Theoretical and numerical analysis of local dispersion models coupled to a discontinuous Galerkin time-domain method for Maxwell's equations

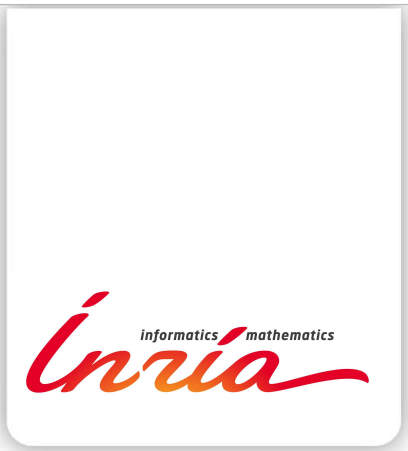
Jonathan Viquerat, Maciej Klemm, Stéphane Lanteri, Claire Scheid

Theoretical and numerical analysis of local dispersion models coupled to a discontinuous Galerkin time-domain method for Maxwell's equations

Jonathan Viquerat, Maciej Klemm, Stéphane Lanteri, Claire Scheid

RESEARCH REPORT
N° 8298
3 Mai 2013
Project-Team Nachos

ISSN 0249-6399 ISRN INRIA/RR--8298--FR+ENG



Theoretical and numerical analysis of local dispersion models coupled to a discontinuous Galerkin time-domain method for Maxwell's equations

Jonathan Viquerat, Maciej Klemm*, Stéphane Lanteri, Claire Scheid†

Project-Team Nachos

Research Report n° 8298 — 3 Mai 2013 — 76 pages

Abstract: This report focuses on a centered-fluxes discontinuous Galerkin method coupled to a second-order Leap-Frog time scheme for the propagation of electromagnetic waves in dispersive media. After a presentation of the physical phenomenon and the classical dispersion models (particularly the Drude one), a generalized dispersive model is introduced. An *a priori* stability and convergence study is lead for the Drude model, as well as in the generalized dispersive case. Eventually, numerical results are presented for various test-cases, highlighting the interest of a proper description of the dispersion phenomenon in metals at the nanoscale.

Key-words: Discontinuous Galerkin method, Maxwell's equations, numerical electromagnetism, dispersive media, nanophotonics, Drude model

* University of Bristol
† Nice-Sophia Antipolis University, J. A. Dieudonné Lab.

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Theoretical and numerical analysis of local dispersion models coupled to a discontinuous Galerkin time-domain method for Maxwell's equations

Résumé : Ce rapport présente une méthode de Galerkin discontinue à flux centrés couplée à un schéma d'avancée en temps de type Leap-Frog d'ordre deux pour la propagation des ondes électromagnétiques dans les milieux dispersifs. Après une présentation du phénomène physique ainsi que des modèles de dispersion les plus classiques (notamment celui de Drude), un modèle de dispersion généralisé est introduit. Une étude de stabilité et de convergence *a priori* est conduite dans le cas du modèle de Drude, ainsi que dans le cas généralisé. Enfin, des résultats numériques sont présentés pour différents cas-tests, mettant en lumière l'intérêt d'une bonne description des phénomènes de dispersion des métaux à l'échelle nanoscopique.

Mots-clés : Méthode de Galerkin discontinue, équations de Maxwell, électromagnétisme numérique, milieux dispersifs, nanophotonique, modèle de Drude.

Contents

List of Figures

List of Tables

1 Photonics and dispersive media

The increasing need in electronic signals processing during the last century came along with an advanced miniaturization of all the associated components. The revolution that followed the advent of the transistor deeply modified the common use of information. In the seventies, the development of the optical fiber led the way to fastest and largest data exchanges by exploiting light flow instead of electron flow. A precise control of the electromagnetic waves (confinment, transmitted frequencies, propagation direction) using wavelength-size systems would permit to replace the electronic-based management of the information. For many fields (*e.g.* optics, informatics, medicine), the benefits withdrawned from such an outcome would be invaluable.

The particular properties required by such devices are artificial, *i.e.* they are not displayed by any natural material. The macroscopic behavior of these "metamaterials" usually results from their nanoscopic structure, which induces particular interactions with propagating waves. Under this vocable are included the negative-index materials [VBSH06], as well as photonic cristals [JJ07], for example.

Then, the fast-growing metamaterials development is due to their innovative features at the macroscopic scale. However, one must not forget that these properties rely on those of "regular" materials (dielectrics, metals, etc.) assembled at the nanoscopic scale. Hence, poorly modelized attributes for the latter would result in imprecisely predicted properties for the resulting metamaterial. Among others, physical dispersion has a great impact on the material/electromagnetic wave interaction; therefore, attention has been called on the various ways of modeling physical dispersion, especially in metals.

1.1 Physical dispersion

1.1.1 Definition

Dispersion is a common phenomenon to all kinds of waves traveling through a material medium : it results from the way the latter reacts to the presence of the wave, therefore affecting its propagation. For a polychromatic wave, it often happens that all the frequencies do not travel at the same speed through the medium : this phenomenon is called dispersion. In order to characterize it, one generaly tries to express a *dispersion relation* that bounds the angular frequency ω to the wave number k.

1.1.2 Dispersion relation

Since it links a temporal parameter ω to a spatial one k, this relation $\omega = f(k)$ describes the allowed modes inside the considered domain. It is often written as :

$$\omega(k) = v(k)k, \tag{1}$$

where $v(k)$ is the speed of the *monochromatic* wave of wavenumber k. This relation is often explicit in the most simple cases, but it can become implicit in more complicated situations.

Let us now consider a polychromatic wave that can be expanded into a sum of monochromatic modes. In the case of a non-dispersive medium, $v(k)$ would remain constant for every value of k, and all the frequencies would travel at the same speed. On the contrary, if the

medium is dispersive, (1) indicates that each frequency travels at a different speed. Therefore, a signal consisting of frequencies that are initially spatially localized will sprawl during its propagation. The description of this phenomenon calls for the definition of two different speeds :

- The speed of the wavefront, usually called *phase velocity* : $v_\phi = \dfrac{\omega(k)}{k}$,
- The speed of the wave envelope, called *group velocity* : $v_g = \dfrac{\partial \omega}{\partial k}$.

These two velocities are equal in the case of a non-dispersive medium. Otherwise, the group velocity is the physical speed of the problem, given it is linked to the speed at which the energy moves through the medium. On the other hand, the phase velocity is more artificial, since it represents the speed at which a point of constant phase moves through the medium. One could imagine situations for which the phase velocity exceeds the speed of light c, without violating any fundamental principle of physics.

We now restrain our study to the case of propagating electromagnetic waves (EMW) through dispersive metals, which is at the heart of to the numerical study that will follow.

1.1.3 Physical origin of the dispersion in metals for EMW

In the presence of a constant electric field, the electrons of a metal are subjected to a Coulomb force which brings them, in a given characteristic time τ_c, to an equilibrium position. This leads to a general electric polarization of the metal, which is usually expressed with the polarization vector $\mathbf{P}$. The latter constitutes an additional term to the *electric displacement* field $\mathbf{D}$: $\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}$. Moreover, $\mathbf{P}$ can be related to $\mathbf{E}$ in homogeneous isotropic media through its *susceptibility* χ such that $\mathbf{P} = \chi \mathbf{E}$. One should now grasp the importance of taking the dispersion effects into account when $\mathbf{P}$ cannot be neglected, since it has a great influence on the permittivity ε of the considered medium, and hence on its optical index.

If one is to consider a variable electric field of given angular frequency ω, the frequency dependence of $\mathbf{P}$ can be intuitively understood : for low enough frequencies, the electrons relaxation time τ is negligible compared to $\frac{1}{\omega}$. Therefore, the electrons dispose of a sufficient amount of time to adapt to the variations of the electric field. However, at higher frequencies, the field varies significantly during the time τ required by the electrons to reach a stable state. Then, the higher the frequency, the shorter the distance traveled by the electrons from their steady state equilibrium, and the lower the polarization. This explains the observed transparency of the metals for very high frequencies EMW. Moreover, one can now easily picture that the electrons contributing to the polarization mainly belong to the conduction band, since they are less bound to the nuclei, and therefore more movable.

It is now possible to introduce the most common modelisations of the disperion phenomenon in metals. Among others, the well-known Drude and Drude-Lorentz models will be presented.

1.2 Drude model

The discovery of the electron in 1897 by Thomson was followed, three years later, by the Drude theory based on the kinetic theory of gases [Dru00]. This particularly simple theory successfully accounts for the optical and thermical properties of some metals. In this

model, the metal is considered as a static lattice of positive ions immersed in a free electrons gas. Those electrons are considered to be the valence electrons of each metallic atom, that got delocalized when put into contact with the potential produced by the rest of the lattice atoms. They are often called *conduction electrons*, and wander freely in the *conduction band*, the core electrons being considered to be strongly bound to the ion core. As a consequence, the Drude model is often described as an *intraband model*. Moreover, several hypothesis are made :

- The electrons description is non-relativistic;
- The only considered interactions are the electron/wave and the electron/ion ones;
- The electron/ion collisions are instantaneous and random events, and their probability of happening during a dt amount of time is equal to $\frac{\mathrm{d}t}{\tau_f}$[1];
- After an electron/ion collision, the new velocity and direction of the electron are independant of those before the collision.

Under these hypotheses, the frequency dependence of the medium permittivity can be deduced from the equations of motion.

1.2.1 Relative permittivity

In the absence of relativistic effects (as the Laplace force), the velocity equation of an electron can be written as follows :

$$\frac{\partial \mathbf{v}}{\partial t} + \gamma_d \mathbf{v} = -\frac{e}{m_e}\mathbf{E}(t), \tag{2}$$

where m_e represents the electron mass, e the electronic charge, and γ_d a coefficient linked to the electron/ion collisions. One could notice that $[\gamma_d] = T^{-1}$, and therefore γ_d matches the definition of the inverse of the mean free path τ_f.

For an harmonic field (*i.e.* $\mathbf{E} = \mathbf{E}_0\, \mathrm{e}^{-i\omega t}$), the electron velocity will be of the form $\mathbf{v}(t) = \mathbf{v}_0\, \mathrm{e}^{-i\omega t}$. Then, (2) leads to :

$$\mathbf{v}_0 = -\frac{e\mathbf{E}_0}{m_e\left(\gamma_d - i\omega\right)}. \tag{3}$$

Combining the current density[2] $\mathbf{J} = -n_e e\mathbf{v}$ with (3) and Ohm's law[3] $\mathbf{J} = \sigma_d \mathbf{E}$, one obtains :

$$\sigma_d\left(\omega\right) = \frac{n_e e^2}{m_e}\frac{1}{\gamma_d - i\omega}. \tag{4}$$

The relative permittivity is linked to its conductivity with the relation $\varepsilon_{r,d}(\omega) = \varepsilon_\infty + \frac{i\sigma_d}{\omega\varepsilon_0}$. Here, ε_∞ represents the core electrons contribution, which is equivalent to approximate the positive ions lattice as a continuous medium of permittivity ε_∞. For the metals successfully described by the Drude model, ε_∞ is close to 1 (taking ε_∞ different from 1 is sometimes refered to as the Drude-Sommerfeld model). Eventually, one gets :

[1] τ_f is the electron mean free path.
[2] n_e stands for the electronic density.
[3] σ_d represents the conductivity of the considered electrons.

$$\varepsilon_{r,d}(\omega) = \varepsilon_\infty - \frac{\omega_d^2}{\omega^2 + i\omega\gamma_d}, \tag{5}$$

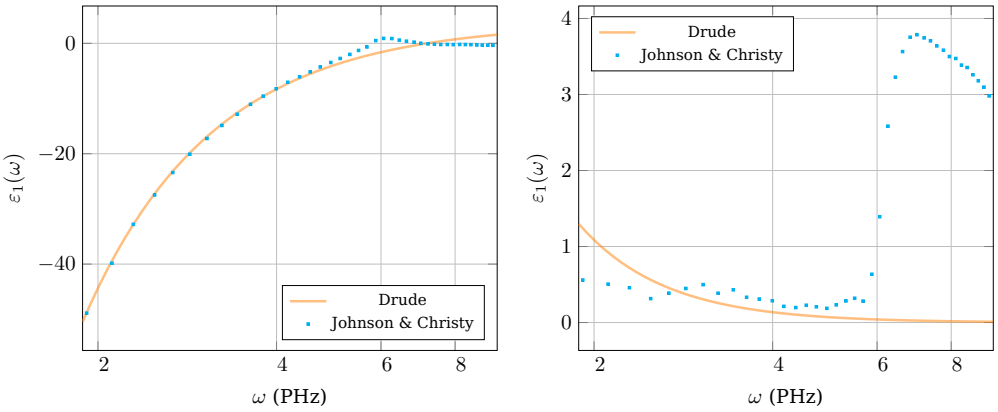
where $\omega_d = \sqrt{\frac{n_e e^2}{m_e \varepsilon_0}}$ is the plasma frequency of the electrons. The latter is linked to a typical relaxation time of slightly perturbed plasma electrons around their thermodynamic equilibrium state. By separating the real and imaginary parts of (5) :

$$\begin{aligned} \varepsilon_{r,d} &= \varepsilon_1 + i\varepsilon_2 \\ &= \varepsilon_\infty - \frac{\omega_d^2}{\gamma_d^2 + \omega^2} + i\frac{\gamma_d \omega_d^2}{\omega\left(\gamma_d^2 + \omega^2\right)}, \end{aligned}$$

and by considering $\omega \ll \gamma_d$ (which remains particularly accurate in the visible frequency field for the metals described by the Drude model), the above expressions can be simplified as :

$$\varepsilon_1(\omega) \simeq \varepsilon_\infty - \frac{\omega_d^2}{\omega^2} \text{ and } \varepsilon_2(\omega) \simeq \frac{\omega_d^2 \gamma_d}{\omega^3},$$

which enlights the role of the parameter γ_d in the dissipation induced by the Drude model[4]. The real and imaginary parts of the Drude permittivity function of the silver are plotted in figure 1, along with experimental curves from Johnson and Christy [JC72]. The used parameters can be found in table 1.



(a) Real part (b) Imaginary part

Figure 1 | Real and imaginary parts of the silver relative permittivity predicted by the Drude model compared to experimental data from Johnson & Christy.

One should notice that, if the real part fits the Drude model prediction, the experimental imaginary parts show features that are not predicted by the model. Those latter root in *interband* phenomenons, *i.e.* core electrons contributions that jump to the conduction band. These features can be better fitted with the Drude-Lorentz model, which is discussed in the following section.

[4] Indeed, γ_d represents the friction experienced by the electrons.

Table 1 | Drude parameters for silver in the [300, 1500] THz frequency range.

ε_∞	ω_d	γ_d
–	*GHz*	*GHz*
3.7362	$1.3871\,10^7$	$4.5154\,10^4$

1.3 Drude-Lorentz model

The initial Drude model describes the behavior of the free electrons contained in the conduction band. Nevertheless, for certain metals (such as noble ones), electronic transitions between valence and conduction band occur in the visible frequency range, giving them their particular colors. The contributions of these electrons can be taken into account in the permittivity function with additional Lorentz terms. Contrary to the Drude-described electrons, the Lorentz ones are, in some way, "bound" to their ion cores. Therefore, it seems logical to reuse the expression (2), including an additional spring term to it :

$$\frac{\partial^2 \mathbf{x}}{\partial t^2} + \gamma_l \frac{\partial \mathbf{x}}{\partial t} + \omega_l^2 \mathbf{x} = -\frac{e}{m_e}\mathbf{E}(t).$$

Following the same development as for the Drude model, one easily obtains the expression of the Lorentz permittivity :

$$\varepsilon_{r,l}(\omega) = -\frac{\Delta\varepsilon\omega_l^2}{\omega^2 - \omega_l^2 + i\omega\gamma_l}.$$

The total permittivity can then be written by adding the Drude and Lorentz terms[5] :

$$\varepsilon_{r,dl}(\omega) = \varepsilon_\infty - \frac{\omega_d^2}{\omega^2 + i\omega\gamma_d} - \frac{\Delta\varepsilon\omega_l^2}{\omega^2 - \omega_l^2 + i\omega\gamma_l}. \tag{6}$$

As done previously, the previous permittivity can be splitted into a real and an imaginary parts :

$$\varepsilon_1(\omega) = \varepsilon_\infty - \frac{\omega_d^2}{\gamma_d^2 + \omega^2} - \frac{\Delta\varepsilon\omega_l^2\left(\omega^2 - \omega_l^2\right)}{\left(\omega^2 - \omega_l^2\right)^2 + \gamma_l^2\omega^2},$$

and

$$\varepsilon_2(\omega) = \frac{\gamma_d\omega_d^2}{\omega\left(\gamma_d^2 + \omega^2\right)} + \frac{\Delta\varepsilon\omega_l^2\gamma_l\omega}{\left(\omega^2 - \omega_l^2\right)^2 + \gamma_l^2\omega^2}.$$

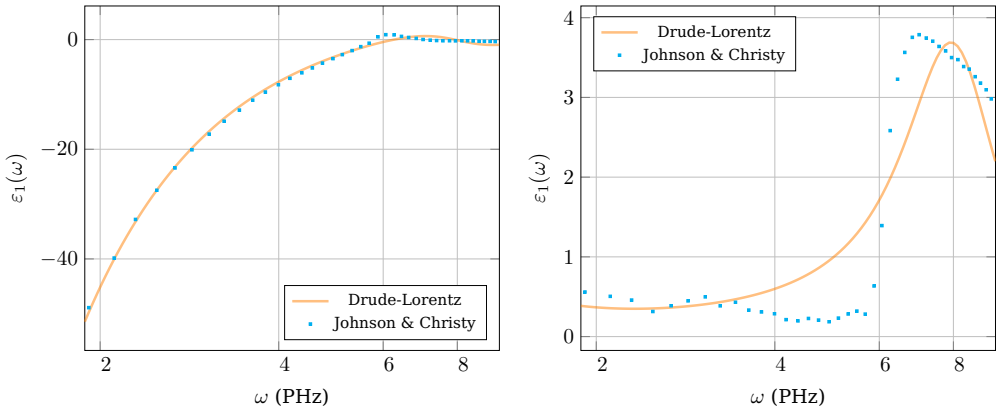
For some metals such as gold, silver or copper, the addition of Lorentz terms brings a much better fit between experimental and theoretical values. In the case of silver, a Drude-Lorentz model with one Lorentz pole brings a much better adequation between experiment and theory, especially in the high-frequency range for the imaginary part, as can be seen on figure 3. The Drude-Lorentz model parameters used for this fitting are summed up in table 2.

[5] It is of course possible to add more Lorentz poles, in order to describe more electronic transition resonances.

Table 2 | Drude-Lorentz parameters for silver in the [300, 1500] THz frequency range.

ε_∞	ω_d	γ_d	$\Delta\varepsilon$	ω_l	γ_l
–	*GHz*	*GHz*	–	*GHz*	*GHz*
2.7311	$1.4084\,10^7$	$6.6786\,10^3$	1.6336	$8.1286\,10^6$	$3.6448\,10^6$

Nevertheless, it is to be understood that Drude and Drude-Lorentz parameters are fitted from experimental data *over a particular range of frequencies*. This has several consequences, such as (i) the variety of parameter sets that can be found in the literature for the different models (particular attention must be paid to the frequency range of interest); (ii) in regions of lower regularity of the permittivity function, additional adequatly-chosen poles can be considered to improve the fitting (the next section presents a way of fitting arbitrary permittivity functions).



(a) Real part (b) Imaginary part

Figure 2 | Real and imaginary parts of the silver relative permittivity predicted by the Drude-Lorentz model compared to experimental data from Johnson & Christy.

1.4 Generalized dispersive model

Given an experimental permittivity function, a Padé approximant would be a convenient analytical coefficient-based function to approach the experimental data. Thus, one could write :

$$\varepsilon_{r,g}(\omega) = \frac{\sum_{i=1}^{n} \alpha_i (j\omega)^i}{\sum_{k=1}^{m} \alpha_k (j\omega)^k}.$$

The fundamental theorem of algebra allows to write the previous expression as a decomposition of a constant, one zero-order pole (ZOP), a set of first-order generalized poles (FOGP), and a set of second-order generalized poles (SOGP) :

$$\varepsilon_{r,g}(\omega) = \varepsilon_\infty - \frac{\sigma}{j\omega} - \sum_{l\in L_1} \frac{a_l}{j\omega - b_l} - \sum_{l\in L_2} \frac{c_l - j\omega d_l}{\omega^2 - e_l + j\omega f_l}, \tag{7}$$

where $\varepsilon_\infty, \sigma, a_l, b_l, c_l, d_l, e_l, f_l$ are real constants, and L_1, L_2 are non-overlapping set of indices. This general writing allows an important flexibility for two reasons : (i) it unifies most of the common dispersion models in a single formulation (the right coefficient choices are presented in table 3 for the Drude, Drude-Lorentz and Debye media), and (ii) it permits to fit any experimental data set in a reasonable number of poles (and thus a reasonable number of coefficients), as will be presented later. One should notice that this approach is very similar to the Critical Points (CP) [VLDC11] and the Complex-Conjugate Pole-Residue Pairs (CCPRP) [HDF06]. In facts, these developments are the same in some mathematical sense, since the three of them exploit the fundamental theorem of algebra. The difference between them lies in the way they choose to use it : the CCPRP and the CP allow complex coefficients in their developments, and can therefore write the decomposition of the permittivity function in terms of single-order poles only, whereas choosing real coefficients leads to a collection of first-order and second-order poles.

Table 3 | Parameters of the generally dispersive permittivity function for the Drude, Drude-Lorentz and Debye models.

Model	a_l	b_l	c_l	d_l	e_l	f_l
Drude	0	0	ω_d^2	0	0	γ_d
Drude-Lorentz	0	0	$\Delta\varepsilon\omega_l^2$	0	ω_l^2	γ_l
Debye	$\Delta\varepsilon\gamma_{de}$	$-\gamma_{de}$	0	0	0	0

The matter of fitting the coefficients of (7) to experimental data remains to be detailed. Various techniques can be used, through which the many existing versions of least square method. Vector fitting techniques [GS99] are also well developed for the CCPRP formulation. In our case, a Simulated Annealing (SA) algorithm has been used to determine the coefficients to use given an experimental data set. The original paper describing this method is [KGV83], but the reader can find extensive presentations and discussions by himself over the internet. Notice that the used algorithm has been written by William L. Goffe, and can be found at `http://ideas.repec.org/c/wpa/wuwppr/9406001.html`.

A set of coefficients has been calculated for the silver in the same range of frequencies as in the previous sections, using two SOGP. The values of the coefficients are presented in table 4, whereas a plot of the real and imaginary parts of the permittivity function is presented in figure 2. This plot is to be compared with the Drude-Lorentz one (see figure 3), which can also be seen as a two second-order poles fitting. Although the general expression (7) is not based on a physical model, it seems to display better fittings properties than classical poles such as Lorentz ones. As in the Lorentz case, additional poles can enhance the precision of the fitting.

1.5 Summary

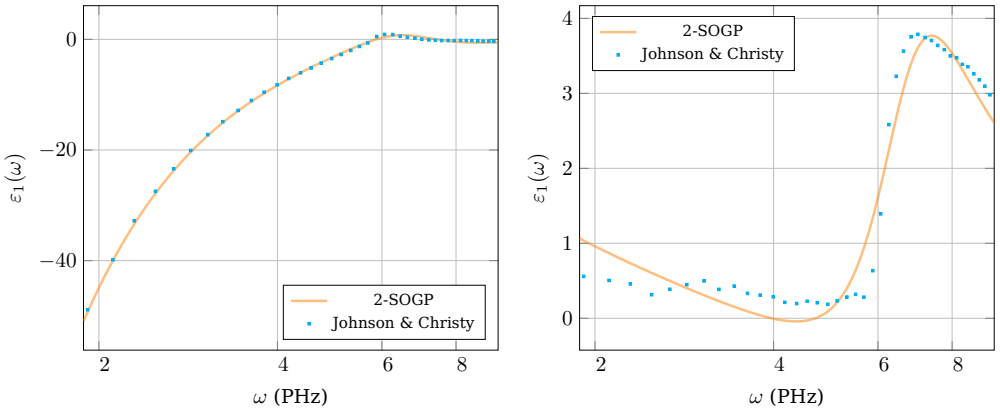
The physical origins of the dispersion for propagating waves have been briefly outlined and common models for EMW dispersion in metals have been presented. A generalized disper-

sion models has been proposed that seems to improve the fitting of permittivity function over arbitrary sets of experimental data.

The next part of this report focuses on the Discontinuous Galerkin (DG) method in the time domain, and presents its formulation in the case of the Maxwell's equations in vacuum. Then, the respective continuous equations and Discontinuous Galerkin Time Domain (DGTD) schemes are deduced for the Drude model, and for the generalized dispersive model.

Table 4 | 2-SOGP parameters for silver in the [300, 1500] THz frequency range.

ε_∞	c_1	d_1	e_1	f_1	c_2	d_2	e_2	f_2
–	GHz^2	GHz	GHz^2	GHz	GHz^2	GHz	GHz^2	GHz
1.2944	$1.8909\,10^{14}$	$2.6584\,10^{6}$	0.0	0.0	$5.6165\,10^{13}$	$1.2005\,10^{7}$	$4.3932\,10^{13}$	$3.1709\,10^{6}$



(a) Real part (b) Imaginary part

Figure 3 | Real and imaginary parts of the silver relative permittivity predicted by the 2-SOGP model compared to experimental data from Johnson & Christy.

2 DGTD method for non-dispersive media

This section focuses on the DGTD method for Maxwell equations. Let $\Omega \subset \mathbb{R}^3$ be a bounded convex domain, and $\mathbf{n}$ the outward normal to its boundary $\partial\Omega$. The electric permittivity and permeability of free space are respectively denoted ε_0 and μ_0. Then, in free space, the Maxwell equations can be written as follows :

$$\begin{cases} \mu_0 \dfrac{\partial \mathbf{H}}{\partial t} &= -\nabla \times \mathbf{E}, \\ \varepsilon_0 \dfrac{\partial \mathbf{E}}{\partial t} &= \nabla \times \mathbf{H} - \mathbf{J}, \end{cases} \tag{8}$$

along with constitutive relations $\mathbf{B} = \mu_0 \mathbf{H}$ and $\mathbf{D} = \varepsilon_0 \mathbf{E}$ and metallic boundary conditions.

