K'|_d \geq \sum_{K' \in \mathcal{S}_K'} |B_d|_d r_{K'}^d \geq \sum_{K' \in \mathcal{S}_K'} |B_d|_d \rho_1^{-d} h_{K'}^d \\ \geq \sum_{K' \in \mathcal{S}_K'} |B_d|_d \rho_1^{-d} \rho_2^{-d} h_K^d = \operatorname{card}(\mathcal{S}_K') |B_d|_d \rho^{-d} h_K^d. \end{aligned}$$

Here, we used (2.2) for the third inequality and (2.3) for the fourth inequality. This yields

$$\operatorname{card}(\mathcal{S}'_K) \le |B_d|_d^{-1}\rho^d.$$

The bound on N_∂ is seen from

$$\operatorname{card}\{F \in \mathcal{F}_h \mid F \subset \partial K\} \le \operatorname{card}\{F' \in \mathcal{F}'_h \mid F' \subset \partial K'\} = (d+1)\operatorname{card}(\mathcal{S}'_K),$$

since every simplex has $d+1$ faces. □

The next lemma gives a comparison of the diameters of neighboring elements.

Lemma 2.15. *Let $\mathcal{T}_\mathcal{H}$ be a shape- and contact-regular mesh sequence. Then, for all $h \in \mathcal{H}$ and all $K, \widehat{K} \in \mathcal{T}_h$ sharing a face F, we have that*

$$\max(h_K, h_{\widehat{K}}) \le \rho \min(h_K, h_{\widehat{K}}),$$

and

$$\rho^{-1}\max(h_K, h_{\widehat{K}}) \le \frac{h_K + h_{\widehat{K}}}{2} \le \rho \min(h_K, h_{\widehat{K}}). \tag{2.4}$$

Proof. We adapt the proof given in [Di Pietro and Ern, 2012, Lemmas 1.42, 1.43]. Let δ_F denote the diameter of F. Clearly, we have

$$\delta_F \le \min(h_K, h_{\widehat{K}}).$$

Let $\mathcal{T}'_h$ be a matching simplicial submesh of $\mathcal{T}_h$ and let $K', \widehat{K}' \in \mathcal{T}'_h$ such that

$$K' \subset K, \qquad \widehat{K}' \subset \widehat{K}, \qquad F' = \partial K' \cap \partial \widehat{K}' \subset F.$$

Then, we have

$$\delta_F \ge \delta_{F'} \ge \max(r_{K'}, r_{\widehat{K}'}) \ge \rho_1^{-1}\max(h_{K'}, h_{\widehat{K}'}) \ge \rho_1^{-1}\rho_2^{-1}\max(h_K, h_{\widehat{K}}),$$

where we applied (2.2), (2.3) for the last two estimates. This gives the first assertion which easily yields (2.4). □

2.2 Approximation spaces: Broken polynomial spaces

We want to approximate the exact solution $\mathbf{u}(t)$ in a finite dimensional function space consisting of piecewise polynomials, i.e. in a **broken polynomial space**.

2.2.1 The spaces $\mathbb{P}_d^k$ and $\mathbb{P}_d^k(\mathcal{T}_h)$

Let $k \in \mathbb{N}_0$ be an integer and $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ be a multi-index. For $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ we use the convention $x^\alpha = \prod_{i=1}^d x_i^{\alpha_i}$. We set

$$A_d^k = \{\alpha \in \mathbb{N}_0^d \mid |\alpha|_{\ell^1} = \sum_{i=1}^d \alpha_i \le k\}.$$

We define the space of polynomials in d variables and of total degree at most k as

$$\mathbb{P}_d^k = \left\{ p : \mathbb{R}^d \to \mathbb{R} \mid \exists (\gamma_\alpha)_{\alpha \in A_d^k} \in \mathbb{R}^{\operatorname{card}(A_d^k)} \text{ s.t. } p(x) = \sum_{\alpha \in A_d^k} \gamma_\alpha x^a \right\},$$

which is of dimension

$$\dim(\mathbb{P}_d^k) = \operatorname{card}(A_d^k) = \binom{k+d}{k} = \frac{(k+d)!}{k!d!}.$$

This follows since the cardinality of A_d^k is the number of multi-indices $\alpha \in \mathbb{N}_0^d$, that satisfy $\sum_{i=1}^d \alpha_i \leq k$. This can be equivalently expressed as

$$\alpha_1 + \cdots + \alpha_d + \alpha_{d+1} = k,$$

with a slack variable $\alpha_{d+1} \in \mathbb{N}_0$. A "stars and bars" argument then gives the result.

We define the **approximation space** (or **dG space**) as the broken polynomial space

$$\mathbb{P}_d^k(\mathcal{T}_h) = \left\{ v \in L^2(\Omega) \mid v|_K \in \mathbb{P}_d^k \text{ for all } K \in \mathcal{T}_h \right\}. \tag{2.5}$$

The space $\mathbb{P}_d^k(\mathcal{T}_h)$ consists of functions which are polynomials on each mesh element but which are allowed to be discontinuous across the mesh faces. $\mathbb{P}_d^k(\mathcal{T}_h)$ is a vector space with dimension

$$\dim(\mathbb{P}_d^k(\mathcal{T}_h)) = \operatorname{card}(\mathcal{T}_h) \cdot \dim(\mathbb{P}_d^k).$$

Definition 2.16. *The L^2-**orthogonal projection** onto $\mathbb{P}_d^k(\mathcal{T}_h)$, $\pi_h : L^2(\Omega) \to \mathbb{P}_d^k(\mathcal{T}_h)$, is defined such that for every $v \in L^2(\Omega)$,*

$$(v - \pi_h v, \varphi_h) = 0, \qquad \text{for all } \varphi_h \in \mathbb{P}_d^k(\mathcal{T}_h). \tag{2.6}$$

For vector fields $\mathbf{V} \in L^2(\Omega)^m$ the projection acts componentwise.

In our later dG discretization we will need L^2-projections which are orthogonal w.r.t. the weighted inner products $(\cdot,\cdot)_\mu$ and $(\cdot,\cdot)_\varepsilon$, respectively. The next lemma shows that, under Assumption 2.5, the L^2-orthogonal projection (2.6) satisfies this. Moreover, the lemma provides a bound on the projection operator.

Lemma 2.17. *For $\mathbf{V} \in L^2(\Omega)^3$ we have that*

$$(\mathbf{V} - \pi_h \mathbf{V}, \varphi_h)_\mu = (\mathbf{V} - \pi_h \mathbf{V}, \varphi_h)_\varepsilon = 0, \qquad \text{for all } \varphi_h \in \mathbb{P}_d^k(\mathcal{T}_h)^3. \tag{2.7}$$

Moreover, we have the following bounds

$$\|\pi_h \mathbf{V}\|_\mu \leq \|\mathbf{V}\|_\mu, \qquad \|\pi_h \mathbf{V}\|_\varepsilon \leq \|\mathbf{V}\|_\varepsilon. \tag{2.8}$$

Proof. For $\varphi_h \in \mathbb{P}_d^k(\mathcal{T}_h)^3$ the restriction $\varphi_h|_K$ only depends on the values of φ_h in K. So, we can deduce

$$(\mathbf{V} - \pi_h \mathbf{V}, \varphi_h)_K = 0, \qquad \text{for all } K \in \mathcal{T}_h,\ \varphi_h \in (\mathbb{P}_d^k)^3 \subset \mathbb{P}_d^k(\mathcal{T}_h)^3,$$

since by (2.6) this holds true for all $\varphi_h \in \mathbb{P}_d^k(\mathcal{T}_h)^3$ with $\varphi_h|_{\widehat{K}} \equiv 0$, $\widehat{K} \neq K$. So, for all $\varphi_h \in \mathbb{P}_d^k(\mathcal{T}_h)^3$ we have that

$$(\mathbf{V} - \pi_h \mathbf{V}, \varphi_h)_\mu = \sum_{K \in \mathcal{T}_h} (\mathbf{V} - \pi_h \mathbf{V}, \varphi_h)_{\mu,K} = \sum_{K \in \mathcal{T}_h} \mu_K (\mathbf{V} - \pi_h \mathbf{V}, \varphi_h)_K = 0,$$

which proves (2.7). The bounds (2.8) are obtained by

$$\|\pi_h \mathbf{V}\|_\mu = \sup_{\substack{\varphi_h \in \mathbb{P}_d^k(\mathcal{T}_h)^3 \\ \|\varphi_h\|_\mu = 1}} (\pi_h \mathbf{V}, \varphi_h)_\mu = \sup_{\substack{\varphi_h \in \mathbb{P}_d^k(\mathcal{T}_h)^3 \\ \|\varphi_h\|_\mu = 1}} (\mathbf{V}, \varphi_h)_\mu \leq \sup_{\substack{\varphi_h \in \mathbb{P}_d^k(\mathcal{T}_h)^3 \\ \|\varphi_h\|_\mu = 1}} \|\mathbf{V}\|_\mu \|\varphi_h\|_\mu = \|\mathbf{V}\|_\mu .$$

Replacing μ by ε shows the corresponding results for the weight ε. □

Alternatively, we could use orthogonal projections w.r.t. the weighted inner products, e.g. π_μ via

$$\left(\mathbf{H} - \pi_\mu \mathbf{H}, \varphi_h\right)_\mu = 0, \qquad \text{for all } \varphi_h \in \mathbb{P}_d^k(\mathcal{T}_h)^3.$$

Then, one can show that $\pi_h \mathbf{H} = \pi_\mu \mathbf{H}$ if the weight function satisfies Assumption 2.5.

Remark 2.18. It is possible to consider other broken polynomial spaces. An important example is the space of polynomials in d variables and of degree at most k in each variable,

$$\mathbb{Q}_d^k = \left\{ p : \mathbb{R}^d \to \mathbb{R} \mid \exists (\gamma_\alpha)_{\alpha \in B_d^k} \in \mathbb{R}^{\operatorname{card}(B_d^k)} \text{ s.t. } p(x) = \sum_{\alpha \in B_d^k} \gamma_\alpha x^a \right\},$$

where

$$B_d^k = \left\{ \alpha \in \mathbb{N}_0^d \mid \max_{i \in \{1,\dots,d\}} \alpha_i \leq k \right\}.$$

This space is used e.g. when working with hexahedra instead of tetrahedra.

2.2.2 Inverse and trace inequality

Next we study properties of $\mathbb{P}_d^k(\mathcal{T}_h)$, which are essential for proving error bounds.

Lemma 2.19 (**Inverse inequality**, cf. [Di Pietro and Ern, 2012, Lemma 1.44])**.** *Let $\mathcal{T}_\mathcal{H}$ be a shape- and contact-regular mesh sequence. Then, for all $h \in \mathcal{H}$, all $v_h \in \mathbb{P}_d^k(\mathcal{T}_h)$, and all $K \in \mathcal{T}_h$, we have that*

$$\| \operatorname{grad} v_h \|_K \leq C'_{\mathrm{inv}} h_K^{-1} \| v_h \|_K. \tag{2.9}$$

The constant C'_{inv} only depends on d, k, and the mesh regularity parameters ρ_1, ρ_2.

Clearly, under the assumptions of Lemma 2.19 we have for all $\mathbf{V}_h \in \mathbb{P}_d^k(\mathcal{T}_h)^3$ that

$$\| \operatorname{curl} \mathbf{V}_h \|_K \leq C_{\mathrm{inv}} h_K^{-1} \| \mathbf{V}_h \|_K, \tag{2.10}$$

where C_{inv} has the same dependences as C'_{inv}.

Lemma 2.20 (**Discrete trace inequality**, cf. [Di Pietro and Ern, 2012, Lemma 1.46])**.** *Let $\mathcal{T}_\mathcal{H}$ be a shape- and contact-regular mesh sequence. Then, for all $h \in \mathcal{H}$, all $v_h \in \mathbb{P}_d^k(\mathcal{T}_h)$, all $K \in \mathcal{T}_h$, and all $F \in \mathcal{F}_h$, $F \subset \partial K$, it holds*

$$\| v_h \|_F \leq C_{\mathrm{tr}} h_K^{-1/2} \| v_h \|_K. \tag{2.11}$$

The constant C_{tr} only depends on d, k, and the mesh regularity parameters ρ_1, ρ_2.

Remark 2.21. The constants C'_{inv} (and thus C_{inv}) and C_{tr} depend on the polynomial degree k. E.g. on triangles, C'_{inv} scales as k^2, whereas C_{tr} scales as $\sqrt{k(k+d)}$, see [Di Pietro and Ern, 2012, Remark 1.47].

2.2.3 Approximation properties

Recall that we are interested in approximating the exact solution $\mathbf{u}(t)$ of Maxwell's equations by a discrete function $\mathbf{u}_h(t)$ in the dG space $\mathbb{P}_d^k(\mathcal{T}_h)^6$. Consequently, the question arises which quality can be achieved by this approximation. It turns out that this depends on the mesh sequence we employ. For this thesis we will focus on mesh sequences which allow the optimal approximation [Di Pietro and Ern, 2012, Definition 1.55]. We will give error bounds in terms of the seminorm on $H^m(K)$, which we denote by $|\cdot|_{m,K} = |\cdot|_{H^m(K)}$.

Definition 2.22. *A mesh sequence $\mathcal{T}_\mathcal{H}$ has **optimal polynomial approximation properties** if, for all $h \in \mathcal{H}$, all $K \in \mathcal{T}_h$, and all polynomial degrees k, there is a linear interpolation operator $\mathcal{I}_K^k : L^2(K) \to \mathbb{P}_d^k(K)$ such that, for all $s \in \{0, \dots, k+1\}$ and all $v \in H^s(K)$, we have that*

$$|v - \mathcal{I}_K^k v|_{m,K} \leq C'_{\mathrm{app}} h_K^{s-m} |v|_{s,K} , \qquad \text{for all } m \in \{0, \dots, s\},$$

with a constant C'_{app} that is independent of both K and h.

This allows to define the following class of mesh sequences.

Definition 2.23. *A shape- and contact-regular mesh sequence $\mathcal{T}_\mathcal{H}$ with optimal polynomial approximation properties is called an **admissible mesh sequence**.*

An important example is that of shape- and contact-regular mesh sequences whose elements are either simplices or parallelotopes. Further examples can be found in [Di Pietro and Ern, 2012, Section 1.4.4].

Assumption 2.24. *For the remaining thesis we assume that $\mathcal{T}_\mathcal{H}$ is an admissible mesh sequence.*

Lemma 2.25 ([Di Pietro and Ern, 2012, Lemmas 1.58, 1.59])**.** *Let π_h be the L^2-orthogonal projection onto $\mathbb{P}_d^k(\mathcal{T}_h)$ defined in (2.6). Then, for all $h \in \mathcal{H}$, all $K \in \mathcal{T}_h$, and all $v \in H^{k+1}(K)$ it holds that*

$$\|v - \pi_h v\|_K \leq C''_{\mathrm{app}} h_K^{k+1} |v|_{k+1,K}. \tag{2.12a}$$

For all $F \in \mathcal{F}_h$, $F \subset \partial K$ we have

$$\|v - \pi_h v\|_F \leq \widehat{C}''_{\mathrm{app}} h_K^{k+1/2} |v|_{k+1,K}. \tag{2.12b}$$

The constants C''_{app} and $\widehat{C}''_{\mathrm{app}}$ are independent of both K and h.

2.3 Broken Sobolev spaces

We already considered polynomial spaces and their broken versions. In this section we introduce a similar concept for the Sobolev spaces $H^m(\Omega)$.

Definition 2.26. *For $m \in \mathbb{N}_0$ we define the **broken Sobolev spaces** as*

$$H^m(\mathcal{T}_h) = \{v \in L^2(\Omega) \mid v|_K \in H^m(K) \text{ for all } K \in \mathcal{T}_h\}.$$

On $H^m(\mathcal{T}_h)$ we define the seminorm and norm

$$|v|_{m,\mathcal{T}_h}^2 = \sum_{K \in \mathcal{T}_h} |v|_{m,K}^2, \qquad \|v\|_{m,\mathcal{T}_h}^2 = \sum_{j=0}^{m} |v|_j^2,$$

respectively. Clearly, for all functions $v \in H^1(\mathcal{T}_h)$ and all elements $K \in \mathcal{T}_h$, the restriction $v|_K \in H^1(K)$ has a well-defined trace on the boundary ∂K. Moreover, the **continuous trace inequality** [Di Pietro and Ern, 2012, Section 1.1.3] yields

$$\|v\|_{\partial K} \leq C_{\mathrm{ctr}} \|v\|_K^{1/2} \|v\|_{1,K}^{1/2} , \qquad \text{for all } K \in \mathcal{T}_h. \tag{2.13}$$

Obviously, the usual Sobolev spaces are subspaces of their broken versions, i.e. for every $m \geq 0$ we have $H^m(\Omega) \subset H^m(\mathcal{T}_h)$. However, the converse inclusion does not hold true. The crucial difference is that functions in $H^1(\mathcal{T}_h)$ might have **nonzero jumps at interfaces** whereas the jumps of a function in $H^1(\Omega)$ at an interface vanish. The next lemma shows that this property characterizes functions in $H^1(\Omega)$.

Lemma 2.27. [Di Pietro and Ern, 2012, Lemma 1.23] *A function $v \in H^1(\mathcal{T}_h)$ belongs to $H^1(\Omega)$ if and only if*

$$[\![v]\!]_F = 0 \qquad \text{for all } F \in \mathcal{F}_h^{\text{int}}.$$

As we have seen in Chapter 1, Maxwell's equations are well-posed in the space $D(\mathcal{C}) = H(\operatorname{curl}, \Omega) \times H_0(\operatorname{curl}, \Omega)$. However, we will assume from now on slightly more regularity, namely that the solution of Maxwell's equations satisfies $\mathbf{u}(t) \in D(\mathcal{C}) \cap H^1(\mathcal{T}_h)^6$. We prefer working in this space since it admits L^2-traces on the faces $F \in \mathcal{F}_h$. Moreover, we need at least this regularity to show convergence of the dG method. In the following, we write $\mathbf{U}_K = \mathbf{U}|_K$ for the restriction of a function $\mathbf{U}$ onto a subset $K \subset \Omega$.

Lemma 2.28. *Let $\mathbf{V} \in H^1(\mathcal{T}_h)^3$ and let ω, $\overline{\omega}$ be given piecewise constant weight functions satisfying (2.1a).*

(a) *For $\varphi \in H^1(\mathcal{T}_h)^3$ we have*

$$\sum_{K \in \mathcal{T}_h} (n_K \times \mathbf{V}_K, \varphi_K)_{\partial K} = \sum_{F \in \mathcal{F}_h^{\text{bnd}}} (n_F \times \mathbf{V}, \varphi)_F - \sum_{F \in \mathcal{F}_h^{\text{int}}} \Big((n_F \times \{\!\{\mathbf{V}\}\!\}_F^{\omega}, [\![\varphi]\!]_F)_F + (n_F \times [\![\mathbf{V}]\!]_F, \{\!\{\varphi\}\!\}_F^{\overline{\omega}})_F \Big). \tag{2.14}$$

(b) *For $\varphi \in C_0^\infty(\Omega)^3$ we have*

$$(\operatorname{curl} \mathbf{V}, \varphi) = (\mathbf{V}, \operatorname{curl} \varphi) - \sum_{F \in \mathcal{F}_h^{\text{int}}} (n_F \times [\![\mathbf{V}]\!]_F, \varphi)_F. \tag{2.15}$$

Proof. (a) By Definitions 2.8 and 2.9 of the face normal n_F and the jump $[\![\cdot]\!]_F$, respectively, we have

$$\begin{aligned} \sum_{K \in \mathcal{T}_h} (n_K \times \mathbf{V}_K, \varphi_K)_{\partial K} &= \sum_{F \in \mathcal{F}_h^{\text{bnd}}} (n_F \times \mathbf{V}, \varphi)_F \\ &\quad + \sum_{F \in \mathcal{F}_h^{\text{int}}} \Big((n_F \times \mathbf{V}_K, \varphi_K)_F - (n_F \times \mathbf{V}_{K_F}, \varphi_{K_F})_F \Big) \\ &= \sum_{F \in \mathcal{F}_h^{\text{bnd}}} (n_F \times \mathbf{V}, \varphi)_F - \sum_{F \in \mathcal{F}_h^{\text{int}}} ([\![(n_F \times \mathbf{V}) \cdot \varphi]\!]_F, 1)_F. \end{aligned}$$

The statement now follows from the identity (2.1b).

(b) The integration by parts formula (1.36) applied on every element K yields

$$\sum_{K \in \mathcal{T}_h} (\operatorname{curl} \mathbf{V}, \varphi)_K = \sum_{K \in \mathcal{T}_h} (\mathbf{V}, \operatorname{curl} \varphi)_K + \sum_{K \in \mathcal{T}_h} (n_K \times \mathbf{V}_K, \varphi_K)_{\partial K}. \tag{2.16}$$

Using (2.14) for the second sum and exploiting that for $\varphi \in C_0^\infty(\Omega)^3$ we have $[\![\varphi]\!]_F = 0$ and $\{\!\{\varphi\}\!\}_F^{\overline{\omega}} = \varphi$ for all $F \in \mathcal{F}_h^{\text{int}}$ and $\varphi|_F = 0$ for all $F \in \mathcal{F}_h^{\text{bnd}}$ proves the result. □

In the next lemma we explore the relation between $H(\operatorname{curl}, \Omega)$ and $H^1(\mathcal{T}_h)^3$. It turns out that functions in $H(\operatorname{curl}, \Omega) \cap H^1(\mathcal{T}_h)^3$ have **vanishing tangential jumps along interfaces**.

Lemma 2.29. *A function $\mathbf{V} \in H^1(\mathcal{T}_h)^3$ belongs to $H(\operatorname{curl}, \Omega)$ if and only if*

$$n_F \times [\![\mathbf{V}]\!]_F = 0 \qquad \text{for all } F \in \mathcal{F}_h^{\text{int}}. \tag{2.17a}$$

Additionally, for $\mathbf{V} \in H_0(\operatorname{curl}, \Omega) \cap H^1(\mathcal{T}_h)^3$ we have that

$$n_F \times \mathbf{V} = 0 \qquad \text{for all } F \in \mathcal{F}_h^{\text{bnd}}. \tag{2.17b}$$

Proof. (a) Let $\mathbf{V} \in H^1(\mathcal{T}_h)^3$. We first prove that (2.17a) implies $\mathbf{V} \in H(\mathrm{curl}, \Omega)$. Inserting (2.17a) into (2.15) we obtain

$$\sum_{K \in \mathcal{T}_h} \big(\mathrm{curl}\, \mathbf{V}, \varphi\big)_K = \sum_{K \in \mathcal{T}_h} \big(\mathbf{V}, \mathrm{curl}\, \varphi\big)_K = \big(\mathbf{V}, \mathrm{curl}\, \varphi\big), \qquad \text{for all } \varphi \in C_0^\infty(\Omega)^3.$$

By Definitions 1.19 and 1.20 this shows $\mathbf{V} \in H(\mathrm{curl}, \Omega)$.

(b) Now we assume that $\mathbf{V} \in H^1(\mathcal{T}_h)^3 \cap H(\mathrm{curl}, \Omega)$ and we choose $\varphi \in C_0^\infty(\Omega)^3$ arbitrarily. By Definitions 1.19 and 1.20, and (2.15) we have

$$\big(\mathbf{V}, \mathrm{curl}\, \varphi\big) = \big(\mathrm{curl}\, \mathbf{V}, \varphi\big) = \big(\mathbf{V}, \mathrm{curl}\, \varphi\big)_\Omega - \sum_{F \in \mathcal{F}_h^{\mathrm{int}}} \big(n_F \times [\![\mathbf{V}]\!]_F, \varphi\big)_F.$$

Thus, we obtain

$$\sum_{F \in \mathcal{F}_h^{\mathrm{int}}} \big(n_F \times [\![\mathbf{V}]\!]_F, \varphi\big)_F = 0 \qquad \text{for all } \varphi \in C_0^\infty(\Omega)^3.$$

Since this holds for arbitrarily chosen functions φ, we can choose it such that the support of φ intersects only a single interface. This shows (2.17a).

(c) To prove (2.17b) we use (2.16) and then (2.14) for $\varphi \in C^\infty(\overline{\Omega})^3$. Then the sum over all $F \in \mathcal{F}_h^{\mathrm{int}}$ vanishes since $[\![\varphi]\!]_F = 0$ and also $n_F \times [\![\mathbf{V}]\!]_F = 0$ by (b). By (1.37) we thus obtain

$$\sum_{F \in \mathcal{F}_h^{\mathrm{bnd}}} \big(n_F \times \mathbf{V}, \varphi\big)_F = 0 \qquad \text{for all } \varphi \in C^\infty(\overline{\Omega})^3.$$

An argument analogous to (b) applied to the boundary faces proves the result. □

CHAPTER 3

Spatial discretization: construction and analysis of the dG method

In the previous chapter we established the underlying discrete setting needed for dG methods. The aim of this chapter is to derive the actual dG space discretization of Maxwell's equations. We start by formulating the unstabilized central fluxes dG discretization and then extend it to the stabilized case leading to an upwind fluxes dG method. We show that the central fluxes discretization preserves the energy conservation of the continuous Maxwell's equations whereas the upwind fluxes discretization leads to a dissipative scheme. Moreover, we provide an error analysis for both space discretization methods. For the central fluxes scheme our arguments rely on the fact that the spatially discretized problem inherits the property of having a unitary group as solution operator as in the continuous case. In contrary, in the upwind fluxes case we need to apply an energy technique in order to profit from the dissipative nature of this space discretization which eventually gives a superior convergence rate compared to the central fluxes case.

As pointed out above we aim in this chapter in deriving the spatial discretization of Maxwell's equations (1.21),

$$\partial_t \mathbf{H}(t) = -\mathcal{C}_{\mathbf{E}}\mathbf{E}(t), \tag{3.1a}$$

$$\partial_t \mathbf{E}(t) = \mathcal{C}_{\mathbf{H}}\mathbf{H}(t) - \varepsilon^{-1}\mathbf{J}(t), \tag{3.1b}$$

$$\mathbf{H}(0) = \mathbf{H}^0, \quad \mathbf{E}(0) = \mathbf{E}^0, \tag{3.1c}$$

or, equivalently,

$$\partial_t \mathbf{u}(t) = \mathcal{C}\mathbf{u}(t) + \mathbf{j}(t), \tag{3.1d}$$

$$\mathbf{u}(0) = \mathbf{u}^0, \tag{3.1e}$$

with a **discontinuous Galerkin (dG) method**. The Maxwell operator $\mathcal{C}$ and the curl operators $\mathcal{C}_{\mathbf{H}}$, $\mathcal{C}_{\mathbf{E}}$ have been defined in (1.22) and (1.40).

3.1 dG spaces

As in the last section we will assume that the solution $\mathbf{u}(t) = (\mathbf{H}(t), \mathbf{E}(t))$ of (3.1) is slightly more regular, namely that for all $t \geq 0$ it satisfies

$$\mathbf{H}(t) \in V_\star^{\mathbf{H}} = D(\mathcal{C}_{\mathbf{H}}) \cap H^1(\mathcal{T}_h)^3, \qquad \mathbf{E}(t) \in V_\star^{\mathbf{E}} = D(\mathcal{C}_{\mathbf{E}}) \cap H^1(\mathcal{T}_h)^3,$$

or, equivalently,

$$\mathbf{u}(t) \in V_\star = V_\star^{\mathbf{H}} \times V_\star^{\mathbf{E}}.$$

We recall some consequences from this assumption. First, all functions $\mathbf{H} \in V_\star^{\mathbf{H}}$, $\mathbf{E} \in V_\star^{\mathbf{E}}$ have well-defined L^2-traces on mesh elements K, i.e. $\mathbf{H}|_{\partial K}, \mathbf{E}|_{\partial K} \in L^2(\partial K)^3$. Moreover, by Lemma 2.29 all vector fields $\mathbf{H} \in V_\star^{\mathbf{H}}$ and $\mathbf{E} \in V_\star^{\mathbf{E}}$ have vanishing tangential jumps, i.e.,

$$n_F \times [\![\mathbf{H}]\!]_F = n_F \times [\![\mathbf{E}]\!]_F = 0, \qquad \text{for all } F \in \mathcal{F}_h^{\text{int}}, \tag{3.2a}$$

and $\mathbf{E}$ has zero tangential components on the boundary, i.e.,

$$n_F \times \mathbf{E} = 0, \qquad \text{for all } F \in \mathcal{F}_h^{\text{bnd}}. \tag{3.2b}$$

In our dG method we want to construct discrete approximations $\mathbf{H}_h(t) \approx \mathbf{H}(t)$, $\mathbf{E}_h(t) \approx \mathbf{E}(t)$ in the broken polynomial space

$$V_h = \mathbb{P}_3^k(\mathcal{T}_h)^3,$$

where the mesh $\mathcal{T}_h$ belongs to an admissible mesh sequence. We refer to V_h as the **discrete solution space** (or just **dG space**) and seek our discrete solution $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h) \in V_h^2$. Note that the discrete solution space is **not** contained in the continuous solution space, $V_h^2 \not\subset V_\star$: Every function $\mathbf{v}_h \in V_h^2$ with nonzero tangential jumps cannot be in $V_\star$ due to Lemma 2.29. This characterizes dG methods as **non-conforming** space discretization schemes. So, we additionally consider the spaces

$$V_{\star,h}^{\mathbf{H}} = V_\star^{\mathbf{H}} + V_h, \qquad V_{\star,h}^{\mathbf{E}} = V_\star^{\mathbf{E}} + V_h, \qquad V_{\star,h} = V_{\star,h}^{\mathbf{H}} \times V_{\star,h}^{\mathbf{E}},$$

which contain both the exact *and* the discrete solutions. Moreover, theses spaces also contain the error function of the dG discretization, which is the difference of the exact solution and the discrete solution. Thus, it will be important that our discretizations of $\mathcal{C}_{\mathbf{H}}$ and $\mathcal{C}_{\mathbf{E}}$ are not only well-defined on V_h but also on $V_\star^{\mathbf{H}}$ and $V_\star^{\mathbf{E}}$ and thus on $V_{\star,h}^{\mathbf{H}}$ and on $V_{\star,h}^{\mathbf{E}}$, respectively.

Observe that finding a solution $\mathbf{H}, \mathbf{E}$ to (3.1) is equivalent to solving the variational problem: Seek $\mathbf{u} = (\mathbf{H}, \mathbf{E}) \in C^1(0,T; L^2(\Omega)^6) \cap C(0,T; V_\star)$ such that and

$$\big(\partial_t \mathbf{H}(t), \phi\big)_\mu = -\big(\mathcal{C}_{\mathbf{E}}\mathbf{E}(t), \phi\big)_\mu\,, \ \big(\mathbf{H}(0), \phi\big)_\mu = \big(\mathbf{H}^0, \phi\big)_\mu, \quad \text{for all } \phi \in L^2(\Omega)^3, \tag{3.3a}$$

$$\big(\partial_t \mathbf{E}(t) - \varepsilon^{-1}\mathbf{J}(t), \psi\big)_\varepsilon = \big(\mathcal{C}_{\mathbf{H}}\mathbf{H}(t), \psi\big)_\varepsilon\,, \quad \big(\mathbf{E}(0), \psi\big)_\varepsilon = \big(\mathbf{E}^0, \psi\big)_\varepsilon, \quad \text{for all } \psi \in L^2(\Omega)^3. \tag{3.3b}$$

The essential task in a space discretization of (3.3) is to discretize the curl-operators $\mathcal{C}_{\mathbf{H}}$ and $\mathcal{C}_{\mathbf{E}}$. We will now derive such discretizations resulting in **discrete curl-operators** $\boldsymbol{\mathcal{C}}_{\mathbf{H}}$ and $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$.

3.2 Central fluxes

In order to define $\boldsymbol{\mathcal{C}}_{\mathbf{H}}$, we consider the continuous curl-operator $\mathcal{C}_{\mathbf{H}}$ tested with a discrete test function: Let $\mathbf{H} \in V_\star^{\mathbf{H}}$ and $\psi_h \in V_h$. Then, by using the integration by parts formula (1.36) we infer that

$$\big(\mathcal{C}_{\mathbf{H}}\mathbf{H}, \psi_h\big)_\varepsilon = \sum_{K \in \mathcal{T}_h} \big(\operatorname{curl}\mathbf{H}, \psi_h\big)_K = \sum_{K \in \mathcal{T}_h} \big(\mathbf{H}, \operatorname{curl}\psi_h\big)_K + \sum_{K \in \mathcal{T}_h} \big(n_K \times \mathbf{H}_K, \psi_K\big)_{\partial K}.$$

Using (2.14) with $\omega = \mu c$, $\overline{\omega} = \varepsilon c$ (the local **impedance** and the local **conductance**, respectively) and (3.2a) we obtain

$$\begin{aligned}\sum_{K \in \mathcal{T}_h} \big(n_K \times \mathbf{H}_K, \psi_K\big)_{\partial K} &= \sum_{F \in \mathcal{F}_h^{\text{bnd}}} \big(n_F \times \mathbf{H}, \psi_h\big)_F - \sum_{F \in \mathcal{F}_h^{\text{int}}} \big(n_F \times \{\!\{\mathbf{H}\}\!\}_F^{\mu c}, [\![\psi_h]\!]_F\big)_F \\ &= \sum_{F \in \mathcal{F}_h^{\text{int}}} \big(\{\!\{\mathbf{H}\}\!\}_F^{\mu c}, n_F \times [\![\psi_h]\!]_F\big)_F - \sum_{F \in \mathcal{F}_h^{\text{bnd}}} \big(\mathbf{H}, n_F \times \psi_h\big)_F.\end{aligned}$$

Here, the last equality was obtained via (A.1). The same computations with $\omega = \varepsilon c$ and $\overline{\omega} = \mu c$ can be carried out for $\mathcal{C}_{\mathbf{E}}$ with the difference that the integrals over boundary faces vanish due to (3.2b). This motivates to define the following discrete versions of the curl-operators $\mathcal{C}_{\mathbf{H}}$, $\mathcal{C}_{\mathbf{E}}$ in their **weak form** (derivatives on the test functions).

Definition 3.1. *We define* $\boldsymbol{\mathcal{C}}_{\mathbf{H}} : V_{\star,h}^{\mathbf{H}} \to V_h$ *such that for all* $\psi_h \in V_h$,

$$\begin{aligned}\big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}, \psi_h\big)_{\varepsilon} = &\sum_{K\in\mathcal{T}_h} \big(\mathbf{H}, \operatorname{curl}\psi_h\big)_K \\ &+ \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \big(\{\!\{\mathbf{H}\}\!\}_F^{\mu c}, n_F \times [\![\psi_h]\!]_F\big)_F - \sum_{F\in\mathcal{F}_h^{\mathrm{bnd}}} \big(\mathbf{H}, n_F\times\psi_h\big)_F\,, \end{aligned} \tag{3.4a}$$

and $\boldsymbol{\mathcal{C}}_{\mathbf{E}} : V_{\star,h}^{\mathbf{E}} \to V_h$ *such that for all* $\phi_h \in V_h$,

$$\big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}, \phi_h\big)_{\mu} = \sum_{K\in\mathcal{T}_h} \big(\mathbf{E}, \operatorname{curl}\phi_h\big)_K + \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \big(\{\!\{\mathbf{E}\}\!\}_F^{\varepsilon c}, n_F\times[\![\phi_h]\!]_F\big)_F\,. \tag{3.4b}$$

We collect $\boldsymbol{\mathcal{C}}_{\mathbf{H}}$ *and* $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$ *in the* ***discrete Maxwell operator****,*

$$\boldsymbol{\mathcal{C}} : V_{\star,h} \to V_h^2, \quad \boldsymbol{\mathcal{C}} = \begin{pmatrix} 0 & -\boldsymbol{\mathcal{C}}_{\mathbf{E}} \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}} & 0 \end{pmatrix}. \tag{3.4c}$$

Observe that by the discontinuous ansatz the respective first terms on the right-hand sides of (3.4a) and (3.4b) do not admit a transfer of information between the mesh elements. This task is performed by the **flux functions**, i.e. by the terms involving the inner products on the interfaces $F \in \mathcal{F}_h^{\mathrm{int}}$. We see that they couple two neighboring elements by using the (weighted) mean of $\mathbf{H}$ and $\mathbf{E}$, respectively. Thus, such a dG discretization is called a **central fluxes** discretization.

The discrete curl-operators can also be stated in their equivalent **strong form** (derivatives on $\mathbf{H}$ and $\mathbf{E}$).

Lemma 3.2. *For* $\mathbf{H} \in V_{\star,h}^{\mathbf{H}}$, $\mathbf{E} \in V_{\star,h}^{\mathbf{E}}$ *and* $\phi_h, \psi_h \in V_h$ *we have that*

$$\big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}, \psi_h\big)_{\varepsilon} = \sum_{K\in\mathcal{T}_h} \big(\operatorname{curl}\mathbf{H}, \psi_h\big)_K + \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \big(n_F\times[\![\mathbf{H}]\!]_F, \{\!\{\psi_h\}\!\}_F^{\varepsilon c}\big)_F\,, \tag{3.5a}$$

and

$$\begin{aligned}\big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}, \phi_h\big)_{\mu} = &\sum_{K\in\mathcal{T}_h} \big(\operatorname{curl}\mathbf{E}, \phi_h\big)_K \\ &+ \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \big(n_F\times[\![\mathbf{E}]\!]_F, \{\!\{\phi_h\}\!\}_F^{\mu c}\big)_F - \sum_{F\in\mathcal{F}_h^{\mathrm{bnd}}} \big(n_F\times\mathbf{E}, \phi_h\big)_F\,. \end{aligned} \tag{3.5b}$$

Proof. This statement follows by applying the integration by parts formula (1.36) to (3.4). □

In the next lemma we examine the discrete curl-operators in more detail:

Lemma 3.3. *The discrete operators* $\boldsymbol{\mathcal{C}}_{\mathbf{H}}$, $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$, *and* $\boldsymbol{\mathcal{C}}$ *have the following properties:*

(a) $\boldsymbol{\mathcal{C}}_{\mathbf{H}}$, $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$, *and* $\boldsymbol{\mathcal{C}}$ *are* ***consistent****, i.e., for* $\mathbf{u} = (\mathbf{H}, \mathbf{E}) \in V_\star$ *we have*

$$\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H} = \pi_h\mathcal{C}_{\mathbf{H}}\mathbf{H}, \qquad \boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E} = \pi_h\mathcal{C}_{\mathbf{E}}\mathbf{E}, \qquad \boldsymbol{\mathcal{C}}\mathbf{u} = \pi_h\mathcal{C}\mathbf{u}. \tag{3.6}$$

(b) *For* $\mathbf{H}_h, \mathbf{E}_h \in V_h$ *we have the* **adjointness** *property*

$$\big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h, \mathbf{E}_h\big)_{\varepsilon} = \big(\mathbf{H}_h, \boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h\big)_{\mu}. \tag{3.7a}$$

(c) *The discrete Maxwell-operator* $\boldsymbol{\mathcal{C}}$ *is* ***skew-adjoint*** *on* V_h^2 *w.r.t.* $(\cdot,\cdot)_{\mu\times\varepsilon}$*, i.e. for all* $\mathbf{u}_h, \mathbf{v}_h \in V_h^2$ *we have*

$$\big(\boldsymbol{\mathcal{C}}\mathbf{u}_h, \mathbf{v}_h\big)_{\mu\times\varepsilon} = -\big(\mathbf{u}_h, \boldsymbol{\mathcal{C}}\mathbf{v}_h\big)_{\mu\times\varepsilon}\,. \tag{3.7b}$$

Note that by (3.7a) the discrete curl-operators inherit the adjointness properties of the continuous curl-operators we proved in Lemma 1.24, but on the discrete space V_h. Furthermore, observe that by (3.7b) it holds that

$$\big(\boldsymbol{\mathcal{C}}\mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} = 0 \qquad \text{for all} \quad \mathbf{u}_h \in V_h^2.$$

Note that by Definition 3.1 an expression like $\big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h, \mathbf{E}\big)_{\varepsilon}$ is only well-defined for $\mathbf{E} \in V_h$ but in general it is not well-defined for $\mathbf{E} \in V_\star^{\mathbf{E}}$. This is the reason why (3.7) only hold true on V_h.

Proof. (a) follows directly from (3.5) by using (3.2).

(b) is seen from from (3.4a) with $\mathbf{H} = \mathbf{H}_h$ and $\psi_h = \mathbf{E}_h$ and (3.5b) with $\mathbf{E} = \mathbf{E}_h$ and $\phi_h = \mathbf{H}_h$.

(c) is a direct consequence of (b). □

Using the central fluxes dG discretization to approximate Maxwell's equations (3.1) we obtain the **semidiscrete problem**: Find $\mathbf{H}_h, \mathbf{E}_h \in C^1(0,T;V_h)$ such that

$$\begin{aligned} \partial_t \mathbf{H}_h(t) &= -\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h(t), \\ \partial_t \mathbf{E}_h(t) &= \boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h(t) - \mathbf{J}_h(t), \\ \mathbf{H}_h(0) &= \mathbf{H}_h^0, \quad \mathbf{E}_h(0) = \mathbf{E}_h^0, \end{aligned} \tag{3.8a}$$

or, more compactly, find $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h) \in C^1(0,T;V_h^2)$ such that

$$\begin{aligned} \partial_t \mathbf{u}_h(t) &= \boldsymbol{\mathcal{C}}\mathbf{u}_h(t) + \mathbf{j}_h(t), \\ \mathbf{u}_h(0) &= \mathbf{u}_h^0, \end{aligned} \tag{3.8b}$$

where

$$\begin{aligned} \mathbf{J}_h(t) &= \pi_h(\varepsilon^{-1}\mathbf{J}(t)), \quad & \mathbf{j}_h(t) &= (0, -\mathbf{J}_h(t)), \\ \mathbf{H}_h^0 &= \pi_h \mathbf{H}^0, & \mathbf{E}_h(0) &= \pi_h \mathbf{E}^0. \end{aligned} \tag{3.8c}$$

Note that the boundary condition $(n \times \mathbf{E}_h(t))|_{\partial\Omega} = 0$ is weakly enforced within the definition of $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$, cf. (3.5b).

We denote the restriction of $\boldsymbol{\mathcal{C}}$ to V_h^2 by

$$\boldsymbol{\mathcal{C}}_h : V_h^2 \to V_h^2, \qquad \boldsymbol{\mathcal{C}}_h = \boldsymbol{\mathcal{C}}|_{V_h^2}.$$

Now, we can prove well-posedness of (3.8).

Theorem 3.4. *The semidiscrete problem* (3.8) *is well-posed, i.e. there is a unique solution* $\mathbf{u}_h \in C^1(0,T;V_h^2)$ *given by*

$$\mathbf{u}_h(t) = e^{t\boldsymbol{\mathcal{C}}_h}\mathbf{u}_h^0 + \int_0^t e^{(t-s)\boldsymbol{\mathcal{C}}_h}\mathbf{j}_h(s)\, ds. \tag{3.9}$$

Moreover, the following stability results hold:

(a) *For* $\mathbf{J}_h \in C(0,T;V_h)$ *we have*

$$\|\mathbf{u}_h(t)\|_{\mu\times\varepsilon} \le \|\mathbf{u}^0\|_{\mu\times\varepsilon} + \frac{1}{\sqrt{\delta}}\int_0^t \|\mathbf{J}(s)\|\, ds, \tag{3.10a}$$

where δ *was defined in* (1.20).

(b) *For* $\mathbf{J}_h \equiv 0$, *we have*

$$\|\mathbf{u}_h(t)\|_{\mu\times\varepsilon} = \|\pi_h\mathbf{u}^0\|_{\mu\times\varepsilon} \le \|\mathbf{u}^0\|_{\mu\times\varepsilon}. \tag{3.10b}$$

By (3.10b) we see that semidiscrete Maxwell's equations stemming from the central fluxes space discretization **conserve the electromagnetic energy** smilar to the continuous Maxwell's equations. In fact, for the semidiscrete solution $\mathbf{u}_h(t) = (\mathbf{H}_h(t), \mathbf{E}_h(t))$ we have

$$\mathcal{E}(\mathbf{H}_h(t), \mathbf{E}_h(t)) = \mathcal{E}(\mathbf{H}_h^0, \mathbf{E}_h^0), \qquad t \ge 0, \tag{3.11}$$

given that $\mathbf{j}_h \equiv 0$.

Proof. Since V_h^2 is finite dimensional, the operator $\boldsymbol{\mathcal{C}}_h$ is bounded, i.e. $\boldsymbol{\mathcal{C}}_h \in \mathcal{L}(V_h^2, V_h^2)$. From (3.7b) we deduce that $\boldsymbol{\mathcal{C}}_h$ is skew-adjoint and thus, by Stone's Theorem (Theorem 1.17), it generates a unitary C_0-group $e^{t\boldsymbol{\mathcal{C}}_h}$ on V_h^2. Together with Theorem 1.9 this yields (3.9). Since $e^{t\boldsymbol{\mathcal{C}}_h}$ is unitary, the equality in (3.10b) holds true. The bound in (3.10b) stems from the boundedness of the projection operator π_h, see (2.8),

$$\|\mathbf{u}_h(0)\|_{\mu\times\varepsilon} = \|\pi_h\mathbf{u}^0\|_{\mu\times\varepsilon} \le \|\mathbf{u}^0\|_{\mu\times\varepsilon}\,.$$

The estimate in (3.10a) is obtained by the triangle inequality, (2.8) and

$$\|\varepsilon^{-1}\mathbf{J}\|_\varepsilon = \|\varepsilon^{-1/2}\mathbf{J}\| \le \frac{1}{\min_{x\in\Omega}(\varepsilon(x)^{1/2})}\|\mathbf{J}\| \le \frac{1}{\delta^{1/2}}\|\mathbf{J}\|. \tag{3.12}$$

Here, we used the assumption that the coefficient ε is uniformly positive. □

In Section (3.4) we will prove that the central fluxes dG discretization is convergent of order h^k. But before, we introduce a stabilization, which will enable us to improve this rate to $h^{k+1/2}$.

3.3 Upwind fluxes

The following definitions are taken from Hesthaven and Warburton [2008] and Hochbruck et al. [2015b]. Further insight in the motivation of the stabilization terms, in particular on the solution of the Riemann problem, can be found there.

On the faces $F \in \mathcal{F}_h$ we define the coefficients

$$a_F = \frac{1}{\varepsilon_K c_K + \varepsilon_{K_F} c_{K_F}}, \quad b_F = \frac{1}{\mu_K c_K + \mu_{K_F} c_{K_F}}, \qquad F \in \mathcal{F}_h^{\text{int}}, \tag{3.13a}$$

$$b_F = \frac{1}{\mu_K c_K}, \qquad F \in \mathcal{F}_h^{\text{bnd}}, \tag{3.13b}$$

where $c_K = 1/\sqrt{\mu_K\varepsilon_K}$ is the speed of light in the element K.

Definition 3.5. *We define the* ***stabilization operators*** $\boldsymbol{\mathcal{S}}_{\mathbf{H}} : V_{\star,h}^{\mathbf{H}} \to V_h$ *such that for all* $\phi_h \in V_h$,

$$\big(\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}, \phi_h\big)_\mu = \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} a_F\big(n_F \times [\![\mathbf{H}]\!]_F, n_F \times [\![\phi_h]\!]_F\big)_F \,, \tag{3.14a}$$

and $\boldsymbol{\mathcal{S}}_{\mathbf{E}} : V_{\star,h}^{\mathbf{E}} \to V_h$ *such that for all* $\psi_h \in V_h$,

$$\big(\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}, \psi_h\big)_\varepsilon = \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} b_F\big(n_F \times [\![\mathbf{E}]\!]_F, n_F \times [\![\psi_h]\!]_F\big)_F + \sum_{F\in\mathcal{F}_h^{\mathrm{bnd}}} b_F\big(n_F \times \mathbf{E}, n_F \times \psi_h\big)_F \,. \tag{3.14b}$$

Moreover, we define

$$\boldsymbol{\mathcal{S}} : V_{\star,h} \to V_h^2, \qquad \boldsymbol{\mathcal{S}} = \begin{pmatrix} \boldsymbol{\mathcal{S}}_{\mathbf{H}} & 0 \\ 0 & \boldsymbol{\mathcal{S}}_{\mathbf{E}} \end{pmatrix}. \tag{3.14c}$$

We introduce the **stabilization parameter** $\alpha \in [0,1]$. The **stabilized dG discretization** of Maxwell's equations (3.1) reads as follows. Seek $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h) \in C^1(0,T;V_h^2)$ such that

$$\begin{aligned} \partial_t \mathbf{H}_h(t) &= -\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h(t) - \alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h(t), \\ \partial_t \mathbf{E}_h(t) &= \boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h(t) - \alpha\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}_h(t) - \mathbf{J}_h(t), \\ \mathbf{H}_h(0) &= \mathbf{H}_h^0, \quad \mathbf{E}_h(0) = \mathbf{E}_h^0, \end{aligned} \tag{3.15a}$$

or, equivalently,

$$\begin{aligned} \partial_t \mathbf{u}_h(t) &= \boldsymbol{\mathcal{C}}\mathbf{u}_h(t) - \alpha\boldsymbol{\mathcal{S}}\mathbf{u}_h(t) + \mathbf{j}_h(t), \\ \mathbf{u}_h(0) &= \mathbf{u}_h^0. \end{aligned} \tag{3.15b}$$

The source term $\mathbf{j}_h$ and the initial value $\mathbf{u}_h^0$ are given in (3.8c). In the context of dG methods employing $\alpha \in (0,1]$ is usually referred to as an **upwind fluxes** dG discretization and the choice $\alpha = 1$ as **the** upwind fluxes dG discretization. For $\alpha = 0$ we retrieve the central fluxes dG scheme.

The stabilization has no physical meaning but is only used for numerical reasons. Thus, it is natural to demand that the stabilization operators vanish when applied to the exact solution because then the extra term does not destroy the consistency of the overall discretization.

Lemma 3.6. *The stabilization operators* $\boldsymbol{\mathcal{S}}_{\mathbf{H}}$, $\boldsymbol{\mathcal{S}}_{\mathbf{E}}$, *and* $\boldsymbol{\mathcal{S}}$ *have the following properties:*

(a) *They are* ***consistent****, i.e. for* $\mathbf{u} = (\mathbf{H}, \mathbf{E}) \in V_\star$ *we have*

$$\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H} = 0, \qquad \boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E} = 0, \qquad \boldsymbol{\mathcal{S}}\mathbf{u} = 0. \tag{3.16}$$

(b) *They are* ***symmetric*** *on* V_h*, i.e. for* $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h)$, $\widehat{\mathbf{u}}_h = (\widehat{\mathbf{H}}_h, \widehat{\mathbf{E}}_h) \in V_h^2$ *we have*

$$\begin{aligned} \big(\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h, \widehat{\mathbf{H}}_h\big)_\mu &= \big(\mathbf{H}_h, \boldsymbol{\mathcal{S}}_{\mathbf{H}}\widehat{\mathbf{H}}_h\big)_\mu \,, \qquad \big(\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}_h, \widehat{\mathbf{E}}_h\big)_\varepsilon = \big(\mathbf{E}_h, \boldsymbol{\mathcal{S}}_{\mathbf{E}}\widehat{\mathbf{E}}_h\big)_\varepsilon \,, \\ \big(\boldsymbol{\mathcal{S}}\mathbf{u}_h, \widehat{\mathbf{u}}_h\big)_{\mu\times\varepsilon} &= \big(\mathbf{u}_h, \boldsymbol{\mathcal{S}}\widehat{\mathbf{u}}_h\big)_{\mu\times\varepsilon} \,. \end{aligned} \tag{3.17}$$

(c) *They are* ***positive semi-definite*** *on* V_h*, i.e. for* $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h) \in V_h^2$ *we have*

$$\big(\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h, \mathbf{H}_h\big)_\mu \geq 0, \qquad \big(\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}_h, \mathbf{E}_h\big)_\varepsilon \geq 0, \qquad \big(\boldsymbol{\mathcal{S}}\mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} \geq 0. \tag{3.18}$$

Proof. (a) follows with (3.2). (b) and (c) follow directly from the definition. □

Since the stabilization operators are symmetric and positive semidefinite, they induce seminorms. We define these seminorms in such a way that they are also well-defined on $V_\star^{\mathbf{H}}$ and $V_\star^{\mathbf{E}}$.

Definition 3.7. *For* $\mathbf{u} = (\mathbf{H}, \mathbf{E}) \in V_{\star,h}$ *we define the* ***seminorms***

$$|\mathbf{H}|^2_{\boldsymbol{\mathcal{S}}_\mathbf{H}} = \sum_{F \in \mathcal{F}_h^{\text{int}}} a_F \|n_F \times [\![\mathbf{H}]\!]_F\|^2_F \ , \tag{3.19a}$$

$$|\mathbf{E}|^2_{\boldsymbol{\mathcal{S}}_\mathbf{E}} = \sum_{F \in \mathcal{F}_h^{\text{int}}} b_F \|n_F \times [\![\mathbf{E}]\!]_F\|^2_F + \sum_{F \in \mathcal{F}_h^{\text{bnd}}} b_F \|n_F \times \mathbf{E}\|^2_F \ , \tag{3.19b}$$

and

$$|\mathbf{u}|^2_{\boldsymbol{\mathcal{S}}} = |\mathbf{H}|^2_{\boldsymbol{\mathcal{S}}_\mathbf{H}} + |\mathbf{E}|^2_{\boldsymbol{\mathcal{S}}_\mathbf{E}} \ . \tag{3.19c}$$

On the discrete space V_h we can represent these seminorms by the stabilization operators. In fact, for $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h) \in V_h^2$ we have that

$$|\mathbf{H}_h|^2_{\boldsymbol{\mathcal{S}}_\mathbf{H}} = \left(\boldsymbol{\mathcal{S}}_\mathbf{H}\mathbf{H}_h, \mathbf{H}_h\right)_\mu \ , \qquad |\mathbf{E}_h|^2_{\boldsymbol{\mathcal{S}}_\mathbf{E}} = \left(\boldsymbol{\mathcal{S}}_\mathbf{E}\mathbf{E}_h, \mathbf{E}_h\right)_\varepsilon \ , \qquad |\mathbf{u}_h|^2_{\boldsymbol{\mathcal{S}}} = \left(\boldsymbol{\mathcal{S}}\mathbf{u}_h, \mathbf{u}_h\right)_{\mu\times\varepsilon} \ . \tag{3.20}$$

Analogously to $\boldsymbol{\mathcal{C}}_h$ we define the restriction of $\boldsymbol{\mathcal{S}}$ to V_h^2 by

$$\boldsymbol{\mathcal{S}}_h : V_h^2 \to V_h^2, \qquad \boldsymbol{\mathcal{S}}_h = \boldsymbol{\mathcal{S}}|_{V_h^2}.$$

Now, we show well-posedness of (3.15).

Theorem 3.8. *The stabilized, semidiscrete problem* (3.15) *is well-posed, i.e. there is a unique solution* $\mathbf{u}_h \in C^1(0,T;V_h^2)$ *given by*

$$\mathbf{u}_h(t) = e^{t(\boldsymbol{\mathcal{C}}_h - \alpha\boldsymbol{\mathcal{S}}_h)}\mathbf{u}_h^0 + \int_0^t e^{(t-s)(\boldsymbol{\mathcal{C}}_h - \alpha\boldsymbol{\mathcal{S}}_h)}\mathbf{j}_h(s)\ ds. \tag{3.21}$$

Moreover, the following stability results hold:

(a) *For* $\mathbf{J}_h \in C(0,T;V_h)$ *we have*

$$\|\mathbf{u}_h(t)\|^2_{\mu\times\varepsilon} + 2\alpha\int_0^t |\mathbf{u}_h(s)|^2_{\boldsymbol{\mathcal{S}}}\ ds \le \mathrm{e}^1\|\mathbf{u}^0\|^2_{\mu\times\varepsilon} + \mathrm{e}^1\frac{T+1}{\delta}\int_0^t \|\mathbf{J}(s)\|^2\ ds, \tag{3.22a}$$

where δ *was defined in* (1.20).

(b) *For* $\mathbf{J}_h \equiv 0$ *we have*

$$\|\mathbf{u}_h(t)\|^2_{\mu\times\varepsilon} + 2\alpha\int_0^t |\mathbf{u}_h(s)|^2_{\boldsymbol{\mathcal{S}}}\ ds = \|\pi_h\mathbf{u}^0\|^2_{\mu\times\varepsilon}2 \le \|\mathbf{u}^0\|^2_{\mu\times\varepsilon}. \tag{3.22b}$$

Note that by (3.22b) the upwind dG discretization does not conserve the electromagnetic energy, but it is a **dissipative** scheme, i.e. it decreases the energy. In fact, we have

$$\mathcal{E}(\mathbf{H}_h(t), \mathbf{E}_h(t)) = \mathcal{E}(\mathbf{H}_h^0, \mathbf{E}_h^0) - \alpha\int_0^t |\mathbf{u}_h(s)|^2_{\boldsymbol{\mathcal{S}}}\ ds, \qquad t \ge 0.$$

The stability parameter $\alpha \in [0,1]$ controls the amount of dissipation. The dissipative term $\alpha\int_0^t |\mathbf{u}_h(s)|^2_{\boldsymbol{\mathcal{S}}}\ ds$ plays a crucial role in our error analysis and in particular in proving the superior convergence rate $h^{k+1/2}$ compared to h^k of a central fluxes discretization.

Proof. We introduce $\widehat{\boldsymbol{\mathcal{C}}}_h = \boldsymbol{\mathcal{C}}_h - \alpha \boldsymbol{\mathcal{S}}_h$, which is a bounded operator since V_h^2 is a finite dimensional vector space, i.e. $\widehat{\boldsymbol{\mathcal{C}}}_h \in \mathcal{L}(V_h^2, V_h^2)$. By (3.7b) and (3.20) we infer that for all $\mathbf{u}_h \in V_h^2$ we have that

$$\big(\widehat{\boldsymbol{\mathcal{C}}}_h \mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} = -\alpha |\mathbf{u}_h|_{\boldsymbol{\mathcal{S}}}^2.$$

Thus, the operator $\widehat{\boldsymbol{\mathcal{C}}}_h$ is dissipative on V_h^2. Moreover, we have for all $\mathbf{u}_h \in V_h^2$,

$$\big((\boldsymbol{\mathcal{I}} - \widehat{\boldsymbol{\mathcal{C}}}_h)\mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} = \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2 + \alpha |\mathbf{u}_h|_{\boldsymbol{\mathcal{S}}}^2 \geq \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2,$$

whence we conclude that $\boldsymbol{\mathcal{I}} - \widehat{\boldsymbol{\mathcal{C}}}_h$ is injective (even coercive) and thus also surjective, $\operatorname{ran}(\boldsymbol{\mathcal{I}} - \widehat{\boldsymbol{\mathcal{C}}}_h) = V_h^2$. Theorem 1.12 now states that $\widehat{\boldsymbol{\mathcal{C}}}_h$ generates a contraction semigroup $e^{t\widehat{\boldsymbol{\mathcal{C}}}_h}$ on V_h^2. So, the unique solution of (3.15) is given by (3.21).

In order to prove (3.22a) and (3.22b) we take the inner-product of (3.15b) with $\mathbf{u}_h(t)$ and use (3.7b) and (3.20) to obtain

$$\frac{1}{2}\frac{d}{dt}\|\mathbf{u}_h(t)\|_{\mu\times\varepsilon}^2 + \alpha |\mathbf{u}_h(t)|_{\boldsymbol{\mathcal{S}}}^2 = \big(\mathbf{j}_h(t), \mathbf{u}_h(t)\big)_{\mu\times\varepsilon}.$$

For vanishing source term $\mathbf{j}_h \equiv 0$ we integrate this identity from 0 to t and get the statement (3.22b). For the bound (3.22a) we apply the Cauchy–Schwarz inequality (A.5) and the weighted Young's inequality (A.2) with weight $\gamma > 0$ to the right hand side of the upper equation, which yields

$$\frac{1}{2}\frac{d}{dt}\|\mathbf{u}_h(t)\|_{\mu\times\varepsilon}^2 + \alpha |\mathbf{u}_h(t)|_{\boldsymbol{\mathcal{S}}}^2 \leq \frac{1}{2\gamma}\|\mathbf{j}_h(t)\|_{\mu\times\varepsilon}^2 + \frac{\gamma}{2}\|\mathbf{u}_h(t)\|_{\mu\times\varepsilon}^2.$$

Integrating from 0 to t gives

$$\|\mathbf{u}_h(t)\|_{\mu\times\varepsilon}^2 + 2\alpha \int_0^t |\mathbf{u}_h(s)|_{\boldsymbol{\mathcal{S}}}^2 \, ds \leq \|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{1}{\gamma}\int_0^t \|\mathbf{j}_h(s)\|_{\mu\times\varepsilon}^2 \, ds + \gamma \int_0^t \|\mathbf{u}_h(s)\|_{\mu\times\varepsilon}^2 \, ds.$$

The continuous Gronwall lemma (Lemma A.1) gives

$$\begin{aligned}\|\mathbf{u}_h(t)\|_{\mu\times\varepsilon}^2 + 2\alpha \int_0^t |\mathbf{u}_h(s)|_{\boldsymbol{\mathcal{S}}}^2 \, ds &\leq \mathrm{e}^{\gamma t}\|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{\mathrm{e}^{\gamma t}}{\gamma}\int_0^t \|\mathbf{j}_h(s)\|_{\mu\times\varepsilon}^2 \, ds \\ &\leq \mathrm{e}^{\gamma t}\|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{\mathrm{e}^{\gamma t}}{\delta\gamma}\int_0^t \|\mathbf{J}(s)\|^2 \, ds,\end{aligned}$$

where the last inequality follows from (3.12). The proof is completed by choosing $\gamma = \frac{1}{T+1}$ and the boundedness of the projection operator π_h. □

Note that the bounds (3.10) and (3.22) of the central fluxes discretization and of the upwind fluxes discretization, respectively, are derived differently. For the central fluxes we used the variation of constants formula and the bound $\|e^{t\boldsymbol{\mathcal{C}}_h}\| = 1$, whereas for the upwind fluxes we applied an energy technique. This different treating will be needed in the now following error analysis and also plays an important role in the time integration.

3.4 Error analysis of the spatial discretization

Let

$$c_\infty = \max_{K\in\mathcal{T}_h} c_K$$

denote the maximum speed of light. Furthermore, let $\mathbf{u}(t) = (\mathbf{H}(t), \mathbf{E}(t)) \in V_\star$ denote the exact solution of (3.1) and let $\mathbf{u}_h(t) = (\mathbf{H}_h(t), \mathbf{E}_h(t)) \in V_h^2$ be the semidiscrete approximation to $\mathbf{u}(t)$ obtained by the central fluxes scheme (3.8) or the upwind scheme (3.15). By considering (3.15) for $\alpha \in [0,1]$ we can do the first steps of the error analysis for both schemes simultaneously. We denote the error by

$$\mathbf{e}(t) = \mathbf{u}(t) - \mathbf{u}_h(t), \tag{3.23a}$$

which we split into a **projection error** and a **dG error**,

$$\mathbf{e}(t) = \mathbf{e}_\pi(t) - \mathbf{e}_h(t) = \begin{pmatrix} \mathbf{H}(t) - \pi_h \mathbf{H}(t) \\ \mathbf{E}(t) - \pi_h \mathbf{E}(t) \end{pmatrix} - \begin{pmatrix} \mathbf{H}_h(t) - \pi_h \mathbf{H}(t) \\ \mathbf{E}_h(t) - \pi_h \mathbf{E}(t) \end{pmatrix}. \tag{3.23b}$$

The projection $\pi_h \mathbf{u}(t)$ is the best approximation of $\mathbf{u}(t)$ in the dG space V_h^2 w.r.t. the L^2-norm and thus $\mathbf{e}_\pi(t)$ is the **best approximation error** in the dG space. The error $\mathbf{e}_h(t)$ describes the error between the best approximation $\pi_h \mathbf{u}(t)$ and the approximation $\mathbf{u}_h(t)$ we obtain from the dG scheme.

We recall that by Assumption 2.24 the mesh $\mathcal{T}_h$ has optimal polynomial approximation properties. Hence, by Lemma 2.25, for $K \in \mathcal{T}_h$, $F \in \mathcal{F}_h$, $F \subset \partial K$ and $\mathbf{H}, \mathbf{E} \in H^{k+1}(K)^3$ there are constants C_{app}, $\widehat{C}_{\text{app}}$ such that the projection errors $\mathbf{e}_\pi = (\mathbf{e}_{\pi,\mathbf{H}}, \mathbf{e}_{\pi,\mathbf{E}})$ satisfy

$$\|\mathbf{e}_{\pi,\mathbf{H}}\|_{\mu,K} \le C_{\text{app}} h_K^{k+1} |\mathbf{H}|_{k+1,K}\,, \quad \|\mathbf{e}_{\pi,\mathbf{E}}\|_{\varepsilon,K} \le C_{\text{app}} h_K^{k+1} |\mathbf{E}|_{k+1,K}\,, \tag{3.24a}$$

and

$$\|\mathbf{e}_{\pi,\mathbf{H}}\|_{\mu,F} \le \widehat{C}_{\text{app}} h_K^{k+1/2} |\mathbf{H}|_{k+1,K}\,, \quad \|\mathbf{e}_{\pi,\mathbf{E}}\|_{\varepsilon,F} \le \widehat{C}_{\text{app}} h_K^{k+1/2} |\mathbf{E}|_{k+1,K}\,. \tag{3.24b}$$

Observe that this already yields an optimal bound for the projection error $\mathbf{e}_\pi$ and consequently, we only have to bound the dG error. To improve readability we omit the arguments of the vector fields whenever possible.

Lemma 3.9. *Let $\alpha \in [0,1]$. Then, the dG error $\mathbf{e}_h = \mathbf{u}_h - \pi_h \mathbf{u}$ of (3.15) satisfies*

$$\partial_t \mathbf{e}_h = \boldsymbol{\mathcal{C}} \mathbf{e}_h - \alpha \boldsymbol{\mathcal{S}} \mathbf{e}_h + \mathbf{d}_\pi, \qquad \mathbf{e}_h(0) = 0, \tag{3.25a}$$

*with a defect $\mathbf{d}_\pi$, called the **space truncation error**, given by*

$$\mathbf{d}_\pi = -\boldsymbol{\mathcal{C}} \mathbf{e}_\pi + \alpha \boldsymbol{\mathcal{S}} \mathbf{e}_\pi. \tag{3.25b}$$

Proof. We proceed in the usual way by inserting the projected exact solution $\pi_h \mathbf{u}$ into the semidiscrete equations (3.15b). This yields

$$\partial_t \pi_h \mathbf{u} = \boldsymbol{\mathcal{C}} \pi_h \mathbf{u} - \alpha \boldsymbol{\mathcal{S}} \pi_h \mathbf{u} + \mathbf{j}_h - \mathbf{d}_\pi, \tag{3.26}$$

with a yet to be determined defect $\mathbf{d}_\pi$. Subtracting (3.26) from (3.15b) shows (3.25a).

It remains to compute $\mathbf{d}_\pi$. Projecting (3.1) onto V_h^2 and using the fact that ∂_t and π_h commute yields

$$\partial_t \pi_h \mathbf{u} = \pi_h \partial_t \mathbf{u} = \pi_h \mathcal{C} \mathbf{u} + \pi_h \mathbf{j} = \boldsymbol{\mathcal{C}} \mathbf{u} - \alpha \boldsymbol{\mathcal{S}} \mathbf{u} + \mathbf{j}_h, \tag{3.27}$$

where we used the consistency properties (3.6), (3.16) of $\boldsymbol{\mathcal{C}}$ and of $\boldsymbol{\mathcal{S}}$, respectively, i.e. $\pi_h \mathcal{C} = \boldsymbol{\mathcal{C}}$ and $\boldsymbol{\mathcal{S}} \mathbf{u} = 0$, and the definition of $\mathbf{j}_h$. Comparing (3.26) and (3.27) finally proves (3.25b). □

The lemma shows that $\mathbf{e}_h$ is the solution to the semidiscrete scheme (3.15b) with source term $\mathbf{j}_h = \mathbf{d}_\pi$ and zero initial value. Hence, we can apply the stability result (3.22), which provides a bound of $\mathbf{e}_h$ in terms of $\mathbf{d}_\pi$. In the next theorem we establish bounds on $\mathbf{d}_\pi$, i.e. bounds on

$\boldsymbol{\mathcal{C}}\mathbf{e}_\pi$ and $\boldsymbol{\mathcal{S}}\mathbf{e}_\pi$. Here, we introduce the following notation for broken ℓ^p-H^m-seminorms scaled with the order q of the spatial approximation by

$$|v|^p_{m,\mathcal{T}_h,p,q} = \sum_{K\in\mathcal{T}_h} h_K^{pq}\,|v|^p_{m,K}. \tag{3.28}$$

Note that for our $H^m(\mathcal{T}_h)$-seminorm we have $|v|_m = |v|_{m,\mathcal{T}_h,2,0}$. see Section 2.3.

Theorem 3.10. *Assume* $\mathbf{u} \in V_\star \cap H^{k+1}(\mathcal{T}_h)^6$. *Then, for all* $\varphi_h \in V_h^2$, $\mathbf{e}_\pi = \mathbf{u} - \pi_h\mathbf{u}$ *satisfies*

$$\left(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon} \le C_\pi |\varphi_h|_{\boldsymbol{\mathcal{S}}} |\mathbf{u}|_{k+1,\mathcal{T}_h,2,k+\frac{1}{2}}, \tag{3.29a}$$

and

$$\left(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon} \le \widehat{C}_\pi \|\varphi_h\|_{\mu\times\varepsilon} |\mathbf{u}|_{k+1,\mathcal{T}_h,2,k}, \tag{3.29b}$$

Moreover, we have

$$\left(\boldsymbol{\mathcal{S}}\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon} \le C_\pi |\varphi_h|_{\boldsymbol{\mathcal{S}}} |\mathbf{u}|_{k+1,\mathcal{T}_h,2,k+\frac{1}{2}}. \tag{3.30}$$

The constants are given by $C_\pi = \left(2N_\partial c_\infty\right)^{1/2}\widehat{C}_{\mathrm{app}}$ *and* $\widehat{C}_\pi = 2\widehat{C}_{\mathrm{app}} C_{\mathrm{tr}} N_\partial c_\infty \rho$.

Remark 3.11. Because the bounds (3.29a), (3.29b) are also valid for $-\varphi_h$, we conclude by $-\left(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon} = \left(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, -\varphi_h\right)_{\mu\times\varepsilon}$ that (3.29a), (3.29b) also hold true for $\left|\left(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon}\right|$. With the same arguments the bound (3.30) also holds true for $\left|\left(\boldsymbol{\mathcal{S}}\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon}\right|$.

Proof. (a) Let $\mathbf{e}_\pi = (\mathbf{e}_{\pi,\mathbf{H}}, \mathbf{e}_{\pi,\mathbf{E}})$ and $\varphi_h = (\phi_h, \psi_h)$. By definition of the inner products, we can write

$$\left(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon} = \left(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{e}_{\pi,\mathbf{E}}, \phi_h\right)_\mu + \left(\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{e}_{\pi,\mathbf{H}}, \psi_h\right)_\varepsilon. \tag{3.31}$$

For arbitrarily chosen weights $\omega_F > 0$, we have by (3.4b)

$$\begin{aligned}
\left(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{e}_{\pi,\mathbf{E}}, \phi_h\right)_\mu &= \sum_{K\in\mathcal{T}_h} \left(\operatorname{curl}\phi_h, \mathbf{e}_{\pi,\mathbf{E}}\right)_K + \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \left(n_F \times [\![\phi_h]\!]_F, \{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\right)_F \\
&= \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \left(n_F \times [\![\phi_h]\!]_F, \{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\right)_F \qquad \text{by (2.7)} \\
&\le \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \|n_F \times [\![\phi_h]\!]_F\|_F \, \|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F \qquad \text{Cauchy-Schwarz in } L^2(F) \\
&\le \left(\sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \omega_F \|n_F \times [\![\phi_h]\!]_F\|_F^2\right)^{1/2} \left(\sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \omega_F^{-1} \|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F^2\right)^{1/2},
\end{aligned} \tag{3.32}$$

where the last inequality stems from the Cauchy–Schwarz inequality (A.3) in $\mathbb{R}^{\mathrm{card}(\mathcal{F}_h^{\mathrm{int}})}$ with weight ω_F. By definition of a_F in (3.13), the second factor can be bounded by

$$\begin{aligned}
\|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F^2 &= a_F^2 \, \|\varepsilon_K c_K \mathbf{e}_{\pi,\mathbf{E}}|_K + \varepsilon_{K_F} c_{K_F} \mathbf{e}_{\pi,\mathbf{E}}|_{K_F}\|_F^2 \\
&\le 2a_F^2 \left(\|\varepsilon_K c_K \mathbf{e}_{\pi,\mathbf{E}}|_K\|_F^2 + \|\varepsilon_{K_F} c_{K_F} \mathbf{e}_{\pi,\mathbf{E}}|_{K_F}\|_F^2\right) \\
&= 2a_F^2 \left(\varepsilon_K c_K^2 \|\mathbf{e}_{\pi,\mathbf{E}}|_K\|_{\varepsilon,F}^2 + \varepsilon_{K_F} c_{K_F}^2 \|\mathbf{e}_{\pi,\mathbf{E}}|_{K_F}\|_{\varepsilon,F}^2\right) \\
&\le 2a_F \left(c_K \|\mathbf{e}_{\pi,\mathbf{E}}|_K\|_{\varepsilon,F}^2 + c_{K_F} \|\mathbf{e}_{\pi,\mathbf{E}}|_{K_F}\|_{\varepsilon,F}^2\right) \\
&\le 2a_F \, \widehat{C}_{\mathrm{app}}^2 c_\infty \left(h_K^{2k+1} |\mathbf{E}|_{k+1,K}^2 + h_{K_F}^{2k+1} |\mathbf{E}|_{k+1,K_F}^2\right).
\end{aligned} \tag{3.33}$$

Here, we applied the triangle inequality, Young's inequality, and used $\|\varepsilon^{1/2}\mathbf{u}\|_F = \|\mathbf{u}\|_{\varepsilon,F}$ and the obvious bounds

$$a_F\varepsilon_K c_K \le 1, \qquad a_F\varepsilon_{K_F} c_{K_F} \le 1. \tag{3.34}$$

The last inequality (3.33) follows from (3.24b).

(b) To prove (3.29a) we choose $\omega_F = a_F$ in (3.32). Then, the first factor in (3.32) is equal to $|\phi_h|_{\boldsymbol{\mathcal{S}}_\mathbf{H}}$. Summing over all interfaces, and recalling that every mesh element K has at most N_∂ faces shows

$$\begin{aligned}\big(\boldsymbol{\mathcal{C}}_\mathbf{E}\mathbf{e}_{\pi,\mathbf{E}},\phi_h\big)_\mu &\le \sqrt{2}\widehat{C}_{\mathrm{app}}c_\infty^{1/2}|\phi_h|_{\boldsymbol{\mathcal{S}}_\mathbf{H}}\Bigg(\sum_{F\in\mathcal{F}_h^{\mathrm{int}}} h_K^{2k+1}|\mathbf{E}|_{k+1,K}^2 + h_{K_F}^{2k+1}|\mathbf{E}|_{k+1,K_F}^2\Bigg)^{1/2}\\ &\le C_\pi|\phi_h|_{\boldsymbol{\mathcal{S}}_\mathbf{H}}|\mathbf{E}|_{k+1,\mathcal{T}_h,2,k+\frac12}\end{aligned}$$

with $C_\pi = \big(2N_\partial c_\infty\big)^{1/2}\widehat{C}_{\mathrm{app}}$.

The same computations carried out for $\boldsymbol{\mathcal{C}}_\mathbf{H}$ show

$$\big(\boldsymbol{\mathcal{C}}_\mathbf{H}\mathbf{e}_{\pi,\mathbf{H}},\psi_h\big)_\varepsilon \le C_\pi|\psi_h|_{\boldsymbol{\mathcal{S}}_\mathbf{E}}\|\mathbf{H}\|_{k+1,\mathcal{T}_h,2,k+\frac12}.$$

Using (3.31), and the Cauchy–Schwarz inequality in $\mathbb{R}^2$ yields

$$\begin{aligned}\big(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi,\varphi_h\big)_{\mu\times\varepsilon} &\le C_\pi\Big(|\phi_h|_{\boldsymbol{\mathcal{S}}_\mathbf{H}}|\mathbf{E}|_{k+1,\mathcal{T}_h,2,k+\frac12} + |\psi_h|_{\boldsymbol{\mathcal{S}}_\mathbf{E}}|\mathbf{H}|_{k+1,\mathcal{T}_h,2,k+\frac12}\Big)\\ &\le C_\pi\big(|\phi_h|_{\boldsymbol{\mathcal{S}}_\mathbf{H}}^2 + |\psi_h|_{\boldsymbol{\mathcal{S}}_\mathbf{E}}^2\big)^{1/2}\big(|\mathbf{E}|_{k+1,\mathcal{T}_h,2,k+\frac12}^2 + |\mathbf{H}|_{k+1,\mathcal{T}_h,2,k+\frac12}^2\big)^{1/2}\\ &= C_\pi|\varphi_h|_{\boldsymbol{\mathcal{S}}}|\mathbf{u}|_{k+1,\mathcal{T}_h,2,k+\frac12}.\end{aligned}$$

(c) To prove (3.29b) we start again from (3.32). To bound the first factor we use $|n_F| = 1$, the triangle inequality, Young's inequality, and subsequently the trace inequality (2.11), to obtain

$$\begin{aligned}\|n_F\times[\![\phi_h]\!]_F\|_F^2 &\le 2\big(\|\phi_h|_K\|_F^2 + \|\phi_h|_{K_F}\|_F^2\big)\\ &\le 2C_{\mathrm{tr}}^2\big(h_K^{-1}\|\phi_h\|_K^2 + h_{K_F}^{-1}\|\phi_h\|_{K_F}^2\big)\\ &= 2C_{\mathrm{tr}}^2\big(\mu_K^{-1}h_K^{-1}\|\phi_h\|_{\mu,K}^2 + \mu_{K_F}^{-1}h_{K_F}^{-1}\|\phi_h\|_{\mu,K_F}^2\big).\end{aligned} \tag{3.35}$$

In (3.32) we now choose the weight as

$$\omega_F = \frac{h_K + h_{K_F}}{2}a_F. \tag{3.36}$$

From the shape- and contact-regularity of the mesh $\mathcal{T}_h$, in fact by using (2.4), we deduce

$$\rho^{-1}a_F \le \omega_F h_K^{-1},\ \omega_F h_{K_F}^{-1} \le \rho a_F. \tag{3.37}$$

This gives

$$\begin{aligned}\omega_F\|n_F\times[\![\phi_h]\!]_F\|_F^2 &\le 2C_{\mathrm{tr}}^2\rho a_F\big(\mu_K^{-1}\|\phi_h\|_{\mu,K}^2 + \mu_{K_F}^{-1}\|\phi_h\|_{\mu,K_F}^2\big)\\ &\le 2C_{\mathrm{tr}}^2c_\infty\rho\|\phi_h\|_{\mu,K\cup K_F}^2,\end{aligned} \tag{3.38}$$

since one can easily show that

$$a_F \le c_K\mu_K, \qquad a_F \le c_{K_F}\mu_{K_F}, \tag{3.39}$$

by definition of $c_K = 1/(\varepsilon_K\mu_K)$. Finally, (3.33) yields

$$\omega_F^{-1}\|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F^2 \le 2\widehat{C}_{\mathrm{app}}^2c_\infty\rho\left(h_K^{2k}|\mathbf{E}|_{k+1,K}^2 + h_{K_F}^{2k}|\mathbf{E}|_{k+1,K_F}^2\right). \tag{3.40}$$

As in (b) one first shows

$$\big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{e}_{\pi,\mathbf{E}},\phi_h\big)_\mu \le 2\widehat{C}_{\mathrm{app}}C_{\mathrm{tr}}N_\partial c_\infty\rho\|\phi_h\|_\mu|\mathbf{E}|_{k+1,\mathcal{T}_h,2,k}$$

by summing over all interfaces, then uses the analog result for $\big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{e}_{\pi,\mathbf{H}},\psi_h\big)_\varepsilon$ to finally prove the desired bound (3.29a).

