CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Nuclear Sciences and Physical Engineering

Department of Physical Electronics
Specialization: Computational physics

Optimal computation of the source terms for modelling high-order harmonic generation in long media

DIPLOMA THESIS

Author: Bc. Tadeáš Němec
Supervisor: prof. Ing. Jiří Limpouch, CSc.
Co-supervisor: Ing. Jan Vabek, Ph.D.
Year: 2024

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA JADERNÁ A FYZIKÁLNĚ INŽENÝRSKÁ
Katedra fyzikální elektroniky

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: **Bc. Tadeáš Němec**

Studijní program: **Fyzikální elektronika**

Specializace: **Počítačová fyzika**

Akademický rok: **2022/2023**

Název práce: (česky) **Optimalizace výpočtu zdrojových členů pro modelování generace vysokých harmonických v dlouhých prostředích**

Název práce: (anglicky) **Optimal computation of the source terms for modelling high-order harmonic generation in long media**

Jazyk práce: **Angličtina**

Cíl práce:

V této práci se zaměříme na optimalizaci výpočtu zdrojových členů pro rovnice popisující elektromagnetické pole v kontextu generace vysokých harmonických frekvencí v plynech, tj. pole řídícího pulzu a vygenerovaného sekundárního záření v důsledku interakce řídícího pulzu s plynem. Zaměříme se na různé metody popisu zdrojového členu: od obvyklých popisů pomocí nelinearit v nelineární optice po plně numerické řešení pomocí Schrödingerovy rovnice.
Dále je v tomto makroskopickém kontextu potřeba daný zdrojový člen počítat mnohokrát pro „podobné“ vstupy. To je potenciálně slibný výchozí bod pro aplikaci neuronových sítí a strojového učení pro výpočet na základě známé třídy přesně spočítaných řešení. V rámci práce se budeme zabývat i touto možností.

Pokyny pro vypracování

1. Seznamte se s výpočtem zdrojových členů pomocí nelineárního popisu Kerrova jevu a ionizace. Seznamte se s výpočtem odpovídajícího zdrojového členu pomocí numerického řešení Schrödingerovy rovnice.
2. Porovnejte tyto dva přístupy výpočtu z pohledu řešení pro jeden daný vstup.

3. Prozkoumejte, jaký je následný vliv tohoto výsledku se započítáním makroskopických efektů, tj. jak je ovlivněn výsledek šíření pulzu v dlouhém plynném médiu.
4. Prozkoumejte možnosti optimalizace výpočtů zdrojového členu pomocí Schrödingerovy rovnice: i) ve frekvenční doméně řídícího pulzu; ii) pro frekvence odpovídající vysokým harmonickým.
5. Pro vypracování předchozích bodů uvažujte zjednodušenou geometrii (1D-TDSE + 1D šíření na ose), diskutujte možnosti rozšíření zahrnující více dimenzí problému.

Doporučená literatura:

1. Finke O., Vábek J., Nevrkla M. et al. Phase-matched high-order harmonic generation in pre-ionized noble gases. Sci Rep 12, 7715,2022.
2. Bergé L., Skupin S., Nuter R.,Kasparian J., Wolf J.-P., Ultrashort filaments of light in weakly ionized, optically transparent media, Rep. Prog. Phys. 70, 1633, 2007.
3. Lagaris, I. E., Likas, A, Fotiadis, D. I., Artificial neural networks for solving ordinary and partial differential equations, IEEE Transactions on Neural Networks, Vol. 9, Iss. 5, p987-1000,1998.

Jméno a pracoviště vedoucího práce:

prof. Ing. Jiří Limpouch, CSc.
Katedra fyzikální elektroniky, Fakulta jaderná a fyzikálně inženýrská ČVUT v Praze

Jméno a pracoviště konzultanta:

Ing. Jan Vábek
Katedra fyzikální elektroniky, Fakulta jaderná a fyzikálně inženýrská ČVUT v Praze
FzÚ AV ČR, v.v.i., ELI-Beamlines

Datum zadání: 12. říjen 2022
Datum odevzdání: 3. květen 2023

Doba platnosti zadání je dva roky od data zadání.

....
Garant programu

....
Vedoucí katedry *Děkan*

V Praze dne 12.10.2022

Prohlášení

Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně a že jsem uvedl veškeré zdroje a použitou literaturu.

V Praze dne 8. 1. 2024 ....

Bc. Tadeáš Němec

Acknowledgements

I would like to thank prof. Ing. Jiří Limpouch, CSc. for supervising the diploma thesis and helping with practical aspects leading to its successful completion. Great acknowledgment belongs to Ing. Jan Vábek, Ph.D. who has been always available for consulting on problems and who brought me into the physics community that studies higher-order harmonic generation. Without his initiative, I would not have gotten into a friendly group in ELI-Beamlines and on an immensely enriching Erasmus internship in Bordeaux, where I had had the privilege to work with Fabrice Catoire. Fabrice Catoire had been immensely helpful during the design of the multi-scale model and clarified many technical bits. Regular online meetings with him and Jan Vábek had advanced the thesis to a significant extent.

Last I would like to thank my family for supporting me during my studies.

Poděkování

Děkuji prof. Ing. Jiřímu Limpouchovi, CSc. za vedení mé diplomové práce a za pomoc s praktickými kroky vedoucími k jejímu zdárnému dokončení. Velké poděkování patří zejména Ing. Janu Vábkovi, Ph.D., jenž byl vždy k dispozici pro konzultaci v případě nejasností a který mne přivedl do fyzikální komunity zabývající se problematikou generace vysokých harmonických frekvencí. Bez jeho iniciaitivy bych se nedostal do přátelské skupiny v ELI-Beamlines a na nesmírně obohacující pracovní stáž Erasmus do Bordeaux, kde jsem měl to privilegium strávit 3 měsíce pod vedením Fabrice Catoire. Fabrice Catoire byl nesmírně nápomocný při návrhu multiškálového modelu a objasnění fungování jeho dílčích částí. Pravidelné online hovory s ním a s Janem Vábkem posouvaly tuto diplomovou práci kupředu významnou měrou.

Závěrem děkuji mé rodině za zázemí a za podporu při studiu.

Bc. Tadeáš Němec

Thesis title:

Optimal computation of the source terms for modelling high-order harmonic generation in long media

Author: Bc. Tadeáš Němec

Study Programme: Physical electronics
Specialization: Computational physics
Type of Thesis: Diploma thesis

Supervisor: **prof. Ing. Jiří Limpouch, CSc.**
Department of Physical Electronics, Faculty of Nuclear Sciences and Physical Engineering
Consultants: **Ing. Jan Vábek, Ph.D.**,
ELI-ERIC (ELI Beamlines facillity)

Abstract: This thesis covers the theory and the practical implementation of a one-dimensional multi-scale model for laser pulse propagation in gaseous media. The macroscopic pulse propagation, governed by the solution of Maxwell's equations in paraxial approximation, is coupled with a one-dimensional time-dependent Schrödinger equation (1D-TDSE) and briefly contrasted with the nonlinear optics description by propagating the pulse in the medium for a distance up to 2.6 mm at various intensities. For a one-to-one correspondence of the two approaches, the 1st- and 3rd-order susceptibilities are extracted directly using the 1D-TDSE. The main goal of the thesis and the motivation behind the multi-scale model is to use the computed numerical fields from the macroscopic propagation for training and evaluation of a custom-designed artificial neural network (ANN), implemented using the Tensorflow library. The ANN offers a possibility to significantly speed up the computation of the source terms for modeling non-linear optical phenomena including higher-order harmonic generation in long media. The preliminary results of the ANN source term modeling are also discussed in the scope of this thesis.

Key words: Ab-initio simulations, multi-scale model, neural networks, high harmonic generation (HHG) in long media

Název:

Optimalizace výpočtu zdrojových členů pro modelování generace vysokých harmonických v dlouhých prostředích

Autor: Bc. Tadeáš Němec

Abstrakt: Tato práce rozebírá teorii a praktickou implementaci jednodimenzionálního multiškálového modelu šíření laserového pulzu plynným prostředím. Makroskopické šíření pulsu, vycházející z řešení Maxwellových rovnic v paraxiální aproximaci, je provázáno s jednodimenzionální časově závislou Schrödingerovou rovnicí (1D-TDSE) a dáno do kontrastu s popisem z nelineární optiky šířením pulsu v prostoru na vzdálenosti až 2,6 mm při různých intenzitách. Pro přímé srovnání těchto dvou přístupů jsou za pomoci 1D-TDSE extrahovány susceptibility 1. a 3. řádu. Hlavním cílem práce a motivací multiškálového modelu je využití napočítaných numerických polí z makroskopického šíření pro trénování a vyhodnocení vlastního návrhu umělé neuronové sítě (ANN), implementované pomocí knihovny Tensorflow. ANN nabízí možnost významně urychlit výpočet zdrojových členů pro modelování nelineárních optických jevů zahrnujících generaci vysokých harmonických frekvencí v dlouhých prostředích. Prvotní výsledky modelování zdrojových členů pomocí ANN jsou předmětem diskuze v této práci.

Klíčová slova: Ab-initio simulace, multiškálový model, neuronové sítě, generace vysokých harmonických frekvencí (HHG) v dlouhých prostředích

Contents

Nomenclature

Symbols and notations

Symbol	Meaning
$\boldsymbol{a}, \mathbf{A}$	Vector a or A respectively
a_i, A_i	i-th component of vector $\boldsymbol{a}$ or $\mathbf{A}$ respectively
∂_ξ	Partial derivative with respect to variable ξ
∂_i	Partial derivative along the x_i component
∇	Gradient operator
Δ	Laplace operator
i	Imaginary unit
$\mathscr{F}[\varphi(t)](\omega)$	Fourier transform of φ
$\hat{a}$	Fourier transform of a
a^*	Complex conjugate of a

Selected physical constants and quantities[1]

Hartree atomic units (abbreviation a.u.) [15] are conveniently utilized in this thesis, for which the values of m_e, e, $\hbar$ and $4\pi\varepsilon_0$ have been set to 1. We bring a brief overview of selected physical quantities and constants that are used in the following text (in SI units and in a.u.).

Quantity	Value [SI]	Value [a.u.]	Meaning
α	$\frac{1}{137.035999}$	$\frac{1}{137.035999}$	Fine-structure constant
c	$2.99792 \cdot 10^{8}$ m·s^{-1}	$\frac{1}{\alpha}$	Speed of light in vacuum
e	$1.60218 \cdot 10^{-19}$ C	1	Elementary charge
E_h	27.21138 eV	1	Hartree energy
m_e	$9.10938 \cdot 10^{-31}$ kg	1	Electron mass
$\hbar$	$1.05457 \cdot 10^{-34}$ J·s	1	Reduced Planck constant
a_0	$5.29177 \cdot 10^{-11}$ m	1	Bohr radius
ε_0	$8.85419 \cdot 10^{-12}$ F·m^{-1}	$\frac{1}{4\pi}$	Vacuum permittivity
μ_0	$4\pi \cdot 10^{-7}$ H·m^{-1}	$\frac{1}{\varepsilon_0 c^2}$	Vacuum permeability
t	$2.41888 \cdot 10^{-17}$ s	1	Time
E_0	$5.14221 \cdot 10^{11}$ V·m^{-1}	1	Electric field
I_0	$3.50945 \cdot 10^{16}$ W·cm^{-2}	1	Field intensity[2]

---

[1] Reference: [1, 114].

[2] Peak intensity at $E = 1$ a.u. ($I = \frac{1}{2}\varepsilon_0 c E^2$).

Definition of the Fourier transform

The definitions of the Fourier transform vary from author to author. In the scope of this thesis, the following definitions of the Fourier transform will be used[3]

$$\mathscr{F}[\varphi(t)](\omega) = \int_{\mathbb{R}} \mathrm{d}t\, \varphi(t) e^{i\omega t}, \tag{1}$$

$$\mathscr{F}^{-1}[\hat{\varphi}(\omega)](t) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{d}\omega\, \hat{\varphi}(\omega) e^{-i\omega t}. \tag{2}$$

The definitions are in line with the standard electrodynamics nomenclature [2] of the forward propagating harmonic wave. We can write the total real field $F(t)$ as a superposition of elementary harmonic fields as follows:

$$F(t) = \mathrm{Re}\left[\sum_{\omega} A_\omega e^{-i\omega t}\right],$$

where A_ω is generally complex amplitude corresponding to the particular frequency ω.

We conclude this section by enlisting some of the properties of the Fourier transform, according to the definitions (1), (2). Let $\varphi(t), \psi(t)$ have finite support on $\mathbb{R}$ and let all of the following Fourier integrals exist, then:

- Parseval's identity: $\int_{\mathbb{R}} \mathrm{d}t |\varphi(t)|^2 = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{d}\omega |\mathscr{F}[\varphi(t)]|^2$
- Identity relation: $\mathscr{F}^{-1}\left[\mathscr{F}\right] = \mathrm{Id}$.
- Derivative theorem: $\mathscr{F}\left[\frac{\mathrm{d}}{\mathrm{d}t}\varphi(t)\right](\omega) = (-i\omega)\mathscr{F}\left[\varphi(t)\right](\omega)$.
- Shift theorem: $\mathscr{F}\left[\varphi(t-t_0)\right] = e^{i\omega t_0}\mathscr{F}\left[\varphi(t)\right]$.
- Convolution theorem:

$$\begin{aligned} \mathscr{F}\left[\varphi(t)\right] \cdot \mathscr{F}\left[\psi(t)\right] &= \mathscr{F}\left[\varphi(t) * \psi(t)\right], \\ \mathscr{F}\left[\varphi(t) \cdot \psi(t)\right] &= \mathscr{F}\left[\varphi(t)\right] * \mathscr{F}\left[\psi(t)\right]. \end{aligned}$$

---

[3] Remark that this definition is not unitary and therefore does not preserve the norm, see the Parseval's identity.

Introduction

The advent of ultrashort, high-intensity laser pulses has spurred the exploitation of highly nonlinear optical phenomena for generating secondary radiation and the advent of the field of attoscience. The attoscience field opened the door to probing and controlling the electron dynamics in atoms thanks to the generation of attosecond pulses [3], particularly through the process of higher-order harmonic generation (HHG). Highlighting the importance of this field of physics, the 2023 Nobel Prize in physics has been awarded *"for experimental methods that generate attosecond pulses of light for the study of electron dynamics in matter"* [115] to Anne L'Huiller, Ferenc Krausz and Pierre Agostini. The current efforts drift towards the optimal generation of secondary radiation and controlling the frequency gain in gas chambers [16, 17, 18]. The linear and nonlinear effects modulating the propagating beam can have detrimental effects on phase matching in long gaseous media. Moreover, the delicate interplay between Kerr focusing and plasma diffusion gives rise to laser filamentation in the medium. The filamentation of femtosecond pulses is also of particular interest due to beam self-guiding and generation of THz radiation ($10^{11} - 10^{13}$ Hz) offering the study of optically non-transparent samples [19].

It is of great interest to the scientific and engineering community to develop robust simulation tools capable of representing all of the linear and nonlinear optical effects. However, it is difficult to assemble individual contributions into an encompassing model. Under some circumstances, we can develop *ad-hoc* models from certain assumptions and phenomenological observations. For the interaction of lasers with neutral particles, i.e. non-ionized or slightly ionized media[4], we may exploit classical models from linear and nonlinear optics or semi-classical models of ionization to fully describe the propagation of the ultra-short laser pulse in gas. The SFA model [20] has been the workhorse for simulating HHG in gases for arbitrary laser fields [111]. However, by construction, the resonances, Kerr effect and other processes of interest for the propagation are not well described in the SFA. Hence the beam propagation must be treated using a separate model. Thus we have two options for obtaining medium response in two distinct frequency regions – perturbative and HHG.

A more comprehensive perspective is to employ the *ab-initio* approach, i.e. using many time-dependent Schrödinger equations (TDSEs) for the computation of source terms in Maxwell's equations. The solution of the TDSE naturally incorporates all of the linear and nonlinear optical effects that are not easily obtainable with the standard nonlinear optics methods, including HHG. For this reason, the coupling of Maxwell's equations with the exact microscopic description is desired for it could unravel a path towards optimal generation of radiation for widespread usage. However, this approach also comes with several drawbacks. As we will show in this thesis, even the 1-dimensional TDSE (1D-TDSE) is the serious bottleneck in the computation of the source terms. Even the most

---

[4] We are not interested in the laser-plasma interaction in the framework of this thesis.

robust 3-dimensional TDSE solvers coupled with fully vectorial descriptions of fields are limited to distances ranging from a few tens of microns [21] to a fraction of millimeters [22].

The motivation behind optimizing the computation of the source terms for harmonic generation in long media is evident. In this thesis, we will explore the idea of optimization using methods of artificial neural networks. Neural networks have shown rapid development in recent years and have been applied in numerous tasks – from text processing large language models [23, 24, 116] or text-to-image networks [117] to physics applications for the inertial confinement fusion optimization [25], subatomic particle discovery [26], physics informed neural networks [27, 28] or finding solutions to the Schrödinger equation [29, 30, 31]. Thanks to the wide availability of neural network software, e.g. Tensorflow [118] or PyTorch [32], designing a custom network is reduced to assembling off-the-shelf components. The acquired datasets from the macroscopic field propagation will be preprocessed and used for training the custom network. The neural network might be suitable for this task due to higher computational efficiency than the 1D-TDSE and its ability to create a universal function approximator [33], assuming enough training data is generated.

The thesis is organized as follows. Chapter 1 provides a theoretical overview of the microscopic picture including the description of atom-laser field interaction and the consequent secondary radiation induced by the fields in the form of HHG. Chapter 2 revolves around two major topics. First, the solution of Maxwell's equations under the unidirectional pulse propagation approximation, which is used for the propagation of the beam, is discussed. Second, the technique for obtaining 1st and 3rd-order susceptibilities, $\chi^{(1)}$ and $\chi^{(3)}$ respectively, from the time-dependent perturbation theory is theoretically examined and then linked to the numerical *ab-initio* computation using the 1D-TDSE. Chapter 3 dives into the numerical methods of 1D-TDSE and Maxwell solver and the *ab-initio* and *ad-hoc* models are defined in the context of multi-scale model. The results for the beam propagation in Argon obtained using the multi-scale model are presented in Chapter 4. Chapter 5 explains the fundamental principles of neural networks and concludes with an assessment of a custom neural network model called TDSE-NN for optimizing the computation of the source term.

The thesis is complemented by two appendices. One revolves around ionization and gives a brief overview of various models for its description and a method of extracting free electron density from the 1D-TDSE is explained. The second appendix elaborates further on some aspects of NN design, NN data preprocessing and empirical observations that were briefly mentioned in the main text.

Chapter 1

Microscopic picture

To assemble the multi-scale model, we must begin with the family of equations and quantities describing the microscopic frame. A significant portion of this chapter revolves around the time-dependent Schrödinger equation (TDSE) governing the dynamics of a system influenced by an external laser field. At the end of the chapter, we take a brief look at high-order harmonic generation (HHG) in gases.

1.1 Atom-field interaction

This section is devoted to the theoretical introduction to the quantum-mechanical description of single atoms under the influence of an external laser field. First, the many-electron problem is discussed briefly and is followed by single-active electron approximation. We then proceed with different gauge descriptions of Hamiltonians and observables under the influence of the laser field.

1.1.1 Many-electron problem

The Hamiltonian describing a system composed of N electrons under the influence of Coulomb potential of the core reads[1] [4]

$$H(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = -\frac{\hbar^2}{2m_e}\sum_i \Delta_i + \sum_i V_{\text{nucleus-electron}}(\boldsymbol{r}_i) + \frac{1}{2}\sum_i \sum_{i \neq j} V_{\text{electron-electron}}(\boldsymbol{r}_{ij}), \tag{1.1}$$

Where $V_{\text{nucleus-electron}}(\boldsymbol{r}_i)$ and $V_{\text{electron-electron}}(\boldsymbol{r}_{ij})$ correspond to the Coulomb interactions of the nucleus with electrons and inter-electron interactions respectively. The solution of the corresponding time-independent Schrödinger equation (TISE)

$$H(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)\psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = E\psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) \tag{1.2}$$

was originally proposed by Hartree [35] and improved later by Slater [36] and Fock [37]. Hartree-Fock solvers have been used for determining ground states of simple molecular and atomic systems and have been the workhorse in quantum chemistry research [38, 39]. An alternative and popular method to solve the many-electron Schrödinger equation relies on the density functional theory (DFT) and its time-dependent extension (TDDFT) [5].

---

[1] The Hamiltonian concerns only the electron-electron and electron-nuclei interactions. However, under the Born-Oppenheimer approximation, also the nuclei-nuclei interactions should be included [34].

1.1.2 Single-active electron approximation

In the framework of this thesis, we treat the individual atoms as fully uncorrelated, meaning atoms are completely isolated systems, and no phenomenological damping is included, thus the effective potential V_{eff} approaches the plain Coulomb potential. Because the potential varies from atom to atom, we keep the effective potential nomenclature for now, V_{eff} will be specified later, see Subsection 3.1.1. For the simplest symmetrical problems such as the simulation of rare gases, solving the many-electron problem with external fields is complicated and computationally expensive. The problem is addressed by employing a method of single-active electron (SAE) approximation. The method reduces the problem into a single electron interaction with external fields and atomic core. Remaining averaged electron contributions introduce effective shielding of the potential – this is embedded within the effective potential V_{eff}. SAE approximation has its clear advantages in computational efficiency compared to TDDFT and provides satisfactory qualitative predictions but may lack quantitative agreement with an exact many-electron description and the polarization character of electromagnetic radiation may also play a role, see [40]. The corresponding field-free Hamiltonian is expressed as follows:

$$H_0 = \frac{1}{2m_e}\boldsymbol{p}^2 + V_{\text{eff}}, \tag{1.3}$$

where the momentum operator is defined as $\boldsymbol{p} = -i\hbar\nabla$. The field-free Hamiltonian H_0 becomes the foundation for the laser-atom interaction. Fields are implemented in the next subsection.

1.1.3 Hamiltonian of an electron in electromagnetic field

Let $(\mathbf{A}(\boldsymbol{r},t), \phi(\boldsymbol{r},t))$ be a pair of electromagnetic potentials fully describing the electric and magnetic fields

$$\mathbf{E} = -\partial_t\mathbf{A} - \nabla\phi, \tag{1.4}$$

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{1.5}$$

The exact form of potentials $(\mathbf{A}, \phi)$ in electromagnetism relies on the selected gauge, depending on the application and convenience. Given the scalar function $\Lambda(\boldsymbol{r},t)$, the potentials are transformed according to gauge Λ as

$$\mathbf{A}^{\Lambda} = \mathbf{A} - \nabla\Lambda, \tag{1.6}$$

$$\phi^{\Lambda} = \phi + \partial_t\Lambda. \tag{1.7}$$

Fields $\mathbf{E}$ and $\mathbf{B}$ given by Eqs. (1.4), (1.5) remain the same regardless of gauge[2].

The Hamiltonian of an electron in an effective Coulomb potential of atom V_{eff} under the influence of the external fields reads

$$H = \frac{1}{2m_e}\left(\boldsymbol{p} + e\mathbf{A}\right)^2 + V_{\text{eff}} - e\phi. \tag{1.8}$$

The dynamic of the system with the varying fields is then governed by the TDSE

$$i\hbar\partial_t\psi = H\psi \tag{1.9}$$

---

[2] Enforcing some approximations within the gauges can cancel the validity of Maxwell's equations, as discussed later in the subsection for the length gauge and dipole approximation, but still provide satisfactory results.

with an initial condition $\psi(t=0)=\psi_0$ where ψ_0 is the ground state given by the TISE $H_0\psi_0 = E_0\psi_0$.

We can apply the Lorenz gauge condition

$$\nabla \cdot \mathbf{A}^V = 0 \quad , \quad \phi^V = 0, \tag{1.10}$$

which yields the *velocity gauge Hamiltonian*

$$H^V = \frac{1}{2m_e}\left(\boldsymbol{p} + e\mathbf{A}^V\right)^2 + V_{\text{eff}}. \tag{1.11}$$

The nomenclature stems from the description of the field interaction using purely vector potential $\mathbf{A}$. The recipe to transform the TDSE (1.9) with Hamiltonian (1.8) into arbitrary gauge Λ is performed via the following unitary operator

$$T_\Lambda = \exp\left(\frac{ie\Lambda}{\hbar}\right), \tag{1.12}$$

where e is the charge of the electron. We then introduce

$$\psi^\Lambda = T_\Lambda \psi, \tag{1.13}$$

and plug ψ into the TDSE (1.9) to obtain the transformed TDSE with respect to a chosen gauge Λ [6]:

$$i\hbar\partial_t\psi^\Lambda = \left(\frac{1}{2m_e}\left(\boldsymbol{p} + e\mathbf{A}^\Lambda\right)^2 + V_{\text{eff}} - e\phi^\Lambda\right)\psi^\Lambda. \tag{1.14}$$

Different gauge descriptions of Hamiltonians (1.14) can be useful for particular implementations of TDSE. In the scope of the thesis, we will be using the *length gauge Hamiltonian* that couples the field interaction via electric field $\mathbf{E}$ directly. We first define the gauge function

$$\Lambda^L = \boldsymbol{r} \cdot \mathbf{A}^V. \tag{1.15}$$

Then we employ the *dipole approximation* $\mathbf{A}^V \approx \mathbf{A}^V(t)$. The transformed potentials, according to Eqs. (1.6), (1.7), now read

$$\mathbf{A}^L = 0, \tag{1.16}$$

$$\phi^L = \boldsymbol{r} \cdot \partial_t \mathbf{A}^V = -\boldsymbol{r} \cdot \mathbf{E}. \tag{1.17}$$

By substituting the potentials into (1.14), we get the length gauge Hamiltonian:

$$H^L = \frac{1}{2m_e}\boldsymbol{p}^2 + V_{\text{eff}} + e\boldsymbol{r} \cdot \mathbf{E}. \tag{1.18}$$

We can conveniently express the Hamiltonian in a separate form

$$H = H_0 + V_I(t), \tag{1.19}$$

where $V_I(t)$ is a time-dependent perturbation of the Hamiltonian by the external fields.

Let us inspect the validity of the length gauge under dipole approximation. One of the direct consequences is the complete omission of the magnetic interaction. However, the magnetic field term can be fully neglected for nonrelativistic laser pulses if we do not account for the spin-magnetic interaction. The dipole approximation also requires sufficiently large wavelengths compared to the overall size of the atom a_0, i.e. $\lambda \gg a_0$. For more in-depth reading on non-physical aspects of length gauge within the dipole approximation see [41].

The length gauge is also used in the SFA and agrees well with experiments[3] [42]. Both gauges yield the same results in the case of the 1D and 3D-TDSE, however, in the 3D-TDSE, they differ in computational efficiency for varying intensities [43]. It can be shown that for certain quantities, e.g. photoelectron spectrum, the convergence of the velocity gauge is much faster in contrast with the length gauge because the latter requires a much larger numerical grid due to long-traveling electrons [44, 45]. In the end, the choice of a suitable gauge often relies on a particular numerical implementation of the TDSE and the desired input field. Since the propagated quantity in the multi-scale model is the electric field, we stick to the length gauge 1D-TDSE.

1.1.4 Gauge invariant position and momentum

The physical observables are gauge-independent quantities. Thus, the expectation values of the operators in length and velocity gauges must be the same. The correspondence will be demonstrated on position and momentum operators.

The gauge invariant momentum is defined as [6]

$$\mathbf{\Pi}^{\Lambda} = \boldsymbol{p} + e\mathbf{A}^{\Lambda}, \tag{1.20}$$

where $\boldsymbol{p}$ is the canonical momentum from the corresponding Hamiltonian, expressed as operator $-i\hbar\nabla$. This yields $\mathbf{\Pi}^V = \boldsymbol{p} + e\mathbf{A}^V$ in the velocity gauge and $\mathbf{\Pi}^L = \boldsymbol{p}$ in the length gauge, since $\mathbf{A}^L = 0$. If we use the respective wavefunctions for the computation of the expectation values, we arrive at the same expectation values, as demonstrated in the following series of equalities:

$$\langle \mathbf{\Pi}^L \rangle_L = \langle \psi^L | \boldsymbol{p} | \psi^L \rangle = \langle \psi^V T_\Lambda^\dagger | \boldsymbol{p} | T_\Lambda \psi^V \rangle = \langle \psi^V | \boldsymbol{p} + e\mathbf{A}^V | \psi^V \rangle = \langle \mathbf{\Pi}^V \rangle_V \,, \tag{1.21}$$

where the momentum operator was applied on $|T_\Lambda \psi^V\rangle$ as follows:

$$\boldsymbol{p} \, |T_\Lambda \psi^V\rangle = -i\hbar\nabla \left(\exp\left(\frac{ie\boldsymbol{r}\cdot\mathbf{A}^V}{\hbar} \right) |\psi^V\rangle \right) = T_\Lambda(\boldsymbol{p} + e\mathbf{A}) \, |\psi^V\rangle \,. \tag{1.22}$$

Analogically for the position operator $\boldsymbol{x}$ in the length and velocity gauge, it can be shown that

$$\langle \boldsymbol{x}^L \rangle_L = \langle \boldsymbol{x} \rangle_L = \langle \boldsymbol{x} \rangle_V = \langle \boldsymbol{x}^V \rangle_V \,. \tag{1.23}$$

Expressions (1.21) and (1.23) are the essential gauge-invariant quantities for the proper physical description of the source terms.

1.2 Microscopic fields

The electrons start to wiggle in the vicinity of the time-varying external fields. The induced averaged electron currents $\langle \boldsymbol{j} \rangle$ act as source term for the wave equation

$$\Delta\mathbf{E} - \frac{1}{c^2}\partial_t^2\mathbf{E} = \frac{1}{\varepsilon_0 c^2}\partial_t \langle \boldsymbol{j} \rangle \,. \tag{1.24}$$

---

[3] There is no satisfactory fit for any field intensity for SFA in the velocity gauge [42].

Focusing only on the generation of a local microscopic field by currents $\langle \boldsymbol{j} \rangle$, we may neglect the Laplacian in (1.24) and write

$$\partial_t^2 \mathbf{E}_{\text{micro}} = -\frac{1}{\varepsilon_0} \partial_t \langle \boldsymbol{j} \rangle . \tag{1.25}$$

The gauge invariant electron current operator is defined in terms of the momentum $\mathbf{\Pi}^\Lambda$ as [6]:

$$\boldsymbol{j}^\Lambda(\boldsymbol{r}, t) = -\frac{e}{m_e} \mathbf{\Pi}^\Lambda(\boldsymbol{r}, t) = -\frac{e}{m_e} \left(\boldsymbol{p} + e\mathbf{A}^\Lambda \right) , \tag{1.26}$$

or in terms of the position operator:

$$\langle \boldsymbol{j}^\Lambda \rangle = -e \partial_t \langle \boldsymbol{x}^\Lambda \rangle \tag{1.27}$$

To calculate the microscopic field $\mathbf{E}_{\text{micro}}$, one must solve Equation (1.25), or solve the following algebraic equation in the frequency domain:

$$\omega^2 \hat{\mathbf{E}}_{\text{micro}} = -\frac{i\omega}{\varepsilon_0} \langle \hat{\boldsymbol{j}}^\Lambda \rangle . \tag{1.28}$$

Computing directly $\langle \boldsymbol{j}^\Lambda \rangle$ requires numerical evaluation of the spatial derivative of the wavefunction. From the numerical point of view, this approach may inherently introduce further errors.

A more suitable approach is to combine classical Hamilton equations with the Ehrenfest theorem and find the corresponding terms. Let us assume the length gauge Hamiltonian (Eq. (1.18)), the classical Hamilton equations yield:

$$\dot{\boldsymbol{r}} = \nabla_{\boldsymbol{p}} H^L = \frac{1}{m_e} \boldsymbol{p} \tag{1.29}$$

$$\dot{\boldsymbol{p}} = -\nabla_{\boldsymbol{r}} H^L = -\nabla V_{\text{eff}} - e\mathbf{E}. \tag{1.30}$$

The Ehrenfest theorem states that the Hamilton equations above also hold for quantum operators assuming the averaged observables $\boldsymbol{r}$ and $\boldsymbol{p}$. In total, we have the following equalities:

$$\langle \ddot{\boldsymbol{r}} \rangle = \frac{1}{m_e} \partial_t \langle \boldsymbol{p} \rangle_L = \frac{1}{m_e} \partial_t \langle \boldsymbol{p} + e\mathbf{A}^V \rangle_V = \frac{1}{m_e} \langle -\nabla V_{\text{eff}} - e\mathbf{E} \rangle_L = -\frac{1}{e} \partial_t \langle \boldsymbol{j} \rangle . \tag{1.31}$$

We see that we may express Equation (1.28) in 3 equivalent forms:

- **Length form**: $\hat{\mathbf{E}}_{\text{micro}} = \frac{e}{\varepsilon_0} \langle \hat{\boldsymbol{r}} \rangle$.
- **Velocity form**: $-i\omega \hat{\mathbf{E}}_{\text{micro}} = \frac{e}{m_e \varepsilon_0} \langle \hat{\boldsymbol{p}} \rangle_L = \frac{e}{m_e \varepsilon_0} \langle \widehat{\boldsymbol{p} + e\mathbf{A}} \rangle_V$.
- **Acceleration form**[4]: $\omega^2 \hat{\mathbf{E}}_{\text{micro}} = \frac{e}{\varepsilon_0} \langle \widehat{\nabla V_{\text{eff}}} \rangle_V = \frac{e}{\varepsilon_0} \langle \widehat{\nabla V_{\text{eff}} + e\mathbf{E}} \rangle_L$.

Even though the 3 equivalent forms theoretically reach the same microscopic field, some forms are preferred over others due to simpler evaluation of the expectation value. For instance, as we will see in the chapter devoted to the numerical aspects of TDSE, the acceleration form and the length form have a simple analytical evaluation without any approximation involved, whereas in the case of the velocity form we need to evaluate the derivative numerically.

Provided sufficiently strong field intensities, the computation of microscopic fields leads to higher-order harmonic generation. The next section gives a brief overview of the process.

---

[4] The equality $\langle \widehat{\nabla V_{\text{eff}}} \rangle_V = \langle \widehat{\nabla V_{\text{eff}} + e\mathbf{E}} \rangle_L$ follows directly from the Hamilton equations for the velocity gauge Hamiltonian (1.11) and comparing the respective acceleration from the length gauge Hamiltonian.

1.3 Higher-order harmonic generation

The process of higher-order harmonic generation (HHG) offers a broadband, coherent, tabletop, flexible source of extreme UV (XUV) radiation [46]. The radiation can be obtained through intense IR laser beam interaction with gaseous targets[5] [3], typically rare or inert gases. The intensity is in the non-relativistic window $I \sim 10^{14}$ W/cm^2 but significant ionization occurs. The non-linear response of the media results in the generation of a very broad spectrum of higher-order harmonic frequencies of the incident beam [48, 49]. Due to the periodicity of the laser field and high spectral bandwidth, a train of attosecond impulses is created, offering many intriguing applications in probing dynamics of valence electron shells in gases and molecules, in pump-probe experiments or condensed matter spectroscopy [50, 51, 52]. The high-quality harmonic beam may also provide interesting seeding applications for other sources of radiation such as free-electron lasers (XFELs) to further improve the amplification and beam convergence of XFELs [53]. Due to low conversion efficiency, many works have been devoted to the study of HHG amplification, which is generally limited by photon reabsorption and phase matching in media [54, 55, 56]. Some promising amplification techniques involve the use of two-color laser fields (e.g. fundamental and second harmonic fields) for optimum phase matching and increasing overall harmonic yield [16, 17], or the recent usage of a capillary discharge to phase-match the harmonic generation [18].

1.3.1 Three-step model

The process of HHG in gas can be intuitively understood from a so-called *three-step model* [57] which is sketched in Figure 1.1. In the first step, the atomic Coulomb barrier is partly suppressed by the external periodic electric fields, raising the probability of the electron tunneling through the barrier. Once freed, the electron is then accelerated by the field away from the parent ion. The electron gains additional momentum up to the point it comes back to the vicinity of the remaining ion and recombine, emitting the energy excess in the form of an XUV photon. The maximum energy of the photon in the case of a linearly polarised driving field is given by the following formula [20]:

$$\hbar\omega_{\max} = I_p + 3.17U_p \quad , \quad U_p = \frac{e^2 E_L^2}{4m_e\omega_L^2}, \tag{1.32}$$

where I_p is the ionisation potential of the atom and U_p ponderomotive potential given by the laser field intensity E_L and frequency ω_L.

Many models explaining the nature of HHG and predicting the higher-order harmonic spectra have been developed over the years, ranging from the simplest models combining Newtonian mechanics with the semi-classical models of ionization [57] to more complex ones relying on the semiclassical approach or the full quantum description using the TDSE. The problem with the TDSE is that we cannot find a closed-form solution without approximations and, in the case of 3D, non-linearly polarized fields, the exact numerical solution is very expensive. The approximations, as explained in the next subsection, allow us to compute the harmonic spectra more efficiently. We briefly introduce the semiclassical Lewenstein model [20] that qualitatively describes the harmonic generation.

---

[5] HHG can also be generated in solids or using relativistic plasma mirrors [47].