DG methods have been originally introduced in 1973 by Reed and Hill [RH73], and has been widely used since in the computational fluid dynamics field. However, their application to the time-domain Maxwell equations are more recent. DG methods can be seen as classical finite element methods (FEM) for which the global continuity of the approximation has been lifted. This implies that the support of each basis function is restrained to a discretization cell, which leads to local formulations implying no large mass matrix inversion in the process. Afterward, connexion between the cells is restored by the use of a numerical flux to evaluate the boundary integrals. The choice of the numerical flux has a great influence on the mathematical properties of the DG discretization, as energy preservation, for example. Both centered [FLLP05] or upwind [HW02] fluxes can be used.

The discontinuity of the approximation makes room for numerous methodologic improvements, such as local approximation orders [Fah09], high parallelization features ([BFLP06], [CCL11]) and the use of non-conformeous [FL10] and hybrid meshes [DLS12], for example. Also, a wide choice of time-integration schemes can be used for the discretization of time derivatives, including Leap-Frog (LF) and Runge-Kutta (RK). As well as local approximation orders can be used, local time-stepping [Pip05] as well as locally implicit formulations [Moy12] are some of the main features that have been studied during the last years in the DG framework. More recently, Space-Time Discontinuous Galerkin methods (STDG) have emerged. These formulations exploit a temporal discretization similar to the spatial one instead of exploiting classical advancing-in-time schemes like LF or RK.

The following section presents the non-dissipative DGTD method originally developped by Fezoui *et al.* in [FLLP05]. Let Ω_h be a discretization of Ω, relying on a quasi-uniform triangulation $\mathcal{T}_h$ verifying $\mathcal{T}_h = \bigcup_{i=1}^{N} T_i$. The internal faces of the discretization are denoted $a_{ik} = T_i \bigcap T_k$, and $\mathbf{n}_{ik}$ is defined as the unit normal vector to the face a_{ik}, oriented from T_i toward T_k. For each cell T_i, $\mathcal{V}_i$ is the set of indices $\{k \mid T_i \bigcap T_k \neq \emptyset\}$. Then, the quasi-uniform assumption implies that :

$$\forall T_i \in \mathcal{T}_h, \forall k \in \mathcal{V}_i, \exists \delta, h_k \leq \delta_h h_i.$$

The semi-discrete fields are denoted $(\mathbf{H}_h, \mathbf{E}_h, \mathbf{J}_h)$, and on each cell T_i the restrictions $(\mathbf{H}_i, \mathbf{E}_i, \mathbf{J}_i) = \left(\mathbf{H}_h|_{T_i}, \mathbf{E}_h|_{T_i}, \mathbf{J}_h|_{T_i}\right)$ are defined.

For each cell, a set of scalar basis functions $(\phi_{ij})_{1 \leqslant j \leqslant d_i}$ is defined, where d_i is the number of degrees of freedom (d.o.f.) per dimension (therefore, a 3D problem implies $3\,d_i$ d.o.f. per cell).

2.1 Weak formulation

It is now possible to write the weak formulation of the problem (8) in a cell T_i. By taking the dot-product of each term with a *vectorial* test function $\boldsymbol{\psi}$ and then integrating over the cell, one obtains the following variational problem :

$$\text{Find } (\mathbf{E},\mathbf{H}) \in H_0(\mathbf{rot},\Omega) \times H(\mathbf{rot},\Omega) \text{ such that } \forall\, \boldsymbol{\psi} \in H(\mathbf{rot},\Omega),$$

$$\begin{cases} \displaystyle\int_{T_i} \mu_0 \frac{\partial \mathbf{H}}{\partial t} \cdot \boldsymbol{\psi} + \int_{T_i} \nabla \times \mathbf{E} \cdot \boldsymbol{\psi} \;=\; \mathbf{0}, \\ \displaystyle\int_{T_i} \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \cdot \boldsymbol{\psi} - \int_{T_i} \nabla \times \mathbf{H} \cdot \boldsymbol{\psi} \;=\; -\int_{T_i} \mathbf{J} \cdot \boldsymbol{\psi}, \end{cases}$$

which can be rewritten as follows using classical vectorial calculus and Green formulae :

$$\begin{cases} \displaystyle\int_{T_i} \mu_0 \frac{\partial \mathbf{H}}{\partial t} \cdot \boldsymbol{\psi} + \int_{T_i} \mathbf{E} \cdot \nabla \times \boldsymbol{\psi} \;=\; \int_{\partial T_i} (\boldsymbol{\psi} \times \mathbf{E}) \cdot \mathbf{n}_i, \\ \displaystyle\int_{T_i} \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \cdot \boldsymbol{\psi} - \int_{T_i} \mathbf{H} \cdot \nabla \times \boldsymbol{\psi} \;=\; -\int_{T_i} \mathbf{J} \cdot \boldsymbol{\psi} - \int_{\partial T_i} (\boldsymbol{\psi} \times \mathbf{H}) \cdot \mathbf{n}_i. \end{cases}$$

One immediatly notices that the previous equality only holds if the boundary terms exist. Considering the properties of the mixed product, the latter becomes :

$$(\boldsymbol{\psi} \times \mathbf{E}) \cdot \mathbf{n}_i = (\mathbf{E} \times \mathbf{n}_i) \cdot \boldsymbol{\psi},$$

which implies that taking $\mathbf{E}$ in $H_0(\mathbf{rot},\Omega)$ requires the normal *and* tangential trace of $\boldsymbol{\psi}$ on ∂T_i to exist. This implies that one should take $\boldsymbol{\psi}$ in $H^1(\Omega)$ instead of $H(\mathbf{rot},\Omega)$.

Remark *:*

A set of d_i scalar basis functions ϕ_{ij} have previously been defined on the cell T_i. Nevertheless, the test functions are to be naturally chosen as vectorial elements. Therefore, one should define three vectorial basis functions (for the three space dimensions) for each scalar one. We now denote :

$$\boldsymbol{\phi}^{\mathbf{1}}_{\boldsymbol{ij}} = \begin{bmatrix} \phi_{ij} \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\phi}^{\mathbf{2}}_{\boldsymbol{ij}} = \begin{bmatrix} 0 \\ \phi_{ij} \\ 0 \end{bmatrix} \text{ and } \boldsymbol{\phi}^{\mathbf{3}}_{\boldsymbol{ij}} = \begin{bmatrix} 0 \\ 0 \\ \phi_{ij} \end{bmatrix}.$$

2.2 Space discretization

2.2.1 Volumic integrals

We now seek the approximations $\mathbf{E}_h$ and $\mathbf{H}_h$ of $\mathbf{E}$ and $\mathbf{H}$ in the following approximation space :

$$V_h = \left\{ v \in \left(L^2(\Omega)\right)^3, v_{|T_i} \in \left(\mathbb{P}_p(T_i)\right)^3 \forall\, T_i \in \mathcal{T}_h \right\}, \tag{9}$$

where $\mathbb{P}_p(T_i)$ is the space of polynomials of maximum degree p on T_i. The contribution of each cell is therefore defined as $\mathbf{E}_i = \mathbf{E}_{h\,|T_i}$, where $\mathbf{E}_i$ is locally expanded as :

$$\mathbf{E}_i = \sum_{d=1}^{3} \sum_{j}^{d_i} E_{ij}^d \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}}$$

We now choose the test functions $\boldsymbol{\psi}$ equal to the $3\,d_i$ vectors $\boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}}$. The spatial discretization of $\displaystyle\int_{T_i} \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \cdot \boldsymbol{\psi}$ then leads to $3\,d_i$ terms of the following form :

$$\int_{T_i} {\varepsilon_0}^T \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} \cdot \boldsymbol{\phi}_{\boldsymbol{il}}^{\boldsymbol{d}} \frac{\partial E_{ij}^d}{\partial t}. \tag{10}$$

The previous term can be cast under a matrix form $\mathbb{M}_i^{\varepsilon_0} \overline{\mathbf{E}}_i$, where the mass matrix is block diagonal of size $3\,d_i \times 3\,d_i$:

$$\mathbb{M}_i^{\varepsilon_0} = \begin{bmatrix} \widetilde{\mathbb{M}}_i^{\varepsilon_0} & 0 & 0 \\ 0 & \widetilde{\mathbb{M}}_i^{\varepsilon_0} & 0 \\ 0 & 0 & \widetilde{\mathbb{M}}_i^{\varepsilon_0} \end{bmatrix}, \text{ with } \left(\widetilde{\mathbb{M}}_i^{\varepsilon_0}\right)_{jl} = \int_{T_i} {\varepsilon_0}^T \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} \cdot \boldsymbol{\phi}_{\boldsymbol{il}}^{\boldsymbol{d}}$$

The vector $\overline{\mathbf{E}}_i$ introduced previously has $3\,d_i$ components, and is defined as follows :

$$\overline{\mathbf{E}}_i = \begin{bmatrix} \left(E_{ij}^1\right)_{1 \leqslant j \leqslant d_i} \\ \left(E_{ij}^2\right)_{1 \leqslant j \leqslant d_i} \\ \left(E_{ij}^3\right)_{1 \leqslant j \leqslant d_i} \end{bmatrix}$$

From now on, the mass matrices with an exponent should be understood as :

$$\left(\widetilde{\mathbb{M}}_i^x\right)_{jl} = \int_{T_i} x^T \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} \cdot \boldsymbol{\phi}_{\boldsymbol{il}}^{\boldsymbol{d}},$$

and their associated block matrices denoted $\mathbb{M}_i^x$.

2.2.2 Surface integrals

Given that **E** and **H** are discontinuous at cell boundaries, the surface integrals need a special treatment. We choose a centered approximation, which permits to rewrite $\displaystyle\int_{\partial T_i} (\boldsymbol{\psi} \times \mathbf{E}) \cdot \mathbf{n}$ as :

$$\sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \left(\boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} \times \frac{\mathbf{E}_i + \mathbf{E}_k}{2} \right) \cdot \mathbf{n}_{ik},$$

with a similar choice for the surface integral involving **H**.

2.2.3 Semi-discrete equations

Taking into account the previous definitions, one obtains the following $6\,d_i$ scalar equations :

$$\begin{cases} \left(\mathbb{M}_i^{\mu_0} \dfrac{\partial \overline{\mathbf{H}}_i}{\partial t}\right)_j + \displaystyle\int_{T_i} \mathbf{E}_i \cdot \nabla \times \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} = \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \left(\boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} \times \dfrac{\mathbf{E}_i + \mathbf{E}_k}{2}\right) \cdot \mathbf{n}_{ik}, \\ \left(\mathbb{M}_i^{\varepsilon_0} \dfrac{\partial \overline{\mathbf{E}}_i}{\partial t}\right)_j - \displaystyle\int_{T_i} \mathbf{H}_i \cdot \nabla \times \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} = -\int_{T_i} \mathbf{J}_i \cdot \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} - \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \left(\boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} \times \dfrac{\mathbf{H}_i + \mathbf{H}_k}{2}\right) \cdot \mathbf{n}_{ik}. \end{cases} \tag{11}$$

Performing an integration by parts on the surface integrals leads to :

$$\begin{cases} \left(\mathbb{M}_i^{\mu_0} \dfrac{\partial \overline{\mathbf{H}}_i}{\partial t}\right)_j = -\dfrac{1}{2} \displaystyle\int_{T_i} \left(\mathbf{E}_i \cdot \nabla \times \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} + \nabla \times \mathbf{E}_i \cdot \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}}\right) + \dfrac{1}{2} \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} \cdot (\mathbf{E}_k \times \mathbf{n}_{ik}), \\ \left(\mathbb{M}_i^{\varepsilon_0} \dfrac{\partial \overline{\mathbf{E}}_i}{\partial t}\right)_j = \dfrac{1}{2} \displaystyle\int_{T_i} \left(\mathbf{H}_i \cdot \nabla \times \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} + \nabla \times \mathbf{H}_i \cdot \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}}\right) \\ \qquad\qquad - \displaystyle\int_{T_i} \mathbf{J}_i \cdot \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} - \dfrac{1}{2} \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} \cdot (\mathbf{H}_k \times \mathbf{n}_{ik}). \end{cases} \tag{12}$$

Reasoning similarly as for (10) permits to rewrite (12) as :

$$\begin{cases} \mathbb{M}_i^{\mu_0} \dfrac{\partial \overline{\mathbf{H}}_i}{\partial t} = -\mathbb{K}_i \overline{\mathbf{E}}_i + \displaystyle\sum_{k \in \mathcal{V}_i} \mathbb{S}_{ik} \overline{\mathbf{E}}_k, \\ \mathbb{M}_i^{\varepsilon_0} \dfrac{\partial \overline{\mathbf{E}}_i}{\partial t} = \mathbb{K}_i \overline{\mathbf{H}}_i - \displaystyle\sum_{k \in \mathcal{V}_i} \mathbb{S}_{ik} \overline{\mathbf{H}}_k - \mathbb{M}_i \overline{\mathbf{J}}_i, \end{cases}$$

where the $3\,d_i \times 3\,d_i$ local stiffness matrix and the $3\,d_i \times 3\,d_k$ surface matrix are respectively defined by :

$$\left(\widetilde{\mathbb{K}}_i\right)_{jl} = \frac{1}{2} \int_{T_i} \left(\boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} \cdot \nabla \times \boldsymbol{\phi}_{\boldsymbol{il}}^{\boldsymbol{d}} + \boldsymbol{\phi}_{\boldsymbol{il}}^{\boldsymbol{d}} \cdot \nabla \times \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}}\right) \text{ and } \left(\widetilde{\mathbb{S}}_{ik}\right)_{jl} = \frac{1}{2} \int_{T_i} \boldsymbol{\phi}_{\boldsymbol{ij}}^{\boldsymbol{d}} \cdot \left(\boldsymbol{\phi}_{\boldsymbol{kl}}^{\boldsymbol{d}} \times \mathbf{n}_{ik}\right).$$

As mentioned previously, the basis functions are chosen in the approximation space V_h. A common choice consists in using Lagrange polynomials, though other choices are possible [CFL10]. A set of $p_i + 1$ interpolation nodes $(x_j)_{0 \leqslant j \leqslant p_i}$ is defined in the cell, and the basis functions are then the Lagrange function $L_k^p(x)$ equal to 1 on the x_k node, and to 0 on all the other $x_j, j \neq k$.

2.3 Time discretization

For the time derivatives, a second-order leap frog scheme (LF2) is used, where the $\overline{\mathbf{E}}_i$ are evaluated at the time station $t^n = n\Delta t$, whereas the $\overline{\mathbf{H}}_i, \overline{\mathbf{J}}_i$ are evaluated at the time station $t^{n+\frac{1}{2}} = \left(n + \frac{1}{2}\right) \Delta t$. This leads to seek the values of $\overline{\mathbf{E}}_i^{n+1}$ and $\overline{\mathbf{H}}_i^{n+\frac{3}{2}}$ when knowing those of $\overline{\mathbf{E}}_i^{n}$, $\overline{\mathbf{H}}_i^{n+\frac{1}{2}}$ and $\overline{\mathbf{J}}_i^{n+\frac{1}{2}}$ with the following discretization :

$$\begin{cases} \dfrac{\mathbb{M}_i^{\mu_0}}{\Delta t}\left(\overline{\mathbf{H}}_i^{\,n+\frac{3}{2}} - \overline{\mathbf{H}}_i^{\,n+\frac{1}{2}}\right) = -\mathbb{K}_i\,\overline{\mathbf{E}}_i^{\,n+1} + \displaystyle\sum_{k\in\mathcal{V}_i} \mathbb{S}_{ik}\,\overline{\mathbf{E}}_k^{\,n+1}, \\ \dfrac{\mathbb{M}_i^{\varepsilon_0}}{\Delta t}\left(\overline{\mathbf{E}}_i^{\,n+1} - \overline{\mathbf{E}}_i^{\,n}\right) = \mathbb{K}_i\,\overline{\mathbf{H}}_i^{\,n+\frac{1}{2}} - \displaystyle\sum_{k\in\mathcal{V}_i} \mathbb{S}_{ik}\,\overline{\mathbf{H}}_k^{\,n+\frac{1}{2}} - \bar{\mathbf{J}}_i^{\,n+\frac{1}{2}}. \end{cases} \tag{13}$$

Remark *: Different time discretization can be used, like fourth-order leap frog (LF4) scheme or second and fourth-order Runge-Kutta (RK2 and RK4) schemes, under their respective stability conditions.*

3 DGTD formulation in dispersive media

The case of Maxwell's equations in vacuum has been presented in the previous section. After a brief review of existing works, the extension of the DGTD formulation to Drude-like media is presented. Then, the same treatment is done for the generalized dispersive model.

3.1 Previous works

Before they were exploited with DGTD methods, dispersion models have been extensively used in the context of FDTD, and a very large amount of references are therefore available on this topic. We only settle here for giving a very few of them, such as [OO06] that presents an ADE-FDTD (Auxiliary Differential Equation) algorithm with the Drude model, or [LSK93] that uses a RC-FDTD (Recursive Convolution) method for Debye, Drude and Lorentz models.

Although the exploitation of dispersion models in the DGTD framework is not as rich as the FDTD one, an important amount of studies have already been conducted. In [LS12], the authors study a DGTD-CF (Centered Fluxes) method for the Maxwell equations coupled with a Debye dispersion model. Stability and convergence are proved, and a bound on the error is given. In [JCZ07], the authors study the 2D Maxwell equations in a Drude-like medium with a DGTD-CF of the fourth order in space, and with a RK4 scheme in time.

A significant number of contributions on the numerical analysis of the schemes for dispersive media models have been made by J. Li, such as [LCE08] and [Li09] to name a few. Numerous applications of the DG method in the area photonics have been issued : [BKN11] presents a DGTD and DGFD (Discontinuous Galerkin Frequency Domain) method as well as realistic cases, and discuss topics like sources treatment and boundary conditions, like absorbing ones (ABC). In [SKNB09], the authors focus on the field enhancements observed in the vicinity of metallic V-shaped nanostructures described by Drude model. A nice overview of the DGTD method coupled with Drude and Drude-Lorentz models can be found in [Die12], where the author exploit a GPU implementation to focuses on the analysis of various nanostructures features.

3.2 DGTD method in Drude-like media

3.2.1 Maxwell-Drude equations

We now consider the case of a frequency-dependent medium, under the hypothesis of a Drude-Sommerfeld model :

$$\varepsilon_r(\omega) = \varepsilon_\infty - \frac{\omega_d^2}{\omega^2 + i\omega\gamma}.$$

Considering a constant permeability and a homogeneous and isotropic medium, one writes the general Maxwell equations as follows :

$$\begin{cases} \nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t}, \\ \nabla \times \mathbf{H} = \dfrac{\partial \mathbf{D}}{\partial t}, \end{cases} \tag{14}$$

along with the following constitutive relations :

$$\begin{cases} \mathbf{D} = \varepsilon_0 \varepsilon_\infty \mathbf{E} + \mathbf{P}, \\ \mathbf{B} = \mu_0 \mathbf{H}. \end{cases} \tag{15}$$

Combining (14) and (15) leads to :

$$\begin{cases} \nabla \times \mathbf{E} = -\mu_0 \dfrac{\partial \mathbf{H}}{\partial t}, \\ \nabla \times \mathbf{H} = \varepsilon_0 \varepsilon_\infty \dfrac{\partial \mathbf{E}}{\partial t} + \dfrac{\partial \mathbf{P}}{\partial t}. \end{cases}$$

In the frequential domain, and under the aforementioned hypothesis, the polarization $\mathbf{P}$ is linked to the electric field through $\hat{\mathbf{P}} = -\frac{\varepsilon_0 \omega_d^2}{\omega^2 + i\gamma_d \omega} \hat{\mathbf{E}}$, where $\hat{\cdot}$ denotes the Fourier transform of the associated field in the temporal domain. An inverse Fourier transform gives :

$$\frac{\partial^2 \mathbf{P}}{\partial t^2} + \gamma_d \frac{\partial \mathbf{P}}{\partial t} = \varepsilon_0 \omega_d^2 \mathbf{E}. \tag{16}$$

By defining the dipolar current vector $\mathbf{J}_p = \dfrac{\partial \mathbf{P}}{\partial t}$, (14)-(16) can be rewritten as follows :

$$\begin{cases} \mu_0 \dfrac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E}, \\ \varepsilon_0 \varepsilon_\infty \dfrac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{H} - \mathbf{J}_p, \\ \dfrac{\partial \mathbf{J}_p}{\partial t} + \gamma_d \mathbf{J}_p = \varepsilon_0 \omega_d^2 \mathbf{E}. \end{cases} \tag{17}$$

3.2.2 Normalization

The previous set of equations is now normalized. A normalized variable is denoted $\widetilde{X}$ when the original variable is X. First, we define vacuum impedance and light velocity :

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \text{ and } c_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}.$$

Then, the following substitutions are applied :

$$\widetilde{\mathbf{H}} = Z_0 \mathbf{H},\ \widetilde{\mathbf{E}} = \mathbf{E},\ \widetilde{\mathbf{J}}_p = Z_0 \mathbf{J}_p,\ \widetilde{t} = c_0 t,\ \widetilde{\gamma}_d = \frac{\gamma_d}{c_0} \text{ and } \widetilde{\omega}_d^2 = \frac{\omega_d^2}{c_0^2}.$$

One can now rewrite (17) in the form :

$$\begin{cases} \dfrac{\mu_0 c_0}{Z_0} \dfrac{\partial \widetilde{\mathbf{H}}}{\partial \widetilde{t}} = -\nabla \times \widetilde{\mathbf{E}}, \\ \varepsilon_0 c_0 Z_0 \varepsilon_\infty \dfrac{\partial \widetilde{\mathbf{E}}}{\partial \widetilde{t}} = \nabla \times \widetilde{\mathbf{H}} - \widetilde{\mathbf{J}}_p, \\ \dfrac{\partial \widetilde{\mathbf{J}}_p}{\partial \widetilde{t}} + \widetilde{\gamma}_d \widetilde{\mathbf{J}}_p = \dfrac{Z_0 \varepsilon_0 c_0^2}{c_0} \widetilde{\omega}_d^2 \widetilde{\mathbf{E}}. \end{cases}$$

Using the equalities $\dfrac{\mu_0 c_0}{Z_0} = 1$ and $\varepsilon_0 c_0 Z_0 = 1$, one gets :

$$\begin{cases} \dfrac{\partial \widetilde{\mathbf{H}}}{\partial \widetilde{t}} &= -\nabla \times \widetilde{\mathbf{E}}, \\ \dfrac{\partial \widetilde{\mathbf{E}}}{\partial \widetilde{t}} &= \dfrac{1}{\varepsilon_\infty}\left(\nabla \times \widetilde{\mathbf{H}} - \widetilde{\mathbf{J}}_p\right), \\ \dfrac{\partial \widetilde{\mathbf{J}}_p}{\partial \widetilde{t}} &= \widetilde{\omega}_d^2 \widetilde{\mathbf{E}} - \widetilde{\gamma}_d \widetilde{\mathbf{J}}_p. \end{cases} \tag{18}$$

It could be noticed that the first two equations are expressed in $V.m^{-2}$, whereas the third one is in $V.m^{-3}$.

Remark *: To simplify the writings, the $\widetilde{X}$ notation for the normalized variables will be omitted from now on.*

3.2.3 DGTD formulation for the Maxwell-Drude equations

The extension of the DGTD method to the Maxwell-Drude equations is straightforward from the classical Maxwell equations discretization (see section 2). The last equation of (18) leads to a simple vectorial equation :

$$\frac{1}{\Delta t}\left(\overline{\mathbf{J}}_i^{\,n+\frac{3}{2}} - \overline{\mathbf{J}}_i^{\,n+\frac{1}{2}}\right) = -\gamma_d \overline{\mathbf{J}}_i^{\,n+\frac{1}{2}} + \omega_d^2 \overline{\mathbf{E}}_i^{\,n+1}.$$

The whole system can then be written as :

$$\begin{cases} \dfrac{\mathbb{M}_i}{\Delta t}\left(\overline{\mathbf{H}}_i^{\,n+\frac{3}{2}} - \overline{\mathbf{H}}_i^{\,n+\frac{1}{2}}\right) &= -\mathbb{K}_i \overline{\mathbf{E}}_i^{\,n+1} + \displaystyle\sum_{k\in\nu_i} \mathbb{S}_{ik} \overline{\mathbf{E}}_k^{\,n+1}, \\ \dfrac{\mathbb{M}_i^{\varepsilon_\infty}}{\Delta t}\left(\overline{\mathbf{E}}_i^{\,n+1} - \overline{\mathbf{E}}_i^{\,n}\right) &= \mathbb{K}_i \overline{\mathbf{H}}_i^{\,n+\frac{1}{2}} - \displaystyle\sum_{k\in\nu_i} \mathbb{S}_{ik} \overline{\mathbf{H}}_k^{\,n+\frac{1}{2}} - \mathbb{M}_i \overline{\mathbf{J}}_i^{\,n+\frac{1}{2}}, \\ \dfrac{1}{\Delta t}\left(\overline{\mathbf{J}}_i^{\,n+\frac{3}{2}} - \overline{\mathbf{J}}_i^{\,n+\frac{1}{2}}\right) &= \omega_d^2 \overline{\mathbf{E}}_i^{\,n+1} - \dfrac{\gamma_d}{2}\left(\overline{\mathbf{J}}_i^{\,n+\frac{3}{2}} + \overline{\mathbf{J}}_i^{\,n+\frac{1}{2}}\right). \end{cases} \tag{19}$$

3.3 DGTD method for the generalized dispersive model

A similar development as the one of the previous section can be made for the general permittivity function (7).