(d) It remains to prove (3.30). By Definition 3.5, the Cauchy–Schwarz inequalities (A.5), (A.3) yield

$$\begin{aligned}\big(\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{e}_{\pi,\mathbf{H}},\phi_h\big)_\mu &\le \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} a_F\|n_F\times[\![\mathbf{e}_{\pi,\mathbf{H}}]\!]_F\|_F\ \|n_F\times[\![\phi_h]\!]_F\|_F\\ &\le |\phi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}}\Big(\sum_{F\in\mathcal{F}_h^{\mathrm{int}}} a_F\|n_F\times[\![\mathbf{e}_{\pi,\mathbf{H}}]\!]_F\|_F^2\Big)^{1/2}.\end{aligned}$$

By $|n_F|=1$, the triangle inequality and Young's inequality (A.6) we infer

$$\begin{aligned}a_F\|n_F\times[\![\mathbf{e}_{\pi,\mathbf{H}}]\!]_F\|_F^2 &\le 2a_F\big(\|\mathbf{e}_{\pi,\mathbf{H}}|_K\|_F^2+\|\mathbf{e}_{\pi,\mathbf{H}}|_{K_F}\|_F^2\big)\\ &= 2a_F\big(\mu_K^{-1}\|\mathbf{e}_{\pi,\mathbf{H}}|_K\|_{\mu,F}^2+\mu_{K_F}^{-1}\|\mathbf{e}_{\pi,\mathbf{H}}|_{K_F}\|_{\mu,F}^2\big)\\ &\le 2\widehat{C}_{\mathrm{app}}^2\big(c_K h_K^{2k+1}|\mathbf{H}|_{k+1,K}^2+c_{K_F}h_{K_F}^{2k+1}|\mathbf{H}|_{k+1,K_F}^2\big).\end{aligned}\tag{3.41}$$

Here, the last inequality follows from (3.24b) and (3.39). Consequently, we have

$$\big(\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{e}_{\pi,\mathbf{H}},\phi_h\big)\le C_\pi|\phi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}}|\mathbf{H}|_{k+1,\mathcal{T}_h,2,k+\frac12}.$$

Analogously, we obtain

$$\big(\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{e}_{\pi,\mathbf{E}},\psi_h\big)\le C_\pi|\psi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{E}}}|\mathbf{E}|_{k+1,\mathcal{T}_h,2,k+\frac12}.$$

Finally, by the Cauchy–Schwarz inequality in $\mathbb{R}^2$ we obtain (3.30). □

From now on, the error analysis in the central fluxes case and in the upwind fluxes case diverge. For the central fluxes error analysis we can only use the bound (3.29b) whereas for the upwind fluxes analysis we can use the bounds (3.29a) and (3.30). This will allow us to prove the superior convergence rate for the upwind fluxes case.

3.4.1 Convergence result for central fluxes

By Lemma 3.9, the error $\mathbf{e}_h(t)$ solves the semidiscrete problem (3.8) with source term $\mathbf{j}_h(t)=-\boldsymbol{\mathcal{C}}\mathbf{e}_\pi(t)$ and initial value $\mathbf{e}_h(0)=0$. Hence, we can apply Theorem 3.4 which shows

$$\|\mathbf{e}_h(t)\|_{\mu\times\varepsilon}\le\int_0^t\|\boldsymbol{\mathcal{C}}\mathbf{e}_\pi(s)\|_{\mu\times\varepsilon}\,ds.\tag{3.42}$$

The **convergence result for the central fluxes dG discretization** is stated in the following theorem.

Theorem 3.12. *Let $\mathbf{u}\in C^1(0,T;L^2(\Omega)^6)\cap C(0,T;D(\mathcal{C})\cap H^{k+1}(\mathcal{T}_h)^6)$ be the solution of Maxwell's equations (3.1) and let $\mathbf{u}_h\in C^1(0,T;V_h^2)$ denote the semidiscrete approximation obtained from the central fluxes dG discretization (3.8). Then, the error $\mathbf{e}(t)=\mathbf{u}(t)-\mathbf{u}_h(t)$ is bounded by*

$$\|\mathbf{e}(t)\|_{\mu\times\varepsilon}\le C_{\mathrm{app}}|\mathbf{u}(t)|_{k+1,\mathcal{T}_h,1,k+1}+\widehat{C}_\pi\int_0^t|\mathbf{u}(s)|_{k+1,\mathcal{T}_h,2,k}\,ds\le Ch^k,$$

where C only depends on $C_{\mathrm{app}},\widehat{C}_\pi$, and $|\mathbf{u}(s)|_{k+1,\mathcal{T}_h}$, $s\in[0,t]$.

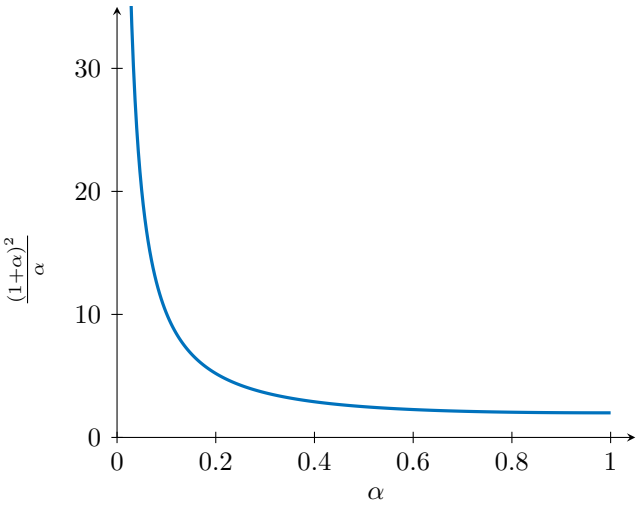


Figure 3.1: Dependence of C_{upw} on α.

Proof. As before we split the error into the projection error and the dG error, i.e., $\mathbf{e} = \mathbf{e}_\pi - \mathbf{e}_h$. The projection error can be bounded with (3.24a). To bound the dG error, we infer from (3.29b)

$$\|\boldsymbol{\mathcal{C}}\mathbf{e}_\pi\|_{\mu\times\varepsilon} = \sup_{\substack{\varphi_h\in V_h^2 \\ \|\varphi_h\|_{\mu\times\varepsilon}=1}} \big(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, \varphi_h\big)_{\mu\times\varepsilon} \leq \widehat{C}_\pi |\mathbf{u}|_{k+1,\mathcal{T}_h,2,k}. \tag{3.43}$$

Inserting this bound into (3.42), we obtain

$$\|\mathbf{e}_h(t)\|_{\mu\times\varepsilon} \leq \widehat{C}_\pi \int_0^t |\mathbf{u}(s)|_{k+1,\mathcal{T}_h,2,k}\, ds.$$

The triangle inequality $\|\mathbf{e}(t)\|_{\mu\times\varepsilon} \leq \|\mathbf{e}_\pi(t)\|_{\mu\times\varepsilon} + \|\mathbf{e}_h(t)\|_{\mu\times\varepsilon}$ completes the proof. □

3.4.2 Convergence result for upwind fluxes

In order to prove the convergence in the upwind fluxes case we apply an energy technique.

Theorem 3.13. *Let* $\mathbf{u} \in C^1(0,T;L^2(\Omega)^6) \cap C(0,T;D(\mathcal{C}) \cap H^{k+1}(\mathcal{T}_h)^6))$ *be the solution of Maxwell's equations* (3.1) *and let* $\mathbf{u}_h \in C^1(0,T;V_h^2)$ *denote the semidiscrete approximation obtained from the upwind fluxes dG discretization* (3.15)*. Then, the error* $\mathbf{e} = \mathbf{u} - \mathbf{u}_h$ *satisfies*

$$\begin{aligned}\|\mathbf{e}(t)\|_{\mu\times\varepsilon}^2 + \alpha \int_0^t |\mathbf{e}_h(s)|_{\boldsymbol{\mathcal{S}}}^2\, ds &\leq C_{\text{app}}^2 |\mathbf{u}(t)|_{k+1,\mathcal{T}_h,1,k+1}^2 + C_{\text{upw}} \int_0^t |\mathbf{u}(s)|_{k+1,\mathcal{T}_h,2,k+\frac{1}{2}}^2\, ds \\ &\leq Ch^{2k+1}.\end{aligned}$$

with $C_{\text{upw}} = C_\pi^2(1+\alpha)^2/\alpha$*. The constant* C *only depends on* $C_{\text{app}}, C_{\text{upw}}$*, and* $|\mathbf{u}(s)|_{k+1,\mathcal{T}_h}$*,* $s \in [0,t]$*.*

Note that the constant C_{upw} depends on the dissipation parameter α, see Figure 3.1. For $\alpha = 1$ we obtain the smallest constant. On the other hand for $\alpha \searrow 0$ the constant explodes and therefore the upper bound is not valid for the case $\alpha = 0$, i.e. for the central fluxes case.

Proof. The energy technique is based on taking the $\mu\times\varepsilon$-inner product of (3.25a) with $\mathbf{e}_h(t)$. Using the skew-symmetry (3.7b) and the definition of the stabilization seminorm (3.20) we obtain

$$\frac{1}{2}\frac{d}{dt}\|\mathbf{e}_h(t)\|_{\mu\times\varepsilon}^2 = -\alpha|\mathbf{e}_h(t)|_{\boldsymbol{\mathcal{S}}}^2 + \big(\mathbf{d}_\pi(t), \mathbf{e}_h(t)\big)_{\mu\times\varepsilon},$$

with $\mathbf{d}_\pi = -\boldsymbol{\mathcal{C}}\mathbf{e}_\pi + \alpha\boldsymbol{\mathcal{S}}\mathbf{e}_\pi$. Integrating from 0 to t and using $\mathbf{e}_h(0) = 0$ yields

$$\|\mathbf{e}_h(t)\|^2_{\mu\times\varepsilon} + 2\alpha\int_0^t |\mathbf{e}_h(s)|^2_{\boldsymbol{\mathcal{S}}}\, ds = 2\int_0^t \big(\mathbf{d}_\pi(s), \mathbf{e}_h(s)\big)_{\mu\times\varepsilon}\, ds. \tag{3.44}$$

From (3.29a) and (3.30) we obtain by Young's inequality with weight $\gamma = \alpha/(1+\alpha)^2$

$$\begin{aligned} 2\big(\mathbf{d}_\pi, \mathbf{e}_h\big)_{\mu\times\varepsilon} &\le 2(1+\alpha)C_\pi|\mathbf{e}_h|_{\boldsymbol{\mathcal{S}}}|\mathbf{u}|_{k+1,\mathcal{T}_h,2,k+\frac12} \\ &\le \gamma(1+\alpha)^2|\mathbf{e}_h|^2_{\boldsymbol{\mathcal{S}}} + \frac{C_\pi^2}{\gamma}|\mathbf{u}|^2_{k+1,\mathcal{T}_h,2,k+\frac12} \\ &= \alpha|\mathbf{e}_h|^2_{\boldsymbol{\mathcal{S}}} + C_{\mathrm{upw}}|\mathbf{u}|^2_{k+1,\mathcal{T}_h,2,k+\frac12}, \end{aligned} \tag{3.45}$$

by the definition of $C_{\mathrm{upw}} = C_\pi^2/\gamma$. Inserting this bound into (3.44) we conclude

$$\|\mathbf{e}_h(t)\|^2_{\mu\times\varepsilon} + \alpha\int_0^t |\mathbf{e}_h(s)|^2_{\boldsymbol{\mathcal{S}}}\, ds \le C_{\mathrm{upw}}\int_0^t |\mathbf{u}(s)|^2_{k+1,\mathcal{T}_h,2,k+\frac12}\, ds\,. \tag{3.46}$$

Because of $\mathbf{e}_h \in V_h^2$ we have $\big(\mathbf{e}_h, \mathbf{e}_\pi\big)_{\mu\times\varepsilon} = 0$, cf. (2.7), and thus we conclude

$$\|\mathbf{e}\|^2_{\mu\times\varepsilon} = \|\mathbf{e}_\pi - \mathbf{e}_h\|^2_{\mu\times\varepsilon} = \|\mathbf{e}_\pi\|^2_{\mu\times\varepsilon} + \|\mathbf{e}_h\|^2_{\mu\times\varepsilon}.$$

The statement now follows from (3.24a) and (3.46). □

3.5 Bounds of the discrete operators

The discrete operators $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$ and $\boldsymbol{\mathcal{C}}_{\mathbf{H}}$ are bounded as operators on the finite dimensional space V_h. Obviously, their bounds depend on the mesh parameter h, namely they tend to infinity for $h \searrow 0$. Next we consider this dependence in more detail. This is necessary to understand the dependence of the CFL condition on h for explicit time integration methods that we will consider in the next chapter.

We use the following short notation inspired from (3.28)

$$\|\mathbf{H}_h\|^p_{\mu,\mathcal{T}_h,p,q} = \sum_{K\in\mathcal{T}_h} h_K^{pq}\|\mathbf{H}_h\|^p_{\mu,K}, \qquad \|\mathbf{E}_h\|^p_{\varepsilon,\mathcal{T}_h,p,q} = \sum_{K\in\mathcal{T}_h} h_K^{pq}\|\mathbf{E}_h\|^p_{\varepsilon,K}. \tag{3.47a}$$

Furthermore, for $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h)$ we write

$$\|\mathbf{u}_h\|^p_{\mu\times\varepsilon,\mathcal{T}_h,p,q} = \|\mathbf{H}_h\|^p_{\mu,\mathcal{T}_h,p,q} + \|\mathbf{E}_h\|^p_{\varepsilon,\mathcal{T}_h,p,q}. \tag{3.47b}$$

Theorem 3.14. *For $\mathbf{H}_h, \mathbf{E}_h \in V_h$ we have the bounds*

$$\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h\|_\mu \le C_{\mathrm{bnd}}c_\infty\|\mathbf{E}_h\|_{\varepsilon,\mathcal{T}_h,2,-1}, \tag{3.48a}$$

and

$$\|\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h\|_\varepsilon \le C_{\mathrm{bnd}}c_\infty\|\mathbf{H}_h\|_{\mu,\mathcal{T}_h,2,-1}. \tag{3.48b}$$

Moreover, the stabilization seminorm is bounded by

$$|\mathbf{u}_h|_{\boldsymbol{\mathcal{S}}} \le \big(\widehat{C}_{\mathrm{bnd}}c_\infty\big)^{1/2}\|\mathbf{u}_h\|_{\mu\times\varepsilon,\mathcal{T}_h,2,-\frac12}. \tag{3.49}$$

The constants are given by $C_{\mathrm{bnd}} = C_{\mathrm{inv}} + 2C_{\mathrm{tr}}^2 N_\partial\rho$ and $\widehat{C}_{\mathrm{bnd}} = 2C_{\mathrm{tr}}^2 N_\partial$.

Proof. (a) We prove (3.48a). The bound (3.48b) can be shown analogously. For $\mathbf{E}_h, \phi_h \in V_h$ we have by (3.4b)

$$\big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h, \phi_h\big)_\mu = \sum_{K\in\mathcal{T}_h} \big(\mathbf{E}_h, \operatorname{curl}\phi_h\big)_K + \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \big(\{\!\{\mathbf{E}_h\}\!\}_F^{\varepsilon c}, n_F\times[\![\phi_h]\!]_F\big)_F\,. \tag{3.50}$$

We bound the two terms on the right-hand side separately. For the first term we apply the Cauchy–Schwarz inequality twice and in between the inverse inequality (2.10) to obtain

$$\begin{aligned}
\sum_{K\in\mathcal{T}_h} \big(\mathbf{E}_h, \operatorname{curl}\phi_h\big)_K &\le C_{\mathrm{inv}} \sum_{K\in\mathcal{T}_h} h_K^{-1}\|\mathbf{E}_h\|_K\|\phi_h\|_K \\
&= C_{\mathrm{inv}} \sum_{K\in\mathcal{T}_h} c_K h_K^{-1}\|\mathbf{E}_h\|_{\varepsilon,K}\|\phi_h\|_{\mu,K} \\
&\le C_{\mathrm{inv}} c_\infty \Big(\sum_{K\in\mathcal{T}_h} \|\phi_h\|_{\mu,K}^2\Big)^{1/2}\Big(\sum_{K\in\mathcal{T}_h} h_K^{-2}\|\mathbf{E}_h\|_{\varepsilon,K}^2\Big)^{1/2} \\
&= C_{\mathrm{inv}} c_\infty \|\phi_h\|_\mu \|\mathbf{E}_h\|_{\varepsilon,\mathcal{T}_h,2,-1}.
\end{aligned} \tag{3.51}$$

For the second term in (3.50), a weighted Cauchy–Schwarz inequality yields

$$\sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \big(\{\!\{\mathbf{E}\}\!\}_F^{\varepsilon c}, n_F\times[\![\phi_h]\!]_F\big)_F \le \Big(\sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \omega_F \|n_F\times[\![\phi_h]\!]_F\|_F^2\Big)^{1/2}\Big(\sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \omega_F^{-1}\|\{\!\{\mathbf{E}_h\}\!\}_F^{\varepsilon c}\|_F^2\Big)^{1/2}.$$

The weight is chosen as in (3.36). By definition of a_F, the discrete trace inequality (2.11), and (3.37) we end up with

$$\begin{aligned}
\omega_F^{-1}\|\{\!\{\mathbf{E}_h\}\!\}_F^{\varepsilon c}\|_F^2 &\le \frac{2c_\infty}{\omega_F} a_F\big(\|\mathbf{E}_h|_K\|_{\varepsilon,F}^2 + \|\mathbf{E}_h|_{K_F}\|_{\varepsilon,F}^2\big) \\
&\le \frac{2C_{\mathrm{tr}}^2 c_\infty}{\omega_F} a_F\big(h_K^{-1}\|\mathbf{E}_h\|_{\varepsilon,K}^2 + h_{K_F}^{-1}\|\mathbf{E}_h\|_{\varepsilon,K_F}^2\big) \\
&\le 2C_{\mathrm{tr}}^2 c_\infty \rho\big(h_K^{-2}\|\mathbf{E}_h\|_{\varepsilon,K}^2 + h_{K_F}^{-2}\|\mathbf{E}_h\|_{\varepsilon,K_F}^2\big).
\end{aligned}$$

Together with (3.38) we obtain the bound

$$\sum_{F\in\mathcal{F}_h^{\mathrm{int}}} \big(\{\!\{\mathbf{E}\}\!\}_F^{\varepsilon c}, n_F\times[\![\phi_h]\!]_F\big)_F \le 2C_{\mathrm{tr}}^2 N_\partial c_\infty \rho \|\phi_h\|_\mu \|\mathbf{E}_h\|_{\varepsilon,\mathcal{T}_h,2,-1}. \tag{3.52}$$

Inserting the estimates (3.51) and (3.52) in (3.50) and using the identity

$$\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h\|_\mu = \sup_{\phi_h\in V_h, \|\phi_h\|_\mu=1} \big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h, \phi_h\big)_\mu$$

yields the statement.

(b) We have $|\mathbf{u}_h|_{\boldsymbol{\mathcal{S}}}^2 = |\mathbf{H}_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}}^2 + |\mathbf{E}_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{E}}}^2$, where by Definition 3.7,

$$|\mathbf{H}_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}}^2 = \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} a_F \|n_F\times[\![\mathbf{H}_h]\!]_F\|_F^2. \tag{3.53}$$

By $|n_F| = 1$, the triangle inequality, Young's inequality, the trace inequality (2.11) and (3.34) we infer

$$\begin{aligned}
a_F\|n_F\times[\![\mathbf{H}_h]\!]_F\|_F^2 &\le 2C_{\mathrm{tr}}^2 a_F\left(\varepsilon_K c_K^2 h_K^{-1}\|\mathbf{H}_h\|_{\mu,K}^2 + \varepsilon_{K_F} c_{K_F}^2 h_{K_F}^{-1}\|\mathbf{H}_h\|_{\mu,K_F}^2\right) \\
&\le 2C_{\mathrm{tr}}^2 c_\infty\left(h_K^{-1}\|\mathbf{H}_h\|_{\mu,K}^2 + h_{K_F}^{-1}\|\mathbf{H}_h\|_{\mu,K_F}^2\right).
\end{aligned}$$

Inserting into (3.53) gives

$$|\mathbf{H}_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}}^2 \le \widehat{C}_{\mathrm{bnd}} c_\infty \|\mathbf{H}_h\|_{\mu,\mathcal{T}_h,2,-\frac{1}{2}}^2.$$

The proof of the bound for $|\mathbf{E}_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{E}}}^2$ is done analogously. □

3.6 Implementation issues

In this section we briefly consider the implementation of the dG discretization.

In order to implement a dG method we first construct a **basis** of the dG space $V_h^2 = (\mathbb{P}_3^k(\mathcal{T}_h))^6$ consisting of piecewise polynomials without any coupling between the elements of $\mathcal{T}_h$. This allows to choose basis functions independently for each element K and to restrict them to K. Hence, we consider a basis of the form

$$\{\varphi_1, \dots, \varphi_{N_h}\} = \left\{\varphi_1^K, \dots, \varphi_{n_h}^K\right\}_{K \in \mathcal{T}_h},$$

where

$$\varphi_\ell^K \in (\mathbb{P}_3^k)^6, \quad \operatorname{supp}(\varphi_\ell^K) = \overline{K}, \qquad \text{for } K \in \mathcal{T}_h, \quad \ell = 1, \dots, n_h.$$

Recall from Section 2.2.1 that the dimension n_h is given by

$$n_h = n_h(k) = \dim\left((\mathbb{P}_3^k)^6\right) = 6\dim(\mathbb{P}_3^k) = (k+3)(k+2)(k+1),$$

and thus the dimension of our basis is given by

$$N_h = n_h \operatorname{card}(\mathcal{T}_h).$$

Note that without any difficulty, one could vary the degree k between the elements. Although n_h is independent of h, we use this notation to reflect the fact that it is a parameter related to the space discretization.

Using this basis, the semidiscrete Maxwell's equations (3.15) can be equivalently stated as

$$\left(\partial_t \mathbf{u}_h, \varphi_\ell\right)_{\mu\times\varepsilon} = \left(\boldsymbol{\mathcal{C}}\mathbf{u}_h, \varphi_\ell\right)_{\mu\times\varepsilon} - \alpha\left(\boldsymbol{\mathcal{S}}\mathbf{u}_h, \varphi_\ell\right)_{\mu\times\varepsilon} + \left(\mathbf{j}_h, \varphi_\ell\right)_{\mu\times\varepsilon}, \quad \ell = 1, \dots, N_h. \tag{3.54}$$

Since $\mathbf{u}_h(t) \in V_h^2$, we can represent it as

$$\mathbf{u}_h(t) = \sum_{m=1}^{N_h} u_m(t)\varphi_m,$$

with **coefficient vector** $u(t) = (u_1(t), \dots, u_{N_h}(t)) \in \mathbb{R}^{N_h}$. Inserting this representation into (3.54) we obtain the following system of ordinary differential equations in $\mathbb{R}^{N_h}$

$$M\ddot{u}(t) = Cu(t) - \alpha S u(t) + j(t). \tag{3.55a}$$

Here,

$$M = \left(\left(\varphi_m, \varphi_\ell\right)_{\mu\times\varepsilon}\right)_{\ell,m=1,\dots,N_h}, \tag{3.55b}$$

denotes the **mass matrix** and

$$C = \left(\left(\boldsymbol{\mathcal{C}}\varphi_m, \varphi_\ell\right)_{\mu\times\varepsilon}\right)_{\ell,m=1,\dots,N_h}, \qquad S = \left(\left(\boldsymbol{\mathcal{S}}\varphi_m, \varphi_\ell\right)_{\mu\times\varepsilon}\right)_{\ell,m=1,\dots,N_h}, \tag{3.55c}$$

denote the **stiffness matrix** and the **stabilization matrix**, respectively. Furthermore, the source term is given as $j = M\widehat{j}$, where $\widehat{j}$ denotes the coefficient vector of $\pi_h\mathbf{j}(t)$, i.e.,

$$\mathbf{j}_h(t) = \pi_h\mathbf{j}(t) = \sum_{m=1}^{N_h} \widehat{j}_m(t)\varphi_m. \tag{3.55d}$$

The localized ansatz of our basis functions reduces the communication between mesh elements and ultimately leads to **block-diagonal** mass matrices (in contrast to conformal finite element

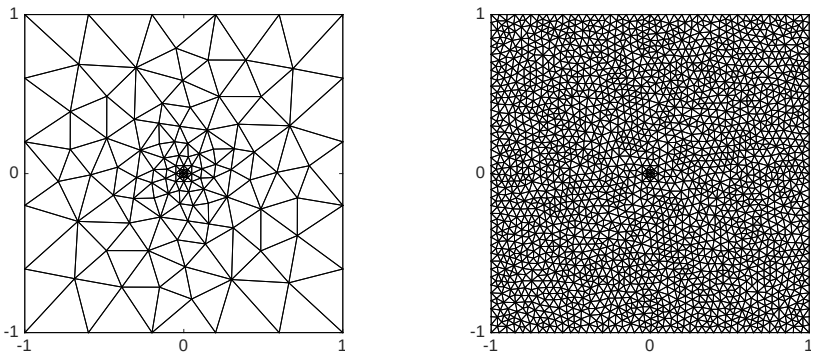


Figure 3.2: Illustration of the mesh refinement: Coarsest mesh $\mathcal{T}_h^{(1)}$ left, finest mesh $\mathcal{T}_h^{(4)}$ right.

j	$\max_{K\in\mathcal{T}_h^{(j)}} h_K$	$\min_{K\in\mathcal{T}_h^{(j)}} h_K$
1	0.2384	0.0125
2	0.1248	0.00625
3	0.0721	0.003125
4	0.0370	0.0015625

Table 3.1: Diameter of the largest and of the smallest mesh element in $\mathcal{T}_h^{(j)}$.

discretizations). However, the choice of the basis functions can have a strong impact on the performance and the accuracy of a dG scheme, in particular, when using high order polynomials. For example, choosing **modal basis functions**, i.e. basis functions being orthogonal w.r.t. $(\cdot,\cdot)_{\mu\times\varepsilon,K}$, leads to a diagonal mass matrix. However, they suffer from the fact that the approximation of the integrals by quadrature formulas is costly. Another popular ansatz is to use **nodal basis functions** associated to a set of nodal points. Usually one uses **Lagrange polynomials** and a set of nodal points leading to good approximation properties. For instance, Gauß-Lobatto points are well suited for rectangular meshes, since they provide a high approximation order and directly allow to evaluate the dG function on the faces of the elements, which is required to compute the fluxes. For general meshes the efficient approximation of integrals is often considered more important than orthogonality. For further insight, we refer to [Di Pietro and Ern, 2012, Section A.2] and [Hesthaven and Warburton, 2008, Section 6.1].

Obviously, the stiffness matrix C and the stabilization matrix S are also sparse, since a coupling between the elements only takes place over common faces. More precisely, nonzero elements can only appear for basis functions φ_m, φ_ℓ satisfying,

$$\operatorname{supp}(\varphi_m)\cap\operatorname{supp}(\varphi_\ell)=K,\ K\in\mathcal{T}_h,\qquad\text{or}\qquad \operatorname{supp}(\varphi_m)\cap\operatorname{supp}(\varphi_\ell)=F,\ F\in\mathcal{F}_h^{\text{int}}.$$

For more details we refer to [Di Pietro and Ern, 2012, Section A.1].

3.7 Numerical examples

Finally, we illustrate our theoretical results by employing the dG space discretization to the TM Maxwell's equations in $\mathbb{R}^2$, see (1.17). As setting we consider a homogeneous medium with $\mu,\varepsilon\equiv 1$ in the square $\Omega=(-1,1)^2$. We use a reference example from Descombes et al. [2013],

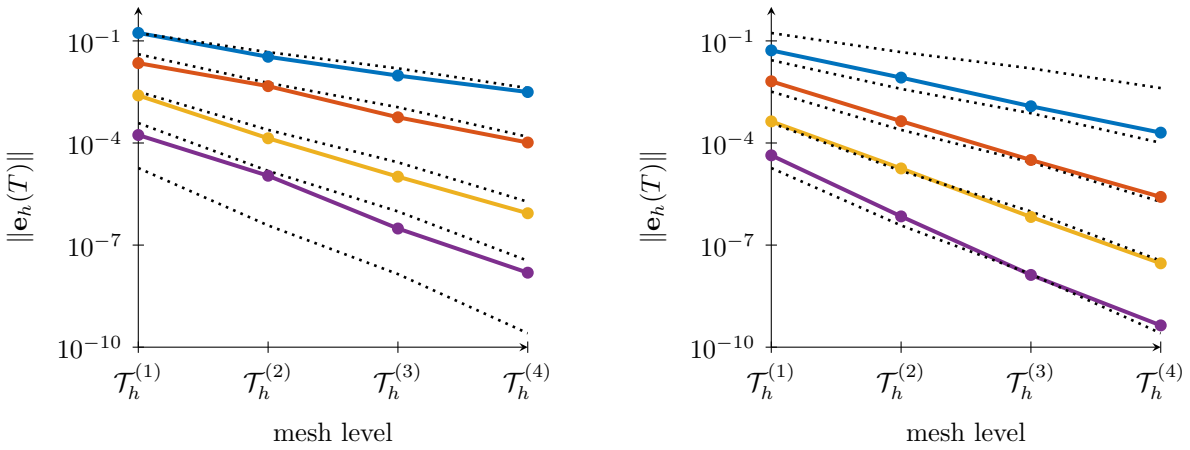


Figure 3.3: Convergence of the dG method w.r.t. the mesh width h. We used central fluxes (left), upwind fluxes with $\alpha = 1$ (right), and polynomial degrees $k = 2$, $k = 3$, $k = 4$, $k = 5$. The dotted lines have slope h^k for $k = 2, \dots, 6$. The final time was $T = 1$.

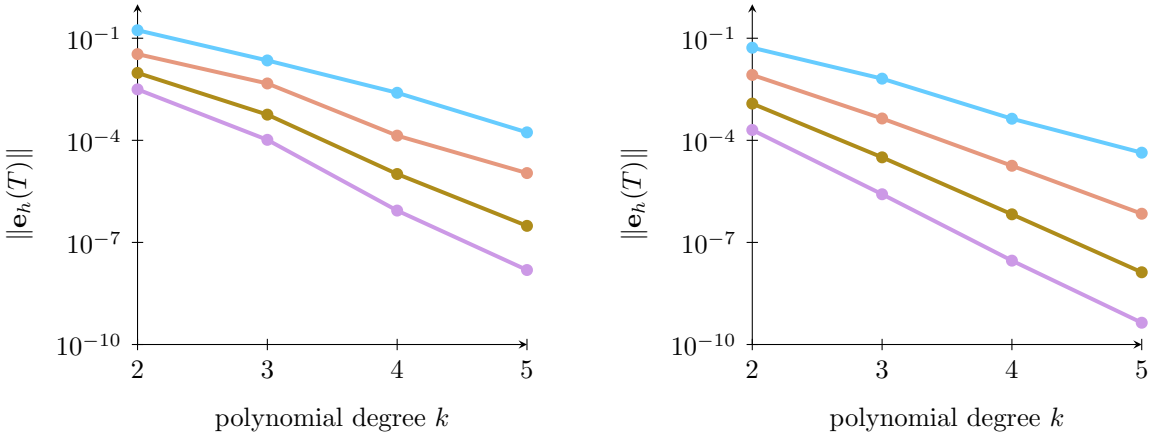


Figure 3.4: Convergence of the dG method w.r.t. the polynomial degree k. We used central fluxes (left), upwind fluxes with $\alpha = 1$ (right), and the mesh levels $\mathcal{T}_h^{(1)}$, $\mathcal{T}_h^{(2)}$, $\mathcal{T}_h^{(3)}$, $\mathcal{T}_h^{(4)}$. The final time was $T = 1$.

namely $\mathbf{u} = (\mathbf{H}_x, \mathbf{H}_y, \mathbf{E}_z)$ with components

$$\begin{aligned} \mathbf{H}_x(t) &= -\pi \sin(\pi x) \cos(\pi y) \exp(t), \\ \mathbf{H}_y(t) &= \pi \cos(\pi x) \sin(\pi y) \exp(t), \\ \mathbf{E}_z(t) &= \sin(\pi x) \sin(\pi y) \exp(t). \end{aligned} \tag{3.56a}$$

This function satisfies Maxwell's equation (1.17) with source term

$$\mathbf{J}_z(t) = -(1 + 2\pi^2) \sin(\pi x) \sin(\pi y) \exp(t). \tag{3.56b}$$

We consider a mesh sequence $\mathcal{T}_h^{(1)}, \dots, \mathcal{T}_h^{(4)}$ of continuously refined meshes. The mesh data can be found in Table 6.7. Plots of the coarsest mesh $\mathcal{T}_h^{(1)}$ and of the finest mesh $\mathcal{T}_h^{(4)}$ are given in Figure 3.2. In Figure 3.3 we plotted the L^2-norm of the error $\mathbf{e}_h(T) = \mathbf{u}_h(T) - \pi_h \mathbf{u}(T)$ which the dG method generates at the final time $T = 1$ when applied to the different mesh levels $\mathcal{T}_h^{(1)}, \dots, \mathcal{T}_h^{(4)}$. For the time integration we used the Verlet method (see Chapter 4) with a small time-step size $\tau = 10^{-5}$ which ensures that the time integration error is negligible. We observe that the central fluxes discretization converges with order k, which is in agreement with

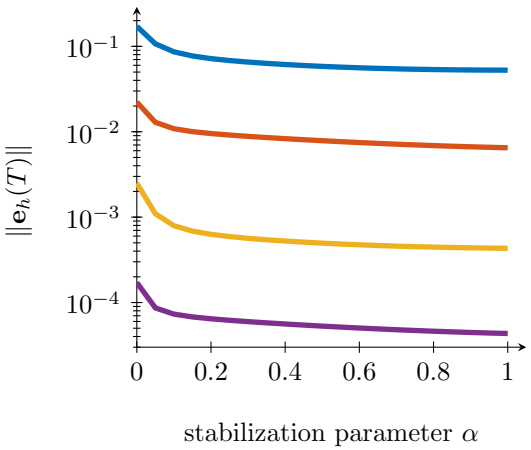


Figure 3.5: Dependence of the error of the dG method on the stabilization parameter α. We used the grid is $\mathcal{T}_h^{(1)}$, the polynomial degrees $k = 2$, $k = 3$, $k = 4$, $k = 5$ and the final time $T = 1.$.

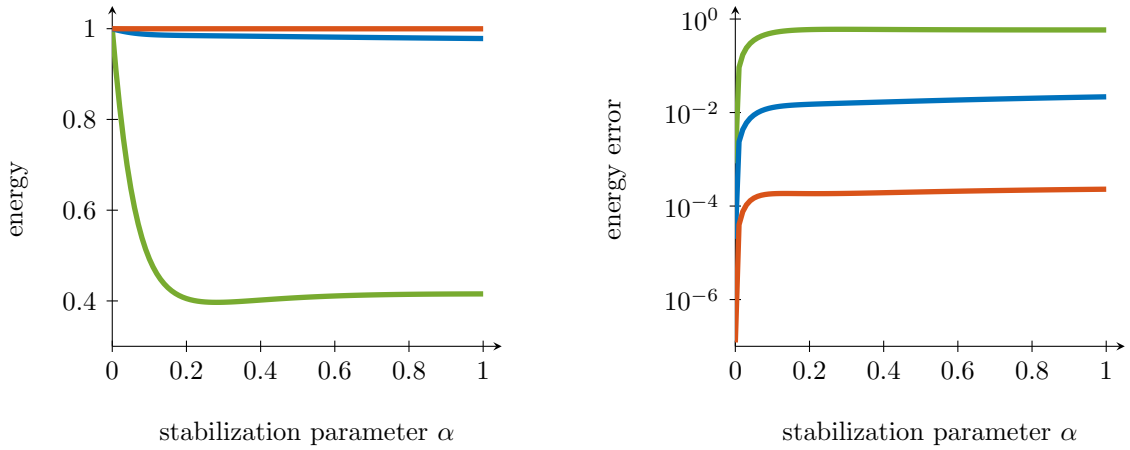


Figure 3.6: Dissipation of the electromagnetic energy. We used the mesh $\mathcal{T}_h^{(1)}$ and polynomial degrees $k = 1$, $k = 2$ and $k = 3$. The final time is $T = 20$.

Theorem 3.12. Moreover, we see that the upwind fluxes discretization is convergent with order $k + 1$ and thus even half an order better than stated in Theorem 3.13. In Figure 3.4 we show the L^2-norm of the error when using different polynomial degrees and a fixed grid in the dG method.

Recall that the constant C_{upw}, which appears in the convergence result for the upwind method, depends on α. This dependence is illustrated in Figure 3.5 where we plotted the error of the dG method for different values of alpha. We see that the choice $\alpha = 1$ yields the smallest error. For the time integration we used a Verlet-type scheme (see Chapter 5) with time-step size $\tau = 10^{-5}$ which yields a small time integration error.

As pointed out in Sections 3.2 and 3.3 the central fluxes dG discretization is energy preserving while the upwind fluxes discretization is dissipative. This is confirmed in Figure 3.5 where we give the electromagnetic energy of the semidiscrete solution obtained from a dG method depending on the stabilization parameter α.

Last, we plotted in Figure 3.7 the eigenvalues of the matrices associated with the central fluxes dG operator $\boldsymbol{\mathcal{C}}$, i.e. $M^{-1}C$, and the eigenvalues of the upwind dG operator $\boldsymbol{\mathcal{C}}-\boldsymbol{\mathcal{S}}$, i.e. $M^{-1}(C-S)$. We see that in the central fluxes case the numerically computed eigenvalues are on (or at

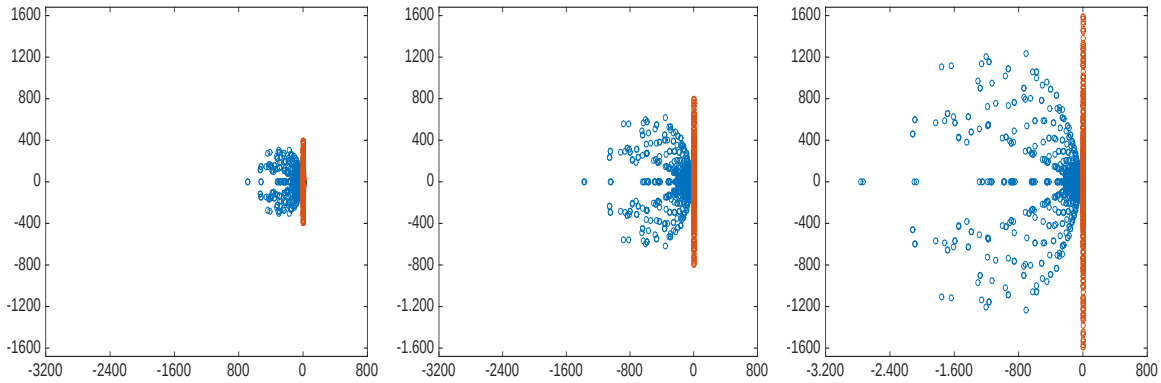


Figure 3.7: Plot of the eigenvalues of the central fluxes matrix $M^{-1}C$ (orange), and of the upwind fluxes matrix $M^{-1}(C-S)$ (blue). We used the polynomial degree $k=1$ and mesh levels $\mathcal{T}_h^{(1)}$ (left), $\mathcal{T}_h^{(2)}$ (middle) and $\mathcal{T}_h^{(3)}$ (right).

least very close to) the imaginary axis. In contrary, we observe that in the upwind fluxes case the eigenvalues moved from the imaginary axis into the left complex half plane. This is due to the dissipative behavior and the improved stability properties of the upwind fluxes discretization. Moreover, we observe that the eigenvalues grow with $(\min_{K\in\mathcal{T}_h} h_K)^{-1}$ which illustrates Theorem 3.14.

CHAPTER 4

Time integration

Let us recall Maxwell's equations from (1.18),

$$\partial_t \mathbf{u}(t) = \mathcal{C}\mathbf{u}(t) + \mathbf{j}(t) \qquad \Leftrightarrow \qquad \begin{aligned} \partial_t \mathbf{H}(t) &= -\mathcal{C}_{\mathbf{E}}\mathbf{E}(t), \\ \partial_t \mathbf{E}(t) &= \mathcal{C}_{\mathbf{H}}\mathbf{H}(t) - \varepsilon^{-1}\mathbf{J}(t), \end{aligned} \tag{4.1}$$

with initial value $\mathbf{u}(0) = \mathbf{u}^0 = (\mathbf{H}^0, \mathbf{E}^0)$, and the semidiscrete evolution equation stemming from their spatial discretization with a dG method

$$\partial_t \mathbf{u}_h(t) = (\boldsymbol{\mathcal{C}} - \alpha\boldsymbol{\mathcal{S}})\mathbf{u}_h(t) + \mathbf{j}_h(t) \qquad \Leftrightarrow \qquad \begin{aligned} \partial_t \mathbf{H}_h(t) &= -\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h(t) - \alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h(t), \\ \partial_t \mathbf{E}_h(t) &= \boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h(t) - \alpha\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}_h(t) - \mathbf{J}_h(t), \end{aligned} \tag{4.2}$$

with $\mathbf{u}_h(0) = \mathbf{u}_h^0 = (\mathbf{H}_h^0, \mathbf{E}_h^0)$, see (3.8) and (3.15). Here, $\alpha = 0$ corresponds to a central fluxes dG scheme, and $\alpha \in (0, 1]$ to an upwind fluxes dG scheme.

In order to obtain a fully discrete numerical scheme we further have to integrate the semidiscrete problem (4.2) in time. This chapter is devoted to this time integration and there are plenty time integrators for this purpose proposed in the literature. For Runge–Kutta (RK) methods let us mention the following references: explicit two or three stage RK methods are analyzed in Burman et al. [2010]. More adapted to the time integration of Maxwell's equations are the low-storage RK schemes from Diehl et al. [2010] and the implicit, algebraically stable and coercive RK methods (such as Gauss and Radau collocation methods) analyzed in Hochbruck and Pažur [2015]. Moreover, there are exponential integrators Hochbruck and Ostermann [2010], Pažur [2013], ADI methods Namiki [1999, 2000], Zhen et al. [2000], Hochbruck et al. [2015a], Krylov subspace methods Hochbruck et al. [2015b] and many others.

In this thesis we focus on two widely applied methods, namely the explicit Verlet (or leap frog) method and the implicit Crank–Nicolson method. These two methods are also the underlying schemes for the locally implicit time integrator studied in Chapter 5. We begin this chapter by introducing the two methods and shortly give an overview of their analysis in the ODE case. Next, we apply the Verlet method and the Crank–Nicolson method as time integrators for the semidiscrete Maxwell's equations emanating from a central fluxes dG discretization. We provide a stability and an error analysis, which is inspired by the analysis in the semidiscrete case. Next, we tackle the upwind fluxes case. The Crank–Nicolson method can be directly

used as a time integrator for this case. Contrary, the Verlet method first has to be modified in order to meet the requirements of the semidiscrete problem stemming from an upwind fluxes dG discretization. Similar to the central fluxes case we provide a stability analysis and an error analysis, but this time it is based on an energy techniques. We conclude this chapter with numerical results.

All time integration methods we analyze in this thesis use equidistant time steps $\tau = T/N_T$ and provide approximations $\mathbf{u}_h^n \approx \mathbf{u}_h(t_n)$, $t_n = n\tau$, $n = 0, \dots, N_T$.

4.1 Time integration for ODEs: 2nd order methods

4.1.1 The Verlet or leap frog method

In this section we construct the **Verlet** or **leap frog** method, cf. Hairer et al. [2006]. It is an **explicit** time integration scheme, which is particularly constructed to integrate second order differential equations of the type

$$\begin{aligned} \ddot{q}(t) &= f(q(t)), \\ q(0) = q^0, \quad \dot{q}(0) &= p^0. \end{aligned} \tag{4.3}$$

Here, $q : \mathbb{R}_+ \to \mathbb{R}^d$ is the searched vector field, q^0, p^0 are given initial values, and $f : \mathbb{R}^d \to \mathbb{R}^d$ is a given force. By introducing $p = \dot{q}$ we can rewrite (4.3) as a first order problem by

$$\dot{p}(t) = f(q(t)), \tag{4.4a}$$
$$\dot{q}(t) = p(t), \tag{4.4b}$$

with initial values $q(0) = q^0$ and $p(0) = p^0$. Note that (4.4) is a **Hamiltonian system** with **Hamiltonian**

$$H(p, q) = \frac{p^2}{2} - F(q),$$

where F is the anti-derivative of f, i.e. $\frac{d}{dq}F(q) = f(q)$.

The Verlet method can be derived in different ways. One option is to interpret it as a collocation method. For given values q^n, q^{n-1} and unknown q^{n+1}, let $\ell \in \mathbb{P}_2$ be the unique interpolation polynomial satisfying

$$\ell(t_j) = q^j, \qquad j = n-1,\ n,\ n+1.$$

The Lagrange form of ℓ is given by

$$\ell(t) = \frac{(t-t_n)(t-t_{n-1})}{2\tau^2} q^{n+1} - \frac{(t-t_{n+1})(t-t_{n-1})}{\tau^2} q^n + \frac{(t-t_{n+1})(t-t_n)}{2\tau^2} q^{n-1}.$$

The unknown approximation q^{n+1} is determined by the **collocation condition**

$$\ddot{\ell}(t_n) = \frac{1}{\tau^2}(q^{n+1} - 2q^n + q^{n-1}) \stackrel{!}{=} f(q^n),$$

cf. Figure 4.1. This yields the **two-step Verlet method**

$$q^{n+1} - 2q^n + q^{n-1} = \tau^2 f(q^n). \tag{4.5}$$

Now, we derive the one-step formulation of the Verlet method from the following central finite difference approximations to $p = \dot{q}$,

$$p^n = \frac{q^{n+1} - q^{n-1}}{2\tau}, \qquad p^{n+1/2} = \frac{q^{n+1} - q^n}{\tau}, \qquad p^{n-1/2} = \frac{q^n - q^{n-1}}{\tau}. \tag{4.6}$$

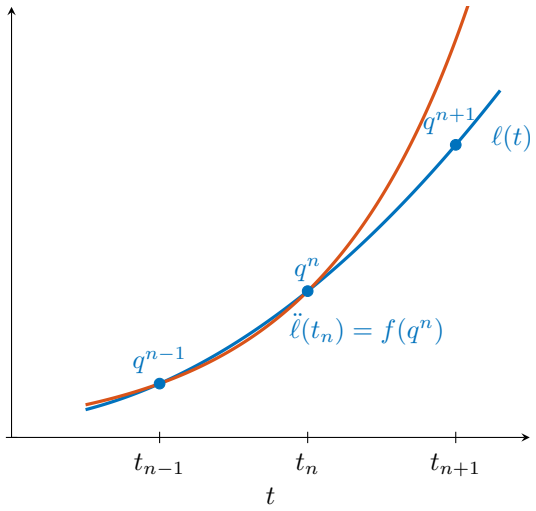


Figure 4.1: Illustration of the collocation condition for the Verlet method.

Clearly, this yields

$$p^{n+1/2} - p^{n-1/2} = \frac{q^{n+1} - 2q^n + q^{n-1}}{\tau}.$$

By using (4.5) we obtain the **one-step Verlet method**,

$$p^{n+1/2} - p^{n-1/2} = \tau f(q^n), \tag{4.7a}$$

$$q^{n+1} - q^n = \tau p^{n+1/2}. \tag{4.7b}$$

Observe that q and p live on a **staggered time grid**, i.e., approximations to q are computed at times t_n and approximations to p at times $t_{n+1/2}$. It is also possible to provide both values at t_n, since by (4.6) we have

$$p^{n+1/2} + p^{n-1/2} = \frac{q^{n+1} - q^{n-1}}{\tau} = 2p^n.$$

Solving either for $p^{n+1/2}$ or for $p^{n-1/2}$ and inserting into (4.7a) yields

$$p^{n+1/2} - p^n = p^n - p^{n-1/2} = \frac{\tau}{2} f(q^n). \tag{4.8}$$

Consequently, we obtain,

$$p^{n+1/2} - p^n = \frac{\tau}{2} f(q^n), \tag{4.9a}$$

$$q^{n+1} - q^n = \tau p^{n+1/2}, \tag{4.9b}$$

$$p^{n+1} - p^{n+1/2} = \frac{\tau}{2} f(q^{n+1}). \tag{4.9c}$$

Except for the first time step, the scheme requires only one evaluation of f per time step, since the evaluation in (4.9a) is already available from the previous time step. Alternatively, one could also use the update formula

$$p^{n+1/2} = 2p^n - p^{n-1/2}, \qquad n \geq 1,$$

which follows from (4.8).

In the following we will always use the one-step formulation (4.9) of the Verlet method.

In order to analyze the stability behavior of the Verlet method we consider the (undamped) **harmonic oscillator** in the scalar case $d = 1$,

$$\ddot{q}(t) = -\omega^2 q(t), \qquad \text{or, equivalently,} \qquad \begin{aligned} \dot{p}(t) &= -\omega^2 q(t), \\ \dot{q}(t) &= p(t), \end{aligned} \tag{4.10}$$

with initial values $q(0) = q^0$, $\dot{q}(0) = p(0) = p^0$, and $\omega \in \mathbb{R}_+$, see also [Hairer et al., 2006, Section I.5.2]. (When (4.10) is used to describe a mass-spring system we have $\omega = (k/m)^{1/2}$ where m is the mass and k is Hooke's constant of the spring). The exact solution of (4.10) is given by

$$\begin{pmatrix} p(t) \\ \omega q(t) \end{pmatrix} = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{pmatrix} \begin{pmatrix} p^0 \\ \omega q^0 \end{pmatrix}.$$

Clearly, we have

$$p(t)^2 + \big(\omega q(t)\big)^2 = \big(p^0\big)^2 + \big(\omega q^0\big)^2, \qquad \text{for all } t \geq 0.$$

Thus, the question arises if also the Verlet method produces a bounded approximation. This is answered in the following lemma.

Lemma 4.1. *Let $0 < \theta \leq 1$. Assume that the time-step size τ satisfies*

$$0 \leq \omega\tau \leq 2\theta. \tag{4.11}$$

If $\theta \in (0, 1)$, the approximation (p^n, q^n) to the solution of (4.10) obtained from the Verlet method (4.9) is bounded and satisfies

$$\big(p^n\big)^2 + (1 - \theta^2)\big(\omega q^n\big)^2 \leq \big(p^0\big)^2 + \big(\omega q^0\big)^2. \tag{4.12a}$$

Moreover, for $\theta = 1$ we have

$$|p^n| + |\omega q^n| \leq (1 + \omega T)|p^0| + |\omega q^0|, \qquad n \leq N_T. \tag{4.12b}$$

A condition on the time-step size like (4.11) is usually referred to as a **C**ourant–**F**riedrichs–**L**ewy (CFL) condition.

Proof. The Verlet method (4.9) applied to (4.10) reads

$$q^{n+1} - q^n = \tau p^{n+1/2} = \tau p^n - \frac{\tau^2\omega^2}{2} q^n, \tag{4.13a}$$

and

$$p^{n+1} - p^n = -\frac{\tau\omega^2}{2}(q^{n+1} + q^n) = -\frac{\tau^2\omega^2}{2} p^n - \frac{\tau\omega^2}{2}(2 - \frac{\tau^2\omega^2}{2}) q^n. \tag{4.13b}$$

Here, the second equality in (4.13a) follows with (4.9a) and (4.13b) is obtained by adding (4.9a) with (4.9c) and inserting (4.13a). We can write (4.13) compactly as

$$\begin{pmatrix} p^{n+1} \\ q^{n+1} \end{pmatrix} = A \begin{pmatrix} p^n \\ q^n \end{pmatrix}, \qquad A = \begin{pmatrix} 1 + \zeta & \frac{\zeta}{\tau}(2 + \zeta) \\ \tau & 1 + \zeta \end{pmatrix}, \qquad \zeta = -\frac{\tau^2\omega^2}{2}. \tag{4.14}$$

The stability of (4.14) is determined by the eigenvalues of A, which are given by

$$\lambda_{1,2} = \zeta + 1 \pm \sqrt{\zeta^2 + 2\zeta}.$$

Now, we discuss the three cases associated with the sign of the term in $\sqrt{\cdot}$.

(1) "$\zeta^2 + 2\zeta > 0$": Because of $\zeta < 0$ this case is equivalent to $\zeta < -2$. Then, for the second eigenvalue we have

$$\lambda_2 = \zeta + 1 - \sqrt{\zeta^2 + 2\zeta} < -1 - \underbrace{\sqrt{\zeta^2 + 2\zeta}}_{>0} < -1.$$

This means that for all τ with

$$\zeta = -\frac{\tau^2\omega^2}{2} < -2 \qquad \Longleftrightarrow \qquad \omega\tau > 2$$

(4.14) possesses an unbounded solution.

(2) "$\zeta^2 + 2\zeta < 0$": Because of $\zeta < 0$ this case is equivalent to $\zeta > -2$. Then, for the eigenvalues we have that

$$\lambda_{1,2} = \zeta + 1 \pm \sqrt{(-1)(-1)(\zeta^2 + 2\zeta)} = \zeta + 1 \pm i\sqrt{-\zeta^2 - 2\zeta},$$

which means that their real and imaginary part are given by

$$\mathrm{Re}(\lambda_{1,2}) = \zeta + 1, \qquad \mathrm{Im}(\lambda_{1,2}) = \pm\sqrt{-\zeta^2 - 2\zeta} \neq 0,$$

respectively. Consequently, we have

$$|\lambda_{1,2}|^2 = (\zeta + 1)^2 - \zeta^2 - 2\zeta = 1, \qquad \lambda_1 \neq \lambda_2.$$

This means that for all τ with

$$\zeta = -\frac{\tau^2\omega^2}{2} > -2 \qquad \Longleftrightarrow \qquad \omega\tau < 2$$

(4.14) possesses a bounded solution. For the bound (4.12a) we use an energy technique. By (4.9a) and (4.9c) we have

$$p^{n+1/2} = \frac{1}{2}(p^{n+1} + p^n) + \frac{\tau}{4}\omega^2(q^{n+1} - q^n).$$

Inserting this into the first equality of (4.13a) we obtain

$$\begin{aligned} q^{n+1} - q^n &= \tau p^{n+1/2} = \frac{\tau}{2}(p^{n+1} + p^n) + \frac{\tau^2}{4}\omega^2(q^{n+1} - q^n), \\ p^{n+1} - p^n &= -\frac{\tau}{2}\omega^2(q^{n+1} + q^n), \end{aligned}$$

where the second equality stems from (4.13b). Multiplying the first line with $\omega^2(q^{n+1} + q^n)$, the second line with $p^{n+1} + p^n$ and adding the resulting equations we get

$$(p^{n+1})^2 - (p^n)^2 + (\omega q^{n+1})^2 - (\omega q^n)^2 = \frac{\tau^2}{4}\omega^2\left((\omega q^{n+1})^2 - (\omega q^n)^2\right).$$

Summing from 0 to n yields

$$(p^n)^2 + (\omega q^n)^2 + \frac{\tau^2}{4}\omega^2(\omega q^0)^2 = (p^0)^2 + (\omega q^0)^2 + \frac{\tau^2}{4}\omega^2(\omega q^n)^2.$$

Employing the CFL condition (4.11) we obtain,

$$(p^N)^2 + (1 - \theta)(\omega q^N)^2 \leq (p^0)^2 + (\omega q^0)^2,$$

which proves (4.12a).

0	0	0
1	1/2	1/2
	1/2	1/2

Table 4.1: Butcher tableau of the Crank–Nicolson method.

(3) "$\zeta^2 + 2\zeta = 0$": This case can only appear for $\zeta = 0$ or $\zeta = -2$. The first case would require $\tau = 0$ or $\omega = 0$, which are both a contradiction to our assumptions. So, we only consider $\zeta = -2$. In this case we get the repeated eigenvalue

$$\lambda_{1,2} = \zeta + 1 = -1.$$

In order to decide if this repeated eigenvalue provides a bounded solution we insert $\zeta = -2$ into (4.14),

$$\begin{pmatrix} p^{n+1} \\ q^{n+1} \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ \tau & -1 \end{pmatrix} \begin{pmatrix} p^n \\ q^n \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ \tau & -1 \end{pmatrix}^{n+1} \begin{pmatrix} p^0 \\ q^0 \end{pmatrix} = \begin{pmatrix} (-1)^{n+1} & 0 \\ \tau(n+1)(-1)^n & (-1)^{n+1} \end{pmatrix} \begin{pmatrix} p^0 \\ q^0 \end{pmatrix}.$$

Taking the absolute value we obtain

$$|p^{n+1}| \le |p^0|, \qquad |q^{n+1}| \le |q^0| + \tau(n+1)|p^0| \le |q^0| + T|p^0|,$$

and consequently the time-step size τ satisfying

$$\zeta = -\frac{\tau^2\omega^2}{2} = -2 \qquad \Longleftrightarrow \qquad \omega\tau = 2$$

yields a bounded solution on finite time intervals, $T < \infty$. □

4.1.2 The Crank–Nicolson method

The **Crank–Nicolson** or implicit trapezoidal rule [Hairer et al., 2006, Section II.1.1], [Hairer and Wanner, 1996, Section IV.3] is an implicit RK scheme with Butcher Tableau given in Table 4.1. We first analyze the Crank–Nicolson method when applied to a general evolution equation in $\mathbb{R}^d$,

$$\begin{aligned} \dot{u}(t) &= F(t, u(t)), \\ u(0) &= u^0, \end{aligned} \tag{4.15}$$

with vector fields $u : \mathbb{R}_+ \to \mathbb{R}^d$ and $F : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$. According to Table 4.1 the Crank–Nicolson time integration reads

$$U^{n1} = u^n, \qquad \dot{U}^{n1} = F(t_n, U^{n1}), \tag{4.16a}$$

$$U^{n2} = u^n + \frac{\tau}{2}(\dot{U}^{n1} + \dot{U}^{n2}), \qquad \dot{U}^{n2} = F(t_{n+1}, U^{n2}), \tag{4.16b}$$

$$u^{n+1} = u^n + \frac{\tau}{2}(\dot{U}^{n1} + \dot{U}^{n2}), \tag{4.16c}$$

where we used the notation of Hochbruck [2015]. Observe that we have

$$U^{n1} = u^n, \qquad U^{n2} = u^{n+1}, \qquad \dot{U}^{n1} + \dot{U}^{n2} = F(t_n, u^n) + F(t_{n+1}, u^{n+1}).$$

As a consequence, the Crank–Nicolson scheme simplifies to

$$u^{n+1} = u^n + \frac{\tau}{2}\Big(F(t_n, u^n) + F(t_{n+1}, u^{n+1})\Big). \tag{4.17}$$

Because we are interested in linear Maxwell's equations, we now consider the Crank–Nicolson method for a linear evolution equation in $\mathbb{R}^d$,

$$\begin{aligned} \dot{u}(t) &= Au(t) + f(t), \\ u(0) &= u^0, \end{aligned} \tag{4.18}$$

where $A \in \mathbb{R}^{d\times d}$ is a matrix with field of values (see (1.26)) in the left complex half-plane, $\mathcal{F}(A) \subset \mathbb{C}^-$. The exact solution of (4.18) is given by the variation of constants formula, c.f. Theorem 1.9,

$$u(t) = e^{tA}u^0 + \int_0^t e^{(t-s)A} f(s)\, ds.$$

Employing the Crank–Nicolson method (4.17) as a time integrator for (4.18) gives the scheme

$$u^{n+1} = u^n + \frac{\tau}{2}A\big(u^{n+1} + u^n\big) + \frac{\tau}{2}\big(f^{n+1} + f^n\big), \tag{4.19}$$

with $f^n = f(t_n)$. Equivalently, we can write this as

$$R_L u^{n+1} = R_R u^n + \frac{\tau}{2}\big(f^{n+1} + f^n\big), \tag{4.20a}$$

or

$$u^{n+1} = Ru^n + \frac{\tau}{2}R_L^{-1}\big(f^{n+1} + f^n\big), \tag{4.20b}$$

with matrices

$$R_L = R_L(\tau A), \qquad R_R = R_R(\tau A), \qquad R = R(\tau A), \tag{4.21a}$$

stemming from the functions

$$R_L(z) = 1 - \frac{z}{2}, \qquad R_R(z) = 1 + \frac{z}{2}, \qquad R(z) = R_L(z)^{-1}R_R(z) = \frac{1 + \frac{z}{2}}{1 - \frac{z}{2}}. \tag{4.21b}$$

$R(z)$ is called the **stability function** of the Crank–Nicolson method. It is the same stability function as the one of the implicit midpoint rule, namely the $(1,1)$-**Padé approximation** of e^z, i.e., numerator and denominator are polynomials of degree one and $e^z - R(z) = \mathcal{O}(z^3)$ for $z \to 0$, cf. [Hairer and Wanner, 1996, Section IV.3]. The stability region associated with this stability function is the left complex half-plane $\mathbb{C}^-$, see [Hairer and Wanner, 1996, Chapter IV] or [Hochbruck, 2015, Chapters 10.3 and 10.6]. Consequently, the Crank–Nicolson method is **A-stable**, but is not L-stable, since $\lim_{z\to\infty} R(z) = -1 \neq 0$.

Lemma 4.2. *Assume that $A \in \mathbb{R}^{d\times d}$ satisfies $\mathcal{F}(A) \subset \mathbb{C}^-$. Then, the approximation obtained from the Crank–Nicolson method (4.19) is bounded by*

$$|u^n| \le |u^0| + \frac{\tau}{2}\sum_{m=0}^{n-1} |f^{m+1} + f^m|. \tag{4.22}$$

Proof. The assumption on the field of values of A ensures

$$|R| \le 1, \qquad |R_L^{-1}| \le 1,$$

since the functions $R(z)$ and $R_L^{-1}(z)$ defined in (4.21) are the stability functions of the Crank–Nicolson method and of the implicit Euler method, respectively. From (4.20b) we deduce that

$$u^{n+1} = R^{n+1}u^0 + \frac{\tau}{2}\sum_{m=0}^{n} R^{n-m}R_L^{-1}\big(f^{m+1} + f^m\big).$$

Taking norms and using the upper bounds on R and R_L^{-1} yields the statement. □

Remark 4.3. For skew-adjoint matrices A the matrix R is unitary and thus, for vanishing source term $f \equiv 0$, the Crank–Nicolson method preserves the norm,

$$|u^n| = |u^0|, \qquad n = 1, 2, \dots.$$

Our later error analysis for the full discretization of Maxwell's equations is based on the ideas of the convergence analysis of the Crank–Nicolson method. It is instructive to recall this analysis also in the ODE case.

4.1.3 Error analysis of the Crank–Nicolson method

In order to compute the error $e^n = u^n - u(t_n)$ of the Crank–Nicolson method we would like to insert the exact solution $u(t)$ of (4.15) into the recursion (4.17) of the Crank–Nicolson method. However, the exact solution does not satisfy this recursion but we obtain

$$u(t_{n+1}) = u(t_n) + \frac{\tau}{2}\big(\dot u(t_n) + \dot u(t_{n+1})\big) - d^n, \qquad d^n = -\tau^2 \delta^n(\dot u), \tag{4.23}$$

where the **defect** d^n is the quadrature error of the trapezoidal rule applied to $\dot u$,

$$\tau^2 \delta^n(g) = \int_{t_n}^{t_{n+1}} g(t)\ dt - \frac{\tau}{2}\big(g(t_{n+1}) + g(t_n)\big). \tag{4.24a}$$

We can express quadrature errors in terms of the Peano kernels, see, e.g., [Hochbruck, 2015, Theorem 1.10]. Hence, we have

$$\delta^n(g) = \int_{t_n}^{t_{n+1}} \frac{(t - t_n)(t - t_{n+1})}{2\tau^2} \ddot g(t)\ dt, \qquad |\delta^n(g)| \le \frac{1}{8} \int_{t_n}^{t_{n+1}} |\ddot g(t)|\ dt, \tag{4.24b}$$

since the Peano kernel of the trapezoidal rule is given by $s(s-1)/2$, cf. [Hochbruck, 2015, Example 1.11]. Subtracting (4.23) from (4.19) and using (4.18) shows that the error e^n satisfies

$$e^{n+1} = e^n + \frac{\tau}{2} A(e^{n+1} + e^n) + d^n, \qquad e^0 = 0. \tag{4.25a}$$

Solving this recursion gives

$$e^{n+1} = R e^n + R_L^{-1} d^{n+1} = \sum_{m=0}^{n} R^{n-m} R_L^{-1} d^m, \tag{4.25b}$$

by definition of R and R_L in (4.21).

Lemma 4.4. *Assume that $A \in \mathbb{R}^{d\times d}$ satisfies $\mathcal{F}(A) \subset \mathbb{C}^-$. Then, the error of the Crank–Nicolson method satisfies*

$$|e^n| \le \frac{\tau^2}{8} \int_0^{t_n} |u^{(3)}(t)|\ dt.$$

Proof. As in the proof of Lemma 4.2 the assumption on the field of values of A ensures $|R| \le 1$ and $|R_L^{-1}| \le 1$. Taking norms in (4.25b) and using the triangle inequality and (4.24b) yields the result. □

1/2	1/2
	1

Table 4.2: Butcher tableau of the implicit midpoint method.

4.1.4 The implicit midpoint method

In this section we consider a very similar method to the Crank–Nicolson method, namely the **implicit midpoint method** [Hairer et al., 2006, Section II.1.1], [Hairer and Wanner, 1996, Section IV.3]. As the Crank–Nicolson method it is an implicit RK scheme, however with one instead of two stages, see its Butcher tableau in Table 4.2. If we apply the implicit midpoint method to the general evolution equation (4.15) we obtain the recursion

$$U^{n1} = u^n + \frac{\tau}{2}\dot{U}^{n1}, \qquad \dot{U}^{n1} = F(t_{n+1/2}, U^{n1}), \tag{4.26a}$$

$$u^{n+1} = u^n + \tau\dot{U}^{n1}, \tag{4.26b}$$

where we abbreviated $t_{n+1/2} = t_n + \tau/2$. Clearly, we have $U^{n1} = (u^{n+1} + u^n)/2$ and hence we can write the implicit midpoint method compactly as

$$u^{n+1} = u^n + \tau F\left(t_{n+1/2}, \frac{u^{n+1} + u^n}{2}\right). \tag{4.27}$$

For the linear evolution equation (4.18) the implicit midpoint scheme is given by

$$u^{n+1} = u^n + \frac{\tau}{2}A\big(u^{n+1} + u^n\big) + \tau f^{n+1/2}, \tag{4.28a}$$

where $f^{n+1/2} = f(t_{n+1/2})$. Using the matrices R_L, R_R (4.21) introduced for the Crank–Nicolson method, we can write (4.28a) equivalently as

$$R_L u^{n+1} = R_R u^n + \tau f^{n+1/2}, \tag{4.28b}$$

or

$$u^{n+1} = R u^n + \tau R_L^{-1} f^{n+1/2}. \tag{4.28c}$$

Comparing (4.28c) with (4.20b), we see that the implicit midpoint method and the Crank–Nicolson method exhibit the same stability function, and only differ in the treatment of the source function f. As a consequence, the implicit midpoint method has the same stability properties as the Crank–Nicolson method. In fact, it is A-stable, but not L-stable. Moreover, for skew-adjoint matrices A it conserves the norm,

$$|u^n| = |u^0|, \qquad n = 1, 2, \dots,$$

In summary, we observe that the Crank–Nicolson method and the implicit midpoint method are closely related. Thus, we focus in this thesis on the Crank–Nicolson method, and only mention how the proofs and techniques can be transferred to the implicit midpoint method.

4.1.5 Error analysis of the implicit midpoint method

In this section we present the error analysis for the implicit midpoint method when applied to the linear evolution equation (4.18). It turns out that it is more involved compared to the Crank–Nicolson method. As a first step, we present an error recursion in the subsequent lemma.

Lemma 4.5. *The error of the implicit midpoint method (4.28) satisfies*

$$e^{n+1} = e^n + \frac{\tau}{2}A\big(e^{n+1} + e^n\big) + \bar{d}^n. \tag{4.29a}$$

The defect $\bar{d}^n$ is given by

$$\bar{d}^n = -\tau^2\bar{\delta}^n(\dot{u}) - \tau^2 A\big(\delta^n(u) - \bar{\delta}^n(u)\big), \qquad \tau^2\bar{\delta}^n(g) = \int_{t_n}^{t_{n+1}} g(t)\ dt - \tau g(t_{n+1/2}), \tag{4.29b}$$

where $\bar{\delta}^n$ is the error of the midpoint quadrature rule and where δ^n is the error of the trapezoidal quadrature rule given in (4.24). We further have

$$\bar{\delta}^n(g) = \int_{t_n}^{t_{n+1/2}} \frac{(t_n - t)^2}{2\tau^2}\ddot{g}(t)\ dt + \int_{t_{n+1/2}}^{t_{n+1}} \frac{(t_{n+1} - t)^2}{2\tau^2}\ddot{g}(t)\ dt, \tag{4.29c}$$

and

$$\big|\bar{\delta}^n(\dot{u})\big| \le \frac{1}{8}\int_{t_n}^{t_{n+1}} |u^{(3)}(t)|\ dt, \qquad \big|A\big(\delta^n(u) - \bar{\delta}^n(u)\big)\big| \le \frac{1}{4}\int_{t_n}^{t_{n+1}} |A\ddot{u}(t)|\ dt. \tag{4.29d}$$

Note that in the Crank–Nicolson method only the defect $d^n = -\tau^2\delta^n(\dot{u})$ appears, whereas the defect $\bar{d}^n$ of the implicit midpoint method involves besides $-\tau^2\bar{\delta}^n(\dot{u})$ additionally $-\tau^2 A\big(\delta^n(u) - \bar{\delta}^n(u)\big)$. We observe that the implicit midpoint method is **only of order 2 if $A\ddot{u}(t)$ can be bounded**. In the literature this assumption is often made, e.g. in [Hochbruck and Pažur, 2015, Theorem 5.4] for Maxwell's equations. However, this is not a desirable condition, since it requires artificial regularity assumptions on the exact solution. In this thesis, we propose a different way that omits additional regularity assumptions. But first we give the proof of the upper lemma.

Proof. We start by inserting the exact solution of (4.18) into the implicit midpoint scheme (4.28b),

$$u(t_{n+1}) = u(t_n) + \frac{\tau}{2}A\big(u(t_{n+1}) + u(t_n)\big) + \tau f^{n+1/2} - \bar{d}^n. \tag{4.30}$$

Subtracting (4.30) from (4.28a) yields the error recursion (4.29a) and it remains to determine the defect $\bar{d}^{n+1}$. Note that we cannot proceed like for the Crank–Nicolson method in the previous section. The reason is that we cannot write (4.30) analog to (4.23), i.e. with derivatives of the exact solution u, since the treatment of the linear part and the treatment of the source term do not match. Instead, we replace the source term $f^{n+1/2}$ according to the linear evolution equation (4.18). Then, (4.30) reads

$$u(t_{n+1}) = u(t_n) + \frac{\tau}{2}A\big(u(t_{n+1}) + u(t_n)\big) - \tau A u(t_{n+1/2}) + \tau\dot{u}(t_{n+1/2}) - \bar{d}^n. \tag{4.31}$$

So, the defect $\bar{d}^{n+1}$ is given by

$$\begin{aligned}
\bar{d}^n =& \frac{\tau}{2}A\big(u(t_{n+1}) + u(t_n)\big) - \tau A u(t_{n+1/2}) + \tau\dot{u}(t_{n+1/2}) - \int_{t_n}^{t_{n+1}} \dot{u}(t)dt \\
=& \frac{\tau}{2}A\big(u(t_{n+1}) + u(t_n)\big) - \tau A u(t_{n+1/2}) - \tau^2\bar{\delta}^n(\dot{u}) \\
=& \frac{\tau}{2}A\big(u(t_{n+1}) + u(t_n)\big) - \int_{t_n}^{t_{n+1}} Au(t)\ dt - \tau A u(t_{n+1/2}) + \int_{t_n}^{t_{n+1}} Au(t)\ dt - \tau^2\bar{\delta}^n(\dot{u}).
\end{aligned}$$

This shows (4.29b). The representation (4.29c) of $\tau^2\bar{\delta}^n(g)$ is obtained by using the Peano kernels.

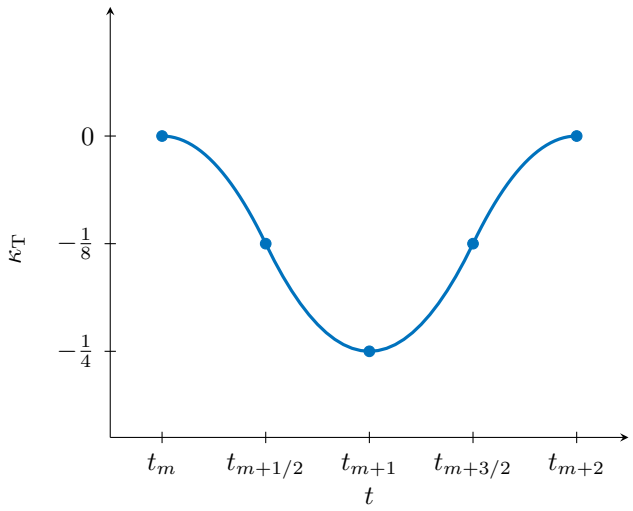


Figure 4.2: Kernel $\kappa_T(t)$of the defect $\bar{\zeta}^{m+1} - \bar{\zeta}^m$.

The first bound in (4.29d) is seen from (4.29b). For the second bound note that by (4.24a), (4.29b) we have that

$$\begin{aligned}\delta^n(u) - \bar{\delta}^n(u) &= -\frac{1}{2\tau}\big(u(t_{n+1}) - 2u(t_{n+1/2}) + u(t_n)\big) \\ &= -\frac{1}{2}\int_{t_{n+1/2}}^{t_{n+1}} \frac{t_{n+1}-t}{\tau}\ddot{u}(t)\,dt + \frac{1}{2}\int_{t_n}^{t_{n+1/2}} \frac{t_n - t}{\tau}\ddot{u}(t)\,dt,\end{aligned} \tag{4.32}$$

where the second equality follows from a Taylor expansion of $u(t_{n+1})$ and of $u(t_n)$ around $t_{n+1/2}$. Taking norms and applying the triangle inequality completes the proof. □

Now, we discuss how we can eliminate the boundedness assumption on $A\ddot{u}(t)$.

Lemma 4.6. *The error e^{n+1} of the implicit midpoint method (4.28) satisfies*

$$e^{n+1} = \bar{\zeta}^n - R^{n+1}\bar{\zeta}^0 - \tau^2 \sum_{m=0}^{n} R^{n-m} R_L^{-1} \bar{\delta}^m(\dot{u}) - \sum_{m=0}^{n-1} R^{n-m}(\bar{\zeta}^{m+1} - \bar{\zeta}^m), \tag{4.33a}$$

where $\bar{\zeta}^m$ is given by

$$\bar{\zeta}^m = \tau\big(\delta^m(u) - \bar{\delta}^m(u)\big), \tag{4.33b}$$

and obeys the bounds

$$\big|\bar{\zeta}^m\big| \le \frac{\tau^2}{4} \max_{t\in[t_m,t_{m+1}]} |\ddot{u}(t)|\,, \qquad \big|\bar{\zeta}^{m+1} - \bar{\zeta}^m\big| \le \frac{\tau^2}{8}\int_{t_m}^{t_{m+2}} |u^{(3)}(t)|\,dt. \tag{4.33c}$$

Proof. First, we write the error recursion (4.29a) with the matrices R_L, R_R,

$$R_L e^{n+1} = R_R e^n + \bar{d}^n = R_R e^n - \tau^2 \bar{\delta}^m(\dot{u}) + (R_L - R_R)\bar{\zeta}^n, \tag{4.34a}$$

where we used $-\tau A = R_L - R_R$ for the second equality. Because R_L is invertible, we can rewrite (4.34a) as

$$e^{n+1} = Re^n - \tau^2 R_L^{-1}\bar{\delta}^m(\dot{u}) + (I - R)\bar{\zeta}^n. \tag{4.34b}$$

Solving this recursion yields

$$e^{n+1} = -\tau^2 \sum_{m=0}^{n} R^{n-m} R_L^{-1}\bar{\delta}^m(\dot{u}) + \sum_{m=0}^{n} R^{n-m}(I - R)\bar{\zeta}^m.$$

The second sum can be rewritten as

$$\begin{aligned}\sum_{m=0}^{n} R^{n-m}(I-R)\bar{\zeta}^m &= \sum_{m=0}^{n} R^{n-m}\bar{\zeta}^m - \sum_{m=-1}^{n-1} R^{n-m}\bar{\zeta}^{m+1} \\ &= R^0\bar{\zeta}^n - R^{n+1}\bar{\zeta}^0 - \sum_{m=0}^{n-1} R^{n-m}(\bar{\zeta}^{m+1}-\bar{\zeta}^m). \end{aligned} \tag{4.34c}$$

This shows (4.33a). The first bound in (4.33c) follows from (4.32) by

$$\left|\bar{\zeta}^m\right| \le \frac{\tau}{4}\int_{t_m}^{t_{m+1}} |\ddot{u}(t)|\ dt \le \frac{\tau^2}{4}\max_{t\in[t_m,t_{m+1}]}|\ddot{u}(t)|\,. \tag{4.35}$$

For the second bound we use a Taylor expansion of $u(t_{m+1})$ and of $u(t_m)$ around $t_{m+1/2}$. Together with (4.32) this implies

$$\begin{aligned}\bar{\zeta}^m &= -\frac{1}{2}\big(u(t_{m+1}) - 2u(t_{m+1/2}) + u(t_m)\big) \\ &= -\frac{\tau^2}{8}\ddot{u}(t_{m+1/2}) - \frac{\tau^2}{2}\int_{t_{m+1/2}}^{t_{m+1}} \frac{(t_{m+1}-t)^2}{2\tau^2}u^{(3)}(t)\ dt + \frac{\tau^2}{2}\int_{t_m}^{t_{m+1/2}} \frac{(t_m-t)^2}{2\tau^2}u^{(3)}(t)\ dt.\end{aligned} \tag{4.36}$$

Because of $\ddot{u}(t_{m+3/2}) - \ddot{u}(t_{m+1/2}) = \int_{t_{m+1/2}}^{t_{m+3/2}} u^{(3)}(t)\ dt$ we have that

$$\bar{\zeta}^{m+1} - \bar{\zeta}^m = \frac{\tau^2}{2}\int_{t_m}^{t_{m+2}} \kappa_{\mathrm{T}}(t)u^{(3)}(t)\ dt,$$

where κ_{T} is given by

$$2\tau^2\kappa_{\mathrm{T}}(t) = \begin{cases} -(t_m-t)^2, & t\in[t_m,t_{m+1/2}], \\ (t_{m+1}-t)^2 - \frac{\tau^2}{2}, & t\in[t_{m+1/2},t_{m+3/2}], \\ -(t_{m+2}-t)^2, & t\in[t_{m+3/2},t_{m+2}]. \end{cases} \tag{4.37}$$

Since $\kappa_{\mathrm{T}}(t)$ is bounded by $1/4$ for $t\in[t_m,t_{m+2}]$, see also Figure 4.2, we obtain the second bound in (4.33c) and the proof is finished. □

We end this section with the convergence result for the implicit midpoint method.