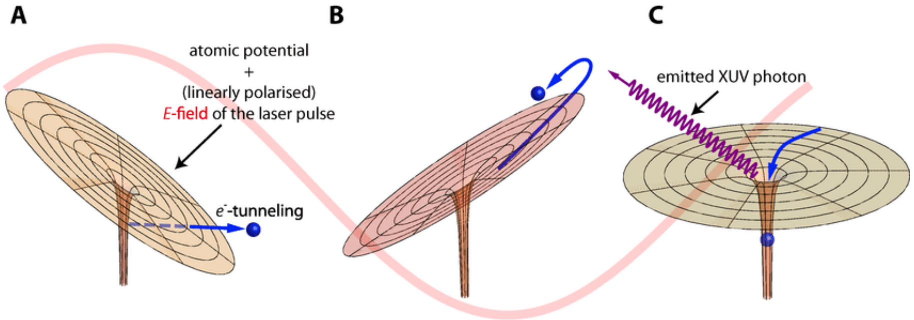

Figure 1.1: Sketch of the three-step model. First the electron tunnels through the potential barrier (A), next it accelerates due to the outer field (B) and finally recollides while the generation of the XUV photon occurs (C). The image is overlayed by the driving electric field (pale red sinusoidal curve). Available from: [58].

1.3.2 Lewenstein model – SFA

The Lewenstein model is based on several key assumptions:

- Single active electron and dipole approximation is employed.
- We assume length gauge Hamiltonian (1.18) with Coulomb potential:

$$H^L = \frac{1}{2m_e}\boldsymbol{p}^2 - \frac{e^2}{4\pi\varepsilon_0 r} + e\boldsymbol{r}\cdot\mathbf{E}. \tag{1.33}$$

- Tunnelling ionization is the dominant photoionization process.
- We fully neglect all bound states except the ground state $|g\rangle$ and the continuum states are not influenced by the Coulomb forces of the ion – this assumption is valid for sufficiently strong fields, hence the term *strong field approximation* (SFA).

The aforementioned assumptions condense into the following *ansatz* for the wavefunction $|\psi(t)\rangle$ [20]

$$|\psi(t)\rangle = e^{iI_pt/\hbar}\left(|g\rangle + \int_{\mathbb{R}^3}\mathrm{d}\boldsymbol{k}\, b(\boldsymbol{k},t)\,|\boldsymbol{k}\rangle\right), \tag{1.34}$$

where $e^{-iI_pt/\hbar}$ corresponds to the evolution operator of the field-free Hamiltonian, $b(\boldsymbol{k},t)$ are the expansion coefficients of the continuum states[6] $|\boldsymbol{k}\rangle$ and $\mathbf{E} = -\partial_t\mathbf{A}$, according to the length gauge.

The goal is to derive the dipole $\mathbf{D}(t) = \langle\psi|\boldsymbol{x}|\psi\rangle$ acting as a source of the microscopic field $\mathbf{E}_{\text{micro}}$. We plug the *ansatz* (1.34) into the Hamiltonian (1.33) and solve the corresponding TDSE. The time evolution is summarized by defining the unitary operator $U(t,t')$ acting on the ground state:

$$|\psi(t)\rangle = U(t,\,-\infty)\,|g\rangle\,. \tag{1.35}$$

We find $b(\boldsymbol{p},t)$ as

$$b(\boldsymbol{p},t) = e^{-iI_pt/\hbar}\,\langle\boldsymbol{k}|U(t,\,-\infty)|g\rangle\,, \tag{1.36}$$

---

[6] The continuum states are effectively described using plane waves in SFA.

to obtain the following well-known dipole formula [20]:

$$\mathbf{D}(t) = -i\int_0^t \mathrm{d}t'\, \underbrace{\boldsymbol{d}(\boldsymbol{k}_0 + \mathbf{A}(t'))\cdot \mathbf{E}(t')}_{\text{Ionization}} \underbrace{\left(\frac{-2\pi i}{t-t'}\right)^{\frac{3}{2}} e^{-iS(\boldsymbol{k}_0,\,t,\,t')}}_{\text{Propagation}} \underbrace{\boldsymbol{d}^*(\boldsymbol{k}_0 + \mathbf{A}(t))}_{\text{Recombination}} + c.c.. \tag{1.37}$$

Here $\boldsymbol{d}(\boldsymbol{p}) = \langle \boldsymbol{p}|\boldsymbol{x}|g\rangle$ is the so-called transition dipole matrix element describing the bounded electron state, $S(\boldsymbol{p}, t, t') = \int_{t'}^{t} \mathrm{d}t''(\boldsymbol{p} + \mathbf{A})^2/2m_e + I_p(t - t')/\hbar$ is the classical action of the electron in an electric field and $\boldsymbol{k}_0$ is a stationary momentum given by the saddle point condition $\nabla S(\boldsymbol{k}_0, t, t') = 0$. If we shine the light on the formula (1.37), the terms provide the quantum analog to the classical three-step model presented in the previous section, as the explanatory notes proposed.

1.3.3 TDSE perspective on HHG

We compare the HHG power spectrum obtained from the SFA model, Eq. (1.37), with the exact ab-initio calculations from the 1D-TDSE solver in the length gauge, see Chapter 3. The driving pulse is a 15-cycle (40fs, FWHM in field) $\cos^2$-envelope field of peak intensity $I = 10^{14}$ W/cm^2 with fundamental wavelength of $\lambda_L = 800$ nm. The target is the hydrogen atom for simplicity. The result power spectra for harmonic orders H of the fundamental field frequency ω_L are plotted in Figure 1.2.

We notice 3 distinct regions within the power spectra. The perturbation region (H < 9) is characterized by an exponential decrease of odd harmonics that could be obtained through perturbation expansion of polarization (Chapter 2). The 9-th harmonic corresponds to the ionization potential energy of atomic hydrogen ($I_p = 13.6$ eV). Next comes the plateau region with harmonics of comparable magnitude that spans from H $= 9 - 21$. The spectrum ends with an exponential decrease of the generated harmonic amplitudes after the cutoff harmonic H $= 21$. The cutoff threshold fits with the classical model prediction in Eq. (1.32). Due to the symmetric target and single color field, the odd harmonics are more pronounced than even harmonics in the harmonic spectrum[7]. While the 1D-TDSE spectrum comprises the three main regions, it differs dramatically from the SFA spectrum in the perturbation region. The Lewenstein model, which describes the wavefunction solution of the TDSE as only composed of the gound state and the continnum states, fails at describing the harmonic spectrum where excited states are involved. The ground state in SFA is typically only described using a symmetric 1s hydrogen orbital with respective ionization potential for the particular gas whereas the 1D-TDSE solution contains the excited states[8]. One then needs to rely on the *ab-initio* description for the lower-order harmonics or use a proper *ad-hoc* model. The wave packet spread in continuum is also partly included in the SFA through the term $(-2\pi i/(t - t'))^{\frac{3}{2}}$ but the TDSE wave packet broadening inherently includes the vicinity of the potential. The main advantage of the SFA is that it can solve a variety of problems involving complex 3D fields. For symmetrical targets, we can plug in an arbitrary polarization of fields and obtain the solution in reasonable time frame compared to the 3D-TDSE.

The 1D-TDSE solution in Fig. 1.2 runs faster than the particular implementation of SFA [111]. For this reason, the SFA model is left for comparison with 1D-TDSE.

---

[7] The length of the pulse plays a role in the broadening of the odd harmonics as well as the interferences of harmonics generated between multiple cycles, see discussion and figure in [112], Subsection 1.1.2.

[8] See excited states in Figure A.2.

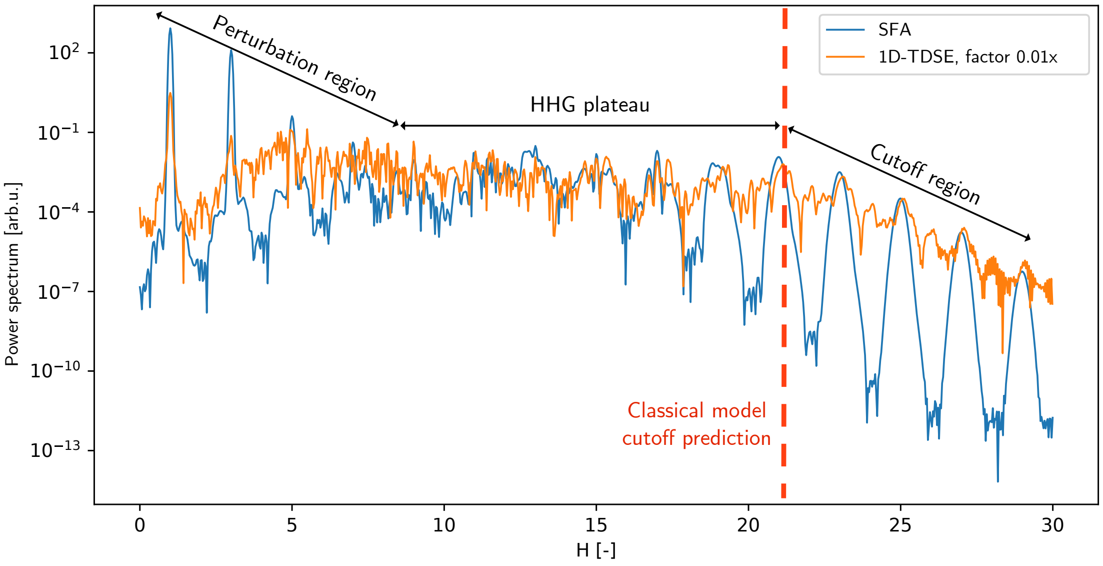

Figure 1.2: Logarithmic power harmonic spectra ($|E_{\mathrm{micro}}(\omega)|^2$) of the microscopic response obtained from the SFA (blue) and from the 1D-TDSE (orange), multiplied by factor 0.01x for correspondence. The SFA data was computed using the SFA solver solving the integral in Equation (1.37) [111] and the same field was applied on the 1D-TDSE solver. The cutoff threshold fits with the classical model prediction in Eq. (1.32) (red vertical dashed line).

Chapter 2

Macroscopic picture

The temporal and spatial profile of high-power laser beams is affected by linear and nonlinear phenomena occurring during the beam-media interaction over a given propagation distance. In this chapter, a one-dimensional, unidirectional pulse propagation equation (UPPE) in the co-moving reference frame describing the evolution of ultrashort laser pulses in transparent media is derived from Maxwell's equations. We introduce the corresponding terms responsible for the pulse modulation into the UPPE and establish the *ad-hoc* and *ab-initio* models. To fulfill the requirement for the qualitative and quantitative comparison of *ad-hoc* and *ab-initio* models, we will discuss the inference of susceptibilities $\chi^{(1)}$ and $\chi^{(3)}$ using the 1D-TDSE.

From this chapter onward, we will rely heavily on the notation to discern temporal and spectral variables that are coupled via Fourier transform, e.g. $\hat{X}(\omega) \overset{\mathscr{F}}{\longleftrightarrow} X(t)$. For more details, see the Nomenclature.

2.1 Propagation equation

We first recall Maxwell's equations with sources for electric and magnetic fields $\mathbf{E}(\boldsymbol{r}, t), \mathbf{B}(\boldsymbol{r}, t)$:

$$\nabla \cdot \mathbf{E} = \frac{\varrho}{\varepsilon_0}, \tag{2.1}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{2.2}$$

$$\nabla \times \mathbf{B} = \varepsilon_0 \mu_0 \partial_t \mathbf{E} + \mu_0 \boldsymbol{j}, \tag{2.3}$$

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B}, \tag{2.4}$$

where $\varrho(\boldsymbol{r}, t), \boldsymbol{j}(\boldsymbol{r}, t)$ are the charge and current densities respectively. Bound and free charges and currents are embedded within ϱ and $\boldsymbol{j}$. Applying the operator $\nabla \times$ on Faraday law (2.4) and using the well known identity $\nabla \times \nabla \equiv \nabla(\nabla \cdot) - \Delta$ in combination with Ampere law (2.3) yields the following wave equation for $\mathbf{E}$ with charges and currents entering as source terms:

$$\Delta \mathbf{E} - \frac{1}{c^2} \partial_t^2 \mathbf{E} = \frac{1}{\varepsilon_0} \nabla \varrho + \mu_0 \partial_t \boldsymbol{j}. \tag{2.5}$$

We assume the volume is filled with neutral particles and no free charges are present. Thus we can write $\nabla \varrho \equiv 0$. The contribution to the propagating field $\mathbf{E}$ comes mainly from the macroscopic time-varying electron current and charge densities. We include them in an encompassing term $\mathbf{F}(\boldsymbol{r},t)$ and substitute to the RHS of the wave equation (2.5) to obtain:

$$\Delta\mathbf{E} - \frac{1}{c^2}\partial_t^2\mathbf{E} = \frac{1}{\varepsilon_0 c^2}\partial_t^2\mathbf{F}. \tag{2.6}$$

The exact implementation of the source term $\mathbf{F}$ varies for the *ad-hoc* and *ab-initio* models. In the case of the *ad-hoc* model, the response of the medium is encoded within the material constants, e.g. susceptibilities $\chi^{(n)}$, and the material response functions – polarization $\mathbf{P}(\mathbf{E})$ or free electron currents $\boldsymbol{j}(\mathbf{E})$. In isotropic media, the polarization can be expanded in the frequency domain as the following series[1]

$$\hat{\mathbf{P}}(\omega,\boldsymbol{r}) = \varepsilon_0\left(\chi^{(1)}\hat{\mathbf{E}} + \chi^{(3)}\widehat{\mathbf{E}^3} + \dots\right)(\omega,\boldsymbol{r}) = \hat{\mathbf{P}}^{\mathrm{L}} + \hat{\mathbf{P}}^{\mathrm{NL}}, \tag{2.7}$$

Remark that due to the choice of a centro-symmetric potential, the expansion (2.7) includes only odd orders. In the scope of the thesis, the *ad-hoc* model is defined by setting[2]

$$\hat{\mathbf{F}}_{\text{ad-hoc}} := \hat{\mathbf{P}}^{\mathrm{L}} + \hat{\mathbf{P}}^{\mathrm{NL}} = \varepsilon_0\left(\chi^{(1)}\hat{\mathbf{E}} + \chi^{(3)}\widehat{\mathbf{E}^3}\right). \tag{2.8}$$

The *ab-initio* response is given by a complementary solver computing the source terms using the TDSE. The averaged microscopic field response $\mathbf{E}_{\text{micro}}$, see the different forms in Section 1.2, is computed at every macroscopic point of the medium. The link between the microscopic quantity ($\mathbf{E}_{\text{micro}}$) and the source term ($\mathbf{F}_{\text{ab-initio}}$) is provided by the formula

$$\mathbf{F}_{\text{ab-initio}} := \mathcal{N}\cdot\varepsilon_0\mathbf{E}_{\text{micro}}, \tag{2.9}$$

where $\mathcal{N}$ is the number gas density.

The following derivation of the propagation equation is then common for both *ad-hoc* and *ab-initio* models. Applying Fourier transform on (2.6) yields the Helmholtz equation for the individual frequency components ω:

$$\Delta\hat{\mathbf{E}} + k_0^2\hat{\mathbf{E}} = -\frac{\omega^2}{\varepsilon_0 c^2}\hat{\mathbf{F}}, \tag{2.10}$$

where the wave number $k_0 = \omega/c$. Remark that a non-relativistic beam is assumed and any atomic or spin magnetism is neglected, hence the Helmholtz equation for the magnetic field can be omitted.

There are many ways how to numerically integrate Maxwell's equations and wave equation (2.5). From Fourier methods to explicit finite difference methods in temporal domain [119] or finite-difference time-domain method on the staggered grid known as the Yee scheme [60]. Since we are only concerned with the forward propagation of the laser pulse in 1D, we can derive a simpler, sufficiently accurate scheme under several approximations. The three-dimensional Helmholtz equation (2.10) will be the starting point for the derivation of the propagation equation.

---

[1] The 3rd order expansion of $\hat{\mathbf{P}}$ implicitly contains third harmonic generation and Kerr effect through the process of three-wave mixing. See details of implementation in Subsection 3.2.2 in the part devoted to *ad-hoc* model.

[2] P^{NL} may include other nonlinear effects such as Raman response but is omitted here. For details see [59], Section 2.2.

2.1.1 Unidirectional pulse propagation equation (UPPE)

Instead of directly integrating the Helmholtz equation (2.10), we further simplify the dimensionality of the problem to 1D by neglecting the transverse effects. This approach will be sufficient for the studies presented in the thesis and it has also been tested for HHG and compared with experiments to reasonable accuracy [54]. The motivation for choosing the simple model is to test the concept of optimizing the computation of TDSE source terms using neural networks. The cases when we can not neglect transverse effects are discussed thoroughly in Section 4.4.

Furthermore, this simplification can be better physically justified in certain cases. For example for hollow-core fibers, we can assume the solution of Helmholtz equation (2.10) in a sperated form [7]

$$\hat{\mathbf{E}}(\boldsymbol{r},\omega) = \boldsymbol{x} S(x,y)\hat{A}(z,\omega)e^{i\beta_0 z}, \tag{2.11}$$

where z is the direction of propagation, $\boldsymbol{x}$ is the field polarization, $S(x,y)$ describes the spatial distribution of the fibre modes, $\hat{A}$ denotes the slowly varying amplitude of the propagating mode and β_0 is the wave number corresponding to the mode. Separation of variables (Eq. (2.11)) leads to equations for $S(x,y)$ and $\hat{A}(z,\omega)$ [7]:

$$\partial_x^2 S + \partial_y^2 S + [\varepsilon(\omega)k_0^2 - \tilde{\beta}]S = 0, \tag{2.12}$$

$$2i\beta_0\partial_z\hat{A} + (\tilde{\beta}^2 - \beta_0^2)\hat{A} = 0. \tag{2.13}$$

The wave number $\tilde{\beta}$ satisfies the eigenvalue equation (2.12) for the fiber modes, for details see [7], and $\varepsilon(\omega)$ is the dielectric constant. If we assume only short propagation distances in the order of millimetres, we can completely disregard mode coupling via the wave number $\tilde{\beta}$. Thus in the first approximation, it is sufficient to compute only the evolution of forward propagating amplitude in z.

Therefore we neglect the transverse component[3] within the Helmholtz equation (2.10):

$$\Delta \equiv \partial_z^2 \quad , \quad \Delta_\perp \equiv 0. \tag{2.14}$$

Using only linearly polarized fields, we can continue with the scalar description of the electric field $\hat{E}(z,\omega)$. With the inclusion of the simplifying conditions, the following equation directly emerges from (2.10):

$$\partial_z^2\hat{E} + k_0^2\hat{E} = -\frac{\omega^2}{\varepsilon_0 c^2}\hat{F}. \tag{2.15}$$

Let us define an operator

$$D^\pm(\omega) = \partial_z \mp ik_0(\omega) \tag{2.16}$$

satisfying the relation

$$D^-(\omega)D^+(\omega) = \partial_z^2 + k_0^2(\omega) \tag{2.17}$$

and assume a general solution of (2.15) as a superposition of forward ($\hat{E}^+$) and backward ($\hat{E}^-$) moving waves, with their corresponding amplitudes $\hat{U}^+$ and $\hat{U}^+$ respectively,

$$\hat{E} = \hat{U}^+ e^{ik_0 z} + \hat{U}^- e^{-ik_0 z}. \tag{2.18}$$

We can neglect the backscattering component of the field $\hat{U}^-$ if the paraxiality condition $|\partial_z^2\hat{U}^+| \ll |k_0(\omega)\partial_z\hat{U}^+|$ applies for the forward propagating envelope $\hat{U}^+$. The paraxial

---

[3] For the details of the transverse Laplacian implementation in free space refer to [59].

approximation holds if $\hat{U}^+$ varies slowly with the propagation distance on scale of one wavelength of the laser beam. Having $|\hat{U}^+| \gg |\hat{U}^-|$, the backscattering operator applied on $\hat{E}$ then yields $D^-(\omega)\hat{E} \approx 2ik_0\hat{E}$ [59]. Helmholtz equation (2.15) simplifies into

$$2ik_0 D^+ \hat{E} = -\frac{\omega^2}{\varepsilon_0 c^2}\hat{F} \tag{2.19}$$

After rearranging the terms, the equation finally reduces into the **unidirectional pulse propagation equation** (UPPE):

$$\partial_z \hat{E} = ik_0\hat{E} + i\frac{\omega}{2\varepsilon_0 c}\hat{F} \tag{2.20}$$

The UPPE (2.20) is the main driving equation of the beam evolution in gas for the *ad-hoc* and *ab-initio* models. However, Eq. (2.20) contains a quickly oscillating term $ik_0\hat{E}$ which is unsuitable for the numerical modeling. We can get rid of it by transforming it into a co-moving frame as demonstrated in the next subsection.

2.1.2 UPPE in co-moving frame

The UPPE, Eq. (2.20), is transformed into a co-moving frame advancing at the phase velocity[4] $v_p = \frac{c}{n_0}$, where n_0 is a refractive index at a fundamental laser wavelength λ_0. The propagating pulse then remains anchored around the center of the computational window, removing the necessity of modifying the numerical grid along the propagation of the solution. We define the retarded time $\tau = t - z/v_p$, referencing the time within the pulse, and the transform of the retarded field then reads:

$$E(z,t) \to E(z,\tau) = E(z, t - z/v_p) \tag{2.21}$$

The derivatives in the co-moving frame are transformed as follows:

$$\partial_z \to \partial_z + \frac{\partial\tau}{\partial z}\frac{\partial}{\partial\tau} = \partial_z - \frac{1}{v_p}\partial_\tau, \tag{2.22}$$

$$\partial_t \to \partial_\tau. \tag{2.23}$$

In the spectral domain, the temporal derivative transforms as $\partial_\tau \to -i\omega$. Applying the transform (2.21) on the UPPE (2.20) yields:

$$\partial_z \hat{E} = i\left(\frac{\omega}{c} - \frac{n_0\omega}{c}\right)\hat{E} + i\frac{\omega}{2\varepsilon_0 c}\hat{F} \tag{2.24}$$

---

[4] The reason why we opt for phase velocity instead of group velocity v_g is both practical and numerical. The group velocity is computed as follows:

$$\frac{1}{v_g} = \left.\frac{\mathrm{d}k}{\mathrm{d}\omega}\right|_{\omega_0} = \frac{1}{v_p} + \frac{\omega_0}{c}\frac{\mathrm{d}n(\omega_0)}{\mathrm{d}\omega} = \frac{1}{v_p} + \Delta\frac{1}{v}.$$

Hence the group velocity can be computed as

$$v_g = \frac{1}{\frac{1}{v_p} + \Delta\frac{1}{v}},$$

however the effect of the correction term $\Delta\frac{1}{v}$ is small compared to $\frac{1}{v_p}$ and secondly we would need to evaluate a derivative of refractive index that can be obtained with certain error from numerical evaluation of $\chi^{(1)}$, see Section 2.2.3. For these reasons, we stick to a frame moving at v_p.

which can be simplified assuming[5]

$$n_0 = \sqrt{1 + \chi_0^{(1)}} \approx 1 + \frac{1}{2}\chi_0^{(1)}. \tag{2.25}$$

We finally define the **UPPE in co-moving frame** as:

$$\partial_z \hat{E} = i\frac{\omega}{2\varepsilon_0 c}\left(\hat{F} - \varepsilon_0 \chi_0^{(1)} \hat{E}\right). \tag{2.26}$$

We should address the fact that we do not a-priori know v_p, respectively $\chi_0^{(1)}$ during the propagation of the pulse. However, the $\chi_0^{(1)}$ coefficient can be extracted directly from the TSDE at a low-intensity field ahead of the propagation itself, as will be the subject of the next section.

Let us discuss why the choice of UPPE is much more advantageous compared to precise Maxwell's equations solvers. Maxwell codes are necessary primarily in cases when the backscattering becomes significant, i.e. out of the paraxial approximation applicability range. Algorithms based on the Yee scheme [60] for Maxwell propagation coupling the TDSE have been developed [61, 21]. However, these schemes are limited to smaller propagation steps and require large computational resources due to the high sampling of the TDSE response. Whereas with UPPE with dominant z-propagation (optically thin gas, capillaries, fibers, etc.), we can use much larger propagation steps with comparable accuracy at a fraction of the cost [59, 62, 63]. All in all, the UPPE in the co-moving frame, Eq. (2.26), will be the workhorse in our simulations as it is suitable for describing high-power femtosecond pulse propagation in gas if we assume a transverse filament size larger than the fundamental wavelength [62].

2.2 Determination of first and third order susceptibilities

To provide a quantitative and qualitative comparison of the *ad-hoc* and *ab-initio* models, it is necessary to correctly determine the values of susceptibilities. Susceptibility $\chi(\omega)$ systematically couples linear dispersion, linear and nonlinear refractive indices (Kerr effect), third harmonic generation (3HG) and other higher order effects. The susceptibility breaks down into linear and non-linear parts corresponding to different terms of perturbation expansion[6]

$$\chi(\omega) = \chi^{\mathrm{L}}(\omega) + \chi^{\mathrm{NL}}(\omega) = \chi^{(1)}(\omega) + \chi^{(3)}(\omega)\hat{\mathbf{E}}^2 + \chi^{(5)}(\omega)\hat{\mathbf{E}}^4 + \dots. \tag{2.27}$$

The subject of determining susceptibilities and related quantities such as non-linear refractive index is a research task highly extending the scope of this thesis. One aspect is the transition between the perturbative and non-perturbative descriptions of non-linear optics. At intensities ranging from $10^{13} - 10^{14}$ W/cm^2 we observe intense Kerr-induced nonlinearities as well as higher-order Kerr effects (HOKE) that affect filamentation of laser beams. In this range the non-perturbative HHG is present and the ionization effects also become significant, which renders the perturbative approach inapplicable. From the ab-initio calculations, we can estimate the boundary intensity at which the standard perturbative approach is applicable. For hydrogen with $I_p = 13.6$ eV interacting with 800 nm

---

[5] Approximation is valid for $\chi_0^{(1)} \ll 1$.

[6] Assuming the central symmetry, the even orders vanish.

field, the intensity threshold is $2 \cdot 10^{13}$ W/cm^2 [64]. Above this threshold, we have to rely on more complicated HOKE descriptions which are still not well understood to this day. The contribution of the HOKE terms and the saturation of the Kerr non-linearity is discussed thoroughly in [65].

In this section, we will refer to the derivation of $\chi^{(1)}$ using the non-stationary perturbation theory and we provide the link with the *ab-initio* approach. We will employ the 1D-TDSE to determine the values of $\chi^{(1)}$ and $\chi^{(3)}$ numerically. To extract the susceptibilities we follow the methods proposed in [65, 66, 67].

2.2.1 Density matrix formalism

We will loosely follow the standard nonlinear optics textbook [8] derivation of $\chi(\omega)$. Instead of using the quantum mechanical formalism presented in Chapter 1, we will employ the alternative density matrix formulation of quantum mechanics to derive the susceptibilities. The density matrix formalism is particularly advantageous for the description of mixed quantum states (e.g. atomic vapors) and phenomenological damping. The presence of damping is reflected in the resonant behavior of susceptibility.

Let $|u_s\rangle$ be a set of eigenstates in an unperturbed system described by Hamiltonian H_0. For each state s applies

$$H_0 |u_s\rangle = E_s |u_s\rangle . \tag{2.28}$$

Next the temporal perturbation $V(t) = -\boldsymbol{\mu}\cdot\mathbf{E}(t)$ is introduced where $\boldsymbol{\mu}$ is a dipole operator defined via charge q and position operator as $\boldsymbol{\mu} = q\boldsymbol{x}$. The wavefunction of the state s in time t may be written as $|\psi_s\rangle = C_s(t) |u_s\rangle$. The temporal evolution of $|\psi_s(t)\rangle$ is governed by the TDSE with the perturbation $V(t)$:

$$i\hbar\partial_t |\psi_s(t)\rangle = (H_0 + V(t)) |\psi_s(t)\rangle . \tag{2.29}$$

The whole ensemble of states can be described using a density matrix

$$\varrho(t) = \sum_s w(s) |\psi_s(t)\rangle \langle\psi_s(t)| , \tag{2.30}$$

mn-th element

$$\varrho_{mn}(t) = \langle u_m|\varrho(t)|u_n\rangle = \sum_s w(s) C_m^{s,*}(t) C_n^s(t), \tag{2.31}$$

where $w(s)$ is a weight, or a probability of a given state[7]. The weights highlight the probabilistic nature of the density matrix.

The dynamics of the density matrix is governed by quantum Liouville equation

$$\partial_t \varrho_{mn}(t) = -\frac{i}{\hbar}[H(t), \varrho(t)]_{mn}. \tag{2.32}$$

From the quantum theory of damping, we assume an isolated system that exchanges energy with a reservoir until it reaches an equilibrium state $\varrho_{mn}^{\mathrm{EQ}}$. Since the interaction results in non-trivial correlations of the wavefunctions that are difficult to analyze analytically, the standard procedure is to average the damping effect by introducing a phenomenological

---

[7] Since w is a probabilistic distribution, $w(s) \in [0; 1]$ and $\sum_s w(s) = 1$.

damping coefficient γ_{mn}, thus allowing to study of each state n independently. We update the Liouville equation into

$$\partial_t \varrho_{mn}(t) = -\frac{i}{\hbar}\left[H(t), \varrho(t)\right]_{mn} - \gamma_{mn}\left(\varrho_{mn}(t) - \varrho_{mn}^{\mathrm{EQ}}\right). \tag{2.33}$$

To solve (2.33) we need to employ the apparatus of time-dependent perturbation theory. Without dwelling in to the detailed description of the method, we refer to Boyd [8], Sections 3.3, 3.4.

2.2.2 Density matrix derivation of linear susceptibility $\chi^{(1)}$

The link between the macroscopic susceptibility $\chi^{(1)}$ and microscopic response of the dipole operator $\langle \boldsymbol{\mu} \rangle$ is via the polarization

$$\hat{\mathbf{P}}(\omega) = \mathcal{N}\left\langle \hat{\boldsymbol{\mu}}(\omega) \right\rangle \approx \varepsilon_0 \chi^{(1)}(\omega)\hat{\mathbf{E}}, \tag{2.34}$$

where $\mathcal{N}$ is the number density.

The expectation value of the dipole operator $\boldsymbol{\mu}$ is obtained in the first-order approximation as the following trace

$$\langle \boldsymbol{\mu} \rangle = \mathrm{Tr}\left(\varrho^{(1)}(t)\boldsymbol{\mu}\right) = \sum_{mn} \varrho^{(1)}_{mn}(t)\boldsymbol{\mu}_{mn}, \tag{2.35}$$

where $\boldsymbol{\mu}_{mn} = \langle u_m| \boldsymbol{\mu} |u_n\rangle$. The explicit form of the first-order expansion term $\varrho^{(1)}_{mn}(t)$, employing the Fourier expansion of $\mathbf{E}(t)$, reads [8]

$$\varrho^{(1)}_{mn}(t) = \frac{\varrho^{(0)}_{nn} - \varrho^{(0)}_{mm}}{\hbar} \sum_{\omega} \frac{\boldsymbol{\mu}_{mn} \cdot \mathbf{E}(\omega) e^{i\omega t}}{(\omega_{mn} + \omega) + i\gamma_{mn}}, \tag{2.36}$$

where $\omega_{mn} = (E_m - E_n)/\hbar$ is the Bohr transition frequency.

The first order susceptibility $\chi^{(1)}$ is then inferred from the Relation (2.34), assuming the initial population is fully in the ground state, as [8]

$$\chi^{(1)}(\omega) = \frac{\mathcal{N}}{3\varepsilon_0\hbar} \sum_m |\boldsymbol{\mu}_{mg}|^2 \omega_{mg} \frac{2}{\omega_{mg}^2 - \omega^2 - 2i\gamma_{mg}\omega}. \tag{2.37}$$

The Equation (2.37) is usually simplified by defining the quantity called *oscillator strength*[8] f_{mg}

$$f_{mg} = \frac{2m_e\omega_{mg}|\boldsymbol{\mu}_{mg}|^2}{3\hbar e^2} \tag{2.38}$$

and it assigns a probabilistic weight to a particular transition from state m to the ground state g. The oscillator strengths are tabulated and, for example of argon, we can refer to [9]. Incorporating oscillator strengths, this enables to rewrite $\chi^{(1)}$ in its most compact form as follows:

$$\chi^{(1)}(\omega) = \omega_p^2 \sum_m f_{mg} \frac{1}{\omega_{mg}^2 - \omega^2 - 2i\gamma_{mg}\omega}, \tag{2.39}$$

where $\omega_p^2 = \frac{\mathcal{N}e^2}{\varepsilon_0 m_e}$ is the plasma frequency. If we knew precisely the oscillator strengths and the phenomenological damping coefficients of the system, the Expression (2.39) would analytically describe the susceptibility in a wide frequency range – including the resonances and their damping through the factor $2i\gamma_{mn}\omega$ in the denominator. However, we will further neglect its contribution due to the constraints given by the 1D-TDSE, as will be discussed in the following subsection.

---

[8] It satisfies the following relations: $\sum_m f_{mg} = 1$, $0 \leq f_{mg} \leq 1$.

2.2.3 Computation of $\chi^{(1)}$ from the 1D-TDSE

As we saw in the previous section, to derive $\chi^{(1)}$ we either have to determine the oscillator strengths and lifetimes of the corresponding transitions, eq. (2.39), or use directly the expectation value of the dipole. We opt for the latter since the expectation value of the dipole operator $\langle\mu\rangle$ is easily obtained from the numerical 1D-TDSE. The details of computing $\langle\mu\rangle$ are presented in Section 3.1.4, Equation (3.26). We will further refer to this approach as the **dipole model**.

We elaborate briefly on the range of applicability of this approach. Firstly, we recall that the implementation of the numerical 1D-TDSE concerns only the mono-atomic problem in the SAE approximation frame. This immediately excludes any lifetimes of the states due to the lack of any cross-correlations between the particles. As a consequence, the dipole model is unable to correctly describe the resonant behavior of the susceptibilities. We infer the range of applicability experimentally by computing the dipoles using the 1D-TDSE for various frequencies of the electric field and analyzing the dispersion curves.

Taking the aforementioned assumptions into account, using $\gamma_{mg} \approx 0$ and $\omega \ll \omega_{mg}$, i.e. the range far from the resonances, we simplify the Expression (2.39) accordingly

$$\chi^{(1)}(\omega) \approx \sum_m \frac{f_{mg}\omega_p^2}{\omega_{mg}^2} \frac{1}{1-\omega^2/\omega_{mg}^2}. \tag{2.40}$$

We apply the Taylor expansion of the expression $1/(1-x)$ on (2.40) and we obtain:

$$\chi^{(1)}(\omega) \approx \sum_m \frac{f_{mg}\omega_p^2}{\omega_{mg}^2} \left(1 + \frac{\omega^2}{\omega_{mg}^2} + \frac{\omega^4}{\omega_{mg}^4} + \dots\right) =: A + B\omega^2 + C\omega^4 + \dots, \tag{2.41}$$

where the expansion coefficients A, B, C encapsulate the oscillator strengths and transition frequencies. Formula (2.41) resembles closely the well-known empirical Cauchy's transmission equation[9]. This expression is particularly useful for determining an interpolation polynomial and for checking the validity of the numerical data.

In the simplest 1D case, we compute numerically $\langle\mu\rangle$ for a linearly polarised field E with a central frequency ω_0 from the 1D-TDSE and we derive $\chi^{(1)}(\omega_0)$ from formula (2.34) as follows:

$$\chi^{(1)}(\omega_0) = \frac{\mathcal{N}\langle\mu(\omega_0)\rangle}{\varepsilon_0 E(\omega_0)}. \tag{2.42}$$

Remark that formula (2.42) is generally complex: $\chi^{(1)}(\omega) = \mathrm{Re}[\chi^{(1)}] + i\mathrm{Im}[\chi^{(1)}]$. The real part relates to the linear refractive index while the imaginary part relates to the medium absorption.

We will proceed with the analysis of the numerical data obtained through expression (2.42) for different cases. Because susceptibility is a density-dependent quantity, we will analyze density-independent quantity $\chi(\omega) = \frac{\langle\mu(\omega)\rangle}{\varepsilon_0 E(\omega)}$ instead[10].

---

[9] Cauchy's transmission equation relates the refractive index n with reciprocal even powers of wavelength λ and the general form is

$$n(\lambda) = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4} + \dots.$$

It is extended by the Sellmeier equation [68].

[10] Consequently the unit of χ is mol^{-1}.

Dispersion of $\chi^{(1)}$ for varying pulse length

We start with the low-intensity regime ($I = 10^{10}$ W/cm^2) with 5 and 10-cycle pulses (FWHM in intensity) and plot real and imaginary parts of $\chi^{(1)}$ for the harmonic range $\mathrm{H} \in (0.2, 10)$ of frequency ω_0 corresponding to $\lambda_0 = 800$ nm. Each simulation took a pulse E with a central frequency ω from the harmonic range and the quantity $\langle \mu(\omega) \rangle$ was obtained from the 1D-TDSE. The execution time of 1D-TDSE was in order of 10^2 seconds and we started with the equidistant spacing of 200 fundamental frequency samples for both pulse lengths. The results are plotted in Figure 2.1.

The two cases differ significantly in the region of the first resonance[11] at harmonic order $\mathrm{H} = 5.7$ ($\mathrm{Im}[\chi^{(1)}] \neq 0$), see Figure 2.1. We notice an increase of amplitude for the longer pulse. This outcome is expected given the fact that the 1D-TDSE does not take into account the lifetimes of states. With the increasing pulse lengths, the amplitude in the resonance would effectively diverge due to the absence of spontaneous damping. We also notice a second strong resonance for $\mathrm{H} = 8.6$ between the ground and 4th excited state.

We can conclude from Figure 2.1 that the applicability region of the chi formula (2.42) should lie below $\mathrm{H} = 4$ where none of the resonances in $\chi^{(1)}$ occur.

