RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Université de Toulouse

Doctorat de l'Université de Toulouse

préparé à l'Université Toulouse III - Paul Sabatier

Développement d'une méthode numérique multiéchelle pour les plasmas atmosphériques et application au contrôle d'écoulements

---

Thèse présentée et soutenue, le 6 décembre 2023 par

Tuan Dung NGUYEN

École doctorale
EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse

Spécialité
Mathématiques et Applications

Unité de recherche
IMT : Institut de Mathématiques de Toulouse

Thèse dirigée par
Christophe BESSE et François ROGIER

Composition du jury
M. Michel MEHRENBERGER, Rapporteur, Aix-Marseille Universite
Mme Marie-Hélène VIGNAL, Examinatrice, Université Toulouse III - Paul Sabatier
M. Konstantinos KOURTZANIDIS, Examinateur, Chemical Process & Energy Resources Institute (CPERI), Centre for Research & Technology Hellas (CERTH)
M. Christophe BESSE, Directeur de thèse, CNRS Toulouse - IMT
M. François ROGIER, Co-directeur de thèse, ONERA
M. Raphaël LOUBÈRE, Président, CNRS, Institut de Mathématiques de Bordeaux

Acknowledgments

This PhD. offered me a chance to appreciate what it takes to become a researcher. It is a long, hard road and only the one that walked through it could really feel how they have changed. For me it was magnificent. And for that reason, I would like to express the greatest gratitude to my two mentors - Prof. Christophe Besse and Dr. François Rogier.

Prof. Besse was the director of this thesis and helped me enormously with many things that was going on during the last three years, whether it was research-related or administration-related. It was fortunate for me to know and work with Prof. Besse, as I learned a great deal from his guidance and demeanors. I truly respect his clarity, transparency and rigor during all the discussions that we had throughout this PhD. candidacy.

Dr. Rogier was my day-to-day mentor at ONERA. It struck me after a while that he was more than a supervisor, that he gave me an impression of being close to a colleague of him. I will always be grateful for his calm and beautiful personality, and I will always cherish his professional as well as personal advice, his encouragements during the hardest times, and the countless hours in our offices trying to figure out ways to deal with problems.

I am honored to have worked with both Prof. Besse and Dr. Rogier. They showed unparalleled patience to me from even before the beginning, when I started the project with a Master internship during the lockdowns. I would also like to extend my gratitude to Dr. Guillaume Dufour, who co-mentored the internship and also helped me a lot, especially with numerical aspects, during my time at ONERA.

I would like to express my sincere gratitude to the two reviewers - Dr. Raphaël Loubère and Prof. Michel Mehrenberger - for having dedicated their time and professional opinions to evaluate my work. I am obliged to thank Dr. Loubère in particular for the fruitful discussions during the conferences in France and Portugal. I would also like to thank the two examiners for having accepted to participate in my defense - Dr. Marie-Hélène Vignal, whom I also consulted for professional advice from time to time, and Dr. Konstantinos Kourtzanidis, who also guided me a lot in the physics of plasma actuators.

I am thankful to all the researchers and peers at ONERA and IMT in Toulouse, CWI in Amsterdam and to other friends in Toulouse that I had a chance to meet. Special mentions go to Hiếu, Hà, Lê Minh, Hà Trang who I have known long before this thesis and Đức Anh with whom I shared many coffee breaks at ONERA.

I am especially grateful to Tú Quỳnh for having been on my side for a decade now. Your sharings and understandings are priceless to me. I cannot imagine how hard it would have been had you not been there for me.

Finally, I dedicate this thesis to my parents and sister. There are absolutely no words in this world that can describe what you guys have done for me. Cảm ơn mọi người vì tất cả.

Contents

Acknowledgments i

Table of Contents iii

List of Figures v

List of Tables vii

List of Symbols 1

Introduction 1

I Physical, Mathematical and Numerical Background on Atmospheric Pressure Discharge 13

1 Notes on Gas Discharge Physics and Plasma Actuators 15

1.1 What is a plasma? . . . 16

1.2 Low-temperature discharges in air . . . 21

1.3 Corona and dielectric barrier discharges . . . 32

2 Mathematical Modeling of Electric Discharge in Air 47

2.1 Macroscopic description of gas discharge dynamics and drift-diffusion-Poisson equations 48

2.2 Boundary conditions . . . 53

2.3 Initial data . . . 55

2.4 Summary of the discharge models . . . 55

2.5 Validity limit and improvements of the LFA model . . . 56

3 Numerical Modeling of Electric Discharge in Air - the COPAIER Plasma Solver 61

3.1 Meshing of the computation domain . . . 62

3.2 Discretization of the conservation laws . . . 63

3.3 Approximation of the potential and the plasma coefficients . . . 67

II Contributions to Mathematical and Numerical Modeling of Atmospheric Pressure Discharge 71

4 High-order Scharfetter-Gummel Schemes: Derivation, Properties and Extension on Two-dimensional Grids 73

4.0 Aperçu . . . 74

4.1 Overview . . . 76

4.2 The Scharfetter-Gummel scheme . . . 78

4.3 Construction of high-order Scharfetter-Gummel schemes . . . 81

4.4 Properties of the point-wise SGCC-p flux . . . 87

4.5 Extension of the SGCC-p flux scheme on two-dimensional grids . . . 92

4.6 High-order reconstruction of particle density . . . 94

4.7 Choices of slope limiters . . . 100

4.8 Numerical verification . . . 105

4.9 Closing remarks . . . 111

4.10 Remarques finales . . . 113

5 High-order Scharfetter-Gummel Schemes: Application to Simulation of Low-temperature Gas Discharges 115

5.0 Aperçu . . . 116

5.1 Overview . . . 116

5.2 Simulations of corona discharge on one-dimensional grids . . . 117

5.3 Simulation of streamer propagation on two-dimensional cartesian grids . . . 123

5.4 Closing remarks . . . 137

5.5 Remarques finales . . . 139

6 Mathematical and Numerical Modeling of Corona Discharge 141

6.0 Aperçu . . . 142

6.1 Overview . . . 143

6.2 Formulation of the gas discharge model with floor density . . . 145
6.3 Existence and uniqueness of the solution of the differential inclusion . . . 147
6.4 Implicit time discretization and treatment of the source terms . . . 160
6.5 Solving the conservation laws . . . 164
6.6 Comparison of algorithms on one-dimensional grids . . . 169
6.7 Closing remarks . . . 174
6.8 Remarques finales . . . 174

7 Simulations of Gas Discharge in Air with the Plasma Solver COPAIER 177
7.0 Aperçu . . . 178
7.1 Overview . . . 178
7.2 Wire-to-wire corona discharge . . . 179
7.3 Needle-to-ring corona discharge . . . 186
7.4 Closing remarks . . . 199
7.5 Remarques finales . . . 199

III Conclusion and Prospects 201

Appendices 209
A Notations on multi-dimensional arrays - tensors . . . 213
B A brief presentation on total variation diminishing schemes . . . 214
C Monotone operators and some useful results from functional analysis . . . 215
D Regularity results for problems (6.2) and (6.3) . . . 217
E Implicit integration of the SGCC schemes . . . 218
F MPI implementation of the Gauss-Seidel algorithm . . . 219

Bibliography 221

List of Figures

1 Écoulement naturel (à gauche) et effet des actionneurs électriques sur l'écoulement (à droite) autour d'un profil aérodynamique, extrait de [55] 1

2 Natural flow (left) and effect of electric actuators on the flow (right) around an airfoil, taken from [55] 7

1.1 Debye shielding of a positive charge (maroon) by negative ones (blue) 18

1.2 Particle displacements from "equilibrium" (left) to charge separation (right) that lead to plasma oscillations 18

1.3 Classification of plasmas based on electron temperature and plasma density, taken from [80] 20

1.4 Townsend ionization coefficients according to the empirical formula (1.5) 22

1.5 Sketch of a Townsend tube 26

1.6 The Paschen curve for a Townsend discharge in air 27

1.7 A typical V-I curve in a Townsend tube 28

1.8 Photographs of some discharge regimes (not necessarily in a Townsend tube) . . . 29

1.9 Photographs of a spark, taken from [107] 29

1.10 Photographs of a positive streamer (left) and a negative streamer (right), taken from [28] 30

1.11 Streamer propagation mechanisms in gap between the anode (A) and the cathode (C). The positive charges are illustrated in maroon, negative charges in blue, photons in green and quasi-neutral ionized gas in purple. 31

1.12 Photograph of a wire-to-wire corona discharge: luminous glow is observed in the vicinity of both wires, but brighter near the smaller one (the stressed electrode). Photo taken from [109]. 33

1.13 Schematics of some corona discharge actuators 33

1.14 Positive and negative corona discharge regimes, taken from [124]. Note that spark was also counted. 33

1.15 Waveform of Trichel pulses at $p = 225$ Torr in *(a)* pure Ar, *(b)* Ar-1%O_2, *(c)* pure N_2, *(d)* N_2-1%O_2 and *(e)* air, taken from [164] . . . 35

1.16 Photograph of a SDBD, taken from [109] . . . 36

1.17 Illustration of some DBD actuators. Electrodes are colored in black, dielectric barriers in blue. . . . 36

1.18 Schematics of a plate-to-plate SDBD actuator, taken from [80] . . . 37

1.19 Velocity profile during a voltage cycle for a wire-to-plate discharge and a plate-to-plate discharge, taken from [42] . . . 37

1.20 Typical electric current during a sinusoidal cycle of a SDBD, taken from [109] . . . 38

1.21 High speed photography of a sinusoidal SDBD in the middle of the negative-going cycle (above) and the positive-going cycle (below), taken from [49] . . . 39

1.22 Total surface charge in a SDBD in function of time, taken from [73]. Note that the left axis measures the surface charge (nC) while the right axis measures the voltage (V). 40

1.23 Surface charges after cutting off the generator of a 15 kV, 5 kHz sinusoidal voltage, taken from [40] . . . 40

1.24 Applied voltage and computed electric current in function of time with the voltage slope $\eta_V = 100$ V μs^{-1}, taken from [19] . . . 42

1.25 Parallel and perpendicular components (to the dielectric surface) of $\boldsymbol{F}_{\mathrm{EHD}}/L$ and $\boldsymbol{S}_{\mathrm{EHD}}/L$ in function of time, taken from [19]. . . . 43

1.26 Applied voltage and computed electric current (in absolute values) in function of time with the voltage slope $\eta_V = -300$ V μs^{-1}, taken from [87] . . . 43

1.27 Applied voltage and computed electric current in function of time during the third voltage period (10 kV, 100 kHz), taken from [81] . . . 44

1.28 Charge per surface unit before (in blue) and after (red) the positive pulse and the first negative pulse, taken from [81]. . . . 45

1.29 Magnitude of the time-averaged EHD force density over one sinusoidal period, taken from [81]. The force in the red sheath around the bare electrode directs towards it due to the absorption of positive ions into the cathode during the negative discharge. This point is coherent with earlier experimental observations that the air entrained by the actuator comes from above the bare electrode [123]. . . . 46

2.1 Computation domains of a SDBD. The electrodes are colored in black, the dielectric barrier in blue and the gas medium in brown. The domain Ω of n_s (left figure) are delimited by thick brown lines, the domain Ω_ϕ of ϕ (right) are delimited by thick black lines. . . . 49

2.2 Nonphysical growth of electron population computed with the LFA model, taken from [138] . . . 56

3.1 Two neighbor cells of a triangular conforming grid . . . 63
3.2 Dual cell (gray) $\Omega_K^{\mathcal{D}}$ of a triangle Ω_K (maroon) . . . 65

4.1 A one-dimensional grid with cell centers illustrated by big dots and cell interfaces illustrated by vertical strokes . . . 78
4.2 Geometric notations . . . 93
4.3 Ω_K and its neighbors . . . 95
4.4 Border cells (maroon) and ghost cells (blue) . . . 96
4.5 Some limiters with the TVD region in maroon and $\beta = 1.5$ for Φ^{MC} . . . 102
4.6 Test 4.1. L^1-norm errors $e_{\Delta x}$ in function of the CPU time. . . . 108
4.7 Test 4.2. Numerical and benchmark solutions at $T = 4 \times 10^{-5}$ for $x \in (0.3, 0.6)$. . . 109

5.1 Schematics of the wire-to-wire actuator . . . 117
5.2 Evolution of the field strength and specie densities of the ^{400}SGCC-2 simulation . . 121
5.3 Comparison of the circuit current I of different schemes and grid sizes, on 0-4 ms (left) and on 110-290 μs (right) . . . 122
5.4 Geometry of the discharge with the computation domain in brown (left) and initial electric field with equipotential lines (right) . . . 124
5.5 (SS) Evolution in time of the electric field strength E (MV m^{-1}) of ^{3}SGCC-2. Visualization of the field is symmetrized by flipping on the axis $r = 0$ and only shown up to $r = 2$ mm. . . . 126
5.6 (SS) Numerical results of ^{3}SGCC-2 except the dashed line on the left figure which is obtained by ^{3}SGCC-1 . . . 127
5.7 (SS) Numerical results of the SG, MUSCL, SGCC-1 and SGCC-2 schemes as well as of the teams CWI [140] and FR [48]. We subtract νt from $L(t)$, with $\nu = 0.05$ cm ns^{-1}, to enhance the difference between the curves. . . . 129
5.8 (SS) Relative errors on the streamer length L generated by the SGCC-1 and SGCC-2 schemes . . . 130
5.9 (SS) $\mathcal{T}^{\text{CPU}}$-$\mathcal{E}^{\text{tot}}$ curves of the SG, MUSCL, SGCC-1 and SGCC-2 schemes with respect to the benchmarks $^{0.8}$CWI and $^{0.8}$FR . . . 131
5.10 (WOL) Numerical results of the SGCC-1,2,3,4,5 schemes as well as of the Strang splitting $^{0.8}$SGCC-2 simulation in section 5.3.2 as benchmark. We subtract νt from $L(t)$, with $\nu = 0.05$ cm ns^{-1}, to enhance the difference between the curves. . . . 132
5.11 (WOL) Relative errors on the streamer length L generated by the SGCC-1,2,3,4,5 schemes . . . 133

5.12 (WL) Numerical results of the SGCC-1,2,3,4,5 schemes as well as of the Strang splitting $^{0.8}$SGCC-2 simulation in section 5.3.2 as benchmark. We subtract νt from $L(t)$, with $\nu = 0.05$ cm ns^{-1}, to enhance the difference between the curves. 134

5.13 (WL) Relative errors on the streamer length L generated by the SGCC-1,2,3,4,5 schemes 134

5.14 (SSL) Numerical results of $^{1.5}$SGCC-1 and $^{1.5}$SGCC-2 at $T = 20$ ns 135

5.15 (SSL) Numerical results of the SG, MUSCL, SGCC-1 and SGCC-2 schemes as well as of the teams CWI [140] and FR [48]. We subtract νt from $L(t)$, with $\nu = 0.03$ cm ns^{-1}, to enhance the difference between the curves. 136

5.16 (SSL) Relative errors on the streamer length L generated by the SGCC-1 and SGCC-2 schemes 137

5.17 (SSL) $\mathcal{T}^{\text{CPU}}$-$\mathcal{E}^{\text{tot}}$ curves of the SG, MUSCL, SGCC-1 and SGCC-2 schemes with respect to the benchmarks $^{0.8}$CWI and $^{0.8}$FR 138

6.1 Rate coefficients of the three-specie, four-reaction kinetic model 162

6.2 Lie splitting: electric current I with different values of $\mathfrak{C}$ 170

6.3 Douglas-Rachford algorithm: electric current I with different values of $\mathfrak{C}$ 171

6.4 Gauss-Seidel algorithm: electric current I with different values of $\mathfrak{C}$ 171

6.5 Standard SG scheme: electric current I with different mesh size $\Delta x_{\min}$ 173

6.6 SGCC-1 scheme: electric current I with different mesh size $\Delta x_{\min}$ 173

6.7 SGCC-2 scheme: electric current I with different mesh size $\Delta x_{\min}$ 173

7.1 Sketch of the geometry of the actuator (not drawn to scale) and the computation domain Ω (inside the solid-line loop) 180

7.2 Grid refinement near the smaller wire (the left one on fig. 7.1) generated by Gmsh [58] 180

7.3 Positive ion density n_p between the electrodes on the symmetry axis 182

7.4 Field strength E between the electrodes on the symmetry axis 182

7.5 EHD force strength f_{EHD} (N m^{-3}) at $t = 1.17$ μs near the anode 182

7.6 EHD force strength f_{EHD} (N m^{-3}) at $t = 2.23$ μs 183

7.7 EHD force strength f_{EHD} (N m^{-3}) at $t = 3.32$ μs 183

7.8 x_1-component f_1 of the EHD force density (N m^{-3}) at $t = 4$ ms (figure reflected on the symmetry axis) 183

7.9 ψ-I curves for $r_1 = 0.1$ mm. The curve of $V_G = 13$ kV & $d = 10$ mm is colored in blue, $V_G = 20$ kV & $d = 20$ mm in orange, $V_G = 37$ kV & $d = 40$ mm in green. . . 184

7.10 ψ-I curves for $d = 10$ mm. The curve of $r_1 = 0.1$ mm & $V_G = 13$ kV is colored in blue, $r_1 = 0.5$ mm & $V_G = 16.5$ kV in orange. 185

7.11 V-I curves for $d = 10$ mm and $r_1 = 0.1$ to 0.5 mm of experiment data (hollow circles) and numerical solutions (squares for $\psi = 10^{10}$, triangles for $\psi = 10^{-11}$ m^{-3}) 185

7.12 Sketch of the needle-to-ring actuator (not drawn to scale), the computation domain is colored in maroon . . . 187

7.13 Grid refinement near the needle tip generated by Gmsh [58] . . . 187

7.14 Wedge-like domain for the computation of $\boldsymbol{u}$ . . . 189

7.15 Positive ion density n_p (m^{-3}) near the needle tip for $\Delta V_G = 12$ kV and $\psi = 10^{11}$ m^{-3} 190

7.16 Radial and axial components f_r and f_z (N m^{-3}) of the EHD force density near the needle tip for $\Delta V_G = 12$ kV and $\psi = 10^{11}$ m^{-3} . . . 190

7.17 Radial and axial components u_r and u_z (m s^{-1}) of the flow velocity near the needle tip obtained with the turbulence model, for $\Delta V_G = 12$ kV and $\psi = 10^{11}$ m^{-3} . . . 191

7.18 Magnitude of flow velocity $u = |\boldsymbol{u}|$ (m s^{-1}) obtained with the laminar model and the turbulence model (eqs. (7.2) to (7.6)), for $\Delta V_G = 12$ kV and $\psi = 10^{11}$ m^{-3} . . . 192

7.19 V-I curves for $d = 20$ mm. Comparison between experiment data [163] and numerical solutions for different values of ψ. The dotted line is the linear fitting of experiment data. . . . 192

7.20 V-I curves for $\psi = 10^{11}$ m^{-3}. Comparison between experiment data [163] (hollow markers) and numerical solutions for different values of d (filled markers). The dotted lines are the linear fittings of experiment data. . . . 193

7.21 Ionic wind profile at $z = -(\sigma + d + h + l)$ (i.e. 5 mm downstream of the ring) for $d = 20$ mm and $\psi = 10^{11}$ m^{-3}. Comparison between experiment data [163] and numerical solutions obtained with laminar and turbulence models for different values of I_t. . . . 194

7.22 V-u curves for $I_t = 30\%$. Comparison between experiment data [163] (hollow markers) and numerical solutions for different values of d (filled markers). . . . 195

7.23 Positive ion density n_p (m^{-3}) near the needle tip, obtained with the LFA-NAS model (left) and the LFA-AS method (right) right before they crash . . . 196

7.24 Evolution of the positive ion density n_p (m^{-3}) near the needle tip during the first ten nanoseconds, obtained with the LFA-AS-FFS method . . . 197

7.25 Circuit current I obtained with the LFA-AS-FFS method . . . 197

7.26 Positive ion density n_p (m^{-3}) near the needle tip, obtained with the LFA-AS-FFS method (left) and the LEA-AS method (right) right before the latter crashes . . . 198

7.27 Comparison between the circuit currents I (log-scale of absolute values) obtained with the LFA-AS-FFS method and the LFA-AS-FFS-SI method . . . 198

28 In this figure the computation domain is partitioned into two subdomains, sharing the thick black line as interface. The triangles which share an edge with the interface are called border cells and are painted in yellow. The other triangles are called interior cells and are painted in red. 219

List of Tables

1 Échelles de temps caractéristiques des phénomènes dans une décharge de plasma, tableau reproduit de [46] . . . 3

2 Characteristic timescales of phenomena in a plasma discharge, reproduced from [46] 9

1.1 $\xi\frac{\nu_p}{\nu_i}$ in function of $\frac{E}{p}$ ($\mathrm{V\ m^{-1}\ Torr^{-1}}$) [166] . . . 24

1.2 Fitting parameters for the three-exponential Helmholtz model [21] . . . 24

2.1 Rate coefficients of the three-specie, four-reaction kinetic model . . . 50

2.2 Rate coefficients of the five-specie, seven-reaction kinetic model (only available for the LFA model) . . . 51

4.1 Gauss-Legendre quadrature rules for $Q = 1$ to 4. . . . 92

4.2 Test 4.1, $D = 10^{-6}$. Convergence order in L^1-norm, errors in function of number of grid cells $\mathcal{N}$. . . . 107

4.3 Test 4.1, $D = 10^{-4}$. Convergence order in L^1-norm. . . . 107

4.4 Test 4.1, $D = 10^{-2}$. Convergence order in L^1-norm. . . . 107

4.5 Test 4.3, **with** limiter. L^1-error and convergence order for $D = 10^{-6}$. . . . 110

4.6 Test 4.3, **with** limiter. L^1-error and convergence order for $D = 10^{-4}$. . . . 110

4.7 Test 4.3, **without** limiter. L^1-error and convergence order for $D = 10^{-6}$. . . . 111

4.8 Test 4.3, **without** limiter. L^1-error and convergence order for $D = 10^{-4}$. . . . 111

5.1 Parameters for one-dimensional wire-to-wire discharge simulations . . . 119

5.2 Parameters of the streamer propagation simulations . . . 125

5.3 (SS) Evolution in time of $\rho^{\max}$ ($\mathrm{C\ m^{-3}}$), R^{ρ} (μm) and $E^{\max}$ ($\mathrm{MV\ m^{-1}}$) obtained with $^{1.5}$SGCC-1, ^{3}SGCC-2 and $^{1.5}$SGCC-2 . . . 127

5.4 (SS) CPU time T^{CPU} (hours), simulation time T (ns) and CPU processor type used in the simulations . . . 130

5.5 (SSL) CPU time T^{CPU} (hours), simulation time T (ns) and CPU processor of the simulations 137

6.1 Averaged number of iterations $\overline{N}$ and CPU time (in minute) of DR and GS algorithms for simulations on a 400-cell grid 172

7.1 Parameters of the wire-to-wire corona discharge simulations 181

7.2 V-I characteristics of experiment data [12] ("data") and numerical solutions for different values of ψ (m^{-3}) 186

7.3 Experimentally measured ignition voltage V_c^e of positive corona discharge for different electrode gaps d 186

7.4 Parameters of needle-to-ring corona discharge simulations 189

7.5 V-I characteristics of experiment data [163] ("data") and numerical solutions for $d = 20$ mm and different values of ψ (m^{-3}) 191

7.6 V-I characteristics of experiment data [163] ("data") and numerical solutions for $\psi = 10^{11}$ m^{-3} and different values of d 193

7.7 V-u characteristics of experiment data [163] ("data") and numerical solutions for $I_t = 30\%$ and different values of d 194

List of Symbols

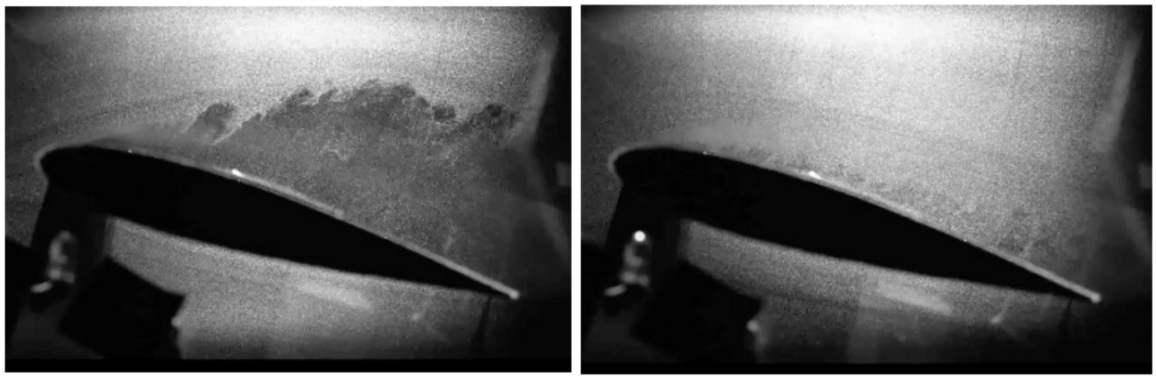
Physics constants		
m_e	Electron mass	$9.109\,383 \times 10^{-31}\,\mathrm{kg}$
k_B	Boltzmann constant	$1.380\,649 \times 10^{-23}\,\mathrm{J\,K^{-1}}$
q	Elementary charge	$1.602\,176 \times 10^{-19}\,\mathrm{C}$
ε_0	Vacuum permittivity	$8.854\,187 \times 10^{-12}\,\mathrm{C\,V^{-1}\,m^{-1}}$
Space and time		
t	Time variable	
T	Simulation time	
$\mathcal{L}$	Number of time levels	
t^l	l^{th} time level, $1 \leq l \leq \mathcal{L}$ ($t^1 = 0$, $t^{\mathcal{L}} = T$)	
$\Delta t^l = t^{l+1} - t^l$	Timestep	
$\boldsymbol{x}$	Space coordinate	
Ω	Computation domain of particle densities	
Ω_ϕ	Computation domain of potential field	
Γ	Boundary of computation domain	
$\Gamma_f, \Gamma_w, \Gamma_s$	Free-flow boundary, wall boundary, symmetric boundary	
$\mathcal{T}$	Grid defined on Ω or Ω_ϕ	
$\mathcal{N} = \mathrm{Card}(\mathcal{T})$	Number of grid cells	
h_K	Circumdiameter of a cell $\Omega_K \in \mathcal{T}$ (triangular grid)	
$h_\mathcal{T} = \max_{K=1,\ldots,\mathcal{N}} h_K$	Grid size	
$\mathcal{E}_K$	Set of edges of the cell Ω_K	
λ_{KL}	Common edge of the cells Ω_K and Ω_L	
$\boldsymbol{x}_K^c$	Center of gravity of Ω_K	
$\boldsymbol{\nu}_{K\lambda}, \boldsymbol{\nu}_{KL}$	Unit outward normal of Ω_K on the edge $\lambda \in \mathcal{E}_K$ or on the edge λ_{KL}	
$\boldsymbol{x}_{K\lambda}, \boldsymbol{x}_{KL}^\lambda$	Midpoint of $\lambda \in \mathcal{E}_K$ or of λ_{KL}	
$\mathcal{V}_K^k$	k^{th}-level neighborhood of Ω_K	
Other symbols		
$\mathfrak{S}$	Set of particle species taken into account in the plasma model	
$n_s(t, \boldsymbol{x})$	Particle density of the specie $s \in \mathfrak{S}$	$\mathrm{m^{-3}}$
$\boldsymbol{f}_s(t, \boldsymbol{x})$	Particle flux density of s	$\mathrm{m^{-2}\,s^{-1}}$
$\boldsymbol{u}_s(t, \boldsymbol{x}), \overline{v}_s$	Drift velocity, mean thermal speed of s	$\mathrm{m\,s^{-1}}$

z_s	Charge number of s	
$\mu_s(t, \boldsymbol{x})$	Mobility coefficient of s	$\mathrm{m^2\,V^{-1}\,s^{-1}}$
$D_s(t, \boldsymbol{x})$	Diffusion coefficient of s	$\mathrm{m^2\,s^{-1}}$
$S_s(t, \boldsymbol{x})$	Production and decay rate (source term) of s	$\mathrm{m^{-3}\,s^{-1}}$
$\overline{\varepsilon}_s(t, \boldsymbol{x})$	Mean energy density of s	$\mathrm{J\,m^{-3}}$
$T_s(t, \boldsymbol{x})$, T_0	Mean particle temperature of s, gas temperature	K
N	Density of neutral particles	$\mathrm{m^{-3}}$
$\alpha(t, \boldsymbol{x})$, $\eta(t, \boldsymbol{x})$	Electron impact ionization, attachment rate coefficients	$\mathrm{m^3\,s^{-1}}$
γ	Secondary emission coefficient	
$\boldsymbol{U}(t, \boldsymbol{x})$	$(n_s(t, \boldsymbol{x}))_{s\in\mathfrak{S}}$	
$\boldsymbol{F}(t, \boldsymbol{x})$	$(\boldsymbol{f}_s(t, \boldsymbol{x}))_{s\in\mathfrak{S}}$	
$\boldsymbol{S}(t, \boldsymbol{x})$	$(S_s(t, \boldsymbol{x}))_{s\in\mathfrak{S}}$	
$\psi(\boldsymbol{x})$	Floor density of electrons ($n_e(t, \boldsymbol{x}) \geq \psi(\boldsymbol{x})$)	$\mathrm{m^{-3}}$
$\rho = q\sum_{s\in\mathfrak{S}} z_s n_s$	Charge density	$\mathrm{C\,m^{-3}}$
$\sigma_{\Gamma_d}(t, \boldsymbol{x})$	Charge surface density on the dielectric surface Γ_d	$\mathrm{C\,m^{-2}}$
$\sigma = q\sum_{s\in\mathfrak{S}} \lvert z_s\rvert \mu_s n_s$	Electrical conductivity	$\mathrm{C\,V^{-1}\,m^{-1}\,s^{-1}}$
$V_G(t)$	Electromotive force of the electric generator	V
$\phi(t, \boldsymbol{x})$	Potential field	V
$\varepsilon_r(\boldsymbol{x})$	Relative permittivity of the medium	$\mathrm{C\,V^{-1}\,m^{-1}}$
$\boldsymbol{E} = -\nabla\phi$, $E = \lvert\boldsymbol{E}\rvert$	Electric field, field strength	$\mathrm{V\,m^{-1}}$
$\widehat{E} = E/N$	Reduced field strength	$\mathrm{V\,m^2}$
$\boldsymbol{j} = \sigma\boldsymbol{E}$	Conduction current density	$\mathrm{C\,m^{-2}\,s^{-1}}$
$I(t)$	Electric current in the circuit	A
R	Resistance of the circuit	Ω
$\boldsymbol{f}_{\mathrm{EHD}} = \rho\boldsymbol{E}$	Electrohydrodynamic force density	$\mathrm{N\,m^{-3}}$

Introduction

Les récents événements météorologiques extrêmes qui se sont produits dans le monde entier suscitent de grandes inquiétudes concernant le changement climatique et renforcent la nécessité de développer des technologies écologiques permettant d'enrayer la tendance croissante à la destruction de l'écosystème. Selon une étude publiée en 2021 dans la revue Atmospheric Environment [89], l'industrie aéronautique est responsable d'environs 3.5% des facteurs de changement climatique liés aux activités humaines entre 2000 et 2018, dont les émissions de dioxyde de carbone (CO_2) et d'oxydes d'azote (NO_x). L'augmentation drastique du trafic de fret et de passagers nécessite donc l'augmentation des performances aérodynamiques et énergétiques des aéronefs.

Un objectif d'optimisation important dans la conception d'un avion est la réduction de la traînée de surface (de frottement). La traînée apparaît comme une force de réaction lorsqu'un objet se déplace dans l'air. Comme l'objet ralentit l'écoulement de l'air, la troisième loi de Newton stipule que l'écoulement de l'air ralentit également l'objet. Lorsque la traînée augmente, le moteur doit augmenter la poussée, c'est-à-dire la force de poussée agissant dans la direction opposée à la traînée, pour maintenir la vitesse de l'avion. La situation est encore plus grave si une couche limite d'écoulement est séparée de la surface de l'avion, par exemple comme le montre fig. 1a, où la vitesse de l'écoulement diminue rapidement jusqu'à devenir nulle ou même s'inverser. La force de traînée étant proportionnelle au taux de diminution de la vitesse de l'écoulement, la séparation de l'écoulement nécessite beaucoup plus de poussée, et donc de carburant, pour l'équilibrer.



(a) Séparation de l'écoulement (b) Rattachement de l'écoulement

Figure 1: Écoulement naturel (à gauche) et effet des actionneurs électriques sur l'écoulement (à droite) autour d'un profil aérodynamique, extrait de [55]

Les études menées au cours des deux dernières décennies ont montré que les dispositifs électriques appelés **actionneurs de plasma** ou **actionneurs électriques** qui génèrent des décharges électriques

continues, alternatives ou pulsées sont capables de modifier les propriétés d'un écoulement et permettent ainsi de contrôler l'écoulement autour d'une surface, comme le montre par exemple fig. 1b où l'écoulement est rattaché à la surface d'un profil aérodynamique. L'énergie électrique fournie par un tel dispositif est utilisée pour transformer localement l'air en un gaz ionisé vaguement appelé **plasma**, qui à son tour cède son moment par des collisions élastiques et crée ainsi un déplacement des molécules d'air. Cet effet est connu sous le nom de **vent ionique**. L'étude des effets aérodynamiques des décharges électriques a été lancée par Roth et al. [128] en 2000 et bien d'autres équipes. Une revue des travaux sur le sujet se trouve dans un excellent article de E. Moreau [109] en 2007.

Depuis lors, l'intérêt pour les actionneurs à plasma s'est accru dans l'industrie aéronautique, non seulement pour le contrôle de l'écoulement, mais aussi pour d'autres applications récentes telles que le vol des drônes [161] ou l'antigivrage [167], en raison de leur simplicité de conception, de leur petite taille, de leur faible coût et de leur facilité d'installation à la surface de l'avion. Toutefois, les effets de ces dispositifs sur le contrôle des flux n'ont été testés avec succès que pour une vitesse de l'écoulement allant jusqu'à $50\ \mathrm{m\ s^{-1}}$ dans des conditions de laboratoire.

Il est évident que la recherche fondamentale sur la formation, le développement, la morphologie et l'interaction du gaz ionisé avec l'air environnant, ainsi que sur ses conséquences aérodynamiques, doit encore être renforcée afin d'améliorer et d'optimiser les performances des actionneurs à plasma. La modélisation numérique, en particulier, est l'une des approches les plus puissantes dont disposent les scientifiques et les ingénieurs et qui pourrait ouvrir la voie à des applications plus larges, voire à la commercialisation de cette technologie au bénéfice de tous. L'intérêt pour les modèles numériques de décharge de gaz n'est ni nouveau ni superflu. Les industriels de l'aéronautique utilisent la simulation numérique pour réduire le cycle de conception des aéronefs. En effet, elle reste peu coûteuse par rapport à la réalisation du prototype.

La modélisation numérique de la décharge électrique dans l'air s'avère être une tâche difficile pour trois raisons principales. La première raison est liée à la **qualité de précision** des solutions fournies par les solveurs numériques. Ce problème a été signalé pour la première fois pour la simulation de décharge par Bœuf et al. [19] où les auteurs ont utilisé des méthodes numériques du premier ordre (faible précision) et ont observé que les solutions calculées sur des maillages grossiers s'écartaient de manière significative de celles évaluées avec une taille de maillage plus petite. Des maillages plus fins impliquent davantage de ressources de calcul et d'efforts à consacrer à l'obtention de solutions de haute qualité, et donc prédictives, ce qui est l'objectif de la modélisation numérique dès le départ.

La deuxième raison découle déjà de la première, à savoir la question du **temps de calcul**. En effet, une décharge électrique est le siège de nombreux phénomènes physiques complexes qui se produisent à des échelles spatiales et temporelles extrêmement disparates et qui doivent toutes être résolues. Le temps caractéristique le plus minime, généralement celui du transport des électrons, peut être 10^9 fois plus petit que l'échelle de temps du vent ionique (voir table 1 par exemple) qui est aussi le temps de simulation typique pour une décharge de plasma afin d'étudier son interaction avec le flux d'air extérieur. La zone de transition entre le plasma et l'air, connue sous le nom de **gaines**, dont la taille est micrométrique, doit également être attentivement prise en compte car elle est le siège d'une partie de la force **électrohydrodynamique** (EHD) qui nous intéresse de calculer dans

les problèmes d'actuation. La décharge de plasma est donc un **problème multi-échelle** car le pas de temps numérique doit se soumettre à la plus petite échelle d'espace et de temps tout en résolvant d'autres échelles plus grandes. Cette exigence représente un énorme défi car le calcul prend un temps excessif, en particulier dans les simulations bidimensionnelles ou tridimensionnelles. À titre de référence, une simulation typique d'une décharge électrique alimentée par une tension sinusoïdale, utilisant le solveur de plasma construit en interne par l'ONERA - le code **COPAIER**[1] [46] - prend entre 1 et 2 mois pour compléter 4 périodes d'une tension de fréquence 100 kHz, ou un temps de simulation de 0.4 ms [82]. Ce calcul a même été parallélisé et effectué sur 100 cœurs de CPU, et la fréquence de simulation était supérieure à une fréquence typique utilisée dans les expériences (1-25 kHz).

Phénomène	Temps caractéristique
Transport des électrons	1 ps à 1 ns
Relaxation du champ électrique	10 ps à 1 μs
Réactions cinétiques	0.1 à 1 μs
Transport de particules lourdes	0.1 à 1 ns
Dynamique des plasmas	10 à 100 ns
Vent ionique	1 ms

Table 1: Échelles de temps caractéristiques des phénomènes dans une décharge de plasma, tableau reproduit de [46]

La troisième raison se trouve dans le **couplage** entre la décharge de plasma et le flux d'extérieur. Dans les écoulements subsoniques, on suppose généralement que ce couplage est unidirectionnel, en ce sens que la décharge peut modifier l'écoulement extérieur, mais que ce dernier n'est pas assez puissant pour influencer les caractéristiques de la décharge. Cette hypothèse est prise en compte dans le chapitre 7 de cette thèse, où la force EHD résultant de la simulation de la décharge est injectée dans le terme source de l'équation de quantité de mouvement de l'écoulement extérieur. D'autre part, dans les écoulements supersoniques, le couplage est fortement bidirectionnel car la vitesse des espèces chargées due au transport de l'écoulement n'est plus négligeable par rapport à la composante de vitesse due à la force électrique. Cependant, le régime supersonique ne sera pas considéré dans cette thèse.

L'objectif de cette thèse est d'aborder les deux premiers problèmes dans la modélisation numérique de la décharge électrique dans l'air, dans le cadre de la description hydrodynamique (macroscopique) du plasma. Le modèle mathématique utilisé consiste d'un système d'équations de dérive-diffusion-réaction (équations de continuité) décrivant le mouvement, la création et la destruction des différentes espèces de particules dans le plasma (électrons, ions, etc.) sous l'influence du champ électrique. Ce système est couplé à l'équation de Poisson du champ électrique, elle-même modifiée par la présence de densités de charges locales. Pour la première question, nous avons développé de nouveaux schémas de flux de haute précision basés sur la méthode classique de Scharfetter-Gummel du premier ordre (faible précision) pour la résolution des équations de dérive-diffusion-réaction. Le deuxième problème

[1] COde PlasmA Instationnaire pour l'aERodynamique

se décompose en fait en deux sous-problèmes qui correspondent à deux régimes de décharge qui pourraient être classés sur l'observation du courant électrique généré par la décharge.

1. Le premier est le régime **couronne** qui se caractérise par un faible courant de l'ordre de 10-100 $\mathrm{mA\ m^{-1}}$ et une faible densité d'électrons de l'ordre de 10^{15}-10^{16} $\mathrm{m^{-3}}$. Les variations de la charge d'espace sont relativement faibles, et les échelles de temps caractéristiques de la relaxation du champ électrique et du transport des ions peuvent être considérées comme longues par rapport à celles des réactions cinétiques et du transport des électrons.

2. Le second est le régime des **microdécharges** qui se caractérise par un courant élevé de l'ordre de 1-10 $\mathrm{A\ m^{-1}}$ et une forte densité d'électrons de l'ordre de 10^{18}-10^{21} $\mathrm{m^{-3}}$. Le champ de charge spatiale modifie de manière significative le champ électrique externe sur des temps caractéristiques très courts. Par conséquent, les temps caractéristiques de la variation du champ électrique et du transport d'électrons sont à peu près du même ordre.

Pour le régime de décharge couronne, nous avons développé une méthode d'intégration temporelle implicite pour résoudre le système d'équations de dérive-diffusion-réaction alors que le champ électrique est considéré comme constant sur la période de temps pendant laquelle les densités d'espèces évoluent. De cette manière, la résolution des équations de continuité et de l'équation de Poisson sont deux problèmes distincts et cette stratégie numérique est donc beaucoup plus simple que la résolution de l'ensemble du système de décharge en une seule fois. La méthode implicite permet de réduire considérablement le temps de calcul par rapport aux méthodes explicites, car elle permet de relâcher la contrainte sur le pas de temps numérique qui est strictement requise dans l'intégration temporelle explicite.

Pour le régime des microdécharges, l'approche précédente n'est plus valable car la variation du champ électrique est trop rapide pour être considérée comme constante au cours de l'évolution des densités. Le système de décharge doit être linéarisé d'une certaine manière afin que les équations de continuité et l'équation de Poisson puissent être mises à jour séparément. Pour ce type de décharge, nous utilisons l'approche semi-implicite préexistante [155, 66] pour la linéarisation de l'équation de Poisson qui a été implémentée dans COPAIER avant le début de la thèse.

Les solveurs numériques développés dans cette thèse ont été construits à l'image du code COPAIER. Ce dernier est un solveur plasma multi-espèces et multi-réactions qui permet d'incorporer autant d'espèces et de réactions cinétiques que nécessaire, fonctionne sur des maillages bidimensionnels structurés et non-structurés et a contribué à la recherche sur la décharge électrique dans l'air par de nombreuses simulations très complexes [81, 82, 99], mais les méthodes numériques dans ce solveur étaient néanmoins explicites en temps.

La structure de ce document est organisée en deux parties principales. La partie I composée des trois premiers chapitres présente quelques notions de physique des décharges de gaz ainsi que les travaux de recherche notables sur les actionneurs de plasma au cours des deux dernières décennies (chapitre 1), la description des modèles mathématiques utilisés dans les simulations de cette thèse (chapitre 2), et un aperçu rapide des méthodes numériques qui sont mises en œuvre dans COPAIER (chapitre 3).

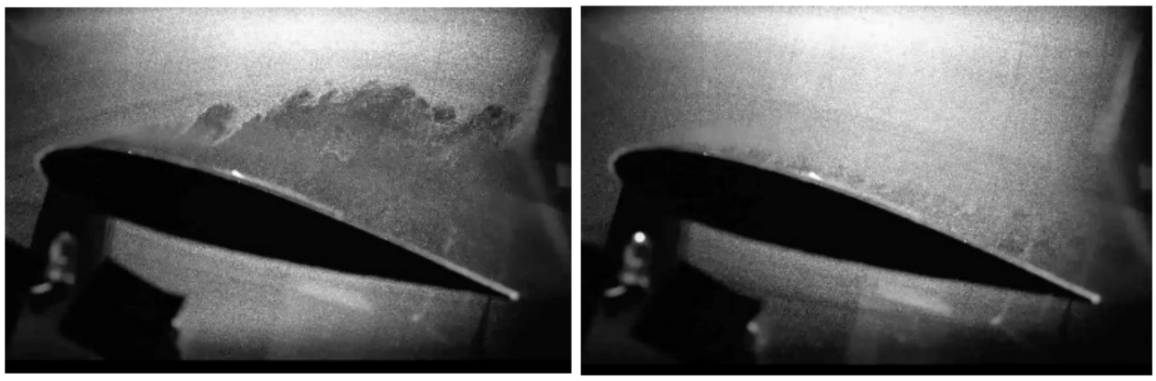
La partie II couvrant les chapitres 4, 5, 6 et 7 présente la contribution de nos travaux à la modélisation numérique de la décharge atmosphérique. Plus précisément, le chapitre 4 introduit la dérivation de schémas de flux de Scharfetter-Gummer d'ordre élevé pour la discrétisation des équations de dérive-diffusion unidimensionnelles, mais discute également de leur extension aux problèmes bidimensionnels et fournit une vérification numérique de l'ordre de convergence pour les cas tests unidimensionnels et bidimensionnels. Dans le chapitre 5, nous appliquons les nouveaux schémas à la simulation d'une décharge couronne et de la propagation d'un streamer positif, qui est un type de microdécharges. Un solveur de plasma en Fortran pour la simulation des décharges sur des maillages bidimensionnels cartésiens a été développé à partir de scratch pour ce chapitre.

Le chapitre 6 aborde la question de la réduction du temps de calcul pour le régime de décharge de la couronne qui a été brièvement mentionné précédemment. Une particularité du modèle de décharge présenté dans ce chapitre est l'inclusion d'une contrainte qui exige que la densité d'électrons soit toujours supérieure à une valeur prescrite. L'approche proposée nécessite un modèle mathématique propre ainsi que des algorithmes spécifiques pour résoudre le système d'équations de continuité. Enfin, le chapitre 7 présente les solutions numériques pour différents cas de test de décharge, obtenues avec la méthode implicite qui a été développée dans le chapitre précédent et mise en œuvre dans le code COPAIER. Cette méthode implicite a été récemment mise en œuvre dans COPAIER, ce qui permet d'augmenter la capacité de simulation du solveur de plasma par un facteur significatif.

Introduction

Recent extreme weather events that have been occurring more frequently around the globe have raised further concerns about climate change and fuels the need of developing environment-friendly technologies that could halt the increasing trend of ecosystem destruction. According to a study in the Atmospheric Environment journal in 2021 [89], the aviation industry is responsible for roughly 3.5% of all drivers of climate change from human activities from 2000 to 2018, including emissions of harmful gases such as carbon dioxide (CO_2) and nitrogen oxides (NO_x). The drastic increase in freight and passenger traffic therefore require significant aircraft performance.

An important optimizing goal in aircraft designing is the reduction of drag on the its surface. Drag appears as a reaction force when any object travels in air. As the object forces the flow to slow down, by the third law of Newton the flow also forces the object to slow down. As drag increases, the engine must add more thrust that is the pushing force acting on the opposite direction of drag to maintain the aircraft speed. The situation is more dire if a flow boundary layer is separated from the aircraft surface, for example as shown in fig. 2a, where the flow velocity reduces rapidly to zero or even reversed. Since the drag force is proportional to the decreasing rate of the flow speed, flow separation requires a lot more thrust, and therefore fuels, to balance out.



(a) Flow separation (b) Flow reattachment

Figure 2: Natural flow (left) and effect of electric actuators on the flow (right) around an airfoil, taken from [55]

Studies carried out over the last two decades have shown that electrical devices called **plasma actuators** or **electric actuators** that generates continuous, alternative or pulsed electric discharges are capable of modifying the properties of a flow and thus enable to control the flow around a surface, for example as shown in fig. 2b where the flow is reattached to the surface of an airfoil. The electrical energy supplied by such a device is used to locally transform air into an ionized gas loosely

called a **plasma**, which in turn gives up its momentum through elastic collisions and thus create a displacement of air molecules. This effect is known as **ionic wind**. The study on aerodynamic effects of electric discharge was initiated by Roth et al. [128] back in 2000 and many other teams. A scientific review on this subject can be found in the excellent article of E. Moreau [109] in 2007.

Since then, the interest in plasma actuators have been growing in the aeronautical industry, not just for flow control but for other recent applications such as UAV flight sustaining [161] or anti-icing [167], because of its simplicity in design, smallness in size, low cost, facility in installing on aircraft surface and eco-friendliness. However, the effects of these devices in flow control have been only successfully tested for air speed up to $50\ \mathrm{m\ s^{-1}}$ in laboratory conditions. Therefore, it could be foreseen that the use of plasma actuators for commercial purposes will not happen in short time.

It is quite clear that fundamental research on the formation, development, morphology and interaction of the ionized gas with the surrounding air as well as its aerodynamic consequences still needs to be supported in order to enhance and optimize the performance of plasma actuators. Numerical modeling, in particular, is one of the most powerful approaches in the hands of scientists and engineers that could pave the way to wider applications or even commercialization of this technology to the benefit of all. The dedication in numerical models for gas discharge is not something new or superfluous. Even well-rooted industrial fields such as aircraft or car manufacturing still heavily involve numerical simulations as highly predictive and complementary tools to experimental works. Most importantly, they are very uncostly. Indeed, destroying an F1 car in simulation would be much more less heartbreaking than crashing my 30 € bike in a tree (unless you are on the Mercedes team).

Numerical modeling of electric discharge in air proves to be not an easy task, though, for three reasons. The first reason ties with the **precision quality** of the solutions provided by the numerical solvers. This issue was first reported for gas discharge simulation by Bœuf et al. [19] where the authors used first-order (low precision) numerical methods and observed that the computed solutions on coarse grids deviated significantly from those evaluated with smaller grid size. Finer grids mean more computation resources and efforts must be dedicated in order to obtain high quality, and therefore predictive, solutions which are the purpose of numerical modeling from the beginning.

The second reason already looms out from the first one that is the issue of **computation time**. Indeed, an electric discharge is the siege of many complex physical phenomena that happen on extremely disparate space and time scales that all need to be resolved. The tiniest characteristic time, usually that of the transport of electrons, could be as low as 10^9 times smaller than the ionic wind timescale (see table 2 for instance) that is also the typical simulation time for a plasma discharge in order to study its interaction with the outer air flow. The transition zone between plasma and air, known as **sheaths**, which are micrometric in size, also needs to be carefully taken into account as they are the siege of part of the **electrohydrodynamic** (EHD) force that we are interested in computing in actuation problems. The plasma discharge is therefore a **multiscale problem** as the numerical time stepping must comply with the smallest space and time scale while solving other greater ones. This requirement presents a huge challenge since the computation takes an excessive amount of time to finish, especially in two-dimensional or three-dimensional simulations.

For reference, a typical simulation of an electric discharge supplied with a sinusoidal voltage, using the ONERA-in-house-built plasma solver - the **COPAIER**[2] code [46] - takes between 1 and 2 months to complete 4 periods of a 100 kHz-frequency voltage, or a simulation time of 0.4 ms [82]. This computation was even parallelized and performed on 100 CPU cores, and the simulation frequency was higher than a typical frequency used in experiments (1-25 kHz).

Phenomenon	Characteristic time
Transport of electrons	1 ps to 1 ns
Electric field relaxation	10 ps to 1 μs
Kinetic reactions	0.1 to 1 μs
Transport of heavy particles	0.1 to 1 ns
Plasma dynamics	10 to 100 ns
Ionic wind	1 ms

Table 2: Characteristic timescales of phenomena in a plasma discharge, reproduced from [46]

The third reason lies in the **coupling** between the plasma discharge and the outer airflow. In subsonic flows, it is widely assumed that this coupling is one-way, in the sense that the discharge can modify the outer flow but the outer is not strong enough to influence the characteristics of the discharge. This assumption is considered in chapter 7 of this thesis, where the EHD force ensued from the discharge simulation is injected into the source term of the momentum equation of the outer flow. On the other hand, in supersonic flows the coupling is strongly two-way as the velocity of charged species due to the flow transport is no longer negligible comparing to the velocity component due to the electric force. However, the supersonic regime will not be considered in this thesis.

The objective of this thesis is to tackle the first two issues in numerical modeling of electric discharge in air, in the framework of hydrodynamic (macroscopic) description of plasma. The underlying mathematical model consists of a system of conservation laws, under the form of drift-diffusion-reaction (continuity) equations, describing the movement, creation and destruction of the various particle species in the plasma (electrons, ions, etc.) under the influence of the electric field. This system is coupled to the Poisson equation of the electric field, which is itself modified by the presence of local charge densities. For the first issue, we have developed novel high-accuracy flux schemes based on the classical first-order (low-accuracy) Scharfetter-Gummel method for the resolution of drift-diffusion-reaction equations. The second issue actually breaks into two subproblems that correspond to two discharge regimes that could be classified based on observation of the electric current generated by the discharge.

1. The first one is the **corona** regime that is characterized by a low current on the order of 10-100 $\mathrm{mA\ m^{-1}}$ and a low electron density on the order of 10^{15}-10^{16} $\mathrm{m^{-3}}$. Space charge variation are relatively weak, and the characteristic electric field relaxation and ion transport timescales can be considered long compared to those of kinetic reactions and transport of electrons.

[2] COde PlasmA Instationnaire pour l'aERodynamique

2. The second one is the **microdischarge** regime that is characterized by a high current on the order of 1-10 $\mathrm{A\ m^{-1}}$ and a high electron density on the order of 10^{18}-10^{21} $\mathrm{m^{-3}}$. The space charge density modifies significantly the external electric field over a very short timescale. Therefore, the characteristic timescales of electric field relaxation and electron transport are on the same order.

For the corona discharge regime, we have developed an implicit time integration method to solve the system of drift-diffusion-reaction equations while the electric field is considered constant on the time window that the specie densities evolve. In this way, the resolution of the continuity equations and the Poisson equation is two separate issues and therefore this numerical strategy is much simpler than solving the whole discharge system all at once. The implicit method allows to significantly reduce the computation time comparing to explicit methods, as it allows to relax the constraint on the numerical timesteps that is strictly required in explicit methods.

For the microdischarge regime, the previous approach is hardly valid since the electric field variation is too rapid that it could not be no longer considered constant as the densities evolve. The discharge system needs to be linearized in a certain way so that the continuity equations and the Poisson equation could be updated separately. For this type of discharge, we use the preexisting semi-implicit approach [155, 66] for linearization of the Poisson equation that was implemented in COPAIER before this thesis.

The numerical solvers developed in this thesis were built in the image of the COPAIER code. The latter is a multi-specie, multi-reaction plasma solver that allows to incorporate as many species as well as kinetic reactions as needed, functions on both two-dimensional structured and unstructured grids and has been contributing to the research on electric discharge in air through many highly complex simulations [81, 82, 99], but the numerical methods in this solver were however explicit in time.

The structure of this document is arranged into two main parts. Part I consisting of the first three chapters will introduce some notions in gas discharge physics as well as notable researches on plasma actuators over the last two decades (chapter 1), the description of mathematical models used in the simulations in this thesis (chapter 2), and a quick glance of the numerical methods that are implemented in COPAIER (chapter 3).

Part II covering chapters 4 to 7 presents the contribution of our works in numerical modeling of atmospheric pressure discharge. More precisely, chapter 4 introduces the derivation of high-order Scharfetter-Gummer flux schemes for the discretization of one-dimensional drift-diffusion equations, but also discusses their extension to two-dimensional problems and provides numerical verification of convergence order for both one-dimensional and two-dimensional test cases. In chapter 5, we apply the novel schemes to simulation of a corona discharge and of the propagation of a positive streamer, which is a type of microdischarges. A Fortran plasma solver for simulation of two-dimensional streamer discharge on cartesian grids has been developed from scratch for this chapter.

Chapter 6 addresses the issue of reducing the computation time for the corona discharge regime that was briefly mentioned before. A particularity of the discharge model in this chapter is the

inclusion of a constraint that requires the electron density to be always larger than a prescribed value. The proposed approach necessitates an appropriate mathematical model as well as some specific algorithms to solve the system of conservation laws. Finally, chapter 7 presents the numerical solutions for different discharge test cases, obtained with the implicit method that was developed in the previous chapter. This implicit method has been recently implemented in COPAIER that allows to increase the simulation capacity of the plasma solver by a significant factor.

Part I

Physical, Mathematical and Numerical Background on Atmospheric Pressure Discharge

Chapter 1

Notes on Gas Discharge Physics and Plasma Actuators

1.1 What is a plasma? . . . 16
1.1.1 Elements of kinetic theory of gases . . . 16
1.1.2 Electric screening and plasma oscillations . . . 17
1.1.3 Ideal plasmas . . . 19
1.1.4 Classifications of plasmas . . . 19
1.1.5 Are ionized gases in flow control always ideal plasmas? . . . 20
1.2 Low-temperature discharges in air . . . 21
1.2.1 Production and decay of particles in low-temperature discharges . . . 21
1.2.2 Discharges in moderate-pressure gases . . . 25
1.2.3 Discharges in atmospheric pressure . . . 29
1.3 Corona and dielectric barrier discharges . . . 32
1.3.1 Corona discharges . . . 32
1.3.2 Surface dielectric barrier discharges . . . 35

1.1 What is a plasma?

Plasma is one of the fundamental states of matter which accounts for roughly 99% the mass of ordinary matter in the observable universe. Fundamental as it is, plasma is a fairly new concept which was only coined in 1929 by Langmuir [88] to describe a form of **ionized gas** that contains "balanced charges of ions and electrons", and was (maybe still is) unpopular comparing to other states of matter such as solid, liquid and gas. On Earth, natural plasmas are only seldom encountered in few examples of exotic phenomena: lightning bolts, auroras, sprites or St. Elmo's fires. It would seem that the lack of occurrences and of technological advances had contributed to the late understanding and applications of plasma physics.

Today, it is established that plasmas constitute a large number of celestial objects including the Sun, stars, white dwarfs, and nebulae. Ionized hydrogen that fills the interstellar media, solar wind that fills the interplanetary space, Van Allen radiation belts and the ionosphere surrounding the Earth are also examples of plasmas in nature. Back to our planet, they have found use in plasma etching, sterilization of medical devices, fluorescence tubes, neon signs, space propulsion, and historically PDP and PALC television screens. Other artificial plasmas being tested in laboratories have enormous potential to become more relevant to human life in a wide range of areas such as water treatment, air purification, water harvesting, thermal management in electronic devices, plasma-assisted combustion, food drying and controlled thermonuclear fusion. Particularly in the field of aeronautics, potential applications of plasma discharge include flow control [109, 20], cold plasma propulsion [161] and anti-icing [167].

So what exactly is plasma and what differentiates it from a normal ionized gas which is, generally speaking, a medium that contains a certain amount of "atoms dissociated into positive ions and negative electrons" [34]. Several authors such as Chen [34] and Moisan & Pelletier [108] extended the definition of Langmuir to characterize a plasma as

"a **macroscopically neutral** gaseous medium with **collective behavior**."

In order to clarify the meaning of this concept, it is necessary to first recall certain notions of kinetic theory as well as some useful parameters in plasma physics.

1.1.1 Elements of kinetic theory of gases

For a particle specie s (e.g. positive ions or electrons[1]), the distribution function $f_s(t, \boldsymbol{x}, \boldsymbol{v})$ at time t, position $\boldsymbol{x} \in \mathbb{R}^3$ and velocity $\boldsymbol{v} \in \mathbb{R}^3$, is defined as follows. Within an infinitesimal element $\mathrm{d}\boldsymbol{v}$ of the velocity space, centered on the velocity $\boldsymbol{v}$, the probability of finding particles of the specie s having a velocity $\boldsymbol{v}'$ contained in that infinitesimal element $\mathrm{d}\boldsymbol{v}$, at time t and position $\boldsymbol{x}$, equals to $f_s(t, \boldsymbol{x}, \boldsymbol{v})\mathrm{d}\boldsymbol{v}$. Since f_s is a probability distribution on the velocity phase, it should be normalized, i.e. $\displaystyle\int_{\mathbb{R}^3} f_s(t, \boldsymbol{x}, \boldsymbol{v})\mathrm{d}\boldsymbol{v} = 1$. If we denote as $n_s(t, \boldsymbol{x})$ the particle density of the specie s, then $n_s f_s \mathrm{d}\boldsymbol{v}$ would be the number of particles per unit volume having a velocity $\boldsymbol{v}'$ contained in $\mathrm{d}\boldsymbol{v}$. Without ambiguity, we shall drop the dependence of f_s on t and $\boldsymbol{x}$.

[1] denoted resp. by the symbols p and e

The mean particle speed and the mean kinetic energy are defined as resp.

$$\overline{v}_s = \int_{\mathbb{R}^3} v f_s(\boldsymbol{v}) \mathrm{d}\boldsymbol{v} \quad \text{and} \quad \overline{\varepsilon}_s = \int_{\mathbb{R}^3} \frac{1}{2} m_s v^2 f_s(\boldsymbol{v}) \mathrm{d}\boldsymbol{v},$$

with $v \equiv |\boldsymbol{v}|$ and m_s the mass of a particle of specie s.

When the gas reaches thermodynamic equilibrium, the velocity distribution of the species follow what is known as the Maxwell-Boltzmann distribution, which is given by

$$f_s(\boldsymbol{v}) = \left(\frac{m_s}{2\pi k_B T_0}\right)^{\frac{3}{2}} \exp\left(-\frac{m_s v^2}{2 k_B T_0}\right),$$

where k_B is the Boltzmann constant and T_0 is the absolute temperature of the gas (same for all species). In this case,

$$\overline{v}_s = \left(\frac{8 k_B T_0}{\pi m_s}\right)^{\frac{1}{2}} \quad \text{and} \quad \overline{\varepsilon}_s = \frac{3}{2} k_B T_0.$$

The mean particle velocity $\overline{\boldsymbol{v}}_s = \int_{\mathbb{R}^3} \boldsymbol{v} f_s(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}$ is zero, though, since f_s is isotropic with respect to $\boldsymbol{v}$. This shows the randomness of particles' movement in a thermal-equilibrium gas, and therefore the close relation between the macroscopic concept of gas temperature and the random movement of particles as well as the mean kinetic energy in the kinetic theory of gases. For this reason, T_0 is often expressed in energy units such as J or eV.

Not all ionized gases are in thermal equilibrium, though, such as the type of gas often used in flow control. In these **non-equilibrium** gases[2], the electron temperature T_e is between 1-10 eV, i.e. 11600-116000 K, whereas the ion temperature T_p is roughly the same as ambient temperature, i.e. $T_p \approx 300$ K. In such circumstances the electron velocity distribution is not exactly Maxwellian[3] because of the temperature disparity, but assumes another function such as the Druyvesteyn distribution [108]. However, the discussion of electron distribution in non-equilibrium gases is out of scope of this thesis. More than often the electron distribution is assumed to be Maxwellian and the derivation of the Debye length introduced later is based on this assumption [34, 108].

1.1.2 Electric screening and plasma oscillations

Of particular importance is the **Debye length** λ_D which is a length measure of the electric screening effect of a charge carrier in the gas. Imagine a positive ion of charge q floating in the medium, then it would attract a cloud of electrons due to the Coulomb force (see fig. 1.1). In a plasma, as the number of electrons is high enough, the electric potential ϕ generated by the ion will be exponentially attenuated as

$$\phi(r) = \frac{q}{4\pi\varepsilon_r\varepsilon_0 r} \exp\left(-\frac{r}{\lambda_D}\right) \quad \text{with} \quad \lambda_D = \left(\frac{\varepsilon_r\varepsilon_0 k_B T_e}{n_e q^2}\right)^{\frac{1}{2}}, \tag{1.1}$$

[2] otherwise known as **low-temperature** or **two-temperature** gases

[3] i.e. f_e is the Maxwell-Boltzmann distribution

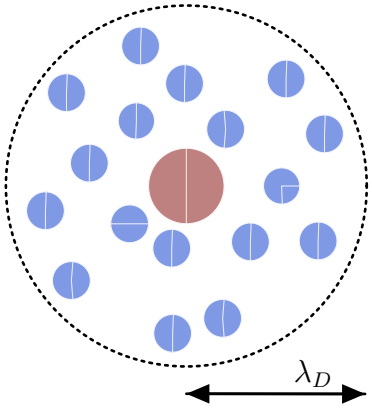


Figure 1.1: Debye shielding of a positive charge (maroon) by negative ones (blue)

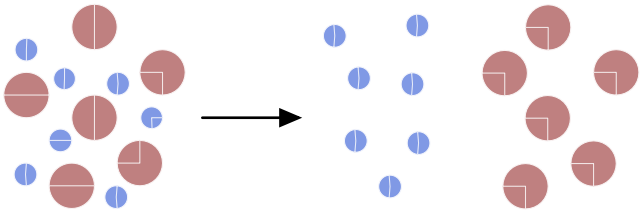


Figure 1.2: Particle displacements from "equilibrium" (left) to charge separation (right) that lead to plasma oscillations

with r the distance from the charge carrier, ε_r the relative permittivity of the medium and ε_0 the vacuum permittivity. The concentration of electrons around the positive charge has a better screening effect in which it supplants the Coulomb potential $\phi^C(r) = \dfrac{q}{4\pi\varepsilon_r\varepsilon_0 r}$ generated by the charge carrier (the ion); in other words, binary particle-particle electric interactions are no longer important on a spatial scale on the order of λ_D.

The number of electrons contained in the Debye sphere, which is the volume of radius λ_D centered at the charge point, is proportional to $n_e\lambda_D^3$. The latter known as the **plasma parameter** can be alternatively written as

$$n_e\lambda_D^3 = \frac{1}{6\pi}\frac{\overline{\varepsilon}_e}{q\phi^C(\lambda_D)}.$$

It is interpreted as the ratio between the electron mean kinetic energy and the Coulomb potential energy generated by the charge carrier. When the electron population inside the Debye sphere is large enough, i.e. $n_e\lambda_D^3 \gg 1$, the electrons' movement in the outer region of the Debye sphere is mostly due to their thermal agitation and is shielded from the electric effect of the charge carrier.

The time scale associated to the Debye length is the **plasma frequency**[4] ω_e given by

$$\omega_e = \left(\frac{n_e q^2}{\varepsilon_r\varepsilon_0 m_e}\right)^{\frac{1}{2}} = \frac{\sqrt{2\pi}}{4}\frac{\overline{v}_e}{\lambda_D}.$$

This parameter characterizes the plasma oscillations which occur when small disturbances in the electric field cause the electrons to move away from the ions (see fig. 1.2). The charge separation

[4] usually is the electron oscillation frequency since electrons are much lighter, thus move faster, than ions

generates a polarizing field on the order of $\frac{n_e q^2}{\varepsilon_r \varepsilon_0}$ that "pulls" the electrons back to where they were, resulting in a collective motion of electrons in order to restore the shielding effect of the Debye spheres in response to small charge displacements.

1.1.3 Ideal plasmas

Now the picture painted in fig. 1.1 is a bit misleading since it gives an impression that the electrons are much more numerous than the ions. In fact, shielding could occur for whatever charge carrier so that an electron potential could also be neutralized by an ion cloud (in this situation the Debye length depends on parameters of ions). Each particle hosts its own Debye sphere while takes part in the same time in others' sphere, so the microscopic picture of an ionized gas is really an agglomeration of overlapping Debye spheres. The mutual electric screening of charge carriers results in a **quasi-neutral** medium on a length scale much larger than λ_D and the electron and ion densities are roughly the same and equal a characteristic density n known as **plasma density**, i.e. $n_e \approx n_p \approx n$. Furthermore, the picture is not static: there is always "runaway" electrons from a sphere to another one if their thermal motion is fast enough to escape the potential well of the center charge, causing oscillations in the electric field.

This picture of ionized gases shows collective interactions between Debye spheres, rather than binary interactions between particles via Coulomb force since the charge carriers are effectively screened. This behavior was deduced from a condition that we have seen in the previous section: $n_e \lambda_D^3 \gg 1$; it is the first condition that differs a plasma from a generally-speaking ionized gas. The full list of criteria defining an (ideal) plasma according to [34, 108] consists of the following points.

1. $n\lambda_D^3 \gg 1$, which is a necessary condition for collective behavior.
2. $\lambda_D \ll L$, where L is a length characterizing the scale of the charge density gradient. This condition is derived from the approximation $\delta n/n \sim (\lambda_D/L)^2$ (see [126]) where δn is a small space charge density caused by charge separation. If $\delta n \ll n$, then the medium remains macroscopically quasi-neutral.
3. $\omega_e \tau_{\text{col}} \gg 1$, where τ_{col} is the smallest time scale of particle collisions. This condition allows the particles to rearrange themselves into Debye spheres and retain collective behavior between successive collisions.

1.1.4 Classifications of plasmas

A wide range of plasmas exist in nature or are artificially created and can be classified based on electron temperature T_e and plasma density n. An incomplete overview of different types of plasma is sketched in fig. 1.3. For low-temperature (non-equilibrium) plasmas in aeronautical applications, T_e ranges typically between 1-10 eV while n is typically found between 10^{15}-10^{21} m^{-3}.

Within non-equilibrium plasmas, the electric field frequency serves to distinguish between *(i)* low-frequency (0-10^4 Hz) and pulsed discharges (except very short pulses); *(ii)* radio-frequency

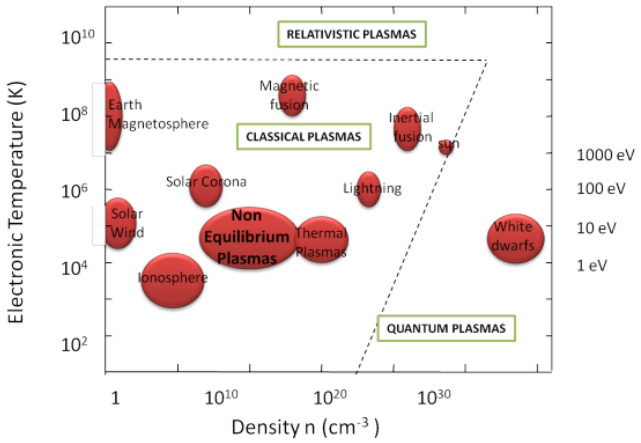


Figure 1.3: Classification of plasmas based on electron temperature and plasma density, taken from [80]

discharges (10^5-10^8 Hz); *(iii)* microwave discharges (10^9-10^{11} Hz); and *(iv)* optical discharges (10^{12}-10^{15} Hz) [126]. The first category consists of non-magnetized gases while in the latter ones, the effect of the magnetic field is more pronouncing and therefore is not neglected. Further classification is based on gas pressure which affects ionization mechanisms and will be addressed with more details in the next sections. In the scope of this thesis, we focus on **low frequency, low-temperature** discharges at **atmospheric pressure**, i.e. $p \approx 1$ atm or $p \approx 760$ Torr where p is the gas pressure.