Lemma 4.7. *Assume that the matrix $A\in\mathbb{R}^{d\times d}$ satisfies $\mathcal{F}(A)\subset\mathbb{C}^-$. Then, the error of the implicit midpoint method satisfies*

$$|e^n| \le \frac{\tau^2}{4}\max_{t\in[t_0,t_1]\cup[t_{n-1},t_n]}|\ddot{u}(t)| + \frac{3\tau^2}{8}\int_0^{t_n}|u^{(3)}(t)|\ dt. \tag{4.38}$$

Note that this bound does **not** involve $A\ddot{u}(t)$.

Proof. As pointed out in the proof of Lemma 4.4 we have $|R|\le 1$ and $|R_L^{-1}|\le 1$. Taking norms in (4.33a) and using the triangle inequality, we infer

$$\left|e^{n+1}\right| \le \left|\bar{\zeta}^n\right| + \left|\bar{\zeta}^0\right| + \tau^2\sum_{m=0}^{n}\left|\bar{\delta}^m(\dot{u})\right| + \sum_{m=0}^{n-1}\left|\bar{\zeta}^{m+1}-\bar{\zeta}^m\right|.$$

Inserting the bounds (4.29d) and (4.33c) concludes the proof. □

4.2 Time integration for Maxwell's equations: central fluxes

In this section we integrate the semidiscrete Maxwell's equations in time by using the Verlet method (4.9) or the Crank–Nicolson method (4.19). Note that if we want to use the Verlet method we have to restrict ourselves to a space discretization using central fluxes, i.e. we can only consider the semidiscrete problem (3.8). This is due to the fact that the Verlet method is only applicable to evolution equations possessing a Hamiltonian structure, see (4.4). However, this is not the case for an upwind fluxes dG discretization (3.15a). Thus, we only consider a central fluxes dG discretization in this section. However, we point out that it is possible to adapt the Verlet method to the upwind fluxes case. This will be discussed in Section 4.3.

We start by stating the Verlet method (4.9) and the Crank–Nicolson method (4.19) when applied to the semidiscrete Maxwell's equations emanating from a central fluxes dG discretization (3.8). The Verlet method yields the recursion

$$\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^n, \tag{4.39a}$$

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \tau\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h^{n+1/2} - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n), \tag{4.39b}$$

$$\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n+1/2} = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1}, \tag{4.39c}$$

and from the Crank–Nicolson method we obtain

$$\mathbf{u}_h^{n+1} - \mathbf{u}_h^n = \frac{\tau}{2}\boldsymbol{\mathcal{C}}(\mathbf{u}_h^{n+1} + \mathbf{u}_h^n) + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n), \tag{4.40}$$

where we abbreviated $\mathbf{u}_h^n = (\mathbf{H}_h^n, \mathbf{E}_h^n)$ and $\mathbf{j}_h^n = (0, -\mathbf{J}_h^n)$.

Remark 4.8. In fact, the scheme (4.39) is a (slight) adaption of the Verlet method (4.9) as proposed in [Verwer, 2011, Equation (2.1)]. It is constructed in such a way that the scheme (4.39) can be interpreted as perturbed Crank–Nicolson method, see Lemma 4.9 below. This will allow us to construct the locally implicit time integrator in Chapter 5. For convenience we refer to (4.39) as the Verlet method in this thesis.

4.2.1 Stability and energy preservation

Adapting (4.20a) to the Maxwell's equations, the Crank–Nicolson method can also be written as

$$\boldsymbol{\mathcal{R}}_L\mathbf{u}_h^{n+1} = \boldsymbol{\mathcal{R}}_R\mathbf{u}_h^n + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n), \tag{4.41a}$$

with operators $\boldsymbol{\mathcal{R}}_L, \boldsymbol{\mathcal{R}}_R : V_h^2 \to V_h^2$ given by

$$\boldsymbol{\mathcal{R}}_L = \boldsymbol{\mathcal{I}} - \frac{\tau}{2}\boldsymbol{\mathcal{C}}, \qquad \boldsymbol{\mathcal{R}}_R = \boldsymbol{\mathcal{I}} + \frac{\tau}{2}\boldsymbol{\mathcal{C}}, \qquad \boldsymbol{\mathcal{C}} = \begin{pmatrix} 0 & -\boldsymbol{\mathcal{C}}_{\mathbf{E}} \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}} & 0 \end{pmatrix}. \tag{4.41b}$$

In the next lemma we show that the Verlet method can also be cast into the form (4.41a) but with perturbed operators $\widehat{\boldsymbol{\mathcal{R}}}_L$ and $\widehat{\boldsymbol{\mathcal{R}}}_R$.

Lemma 4.9. *The Verlet method (4.39) can be written as*

$$\widehat{\boldsymbol{\mathcal{R}}}_L\mathbf{u}_h^{n+1} = \widehat{\boldsymbol{\mathcal{R}}}_R\mathbf{u}_h^n + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n), \tag{4.42a}$$

with operators $\widehat{\boldsymbol{\mathcal{R}}}_L, \widehat{\boldsymbol{\mathcal{R}}}_R : V_h^2 \to V_h^2$ defined by

$$\widehat{\boldsymbol{\mathcal{R}}}_L = \boldsymbol{\mathcal{R}}_L - \frac{\tau^2}{4}\boldsymbol{\mathcal{D}}, \qquad \widehat{\boldsymbol{\mathcal{R}}}_R = \boldsymbol{\mathcal{R}}_R - \frac{\tau^2}{4}\boldsymbol{\mathcal{D}}, \qquad \boldsymbol{\mathcal{D}} = \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{\mathcal{C}}_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}} \end{pmatrix}. \tag{4.42b}$$

Proof. Adding (4.39a) and (4.39c) we obtain

$$\mathbf{H}_h^{n+1} - \mathbf{H}_h^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{E}_h^{n+1} + \mathbf{E}_h^n). \tag{4.43a}$$

This is the first component of (4.42a). For the second component, we subtract (4.39c) from (4.39a):

$$\mathbf{H}_h^{n+1/2} = \frac{1}{2}(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) + \frac{\tau}{4}\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{E}_h^{n+1} - \mathbf{E}_h^n).$$

Inserting this into (4.39b) yields

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) + \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{E}_h^{n+1} - \mathbf{E}_h^n) - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n), \tag{4.43b}$$

which is the second component of (4.42a). □

The next lemma gives fundamental properties of the operators $\boldsymbol{\mathcal{R}}_L$, $\boldsymbol{\mathcal{R}}_R$, $\widehat{\boldsymbol{\mathcal{R}}}_L$ and $\widehat{\boldsymbol{\mathcal{R}}}_R$.

Lemma 4.10. *Let $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h), \widehat{\mathbf{u}}_h \in V_h^2$. The operators $\boldsymbol{\mathcal{R}}_L, \boldsymbol{\mathcal{R}}_R$ have the following properties:*

$$\big(\boldsymbol{\mathcal{R}}_L\mathbf{u}_h, \widehat{\mathbf{u}}_h\big)_{\mu\times\varepsilon} = \big(\mathbf{u}_h, \boldsymbol{\mathcal{R}}_R\widehat{\mathbf{u}}_h\big)_{\mu\times\varepsilon}, \tag{4.44a}$$

$$\big(\boldsymbol{\mathcal{R}}_L\mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} = \big(\boldsymbol{\mathcal{R}}_R\mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} = \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2, \tag{4.44b}$$

$$\|\boldsymbol{\mathcal{R}}_L^{-1}\|_{\mu\times\varepsilon} \leq 1. \tag{4.44c}$$

Moreover, for the $\widehat{\boldsymbol{\mathcal{R}}}_L, \widehat{\boldsymbol{\mathcal{R}}}_R$ operators we have that

$$\big(\widehat{\boldsymbol{\mathcal{R}}}_L\mathbf{u}_h, \widehat{\mathbf{u}}_h\big)_{\mu\times\varepsilon} = \big(\mathbf{u}_h, \widehat{\boldsymbol{\mathcal{R}}}_R\widehat{\mathbf{u}}_h\big)_{\mu\times\varepsilon}, \tag{4.45a}$$

$$\big(\widehat{\boldsymbol{\mathcal{R}}}_L\mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} = \big(\widehat{\boldsymbol{\mathcal{R}}}_R\mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} = \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2 - \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h\|_{\mu}^2. \tag{4.45b}$$

Proof. The statements (4.44a), (4.44b), (4.45a) and (4.45b) follow directly from the adjointness property (3.7) of the discrete curl-operators.

By (4.44b) we see that $\boldsymbol{\mathcal{R}}_L$ is injective (and thus bijective). In fact, we have

$$\|\boldsymbol{\mathcal{R}}_L\mathbf{u}_h\|_{\mu\times\varepsilon} = \sup_{\mathbf{v}_h\in V_h^2} \frac{\big(\boldsymbol{\mathcal{R}}_L\mathbf{u}_h, \mathbf{v}_h\big)_{\mu\times\varepsilon}}{\|\mathbf{v}_h\|_{\mu\times\varepsilon}} \geq \frac{\big(\boldsymbol{\mathcal{R}}_L\mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon}}{\|\mathbf{u}_h\|_{\mu\times\varepsilon}} = \|\mathbf{u}_h\|_{\mu\times\varepsilon}.$$

This implies that $\|\boldsymbol{\mathcal{R}}_L\|_{\mu\times\varepsilon} \geq 1$ and by setting $\mathbf{v}_h = \boldsymbol{\mathcal{R}}_L\mathbf{u}_h$ we obtain

$$\|\boldsymbol{\mathcal{R}}_L^{-1}\mathbf{v}_h\|_{\mu\times\varepsilon} \leq \|\mathbf{v}_h\|_{\mu\times\varepsilon},$$

which proves (4.44c). □

As a consequence of this lemma, we can write (4.41a) as

$$\mathbf{u}_h^{n+1} = \boldsymbol{\mathcal{R}}\mathbf{u}_h^n + \frac{\tau}{2}\boldsymbol{\mathcal{R}}_L^{-1}\big(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n\big), \qquad \text{where} \qquad \boldsymbol{\mathcal{R}} = \boldsymbol{\mathcal{R}}_L^{-1}\boldsymbol{\mathcal{R}}_R. \tag{4.46a}$$

Solving this recursion yields

$$\mathbf{u}_h^{n+1} = \boldsymbol{\mathcal{R}}^{n+1}\mathbf{u}_h^0 + \frac{\tau}{2}\sum_{m=0}^{n}\boldsymbol{\mathcal{R}}^{n-m}\boldsymbol{\mathcal{R}}_L^{-1}\big(\mathbf{j}_h^{m+1} + \mathbf{j}_h^m\big). \tag{4.46b}$$

Note the similarity of this recursion with the semidiscrete approximation,

$$\mathbf{u}_h(t_{n+1}) = e^{t_{n+1}\boldsymbol{\mathcal{C}}}\mathbf{u}_h^0 + \int_0^{t_{n+1}} e^{(t_{n+1}-t)\boldsymbol{\mathcal{C}}}\mathbf{j}_h(t)\,dt,$$

see Theorem 3.4. As explained in Section 4.1.2, we observe that the Crank–Nicolson method employs a $(1,1)$-Padé approximation to the exponential function, i.e.

$$\boldsymbol{\mathcal{R}} = (\boldsymbol{\mathcal{I}} - \frac{\tau}{2}\boldsymbol{\mathcal{C}})^{-1}(\boldsymbol{\mathcal{I}} + \frac{\tau}{2}\boldsymbol{\mathcal{C}}) \approx e^{\tau\boldsymbol{\mathcal{C}}}.$$

In Theorem 3.4 we showed the stability of the semidiscrete approximation. In the central fluxes case this proof is based on the skew-adjointness of the discretized Maxwell operator $\boldsymbol{\mathcal{C}}$ and the resulting unitary property of the group it generates,

$$\|e^{t\boldsymbol{\mathcal{C}}}\|_{\mu\times\varepsilon} = 1.$$

In the next lemma we show that this property is preserved by the operator $\boldsymbol{\mathcal{R}}$. In fact, $\boldsymbol{\mathcal{R}}$ is a **Cayley transform**.

Lemma 4.11. *The operator $\boldsymbol{\mathcal{R}}$ is an isometry on V_h^2, i.e.,*

$$\|\boldsymbol{\mathcal{R}}\mathbf{u}_h\|_{\mu\times\varepsilon} = \|\mathbf{u}_h\|_{\mu\times\varepsilon}, \qquad \|\boldsymbol{\mathcal{R}}\|_{\mu\times\varepsilon} = 1. \tag{4.47}$$

Proof. By (4.44b), $\boldsymbol{\mathcal{R}}_L\boldsymbol{\mathcal{R}} = \boldsymbol{\mathcal{R}}_R$, and then multiple times (4.44a) we have

$$\begin{aligned}
\|\boldsymbol{\mathcal{R}}\mathbf{u}_h\|_{\mu\times\varepsilon}^2 &= \big(\boldsymbol{\mathcal{R}}_L\boldsymbol{\mathcal{R}}\mathbf{u}_h, \boldsymbol{\mathcal{R}}\mathbf{u}_h\big)_{\mu\times\varepsilon} \\
&= \big(\boldsymbol{\mathcal{R}}_R\mathbf{u}_h, \boldsymbol{\mathcal{R}}_L^{-1}\boldsymbol{\mathcal{R}}_R\mathbf{u}_h\big)_{\mu\times\varepsilon} \\
&= \big(\mathbf{u}_h, \boldsymbol{\mathcal{R}}_R\mathbf{u}_h\big)_{\mu\times\varepsilon} \\
&= \big(\boldsymbol{\mathcal{R}}_L\mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} \\
&= \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2.
\end{aligned}$$

This is the desired equality. □

As a consequence of this lemma the Crank–Nicolson method inherits the properties of the central fluxes semidiscrete approximation given in Theorem 3.4: In the following two corollaries we show that the Crank–Nicolson method is energy preserving and that it is stable with a bound analog to (3.10a).

Corollary 4.12. *For vanishing source term $\mathbf{J}_h \equiv 0$, the approximation obtained from the central fluxes dG discretization and the Crank–Nicolson method (4.40) conserves the electromagnetic energy, i.e.,*

$$\mathcal{E}(\mathbf{H}_h^n, \mathbf{E}_h^n) = \mathcal{E}(\mathbf{H}_h^0, \mathbf{E}_h^0), \qquad n = 1, 2, \dots\ .$$

Proof. For $\mathbf{J}_h \equiv 0$, we have $\mathbf{u}_h^n = \boldsymbol{\mathcal{R}}^n\mathbf{u}_h^0$, see (4.46b). The statement follows with Lemma 4.11, since $\mathcal{E}(\mathbf{H}_h, \mathbf{E}_h) = \frac{1}{2}\|\mathbf{u}_h\|_{\mu\times\varepsilon}^2$. □

Corollary 4.13. *The approximation $\mathbf{u}_h^n$ obtained from the central fluxes dG discretization and the Crank–Nicolson method (4.40) is bounded by*

$$\|\mathbf{u}_h^n\|_{\mu\times\varepsilon} \le \|\mathbf{u}^0\|_{\mu\times\varepsilon} + \frac{\tau}{2\sqrt{\delta}}\sum_{m=0}^{n-1}\|\mathbf{J}^{m+1} + \mathbf{J}^m\|, \qquad n = 1, 2, \dots\ . \tag{4.48}$$

Proof. Taking the norm of (4.46b) and using the triangle inequality, (4.44c), (4.47) and (3.12) gives the statement. □

Now, we turn to the Verlet method. In contrary to the operator $\boldsymbol{\mathcal{R}}_L$ from the Crank–Nicolson method, the operator $\widehat{\boldsymbol{\mathcal{R}}}_L$ associated with the Verlet method is not unconditionally invertible. In fact, we need to ensure the following condition to guarantee its invertibility: Let $0 < \widehat{\theta} < 1$ be an arbitrary but fixed parameter. Then, the **CFL condition of the Verlet method** reads

$$\tau \leq \frac{2\widehat{\theta}}{C_{\mathrm{bnd}} c_\infty} \min_{K \in \mathcal{T}_h} h_K, \tag{4.49}$$

where C_{bnd} was defined in Theorem 3.14 and c_∞ is given by $c_\infty = \max_{K\in\mathcal{T}_h} c_K$. The next lemma states that if (4.49) is satisfied $\widehat{\boldsymbol{\mathcal{R}}}_L$ is invertible and $\big(\widehat{\boldsymbol{\mathcal{R}}}_L \cdot, \cdot\big)$ defines a norm which is equivalent to the weighted L^2-norm $\|\cdot\|_{\mu\times\varepsilon}$ (where one of the constants depends on $\widehat{\theta}$).

Lemma 4.14. *Let $\mathbf{u}_h \in V_h^2$ and assume that the CFL condition (4.49) is satisfied with a $\widehat{\theta} \in (0,1)$. Then, we have*

$$(1-\widehat{\theta}^2)\|\mathbf{u}_h\|_{\mu\times\varepsilon}^2 \leq \big(\widehat{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} \leq \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2. \tag{4.50}$$

In particular, $\widehat{\boldsymbol{\mathcal{R}}}_L$ is invertible with bound

$$\|\widehat{\boldsymbol{\mathcal{R}}}_L^{-1}\|_{\mu\times\varepsilon} \leq C_{\mathrm{stb}}, \qquad C_{\mathrm{stb}} = (1-\widehat{\theta}^2)^{-1}. \tag{4.51}$$

Proof. The upper bound in (4.50) follows immediatly from (4.45b). For the lower bound we use Theorem 3.14 and the CFL condition (4.49) to infer

$$\frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h\|_\mu^2 \leq \frac{\tau^2}{4} C_{\mathrm{bnd}}^2 c_\infty^2 \|\mathbf{E}_h\|_{\varepsilon,\mathcal{T}_h,2,-1}^2 \leq \widehat{\theta}^2 \|\mathbf{E}_h\|_\varepsilon^2 \leq \widehat{\theta}^2 \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2.$$

Together with (4.45b) this proves (4.50).

In order to bound $\widehat{\boldsymbol{\mathcal{R}}}_L^{-1}$ we proceed as for the Crank–Nicolson scheme. In fact, we have

$$\|\widehat{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h\|_{\mu\times\varepsilon} = \sup_{\mathbf{v}_h \in V_h^2} \frac{\big(\widehat{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h, \mathbf{v}_h\big)_{\mu\times\varepsilon}}{\|\mathbf{v}_h\|_{\mu\times\varepsilon}} \geq \frac{\big(\widehat{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon}}{\|\mathbf{u}_h\|_{\mu\times\varepsilon}} \geq (1-\widehat{\theta}^2)\|\mathbf{u}_h\|_{\mu\times\varepsilon}.$$

Hence, $\widehat{\boldsymbol{\mathcal{R}}}_L$ is an isomorphism on V_h^2. Setting $\mathbf{v}_h = \widehat{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h$ proves (4.51). □

This lemma enables us to write the Verlet method (4.42a) as

$$\mathbf{u}_h^{n+1} = \widehat{\boldsymbol{\mathcal{R}}}^{n+1}\mathbf{u}_h^0 + \frac{\tau}{2}\sum_{m=0}^{n} \widehat{\boldsymbol{\mathcal{R}}}^{n-m}\widehat{\boldsymbol{\mathcal{R}}}_L^{-1}\big(\mathbf{j}_h^{m+1} + \mathbf{j}_h^m\big), \qquad \widehat{\boldsymbol{\mathcal{R}}} = \widehat{\boldsymbol{\mathcal{R}}}_L^{-1}\widehat{\boldsymbol{\mathcal{R}}}_R, \tag{4.52}$$

if the time step τ satisfies the CFL condition (4.49). Analogously to the bound (4.47) for the Crank–Nicolson method, we need a bound on powers of $\widehat{\boldsymbol{\mathcal{R}}}$.

Lemma 4.15. *Assume that the CFL condition (4.49) is satisfied with a $\widehat{\theta} \in (0,1)$. Then, for all $m \in \mathbb{N}$ and for all $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h) \in V_h^2$ we have the bound*

$$\|\widehat{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h\|_{\mu\times\varepsilon}^2 \leq C_{\mathrm{stb}}\Big(\|\mathbf{u}_h\|_{\mu\times\varepsilon}^2 - \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h\|_\mu^2\Big). \tag{4.53a}$$

In particular, it holds that

$$\|\widehat{\boldsymbol{\mathcal{R}}}^m\|_{\mu\times\varepsilon} \leq C_{\mathrm{stb}}^{1/2}. \tag{4.53b}$$

Note that contrary to the bound (4.47) for the Crank–Nicolson method, the bound (4.53) depends on the CFL parameter $\widehat{\theta}$. Moreover, we see that this bound cannot hold true if the CFL condition is harmed since $C_{\mathrm{stb}} \to \infty$ for $\widehat{\theta} \nearrow 1$.

Proof. As in the proof of Lemma 4.11, an induction argument shows

$$\big(\widehat{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} = \big(\widehat{\boldsymbol{\mathcal{R}}}_L \widehat{\boldsymbol{\mathcal{R}}} \mathbf{u}_h, \widehat{\boldsymbol{\mathcal{R}}} \mathbf{u}_h\big)_{\mu\times\varepsilon} = \ldots = \big(\widehat{\boldsymbol{\mathcal{R}}}_L \widehat{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h, \widehat{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h\big)_{\mu\times\varepsilon}, \qquad m = 1, 2, \ldots \,.$$

Together with (4.50) and (4.45b) this implies

$$(1-\widehat{\theta}^2)\|\widehat{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h\|_{\mu\times\varepsilon}^2 \le \big(\widehat{\boldsymbol{\mathcal{R}}}_L \widehat{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h, \widehat{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h\big)_{\mu\times\varepsilon} = \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2 - \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}} \mathbf{E}_h\|_{\mu}^2,$$

$m = 1, 2, \ldots$, which completes the proof. □

In the next corollary we prove that the Verlet method preserves a perturbed electromagnetic energy.

Corollary 4.16. *Assume that the CFL condition (4.49) is satisfied with parameter $\widehat{\theta} \in (0,1)$. Then, for $\mathbf{J}_h \equiv 0$, the approximation $\mathbf{u}_h^n = (\mathbf{H}_h^n, \mathbf{E}_h^n)$ obtained from the scheme (4.39) conserves the perturbed electromagnetic energy*

$$\widehat{\mathcal{E}}(\mathbf{H}_h, \mathbf{E}_h) = \mathcal{E}(\mathbf{H}_h, \mathbf{E}_h) - \frac{\tau^2}{8}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}} \mathbf{E}_h\|_{\mu}^2, \tag{4.54}$$

i.e., $\widehat{\mathcal{E}}(\mathbf{H}_h^n, \mathbf{E}_h^n) = \widehat{\mathcal{E}}(\mathbf{H}_h^0, \mathbf{E}_h^0)$, $n = 1, 2, \ldots$.

Proof. For $\mathbf{J}_h \equiv 0$ the Verlet method reads $\mathbf{u}_h^n = \widehat{\boldsymbol{\mathcal{R}}}^n \mathbf{u}_h^0$, see (4.52). Thus, the proof of the previous lemma shows that

$$\big(\widehat{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h^n, \mathbf{u}_h^n\big)_{\mu\times\varepsilon} = \big(\widehat{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h^0, \mathbf{u}_h^0\big)_{\mu\times\varepsilon}.$$

The statement then follows from (4.45b). □

We conclude this section with the stability result for the Verlet method.

Corollary 4.17. *Assume that the CFL condition (4.49) is satisfied with parameter $\widehat{\theta} \in (0,1)$. Then, the approximation $\mathbf{u}_h^n$ obtained from the Verlet method (4.39) is bounded by*

$$\|\mathbf{u}_h^n\|_{\mu\times\varepsilon} \le C_{\mathrm{stb}}^{1/2}\|\mathbf{u}^0\|_{\mu\times\varepsilon} + C_{\mathrm{stb}}^{3/2}\frac{\tau}{2\sqrt{\delta}}\sum_{m=0}^{n-1}\|\mathbf{J}^{m+1} + \mathbf{J}^m\|. \tag{4.55}$$

Proof. Taking the norm of (4.52) and using the triangle inequality, (4.51), (4.53) and (3.12) gives the statement. □

4.2.2 Full discretization errors

Let $\mathbf{u}^n = (\mathbf{H}^n, \mathbf{E}^n) = (\mathbf{H}(t_n), \mathbf{E}(t_n))$ be the exact solution of (4.1) at time t_n and denote by $\mathbf{u}_h^n = (\mathbf{H}_h^n, \mathbf{E}_h^n) \approx \mathbf{u}^n$ the approximation obtained by the central fluxes dG discretization in combination with the Verlet method (4.39) or with the Crank–Nicolson method (4.40). The full discretization error is given by

$$\mathbf{e}^n = \begin{pmatrix} \mathbf{e}_{\mathbf{H}}^n \\ \mathbf{e}_{\mathbf{E}}^n \end{pmatrix} = \begin{pmatrix} \mathbf{H}^n - \mathbf{H}_h^n \\ \mathbf{E}^n - \mathbf{E}_h^n \end{pmatrix}. \tag{4.56a}$$

As in Chapter 3 we split it into

$$\mathbf{e}^n = \mathbf{e}_\pi^n - \mathbf{e}_h^n = \begin{pmatrix} \mathbf{H}^n - \pi_h \mathbf{H}^n \\ \mathbf{E}^n - \pi_h \mathbf{E}^n \end{pmatrix} - \begin{pmatrix} \mathbf{H}_h^n - \pi_h \mathbf{H}^n \\ \mathbf{E}_h^n - \pi_h \mathbf{E}^n \end{pmatrix}. \tag{4.56b}$$

So, $\mathbf{e}_\pi^n$ contains the projection error and $\mathbf{e}_h^n$ contains the dG error and the time integration error. The projection error has already been studied in Chapter 3, cf. (3.24a) and (3.43). Hence, we can focus on $\mathbf{e}_h^n$. In the next lemma we prove that $\mathbf{e}_h^n$ satisfies a perturbed version of the Crank–Nicolson recursion (4.41a).

Lemma 4.18. *Let $\mathbf{u} \in C\big(0,T;V_\star\big) \cap C^3\big(0,T;L^2(\Omega)^6\big)$ be the exact solution of (4.1). The error $\mathbf{e}_h^n$ defined in (4.56b) satisfies*

$$\boldsymbol{\mathcal{R}}_L \mathbf{e}_h^{n+1} = \boldsymbol{\mathcal{R}}_R \mathbf{e}_h^n + \mathbf{d}^n, \tag{4.57}$$

if we employ the Crank–Nicolson method. The defect $\mathbf{d}^n = \mathbf{d}_\pi^n + \mathbf{d}_h^n$ is given by

$$\mathbf{d}_\pi^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}(\mathbf{e}_\pi^{n+1} + \mathbf{e}_\pi^n), \qquad \mathbf{d}_h^n = -\tau^2 \pi_h \delta^n(\partial_t \mathbf{u}), \tag{4.58}$$

where δ^n denotes the quadrature error of the trapezoidal rule given in (4.24).

Proof. The defects are obtained by inserting the projected exact solution into the numerical scheme (4.40). This yields

$$\pi_h(\mathbf{u}^{n+1} - \mathbf{u}^n) = \frac{\tau}{2}\boldsymbol{\mathcal{C}}\pi_h(\mathbf{u}^{n+1} + \mathbf{u}^n) + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) - \mathbf{d}^n. \tag{4.59}$$

Subtracting this equation from (4.40) proves (4.57).

It remains to determine the defect $\mathbf{d}^n$. By (4.23) we have

$$\mathbf{u}^{n+1} - \mathbf{u}^n = \frac{\tau}{2}(\partial_t \mathbf{u}^{n+1} + \partial_t \mathbf{u}^n) + \tau^2 \delta^n(\partial_t \mathbf{u}). \tag{4.60}$$

Moreover, (3.27) shows that $\mathbf{u}$ satisfies

$$\pi_h \partial_t \mathbf{u}(t) = \boldsymbol{\mathcal{C}}\mathbf{u}(t) + \mathbf{j}_h(t).$$

Projecting (4.60) onto V_h^2 and inserting the last identity, we infer

$$\pi_h(\mathbf{u}^{n+1} - \mathbf{u}^n) = \frac{\tau}{2}\boldsymbol{\mathcal{C}}(\mathbf{u}^{n+1} + \mathbf{u}^n) + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) + \tau^2 \pi_h \delta^n(\partial_t \mathbf{u}). \tag{4.61}$$

Together with (4.59) this yields the desired representation (4.58). □

Next, we give the error recursion for the Verlet method.

Lemma 4.19. *Under the assumptions of Lemma 4.18 the error $\mathbf{e}_h^n$ satisfies*

$$\widehat{\boldsymbol{\mathcal{R}}}_L \mathbf{e}_h^{n+1} = \widehat{\boldsymbol{\mathcal{R}}}_R \mathbf{e}_h^n + \widehat{\mathbf{d}}^n, \tag{4.62}$$

if we use the Verlet method as a time integrator. The defect $\widehat{\mathbf{d}}^n = \widehat{\mathbf{d}}_\pi^n + \widehat{\mathbf{d}}_h^n$ is given by

$$\widehat{\mathbf{d}}_\pi^n = \mathbf{d}_\pi^n - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n) \end{pmatrix}, \tag{4.63a}$$

and

$$\widehat{\mathbf{d}}_h^n = \mathbf{d}_h^n - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}\pi_h \Delta_{\mathbf{H}}^n \end{pmatrix}, \qquad \Delta_{\mathbf{H}}^n = \partial_t \mathbf{H}^{n+1} - \partial_t \mathbf{H}^n = \int_{t_n}^{t_{n+1}} \partial_t^2 \mathbf{H}(t)\, dt. \tag{4.63b}$$

Proof. The identity (4.59) for the projected exact solution is equivalent to

$$\boldsymbol{\mathcal{R}}_L \pi_h \mathbf{u}^{n+1} = \boldsymbol{\mathcal{R}}_R \pi_h \mathbf{u}^n + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) - \mathbf{d}^n. \tag{4.64}$$

Inserting again the projected exact solution into the Verlet scheme yields

$$\widehat{\boldsymbol{\mathcal{R}}}_L \pi_h \mathbf{u}^{n+1} = \widehat{\boldsymbol{\mathcal{R}}}_R \pi_h \mathbf{u}^n + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) - \widehat{\mathbf{d}}^n. \tag{4.65}$$

Subtracting both equations and using (4.42b), we thus have

$$\begin{aligned}\widehat{\mathbf{d}}^n &= \mathbf{d}^n + \frac{\tau^2}{4}\boldsymbol{\mathcal{D}}\pi_h(\mathbf{u}^{n+1} - \mathbf{u}^n),\\ &= \mathbf{d}^n + \frac{\tau^2}{4}\boldsymbol{\mathcal{D}}\big(\mathbf{u}^{n+1} - \mathbf{u}^n - (\mathbf{e}_\pi^{n+1} - \mathbf{e}_\pi^n)\big),\end{aligned}$$

whose components read

$$\widehat{\mathbf{d}}_{\mathbf{H}}^n = \mathbf{d}_{\mathbf{H}}^n, \qquad \widehat{\mathbf{d}}_{\mathbf{E}}^n = \mathbf{d}_{\mathbf{E}}^n + \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{E}^{n+1} - \mathbf{E}^n - (\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n)\big).$$

By the consistency of the discrete curl-operators, cf. (3.6), we can write

$$\boldsymbol{\mathcal{C}}_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{E}^{n+1} - \mathbf{E}^n) = \boldsymbol{\mathcal{C}}_{\mathbf{H}}\pi_h \mathcal{C}_{\mathbf{E}}(\mathbf{E}^{n+1} - \mathbf{E}^n) = -\boldsymbol{\mathcal{C}}_{\mathbf{H}}\pi_h(\partial_t \mathbf{H}^{n+1} - \partial_t \mathbf{H}^n) = -\boldsymbol{\mathcal{C}}_{\mathbf{H}}\pi_h \Delta_{\mathbf{H}}^n.$$

Here, the second equality is obtained via Maxwell's equations (1.21), in particular by differentiating $\partial_t \mathbf{H} = -\mathcal{C}_{\mathbf{E}}\mathbf{E}$ w.r.t. t. This yields

$$\widehat{\mathbf{d}}^n = \mathbf{d}^n - \frac{\tau^2}{4}\begin{pmatrix}0\\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}\pi_h \Delta_{\mathbf{H}}^n\end{pmatrix} - \frac{\tau^2}{4}\begin{pmatrix}0\\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big)\end{pmatrix},$$

which completes the proof. □

Solving the error recursions for the Crank–Nicolson method and the Verlet method, respectively, while exploiting $\mathbf{e}_h^0 = 0$, shows that the errors satisfy

$$\text{Crank–Nicolson}: \quad \mathbf{e}_h^{n+1} = \sum_{m=0}^{n} \boldsymbol{\mathcal{R}}^{n-m}\boldsymbol{\mathcal{R}}_L^{-1}\mathbf{d}^m, \tag{4.66}$$

and, under the CFL condition (4.49),

$$\text{Verlet}: \quad \mathbf{e}_h^{n+1} = \sum_{m=0}^{n} \widehat{\boldsymbol{\mathcal{R}}}^{n-m}\widehat{\boldsymbol{\mathcal{R}}}_L^{-1}\widehat{\mathbf{d}}^m. \tag{4.67}$$

Since we already established bounds on $\boldsymbol{\mathcal{R}}^m$, $\boldsymbol{\mathcal{R}}_L^{-1}$, $\widehat{\boldsymbol{\mathcal{R}}}^m$ and $\widehat{\boldsymbol{\mathcal{R}}}_L^{-1}$, it remains to prove bounds on the defects $\mathbf{d}^m$ and $\widehat{\mathbf{d}}^m$.

Lemma 4.20. *Let $\mathbf{u} \in C\big(0,T;D(\mathcal{C}) \cap H^{k+1}(\mathcal{T}_h)^6\big) \cap C^3\big(0,T;L^2(\Omega)^6\big)$ be the exact solution of (4.1). Then, the following bounds hold true,*

$$\|\mathbf{d}_\pi^n\|_{\mu\times\varepsilon} \le \widehat{C}_\pi \frac{\tau}{2}|\mathbf{u}^{n+1} + \mathbf{u}^n|_{k+1,\mathcal{T}_h,2,k}, \qquad \|\mathbf{d}_h^n\|_{\mu\times\varepsilon} \le \frac{\tau^2}{8}\int_{t_n}^{t_{n+1}} \|\partial_t^3 \mathbf{u}(t)\|_{\mu\times\varepsilon}\, dt. \tag{4.68a}$$

Moreover, if the CFL condition (4.49) is fulfilled, it holds that

$$\|\widehat{\mathbf{d}}_\pi^n\|_{\mu\times\varepsilon} \le \widehat{C}_\pi \frac{\tau}{2}\big(|\mathbf{u}^{n+1} + \mathbf{u}^n|_{k+1,\mathcal{T}_h,2,k} + |\mathbf{E}^{n+1} - \mathbf{E}^n|_{k+1,\mathcal{T}_h,2,k}\big). \tag{4.68b}$$

If we assume more regularity for $\mathbf{H}$, in particular $\mathbf{H} \in C^2(0,T;V_\star^{\mathbf{H}})$, we obtain,

$$\|\widehat{\mathbf{d}}_h^n\|_{\mu\times\varepsilon} \le \frac{\tau^2}{8}\int_{t_n}^{t_{n+1}} \Big(\|\partial_t^3 \mathbf{u}(t)\|_{\mu\times\varepsilon} + \frac{2}{\sqrt{\delta}}\|\partial_t^2 \mathbf{H}(t)\|_{H(\mathrm{curl},\Omega)} + 2\widehat{C}_\pi |\partial_t^2 \mathbf{H}(t)|_{1,\mathcal{T}_h,2}\Big)\, dt. \tag{4.68c}$$

Proof. (a) Using (4.58), the bound on the projection defect $\mathbf{d}_\pi^n$ follows from (3.43) and the bound on $\mathbf{d}_h^n$ from (4.24b).

(b) For the bound (4.68b) observe that $\boldsymbol{\mathcal{C}}_\mathbf{E}\big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big) \in V_h$ and consequently we can apply Theorem 3.14. This gives

$$\begin{aligned}\frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_\mathbf{H}\boldsymbol{\mathcal{C}}_\mathbf{E}\big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big)\|_\varepsilon &\le \frac{\tau^2}{4} C_{\mathrm{bnd}} c_\infty \|\boldsymbol{\mathcal{C}}_\mathbf{E}\big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big)\|_{\mu,\mathcal{T}_h,2,-1} \\ &\le \frac{\tau}{2}\widehat{\theta}\|\boldsymbol{\mathcal{C}}_\mathbf{E}\big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big)\|_\mu \\ &\le \widehat{C}_\pi \frac{\tau}{2}|\mathbf{E}^{n+1} - \mathbf{E}^n|_{k+1,\mathcal{T}_h,2,k}, \end{aligned} \tag{4.69}$$

where the last inequality follows from (3.29b).

(c) In order to prove (4.68c) we decompose $\pi_h\Delta_\mathbf{H}^n = \Delta_\mathbf{H}^n - \Delta_\pi^n$, where $\Delta_\mathbf{H}^n$ is given by (4.63b) and Δ_π^n is defined as

$$\Delta_\pi^n = \Delta_\mathbf{H}^n - \pi_h\Delta_\mathbf{H}^n = \int_{t_n}^{t_{n+1}} \partial_t^2 \mathbf{e}_{\pi,\mathbf{H}}(t)\, dt.$$

By the regularity assumption on $\mathbf{H}$ we have $n_F \times [\![\partial_t^2\mathbf{H}]\!]_F = 0$ for all $F \in \mathcal{F}_h^{\mathrm{int}}$ and thus the strong form (3.5a) of the discrete curl-operator implies

$$\big(\boldsymbol{\mathcal{C}}_\mathbf{H}(\partial_t^2\mathbf{H}), \psi_h\big)_\varepsilon = \sum_{K\in\mathcal{T}_h} \big(\mathrm{curl}(\partial_t^2\mathbf{H}), \psi_h\big)_K \le \frac{1}{\delta^{1/2}}\|\partial_t^2\mathbf{H}\|_{H(\mathrm{curl},\Omega)}\|\psi_h\|_\varepsilon.$$

This shows

$$\|\boldsymbol{\mathcal{C}}_\mathbf{H}(\partial_t^2\mathbf{H})\|_\varepsilon \le \frac{1}{\delta^{1/2}}\|\partial_t^2\mathbf{H}\|_{H(\mathrm{curl},\Omega)},$$

and

$$\|\boldsymbol{\mathcal{C}}_\mathbf{H}\Delta_\mathbf{H}^n\|_\varepsilon \le \frac{1}{\delta^{1/2}}\int_{t_n}^{t_{n+1}} \|\partial_t^2\mathbf{H}(t)\|_{H(\mathrm{curl},\Omega)}\, dt.$$

Finally, by (3.29b) for $k = 0$ we have

$$\|\boldsymbol{\mathcal{C}}_\mathbf{H}\Delta_\pi^n\|_\varepsilon \le \int_{t_n}^{t_{n+1}} \|\boldsymbol{\mathcal{C}}_\mathbf{H}(\partial_t^2\mathbf{e}_{\pi,\mathbf{H}}(t))\|_\varepsilon\, dt \le \widehat{C}_\pi \int_{t_n}^{t_{n+1}} |\partial_t^2\mathbf{H}(t)|_{1,\mathcal{T}_h,2}\, dt.$$

This completes the proof. □

With the bounds of Lemma 4.20 at hand we can already prove the **fully discrete convergence result** for the **Crank–Nicolson method**.

Theorem 4.21. *Let $\mathbf{u} \in C\big(0,T;D(\mathcal{C})\cap H^{k+1}(\mathcal{T}_h)^6\big)\cap C^3\big(0,T;L^2(\Omega)^6\big)$ be the exact solution of (4.1). Then, the error of the central fluxes dG discretization and the Crank–Nicolson scheme (4.40) satisfies*

$$\begin{aligned}\|\mathbf{u}^n - \mathbf{u}_h^n\|_{\mu\times\varepsilon} &\le C_{\mathrm{app}}|\mathbf{u}^n|_{k+1,\mathcal{T}_h,1,k+1} \\ &\quad + \widehat{C}_\pi\frac{\tau}{2}\sum_{m=0}^{n-1}|\mathbf{u}^{m+1} + \mathbf{u}^m|_{k+1,\mathcal{T}_h,2,k} + \frac{\tau^2}{8}\int_0^{t_n}\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}\, dt. \\ &\le C\Big(h^k + \tau^2\Big).\end{aligned}$$

The constant C only depends on C_{app}, $\widehat{C}_\pi$, $|\mathbf{u}(t)|_{k+1,\mathcal{T}_h}$, and $\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}$, $t \in [0,t_n]$.

Proof. Our proof is a discrete counterpart to the convergence proof of Theorem 3.12 for the semidiscrete central fluxes discretization with the following differences: Instead of the bound on the semigroup $\|e^{(t_{n+1}-s)\boldsymbol{\mathcal{C}}}\|_{\mu\times\varepsilon} \leq 1$ we now use the operator bound $\|\boldsymbol{\mathcal{R}}^{n-m}\boldsymbol{\mathcal{R}}_L^{-1}\|_{\mu\times\varepsilon} \leq 1$, cf. (4.44c) and (4.47). The time-integral over the defect $\int_0^{t_{n+1}} \boldsymbol{\mathcal{C}}\mathbf{e}_\pi(t)\,dt$ is replaced by the discrete integral $\frac{\tau}{2}\sum_{m=0}^{n} \boldsymbol{\mathcal{C}}(\mathbf{e}_\pi^{m+1} + \mathbf{e}_\pi^m)$. In addition, the full discretization error now involves the quadrature error $\mathbf{d}_h^n$.

We take the norm of (4.66) and apply the triangle inequality, and use $\mathbf{d}^n = \mathbf{d}_\pi^m + \mathbf{d}_h^m$ to obtain

$$\begin{aligned}\|\mathbf{e}_h^{n+1}\|_{\mu\times\varepsilon} \leq \sum_{m=0}^{n} \|\mathbf{d}^m\|_{\mu\times\varepsilon} &\leq \sum_{m=0}^{n} \left(\|\mathbf{d}_\pi^m\|_{\mu\times\varepsilon} + \|\mathbf{d}_h^m\|_{\mu\times\varepsilon}\right) \\ &\leq \widehat{C}_\pi \frac{\tau}{2} \sum_{m=0}^{n} |\mathbf{u}^{m+1} + \mathbf{u}^m|_{k+1,\mathcal{T}_h,2,k} + \frac{\tau^2}{8} \sum_{m=0}^{n} \int_{t_m}^{t_{m+1}} \|\partial_t^3 \mathbf{u}(t)\|_{\mu\times\varepsilon}\,dt,\end{aligned}$$

where the bounds on the defects were taken from Lemma 4.20. For the full discretization error recall $\mathbf{e}^n = \mathbf{e}_\pi^n - \mathbf{e}_h^n$ and use (3.24a) for the projection error. □

It is possible to prove an analogo convergence result for the Verlet method based on the bounds of Lemma 4.20. However, we would like to stress that we can relax the regularity assumption for $\mathbf{H}$ which we used to prove (4.68c). The different technique for this proof is mandatory for the locally implicit time integrator we consider in the next chapter since a result like (4.68c) is not available in this case. A key observation is that for all $\mathbf{H}_h \in V_h$ we have that

$$\begin{pmatrix} 0 \\ -\tau\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h \end{pmatrix} = \begin{pmatrix} 0 & \tau\boldsymbol{\mathcal{C}}_{\mathbf{E}} \\ -\tau\boldsymbol{\mathcal{C}}_{\mathbf{H}} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{H}_h \\ 0 \end{pmatrix} = -\tau\boldsymbol{\mathcal{C}} \begin{pmatrix} \mathbf{H}_h \\ 0 \end{pmatrix} = (\widehat{\boldsymbol{\mathcal{R}}}_L - \widehat{\boldsymbol{\mathcal{R}}}_R) \begin{pmatrix} \mathbf{H}_h \\ 0 \end{pmatrix}.$$

Now, consider the defect $\widehat{\mathbf{d}}^n = \widehat{\mathbf{d}}_\pi^n + \widehat{\mathbf{d}}_h^n$ defined in Lemma 4.19. Using the previous identity, we can write

$$\widehat{\mathbf{d}}_h^n = \mathbf{d}_h^n - \frac{\tau^2}{4} \begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}\pi_h\Delta_{\mathbf{H}}^n \end{pmatrix} = \mathbf{d}_h^n + (\widehat{\boldsymbol{\mathcal{R}}}_L - \widehat{\boldsymbol{\mathcal{R}}}_R)\widehat{\boldsymbol{\xi}}^n, \qquad \widehat{\boldsymbol{\xi}}^n = \begin{pmatrix} \widehat{\boldsymbol{\xi}}_{\mathbf{H}}^n \\ \widehat{\boldsymbol{\xi}}_{\mathbf{E}}^n \end{pmatrix} = \frac{\tau}{4} \begin{pmatrix} \pi_h\Delta_{\mathbf{H}}^n \\ 0 \end{pmatrix}. \tag{4.70a}$$

This enables us to split the defect further into

$$\widehat{\mathbf{d}}^n = \widehat{\boldsymbol{\eta}}^n + (\widehat{\boldsymbol{\mathcal{R}}}_L - \widehat{\boldsymbol{\mathcal{R}}}_R)\widehat{\boldsymbol{\xi}}^n, \qquad \widehat{\boldsymbol{\eta}}^n = \widehat{\mathbf{d}}_\pi^n + \mathbf{d}_h^n. \tag{4.70b}$$

The advantage of this splitting is that $\widehat{\boldsymbol{\eta}}^n$ can be bounded by Lemma 4.20 and that we can exploit that the error recursion involves terms of the form

$$\widehat{\boldsymbol{\mathcal{R}}}_L^{-1}\widehat{\mathbf{d}}^n = \widehat{\boldsymbol{\mathcal{R}}}_L^{-1}\widehat{\boldsymbol{\eta}}^n + (\boldsymbol{\mathcal{I}} - \widehat{\boldsymbol{\mathcal{R}}})\widehat{\boldsymbol{\xi}}^n. \tag{4.71}$$

This is detailed in the following **fully discrete convergence result** for the **Verlet method**.

Theorem 4.22. *Let $\mathbf{u} \in C\big(0,T; D(\mathcal{C}) \cap H^{k+1}(\mathcal{T}_h)^6\big) \cap C^3\big(0,T; L^2(\Omega)^6\big)$ be the exact solution of (4.1). Moreover, assume that the CFL condition (4.49) is satisfied with $\widehat{\theta} \in (0,1)$. Then, the error of the central fluxes dG discretization and the Verlet scheme (4.39) satisfies*

$$\begin{aligned}\|\mathbf{u}^n - \mathbf{u}_h^n\|_{\mu\times\varepsilon} \leq\; & C_{\text{app}}|\mathbf{u}^n|_{k+1,\mathcal{T}_h,1,k+1} \\ &+ C_{\text{stb}}^{3/2}\widehat{C}_\pi \frac{\tau}{2} \sum_{m=0}^{n-1} \Big(|\mathbf{u}^{m+1} + \mathbf{u}^m|_{k+1,\mathcal{T}_h,2,k} + |\mathbf{E}^{m+1} - \mathbf{E}^m|_{k+1,\mathcal{T}_h,2,k}\Big) \qquad (4.72a) \\ &+ (1 + C_{\text{stb}}^{1/2})\frac{\tau^2}{4} \max_{t\in[0,t_n]} \|\partial_t^2 \mathbf{H}(t)\|_\mu \\ &+ C_{\text{stb}}^{1/2}(4 + C_{\text{stb}})\frac{\tau^2}{8} \int_0^{t_n} \|\partial_t^3 \mathbf{u}(t)\|_{\mu\times\varepsilon}\,dt \\ \leq\; & C\Big(h^k + \tau^2\Big). \qquad (4.72b)\end{aligned}$$

The constant C only depends on C_{app}, $\widehat{C}_\pi$, $\widehat{\theta}$, $|\mathbf{u}(t)|_{k+1,\mathcal{T}_h}$, $\|\partial_t^2\mathbf{H}(t)\|_\mu$, and $\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}$, $t\in[0,t_n]$.

Proof. The proof is done in three steps.

(a) First, we rewrite the error recursion such that the terms involving $\widehat{\boldsymbol{\xi}}^m$ within a sum only appear as differences of consecutive values.

Employing (4.71) we obtain

$$\mathbf{e}_h^{n+1} = \sum_{m=0}^{n} \widehat{\boldsymbol{\mathcal{R}}}^{n-m}\widehat{\boldsymbol{\mathcal{R}}}_L^{-1}\widehat{\mathbf{d}}^m = \sum_{m=0}^{n} \widehat{\boldsymbol{\mathcal{R}}}^{n-m}\widehat{\boldsymbol{\mathcal{R}}}_L^{-1}\widehat{\boldsymbol{\eta}}^m + \sum_{m=0}^{n} \widehat{\boldsymbol{\mathcal{R}}}^{n-m}(\boldsymbol{\mathcal{I}}-\widehat{\boldsymbol{\mathcal{R}}})\widehat{\boldsymbol{\xi}}^m.$$

Using (4.34c) for the second sum shows

$$\mathbf{e}_h^{n+1} = \widehat{\boldsymbol{\xi}}^n - \widehat{\boldsymbol{\mathcal{R}}}^{n+1}\widehat{\boldsymbol{\xi}}^0 + \sum_{m=0}^{n} \widehat{\boldsymbol{\mathcal{R}}}^{n-m}\widehat{\boldsymbol{\mathcal{R}}}_L^{-1}\widehat{\boldsymbol{\eta}}^m - \sum_{m=0}^{n-1} \widehat{\boldsymbol{\mathcal{R}}}^{n-m}(\widehat{\boldsymbol{\xi}}^{m+1} - \widehat{\boldsymbol{\xi}}^m). \tag{4.73}$$

(b) Next, we prove a bound on $\widehat{\boldsymbol{\xi}}^{m+1} - \widehat{\boldsymbol{\xi}}^m$. By definition (4.63b) of $\Delta_{\mathbf{H}}^n$ we observe that

$$\widehat{\boldsymbol{\xi}}_{\mathbf{H}}^{n+1} - \widehat{\boldsymbol{\xi}}_{\mathbf{H}}^{n} = \frac{\tau}{4}\pi_h\big(\partial_t\mathbf{H}^{n+2} - 2\partial_t\mathbf{H}^{n+1} + \partial_t\mathbf{H}^n\big).$$

A Taylor expansion of $\partial_t\mathbf{H}^{n+1}$ at t_n and t_{n+2}, respectively, yields

$$\partial_t\mathbf{H}^{n+1} = \partial_t\mathbf{H}^{n+1\pm1} \mp \tau\partial_t^2\mathbf{H}^{n+1\pm1} + \int_{t_{n+1\pm1}}^{t_{n+1}} (t_{n+1}-t)\partial_t^3\mathbf{H}(t)dt.$$

Adding both equations implies

$$\begin{aligned}\widehat{\boldsymbol{\xi}}_{\mathbf{H}}^{n+1} - \widehat{\boldsymbol{\xi}}_{\mathbf{H}}^{n} &= \frac{\tau}{4}\pi_h\Big(\tau\partial_t^2\big(\mathbf{H}^{n+2}-\mathbf{H}^n\big) - \int_{t_n}^{t_{n+2}} |t_{n+1}-t|\,\partial_t^3\mathbf{H}(t)\;dt\Big)\\ &= \frac{\tau^2}{4}\int_{t_n}^{t_{n+2}}\left(1-\frac{|t_{n+1}-t|}{\tau}\right)\pi_h\partial_t^3\mathbf{H}(t)\;dt.\end{aligned}$$

Taking the norm, using the triangle inequality and observing that the kernel of the integral is bounded by 1 yields

$$\|\widehat{\boldsymbol{\xi}}_{\mathbf{H}}^{n+1} - \widehat{\boldsymbol{\xi}}_{\mathbf{H}}^{n}\|_\mu \le \frac{\tau^2}{4}\int_{t_n}^{t_{n+2}} \|\partial_t^3\mathbf{H}(t)\|_\mu\;dt. \tag{4.74}$$

(c) Finally, we combine the results of (a) and (b) by taking norms in the error recursion (4.73), using the triangle inequality, (4.51), and (4.53b). This yields

$$\|\mathbf{e}_h^n\|_{\mu\times\varepsilon} \le \|\widehat{\boldsymbol{\xi}}^{n-1}\|_{\mu\times\varepsilon} + C_{\mathrm{stb}}^{1/2}\|\widehat{\boldsymbol{\xi}}^0\|_{\mu\times\varepsilon} + C_{\mathrm{stb}}^{3/2}\sum_{m=0}^{n-1}\|\widehat{\boldsymbol{\eta}}^m\|_{\mu\times\varepsilon} + C_{\mathrm{stb}}^{1/2}\sum_{m=0}^{n-2}\|\widehat{\boldsymbol{\xi}}^{m+1}-\widehat{\boldsymbol{\xi}}^m\|_{\mu\times\varepsilon}.$$

Observe that by (4.70a) and (4.63b) the first two terms can be bounded by

$$\|\widehat{\boldsymbol{\xi}}^{n-1}\|_{\mu\times\varepsilon} \le \frac{\tau^2}{4}\max_{t\in[t_{n-1},t_n]}\|\partial_t^2\mathbf{H}(t)\|_\mu, \qquad \|\widehat{\boldsymbol{\xi}}^0\|_{\mu\times\varepsilon} \le \frac{\tau^2}{4}\max_{t\in[0,\tau]}\|\partial_t^2\mathbf{H}(t)\|_\mu. \tag{4.75}$$

By (4.70b) and Lemma 4.20 we have

$$\begin{aligned}\|\widehat{\boldsymbol{\eta}}^n\|_{\mu\times\varepsilon} &= \|\widehat{\mathbf{d}}_\pi^n + \mathbf{d}_h^n\|_{\mu\times\varepsilon}\\ &\le \widehat{C}_\pi\frac{\tau}{2}\big(|\mathbf{u}^{n+1}+\mathbf{u}^n|_{k+1,\mathcal{T}_h,2,k} + |\mathbf{E}^{n+1}-\mathbf{E}^n|_{k+1,\mathcal{T}_h,2,k}\big) + \frac{\tau^2}{8}\int_{t_n}^{t_{n+1}}\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}\;dt.\end{aligned}$$

Finally, (4.74) gives

$$\sum_{m=0}^{n-2} \|\widehat{\boldsymbol{\xi}}^{m+1} - \widehat{\boldsymbol{\xi}}^{m}\|_{\mu\times\varepsilon} \leq \frac{\tau^2}{4} \sum_{m=0}^{n-2} \int_{t_m}^{t_{m+2}} \|\partial_t^3 \mathbf{H}(t)\|_\mu \, dt \leq \frac{\tau^2}{2} \int_0^{t_n} \|\partial_t^3 \mathbf{H}(t)\|_\mu \, dt.$$

This proves the desired bound on $\mathbf{e}_h^n$. The stated bound on the full discretization error $\mathbf{e}^n = \mathbf{e}_\pi^n - \mathbf{e}_h^n$ is then obtained from the bounds (3.24a) for the projection error $\mathbf{e}_\pi^n$. □

4.3 Time integration for Maxwell's equations: upwind fluxes

In the previous section we used the Verlet method or the Crank–Nicolson method to integrate the semidiscrete Maxwell's equations (3.8) stemming from a central fluxes dG discretization. Now, we turn to the semidiscrete Maxwell's equations (3.15) arising from an upwind fluxes discretization. Since the Crank–Nicolson method is a RK scheme, it can be applied to every (first order) evolution equation. For the semidiscrete upwind fluxes Maxwell's equations it reads

$$\mathbf{u}_h^{n+1} - \mathbf{u}_h^n = \frac{\tau}{2}(\boldsymbol{\mathcal{C}} - \alpha\boldsymbol{\mathcal{S}})(\mathbf{u}_h^{n+1} + \mathbf{u}_h^n) + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n), \tag{4.76}$$

or, by using the opperators $\boldsymbol{\mathcal{R}}_L$ and $\boldsymbol{\mathcal{R}}_R$,

$$\boldsymbol{\mathcal{R}}_L \mathbf{u}_h^{n+1} - \boldsymbol{\mathcal{R}}_R \mathbf{u}_h^n = -\frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}(\mathbf{u}_h^{n+1} + \mathbf{u}_h^n) + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n). \tag{4.77}$$

On the contrary, the Verlet method is a time integration scheme designed for Hamiltonian systems. Regarding the upwind Maxwell's equations we see that they do not fit into this class. So, we need to adapt the Verlet method. A first idea could be to treat the stabilization operators as in the Crank–Nicolson method. This yields the scheme

$$\begin{aligned}
\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n &= -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^n - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h^n, \\
\mathbf{E}_h^{n+1} - \mathbf{E}_h^n &= \tau\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h^{n+1/2} - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{E}}(\mathbf{E}_h^{n+1} + \mathbf{E}_h^n) - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n), \\
\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n+1/2} &= -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1} - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h^{n+1}.
\end{aligned}$$

Observe that we end up with an implicit scheme which is not a desired property for a Verlet-type integrator. However, we can retrieve an explicit scheme by approximating the implicit terms by $\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}_h^{n+1} \approx \boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}_h^n$ and $\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h^{n+1} \approx \boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h^n$. This results in the scheme

$$\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^n - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h^n, \tag{4.78a}$$

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \tau\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h^{n+1/2} - \tau\alpha\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}_h^n - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n), \tag{4.78b}$$

$$\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n+1/2} = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1} - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h^n, \tag{4.78c}$$

which we will work with in this thesis. We note that related ideas have been presented in Alvarez et al. [2014] and Montseny et al. [2008] when working with the Verlet method on a staggered time grid. In the notation with the operators $\widehat{\boldsymbol{\mathcal{R}}}_L$ and $\widehat{\boldsymbol{\mathcal{R}}}_R$ the scheme (4.78) reads

$$\widehat{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h^{n+1} - \widehat{\boldsymbol{\mathcal{R}}}_R \mathbf{u}_h^n = -\tau\alpha\boldsymbol{\mathcal{S}}\mathbf{u}_h^n + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n). \tag{4.79}$$

4.3.1 Stability and energy dissipation

Contrary to Section 4.2.1 our stability analysis (and our later error analysis) are based on an energy technique. This is in accordance with the semidiscrete case, where we also used this technique, see Theorems 3.8 and 3.13. We start by giving an energy identity for the Crank–Nicolson and the Verlet method.

Lemma 4.23. *The approximation* $\mathbf{u}_h^n$ *obtained from the Crank–Nicolson method* (4.77) *satisfies*

$$\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n}|\mathbf{u}_h^{m+1}+\mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2 = \|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{\tau}{2}\sum_{m=0}^{n}\big(\mathbf{j}_h^{m+1}+\mathbf{j}_h^m, \mathbf{u}_h^{m+1}+\mathbf{u}_h^m\big)_{\mu\times\varepsilon}. \tag{4.80}$$

The approximation $\mathbf{u}_h^n = (\mathbf{H}_h^n, \mathbf{E}_h^n)$ *obtained from the Verlet method* (4.79) *satisfies*

$$\begin{aligned}&\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 - \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1}\|_\mu^2 - \alpha\frac{\tau}{2}|\mathbf{u}_h^{n+1}|_{\boldsymbol{\mathcal{S}}}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n}|\mathbf{u}_h^{m+1}+\mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2\\ &= \|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 - \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^0\|_\mu^2 - \alpha\frac{\tau}{2}|\mathbf{u}_h^0|_{\boldsymbol{\mathcal{S}}}^2 + \frac{\tau}{2}\sum_{m=0}^{n}\big(\mathbf{j}_h^{m+1}+\mathbf{j}_h^m, \mathbf{u}_h^{m+1}+\mathbf{u}_h^m\big)_{\mu\times\varepsilon}.\end{aligned} \tag{4.81}$$

Proof. (a) In order to prove the identity for the Crank–Nicolson method we take the $\mu\times\varepsilon$-inner product of (4.77) with $\mathbf{u}_h^{n+1}+\mathbf{u}_h^n$ and use the definition of $|\cdot|_{\boldsymbol{\mathcal{S}}}$, see (3.20), to obtain

$$\big(\boldsymbol{\mathcal{R}}_L\mathbf{u}_h^{n+1} - \boldsymbol{\mathcal{R}}_R\mathbf{u}_h^n, \mathbf{u}_h^{n+1}+\mathbf{u}_h^n\big)_{\mu\times\varepsilon} = -\frac{\tau}{2}\alpha|\mathbf{u}_h^{n+1}+\mathbf{u}_h^n|_{\boldsymbol{\mathcal{S}}}^2 + \frac{\tau}{2}\big(\mathbf{j}_h^{n+1}+\mathbf{j}_h^n, \mathbf{u}_h^{n+1}+\mathbf{u}_h^n\big)_{\mu\times\varepsilon}.$$

The adjointness of $\boldsymbol{\mathcal{R}}_L$ and $\boldsymbol{\mathcal{R}}_R$ given in (4.44a), and furthermore (4.44b) imply

$$\begin{aligned}\big(\boldsymbol{\mathcal{R}}_L\mathbf{u}_h^{n+1} - \boldsymbol{\mathcal{R}}_R\mathbf{u}_h^n, \mathbf{u}_h^{n+1}+\mathbf{u}_h^n\big)_{\mu\times\varepsilon} &= \big(\boldsymbol{\mathcal{R}}_L\mathbf{u}_h^{n+1}, \mathbf{u}_h^{n+1}\big)_{\mu\times\varepsilon} - \big(\boldsymbol{\mathcal{R}}_R\mathbf{u}_h^n, \mathbf{u}_h^n\big)_{\mu\times\varepsilon}\\ &= \|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 - \|\mathbf{u}_h^n\|_{\mu\times\varepsilon}^2.\end{aligned}$$

Thus, we conclude

$$\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 - \|\mathbf{u}_h^n\|_{\mu\times\varepsilon}^2 = -\frac{\tau}{2}\alpha|\mathbf{u}_h^{n+1}+\mathbf{u}_h^n|_{\boldsymbol{\mathcal{S}}}^2 + \frac{\tau}{2}\big(\mathbf{j}_h^{n+1}+\mathbf{j}_h^n, \mathbf{u}_h^{n+1}+\mathbf{u}_h^n\big)_{\mu\times\varepsilon},$$

and by summing this identity we obtain the desired result.

(b) By analog arguments we obtain for the Verlet method

$$\begin{aligned}&\big(\widehat{\boldsymbol{\mathcal{R}}}_L\mathbf{u}_h^{n+1}, \mathbf{u}_h^{n+1}\big)_{\mu\times\varepsilon} - \big(\widehat{\boldsymbol{\mathcal{R}}}_R\mathbf{u}_h^n, \mathbf{u}_h^n\big)_{\mu\times\varepsilon}\\ &\qquad = -\tau\alpha\big(\boldsymbol{\mathcal{S}}\mathbf{u}_h^n, \mathbf{u}_h^{n+1}+\mathbf{u}_h^n\big)_{\mu\times\varepsilon} + \frac{\tau}{2}\big(\mathbf{j}_h^{n+1}+\mathbf{j}_h^n, \mathbf{u}_h^{n+1}+\mathbf{u}_h^n\big)_{\mu\times\varepsilon}.\end{aligned} \tag{4.82}$$

Using the symmetry of $\boldsymbol{\mathcal{S}}$, we have

$$\begin{aligned}\big(\boldsymbol{\mathcal{S}}\mathbf{u}_h^n, \mathbf{u}_h^{n+1}+\mathbf{u}_h^n\big)_{\mu\times\varepsilon} &= \frac{1}{2}\big(\boldsymbol{\mathcal{S}}(\mathbf{u}_h^{n+1}+\mathbf{u}_h^n), \mathbf{u}_h^{n+1}+\mathbf{u}_h^n\big)_{\mu\times\varepsilon} - \frac{1}{2}\big(\boldsymbol{\mathcal{S}}(\mathbf{u}_h^{n+1}-\mathbf{u}_h^n), \mathbf{u}_h^{n+1}+\mathbf{u}_h^n\big)_{\mu\times\varepsilon}\\ &= \frac{1}{2}|\mathbf{u}_h^{n+1}+\mathbf{u}_h^n|_{\boldsymbol{\mathcal{S}}}^2 \qquad - \frac{1}{2}\big(|\mathbf{u}_h^{n+1}|_{\boldsymbol{\mathcal{S}}}^2 - |\mathbf{u}_h^n|_{\boldsymbol{\mathcal{S}}}^2\big).\end{aligned}$$

Inserting this identity into (4.82) and further using (4.45a), (4.45b) and summing yields the statement. □

This lemma implies that the combination of the upwind fluxes dG space discretization and the Crank–Nicolson method is a dissipative scheme. Clearly, this implies unconditional stability. For the Verlet method we again need a CFL condition to ensure stability.

Corollary 4.24. *For vanishing source term $\mathbf{J}_h \equiv 0$, the Crank–Nicolson method is dissipative. More precisely, we have*

$$\mathcal{E}(\mathbf{H}_h^n, \mathbf{E}_h^n) = \mathcal{E}(\mathbf{H}_h^0, \mathbf{E}_h^0) - \alpha\frac{\tau}{4}\sum_{m=0}^{n-1} |\mathbf{u}_h^{m+1} + \mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2, \qquad n = 1, 2, \dots . \tag{4.83}$$

For the Verlet method we have

$$\widehat{\mathcal{E}}_{\mathrm{upw}}(\mathbf{H}_h^n, \mathbf{E}_h^n) = \widehat{\mathcal{E}}_{\mathrm{upw}}(\mathbf{H}_h^0, \mathbf{E}_h^0) - \alpha\frac{\tau}{4}\sum_{m=0}^{n-1} |\mathbf{u}_h^{m+1} + \mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2, \qquad n = 1, 2, \dots , \tag{4.84}$$

where the perturbed electromagnetic energy $\widehat{\mathcal{E}}_{\mathrm{upw}}$ is defined as

$$\widehat{\mathcal{E}}_{\mathrm{upw}}(\mathbf{H}_h, \mathbf{E}_h) = \widehat{\mathcal{E}}(\mathbf{H}_h, \mathbf{E}_h) - \alpha\frac{\tau}{4}|\mathbf{u}_h|_{\boldsymbol{\mathcal{S}}}^2.$$

The next corollary gives the stability result for the Crank–Nicolson method.

Corollary 4.25. *The approximation $\mathbf{u}_h^n$ obtained from the upwind fluxes dG discretization and the Crank–Nicolson method (4.77) is bounded by*

$$\|\mathbf{u}_h^n\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n-1} |\mathbf{u}_h^{m+1} + \mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2 \le \mathrm{e}^{3/2}\|\mathbf{u}^0\|_{\mu\times\varepsilon}^2 + \mathrm{e}^{3/2}\frac{T+1}{\delta}\frac{\tau}{4}\sum_{m=0}^{n-1} \|\mathbf{J}^{m+1} + \mathbf{J}^m\|^2, \tag{4.85}$$

for $n = 1, 2, \dots, N_T$.

Proof. We apply the Cauchy–Schwarz inequality and the weighted Young's inequality with weight $\gamma > 0$ to (4.80). This yields

$$\begin{aligned}\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n} |\mathbf{u}_h^{m+1} + \mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2 \\ \le \|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{\tau}{2}\sum_{m=0}^{n}\left(\frac{1}{2\gamma}\|\mathbf{j}_h^{m+1} + \mathbf{j}_h^m\|_{\mu\times\varepsilon}^2 + \frac{\gamma}{2}\|\mathbf{u}_h^{m+1} + \mathbf{u}_h^m\|_{\mu\times\varepsilon}^2\right).\end{aligned}$$

Applying the triangle inequality and Young's inequality to the last term, we obtain

$$\begin{aligned}\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n} |\mathbf{u}_h^{m+1} + \mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2 \le \|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{\tau}{4\gamma}\sum_{m=0}^{n} \|\mathbf{j}_h^{m+1} + \mathbf{j}_h^m\|_{\mu\times\varepsilon}^2 \\ + \gamma\frac{\tau}{2}\sum_{m=0}^{n}\left(\|\mathbf{u}_h^{m+1}\|_{\mu\times\varepsilon}^2 + \|\mathbf{u}_h^m\|_{\mu\times\varepsilon}^2\right).\end{aligned}$$

Now, we choose the weight $\gamma = 1/(T+1)$. This enables us to apply a variant of the discrete Gronwall inequality given in Lemma A.2,

$$\|\mathbf{u}_h^n\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n-1} |\mathbf{u}_h^{m+1} + \mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2 \le \exp\left(\frac{3}{2}\frac{t_n}{T+1}\right)\left(\|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + (T+1)\frac{\tau}{4}\sum_{m=0}^{n-1} \|\mathbf{j}_h^{m+1} + \mathbf{j}_h^m\|_{\mu\times\varepsilon}^2\right).$$

Clearly, we have $t_n/(T+1) \le 1$ and the proof is complete. □

As mentioned above we need a CFL condition for the Verlet method. Since we integrate the stabilization operators explicitly in time, they contribute to the CFL condition and we thus need a stronger condition compared to the one for the central fluxes (4.49). In particular, we need the following **CFL condition for the upwind fluxes Verlet scheme**,

$$\tau \leq \frac{2\widehat{\theta}}{C_{\mathrm{bnd}} c_\infty} \min_{K\in\mathcal{T}_h} h_K, \tag{4.86a}$$

with a fixed parameter $0 < \widehat{\theta} < 1$ which satisfies the condition

$$\widehat{\theta}_{\mathrm{upw}} := \widehat{\theta}^2 + \alpha\widehat{\theta} < 1. \tag{4.86b}$$

Note that the CFL condition depends on the stabilization parameter α. For larger α we get a stricter condition.

Corollary 4.26. *Assume that the CFL condition (4.86) is satisfied. Then, the approximation $\mathbf{u}_h^n$ obtained from the upwind fluxes dG discretization and the Verlet method (4.79) is bounded by*

$$\begin{aligned}(1-\widehat{\theta}_{\mathrm{upw}})\|\mathbf{u}_h^n\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n-1}|\mathbf{u}_h^{m+1}+\mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2 \\ \leq \mathrm{e}^{3/2}\|\mathbf{u}^0\|_{\mu\times\varepsilon}^2 + \mathrm{e}^{3/2}\frac{T+1}{\delta(1-\widehat{\theta}_{\mathrm{upw}})}\frac{\tau}{4}\sum_{m=0}^{n-1}\|\mathbf{J}^{m+1}+\mathbf{J}^m\|^2,\end{aligned} \tag{4.87}$$

for $n = 1, 2, \ldots, N_T$.

Observe that the bound deteriorates for $\widehat{\theta}_{\mathrm{upw}} \nearrow 1$.

Proof. By (4.81) we have

$$\begin{aligned}\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n}|\mathbf{u}_h^{m+1}+\mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2 \leq \|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1}\|_\mu^2 + \alpha\frac{\tau}{2}|\mathbf{u}_h^{n+1}|_{\boldsymbol{\mathcal{S}}}^2 \\ + \frac{\tau}{4\gamma}\sum_{m=0}^{n}\|\mathbf{j}_h^{m+1}+\mathbf{j}_h^m\|_{\mu\times\varepsilon}^2 \\ + \gamma\frac{\tau}{2}\sum_{m=0}^{n}\left(\|\mathbf{u}_h^{m+1}\|_{\mu\times\varepsilon}^2 + \|\mathbf{u}_h^m\|_{\mu\times\varepsilon}^2\right),\end{aligned}$$

see the proof of Corollary 4.25. Applying the boundedness results for $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$ and $|\cdot|_{\boldsymbol{\mathcal{S}}}$ obtained in Theorem 3.14 and the CFL condition (4.86), we infer

$$\begin{aligned}\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n}|\mathbf{u}_h^{m+1}+\mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2 \\ \leq \|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \widehat{\theta}^2\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\widehat{\theta}\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 \\ + \frac{\tau}{4\gamma}\sum_{m=0}^{n}\|\mathbf{j}_h^{m+1}+\mathbf{j}_h^m\|_{\mu\times\varepsilon}^2 + \gamma\frac{\tau}{2}\sum_{m=0}^{n}\left(\|\mathbf{u}_h^{m+1}\|_{\mu\times\varepsilon}^2 + \|\mathbf{u}_h^m\|_{\mu\times\varepsilon}^2\right).\end{aligned}$$

We choose the weight $\gamma = (1-\widehat{\theta}_{\mathrm{upw}})/(T+1)$, which enables us to apply the discrete Gronwall lemma (Lemma A.2 with $\lambda = 1/(T+1)$). This yields

$$\begin{aligned}(1-\widehat{\theta}_{\mathrm{upw}})\|\mathbf{u}_h^n\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n-1}|\mathbf{u}_h^{m+1}+\mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}}^2 \\ \leq \exp\left(\frac{3}{2}\frac{t_n}{T+1}\right)\left(\|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{T+1}{1-\widehat{\theta}_{\mathrm{upw}}}\frac{\tau}{4}\sum_{m=0}^{n-1}\|\mathbf{j}_h^{m+1}+\mathbf{j}_h^m\|_{\mu\times\varepsilon}^2\right),\end{aligned}$$

which completes the proof. □

4.3.2 Full discretization errors

In the previous section we established the stability of the upwind fluxes dG space discretization in combination with the Crank–Nicolson time integration or with the Verlet scheme. Now, we turn to the error analysis. We restrict ourselves to the Crank–Nicolson method because its convergence result is relatively straightforward, whereas the result for the Verlet method is more involved. The result for the Verlet scheme will be given in the next chapter as a special case of a result for the locally implicit scheme.

Similar to semidiscrete case the use of an upwind fluxes discretization improves the spatial convergence to order $k+1/2$ compared with order k in the central fluxes case. Another similarity is that our analysis is again based on an energy method, compare Theorem 3.13. The convergence result for the Crank–Nicolson method is relatively straightforward, whereas the result for the Verlet method is more involved.

As in (4.56b) we split the error into $\mathbf{e}^n = \mathbf{e}_\pi^n - \mathbf{e}_h^n$.

Lemma 4.27. *Let* $\mathbf{u} \in C\big(0,T;V_\star\big) \cap C^3\big(0,T;L^2(\Omega)^6\big)$ *be the exact solution of* (4.1). *The error* $\mathbf{e}_h^n$ *satisfies*

$$\boldsymbol{\mathcal{R}}_L \mathbf{e}_h^{n+1} - \boldsymbol{\mathcal{R}}_R \mathbf{e}_h^n = -\frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}(\mathbf{e}_h^{n+1} + \mathbf{e}_h^n) + \mathbf{d}_{\mathrm{upw}}^n, \tag{4.88}$$

if we employ the Crank–Nicolson method. The defect $\mathbf{d}_{\mathrm{upw}}^n = \mathbf{d}_{\pi,\mathrm{upw}}^n + \mathbf{d}_h^n$ *is given by*

$$\mathbf{d}_{\pi,\mathrm{upw}}^n = \mathbf{d}_\pi^n + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}(\mathbf{e}_\pi^{n+1} + \mathbf{e}_\pi^n) = -\frac{\tau}{2}(\boldsymbol{\mathcal{C}} - \alpha\boldsymbol{\mathcal{S}})(\mathbf{e}_\pi^{n+1} + \mathbf{e}_\pi^n), \tag{4.89}$$

where $\mathbf{d}_\pi^n$ *and* $\mathbf{d}_h^n$ *were defined in* (4.58).

Proof. The defects are obtained by inserting the projected exact solution into the numerical scheme (4.76),

$$\pi_h(\mathbf{u}^{n+1} - \mathbf{u}^n) = \frac{\tau}{2}(\boldsymbol{\mathcal{C}} - \alpha\boldsymbol{\mathcal{S}})\big(\pi_h(\mathbf{u}^{n+1} + \mathbf{u}^n)\big) + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) - \mathbf{d}_{\mathrm{upw}}^n. \tag{4.90}$$

By (4.61) and the consistency of the stabilization operator (3.16) the exact solution $\mathbf{u}$ satisfies

$$\pi_h(\mathbf{u}^{n+1} - \mathbf{u}^n) = \frac{\tau}{2}(\boldsymbol{\mathcal{C}} - \alpha\boldsymbol{\mathcal{S}})(\mathbf{u}^{n+1} + \mathbf{u}^n) + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) + \tau^2\pi_h\delta^n(\partial_t\mathbf{u}).$$

Subtracting this from (4.90) gives

$$\mathbf{d}_{\mathrm{upw}}^n = -\frac{\tau}{2}(\boldsymbol{\mathcal{C}} - \alpha\boldsymbol{\mathcal{S}})(\mathbf{e}_\pi^{n+1} + \mathbf{e}_\pi^n) - \tau^2\pi_h\delta^n(\partial_t\mathbf{u}) = \mathbf{d}_\pi^n + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}(\mathbf{e}_\pi^{n+1} + \mathbf{e}_\pi^n) + \mathbf{d}_h^n,$$

by definition of $\mathbf{d}_\pi^n$ and $\mathbf{d}_h^n$ in (4.58). This proves the statement. □

The **convergence result** for the **upwind fluxes dG discretization in combination with the Crank–Nicolson scheme** reads as follows.