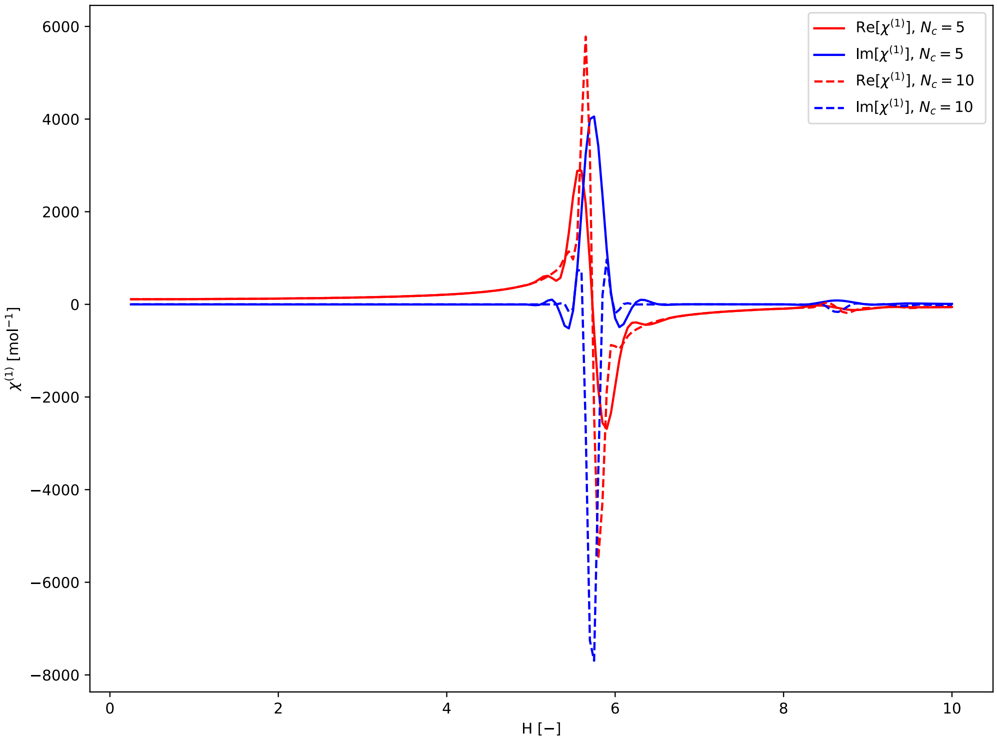

Figure 2.1: Dispersion of real and imaginary parts of $\chi^{(1)}$ for harmonic orders H of ω_0 ($\lambda_0 = 800$ nm) in range $(0.2, 10)\omega_0$, 5 cycle pulse (full line) and 10 cycle pulse (dashed line).

---

[11] This is the resonance between the ground state and the first excited state. See Figure A.2 in Appendix A with the wavefunctions and their corresponding energies.

Kramers-Kronig relations

In this context, it is interesting to briefly discuss the validity of the Kramers-Kronig relations. These relations, fundamental to linear optics, state that for a real, analytical quantity $\chi^{(1)}$ there is a direct relation between the real and imaginary part of its Fourier transform. We should be able to extract $\mathrm{Re}[\chi^{(1)}(\omega)]$ from $\mathrm{Im}[\chi^{(1)}(\omega)]$ and vice versa by evaluating the following principal value integrals [8]:

$$\mathrm{Re}[\chi^{(1)}(\omega)] = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} \mathrm{d}\omega' \frac{\mathrm{Im}[\chi^{(1)}(\omega')]}{\omega' - \omega} = -H\left(\mathrm{Im}[\chi^{(1)}]\right), \tag{2.43}$$

$$\mathrm{Im}[\chi^{(1)}(\omega)] = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} \mathrm{d}\omega' \frac{\mathrm{Re}[\chi^{(1)}(\omega')]}{\omega' - \omega} = H\left(\mathrm{Re}[\chi^{(1)}]\right), \tag{2.44}$$

where

$$H\left(u(\omega)\right) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} \mathrm{d}\omega' \frac{u(\omega')}{\omega - \omega'} \tag{2.45}$$

is an integral transform called *Hilbert transform.*

We employed the Python library SciPy [69] and its implementation of Hilbert transform to check the validity of the Kramers-Kronig relations. The results for a 5-cycle pulse are depicted in Figure 2.2. The trends of $\mathrm{Re}[\chi^{(1)}]$ and $(-1)H\left(\mathrm{Im}[\chi^{(1)}]\right)$, respectively $\mathrm{Im}[\chi^{(1)}]$ and $H\left(\mathrm{Re}[\chi^{(1)}]\right)$ agree reasonably well in the resonance region $\mathrm{H} \in (5, 7)$. The discrepancy between the two is attributed to the evaluation of the Hilbert transform only on a non-negative, finite interval of frequencies. However, the rough demonstration of the Kramers-Kronig relations indicates the validity of the proposed approach of $\chi^{(1)}$ computation.

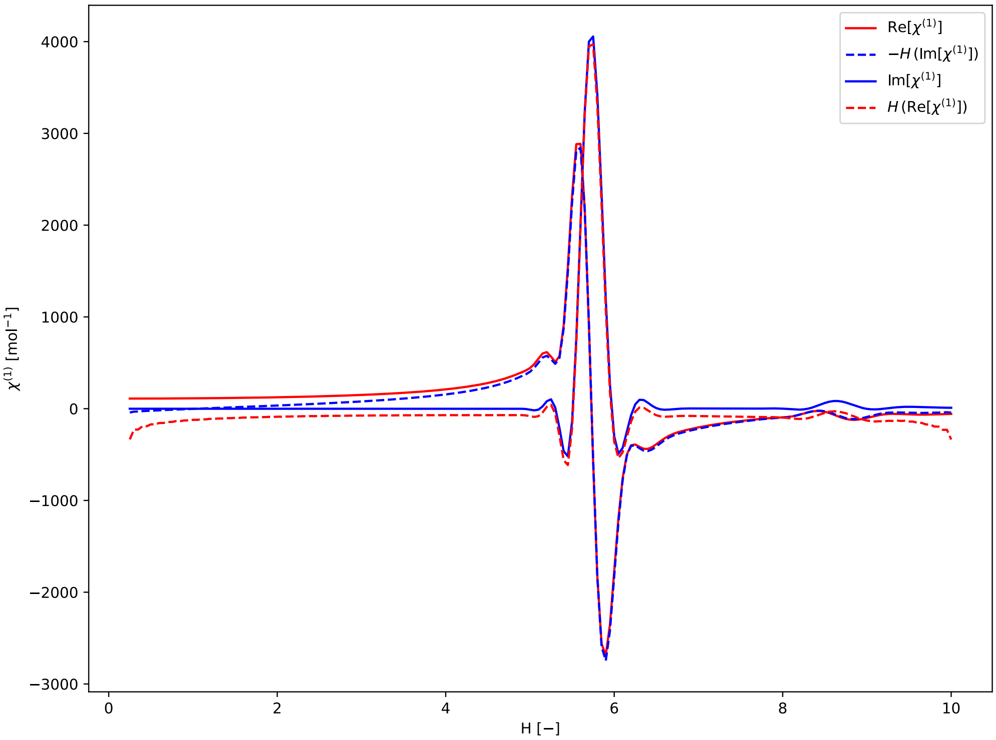

Figure 2.2: Validity of the Kramers-Kronig relations. On display the $\chi^{(1)}$ dispersion, corresponding to a 5 cycle pulse, for harmonic orders of ω_0 ($\lambda_0 = 800$ nm) in range $(0.2, 10)$.

Polynomial fit of $\chi^{(1)}$

We will now focus on a polynomial interpolation of $\chi^{(1)}$ from the numerical 1D-TDSE data. The number of cycles N_c was initially set to 5 cycles FWHM at low intensity $I = 10^{10}$ W/cm^2. The central frequency of the pulse ranged between $(0.5\,, 4.5\,)\omega_0$ and we performed computation for 50 frequency samples given by the Chebyshev distribution for optimal interpolation[12]. Next, we interpolated the data points by a 6th-degree polynomial using the least squares method and plotted the result for range[13] $(0, 3.5\,)\omega_0$ in Figure 2.3. We restricted ourselves for an interval below $3.5\,\omega_0$ since we neglected any non-linear effects above the third harmonic harmonic generation.

The result of the polynomial fit for $N_c = 5$, rounded to the first two significant decimal points, is as follows:

$$\chi^{(1)}(\omega) = 110.45 + 0.067\left(\frac{\omega}{\omega_0}\right) + 3.34\left(\frac{\omega}{\omega_0}\right)^2 - 0.039\left(\frac{\omega}{\omega_0}\right)^3 + \\ + 0.055\left(\frac{\omega}{\omega_0}\right)^4 + 0.0039\left(\frac{\omega}{\omega_0}\right)^5 + 0.011\left(\frac{\omega}{\omega_0}\right)^6 \qquad (2.46)$$

with coefficient of determination $\mathrm{R}^2 \doteq 0.9909$. Coefficients were normalized to the frequency ω_0 corresponding to $\lambda_0 = 800$ nm. While the expansion (2.46) contains also odd frequency powers, in contrast with the expected theoretical expansion (2.41), plotting only the even powers of the fit (2.46) overlays the full fit well.

The question remains how the fit behaves with increasing number of cycles N_c at FWHM, i.e. pulse length, of the driving field. We assume the same frequency sampling as for the 5-cycle pulse and simulate for $N_c = 10$ and $N_c = 20$. In Figure 2.3 we notice a downward shift of the $\chi^{(1)}$ fit, however the fits for $N_c = 10$ (red curve) and $N_c = 20$ (violet dashed curve) sit almost on each other with almost the same $\mathrm{R}^2 \doteq 0.9984$. It is expected[14] that with increasing number of cycles the $\chi^{(1)}$ fit coefficients reach the value independent of pulse length. The remaining fitting error might be caused by insufficient sampling of the fundamental frequency, but for $N_c < 10$ the coefficients are still pulse length-dependent.

Finally, we compare the numerical results obtained from the 1D-TDSE with the analytical fit of $\chi^{(1)}$ for argon at room pressure and temperature by Dalgarno et al. [70]. The model is based on the non-resonant expansion of the equation for $\chi^{(1)}$ with oscillator strengths, Eq. (2.39). It combines the expansion with constants retrieved from the experimental measurements of argon. Because the model employs full 3D geometry, it is not directly comparable to our usage of 1D-TDSE[15], but it is left here for reference. Overall, we see that the 1D-TDSE results have stronger dispersion compared to the Dalgarno model. If we plug the susceptibilities into the formula for refractive index $n \approx 1 + \mathcal{N}\chi^{(1)}/2$ assuming atomic density at room temperature and pressure $\mathcal{N} \doteq 3.6 \cdot 10^{-6}$ a.u., for argon we obtain at $\lambda = 800$ nm the index $n \doteq 1.000204$ which is not far from the experimental value $n = 1.000265$ [71] and the Dalgarno fit retrieved value $n = 1.000257$.

---

[12] The Chebyshev nodes for N data points for an arbitrary range (a, b) are given according to the formula: $x_k = \frac{1}{2}(a+b) + \frac{1}{2}(b-a)\cos\left(\frac{2k-1}{2n}\pi\right)$, $k = 1\,,...,\, N$.

[13] Remark that $\mathrm{Re}[\chi^{(1)}]$ is an odd function and thus can be applied on negative frequencies.

[14] The susceptibility formula for $\chi^{(1)}$, Eq. (2.39), assumes monochromatic waves, i.e. infinite pulse. By extending the pulse length in the 1D-TDSE, we asymptotically approach closer to the exact value. However, there is still some margin of error since we don't account for the resonances in the 1D-TDSE.

[15] The 1D-TDSE populates the excited states at different rates compared to the 3D-TDSE and the potentials are modeled differently in 1D and 3D. These are some of the key factors for the discrepancy.

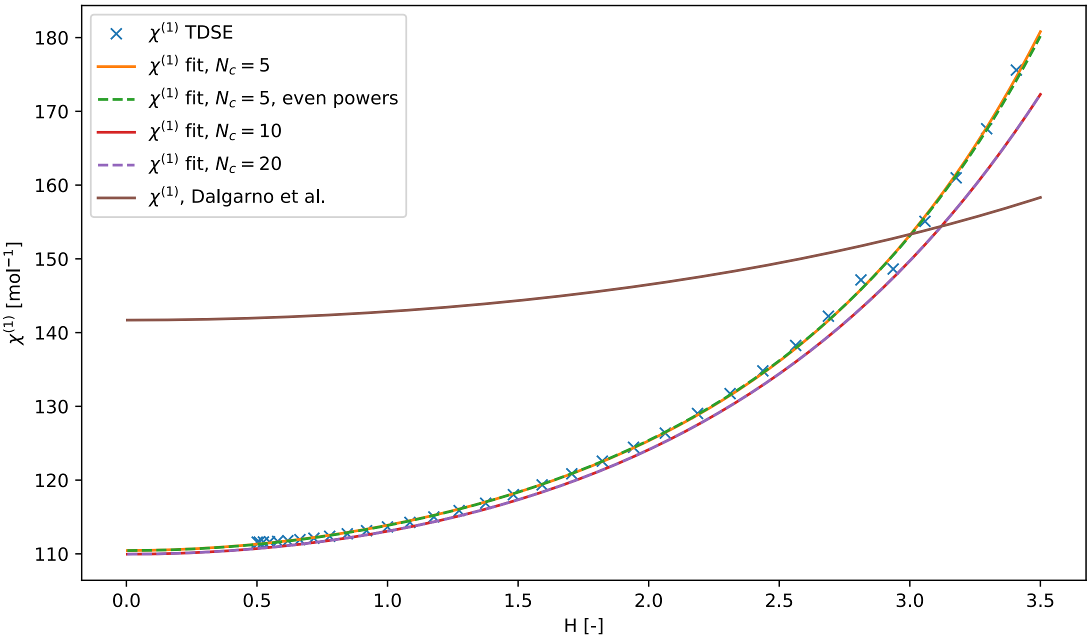

Figure 2.3: Least squares fit of $\chi^{(1)}$ (orange curve) from the numerical data (blue cross) for the harmonic range $(0, 3.5)$ of a 5-cycle FWHM field. The fit is overlayed by polynomial fit of $\chi^{(1)}$ keeping only the even powers (green dashed curve). Next, polynomial $\chi^{(1)}$ fits for $N_c = 10$ (red curve) and $N_c = 20$ (violet dashed curve) are displayed. Finally, the $\chi^{(1)}$ fit computed using the Dalgarno et al. [70] model is plotted (brown curve).

In the rest of the thesis, we will employ the fit of $\chi^{(1)}(\omega)$ for the 5-cycle pulse, Eq. (2.46), because simulating for longer pulses is very time-consuming from the 1D-TDSE standpoint. Since we want to acquire a lot of data for training the neural network in a reasonable time frame and tailor its inputs to be as small as possible, 5-cycle pulses are sufficient. In addition to the reasonable R^2 coefficient, we may consider the polynomial fit (2.46) adequate for the implementation to the Maxwell solver. The 5-cycle pulses were also used to retrieve the $\chi^{(3)}$ coefficients from the 1D-TDSE.

2.2.4 Computation of $\chi^{(3)}$ from the 1D-TDSE

Extraction of $\chi^{(3)}$ gives us a way to quantify the Kerr effect and third harmonic generation in the *ad-hoc* model. We could follow a similar procedure proposed in Subsection 2.2.2 by including higher-order terms of $\varrho(t)$ expansion in the evaluation of the expectation value of $\langle \boldsymbol{\mu} \rangle$. However, the final expression contains a product of 4 matrix elements of $\boldsymbol{\mu}$ and summation over 3 indices and can be found for example in Boyd, Section 3.7 [8]. Instead, we realize that we can expand susceptibility, assuming instantaneous response, as

$$\chi(\omega, I) = \chi^{(1)}(\omega) + \chi^{(3)}(\omega) \cdot I. \tag{2.47}$$

If we treat χ as a linear function of intensity, we can extract $\chi^{(3)}(\omega)$ by computing the quantity

$$\chi(\omega, I) = \frac{\langle \mu \rangle}{\varepsilon_0 E(\omega, I)} \tag{2.48}$$

for various frequencies at different intensities until the threshold where the expansion still makes physical sense. To draw the line for the threshold intensity, we once again selected a 5-cycle (FWHM) 800 nm pulse of different peak intensities ranging from 10^{10} - 10^{14} W/cm^2. The intensity samples are again given by the Chebyshev nodes distribution. We fitted the data, computed with Eq. (2.48), using a first-degree polynomial in peak intensity I, Eq. (2.47), and the result is depicted in Figure 2.4. For better representation, the data is plotted in a log scale on the intensity axis.

The fit and the numerical χ values are plotted alongside the populations of the bound states and continuum after the pulse propagation to determine the ionization regime threshold. It is reasonable to draw the line for the *ad-hoc* regime below the intensity 10^{13} W/cm^2 as the depopulation of the ground state and the population of the other excited states ($n > 1$) rises exponentially with increasing intensity after this threshold. As seen from Figure 2.4, the population of the continuum comes mostly at the expense of the depletion of the ground state. We notice that around $I = 3 \cdot 10^{13}$ W/cm^2 the perturbation regime crumbles as the χ curve (blue curve) is no longer linearly dependent on intensity (red dotted curve), hinting HOKE terms involvement. This observation is in line with results from the literature [64]. Therefore, we can set the intensity 10^{13} W/cm^2 as the threshold for the $\chi^{(3)}$ computation. The ionized fraction computed at the end of the pulse using the extended ADK model [72, 73, 74], for high-intensity fields including tunneling, is plotted for reference in Figure 2.4 (orange curve). The details of the continuum population computation and models of ionization are discussed in Appendix A.

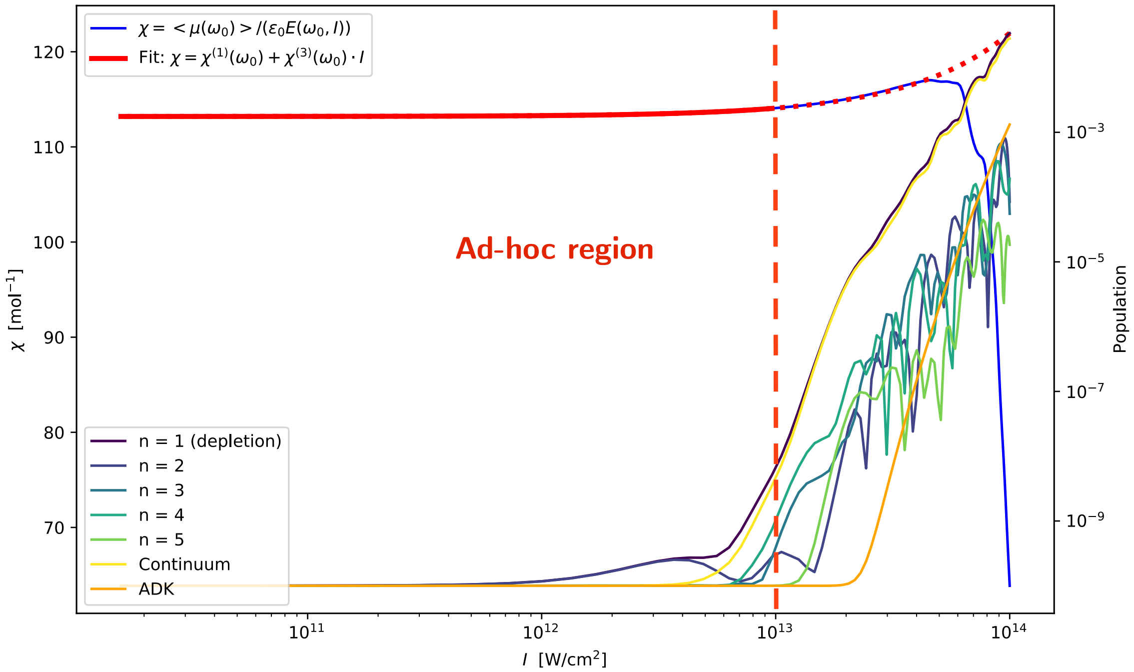

Figure 2.4: On the primary axis we see a linear fit of $\chi(\omega_0)$ (red curve) as a function of intensity (log-scale) at the fundamental frequency ω_0 from the numerical $\chi(\omega_0)$ data (blue curve). The secondary axis shows the ground state depletion ($n = 1$ depletion curve) and population of the first four excited states ($n > 1$ curves) at the end of the pulse propagation. The continuum population draws the line of the ionization regime in which it is meaningless to assume perturbation expansion. The region left to the red dashed vertical line denotes the region where the *ad-hoc* description is still meaningful. The ADK rate is plotted for reference (orange curve).

Finally, a set of 1D-TDSEs for 30 intensity samples in range $(10^{10}, 10^{13})$ W/cm^2 and 51 frequency samples in range $(0.5, 4.5)\,\omega_0$ was computed, totalling 1530 1D-TDSE computations. For each frequency ω in the range, the value of $\chi^{(3)}$ was extracted according to the expected fit (2.47). We again restricted ourselves on range $(0, 3.5)\,\omega_0$ for the interpolation. The resulting plot is depicted in Figure 2.5 where the interpolated datapoints (blue cross) are fitted using a 14th-degree polynomial in ω with $\mathrm{R}^2 = 0.9624$.

Looking closely at Figure 2.5, we notice significant oscillations centered around $\omega \approx 1.9\,\omega_0$ and $\omega \approx 2.9\,\omega_0$ (blue dashed curve) which are smeared out by the fit. We presume these resonance-like oscillations are caused by the non-linear process of three-wave mixing. Three waves of $\omega \approx 1.9\,\omega_0$ and $\omega \approx 2.9\,\omega_0$ sum into resonance frequencies $\omega \approx 5.7\,\omega_0$ and $\omega \approx 8.6\,\omega_0$, see Figure 2.1.

We sanity-check the results by plugging the numbers into Equation (2.5). For argon with atomic density at room temperature and pressure ($\mathcal{N} \doteq 3.6 \cdot 10^{-6}$ a.u.), the 3rd order susceptibility for 800 nm field is $\chi^{(3)} \doteq 3.5812 \cdot 10^{-26}\ \mathrm{m}^2/\mathrm{V}^2$. All in all, the $\chi^{(3)}$ fit is the final piece in the puzzle of assembling the multi-scale model. The practical implementation of the multi-scale model incorporating the susceptibilities follows in the next chapter.

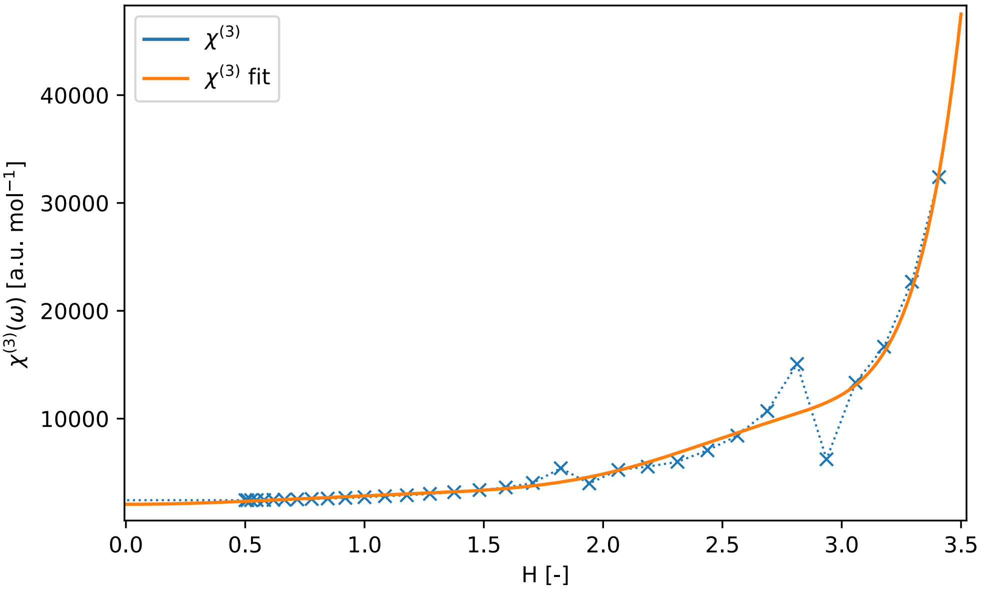

Figure 2.5: Least squares fit of $\chi^{(3)}$ (orange curve) from the numerical data (blue cross) for the harmonic range $(0, 3.5)$. The blue dashed line interconnects the data points for a clearer discrepancy between the datapoints and fit.

Chapter 3

Implementation of the multi-scale model

We first discuss the numerical methods employed within each part of the multiscale model separately – microscopic, governed by 1D-TDSE, and macroscopic Maxwell solver, guided by the unidirectional pulse propagation equation (UPPE) in the co-moving frame, defined in Subsection 2.1.2. Second, we define exactly the *ad-hoc* and *ab-initio* models source terms that will be the subjects for further comparison.

3.1 1D-TDSE

The numerical aspects of the 1D-TDSE are discussed in detail in this section. The scope of the problem is the following:

(i) Find a solution for the eigenproblem of an electron in an effective potential, i.e. solving the time-independent Schrödinger equation (TISE):

$$H_0\psi_0 = E_0\psi_0. \tag{3.1}$$

(ii) Time propagation of the initial wavefunction according to an external electric field encapsulated in the time-dependent interaction potential $V_I(t)$ using TDSE:

$$i\partial_t\psi(t) = \left(H_0 + V_I(t)\right)\psi(t) \tag{3.2}$$

with the initial condition $\psi(t=0) = \psi_0$.

(iii) Computation of expectation values of gauge-invariant observables A at each time step:

$$\langle A(t)\rangle = \langle\psi(t)|\, A\, |\psi(t)\rangle \tag{3.3}$$

We will examine the steps in greater detail later but it is first necessary to establish the representation of the wavefunction and operators for the employment of numerical methods. There exist numerous ways how to solve the TDSE and represent the wavefunctions. We can split them into two families: spectral and grid methods.

Spectral methods take a suitable basis of smooth orthonormal functions $\{\varphi_i\}_{i\in\mathbb{N}} \in \mathscr{H}(\mathbb{R})$, where $\mathscr{H}(\mathbb{R})$ is the Hilbert space, and the solution of the TDSE breaks down to finding time-dependent coefficients of the corresponding basis functions [75]. The basis is usually composed of eigenstates of the field-free Hamiltonian, Gaussian functions or any other suitable smooth functions. It is generally challenging to evaluate the matrix elements of Hamiltonian. However, we do not need as many eigenfunctions for the convergence using the spectral method.

The second family is more versatile and straightforward and that is the grid method. Grid methods heavily rely on the spatio-temporal discretization of wavefunction and operators. The advantage is then an intuitive understanding of the numerical TDSE solution, the disadvantage is the requirement for a higher number of discretization points for convergence[1].

Throughout the thesis, we use the simplest one-dimensional grid method for solving the 1D-TDSE. Concrete implementation of the method is presented in the following subsections.

3.1.1 Wavefunction and operators representation – grid method

We propose the following discretizations for the 1D-TDSE represented on the grid:

- **Spatial grid**:
  Symmetric grid centred around 0, $x_i = i \cdot dx$, $i \in \{-N, \ldots, N\}$. A symmetric grid is important to resolve the parity of the wavefunction and hermiticity of operators.

- **Dot product**:
  The dot product $\langle \bullet | \bullet \rangle$ over the bounded discretized Hilbert space $\mathscr{H}\left((-N \cdot dx, N \cdot dx)\right)$ is defined for $a, b \in \mathscr{H}\left((-N \cdot dx, N \cdot dx)\right)$ as

  $$\langle a | b \rangle = (a^*)^T \cdot b = \sum_i a_i^* \cdot b_i. \tag{3.4}$$

- **Wavefunction**:
  The wavefunction $\psi \in \mathscr{H}\left((-N \cdot dx, N \cdot dx)\right)$ is discretized on every point of the spatial grid, we denote $\psi_i \equiv \psi(x_i)$, $i \in \{-N, \ldots, N\}$. We impose the Dirichlet boundary condition[2] $\psi_{-N} = \psi_N \equiv 0$.

- **Derivatives**:
  The hermiticity of the momentum operator $p = -i\partial_x$ must be satisfied. For a constant step size dx we can define the finite difference discretization of the operator using the central difference as follows:

  $$p\psi_i = -i\frac{\psi_{i+1} - \psi_{i-1}}{2dx} + o(dx^2). \tag{3.5}$$

---

[1] A notable mention is the discrete variable representation (DVR) method [76] which represents the wavefunction on the spatial grid in terms of a set of local smooth basis functions while keeping the operators on the grid. DVR in this sense closely resembles the finite element method [10]. The basis functions on each subinterval of the grid are defined for easy evaluation of the grid operators. DVR requires much less computational grid points while maintaining a high level of accuracy and the operators can still be represented as sparse matrices.

[2] The requirement from quantum mechanics is that the wavefunction asymptotically $\psi \xrightarrow{\pm\infty} 0$. The caveat regarding the numerical implementation is discussed in greater detail later in the subsection.

We can rewrite the derivative operator in matrix representation as

$$p = -\frac{i}{2dx}\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & & \\ 0 & -1 & 0 & \ddots & \\ \vdots & & \ddots & \ddots & 1 \\ 0 & & & -1 & 0 \end{pmatrix} \tag{3.6}$$

We can see that by transposing and complex conjugating the operator matrix (3.6) the hermiticity of the derivative operator, i.e. $p = p^\dagger$, is thus satisfied[3]. The symmetric second derivative operator $p^2 = -\partial_x^2$ is analogical:

$$p^2 = \frac{1}{dx^2}\begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & \ddots & \\ \vdots & & \ddots & \ddots & -1 \\ 0 & & & -1 & 2 \end{pmatrix} + o(dx^2) =: -\Delta \tag{3.7}$$

- **Potential**:
  The problem with an exact Coulomb potential for centrosymmetric atoms $V_C = -\frac{1}{r}$ is the non-integrable pole in 1D for $r = 0$. To overcome this issue, the effective soft core potential V_{eff} is introduced as:

  $$V_{\text{eff}} = -\frac{1}{\sqrt{x^2 + a^2}}, \tag{3.8}$$

  where a is an atomic parameter which sets the ionization potential of the atom[4]. The values of a for various rare gases may be found in [112], Tab. 7.1. For a more elaborate analysis of suitability and generalization of soft-core potentials refer to [77]. The potential is represented as a diagonal matrix in the grid approach:

  $$V_{\text{eff}} = \operatorname{diag}(V_{-N}, V_{-N+1}, \dots, V_N) \quad , \quad V_i \equiv V_{\text{eff}}(x_i). \tag{3.9}$$

The finite differences are by definition 2nd order accurate, however, the accuracy of the second derivative can be increased using the Numerov method. For an ODE

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = f(x, y) =: \Delta y \tag{3.10}$$

the Numerov method gives 6th order accurate finite difference discretization

$$y_{i+1} - 2y_i + y_{i-1} = \frac{dx^2}{12}\left(f_{i+1} + 10 f_i + f_{i-1}\right) + o(dx^6). \tag{3.11}$$

Analogically to the differential operators, we can define the matrix representation of the Numerov matrix

$$M_2 = \frac{1}{12}\begin{pmatrix} 10 & 1 & 0 & \dots & 0 \\ 1 & 10 & 1 & & \\ 0 & 1 & 10 & \ddots & \\ \vdots & & \ddots & \ddots & 1 \\ 0 & & & 1 & 10 \end{pmatrix}. \tag{3.12}$$

---

[3] Remark that we need to also include the boundary condition to fulfill hermiticity for the definition of the operator p on a finite interval, see Example 4.2.5 in [11].

[4] The value of a for argon is $a = 1.1893$, Tab. 7.1 [112].

The accuracy of the central difference second derivative Δ, Eq. (3.7), can be improved by 4 orders and the improved Numerov 2nd derivative Δ_N is:

$$\Delta_N := M_2^{-1}\Delta \tag{3.13}$$

In total, we can write TISE in discretized, 6th-order accurate matrix representation as follows:

$$H_0\psi = \left(-\frac{1}{2}M_2^{-1}\Delta + V_{\text{eff}}\right)\psi = E\psi. \tag{3.14}$$

We notice that all the aforementioned matrix operators are hermitian (symmetric) and tridiagonal[5]. The tridiagonal system of algebraic equations can be computed using the Thomas algorithm and this feature is exploited throughout the computation.

Remark that the boundary $\psi_0 \equiv \psi_N = 0$. This effectively implies a reflective boundary[6] condition which, if the grid size is insufficient, can disrupt the wavefunction propagation. However, the design of the absorber for the grid method is a difficult problem on its own and simple exponential damping also causes reflections. One notable example is the exterior complex scaling, see [78]. We do not use any absorber for the 1D-TDSE in the scope of this thesis.

3.1.2 Eigenproblem

The first step is to obtain the ground state (GS) wavefunction and energy from the TISE. An iterative method is used based on the following resolvent operator [11]:

$$R = (E - H_0)^{-1}, \tag{3.15}$$

We set an initial guess for energy E and wavefunction ψ and apply the resolvent iteratively until the convergence criteria are met. This approach is significantly faster than doing the full diagonalization of the Hamiltonian H_0. The computation of the ground state ψ_0 and initial energy E_0 is summarized in Algorithm 1. The initial values for GS wavefunction are $\psi_i \equiv 1$, $\forall i$, and E_0 is set closest to the desired corresponding ground state energy, usually[7] $E_0 \equiv -1$ a.u. Remark that while the resolvent method converges fast towards the eigenstate, it is only capable of finding a single GS wavefunction ψ_0 and energy E_0 at the time. We set parameter $\delta = 10^{-15}$ which gave reasonable convergence for our purposes. The ground state and excited states computed using the resolvent iterative method are depicted in Figure A.2 in Appendix A.

3.1.3 Time propagation

The time evolution of the ground state wavefunction ψ_0 is governed by a unitary evolution operator $U(t_0, t)$ such that $\psi(t) = U(t_0, t)\psi_0$ where $\psi(t)$ is the exact solution to the TDSE.

---

[5] We can expand the system of equations (3.14) by the Numerov matrix M_2 to obtain:

$$\left(-\frac{1}{2}\Delta + M_2 V_{\text{eff}}\right)\psi = E M_2 \psi$$

which becomes a tridiagonal system.

[6] Physically the effect of a finite box translates to $V(|x| > x_{\max}) \to +\infty$.

[7] The goal is to iteratively find the GS given the spatial discretization. The energy values are well known, see Tab. 7.1 [112], and for argon, the ground state energy is $E_0 = -0.5792$ a.u..

Algorithm 1 Ground state search using resolvent

```
$E \leftarrow E_{\text{guess}}$
$\psi \leftarrow \psi_{\text{guess}}$
while $|E_{\text{old}} - E| > \delta$ do
    $E_{\text{old}} \leftarrow E$
    Projection: $\psi' \leftarrow R\psi$
    Normalize: $\psi \leftarrow \frac{\psi'}{||\psi'||}$
    Eigenenergy: $E \leftarrow \frac{\langle\psi|H_0|\psi\rangle}{||\psi||^2}$
end while
```

A formal solution of the TDSE for the operator $U(t, t_0)$ with time-dependent Hamiltonian $H(t)$ is

$$U(t, t_0) = \mathcal{P} \exp\left(-i \int_{t_0}^{t} H(t')\mathrm{d}t'\right) = \mathcal{P} \exp\left(-i \int_{t_0}^{t} (H_0 + V_I(t'))\mathrm{d}t'\right), \tag{3.16}$$

Where $\mathcal{P}$ refers to the path-ordering operator[8]. Assuming a small temporal evolution step dt and incremental evolution $U(t, t + dt)$, we can approximate the formal solution by[9]

$$U(t, t + dt) = \exp\left(-i \int_{t}^{t+dt} H(t')\mathrm{d}t'\right) \approx \exp\left(-i(H_0 + V_I)dt\right). \tag{3.17}$$

The next approximation revolves around splitting the exponential operator (3.17). The Baker-Hausdorff formula for exponential operators states that for operators A, B the following expansion applies [6]:

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\dots}. \tag{3.18}$$

Since H_0 and V_I are generally non-commuting operators, applying Baker-Hausdorff formula on (3.17) yields

$$U(t, t + dt) = \exp\left(-iH_0 dt\right) \exp\left(-iV_I dt\right) \exp\left(\frac{i}{2}[H_0, V_I]dt^2 + \dots\right). \tag{3.19}$$

For a very small dt we neglect the commutator term $\exp\left(\frac{i}{2}[H_0, V_I]dt^2 + \dots\right) \approx 1$ and perform operator splitting

$$U(t, t + dt) \approx \exp\left(-iH_0 dt\right) \exp\left(-iV_I dt\right) =: U_0(t, t + dt) \cdot U_I(t, t + dt). \tag{3.20}$$

Operators $U_0(t, t + dt)$ and $U_I(t, t + dt)$ are treated separately.

---

[8] Due to non-commutativity of the time-dependent Hamiltonian, a so-called ordered exponential is defined and can be written in terms of an infinite sum

$$\mathcal{P} \exp\left(\int_{t_0}^{t} H(t')\mathrm{d}t'\right) = \sum_{n=0}^{\infty} \int_{t_0}^{t} \dots \int_{t_0}^{t} H(t'_1) \dots H(t'_n)\mathrm{d}t'_1 \dots \mathrm{d}t'_n.$$

For infinitesimally small time evolution step dt the path-ordering operator is converted into the product of individual exponential operators, as discussed further.

[9] We employ a composition property of the evolution operator

$$U(t_2, t_0) = U(t_2, t_1)U(t_1, t_0) \quad , \quad t_2 > t_1 > t_0$$

so that we can simply discretize the evolution of $U(t, t_0)$.