1.1.5 Are ionized gases in flow control always ideal plasmas?

The answer is no. A counter-example is the corona discharge (see sections 1.3, 5.2 and 7.2) where there is no region of charge balance so the quasi-neutrality condition is violated. An example of ideal plasma, though, is the body of a streamer (see sections 1.3 and 5.3) which is a quasi-neutral thin conductive channel, millimeters long and 0.1-1 mm wide [28]. The particle density and electron temperature are typically $n = 10^{19}$ m^{-3} and $T_e = 11600$ K. So with $\varepsilon_r = 1$ and $\tau_{\text{col}} = 10^{-10}$ s [123], from eq. (1.1) and section 1.1.2 we have $n\lambda_D^3 \approx 130$, $\lambda_D \approx 2.35$ μm and $\omega_e \tau_{\text{col}} \approx 18$. Therefore, the ionized gas inside the streamer body is by definition an ideal plasma.

In the literature, the term plasma is frequently used in the loose sense, that means a "plasma" could refer to an ionized gas not strictly being an ideal plasma according to the criteria of section 1.1.3. This practice could be sometimes confusing for readers. For example, a plasma actuator[5] does not actually always generate plasmas as in low-voltage, low-current active mode, it could only produce corona discharges.

[5] a small electric device capable of generating ionized gas (see section 1.3)

1.2 Low-temperature discharges in air

1.2.1 Production and decay of particles in low-temperature discharges

A feature in the dynamics of ionized gases that distinguishes them from fluid mechanics is the role of chemical reactions that constantly change the composition of the gas. These plasma reactions, as they are designated, are the consequences of inelastic collisions between particles in which the total kinetic energy of the reactants is not conserved. Some of the kinetic energy could be absorbed in one of the reactants, potentially causing the particle to have an internal energy higher than the ground state and become excited. The amount of supplied kinetic energy could eventually exceed the ionization potential of the target particle and cause one or more electrons to cease to be bound to the particle, and the particle becomes ionized. Conversely, some of the reactions are associative in the sense that the reactants combine to form new species.

A plasma reaction is characterized by the reactants, the products and the **rate coefficient**. More precisely, if S, Q are the number of reactants and products, k is the rate coefficient and the chemical equation reads as

$$a_1 R_1 + \cdots + a_S R_S \to b_1 P_1 + \cdots + b_Q P_Q, \tag{1.2}$$

with a_s ($s = 1, \ldots, S$) the stoichiometric coefficient of the reactant R_s and b_q ($q = 1, \ldots, Q$) the stoichiometric coefficient of the product P_q, then the rate of production of P_q is

$$\frac{\mathrm{d}n_{P_q}}{\mathrm{d}t} = b_q k n_{R_1}^{a_1} \ldots n_{R_S}^{a_S}, \tag{1.3}$$

and the rate of decay of R_s is

$$\frac{\mathrm{d}n_{R_s}}{\mathrm{d}t} = -a_s k n_{R_1}^{a_1} \ldots n_{R_S}^{a_S}. \tag{1.4}$$

Here n_P denotes the particle density of the specie P.

In non-equilibrium discharges in air, the number of reactants as well as products is very high given the obvious complex chemical composition of air. For low-temperature gases with $200\ \mathrm{K} \leq T_0 \leq 500\ \mathrm{K}$, the **kinetic model**[6], which is the set of chemical equations of the form (1.2) as well as their associated rate coefficient, could contain as many as 450 reactions [79]. In this section, we only introduce some of the most important plasma reactions in air.

Electron impact ionization

Ionization of particles by electron impact is the most important charge generation mechanism in the bulk of a gas discharge [126]. The generic form of the chemical equation of this reaction reads as

$$e + M \to 2e + M^+,$$

where e denotes an electron, M denotes an electrically neutral particle in air and M^+ denotes an ion that carries one positive charge. The associated rate coefficient is usually denoted as α ($\mathrm{m^3\ s^{-1}}$).

[6] or kinetic scheme

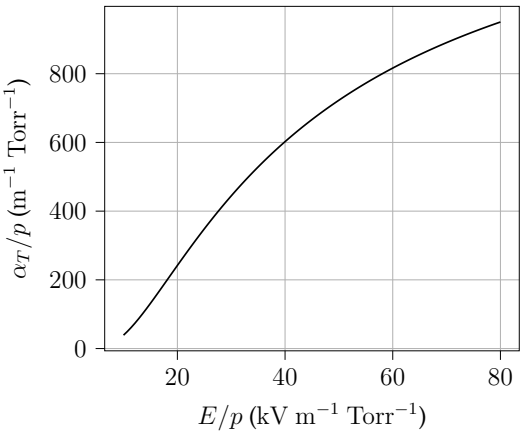


Figure 1.4: Townsend ionization coefficients according to the empirical formula (1.5)

The ionization reaction is also characterized by the **ionization frequency** $\nu_i = \alpha N$ (s^{-1}) in which N denotes the particle density of air. This frequency gives the number of ionization events per second produced by an electron. If ν_i is constant and the decay of electron population is neglected, then the electron density proliferates exponentially as $n_e(t) = n_e^0 \exp(\nu_i t)$ from a seed (initial) density n_e^0. Such rapid multiplications of charge is called an **electron avalanche**.

Another widely used coefficient is the **Townsend ionization coefficient**, which is defined as $\alpha_T = \dfrac{\nu_i}{\mu_e E}$ (m^{-1}) where μ_e ($\mathrm{m}^2\ \mathrm{V}^{-1}\ \mathrm{s}^{-1}$) is the mobility coefficient of electrons (see section 2.1.1) and E is the electric field strength. The Townsend coefficient gives the number of ionization events produced by an electron on a $1\ \mathrm{m}$ path along the field, and conversely, α_T^{-1} characterizes the mean free path covered by an electron before an ionization event.

The Townsend coefficient can be estimated for instance by a well-known empirical formula [126, Chapter 4] suggested by J. S. Townsend,

$$\frac{\alpha_T}{p} \approx A \exp\left(-\frac{Bp}{E}\right), \tag{1.5}$$

with p the gas pressure and E the electric field strength. The parameters A and B are determined by fitting the experiment data on a applicability region of E/p. In air, $A = 1500\ \mathrm{m}^{-1}\ \mathrm{Torr}^{-1}$, $B = 36.5\ \mathrm{kV\ m}^{-1}\ \mathrm{Torr}^{-1}$ and the applicability region is 10 to 80 $\mathrm{kV\ m}^{-1}\ \mathrm{Torr}^{-1}$. The curve (1.5) is shown in fig. 1.4 for the sake of illustration.

Electron attachment

Electron attachment is an important electron removal mechanism in gases that contain electronegative atoms such as oxygen. Since negative ions are too massive to gain sufficient kinetic energy produce ionization, attachment is a process that impedes the multiplication of charges as well as the sustaining of the ionized state of the gas.

The most common attachment processes encountered in electronegative gases are

- direct attachment in which an electron attaches directly to a neutral particle to form a negative

ion, symbolically represented by

$$e + N \to N^-;$$

- dissociative attachment in which a molecule split into their constituent atoms and one of these captures an electron to form a negative ion, symbolically represented by

$$e + AB \to A + B^-.$$

Similar to the electron ionization, the attachment process is characterized by the **attachment coefficient** η, the **attachment frequency** $\nu_a \equiv \eta N$ as well as the **Townsend attachment coefficient** $\alpha_T = \dfrac{\nu_a}{\mu_e E}$. The variation in time of the electron population is $n_e(t) = n_e^0 \exp\left((\nu_i - \nu_a)t\right)$. Therefore, charge multiplication is attenuated whenever $\nu_i < \nu_a$.

Photoionization

This ionization mechanism originates from the radiations that are emitted from the electron impact ionization, excitation as well as deexcitation processes in the gas. Some of these photons could have an energy $\hbar\omega$[7] that exceeds the ionization potential of an atom and ionize it. Although the photoionization process cannot compete with the electron impact ionization under discharge conditions [126, Chapter 4], it could play an important role in certain situations, as they do in supplying seed electrons for electron avalanches in streamer propagation (see [126, Chapter 12] for theoretical study and [139, 7] for numerical evidences).

For low-temperature discharges in air, a photoionization model was proposed by Zheleznyak et al. [166] for photons with wavelength between 98 and 102.5 nm. In this interval, the main source of ionizing photons comes from the deexcitation of nitrogen molecules after inelastic collisions with electrons, and the main source of photoelectrons[8] are oxygen molecules after absorbing the ionizing radiations.

The number of photoionization events per cubic meter per second, $S_{\text{ph}}(\boldsymbol{x})$, can be computed with an integral model in the following way,

$$S_{\text{ph}}(\boldsymbol{x}) = \int_\Omega \frac{g(|\boldsymbol{x}' - \boldsymbol{x}|)h(\boldsymbol{x}')}{4\pi|\boldsymbol{x}' - \boldsymbol{x}|^2}\mathrm{d}\boldsymbol{x}', \tag{1.6}$$

where

$$h(\boldsymbol{x}) = \frac{p_q}{p + p_q}\xi\frac{\nu_p}{\nu_i}\nu_i n_e, \tag{1.7}$$

$$\frac{g(R)}{p_{O_2}} = -\frac{\exp(-\chi_{\max}p_{O_2}R) - \exp(-\chi_{\min}p_{O_2}R)}{p_{O_2}R\left(\ln(\chi_{\max}) - \ln(\chi_{\min})\right)}, \tag{1.8}$$

for $R > 0$. Here, Ω is the discharge volume, $\boldsymbol{x} \in \Omega$, p_q is the pressure arising from quenching by heavy particles[9], ξ is the average photoionization efficiency, ν_p is the number of ionizing photons created

[7] $\hbar$ is the reduced Plank constant, ω is the photon frequency

[8] electrons ensued from photoionization

[9] $p_q = 30$ Torr in air [166]

$\frac{E}{p} \times 10^3$	$\xi\frac{\nu_p}{\nu_i} \times 10^{-2}$
3	5
5	12
10	8
20	6

Table 1.1: $\xi\frac{\nu_p}{\nu_i}$ in function of $\frac{E}{p}$ (V m^{-1} Torr^{-1}) [166]

q	A_q (m^{-2} Torr^{-2})	λ_q (m^{-1} Torr^{-1})
1	1.986	5.53
2	51	14.6
3	4886	89

Table 1.2: Fitting parameters for the three-exponential Helmholtz model [21]

by an electron per second in the absence of quenching and p_{O_2} is the partial pressure of oxygen in the gas; $\chi_{\max} = 200$ and $\chi_{\min} = 3.5$ m^{-1} Torr^{-1} are the maximum and minimum absorption cross sections of oxygen corresponding to the spectrum 98-102.5 nm. The quantity $\xi\nu_p/\nu_i$ depends on the reduced field strength E/p and is given in table 1.1.

The photoionization rate $S_{\text{ph}}(\boldsymbol{x})$ can also be evaluated with an approximation model of (1.6)-(1.8) proposed by Bourdon et al. [21]. The idea is to consider the fitting,

$$\frac{g(R)}{p_{O_2}^2 R} \approx \sum_{q=1}^{Q} A_q \exp(-\lambda_q p_{O_2} R),$$

for $10^{-2} < p_{O_2}R < 0.6$ Torr m, with A_q and λ_q the fitting parameters. As a results, the approximating photoionization rate reads as,

$$S_{\text{ph}}^Q(\boldsymbol{x}) = \sum_{q=1}^{Q} S_{ph,q}(\boldsymbol{x}), \quad S_{ph,q}(\boldsymbol{x}) = \int_\Omega \frac{A_q p_{O_2}^2 \exp\left(-\lambda_q p_{O_2}|\boldsymbol{x}' - \boldsymbol{x}|\right) h(\boldsymbol{x}')}{4\pi|\boldsymbol{x}' - \boldsymbol{x}|}\mathrm{d}\boldsymbol{x}',$$

where the terms $S_{ph,q}$ satisfy the Helmholtz equations,

$$\Delta S_{ph,q}(\boldsymbol{x}) - \left(\lambda_q p_{O_2}\right)^2 S_{ph,q}(\boldsymbol{x}) = -A_q p_{O_2}^2 h(\boldsymbol{x}). \tag{1.9}$$

In [21], the fitting with $Q = 3$ was shown to agree well with the data of [166]. The fitting parameters of this three-exponential Helmholtz model is shown in table 1.2.

The Helmholtz model is interesting in the computation resource point of view. Indeed, in two dimensions, the Helmholtz equations could be solved using a direct linear solver, for example the LU factorization [125] and then forward-backward substitutions. But the LU factorization could be done once for all at the initiation of the simulation if the grid does not change, since the fitting parameters are already known. In this case, the complexity of computing S_{ph}^Q is just $\mathcal{O}(\mathcal{N}^{1.5})$ where $\mathcal{N}$ is the number of unknowns, comparing to $\mathcal{O}(\mathcal{N}^2)$ to compute S_{ph} by the Zheleznyak model (1.6)-(1.8).

For more details on the photoionization process as well as the boundary conditions for the equations, we refer to [131, 112, 21, 29, 118] and the references therein.

Secondary emission

This mechanism is caused by the impact of heavy particles on a cold cathode[10] or a dielectric material that is capable of knocking out some electrons from the surface. Secondary emission is the source of seed electrons for producing and sustaining direct-current discharges between conductors separated by a small gap [126, Chapter 4].

Among the incident particles, the secondary emission generated by positive ions is the most important and is characterized by a **secondary emission coefficient** γ. This coefficient depends strongly on various factors, namely the material of the conductor, the type of incident particles and the impurities on the impact surface. For the numerical simulations in this thesis, though, γ is typically fixed at 10^{-4}.

1.2.2 Discharges in moderate-pressure gases

The first studies of the initiation of a discharge, or **electric breakdown**, were carried out by J. S. Townsend in the 1930s [132]. The discharge device (see fig. 1.5), the "Townsend tube", is composed of two metal plates A (anode) and C (cathode), wired to a direct-current generator and an ohmic resistance R to limit the electric current and impede the formation of an electric arc. The plates are put in a tube containing a gas and the gas pressure p can be modulated.

When the voltage V between the electrodes is sufficiently high, a moderately high current I, typically between 10^{-10} to 10^{-5} A, appears in the circuit due to the absorption of charges to the electrodes. The breakdown voltage - the necessary potential difference to create this current - as well as the nature of the discharge are primarily determined by the inter-electrode distance d and the gas pressure p. The experiments of Townsend were mostly performed under moderately low pressure $p < 100$ Torr and moderately large product pd, $pd < 2$ Torr m. Under these conditions, the electric breakdown occupies the entire gap volume with a relatively homogeneous electric field $E = V/d$. The discharge is triggered by the so-called **Townsend mechanism**.

Breakdown of gases in a constant homogeneous field and Paschen curves

The Townsend breakdown mechanism is based on the principle of multiple electron avalanches that can be described as follows. Assume that there are three species of charged particles in the gas: electrons (e), positive ions (p) and negative ions (n), and the evolution in time of the particle densities is dictated by the electron impact ionization and attachment processes,

$$\frac{\mathrm{d}n_e}{\mathrm{d}t} = (\nu_i - \nu_a)n_e, \quad \frac{\mathrm{d}n_p}{\mathrm{d}t} = \nu_i n_e, \quad \frac{\mathrm{d}n_n}{\mathrm{d}t} = \nu_a n_e.$$

Then the electron and positive ions densities are given by

$$n_e(t) = n_e^0 \exp\left((\nu_i - \nu_a)t\right), \quad n_p(t) = \frac{\nu_i}{\nu_i - \nu_a} n_e^0 \left(\exp\left((\nu_i - \nu_a)t\right) - 1\right).$$

[10] thermal effects are not taken into account

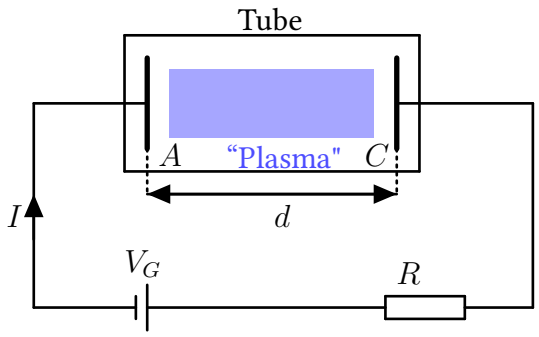


Figure 1.5: Sketch of a Townsend tube

An electron emitted from the cathode surface would create in total $\frac{\alpha_T}{\alpha_T - \eta_T}\left(\exp\left((\alpha_T - \eta_T)d\right) - 1\right)$ positive ions on its way to the anode. This ion population arrive at the cathode and set off secondary emission which generates new $C_r = \gamma\frac{\alpha_T}{\alpha_T - \eta_T}\left(\exp\left((\alpha_T - \eta_T)d\right) - 1\right)$ electrons. C_r is also known as the **electron reproduction coefficient**. It must be noted that the whole process can only take place if in the beginning there is a population n_e^0 of **primary electrons** in the discharge gap, for example the electrons emitted from the cathode surface due to high-energy incident cosmic radiations. Therefore, at the stationary state $t \to \infty$, the total electron density at the cathode $n_{e,C}$ would be

$$n_{e,C} = n_e^0 + n_{e,C}C_r.$$

On the other hand, the electron population $n_{e,A}$ that arrive at the anode (always in the stationary state) would be $n_{e,C}$ amplified by a factor $\exp\left((\alpha_T - \eta_T)d\right)$. Therefore, the total current I appears in the electric circuit would be

$$I = I_0\frac{\exp\left((\alpha_T - \eta_T)d\right)}{1 - C_r} \tag{1.10}$$

where I_0 is the current generated by the primary electrons. It is noted that the negative ions also contribute to the current at the anode, but since their drift velocity is much smaller than the electrons, their contribution can be neglected.

A formula of the type (1.10) was first derived by Townsend to formally explain the breakdown process in a discharge. When the voltage V between the electrodes is below a certain breakdown voltage V_b where V_b is such that $C_r = 1$, then $1 - C_r > 0$ since $\alpha_T - \eta_T$ only increases with V by experimental observations, and the circuit is closed only if there is a preexisting current I_0. On the contrary, if $V = V_b$ and consequently $C_r = 1$, then eq. (1.10) stops making sense. In fact, each primary electron would create C_r/γ positive ions and thus induce exactly one electron from secondary emission. As the process is repeated until stationary state is reached, there is no longer need to take into account the seed electrons. As a result, there is no longer need of the current I_0 to sustain the discharge current I.

As a convention, the discharge is called **non-self-sustaining** if the circuit current I depends on the preexisting current I_0 in the manner of eq. (1.10) and if this is not the case, the discharge is

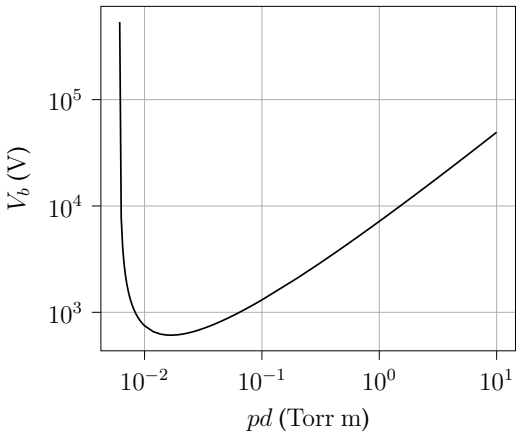


Figure 1.6: The Paschen curve for a Townsend discharge in air

self-sustaining. The transition point $V = V_b$ can be interpreted as the onset condition of breakdown [126, Chapter 7] in a homogeneous field where the Townsend mechanism is predominant.

The breakdown voltage V_b can be determined from the condition $C_r = 1$ with $C_r = \gamma \dfrac{\alpha_T}{\alpha_T - \eta_T} \left(\exp\left((\alpha_T - \eta_T)d\right) - 1\right)$. In the simplest case, with $\eta_T = 0$ (although this is not true since air is an electronegative gas), $E = V/d$ and α_T given by the empirical law (1.5), we obtain

$$V_b = \frac{B(pd)}{\ln(Apd) - \ln\left(\ln\left(\frac{1}{\gamma} + 1\right)\right)}, \tag{1.11}$$

where A, B are the constants from eq. (1.5). The curves $V_b(pd)$ are known as **Paschen curves** and are usually obtained experimentally or numerically. In [126, Chapter 7], we can find numerous Paschen curves for different types of gas. In fig. 1.6, the Paschen curve corresponding to eq. (1.11) with $\gamma = 10^{-4}$ is shown for the sake of illustration.

Regimes of self-sustaining discharge and V-I characteristics

The process of finding the breakdown voltage V_b was performed by increasing gradually the gap voltage V across the electrodes, or more precisely the electromotive force V_G of the power generator (see fig. 1.5), determined by $V_G = V + IR$. As V_G increases, the circuit current I also increases but V does not necessarily and behaves differently for each value range of I. The dependence of gap voltage on the current is known as the ***V*-*I* characteristics** of the discharge and each part of this $V(I)$ curve corresponds to a discharge regime with specific features and operating conditions. A typical V-I characteristics in a Townsend tube is illustrated in fig. 1.7.

We distinguish some four discharge regimes from observation of the V-I curve.

1. **Non-self-sustaining discharge.** This regime corresponding to the segment AB precedes the electric breakdown in the discharge gap. The current is very small $I < 10^{-10}$ A and $V < V_b$. The reproduction of electrons by avalanches is not enough to replace primary electrons.

2. **Townsend dark discharge.** As the electromotive force V_G increases, the gap voltage at some point exceeds the breakdown voltage, the electrons multiplies rapidly to the point that supplant

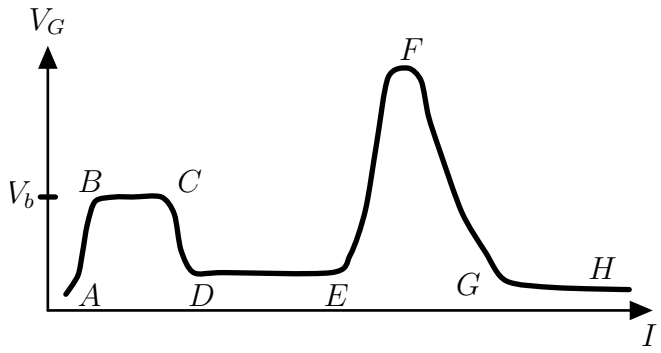


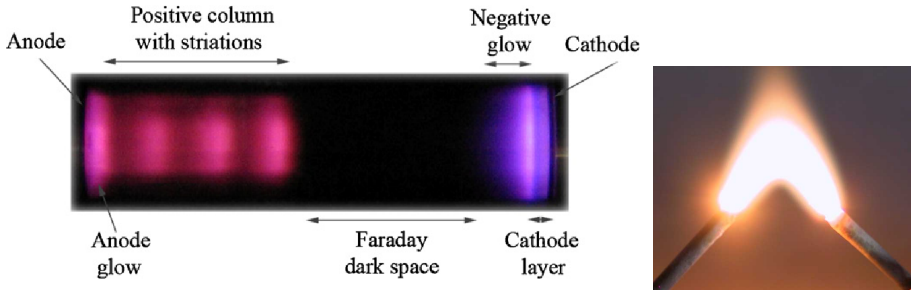
Figure 1.7: A typical V-I curve in a Townsend tube

primary electrons and allow a self-sustaining discharge. As RI is raised, the gap voltage gradually decreases and when V drops to V_b, the current ceases to grow and the discharge reaches a stationary state. The current falls between 10^{-10}-10^{-5} A and the regime corresponds to the segment BC on fig. 1.7.

3. **Glow discharge.** This regime (see section 1.2.2), corresponding to the curve between the points C and F, is characterized by a distorted electric field due to strong charge separation, contrarily to the first two regimes where the field is relatively homogeneous. The field strength concentrates near the cathode in a small region called the cathode fall while it is weaker elsewhere on the gap, which favors the growth of the current since only a small voltage across the cathode fall is required to sustain the avalanches. This explains the plateau DE being lower than the breakdown voltage V_b.

   The region CD corresponds to the transition from a Townsend dark discharge to a glow discharge where the field gradually self-organizes into the configuration described above. This regime is subcategorized as the **subnormal** glow discharge. The **normal** glow discharge (the segment DE), is characterized by luminous spots on the cathode surface which gives the name to the discharge regime. These glow spots concentrates around the impurities on the conductor surface that locally enhance the field, hence intensify the excitation-deexcitation reactions, releasing radiations in the visible range in the process. A specific property of the normal glow regime is that the gap voltage as well as the current density, that is the current per unit surface, in these spots is virtually independent on the total current I. As a result, the glows expand geographically as I increases until they cover all the cathode surface. Once this happens, the voltage increases with the current and the discharge enters the segment EF, the **abnormal** glow regime. Overall, the current of a glow discharge varies between 10^{-4}-10^{-1} A.

4. **Arc discharge.** As the current reach about 1 A, the voltage cascades down the curve FH and an electric arc develops (see fig. 1.8b). Highly-conductive plasma channels begin to form and short-circuits the two electrodes. This regime marks the transition from a low-temperature to a high-temperature discharge.



(a) Glow discharge, taken from [95]

(b) Arc discharge

Figure 1.8: Photographs of some discharge regimes (not necessarily in a Townsend tube)

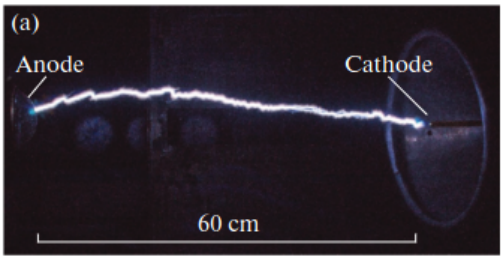


Figure 1.9: Photographs of a spark, taken from [107]

1.2.3 Discharges in atmospheric pressure

The Townsend mechanism is well suited to describe electric breakdown in a tube with $p < 100$ Torr and $pd < 2$ Torr m. In these circumstances, breakdown occurs in a homogeneous field and the cathode plays a crucial role as the principal source of electron reproduction for multiple avalanche events.

The situation is usually very different at atmospheric pressure $p = 760$ Torr and for gaps $d > 0.06$ m or roughly $pd > 40$ Torr m. Firstly, the structure of the field is highly distorted, intertwined with spatial inhomogeneities in the distribution of charges: the discharge develops into very thin channels, called **microdischarges**, each of them could spread into numerous branches. These microdischarges have very short lifetime and could eventually pierce through the whole gap and bridge the electrodes; in this case the microdischarge is known as a **spark** (see fig. 1.9). Secondly, the role of the cathode is irrelevant to the breakdown in such situations. Indeed, the microdischarges advance much faster along the gap than is predicted by the theory of Townsend. The electric current exhibits "strange" intermittent pulses of duration 100 times shorter than the characteristic drift time of ions from the anode to the cathode and thus confirms the invalidity of this theory [132].

It is not clear when the Townsend theory fails and the microdischarges prevail. In [126], it is asserted that the Townsend mechanism could still be realized in air for $pd < 40$ Torr m, but sparks could also appear for $pd > 10$ Torr m. It is clear though that discharges in aeronautical applications do not follow the Townsend mechanism. Indeed, the plasma actuators have typically more complex shape than the two parallel metal plates in a Townsend tube. Therefore, we have to deal with extremely inhomogeneous field right at the beginning.

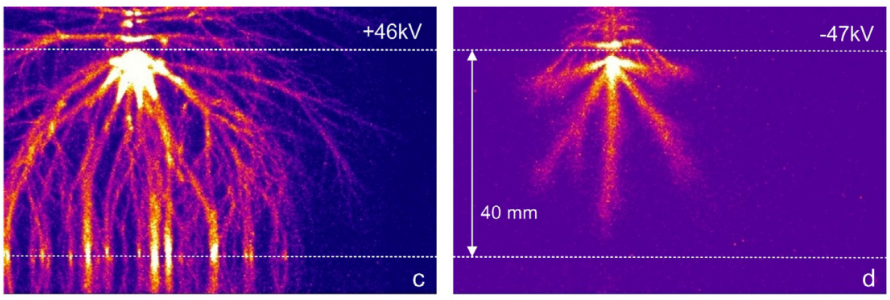


Figure 1.10: Photographs of a positive streamer (left) and a negative streamer (right), taken from [28]

The theory of streamers

Streamer is a specific type of microdischarges that was first theorized by Meek, Loeb and Raether [126, Chapter 12] in the late 1930s to explain the irregular observations in atmospheric discharge in long gaps. The appearance of a streamer is preceded by electron avalanches known as the **primary avalanches**. The difference in mobility between electrons and ions creates a space charge imbalance that increases as the avalanche develops. When certain conditions is met, the ambipolar field[11] due to the charge imbalance begins to surpass the preexisting field generated by the gap voltage, and an "ionization wave" is initiated from the existing charges and plunges into the gap, leaving along a thin, micrometers-wide channel that could possibly branch. Typically, the length, width and propagation velocity of a streamer are on the order of $10\ \mathrm{mm}$, $0.1\text{-}1\ \mathrm{mm}$ and $10^5\text{-}10^6\ \mathrm{m\ s^{-1}}$.

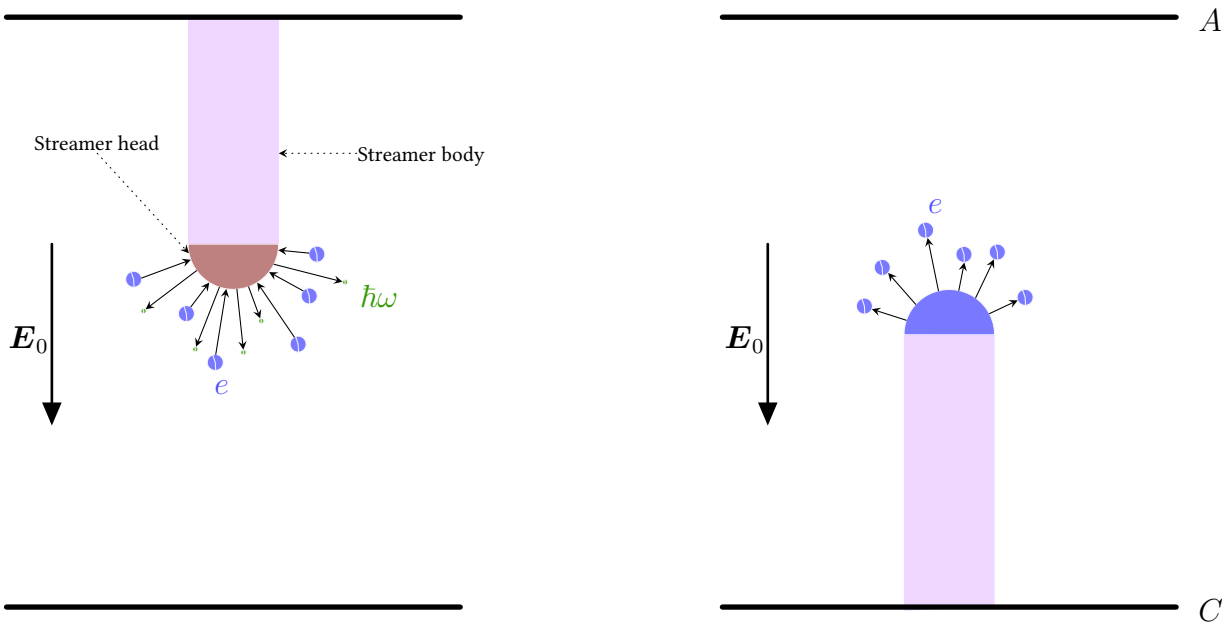
Streamers can be classified into two types: **cathode-directed/positive** streamers and **anode-directed/negative** streamers (see fig. 1.10). The condition for which a streamer is positive or negative depends on the polarity of the excessive charge left behind after the avalanches, and each type of streamer advances along the gap by a different mechanism.

The propagation of a positive streamer is shown schematically in fig. 1.11a. The formation of the streamer starts with a large-gradient layer of positive ions that generates a local ambipolar field that points in the direction of the cathode. If this local field strength surpasses the preexisting field E_0, the resulting total field will be heavily distorted around this large-gradient layer, or **streamer head**, and accelerates electrons in the gap towards it, initiating **secondary avalanches** and creating more electrons as well as positive ions in front of the streamer head. The newborn electrons then move into it, neutralizing electrically the excessive charges in the layer while leaving behind a population of positive charges. The streamer head is thus "displaced" towards the cathode and a thin, quasi-neutral, highly-conductive "trace" is left behind which lengthens gradually and is known as the **streamer body**. The charge density inside the streamer body is typically on the order of $10^{19}\text{-}10^{21}\ \mathrm{m^{-3}}$.

The propagation of a positive streamer requires a seemingly external source of electrons to initiate the secondary avalanches, since the electrons are pulled into the streamer. In fact, these seed electrons are generated by highly-energized radiations emitted from previous avalanches via ionization and deexcitation processes. Hence, photoionization plays a essential role in this case.

A negative streamer, on the other hand, develops and advances along the gap in a somewhat

[11] electric field generated by negative and positive charges moving in opposite directions



(a) Positive streamer

(b) Negative streamer

Figure 1.11: Streamer propagation mechanisms in gap between the anode (A) and the cathode (C). The positive charges are illustrated in maroon, negative charges in blue, photons in green and quasi-neutral ionized gas in purple.

reverse manner. The charge polarity in the streamer head is negative and is formed by electrons and negative ions. If the resulting total field is strong enough, the electrons accelerate outwards from the streamer head and initiate secondary avalanches, then join the newborn positive ions to form a quasi-neutral gas. Contrarily to a positive streamer, a propagation mechanism without photoionization is presumable possible in the case of a negative streamer, since the electrons are created and pushed away from the streamer head, not absorbed into it.

Streamers have been the research subject of gas discharge physics as well as computational physics for many years. We refer to some references and the references therein, such as [98, 126, 121, 28, 114] for much more detailed theoretical studies and [84, 85, 86, 97, 117, 100, 48, 102] for numerical studies.

Breakdown or ignition?

The two terminologies are sometimes used interchangeably, but technically they describe different stages of a discharge. While ignition refers to the transition of a non-self-sustaining to a self-sustaining discharge, the term breakdown is used when an insulating medium becomes (totally) conductive. In the case of Townsend mechanism, the two phenomena rather coincides since ignition happens when avalanches are strong enough to efficiently reproduce electrons, and these avalanches occur on the whole gap which means that the gas is fully ionized. But in the microdischarge mechanism, this is not always the case.

Streamer is self-sustaining, so its formation marks an ignition of discharge, but it might only pierce through a partial portion of the electrode gap. The growth of a streamer depends strongly

on the applied field strength E_0. Indeed, experiments had shown that the stronger the external field is, the longer the streamer is and the faster it propagates [126, Chapter 12] as the preexisting field is the source of energy of the streamer to balance against the resistance of air. Breakdown can only happen when a streamer reaches the opposing electrode and consequently acts as a conductive channel between the two conductors. In this case, the streamer transforms into a **spark discharge** - a precursor of an electric arc.

1.3 Corona and dielectric barrier discharges

Not all ignition processes in atmospheric pressure and in large gaps manifest as microdischarges. In strongly non-uniform fields between a small wire and a plate, or a point and a plate, etc., a low-current, weakly luminous discharge called **corona discharge** could develop for moderate voltages with a charge density on the order of 10^{15}-10^{17} m^{-3}. If the voltage is higher, the microdischarges could occur and potentially transform into a spark, but the situation could be also reversed: microdischarges could be ignited in the first place and then dial down to a corona discharge.

Sparks are often avoided in aeronautical applications since they could potentially heat the conductors to the point of damaging them. To prevent the formation of a spark discharge, a practice frequently employed is to put a dielectric material between the electrodes. In this situation we have a **dielectric barrier discharge** (DBD), which is one of the most extensively studied discharge configurations in aeronautical applications.

1.3.1 Corona discharges

A corona discharge is characterized by a luminous glow (see fig. 1.12) around a metallic wire, point, needle or any conductor with a small radius curvature which is called a **stressed electrode**. Because of the distorted geometry, the electric field is strongly enhanced in the vicinity of the stressed electrode, hence charged particles are mainly produced there by ionization processes and then drift towards the opposite-sign electrode along the weaker field in the inter-electrode gap. Figure 1.13 shows the schematics of some corona discharge actuators.

Corona discharge belongs to the group of self-sustaining discharges, which means that the reproduction of electrons in the region of enhanced field around the stressed electrode is efficient enough to replace preexisting sources of electrons. The mechanism of electron multiplication depends

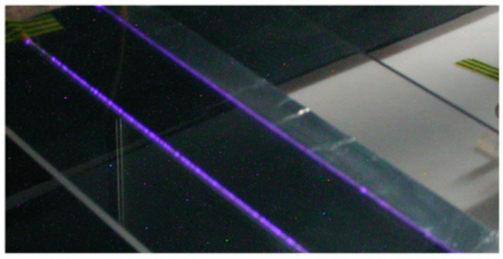
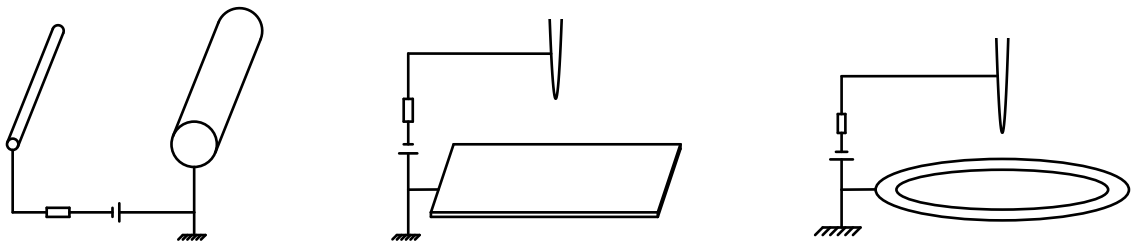


Figure 1.12: Photograph of a wire-to-wire corona discharge: luminous glow is observed in the vicinity of both wires, but brighter near the smaller one (the stressed electrode). Photo taken from [109].



(a) Wire-to-wire (b) Point-to-plate (c) Needle-to-ring

Figure 1.13: Schematics of some corona discharge actuators

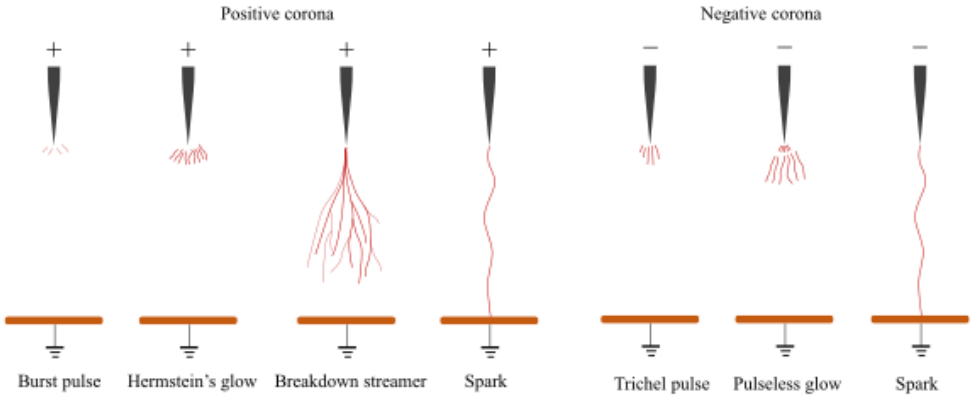


Figure 1.14: Positive and negative corona discharge regimes, taken from [124]. Note that spark was also counted.

essentially, on the other hand, on the polarity of the stressed electrode.

Positive corona discharges

The stressed electrode is the anode, and three discharge regimes appear depending on the applied voltage (see fig. 1.14). The **flashing corona/burst pulse** regime observed at low voltages (5-9.3 kV for a point-to-plane actuator [126, Chapter 12]) is associated to successive scintillating bursts characterized by weak current pulses on the order of $1\ \mu\mathrm{A}$ with frequency up to $10\ \mathrm{kHz}$. Electron avalanches is triggered near the anode and develops towards the cathode. The electrons, due to their light mass, rapidly surround the stressed electrode and shield the gap from the anode potential, leaving behind a residual positive charge. Since the voltage is low from the start, the avalanches do not create enough positive ions to initiate the formation of a streamer. Eventually, the electric field becomes too weak to sustain the charge multiplication, the positive ions drift towards the cathode while the electrons as well as the negative ions are absorbed to the anode, causing the anode shielding to fade away so that new avalanches are produced. The sources of electrons are photoionization, as well as secondary emission from the cathode if the gap is not too large.

For moderate voltages (9.3-16 kV for a point-to-plane actuator), the ignition mechanism is the same as before. However, after the first avalanches and anode shielding, the field is strong enough to continuously sustain the ionization process around the stressed electrode. At the same time, the positive ions keep on with drifting to the cathode. After a while, the densities of charged particles reach a steady state, the current is stabilized and could go up to $10\ \mu\mathrm{A}$. This regime is known as

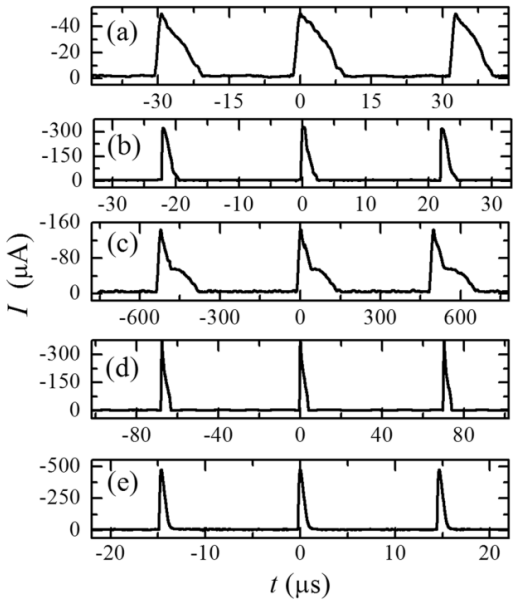


Figure 1.15: Waveform of Trichel pulses at $p = 225$ Torr in *(a)* pure Ar, *(b)* Ar-1%O_2, *(c)* pure N_2, *(d)* N_2-1%O_2 and *(e)* air, taken from [164]

Hermstein's glow [124].

For high voltages (16-29 kV for a point-to-plane actuator), the discharge becomes pulsed again but this time for a different reason. In fact, the applied field is so strong in this case that it generates successive streamers along the gap. This is the **positive streamer regime** which is the precursor of a spark discharge.

Negative corona discharges

Similar to positive corona discharges, negative corona discharges develop from electron avalanches near the stressed electrode which acts as the cathode in this case. Due to the high field, the newborn positive ions drift rapidly towards the cathode and extract electrons from the surface by secondary emission. Two discharge regimes can be observed depending on the gap voltage.

For low voltages, we observe intermittent pulses with a repetition frequency much higher than in the flashing corona regime of a positive corona discharge. These are known as **Trichel pulses**; in air, their frequency is on the order of 1 MHz. The Trichel pulses have been long thought to appear only in electronegative gases such as air, based on the theory of negative corona proposed by R. Morrow [110] back in 1985. However, recent experiment data from Ouyang et al. [164] showed that this discharge regime can also exist in electropositive gases such as (pure) argon and nitrogen. The peak value of pulses in the latter type of gas, nevertheless, are much smaller and the pulse shape are much broader than in electronegative gases (see fig. 1.15); hence, they are more difficult to be observed and measured.

According to the new theory laid out in [164], the positive ions play the dominant role in the mechanism of Trichel pulses, contrarily to the theory in [110] where the negative ions were more of the central figure. The rising time of a Trichel pulse corresponds to the build up of a positive

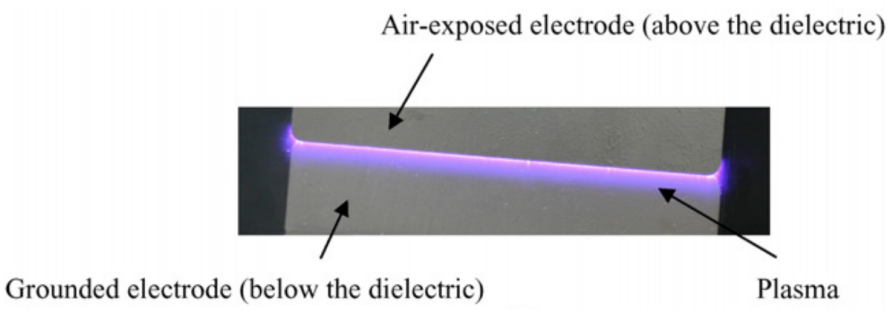


Figure 1.16: Photograph of a SDBD, taken from [109]

charge layer around the cathode, or **cathode sheath**, as electrons are multiplied and rapidly expelled outwards. The decay time of the pulse corresponds to the absorption of positive ions by the cathode and at the same time, the cathode is also shielded, resulting in the disruption of electron avalanches. Because of secondary emission from the cathode surface, the reappearance of a Trichel pulse is imminent. In an electronegative gas, the presence of negative ions has a double effect. On one hand, the local field generated by the accumulation of negative charges is opposite to the preexisting one, so it reduces the cathode shielding by positive ions, leading to higher amplitude of Trichel pulses. On the other hand, the presence of negative charges helps accelerate the neutralization process of positive charges and consequently reduces the pulse duration.

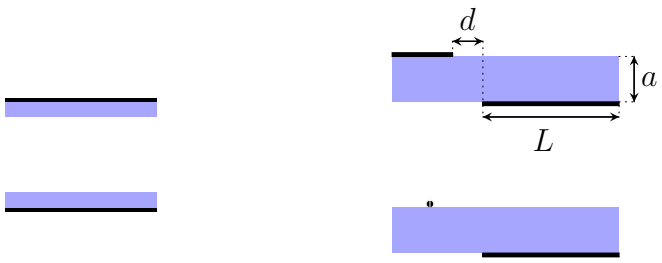
It should be noted that the amplitude of the first pulse is frequently much larger than other Trichel pulses, which was first raised in [41] and could be attested many times in the literature, for example in fig. 1.26.

Finally, as the applied voltage increases, ionization is amplified and the discharge transitions to a **pulseless glow regime** where the current is stabilized and the discharge transforms into a steady glow.

1.3.2 Surface dielectric barrier discharges

As mentioned before, dielectric layers are frequently utilized in aeronautical applications to prevent spark formation that could lead to the erosion of electrodes. Dielectric barrier discharges (DBD) are subcategorized into *(i)* **volume DBD** in which charged particles move in bulk while the electrodes are covered with dielectric materials; and *(ii)* **surface DBD** (SDBD) where the particles move along the dielectric surface. A photograph of a SDBD is shown in fig. 1.16. Simple illustration of some DBD actuators is shown in figs. 1.17a and 1.17b.

If DC generators are used in the presence of dielectric material, the discharge would likely attenuate after while due to the accumulation of charges on the dielectric surface facing the electrode carrying the same-sign voltage, therefore the electrodes are electrically shielded from each other. This is not perfectly true, however, if the voltage is too high, giving rise to a spark slipping aside or even piercing through the dielectric materials in a volume DBD, or creeping from an electrode along the dielectric surface to join the other one in a SDBD [126, 45]. Such situations are of course not preferable. Therefore, AC generators are commonly used in DBD to sustain a spark-free discharge



(a) A plate-to-plate volume DBD actuator

(b) Surface DBD actuators

Figure 1.17: Illustration of some DBD actuators. Electrodes are colored in black, dielectric barriers in blue.

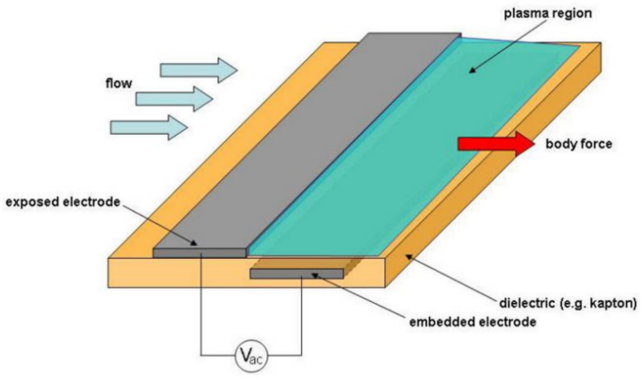


Figure 1.18: Schematics of a plate-to-plate SDBD actuator, taken from [80]

[51].

Although volume DBD actuators were invented at first as early as 1933 [158], recently SDBD actuators have been extensively studied since they could be easily installed on target surfaces such as airfoil for aeronautical applications. Figure 1.18 shows the schematics of a plate-to-plate SDBD actuator, first developed by Roth et al. [128], which consists of two metal plates flush-mounted on a dielectric barrier. One of them is insulated in a dielectric material to prevent unwanted ionization reactions, the other one is exposed in air.

The electrical and mechanical effects of a SDBD actuator depend strongly on different parameters such as electrode gap, dielectric thickness, dielectric material, AC waveform as well as the shape of electrodes [109]. Figure 1.19 shows the velocity profile of the ionic wind measured from experiments [42], of a plate-to-plate and a wire-to-plate actuators (resp. the upper and lower schematics in fig. 1.17b). The AC voltage has a sinusoidal waveform with $1\ \mathrm{kHz}$ frequency and $40\ \mathrm{kV}$ amplitude. The difference between the velocity profiles is clear: while the flow speed of the plate-to-plate discharge decreases during the positive-going cycle[12] and increases during the negative-going cycle[13], for the wire-to-plate discharge it increases during both cycles and only decreases roughly at the

[12] i.e. for $\mathrm{d}V_G/\mathrm{d}t > 0$
[13] i.e. for $\mathrm{d}V_G/\mathrm{d}t < 0$

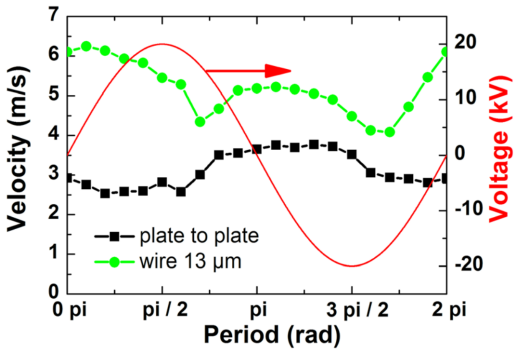


Figure 1.19: Velocity profile during a voltage cycle for a wire-to-plate discharge and a plate-to-plate discharge, taken from [42]

transition points between the two cycles.

Another factor that influences the mechanism of a SDBD is the AC waveform. For long-rising voltages, i.e. the slope $\mathrm{d}V_G/\mathrm{d}t$ is small (in absolute value), with moderate peak value such as an 1-10 kHz-frequency, 5-30 kV-amplitude sinusoidal voltage, the thrusts produced by the actuator is likely generated by momentum transfer from heavy ions to neutral particles. In other words, the effect of ionic wind is predominant in this type of discharge whereas the effect of air heating due to energy transfer from the electric field and particle collisions can be neglected, as pointed out by Enloe et al. [51].

The situation is opposite for short-rising voltages (large $\mathrm{d}V_G/\mathrm{d}t$) as the ionic wind is measured quasi-zero and the discharge effect on a flow relies heavily on gas heating mechanisms [146]. Indeed, numerical simulations with pulse voltages, with rise and decay time on the order of 10 ns, amplitude of 14-42 kV and repetition frequency of a few kHz [146, 18], confirmed that the aerodynamic effect of the actuator came from the formation and propagation of shock waves near the bare electrode tip, which are generated by a local pressure of a few kPa due to a sharp increase of gas temperature caused by fast energy deposition from the discharge. The shock waves propagate at the speed 450 m s^{-1} during the first 100 ns then quickly slow down to roughly the sound speed (about 350 m s^{-1}).

In this section, we give a quick review of **SDBDs** that are produced with **long-rising AC** voltages and with **plate-to-plate** actuators as shown in fig. 1.18.

Electrical properties

Figure 1.20, extracted from [109], shows the current measurement during a sinusoidal cycle with 20 kV-amplitude and 300 Hz-frequency. It is revealed that the discharge nature differs significantly when voltage slope changes its sign. The current during the positive-going cycle exhibits lots of high-value peaks which indicates that the discharge is mainly composed of microdischarges, while it seems to be more homogeneous during the negative-going cycle. High-speed photography using a photomultiplier tube from [49] (see fig. 1.21) validates the observations from the current measurement. Indeed, the discharge structure in the negative-going cycle is visibly more diffusive and uniform than the filamentary structures that appear in the positive-going cycle. This phenomenon was explained in [39] by an argument on the electron sources in the following way: during the negative-going cycle,

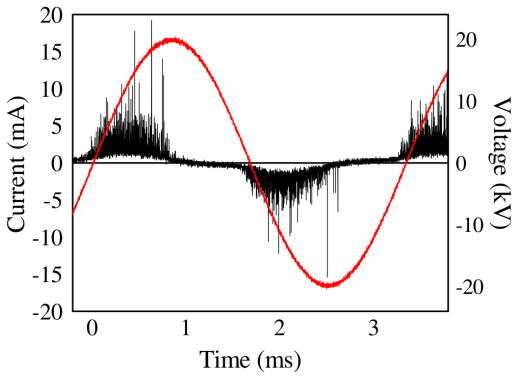


Figure 1.20: Typical electric current during a sinusoidal cycle of a SDBD, taken from [109]

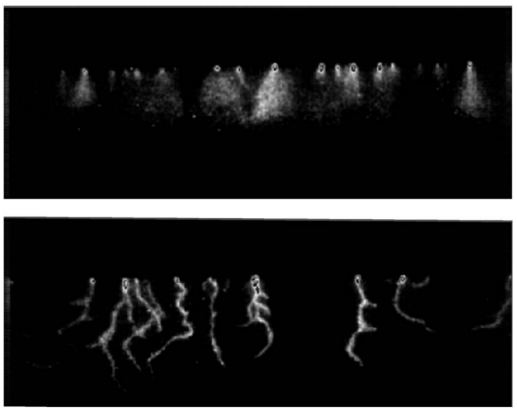


Figure 1.21: High speed photography of a sinusoidal SDBD in the middle of the negative-going cycle (above) and the positive-going cycle (below), taken from [49]

the electrons originate from the bare electron which is an infinite and instantaneous source that facilitates the ignition process. On the other hand, the source of electrons during the positive-going cycle are the secondary emissions from the dielectric surface which are apparently not continuous; the electrons in this cycle then come in the form of fewer, larger microdischarges. The different nature of discharge in each half-cycle was also observed earlier in [51].

While it seems that each half-cycle of a sinusoidal SDBD could be treated simply as a corona discharge, the situation is much more complicated due to the interaction of the discharge with the dielectric barrier. Opaits et al. [116] measured the dielectric surface charging after $15\ \mathrm{s}$ of discharge with a $5\ \mathrm{kV}$-amplitude, $3\ \mathrm{kHz}$-frequency sinusoidal voltage and discovered that the time-averaged surface charge is always positive. An explanation was provided which stated that the electron mobility is much larger than the positive ions', so that the electron current that flows into the bare electrode during the positive-going cycle is much larger and consequently the residual charge after each sinusoidal cycle is positive. They also discovered that while the ionization region only extended some few millimeters downstream of the bare electrode, the dielectric charge extended centimeters far more downstream. This could be explained considering the facts that the ionized gas outside the electrode sheath is mostly an ion-ion mixture due to the dominance of electron attachment while the ion-ion recombination characteristic time is on the order of milliseconds, combining with a jet velocity of several $\mathrm{m\ s^{-1}}$, translate into several centimeters of downstream surface charge.

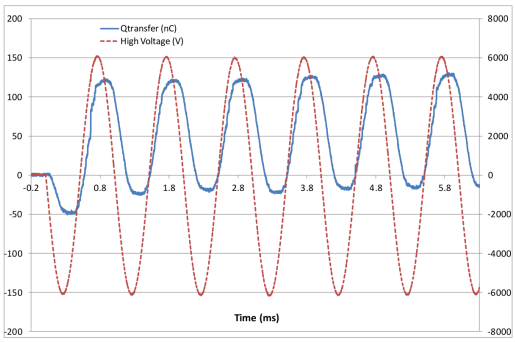


Figure 1.22: Total surface charge in a SDBD in function of time, taken from [73]. Note that the left axis measures the surface charge (nC) while the right axis measures the voltage (V).

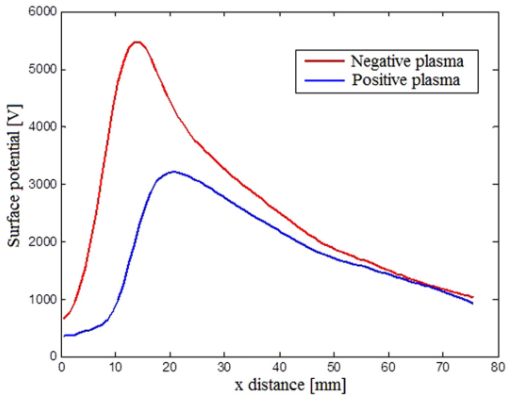


Figure 1.23: Surface charges after cutting off the generator of a 15 kV, 5 kHz sinusoidal voltage, taken from [40]

Hong et al. [73] considered the discharge as a variable current source in series with a resistance to measure the total charge deposited on the dielectric surface as a function of time. Their results, with a 3 kV, 1 kHz-voltage, showed that the surface charge varies periodically with the same frequency as the voltage but with a phase delay about 0.1 ms (see fig. 1.22). The surface charge is positive after the positive-going phase and negative otherwise, but the peak positive charge is much more significant than the peak negative one which is consistent with the measures of [116].

Critofolini et al. [40] measured the punctual surface charge after cutting off the generator at each mid-cycle when the voltage is zero. They found out, with a 15 kV, 5 kHz-voltage, that the surface charge of the "positive plasma" (i.e. after cutting off at the positive-going mid-cycle) is only half of that of the "negative plasma" (see fig. 1.23), and they are both positive which is also consistent with [116]. They gave an explanation that "high energetic electrons emitted from the cathode cause secondary emission thus generate positive holes on the dielectric surface". But it seems more convincing to look at fig. 1.22, and reason on the fact that at the cut-off moment at the positive-going mid-cycle (at zero voltage), the surface charge is still negative, so the residual discharge could only veneer the surface with few positive ions. On the other hand, at the cut-off moment at the negative-going mid-cycle, the surface charge is still at peak, so the surface charge measured afterwards is still high.

The maximum positive potential on the dielectric surface measured in [116, 40] was around 30%

the voltage amplitude, that is large enough to distorted the applied field and strongly influences the SDBD mechanism. We shall see in the next section the mechanical effects of each discharge half-cycle.

Mechanical effects

The mechanical effects induced by a SDBD actuator could be expressed by three parameters: ionic wind velocity, electric power and thrusts [109].

Porter et al. [123] conducted a measurement of thrusts in a wind tunnel using a momentum balance that was placed some centimeters downstream of the bare electrode. They showed that the time-averaged thrusts as well as the electric power for a constant-amplitude voltage (5-10 kV) are proportional to the frequency in the range 5-20 kHz. This suggests that the efficiency of the actuator, defined as the ratio of thrusts to electric power, is independent of the voltage frequency for a constant-amplitude voltage. Enloe et al. [51, 50], on the other hand, deduced from measurements by a mass balance, that thrusts as well as electric power are proportional to $V_{G,\max}^{3.5}$ for a fixed frequency between 1-10 kHz, where $V_{G,\max}$ is the voltage amplitude in the range 5-10 kV. This suggests that the actuator efficiency is also independent of the voltage amplitude. Further experiments from these works pointed out that thrusts and subsequently the actuator efficiency begin to roll off for higher voltages and/or higher frequencies where the discharge displays more filamentary structures. Therefore, a uniform, filament-free discharge in the voltage range 5-10 kV and frequency range 1-10 kHz seems to be the most efficient mean of transferring momentum from the discharge to the air.

Enloe et al. [51] indicated at the same time that the negative-going cycle produces greater thrusts than the positive-going one (in air). This phenomenon could be explained by looking at the dielectric surface charge at the onset of each half-cycle, for example fig. 1.22. Indeed, the dielectric surface is heavily positively charged at the time the negative-going cycle begins, thus it could sustain the discharge during a long time since the voltage difference is huge. On the other hand, the surface is negatively charged at the start of the positive-going cycle but its charge amplitude is much smaller, hence the condition during this cycle is less favored to sustain a discharge. It must be noted that the thrusts produced during the negative-going cycle is due to the presence of negative ions. Indeed, removing oxygen (an electronegative agent) from the air surrounding the actuator, although only changes slightly the discharge current, but could result in a dramatic reduction of thrusts [38].

A frequently discussed topic ensued from thrusts disparity between the two cycles is whether the direction of the thrusts is the same for both. This led to the theories of push-push and push-pull thrusts, where the net momentum transfer during both cycles points away from the bare electrode in the first scenario, whereas it is in the opposite direction during the positive-going cycle in the second scenario. Some experiments have been conducted without always being consistent with each other. For instance, the results of [123] favored the push-pull scenario, while [49] supported the push-push one. In the end, the push-push scenario seems to prevail according to [39] as evidences in its favor are more numerous. But it is worth noting that the small push or pull produced during the positive-going cycle accounts for less than 10% of that produced by the negative-going one [123, 49], so it does not really have a remarkable aerodynamic effect.

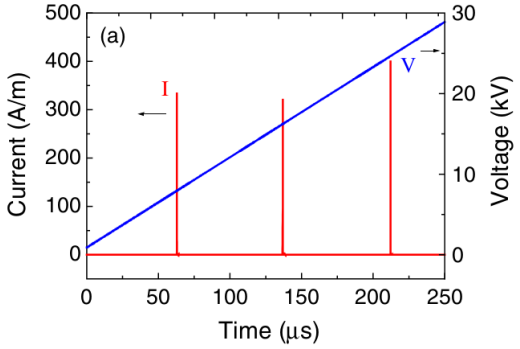


Figure 1.24: Applied voltage and computed electric current in function of time with the voltage slope $\eta_V = 100$ V μs^{-1}, taken from [19]

Numerical simulations

There have been a large array of numerical works since the late 90s in order to shed light on the mechanism of SDBD. Boeuf et al. [19] considered SDBD in **pure N_2** at atmospheric pressure with linearly increasing positive voltages (positive ramp voltages) and found out that the discharge consists of large-amplitude current pulses during which streamers are formed and spread along the dielectric surface, both away from (cathode-directed) and towards (anode-directed) the bare electrode. The cathode-directed streamer propagates several millimeters with a velocity on the order of 10^5 m s^{-1}. Between two pulses, the discharge recedes to a low-amplitude current regime during which the positive ions are drifted away from the bare electrode similarly to a positive corona discharge. Figure 1.24 shows the evolution of the computed current per unit electrode length[14] I/L in time during the voltage rise with a voltage slope $\eta_V = \mathrm{d}V_G/\mathrm{d}t = 100$ V μs^{-1}. We can clearly observe three current peaks between 300-400 A m^{-1} corresponding to the streamer regime and low-amplitude current regions (around 0.1 A m^{-1}) corresponding to the corona regime. After each streamer decay, the dielectric surface receives a huge amount of positive charges which impede the immediate formation of another streamer, but as the voltage increases still, streamer ignition happens again.

In [19], the EHD force density[15], the total EHD force in time and the total EHD force over time is resp. evaluated in the following way,

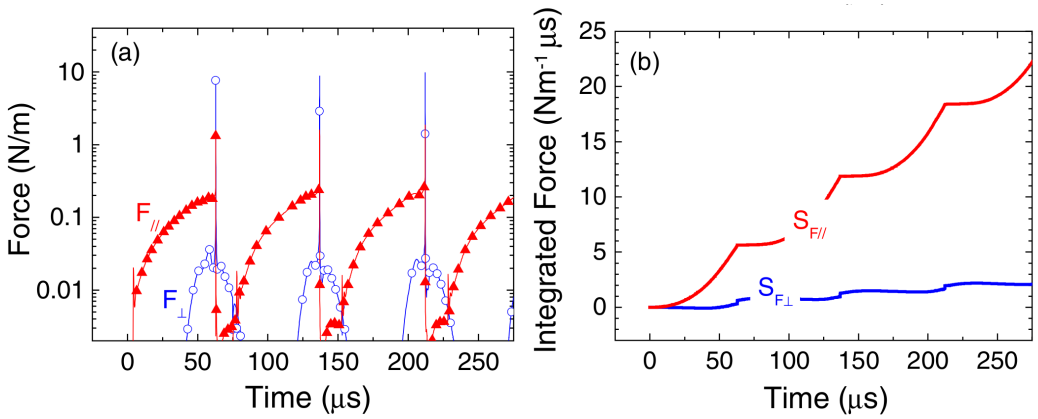
$$\boldsymbol{f}_{\mathrm{EHD}}(t,\boldsymbol{x}) \equiv \rho(t,\boldsymbol{x})\boldsymbol{E}(t,\boldsymbol{x}), \quad \boldsymbol{F}_{\mathrm{EHD}}(t) \equiv \int_\Omega \boldsymbol{f}_{\mathrm{EHD}}(t,\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \quad \boldsymbol{S}_{\mathrm{EHD}}(t) \equiv \int_0^t \boldsymbol{F}_{\mathrm{EHD}}(s)\mathrm{d}s$$

with ρ the total charge density, $\boldsymbol{E}$ the electric field and Ω the computation domain. During the streamer phase, $\boldsymbol{F}_{\mathrm{EHD}}$ is much larger comparing to during the corona phase (see fig. 1.25a), but overall the contribution to the total force over time of the corona phase is much larger than the streamer phase (see fig. 1.25b), since the latter occurs on a time scale very short (on the order of 10 ns) comparing to the former (on the order of 10 μs).

Lagmich et al. [87] subsequently extended the work of [19] in **ambient air**. The characteristics

[14] without ambiguity, we simply address this quantity as "current" for short

[15] needs to be distinguished from thrusts, as the latter is the resulting force after taking into account the shear force on the dielectric surface, acting on the opposite direction of the EHD force [123]



(a) $\boldsymbol{F}_{\mathrm{EHD}}/L$ (b) $\boldsymbol{S}_{\mathrm{EHD}}/L$

Figure 1.25: Parallel and perpendicular components (to the dielectric surface) of $\boldsymbol{F}_{\mathrm{EHD}}/L$ and $\boldsymbol{S}_{\mathrm{EHD}}/L$ in function of time, taken from [19].

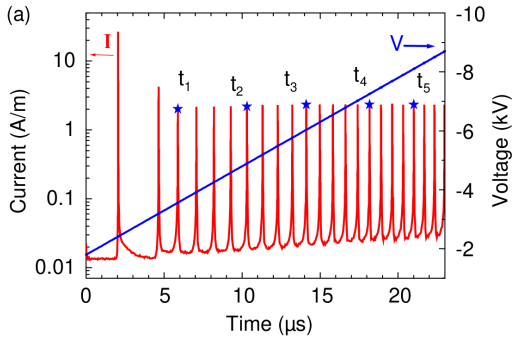


Figure 1.26: Applied voltage and computed electric current (in absolute values) in function of time with the voltage slope $\eta_V = -300$ V μs^{-1}, taken from [87]

of the discharge for a positive ramp voltage is the same as in pure nitrogen. The expansion of the positive ion cloud during the corona phase, with a characteristic speed of 100 m s^{-1}, is conditioned by the extension of the discharge during the streamer phase as the dielectric surface portion under the streamer is positively charged after the streamer decay. This point has been more clarified in [81]. On the other hand, for a linearly decreasing negative voltage (negative ramp voltage), the current exhibits high-frequency pulses (about 1 MHz) where the current amplitude is on the order of 10 A m^{-1} during the first pulse and 1 A m^{-1} during the subsequent pulses (see fig. 1.26). This behavior is similar to the Trichel pulses during a negative corona discharge. The expansion of the negative ion cloud is a "creeping" process along the dielectric surface.