Theorem 4.28. *Let* $\mathbf{u} \in C\big(0,T;D(\mathcal{C}) \cap H^{k+1}(\mathcal{T}_h)^6\big) \cap C^3\big(0,T;L^2(\Omega)^6\big)$ *be the exact solution of* (4.1). *Then, the error of the upwind fluxes dG discretization and the Crank–Nicolson scheme*

(4.76) *satisfies*

$$\begin{aligned}\|\mathbf{u}^n-\mathbf{u}_h^n\|_{\mu\times\varepsilon}^2+\alpha\frac{\tau}{4}\sum_{m=0}^{n-1}|\mathbf{e}_h^{m+1}+\mathbf{e}_h^m|_{\boldsymbol{\mathcal{S}}}^2 \le& C_{\mathrm{app}}^2|\mathbf{u}^n|_{k+1,\mathcal{T}_h,1,k+1}^2\\ &+\tau^4\frac{\mathrm{e}^{3/2}}{64}(T+1)\int_0^{t_n}\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}^2\,dt\\ &+\frac{\mathrm{e}^{3/2}}{4}C_{\mathrm{upw}}\tau\sum_{m=0}^{n-1}|\mathbf{u}^{m+1}+\mathbf{u}^m|_{k+1,\mathcal{T}_h,2,k+\frac12}^2\\ \le& C\Big(h^{2k+1}+\tau^4\Big).\end{aligned}\tag{4.91}$$

The constant C only depends on C_{app}, C_π, $(1+\alpha)^2/\alpha$, T, $|\mathbf{u}(t)|_{k+1,\mathcal{T}_h}$, and $\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}$, $t\in[0,t_n]$.

Proof. By (4.88), Lemma 4.23 with $\frac{\tau}{2}(\mathbf{j}_h^{m+1}+\mathbf{j}_h^m)$ replaced by $\mathbf{d}_{\mathrm{upw}}^m$, and $\mathbf{e}_h^0=0$, the error $\mathbf{e}_h^n$ satisfies

$$\|\mathbf{e}_h^{n+1}\|_{\mu\times\varepsilon}^2+\alpha\frac{\tau}{2}\sum_{m=0}^{n}|\mathbf{e}_h^{m+1}+\mathbf{e}_h^m|_{\boldsymbol{\mathcal{S}}}^2=\sum_{m=0}^{n}\big(\mathbf{d}_{\mathrm{upw}}^m,\mathbf{e}_h^{m+1}+\mathbf{e}_h^m\big)_{\mu\times\varepsilon}.\tag{4.92}$$

From Lemma 4.27 we have

$$\mathbf{d}_{\mathrm{upw}}^m=\mathbf{d}_h^m-\frac{\tau}{2}(\boldsymbol{\mathcal{C}}-\alpha\boldsymbol{\mathcal{S}})(\mathbf{e}_\pi^{m+1}+\mathbf{e}_\pi^m).$$

For the first term on the right-hand side of (4.92) we have

$$\big(\mathbf{d}_h^m,\mathbf{e}_h^{m+1}+\mathbf{e}_h^m\big)_{\mu\times\varepsilon}\le\frac{T+1}{\tau}\|\mathbf{d}_h^m\|_{\mu\times\varepsilon}^2+\frac{1}{T+1}\frac{\tau}{4}\|\mathbf{e}_h^{m+1}+\mathbf{e}_h^m\|_{\mu\times\varepsilon}^2.$$

The bound (4.68a) for $\mathbf{d}_h^m$ and the Cauchy–Schwarz inequality imply

$$\|\mathbf{d}_h^m\|_{\mu\times\varepsilon}^2\le\frac{\tau^4}{64}\Big(\int_{t_m}^{t_{m+1}}\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}\,dt\Big)^2\le\frac{\tau^5}{64}\int_{t_m}^{t_{m+1}}\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}^2\,dt.$$

For the second term we obtain from Theorem 3.10 and Remark 3.11

$$\begin{aligned}-\big((\boldsymbol{\mathcal{C}}-\alpha\boldsymbol{\mathcal{S}})(\mathbf{e}_\pi^{m+1}+\mathbf{e}_\pi^m),\,\mathbf{e}_h^{m+1}+\mathbf{e}_h^m\big)_{\mu\times\varepsilon}&\\ \le C_\pi(1+\alpha)|\mathbf{e}_h^{m+1}+\mathbf{e}_h^m|_{\boldsymbol{\mathcal{S}}}|\mathbf{u}^{m+1}+\mathbf{u}^m|_{k+1,\mathcal{T}_h,2,k+\frac12}&\\ \le\frac{\gamma}{2}(1+\alpha)^2|\mathbf{e}_h^{m+1}+\mathbf{e}_h^m|_{\boldsymbol{\mathcal{S}}}^2+\frac{C_\pi^2}{2\gamma}|\mathbf{u}^{m+1}+\mathbf{u}^m|_{k+1,\mathcal{T}_h,2,k+\frac12}^2.&\end{aligned}$$

Choosing $\gamma=\alpha/(1+\alpha)^2$ we conclude

$$\begin{aligned}\|\mathbf{e}_h^{n+1}\|_{\mu\times\varepsilon}^2+\alpha\frac{\tau}{4}\sum_{m=0}^{n}|\mathbf{e}_h^{m+1}+\mathbf{e}_h^m|_{\boldsymbol{\mathcal{S}}}^2\le&\frac{1}{T+1}\frac{\tau}{2}\sum_{m=0}^{n}\Big(\|\mathbf{e}_h^{m+1}\|_{\mu\times\varepsilon}^2+\|\mathbf{e}_h^m\|_{\mu\times\varepsilon}^2\Big)\\ &+\tau\sum_{m=0}^{n}\Big(\frac{T+1}{\tau^2}\|\mathbf{d}_h^m\|_{\mu\times\varepsilon}^2+\frac{C_{\mathrm{upw}}}{4}|\mathbf{u}^{m+1}+\mathbf{u}^m|_{k+1,\mathcal{T}_h,2,k+\frac12}^2\Big),\end{aligned}$$

where we used $C_\pi^2/\gamma=C_{\mathrm{upw}}$, see Theorem 3.13. Since $\tau/(T+1)\le3/2$, the discrete Gronwall lemma (Lemma A.2) shows

$$\begin{aligned}\|\mathbf{e}_h^n\|_{\mu\times\varepsilon}^2+\alpha\frac{\tau}{4}\sum_{m=0}^{n-1}|\mathbf{e}_h^{m+1}+\mathbf{e}_h^m|_{\boldsymbol{\mathcal{S}}}^2\le&\tau^4\frac{\mathrm{e}^{3/2}}{64}(T+1)\int_0^{t_n}\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}^2\,dt\\ &+\frac{\mathrm{e}^{3/2}}{4}C_{\mathrm{upw}}\tau\sum_{m=0}^{n-1}|\mathbf{u}^{m+1}+\mathbf{u}^m|_{k+1,\mathcal{T}_h,2,k+\frac12}^2.\end{aligned}$$

Since $\big(\mathbf{e}_\pi, \mathbf{e}_h\big)_{\mu\times\varepsilon} = 0$, the result now follows from

$$\|\mathbf{e}^n\|_{\mu\times\varepsilon}^2 = \|\mathbf{e}_\pi^n\|_{\mu\times\varepsilon}^2 + \|\mathbf{e}_h^n\|_{\mu\times\varepsilon}^2,$$

and (3.24a). □

4.4 Time integration for Maxwell's equations: Implicit midpoint method

In this section we briefly present how the above developed techniques can be extended for the analysis of the implicit midpoint method in combination with a central fluxes dG space discretization. In this case the implicit midpoint method (4.28a) reads

$$\mathbf{u}_h^{n+1} - \mathbf{u}_h^n = \frac{\tau}{2}\boldsymbol{\mathcal{C}}\big(\mathbf{u}_h^{n+1} + \mathbf{u}_h^n\big) + \frac{\tau}{2}\mathbf{j}_h^{n+1/2}, \tag{4.93a}$$

or, equivalently,

$$\boldsymbol{\mathcal{R}}_L\mathbf{u}_h^{n+1} = \boldsymbol{\mathcal{R}}_R\mathbf{u}_h^n + \frac{\tau}{2}\mathbf{j}_h^{n+1/2}. \tag{4.93b}$$

The operators $\boldsymbol{\mathcal{R}}_L$ and $\boldsymbol{\mathcal{R}}_R$ are the same as for the Crank–Nicolson method, see (4.41b). We emphasize again that the implicit midpoint method and the Crank–Nicolson scheme differ only (for linear problems) in the treatment of the source term $\mathbf{j}_h$. As a consequence, the implicit midpoint method inherits the stability and energy conservation properties of the Crank–Nicolson method shown in Section 4.2.1.

Corollary 4.29. *For the approximation $\mathbf{u}_h^n = (\mathbf{H}_h^n, \mathbf{E}_h^n)$ of the central fluxes dG discretization and the implicit midpoint method (4.93) we have the following stability bound*

$$\|\mathbf{u}_h^n\|_{\mu\times\varepsilon} \le \|\mathbf{u}^0\|_{\mu\times\varepsilon} + \frac{\tau}{\sqrt{\delta}}\sum_{m=0}^{n-1}\|\mathbf{J}^{m+1/2}\|, \qquad n = 1, 2, \dots\ . \tag{4.94}$$

Moreover, for vanishing source term $\mathbf{J}_h \equiv 0$, the electromagnetic energy is conserved over time, i.e.,

$$\mathcal{E}(\mathbf{H}_h^n, \mathbf{E}_h^n) = \mathcal{E}(\mathbf{H}_h^0, \mathbf{E}_h^0), \qquad n = 1, 2, \dots\ .$$

Comprising the ideas of Sections 4.1.5 and 4.2.2 we obtain the following error recursion.

Lemma 4.30. *Let $\mathbf{u} \in C\big(0, T; V_\star\big) \cap C^3\big(0, T; L^2(\Omega)^6\big)$ be the exact solution of (4.1). The error $\mathbf{e}_h^n$ of the central fluxes dG discretization and the implicit midpoint rule (4.93) satisfies*

$$\boldsymbol{\mathcal{R}}_L\mathbf{e}_h^{n+1} = \boldsymbol{\mathcal{R}}_R\mathbf{e}_h^n + \bar{\mathbf{d}}^n, \qquad \bar{\mathbf{d}}^n = \mathbf{d}_\pi^n + \bar{\mathbf{d}}_h^n. \tag{4.95a}$$

The projection defect $\mathbf{d}_\pi^n$ was defined in (4.58) and the quadrature defect $\bar{\mathbf{d}}_h^n$ is given by

$$\bar{\mathbf{d}}_h^n = -\tau^2\pi_h\bar{\delta}^n(\partial_t\mathbf{u}) - \tau^2\boldsymbol{\mathcal{C}}\big(\delta^n(\mathbf{u}) - \bar{\delta}^n(\mathbf{u})\big). \tag{4.95b}$$

Here, δ^n and $\bar{\delta}^n$ are the quadrature errors of the trapezoidal rule and the midpoint rule, respectively, given in (4.24a) and (4.29b). The defect $\bar{\mathbf{d}}^n$ can be expressed as

$$\bar{\mathbf{d}}^n = \bar{\mathbf{d}}_\pi^n - \tau^2\pi_h\bar{\delta}^n(\partial_t\mathbf{u}) + (\boldsymbol{\mathcal{R}}_L - \boldsymbol{\mathcal{R}}_R)\pi_h\bar{\boldsymbol{\zeta}}^n, \tag{4.95c}$$

with

$$\bar{\mathbf{d}}_\pi^n = -\tau\boldsymbol{\mathcal{C}}\mathbf{e}_\pi^{n+1/2}, \qquad \bar{\boldsymbol{\zeta}}^n = \tau\big(\delta^n(\mathbf{u}) - \bar{\delta}^n(\mathbf{u})\big). \tag{4.95d}$$

Remark 4.31. Note that a straightforward bound on $\tau^2\boldsymbol{\mathcal{C}}(\delta^n(\mathbf{u}) - \bar{\delta}^n(\mathbf{u}))$ of order τ^3 requires the **regularity assumption** $\partial_t^2\mathbf{u} \in D(\mathcal{C})$. Under this assumption we could achieve the bound

$$\|\tau^2\boldsymbol{\mathcal{C}}(\delta^n(\mathbf{u}) - \bar{\delta}^n(\mathbf{u}))\|_{\mu\times\varepsilon} \le \tau^2\|\pi_h\mathcal{C}(\delta^n(\mathbf{u}) - \bar{\delta}^n(\mathbf{u}))\|_{\mu\times\varepsilon} \le \frac{\tau^2}{4}\int_{t_n}^{t_{n+1}} \|\partial_t^2\mathbf{u}(t)\|_{H(\mathrm{curl},\Omega)}\, dt.$$

Here, the first inequality follows by the consistency of $\mathcal{C}$, see (3.6), and the second inequality with (4.32). However, in Section 4.1.5 we derived a technique **to omit this assumption** and which enables us to prove the convergence result in Theorem 4.32 below without it.

Proof. (a) We insert the projection of the exact solution $\mathbf{u}$ into the implicit midpoint scheme (4.93),

$$\pi_h\big(\mathbf{u}^{n+1} - \mathbf{u}^n\big) = \frac{\tau}{2}\boldsymbol{\mathcal{C}}\pi_h\big(\mathbf{u}^{n+1} + \mathbf{u}^n\big) + \tau\mathbf{j}_h^{n+1/2} - \bar{\mathbf{d}}^n. \tag{4.96}$$

Subtracting this equation from (4.93) yields

$$\mathbf{e}_h^{n+1} - \mathbf{e}_h^n = \frac{\tau}{2}\boldsymbol{\mathcal{C}}\big(\mathbf{e}_h^{n+1} + \mathbf{e}_h^n\big) + \bar{\mathbf{d}}^n,$$

which proves (4.95a). In order to compute the defect $\bar{\mathbf{d}}^n$ we substitute $\mathbf{j}_h^{n+1/2}$ in (4.96) via (3.27) with $\alpha = 0$,

$$\pi_h\big(\mathbf{u}^{n+1} - \mathbf{u}^n\big) = \frac{\tau}{2}\boldsymbol{\mathcal{C}}\pi_h\big(\mathbf{u}^{n+1} + \mathbf{u}^n\big) + \frac{\tau}{2}\pi_h\partial_t\mathbf{u}^{n+1/2} - \tau\boldsymbol{\mathcal{C}}\mathbf{u}^{n+1/2} - \bar{\mathbf{d}}^n.$$

Thus, the defect is given by

$$\begin{aligned}\bar{\mathbf{d}}^n = &-\frac{\tau}{2}\boldsymbol{\mathcal{C}}\big(\mathbf{e}_\pi^{n+1} + \mathbf{e}_\pi^n\big) + \pi_h\partial_t\mathbf{u}^{n+1/2} - \int_{t_n}^{t_{n+1}} \pi_h(\partial_t\mathbf{u}(t))\, dt\\ &- \tau\boldsymbol{\mathcal{C}}\mathbf{u}^{n+1/2} + \int_{t_n}^{t_{n+1}} \boldsymbol{\mathcal{C}}\mathbf{u}(t)\, dt + \frac{\tau}{2}\boldsymbol{\mathcal{C}}\big(\mathbf{u}^{n+1} + \mathbf{u}^n\big) - \int_{t_n}^{t_{n+1}} \boldsymbol{\mathcal{C}}\mathbf{u}(t)\, dt,\end{aligned}$$

which shows (4.95b).

(b) For the splitting (4.95c) observe that by that $-\tau\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{R}}_L - \boldsymbol{\mathcal{R}}_R$, cf. Lemma 4.6, we have

$$\bar{\mathbf{d}}^n = \mathbf{d}_\pi^n - \tau^2\pi_h\bar{\delta}^n(\partial_t\mathbf{u}) - \tau\boldsymbol{\mathcal{C}}\bar{\boldsymbol{\zeta}}^n = \mathbf{d}_\pi^n - \tau\boldsymbol{\mathcal{C}}(\boldsymbol{\mathcal{I}} - \pi_h)\bar{\boldsymbol{\zeta}}^n - \tau^2\pi_h\bar{\delta}^n(\partial_t\mathbf{u}) + (\boldsymbol{\mathcal{R}}_L - \boldsymbol{\mathcal{R}}_R)\pi_h\bar{\boldsymbol{\zeta}}^n.$$

Then,

$$\mathbf{d}_\pi^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}\big(\mathbf{e}_\pi^{n+1} + \mathbf{e}_\pi^n\big), \qquad \tau\boldsymbol{\mathcal{C}}(\boldsymbol{\mathcal{I}} - \pi_h)\bar{\boldsymbol{\zeta}}^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}\big(\mathbf{e}_\pi^{n+1} - 2\mathbf{e}_\pi^{n+1/2} + \mathbf{e}_\pi^n\big),$$

see (4.58) and (4.36), yield (4.95c), (4.95d). □

Theorem 4.32. *Assume that the exact solution of (4.1) satisfies* $\mathbf{u} \in C\big(0,T;D(\mathcal{C})\cap H^{k+1}(\mathcal{T}_h)^6\big)\cap C^3\big(0,T;L^2(\Omega)^6\big)$. *Then, the full discretization error of the central fluxes dG discretization and the implicit midpoint rule (4.93) is bounded by*

$$\begin{aligned}\|\mathbf{u}^n - \mathbf{u}_h^n\|_{\mu\times\varepsilon} &\le C_{\mathrm{app}}|\mathbf{u}^n|_{k+1,\mathcal{T}_h,1,k+1} + \widehat{C}_\pi\tau\sum_{m=0}^{n-1}|\mathbf{u}^{m+1/2}|_{k+1,\mathcal{T}_h,2,k}\\ &\quad + \frac{\tau^2}{2}\left(\max_{[t_0,t_1]\cup[t_{n-1},t_n]}\|\partial_t^2\mathbf{u}(t)\|_{\mu\times\varepsilon} + \frac{3}{4}\int_0^{t_n}\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}\, dt\right).\\ &\le C\Big(h^k + \tau^2\Big).\end{aligned}$$

Proof. Employing the splitting (4.95c) in the error recursion (4.95a) we infer

$$\mathbf{e}_h^{n+1} = \pi_h\bar{\boldsymbol{\zeta}}^n - \boldsymbol{\mathcal{R}}^{n+1}\pi_h\bar{\boldsymbol{\zeta}}^0 + \sum_{m=0}^{n}\boldsymbol{\mathcal{R}}^{n-m}\boldsymbol{\mathcal{R}}_L^{-1}\big(\bar{\mathbf{d}}_\pi^m - \tau^2\pi_h\bar{\delta}^m(\partial_t\mathbf{u})\big) - \sum_{m=0}^{n}\boldsymbol{\mathcal{R}}^{n-m}\pi_h\big(\bar{\boldsymbol{\zeta}}^{m+1} - \bar{\boldsymbol{\zeta}}^m\big),$$

The assertion now follows with (3.43), (4.29d), (4.33c), (4.35), and the bounds (4.44c) and (4.47) on $\boldsymbol{\mathcal{R}}_L^{-1}$ and $\boldsymbol{\mathcal{R}}$, respectively. (The bounds are applicable since all defects are elements of V_h^2). □

Remark 4.33. Note the similarity of the upper convergence proof for the implicit midpoint method to the convergence proof for the Verlet method, i.e. to the proof of Theorem 4.22. In both proofs we use that problematic part of the defect can be represented by using $\boldsymbol{\mathcal{R}}_L - \boldsymbol{\mathcal{R}}_R$ (implicit midpoint) or $\widehat{\boldsymbol{\mathcal{R}}}_L - \widehat{\boldsymbol{\mathcal{R}}}_R$ (Verlet). This enables us to achieve a convergence result with less regularity assumptions on the exact solution than a naive approach.

4.5 Implementation and numerical results

We end this chapter with the discussion of some implementation issues of the Verlet method and of the Crank–Nicolson method and subsequently give numerical results confirming our theoretical considerations.

4.5.1 Implementation

We begin by discussing the implementation and the costs of the Verlet and of the Crank–Nicolson method. The central fluxes dG discretization in combination with the Verlet method (4.39) only needs one evaluation of $\boldsymbol{\mathcal{C}}_\mathbf{H}$ in (4.39b) and one evalution of $\boldsymbol{\mathcal{C}}_\mathbf{E}$ in (4.39c). The computation in (4.39a) only has to be carried out for $n = 1$ and then can be replaced by

$$\mathbf{H}_h^{n+1/2} = 2\mathbf{H}_h^n - \mathbf{H}_h^{n-1/2}, \qquad n = 2, 3, \dots\ . \tag{4.97}$$

Alternatively, we can store $\boldsymbol{\mathcal{C}}_\mathbf{E}\mathbf{E}_h^{n+1}$ in (4.39c) and use it to compute (4.39a) in the next step, i.e.,

$$\mathbf{H}_h^{n+3/2} = \mathbf{H}_h^{n+1} - \frac{\tau}{2}\boldsymbol{\mathcal{C}}_\mathbf{E}\mathbf{E}_h^{n+1}, \qquad n = 2, 3, \dots\ .$$

For the upwind fluxes dG discretization together with the Verlet method, we cannot use (4.97), but by storing $\boldsymbol{\mathcal{C}}_\mathbf{E}\mathbf{E}_h^{n+1}$ in (4.78c) we can save one matrix-vector multiplication in (4.78a), since

$$\mathbf{H}_h^{n+3/2} = \mathbf{H}_h^{n+1} - \frac{\tau}{2}\boldsymbol{\mathcal{C}}_\mathbf{E}\mathbf{E}_h^{n+1} - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_\mathbf{H}\mathbf{H}_h^{n+1}, \qquad n = 2, 3, \dots\ .$$

Thus, we need one evaluation of each $\boldsymbol{\mathcal{C}}_\mathbf{E}$, $\boldsymbol{\mathcal{C}}_\mathbf{H}$, $\boldsymbol{\mathcal{S}}_\mathbf{H}$ and $\boldsymbol{\mathcal{S}}_\mathbf{E}$.

For the Crank–Nicolson method the main effort lies in the solution of a linear system. When using a central fluxes dG discretization this linear system reads

$$\boldsymbol{\mathcal{R}}_L\mathbf{u}_h^{n+1} = \mathbf{b}_h^n, \qquad \mathbf{b}_h^n = \boldsymbol{\mathcal{R}}_R\mathbf{u}_h^n + \frac{\tau}{2}\big(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n\big), \tag{4.98a}$$

see (4.40). This is a linear system on all degrees of freedom (dof) of the combined field $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h)$. By using a Schur decomposition one can show that (4.98a) is equivalent to

$$\begin{pmatrix} \boldsymbol{\mathcal{I}} & \frac{\tau}{2}\boldsymbol{\mathcal{C}}_\mathbf{E} \\ 0 & \boldsymbol{\mathcal{L}} \end{pmatrix}\begin{pmatrix} \mathbf{H}_h^{n+1} \\ \mathbf{E}_h^{n+1} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_\mathbf{H}^n \\ \mathbf{b}_\mathbf{E}^n + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_\mathbf{H}\mathbf{b}_\mathbf{H}^n \end{pmatrix}, \tag{4.98b}$$

where the right-hand side consists of

$$\mathbf{b}_{\mathbf{H}}^n = \mathbf{H}_h^n - \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^n, \qquad \mathbf{b}_{\mathbf{E}}^n = \mathbf{E}_h^n + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h^n - \frac{\tau}{2}\left(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n\right), \tag{4.98c}$$

and where $\boldsymbol{\mathcal{L}}$ is the **Schur complement** of $\boldsymbol{\mathcal{I}} - \frac{\tau}{2}\boldsymbol{\mathcal{C}}$ given by

$$\boldsymbol{\mathcal{L}} = \boldsymbol{\mathcal{I}} + \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}. \tag{4.98d}$$

We see that (4.98b) only requires the solution of a linear system on the dof of the electric field $\mathbf{E}_h$,

$$\boldsymbol{\mathcal{L}}\mathbf{E}_h^{n+1} = \mathbf{b}_{\mathbf{E}}^n + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{b}_{\mathbf{H}}^n,$$

and the magnetic field then can be explicitly updated via

$$\mathbf{H}_h^{n+1} = \mathbf{b}_{\mathbf{H}}^n - \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1}.$$

In the case of an upwind fluxes dG discretization the Crank–Nicolson method reads

$$(\boldsymbol{\mathcal{R}}_L + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}})\mathbf{u}_h^{n+1} = \widehat{\mathbf{b}}_h^n, \qquad \widehat{\mathbf{b}}_h^n = \mathbf{b}_h^n - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}\mathbf{u}_h^n, \tag{4.99a}$$

where $\mathbf{b}_h^n = (\mathbf{b}_{\mathbf{H}}^n, \mathbf{b}_{\mathbf{E}}^n)$. A Schur decomposition similar to (4.98b) yields

$$\begin{pmatrix} \boldsymbol{\mathcal{I}} + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}} & \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}} \\ 0 & \boldsymbol{\mathcal{L}}_{\text{upw}} \end{pmatrix} \begin{pmatrix} \mathbf{H}_h^{n+1} \\ \mathbf{E}_h^{n+1} \end{pmatrix} = \begin{pmatrix} \widehat{\mathbf{b}}_{\mathbf{H}}^n \\ \widehat{\mathbf{b}}_{\mathbf{E}}^n + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}(\boldsymbol{\mathcal{I}} + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}})^{-1}\widehat{\mathbf{b}}_{\mathbf{H}}^n \end{pmatrix}, \tag{4.99b}$$

with Schur complement

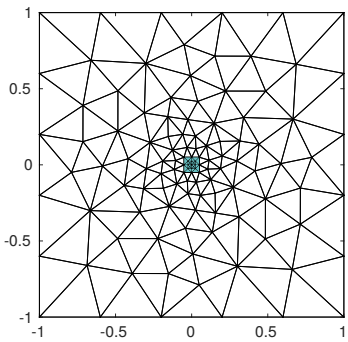
$$\boldsymbol{\mathcal{L}}_{\text{upw}} = \boldsymbol{\mathcal{I}} + \frac{\tau}{2}\boldsymbol{\mathcal{S}}_{\mathbf{E}} + \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}_{\mathbf{H}}(\boldsymbol{\mathcal{I}} + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}})^{-1}\boldsymbol{\mathcal{C}}_{\mathbf{E}}. \tag{4.99c}$$

If we want to use a direct linear solver, working with the system (4.99b) requires the computation of the Schur complement $\boldsymbol{\mathcal{L}}_{\text{upw}}$. Because this needs the inversion (and the storage of the inverse) of the matrix associated with $\boldsymbol{\mathcal{I}} + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}$, it cannot be carried out efficiently. Thus, it is preferable to solve the linear system (4.99a). On the other hand, for an iterative solver the formulation (4.99b) might be beneficial. For this type of solver we only need matrix-vector multiplications with $\boldsymbol{\mathcal{L}}_{\text{upw}}$. This only requires the solution of linear system involving $\boldsymbol{\mathcal{I}} + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}$ (and not the inversion of $\boldsymbol{\mathcal{I}} + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}$) which might be possible with a direct solver.

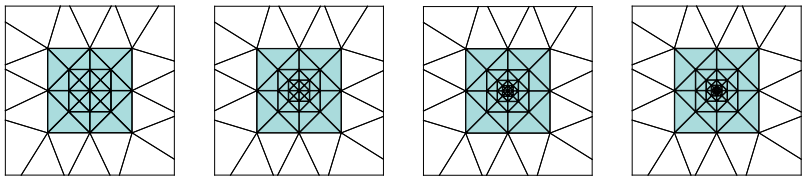
Last, let us comment on the mass matrices which enter in our time integration schemes when working with a representation of our semidiscrete equation w.r.t. a basis of V_h, see Section 3.6 and in particular (3.55). The mass matrices enter in the right-hand side of the Verlet methods (4.39), (4.78) and of the Crank–Nicolson methods (4.40), (4.76). Since in dG methods the mass matrices are block-diagonal, they can be inverted at low costs. Thus, the fully explicit nature of the Verlet methods is preserved and also the Schur decomposition for the central fluxes Crank–Nicolson method can be carried out.

4.5.2 Numerical results

We consider the example from Section 3.7. Our aim in this section is to observe the temporal convergence of the Verlet method and of the Crank–Nicolson method. If we use the mesh sequence $\mathcal{T}_h^{(1)}, \ldots, \mathcal{T}_h^{(4)}$ from Section 3.7, the CFL condition of the Verlet method only allows us to use such tiny time-step sizes that the space discretization error is already dominant and



(a) Initial mesh $\mathcal{T}_h^{(1)}$ corresponding to $\mathcal{T}_{h,\mathrm{CFL}}^{(1)}$. The square $[-0.05, 0.05]^2$ is marked in green.



(b) Refinement of the elements in $[-0.05, 0.05]^2$. This correspond to the meshes $\mathcal{T}_{h,\mathrm{CFL}}^{(1)}, \dots, \mathcal{T}_{h,\mathrm{CFL}}^{(4)}$.

Figure 4.3: Mesh family $\mathcal{T}_{h,\mathrm{CFL}}^{(j)}$.

j	$\max_{K\in\mathcal{T}_{h,\mathrm{CFL}}^{(j)}} h_K$	$\min_{K\in\mathcal{T}_{h,\mathrm{CFL}}^{(j)}} h_K$
1	0.2384	0.0125
2	0.2384	0.00625
3	0.2384	0.003125
4	0.2384	0.0015625

(a) Mesh parameters of $\mathcal{T}_{h,\mathrm{CFL}}^{(j)}$.

j	$\max_{K\in\mathcal{T}_{h,\tau}^{(j)}} h_K$	$\min_{K\in\mathcal{T}_{h,\tau}^{(j)}} h_K$
1	0.2384	0.0125
2	0.1248	0.0125
3	0.0721	0.0125
4	0.0370	0.0125

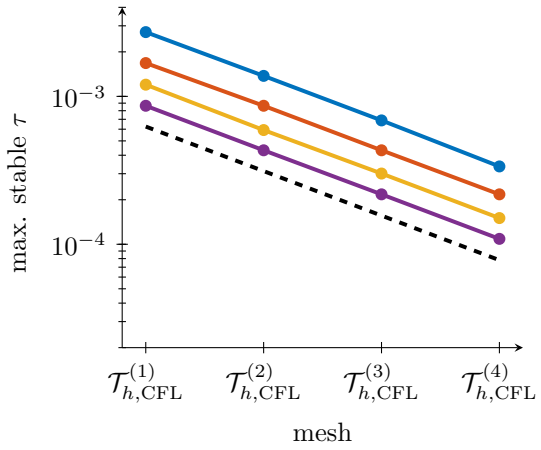
(b) Mesh parameters of $\mathcal{T}_{h,\tau}^{(j)}$.

Table 4.3: Maximum and minimum diameter of the mesh elements in $\mathcal{T}_{h,\mathrm{CFL}}^{(j)}$ and in $\mathcal{T}_{h,\tau}^{(j)}$.

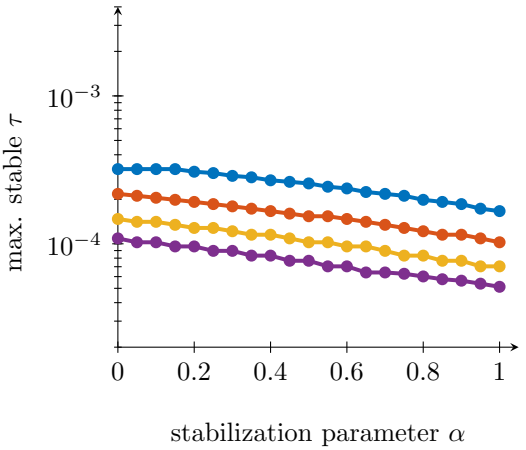
we cannot conclude about the time discretization error. Thus, we use different mesh sequences for the following numerical experiments. In order to examine the CFL condition we use the mesh $\mathcal{T}_h^{(1)}$ as an initial mesh and then only refine the elements in the square $[-0.05, 0.05]^2$, see Figure 4.3. This yields a mesh sequence $\mathcal{T}_{h,\mathrm{CFL}}^{(1)}, \dots, \mathcal{T}_{h,\mathrm{CFL}}^{(4)}$ with parameters given in Table 4.3a. For the confirmation of the temporal convergence we start again with $\mathcal{T}_h^{(1)}$ and then refine all mesh elements in $[-1, 1]^2 \backslash [-0.05, 0.05]^2$. We call the resulting mesh sequence $\mathcal{T}_{h,\tau}^{(1)}, \dots, \mathcal{T}_{h,\tau}^{(4)}$. The mesh parameters can be found in Table 4.3b and a plot of $\mathcal{T}_{h,\tau}^{(1)}$ and $\mathcal{T}_{h,\tau}^{(4)}$ in Figure 3.2.

We start with the validation of our theoretical results with the CFL condition of the Verlet method. We used a central fluxes dG space discretization of Maxwell's equation with different polynomial degrees k and different mesh levels $\mathcal{T}_{h,\mathrm{CFL}}^{(j)}$. We ran our simulation with final time $T = 1$ with decreasing time step τ until our numerical solution became stable. In Figure 4.4a we plotted these maximum stable time-step sizes. We see that the decrease of the maximum stable time-step size matches the minimum mesh element diameter as stated by the CFL condition (4.49). Next, we turn to the Verlet method when applied to an upwind fluxes dG method. In this case the CFL condition depends on the stabilization parameter α, see (4.86). In fact, it gets stricter for a larger α, and this can be observed in Figure 4.4b.

In order to examine the temporal convergence of the Crank–Nicolson method and of the Verlet



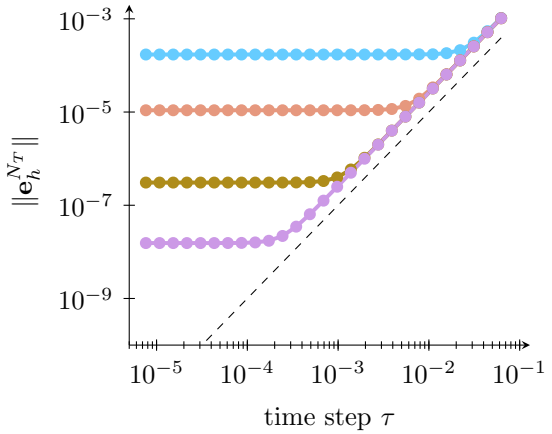
(a) Maximum stable time step of the Verlet method when applied to a central fluxes dG discretization. The black dashed line represents the slope $0.05 \min_{K \in \mathcal{T}^{(j)}_{h,\mathrm{CFL}}} h_K$.



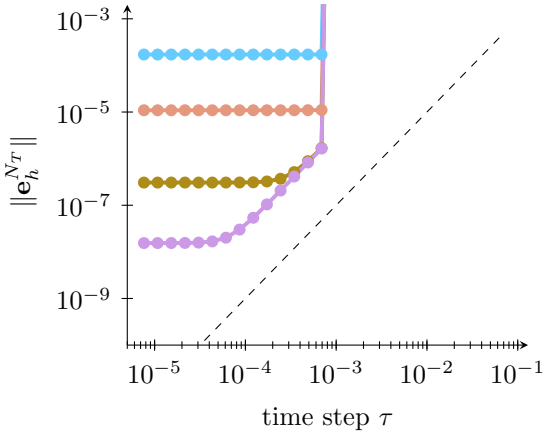
(b) Maximum stable time step of the Verlet method when applied to an upwind fluxes dG discretization with stabilization parameter α. We used the grid $\mathcal{T}^{(4)}_{h,\mathrm{CFL}}$.

Figure 4.4: Maximum stable time-step size of the Verlet method. The polynomial degrees in the dG space discretization are $k = 2$, $k = 3$, $k = 4$, $k = 5$.

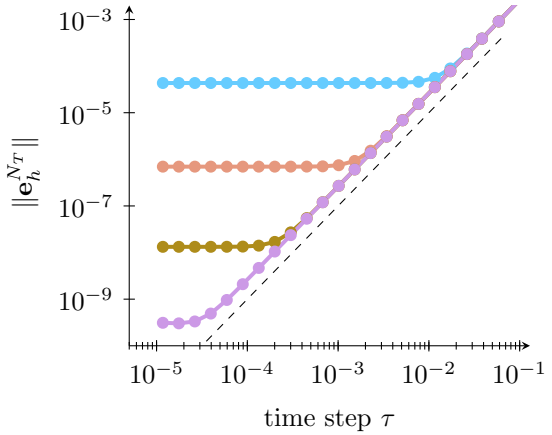
method we use the mesh sequence $\mathcal{T}^{(j)}_{h,\tau}$ and the polynomial degree $k = 5$. This rather high polynomial degree ensures that the space discretization error is small enough such that the time discretization error is dominant. In Figure 4.5 we give the graphs of the error $\mathbf{e}_h^{N_T} = \mathbf{u}_h^{N_T} - \pi_h \mathbf{u}(T)$ measured in the L^2-norm at the final time $T = t_{N_T} = 1$. They confirm the convergence order two in the time variable as proven in Theorems 4.21, 4.22 and 4.28 (the proof for the upwind fluxes Verlet method is postponed to Chapter 5). Comparing Figures 4.5a, 4.5b with Figures 4.5c, 4.5d we see again the superior space convergence of the upwind fluxes dG method compared to the central fluxes dG method, see also Section 3.7.



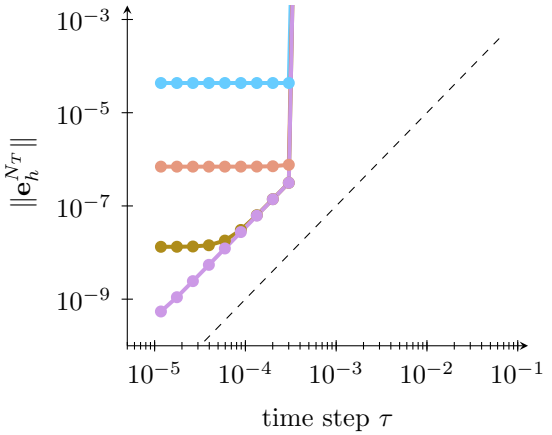
(a) Crank–Nicolson method in combination with a central fluxes dG space discretization.



(b) Verlet method in combination with a central fluxes dG space discretization.



(c) Crank–Nicolson method in combination with an upwind fluxes dG space discretization with $\alpha = 1$.



(d) Verlet method in combination with an upwind fluxes dG space discretization with $\alpha = 1$.

Figure 4.5: Temporal convergence of the Crank–Nicolson method and of the Verlet method. The final time is $T = t_{N_T} = 1$ and the polynomial degree is $k = 5$. We used the meshes $\mathcal{T}_{h,\tau}^{(1)}$, $\mathcal{T}_{h,\tau}^{(2)}$, $\mathcal{T}_{h,\tau}^{(3)}$ and $\mathcal{T}_{h,\tau}^{(4)}$. The black dashed line represents slope $\tau^2/10$.

CHAPTER 5

Locally implicit time integration

Let us recall Maxwell's equations (1.21),

$$\partial_t \mathbf{u}(t) = \mathcal{C}\mathbf{u}(t) + \mathbf{j}(t), \qquad \text{or, equivalently,} \qquad \begin{aligned} \partial_t \mathbf{H}(t) &= -\mathcal{C}_{\mathbf{E}}\mathbf{E}(t), \\ \partial_t \mathbf{E}(t) &= \mathcal{C}_{\mathbf{H}}\mathbf{H}(t) - \varepsilon^{-1}\mathbf{J}(t), \end{aligned} \tag{5.1}$$

with initial values $\mathbf{u}(0) = \mathbf{u}^0 = (\mathbf{H}^0, \mathbf{E}^0)$. Many applications require a space discretization with a **locally refined** spatial mesh, i.e. a mesh which consists mostly of coarse elements but also contains a few (very) fine elements.

5.1 Examples and overview

Let us give some examples which require such a locally refined mesh: If the domain Ω contains tiny geometric features, e.g. narrow areas as in Figure 5.1 or a barrier with a small gap as in Figure 5.2, the mesh has to be adapted to this situation which might only be possible with some small elements. As another example, observe that the convergence rate of the spatially discretized Maxwell's equations depends on the regularity of the exact solution, see Theorems 3.12 and 3.13. However, in many situations the exact solution is known to be of low regularity and thus a spatial discretization on a quasi-uniform grid fails to provide an optimal convergence rate. Examples where this phenomenon appears are domains Ω with reentrant corners, see e.g. Figure 5.3a. In such situations one can restore the optimal convergence rate by using a locally refined grid around the subdomains where the solution is of low regularity, see Figure 5.3b. For further insight we refer to Costabel and Dauge [2000], Dörfler [2013] and Nochetto et al. [2009]. As a third example let us mention the situation where the material

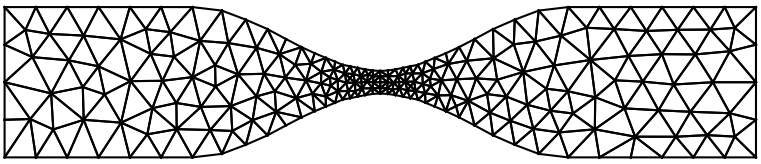
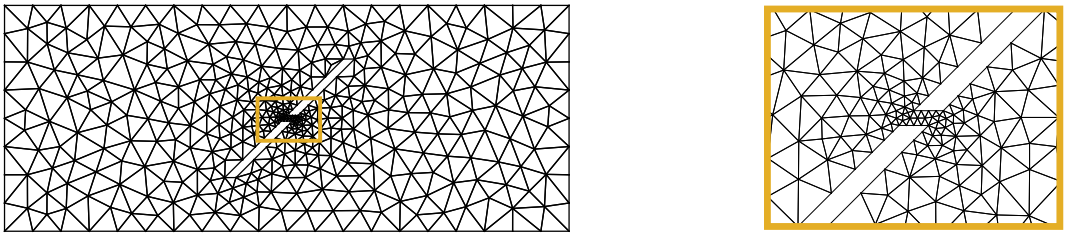


Figure 5.1: Deformed mesh.



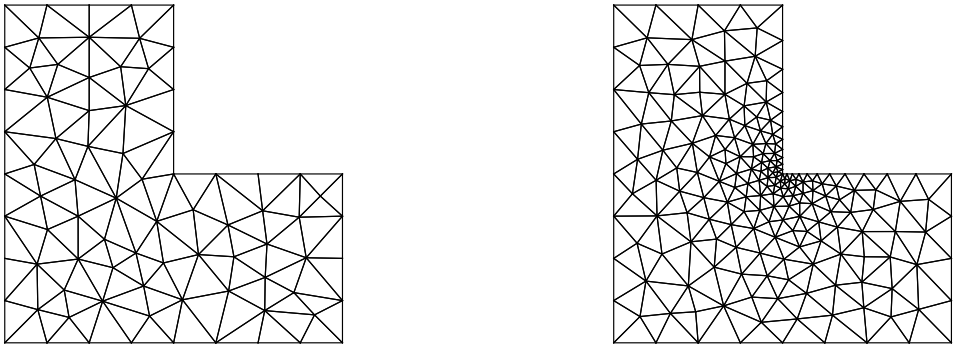
(a) Whole mesh. (b) Zoom on the gap in the barrier.

Figure 5.2: Rectangular mesh with a barrier inside that possesses a small gap in the middle.

coefficients ε and μ vary on a small spatial scale. In Figure 5.4 we depict an (adapted) example from Busch et al. [2011], where we have a ring resonator and two wave guides with permittivity $\varepsilon = 9$ in a domain which is covered with vacuum ($\varepsilon = 1$). Because our dG method requires constant material coefficients on each mesh element, we have to resolve the small (vacuum) gap between the ring wave guide and the straight wave guides with a few very tiny mesh elements.

In summary, we see that there are many situations demanding for a space discretization with a locally refined mesh. This yields a semidiscrete scheme approximating the exact solution of Maxwell's equations. For a fully discrete approximation we then have to integrate this semidiscrete scheme in time. It turns out that this is a challenging task and standard time integration methods fail to be efficient. In Chapter 4 we have seen two popular time integrators for Maxwell's equations representing the two basic classes of available time integration methods. On the one hand, we considered the Verlet method which is an example for explicit time integrators. On the other hand, the Crank–Nicolson method belongs to the class of implicit time integration schemes. Independent of the class of time integrators we want to use the **optimal time-step size**. This means we want to use the time step such that the spatial discretization error and the time integration error are (approximately) of the same size. Using a bigger time-step size results in an approximation which is not of the best possible quality (w.r.t. space discretization) while smaller time steps do not yield a better approximation but come at the cost of having to compute more time steps than necessary. So, we can conclude that using the optimal time-step size is the most efficient choice. In the particular case of locally refined meshes the space discretization error is dominated by the contribution of the coarse elements and consequently we have a rather large optimal time step size. The problem of explicit methods is that their CFL condition becomes very restrictive when we work with a locally refined spatial mesh. In fact, we are forced to use a time-step size which is considerable smaller than the optimal time-step size. This renders explicit methods inefficient for locally refined meshes. We illustrate this effect with the example of the ring resonator from above. In Figure 5.5 we give the full discretization error versus the time-step size of the Verlet method for an example using the mesh of the ring resonator. As comparison we plotted the error of the Crank–Nicolson method which indicates the space discretization errors (the plateaus in Figure 5.5) and which we use to determine the optimal time-step size. We see that due to the restrictive CFL condition we need to apply the Verlet scheme with a far too small time-step size, at least for polynomial degrees $k = 1, 2, 3$, see also Table 5.1. For higher polynomial degrees the situation seems to be better and it is possible to use the Verlet method with the optimal time-step size. However, we point out that we used a C^∞ solution for this example. This allows to access the small errors we observe in Figure 5.5 for $k = 4, 5$. However, we point out that for realistic, low regularity examples this might not be the case.

If we employ an A-stable implicit time integrator, we avoid a CFL condition. However, these methods come with the drawback that we have to solve a large linear system in each time step. In fact, this linear system involves all degrees of freedom in the spatial grid. For many



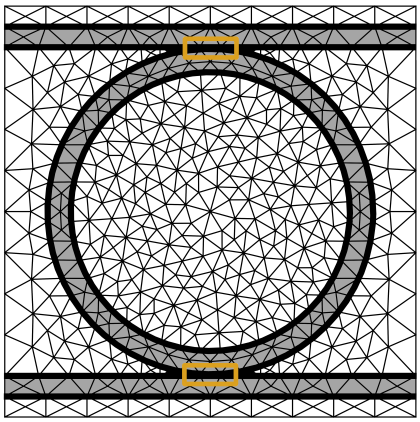
(a) Quasi-uniform mesh. (b) Locally refined mesh.

Figure 5.3: Mesh with reentrant corner.

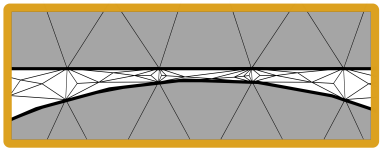
applications this is not feasible anymore. In particular, if only a small amount of the mesh elements are small, and thus require an implicit scheme, a fully implicit method is too costly or not even possible to realize for 3D problems.

In the literature several methods have been proposed as a remedy to this problem. In Diehl et al. [2010] the authors consider explicit low storage RK methods. They use the stabilization parameter α in the upwind fluxes dG discretization to tune the spectrum of the dG operator $\boldsymbol{\mathcal{C}} - \alpha\boldsymbol{\mathcal{S}}$ such that it better fits the stability region of the low storage RK methods. This allows for larger time-step sizes. Another approach are **explicit local time stepping methods** initially proposed in Diaz and Grote [2009] for the second order wave equation. Based on the explicit Verlet method, the authors construct a time integrator which uses a small time-step size on the small elements in the spatial grid while treating the coarse elements with a big time step. In numerical examples it is shown that the CFL condition of the resulting scheme only depends on the coarse part of the grid. An extension of this work to Maxwell's equations is given in Grote and Mitkova [2010]. Moreover, in Grote and Mitkova [2013] the authors derived explicit local time stepping methods of arbitrarily high order based on Adams multistep methods for the damped wave equation. Hochbruck and Ostermann [2011] showed that these methods can be interpreted as a particular implementation of exponential multistep methods (where actions of the matrix exponentials are replaced by approximations gained from explicit multistep methods). Moreover, in Demirel et al. [2015], the ideas of optimizing the stability region with respect to the shape of the field of values of the given discrete operator was used to construct optimized predictor corrector schemes which outperform the low storage RK schemes of Diehl et al. [2010]. In Grote et al. [2015] and Mehlin [2015] explicit local time stepping schemes based on explicit RK and low storage explicit RK methods instead of the Verlet method were derived. Currently, multi-level explicit local time stepping methods have been proposed. These methods take into account that a spatial mesh might consist of different areas with varying diameters of the elements. Thus, every area is treated with an adapted time-step size. In Diaz and Grote [2015] the multi-level local time stepping scheme is based on the Verlet method and in Almquist and Mehlin [2016] on RK methods.

In this thesis we consider a different approach to integrate the semidiscrete Maxwell's equations disposing locally refined meshes, namely **locally implicit time integrators**. The underlying idea of these methods is to treat the fine mesh elements in the spatial grid with an implicit time integrator, thus avoiding a restrictive CFL condition, while employing an explicit time integration scheme for the remaining coarse elements. In Piperno [2006] the author proposed such a locally implicit scheme for the homogeneous semidiscrete Maxwell's equations comprising the explicit Verlet method and the implicit midpoint method (or Crank–Nicolson method, which is the same in the homogeneous case). However, in Moya [2012] it is shown that this locally



(a) The ring resonator and the wave guides (grey areas) are separated by a small gap.



(b) Gap between the ring resonator and the wave guides.

Figure 5.4: Mesh of a ring resonator. In the white areas we have $\varepsilon = 1$ and in the grey areas $\varepsilon = 9$.

implicit method fails to retain the second order temporal convergence of the underlying schemes (unless unnatural regularity assumptions for the exact solution are demanded) and is only of order one. A different combination of the Verlet method and the Crank–Nicolson method was proposed and analyzed by Verwer [2011]. Further insight into this method and numerical examples were provided in Descombes et al. [2013] and extended in Descombes et al. [2016, 2017] to dispersive media with the focus on biological tissues. In the two papers Verwer [2011] and Descombes et al. [2013] the authors have proven that the proposed locally implicit method is second order convergent in time (Verwer [2011]) and only exhibits a CFL condition involving the coarse parts of the mesh (Descombes et al. [2013]). However, in both papers the construction and the analysis of the locally implicit method is based on a formulation of the spatially discretized Maxwell's equations as an ODE, i.e. as (3.55). As a consequence the locally implicit method is based on a splitting of the stiffness matrix C in order to assign the spatial degrees of freedom to the explicit and implicit time integration. It is left unclear how the spatial mesh has to be split and which mesh elements enter in the CFL condition. Moreover, the error analysis based on the ODE formulation given in Verwer [2011] exhibits constants depending on the spatial grid and as a consequence this analysis deteriorates if the mesh parameter h tends to 0. This means that the given error analysis is only valid in a non-stiff regime. It is based on an infinite Taylor expansion of the exact solution and unfortunately lacks the analysis of remainder terms and spatial discretization errors. Last, let us point out that the mentioned references only cover the case of the semidiscrete Maxwell's equations stemming from a central fluxes dG space discretization.

In this thesis we aim at complementing the previous work in Descombes et al. [2013], Verwer [2011]. In the following we will present a locally implicit scheme based on the ideas of Verwer [2011]. In contrary to the previous work we formulate the locally implicit scheme as a time integrator for the semidiscrete Maxwell equations in the variational formulation (3.8) or (3.15). We provide a splitting of the spatial mesh into parts which have to be treated implicitly and parts which can be treated explicitly. We already point out that this splitting does not coincide with the splitting of the mesh in coarse and fine elements, if we want to obtain a CFL condition which solely depends on the coarse mesh elements. We proceed in two steps: First, we formulate the locally implicit method for the semidiscrete Maxwell equations stemming from a central fluxes dG method. We give a stability and error analysis which is also valid in a non-stiff regime, i.e. an analysis exhibiting only constants independent of the spatial mesh. Our work is based on the techniques we presented in Chapter 4, in particular in Section 4.2, for the fully implicit

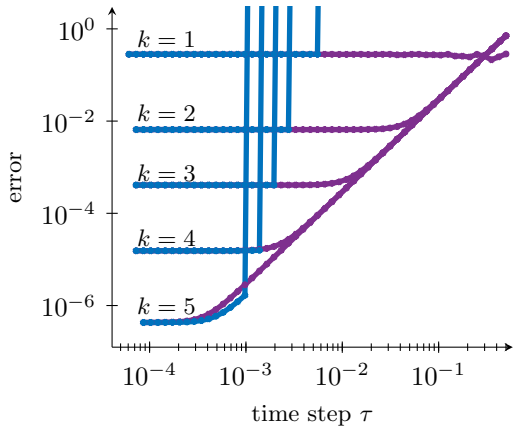


Figure 5.5: Full discretization error versus time-step size. We used the polynomial degree k in the space discretization and the Verlet method and the Crank–Nicolson method as time integrators. The exact solution is a C^∞ function.

k	1	2	3	4	5
Verlet time-step size	0.0055	0.0028	0.002	0.0014	0.00035
optimal time-step size	0.1050	0.0313	0.009	0.0014	0.00035

Table 5.1: Optimal time-step sizes.

Crank–Nicolson method and the fully explicit Verlet method. A compact version of the results can be found in Hochbruck and Sturm [2016]. Subsequently, we turn to the case of an upwind fluxes dG discretization. We show how the locally implicit scheme from the central fluxes case can be adapted to this situation and then present a stability and an error analysis. Again, we obtain a scheme with a CFL condition which only depends on the coarse mesh elements and an error analysis independent of the spatial mesh. A summary of these results can be found in Hochbruck and Sturm [2017].

5.2 Splitting of the mesh

As pointed out above we are interested in locally refined meshes, i.e. we deal with grids which are split into a coarse and a fine part

$$\mathcal{T}_h = \mathcal{T}_{h,c} \dot\cup \mathcal{T}_{h,f},$$

where the number of fine elements is small compared to the number of coarse elements,

$$0 < \operatorname{card}(\mathcal{T}_{h,f}) \ll \operatorname{card}(\mathcal{T}_{h,c}).$$

Clearly, the locally implicit time integrator has to treat the fine elements in $\mathcal{T}_{h,f}$ implicitly to that they enter the CFL condition. However, if we recall Definition 3.1 of the discrete curl-operators, we observe that each mesh element couples with its neighbors, i.e. with all elements with whom it shares a face. As a consequence, we cannot only treat the fine elements in $\mathcal{T}_{h,f}$ implicitly, but we also need to include their neighbors in the implicit time integration. All remaining elements can be treated explicitly. We fix this observation in the following definition.

Definition 5.1. *We partition the mesh $\mathcal{T}_h$ into an implicitly and an explicitly treated part defined by*

$$\mathcal{T}_{h,i} = \{K \in \mathcal{T}_h \mid \exists K_f \in \mathcal{T}_{h,f} \ : \ \left|\partial K \cap \partial K_f\right|_{d-1} \neq 0\}, \qquad \text{and} \qquad \mathcal{T}_{h,e} = \mathcal{T}_h \setminus \mathcal{T}_{h,i}, \quad (5.2a)$$

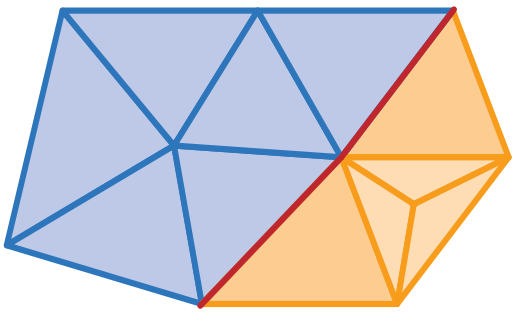


Figure 5.6: Example of the splitting of the mesh elements. The coarse elements from $\mathcal{T}_{h,c}$ are blue and dark orange. The fine elements from $\mathcal{T}_{h,f}$ are light orange. Explicitly treated elements from $\mathcal{T}_{h,e}$ are blue, implicitly treated elements from $\mathcal{T}_{h,i}$ are orange. The coarse, but implicitly treated elements from $\mathcal{T}_{h,ci}$ are dark orange. The faces in $\mathcal{F}_{h,e}^{\text{int}}$ are blue, the faces in $\mathcal{F}_{h,i}^{\text{int}}$ are orange and the faces in $\mathcal{F}_{h,ci}^{\text{int}}$ are red.

respectively. Here, $|\cdot|_{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure. Furthermore, we denote the set of implicitly treated elements which share a face with at least one explicitly treated element by

$$\mathcal{T}_{h,ci} = \{K_i \in \mathcal{T}_{h,i} \mid \exists K_e \in \mathcal{T}_{h,e} \ : \ \left|\partial K_e \cap \partial K_i\right|_{d-1} \neq 0\}. \tag{5.2b}$$

Note that the explicitly treated set only contains coarse elements. In contrast, the implicitly treated set does not only contain fine elements but also their coarse neighbors. Furthermore, all elements in $\mathcal{T}_{h,ci}$ are coarse although they are treated implicitly (as suggested by the index ci):

$$\mathcal{T}_{h,e} \subset \mathcal{T}_{h,c}, \qquad \mathcal{T}_{h,f} \subset \mathcal{T}_{h,i}, \qquad \mathcal{T}_{h,i} \cap \mathcal{T}_{h,c} \neq \emptyset, \qquad \mathcal{T}_{h,ci} \subset \mathcal{T}_{h,c} \cap \mathcal{T}_{h,i}.$$

An example for the sets is given in Figure 5.6.

Definition 5.2. *The set of interfaces is partitioned into*

$$\mathcal{F}_h^{\text{int}} = \mathcal{F}_{h,i}^{\text{int}} \,\dot\cup\, \mathcal{F}_{h,e}^{\text{int}} \,\dot\cup\, \mathcal{F}_{h,ci}^{\text{int}}, \tag{5.3a}$$

where $\mathcal{F}_{h,i}^{\text{int}}$ contains the faces between implicitly treated elements, $\mathcal{F}_{h,e}^{\text{int}}$ the faces between explicitly treated elements and $\mathcal{F}_{h,ci}^{\text{int}}$ the faces bordering an explicitly and an implicitly treated element. Furthermore, we write

$$\mathcal{F}_{h,c}^{\text{int}} = \mathcal{F}_{h,e}^{\text{int}} \,\dot\cup\, \mathcal{F}_{h,ci}^{\text{int}}. \tag{5.3b}$$

Moreover, we split the set of boundary faces into

$$\mathcal{F}_h^{\text{bnd}} = \mathcal{F}_{h,i}^{\text{bnd}} \,\dot\cup\, \mathcal{F}_{h,e}^{\text{bnd}}. \tag{5.4}$$

An example for the splitting of the mesh faces can be found in Figure 5.6. It is important to observe that the set $\mathcal{F}_{h,c}^{\text{int}}$ only contains faces bordering two coarse elements. We use the convention that for a face $F \in \mathcal{F}_{h,ci}^{\text{int}}$ the normal n_F is directed from the implicit element K_i towards the explicit element K_e, see Figure 5.7.

In our locally implicit time integrator we will assign the mesh elements to the explicit or implicit time integration with cut-off functions χ_i and χ_e,

$$\chi_i v(x) = \begin{cases} v(x), & x \in K_i,\ K_i \in \mathcal{T}_{h,i}, \\ 0, & x \in K_e,\ K_e \in \mathcal{T}_{h,e}, \end{cases} \qquad \chi_e v(x) = \begin{cases} 0, & x \in K_i,\ K_i \in \mathcal{T}_{h,i}, \\ v(x), & x \in K_e,\ K_e \in \mathcal{T}_{h,e}. \end{cases}$$

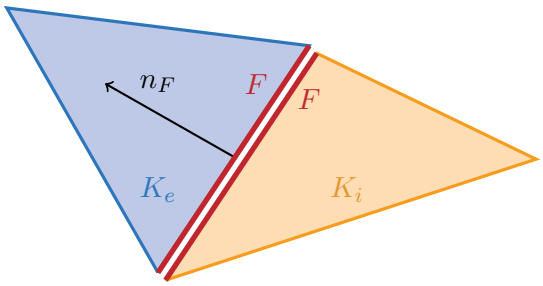


Figure 5.7: Convention for the face normal n_F in the case where the face F borders an explicit element K_e and an implicit element K_i, i.e. if $F \in \mathcal{F}^{\text{int}}_{h,ci}$.

Because the cut-off functions are matched to the mesh elements, their application to a broken polynomial yields again a broken polynomial, i.e.

$$\chi_e v_h,\ \chi_i v_h \in V_h, \qquad \text{for all } v_h \in V_h. \tag{5.5}$$

We recall from Assumption 2.24 that our mesh $\mathcal{T}_h$ is shape- and contact-regular with parameter ρ. Clearly, also the coarse part of the mesh $\mathcal{T}_{h,c}$ is shape- and contact-regular, but with parameter $\rho_c \leq \rho$ and for locally refined meshes we might have $\rho_c \ll \rho$. As a consequence the upper and lower bound for the ratio of the diameters of neighboring elements (2.4) holds true with this parameter,

$$\rho_c^{-1} \max(h_K, h_{\widehat{K}}) \leq \frac{h_K + h_{\widehat{K}}}{2} \leq \rho_c \min(h_K, h_{\widehat{K}}), \qquad h_K, h_{\widehat{K}} \in \mathcal{T}_{h,c}. \tag{5.6}$$

Moreover, the constants C_{inv} and C_{tr} in the inverse and the trace inequality (2.10) and (2.11), respectively, only depend on ρ_c on the coarse mesh $\mathcal{T}_{h,c}$. We denote these constants by $C_{\text{inv},c}$ and $C_{\text{tr},c}$. In our later analysis of the locally implicit scheme we show that its bounds only depend on these constants, i.e. on the mesh regularity of the coarse part $\mathcal{T}_{h,c}$ but not on the fine part $\mathcal{T}_{h,f}$.

5.3 Central fluxes

In Chapter 4 we saw that for the time integration of the semidiscrete Maxwell's equations we have to distinguish whether the space discretization stems from a central fluxes dG method or from an upwind fluxes dG method. The same holds true for the locally implicit time integration. So, in this section we focus on the locally implicit time integration for the space semidiscrete Maxwell's equations obtained from a central fluxes dG method, i.e.,

$$\begin{aligned} \partial_t \mathbf{H}_h(t) &= -\boldsymbol{\mathcal{C}}_{\mathbf{E}} \mathbf{E}_h(t), \\ \partial_t \mathbf{E}_h(t) &= \boldsymbol{\mathcal{C}}_{\mathbf{H}} \mathbf{H}_h(t) - \mathbf{J}_h(t), \\ \mathbf{H}_h(0) &= \mathbf{H}_h^0, \quad \mathbf{E}_h(0) = \mathbf{E}_h^0, \end{aligned} \tag{5.7}$$

see (3.8). Adapting the locally implicit scheme of Verwer [2011] we will blend the explicit Verlet method and the implicit Crank–Nicolson method, which we analyzed in Chapter 4. So, let us begin by recalling these methods. Employing the Verlet method as a time integrator for (5.7) yields the recursion

$$\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n = -\frac{\tau}{2} \boldsymbol{\mathcal{C}}_{\mathbf{E}} \mathbf{E}_h^n, \tag{5.8a}$$

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \tau \boldsymbol{\mathcal{C}}_{\mathbf{H}} \mathbf{H}_h^{n+1/2} - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n), \tag{5.8b}$$

$$\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n+1/2} = -\frac{\tau}{2} \boldsymbol{\mathcal{C}}_{\mathbf{E}} \mathbf{E}_h^{n+1}, \tag{5.8c}$$

and the Crank–Nicolson scheme for (5.7) is given by

$$\mathbf{H}_h^{n+1} - \mathbf{H}_h^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{E}_h^{n+1} + \mathbf{E}_h^n), \tag{5.9a}$$

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n), \tag{5.9b}$$

see (4.39) and (4.40), respectively.

5.3.1 Construction of the locally implicit method

As a first step in the construction of our locally implicit method we observe that the Crank–Nicolson method (5.9) can be cast into the form (5.8) of the Verlet method by splitting the update formula of $\mathbf{H}_h^{n+1}$ into two half steps. In fact, we can write the Crank–Nicolson recursion equivalently as

$$\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^n, \tag{5.10a}$$

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n), \tag{5.10b}$$

$$\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n+1/2} = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1}. \tag{5.10c}$$

This motivates a combination of the Verlet method and of the Crank–Nicolson method given in the following **locally implicit scheme**

$$\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^n, \tag{5.11a}$$

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \tau\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e\mathbf{H}_h^{n+1/2} + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}^i(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n), \tag{5.11b}$$

$$\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n+1/2} = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1}. \tag{5.11c}$$

Here, $\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e$ and $\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e$ denote the (yet to be determined) discrete curl-operators associated with the explicit mesh elements in $\mathcal{T}_{h,e}$ and analogously $\boldsymbol{\mathcal{C}}_{\mathbf{H}}^i$ and $\boldsymbol{\mathcal{C}}_{\mathbf{E}}^i$ the ones associated with $\mathcal{T}_{h,i}$. We construct these **split discrete curl-operators** by the following observations: First, it is natural to enforce that adding the split discrete curl-operators yields the original discrete curl-operators (acting on the full mesh), i.e.

$$\boldsymbol{\mathcal{C}}_{\mathbf{H}} = \boldsymbol{\mathcal{C}}_{\mathbf{H}}^i + \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e, \qquad \boldsymbol{\mathcal{C}}_{\mathbf{E}} = \boldsymbol{\mathcal{C}}_{\mathbf{E}}^i + \boldsymbol{\mathcal{C}}_{\mathbf{E}}^e. \tag{5.12}$$

Further insight is gained by casting the scheme (5.11) into the familiar notation using modified versions of the operators $\boldsymbol{\mathcal{R}}_L$ and $\boldsymbol{\mathcal{R}}_R$. As before, we write

$$\mathbf{u}_h^n = \begin{pmatrix} \mathbf{H}_h^n \\ \mathbf{E}_h^n \end{pmatrix}, \qquad \mathbf{j}_h^n = \begin{pmatrix} 0 \\ -\mathbf{J}_h^n \end{pmatrix},$$

and recall that

$$\boldsymbol{\mathcal{R}}_L = \boldsymbol{\mathcal{I}} - \frac{\tau}{2}\boldsymbol{\mathcal{C}}, \qquad \boldsymbol{\mathcal{R}}_R = \boldsymbol{\mathcal{I}} + \frac{\tau}{2}\boldsymbol{\mathcal{C}}, \qquad \boldsymbol{\mathcal{C}} = \begin{pmatrix} 0 & -\boldsymbol{\mathcal{C}}_{\mathbf{E}} \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}} & 0 \end{pmatrix}.$$

Lemma 5.3. *The recursion (5.11) of the locally implicit scheme can be written as*

$$\widetilde{\boldsymbol{\mathcal{R}}}_L\mathbf{u}_h^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}}_R\mathbf{u}_h^n + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n), \tag{5.13a}$$

with operators $\widetilde{\boldsymbol{\mathcal{R}}}_L$, $\widetilde{\boldsymbol{\mathcal{R}}}_R : V_h^2 \to V_h^2$ defined by

$$\widetilde{\boldsymbol{\mathcal{R}}}_L = \boldsymbol{\mathcal{R}}_L - \frac{\tau^2}{4}\boldsymbol{\mathcal{D}}^e, \qquad \widetilde{\boldsymbol{\mathcal{R}}}_R = \boldsymbol{\mathcal{R}}_R - \frac{\tau^2}{4}\boldsymbol{\mathcal{D}}^e, \qquad \boldsymbol{\mathcal{D}}^e = \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e\boldsymbol{\mathcal{C}}_{\mathbf{E}} \end{pmatrix}. \tag{5.13b}$$

Proof. The first component of (5.13a) is obtained by adding (5.11a) and (5.11c). For the second component we subtract (5.11c) from (5.11a):

$$\mathbf{H}_h^{n+1/2} = \frac{1}{2}(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) + \frac{\tau}{4}\mathcal{C}_{\mathbf{E}}(\mathbf{E}_h^{n+1} - \mathbf{E}_h^n).$$

Inserting this into (5.11b) we infer

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \frac{\tau}{2}\mathcal{C}_{\mathbf{H}}(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) + \frac{\tau^2}{4}\mathcal{C}_{\mathbf{H}}^e\mathcal{C}_{\mathbf{E}}(\mathbf{E}_h^{n+1} - \mathbf{E}_h^n) - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n), \tag{5.14}$$

by using $\mathcal{C}_{\mathbf{H}}^e + \mathcal{C}_{\mathbf{H}}^i = \mathcal{C}_{\mathbf{H}}$, see (5.12). □

We saw in Chapters 3 and 4 that the adjointness of $\mathcal{C}_{\mathbf{H}}$ and $\mathcal{C}_{\mathbf{E}}$,see (3.7a), is crucial for the well-posedness, the stability and the convergence of the space semi-discretization and of the full space and time discretization. So, we require this property for the explicit and the implicit split curl-operators, respectively, i.e. for all $\mathbf{H}_h, \mathbf{E}_h \in V_h$,

$$\left(\mathcal{C}_{\mathbf{H}}^i\mathbf{H}_h, \mathbf{E}_h\right)_\varepsilon = \left(\mathbf{H}_h, \mathcal{C}_{\mathbf{E}}^i\mathbf{E}_h\right)_\mu, \qquad \left(\mathcal{C}_{\mathbf{H}}^e\mathbf{H}_h, \mathbf{E}_h\right)_\varepsilon = \left(\mathbf{H}_h, \mathcal{C}_{\mathbf{E}}^e\mathbf{E}_h\right)_\mu. \tag{5.15}$$

Ensuring the conditions (5.12) and (5.15) leaves us with the choice of using either

$$\mathcal{C}_{\mathbf{H}}^b = \mathcal{C}_{\mathbf{H}} \circ \chi_b,\ \mathcal{C}_{\mathbf{E}}^b = \chi_b \circ \mathcal{C}_{\mathbf{E}}, \qquad \text{or} \qquad \mathcal{C}_{\mathbf{H}}^b = \chi_b \circ \mathcal{C}_{\mathbf{H}},\ \mathcal{C}_{\mathbf{E}}^b = \mathcal{C}_{\mathbf{E}} \circ \chi_b, \qquad b \in \{i, e\}.$$

If we also want to preserve the adjointness properties of the operators $\boldsymbol{\mathcal{R}}_L$, $\boldsymbol{\mathcal{R}}_R$ of Crank–Nicolson method and of the operators $\widehat{\boldsymbol{\mathcal{R}}}_L$, $\widehat{\boldsymbol{\mathcal{R}}}_R$ of the Verlet method given in Lemma 4.10, then the only possible option is the first one. It is easy to see that for the second option the adjointness of $\widetilde{\boldsymbol{\mathcal{R}}}_L$ and $\widetilde{\boldsymbol{\mathcal{R}}}_R$ is lost.

Definition 5.4. *We define the split discrete curl-operators as*

$$\mathcal{C}_{\mathbf{H}}^i = \mathcal{C}_{\mathbf{H}} \circ \chi_i, \qquad \mathcal{C}_{\mathbf{H}}^e = \mathcal{C}_{\mathbf{H}} \circ \chi_e, \tag{5.16a}$$

and

$$\mathcal{C}_{\mathbf{E}}^i = \chi_i \circ \mathcal{C}_{\mathbf{E}}, \qquad \mathcal{C}_{\mathbf{E}}^e = \chi_e \circ \mathcal{C}_{\mathbf{E}}. \tag{5.16b}$$

This definition immediately yields

$$\mathcal{C}_{\mathbf{H}}^e\mathcal{C}_{\mathbf{E}} = \mathcal{C}_{\mathbf{H}}^e\mathcal{C}_{\mathbf{E}}^e \qquad \text{and} \qquad \mathcal{D}^e = \begin{pmatrix} 0 & 0 \\ 0 & \mathcal{C}_{\mathbf{H}}^e\mathcal{C}_{\mathbf{E}}^e \end{pmatrix}. \tag{5.17}$$

In combination with (5.15) and Lemma 4.10 we obtain the following result.

Lemma 5.5. *Let $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h), \widetilde{\mathbf{u}}_h \in V_h^2$. Then, the operators $\widetilde{\boldsymbol{\mathcal{R}}}_L$ and $\widetilde{\boldsymbol{\mathcal{R}}}_R$ have the following properties:*

$$\left(\widetilde{\boldsymbol{\mathcal{R}}}_L\mathbf{u}_h, \widetilde{\mathbf{u}}_h\right)_{\mu\times\varepsilon} = \left(\mathbf{u}_h, \widetilde{\boldsymbol{\mathcal{R}}}_R\widetilde{\mathbf{u}}_h\right)_{\mu\times\varepsilon}, \tag{5.18a}$$

$$\left(\widetilde{\boldsymbol{\mathcal{R}}}_L\mathbf{u}_h, \mathbf{u}_h\right)_{\mu\times\varepsilon} = \left(\widetilde{\boldsymbol{\mathcal{R}}}_R\mathbf{u}_h, \mathbf{u}_h\right)_{\mu\times\varepsilon} = \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2 - \frac{\tau^2}{4}\|\mathcal{C}_{\mathbf{E}}^e\mathbf{E}_h\|_\mu^2. \tag{5.18b}$$

As we demanded, we obtain the same adjointness property as satisfied by $\boldsymbol{\mathcal{R}}_L$ and $\boldsymbol{\mathcal{R}}_R$ and by $\widehat{\boldsymbol{\mathcal{R}}}_L$ and $\widehat{\boldsymbol{\mathcal{R}}}_R$, compare (4.44a) and (4.45a) with (5.18a). Moreover, the property (5.18b) is the same as (4.45b) for the Verlet method but where $\mathcal{C}_{\mathbf{E}}$ is replaced by $\mathcal{C}_{\mathbf{E}}^e$.

5.3.2 Alternative construction of the locally implicit method

In this section we briefly present an alternative construction of the locally implicit scheme (5.11) based on the two step formulations of the Verlet method and of the Crank–Nicolson method. Since we slightly adapted the Verlet method (4.9) in respect of the inhomogeneity in order to fit Maxwell's equations, we cannot use the two step formulation (4.5) but derive a suitable alternative now. Observe that by (5.8a) and (5.8c) we have for the Verlet method

$$\begin{aligned} \mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n &= -\frac{\tau}{2}\mathcal{C}_{\mathbf{E}}\mathbf{E}_h^n, \\ \mathbf{H}_h^n - \mathbf{H}_h^{n-1/2} &= -\frac{\tau}{2}\mathcal{C}_{\mathbf{E}}\mathbf{E}_h^n, \end{aligned} \qquad \text{and thus} \qquad \mathbf{H}_h^{n+1/2} - \mathbf{H}_h^{n-1/2} = -\tau\mathcal{C}_{\mathbf{E}}\mathbf{E}_h^n. \tag{5.19}$$

Moreover, by (5.8b) we have

$$\begin{aligned} \mathbf{E}_h^n - \mathbf{E}_h^{n-1} &= \tau\mathcal{C}_{\mathbf{H}}\mathbf{H}_h^{n-1/2} - \frac{\tau}{2}\big(\mathbf{J}_h^n + \mathbf{J}_h^{n-1}\big), \\ \mathbf{E}_h^{n+1} - \mathbf{E}_h^n &= \tau\mathcal{C}_{\mathbf{H}}\mathbf{H}_h^{n+1/2} - \frac{\tau}{2}\big(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n\big), \end{aligned}$$

and by subtracting these two equations we obtain

$$\mathbf{E}_h^{n+1} - 2\mathbf{E}_h^n + \mathbf{E}_h^{n-1} = \tau\mathcal{C}_{\mathbf{H}}\big(\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^{n-1/2}\big) - \frac{\tau}{2}\big(\mathbf{J}_h^{n+1} - \mathbf{J}_h^{n-1}\big). \tag{5.20}$$

Inserting this in (5.19) we obtain the Verlet method in the desired two step formulation,

$$\mathbf{E}_h^{n+1} - 2\mathbf{E}_h^n + \mathbf{E}_h^{n-1} = -\tau^2\mathcal{C}_{\mathbf{H}}\mathcal{C}_{\mathbf{E}}\mathbf{E}_h^n - \frac{\tau}{2}(\mathbf{J}_h^{n+1} - \mathbf{J}_h^{n-1}). \tag{5.21}$$

Next, we rewrite the Crank–Nicolson method as a two step scheme. By (5.9a) we have

$$\begin{aligned} \mathbf{H}_h^n - \mathbf{H}_h^{n-1} &= -\frac{\tau}{2}\mathcal{C}_{\mathbf{E}}\big(\mathbf{E}_h^n + \mathbf{E}_h^{n-1}\big), \\ \mathbf{H}_h^{n+1} - \mathbf{H}_h^n &= -\frac{\tau}{2}\mathcal{C}_{\mathbf{E}}\big(\mathbf{E}_h^{n+1} + \mathbf{E}_h^n\big), \end{aligned} \qquad \text{and thus} \qquad \mathbf{H}_h^{n+1} - \mathbf{H}_h^{n-1} = -\frac{\tau}{2}\mathcal{C}_{\mathbf{E}}\big(\mathbf{E}_h^{n+1} + 2\mathbf{E}_h^n + \mathbf{E}_h^{n-1}\big). \tag{5.22}$$

Analogously to (5.20), we obtain from (5.9b),

$$\mathbf{E}_h^{n+1} - 2\mathbf{E}_h^n + \mathbf{E}_h^{n-1} = \frac{\tau}{2}\mathcal{C}_{\mathbf{H}}\big(\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n-1}\big) - \frac{\tau}{2}\big(\mathbf{J}_h^{n+1} - \mathbf{J}_h^{n-1}\big).$$

Inserting (5.22) into the last equation, we obtain the Crank–Nicolson method in the two step formulation,

$$\mathbf{E}_h^{n+1} - 2\mathbf{E}_h^n + \mathbf{E}_h^{n-1} = -\frac{\tau^2}{4}\mathcal{C}_{\mathbf{H}}\mathcal{C}_{\mathbf{E}}\big(\mathbf{E}_h^{n+1} + 2\mathbf{E}_h^n + \mathbf{E}_h^{n-1}\big) - \frac{\tau}{2}(\mathbf{J}_h^{n+1} - \mathbf{J}_h^{n-1}). \tag{5.23}$$

Now, we combine the Verlet method (5.21) and the Crank–Nicolson method (5.23) to the following locally implicit scheme,

$$\mathbf{E}_h^{n+1} - 2\mathbf{E}_h^n + \mathbf{E}_h^{n-1} = -\tau^2\mathcal{C}_{\mathbf{H}}^e\mathcal{C}_{\mathbf{E}}^e\mathbf{E}_h^n - \frac{\tau^2}{4}\mathcal{C}_{\mathbf{H}}^i\mathcal{C}_{\mathbf{E}}^i\big(\mathbf{E}_h^{n+1} + 2\mathbf{E}_h^n + \mathbf{E}_h^{n-1}\big) - \frac{\tau}{2}(\mathbf{J}_h^{n+1} - \mathbf{J}_h^{n-1}).$$

As in Section 5.3.1 we demand that adding the split discrete curl-operators restores the full discrete curl-operators, i.e. (5.12). Next, we consider the adjointness of $\mathcal{C}_{\mathbf{H}}$ and $\mathcal{C}_{\mathbf{E}}$,see (3.7a). For the composition of $\mathcal{C}_{\mathbf{H}}$ and $\mathcal{C}_{\mathbf{E}}$ appearing in both the Verlet and the Crank–Nicolson method this means

$$\big(\mathcal{C}_{\mathbf{H}}\mathcal{C}_{\mathbf{E}}\mathbf{E}_h, \widehat{\mathbf{E}}_h\big)_\varepsilon = \big(\mathbf{E}_h, \mathcal{C}_{\mathbf{H}}\mathcal{C}_{\mathbf{E}}\widehat{\mathbf{E}}_h\big)_\varepsilon, \qquad \mathbf{E}_h, \widehat{\mathbf{E}}_h \in V_h.$$

So, our second requirement on the split curl-operators is that they satisfy

$$\big(\mathcal{C}_{\mathbf{H}}^i\mathcal{C}_{\mathbf{E}}^i\mathbf{E}_h, \widehat{\mathbf{E}}_h\big)_\varepsilon = \big(\mathbf{E}_h, \mathcal{C}_{\mathbf{H}}^i\mathcal{C}_{\mathbf{E}}^i\widehat{\mathbf{E}}_h\big)_\varepsilon, \qquad \big(\mathcal{C}_{\mathbf{H}}^e\mathcal{C}_{\mathbf{E}}^e\mathbf{E}_h, \widehat{\mathbf{E}}_h\big)_\varepsilon = \big(\mathbf{E}_h, \mathcal{C}_{\mathbf{H}}^e\mathcal{C}_{\mathbf{E}}^e\widehat{\mathbf{E}}_h\big)_\varepsilon,$$

for all $\mathbf{E}_h, \widehat{\mathbf{E}}_h \in V_h$. It is easy to check that only the split discrete curl-operators as defined in Definition 5.4 satisfy both properties.