Operator $U_0(t, t + dt)$ is treated using the Crank-Nicolson method. Crank-Nicolson (CN) is an unconditionally stable, second-order numerical scheme based on averaging explicit and implicit Euler methods for ODEs. For TDSE $i\dot{\psi} = H_0\psi$ the CN scheme, denoting $\psi(t) \equiv \psi^n$ and $\psi(t + dt) \equiv \psi^{n+1}$, is the following:

$$i\frac{\psi^{n+1} - \psi^n}{dt} = \frac{1}{2}\left(H_0\psi^{n+1} + H_0\psi^n\right). \tag{3.21}$$

Rearranging scheme (3.21) by explicitly expressing ψ^{n+1} gives[10]

$$\psi^{n+1} = \left(1 + \frac{i}{2}H_0 dt\right)^{-1}\left(1 - \frac{i}{2}H_0 dt\right)\psi^n \tag{3.22}$$

where we identify unitary operator[11] U_0 as

$$U_0(t, t + dt) = \left(1 + \frac{i}{2}H_0 dt\right)^{-1}\left(1 - \frac{i}{2}H_0 dt\right) =: U_0^{n+1}. \tag{3.23}$$

The discretization of H_0 within U_0 is according to (3.14) and the inverse matrix $\left(1 + \frac{i}{2}H_0 dt\right)^{-1}$ is computed using Thomas algorithm.

For the length gauge Hamiltonian the application of the interaction term U_I is straightforward since

$$V_I(t) = x \cdot E(t) \quad , \quad U_I(t, t + dt) = \exp\left(-iV_I(t)dt\right) =: U_I^{n+1} \tag{3.24}$$

which essentially results in a multiplication by the complex exponential function.

The wavefunction ψ in step n is propagated in time to the next step $n + 1$ as

$$\psi^{n+1} = U_I^{n+1} \cdot U_0^{n+1} \cdot \psi^n. \tag{3.25}$$

In total, the propagation scheme (3.25) is order $o(dx^6, dt^2)$ accurate[12].

3.1.4 Expectation values

We can directly compute the expectation value of operators[13] x and ∇V using the wavefunction ψ at time t as

$$\langle x(t)\rangle = \langle\psi(t)|x|\psi(t)\rangle = \sum_i |\psi_i(t)|^2 x_i, \tag{3.26}$$

$$\langle\nabla V(t)\rangle = \langle\psi(t)|\nabla V|\psi(t)\rangle = \sum_i |\psi_i(t)|^2 \nabla V_i \tag{3.27}$$

---

[10] Remark that we can expand the numerator and denominator of the equation by the Numerov matrix M_2 to skip the computation of inverse M_2^{-1} as written in the unperturbed Hamiltonian in Eq. (3.14). See the exact implementation in [112], Eq. (7.39).

[11] It can be shown directly that operator U_0 is indeed unitary, i.e. $U_0^\dagger U_0 = 1$.

[12] We increase the temporal accuracy by an order of magnitude by using a Trotter formula

$$e^{\lambda(A+B)} = \lim_{m\to}(e^{\lambda A/m}e^{\lambda B/m})^m$$

to split the operator (3.17) symmetrically as

$$e^{-iV_I dt/2}e^{-iH_0 dt}e^{-iV_I dt/2}.$$

During the temporal evolution, the operators $e^{-iV_I dt/2}$ get multiplied from the application of the split operator in the previous step, so the Trotter formula in scheme (3.25) is hidden in the propagation of the wavefunction.

[13] The dipole operator μ used in the previous chapter is defined as $\mu = -x$ so the sign is simply flipped.

where

$$\nabla V_i = \frac{x_i}{(x_i^2 + a^2)^{3/2}}. \tag{3.28}$$

The population of the ground state, respectively the ground state depletion, is defined, given the initial ground state ψ_0, as the following dot product[14]

$$|\langle\psi_0|\psi(t)\rangle|^2 = \left|\sum_i \psi_{0,i}\psi_i^*(t)\right|^2. \tag{3.29}$$

For the length gauge Hamiltonian, the gauge independent source terms are:

$$\partial_t \langle j\rangle^L = -\langle\ddot{x}\rangle \quad , \quad \ldots\text{Length form} \tag{3.30}$$

$$\partial_t \langle j\rangle^L = -\langle\dot{p}\rangle \quad , \quad \ldots\text{Velocity form} \tag{3.31}$$

$$\partial_t \langle j\rangle^L = \langle\nabla V\rangle + E \quad , \quad \ldots\text{Acceleration form,} \tag{3.32}$$

where E is the electric field. Remark that in practice we compute $\langle x\rangle$ and $\langle\nabla V\rangle$ and the derivatives are added in the frequency domain by multiplication of the source term with $-i\omega$. The velocity form (3.31) is unsuitable as a source term due to the evaluation of the momentum operator $\langle p\rangle = \langle\psi| - i\nabla|\psi\rangle$ numerically.

3.2 Maxwell solver – UPPE

First, we recall the UPPE in the co-moving frame to be solved for both *ad-hoc* and *ab-initio* models (Eq. (2.26)):

$$\partial_z\hat{E} = i\frac{\omega}{2\varepsilon_0 c}\left(\hat{F} - \varepsilon_0\chi_0^{(1)}\hat{E}\right), \tag{3.33}$$

where $\hat{E}$ is the propagated electric field, $\chi_0^{(1)}$ is linear susceptibility at $\lambda = 800$nm, obtained from the susceptibility computation discussed in Section 2.2.3, and $\hat{F}$ is the source term that differs for *ad-hoc* and *ab-initio*. Remark that all the quantities used are converted to atomic units to keep the same units across the macroscopic and microscopic solvers.

3.2.1 Propagation equations

Because the UPPE (3.33) is in the frequency domain, we effectively solve the following ODE for all frequency components ω independently as:

$$\partial_z\hat{E}(z;\omega) = f\left(z, \hat{E};\omega\right), \tag{3.34}$$

$$f\left(z, \hat{E};\omega\right) = i\frac{\omega}{2\varepsilon_0 c}\left(\hat{F}\left(z, \hat{E};\omega\right) - \varepsilon_0\chi_0^{(1)}\hat{E}\left(z;\omega\right)\right). \tag{3.35}$$

We grabbed the 4th order Runge-Kutta (RK4) to solve ODE (3.34). The explicit RK4 for a step size dz is given by[15]:

$$\hat{E}(z+dz) = \hat{E}(z) + \frac{dz}{6}\left(k_1 + 2k_2 + 2k_3 + k4\right) + o(dz^5) \tag{3.36}$$

---

[14] This quantity is not gauge-invariant. A proper spectral way to interpret it as the depletion of the ground state in the length gauge is shown in [79].

[15] $\hat{E}$ and $\hat{F}$ are automatically assumed as ω-dependent.

where

$$k_1 = f\left(z, \hat{E}(z)\right), \tag{3.37}$$

$$k_2 = f\left(z + dz/2, \hat{E}(z) + dz \cdot \frac{k_1}{2}\right), \tag{3.38}$$

$$k_3 = f\left(z + dz/2, \hat{E}(z) + dz \cdot \frac{k_2}{2}\right), \tag{3.39}$$

$$k_4 = f\left(z + dz, \hat{E}(z) + dz \cdot k_3\right). \tag{3.40}$$

To summarize, we are solving an initial value problem of a laser pulse $\hat{E}_0 = \hat{E}(z = 0, \omega)$ propagating through a gas to distance $z = L$ and the goal is to obtain $\hat{E}(z = L, \omega)$ using the RK4 scheme. At each Runge-Kutta step, the most expensive part is the evaluation of f, respectively the source term $\hat{F}$. Once $\hat{E}$ is computed for every ω, the field must be first converted into the temporal domain at every intermediate step to evaluate F and then back into the spectral domain to be used in the next step. The pseudocode scheme is written in Algorithm 2. We notice that forward and inverse fast Fourier transform (FFT and IFFT) is computed 8 times and source term F is computed 4 times per dz. The Algorithm 2 was fully implemented in C programming language and FFTW3 library [80] for computing fast Fourier transforms was used.

3.2.2 Source terms

We can finally implement the *ad-hoc* and *ab-initio* models source terms. We recall the wave equation for electric field with sources (2.6):

$$\Delta E - \frac{1}{c^2}\partial_t^2 E = \frac{1}{\varepsilon_0 c^2}\partial_t^2 F. \tag{3.41}$$

The only difference between the models is now the source term F. The source term of *ad-hoc* $F_{\text{ad-hoc}}$ implements the medium response through material constants (susceptibilities $\chi^{(1)}, \chi^{(3)}$) and response functions (polarisation $P(t)$) whereas the *ab-initio* source term $F_{\text{ab-initio}}$ is computed via complementary 1D-TDSE solver. Remark that the spatial grid for the 1D-TDSE, on which we compute the evolution of the wavefunction, is different from the grid on which we propagate the electric field.

After identifying the source term F, the pivotal equation is the UPPE in co-moving frame (2.26):

$$\partial_z \hat{E} = i\frac{\omega}{2\varepsilon_0 c}\left(\hat{F} - \varepsilon_0 \chi_0^{(1)} \hat{E}\right). \tag{3.42}$$

Linear susceptibility has the same numerical value for both models: $\chi_0^{(1)} = \mathcal{N}\chi^{(1)}(\omega_0)$ where $\mathcal{N}$ is the medium density and $\chi^{(1)}(\omega_0)$ is obtained from the polynomial fit (2.46) at 5 cycles of the field.

Ab-initio **model**

The *ab-initio* source term enters the wave equation (3.41) in terms of the averaged microscopic fields in a volume (Eq. (2.9)) as

$$F_{\text{ab-initio}} := \mathcal{N} \cdot \varepsilon_0 E_{\text{micro}} \tag{3.43}$$

Algorithm 2 Maxwell solver

```
$E \leftarrow E(z=0)$
**FFT**: $\hat{E} \leftarrow \mathscr{F}[E]$
**while** $z \leq L$ **do**
    /*** First step ***/
    **Evaluate**: $F(z, E)$
    **FFT**: $\hat{F} \leftarrow \mathscr{F}[F(E)]$
    $k_1 \leftarrow f(\hat{F}, \hat{E})$
    $\hat{E}_{k_1} \leftarrow \hat{E} + dz \cdot \frac{k_1}{2}$
    **IFFT**: $E_{k_1} \leftarrow \mathscr{F}^{-1}\left[\hat{E}_{k_1}\right]$
    /*** Second step ***/
    **Evaluate**: $F(z + dz/2, E_{k_1})$
    **FFT**: $\hat{F} \leftarrow \mathscr{F}[F(E_{k_1})]$
    $k_2 \leftarrow f(\hat{F}, \hat{E}_{k_1})$
    $\hat{E}_{k_2} \leftarrow \hat{E} + dz \cdot \frac{k_2}{2}$
    **IFFT**: $E_{k_2} \leftarrow \mathscr{F}^{-1}\left[\hat{E}_{k_2}\right]$
    /*** Third step ***/
    **Evaluate**: $F(z + dz/2, E_{k_2})$
    **FFT**: $\hat{F} \leftarrow \mathscr{F}[F(E_{k_2})]$
    $k_3 \leftarrow f(\hat{F}, \hat{E}_{k_2})$
    $\hat{E}_{k_3} \leftarrow \hat{E} + dz \cdot k_3$
    **IFFT**: $E_{k_3} \leftarrow \mathscr{F}^{-1}\left[\hat{E}_{k_3}\right]$
    /*** Final step ***/
    **Evaluate**: $F(z + dz, E_{k_3})$
    **FFT**: $\hat{F} \leftarrow \mathscr{F}[F(E_{k_3})]$
    $k_4 \leftarrow f(\hat{F}, \hat{E}_{k_3})$
    $\hat{E} \leftarrow \hat{E} + \frac{dz}{6}(k_1 + 2k_2 + 2k_3 + k4)$
    **IFFT**: $E \leftarrow \mathscr{F}^{-1}\left[\hat{E}\right]$
    **Step**: $z = z + dz$
**end while**
```

The microscopic field E_{micro} is expressed in terms of the averaged current $\langle j \rangle$ in the frequency domain as (recalling Eq. (1.28))

$$\hat{E}_{\text{micro}} = -\frac{i}{\omega\varepsilon_0}\langle\hat{j}\rangle\,. \tag{3.44}$$

Combining Eqs. (3.43) and (3.44) yields the UPPE in co-moving frame (3.42) for the *ab-initio* model:

$$\partial_z\hat{E} = i\frac{\omega}{2\varepsilon_0 c}\left(\frac{i}{\omega}\mathcal{N}\langle\hat{j}\rangle - \varepsilon_0\chi_0^{(1)}\hat{E}\right). \tag{3.45}$$

We recall $\hat{E}_{\text{micro}}$ can be written in 3 equivalent forms – length, velocity and acceleration, see Section 1.2. For length and acceleration forms, the current reads:

$$\langle\hat{j}\rangle = i\omega\varepsilon_0\hat{E}_{\text{micro}} = \underbrace{i\omega\widehat{\langle x\rangle}}_{\text{Length form}} = \underbrace{i\frac{1}{\omega}\widehat{\langle\nabla V_{\text{eff}} + E\rangle}}_{\text{Acceleration form}} \tag{3.46}$$

Plugging $\langle \hat{j} \rangle$ from (3.46) into (3.45) yields the final UPPE for the *ab-initio* model:

$$\partial_z \hat{E} = i\frac{\omega}{2\varepsilon_0 c}\left(\frac{1}{\omega^2}\mathcal{N}\widehat{\langle -\nabla V_{\text{eff}} - E\rangle} - \varepsilon_0 \chi_0^{(1)} \hat{E}\right) \qquad \ldots \text{Acceleration form,} \tag{3.47}$$

$$\partial_z \hat{E} = i\frac{\omega}{2\varepsilon_0 c}\left(\mathcal{N}\widehat{\langle -x\rangle} - \varepsilon_0 \chi_0^{(1)} \hat{E}\right) \qquad \ldots \text{Length form,} \tag{3.48}$$

where we identify $\hat{F}_{\text{ab-initio}}$ according to the Algorithm 2 as

$$\hat{F}_{\text{ab-initio}}(E) = \frac{1}{\omega^2}\mathcal{N}\widehat{\langle -\nabla V_{\text{eff}} - E\rangle} = \mathcal{N}\widehat{\langle -x\rangle} \tag{3.49}$$

in the acceleration and length forms respectively where $\langle \nabla V_{\text{eff}} \rangle$ and $\langle x \rangle$ are observables computed using the 1D-TDSE for the field E.

Ad-hoc **model**

We start from the same point with the derivation of the *ad-hoc* model by identifying $\hat{F}_{\text{ad-hoc}}$ according to nonlinear optics as follows:

$$\hat{F}_{\text{ad-hoc}}(E) = \hat{P}^{\text{L}}(E) + \hat{P}^{\text{NL}}(E) = \varepsilon_0\left(\chi^{(1)}(\omega)\hat{E} + \chi^{(3)}(\omega)\widehat{E^3}\right) \tag{3.50}$$

where we took into consideration only linear dispersion and 3rd-order nonlinearities (Kerr and 3HG) since we assume a gas with central inner symmetry.

Let us briefly discuss the modeling of Kerr and 3rd harmonic. In the standard textbook [8] descriptions of higher order nonlinearities, we assume real uniform waves[16] with single frequency component ω, i.e. $E(t) = \text{Re}\left[A_\omega e^{-i\omega t}\right]$, where A_ω is the corresponding amplitude, generally complex. Picking the 3rd order nonlinearity, it is meaningful to include 3 waves

$$E(t) = \text{Re}\left[A_{\omega_1} e^{-i\omega_1 t} + A_{\omega_2} e^{-i\omega_2 t} + A_{\omega_3} e^{-i\omega_3 t}\right] \tag{3.51}$$

interacting through 3-wave mixing as $E^3(t)$ yielding in total 28 nonlinear 3rd order effects such as sum-frequency generation, subharmonic generation and other frequency combinations including 3HG and also Kerr effect. It is impractical to analyze a combination of more than 3 waves for 3rd-order nonlinearities. However, for numerical simulations, it is sufficient to model them by computing the Fourier transform of the field $E^3(t)$, giving $\widehat{E^3}$. We then multiply by the corresponding proportionality factor in the Fourier space – $\chi^{(3)}(\omega)$, which quantifies the Kerr nonlinear refractive index

$$n_2 = \frac{3}{8n_0}\chi^{(3)}(\omega) \tag{3.52}$$

and the magnitude of the 3HG as

$$|A_{3\omega}| = \frac{1}{4}\chi^{(3)}(3\omega)|A_\omega|^3. \tag{3.53}$$

In total, the $\chi^{(3)}$ factor is split for a single frequency component ω between Kerr and 3HG schematically as follows[17]

$$\text{Re}\left[\chi^{(3)}\widehat{E^3}\right](\omega) \sim \chi^{(3)}(\omega)\text{Re}\left[\underbrace{\frac{3}{4}|A_\omega|^2 A_\omega e^{-i\omega t}}_{\text{Kerr: } \omega-\omega+\omega=\omega} + \underbrace{\frac{1}{4}A_{\omega/3}^3 e^{-i\omega t}}_{\text{3HG: } \omega/3+\omega/3+\omega/3=\omega}\right]. \tag{3.54}$$

---

[16] The electric field in the temporal domain is strictly real quantity.

[17] For simplicity, we assume only the particular solutions for Kerr and 3HG at single frequency component ω.

The shape of Eq. (3.54) is counter-intuitive if we compare it with standard nonlinear optics where we have a fundamental frequency component ω and observe the nonlinear phenomena emerging from the fundamental wave (Kerr, 3HG). Here, the modeling of the nonlinear effects is approached from the other side, i.e. what effects contribute to a single frequency component ω – Kerr and subharmonic frequency.

Finally plugging the *ad-hoc* source term $F_{\text{ad-hoc}}$ (3.50) into UPPE (3.42) gives the UPPE for the *ad-hoc* model

$$\partial_z \hat{E} = i\frac{\omega}{2c}\left((\chi^{(1)}(\omega) - \chi_0^{(1)})\hat{E} + \chi^{(3)}(\omega)\widehat{E^3}\right) \tag{3.55}$$

It should be stressed that the *ad-hoc* model proposed in the scope of the thesis, Eq. (3.55), is not universally applicable for the whole range of laser pulses, see $\chi^{(1)}(\omega)$ fit in Figure 2.3, or for frequencies above H = 3.5. The harmonic threshold is given by the absence of suitable resonance damping within the 1D-TDSE, see discussion in Subsection 2.2.3. In the end, the pulse length was fixed in the following examples to 5 cycles FWHM due to the computational complexity of source term evaluation via 1D-TDSE, as we wanted to compare the two models. Hence the model will not be used for an exhaustive study of different laser parameters but only for different intensities. The aim is to easily generate large amounts of numerical data to feed a neural network designed in Chapter 5

3.2.3 Finite box size considerations

A few last issues regarding the finite numerical temporal grid must be addressed. The FFT introduces several numerical artifacts which, if not treated correctly, may completely invalidate the result. If we apply FFT on a temporal signal and then perform IFFT, the last and first bins of the temporal signal F are equal, i.e. $F_0 \equiv F_N$, effectively introducing the periodic boundary condition. Hence, any forward propagating frequency component may at some point reach the start of the pulse which has a detrimental effect on the whole propagation. This is particularly harmful for the *ab-initio* model where the generation of the resonant frequencies can produce a positive feedback loop and exponentially increase the energy of the system. Because the resonant frequencies propagate slower than the fundamental frequency, they propagate faster within the numerical temporal grid towards the pulse tail to reappear later in the pulse front. To prevent this, a soft filter is introduced at the edges of the numerical window, as proposed in [63]. Physically, the temporal filter introduces a global ad-hoc lifetime of the resonances, indiscriminate of particular frequencies.

The soft filter is an asymmetric temporal window composed of two error functions damping the pulse front and tail in the temporal domain by 5 orders of magnitude. The filter along with an example of a generated source term $-\langle \nabla V \rangle - E$ for a 5-cycle (FWHM), 800 nm Gaussian pulse of peak intensity $I = 10^{10}$ W/cm^2, is depicted in Figure 3.1. The filter was fine-tuned carefully so that the spectrum is affected as little as possible[18]. Source term spectra are plotted in Figure 3.2. The filter effectively enhances the source term spectrum by highlighting the lower intensity features such as 5th and 7th harmonic generation and resonances located around H $\approx$ 5.7 and H $\approx$ 8.6. We can notice a slight artifact of the filter in the close-up in Figure 3.2 for H $\approx$ 0.7 and H $\approx$ 1.3. All in all, the soft filter is well suited for its purpose and serves an optimal compromise between stabilizing the

---

[18] The temporal filter f interacts with spectrum as a convolution of the Fourier signal $\hat{F}$ with $\hat{f}$ as a consequence of the convolution theorem, see Nomenclature.

algorithm and retaining key spectral features. The filter was used only on the *ab-initio* model because of the inherent problems with resonances.

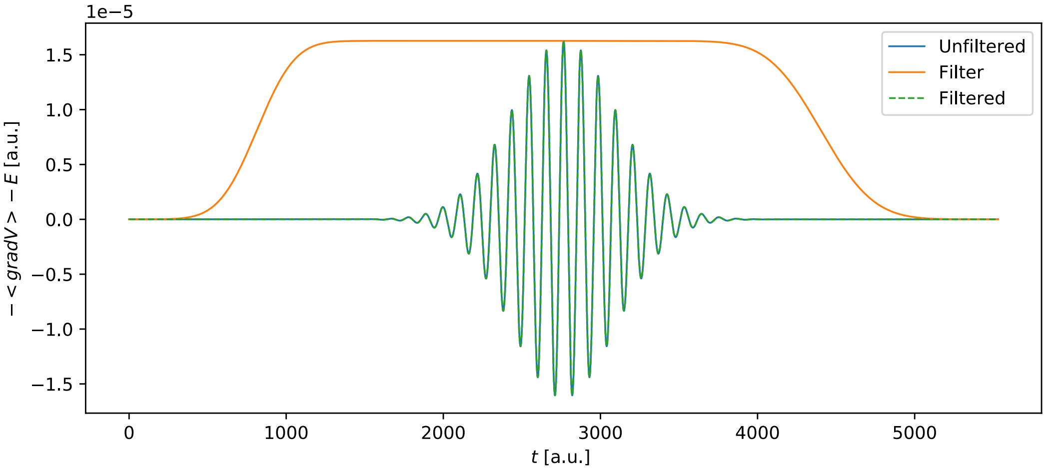

Figure 3.1: Source term $-\langle \nabla V \rangle - E$ for a 5-cycle (FWHM) Gaussian pulse with and without the temporal filter. The temporal filter is plotted alongside the source terms (orange curve)

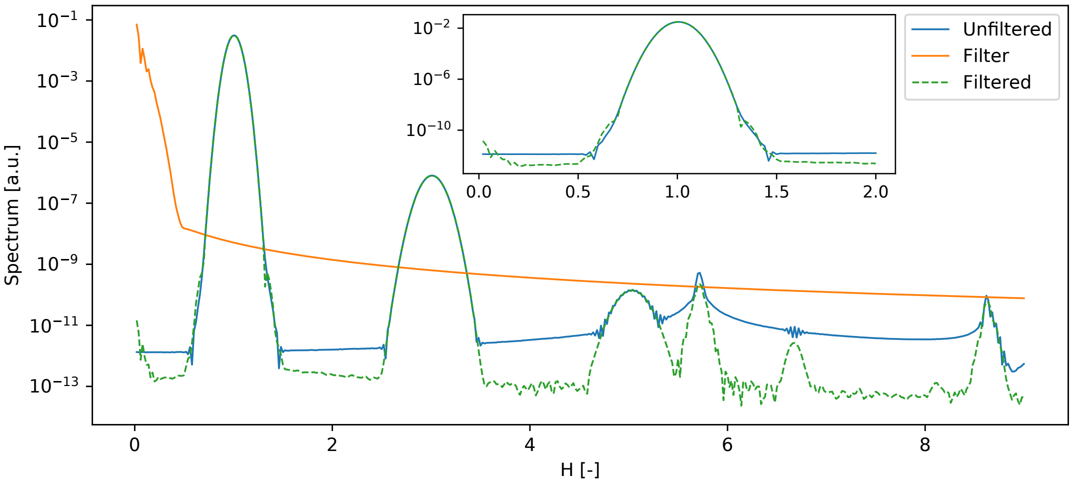

Figure 3.2: Spectrum of the source term $-\langle \nabla V \rangle - E$ for a 5-cycle (FWHM) Gaussian pulse with and without the temporal filter. The frequency response of the filter is plotted alongside the source terms (orange curve) and the filtered spectrum is essentially a convolution of the filter with the unfiltered signal in the spectral domain. A close-up graph of the filtered and unfiltered source terms around H = 1 is shown at the top of the figure.

Chapter 4

Results of the multi-scale model

The results from the multi-scale model are presented in this chapter. First, we look at the direct comparison of the *ad-hoc* and *ab-initio* models for low-intensity fields and compare key features both qualitatively and quantitatively. Next, we discuss the beam propagation using the *ab-initio* model for various orders of laser intensity and we examine the spectral features occurring during the propagation. We conclude with the discussion regarding performance and extension of the multi-scale model to higher dimensions.

Convergence of the multi-scale Maxwell solver has been checked for all of the following simulations by halving the propagation step and comparing the result spectra. The spatial grid resolution for the 1D-TDSE was set to $dx = 0.4$ a.u. and the size of the grid was chosen so that the wavefunction had no reflections from the boundary and was near-zero at the boundary. We used 512 temporal grid points per cycle (1 period ≈ 2.68 fs) of the 800 nm fundamental wavelength and the box size was set to $10 \cdot \tau_{\mathrm{FWHM}}$ where τ_{FWHM} is the FWHM of the Gaussian pulse, w.r.t. intensity. For a 5-cycle FWHM pulse, this translates to a 134 fs box. The number density of argon was set in all the following examples to[1] $0.02504 \cdot 10^{27}$ m^{-3} corresponding to the ambient room pressure and temperature.

4.1 Comparison of *ad-hoc* and *ab-initio*

4.1.1 Short propagation distance

We compared *ad-hoc* and *ab-initio* models for 800 nm 5-cycle (FWHM) pulses at low peak pulse intensity $I = 10^{10}$ W/cm^2. The pulse was propagated through argon for distance $L \doteq 26$ μm, resulting in 10 full RK4 iterations of Maxwell solver with step[2] size $dz \doteq 2.6\,\mu$m of distance. The result spectra for low-intensity pulse are shown in Figure 4.1.

---

[1] Numerical parameter $\mathcal{N} \doteq 3.6 \cdot 10^{-6}$ a.u..

[2] Numerical parameter $dz = 50000$ a.u..

To no surprise, we immediately notice richer spectrum structure within the *ab-initio* model description, including 5th and 7th harmonic generation (HG), resonances and Rydberg states[3]. On the other hand, the *ad-hoc* model as defined in Section 3.2.2, Eq. (3.55) is capable of properly describing only the 1st linear and 3rd order nonlinear effects and upto H = 3.5 due to the $\chi^{(1)}$ and $\chi^{(3)}$ fits. Further, the two models quantitatively fit reasonably well around the 3rd HG (3HG). This is attributed to the fact that we assumed non-constant susceptibility $\chi^{(3)}$, see fit in Figure 2.5. We can see the comparison of using dispersionless $\chi^{(3)}$, computed only for fundamental pulse frequency ω_0, against frequency dependent $\chi^{(3)}$ in Figure 4.2. As we saw in Figure 2.5, the approximate ratio $\chi^{(3)}(3\omega_0) \approx 3\chi^{(3)}(\omega_0)$ applies and the factor of 3 distinguishes the 3rd HG between orange (dispersion-free $\chi^{(3)}$) and green ($\chi^{(3)}(\omega)$) curves in Figure 4.2. This observation further justifies the usage of the *ad-hoc* model and non-constant $\chi^{(3)}(\omega)$ interpolation (Subsection 2.2.4). Remark that we have extrapolated $\chi^{(3)}(\omega > 3.5\omega_0) \equiv \chi^{(3)}(3.5\omega_0)$ in the *ad-hoc* results to avoid computing $\chi^{(3)}(\omega)$ for the resonance frequencies. Linear susceptibility $\chi^{(1)}(\omega)$ is extrapolated in the same fashion.

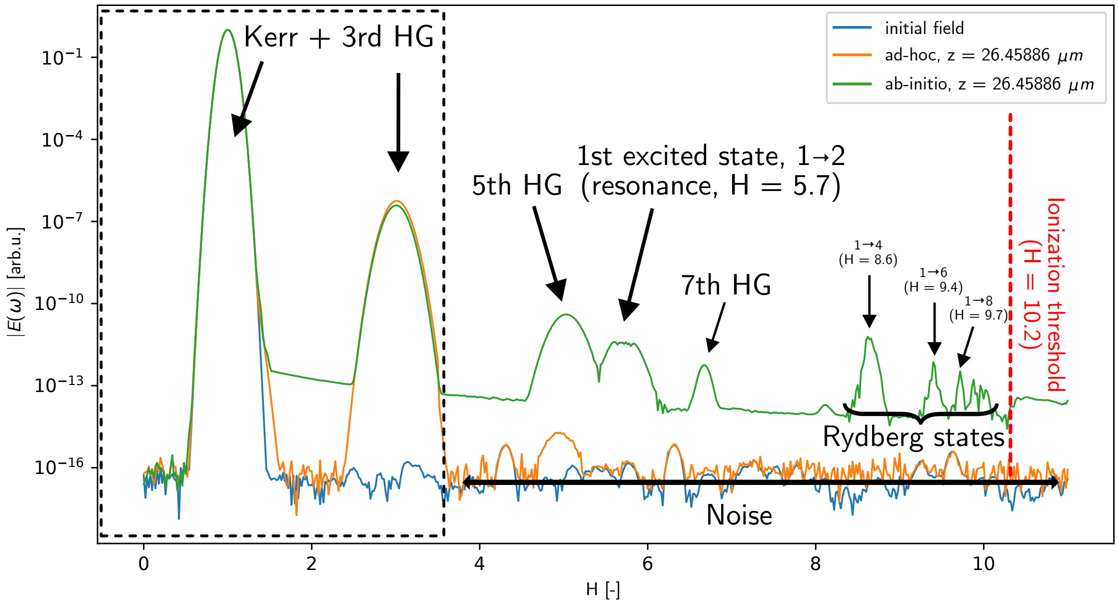

Figure 4.1: Comparison of *ad-hoc* and *ab-initio* models after the propagation of 800 nm, 5-cycle (FWHM) beam of intensity $I = 10^{10}$ W/cm^2 in argon. The propagation distance is $L \doteq 26$ μm, resulting in a total of 10 iterations of the RK4 Maxwell solver. The initial field spectrum is plotted for reference (blue curve). The dashed box roughly emphasizes the range of applicability of the *ad-hoc* model. In the *ab-initio* spectrum, we notice 5th and 7th HG as well as resonances for the 1st excited state at H = 5.7 (transition 1 → 2) and higher excited states (Rydberg states).

---

[3] We refer here to Rydberg states as the transitions from $n = 1$ to $n \geq 4$ and the first resonance as the transition $n = 1$ and $m = 2$. We compute the normalized Bohr transition frequencies for the allowed transitions as the difference of energy of states n, m

$$H = |E_n - E_m|/\omega_0,$$

where ω_0 is the fundamental laser frequency. To determine the allowed transitions, we examine the symmetry of the transitions using the transition moment integral in 1D [9]:

$$\int \psi_n^* x \psi_m \mathrm{d}x,$$

where only the product of odd and even wavefunctions yields a non-zero moment. Thus we see from Figure A.2 that for $n = 1$ the allowed transitions are $m = 2k, k \in \mathbb{N}$.

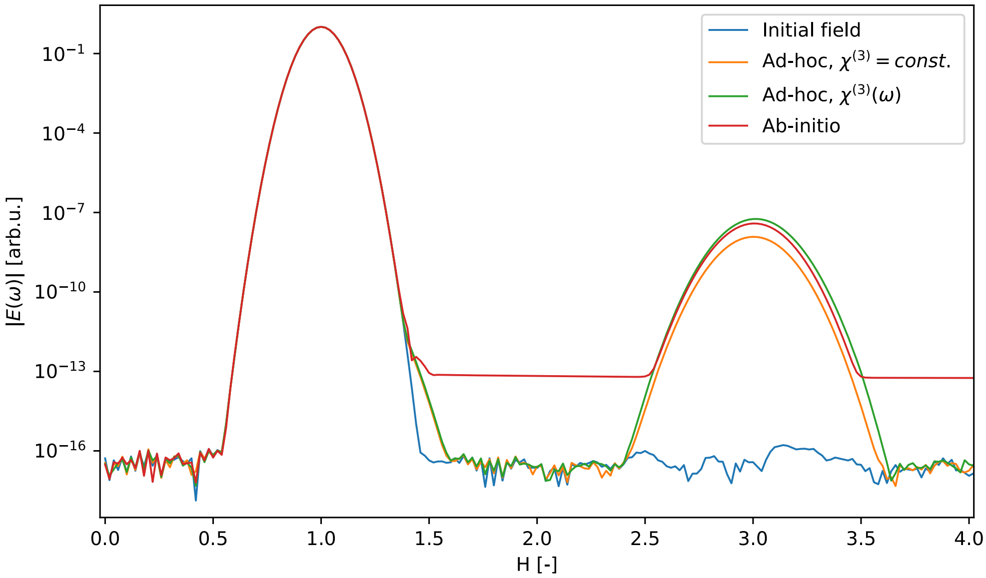

Figure 4.2: Comparison of *ad-hoc* and *ab-initio* models after a single RK4 Maxwell solver iteration for 800 nm, 5-cycle (FWHM) beam of intensity $I = 10^{10}$ W/cm^2 in argon. The initial field spectrum is plotted for reference (blue curve).

4.1.2 Long propagation distance

We compared *ad-hoc* and *ab-initio* models for 800 nm 5-cycle (FWHM) pulses of various intensities – low intensity $I = 10^{10}$ W/cm^2 and intensity below ionisation threshold $I = 10^{13}$ W/cm^2. The pulse was propagated through argon for length $L \doteq 2.6$ mm, resulting in 1000 RK4 iterations of Maxwell solver with step size 2.6 μm.

Low intensity field

The result electric field spectra for low-intensity pulse are shown in Figure 4.3. The first thing to notice is the 5th HG within the *ad-hoc* propagation. Even though there is no 5th-order term description, under favorable phase matching conditions, the 5th harmonic can be also generated via 3-wave mixing: $5\omega = \omega + \omega + 3\omega$. One odd feature is the sudden decrease of amplitude around H = 3.2, see detail in Figure 4.3. The cause of this decrease is unknown. Artificial extrapolation of the susceptibility fit threshold to H = 4.5 has not solved the problem. We can see spatially resolved electric field spectra for the full propagation of *ad-hoc* model in Figure 4.4 that clearly shows the occurrence of the spectral dip in 3rd harmonic after 1.5 mm propagation distance. Doubling the numerical box size to 268 fs and increasing the propagation distance to 5.2 mm still contained the drop as depicted in Figure 4.5. Moreover, another drop reappears at a propagation distance of 3.7 mm.

Observing the spectral colormap for the full *ab-initio* propagation in Figure 4.6, the 5th harmonic in the *ab-initio* spectrum has been overcharged by the strong dispersion of the resonance centered around H = 5.7 which affects also the adjacent frequencies. The resonance appears to spread into frequencies H = 6 − 8 where it creates interference patterns. The Rydberg frequencies disperse slightly during the propagation as seen for the frequency H = 8.6.

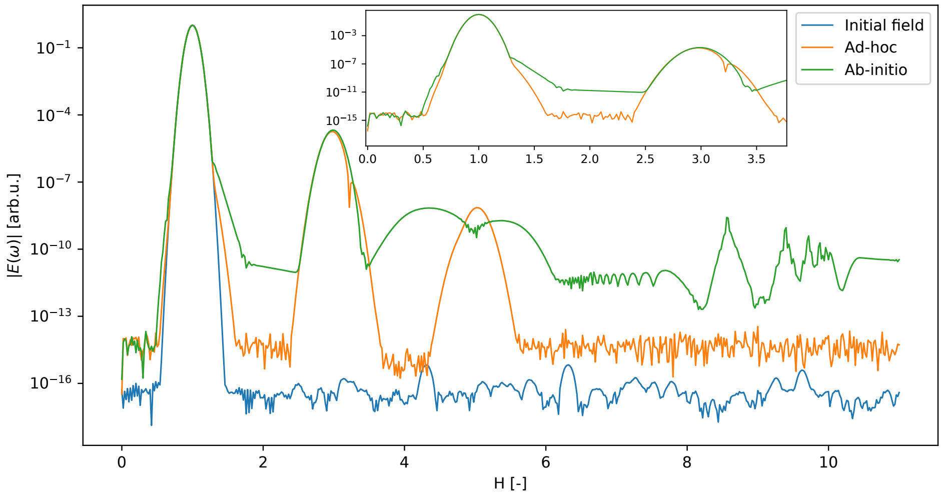

Figure 4.3: Comparison of *ad-hoc* and *ab-initio* electric fields spectra after the propagation of 800 nm, 5-cycle beam of intensity $I = 10^{10}$ W/cm^2 in argon. The propagation distance is $L \doteq 2.6$ mm, resulting in a total of 1000 iterations of the RK4 Maxwell solver. The initial field spectrum is plotted for reference (blue curve).