Boeuf et al. [15] later extended the work of [87] for sinusoidal SDBD in ambient air. For a 15 kV, 10 kHz-voltage, the current characteristics during the positive-going and negative-going phases are resp. similar to the positive and negative ramp voltages observed in [19, 87]. The EHD force density $\boldsymbol{f}_{\mathrm{EHD}}$ during the negative-going phase seems to be smoothly distributed in space and time and extends up to 5 mm downstream of the bare electrode. On the other hand, $\boldsymbol{f}_{\mathrm{EHD}}$ during the positive-going phase is higher but is confined in small intermittent time windows and only extends up to 1-3 mm along the dielectric surface, reflecting the filamentary nature of the discharge in this

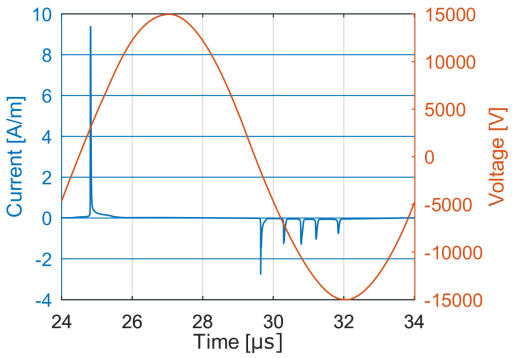


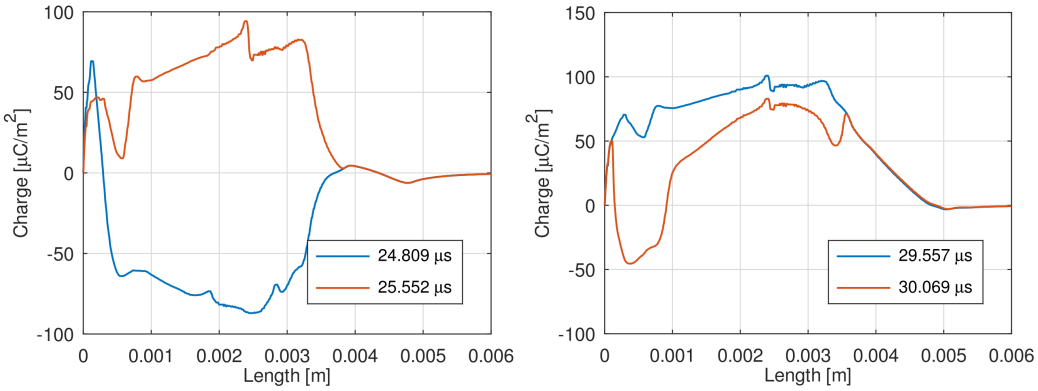
Figure 1.27: Applied voltage and computed electric current in function of time during the third voltage period (10 kV, 100 kHz), taken from [81]

phase.

Kourtzanidis et al. [81] further shed light on the mechanism of sinusoidal SDBD. Figure 1.27 shows the computed current during the third period of a 10 kV, 100 kHz sinusoidal voltage, when the discharge as well as the mutual influence between two phases are well established. The current consists of a large peak in the positive-going phase due to the propagation of a streamer and 5 smaller peaks in the negative-going phase similar to the Trichel pulses. The dielectric surface charging before and after each pulse was emphasized in this paper.

Figures 1.28a and 1.28b show the number of charges per surface unit before and after the positive pulse and the first negative pulse. During the positive-going phase, a streamer is formed and propagates along, but 10-20 μm above, the dielectric surface during tens of nanoseconds and expands up to 3.5 mm downstream of the bare electrode (corresponds to the length of the negative-charged portion of the surface). The positive ion and electron densities inside the streamer body is as high as 10^{21} m^{-3}. After the streamer stops growing and starts decaying, the electrons quickly drift towards the anode and the positive ions attach themselves to the dielectric surface, which explains the sudden change of sign of the surface charge after the positive pulse. The positive charge region ends in particular at a position around 3.5 mm on the dielectric surface (see fig. 1.28a) which corresponds to the maximum expansion of the decayed streamer. This configuration conditions the morphology of the ionic wind during the corona phase, as mentioned in [87], since the positive-charged surface portion would "nudge" the positive ions towards the lightly-charged portion up to 5 mm. The positive-charged portion, in addition, does not provide a favorable condition for the ignition of another streamer.

On the contrary, during a negative pulse, a negative streamer is formed but quickly collapses into the bare electrode (cathode). Provided the short lifetime of the microdischarge, the electron density unleashed onto the dielectric surface is not particularly high so the surface charge after the first pulse only slightly changes and is still positive beyond 1 mm from the bare electrode (see fig. 1.28b). As a result, the ignition condition is still in favor for another microdischarge. Furthermore, the positive surface charge around 3.5 mm from the bare electrode acts like a virtual anode (see [81, Figure 15]) that attracts negative ions: the negative ion cloud expands in a "creeping" way along the surface. In absence of powerful streamer(s), the charge density (consequently the EHD force density) during



(a) Positive pulse (b) Negative pulse

Figure 1.28: Charge per surface unit before (in blue) and after (red) the positive pulse and the first negative pulse, taken from [81].

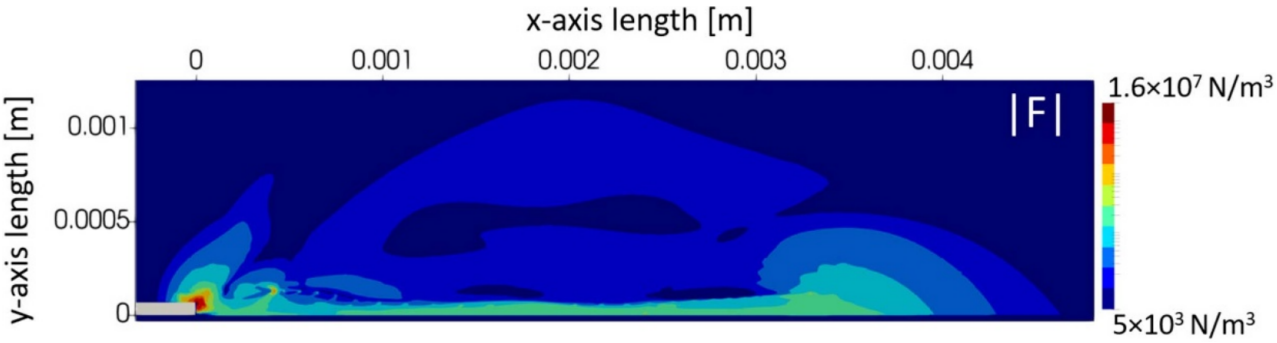


Figure 1.29: Magnitude of the time-averaged EHD force density over one sinusoidal period, taken from [81]. The force in the red sheath around the bare electrode directs towards it due to the absorption of positive ions into the cathode during the negative discharge. This point is coherent with earlier experimental observations that the air entrained by the actuator comes from above the bare electrode [123].

the negative-going phase is more smoothly distributed than the positive-going phase as previously mentioned.

The mutual relation between the positive-going and negative-going cycles is thus remarkable, as it seems that the geographical expansion of the positive streamer determines, to a certain degree, the positive charge distribution on the dielectric surface after the positive-going cycle, which later affects the morphology of the discharge during the negative-going cycle. The negative surface charge in its turn conditions the extension of the positive streamer. The mutual influences between the SDBD cycles balance themselves into a causal loop (as it seems since the analysis was done on the third voltage period) that makes a SDBD much more complex than two separate ramp discharges. Although the contribution to the EHD force of the positive streamer is negligible comparing to the negative discharge, it sure does have an effect on the morphology of the EHD force. Figure 1.29 shows that the time-averaged EHD force density has a elongated shape over the dielectric surface and its maximum length is about $4\ \mathrm{mm}$, approximately the streamer length.

Chapter 2

Mathematical Modeling of Electric Discharge in Air

2.1 Macroscopic description of gas discharge dynamics and drift-diffusion-Poisson equations . . . 48
2.1.1 Local field approximation (LFA) model . . . 49
2.1.2 Local energy approximation (LEA) model . . . 53
2.2 Boundary conditions . . . 53
2.2.1 For the conservation laws . . . 53
2.2.2 For the Poisson equation . . . 54
2.2.3 Wall potential and circuit current . . . 54
2.3 Initial data . . . 55
2.4 Summary of the discharge models . . . 55
2.5 Validity limit and improvements of the LFA model . . . 56
2.5.1 Derivation of the LFA model from the LEA model . . . 57
2.5.2 Improvements of the ionization source term for the LFA model . . . 59

2.1 Macroscopic description of gas discharge dynamics and drift-diffusion-Poisson equations

The dynamics of gas particles can be described on different spatial scales. On the microscopic scale, the Klimontovich equation [74] keeps track of the motion of each particle of the same specie (in the framework of classical mechanics). The complete description requires the knowledge of position and momentum of all constituent particles, which is apparently not an attainable work. On the mesoscopic scale, this point-wise information (on the position and momentum) is "smoothed out" by a complex averaging process that evolves some manipulations with statistical mechanics and in the end we only need to keep track of the probability distribution $f_s(t, \boldsymbol{x}, \boldsymbol{v})$ of a particle specie s that was briefly mentioned in section 1.1. The evolution of $f_s(t, \boldsymbol{x}, \boldsymbol{v})$ is described by so-called kinetic equations and we refer to [74, Chapter 2] for the passage from the Klimontovich equations to the kinetic equations.

The computation of kinetic equations is also not attractive from the numerical resource point of view since the phase space $(\boldsymbol{x}, \boldsymbol{v})$ is obviously much larger than the space of $\boldsymbol{x}$. For this reason, numerical simulations of gas discharge (at least for low-temperature plasmas in air) are almost exclusively conducted on the macroscopic scale for industry-oriented applications. The unknowns are now the particle density $n_s(t, \boldsymbol{x})$ of specie s, the mean velocity $\boldsymbol{u}_s(t, \boldsymbol{x})$, the mean thermal energy $\overline{\varepsilon}_s(t, \boldsymbol{x})$, etc. which are the average of the moments of $f_s(t, \boldsymbol{x}, \boldsymbol{v})$ on the velocity space (see section 1.1).

Among a huge variety of macroscopic models that could be derived from the Navier-Stokes-Maxwell equations, the **drift-diffusion-Poisson equations** are in particular interesting since they have a simple structure (the unknowns are only the $n_s(t, \boldsymbol{x})$ and $\overline{\varepsilon}_s(t, \boldsymbol{x})$) but still "rich" enough to produce simulation results that are coherent, or at least not contradictory, to experiment measurements. In fact, all the numerical works mentioned in section 1.3 on the SDBD were performed with drift-diffusion-Poisson equations. We refer to [43, 35] for the derivation of drift-diffusion equations from kinetic equations[1] and to [44] for the derivation of drift-diffusion-Poisson equations from Navier-Stokes-Maxwell equations.

In this thesis, two drift-diffusion-Poisson models are considered. In the first one known as *(i)* **local field approximation model**, the unknowns are only $n_s(t, \boldsymbol{x})$ and $\overline{\varepsilon}_s(t, \boldsymbol{x})$ are assumed to be constant. The second one, *(ii)* **local energy approximation model**, takes into account an extra equation of the mean electron energy $\overline{\varepsilon}_e(t, \boldsymbol{x})$.

In the following sections, we assume that the particle species s are contained inside a bounded domain $\Omega \subset \mathbb{R}^d$ $(d = 1, 2)$ with boundary $\Gamma \equiv \overline{\Omega} \setminus \Omega$ and evolves on a time interval $(0, T)$ where $T > 0$. The computation domain of the potential field, Ω_ϕ, is in general larger than Ω since it needs to contain the dielectric barrier, so we have $\Omega \subset \Omega_\phi$. Figures 2.1a and 2.1b schematize the computation domains of the densities and of the potential field for simulation of an SDBD.

[1] it must be noted that the derivation is based on the assumption that the temperature of heavy particles is equal to that of electrons, i.e. $T_0 = T_e$



(a) Ω (b) Ω_ϕ

Figure 2.1: Computation domains of a SDBD. The electrodes are colored in black, the dielectric barrier in blue and the gas medium in brown. The domain Ω of n_s (left figure) are delimited by thick brown lines, the domain Ω_ϕ of ϕ (right) are delimited by thick black lines.

2.1.1 Local field approximation (LFA) model

Conservation laws

The LFA model is composed of a set of mass conservation laws coupled with the Gauss law that describes the relation between the electric charge density to the potential field. The conservation laws are expressed in the form of drift-diffusion equations and describe the evolution of the particle density of species. Let us denote as $\mathfrak{S}$ the set of all species. The conservation law of the specie $s \in \mathfrak{S}$ is written in the following way,

$$\begin{cases} \partial_t n_s(t,\boldsymbol{x}) + \nabla \cdot \boldsymbol{f}_s\left(\boldsymbol{E}(t,\boldsymbol{x}), \widehat{E}(t,\boldsymbol{x}), n_s(t,\boldsymbol{x})\right) = S_s\left(\widehat{E}(t,\boldsymbol{x}), (n_s(t,\boldsymbol{x}))_{s\in\mathfrak{S}}\right), \\ \boldsymbol{f}_s(\boldsymbol{E}, \widehat{E}, n_s) = \operatorname{sign}(z_s)\mu_s(\widehat{E})\boldsymbol{E}n_s - D_s(\widehat{E})\nabla n_s, \end{cases} \tag{2.1}$$

for $(t,\boldsymbol{x}) \in (0,T)\times\Omega$, where $n_s(t,\boldsymbol{x})$ is the **particle density**, $\boldsymbol{f}_s(\boldsymbol{E},\widehat{E},n_s)$ is the **particle flux density** and $S_s(\widehat{E},(n_s)_{s\in\mathfrak{S}})$ is the production and decay source term of the specie s. In the second equation, which is also known as the **flux equation**, z_s is the charge number of the specie s, $\boldsymbol{E}(t,\boldsymbol{x})$ is the electric field, $\mu_s(\widehat{E}) > 0$ and $D_s(\widehat{E}) > 0$ are resp. the mobility and diffusion coefficients of s which are function of the reduced field strength $\widehat{E} \equiv |\boldsymbol{E}|/N$, where N is the neutral particle density.

S_s is the sum of source terms like the right-hand sides of eqs. (1.3) and (1.4) and the number of terms depends on the kinetic model that is used to describe the chemical reactions in the discharge. For example, we consider two kinetic schemes from [15] and [54] that will be used in the simulations in chapter 7.

Example 2.1 (Kinetic scheme from [15]). *The set of species consists of electrons (denoted as e and $z_e = -1$), generic positive ions (p, $z_p = +1$) and generic negative ions (n, $z_n = -1$), i.e. $\mathfrak{S} = \{e, p, n\}$.*

The plasma-reactions are

$$\begin{cases} \textit{electron impact ionization} & e + M \to 2e + p, \\ \textit{electron-neutral attachment} & e + M \to n, \\ \textit{electron-ion recombination} & e + p \to M, \\ \textit{ion-ion recombination} & n + p \to 2M, \end{cases}$$

where M stands for a neutral particle. The rate coefficients of this three-specie, four-reaction kinetic model are described in table 2.1 and are function of $\widehat{E}$.

Number	Reaction	Rate coefficient	Data/Evaluation	Reference
1	$e + M \to 2e + p$	α ($\mathrm{m}^3\ \mathrm{s}^{-1}$)	by BOLSIG+	[67]
2	$e + M \to n$	η ($\mathrm{m}^3\ \mathrm{s}^{-1}$)	by BOLSIG+	[67]
3	$e + p \to M$	k_{ep} ($\mathrm{m}^3\ \mathrm{s}^{-1}$)	2×10^{-13}	[15]
4	$n + p \to 2M$	k_{np} ($\mathrm{m}^3\ \mathrm{s}^{-1}$)	1.7×10^{-13}	[15]

Table 2.1: Rate coefficients of the three-specie, four-reaction kinetic model

The kinetic source terms, without photoionization, read as,

$$\begin{cases} S_e(\widehat{E}, n_e, n_p) & = \left(\alpha(\widehat{E}) - \eta(\widehat{E})\right) N n_e - k_{ep}(\widehat{E}) n_e n_p, \\ S_p(\widehat{E}, n_e, n_p, n_n) & = \alpha(\widehat{E}) N n_e - k_{ep}(\widehat{E}) n_e n_p - k_{np}(\widehat{E}) n_n n_p, \\ S_n(\widehat{E}, n_e, n_p, n_n) & = \eta(\widehat{E}) N n_e - k_{np}(\widehat{E}) n_n n_p. \end{cases}$$

If photoionization is taken into account, the source term S_{ph} given by eq. (1.9) is added in to the electron and positive ion source terms, i.e.

$$\begin{cases} S_e(\widehat{E}, n_e, n_p) & = \left(\alpha(\widehat{E}) - \eta(\widehat{E})\right) N n_e - k_{ep}(\widehat{E}) n_e n_p + S_{\mathrm{ph}}, \\ S_p(\widehat{E}, n_e, n_p, n_n) & = \alpha(\widehat{E}) N n_e - k_{ep}(\widehat{E}) n_e n_p - k_{np}(\widehat{E}) n_n n_p + S_{\mathrm{ph}}, \\ S_n(\widehat{E}, n_e, n_p, n_n) & = \eta(\widehat{E}) N n_e - k_{np}(\widehat{E}) n_n n_p. \end{cases}$$

Example 2.2 (Kinetic scheme from [54]). *The set of species consists of electrons (e), generic positive ions (p), oxygen negative ions (O^-, $z_{O^-} = -1$), dioxygen negative ions (O_2^-, $z_{O_2^-} = -1$) and ozone*

negative ions (O_3^-, $z_{O_3^-} = -1$), i.e. $\mathfrak{S} = \{ e, p, O^-, O_2^-, O_3^- \}$. The plasma-reactions are

$$\begin{cases} \textit{electron impact ionization} & e + M \to 2e + p, \\ \textit{electron-neutral attachment} & e + O_2 \to O^- + O, \\ \textit{three-body attachment} & e + O_2 + M \to O_2^- + M, \\ O_2^-\textit{-electron detachment} & O_2^- + M \to e + O_2 + M, \\ O^-\textit{-electron detachment} & O^- + N_2 \to e + N_2O, \\ \textit{two-body charge transfer} & O^- + O_2 \to O + O_2^-, \\ \textit{three-body charge transfer} & O^- + O_2 + M \to O_3^- + N, \end{cases}$$

where the O_2 and N_2 densities are fixed at $N_{O_2} = 20\%N$ and $N_{N_2} = 80\%N$. The rate coefficients of this five-specie, seven-reaction kinetic model are described in table 2.2.

Number	Reaction	Rate coefficient	Data/Evaluation	Reference
1	$e + M \to 2e + p$	α ($\mathrm{m^3\,s^{-1}}$)	$\begin{cases} 4.1\times10^5 \exp\left(-\frac{680\times10^{-21}}{E/N}\right)\mu_e E/N & \text{if} \quad E/N < 186\times10^{-21}\mathrm{Vm^2} \\ \left(1+\frac{6\times10^{-57}}{(E/N)^3}\right) 12.5\times10^5 \exp\left(-\frac{1010\times10^{-21}}{E/N}\right)\mu_e E/N & \text{otherwise} \end{cases}$	[54]
2	$e + O_2 \to O^- + O$	η ($\mathrm{m^3\,s^{-1}}$)	$4.3\times10^3 \exp\left(-1.05\left\|5.3 - \ln(E/N) - 21\ln(10)\right\|^3\right)\mu_e E/N$	[54]
3	$e + O_2 + M \to O_2^- + M$	η_{3b} ($\mathrm{m^6\,s^{-1}}$)	$1.59\times10^{-44}\mu_e(E/N)^{-0.1}$	[54]
4	$O_2^- + M \to e + O_2 + M$	k_{d-O2} ($\mathrm{m^3\,s^{-1}}$)	$1.24\times10^{-17}\exp\left(-\left(\frac{179\times10^{-21}}{8.8\times10^{-21}+E/N}\right)^2\right)$	[54]
5	$O^- + N_2 \to e + N_2O$	k_{d-O} ($\mathrm{m^3\,s^{-1}}$)	$1.16\times10^{-18}\exp\left(-\left(\frac{48.9\times10^{-21}}{11\times10^{-21}+E/N}\right)^2\right)$	[54]
6	$O^- + O_2 \to O + O_2^-$	k_{ct-2b} ($\mathrm{m^3\,s^{-1}}$)	$6.96\times10^{-17}\exp\left(-\left(\frac{198\times10^{-21}}{5.6\times10^{-21}+E/N}\right)^2\right)$	[54]
7	$O^- + O_2 + M \to O_3^- + M$	k_{ct-3b} ($\mathrm{m^6\,s^{-1}}$)	$1.1\times10^{-42}\exp\left(-\left(\frac{E/N}{65\times10^{-21}}\right)^2\right)$	[54]

Table 2.2: Rate coefficients of the five-specie, seven-reaction kinetic model (only available for the LFA model)

The kinetic source terms of charged species read as,

$$\begin{cases} S_e\left(\widehat{E}, n_e, n_{O^-}, n_{O_2^-}\right) & = \left(\alpha(\widehat{E}) - \eta(\widehat{E}) - \eta_{3b}(\widehat{E})N_{O_2}\right) N n_e \\ & \qquad + k_{d-O2}(\widehat{E}) N n_{O_2^-} + k_{d-O}(\widehat{E}) N_{N_2} n_{O^-}, \\ S_p(\widehat{E}, n_e) & = \alpha(\widehat{E}) N n_e, \\ S_{O^-}(\widehat{E}, n_e, n_{O^-}) & = \eta(\widehat{E}) N n_e \\ & \qquad - \left(k_{d-O}(\widehat{E})\frac{N_{N_2}}{N} + k_{ct-2b}(\widehat{E})\frac{N_{O_2}}{N} + k_{ct-3b}(\widehat{E})N_{O_2}\right) N n_{O^-}, \\ S_{O_2^-}\left(\widehat{E}, n_e, n_{O^-}, n_{O_2^-}\right) & = \eta_{3b}(\widehat{E}) N_{O_2} N n_e - k_{d-O2}(\widehat{E}) N n_{O_2^-} + k_{ct-2b}(\widehat{E}) N_{O_2} n_{O^-}, \\ S_{O_3^-}(\widehat{E}, n_{O^-}) & = k_{ct-3b}(\widehat{E}) N_{O_2} N n_{O^-}. \end{cases}$$

The name of the model - local field approximation - comes from the fact that the drift-diffusion coefficients μ_s, D_s as well as the reaction coefficients $\frac{\alpha}{N}$, $\frac{\eta}{N}$ are all functions of the **reduced field strength** $\widehat{E}$.

The floor density hypothesis

In practice, the electron density is assumed to be always larger than a user-defined density $\psi(\boldsymbol{x})$, called the **floor density**, which represents the smallest electron density that exist because of various plasma-chemical processes that are unaccounted for in the discharge model. The floor density hypothesis means that the constraint

$$n_e(t,\boldsymbol{x}) \geq \psi(\boldsymbol{x}), \quad (t,\boldsymbol{x}) \in (0,T)\times\Omega,$$

is imposed on the electron density. This point will be addressed carefully in chapter 6.

Poisson equation

The potential field variation does not only depend on the (volumetric) charge density, but also on the charges accumulated on the dielectric surface $\Gamma_d \equiv \overline{\Omega}\cap\overline{\Omega_\phi\setminus\Omega}$ (see fig. 2.1b). The surface Γ_d can be described by the equation $g_{\Gamma_d}(\boldsymbol{x})=0$ in Cartesian coordinates, where the parametric function g is chosen such that $g_{\Gamma_d}(\boldsymbol{x})>0$ for $\boldsymbol{x}\in\Omega$ and $g_{\Gamma_d}(\boldsymbol{x})<0$ for $\boldsymbol{x}\in\Omega_\phi\setminus\overline{\Omega}$. The accumulated charges on Γ_d per surface unit, σ_{Γ_d}, is given by

$$\begin{cases} \sigma_{\Gamma_d}(t,\boldsymbol{x}) = -\displaystyle\int_0^t \boldsymbol{j}(s,\boldsymbol{x})\cdot\frac{\nabla g_{\Gamma_d}(\boldsymbol{x})}{|\nabla g_{\Gamma_d}(\boldsymbol{x})|}\delta(g_{\Gamma_d}(\boldsymbol{x}))\mathrm{d}s, & (t,\boldsymbol{x})\in(0,T)\times\Omega_\phi,\\ \boldsymbol{j}(t,\boldsymbol{x}) = q\displaystyle\sum_{s\in\mathfrak{S}} z_s\boldsymbol{f}_s(t,\boldsymbol{x}), & (t,\boldsymbol{x})\in(0,T)\times\Omega,\\ \boldsymbol{j}(t,\boldsymbol{x}) = 0, & (t,\boldsymbol{x})\in(0,T)\times(\Omega_\phi\setminus\overline{\Omega}), \end{cases} \tag{2.2}$$

where $\delta(x)$ is the delta function.

Finally, the Gauss law is expressed in the form of a Poisson equation which is written in the following way,

$$\begin{cases} -\nabla\cdot(\varepsilon_r(\boldsymbol{x})\varepsilon_0\nabla\phi(t,\boldsymbol{x})) = \rho(t,\boldsymbol{x}) + \sigma_{\Gamma_d}(t,\boldsymbol{x})|\nabla g_{\Gamma_d}(\boldsymbol{x})|, & (t,\boldsymbol{x})\in(0,T)\times\Omega_\phi,\\ \rho(t,\boldsymbol{x}) = q\displaystyle\sum_{s\in\mathfrak{S}} z_s n_s(t,\boldsymbol{x}), & (t,\boldsymbol{x})\in(0,T)\times\Omega,\\ \rho(t,\boldsymbol{x}) = 0, & (t,\boldsymbol{x})\in(0,T)\times(\Omega_\phi\setminus\overline{\Omega}), \end{cases} \tag{2.3}$$

where $\varepsilon_r(\boldsymbol{x})$ is the relative permittivity of the domain, ε_0 is the vacuum permittivity, $\phi(t,\boldsymbol{x})$ is the **potential field**, $\rho(t,\boldsymbol{x})$ is the electric charge density and q is the elementary charge. The second term on the right-hand side of the first equation is the equivalent volume charge density of a surface distribution of charges [113]. The **electric field** is computed from the potential as

$$\boldsymbol{E} = -\nabla\phi. \tag{2.4}$$

2.1.2 Local energy approximation (LEA) model

The LEA model has basically the same structure as the LFA model except that the coefficients such as μ_s, D_s, $\frac{\alpha}{N}$, $\frac{\eta}{N}$ now are functions of the **mean electron energy** $\overline{\varepsilon}_e$ instead of the reduced field strength, and the evolution of the **electron energy density** $n_\varepsilon \equiv n_e\overline{\varepsilon}_e$ is taken into account in addition[2].

The conservation law of n_ε is also written in the form of a drift-diffusion equation,

$$\begin{cases} \partial_t n_\varepsilon(t,\boldsymbol{x}) + \nabla\cdot\boldsymbol{f}_\varepsilon\left(\boldsymbol{E}(t,\boldsymbol{x}),\overline{\varepsilon}_e(t,\boldsymbol{x}),n_\varepsilon(t,\boldsymbol{x})\right) = S_\varepsilon(\boldsymbol{E}(t,\boldsymbol{x}),\overline{\varepsilon}_e(t,\boldsymbol{x}),n_e(t,\boldsymbol{x})), \\ \boldsymbol{f}_\varepsilon(\boldsymbol{E},\overline{\varepsilon}_e,n_\varepsilon) = -\mu_\varepsilon(\overline{\varepsilon}_e)\boldsymbol{E}n_\varepsilon - D_\varepsilon(\overline{\varepsilon}_e)\nabla n_\varepsilon, \end{cases} \tag{2.5}$$

with (see [64]),

$$\begin{cases} \mu_\varepsilon(\overline{\varepsilon}_e) = \dfrac{5}{3}\mu_e(\overline{\varepsilon}_e), \quad D_\varepsilon(\overline{\varepsilon}_e) = \dfrac{5}{3}D_e(\overline{\varepsilon}_e), \\ S_\varepsilon(\boldsymbol{E},\overline{\varepsilon}_e,n_e) = -q\boldsymbol{f}_e(\boldsymbol{E},\overline{\varepsilon}_e,n_e)\cdot\boldsymbol{E} - k_\varepsilon(\overline{\varepsilon}_e)Nn_e. \end{cases} \tag{2.6}$$

The two term of the energy source term represent resp. mean energy gain from heating by the electric field (Joule heat) and mean energy loss in collisions[3].

2.2 Boundary conditions

2.2.1 For the conservation laws

The boundary conditions for the conservation laws are imposed on the particle flux $\boldsymbol{f}_s$. Different boundary conditions exist and each is defined on a non-overlapping boundary portion Γ_k. The union of these portions is the domain boundary Γ: $\Gamma = \bigcup_k \Gamma_k$ (see fig. 2.1a).

Boundary conditions for particle density have been the subject of still-ongoing debates but to our knowledge, quite a few formulations have been investigated [16, 65, 59] without definitive conclusions. This subject is not in the scope of this thesis. In ONERA's plasma solver COPAIER, we use a set of boundary conditions that are described below. In our opinion, the effects of boundary conditions on the discharge, especially on the formation and evolution of electrode sheaths and/or of microdischarges need to be carefully clarified in future works.

1. On free-flow boundaries Γ_f, which are fictive interfaces between the discharge domain and the outer air space, we assume that they are far enough from the "core" discharge so that the density gradients are null, i.e.

$$f_s \equiv \boldsymbol{f}_s\cdot\boldsymbol{\nu} = n_s\boldsymbol{u}_s\cdot\boldsymbol{\nu}, \quad s\in\mathfrak{S}, \tag{2.7}$$

   where $\boldsymbol{\nu}(\boldsymbol{x})$ is the unit outward normal of Ω at $\boldsymbol{x}\in\Gamma$ and $\boldsymbol{u}_s \equiv \mathrm{sign}(z_s)\mu_s\boldsymbol{E}$ is the drift velocity[4] of specie s.

[2] we have $\overline{\varepsilon}_e \in \mathfrak{S}$, but whenever $\overline{\varepsilon}_e$ passes into subscripts, the notations are simplified to, for example, μ_ε, D_ε, etc.
[3] note that the unit of k_ε is $\mathrm{J\ m^3\ s^{-1}}$
[4] by convention, $z_\varepsilon = -1$

2. On symmetric boundaries Γ_s, which are fictive interfaces between symmetric discharge domains, the boundary conditions are

$$f_s = 0, \quad s \in \mathfrak{S}. \tag{2.8}$$

3. On wall boundaries Γ_w, which are physical interfaces between the discharge domain and the electrodes or dielectric barriers, we assume that there is no particle reflection from the walls and only electron secondary emission due to ion bombardment is involved. The boundary conditions read as

$$\begin{cases} f_s = n_s \max\left(\boldsymbol{u}_s \cdot \boldsymbol{\nu}, 0\right) + \dfrac{\overline{v}_s}{2} n_s, & s \in \mathfrak{S} \setminus \{ e, \overline{\varepsilon}_e \}, \\ f_e = n_e \max\left(\boldsymbol{u}_e \cdot \boldsymbol{\nu}, 0\right) + \dfrac{\overline{v}_e}{2} n_e - 2\gamma f_p, \\ f_\varepsilon = n_\varepsilon \max\left(\boldsymbol{u}_\varepsilon \cdot \boldsymbol{\nu}, 0\right) + \dfrac{\overline{v}_\varepsilon}{2} n_\varepsilon - 2\overline{\varepsilon}_w \gamma f_p, \end{cases} \tag{2.9}$$

where $\overline{v}_s \equiv \left(\dfrac{8 k_B T_s}{\pi m_s}\right)^{\frac{1}{2}}$ for $\mathfrak{S} \setminus \{ \overline{\varepsilon}_e \}$ and $\overline{v}_\varepsilon = \dfrac{5}{2}\overline{v}_e$ are mean thermal velocities, T_s is the mean kinetic temperature, m_s is the particle mass of the specie s and $\overline{\varepsilon}_w = 2$ eV, meaning that an electron emitted from the wall has an energy of 2 eV. Here, $T_s = 300$ K for $\mathfrak{S} \setminus \{ e, \overline{\varepsilon}_e \}$ (room temperature); $T_e = 11600\ K$ (or equivalent to 1 eV in energy) in the LFA model and $T_e = \dfrac{2}{3}\dfrac{\overline{\varepsilon}_e}{k_B}$ in the LEA model.

2.2.2 For the Poisson equation

Different boundary conditions exist and each is defined on a non-overlapping boundary portion Γ_k. The union of these portions is the domain boundary Γ_ϕ: $\Gamma_\phi = \bigcup_k \Gamma_k$ (see fig. 2.1b). The boundary conditions for the Poisson equation are described in the following way,

$$\begin{cases} \nabla\phi \cdot \boldsymbol{\nu} = 0 & \text{on } \Gamma_f \cup \Gamma_s, \\ \phi = \phi_w & \text{on } \Gamma_w, \end{cases} \tag{2.10}$$

where ϕ_w is the **wall potential** which might not be the same for different walls (for example anode has different potential than cathode).

2.2.3 Wall potential and circuit current

The physical model needs to be closed by determining the potential ϕ_w. In the simplest configuration[5] (see fig. 2.1b) that we have one stressed electrode (corresponding to the boundary $\Gamma_{w,1}$) and one non-stressed electrode ($\Gamma_{w,2} = \Gamma_w \setminus \Gamma_{w,1}$) in series with a power generator V_G and an ohmic resistance R, the non-stressed electrode being grounded ($\phi_w = 0$ on $\Gamma_{w,2}$), we only have to compute ϕ_w on $\Gamma_{w,1}$ which is given by Ohm's law[6],

$$\phi_w(t) = V_G(t) - RI(t).$$

[5] that all the simulations in this thesis are set in

[6] here V_G is a function of t since it could be an alternative-current generator

Here I is the **circuit current** that appears due to the movement of charges as well as the rate of change of the electric field. The current I is determined by the Sato formula [129, 105]

$$I(t) = \int_\Omega \left(\varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}(t, \boldsymbol{x}) + \boldsymbol{j}(t, \boldsymbol{x}) \right) \cdot \nabla \widetilde{\phi}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{2.11}$$

with

$$\boldsymbol{j}(t, \boldsymbol{x}) = \sum_{s \in \mathfrak{S}} \boldsymbol{j}_s(t, \boldsymbol{x}), \quad \boldsymbol{j}_s(t, \boldsymbol{x}) = q z_s \boldsymbol{f}_s(t, \boldsymbol{x}),$$

while $\widetilde{\phi}(\boldsymbol{x})$ is a function that depends only on the geometry of the actuator and satisfies

$$\begin{cases} \Delta \widetilde{\phi} = 0 & \text{in } \Omega, \\ \widetilde{\phi} = 1 & \text{on } \Gamma_{w,1}, \\ \widetilde{\phi} = 0 & \text{on } \Gamma_{w,2}. \end{cases}$$

Further algebraic manipulations that can be consulted in [105] show that ϕ_w (on $\Gamma_{w,1}$) is the solution of the following differential equation,

$$\left(\varepsilon_0 R \int_\Omega \left| \nabla \widetilde{\phi} \right|^2 \mathrm{d}\boldsymbol{x} \right) \frac{\mathrm{d}\phi_w}{\mathrm{d}t} + \phi_w = V_G - R \int_\Omega \boldsymbol{j} \cdot \nabla \widetilde{\phi} \mathrm{d}\boldsymbol{x}. \tag{2.12}$$

2.3 Initial data

In open air, the level of pre-ionization charged species is dictated by natural radioactive decays, especially those of radon [117]. The appearance rate of an electron-positive ion pair is up to 10^9 m^{-3} s^{-1} which yields an equilibrium density of 10^9-10^{10} m^{-3}. On the contrary, the seed negative ion density is usually considered negligible.

For the gas composition in example 2.1 ($\mathfrak{S} = \{ e, p, n \}$), we could set

$$\begin{cases} n_e^0(\boldsymbol{x}) = n_p^0(\boldsymbol{x}) = 10^9 \ \mathrm{m}^{-3}, \\ n_n^0(\boldsymbol{x}) = 0, \end{cases}$$

where $n_s^0(\boldsymbol{x})$ is the initial particle density of s. For example 2.2 ($\mathfrak{S} = \{ e, p, O^-, O_2^-, O_3^- \}$),

$$\begin{cases} n_e^0 = n_p^0 = 10^9 \ \mathrm{m}^{-3}, \\ n_{O^-}^0 = n_{O_2^-}^0 = n_{O_3^-}^0 = 0. \end{cases}$$

2.4 Summary of the discharge models

From eqs. (2.1), (2.3) and (2.5), the discharge system of equations can be written in a more compact form in the following way,

$$\begin{cases} \partial_t \boldsymbol{U}(t, \boldsymbol{x}) + \nabla \cdot \boldsymbol{F}\big((\boldsymbol{E}(t, \boldsymbol{x}), w(t, \boldsymbol{x}), \boldsymbol{U}(t, \boldsymbol{x}) \big) = \boldsymbol{S}(\boldsymbol{E}, w, \boldsymbol{U}), & (t, \boldsymbol{x}) \in (0, T) \times \Omega, \\ -\nabla \cdot (\varepsilon_r(\boldsymbol{x}) \varepsilon_0 \nabla \phi(t, \boldsymbol{x})) = \rho(t, \boldsymbol{x}) + \sigma_{\Gamma_d}(t, \boldsymbol{x}) |\nabla g_{\Gamma_d}(\boldsymbol{x})|, & (t, \boldsymbol{x}) \in (0, T) \times \Omega_\phi, \end{cases} \tag{2.13}$$

where

- $\boldsymbol{U} \equiv (n_s)_{s\in\mathfrak{S}}$ (for the LEA model, the mean electron energy is also a "specie", i.e. $\overline{\varepsilon}_e \in \mathfrak{S}$[7]), $\boldsymbol{F} \equiv (\boldsymbol{f}_s)_{s\in\mathfrak{S}}$, $\boldsymbol{S} \equiv (S_s)_{s\in\mathfrak{S}}$;
- w implies the dependence of the mobility, diffusion and reaction coefficients on *(i)* $w = \widehat{E}$ in the **LFA model** or *(ii)* $w = \overline{\varepsilon}_e$ in the **LEA model**;
- $\boldsymbol{S}$ is given in example 2.1 or example 2.2 and eq. (2.6);
- ρ, σ_{Γ_d} and $\boldsymbol{E}$ are resp. given in eq. (2.3), eq. (2.2) and eq. (2.4).

The system (2.13) is completed with the boundary conditions in eqs. (2.7) to (2.9) for the conservation laws as well as in eqs. (2.10) and (2.12) for the Poisson equation and with the initial data described in section 2.3.

2.5 Validity limit and improvements of the LFA model

In certain specific discharge conditions such as when large-gradient charge density layers are present in the densities or the field, numerical simulations in the framework of the LFA model exhibit some nonphysical properties that can be attributed to the validity limit of the model. In [138], a simulation example was made in a one-dimensional test case where the discharge was initiated in a two-component gas (positive ions and electrons) between a 4 mm gap. The initial population of each charged specie was 10^{20} m^{-3} on the 2 mm left half of the domain and 0 on the right half. A high background field of 10 MV m^{-1}, pointing from left to right, was applied between the gap.

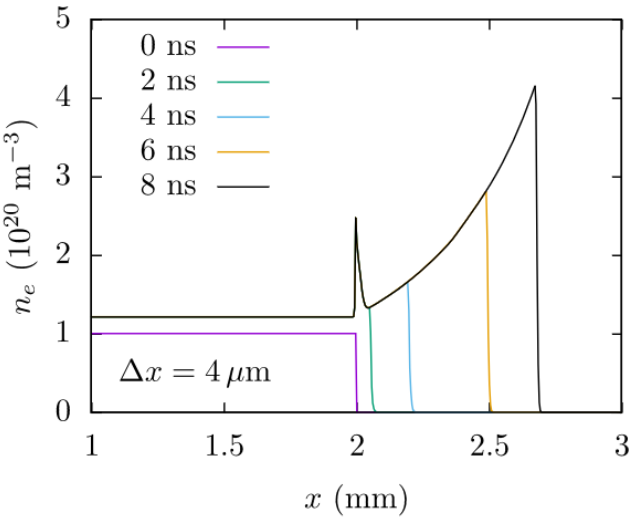


Figure 2.2: Nonphysical growth of electron population computed with the LFA model, taken from [138]

Under these conditions, the electrons should be evacuated to the left and we would not expect any strong ionization process on the 2 mm-right half of the domain. This proves to be not true as we observe the evolution in time of the electron density in fig. 2.2. In fact, electrons from the left

[7] but we write n_ε instead of $n_{\overline{\varepsilon}_e}$

half-domain could escape into the right half due to diffusion, and since the ionization coefficient depends on the field strength in the LFA model, the ionization process rapidly takes place on the right half-domain. This phenomenon could not happen in real world since the escaping electrons (from the left half-domain) will lose their energy moving against the field and cannot effectively ionize the gas as predicted by the simulation [135]. Moreover, the increasing charge density can severely limit the numerical timestep (see chapter 3). It must be also noted that refining the grid did not put off this undesirable behavior, so it is not a numerical artifact but a manifest of the validity limit of the LFA model. Further evidences of the latter can be found in chapter 7 via the simulation results of a microdischarge approaching an electrode surface.

This section presents an analysis of the LFA model that unveils the reason why it does not work in large-gradient density, high field condition. This analysis follows the derivation of the LFA model from the LEA model that was carried out by Arcese et al. [4]. Some amendments of the ionization coefficient for the LFA model [135, 138] which attempt to erase the nonphysical density growth problem will be briefly introduced in section 2.5.2.

2.5.1 Derivation of the LFA model from the LEA model

Since the mean electron energy is concerned in particular, we are interested in the electron particle and energy drift-diffusion equations, and, for simplicity, we only take into account the electron impact ionization source term. The considered equations for the analysis reads as

$$\begin{cases} \partial_t n_e - \nabla \cdot (\mu_e \boldsymbol{E} n_e + D_e \nabla n_e) = \alpha N n_e, \\ \partial_t n_\varepsilon - \nabla \cdot \left(\frac{5}{3} \mu_e \boldsymbol{E} n_e + \frac{5}{3} D_e \nabla n_e \right) = q(\mu_e \boldsymbol{E} n_e + D_e \nabla n_e) \cdot \boldsymbol{E} - k_\varepsilon N n_e. \end{cases} \tag{2.14}$$

Let $\nu_i \equiv \alpha N$ (ionization frequency) and $\nu_\varepsilon \equiv \dfrac{k_\varepsilon N}{\overline{\varepsilon}_e}$ (inverse of the **electron energy relaxation time**). As $n_\varepsilon = n_e \overline{\varepsilon}_e$, system (2.14) is equivalent to

$$\begin{cases} \partial_t n_e - \nabla \cdot (\mu_e \boldsymbol{E} n_e + D_e \nabla n_e) = \nu_i n_e, \\ \partial_t \overline{\varepsilon}_e - \dfrac{1}{n_e} \nabla \cdot \left(\dfrac{5}{3} \mu_e \boldsymbol{E} n_e + \dfrac{5}{3} D_e \nabla n_e \right) + \dfrac{\overline{\varepsilon}_e}{n_e} \nabla \cdot (\mu_e \boldsymbol{E} n_e + D_e \nabla n_e) \\ \qquad\qquad = q \left(\mu_e \boldsymbol{E} + D_e \dfrac{\nabla n_e}{n_e} \right) \cdot \boldsymbol{E} - (\nu_i + \nu_\varepsilon) \overline{\varepsilon}_e. \end{cases} \tag{2.15}$$

In order to properly rescale (2.15) and obtain meaningful information from its dimensionless form, it is important to well define some characteristic parameters in our problem. As ionization and diffusion processes were specifically mentioned, we choose for the characteristic time scale t_0 and space scale x_0 to be

$$t_0 = \frac{1}{\nu_i}, \quad x_0 = \sqrt{\frac{D_0}{\nu_i}},$$

with D_0 the characteristic diffusion coefficient. The characteristic velocity is therefore $u_0 = \sqrt{D_0 \nu_i}$. Furthermore, let n_0, μ_0, E_0, $\overline{\varepsilon}_0$, $\nu_{i,0}$ and $\nu_{\varepsilon,0}$ be resp. the characteristic values of n_e, μ_e, E, $\overline{\varepsilon}_e$, ν_i and

ν_ε. In particular, we also define the characteristic drift velocity as $u_0^d = \mu_0 E_0$. Using the Einstein relation [126, Chapter 2] as the approximation formula of the diffusion equation which reads as

$$D_e = \frac{k_B T_e \mu_e}{q},$$

we have

$$\xi \equiv \frac{u_0^d}{u_0} = \frac{qE_0}{k_B T_e}\sqrt{\frac{D_0}{\nu_i}}.$$

We proceed to introduce a rescaling in system (2.15) by setting

$$\begin{gathered} t = t_0\widetilde{t}, \quad x = x_0\widetilde{x}, \quad n_e = n_0\widetilde{n}_e, \quad \mu_e = \mu_0\widetilde{\mu}_e, \quad \boldsymbol{E} = E_0\widetilde{\boldsymbol{E}}, \\ D_e = D_0\widetilde{D}_e, \quad \nu_i = \nu_{i,0}\widetilde{\nu}_i, \quad \overline{\varepsilon}_e = \overline{\varepsilon}_0\widetilde{\overline{\varepsilon}}_e, \quad \nu_i + \nu_\varepsilon = (\nu_{i,0} + \nu_{\varepsilon,0})\widetilde{(\nu_i + \nu_\varepsilon)}. \end{gathered}$$

Apparently, the tilded values are on the order of 1. Using this change of variables and then omitting the tildes, system (2.15) becomes

$$\begin{cases} \partial_t n_e - \nabla\cdot(\xi\mu_e\boldsymbol{E}n_e + D_e\nabla n_e) = \nu_i n_e, \\ \partial_t\overline{\varepsilon}_e - \dfrac{1}{n_e}\nabla\cdot\left(\dfrac{5}{3}\xi\mu_e\boldsymbol{E}n_e + \dfrac{5}{3}D_e\nabla n_e\right) + \dfrac{\overline{\varepsilon}_e}{n_e}\nabla\cdot(\xi\mu_e\boldsymbol{E}n_e + D_e\nabla n_e) \\ \qquad\qquad = \dfrac{1}{\delta}\left(\left(\xi\mu_e\boldsymbol{E} + D_e\dfrac{\nabla n_e}{n_e}\right)\cdot\boldsymbol{E} - \chi(\nu_i + \nu_\varepsilon)\overline{\varepsilon}_e\right). \end{cases} \tag{2.16}$$

with $\delta \equiv \dfrac{\overline{\varepsilon}_0}{qE_0x_0}$ and $\chi \equiv \dfrac{\overline{\varepsilon}_0(\nu_{i,0} + \nu_{\varepsilon,0})}{qE_0u_0}$.

Near a streamer head, the characteristic values are typically $E_0 = 10^7$ V m^{-1}, $T_e = 10^5$ K (10 eV), $\nu_{i,0} = 10^{11}$ s^{-1} and $D_0 = 10^{-2}$ m^2 s^{-1}. Therefore, $\xi \approx 0.38$.

If we assume that $\delta \ll 1$, the second equation of (2.16) is reduced to

$$\left(\xi\mu_e\boldsymbol{E} + D_e\frac{\nabla n_e}{n_e}\right)\cdot\boldsymbol{E} = \chi(\nu_i + \nu_\varepsilon)\overline{\varepsilon}_e. \tag{2.17}$$

As both sides of this equation are positive (they correspond resp. to the energy gain and loss terms), it is necessary that ξ is on the order of 1, which means that we can choose $\overline{\varepsilon}_0 \propto \dfrac{qE_0u_0}{\nu_{i,0} + \nu_{\varepsilon,0}}$ and consequently $\delta \propto \dfrac{\nu_{i,0}}{\nu_{i,0} + \nu_{\varepsilon,0}}$. Moreover, eq. (2.17) tells us that if μ_e, $\dfrac{\nu_i}{N^2}$ and $\dfrac{\nu_\varepsilon}{N^2}$ are functions of $\overline{\varepsilon}_e$, as well as if $\dfrac{\nabla n_e}{n_e}$ is neglected, then a dependence law of $\overline{\varepsilon}_e$ on $\widehat{E}$ could be figured out.

The LFA model thus can be viewed as the formal limit of the LEA model as $\delta \to 0$ under the hypothesis $\nabla n_e/n_e \approx 0$, the latter is clearly not satisfied near a streamer head (numerical evidences in section 5.3). Furthermore, we have the estimation $\nu_{\varepsilon,0} \approx \sqrt{\dfrac{m_e}{m_0}}\nu_{m,0}$ where m_0 is the typical mass of a neutral particle[8] while $\nu_{m,0}$ is the characteristic effective collision frequency for momentum transfer, $\nu_{m,0} \approx 3\times10^{12}$ s^{-1} [126, Chapter 2]. So $\nu_{\varepsilon,0} \approx 4\times10^8$ s^{-1} which is much smaller than ν_i near a streamer head so $\delta \approx 1$ in fact. This analysis has therefore shown that the LFA model is not valid in certain discharge conditions such as streamers and will raise some numerical problems in certain configurations (see chapter 7).

[8] take nitrogen for example: $m_0 \approx 4.7\times10^{-23}$ g

2.5.2 Improvements of the ionization source term for the LFA model

Due to the drawback of the LFA model that could lead to nonphysical growth of electric charge, the simple remedy is to switch to the LEA model. Nevertheless, some authors still prefer to work with the LFA model because of its simplicity and attractiveness in term of computational cost (one less equation to solve). This leads to some attempts to correct the ionization rate coefficient α since the ionization process are mainly responsible for charge production. As $\frac{\alpha}{N}$, a function of $\widehat{E}$ in the LFA model, is computed, for example, by the application BOLSIG+ [67] where the gas medium as well as the field are assumed to be homogeneous, these attempts are somewhat justified.

Soloviev & Krivtsov [135] are perhaps the first to propose a method to recompute the ionization rate coefficient. Their approach was derived from the dimensional form of eq. (2.17) which reads as

$$q\left(\mu_e \boldsymbol{E} + D_e \frac{\nabla n_e}{n_e}\right) \cdot \boldsymbol{E} = (\nu_i + \nu_\varepsilon)\overline{\varepsilon}_e.$$

The coefficients that we use by default in the LFA model that are denoted here as $\widehat{\nu}_i$ and $\widehat{\nu}_\varepsilon$ are computed by neglecting the density gradient term, i.e.

$$q\mu_e E^2 = (\widehat{\nu}_i + \widehat{\nu}_\varepsilon)\overline{\varepsilon}_e.$$

Therefore, we have

$$(\nu_i + \nu_\varepsilon)\overline{\varepsilon}_e = (\widehat{\nu}_i + \widehat{\nu}_\varepsilon)\overline{\varepsilon}_e \left(1 + \frac{D_e \nabla n_e \cdot \boldsymbol{E}}{\mu_e E^2 n_e}\right),$$

and we set

$$\nu_i = \left(1 + \frac{D_e \nabla n_e \cdot \boldsymbol{E}}{\mu_e E^2 n_e}\right) \widehat{\nu}_i. \tag{2.18}$$

J. Teunissen [138] proposed another ionization frequency in the following way,

$$\nu_i = \frac{|\mu_e n_e \boldsymbol{E} + D_e \nabla n_e|}{\mu_e n_e E} \widehat{\nu}_i.$$

This approach coincides with Soloviev & Krivtsov's in the case where ∇n_e aligns with $\boldsymbol{E}$. If ∇n_e acts on the opposite direction of $\boldsymbol{E}$, for example in the situation observed in fig. 2.2, then $\nu_i < \widehat{\nu}_i$ and we would expect less ionization in the region where it is not supposed to be strong. On the other hand, if ∇n_e acts on the same direction of $\boldsymbol{E}$ then $\nu_i > \widehat{\nu}_i$. Therefore, [138] further proposed to set

$$\nu_i = \min\left(1, \frac{|\mu_e n_e \boldsymbol{E} + D_e \nabla n_e|}{\mu_e n_e E}\right) \widehat{\nu}_i. \tag{2.19}$$

Comparing to eq. (2.18), there is less legit argument for the derivation of eq. (2.19). Nevertheless, the latter approach will be used in the simulations in chapter 7.

Chapter 3

Numerical Modeling of Electric Discharge in Air - the COPAIER Plasma Solver

3.1 Meshing of the computation domain . . . 62
3.2 Discretization of the conservation laws . . . 63
3.2.1 Discretization in space . . . 63
3.2.2 Flux approximation . . . 64
3.2.3 Discretization in time . . . 66
3.2.4 The CFL condition . . . 67
3.3 Approximation of the potential and the plasma coefficients . . . 67
3.3.1 Discretization of the Poisson equation . . . 67
3.3.2 Computation of the drift-diffusion-reaction coefficients . . . 67
3.3.3 The dielectric relaxation time . . . 68
3.3.4 Semi-implicit method for the computation of the potential . . . 69

In this chapter, we present some key features of ONERA's in-house plasma solver COPAIER [46], conceived and developed for two-dimensional simulations of electric discharge in air.

3.1 Meshing of the computation domain

Let Ω be the computation domain of the specie densities (an open bounded domain of $\mathbb{R}^2$). We start with the definition of a **grid** (or **mesh**) on Ω while assuming at first that Ω is a polygonal domain. In COPAIER, structured (rectangular) or unstructured (triangular) grids or the mixing of both can be used.

Definition 3.1. *Let Ω be an open connected polygonal of $\mathbb{R}^2$. A **conforming** grid of $\overline{\Omega}$ is a set $\mathcal{T}$ of open polygonal domains (or **cells**) $(\Omega_K)_{1\leq K\leq \mathcal{N}}$ with $\mathcal{N} = \mathrm{Card}(\mathcal{T})$, that satisfies the following conditions.*

1. *Ω_K are convex.*
2. *No three or more vertices of an Ω_K are aligned.*
3. *$\Omega_K \subset \Omega$ and $\Omega = \bigcup_{K=1}^{\mathcal{N}} \Omega_K$.*
4. *For any two polygons Ω_K and Ω_L of $\mathcal{T}$, the intersection $\overline{\Omega}_K \cap \overline{\Omega}_L$ is either empty or a common vertex or a common segment λ_{KL} (or indifferently λ_{LK}) such that λ_{KL} is an **edge** of both Ω_K and Ω_L.*

*The **nodes** of the grid $\mathcal{T}$ are the vertices of the polygons $\Omega_K \in \mathcal{T}$. By convention, for each $\Omega_K \in \mathcal{T}$, we denote as h_K the diameter of the smallest circle that contains Ω_K, and*

$$h_{\mathcal{T}} = \max_{K=1,\dots,\mathcal{N}} h_K.$$

We also denote as $\mathcal{E}_K$ the set of edges of $\Omega_K \in \mathcal{T}$, and $\mathcal{E}_{\mathcal{T}} = \bigcup_{K=1}^{\mathcal{N}} \mathcal{E}_K$.

The third condition of definition 3.1 which characterized the conforming property of $\mathcal{T}$, ensures that there is no hanging node in the grid, meaning that there is no vertex of a cell of $\mathcal{T}$ that lies on an edge of another cell.

For an edge $\lambda \in \mathcal{E}_K$, we denote as resp. $\boldsymbol{x}_{K\lambda}$ and $\boldsymbol{\nu}_{K\lambda}$ the midpoint of λ and the unit outward normal of Ω_K on λ. In the case that λ is the common edge of Ω_K and its neighbor Ω_L (see for example fig. 3.1), we can use the notations $\boldsymbol{x}^{\lambda}_{KL}$ (or indifferently $\boldsymbol{x}^{\lambda}_{LK}$) and $\boldsymbol{\nu}_{KL}$ instead of $\boldsymbol{x}_{K\lambda}$ and $\boldsymbol{\nu}_{K\lambda}$. Finally, the center of gravity $\boldsymbol{x}^c_K$ of Ω_K is defined as

$$\boldsymbol{x}^c_K = \frac{1}{|\Omega_K|}\int_{\Omega_K} \boldsymbol{x}\mathrm{d}\boldsymbol{x},$$

where $|\Omega_K| \equiv \int_{\Omega_K} \mathrm{d}\boldsymbol{x}$ is the surface of Ω_K.

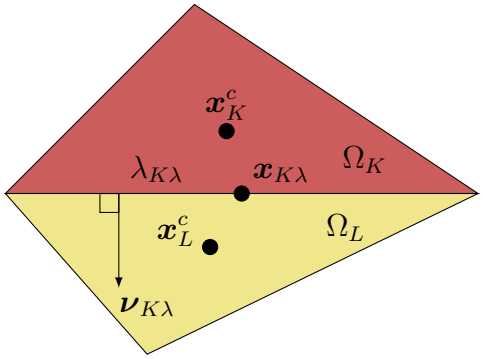


Figure 3.1: Two neighbor cells of a triangular conforming grid

Remark 3.1 ([3, Chapter 6]). *If Ω is not a polygonal domain but its boundary is sufficiently smooth, we start by approximating Ω by a polygonal domain Ω_p and then mesh Ω_p with a grid $\mathcal{T}$. We can choose Ω_p and $\mathcal{T}$ in such a way that*

$$\operatorname{dist}(\Gamma, \Gamma_p) \leq Ch_{\mathcal{T}},$$

where $\Gamma = \overline{\Omega} \setminus \Omega$, $\Gamma_p = \overline{\Omega}_p \setminus \Omega_p$ and $C > 0$ is a constant only depending on the boundary Γ.

For the computation domain Ω_ϕ of the potential, we mesh the subdomain $\Omega_\phi \setminus \overline{\Omega}$ with a conforming grid $\widetilde{\mathcal{T}}$, such that the union $\mathcal{T} \cup \widetilde{\mathcal{T}}$ is also a conforming grid on Ω_ϕ.

The meshing work for all the simulations of COPAIER is done by the free mesh-generator **Gmsh** [58].

3.2 Discretization of the conservation laws

3.2.1 Discretization in space

In COPAIER, we employ the **finite volume method** for the spatial approximation of the conservation laws. Let us assume that the densities $\boldsymbol{U}$ and the source terms $\boldsymbol{S}$ in system (2.13) are locally integrable. Integrating the continuity equations in (2.13) on each $\Omega_K \in \mathcal{T}$ and dividing it by $|\Omega_K|$, we have

$$\frac{\mathrm{d}\overline{\boldsymbol{U}}_K(t)}{\mathrm{d}t} + \frac{1}{|\Omega_K|} \sum_{\lambda \in \mathcal{E}_K} \int_\lambda \boldsymbol{F}(t, \boldsymbol{x}) \cdot \boldsymbol{\nu}_{K\lambda} \mathrm{d}\boldsymbol{l} = \overline{\boldsymbol{S}}_K(t), \tag{3.1}$$

where $\overline{\boldsymbol{U}}_K \equiv (\overline{n}_{s,K})_{s\in\mathfrak{S}}$, $\overline{v}_K \equiv \dfrac{1}{|\Omega_K|} \displaystyle\int_{\Omega_K} v(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$ for any locally integrable function $v(\boldsymbol{x})$, $\boldsymbol{F} \cdot \boldsymbol{\nu}_{K\lambda} \equiv (\boldsymbol{f}_s \cdot \boldsymbol{\nu}_{K\lambda})_{s\in\mathfrak{S}}$ and $\mathrm{d}\boldsymbol{l}$ is the unit length element.

Since $\boldsymbol{S}$ is usually nonlinear in $\boldsymbol{U}$ (see examples 2.1 and 2.2), we can make some approximations in $\overline{\boldsymbol{S}}_K$ so that the resulted source term depends on $\overline{\boldsymbol{U}}_K$ and the computation of $\overline{\boldsymbol{U}}_K$ would be easier.

Taking for instance the kinetic scheme in example 2.1, we can use the following approximations

$$\begin{cases} (\overline{\alpha n_e})_K(t) \approx \alpha(t, \boldsymbol{x}_K^c)\overline{n}_{e,K}(t), \\ (\overline{\eta n_e})_K(t) \approx \eta(t, \boldsymbol{x}_K^c)\overline{n}_{e,K}(t), \\ (\overline{k_{ep} n_e n_p})_K(t) \approx k_{ep}(t, \boldsymbol{x}_K^c)\overline{n}_{e,K}(t)\overline{n}_{p,K}(t), \\ (\overline{k_{np} n_n n_p})_K(t) \approx k_{np}(t, \boldsymbol{x}_K^c)\overline{n}_{n,K}(t)\overline{n}_{p,K}(t). \end{cases} \tag{3.2}$$

For the mean energy source term S_ε, we refer to section 6.4.2 for details.

For a specie $s \in \mathfrak{S}$, the flux-integral $\displaystyle\int_\lambda \boldsymbol{f}_s(t, \boldsymbol{x}) \cdot \boldsymbol{\nu}_{K\lambda} \mathrm{d}\boldsymbol{l}$ is evaluated by the midpoint rule and its approximant reads as

$$|\lambda| \widehat{f}_{s,K\lambda}(t, \boldsymbol{x}_{K\lambda}) \equiv |\lambda| \boldsymbol{f}_s(t, \boldsymbol{x}_{K\lambda}) \cdot \boldsymbol{\nu}_{K\lambda}. \tag{3.3}$$

Therefore, with eqs. (3.2) and (3.3) we replace eq. (3.1) for each $\Omega_K \in \mathcal{T}$ by (we keep the same notation for $\overline{\boldsymbol{U}}_K$)

$$\frac{\mathrm{d}\overline{\boldsymbol{U}}_K(t)}{\mathrm{d}t} + \frac{1}{|\Omega_K|} \sum_{\lambda \in \mathcal{E}_K} |\lambda| \widehat{F}_{K\lambda}(t, \boldsymbol{x}_{K\lambda}) = \widehat{\boldsymbol{S}}_K \left(\boldsymbol{E}(t, \boldsymbol{x}_K^c), w(t, \boldsymbol{x}_K^c), \overline{\boldsymbol{U}}_K(t) \right), \tag{3.4}$$

where $|\lambda| \equiv \displaystyle\int_\lambda \mathrm{d}\boldsymbol{l}$, $\widehat{F}_{K\lambda} \equiv (\widehat{f}_{s,K\lambda})_{s\in\mathfrak{S}}$ and $\widehat{\boldsymbol{S}}_K$ is obtained by replacing resp. $\alpha(t, \boldsymbol{x}_K^c)\overline{n}_e(t)$, etc. for $\overline{\alpha n_e}(t)$, etc. in $\overline{\boldsymbol{S}}_K$.

Remark 3.2. *Since $\widehat{F}_{K\lambda}$, $\alpha(t, \boldsymbol{x}_K^c)$ and $\overline{n}_{e,K}\overline{n}_{p,K}$ are resp. second-order approximations in h_K of $\displaystyle\int_\lambda \boldsymbol{F}(t, \boldsymbol{x}) \cdot \boldsymbol{\nu}_{K\lambda} \mathrm{d}\boldsymbol{l}$, $\alpha(\boldsymbol{x})$ and $(\overline{n_e n_p})_K$ on Ω_K, we have in fact added a numerical error $\mathcal{O}(h_{\mathcal{T}}^2)$ on the passage from eq. (3.1) to eq. (3.4).*

3.2.2 Flux approximation

The flux $\widehat{f}_{s,K\lambda}$ can be written as the sum of the drift flux $\widehat{f}^u_{s,K\lambda}(t) \equiv u_{s,K\lambda}(t) n_s(t, \boldsymbol{x}_{K\lambda})$ and the diffusion flux $\widehat{f}^D_{s,K\lambda}(t) \equiv -D_{s,K\lambda}(t) \nabla n_s(t, \boldsymbol{x}_{K\lambda}) \cdot \boldsymbol{\nu}_{K\lambda}$ (see eq. (2.1)) with $u_{s,K\lambda}(t) \equiv \boldsymbol{u}_s(t, \boldsymbol{x}_{K\lambda}) \cdot \boldsymbol{\nu}_{K\lambda}$, $\boldsymbol{u}_s(t, \boldsymbol{x}) \equiv \mathrm{sign}(z_s)\mu_s(t, \boldsymbol{x})\boldsymbol{E}(t, \boldsymbol{x})$ and $D_{s,K\lambda}(t) = D_s(t, \boldsymbol{x}_{K\lambda})$.

Let us denote as Ω_L the neighbor cell of Ω_K that shares the common edge λ. The drift flux is approximated by the central-difference scheme

$$\widehat{f}^D_{s,KL} \approx \overline{f}^D_{s,KL} \equiv -D_{s,K\lambda} \frac{\overline{n}_{s,L} - \overline{n}_{s,K}}{d_{KL}}, \tag{3.5}$$

where d_{KL} (or indifferently d_{LK}) equals $|\boldsymbol{x}_K^c - \boldsymbol{x}_L^c|$.

The drift flux $\widehat{f}^u_{s,K\lambda}$ is computed using either a first-order upwind scheme or a second-order MUSCL scheme. For the latter, the density gradients are estimated with a dual-mesh algorithm [46] on **triangular** grids. The first-order upwind scheme reads as

$$\widehat{f}^u_{s,KL} \approx \overline{f}^{u,1}_{s,KL} \equiv \max(u_{s,KL}, 0)\overline{n}_{s,K} + \min(u_{s,KL}, 0)\overline{n}_{s,L}. \tag{3.6}$$

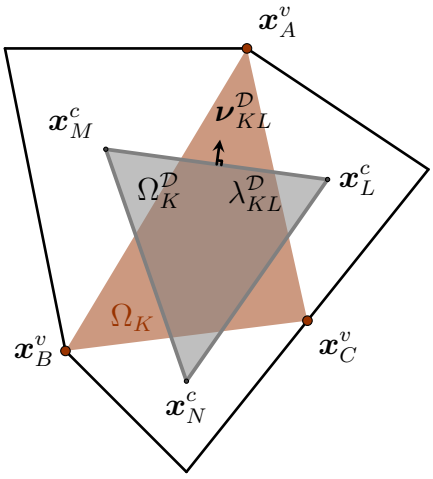


Figure 3.2: Dual cell (gray) $\Omega_K^{\mathcal{D}}$ of a triangle Ω_K (maroon)

For the MUSCL scheme, $\widehat{f}_{s,K\lambda}^u$ is computed in the following way,

$$\widehat{f}_{s,KL}^u \approx \overline{f}_{s,KL}^{u,2} \equiv \max(u_{s,KL},0)\left(\overline{n}_{s,K} + \phi_{s,K}\mathfrak{D}n_{s,K}\cdot(\boldsymbol{x}_{KL}^{\lambda} - \boldsymbol{x}_K^c)\right) + \min(u_{s,KL},0)\left(\overline{n}_{s,L} + \phi_{s,L}\mathfrak{D}n_{s,L}\cdot(\boldsymbol{x}_{LK}^{\lambda} - \boldsymbol{x}_L^c)\right), \quad (3.7)$$

where $\mathfrak{D}n_{s,K}(t)$ is an approximation of the gradient $\nabla n_s(t,\boldsymbol{x}_K^c)$ while $\phi_{s,K}(t)$ is a slope limiter that is added to avoid nonphysical oscillations.

In order to compute $\mathfrak{D}n_{s,K}$, let us define the **dual cell** $\Omega_K^{\mathcal{D}}$ of $\Omega_K \in \mathcal{T}$ as the triangle whose vertices are the centers of gravity of the three neighbor cells Ω_L, Ω_M and Ω_N of Ω_K. Let us denote as $\lambda_{qr}^{\mathcal{D}}$ the segment joining $\boldsymbol{x}_q^c$ and $\boldsymbol{x}_r^c$, and as $\boldsymbol{\nu}_{qr}^{\mathcal{D}}$ the unit outward normal of $\Omega_K^{\mathcal{D}}$ on $\lambda_{qr}^{\mathcal{D}}$ with $q,r \in \{\, L,M,N \,\}$, $q \neq r$ (see fig. 3.2). The approximant $\mathfrak{D}n_{s,K}$ reads as [10]

$$\mathfrak{D}n_{s,K} = \frac{1}{2|\Omega_K^{\mathcal{D}}|}\sum_{q,r\in\{\,L,M,N\,\},q\neq r}(\overline{n}_{s,q}+\overline{n}_{s,r})\lambda_{qr}^{\mathcal{D}}\boldsymbol{\nu}_{qr}^{\mathcal{D}}.$$

In order to compute $\phi_{s,K}$, let us denote as $\boldsymbol{x}_A^v$, $\boldsymbol{x}_B^v$ and $\boldsymbol{x}_C^v$ the three vertices of Ω_K (see fig. 3.2),

$$\overline{n}_{s,K}^{\max} \equiv \max_{q\in\{\,K,L,M,N\,\}}\overline{n}_{s,q}, \quad \overline{n}_{s,K}^{\min} \equiv \min_{q\in\{\,K,L,M,N\,\}}\overline{n}_{s,q},$$
$$n_{s,K,r} \equiv \overline{n}_{s,K} + \mathfrak{D}n_{s,K}\cdot(\boldsymbol{x}_r^v - \boldsymbol{x}_K^c), \quad r \in \{\, A,B,C \,\}.$$

Let us define, for $r \in \{\, A,B,C \,\}$ [6],

$$\phi_{s,K,r} = \begin{cases} \min\left(1, \dfrac{\overline{n}_{s,K}^{\max} - \overline{n}_{s,K}}{n_{s,K,r} - \overline{n}_{s,K}}\right), & \text{if } n_{s,K,r} - \overline{n}_{s,K} > 0, \\ \min\left(1, \dfrac{\overline{n}_{s,K}^{\min} - \overline{n}_{s,K}}{n_{s,K,r} - \overline{n}_{s,K}}\right), & \text{if } n_{s,K,r} - \overline{n}_{s,K} < 0, \\ 1, & \text{otherwise.} \end{cases}$$

Then the limiter $\phi_{s,K}$ is evaluated in the following way,

$$\phi_{s,K} = \min_{r\in\{\,A,B,C\,\}}\phi_{s,K,r}.$$

3.2.3 Discretization in time

With the numerical flux $\overline{f}_{s,K\lambda}\left(w(t,\boldsymbol{x}_{K\lambda}),\overline{\boldsymbol{U}}_K(t)\right)$ as the sum of (3.5) and (3.6) or (3.7), the system of equations (3.4) is transformed into

$$\begin{cases} \dfrac{\mathrm{d}\overline{\boldsymbol{U}}_K(t)}{\mathrm{d}t} = \mathcal{G}(t,\overline{\boldsymbol{U}}_K(t)), \\ \mathcal{G}(t,\overline{\boldsymbol{U}}_K(t)) = \widehat{\boldsymbol{S}}_K\left(\boldsymbol{E}(t,\boldsymbol{x}_K^c), w(t,\boldsymbol{x}_K^c), \overline{\boldsymbol{U}}_K(t)\right) \\ \qquad\qquad -\dfrac{1}{|\Omega_K|}\displaystyle\sum_{\lambda\in\mathcal{E}_K} |\lambda|\overline{F}_{K\lambda}\left(\boldsymbol{E}(t^l,\boldsymbol{x}_{K\lambda}), w(t,\boldsymbol{x}_{K\lambda}), \overline{\boldsymbol{U}}_K(t)\right), \end{cases} \tag{3.8}$$

with $\overline{F}_{K\lambda} = (\overline{f}_{s,K\lambda})_{s\in\mathfrak{S}}$.