5.3.3 Bounds of the explicit discrete curl-operators

In this section we transfer the bounds on the full discrete curl-operators given in Theorem 3.14 to the explicit curl-operators. A crucial observation is that the explicit curl-operators can be bounded independently of the fine mesh. This is the essential ingredient for our proof that the locally implicit scheme possesses a CFL condition which solely depends on the coarse part of the mesh. Let

$$c_{\infty,c} = \max_{K\in\mathcal{T}_{h,c}} c_K \tag{5.24}$$

denote the maximum speed of light in the coarse grid.

Theorem 5.6. *For $\mathbf{H}_h, \mathbf{E}_h \in V_h$ we have the bounds*

$$\|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\mathbf{E}_h\|_\mu \le C_{\mathrm{bnd},c}c_{\infty,c}\|\mathbf{E}_h\|_{\varepsilon,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,-1}, \tag{5.25a}$$

and

$$\|\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\mathbf{H}_h\|_\varepsilon \le C_{\mathrm{bnd},c}c_{\infty,c}\|\mathbf{H}_h\|_{\mu,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,-1}, \tag{5.25b}$$

where the constant is given by $C_{\mathrm{bnd},c} = C_{\mathrm{inv},c} + 2C^2_{\mathrm{tr},c}N_\partial\rho_c$.

Although our proof follows mainly the one of Theorem 3.14 we give it here in detail such that it can be retraced without detailed knowledge from Chapter 3.

Proof. We only prove (5.25a) since the bound (5.25b) can be shown analogously. For $\mathbf{E}_h$, $\phi_h \in V_h$ we have by (3.4b) and (5.16b),

$$\begin{aligned}\big(\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\mathbf{E}_h, \phi_h\big)_\mu &= \big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h, \chi_e\phi_h\big)_\mu \\ &= \sum_{K\in\mathcal{T}_{h,e}} \big(\mathbf{E}_h, \operatorname{curl}\phi_h\big)_K + \sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} \big(\{\!\{\mathbf{E}_h\}\!\}^{\varepsilon c}_F, n_F\times[\![\chi_e\phi_h]\!]_F\big)_F\,.\end{aligned} \tag{5.26}$$

We bound the two terms on the right-hand side separately. For the first term we apply the Cauchy–Schwarz inequality twice and in between the inverse inequality (2.10) on the coarse mesh $\mathcal{T}_{h,c}$ to obtain

$$\begin{aligned}\sum_{K\in\mathcal{T}_{h,e}} \big(\mathbf{E}_h, \operatorname{curl}\phi_h\big)_K &\le C_{\mathrm{inv},c}\sum_{K\in\mathcal{T}_{h,e}} h_K^{-1}\|\mathbf{E}_h\|_K\|\phi_h\|_K \\ &= C_{\mathrm{inv},c}\sum_{K\in\mathcal{T}_{h,e}} c_K h_K^{-1}\|\mathbf{E}_h\|_{\varepsilon,K}\|\phi_h\|_{\mu,K} \\ &\le C_{\mathrm{inv},c}c_{\infty,c}\Big(\sum_{K\in\mathcal{T}_{h,e}}\|\phi_h\|^2_{\mu,K}\Big)^{1/2}\Big(\sum_{K\in\mathcal{T}_{h,e}} h_K^{-2}\|\mathbf{E}_h\|^2_{\varepsilon,K}\Big)^{1/2} \\ &\le C_{\mathrm{inv},c}c_{\infty,c}\|\phi_h\|_\mu\|\mathbf{E}_h\|_{\varepsilon,\mathcal{T}_{h,e},2,-1}.\end{aligned} \tag{5.27}$$

For the second term in (5.26), a weighted Cauchy–Schwarz inequality yields

$$\sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} \big(\{\!\{\mathbf{E}_h\}\!\}^{\varepsilon c}_F, n_F\times[\![\chi_e\phi_h]\!]_F\big)_F \le \Big(\sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} \omega_F\|n_F\times[\![\chi_e\phi_h]\!]_F\|^2_F\Big)^{1/2}\Big(\sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} \omega_F^{-1}\|\{\!\{\mathbf{E}_h\}\!\}^{\varepsilon c}_F\|^2_F\Big)^{1/2}. \tag{5.28}$$

To bound the first factor on the right-hand side, we use $|n_F| = 1$, the triangle inequality, Young's inequality, and subsequently the trace inequality (2.11) on the coarse mesh $\mathcal{T}_{h,c}$, to obtain

$$\begin{aligned}\|n_F\times[\![\chi_e\phi_h]\!]_F\|^2_F &\le 2\big(\|\chi_e\phi_h|_K\|^2_F + \|\chi_e\phi_h|_{K_F}\|^2_F\big) \\ &\le 2C^2_{\mathrm{tr},c}\big(h_K^{-1}\|\chi_e\phi_h\|^2_K + h_{K_F}^{-1}\|\chi_e\phi_h\|^2_{K_F}\big) \\ &= 2C^2_{\mathrm{tr},c}\big(\mu_K^{-1}h_K^{-1}\|\chi_e\phi_h\|^2_{\mu,K} + \mu_{K_F}^{-1}h_{K_F}^{-1}\|\chi_e\phi_h\|^2_{\mu,K_F}\big).\end{aligned} \tag{5.29}$$

Now, we choose the weight ω_F as

$$\omega_F = \frac{h_K + h_{K_F}}{2} a_F, \qquad \text{where} \qquad a_F = \frac{1}{\varepsilon_K c_K + \varepsilon_{K_F} c_{K_F}}. \tag{5.30}$$

From the shape- and contact-regularity of the coarse mesh $\mathcal{T}_{h,c}$, in particular by applying (5.6), we obtain

$$\rho_c^{-1} a_F \le \omega_F h_K^{-1},\ \omega_F h_{K_F}^{-1} \le \rho_c a_F, \qquad \text{for all } K, K_F \in \mathcal{T}_{h,c}. \tag{5.31}$$

Moreover, it is easy to see, that we have

$$a_F \le c_K \mu_K, \qquad a_F \le c_{K_F} \mu_{K_F}. \tag{5.32}$$

By (5.29), (5.31) and subsequently (5.32) we infer

$$\begin{aligned} \omega_F \|n_F \times [\![\chi_e \phi_h]\!]_F\|_F^2 &\le 2C_{\mathrm{tr},c}^2 \rho_c a_F \big(\mu_K^{-1} \|\chi_e \phi_h\|_{\mu,K}^2 + \mu_{K_F}^{-1} \|\chi_e \phi_h\|_{\mu,K_F}^2\big) \\ &\le 2C_{\mathrm{tr},c}^2 c_{\infty,c} \rho_c \|\chi_e \phi_h\|_{\mu,K\cup K_F}^2. \end{aligned} \tag{5.33}$$

For the second factor on the right-hand side of (5.28) we use the triangle inequality, Young's inequality and the trace inequality (2.11) on the coarse mesh $\mathcal{T}_{h,c}$. This yields

$$\begin{aligned} \omega_F^{-1} \|\{\!\{\mathbf{E}_h\}\!\}_F^{\varepsilon c}\|_F^2 &\le 2a_F \omega_F^{-1} \big(c_K \|\mathbf{E}_h|_K\|_{\varepsilon,F}^2 + c_{K_F} \|\mathbf{E}_h|_{K_F}\|_{\varepsilon,F}^2\big) \\ &\le 2C_{\mathrm{tr},c}^2 a_F c_{\infty,c} \omega_F^{-1} \big(h_K^{-1} \|\mathbf{E}_h\|_{\varepsilon,K}^2 + h_{K_F}^{-1} \|\mathbf{E}_h\|_{\varepsilon,K_F}^2\big) \\ &\le 2C_{\mathrm{tr},c}^2 c_{\infty,c} \rho_c \big(h_K^{-2} \|\mathbf{E}_h\|_{\varepsilon,K}^2 + h_{K_F}^{-2} \|\mathbf{E}_h\|_{\varepsilon,K_F}^2\big). \end{aligned} \tag{5.34}$$

Here, we further used the obvious bounds

$$a_F \varepsilon_K c_K \le 1, \qquad a_F \varepsilon_{K_F} c_{K_F} \le 1, \tag{5.35}$$

in the first inequality and (5.31) for the last inequality. Inserting (5.33) and (5.34) in (5.28) we obtain

$$\begin{aligned} \sum_{F \in \mathcal{F}_{h,c}^{\mathrm{int}}} \big(\{\!\{\mathbf{E}_h\}\!\}_F^{\varepsilon c}, n_F \times [\![\chi_e \phi_h]\!]_F\big)_F &\le 2C_{\mathrm{tr},c}^2 N_\partial c_{\infty,c} \rho_c \|\phi_h\|_{\mu,\mathcal{T}_{h,e}} \|\mathbf{E}_h\|_{\varepsilon,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,-1} \\ &\le 2C_{\mathrm{tr},c}^2 N_\partial c_{\infty,c} \rho_c \|\phi_h\|_\mu \|\mathbf{E}_h\|_{\varepsilon,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,-1}. \end{aligned} \tag{5.36}$$

Last, we insert the bounds (5.27) and (5.36) in (5.26) and use the identity

$$\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e \mathbf{E}_h\|_\mu = \sup_{\phi_h \in V_h, \|\phi_h\|_\mu = 1} \big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e \mathbf{E}_h, \phi_h\big)_\mu.$$

This proves the statement. □

So far, the split discrete curl-operators inherited the properties of the full operators. By the construction of $\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e$ and $\boldsymbol{\mathcal{C}}_{\mathbf{E}}^i$ this also holds true for the consistency property (3.6). In fact, for $\mathbf{E} \in V_\star^{\mathbf{E}}$ we have that

$$\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e \mathbf{E} = \chi_e(\pi_h \mathcal{C}_{\mathbf{E}} \mathbf{E}), \qquad \boldsymbol{\mathcal{C}}_{\mathbf{E}}^i \mathbf{E} = \chi_i(\pi_h \mathcal{C}_{\mathbf{E}} \mathbf{E}). \tag{5.37}$$

Clearly, this yields the bounds

$$\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e \mathbf{E}\|_\mu \le \delta^{-1/2} \|\mathbf{E}\|_{H(\mathrm{curl},\Omega^e)}, \qquad \|\boldsymbol{\mathcal{C}}_{\mathbf{E}}^i \mathbf{E}\|_\mu \le \delta^{-1/2} \|\mathbf{E}\|_{H(\mathrm{curl},\Omega^i)}, \qquad \text{for all } \mathbf{E} \in V_\star^{\mathbf{E}}, \tag{5.38}$$

with δ from (1.20). Here, Ω^e and Ω^i correspond to the explicitly and implicitly treated part of the domain Ω, respectively, i.e.,

$$\Omega^e = \bigcup_{K \in \mathcal{T}_{h,e}} \overline{K}, \qquad \Omega^i = \bigcup_{K \in \mathcal{T}_{h,i}} \overline{K}.$$

Unfortunately, a uniform bound like (5.38) cannot by obtained for $\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e$ and $\boldsymbol{\mathcal{C}}_{\mathbf{H}}^i$, but only one involving $h_K^{-1/2}$.

Lemma 5.7. *For $\mathbf{H} \in V_\star^{\mathbf{H}}$ we have the bound*

$$\|\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \mathbf{H}\|_\varepsilon \leq \delta^{-1/2} \|\mathbf{H}\|_{H(\mathrm{curl},\Omega^e)} + C'_{\mathrm{bnd},c} \delta^{-1/2} \left(\sum_{F \in \mathcal{F}_{h,ci}^{\mathrm{int}}} h_{K_e}^{-1} \|\mathbf{H}\|_{1,K_e}^2 \right)^{1/2}, \tag{5.39}$$

where K_e denotes the explicit element corresponding to a face $F \in \mathcal{F}_{h,ci}^{\mathrm{int}}$ and the constant is given by $C'_{\mathrm{bnd},c} = \sqrt{2} C_{\mathrm{ctr},c} C_{\mathrm{tr},c} N_\partial \rho_c$.

Proof. Let K_e and K_i denote the explicit and implicit element corresponding to a face $F \in \mathcal{F}_{h,c}^{\mathrm{int}}$, respectively, see Figure 5.7. Note that both elements are coarse, $K_e, K_i \in \mathcal{T}_{h,c}$. Employing $\mathbf{H} \in V_\star^{\mathbf{H}}$ and $\psi_h \in V_h$ in (3.5a), we have

$$\begin{aligned}
\big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \mathbf{H}, \psi_h\big)_\varepsilon &= \sum_{K \in \mathcal{T}_h} \big(\mathrm{curl}(\chi_e \mathbf{H}), \psi_h\big)_K + \sum_{F \in \mathcal{F}_h^{\mathrm{int}}} \big(n_F \times [\![\chi_e \mathbf{H}]\!]_F, \{\!\{\psi_h\}\!\}_F^{\varepsilon c}\big)_F \\
&= \sum_{K \in \mathcal{T}_{h,e}} \big(\mathrm{curl}\, \mathbf{H}, \psi_h\big)_K + \sum_{F \in \mathcal{F}_{h,e}^{\mathrm{int}}} \big(n_F \times [\![\mathbf{H}]\!]_F, \{\!\{\psi_h\}\!\}_F^{\varepsilon c}\big)_F \\
&\quad + \sum_{F \in \mathcal{F}_{h,ci}^{\mathrm{int}}} \big(n_F \times [\![\chi_e \mathbf{H}]\!]_F, \{\!\{\psi_h\}\!\}_F^{\varepsilon c}\big)_F \\
&= \sum_{K \in \mathcal{T}_{h,e}} \big(\mathrm{curl}\, \mathbf{H}, \psi_h\big)_K + \sum_{F \in \mathcal{F}_{h,ci}^{\mathrm{int}}} \big(n_F \times \mathbf{H}|_{K_e}, \{\!\{\psi_h\}\!\}_F^{\varepsilon c}\big)_F.
\end{aligned}$$

Here, for the last equality we exploited that by (3.2a) for all $\mathbf{H} \in V_\star^{\mathbf{H}} = D(\mathfrak{C}_{\mathbf{H}}) \cap H^1(\mathcal{T}_h)^3$ it holds that

$$n_F \times [\![\mathbf{H}]\!]_F = 0, \qquad \text{for all } F \in \mathcal{F}_h^{\mathrm{int}}.$$

For the first term we obtain from the Cauchy–Schwarz inequality

$$\sum_{K \in \mathcal{T}_{h,e}} \big(\mathrm{curl}\, \mathbf{H}, \psi_h\big)_K \leq \|\,\mathrm{curl}\, \mathbf{H}\|_{\mathcal{T}_{h,e}} \|\psi_h\|_{\mathcal{T}_{h,e}} \leq \delta^{-1/2} \|\mathbf{H}\|_{H(\mathrm{curl},\Omega^e)} \|\psi_h\|_\varepsilon.$$

For the second term we use the Cauchy–Schwarz inequality with weight $\widehat{\omega}_F = (h_{K_e} + h_{K_i})/2$ and $|n_F| = 1$, to obtain

$$\sum_{F \in \mathcal{F}_{h,ci}^{\mathrm{int}}} \big(n_F \times \mathbf{H}|_{K_e}, \{\!\{\psi_h\}\!\}_F^{\varepsilon c}\big)_F \leq \left(\sum_{F \in \mathcal{F}_{h,ci}^{\mathrm{int}}} \widehat{\omega}_F^{-1} \|\mathbf{H}|_{K_e}\|_F^2 \right)^{1/2} \left(\sum_{F \in \mathcal{F}_{h,ci}^{\mathrm{int}}} \widehat{\omega}_F \|\{\!\{\psi_h\}\!\}_F^{\varepsilon c}\|_F^2 \right)^{1/2}.$$

By the continuous trace inequality (2.13) on the coarse mesh $\mathcal{T}_{h,c}$ and (5.6) we infer

$$\widehat{\omega}_F^{-1} \|\mathbf{H}|_{K_e}\|_F^2 \leq C_{\mathrm{ctr},c}^2 \widehat{\omega}_F^{-1} \|\mathbf{H}\|_{1,K_e}^2 \leq C_{\mathrm{ctr},c}^2 \rho_c h_{K_e}^{-1} \|\mathbf{H}\|_{1,K_e}^2.$$

By the triangle inequality, Young's inequality and the discrete trace inequality (2.11) on the coarse mesh, we have

$$\begin{aligned}
\widehat{\omega}_F \|\{\!\{\psi_h\}\!\}_F^{\varepsilon c}\|_F^2 \leq \widehat{\omega}_F \|\{\!\{\psi_h\}\!\}_F\|_F^2 &\leq 2\widehat{\omega}_F \big(\varepsilon_{K_e} \|\psi_h\|_{\varepsilon,F}^2 + \varepsilon_{K_i} \|\psi_h\|_{\varepsilon,F}^2\big) \\
&\leq 2C_{\mathrm{tr},c}^2 \widehat{\omega}_F \big(\varepsilon_{K_e} h_{K_e}^{-1} \|\psi_h\|_{\varepsilon,K_e}^2 + \varepsilon_{K_i} h_{K_i}^{-1} \|\psi_h\|_{\varepsilon,K_i}^2\big) \\
&\leq 2C_{\mathrm{tr},c}^2 \rho_c \delta^{-1} \big(\|\psi_h\|_{\varepsilon,K_e}^2 + \|\psi_h\|_{\varepsilon,K_i}^2\big).
\end{aligned}$$

Here, we further used (5.35) in the first inequality and

$$\widehat{\omega}_F h_{K_e}^{-1},\ \widehat{\omega}_F h_{K_i}^{-1} \leq \rho_c,$$

see (5.30) and (5.31). In summary, we have

$$\sum_{F\in\mathcal{F}_{h,ci}^{\mathrm{int}}} \big(n_F \times \mathbf{H}|_{K_e}, \{\!\{\psi_h\}\!\}_F^{\varepsilon c}\big)_F \le \sqrt{2} C_{\mathrm{ctr},c} C_{\mathrm{tr},c} N_\partial \rho_c \delta^{-1/2} \|\psi_h\|_\varepsilon \left(\sum_{F\in\mathcal{F}_{h,ci}^{\mathrm{int}}} h_{K_e}^{-1} \|\mathbf{H}\|_{1,K_e}^2 \right)^{1/2}.$$

Applying

$$\|\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \mathbf{H}\|_\varepsilon = \sup_{\psi_h\in V_h, \|\psi_h\|_\varepsilon = 1} \big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \mathbf{H}, \psi_h\big)_\varepsilon,$$

gives the statement. □

5.3.4 Analysis of the locally implicit method

In this section we prove the well-posedness and the stability of the locally implicit scheme (5.11) under a CFL condition that solely depends on the size of the mesh elements in the coarse mesh $\mathcal{T}_{h,c}$: Let $0 < \widetilde{\theta} < 1$ be an arbitrary but fixed parameter. Then, the **CFL condition of the locally implicit scheme** reads

$$\tau \le \frac{2\widetilde{\theta}}{C_{\mathrm{bnd},c} c_{\infty,c}} \min_{K\in\mathcal{T}_{h,c}} h_K, \tag{5.40}$$

where $C_{\mathrm{bnd},c}$ was defined in Theorem 5.6 and $c_{\infty,c}$ in (5.24).

We have seen in Section 4.2.1 that the CFL condition of the Verlet method (4.49) ensures the invertibility of the operator $\widehat{\boldsymbol{\mathcal{R}}}_L$ and the boundedness of $\widehat{\boldsymbol{\mathcal{R}}}^m = (\widehat{\boldsymbol{\mathcal{R}}}_L^{-1}\widehat{\boldsymbol{\mathcal{R}}}_R)^m$ for all $m \in \mathbb{N}$, see Lemmas 4.14 and 4.15. The same holds true for the analog operators of the locally implicit method but under the weakened CFL condition (5.40).

Lemma 5.8. *Let $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h) \in V_h^2$ and assume that the CFL condition (5.40) is satisfied with a $\widetilde{\theta} \in (0,1)$. Then, we have*

$$(1-\widetilde{\theta}^2)\|\mathbf{u}_h\|_{\mu\times\varepsilon}^2 \le \big(\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} \le \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2. \tag{5.41}$$

In particular, $\widetilde{\boldsymbol{\mathcal{R}}}_L$ is invertible with bound

$$\|\widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}\|_{\mu\times\varepsilon} \le C_{\mathrm{stb},c}, \qquad C_{\mathrm{stb},c} = (1-\widetilde{\theta}^2)^{-1}. \tag{5.42}$$

Moreover, for all $m \in \mathbb{N}$, $\widetilde{\boldsymbol{\mathcal{R}}} = \widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}\widetilde{\boldsymbol{\mathcal{R}}}_R$ satisfies

$$\|\widetilde{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h\|_{\mu\times\varepsilon}^2 \le C_{\mathrm{stb},c}\Big(\|\mathbf{u}_h\|_{\mu\times\varepsilon}^2 - \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e \mathbf{E}_h\|_\mu^2\Big) \qquad \text{and} \qquad \|\widetilde{\boldsymbol{\mathcal{R}}}^m\|_{\mu\times\varepsilon} \le C_{\mathrm{stb},c}^{1/2}. \tag{5.43}$$

Proof. For this proof we follow the ones of Lemmas 4.14 and 4.15 .

The upper bound in (5.41) is clear by (5.18b). For the lower bound we use Theorem 5.6 and the CFL condition (5.40) to infer

$$\frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e \mathbf{E}_h\|_\mu^2 \le \frac{\tau^2}{4} C_{\mathrm{bnd},c}^2 c_{\infty,c}^2 \|\mathbf{E}_h\|_{\varepsilon,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,-1}^2 \le \widetilde{\theta}^2 \|\mathbf{E}_h\|_{\varepsilon,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci}}^2 \le \widetilde{\theta}^2 \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2.$$

Together with (5.18b) this proves the lower bound.

In order to bound $\widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}$ we use

$$\|\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h\|_{\mu\times\varepsilon} = \sup_{\mathbf{v}_h\in V_h^2} \frac{\big(\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h, \mathbf{v}_h\big)_{\mu\times\varepsilon}}{\|\mathbf{v}_h\|_{\mu\times\varepsilon}} \ge \frac{\big(\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon}}{\|\mathbf{u}_h\|_{\mu\times\varepsilon}} \ge (1-\widetilde{\theta}^2)\|\mathbf{u}_h\|_{\mu\times\varepsilon},$$

which follows from (5.41). Consequently, $\widetilde{\boldsymbol{\mathcal{R}}}_L$ is an isomorphism on V_h^2 and by setting $\mathbf{v}_h = \widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h$ we obtain the first bound in (5.43).

As in the proof of Lemma 4.11, an induction argument shows

$$\big(\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} = \big(\widetilde{\boldsymbol{\mathcal{R}}}_L \widetilde{\boldsymbol{\mathcal{R}}} \mathbf{u}_h, \widetilde{\boldsymbol{\mathcal{R}}} \mathbf{u}_h\big)_{\mu\times\varepsilon} = \ldots = \big(\widetilde{\boldsymbol{\mathcal{R}}}_L \widetilde{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h, \widetilde{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h\big)_{\mu\times\varepsilon}, \qquad m = 1, 2, \ldots .$$

Together with (5.18b) and (5.41) this implies

$$(1-\widetilde{\theta}^2)\|\widetilde{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h\|_{\mu\times\varepsilon}^2 \leq \big(\widetilde{\boldsymbol{\mathcal{R}}}_L \widetilde{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h, \widetilde{\boldsymbol{\mathcal{R}}}^m \mathbf{u}_h\big)_{\mu\times\varepsilon} = \|\mathbf{u}_h\|_{\mu\times\varepsilon}^2 - \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e \mathbf{E}_h\|_\mu^2,$$

$m = 1, 2, \ldots$, which completes the proof. □

This lemma enables us to write the locally implicit scheme (5.13a) as

$$\mathbf{u}_h^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}} \mathbf{u}_h^n + \frac{\tau}{2}\widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}\big(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n\big) = \widetilde{\boldsymbol{\mathcal{R}}}^{n+1}\mathbf{u}_h^0 + \frac{\tau}{2}\sum_{m=0}^{n} \widetilde{\boldsymbol{\mathcal{R}}}^{n-m}\widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}\big(\mathbf{j}_h^{m+1} + \mathbf{j}_h^m\big), \tag{5.44}$$

if the time step τ satisfies the CFL condition (5.40). This representation of the locally implicit method together with Lemma 5.8 allows us to prove that the locally implicit method preserves a perturbed electromagnetic energy and furthermore it allows us to prove the stability of the scheme. We give these results in the subsequent two corollaries.

Corollary 5.9. *Assume that the CFL condition (5.40) is satisfied with parameter $\widetilde{\theta} \in (0,1)$. Then, for vanishing source term $\mathbf{J}_h \equiv 0$, the approximation $\mathbf{u}_h^n = (\mathbf{H}_h^n, \mathbf{E}_h^n)$ obtained from the locally implicit scheme (5.11) conserves the perturbed electromagnetic energy*

$$\widetilde{\mathcal{E}}(\mathbf{H}_h, \mathbf{E}_h) = \mathcal{E}(\mathbf{H}_h, \mathbf{E}_h) - \frac{\tau^2}{8}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e \mathbf{E}_h\|_\mu^2, \tag{5.45}$$

i.e., $\widetilde{\mathcal{E}}(\mathbf{H}_h^n, \mathbf{E}_h^n) = \widetilde{\mathcal{E}}(\mathbf{H}_h^0, \mathbf{E}_h^0)$, $n = 1, 2, \ldots$.

Proof. For $\mathbf{J}_h \equiv 0$ the locally implicit method reads $\mathbf{u}_h^n = \widetilde{\boldsymbol{\mathcal{R}}}^n \mathbf{u}_h^0$, see (5.44). Thus, the proof of the previous lemma shows that

$$\big(\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h^n, \mathbf{u}_h^n\big)_{\mu\times\varepsilon} = \big(\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h^0, \mathbf{u}_h^0\big)_{\mu\times\varepsilon}. \tag{5.46}$$

The statement then follows from (5.18b). □

Note that the locally implicit method conserves the same perturbed energy as the Verlet method, but involving the explicit discrete curl-operator $\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e$ instead of the full discrete curl-operator $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$, see (4.54).

The next corollary adresses the stability of the locally implicit scheme.

Corollary 5.10. *Assume that the CFL condition (5.40) is satisfied with parameter $\widetilde{\theta} \in (0,1)$. Then, the approximation $\mathbf{u}_h^n$ obtained from the locally implicit method (5.11) is bounded by*

$$\|\mathbf{u}_h^n\|_{\mu\times\varepsilon} \leq C_{\mathrm{stb},c}^{1/2}\|\mathbf{u}^0\|_{\mu\times\varepsilon} + C_{\mathrm{stb},c}^{3/2}\frac{\tau}{2\sqrt{\delta}}\sum_{m=0}^{n-1}\|\mathbf{J}^{m+1} + \mathbf{J}^m\|. \tag{5.47}$$

Proof. Taking the norm of (5.44) and using the triangle inequality, (5.42), (5.43) and (3.12) gives the statement. □

We observe that the locally implicit scheme satisfies a stability bound analogous to the one of the Verlet method. The difference between the two schemes is that the locally implicit scheme only requires a CFL condition on the coarse mesh to ensure stability and that the stability bound involves $C_{\mathrm{bnd},c}$ rather than C_{bnd}, cf. (4.55).

5.3.5 Error analysis of the locally implicit scheme

For our error analysis we recall some notations from Chapter 4. By $\mathbf{u}^n = (\mathbf{H}^n, \mathbf{E}^n) = (\mathbf{H}(t_n), \mathbf{E}(t_n))$ we denote the exact solution of Maxwell's equations (5.1) at time t_n and by $\mathbf{u}_h^n = (\mathbf{H}_h^n, \mathbf{E}_h^n) \approx \mathbf{u}^n$ we denote the approximation obtained by the central fluxes dG discretization and the locally implicit scheme (5.11). As before, we split the full discretization error into

$$\mathbf{e}^n = \mathbf{u}^n - \mathbf{u}_h^n = (\mathbf{u}^n - \pi_h \mathbf{u}^n) - (\mathbf{u}_h^n - \pi_h \mathbf{u}^n) = \mathbf{e}_\pi^n - \mathbf{e}_h^n. \tag{5.48}$$

We note that we already obtained a bound on the projection error $\mathbf{e}_\pi^n$ in Chapter 3, cf. (3.24a). In the next lemma we provide a recursion for the remaining error $\mathbf{e}_h^n$. It turns out that it satisfies a perturbed version of the locally implicit scheme (5.11).

Lemma 5.11. *Let $\mathbf{u} \in C\big(0,T;V_\star\big) \cap C^3\big(0,T;L^2(\Omega)^6\big)$ be the exact solution of (5.1). The error $\mathbf{e}_h^n$ satisfies*

$$\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{e}_h^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}}_R \mathbf{e}_h^n + \widetilde{\mathbf{d}}^n. \tag{5.49}$$

The defect $\widetilde{\mathbf{d}}^n = \widetilde{\mathbf{d}}_\pi^n + \widetilde{\mathbf{d}}_h^n$ is given by

$$\widetilde{\mathbf{d}}_\pi^n = \mathbf{d}_\pi^n - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \boldsymbol{\mathcal{C}}_{\mathbf{E}} \big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big) \end{pmatrix}, \qquad \widetilde{\mathbf{d}}_h^n = \mathbf{d}_h^n - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \pi_h \Delta_{\mathbf{H}}^n \end{pmatrix}, \tag{5.50}$$

where $\mathbf{d}_\pi^n$, $\mathbf{d}_h^n$ were defined in (4.58) and $\Delta_{\mathbf{H}}^n$ was defined in (4.63b).

Proof. We follow the proof of Lemma 4.19. In (4.59) we obtained the following recursion for the projected exact solution,

$$\boldsymbol{\mathcal{R}}_L \pi_h \mathbf{u}^{n+1} = \boldsymbol{\mathcal{R}}_R \pi_h \mathbf{u}^n + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) - \mathbf{d}^n. \tag{5.51}$$

If we insert the projected exact solution into the locally implicit scheme (5.11), we obtain

$$\widetilde{\boldsymbol{\mathcal{R}}}_L \pi_h \mathbf{u}^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}}_R \pi_h \mathbf{u}^n + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) - \widetilde{\mathbf{d}}^n. \tag{5.52}$$

Subtracting (5.52) from (5.13a) yields the stated recursion (5.49) and it remains to determine the defect $\widetilde{\mathbf{d}}^n$. We subtract (5.52) from (5.51),

$$\begin{aligned}
\widetilde{\mathbf{d}}^n &= \mathbf{d}^n + (\boldsymbol{\mathcal{R}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_L)\pi_h \mathbf{u}^{n+1} - (\boldsymbol{\mathcal{R}}_R - \widetilde{\boldsymbol{\mathcal{R}}}_R)\pi_h \mathbf{u}^n \\
&= \mathbf{d}^n + \frac{\tau^2}{4}\boldsymbol{\mathcal{D}}^e \pi_h (\mathbf{u}^{n+1} - \mathbf{u}^n), \\
&= \mathbf{d}^n + \frac{\tau^2}{4}\boldsymbol{\mathcal{D}}^e \big(\mathbf{u}^{n+1} - \mathbf{u}^n - (\mathbf{e}_\pi^{n+1} - \mathbf{e}_\pi^n)\big).
\end{aligned}$$

Here, the second equality follows by the definitions of the operators $\boldsymbol{\mathcal{R}}_L$, $\boldsymbol{\mathcal{R}}_R$ and $\widetilde{\boldsymbol{\mathcal{R}}}_L$, $\widetilde{\boldsymbol{\mathcal{R}}}_R$, see (4.41b) and (5.13b), respectively. The components of $\widetilde{\mathbf{d}}^n = (\widetilde{\mathbf{d}}_{\mathbf{H}}^n, \widetilde{\mathbf{d}}_{\mathbf{E}}^n)$ are given by

$$\widetilde{\mathbf{d}}_{\mathbf{H}}^n = \mathbf{d}_{\mathbf{H}}^n, \qquad \widetilde{\mathbf{d}}_{\mathbf{E}}^n = \mathbf{d}_{\mathbf{E}}^n + \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \boldsymbol{\mathcal{C}}_{\mathbf{E}} \big(\mathbf{E}^{n+1} - \mathbf{E}^n - (\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n)\big).$$

From the consistency of the discrete curl-operators, cf. (3.6), we conclude

$$\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \boldsymbol{\mathcal{C}}_{\mathbf{E}} (\mathbf{E}^{n+1} - \mathbf{E}^n) = \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \pi_h \mathcal{C}_{\mathbf{E}} (\mathbf{E}^{n+1} - \mathbf{E}^n) = -\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \pi_h (\partial_t \mathbf{H}^{n+1} - \partial_t \mathbf{H}^n) = -\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \pi_h \Delta_{\mathbf{H}}^n.$$

Here, the second equality is obtained via Maxwell's equations (1.21), in particular by differentiating $\partial_t \mathbf{H} = -\mathcal{C}_{\mathbf{E}} \mathbf{E}$ w.r.t. t. Combining the last three equations we end up with

$$\widetilde{\mathbf{d}}^n = \mathbf{d}^n - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \pi_h \Delta_{\mathbf{H}}^n \end{pmatrix} - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \boldsymbol{\mathcal{C}}_{\mathbf{E}} \big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big) \end{pmatrix},$$

which completes the proof. □

If we assume that the CFL condition (5.40) is satisfied, we can rewrite the error recursion (5.49) as

$$\mathbf{e}_h^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}}\mathbf{e}_h^n + \widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}\widetilde{\mathbf{d}}^n = \sum_{m=0}^{n} \widetilde{\boldsymbol{\mathcal{R}}}^{n-m}\widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}\widetilde{\mathbf{d}}^m, \tag{5.53}$$

since $\mathbf{e}_h^0 = 0$. Now, we give a bound on the defect $\widetilde{\mathbf{d}}^m$.

Lemma 5.12. *Let* $\mathbf{u} \in C\big(0,T;D(\mathcal{C}) \cap H^{k+1}(\mathcal{T}_h)^6\big) \cap C^3\big(0,T;L^2(\Omega)^6\big)$ *be the exact solution of* (5.1)*. If the CFL condition* (5.40) *is fulfilled, we have that*

$$\|\widetilde{\mathbf{d}}_\pi^n\|_{\mu\times\varepsilon} \leq \widehat{C}_\pi \frac{\tau}{2}\big(|\mathbf{u}^{n+1} + \mathbf{u}^n|_{k+1,\mathcal{T}_h,2,k} + |\mathbf{E}^{n+1} - \mathbf{E}^n|_{k+1,\mathcal{T}_h,2,k}\big). \tag{5.54a}$$

If we assume more regularity for $\mathbf{H}$*, in particular* $\mathbf{H} \in C^2(0,T;V_\star^{\mathbf{H}})$*, we obtain*

$$\|\widetilde{\mathbf{d}}_h^n\|_{\mu\times\varepsilon} \leq \frac{\tau^2}{8}\int_{t_n}^{t_{n+1}} \|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon} + C|\partial_t^2\mathbf{H}(t)|_{1,\mathcal{T}_{h,e}}\,dt + C\tau^{3/2}\int_{t_n}^{t_{n+1}} \|\partial_t^2\mathbf{H}(t)\|_{1,\mathcal{T}_{h,e}}\,dt. \tag{5.54b}$$

The constant C *depends on* $C_{\mathrm{bnd},c}$, $C'_{\mathrm{bnd},c}$, C_{app}, $\widehat{C}_\pi$, $c_{\infty,c}$, *and* δ.

If we compare the bounds from Lemma 5.12 with those of Lemma 4.20, we observe that the bound on $\widetilde{\mathbf{d}}_\pi^n$ is of the correct order, namely k in the space variable. However, the defect $\widetilde{\mathbf{d}}_h^n$ is only of order 2.5 in time compared to the order 3 of the defects $\mathbf{d}_h^n$ and $\widehat{\mathbf{d}}_h^n$ of the Crank–Nicolson and the Verlet method, respectively. If we would use this bound on $\widetilde{\mathbf{d}}_h^n$, we could only prove a temporal convergence order of 1.5 for the locally implicit method. This would imply that the locally implicit method suffers from an **order reduction** in the temporal convergence. If we consider the proof below, we see that the problem of the reduced order of $\widetilde{\mathbf{d}}_h^n$ lies in the loss of the consistency of the explicit curl-operator $\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e$, cf. Section 5.3.3 and in particular Lemma 5.7. However, we point out that the locally implicit scheme (5.11) **does not suffer from an order reduction**, which we will prove in the following. Yet, we first give the proof of the lemma.

Proof. This proof follows the proof of Lemma 4.20.

(a) The first part of the defect $\widetilde{\mathbf{d}}_\pi^n$ was already bounded in Lemma 4.20. For the second part observe that in the last term $\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{e}_{\pi,\mathbf{E}}^{m+1} - \mathbf{e}_{\pi,\mathbf{E}}^m) \in V_h$ and consequently we can apply Theorem 5.6. This yields

$$\begin{aligned}
\frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big)\|_\varepsilon &\leq \frac{\tau^2}{4}C_{\mathrm{bnd},c}c_{\infty,c}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big)\|_{\mu,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,-1} \\
&\leq \frac{\tau}{2}\widetilde{\theta}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big)\|_{\mu,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci}} \\
&\leq \frac{\tau}{2}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n\big)\|_\mu \\
&\leq \widehat{C}_\pi\frac{\tau}{2}|\mathbf{E}^{n+1} - \mathbf{E}^n|_{k+1,\mathcal{T}_h,2,k}.
\end{aligned} \tag{5.55}$$

Here, the second inequality is obtained via the CFL condition (5.40) and the last inequality follows from (3.29b).

(b) Next, we consider the two terms of the defect $\widetilde{\mathbf{d}}_h^n$ given in (5.50). In Lemma 4.20 we already derived a bound for the first part. For the second part we decompose $\pi_h\Delta_{\mathbf{H}}^n = \Delta_{\mathbf{H}}^n - \Delta_\pi^n$, where Δ_π^n is defined as

$$\Delta_\pi^n = \Delta_{\mathbf{H}}^n - \pi_h\Delta_{\mathbf{H}}^n = \int_{t_n}^{t_{n+1}} \partial_t^2\mathbf{e}_{\pi,\mathbf{H}}(t)\,dt.$$

We have

$$\|\mathcal{C}^e_{\mathbf{H}}\Delta^n_\pi\|_\varepsilon \le \int_{t_n}^{t_{n+1}} \|\mathcal{C}^e_{\mathbf{H}}(\partial_t^2 \mathbf{e}_{\pi,\mathbf{H}}(t))\|_\varepsilon \, dt \le \widehat{C}_\pi \int_{t_n}^{t_{n+1}} |\partial_t^2 \mathbf{H}(t)|_{1,\mathcal{T}_{h,e}} \, dt.$$

Here, the last inequality is obtained by the following computations:

$$\begin{aligned}\big(\mathcal{C}^e_{\mathbf{H}}\mathbf{e}_{\pi,\mathbf{H}}, \psi_h\big)_\varepsilon = \big(\mathcal{C}_{\mathbf{H}}(\chi_e\mathbf{H} - \chi_e\pi_h\mathbf{H}), \psi_h\big)_\varepsilon &= \big(\mathcal{C}_{\mathbf{H}}(\chi_e\mathbf{H} - \pi_h\chi_e\mathbf{H}), \psi_h\big)_\varepsilon \\ &= \big(\mathcal{C}_{\mathbf{H}}\mathbf{e}_{\pi,\chi_e\mathbf{H}}, \psi_h\big)_\varepsilon \\ &\le \widehat{C}_\pi \|\psi_h\|_\varepsilon |\chi_e\mathbf{H}|_{k+1,\mathcal{T}_h,2,k} \\ &= \widehat{C}_\pi \|\psi_h\|_\varepsilon |\mathbf{H}|_{k+1,\mathcal{T}_{h,e},2,k}.\end{aligned}$$

In the second equality we were allowed to interchange χ_e and π_h, since χ_e is matched to the mesh elements in the spatial grid $\mathcal{T}_h$. The last inequality is obtained via (3.29b). Moreover, we have

$$\|\mathcal{C}^e_{\mathbf{H}}\Delta^n_{\mathbf{H}}\|_\varepsilon \le \int_{t_n}^{t_{n+1}} \|\mathcal{C}^e_{\mathbf{H}}(\partial_t^2 \mathbf{H}(t))\|_\varepsilon \, dt,$$

and Lemma 5.7 yields

$$\begin{aligned}\frac{\tau^2}{4}\|\mathcal{C}^e_{\mathbf{H}}(\partial_t^2\mathbf{H})\|_\varepsilon &\le \frac{\tau^2}{4\delta^{1/2}}\|\partial_t^2\mathbf{H}\|_{H(\mathrm{curl},\Omega)} + C'_{\mathrm{bnd},c}\frac{\tau^{3/2}}{4\delta^{1/2}}\left(\sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,ci}} \tau h_{K_e}^{-1}\|\partial_t^2\mathbf{H}\|^2_{1,K_e}\right)^{1/2} \\ &\le \frac{\tau^2}{4\delta^{1/2}}\|\partial_t^2\mathbf{H}\|_{H(\mathrm{curl},\Omega)} + \frac{C'_{\mathrm{bnd},c}}{(C_{\mathrm{bnd},c}c_{\infty,c})^{1/2}}\frac{\tau^{3/2}}{2\sqrt{2}\delta^{1/2}}\|\partial_t^2\mathbf{H}\|_{1,\mathcal{T}_{h,e}},\end{aligned}$$

because of the CFL condition (5.40). This completes the proof. □

Recalling the error analysis for the Verlet method in Section 4.2.2 we now rewrite the second term in the defect $\widetilde{\mathbf{d}}^n_h$. A crucial observation is that, by the definition of the split discrete curl-operator $\mathcal{C}^e_{\mathbf{H}} = \mathcal{C}_{\mathbf{H}} \circ \chi_e$, we can transfer the idea from the Verlet method to the locally implicit case. In fact, for all $\mathbf{H}_h \in V_h$, we have that

$$\begin{pmatrix} 0 \\ -\tau\mathcal{C}^e_{\mathbf{H}}\mathbf{H}_h \end{pmatrix} = \begin{pmatrix} 0 & \tau\mathcal{C}_{\mathbf{E}} \\ -\tau\mathcal{C}_{\mathbf{H}} & 0 \end{pmatrix}\begin{pmatrix} \chi_e\mathbf{H}_h \\ 0 \end{pmatrix} = -\tau\mathcal{C}\begin{pmatrix} \chi_e\mathbf{H}_h \\ 0 \end{pmatrix} = (\widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R)\begin{pmatrix} \chi_e\mathbf{H}_h \\ 0 \end{pmatrix}. \tag{5.56}$$

Using this identity we can write the defect $\widetilde{\mathbf{d}}^n_h$ as

$$\widetilde{\mathbf{d}}^n_h = \mathbf{d}^n_h - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \mathcal{C}^e_{\mathbf{H}}\pi_h\Delta^n_{\mathbf{H}} \end{pmatrix} = \mathbf{d}^n_h + (\widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R)\widetilde{\boldsymbol{\xi}}^n, \qquad \widetilde{\boldsymbol{\xi}}^n = \begin{pmatrix} \widetilde{\boldsymbol{\xi}}^n_{\mathbf{H}} \\ \widetilde{\boldsymbol{\xi}}^n_{\mathbf{E}} \end{pmatrix} = \frac{\tau}{4}\begin{pmatrix} \chi_e\pi_h\Delta^n_{\mathbf{H}} \\ 0 \end{pmatrix}, \tag{5.57a}$$

and split the defect $\widetilde{\mathbf{d}}^n = \widetilde{\mathbf{d}}^n_\pi + \widetilde{\mathbf{d}}^n_h$ into

$$\widetilde{\mathbf{d}}^n = \widetilde{\boldsymbol{\eta}}^n + (\widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R)\widetilde{\boldsymbol{\xi}}^n, \qquad \widetilde{\boldsymbol{\eta}}^n = \widetilde{\mathbf{d}}^n_\pi + \mathbf{d}^n_h. \tag{5.57b}$$

Inserting this splitting into the error recursion (5.53), we obtain

$$\mathbf{e}^{n+1}_h = \widetilde{\boldsymbol{\xi}}^n - \widetilde{\boldsymbol{\mathcal{R}}}^{n+1}\widetilde{\boldsymbol{\xi}}^0 + \sum_{m=0}^{n}\widetilde{\boldsymbol{\mathcal{R}}}^{n-m}\widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}\widetilde{\boldsymbol{\eta}}^m - \sum_{m=0}^{n-1}\widetilde{\boldsymbol{\mathcal{R}}}^{n-m}(\widetilde{\boldsymbol{\xi}}^{m+1} - \widetilde{\boldsymbol{\xi}}^m), \tag{5.58}$$

with the same computations as for (4.73). Now we have all ingredients to prove our main result, namely the **convergence result of the full discretization** of the **locally implicit scheme**.

Theorem 5.13. *Let* $\mathbf{u} \in C\big(0,T;D(\mathcal{C})\cap H^{k+1}(\mathcal{T}_h)^6\big)\cap C^3\big(0,T;L^2(\Omega)^6\big)$ *be the exact solution of* (5.1)*. Moreover, assume that the CFL condition* (5.40) *is satisfied with* $\widetilde{\theta}\in(0,1)$*. Then, the error of the central fluxes dG discretization and the locally implicit scheme* (5.11) *satisfies*

$$\begin{aligned}\|\mathbf{u}^n-\mathbf{u}_h^n\|_{\mu\times\varepsilon}\leq& C_{\mathrm{app}}|\mathbf{u}^n|_{k+1,\mathcal{T}_h,1,k+1}\\&+C_{\mathrm{stb},c}^{3/2}\widehat{C}_\pi\frac{\tau}{2}\sum_{m=0}^{n-1}\Big(|\mathbf{u}^{m+1}+\mathbf{u}^m|_{k+1,\mathcal{T}_h,2,k}+|\mathbf{E}^{m+1}-\mathbf{E}^m|_{k+1,\mathcal{T}_h,2,k}\Big)\\&+(1+C_{\mathrm{stb},c}^{1/2})\frac{\tau^2}{4}\max_{t\in[0,t_n]}\|\partial_t^2\mathbf{H}(t)\|_{\mu,\mathcal{T}_{h,e}}\\&+C_{\mathrm{stb},c}^{1/2}(4+C_{\mathrm{stb},c})\frac{\tau^2}{8}\int_0^{t_n}\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}\,dt\\\leq& C\Big(h^k+\tau^2\Big).\end{aligned}\tag{5.59}$$

The constant C *only depends on* C_{app}, $\widehat{C}_\pi$, $\widetilde{\theta}$, $|\mathbf{u}(t)|_{k+1,\mathcal{T}_h}$, $\|\partial_t^2\mathbf{H}(t)\|_\mu$ *and* $\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}$, $t\in[0,t_n]$.

Proof. We take norms in (5.58), use the triangle inequality, (5.42) and (5.43), which gives

$$\|\mathbf{e}_h^n\|_{\mu\times\varepsilon}\leq\|\widetilde{\boldsymbol{\xi}}^{n-1}\|_{\mu\times\varepsilon}+C_{\mathrm{stb},c}^{1/2}\|\widetilde{\boldsymbol{\xi}}^0\|_{\mu\times\varepsilon}+C_{\mathrm{stb},c}^{3/2}\sum_{m=0}^{n-1}\|\widetilde{\boldsymbol{\eta}}^m\|_{\mu\times\varepsilon}+C_{\mathrm{stb}}^{1/2}\sum_{m=0}^{n-2}\|\widetilde{\boldsymbol{\xi}}^{m+1}-\widetilde{\boldsymbol{\xi}}^m\|_{\mu\times\varepsilon}.$$

The defect $\widetilde{\boldsymbol{\eta}}^m$ can be bounded with (4.68a) and (5.54a),

$$\|\widetilde{\boldsymbol{\eta}}^m\|_{\mu\times\varepsilon}\leq\widehat{C}_\pi\frac{\tau}{2}\big(|\mathbf{u}^{m+1}+\mathbf{u}^m|_{k+1,\mathcal{T}_h,2,k}+|\mathbf{E}^{m+1}-\mathbf{E}^m|_{k+1,\mathcal{T}_h,2,k}\big)+\frac{\tau^2}{8}\int_{t_m}^{t_{m+1}}\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}\,dt.$$

For $\widetilde{\boldsymbol{\xi}}^m$, observe that $\widetilde{\boldsymbol{\xi}}^m=\chi_e\widehat{\boldsymbol{\xi}}^m$, where $\widehat{\boldsymbol{\xi}}^m$ was defined in (4.70a). From (4.74) and (4.75) we infer

$$\|\widetilde{\boldsymbol{\xi}}^{m+1}-\widetilde{\boldsymbol{\xi}}^m\|_{\mu\times\varepsilon}=\|\widetilde{\boldsymbol{\xi}}_{\mathbf{H}}^{m+1}-\widetilde{\boldsymbol{\xi}}_{\mathbf{H}}^m\|_\mu=\|\widehat{\boldsymbol{\xi}}_{\mathbf{H}}^{m+1}-\widehat{\boldsymbol{\xi}}_{\mathbf{H}}^m\|_{\mu,\mathcal{T}_{h,e}}\leq\frac{\tau^2}{4}\int_{t_m}^{t_{m+2}}\|\partial_t^3\mathbf{H}(t)\|_{\mu,\mathcal{T}_{h,e}}\,dt,\tag{5.60}$$

and

$$\|\widetilde{\boldsymbol{\xi}}^{n-1}\|_{\mu\times\varepsilon}\leq\frac{\tau^2}{4}\max_{t\in[t_{n-1},t_n]}\|\partial_t^2\mathbf{H}(t)\|_{\mu,\mathcal{T}_{h,e}},\qquad\|\widetilde{\boldsymbol{\xi}}^0\|_{\mu\times\varepsilon}\leq\frac{\tau^2}{4}\max_{t\in[0,\tau]}\|\partial_t^2\mathbf{H}(t)\|_{\mu,\mathcal{T}_{h,e}}.\tag{5.61}$$

The result now follows by applying the triangle inequality to the full discretization error $\mathbf{e}^n=\mathbf{e}_\pi^n-\mathbf{e}_h^n$, and using (3.24a) for the projection error. □

5.4 Upwind fluxes

In this section we extend the locally implicit scheme (5.11) from a central fluxes dG space discretization to an upwind fluxes dG method. We recall that an upwind fluxes dG discretization of Maxwell's equations reads

$$\begin{aligned}\partial_t\mathbf{H}_h(t)&=-\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h(t)-\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h(t),\\\partial_t\mathbf{E}_h(t)&=\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h(t)-\alpha\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}_h(t)-\mathbf{J}_h(t),\\\mathbf{H}_h(0)&=\mathbf{H}_h^0,\quad\mathbf{E}_h(0)=\mathbf{E}_h^0,\end{aligned}\tag{5.62}$$

see (3.15).

5.4.1 Construction of the locally implicit method

In Section 4.3 we presented the Crank–Nicolson method when applied to (5.62),

$$\begin{aligned}
\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n &= -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^n - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h^n,\\
\mathbf{E}_h^{n+1} - \mathbf{E}_h^n &= \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{E}}(\mathbf{E}_h^{n+1} + \mathbf{E}_h^n) - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n),\\
\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n+1/2} &= -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1} - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h^{n+1},
\end{aligned} \tag{5.63}$$

and an adaption of the Verlet method for (5.62),

$$\begin{aligned}
\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n &= -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^n - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h^n,\\
\mathbf{E}_h^{n+1} - \mathbf{E}_h^n &= \tau\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{H}_h^{n+1/2} - \tau\alpha\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}_h^n - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n),\\
\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n+1/2} &= -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1} - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}_h^n,
\end{aligned} \tag{5.64}$$

see (4.76) and (4.78), respectively. We recall that by our construction of the scheme (5.64), it is fully explicit, which is a desired property for a Verlet-type method. However, it comes with the drawback that the stabilization operators contribute to the CFL condition of the method, see (4.86) and how it enters the proof of Corollary 4.26. Recalling Definition 3.5 of the stabilization operators, we see that they involve every mesh element in the spatial grid. As a consequence, we cannot use the full stabilization operators in our locally implicit scheme, since this would lead to an integrator with a CFL condition depending on the whole mesh $\mathcal{T}_h$. As a remedy we propose to use in place of the full stabilization operators $\boldsymbol{\mathcal{S}}_{\mathbf{H}}$ and $\boldsymbol{\mathcal{S}}_{\mathbf{E}}$ explicit versions $\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e$ and $\boldsymbol{\mathcal{S}}_{\mathbf{E}}^e$ of these operators. In summary, we base our locally implicit time integrator on the Verlet scheme (5.64), since it is fully explicit, replace the full stabilization operators by their explicit (yet to be defined) counterparts, and incorporate the Crank–Nicolson method analogous to the central fluxes locally implicit method. The resulting **locally implicit scheme for an upwind fluxes dG discretization** then reads

$$\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^n - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e\mathbf{H}_h^n, \tag{5.65a}$$

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \tau\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e\mathbf{H}_h^{n+1/2} + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}^i(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) - \tau\alpha\boldsymbol{\mathcal{S}}_{\mathbf{E}}^e\mathbf{E}_h^n - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n), \tag{5.65b}$$

$$\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n+1/2} = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1} - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e\mathbf{H}_h^n. \tag{5.65c}$$

It remains to define the explicit stabilization operators, which we do in the next section.

5.4.2 The explicit stabilization operators

Recall from Lemma 3.6 that the full stabilization operators given by

$$\begin{aligned}
\left(\boldsymbol{\mathcal{S}}_{\mathbf{H}}\mathbf{H}, \phi_h\right)_\mu &= \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} a_F\left(n_F\times[\![\mathbf{H}]\!]_F, n_F\times[\![\phi_h]\!]_F\right)_F,\\
\left(\boldsymbol{\mathcal{S}}_{\mathbf{E}}\mathbf{E}, \psi_h\right)_\varepsilon &= \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} b_F\left(n_F\times[\![\mathbf{E}]\!]_F, n_F\times[\![\psi_h]\!]_F\right)_F + \sum_{F\in\mathcal{F}_h^{\mathrm{bnd}}} b_F\left(n_F\times\mathbf{E}, n_F\times\psi_h\right)_F,
\end{aligned}$$

are consistent, symmetric, and positive semi-definite. They solely take values of the functions on faces into account. Hence, it is natural to construct explicit stabilization operators by replacing the sums over all faces by sums over faces belonging to explicit elements, i.e., by the sets $\mathcal{F}_{h,c}^{\mathrm{int}}$ and $\mathcal{F}_{h,e}^{\mathrm{bnd}}$, cf. Definition 5.2. We fix this idea in the following definition.

Definition 5.14. *We define the* ***explicit stabilization operators*** $\boldsymbol{\mathcal{S}}^e_{\mathbf{H}} : V^{\mathbf{H}}_{\star,h} \to V_h$ *such that for all* $\phi_h \in V_h$,

$$\big(\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\mathbf{H}, \phi_h\big)_\mu = \sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} a_F\big(n_F \times [\![\mathbf{H}]\!]_F, n_F \times [\![\phi_h]\!]_F\big)_F \ , \tag{5.66a}$$

and $\boldsymbol{\mathcal{S}}^e_{\mathbf{E}} : V^{\mathbf{E}}_{\star,h} \to V_h$ *such that for all* $\psi_h \in V_h$,

$$\big(\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}\mathbf{E}, \psi_h\big)_\varepsilon = \sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} b_F\big(n_F \times [\![\mathbf{E}]\!]_F, n_F \times [\![\psi_h]\!]_F\big)_F + \sum_{F\in\mathcal{F}^{\mathrm{bnd}}_{h,e}} b_F\big(n_F \times \mathbf{E}, n_F \times \psi_h\big)_F \ , \tag{5.66b}$$

where a_F *and* b_F *were defined in* (3.13). *Moreover, we define*

$$\boldsymbol{\mathcal{S}}^e : V_{\star,h} \to V_h^2, \qquad \boldsymbol{\mathcal{S}}^e = \begin{pmatrix} \boldsymbol{\mathcal{S}}^e_{\mathbf{H}} & 0 \\ 0 & \boldsymbol{\mathcal{S}}^e_{\mathbf{E}} \end{pmatrix}. \tag{5.66c}$$

The explicit stabilization operators share important properties with their full counterparts.

Lemma 5.15. *The stabilization operators* $\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}$ *and* $\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}$ *have the following properties:*

(a) *They are* ***consistent****, i.e. for* $\mathbf{u} = (\mathbf{H}, \mathbf{E}) \in V_\star$ *we have*

$$\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\mathbf{H} = 0, \qquad \boldsymbol{\mathcal{S}}^e_{\mathbf{E}}\mathbf{E} = 0, \qquad \boldsymbol{\mathcal{S}}^e\mathbf{u} = 0. \tag{5.67}$$

(b) *They are* ***symmetric*** *on* V_h*, i.e. for* $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h)$, $\widehat{\mathbf{u}}_h = (\widehat{\mathbf{H}}_h, \widehat{\mathbf{E}}_h) \in V_h^2$ *they satisfy*

$$\begin{aligned} \big(\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\mathbf{H}_h, \widehat{\mathbf{H}}_h\big)_\mu &= \big(\mathbf{H}_h, \boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\widehat{\mathbf{H}}_h\big)_\mu \ , \qquad \big(\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}\mathbf{E}_h, \widehat{\mathbf{E}}_h\big)_\varepsilon = \big(\mathbf{E}_h, \boldsymbol{\mathcal{S}}^e_{\mathbf{E}}\widehat{\mathbf{E}}_h\big)_\varepsilon \ , \\ \big(\boldsymbol{\mathcal{S}}^e\mathbf{u}_h, \widehat{\mathbf{u}}_h\big)_{\mu\times\varepsilon} &= \big(\mathbf{u}_h, \boldsymbol{\mathcal{S}}^e\widehat{\mathbf{u}}_h\big)_{\mu\times\varepsilon} \ . \end{aligned} \tag{5.68}$$

(c) *The are* ***positive semi-definite*** *on* V_h*, i.e. for* $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h) \in V_h^2$ *it holds that*

$$\big(\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\mathbf{H}_h, \mathbf{H}_h\big)_\mu \geq 0, \qquad \big(\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}\mathbf{E}_h, \mathbf{E}_h\big)_\varepsilon \geq 0, \qquad \big(\boldsymbol{\mathcal{S}}^e\mathbf{u}_h, \mathbf{u}_h\big)_{\mu\times\varepsilon} \geq 0. \tag{5.69}$$

Proof. Analogous to Lemma 3.6. □

Remark 5.16. It is easy to see that it is not possible to define stabilization operators in a similar way as the discrete curl-operators by means of cut-off functions such that they inherit all properties in the previous lemma. On the other hand, splitting the discrete curl-operators $\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}$, $\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}$, $\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}$ and $\boldsymbol{\mathcal{C}}^i_{\mathbf{E}}$ as in Definition 5.14 by replacing the full set of faces in the full operators (3.4) by the sets of faces boardering explicit (for $\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}$, $\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}$) or implicit elements (for $\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}$, $\boldsymbol{\mathcal{C}}^i_{\mathbf{E}}$) leads to operators losing the adjointness property (5.15).

As for the full stabilization operators, we associate a seminorm with the explicit stabilization operators.

Definition 5.17. *For* $\mathbf{u} = (\mathbf{H}, \mathbf{E}) \in V_{\star,h}$ *we define the* ***seminorms***

$$|\mathbf{H}|^2_{\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}} = \sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} a_F \|n_F \times [\![\mathbf{H}]\!]_F\|^2_F \ , \tag{5.70a}$$

$$|\mathbf{E}|^2_{\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}} = \sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} b_F \|n_F \times [\![\mathbf{E}]\!]_F\|^2_F + \sum_{F\in\mathcal{F}^{\mathrm{bnd}}_{h,e}} b_F \|n_F \times \mathbf{E}\|^2_F \ . \tag{5.70b}$$

Moreover, we set

$$|\mathbf{u}|^2_{\boldsymbol{\mathcal{S}}^e} = |\mathbf{H}|^2_{\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}} + |\mathbf{E}|^2_{\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}} \ . \tag{5.70c}$$

Note that for $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h) \in V_h^2$ we can represent these seminorms with the stabilization operators by

$$|\mathbf{H}_h|^2_{\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}} = \left(\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\mathbf{H}_h, \mathbf{H}_h\right)_\mu , \qquad |\mathbf{E}_h|^2_{\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}} = \left(\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}\mathbf{E}_h, \mathbf{E}_h\right)_\varepsilon , \qquad |\mathbf{u}_h|^2_{\boldsymbol{\mathcal{S}}^e} = \left(\boldsymbol{\mathcal{S}}^e\mathbf{u}_h, \mathbf{u}_h\right)_{\mu\times\varepsilon} . \tag{5.71}$$

We conclude this section by transferring the results from Theorems 3.10 and 3.14 to the explicit stabilization operators.

Theorem 5.18. *For $\mathbf{u}_h \in V_h^2$ we have the bound*

$$|\mathbf{u}_h|_{\boldsymbol{\mathcal{S}}^e} \leq \left(\widehat{C}_{\mathrm{bnd},c} c_{\infty,c}\right)^{1/2} \|\mathbf{u}_h\|_{\mu\times\varepsilon,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,-\frac{1}{2}}, \tag{5.72}$$

with constant $\widehat{C}_{\mathrm{bnd},c} = 2C^2_{\mathrm{tr},c}N_\partial$.

Proof. For $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h)$ we have $|\mathbf{u}_h|^2_{\boldsymbol{\mathcal{S}}^e} = |\mathbf{H}_h|^2_{\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}} + |\mathbf{E}_h|^2_{\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}}$, where by Definition 5.17

$$|\mathbf{H}_h|^2_{\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}} = \sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} a_F \|n_F \times [\![\mathbf{H}_h]\!]_F\|^2_F. \tag{5.73}$$

By $|n_F| = 1$, the triangle inequality, Young's inequality, the trace inequality (2.11) on the coarse mesh $\mathcal{T}_{h,c}$, and (5.35) we infer

$$\begin{aligned} a_F\|n_F \times [\![\mathbf{H}_h]\!]_F\|^2_F &\leq 2C^2_{\mathrm{tr},c}a_F\left(\varepsilon_K c^2_K h_K^{-1}\|\mathbf{H}_h\|^2_{\mu,K} + \varepsilon_{K_F}c^2_{K_F}h^{-1}_{K_F}\|\mathbf{H}_h\|^2_{\mu,K_F}\right) \\ &\leq 2C^2_{\mathrm{tr},c}c_{\infty,c}\left(h_K^{-1}\|\mathbf{H}_h\|^2_{\mu,K} + h^{-1}_{K_F}\|\mathbf{H}_h\|^2_{\mu,K_F}\right). \end{aligned}$$

Inserting this bound into (5.73) gives

$$|\mathbf{H}_h|^2_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}} \leq \widehat{C}_{\mathrm{bnd},c}c_{\infty,c}\|\mathbf{H}_h\|^2_{\mu,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,-\frac{1}{2}}.$$

The proof of the bound for $|\mathbf{E}_h|^2_{\boldsymbol{\mathcal{S}}_{\mathbf{E}}}$ is done analogously. □

Theorem 5.19. *Let $\mathbf{u} \in V_\star \cap H^{k+1}(\mathcal{T}_h)^6$. Then, for all $\varphi_h \in V_h^2$, the projection error $\mathbf{e}_\pi = \mathbf{u} - \pi_h\mathbf{u}$ satisfies*

$$\begin{aligned} \left(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon} \leq& C_{\pi,c}|\varphi_h|_{\boldsymbol{\mathcal{S}}^e}|\mathbf{u}|_{k+1,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,k+\frac{1}{2}} \\ &+ \widehat{C}_\pi\|\varphi_h\|_{\mu\times\varepsilon,\mathcal{T}_{h,i}}|\mathbf{u}|_{k+1,\mathcal{T}_{h,i},2,k}, \end{aligned} \tag{5.74}$$

and

$$\left(\boldsymbol{\mathcal{S}}^e\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon} \leq C_{\pi,c}|\varphi_h|_{\boldsymbol{\mathcal{S}}^e}|u|_{k+1,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,k+\frac{1}{2}}. \tag{5.75}$$

The constants are given by $C_{\pi,c} = \left(2N_\partial c_{\infty,c}\right)^{1/2}\widehat{C}_{\mathrm{app}}$ and $\widehat{C}_\pi = 2\widehat{C}_{\mathrm{app}}C_{\mathrm{tr}}N_\partial c_\infty\rho$.

Remark 5.20. (a) The bound (5.74) combines the results (3.29a) and (3.29b) for the full discrete curl-operator $\boldsymbol{\mathcal{C}}$ which we used in the convergence proofs in the upwind fluxes case and in the central fluxes case, respectively. On the elements that are stabilized by $\boldsymbol{\mathcal{S}}^e$, we can use a bound similar to (3.29a) and obtain the higher convergence rate $k + 1/2$ in the spatial variable. On the remaining elements we are forced to use an estimate like (3.29b), which leaves us only with convergence order k. The result (5.75) is the counterpart of (3.30) for the explicit stabilization operator $\boldsymbol{\mathcal{S}}^e$ instead of the full stabilization operator $\boldsymbol{\mathcal{S}}$.

(b) Both $|\mathbf{u}|_{k+1,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,k+\frac{1}{2}}$ and $|\mathbf{u}|_{k+1,\mathcal{T}_{h,i},2,k}$ involve $\mathbf{u}$ on the set $\mathcal{T}_{h,ci}$. In fact, the former involves $|\mathbf{u}|_{k+1,\mathcal{T}_{h,ci},2,k+\frac{1}{2}}$ and the latter involves $|\mathbf{u}|_{k+1,\mathcal{T}_{h,ci},2,k}$. This results in the convergence rate k on the (very few) coarse elements in $\mathcal{T}_{h,ci}$. It also might happen that a very small amount of coarse mesh elements belongs to $\mathcal{T}_{h,i} \setminus \mathcal{T}_{h,ci}$ (e.g. if a coarse mesh element possesses

only fine neighbors). Consequently, we only obtain the convergence rate $k+1/2$ on the set of explicitly treated elements $\mathcal{T}_{h,e}$ rather than on the whole set of coarse elements $\mathcal{T}_{h,c}$. However, an advantage of dG methods is their flexibility in choosing a different polynomial degree on each mesh element. As a consequence, if we choose the polynomial degree $k+1$ on the (small number of) mesh elements in $\mathcal{T}_{h,c} \cap \mathcal{T}_{h,i}$, we obtain the rate $k+1/2$ on the whole coarse set. Particularly, we obtain

$$\left(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon} \leq C_{\pi,c}|\varphi_h|_{\boldsymbol{\mathcal{S}}^e}|\mathbf{u}|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}} + \widehat{C}_\pi\|\varphi_h\|_{\mu\times\varepsilon,\mathcal{T}_{h,i}}|\mathbf{u}|_{k+1,\mathcal{T}_{h,f},2,k}, \tag{5.76}$$

and

$$\left(\boldsymbol{\mathcal{S}}^e\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon} \leq C_{\pi,c}|\varphi_h|_{\boldsymbol{\mathcal{S}}^e}|u|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}. \tag{5.77}$$

(c) In the following we will use (for a shorter notation) the bounds (5.74) and (5.75) w.r.t. the set $\mathcal{T}_{h,c}$ instead of $\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci}$, and leave it to the reader to recall that by the idea from (b) they can be sharpened to (5.76) and (5.77), respectively.

Proof. (a) We start with the proof of (5.74): For $\mathbf{e}_\pi = (\mathbf{e}_{\pi,\mathbf{H}}, \mathbf{e}_{\pi,\mathbf{E}})$ and $\varphi_h = (\phi_h, \psi_h)$ we have

$$\left(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, \varphi_h\right)_{\mu\times\varepsilon} = \left(\boldsymbol{\mathcal{C}}_\mathbf{E}\mathbf{e}_{\pi,\mathbf{E}}, \phi_h\right)_\mu + \left(\boldsymbol{\mathcal{C}}_\mathbf{H}\mathbf{e}_{\pi,\mathbf{H}}, \psi_h\right)_\varepsilon. \tag{5.78}$$

By Definition 3.1 of $\boldsymbol{\mathcal{C}}_\mathbf{E}$ and since the projection error $\mathbf{e}_{\pi,\mathbf{E}}$ is orthogonal on V_h (cf. Definition 2.16), we deduce that

$$\begin{aligned}
\left(\boldsymbol{\mathcal{C}}_\mathbf{E}\mathbf{e}_{\pi,\mathbf{E}}, \phi_h\right)_\mu &= \sum_{K\in\mathcal{T}_h}\left(\operatorname{curl}\phi_h, \mathbf{e}_{\pi,\mathbf{E}}\right)_K + \sum_{F\in\mathcal{F}_h^{\mathrm{int}}}\left(n_F\times[\![\phi_h]\!]_F, \{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\right)_F \\
&= \sum_{F\in\mathcal{F}_h^{\mathrm{int}}}\left(n_F\times[\![\phi_h]\!]_F, \{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\right)_F \\
&\leq \sum_{F\in\mathcal{F}_{h,c}^{\mathrm{int}}}\|n_F\times[\![\phi_h]\!]_F\|_F\ \|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F + \sum_{F\in\mathcal{F}_{h,i}^{\mathrm{int}}}\|n_F\times[\![\phi_h]\!]_F\|_F\ \|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F.
\end{aligned} \tag{5.79}$$

Here, we used the splitting of the mesh faces $\mathcal{F}_h^{\mathrm{int}} = \mathcal{F}_{h,c}^{\mathrm{int}} \dot\cup \mathcal{F}_{h,i}^{\mathrm{int}}$ from Definition 5.2 and the Cauchy–Schwarz inequality. By Definition 2.9 of the weighted averages we have

$$\begin{aligned}
\|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F^2 &= a_F^2\,\|\varepsilon_K c_K\mathbf{e}_{\pi,\mathbf{E}}|_K + \varepsilon_{K_F}c_{K_F}\mathbf{e}_{\pi,\mathbf{E}}|_{K_F}\|_F^2 \\
&\leq 2a_F^2\left(\|\varepsilon_K c_K\mathbf{e}_{\pi,\mathbf{E}}|_K\|_F^2 + \|\varepsilon_{K_F}c_{K_F}\mathbf{e}_{\pi,\mathbf{E}}|_{K_F}\|_F^2\right) \\
&= 2a_F^2\left(\varepsilon_K c_K^2\|\mathbf{e}_{\pi,\mathbf{E}}|_K\|_{\varepsilon,F}^2 + \varepsilon_{K_F}c_{K_F}^2\|\mathbf{e}_{\pi,\mathbf{E}}|_{K_F}\|_{\varepsilon,F}^2\right) \\
&\leq 2a_F\left(c_K\|\mathbf{e}_{\pi,\mathbf{E}}|_K\|_{\varepsilon,F}^2 + c_{K_F}\|\mathbf{e}_{\pi,\mathbf{E}}|_{K_F}\|_{\varepsilon,F}^2\right) \\
&\leq 2a_F\,\widehat{C}_{\mathrm{app}}^2\left(c_K h_K^{2k+1}|\mathbf{E}|_{k+1,K}^2 + c_{K_F}h_{K_F}^{2k+1}|\mathbf{E}|_{k+1,K_F}^2\right).
\end{aligned} \tag{5.80}$$

Here, we applied the triangle inequality, Young's inequality, and (5.35).

From now, the two sums in (5.79) have to be treated differently.

(b) By the Cauchy–Schwarz inequality (A.3) in $\mathbb{R}^{\mathrm{card}(\mathcal{F}_{h,c}^{\mathrm{int}})}$ with weight a_F, we obtain

$$\begin{aligned}
&\sum_{F\in\mathcal{F}_{h,c}^{\mathrm{int}}} \|n_F\times[\![\phi_h]\!]_F\|_F\,\|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F \\
&\qquad\le \Big(\sum_{F\in\mathcal{F}_{h,c}^{\mathrm{int}}} a_F\|n_F\times[\![\phi_h]\!]_F\|_F^2\Big)^{1/2}\Big(\sum_{F\in\mathcal{F}_{h,c}^{\mathrm{int}}} a_F^{-1}\|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F^2\Big)^{1/2} \\
&\qquad\le 2^{1/2}\widehat{C}_{\mathrm{app}}|\phi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e}\Big(\sum_{F\in\mathcal{F}_{h,c}^{\mathrm{int}}} c_K h_K^{2k+1}|\mathbf{E}|_{k+1,K}^2 + c_{K_F}h_{K_F}^{2k+1}|\mathbf{E}|_{k+1,K_F}^2\Big)^{1/2} \\
&\qquad\le (2N_\partial c_{\infty,c})^{1/2}\widehat{C}_{\mathrm{app}}|\phi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e}|\mathbf{E}|_{k+1,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,k+\frac12} \\
&\qquad= C_{\pi,c}|\phi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e}|\mathbf{E}|_{k+1,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,k+\frac12}. \qquad (5.81)
\end{aligned}$$

For the second inequality we used the Definition 5.17 of the stabilization seminorm and (5.80).

(c) Again the Cauchy–Schwarz inequality (A.3) in $\mathbb{R}^{\mathrm{card}(\mathcal{F}_{h,i}^{\mathrm{int}})}$ implies

$$\begin{aligned}
&\sum_{F\in\mathcal{F}_{h,i}^{\mathrm{int}}} \|n_F\times[\![\phi_h]\!]_F\|_F\,\|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F \\
&\qquad\le \Big(\sum_{F\in\mathcal{F}_{h,i}^{\mathrm{int}}} \omega_F\|n_F\times[\![\phi_h]\!]_F\|_F^2\Big)^{1/2}\Big(\sum_{F\in\mathcal{F}_{h,i}^{\mathrm{int}}} \omega_F^{-1}\|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F^2\Big)^{1/2},
\end{aligned}$$

with a weight $\omega_F = a_F(h_K+h_{K_F})/2$ as in (3.36). Note that $\mathcal{F}_{h,i}^{\mathrm{int}}$ (also) contains faces bordering mesh elements from the fine set $\mathcal{T}_{h,f}$. Thus, in this part of the proof we need the shape- and contact-regularity of the whole mesh $\mathcal{T}_h$, i.e. (2.4). In fact, we can now use Part (c) of the proof of Theorem 3.10 where we proved the bounds (3.38) and (3.40). This yields

$$\sum_{F\in\mathcal{F}_{h,i}^{\mathrm{int}}} \|n_F\times[\![\phi_h]\!]_F\|_F\,\|\{\!\{\mathbf{e}_{\pi,\mathbf{E}}\}\!\}_F^{\varepsilon c}\|_F \le \widehat{C}_\pi\|\phi_h\|_{\mu,\mathcal{T}_{h,i}}|\mathbf{E}|_{k+1,\mathcal{T}_{h,i},2,k}. \qquad (5.82)$$

Inserting (5.81) and (5.82) into (5.79), we finally obtain

$$\big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{e}_{\pi,\mathbf{E}},\phi_h\big)_\mu \le C_{\pi,c}|\phi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e}|\mathbf{E}|_{k+1,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,k+\frac12} + \widehat{C}_\pi\|\phi_h\|_{\mu,\mathcal{T}_{h,i}}|\mathbf{E}|_{k+1,\mathcal{T}_{h,i},2,k}.$$

Analog computations show

$$\big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}\mathbf{e}_{\pi,\mathbf{H}},\psi_h\big)_\varepsilon \le C_{\pi,c}|\psi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{E}}^e}|\mathbf{H}|_{k+1,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,k+\frac12} + \widehat{C}_\pi\|\psi_h\|_{\varepsilon,\mathcal{T}_{h,i}}|\mathbf{H}|_{k+1,\mathcal{T}_{h,i},2,k},$$

whence the asserted bound (5.74) is obtained by (5.78) and the Cauchy–Schwarz inequality in $\mathbb{R}^2$.