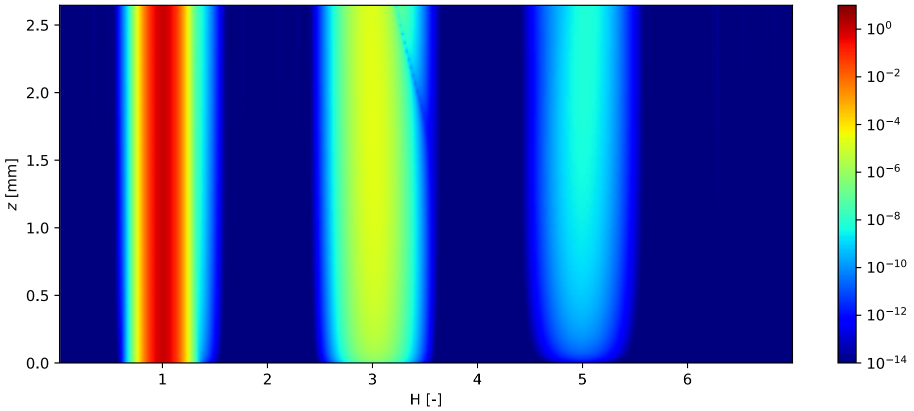

Figure 4.4: Resolved *ad-hoc* electric field spectra for propagation of 800 nm, 5-cycle (FWHM) beam of intensity $I = 10^{10}$ W/cm^2 in argon. The propagation distance is $L \doteq 2.6$ mm.

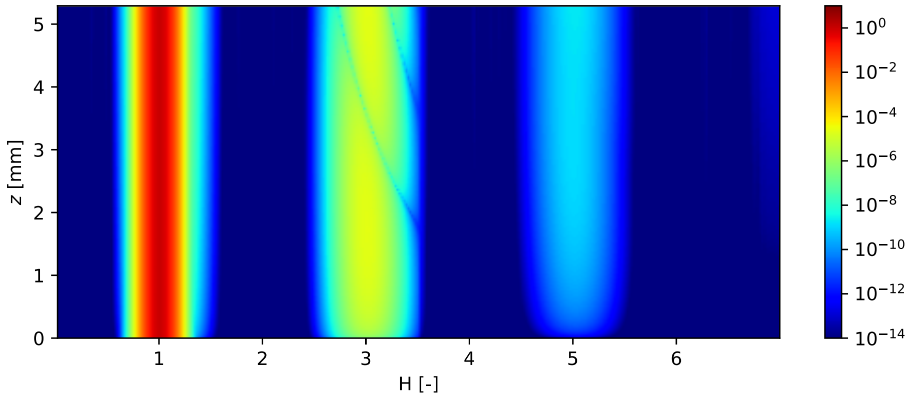

Figure 4.5: Resolved *ad-hoc* electric field spectra for propagation of 800 nm, 5-cycle (FWHM) beam of intensity $I = 10^{10}$ W/cm^2 in argon. The propagation distance is $L \doteq 5.2$ mm.

One last attribute of the *ab-initio* spectrum is the broader fundamental field component as opposed to *ad-hoc*, see detail in Figure 4.3 for H = 1.3 – 1.8. This broadening is caused by the convolution of the temporal filter with the source term signal in the frequency domain, see Subsection 3.2.3. Remark that the temporal filter is not applied in the *ad-hoc* model. Slight modifications of the temporal filter have a significant effect on the spectrum as depicted in Figure 4.7. The comparison was made for 13 μm propagation in argon at intensity $I = 10^{10}$ W/cm^2 at ambient temperature and pressure. We can see that the hard filter with steeper damping introduces a stronger broadening of the fundamental frequency in contrast with the smoother filter. This artificial broadening accumulates over time, which explains the broadening visible in 4.3. At higher intensities, the artificial broadening is overcharged by the Kerr-induced self-phase modulation, as will be discussed further.

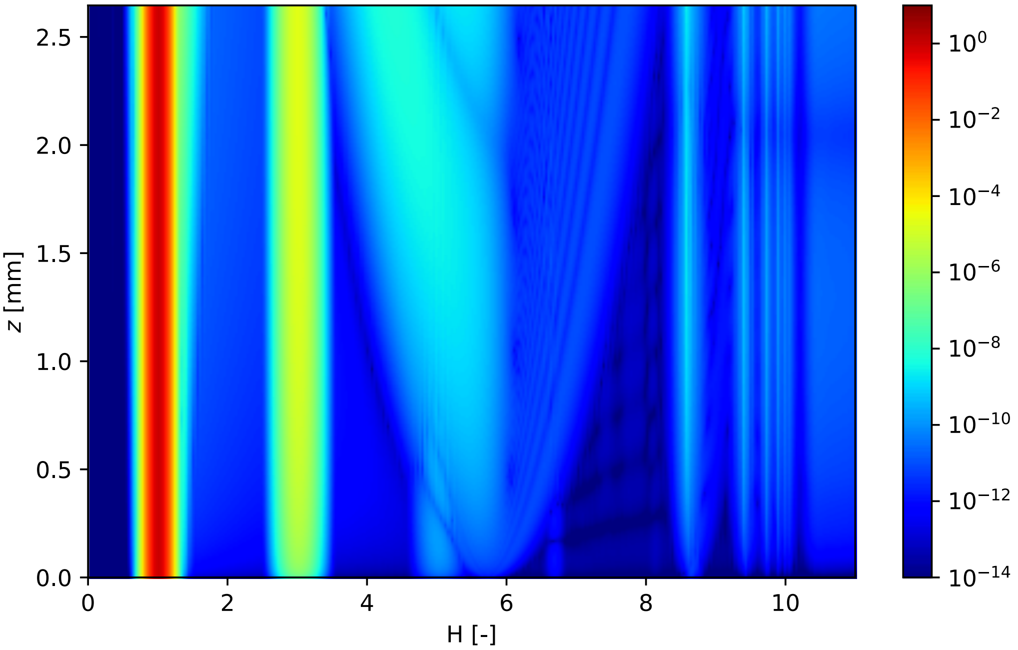

Figure 4.6: Resolved *ab-initio* electric field spectra for propagation of 800 nm, 5-cycle (FWHM) beam of intensity $I = 10^{10}$ W/cm^2 in argon. The propagation distance is $L \doteq 2.6$ mm.

Higher intensity field

The electric field spectra resulting from the propagation of the high-intensity pulse (10^{13} W/cm^2) are shown in Figure 4.8. The validity of the *ad-hoc* is confirmed once again by looking at the comparable magnitude of the 3rd HG and also the Kerr-induced effect of self-phase modulation causing the fundamental frequency broadening. In addition, the *ad-hoc* model includes the 7th HG, which can be again explained through three-wave mixing $7\omega = \omega + \omega + 5\omega$ or $7\omega = 3\omega + 3\omega + \omega$, and 9th HG. The increased amplitude of the 5th harmonic in *ad-hoc* compared to *ab-initio* is caused by our modelling of $\chi^{(3)}(5\omega_0) \equiv \chi^{(3)}(3.5\omega)$. The harmonics H > 4 are damped by a smooth sigmoid filter in the frequency domain[4].

---

[4] We apply the following sigmoid filter for positive frequencies (analogically for the negative frequencies with opposite argument):

$$\mathrm{f}(\omega) = 1 - \frac{1}{1 + \exp\left(-5(\omega - \omega_0)\right)/\omega_0}.$$

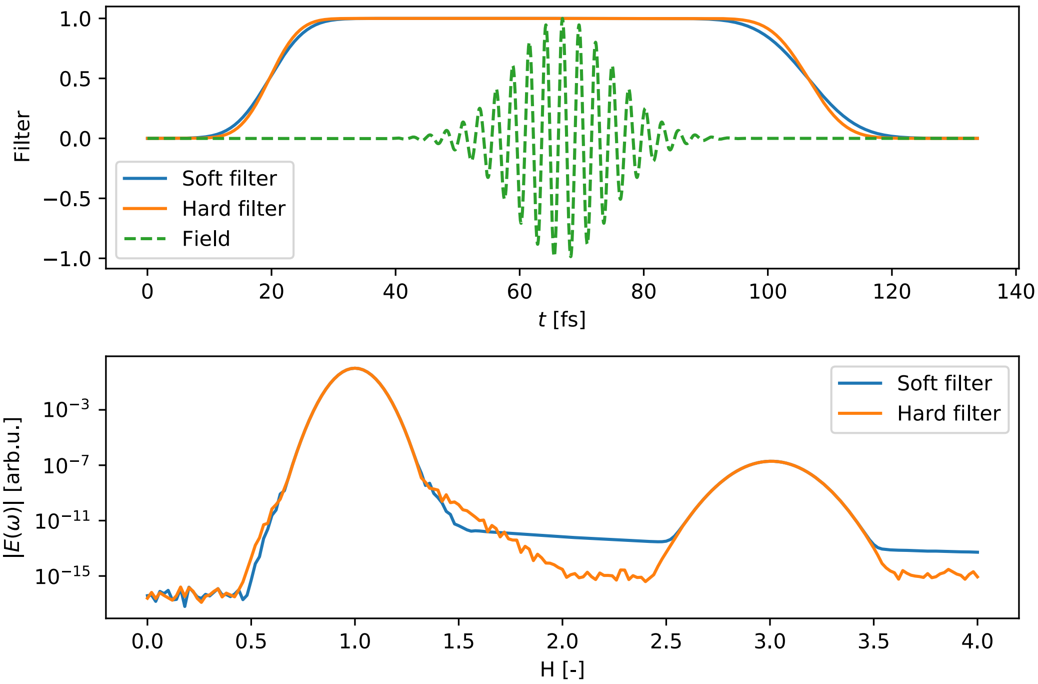

Figure 4.7: The comparison of ”soft” and ”hard” temporal filters is depicted in the top subfigure. The terms ”hard” and ”soft” relate to the steepness of the filter. A 5-cycle (FWHM) driving pulse is plotted for reference (green dashed curve). The result electric field spectra for the propagation of 800 nm 5-cycle beam of intensity $I = 10^{10}$ W/cm^2 in argon are plotted in the bottom subfigure. The propagation distance is $L \doteq 13\,\mu$m.

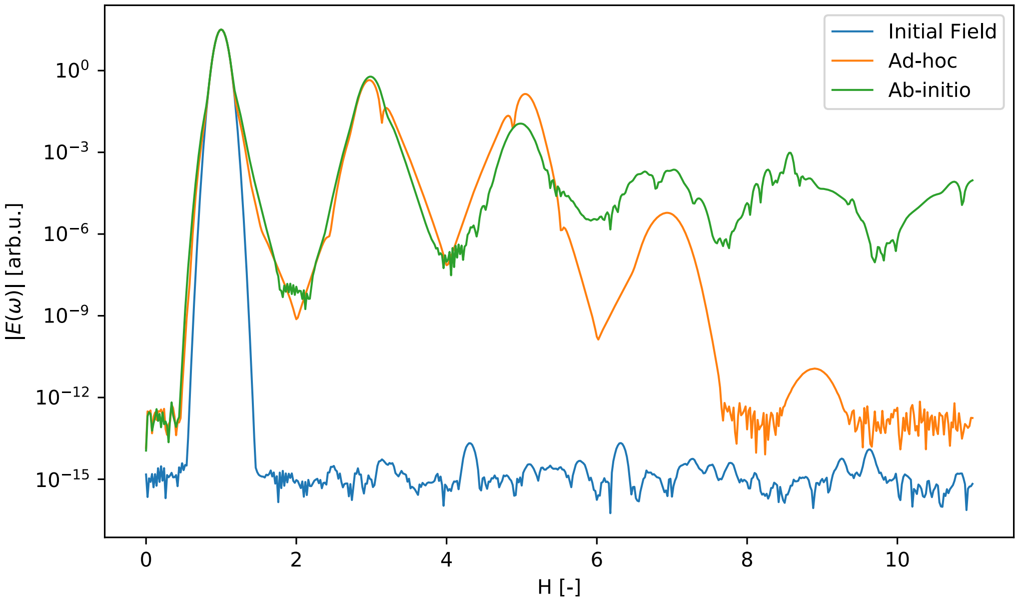

Figure 4.8: Comparison of *ad-hoc* and *ab-initio* electric field spectra after the propagation of 800 nm, 5-cycle (FWHM) beam of intensity $I = 10^{13}$ W/cm^2 through argon. The propagation distance is $L \doteq 2.6$ mm, resulting in a total of 1000 iterations of the RK4 Maxwell solver. The initial field spectrum is plotted for reference (blue curve).

We can look at the fully resolved spectra for the *ab-initio* model in Figure 4.9. We notice the generation of odd harmonics up to H = 21. The resonance at H = 5.7 likely interferes with the 5th harmonic for the first 0.5 mm of propagation distance. The narrow Rydberg state[5] located at H = 8.6 blends with the 9th harmonic but is clearly defined throughout the propagation. The origin of the periodic interference patterns in Figure 4.9 occurring in the spectrum for 9th harmonics up to 21st with a period of $\approx 150\mu$m remains unclear. The fact that we notice similar interference patterns with the same period also in Figure 4.6 in frequency range H = 6 − 8 drives us to a suspicion that there either might be some physical feedback loop for the harmonic generation or a numerical artifact in the Maxwell solver at play. The location of the filaments does not change by increasing the Maxwell solver propagation step. However, halving the gas density doubled the interference pattern period. The *ab-initio* simulations in an article by Berti et al. [63] investigated the harmonics generation at propagation distances up to 5 mm and their generated harmonics exhibit similar yield variations to our simulations, but with periods exceeding a few millimeters at the same pressure. Inspecting closely our results, Figure 4.9, the interference pattern sizes are smaller for higher harmonics. This concludes that the filaments may be caused by the harmonics interaction with each other during the propagation through complex phase matching – affected by the phase of the generated harmonic and nonlinear index [63]. A closer inspection by computing coherence length of q-th harmonic [8]

$$L_{\mathrm{coh},\,q} = \left|\frac{\pi}{\Delta k}\right| \quad , \quad \Delta k = qk_{\omega_0} - k_{q\cdot\omega_0} = \frac{q}{\lambda_0}\left(n(\omega_0) - n(q\omega_0)\right) \tag{4.1}$$

resulted for harmonic[6] H = 15 to value $L_{\mathrm{coh},\,15} \doteq 1$ mm which is roughly 6 times higher than the observed period $150\,\mu$m. Overlaying the gaps between the interference patterns in Figure 4.9 by function $1/q$ (white dashed line), where q is the harmonic order, indicates the reciprocal relationship of the coherence length L_{coh} on the harmonic order q. It has been also demonstrated that the spot size increases with decreasing atomic density. This indicates that the root cause of the interference patterns should be attributed to the coherence length, yet it is unclear why the theoretical value of L_{coh} does not quantitatively agree with the simulations.

4.2 Propagation of pulse at HHG intensity

We conclude the examples with the propagation of a 5-cycle FWHM pulse of intensity $I = 10^{14}$ W/cm^2 with the fundamental wavelength of 800 nm. These parameters already correspond to the HHG regime, thus we expect the plateau above the ionization potential of argon (H > 10) and the cut-off after H = 23, see Section 1.3. Because we are out of the applicability range for *ad-hoc* model, we present only the *ab-initio* results.

The propagation distance was set to $L \doteq 1.3$ mm. The electric field at the end of the propagation is depicted in Figure 4.10 and the corresponding HHG spectrum is shown in Figure 4.11. The attosecond pulse train generated in the process is in Figure 4.12. The pulse energy dissipated into the HHG during the propagation as the peak intensity of the fundamental field component dropped to $I_{\omega_0} = 1.539 \cdot 10^{13}$ W/cm^2. The spectrum indeed contains harmonic plateau from H = 10.2 up to cut-off frequency H = 23, as predicted by the three-step model [57, 20]. We notice also Rydberg states peaks in Figure 4.11. However, we see a local dip of the harmonic yield around the first resonance (H = 5.7). The

---

[5] See Fig. 4.1.

[6] The index $n(15\omega_0)$ was computed using the dipole method presented in Subsection 2.2.3.

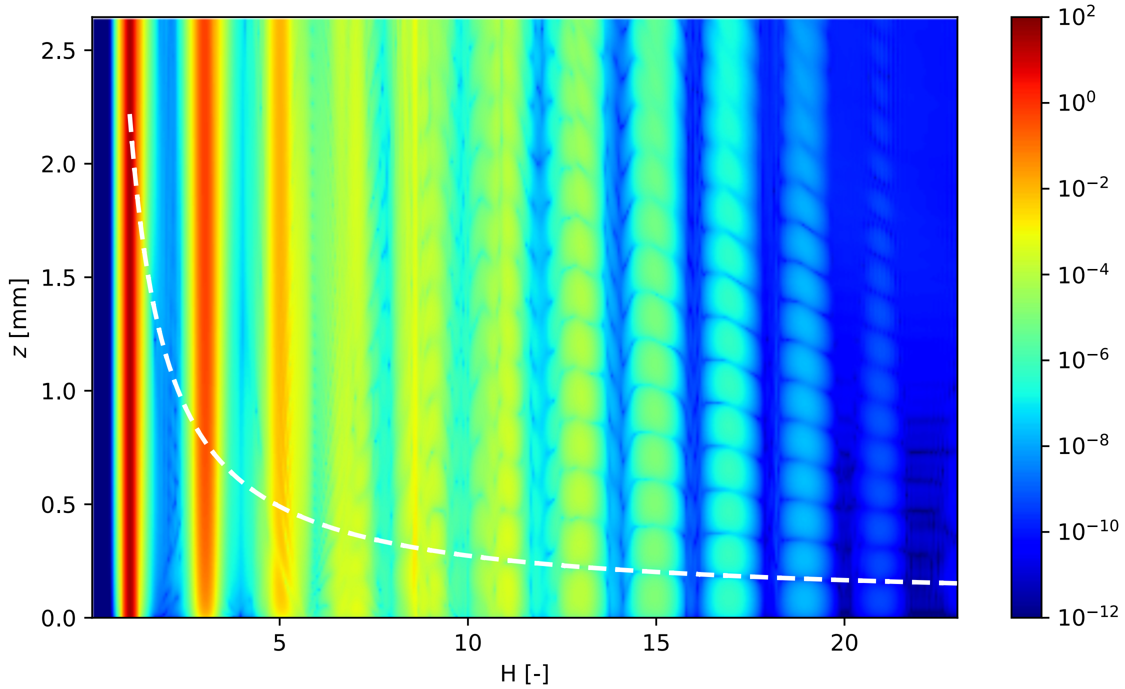

Figure 4.9: Resolved *ab-initio* electric field spectra for propagation of 800 nm, 5-cycle (FWHM) beam of peak intensity $I = 10^{13}$ W/cm^2 through argon. The propagation distance is $L \doteq 2.6$ mm. The interference spot gaps are overlaid by the function $1/q$ (white dashed curve), where q is the harmonic order.

3rd and 5th harmonics are well discernible but starting with the 5th harmonic up to $\mathrm{H} = 10$, the perturbation region is smeared out significantly due to interference. Spatially resolved harmonic spectra for the whole propagation are shown in Figure 4.13. The interference patterns can be spotted again in harmonics $\mathrm{H} < 11$ within the resolved spectra.

The *ab-initio* model demonstrates the capability to generate higher-order harmonics in macroscopic medium and retains the key features according to the theory [57, 20], Fig. 4.11. Moreover, the creation of an attosecond pulse train has been demonstrated in Figure 4.12. From the theory we would expect clearly defined peaks for odd harmonics but the plateau profile in Figure 4.11 appears ”chaotic”. This is attributed to the fact that we assume only the 1D geometry and examine the field in the near-field region, i.e. we take the field straight at the end of the propagation cell. In order to properly model the macroscopic HHG, we need to compute the coherent sum of individual fields from the 2D emitting plane in free space which filters out the incoherent contributions so that we obtain a much ”cleaner” harmonics profile[7] [81]. Thus the fully macroscopic, multi-scale HHG model requires, in addition, 2D propagation in medium and resolution in the far-field region where the individual emitters from the plane coherently interfere.

---

[7] See Fig. 1.2 in [112].

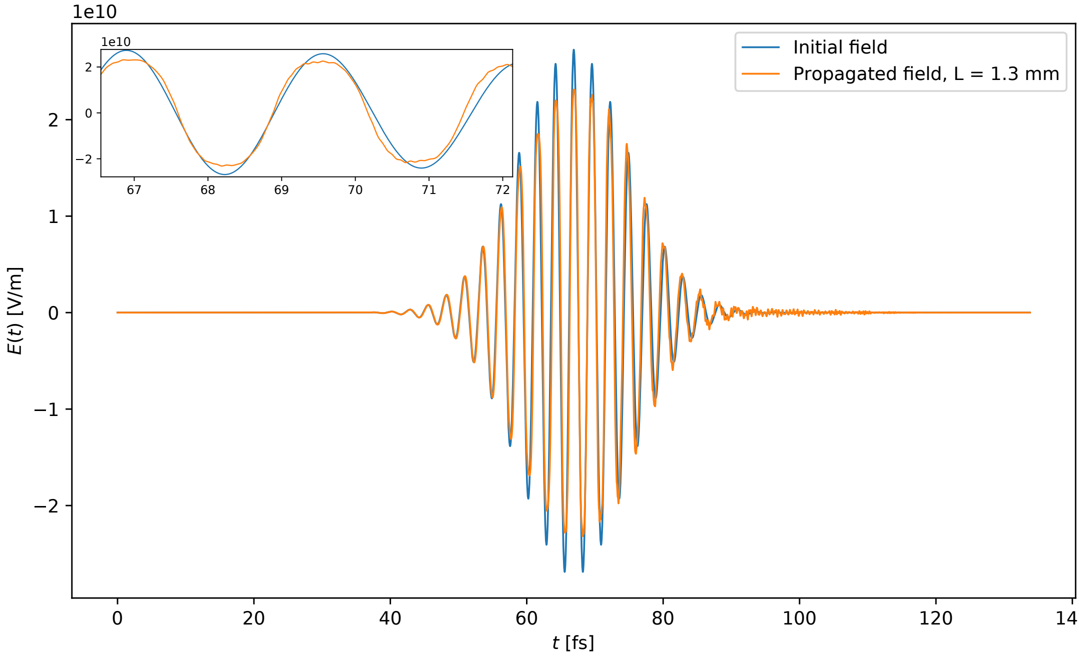

Figure 4.10: Electric field, corresponding to peak intensity $I = 10^{14}$ W/cm^2, computed using *ab-initio* model after the propagation of $L \doteq 1.3$ mm in argon. The detailed view of the fields near the pulse peak for 2 optical cycles is shown in the top left corner.

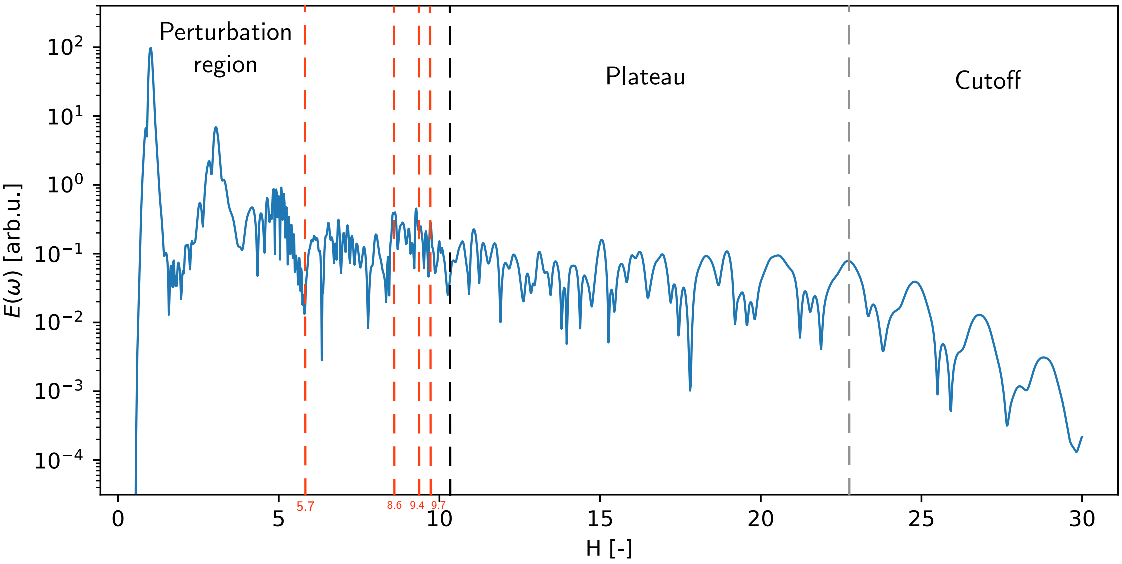

Figure 4.11: Electric field spectra, corresponding to peak intensity $I = 10^{14}$ W/cm^2, computed using *ab-initio* model after the propagation of $L \doteq 1.3$ mm in argon. Red vertical dashed lines mark the first excited state (H = 5.7) and the Rydberg states. Black and grey vertical dashed lines distinguish the three main regions of the HHG spectrum: perturbation region, plateau (H = 10.2 – 23) and cut-off.

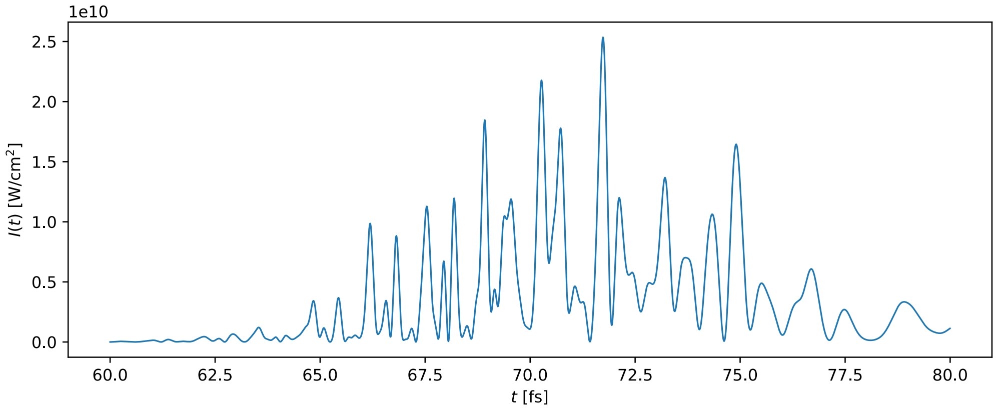

Figure 4.12: Attosecond pulse train generated after the propagation of $L \doteq 1.3$ mm in argon. The averaged intensity profile was obtained using bandpass filtering of the harmonic spectrum, Fig. 4.11, in range H = 10 − 25.

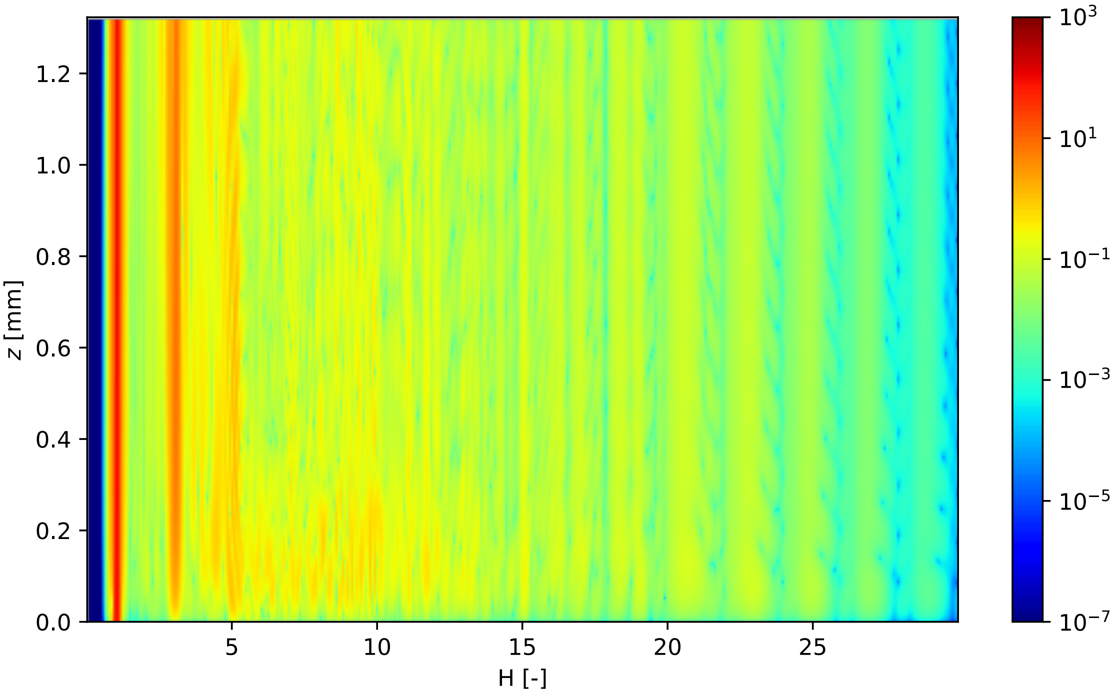

Figure 4.13: Resolved *ab-initio* electric field spectra for propagation of 800 nm, 5-cycle (FWHM) beam of peak intensity $I = 10^{14}$ W/cm^2 through argon. The propagation distance is $L \doteq 1.3$ mm.

4.3 Performance of the multi-scale model

We have already seen the qualitative comparison of the *ab-initio* and *ad-hoc* models results. We will now assess their performances. Since the task of the thesis is to optimize the computation of the source terms, the first step is to determine the bottlenecks of the algorithm through benchmarks.

The benchmarks of the models are displayed in Table 4.1. While the performance of the *ad-hoc* model remains the same for different intensities, the *ab-initio* performance depends on various factors. Maintaining the same temporal sampling of the pulse in the time domain, the higher peak intensity of the driver field requires a larger spatial grid for the proper convergence of the 1D-TDSE. Thus the critical bottleneck of the propagation is the evaluation of the 1D-TDSE, especially for the high-intensity regime (HHG intensities). We see that the *ad-hoc* model evaluation is almost 3 to 4 orders of magnitude faster than the *ab-initio* for the same range of intensities $10^{10} - 10^{13}$ W/cm^2.

Model / intensity	TDSE resolution	RK4 iterations (distance)	Time per TDSE [s]	Time total [h:m:s]
Ab-initio / 10^{10} W/cm^2	2000	1000 (2.6 mm)	10	11h 4m 25s
Ad-hoc / 10^{10} W/cm^2	-	1000 (2.6 mm)	-	68s
Ab-initio/ 10^{13} W/cm^2	25000	1000 (2.6 mm)	120	142h 57m 5s
Ad-hoc / 10^{13} W/cm^2	-	1000 (2.6 mm)	-	69s
Ab-initio / 10^{14} W/cm^2	40000	500 (1.3 mm)	200	118h 0m 33s

Table 4.1: Multi-scale model benchmarks for different models and intensities. One iteration of RK4 corresponds to 4 evaluations of the source term, respectively 4 1D-TDSEs. One spatial step $dz = 50000$ a.u. ($\approx 2.6\,\mu$m). The 1D-TDSE spatial grid resolution must be higher for high-intensity fields due to wavefunction reflection at the boundary. The temporal window resolution was set to 25601 points in total at 512 points per optical cycle ($\doteq$ 134 fs) in all cases. The benchmarks were obtained on an Intel Xeon Gold 5218 CPU.

Because the multi-scale model was restricted to 1D propagation and coupled with the Maxwell solver, we can not easily parallelize the computation of the source terms. The Maxwell solver (Algorithm 2) is serial because of the evaluation of different k-coefficients using the preceding field and source term. The C implementation of FFT in FFTW3 [80], computed 8 times per RK4 iteration, is already very efficient even for arrays other than factor 2 sizes. Any collective multi-core operation might not provide reasonable gain due to the communication bottleneck. The same argument applies to the 1D-TDSE. Significant optimization of the grid 1D-TDSE might be the implementation of a suitable wavefunction absorber, e.g. ECS [78].

In the end, the aim is to find a computationally inexpensive model that utilizes the speed of the *ad-hoc* model with the accuracy of *ab-initio.* The model should couple the Maxwell solver with accurate source term computation. A promising option is to approximate the 1D-TDSE source terms using an artificial neural network. We will examine the technique in the next chapter.

4.4 Extension of the multi-scale model to higher dimensions

One-dimensional propagation code is not physically accurate in free space and in cases of strong focusing induced by high-power laser beams. The transverse effects become significant at short distances, therefore the transverse Laplacian must be included. In macroscopic HHG, many features are directly linked with transverse effects and the properties of the driving laser beam. One example is achieving optics-less focusing of HHG in the far-field region by adjusting the focus of the incoming beam [82]. In the far-field, controlling the wavefront of a high-harmonic beam can result in optics-free spectral filtering of individual harmonics [83]. Thus, as outlined at the end of Section 4.2, we need to include two additional components for the truly macroscopic model: transverse effects and far-field.

Far-field contribution can be computed by evaluation of the diffraction integral [112] or under the thin target approximation[8] using the Hankel transform [81]. As an example of a multi-dimensional code with transverse effects, we can mention CUPRAD [113]. CUPRAD is a two-dimensional code employing the radial symmetry of the medium based on the UPPE in the co-moving frame with a transverse field component. CUPRAD has been successfully coupled with the 1D-TDSE code for the study of HHG in long media [112, 18]. The mechanism of coupling is different in CUPRAD since the propagated fields are not influenced directly by the TDSE source terms during the propagation. Instead, an ad-hoc description was used to compute the fields. All in all, the approach with the CUPRAD-TDSE coupling agrees well with the experimental observations [18].

In principle, there is no obstacle to coupling 1D- or 3D-TDSE with a 1D propagation scheme. Experimentally we could send a prepulse for polarising the atoms in the medium in conjunction with the driving pulse field polarisation. However, for advanced problems such as the study of optical gating, a technique in HHG for the generation of attosecond impulses [12], or exploring the contribution of individual orbitals for HHG we would benefit from a fully vectorial 3D description of fields coupled with a 3D-TDSE solver [21, 22, 84].

---

[8] Gas cell smaller than the Rayleigh length of the driving beam.

Chapter 5

Multi-scale model optimization using neural networks

The usage of 1D-TDSE for the computation of the source terms has been shown in the previous chapter as the critical bottleneck in the multi-scale model. In this thesis, we set a goal of developing a neural network (NN) that takes numerical electric fields and outputs the corresponding source terms mimicking the result of the exact TDSE computation. In Section 5.1 we provide an overview of NN principles necessary for understanding the design of a custom network for this thesis. In Section 5.2 we propose a data pre-processing pipeline and establish the design of a proprietary NN for computing the 1D-TDSE source terms. The results are showcased in Section 5.3 and we discuss the outlook of using NNs in Section 5.4. Since the NN research has been rapidly evolving in recent decades, it is useful to introduce first some historical context with examples of neural network applications. The general overview will motivate the selection of appropriate network architecture for this thesis.

The task of creating artificial intelligence (AI) has marveled the minds of scientists since the conception of Turing machines. AI is a broad term for a branch of research encompassing the conception and design of machines and algorithms capable of mimicking intelligent behavior. They vary from the simplest decision trees or Bayesian statistics [13] to Boltzmann machines [85] and artificial neural networks [14]. The idea of artificial neural networks arose from the area of neuroscience. The first concept of non-learning NN mimicking brain function was proposed in 1943 by McCulloch and Pitts [86]. Since then, many models have advanced on the concept and the term machine learning (ML) was established for the perceptron in 1958 [87]. ML is a class of statistical and computational methods responsible for training NNs. Training can be supervised – the network learns the relationships between inputs and outputs on correctly labeled data, or unsupervised – the network is trained on unlabelled data. Notable mentions of unsupervised training are generative adversarial networks for image generation [88] or anomaly detection networks [89], which are based on the encoder-decoder architecture [90]. Due to their variety and large application potential, NNs have gained significant attention in recent years. With the advent of so-called large language models (LLMs) such as GPT-4 by OpenAI [23] or Google PaLM [24] and more recent multi-modal Gemini model [116], the NN models have become capable of interpreting human-written prompts and offering solutions to abstract problems. Powered by OpenAI's DALL-E 3 [117] image generator, GPT-4 is also capable of image generation from a user prompt.

In physics, NNs have been employed in numerous ways. The ability of NNs to quickly process vast amounts of data found its use in the domain of experimental physics where the measured data flux is overwhelming. Instead of using difficult decision trees, where the programmer must set the ranges manually, the ML algorithm can learn from the curated experimental data to provide superior analysis. NNs have been successfully applied for distinguishing unusual signals hinting new particle observations in CERN experiments [26]. NNs also drive the inertial confinement fusion research by exploiting its optimization capabilities for designing optimal fusion capsules or predicting neutron flux, both increasing the energy gain [25]. NNs have found use in atomic and molecular science for the search of the ground state of many-electron wavefunction from the notoriously difficult time-independent Schrödinger equation [29, 30], opening the opportunities for solving large-scale ab-initio problems of atoms and molecules with more efficiency. Some works were devoted to providing directly a numerical solution of the TDSE by propagating the wavefunction for arbitrary time-dependent potential using NN [31]. Last but not least, the physics applications were expanded by the introduction of so-called physics-informed neural networks (PINNs) [27, 28]. The concept of PINN has been demonstrated on various physical problems [27, 28], a notable example is the NSFnets model for solving incompressible Navier-Stokes equations [91].