Let $T>0$, $0=t^0<t^1<\cdots<t^{\mathcal{L}}=T$ be an increasing series of real numbers call **time levels** and $\Delta t^l \equiv t^{l+1}-t^l>0$ $(l=0,\ldots,\mathcal{L}-1)$ be the **timesteps**. We can even simplify more eq. (3.8) by making the approximations $\boldsymbol{E}(t,\boldsymbol{x})\approx\boldsymbol{E}(t^l,\boldsymbol{x})$ and $w(t,\boldsymbol{x})\approx w(t^l,\boldsymbol{x})$ for $t\in[t^l,t^{l+1})$. For each l, eq. (3.8) is then replaced by

$$\begin{cases} \dfrac{\mathrm{d}\overline{\boldsymbol{U}}_K(t)}{\mathrm{d}t} = \mathcal{G}^l(\overline{\boldsymbol{U}}_K(t)), \quad t\in[t^l,t^{l+1}) \\ \mathcal{G}^l(\overline{\boldsymbol{U}}_K(t)) = \widehat{\boldsymbol{S}}_K\left(\boldsymbol{E}(t^l,\boldsymbol{x}_K^c), w(t^l,\boldsymbol{x}_K^c), \overline{\boldsymbol{U}}_K(t)\right) \\ \qquad\qquad -\dfrac{1}{|\Omega_K|}\displaystyle\sum_{\lambda\in\mathcal{E}_K} |\lambda|\overline{F}_{K\lambda}\left(\boldsymbol{E}(t^l,\boldsymbol{x}_{K\lambda}), w(t^l,\boldsymbol{x}_{K\lambda}), \overline{\boldsymbol{U}}_K(t)\right), \end{cases} \tag{3.9}$$

Now we can discretize eq. (3.9) with a chosen time scheme (see examples 3.1 to 3.4). For $l\geq 0$, we denote as $\overline{\boldsymbol{U}}_K^l$ an approximation of $\overline{\boldsymbol{U}}_K(t^l)$[1].

Example 3.1 (Forward Euler scheme)**.** *$\overline{\boldsymbol{U}}_K^{l+1}$ is computed in the following way,*

$$\overline{\boldsymbol{U}}_K^{l+1} = \overline{\boldsymbol{U}}_K^l + \Delta t^l \mathcal{G}^l(\overline{\boldsymbol{U}}_K^l).$$

Example 3.2 (Runge-Kutta-Heun scheme)**.**

$$\begin{cases} \overline{\boldsymbol{U}}_K^{l+\frac{1}{3}} = \overline{\boldsymbol{U}}_K^l + \Delta t^l \mathcal{G}^l(\overline{\boldsymbol{U}}_K^l), \\ \overline{\boldsymbol{U}}_K^{l+\frac{2}{3}} = \overline{\boldsymbol{U}}_K^{l+\frac{1}{3}} + \Delta t^l \mathcal{G}^l(\overline{\boldsymbol{U}}_K^{l+\frac{1}{3}}), \\ \overline{\boldsymbol{U}}_K^{l+1} = \dfrac{1}{2}\left(\boldsymbol{U}_K^l + \boldsymbol{U}_K^{l+\frac{2}{3}}\right). \end{cases}$$

Example 3.3 (Third-order strong stability-preserving (SSP) scheme [60])**.**

$$\begin{cases} \overline{\boldsymbol{U}}_K^{l+\frac{1}{4}} = \overline{\boldsymbol{U}}_K^l + \Delta t^l \mathcal{G}^l(\overline{\boldsymbol{U}}_K^l), \\ \overline{\boldsymbol{U}}_K^{l+\frac{2}{4}} = \overline{\boldsymbol{U}}_K^{l+\frac{1}{4}} + \Delta t^l \mathcal{G}^l(\overline{\boldsymbol{U}}_K^{l+\frac{1}{4}}), \\ \overline{\boldsymbol{U}}_K^{l+\frac{3}{4}} = \dfrac{3}{4}\overline{\boldsymbol{U}}_K^l + \dfrac{1}{4}\overline{\boldsymbol{U}}_K^{l+\frac{2}{4}} + \mathcal{G}^l\left(\dfrac{3}{4}\overline{\boldsymbol{U}}_K^l + \dfrac{1}{4}\overline{\boldsymbol{U}}_K^{l+\frac{2}{4}}\right), \\ \overline{\boldsymbol{U}}_K^{l+1} = \dfrac{1}{3}\boldsymbol{U}_K^l + \dfrac{2}{3}\boldsymbol{U}_K^{l+\frac{3}{4}} \end{cases}$$

Example 3.4 (Backward Euler scheme)**.**

$$\overline{\boldsymbol{U}}_K^{l+1} = \overline{\boldsymbol{U}}_K^l + \Delta t^l \mathcal{G}^l(\overline{\boldsymbol{U}}_K^{l+1}).$$

[1] $\overline{\boldsymbol{U}}_K^0$ is simply the average of $\boldsymbol{U}^0(\boldsymbol{x})$ on Ω_K

3.2.4 The CFL condition

It is well known that the use of explicit time schemes like those in examples 3.1 to 3.4 imposes a constraint on the numerical timesteps Δt^l called **CFL condition** [90, 91]. In practice, it is expressed in the following way,

$$\Delta t^l \leq \mathfrak{C} \min_{K=1,\dots,\mathcal{N}} \left(\frac{h_K^2}{u_{e,K}^l h_K + 2D_{e,K}^l} \right), \tag{3.10}$$

where $\mathfrak{C} \leq 1$ is a user-defined parameter, $u_{e,K}^l$ and $D_{e,K}^l$ are resp. approximations of $|\boldsymbol{u}_e(t^l, \boldsymbol{x}_K^c)|$ and $D_e(t^l, \boldsymbol{x}_K^c)$. The CFL condition only depends on electrons since they are the lightest specie so their transport-diffusion timescale is much shorter than other heavy species.

3.3 Approximation of the potential and the plasma coefficients

3.3.1 Discretization of the Poisson equation

In COPAIER, the Poisson equation of the potential (2.3) is discretized using the **P1-Lagrange finite element method**. Let $\phi^l(\boldsymbol{x})$ be an approximation of $\phi(t^l, \boldsymbol{x})$. Firstly, we approximate the right-hand-side of the first equation in (2.3) by the cell-averaged charge density. Equation (2.3) is replaced by

$$\begin{cases} -\nabla \cdot \big(\varepsilon_r(\boldsymbol{x})\varepsilon_0 \nabla \phi^l(\boldsymbol{x})\big) = \overline{\rho}_K^l, & \boldsymbol{x} \in \Omega_K, \\ \overline{\rho}_K^l = q \sum_{s \in \mathfrak{S}} z_s \overline{n}_{s,K}^l, & \text{if } \Omega_K \in \mathcal{T}, \\ \overline{\rho}_K^l = 0, & \text{if } \Omega_K \in \widetilde{\mathcal{T}}, \end{cases} \tag{3.11}$$

where we recall that $\widetilde{\mathcal{T}}$ is the conforming grid defined on the subdomain $\Omega_\phi \setminus \Omega$.

The Poisson equation is then multiplied with the finite element basis functions which allow to characterize ϕ^l by its values at the grid **nodes**. The process of solving eq. (3.11) by the finite element method can be found in many textbooks on this subject, such as [3]. The rigid matrix resulting from the discretization of eq. (3.11) is factorized using the **LU decomposition** method [125]. But since the stiffness matrix only depends on the grid (which is unchanged), its factorization is done at the beginning and stored throughout the simulation. Therefore, eq. (3.11) is solved each time merely by forward and backward substitutions.

3.3.2 Computation of the drift-diffusion-reaction coefficients

The drift-diffusion coefficients $u_{s,K\lambda}$ and $D_{s,K\lambda}$ in eqs. (3.5) to (3.7) require an approximation of the field $\boldsymbol{E}$ at edge centers $\boldsymbol{x}_{K\lambda}$, while the reaction coefficients α, η, etc. (for the LFA model) in the right-hand side of eq. (3.2) require an approximation of the field strength E at cell centers $\boldsymbol{x}_K^c$. These quantities of the field can be easily evaluated in the finite element framework as the solution $\phi^l(\boldsymbol{x})$ of eq. (3.11) can be expressed using the finite element basis functions and the approximant $\boldsymbol{E}^l(\boldsymbol{x})$ of

the field $\boldsymbol{E}(t^l, \boldsymbol{x})$ simply is

$$\boldsymbol{E}^l(\boldsymbol{x}) = -\nabla\phi^l(\boldsymbol{x}).$$

3.3.3 The dielectric relaxation time

When solving the discharge system with an explicit time scheme, for instance the forward Euler method in example 3.1, the numerical timesteps Δt^l must also respect a constraint related to the **dielectric relaxation time** [63], which will be introduced in the following.

Let us go back to the system eqs. (2.1) and (2.3) and assume, for practical reason, that $D_s = 0$ for all species s. By definition of the charge density ρ in eq. (2.3), taking the time derivative of ρ, using the continuity equations eq. (2.1) and the fact that $\sum_{s\in\mathfrak{S}} z_s S_s = 0$, we have

$$\partial_t\rho + \nabla\cdot\left(q\left(\sum_{s\in\mathfrak{S}}|z_s|\mu_s n_s\right)\boldsymbol{E}\right) = 0. \tag{3.12}$$

We define the **conductivity** of the gas medium (which is always positive) as $\sigma \equiv q\sum_{s\in\mathfrak{S}}|z_s|\mu_s n_s$. The time-discrete version of eq. (3.12) (using the forward Euler method) reads as

$$\rho^{l+1} - \rho^l + \Delta t^l\nabla\cdot(\sigma^l\boldsymbol{E}^l) = 0, \quad l = 1,\dots,\mathcal{L},$$

where $\rho^l(\boldsymbol{x})$ is an approximation of $\rho(t^l,\boldsymbol{x})$, etc.

With this equation and the Poisson equation (2.3), we have

$$\nabla\cdot(\varepsilon_r\varepsilon_0\boldsymbol{E}^{l+1}) = \nabla\cdot((\varepsilon_r\varepsilon_0 - \Delta t^l\sigma^l\mathbb{I}_\Omega)\boldsymbol{E}^l),$$

with $\mathbb{I}_\Omega$ the indicator function of the domain Ω. Then multiplying this equation with ϕ^{l+1}, integrating on Ω_ϕ and using integration by parts lead to

$$\int_{\Omega_\phi}\varepsilon_r\varepsilon_0|\boldsymbol{E}^{l+1}|^2\mathrm{d}\boldsymbol{x} = \int_{\Omega_\phi}(\varepsilon_r\varepsilon_0 - \Delta t^l\sigma^l\mathbb{I}_\Omega)\boldsymbol{E}^l\cdot\boldsymbol{E}^{l+1}\mathrm{d}\boldsymbol{x}.$$

Here we have made the boundary terms vanish since the boundary conditions are either Dirichlet for ϕ^l or homogeneous Neumann for $\boldsymbol{E}^l$ (see section 2.2). Using the Cauchy-Schwarz inequality, we obtain an estimate on the time-discrete total electrical energy $\mathcal{E}_\phi(t) \equiv \int_{\Omega_\phi}\varepsilon_r\varepsilon_0 E^2\mathrm{d}\boldsymbol{x}$ as follows,

$$(\mathcal{E}_\phi^{l+1})^{\frac{1}{2}} \le \sup_{x\in\Omega_\phi}\left|1 - \Delta t^l\frac{\sigma^l\mathbb{I}_\Omega}{\varepsilon_r\varepsilon_0}\right|(\mathcal{E}_\phi^l)^{\frac{1}{2}}. \tag{3.13}$$

It can be shown that $\mathcal{E}_\phi$ is non-increasing. Indeed, by substituting eq. (3.12) in the time derivative of eq. (2.3), we obtain the local conservation of the total current (displacement current plus conducting current) which reads as follows,

$$\nabla\cdot(\varepsilon_r\varepsilon_0\partial_t\boldsymbol{E} + \sigma\boldsymbol{E}) = 0.$$

Multiplying this equation with ϕ and integrating on Ω_ϕ leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_\phi}\varepsilon_r\varepsilon_0 E^2\mathrm{d}\boldsymbol{x}+\int_{\Omega_\phi}\sigma E^2\mathrm{d}\boldsymbol{x}=0,$$

where we have also made the boundary terms vanish (no outer energy source). We then have $\frac{\mathrm{d}\mathcal{E}_\phi}{\mathrm{d}t}\leq 0$ since $\sigma\geq 0$.

Therefore, a necessary condition ensuring that the time-discrete energy $\mathcal{E}_\phi^l$ is non-increasing can be derived from eq. (3.13) that is

$$\sup_{x\in\Omega_\phi}\left|1-\Delta t^l\frac{\sigma^l\mathbb{I}_\Omega}{\varepsilon_r\varepsilon_0}\right|\leq 1.$$

As σ^l, ε_0, ε_r are positive and $\varepsilon_r=1$ on Ω (air), this condition takes the form of a constraint on the timestep,

$$\Delta t^l\leq\frac{2\varepsilon_0}{\sup_{x\in\Omega}\sigma^l}. \tag{3.14}$$

In practice, it has been numerically observed that if this constraint on the timesteps is violated, then the simulation will in fact become unstable.

The quantity on the right-hand side of eq. (3.14) is the so-called dielectric relaxation time. In practical implementation, we use the approximants σ_K^l of $\sigma^l(\boldsymbol{x}_K^c)$ and compute the timesteps as follows,

$$\Delta t^l\leq\mathfrak{C}\frac{2\varepsilon_0}{\max_{K=1,\dots,\mathcal{N}}\sigma_K^l}\equiv\Delta t_\phi^l, \tag{3.15}$$

where $\mathfrak{C}\leq 1$ is a user-defined parameter. In order to evaluate σ_K^l, we compute the drift-diffusion coefficients from the field $\boldsymbol{E}^l(\boldsymbol{x}_K^c)$ (see section 3.3.2) and the gradients at cell centers $\boldsymbol{x}_K^c$ using the dual-cell reconstruction (see section 3.2.2).

3.3.4 Semi-implicit method for the computation of the potential

It is straightforward that if we use an implicit time scheme for the time-discretization of eq. (3.12), for instance the backward Euler method in example 3.4, then the electric field at t^{l+1} satisfies

$$\nabla\cdot((\varepsilon_r\varepsilon_0+\Delta t^l\sigma^{l+1}\mathbb{I}_\Omega)\boldsymbol{E}^{l+1})=\nabla\cdot(\varepsilon_r\varepsilon_0\boldsymbol{E}^l). \tag{3.16}$$

Therefore, the total electrical energy is immediately non-increasing as $\mathcal{E}_\phi^{l+1}\leq\mathcal{E}_\phi^l$ since $\sigma^{l+1}\geq 0$. However, eq. (3.16) is usually not simple to solve because of the presence of the implicit conductivity σ^{l+1}. In order to overcome this technical difficulty, a **semi-implicit approach** [155, 16, 63] was proposed in which the conductivity is taken into account explicitly, i.e. eq. (3.16) is replaced by

$$\nabla\cdot((\varepsilon_r\varepsilon_0+\Delta t^l\sigma^l\mathbb{I}_\Omega)\boldsymbol{E}^{l+1})=\nabla\cdot(\varepsilon_r\varepsilon_0\boldsymbol{E}^l). \tag{3.17}$$

Equation (3.17) allows to compute the field with a timestep as 50 times larger as the dielectric relaxation time [16, 63]. However, the LU factorization is required at each time t^l since the stiffness matrix of eq. (3.17) changes each time. Therefore, the trade-off between larger timesteps and higher complexity of the numerical method needs to be carefully balanced so that the use of the semi-implicit scheme will not be counterproductive.

Part II

Contributions to Mathematical and Numerical Modeling of Atmospheric Pressure Discharge

Chapter 4

High-order Scharfetter-Gummel Schemes: Derivation, Properties and Extension on Two-dimensional Grids*

4.0 Aperçu . . . 74
4.1 Overview . . . 76
4.2 The Scharfetter-Gummel scheme . . . 78
4.3 Construction of high-order Scharfetter-Gummel schemes . . . 81
4.4 Properties of the point-wise SGCC-p flux . . . 87
4.4.1 Drift and diffusion limits . . . 87
4.4.2 Flux consistency . . . 90
4.5 Extension of the SGCC-p flux scheme on two-dimensional grids . . . 92
4.5.1 Flux integration on cell edges . . . 92
4.5.2 Edge-normal SGCC-p flux . . . 93
4.6 High-order reconstruction of particle density . . . 94
4.6.1 Cell neighborhoods . . . 95
4.6.2 Ghost cells and boundary conditions . . . 95
4.6.3 A reconstruction technique . . . 96
4.6.4 Approximation error . . . 99
4.6.5 The discrete SGCC-p flux . . . 100
4.7 Choices of slope limiters . . . 100
4.7.1 For piece-wise linear reconstruction on one-dimensional grids . . . 100
4.7.2 For piece-wise parabolic reconstruction on one-dimensional grids . . . 103
4.7.3 A limiter on two-dimensional grids . . . 104
4.8 Numerical verification . . . 105
4.8.1 On one-dimensional grids . . . 105
4.8.2 On two-dimensional cartesian grids . . . 109
4.9 Closing remarks . . . 111
4.10 Remarques finales . . . 113

*parts of this chapter, in altered form, has been published in [144]: *N Tuan Dung, C Besse, and F Rogier. "High-order Scharfetter-Gummel-based schemes and applications to gas discharge modeling". In: Journal of Computational Physics 461 (2022), p. 111196.*

4.0 Aperçu

Dans ce chapitre, nous présentons une généralisation d'ordre élevé du schéma de Scharfetter-Gummel (SG), proposé dans [130], qui est utilisé pour résoudre les équations de dérive-diffusion qui apparaissent dans le modèle de décharge (2.13). Le schéma de Scharfetter-Gummel est largement utilisé dans la simulation des décharges dans l'air. De plus, il est peut-être plus connu dans le domaine des semi-conducteurs, d'où l'existence de nombreuses études et variations du schéma SG, par exemple [101, 31, 30, 32, 52, 33, 13, 14, 122] . Cependant, à notre connaissance, aucun de ces travaux n'a conduit à la dérivation d'un schéma de type SG qui soit plus précis que le second ordre.

Les équations de dérive-diffusion sont également largement présentes dans de nombreux autres domaines scientifiques, tels que la mécanique des fluides. Pour la discrétisation de ces équations dans l'espace, il existe un grand nombre d'études disponibles qui entrent dans la catégorie des méthodes de volumes finis. Sans vouloir être exhaustif, nous n'en mentionnerons que quelques-unes. La méthode la plus standard a été proposée par Eymard et al. [53], qui ont utilisé le schéma décentré du premier ordre pour discrétiser le flux de dérive et le schéma de différence-centrale pour le flux de diffusion. Une technique plus intéressante qui produit des solutions numériques de meilleure qualité est celle des schémas MUSCL qui ont été proposés par B. van Leer dans ses articles pionniers [148, 149, 150, 151, 152], qui s'appuient sur la reconstruction du gradient de densité pour obtenir une plus grande précision numérique. Ces schémas numériques ont été principalement conçus pour les problèmes hyperboliques (sans diffusion) et dédiés aux problèmes de capture des chocs. Il y a en fait une très grande recherche sur les méthodes d'ordre élevé pour approcher le flux de dérive. Sweby [137], Harten [70] et d'autres auteurs ont ensuite étendu les travaux de B. van Leer et étudié systématiquement la classe des méthodes de variation totale décroissante (TVD) qui utilisent des limiteurs de flux ou des limiteurs de pente pour supprimer les oscillations dans les solutions numériques qui pourraient déstabiliser la simulation. Colella et Woodward [37] ont mis en œuvre le limiteur PPM pour la reconstruction quadratique. Plus tard, Harten et al. [71, 72] ont introduit une technique pour construire des schémas d'ordre arbitrairement élevé qui sont exempts d'oscillations nonphysiques. Leurs travaux ont ouvert la voie aux recherches toujours en cours sur les méthodes d'ENO/WENO et d'autres schémas de ce type (voir [134, 143] et les références qui y figurent). D'autres développements sur les processus de reconstruction d'ordre élevé de la densité peuvent être trouvés, par exemple, dans [9] pour la reconstruction par moindres carrés, ou [69] pour la reconstruction itérative et adaptée pour faire du MPI. Ces schémas d'ordre élevé souffrent d'oscillations parasites, en particulier en présence de discontinuités ou de couches de gradient fort dans la densité. A notre connaissance, il n'existe pas aujourd'hui de technique unifiée de limitation pour les schémas d'ordre arbitrairement élevé de la même manière que la technique MUSCL pour les schémas du second ordre. Quelques tentatives ont été faites, comme dans [162] où les auteurs ont appliqué successivement les limiteurs TVD sur les approximants des dérivées de la densité. Les techniques mentionnées jusqu'à présent sont généralement appelées **limitation a priori**, car les limiteurs sur la densité reconstruite sont appliqués avant d'avancer dans le temps. Les autres sont les techniques de **limitation a posteriori**, qui ont été popularisées par la méthode MOOD [36], qui permet d'avancer dans le temps avec une reconstruction d'ordre élevé de la densité, puis de recalculer

la solution avec une reconstruction d'ordre inférieur sur les cellules où présentent des oscillations parasites. D'autre part, le développement de méthodes de haute précision pour la discrétisation du flux de diffusion est moins souligné, à notre connaissance. Nous renvoyons à [115, 153] pour la discussion sur ce sujet.

Une classe particulière de méthodes numériques spécifiques aux équations de dérive-diffusion sont les **schémas exponentiels**. Allen & Southwell [56] et A. M. Il'in [75] ont introduit un schéma de différences finies qui résout exactement les équations de diffusion-dérive stationnaires avec des coefficients constants, ce qui signifie que la solution numérique se coïncide avec la vraie solution de l'équation différentielle sur les points de maillage. Scharfetter & Gummel [130] ont introduit la même année qu'A. M. Il'in un schéma de différences finies basé sur une intégration locale du flux de particules entre les centres des cellules du maillage. Bien que ces différentes approches aient abouti à l'obtention du même schéma, les philosophies qui sous-tendent la dérivation de ce schéma sont différentes les unes des autres. Comme cette dernière approche facilite la construction d'un schéma de volumes finis puisqu'elle raisonne sur le flux de densité, nous préférons le nom de Scharfetter-Gummel (SG) pour désigner cette méthode numérique. Le schéma SG est uniformément convergent au premier ordre pour les équations de dérive-diffusion-réaction stationnaires en général, ce qui signifie que sa constante d'erreur ne dépend pas du gradient de la solution, un résultat qui a été prouvé dans [75] par une méthode à deux maillages et dans [127] avec une étude asymptotique de la solution de l'équation différentielle. E. C. Gartland [57] a proposé en outre une famille de schémas uniformément convergents d'ordre arbitrairement élevé - les schémas HODIE - mais qui nécessitent des points d'évaluation auxiliaires à l'intérieur d'un stencil compact. Pour une bonne lecture sur les schémas exponentiels, nous renvoyons au livre [127]. Plus récemment, ten Thije Boonkkamp & Anthonissen [142] ont introduit une amélioration du schéma SG qui a été dérivé en prenant en compte en plus le terme source de l'équation de dérive-diffusion, pas seulement l'équation de flux comme dans la dérivation du schéma SG. Le flux résultant est une combinaison de deux composante : une composante "homogène" qui est similaire au flux SG et une composante "inhomogène" qui intègre le terme source. Liu et al. [96] ont ensuite prouvé que ce schéma est uniformément convergent au second ordre.

Comme nous le pouvons observer dans [19] et dans nos simulations dans le chapitre 5, le schéma SG fournit généralement des solutions de simulation de mauvaise qualité sur des maillages grossiers, car sa précision dégénère au premier ordre si le régime de dérive est prédominant. Par conséquent, les schémas d'ordre élevé tels que les méthodes MUSCL sont fréquemment utilisés à la place dans la simulation de la décharge de gaz, par exemple dans [106, 46, 48, 140], pour la discrétisation du flux de dérive. Au contraire, l'un des intérêts du SG est qu'il discrétise en même temps les flux de dérive et de diffusion, de sorte que s'il existe une généralisation d'ordre élevé du schéma SG, il pourrait s'agir d'un moyen simple de résoudre les équations de dérive-diffusion.

La structure de ce chapitre est la suivante : la section 4.2 présente brièvement le schéma de Scharfetter-Gummel (le schéma SG "standard"), la section 4.3 décrit la construction, sur des maillages unidimensionnelles, de nouvelles méthodes d'ordre élevé, appelées les schémas de **Scharfetter-Gummel avec correction du current** (SGCC), qui sont en fait une généralisation d'ordre élevé

du schéma SG standard. Ensuite, dans la section 4.4, nous étudions certaines de leurs propriétés, à savoir leur comportement dans les limites de dérive et de diffusion ainsi que leur consistence de flux. Dans les sections 4.5 à 4.7, nous discutons de l'extension des nouveaux schémas sur des maillages bidimensionnelles ainsi que des techniques de limitation pour s'assurer que les solutions numériques sont exemptes d'oscillations parasites. Enfin, queslques cas de test simples sont proposés dans la section 4.8 pour vérifier l'ordre de convergence numérique des schémas SGCC.

4.1 Overview

In this chapter, we introduce a high-order generalization of the Scharfetter-Gummel (SG) flux scheme, proposed in [130], that is employed to solve the drift-diffusion equations that appear in the discharge model (2.13). The Scharfetter-Gummel scheme has been widely used in computational gas discharge physics. However, it is perhaps more well known in the semiconductor field, hence there exists numerous studies as well as variations of the SG scheme in this domain, for example [101, 31, 30, 32, 52, 33, 13, 14, 122] . However, to our knowledge, none of the works has lead to the derivation of a SG-like scheme that is more accurate than second-order.

Drift-diffusion equations are also widely present in many other science fields, such as fluid mechanics. For the space discretization of these equations, there are a huge array of available studies that fall into the category of finite volume methods. Without trying to be complete, we only mention a few of them. The most standard method was proposed by Eymard et al. [53], who used the first-order upwind scheme to discretize the drift flux and the central-difference scheme for the diffusion flux. A more interesting technique that produces better-quality numerical solutions is the well known MUSCL schemes that were put forth by B. van Leer in his pioneer series of papers [148, 149, 150, 151, 152], which relies on the reconstruction of the density gradient to gain more numerical precision. These numerical schemes were mainly designed for hyperbolic problems (no diffusion) and dedicated to shock capturing problems. There is actually a very large research on high-order methods to approximate the drift flux. Sweby [137], Harten [70] and other authors later extended the work of B. van Leer and systematically studied the class of total-variation diminishing (TVD) methods that feature the use of flux limiters or slope limiters to suppress nonphysical oscillations in the numerical solutions that could destabilize the simulation. Colella and Woodward [37] developed the so-called PPM limiter for quadratic reconstruction. Later on, Harten et al. [71, 72] introduced a technique to construct arbitrarily high-order schemes that are free of spurious oscillations. Their work paved the way for the still on-going research on the ENO/WENO methods and other related schemes (see [134, 143] and the references therein). Further developments on high-order density reconstruction techniques can be found in, for example, [9] for least-squares reconstruction, or [69] for iterative and MPI-friendly reconstruction. Such high-order schemes suffer spurious oscillations, especially in the presence of discontinuities or large-gradient layers in the density. To our knowledge, there is until now no unified framework of limiting technique for arbitrarily high-order schemes in the same fashion as the MUSCL technique for second-order schemes. Some attempts have been made, such as in [162] where the authors applied successively the TVD limiters on the approximants of the density derivatives. The techniques that have been mentioned so far are generally dubbed as

a priori limiting, since the limiters on the reconstructed density are applied before advancing in time. Their counterparts are the **a posteriori limiting** techniques, which was popularized by the MOOD method [36], that allows the advance in time with a high-order reconstruction of the density and then recomputing the solution with lower-order reconstruction on cells that exhibit spurious oscillations. On the other hand, the development of high-accuracy methods for the discretization of the diffusion flux are less emphasized though, to our knowledge. We refer to [115, 153] for the discussion on this subject.

A particular class of numerical methods specific to drift-diffusion equations are the **exponentially fitted schemes**. Allen & Southwell [56] and A. M. Il'in [75] introduced a finite difference scheme that solves exactly the stationary drift-diffusion equations with constant coefficients, meaning that the numerical solution fits exactly the true solution of the differential equation on grid points. Scharfetter & Gummel [130] introduced in the same year as A. M. Il'in a finite difference scheme which is based on a local integration of the density flux between cell centers. Although these different approaches ended up in obtaining the same scheme, the philosophies behind the derivation of this scheme are different from each other. Since the latter approach somewhat facilitates the construction of a finite volume scheme since it reasons on the particle flux, we prefer the name Scharfetter-Gummel (SG) to refer to this numerical method. The SG scheme is uniformly first-order convergent for general steady drift-diffusion-reaction equations, meaning that its error constant does not depend on the gradient of the solution, a result that was proved in [75] by a two-grid method and in [127] with an asymptotic study of the solution of the differential equation. E. C. Gartland [57] proposed in addition a family of uniformly convergent schemes of arbitrary order - the HODIE schemes - but they require auxiliary evaluation points within a compact stencil. For a good reading on the exponentially fitted schemes, we refer to the book [127]. More recently, ten Thije Boonkkamp & Anthonissen [142] introduced an improvement of the SG scheme that was derived by taking into account in addition the source term of the stationary drift-diffusion equation, not just the flux equation as in the derivation of the SG scheme. The resulting flux is a combination of two parts: a "homogeneous" component which is similar to the SG flux and an "inhomogeneous" part which integrates the source term. Liu et al. [96] later proved that this scheme is uniformly second-order convergent.

As observed in [19] and in our simulations later in chapter 5, the SG scheme usually provides poor-quality simulation solutions on coarse grids since its accuracy degenerates to first order if the drift regime is predominant. Therefore, high-order schemes such as the MUSCL methods are frequently used instead in simulation of gas discharge, for example in [106, 46, 48, 140]. On the contrary, an interest of SG is that it discretizes at the same time the drift and diffusion fluxes, so if a high-order generalization of the SG scheme exists, it could be a simple way to solve drift-diffusion equations.

The structure of this chapter is arranged as follows: section 4.2 briefly introduces the Scharfetter-Gummel flux scheme (the "standard" SG scheme), section 4.3 describes the derivation, on one-dimensional grids, of new high-order methods that are coined **Scharfetter-Gummel schemes with correction of current** (SGCC), which are in fact a high-order generalization of the SG scheme. Then in section 4.4, we study some of their properties, namely their behavior in the drift and diffusion limits

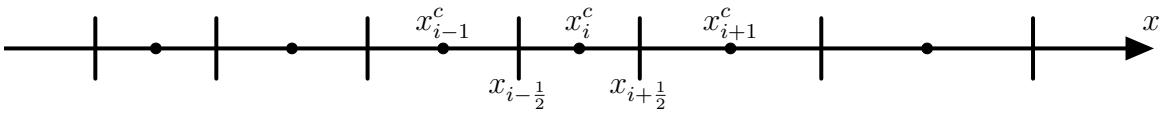


Figure 4.1: A one-dimensional grid with cell centers illustrated by big dots and cell interfaces illustrated by vertical strokes

as well as their flux consistency. In sections 4.5 to 4.7, we discuss the extension of the novel schemes on two-dimensional grids as well as limiting techniques to ensure that the numerical solutions are oscillation-free. Finally, some academic tests are proposed in section 4.8 to verify the numerical convergence order of the SGCC schemes.

4.2 The Scharfetter-Gummel scheme

Let us first consider a one-dimensional domain Ω, $T > 0$ and a drift-diffusion equation defined on Ω with, for simplicity, **constant** (both in space and time) drift velocity u and diffusion coefficient D,

$$\begin{cases} \partial_t n(t,x) + \partial_x f(t,x) = 0, \quad (t,x) \in (0,T) \times \Omega, \\ f(t,x) = un(t,x) - D\partial_x n(t,x). \end{cases} \tag{4.1}$$

The domain Ω is partitioned into $\mathcal{N}$ non-overlapping adjacent intervals, or cells, whose centers are enumerated increasingly from left to right with integers, while the cell interfaces are enumerated with half-integers (see fig. 4.1). The position of cell centers and interfaces are denoted as resp. x_i^c for $i = 1, \dots, \mathcal{N}$ and $x_{i+\frac{1}{2}}$ for $i = 0, \dots, \mathcal{N}$. In the finite volume framework, we search for an approximation of the particle flux density $f(t,x)$ at each interface $x_{i+\frac{1}{2}}$.

The idea of Scharfetter & Gummel [130] is to replace the flux $f(t^l, x)$, at each time t^l and for $x \in (x_i^c, x_{i+1}^c)$, with a constant $f_{i+\frac{1}{2}}^{l|0}$, $i = 1, \dots, \mathcal{N}-1$, and then to solve the second equation of eq. (4.1), called **flux equation**, with $f_{i+\frac{1}{2}}^{l|0}$ on the left-hand side, while considering two constants n_i^l and n_{i+1}^l as boundary values for n at x_i^c and x_{i+1}^c. For numerical implementation, it is practical to choose the cell-averaged values $\overline{n}_i^l$, $\overline{n}_{i+1}^l$ for n_i^l, n_{i+1}^l. We also drop the time dependence of n as well as the upper index l without ambiguity since the problem invokes only x. In summary, we want to solve the following ordinary differential problem for $f_{i+\frac{1}{2}}^{|0}$ on each interval (x_i^c, x_{i+1}^c),

$$\begin{cases} un(x) - Dn'(x) = f_{i+\frac{1}{2}}^{|0}, \quad x \in (x_i^c, x_{i+1}^c), \\ n(x_i^c) = n_i, \quad n(x_{i+1}^c) = n_{i+1}. \end{cases} \tag{4.2}$$

The first equation of (4.2) is equivalent to $f_{i+\frac{1}{2}}^{|0} = -D \exp\left(\frac{u}{D}x\right)\left(n(x)\exp\left(-\frac{u}{D}x\right)\right)'$, so we have

$$\frac{f_{i+\frac{1}{2}}^{|0}}{u}\left(\exp\left(-\frac{u}{D}x_i^c\right) - \exp\left(-\frac{u}{D}x_{i+1}^c\right)\right) = n(x_i^c)\exp\left(-\frac{u}{D}x_i^c\right) - n(x_{i+1}^c)\exp\left(-\frac{u}{D}x_{i+1}^c\right).$$

Multiplying this equation with $\exp\left(\frac{u}{D}x_{i+\frac{1}{2}}\right)$ and defining $\kappa_{i+\frac{1}{2}} \equiv \frac{x^c_{i+1} - x_{i+\frac{1}{2}}}{\Delta x_{i+\frac{1}{2}}}$, $\chi_{i+\frac{1}{2}} \equiv \frac{x_{i+\frac{1}{2}} - x^c_i}{\Delta x_{i+\frac{1}{2}}}$ with $\Delta x_{i+\frac{1}{2}} \equiv x^c_{i+1} - x^c_i$ yield

$$\frac{f^{|0}_{i+\frac{1}{2}}}{u}\left(\exp\left(\frac{u\Delta x_{i+\frac{1}{2}}}{D}\chi_{i+\frac{1}{2}}\right) - \exp\left(-\frac{u\Delta x_{i+\frac{1}{2}}}{D}\kappa_{i+\frac{1}{2}}\right)\right) = n_i \exp\left(\frac{u\Delta x_{i+\frac{1}{2}}}{D}\chi_{i+\frac{1}{2}}\right) - n_{i+1}\exp\left(-\frac{u\Delta x_{i+\frac{1}{2}}}{D}\kappa_{i+\frac{1}{2}}\right).$$

Using $\kappa_{i+\frac{1}{2}} + \chi_{i+\frac{1}{2}} = 1$ and defining the **numerical Péclet number** as $\mathfrak{p}_{i+\frac{1}{2}} \equiv -\frac{u\Delta x_{i+\frac{1}{2}}}{D}$, we have

$$f^{|0}_{i+\frac{1}{2}} = u\left(n_i \frac{1}{1 - \mathrm{e}^{\mathfrak{p}_{i+\frac{1}{2}}}} - n_{i+1}\frac{1}{\mathrm{e}^{-\mathfrak{p}_{i+\frac{1}{2}}} - 1}\right).$$

Definition 4.1 (SG scheme). *We define the **Bernoulli function** as* $\mathcal{B}(\mathfrak{p}) \equiv \frac{\mathfrak{p}}{\mathrm{e}^{\mathfrak{p}} - 1}$ *and so the approximate flux reads as*

$$f^{|0}_{i+\frac{1}{2}} = \frac{D}{\Delta x_{i+\frac{1}{2}}}\left(\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}})n_i - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}})n_{i+1}\right). \tag{4.3}$$

The SG flux can also be recast into a different form. Firstly, we have

$$\begin{aligned} f^{|0}_{i+\frac{1}{2}} &= \frac{D}{\Delta x_{i+\frac{1}{2}}}\left[\left(\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}}) - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}})\right)n_i - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}})(n_{i+1} - n_i)\right] \\ &= \frac{D}{\Delta x_{i+\frac{1}{2}}}\left[\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}})(n_i - n_{i+1}) - \left(\mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}}) - \mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}})\right)n_{i+1}\right]. \end{aligned}$$

As $\mathcal{B}(-\mathfrak{p}) - \mathcal{B}(\mathfrak{p}) = \mathfrak{p}$ and $\mathcal{B}(-\mathfrak{p}) + \mathcal{B}(\mathfrak{p}) = \mathfrak{p}\coth\left(\frac{\mathfrak{p}}{2}\right)$, summing the two lines on the right-hand side and dividing by 2 yield

$$f^{|0}_{i+\frac{1}{2}} = u\frac{n_i + n_{i+1}}{2} - D\varphi(\mathfrak{p}_{i+\frac{1}{2}})\frac{n_{i+1} - n_i}{\Delta x_{i+\frac{1}{2}}},$$

with $\varphi(\mathfrak{p}) \equiv \frac{\mathfrak{p}}{2}\coth\left(\frac{\mathfrak{p}}{2}\right)$. Therefore, the SG flux can be interpreted as a central-difference flux with a modified diffusion coefficient $D\varphi(\mathfrak{p}_{i+\frac{1}{2}})$ that depends on the grid size. Moreover, since $\varphi(\mathfrak{p}) \to 1$ as $\mathfrak{p} \to 0$ and $\mathfrak{p}_{i+\frac{1}{2}} \to 0$ as $\Delta x_{i+\frac{1}{2}} \to 0$, it is easy to see that the SG flux $f^{|0}_{i+\frac{1}{2}}$ is a second-order approximation of the real flux $f(x_{i+\frac{1}{2}})$ on uniform grids.

An interesting property of the Bernoulli function is that $\mathcal{B}(\mathfrak{p}) \sim -\mathfrak{p}$ as $\mathfrak{p} \to -\infty$ and $\mathcal{B}(\mathfrak{p}) \to 0$ as $\mathfrak{p} \to +\infty$. Therefore, in the diffusion limit $\mathfrak{p} \to 0$, the SG scheme behaves as the central-difference scheme in a pure diffusion problem, which is second-order on uniform grids,

$$f^{|0}_{i+\frac{1}{2}} \to -D\frac{n_{i+1} - n_i}{\Delta x_{i+\frac{1}{2}}}.$$

On the contrary, in the drift limit $\mathfrak{p} \to -\infty$ (assume that $u > 0$), the SG flux behaves as the first-order upwind scheme in a pure drift problem,

$$f^{|0}_{i+\frac{1}{2}} \to u n_i.$$

As a consequence, the numerical precision of the SG flux deteriorates quickly on uniform grids as the Péclet number increases (in absolute value). This is an obstacle for simulation of drift-dominated problems because in these cases, extremely refined grids are required to obtain high-accuracy numerical results, which are costly in terms of CPU memory as well as CPU time.

Remark 4.1. *A distinction should be made when we choose $n(x_i^c)$ and $n(x_{i+1}^c)$, or $\overline{n}_i$ and $\overline{n}_{i+1}$ for the boundary values n_i and n_{i+1}. In the former case, the approximate flux, still denoted as $f^{|0}_{i+\frac{1}{2}}$, is a function of point-wise values of the real density n; for this reason, we call it the **point-wise flux**. In the latter case, we denote the approximate flux as $\overline{f}^{|0}_{i+\frac{1}{2}}$ to emphasize its dependence on the numerical cell-averaged density; it is alternatively known as the **discrete flux**.*

When solving the first equation of (4.1) with the SG scheme and an explicit time integration method, e.g. the forward Euler scheme, it is necessary to impose a CFL condition on the numerical timestep $\Delta t^l = t^{l+1} - t^l$ in the same fashion as (3.10). The discrete equation reads as

$$\frac{\overline{n}_i^{l+1} - \overline{n}_i^l}{\Delta t^l} + \frac{\overline{f}^{l|0}_{i+\frac{1}{2}} - \overline{f}^{l|0}_{i-\frac{1}{2}}}{\Delta x_i} = 0,$$

with $\Delta x_i \equiv x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$. In other words,

$$\begin{aligned}\overline{n}_i^{l+1} = &\left(1 - \frac{\Delta t^l}{\Delta x_i}\frac{D\varphi(\mathfrak{p}_{i+\frac{1}{2}})}{\Delta x_{i+\frac{1}{2}}} - \frac{\Delta t^l}{\Delta x_i}\frac{D\varphi(\mathfrak{p}_{i-\frac{1}{2}})}{\Delta x_{i-\frac{1}{2}}}\right)\overline{n}_i^l \\ &+ \frac{\Delta t^l}{\Delta x_i}\left(\frac{D\varphi(\mathfrak{p}_{i+\frac{1}{2}})}{\Delta x_{i+\frac{1}{2}}} - \frac{u}{2}\right)\overline{n}_{i+1}^l + \frac{\Delta t^l}{\Delta x_i}\left(\frac{D\varphi(\mathfrak{p}_{i-\frac{1}{2}})}{\Delta x_{i-\frac{1}{2}}} - \frac{u}{2}\right)\overline{n}_{i-1}^l.\end{aligned}$$

Since $\dfrac{D\varphi(\mathfrak{p}_{i+\frac{1}{2}})}{\Delta x_{i+\frac{1}{2}}} - \dfrac{u}{2} = \dfrac{D}{\Delta x_{i+\frac{1}{2}}}\left(\varphi(\mathfrak{p}_{i+\frac{1}{2}}) + \dfrac{\mathfrak{p}_{i+\frac{1}{2}}}{2}\right)$ and $\varphi(\mathfrak{p}) + \dfrac{\mathfrak{p}}{2} > 0$ for any $\mathfrak{p}$, the condition for which the scheme is positive is

$$\Delta t^l \le \left(\frac{D\varphi(\mathfrak{p}_{i+\frac{1}{2}})}{\Delta x_{i+\frac{1}{2}}} + \frac{D\varphi(\mathfrak{p}_{i-\frac{1}{2}})}{\Delta x_{i-\frac{1}{2}}}\right)^{-1}\Delta x_i, \quad \forall i = 1, \dots, \mathcal{N}. \tag{4.4}$$

Remark 4.2. *In its simplest form with homogeneous grid size, i.e. $\Delta x_{i+\frac{1}{2}} = \Delta x$, (4.4) reads as $\Delta t^l \le \dfrac{\Delta x}{u}\tanh\left(\dfrac{u\Delta x}{2D}\right)$. In the diffusion limit $u\Delta x \ll D$, $\tanh\left(\dfrac{u\Delta x}{2D}\right) \sim \dfrac{u\Delta x}{2D}$ so the CFL condition $\Delta t^l \le \dfrac{\Delta x^2}{2D}$ reflects the domination of the diffusion term. On the contrary, in the drift limit $u\Delta x \gg D$, $\tanh\left(\dfrac{u\Delta x}{2D}\right) \approx 1$ so the CFL condition $\Delta t^l \le \dfrac{\Delta x}{u}$ reflects the domination of the drift term.*

4.3 Construction of high-order Scharfetter-Gummel schemes

The idea is to replace the flux $f(t^l, x)$, at each time t^l and for $x \in (x_i^c, x_{i+1}^c)$, with a **polynomial** $F_{i+\frac{1}{2}}^{l|p}(x)$ of degree p, $i = 1, \ldots, \mathcal{N}-1$, and then to solve the flux equation in eq. (4.1) with $F_{i+\frac{1}{2}}^{l|p}$ on the left-hand side. $F_{i+\frac{1}{2}}^{l|p}$ can be written in a general form as[1]

$$F_{i+\frac{1}{2}}^{|p}(x) = f_{i+\frac{1}{2}}^{[0|p]} + f_{i+\frac{1}{2}}^{[1|p]}\left(x - x_{i+\frac{1}{2}}\right) + \frac{1}{2}f_{i+\frac{1}{2}}^{[2|p]}\left(x - x_{i+\frac{1}{2}}\right)^2 + \cdots + \frac{1}{p!}f_{i+\frac{1}{2}}^{[p|p]}\left(x - x_{i+\frac{1}{2}}\right)^p,$$

where $f_{i+\frac{1}{2}}^{[q|p]}$, for $q = 0, \ldots, p$, are real numbers that we need to solve for. In the following, we omit the upper index p in $f_{i+\frac{1}{2}}^{[q|p]}$ without ambiguity.

Now if we make an assumption that n is a smooth function of x and derive the flux equation in eq. (4.1) q times with respect to x, with $F_{i+\frac{1}{2}}^{|p}$ on the left-hand side, then the result is[2]

$$un^{(q)}(x) - Dn^{(q+1)}(x) = \left(F_{i+\frac{1}{2}}^{|p}(x)\right)^{(q)} = \sum_{m=q}^{p} \frac{1}{(m-q)!} f_{i+\frac{1}{2}}^{[m]}\left(x - x_{i+\frac{1}{2}}\right)^{m-q},$$

for $q = 0, \ldots, p$, where $n^{(q)}$ is the q^{th}-derivative (with respect to x) of n. Following the derivation of the SG scheme of section 4.2, we prescribe the boundary values $n_i^{(q)}$ and $n_{i+1}^{(q)}$ for $n^{(q)}(x)$ at x_i^c and x_{i+1}^c, for each $q = 0, \ldots, p$. Therefore, we end up at solving in total $p+1$ ordinary differential problems on each interval (x_i^c, x_{i+1}^c),

$$\begin{cases} un^{(q)}(x) - Dn^{(q+1)}(x) = \displaystyle\sum_{m=q}^{p} \frac{1}{(m-q)!} f_{i+\frac{1}{2}}^{[m]}\left(x - x_{i+\frac{1}{2}}\right)^{m-q}, \\ n^{(q)}(x_i^c) = n_i^{(q)}, \quad n^{(q)}(x_{i+1}^c) = n_{i+1}^{(q)}. \end{cases} \tag{4.5}$$

Let us begin with $q = p$ in eq. (4.5). This is the simplest problem since the right-hand side of the first equation is a constant, i.e.

$$\begin{cases} un^{(p)}(x) - Dn^{(p+1)}(x) = f_{i+\frac{1}{2}}^{[p]}, \\ n^{(p)}(x_i^c) = n_i^{(p)}, \quad n^{(p)}(x_{i+1}^c) = n_{i+1}^{(p)}. \end{cases}$$

This single case is very much similar to the SG problem (4.2), so the computation of $f_{i+\frac{1}{2}}^{[p]}$ follows strictly the same algebraic manipulations. Thus, the expression of $f_{i+\frac{1}{2}}^{[p]}$ echoes eq. (4.3), which is

$$f_{i+\frac{1}{2}}^{[p]} = \frac{D}{\Delta x_{i+\frac{1}{2}}}\left(\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}})n_i^{(p)} - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}})n_{i+1}^{(p)}\right). \tag{4.6}$$

Let us continue with $q = p-1$. The flux equation of eq. (4.5) in this case reads as

$$-D\exp\left(\frac{u}{D}x\right)\left(n^{(p-1)}(x)\exp\left(-\frac{u}{D}x\right)\right)' = f_{i+\frac{1}{2}}^{[p-1]} + f_{i+\frac{1}{2}}^{[p]}\left(x - x_{i+\frac{1}{2}}\right).$$

[1] omitting the upper index l without ambiguity

[2] dropping the time dependence of n for the sake of simplicity

Multiplying this equation with $\frac{1}{D}\exp\left(-\frac{u}{D}\left(x-x_{i+\frac{1}{2}}\right)\right)$ and integrating on (x_i^c, x_{i+1}^c) yield

$$\begin{aligned}
&n_i^{(p-1)}\exp\left(-\mathfrak{p}_{i+\frac{1}{2}}\chi_{i+\frac{1}{2}}\right) - n_{i+1}^{(p-1)}\exp\left(\mathfrak{p}_{i+\frac{1}{2}}\kappa_{i+\frac{1}{2}}\right)\\
&= \frac{f_{i+\frac{1}{2}}^{[p-1]}}{D}\int_{x_i^c}^{x_{i+1}^c}\exp\left(-\frac{u}{D}\left(x-x_{i+\frac{1}{2}}\right)\right)\mathrm{d}x + \frac{f_{i+\frac{1}{2}}^{[p]}}{D}\int_{x_i^c}^{x_{i+1}^c}\exp\left(-\frac{u}{D}\left(x-x_{i+\frac{1}{2}}\right)\right)\left(x-x_{i+\frac{1}{2}}\right)\mathrm{d}x.
\end{aligned}$$

Using $\int_{x_i^c}^{x_{i+1}^c}\exp\left(-\frac{u}{D}\left(x-x_{i+\frac{1}{2}}\right)\right)\mathrm{d}x = \frac{D}{u}\left(\exp\left(-\mathfrak{p}_{i+\frac{1}{2}}\chi_{i+\frac{1}{2}}\right) - \exp\left(\mathfrak{p}_{i+\frac{1}{2}}\kappa_{i+\frac{1}{2}}\right)\right)$ and making the variable change $\xi = \frac{x-x_{i+\frac{1}{2}}}{\Delta x_{i+\frac{1}{2}}}$, we have

$$f_{i+\frac{1}{2}}^{[p-1]} = \frac{D}{\Delta x_{i+\frac{1}{2}}}\left(\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}})n_i^{(p-1)} - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}})n_{i+1}^{(p-1)}\right) - \psi_{i+\frac{1}{2}}^{[1]}\Delta x_{i+\frac{1}{2}}f_{i+\frac{1}{2}}^{[p]}, \tag{4.7}$$

where $\psi_{i+\frac{1}{2}}^{[1]} \equiv \left(\int_{-\chi_{i+\frac{1}{2}}}^{\kappa_{i+\frac{1}{2}}}\exp\left(\mathfrak{p}_{i+\frac{1}{2}}\xi\right)\mathrm{d}\xi\right)^{-1}\int_{-\chi_{i+\frac{1}{2}}}^{\kappa_{i+\frac{1}{2}}}\exp\left(\mathfrak{p}_{i+\frac{1}{2}}\xi\right)\xi\mathrm{d}\xi.$

Repeating this process for $q = p-2, \dots, 0$, we obtain a recurrence formula of $f_{i+\frac{1}{2}}^{[q]}$ as follows,

$$f_{i+\frac{1}{2}}^{[q]} = \frac{D}{\Delta x_{i+\frac{1}{2}}}\left(\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}})n_i^{(q)} - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}})n_{i+1}^{(q)}\right) - \sum_{m=q+1}^{p}\frac{\psi_{i+\frac{1}{2}}^{[m-q]}}{(m-q)!}\left(\Delta x_{i+\frac{1}{2}}\right)^{m-q}f_{i+\frac{1}{2}}^{[m]}, \tag{4.8}$$

for $q = p-1, \dots, 0$, while the case $q = p$ was treated in eq. (4.6). Here the "ψ-factors" $\psi_{i+\frac{1}{2}}^{[q]}$ are defined as

$$\psi_{i+\frac{1}{2}}^{[q]} = \left(\int_{-\chi_{i+\frac{1}{2}}}^{\kappa_{i+\frac{1}{2}}}\exp\left(\mathfrak{p}_{i+\frac{1}{2}}\xi\right)\mathrm{d}\xi\right)^{-1}\int_{-\chi_{i+\frac{1}{2}}}^{\kappa_{i+\frac{1}{2}}}\exp\left(\mathfrak{p}_{i+\frac{1}{2}}\xi\right)\xi^q\mathrm{d}\xi. \tag{4.9}$$

By integration by parts, we can also derive a recurrence formula for them which reads as

$$\begin{cases}
\psi_{i+\frac{1}{2}}^{[0]} = 1,\\
\psi_{i+\frac{1}{2}}^{[q]} = \dfrac{\mathrm{e}^{\mathfrak{p}_{i+\frac{1}{2}}}\kappa_{i+\frac{1}{2}}^q - \left(-\chi_{i+\frac{1}{2}}\right)^q}{\mathrm{e}^{\mathfrak{p}_{i+\frac{1}{2}}} - 1} - \dfrac{q}{\mathfrak{p}_{i+\frac{1}{2}}}\psi_{i+\frac{1}{2}}^{[q-1]}, \quad q = 1, \dots, p.
\end{cases} \tag{4.10}$$

Remark 4.3. *In practical implementation, we need to take the Taylor development of $\psi^{[q]}$ (omitting the subscript $i+\frac{1}{2}$ here for simplicity) to avoid division by zero if $\mathfrak{p}$ is close to zero, i.e. $|\mathfrak{p}| < \mathfrak{p}_{\min}$ where $\mathfrak{p}_{\min} > 0$ is a small parameter. For example, we have*

$$\begin{aligned}
\psi^{[1]} &= \frac{1}{\mathrm{e}^{\mathfrak{p}}-1} - \frac{1}{\mathfrak{p}} + \frac{1}{2} + \frac{\kappa-\chi}{2},\\
\psi^{[1]} &\underset{0}{\sim} \frac{\kappa-\chi}{2} + \frac{\mathfrak{p}}{12} - \frac{\mathfrak{p}^3}{720} + \frac{\mathfrak{p}^5}{30240} + \mathcal{O}(\mathfrak{p}^7).
\end{aligned}$$

For the simulations in sections 4.8 and 6.6 and chapters 5 and 7, we set $\mathfrak{p}_{\min} = 0.1$.

Let us get back to eqs. (4.6) to (4.8). From the first two equations we have

$$f_{i+\frac{1}{2}}^{[p-1]} = \frac{D}{\Delta x_{i+\frac{1}{2}}} \left(\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}}) Q_{i+\frac{1}{2},i}^{[p-1]} - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}}) Q_{i+\frac{1}{2},i+1}^{[p-1]} \right), \tag{4.11}$$

with $Q_{i+\frac{1}{2},r}^{[p-1]} \equiv n_r^{(p-1)} - \psi_{i+\frac{1}{2}}^{[1]} \Delta x_{i+\frac{1}{2}} n_r^{(p)}$, $r = i, i+1$. This example and the structure of the recurrence formula (4.8) show that $f_{i+\frac{1}{2}}^{[q]}$ can be written in a more compact expression as laid out in lemma 4.1.

Lemma 4.1. *For $q = p, \dots, 0$,*

$$f_{i+\frac{1}{2}}^{[q]} = \frac{D}{\Delta x_{i+\frac{1}{2}}} \left(\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}}) Q_{i+\frac{1}{2},i}^{[q]} - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}}) Q_{i+\frac{1}{2},i+1}^{[q]} \right), \tag{4.12}$$

where $Q_{i+\frac{1}{2},r}^{[q]} \equiv \sum_{m=q}^{p} \Theta_{i+\frac{1}{2}}^{[m|q]} n_r^{(m)}$ for $r = i, i+1$ and $\Theta_{i+\frac{1}{2}}^{[m|q]}$ satisfies the following relation[3] for fixed q,

$$\begin{cases} \Theta_{i+\frac{1}{2}}^{[q|q]} = 1, \\ \Theta_{i+\frac{1}{2}}^{[m|q]} = -\sum_{k=q+1}^{m} \dfrac{\psi_{i+\frac{1}{2}}^{[k-q]}}{(k-q)!} \left(\Delta x_{i+\frac{1}{2}} \right)^{k-q} \Theta_{i+\frac{1}{2}}^{[m|k]}, \quad m > q. \end{cases} \tag{4.13}$$

Moreover, $\Theta_{i+\frac{1}{2}}^{[m|q]} = \Theta_{i+\frac{1}{2}}^{[m-q|0]}$ for $m > q$.

Remark 4.4. *$Q_{i+\frac{1}{2},r}^{[q]}$ is a simplified notation; the full notations should be $Q_{i+\frac{1}{2},r}^{[q|p]}$. On the contrary, we shall see in lemma 4.2 that $\Theta_{i+\frac{1}{2}}^{[m|q]}$ does not depend on p.*

Proof. Equations (4.12) and (4.13) hold true for $q = p, p-1$ as shown in eqs. (4.6) and (4.11). Assume that they also hold true for a certain $q+1$ with $q < p$, then from eq. (4.8) we have

$$f_{i+\frac{1}{2}}^{[q]} = \frac{D}{\Delta x_{i+\frac{1}{2}}} \left(\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}}) \tilde{Q}_{i+\frac{1}{2},i}^{[q]} - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}}) \tilde{Q}_{i+\frac{1}{2},i+1}^{[q]} \right),$$

where we have defined $\tilde{Q}_{i+\frac{1}{2},r}^{[q]} \equiv n_r^{(q)} - \sum_{k=q+1}^{p} \dfrac{\psi_{i+\frac{1}{2}}^{[k-q]}}{(k-q)!} \left(\Delta x_{i+\frac{1}{2}} \right)^{k-q} Q_{i+\frac{1}{2},r}^{[k]}$ for $r = i, i+1$. Using $Q_{i+\frac{1}{2},r}^{[k]} = \sum_{m=k}^{p} \Theta_{i+\frac{1}{2}}^{[m|k]} n_r^{(m)}$, for $k = q+1, \dots, p$, we can further write

$$\begin{aligned} \tilde{Q}_{i+\frac{1}{2},r}^{[q]} &= n_r^{(q)} - \sum_{k=q+1}^{p} \sum_{m=k}^{p} \frac{\psi_{i+\frac{1}{2}}^{[k-q]}}{(k-q)!} \left(\Delta x_{i+\frac{1}{2}} \right)^{k-q} \Theta_{i+\frac{1}{2}}^{[m|k]} n_r^{(m)} \\ &= n_r^{(q)} - \sum_{m=q+1}^{p} \left(\sum_{k=q+1}^{m} \frac{\psi_{i+\frac{1}{2}}^{[k-q]}}{(k-q)!} \left(\Delta x_{i+\frac{1}{2}} \right)^{k-q} \Theta_{i+\frac{1}{2}}^{[m|k]} \right) n_r^{(m)}. \end{aligned}$$

[3] note that $m \geq q$ here

Since the right-hand side is a linear combination of $n_r^{(m)}$, $m = q, \dots, p$, we can set $\tilde{Q}^{[q]}_{i+\frac{1}{2},r} = \sum_{m=q}^{p} \Theta^{[m|q]}_{i+\frac{1}{2}} n_r^{(m)}$ and a simple comparison with the above equation confirms the relation (4.13) as well as the equivalence between $\tilde{Q}^{[q]}_{i+\frac{1}{2},r}$ and $Q^{[q]}_{i+\frac{1}{2},r}$.

Finally, it is evident that $\Theta^{[q|q]}_{i+\frac{1}{2}} = \Theta^{[0|0]}_{i+\frac{1}{2}} (= 1)$ for all $q \geq 0$. Assume that the equality $\Theta^{[q+k|q]}_{i+\frac{1}{2}} = \Theta^{[k|0]}_{i+\frac{1}{2}}$ holds for all $q \geq 0$ and for all k strictly smaller than a certain natural number m. Then from the second equation of (4.13) we have

$$\Theta^{[q+m|q]}_{i+\frac{1}{2}} = -\sum_{k=q+1}^{q+m} \frac{\psi^{[k-q]}_{i+\frac{1}{2}}}{(k-q)!} \left(\Delta x_{i+\frac{1}{2}}\right)^{k-q} \Theta^{[q+m|k]}_{i+\frac{1}{2}} = -\sum_{k=q+1}^{q+m} \frac{\psi^{[k-q]}_{i+\frac{1}{2}}}{(k-q)!} \left(\Delta x_{i+\frac{1}{2}}\right)^{k-q} \Theta^{[m|k-q]}_{i+\frac{1}{2}},$$

since $q + m - k < m$ for $k = q+1, \dots, q+m$. Using the index change $h = k - q$ and eq. (4.13) yield

$$\Theta^{[q+m|q]}_{i+\frac{1}{2}} = -\sum_{h=1}^{m} \frac{\psi^{[h]}_{i+\frac{1}{2}}}{h!} \left(\Delta x_{i+\frac{1}{2}}\right)^{h} \Theta^{[m|h]}_{i+\frac{1}{2}} = \Theta^{[m|0]}_{i+\frac{1}{2}}. \qquad \blacksquare$$

The final point of lemma 4.1 asserts that only $\Theta^{[q|0]}_{i+\frac{1}{2}}$, for $q \geq 0$, need to be computed. This point is addressed in lemma 4.2.

Lemma 4.2. *Let $I_m(q)$ denote the set of strictly positive m-tuples[4] $(x_r)_{r=1,\dots,m}$ whose sum of elements equals q, i.e.*

$$I_m(q) = \left\{ \boldsymbol{X} = (x_r)_{r=1,\dots,m} \middle| x_r \in \mathbb{N}^*, \sum_{r=1}^{m} x_r = q \right\}.$$

Then

$$\begin{cases} \Theta^{[0|0]}_{i+\frac{1}{2}} = 1, \\ \Theta^{[q|0]}_{i+\frac{1}{2}} = \left(\Delta x_{i+\frac{1}{2}}\right)^{q} \sum_{m=1}^{q} (-1)^m \sum_{\boldsymbol{X} \in I_m(q)} \prod_{x_r \in \boldsymbol{X}} \frac{\psi^{[x_r]}_{i+\frac{1}{2}}}{x_r!}, \quad q > 0. \end{cases} \tag{4.14}$$

Proof. Let us first note that $I_1(1) = \{ (1) \}$, $I_1(2) = \{ (2) \}$ and $I_2(2) = \{ (1,1) \}$. Then eq. (4.14) holds for $q = 1, 2$ since from eq. (4.13) we have

$$\Theta^{[1|0]}_{i+\frac{1}{2}} = \Delta x_{i+\frac{1}{2}} \left(-\psi^{[1]}_{i+\frac{1}{2}}\right),$$

$$\Theta^{[2|0]}_{i+\frac{1}{2}} = -\psi^{[1]}_{i+\frac{1}{2}} \Delta x_{i+\frac{1}{2}} \Theta^{[1|0]}_{i+\frac{1}{2}} - \frac{\psi^{[2]}_{i+\frac{1}{2}}}{2} \left(\Delta x_{i+\frac{1}{2}}\right)^2 = \left(\Delta x_{i+\frac{1}{2}}\right)^2 \left(-\frac{\psi^{[2]}_{i+\frac{1}{2}}}{2} + \left(\psi^{[1]}_{i+\frac{1}{2}}\right)^2 \right).$$

[4] an m-tuple is a ordered list of m integer numbers that may contain multiple instances of a number, for example $(1,1,2)$ is a 3-tuple; two m-tuples $(x_{r,1})_{r=1,\dots,m}$ and $(x_{r,2})_{r=1,\dots,m}$ are equivalent if and only if $x_{r,1} = x_{r,2}$ for all $r = 1, \dots, m$

Assume that eq. (4.14) also holds for all $k = 0, \dots, q-1$ with a certain $q \geq 1$. Then from eq. (4.13) we have

$$\Theta_{i+\frac{1}{2}}^{[q|0]} = -\sum_{h=1}^{q} \frac{\psi_{i+\frac{1}{2}}^{[h]}}{h!} \left(\Delta x_{i+\frac{1}{2}}\right)^h \Theta_{i+\frac{1}{2}}^{[q|h]} = -\sum_{k=0}^{q-1} \frac{\psi_{i+\frac{1}{2}}^{[q-k]}}{(q-k)!} \left(\Delta x_{i+\frac{1}{2}}\right)^{q-k} \Theta_{i+\frac{1}{2}}^{[k|0]}, \tag{4.15}$$

where we have made the index change $k = q - h$ and used $\Theta_{i+\frac{1}{2}}^{[q|q-k]} = \Theta_{i+\frac{1}{2}}^{[k|0]}$. Using eq. (4.14) for $k = 0, \dots, q-1$, we have

$$\begin{aligned}
\Theta_{i+\frac{1}{2}}^{[q|0]} &= -\frac{\psi_{i+\frac{1}{2}}^{[q]}}{q!} \left(\Delta x_{i+\frac{1}{2}}\right)^q - \sum_{k=1}^{q-1} \frac{\psi_{i+\frac{1}{2}}^{[q-k]}}{(q-k)!} \left(\Delta x_{i+\frac{1}{2}}\right)^q \sum_{m=1}^{k} (-1)^m \sum_{\boldsymbol{Y} \in I_m(k)} \prod_{x_r \in \boldsymbol{Y}} \frac{\psi_{i+\frac{1}{2}}^{[x_r]}}{x_r!} \\
&= \left(\Delta x_{i+\frac{1}{2}}\right)^q \left(-\frac{\psi_{i+\frac{1}{2}}^{[q]}}{q!} - \sum_{m=1}^{q-1} (-1)^m \sum_{k=m}^{q-1} \frac{\psi_{i+\frac{1}{2}}^{[q-k]}}{(q-k)!} \sum_{\boldsymbol{Y} \in I_m(k)} \prod_{x_r \in \boldsymbol{Y}} \frac{\psi_{i+\frac{1}{2}}^{[x_r]}}{x_r!} \right) \\
&= \left(\Delta x_{i+\frac{1}{2}}\right)^q \left(-\frac{\psi_{i+\frac{1}{2}}^{[q]}}{q!} + \sum_{m=2}^{q} (-1)^m \sum_{k=m-1}^{q-1} \frac{\psi_{i+\frac{1}{2}}^{[q-k]}}{(q-k)!} \sum_{\boldsymbol{Y} \in I_{m-1}(k)} \prod_{x_r \in \boldsymbol{Y}} \frac{\psi_{i+\frac{1}{2}}^{[x_r]}}{x_r!} \right)
\end{aligned} \tag{4.16}$$

Now the question is how to list all the m-tuples $\boldsymbol{X}$ that belong to the set $I_m(q)$, for $m = 2, \dots, q$. There are at least two ways to do so. The first one is a no-thinker: we just do it directly! The alternative way is more sophisticated and we need to do it step by step.

1. Firstly, we fix an integer k and list all the $(m-1)$-tuples $\boldsymbol{Y} = (x_r)_{r=1,\dots,m-1}$ that belong to the set $I_{m-1}(k)$. Since all elements of $\boldsymbol{Y}$ are strictly positive integers, k must be greater or equal to $m-1$.

2. Then we set $x_m = q - k$ and add it to each $(m-1)$-tuple $\boldsymbol{Y}$ of $I_{m-1}(k)$ to build a m-tuple $\boldsymbol{X} = (x_r)_{r=1,\dots,m}$. For such $\boldsymbol{X}$ belongs to $I_m(k)$, x_m must be greater or equal to 1, so k must be lesser or equal to q.

3. We repeat the two previous steps for all $k = m-1, \dots, q-1$.

For two different integers k_1 and k_2, such tuple construction process never yields redundant m-tuples $\boldsymbol{X}_1 = (x_{r,1})_{r=1,\dots,m}$ and $\boldsymbol{X}_2 = (x_{r,2})_{r=1,\dots,m}$ such that $\boldsymbol{Y}_1 = (x_{r,1})_{r=1,\dots,m-1} \in I_{m-1}(k_1)$ and $\boldsymbol{Y}_2 = (x_{r,2})_{r=1,\dots,m-1} \in I_{m-1}(k_2)$, since $x_{m,1} \neq x_{m,2}$.