3.3.1 Continuous equations

As in the Drude case, polarizations and currents are introduced to account for the dispersive behavior in the temporal domain. The (normalized) continuous equations are therefore :

$$
\begin{cases}
\dfrac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E}, \\
\dfrac{\partial \mathbf{E}}{\partial t} = \dfrac{1}{\varepsilon_\infty}\left(\nabla \times \mathbf{H} - \left(\mathbf{J}_0 + \displaystyle\sum_{l \in L_1} \mathbf{J}_l + \sum_{l \in L_2} \mathbf{J}_l\right)\right), \\
\mathbf{J}_0 = \sigma \mathbf{E}, \\
\mathbf{J}_l = a_l \mathbf{E} - b_l \mathbf{P}_l & \forall l \in L_1, \\
\dfrac{\partial \mathbf{P}_l}{\partial t} = \mathbf{J}_l & \forall l \in L_1, \\
\dfrac{\partial \mathbf{J}_l}{\partial t} = d_l \dfrac{\partial \mathbf{E}}{\partial t} + c_l \mathbf{E} - f_l \mathbf{J}_l - e_l \mathbf{P}_l & \forall l \in L_2, \\
\dfrac{\partial \mathbf{P}_l}{\partial t} = \mathbf{J}_l & \forall l \in L_2.
\end{cases}
\tag{20}
$$

3.3.2 DGTD formulation

In this formulation, the currents are evaluated at the even time stations, whereas the polarizations are evaluated at the odd ones. The resulting scheme in time is, as well as in the Drude case, of order two (see section 5.2). As one might notice, some fields have been substituted to avoid unnecessary calculations :

$$
\begin{cases}
\dfrac{\mathbb{M}_i}{\Delta t}\left(\overline{\mathbf{H}}_i^{n+\frac{3}{2}} - \overline{\mathbf{H}}_i^{n+\frac{1}{2}}\right) = -\mathbb{K}_i \overline{\mathbf{E}}_i^{n+1} + \displaystyle\sum_{k \in \nu_i} \mathbb{S}_{ik} \overline{\mathbf{E}}_k^{n+1}, \\
\dfrac{\mathbb{M}_i}{\Delta t}\left(\overline{\mathbf{E}}_i^{n+1} - \overline{\mathbf{E}}_i^{n}\right) = \dfrac{1}{\varepsilon_\infty}\left(\mathbb{K}_i \overline{\mathbf{H}}_i^{n+\frac{1}{2}} - \displaystyle\sum_{k \in \nu_i} \mathbb{S}_{ik} \overline{\mathbf{H}}_k^{n+\frac{1}{2}}\right. \\
\qquad\qquad - \overline{\sigma}\dfrac{\mathbb{M}_i}{2}\left(\overline{\mathbf{E}}_i^{n+1} + \overline{\mathbf{E}}_i^{n}\right) + \mathbb{M}_i \displaystyle\sum_{l \in L_1} b_l \overline{\mathbf{P}}_{l,i}^{n+\frac{1}{2}} \\
\qquad\qquad \left. - \dfrac{\mathbb{M}_i}{2} \displaystyle\sum_{l \in L_2}\left(\overline{\mathbf{J}}_{l,i}^{n+1} + \overline{\mathbf{J}}_{l,i}^{n}\right)\right), \\
\dfrac{1}{\Delta t}\left(\overline{\mathbf{P}}_{l,i}^{n+\frac{3}{2}} - \overline{\mathbf{P}}_{l,i}^{n+\frac{1}{2}}\right) = a_l \overline{\mathbf{E}}_i^{n+1} - \dfrac{b_l}{2}\left(\overline{\mathbf{P}}_{l,i}^{n+\frac{3}{2}} + \overline{\mathbf{P}}_{l,i}^{n+\frac{1}{2}}\right) \quad \forall l \in L_1, \\
\dfrac{1}{\Delta t}\left(\overline{\mathbf{P}}_{l,i}^{n+\frac{3}{2}} - \overline{\mathbf{P}}_{l,i}^{n+\frac{1}{2}}\right) = \overline{\mathbf{J}}_{l,i}^{n+1} \quad \forall l \in L_2, \\
\dfrac{1}{\Delta t}\left(\overline{\mathbf{J}}_{l,i}^{n+1} - \overline{\mathbf{J}}_{l,i}^{n}\right) = \dfrac{d_l}{\Delta t}\left(\overline{\mathbf{E}}_i^{n+1} - \overline{\mathbf{E}}_i^{n}\right) + \dfrac{c_l}{2}\left(\overline{\mathbf{E}}_i^{n+1} + \overline{\mathbf{E}}_i^{n}\right), \\
\qquad\qquad - \dfrac{f_l}{2}\left(\overline{\mathbf{J}}_{l,i}^{n+1} + \overline{\mathbf{J}}_{l,i}^{n}\right) - e_l \overline{\mathbf{P}}_{l,i}^{n+\frac{1}{2}} \quad \forall l \in L_2.
\end{cases}
\tag{21}
$$

where $\overline{\sigma} = \sigma + \displaystyle\sum_{l \in L_1} a_l$. The resulting updating scheme is presented in figure 4.

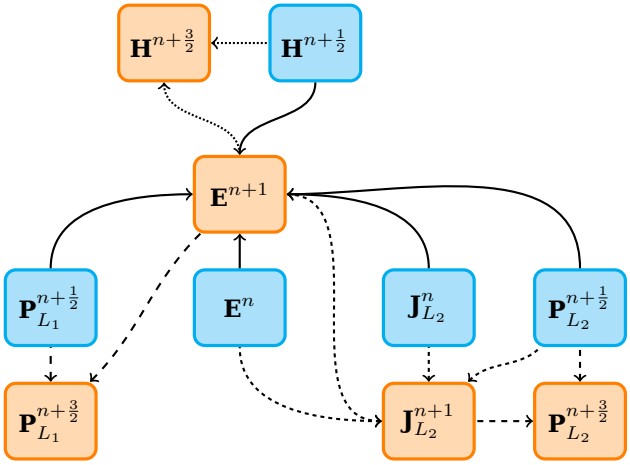


Figure 4 | Update scheme in the generally dispersive case.

4 Theoretical study of the Maxwell-Drude equations

A stability and convergence study of the Maxwell-Drude equations is now presented. This part is highly inspired from [LS12].

4.1 Stability study of the Maxwell-Drude equations

4.1.1 Continuous equations

First, the energy associated to the differential system (18) is defined at a given time t :

$$\xi(t) = \frac{1}{2}\left(\|\mathbf{H}(t)\|^2_{L^2(\Omega)} + \varepsilon_\infty \|\mathbf{E}(t)\|^2_{L^2(\Omega)} + \frac{1}{\omega_d^2} \left\|\mathbf{J}_p(t)\right\|^2_{L^2(\Omega)} \right). \tag{22}$$

Assuming that $(\mathbf{E}, \mathbf{H})$ have a sufficient regularity (*i.e.* $(\mathbf{E}, \mathbf{H}) \in H_0(\mathbf{rot}, \Omega) \times H(\mathbf{rot}, \Omega)$ in space, and C^1 in time, for example), one would like to prove that this energy decreases with time. The L^2 scalar product on Ω of each equation leads to :

$$\frac{\partial \xi}{\partial t} = -\int_\Omega (\nabla \times \mathbf{E}) \cdot \mathbf{H} + \int_\Omega (\nabla \times \mathbf{H}) \cdot \mathbf{E} - \frac{\gamma_d}{\omega_d^2} \int_\Omega \mathbf{J}_p \cdot \mathbf{J}_p.$$

The Green formula applied to the first term implies :

$$\int_\Omega (\nabla \times \mathbf{E}) \cdot \mathbf{H} = \int_\Omega (\nabla \times \mathbf{H}) \cdot \mathbf{E} - \int_{\partial\Omega} (\mathbf{H} \times \mathbf{E}) \cdot \mathbf{n}.$$

Using the property of the mixed product, and given the metallic boundary conditions, the boundary integral is equal to zero. Therefore :

$$\frac{\partial \xi}{\partial t} = -\frac{\gamma_d}{\omega_d^2} \left\|\mathbf{J}_p(t)\right\|^2_{L^2(\Omega)} \leq 0 \tag{23}$$

4.1.2 Semi-discrete case

We now seek the approximated fields in the $\mathbf{V}_h$ space that has been previously defined in equation (9). Moreover, $\forall \mathbf{W}_h \in \mathbf{V}_h$, the following notations are used :

- The restriction $\mathbf{W}_i = \mathbf{W}_h|_{T_i}$,
- The mean of $\mathbf{W}_h$ across an interface between two cells T_i and T_k :

$$\{\mathbf{W}_h\}_{ik} = \frac{\mathbf{W}_i|_{a_{ik}} + \mathbf{W}_k|_{a_{ik}}}{2},$$

- The tangential jump of $\mathbf{W}_h$ through an interface :

$$[\![\mathbf{W}_h]\!]_{ik} = \left(\mathbf{W}_k|_{a_{ik}} - \mathbf{W}_i|_{a_{ik}}\right) \times \mathbf{n}_{ik}.$$

We focus on the differential system (18). Taking the L^2 scalar product on T_i of each equation with test functions φ, ψ, ϕ sufficiently regular leads to :

$$\begin{cases} \displaystyle\int_{T_i} \frac{\partial \mathbf{H}}{\partial t} \cdot \varphi &= -\displaystyle\int_{T_i} \mathbf{E} \cdot (\nabla \times \varphi) + \int_{\partial T_i} \varphi \cdot (\mathbf{E} \times \mathbf{n}) \,, \\ \varepsilon_\infty \displaystyle\int_{T_i} \frac{\partial \mathbf{E}}{\partial t} \cdot \psi &= \displaystyle\int_{T_i} \mathbf{H} \cdot (\nabla \times \psi) - \int_{\partial T_i} (\psi \times \mathbf{H}) \cdot \mathbf{n} - \int_{T_i} \mathbf{J}_p \cdot \psi, \\ \dfrac{1}{\omega_d^2} \displaystyle\int_{T_i} \frac{\partial \mathbf{J}_p}{\partial t} \cdot \phi &= \displaystyle\int_{T_i} \mathbf{E} \cdot \phi - \frac{\gamma_d}{\omega_d^2} \int_{T_i} \mathbf{J}_p \cdot \phi. \end{cases}$$

One defines the semi-discrete fields $(\mathbf{H}_h, \mathbf{E}_h, \mathbf{J}_h)$ as solutions of the following weak formulation : $\forall\, (\varphi_h, \psi_h, \phi_h) \in \mathbf{V}_h^3, \forall t \in [0,T], \forall i \in [0, N_T]$,

$$\begin{cases} \displaystyle\int_{T_i} \frac{\partial \mathbf{H}_h}{\partial t} \cdot \varphi_h &= -\displaystyle\int_{T_i} \mathbf{E}_h \cdot (\nabla \times \varphi_h) + \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \varphi_h \cdot (\{\mathbf{E}_h\}_{ik} \times \mathbf{n}) \,, \\ \varepsilon_\infty \displaystyle\int_{T_i} \frac{\partial \mathbf{E}_h}{\partial t} \cdot \psi_h &= \displaystyle\int_{T_i} \mathbf{H}_h \cdot (\nabla \times \psi_h) - \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} (\psi_h \times \{\mathbf{H}_h\}_{ik}) \cdot \mathbf{n} - \int_{T_i} \mathbf{J}_h \cdot \psi_h, \\ \dfrac{1}{\omega_d^2} \displaystyle\int_{T_i} \frac{\partial \mathbf{J}_h}{\partial t} \cdot \phi_h &= \displaystyle\int_{T_i} \mathbf{E}_h \cdot \phi_h - \frac{\gamma_d}{\omega_d^2} \int_{T_i} \mathbf{J}_h \cdot \phi_h. \end{cases} \tag{24}$$

and the associated semi-discrete energy ξ_h is therefore defined, according to the definition of ξ, as follows :

$$\xi_h = \frac{1}{2} \left(\|\mathbf{H}_h\|^2_{L^2(\Omega)} + \varepsilon_\infty \|\mathbf{E}_h\|^2_{L^2(\Omega)} + \frac{1}{\omega_d^2} \|\mathbf{J}_h\|^2_{L^2(\Omega)} \right).$$

Taking $(\varphi_h, \psi_h, \phi_h) = (\mathbf{H}_h, \mathbf{E}_h, \mathbf{J}_h)$, and then summing the three equations of (24) together and over all cells, one gets :

$$\begin{aligned} \frac{\partial \xi_h}{\partial t} &= \int_\Omega (\mathbf{H}_h \cdot (\nabla \times \mathbf{E}_h) - \mathbf{E}_h \cdot (\nabla \times \mathbf{H}_h)) - \frac{\gamma_d}{\omega_d^2} \int_\Omega \mathbf{J}_h \cdot \mathbf{J}_h \\ &+ \int_{\mathcal{F}_{int}} \{\mathbf{H}_h\} [\![\mathbf{E}_h]\!] - \int_{\mathcal{F}_{int}} \{\mathbf{E}_h\} [\![\mathbf{H}_h]\!] + \int_{\partial\Omega} \mathbf{E}_h \cdot (\mathbf{H}_h \times \mathbf{n}) \,. \end{aligned}$$

A final integration by parts[6] leads to :

$$\frac{\partial \xi_h}{\partial t} = -\frac{\gamma_d}{\omega_d^2} \|\mathbf{J}_h\|^2_{L^2(\Omega)} \,. \tag{25}$$

Hence, $\xi_h(t) \leq \xi_h(0)$, and the stability of the semi-discrete scheme is ensured. For the rest of this study, we need to write the semi-discrete formulation over the whole domain. Therefore, let $\mathbf{U} = (X, Y, Z)$ and $\mathbf{U}' = (X', Y', Z')$, and let the following bilinear forms :

$$\begin{aligned} m\left(\mathbf{U}, \mathbf{U}'\right) &= \int_\Omega X \cdot X' + \varepsilon_\infty \int_\Omega Y \cdot Y' + \frac{1}{\omega_d^2} \int_\Omega Z \cdot Z' \\ a\left(\mathbf{U}, \mathbf{U}'\right) &= \int_\Omega (X \cdot (\nabla \times Y') - Y \cdot (\nabla \times X')) - \int_\Omega Z \cdot Y' + \int_\Omega Y \cdot Z' - \frac{\gamma_d}{\omega_d^2} \int_\Omega Z \cdot Z' \\ b\left(\mathbf{U}, \mathbf{U}'\right) &= \int_{\mathcal{F}_{\text{int}}} \{X\} [\![Y']\!] - \int_{\mathcal{F}_{\text{int}}} \{Y\} [\![X']\!] + \int_{\partial\Omega} Y' \cdot (X \times \mathbf{n}) \,. \end{aligned}$$

[6] Realized on each cell T_i and then summed over the whole domain.

Let $\mathbf{U}_h \in \mathbf{V}_h$ be the solution of the semi-discrete problem. Given thee latter definitions, the semi-discrete formulation can be rewritten on the entire domain as :

$$m\left(\frac{\partial \mathbf{U}_h}{\partial t}, \mathbf{U}'_h\right) = a\left(\mathbf{U}_h, \mathbf{U}'_h\right) + b\left(\mathbf{U}_h, \mathbf{U}'_h\right), \forall \mathbf{U}_h \in \mathbf{V}_h^6. \tag{26}$$

Moreover, the solution $\mathbf{U}$ of the continuous equations verifies :

$$m\left(\frac{\partial \mathbf{U}}{\partial t}, \mathbf{U}'_h\right) = a\left(\mathbf{U}, \mathbf{U}'_h\right) + b\left(\mathbf{U}, \mathbf{U}'_h\right), \forall \mathbf{U}_h \in \mathbf{V}_h^6. \tag{27}$$

4.1.3 Fully discrete scheme

A second-order leap-frog (LF2) time discretization is chosen. The electric field on one hand, and the magnetic field and polarization current on the other hand are evaluated respectively at the even and odd time stations. The fully discrete scheme can then be written as follows :

$$\begin{cases} \displaystyle\int_{T_i} \frac{\mathbf{H}_i^{n+\frac{3}{2}} - \mathbf{H}_i^{n+\frac{1}{2}}}{\Delta t} \cdot \varphi_h &= -\displaystyle\int_{T_i} \mathbf{E}_i^{n+1} \cdot (\nabla \times \varphi_h) + \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \varphi_h \cdot \left(\left\{\mathbf{E}_h^{n+1}\right\}_{ik} \times \mathbf{n}_{ik}\right), \\ \varepsilon_\infty \displaystyle\int_{T_i} \frac{\mathbf{E}_i^{n+1} - \mathbf{E}_i^{n}}{\Delta t} \cdot \psi_h &= \displaystyle\int_{T_i} \mathbf{H}_i^{n+\frac{1}{2}} \cdot (\nabla \times \psi_h) - \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \left(\psi_h \times \left\{\mathbf{H}_i^{n+\frac{1}{2}}\right\}_{ik}\right) \cdot \mathbf{n}_{ik} \\ & \quad - \displaystyle\int_{T_i} \mathbf{J}_i^{n+\frac{1}{2}} \cdot \psi_h, \\ \dfrac{1}{\omega_d^2} \displaystyle\int_{T_i} \frac{\mathbf{J}_i^{n+\frac{3}{2}} - \mathbf{J}_i^{n+\frac{1}{2}}}{\Delta t} \cdot \phi_h &= \displaystyle\int_{T_i} \mathbf{E}_i^{n+1} \cdot \phi_h - \frac{\gamma_d}{\omega_d^2} \int_{T_i} \frac{\mathbf{J}_i^{n+\frac{3}{2}} + \mathbf{J}_i^{n+\frac{1}{2}}}{2} \cdot \phi_h, \end{cases} \tag{28}$$

and we define the associated energy on the cell T_i by :

$$\xi_i^n = \frac{1}{2}\left(\int_{T_i} \mathbf{H}_i^{n+\frac{1}{2}} \cdot \mathbf{H}_i^{n-\frac{1}{2}} + \varepsilon_\infty \int_{T_i} \mathbf{E}_i^n \cdot \mathbf{E}_i^n + \frac{1}{\omega_d^2} \int_{T_i} \mathbf{J}_i^{n+\frac{1}{2}} \cdot \mathbf{J}_i^{n-\frac{1}{2}}\right) \tag{29}$$

The total energy at a given time $t_n = n\Delta t$ is calculated as :

$$\xi^n = \sum_{i=0}^{N_T} \xi_i^n$$

Since they will be useful later, we remind to the reader the following classical inequalities [BS08].

Lemma 1. *let T_i be a cell of the mesh $\mathcal{T}_h$. Then, $\exists C > 0$, such that $\forall\, u \in \mathbb{P}_p(T_i)$:*

$$\|\nabla \times u\|_{T_i} \leq \frac{C}{h} \|u\|_{T_i}, \tag{30}$$

and

$$\|u\|_{a_{ik}}^2 \leq \frac{C}{h} \|u\|_{T_i}^2, \tag{31}$$

where $\|\cdot\|_{T_i}$ *and* $\|\cdot\|_{a_{ik}}$ *represent the* L^2 *norms defined respectively on the cell* T_i *and on the face of the cell* a_{ik}*.*

Proposition 1 (Stability)**.** *The formulation (28) is stable under the following condition :*

$$\Delta t < \min\left(\frac{h}{C}, \frac{2}{\omega_d + \gamma_d}, \frac{4\varepsilon_\infty}{\frac{C}{h} - \omega_d}\right) \tag{32}$$

Proof. One would like to know under which conditions the energy ξ^n can be written as a definite positive form of the variables $\mathbf{H}^{n-\frac{1}{2}}$, $\mathbf{E}^n$ and $\mathbf{J}^{n-\frac{1}{2}}$. To do so, we seek a lower bound of ξ^n by using the equations of the system (28) to replace the occurences of $\mathbf{H}^{n+\frac{1}{2}}$ and $\mathbf{H}^{n-\frac{1}{2}}$ with the help of well-chosen test functions. More accurately, the following substitutions are used :

- The first equation of (28) is used at time t_n with $\varphi_h = \mathbf{H}_i^{n-\frac{1}{2}}$;
- The second one at t_n with $\psi_h = \mathbf{E}_i^n$;
- The third one at t_n with $\phi_h = \mathbf{J}_i^{n-\frac{1}{2}}$.

These lead to :

$$\begin{aligned}
\xi_i^n &= \frac{1}{2}\left(\left\|\mathbf{H}_i^{n-\frac{1}{2}}\right\|_{T_i}^2 + \varepsilon_\infty \left\|\mathbf{E}_i^n\right\|_{T_i}^2 + \frac{\alpha}{\beta\omega_d^2}\left\|\mathbf{J}_i^{n-\frac{1}{2}}\right\|_{T_i}^2 - \Delta t \int_{T_i}\left(\nabla\times\mathbf{H}_i^{n-\frac{1}{2}}\right)\cdot\mathbf{E}_i^n\right. \\
&+ \left.\frac{\Delta t}{\beta}\int_{T_i}\mathbf{E}_i^n\cdot\mathbf{J}_i^{n-\frac{1}{2}} + \Delta t\sum_{k\in\mathcal{V}_i}\int_{a_{ik}}\mathbf{H}_i^{n-\frac{1}{2}}\cdot\left(\{\mathbf{E}_h^n\}_{ik}\times\mathbf{n}_{ik}\right)\right),
\end{aligned}$$

where $\alpha = 1 - \frac{\Delta t\gamma_d}{2}$ and $\beta = 1 + \frac{\Delta t\gamma_d}{2}$. Splitting the curl term in two parts, and integrating by parts one out of the two, one gets (the power indices are omitted up to the end of this proof) :

$$\begin{aligned}
\xi_i^n &= \frac{1}{2}\left(\|\mathbf{H}_i\|_{T_i}^2 + \varepsilon_\infty\|\mathbf{E}_i\|_{T_i}^2 + \frac{\alpha}{\beta\omega_d^2}\|\mathbf{J}_i\|_{T_i}^2\right. \\
&- \frac{\Delta t}{2}\int_{T_i}\left(\mathbf{E}_i\cdot(\nabla\times\mathbf{H}_i) + \mathbf{H}_i\cdot(\nabla\times\mathbf{E}_i)\right) + \frac{\Delta t}{\beta}\int_{T_i}\mathbf{E}_i\cdot\mathbf{J}_i \\
&+ \left.\Delta t\sum_{k\in\mathcal{V}_i}\int_{a_{ik}}\mathbf{H}_i\cdot\left(\{\mathbf{E}_h\}_{ik}\times\mathbf{n}_{ik}\right) - \frac{\Delta t}{2}\sum_{k\in\mathcal{V}_i}\int_{a_{ik}}\left(\mathbf{H}_i\times\mathbf{E}_i\right)\cdot\mathbf{n}_{ik}\right).
\end{aligned}$$

The surface integrals then rearrange as follows, given the metallic boundary conditions at the boundaries of the domain :

$$\begin{aligned}
&\Delta t\sum_{k\in\mathcal{V}_i}\int_{a_{ik}}\mathbf{H}_i\cdot\left(\{\mathbf{E}_h\}_{ik}\times\mathbf{n}_{ik}\right) - \frac{\Delta t}{2}\sum_{k\in\mathcal{V}_i}\int_{a_{ik}}\left(\mathbf{H}_i\times\mathbf{E}_i\right)\cdot\mathbf{n}_{ik} \\
&= \frac{\Delta t}{2}\sum_{k\in\mathcal{F}_{int}\bigcap\mathcal{V}_i}\int_{a_{ik}}\mathbf{H}_i\cdot\left(\mathbf{E}_k\times\mathbf{n}_{ik}\right) - \frac{\Delta t}{2}\sum_{k\in\partial\Omega\bigcap\mathcal{V}_i}\int_{a_{ik}}\mathbf{H}_i\cdot\left(\mathbf{E}_i\times\mathbf{n}_{ik}\right).
\end{aligned}$$

The different integrals are then bounded as follows using (30) and (31) :

$$\begin{aligned}\left|\int_{T_i}\left(\mathbf{E}_i\cdot(\nabla\times\mathbf{H}_i)+\mathbf{H}_i\cdot(\nabla\times\mathbf{E}_i)\right)\right| &\leq \frac{C}{h}\|\mathbf{E}_i\|_{T_i}\|\mathbf{H}_i\|_{T_i},\\ \left|\int_{a_{ik}}\mathbf{H}_i\cdot(\mathbf{E}_i\times\mathbf{n}_{ik})\right| &\leq \frac{C}{h}\|\mathbf{E}_i\|_{T_i}\|\mathbf{H}_i\|_{T_i},\\ \left|\int_{T_i}\mathbf{E}_i\cdot\mathbf{J}_i\right| &\leq \|\mathbf{E}_i\|_{T_i}\|\mathbf{J}_i\|_{T_i}.\end{aligned}$$

Using the previous relations as well as the classical inequality $ab\leq\frac{1}{2}\left(a^2+b^2\right)$, one gets :

$$\begin{aligned}\xi_i^n &\geq \frac{1}{2}\left(\|\mathbf{H}_i\|_{T_i}^2+\varepsilon_\infty\|\mathbf{E}_i\|_{T_i}^2+\frac{\alpha}{\beta\omega_d^2}\|\mathbf{J}_i\|_{T_i}^2\right)\\ &- \frac{\Delta t}{8}\frac{C}{h}\left(\|\mathbf{E}_i\|_{T_i}^2+\|\mathbf{H}_i\|_{T_i}^2\right)-\frac{\Delta t}{4}\frac{\omega_d}{\beta}\left(\|\mathbf{E}_i\|_{T_i}^2+\frac{1}{\omega_d^2}\|\mathbf{J}_i\|_{T_i}^2\right)\\ &- \frac{\Delta t}{8}\frac{C}{h}\left(\sum_{a_{ik}\in\mathcal{F}_{int}}\left(\|\mathbf{H}_i\|_{T_i}^2+\|\mathbf{E}_k\|_{T_k}^2\right)+\sum_{a_{ik}\in\partial\Omega}\left(\|\mathbf{H}_i\|_{T_i}^2+\|\mathbf{E}_i\|_{T_i}^2\right)\right).\end{aligned}\tag{33}$$

In order to make the CFL conditions more readable, it is assumed that :

$$\Delta t\leq\frac{2}{\gamma_d},\tag{34}$$

which implies :

$$1\geqslant\frac{1}{\beta}\geqslant\frac{1}{2}.$$

Therefore, adjusting C, (33) can be rewritten as :

$$\begin{aligned}\xi_i^n &\geq \frac{1}{2}\left(1-\frac{C\Delta t}{4h}\right)\|\mathbf{H}_i\|_{T_i}^2+\frac{1}{2}\left(\varepsilon_\infty-\frac{C\Delta t}{4h}-\frac{\Delta t\omega_d}{4}\right)\|\mathbf{E}_i\|_{T_i}^2\\ &- \frac{C\Delta t}{8h}\sum_{k\in\mathcal{V}_i}\|\mathbf{E}_k\|_{T_k}^2+\frac{1}{4\omega_d^2}\left(1-\frac{\gamma_d\Delta t}{2}-\frac{\omega_d\Delta t}{2}\right)\|\mathbf{J}_i\|_{T_i}^2.\end{aligned}$$

Then, summing over all the cells and adjusting C, one obtains an inequality involving ξ^n :

$$\xi^n\geq\frac{1}{2}\left(1-\frac{C\Delta t}{h}\right)\|\mathbf{H}\|_\Omega^2+\frac{1}{2}\left(\varepsilon_\infty-\frac{\Delta t}{4}\left(\frac{C}{h}-\omega_d\right)\right)\|\mathbf{E}\|_\Omega^2+\frac{1}{4\omega_d^2}\left(1-\frac{\Delta t}{2}\left(\gamma_d+\omega_d\right)\right)\|\mathbf{J}\|_\Omega^2.$$

Each one of the three induced conditions is now considered separately :

$$\Delta t<\frac{h}{C},\ \Delta t<\frac{4\varepsilon_\infty}{\frac{C}{h}-\omega_d},\ \Delta t<\frac{2}{\omega_d+\gamma_d}.$$

One should eventually notice that the condition (34) is contained in the last condition above-written, which leads to the desired result.