Since we hadn't had much previous experience with NNs, we started with a simple design to understand the limitations of employing NNs for our task. The design of our purpose-built NN has led to many trial-and-error situations concerning data preprocessing and the NN design architecture. Some points are discussed in greater detail in Appendix B. Multiple NN architectures were assessed. The most straightforward is the multilayer perceptron architecture (MLP), however, the feature extraction of MLP is not very sophisticated. The ground-breaking LeNet architecture [92] established convolution neural networks (CNN) fused with MLP network and showcased better feature extraction en par with higher predictive accuracy. The design was later polished in AlexNet [93], one of the most influential computer vision networks. Thus, the choice of NN architecture in the thesis shifted towards CNN architecture since we wanted to extract information from contiguous data.

5.1 Fundamentals of neural networks

The general prospect of CNN follows a straightforward pipeline. Given a set of inputs $\boldsymbol{x}$, also called the **features**, a supervised learning algorithm propagates the input through a series of convolution filters into a network of interconnected layers of nodes, also called artificial neurons. As the input data propagates through the network, nonlinearities are applied and important data features are extracted similarly as organic brains process information. The output $\boldsymbol{y}$, also called the **label**, is inferred from the network after the propagation. The accuracy of the prediction is assessed using an appropriate metric by comparing the NN output with the correctly labeled data, i.e. the desired outcome. Finally, the ML algorithm ensures the network iteratively adjusts its parameters to improve the network. Let us discuss in detail the inner structure of a NN.

5.1.1 Perceptron

The smallest unit of NN is the perceptron. Perceptron resembles a neuron cell – it takes multiple signals, each with attributed **weight $\boldsymbol{w}$** (strength of the connection) and **bias**[1] b, and processes all incoming signals into a single output value. A single perceptron is sketched in Figure 5.1.

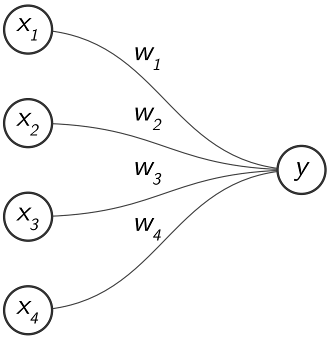

Figure 5.1: Single perceptron with 4 inputs, each input has a weight attributed to it. Image generated using NN-SVG tool [94], edited.

We can mathematically describe the action of perceptron as follows: for a vector of inputs $\boldsymbol{x}$ and weights $\boldsymbol{w}$ and bias b, the perceptron applies an operation f on a weighted average of the inputs with bias[2]. The output is then

$$y(\boldsymbol{x}) = f\left(\sum_i w_i \cdot x_i + b\right). \tag{5.1}$$

We call the scalar function $f : \mathbb{R} \to \mathbb{R}$ the **activation function**. Activation may be arbitrary and should have a simple evaluation and derivative, however over the years the selection has crystallized into a few most used activations. The most prevalent ones are[3]

$$\mathrm{Id}(x) := x \qquad \mathrm{Range}(\mathrm{Id}) \in \mathbb{R}, \tag{5.2}$$

$$\mathrm{ReLU}(x) := \begin{cases} x \dots x > 0 \\ 0 \dots x \leq 0 \end{cases} \qquad \mathrm{Range}(\mathrm{ReLU}) \in \mathbb{R}_0^+, \tag{5.3}$$

$$\mathrm{sigmoid}(x) := \frac{1}{1+e^{-x}} \qquad \mathrm{Range}(\mathrm{sigmoid}) \in (0,1), \tag{5.4}$$

$$\tanh(x) := \frac{e^x - e^{-x}}{e^x + e^{-x}} \qquad \mathrm{Range}(\tanh) \in (-1,1). \tag{5.5}$$

The activations are depicted in Figure 5.2. If we used only an identity activation $f = \mathrm{Id}$ instead of a nonlinear activation, the expression (5.1) would yield a linear regression equation. This can be sufficient as long as the relationship between $\boldsymbol{x}$ and y is indeed linear, but this is not generally the case. That is why we have to introduce nonlinearities in the form of nonlinear activation functions. In statistics, this class of problems is generally tackled using a method of least squares, we just have to assume the type of relationship between $\boldsymbol{x}$ and y. The power of NNs is that we can actually keep the linear argument and

---

[1] The bias adjusts the output of perceptron to fit the training data accurately. In many cases, training only on the input data would yield incorrect results. The key role of bias is to provide and adjust a starting point, independent of the input data.

[2] In mathematical terms this is also called the affine mapping.

[3] ReLU stands for **Re**ctified **L**inear **U**nit.

opt for a nonlinear activation for nonlinear problems, thereby we can blindly retrieve the relationship without assuming a particular function to fit our data. The downside is that we can easily overfit the data.

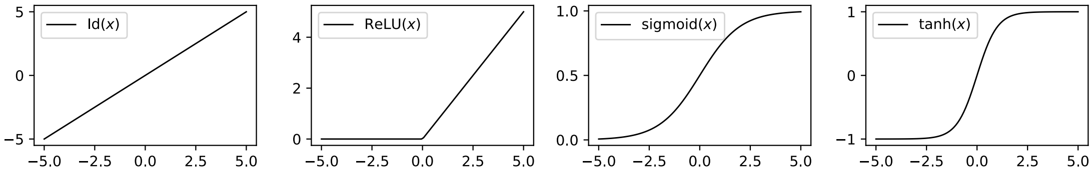

Figure 5.2: Activation functions from Eqs. (5.2)-(5.5).

A single perceptron is only suitable for binary classification problems, say, for a set of features $\boldsymbol{x}$, the perceptron can only yield a single binary outcome. For example, *true* or *false* classification of an object given a set of features $\boldsymbol{x}$. That is why, in practice, many perceptrons are stacked in a single layer and we pile many layers consecutively into a multilayer perceptron network.

5.1.2 Multilayer perceptron network

The multilayer perceptron (MLP) network incorporates multiple layers of perceptrons to increase the network's predicting capabilities. Apart from the possibility of implementing more classifiers or detailed regression analysis, the depth of the network also improves the prediction power by identifying stronger relationships between the data fed through. However, with the increasing size of the MLP comes the greater computational cost of training the network. A simple MLP network is depicted in Figure 5.3.

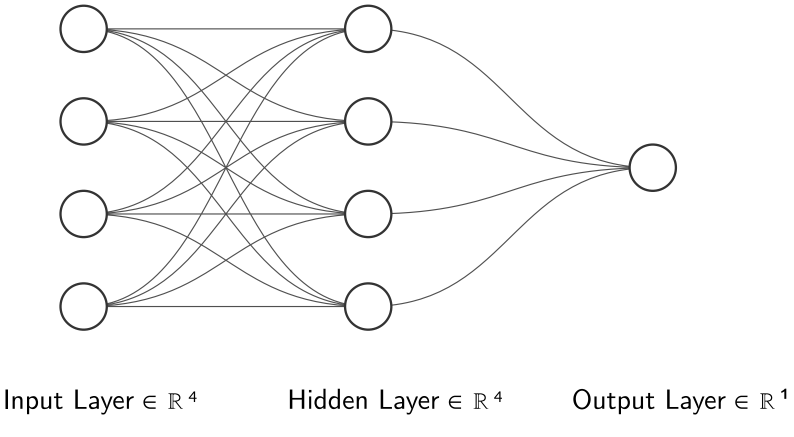

Figure 5.3: MLP network composed of an input layer with 4 features, 1 hidden layer with 4 perceptrons, and an output layer with a single perceptron. Image generated using NN-SVG tool [94].

We improved the perceptron architecture by adding a single hidden layer $\boldsymbol{h}$. This hidden layer is composed of 4 independent perceptrons. The outputs h_i are then fed into the output layer o. Mathematically, we can describe the transfer of features as follows[4]

$$h_i(\boldsymbol{x}) = f\left(\sum_j w_{ij}^{\text{hidden}} \cdot x_j + b_i^{\text{hidden}}\right), \tag{5.6}$$

$$o(\boldsymbol{x}) = (f \circ \boldsymbol{h})(\boldsymbol{x}) = f\left(\sum_i w_i^{\text{output}} \cdot h_i(\boldsymbol{x}) + b^{\text{output}}\right), \tag{5.7}$$

where f is arbitrary activation function from (5.2)-(5.5), $w_{ij} \in \mathbb{R}^{4\times 4}$ is a matrix of weights and $\boldsymbol{b}$ a vector of biases. There are in total $N_{\text{input}} \times N_{\text{hidden}} + N_{\text{hidden}} \times N_{\text{output}} + \text{size}(\boldsymbol{b}^{\text{hidden}}) + 1 = 16 + 4 + 4 + 1 = 25$ adjustable parameters.

The output o can also have an arbitrary size. This is useful if we want to create a model classifying more than two outcomes. Another possibility is to approximate an arbitrary function $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$. Having a set of inputs $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N | \boldsymbol{x}_i \in \mathbb{R}^n\}$ and a known set of corresponding outputs $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N | \boldsymbol{y}_i \in \mathbb{R}^m\}$, we can design and train a neural network with an input layer $\in \mathbb{R}^n$, an output layer $\in \mathbb{R}^m$ and an arbitrary number of hidden layers to obtain a function $\boldsymbol{g}$ approximator. Feedforward NNs have been shown and proven to be universal approximators of any Borel measurable function from finite-dimensional spaces to another with arbitrary accuracy, provided the network is reasonably designed. This is summarized as *universal approximation theorem* in NNs [33]. Remark that the universal approximation theorem generally does not apply to extrapolation.

5.1.3 NN learning and evaluation

To train a NN, we generally provide **training** and **validation** sets. The NN is first trained on the training set and the success of training is evaluated on the validation set. In essence, the neural network needs to adjust the weights of individual connections and biases according to the expected correct output through the process of supervised learning.

Let $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N | \boldsymbol{x}_i \in \mathbb{R}^n\}_{\text{train}}$ be the training set of features and $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N | \boldsymbol{y}_i \in \mathbb{R}^m\}_{\text{train}}$ be the training set of correct labels. Say the map $\mathcal{N} : (\mathbb{R}^n | \boldsymbol{w}, \boldsymbol{b}) \to \mathbb{R}^m$ maps the input of the MLP network into predicted labels $\boldsymbol{y}_i^{\text{pred}}$, which depend on trained parameters $(\boldsymbol{w}, \boldsymbol{b})$. We call this part the *forward propagation.*

To quantify the error between the predicted and correct labels $\boldsymbol{y}_i^{\text{true}}$, we need to establish a loss function $\mathcal{L}(\boldsymbol{x}|\boldsymbol{w}, \boldsymbol{b})$. We can evaluate the error with mean squared error (MSE) loss function[5]

$$\mathcal{L}(\boldsymbol{x}|\boldsymbol{w}, \boldsymbol{b}) := \text{MSE}(\boldsymbol{x}) = \frac{1}{m}\sum_{i=1}^{m}\left(y_i^{\text{pred}} - y_i^{\text{true}}\right)^2. \tag{5.8}$$

The learning algorithm tries to minimize MSE (5.8) between predicted and correct labels using the *backpropagation* algorithm. The backpropagation is based on the evaluation of the derivative of MSE with respect to the network parameters $(\boldsymbol{w}, \boldsymbol{b})$:

$$\nabla_{(\boldsymbol{w}, \boldsymbol{b})}\mathcal{L} = 0. \tag{5.9}$$

---

[4] Remark that for nonlinear problems we need to add nonlinear activations into the hidden layer. Composing the layers with identity activation $\text{Id}(\boldsymbol{x})$ results in multiple affine mappings which yields another affine mapping.

[5] A list of other loss functions can be found in [14]. MSE is suitable for function approximators. The categorical cross entropy loss function is an example of loss suitable for classification problems.

Because the map $\mathcal{N}$ comprises many layers of neurons with different weights, the chain rule for derivatives must be applied. Say for the simple example of MLP network in Figure 5.3, the chain rule w.r.t. single weight w_{ij} and bias b_i from the hidden layer looks as follows:

$$\partial_{w_{ij}}\mathcal{L}(\boldsymbol{x}) = (\nabla\mathrm{MSE}\cdot\nabla o\cdot\nabla\boldsymbol{h}\cdot x_j)(\boldsymbol{x}), \tag{5.10}$$

$$\partial_{b_i}\mathcal{L}(\boldsymbol{x}) = (\nabla\mathrm{MSE}\cdot\nabla o\cdot\nabla\boldsymbol{h})(\boldsymbol{x}), \tag{5.11}$$

where x_j is the input value corresponding to the weight w_{ij} in the hidden layer. Remark that the most numerically expensive part of the chain rule is the evaluation of derivatives of activations. The rest of the evaluation is just plain tensor multiplication.

To obtain a solution to (5.9), we employ iterative gradient methods for finding the minima of functions. The common algorithm is gradient descent (GD) for which we set an initial $w^{(0)}$ (typically random) and adjust it according to the steepest descent towards the optimum[6] as:

$$\boldsymbol{w}^{(1)} := \boldsymbol{w}^{(0)} - \gamma\frac{1}{N}\sum_{i}^{N}\nabla_{\boldsymbol{w}}\mathcal{L}(\boldsymbol{x}_i), \tag{5.12}$$

where γ is a parameter called the **learning rate**. The simplest GD has, however, a poor performance because of the computation of the derivative for each input $\boldsymbol{x}_i$ at once before updating w. Therefore more sophisticated optimization algorithms are used. We may instead approximate the descent $\nabla_w\mathcal{L}(\boldsymbol{x})$ with an average over a randomly selected mini-batch of M $(M < N)$ training examples $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ as:

$$\frac{1}{M}\sum_{i}^{M}\nabla_{\boldsymbol{w}}\mathcal{L}(\boldsymbol{x}_i) =: \boldsymbol{g} \tag{5.13}$$

and then use this quantity to update the weight w for one iteration as

$$\boldsymbol{w}^{(1)} := \boldsymbol{w}^{(0)} - \gamma\boldsymbol{g}. \tag{5.14}$$

The quantity M is called the **batch size** and plays a key role in the converging properties. Typically, the batch size is set to 32, 64 or 128. We call this algorithm the mini-batch stochastic gradient descent (SGD). The advantage of using the mini-batch SGD is that it updates the weights after just M derivatives instead of computing the derivatives for the whole dataset. Another advantage is that mini-batch computation can exploit CPU/GPU vectorization, further increasing the speed.

The backpropagation algorithm hence updates the weights after selecting M random samples according to (5.14), then selects another M samples from the remaining dataset until it sweeps through the entire dataset, this is called the **epoch**. The network $\mathcal{N}$ is trained for multiple epochs until the desired convergence is reached.

Note that mini-batch SGD is succeeded by various more advanced optimizers employing various techniques for reaching better convergence. Some of these techniques involve the implementation of momentum vector [95] (or multiple vectors) slightly shifting the weight at each update. Other optimizers also implement adaptive learning rate [96]. This can be schematically summarized as

$$\boldsymbol{w}^{(1)} := \boldsymbol{w}^{(0)} + \beta\boldsymbol{v} - \gamma(t)\boldsymbol{g}, \tag{5.15}$$

---

[6] Remark that the convergence towards the global optimum is generally not assured. Reaching global optimum requires, aside from the robust optimizer, also suitable network architecture and representative training data with a large variety.

where $\boldsymbol{v}$ is the momentum vector, β is the momentum vector parameter and $\gamma(t)$ denotes the adaptive learning rate. Parameters β and γ are set in the beginning to some default value and $\boldsymbol{v}$ and $\gamma(t)$ are algorithmically estimated from the loss and the initial parameters, depending on the specific implementation. Both momentum and adaptive steps are implemented within Adam optimizer [96], currently one of the most used optimizers for NN training.

The number of epochs, batch size and initial learning rate are one of many tuning parameters we call the **hyperparameters**[7]. The question naturally arises of how to optimally choose the hyperparameters and tune the network size or layers to obtain the best fit for our input data. The task of every NN design is to minimize the test error and also the generalization error on previously unseen data $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N | \boldsymbol{x}_i \in \mathbb{R}^n\}_{\text{test}}$, $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_N | \boldsymbol{y}_i \in \mathbb{R}^m\}_{\text{test}}$ (validation dataset), chosen from the same data distribution as the training set. During the testing of the NN, the weights are no longer updated and we call this the **inference phase**. The error during the training phase and the inference phase is quantified using the loss function and **metric**. For vectors of real-valued outputs $\boldsymbol{y} \in \mathbb{R}^m$, the mean absolute percentage error metric is defined as:

$$\text{MAPE} = \frac{1}{N} \sum_{i}^{N} \left| \frac{\boldsymbol{y}_i^{\text{true}} - \boldsymbol{y}_i^{\text{pred}}}{\boldsymbol{y}_i^{\text{true}}} \right| . \tag{5.16}$$

We will employ MAPE for the assessment of the custom NN performance.

5.1.4 NN regularization

The challenge of each NN is to find a good balance between *underfitting*, the model is too weak or has bad performance, and *overfitting*, the model is too strong and does not generalize. Underfitting may be fixed by increasing the number of epochs or increasing the model *capacity* – what the model is able to represent at the given size and what is it capable of learning. The capacity practically translates to the size of the network in terms of hidden layers and the number of trainable parameters. Overfitting increases generalization errors and is generally tackled using a process called **regularization**. The dependence of the error on the model capacity is plotted in Figure 5.4.

There are various ways how to regularize NN. The easiest to implement into the model and very robust is the usage of a **dropout** layer. Given a rational fraction, say $p = 0.5$, then 50% of connections between layers are randomly off and the remaining 50% of features actually propagate to another layer. **Batch normalization** is a powerful technique that normalizes the passing batch of data by subtracting the batch mean and dividing by its variance. The normalized output is then scaled and shifted by trainable parameters. This significantly stabilizes the training process as it can mitigate any internal covariate shifts within the network. Another regularization technique involves adding random **gaussian noise** into the data after each pass. **Early stopping** the learning process after fewer epochs may also decrease the generalization error. Artificial augmentation of the training dataset (noise addition, smarter feature extraction) or increasing the number of test examples can improve the overall performance too.

The network predictive capacity can be further improved by the addition of skipped connections. Skipped connection is a powerful tool first implemented in ResNet architecture [97]. The deeper the network, the harder it is to train due to the *vanishing gradient*

---

[7] Not to confuse with weights and biases of the NN which are the model parameters.

problem, i.e. the gradients become so small during the backpropagation that the performance of the network decreases. A skipped connection is sketched in Figure 5.5. The key component of skipped connection is the **add layer** that performs elementwise addition of the first input layer with the output of the second hidden layer. The add layer does not have any trainable parameters but the data can be fed to another hidden layer. A skipped connection serves as a "neural network highway" in the sense that some features propagate faster through the network. This enables to increase dramatically the number of layers while gaining further performance and capacity [97].

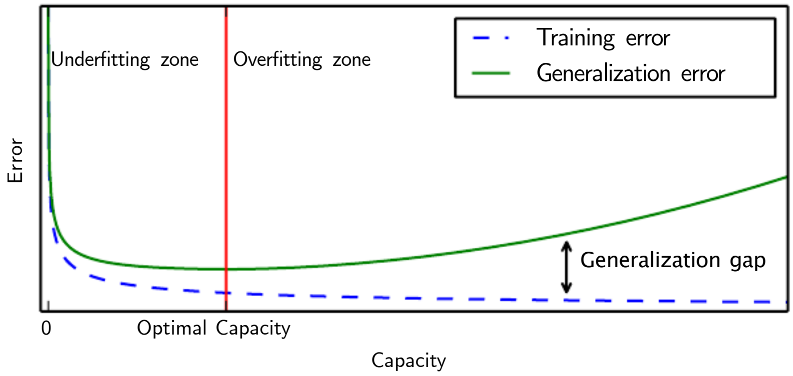

Figure 5.4: Relationship between model capacity and error. The tedious task is to obtain the "sweet spot" between underfitting and overfitting. The generalization gap increases for models with a higher number of parameters[8]. We also notice that the decreasing training error (loss error) does not always provide the decisive measure for the performance of the model. Obtained from [14], edited.

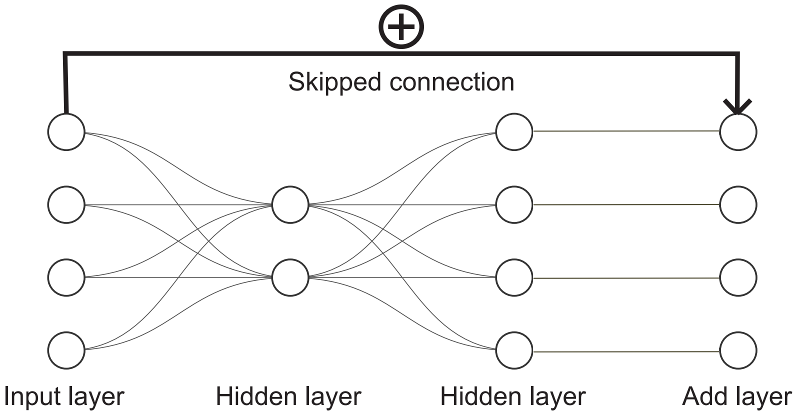

Figure 5.5: A sketch of a simple skipped connection. Image generated using NN-SVG tool [94], edited.

5.1.5 Convolution neural network

Convolution neural network (CNN) tackles the problem of too many input parameters by first extracting the significant features from the input data using convolution. One

---

[8] This claim does not apply generally. As we increase the model size above a certain threshold, the performance suddenly improves, see Figure 1 from [98]. This feature is called "double-descent" and is the underlying reason why modern ANNs have become so large with millions of parameters. The common knowledge within the NN community is that larger networks generally perform better [98].

dimensional convolution of an input vector $\boldsymbol{x}$ of size N and convolution kernel K of size 3 yields[9]

$$(\boldsymbol{x} * \mathbf{K})_i = \sum_{j=1}^{3} K_j \cdot x_{i+j-2} \quad , \quad i = 2, \ldots, N-1 \tag{5.17}$$

The kernel contains particular predefined weights for optimum feature extraction. Moreover, we can also set an output dimensionality in the same way as in the MLP. Hence, for each i-th convolution we have multiple **output channels** c so the total output of a single 1D convolution layer would be a matrix of size $(N-2, c)$. Sometimes the activation function is added just after the result of the convolution. The output matrix is the same as if without an activation but every matrix element has been fed through the activation. Now we need to introduce **pooling layer** to decrease the size of the convolution output. For example, the maximum pooling layer sweeps through the convolution and finds the maximum of 3 consecutive points in the output that is further propagated in the network. The output matrix size is then $((N-2)/3, c)$. We may also implement an average pooling layer that computes the average of 3 consecutive points instead.

We usually stack many convolution and pooling layers to reduce the final output size. Finally, the MLP is added at the end, however, the output tensor must be flattened first. The flattening layer joins the rows of the tensor together into a single 1-dimensional vector, that can be further processed in the MLP. The whole 1D-CNN is sketched in Figure 5.6.

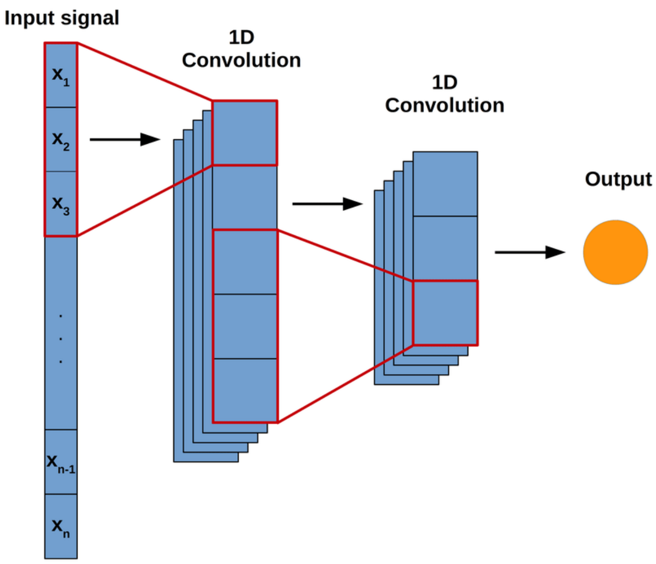

Figure 5.6: One-dimensional CNN with an input layer, convolution layer with a kernel of size 3 with 5 output channels, pooling layer, and after the flattening layer comes finally MLP with a single output. Obtained from [99].

The advantages of using CNN are namely the reduction of the input parameters (mainly for image processing networks) and better local feature extraction due to accounting for local interactions. It has been also shown that the CNNs can have a regularizing effect [100].

---

[9] If the vector $\boldsymbol{x}$ has size N, then the output convolved vector has size $N-2$.

5.1.6 Overview of the NN design

We can now summarize the key aspects of the network and supplement them with an example code from Tensorflow [118]. From this moment we refer to vectors, matrices and higher-order tensors simply as tensors of particular shape.

Example NN

We start with loading the corresponding modules from Tensorflow and defining a sequential model simply called "model". The sequential model builds chronologically the custom NN as we add the layers in their respective order. We define the input layer of the model by taking the input 1D-tensor $\boldsymbol{x}$ of shape (1, 100):

```
# Language: Python
from tensorflow.keras.layers import *
from tensorflow.keras import Sequential
# Define model
model = Sequential()
# Define input layer
input_layer = InputLayer((1,100))
# Add layer to model
model.add(input_layer)
```

which is followed by adding two convolution layers with kernel sizes 3 and 4, respectively 16 output channels (filters). In between we define the maximum pooling layers with the kernel of size 3.

```
convolution1 = Conv1D(filters=4, kernel_size=3)
convolution2 = Conv1D(filters=16, kernel_size=3)
max_pooling1 = MaxPooling1D(pool_size=3)
max_pooling2 = MaxPooling1D(pool_size=3)

model.add(convolution1)
model.add(max_pooling1)
model.add(convolution2)
model.add(max_pooling2)
```

The Tensorflow API adjusts the kernels to fit the particular shape of the input and the desired output. Thus the kernel of the first convolution layer is set to the tensor of shape (3, 4). The weights of the convolution kernel are set automatically. The result of the first convolution is a tensor of shape (98, 4) which is then reduced within the first pooling layer to size (32, 4).

Once we reach the end of the CNN pipeline (Fig. 5.6), we implement the flattening layer to convert the final 2D tensor into a 1D tensor to be fed into the MLP network, see Figure 5.3. We implement two consecutive hidden dense layers with 20, respectively 40 perceptrons (Fig. 5.1).

```
flatten = Flatten()
hidden_dense1 = Dense(units=20, activation="tanh")
hidden_dense2 = Dense(units=40, activation="tanh")
dropout = Dropout(rate=0.5)

model.add(flatten)
model.add(hidden_dense1)
model.add(dropout)
model.add(hidden_dense2)
```

Each of the 40 perceptrons in the second dense layer has 20 inputs from the preceding dense layer and 21 trainable parameters (20 weights and 1 bias), see Figure 5.3. The same logic applies to the first dense layer. The $\tanh(x)$ activation function has been selected, Eq (5.5). A dropout layer with a 50% dropout rate has been added between the dense layers for regularizing the network. Alternative regularizers are Gaussian noise or batch normalization layers.

Finally, the output layer of size 100 with identity activation is implemented

```
output = Dense(100)

model.add(output)
```

which summarizes our simple example model. We can now configure the model for training by specifying the loss (MSE, Eq. (5.8)), metrics (MAPE, Eq. (5.16)) and optimizer (Adam [96]) as follows[10]

```
model.compile(optimizer='Adam', loss='MSE', metrics='MAPE')
```

The training datasets $(\boldsymbol{x}_{\text{train}}, \boldsymbol{y}_{\text{train}})$ respectively validation datasets $(\boldsymbol{x}_{\text{test}}, \boldsymbol{y}_{\text{test}})$ are supplied as tensors of shape $(N_{\text{train}}, 100)$ respectively $(N_{\text{test}}, 100)$. We fit the model with batch size equal to 64 on the training data for 20 epochs and validate afterward:

```
# Training the network
model.fit(x_train, y_train, batch_size=64, epochs=20)
# Validating the network
model.evaluate(x_test, y_test)
```

---

[10] See the exact API documentation for properly defining the method attributes.

5.2 NN data pipeline and design for TDSE output estimation

The data pipeline for feeding the NN is presented in this section. The network is fed by the numerical electric fields and predicts the 1D-TDSE source term $\langle -\nabla V \rangle - E$. The datasets containing fields and source terms were acquired using the *ab-initio* multi-scale model. During the propagation of the multi-scale code, every RK4 step was stored to increase the dataset size by a factor of 4, since the 1D-TDSE is evaluated 4 times for each RK4 iteration. The input data needs to be preprocessed first before it can be fed to the network. The output of the network must be established and the nature of the input and output data has to be reflected in the overall design of the network. We will now go through the individual steps.

5.2.1 Data preprocessing

Data reduction – low pass filter

A single output source term from the *ab-initio* model comprises 25601 bins of data – given by the temporal sampling of the 1D-TDSE solver, see Subsection 3.1.4. If we decided to plug in all of the data points into a network with a hidden layer of size 1024, this would yield at least 10^7 trainable parameters. This scale is both practically and computationally very inefficient. However, we can exploit the knowledge of the structure of the source term and the driving electric field to drastically reduce the size of the input layer.

From the physics models, it is well known what range of frequencies is generated for certain field intensities. If we are focused on HHG, for example, we can neglect the frequencies above the cutoff (for argon at intensity 10^{14} W/cm^2 $\mathrm{H}_{\mathrm{cutoff}} \approx 23$). For lower-intensity fields where the ionization has not become significant, we can set the frequency range even smaller. Hence the first component of the data preparation pipeline is a low-pass filter and all the data points above certain harmonic are neglected. In the temporal domain, cutting of higher frequencies results in the coarsening of the pulse. However, because the UPPE operates in the frequency domain and we are interested mainly in the spectral properties of the result field, reducing the size of the problem in the frequency domain does not affect the propagation of the field as long as we have adequate sampling. Using this approach we can reduce the size of the input and output data by a factor of 10.

Feature extraction – real and imaginary components of the spectrum

Because we are mainly interested in the spectral features – which represent different orders orders of magnitude in the frequency domain – we keep the inputs and the outputs of the NN in the spectral domain. If we want to transform the output back into the temporal domain, we need the full information of the spectrum, i.e. the amplitude and phase. Because the field and source term are real, the negative frequency components are redundant (complex conjugate). For inputs and outputs, we stick to the real and imaginary parts of the positive spectrum with lowpass filtering. Using amplitude and phase for input and output description appeared problematic, see Appendix B, Section B.1. We still have to account for the fact that the intensities of spectral components vary by orders of magnitude. Another tried and tested approach was to compute odd roots of the individual components, see Appendix B, Section B.2. This idea was also eventually dropped and we settled with the unmodified amplitudes. Last but not least, to enhance the spectral features, the temporal error function filter presented in Subsection 3.2.3 was applied before the spectral decomposition and before the low pass filtering.

The output and input spectrum is phase-shifted before the decomposition into real and imaginary parts. Every component of a Fourier signal F_ω is multiplied by factor $e^{-i\omega T/2}$ where T is the pulse length in time. The phase shift, according to the Fourier shift theorem, basically pushes the field in the temporal domain by $T/2$. Due to the nature of FFT[11], the phase shift reduces fast oscillations within the real and imaginary parts of the spectra, as depicted in Figure 5.7. As a result, we have nicely continuous curves better suited for features extraction.

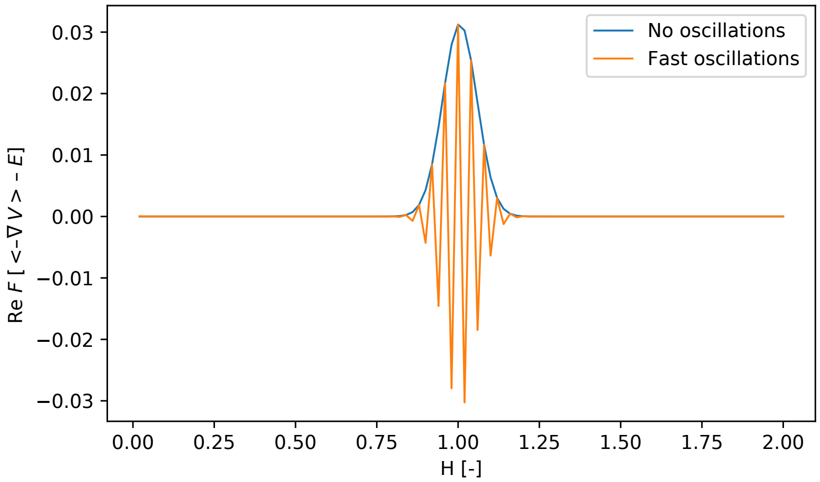

Figure 5.7: The effect of the phase shift $e^{-i\omega T/2}$ on the real part of the 1D-TDSE source term, around the fundamental field component. By application of this transform we effectively get rid of the fast oscillations (orange curve) and obtain a smooth curve with the same amplitude and phase (blue curve).

Data normalization

Normalization is a standard data preprocessing procedure in NNs. NNs generally perform better with normalized inputs for many reasons. Activations may saturate more easily, implying almost zero gradients for weight computation. This results in much slower convergence of the model or the model reaches only local minima. Having the input in a predefined range, say $(-1, 1)$, the activation behaves almost linearly in this region and has the steepest gradient which further improves the convergence.

The input data was normalized according to the min-max normalization[12] in range $(-1, 1)$:

$$x_{\text{norm}} = \frac{2\,(x - \min(x))}{\max(x) - \min(x)} - 1, \tag{5.18}$$

---

[11] Many implementations of FFT [80] assume the temporal signal centered around 0. The extra phase $e^{-i\omega T/2}$ effectively shifts the signal around zero.

[12] There exist other normalization techniques such as mean normalization, z-score normalization, or logarithmic normalization. Z-score and mean normalizations performed comparatively for our data.

where min(x) and max(x) correspond to the minimum and the maximum of the whole training dataset.

Summary

The data preprocessing pipeline including temporal filtering, low pass filter (downsampling) and separation of the spectrum into real and imaginary parts is sketched in Figure 5.8. The custom pipeline was fully implemented in Python.

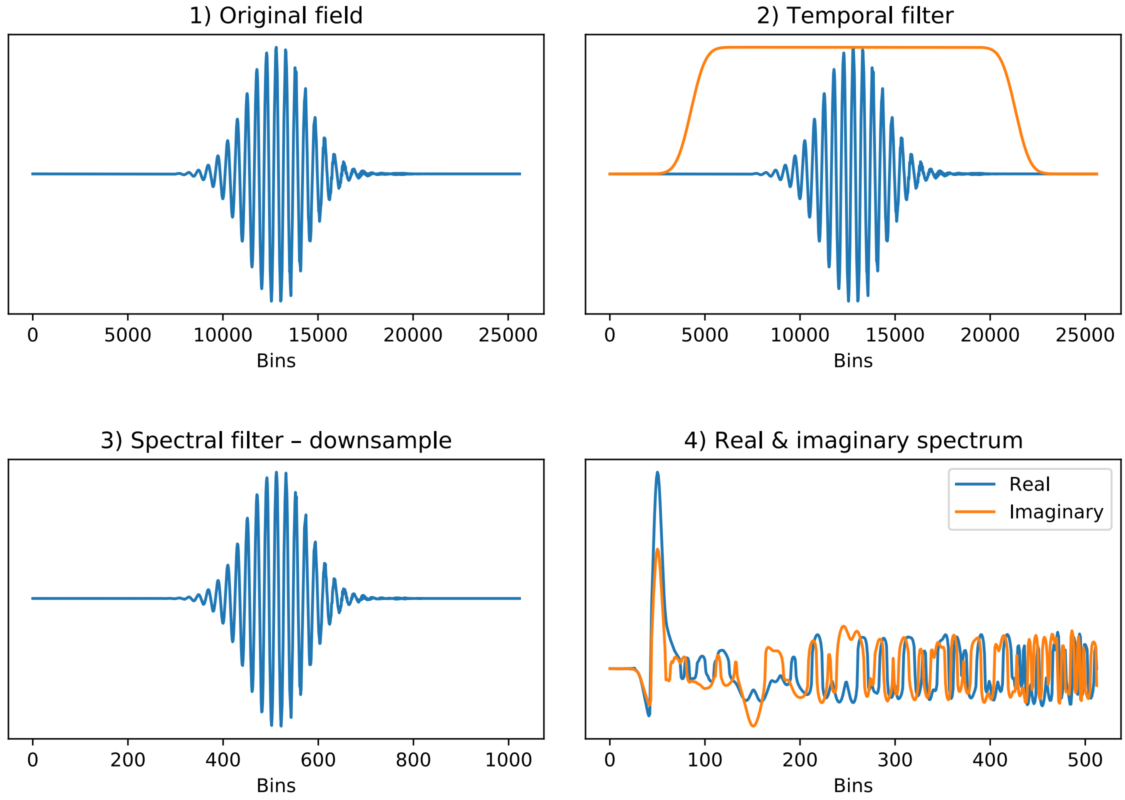

Figure 5.8: Data preparation pipeline. The electric field, Subfigure 1), is multiplied by the temporal filter, Subfig. 2) and downsampled by cutting of higher frequencies, Subfig. 3), decreasing the number of bins roughly 20 times. Finally, the spectrum spectrum is separated into real and imaginary parts, Subfig. 4 (artificially enhanced features for clarity using 3rd root, see Appendix B).

5.2.2 Neural network design

A custom neural network incorporating convolution and dense layers for predicting source terms of the 1D-TDSE, given the numerical electric field input, has been implemented. Python library Tensorflow was used to assemble the model. The model will now be referred to as **TDSE-NN** in the following text. TDSE-NN full design is depicted in Figure 5.9. The design is a combination of CNN and MLP networks.