The equivalence of the two listing-methods allows us to deduce that

$$\sum_{\boldsymbol{X} \in I_m(q)} \prod_{x_r \in \boldsymbol{X}} \frac{\psi_{i+\frac{1}{2}}^{[x_r]}}{x_r!} = \sum_{k=m-1}^{q-1} \frac{\psi_{i+\frac{1}{2}}^{[q-k]}}{(q-k)!} \sum_{\boldsymbol{Y} \in I_{m-1}(k)} \prod_{x_r \in \boldsymbol{Y}} \frac{\psi_{i+\frac{1}{2}}^{[x_r]}}{x_r!}, \tag{4.17}$$

for $m = 2, \dots, q$. The case $m = 1$ is trivial since $I_1(q) = \{ q \}$. Hence, inserting this equality to the last line of eq. (4.16) would yield

$$\begin{aligned}\Theta_{i+\frac{1}{2}}^{[q|0]} &= \left(\Delta x_{i+\frac{1}{2}}\right)^q \left(-\frac{\psi_{i+\frac{1}{2}}^{[q]}}{q!} + \sum_{m=2}^{q} (-1)^m \sum_{\boldsymbol{X} \in I_m(q)} \prod_{x_r \in \boldsymbol{X}} \frac{\psi_{i+\frac{1}{2}}^{[x_r]}}{x_r!} \right) \\ &= \left(\Delta x_{i+\frac{1}{2}}\right)^q \sum_{m=1}^{q} (-1)^m \sum_{\boldsymbol{X} \in I_m(q)} \prod_{x_r \in \boldsymbol{X}} \frac{\psi_{i+\frac{1}{2}}^{[x_r]}}{x_r!}. \end{aligned}$$ ■

So now with eqs. (4.12) and (4.14) we have a complete description of the approximation polynomial $F_{i+\frac{1}{2}}^{|p}(x)$. However, we actually only need the value of the flux at the interface $x_{i+\frac{1}{2}}$, i.e. $F_{i+\frac{1}{2}}^{|p}(x_{i+\frac{1}{2}})$. This is called the p-degree[5] **Scharfetter-Gummel flux with current correction** (SGCC-p) and is denoted as $f_{i+\frac{1}{2}}^{|p}$.

Definition 4.2 (SGCC-p scheme)**.** *To sum up the principal result of this section,*

$$f_{i+\frac{1}{2}}^{|p} = f_{i+\frac{1}{2}}^{[0|p]} = \frac{D}{\Delta x_{i+\frac{1}{2}}} \left(\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}}) Q_{i+\frac{1}{2},i}^{|p} - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}}) Q_{i+\frac{1}{2},i+1}^{|p} \right) \tag{4.18}$$

where $Q_{i+\frac{1}{2},r}^{|p} = \sum_{m=0}^{p} \left(\Delta x_{i+\frac{1}{2}}\right)^m W_{i+\frac{1}{2}}^{[m]} n_r^{(m)}$ *for* $r = i, i+1$*, the "W-functions"* $W_{i+\frac{1}{2}}^{[m]} \equiv \dfrac{\Theta_{i+\frac{1}{2}}^{[m|0]}}{\left(\Delta x_{i+\frac{1}{2}}\right)^m}$ *are defined in the following way,*

$$\begin{cases} W_{i+\frac{1}{2}}^{[0]} = 1, \\ W_{i+\frac{1}{2}}^{[m]} = \sum_{k=1}^{m} (-1)^k \sum_{\boldsymbol{X} \in I_k(m)} \prod_{x_r \in \boldsymbol{X}} \frac{\psi_{i+\frac{1}{2}}^{[x_r]}}{x_r!}, \quad m = 1, \dots, p, \end{cases} \tag{4.19}$$

and the ψ*-functions* $\psi_{i+\frac{1}{2}}^{[x_r]}$ *are defined in eq.* (4.10)*.*

Remark 4.5. *From definitions 4.1 and 4.2, it is straightforward to see that the standard SG scheme is in fact the SGCC-*0 *scheme. Therefore, the SGCC-*p *scheme is indeed a high-order generalization of the SG scheme.*

Remark 4.6. *In the case that* u *and/or* D *are smooth functions of* x*, we might replace* u *and* D *in eq.* (4.18) *with resp. the approximants* $u_{i+\frac{1}{2}}$ *and* $D_{i+\frac{1}{2}}$ *of* $u(x_{i+\frac{1}{2}})$ *and* $D(x_{i+\frac{1}{2}})$*. The numerical flux then features* $u_{i+\frac{1}{2}}$ *and* $D_{i+\frac{1}{2}}$ *instead of* u *and* D*.*

It is noticeable that the SGCC-p scheme (4.18) has the very same structure as the SG scheme (4.3). Therefore, it can be interpreted as the sum of a drift flux and a diffusion flux, i.e.

$$f_{i+\frac{1}{2}}^{|p} = u \frac{Q_{i+\frac{1}{2},i}^{|p} + Q_{i+\frac{1}{2},i+1}^{|p}}{2} - D\varphi(\mathfrak{p}_{i+\frac{1}{2}}) \frac{Q_{i+\frac{1}{2},i+1}^{|p} - Q_{i+\frac{1}{2},i}^{|p}}{\Delta x_{i+\frac{1}{2}}}.$$

[5] not to be confused with the precision order

Moreover, the SGCC-p flux can also be interpreted as a sum of the SG flux $f^{|0}_{i+\frac{1}{2}}$ and a residual flux, i.e.

$$f^{|p}_{i+\frac{1}{2}} = f^{|0}_{i+\frac{1}{2}} + \frac{D}{\Delta x_{i+\frac{1}{2}}}\left(\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}})\left(Q^{|p}_{i+\frac{1}{2},i} - n_i\right) - \mathcal{B}(-\mathfrak{p}_{i+\frac{1}{2}})\left(Q^{|p}_{i+\frac{1}{2},i+1} - n_{i+1}\right)\right).$$

We shall see in the next section that $f^{|p}_{i+\frac{1}{2}}$ is a $(p+1)^{\text{th}}$-order approximation of the real flux $f(x_{i+\frac{1}{2}})$, so the residual flux plays a role of a high-order correction added to the SG flux (or current, if we deal with motion of charges instead of particles); hence the name of the scheme.

Remark 4.7. *Finally, a distinction should be made when we choose $n^{(q)}(x^c_i)$ and $n^{(q)}(x^c_{i+1})$, or $\overline{n}^{(q)}_i$ and $\overline{n}^{(q)}_{i+1}$ in the place of $n^{(q)}_i$ and $n^{(q)}_{i+1}$, where $\overline{n}^{(q)}_i$ is a numerical approximation of $n^{(q)}(x^c_i)$ that can be computed via a density reconstruction technique. In the former case, the approximate flux, still denoted as $f^{|p}_{i+\frac{1}{2}}$, is a function of point-wise values of the real density derivatives $n^{(q)}$; we call it the **point-wise flux**. In the latter case, we denote the approximate flux as $\overline{f}^{|p}_{i+\frac{1}{2}}$, alternatively known as the **discrete flux**.*

4.4 Properties of the point-wise SGCC-p flux

In this section, the velocity u and the diffusion coefficient D are always assumed to be **constant**. In the case u and D are functions of x (see remark 4.6), the approximation error of u and D at the interfaces need to be accounted as it might alter the consistency order of the flux.

4.4.1 Drift and diffusion limits

It is instructive to inquire the behavior of the SGCC-p flux $f^{|p}_{i+\frac{1}{2}}$ in the diffusion limit $\mathfrak{p}_{i+\frac{1}{2}} \to 0$, as well as in the drift limit $\mathfrak{p}_{i+\frac{1}{2}} \to \pm\infty$, as we did for the SG flux in section 4.2. Let us at first cite in lemma 4.3 some useful properties of the ψ-functions (4.10) as well as the W-functions (4.19).

Lemma 4.3. *For $q \geq 0$,*

$$\psi^{[q]}_{i+\frac{1}{2}} \to (-\chi_{i+\frac{1}{2}})^q, \quad W^{[q]}_{i+\frac{1}{2}} \to \frac{\chi^q_{i+\frac{1}{2}}}{q!} \quad \text{as } \mathfrak{p}_{i+\frac{1}{2}} \to -\infty, \tag{4.20}$$

$$\psi^{[q]}_{i+\frac{1}{2}} \to \kappa^q_{i+\frac{1}{2}}, \quad W^{[q]}_{i+\frac{1}{2}} \to \frac{(-\kappa_{i+\frac{1}{2}})^q}{q!} \quad \text{as } \mathfrak{p}_{i+\frac{1}{2}} \to +\infty, \tag{4.21}$$

$$\psi^{[q]}_{i+\frac{1}{2}} \to \frac{\kappa^{q+1}_{i+\frac{1}{2}} - (-\chi_{i+\frac{1}{2}})^{q+1}}{q+1}, \quad W^{[q]}_{i+\frac{1}{2}} \to \mathfrak{W}^{[q]}_{i+\frac{1}{2}} \quad \text{as } \mathfrak{p}_{i+\frac{1}{2}} \to 0, \tag{4.22}$$

where $\mathfrak{W}^{[k]}_{i+\frac{1}{2}}$ satisfy $\displaystyle\sum_{k=0}^{q} \frac{\kappa^{q-k+1}_{i+\frac{1}{2}} - (-\chi_{i+\frac{1}{2}})^{q-k+1}}{(q-k+1)!}\mathfrak{W}^{[k]}_{i+\frac{1}{2}} = 0$ for $q \geq 1$.

Proof. For the ψ-functions $\psi^{[q]}_{i+\frac{1}{2}}$, the case $q = 0$ is trivial. Assume that eqs. (4.20) and (4.21) hold for $\psi^{[q]}_{i+\frac{1}{2}}$ for a certain $q \geq 0$, then the limits of $\psi^{[q]}_{i+\frac{1}{2}}$ as $\mathfrak{p}_{i+\frac{1}{2}} \to \pm\infty$ are finite. Then from eq. (4.10) it is

straightforward that eqs. (4.20) and (4.21) also hold for $\psi^{[q+1]}_{i+\frac{1}{2}}$. On the other hand, using eq. (4.9) we have[6]

$$\psi^{[q]}_{i+\frac{1}{2}} \to \frac{\int_{-\chi_{i+\frac{1}{2}}}^{\kappa_{i+\frac{1}{2}}} \xi^q \mathrm{d}\xi}{\int_{-\chi_{i+\frac{1}{2}}}^{\kappa_{i+\frac{1}{2}}} \mathrm{d}\xi} = \frac{\kappa^{q+1}_{i+\frac{1}{2}} - (-\chi_{i+\frac{1}{2}})^{q+1}}{q+1}, \quad \text{as } \mathfrak{p}_{i+\frac{1}{2}} \to 0.$$

For the W-functions $W^{[q]}_{i+\frac{1}{2}}$, the case $q = 0$ is trivial for eqs. (4.20) and (4.21). Let us remark that $W^{[q]}_{i+\frac{1}{2}} = \dfrac{\Theta^{[p|p-q]}_{i+\frac{1}{2}}}{\left(\Delta x_{i+\frac{1}{2}}\right)^q}$ for $0, \dots, p$. Therefore from eq. (4.15) we have, for $q = 1, \dots, p$,

$$W^{[q]}_{i+\frac{1}{2}} = -\sum_{k=0}^{q-1} \frac{\psi^{[q-k]}_{i+\frac{1}{2}}}{(q-k)!} W^{[k]}_{i+\frac{1}{2}}. \tag{4.23}$$

Assume that eqs. (4.20) and (4.21) hold for $W^{[k]}_{i+\frac{1}{2}}$ for all k smaller than a certain $q > 0$. So now we have

$$W^{[q]}_{i+\frac{1}{2}} \to -\sum_{k=0}^{q-1} \frac{(-\chi_{i+\frac{1}{2}})^{q-k}}{(q-k)!} \frac{\chi^k_{i+\frac{1}{2}}}{k!} = -\frac{\chi^q_{i+\frac{1}{2}}}{q!} \sum_{k=0}^{q} \frac{q!}{(q-k)!k!} 1^k (-1)^{q-k} + \frac{\chi^q_{i+\frac{1}{2}}}{q!} = \frac{\chi^q_{i+\frac{1}{2}}}{q!},$$

as $\mathfrak{p}_{i+\frac{1}{2}} \to -\infty$. Similarly, we have

$$W^{[q]}_{i+\frac{1}{2}} \to -\sum_{k=0}^{q-1} \frac{\kappa^{q-k}_{i+\frac{1}{2}}}{(q-k)!} \frac{(-\kappa_{i+\frac{1}{2}})^k}{k!}$$
$$= -\frac{\kappa^q_{i+\frac{1}{2}}}{q!} \sum_{k=0}^{q} \frac{q!}{(q-k)!k!} 1^{q-k} (-1)^k + \frac{\kappa^q_{i+\frac{1}{2}}}{q!} (-1)^q = \frac{(-\kappa_{i+\frac{1}{2}})^q}{q!},$$

as $\mathfrak{p}_{i+\frac{1}{2}} \to +\infty$. Finally, the case $\mathfrak{p}_{i+\frac{1}{2}} \to 0$ is straightforward by substitution of eq. (4.22) in eq. (4.23). ■

Now we can derive the diffusion and drift limits of the SGCC-p flux $f^{|p}_{i+\frac{1}{2}}$, with the assumption that n is a smooth function of x.

Proposition 4.1. *In drift limits, the SGCC-p flux $f^{|p}_{i+\frac{1}{2}}$ behaves as a $(p+1)^{th}$-order upwind scheme in a pure drift problem. More precisely,*

$$f^{|p}_{i+\frac{1}{2}} \to u \sum_{q=0}^{p} \frac{\left(x_{i+\frac{1}{2}} - x^c_i\right)^q}{q!} n^{(q)}(x^c_i), \quad \textit{as } \mathfrak{p}_{i+\frac{1}{2}} \to -\infty,$$
$$f^{|p}_{i+\frac{1}{2}} \to u \sum_{q=0}^{p} \frac{\left(x_{i+\frac{1}{2}} - x^c_{i+1}\right)^q}{q!} n^{(q)}(x^c_{i+1}), \quad \textit{as } \mathfrak{p}_{i+\frac{1}{2}} \to +\infty.$$

[6] recall that $\kappa_{i+\frac{1}{2}} \equiv \dfrac{x^c_{i+1} - x_{i+\frac{1}{2}}}{\Delta x_{i+\frac{1}{2}}}$, $\chi_{i+\frac{1}{2}} \equiv \dfrac{x_{i+\frac{1}{2}} - x^c_i}{\Delta x_{i+\frac{1}{2}}}$ and $\kappa_{i+\frac{1}{2}} + \chi_{i+\frac{1}{2}} = 1$

In the diffusion limit $\mathfrak{p}_{i+\frac{1}{2}} \to 0$, $f^{|p}_{i+\frac{1}{2}}$ behaves as a $(p+1)^{th}$-order central scheme in a pure diffusion problem. More precisely,

$$f^{|p}_{i+\frac{1}{2}} \sim Dn'(x_{i+\frac{1}{2}}) + \mathcal{O}\left(\left(\Delta x_{i+\frac{1}{2}}\right)^{p+1}\right).$$

Proof. Using $\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}}) \sim -\mathfrak{p}_{i+\frac{1}{2}}$ as $\mathfrak{p}_{i+\frac{1}{2}} \to -\infty$, $\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}}) \to 0$ as $\mathfrak{p}_{i+\frac{1}{2}} \to +\infty$ and eqs. (4.12), (4.20) and (4.21) yields

$$f^{|p}_{i+\frac{1}{2}} \to u\sum_{q=0}^{p}\left(\Delta x_{i+\frac{1}{2}}\right)^q \frac{\chi^q_{i+\frac{1}{2}}}{q!} n^{(q)}(x^c_i) = u\sum_{q=0}^{p}\frac{\left(x_{i+\frac{1}{2}} - x^c_i\right)^q}{q!} n^{(q)}(x^c_i), \quad \text{as } \mathfrak{p}_{i+\frac{1}{2}} \to -\infty,$$

$$f^{|p}_{i+\frac{1}{2}} \to u\sum_{q=0}^{p}\left(\Delta x_{i+\frac{1}{2}}\right)^q \frac{(-\kappa_{i+\frac{1}{2}})^q}{q!} n^{(q)}(x^c_{i+1}) = u\sum_{q=0}^{p}\frac{\left(x_{i+\frac{1}{2}} - x^c_{i+1}\right)^q}{q!} n^{(q)}(x^c_{i+1}), \quad \text{as } \mathfrak{p}_{i+\frac{1}{2}} \to +\infty.$$

It is straightforward that the right-hand side of each equation is the $(p+1)^{\text{th}}$-order Taylor expansion of $un(x)$ at x^c_i and x^c_{i+1} respectively, evaluated at $x_{i+\frac{1}{2}}$.

For the case $\mathfrak{p}_{i+\frac{1}{2}} \to 0$, using $\mathcal{B}(\mathfrak{p}_{i+\frac{1}{2}}) \to 1$ as $\mathfrak{p}_{i+\frac{1}{2}} \to 0$ and eqs. (4.12) and (4.22) we have

$$f^{|p}_{i+\frac{1}{2}} \to -\frac{D}{\Delta x_{i+\frac{1}{2}}}\sum_{q=0}^{p}\left(\Delta x_{i+\frac{1}{2}}\right)^q \mathfrak{W}^{[q]}_{i+\frac{1}{2}}\left(n^{(q)}(x^c_{i+1}) - n^{(q)}(x_i)\right).$$

Substituting the following Taylor developments,

$$\begin{aligned} n^{(q)}(x^c_{i+1}) &= n^{(q)}(x_{i+\frac{1}{2}}) + \sum_{m=1}^{p-q+1}\left(\Delta x_{i+\frac{1}{2}}\right)^m \frac{\kappa^m_{i+\frac{1}{2}}}{m!} n^{(q+m)}(x_{i+\frac{1}{2}}) + \mathcal{O}\left(\left(\Delta x_{i+\frac{1}{2}}\right)^{p-q+2}\right), \\ n^{(q)}(x_i) &= n^{(q)}(x_{i+\frac{1}{2}}) + \sum_{m=1}^{p-q+1}\left(\Delta x_{i+\frac{1}{2}}\right)^m \frac{(-\chi_{i+\frac{1}{2}})^m}{m!} n^{(q+m)}(x_{i+\frac{1}{2}}) + \mathcal{O}\left(\left(\Delta x_{i+\frac{1}{2}}\right)^{p-q+2}\right), \end{aligned} \tag{4.24}$$

to the above equation, we have

$$f^{|p}_{i+\frac{1}{2}} \sim -\frac{D}{\Delta x_{i+\frac{1}{2}}}\sum_{q=0}^{p}\left(\Delta x_{i+\frac{1}{2}}\right)^q \mathfrak{W}^{[q]}_{i+\frac{1}{2}}\sum_{m=1}^{p-q+1}\left(\Delta x_{i+\frac{1}{2}}\right)^m \frac{\kappa^m_{i+\frac{1}{2}} - (-\chi_{i+\frac{1}{2}})^m}{m!} n^{(q+m)}(x_{i+\frac{1}{2}}) + \mathcal{O}\left(\left(\Delta x_{i+\frac{1}{2}}\right)^{p+1}\right).$$

Making the index change $k = m + q - 1$ and switching the two summations yield

$$f^{|p}_{i+\frac{1}{2}} \sim -\frac{D}{\Delta x_{i+\frac{1}{2}}}\sum_{k=0}^{p}\left(\Delta x_{i+\frac{1}{2}}\right)^{k+1} n^{(k+1)}(x_{i+\frac{1}{2}})\sum_{q=0}^{k}\mathfrak{W}^{[q]}_{i+\frac{1}{2}}\frac{\kappa^{k-q+1}_{i+\frac{1}{2}} - (-\chi_{i+\frac{1}{2}})^{k-q+1}}{(k-q+1)!} + \mathcal{O}\left(\left(\Delta x_{i+\frac{1}{2}}\right)^{p+1}\right).$$

Finally, using the last point of lemma 4.3 for $k > 0$, we have

$$f^{|p}_{i+\frac{1}{2}} \sim -Dn'(x_{i+\frac{1}{2}}) + \mathcal{O}\left(\left(\Delta x_{i+\frac{1}{2}}\right)^{p+1}\right), \quad \text{as } \mathfrak{p}_{i+\frac{1}{2}} \to 0. \qquad \blacksquare$$

Corollary 4.1. *In particular, if $\chi_{i+\frac{1}{2}} = \kappa_{i+\frac{1}{2}}$ and p is even, then we have a super-convergence in the case $\mathfrak{p}_{i+\frac{1}{2}} \to 0$, since*

$$f^{|p}_{i+\frac{1}{2}} \sim Dn'(x_{i+\frac{1}{2}}) + \mathcal{O}\left(\left(\Delta x_{i+\frac{1}{2}}\right)^{p+2}\right).$$

Proof. The first observation is that $\mathfrak{W}^{[2l-1]}_{i+\frac{1}{2}} = 0$ for $l \geq 1$. Indeed, from the last equality in lemma 4.3 with $q = 2l-1$, $l \geq 1$, and the fact that $\chi_{i+\frac{1}{2}} = \kappa_{i+\frac{1}{2}}$, we have

$$\sum_{r=1}^{l} \mathfrak{W}^{[2r-1]}_{i+\frac{1}{2}} \frac{2\kappa^{2l-2r+1}_{i+\frac{1}{2}}}{(2l-2r+1)!} = 0.$$

Therefore, by substituting successively $l = 1, 2, \ldots$, it is straightforward that $\mathfrak{W}^{[2l-1]}_{i+\frac{1}{2}} = 0$.

Now let us proceed as in the proof of proposition 4.1, except that we take the Taylor expansion up to the $p-q+2$ degree in eq. (4.24). As a result, we have

$$f^{|p}_{i+\frac{1}{2}} \sim -Dn'(x_{i+\frac{1}{2}}) + \left(\Delta x_{i+\frac{1}{2}}\right)^{p+2} n^{(p+2)}(x_{i+\frac{1}{2}}) \sum_{q=0}^{p} \mathfrak{W}^{[q]}_{i+\frac{1}{2}} \frac{\kappa^{p+2-q}_{i+\frac{1}{2}} - (-\chi_{i+\frac{1}{2}})^{p+2-q}}{(p+2-q)!} + \mathcal{O}\left(\left(\Delta x_{i+\frac{1}{2}}\right)^{p+2}\right),$$

but $\sum_{q=0}^{p} \mathfrak{W}^{[q]}_{i+\frac{1}{2}} \frac{\kappa^{p+2-q}_{i+\frac{1}{2}} - (-\chi_{i+\frac{1}{2}})^{p+2-q}}{(p+2-q)!} = -\mathfrak{W}^{[p+1]}_{i+\frac{1}{2}} = 0$ since p is even. $\blacksquare$

4.4.2 Flux consistency

In this section, we investigate the consistency of the SGCC-p flux. More specifically, we want to see if the difference $\left|f^{|p}_{i+\frac{1}{2}} - f(x_{i+\frac{1}{2}})\right|$ decreases with the grid size and how fast if it does.

Let us at first remark an interesting relation between the approximation polynomial $F^{|p}_{i+\frac{1}{2}}(x)$ and the real flux $f(x)$ that emerges directly from eqs. (4.1) and (4.5),

$$\int_{x^c_i}^{x^c_{i+1}} \exp\left(-\frac{u}{D}x\right) \left(F^{|p}_{i+\frac{1}{2}}\right)^{(q)}(x)\mathrm{d}x = \int_{x^c_i}^{x^c_{i+1}} \exp\left(-\frac{u}{D}x\right) f^{(q)}(x)\mathrm{d}x, \quad q = 0, \ldots, p.$$

This property of $F^{|p}_{i+\frac{1}{2}}(x)$ allows us to derive the flux consistency of the SGCC-p scheme in the following proposition.

Proposition 4.2. *Assume that n is a smooth function of x. Then the SGCC-p scheme is consistent in the sense that*

$$\left|f^{|p}_{i+\frac{1}{2}} - f(x_{i+\frac{1}{2}})\right| < M_{i+\frac{1}{2}}(\mathrm{e}-1)^p \varsigma^{p+1}_{i+\frac{1}{2}} \left(\Delta x_{i+\frac{1}{2}}\right)^{p+1},$$

where $M_{i+\frac{1}{2}} = \sup_{x\in(x_i^c,x_{i+1}^c)} \left|un^{(p+1)}(x) - Dn^{(p+2)}(x)\right|$ *and* $\varsigma_{i+\frac{1}{2}} = \max\left(\kappa_{i+\frac{1}{2}}, \chi_{i+\frac{1}{2}}\right)$.

Proof. Let $g(x) = F_{i+\frac{1}{2}}^{|p}(x) - f(x)$ for $x \in (x_i^c, x_{i+1}^c)$ which satisfies $\int_{x_i^c}^{x_{i+1}^c} \exp\left(-\frac{u}{D}x\right) g^{(q)}(x)\mathrm{d}x = 0$ for $q = 0, \ldots, p$. For each q, the Taylor expansion of $g^{(q)}(x)$ at $x_{i+\frac{1}{2}}$ which stops at the term $g^{(p)}(x_{i+\frac{1}{2}})$ reads as,

$$g^{(q)}(x) = \sum_{m=q}^{p} \frac{g^{(m)}(x_{i+\frac{1}{2}})}{(m-q)!}\left(x - x_{i+\frac{1}{2}}\right)^{m-q} + R^{[q|p]}(x), \tag{4.25}$$

where the remainder $R^{[q|p]}(x)$ in the Lagrange form is defined as

$$R^{[q|p]}(x) = \frac{g^{(p+1)}(x^{[q]})}{(p-q+1)!}\left(x - x_{i+\frac{1}{2}}\right)^{p-q+1}, \tag{4.26}$$

for some real number $x^{[q]}$ between x and $x_{i+\frac{1}{2}}$. From the fact that $\int_{x_i^c}^{x_{i+1}^c} \exp\left(-\frac{u}{D}x\right) g^{(q)}(x)\mathrm{d}x = 0$ and eq. (4.25), we have

$$\begin{aligned} G\left|g^{(q)}(x_{i+\frac{1}{2}})\right| &\le \sum_{m=q+1}^{p} \frac{\left|g^{(m)}(x_{i+\frac{1}{2}})\right|}{(m-q)!}\left(\Delta x_{i+\frac{1}{2}}\right)^{m-q} \int_{x_i^c}^{x_{i+1}^c} \exp\left(-\frac{u}{D}x\right)\left|\frac{x - x_{i+\frac{1}{2}}}{\Delta x_{i+\frac{1}{2}}}\right|^{m-q}\mathrm{d}x \\ &\quad + \int_{x_i^c}^{x_{i+1}^c} \exp\left(-\frac{u}{D}x\right)\left|R^{[q|p]}(x)\right|\mathrm{d}x \\ &\le \sum_{m=q+1}^{p} \frac{\left|g^{(m)}(x_{i+\frac{1}{2}})\right|}{(m-q)!} G\varsigma_{i+\frac{1}{2}}^{m-q}\left(\Delta x_{i+\frac{1}{2}}\right)^{m-q} + \int_{x_i^c}^{x_{i+1}^c} \exp\left(-\frac{u}{D}x\right)\left|R^{[q|p]}(x)\right|\mathrm{d}x, \end{aligned} \tag{4.27}$$

where $G = \int_{x_i^c}^{x_{i+1}^c} \exp\left(-\frac{u}{D}x\right)\mathrm{d}x > 0$ and $\varsigma_{i+\frac{1}{2}} = \max\left(\kappa_{i+\frac{1}{2}}, \chi_{i+\frac{1}{2}}\right)$. From eq. (4.26), an estimate on $\left|R^{[q|p]}(x)\right|$ reads as

$$\left|R^{[q|p]}(x)\right| \le \frac{M_{i+\frac{1}{2}}}{(p-q+1)!}\left(\Delta x_{i+\frac{1}{2}}\right)^{p-q+1}\left|\frac{x - x_{i+\frac{1}{2}}}{\Delta x_{i+\frac{1}{2}}}\right|^{p-q+1} \le \frac{M_{i+\frac{1}{2}}\varsigma_{i+\frac{1}{2}}^{p-q+1}}{(p-q+1)!}\left(\Delta x_{i+\frac{1}{2}}\right)^{p-q+1}, \tag{4.28}$$

where $M_{i+\frac{1}{2}} = \sup_{x\in(x_i^c,x_{i+1}^c)} \left|g^{(p+1)}(x)\right| = \sup_{x\in(x_i^c,x_{i+1}^c)} \left|f^{(p+1)}(x)\right|$ since $\left(F_{i+\frac{1}{2}}^{|p}\right)^{(p+1)}(x) = 0$.

Now in the case $q = p$, it is straightforward from eqs. (4.27) and (4.28) that $\left|g^{(p)}(x_{i+\frac{1}{2}})\right| \le M_{i+\frac{1}{2}}\varsigma_{i+\frac{1}{2}}\Delta x_{i+\frac{1}{2}}$. Assume the ansatz that $\left|g^{(m)}(x_{i+\frac{1}{2}})\right| < M_{i+\frac{1}{2}}(\mathrm{e}-1)^{p-m}\varsigma_{i+\frac{1}{2}}^{p-m+1}\left(\Delta x_{i+\frac{1}{2}}\right)^{p-m+1}$ for $m = p, \ldots, q+1$ with a certain $q < p$, then combining with eqs. (4.27) and (4.28) we deduce that

$$\begin{aligned} \left|g^{(q)}(x_{i+\frac{1}{2}})\right| &< \sum_{m=q+1}^{p} \frac{M_{i+\frac{1}{2}}(\mathrm{e}-1)^{p-m}\varsigma_{i+\frac{1}{2}}^{p-q+1}}{(m-q)!}\left(\Delta x_{i+\frac{1}{2}}\right)^{p-q+1} + \frac{M_{i+\frac{1}{2}}\varsigma_{i+\frac{1}{2}}^{p-q+1}}{(p-q+1)!}\left(\Delta x_{i+\frac{1}{2}}\right)^{p-q+1} \\ &< M_{i+\frac{1}{2}}(\mathrm{e}-1)^{p-q-1}\varsigma_{i+\frac{1}{2}}^{p-q+1}\left(\Delta x_{i+\frac{1}{2}}\right)^{p-q+1}\sum_{m=1}^{p-q+1}\frac{1}{m!} < M_{i+\frac{1}{2}}(\mathrm{e}-1)^{p-q}\varsigma_{i+\frac{1}{2}}^{p-q+1}\left(\Delta x_{i+\frac{1}{2}}\right)^{p-q+1}. \end{aligned}$$

Therefore, the ansatz holds for every $q = p, \ldots, 0$, and in particular for $q = 0$. ∎

4.5 Extension of the SGCC-p flux scheme on two-dimensional grids

We have discussed the construction of the SG and SGCC-p schemes on one-dimensional grids in sections 4.2 and 4.3. Let us consider now a two-dimensional polygonal domain Ω as well as a grid $\mathcal{T}$ defined on Ω, with the notations introduced in section 3.1.

The extension of the SGCC schemes on two-dimensional grids is not straightforward since the segment $(\boldsymbol{x}_K^c \boldsymbol{x}_L^c)$ between the centers of gravity of two neighbor cells Ω_K and Ω_L is not always parallel to the normal $\boldsymbol{\nu}_{KL}$, so there is a lack of consistency if we evaluate the edge-normal flux $\boldsymbol{f} \cdot \boldsymbol{\nu}_{KL}$ on this segment [53, Chapter 3]. In this section, we present a technique of evaluating the normal flux-integral $\displaystyle\int_{\lambda_{KL}} \boldsymbol{f} \cdot \boldsymbol{\nu}_{KL} \mathrm{d}\boldsymbol{l}$ in a general setting (structured and unstructured grids). We note that this technique will be tested on **cartesian grids** in sections 4.8 and 5.3 and partly tested (only for the first-order SG scheme) on **triangular grids** with ONERA's own plasma solver COPAIER in chapter 7.

4.5.1 Flux integration on cell edges

We use the Gauss-Legendre quadrature [1] to evaluate the flux-integral $\displaystyle\int_{\lambda_{KL}} \boldsymbol{f}(\boldsymbol{x}) \cdot \boldsymbol{\nu}_{KL} \mathrm{d}\boldsymbol{l}$. For the integration of a smooth function $g(x)$ over the interval $[-1, 1]$, the quadrature rule takes the form $\displaystyle\int_{-1}^{1} g(x)\mathrm{d}x \approx \sum_{q=1}^{Q} w_q g\left(x_q\right)$ where Q is the number of sample points, w_q are the quadrature weights and x_q are the roots of the Q-degree Legendre polynomial. The set of w_q and x_q is the unique that allows to exactly integrate $(2Q-1)$-degree polynomials. The Gauss-Legendre quadrature rules for $Q = 1$ to 4 are listed in table 4.1.

Let $\boldsymbol{x}_{KL,l}^{\lambda}$ and $\boldsymbol{x}_{KL,r}^{\lambda}$ be the coordinates of the two end-points of the edge λ_{KL}. Then the coordinates of the q^{th} quadrature point on λ_{KL} is

$$\boldsymbol{x}_{KL,q}^{\lambda} \equiv \frac{\boldsymbol{x}_{KL,r}^{\lambda} - \boldsymbol{x}_{KL,l}^{\lambda}}{2} x_q + \frac{\boldsymbol{x}_{KL,l}^{\lambda} + \boldsymbol{x}_{KL,r}^{\lambda}}{2},$$

and the quadrature rule reads as

$$\int_{\lambda_{KL}} \boldsymbol{f}(\boldsymbol{x}) \cdot \boldsymbol{\nu}_{KL} \mathrm{d}\boldsymbol{l} \approx \frac{|\lambda_{KL}|}{2} \sum_{q=1}^{Q} w_q \boldsymbol{f}\left(\boldsymbol{x}_{KL,q}^{\lambda}\right) \cdot \boldsymbol{\nu}_{KL}.$$

For the following sections, let $\boldsymbol{x}_{KL,q}^{\Omega}$ (resp. $\boldsymbol{x}_{LK,q}^{\Omega}$) be the coordinates of the intersection between an edge of Ω_K (resp. Ω_L) other than λ_{KL} and the line orthogonal to λ_{KL} passing through $\boldsymbol{x}_{KL,q}^{\lambda}$ such that the segment $(\boldsymbol{x}_{KL,q}^{\lambda}, \boldsymbol{x}_{KL,q}^{\Omega})$ (resp. $(\boldsymbol{x}_{LK,q}^{\lambda}, \boldsymbol{x}_{LK,q}^{\Omega})$) lies entirely within Ω_K (resp. Ω_L) (see fig. 4.2). The evaluation points of the flux $\boldsymbol{f}\left(\boldsymbol{x}_{KL,q}^{\lambda}\right) \cdot \boldsymbol{\nu}_{KL}$ (analogous to the points x_i and x_{i+1} in section 4.3 for one-dimensional grids[7]) are defined in the following way,

$$\boldsymbol{x}_{KL,q} \equiv \frac{\boldsymbol{x}_{KL,q}^{\Omega} + \boldsymbol{x}_{KL,q}^{\lambda}}{2}, \quad \boldsymbol{x}_{LK,q} \equiv \frac{\boldsymbol{x}_{LK,q}^{\Omega} + \boldsymbol{x}_{KL,q}^{\lambda}}{2}.$$

[7] by this logic, $\boldsymbol{x}_{KL,q}^{\lambda}$ is analogous to $x_{i+\frac{1}{2}}$

Q	x_q	w_q
1	0	2
2	$\pm\frac{1}{\sqrt{3}}$	1
3	0 $\pm\sqrt{\frac{3}{5}}$	$\frac{8}{9}$ $\frac{5}{9}$
4	$\pm\sqrt{\frac{3}{7}-\frac{2}{7}\sqrt{\frac{6}{5}}}$ $\pm\sqrt{\frac{3}{7}+\frac{2}{7}\sqrt{\frac{6}{5}}}$	$\frac{18+\sqrt{30}}{36}$ $\frac{18-\sqrt{30}}{36}$

Table 4.1: Gauss-Legendre quadrature rules for $Q = 1$ to 4.

4.5.2 Edge-normal SGCC-p flux

We now evaluate $\boldsymbol{f}\left(\boldsymbol{x}^{\lambda}_{KL,q}\right)\cdot\boldsymbol{\nu}_{KL}$ for $q = 1,\dots,Q$ by following the flux evaluation technique in section 4.3 on the one-dimensional coordinates system defined by the line $(\boldsymbol{x}_{KL,q}\boldsymbol{x}_{LK,q})$, the vector $\boldsymbol{\nu}_{KL}$ and the origin $\boldsymbol{x}^{\lambda}_{KL,q}$.

Let $\aleph(y)\equiv n\left(\boldsymbol{x}^{\lambda}_{KL,q}+y\boldsymbol{\nu}_{KL}\right)$ be the restriction of $n(\boldsymbol{x})$ on this coordinates system with $y\in\mathbb{R}$. Let $\mathbf{D}^m n(\boldsymbol{x})\in\mathcal{S}^m\left(\mathbb{R}^2\right)$[8] be the m^{th}-derivative tensor of n, i.e. $\left(\mathbf{D}^m n(\boldsymbol{x})\right)_{k_1\dots k_m}=\dfrac{\partial^m n(\boldsymbol{x})}{\partial x_{k_1}\dots\partial x_{k_m}}$ with $k_r\in\{\,1,2\,\}$. Then the m^{th}-derivative of $\aleph$ reads as

$$\aleph^{(m)}_{KL,q}(y)=\mathbf{D}^m n\left(\boldsymbol{x}^{\lambda}_{KL,q}+y\boldsymbol{\nu}_{KL}\right)\bullet\boldsymbol{\nu}^{\otimes m}_{KL}. \tag{4.29}$$

Let $u_{KL}=\boldsymbol{u}\cdot\boldsymbol{\nu}_{KL}$, $\Delta x_{KL,q}=|\boldsymbol{x}_{KL,q}-\boldsymbol{x}_{LK,q}|$ and

$$\chi_{KL,q}=\frac{\left|\boldsymbol{x}_{KL,q}-\boldsymbol{x}^{\lambda}_{KL,q}\right|}{\Delta x_{KL,q}},\quad \kappa_{KL,q}=\frac{\left|\boldsymbol{x}_{LK,q}-\boldsymbol{x}^{\lambda}_{KL,q}\right|}{\Delta x_{KL,q}},\quad \mathfrak{p}_{KL,q}=-\frac{u_{KL}\Delta x_{KL,q}}{D}. \tag{4.30}$$

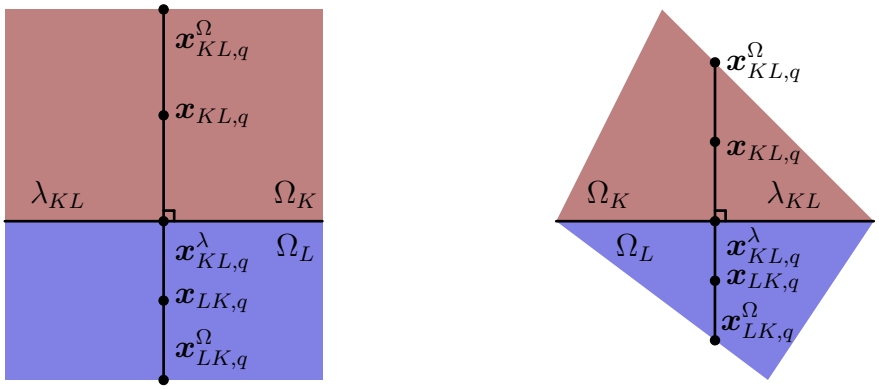
The normal component to λ_{KL} of the point-wise SGCC-p flux reads as follows,

$$f^{|p}_{KL,q}=\frac{D_{KL,q}}{\Delta x_{KL,q}}\left(\mathcal{B}(\mathfrak{p}_{KL,q})Q^{|p}_{KL,q}-\mathcal{B}(-\mathfrak{p}_{KL,q})Q^{|p}_{LK,q}\right), \tag{4.31}$$

where, remarking that

$$\aleph^{(m)}_{KL,q}\left(-\Delta x_{KL,q}\chi_{KL,q}\right)=\mathbf{D}^m n\left(\boldsymbol{x}_{KL,q}\right)\bullet\boldsymbol{\nu}^{\otimes m}_{KL},\quad \aleph^{(m)}_{KL,q}\left(\Delta x_{KL,q}\kappa_{KL,q}\right)=\mathbf{D}^m n\left(\boldsymbol{x}_{LK,q}\right)\bullet\boldsymbol{\nu}^{\otimes m}_{KL},$$

[8] $\mathcal{S}^m\left(\mathbb{R}^2\right)$ is the space of symmetric m-dimensional arrays, or symmetric tensors of rank m, with 2 elements in each dimension (see appendix A)



(a) A rectangular (structured) grid

(b) A triangular (unstructured) grid

Figure 4.2: Geometric notations

we have

$$Q^{|p}_{KL,q} = \sum_{m=0}^{p} (\Delta x_{KL,q})^m \, W^{[m]}_{KL,q} \aleph^{(m)}_{KL,q} \left(-\Delta x_{KL,q} \chi_{KL,q}\right),$$

$$Q^{|p}_{LK,q} = \sum_{m=0}^{p} (\Delta x_{KL,q})^m \, W^{[m]}_{KL,q} \aleph^{(m)}_{KL,q} \left(\Delta x_{KL,q} \kappa_{KL,q}\right)$$

In the above expressions, the ψ-functions $\psi^{[m]}_{KL,q}$, hidden in $W^{[m]}_{KL,q}$, are computed by replacing $\chi_{KL,q}$, $\kappa_{KL,q}$ and $\mathfrak{p}_{KL,q}$ for resp. $\chi_{i+\frac{1}{2}}$, $\kappa_{i+\frac{1}{2}}$ and $\mathfrak{p}_{i+\frac{1}{2}}$ in eq. (4.9). The W-functions $W^{[m]}_{KL,q}$ are computed from eq. (4.19) with $\psi^{[m]}_{KL,q}$ in place of $\psi^{[m]}_{i+\frac{1}{2}}$.

Remark 4.8. *In the case $\boldsymbol{u}$ and/or D are smooth functions of $\boldsymbol{x}$, we might replace u_{KL} and D in the Péclet number $\mathfrak{p}_{KL,q}$ in eq. (4.30) with resp. certain approximants $u_{KL,q}$ and $D_{KL,q}$ of $\boldsymbol{u}(\boldsymbol{x}^{\lambda}_{KL,q}) \cdot \nu_{KL}$ and $D(\boldsymbol{x}^{\lambda}_{KL,q})$.*

In the next step, the point-wise values of the real derivatives $\aleph^{(m)}$ have to be replaced with their numerical approximations for practical implementation of the flux schemes. More precisely, we need to compute the approximants of $\mathbf{D}^m n$, since in the finite volume method, the discrete derivatives of n are not stored, but only the cell-averaged values of density are. A derivative reconstruction technique will be the subject of the next section.

4.6 High-order reconstruction of particle density

Parts of this section is an adaptation from the paper [69]. It is edited for the sake of continuity of the workflow of this chapter. We emphasize that the reconstruction technique will be only tested on **cartesian grids** in sections 4.8 and 5.3, but in general it can be imagined on any conforming grids.

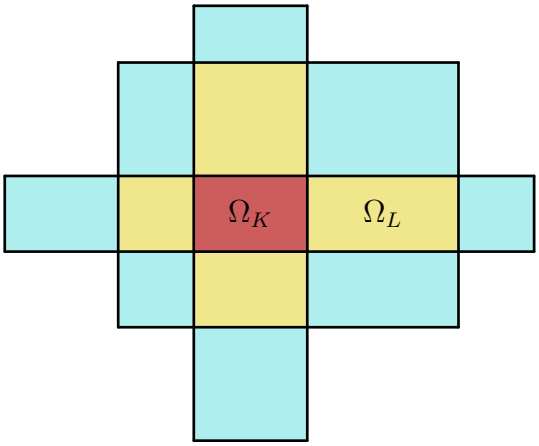


Figure 4.3: Ω_K and its neighbors

4.6.1 Cell neighborhoods

Let $\Omega_K \in \mathcal{T}$[9]. We define the **first-level neighborhood** of Ω_K, denoted as $\mathcal{V}_K^1$, as the set of elements of $\mathcal{T}$ that share an edge with Ω_K. Subsequently, we define the **second-level neighborhood**[10] of Ω_K, denoted as $\mathcal{V}_K^2$, as the set of all elements of $\mathcal{V}_K^1$ as well as those belonging to the first-level neighborhood of any element of $\mathcal{V}_K^1$, excluding Ω_K. In fig. 4.3, the elements of $\mathcal{V}_K^1$ are colored in yellow, while the elements of $\mathcal{V}_K^2$ are colored in yellow and blue. More generally, the k^{th}**-level neighborhood** of Ω_K is defined in the following way,

$$\begin{cases} \mathcal{V}_K^1 = \left\{ \Omega_L \in \mathcal{T} \mid \exists \lambda \in \mathcal{E}_K, \overline{\Omega}_L \cap \overline{\Omega}_K = \lambda \right\}, \\ \mathcal{V}_K^k = \displaystyle\bigcup_{\Omega_L \in \mathcal{V}_K^{k-1}} \mathcal{V}_L^1 \setminus \Omega_K, \quad k > 1. \end{cases}$$

We also employ the notations $V^k \equiv \text{Card}\left(\mathcal{V}_K^k\right)$[11] and $\overline{\mathcal{V}}_K^k \equiv \mathcal{V}_K^k \cup \{ \Omega_K \}$.

4.6.2 Ghost cells and boundary conditions

The definition of the neighborhood of a border cell[12] or a cell near the boundaries becomes problematic since the area of the neighborhood could extend beyond the computation domain. One simple approach to overcome this problem is to extend the computation domain to include a few additional cells beyond the boundaries, called **ghost cells**. For cartesian grids, we can construct a ghost cell by taking the symmetric image of a cell near a boundary Γ_k with respect to Γ_k. For example, in fig. 4.4, the ghost cells Ω_K^g, Ω_L^g, Ω_M^g and Ω_N^g are resp. the symmetry images of Ω_K, Ω_L, Ω_M and Ω_N. The ghost cells are imagined in this way until the neighborhood of each cell of $\mathcal{T}$ is well defined.

The values on the ghost cells are set at the beginning of each time level and depend on the type

[9] we refer again to section 3.1 for grid notations

[10] by convention, the second-level neighborhood is the same as the first-level neighborhood on **one-dimensional grids**

[11] assuming of course that every cell has the same number of edges, i.e. no mixing triangles/quadrilaterals

[12] a cell Ω_K is a **border cell** if there exists $\lambda \in \mathcal{E}_K$ such that $\lambda \subseteq \Gamma = \overline{\Omega} \setminus \Omega$

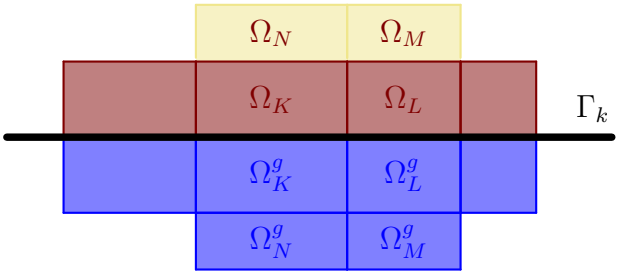


Figure 4.4: Border cells (maroon) and ghost cells (blue)

of boundary condition (BC) on Γ_k. If it is the homogeneous Dirichlet BC (for example if Γ_k is the wall boundary[13], i.e. $\Gamma_k = \Gamma_w$ (see section 2.2), the charge density inside the wall is assumed to be zero), then the density values on the ghost cells are zero. If it is the homogeneous Neumann BC (for example if $\Gamma_k = \Gamma_f$), then we set $\overline{n}_{K^g} = \overline{n}_{N^g} = \overline{n}_K$ and $\overline{n}_{L^g} = \overline{n}_{M^g} = \overline{n}_L$, where $\overline{n}_{K^g}$ is the density value on Ω_K^g. Finally, if Γ_k is the symmetry axis, then $\overline{n}_{K^g} = \overline{n}_K$, $\overline{n}_{L^g} = \overline{n}_L$, $\overline{n}_{M^g} = \overline{n}_M$ and $\overline{n}_{N^g} = \overline{n}_N$.

We note that this approach introduces first-order errors into the values on border cells that could potentially propagate into the interior cells [91, Chapter 7]. High-order extrapolations for density values on ghost cells were not addressed in this thesis and could be one of future works.

4.6.3 A reconstruction technique

Let $n(\boldsymbol{x})$ be a locally integrable function defined on Ω, i.e. $n \in L^1_{\text{loc}}(\Omega)$. Let $\mathcal{N} = \text{Card}(\mathcal{T})$ and $p > 0$ be a positive integer. Let $\mathbb{R}^{\mathcal{N}}_{|K}$ be the set of all $\mathcal{N}$-component arrays $\overline{\mathfrak{n}}$ with $\overline{\mathfrak{n}}_L = 0$ for $\Omega_L \notin \overline{\mathcal{V}}^p_K$.

Definition 4.3. *For each Ω_K an element of $\mathcal{T}$, the associated **local average operator** $\mathcal{P}^p_K$ is defined as*

$$\begin{aligned} \mathcal{P}^p_K : \quad & L^1_{\text{loc}}(\Omega) \to \mathbb{R}^{\mathcal{N}}_{|K} \\ & n(\boldsymbol{x}) \mapsto \overline{\mathfrak{n}} \quad \textit{with} \begin{cases} \overline{\mathfrak{n}}_L = \overline{n}_L, & \textit{if } \Omega_L \in \overline{\mathcal{V}}^p_K, \\ \overline{\mathfrak{n}}_L = 0, & \textit{otherwise.} \end{cases} \end{aligned}$$

As stated before, the writing of the discrete SGCC-p flux requires approximants of the derivatives of $n(\boldsymbol{x})$, but the information of $n(\boldsymbol{x})$ is only stored in its cell averages $\overline{n}_K$. In this section, on each neighborhood $\overline{\mathcal{V}}^p_K$, also called **reconstruction stencil**, we search for a polynomial approximation of $n(\boldsymbol{x})$ using the cell values $\overline{n}_K$ which possesses the properties of p-**exactness** and **local mass conservation** that are introduced in the following definitions.

Definition 4.4. *For each Ω_K, the associated **local reconstruction operator** $\mathcal{R}^p_K$ is defined as*

$$\mathcal{R}^p_K : \quad \mathbb{R}^{\mathcal{N}}_{|K} \to \mathbb{P}^p\left(\overline{\mathcal{V}}^p_K\right) \tag{4.32}$$

$$\overline{\mathfrak{n}} \mapsto \mathcal{N}(\boldsymbol{x}) = \sum_{m=0}^{p} \frac{1}{m!} \mathfrak{D}^{[m|p]} n_K \bullet (\boldsymbol{x} - \boldsymbol{x}^c_K)^{\otimes m} \tag{4.33}$$

[13] but note that in section 5.3, though, the BC on the wall is homogeneous Neumann

where $\bullet$ *is the contraction operator between two same-rank tensors (see appendix A) and* $\mathfrak{D}^{[m|p]}n_K \in \mathcal{S}^m\left(\mathbb{R}^2\right)$ *which depends linearly on* $\overline{\mathfrak{n}}$*, i.e.*

$$\mathfrak{D}^{[m|p]}n_K = \sum_{\Omega_L \in \overline{\mathcal{V}}_K^p} \mathbf{d}_{KL}^{[m|p]} \overline{n}_L, \tag{4.34}$$

with $\mathbf{d}_{KL}^{[m|p]} \in \mathcal{S}^m\left(\mathbb{R}^2\right)$*, independent of* $\overline{\mathfrak{n}}$*.*

Definition 4.5 (p-exactness [69])**.** *A reconstruction operator* $\mathcal{R}_K^p : \mathbb{R}_{|K}^{\mathcal{N}} \to \mathbb{P}^p\left(\overline{\mathcal{V}}_K^p\right)$ *is called* p*-exact on the neighborhood* $\overline{\mathcal{V}}_K^p$ *if* $\mathcal{R}_K^p$ *is the left inverse of the restriction of* $\mathcal{P}_K^p$ *on* $\mathbb{P}^p\left(\overline{\mathcal{V}}_K^p\right)$*, i.e.*

$$\mathcal{R}_K^p \circ \mathcal{P}^p_{K|\mathbb{P}^p\left(\overline{\mathcal{V}}_K^p\right)} = \mathrm{Id}_{\mathbb{P}^p\left(\overline{\mathcal{V}}_K^p\right)}.$$

The existence of such $\mathcal{R}_K^p$ exceeds the scope of this thesis and we refer to [69] for the discussion on this subject. But a necessary condition for the existence of the left inverse of $\mathcal{P}_K^p$ on $\mathbb{P}^p\left(\overline{\mathcal{V}}_K^p\right)$ is that $\mathrm{Card}\left(\overline{\mathcal{V}}_K^p\right) \geq \mathrm{Card}\left(\mathcal{P}_K^p\right)$. This explains why the reconstruction stencil needs to be enlarged as the polynomial degree increases.

In the following, we assume that the p-exact reconstruction $\mathcal{R}_K^p$ exists and focus on the technique of computing this operator. In particular, for a non-zero **constant** function $n(\boldsymbol{x})$ on Ω, we would have $\mathfrak{D}^{[m|p]}n_K = 0$, i.e. $\sum_{\Omega_L \in \overline{\mathcal{V}}_K^p} \mathbf{d}_{KL}^{[m|p]} = 0$, i.e. $\sum_{\Omega_L \in \mathcal{V}_K^p} \mathbf{d}_{KL}^{[m|p]} = -\mathbf{d}_{KK}^{[m|p]}$. Therefore,

$$\mathfrak{D}^{[m|p]}n_K = \sum_{\Omega_L \in \mathcal{V}_K^p} \mathbf{d}_{KL}^{[m|p]} \left(\overline{n}_L - \overline{n}_K\right). \tag{4.35}$$

Definition 4.6 (Local mass conservation)**.** *Let* $n \in L^1_{\mathrm{loc}}(\Omega)$ *and* $\mathcal{N}_K^{|p} = \mathcal{R}_K^p \circ \mathcal{P}_K^p n$*. The reconstruction operator* $\mathcal{R}_K^p : \mathbb{R}_{|K}^{\mathcal{N}} \to \mathbb{P}^p\left(\overline{\mathcal{V}}_K^p\right)$ *conserves mass locally if*

$$\frac{1}{|\Omega_K|}\int_{\Omega_K} \mathcal{N}_K^{|p}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \overline{n}_K = \frac{1}{|\Omega_K|}\int_{\Omega_K} n(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

Therefore, if $\mathcal{R}_K^p : \mathbb{R}_{|K}^{\mathcal{N}} \to \mathbb{P}^p\left(\overline{\mathcal{V}}_K^p\right)$ conserves mass locally then from eq. (4.33) we have $\mathfrak{D}^{[0|p]}n_K = \overline{n}_K - \sum_{m=1}^{p} \frac{1}{m!}\mathfrak{D}^{[m|p]}n_K \bullet \boldsymbol{h}_{KK}^{[m]}$ with $\boldsymbol{h}_{KL}^{[m]} \equiv \frac{1}{|\Omega_L|}\int_{\Omega_L} \left(\boldsymbol{x} - \boldsymbol{x}_K^c\right)^{\otimes m} \mathrm{d}\boldsymbol{x}$. Consequently,

$$\mathcal{N}_K^{|p}(\boldsymbol{x}) = \overline{n}_K + \sum_{m=1}^{p} \frac{1}{m!}\mathfrak{D}^{[m|p]}n_K \bullet \left[\left(\boldsymbol{x} - \boldsymbol{x}_K^c\right)^{\otimes m} - \boldsymbol{h}_{KK}^{[m]}\right]. \tag{4.36}$$

In the rest of this section, we assume that $n(\boldsymbol{x})$ is a polynomial, i.e. $n \in \mathbb{P}^p(\Omega)$. Then for $\Omega_L \in \overline{\mathcal{V}}_K^p$,

$$\overline{n}_L = n\left(\boldsymbol{x}_K^c\right) + \sum_{m=1}^{p} \frac{1}{m!}\mathbf{D}^m n\left(\boldsymbol{x}_K^c\right) \bullet \boldsymbol{h}_{KL}^{[m]}, \tag{4.37}$$

and so for $\Omega_L \in \mathcal{V}_K^p$,

$$\overline{n}_L - \overline{n}_K = \sum_{m=1}^{p} \frac{1}{m!}\mathbf{D}^m n\left(\boldsymbol{x}_K^c\right) \bullet \left(\boldsymbol{h}_{KL}^{[m]} - \boldsymbol{h}_{KK}^{[m]}\right). \tag{4.38}$$

Let us now compute the derivative approximants $\mathfrak{D}^{[m|p]}n_K$. Since $\mathbf{D}^m n\left(\boldsymbol{x}_K^c\right)$ and $\boldsymbol{h}_{KL}^{[m]} - \boldsymbol{h}_{KK}^{[m]}$ are elements of $\mathcal{S}^m$, eq. (4.38) could be explicitly rewritten in the following way,

$$\begin{aligned}\overline{n}_L - \overline{n}_K &= \sum_{m=1}^{p} \frac{1}{m!} \sum_{\boldsymbol{i}\in\overline{\mathcal{J}}^m} \left(\mathbf{D}^m n\left(\boldsymbol{x}_K^c\right)\right)_{\boldsymbol{i}} \left(\boldsymbol{h}_{KL}^{[m]} - \boldsymbol{h}_{KK}^{[m]}\right)_{\boldsymbol{i}} \\ &= \sum_{m=1}^{p} \sum_{\boldsymbol{i}\in\mathcal{J}^m} \frac{1}{\mathfrak{1}(\boldsymbol{i})!(m-\mathfrak{1}(\boldsymbol{i}))!} \left(\mathbf{D}^m n\left(\boldsymbol{x}_K^c\right)\right)_{\boldsymbol{i}} \left(\boldsymbol{h}_{KL}^{[m]} - \boldsymbol{h}_{KK}^{[m]}\right)_{\boldsymbol{i}} \\ &= \sum_{m=1}^{p} \sum_{\boldsymbol{i}\in\mathcal{J}^m} \left(\mathbf{D}^m n\left(\boldsymbol{x}_K^c\right)\right)_{\boldsymbol{i}} \left(\widetilde{\boldsymbol{h}_{KL}^{[m]} - \boldsymbol{h}_{KK}^{[m]}}\right)_{\boldsymbol{i}}, \end{aligned} \tag{4.39}$$

where $\overline{\mathcal{J}}^m$ is the set of indices of a $\mathcal{S}^m\left(\mathbb{R}^2\right)$-tensor, $\mathcal{J}^m$ and the tilded version of a $\mathcal{S}^m\left(\mathbb{R}^2\right)$-tensor are defined in appendix A.

Let $d^m \equiv \dim\left(\mathbb{P}^m\left(\mathbb{R}^2\right)\right) - 1 = \dfrac{(m+1)(m+2)}{2} - 1$. We define the **sorting operator** Σ^p as follows,

$$\begin{aligned}\Sigma^p : \quad \mathcal{S}^1 \times \cdots \times \mathcal{S}^p &\to \mathbb{R}^{d^p} \\ \left(\boldsymbol{a}^1, \ldots, \boldsymbol{a}^p\right) &\mapsto A\end{aligned}$$

Here A is an array such that $A_{l+d^{m-1}}$, for $l = 1, \ldots, m+1$ and $m = 1, \ldots, p$, is the l^{th} element of $\boldsymbol{a}^m$ that is stored[14]; the stored elements of $\boldsymbol{a}^m$ are sorted by an order induced by $\mathcal{J}^m$.

Example 4.1. *For* $p = 3$, $A = \begin{pmatrix} a_1^1 & a_2^1 & a_{11}^2 & a_{12}^2 & a_{22}^2 & a_{111}^3 & a_{112}^3 & a_{122}^3 & a_{222}^3 \end{pmatrix}^t$.

With this notation, we define the following arrays,

$$\begin{aligned} U_K^p &\equiv \Sigma^p\left(\mathbf{D}^1 n\left(\boldsymbol{x}_K^c\right), \ldots, \mathbf{D}^p n\left(\boldsymbol{x}_K^c\right)\right), \\ D_{KL}^p &\equiv \Sigma^p\left(\mathbf{d}_{KL}^{[1|p]}, \ldots, \mathbf{d}_{KL}^{[p|p]}\right), \\ H_{KL}^p &\equiv \Sigma^p\left(\widetilde{\boldsymbol{h}_{KL}^{[1]} - \boldsymbol{h}_{KK}^{[1]}}, \ldots, \widetilde{\boldsymbol{h}_{KL}^{[p]} - \boldsymbol{h}_{KK}^{[p]}}\right), \\ H_K^p(\boldsymbol{x}) &\equiv \Sigma^p\left(\widetilde{(\boldsymbol{x} - \boldsymbol{x}_K^c)^{\otimes 1}} - \boldsymbol{h}_{KK}^{[1]}, \ldots, \widetilde{(\boldsymbol{x} - \boldsymbol{x}_K^c)^{\otimes p}} - \boldsymbol{h}_{KK}^{[p]}\right). \end{aligned}$$

Moreover, let $\mathcal{D}_K^p$ be a $d^p \times V_K^p$ matrix and $\mathcal{H}_K^p$ be a $V_K^p \times d^p$ matrix with $V_K^p \equiv \operatorname{Card}\left(\mathcal{V}_K^p\right)$, such that

$$\mathcal{D}_K^p = \left(D_{K1}^p \ldots D_{KV_K^p}^p\right), \quad \mathcal{H}_K^p = \left(H_{K1}^p \ldots H_{KV_K^p}^p\right)^t.$$

Now eq. (4.39) re-reads as $\overline{n}_L - \overline{n}_K = U_K^p \cdot H_{KL}^p$. We substitute this into eq. (4.35) and then eq. (4.36) to obtain

$$\left(\mathcal{R}_K^p \circ \mathcal{P}_K^p n\right)(\boldsymbol{x}) = \mathcal{N}_K^{|p}(\boldsymbol{x}) = \overline{n}_K + \left(\mathcal{D}_K^p \mathcal{H}_K^p\right) U_K^p \cdot H_K^p(\boldsymbol{x}).$$

On the other hand, from eq. (4.37) we deduce that

$$n(\boldsymbol{x}) = \overline{n}_K + U_K^p \cdot H_K^p(\boldsymbol{x}).$$

[14] not all 2^m elements of $\boldsymbol{a}^m$ are stored since $\boldsymbol{a}^m$ is a symmetric tensor

Hence, if $\mathcal{R}_K^p \circ \mathcal{P}_K^p = \mathrm{Id}_{\mathbb{P}^p(\overline{\mathcal{V}}_K^p)}$ then it is sufficient that $\mathcal{D}_K^p \mathcal{H}_K^p = \mathcal{I}^p$, where $\mathcal{I}^p$ is the $d^p \times d^p$ identity matrix. The solutions $\mathcal{D}_K^p$ of this problem are in general not unique since $V_K^p \geq d^p$. A particular solution known as the least-square solution is the **Moore-Penrose inverse** of $\mathcal{H}_K^p$ which minimizes the Frobenius norm $\sqrt{\mathrm{Tr}\left(\mathcal{D}_K^p\right)^t \mathcal{D}_K^p}$ [68] and reads as

$$\mathcal{D}_K^p = \left(\left(\mathcal{H}_K^p\right)^t \mathcal{H}_K^p\right)^{-1} \left(\mathcal{H}_K^p\right)^t. \tag{4.40}$$

Finally, the derivative approximants $\mathfrak{D}^{[m|p]} n_K$ can be evaluated from the **reconstruction matrix** $\mathcal{D}_K^p$ according to eq. (4.34). We note that the **grid matrix** $\mathcal{H}_K^p$ only depends on the grid parameters (distance, cell size, etc.). Therefore, if the grid is fixed then $\mathcal{H}_K^p$, and consequently $\mathcal{D}_K^p$, could be computed once for all at the beginning of the simulation.

4.6.4 Approximation error

In the more general case where $n(\boldsymbol{x})$ is a smooth function on Ω, we have the following result which is inspired from [69].

Theorem 4.1. *Let $p > 0$ be an integer and make the following assumptions.*

1. *Each component of $\mathbf{D}^{p+1} n$ is bounded on Ω.*
2. *There exists a constant $C_0 > 0$ independent of $h_{\mathcal{T}}$[15] such that for all $\Omega_K \in \mathcal{T}$ and $m = 0, \ldots, p$, the reconstruction operator satisfies, for any function $g \in L^1_{\mathrm{loc}}(\Omega)$,*

$$\left\|\mathbf{D}^m \left(\mathcal{R}_K^p \circ \mathcal{P}_K^p g\right)\right\|_{\left(L^\infty(\mathcal{V}_K^p)\right)^{2^m}} \leq \frac{C_0}{h_{\mathcal{T}}^m} |\mathcal{P}_K^p g|,$$

where $|\mathcal{P}_K^p g|$ is the Euclidean norm of the array $\mathcal{P}_K^p g$.

Then there exists a constant $C > 0$ independent of $h_{\mathcal{T}}$ such that for all $\Omega_K \in \mathcal{T}$ and $m = 0, \ldots, p$,

$$\left\|\mathbf{D}^m \left(\mathcal{R}_K^p \circ \mathcal{P}_K^p n\right) - \mathbf{D}^m n\right\|_{\left(L^\infty(\mathcal{V}_K^p)\right)^{2^m}} \leq C h_{\mathcal{T}}^{p-m+1}. \tag{4.41}$$

Proof. We follow the proof of Theorem 3.9 in [69]. Apply Taylor's theorem to $n(\boldsymbol{x})$ at the point $\boldsymbol{x}_K^c$ we have $n(\boldsymbol{x}) = p(\boldsymbol{x}) + r(\boldsymbol{x})$ with

$$p(\boldsymbol{x}) = \sum_{m=0}^{p} \frac{1}{m!} \mathbf{D}^m n\left(\boldsymbol{x}_K^c\right) \bullet \left(\boldsymbol{x} - \boldsymbol{x}_K^c\right)^{\otimes m},$$

and there exists a constant $C_1 > 0$ independent of $h_{\mathcal{T}}$ such that $\left\|\mathbf{D}^m r\right\|_{\left(L^\infty(\mathcal{V}_K^p)\right)^{2^m}}$ is bounded by $C_1 h_{\mathcal{T}}^{p-m+1}$ for $m = 0, \ldots, p$, since $\mathbf{D}^{p+1} n$ is bounded on Ω.

[15] definition in section 3.1

Since $\mathcal{R}_K^p$ is exact for all polynomials $\mathbb{P}^p\left(\mathbb{R}^2\right)$, we have $\mathcal{R}_K^p \circ \mathcal{P}_K^p n = p + \mathcal{R}_K^p \circ \mathcal{P}_K^p r$ and as a result $\mathbf{D}^m\left(\mathcal{R}_K^p \circ \mathcal{P}_K^p n\right) = \mathbf{D}^m p + \mathbf{D}^m\left(\mathcal{R}_K^p \circ \mathcal{P}_K^p r\right)$. Consequently,

$$\begin{aligned}\left\|\mathbf{D}^m\left(\mathcal{R}_K^p \circ \mathcal{P}_K^p n\right) - \mathbf{D}^m n\right\|_{\left(L^\infty(\mathcal{V}_K^p)\right)^{2^m}} &= \left\|\mathbf{D}^m\left(\mathcal{R}_K^p \circ \mathcal{P}_K^p r\right) - \mathbf{D}^m r\right\|_{\left(L^\infty(\mathcal{V}_K^p)\right)^{2^m}} \\ &\leq \frac{C_0}{h_{\mathcal{T}}^m}|\mathcal{P}_K^p r| + C_1 h_{\mathcal{T}}^{p-m+1}.\end{aligned}$$

Since[16] $|\mathcal{P}_K^p r| \leq V_K^p C_1 h_{\mathcal{T}}^{p+1}$, it is straightforward that

$$\left\|\mathbf{D}^m\left(\mathcal{R}_K^p \circ \mathcal{P}_K^p n\right) - \mathbf{D}^m n\right\|_{\left(L^\infty(\mathcal{V}_K^p)\right)^{2^m}} \leq \left(V_K^p C_1 C_0 + C_1\right) h_{\mathcal{T}}^{p-m+1}. \qquad \blacksquare$$

4.6.5 The discrete SGCC-p flux

We denote $\mathcal{N}_K^{|p}(\boldsymbol{x}) = \mathcal{R}_K^p \circ \mathcal{P}_K^p n(\boldsymbol{x})$. Let us define, as the discrete counterpart of eq. (4.29),

$$\overline{\aleph}_{KL,q}^{(m)}(y) \equiv \mathbf{D}^m \mathcal{N}_K^{|p}\left(\boldsymbol{x}_{KL,q}^{\lambda} + y\boldsymbol{\nu}_{KL}\right) \bullet \boldsymbol{\nu}_{KL}^{\otimes m}, \quad y \in \mathbb{R},$$

and replace $\aleph_{KL,q}^{(m)}$ in section 4.5.2 and eq. (4.31) with $\overline{\aleph}_{KL,q}^{(m)}$. We achieve the discrete SGCC-p flux $\overline{f}_{KL,q}^{|p}$ which reads as

$$\begin{cases}\overline{f}_{KL,q}^{|p} = \dfrac{D_{KL,q}}{\Delta x_{KL,q}}\left(\mathcal{B}(\mathfrak{p}_{KL,q})\overline{Q}_{KL,q}^{|p} - \mathcal{B}(-\mathfrak{p}_{KL,q})\overline{Q}_{LK,q}^{|p}\right), \\ \overline{Q}_{KL,q}^{|p} = \displaystyle\sum_{m=0}^{p}\left(\Delta x_{KL,q}\right)^m W_{KL,q}^{[m]}\overline{\aleph}_{KL,q}^{(m)}\left(-\Delta x_{KL,q}\chi_{KL,q}\right), \\ \overline{Q}_{LK,q}^{|p} = \displaystyle\sum_{m=0}^{p}\left(\Delta x_{KL,q}\right)^m W_{KL,q}^{[m]}\overline{\aleph}_{KL,q}^{(m)}\left(\Delta x_{KL,q}\kappa_{KL,q}\right).\end{cases}$$

4.7 Choices of slope limiters

4.7.1 For piece-wise linear reconstruction on one-dimensional grids

Consider the grid in fig. 4.1 and an a cell $\Omega_i \equiv (x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$ such that $x_{i-\frac{1}{2}}$, $x_{i+\frac{1}{2}}$ do not coincide with the domain boundaries. For linear reconstruction, i.e. $p = 1$, the grid matrix reads as $\mathcal{H}_i^1 = \begin{pmatrix} x_{i+1} - x_i \\ x_{i-1} - x_i \end{pmatrix} = \begin{pmatrix} \Delta x_{i+\frac{1}{2}} \\ -\Delta x_{i-\frac{1}{2}} \end{pmatrix}$, and so from eq. (4.40) the reconstruction matrix is

$$\mathcal{D}_i^1 = \left(\frac{\Delta x_{i+\frac{1}{2}}}{\left(\Delta x_{i-\frac{1}{2}}\right)^2 + \left(\Delta x_{i+\frac{1}{2}}\right)^2}, -\frac{\Delta x_{i-\frac{1}{2}}}{\left(\Delta x_{i-\frac{1}{2}}\right)^2 + \left(\Delta x_{i+\frac{1}{2}}\right)^2}\right).$$

Plugging this into eq. (4.35) we derive the gradient approximant as follows,

$$\mathfrak{D}^{[1|1]}n_i = \frac{\Delta x_{i+\frac{1}{2}}\left(\overline{n}_{i+1} - \overline{n}_i\right)}{\left(\Delta x_{i-\frac{1}{2}}\right)^2 + \left(\Delta x_{i+\frac{1}{2}}\right)^2} + \frac{\Delta x_{i-\frac{1}{2}}\left(\overline{n}_i - \overline{n}_{i-1}\right)}{\left(\Delta x_{i-\frac{1}{2}}\right)^2 + \left(\Delta x_{i+\frac{1}{2}}\right)^2}. \tag{4.42}$$

[16] recall that $V_K^p = \mathrm{Card}\left(\mathcal{V}_K^p\right)$

It is well known that this central-difference approximation generates numerical solutions with spurious oscillations whenever there is a large-gradient layer in the density, typically in drift-dominant flows, i.e. for large Péclet numbers $\mathfrak{p}$. For stability reason, we are interested in a special class of numerical methods that are **total-variation diminishing** (TVD), being known for their ability to suppress the undesirable oscillations [90]. We refer to [90, 53, 91] for a detailed lecture, or to appendix B for a brief presentation on this subject.