□

Proposition 2 (Bound of the discrete energy)**.** *Under CFL condition, the discrete energy ξ^n can be bounded in the following way :*

$$\xi^n \leq \frac{\xi^0}{\left(\frac{1-\theta}{1+\theta}\right)^n} \ \forall n \in \mathbb{N}^*$$

with $\theta \geq 0$.

Proof. In order to bound the discrete energy, one would first be interested in knowing the sign of $\xi^{n+1} - \xi^n \,, \forall\, n \geq 0$. To do so, the equations of (28) are used at different times and for different test functions :

- The first one at time t_n and time t_{n+1} with $\varphi_h = \mathbf{H}_i^{n+\frac{1}{2}}$;
- The second one at time t_n with $\psi_h = \mathbf{E}_i^{\left[n+\frac{1}{2}\right]}$;
- The last one at time t_n and time t_{n+1} with $\phi_h = \mathbf{J}_i^{n+\frac{1}{2}}$.

Combining these different expressions yields :

$$\begin{aligned}
\xi_i^{n+1} - \xi_i^n &= \Delta t \int_{T_i} \left(\mathbf{H}_i^{n+\frac{1}{2}} \cdot \left(\nabla \times \mathbf{E}_i^{\left[n+\frac{1}{2}\right]} \right) - \mathbf{E}_i^{\left[n+\frac{1}{2}\right]} \cdot \left(\nabla \times \mathbf{H}_i^{n+\frac{1}{2}} \right) \right) \\
&+ \Delta t \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \left(\mathbf{H}_i^{n+\frac{1}{2}} \cdot \left(\mathbf{E}_i^{\left[n+\frac{1}{2}\right]} \times \mathbf{n}_{ik} \right) - \mathbf{E}_i^{\left[n+\frac{1}{2}\right]} \cdot \left(\mathbf{H}_i^{n+\frac{1}{2}} \times \mathbf{n}_{ik} \right) \right) \\
&+ \Delta t \int_{T_i} \mathbf{J}_i^{n+\frac{1}{2}} \cdot \left(\frac{1}{2\beta} \mathbf{E}_i^{n+1} + \frac{1}{2\alpha} \mathbf{E}_i^n - \mathbf{E}_i^{\left[n+\frac{1}{2}\right]} \right) - \frac{\alpha^2 - \beta^2}{2\alpha\beta\omega_d^2} \int_{T_i} \mathbf{J}_i^{n+\frac{1}{2}} \cdot \mathbf{J}_i^{n+\frac{1}{2}}.
\end{aligned}$$

Integrating by parts the curl terms twice yields :

$$\begin{aligned}
\xi_i^{n+1} - \xi_i^n &= -\frac{\Delta t}{2} \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \left(\mathbf{E}_i^{\left[n+\frac{1}{2}\right]} \times \mathbf{H}_k^{n+\frac{1}{2}} + \mathbf{E}_k^{\left[n+\frac{1}{2}\right]} \times \mathbf{H}_i^{n+\frac{1}{2}} \right) \cdot \mathbf{n}_{ik} \\
&+ \underbrace{\frac{\Delta t^2 \gamma_d}{4} \int_{T_i} \mathbf{J}_i^{n+\frac{1}{2}} \cdot \left(-\frac{\mathbf{E}_i^{n+1}}{1 + \frac{\Delta t \gamma_d}{2}} + \frac{\mathbf{E}_i^n}{1 - \frac{\Delta t \gamma_d}{2}} \right)}_{\eta} + \frac{1}{\omega_d^2} \underbrace{\frac{\Delta t \gamma_d}{\frac{\Delta t^2 \gamma_d^2}{4} - 1}}_{\theta_0} \left\| \mathbf{J}_i^{n+\frac{1}{2}} \right\|_{T_i}^2.
\end{aligned}$$

If $\Delta t \leq \frac{2}{\gamma_d}$, then $\theta_0 < 0$ and $|\theta| < |\theta_1| = \left| \frac{2}{\frac{\Delta t^2 \gamma_d^2}{4} - 1} \right|$. One can obtain the following bound for η :

$$\begin{aligned}
|\eta| &\leq \frac{\Delta t}{4} |\theta_1| \left(\beta \left\| \mathbf{J}_i^{n+\frac{1}{2}} \right\|_{T_i} \left\| \mathbf{E}_i^n \right\|_{T_i} + \alpha \left\| \mathbf{J}_i^{n+\frac{1}{2}} \right\|_{T_i} \left\| \mathbf{E}_i^{n+1} \right\|_{T_i} \right) \\
&\leq \frac{\Delta t \omega_d}{2} |\theta_1| \, \beta \frac{\left\| \mathbf{J}_i^{n+\frac{1}{2}} \right\|_{T_i}}{\omega_d} \underbrace{\max \left(\left\| \mathbf{E}_i^n \right\|_{T_i}, \left\| \mathbf{E}_i^{n+1} \right\|_{T_i} \right)}_{\rho_i} \\
&\leq \frac{\Delta t \omega_d}{2} |\theta_1| \left(\frac{\left\| \mathbf{J}_i^{n+\frac{1}{2}} \right\|_{T_i}^2}{\omega_d^2} + \rho_i^2 \right).
\end{aligned}$$

where $\alpha = 1 + \dfrac{\Delta t \gamma_d}{2}$ and $\alpha = 1 - \dfrac{\Delta t \gamma_d}{2}$. As specified earlier, if $\Delta t \leq \frac{2}{\gamma_d}$, then one has :

$$\beta \leq 2 \text{ and } \alpha \leq 1.$$

Then, by summing the contributions over all the mesh cells, and by considering metallic boundaries, it is possible to write :

$$\begin{aligned}
\xi^{n+1} - \xi^n &\leq \sum_{i=0}^{N_T} \left(\frac{\Delta t \omega_d \left|\theta_1\right|}{2} \left(\frac{\left\| \mathbf{J}_i^{n+\frac{1}{2}} \right\|_{T_i}^2}{\omega_d^2} + \rho_i^2 \right) + \underbrace{\frac{\theta_0}{\omega_d^2} \left\| \mathbf{J}_i^{n+\frac{1}{2}} \right\|_{T_i}^2}_{\leq 0} \right) \\
&\leq \sum_{i=0}^{N_T} \left|\theta_1\right| \frac{\Delta t \omega_d}{2} \left(\frac{\left\| \mathbf{J}_i^{n+\frac{1}{2}} \right\|_{T_i}^2}{\omega_d^2} + \rho_i^2 \right).
\end{aligned}$$

Let C be a generic constant. Under CFL condition, the first term involving $\sum_{i=1}^{N_T} \left\| \mathbf{J}_i^{n+\frac{1}{2}} \right\|_{T_i}^2$ is bounded by $C\xi^{n+1}$, whereas the $\sum_{i=1}^{N_T} \rho_i^2$ one is bounded by $C\left(\xi^n + \xi^{n+1}\right)$. This yields :

$$\xi^{n+1} \leq \left|\theta_1\right| \Delta t \omega_d \xi^{n+1} + \left(1 + \frac{\left|\theta_1\right| \Delta t \omega_d}{2} \right) \xi^n.$$

The latter leads to :

$$\xi^{n+1} \leq \frac{1 + \frac{\left|\theta_1\right| \Delta t \omega_d}{2}}{1 - \left|\theta_1\right| \Delta t \omega_d} \xi^n \leq \frac{1 + \left|\theta_1\right| \Delta t \omega_d}{1 - \left|\theta_1\right| \Delta t \omega_d} \xi^n = \frac{1+\theta}{1-\theta} \xi^n,$$

with $\theta \geq 0$. It is also required that $\theta \leq 1$, which can be shown to be equivalent to :

$$\Delta t \leq \frac{\gamma_d^2}{2} \left(-2\omega_d + \sqrt{4\omega_d^2 + \gamma_d^2} \right).$$

The latter condition can be proved to be less restrictive than the $\Delta t \leq \frac{2}{\gamma_d}$. Therefore, we have obtained the expected result.

□

4.2 Convergence of the fully discrete Maxwell-Drude DG formulation

In this section, the convergence of the Maxwell-Drude DG scheme is proven, after a preliminary convergence result about the semi-discrete formulation. We start with the following lemma.

Lemma 2. *Let T_i be a cell of the mesh $\mathcal{T}_h$, and $\pi_h(\cdot)$ a linear continuous projector from $H^{s+1}(T_i)$ onto $\mathbb{P}_p(T_i)$ with $s \geq 0$ and $k \geq 1$. Then, for $u \in H^{s+1}(T_i)$ and $m = 0, 1$:*

$$\left| u - \pi_h(u) \right|_{m,T_i} \leq C h_{T_i}^{\min(s,p)+1-m} \left\| u \right\|_{s+1,T_i}, \tag{35}$$

and

$$\left\|u-\pi_h\left(u\right)\right\|_{0,\partial T_i} \leq C h_{T_i}^{\min(s,p)+\frac{1}{2}} \left\|u\right\|_{s+1,T_i}. \tag{36}$$

4.2.1 Convergence of the semi-discrete formulation

Theorem 1 (Convergence of the semi-discrete formulation)**.** *Let $\left(\boldsymbol{H},\boldsymbol{E},\boldsymbol{J}_p\right)$ be the solution of (17) and $\left(\boldsymbol{H}_h,\boldsymbol{E}_h,\boldsymbol{J}_h\right) \in \mathcal{C}^1\left(\left[0,T\right],\mathbf{V}_h^3\right)$ the semi-discrete solution of (24). If $\left(\boldsymbol{H},\boldsymbol{E},\boldsymbol{J}_p\right) \in \mathcal{C}^0\left(\left[0,T\right],H^{s+1}\left(\Omega\right)^9\right)$ for $s \geq 0$, then there exists $C \geq 0$ independent of h such that :*

$$\max_{t\in[0,T]} \gamma(t)^{\frac{1}{2}} \leq C h^{\min(s,p)} \left\|\left(\boldsymbol{H},\boldsymbol{E},\boldsymbol{J}_p\right)\right\|_{\mathcal{C}^0\left(\left[0,T\right],H^{s+1}(\Omega)^9\right)},$$

with :

$$\gamma(t) = \left\|\pi_h\left(\boldsymbol{H}\right)-\boldsymbol{H}_h\right\|_{L^2(\Omega)}^2 + \left\|\pi_h\left(\boldsymbol{E}\right)-\boldsymbol{E}_h\right\|_{L^2(\Omega)}^2 + \left\|\pi_h\left(\boldsymbol{J}_p\right)-\boldsymbol{J}_h\right\|_{L^2(\Omega)}^2 \ \forall t \in [0,T].$$

Proof. The orthogonal L^2 projection on the $\mathbf{V}_h^3$ space is defined as $\pi_h\left(\mathbf{U}\right) = \left(\pi_h\left(\mathbf{H}\right),\pi_h\left(\mathbf{E}\right),\right.$ $\left.\pi_h\left(\mathbf{J}_p\right)\right)$, and the consistency error as $\varepsilon(t) = \frac{1}{2} m\left(\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h,\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right)$. Then, it is easy to bound ξ in the following way :

$$\varepsilon(t) \geq \frac{1}{2}\min\left(1,\varepsilon_\infty,\frac{1}{\omega_d^2}\right)\left(\left\|\pi_h\left(\mathbf{H}-\mathbf{H}_h\right)\right\|_{L^2(\Omega)}^2 + \left\|\pi_h\left(\mathbf{E}-\mathbf{E}_h\right)\right\|_{L^2(\Omega)}^2 + \left\|\pi_h\left(\mathbf{J}_p-\mathbf{J}_h\right)\right\|_{L^2(\Omega)}^2\right)$$

In addition, given that $\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h \in V_h$, one could easily prove that :

$$m\left(\pi_h\left(\frac{\partial \mathbf{U}}{\partial t}\right)-\frac{\partial \mathbf{U}}{\partial t},\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right) = 0, \tag{37}$$

$$a\left(\pi_h\left(\frac{\partial \mathbf{U}}{\partial t}\right)-\frac{\partial \mathbf{U}}{\partial t},\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right) = 0. \tag{38}$$

Since m, a and b are bilinear forms, and given the equalities (26 - 27 - 37 - 38), it is possible to write (the proof is exactly the same as in [LS12], pp 12 – 13) :

$$\begin{aligned} m\left(\frac{\partial \pi_h\left(\mathbf{U}\right)}{\partial t}-\frac{\partial \mathbf{U}_h}{\partial t},\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right) &= a\left(\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h,\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right) \\ &+ b\left(\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h,\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right) \\ &+ b\left(\mathbf{U}-\pi_h\left(\mathbf{U}\right),\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right). \end{aligned}$$

Given what has already been proven in the stability study of the semi-discrete formulation, it is straightforward to prove that we have :

$$a\left(\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h,\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right) + b\left(\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h,\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right) = -\frac{\gamma_d}{\omega_d^2}\left\|\pi_h\left(\mathbf{J}_p\right)-\mathbf{J}_h\right\|_\Omega^2.$$

Then, the term $b\left(\mathbf{U}-\pi_h\left(\mathbf{U}\right),\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right)$ remains to be bounded. Let us write :

$$
\begin{aligned}
b\left(\mathbf{U}-\pi_h\left(\mathbf{U}\right),\pi_h\left(\mathbf{U}\right)-\mathbf{U}_h\right) &= \underbrace{\int_{\mathcal{F}_{\text{int}}}\left\{\mathbf{H}-\pi_h\left(\mathbf{H}\right)\right\}\left[\!\left[\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right]\!\right]}_{\zeta} \\
&- \underbrace{\int_{\mathcal{F}_{\text{int}}}\left\{\mathbf{E}-\pi_h\left(\mathbf{E}\right)\right\}\left[\!\left[\pi_h\left(\mathbf{H}\right)-\mathbf{H}_h\right]\!\right]}_{\mu} \\
&+ \underbrace{\int_{\partial\Omega}\left(\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right)\cdot\left(\left(\mathbf{H}-\pi_h\left(\mathbf{H}\right)\right)\times\mathbf{n}\right)}_{\kappa}.
\end{aligned}
$$

First, one considers the ζ term. Applying the Cauchy-Schwarz inequality leads to :

$$
\begin{aligned}
|\zeta| &\leq \left|\sum_{a_{ik}\in\mathcal{F}_{\text{int}}}\int_{a_{ik}}\left\{\mathbf{H}-\pi_h\left(\mathbf{H}\right)\right\}_{ik}\left[\!\left[\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right]\!\right]_{ik}\right| \\
&\leq \sum_{a_{ik}\in\mathcal{F}_{\text{int}}}\left|\int_{a_{ik}}\left\{\mathbf{H}-\pi_h\left(\mathbf{H}\right)\right\}_{ik}\left[\!\left[\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right]\!\right]_{ik}\right| \\
&\leq \sum_{a_{ik}\in\mathcal{F}_{\text{int}}}\left(\int_{a_{ik}}\left|\left\{\mathbf{H}-\pi_h\left(\mathbf{H}\right)\right\}_{ik}\right|^2\right)^{\frac{1}{2}}\left(\int_{a_{ik}}\left|\left[\!\left[\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right]\!\right]_{ik}\right|^2\right)^{\frac{1}{2}}.
\end{aligned}
$$

The first sum can be bounded as follows :

$$
\begin{aligned}
\int_{a_{ik}}\left|\left\{\mathbf{H}-\pi_h\left(\mathbf{H}\right)\right\}_{ik}\right|^2 &\leq C\left(\left\|\mathbf{H}_i-\pi_h\left(\mathbf{H}_i\right)\right\|^2_{L^2(a_{ik})}+\left\|\mathbf{H}_k-\pi_h\left(\mathbf{H}_k\right)\right\|^2_{L^2(a_{ik})}\right) \\
&\leq C\left(\left\|\mathbf{H}_i-\pi_h\left(\mathbf{H}_i\right)\right\|^2_{L^2(\partial T_i)}+\left\|\mathbf{H}_k-\pi_h\left(\mathbf{H}_k\right)\right\|^2_{L^2(\partial T_k)}\right) \\
&\leq C\left(\left(h_{T_i}^{\min(s,p)+\frac{1}{2}}\left\|\mathbf{H}\right\|_{s+1,T_i}\right)^2+\left(h_{T_k}^{\min(s,p)+\frac{1}{2}}\left\|\mathbf{H}\right\|_{s+1,T_k}\right)^2\right) \\
&\leq C\delta^{2\min(s,p)+1}h_{T_i}^{2\min(s,p)+1}\left(\left\|\mathbf{H}\right\|^2_{s+1,T_i}+\left\|\mathbf{H}\right\|^2_{s+1,T_k}\right),
\end{aligned}
$$

the last inequality resulting from the quasi-uniform assumption made on the triangulation of the domain. Bounding the second integral can be done almost the same way, and requires the inverse inequalities presented earlier :

$$
\int_{a_{ik}}\left|\left[\!\left[\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right]\!\right]_{ik}\right|^2\leq C\delta^{-1}h_{T_i}^{-1}\left(\left\|\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right\|^2_{0,T_i}+\left\|\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right\|^2_{0,T_k}\right).
$$

Then, by adjusting the constant C, one obtains :

$$
\begin{aligned}
|\zeta| &\leq C\delta^{\min(s,p)}\sum_{(i,k)|a_{ik}\in\mathcal{F}_{\text{int}}}h_{T_i}^{\min(s,p)}\left(\left\|\mathbf{H}\right\|^2_{s+1,T_i}+\left\|\mathbf{H}\right\|^2_{s+1,T_k}\right)^{\frac{1}{2}}\left(\left\|\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right\|^2_{0,T_i}+\left\|\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right\|^2_{0,T_k}\right)^{\frac{1}{2}} \\
&\leq C\delta^{\min(s,p)}h^{\min(s,p)}\sum_{(i,k)|a_{ik}\in\mathcal{F}_{\text{int}}}\left(\left\|\mathbf{H}\right\|^2_{s+1,T_i}+\left\|\mathbf{H}\right\|^2_{s+1,T_k}\right)^{\frac{1}{2}}\left(\left\|\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right\|^2_{0,T_i}+\left\|\pi_h\left(\mathbf{E}\right)-\mathbf{E}_h\right\|^2_{0,T_k}\right)^{\frac{1}{2}},
\end{aligned}
$$

where $h = \max_{T_i} h_{T_i}$. Eventually, one must consider the sum with respect to all the internal faces, which can be rewritten as follows by adjusting once again the generic constant C :

$$|\zeta| \leq C\delta^{\min(s,p)} h^{\min(s,p)} \sum_{T_i} \|\mathbf{H}\|_{s+1,T_i} \|\pi_h(\mathbf{E}) - \mathbf{E}_h\|_{0,T_i}.$$

Bounding μ can be done the exact same way :

$$|\mu| \leq C\delta^{\min(s,p)} h^{\min(s,p)} \sum_{T_i} \|\mathbf{E}\|_{s+1,T_i} \|\pi_h(\mathbf{H}) - \mathbf{H}_h\|_{0,T_i}.$$

Moreover, one could easily prove that $\kappa = 0$, given the metallic boundary condition on $\partial\Omega$. Therefore :

$$\begin{aligned}
|\zeta| + |\mu| &\leq \sum_{T_i} C\delta^{\min(s,p)} h_{T_i}^{\min(s,p)} \left(\|\mathbf{H}\|_{s+1,T_i} \|\pi_h(\mathbf{E}) - \mathbf{E}_h\|_{0,T_i} + \|\mathbf{E}\|_{s+1,T_i} \|\pi_h(\mathbf{H}) - \mathbf{H}_h\|_{0,T_i} \right) \\
&\leq C \max_{T_i} \delta^{\min(s,p)} h_{T_i}^{\min(s,p)} \left[\left(\sum_{T_i} \|\mathbf{H}\|_{s+1,T_i}^2 \right)^{\frac{1}{2}} \left(\sum_{T_i} \|\pi_h(\mathbf{E}) - \mathbf{E}_h\|_{0,T_i}^2 \right)^{\frac{1}{2}} \right. \\
&+ \left. \left(\sum_{T_i} \|\mathbf{E}\|_{s+1,T_i}^2 \right)^{\frac{1}{2}} \left(\sum_{T_i} \|\pi_h(\mathbf{H}) - \mathbf{H}_h\|_{0,T_i}^2 \right)^{\frac{1}{2}} \right] \\
&\leq C\delta^{\min(s,p)} h^{\min(s,p)} \left(\|\mathbf{H}\|_{s+1,\Omega} \|\pi_h(\mathbf{E}) - \mathbf{E}_h\|_{0,\Omega} + \|\mathbf{E}\|_{s+1,\Omega} \|\pi_h(\mathbf{H}) - \mathbf{H}_h\|_{0,\Omega} \right) \\
&\leq C\delta^{\min(s,p)} h^{\min(s,p)} \|(\mathbf{H},\mathbf{E})\|_{s+1,\Omega} \left(\|\pi_h(\mathbf{E}) - \mathbf{E}_h\|_{0,\Omega}^2 + \|\pi_h(\mathbf{H}) - \mathbf{H}_h\|_{0,\Omega}^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

Since $\varepsilon(t) = \frac{1}{2} \int_0^t m\left(\frac{\partial(\pi_h(\mathbf{U}) - \mathbf{U}_h)}{\partial s}, \pi_h(\mathbf{U}) - \mathbf{U}_h \right) \mathrm{d}s$, and assuming that $\varepsilon(0) = 0$, one can write :

$$\begin{aligned}
\varepsilon(t) &= \frac{1}{2} \int_0^t -\frac{\gamma_d}{\omega_d^2} \|\mathbf{J}_h\|_\Omega^2 \,\mathrm{d}s + \frac{1}{2} \int_0^t b(\mathbf{U} - \pi_h(\mathbf{U}), \pi_h(\mathbf{U}) - \mathbf{U}_h) \,\mathrm{d}s \\
&\leq \frac{1}{2} \int_0^t b(\mathbf{U} - \pi_h(\mathbf{U}), \pi_h(\mathbf{U}) - \mathbf{U}_h) \,\mathrm{d}s.
\end{aligned}$$

In regard of the previous results, one obtains :

$$\begin{aligned}
\varepsilon(t) &\leq Ch^{\min(s,p)} \int_0^t \|(\mathbf{H},\mathbf{E})\|_{s+1,\Omega} \left(\|\pi_h(\mathbf{E}) - \mathbf{E}_h\|_{0,\Omega}^2 + \|\pi_h(\mathbf{H}) - \mathbf{H}_h\|_{0,\Omega}^2 \right)^{\frac{1}{2}} \mathrm{d}s \\
&\leq Ch^{\min(s,p)} \int_0^t \left\| \left(\mathbf{H},\mathbf{E},\mathbf{J}_p\right) \right\|_{s+1,\Omega} \gamma(s)^{\frac{1}{2}} \mathrm{d}s,
\end{aligned}$$

therefore :

$$\gamma(t) \leq Ch^{\min(s,p)} \upsilon \left\| \left(\mathbf{H},\mathbf{E},\mathbf{J}_p\right) \right\|_{\mathcal{C}^0\left([0,T],H^{s+1}(\Omega)^3\right)},$$

where $\upsilon = \max_{[0,T]} \gamma(t)^{\frac{1}{2}}$. Taking the maximum value of the left hand side over $[0,T]$ permits to conclude.