The input layer takes the already preprocessed and normalized data of size $[N, 2]$, Fig. 5.8. The size N of the input array is set fixed in the beginning according to the desired frequency cutoff. The data is then fed through the cascade of 1D convolution and max pooling layers. Convolution layers have kernels of size 3 with $\tanh(x)$ activation and the number of channels doubles each layer. The data tensors must then be flattened before advancing into the MLP section. A batch normalization layer is introduced here to improve the robustness of the model along with a Gaussian noise layer, further regularizing the network. The MLP is based on the compression-decompression scheme – the dense layers shrink the number of nodes from 128 to 32 and then back to 128. Each dense layer is equipped with[13] $\tanh(x)$ activation and dropout layers are added in between the dense layers for increased regularization.

A skipped connection was implemented right after the first dense layer and then added to the decompressed data. After a final dropout layer comes an output dense layer with identity activation outputting an array of size $2 \times M$, where M is the length of the real or imaginary part of the positive spectrum, i.e. the output array looks as follows:

$$\boldsymbol{y}_{\text{output}} = [\mathrm{Re}[\hat{y}_1], \mathrm{Re}[\hat{y}_2], \ldots, \mathrm{Re}[\hat{y}_M], \mathrm{Im}[\hat{y}_1], \mathrm{Im}[\hat{y}_2], \ldots, \mathrm{Im}[\hat{y}_M]] \tag{5.19}$$

which can be then reconstructed using the IFFT back into the temporal domain (the negative frequencies are hermitian).

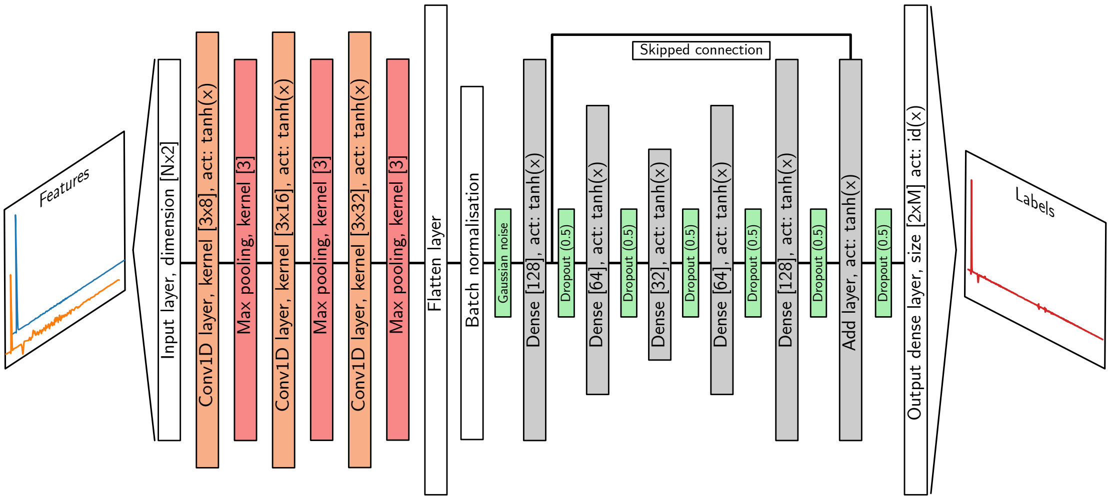

Figure 5.9: TDSE-NN Neural network design.

5.3 Results

The results obtained using the TDSE-NN model are presented in this section. TDSE-NN was trained on the data acquired from the multi-scale model. The data was preprocessed as proposed in Subsection 5.2.1. Low intensity field (10^{10} W/cm^2), see Subsection 4.1.2, and high intensity field (10^{14} W/cm^2), see Section 4.2, simulations were chosen for the assessment of the TDSE-NN. In total from a single simulation run, we had 2000 data samples for high-intensity fields and 4000 samples for low-intensity fields. Each sample comprises the pre-processed input electric field and the output in the form of the 1D-TDSE source term. The ratio of the training and validation data was established at 80:20%. Therefore the shape of the full training features dataset for low-intensity field (10^{10} W/cm^2) is $(3200, N, 2)$ and the training labels dataset shape $(3200, 2 \times M)$, or $(800, N, 2)$ and $(800, 2 \times M)$ respectively for the low-intensity field validation datasets. The shapes of the datasets for high-intensity field simulation (10^{14} W/cm^2) are analogical.

---

[13] For the details about the choice of activations see Section B.3 in Appendix B.

The datasets were initially randomly shuffled using the Numpy [101] shuffle method along the first dimension to further mitigate data bias during the training.

5.3.1 TDSE-NN for low-intensity field

We begin with the dataset for low peak intensity (10^{10} W/cm^2), Subsection 4.1.2. We do the subsampling of the input data to[14] 1025 bins including frequencies up to H $\approx$ 10 corresponding to the ionization potential of argon. The real and imaginary parts of the input and output spectra have sizes $N = M = 513$. The entire TDSE-NN comprises of a total 231,306 parameters (230,154 trainable/1,152 non-trainable). The batch size was fixed to 64.

Model accuracy and number of epochs

We first examine the accuracy of the model for 3 different amounts of training data corresponding to the same intensity and propagation distance. The number of epochs was fixed to 200 and the TDSE-NN was trained and validated on the datasets of different sizes. The training and validation losses (MSE, Eq. (5.8)) along with metrics (MAPE, Eq. (5.16)) during the training are depicted in Figure 5.10. The figure clearly shows that training with more data decreased the number of epochs necessary for the training but did not significantly improve the accuracy of the model. Moreover, the model with the largest dataset (green curve) actually starts overfitting after 100 epochs and its accuracy decreases. Looking at the result spectra obtained for the same input from the models at different dataset sizes in Figure 5.11, we can notice the overfitting of the model by examining the noisy red curve for range H $= 3 - 6$ that differs from the rest (green and orange). We also notice a significant discrepancy between the predictions and the correct spectrum for H $= 4 - 8$.

Model results

We now examine the TDSE-NN model which was trained using the full dataset (3200 train, 800 validation) with the implementation of early stopping criteria. The criteria were chosen such that it stopped the training if the model loss had not improved after 25 consecutive epochs. In total, the model was trained for 53 epochs. The final validation loss was $\mathcal{L}_{\text{val}} \doteq 2.29 \cdot 10^{-11}$ and validation metric $\text{MAPE}_{\text{val}} = 2.06\%$. The training took around 3 min on a quad-core Intel i7-6700HQ. Inference for a single input took around 40 ms. The result spectra for 2 different inputs are plotted in Figure 5.12. The predictions again differ significantly in region H $= 4 - 8$. However, the dominant spectral features such as the fundamental and third harmonic components or the Rydberg states quantitatively match the values to some small margin of error. It should be stressed that the labels and features contain details of scales spanning 10 orders of magnitude as no data scaling technique has been used.

Figure 5.12 fills the suspicion that the training data is biased to a particular spectral shape since the network outputs similar spectra in range H $= 4 - 6$ for two different inputs. We base our claim on looking at the resolved harmonic spectra for the full propagated distance

---

[14] The odd number of bins is selected deliberately so that the positive signal in the frequency domain always has exactly $(N + 1)/2$ bins.

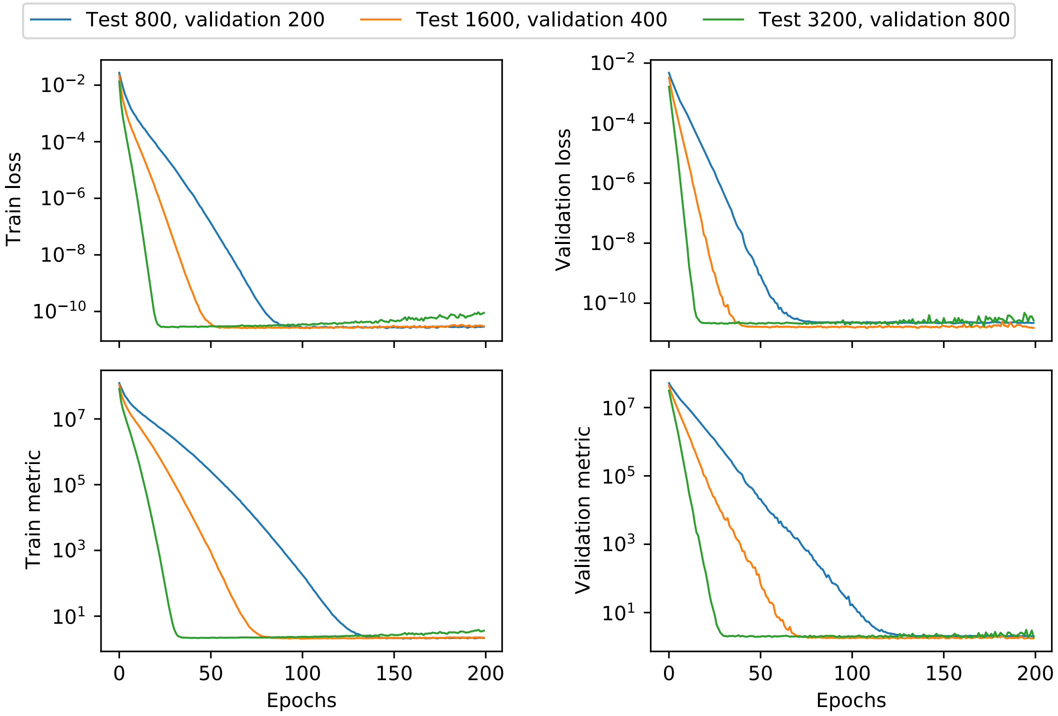

Figure 5.10: The result mean squared error loss (MSE) and mean absolute percentage error metric (MAPE) during the training of TDSE-NN for 200 epochs for varying amounts of training and validation data.

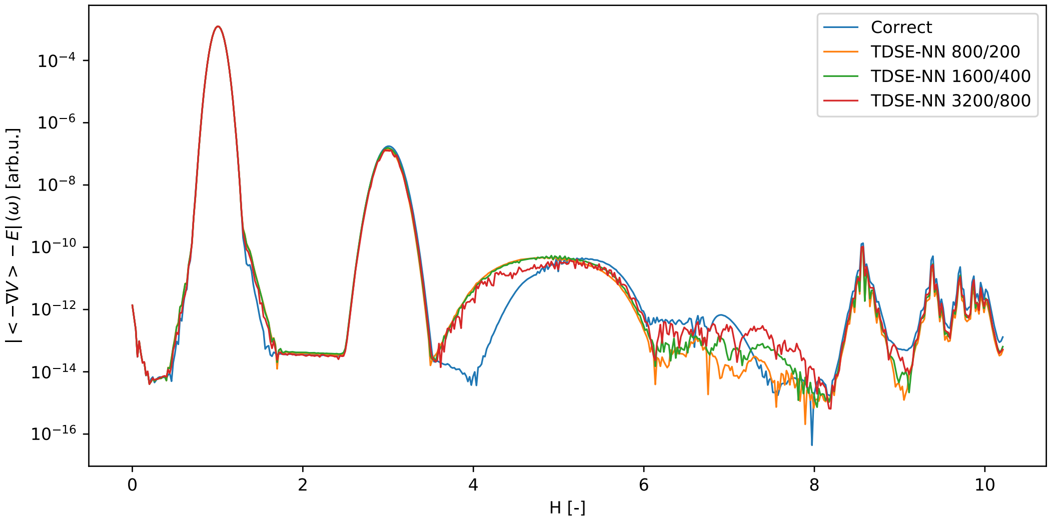

Figure 5.11: The correct result harmonic spectrum of the *ab-initio* source term (blue) is overlaid by the TDSE-NN results after the training of TDSE-NN for 200 epochs for varying amounts of training and validation data (training/validation) for the same input.

in Figure 4.6. We notice that the spectra for distances 0.8 – 2.0 mm do not contain much variety. Since it takes up to 50% of the whole dataset, the bias is likely caused by a uniform selection of data samples from this region of space. The data bias problem for the low-intensity dataset could be reduced by balancing the selection of data from Figure 4.6.

The predicted source term was reconstructed in the time domain using the IFFT and

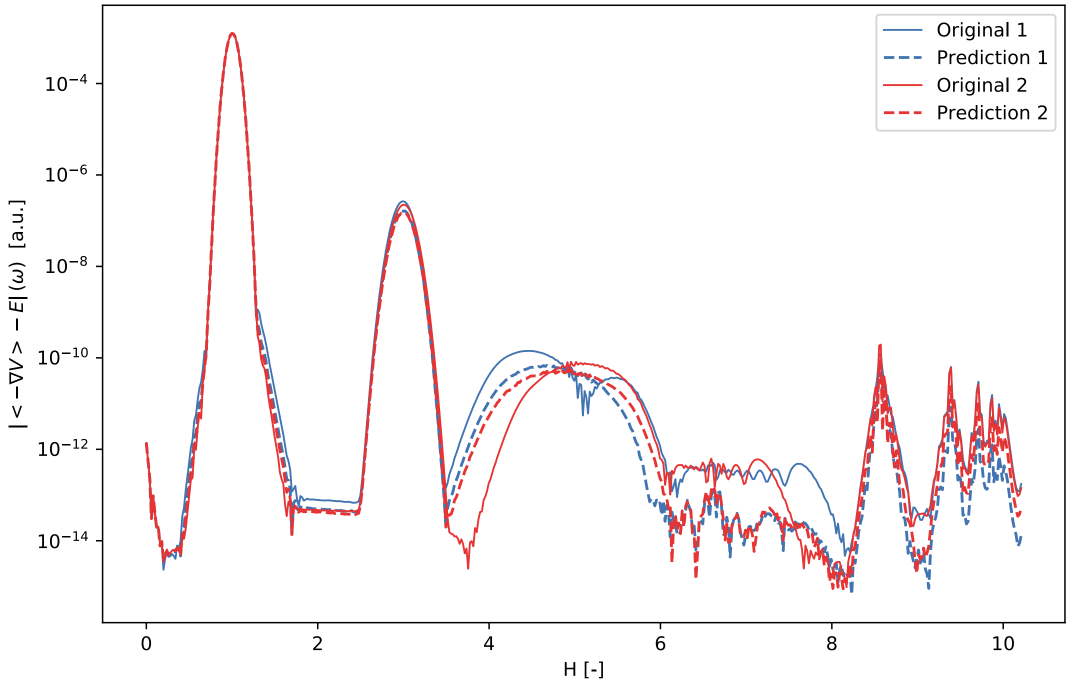

Figure 5.12: The result spectra of the source terms for two different field inputs from the TDSE-NN model (dashed line) along with the correct spectra (solid line).

plotted along with the correct source term in Figure 5.13. The envelope of the NN prediction has been decently preserved. Inspecting the detailed subplot, we notice that the NN prediction is slightly shifted to the left by a few atomic units of time. The shift is caused by an incorrect phase of the fundamental frequency as depicted in Figure 5.14.

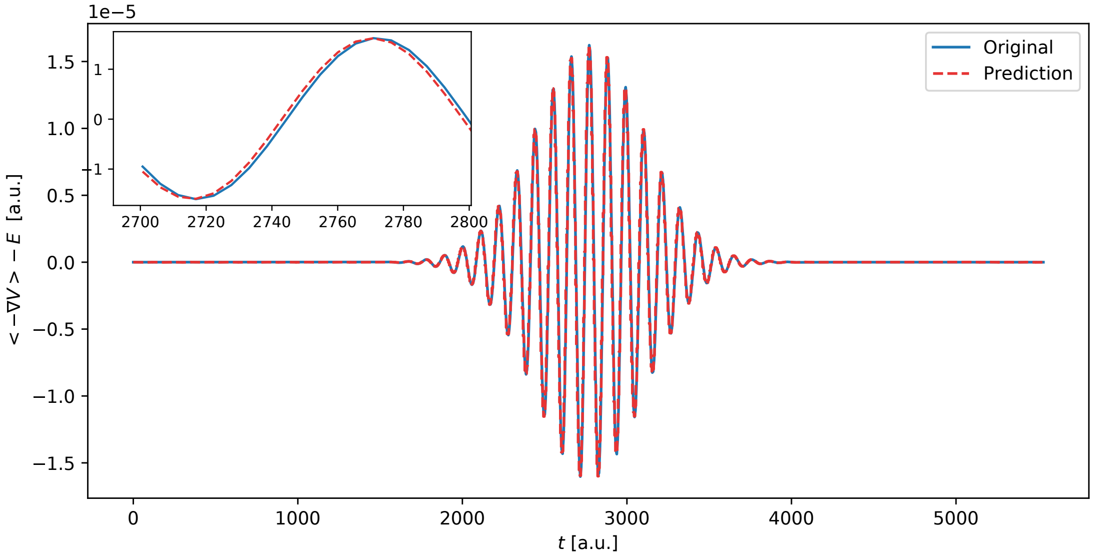

Figure 5.13: The result reconstructed source term in the time domain for single input from the TDSE-NN model (dashed line) along with the correct source term (solid line).

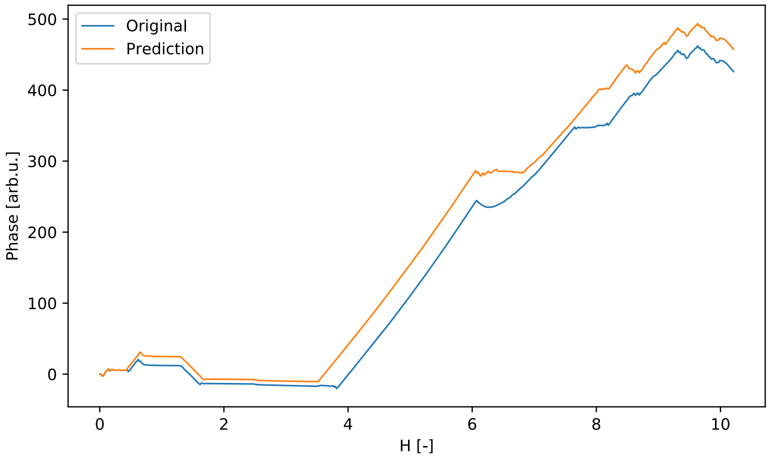

Figure 5.14: The unwrapped phases of the source terms $\langle -\nabla V - E \rangle$ for different harmonic orders.

5.3.2 TDSE-NN for high-intensity field

The TDSE-NN model was trained on the full dataset with high-intensity data (1600 train, 400 validation) for 100 epochs. The batch size was fixed to 64. The size of the input feature vector was $(1000, 2)$ and the output labels had a size of 2000 bins, capable of describing spectral features up to H $\approx$ 20. The model comprises 432,984 parameters (430,680 trainable/2,304 non-trainable). The final validation loss was $\mathcal{L}_{\text{val}} \doteq 6.36 \cdot 10^{-4}$ and validation metric $\text{MAPE}_{\text{val}} = 862.70\%$. The training took around 1 min 20 s on a quad-core Intel i7-6700HQ. Inference for a single input took around 80 ms.

The correct and predicted spectra for a single input are plotted in Figure 5.15. The lack of spectral overlap is quantified by the high error rate of the validation metric. The network prediction spectrum does not fit the correct source term in region H $= 3.5 - 7.5$. On the other hand, the profile of the predicted source term at least follows, with a considerable margin of error, the spectral features above H $= 7.5$. The root of the high validation MAPE can be spotted by inspecting Figure 5.16. The losses and metrics are evaluated exactly on the differences between the blue (correct) and red (prediction) curves. Last we plot the reconstructed temporal evolution of the source term in a smaller temporal window in Figure 5.17. The prediction differs from the correct result substantially after $t = 2500$ a.u.

We investigate the reasons for the high prediction inaccuracy by inspecting the evolution of losses and metrics during the training of the network, Figure 5.18. We notice that while the test loss and metric (blue solid and dotted lines) monotonously decrease, the same cannot be said about the validation loss and metric. We observe a local minimum around epoch 16, then the errors soar by a factor of 5 and steadily start to decrease after epoch 40. According to [98], this so-called ”double-descent” phenomenon can result in two conclusions. Either the number of epochs is low or the capacity of the model is insufficient for the problem. During the experimentation with optimizing the training duration, it has been shown that increasing the number of epochs did not yield better results. This brings us to another conclusion that the TDSE-NN has low capacity, i.e. the size of the network is insufficient. Therefore in order to get better results, we should probably add more hidden layers or increase the size of the current layers.

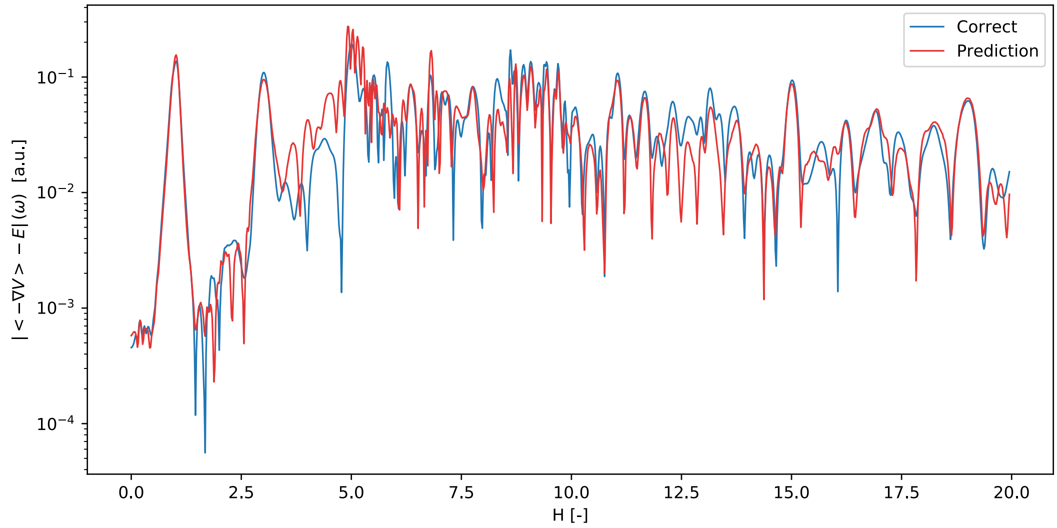

Figure 5.15: The result spectra of the source terms for the same numerical field input from the TDSE-NN model (red) along with the correct spectrum (blue).

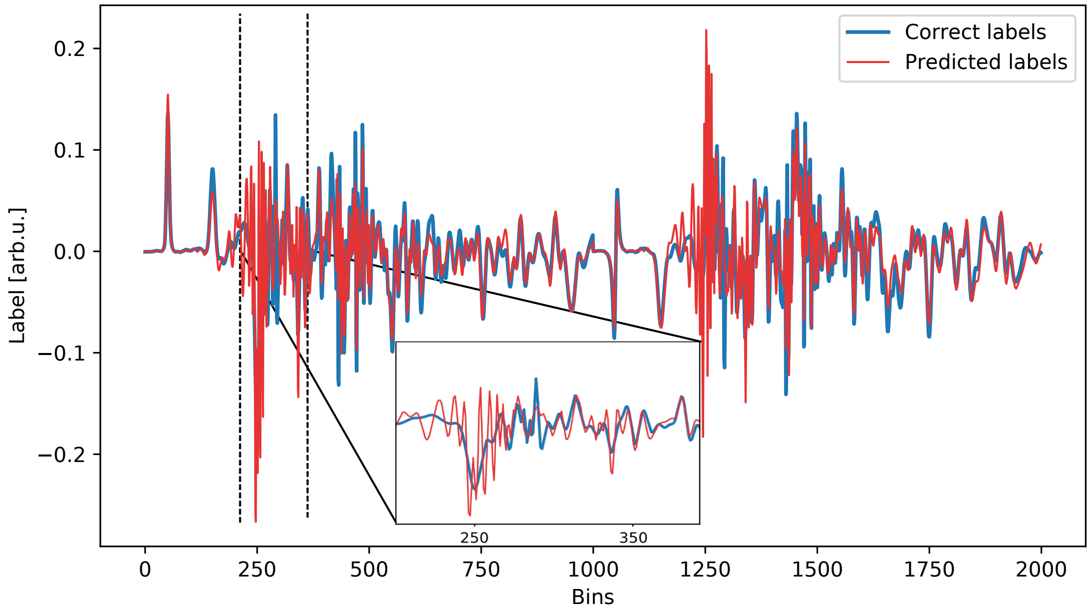

Figure 5.16: The correct (blue) and predicted labels (red) from the TDSE-NN model for a single input. We inset a detailed comparison corresponding to the region marked by the black dashed vertical lines. Remark that labels comprise real and imaginary parts of the source term spectrum for the first, respectively second part of the full array. If we join the real and imaginary parts together and plot the modulus in log-scale, we get exactly Figure 5.15.

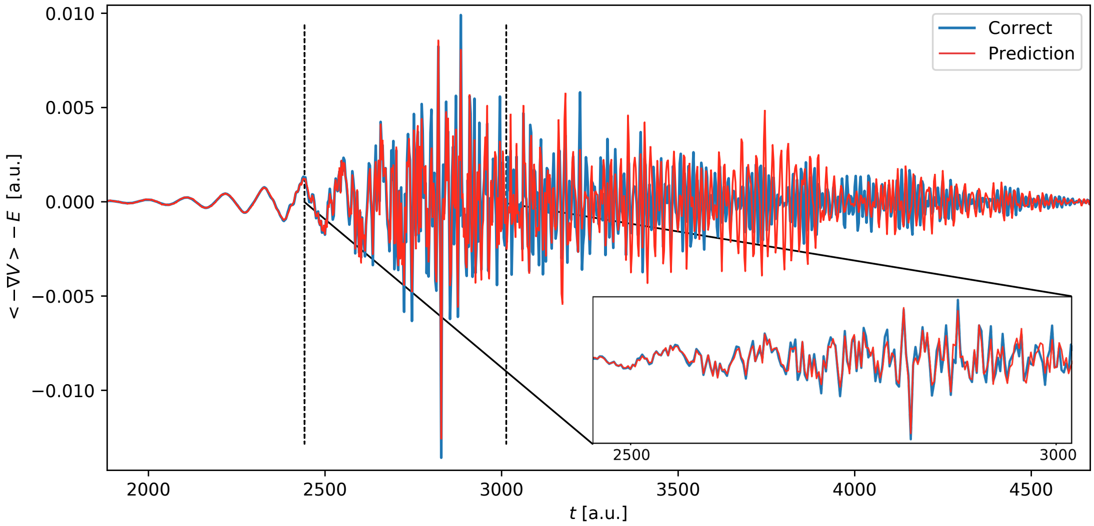

Figure 5.17: The result reconstructed source term in the time domain for single input from the TDSE-NN model (red) on top of the correct source term (blue). We inset a detailed comparison corresponding to the region marked by the black dashed vertical lines.

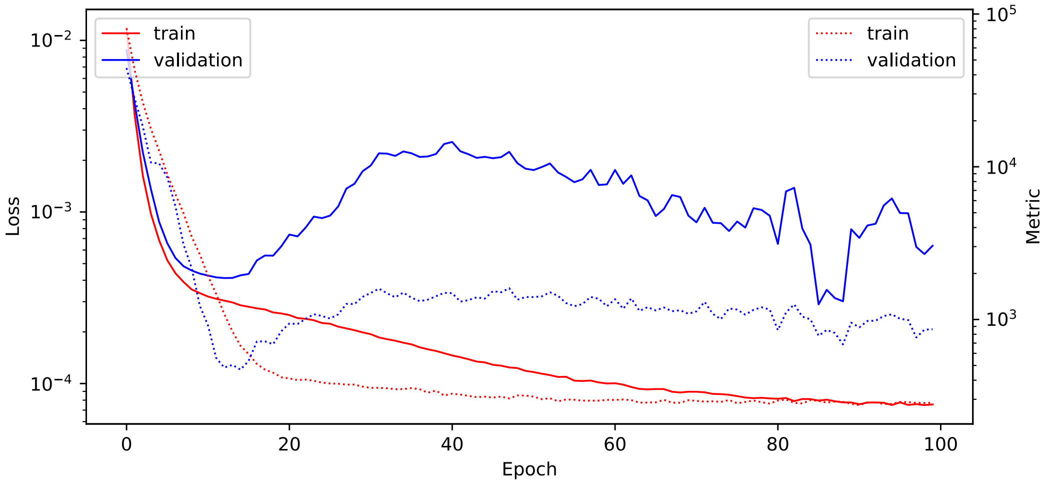

Figure 5.18: The result mean squared error loss for train and validation datasets, on the primary y-axis, (solid lines) and mean absolute percentage error metric for train and validation datasets, on the secondary y-axis, (dotted lines) during the training of TDSE-NN for 100 epochs.

5.4 Outlook and further improvements

Let us discuss some possible design improvements of TDSE-NN and outline further research paths. The general goal was to test the feasibility of using the NN for source term computation from numerical electric fields. Therefore we naturally began with a simple and rather small network[15] to establish the starting point for further development.

Our TDSE-NN model was designed to work in the frequency domain. However, there are no fundamental issues for it to operate in the temporal domain of the source term. In principle, the neural network is indifferent to the nature of the input and output data as long as it can find some deterministic relationship between the two. TDSE-NN in the temporal domain would likely benefit from a correct phase of the fundamental field component. On the other hand, inferring the spectrum from the temporal output is more challenging as the model must sample globally the evolution of the field with a factor of 10 accuracy to output correct spectra.

An idea of how to implement the temporal component into the prediction of the network was to evaluate MSE loss both in frequency and temporal domain and minimize their sum. The approach was tested but due to the high cost of the custom loss function evaluation and the lack of deep experience with Tensorflow API, the idea was dropped for later developments. An alternative approach borrows inspiration from audio processing neural networks and analysis of the input electric field in the time-frequency domain [103, 120]. By implementing short-time Fourier transforms we obtain a spectrogram that binds together frequencies occurring at a given instant of the signal. The spectrogram gets further processed by 2D convolution layers. The audio processing NNs are well accustomed to working with frequencies spanning tens to hundreds of decibels while retaining the phase of the signal. This coincides with the same task as the TDSE-NN is trying to achieve.

Naturally, a question emerges about how to tackle variable-sized outputs and inputs. We are not aware of any NN architecture other than RNN or LSTM [104, 105] that could process variable-sized inputs and outputs. On the other hand, we are uncertain these architectures would yield superior and cost-effective results anyway, see Appendix B, Section B.4. A possible option within the realm of proposed networks is to set a sufficiently large, fixed-sized input and output to encompass the required validity range for the desired pulse lengths and frequencies. Further data augmentation is also possible by neglecting the parts of the pulse (e.g. pulse front) where we do not expect any substantial harmonic yields. Alternatively, we could keep only the individual phases and amplitudes of the generated harmonics in a single frequency point to push the data compression even further. Instead of looking at the whole pulse, we might also try to generate more detailed snapshots of the numerical fields in time and estimate the generation based on these snapshots. The physical reasoning behind this idea is the assumption that the source term generation could be influenced just by a few preceding cycles.

Last but not least, the 1D-TDSE approximator needs large quantities of representative data. The examples showcased in the thesis concerned a tiny portion of possible input pulses, mainly given by constraints of the multi-scale model. More emphasis should be placed on including pulses with varying peak intensities, pulse envelopes, central frequencies, and even chirped pulses. Generally the larger the dataset for a network with sufficient capacity, the better the outcome. The network, by construction, cannot correctly describe anything out of the scope it has been trained on [33].

---

[15] Networks dealing with similar physical problems can take days to train even on powerful GPUs [102].

Conclusion

The thesis covered the physics of microscopic and macroscopic scales for establishing a solid foundation for assembling the multi-scale *ad-hoc* and *ab-initio* models. The multi-scale model was based on the unidirectional pulse propagation equation with varying source terms and was solved using the 4th-order Runge-Kutta scheme. Implementation of the *ab-initio* model required getting deeply familiar with a numerical 1D-TDSE. Defining the *ad-hoc* model required an understanding of the chi-process and obtaining the 1st- and 3rd-order susceptibilities based on the 1D-TDSE calculations.

The susceptibilities $\chi^{(1)}$ and $\chi^{(3)}$ for argon in the laser harmonic range $\mathrm{H} = 0 - 3.5$ were successfully extracted using the dipole model. The intensity threshold for the interpolation was set to $I = 10^{13}$ W/cm^2 which was determined by the ionization rate. The validity of the polynomial fits of $\chi^{(1)}$ and $\chi^{(3)}$ implemented into the *ad-hoc* model was checked against the *ab-initio* model at intensities ranging from 10^{10} to 10^{13} W/cm^2 for propagation at atmospheric pressure in argon for 26 μm upto 2.6 mm. The validity of the *ad-hoc* model was constrained on 5-cycle (FWHM) pulses and frequency range $\mathrm{H} = 0 - 3.5$. The *ad-hoc* model results demonstrated Kerr-induced self-phase modulation at intensity 10^{13} W/cm^2 and 3rd harmonic generation, both quantitatively and qualitatively comparable with the *ab-initio* model. The *ab-initio* model results displayed interference patterns in the spatially resolved harmonic spectrum. While these features may be partly attributed to physical interaction between the individual harmonics, more in-depth testing is necessary to rule out numerical artifacts.

The critical bottleneck in the evaluation of the *ab-initio* model is the computation of the source term using the 1D-TDSE. A custom NN model called TDSE-NN was presented to substitute the source term computation. TDSE-NN combined convolution and multilayer perceptron NN architectures and took a numerical electric field in the frequency domain as its input to produce the source term for the macroscopic model. The network was trained on the propagation data from the *ab-initio* model for intensities 10^{10} and 10^{14} W/cm^2. The training on the low-intensity dataset (10^{10} W/cm^2, 2.6 mm) revealed signs of data bias during the inference phase. Moreover, the phase of the result source term was slightly off, leading to the shift of the field in the temporal domain. The best-achieved validation loss and mean absolute percentage error metric of the model for low-intensity fields was $2.29 \cdot 10^{-11}$ respectively 2.06%. TDSE-NN was also trained on a wider spectral range in the case of the higher intensity field with the best validation loss and metric being $6.26 \cdot 10^{-4}$ respectively 862.70%. While some of the spectral features for higher-order harmonics were predicted reasonably well, the model was unable to cope with lower harmonics. The likely reason for the high inaccuracy of the model was the insufficient capacity of the network.

The greatest benefit of this thesis was to examine the viability of using a simple neural network in place of the TDSE solver. This research is still in a very preliminary stage but the potential of NNs should not be understated and is certainly worth further examination. As demonstrated in a recent article by Pablos-Marín et. al. [102], NNs are capable of predicting higher-order harmonics response in gas mimicking 3D-TDSE with good accuracy. However, the model proposed in this thesis goes one step further by incorporating a full numerical field as the input for the NN. While the evaluation of a single source term took in the range of $10 - 100$ s, the NN inference for a single input took in the range of $10 - 100$ ms, offering the opportunity to speed up the multi-scale model by 2 to 3 orders of magnitude!

The future outlook is to focus on increasing the capacity of the network, possibly by introducing multiple skipped connections in the MLP section. More emphasis should be put on reducing the data bias through careful selection of samples from the macroscopic propagation or by the generation of datasets with increased variability of field parameters. To change the network design completely, we could borrow inspiration from the audio processing neural networks [103, 120] and analyze the input electric field in the time-frequency domain using the short-time Fourier transform. The pinnacle of the NN model would be the option to predict the 1D-TDSE source terms for arbitrarily sized and shaped pulses with various intensities and central frequencies.

References

Monographs

1. HUBA, J. *NRL: Plasma Formulary.* Naval Research Laboratory, 2004. Available also from: `https://www.researchgate.net/publication/235025211_NRL_Plasma_Formulary`.
2. JACKSON, J. D. *Classical electrodynamics.* 3rd ed. New York, NY: Wiley, 1999. ISBN 9780471309321.
3. LIN, C. D.; LE, A.-T.; JIN, C.; WEI, H. *Attosecond and Strong-Field Physics: Principles and Applications.* Cambridge University Press, 2018. ISBN 978-1-107-19776-3.
4. HAKEN, H.; WOLF, H.; BREWER, W. *The Physics of Atoms and Quanta: Introduction to Experiments and Theory.* Berlin: Springer, 2005. ISBN 978-3-540-20807-5. Available from DOI: `10.1007/3-540-29281-0`.
5. MARQUES, M. A. L.; ULLRICH, C. A.; NOGUEIRA, F.; RUBIO, A.; BURKE, K.; GROSS, E. K. U. *Time-Dependent Density Functional Theory.* Springer Berlin Heidelberg, 2006. Available from DOI: `10.1007/b11767107`.
6. BALLENTINE, L. E. *Quantum Mechanics: A Modern Development.* 2nd ed. Singapore, Singapore: World Scientific Publishing, 1998. ISBN 9810241054.
7. AGRAWAL, G. *Nonlinear Fiber Optics.* Chapter 2 - Pulse Propagation in Fibers. 5th ed. Boston: Academic Press, 2013. Optics and Photonics. ISSN 15575837. Available from DOI: `10.1016/B978-0-12-397023-7.00002-4`.
8. BOYD, R. W. *Nonlinear Optics.* 4th ed. Elsevier Science, 2020. Available from DOI: `10.1016/c2015-0-05510-1`.
9. WIESE, W. L.; MILES, B. M.; SMITH, M. W. *Atomic transition probabilities, Vol. 2: Sodium Through Calcium; A Critical Data Compilation.* 1st ed. Washington: National Standard Reference Data System, 1969.
10. LIU, G.R.; QUEK, S.S. (eds.). *The Finite Element Method (Second Edition).* The Finite Element Method. Oxford: Butterworth-Heinemann, 2014. ISBN 978-0-08-098356-1. Available from DOI: `10.1016/B978-0-08-098356-1.00014-X`.
11. BLANK, J.; EXNER, P.; HAVLICEK, M. *Hilbert Space Operators in Quantum Physics.* 2nd ed. Berlin: Springer, 2008. ISBN 978-1-4020-8869-8. Available from DOI: `10.1007/978-1-4020-8870-4`.
12. CALEGARI, F.; LIU, C.; LUCCHINI, M.; SANSONE, G.; NISOLI, M. *Progress in Ultrafast Intense Laser Science: Volume IX.* Gating Techniques for Shaping of Attosecond Pulses. Berlin: Springer Berlin Heidelberg, 2013. ISBN 978-3-642-35052-8. Available from DOI: `10.1007/978-3-642-35052-8_4`.
13. BOLSTAD, William; CURRAN, James. *Introduction to Statistical Science.* John Wiley and Sons, 2016. ISBN 9781118091562. Available from DOI: `10.1002/9781118593165.ch1`.
14. GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning.* Illustrated. MIT Press, 2016. ISBN 978-0262035613.