The linear scheme with the slope in eq. (4.42) is not TVD since it would contradict the fact that any linear TVD scheme is at most first-order [90, Theorem 16.1]. In order to achieve a TVD method while still retaining second-order accuracy for smooth solutions, one possible choice is to use a **TVD slope limiter** $\Phi(\theta)$ so that the linear reconstruction of $n(t^l, x)$ on the cell Ω_i reads as

$$\mathcal{N}_i^{l|1}(x) = \overline{n}_i^l + \frac{\Delta x_i}{2} \frac{\overline{n}_{i+1}^l - \overline{n}_i^l}{\Delta x_{i+\frac{1}{2}}} \Phi\left(\theta_i^l\right),$$

for the case $u > 0$, where $\theta_i \equiv \dfrac{\overline{n}_i^l - \overline{n}_{i-1}^l}{\overline{n}_{i+1}^l - \overline{n}_i^l}$. In the case $u < 0$, we simply switch to

$$\mathcal{N}_i^{l|1}(x) = \overline{n}_i^l + \frac{\Delta x_i}{2} \frac{\overline{n}_i^l - \overline{n}_{i-1}^l}{\Delta x_{i-\frac{1}{2}}} \Phi\left(\theta_i^l\right),$$

with $\theta_i^l \equiv \dfrac{\overline{n}_{i+1}^l - \overline{n}_i^l}{\overline{n}_i^l - \overline{n}_{i-1}^l}$. So for now on we assume that $u > 0$.

Let us assume now that the grid is **uniform**, i.e. $\Delta x_i = \Delta x > 0$. The notion of TVD slope limiters as well as some examples of limiters that will be used in section 4.8 and chapter 5 are introduced below.

Definition 4.7 ([137]). *A slope limiter $\Phi(\theta)$ is called TVD if* $\begin{cases} 0 \le \dfrac{\Phi(\theta)}{\theta} \le 2 \\ 0 \le \Phi(\theta) \le 2 \end{cases}$ *. In this case, the graph of Φ locates within the **TVD region**.*

Example 4.2. *Some examples of TVD slope limiters [91, Chapter 6] (see fig. 4.5).*

1. *Minmod* $\quad \Phi^{\mathrm{M}}(\theta) = \max(0, \min(1, \theta))$.
2. *Superbee* $\quad \Phi^{\mathrm{S}}(\theta) = \max(0, \min(1, 2\theta), \min(2, \theta))$.
3. *Monotonized central-difference (MC)* $\quad \Phi^{\mathrm{MC}}(\theta) = \max\left(0, \min\left(\dfrac{1+\theta}{2}, \beta, \beta\theta\right)\right)$ *with* $1 \le \beta \le 2$.

We now investigate whether the SGCC-1 scheme combining with TVD slope limiters are actually TVD.

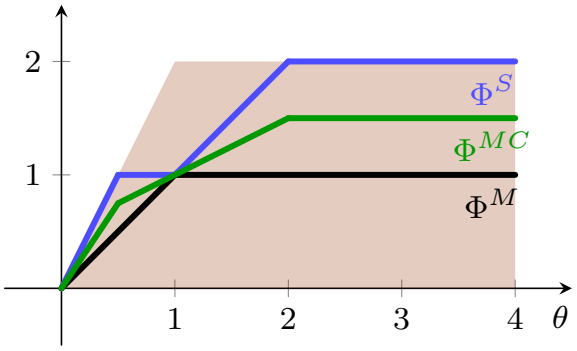


Figure 4.5: Some limiters with the TVD region in maroon and $\beta = 1.5$ for Φ^{MC}

Definition 4.8. *The TVD-SGCC-1 flux schemes writes as follows,*

$$\widehat{f}^{l|1}_{i+\frac{1}{2}} = \frac{D}{\Delta x}\left(\mathcal{B}(\mathfrak{p})\left(n^l_i + \Delta x W^{[1]}\frac{\overline{n}^l_{i+1} - \overline{n}^l_i}{\Delta x}\Phi\left(\theta^l_i\right)\right)\right.$$
$$\left. - \mathcal{B}(-\mathfrak{p})\left(n^l_{i+1} + \Delta x W^{[1]}\frac{\overline{n}^l_{i+2} - \overline{n}^l_{i+1}}{\Delta x}\Phi\left(\theta^l_{i+1}\right)\right)\right), \quad (4.43)$$

with $\mathfrak{p} = -\dfrac{u\Delta x}{D}$ *and* $W^{[1]} = -\psi^{[1]} = -\dfrac{1}{2}\dfrac{\mathrm{e}^{\mathfrak{p}}+1}{\mathrm{e}^{\mathfrak{p}}-1} + \dfrac{1}{\mathfrak{p}}$.

The following result is due to Harten [70].

Theorem 4.2. *Suppose that a numerical scheme could be written in the form*

$$\overline{n}^{l+1}_i = \overline{n}^l_i + A^l_i\left(\overline{n}^l_{i+1} - \overline{n}^l_i\right) - B^l_{i-1}\left(\overline{n}^l_i - \overline{n}^l_{i-1}\right). \quad (4.44)$$

If $A^l_i \geq 0$, $B^l_i \geq 0$ *and*

- $A^l_i + B^l_i \leq 1$, *then the scheme is TVD;*
- $A^l_i + B^l_{i-1} \leq 1$, *then* $\|\overline{n}\left(t^{l+1},\cdot\right)\|_{L^\infty(\mathbb{R})} \leq \|\overline{n}\left(t^l,\cdot\right)\|_{L^\infty(\mathbb{R})}$.

Substituting eq. (4.43) into eq. (9) yields a numerical scheme in the form of eq. (4.44) with

$$A^l_i = a\left(1 + W^{[1]}\frac{\Phi\left(\theta^l_{i+1}\right)}{\theta^l_{i+1}}\right), \quad B^l_{i-1} = b\left(1 - W^{[1]}\Phi\left(\theta^l_{i-1}\right)\right) + (a+b)W^{[1]}\frac{\Phi\left(\theta^l_i\right)}{\theta^l_i},$$
$$a = \frac{D\Delta t}{(\Delta x)^2}\mathcal{B}(-\mathfrak{p}), \quad b = \frac{D\Delta t}{(\Delta x)^2}\mathcal{B}(\mathfrak{p}).$$

Based on theorem 4.2, we can demonstrate that the TVD-SGCC-1 schemes are in fact TVD.

Proposition 4.3. *Assume that* $\overline{n}(t^0,\cdot) \in \mathrm{BV}(\mathbb{R})$[17]. *Then the scheme* (4.44) *with the numerical flux* (4.43) *is TVD and satisfies* $\|\overline{n}\left(t^{l+1},\cdot\right)\|_{L^\infty(\mathbb{R})} \leq \|\overline{n}\left(t^l,\cdot\right)\|_{L^\infty(\mathbb{R})}$, *for* $l = 0,\ldots,\mathcal{L}-1$, *if the following CFL condition is respected,*

$$\Delta t \leq \frac{\Delta x}{2u}\tanh\left(\frac{u\Delta x}{2D}\right) \equiv \Delta t_{\mathrm{CFL}}. \quad (4.45)$$

[17] the definition of the $\mathrm{BV}(\mathbb{R})$ space can be found in appendix B

Proof. Let us first notice that $a, b > 0$ and $0 \leq W^{[1]} \leq \frac{1}{2}$ since $\mathfrak{p} < 0$ (we have assumed that $u > 0$). With $\Phi\left(\theta_i^l\right)$ being a TVD slope limiter (see definition 4.7), it is straightforward that $A_i^l, B_i^l \geq 0$.

The inequality $A_i^l + B_i^l \leq 1$ is equivalent to

$$1 - \frac{b}{a+b} W^{[1]} \Phi\left(\theta_i^l\right) + \frac{2a+b}{a+b} W^{[1]} \frac{\Phi\left(\theta_{i+1}^l\right)}{\theta_{i+1}^l} \leq \frac{1}{a+b}.$$

We have the following estimates,

$$0 \leq \frac{b}{a+b} W^{[1]} \Phi\left(\theta_i^l\right) \leq \frac{2b}{a+b} W^{[1]}, \quad 0 \leq \frac{2a+b}{a+b} W^{[1]} \frac{\Phi\left(\theta_{i+1}^l\right)}{\theta_{i+1}^n} \leq 2\frac{2a+b}{a+b} W^{[1]}.$$

It can be shown that $\frac{2b}{a+b} W^{[1]} \leq 1$ and $2\frac{2a+b}{a+b} W^{[1]}$ is decreasing as well as bounded from above by 1 for $\mathfrak{p} < 0$, so

$$\left| \frac{b}{a+b} W^{[1]} \Phi\left(\theta_i^l\right) - \frac{2a+b}{a+b} W^{[1]} \frac{\Phi\left(\theta_{i+1}^l\right)}{\theta_{i+1}^l} \right| \leq 1,$$

and hence,

$$1 - \frac{b}{a+b} W^{[1]} \Phi\left(\theta_i^l\right) + \frac{2a+b}{a+b} W^{[1]} \frac{\Phi\left(\theta_{i+1}^l\right)}{\theta_{i+1}^l} \leq 2.$$

Thus, a sufficient condition for which the scheme is TVD, according to theorem 4.2, is that $a + b \leq \frac{1}{2}$ which is equivalent to the CFL condition (4.45)[18].

Finally, the inequality $A_i^l + B_{i-1}^l \leq 1$ is proven in the same manner since it is equivalent to

$$1 - \frac{b}{a+b} W^{[1]} \Phi\left(\theta_{i-1}^l\right) + \frac{a}{a+b} W^{[1]} \frac{\Phi\left(\theta_{i+1}^l\right)}{\theta_{i+1}^l} + W^{[1]} \frac{\Phi\left(\theta_i^l\right)}{\theta_i^l} \leq \frac{1}{a+b}.$$

Therefore, from theorem 4.2 we have $\|\overline{n}\left(t^{l+1}, \cdot\right)\|_{L^\infty(\mathbb{R})} \leq \|\overline{n}\left(t^l, \cdot\right)\|_{L^\infty(\mathbb{R})}$ under the condition (4.45). ■

4.7.2 For piece-wise parabolic reconstruction on one-dimensional grids

For $p = 2$, the grid matrix reads as

$$\mathcal{H}_i^2 = \begin{pmatrix} \Delta x_{i+\frac{1}{2}} & \frac{1}{6}\Delta x_{i+\frac{1}{2}} \Delta x_{i,i+1} \\ -\Delta x_{i-\frac{1}{2}} & \frac{1}{6}\Delta x_{i-\frac{1}{2}} \Delta x_{i,i-1} \end{pmatrix}$$

with $\Delta x_{i,i\pm1} \equiv \Delta x_i + 2\Delta x_{i\pm1}$. Thus, the reconstruction matrix is

$$\mathcal{D}_i^2 = \left(\mathcal{H}_i^2\right)^{-1} = \frac{1}{\Delta x_{i+\frac{1}{2}} \Delta x_{i-\frac{1}{2}} \left(\Delta x_{i,i+1} + \Delta x_{i,i-1}\right)} \begin{pmatrix} \Delta x_{i-\frac{1}{2}} \Delta x_{i,i-1} & 6\Delta x_{i-\frac{1}{2}} \\ -\Delta x_{i+\frac{1}{2}} \Delta x_{i,i+1} & 6\Delta x_{i+\frac{1}{2}} \end{pmatrix}$$

[18] recall that $\mathcal{B}(\mathfrak{p}) + \mathcal{B}(-\mathfrak{p}) = \mathfrak{p} \coth\left(\frac{\mathfrak{p}}{2}\right)$

Plugging this into eq. (4.35) we derive the derivative approximants as follows,

$$\mathfrak{D}^{[1|2]}n_i = \frac{\Delta x_{i-\frac{1}{2}}\Delta x_{i,i-1}\left(\overline{n}_{i+1}-\overline{n}_i\right)+\Delta x_{i+\frac{1}{2}}\Delta x_{i,i+1}\left(\overline{n}_i-\overline{n}_{i-1}\right)}{\Delta x_{i+\frac{1}{2}}\Delta x_{i-\frac{1}{2}}\left(\Delta x_{i,i+1}+\Delta x_{i,i-1}\right)},$$
$$\mathfrak{D}^{[2|2]}n_i = 6\frac{\Delta x_{i-\frac{1}{2}}\left(\overline{n}_{i+1}-\overline{n}_i\right)-\Delta x_{i+\frac{1}{2}}\left(\overline{n}_i-\overline{n}_{i-1}\right)}{\Delta x_{i+\frac{1}{2}}\Delta x_{i-\frac{1}{2}}\left(\Delta x_{i,i+1}+\Delta x_{i,i-1}\right)}.$$

We use an extension of the MC limiter to piece-wise parabolic reconstruction [162] which is a sequential application of the MC limiter to the second derivative approximant and then to the first derivative approximant. This approach is a particular case of the parameter-free generalized moment limiter for high-order methods (GML) proposed in [162]. We define as such, for $1 \leq \beta \leq 2$,

$$\widehat{\mathfrak{D}}^{[2|2]}n_i \equiv \text{minmod}\left(\mathfrak{D}^{[2|2]}n_i, \beta\frac{\mathfrak{D}^{[1|2]}n_{i+1}-\mathfrak{D}^{[1|2]}n_i}{\Delta x_{i+\frac{1}{2}}}, \beta\frac{\mathfrak{D}^{[1|2]}n_i-\mathfrak{D}^{[1|2]}n_{i-1}}{\Delta x_{i-\frac{1}{2}}}\right),$$
$$\widehat{\mathfrak{D}}^{[1|2]}n_i \equiv \text{minmod}\left(\mathfrak{D}^{[1|2]}n_i, \beta\frac{\overline{n}_{i+1}-\overline{n}_i}{\Delta x_{i+\frac{1}{2}}}, \beta\frac{\overline{n}_i-\overline{n}_{i-1}}{\Delta x_{i-\frac{1}{2}}}\right).$$

To our knowledge, there is not yet a theoretical result on the stability or convergence of this limiting approach. However, there have been numerical evidences showing that this scheme is stable for some simple convection problems with sharp discontinuities and exhibits high accuracy near the shocks [162]. Furthermore, even though some overshoots of the numerical solutions were observed near the discontinuity, their amplitude is very small and do not transform into spurious oscillations [154].

4.7.3 A limiter on two-dimensional grids

On two-dimensional grid, we extend the slope limiter of Barth & Jespersen [10] for high-order reconstruction polynomials and flux integration over the edges. For a cell Ω_K, let

$$\overline{n}_K^{\max} \equiv \max_{\Omega_L \in \overline{\mathcal{V}}_K^1} \overline{n}_L, \quad \overline{n}_K^{\min} \equiv \min_{\Omega_L \in \overline{\mathcal{V}}_K^1} \overline{n}_L$$

be resp. the local maximum and local minimum discrete values of the density on the neighborhood $\overline{\mathcal{V}}_K^1$. The idea is to calibrate the reconstruction polynomial such that its values at any flux evaluation point on the edges of Ω_K, i.e. the points $\boldsymbol{x}_{KL,q}^{\lambda}$ (see fig. 4.2), for every $\Omega_L \in \mathcal{V}_K^1$, fall between $\overline{n}_K^{\min}$ and $\overline{n}_K^{\max}$.

More precisely, let us consider a Q-point quadrature rule (see section 4.5), $\overline{\mathbf{n}} = (\overline{n}_K)_{\Omega_K \in \mathcal{T}}$ and $\mathcal{N}_K^{|p}(\boldsymbol{x}) = \left(\mathcal{R}_K^p \overline{\mathbf{n}}\right)(\boldsymbol{x})$. For every point $\boldsymbol{x}_{KL,q}^{\lambda}$, we define the "local" limiter $\Phi_{KL,q}$ in the following way,

$$\Phi_{KL,q} = \begin{cases} \min\left(1, \dfrac{\overline{n}_K^{\max}-\overline{n}_K}{\mathcal{N}_K^{|p}\left(\boldsymbol{x}_{KL,q}^{\lambda}\right)-\overline{n}_K}\right) & \text{if} \quad \mathcal{N}_K^{|p}\left(\boldsymbol{x}_{KL,q}^{\lambda}\right) > \overline{n}_K, \\ \min\left(1, \dfrac{\overline{n}_K^{\min}-\overline{n}_K}{\mathcal{N}_K^{|p}\left(\boldsymbol{x}_{KL,q}^{\lambda}\right)-\overline{n}_K}\right) & \text{if} \quad \mathcal{N}_K^{|p}\left(\boldsymbol{x}_{KL,q}^{\lambda}\right) < \overline{n}_K, \\ 1 & \text{otherwise}, \end{cases}$$

and then define the limiter Φ_K as the minimum of these $\Phi_{KL,q}$, i.e.

$$\Phi_K \equiv \min_{\Omega_L \in \mathcal{V}_K^1} \min_{q=1,\dots,Q} \Phi_{KL,q}.$$

Subsequently, the limited version of the derivative approximants read as

$$\widehat{\mathfrak{D}}^{[m|p]} n_K = \Phi_K \mathfrak{D}^{[m|p]} n_K,$$

and therefore from eq. (4.36), the limited version of the reconstructed density reads as

$$\widehat{\mathcal{N}}_K^{|p}(\boldsymbol{x}) = \overline{n}_K + \sum_{m=1}^{p} \frac{1}{m!} \widehat{\mathfrak{D}}^{[m|p]} n_K \bullet \left[(\boldsymbol{x} - \boldsymbol{x}_K)^{\otimes m} - \boldsymbol{h}_{KK}^{[m]} \right] = \overline{n}_K + \Phi_K \left(\mathcal{N}_K^{|p}(\boldsymbol{x}) - \overline{n}_K \right).$$

We can easily check that $\overline{n}_K^{\min} \leq \widehat{\mathcal{N}}_K^{|p}\left(\boldsymbol{x}_{KL,q}^{\lambda}\right) \leq \overline{n}_K^{\max}$ for every L, q, so we could at least avoid the overshoots or undershoots of numerical solutions near large-gradient layers due to the reconstruction process. However, there is a drawback of this limiter which is that if $\overline{n}_K$ is a local maximum or a local minimum, i.e. $\overline{n}_K = \overline{n}_K^{\max}$ or $\overline{n}_K = \overline{n}_K^{\min}$, then $\Phi_K = 0$ and we lost the high-order precision on the cell Ω_K since $\widehat{\mathcal{N}}_K^{|p}(\boldsymbol{x}) = \overline{n}_K$. This loss of accuracy will be observed in section 4.8. Furthermore, at this point there is no theoretical proof that the numerical method ensued from this limiting technique is stable and convergent; the only evidences are numerical observations which will be presented in sections 4.8 and 5.3.

Remark 4.9. *In practical implementation with variable u, D and non-uniform grids, the timesteps are limited by the CFL condition of electrons,*

$$\Delta t^l \leq \mathfrak{C} \min_{K=1,\dots,\mathcal{N}} \left(\frac{h_K}{u_{e,K}^l} \tanh\left(\frac{u_{e,K}^l h_K}{2 D_{e,K}^l} \right) \right) \equiv \mathfrak{C} \Delta t_e^l, \tag{4.46}$$

where $u_{e,K}^l$, $D_{e,K}^l$ are resp. the approximations of $|\boldsymbol{u}_e(t^l, \boldsymbol{x}_K^c)|$, $D_e(t^l, \boldsymbol{x}_K^c)$ and $\mathfrak{C} \leq \frac{1}{2}$ is a user-defined parameter. We also define the CFL condition of ions that will be useful for the time-implicit simulations in chapter 7,

$$\Delta t^l \leq \mathfrak{C} \min_{s \in \mathfrak{S} \setminus \{ e, \overline{e}_e \}} \min_{K=1,\dots,\mathcal{N}} \left(\frac{h_K}{u_{s,K}^l} \tanh\left(\frac{u_{s,K}^l h_K}{2 D_{s,K}^l} \right) \right) \equiv \mathfrak{C} \Delta t_{\text{ion}}^l. \tag{4.47}$$

4.8 Numerical verification

4.8.1 On one-dimensional grids

We begin the numerical tests with two one-dimensional examples to verify the spatial convergence order, only for the SGCC-1 and SGCC-2 schemes. The superbee as well as minmod limiters (see section 4.7.1) are employed for the first scheme while the GML limiter (see section 4.7.2) is used for the latter. We also include numerical results of the standard SG method as well as the second-order

MUSCL-central-difference scheme [90] for comparison with the novel methods. The MUSCL-central-difference flux scheme reads as follows, assuming that $u > 0$,

$$f_{i+\frac{1}{2}}^{\text{MUSCL}} = u\left(n_i + \frac{\Delta x_i}{2}\frac{n_{i+1}-n_i}{\Delta x_{i+\frac{1}{2}}}\Phi^M(\theta_i)\right) - D\frac{n_{i+1}-n_i}{\Delta x_{i+\frac{1}{2}}}.$$

Throughout this section, we consider uniform grids and the third-order SSP-Runge-Kutta scheme[19] is used for time discretization. The timestep is set to $\mathfrak{C}\Delta t_{\text{CFL}}$ (see eq. (4.45)) with $\mathfrak{C} = 0.8$. The first test is the ideal transport-diffusion of a Gaussian hat with constant diffusion rate and drift velocity. The other one taken from [83] is the transport-diffusion of a hyperbolic tangent profile.

Test 4.1 (moving Gaussian hat). *We consider $(0,T)\times\Omega = (0,0.25)\times(0,1)$, a constant drift velocity $u = 1$, a constant diffusion rate D taking three different values of 10^{-2}, 10^{-4}, 10^{-6} as well as the homogeneous Neumann condition on the boundaries. The initial datum is*

$$n^0(x) = \exp\left(-\frac{(x-x_0)^2}{2\sigma}\right),$$

with $x_0 = 0.25$ and $\sigma = 10^{-4}$. Since $n^0(x)$ is flat and near-zero near the domain boundaries, the exact solution of eq. (4.1) with this initial datum could be considered as

$$n^{\text{ex}}(t,x) = \left(\frac{\sigma}{2Dt+\sigma}\right)^{\frac{1}{2}}\exp\left(-\frac{(x-x_0-ut)^2}{4Dt+2\sigma}\right).$$

When the simulations are finished, we evaluate the discretization error in the L^1-norm of the numerical solutions at the last iteration $\mathcal{L} = \dfrac{T}{\Delta t}$. The discretization error of a scheme S reads as follows,

$$e_{\Delta x}^S \equiv \|\overline{n}^{\mathcal{L}} - n^{\text{ex}}(T,\cdot)\|_{L^1(\Omega)}.$$

An estimation of the **convergence order** is then given by $\log_2\left(\dfrac{e_{\Delta x}^S}{e_{\frac{\Delta x}{2}}^S}\right)$. We also evaluate the error ratio,

$$r_{\Delta x}^S \equiv \frac{e_{\Delta x}^S}{e_{\Delta x}^{\text{SGCC-2}}},$$

for more clarity in lecture.

The numerical results are grouped in tables 4.2 to 4.4. We could observe that when the diffusion is large, e.g. $D = 10^{-2}$, the SG, SGCC-1 and MUSCL schemes are second-order. Meanwhile, the SGCC-2 scheme is fourth-order, which is predicted by corollary 4.1.

The convergence orders deteriorate with the diffusion rate. The SG scheme is at most first-order since the drift is dominant. On the contrary the SGCC-1 and MUSCL schemes maintain more or less a high convergence order. Overall, SGCC-1 are more accurate than MUSCL in this test case. It is worth noted that the embedded TVD limiters are first-order at extrema, so while the grid is sparse it is comprehensible that these schemes are first-order. The same observation and argument could be

$\mathcal{N}$	SG			SGCC-1			MUSCL			SGCC-2	
	$e_{\Delta x}$	order	$r_{\Delta x}$	$e_{\Delta x}$	order	$r_{\Delta x}$	$e_{\Delta x}$	order	$r_{\Delta x}$	$e_{\Delta x}$	order
100	3.26e-2	_	1.9	1.62e-2	_	0.95	1.63e-2	_	0.95	1.71e-2	_
200	2.79e-2	0.23	4.07	4.96e-3	1.72	0.72	4.95e-3	1.72	0.72	6.85e-3	1.32
400	2.22e-2	0.33	11.6	3e-3	0.72	1.56	3e-3	0.72	1.56	1.92e-3	1.84
800	1.64e-2	0.43	56.9	1.35e-3	1.15	4.69	1.35e-3	1.15	4.69	2.88e-4	2.73
1600	1.11e-2	0.56	309	6.98e-4	0.96	19.4	7.02e-4	0.94	19.6	3.59e-5	3.00
3200	6.9e-3	0.69	1516	2.49e-4	1.49	54.7	2.51e-4	1.48	55.2	4.55e-6	2.98

Table 4.2: Test 4.1, $D = 10^{-6}$. Convergence order in L^1-norm, errors in function of number of grid cells $\mathcal{N}$.

$\mathcal{N}$	SG			SGCC-1			MUSCL			SGCC-2	
	$e_{\Delta x}$	order	$r_{\Delta x}$	$e_{\Delta x}$	order	$r_{\Delta x}$	$e_{\Delta x}$	order	$r_{\Delta x}$	$e_{\Delta x}$	order
100	2.95e-2	_	2.2	1.24e-2	_	0.93	1.25e-2	_	0.93	1.34e-2	_
200	2.43e-2	0.28	5.52	2.99e-3	2.05	0.68	3.06e-3	2.03	0.7	4.4e-3	1.6
400	1.82e-2	0.42	19.5	1.7e-3	0.82	1.81	1.77e-3	0.79	1.89	9.35e-4	2.23
800	1.2e-2	0.6	99.2	8.36e-4	1.02	6.9	1.01e-3	0.81	8.35	1.21e-4	2.95
1600	6.49e-3	0.89	467	2.03e-4	2.04	14.6	3.22e-4	1.64	23.2	1.39e-5	3.12
3200	2.56e-3	1.34	1925	4.28e-5	2.24	32.2	8.83e-5	1.86	66.4	1.33e-6	3.39

Table 4.3: Test 4.1, $D = 10^{-4}$. Convergence order in L^1-norm.

$\mathcal{N}$	SG			SGCC-1			MUSCL			SGCC-2	
	$e_{\Delta x}$	order	$r_{\Delta x}$	$e_{\Delta x}$	order	$r_{\Delta x}$	$e_{\Delta x}$	order	$r_{\Delta x}$	$e_{\Delta x}$	order
100	9.73e-4	_	57.2	2.25e-4	_	13.2	4.89e-4	_	28.8	1.7e-5	_
200	2.54e-4	1.94	244	5.63e-5	2	54.1	1.71e-4	1.52	164	1.04e-6	4.03
400	6.42e-5	1.98	1017	1.4e-5	2.01	222	4.58e-5	1.9	726	6.31e-8	4.04
800	1.61e-5	2	4128	3.49e-6	2	895	1.16e-5	1.98	2974	3.9e-9	4.02
1600	4.02e-6	2	16475	8.71e-7	2	3570	2.93e-6	1.99	12008	2.44e-10	4
3200	1.01e-6	2	63522	2.18e-7	2	13711	7.33e-7	2	46101	1.59e-11	3.94

Table 4.4: Test 4.1, $D = 10^{-2}$. Convergence order in L^1-norm.

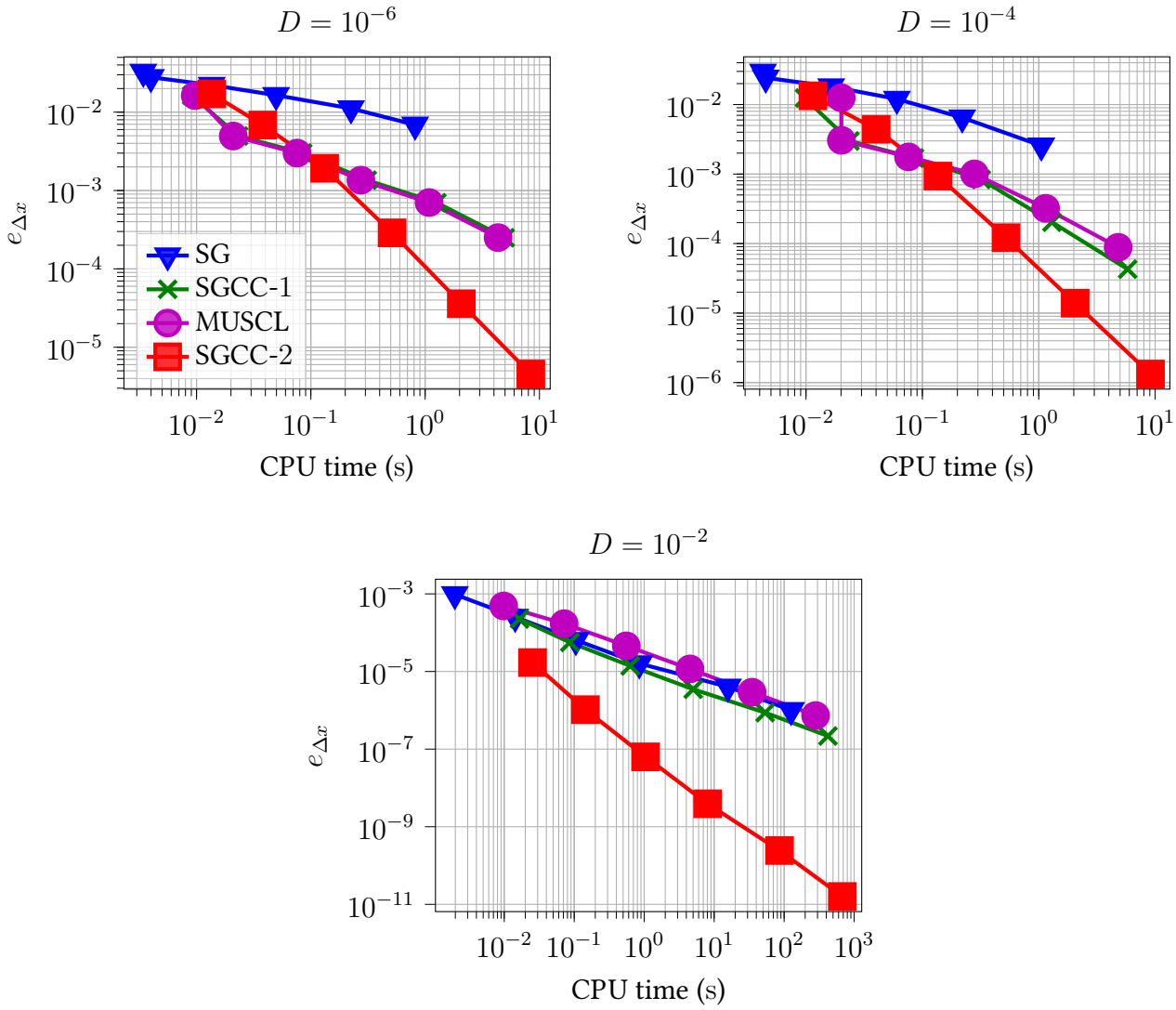


Figure 4.6: Test 4.1. L^1-norm errors $e_{\Delta x}$ in function of the CPU time.

applied for the SGCC-2 scheme combining with the GML limiter. But overall, the SGCC-2 scheme is third-order in drift-dominant regimes.

Figure 4.6 displays the L^1-norm errors $e^S_{\Delta x}$ in function of the CPU time. Apparently, the greater number of cells $\mathcal{N}$ we have, the more CPU time the simulation costs. We observe clearly that for $D = 10^{-6}$, $D = 10^{-4}$ and $\mathcal{N}$ greater than 400 (corresponding to the third node from the left on each curve), the computation time of higher-order schemes are shorter than lower-order schemes for the same level of precision. For $D = 10^{-2}$, the performance of SGCC-2 on $\mathcal{N} = 400$ grid cells (third node from the left on the red curve) is even better than lower-order schemes on $\mathcal{N} = 3200$ elements (last nodes on other curves). These indicate that the complexity of high-order schemes is well compensated by their enhanced accuracy.

Test 4.2 (moving canyon). *We take* $(0, T) \times \Omega = (0, 4 \times 10^{-5}) \times (0, 1)$, $\mathcal{N} = 200$, $u(x) = -ax$,

[19] see example 3.3

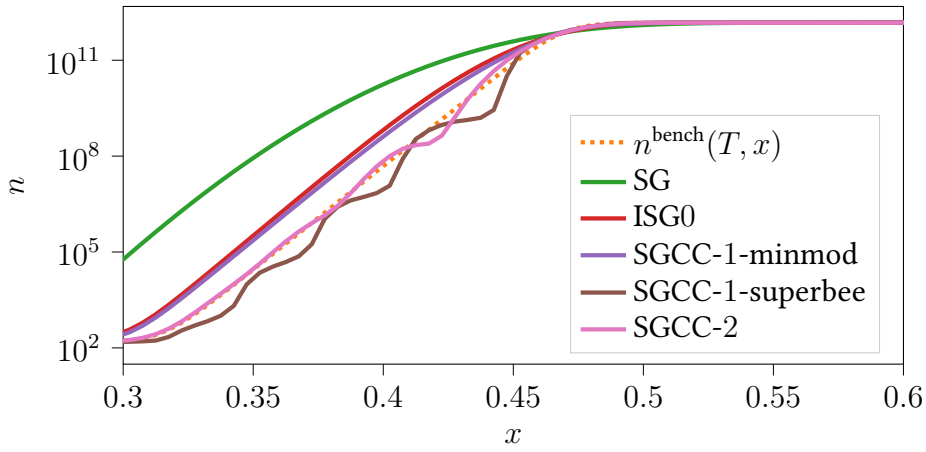


Figure 4.7: Test 4.2. Numerical and benchmark solutions at $T = 4 \times 10^{-5}$ for $x \in (0.3, 0.6)$.

$D = 1$ *and*

$$n^0(x) = a_1 + \frac{a_2}{2}\left(1 + \tanh\left(\frac{x - x_0}{\sigma}\right)\right),$$

with $a = 10^4$, $a_1 = 10^2$, $a_2 = 10^{12}$, $x_0 = 0.7$ *and* $\sigma = 0.02$. *If we neglect the diffusion term, the solution to the resulting drift equation, which could serve as a benchmark since* $D \ll u\Delta x$, *is given in the following way,*

$$n^{\text{bench}}(t,x) = n^0\left(x\mathrm{e}^{at}\right)\mathrm{e}^{at}.$$

In fig. 4.7, we illustrate the numerical solutions at the final time T of the SG-based schemes. We remark that the assumptions that guarantee TVD solutions in section 4.7.1 (constant cell size and drift velocity) do not hold in this test. We also include the solution of Kulikovsky's improved Scharfetter-Gummel scheme (ISG0) with $\epsilon = 0.01$ (see [83]). It does not exhibit any oscillations near the large-gradient layer, but this is usually not guaranteed.

The numerical solution of the MUSCL scheme are very close to SGCC-1 so it is not displayed. The SGCC-1-superbee scheme seems to be more compressing since it generates a solution that is wavy in the large-gradient layer. On the other hand, the SGCC-1-minmod is more diffusing as it leads to a more smooth-looking solution. The SG scheme is too diffusive as it smears the large-gradient layer a lot more than other schemes. Finally, the SGCC-2 scheme also produces some ripples in the large-gradient layer but overall, this numerical solution is the closest to the benchmark solution $n^{\text{bench}}(t,x)$.

4.8.2 On two-dimensional cartesian grids

We extend test 4.1 on uniform $\mathcal{N} \times \mathcal{N}$ square grids with $\mathcal{N}$ elements on each direction. In this section, we verify the convergence order of the SGCC-p schemes for $p = 1, 2, 3, 4$ in two cases: *(i)* with the slope limiter Φ_K defined in section 4.7.3 and *(ii)* without limiter, i.e. $\Phi_K = 0$. The time scheme is always the third-order SSP-Runge-Kutta method and the timestep is set to $\mathfrak{C}\Delta t_{\text{CFL}}$ (see eq. (4.45)) with $\mathfrak{C} = 0.8$.

Test 4.3 (2D Gaussian hat). *We consider* $(0,T)\times\Omega = (0,0.5)\times(0,1)^2$*, a constant drift velocity* $\boldsymbol{u} = (u_1,u_2)^t = (0.5,0.25)^t$*, a constant diffusion rate* D *taking two different values of* $10^{-4}, 10^{-6}$ *as well as the homogeneous Neumann condition on the boundaries. The initial datum is*

$$n^0(\boldsymbol{x}) = \exp\left(-\frac{(x_1-a)^2}{\sigma_1} - \frac{(x_2-b)^2}{\sigma_2}\right),$$

with $\boldsymbol{x} = (x_1,x_2)^t$*,* $a = 0.3$*,* $b = 0.5$*,* $\sigma_1 = 0.004$ *and* $\sigma_2 = 0.003$*. The solution of the dimensional version of eq. (4.1),*

$$\partial_t n + \nabla\cdot(\boldsymbol{u}n - D\nabla n) = 0,$$

with this initial datum reads as

$$n^{\mathrm{ex}}(t,\boldsymbol{x}) = \left(\frac{\sigma_1\sigma_2}{(4Dt+\sigma_1)(4Dt+\sigma_2)}\right)^{\frac{1}{2}} \exp\left(-\frac{(x_1-a-u_1t)^2}{4Dt+\sigma_1} - \frac{(x_2-b-u_2t)^2}{4Dt+\sigma_2}\right).$$

In the first case, the numerical results **with** the limiter Φ_K defined in section 4.7.3 are grouped in tables 4.5 and 4.6. We could observe that when the diffusion is large, e.g. $D = 10^{-4}$, all the schemes except SGCC-4 exhibits good convergence order, i.e. $p+1$, which is coherent with the flux consistency analysis in section 4.4. However, when the diffusion coefficient decreases, their convergence rates also drop. Specifically for the SGCC-3 scheme with $D = 10^{-6}$ and the SGCC-4 scheme with both values of D, the rate is not optimal as we lose at least an order of convergence. This phenomenon is more severe for the SGCC-4 scheme where its convergence rate is only roughly that of the SGCC-3 scheme when $D = 10^{-6}$. A possible explanation of this situation is the fact that the limiter Φ_K is inactive in local maximum of the solution as remarked in section 4.7.3. Therefore, the sharper the shape of the solution is, i.e. the smaller D is, the less accurate the numerical methods are.

$\mathcal{N}$	SG		SGCC-1		SGCC-2		SGCC-3		SGCC-4	
	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order
64	4.6e-3		7.4e-4		9.0e-4		5.0e-4		7.0e-4	
96	3.4e-3	0.75	3.3e-4	2.04	3.3e-4	2.52	1.7e-4	3.21	1.7e-4	3.44
144	2.5e-3	0.82	1.4e-4	2.10	9.9e-5	2.94	3.7e-6	3.24	4.1e-5	3.58
216	1.7e-3	0.87	6.0e-5	2.07	3.0e-5	2.94	1.0e-5	3.10	1.2e-5	3.09
324	1.2e-3	0.91	2.6e-5	2.08	9.2e-6	2.91	2.9e-6	3.18	3.2e-6	3.19

Table 4.5: Test 4.3, **with** limiter. L^1-error and convergence order for $D = 10^{-6}$.

In the second case, the numerical results **without** slope limiters, i.e. $\Phi_K = 0$, are grouped in tables 4.7 and 4.8. In this test, it happens that the unlimited reconstruction works fine without generating any oscillations or instabilities, plausibly because the solution is smooth and the grid size is small enough. The interesting remark is that the order of converge for all schemes agrees well

$\mathcal{N}$	SG		SGCC-1		SGCC-2		SGCC-3		SGCC-4	
	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order
64	4.4e-3		6.6e-4		8.4e-4		4.4e-4		6.4e-4	
96	3.2e-3	0.80	2.7e-4	2.20	2.9e-4	2.59	1.1e-4	3.42	1.5e-4	3.62
144	2.2e-3	0.89	1.0e-4	2.37	8.8e-5	2.99	2.6e-5	3.61	3.0e-5	3.93
216	1.5e-3	0.97	3.8e-5	2.47	2.5e-5	3.05	5.6e-6	3.73	6.3e-5	3.81
324	9.8e-4	1.07	1.2e-5	2.73	7.2e-6	3.11	1.1e-6	3.99	1.2e-6	4.11

Table 4.6: Test 4.3, **with** limiter. L^1-error and convergence order for $D = 10^{-4}$.

with the flux consistency analysis and no convergence rate deterioration is observed even when D decreases. This confirms that the use of slope limiters, at least for the one defined in section 4.7.3, has a negative effect on the accuracy of the numerical solution, especially if the latter possesses many local extrema or large-gradient layers.

$\mathcal{N}$	SG		SGCC-1		SGCC-2		SGCC-3		SGCC-4	
	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order
64	4.6e-3		6.5e-4		9.3e-4		2.9e-4		5.7e-4	
96	3.4e-3	0.75	2.8e-4	2.09	3.2e-4	2.67	5.2e-5	4.18	1.1e-4	4.16
144	2.5e-3	0.82	1.2e-4	2.09	9.9e-5	2.85	9.1e-6	4.31	1.6e-5	4.65
216	1.7e-3	0.87	5.2e-5	2.06	3.0e-5	2.94	1.6e-6	4.24	2.3e-6	4.86
324	1.2e-3	0.91	2.3e-5	2.03	9.0e-6	2.98	3.0e-7	4.15	3.0e-7	4.94

Table 4.7: Test 4.3, **without** limiter. L^1-error and convergence order for $D = 10^{-6}$.

Finally, we note that the discretization errors of the SGCC-4 scheme are always larger than those of the SGCC-3 scheme for both cases, with or without limiter. This shows that the convergence order does not absolutely guarantee an accurate numerical solution since we often work with a fixed grid size but not with multiple decreasing grid sizes. The precision of a numerical method relies on many other factors such as the error constant which depends on the derivatives of the solution.

4.9 Closing remarks

In this chapter, we have proposed a generalization of the classical Scharfetter-Gummel flux scheme into a set of high-order methods collectively known as the Scharfetter-Gummel schemes with correction of current (SGCC) that are employed to solve the drift-diffusion equations which appear frequently is gas discharge modeling. The derivation of the new flux(es) was based on a polynomial matching of the particle flux (the continuous flux) in the neighborhood of each cell interface. The degree p of the matching polynomial could be a priori as large as needed.

$\mathcal{N}$	SG		SGCC-1		SGCC-2		SGCC-3		SGCC-4	
	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order	$e_{\Delta x}$	order
64	4.4e-3		5.8e-4		8.6e-4		2.5e-4		5.2e-4	
96	3.2e-3	0.80	2.3e-4	2.25	2.9e-4	2.72	4.4e-5	4.31	9.4e-5	4.23
144	2.2e-3	0.89	9.0e-5	2.33	8.8e-5	2.92	7.0e-6	4.54	1.4e-5	4.73
216	1.5e-3	0.97	3.4e-5	2.43	2.6e-5	3.04	1.1e-6	4.64	1.8e-6	4.97
324	9.8e-4	1.07	1.1e-5	2.66	7.2e-6	3.12	1.5e-7	4.83	2.3e-7	5.11

Table 4.8: Test 4.3, **without** limiter. L^1-error and convergence order for $D = 10^{-4}$.

The SGCC-p flux has been theoretically demonstrated to be consistent with the continuous flux and the consistency order is $p+1$. The new schemes inherit a special property of the standard SG method in which they switch automatically between a high-order upwind scheme for the discretization of the drift flux and a high-order central-difference scheme for the discretization of the diffusion flux, depending on the nature of the flow regime (drift-dominant or diffusion-dominant). Therefore, the interests of the SGCC methods are twofold: on one hand, they allow to gain more precision and on the other hand, they provide a unified framework for the discretization of both the drift and diffusion operators.

The derivation of the novel schemes is intrinsically an exponential interpolation between the particle densities at two points on different sides of a cell interface. It is by all means a one-dimensional process. Therefore, it is natural to extend the SGCC schemes in two-dimensional setting (probably multi-dimensional as well) in the finite volume framework, since the edge-normal fluxes play a central role in the finite volume method. The extension is not straightforward though and we have proposed an interpolation method that allows to evaluate the numerical flux in the normal direction to a cell edge.

The high-order SGCC fluxes are function of the derivatives of the particle density as similar to high-order upwind schemes which use the Taylor expansion of the density to gain more numerical precision. This dependency necessitates thus a polynomial reconstruction of the density from (the approximants of) its cell-averaged values. Moreover, slope limiting techniques are also a priori required to ensure the stability of the simulation since high-order linear schemes usually generate spurious oscillations on the numerical solutions. The reconstruction aspect was addressed by a least-square algorithm and the slope limiting was addressed with the use of TVD limiters for linear reconstruction and GML limiters for quadratic reconstruction in one dimension, and a Barth-Jespersen-inspired limiter on two-dimensional grids.

The workflow presented in this chapter enables the implementation of the SGCC methods in two proto-plasma solvers, one for one-dimensional tests and the other for two-dimensional simulations. The two solvers were tested against some simple drift-diffusion problems and showed adequate convergence results that were coherent with the theoretical consistency order of the SGCC schemes, except in certain cases where the (a priori) limiters were strongly active, such as in drift-dominant

regimes where the density gradients are steep. But as the limiters were turned off, the convergence order regained as expected. The reduction of convergence order could potentially be avoided with a posteriori limiting techniques such as MOOD [36] if one could find a way to detect "trouble" cells where the numerical solution is contaminated with nonphysical oscillations.

In the next chapter, the two proto-plasma solvers are updated to simulate fully-developed plasma discharges. We shall investigate the capability as well as the interests of the high-order SGCC schemes in more complex simulation settings.

4.10 Remarques finales

Dans ce chapitre, nous avons proposé une généralisation du schéma de flux standard de Scharfetter-Gummel en un ensemble de méthodes d'ordre élevé connues collectivement sous le nom des schémas de Scharfetter-Gummel avec la correction du courant (SGCC) qui sont utilisées pour résoudre les équations de dérive-diffusion qui apparaissent fréquemment dans la modélisation des décharges de gaz. La dérivation des nouveaux flux est basée sur une approximation polynomiale du flux de particules (le flux continu) dans le voisinage de chaque interface d'une cellule. Le degré p du polynôme d'approximation peut être a priori aussi grand que nécessaire.

Il a été démontré théoriquement que le flux SGCC-p est consistent avec le flux continu et que l'ordre de consistance est de $p + 1$. Les nouveaux schémas héritent d'une propriété spéciale de la méthode SG standard qui permet de passer automatiquement d'un schéma décentré d'ordre élevé pour la discrétisation du flux de dérive à un schéma de différence centrale d'ordre élevé pour la discrétisation du flux de diffusion, en fonction de la nature du régime d'écoulement (à la dérive dominante ou à la diffusion dominante). L'intérêt des méthodes SGCC est donc double : d'une part, elles permettent de gagner en précision numérique et, d'autre part, elles fournissent un cadre unifié pour la discrétisation des opérateurs de dérive et de diffusion.

La dérivation des nouveaux schémas est intrinsèquement une interpolation exponentielle entre la densité en deux points situés de part et d'autre d'une interface d'une cellule. Il s'agit donc d'un processus unidimensionnel. Par conséquent, il est naturel d'étendre les schémas SGCC dans sur des maillages bidimensionnels (probablement multidimensionnels aussi) dans le cadre de la méthode des volumes finis, puisque les flux normaux de bord jouent un rôle central. L'extension n'est cependant pas évident et nous avons proposé une méthode d'interpolation qui permet d'évaluer le flux numérique dans la direction normale au bord d'une cellule.

Les flux SGCC d'ordre élevé sont fonctions des dérivées de la densité des particules, à l'instar des schémas décentrés d'ordre élevé qui utilisent l'expansion de Taylor de la densité pour gagner en précision numérique. Cette dépendance nécessite donc une reconstruction polynomiale de la densité à partir (des approximants) de ses valeurs moyennes sur les cellules. En outre, des techniques de limitation des pentes sont également nécessaires pour garantir la stabilité de la simulation, étant donné que les schémas linéaires d'ordre élevé entraînent généralement des oscillations parasites sur les solutions numériques. La reconstruction a été traité par un algorithme des moindres carrés et la limitation des pentes a été traitée par les limiteurs de TVD pour la reconstruction linéaire et les

limiteurs de GML pour la reconstruction quadratique en une dimension, et un limiteur inspiré de Barth-Jespersen en deux dimensions.

Le travail présenté dans ce chapitre permet la mise en œuvre des méthodes SGCC dans deux proto-solveurs de plasma, l'un pour les simulations unidimensionnelles et l'autre pour les simulations bidimensionnelles. Les deux proto-solveurs ont été testés sur quelques problèmes simples de dérive-diffusion et ont montré des résultats de convergence adéquats qui étaient cohérents avec l'ordre de consistance théorique des schémas SGCC, sauf dans certains cas où les limiteurs (a priori) étaient fortement actifs, comme dans les régimes dominés par la dérive où les gradients de densité sont raids. Mais lorsque les limiteurs ont été désactivés, l'ordre de convergence s'est rétabli comme prévu. La réduction de l'ordre de convergence pourrait potentiellement être évitée en utilisant des techniques de limitation a posteriori telles que le MOOD [36] si l'on pouvait trouver un moyen de détecter les cellules "problématiques" où la solution numérique est contaminée par des oscillations parasites.

Dans le chapitre suivant, les deux proto-solveurs plasma sont mis à niveau pour simuler des décharges de plasma complètes. Nous étudierons les capacités ainsi que les intérêts que les schémas SGCC d'ordre élevé apportent dans des simulations plus complexes.

Chapter 5

High-order Scharfetter-Gummel Schemes: Application to Simulation of Low-temperature Gas Discharges*

5.0 Aperçu . . . 116
5.1 Overview . . . 116
5.2 Simulations of corona discharge on one-dimensional grids . . . 117
5.2.1 Description of the test case and numerical parameters . . . 117
5.2.2 Definition of steady-state discharge based on electric current . . . 119
5.2.3 Numerical results . . . 120
5.3 Simulation of streamer propagation on two-dimensional cartesian grids . . . 123
5.3.1 Description of the test case and numerical parameters . . . 123
5.3.2 Numerical results with high initial density, direction splitting and slope limiters (SS simulations) . . . 125
5.3.3 Numerical results with high initial density, without direction splitting and without slope limiters (WOL simulations) . . . 131
5.3.4 Numerical results with high initial density, without direction splitting and with slope limiters (WL simulations) . . . 133
5.3.5 Numerical results with low initial density, direction splitting and slope limiters (SSL simulations) . . . 133
5.4 Closing remarks . . . 137
5.5 Remarques finales . . . 139

*parts of this chapter, in altered form, has been published in [144]: *N Tuan Dung, C Besse, and F Rogier. "High-order Scharfetter-Gummel-based schemes and applications to gas discharge modeling". In: Journal of Computational Physics 461 (2022), p. 111196.*

5.0 Aperçu

Ce chapitre poursuit le travail du chapitre 4 en étudiant la capacité de la méthode SGCC, jusqu'à l'ordre 6, dans la simulation de certaines décharges électriques complexes dans l'air, à savoir une décharge couronne fil-fil et la propagation d'un streamer positif. À notre connaissance, il n'existe aucune recherche sur les simulations de décharges dans la littérature où la discrétisation des équations de continuité a une précision supérieure à l'ordre 3. Les résultats numériques de ce chapitre sont obtenus avec nos propres solveurs de plasma.

Nous devons cependant souligner que l'équation de Poisson est résolue avec des schémas du premier ordre tels que la méthode des éléments finis P1-Lagrange dans la section 5.2 et une méthode des volumes finis du premier ordre dans la section 5.3. Par conséquent, le schéma numérique global est du premier ordre. Néanmoins, il semble que l'erreur de discrétisation de l'équation de Poisson ait peu d'influence sur la qualité des solutions numériques des décharges. D'après les résultats des sections 5.2 et 5.3, l'effet de la précision numérique est essentiellement dû aux équations de dérive-diffusion.

Dans la section 5.2, nous fournissons une validation numérique des nouveaux schémas SGCC-1 et SGCC-2 dans des simulations unidimensionnelles d'une décharge couronne fil-fil. Les résultats numériques des schémas SG standard et MUSCL sont également inclus pour comparaison. Dans la section 5.3, nous étudions les schémas SGCC-p, pour $p = 1, 2, 3, 4, 5$, dans des simulations bidimensionnelles de la propagation d'un streamer positif à travers différents cas de test - avec une densité initiale d'électrons élevée ou faible, avec ou sans le splitting directionnel, et avec ou sans limiteurs des pentes. Les résultats numériques des schémas SG et MUSCL sont parfois inclus pour comparaison. Nous nous concentrons également sur la pertinence physique des solutions numériques, la comparaison entre les schémas de différents ordres de précision ainsi que l'efficacité du calcul au niveau de l'équilibre entre la précision numérique et le temps de calcul.

5.1 Overview

This chapter continues the work of chapter 4 by investigating the capability of the SGCC method, up to the sixth-order scheme, in the simulation of some complex electric discharges in air, namely a wire-to-wire corona discharge and the propagation of a positive streamer. As far as we are aware, there had not existed any research on atmospheric gas discharge simulations in the literature where the discretization of the specie continuity equations has a precision higher than third-order. The numerical results of this chapter are obtained with self-implemented codes.

We have to stress, though, that the Poisson equation is solved with first-order schemes such as the P1-Lagrange finite element method in section 5.2 and a first-order finite volume method in section 5.3. Therefore, the global numerical scheme is first-order. Nevertheless, it seems that the discretization error on the Poisson equation has little influence on the quality of numerical solutions of the discharges. Based on the results of sections 5.2 and 5.3, the effect of numerical precision is essentially due to the drift-diffusion equations.

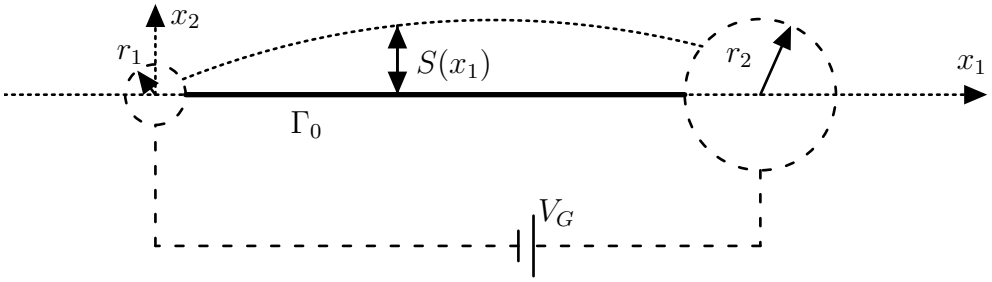


Figure 5.1: Schematics of the wire-to-wire actuator

In section 5.2, we provide numerical validation of the new SGCC-1 and SGCC-2 schemes in one-dimensional simulations of a wire-to-wire corona discharge. Numerical results of the standard SG and MUSCL schemes are also included for comparison. In section 5.3, we study the SGCC-p schemes, for $p = 1, 2, 3, 4, 5$, in two-dimensional simulations of the propagation of a positive streamer through different test cases - with high or low electron initial density, with or without directional splitting, and with or without slope limiters. Numerical results of the standard SG and MUSCL schemes are sometimes included for comparison. We also focus on the physics relevance of the numerical solutions, the comparison between schemes of different precision orders as well as the computation efficiency in terms of trade-off between numerical accuracy and CPU time.

5.2 Simulations of corona discharge on one-dimensional grids

5.2.1 Description of the test case and numerical parameters

The sketch of the electric actuator is illustrated in 5.1. It is composed of a small wire of radius $r_1 = 0.35$ mm and a bigger wire $r_2 = 1$ mm, parallel to each other and $d = 40$ mm apart. Each wire is $L^{\text{wire}} = 16$ cm long and there is no presence of other dielectric materials other than air ($\varepsilon_r = 1$). We apply a voltage $V_G = 40$ kV on the smaller wire (the stressed electrode) and the larger wire is grounded.

If we suppose that the plasma is uniform along the wires then the three-dimensional problem is transformed into a two-dimensional one. In order to further transform it into a **one-dimensional** problem, a quasi-2D model was developed in [106] in which we suppose that the plasma is uniformly contained in a "bumpy" layer between the electrodes, bounded by the segment $\Gamma_0 \equiv \{ (x_1, 0) \mid x_1 \in (r_1, r_1 + d) \}$ from the bottom and the graph of a function $S(x_1) > 0$ from the top (see fig. 5.1). $S(x_1)$ is computed from the relation

$$\frac{\mathrm{d}}{\mathrm{d}x_1} \left(\boldsymbol{E}_1^{\text{ext}}(x_1, 0) S(x_1) \right) = 0, \tag{5.1}$$

where $\boldsymbol{E}^{\text{ext}}(x_1, x_2) = \left(\boldsymbol{E}_1^{\text{ext}}(x_1, x_2), \boldsymbol{E}_2^{\text{ext}}(x_1, x_2) \right)^t = -\nabla \phi^{\text{ext}}(x_1, x_2)$ is the analytic expression of

the electrostatic field generated by the voltage between the electrodes, i.e. the solution of

$$\begin{cases} -\nabla \cdot \left(\varepsilon_0 \nabla \phi^{\text{ext}}(x_1, x_2)\right) = 0, & \\ \phi = V_G, & \text{on} \quad S_1, \\ \phi = 0, & \text{on} \quad S_2, \end{cases} \tag{5.2}$$

where S_1, S_2 are resp. the smaller and larger wire surfaces. In this wire-to-wire configuration, the expression of $\phi^{\text{ext}}(x_1, x_2)$ is given in [92] in the following way,

$$\phi^{\text{ext}}(x_1, x_2) = V_G - \frac{V_G}{2\ln\left(\frac{A}{r_1}\right)} \ln\left(\frac{(x_1 - Br_1)^2 + x_2^2}{(Bx_1 - r_1)^2 + (Bx_2)^2}\right), \tag{5.3}$$

with

$$A = \frac{CD - r_1^2 - \sqrt{(C^2 - r_1^2)\,(D^2 - r_1^2)}}{C - D} > 0, \tag{5.4}$$

$$B = \frac{CD + r_1^2 + \sqrt{(C^2 - r_1^2)\,(D^2 - r_1^2)}}{r_1(C + D)} > 0, \tag{5.5}$$

$$C = r_1 + d, \quad D = r_1 + d + 2r_2. \tag{5.6}$$

In the next parts, we simply write x instead of x_1. Since the elementary volume unit in this quasi-2D setting is equal to $S(x)L^{\text{wire}}\mathrm{d}x$, the divergence of a field $F(x)$ reads as $\nabla \cdot F = \dfrac{1}{S}\dfrac{\mathrm{d}(SF)}{\mathrm{d}x}$. Therefore, eq. (5.1) is just a rewriting of the first equation of (5.2). By some algebraic manipulations, from eqs. (5.1) and (5.3) we deduce that

$$S(x) = G\,(Bx - r_1)\,(Br_1 - x)\,,$$

where G is a positive constant. Moreover, a straightforward computation shows that S reaches its maximum value $S^{\max}$ at $x^{\max} = \dfrac{(B^2+1)r_1}{2B}$ and $S^{\max} = S\left(x^{\max}\right) = G\dfrac{(B^2-1)^2\, r_1^2}{4B}$. Therefore,

$$S(x) = \frac{4BS^{\max}}{(B^2-1)^2\, r_1^2}\,(Bx - r_1)\,(Br_1 - x)\,.$$

In the simulations of this section, we use the prescribed value $S^{\max} = 5\ \mathrm{mm}$. The two-dimensional problem is thus transformed in to a one-dimensional one, where Γ_0 serves as the computation domain.

A constant floor density ψ is imposed on the electron density (see section 2.1.1). The justification of this practice will be presented in details in chapter 6. In the following simulations, we set $\psi = 10^9$ m^{-3}. This value is also used for the initial data: $n_s^0 = 10^9\ \mathrm{m}^{-3}$.

For this one-dimensional test, we consider the kinetic scheme of example 2.1 ($\mathfrak{S} = \{\, e, p, n \,\}$) without photoionization. Additionally, we assume that $D_p = D_n = 0$ for simplicity. Using the variable change $v_s(t,x) \equiv S(x)n_s(t,x)$, the one-dimensional discharge model now reads as

$$\begin{cases} \partial_t \boldsymbol{V} + \partial_x \boldsymbol{F}(\boldsymbol{V}) = \boldsymbol{S}(\boldsymbol{V}), \\ -\partial_x(\varepsilon_0 S \partial_x \phi) = S\rho, \end{cases}$$

where $\boldsymbol{V} = \begin{pmatrix} v_e \\ v_p \\ v_n \end{pmatrix}$, $\boldsymbol{F}(\boldsymbol{V}) = \begin{pmatrix} u_e^S v_e - D_e \partial_x v_e \\ u_p v_p \\ u_n v_n \end{pmatrix}$, $u_e^S = -\mu_e \boldsymbol{E} + D_e \dfrac{S'}{S}$ and $\boldsymbol{S}(\boldsymbol{V}) = \begin{pmatrix} S_e \\ S_p \\ S_n \end{pmatrix}$ with

$$\begin{cases} S_e = (\alpha - \eta) v_e - \dfrac{k_{ep}}{S} v_e v_p, \\ S_p = \alpha v_e - \dfrac{k_{ep}}{S} v_e v_p - \dfrac{k_{np}}{S} v_n v_p, \\ S_n = \eta v_e - \dfrac{k_{np}}{S} v_n v_p. \end{cases}$$

The computation domain $\Gamma_0 = (r_1, r_1 + d)$ is partitioned into $\mathcal{N}$ non-overlapping intervals (cells) such that the left interface of the first cell and the right interface of the last cell coincide resp. with the anode surface and the cathode surface. The cell sizes satisfy a refinement criterion in the following way,

$$\Delta x_i = aS\left(r_1 + \left(i - \frac{1}{2}\right)\frac{d}{\mathcal{N}}\right), \quad i = 1, \ldots, \mathcal{N},$$

where a is a positive constant. Therefore, since $\displaystyle\sum_{i=1}^{\mathcal{N}} \Delta x_i = d$, we have

$$\Delta x_i = d \frac{S\left(r_1 + \left(i - \frac{1}{2}\right)\frac{d}{\mathcal{N}}\right)}{\sum_{l=1}^{\mathcal{N}} S\left(r_1 + \left(l - \frac{1}{2}\right)\frac{d}{\mathcal{N}}\right)}.$$

This refinement technique is employed to capture correctly the discharge dynamics in the electrode sheaths and relax the cell size in the middle region (outside the electrode sheaths) where no ionization process takes place and the charged species only drift along electric field. Indeed, the cell size Δx_i becomes small whenever the field strength is large as $S(x)$ is proportional to the inverse of the latter.

Since ions weigh much heavier than electrons, their drifting time is also much longer. As we use explicit time integration in this section, the CFL conditions of species are disparate. Furthermore, the charge density in the corona discharge is not extremely high (around 10^{15}-10^{18} m^{-3}), hence the dielectric relaxation time is much larger than the CFL conditions. Therefore, a sub-cycling strategy [46] is employed to boost the CPU time. The second-order Runge-Kutta-Heun method (see example 3.2) is used for time integration. The CFL condition satisfies eq. (4.46) with $\mathfrak{C} = 0.49$. The P_1-Lagrange finite element method is employed to solve the Poisson equation. The second-order **MUSCL scheme** is used to solve the **ion transport equations** and the **SGCC-1, SGCC-2, MUSCL and SG schemes** are used to solve the **electron continuity equation**.

Finally, the simulation time of the discharge is $T = 4$ ms. All the geometric and numerical parameters are gathered in table 5.1.

5.2.2 Definition of steady-state discharge based on electric current

We define the steady state of a corona discharge as when the value of the circuit current I (see eq. (2.11)) does not change more than $10^{-4}\%$ between two successive time levels. More precisely,

r_1	0.35 mm	V_G	40 kV	ψ	10^9 m^{-3}
r_2	1 mm	$S^{\max}$	5 mm	n_s^0	10^9 m^{-3}
L^{wire}	16 cm	$\mathcal{N}$	400, 800, 3200	T	4 ms
d	40 mm	$\mathfrak{C}$	0.49		

Table 5.1: Parameters for one-dimensional wire-to-wire discharge simulations

if $I(t^l), I(t^{l+1})$ are the computed values of the electric current at resp. t^l, t^{l+1}, the steady state is reached when

$$\left|\frac{I(t^{l+1}) - I(t^l)}{I(t^l)}\right| < 10^{-6}.$$

For this wire-to-wire test case, it happens that the steady state exists, or at least could be observed numerically. The simulation time $T = 4$ ms is chosen since the steady state is reached at this instant in all the simulations of this section.

5.2.3 Numerical results

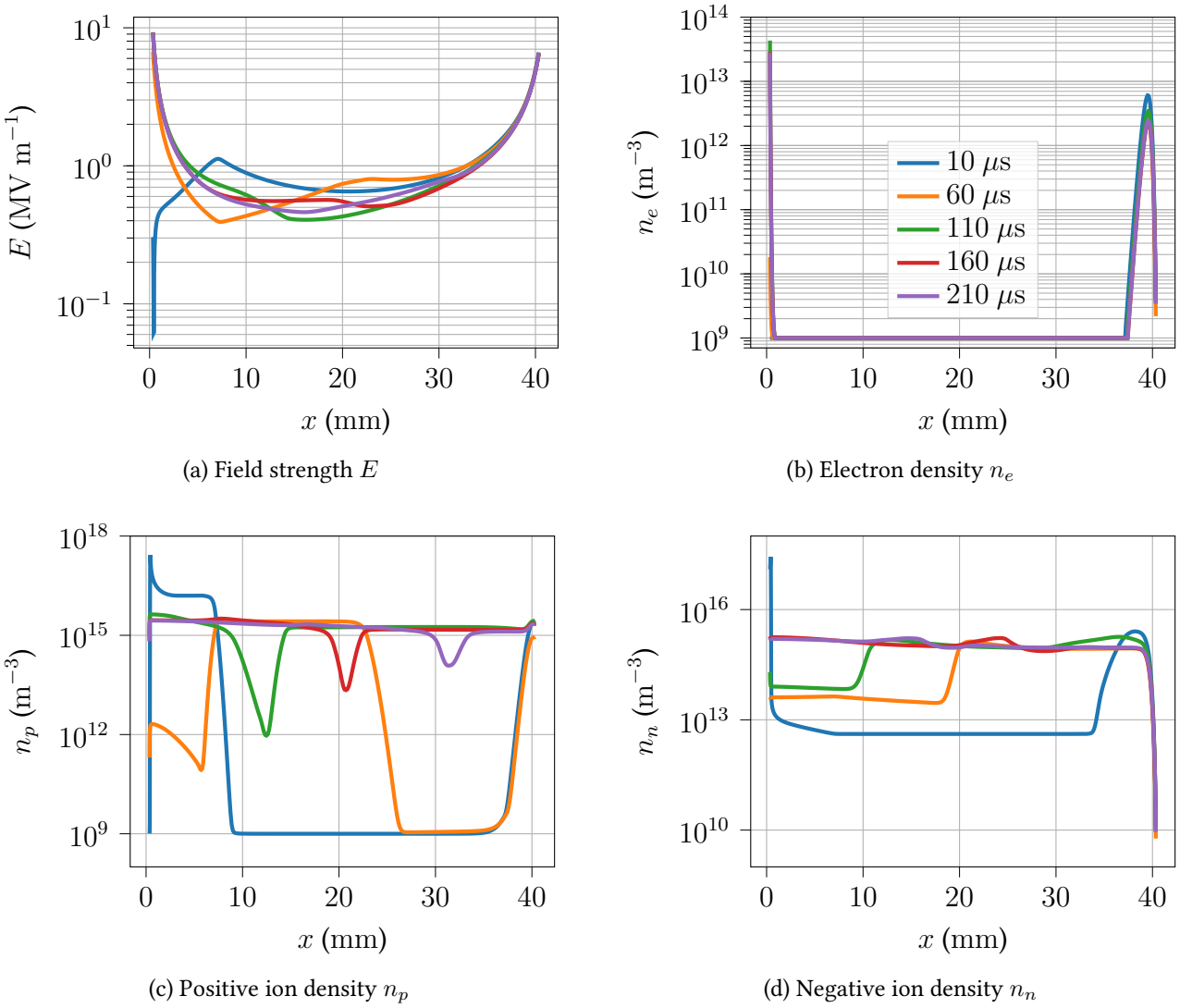
We employ the notation ${}^{\mathcal{N}}\mathrm{S}$ to refer to a simulation with the scheme S that is used to solve the electron continuity equation, on the $\mathcal{N}$-cell grid. For example, the simulation with SGCC-2 on the 400-cell grid is denoted as ^{400}SGCC-2.