□

4.2.2 Convergence of the fully discrete formulation

We consider the fully discrete formulation presented previously in equation (28). We choose a constant time step Δt verifying the CFL condition (32) and such that $N\Delta t = T$, where N is the total number of timesteps. Two classical results about Taylor-Lagrange expansions will be required :

Lemma 3. *Let be* $U \in \mathcal{C}^3\left([t_n, t_{n+1}]\right)$*, then* $\exists\, (c_n, c_{n+1}) \in]t_n, t_{n+\frac{1}{2}}[\times]t_{n+\frac{1}{2}}, t_{n+1}[$ *and* $(d_n, d_{n+1}) \in]t_n, t_{n+\frac{1}{2}}[\times]t_{n+\frac{1}{2}}, t_{n+1}[$ *such that :*

$$U_h\left(t_{n+1}\right) - U_h\left(t_n\right) = \Delta t \frac{\partial U_h}{\partial t}\left(t_{n+\frac{1}{2}}\right) + \frac{\Delta t^3}{28}\left(\frac{\partial^3 U_h}{\partial t^3}\left(c_{n+1}\right) + \frac{\partial^3 U_h}{\partial t^3}\left(c_n\right)\right),$$

and

$$\frac{1}{2}\left(U_h\left(t_{n+1}\right) + U_h\left(t_n\right)\right) = U_h\left(t_{n+\frac{1}{2}}\right) + \frac{\Delta t^2}{16}\left(\frac{\partial^2 U_h}{\partial t^2}\left(d_{n+1}\right) + \frac{\partial^2 U_h}{\partial t^2}\left(d_n\right)\right).$$

Theorem 2 (Convergence of the fully discrete formulation)**.** *Let be :*

$$\left(\boldsymbol{H}, \boldsymbol{E}, \boldsymbol{J}_p\right) \in \mathcal{C}^3\left(\left[0,T\right], L^2\left(\Omega\right)^9\right) \bigcap \mathcal{C}^0\left(\left[0,T\right], H^{s+1}\left(\Omega\right)^9\right).$$

Under a CFL condition as in (32), the following error estimate holds :

$$\begin{aligned}
&\max_{n\in[0,N]}\left(\left\|\boldsymbol{H}\left(t_{n+\frac{1}{2}}\right) - \boldsymbol{H}_h^{n+\frac{1}{2}}\right\|^2_{L^2(\Omega)^3} + \left\|\boldsymbol{E}\left(t_n\right) - \boldsymbol{E}_h^n\right\|^2_{L^2(\Omega)^3}\right.\\
&\qquad\left. + \left\|\boldsymbol{J}_p\left(t_{n+\frac{1}{2}}\right) - \boldsymbol{J}_h^{n+\frac{1}{2}}\right\|^2_{L^2(\Omega)^3}\right)^{\frac{1}{2}}\\
&\leq C\left(\Delta t^2 + h^{\min(s,k)}\right)\left(\left\|\left(\boldsymbol{H},\boldsymbol{E},\boldsymbol{J}_p\right)\right\|_{\mathcal{C}^3\left([0,T],L^2(\Omega)^9\right)} + \left\|\left(\boldsymbol{H},\boldsymbol{E},\boldsymbol{J}_p\right)\right\|_{\mathcal{C}^0\left([0,T],H^{s+1}(\Omega)^9\right)}\right)
\end{aligned}$$

Proof. We define the consistency error as follows :

$$\varepsilon_h^{n+1} = \left(\left\|\mathbf{E}_h\left(t_{n+1}\right) - \widetilde{\mathbf{E}}_h^{n+1}\right\|^2_{L^2(\Omega)} + \left\|\mathbf{H}_h\left(t_{n+\frac{3}{2}}\right) - \widetilde{\mathbf{H}}_h^{n+\frac{3}{2}}\right\|^2_{L^2(\Omega)} + \left\|\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right) - \widetilde{\mathbf{J}}_h^{n+\frac{3}{2}}\right\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}},$$

where $\widetilde{\mathbf{E}}_h^{n+1}, \widetilde{\mathbf{H}}_h^{n+\frac{3}{2}}$ and $\widetilde{\mathbf{J}}_h^{n+\frac{3}{2}}$ are defined as :

$$\begin{cases}
\displaystyle\int_{T_i} \frac{\widetilde{\mathbf{H}}_h^{n+\frac{3}{2}} - \mathbf{H}_h\left(t_{n+\frac{1}{2}}\right)}{\Delta t}\cdot\varphi_h &= -\displaystyle\int_{T_i}\mathbf{E}_h\left(t_{n+1}\right)\cdot\left(\nabla\times\varphi_h\right) + \sum_{k\in\mathcal{V}_i}\int_{a_{ik}}\varphi_h\cdot\left(\left\{\mathbf{E}_h\left(t_{n+1}\right)\right\}_{ik}\times\mathbf{n}_{ik}\right),\\
\varepsilon_\infty\displaystyle\int_{T_i}\frac{\widetilde{\mathbf{E}}_h^{n+1} - \mathbf{E}_h\left(t_n\right)}{\Delta t}\cdot\psi_h &= \displaystyle\int_{T_i}\mathbf{H}_h\left(t_{n+\frac{1}{2}}\right)\cdot\left(\nabla\times\psi_h\right) - \int_{T_i}\mathbf{J}_h\left(t_{n+\frac{1}{2}}\right)\cdot\psi_h\\
&\quad - \displaystyle\sum_{k\in\mathcal{V}_i}\int_{a_{ik}}\left(\psi_h\times\left\{\mathbf{H}_h\left(t_{n+\frac{1}{2}}\right)\right\}_{ik}\right)\cdot\mathbf{n}_{ik},\\
\displaystyle\frac{1}{\omega_d^2}\int_{T_i}\frac{\widetilde{\mathbf{J}}_h^{n+\frac{3}{2}} - \mathbf{J}_h\left(t_{n+\frac{1}{2}}\right)}{\Delta t}\cdot\phi_h &= \displaystyle\int_{T_i}\mathbf{E}_h\left(t_{n+1}\right)\cdot\phi_h - \frac{\gamma_d}{\omega_d^2}\int_{T_i}\frac{\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right) + \mathbf{J}_h\left(t_{n+\frac{1}{2}}\right)}{2}\cdot\phi_h.
\end{cases}$$

The following developments mainly focus on the $\mathbf{J}$ equation, since the treatment of the equation for the update of $\mathbf{E}$ and $\mathbf{H}$ is very similar to the one done in [LS12]. One could rewrite the last equation as follows :

$$\begin{aligned}&\frac{1}{\omega_d^2\Delta t}\int_{T_i}\left(\widetilde{\mathbf{J}}_h^{n+\frac{3}{2}}-\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right)+\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right)-\mathbf{J}_h\left(t_{n+\frac{1}{2}}\right)\right)\cdot\phi_h\\&=\int_{T_i}\mathbf{E}_h\left(t_{n+1}\right)\cdot\phi_h-\frac{\gamma_d}{\omega_d^2}\int_{T_i}\frac{\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right)+\mathbf{J}_h\left(t_{n+\frac{1}{2}}\right)}{2}\cdot\phi_h.\end{aligned}$$

The semi-discrete equation involving $\mathbf{J}_h$ used at time $t=t_{n+1}$ reads as :

$$\frac{1}{\omega_d^2}\int_{T_i}\frac{\partial\mathbf{J}_h}{\partial t}\left(t_{n+1}\right)\cdot\phi_h=\int_{T_i}\mathbf{E}_h\left(t_{n+1}\right)\cdot\phi_h-\frac{\gamma_d}{\omega_d^2}\int_{T_i}\mathbf{J}_h\left(t_{n+1}\right)\cdot\phi_h.$$

Then, substituting the $\mathbf{E}_h$ term in the fully discrete equation leads to :

$$\begin{aligned}&\frac{1}{\omega_d^2\Delta t}\int_{T_i}\left(\widetilde{\mathbf{J}}_h^{n+\frac{3}{2}}-\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right)+\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right)-\mathbf{J}_h\left(t_{n+\frac{1}{2}}\right)\right)\cdot\phi_h\\&=\frac{1}{\omega_d^2}\int_{T_i}\frac{\partial\mathbf{J}_h}{\partial t}\left(t_{n+1}\right)\cdot\phi_h+\frac{\gamma_d}{\omega_d^2}\int_{T_i}\mathbf{J}_h\left(t_{n+1}\right)\cdot\phi_h-\frac{\gamma_d}{\omega_d^2}\int_{T_i}\frac{\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right)+\mathbf{J}_h\left(t_{n+\frac{1}{2}}\right)}{2}\cdot\phi_h.\end{aligned}$$

Exploiting the Taylor-Lagrange equalities and notations from lemma 3 gives :

$$\begin{aligned}&\frac{1}{\Delta t}\int_{T_i}\left(\widetilde{\mathbf{J}}_i^{n+\frac{3}{2}}-\mathbf{J}_i\left(t_{n+\frac{3}{2}}\right)\right)\cdot\phi_h+\frac{\Delta t^2}{28}\int_{T_i}\left(\frac{\partial^3\mathbf{J}_h}{\partial t^3}\left(c_{n+1}\right)+\frac{\partial^3\mathbf{J}_h}{\partial t^3}\left(c_n\right)\right)\\&+\frac{\Delta t^2}{16}\int_{T_i}\left(\frac{\partial^2\mathbf{J}_h}{\partial t^2}\left(d_{n+1}\right)+\frac{\partial^2\mathbf{J}_h}{\partial t^2}\left(d_n\right)\right)=0,\end{aligned}$$

therefore :

$$\left\|\widetilde{\mathbf{J}}_h^{n+\frac{3}{2}}-\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right)\right\|_{L^2(\Omega)}\leq C\Delta t^3\left(\left\|\frac{\partial^2\mathbf{J}_h}{\partial t^2}(t)\right\|_{L^2(\Omega)}+\left\|\frac{\partial^3\mathbf{J}_h}{\partial t^3}(t)\right\|_{L^2(\Omega)}\right).$$

Following the same idea for the other two fields allows to write :

$$\left|\varepsilon_h^n\right|\leq C\Delta t^3\left\|\mathbf{U}\right\|_{\mathcal{C}^3\left([0,T],L^2(\Omega)^9\right)}. \tag{39}$$

One could now define $\hat{j}_i^n\left(\phi_h\right)=\dfrac{1}{\omega_d^2\Delta t}\displaystyle\int_{T_i}\left(\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right)-\widetilde{\mathbf{J}}_h^{n+\frac{3}{2}}\right)\cdot\phi_h$ as well as $\hat{e}_i^n$ and $\hat{h}_i^n$ associated to their related fields. One has :

$$\begin{aligned}&\frac{1}{\omega_d^2\Delta t}\int_{T_i}\left(\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right)-\mathbf{J}_h\left(t_{n+\frac{1}{2}}\right)\right)\cdot\phi_h+\frac{\gamma_d}{\omega_d^2}\int_{T_i}\frac{\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right)+\mathbf{J}_h\left(t_{n+\frac{1}{2}}\right)}{2}\cdot\phi_h\\&\qquad=\int_{T_i}\mathbf{E}_h\left(t_{n+1}\right)\cdot\phi_h+\hat{j}_i^n(\phi_h).\end{aligned}$$

We define $\hat{j}_h^n(\phi_h) = \sum_{i=0}^{N_T} \hat{j}_i^n(\phi_h) = \frac{1}{\omega_d^2 \Delta t} \int_\Omega \left(\mathbf{J}_h\left(t_{n+\frac{3}{2}}\right) - \widetilde{\mathbf{J}}_h^{n+\frac{3}{2}} \right) \cdot \phi_h$, which can be easily bounded as follows :

$$\left| \hat{j}_h^n(\phi_h) \right| \leq \frac{C}{\Delta t} \left\| \widetilde{\mathbf{J}}_h^{n+\frac{3}{2}} - \mathbf{J}_h\left(t_{n+\frac{3}{2}}\right) \right\|_{L^2(\Omega)} \|\phi_h\|_{L^2(\Omega)}.$$

Following the same development for $\hat{e}_h^n$ and $\hat{h}_h^n$, one eventually proves that, $|||\cdot|||$ being the operator norm associated to the considered linear forms on $L^2(\Omega)^3$:

$$||| \hat{j}_h^n ||| + ||| \hat{h}_h^n ||| + ||| \hat{e}_h^n ||| \leq C\Delta t^2 \|\mathbf{U}\|_{\mathcal{C}^3([0,T],L^2(\Omega)^9)}. \tag{40}$$

One now focuses on the error energy. One defines $\mathbb{J}_h^{n+\frac{1}{2}} = \mathbf{J}_h\left(t_{n+\frac{1}{2}}\right) - \mathbf{J}_h^{n+\frac{1}{2}}$, as well as their **E** and **H** counterparts $\mathbb{H}_h^{n+\frac{1}{2}}$ and $\mathbb{E}_h^n$. The error energy then writes as :

$$\hat{\varepsilon}_i^n = \frac{1}{2}\left(\varepsilon_\infty \|\mathbb{E}_i^n\|_{L^2(\Omega)}^2 + \left\langle \mathbb{H}_i^{n-\frac{1}{2}}, \mathbb{H}_i^{n+\frac{1}{2}} \right\rangle_{T_i} + \frac{1}{\omega_d^2} \left\langle \mathbb{J}_i^{n-\frac{1}{2}}, \mathbb{J}_i^{n+\frac{1}{2}} \right\rangle_{T_i} \right),$$

where $\mathbb{J}_i^{n+\frac{1}{2}}$ verifies :

$$\frac{1}{\omega_d^2 \Delta t} \int_{T_i} \left(\mathbb{J}_i^{n+\frac{3}{2}} - \mathbb{J}_i^{n+\frac{1}{2}} \right) \cdot \phi_h = \int_{T_i} \mathbb{E}_i^{n+1} \cdot \phi_h - \frac{\gamma_d}{\omega_d^2} \int_{T_i} \frac{\mathbb{J}_i^{n+\frac{3}{2}} + \mathbb{J}_i^{n+\frac{1}{2}}}{2} \cdot \phi_h + \hat{j}_i^n.$$

Combinations similar to the ones in the stability study prove that, under a CFL-like condition :

$$\begin{aligned}\hat{\varepsilon}^{n+1} - \hat{\varepsilon}^n = \sum_{i \in [0,N_T]} \Bigg[& \frac{\Delta t^2 \gamma_d}{4} \int_{T_i} \mathbb{J}_i^{n+\frac{1}{2}} \cdot \left(-\frac{\mathbb{E}_i^{n+1}}{1+\frac{\Delta t \gamma_d}{2}} + \frac{\mathbb{E}_i^n}{1-\frac{\Delta t \gamma_d}{2}} \right) + \frac{1}{\omega_d^2} \frac{\Delta t \gamma_d}{\frac{\Delta t^2 \gamma_d^2}{4} - 1} \left\| \mathbb{J}_i^{n+\frac{1}{2}} \right\|_{T_i}^2 \Bigg] \\ & + \hat{e}_h^n\left(\mathbb{E}_h^{\left[n+\frac{1}{2}\right]} \right) + \frac{1}{2}\left(\hat{h}_h^n\left(\mathbb{H}_h^{n+\frac{1}{2}} \right) + \hat{h}_h^{n+1}\left(\mathbb{H}_h^{n+\frac{1}{2}} \right) \right) + \frac{1}{2}\left(\hat{j}_h^n\left(\mathbb{J}_h^{n+\frac{1}{2}} \right) + \hat{j}_h^{n+1}\left(\mathbb{J}_h^{n+\frac{1}{2}} \right) \right).\end{aligned}$$

Following the same ideas, one obtains :

$$\left(\|\mathbb{H}_h^n\|_{L^2(\Omega)}^2 + \|\mathbb{E}_h^n\|_{L^2(\Omega)}^2 + \|\mathbb{J}_h^n\|_{L^2(\Omega)}^2 \right)^{\frac{1}{2}} \leq C\Delta t^2 \|\mathbf{U}\|_{\mathcal{C}^3([0,T],L^2(\Omega)^9)}. \tag{41}$$

It only remains to put together the theorem (3) along with (47) and the lemma 2 :

$$\begin{aligned}\|\mathbf{U}(t_n) - \mathbf{U}_h^n\|_{L^2(\Omega)^9} &\leq Ch^{\min(s,k)} \|\mathbf{U}\|_{\mathcal{C}^0([0,T],L^2(\Omega)^9)} + Ch^{\min(s,k)} \|\mathbf{U}\|_{\mathcal{C}^0\left([0,T],H^{s+1}(\Omega)^9\right)} \\ &+ C\Delta t^2 \|\mathbf{U}\|_{\mathcal{C}^3([0,T],L^2(\Omega)^9)}.\end{aligned}$$

Taking the maximum of the latter over all $n \in [0, N]$ then leads to the desired result.

□

5 Theoretical study of the generalized dispersive model

5.1 Stability study of the generalized dispersive formulation

5.1.1 Continuous equations

Let N be the total number of fields involved :

$$N = 2 + \text{card}\,(L_1) + 2\text{card}\,(L_2)\,.$$

We define the energy associated to the system (20) at a given time t as follows :

$$\begin{aligned}\xi(t) &= \frac{1}{2}\left(\|\mathbf{H}(t)\|^2_{L^2(\Omega)} + \varepsilon_\infty \|\mathbf{E}(t)\|^2_{L^2(\Omega)} + \sum_{l\in L_1}\frac{b_l}{a_l}\|\mathbf{P}_l(t)\|^2_{L^2(\Omega)}\right.\\ &+ \left.\sum_{l\in L_2}\frac{e_l}{c_l+d_l f_l}\|\mathbf{P}_l(t)\|^2_{L^2(\Omega)} + \sum_{l\in L_2}\frac{1}{c_l+d_l f_l}\|\mathbf{J}_l(t)-d_l\mathbf{E}(t)\|^2_{L^2(\Omega)}\right).\end{aligned}$$

Assuming a sufficient regularity for each of the fields, using equation (20) leads to :

$$\begin{aligned}\frac{\partial\xi}{\partial t} &= -\left(\sigma + \sum_{l\in L_1} a_l + \sum_{l\in L_2}\frac{c_l d_l}{c_l+d_l f_l}\right)\|\mathbf{E}\|^2_{L^2(\Omega)}\\ &- \sum_{l\in L_2}\frac{f_l}{c_l+d_l f_l}\|\mathbf{J}_l\|^2_{L^2(\Omega)} + \sum_{l\in L_2}\frac{e_l d_l}{c_l+d_l f_l}\mathbf{E}\cdot\mathbf{P}_l\\ &- \sum_{l\in L_1}\frac{b_l^2}{a_l}\|\mathbf{P}\|^2_{L^2(\Omega)} + \sum_{l\in L_1} 2b_l\mathbf{E}\cdot\mathbf{P}_l.\end{aligned}$$

The L_1 terms can be rewritten the following way :

$$-a_l\|\mathbf{E}\|^2_{L^2(\Omega)} - \frac{b_l^2}{a_l}\|\mathbf{P}_l\|^2_{L^2(\Omega)} + 2b_l\mathbf{E}\cdot\mathbf{P}_l = -\frac{b_l^2}{a_l}\left\|\frac{a_l}{b_l}\mathbf{E}-\mathbf{P}_l\right\|^2_{L^2(\Omega)}.$$

Then, every term is negative, except for the $\mathbf{E}\cdot\mathbf{P}_l$ one. The latter can be bounded in the following way :

$$\sum_{l\in L_2}\frac{e_l d_l}{c_l+d_l f_l}\mathbf{E}\cdot\mathbf{P}_l \le \alpha\xi,$$

where

$$\alpha = \frac{1}{\sqrt{\varepsilon_\infty}}\max_{L_2}\left(d_l\sqrt{\frac{e_l}{c_l+d_l f_l}}\right)\xi.$$

One might notice that $d_l = 0$ or $e_l = 0$ implies a decreasing energy. Although it is not provided in the current study, a physical interpretation of the d_l parameter might be of interest for future fitting improvements, since it appears to be responsible of the non-decreasing energy. Then :

$$\frac{\partial \xi}{\partial t} \leqslant \alpha \xi$$

which directly implies that, $\forall t \in [0,T]$:

$$\xi(t) \leqslant \xi(0) \exp(\alpha t).$$

5.1.2 Semi-discrete case

From now on, subscript indices l will be omitted in the sums over L_1 or L_2 sets, and a vector field written $\mathbf{A}_{L_u,h}$ must be understood as $\mathbf{A}_{l \in L_u,h}$. We define the semi-discrete fields $\left(\mathbf{H}_h, \mathbf{E}_h, \mathbf{P}_{L_1,h}, \mathbf{P}_{L_2,h}, \mathbf{J}_{L_2,h}\right)$ as solutions of the following weak formulation : $\forall \left(\varphi_h, \psi_h, \phi_h, \kappa_h, \Pi_h\right) \in \mathbf{V}_h^3, \forall t \in [0,T], \forall i \in [0, N_T]$,

$$\begin{cases} \displaystyle\int_{T_i} \frac{\partial \mathbf{H}_h}{\partial t} \cdot \varphi_h & = \displaystyle -\int_{T_i} \mathbf{E}_h \cdot (\nabla \times \varphi_h) + \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \varphi_h \cdot \left(\{\mathbf{E}_h\}_{ik} \times \mathbf{n}\right), \\ \varepsilon_\infty \displaystyle\int_{T_i} \frac{\partial \mathbf{E}_h}{\partial t} \cdot \psi_h & = \displaystyle\int_{T_i} \mathbf{H}_h \cdot (\nabla \times \psi_h) - \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} (\psi_h \times \{\mathbf{H}_h\}_{ik}) \cdot \mathbf{n} \\ & \displaystyle - \int_{T_i} \sigma \mathbf{E}_h \cdot \psi_h - \sum_{L_1} \int_{T_i} (a_l \mathbf{E}_h - \mathbf{P}_{l,h}) \cdot \psi_h \\ & \displaystyle - \sum_{L_2} \int_{T_i} \mathbf{J}_{l,h} \cdot \psi_h, \\ \displaystyle\int_{T_i} \frac{\partial \mathbf{P}_{l,h}}{\partial t} \cdot \phi_h & = \displaystyle\int_{T_i} a_l \mathbf{E}_h \cdot \phi_h - \int_{T_i} b_l \mathbf{P}_{l,h} \cdot \phi_h \quad l \in L_1, \\ \displaystyle\int_{T_i} \frac{\partial \mathbf{P}_{l,h}}{\partial t} \cdot \kappa_h & = \displaystyle\int_{T_i} \mathbf{J}_{l,h} \cdot \kappa_h \quad l \in L_2, \\ \displaystyle\int_{T_i} \left(\frac{\partial \mathbf{J}_{l,h}}{\partial t} - d_l \frac{\partial \mathbf{E}_h}{\partial t} \right) \cdot \Pi_h & = \displaystyle\int_{T_i} c_l \mathbf{E}_h \cdot \Pi_h - \int_{T_i} f_l \mathbf{J}_{l,h} \cdot \Pi_h - \int_{T_i} e_l \mathbf{P}_{l,h} \cdot \Pi_h \quad l \in L_2. \end{cases} \tag{42}$$

Following what has been done in the Drude case for the treatment of the boundaries, as well as the ideas of the continuous equations, one easily obtains :

$$\xi_h(t) \leqslant \xi_h(0) \exp(\alpha t).$$

5.1.3 Fully discrete scheme

The time discretization of the considered system can be written as follows :

$$
\begin{cases}
\displaystyle\int_{T_i} \frac{\mathbf{H}_i^{n+\frac{3}{2}} - \mathbf{H}_i^{n+\frac{1}{2}}}{\Delta t} \cdot \varphi_h &= -\displaystyle\int_{T_i} \mathbf{E}_i^{n+1} \cdot (\nabla \times \varphi_h) + \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \varphi_h \cdot \left(\left\{ \mathbf{E}_h^{n+1} \right\}_{ik} \times \mathbf{n}_{ik} \right), \\
\varepsilon_\infty \displaystyle\int_{T_i} \frac{\mathbf{E}_i^{n+1} - \mathbf{E}_i^{n}}{\Delta t} \cdot \psi_h &= \displaystyle\int_{T_i} \mathbf{H}_i^{n+\frac{1}{2}} \cdot (\nabla \times \psi_h) - \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \left(\psi_h \times \left\{ \mathbf{H}_i^{n+\frac{1}{2}} \right\}_{ik} \right) \cdot \mathbf{n}_{ik} \\
&\quad - \left(\sigma + \displaystyle\sum_{l \in L_1} a_l \right) \int_{T_i} \mathbf{E}_i^{[n+\frac{1}{2}]} \cdot \psi_h + \sum_{l \in L_1} \int_{T_i} \mathbf{P}_{l,i}^{n+\frac{1}{2}} \cdot \psi_h \\
&\quad - \displaystyle\sum_{l \in L_2} \int_{T_i} \mathbf{J}_{l,i}^{[n+\frac{1}{2}]} \cdot \psi_h, \\
\displaystyle\int_{T_i} \frac{\mathbf{P}_{l,i}^{n+\frac{3}{2}} - \mathbf{P}_{l,i}^{n+\frac{1}{2}}}{\Delta t} \cdot \phi_h &= \displaystyle\int_{T_i} \mathbf{E}_i^{n+1} \cdot \phi_h - b_l \int_{T_i} \frac{\mathbf{P}_{l,i}^{n+\frac{3}{2}} + \mathbf{P}_{l,i}^{n+\frac{1}{2}}}{2} \cdot \phi_h \quad l \in L_1, \\
\displaystyle\int_{T_i} \frac{\mathbf{P}_{l,i}^{n+\frac{3}{2}} - \mathbf{P}_{l,i}^{n+\frac{1}{2}}}{\Delta t} \cdot \kappa_h &= \displaystyle\int_{T_i} \mathbf{J}_{l,i}^{n+1} \cdot \kappa_h \quad l \in L_2, \\
\displaystyle\int_{T_i} \frac{\left(\mathbf{J}_{l,i} - d_l \mathbf{E}_i\right)^{n+1} - \left(\mathbf{J}_{l,i} - d_l \mathbf{E}_i\right)^{n}}{\Delta t} \cdot \Pi_h &= \displaystyle\int_{T_i} \mathbf{E}_i^{[n+\frac{1}{2}]} \cdot \Pi_h - f_l \int_{T_i} \mathbf{J}_{l,i}^{[n+\frac{1}{2}]} \cdot \Pi_h \\
&\quad - e_l \displaystyle\int_{T_i} \mathbf{P}_{l,i}^{n+\frac{1}{2}} \cdot \Pi_h \quad l \in L_2,
\end{cases}
\tag{43}
$$

and its associated energy for the cell T_i is :

$$
\begin{aligned}
\xi_i^n &= \frac{1}{2} \left(\int_{T_i} \mathbf{H}_i^{n+\frac{1}{2}} \cdot \mathbf{H}_i^{n-\frac{1}{2}} + \varepsilon_\infty \int_{T_i} \mathbf{E}_i^n \cdot \mathbf{E}_i^n + \sum_{l \in L_1} \frac{a_l}{b_l} \int_{T_i} \mathbf{P}_{l,i}^{n+\frac{1}{2}} \cdot \mathbf{P}_{l,i}^{n-\frac{1}{2}} \right. \\
&+ \left. \sum_{l \in L_2} \frac{1}{c_l + d_l f_l} \int_{T_i} \left(\mathbf{J}_{l,i} - d_l \mathbf{E}_i\right)^n \cdot \left(\mathbf{J}_{l,i} - d_l \mathbf{E}_i\right)^n + \sum_{l \in L_2} \frac{e_l}{c_l + d_l f_l} \int_{T_i} \mathbf{P}_{l,i}^{n+\frac{1}{2}} \cdot \mathbf{P}_{l,i}^{n-\frac{1}{2}} \right).
\end{aligned}
$$

Proposition 3 (Stability)**.** *The formulation (43) is stable under the following condition :*

$$
\Delta t \leq \min \left(\frac{h}{C}, \min \left(\sqrt{\frac{2}{e_l}}, \frac{1}{d_l}, \frac{2}{a_l + b_l}, \frac{\varepsilon_\infty}{\frac{C}{h} + \frac{1}{2} \sum_{L_1} \frac{a_l^2}{2 b_l} + \frac{1}{2} \sum_{L_2} \frac{e_l d_l}{c_l + d_l f_l}} \right) \right). \tag{44}
$$

Proof. An adequate choice of test functions for the scheme (43) similar to what was done in the Drude case allows us to write :

$$
\int_{T_i} \mathbf{P}_{l,i}^{n+\frac{1}{2}} \cdot \mathbf{P}_{l,i}^{n-\frac{1}{2}} = \frac{1}{1 + \square_l} \left((1 - \square_l) \left\| \mathbf{P}_{l,i}^{n-\frac{1}{2}} \right\|_{T_i}^2 + a_l \Delta t \int_{T_i} \mathbf{E}_i^n \cdot \mathbf{P}_{l,i}^{n-\frac{1}{2}} \right),
$$

for $l \in L_1$, with the definition $\square_l = \frac{b_l \Delta t}{2}$. For $l \in L_2$, one has :

$$
\int_{T_i} \mathbf{P}_{l,i}^{n+\frac{1}{2}} \cdot \mathbf{P}_{l,i}^{n-\frac{1}{2}} = \left\| \mathbf{P}_{l,i}^{n-\frac{1}{2}} \right\|_{T_i}^2 + \Delta t \int_{T_i} \mathbf{J}_{l,i}^n \cdot \mathbf{P}_{l,i}^{n-\frac{1}{2}}.
$$

The treatment of the **H** term is exactly identical to what has been donne in the Drude section. Therefore :

$$\begin{aligned}
\xi_i^n &= \frac{1}{2}\left(\|\mathbf{H}_i\|_{T_i}^2 + \varepsilon_\infty \|\mathbf{E}_i\|_{T_i}^2 + \sum_{l\in L_1} \frac{a_l(1-\square_l)}{b_l(1+\square_l)} \|\mathbf{P}_{l,i}\|_{T_i}^2 + \sum_{l\in L_2} \frac{e_l}{c_l+d_l f_l} \|\mathbf{P}_{l,i}\|_{T_i}^2 \right. \\
&+ \sum_{l\in L_2} \frac{1}{c_l+d_l f_l} \left\|\mathbf{J}_{l,i} - d_l \mathbf{E}_i\right\|_{T_i}^2 + \Delta t \sum_{k\in\mathcal{V}_i} \int_{a_{ik}} \mathbf{H}_i \cdot (\{\mathbf{E}_h\}_{ik} \times \mathbf{n}_{ik}) - \frac{\Delta t}{2} \sum_{k\in\mathcal{V}_i} \int_{a_{ik}} (\mathbf{H}_i \times \mathbf{E}_i)\cdot \mathbf{n}_{ik} \\
&+ \left. \sum_{l\in L_2} \frac{e_l}{c_l+d_l f_l} \Delta t \int_{T_i} \mathbf{J}_{l,i}\cdot\mathbf{P}_{l,i} + \sum_{l\in L_1} \frac{a_l}{b_l(1+\square_l)} a_l \Delta t \int_{T_i} \mathbf{E}_i \cdot \mathbf{P}_{l,i} \right).
\end{aligned}$$

If one assumes that $\Delta t \leq \frac{2}{b_l}\ \forall l \in L_1$, then $\frac{1}{1+\square_l} \geq \frac{1}{2}$. The minoration of ξ_i^n is similar to the Drude case, and one eventually obtains :

$$\begin{aligned}
\xi^n &\geq \frac{1}{2}\left(1 - \frac{C\Delta t}{h}\right) \|\mathbf{H}\|_\Omega^2 \\
&+ \frac{1}{2}\left(\varepsilon_\infty - \frac{C\Delta t}{h} - \frac{1}{2}\sum_{L_1} \frac{a_l^2 \Delta t}{2b_l} - \frac{1}{2}\sum_{L_2} \frac{e_l d_l \Delta t}{c_l + d_l f_l}\right) \|\mathbf{E}\|_\Omega^2 \\
&+ \frac{1}{2}\sum_{L_2} \frac{1}{c_l+d_l f_l}\left(1 - \frac{e_l \Delta t^2}{2}\right) \|\mathbf{J}_l - d_l \mathbf{E}\|_\Omega^2 \\
&+ \frac{1}{2}\sum_{L_2} \frac{e_l}{c_l+d_l f_l} (1 - \Delta t d_l) \|\mathbf{P}_l\|_\Omega^2 \\
&+ \frac{1}{2}\sum_{L_1} \frac{a_l}{2b_l}\left(1 - \square_l - \frac{a_l \Delta t}{2}\right) \|\mathbf{P}_l\|_\Omega^2 .
\end{aligned}$$

which leads to the desired result.