Articles, proceedings and inproceedings

15. HARTREE, D. R. The wave mechanics of an atom with a non-coulomb central field. Part I. Theory and methods. *Mathematical Proceedings of the Cambridge Philosophical Society.* 1928, **24**(4), 89–110. Available from DOI: `10.1017/S0305004100011919`.
16. WATANABE, S.; KONDO, K.; NABEKAWA, Y.; SAGISAKA, A.; KOBAYASHI, Y. Two-Color Phase Control in Tunneling Ionization and Harmonic Generation by a Strong Laser Field and Its Third Harmonic. *Physical review letters.* 1994, **73**, 2692–2695. Available from DOI: `10.1103/PhysRevLett.73.2692`.
17. JIN, C.; WANG, G.; WEI, H.; LE, A.-T.; LIN, C. D. Waveforms for optimal sub-keV high-order harmonics with synthesized two- or three-colour laser fields. *Nature Communications.* 2014, **5**(1). ISSN 2041-1723. Available from DOI: `10.1038/ncomms5003`.
18. FINKE, O. et al. Phase-matched high-order harmonic generation in pre-ionized noble gases. *Scientific Reports.* 2022, **12**(1). Available from DOI: `10.1038/s41598-022-11313-6`.
19. AMICO, C D et al. Forward THz radiation emission by femtosecond filamentation in gases: theory and experiment. *New Journal of Physics.* 2008, **10**(1), 013015. Available from DOI: `10.1088/1367-2630/10/1/013015`.
20. LEWENSTEIN, M.; BALCOU, Ph.; IVANOV, M. Yu.; L'HUILLIER, A.; CORKUM, P. B. Theory of high-harmonic generation by low-frequency laser fields. *Physical Review A.* 1994, **49**(3), 2117–2132. ISSN 1050-2947. Available from DOI: `10.1103/PhysRevA.49.2117`.
21. LORIN, E.; CHELKOWSKI, S.; BANDRAUK, A. The WASP Model: A Micro-Macro System of Wave-Schrodinger-Plasma Equations for Filamentation. *Communications in Computational Physics.* 2010, **9**. Available from DOI: `10.4208/cicp.071009.300410a`.
22. LORIN, E.; BANDRAUK, A. D. Efficient and accurate numerical modeling of a micro–macro nonlinear optics model for intense and short laser pulses. *Journal of Computational Science.* 2012, **3**(3), 159–168. ISSN 1877-7503. Available from DOI: `10.1016/j.jocs.2011.12.004`. Scientific Computation Methods and Applications.
23. OPENAI. GPT-4 Technical Report. *arXiv.* 2023. Available from DOI: `10.48550/arXiv.2303.08774`.
24. CHOWDHERY, A. et al. PaLM: Scaling Language Modeling with Pathways. *arXiv.* 2022. Available from DOI: `10.48550/arXiv.2204.02311`.
25. CHENG, B.; BRADLEY, P. A. What Machine Learning Can and Cannot Do for Inertial Confinement Fusion. *Plasma.* 2023, **6**(2), 334–344. ISSN 2571-6182. Available from DOI: `10.3390/plasma6020023`.
26. THE CMS COLLABORATION. A deep neural network to search for new long-lived particles decaying to jets. *Machine Learning: Science and Technology.* 2020, **1**(3), 035012. Available from DOI: `10.1088/2632-2153/ab9023`.
27. RAISSI, M.; PERDIKARIS, P.; KARNIADAKIS, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics.* 2019, **378**, 686–707. ISSN 0021-9991. Available from DOI: `10.1016/j.jcp.2018.10.045`.
28. KARNIADAKIS, G. E.; KEVREKIDIS, I. G.; LU, L.; PERDIKARIS, P.; WANG, S.; YANG, L. Physics-informed machine learning. *Nature Reviews Physics.* 2021, **3**(6), 422–440. ISSN 2522-5820. Available from DOI: `10.1038/s42254-021-00314-5`.
29. HAN, J.; ZHANG, L.; E, W. Solving many-electron Schrödinger equation using deep neural networks. *Journal of Computational Physics.* 2019, **399**, 108929. ISSN 0021-9991. Available from DOI: `10.1016/j.jcp.2019.108929`.
30. HERMANN, J.; SPENCER, J.; CHOO, K.; MEZZACAPO, A.; FOULKES, W. M. C.; PFAU, D.; CARLEO, G.; NOÉ, F. Ab initio quantum chemistry with neural-network wavefunctions. *Nature Reviews Chemistry.* 2023, **7**(10), 692–709. ISSN 2397-3358. Available from DOI: `10.1038/s41570-023-00516-8`.
31. SECOR, M.; SOUDACKOV, A. V.; HAMMES-SCHIFFER, S. Artificial Neural Networks as Propagators in Quantum Dynamics. *The Journal of Physical Chemistry Letters.* 2021, **12**(43), 10654–10662. Available from DOI: `10.1021/acs.jpclett.1c03117`.
32. PASZKE, A. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv.* 2019. Available from DOI: `10.48550/arXiv.1912.01703`.

33. HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural Networks.* 1989, **2**(5), 359–366. ISSN 0893-6080. Available from DOI: `10.1016/0893-6080(89)90020-8`.

34. ZIELINSKI, T. J.; HARVEY, E.; SWEENEY, R.; HANSON, D. M. Quantum States of Atoms and Molecules. *Journal of Chemical Education.* 2005, **82**(12), 1880. Available from DOI: `10.1021/ed082p1880.2`.

35. HARTREE, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Mathematical Proceedings of the Cambridge Philosophical Society.* 1928, **24**(1), 111–132. Available from DOI: `10.1017/S0305004100011920`.

36. SLATER, J. C. The Self Consistent Field and the Structure of Atoms. *Physical Review.* 1928, **32**, 339–348. Available from DOI: `10.1103/PhysRev.32.339`.

37. FOCK, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik.* 1930, **61**(1-2), 126–148. Available from DOI: `10.1007/bf01340294`.

38. FROESE FISCHER, C. General Hartree-Fock program. *Computer Physics Communications.* 1987, **43**(3), 355–365. ISSN 0010-4655. Available from DOI: `10.1016/0010-4655(87)90053-1`.

39. MORRISON, J. C.; STEFFEN, K.; PANTOJA, B.; NAGAIYA, A.; KOBUS, J.; ERICSSON, T. Numerical Methods for Solving the Hartree-Fock Equations of Diatomic Molecules II. *Communications in Computational Physics.* 2016, **19**(3), 632–647. Available from DOI: `10.4208/cicp.101114.170615a`.

40. AWASTHI, M.; VANNE, Y. V.; SAENZ, A.; CASTRO, A.; DECLEVA, P. Single-active-electron approximation for describing molecules in ultrashort laser pulses and its application to molecular hydrogen. *Physical Review A.* 2008, **77**, 063403. Available from DOI: `10.1103/PhysRevA.77.063403`.

41. REISS, H. Altered Maxwell equations in the length gauge. *Journal of Physics B: Atomic, Molecular and Optical Physics.* 2013, **46**. Available from DOI: `10.1088/0953-4075/46/17/175601`.

42. GAZIBEGOVIC-BUSULADZIC, A.; MILOSEVIC, D.B.; BECKER, W. Gauge dependence of the strong-field approximation: Theory vs. experiment for photodetachment of F-. *Optics Communications.* 2007, **275**(1), 116–122. ISSN 0030-4018. Available from DOI: `10.1016/j.optcom.2007.03.009`.

43. HAN, Y.-C.; MADSEN, L. Comparison between length and velocity gauges in quantum simulations of high-order harmonic generation. *Physical Review A.* 2010, **81**. Available from DOI: `10.1103/PhysRevA.81.063430`.

44. MILOŠEVIĆ, D B; PAULUS, G G; BAUER, D; BECKER, W. Above-threshold ionization by few-cycle pulses. *Journal of Physics B: Atomic, Molecular and Optical Physics.* 2006, **39**(14), R203. Available from DOI: `10.1088/0953-4075/39/14/R01`.

45. CORMIER, E.; LAMBROPOULOS, P. Optimal gauge and gauge invariance in non-perturbative time-dependent calculation of above-threshold ionization. *Journal of Physics B: Atomic, Molecular and Optical Physics.* 1996, **29**(9), 1667. Available from DOI: `10.1088/0953-4075/29/9/013`.

46. AGOSTINI, P.; DIMAURO, L. F. The physics of attosecond light pulses. *Reports on Progress in Physics.* 2004, **67**(6), 813–855. Available from DOI: `10.1088/0034-4885/67/6/r01`.

47. BUCKSBAUM, P. H. Sources and Science of Attosecond Light. *Opt. Photon. News.* 2015, **26**(5), 28–35. Available from DOI: `10.1364/OPN.26.5.000028`.

48. MCPHERSON, A.; GIBSON, G.; JARA, H.; JOHANN, U.; LUK, T. S.; MCINTYRE, I. A.; BOYER, K.; RHODES, C. K. Studies of multiphoton production of vacuum-ultraviolet radiation in the rare gases. *Journal of the Optical Society of America B.* 1987, **4**(4). ISSN 0740-3224. Available from DOI: `10.1364/JOSAB.4.000595`.

49. FERRAY, M.; L'HUILLIER, A.; LI, X. F.; LOMPRE, L. A.; MAINFRAY, G.; MANUS, C. Multiple-harmonic conversion of 1064 nm radiation in rare gases. *Journal of Physics B: Atomic, Molecular and Optical Physics.* 1988, **21**(3), L31–L35. ISSN 0953-4075. Available from DOI: `10.1088/0953-4075/21/3/001`.

50. GOULIELMAKIS, E. et al. Real-time observation of valence electron motion. *Nature.* 2010, **466**, 739–43. Available from DOI: `10.1038/nature09212`.

51. HAESSLER, S. et al. Phase-resolved attosecond near-threshold photoionization of molecular nitrogen. *Physical Review A.* 2009, **80**, 011404. Available from DOI: `10.1103/PhysRevA.80.011404`.

52. CAVALIERI, A et al. Attosecond spectroscopy in condensed matter. *Nature.* 2007, **449**, 1029–32. Available from DOI: `10.1038/nature06229`.

53. HE, Xinkui et al. Spatial and spectral properties of the high-order harmonic emission in argon for seeding applications. *Physical Review A*. 2009, **79**, 063829. Available from DOI: `10.1103/PhysRevA.79.063829`.

54. CONSTANT, E.; GARZELLA, D.; BREGER, P.; MÉVEL, E.; DORRER, C.; LE BLANC, C.; SALIN, F.; AGOSTINI, P. Optimizing High Harmonic Generation in Absorbing Gases: Model and Experiment. *Physical Review Letters*. 1999, **82**, 1668–1671. Available from DOI: `10.1103/PhysRevLett.82.1668`.

55. KAZAMIAS, S. et al. Global Optimization of High Harmonic Generation. *Physical Review Letters*. 2003, **90**, 193901. Available from DOI: `10.1103/PhysRevLett.90.193901`.

56. YAKOVLEV, V. S.; IVANOV, M.; KRAUSZ, F. Enhanced phase-matching for generation of soft X-ray harmonics and attosecond pulses in atomic gases. *Optics Express*. 2007, **15**(23), 15351–15364. Available from DOI: `10.1364/OE.15.015351`.

57. CORKUM, P. B. Plasma perspective on strong field multiphoton ionization. *Physical Review Letters*. 1993, **71**, 1994–1997. Available from DOI: `10.1103/PhysRevLett.71.1994`.

58. KERN, C.; ZUERCH, M.; SPIELMANN, C. Limitations of Extreme Nonlinear Ultrafast Nanophotonics. *Nanophotonics*. 2015, **4**. Available from DOI: `10.1515/nanoph-2015-0013`.

59. BERGÉ, L; SKUPIN, S; NUTER, R; KASPARIAN, J; WOLF, J-P. Ultrashort filaments of light in weakly ionized, optically transparent media. *Reports on Progress in Physics*. 2007, **70**(10), 1633–1713. Available from DOI: `10.1088/0034-4885/70/10/r03`.

60. YEE, K. S. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation*. 1966, **14**, 302–307.

61. LORIN, E.; CHELKOWSKI, S.; BANDRAUK, A. A numerical Maxwell–Schrödinger model for intense laser–matter interaction and propagation. *Computer Physics Communications*. 2007, **177**(12), 908–932. ISSN 0010-4655. Available from DOI: `10.1016/j.cpc.2007.07.005`.

62. KOLESIK, M.; MOLONEY, J. V. Nonlinear optical pulse propagation simulation: From Maxwell's to unidirectional equations. *Physical Review E*. 2004, **70**, 036604. Available from DOI: `10.1103/PhysRevE.70.036604`.

63. BERTI, N.; BÉJOT, P.; CORMIER, E.; KASPARIAN, J.; FAUCHER, O.; WOLF, J.-P. Ab initio calculations of laser-atom interactions revealing harmonics feedback during macroscopic propagation. *Physical Review A*. 2019, **99**, 061401. Available from DOI: `10.1103/PhysRevA.99.061401`.

64. SPOTT, A.; JARO Ń-BECKER, A.; BECKER, A. Ab initio and perturbative calculations of the electric susceptibility of atomic hydrogen. *Physical Review A*. 2014, **90**, 013426. Available from DOI: `10.1103/PhysRevA.90.013426`.

65. BÉJOT, P.; CORMIER, E.; HERTZ, E.; LAVOREL, B.; KASPARIAN, J.; WOLF, J.-P.; FAUCHER, O. High-Field Quantum Calculation Reveals Time-Dependent Negative Kerr Contribution. *Physical Review Letters*. 2013, **110**, 043902. Available from DOI: `10.1103/PhysRevLett.110.043902`.

66. KÖHLER, C.; GUICHARD, R.; LORIN, E.; CHELKOWSKI, S.; BANDRAUK, A. D.; BERGÉ, L.; SKUPIN, S. Saturation of the nonlinear refractive index in atomic gases. *Physical Review A*. 2013, **87**, 043811. Available from DOI: `10.1103/PhysRevA.87.043811`.

67. NURHUDA, M.; SUDA, A.; MIDORIKAWA, K. Generalization of the Kerr effect for high intensity, ultrashort laser pulses. *New Journal of Physics*. 2008, **10**(5), 053006. Available from DOI: `10.1088/1367-2630/10/5/053006`.

68. SELLMEIER, W. Ueber die durch die Aetherschwingungen erregten Mitschwingungen der Körpertheilchen und deren Rückwirkung auf die ersteren, besonders zur Erklärung der Dispersion und ihrer Anomalien. *Annalen der Physik*. 1872, **223**(11), 386–403. Available from DOI: `10.1002/andp.18722231105`.

69. VIRTANEN, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*. 2020, **17**, 261–272. Available from DOI: `10.1038/s41592-019-0686-2`.

70. DALGARNO, A.; KINGSTON, A. E.; BATES, D. R. The refractive indices and Verdet constants of the inert gases. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*. 1960, **259**(1298), 424–431. Available from DOI: `10.1098/rspa.1960.0237`.

71. PECK, E. R.; FISHER, D. J. Dispersion of Argon. *Journal of the Optical Society of America*. 1964, **54**(11), 1362–1364. Available from DOI: `10.1364/JOSA.54.001362`.

72. AMMOSOV, M. V.; DELONE, N. B.; KRAINOV, V. P. Tunnel Ionization Of Complex Atoms And Atomic Ions In Electromagnetic Field. In: ALCOCK, John A. (ed.). *High Intensity Laser Processes.* SPIE, 1986, vol. 0664, pp. 138–141. Available from DOI: `10.1117/12.938695`.

73. TONG, X. M.; ZHAO, Z. X.; LIN, C. D. Theory of molecular tunneling ionization. *Physical Review A.* 2002, **66**, 033402. Available from DOI: `10.1103/PhysRevA.66.033402`.

74. TONG, X M; LIN, C D. Empirical formula for static field ionization rates of atoms and molecules by lasers in the barrier-suppression regime. *Journal of Physics B: Atomic, Molecular and Optical Physics.* 2005, **38**(15), 2593. Available from DOI: `10.1088/0953-4075/38/15/001`.

75. FEIT, M.D; FLECK, J.A; STEIGER, A. Solution of the Schrödinger equation by a spectral method. *Journal of Computational Physics.* 1982, **47**(3), 412–433. ISSN 0021-9991. Available from DOI: `10.1016/0021-9991(82)90091-2`.

76. RESCIGNO, T. N.; MCCURDY, C. W. Numerical grid methods for quantum-mechanical scattering problems. *Physical Review A.* 2000, **62**, 032706. Available from DOI: `10.1103/PhysRevA.62.032706`.

77. HALL, R.; SAAD, N.; SEN, K.; CIFTCI, H. Energies and wave functions for a soft-core Coulomb potential. *Physical Review A.* 2009, **80**, 032507. Available from DOI: `10.1103/PhysRevA.80.032507`.

78. HE, F.; RUIZ, C.; BECKER, A. Absorbing boundaries in numerical solutions of the time-dependent Schrödinger equation on a grid using exterior complex scaling. *Physical Review A.* 2007, **75**, 053407. Available from DOI: `10.1103/PhysRevA.75.053407`.

79. VABEK, J.; BACHAU, H.; CATOIRE, F. Ionization dynamics and gauge invariance. *Physical Review A.* 2022, **106**, 053115. Available from DOI: `10.1103/PhysRevA.106.053115`.

80. FRIGO, M.; JOHNSON, S.G. The Design and Implementation of FFTW3. *Proceedings of the IEEE.* 2005, **93**(2), 216–231. Available from DOI: `10.1109/JPROC.2004.840301`.

81. CATOIRE, F. et al. Complex structure of spatially resolved high-order-harmonic spectra. *Phys. Rev. A.* 2016, **94**, 063401. Available from DOI: `10.1103/PhysRevA.94.063401`.

82. QUINTARD, L. et al. Optics-less focusing of XUV high-order harmonics. *Science Advances.* 2019, **5**(4), eaau7175. Available from DOI: `10.1126/sciadv.aau7175`.

83. VEYRINAS, K.; VABEK, J.; VALENTIN, C.; DESCAMPS, D.; PIEJOT, C.; BURGY, F.; CONSTANT, E.; MEVEL, E.; CATOIRE, F. Spectral filtering of high-order harmonics via optics-free focusing. *Opt. Express.* 2021, **29**(19), 29813–29827. Available from DOI: `10.1364/OE.436086`.

84. LORIN, E.; CHELKOWSKI, S.; ZAOUI, E.; BANDRAUK, A. Maxwell – Schrödinger – Plasma (MASP) model for laser–molecule interactions: Towards an understanding of filamentation with intense ultrashort pulses. *Physica D: Nonlinear Phenomena.* 2012, **241**(12), 1059–1071. ISSN 0167-2789. Available from DOI: `10.1016/j.physd.2012.02.013`.

85. SHERRINGTON, David; KIRKPATRICK, Scott. Solvable Model of a Spin-Glass. *Physical Review Letters.* 1975, **35**, 1792–1796. Available from DOI: `10.1103/PhysRevLett.35.1792`.

86. MCCULLOCH, W. S.; PITTS, W. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics.* 1943, **5**. Available from DOI: `10.1007/BF02478259`.

87. ROSENBLATT, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review.* 1958, **65**, 386–408. Available from DOI: `10.1037/h0042519`.

88. GOODFELLOW, I. J. et al. Generative Adversarial Networks. *arXiv.* 2014. Available from DOI: `10.48550/arXiv.1406.2661`.

89. ZHANG, C et al. A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data. *arXiv.* 2018. Available from DOI: `10.48550/arXiv.1811.08055`.

90. VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L; POLOSUKHIN, I. Attention Is All You Need. *arXiv.* 2017. Available from DOI: `10.48550/arXiv.1706.03762`.

91. JIN, X.; CAI, S.; LI, H.; KARNIADAKIS, G. E. NSFnets (Navier-Stokes flow nets): Physics-informed neural networks for the incompressible Navier-Stokes equations. *Journal of Computational Physics.* 2021, **426**, 109951. ISSN 0021-9991. Available from DOI: `10.1016/j.jcp.2020.109951`.

92. LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE.* 1998, **86**(11), 2278–2324. Available from DOI: `10.1109/5.726791`.

93. KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM.* 2017, **60**(6), 84–90. ISSN 0001-0782. Available from DOI: `10.1145/3065386`.

94. LENAIL, A. NN-SVG: Publication-Ready Neural Network Architecture Schematics. *Journal of Open Source Software.* 2019, **4**(33), 747. Available from DOI: `10.21105/joss.00747`.

95. GOH, G. Why Momentum Really Works. *Distill.* 2017. Available from DOI: `10.23915/distill.00006`.

96. KINGMA, D.; BA, J. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations.* 2014. Available from DOI: `10.48550/arXiv.1412.6980`.

97. HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep Residual Learning for Image Recognition. *arXiv.* 2015. Available from DOI: `10.48550/arXiv.1512.03385`.

98. NAKKIRAN, P.; KAPLUN, G.; BANSAL, Y.; YANG, T.; BARAK, B.; SUTSKEVER, I. Deep Double Descent: Where Bigger Models and More Data Hurt. *arXiv.* 2019. Available from DOI: `10.48550/arXiv.1912.02292`.

99. CHAERUN NISA, E.; KUAN, Y.-D. Comparative Assessment to Predict and Forecast Water-Cooled Chiller Power Consumption Using Machine Learning and Deep Learning Algorithms. *Sustainability.* 2021, **13**, 744. Available from DOI: `10.3390/su13020744`.

100. ULYANOV, D.; VEDALDI, A.; LEMPITSKY, V. Deep Image Prior. *International Journal of Computer Vision.* 2020, **128**(7), 1867–1888. ISSN 1573-1405. Available from DOI: `10.1007/s11263-020-01303-4`.

101. HARRIS, C. R. et al. Array programming with NumPy. *Nature.* 2020, **585**(7825), 357–362. Available from DOI: `10.1038/s41586-020-2649-2`.

102. PABLOS-MARÍN, J. M.; SERRANO, J.; HERNÁNDEZ-GARCÍA, C. Simulating macroscopic high-order harmonic generation driven by structured laser beams using artificial intelligence. *Computer Physics Communications.* 2023, **291**, 108823. ISSN 0010-4655. Available from DOI: `10.1016/j.cpc.2023.108823`.

103. PURWINS, H.; LI, B.; VIRTANEN, T.; SCHLÜTER, J.; CHANG, S.-Y.; SAINATH, T. Deep Learning for Audio Signal Processing. *IEEE Journal of Selected Topics in Signal Processing.* 2019, **13**(2), 206–219. Available from DOI: `10.1109/JSTSP.2019.2908700`.

104. SHERSTINSKY, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *CoRR.* 2018, **abs/1808.03314**. Available from DOI: `10.48550/arXiv.1808.03314`.

105. HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. *Neural Computation.* 1997, **9**(8), 1735–1780. ISSN 0899-7667. Available from DOI: `10.1162/neco.1997.9.8.1735`.

106. PERELOMOV, A.; POPOV, V. S.; TERENTEV, M. Ionization of Atoms in an Alternating Electric Field. *Journal of Experimental and Theoretical Physics.* 1966, **23**, 207.

107. CATOIRE, F.; SILVA, R.; RIVIÈRE, P.; BACHAU, H.; MARTÍN, F. Molecular resolvent-operator method: Electronic and nuclear dynamics in strong-field ionization. *Physical Review A.* 2014, **89**, 023415. Available from DOI: `10.1103/PhysRevA.89.023415`.

108. SAK, H.; SENIOR, A. W.; BEAUFAYS, F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: *INTERSPEECH.* 2014, pp. 338–342.

109. GRAVES, A.; SCHMIDHUBER, J. Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks. In: *Advances in Neural Information Processing Systems.* Curran Associates, Inc., 2008, vol. 21.

110. VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. Attention is All you Need. In: *Advances in Neural Information Processing Systems.* Curran Associates, Inc., 2017, vol. 30. Available also from: `https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf`.

Theses

111. NĚMEC, T. *Generace vysokých harmonických frekvencí pomocí dvoubarevného pole.* Praha, 2021. Bachelor thesis. ČVUT FJFI.

112. VÁBEK, J. *Multiscale approach to the description of high-harmonics generation in gases*. Prague, 2022. Doctoral thesis. ČVUT FJFI, Université de Bordeaux.

113. SKUPIN, S. *Nonlinear Dynamics of Trapped Beams*. Jena, 2005. Doctoral thesis. Physikalisch-Astronomischen Fakultät der Friedrich-Schiller-Universität Jena.

Online sources

114. *Atomic and Molecular Calculations are Expressed in Atomic Units* [online]. LibreTexts, 2020 [visited on 2021-05-06]. Available from: `https://chem.libretexts.org/@go/page/210838`.

115. *Experiments with light capture the shortest of moments* [online]. The Nobel Prize - The Royal Swedish Academy of Sciences, 2023 [visited on 2023-11-05]. Available from: `https://nobelprize.org/prizes/physics/2023/press-release`.

116. *Welcome to the Gemini era* [online]. Google DeepMind, 2023 [visited on 2023-12-13]. Available from: `https://deepmind.google/technologies/gemini/#introduction`.

117. *DALL-E 3* [online]. OpenAI, 2023 [visited on 2023-10-21]. Available from: `https://openai.com/dall-e-3`.

118. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems* [online]. Tensorflow, 2015 [visited on 2023-12-13]. Available from: `https://www.tensorflow.org/`.

119. *A short introduction to numerical methods for Maxwell's equations* [online]. Christian Wieners, 2008 [visited on 2023-12-28]. Available from: `https://www.math.kit.edu/user/~wieners/MaxwellCourse.pdf`.

120. *Audio Deep Learning Made Simple (Part 1): State-of-the-Art Techniques* [online]. Medium – Towards Data Science, 2021 [visited on 2023-11-28]. Available from: `https://towardsdatascience.com/audio-deep-learning-made-simple-part-1-state-of-the-art-techniques-da1d3dff2504`.

Appendix A

Retrieving ionization from the 1D-TDSE

The problem of modeling of photoionization rate expressions such as ADK [72] or PPT [106] is their lack of validity across a wide intensity range. To precisely retrieve ionization, we can employ the 1D-TDSE. One option is to retrieve the free electron density in time using the resolvent operator method (ROM) [107] for various ionization regimes [79]. The application of the method for the tunneling regime is depicted in Figure A.1. The second option is to compute the free electron density from the wavefunction directly once the pulse propagated through the medium. This refers to the yellow continuum curve depicted in Figure 2.4. Let us briefly introduce the method.

The final wavefunction at the end of the pulse $|\psi_f\rangle$ is composed of bound (b) and continuum (c) states[1]

$$|\psi_f\rangle = |\psi_b\rangle + |\psi_c\rangle \,. \tag{A.1}$$

By projecting $|\psi_f\rangle$ onto the bound states labelled with principal quantum numbers $n \in \{1, 2, \ldots, N\}$ we get the population of electrons in continuum as

$$p_c = 1 - \langle\psi_f| \, P_b \, |\psi_f\rangle \,, \tag{A.2}$$

where the projector onto the bound states is defined as

$$P_b = \sum_{n=1}^{N} |\psi_n\rangle \, \langle\psi_n| \,. \tag{A.3}$$

The population density of the individual state n is given as

$$p_n = |\, \langle\psi_n|\psi_f\rangle \,|^2. \tag{A.4}$$

For the computation of the continuum density, the first 16 bound states were used (see Algorithm 1 for details). The first 5 bound states of argon and their corresponding energies are plotted in Figure A.2. The important conclusion is that the ionization rate computed using the modified ADK [72, 73, 74], depicted in Figure 2.4, copies the continuum population trend. The factor difference is attributed to the fact that the ADK rate assumes a 3D geometry.

---

[1] Projection of the wavefunction on bound and continuum states requires extra caution depending on the employed gauge, see [79].

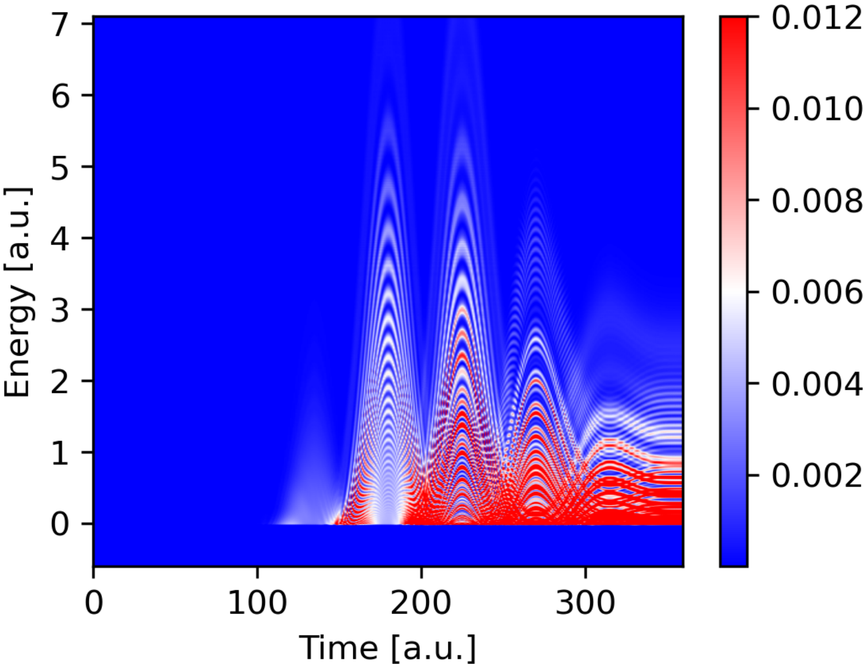

Figure A.1: Time-dependent probability density resolved in energy computed using ROM. Computed for a 4-cycle $\sin^2$-pulse of central frequency ω_0 and field amplitude $E_0 = 0.14$ a.u.. The figure shows the energy distribution of electrons in the continuum in time. The first 16 bound states were subtracted from the wavefunction. We recreated Figure 6 from [79] using the 1D-TDSE solver, introduced in 3.1, to illustrate the method.

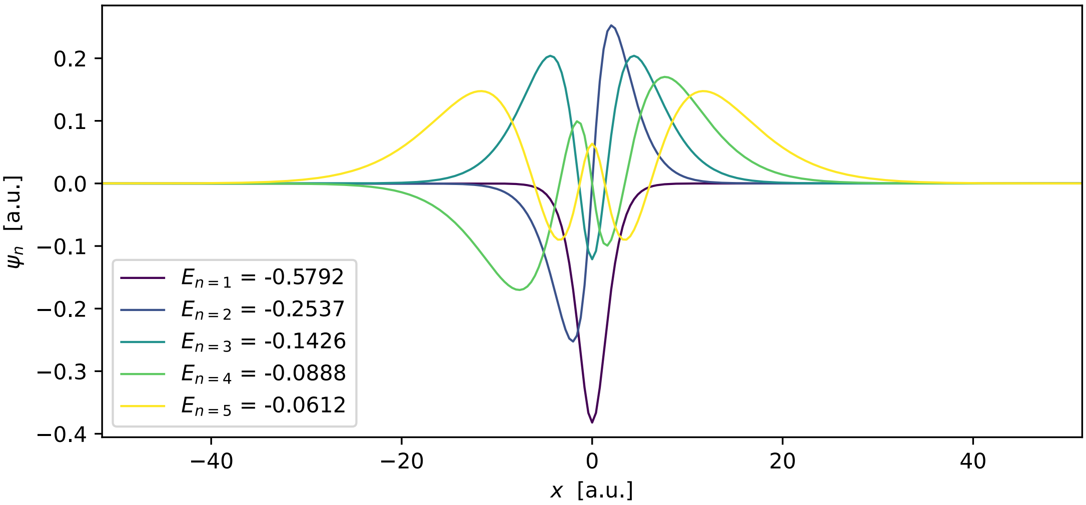

Figure A.2: Normalised first 5 bound states of argon with their corresponding energies in atomic units.

Appendix B

Neural network dead ends

In this appendix, we elaborate further on some tested approaches and problems encountered along the way when dealing with the neural network design.

B.1 Amplitude and phase training

Separating explicitly the amplitude and phase of the positive spectrum appeared as a viable option from the start. The reason is the ability to enhance the low-amplitude features for better comprehension by the network. The idea is as follows. Say we have a field F and its corresponding spectrum $\hat{F}$. We can then write field F in Fourier terms as:

$$F = \sum_{\omega} \hat{F}_{\omega} e^{-i\omega t} = \sum_{\omega} |\hat{F}_{\omega}| e^{-i\varphi_{\omega}} \tag{B.1}$$

where we use the amplitude $|\hat{F}_{\omega}|$ and phase φ_{ω} for the inputs and outputs. Moreover, using the following identity

$$|\hat{F}_{\omega}| = \exp\left(\log\left(|\hat{F}_{\omega}|\right)\right) \tag{B.2}$$

we can pick the logarithmically scaled amplitude $\log\left(|\hat{F}_{\omega}|\right)$, which enhances the low-amplitude features, along with the phase φ. Because the phase of the FFT is discontinuous, between $(-\pi, \pi)$, we used the phase unwrapping to create a continuous phase curve.

The drawback of this procedure was the inability to train such a network that would be sensitive enough on the phase that is well defined only where the amplitude of the harmonic peak is meaningful, see 4.1. The phase of the individual field components is crucial for the determination of the moment when a particular frequency is generated. Our initial designs of NNs were incapable of dealing with this task.

B.2 N-th root feature extraction

Using the real and imaginary parts of the spectrum comes with the issue of having the spectral features vary by several orders of magnitude. The way to tackle this problem is to provide a nonlinear transform of computing odd roots of individual parts. The network is then fed by a pair of inputs $(\sqrt[n]{\mathrm{Re}[F]}, \sqrt[n]{\mathrm{Im}[F]})$ where n is odd integer. Assuming the definition of a square root with odd roots for negative numbers, there is no problem with the application of this idea. Applying the same principle on the outputs and applying n-th power on labels then provides the unscaled results. However, during the testing and experimentation, it became obvious that such an approach disproportionally enhances noise and induces oscillations, further degrading the feature extraction which was the goal in the first place. Some spectral features in predicted labels were significantly deformed by the inverse transform due to the high nonlinearity of the n-th root.

B.3 Choice of activations

As presented in the theoretical overview, Section 5.1.1, we have multiple choices for the activation functions. ReLU is very popular due to its simplicity and efficiency for larger networks composed of hundreds of layers. However, during the testing, it proved to be incapable of coping with our highly nonlinear problem. Sigmoid is sufficiently nonlinear but converges slowly and has a limited output range of $(0, 1)$. For these reasons, we settled with using the $\tanh(x)$ activation. With high nonlinearity and range $(-1, 1)$ it is the best candidate for solving the problem. The activation for the output layer must be selected with special caution as we need to retain an adequate output range. The only suitable candidates are $\tanh(x)$ or $\mathrm{id}(x)$ but we stick to the latter simply because we do not want to restrict the output range in any way.

B.4 Assessment of TDSE-NN designs and trials and errors

Numerous NN designs and improvements were tried and tested:

- **Deep MLP**: MLP network (no CNN) with many dense layers (up to 10) of size around 100 nodes. Problems: longer training time, high generalization error, very high dropout rate necessary (around $0.8 - 0.9$).
- **Shallow MLP**: MLP network with a few dense layers of higher size (≥ 256). Problems: longer training time, inability to retrieve features from the data, and the output was extremely noisy.
- **Deep CNN**: CNN network with more than 3 1D convolution layers. Problem: poor performance – the output data of the convolution was lacking useful information.
- **Compression-decompression with high compression ratio**: compression ratio in the compression-decompression part of MLP higher than 4:1 or narrower bottleneck (layer size < 32) resulted in a very noisy output.
- **Gaussian noise**: setting higher values of noise ($\mu > 0.2$) resulted in poor training and inference of the model.

During the experimentation with various models, it was found that CNN regularized the network as it extracted some lower-level features. Another improvement was made by the implementation of the batch normalization layer right after the flattening layer. Normalizing the data in the range $(-1, 1)$ also made significant progress in training and inferring.

Using the skipped connection was a huge improvement. The previous designs were very insensitive to the phase of the input field. This resulted in a significant temporal shift of the NN output in the temporal domain. A Skipped connection was able to keep some of the very important features in the data during the forward propagation. As a result, the network generalized better.

So far we have discussed only a limited class of neural network architectures. We briefly examined recurrent neural networks (RNN) and their successors long short-term memory (LSTM) networks which are types of bi-directional neural networks where the data flows in both directions [104, 105]. They are mainly used to predict sequences of data given the sequences of variable lengths and are most suited for natural language processing tasks [108] or handwriting recognition [109]. RNNs and LSTMs were overthrown by transformer architecture [110]. Transformers do not require any recurrent units while reaching better performance with natural language processing tasks which has led to the rise of large language models [116, 23]. However, RNNs, LSTMs and transformers are typically too large, overly complicated models and costly to train. Therefore we decided to stick with conservative usage of CNNs and MLPs.