(d) We proceed with proving the bound (5.75): By Definition 5.14, the Cauchy–Schwarz inequalities in $L^2(F)$ and in $\mathbb{R}^{\mathrm{card}(\mathcal{F}_{h,c}^{\mathrm{int}})}$ we have

$$\begin{aligned}
\big(\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e\mathbf{e}_{\pi,\mathbf{H}},\phi_h\big)_\mu &\le \sum_{F\in\mathcal{F}_{h,c}^{\mathrm{int}}} a_F\|n_F\times[\![\mathbf{e}_{\pi,\mathbf{H}}]\!]_F\|_F\,\|n_F\times[\![\phi_h]\!]_F\|_F \\
&\le |\phi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e}\Big(\sum_{F\in\mathcal{F}_{h,c}^{\mathrm{int}}} a_F\|n_F\times[\![\mathbf{e}_{\pi,\mathbf{H}}]\!]_F\|_F^2\Big)^{1/2}.
\end{aligned}$$

Using (3.41) we have

$$\big(\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e\mathbf{e}_{\pi,\mathbf{H}},\phi_h\big) \le C_{\pi,c}|\phi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}}|\mathbf{H}|_{k+1,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,k+\frac12},$$

and analogously we obtain

$$\big(\boldsymbol{\mathcal{S}}_{\mathbf{E}}^e\mathbf{e}_{\pi,\mathbf{E}},\psi_h\big) \le C_{\pi,c}|\psi_h|_{\boldsymbol{\mathcal{S}}_{\mathbf{E}}}|\mathbf{E}|_{k+1,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,k+\frac12}.$$

Finally, by the Cauchy–Schwarz inequality in $\mathbb{R}^2$ we get the desired bound (5.75). □

5.4.3 Interlude: The semidiscrete problem with explicit stabilization

We briefly consider the spatial semi-discretization of Maxwell's equations with the explicit stabilization operator, i.e.,

$$\begin{aligned} \partial_t \mathbf{u}_h(t) &= \boldsymbol{\mathcal{C}}\mathbf{u}_h(t) - \alpha\boldsymbol{\mathcal{S}}^e\mathbf{u}_h(t) + \mathbf{j}_h(t), \\ \mathbf{u}_h(0) &= \mathbf{u}_h^0, \end{aligned} \tag{5.83}$$

since the analysis of this semidiscrete problem gives insight into how the fully discrete analysis has to be carried out. For the stability analysis we can use the ideas for the fully stabilized dG discretization presented Section 3.3.

Theorem 5.21. *We have the following stability result for the solution* $\mathbf{u}_h$ *of* (5.83)*:*

(a) *For* $\mathbf{J}_h \in C(0,T;V_h)$ *we have*

$$\|\mathbf{u}_h(t)\|_{\mu\times\varepsilon}^2 + 2\alpha\int_0^t |\mathbf{u}_h(s)|_{\boldsymbol{\mathcal{S}}^e}^2\, ds \le \mathrm{e}^1\|\mathbf{u}^0\|_{\mu\times\varepsilon}^2 + \mathrm{e}^1\frac{T+1}{\delta}\int_0^t \|\mathbf{J}(s)\|^2\, ds, \tag{5.84a}$$

where δ *was defined in* (1.20).

(b) *For* $\mathbf{J}_h \equiv 0$ *we have*

$$\|\mathbf{u}_h(t)\|_{\mu\times\varepsilon}^2 + 2\alpha\int_0^t |\mathbf{u}_h(s)|_{\boldsymbol{\mathcal{S}}^e}^2\, ds = \|\pi_h\mathbf{u}^0\|_{\mu\times\varepsilon} \le \|\mathbf{u}^0\|_{\mu\times\varepsilon}. \tag{5.84b}$$

Proof. The statement can be proved exactly as Theorem 3.8. □

Note that by (5.84b) the explicitly stabilized upwind dG discretization is (as the fully stabilized upwind fluxes discretization) dissipative, but only with respect to the explicit stabilization seminorm. In fact, we have

$$\mathcal{E}(\mathbf{H}_h(t),\mathbf{E}_h(t)) = \mathcal{E}(\mathbf{H}_h^0,\mathbf{E}_h^0) - \alpha\int_0^t |\mathbf{u}_h(s)|_{\boldsymbol{\mathcal{S}}^e}^2\, ds, \qquad t \ge 0.$$

As in the fully stabilized case, the stability parameter $\alpha \in [0,1]$ controls the amount of dissipation. For the error analysis it turns out that (both in the semidiscrete and in the fully discrete case) we both need techniques applied in the central fluxes analysis and techniques used for the fully stabilized upwind fluxes analysis. (Roughly speaking we need the central fluxes techniques on the implicit part $\mathcal{T}_{h,i}$ and the upwind fluxes technique on the explicit part $\mathcal{T}_{h,e}$). This was already done in Theorem 5.19 which combines these two worlds. First, we give an error representation.

Lemma 5.22. *Let* $\alpha \in [0,1]$*. Then, the error* $\mathbf{e}_h = \mathbf{u}_h - \pi_h\mathbf{u}$ *of* (5.83) *satisfies*

$$\partial_t\mathbf{e}_h = \boldsymbol{\mathcal{C}}\mathbf{e}_h - \alpha\boldsymbol{\mathcal{S}}^e\mathbf{e}_h + \mathbf{d}_\pi^e, \qquad \mathbf{e}_h(0) = 0, \tag{5.85a}$$

with defect

$$\mathbf{d}_\pi^e = -\boldsymbol{\mathcal{C}}\mathbf{e}_\pi + \alpha\boldsymbol{\mathcal{S}}^e\mathbf{e}_\pi. \tag{5.85b}$$

Proof. The proof of Lemma 3.9 can be transferred from the full stabilization $\boldsymbol{\mathcal{S}}$ to the current case of the explicit stabilization $\boldsymbol{\mathcal{S}}^e$, since $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{S}}^e$ share the same properties. □

We end this interlude with the convergence result for the semidiscrete problem (5.83).

Theorem 5.23. *Let* $\mathbf{u} \in C^1\big(0,T;L^2(\Omega)^6\big) \cap C\big(0,T;D(\mathcal{C}) \cap H^{k+1}(\mathcal{T}_h)^6\big)$ *be the solution of Maxwell's equations* (5.1) *and let* $\mathbf{u}_h \in C^1\big(0,T;V_h^2\big)$ *denote the semidiscrete approximation obtained from the (explicitly stabilized) upwind fluxes dG discretization* (5.83)*. Then, the error* $\mathbf{e} = \mathbf{u} - \mathbf{u}_h$ *satisfies*

$$\begin{aligned}\|\mathbf{e}(t)\|_{\mu\times\varepsilon}^2 + \alpha\int_0^t |\mathbf{e}_h(s)|_{\boldsymbol{\mathcal{S}}^e}^2\, ds \leq & C_{\text{app}}^2 |\mathbf{u}(t)|_{k+1,\mathcal{T}_h,2,k+1}^2 \\ & + \mathrm{e}^1 \widetilde{C}_{\text{upw},c} \int_0^t |\mathbf{u}(s)|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2\, ds \\ & + \mathrm{e}^1 \widehat{C}_\pi^2 (T+1) \int_0^t |\mathbf{u}(s)|_{k+1,\mathcal{T}_{h,i},2,k}^2\, ds \\ \leq & C\Big(\max_{K\in\mathcal{T}_{h,e}} h_K^{2k+1} + (T+1) \max_{K\in\mathcal{T}_{h,i}} h_K^{2k} \Big). \end{aligned} \tag{5.86}$$

where $\widetilde{C}_{\text{upw},c}$ *is given by* $\widetilde{C}_{\text{upw},c} = 2C_{\pi,c}^2(1+\alpha^2)/\alpha$*. Moreover, the constant* C *only depends on* C_{app}*,* $\widetilde{C}_{\text{upw},c}$*,* $C_{\pi,c}$*, and* $|\mathbf{u}(s)|_{k+1,\mathcal{T}_h}$*,* $s \in [0,t]$*.*

Similar to the full upwind case, the constant $\widetilde{C}_{\text{upw},c}$ depends on the stabilization parameter $\alpha \in (0,1]$. For $\alpha = 1$ we obtain the smallest constant and for $\alpha \searrow 0$ the bound (5.86) deteriorates.

Proof. We use an energy technique to prove this result, i.e. we take the $\mu\times\varepsilon$-inner product of (5.85a) with $\mathbf{e}_h(t)$. Then, by (5.85b), the skew-symmetry of the discrete Maxwell operator $\boldsymbol{\mathcal{C}}$, see (3.7b), and the property (5.71) of explicit stabilization seminorm, we obtain

$$\frac{1}{2}\frac{d}{dt}\|\mathbf{e}_h(t)\|_{\mu\times\varepsilon}^2 = -\alpha|\mathbf{e}_h(t)|_{\boldsymbol{\mathcal{S}}^e}^2 + \big(\mathbf{d}_\pi^e(t), \mathbf{e}_h(t)\big)_{\mu\times\varepsilon}.$$

Integrating from 0 to t and using $\mathbf{e}_h(0) = 0$ yields

$$\|\mathbf{e}_h(t)\|_{\mu\times\varepsilon}^2 + 2\alpha\int_0^t |\mathbf{e}_h(s)|_{\boldsymbol{\mathcal{S}}^e}^2\, ds = 2\int_0^t \big(\mathbf{d}_\pi^e(s), \mathbf{e}_h(s)\big)_{\mu\times\varepsilon}\, ds. \tag{5.87}$$

Using Young's inequality in the bounds obtained in Theorem 5.19 we conclude that for arbitrary $\gamma_1, \gamma_2 > 0$ it holds

$$\big(\boldsymbol{\mathcal{C}}\mathbf{e}_\pi, \varphi_h\big)_{\mu\times\varepsilon} \leq \gamma_1|\varphi_h|_{\boldsymbol{\mathcal{S}}^e}^2 + \gamma_2\|\varphi_h\|_{\mu\times\varepsilon,\mathcal{T}_{h,i}}^2 + \frac{C_{\pi,c}^2}{4\gamma_1}|\mathbf{u}|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2 + \frac{\widehat{C}_\pi^2}{4\gamma_2}|\mathbf{u}|_{k+1,\mathcal{T}_{h,i},2,k}^2, \tag{5.88}$$

and

$$\big(\alpha\boldsymbol{\mathcal{S}}^e\mathbf{e}_\pi, \varphi_h\big)_{\mu\times\varepsilon} \leq \gamma_1\alpha^2|\varphi_h|_{\boldsymbol{\mathcal{S}}^e}^2 + \frac{C_{\pi,c}^2}{4\gamma_1}|\mathbf{u}|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2. \tag{5.89}$$

Because we have $\mathbf{e}_h \in V_h^2$, we can apply bounds (5.88) and (5.89) to $\big(\mathbf{d}_\pi^e, \mathbf{e}_h\big)_{\mu\times\varepsilon}$ in (5.87) which imply

$$\begin{aligned} 2\big(\mathbf{d}_\pi^e, \mathbf{e}_h\big)_{\mu\times\varepsilon} &\leq 2(1+\alpha^2)\gamma_1|\mathbf{e}_h|_{\boldsymbol{\mathcal{S}}^e}^2 + 2\gamma_2\|\mathbf{e}_h\|_{\mu\times\varepsilon,\mathcal{T}_{h,i}}^2 + \frac{C_{\pi,c}^2}{\gamma_1}|\mathbf{u}|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2 + \frac{\widehat{C}_\pi^2}{2\gamma_2}|\mathbf{u}|_{k+1,\mathcal{T}_{h,i},2,k}^2 \\ &\leq \alpha|\mathbf{e}_h|_{\boldsymbol{\mathcal{S}}^e}^2 + 2\gamma_2\|\mathbf{e}_h\|_{\mu\times\varepsilon}^2 + \widetilde{C}_{\text{upw},c}|\mathbf{u}|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2 + \frac{\widehat{C}_\pi^2}{2\gamma_2}|\mathbf{u}|_{k+1,\mathcal{T}_{h,i},2,k}^2, \end{aligned}$$

where in the second inequality we chose $\gamma_1 = \alpha/\big(2(1+\alpha^2)\big)$ and used $\|\mathbf{e}_h\|_{\mu\times\varepsilon,\mathcal{T}_{h,i}} \le \|\mathbf{e}_h\|_{\mu\times\varepsilon}$. Inserting this bound into (5.87) we conclude

$$\|\mathbf{e}_h(t)\|^2_{\mu\times\varepsilon} + \alpha\int_0^t |\mathbf{e}_h(s)|^2_{\boldsymbol{\mathcal{S}}^e}\,ds \le 2\gamma_2\int_0^t \|\mathbf{e}_h(s)\|^2_{\mu\times\varepsilon}\,ds + \int_0^t \Big(\widetilde{C}_{\mathrm{upw},c}\,|\mathbf{u}(s)|^2_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}} + \frac{\widehat{C}_\pi^2}{2\gamma_2}\,|\mathbf{u}(s)|^2_{k+1,\mathcal{T}_{h,i},2,k}\Big)\,ds.$$

Next, we apply the continuous Gronwall lemma (Lemma A.1), which results in

$$\|\mathbf{e}_h(t)\|^2_{\mu\times\varepsilon} + \alpha\int_0^t |\mathbf{e}_h(s)|^2_{\boldsymbol{\mathcal{S}}^e}\,ds \le \mathrm{e}^{2\gamma_2 t}\int_0^t \Big(\widetilde{C}_{\mathrm{upw},c}\,|\mathbf{u}(s)|^2_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}} + \frac{\widehat{C}_\pi^2}{2\gamma_2}\,|\mathbf{u}(s)|^2_{k+1,\mathcal{T}_{h,i},2,k}\Big)\,ds.$$

Finally, choosing the weight $\gamma_2 = (T+1)/2$ and using $\|\mathbf{e}\|^2_{\mu\times\varepsilon} = \|\mathbf{e}_\pi\|^2_{\mu\times\varepsilon} + \|\mathbf{e}_h\|^2_{\mu\times\varepsilon}$, together with (3.24a) yields the desired statement. □

Note that in contrast to the convergence proof in the fully stabilized case (cf. Theorem 3.13), this proof requires a Gronwall lemma.

5.4.4 Analysis of the locally implicit method: Stability and energy dissipation

Our first step in the analysis of our locally implicit upwind method (5.65) consists in casting it into a compact form with the operators $\widetilde{\boldsymbol{\mathcal{R}}}_L$ and $\widetilde{\boldsymbol{\mathcal{R}}}_R$ of the locally implicit central fluxes scheme.

Lemma 5.24. *The locally implicit scheme (5.65) is equivalent to*

$$\widetilde{\boldsymbol{\mathcal{R}}}_L\mathbf{u}_h^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}}_R\mathbf{u}_h^n - \tau\alpha\boldsymbol{\mathcal{S}}^e\mathbf{u}_h^n + \frac{\tau}{2}\big(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n\big), \tag{5.90}$$

where the operators $\widetilde{\boldsymbol{\mathcal{R}}}_L$ and $\widetilde{\boldsymbol{\mathcal{R}}}_R$ were defined in (5.13b).

Proof. Adding (5.65a) and (5.65c) yields

$$\mathbf{H}_h^{n+1} - \mathbf{H}_h^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{E}_h^{n+1} + \mathbf{E}_h^n\big) - \tau\alpha\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\mathbf{H}_h^n,$$

which is the first component of (5.90). For the second component we subtract (5.65c) from (5.65a):

$$\mathbf{H}_h^{n+1/2} = \frac{1}{2}\big(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n\big) + \frac{\tau}{4}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{E}_h^{n+1} - \mathbf{E}_h^n\big).$$

Inserting this into (5.65b) we infer

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) + \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{E}_h^{n+1} - \mathbf{E}_h^n) - \tau\alpha\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}\mathbf{E}_h^n - \frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n),$$

by using $\boldsymbol{\mathcal{C}}^e_{\mathbf{H}} + \boldsymbol{\mathcal{C}}^i_{\mathbf{H}} = \boldsymbol{\mathcal{C}}_{\mathbf{H}}$, see (5.12). □

Next, we give an energy identity.

Lemma 5.25. *The approximation $\mathbf{u}_h^n = (\mathbf{H}_h^n, \mathbf{E}_h^n)$ obtained from the locally implicit method (5.90) satisfies*

$$\begin{aligned}&\|\mathbf{u}_h^{n+1}\|^2_{\mu\times\varepsilon} - \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\mathbf{E}_h^{n+1}\|^2_\mu - \alpha\frac{\tau}{2}|\mathbf{u}_h^{n+1}|^2_{\boldsymbol{\mathcal{S}}^e} + \alpha\frac{\tau}{2}\sum_{m=0}^n |\mathbf{u}_h^{m+1} + \mathbf{u}_h^m|^2_{\boldsymbol{\mathcal{S}}^e}\\ &= \|\mathbf{u}_h^0\|^2_{\mu\times\varepsilon} - \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\mathbf{E}_h^0\|^2_\mu - \alpha\frac{\tau}{2}|\mathbf{u}_h^0|^2_{\boldsymbol{\mathcal{S}}^e} + \frac{\tau}{2}\sum_{m=0}^n \big(\mathbf{j}_h^{m+1} + \mathbf{j}_h^m, \mathbf{u}_h^{m+1} + \mathbf{u}_h^m\big)_{\mu\times\varepsilon}.\end{aligned} \tag{5.91}$$

Proof. Analogous to the proof of Lemma 4.23. □

This lemma implies that the upwind fluxes locally implicit scheme is dissipative w.r.t a perturbed electromagnetic energy.

Corollary 5.26. *For vanishing source term* $\mathbf{J}_h \equiv 0$*, the approximation* $\mathbf{u}_h^n = (\mathbf{H}_h^n, \mathbf{E}_h^n)$ *of the locally implicit method* (5.90) *satisfies*

$$\widetilde{\mathcal{E}}_{\mathrm{upw}}(\mathbf{H}_h^n, \mathbf{E}_h^n) = \widetilde{\mathcal{E}}_{\mathrm{upw}}(\mathbf{H}_h^0, \mathbf{E}_h^0) - \alpha\frac{\tau}{4}\sum_{m=0}^{n-1} |\mathbf{u}_h^{m+1} + \mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}^e}^2, \qquad n = 1, 2, \dots, \tag{5.92}$$

where the perturbed electromagnetic energy $\widetilde{\mathcal{E}}_{\mathrm{upw}}$ *is defined as*

$$\widetilde{\mathcal{E}}_{\mathrm{upw}}(\mathbf{H}_h, \mathbf{E}_h) = \widetilde{\mathcal{E}}(\mathbf{H}_h, \mathbf{E}_h) - \alpha\frac{\tau}{4}|\mathbf{u}_h|_{\boldsymbol{\mathcal{S}}^e}^2.$$

Next, we address the stability of our locally implicit scheme. As for the upwind fluxes Verlet method we will need a tightened CFL condition compared to the central fluxes case since the stabilization enters the CFL condition (compare the two CFL conditions (4.49) and (4.86) for the central fluxes and the upwind fluxes Verlet method, respectively). The upwind fluxes Verlet method involves the full stabilization operators and thus their contribution in the CFL condition comprises all mesh elements in the spatial grid. In contrary, we constructed the upwind fluxes locally implicit method with the explicit stabilization operators and as a consequence we obtain a CFL condition involving only the coarse grid elements. In particular, the **CFL condition for the upwind fluxes locally implicit scheme** reads,

$$\tau \leq \frac{2\widetilde{\theta}}{C_{\mathrm{bnd},c}c_{\infty,c}} \min_{K\in\mathcal{T}_{h,c}} h_K, \tag{5.93a}$$

with a fixed parameter $0 < \widetilde{\theta} < 1$ which satisfies

$$\widetilde{\theta}_{\mathrm{upw}} := \widetilde{\theta}^2 + \alpha\widetilde{\theta} < 1. \tag{5.93b}$$

Note that the CFL condition depends on the stabilization parameter α. For larger α we obtain a method with a stricter CFL condition.

Corollary 5.27. *Assume that the CFL condition* (5.93) *is satisfied. Then, the approximation* $\mathbf{u}_h^n$ *obtained from the upwind fluxes dG discretization and the locally implicit method* (5.90) *is bounded by*

$$\begin{aligned}(1-\widetilde{\theta}_{\mathrm{upw}})\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n}|\mathbf{u}_h^{m+1}+\mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}^e}^2 \\ \leq \mathrm{e}^{3/2}\left(\|\mathbf{u}^0\|_{\mu\times\varepsilon}^2 + \frac{T+1}{\delta(1-\widetilde{\theta}_{\mathrm{upw}})}\frac{\tau}{4}\sum_{m=0}^{n}\|\mathbf{J}^{m+1}+\mathbf{J}^m\|^2\right),\end{aligned} \tag{5.94}$$

for $n = 1, 2, \dots, N_T$.

Observe that the bound deteriorates for $\widetilde{\theta}_{\mathrm{upw}} \nearrow 1$.

Proof. We apply the Cauchy–Schwarz inequality and the weighted Young's inequality with weight $\gamma > 0$ to (5.91),

$$\begin{aligned}\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n}|\mathbf{u}_h^{m+1}+\mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}^e}^2 \le \|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e\mathbf{E}_h^{n+1}\|_\mu^2 + \alpha\frac{\tau}{2}|\mathbf{u}_h^{n+1}|_{\boldsymbol{\mathcal{S}}^e}^2 \\ + \frac{\tau}{4\gamma}\sum_{m=0}^{n}\frac{1}{2\gamma}\|\mathbf{j}_h^{m+1}+\mathbf{j}_h^m\|_{\mu\times\varepsilon}^2 \\ + \gamma\frac{\tau}{2}\sum_{m=0}^{n}\big(\|\mathbf{u}_h^{m+1}\|_{\mu\times\varepsilon}^2 + \|\mathbf{u}_h^m\|_{\mu\times\varepsilon}^2\big). \qquad (5.95)\end{aligned}$$

By the boundedness results for $\boldsymbol{\mathcal{C}}^e$ and $|\cdot|_{\boldsymbol{\mathcal{S}}^e}$ obtained in Theorems 5.6, 5.18 in combination with the CFL condition (5.93), we infer

$$\frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e\mathbf{E}_h^{n+1}\|_\mu^2 + \alpha\frac{\tau}{2}|\mathbf{u}_h^{n+1}|_{\boldsymbol{\mathcal{S}}^e}^2 \le \widetilde{\theta}_{\mathrm{upw}}\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2, \qquad (5.96)$$

since $\widehat{C}_{\mathrm{bnd},c} \le C_{\mathrm{bnd},c}$. Inserting the last inequality in (5.95) shows

$$\begin{aligned}(1-\widetilde{\theta}_{\mathrm{upw}})\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n}|\mathbf{u}_h^{m+1}+\mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}^e}^2 \le \|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{\tau}{4\gamma}\sum_{m=0}^{n}\|\mathbf{j}_h^{m+1}+\mathbf{j}_h^m\|_{\mu\times\varepsilon}^2 \\ + \gamma\frac{\tau}{2}\sum_{m=0}^{n}\left(\|\mathbf{u}_h^{m+1}\|_{\mu\times\varepsilon}^2 + \|\mathbf{u}_h^m\|_{\mu\times\varepsilon}^2\right).\end{aligned}$$

We choose the weight $\gamma = (1-\widetilde{\theta}_{\mathrm{upw}})/(T+1)$, so that the discrete Gronwall lemma is applicable (Lemma A.2 with $\lambda = 1/(T+1)$). This yields

$$\begin{aligned}(1-\widetilde{\theta}_{\mathrm{upw}})\|\mathbf{u}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n}|\mathbf{u}_h^{m+1}+\mathbf{u}_h^m|_{\boldsymbol{\mathcal{S}}^e}^2 \\ \le \exp\left(\frac{3}{2}\frac{t_{n+1}}{T+1}\right)\left(\|\mathbf{u}_h^0\|_{\mu\times\varepsilon}^2 + \frac{T+1}{1-\widetilde{\theta}_{\mathrm{upw}}}\frac{\tau}{4}\sum_{m=0}^{n}\|\mathbf{j}_h^{m+1}+\mathbf{j}_h^m\|_{\mu\times\varepsilon}^2\right),\end{aligned}$$

which finishes the proof. □

5.4.5 Error analysis of the locally implicit method

As in Chapter 4 we split the full discretization error into $\mathbf{e}^n = \mathbf{e}_\pi^n - \mathbf{e}_h^n$, where the error $\mathbf{e}_h^n$ satisfies the recursion of the locally implicit method (5.90) but with defects instead of the source term $\mathbf{j}_h$. The next lemma gives the details.

Lemma 5.28. *Let $\mathbf{u} \in C\big(0,T;V_\star\big) \cap C^3\big(0,T;L^2(\Omega)^6\big)$ be the exact solution of (5.1). Then, the error $\mathbf{e}_h^n$ of the locally implicit scheme (5.90) satisfies*

$$\widetilde{\boldsymbol{\mathcal{R}}}_L\mathbf{e}_h^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}}_R\mathbf{e}_h^n - \tau\alpha\boldsymbol{\mathcal{S}}^e\mathbf{e}_h^n + \widetilde{\mathbf{d}}_{\mathrm{upw}}^n, \qquad (5.97a)$$

where the defect $\widetilde{\mathbf{d}}_{\mathrm{upw}}^n = \widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^n + \widetilde{\mathbf{d}}_h^n$ is given by

$$\widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^n = \widetilde{\mathbf{d}}_\pi^n + \tau\alpha\boldsymbol{\mathcal{S}}^e\mathbf{e}_\pi^n, \qquad (5.97b)$$

and where $\widetilde{\mathbf{d}}_\pi^n$ and $\widetilde{\mathbf{d}}_h^n$ were defined in (5.50).

Proof. First, we observe that the locally implicit method (5.90) can be written as

$$\mathbf{u}_h^{n+1} - \mathbf{u}_h^n = \frac{\tau}{2}\boldsymbol{\mathcal{C}}\big(\mathbf{u}_h^{n+1} + \mathbf{u}_h^n\big) - \tau\alpha\boldsymbol{\mathcal{S}}^e\mathbf{u}_h^n + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) + \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{E}_h^{n+1} - \mathbf{E}_h^n)\end{pmatrix}.$$

Next, we insert the projected exact solution into this form of the locally implicit scheme,

$$\begin{aligned}\pi_h(\mathbf{u}^{n+1} - \mathbf{u}^n) =& \frac{\tau}{2}\boldsymbol{\mathcal{C}}\pi_h\big(\mathbf{u}^{n+1} + \mathbf{u}^n\big) - \tau\alpha\boldsymbol{\mathcal{S}}^e\pi_h\mathbf{u}^n \\ &+ \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) + \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e\boldsymbol{\mathcal{C}}_{\mathbf{E}}\pi_h(\mathbf{E}^{n+1} - \mathbf{E}^n)\end{pmatrix} - \widetilde{\mathbf{d}}_{\mathrm{upw}}^n.\end{aligned}$$

Subtracting these two equations yields (5.97a). It remains to determine the defect $\widetilde{\mathbf{d}}_{\mathrm{upw}}^n$. By (4.61) and the consistency of the explicit stabilization operator (5.67) the exact solution $\mathbf{u}$ satisfies

$$\pi_h(\mathbf{u}^{n+1} - \mathbf{u}^n) = \frac{\tau}{2}\boldsymbol{\mathcal{C}}(\mathbf{u}^{n+1} + \mathbf{u}^n) - \tau\alpha\boldsymbol{\mathcal{S}}^e\mathbf{u}^n + \frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n) + \tau^2\pi_h\delta^n(\partial_t\mathbf{u}).$$

Subtracting the last two equations yields

$$\widetilde{\mathbf{d}}_{\mathrm{upw}}^n = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}(\mathbf{e}_\pi^{n+1} + \mathbf{e}_\pi^n) + \tau\alpha\boldsymbol{\mathcal{S}}^e\mathbf{e}_\pi^n - \tau^2\pi_h\delta^n(\partial_t\mathbf{u}) + \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e\boldsymbol{\mathcal{C}}_{\mathbf{E}}\pi_h(\mathbf{E}^{n+1} - \mathbf{E}^n)\end{pmatrix}.$$

Finally, using

$$\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e\boldsymbol{\mathcal{C}}_{\mathbf{E}}\pi_h(\mathbf{E}^{n+1} - \mathbf{E}^n) = -\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e\pi_h\Delta_{\mathbf{H}}^n - \boldsymbol{\mathcal{C}}_{\mathbf{H}}^e\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n),$$

see the proof of Lemma 4.19, gives the desired expression for the defect $\widetilde{\mathbf{d}}_{\mathrm{upw}}^n$. □

By (5.50) the defect $\widetilde{\mathbf{d}}_h^n$ consists of two parts, where the first one does not cause problems, but the second part cannot be bounded straightforwardly, since it suffers from an order reduction from τ^3 to $\tau^{2.5}$, see Lemma 5.12. As a cure we presented in (5.57) the idea of splitting the defect $\widetilde{\mathbf{d}}_h^n$ into

$$\widetilde{\mathbf{d}}_h^n = \mathbf{d}_h^n + (\widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R)\widetilde{\boldsymbol{\xi}}^n, \qquad \widetilde{\boldsymbol{\xi}}^n = \frac{\tau}{4}\begin{pmatrix}\chi_e\pi_h\Delta_{\mathbf{H}}^n \\ 0\end{pmatrix}. \tag{5.98}$$

In the following we will use this idea, but we cannot follow the further steps from the central fluxes case. In contrary to the central fluxes case we apply an energy technique in order to obtain the improved spatial convergence order $k+1/2$ in the spatial variable as in the semidiscrete case, see Section 3.4.2, and as in the fully discrete case with the Crank–Nicolson time integration, see Section 4.3.2. However, it turns out that even the energy technique applied directly to (5.97), (5.98) fails to give the desired temporal convergence order. The essential idea – besides the energy technique – is to consider a modified error $\widetilde{\mathbf{e}}_h^n$ instead of $\mathbf{e}_h^n$. A related idea has been presented in Verwer [2011]. In the following lemma we introduce this modified error and give the associated error recursion.

Lemma 5.29. *Let $\mathbf{u} \in C\big(0,T;V_\star\big) \cap C^3\big(0,T;L^2(\Omega)^6\big)$ be the exact solution of (5.1) and let $\mathbf{e}_h^n$ be the error of the locally implicit scheme (5.90). Then, the modified error*

$$\widetilde{\mathbf{e}}_h^n = \mathbf{e}_h^n - \widetilde{\boldsymbol{\xi}}^{n-1},\ n \geq 1, \qquad \widetilde{\mathbf{e}}_h^0 = \mathbf{e}_h^0 = 0, \tag{5.99a}$$

satisfies

$$\widetilde{\boldsymbol{\mathcal{R}}}_L\widetilde{\mathbf{e}}_h^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}}_R\widetilde{\mathbf{e}}_h^n - \tau\alpha\boldsymbol{\mathcal{S}}^e\widetilde{\mathbf{e}}_h^n + \widetilde{\widetilde{\mathbf{d}}}_{\mathrm{upw}}^n, \qquad n \geq 0, \tag{5.99b}$$

with defect

$$\widetilde{\widetilde{\mathbf{d}}}_{\mathrm{upw}}^n = \begin{cases} \widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^0 + \mathbf{d}_h^0 - \widetilde{\boldsymbol{\mathcal{R}}}_R\widetilde{\boldsymbol{\xi}}^0, & n = 0, \\ \widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^n + \mathbf{d}_h^n - \widetilde{\boldsymbol{\mathcal{R}}}_R(\widetilde{\boldsymbol{\xi}}^n - \widetilde{\boldsymbol{\xi}}^{n-1}) - \tau\alpha\boldsymbol{\mathcal{S}}^e\widetilde{\boldsymbol{\xi}}^{n-1}, & n \geq 1. \end{cases} \tag{5.99c}$$

An important observation is that by (5.5) we have $\widetilde{\boldsymbol{\xi}}^n \in V_h^2$ and thus $\widetilde{\mathbf{e}}_h^n \in V_h^2$.

In (5.60) we have seen that the difference $\widetilde{\boldsymbol{\xi}}^n - \widetilde{\boldsymbol{\xi}}^{n-1}$ allows us to gain an order τ which avoided the order reduction in the temporal convergence order in the central fluxes case. The same should hold true in the upwind fluxes case, which motivates the modified error recursion.

Proof. We employ the splitting (5.98) of $\widetilde{\mathbf{d}}_h^n$ in (5.97a), which yields

$$\begin{aligned}
\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{e}_h^{n+1} &= \widetilde{\boldsymbol{\mathcal{R}}}_R \mathbf{e}_h^n - \tau\alpha \boldsymbol{\mathcal{S}}^e \mathbf{e}_h^n + \widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^n + \mathbf{d}_h^n + (\widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R)\widetilde{\boldsymbol{\xi}}^n \\
&= \widetilde{\boldsymbol{\mathcal{R}}}_R \mathbf{e}_h^n - \tau\alpha \boldsymbol{\mathcal{S}}^e \mathbf{e}_h^n + \widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^n + \mathbf{d}_h^n + \widetilde{\boldsymbol{\mathcal{R}}}_L \widetilde{\boldsymbol{\xi}}^n - \widetilde{\boldsymbol{\mathcal{R}}}_R(\widetilde{\boldsymbol{\xi}}^n - \widetilde{\boldsymbol{\xi}}^{n-1}) - \widetilde{\boldsymbol{\mathcal{R}}}_R \widetilde{\boldsymbol{\xi}}^{n-1}, \qquad n \geq 1, \\
\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{e}_h^1 &= \widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^0 + \mathbf{d}_h^0 + (\widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R)\widetilde{\boldsymbol{\xi}}^0,
\end{aligned}$$

since $\mathbf{e}_h^0 = 0$. This is equivalent to

$$\begin{aligned}
\widetilde{\boldsymbol{\mathcal{R}}}_L(\mathbf{e}_h^{n+1} - \widetilde{\boldsymbol{\xi}}^n) &= \widetilde{\boldsymbol{\mathcal{R}}}_R(\mathbf{e}_h^n - \widetilde{\boldsymbol{\xi}}^{n-1}) - \tau\alpha \boldsymbol{\mathcal{S}}^e \mathbf{e}_h^n + \widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^n + \mathbf{d}_h^n - \widetilde{\boldsymbol{\mathcal{R}}}_R(\widetilde{\boldsymbol{\xi}}^n - \widetilde{\boldsymbol{\xi}}^{n-1}), \qquad n \geq 1, \\
\widetilde{\boldsymbol{\mathcal{R}}}_L(\mathbf{e}_h^1 - \widetilde{\boldsymbol{\xi}}^0) &= \widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^0 + \mathbf{d}_h^0 - \widetilde{\boldsymbol{\mathcal{R}}}_R \widetilde{\boldsymbol{\xi}}^0,
\end{aligned}$$

which ends the proof. □

The error $\widetilde{\mathbf{e}}_h^n$ satisfies the recursion (5.90) of the locally implicit scheme with defect $\widetilde{\widetilde{\mathbf{d}}}_{\mathrm{upw}}^n$ instead of the source terms $\frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n)$. Hence, we can apply Lemma 5.25 and obtain

$$\begin{aligned}
\|\widetilde{\mathbf{e}}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n} |\widetilde{\mathbf{e}}_h^{m+1} + \widetilde{\mathbf{e}}_h^m|_{\boldsymbol{\mathcal{S}}^e}^2 &= \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{E}}^e \widetilde{\mathbf{e}}_{h,\mathbf{E}}^{n+1}\|_\mu^2 + \alpha\frac{\tau}{2}|\widetilde{\mathbf{e}}_h^{n+1}|_{\boldsymbol{\mathcal{S}}^e}^2 \\
&\quad + \big(\widetilde{\widetilde{\mathbf{d}}}_{\mathrm{upw}}^0, \widetilde{\mathbf{e}}_h^1\big)_{\mu\times\varepsilon} + \sum_{m=1}^{n} \big(\widetilde{\widetilde{\mathbf{d}}}_{\mathrm{upw}}^m, \widetilde{\mathbf{e}}_h^{m+1} + \widetilde{\mathbf{e}}_h^m\big)_{\mu\times\varepsilon},
\end{aligned}$$

where we used $\widetilde{\mathbf{e}}_h^0 = 0$. By the boundedness results for $\boldsymbol{\mathcal{C}}^e$ and $\boldsymbol{\mathcal{S}}^e$ obtained in Theorems 5.6 and 5.18 in combination with the CFL condition (5.93), we infer

$$\begin{aligned}
(1 - \widetilde{\theta}_{\mathrm{upw}})\|\widetilde{\mathbf{e}}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{2}\sum_{m=0}^{n} |\widetilde{\mathbf{e}}_h^{m+1} + \widetilde{\mathbf{e}}_h^m|_{\boldsymbol{\mathcal{S}}^e}^2 &\leq \big(\widetilde{\widetilde{\mathbf{d}}}_{\mathrm{upw}}^0, \widetilde{\mathbf{e}}_h^1\big)_{\mu\times\varepsilon} \\
&\quad + \sum_{m=1}^{n} \big(\widetilde{\widetilde{\mathbf{d}}}_{\mathrm{upw}}^m, \widetilde{\mathbf{e}}_h^{m+1} + \widetilde{\mathbf{e}}_h^m\big)_{\mu\times\varepsilon}, \qquad (5.100)
\end{aligned}$$

see also (5.96). So, we have to bound the defects in the form $\big(\widetilde{\widetilde{\mathbf{d}}}_{\mathrm{upw}}^m, \varphi_h\big)_{\mu\times\varepsilon}$. For the sake of readability we give these bounds with respect to a generic constant C, which depends on $C_{\pi,c}$, $\widehat{C}_\pi$, C_{ctr}, $\widehat{C}_{\mathrm{app}}$, $C_{\mathrm{bnd},c}$ and $c_{\infty,c}$, but is independent of τ, h_K and α. Moreover, we introduce two weights $\gamma_1, \gamma_2 > 0$ which we will choose in our main theorem (Theorem 5.35).

We start with the projection error $\widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^n$.

Lemma 5.30. *Assume that the exact solution of Maxwell's equations satisfies* $\mathbf{u} = (\mathbf{H}, \mathbf{E}) \in C\big(0,T; H^{k+1}(\mathcal{T}_h)^6\big)$ *and* $\mathbf{E} \in C^1\big(0,T; H^{k+1}(\mathcal{T}_{h,c})^3\big)$ *and moreover assume that the CFL condition (5.93) is satisfied, Then, for all* $\varphi_h \in V_h^2$ *we have the bound*

$$\begin{aligned}
\big(\widetilde{\mathbf{d}}_{\pi,\mathrm{upw}}^n, \varphi_h\big)_{\mu\times\varepsilon} &\leq (1+\alpha^2)\gamma_1\tau|\varphi_h|_{\boldsymbol{\mathcal{S}}^e}^2 + 2\gamma_2\tau\|\varphi_h\|_{\mu\times\varepsilon}^2 \\
&\quad + \frac{C}{\gamma_1}\tau\Big(\big|\mathbf{u}^{n+1} + \mathbf{u}^n\big|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2 + |\mathbf{u}^n|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2\Big) \\
&\quad + \frac{C}{\gamma^2}\tau\Big(\big|\mathbf{u}^{n+1} + \mathbf{u}^n\big|_{k+1,\mathcal{T}_{h,i},2,k}^2 + \int_{t_n}^{t_{n+1}} |\partial_t \mathbf{E}(t)|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2 \, dt\Big).
\end{aligned}$$

Proof. We recall that by (4.58), (5.50) and (5.97b) the defect $\widetilde{\mathbf{d}}^n_{\pi,\mathrm{upw}}$ is given by

$$\widetilde{\mathbf{d}}^n_{\pi,\mathrm{upw}} = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}\big(\mathbf{e}^{n+1}_{\pi}+\mathbf{e}^n_{\pi}\big)+\tau\alpha\boldsymbol{\mathcal{S}}^e\mathbf{e}^n_{\pi}-\frac{\tau^2}{4}\begin{pmatrix}0\\ \boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{e}^{n+1}_{\pi,\mathbf{E}}-\mathbf{e}^n_{\pi,\mathbf{E}}\big)\end{pmatrix}.$$

The first two terms can be bounded by (5.88) and (5.89). For the third term of $\widetilde{\mathbf{d}}^n_{\pi,\mathrm{upw}}$ we use the Cauchy–Schwarz inequality and the weighted Young's inequality to obtain

$$\begin{aligned}\frac{\tau^2}{4}\left(\begin{pmatrix}0\\ \boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{e}^{n+1}_{\pi,\mathbf{E}}-\mathbf{e}^n_{\pi,\mathbf{E}}\big)\end{pmatrix},\varphi_h\right)_{\mu\times\varepsilon} &\le \gamma_2\tau\|\varphi_h\|^2_{\mu\times\varepsilon}+\frac{\tau^3}{64\gamma_2}\|\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{e}^{n+1}_{\pi,\mathbf{E}}-\mathbf{e}^n_{\pi,\mathbf{E}}\big)\|^2_{\varepsilon}\\ &\le \gamma_2\tau\|\varphi_h\|^2_{\mu\times\varepsilon}+\frac{\tau}{16\gamma_2}\|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\big(\mathbf{e}^{n+1}_{\pi,\mathbf{E}}-\mathbf{e}^n_{\pi,\mathbf{E}}\big)\|^2_{\mu}. \qquad (5.101)\end{aligned}$$

Here, we used $\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}=\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}$, the boundedness result for $\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}$ from Theorem 5.6, and the CFL condition (5.93) for the second inequality. So, we need to bound a term of the type $\|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\mathbf{e}_{\pi,\mathbf{E}}\|_\mu$. Therefore, note that by (5.74) (cf. the proof of the associated theorem), we have

$$\big(\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\mathbf{e}_{\pi,\mathbf{E}},\phi_h\big)_\mu=\big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{e}_{\pi,\mathbf{E}},\chi_e\phi_h\big)_\mu\le C_{\pi,c}|\chi_e\phi_h|_{\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}}\,|\mathbf{E}|_{k+1,\mathcal{T}_{h,c},2,k+\frac12},$$

since $\|\chi_e\phi_h\|_{\mu,\mathcal{T}_{h,i}}=0$. Next, we use the boundedness of $|\cdot|_{\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}}$, see (5.72) and the associated proof, to obtain

$$\begin{aligned}\tau^{1/2}\big(\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\mathbf{e}_{\pi,\mathbf{E}},\phi_h\big)_\mu &\le \tau^{1/2}C_{\pi,c}(\widehat{C}_{\mathrm{bnd},c}c_{\infty,c})^{1/2}\|\chi_e\phi_h\|_{\mu,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,-1/2}\,|\mathbf{E}|_{k+1,\mathcal{T}_{h,c},2,k+\frac12}\\ &\le \sqrt{2}C_{\pi,c}\|\phi_h\|_\mu\,|\mathbf{E}|_{k+1,\mathcal{T}_{h,c},2,k+\frac12}.\end{aligned}$$

The second inequality follows with the CFL condition (5.93) and the fact that $\widehat{C}_{\mathrm{bnd},c}/C_{\mathrm{bnd},c}\le 1$. This yields

$$\tau^{1/2}\|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\mathbf{e}_{\pi,\mathbf{E}}\|_\mu\le\sqrt{2}C_{\pi,c}\,|\mathbf{E}|_{k+1,\mathcal{T}_{h,c},2,k+\frac12},$$

and we can conclude

$$\frac{\tau}{16\gamma_2}\|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\big(\mathbf{e}^{n+1}_{\pi,\mathbf{E}}-\mathbf{e}^n_{\pi,\mathbf{E}}\big)\|^2_\mu\le\frac{C^2_{\pi,c}}{8\gamma_2}\,\big|\mathbf{E}^{n+1}-\mathbf{E}^n\big|^2_{k+1,\mathcal{T}_{h,c},2,k+\frac12}.$$

Inserting this bound into (5.101) and using that

$$\begin{aligned}\big|\mathbf{E}^{n+1}-\mathbf{E}^n\big|^2_{k+1,\mathcal{T}_{h,c},2,k+\frac12}&=\left|\int_{t_n}^{t_{n+1}}\partial_t\mathbf{E}(t)\,dt\right|^2_{k+1,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,k+\frac12}\\ &\le\tau\int_{t_n}^{t_{n+1}}|\partial_t\mathbf{E}(t)|^2_{k+1,\mathcal{T}_{h,c},2,k+\frac12}\,dt,\end{aligned}$$

finishes the proof. □

Remark 5.31. In comparison to the central fluxes locally implicit method we need an additional regularity assumption, namely that $\mathbf{E}\in C^1\big(0,T;H^{k+1}(\mathcal{T}_{h,c})^3\big)$. The reason lies in the bound (5.101). It is possible to change this bound to

$$\begin{aligned}\frac{\tau^2}{4}\left(\begin{pmatrix}0\\ \boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{e}^{n+1}_{\pi,\mathbf{E}}-\mathbf{e}^n_{\pi,\mathbf{E}}\big)\end{pmatrix},\varphi_h\right)_{\mu\times\varepsilon}&=\frac{\tau^2}{4}\big(\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\big(\mathbf{e}^{n+1}_{\pi,\mathbf{E}}-\mathbf{e}^n_{\pi,\mathbf{E}}\big),\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\psi_h\big)_\mu\\ &\le\frac{\tau^2}{4}C_{\pi,c}|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}\psi_h|_{\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}}|\mathbf{E}^{n+1}-\mathbf{E}^n|_{k+1,\mathcal{T}_{h,c},2,k+\frac12},\end{aligned}$$

which yields the right convergence order $k+1/2$ in the spatial variable without a regularity assumption on $\partial_t\mathbf{E}$. However, this bound implies that a term $|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}(\mathbf{e}^{n+1}_h+\mathbf{e}^n_h)|_{\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}}$ enters the right-hand side of our error recursion and our locally implicit scheme does not provide a dissipative term of this form to cancel it. Thus, this ansatz is not usable.

Next, we address the defect $\widetilde{\boldsymbol{\mathcal{R}}}_R(\widetilde{\boldsymbol{\xi}}^n - \widetilde{\boldsymbol{\xi}}^{n-1})$.

Lemma 5.32. *Let the exact solution of Maxwell's equations satisfy $\mathbf{H} \in C^3\big(0,T;L^2(\Omega)^3\big)$ and moreover assume that the CFL condition (5.93) holds. Then, for all $\varphi_h \in V_h^2$ and all $n \geq 1$, we have that*

$$\big(\widetilde{\boldsymbol{\mathcal{R}}}_R(\widetilde{\boldsymbol{\xi}}^n - \widetilde{\boldsymbol{\xi}}^{n-1}), \varphi_h\big)_{\mu\times\varepsilon} \leq \gamma_2\tau\|\varphi_h\|_{\mu\times\varepsilon}^2 + \frac{C}{\gamma_2}\tau^4 \int_{t_{n-1}}^{t_{n+1}} \|\partial_t^3 \mathbf{H}(t)\|_{\mu,\mathcal{T}_{h,c}}^2 \, dt.$$

Proof. Comparing (5.98) with (4.70a) we see that

$$\widetilde{\boldsymbol{\xi}}^n = \chi_e \widehat{\boldsymbol{\xi}}^n = \frac{\tau}{4}\begin{pmatrix} \chi_e \pi_h \Delta_{\mathbf{H}}^n \\ 0 \end{pmatrix}. \tag{5.102}$$

For general functions $\mathbf{H}_h \in V_h$ and $\varphi_h = (\phi_h, \psi_h) \in V_h^2$ we have by the definitions of $\widetilde{\boldsymbol{\mathcal{R}}}_L$ in (5.13b) and of $\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e$ in (5.16a) that

$$\begin{aligned}
\left(\widetilde{\boldsymbol{\mathcal{R}}}_R\begin{pmatrix} \chi_e \mathbf{H}_h \\ 0 \end{pmatrix}, \varphi_h\right)_{\mu\times\varepsilon} &= \big(\chi_e \mathbf{H}_h, \phi_h\big)_\mu + \frac{\tau}{2}\big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \mathbf{H}_h, \psi_h\big)_\varepsilon \\
&\leq \gamma_2\tau\|\varphi_h\|_{\mu\times\varepsilon}^2 + \frac{1}{4\gamma_2\tau}\Big(\|\mathbf{H}_h\|_{\mu,\mathcal{T}_{h,e}}^2 + \frac{\tau^2}{4}\|\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \mathbf{H}_h\|_\varepsilon^2\Big) \\
&\leq \gamma_2\tau\|\varphi_h\|_{\mu\times\varepsilon}^2 + \frac{1}{2\gamma_2\tau}\|\mathbf{H}_h\|_{\mu,\mathcal{T}_{h,c}}^2.
\end{aligned} \tag{5.103}$$

Here, the first inequality is obtained by the Cauchy–Schwarz inequality and the weighted Young's inequality, and the second inequality follows from the boundedness result for $\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e$, i.e. (5.25b), and the CFL condition (5.93). Using this, we have

$$\begin{aligned}
\big(\widetilde{\boldsymbol{\mathcal{R}}}_R(\widetilde{\boldsymbol{\xi}}^n - \widetilde{\boldsymbol{\xi}}^{n-1}), \varphi_h\big)_{\mu\times\varepsilon} &= \left(\widetilde{\boldsymbol{\mathcal{R}}}_R\begin{pmatrix} \chi_e(\widehat{\xi}_{\mathbf{H}}^n - \widehat{\xi}_{\mathbf{H}}^{n-1}) \\ 0 \end{pmatrix}, \varphi_h\right)_{\mu\times\varepsilon} \\
&\leq \gamma_2\tau\|\varphi_h\|_{\mu\times\varepsilon}^2 + \frac{1}{2\gamma_2\tau}\|\widehat{\xi}_{\mathbf{H}}^n - \widehat{\xi}_{\mathbf{H}}^{n-1}\|_\mu^2 \\
&\leq \gamma_2\tau\|\varphi_h\|_{\mu\times\varepsilon}^2 + \frac{\tau^4}{32\gamma_2}\int_{t_{n-1}}^{t_{n+1}} \|\partial_t^3 \mathbf{H}(t)\|_{\mu,\mathcal{T}_{h,c}}^2 \, dt.
\end{aligned}$$

Here, the second inequality follows via (4.74) and the proof is complete. □

In the subsequent lemma we provide a bound on $\tau\alpha\boldsymbol{\mathcal{S}}^e\widetilde{\boldsymbol{\xi}}^{n-1}$.

Lemma 5.33. *Let $\mathbf{H} \in C^2\big(0,T;H^{\max(1,k-1)}(\mathcal{T}_{h,e})^3\big)$ and assume that the CFL condition (5.93) is satisfied. Then, for all $\varphi_h \in V_h^2$ and all $n \geq 1$, we have that*

$$\begin{aligned}
\big(\tau\alpha\boldsymbol{\mathcal{S}}^e\widetilde{\boldsymbol{\xi}}^{n-1}, \varphi_h\big)_{\mu\times\varepsilon} \leq{}& \gamma_1\alpha^2\tau|\varphi_h|_{\boldsymbol{\mathcal{S}}}^2 \\
&+ \frac{C}{\gamma_1}\int_{t_{n-1}}^{t_n}\Big(\tau^4\|\mu\partial_t^2\mathbf{H}(t)\|_{1,\mathcal{T}_{h,e}}^2 + |\partial_t^2\mathbf{H}(t)|_{\max(1,k-1),\mathcal{T}_{h,e},2,k+\frac{1}{2}}^2\Big)\, dt.
\end{aligned}$$

Proof. By using (5.102), the Cauchy–Schwarz inequality and Young's inequality we obtain

$$\tau\alpha\big(\boldsymbol{\mathcal{S}}^e\widetilde{\boldsymbol{\xi}}^{n-1}, \varphi_h\big)_{\mu\times\varepsilon} \leq \tau\alpha|\widetilde{\boldsymbol{\xi}}^{n-1}|_{\boldsymbol{\mathcal{S}}^e}|\varphi_h|_{\boldsymbol{\mathcal{S}}^e} \leq \gamma_1\alpha^2\tau|\varphi_h|_{\boldsymbol{\mathcal{S}}^e}^2 + \frac{\tau^3}{64\gamma_1}|\chi_e\pi_h\Delta_{\mathbf{H}}^{n-1}|_{\boldsymbol{\mathcal{S}}_{\mathbf{H}}^e}^2.$$

In the second term we decompose $\pi_h\Delta_{\mathbf{H}}^{n-1} = \Delta_{\mathbf{H}}^{n-1} - \Delta_\pi^{n-1}$, where

$$\Delta_{\mathbf{H}}^{n-1} = \int_{t_{n-1}}^{t_n} \partial_t^2\mathbf{H}(t)\, dt, \qquad \Delta_\pi^{n-1} = \int_{t_{n-1}}^{t_n} \partial_t^2\mathbf{e}_{\pi,\mathbf{H}}(t)\, dt\,,$$

see part (c) of the proof of Lemma 4.20. By the definition of $|\cdot|_{\mathcal{S}^e_{\mathbf{H}}}$, see (5.70a), we have

$$\begin{aligned}\frac{\tau^3}{64\gamma_1}|\chi_e\pi_h\Delta^{n-1}_{\mathbf{H}}|^2_{\mathcal{S}^e_{\mathbf{H}}} &= \frac{\tau^3}{64\gamma_1}\sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} a_F\|n_F\times[\![\chi_e\pi_h\Delta^{n-1}_{\mathbf{H}}]\!]_F\|^2_F\\ &\le \frac{\tau^3}{32\gamma_1}\sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,c}} a_F\big(\|[\![\chi_e\Delta^{n-1}_{\mathbf{H}}]\!]_F\|^2_F + \|[\![\chi_e\Delta^{n-1}_{\pi}]\!]_F\|^2_F\big).\end{aligned} \tag{5.104}$$

Here, the inequality is obtained via the splitting of $\pi_h\Delta^{n-1}_{\mathbf{H}}$, the triangle inequality, Young's inequality and $|n_F|=1$. We bound the two terms separately. For the subsequent calculations it is important to recall that the set $\mathcal{F}^{\mathrm{int}}_{h,c}$ only contains faces bordering coarse elements. So, for the remaining proof let $F\in\mathcal{F}^{\mathrm{int}}_{h,c}$, which yields $K, K_F\in\mathcal{T}_{h,c}$.

(a) For the first term the Cauchy–Schwarz inequality in L^2 yields

$$\begin{aligned}a_F\|[\![\chi_e\Delta^{n-1}_{\mathbf{H}}]\!]_F\|^2_F &\le a_F\tau\int_{t_{n-1}}^{t_n}\|[\![\chi_e\partial_t^2\mathbf{H}(t)]\!]_F\|^2_F\,dt\\ &\le 2C^2_{\mathrm{ctr}}a_F\tau\int_{t_{n-1}}^{t_n}\mu_K^{-1}\|\chi_e(\mu\partial_t^2\mathbf{H}(t))\|^2_{1,K}+\mu_{K_F}^{-1}\|\chi_e(\mu\partial_t^2\mathbf{H}(t))\|^2_{1,K_F}\,dt\\ &\le 2C^2_{\mathrm{ctr}}c_{\infty,c}\tau\int_{t_{n-1}}^{t_n}\|\chi_e(\mu\partial_t^2\mathbf{H}(t))\|^2_{1,K\cup K_F}\,dt.\end{aligned}$$

Here, we used the triangle inequality, Young's inequality and the continuous trace inequality (2.13) for the second inequality, and (5.32) for the third inequality.

(b) For the second term we have

$$\begin{aligned}a_F\|[\![\chi_e\Delta^{n-1}_{\pi}]\!]_F\|^2_F &\le a_F\tau\int_{t_{n-1}}^{t_n}\|[\![\chi_e\partial_t^2\mathbf{e}_{\pi,\mathbf{H}}(t)]\!]_F\|^2_F\,dt\\ &\le 2c_{\infty,c}\tau\int_{t_{n-1}}^{t_n}\|\chi_e\partial_t^2\mathbf{e}_{\pi,\mathbf{H}}(t)|_K\|^2_{\mu,F}+\|\chi_e\partial_t^2\mathbf{e}_{\pi,\mathbf{H}}(t)|_{K_F}\|^2_{\mu,F}\,dt,\end{aligned}$$

where the second inequality is obtained via the triangle inequality, Young's inequality and (5.32). Let $\widetilde{k}=\max(1,k-1)$, then the regularity assumptions on $\partial_t^2\mathbf{H}$ together with (3.24b) imply

$$\begin{aligned}\|\partial_t^2\mathbf{e}_{\pi,\mathbf{H}}|_K\|^2_{\mu,F}\le\widehat{C}^2_{\mathrm{app}}h_K^{2\widetilde{k}-1}|\partial_t^2\mathbf{H}|^2_{\widetilde{k},K} &= \widehat{C}^2_{\mathrm{app}}\tau^{-4}\tau^4h_K^{-4}h_K^{2\widetilde{k}+3}|\partial_t^2\mathbf{H}|^2_{\widetilde{k},K}\\ &\le\frac{16\widehat{C}^2_{\mathrm{app}}}{C^4_{\mathrm{bnd},c}c^4_{\infty,c}}\tau^{-4}h_K^{2\widetilde{k}+3}|\partial_t^2\mathbf{H}|^2_{\widetilde{k},K}.\end{aligned}$$

For the last inequality we used the CFL condition (5.93). Hence, we end up with

$$a_F\|[\![\chi_e\Delta^{n-1}_{\pi}]\!]_F\|^2_F\le 32\frac{\widehat{C}^2_{\mathrm{app}}}{C^4_{\mathrm{bnd},c}c^3_{\infty,c}}\tau^{-3}\int_{t_{n-1}}^{t_n}h_K^{2k+1}|\chi_e\partial_t^2\mathbf{H}(t)|^2_{\widetilde{k},K}+h_{K_F}^{2k+1}|\chi_e\partial_t^2\mathbf{H}(t)|^2_{\widetilde{k},K_F}\,dt.$$

where we used $h_K^{2\widetilde{k}+3}\le h_K^{2k+1}$. (This holds true with in the case $k>1$ and in the case $k=1$ for $h_K\le 1$, i.e. the relevant case for a convergence proof. If $h_K>1$, an additionally constant $|\Omega|^2_d$ enters this bound.)

(c) Inserting the results from (a) and (b) in (5.104), we infer

$$\begin{aligned}\frac{\tau^3}{64\gamma_1}|\chi_e\pi_h\Delta^{n-1}_{\mathbf{H}}|^2_{\mathcal{S}^e_{\mathbf{H}}} &\le \frac{C^2_{\mathrm{ctr}}c_{\infty,c}}{16\gamma_1}\tau^4\int_{t_{n-1}}^{t_n}\|\mu\partial_t^2\mathbf{H}(t)\|^2_{1,\mathcal{T}_{h,e}}\,dt\\ &\quad+\frac{\widehat{C}^2_{\mathrm{app}}}{C^4_{\mathrm{bnd},c}c^3_{\infty,c}\gamma_1}\int_{t_{n-1}}^{t_n}|\partial_t^2\mathbf{H}(t)|_{\max(1,k-1),\mathcal{T}_{h,e},2,k+\frac{1}{2}}\,dt,\end{aligned}$$

which is the desired bound. □

It remains to establish a bound on $\widetilde{\widetilde{\mathbf{d}}}^0_{\mathrm{upw}}$. It will be essential that this defect appears as inner product with the local error $\widetilde{\mathbf{e}}^1_h$.

Lemma 5.34. *Let the exact solution* $\mathbf{u} = (\mathbf{H}, \mathbf{E})$ *of Maxwell's equations satisfy*

$$\mathbf{u} \in C^3(0,T;L^2(\Omega)^6) \cap C\big(0,T;H^{k+1}(\mathcal{T}_h)^6\big), \qquad \mathbf{E} \in C^1\big(0,T;H^{k+1}(\mathcal{T}_{h,c})^3\big),$$

and moreover assume that the CFL condition (5.93) *holds. Then, the following bound holds true*

$$\begin{aligned}\big(\widetilde{\widetilde{\mathbf{d}}}^0_{\mathrm{upw}}, \widetilde{\mathbf{e}}^1_h\big)_{\mu\times\varepsilon} \leq & C|\mathbf{u}^1 + \mathbf{u}^0|^2_{k+1,\mathcal{T}_h,2,k+1} \\ & + C\tau^4 \max_{t\in[0,\tau]} \|\partial_t^2 \mathbf{H}(t)\|^2_{\mu,\mathcal{T}_{h,c}} + C\tau^4 \int_0^\tau \|\partial_t^3 \mathbf{u}(t)\|^2_{\mu\times\varepsilon}\, dt \\ & + C\tau \left|\mathbf{u}^0\right|^2_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}} + C\tau \left|\mathbf{u}^1 - \mathbf{u}^0\right|^2_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}.\end{aligned}$$

Proof. By (5.99b), $\widetilde{\mathbf{e}}^0_h = 0$ and subsequently (5.18b) we have

$$\big(\widetilde{\widetilde{\mathbf{d}}}^0_{\mathrm{upw}}, \widetilde{\mathbf{e}}^1_h\big)_{\mu\times\varepsilon} = \big(\widetilde{\boldsymbol{\mathcal{R}}}_L \widetilde{\mathbf{e}}^1_h, \widetilde{\mathbf{e}}^1_h\big)_{\mu\times\varepsilon} \leq \|\widetilde{\mathbf{e}}^1_h\|^2_{\mu\times\varepsilon}.$$

Under the CFL condition the operator $\widetilde{\boldsymbol{\mathcal{R}}}_L$ is invertible and thus we obtain from (5.99b)

$$\widetilde{\mathbf{e}}^1_h = \widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}(\widetilde{\mathbf{d}}^0_{\pi,\mathrm{upw}} + \mathbf{d}^0_h) + \widetilde{\boldsymbol{\mathcal{R}}}\widetilde{\boldsymbol{\xi}}^0.$$

Observe that by (5.97b), (5.50) and (4.58) we can write the projection defect as

$$\begin{aligned}\widetilde{\mathbf{d}}^0_{\pi,\mathrm{upw}} &= -\frac{\tau}{2}\boldsymbol{\mathcal{C}}(\mathbf{e}^1_\pi + \mathbf{e}^0_\pi) + \tau\alpha\boldsymbol{\mathcal{S}}^e\mathbf{e}^0_\pi - \frac{\tau^2}{4}\begin{pmatrix}0\\ \boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}(\mathbf{e}^1_{\pi,\mathbf{E}} - \mathbf{e}^0_{\pi,\mathbf{E}})\end{pmatrix} \\ &= \frac{1}{2}(\widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R)(\mathbf{e}^1_\pi + \mathbf{e}^0_\pi) + \tau\alpha\boldsymbol{\mathcal{S}}^e\mathbf{e}^0_\pi + \frac{\tau}{4}(\widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R)\begin{pmatrix}\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}(\mathbf{e}^1_{\pi,\mathbf{E}} - \mathbf{e}^0_{\pi,\mathbf{E}})\\ 0\end{pmatrix}.\end{aligned}$$

Here, the second equality follows by $\widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R = -\frac{\tau}{2}\boldsymbol{\mathcal{C}}$ and (5.56). Using Lemma 5.8 we infer

$$\begin{aligned}\|\widetilde{\mathbf{e}}^1_h\|_{\mu\times\varepsilon} \leq & \|\widetilde{\boldsymbol{\mathcal{R}}}\|_{\mu\times\varepsilon}\|\widetilde{\boldsymbol{\xi}}^0\|_{\mu\times\varepsilon} \\ & + \|\widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}\|_{\mu\times\varepsilon}\Big(\tau\|\boldsymbol{\mathcal{S}}^e\mathbf{e}^0_\pi\|_{\mu\times\varepsilon} + \|\mathbf{d}^0_h\|_{\mu\times\varepsilon}\Big) \\ & + \|\boldsymbol{\mathcal{I}} - \widetilde{\boldsymbol{\mathcal{R}}}\|_{\mu\times\varepsilon}\Big(\frac{1}{2}\|\mathbf{e}^1_\pi + \mathbf{e}^0_\pi\|_{\mu\times\varepsilon} + \frac{\tau}{4}\|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}(\mathbf{e}^1_{\pi,\mathbf{E}} - \mathbf{e}^0_{\pi,\mathbf{E}})\|_\mu\Big) \\ \leq & \, C^{1/2}_{\mathrm{stb},c}\|\widetilde{\boldsymbol{\xi}}^0\|_{\mu\times\varepsilon} \\ & + C_{\mathrm{stb},c}\Big(\tau\|\boldsymbol{\mathcal{S}}^e\mathbf{e}^0_\pi\|_{\mu\times\varepsilon} + \|\mathbf{d}^0_h\|_{\mu\times\varepsilon}\Big) \\ & + (1 + C^{1/2}_{\mathrm{stb},c})\Big(\frac{1}{2}\|\mathbf{e}^1_\pi + \mathbf{e}^0_\pi\|_{\mu\times\varepsilon} + \frac{\tau}{4}\|\boldsymbol{\mathcal{C}}^e_{\mathbf{E}}(\mathbf{e}^1_{\pi,\mathbf{E}} - \mathbf{e}^0_{\pi,\mathbf{E}})\|_\mu\Big).\end{aligned}$$

The first term can be bounded with (5.61), the third term with (4.68a), the fourth term with (3.24a) and the last one as in the proof of Lemma 5.30. For the remaining defect we use (5.75), (5.72) and the CFL condition

$$\begin{aligned}\tau\big(\boldsymbol{\mathcal{S}}^e\mathbf{e}_\pi, \varphi_h\big)_{\mu\times\varepsilon} &\leq C_{\pi,c}\tau|\varphi_h|_{\boldsymbol{\mathcal{S}}^e}|u|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}} \\ &\leq C_{\pi,c}(\widehat{C}_{\mathrm{bnd},c}c_{\infty,c})^{1/2}\tau\|\varphi_h\|_{\mu\times\varepsilon,\mathcal{T}_{h,e}\cup\mathcal{T}_{h,ci},2,-1/2}|u|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}} \\ &\leq \sqrt{2}C_{\pi,c}\tau^{1/2}\|\varphi_h\|_{\mu\times\varepsilon}|u|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}},\end{aligned}$$

see also the proof of Lemma 5.30. As a consequence, we have

$$\tau\|\boldsymbol{\mathcal{S}}^e\mathbf{e}_\pi^0\|_{\mu\times\varepsilon} \le \sqrt{2}C_{\pi,c}\tau^{1/2}|\mathbf{u}^0|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}.$$

This concludes the proof. □

Now, we have all ingredients to prove the **convergence result for the full discretization** of the **upwind fluxes locally implicit method**.

Theorem 5.35. *Assume that the exact solution* $\mathbf{u} = (\mathbf{H}, \mathbf{E})$ *of Maxwell's equations* (5.1) *satisfies*

$$\mathbf{u} \in C\big(0,T; D(\mathcal{C}) \cap H^{k+1}(\mathcal{T}_h)^6\big) \cap C^3\big(0,T; L^2(\Omega)^6\big),$$

and

$$\mathbf{E} \in C^1\big(0,T; H^{k+1}(\mathcal{T}_{h,c})^3\big), \qquad \mathbf{H} \in C^2\big(0,T; H^{\max(1,k-1)}(\mathcal{T}_{h,e})^3\big).$$

Moreover, assume that the CFL condition (5.93) *holds true with* $\widetilde{\theta}_{\mathrm{upw}} \in (0,1)$*, and assume that* $n\tau \le T$*. Then, the error of the upwind fluxes dG discretization and the locally implicit scheme* (5.65) *satisfies*

$$\begin{aligned}
&\|\mathbf{u}^n - \mathbf{u}_h^n\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{4}\sum_{m=0}^{n-1}|\widetilde{\mathbf{e}}_h^{m+1} + \widetilde{\mathbf{e}}_h^m|_{\boldsymbol{\mathcal{S}}^e}^2 \\
&\qquad \le C\big(|\mathbf{u}^n|_{k+1,\mathcal{T}_h,1,k+1}^2 + |\mathbf{u}^1 + \mathbf{u}^0|_{k+1,\mathcal{T}_h,1,k+1}^2 + \tau^4 \max_{t\in[0,t_n]}\|\partial_t^2\mathbf{H}(t)\|_{\mu,\mathcal{T}_{h,e}}^2\big) \\
&\qquad\quad + C\tau^4\int_0^{t_n}\big(\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}^2 + \|\mu\partial_t^2\mathbf{H}(t)\|_{1,\mathcal{T}_{h,e}}^2\big)\,dt \\
&\qquad\quad + C\tau\sum_{m=0}^{n}\Big(|\mathbf{u}^m|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2 + |\mathbf{u}^m|_{k+1,\mathcal{T}_{h,i},2,k}^2\Big) \\
&\qquad\quad + C\int_0^{t_n}\Big(|\partial_t\mathbf{E}(t)|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2 + |\partial_t^2\mathbf{H}(t)|_{\max(1,k-1),\mathcal{T}_{h,e},2,k+\frac{1}{2}}^2\Big)\,dt \\
&\qquad \le C\Big(\max_{K\in\mathcal{T}_{h,e}} h_K^{2k+1} + \max_{K\in\mathcal{T}_{h,i}} h_K^{2k} + \tau^4\Big).
\end{aligned}$$

The constant C *only depends on* C_{app}, $C_{\pi,c}$, $\widehat{C}_\pi$, C_{ctr}, $\widehat{C}_{\mathrm{app}}$, $C_{\mathrm{bnd},c}$, $1/(1-\widetilde{\theta}_{\mathrm{upw}})$, *and from* T, $(1+\alpha^2)/\alpha$, *and moreover from* $|\mathbf{u}(t)|_{k+1,\mathcal{T}_h}$, $|\partial_t\mathbf{E}(t)|_{k+1,\mathcal{T}_{h,c}}$, $|\partial_t^2\mathbf{H}(t)|_{\max(1,k-1),\mathcal{T}_{h,c}}$ *and* $\|\partial_t^2\mathbf{H}(t)\|_\mu$, $\|\partial_t^3\mathbf{u}(t)\|_{\mu\times\varepsilon}$, $t\in[0,t_n]$.

Remark 5.36. We recall Remark 5.20: In dG methods we have the freedom to choose the polynomial degree differently on every mesh element. Thus, by choosing degree $k+1$ on the (very few) elements in $\mathcal{T}_{h,c}\cap\mathcal{T}_{h,i}$, we obtain the convergence rate

$$\|\mathbf{u}^n - \mathbf{u}_h^n\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{4}\sum_{m=0}^{n-1}|\widetilde{\mathbf{e}}_h^{m+1} + \widetilde{\mathbf{e}}_h^m|_{\boldsymbol{\mathcal{S}}^e}^2 \le C\Big(\max_{K\in\mathcal{T}_{h,c}} h_K^{2k+1} + \max_{K\in\mathcal{T}_{h,f}} h_K^{2k} + \tau^4\Big).$$

This is the desired rate $k+1/2$ on the coarse elements and k on the fine elements.

Proof. The full discretization error is given by $\mathbf{e}^n = \mathbf{e}_\pi^n - \widetilde{\mathbf{e}}_h^n - \widetilde{\boldsymbol{\xi}}^{n-1}$. Using $\big(\mathbf{e}_\pi, \varphi_h\big)_{\mu\times\varepsilon} = 0$, and the triangle inequality and Young's inequality we infer

$$\|\mathbf{e}^n\|_{\mu\times\varepsilon}^2 \le \|\mathbf{e}_\pi^n\|_{\mu\times\varepsilon}^2 + 2\|\widetilde{\mathbf{e}}_h^n\|_{\mu\times\varepsilon}^2 + 2\|\widetilde{\boldsymbol{\xi}}^{n-1}\|_{\mu\times\varepsilon}^2.$$

The first and the last term can be bounded with (3.24a) and (5.61), respectively, which yields

$$\|\mathbf{e}_\pi^n\|_{\mu\times\varepsilon}^2 \leq C_{\text{app}}^2 |\mathbf{u}|_{k+1,\mathcal{T}_h,2,k+1}, \qquad \|\widetilde{\boldsymbol{\xi}}^{n-1}\|_{\mu\times\varepsilon}^2 \leq \frac{\tau^4}{16} \max_{t\in[t_{n-1},t_n]} \|\partial_t^2 \mathbf{H}(t)\|_{\mu,\mathcal{T}_{h,e}}^2.$$

For the remaining error $\widetilde{\mathbf{e}}_h^n$ we have the bound (5.100) and inserting

$$\left(\mathbf{d}_h^n, \varphi_h\right)_{\mu\times\varepsilon} \leq \gamma_2\tau \|\varphi_h\|_{\mu\times\varepsilon}^2 + \frac{C}{\gamma_2}\tau^4 \int_{t_n}^{t_{n+1}} \|\partial_t^3 \mathbf{u}(t)\|_{\mu\times\varepsilon}^2 \, dt,$$

and the results from Lemmas 5.30–5.34 with $\gamma_1 = \alpha/2(1+2\alpha^2)$ we obtain

$$\begin{aligned}
(1-\widetilde{\theta}_{\text{upw}})\|\widetilde{\mathbf{e}}_h^{n+1}\|_{\mu\times\varepsilon}^2 + \alpha\frac{\tau}{4}\sum_{m=0}^{n} |\widetilde{\mathbf{e}}_h^{m+1} + \widetilde{\mathbf{e}}_h^m|_{\boldsymbol{\mathcal{S}}^e}^2 \\
&\leq 3\gamma_2\tau \sum_{m=1}^{n} \|\widetilde{\mathbf{e}}_h^{m+1} + \widetilde{\mathbf{e}}_h^m\|_{\mu\times\varepsilon}^2 \\
&\quad + C|\mathbf{u}^1 + \mathbf{u}^0|_{k+1,\mathcal{T}_h,2,k+1}^2 \\
&\quad + C\tau^4 \max_{t\in[0,\tau]} \|\partial_t^2 \mathbf{H}(t)\|_{\mu,\mathcal{T}_{h,c}}^2 \\
&\quad + \frac{C}{\gamma_2}\tau^4 \int_0^{t_{n+1}} \|\partial_t^3 \mathbf{u}(t)\|_{\mu\times\varepsilon}^2 \, dt \\
&\quad + \frac{C}{\gamma_1}\tau \sum_{m=0}^{n} \left(\left|\mathbf{u}^{m+1} + \mathbf{u}^m\right|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2 + |\mathbf{u}^m|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2 \right) \\
&\quad + \frac{C}{\gamma_2}\tau \sum_{m=0}^{n} \left|\mathbf{u}^{n+1} + \mathbf{u}^n\right|_{k+1,\mathcal{T}_{h,i},2,k}^2 + \frac{C}{\gamma_2}\tau \int_0^{t_{n+1}} |\partial_t \mathbf{E}(t)|_{k+1,\mathcal{T}_{h,c},2,k+\frac{1}{2}}^2 \, dt \\
&\quad + \frac{C}{\gamma_1} \int_0^{t_n} \left(\tau^4 \|\mu\partial_t^2 \mathbf{H}(t)\|_{1,\mathcal{T}_{h,e}}^2 + |\partial_t^2 \mathbf{H}(t)|_{\max(1,k-1),\mathcal{T}_{h,e},2,k+\frac{1}{2}}^2 \right) dt.
\end{aligned}$$

By the triangle inequality, Young's inequality and by choosing the weight $\gamma_2 = \frac{1-\widetilde{\theta}_{\text{upw}}}{12(T+1)}$ we have

$$3\gamma_2\tau\|\widetilde{\mathbf{e}}_h^{m+1} + \widetilde{\mathbf{e}}_h^m\|_{\mu\times\varepsilon}^2 \leq \frac{1-\widetilde{\theta}_{\text{upw}}}{T+1}\frac{\tau}{2}\left(\|\widetilde{\mathbf{e}}_h^{m+1}\|_{\mu\times\varepsilon}^2 + \|\widetilde{\mathbf{e}}_h^m\|_{\mu\times\varepsilon}^2\right),$$

and thus the discrete Gronwall lemma (Lemma A.2) yields the result. □

5.5 The locally implicit scheme and the implicit midpoint method

In this section we present the locally implicit scheme when the implicit time integration is carried out with the implicit midpoint method instead of the Crank–Nicolson method. We restrict ourselves in this section to a central fluxes dG space discretization.

The only difference of the Crank–Nicolson and the implicit midpoint time integration for the semidiscrete Maxwell's equations stemming from a central fluxes dG discretization (3.8) is the treatment of the source term, see (4.40) and (4.93). So, by substituting $\frac{\tau}{2}(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n)$ to $\tau\mathbf{j}_h^n$ we change from the Crank–Nicolson scheme to the implicit midpoint method. The same holds true for the locally implicit scheme (5.11), i.e. our **implicit midpoint method locally implicit**

scheme reads

$$\mathbf{H}_h^{n+1/2} - \mathbf{H}_h^n = -\frac{\tau}{2}\mathcal{C}_{\mathbf{E}}\mathbf{E}_h^n, \tag{5.105a}$$

$$\mathbf{E}_h^{n+1} - \mathbf{E}_h^n = \tau\mathcal{C}_{\mathbf{H}}^e\mathbf{H}_h^{n+1/2} + \frac{\tau}{2}\mathcal{C}_{\mathbf{H}}^i(\mathbf{H}_h^{n+1} + \mathbf{H}_h^n) - \tau\mathbf{J}_h^{n+1/2}, \tag{5.105b}$$

$$\mathbf{H}_h^{n+1} - \mathbf{H}_h^{n+1/2} = -\frac{\tau}{2}\mathcal{C}_{\mathbf{E}}\mathbf{E}_h^{n+1}, \tag{5.105c}$$

or, more compactly,

$$\widetilde{\boldsymbol{\mathcal{R}}}_L\mathbf{u}_h^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}}_R\mathbf{u}_h^n + \tau\mathbf{j}_h^{n+1/2}, \tag{5.106}$$

where the operators $\widetilde{\boldsymbol{\mathcal{R}}}_L$ and $\widetilde{\boldsymbol{\mathcal{R}}}_R$ were defined in (5.13b). Clearly, the implicit midpoint locally implicit scheme (5.105) satisfies the same stability and energy conservation properties as the (Crank–Nicolson) locally implicit scheme (5.11), see Corollaries 5.9, 5.10.

Corollary 5.37. *Assume that the CFL condition (5.40) is satisfied with parameter $\widetilde{\theta} \in (0,1)$. Then, the approximation $\mathbf{u}_h^n$ obtained from the implicit midpoint locally implicit method (5.105) is bounded by*

$$\|\mathbf{u}_h^n\|_{\mu\times\varepsilon} \le C_{\mathrm{stb},c}^{1/2}\|\mathbf{u}^0\|_{\mu\times\varepsilon} + C_{\mathrm{stb},c}^{3/2}\frac{\tau}{\sqrt{\delta}}\sum_{m=0}^{n-1}\|\mathbf{J}^{m+1/2}\|. \tag{5.107}$$

Moreover, for vanishing source term $\mathbf{J}_h \equiv 0$, the perturbed electromagnetic energy $\widetilde{\mathcal{E}}(\mathbf{H}_h, \mathbf{E}_h)$ defined in (5.45) is conserved,

$$\widetilde{\mathcal{E}}(\mathbf{H}_h^n, \mathbf{E}_h^n) = \widetilde{\mathcal{E}}(\mathbf{H}_h^0, \mathbf{E}_h^0), \qquad n = 1, 2, \dots .$$

Now, we turn to the error analysis of (5.105). First, we present its error recursion.