□

Proposition 4 (Bound on the discrete energy)**.** *Under CFL condition, the discrete energy ξ^n can be bounded in the following way :*

$$\xi^n \leq \frac{\xi^0}{\left(\frac{1-\theta}{1+2\theta}\right)^n},$$

with $\theta \geq 0$.

Proof. Following the same ideas as for the Drude model, one obtains :

$$
\begin{aligned}
\xi_i^{n+1} - \xi_i^n &= -\frac{\Delta t}{2} \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \left(\mathbf{E}_i^{\left[n+\frac{1}{2}\right]} \cdot \mathbf{H}_k^{n+\frac{1}{2}} + \mathbf{E}_k^{\left[n+\frac{1}{2}\right]} \cdot \mathbf{H}_i^{n+\frac{1}{2}} \right) \\
&- \Delta t \left(\sigma + \sum_{L_1} a_l \right) \left\| \mathbf{E}^{[n+\frac{1}{2}]} \right\|_{T_i}^2 \\
&+ \Delta t \sum_{L_1} \left(b_l + \frac{a_l^2}{b_l \left(1 + b_l \Delta t\right)} \right) \int_{T_i} \mathbf{P}_{l,i}^{n+\frac{1}{2}} \cdot \mathbf{E}_i^{[n+\frac{1}{2}]} \\
&- \Delta t \sum_{L_2} \int_{T_i} \mathbf{J}_{l,i}^{[n+\frac{1}{2}]} \cdot \mathbf{E}_i^{[n+\frac{1}{2}]} \\
&+ \Delta t \sum_{L_2} \frac{e_l}{c_l + d_l f_l} \int_{T_i} \mathbf{J}_{l,i}^{[n+\frac{1}{2}]} \cdot \mathbf{P}_{l,i}^{n+\frac{1}{2}} \\
&+ \Delta t \sum_{L_2} \frac{1}{c_l + d_l f_l} \int_{T_i} \left(c_l \mathbf{E}_i^{[n+\frac{1}{2}]} - f_l \mathbf{J}_{l,i}^{[n+\frac{1}{2}]} - e_l \mathbf{P}_{l,i}^{n+\frac{1}{2}} \right) \cdot \left(\mathbf{J}_{l,i} - d_l \mathbf{E}_i \right)^{[n+\frac{1}{2}]},
\end{aligned}
$$

which simplifies in

$$
\begin{aligned}
\xi_i^{n+1} - \xi_i^n &= -\frac{\Delta t}{2} \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \left(\mathbf{E}_i^{\left[n+\frac{1}{2}\right]} \cdot \mathbf{H}_k^{n+\frac{1}{2}} + \mathbf{E}_k^{\left[n+\frac{1}{2}\right]} \cdot \mathbf{H}_i^{n+\frac{1}{2}} \right) \\
&- \Delta t \left(\sigma + \sum_{L_1} a_l + \sum_{L_2} \frac{c_l d_l}{c_l + d_l f_l} \right) \left\| \mathbf{E}_i^{[n+\frac{1}{2}]} \right\|_{T_i}^2 \\
&- \Delta t \sum_{L_2} \frac{f_l}{c_l + d_l f_l} \left\| \mathbf{J}_{l,i}^{[n+\frac{1}{2}]} \right\|_{T_i}^2 \\
&+ \Delta t \sum_{L_1} \left(b_l + \frac{a_l^2}{b_l \left(1 + b_l \Delta t\right)} \right) \int_{T_i} \mathbf{P}_{l,i}^{n+\frac{1}{2}} \cdot \mathbf{E}_i^{[n+\frac{1}{2}]} \\
&+ \Delta t \sum_{L_2} \frac{d_l e_l}{c_l + d_l f_l} \int_{T_i} \mathbf{P}_{l,i}^{n+\frac{1}{2}} \cdot \mathbf{E}_i^{[n+\frac{1}{2}]}.
\end{aligned}
$$

Summing over all the cells and then bounding the scalar products the same way as before gives (remind that $\frac{1}{1+b_l \Delta t} \leq 1$), under CFL condition :

$$
\begin{aligned}
\xi_i^{n+1} - \xi_i^n &\leq \frac{\Delta t}{2} \sum_{L_1} b_l \left(1 + \frac{a_l^2}{b_l^2} \right) \left(\left\| \mathbf{P}_l^{n+\frac{1}{2}} \right\|_\Omega^2 + \left\| \mathbf{E}^{[n+\frac{1}{2}]} \right\|_\Omega^2 \right) \\
&+ \frac{\Delta t}{2} \sum_{L_2} \frac{d_l e_l}{c_l + d_l f_l} \left(\left\| \mathbf{P}_l^{n+\frac{1}{2}} \right\|_\Omega^2 + \left\| \mathbf{E}^{[n+\frac{1}{2}]} \right\|_\Omega^2 \right) \\
&\leq \theta \left(2\xi^n + \xi^{n+1} \right),
\end{aligned}
$$

with $\theta = \frac{\Delta t}{2} \left(\sum_{L_1} b_l \left(1 + \frac{a_l^2}{b_l^2} \right) + \sum_{L_2} \frac{d_l e_l}{c_l + d_l f_l} \right)$. The last inequality then leads to the result.

□

5.2 Convergence of the fully discrete generalized dispersive DG formulation

5.2.1 Convergence of the semi-discrete formulation

As in the Drude case, we start by defining the following bilinear forms. Let $\mathbf{U} = (V, W, X, Y, Z)$ and $\mathbf{U}' = (V', W', X', Y', Z')$, then :

$$
\begin{aligned}
m\left(\mathbf{U}, \mathbf{U}'\right) &= \int_\Omega V \cdot V' + \varepsilon_\infty \int_\Omega W \cdot W' + \sum_{L_1} \frac{b_l}{a_l} \int_\Omega X \cdot X' \\
&+ \sum_{L_2} \frac{1}{c_l + d_l f_l} \int_\Omega (Y - d_l W) \cdot (Y' - d_l W') + \sum_{L_2} \frac{1}{c_l + d_l f_l} \int_\Omega Z \cdot Z', \\
a\left(\mathbf{U}, \mathbf{U}'\right) &= \int_\Omega (V \cdot (\nabla \times W') - W \cdot (\nabla \times V')) - \int_\Omega \sigma W \cdot W' - \sum_{L_1} \int_\Omega (a_l W \cdot W' - b_l X \cdot W') \\
&+ \sum_{L_1} \int_\Omega (a_l W \cdot X' - b_l X \cdot X') - \sum_{L_2} \int_\Omega Y \cdot W' + \sum_{L_2} \int_\Omega Y \cdot Z' \\
&+ \sum_{L_2} \int_\Omega (c_l W - f_l Y - e_l Z) \cdot (Y' - d_l W'), \\
b\left(\mathbf{U}, \mathbf{U}'\right) &= \int_{\mathcal{F}_{\text{int}}} \{V\} [\![W']\!] - \int_{\mathcal{F}_{\text{int}}} \{W\} [\![V']\!] + \int_{\partial\Omega} W' \cdot (V \times \mathbf{n}).
\end{aligned}
$$

The semi-discrete scheme over the entire domain can be written as :

$$
m\left(\frac{\partial \mathbf{U}_h}{\partial t}, \mathbf{U}'_h\right) = a\left(\mathbf{U}_h, \mathbf{U}'_h\right) + b\left(\mathbf{U}_h, \mathbf{U}'_h\right).
$$

Moreover, the solution $\mathbf{U}$ of the continuous equations being solution of the semi-discrete scheme, one has :

$$
m\left(\frac{\partial \mathbf{U}}{\partial t}, \mathbf{U}'_h\right) = a\left(\mathbf{U}, \mathbf{U}'_h\right) + b\left(\mathbf{U}, \mathbf{U}'_h\right).
$$

Theorem 3 (Convergence of the semi-discrete formulation)**.** *Let :*

$$
\left(\boldsymbol{H}, \boldsymbol{E}, (\boldsymbol{P}_l)_{l\in L_1}, (\boldsymbol{P}_l)_{l\in L_2}, (\boldsymbol{J}_l - d_l \boldsymbol{E})_{l\in L_2}\right)
$$

be the solution of (20), and :

$$
\left(\boldsymbol{H}_h, \boldsymbol{E}_h, (\boldsymbol{P}_{h,l})_{l\in L_1}, (\boldsymbol{P}_{h,l})_{l\in L_2}, \left(\boldsymbol{J}_{h,l} - d_l \boldsymbol{E}\right)_{l\in L_2}\right) \in \mathcal{C}^1\left([0,T], \boldsymbol{V}_h^N\right)
$$

the semi-discrete solution of the associated semi-discrete formulation. If :

$$
\left(\boldsymbol{H}, \boldsymbol{E}, (\boldsymbol{P}_l)_{l\in L_1}, (\boldsymbol{P}_l)_{l\in L_2}, (\boldsymbol{J}_l - d_l \boldsymbol{E})_{l\in L_2}\right) \in \mathcal{C}^0\left([0,T], H^{s+1}(\Omega)^{3N}\right)
$$

for $s \geq 0$, then there exists $C \geq 0$ independent of h such that :

$$
\gamma(t)^{\frac{1}{2}} \leq C h^{\min(s,p)} \left\| \left(\boldsymbol{H}, \boldsymbol{E}, (\boldsymbol{P}_l)_{l\in L_1}, (\boldsymbol{P}_l)_{l\in L_2}, (\boldsymbol{J}_l - d_l \boldsymbol{E})_{l\in L_2}\right) \right\|_{\mathcal{C}^0\left([0,T], H^{s+1}(\Omega)^{3N}\right)} e^{\alpha t},
$$

where α has been defined in section 5.1.1, and :

$$\begin{aligned}\gamma(t) &= \|\pi_h(\boldsymbol{H}) - \boldsymbol{H}_h\|^2_{L^2(\Omega)} + \|\pi_h(\boldsymbol{E}) - \boldsymbol{E}_h\|^2_{L^2(\Omega)} \\ &+ \sum_{l\in L_1} \left\|\pi_h\left((\boldsymbol{P}_l)_{l\in L_1}\right) - (\boldsymbol{P}_{h,l})_{l\in L_1}\right\|^2_{L^2(\Omega)} + \sum_{l\in L_2} \left\|\pi_h\left((\boldsymbol{P}_l)_{l\in L_2}\right) - (\boldsymbol{P}_{h,l})_{l\in L_2}\right\|^2_{L^2(\Omega)} \\ &+ \sum_{l\in L_2} \left\|\pi_h\left((\mathbf{J}_l - d_l\boldsymbol{E})_{l\in L_2}\right) - \left(\mathbf{J}_{h,l} - d_l\boldsymbol{E}_h\right)_{l\in L_2}\right\|^2_{L^2(\Omega)}.\end{aligned}$$

Proof. Like previously, we have :

$$\varepsilon(t) = \frac{1}{2}\int_0^t m\left(\frac{\partial\left(\pi_h(\mathbf{U}) - \mathbf{U}_h\right)}{\partial s}, \pi_h(\mathbf{U}) - \mathbf{U}_h\right)\mathrm{d}s,$$

where $\mathbf{U}$ is the solution of the continuous problem. Following what has been done in the Drude case, we have :

$$\begin{aligned}m\left(\frac{\partial\pi_h(\mathbf{U})}{\partial t} - \frac{\partial\mathbf{U}_h}{\partial t}, \pi_h(\mathbf{U}) - \mathbf{U}_h\right) &= a\left(\pi_h(\mathbf{U}) - \mathbf{U}_h, \pi_h(\mathbf{U}) - \mathbf{U}_h\right) \\ &+ b\left(\pi_h(\mathbf{U}) - \mathbf{U}_h, \pi_h(\mathbf{U}) - \mathbf{U}_h\right) \\ &+ b\left(\mathbf{U} - \pi_h(\mathbf{U}), \pi_h(\mathbf{U}) - \mathbf{U}_h\right).\end{aligned}$$

As highlighted in section 5.1.1, the d_l parameter prevents the energy to be decreasing. Indeed, unlike the Drude case, one has :

$$\begin{aligned}&a\left(\pi_h(\mathbf{U}) - \mathbf{U}_h, \pi_h(\mathbf{U}) - \mathbf{U}_h\right) + b\left(\pi_h(\mathbf{U}) - \mathbf{U}_h, \pi_h(\mathbf{U}) - \mathbf{U}_h\right) \\ &\qquad = \sum_{L_2}\int_\Omega e_l d_l\left(\pi_h(Z) - Z_h\right)\cdot\left(\pi_h(W) - W_h\right),\end{aligned}$$

which can be bounded by a Cauchy-Schwarz inequality :

$$a\left(\pi_h(\mathbf{U}) - \mathbf{U}_h, \pi_h(\mathbf{U}) - \mathbf{U}_h\right) + b\left(\pi_h(\mathbf{U}) - \mathbf{U}_h, \pi_h(\mathbf{U}) - \mathbf{U}_h\right) \leq \alpha\varepsilon,$$

where α has been defined in section 5.1.1. The bounding of the term $b\left(\mathbf{U} - \pi_h(\mathbf{U}), \pi_h(\mathbf{U}) - \mathbf{U}_h\right)$ is strictly identical to what has been done for the Drude convergence proof. Therefore :

$$\begin{aligned}\varepsilon'(t) &\leq \alpha\varepsilon(t) + Ch^{\min(s,p)}\|(\mathbf{H},\mathbf{E})\|_{s+1,\Omega}\left(\|\pi_h(\mathbf{E}) - \mathbf{E}_h\|^2_{0,\Omega} + \|\pi_h(\mathbf{H}) - \mathbf{H}_h\|^2_{0,\Omega}\right)^{\frac{1}{2}} \\ &\leq \alpha\varepsilon(t) + Ch^{\min(s,p)}\left\|\left(\mathbf{H},\mathbf{E},(\mathbf{P}_l)_{l\in L_1},(\mathbf{P}_l)_{l\in L_2},(\mathbf{J}_l - d_l\mathbf{E})_{l\in L_2}\right)\right\|_{s+1,\Omega}\gamma(t)^{\frac{1}{2}} \\ &\leq \alpha\varepsilon(t) + Ch^{\min(s,p)}\left\|\left(\mathbf{H},\mathbf{E},(\mathbf{P}_l)_{l\in L_1},(\mathbf{P}_l)_{l\in L_2},(\mathbf{J}_l - d_l\mathbf{E})_{l\in L_2}\right)\right\|_{s+1,\Omega}\varepsilon(t)^{\frac{1}{2}}.\end{aligned}$$

Dividing by $\varepsilon(t)^{\frac{1}{2}}$, one obtains :

$$\frac{\varepsilon'(t)}{\varepsilon(t)^{\frac{1}{2}}} \leq \alpha\varepsilon(t)^{\frac{1}{2}} + Ch^{\min(s,p)}\left\|\left(\mathbf{H},\mathbf{E},(\mathbf{P}_l)_{l\in L_1},(\mathbf{P}_l)_{l\in L_2},(\mathbf{J}_l - d_l\mathbf{E})_{l\in L_2}\right)\right\|_{s+1,\Omega}.$$

Then, a Grönwall inequality gives, assuming that $\varepsilon(0) = 0$:

$$\varepsilon(t)^{\frac{1}{2}} \leq \frac{1}{\alpha} C h^{\min(s,p)} \left\| \left(\mathbf{H}, \mathbf{E}, (\mathbf{P}_l)_{l\in L_1}, (\mathbf{P}_l)_{l\in L_2}, (\mathbf{J}_l - d_l \mathbf{E})_{l\in L_2} \right) \right\|_{s+1,\Omega} \left(\mathrm{e}^{\alpha t} - 1 \right)$$

Therefore :

$$\gamma(t)^{\frac{1}{2}} \leq C h^{\min(s,p)} \delta \left\| \left(\mathbf{H}, \mathbf{E}, (\mathbf{P}_l)_{l\in L_1}, (\mathbf{P}_l)_{l\in L_2}, (\mathbf{J}_l - d_l \mathbf{E})_{l\in L_2} \right) \right\|_{\mathcal{C}^0\left([0,T], H^{s+1}(\Omega)^{3N}\right)} \mathrm{e}^{\alpha t},$$

which is the expected result.

□

5.2.2 Convergence of the fully discrete formulation

Theorem 4 (Convergence of the fully discrete formulation)**.** *Let be :*

$$\left(\boldsymbol{H}, \boldsymbol{E}, (\boldsymbol{P}_l)_{l\in L_1}, (\boldsymbol{P}_l)_{l\in L_2}, (\boldsymbol{J}_l - d_l \boldsymbol{E})_{l\in L_2} \right) \in \mathcal{C}^3\left([0,T], L^2(\Omega)^{3N} \right) \bigcap \mathcal{C}^0\left([0,T], H^{s+1}(\Omega)^{3N} \right).$$

Under the CFL condition (44), the following error estimate holds :

$$\begin{aligned}
&\max_{n\in[0,N]} \left(\left\| \boldsymbol{H}\left(t_{n+\frac{1}{2}}\right) - \boldsymbol{H}_h^{n+\frac{1}{2}} \right\|^2_{L^2(\Omega)^3} + \| \boldsymbol{E}(t_n) - \boldsymbol{E}_h^n \|^2_{L^2(\Omega)^3} + \sum_{L_1} \left\| \boldsymbol{P}_{l\in L_1}\left(t_{n+\frac{1}{2}}\right) - \boldsymbol{P}_{h,l\in L_1}^{n+\frac{1}{2}} \right\|^2_{L^2(\Omega)^3} \right. \\
&\left. + \sum_{L_2} \left\| \boldsymbol{P}_{l\in L_2}\left(t_{n+\frac{1}{2}}\right) - \boldsymbol{P}_{h,l\in L_2}^{n+\frac{1}{2}} \right\|^2_{L^2(\Omega)^3} + \sum_{L_2} \left\| (\boldsymbol{J}_l - d_l \boldsymbol{E})_{l\in L_2}(t_n) - \left(\boldsymbol{J}_{h,l} - d_l \boldsymbol{E}_h\right)^n_{l\in L_2} \right\|^2_{L^2(\Omega)^3} \right)^{\frac{1}{2}} \\
&\leq C\left(\Delta t^2 + h^{\min(s,p)}\right) \left(\left\| \left(\boldsymbol{H}, \boldsymbol{E}, (\boldsymbol{P}_l)_{l\in L_1}, (\boldsymbol{P}_l)_{l\in L_2}, (\boldsymbol{J}_l - d_l \boldsymbol{E})_{l\in L_2} \right) \right\|_{\mathcal{C}^3\left([0,T], L^2(\Omega)^{3N}\right)} \right. \\
&\left. + \left\| \left(\boldsymbol{H}, \boldsymbol{E}, (\boldsymbol{P}_l)_{l\in L_1}, (\boldsymbol{P}_l)_{l\in L_2}, (\boldsymbol{J}_l - d_l \boldsymbol{E})_{l\in L_2} \right) \right\|_{\mathcal{C}^0\left([0,T], H^{s+1}(\Omega)^{3N}\right)} \right).
\end{aligned}$$

Proof. The consistency error is defined as follows :

$$\begin{aligned}
\varepsilon_h^{n+1} &= \left(\left\| \mathbf{E}_h(t_{n+1}) - \widetilde{\mathbf{E}}_h^{n+1} \right\|^2_{L^2(\Omega)} + \left\| \mathbf{H}_h\left(t_{n+\frac{3}{2}}\right) - \widetilde{\mathbf{H}}_h^{n+\frac{3}{2}} \right\|^2_{L^2(\Omega)} \right. \\
&+ \sum_{L_1} \left\| \mathbf{P}_{h,l\in L_1}\left(t_{n+\frac{3}{2}}\right) - \widetilde{\mathbf{P}}_{h,l\in L_1}^{n+\frac{3}{2}} \right\|^2_{L^2(\Omega)} + \sum_{L_2} \left\| \mathbf{P}_{h,l\in L_2}\left(t_{n+\frac{3}{2}}\right) - \widetilde{\mathbf{P}}_{h,l\in L_2}^{n+\frac{3}{2}} \right\|^2_{L^2(\Omega)} \\
&+ \left. \sum_{L_2} \left\| \left(\mathbf{J}_{h,l} - d_l \mathbf{E}_h\right)_{l\in L_2}(t_{n+1}) - \left(\widetilde{\mathbf{J}}_{h,l} - d_l \widetilde{\mathbf{E}}_h\right)^{n+1}_{l\in L_2} \right\|^2_{L^2(\Omega)} \right)^{\frac{1}{2}},
\end{aligned}$$

where $\widetilde{\mathbf{E}}_h^{n+1}, \widetilde{\mathbf{H}}_h^{n+\frac{3}{2}}, \widetilde{\mathbf{P}}_{h,l\in L_1}^{n+\frac{3}{2}}, \widetilde{\mathbf{P}}_{h,l\in L_2}^{n+\frac{3}{2}}$ and $\left(\widetilde{\mathbf{J}}_{h,l} - d_l \widetilde{\mathbf{E}}_h\right)^{n+1}_{l\in L_2}$ are defined as :

$$
\left\{
\begin{array}{rcl}
\displaystyle\int_{T_i} \frac{\widetilde{\mathbf{H}}_i^{n+\frac{3}{2}} - \mathbf{H}_i\left(t_{n+\frac{1}{2}}\right)}{\Delta t} \cdot \varphi_h &=& \displaystyle -\int_{T_i} \mathbf{E}_i\left(t_{n+1}\right) \cdot \left(\nabla \times \varphi_h\right) + \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \varphi_h \cdot \left(\left\{\mathbf{E}_h\left(t_{n+1}\right)\right\}_{ik} \times \mathbf{n}_{ik}\right), \\
\varepsilon_\infty \displaystyle\int_{T_i} \frac{\widetilde{\mathbf{E}}_i^{n+1} - \mathbf{E}_i\left(t_n\right)}{\Delta t} \cdot \psi_h &=& \displaystyle \int_{T_i} \mathbf{H}_i\left(t_{n+\frac{1}{2}}\right) \cdot \left(\nabla \times \psi_h\right) - \sum_{k \in \mathcal{V}_i} \int_{a_{ik}} \left(\psi_h \times \left\{\mathbf{H}_i\left(t_{n+\frac{1}{2}}\right)\right\}_{ik}\right) \cdot \mathbf{n}_{ik} \\
& & \displaystyle -\left(\sigma + \sum_{L_1} a_l\right) \int_{T_i} \frac{\mathbf{E}_i\left(t_{n+1}\right) + \mathbf{E}_i\left(t_n\right)}{2} \cdot \psi_h + \sum_{L_1} \int_{T_i} \mathbf{P}_{l,i}\left(t_{n+\frac{1}{2}}\right) \cdot \psi_h \\
& & \displaystyle -\sum_{L_2} \int_{T_i} \frac{\mathbf{J}_{l,i}\left(t_{n+1}\right) + \mathbf{J}_{l,i}\left(t_n\right)}{2} \cdot \psi_h, \\
\displaystyle\int_{T_i} \frac{\widetilde{\mathbf{P}}_{l,i}^{n+\frac{3}{2}} - \mathbf{P}_{l,i}\left(t_{n+\frac{1}{2}}\right)}{\Delta t} \cdot \phi_h &=& \displaystyle \int_{T_i} \mathbf{E}_i\left(t_{n+1}\right) \cdot \phi_h - b_l \int_{T_i} \frac{\mathbf{P}_{l,i}\left(t_{n+\frac{3}{2}}\right) + \mathbf{P}_{l,i}\left(t_{n+\frac{1}{2}}\right)}{2} \cdot \phi_h \quad l \in L_1, \\
\displaystyle\int_{T_i} \frac{\widetilde{\mathbf{P}}_{l,i}^{n+\frac{3}{2}} - \mathbf{P}_{l,i}\left(t_{n+\frac{1}{2}}\right)}{\Delta t} \cdot \kappa_h &=& \displaystyle \int_{T_i} \mathbf{J}_{l,i}\left(t_{n+1}\right) \cdot \kappa_h \quad l \in L_2, \\
\displaystyle\int_{T_i} \frac{\left(\widetilde{\mathbf{J}}_{l,i} - d_l \widetilde{\mathbf{E}}_i\right)^{n+1} - \left(\widetilde{\mathbf{J}}_{l,i} - d_l \widetilde{\mathbf{E}}_i\right)\left(t_n\right)}{\Delta t} \cdot \Pi_h &=& \displaystyle \int_{T_i} \frac{\mathbf{E}_i\left(t_{n+1}\right) + \mathbf{E}_i\left(t_n\right)}{2} \cdot \Pi_h \\
& & \displaystyle -f_l \int_{T_i} \frac{\mathbf{J}_{l,i}\left(t_{n+1}\right) + \mathbf{J}_{l,i}\left(t_n\right)}{2} \cdot \Pi_h \\
& & \displaystyle -e_l \int_{T_i} \mathbf{P}_{l,i}\left(t_{n+\frac{1}{2}}\right) \quad l \in L_2.
\end{array}
\right.
$$