At $t = 0$, a potential $V_G = 40$ kV is applied on the anode and abruptly causes a breakdown. In figs. 5.2a to 5.2d, we display the evolution of the field strength and the density of charged species within the first 210 μs of the discharge, obtained by the ^{400}SGCC-2 simulation. The field strength at the anode surface (near $x = 0$) is over 10 MV m^{-1} and at the cathode surface (near $x = 40$ mm) is about 4 MV m^{-1} which surpasses the critical ignition value around 3.75 MV m^{-1}, thus we expect ionization in the vicinity of both electrodes but mainly at the anode.

Within the first 10 μs, multiple electron avalanches occur ahead the anode, producing a high-density ion cloud with $n_p = n_n = 2 \times 10^{17}$ m^{-3}. Meanwhile, the electrons are all absorbed by the anode. The field drops abruptly because of the electric screening of the anode due to a large concentration of negative charges. At $t = 1$ μs, we observe a current surge of about 200 μA (see fig. 5.3a) due to steep variation of the field as well as absorption of electrons by the anode. At $t = 10$ μs, the current drops significantly to about 25 μA.

After $t = 10$ μs we enter the ion collection phase [109] where the ions drift towards the opposite-sign electrode. As the negative-charge cloud is gradually absorbed, the screening effect around the anode fades away and the field strength gradually increases.

From $t = 60$ μs and on, the field is only slightly distorted in the middle region due to ion drifting while electron impact ionization takes place stably in the anode sheath, forming a positive corona. The newly produced positive ions follow the previously departing ions to the cathode (see fig. 5.2c) where a small "sink" separates the two ion clouds that will slightly decreases the current at $t = 250$ μs (see fig. 5.3a), which corresponds roughly to the instant when the sink arrives at the cathode.



(a) Field strength E

(b) Electron density n_e

(c) Positive ion density n_p

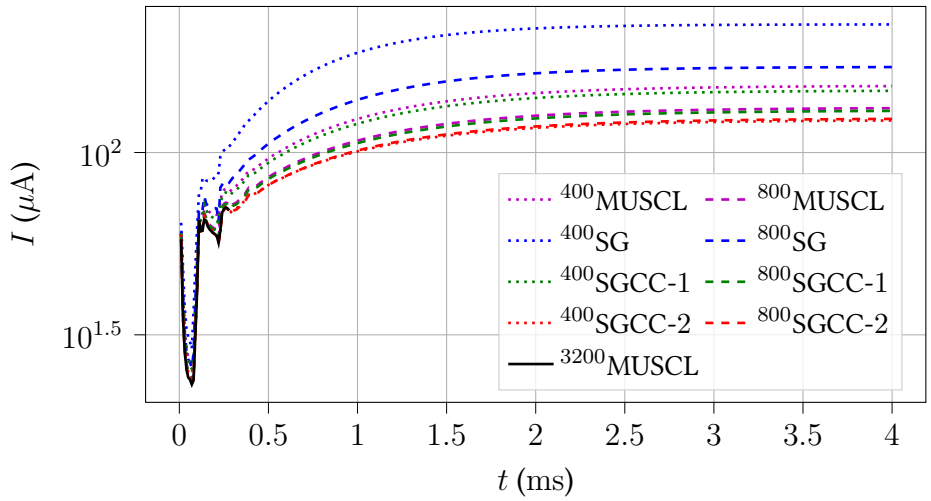
(d) Negative ion density n_n

Figure 5.2: Evolution of the field strength and specie densities of the ^{400}SGCC-2 simulation

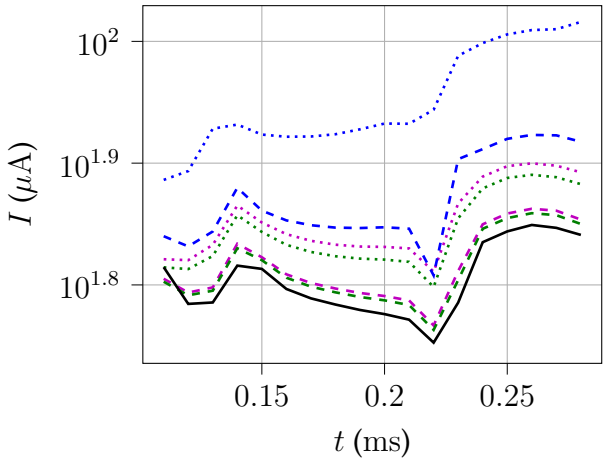
Since ionization is no longer disrupted, the electron density within the anode sheath gradually increases (see fig. 5.2b) and so does the circuit current I. On the other side, weak ionization and electron secondary emission also form a cathode sheath. A population of negative ions appear near the cathode at the beginning of the simulation due to high concentration of electrons, then drift towards the anode. At the same time, electron attachment gradually increases the negative ion density in front of the ion wave (see fig. 5.2d). From about $t = 500$ μs and on, the ion populations in the middle region slightly increase until they reach a steady state.

In figs. 5.3a and 5.3c, we compare the numerical currents obtained by different flux schemes applied to the discretization of the electron continuity equation. We mainly present the results on the 400 and 800-cell grids, since the CPU time of a simulation on the 3200-cell grid would take months to finish. Therefore, we only present the result of ^{3200}MUSCL up to $t = 290$ μs as a benchmark for other cases.

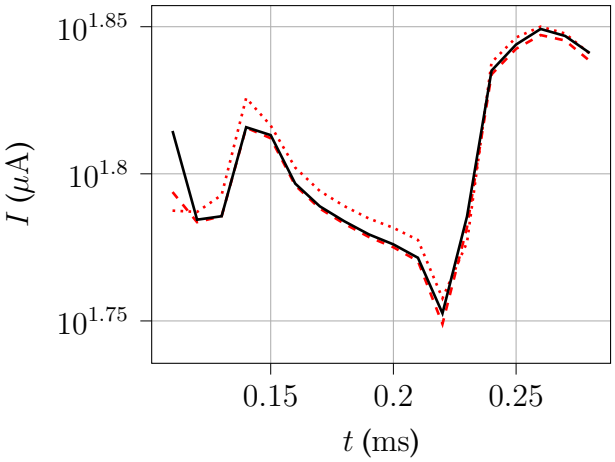
We observe a tendency in which the current decreases as the grid is refined. However, the standard SG scheme overestimates tremendously the current intensity even on the 800-cell grid,



(a) $t = 0$-4 ms



(b) $t = 0.11$-0.29 ms



(c) $t = 0.11$-0.29 ms

Figure 5.3: Comparison of the circuit current I of different schemes and grid sizes, on 0-4 ms (left) and on 110-290 μs (right)

comparing to the benchmark curve (in black) as well as other second-order schemes on the 400-cell grid, which demonstrates its low numerical precision. The SGCC-1 scheme produces results very close to the MUSCL scheme since they have the same convergence order. Interestingly, SGCC-2 gives virtually the same current on both 400-cell and 800-cell grids which also agree well with the benchmark curve before $t = 290$ μs (see fig. 5.3c).

Figure 5.3c displays a zoom on the currents from $t = 110$ μs to 290 μs. Although MUSCL and SGCC-1 produce fairly good results comparing to the standard SG scheme, on the 400-cell grid the maximal discrepancy between the corresponding currents and the benchmark current is about 10 μA, or 15% of the benchmark current. On the contrary, the result of ^{400}SGCC-2 is exceptionally close to that of ^{3200}MUSCL and even better than the results of ^{800}MUSCL and ^{800}SGCC-1. This shows the advantage of high-order schemes since they allow simulations to be well performed on coarser grids.

5.3 Simulation of streamer propagation on two-dimensional cartesian grids

5.3.1 Description of the test case and numerical parameters

The second test case taken from [7] is the propagation of a positive streamer. This reference provides a comparison between numerical solutions of six simulation codes from six different research groups from China and Europe, and serves as a benchmark document for code verification in streamer simulations. Based on the quality of their performance, we use the results of the groups **CWI** (the Netherlands) [141] and **FR** (France) [48] as benchmarks to compare with our simulation results. CWI used a 5-point central difference scheme whereas FR used the Fourier transform and a first-order finite volume method to solve the Poisson equation. Both teams used the MUSCL scheme to solve the electron continuity equation in which the CWI team used Koren's limiter [78] and the FR team used the superbee limiter. The CWI group employed in particular an adaptive mesh refinement technique in their plasma solver, allowing the CPU time to be cut short to matter of minutes.

The cylindrical coordinates are used and the discharge is assumed to be symmetric in the azimuthal direction. Therefore, the three-dimensional problem reduces to a two-dimensional one on radial and axial coordinates (r, z). The computation domain Ω is the square $(0, d)^2$ with $d = 1.25$ cm, hence the left boundary of Ω aligns with the symmetric axis $r = 0$ (see fig. 5.4a). Two planar electrodes are placed resp. at $z = 0$ and $z = d$. A potential $V_G = 18.75$ kV is applied on the upper electrode while the lower one is grounded.

The numerical model considered here only takes into account two species: electrons (e) and positive ions (p). The simulation time T is 13-24 ns, so the drift and diffusion of ions can be neglected, i.e. $\mu_p = d_p = 0$, since they are much heavier than the electrons and considered unmoved on this short time scale. Only two plasma-chemical reactions are considered in this test case: electron impact ionization $e + N \to 2e + p$ and electron attachment $e + N \to n$. Thus, the discharge model reads as follows,

$$\begin{cases} \partial_t \boldsymbol{U}(r,z) + \nabla \cdot \boldsymbol{F}(\boldsymbol{E}(r,z), \widehat{E}(r,z), \boldsymbol{U}(r,z)) = \boldsymbol{S}(\widehat{E}(r,z), \boldsymbol{U}(r,z)), \\ -\nabla \cdot (\varepsilon_0 \nabla \phi(r,z)) = \rho(r,z), \\ \rho(r,z) = q(n_p(r,z) - n_e(r,z)), \quad \boldsymbol{E}(r,z) = -\nabla \phi(r,z), \end{cases}$$

with $\boldsymbol{U} = (n_e, n_p)^t$, $\boldsymbol{F} = (-\mu_e \mathbf{E} n_e - D_e \nabla n_e, 0)^t$ and $\boldsymbol{S} = ((\alpha - \eta) N n_e, (\alpha - \eta) N n_e)^t$. The coefficients μ_e, D_e, α and η depend only on the reduced field strength $\widehat{E} = E/N$ with $N = 2.414 \times 10^{25}$ m^{-3} and are given analytically in [7].

The Dirichlet conditions $\phi = V_G$, $\phi = 0$ as well as the homogeneous Neumann condition are applied for the potential resp. on the top, bottom and right boundaries of Ω. For the electron density, the homogeneous Neumann condition is applied on all boundaries except the left one.

In order to initiate the formation of a streamer, an initial population of positive ions is injected

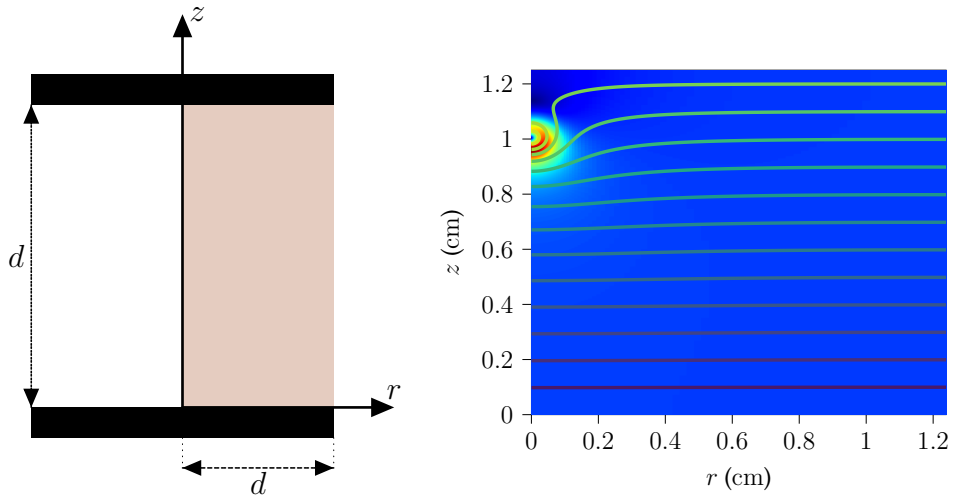


Figure 5.4: Geometry of the discharge with the computation domain in brown (left) and initial electric field with equipotential lines (right)

between the gap in the following way,

$$n_p^0(r,z) = \max\left(N^0 \exp\left(-\frac{r^2+(z-z_0)^2}{\sigma^2}\right), n^0\right), \tag{5.7}$$

with $N^0 = 5\times10^{18}$ m^{-3}, $\sigma = 0.4$ mm, $z_0 = 1$ cm and $n^0 = 10^9$ or 10^{13} m^{-3}. For the electrons, $n_e^0(r,z) = n^0$. These initial data generate a high electric field strength (see fig. 5.4b) that facilitates the primary avalanches.

In this section, the grids are partitioned into rectangular cells so that direction splitting of spatial derivative operators is straightforward. In sections 5.3.2 and 5.3.5, we use the Strang splitting [136] to solve the electron continuity equation. The advantage of this method is that it allows us to deal with a two-dimensional problem with the numerical methods inherited from section 5.2 to solve one-dimensional problems, so it is quite simple to implement. Then in sections 4.6 and 4.7, we study the numerical schemes in "full" two dimensions (i.e. without direction splitting).

The grid consists of $\mathcal{N}_r \times \mathcal{N}_z$ cells, with $\mathcal{N}_r$ cells on the r-direction and $\mathcal{N}_z$ on the z-direction. In the axial direction, the grid size is always constant whereas in the radial direction, it is constant from $r = 0$ to $R_0 = 0.6$ mm and gradually becomes larger for $r > 0.6$ mm. The minimum grid size in the r-direction $\Delta r^{\min}$ is chosen close to the grid size in the z-direction Δz. Furthermore, $\mathcal{N}_z$ is always a power of 2, so that the Poisson's equation could be tensorized by the Fourier transform in the z-direction. In the radial direction, the resulting one-dimensional elliptic equation corresponding to each frequency is solved with a first-order finite volume scheme (see section 3.3).

The third-order SSP-Runge-Kutta method is used for time discretization. The CFL condition satisfies eq. (4.46) with $\mathfrak{C} = 0.4$. All geometric and numerical parameters are gathered in table 5.2.

The numerical results are presented as follows. In section 5.3.2, the simulations are conducted with $n^0 = 10^{13}$ m^{-3} and Strang splitting combining with the SG, MUSCL, SGCC-1 or SGCC-2 for flux approximation[1]. The superbee limiter is used in the MUSCL as well as SGCC-1 schemes and

[1] referred to as "SS" simulations for Strang Splitting

d	1.25 cm	$\mathcal{N}_z$	2^{12}, 2^{13}, 2^{14}	n_p^0	eq. (5.7)
V_G	18.75 kV	$\mathcal{N}_r$	262, 400, 600	n_e^0	10^9, 10^{13} m^{-3}
T	13 to 24 ns	$\Delta r^{\min}$	3, 1.5, 0.8 μm		
R_0	0.6 mm	$\mathfrak{C}$	0.4		

Table 5.2: Parameters of the streamer propagation simulations

the GML limiter is used in the SGCC-2 scheme. In section 5.3.3[2], we solve the electron continuity equation without direction splitting and without slope limiters, for $n^0 = 10^{13}$ m^{-3}. Section 5.3.3[3] is almost identical to section 5.3.4 but with the Barth-Jespersen limiter introduced in section 4.7.3. Finally, Section 5.3.5[4] is identical to section 5.3.2 but for $n^0 = 10^9$ m^{-3}.

Throughout these sections, we employ the notation ${}^{\Delta r^{\min}}$S to refer to a simulation with the scheme S on the grid having a minimum size $\Delta r^{\min}$ (μm). For example, the simulation with SGCC-2 on the 3 μm-grid is denoted as ^{3}SGCC-2.

5.3.2 Numerical results with high initial density, direction splitting and slope limiters (SS simulations)

In fig. 5.5, we display the evolution of the field strength $E = |\mathbf{E}|$, obtained by the ^{3}SGCC-2 simulation up to the simulation time $T = 16$ ns. What can be observed is that the initial Gaussian distribution of positive ions enhances the electric field and transforms into a streamer within the first 4 ns, which then propagates in the direction of the cathode (the lower electrode). The space charge in the streamer head is positive which affirms that this is a positive streamer and the field strength in this region exceeds 14 MV m^{-1} after $t = 4$ ns (see fig. 5.6b) which is well beyond the ignition field strength in air around 2.5 MV m^{-1}. It is the electron avalanches taking place in the streamer head that play a decisive role of the streamer advancement [126]. On the way to the cathode, the streamer leaves behind a very weak-field column, known as streamer body, manifested by a dark-colored trace in fig. 5.5. This is the quasi-neutral region consisting of free electrons and positive ions that form a plasma conducting channel.

Figure 5.6b shows the axial component E_z of $\boldsymbol{E}$ on the symmetry axis in dotted lines as well as the space charge ρ in solid lines. A large-gradient layer of E_z propagates along the axis with a narrow but highly concentrated positive space charge distribution. In a cathode-directed streamer, this positive-charge acts as a beacon that attracts and accelerates free electrons from outside the streamer body so that the incident electrons have enough energy to ionize the gas in front of the streamer head. Electrons produced from the avalanches then move inwards the streamer body, creating a deficit of negative charges downstream of the streamer head. As a result, one could observe that the streamer head advances progressively towards the cathode. In fig. 5.6a, the evolution of electron density n_e on the axis $r = 0$ is shown. Within the first 4 ns, the density increases exponentially until

[2] referred to as "WOL" for WithOut Limiters
[3] referred to as "WL" for With Limiters
[4] referred to as "SSL" for Strang Splitting with Low n^0

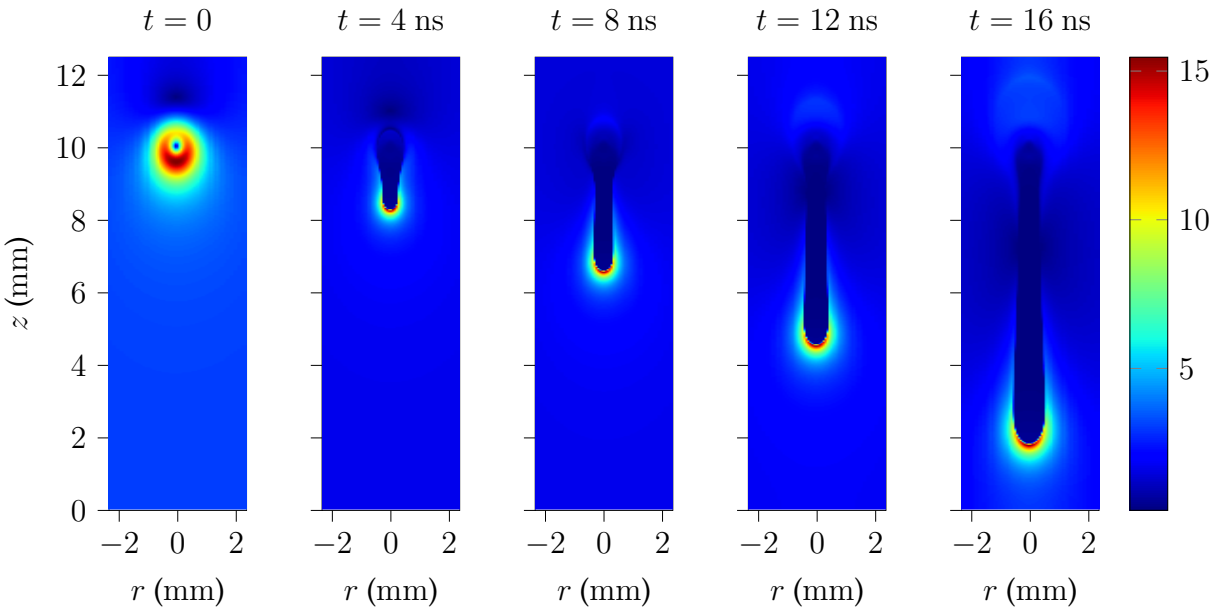


Figure 5.5: (SS) Evolution in time of the electric field strength E (MV m^{-1}) of ^{3}SGCC-2. Visualization of the field is symmetrized by flipping on the axis $r = 0$ and only shown up to $r = 2$ mm.

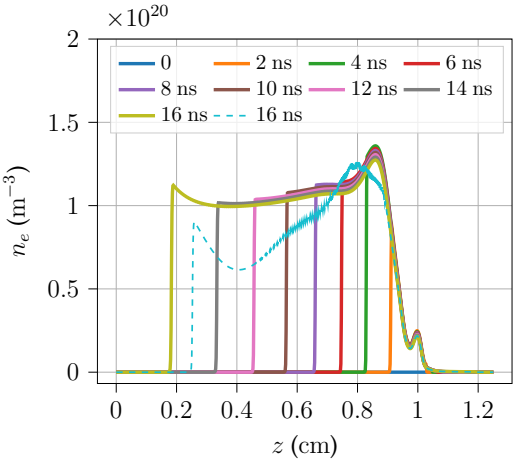
it reaches about 1.4×10^{20} m^{-3}. This phase corresponds to the formation phase of the streamer. Beyond $t = 4$ ns, the electron density is maintained between 10^{20} to 1.2×10^{20} m^{-3} in the streamer body; this is the stable propagation phase.

It is interesting to note that when the propagation is stable, the peak space charge $\rho^{\max}$ diminishes in time but the width of the streamer head, denoted as R^{ρ}, increases. R^{ρ} is measured as the distance between two point z_1, z_2 on the axis such that $\rho_{|z_1|} = \rho_{|z_2|} = 1\%\rho^{\max}$. Moreover, the maximum field strength on the symmetry axis, $E^{\max}(t) \equiv \max_{z} \left|E_{z|r=0}(t, \boldsymbol{x})\right|$, which locates in the streamer head is almost independent of time. This indicates that charge concentration in the layer should be more or less conserved in time. This conjecture was put forth and numerically verified in [85]. It was postulated that the maximum field strength and the peak space charge were related in the following way,

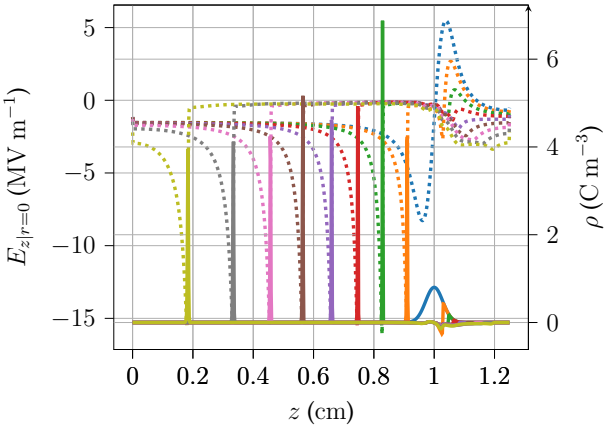
$$E^{\max} \approx \frac{\rho^{\max} R^{\rho}}{3\varepsilon_0} - E_L,$$

for some (almost constant) field strength E_L.

In table 5.3, we list the values of $E^{\max}$, $\rho^{\max}$ and R^{ρ} at different times t from the $^{1.5}$SGCC-1, ^{3}SGCC-2 and $^{1.5}$SGCC-2 simulations. We exclude the ^{3}SGCC-1 simulation since the electron density in this case is contaminated with seemingly nonphysical oscillations (see fig. 5.6a). The same phenomenon is observed in the ^{3}MUSCL simulation. Since these two simulations did carry on until they finished, we suspect that the nature of these oscillations is not linked to the numerical schemes but to the coarse grid size ($\Delta r^{\min} = 3$ μm) that failed to capture correctly the ionization process and/or limit effectively the numerical diffusion. The data in table 5.3 show that E_L is around 2 MV m^{-1} and the variations in time of $\rho^{\max} R^{\rho}$ and $E^{\max}$ are always less than 10%, which agree with the observations in [85].



(a) Electron density n_e on the symmetry axis



(b) Axial component E_z of the electric field (dotted lines, left axis) and space charge ρ (solid lines, right axis) on the symmetry axis $r = 0$

Figure 5.6: (SS) Numerical results of ^{3}SGCC-2 except the dashed line on the left figure which is obtained by ^{3}SGCC-1

	$^{1.5}$SGCC-1				^{3}SGCC-2				$^{1.5}$SGCC-2			
	4 ns	8 ns	12 ns	16 ns	4 ns	8 ns	12 ns	16 ns	4 ns	8 ns	12 ns	16 ns
R^{ρ}	63	75	95	122	61	73	89	116	61	72	93	124
$\rho^{\max}$	7.7	6.2	4.8	3.9	7.5	6.1	4.9	4	8.2	6.5	4.9	4
$\frac{R^{\rho} \times \rho^{\max}}{3\varepsilon_0}$	18.1	17.4	17.2	17.9	17.3	16.9	16.3	17.5	18.7	17.4	17.2	18.4
$E^{\max}$	16.3	15.6	15.1	15.6	16.1	15.4	15	15.5	16.7	15.8	15.3	15.9
E_L	1.8	1.8	2.1	2.3	1.2	1.5	1.3	2	2	1.6	1.9	2.5

Table 5.3: (SS) Evolution in time of $\rho^{\max}$ (C m^{-3}), R^{ρ} (μm) and $E^{\max}$ (MV m^{-1}) obtained with $^{1.5}$SGCC-1, ^{3}SGCC-2 and $^{1.5}$SGCC-2

We now compare the new SGCC-1 and SGCC-2 schemes with other existing methods, namely the standard SG and MUSCL schemes. Figure 5.7 show resp. the streamer length $L(t)$ as well as the maximum field strength $E^{\max}(t)$ as a function of $L(t)$. The streamer length is defined as

$$L(t) = z_0 - z^{\max}(t),$$

where $z^{\max}(t) \equiv \underset{z}{\operatorname{argmax}} \left|E_{z|r=0}(t)\right|$. The simulations that we use as benchmarks from the CWI and FR groups were performed on the 0.8 μm-grid, and are denoted subsequently as $^{0.8}$CWI and $^{0.8}$FR. Comparing to these benchmark results, the relative error on the streamer length generated by any of our simulations is computed in the following way,

$$\mathcal{E}(t) = \frac{\left|L(t) - L^{\text{benchmark}}(t)\right|}{L^{\text{benchmark}}(t)}, \tag{5.8}$$

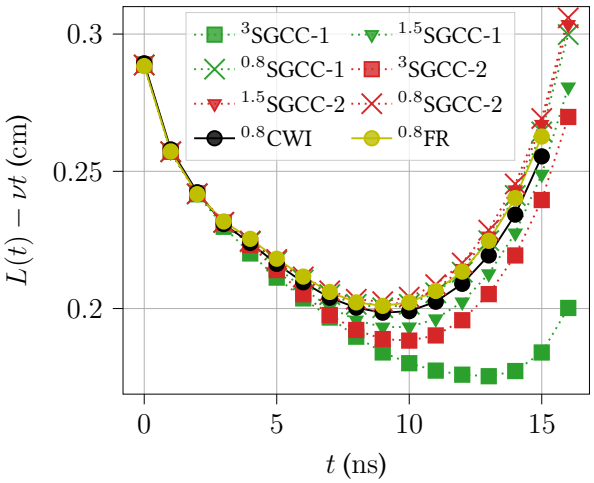
where $L^{\text{benchmark}}(t)$ is the streamer length produced by one of the benchmark simulations, i.e. $^{0.8}$CWI or $^{0.8}$FR. The relative errors for the SGCC-1 and SGCC-2 schemes are presented in fig. 5.8.

At first, we remark from fig. 5.7 that the SG scheme overestimates both the streamer length and the maximum field strength and since the streamer propagates too fast, it reaches the cathode within only 13 to 15 ns comparing to 16 ns in simulations with other schemes. Even on the 0.8 μm-grid, the blue curves of the SG schemes deviate visibly too much from the benchmarks. Similar overestimation behavior of the SG scheme was also observed in [29] which highlights the lack of numerical precision of this method.

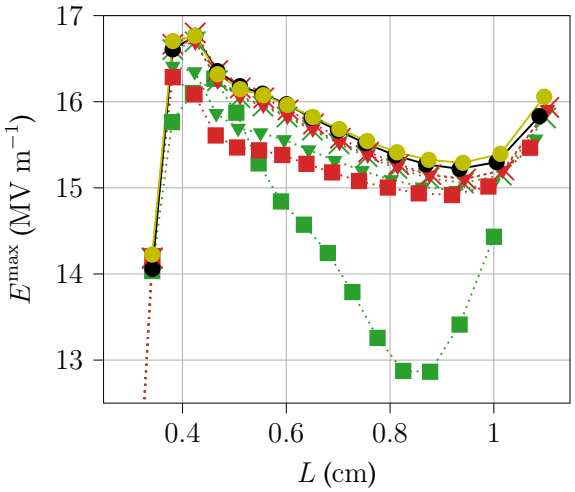
The MUSCL and SGCC-1 schemes in the other hand produce much more accurate results. On the 3 μm-grid, the largest relative error generated by both schemes is less than 10% with respect to both benchmarks, while on the 0.8 μm-grid they are even less than 1% (see fig. 5.8). We remark that on the 3 μm-grid, both schemes produce $E^{\max}$ that is not quasi-constant in time as speculated by [85]. We suspect that this might be linked to the fact that numerical results on the 3 μm-grid of these schemes are contaminated with nonphysical oscillations (see again fig. 5.6a). Therefore, it is recommended to use better-resolution grid to obtain high-quality results. But in general, both second-order schemes show better accuracy than the SG scheme and the difference between their numerical results and the benchmarks decreases with the grid size (see figs. 5.7 and 5.8).

The final method, SGCC-2, show finest results with relative errors never exceeding 3% (see fig. 5.8). Moreover, in terms of maximum field strength, we have seen in table 5.3 and now from fig. 5.7 that $E^{\max}$ is quite steady, even on the 3 μm-grid, during the stable propagation phase (corresponding to $t > 4$ ns or $L(t) > 0.4$ cm). Nevertheless, we remark that the relative error of $^{0.8}$SGCC-2 turns out to be almost always larger than $^{1.5}$SGCC-2 with respect to both benchmarks - an undesirable result in any mesh convergence test, but the relative errors almost never exceed 1%. This might be explained by the fact that *(i)* the benchmarks used different discretizations of the Poisson equation, *(ii)* different approximations of coefficients on cell edges, *(iii)* the low-order approximation of the Poisson equation, etc. or *(iv)* because the flux schemes used in the benchmarks are second-order, therefore their solutions are less accurate than those of the third-order SGCC-2 scheme.

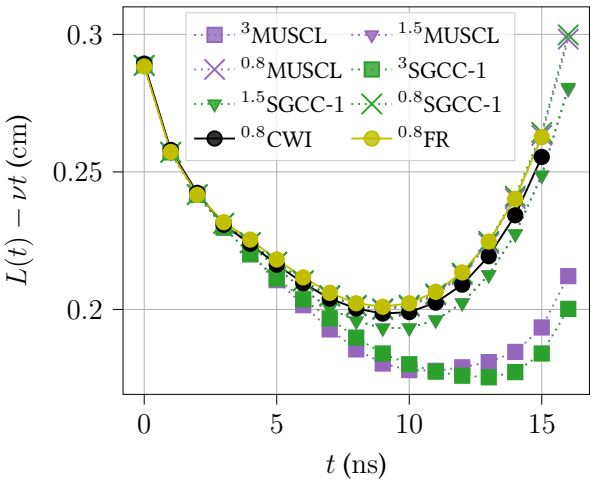
In terms of computation costs, table 5.4 shows the CPU time, the simulation time as well as the CPU processor type used in each test case. Note that the simulation times of the SG scheme are not



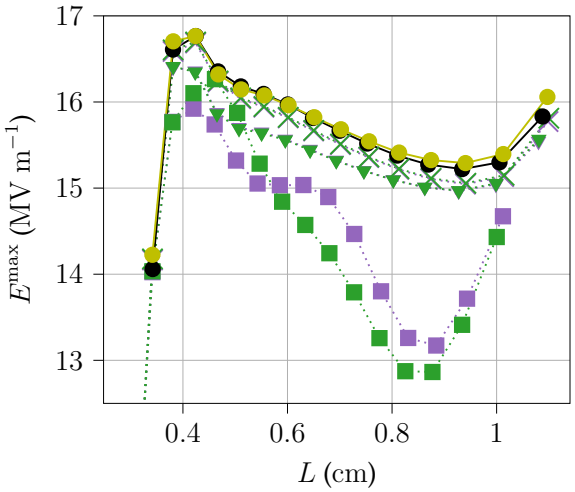
(a) Streamer length L (SGCC-1 & SGCC-2)



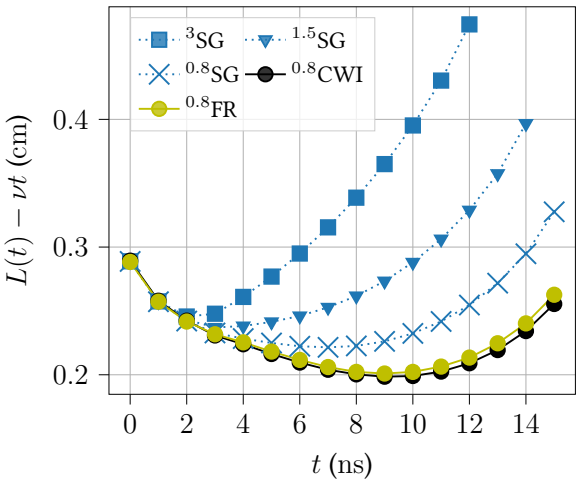
(b) Maximum field strength $E^{\max}$ (SGCC-1 & SGCC-2)



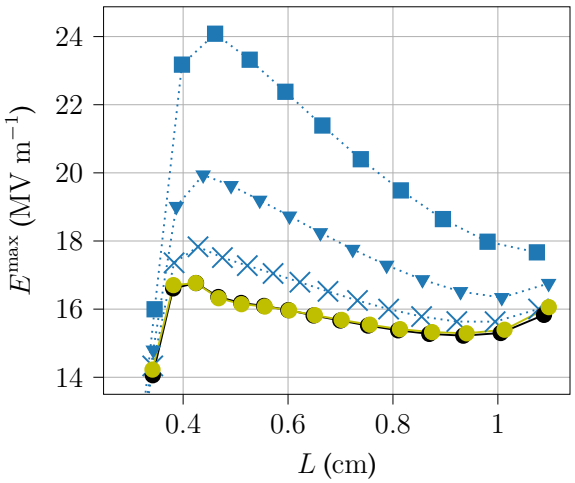
(c) L (SGCC-1 & MUSCL)



(d) $E^{\max}$ (SGCC-1 & MUSCL)

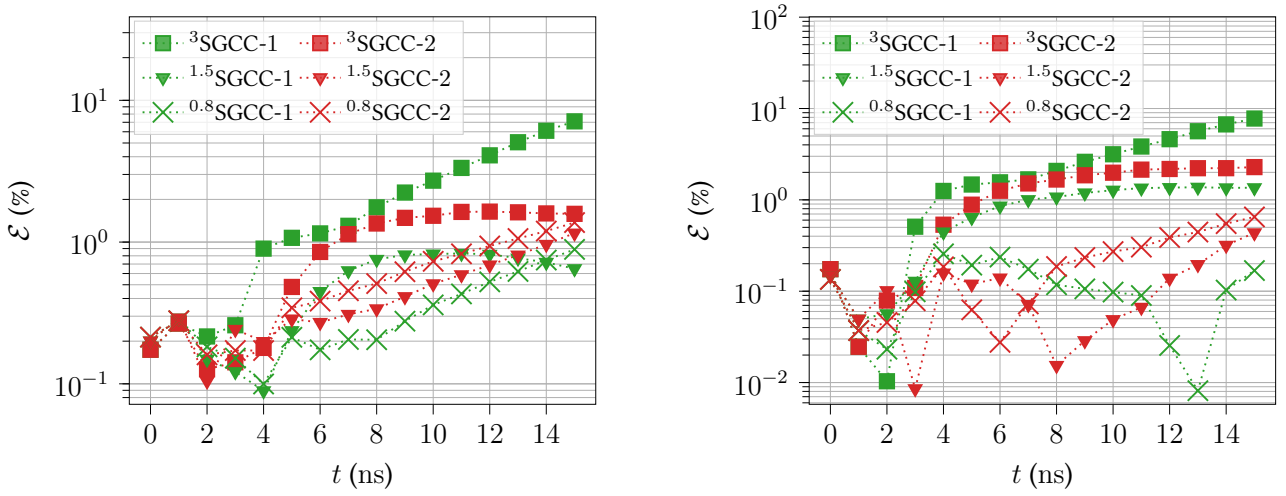


(e) L (SG)



(f) $E^{\max}$ (SG)

Figure 5.7: (SS) Numerical results of the SG, MUSCL, SGCC-1 and SGCC-2 schemes as well as of the teams CWI [140] and FR [48]. We subtract νt from $L(t)$, with $\nu = 0.05$ cm ns^{-1}, to enhance the difference between the curves.



(a) With respect to $^{0.8}$CWI

(b) With respect to $^{0.8}$FR

Figure 5.8: (SS) Relative errors on the streamer length L generated by the SGCC-1 and SGCC-2 schemes

$\Delta r^{\min}$	SG		SGCC-1		SGCC-2		MUSCL		CPU processor
	T^{CPU}	T	T^{CPU}	T	T^{CPU}	T	T^{CPU}	T	
3 μm	2.1	12	2.9	16	3.3	16	2.7	16	i5-10210U @ 1.60GHz
1.5 μm	40	14	44	16	111	16	41	16	i5-10210U @ 1.60GHz
0.8 μm	187	15	334	16	768	16	327	16	E5-4650 @ 2.70GHz

Table 5.4: (SS) CPU time T^{CPU} (hours), simulation time T (ns) and CPU processor type used in the simulations

the same as other schemes since it is very diffusive that the streamer reaches the cathode before $T = 16$ ns.

In order to study the efficiency of the schemes in terms of trade-off between CPU time and numerical accuracy, we evaluate the total relative error $\mathcal{E}^{\mathrm{tot}}$ of each scheme with respect to benchmark results in function of the effective CPU time $\mathcal{T}^{\mathrm{CPU}}$, both defined in the following way,

$$\mathcal{E}^{\mathrm{tot}} = \left| \frac{\sum_i L(t_i) - L^{\mathrm{benchmark}}(t_i)}{\sum_i L^{\mathrm{benchmark}}(t_i)} \right|, \tag{5.9}$$

$$\mathcal{T}^{\mathrm{CPU}} = \frac{T^{\mathrm{ref}}}{T} T^{\mathrm{CPU}} \times \text{ CPU clock rate (GHz)}, \tag{5.10}$$

where $T^{\mathrm{ref}} = 16$ ns, T is measured in ns, $t_i \in \{ 1 \text{ ns}, 2 \text{ ns}, \ldots, T \}$ and the CPU clock rates are given in table 5.4.

The $\mathcal{T}^{\mathrm{CPU}}$-$\mathcal{E}^{\mathrm{tot}}$ curves are displayed in fig. 5.9[5], showing that the relative errors of the ^{3}SGCC-2 simulation (2 to 3%) are closed to those of $^{1.5}$SGCC-1 (3 to 4%) but the CPU time of the former is only 4% of the latter. Similarly, the relative errors of $^{1.5}$SGCC-2 (0.4 to 2%) are comparable to those of $^{0.8}$SGCC-1 (0.3 to 1%) but the CPU time of the former is one-fifth of the latter. Therefore, the

[5] note that simulations on more refined grids cost more CPU time

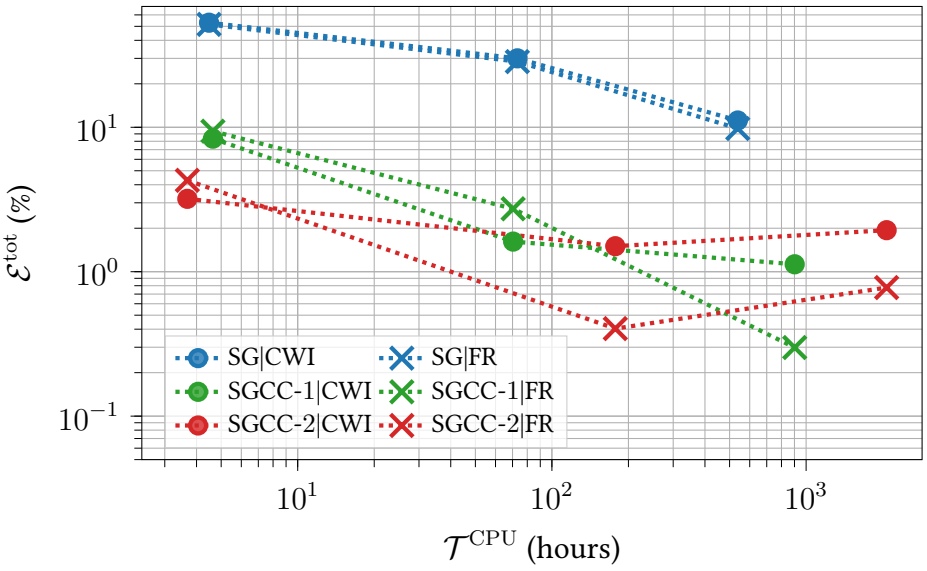


Figure 5.9: (SS) $\mathcal{T}^{\text{CPU}}$-$\mathcal{E}^{\text{tot}}$ curves of the SG, MUSCL, SGCC-1 and SGCC-2 schemes with respect to the benchmarks $^{0.8}$CWI and $^{0.8}$FR

third-order SGCC-2 scheme is more efficient than the second-order SGCC-1, at least for the streamer discharge in this section.

In summary, all four numerical schemes demonstrate their capability to numerically reproduce a complex and dynamic electric discharge and exhibit mesh convergence with decreasing grid size. Among them, the SGCC-2 scheme proves to be the most accurate and efficient method. However, this assertion only holds for these simulations using the Strang splitting, as we shall see in sections 5.3.3 and 5.3.4 that it is no longer true.

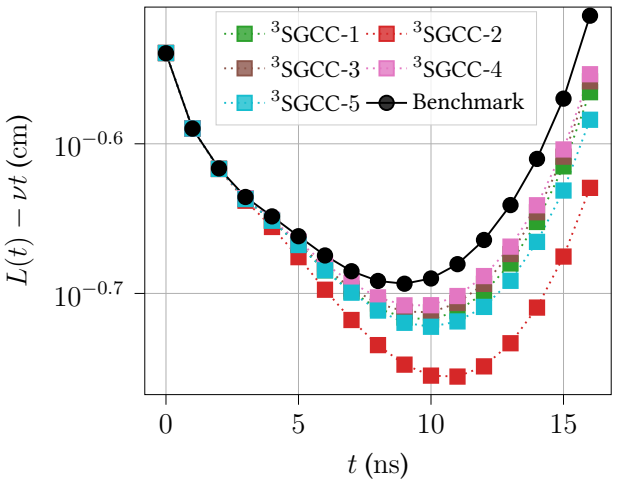
5.3.3 Numerical results with high initial density, without direction splitting and without slope limiters (WOL simulations)

We investigate here the numerical results of the SGCC-p schemes with $p = 1$ to 5. The initial density n^0 is fixed to 10^{13} m^{-3} and the electron continuity equation is directly discretized in two dimensions. No slope limiters are used since as it turns out, the simulations in this section could be conducted without any limiting techniques.

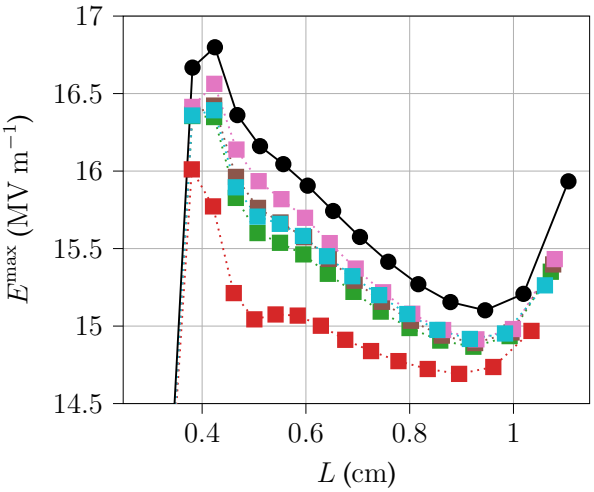
For the benchmark solutions, we use instead the results of the **Strang splitting $^{0.8}$SGCC-2 method** in section 5.3.2 instead of the CWI and FR groups. The evolution in time of the streamer length L as well as the maximum field strength $E^{\max}$ are shown in fig. 5.10. Simulations are carried out on the 3 μm-grid and 1.5 μm-grid.

The relative errors defined in eq. (5.8) for the SGCC-1,2,3,4,5 schemes are presented in fig. 5.11.

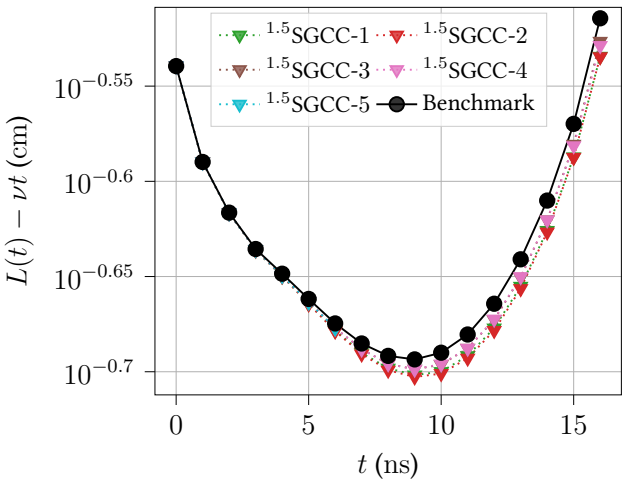
In contrast to section 5.3.2, the two-dimensional ^{3}SGCC-1 simulation yields a more stable maximum field strength during the propagation phase (see fig. 5.10). In particular, the SGCC-1 scheme is even more accurate than the SGCC-2 scheme on the 3 μm-grid, with the relative error on the streamer length around 3% max, comparing to 7% max of the latter (see fig. 5.11). Another surprise



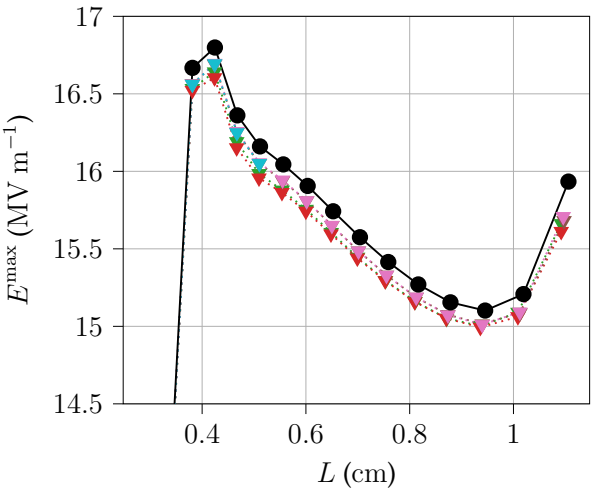
(a) Streamer length L (3 μm-grid)



(b) Maximum field strength $E^{\max}$ (3 μm-grid)

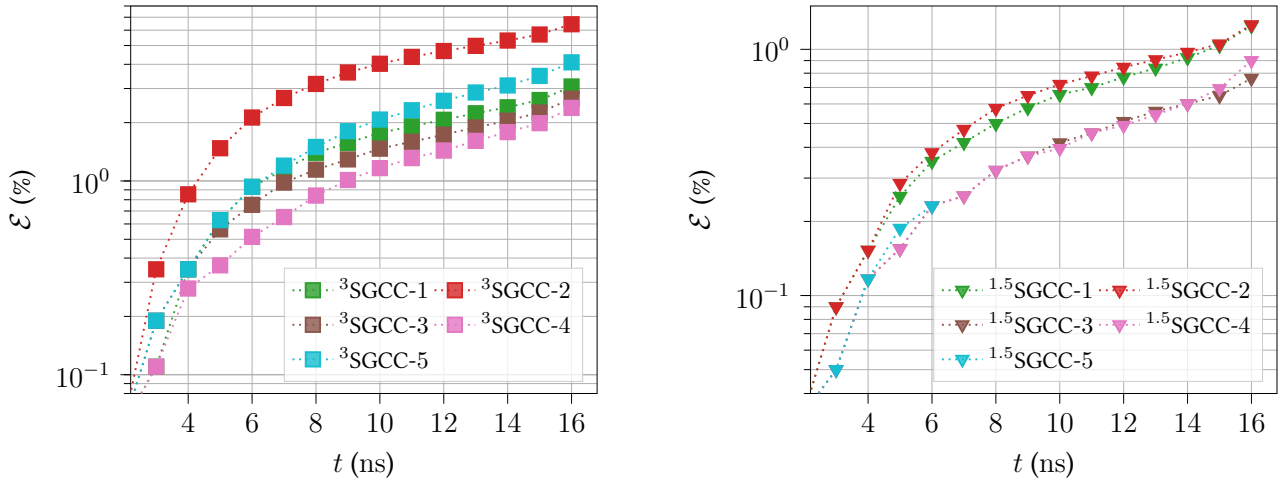


(c) L (1.5 μm-grid)



(d) $E^{\max}$ (1.5 μm-grid)

Figure 5.10: (WOL) Numerical results of the SGCC-1,2,3,4,5 schemes as well as of the Strang splitting $^{0.8}$SGCC-2 simulation in section 5.3.2 as benchmark. We subtract νt from $L(t)$, with $\nu = 0.05$ cm ns^{-1}, to enhance the difference between the curves.



(a) 3 μm-grid

(b) 1.5 μm-grid

Figure 5.11: (WOL) Relative errors on the streamer length L generated by the SGCC-1,2,3,4,5 schemes

is that the SGCC-5 scheme is also not as good as the SGCC-2 scheme on the 3 μm-grid.

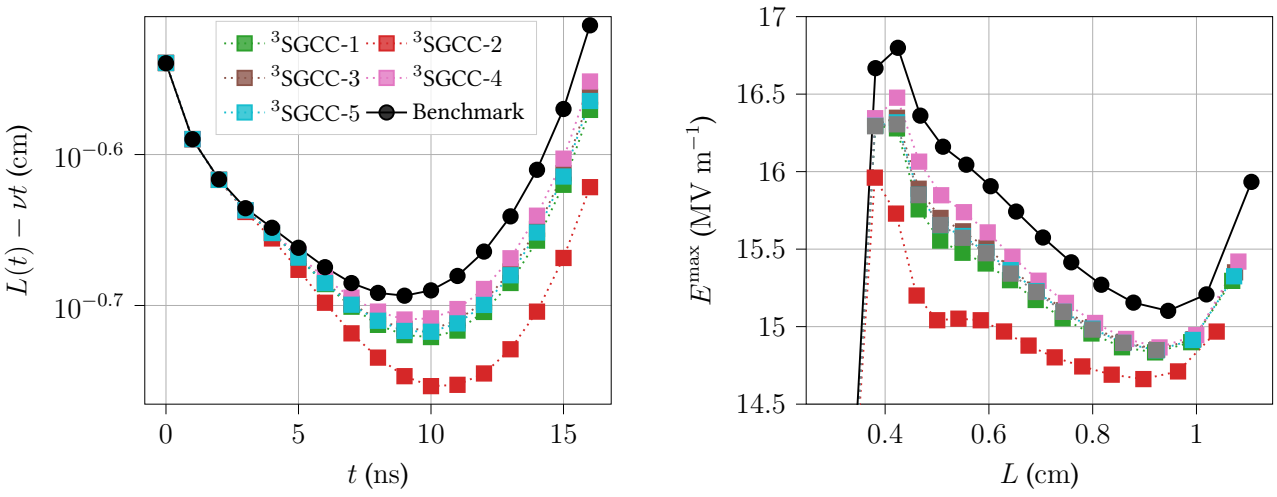
However, on the 1.5 μm-grid, all the curves are visibly very close to each other as well as to the benchmark curves in fig. 5.10, with the cumulative error (at the end) around 0.6 to 1%. The relative errors of the SGCC-2 scheme become almost identical to the SGCC-1 scheme. Although we do not have the numerical results of the SGCC-5 scheme for $t > 6$ ns, it seems from fig. 5.11 that the relative errors of $^{1.5}$SGCC-5 are close to $^{1.5}$SGCC-3 and $^{1.5}$SGCC-4, which are well smaller than $^{1.5}$SGCC-1 and $^{1.5}$SGCC-2. This is consistent with the convergence order of the schemes.

5.3.4 Numerical results with high initial density, without direction splitting and with slope limiters (WL simulations)

The use of Barth-Jespersen limiter described in section 4.7.3 does not change much the numerical solutions, at least on the 3 μm-grid (see fig. 5.12). The discrepancy on the streamer length (fig. 5.13) is roughly the same as the test case in section 5.3.3, with a minor exception that the relative errors of the SGCC-5 scheme are smaller than SGCC-1 in this section.

5.3.5 Numerical results with low initial density, direction splitting and slope limiters (SSL simulations)

In this section, we present the numerical results of the same schemes, on the same grids as in section 5.3.2 but with lower initial density $n^0 = 10^9$ m^{-3}. Figure 5.14 shows that the electron density n_e of this case is almost as twice as in section 5.3.2 (see fig. 5.6a). Comparing to the latter, the simulations in this section are more challenging due to the steeper field gradient in the streamer head and the streamer takes longer, roughly $T = 24$ ns, to cross the inter-electrode gap. Indeed, on the 3 μm-grid, only the ^{3}SGCC-2 simulation succeeds to reach the cathode, whereas ^{3}SG stops at $t = 3$ ns while ^{3}SGCC-1 and ^{3}MUSCL both crash at $t = 10$ ns due to the amplification of nonphysical oscillations that were similarly observed in fig. 5.6a. From fig. 5.14, we note that these oscillations



(a) Streamer length L

(b) Maximum field strength $E^{\max}$

Figure 5.12: (WL) Numerical results of the SGCC-1,2,3,4,5 schemes as well as of the Strang splitting $^{0.8}$SGCC-2 simulation in section 5.3.2 as benchmark. We subtract νt from $L(t)$, with $\nu = 0.05$ cm ns^{-1}, to enhance the difference between the curves.

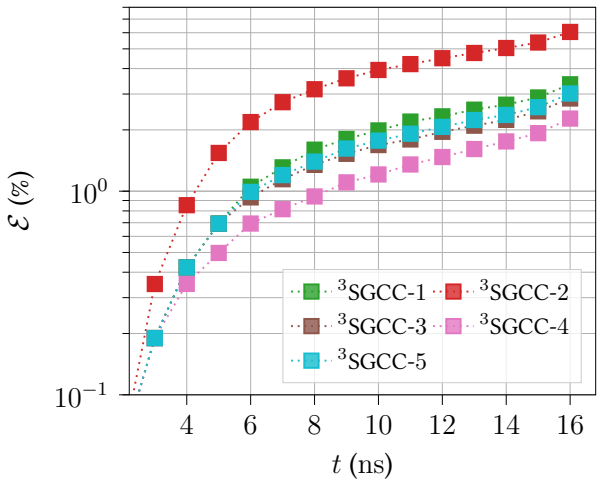
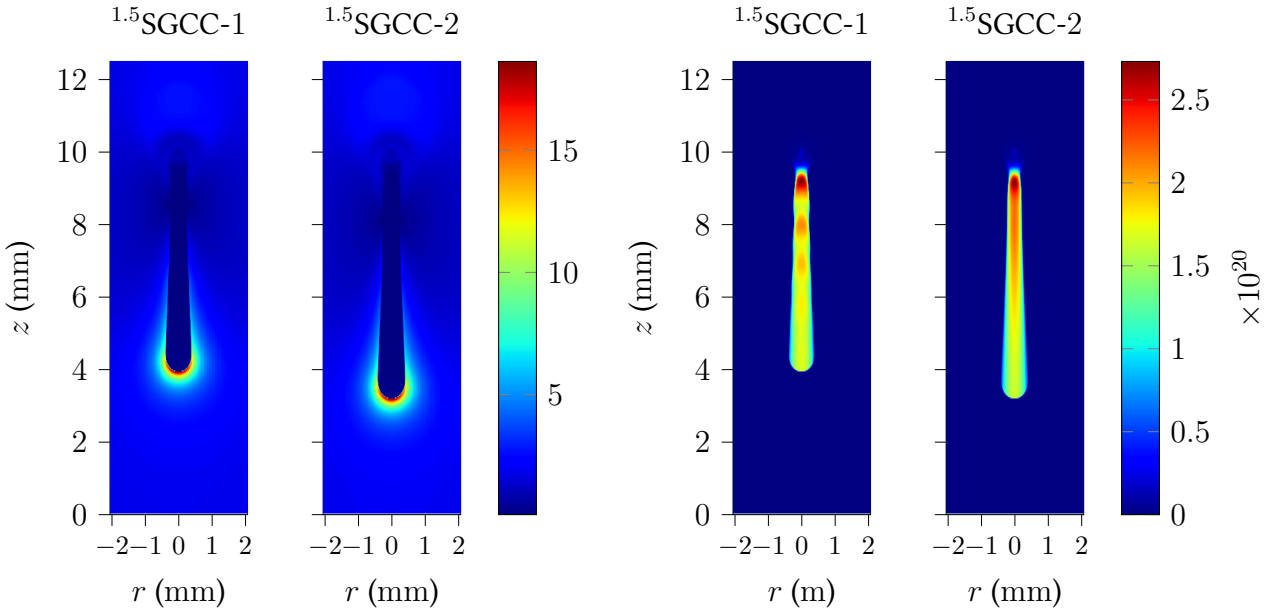


Figure 5.13: (WL) Relative errors on the streamer length L generated by the SGCC-1,2,3,4,5 schemes



(a) Field strength E (MV m^{-1})

(b) Electron density n_e (m^{-3})

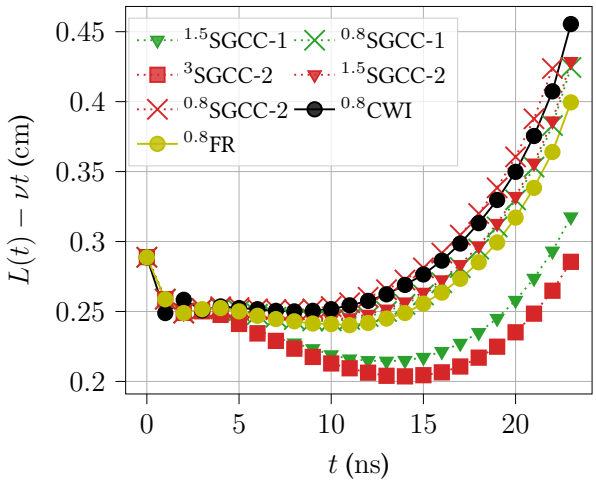
Figure 5.14: (SSL) Numerical results of $^{1.5}$SGCC-1 and $^{1.5}$SGCC-2 at $T = 20$ ns

even persist for $^{1.5}$SGCC-1 (as well as for $^{1.5}$MUSCL but not shown), as the electron density as well as the field strength exhibit a wavy structure on the two laterals of the streamer body. Additionally, we also observe intermittent peaks of the electron density inside the streamer body. On the contrary, the results of $^{1.5}$SGCC-2 are "smooth". On the 3 μm-grid, the SGCC-2 scheme (not shown here) behaves in the same way as $^{1.5}$SGCC-1 and also manages to stay stable until the end of simulation.

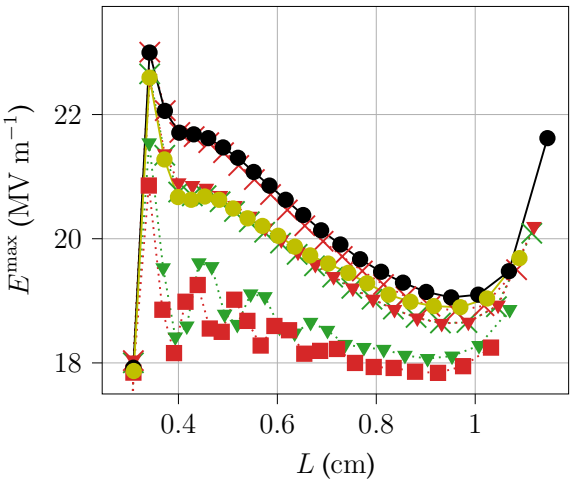
The streamer length $L(t)$ and the maximum field strength $E^{\max}$ are shown in fig. 5.15. The benchmarks are again the solutions of the **CWI** and **FR groups**. As in section 5.3.2, the SG scheme is always too diffusive that it provides inadequate results. On the 1.5 μm-grid it crashes at $t = 10$ ns, while on the 0.8 μm-grid it overestimates L as well as $E^{\max}$ and the streamer already reaches the cathode at $t = 20$ ns.

The relative errors defined in eq. (5.8) for the SGCC-1 and SGCC-2 schemes are presented in fig. 5.16.

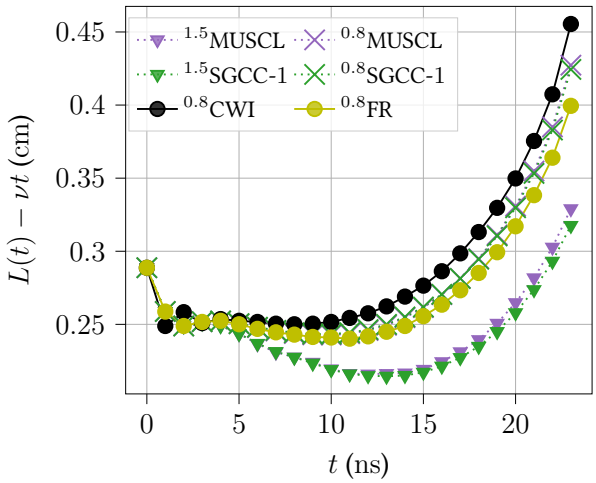
On fig. 5.15, we observe the aforementioned oscillations that exhibit on the maximum field strength of ^{3}SGCC-2, $^{1.5}$SGCC-1 and $^{1.5}$MUSCL, which disappear once the grid size decreases. From figs. 5.15 and 5.16, we remark that the numerical results of ^{3}SGCC-2 are comparable to $^{1.5}$SGCC-1 as well as $^{1.5}$MUSCL, having relative errors around 10% with respect to both benchmarks. On the other hand, $^{1.5}$SGCC-2 are roughly the same as $^{0.8}$SGCC-1 and $^{0.8}$MUSCL with the maximum relative errors around 2 to 3%. The relative errors of $^{0.8}$SGCC-2 with respect to the $^{0.8}$CWI benchmark are smaller that those of $^{1.5}$SGCC-2; however, they become mysteriously closer as the simulations advance in time. With respect to the $^{0.8}$FR benchmark, the relative errors of $^{0.8}$SGCC-2 are definitely larger than $^{1.5}$SGCC-2, suggesting that other sources of numerical error (for example from the Poisson equation) are relevant or because that the numerical result of $^{0.8}$FR has not come close to the mesh-converged solution.



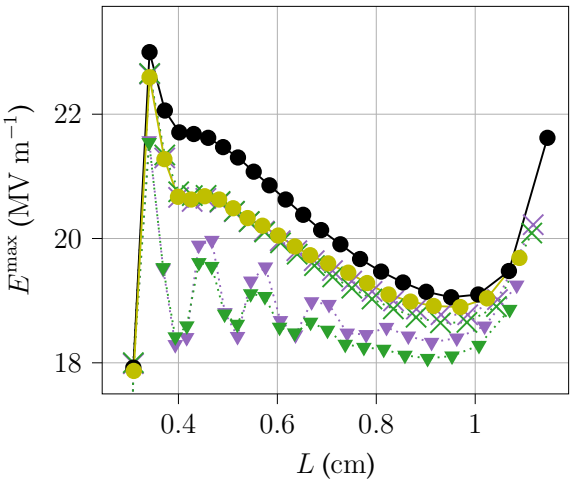
(a) Streamer length L



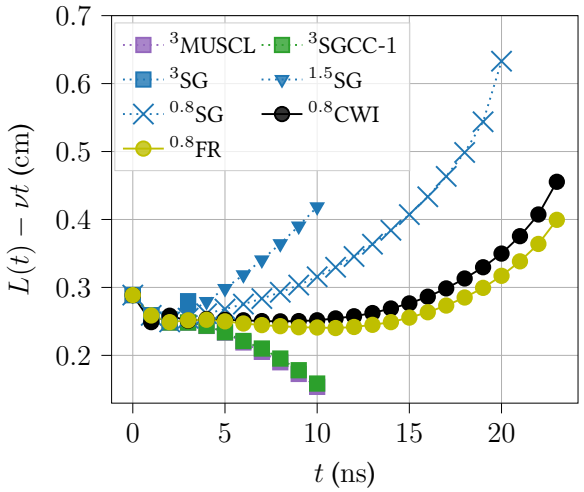
(b) Maximum field strength $E^{\max}$



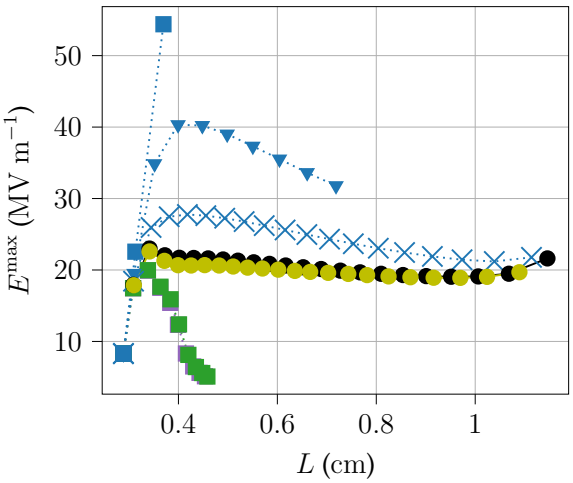
(c) L



(d) $E^{\max}$

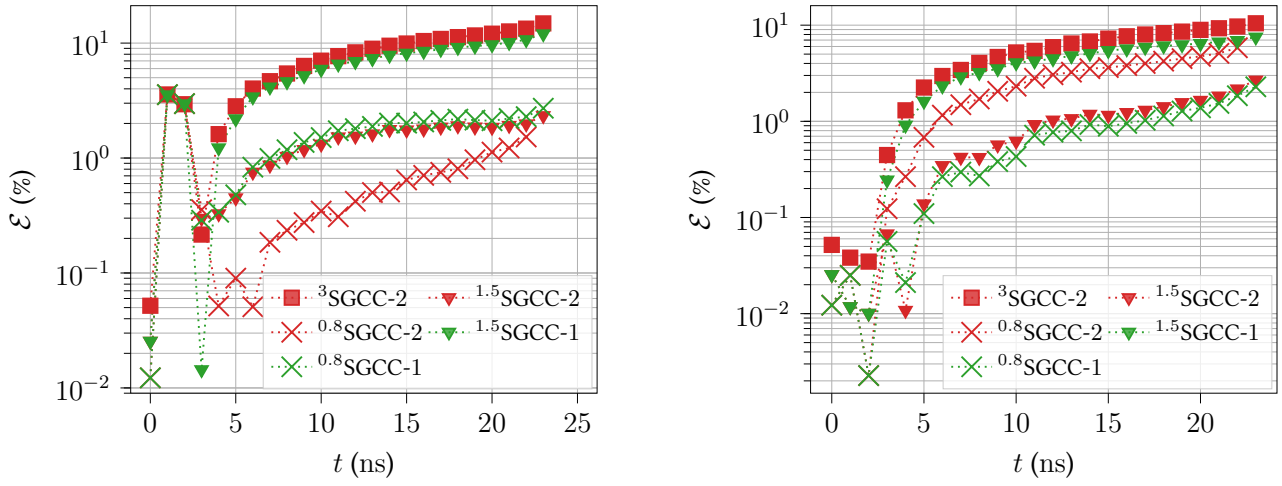


(e) L



(f) $E^{\max}$

Figure 5.15: (SSL) Numerical results of the SG, MUSCL, SGCC-1 and SGCC-2 schemes as well as of the teams CWI [140] and FR [48]. We subtract νt from $L(t)$, with $\nu = 0.03$ cm ns^{-1}, to enhance the difference between the curves.