Lemma 5.38. *Let $\mathbf{u} \in C\big(0,T;V_\star\big) \cap C^3\big(0,T;L^2(\Omega)^6\big)$ be the exact solution of (5.1). The error $\mathbf{e}_h^n$ of the implicit midpoint locally implicit scheme (5.105) satisfies*

$$\widetilde{\boldsymbol{\mathcal{R}}}_L\mathbf{e}_h^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}}_R\mathbf{e}_h^n + \bar{\bar{\mathbf{d}}}^n, \qquad \bar{\bar{\mathbf{d}}}^n = \widetilde{\mathbf{d}}_\pi^n + \bar{\bar{\mathbf{d}}}_h^n. \tag{5.108a}$$

The projection defect $\widetilde{\mathbf{d}}_\pi^n$ was defined in (5.50) and the quadrature defect $\bar{\bar{\mathbf{d}}}_h^n$ is given by

$$\bar{\bar{\mathbf{d}}}_h^n = \bar{\mathbf{d}}_h^n - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \mathcal{C}_{\mathbf{H}}^e\pi_h\Delta_{\mathbf{H}}^n \end{pmatrix}, \tag{5.108b}$$

where $\bar{\mathbf{d}}_h^n$ was introduced in (4.95b). The defect can be written as

$$\bar{\bar{\mathbf{d}}}^n = \bar{\bar{\mathbf{d}}}_\pi^n - \tau^2\pi_h\bar{\delta}^n(\partial_t\mathbf{u}) + (\widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R)(\pi_h\bar{\boldsymbol{\zeta}}^n + \widetilde{\boldsymbol{\xi}}^n), \tag{5.108c}$$

with

$$\bar{\bar{\mathbf{d}}}_\pi^n = \bar{\mathbf{d}}_\pi^n - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \mathcal{C}_{\mathbf{H}}^e\mathcal{C}_{\mathbf{E}}(\mathbf{e}_{\pi,\mathbf{E}}^{n+1} - \mathbf{e}_{\pi,\mathbf{E}}^n) \end{pmatrix}, \tag{5.108d}$$

where $\bar{\mathbf{d}}_\pi^n$ and $\bar{\boldsymbol{\zeta}}^n$ were defined in (4.95d) and $\widetilde{\boldsymbol{\xi}}^n$ was defined in (5.57a).

Proof. (a) In (4.96) we obtained the recursion

$$\boldsymbol{\mathcal{R}}_L\pi_h\mathbf{u}^{n+1} = \boldsymbol{\mathcal{R}}_R\pi_h\mathbf{u}^n + \tau\mathbf{j}_h^{n+1/2} - \bar{\mathbf{d}}^n \tag{5.109}$$

for the projection of the exact solution $\mathbf{u}$. By inserting $\pi_h\mathbf{u}$ into the implicit midpoint locally implicit scheme (5.106) we have

$$\widetilde{\boldsymbol{\mathcal{R}}}_L\pi_h\mathbf{u}^{n+1} = \widetilde{\boldsymbol{\mathcal{R}}}_R\pi_h\mathbf{u}^n + \tau\mathbf{j}_h^{n+1/2} - \bar{\bar{\mathbf{d}}}^n. \tag{5.110}$$

Subtracting (5.110) from (5.106) yields the error recursion (5.108a) where the defect $\bar{\bar{\mathbf{d}}}^n$ yet has to determined. This is achieved by subtracting (5.110) from (5.109) which implies

$$\begin{aligned}\bar{\bar{\mathbf{d}}}^n &= \bar{\mathbf{d}}^n + (\boldsymbol{\mathcal{R}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_L)\pi_h \mathbf{u}^{n+1} - (\boldsymbol{\mathcal{R}}_R - \widetilde{\boldsymbol{\mathcal{R}}}_R)\pi_h \mathbf{u}^n \\ &= \bar{\mathbf{d}}^n - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}^e_{\mathbf{H}} \pi_h \Delta^n_{\mathbf{H}} \end{pmatrix} - \frac{\tau^2}{4}\begin{pmatrix} 0 \\ \boldsymbol{\mathcal{C}}^e_{\mathbf{H}} \boldsymbol{\mathcal{C}}_{\mathbf{E}} (\mathbf{e}^{n+1}_{\pi,\mathbf{E}} - \mathbf{e}^n_{\pi,\mathbf{E}}) \end{pmatrix}.\end{aligned}$$

Here, the second equality follows analog to the proof of Lemma 5.11.

(b) The representation (5.108c) of the defect $\bar{\bar{\mathbf{d}}}^n$ follows with (4.95c), $\boldsymbol{\mathcal{R}}_L - \boldsymbol{\mathcal{R}}_R = \widetilde{\boldsymbol{\mathcal{R}}}_L - \widetilde{\boldsymbol{\mathcal{R}}}_R$, and (5.57a). □

Now, we state the convergence result.

Theorem 5.39. *Let* $\mathbf{u} \in C\big(0,T; D(\mathcal{C}) \cap H^{k+1}(\mathcal{T}_h)^6\big) \cap C^3\big(0,T; L^2(\Omega)^6\big)$ *be the exact solution of* (5.1). *Moreover, assume that the CFL condition* (5.40) *is satisfied with* $\widetilde{\theta} \in (0,1)$. *Then, the error of the central fluxes dG discretization and the implicit midpoint locally implicit scheme* (5.105) *is bounded by*

$$\begin{aligned}\|\mathbf{u}^n - \mathbf{u}^n_h\|_{\mu\times\varepsilon} \leq & C_{\mathrm{app}} |\mathbf{u}^n|_{k+1,\mathcal{T}_h,1,k+1} \\ & + C^{3/2}_{\mathrm{stb},c} \widehat{C}_\pi \tau \sum_{m=0}^{n-1} \Big(|\mathbf{u}^{m+1/2}|_{k+1,\mathcal{T}_h,2,k} + \frac{1}{2} |\mathbf{E}^{m+1} - \mathbf{E}^m|_{k+1,\mathcal{T}_h,2,k} \Big) \\ & + (1 + C^{1/2}_{\mathrm{stb},c}) \frac{\tau^2}{4} \max_{t\in[0,t_n]} \|\partial_t^2 \mathbf{u}(t)\|_{\mu\times\varepsilon} \\ & + C^{1/2}_{\mathrm{stb},c} (4 + C_{\mathrm{stb},c}) \frac{\tau^2}{8} \int_0^{t_n} \|\partial_t^3 \mathbf{u}(t)\|_{\mu\times\varepsilon}\, dt \\ \leq & C\Big(h^k + \tau^2\Big).\end{aligned}$$

Proof. For the full discretization error $\mathbf{e}^n = \mathbf{e}^n_\pi - \mathbf{e}^n_h$ we have by (3.24a)

$$\|\mathbf{e}^n\|_{\mu\times\varepsilon} = \|\mathbf{e}^n_\pi\|_{\mu\times\varepsilon} + \|\mathbf{e}^n_h\|_{\mu\times\varepsilon} \leq C_{\mathrm{app}} |\mathbf{u}^n|_{k+1,\mathcal{T}_h,1,k+1} + \|\mathbf{e}^n_h\|_{\mu\times\varepsilon},$$

since $\big(\mathbf{e}^n_\pi, \mathbf{e}^n_h\big)_{\mu\times\varepsilon} = 0$. Under the assumption of the CFL condition the operator $\widetilde{\boldsymbol{\mathcal{R}}}_L$ is invertible and by solving the recursion (5.108a) we obtain

$$\begin{aligned}\mathbf{e}^{n+1}_h = & \pi_h \bar{\boldsymbol{\zeta}}^n + \widetilde{\boldsymbol{\xi}}^n - \boldsymbol{\mathcal{R}}^{n+1}\big(\pi_h \bar{\boldsymbol{\zeta}}^0 + \widetilde{\boldsymbol{\xi}}^0\big) \\ & + \sum_{m=0}^{n} \boldsymbol{\mathcal{R}}^{n-m} \widetilde{\boldsymbol{\mathcal{R}}}_L^{-1} \big(\bar{\bar{\mathbf{d}}}^m_\pi - \tau^2 \pi_h \bar{\delta}^n(\partial_t \mathbf{u}) \big) - \sum_{m=0}^{n-1} \widetilde{\boldsymbol{\mathcal{R}}}^{n-m} \big(\pi_h(\bar{\boldsymbol{\zeta}}^{m+1} - \bar{\boldsymbol{\zeta}}^m) + \widetilde{\boldsymbol{\xi}}^{m+1} - \widetilde{\boldsymbol{\xi}}^m \big),\end{aligned}$$

compare (5.58). The assertion now follows with the bounds (5.42) and (5.43) on $\widetilde{\boldsymbol{\mathcal{R}}}_L^{-1}$ and $\widetilde{\boldsymbol{\mathcal{R}}}^m$, respectively, and with the bounds (3.43), (4.29d), (4.33c), (4.35),(5.55), (5.60) and (5.61). □

CHAPTER 6

Implementation and numerical results

Our last chapter is dedicated to the efficient numerical implementation of the central fluxes and the upwind fluxes locally implicit schemes. This issue will be discussed in Sections 6.1 and 6.2. Moreover, in Section 6.3 we provide numerical examples illustrating the efficiency, the sole dependence of the CFL condition on the coarse mesh elements, and the theoretical convergence order we obtained in the previous sections.

6.1 Efficient formulation of the locally implicit schemes

Given $\mathbf{u}_h^n$ our locally implicit schemes (5.11) and (5.65) require the solution of the following linear system

$$\widetilde{\boldsymbol{\mathcal{R}}}_L \mathbf{u}_h^{n+1} = \widetilde{\mathbf{b}}_h^n(\alpha), \qquad \widetilde{\mathbf{b}}_h^n(\alpha) = \widetilde{\boldsymbol{\mathcal{R}}}_R \mathbf{u}_h^n - \tau\alpha \boldsymbol{\mathcal{S}}^e \mathbf{u}_h^n + \frac{\tau}{2}\big(\mathbf{j}_h^{n+1} + \mathbf{j}_h^n\big), \tag{6.1}$$

to compute $\mathbf{u}_h^{n+1}$. For $\alpha = 0$ we obtain the right-hand side of the central fluxes locally implicit time integrator and for $\alpha \in (0, 1]$ the right-hand side of the upwind fluxes locally implicit scheme. In particular, both schemes exhibit the **same left-hand side** of the linear system which has to be solved in every time step. In the following we will drop the argument α in $\mathbf{b}_h^n$ and only write it if necessary.

At first glance (6.1) seems to be a linear system of all degrees of freedom (dof) of $\mathbf{u}_h = (\mathbf{H}_h, \mathbf{E}_h)$. However, similar to the Crank–Nicolson method in Section 4.5.1, we can reduce this system by using a Schur decomposition. This yields the equivalent linear system

$$\begin{pmatrix} \boldsymbol{\mathcal{I}} & \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}} \\ 0 & \widetilde{\boldsymbol{\mathcal{L}}} \end{pmatrix} \begin{pmatrix} \mathbf{H}_h^{n+1} \\ \mathbf{E}_h^{n+1} \end{pmatrix} = \begin{pmatrix} \widetilde{\mathbf{b}}_{\mathbf{H}}^n \\ \widetilde{\mathbf{b}}_{\mathbf{E}}^n + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}} \widetilde{\mathbf{b}}_{\mathbf{H}}^n \end{pmatrix}, \tag{6.2a}$$

where the right-hand side reads

$$\widetilde{\mathbf{b}}_{\mathbf{H}}^n = \mathbf{H}_h^n - \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}} \mathbf{E}_h^n - \tau\alpha \boldsymbol{\mathcal{S}}_{\mathbf{H}}^e \mathbf{H}_h^n, \tag{6.2b}$$

$$\widetilde{\mathbf{b}}_{\mathbf{E}}^n = \mathbf{E}_h^n + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}} \mathbf{H}_h^n - \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}_{\mathbf{H}}^e \boldsymbol{\mathcal{C}}_{\mathbf{E}} \mathbf{E}_h^n - \tau\alpha \boldsymbol{\mathcal{S}}_{\mathbf{E}}^e \mathbf{E}_h^n - \frac{\tau}{2}\big(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n\big), \tag{6.2c}$$

and where the Schur complement is given by

$$\widetilde{\boldsymbol{\mathcal{L}}} = \boldsymbol{\mathcal{I}} - \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}} + \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}} = \boldsymbol{\mathcal{I}} + \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}}. \tag{6.2d}$$

Note that solving (6.2a) only requires to solve a linear system on the dof of the electric field $\mathbf{E}_h$. Comparing (6.2d) with the Schur complement $\boldsymbol{\mathcal{L}}$ of the central fluxes Crank–Nicolson method (4.98b), we see that both are linear systems on all dof of the electric field, and they only differ in the fact that the locally implicit methods involve $\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}$ whereas the Crank–Nicolson method uses $\boldsymbol{\mathcal{C}}_{\mathbf{H}}$. A detailed discussion why the locally implicit methods can be implemented far more efficiently is provided in the next section. Moreover, by construction, the upwind fluxes locally implicit scheme integrates the stabilization operators explicitly. Hence, its linear system does not involve these operators – in contrast to the upwind fluxes Crank–Nicolson method, see (4.99c). This is of great computational advantage, in particular if we want to use direct solvers, see the discussion in Section 4.5.1.

Next, we discuss the right-hand side of (6.2a). Note that we have $\mathbf{H}_h^{n+1/2} = \widetilde{\mathbf{b}}^n_{\mathbf{H}} + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\mathbf{H}_h^n$, and thus

$$\begin{aligned}\widetilde{\mathbf{b}}^n_{\mathbf{E}} + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\widetilde{\mathbf{b}}^n_{\mathbf{H}} &= \mathbf{E}_h^n - \tau\alpha\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}\mathbf{E}_h^n \\ &\quad + \tau\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\mathbf{H}_h^{n+1/2} + \frac{\tau}{2}\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}\big(\mathbf{H}_h^{n+1/2} + \mathbf{H}_h^n - \frac{\tau}{2}\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\mathbf{H}_h^n\big) - \frac{\tau}{2}\big(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n\big).\end{aligned} \tag{6.3}$$

In the central fluxes case this simplifies to

$$\widetilde{\mathbf{b}}^n_{\mathbf{E}}(0) + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\widetilde{\mathbf{b}}^n_{\mathbf{H}}(0) = \mathbf{E}_h^n + \tau\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}\mathbf{H}_h^{n+1/2} + \frac{\tau}{2}\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}\big(\mathbf{H}_h^{n+1/2} + \mathbf{H}_h^n\big) - \frac{\tau}{2}\big(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n\big), \tag{6.4}$$

where we used (5.11a).

We summarize the computations needed to perform the time step from $\mathbf{u}_h^n$ to $\mathbf{u}_h^{n+1}$ with the locally implicit methods in Algorithm 6.1.

Algorithm 6.1 Update from $\mathbf{H}_h^n$, $\mathbf{E}_h^n$ to $\mathbf{H}_h^{n+1}$, $\mathbf{E}_h^{n+1}$ in the locally implicit methods

Given $\mathbf{H}_h^n$, $\mathbf{E}_h^n$:

1: Compute $\mathbf{H}_h^{n+1/2} = \widetilde{\mathbf{b}}^n_{\mathbf{H}} + \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\mathbf{H}_h^n$ by (6.2b)

2: Compute $\widetilde{\mathbf{b}}^n_{\mathbf{E}} + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\widetilde{\mathbf{b}}^n_{\mathbf{H}}$ by (6.3) (upwind fluxes) or by (6.4) (central fluxes)

3: Solve $\widetilde{\boldsymbol{\mathcal{L}}}\mathbf{E}_h^{n+1} = \widetilde{\mathbf{b}}^n_{\mathbf{E}} + \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\widetilde{\mathbf{b}}^n_{\mathbf{H}}$ with $\widetilde{\boldsymbol{\mathcal{L}}}$ given in (6.2d)

4: Update $\mathbf{H}_h^{n+1} = \mathbf{H}_h^{n+1/2} - \frac{\tau}{2}\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1} - \frac{\tau}{2}\alpha\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}\mathbf{H}_h^n$

Note that for the central fluxes case the computation of $\mathbf{H}_h^{n+1/2}$ in Line 1 can be replaced by $\mathbf{H}_h^{n+3/2} = 2\mathbf{H}_h^{n+1} - \mathbf{H}_h^n$ for $n \geq 1$. In the upwind fluxes case the value $\boldsymbol{\mathcal{C}}_{\mathbf{E}}\mathbf{E}_h^{n+1}$ might be saved in order to use it in Line 1 for the next step. In summary, carrying out one step of the central fluxes locally implicit method needs one matrix-vector multiplication with (the matrices associated with) $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$, one with $\boldsymbol{\mathcal{C}}^e_{\mathbf{H}}$, one with $\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}$ and the solution of a linear system involving $\widetilde{\boldsymbol{\mathcal{L}}}$. The upwind locally implicit method additionally needs one matrix-vector multiplication with $\boldsymbol{\mathcal{S}}^e_{\mathbf{H}}$ and one with $\boldsymbol{\mathcal{S}}^e_{\mathbf{E}}$.

As indicated above, our next aim is to analyze the left-hand sides $\widetilde{\boldsymbol{\mathcal{L}}}$ and $\boldsymbol{\mathcal{L}}$ of the locally implicit scheme and the Crank–Nicolson method, respectively, and show why (and in which setting) the locally implicit scheme can be implemented more efficiently than the Crank–Nicolson method.

	$m \in \{1,\dots,\frac{N_h}{2}\}$ $\varphi_m = \begin{pmatrix}\phi_m\\0\end{pmatrix}$	$m \in \{\frac{N_h}{2}+1,\dots,N_h\}$ $\varphi_m = \begin{pmatrix}0\\\psi_m\end{pmatrix}$
$\ell \in \{1,\dots,\frac{N_h}{2}\}$ $\varphi_\ell = \begin{pmatrix}\phi_\ell\\0\end{pmatrix}$	$*$	0
$\ell \in \{\frac{N_h}{2}+1,\dots,N_h\}$ $\varphi_\ell = \begin{pmatrix}0\\\psi_\ell\end{pmatrix}$	0	$*$

Table 6.1: Zero and possibly nonzero entries of the mass matrix $M_{\ell,m} = \big(\varphi_m,\varphi_\ell\big)_{\mu\times\varepsilon} = \big(\phi_m,\phi_\ell\big)_\mu - \big(\psi_m,\psi_\ell\big)_\varepsilon$.

	$m \in \{1,\dots,\frac{N_h}{2}\}$ $\varphi_m = \begin{pmatrix}\phi_m\\0\end{pmatrix}$	$m \in \{\frac{N_h}{2}+1,\dots,N_h\}$ $\varphi_m = \begin{pmatrix}0\\\psi_m\end{pmatrix}$
$\ell \in \{1,\dots,\frac{N_h}{2}\}$ $\varphi_\ell = \begin{pmatrix}\phi_\ell\\0\end{pmatrix}$	0	$*$
$\ell \in \{\frac{N_h}{2}+1,\dots,N_h\}$ $\varphi_\ell = \begin{pmatrix}0\\\psi_\ell\end{pmatrix}$	$*$	0

Table 6.2: Zero and possibly nonzero entries of the stiffness matrix $C_{\ell,m} = \big(\boldsymbol{\mathcal{C}}\varphi_m,\varphi_\ell\big)_{\mu\times\varepsilon} = \big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}\phi_m,\psi_\ell\big)_\varepsilon - \big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\psi_m,\phi_\ell\big)_\mu$.

6.2 Efficient numerical implementation

In this section we want to compare the linear systems of the (central fluxes and upwind fluxes) locally implicit scheme $\widetilde{\boldsymbol{\mathcal{L}}}$ with the linear system of the central fluxes Crank–Nicolson method $\boldsymbol{\mathcal{L}}$. We recall from (6.2d) and (4.98b) that we have

$$\widetilde{\boldsymbol{\mathcal{L}}} = \boldsymbol{\mathcal{I}} + \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}\boldsymbol{\mathcal{C}}^i_{\mathbf{E}}, \qquad\qquad \boldsymbol{\mathcal{L}} = \boldsymbol{\mathcal{I}} + \frac{\tau^2}{4}\boldsymbol{\mathcal{C}}_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}},$$

where we used $\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}\boldsymbol{\mathcal{C}}_{\mathbf{E}} = \boldsymbol{\mathcal{C}}^i_{\mathbf{H}}\boldsymbol{\mathcal{C}}^i_{\mathbf{E}}$ for $\widetilde{\boldsymbol{\mathcal{L}}}$. In order to analyze the costs associated with these linear systems we have to examine $\widetilde{\boldsymbol{\mathcal{L}}}$ and $\boldsymbol{\mathcal{L}}$ in a representation w.r.t. a basis of V_h^2. As in Section 3.6 we consider the basis

$$\{\varphi_1,\dots,\varphi_{N_h}\}, \qquad \varphi_\ell = \begin{pmatrix}\phi_\ell\\\psi_\ell\end{pmatrix},$$

and recall that the support of these basis functions consists of a single mesh element,

$$\operatorname{supp}(\varphi_\ell) \subset \overline{K}, \qquad \text{for a } K \in \mathcal{T}_h.$$

A natural ordering of the basis functions is given by

$$\begin{pmatrix}\phi_1\\0\end{pmatrix},\dots,\begin{pmatrix}\phi_{N_h/2}\\0\end{pmatrix},\begin{pmatrix}0\\\psi_{N_h/2+1}\end{pmatrix},\dots,\begin{pmatrix}0\\\psi_{N_h}\end{pmatrix}, \tag{6.5}$$

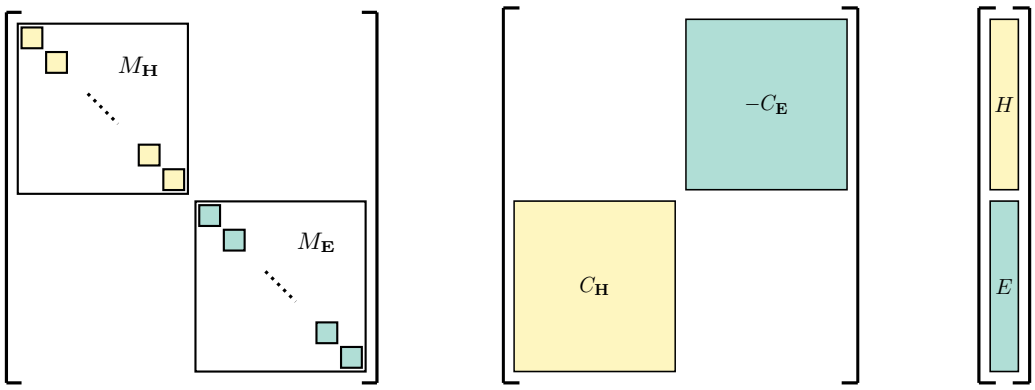


Figure 6.1: Structure of the mass matrix M (left), the stiffness matrix C (middle) and the coefficient vector u (right) for the ordering (6.5) of the basis functions.

i.e. we first take the basis functions for the **H** components and subsequently the ones for the **E** components. We recall that by (3.55b) and (3.55c) the entries of the mass matrix M and of the stiffness matrix C are given by

$$M_{\ell,m} = \left(\varphi_m, \varphi_\ell\right)_{\mu\times\varepsilon} = \left(\phi_m, \phi_\ell\right)_{\mu} + \left(\psi_m, \psi_\ell\right)_{\varepsilon},$$

and

$$C_{\ell,m} = \left(\boldsymbol{\mathcal{C}}\varphi_m, \varphi_\ell\right)_{\mu\times\varepsilon} = \left(\boldsymbol{\mathcal{C}}_{\mathbf{H}}\phi_m, \psi_\ell\right)_{\varepsilon} + \left(-\boldsymbol{\mathcal{C}}_{\mathbf{E}}\psi_m, \phi_\ell\right)_{\mu},$$

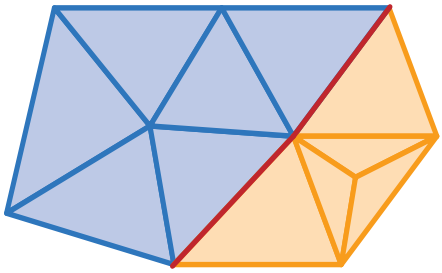
respectively. The structures of these matrices are given in Tables 6.1, 6.2 and visualized in Figure 6.1. In this figure we already indicated the block structure of the mass matrix. In fact, the mass matrix is block-diagonal where the block size corresponds to the number of dof in one spatial mesh element. As we can observe in Figure 6.1 both the mass and the stiffness matrix have a block structure

$$M = \begin{pmatrix} M_{\mathbf{H}} & 0 \\ 0 & M_{\mathbf{E}} \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & -C_{\mathbf{E}} \\ C_{\mathbf{H}} & 0 \end{pmatrix}.$$

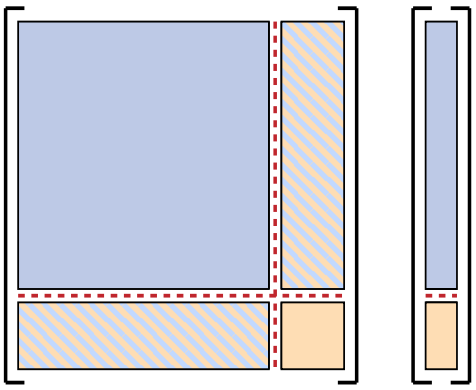
Comparing (3.15b) with (3.55a) we conclude that the operator $\boldsymbol{\mathcal{C}}$ corresponds to the matrix $M^{-1}C$ and vice versa. Thus, the operators $\boldsymbol{\mathcal{C}}_{\mathbf{H}}$ and $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$ correspond to the matrices $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}$ and $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}$, respectively.

Our next aim is to derive the structure of the matrices associated with $\boldsymbol{\mathcal{C}}_{\mathbf{H}}^i$ and $\boldsymbol{\mathcal{C}}_{\mathbf{E}}^i$, which we denote by $C_{\mathbf{H}}^i$ and $C_{\mathbf{E}}^i$, respectively. For this purpose we decompose the stiffness matrices $C_{\mathbf{H}}$ and $C_{\mathbf{E}}$ into explicitly and implicitly treated elements. By (3.5), (5.2), (5.3a) and (5.4) this reads

$$\begin{aligned}(C_{\mathbf{H}})_{\ell,m} = &\sum_{K\in\mathcal{T}_{h,i}} \left(\operatorname{curl}\phi_m, \psi_\ell\right)_{\varepsilon,K} + \sum_{F\in\mathcal{F}_{h,i}^{\mathrm{int}}} \left(n_F \times [\![\phi_m]\!]_F, \{\!\{\psi_\ell\}\!\}_F^{\varepsilon c}\right)_F \\ &+ \sum_{F\in\mathcal{F}_{h,ci}^{\mathrm{int}}} \left(n_F \times [\![\phi_m]\!]_F, \{\!\{\psi_\ell\}\!\}_F^{\varepsilon c}\right)_F \\ &+ \sum_{K\in\mathcal{T}_{h,e}} \left(\operatorname{curl}\phi_m, \psi_\ell\right)_{\varepsilon,K} + \sum_{F\in\mathcal{F}_{h,e}^{\mathrm{int}}} \left(n_F \times [\![\phi_m]\!]_F, \{\!\{\psi_\ell\}\!\}_F^{\varepsilon c}\right)_F,\end{aligned}$$



(a) Splitting of the mesh elements into $\mathcal{T}_{h,e}$ (blue) and $\mathcal{T}_{h,i}$ (orange).



(b) Left: Structure of full stiffness matrices $C_{\mathbf{H}}$ and $C_{\mathbf{E}}$. Explicit element coupling: $\mathcal{T}_{h,e}$ to $\mathcal{T}_{h,e}$ (blue), implicit element coupling: $\mathcal{T}_{h,i}$ to $\mathcal{T}_{h,i}$ (orange), explicit-implicit element coupling: $\mathcal{T}_{h,e}$ to $\mathcal{T}_{h,i}$ (blue-orange striped). Right: Coefficient vector.

Figure 6.2: (a) Splitting of the mesh into explicitly and implicitly treated elements. (b) Stiffness matrix and coefficient vector corresponding to the ordering (6.6) of the basis functions.

	$m \in S_e$ $\operatorname{supp}(\chi_i\phi_m) \subset \mathcal{T}_{h,e}$	$m \in S_i$ $\operatorname{supp}(\chi_i\phi_m) \subset \mathcal{T}_{h,i}$
$\ell \in \frac{N_h}{2} + S_e$ $\operatorname{supp}\psi_\ell \subset \mathcal{T}_{h,e}$	0	$*$
$\ell \in \frac{N_h}{2} + S_i$ $\operatorname{supp}\psi_\ell \subset \mathcal{T}_{h,i}$	0	$*$

Table 6.3: Zero and possibly nonzero entries of the implicit stiffness matrix $(C^i_{\mathbf{H}})_{\ell,m} = \left(\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}\phi_m, \psi_\ell\right)_\varepsilon$.

and

$$
\begin{aligned}
(C_{\mathbf{E}})_{\ell,m} = &\sum_{K\in\mathcal{T}_{h,i}} \left(\operatorname{curl}\psi_m, \phi_\ell\right)_{\mu,K} + \sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,i}} \left(n_F \times [\![\psi_m]\!]_F, \{\!\{\phi_\ell\}\!\}^{\mu c}_F\right)_F - \sum_{F\in\mathcal{F}^{\mathrm{bnd}}_{h,i}} \left(n_F\times\psi_m, \phi_\ell\right)_F \\
&+ \sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,ci}} \left(n_F \times [\![\psi_m]\!]_F, \{\!\{\phi_\ell\}\!\}^{\mu c}_F\right)_F \\
&+ \sum_{K\in\mathcal{T}_{h,e}} \left(\operatorname{curl}\psi_m, \phi_\ell\right)_{\mu,K} + \sum_{F\in\mathcal{F}^{\mathrm{int}}_{h,e}} \left(n_F \times [\![\psi_m]\!]_F, \{\!\{\phi_\ell\}\!\}^{\mu c}_F\right)_F - \sum_{F\in\mathcal{F}^{\mathrm{bnd}}_{h,e}} \left(n_F\times\psi_m, \phi_\ell\right)_F.
\end{aligned}
$$

Here the respective first and third lines only involve basis functions with support on $\mathcal{T}_{h,i}$ and $\mathcal{T}_{h,e}$, respectively, and the second lines contain the coupling between these two sets. It is natural to order the basis functions corresponding to their belonging to the sets $\mathcal{T}_{h,e}$ and $\mathcal{T}_{h,i}$. Let $N_e = n_h \operatorname{card}(\mathcal{T}_{h,e})$ denote the dof in the explicitly treated part $\mathcal{T}_{h,e}$, where $n_h = \dim\left((\mathbb{P}^k_3)^6\right)$, see Section 3.6. We introduce the sets

$$
S_e = \{1, \dots, N_e/2\}, \qquad S_i = \{N_e/2+1, \dots, N_h/2\},
$$

and order our basis functions by

$$
\begin{pmatrix}\phi_\ell\\0\end{pmatrix}_{\ell\in S_e}, \begin{pmatrix}\phi_\ell\\0\end{pmatrix}_{\ell\in S_i}, \begin{pmatrix}0\\\psi_\ell\end{pmatrix}_{\ell\in N_h/2+S_e}, \begin{pmatrix}0\\\psi_\ell\end{pmatrix}_{\ell\in N_h/2+S_i}, \tag{6.6}
$$

	$m \in \frac{N_h}{2} + S_e$ $\operatorname{supp}\psi_m \subset \mathcal{T}_{h,e}$	$m \in \frac{N_h}{2} + S_i$ $\operatorname{supp}\psi_m \subset \mathcal{T}_{h,i}$
$\ell \in S_e$ $\operatorname{supp}(\chi_i\phi_\ell) \subset \mathcal{T}_{h,e}$	0	0
$\ell \in S_i$ $\operatorname{supp}(\chi_i\phi_\ell) \subset \mathcal{T}_{h,i}$	*	*

Table 6.4: Zero and possibly nonzero entries of the implicit stiffness matrix $(C^i_{\mathbf{E}})_{\ell,m} = \big(\boldsymbol{\mathcal{C}}^i_{\mathbf{E}}\psi_m, \phi_\ell\big)_\mu$.

such that they satisfy

$$\operatorname{supp}(\varphi_\ell) = \operatorname{supp}\left(\begin{pmatrix}\phi_\ell\\ \psi_\ell\end{pmatrix}\right) \subset \overline{K}, K \in \mathcal{T}_{h,e}, \qquad \text{for } \ell \in \big(S_e \cup (N_h/2 + S_e)\big),$$
$$\operatorname{supp}(\varphi_\ell) = \operatorname{supp}\left(\begin{pmatrix}\phi_\ell\\ \psi_\ell\end{pmatrix}\right) \subset \overline{K}, K \in \mathcal{T}_{h,i}, \qquad \text{for } \ell \in \big(S_i \cup (N_h/2 + S_i)\big).$$

This ordering of the basis functions yields a stiffness matrix as depicted in Figure 6.2b. Recalling Definition 5.4 we obtain

$$(C^i_{\mathbf{H}})_{\ell,m} = \big(\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}\phi_m, \psi_\ell\big)_\varepsilon = \big(\boldsymbol{\mathcal{C}}_{\mathbf{H}}(\chi_i\phi_m), \psi_\ell\big)_\varepsilon, \qquad (C^i_{\mathbf{E}})_{\ell,m} = \big(\boldsymbol{\mathcal{C}}^i_{\mathbf{E}}\psi_m, \phi_\ell\big)_\mu = \big(\boldsymbol{\mathcal{C}}_{\mathbf{E}}\psi_m, \chi_i\phi_\ell\big)_\mu,$$

and in combination with the upper decomposition of $C_{\mathbf{H}}$ and $C_{\mathbf{E}}$ and the convention about the face normal n_F for $F \in \mathcal{F}^{\text{int}}_{h,ci}$ (see Figure 5.7) we infer

$$\begin{aligned}(C^i_{\mathbf{H}})_{\ell,m} = &\sum_{K\in\mathcal{T}_{h,i}} \big(\operatorname{curl}\phi_m, \psi_\ell\big)_{\varepsilon,K} + \sum_{F\in\mathcal{F}^{\text{int}}_{h,i}} \big(n_F \times [\![\phi_m]\!]_F, \{\!\{\psi_\ell\}\!\}^{\varepsilon c}_F\big)_F \\ &+ \sum_{F\in\mathcal{F}^{\text{int}}_{h,ci}} \big((\phi_m)|_{K_i}, \{\!\{\psi_\ell\}\!\}^{\varepsilon c}_F\big)_F, \end{aligned} \tag{6.7a}$$

and

$$\begin{aligned}(C^i_{\mathbf{E}})_{\ell,m} = &\sum_{K\in\mathcal{T}_{h,i}} \big(\operatorname{curl}\psi_m, \phi_\ell\big)_{\mu,K} + \sum_{F\in\mathcal{F}^{\text{int}}_{h,i}} \big(n_F \times [\![\psi_m]\!]_F, \{\!\{\phi_\ell\}\!\}^{\mu c}_F\big)_F - \sum_{F\in\mathcal{F}^{\text{bnd}}_{h,i}} \big(n_F \times \psi_m, \phi_\ell\big)_F \\ &+ \sum_{F\in\mathcal{F}^{\text{int}}_{h,ci}} \mu_{K_i} c_{K_i} b_F \big(n_F \times [\![\psi_m]\!]_F, (\phi_\ell)|_{K_i}\big)_F. \end{aligned} \tag{6.7b}$$

This means we have

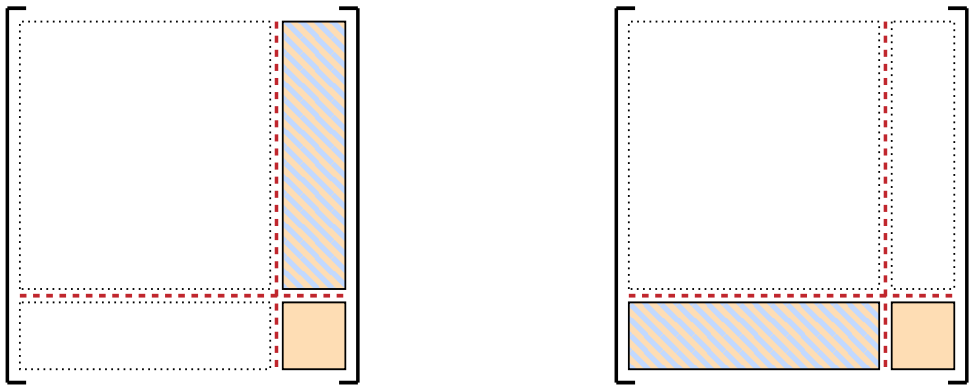
$$(C^i_{\mathbf{H}})_{\ell,m} = 0 \quad \text{for } m \in S_e, \qquad (C^i_{\mathbf{E}})_{\ell,m} = 0 \quad \text{for } \ell \in S_e.$$

We collect these results in Tables 6.3 and 6.4 and illustrate the structure of $C^i_{\mathbf{H}}$ and $C^i_{\mathbf{E}}$ in Figure 6.3. We observe the result of the different definitions of $\boldsymbol{\mathcal{C}}^i_{\mathbf{H}}$ and $\boldsymbol{\mathcal{C}}^i_{\mathbf{E}}$, compare (5.16a) with (5.16b): on the one hand $C^i_{\mathbf{H}}$ belongs to a splitting $C_{\mathbf{H}} = C^i_{\mathbf{H}} + C^e_{\mathbf{H}}$ w.r.t. columns of $C_{\mathbf{H}}$ and on the other hand $C^i_{\mathbf{E}}$ stems from a splitting $C_{\mathbf{E}} = C^i_{\mathbf{E}} + C^e_{\mathbf{E}}$ w.r.t. rows of $C_{\mathbf{E}}$.

Now, we can examine the left-hand sides $\widetilde{\boldsymbol{\mathcal{L}}}$ and $\boldsymbol{\mathcal{L}}$ of our linear systems. They read

$$\widetilde{L} = I + \frac{\tau^2}{4} M^{-1}_{\mathbf{E}} C^i_{\mathbf{H}} M^{-1}_{\mathbf{H}} C^i_{\mathbf{E}}, \qquad L_{\text{cf}} = I + \frac{\tau^2}{4} M^{-1}_{\mathbf{E}} C_{\mathbf{H}} M^{-1}_{\mathbf{H}} C_{\mathbf{E}}. \tag{6.8}$$

Since the mass matrices are block-diagonal they do not change the structure of the stiffness matrices but only can change the sparsity of the nonzero blocks. In Figure 6.4a we illustrate



(a) Structure of $C^i_{\mathbf{H}}$. (b) Structure of $C^i_{\mathbf{E}}$.

Figure 6.3: Structure of the implicit stiffness matrices $C^i_{\mathbf{H}}$ (left) and $C^i_{\mathbf{E}}$ (right) associated with the splitting of the spatial mesh into implicitly treated elements $\mathcal{T}_{h,i}$ and explicitly treated elements $\mathcal{T}_{h,e}$ and the corresponding sorting of the basis functions (6.6). Implicit element coupling: $\mathcal{T}_{h,i}$ to $\mathcal{T}_{h,i}$ (orange), explicit-implicit element coupling: $\mathcal{T}_{h,e}$ to $\mathcal{T}_{h,i}$ (blue-orange striped).

the structure of L_{cf} and in Figure 6.4b the structure of $\widetilde{L}$. Observe that we obtain the same pattern (although more sparse) for $\widetilde{L}$ than for L_{cf}. Consequently, we cannot deduce the gain in efficiency of the locally implicit schemes compared to the Crank–Nicolson method. The essential idea to recognize how the locally implicit scheme can be implemented more efficiently than the Crank–Nicolson method is an additional distinction of the basis functions, which we will discuss next.

We introduce the following partition of the set of explicitly treated mesh elements $\mathcal{T}_{h,e}$,

$$\begin{aligned}\mathcal{T}^e_{h,e} &= \{K_e \in \mathcal{T}_{h,e} \mid \forall K_i \in \mathcal{T}_{h,i} : \ \left|K_e \cap K_i\right|_{d-1} = 0\},\\ \mathcal{T}^i_{h,e} &= \{K_e \in \mathcal{T}_{h,e} \mid \exists K_i \in \mathcal{T}_{h,i} : \ \left|K_e \cap K_i\right|_{d-1} \neq 0\},\end{aligned}$$

i.e. we divide $\mathcal{T}_{h,e}$ into the set of explicitly treated elements which only have explicitly treated neighbors, and into the set of explicitly treated elements which possess at least one implicitly integrated neighbor. Note that $\mathcal{T}^i_{h,e} = \mathcal{T}_{h,ci}$, but we prefer to write $\mathcal{T}^i_{h,e}$ for a consistent notation. Analogously, we partition $\mathcal{T}_{h,i}$ into

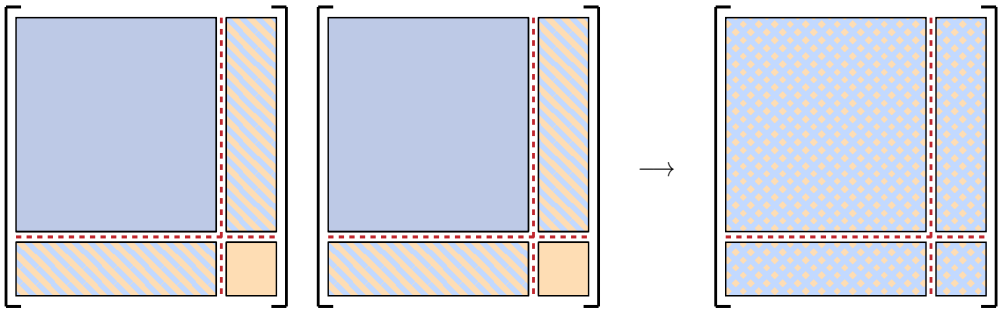
$$\begin{aligned}\mathcal{T}^e_{h,i} &= \{K_i \in \mathcal{T}_{h,i} \mid \exists K_e \in \mathcal{T}_{h,e} : \ \left|K_i \cap K_e\right|_{d-1} \neq 0\},\\ \mathcal{T}^i_{h,i} &= \{K_i \in \mathcal{T}_{h,i} \mid \forall K_e \in \mathcal{T}_{h,e} : \ \left|K_i \cap K_e\right|_{d-1} = 0\}.\end{aligned}$$

The first set contains all implicitly treated elements which only have implicitly integrated neighbors, whereas the second set collects the implicitly treated elements which exhibit at least one explicitly treated neighbor. An example for these sets is given in Figure 6.5a. We denote with

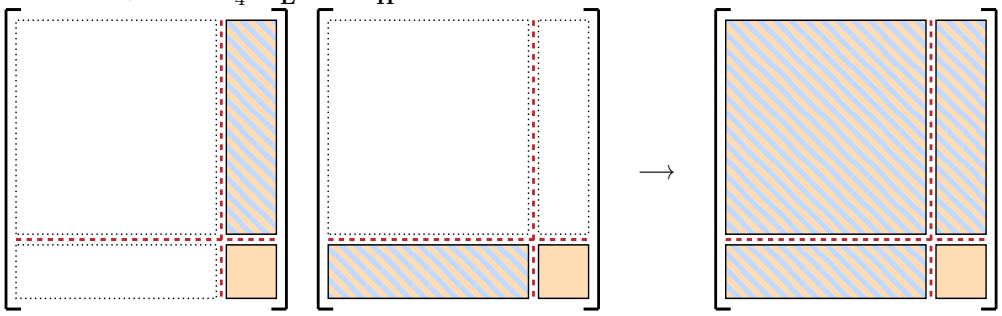
$$\begin{aligned}S^e_e &= \{1,\dots,N^e_e/2\}, & S^i_e &= \{N^e_e/2+1,\dots,N_e/2\},\\ S^e_i &= \{N_e/2+1,\dots,N_e/2+N^e_i/2\}, & S^i_i &= \{N_e/2+N^e_i/2+1,\dots,N_h/2\},\end{aligned}$$

where $N^e_e = n_h\mathrm{card}(\mathcal{T}^e_{h,e})$ and $N^e_i = n_h\mathrm{card}(\mathcal{T}^e_{h,i})$ denote the dof in $\mathcal{T}^e_{h,e}$ and in $\mathcal{T}^e_{h,i}$, respectively, and sort our basis functions by

$$\begin{pmatrix}\phi_\ell\\0\end{pmatrix}_{\ell\in S^e_e}, \begin{pmatrix}\phi_\ell\\0\end{pmatrix}_{\ell\in S^i_e}, \begin{pmatrix}\phi_\ell\\0\end{pmatrix}_{\ell\in S^e_i}, \begin{pmatrix}\phi_\ell\\0\end{pmatrix}_{\ell\in S^i_i}, \tag{6.9a}$$

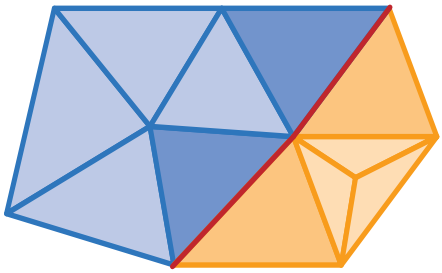


(a) Structure of $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}$, $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}$ and of the left-hand side of the linear system in the Crank–Nicolson method $L_{\mathrm{cf}} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1}C_{\mathbf{H}}M_{\mathbf{H}}^{-1}C_{\mathbf{E}}$.

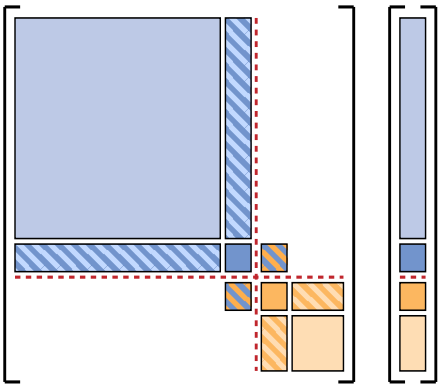


(b) Structure of $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}^{i}$, $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}^{i}$ and of the left-hand side of the linear system in the locally implicit method $\widetilde{L} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1}C_{\mathbf{H}}^{i}M_{\mathbf{H}}^{-1}C_{\mathbf{E}}^{i}$.

Figure 6.4: Structure of the left-hand side of the linear systems in case of the (central fluxes) Crank–Nicolson method and in case of the locally implicit methods.



(a) Splitting of the mesh elements into $\mathcal{T}_{h,e}^{e}$ (light blue), $\mathcal{T}_{h,e}^{i}$ (dark blue), $\mathcal{T}_{h,i}^{e}$ (dark orange) and $\mathcal{T}_{h,i}^{i}$ (light orange).



(b) Left: Structure of full stiffness matrices $C_{\mathbf{H}}$ and $C_{\mathbf{E}}$. Explicit element coupling: $\mathcal{T}_{h,e}^{e}$ to $\mathcal{T}_{h,e}^{e}$ (light blue), $\mathcal{T}_{h,e}^{i}$ to $\mathcal{T}_{h,e}^{i}$ (dark blue) and $\mathcal{T}_{h,e}^{e}$ to $\mathcal{T}_{h,e}^{i}$ (blue, striped). Implicit element coupling: $\mathcal{T}_{h,i}^{i}$ to $\mathcal{T}_{h,i}^{i}$ (light orange), $\mathcal{T}_{h,i}^{e}$ to $\mathcal{T}_{h,i}^{e}$ (dark orange) and $\mathcal{T}_{h,i}^{i}$ to $\mathcal{T}_{h,i}^{e}$ (orange, striped). Explicit-implicit element coupling: $\mathcal{T}_{h,e}^{i}$ to $\mathcal{T}_{h,i}^{e}$ (blue-orange striped). Right: Coefficient vector.

Figure 6.5: (a) Splitting of the mesh into explicitly integrated elements with only explicit neighbors and with at least one implicit neighbor, and implicitly treated elements with only implicit neighbors and with at least one explicit neighbor. (b) Stiffness matrix and coefficient vector corresponding to the ordering (6.9) of the basis.

	$m \in S_e^e$ $\operatorname{supp}(\chi_i\phi_m) \subset \mathcal{T}_{h,e}^e$	$m \in S_e^i$ $\operatorname{supp}(\chi_i\phi_m) \subset \mathcal{T}_{h,e}^i$	$m \in S_i^e$ $\operatorname{supp}(\chi_i\phi_m) \subset \mathcal{T}_{h,e}^i$	$m \in S_i^i$ $\operatorname{supp}(\chi_i\phi_m) \subset \mathcal{T}_{h,i}^i$
$\ell \in \frac{N_h}{2} + S_e^e$ $\operatorname{supp}\psi_\ell \subset \mathcal{T}_{h,e}^e$	0	0	0	0
$\ell \in \frac{N_h}{2} + S_e^i$ $\operatorname{supp}\psi_\ell \subset \mathcal{T}_{h,e}^i$	0	0	$*$	0
$\ell \in \frac{N_h}{2} + S_i^e$ $\operatorname{supp}\psi_\ell \subset \mathcal{T}_{h,i}^e$	0	0	$*$	$*$
$\ell \in \frac{N_h}{2} + S_i^i$ $\operatorname{supp}\psi_\ell \subset \mathcal{T}_{h,i}^i$	0	0	$*$	$*$

Table 6.5: Zero and possibly nonzero entries of the implicit stiffness matrix $(C_{\mathbf{H}}^i)_{\ell,m} = \left(\boldsymbol{\mathcal{C}}_{\mathbf{H}}^i \phi_m, \psi_\ell\right)_\varepsilon$.

	$m \in \frac{N_h}{2} + S_e^e$ $\operatorname{supp}\psi_m \subset \mathcal{T}_{h,e}^e$	$m \in \frac{N_h}{2} + S_e^i$ $\operatorname{supp}\psi_m \subset \mathcal{T}_{h,e}^i$	$m \in \frac{N_h}{2} + S_i^e$ $\operatorname{supp}\psi_m \subset \mathcal{T}_{h,e}^i$	$m \in \frac{N_h}{2} + S_i^i$ $\operatorname{supp}\psi_m \subset \mathcal{T}_{h,i}^i$
$\ell \in S_e^e$ $\operatorname{supp}(\chi_i\phi_\ell) \subset \mathcal{T}_{h,e}^e$	0	0	0	0
$\ell \in S_e^i$ $\operatorname{supp}(\chi_i\phi_\ell) \subset \mathcal{T}_{h,e}^i$	0	0	0	0
$\ell \in S_i^e$ $\operatorname{supp}(\chi_i\phi_\ell) \subset \mathcal{T}_{h,i}^e$	0	$*$	$*$	$*$
$\ell \in S_i^i$ $\operatorname{supp}(\chi_i\phi_\ell) \subset \mathcal{T}_{h,i}^i$	0	0	$*$	$*$

Table 6.6: Zero and possibly nonzero entries of the implicit stiffness matrix $(C_{\mathbf{E}}^i)_{\ell,m} = \left(\boldsymbol{\mathcal{C}}_{\mathbf{E}}^i \psi_m, \phi_\ell\right)_\mu$.

and

$$\begin{pmatrix} 0 \\ \psi_\ell \end{pmatrix}_{\ell \in N_h/2 + S_e^e}, \begin{pmatrix} 0 \\ \psi_\ell \end{pmatrix}_{\ell \in N_h/2 + S_e^i}, \begin{pmatrix} 0 \\ \psi_\ell \end{pmatrix}_{\ell \in N_h/2 + S_i^e}, \begin{pmatrix} 0 \\ \psi_\ell \end{pmatrix}_{\ell \in N_h/2 + S_i^i}. \tag{6.9b}$$

This ordering gives rise to stiffness matrices $C_{\mathbf{H}}$, $C_{\mathbf{E}}$ with a structure as depicted in Figure 6.5b. By (6.7) the implicit stiffness matrices satisfy

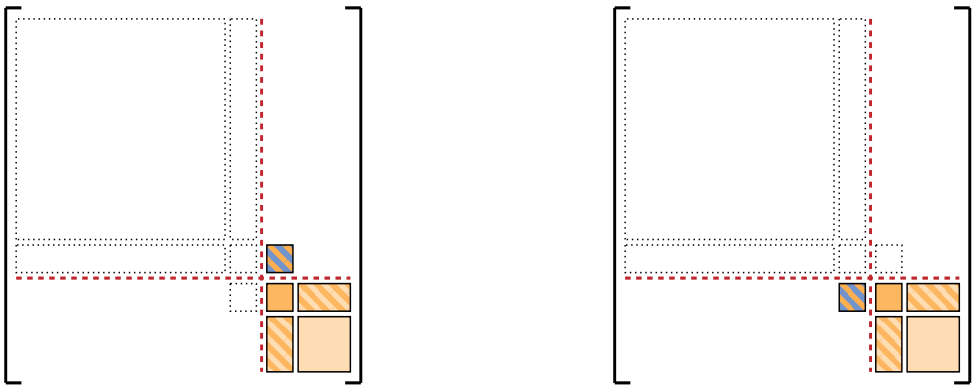
$$(C_{\mathbf{H}}^i)_{\ell,m} = 0 \text{ for } m \in (S_e^i \cup S_e^e), \qquad (C_{\mathbf{E}}^i)_{\ell,m} = 0 \text{ for } \ell \in (S_e^i \cup S_e^e).$$

Moreover, we have

$$(C_{\mathbf{H}}^i)_{\ell,m} = 0 \text{ for } m \in S_i^e,\ \ell \in (\frac{N_h}{2} + S_e^e), \qquad (C_{\mathbf{H}}^i)_{\ell,m} = 0 \text{ for } m \in S_i^i,\ \ell \in \big(\frac{N_h}{2} + (S_e^e \cup S_e^i)\big),$$
$$(C_{\mathbf{E}}^i)_{\ell,m} = 0 \text{ for } \ell \in S_i^e,\ m \in (\frac{N_h}{2} + S_e^e), \qquad (C_{\mathbf{E}}^i)_{\ell,m} = 0 \text{ for } \ell \in S_i^i,\ m \in \big(\frac{N_h}{2} + (S_e^e \cup S_e^i)\big),$$

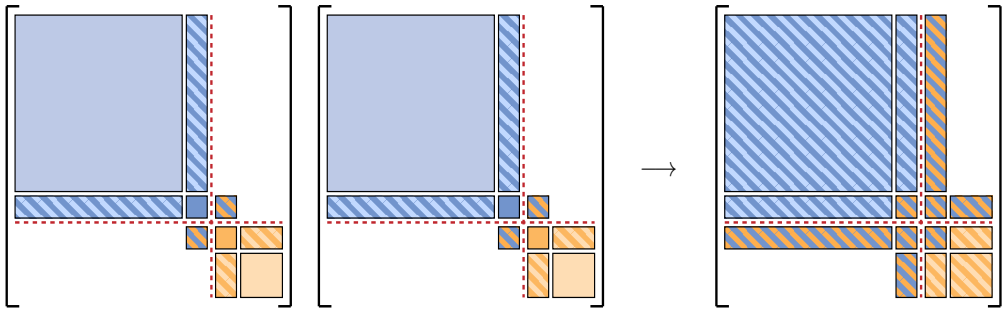
since basis functions with support in an element of $\mathcal{T}_{h,i}^e$ do not couple with basis functions whose support lie in an element of $\mathcal{T}_{h,e}^e$. Basis functions with support in a subset of an element of $\mathcal{T}_{h,i}^i$ do not couple with elements with support in an element of $\mathcal{T}_{h,e}^e \cup \mathcal{T}_{h,e}^i$. We collect this in Tables 6.5 and 6.6 and illustrate the structure of $C_{\mathbf{H}}^i$ and $C_{\mathbf{E}}^i$ in Figure 6.6.

Finally, we give the structure of L_{cf} and $\widetilde{L}$ in Figure 6.7. Observe that for the Crank–Nicolson method L_{cf} involves all dof of the **E**-field in the whole spatial mesh $\mathcal{T}_h$. In contrary, for the locally implicit method $\widetilde{L}$ exhibits the identity matrix on the dof in $\mathcal{T}_{h,e}^e$, i.e. we do **not**

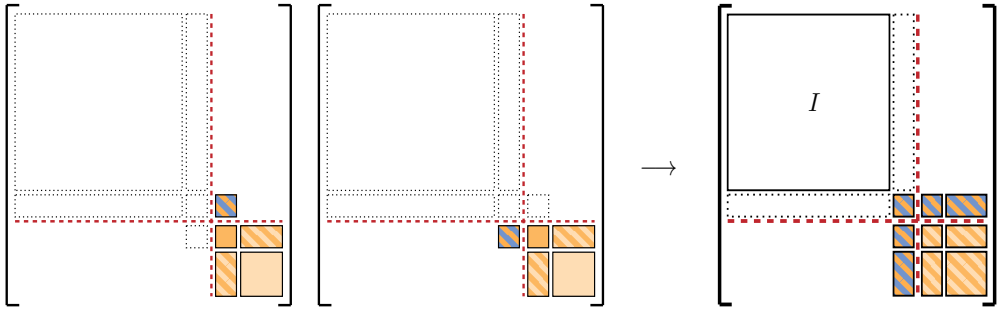


(a) Structure of $C^i_{\mathbf{H}}$. (b) Structure of $C^i_{\mathbf{E}}$.

Figure 6.6: Structure of the implicit stiffness matrices $C^i_{\mathbf{H}}$ (left) and $C^i_{\mathbf{E}}$ (right) associated with the splitting of the spatial mesh into $\mathcal{T}^i_{h,i}$, $\mathcal{T}^e_{h,i}$, $\mathcal{T}^i_{h,e}$, $\mathcal{T}^e_{h,e}$ and the corresponding sorting of the basis functions (6.9). Implicit element coupling: $\mathcal{T}^i_{h,i}$ to $\mathcal{T}^i_{h,i}$ (light orange), $\mathcal{T}^e_{h,i}$ to $\mathcal{T}^e_{h,i}$ (dark orange), $\mathcal{T}^i_{h,i}$ to $\mathcal{T}^i_{h,e}$ (orange striped). Explicit-implicit element coupling: $\mathcal{T}^i_{h,e}$ to $\mathcal{T}^e_{h,i}$ (blue-orange striped).

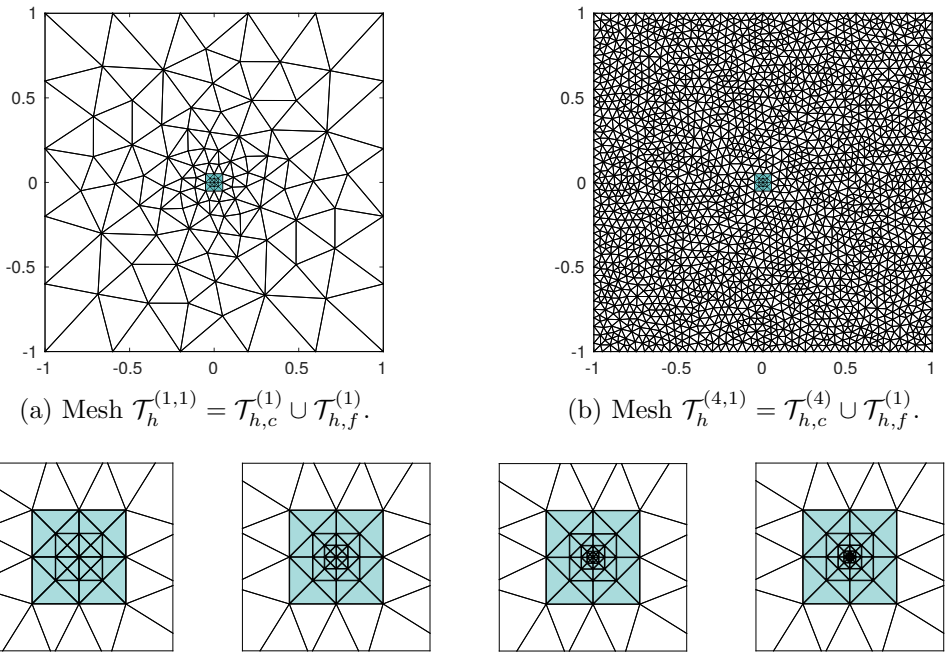


(a) Structure of $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}$, $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}$ and of the left-hand side of the linear system in the Crank–Nicolson method $L_{\mathrm{cf}} = I + \frac{\tau^2}{4}M_{\mathbf{E}}^{-1}C_{\mathbf{H}}M_{\mathbf{H}}^{-1}C_{\mathbf{E}}$.



(b) Structure of $M_{\mathbf{E}}^{-1}C^i_{\mathbf{H}}$, $M_{\mathbf{H}}^{-1}C^i_{\mathbf{E}}$ and of the left-hand side of the linear system in the locally implicit method $\widetilde{L} = I + \frac{\tau^2}{4}M_{\mathbf{E}}^{-1}C^i_{\mathbf{H}}M_{\mathbf{H}}^{-1}C^i_{\mathbf{E}}$.

Figure 6.7: Structure of the left-hand side of the linear system in case of the (central fluxes) Crank–Nicolson method and in case of the locally implicit methods. For the Crank–Nicolson method the left-hand side involves all dof in the spatial mesh, whereas for the locally implicit methods only the dof from the elements from $\mathcal{T}^i_{h,i} \cup \mathcal{T}^e_{h,i} \cup \mathcal{T}^i_{h,e} = \mathcal{T}_{h,i} \cup \mathcal{T}_{h,ci}$ enter in the left-hand side.



(a) Mesh $\mathcal{T}_h^{(1,1)} = \mathcal{T}_{h,c}^{(1)} \cup \mathcal{T}_{h,f}^{(1)}$.

(b) Mesh $\mathcal{T}_h^{(4,1)} = \mathcal{T}_{h,c}^{(4)} \cup \mathcal{T}_{h,f}^{(1)}$.

(c) From left to right $\mathcal{T}_{h,f}^{(1)}, \dots, \mathcal{T}_{h,f}^{(4)}$.

Figure 6.8: Illustration of the two types of mesh refinements yielding the grid $\mathcal{T}_h^{(j,\ell)} = \mathcal{T}_{h,c}^{(j)} \cup \mathcal{T}_{h,f}^{(\ell)}$. In the upper figures the coarse part of the mesh is refined whereas in the lower plots we refined the fine part.

have to solve a linear system on these elements. This means that the implicit part of the locally implicit scheme boils down to solving a linear system only on the dof of the **E**-field stemming from the mesh elements in $\mathcal{T}_{h,i}^i \cup \mathcal{T}_{h,i}^e \cup \mathcal{T}_{h,e}^i = \mathcal{T}_{h,i} \cup \mathcal{T}_{h,ci}$, i.e. a **linear system on the implicitly treated mesh elements and their neighbors**. By Definition 5.1 the set $\mathcal{T}_{h,i} \cup \mathcal{T}_{h,ci}$ consists of all fine elements in $\mathcal{T}_{h,f}$, their neighbors and the neighbors of the neighbors. Now, recalling that we are interested in locally refined meshes, i.e. meshes for which $\operatorname{card}(\mathcal{T}_{h,f}) \ll \operatorname{card}(\mathcal{T}_{h,c})$, we see that the cost for solving the linear system in the locally implicit method is far smaller than the cost for solving the linear system in the Crank–Nicolson method. We summarize this again by

$$\text{dof Crank–Nicolson} = \tfrac{1}{2}\mathbf{n_h}\operatorname{card}(\mathcal{T}_\mathbf{h}) \gg \tfrac{1}{2}\mathbf{n_h}\operatorname{card}(\mathcal{T}_{\mathbf{h,i}} \cup \mathcal{T}_{\mathbf{h,ci}}) = \text{dof locally implicit.}$$

6.3 Numerical results

In this last section we numerically examine the central fluxes and the upwind fluxes locally implicit schemes (5.11) and (5.65), respectively. For comparison we consider the Verlet methods (4.39) and (4.78), and the Crank–Nicolson schemes (4.40) and (4.76).

We illustrate the different aspects of our theoretical results with the help of three examples. As a first example we look at a rather hypothetical scenario that allows us to exactly control the mesh parameters and thus confirm the CFL conditions and the convergence rates of our locally implicit schemes. The second example is an (adapted) example from nanophotonics, cf. Busch et al. [2011], namely the ring resonator of Section 5.1. Besides the repeated confirmation of our convergence results we show with this example the efficient numerical implementation of

j	$h^{(j)}_{\max,c}$	$h^{(j)}_{\min,c}$	$h^{(j)}_{\min,c}/h^{(j-1)}_{\min,c}$
1	0.2384	0.0336	-
2	0.1248	0.0268	0.80
3	0.0721	0.0257	0.96
4	0.0370	0.0209	0.81

(a) Largest and smallest diameter of the elements in $\mathcal{T}^{(j)}_{h,c}$ and refinement factor.

ℓ	$h^{(\ell)}_{\max,f}$	$h^{(\ell)}_{\min,f}$	$h^{(\ell)}_{\min,f}/h^{(\ell-1)}_{\min,f}$
1	0.025	0.0125	-
2	0.025	0.00625	0.5
3	0.025	0.003125	0.5
4	0.025	0.0015625	0.5

(b) Largest and smallest diameter of the elements in $\mathcal{T}^{(\ell)}_{h,f}$ and refinement factor.

j	$h^{(j,\ell)}_{\max,e}$	$h^{(j,\ell)}_{\min,e}$
1	0.2384	0.0376
2	0.1248	0.0277
3	0.0721	0.0272
4	0.0370	0.0209

(c) Largest and smallest diameter of the elements in $\mathcal{T}^{(j,\ell)}_{h,e}$. Valid for all $\ell = 1, \dots, 4$.

j	$h^{(j,\ell)}_{\max,i}$
1	0.0372
2	0.0305
3	0.0333
4	0.0291

ℓ	$h^{(\ell)}_{\min,i}$
1	0.0125
2	0.00625
3	0.003125
4	0.0015625

(d) Largest and smallest diameter of the elements in $\mathcal{T}^{(j,\ell)}_{h,i}$. The left table is valid for all $\ell = 1, \dots, 4$, and the right table is valid for all $j = 1, \dots, 4$.

Table 6.7: Mesh parameters of $\mathcal{T}^{(j,\ell)}_h$. The sets $\mathcal{T}^{(j)}_{h,c}$, $\mathcal{T}^{(\ell)}_{h,f}$ collect the coarse and the fine mesh elements, and the sets $\mathcal{T}^{(j,\ell)}_{h,e}$, $\mathcal{T}^{(j,\ell)}_{h,i}$ collect the explicitly and the implicitly treated elements, respectively.

the locally implicit schemes. Moreover, the first example covers the case of inhomogeneous Maxwell's equations ($\mathbf{J} \neq 0$), whereas the second one treats consider the homogeneous problem ($\mathbf{J} \equiv 0$). In our third example we apply the locally implicit time integrators to a larger locally refined mesh (compared to the two previous examples) to show their ability to treat huge problems. The idea for this mesh is taken from Grote et al. [2015].

All our examples work with the 2D TM Maxwell's equations (1.17) and their implementation is carried out with an extended version of the `matlab` codes for the dG space discretization provided by Hesthaven and Warburton [2008]. The implementation of the dG space discretization and the locally implicit time integrators for the 3D Maxwell's equations (1.18) is beyond the scope of this thesis, which focus lies on the theoretical results. However, we mention that the realization of this implementation with the software package `deal.II` is ongoing work and extensive numerical experiments, in particular in comparison with explicit local time stepping methods, will be presented elsewhere.

Our mesh data is available upon request by software@waves.kit.edu.

6.3.1 Numerical example 1: Test scenario

As mentioned above, our goal of this first example is to examine the CFL condition and spatial and temporal convergence of the locally implicit methods. As in Section 3.7 we consider $\Omega = (-1, 1)^2$ with constant material coefficients $\mu, \varepsilon \equiv 1$ and the reference solution (3.56).

Mesh sequences

For our spatial discretization we consider the following family of grids: Our initial mesh is the grid $\mathcal{T}_h^{(1)}$ from Section 3.7, see Figure 3.2. The fine part $\mathcal{T}_{h,f}^{(1)}$ of this mesh consists of the elements in the green square $[-0.05, 0.05]^2$ and the coarse part $\mathcal{T}_{h,c}^{(1)}$ of the remaining elements in $[-1,1]^2 \setminus [-0.05, 0.05]^2$, see Figure 6.8a.

We refine our initial mesh in two different ways: In each refinement step, we either refine the coarse elements in $\mathcal{T}_{h,c}^{(1)}$ or the fine elements in $\mathcal{T}_{h,f}^{(1)}$. We denote the resulting meshes by $\mathcal{T}_{h,c}^{(j)}$ and $\mathcal{T}_{h,f}^{(\ell)}$, respectively, where the parameters j and ℓ refer to the number of refinements. We denote by $\mathcal{T}_h^{(j,\ell)}$ the complete mesh composed of $\mathcal{T}_{h,c}^{(j)}$ and $\mathcal{T}_{h,f}^{(\ell)}$. In Figure 6.8b we plotted the mesh $\mathcal{T}_h^{(4,1)}$, and in Figure 6.8c the meshes $\mathcal{T}_{h,f}^{(\ell)}$, $\ell = 1, \dots, 4$. By Definition 5.1 we treat the elements in $\mathcal{T}_{h,f}^{(\ell)}$ and their neighbors (wich are elements of $\mathcal{T}_{h,c}^{(j)}$) implicitly and all remaining elements explicitly. We call the respective sets $\mathcal{T}_{h,i}^{(j,\ell)}$ and $\mathcal{T}_{h,e}^{(j,\ell)}$. Moreover, we denote by

$$
\begin{aligned}
h_{\min,c}^{(j)} &= \min_{K \in \mathcal{T}_{h,c}^{(j)}} h_K, & h_{\max,c}^{(j)} &= \max_{K \in \mathcal{T}_{h,c}^{(j)}} h_K, \\
h_{\min,f}^{(\ell)} &= \min_{K \in \mathcal{T}_{h,f}^{(\ell)}} h_K, & h_{\max,f}^{(\ell)} &= \max_{K \in \mathcal{T}_{h,f}^{(\ell)}} h_K, \\
h_{\min,b}^{(j,\ell)} &= \min_{K \in \mathcal{T}_{h,b}^{(j,\ell)}} h_K, & h_{\max,b}^{(j,\ell)} &= \max_{K \in \mathcal{T}_{h,b}^{(j,\ell)}} h_K, & b &\in \{e, i\},
\end{aligned}
$$

the diameter of the smallest and of the largest element in $\mathcal{T}_{h,c}^{(j)}$, $\mathcal{T}_{h,f}^{(\ell)}$, $\mathcal{T}_{h,e}^{(j,\ell)}$ and $\mathcal{T}_{h,i}^{(j,\ell)}$, respectively. In Table 6.7 we collect these mesh parameters as well as the refinement factors of the diameters (when changing from one mesh level to the next one).

Remark 6.1. Note that the mesh sequence $\mathcal{T}_h^{(j,j)}$ corresponds to the sequence $\mathcal{T}_h^{(j)}$ from Section 3.7. The mesh sequences $\mathcal{T}_{h,\mathrm{CFL}}^{(\ell)}$ and $\mathcal{T}_{h,\tau}^{(j)}$ agree with $\mathcal{T}_h^{(1,\ell)}$ and $\mathcal{T}_h^{(j,1)}$, respectively.

CFL condition

We begin with the validation of our theoretical results by examining the CFL condition. For our locally implicit schemes we are interested in confirming two points: First, that the CFL condition is independent of the fine part of the mesh, see (5.40) and (5.93a), and second that a larger stabilization parameter α induces a stricter CFL condition, see (5.93b).

In view of the first point, we ran our numerical experiment with all meshes $\mathcal{T}_h^{(j,\ell)}$, $j, \ell = 1, \dots 4$, polynomial degree $k = 2$, final time $T = t_{N_T} = 1$ and decreased the time-step size τ until the numerical solution became stable. We denote this time step with $\tau_{\max}^{(j,\ell)}$ and give its values in Tables 6.8 and 6.9. We clearly confirm that for both the central fluxes and the upwind fluxes locally implicit method the maximum stable time-step size does not depend on the fine mesh level, i.e. $\tau_{\max}^{(j,\ell)}$ in Tables 6.8a and 6.9a is independent of the fine mesh level ℓ. On the other hand, the coarse mesh elements do enter in the CFL condition, which is seen in the decreasing of $\tau_{\max}^{(j,\ell)}$ when we refine the coarse mesh level j. We observe that the factor of which the maximum stable time-step size reduces matches well the refinement factor of the coarse mesh elements given in Table 6.7a. Moreover, we observe that central fluxes locally implicit method possesses a less strict CFL condition than the upwind fluxes locally implicit scheme. This stems from the explicit time integration of the stabilization operators in the upwind fluxes scheme which

$\tau_{\max}^{(j,l)}$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\tau_{\max}^{(j,\ell)}/\tau_{\max}^{(j-1,\ell)}$
$j = 1$	0.0120	0.0120	0.0120	0.0120	-
$j = 2$	0.0088	0.0088	0.0088	0.0088	0.73
$j = 3$	0.0072	0.0072	0.0072	0.0072	0.82
$j = 4$	0.0062	0.0062	0.0062	0.0062	0.86

(a) Central fluxes locally implicit.

$\tau_{\max}^{(j,\ell)}$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$
$j = 1$	0.00272	0.00138	0.000688	0.000336
$j = 2$	0.00272	0.00138	0.000688	0.000336
$j = 3$	0.00272	0.00138	0.000688	0.000336
$j = 4$	0.00272	0.00138	0.000688	0.000336
$\tau_{\max}^{(j,\ell)}/\tau_{\max}^{(j,\ell-1)}$	-	0.51	0.5	0.49

(b) Central fluxes Verlet.

Table 6.8: Largest stable time steps $\tau_{\max}^{(j,\ell)}$ for the mesh $\mathcal{T}_h^{(j,\ell)}$ and ratio of largest time steps. We used a central fluxes space discretization with polynomial degree $k = 2$. The final time was $T = t_{N_T} = 1$.

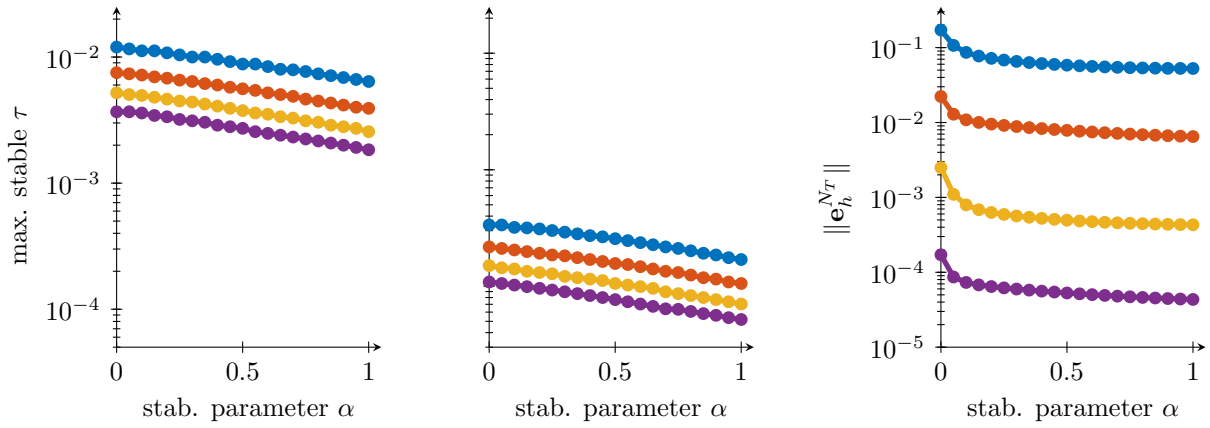
$\tau_{\max}^{(j,\ell)}$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\tau_{\max}^{(j,\ell)}/\tau_{\max}^{(j-1,\ell)}$
$j = 1$	0.00648	0.00648	0.00648	0.00648	-
$j = 2$	0.00448	0.00448	0.00448	0.00448	0.69
$j = 3$	0.00360	0.00360	0.00360	0.00360	0.8
$j = 4$	0.00320	0.00320	0.00320	0.00320	0.89

(a) Locally implicit upwind fluxes.

$\tau_{\max}^{(j,\ell)}$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$
$j = 1$	0.00141	0.000688	0.000336	0.000170
$j = 2$	0.00139	0.000688	0.000336	0.000170
$j = 3$	0.00139	0.000688	0.000336	0.000170
$j = 4$	0.00139	0.000688	0.000336	0.000170
$\tau_{\max}^{(j,\ell)}/\tau_{\max}^{(j,\ell-1)}$	-	0.49	0.49	0.51

(b) Verlet upwind fluxes.

Table 6.9: Largest stable time steps $\tau_{\max}^{(j,\ell)}$ for the mesh $\mathcal{T}_h^{(j,\ell)}$ and ratio of largest time steps. We used an upwind fluxes ($\alpha = 1$) space discretization with polynomial degree $k = 2$. The final time was $T = t_{N_T} = 1$.



(a) Maximum stable time step τ for the locally implicit method.

(b) Maximum stable time step τ for the Verlet method.

(c) Error of the locally implicit method with time step $\tau = 10^{-4}$.

Figure 6.9: Dependence of the maximum stable time step and of the error on the stabilization parameter α. We used the mesh $\mathcal{T}_h^{(1,4)}$, the final time $T = t_{N_T} = 1$ and polynomial degrees $k = 2$, $k = 3$, $k = 4$ and $k = 5$.

additionally enters the CFL condition. This can be seen by the condition $\widetilde{\theta}^2 < 1$, which is needed for the central fluxes locally implicit scheme, whereas the upwind fluxes scheme requires $\widetilde{\theta}^2 + \alpha\widetilde{\theta} < 1$, see (5.40) and (5.93). Last, we give as comparison in Tables 6.8b and 6.9b the maximum stable time steps for the central fluxes and for the upwind fluxes Verlet method, respectively. We see that, in contrary to the locally implicit methods, the CFL condition of the Verlet methods does depend on the fine mesh levels. We also observe that the reduction factor of the maximum stable time-step size matches the refinement factor of the fine mesh elements, see Table 6.7a.

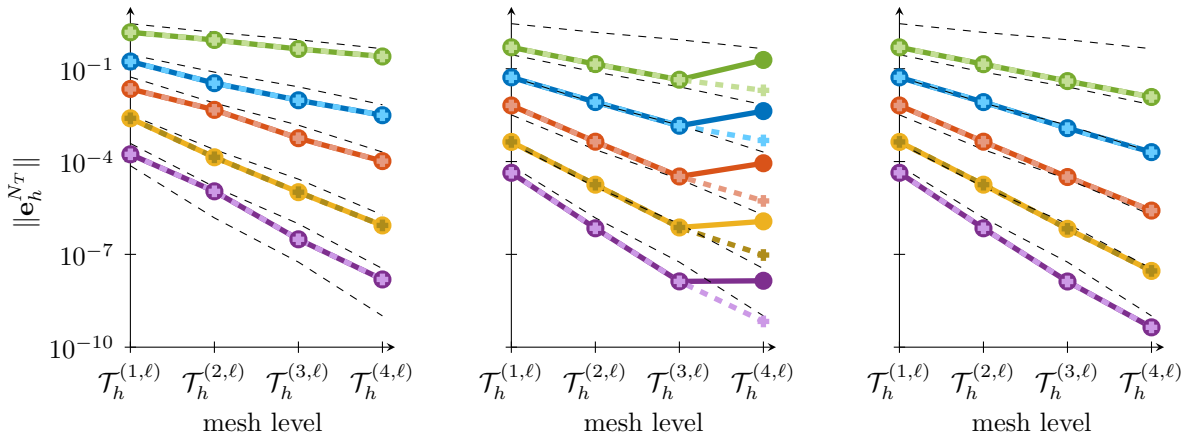
Next, we investigate the dependence of the maximum stable time step on the stabilization parameter $\alpha \in [0, 1]$. In Figures 6.9a and 6.9b we give the maximum stable time-step size we observe in our numerical experiments in dependence on α for the locally implicit method and for the Verlet method. We again confirm that the central fluxes schemes posses the largest maximum stable time step. Moreover, we validate that a larger stabilization parameter α leads to a smaller maximum stable time-step size (i.e. a stricter CFL condition) as predicted by (5.93). This might indicate that the full upwind choice $\alpha = 1$ is not be the best choice. However, we emphasize that the error also depends on α since the error constant scales with $(1 + \alpha^2)/\alpha$, see Theorem 5.35 (and also Theorems 4.28, 3.13 and Figures 3.1, 3.5 for the semidiscrete case and the fully discrete case with the Crank–Nicolson scheme). This is illustrated in Figure 6.9c where we give the error $\mathbf{e}_h^{N_T} = \mathbf{u}_h^{N_T} - \pi_h \mathbf{u}(T)$ at the final time $T = t_{N_T} = 1$ measured in the L^2-norm $\|\cdot\|$ in dependence of the stabilization parameter α. Note that in the here considered case $\mu, \varepsilon \equiv 1$ we have that $\|\cdot\|_{\mu\times\varepsilon} = \|\cdot\|$. We observe that the choice $\alpha = 1$ yields the smallest error. So, we have to carefully choose $\alpha \in [0, 1]$ in order to balance the CFL condition and the error size.