Following the same procedure as for the Drude case, one can prove that :

$$
\left|\varepsilon_h^n\right| \leq C \Delta t^3 \left\|\mathbf{U}\right\|_{\mathcal{C}^3\left([0,T], L^2(\Omega)^{15}\right)}, \tag{45}
$$

which implies :

$$
\left|\!\left|\!\left| \hat{h}_h^n \right|\!\right|\!\right| + \left|\!\left|\!\left| \hat{e}_h^n \right|\!\right|\!\right| + \left|\!\left|\!\left| \hat{p}_{h,l\in L_1}^n \right|\!\right|\!\right| + \left|\!\left|\!\left| \hat{p}_{h,l\in L_2}^n \right|\!\right|\!\right| + \left|\!\left|\!\left| \hat{j}_{h,l\in L_2}^n - d_l \hat{e}_h^n \right|\!\right|\!\right| \leq C \Delta t^2 \left\|\mathbf{U}\right\|_{\mathcal{C}^3([0,T], L^2(\Omega)^{15})}. \tag{46}
$$

where the natural definitions of $\hat{p}_{h,l\in L_1}^n$, $\hat{p}_{h,l\in L_2}^n$ and $\hat{j}_{h,l\in L_2}^n - d_l \hat{e}_h^n$ are coherent with the one given in the Drude case. The error energy is defined as :

$$
\begin{array}{rcl}
\hat{\varepsilon}_i^n &=& \displaystyle \frac{1}{2}\left(\varepsilon_\infty \left\|\mathbb{E}_i^n\right\|_{L^2(\Omega)}^2 + \left\langle \mathbb{H}_i^{n-\frac{1}{2}}, \mathbb{H}_i^{n+\frac{1}{2}} \right\rangle_{T_i} + \sum_{L_1} \frac{a_l}{b_l} \left\langle \mathbb{P}_{l,i}^{n-\frac{1}{2}}, \mathbb{P}_{l,i}^{n+\frac{1}{2}} \right\rangle_{T_i}\right. \\
&+& \displaystyle \left. \sum_{L_2} \frac{e_l}{c_l + d_l f_l} \left\langle \mathbb{P}_{l,i}^{n-\frac{1}{2}}, \mathbb{P}_{l,i}^{n+\frac{1}{2}} \right\rangle_{T_i} + \sum_{L_2} \frac{1}{c_l + d_l f_l} \left\| \mathbb{J}_{l,i}^n - d_l \mathbb{E}_i^n \right\|_{L^2(\Omega)}^2 \right),
\end{array}
$$

where the notations $\mathbb{U}_i^n$ has been defined to the Drude case. Under CFL condition, one readily gets :

$$\begin{aligned}\hat{\varepsilon}^{n+1} - \hat{\varepsilon}^{n} = \sum_{i\in[0,N_T]} \Delta t \Bigg[& \sum_{L_1} \left(b_l + \frac{a_l^2}{b_l\left(1+b_l\Delta t\right)} \right) \int_{T_i} \mathbb{P}_{l,i}^{n+\frac{1}{2}} \cdot \mathbb{E}_i^{[n+\frac{1}{2}]} + \sum_{L_2} \frac{d_l e_l}{c_l + d_l f_l} \int_{T_i} \mathbb{P}_{l,i}^{n+\frac{1}{2}} \cdot \mathbb{E}_i^{[n+\frac{1}{2}]} \\ & - \left(\sigma + \sum_{L_1} a_l + \sum_{L_2} \frac{c_l d_l}{c_l + d_l f_l} \right) \left\| \mathbb{E}_i^{[n+\frac{1}{2}]} \right\|_{T_i}^2 - \sum_{L_2} \frac{f_l}{c_l + d_l f_l} \left\| \mathbb{J}_{l,i}^{[n+\frac{1}{2}]} \right\|_{T_i}^2 \Bigg] + \hat{e}_h^n \left(\mathbb{E}_h^{\left[n+\frac{1}{2}\right]} \right) \\ & + \frac{1}{2} \left(\hat{h}_h^n \left(\mathbb{H}_h^{n+\frac{1}{2}} \right) + \hat{h}_h^{n+1} \left(\mathbb{H}_h^{n+\frac{1}{2}} \right) \right) + \frac{1}{2} \sum_{L_1} \left(\hat{p}_{h,l}^n \left(\mathbb{P}_{h,l}^{n+\frac{1}{2}} \right) + \hat{p}_{h,l}^{n+1} \left(\mathbb{P}_{h,l}^{n+\frac{1}{2}} \right) \right) \\ & + \frac{1}{2} \sum_{L_2} \left(\hat{p}_{h,l}^n \left(\mathbb{P}_{h,l}^{n+\frac{1}{2}} \right) + \hat{p}_{h,l}^{n+1} \left(\mathbb{P}_{h,l}^{n+\frac{1}{2}} \right) \right) + \frac{1}{2} \sum_{L_2} \left(\hat{j}_{h,l}^n - d_l \hat{e}_h^n \right) \left(\left(\mathbb{J}_{l,i} - d_l \mathbb{E}_i \right)^{\left[n+\frac{1}{2}\right]} \right).\end{aligned}$$

Considering inequality (46) as well as classical bounding methods, the following result is obtained :

$$\begin{aligned}\Bigg(& \left\| \mathbb{H}_h^n \right\|_{L^2(\Omega)}^2 + \left\| \mathbb{E}_h^n \right\|_{L^2(\Omega)}^2 + \left\| \mathbb{P}_{h,l\in L_1}^n \right\|_{L^2(\Omega)}^2 + \left\| \mathbb{P}_{h,l\in L_2}^n \right\|_{L^2(\Omega)}^2 \\ & + \left\| \left(\mathbb{J}_{h,l} - d_l \mathbb{E}_h \right)_{l\in L_2}^n \right\|_{L^2(\Omega)}^2 \Bigg)^{\frac{1}{2}} \leq C\Delta t^2 \left\| \mathbf{U} \right\|_{\mathcal{C}^3([0,T],L^2(\Omega)^{3N})}.\end{aligned} \tag{47}$$

The end of the proof is then similar to the Drude case.

□

6 Numerical results for the Drude model

This section presents some simulation results for different situations : first, an artificially-built case is considered to validate the implementation of the method. Then, a physical situation is considered, and comparisons are made against theoretical results.

6.1 Artificial validation case

6.1.1 Presentation

A cubic cavity of length L with metallic walls is considered. The goal is to simulate a standing wave in the cavity, given a *volumic* source. To do so, the analytical solution is built from the classical cavity modes. For the **E** and **H** fields, the following expressions are looked for

$$\mathbf{E}(\mathbf{x},t) = \begin{pmatrix} -f(\mathbf{x}) \\ 0 \\ g(\mathbf{x}) \end{pmatrix} \cos\left(x_k t\right),$$

$$\mathbf{H}(\mathbf{x},t) = \begin{pmatrix} h(\mathbf{x}) \\ i(\mathbf{x}) \\ j(\mathbf{x}) \end{pmatrix} \sin\left(x_k t\right),$$

where $x_m = \frac{n_k \pi}{L}$, n_k representing the number of the mode. We used :

$$\begin{cases} f(x,y,z) &= \cos(x_m x)\sin(x_m y)\sin(x_m z), \\ g(x,y,z) &= \sin(x_m x)\sin(x_m y)\cos(x_m z), \\ h(x,y,z) &= \sin(x_m x)\cos(x_m y)\cos(x_m z), \\ i(x,y,z) &= \cos(x_m x)\sin(x_m y)\cos(x_m z), \\ j(x,y,z) &= \cos(x_m x)\cos(x_m y)\sin(x_m z). \end{cases}$$

Given that the previous modes cannot be solutions of (18), a volumic source is added in the second equation :

$$\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon_\infty}\left(\nabla \times \mathbf{H} - \mathbf{J}_p + \mathbf{J}_s\right), \tag{48}$$

where the source current $\mathbf{J}_s$ is defined as :

$$\mathbf{J}_s(\mathbf{x},t) = \begin{pmatrix} \left(\beta \sin(x_k t) - \gamma_d \cos(x_k t)\right) f(\mathbf{x}) \\ 0 \\ \left(-\beta \sin(x_k t) + \gamma_d \cos(x_k t)\right) g(\mathbf{x}) \end{pmatrix}, \tag{49}$$

assuming that $\mathbf{J}_p(x,y,z,0)$ is taken equal to zero. The parameters used in the previous expression are defined as follows :

$$\alpha = \frac{\omega_d^2}{x_k^2 + \gamma_d^2},$$

$$\beta = \varepsilon_\infty x_k - 3\frac{x_m^2}{x_k} - \alpha x_k.$$

Being given these expressions, the analytical solution for the polarization current is :

$$\mathbf{J}_p(\mathbf{x},t) = \begin{pmatrix} -\alpha\left(x_k \sin(x_k t) + \gamma_d \cos(x_k t)\right) f(\mathbf{x}) \\ 0 \\ \alpha\left(x_k \sin(x_k t) + \gamma_d \cos(x_k t)\right) g(\mathbf{x}) \end{pmatrix}. \tag{50}$$

The chosen Drude parameters for this case are the ones given in section 1.2.1, table 1. The (normalized) maximal time t_{max}, (here $2.0\,10^{-6}$), and the cavity length are chosen in order to simulate two temporal periods. The spatial period is related to the parameter n_k, that describes the spatial shape of the mode. Eventually, the frequency f is taken equal to $3.0\,10^5$ GHz.

6.1.2 Results

We denote $\left(\mathbf{E},\mathbf{H},\mathbf{J}_p\right)$ the exact solution of (48), and $(\mathbf{E}_h,\mathbf{H}_h,\mathbf{J}_h)$ the calculated solution. The results obtained from the DG calculation are coherent and close to the analytical ones : one can refer to figure 5 for some exact and calculated field plots. At a given time t_n, the total L^2 error is :

$$e_h^n = \left(\left\| \mathbf{H}\left(t_{n+\frac{1}{2}}\right) - \mathbf{H}_h^{n+\frac{1}{2}} \right\|^2_{L^2(\Omega)} + \varepsilon_\infty \left\| \mathbf{E}\left(t_n\right) - \mathbf{E}_h^n \right\|^2_{L^2(\Omega)} \right. \tag{51}$$

$$\left. + \frac{1}{\omega_d^2} \left\| \mathbf{J}_p\left(t_{n+\frac{1}{2}}\right) - \mathbf{J}_h^{n+\frac{1}{2}} \right\|^2_{L^2(\Omega)} \right)^{\frac{1}{2}} \tag{52}$$

For a given time step Δt verifying the CFL condition, the convergence order is calculated with DG-$\mathbb{P}_1$ and DG-$\mathbb{P}_2$ methods. The results are summarized in tables 5(a) and 5(b) respectively. One should notice that, for higher approximation orders, the convergence order should remain bounded to 2, given second-order accuracy of the LF2 time approximation. A visual representation of the time evolution L^2 error for different orders of approximation is displayed on figure 6.
This case permits to validate the dispersive media DG code for the Drude model. Nevertheless, it does not represent any physical situation. The following sections will therefore present more realistic cases.

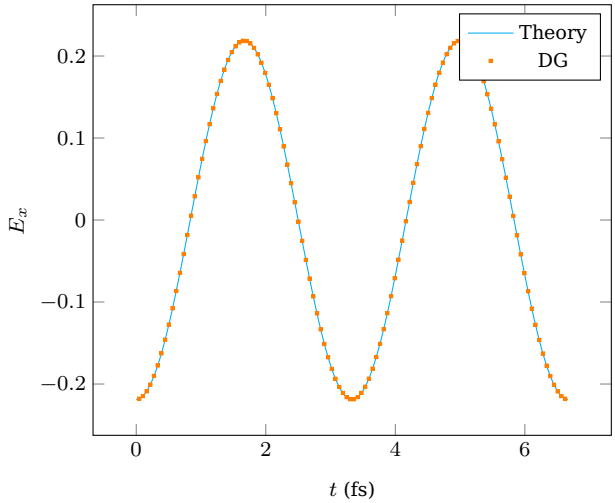


Figure 5 | Exact and calculated E_x fields with a $\mathbb{P}_1$ approximation for the artificial case.

Table 5 | Convergence orders with $\mathbb{P}_1$ and $\mathbb{P}_2$ approximations. The total L^2 error is calculated with the discrete equivalent of expression (51).

(a) Convergence rate with $\mathbb{P}_1$

Refinement	Convergence rate
$\frac{1}{50}$	–
$\frac{1}{75}$	1.2575
$\frac{1}{100}$	1.1197
$\frac{1}{125}$	1.1000
$\frac{1}{150}$	1.0614

(b) Convergence rate with $\mathbb{P}_2$

Refinement	Convergence rate
$\frac{1}{25}$	–
$\frac{1}{50}$	2.2004
$\frac{1}{75}$	2.0826
$\frac{1}{100}$	2.0366
$\frac{1}{125}$	2.0432

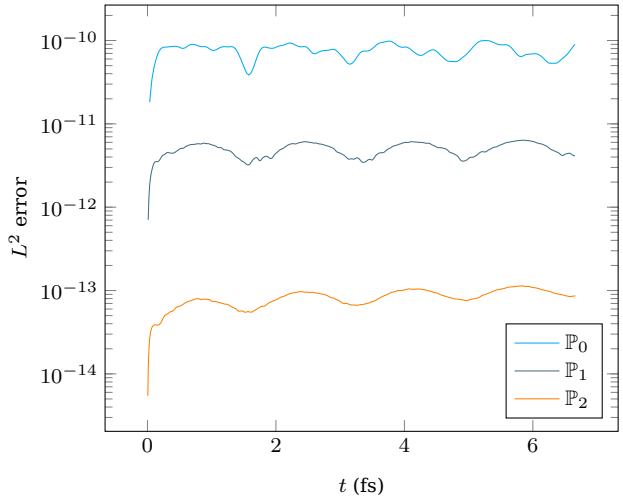


Figure 6 | L^2 error for different orders of approximation.

6.2 Near-field enhancement of a gold nanosphere

6.2.1 Presentation

A gold sphere of radius R and centered at $(0,0,0)$ immersed in vacuum is considered, within a spherical domain. Silver-Muller absorbing boundary conditions (ABC) are used, and the sphere is illuminated with a sinusoidal plane wave propagating in the $\hat{\mathbf{z}}$ direction, which amplitude is modulated in time with a gaussian profile. Its central frequency is denoted as f_c, and τ represents its initial phase delay :

$$\mathbf{E}_{inc}(t) = \sin\left(2\pi f_c(t-4\tau)\right)\ \mathrm{e}^{-\left(\frac{t-4\tau}{\tau}\right)^2}\mathbf{e}_x \tag{53}$$

The permittivity of gold is considered to follow a Drude model, which parameters are summed up in table 6. The situation is represented on figure 7. In this case, analytical solutions exist in the frequency domain. Their expressions result from the Mie theory, which is extensively described in [vdH81]. The nearfield is here considered, and the sphere radius is taken equal to 20 nm.

Table 6 | Parameters set for the gold nanosphere case.

ε_∞	ω_0	γ_d	f_c	τ
–	*GHz*	*GHz*	*GHz*	*s*
1	$1.19\,10^7$	$1.41\,10^5$	$4.5\,10^5$	$2\,10^{-15}$

6.2.2 Results

We focus on the amplitude of the total field in the vicinity of the gold nanosphere. In this case, the Mie theory predicts an enhanced field at the poles of the sphere. The reference

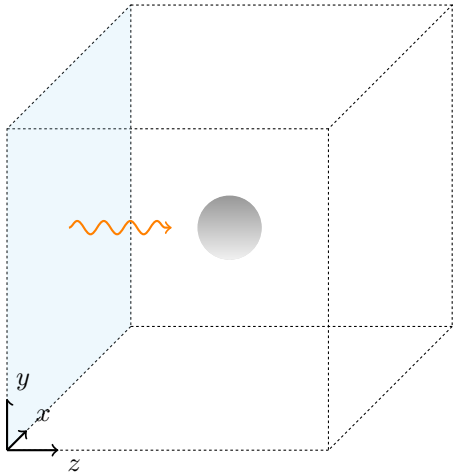


Figure 7 | Physical situation : gold nanosphere illuminated with a plane wave.

solution used is given by a Matlab script[7] which also exploits a Drude model. The reference solution being a frequency domain one, the DFT of the DG solution at the central frequency is calculated at each point of the domain.

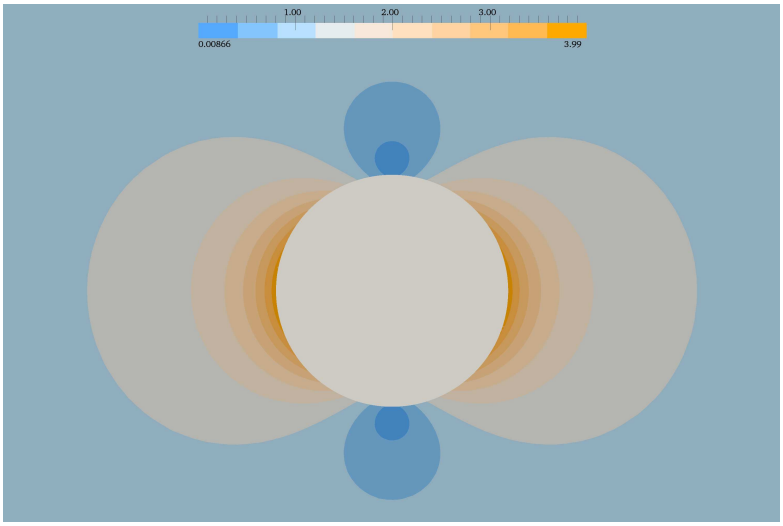
As can be seen on figure 8, there is a very good agreement between the reference and numerical solutions, regarding both the maximum field value and the spatial extension of the resonance lobes. A more precise analysis of the results can be made by looking at the evolution of the modulus of the electric field along the contour line of equation $\{y = 0, z = 0\}$, which crosses the nanosphere at its center along the $x+$ direction. Such plots have been made for various mesh refinements and orders of interpolation, and are presented in figure 9. The L^1 error levels for these plots can be found on table 7, along with the meshes caracteristics. One could see that increasing refinement and interpolation order lead to a sharper field jump at the sphere interface. It also seems that neither of them manage to obtain a significantly better fit of the decreasing field outside of the sphere. This might be a consequence of the first-order geometrical approximation of the sphere, but that point remains to be verified.

Table 7 | L^1 errors for various meshes and orders of approximation in the case of the gold nanosphere nearfield enhancement.

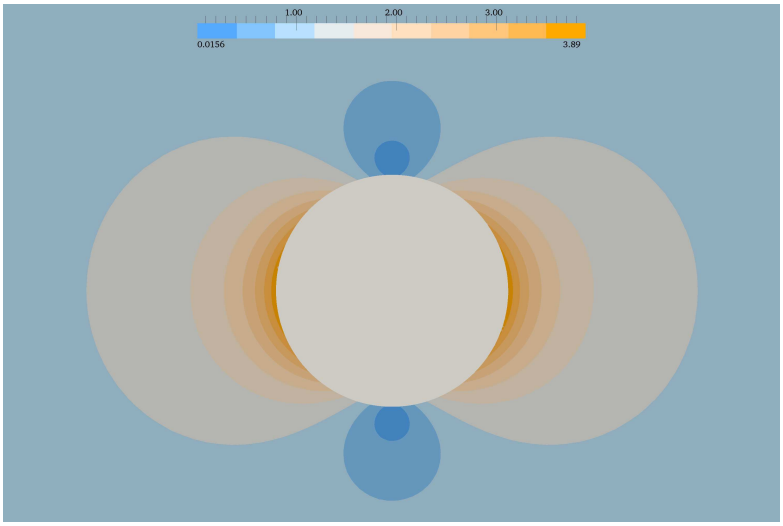
Mesh	n_s	n_t	r	$L^1_{\mathbb{P}_1}$	$L^1_{\mathbb{P}_2}$
M1	26128	153517	17.66	$1.1081\,10^{-8}$	$8.3051\,10^{-9}$
M2	146818	881154	19.77	$7.6520\,10^{-9}$	$6.6986\,10^{-9}$
M3	389955	2338433	24.77	$6.6025\,10^{-9}$	–

As expected, increasing the interpolation order from 1 to 2 implies a roughly comparable relative increase in the computational cost (approximately 300%). On the $M1$ mesh, this leads to a 25% drop in the L^1 error, whereas on the $M2$ mesh, it only lowers of 12%. This

[7] Code developed by Guangran Kevin Zhu, available at http://www.mathworks.com/matlabcentral/fileexchange/31119-sphere-scattering

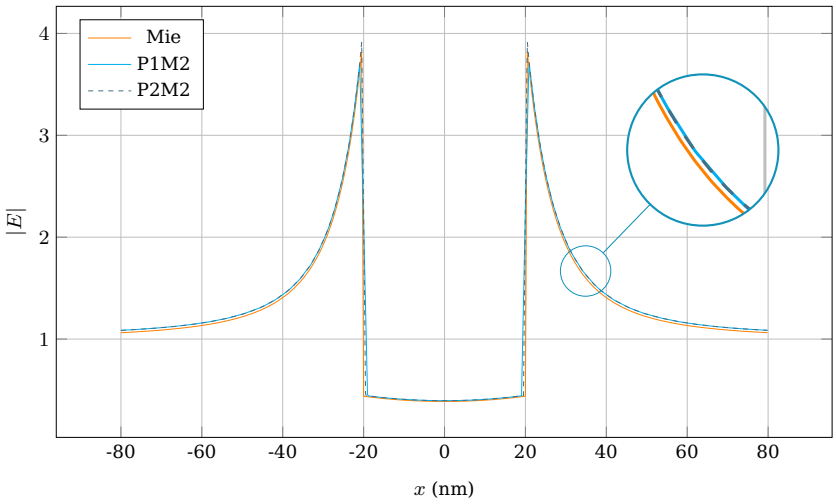


(a) DG solution

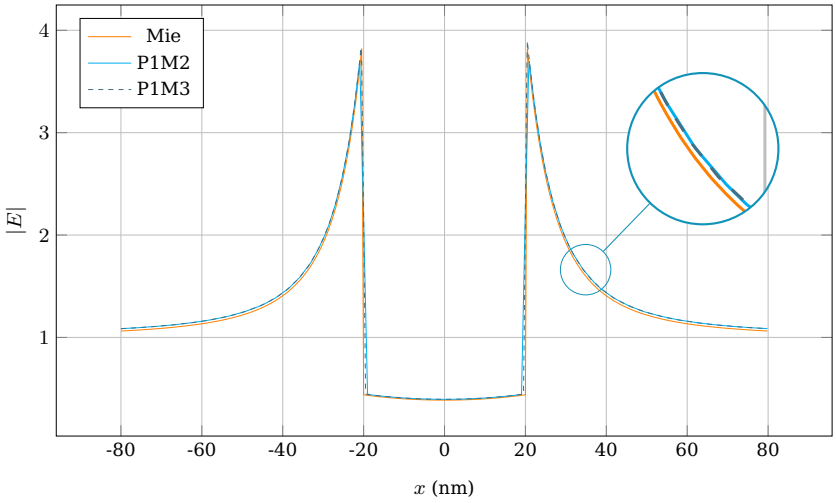


(b) Mie solution

Figure 8 | Comparison of DG and Mie solutions for the near-field enhancement of a gold nanosphere.



(a) Increasing order of approximation

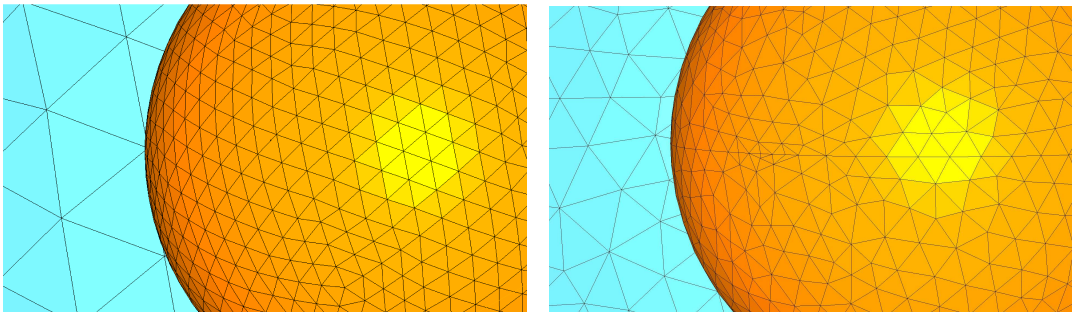


(b) Increasing mesh refinement

Figure 9 | Mie and DG 1D plot of the electric field modulus across the dispersive gold nanosphere for various meshes and approximation ordrers.

is a logical consequence of the first-order approximation of the geometry of the sphere by straight-edged tetrahedrons, and the use of curvilinear elements will be considered in a future work. Another improvement could consist in a p-adaptative formulation, which would get the best from both high interpolation orders and mesh refinement, using large cells with high p values away from the interface, and small cells with low p values in the vicinity of the interface.