Spatial convergence

Our next aim is to validate the spatial convergence rates

$$\max_{K\in\mathcal{T}_h} h_K^k \qquad \text{and} \qquad \max_{K\in\mathcal{T}_{h,e}} h_K^{k+1/2} + \max_{K\in\mathcal{T}_{h,i}} h_K^k$$

proven in Theorems 5.13 and 5.35 for the central fluxes and for the upwind fluxes locally



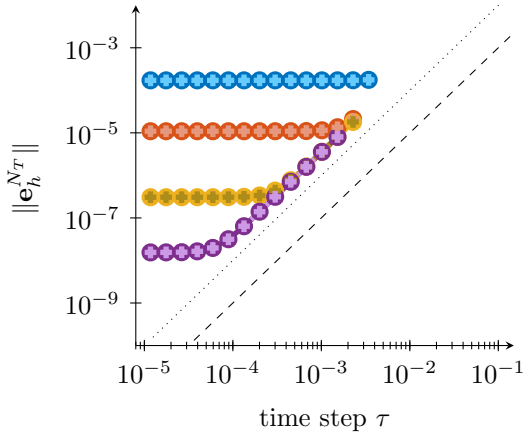
(a) Central fluxes locally implicit. (b) Upwind fluxes locally implicit. (c) Upwind fluxes Verlet.

Figure 6.10: Spatial convergence. We used the final time $T = t_{N_T} = 1$ and the time step $\tau = 10^{-5}$. We employed the polynomial degrees $k = 1$, $k = 2$, $k = 3$, $k = 4$ and $k = 5$. For the solid lines with $\bullet$ markers we used the fine mesh level $\ell = 1$ and for the dashed lines with $+$ markers we used $\ell = 4$. The black dashed lines have slope h^k for $k = 1, \dots, 6$.

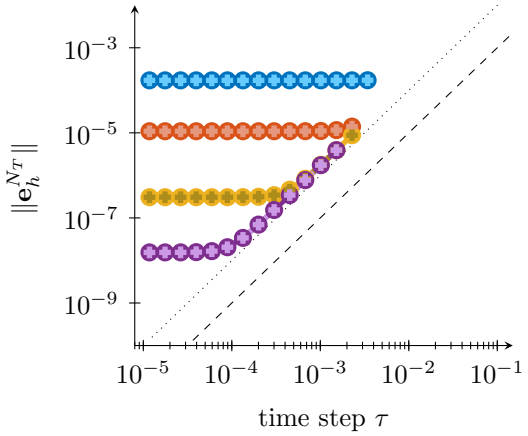
implicit method, respectively. For this purpose we ran our simulation with all coarse mesh levels $j = 1, \dots, 4$, two different fine levels $\ell = 1, 4$ and different polynomial degrees k until the final time $T = t_{N_T} = 1$. We used the small time step $\tau = 10^{-5}$ such that the spatial error dominates over the temporal error. We give the resulting error $\mathbf{e}_h^{N_T} = \mathbf{u}_h^{N_T} - \pi_h \mathbf{u}(t_{N_T})$ measured in the L^2-norm in Figure 6.10 for the central fluxes locally implicit scheme, the upwind fluxes ($\alpha = 1$) locally implicit scheme and the upwind fluxes ($\alpha = 1$) Verlet method. The first figure confirms the spatial convergence rate of the central fluxes locally implicit scheme when the coarse grid $\mathcal{T}_{h,c}$ is refined. Moreover, we do not observe a decrease of the error when the fine elements in $\mathcal{T}_{h,f}$ are refined (the solid and the dashed lines in Figure 6.10a coincide). This is plausible because the contribution of the few small elements in $\mathcal{T}_{h,f}$ to the total error is negligible. In Figure 6.10b we observe that the spatial error of the upwind fluxes locally implicit scheme decreases with order h^{k+1} for the mesh sequences $\mathcal{T}_h^{(1,1)}, \dots, \mathcal{T}_h^{(3,1)}$ and $\mathcal{T}_h^{(1,4)}, \dots \mathcal{T}_h^{(4,4)}$. This confirms the convergence rate $\max_{K \in \mathcal{T}_{h,e}} h_K^{k+1/2}$ (we even get the better rate $k+1$) because for these meshes the error stemming from the explicitly treated elements are dominant over the error arising from the implicitly treated elements (and decreasing only with order k), see the mesh element sizes in Tables 6.7c and 6.7d. However, for the mesh $\mathcal{T}_h^{(4,1)}$, the elements in the explicitly treated set and the larger elements in the implicitly treated set are of the same size. As a consequence we observe the rate $\max_{K \in \mathcal{T}_{h,i}} h_K^k$ (stemming from the unstabilized implicitly integrated part of the mesh) which spoils the upwind fluxes rate $k + 1/2$. However, we point out that the mesh $\mathcal{T}_h^{(4,1)}$ is **not** a locally refined mesh, see Figure 6.8b, and thus a locally implicit time integrator is not appropriate. In contrary, the meshes $\mathcal{T}_h^{(j,4)}$ are locally refined and we observe that the upwind fluxes locally implicit scheme works very well. This is emphasized by comparing it to the error of the fully stabilized upwind fluxes Verlet method given in Figure 6.10c. We see that the locally implicit scheme and the Verlet method yield an error with the same accuracy (for the locally refined meshes $\mathcal{T}_h^{(j,4)}$).

Temporal convergence

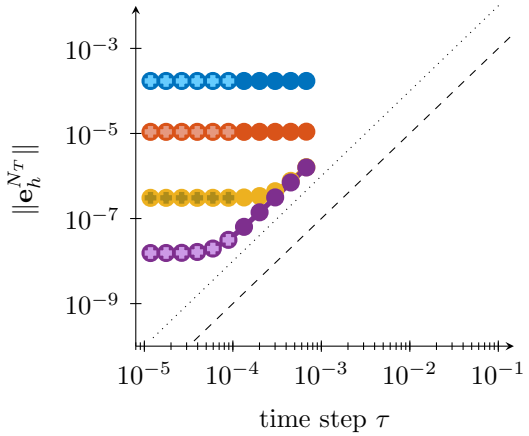
Finally, we confirm the temporal convergence of our locally implicit methods. We employ the polynomial degree $k = 5$ in our dG space discretization so that (at least for larger time-step



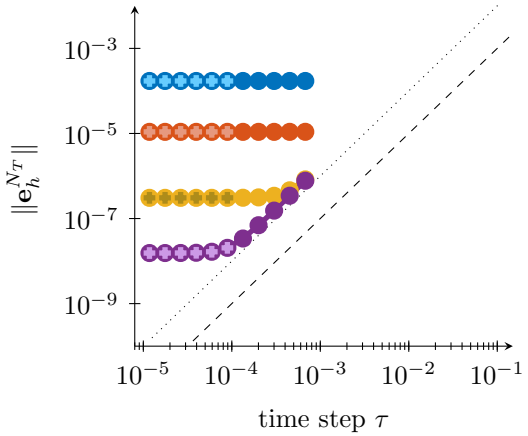
(a) Locally implicit method with treatment of source term analog to the Crank–Nicolson method.



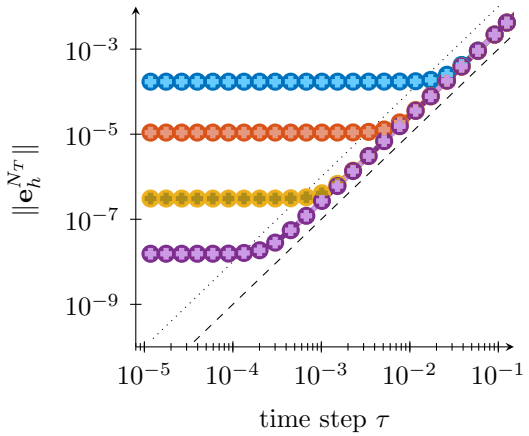
(b) Locally implicit method with treatment of source term analog to the implicit midpoint method.



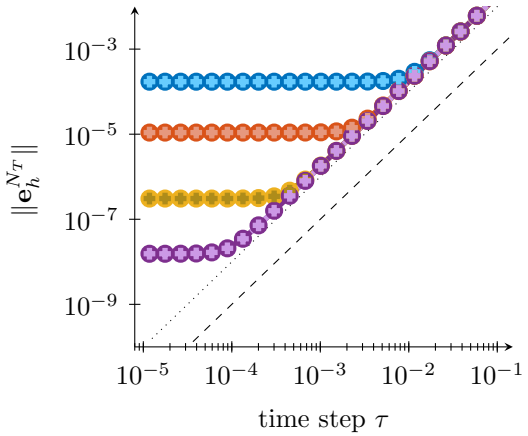
(c) Verlet method with treatment of source term analog to the Crank–Nicolson method.



(d) Verlet method with treatment of source term analog to the implicit midpoint method.

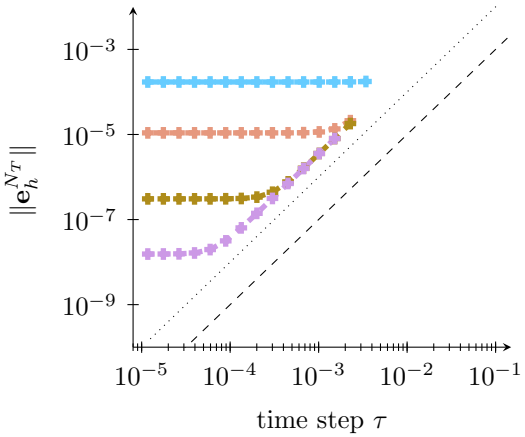


(e) Crank–Nicolson method.

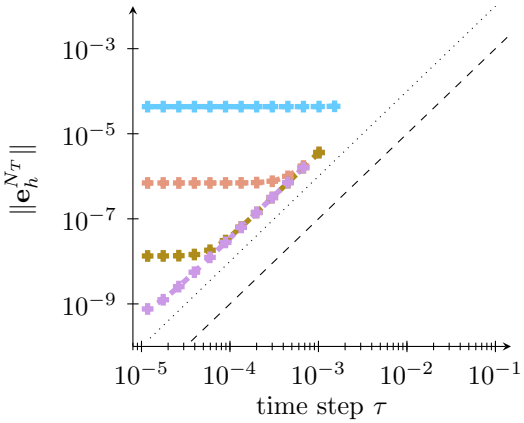


(f) Implicit midpoint method.

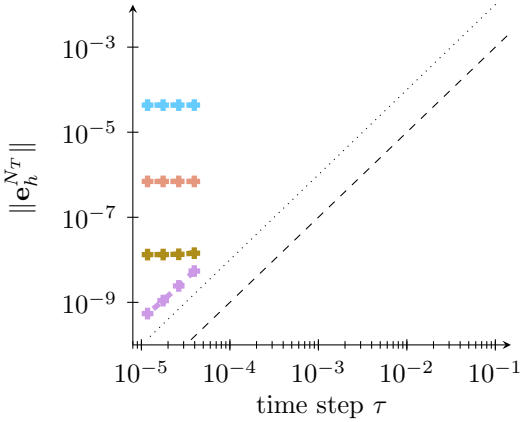
Figure 6.11: Temporal convergence. For the space discretization we employed the central fluxes method with polynomial degree $k = 5$ and used the meshes $\mathcal{T}_h^{(1,\ell)}$, $\mathcal{T}_h^{(2,\ell)}$, $\mathcal{T}_h^{(3,\ell)}$ and $\mathcal{T}_h^{(4,\ell)}$. For the solid lines with $\bullet$ markers we chose the fine mesh level $\ell = 1$ and for the dashed lines with $+$ markers $\ell = 4$. The black dotted line has slope τ^2 and the black dashed line has slope $1/10\tau^2$. The final time was $T = t_{N_T} = 1$



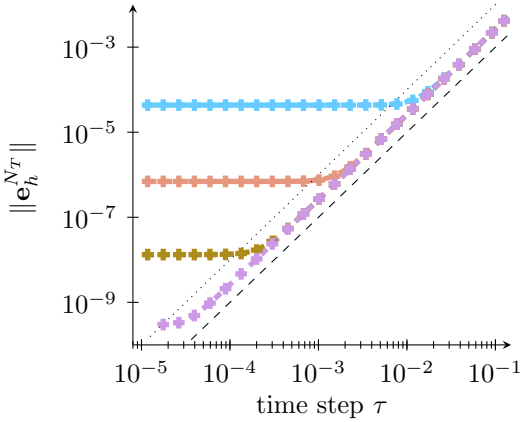
(a) Central fluxes locally implicit, see Figure 6.11a.



(b) Upwind fluxes locally implicit ($\alpha = 1$).



(c) Upwind fluxes Verlet ($\alpha = 1$).



(d) Upwind fluxes Crank–Nicolson ($\alpha = 1$).

Figure 6.12: Temporal convergence. For the space discretization we employed polynomial degree $k = 5$ and used the meshes $\mathcal{T}_h^{(1,4)}$, $\mathcal{T}_h^{(2,4)}$, $\mathcal{T}_h^{(3,4)}$ and $\mathcal{T}_h^{(4,4)}$. The black dotted line has slope τ^2 and the black dashed line has slope $1/10\tau^2$.. The final time was $T = t_{N_T} = 1$

sizes) the time integration error dominates over the space discretization error.

We start with a central fluxes space discretization and the associated time integration schemes. In Figure 6.11a we give the L^2-norm of the error $\mathbf{e}_h^{N_T}$ of the locally implicit scheme at the final time $T = t_{N_T} = 1$ for the meshes $\mathcal{T}_h^{(1,\ell)}, \dots, \mathcal{T}_h^{(4,\ell)}$, with fine mesh levels $\ell = 1, 4$. We only plotted the errors for time-step sizes that yield a stable numerical solution. We clearly observe that our locally implicit method converges with order two in the time-step size, which illustrates the convergence result of Theorem 5.13. Moreover, we see that the mesh levels do not influence the temporal convergence (all errors decrease with the same rate of around τ^2). This confirms that the error constant in Theorem 5.13 does not depend on the spatial mesh and thus our convergence result does not deteriorate if the mesh width h goes to zero. As comparison we give in Figures 6.11c and 6.11e the errors of the Verlet and of the Crank–Nicolson method. In the first figure we observe again the stricter CFL condition of the Verlet method and that both the locally implicit method and the Verlet method converge with the same error constants. In contrary, we deduce from Figure 6.11e that the Crank–Nicolson method enjoys a slightly better error constant. In Sections 4.4 and 5.5 we discussed the implicit midpoint time integrator and the locally implicit scheme based on the implicit midpoint method instead of the Crank–Nicolson method. In Figures 6.11b and 6.11f we give the plots of the errors of these two methods. We see that

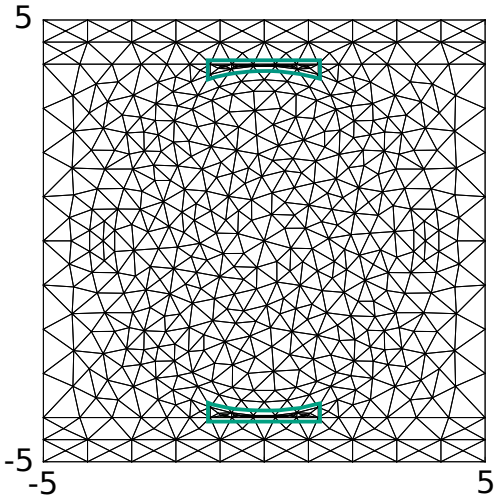


Figure 6.13: Mesh $\mathcal{T}_h$ of the ring resonator. The elements in the green marked area belong to the fine set $\mathcal{T}_h^f$ and all remaining elements are assigned to the coarse set $\mathcal{T}_h^c$.

both converge with order two in the time-step size. Comparing the implicit midpoint locally implicit scheme with the (Crank–Nicolson) locally implicit scheme we see that the former has a slightly better error constant. On the other hand, the implicit midpoint method has a larger error constant than the Crank–Nicolson scheme. For completeness, we give in Figure 6.11d a modified Verlet method which emanates from the original Verlet method (4.39) if we replace the "Crank–Nicolson" treatment of the source terms, i.e. $\frac{\tau}{2}(\mathbf{J}_h^{n+1} + \mathbf{J}_h^n)$, with the "implicit midpoint" treatment $\tau \mathbf{J}_h^{n+1/2}$. The result is the same as for the locally implicit method, namely the error constant slightly improves.

We end this subsection by considering an upwind fluxes space discretization with stabilization parameter $\alpha = 1$. The polynomial degree is again $k = 5$ and the final time is $T = t_{N_T} = 1$. In Figure 6.12b we plotted the error of the upwind fluxes locally implicit scheme for the meshes $\mathcal{T}_h^{(1,4)}, \dots, \mathcal{T}_h^{(4,4)}$. In order to relate the results we give in Figures 6.11a, 6.12c and 6.12d the errors of the central fluxes locally implicit, of the upwind fluxes Verlet and of the Crank–Nicolson method, respectively. First of all, we confirm the temporal convergence result of Theorem 5.13, namely that the upwind fluxes locally implicit method is of order two in the time step. Comparing with the errors of the central fluxes method we again observe the improved spatial convergence rate – the plateaus of the error lines indicating the spatial error are on smaller values for the upwind fluxes locally implicit method than for the central fluxes locally implicit method. Moreover, we see that the integration of the explicit stabilization operator does not spoil the temporal convergence. By comparing the errors of the upwind fluxes locally implicit method with the errors of the upwind fluxes Verlet method we observe that they both converge with the same temporal order and that the Verlet method has a more severe CFL condition.

6.3.2 Numerical example 2: ring resonator

In our second example we consider the ring resonator of Figure 5.4 in the domain $\Omega = (-5, 5)^2$. A crucial difference for this section is that we assume that the entire domain is covered in vacuum, i.e. we have $\mu, \varepsilon \equiv 1$. This is based on two reasons. On the one hand we are mostly interested in the effects of the spatial mesh on our locally implicit time integrators and on the other hand we prefer to have an exact solution available. In vacuum such an exact solution is

$r_{c,\max}$	$r_{c,\min}$	$r_{f,\max}$	$r_{f,\min}$
0.2545	0.0640	0.0424	0.0058

(a) Mesh parameters of $\mathcal{T}_h^c$ and $\mathcal{T}_h^f$.

$r_{e,\max}$	$r_{e,\min}$	$r_{i,\max}$	$r_{i,\min}$
0.2545	0.0640	0.1527	0.0058

(b) Mesh parameters of $\mathcal{T}_h^e$ and $\mathcal{T}_h^i$.

$\widehat{r}_{e,\max}$	$\widehat{r}_{e,\min}$	$\widehat{r}_{i,\max}$	$\widehat{r}_{i,\min}$
0.2545	0.0640	0.0424	0.0058

(c) Mesh parameters of $\widehat{\mathcal{T}}_h^e$ and $\widehat{\mathcal{T}}_h^i$.

Table 6.10: Mesh parameters of the ring resonator mesh and its decomposition: largest and smallest inner radius.

given by the cavity solution $\mathbf{u} = (\mathbf{H}_x, \mathbf{H}_y, \mathbf{E}_z)$,

$$\begin{aligned} \mathbf{H}_x(t) &= -\frac{\pi}{5\omega}\sin(\frac{\pi}{5}x)\cos(\frac{\pi}{5}y)\sin(\omega t), \\ \mathbf{H}_y(t) &= \frac{\pi}{5\omega}\cos(\frac{\pi}{5}x)\sin(\frac{\pi}{5}y)\sin(\omega t), \\ \mathbf{E}_z(t) &= \sin(\frac{\pi}{5}x)\sin(\frac{\pi}{5}y)\cos(\omega t), \end{aligned} \tag{6.10}$$

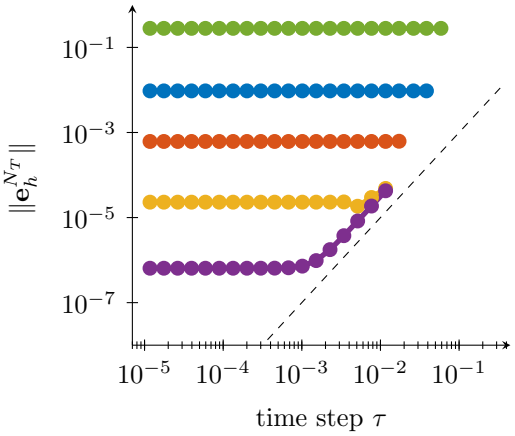
where $\omega = \sqrt{2}\pi/5$. This cavity solution satisfies the homogeneous Maxwell's equation (1.17), i.e. with source term $\mathbf{J}_z \equiv 0$ [Hesthaven and Warburton, 2008, Section 6.5].

Mesh

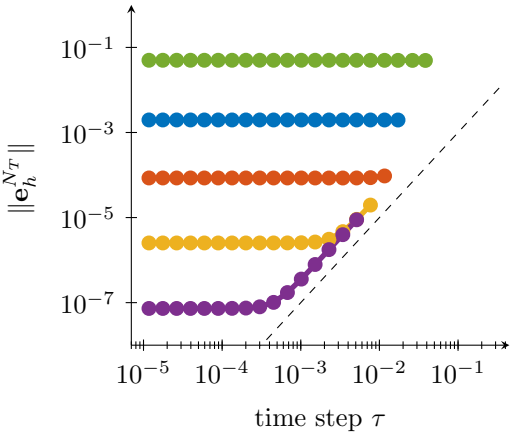
As usual we denote the mesh of the ring resonator with $\mathcal{T}_h$. In Figure 6.13 we give a plot of the spatial grid where the mesh elements in the green marked regions are assigned to the fine set $\mathcal{T}_h^f$ and all remaining elements belong to the coarse set $\mathcal{T}_h^c$. Due to the particular form of the mesh elements in the gap between the ring resonator and the wave guides, which are long but flat, we decided to determine the fine set by the inner radius (i.e. the radius of the largest ball inscribed in a mesh element) of the mesh elements and not by the diamater. Following Definition 5.1 we treat the elements in $\mathcal{T}_h^f$ and their neighbors implicitly and all other elements explicitly. Let us denote these sets with $\mathcal{T}_h^i$ and $\mathcal{T}_h^e$, respectively. In order to show the effect if the neighbors of the fine elements are **not** included into the set of implicitly treated elements, we also consider the choices $\widehat{\mathcal{T}}_h^i = \mathcal{T}_h^f$ and $\widehat{\mathcal{T}}_h^e = \mathcal{T}_h^c$. In Table 6.10 we give the associated mesh parameters, where we denote by $r_{b,\max}$ and $r_{b,\min}$ the largest and the smallest inner radius of $\mathcal{T}_h^b$, $b \in \{c, f\}$, and analog for $\widehat{\mathcal{T}}_h^b$.

Convergence and CFL condition

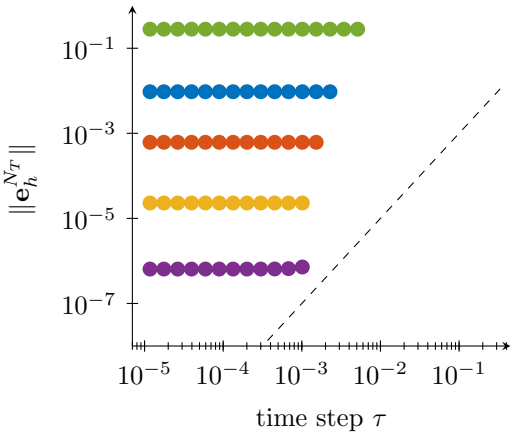
We evaluate the quality of our locally implicit schemes for this example by using different polynomial degrees k in the space discretization and running the simulation with different time-step sizes τ until the final time $T = t_{N_T} = 1$. The L^2-norm of the resulting errors is given in Figure 6.14 for the locally implicit method combined with a central fluxes space discretization and with an upwind fluxes space discretization with stabilization parameter $\alpha = 1$. Moreover, we provide in this figure the errors of the Verlet and of the Crank–Nicolson method. First of all, we confirm the temporal convergence order two for our locally implicit schemes and that they converge with the same rate as the Crank–Nicolson method. By comparing the plateaus of the error lines of Figures 6.14a and 6.14b, which indicate the space discretization error, we



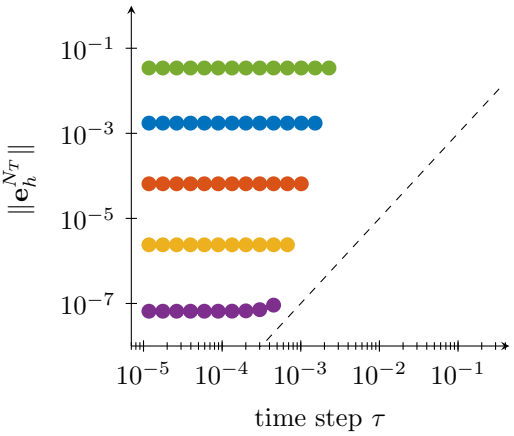
(a) Central fluxes locally implicit.



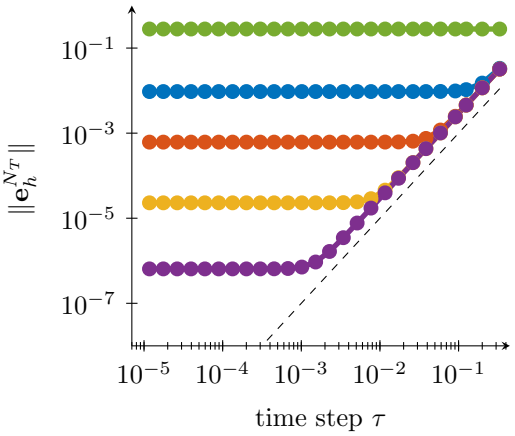
(b) Upwind fluxes locally implicit ($\alpha = 1$).



(c) Central fluxes Verlet.



(d) Upwind fluxes Verlet ($\alpha = 1$).



(e) Central fluxes Crank–Nicolson.

Figure 6.14: Temporal convergence. The implicit and explicit time integration of the mesh elements in $\mathcal{T}_h$ of the locally implicit schemes are based on the sets $\mathcal{T}_h^i$ and $\mathcal{T}_h^e$. For the space discretization we employed polynomial degrees $k = 1$, $k = 2$, $k = 3$, $k = 4$ and $k = 5$. The black dashed line has slope $1/10\tau^2$. The final time was $T = t_{N_T} = 1$.

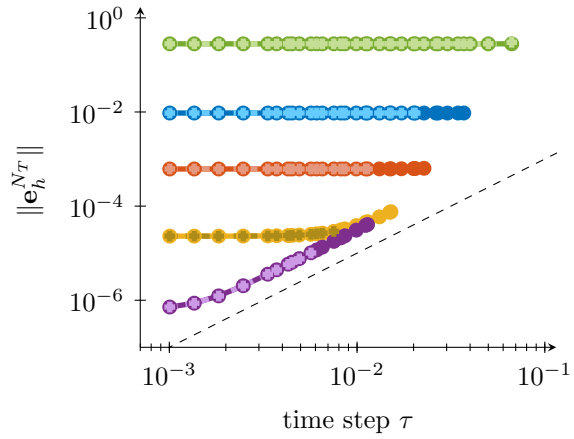


Figure 6.15: Temporal convergence. For the space discretization we employed the central fluxes method and polynomial degrees $k = 1$, $k = 2$, $k = 3$, $k = 4$ and $k = 5$. For the solid lines with • markers we used the locally implicit time integrator based on the sets $\mathcal{T}_h^i$ and $\mathcal{T}_h^e$. For the dashed lines with + markers we used the locally implicit method based on the sets $\widehat{\mathcal{T}}_h^i$ and $\widehat{\mathcal{T}}_h^e$, i.e. we did not include the coarse neighbors of the fine elements into the implicitly treated set of mesh elements. The black dashed line has slope $1/10\tau^2$. The final time was $T = t_{N_T} = 1$.

	$k=1$	$k=5$
$\mathrm{nz}(M_{\mathbf{E}}^{-1}C_{\mathbf{H}})$	56.697	1.744.759
$\mathrm{nz}(M_{\mathbf{E}}^{-1}C_{\mathbf{H}}^i)$	6.995	216.207
$\frac{\mathrm{nz}(M_{\mathbf{E}}^{-1}C_{\mathbf{H}})}{\mathrm{nz}(M_{\mathbf{E}}^{-1}C_{\mathbf{H}}^i)}$	12.3 %	12.4 %

	$k=1$	$k=5$
$\mathrm{nz}(M_{\mathbf{H}}^{-1}C_{\mathbf{E}})$	56.700	1.744.561
$\mathrm{nz}(M_{\mathbf{H}}^{-1}C_{\mathbf{E}}^i)$	6.996	216.171
$\frac{\mathrm{nz}(M_{\mathbf{H}}^{-1}C_{\mathbf{E}})}{\mathrm{nz}(M_{\mathbf{H}}^{-1}C_{\mathbf{E}}^i)}$	12.4 %	12.4 %

Table 6.11: Number of nonzero elements in the matrices $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}$, $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}$ associated with the full discrete curl-operators $\boldsymbol{\mathcal{C}}_{\mathbf{H}}$, $\boldsymbol{\mathcal{C}}_{\mathbf{E}}$, respectively, and in the matrices $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}^i$, $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}^i$ associated with the split implicit curl-operators $\boldsymbol{\mathcal{C}}_{\mathbf{H}}^i$, $\boldsymbol{\mathcal{C}}_{\mathbf{E}}^i$, respectively.

again confirm the improved spatial convergence of the upwind fluxes locally implicit method compared to the central fluxes locally implicit scheme. Moreover, Figures 6.14a – 6.14d prove the considerable relaxed CFL condition of the locally implicit schemes in comparison with the Verlet methods.

Last, we give in Figure 6.15 the errors of the central fluxes locally implicit scheme once with the right choice (i.e. the choice in accordance with Definition 5.1) of the implicitly and explicitly treated elements $\mathcal{T}_h^i$ and $\mathcal{T}_h^e$, and once with the wrong choice $\widehat{\mathcal{T}}_h^i$ and $\widehat{\mathcal{T}}_h^e$. We observe that the spatial and the temporal errors are not spoiled. However, we do observe that the CFL condition gets stricter if we do not treat the coarse neighbors of our fine elements implicitly.

Structure of linear system

In Sections 6.1 and 6.2 we elaborated the ideas how the locally implicit schemes can be implemented efficiently. Now, we illustrate these considerations by our numerical results, i.e. we examine the structure of the mass and the stiffness matrix, and of the linear system that has to be solved in each time step. Note that for the 2D TM Maxwell's equations the Maxwell

	$k=1$	$k=5$
$\mathrm{nz}(L_{\mathrm{cf}})$	74.508	2.644.836
$\mathrm{nz}(\widetilde{L})$	9.348	333.123
$\frac{\mathrm{nz}(L_{\mathrm{cf}})}{\mathrm{nz}(\widetilde{L})}$	12.5 %	12.6 %

Table 6.12: Number of nonzero elements in the left-hand sides $L_{\mathrm{cf}} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1} C_{\mathbf{H}} M_{\mathbf{H}}^{-1} C_{\mathbf{E}}$ and $\widetilde{L} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1} C_{\mathbf{H}}^{i} M_{\mathbf{H}}^{-1} C_{\mathbf{E}}^{i}$ of the Crank–Nicolson method and the locally implicit schemes, respectively. For the values corresponding to $\widetilde{L}$ we omitted the block which can be solved explicitly, i.e. the blue identity in Figure 6.18b.

operator reads

$$\mathcal{C} = \begin{pmatrix} 0 & -\mathcal{C}_{\mathbf{E}} \\ \mathcal{C}_{\mathbf{H}} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -\partial_y \\ 0 & 0 & \partial_x \\ -\partial_y & \partial_x & 0 \end{pmatrix}, \qquad \mathcal{C}_{\mathbf{E}} = \begin{pmatrix} \partial_y \\ -\partial_x \end{pmatrix}, \qquad \mathcal{C}_{\mathbf{H}} = \begin{pmatrix} -\partial_y & \partial_x \end{pmatrix}.$$

The matrices associated with $\mathcal{C}_{\mathbf{E}}$ and $\mathcal{C}_{\mathbf{H}}$, i.e. $M_{\mathbf{E}}^{-1} C_{\mathbf{H}}$ and $M_{\mathbf{H}}^{-1} C_{\mathbf{E}}$, inherit this structure.

As in Section 6.2 we order our dof such that our coefficient vector first contains the dof stemming from $\mathcal{T}_{h,e}^{e}$, then from $\mathcal{T}_{h,e}^{i}$, then from$\mathcal{T}_{h,i}^{e}$ and finally the ones from $\mathcal{T}_{h,i}^{i}$. In Figures 6.16 and 6.17 we give the structure (the nonzero elements) of the full matrices $M_{\mathbf{E}}^{-1} C_{\mathbf{H}}$, $M_{\mathbf{H}}^{-1} C_{\mathbf{E}}$ and of the implicit split matrices $M_{\mathbf{E}}^{-1} C_{\mathbf{H}}^{i}$, $M_{\mathbf{H}}^{-1} C_{\mathbf{E}}^{i}$. Moreover, we give the number of the nonzero entries in Table 6.11. First of all, we observe that the implicit matrices are considerable more sparse than the full matrices (only 12% of the nonzero entries). Moreover, we confirm the theoretical structures of the matrices we gave in Figures 6.5b and 6.6. In particular, the implicit split matrices in Figure 6.17 only depend on the implicit dof from $\mathcal{T}_{h,i}$ and on their explicit neighbors from $\mathcal{T}_{h,e}^{i}$. Last, in Figure 6.18 we give the structure of the left-hand sides

$$\widetilde{L} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1} C_{\mathbf{H}}^{i} M_{\mathbf{H}}^{-1} C_{\mathbf{E}}^{i}, \qquad L_{\mathrm{cf}} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1} C_{\mathbf{H}} M_{\mathbf{H}}^{-1} C_{\mathbf{E}},$$

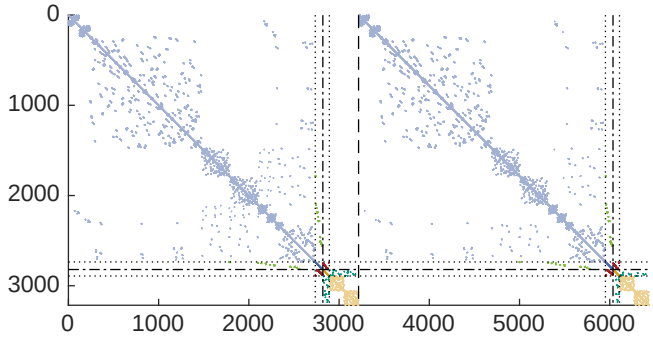
of the locally implicit schemes of the central fluxes Crank–Nicolson method, respectively, see (6.8). We see that $\widetilde{L}$ is considerably more sparse than L_{cf}, see also Table 6.12, and moreover that $\widetilde{L}$ only consists of the identity matrix for the dof associated with mesh elements in $\mathcal{T}_{h,e}^{e}$. This means that the linear system for the locally implicit schemes is only imposed on the dof of $\mathcal{T}_{h,e}^{i} \cup \mathcal{T}_{h,i}^{i}$, see Figure 6.18d. In contrary, the linear system for the Crank–Nicolson method involves all dof of $\mathcal{T}_h$.

6.3.3 Numerical example 3: rectangular mesh with barrier

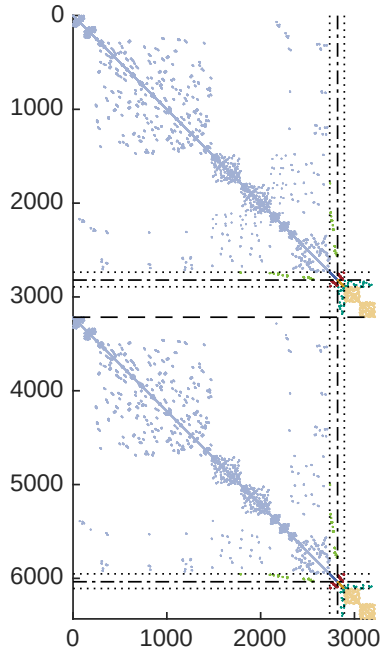
In our last example we study the performance of our locally implicit time integrators for a larger example (compared to the previous two examples). For the spatial discretization we used the mesh shown in Figure 6.19a, where we assign the red marked elements in Figure 6.19b to the fine part. As a polynomial degree in the dG method we chose $k = 6$.

For this example the left-hand sides of our locally implicit schemes and the central fluxes Crank–Nicolson method have

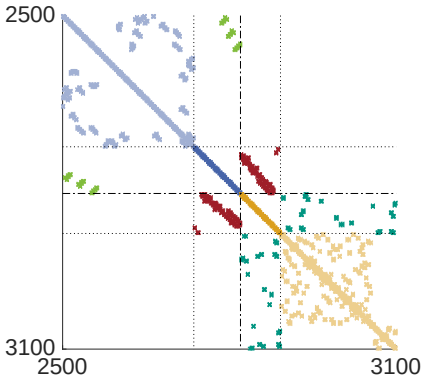
$$\mathrm{nz}(\widetilde{L}) = 676.630, \qquad \mathrm{nz}(L_{\mathrm{cf}}) = 18.191.637$$



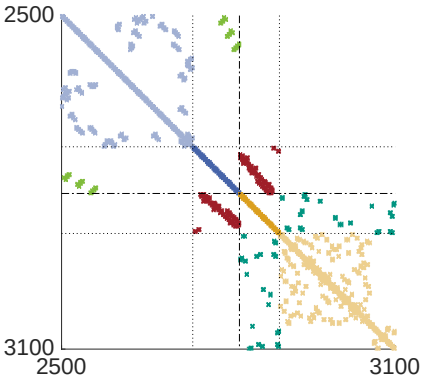
(a) Nonzero entries of the matrix $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}$.



(b) Nonzero entries of the matrix $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}$.

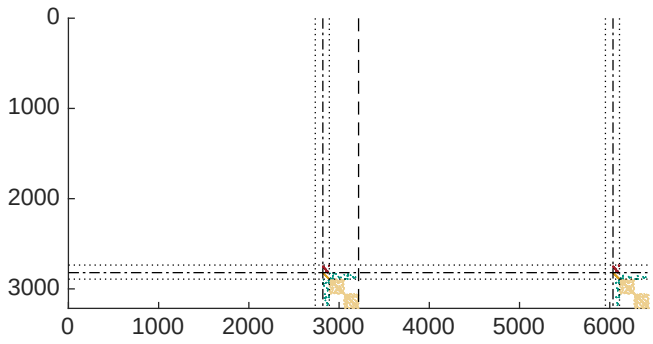


(c) Zoom on nonzero entries of the matrix $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}$.

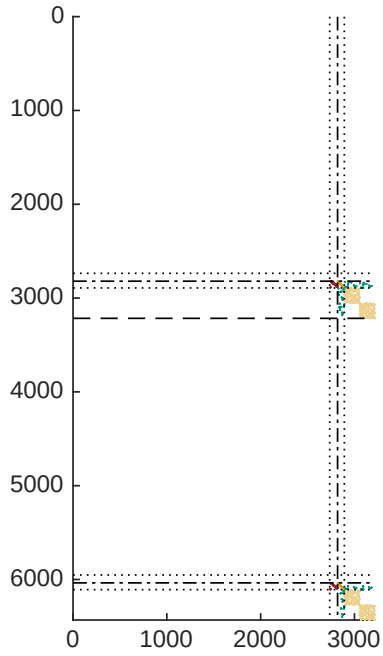


(d) Zoom on nonzero entries of the matrix $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}$.

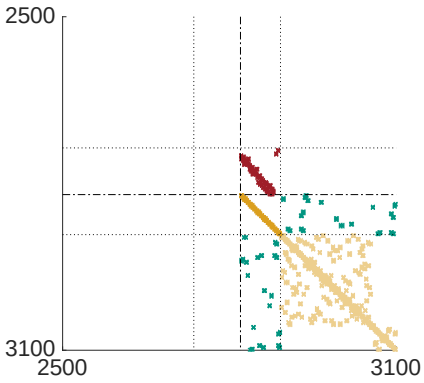
Figure 6.16: Structure of $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}$ and $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}$. We have the following coupling of the dof: Coupling between explicit elements: $\mathcal{T}_{h,e}^{e}$ to $\mathcal{T}_{h,e}^{e}$ (light blue), $\mathcal{T}_{h,e}^{i}$ to $\mathcal{T}_{h,e}^{i}$ (dark blue), $\mathcal{T}_{h,e}^{e}$ to $\mathcal{T}_{h,e}^{i}$ (pale green). Coupling between implicit elements: $\mathcal{T}_{h,i}^{e}$ to $\mathcal{T}_{h,i}^{e}$ (dark orange), $\mathcal{T}_{h,i}^{i}$ to $\mathcal{T}_{h,i}^{i}$ (light orange), $\mathcal{T}_{h,i}^{i}$ to $\mathcal{T}_{h,i}^{e}$ (green). Coupling between explicit and implicit elements: $\mathcal{T}_{h,e}^{i}$ to $\mathcal{T}_{h,i}^{e}$ (red) .



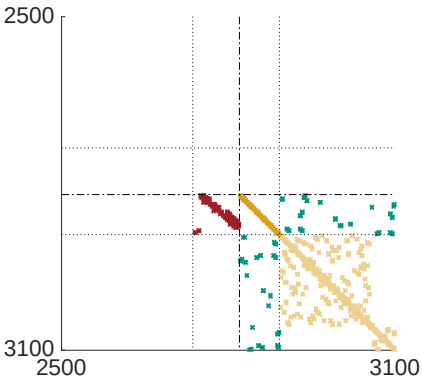
(a) Nonzero entries of the matrix $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}^{i}$.



(b) Nonzero entries of the matrix $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}^{i}$.

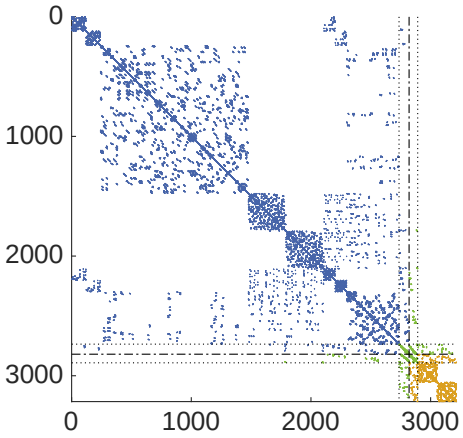


(c) Zoom on nonzero entries of the matrix $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}^{i}$.

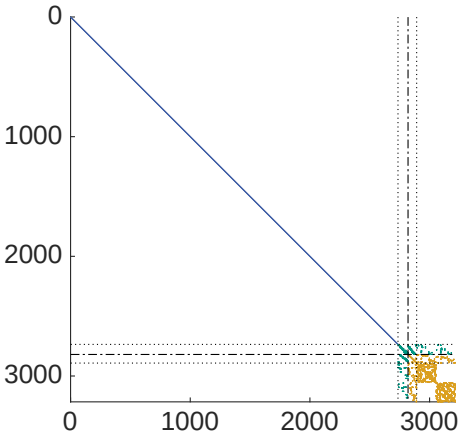


(d) Zoom on nonzero entries of the matrix $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}^{i}$.

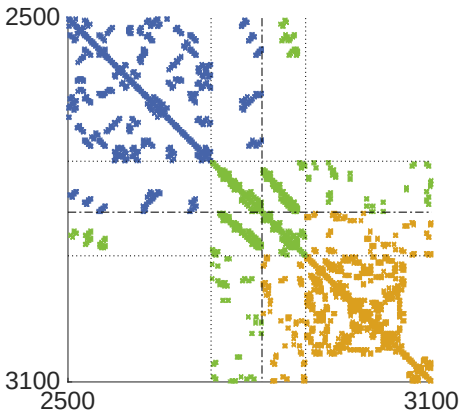
Figure 6.17: Structure of $M_{\mathbf{E}}^{-1}C_{\mathbf{H}}^{i}$ and $M_{\mathbf{H}}^{-1}C_{\mathbf{E}}^{i}$. We have the following coupling of the dof: Coupling between implicit elements: $\mathcal{T}_{h,i}^{e}$ to $\mathcal{T}_{h,i}^{e}$ (dark orange), $\mathcal{T}_{h,i}^{i}$ to $\mathcal{T}_{h,i}^{i}$ (light orange), $\mathcal{T}_{h,i}^{i}$ to $\mathcal{T}_{h,i}^{e}$ (green). Coupling between explicit and implicit elements: $\mathcal{T}_{h,e}^{i}$ to $\mathcal{T}_{h,i}^{e}$ (red) .



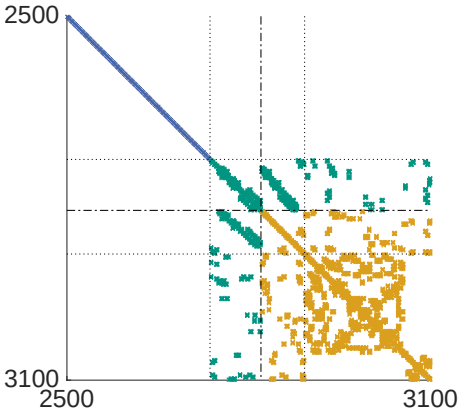
(a) Nonzero entries of the matrix $L_{\text{cf}} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1} C_{\mathbf{H}} M_{\mathbf{H}}^{-1} C_{\mathbf{E}}$.



(b) Nonzero entries of the matrix $\widetilde{L} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1} C_{\mathbf{H}}^{i} M_{\mathbf{H}}^{-1} C_{\mathbf{E}}^{i}$. The left upper block stems only from I.

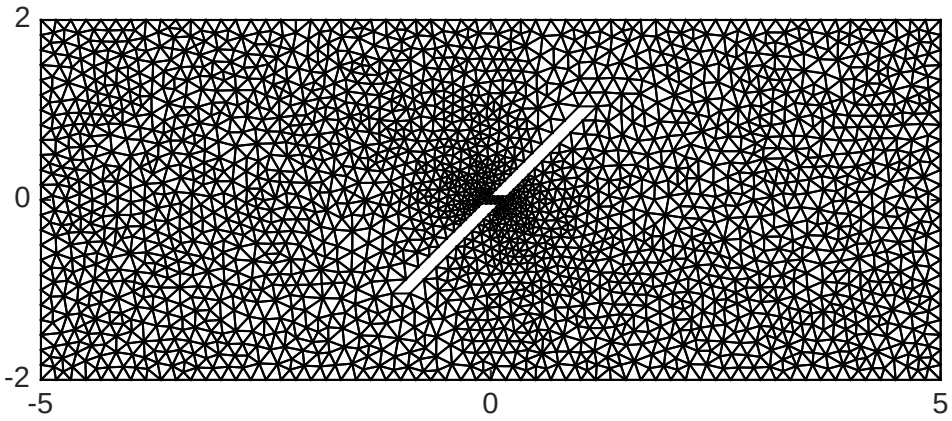


(c) Zoom on nonzero entries of the matrix $L_{\text{cf}} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1} C_{\mathbf{H}} M_{\mathbf{H}}^{-1} C_{\mathbf{E}}$.

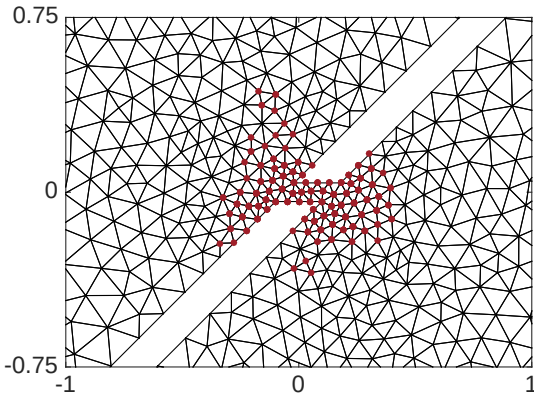


(d) Zoom on nonzero entries of the matrix $\widetilde{L} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1} C_{\mathbf{H}}^{i} M_{\mathbf{H}}^{-1} C_{\mathbf{E}}^{i}$.

Figure 6.18: Structure of $L_{\text{cf}} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1} C_{\mathbf{H}} M_{\mathbf{H}}^{-1} C_{\mathbf{E}}$ and $\widetilde{L} = I + \frac{\tau^2}{4} M_{\mathbf{E}}^{-1} C_{\mathbf{H}}^{i} M_{\mathbf{H}}^{-1} C_{\mathbf{E}}^{i}$. The blue entries only depend on the explicitly integrated mesh elements, the orange entries only depend on the implicitly integrated mesh elements, the pale green entries depend on both the explicitly and the implicitly treated elements. The green entries depend on both the explicitly treated elements in $\mathcal{T}_{h,e}^{i}$ and the implicitly treated elements in $\mathcal{T}_{h,i}^{e}$.



(a) Whole mesh.



(b) Zoom on the fine part of the mesh (red marked elements).

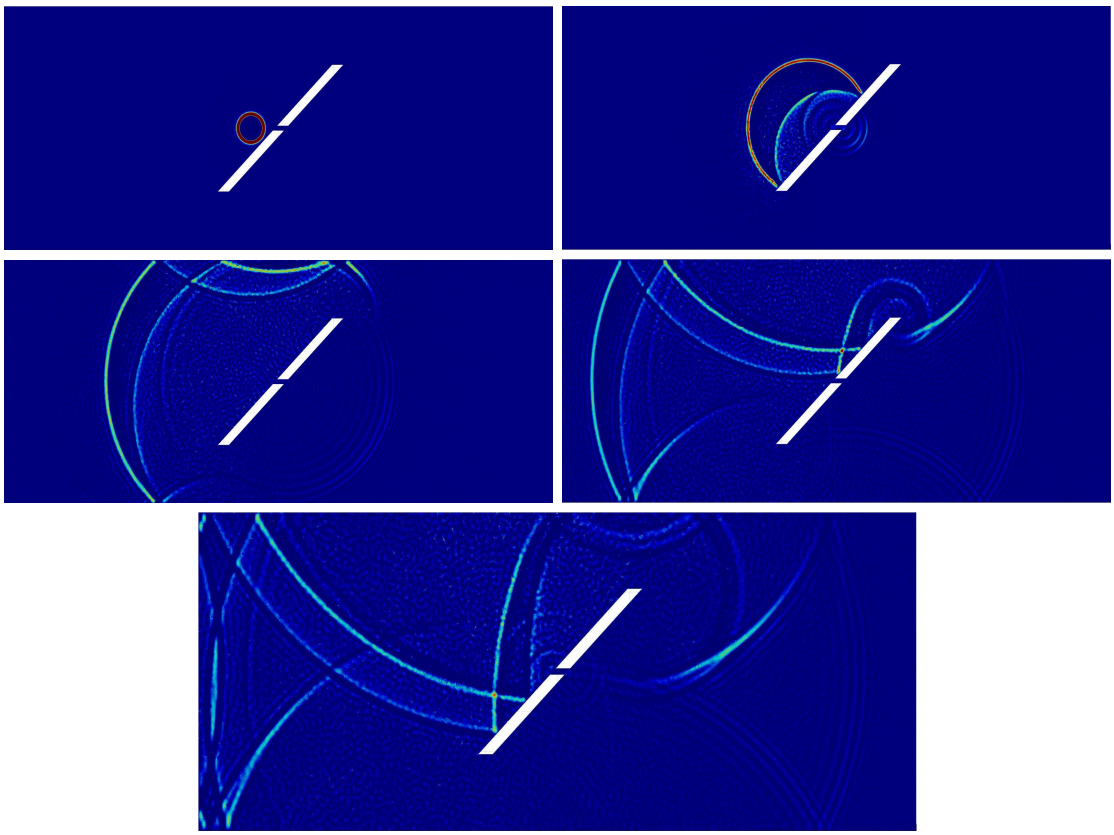
Figure 6.19: Mesh of the rectangular domain with a barrier inside which possesses a small gap in the middle. The maximum and minimum diameters of the mesh elements are given by 0.13 and 0.023, respectively.

nonzero entries, respectively. For the nonzero values of $\widetilde{L}$ we omitted the identity in the explicit part of $\widetilde{L}$ (the blue identity block in Figure 6.18b). We observe that by using the locally implicit schemes we can reduce the size of the linear system considerable, in fact, we only have to solve a linear system with 3.7 % of the nonzero elements compared with the Crank–Nicolson method.

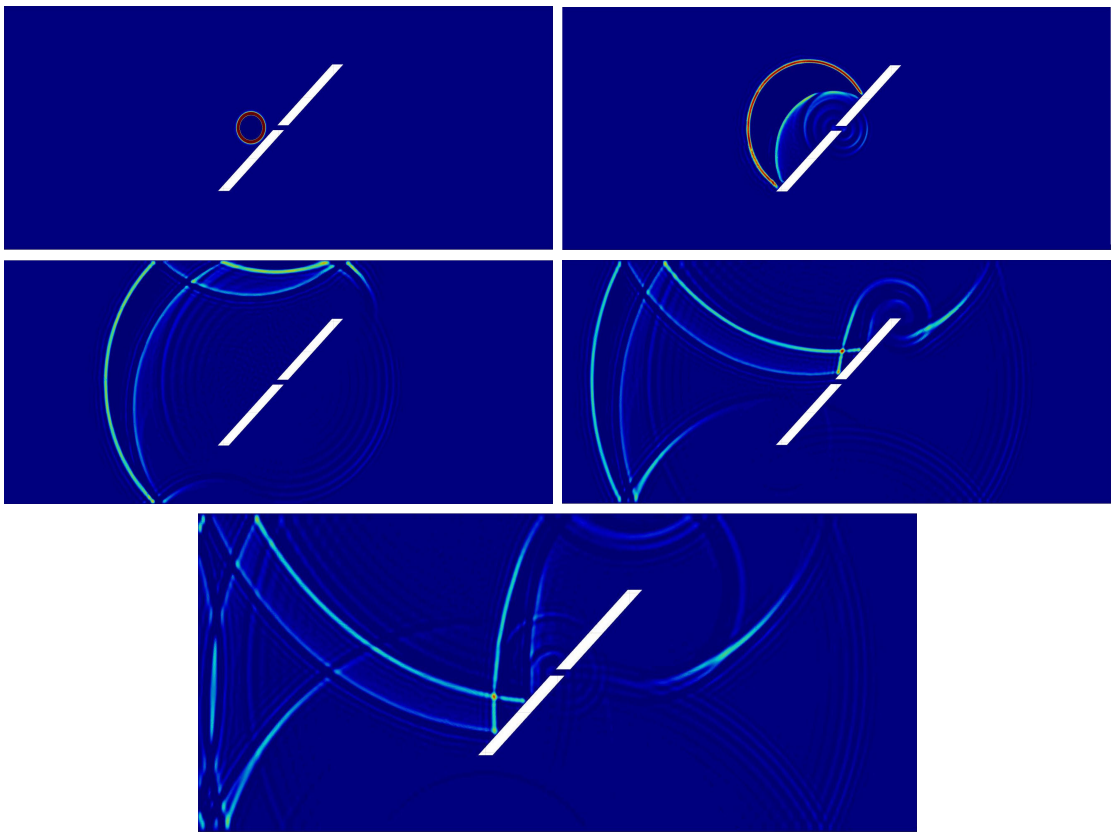
We ran our simulation with the central fluxes and the upwind fluxes ($\alpha = 1$) locally implicit schemes until the final time $T = 6$. As initial value we chose

$$\mathbf{H}_x \equiv 0, \qquad \mathbf{H}_y \equiv 0, \qquad \mathbf{E}_z = \exp\big(-1000(x+0.5)^2 + y^2\big).$$

In Figure 6.20 we give snapshots of the electric field from our simulation. We can nicely observe how the small gap affects the solution. Moreover, we see the improved properties of the upwind fluxes dG discretization. On the one hand we obtained a more detailed approximation while on the other hand we avoid artefacts as exhibited in the central fluxes case, see in particular the snapshot at time $t = 5.02$.



(a) Central fluxes.



(b) Upwind fluxes with stabilization parameter $\alpha = 1$.

Figure 6.20: Snapshots of the electric field $\mathbf{E}_h^n$ at times $t_n = 0.44, 1.31, 4.15, 5.02$. We used a dG method with central fluxes (upper plots) or upwind fluxes with $\alpha = 1$ (lower plots), the polynomial degree $k = 6$ and the locally implicit time integrators with time step $\tau = 0.0011$.

CHAPTER 7

Conclusion and outlook

In this thesis we presented and analyzed an efficient numerical method to discretize the linear Maxwell's equations on a locally refined spatial mesh. This scheme comprises a discontinuous Galerkin (dG) space discretization and a locally implicit time integrator.

We based our idea on a paper of Verwer [2011] and adapted the therein proposed locally implicit scheme to a variational formulation of the dG space discretization. We showed that this scheme can be interpreted as a pertubation of the Crank–Nicolson method. In order to analyze it we developed a novel technique inspired by the variation of constants formula and the boundedness of the solution groups of the continuous and semidiscrete Maxwell's equations. We are confident that this technique can be employed in a wide field in the analysis of time integration methods for PDEs.

Moreover, we succeeded in extending the locally implicit scheme from an unstabilized central fluxes dG discretization to a stabilized upwind fluxes dG method. This provides an improved stability behavior and a higher spatial convergence rate. For the analysis of this method we had to apply a completely different technique than in the central fluxes case, namely an energy technique.

Last, we showed how the locally implicit scheme can be implemented efficiently and verified our theoretical results with numerical experiments. These examples clearly show the improved CFL condition of our locally implicit method compared to the standard explicit time integrator for Maxwell's equations – the Verlet method.

As a byproduct of our work we provide a rigorous stability and error analysis for both the Crank–Nicolson method and the Verlet scheme.

Further extensions of this thesis comprise the application of the locally implicit time integrators to other PDEs such as the wave equation and to Maxwell's equations in anisotropic or even nonlinear materials.

Bibliography

R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics*. Elsevier/Academic Press, Amsterdam, 2. ed., repr. edition, 2008. ISBN 978-0-12-044143-3; 0-12-044143-8. URL http://www.sciencedirect.com/science/bookseries/00798169/140.

M. Almquist and M. Mehlin. Multi-level local time-stepping methods of Runge-Kutta type for wave equations. CRC 1173-Preprint 2016/16, Karlsruhe Institute of Technology, 2016. URL http://www.waves.kit.edu/downloads/CRC1173_Preprint_2016-16.pdf.

J. Alvarez, L. D. Angulo, M. R. Cabello, A. Rubio Bretones, and S. G. Garcia. An analysis of the leap-frog discontinuous Galerkin method for Maxwell's equations. *Microwave Theory and Techniques, IEEE Transactions on*, 62(2):197–207, Feb 2014. ISSN 0018-9480. URL http://dx.doi.org/10.1109/TMTT.2013.2295775.

E. Burman, A. Ern, and M. A. Fernández. Explicit Runge-Kutta schemes and finite elements with symmetric stabilization for first-order linear PDE systems. *SIAM J. Numer. Anal.*, 48 (6):2019–2042, 2010. ISSN 0036-1429. URL http://dx.doi.org/10.1137/090757940.

K. Busch, G. von Freymann, S. Linden, S. F. Mingaleev, L. Tkeshelashvili, and M. Wegener. Periodic nanostructures for photonics. *Physics Reports*, 444(36):101–202, 2007. ISSN 0370-1573. URL http://dx.doi.org/10.1016/j.physrep.2007.02.011.

K. Busch, M. König, and J. Niegemann. Discontinuous galerkin methods in nanophotonics. *Laser & Photonics Reviews*, 5(6):773–809, 2011. ISSN 1863-8899. URL http://dx.doi.org/10.1002/lpor.201000045.

M. Costabel and M. Dauge. Singularities of electromagnetic fields in polyhedral domains. *Archive for Rational Mechanics and Analysis*, 151(3):221–276, 2000. ISSN 0003-9527, 1432-0673. URL http://dx.doi.org/10.1007/s002050050197.

A. Demirel, J. Niegemann, K. Busch, and M. Hochbruck. Efficient multiple time-stepping algorithms of higher order. *J. Comput. Phys.*, 285:133–148, 2015. ISSN 0021-9991. URL http://dx.doi.org/10.1016/j.jcp.2015.01.018.

S. Descombes, S. Lanteri, and L. Moya. Locally implicit time integration strategies in a discontinuous Galerkin method for Maxwell's equations. *J. Sci. Comput.*, 56(1):190–218, 2013. ISSN 0885-7474. URL http://dx.doi.org/10.1007/s10915-012-9669-5.

S. Descombes, S. Lanteri, and L. Moya. Locally implicit discontinuous Galerkin time domain method for electromagnetic wave propagation in dispersive media applied to numerical dosimetry in biological tissues. *SIAM Journal on Scientific Computing*, pages A2611–A2633, 2016. ISSN 1064-8275. URL http://dx.doi.org/10.1137/15M1010282.

S. Descombes, S. Lanteri, and L. Moya. Temporal convergence analysis of a locally implicit discontinuous Galerkin time domain method for electromagnetic wave propagation in dispersive media. *Journal of Computational and Applied Mathematics*, 316:122–132, 2017. ISSN 0377-0427. URL http://dx.doi.org/10.1016/j.cam.2016.09.038.

D. A. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*, volume 69 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Heidelberg, 2012. ISBN 978-3-642-22979-4. URL http://dx.doi.org/10.1007/978-3-642-22980-0.

J. Diaz and M. J. Grote. Energy conserving explicit local time stepping for second-order wave equations. *SIAM J. Sci. Comput.*, 31(3):1985–2014, 2009. ISSN 1064-8275. URL http://dx.doi.org/10.1137/070709414.

J. Diaz and M. J. Grote. Multi-level explicit local time-stepping methods for second-order wave equations. *Comput. Methods Appl. Mech. Engrg.*, 291:240–265, 2015. ISSN 0045-7825. URL http://dx.doi.org/10.1016/j.cma.2015.03.027.

R. Diehl, K. Busch, and J. Niegemann. Comparison of low-storage Runge-Kutta schemes for discontinuous Galerkin time-domain simulations of Maxwell's equations. *Journal of Computational and Theoretical Nanoscience*, 7(8):1572–1580, 2010. ISSN 1546-1955. URL http://dx.doi.org/10.1166/jctn.2010.1521.

W. Dörfler. *Numerical methods for partial differential equations: Adaptive finite element methods.* Arbeitsgruppe Numerik partieller Differentialgleichungen, Institut für Angewandte und Numerische Mathematik, 2013.

W. Dörfler, A. Lechleiter, M. Plum, G. Schneider, and C. Wieners. *Photonic crystals. Mathematical analysis and numerical approximation.* Berlin: Springer, 2011. ISBN 978-3-0348-0112-6/pbk; 978-3-0348-0113-3/ebook. URL http://dx.doi.org/10.1007/978-3-0348-0113-3.

E. Emmrich. Discrete versions of Gronwall's lemma and their application to the numerical analysis of parabolic problems. Preprint no. 637, Fachbereich Mathematik, TU Berlin, 1999. URL https://www.math.uni-bielefeld.de/~emmrich/public/prepA.pdf.

K.-J. Engel and R. Nagel. *One-parameter semigroups for linear evolution equations*, volume 194 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000. ISBN 0-387-98463-1. URL https://www.springer.com/de/book/9780387984636.

H. Fahs. High-order leap-frog based discontinuous Galerkin method for the time-domain Maxwell equations on non-conforming simplicial meshes. *Numer. Math. Theory Methods Appl.*, 2(3):275–300, 2009. ISSN 1004-8979. URL http://dx.doi.org/10.1016/j.cam.2009.05.015.

M. J. Grote and T. Mitkova. Explicit local time-stepping methods for Maxwell's equations. *J. Comput. Appl. Math.*, 234(12):3283–3302, 2010. ISSN 0377-0427. URL http://dx.doi.org/10.1016/j.cam.2010.04.028.

M. J. Grote and T. Mitkova. High-order explicit local time-stepping methods for damped wave equations. *J. Comput. Appl. Math.*, 239:270–289, 2013. ISSN 0377-0427. URL http://dx.doi.org/10.1016/j.cam.2012.09.046.

M. J. Grote, M. Mehlin, and T. Mitkova. Runge-Kutta-based explicit local time-stepping methods for wave propagation. *SIAM J. Sci. Comput.*, 37(2):A747–A775, 2015. ISSN 1064-8275. URL http://dx.doi.org/10.1137/140958293.

E. Hairer and G. Wanner. *Solving ordinary differential equations II. Stiff and differential-algebraic problems*, volume 14 of *Springer Series in Computational Mathematics*. Springer, Berlin, Heidelberg, 2nd edition, 1996. URL https://www.springer.com/de/book/9783540604525.

E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration : structure-preserving algorithms for ordinary differential equations*. Springer series in computational mathematics ; 31. Springer, Berlin, 2. ed. edition, 2006. ISBN 3-540-30663-3; 978-3-540-30663-4. URL https://link.springer.com/book/10.1007%2F3-540-30666-8.

J. S. Hesthaven and T. Warburton. *Nodal discontinuous Galerkin methods*, volume 54 of *Texts in Applied Mathematics*. Springer, New York, 2008. ISBN 978-0-387-72065-4. URL http://dx.doi.org/10.1007/978-0-387-72067-8.

M. Hochbruck. *Skriptum zur Vorlesung Numerik I, II und Numerik von Differentialgleichungen*. WS 2013/14 – WS 2014/15. Arbeitsgruppe Numerik, Institut für Angewandte und Numerische Mathematik, Karlsruher Institut für Technologie, 2015. URL https://na.math.kit.edu/download/teaching/2014w/nummetdgl/skript/skript.pdf.

M. Hochbruck and A. Ostermann. Exponential integrators. *Acta Numer.*, 19:209–286, 2010. ISSN 0962-4929. URL http://dx.doi.org/10.1017/S0962492910000048.

M. Hochbruck and A. Ostermann. Exponential multistep methods of Adams-type. *BIT*, 51(4):889–908, 2011. ISSN 0006-3835. URL http://dx.doi.org/10.1007/s10543-011-0332-6.

M Hochbruck and T. Pažur. Implicit Runge–Kutta methods and discontinuous Galerkin discretizations for linear Maxwell's equations. *SIAM J. Numer. Anal.*, 53(1):485–507, 2015. URL http://dx.doi.org/10.1137/130944114.

M. Hochbruck and A. Sturm. Error analysis of a second-order locally implicit method for linear Maxwell's equations. *SIAM J. Numer. Anal.*, 54(5):3167–3191, 2016. ISSN 0036-1429. doi: 10.1137/15M1038037. URL http://dx.doi.org/10.1137/15M1038037.

M. Hochbruck and A. Sturm. Upwind discontinuous Galerkin space discretization and locally implicit time integration for linear Maxwell's equations. CRC 1173-Preprint 2017/4, Karlsruhe Institute of Technology, 2017. URL http://www.waves.kit.edu/downloads/CRC1173_Preprint_2017-4.pdf.

M. Hochbruck, T. Jahnke, and R. Schnaubelt. Convergence of an ADI splitting for Maxwell's equations. *Numerische Mathematik*, 129:535–561, 2015a. URL http://dx.doi.org/10.1007/s00211-014-0642-0.

M. Hochbruck, T. Pažur, A. Schulz, E. Thawinan, and C. Wieners. Efficient time integration for discontinuous Galerkin approximations of linear wave equations. *ZAMM*, 95(3):237–259, 2015b. URL http://dx.doi.org/10.1002/zamm.201300306.

B. Jacob and H. J. Zwart. *Linear port-Hamiltonian systems on infinite-dimensional spaces*, volume 223 of *Operator Theory: Advances and Applications*. Birkhäuser/Springer Basel AG, Basel, 2012. ISBN 978-3-0348-0398-4. URL http://dx.doi.org/10.1007/978-3-0348-0399-1.

A. Kirsch and F. Hettlich. *The mathematical theory of time-harmonic Maxwell's equations*, volume 190 of *Applied Mathematical Sciences*. Springer, Cham, 2015. ISBN 978-3-319-11085-1; 978-3-319-11086-8. URL http://dx.doi.org/10.1007/978-3-319-11086-8.

M. Mehlin. *Efficient explicit time integration for the simulation of acoustic and electromagnetic waves*. PhD thesis, University of Basel, 2015. URL http://edoc.unibas.ch/38308/.

P. Monk. *Finite element methods for Maxwell's equations*. Numerical Mathematics and Scientific Computation. Oxford University Press, New York, 2003. ISBN 0-19-850888-3. URL http://dx.doi.org/10.1093/acprof:oso/9780198508885.001.0001.

E. Montseny, S. Pernet, X. Ferriéres, 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.*, 227(14):6795–6820, 2008. ISSN 0021-9991. URL http://dx.doi.org/10.1016/j.jcp.2008.03.032.

L. Moya. Temporal convergence of a locally implicit discontinuous Galerkin method for Maxwell's equations. *ESAIM Math. Model. Numer. Anal.*, 46(5):1225–1246, 2012. ISSN 0764-583X. URL http://dx.doi.org/10.1051/m2an/2012002.

T. Namiki. A new FDTD algorithm based on alternating-direction implicit method. *IEEE Transactions on Microwave Theory and Techniques*, 47(10):2003–2007, oct 1999. ISSN 0018-9480. URL http://dx.doi.org/10.1109/22.795075.

T. Namiki. 3-D ADI-FDTD method-unconditionally stable time-domain algorithm for solving full vector Maxwell's equations. *IEEE Transactions on Microwave Theory and Techniques*, 48 (10):1743–1748, oct 2000. ISSN 0018-9480. URL http://dx.doi.org/10.1109/22.873904.

J.-C. Nédélec. Mixed finite elements in $\mathbb{R}^3$. *Numer. Math.*, 35(3):315–341, 1980. ISSN 0029-599X. URL http://dx.doi.org/10.1007/BF01396415.

J. Niegemann. *Higher-order methods for solving Maxwell's equations in the time-domain*. PhD thesis, Karlsruhe Institute of Technology, 2009. URL https://publikationen.bibliothek.kit.edu/1000011812.

R. H. Nochetto, K. G. Siebert, and A. Veeser. Theory of adaptive finite element methods: An introduction. In *Multiscale, Nonlinear and Adaptive Approximation*, pages 409–542. Springer, Berlin, Heidelberg, 2009. URL http://link.springer.com/chapter/10.1007/978-3-642-03413-8_12.

T. Pažur. *Error analysis of implicit and exponential time integration of linear Maxwell's equations*. PhD thesis, Karlsruhe Institute of Technology, 2013. URL https://publikationen.bibliothek.kit.edu/1000038617.

A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*, volume 44 of *Applied Mathematical Sciences*. Springer, New York, 1983. ISBN 9780387908458. URL https://doi.org/10.1007/978-1-4612-5561-1.

S. Piperno. Symplectic local time-stepping in non-dissipative DGTD methods applied to wave propagation problems. *M2AN Math. Model. Numer. Anal.*, 40(5):815–841 (2007), 2006. ISSN 0764-583X. URL http://dx.doi.org/10.1051/m2an:2006035.

M. Pototschnig, J. Niegemann, L. Tkeshelashvili, and K. Busch. Time-domain simulations of the nonlinear Maxwell equations using operator-exponential methods. *IEEE Transactions on Antennas and Propagation*, 57(2):475–483, February 2009. ISSN 0018-926X. URL http://dx.doi.org/10.1109/TAP.2008.2011181.

W. H. Reed and T. R. Hill. Triangular mesh methods for the neutron transport equation. Los Alamos Scientific Laboratory Report LA-UR–73-479, Los Alamos Scientific Lab., N.Mex. (USA), 1973. URL https://www.osti.gov/scitech/biblio/4491151.

R. Schnaubelt. *Lecture notes "Evolution Equations"*. WS 2010/11. Workgroup Functional Analysis, Institute for Analysis, Karlsruhe Institute of Technology, 2010–2011. URL http://www.math.kit.edu/iana3/~schnaubelt/media/evgl-skript.pdf.

R. Schnaubelt. *Lecture notes "Operator Semigroups and Dispersive Equations"*. WS 2012/13. Workgroup Functional Analysis, Institute for Analysis, Karlsruhe Institute of Technology, 2012–2013. URL http://www.math.kit.edu/iana3/~schnaubelt/media/isem16-skript.pdf.

R. Schnaubelt. *Lecture notes "Spectral Theory"*. SS 2015. Workgroup Functional Analysis, Institute for Analysis, Karlsruhe Institute of Technology, 2015. URL http://www.math.kit.edu/iana3/~schnaubelt/media/st-skript15.pdf.

J. G. Verwer. Component splitting for semi-discrete Maxwell equations. *BIT*, 51(2):427–445, 2011. ISSN 0006-3835. URL http://dx.doi.org/10.1007/s10543-010-0296-y.

K. S. Yee. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation*, 14(3):302–307, May 1966. ISSN 0018-926X. URL http://dx.doi.org/10.1109/TAP.1966.1138693.

F. Zhen, Z. Chen, and J. Zhang. Toward the development of a three-dimensional unconditionally stable finite-difference time-domain method. *IEEE Transactions on Microwave Theory and Techniques*, 48(9):1550–1558, sep 2000. ISSN 0018-9480. URL http://dx.doi.org/10.1109/22.869007.

APPENDIX A

Auxiliary results and identities

In this appendix we give auxiliary results which we use throughout the thesis.

For the **scalar triple product** of three vectors a, $b, c \in \mathbb{R}^3$ we have that

$$(a \times b) \cdot a = (a \times b) \cdot b = 0,$$

i.e. $a \times b \perp a, b$. Furthermore, the scalar triple product can be expressed as $(a \times b) \cdot c = \det(a, b, c)$ and satisfies the identity

$$(a \times b) \cdot c = -(a \times c) \cdot b. \tag{A.1}$$

Next, we state some useful inequalities:

Let $a, b \geq 0$ be two non-negative numbers and $\gamma > 0$ be a positive weight. The **weighted Young's inequality** states that

$$ab \leq \frac{\gamma}{2} a^2 + \frac{1}{2\gamma} b^2. \tag{A.2}$$

For two vectors $a, b \in \mathbb{R}^n$ the **Cauchy-Schwarz inequality** (in $\mathbb{R}^n$) gives that

$$a \cdot b \leq |a||b| \qquad \Longleftrightarrow \qquad \sum_{m=1}^{n} a_m b_m \leq \left(\sum_{m=1}^{n} a_m^2 \right)^{1/2} \left(\sum_{m=1}^{n} b_m^2 \right)^{1/2}. \tag{A.3}$$

Let $v, w \in L^p(D)$, $p \in [1, \infty]$. The **Minkowski inequality** (in L^p) yields that

$$\|v + w\|_{L^p(D)} \leq \|v\|_{L^p(D)} + \|w\|_{L^p(D)}. \tag{A.4}$$

We will refer to this inequality by the **triangle inequality** (in L^p).

Let $v, w \in L^2(D)$. Then, $vw \in L^1(D)$ and the **Cauchy-Schwarz inequality** (in L^2) ensures that

$$\big(v, w\big)_D \leq \|v\|_D \|w\|_D. \tag{A.5}$$

Let $v, w \in L^2(D)$. By combining the triangle and Young's inequality with weight $\gamma = 1$ we obtain that,

$$\|v + w\|_D^2 \leq 2\big(\|v\|_D^2 + \|w\|_D^2\big). \tag{A.6}$$

In the following lemma we give a **modification** of the **continuous Gronwall lemma**.

Lemma A.1. *Let $\lambda \geq 0$ and $a \in L^\infty(0,T)$. Moreover, let $b \in C(0,T)$ be a monotonically increasing and $c \in L^\infty(0,T)$ be a non-negative function. If*

$$a(t) + c(t) \leq b(t) + \lambda \int_0^t a(s)\, ds, \qquad \text{a.e. in } [0,T] \tag{A.7}$$

is satisfied, then there holds

$$a(t) + c(t) \leq e^{\lambda t} b(t). \tag{A.8}$$

Proof. Since c is non-negative, we can estimate (A.7) further by

$$a(t) + c(t) \leq b(t) + \lambda \int_0^t a(s) + c(s)\, ds, \qquad \text{a.e. in } [0,T].$$

The continous Gronwall lemma [Emmrich, 1999, Proposition 2.1] gives the assertion. □

Next, we give a **modified discrete Gronwall lemma**.

Lemma A.2. *Let $\lambda \geq 0$, $\tau > 0$ and $\frac{\lambda}{2}\tau < 1$. Furthermore, let $\{a_n\}$, $\{b_n\} \subset \mathbb{R}$ be two sequences satisfying $a_0 \leq b_0$, $\{c_n\} \subset \mathbb{R}_+$ be a non-negative sequence and*

$$a_{n+1} + c_{n+1} \leq b_{n+1} + \lambda \frac{\tau}{2} \sum_{m=0}^{n} (a_{m+1} + a_m). \tag{A.9}$$

Then, if $\{b_n\}$ is monotonically increasing, there holds

$$a_n + c_n \leq \left(\frac{1 + \lambda \frac{\tau}{2}}{1 - \lambda \frac{\tau}{2}} \right)^n b_n. \tag{A.10}$$

If in addition, $\lambda\tau \leq \frac{3}{2}$, then

$$a_n + c_n \leq e^{\frac{3}{2}\lambda n \tau} b_n. \tag{A.11}$$

Proof. Since $\{c_n\}$ is a non-negative sequence we get from (A.9)

$$a_{n+1} + c_{n+1} \leq b_{n+1} + \lambda \frac{\tau}{2} \sum_{m=0}^{n} (a_{m+1} + c_{m+1} + a_m + c_m),$$

whence the statement (A.10) follows from [Emmrich, 1999, Proposition 4.1]. The bound (A.11) follows by

$$\frac{1+x}{1-x} \leq e^{3x}, \quad x \in [0, \frac{3}{4}].$$

□