***Remark** : It has been noticed that a good numerical approximation of the field jump across the sphere interface also relies on a good homogeneity of the surface mesh of the sphere. The presented results were made using a particularly homogenous surface mesh for the sphere, as can be seen on figure 10(a). An example of a poor quality surface mesh is also shown.*



(a) Homogeneous surfacic mesh (b) Poor quality surfacic mesh

Figure 10 | Comparison of two surfacic meshes for the gold nanosphere.

6.2.3 Conclusion

On a physical point of view, it should be noted that the resonance phenomenon presented hereabove is a consequence of the dispersive behaviour of the metal only, and therefore does not appear when using a PEC sphere. These kinds of resonances can be combined in numerous ways (in [Tei08], the reader can find the description of a L-junction device made of a linear assembly of metallic spheres; an extensive description of split-ring resonators is presented in [Die12]; although, the latter references only represent a fraction of the litterature dealing with these kind of devices, and the reader can find many more without difficulties) with various geometries to achieve useful nanoscale devices. This textbook case, limited to a simple model and geometry, opens the way to more complicated geometries, models and methodology developments. While the following section is devoted to a generalized dispersive model, curvilinear elements and p-adaptative formulations will be at the center of future efforts.

7 Reflection coefficient of a silver slab described by various dispersion models

7.1 Presentation

This section presents a validation case for the DGTD formulation presented in section 3.3.2. As shown on figure 11, a silver slab is illuminated by a plane wave, described by equation (53). The central frequency f_c and initial phase delay τ are respectively chosen equal to 900 THz and $4\,10^{-16}s$ in order to obtain an acceptable spectral density over the whole chosen frequency range, which is $[300, 1500]$ THz. The slab is parallelepipedic, its side length being equal to 150 nm, and its thickness l to 10 nm.

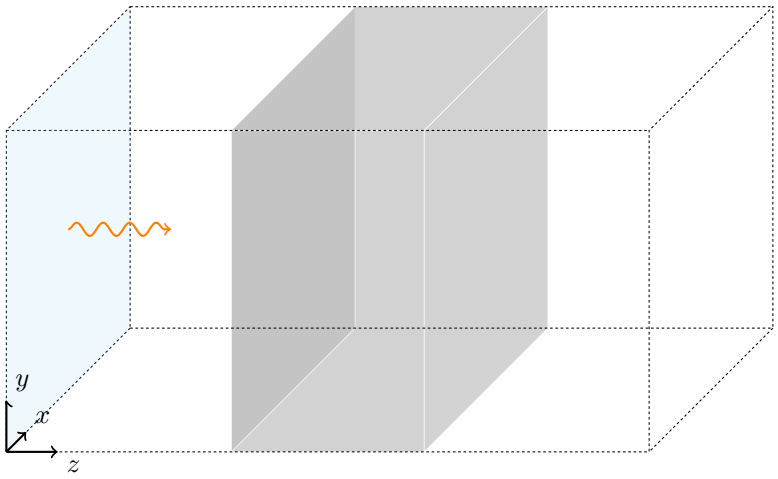


Figure 11 | Physical situation : silver slab illuminated with a plane wave.

The incident field on the slab is noted $\mathbf{E}_i$, and the reflected field $\mathbf{E}_r$. The reflection coefficient is then defined as :

$$R = \frac{|\mathbf{E}_r|}{|\mathbf{E}_i|}.$$

For a monochromatic wave of given frequency f_m and amplitude 1, the reflection coefficient can be calculated analytically. Let $\varepsilon_s(f_c) and \mu_s(f_c)$ be the permittivity and permeability of the slab at frequency f_c, and ε_0, μ_0 those of vacuum. Their respective impedances are therefore expressed as :

$$Z_s(f_c) = \sqrt{\frac{\mu_s}{\varepsilon_s}}(f_c),\ Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}}.$$

A few lines of hand calculation to write the matching conditions of the fields at the two interfaces between the slab and the vacuum leads to :

$$R(f_c) = \frac{g_{0s} + g_{s0}\mathrm{e}^{-2ikl}}{1 + g_{0s}g_{s0}\mathrm{e}^{-2ikl}},$$

where :

$$g_{0s} = \frac{Z_s - Z_0}{Z_s + Z_0} = -g_{s0},$$

and k is the exctinction coefficient of the material, *e.g.* $k(f_c) = \varepsilon_0 \varepsilon_2(f_c)$.

During the simulation, the time evolution of the diffracted field is recorded at a given probe point. Then, its normalized Fourier transform is calculated and compared to the analytical model. The dispersive behavior of the silver slab is modelized successively by a Drude model, a Drude-Lorentz model, a 2-SOGP model and a 4-SOGP model. The coefficients for these models and several others, calculated by the SA algorithm, can be found in the appendix A. For each of these models, the analytical reflection spectrum is calculated by the classical impedance formulas described above. Therefore, each numerical solution is compared to its exact solution, and its quality assessed through the calculation of the L^1 error over the reflection spectrum. Moreover, several of these results are also compared with the exact solution of the problem obtained using the raw Johnson & Christy data [JC72].

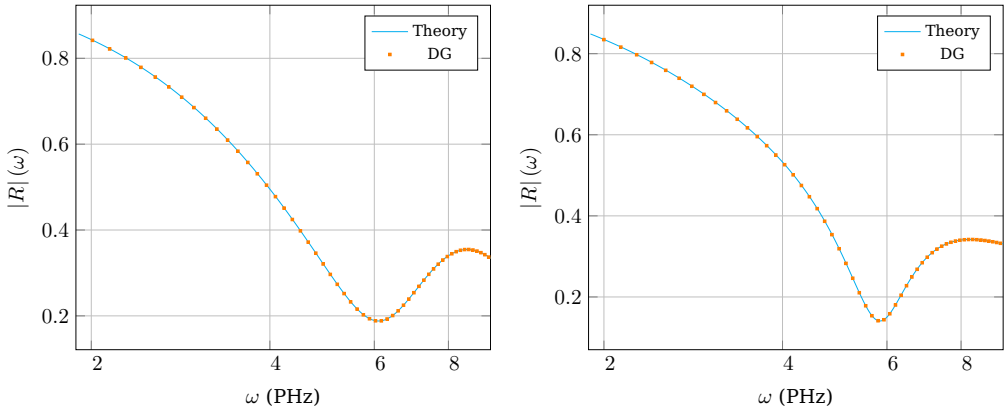
7.2 Results

For all the following results, the used mesh consists of 58,826 vertices and 318,318 tetrahedrons, and a $\mathbb{P}_2$ approximation has been used. We first assess the method by comparing of each model with its own exact solution. The calculated errors can be found in table 8, and plots for the Drude-Lorentz and the 2-SOGP models are presented on figure 12.

Table 8 | L^1 errors, computational times and allocated memory for various dispersive model in the silver slab case.

Dispersion model	L^1 error	$t(s)$	$\frac{t}{t_r}$	$\frac{m}{m_r}$
Vacuum	–	3267	1	1
Drude	0.0804	4300	1.316	1.151
Drude-Lorentz	0.0765	4366	1.336	1.305
2-SOGP	0.0820	4340	1.328	1.305
4-SOGP	0.0941	4571	1.399	1.608

One immediatly notices a very good agreement between the numerical and the exact solution over the whole frequency range, and a steady level of error regarding the number of second-order poles used. The impact of the number of poles on the computational time and on the amount of memory allocated was also assessed, and a visual representation of its influence is presented on figure 13. The time and memory values are normalized respectively by a reference time, t_r, and a reference memory value, m_r, which corresponds to the zero-SOGP case, *e.g.* all the computational domain is made of vacuum. The displayed computational times correspond to a 64 CPU cores parallel case, whereas the allocated memory values were recorded in a sequential case. For the above-described mesh and for a physical calculation time of $2\,10^{-1}s$, one obtains $t_r = 3267s$ and $m_r = 958$ Mo. Several things are to be noticed about this plot : first, the additional computational time when switching from non-dispersive to dispersive behaviour is quite high (approximately 30%). On the contrary, the time cost of each additional pole is almost negligible. This might seem illogical at first sight, but can be easily explained : when a dispersive material is



(a) Drude-Lorentz model (b) 2-SOGP model

Figure 12 | Reflection coefficient of a silver slab described by 12(a) a Drude-Lorentz model and 12(b) a 2-SOGP model.

considered, the fields update requires a loop over all the tetrahedrons of the domain. For each of these, a test must be done to determine wether or not the current tetrahedron is in a dispersive domain, and if yes in which one. Considering the high cost of an `IF` statement done for each tetrahedron at each timestep, the additional cost observed on figure 13 is now understandable. The memory allocation rises regularly with the addition of poles, and the cost of a single SOGP can be evaluated to roughly 15 % of the reference memory occupation m_r. An improved implementation is currently under study, that should particularly reduce the extra computational time required when dispersive materials are considered.

A last comparison is made, showing the main interest of the generalized dispersive model : on figure 14, the plots of silver imaginary part permittivity and the predicted reflectance spectra are presented aside for J&C, 2-SOGP and 4-SOGP data. The positive impact of a good fitting of the material properties is particularly visible here at the resonance frequency, since the relative error on the amplitude of the resonance drops from 90% for 2-SOGP fitting to roughly 33% for 4-SOGP, for an extra computational time of 5% only.

7.3 Conclusion

The interest of using a generalized dispersive model has been shown in the particularly simple situation of a reflective slab. It has been shown that very good approximations of the real dispersive behaviour of metals such as silver can be obtained with a limited number of poles, and the cost of such improvements have been assessed. As an aside, improved implementations of the model might be worth considering in order to reduce the latter. The interest of mixed ZOGP/FOGP/SOGP fitting of permittivity functions remain to be evaluated.

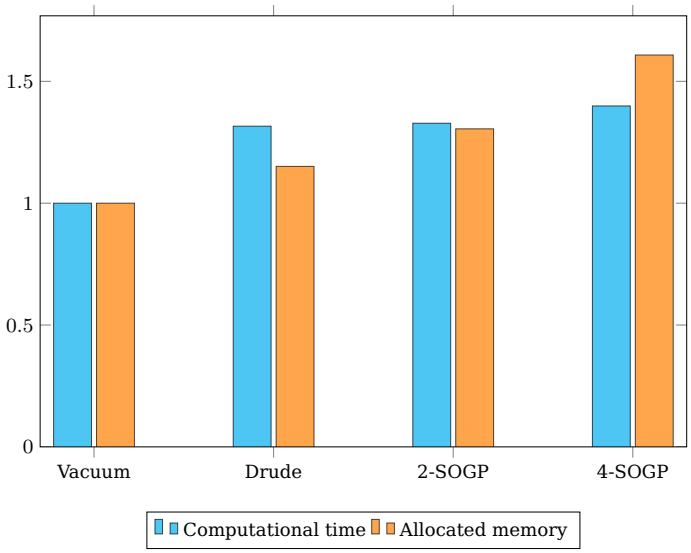
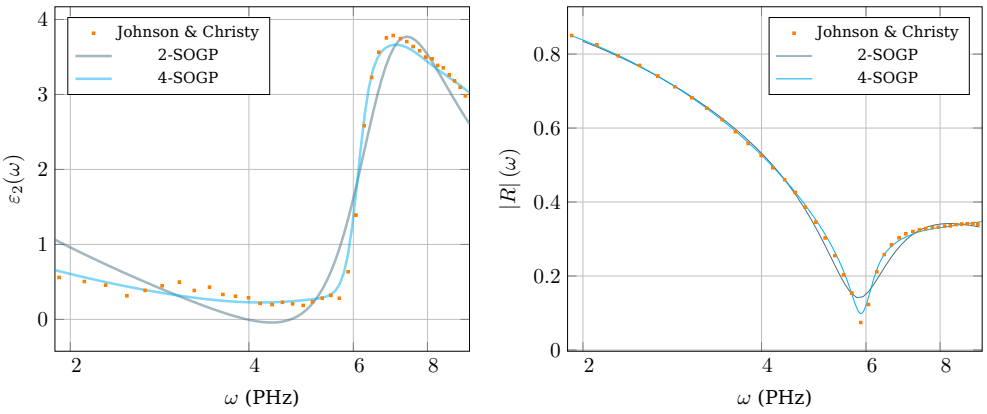


Figure 13 | Computational time and memory allocation for the generalized dispersive formulation for various models.



(a) Permittivity (imaginary part) (b) Calculated reflection spectra

Figure 14 | Comparison of the fitting of imaginary part of the permittivity silver by 2-SOGP and 4-SOGP and its impact on the prediction of the reflectance spectrum.

8 Conclusion

This document presents the theoretical and numerical aspects of the Maxwell equations coupled to (i) a Drude model and (ii) a generalized dispersive model, discretized by a DGTD method. The physical dispersion basics are first presented, followed by a description of the DGTD method and the resulting formulation in cases (i) and (ii). Then, *a priori* stability and convergence results are proved for both models. Eventually, three numerical experiments are led, demonstrating the interest of correctly describing the dispersive behavior of metals in nanoscale devices.

Computing the interaction of electromagnetic waves in the THz range with nanoscale devices have proved to be a greedy kind of problems : considering a 50 nm-sized device and a 600 THz incident plane wave (typical values of these kinds of problems), one readily finds that the classical criterion of meshing with a precision of roughly $h \simeq \frac{\lambda}{10}$ gives $h \simeq 50$ nm, which is the size of the considered scatterer. This draws several conclusions : considering the latter statement, one would be tempted to mesh the computational domain with a high discrepancy in the cell sizes between the diffracting object and the vacuum (or other medium) that surrounds it : this is to be taken with care, for this could be the cause of spurious reflections and of a very small timestep. In many situations, a good way of overcoming such problems is to use a higher geometrical approximation of the considered object : this would allow to loosen the mesh and to put higher orders of approximation to good use. Furthermore, such problems leading to very high computational efforts, high performance computing (HPC) options are to be considered with even more interest. These strategies will be at the center of future efforts.

A Coefficients for the generalized dispersive formulation

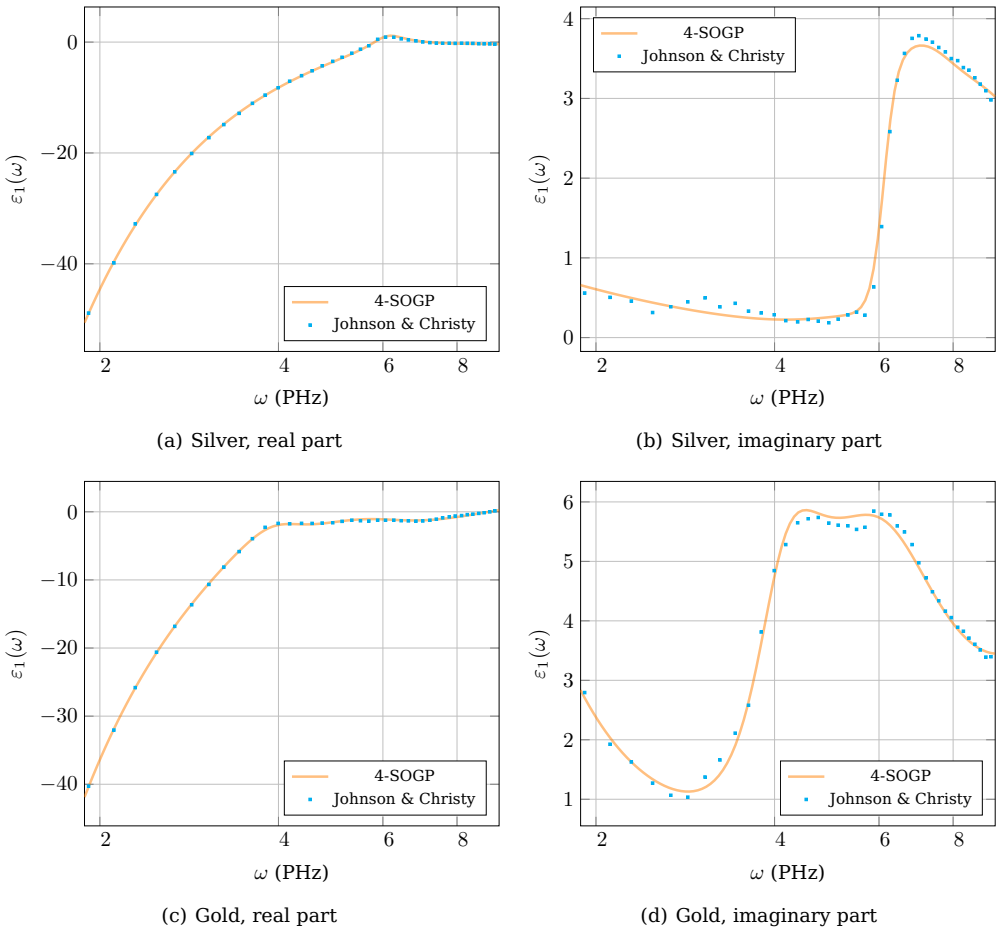
This appendix provides dispersion coefficients for the generalized dispersive formulation for silver and gold over the $[300, 1500]$ THz frequency range. The experimental data, from [JC72], is fitted to different models with a simulated annealing algorithm [KGV83]. For each of the two metals, Drude, Drude-Lorentz, 2-SOGP and 4-SOGP coefficients are given. To ease the reading, special units are used : ω's, γ's, d's and f's are given in PHz, whereas c's and e's are given in PHz^2. Plots of the 4-SOGP permittivities are displayed in figure 15.

Table 9 | Coefficients of various dispersive models for silver.

Parameters	Drude	Drude-Lorentz	2-SOGP	4-SOGP
ε_∞	3.7362	2.7311	1.2944	0.95798
ω_d	13.871	14.084	–	–
γ_d	0.045154	0.0066786	–	–
$\Delta\varepsilon$	–	1.6336	–	–
ω_l	–	8.1286	–	–
γ_l	–	3.6448	–	–
c_1	–	–	189.09	190.69
d_1	–	–	2.6584	1.4784
e_1	–	–	0.0	0.0
f_1	–	–	0.0	0.0
c_2	–	–	56.165	0.020329
d_2	–	–	12.005	2.0383
e_2	–	–	43.932	37.357
f_2	–	–	3.1709	0.96842
c_3	–	–	–	31.345
d_3	–	–	–	11.791
e_3	–	–	–	72.355
f_3	–	–	–	5.0129
c_4	–	–	–	83.642
d_4	–	–	–	0.0
e_4	–	–	–	53.332
f_4	–	–	–	3.8829

Table 10 | Coefficients of various dispersive models for gold.

Parameters	Drude	Drude-Lorentz	2-SOGP	4-SOGP
ε_∞	3.2629	3.6793	0.90746	0.0
ω_d	12.147	13.456	–	–
γ_d	0.24304	0.0	–	–
$\Delta\varepsilon$	–	5.1899	–	–
ω_l	–	6.3681	–	–
γ_l	–	5.7923	–	–
c_1	–	–	160.20	17.832
d_1	–	–	9.3741	4.7977
e_1	–	–	0.0	16.322
f_1	–	–	0.0	1.4611
c_2	–	–	17.949	169.60
d_2	–	–	20.146	0.0
e_2	–	–	15.020	0.025319
f_2	–	–	2.8094	0.11004
c_3	–	–	–	449.06
d_3	–	–	–	0.0
e_3	–	–	–	208.43
f_3	–	–	–	6.2893
c_4	–	–	–	83.521
d_4	–	–	–	4.2579
e_4	–	–	–	37.447
f_4	–	–	–	3.8276



(a) Silver, real part

(b) Silver, imaginary part

(c) Gold, real part

(d) Gold, imaginary part

Figure 15 | Real and imaginary parts of the silver and gold relative permittivity predicted by the 4-SOGP model compared to experimental data from Johnson & Christy.

References

[BFLP06] Marc Bernacki, Loula Fezoui, Stephane Lanteri, and Serge Piperno. Parallel discontinuous Galerkin unstructured mesh solvers for the calculation of three-dimensional wave propagation problems. *Applied Mathematical Modelling*, 30 : 744 – 763, 2006.

[BKN11] Kurt Busch, Michael König, and Jens Niegemann. Discontinuous Galerkin methods in nanophotonics. *Laser Photonics Review*, pages 1 – 37, 2011.

[BS08] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer, 2008.

[CCL11] T. Cabel, J. Charles, and S. Lanteri. Multi-GPU acceleration of a DGTD method for modeling human exposure to electromagnetic waves. Technical report, INRIA Sophia Antipolis-Méditerranée, équipe NACHOS, 2011.

[CFL10] Joseph Charles, Loula Fezoui, and Stephane Lanteri. Étude numérique d'interpolations polynomiales dans une méthode Galerkin discontinue pour la résolution numérique des équations de Maxwell instationnaires 1D. Technical report, INRIA Sophia Antipolis-Méditerranée, équipe NACHOS, 2010.

[Die12] Richard Timothy Helmut Diehl. *Analysis of metallic nanostructures by a discontinuous Galerkin time-domain Maxwell solver on graphics processing units*. PhD thesis, Karlsruher Instituts für Technologie, 2012.

[DLS12] Clement Durochat, Stephane Lanteri, and Claire Scheid. High order non-conforming multi-element discontinuous Galerkin method for time-domain electromagnetics. *Proc. of the 2012 International Conference on Electromagnetics in Advanced Applications (ICEAA '12)*, pages 379 – 382, 2012.

[Dru00] P. Drude. Zur elektronentheorie der metalle. *Annalen der Physik*, 306 : 566 – 613, 1900.

[Fah09] Hassan Fahs. Development of a hp-like discontinuous Galerkin time-domain method on non-conforming simplical meshes for electromagnetic wave propagation. *International Journal of Numerical Analysis and Modeling*, 6 : 193 – 216, 2009.

[FL10] Hassan Fahs and Stephane Lanteri. A high-order non-conforming discontinuous Galerkin method for time-domain electromagnetics. *Journal of Computational and Applied Mathematics*, 234 : 1088 – 1096, 2010.

[FLLP05] Loula Fezoui, Stéphane Lanteri, Stéphanie Lohrengel, and Serge Piperno. Convergence and stability of a discontinuous Galerkin time-domain method for the 3D heterogeneous Maxwell equation on unstructured meshes. *ESAIM : Mathematical Modelling and Numerical Analysis*, 39 : 1149 – 1176, 2005.

[GS99] B. Gustavsen and A. Semlyen. Rational approximation of frequency domain responses by vector fitting. *IEEE Transactions on Power Delivery*, 14 : 1052 – 1061, 1999.

[HDF06] M. Han, R. W. Dutton, and S. Fan. Model dispersive media in FDTD method with complex-conjudate pole-residue pairs. *IEEE Microwave and Wireless Components Letters*, 16 : 119 – 121, 2006.

[HW02] J. S. Hesthaven and T. Warburton. Nodal high-order methods on unstructured grids. *Journal of Computational Physics*, pages 186 – 221, 2002.

[JC72] P. B. Johnson and R. W. Christy. Optical constants of the noble metals. *Physical Review B*, 6 : 4370 – 4379, 1972.

[JCZ07] Xia Ji, Wei Cai, and Pingwen Zhang. High-order DGDT methods for dispersive Maxwell's equations and modelling of silver nanowire coupling. *International Journal for Numerical Methods in Engineering*, 69 : 308 – 325, 2007.

[JJ07] John D. Joannopoulosm and Steven G. Johnson. *Photonic Crystals, Molding the Flow of Light*. Princeton University Press, second edition, 2007.

[KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220 : 671 – 680, 1983.

[LCE08] Jichung Li, Yitung Chen, and Valjean Elander. Mathematical and numerical study of wave propagation in negative-index material. *Computer Methods in Applied Mechanics and Engineering*, 197 : 3976 – 3987, 2008.

[Li09] Jichung Li. Numerical convergence and physical fidelity analysis for maxwell's equations in metamaterials. *Computer Methods in Applied Mechanics and Engineering*, 198 : 3161 – 3172, 2009.

[LS12] Stephane Lanteri and Claire Scheid. Convergence of a discontinuous Galerkin scheme for the mixed time domain Maxwell's equations in dispersive media. *IMA Journal of Numerical Analysis*, 2012.

[LSK93] R. Luebbers, D. Steich, and K. Kunz. FDTD calculation of scattering from frequency-dependent materials. *IEEE Transactions on Antennas and Propagation*, 41 : 1249 – 1257, 1993.

[Moy12] Ludovic Moya. Temporal convergence of a locally implicit discontinuous Galerkin method for Maxwell's equations. *ESAIM : Mathematical Modelling and Numerical Analysis*, pages 1225 – 1246, 2012.

[OO06] M. Okoniewski and E. Okoniewska. Drude dispersion in ADE FDTD revisited. *Electronic Letters*, 42, 2006.

[Pip05] Serge Piperno. Symplectic local time-stepping in non-dissipative DGTD methods applied to wave propagation problems. Technical report, Inria Sophia Antipolis, Project-team Caïman, 2005.

[RH73] W. H. Reed and T. R. Hill. Triangular mesh method for the neutron transport equation. Technical report, Los Alamos National Laboratory, 1973.

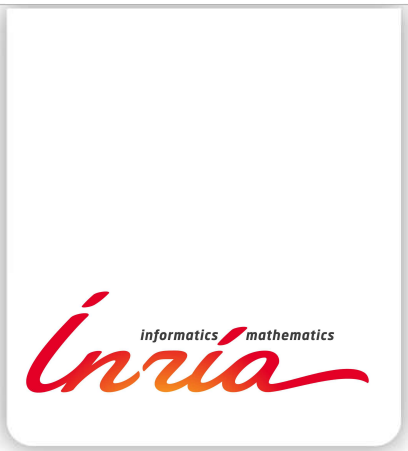
[SKNB09] Kai Stannigel, Michael König, Jens Niegemann, and Kurt Busch. Discontinuous Galerkin time-domain computations of metallic nanostructures. *Optics Express*, 17 : 14934 – 14947, 2009.

[Tei08] F. L. Teixeira. Time-domain finite-difference and finite-element methods for Maxwell equations in complex media. *IEEE Transactions on antennas and propagation*, 56 : 2150 – 2166, 2008.

[VBSH06] Victor Veselago, Leonid Braginsky, Valery Shklover, and Christian Hafner. Negative refractive index materials. *Journal of Computational and Theoretical Nanoscience*, 3 : 1 – 30, 2006.

[vdH81] H.C. van de Hulst. *Light scattering by small particles*. Dover Pulibcations, Inc., 1981.

[VLDC11] A. Vial, T. Laroche, M. Dridi, and L. Le Cunff. A new model of dispersion for metals leading to a more accurate modeling of plasmonic structures using the FDTD method. *Applied Physics A*, 103 : 849 – 853, 2011.



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399