(a) With respect to $^{0.8}$CWI

(b) With respect to $^{0.8}$FR

Figure 5.16: (SSL) Relative errors on the streamer length L generated by the SGCC-1 and SGCC-2 schemes

$\Delta r^{\min}$	SG		SGCC-1		SGCC-2		MUSCL		CPU processor
	T^{CPU}	T	T^{CPU}	T	T^{CPU}	T	T^{CPU}	T	
3 μm					8.2	24			i5-10210U @ 1.60GHz
1.5 μm			48	24	62	24	44	24	E7-8890 @ 2.20GHz
0.8 μm	576	20	797	24	2755	23	1000	24	X7550 @ 2.00GHz

Table 5.5: (SSL) CPU time T^{CPU} (hours), simulation time T (ns) and CPU processor of the simulations

Table 5.5 shows the CPU time, the simulation time as well as the CPU processor used in each simulation, except for ^{3}SG, ^{3}SGCC-1, ^{3}MUSCL and $^{1.5}$SG which crash before the streamer reaches the cathode.

Figure 5.17 shows the $\mathcal{T}^{\text{CPU}}$-$\mathcal{E}^{\text{tot}}$ curves of the schemes SGCC-1 and SGCC-2, with $T^{\text{ref}} = 24$ ns in eq. (5.10). It follows from the visualization that the total relative errors of ^{3}SGCC-2 are roughly the same as $^{1.5}$SGCC-1 but the CPU time of the former is only 13% of the latter. Similarly, the CPU time of $^{1.5}$SGCC-2 is one-tenth of $^{0.8}$SGCC-1 for the same level of precision. Therefore, the third-order SGCC-2 scheme is again more efficient than the second-order SGCC-1 for the streamer discharge in this section.

5.4 Closing remarks

This chapter investigated the capability of the SGCC schemes in the simulation of some complex electric discharges in air. The numerical results of this chapter were obtained with self-implemented codes: one for the corona discharge in section 5.2 and the other for the positive streamer propagation in section 5.3. In particular, our code for the streamer simulation was verified against the works of two research groups from Europe that had been published in [7].

Evidences in sections 5.2, 5.3.2 and 5.3.5 show the accuracy gain with the high-order Scharfetter-

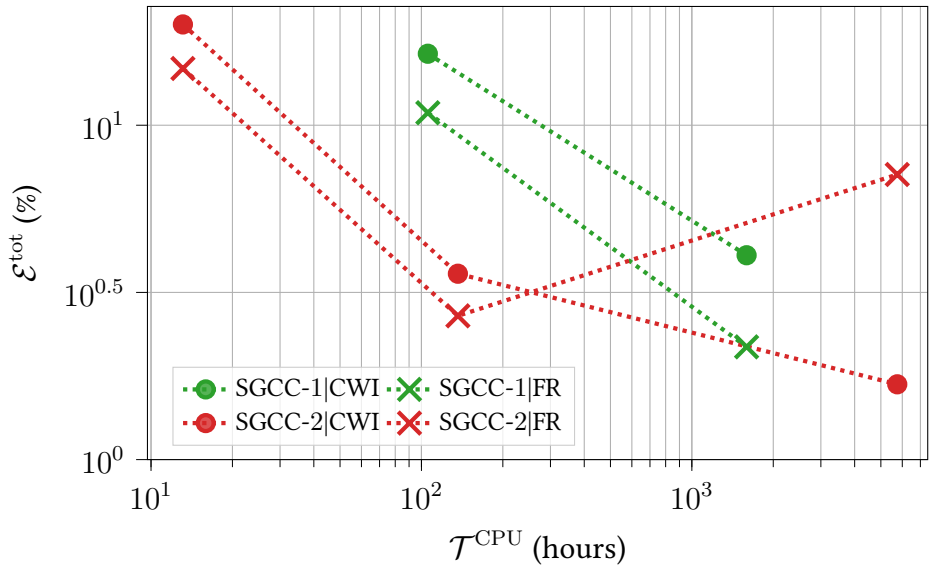


Figure 5.17: (SSL) $\mathcal{T}^{\text{CPU}}$-$\mathcal{E}^{\text{tot}}$ curves of the SG, MUSCL, SGCC-1 and SGCC-2 schemes with respect to the benchmarks $^{0.8}$CWI and $^{0.8}$FR

Gummel schemes since with the same grid size, the quality of the numerical solutions is better as the precision order is increased. Particularly, sections 5.3.2 and 5.3.5 also show that the third-order SGCC-2 scheme is more efficient than the second-order SGCC-1 in terms of trade-off between numerical accuracy and CPU time. Moreover, section 5.3.5 highlights the robustness of the SGCC-2 method as it succeeds to reach the simulation time, even on the coarsest grid, as opposed to the SG or SGCC-1 schemes. The simulations in these sections used the Strang splitting to discretize the electron continuity equation.

The numerical results of sections 5.3.3 and 5.3.4 validate the SGCC-1,2,3,4,5 schemes, but also show on the contrary to sections 5.2, 5.3.2 and 5.3.5 that the SGCC-2 scheme yields less accurate solutions comparing to SGCC-1 on the 3 μm-grid. A possible explanation for this observation is that the constant error of SGCC-2 is greater than that of SGCC-1. On the 1.5μm-grid though, the two schemes perform equivalently well.

We remark again that the Poisson equation was solved with first-order schemes throughout this chapter, so the global precision is first-order. Nevertheless, it seems that the quality of the numerical results depends essentially on the discretization of the electron continuity equation, as we observe from sections 5.2 and 5.3 that with the same grid size, the solutions came closer to the benchmarks as the precision order was increased. The influence of the Poisson equation should not be totally disregarded, though, without any hard evidences in certain unwanted situations such as the rise in total relative error of the $^{0.8}$SGCC-2 simulation in sections 5.3.2 and 5.3.5 and the under-performance of the SGCC-2,5 schemes on the 3 μm-grid in sections 5.3.3 and 5.3.4. Therefore, we think that it will be interesting if a study with high-order approximation of the Poisson equation is carried out in the future.

A general take-away from this chapter is that high-order schemes allow us to save time as they could be performed on coarse grids and still yield satisfactory solutions. However, their combination

with explicit time integration still keeps them away from large-scale simulations, as it took **hours to months** to complete a corona discharge simulation in **one dimension** (section 5.2), or **hours** to simulate some **tens of nanoseconds** of a streamer discharge (section 5.3). Consequently, the main goal of the next chapter is to develop an implicit time integration strategy that is able to considerably shorten the CPU time of discharge simulations.

5.5 Remarques finales

Dans ce chapitre, on a étudié la capacité des schémas SGCC à simuler certaines décharges électriques complexes dans l'air. Les résultats numériques de ce chapitre ont été obtenus avec les codes réalisés de scratch : l'un pour la décharge couronne dans la section 5.2 et l'autre pour la propagation de streamers positifs dans la section 5.3. En particulier, nos codes ont été vérifiés contre les travaux de deux groupes de recherche européens qui avaient été publiés dans [7].

Les preuves dans les sections 5.2, 5.3.2 et 5.3.5 montrent le gain de précision avec les schémas de Scharfetter-Gummel d'ordre élevé puisqu'avec la même taille de maillage, la qualité des solutions numériques est meilleure à mesure que l'ordre de précision augmente. En particulier, les sections 5.3.2, 5.3.5 et 5.3.5 montrent également que le schéma SGCC-2 du troisième ordre est plus efficace que le SGCC-1 du deuxième ordre en termes de l'équilibre entre la précision numérique et le temps de CPU. En outre, la section 5.3.5 met en évidence la robustesse de SGCC-2 puisqu'elle parvient à atteindre la fin de simulation, même sur le maillage le plus grossier, contrairement aux SG et SGCC-1. Les simulations présentées dans ces sections ont utilisé le splitting de Strang pour discrétiser l'équation de continuité des électrons.

Les résultats numériques des sections 5.3.3 et 5.3.4 valident les schémas SGCC-1,2,3,4,5, mais montrent également, contrairement aux sections 5.2, 5.3.2 et 5.3.5, que SGCC-2 produit une solution moins précise que SGCC-1 sur le maillage de $3\ \mu\mathrm{m}$. Cette observation peut s'expliquer par le fait que la constante d'erreur de SGCC-2 est plus importante que celle de SGCC-1. Sur le maillage de $1.5\ \mu\mathrm{m}$ cependant, les deux schémas ont des performances équivalentes.

Nous remarquons à nouveau que l'équation de Poisson a été résolue avec des schémas du premier ordre tout au long de ce chapitre, de sorte que la précision globale est du premier ordre. Néanmoins, il semble que la qualité des résultats numériques dépende essentiellement de la discrétisation de l'équation de continuité des électrons, car nous observons dans les sections 5.2 et 5.3 qu'avec la même taille de maillage, les solutions se rapprochent des références à mesure que l'ordre de précision augmente. L'influence de l'équation de Poisson ne doit cependant pas être totalement ignorée, en l'absence de preuves tangibles dans certaines situations indésirables telles que l'augmentation de l'erreur relative totale de la simulation ^{0,8}SGCC-2 dans la sections 5.3.2 and 5.3.5 et la sous-performance des schémas SGCC-2,5 sur le maillage de $3\ \mu\mathrm{m}$ dans la sections 5.3.3 and 5.3.4. Par conséquent, nous pensons qu'il serait intéressant de réaliser à l'avenir une étude avec une approximation d'ordre élevé de l'équation de Poisson.

Il ressort de ce chapitre que les schémas d'ordre élevé nous permettent de gagner du temps, car ils peuvent être lancés sur des maillages grossiers tout en produisant des solutions satisfaisantes.

Cependant, leur combinaison avec une intégration temporelle explicite les tient encore à l'écart des simulations à grande échelle, car il a fallu **des heures à des mois** pour terminer une simulation de décharge de couronne dans **une dimension** (section 5.2), ou **des heures** pour simuler quelques **dizaines de nanosecondes** d'une décharge de streamer (section 5.3). Par conséquent, l'objectif principal du chapitre suivant est de développer une stratégie d'intégration temporelle implicite capable de réduire considérablement le temps de CPU des simulations.

Chapter 6

Mathematical and Numerical Modeling of Corona Discharge*

6.0 Aperçu . . . 142
6.1 Overview . . . 143
6.2 Formulation of the gas discharge model with floor density . . . 145
6.3 Existence and uniqueness of the solution of the differential inclusion . . . 147
6.3.1 Preliminaries . . . 148
6.3.2 A penalization problem . . . 150
6.3.3 Weak solutions . . . 151
6.3.4 Proof of theorem 6.2 . . . 152
6.3.5 Proof of proposition 6.1 . . . 157
6.3.6 Proof of theorem 6.3 . . . 157
6.4 Implicit time discretization and treatment of the source terms . . . 160
6.4.1 Density-field decoupling strategy for simulation of corona discharge 160
6.4.2 Treatment of the source terms . . . 161
6.5 Solving the conservation laws . . . 164
6.5.1 Time-discrete system of conservation laws . . . 164
6.5.2 Fully discrete system of conservation laws . . . 164
6.5.3 Lie splitting . . . 166
6.5.4 Douglas-Rachford algorithm . . . 167
6.5.5 Gauss-Seidel algorithm . . . 169
6.6 Comparison of algorithms on one-dimensional grids . . . 169
6.6.1 Conservation of steady state . . . 169
6.6.2 Performance comparison of the DR and GS algorithms . . . 171
6.6.3 Mesh convergence . . . 172
6.7 Closing remarks . . . 174
6.8 Remarques finales . . . 174

*parts of this chapter, in altered form, has been published in [145]: *N Tuan Dung, C Besse, and F Rogier. "An implicit time integration approach for simulation of corona discharges". In: Computer Physics Communications 294 (2024), p. 108906.*

6.0 Aperçu

L'objectif de ce chapitre est de concevoir une stratégie d'intégration implicite en temps pour la simulation de la décharge couronne.

Comme nous l'avons déjà mentionné, la simulation de la décharge couronne s'avère très difficile en termes de temps de CPU car il existe une grande disparité entre les échelles de temps caractéristiques de plusieurs phénomènes au sein de la décharge. En effet, l'échelle de temps du régime couronne est de l'ordre de 10^{-4} s, mais l'échelle de temps du transport d'électrons n'est que de l'ordre des picosecondes. La résolution du modèle de décharge (2.13) avec une méthode explicite en temps est exhaustive car les pas de temps numériques sont limités par la condition de CFL liée au transport d'électrons.

Certains efforts numériques ont été réalisés pour réduire le temps de calcul. Par exemple, Adamiak & Atten [2] ont proposé un modèle simplifié et stationnaire composé de l'équation de Poisson et d'une équation non linéaire pour la charge d'espace, au détriment de la capture de la dynamique de décharge. Seimandi et al. [132] ont proposé des modèles asymptotiques pour réduire la dimension du problème (de 2D à 1D) dans les gaines d'électrodes. Dufour & Rogier [46] ont développé un algorithme de sous-cyclage dans le solveur de plasma COPAIER de l'ONERA.

Il convient de mentionner qu'il existe des méthodes implicites qui ont été appliquées à la simulation des décharges de gaz dans l'air, mais pas nécessairement pour le régime couronne. Ventzek et al. [155] ont proposé un schéma semi-implicite qui a permis de relâcher la contrainte de pas de temps liée au temps de relaxation diélectrique. Hagelaar & Kroesen [66] ont proposé un schéma similaire qui prend en compte l'équation de l'énergie moyenne des électrons. Il existe d'autres méthodes numériques telles que les schémas entièrement implicites [11, 119, 156] ou une mise-à-jour asynchrone des variables [147], mais leurs développements visent généralement à traiter les microdécharges où le temps de relaxation diélectrique miniature est miniature. Par conséquent, leur application aux décharges couronnes, où le temps de relaxation diélectrique est beaucoup plus important, ne semble pas être bénéfique.

Un obstacle au développement de la méthode implicite présentée dans ce chapitre est la contrainte de maintenir la densité électronique n_e toujours supérieure à une densité prescrite qui est désignée par $\psi(\boldsymbol{x})$.

Cette densité prescrite est appelée **densité de fond**, et même si elle dépend a priori de $\boldsymbol{x}$, ψ est souvent supposée constante. L'imposition de la contrainte $n_e(t, \boldsymbol{x}) \geq \psi(\boldsymbol{x})$ peut être interprétée comme une tentative d'équilibre entre la nécessité de produire des solutions numériques qui s'accordent bien avec les mesures expérimentales et la simplicité du modèle de décharge afin de limiter la complexité numérique. En fait, le schéma cinétique de la décharge dans les simulations dédiées au contrôle d'écoulements tend à être simplifié autant que possible, puisqu'il n'est pas nécessaire de surreprésenter les réactions chimiques qui se produisent sur des échelles de temps beaucoup plus petites que celle du vent ionique. Par exemple, l'équipe de J.-P. Bœuf a utilisé un modèle cinétique à trois espèces et quatre réactions (décrit dans example 2.1) dans ses études numériques [17, 19, 15, 87, 146]. En imposant $n_e \geq \psi$, on peut limiter le modèle cinétique à un modèle simple, mais aussi ajuster

les solutions numériques aux données expérimentales en choisissant une valeur raisonnable pour ψ. Cette approche de la densité de fond a été utilisée dans COPAIER pour reproduire les courants de circuit mesurés dans une décharge couronne entre un fil et un profil aérodynamique [99], où ψ a été pris entre 10^{11}-10^{12} m^{-3}.

D'autres approches ont également été adoptées pour calibrer le courant numérique. Le modèle proposé par Adamiak & Atten [2] a été validé par rapport aux données expérimentales pour différentes géométries d'actionneur telles que le point-plan [2], l'aiguille-plan [165], le fil-cylindre-plaque [103] et l'aiguille-anneau [159]. Un modèle similaire, mais dépendant du temps, a été proposé par Guan et al. [62] mais la condition limite pour l'équation de charge dépend du courant électrique expérimental.

La résolution du système de décharge (2.13) et l'imposition de la contrainte $n_e \geq \psi$ semblent simples à première vue : à l'instant t^l, nous calculons l'approximant discret $(\overline{n}_{e,K}^{l+1})_{K=1,\dots,\mathcal{N}}$ de $n_e(t^{l+1}, \boldsymbol{x})$ puis nous prenons $\overline{n}_{e,K}^{l+1} := \max(\overline{n}_{e,K}^{l+1}, \psi)$ pour chaque $K = 1, \dots, \mathcal{N}$. Cette méthode fonctionne bien avec les schémas explicites en temps, mais beaucoup moins bien avec les schémas implicites, comme on le verra dans section 6.6, car elle ne conserve pas le courant en régime stationnaire (si la décharge en régime stationnaire existe).

Afin de concevoir une méthode d'intégration implicite "correcte", nous proposons dans ce chapitre une reformulation mathématique du modèle de décharge de sorte que la contrainte sur la densité électronique apparaisse dans la loi de conservation des électrons. La dérivation formelle de cette reformulation du modèle est présentée dans la section 6.2. La section 6.3 réexamine la question de l'existence et de l'unicité de la solution de la loi de conservation des électrons proposée sous certaines hypothèses spécifiques. La section 6.4 discute une stratégie de discrétisation du modèle de plasma afin que les lois de conservation des particules soient découplées de l'équation de Poisson ainsi que les traitements spécifiques pour les termes sources. Ensuite, nous étudions dans la section 6.5 quelques algorithmes pour résoudre numériquement le modèle de décharge proposé. Comme le problème est non linéaire en raison de la présence de la contrainte de densité électronique, les méthodes directes ne sont pas appropriées et, par conséquent, des méthodes de splitting telles que le splitting de Lie ou des méthodes itératives telles que les algorithmes de Douglas-Rachford ou de Gauss-Seidel sont employées. Enfin, des tests numériques sont effectués dans la section 6.6 pour évaluer la performance de ces algorithmes.

6.1 Overview

The goal of this chapter is to devise an implicit time integration strategy for the simulation of corona discharge.

As mentioned earlier, simulation of corona discharge proves to be very challenging in terms of CPU time [111] since there is a great disparity among characteristic time scales of several phenomena within a discharge. Indeed, the time scale of corona regime is on the order of 10^{-4} s, but the time scale of electron transport is only on the order of picoseconds. The resolution of the discharge model (2.13) with an explicit time method is exhaustive because the numerical timesteps are limited by the CFL condition related to electron transport.

Some numerical efforts have been made to reduce the CPU time. For instance, Adamiak & Atten [2] proposed a simplified, stationary model consisting of the Poisson equation and a nonlinear equation for the space charge, at the expense of capturing the discharge dynamics. Seimandi et al. [132] proposed some asymptotic models to reduce the dimension of the problem (2D to 1D) in the electrode sheaths. Dufour & Rogier [46] developed a sub-cycling algorithm in ONERA's plasma solver COPAIER.

It is worth mentioning that there exists implicit methods that have been applied in simulation of gas discharge in air, but not necessarily suitable to the corona regime. Ventzek et al. [155] proposed a semi-implicit scheme which allowed to relax the timestep constraint related to the dielectric relaxation time. Hagelaar & Kroesen [66] proposed a similar scheme that takes into account the electron mean energy equation. There are other numerical methods such as fully-coupled implicit schemes [11, 119, 156] or asynchronous time stepping [147], but their developments usually aim to deal with the miniature dielectric relaxation time in microdischarges. Hence, their application in corona discharge, where the dielectric relaxation time is much larger, does not seem to be beneficial.

An obstacle to the development of the implicit method in this chapter is the requirement to keep the electron density n_e always higher than a prescribed density that is denoted as $\psi(\boldsymbol{x})$.

This prescribed density is called **floor density**, and even though it depends a priori on $\boldsymbol{x}$, ψ is frequently assumed to be constant. The imposition of the constraint $n_e(t, \boldsymbol{x}) \geq \psi(\boldsymbol{x})$ can be interpreted as an attempt to balance the need of producing numerical solutions that agree well with experiment measurements, and the simplicity of the discharge model to keep the numerical complexity low. As a matter of fact, kinetic schemes in simulations dedicated to flow control tend to be simplified as much as possible, since there is no need to over-represent the chemical reactions which occur on much smaller time scales than that of ionic wind. For instance, the team of J.-P. Bœuf used a three-specie, four-reaction kinetic model (described in example 2.1) in their numerical studies [17, 19, 15, 87, 146]. By imposing $n_e \geq \psi$, one can limit the kinetic model to a simple one, but also can fit the numerical solutions with experiment data by choosing a reasonable value for ψ. This floor density approach was employed in COPAIER to reproduce measured circuit currents in a corona discharge between a wire and an airfoil [99], and ψ was determined to be between 10^{11}-10^{12} m^{-3}.

There has also been other approaches to calibrate the numerical current. The model proposed by Adamiak & Atten [2] have been validated against experiment data for different actuator geometries such as point-to-plane [2], needle-to-plane [165], wire-cylinder-plate [103] and needle-to-ring [159]. A similar model, but time-dependent, was proposed by Guan et al. [62] but the boundary condition for the charge equation depends on experimental electric currents.

The resolution of the discharge system (2.13) and the imposition of the constraint $n_e \geq \psi$ seems to be simple at first glance: at time t^l, we compute the discrete approximants $(\overline{n}_{e,K}^{l+1})_{K=1,\dots,\mathcal{N}}$ of $n_e(t^{l+1}, \boldsymbol{x})$, then set $\overline{n}_{e,K}^{l+1} := \max(\overline{n}_{e,K}^{l+1}, \psi)$ for every $K = 1, \dots, \mathcal{N}$. This method works well with explicit time schemes but not so much with implicit schemes as will be shown in section 6.6, since it does not conserve the steady-state current (if the steady-state discharge exists).

In order to devise a "correct" implicit time integration method, in this chapter, we propose a

mathematical reformulation of the discharge model so that the constraint on the electron density appears in the electron conservation law. The formal derivation of this model reformulation is presented in section 6.2. Section 6.3 revisits the question of existence and uniqueness of the solution of the proposed electron conservation law under some specific hypotheses. Section 6.4 discusses a discretization strategy of the plasma model so that the particle conservation laws are decoupled from the Poisson equation as well as specific treatments for the source terms. Then, we investigate in section 6.5 some algorithms to solve (numerically) the proposed discharge model. Since the problem is nonlinear due to the presence of the electron density constraint, direct methods are not suitable to use and hence, splitting methods such as Lie operator splitting or iterative methods such as Douglas-Rachford or Gauss-Seidel algorithms are employed. Finally, numerical tests are performed in section 6.6 to evaluate the performance of these algorithms.

6.2 Formulation of the gas discharge model with floor density

Let us consider an open bounded domain $(0,T)\times\Omega\subset\mathbb{R}\times\mathbb{R}^d$ $(d=1,2)$. As first mentioned in sections 5.2 and 5.3, we introduce a positive function defined on Ω,

$$\psi(\boldsymbol{x})\geq 0,\quad \boldsymbol{x}\in\Omega,$$

which is the floor density, to represent the smallest electron density that exists because of various plasma-chemical processes unaccounted for by the discharge model. We enforce the following constraint on the electron density,

$$n_e(t,\boldsymbol{x})\geq\psi(\boldsymbol{x}),\quad \forall(t,\boldsymbol{x})\in(0,T)\times\Omega, \tag{6.1}$$

which also satisfies a priori the equation, recalling from eq. (2.1),

$$\begin{cases}\partial_t n_e+\nabla\cdot\boldsymbol{f}_e=S_e,\\ \boldsymbol{f}_e=\boldsymbol{u}_e n_e-D_e\nabla n_e,\end{cases} \tag{6.2}$$

where $\boldsymbol{u}_e$, D_e and S_e are resp. the electron drift velocity, diffusion coefficient and source term.

The problem as a whole is ill-posed since the solution of eq. (6.2) does not necessarily respect the constraint (6.1). Therefore, a new reformulation of the problem is required. In the next parts, we consider a Hilbert space $\mathcal{H}$ of functions defined on Ω.

Definition 6.1 (Characteristic functions)**.** *Let $\mathcal{K}$ be a non-empty closed convex subset of $\mathcal{H}$. The characteristic function $I_{\mathcal{K}}$ of $\mathcal{K}$ is defined in the following way,*

$$\forall g\in\mathcal{H},\quad I_{\mathcal{K}}(g)=\begin{cases}0 & \text{if}\quad g\in\mathcal{K},\\ +\infty & \text{if}\quad g\notin\mathcal{K}.\end{cases}$$

Definition 6.2 (Subdifferentials [5])**.** *Let G be a function from $\mathcal{H}$ to $\mathbb{R}\cup\{+\infty\}$ such that $\mathrm{dom}(G)\neq\emptyset$*[1]*. We define, for $g\in\mathrm{dom}(G)$ and $h\in\mathcal{H}$,*

$$DG(g)(h)=\lim_{\varepsilon\to 0^+}\frac{G(g+\varepsilon h)-G(g)}{\varepsilon}\in[-\infty,+\infty],$$

[1] $\mathrm{dom}(G)$ is the domain of G

which is known as the derivative from the right of G *at* g *in the direction of* h*. The subdifferential of* G *at* g *is the closed convex subset*

$$\partial G(g) = \{\, v \in \mathcal{H}^* \mid \forall h \in \mathcal{H}, \langle v, h\rangle \leq DG(g)(h) \,\}\,, \tag{6.3}$$

where $\langle \cdot, \cdot \rangle$ *is the duality pairing between* $\mathcal{H}$ *and its topological dual* $\mathcal{H}^*$*. The elements* v *of* $\partial G(g)$ *are called subgradients of* G *at* g*.*

Furthermore, if G *is a convex functional, then*

$$\partial G(g) = \{\, v \in \mathcal{H}^* \mid \forall h \in \mathcal{H}, G(g) + \langle v, h - g\rangle \leq G(h) \,\}\,.$$

Let us denote as $\mathcal{K}_\psi$ the closed convex set of functions in $\mathcal{H}$ that satisfy the constraint eq. (6.1), i.e.

$$\mathcal{K}_\psi = \{\, g \in \mathcal{H} \mid g(\boldsymbol{x}) \geq \psi(\boldsymbol{x}) \quad \text{for a.e. } \boldsymbol{x} \in \Omega \,\}\ ^{2}.$$

We could easily verify that $I_{\mathcal{K}_\psi}$ is a convex function on $\mathcal{H}$ and thus, from definition 6.2 we have

$$\partial I_{\mathcal{K}_\psi}(g) = \begin{cases} \emptyset & \text{if} \quad g \notin \mathcal{K}_\psi, \\ \{\, 0 \,\} & \text{if} \quad g > \psi, \\ \{\, v \in \mathcal{H}^* \mid \forall h \in \mathcal{K}_\psi, \langle v, h - g\rangle \leq 0 \,\} & \text{otherwise.} \end{cases} \tag{6.4}$$

As we have discussed, the constraint eq. (6.1) may not be compatible with eq. (6.2), but if we assume anyway, for now, that the solution n_e of eq. (6.2) satisfies the constraint (6.1), i.e. $n_e(t, \cdot) \in \mathcal{K}_\psi$ for a.e. t, then we would formally have

$$S_e(t, \cdot) - \partial_t n_e(t, \cdot) - \nabla \cdot \boldsymbol{f}_e(t, \cdot) + n_e(t, \cdot) = n_e(t, \cdot) \geq \psi. \tag{6.5}$$

This relations suggest that the projection of $S_e(t, \cdot) - \partial_t n_e(t, \cdot) - \nabla \cdot \boldsymbol{f}_e(t, \cdot) + n_e(t, \cdot)$ on $\mathcal{K}_\psi$ is in fact $n_e(t, \cdot)$. On the other hand, we have the following functional analysis result from [27].

Theorem 6.1. *Let* $\mathcal{H}$ *be a Hilbert space endowed with an inner product* $(\cdot, \cdot)$ *and* $\mathcal{K} \subset \mathcal{H}$ *be a non-empty closed convex. Then for all* $h \in \mathcal{H}$*, there exists a unique* $g \in \mathcal{K}$*, called the projection of* h *on* $\mathcal{K}$*, such that*

$$|h - g| = \min_{v \in \mathcal{K}} |h - v|,$$

where $|\cdot|$ *is the norm induced from* $(\cdot, \cdot)$*. Furthermore,* g *is characterized by the property*

$$(h - g, v - g) \leq 0, \quad \forall v \in \mathcal{K}. \tag{6.6}$$

[2] without ambiguity, we write $g \geq \psi$

Thus from eqs. (6.5) and (6.6),

$$\big(S_e(t,\cdot) - \partial_t n_e(t,\cdot) - \nabla\cdot \boldsymbol{f}_e(t,\cdot) + n_e(t,\cdot) - n_e(t,\cdot), v - n_e(t,\cdot)\big) \leq 0,$$

for a.e. $t \in (0,T)$ and for all $v \in \mathcal{K}_\psi$. By using the Riesz representation theorem to identify $S_e(t,\cdot) - \partial_t n_e(t,\cdot) - \nabla\cdot \boldsymbol{f}_e(t,\cdot)$ with a (unique) element of $\mathcal{H}^*$, this inequality and eq. (6.4) formally imply that $S_e(t,\cdot) - \partial_t n_e(t,\cdot) - \nabla\cdot \boldsymbol{f}_e(t,\cdot)$ is an element of the set $\partial I_{\mathcal{K}_\psi}(n_e(t,\cdot))$. Let us define $S_\psi(n_e(t,\cdot)) \equiv -\partial I_{\mathcal{K}_\psi}(n_e(t,\cdot)) \equiv \{ -v \mid v \in \partial I_{\mathcal{K}_\psi}(n_e(t,\cdot)) \}$. The electron conservation law that integrates the floor density constraint now transforms into a differential inclusion [8] which writes as

$$\partial_t n_e + \nabla\cdot \boldsymbol{f}_e \in S_e + S_\psi, \tag{6.7}$$

where $S_e + S_\psi$ denotes the set $\{ v + S_e \mid v \in S_\psi \}$.

6.3 Existence and uniqueness of the solution of the differential inclusion

We remark that the sign of S_e and $\partial_t n_e + \nabla\cdot \boldsymbol{f_e}$ in inclusion (6.7) is not arbitrarily chosen (we could have had otherwise $S_e \in \partial_t n_e + \nabla\cdot \boldsymbol{f_e} + S_\psi$). The existence and uniqueness of the solution of this nonlinear problem have been the subject of numerous studies in the field of monotone operators [25] or variational inequalities [93, 24, 76], and will be revisited in this section.

In [93], the existence and uniqueness of the solution of (6.7) was demonstrated with the use of a penalization problem, which is obtained by replacing the operator $-S_\psi(n_e(t,\cdot))$ in (6.7) with the operator

$$\frac{1}{\zeta}\widehat{B}(n_e(t,\cdot)) \equiv -\frac{1}{\zeta}(\psi - n_e(t,\cdot))^+, \tag{6.8}$$

which is in fact the Yosida approximation of $\partial I_{\mathcal{K}_\psi}$, with $(\psi - n_e(t,\cdot))^+$ the positive part[3] of $\psi - n_e(t,\cdot)$. In this section, we instead replace $-S_\psi$ with

$$\frac{1}{\zeta}B(n_e(t,\cdot)) \equiv \frac{1}{2\zeta}\left(\operatorname{sign}(n_e(t,\cdot) - \psi) - 1\right),$$
$$\operatorname{sign}(g) = \left\{ h \in \mathcal{H} \mid h(\boldsymbol{x}) \in \begin{cases} \{-1\}, & \text{if } g(\boldsymbol{x}) < 0 \\ [-1,1], & \text{if } g(\boldsymbol{x}) = 0 \\ \{1\}, & \text{if } g(\boldsymbol{x}) > 0 \end{cases} \text{for a.e. } \boldsymbol{x} \in \Omega \right\}. \tag{6.9}$$

In order to compare the two penalization methods in a simple way, let us assume that $\psi \in \mathbb{R}$ and consider the following ordinary differential inclusion problem,

$$\begin{cases} \dfrac{\mathrm{d}g(t)}{\mathrm{d}t} + \partial I_{\mathcal{K}_\psi}(g(t)) \ni 0, \quad t > 0, \\ g(0) = g_0 \in \mathbb{R}. \end{cases} \tag{6.10}$$

[3] $g^+(\boldsymbol{x}) = \max(g(\boldsymbol{x}), 0)$ for a.e. $\boldsymbol{x} \in \Omega$

The penalization problems corresponding to the penalizing operators (6.8) and (6.9) are resp.

$$\begin{cases} \dfrac{\mathrm{d}g(t)}{\mathrm{d}t} - \dfrac{1}{\zeta}(\psi - g(t))^+ = 0, \quad t > 0, \\ g(0) = g_0 \in \mathbb{R}, \end{cases} \tag{6.11}$$

and

$$\begin{cases} \dfrac{\mathrm{d}g(t)}{\mathrm{d}t} + \dfrac{1}{2\zeta}\left(\operatorname{sign}(g(t) - \psi) - 1\right) \ni 0, \quad t > 0, \\ g(0) = g_0 \in \mathbb{R}. \end{cases} \tag{6.12}$$

The solutions of the problems (6.10), (6.11) and (6.12) are resp.[4]

$$g(t) = \max(g_0, \psi),$$

$$\widehat{g}_\zeta(t) = \begin{cases} g_0, & \text{if } g_0 \geq \psi, \\ (g_0 - \psi)\mathrm{e}^{-\frac{t}{\zeta}} + \psi, & \text{if } g_0 < \psi, \end{cases}$$

$$g_\zeta(t) = \begin{cases} g_0, & \text{if } g_0 \geq \psi, \\ \max\left(g_0 + \dfrac{t}{\zeta}, \psi\right), & \text{if } g_0 < \psi. \end{cases}$$

We have then $\widehat{g}_\zeta \to g$, $g_\zeta \to g$ in $L^2(0, +\infty)$ as $\zeta \to 0$, but if $g_0 < \psi$ then $g_\zeta(t) = \psi$ after a finite time t_0 (more precisely $t_0 = \zeta(\psi - g_0)$) while $\widehat{g}_\zeta$ is always strictly smaller than ψ. This shows the difference between the proposed penalizing operator (6.9) and the "classic" penalizing operator (6.8).

6.3.1 Preliminaries

Let $\mathcal{H} = L^2(\Omega)$ be endowed with the inner product $(g, h)_\mathcal{H} \equiv \displaystyle\int_\Omega g(\boldsymbol{x})h(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ with $g, h \in \mathcal{H}$, and $\mathcal{V}$ be a Hilbert subspace of $\mathcal{H}$ such that $\mathcal{V} \subset \mathcal{H}$ with dense and compact embedding. We identify the linear functionals on $\mathcal{V}$ with $(\cdot, \cdot)_\mathcal{H}$ so that the embedding $\mathcal{H} \subset \mathcal{V}^*$ is also dense and compact [26, Chapter V]. Let $|\cdot|_\mathcal{H}$ be the norm induced by $(\cdot, \cdot)_\mathcal{H}$ on $\mathcal{H}$, $\|\cdot\|_\mathcal{V}$ be a norm on $\mathcal{V}$ and $\langle \cdot, \cdot \rangle$ the dual pairing between $\mathcal{V}$ and $\mathcal{V}^*$.

Let $\mathcal{K}$ be a closed convex of $\mathcal{V}$. We set $V = L^2(0, T; \mathcal{V})$, $H = L^2(0, T; \mathcal{H})$, $V^* = L^2(0, T; \mathcal{V}^*)$ and

$$K = \{\varphi \in V \mid \varphi(t) \in \mathcal{K} \text{ for a.e. } t \in (0, T)\}.$$

The differential inclusion (6.7) is addressed in a more general setting in the following way.

Let $a(\cdot, \cdot)$ be a continuous and coercive bilinear form on $\mathcal{V}$, i.e. there exists constants $M, C_1, C_2 > 0$ such that for all $g, h \in \mathcal{V}$,

$$\begin{aligned} |a(g, h)| &\leq M\|g\|_\mathcal{V}\|h\|_\mathcal{V}, \\ a(g, g) &\geq C_1\|g\|_\mathcal{V}^2 - C_2|g|_\mathcal{H}^2. \end{aligned} \tag{6.13}$$

[4] the solution of (6.10) is derived in section 6.5.3

Given $n_0 \in \mathcal{H}$, we aim to find a function $n \in V$ such that

$$\begin{cases} \forall g \in K, \quad \left\langle \dfrac{\mathrm{d}n}{\mathrm{d}t}, g - n \right\rangle + a(n, g - n) \geq 0, \quad \text{a.e. on } (0, T), \\ n(0) = n_0. \end{cases} \tag{6.14}$$

To this problem we associate the notions of strong and weak solutions.

Definition 6.3 (Strong solutions [24]). *A function n is called a strong solution of (6.14) if $n \in K$, $\dfrac{dn}{dt} \in V^*$ and n satisfies (6.14) for a.e. $t \in (0, T)$.*

Definition 6.4 (Weak solutions [24]). *A function n is called a weak solution of (6.14) if $n \in K$ and for all $g(t, \boldsymbol{x}) \in K$ such that $\dfrac{\mathrm{d}g}{\mathrm{d}t} \in V^*$, n satisfies*

$$\int_0^T \left(\left\langle \frac{\mathrm{d}g}{\mathrm{d}t}, g - n \right\rangle + a(n, g - n) \right) \mathrm{d}t \geq -\frac{1}{2} |g(0, \cdot) - n_0|_{\mathcal{H}}^2. \tag{6.15}$$

Let $A \in \mathcal{L}(\mathcal{V}, \mathcal{V}^*)$[5] defined by $\langle An, g \rangle = a(n, g)$, for $n, g \in \mathcal{V}$. The restriction of A into $\mathcal{H}$, still denoted as A, is characterized by $\mathrm{dom}(A) = \{g \in \mathcal{V} \mid Ag \in \mathcal{H}\}$.

Remark 6.1. *For inclusion (6.7), we suppose for simplicity that $S_e = (\alpha - \eta) N n_e$ with α, η the ionization and attachment coefficients and N the neutral particle density. Let $\mathcal{V} = H_0^1(\Omega)$ with the usual norm $\|g\|_{\mathcal{V}}^2 = |g|_{\mathcal{H}}^2 + \sum_{i=1}^{d} \left| \dfrac{\partial g}{\partial x_i} \right|_{\mathcal{H}}^2$, $\psi \in \mathcal{V}$ and $\mathcal{K} = \mathcal{K}_\psi$. Assume that $\boldsymbol{u}_e \in V^d$, $D_e \in V$ and $\alpha - \eta \in H$. From definition (6.4), inclusion (6.7) is equivalent to the following* ***variational inequality****,*

$$\forall g \in K, \quad \left\langle \frac{\mathrm{d}n_e}{\mathrm{d}t}, g - n \right\rangle + \langle \nabla \cdot (\boldsymbol{u}_e n_e - D_e \nabla n_e) - (\alpha - \eta) N n_e, g - n_e \rangle \geq 0.$$

Therefore, $An = \nabla \cdot (\boldsymbol{u}_e n_e - D_e \nabla n_e) - (\alpha - \eta) N n_e$ and $\mathrm{dom}(A) = \mathcal{H}_0^2(\Omega)$. By using integration by parts, we can identify $a(\cdot, \cdot)$ as

$$a(n, g) = \int_\Omega (-n \boldsymbol{u}_e \cdot \nabla g + D_e \nabla n \cdot \nabla g - (\alpha - \eta) N n g) \, \mathrm{d}\boldsymbol{x}.$$

Additionally, if we assume that $\boldsymbol{u}_e \in V^d \cap L^\infty \left(0, T; L^\infty(\Omega)^d\right)$, $D_e \in V \cap L^\infty(0, T; L^\infty(\Omega))$ with $D_e(t, \boldsymbol{x}) \geq c_D > 0$ a.e. on $(0, T) \times \Omega$ where c_D is a constant and $\alpha - \eta \in H \cap L^\infty(0, T; L^\infty(\Omega))$, we can verify that $a(\cdot, \cdot)$ is continuous on $\mathcal{V} \times \mathcal{V}$ since

$$|a(n, g)| \leq \left(|\boldsymbol{u}_e|_{L^\infty\left(0,T;L^\infty(\Omega)^d\right)} + |D_e|_{L^\infty(0,T;L^\infty(\Omega))} + |\alpha - \eta|_{L^\infty(0,T;L^\infty(\Omega))} N \right) \|n\|_{\mathcal{V}} \|g\|_{\mathcal{V}},$$

for any $n, g \in \mathcal{V}$. Further more, $a(\cdot, \cdot)$ is coercive on $\mathcal{V}$ according to definition (6.13). Indeed, for any $g \in \mathcal{V}$,

$$a(g, g) \geq -|\boldsymbol{u}_e|_{L^\infty\left(0,T;L^\infty(\Omega)^d\right)} |g|_{\mathcal{H}} \|g\|_{\mathcal{V}} + c_D C \|g\|_{\mathcal{V}}^2 - |\alpha - \eta|_{L^\infty(0,T;L^\infty(\Omega))} N |g|_{\mathcal{H}}^2,$$

[5] the space of linear maps from $\mathcal{V}$ to $\mathcal{V}^*$

where C is the positive constant from the Poincaré inequality. Thus, using the Cauchy-Schwarz inequality, we have

$$a(g,g) \geq \frac{c_D C}{2}\|g\|_{\mathcal{V}}^2 - \left(|\alpha-\eta|_{L^\infty(0,T;L^\infty(\Omega))} N + \frac{|\boldsymbol{u}_e|^2_{L^\infty\left(0,T;L^\infty(\Omega)^d\right)}}{2c_D C} \right) |g|_{\mathcal{H}}^2.$$

6.3.2 A penalization problem

For a small parameter $\zeta > 0$, we introduce the following penalization problem to (6.14),

$$\begin{cases} \dfrac{\mathrm{d}n}{\mathrm{d}t} + An + \dfrac{1}{\zeta} B(n) \ni 0 \quad \text{a.e. on } (0,T), \\ n(0) = n_0, \end{cases} \tag{6.16}$$

with $n_0 \in \mathcal{H}$ and $B(n)$ defined in (6.9).

The corresponding variational formulation of problem (6.16) writes as follows,

$$\begin{cases} \forall g \in \mathcal{V}, \quad \left\langle \dfrac{\mathrm{d}n}{\mathrm{d}t}, g \right\rangle + a(n,g) + \dfrac{1}{\zeta}\left(B(n), g\right)_{\mathcal{H}} \ni 0 \quad \text{a.e. on } (0,T), \\ n(0) = n_0, \end{cases} \tag{6.17}$$

where the last scalar product in the first equation is a notation abuse; it should be understood that $\left(B(n), g\right)_{\mathcal{H}} = \{\, (h,g)_{\mathcal{H}} \mid h \in B(n) \,\}$.

In addition, we can verify that $B(g)$ is the subdifferential of the functional G defined on $\mathcal{H}$ as follows,

$$G(g) = \int_\Omega (\psi - g)^+ \mathrm{d}\boldsymbol{x}, \tag{6.18}$$

where $(\psi - g)^+$ is the positive part of $\psi - g$. Indeed, for all $h \in B(g)$ and by (6.9),

$$(h, g-\psi)_{\mathcal{H}} = \frac{1}{2}\int_\Omega \left(|g-\psi| - (g-\psi)\right) \mathrm{d}\boldsymbol{x} = G(g).$$

Hence, for all $v \in \mathcal{H}$,

$$\begin{aligned} (h, v-g)_{\mathcal{H}} + G(g) = (h, v-\psi)_{\mathcal{H}} &= (h, (\psi-v)^-)_{\mathcal{H}} - (h, (\psi-v)^+)_{\mathcal{H}}, \\ &\leq (-h, (\psi-v)^+)_{\mathcal{H}} \leq \int_\Omega (\psi-v)^+ \mathrm{d}\boldsymbol{x} = G(v). \end{aligned}$$

where $(\psi-g)^-$ is the negative part[6] of $\psi - g$. The first inequality is due to the fact that $h(\boldsymbol{x}) \leq 0$ and the second is due to $h(\boldsymbol{x}) \geq -1$ for a.e. $\boldsymbol{x} \in \Omega$. Since G is convex, proper and lower semi-continuous, B is a maximal monotone operator with $\mathrm{dom}(B) = \mathcal{H}$ [25, Lemma 2.1].

[6] for $g \in \mathcal{H}$, $g^-(\boldsymbol{x}) = \max(-g(\boldsymbol{x}), 0)$ for a.e. $\boldsymbol{x} \in \Omega$

6.3.3 Weak solutions

The existence and uniqueness of the solution of the penalization problem (6.17) under reasonable conditions are presented in the following result (the proof is postponed to section 6.3.4).

Theorem 6.2. *Given $n_0 \in \mathcal{H}$ and $\zeta > 0$. Then the penalization problem (6.17) has a unique solution $n_\zeta \in H^1(0,T;\mathcal{V}^*) \cap V$ and the following estimate holds a.e. on $(0,T)$ and for all $g \in \mathcal{K}$,*

$$|n_\zeta(t,\cdot) - g|_{\mathcal{H}}^2 + \int_0^t \left(\frac{3C_1}{2} \|n_\zeta(s,\cdot) - g\|_{\mathcal{V}}^2 + \frac{2}{\zeta} \int_\Omega (\psi - n_\zeta(s,\cdot))^+ \mathrm{d}x \right) \mathrm{d}s \\ \leq \exp(2C_2 T) \left(|n_0 - g|_{\mathcal{H}}^2 + \frac{2M^2 T}{C_1} \|g\|_{\mathcal{V}}^2 \right), \quad (6.19)$$

where M, C_1, C_2 are the constants from (6.13).

A special property of the solution of (6.17) is that it is larger than ψ a.e. after a finite time, given the right conditions, even if the initial datum does not necessarily respect this constraint.

Proposition 6.1. *Assume that $\psi \in \mathrm{dom}(A)$, $n_0 \in \mathcal{H}$ such that $(\psi - n_0)^+ \in L^\infty(\Omega)$ and, furthermore,*

$$\forall g \in \mathcal{V}, \quad a(g, g^+) \geq C_1 \|g^+\|_{\mathcal{V}}^2 - C_2 |g^+|_{\mathcal{H}}^2, \quad (6.20)$$

where g^+ is the positive part of g and C_1, C_2 are the constants from (6.13).

For $\zeta > 0$, let n_ζ be the solution of the penalization problem (6.17). We have the following results,

- *if $n_0 \in \mathcal{K}$, then for any $\zeta > 0$ such that $\zeta A\psi \leq 1$ a.e. on Ω, $n_\zeta \geq \psi$ a.e. on $(0,T) \times \Omega$;*
- *if $n_0 \in \mathcal{H} \setminus \mathcal{K}$, given $\beta > 0$, then for any $\zeta > 0$ such that $\zeta A\psi \leq 1 - \beta$ a.e. on Ω, there exists a finite time $T_0 \geq 0$ such that $n_\zeta \geq \psi$ a.e. on $[T_0, T) \times \Omega$.*

The proof of this proposition is postponed to section 6.3.5.

Remark 6.2. *For the discharge problem (6.7) and ψ constant, the condition $\zeta A\psi \leq 1 - \beta$ a.e. is equivalent to $\zeta(\eta - \alpha)N\psi \leq 1 - \beta$ a.e. (see remark 6.1). Therefore, the penalization parameter ζ must satisfy $\zeta \leq \frac{1-\beta}{N} \left(\max\left(\sup_\Omega \eta - \alpha \right), 0 \right)^{-1}$ so that the penalizing source term $B(n)$ in (6.16) could be "strong" enough to fight back the depletion of electrons due to attachment.*

In turn of the variational problem (6.14), the existence and uniqueness of its **weak** solution are the subject of the following theorem (the proof is postponed to section 6.3.6).

Theorem 6.3. *Given $n_0 \in \mathcal{H}$. Then the variational problem (6.14) has a unique weak solution n in the sense of definition 6.4. Let n_ζ denote the solution in theorem 6.2 with $\zeta > 0$, then $n_\zeta \rightharpoonup n$ in V and $n_\zeta \to n$ in H as $\zeta \to 0$.*

6.3.4 Proof of theorem 6.2

Proof. The proof is divided into three parts: the first one shows the uniqueness of the solution, the second one demonstrates the existence of the solution for $n_0 \in \mathcal{V}$ and the last part concludes the existence of the solution for $n_0 \in \mathcal{H}$.

(i) Let n_ζ (resp. $\widetilde{n}_\zeta$) $\in H^1(0,T;\mathcal{V}^*) \cap V$ satisfy (6.17) with initial condition n_0 (resp. $\widetilde{n}_0$), then choosing $g = n_\zeta(t,\cdot) - \widetilde{n}_\zeta(t,\cdot) \in \mathcal{V}$ in (6.17), there exists $h_\zeta \in B\left(n_\zeta(t,\cdot)\right)$ and $\widetilde{h}_\zeta \in B\left(\widetilde{n}_\zeta(t,\cdot)\right)$ such that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left|n_\zeta - \widetilde{n}_\zeta\right|_{\mathcal{H}}^2 + a\left(n_\zeta - \widetilde{n}_\zeta, n_\zeta - \widetilde{n}_\zeta\right) + \frac{1}{\zeta}\left(h_\zeta - \widetilde{h}_\zeta, n_\zeta - \widetilde{n}_\zeta\right)_{\mathcal{H}} = 0.$$

Since B is monotone[7], $\left(h_\zeta - \widetilde{h}_\zeta, n_\zeta - \widetilde{n}_\zeta\right)_{\mathcal{H}} \geq 0$, so combining with (6.13) we deduce that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left|n_\zeta - \widetilde{n}_\zeta\right|_{\mathcal{H}}^2 + 2C_1\|n_\zeta - \widetilde{n}_\zeta\|_{\mathcal{V}}^2 - 2C_2|n_\zeta - \widetilde{n}_\zeta|_{\mathcal{H}}^2 \leq 0.$$

With $w(t) = \exp(2C_2 t)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{|n_\zeta - \widetilde{n}_\zeta|_{\mathcal{H}}^2}{w} \leq -\frac{2C_1\|n_\zeta - \widetilde{n}_\zeta\|_{\mathcal{V}}^2}{w},$$

therefore for a.e. $t \in (0,T)$,

$$|n_\zeta(t,\cdot) - \widetilde{n}_\zeta(t,\cdot)|_{\mathcal{H}}^2 + 2C_1\int_0^t \|n_\zeta(s,\cdot) - \widetilde{n}_\zeta(s,\cdot)\|_{\mathcal{V}}^2 \mathrm{d}s \leq \exp(2C_2T)|n_0 - \widetilde{n}_0|_{\mathcal{H}}^2. \tag{6.21}$$

It is clear that if $n_0 = \widetilde{n}_0$ a.e. on Ω then $n_\zeta = \widetilde{n}_\zeta$ a.e. on $(0,T)\times\Omega$.

(ii) Let us show now the existence of the solution of (6.17) with $n_0 \in \mathcal{V}$ by adapting the proof of Ito & Kunisch [76, Theorem 1]. We consider the implicit Euler discretization for (6.16) with a constant timestep $\Delta t > 0$,

$$\frac{n_\zeta^l - n_\zeta^{l-1}}{\Delta t} + An_\zeta^l + \frac{1}{\zeta}B(n_\zeta^l) \ni 0, \tag{6.22}$$

with $n_\zeta^0 = n_0 \in \mathcal{V}$, $l = 1,\ldots,\mathcal{L}$ and $\mathcal{L} = \dfrac{T}{\Delta t}$. The bilinear form $(\cdot,\cdot)_{\mathcal{H}} + 2\Delta t a(\cdot,\cdot)$ with $\Delta t \leq \dfrac{1}{2C_2}$ is continuous, coercive on $\mathcal{V}$ so by the Lax-Milgram theorem [26, Chapter V], $I + 2\Delta t A$ is a maximal monotone operator of $\mathcal{H}$. Therefore in the light of lemma C.1 with $M := \dfrac{2\Delta t}{\zeta}B$ and $N := \mathrm{Id} + 2\Delta t A$ (note that $\mathrm{dom}(M) = \mathcal{H}$), the operator $\mathrm{Id} + 2\Delta t A + \dfrac{2\Delta t}{\zeta}B$ is maximal monotone on $\mathcal{H}$. Consequently, for each $l > 0$ there exists (a unique) $n_\zeta^l \in \mathrm{dom}(A) \subset \mathcal{V}$ that satisfies (6.22).

Multiplying (6.22) with $n_\zeta^l - g \in \mathcal{V}$, $g \in \mathcal{K}$, we have

$$\left(\frac{n_\zeta^l - n_\zeta^{l-1}}{\Delta t}, n_\zeta^l - g\right)_{\mathcal{H}} + a(n_\zeta^l, n_\zeta^l - g) + \frac{1}{\zeta}(B(n_\zeta^l), n_\zeta^l - g)_{\mathcal{H}} \ni 0,$$

[7] for the definition of monotone and maximal monotone operators, see appendix C

i.e. in particular, there exists $h_\zeta^l \in B(n_\zeta^l)$ such that

$$\left(\frac{n_\zeta^l - n_\zeta^{l-1}}{\Delta t}, n_\zeta^l - g\right)_{\mathcal{H}} + a(n_\zeta^l, n_\zeta^l - g) + \frac{1}{\zeta}(h_\zeta^l, n_\zeta^l - g)_{\mathcal{H}} = 0. \tag{6.23}$$

Since

$$2(n_\zeta^l - n_\zeta^{l-1}, n_\zeta^l - g)_{\mathcal{H}} = |n_\zeta^l - n_\zeta^{l-1}|_{\mathcal{H}}^2 + |n_\zeta^l - g|_{\mathcal{H}}^2 - |n_\zeta^{l-1} - g|_{\mathcal{H}}^2,$$

$$\begin{aligned} a(n_\zeta^l, n_\zeta^l - g) &\geq a(n_\zeta^l - g, n_\zeta^l - g) - |a(g, n_\zeta^l - g)| \\ &\geq C_1\|n_\zeta^l - g\|_{\mathcal{V}}^2 - C_2|n_\zeta^l - g|_{\mathcal{H}}^2 - M\|g\|_{\mathcal{V}}\|n_\zeta^l - g\|_{\mathcal{V}} \\ &\geq \frac{3C_1}{4}\|n_\zeta^l - g\|_{\mathcal{V}}^2 - C_2|n_\zeta^l - g|_{\mathcal{H}}^2 - \frac{M^2}{C_1}\|g\|_{\mathcal{V}}^2, \end{aligned} \tag{6.24}$$

and since $g \geq \psi$, $h_\zeta^l \leq 0$ a.e. on Ω,

$$(h_\zeta^l, n_\zeta^l - g)_{\mathcal{H}} = \int_\Omega (\psi - n_\zeta^l)^+ \mathrm{d}\boldsymbol{x} + \int_\Omega h_\zeta^l(\psi - g)\mathrm{d}\boldsymbol{x} \geq \int_\Omega (\psi - n_\zeta^l)^+ \mathrm{d}\boldsymbol{x},$$

we have, by multiplying eq. (6.23) with $2\Delta t$,

$$\begin{aligned} &(1 - 2C_2\Delta t)|n_\zeta^l - g|_{\mathcal{H}}^2 - |n_\zeta^{l-1} - g|_{\mathcal{H}}^2 + |n_\zeta^l - n_\zeta^{l-1}|_{\mathcal{H}}^2 \\ &\qquad + \frac{3C_1\Delta t}{2}\|n_\zeta^l - g\|_{\mathcal{V}}^2 + \frac{2\Delta t}{\zeta}\int_\Omega (\psi - n_\zeta^l)^+ \mathrm{d}\boldsymbol{x} \leq \frac{2M^2\Delta t}{C_1}\|g\|_{\mathcal{V}}^2. \end{aligned} \tag{6.25}$$

Let $w^l = (1 - 2C_2\Delta t)^{-l}$ for $l = 0, \ldots, \mathcal{L}$. For any arbitrary sequence of real numbers $(v^l)_{l=0,\ldots,\mathcal{L}}$, we have

$$\frac{v^l}{w^l} - \frac{v^{l-1}}{w^{l-1}} = \frac{(1 - 2C_2\Delta t)v^l - v^{l-1}}{w^{l-1}}.$$

Using this equality with $v^l = |n_\zeta^l - g|_{\mathcal{H}}^2$ and the fact that $1 = w^0 \leq w^l \leq w^{\mathcal{L}} = \left(1 - 2C_2\frac{T}{\mathcal{L}}\right)^{-\mathcal{L}} \leq \exp(2C_2T)$, while dividing (6.25) by w^l and summing it over $k = 1, \ldots, l$, we now have, for all $l = 1, \ldots, \mathcal{L}$,

$$\begin{aligned} &\sum_{k=1}^{l}\left(\frac{|n_\zeta^k - g|_{\mathcal{H}}^2}{w^k} - \frac{|n_\zeta^{k-1} - g|_{\mathcal{H}}^2}{w^{k-1}} + \frac{|n_\zeta^k - n_\zeta^{k-1}|_{\mathcal{H}}^2}{w^k} + \frac{3C_1\Delta t}{2w^k}\|n_\zeta^k - g\|_{\mathcal{V}}^2\right. \\ &\qquad \left. + \frac{2\Delta t}{\zeta w^k}\int_\Omega (\psi - n_\zeta^k)^+ \mathrm{d}\boldsymbol{x}\right) \leq \sum_{k=1}^{l}\frac{2M^2\Delta t}{C_1 w^k}\|g\|_{\mathcal{V}}^2 \leq \frac{2M^2 l\Delta t}{C_1}\|g\|_{\mathcal{V}}^2 \leq \frac{2M^2T}{C_1}\|g\|_{\mathcal{V}}^2. \end{aligned}$$

Therefore, using the fact that $\frac{w^l}{w^k} \geq 1$ for $k \leq l$,

$$\begin{aligned} &|n_\zeta^l - g|_{\mathcal{H}}^2 + \sum_{k=1}^{l}|n_\zeta^k - n_\zeta^{k-1}|_{\mathcal{H}}^2 + \Delta t\sum_{k=1}^{l}\left(\frac{3C_1}{2}\|n_\zeta^k - g\|_{\mathcal{V}}^2 + \frac{2}{\zeta}\int_\Omega (\psi - n_\zeta^k)^+ \mathrm{d}\boldsymbol{x}\right) \\ &\qquad \leq w^l\left(\frac{|n_0 - g|_{\mathcal{H}}^2}{w^0} + \frac{2M^2T}{C_1}\|g\|_{\mathcal{V}}^2\right) \leq \exp(2C_2T)\left(|n_0 - g|_{\mathcal{H}}^2 + \frac{2M^2T}{C_1}\|g\|_{\mathcal{V}}^2\right). \end{aligned} \tag{6.26}$$

Let us define the functions $n_\zeta^{\Delta t}(t,\boldsymbol{x})$, $\widehat{n}_\zeta^{\Delta t}(t,\boldsymbol{x})$ and $h_\zeta^{\Delta t}(t,\boldsymbol{x})$ in the following way,

$$\begin{cases} n_\zeta^{\Delta t}(t,\cdot) = n_\zeta^l \\ \widehat{n}_\zeta^{\Delta t}(t,\cdot) = n_\zeta^l + \dfrac{t-l\Delta t}{\Delta t}(n_\zeta^l - n_\zeta^{l-1}) \qquad \text{for a.e. } t \in ((l-1)\Delta t, l\Delta t], \quad l \geq 1. \\ h_\zeta^{\Delta t}(t,\cdot) = h_\zeta^l \end{cases} \tag{6.27}$$

Inequality (6.26) implies that $\Delta t \sum_{l=1}^{\mathcal{L}} \|n_\zeta^k - g\|_{\mathcal{V}}^2$ is bounded independently of Δt, and since $n_0, n_\zeta^l, g \in \mathcal{V}$ for $l = 1, \dots, \mathcal{L}$ as well as

$$\begin{aligned} \int_0^T \|\widehat{n}_\zeta^{\Delta t}(t,\cdot)\|_{\mathcal{V}}^2 \mathrm{d}t &= \sum_{l=1}^{\mathcal{L}} \int_{(l-1)\Delta t}^{l\Delta t} \left(\|n_\zeta^l\|_{\mathcal{V}}^2 + \left(\frac{t-l\Delta t}{\Delta t}\right)^2 \|n_\zeta^l - n_\zeta^{l-1}\|_{\mathcal{V}}^2 \right) \mathrm{d}t \\ &= \Delta t \sum_{l=1}^{\mathcal{L}} \left(\|n_\zeta^l\|_{\mathcal{V}}^2 + \frac{1}{3}\|n_\zeta^l - n_\zeta^{l-1}\|_{\mathcal{V}}^2 \right) \\ &\leq \Delta t \sum_{l=1}^{\mathcal{L}} \left(\|n_\zeta^l - g\|_{\mathcal{V}}^2 + \|g\|_{\mathcal{V}}^2 + \frac{1}{3}\|n_\zeta^l - g\|_{\mathcal{V}}^2 + \frac{1}{3}\|n_\zeta^{l-1} - g\|_{\mathcal{V}}^2 \right) \\ &= T\|g\|_{\mathcal{V}}^2 + \frac{4\Delta t}{3}\|n_\zeta^{\mathcal{L}} - g\|_{\mathcal{V}}^2 + \frac{\Delta t}{3}\|n_0 - g\|_{\mathcal{V}}^2 + \Delta t \sum_{l=1}^{\mathcal{L}-1} \left(\frac{5}{3}\|n_\zeta^l - g\|_{\mathcal{V}}^2 \right), \end{aligned}$$

we deduce that $\widehat{n}_\zeta^{\Delta t}$ is bounded in V independently of Δt. Moreover, from inclusion (6.22) we see that for any $v \in \mathcal{V}$,

$$\left\langle \frac{n_\zeta^l - n_\zeta^{l-1}}{\Delta t}, v \right\rangle \leq |a(n_\zeta^l, v)| + \frac{1}{\zeta}|(h_\zeta^l, v)_{\mathcal{H}}| \leq M\|n_\zeta^l\|_{\mathcal{V}}\|v\|_{\mathcal{V}} + \frac{1}{\zeta}|\Omega|^{\frac{1}{2}}|v|_{\mathcal{H}}, \tag{6.28}$$

with $h_\zeta^l \in B(n_\zeta^l)$ and since $|h_\zeta^l| \leq 1$ a.e. on Ω. Since $\mathcal{V}$ is continuously embedded in $\mathcal{H}$, there exists a constant $C > 0$ such that $|v|_{\mathcal{H}} \leq C\|v\|_{\mathcal{V}}$ for any v. Thus, $\widehat{n}_\zeta^{\Delta t}$ is bounded in $H^1(0,T;\mathcal{V}^*)$ uniformly in Δt (since Ω is bounded) and consequently, for any sequence $\left(\widehat{n}_\zeta^{\Delta t}\right)_{\Delta t}$ with fixed ζ and $\Delta t \to 0$, there exists a subsequence, still denoted as the same, such that $\widehat{n}_\zeta^{\Delta t}$ converge weakly to a function n_ζ in $H^1(0,T;\mathcal{V}^*) \cap V$ as $\Delta t \to 0$. Using theorem C.3 (Aubin-Lions-Simon) with $B_0 = \mathcal{V}$, $B_1 = \mathcal{H}$ and $B_2 = \mathcal{V}^*$, we have in addition that $\widehat{n}_\zeta^{\Delta t}$ converge strongly to n_ζ in H as $\Delta t \to 0$.

Now providing that

$$\int_0^T |n_\zeta^{\Delta t} - \widehat{n}_\zeta^{\Delta t}|_{\mathcal{H}}^2 \mathrm{d}t = \frac{\Delta t}{3} \sum_{l=1}^{\mathcal{L}} |n_\zeta^l - n_\zeta^{l-1}|_{\mathcal{H}}^2 \to 0,$$

as $\Delta t \to 0$ (a consequence of estimate (6.26)), we also have $n_\zeta^{\Delta t} \to n_\zeta$ in H. Furthermore, as $\Delta t \to 0$,

$$\begin{aligned} \int_0^T \int_\Omega \left| (\psi - n_\zeta^{\Delta t})^+ - (\psi - n_\zeta)^+ \right| \mathrm{d}\boldsymbol{x}\mathrm{d}t &\leq \int_0^T \int_\Omega \left| n_\zeta^{\Delta t} - n_\zeta \right| \mathrm{d}\boldsymbol{x}\mathrm{d}t \\ &\leq |\Omega|^{\frac{1}{2}} \int_0^T |n_\zeta^{\Delta t} - n_\zeta|_{\mathcal{H}} \mathrm{d}t \leq T^{\frac{1}{2}}|\Omega|^{\frac{1}{2}} \int_0^T |n_\zeta^{\Delta t} - n_\zeta|_{\mathcal{H}}^2 \mathrm{d}t \to 0, \end{aligned} \tag{6.29}$$

we conclude, by letting $\Delta t \to 0$ in estimate (6.26) and using the fact that $\|n_\zeta(t,\cdot) - g\|_{\mathcal{V}} \leq \liminf_{\Delta t \to 0} \|n_\zeta^{\Delta t}(t,\cdot) - g\|_{\mathcal{V}}$ a.e. on $(0,T)$, that estimate (6.19) holds for n_ζ.

On the other hand, $h_\zeta^{\Delta t}(t,\cdot)$ is bounded in $L^\infty(\Omega)$ a.e. on $(0,T)$ and in particular in $\mathcal{H}$ since Ω is bounded. As a consequence, we can extract a subsequence for fixed ζ, still denoted as $\left(h_\zeta^{\Delta t}(t,\cdot)\right)_{\Delta t}$, such that $h_\zeta^{\Delta t}(t,\cdot) \rightharpoonup y_\zeta$ in $\mathcal{H}$ as $\Delta t \to 0$. We now check if $y_\zeta \in B(n_\zeta(t,\cdot))$. By property (6.2) of subdifferentials as well as definition (6.18) of the functional G, we have

$$\forall v \in \mathcal{H}, \quad G(v) - G(n_\zeta^{\Delta t}(t,\cdot)) \geq (h_\zeta^{\Delta t}(t,\cdot), v - n_\zeta^{\Delta t}(t,\cdot))_{\mathcal{H}}.$$

Using (6.29) and letting $\Delta t \to 0$, we have

$$\forall v \in \mathcal{H}, \quad G(v) - G(n_\zeta(t,\cdot)) \geq (y_\zeta, v - n_\zeta(t,\cdot))_{\mathcal{H}},$$

i.e. $y_\zeta \in B(n_\zeta(t,\cdot))$. Therefore, by multiplying inclusion (6.22) with any $v \in \mathcal{V}$, i.e.

$$\left\langle \frac{\mathrm{d}\widehat{n}_\zeta^{\Delta t}(t,\cdot)}{\mathrm{d}t}, v \right\rangle + a\left(n_\zeta^{\Delta t}(t,\cdot), v\right) + \frac{1}{\zeta}\left(h_\zeta^{\Delta t}(t,\cdot), v\right)_{\mathcal{H}} = 0,$$

and letting $\Delta t \to 0$, we conclude that inclusion (6.17) holds a.e. on $(0,T)$ for $n_0 \in \mathcal{V}$.

(iii) Finally, for each $n_0 \in \mathcal{H}$ there exists a sequence $(n_{0,p})_p$ such that $n_{0,p} \in \mathcal{V}$ and $n_{0,p} \to n_0$ in $\mathcal{H}$ as $p \to \infty$ since $\mathcal{V}$ is dense in $\mathcal{H}$. Let $n_{\zeta,p}$ be the (unique) solution of inclusion (6.17) corresponding to the initial datum $n_{0,p}$. From estimate (6.21), we infer that $(n_{\zeta,p})_p$ is a Cauchy sequence in V, therefore there exists an $n_\zeta \in V$ such that $n_{\zeta,p} \to n_\zeta$ in V as $p \to \infty$ and consequently $\frac{\mathrm{d}n_{\zeta,p}}{\mathrm{d}t} \to \frac{\mathrm{d}n_\zeta}{\mathrm{d}t}$ in the distributional sense.

Moreover, since $\frac{\mathrm{d}n_{\zeta,p}}{\mathrm{d}t}$ is bounded in $L^2(0,T;\mathcal{V}^*)$ for fixed ζ (see estimate (6.28)), there exists a function $g \in L^2(0,T;\mathcal{V}^*)$ such that $\frac{\mathrm{d}n_{\zeta,p}}{\mathrm{d}t} \rightharpoonup g$ in $L^2(0,T;\mathcal{V}^*)$. Hence, we identify $\frac{\mathrm{d}n_\zeta}{\mathrm{d}t} = g$ and as a result, $n_\zeta \in H^1(0,T;\mathcal{V}^*) \cap V$. We recover (6.17) and (6.19) for n_ζ by just letting $p \to \infty$. ■

Corollary 6.1. *Similar to the case $n_0 \in \mathcal{V}$, in the case $n_0 \in \mathcal{H}$ we also have $n_\zeta^{\Delta t} \to n_\zeta$ in H as $\Delta t \to 0$, where $n_\zeta^{\Delta t}$ is the piecewise-constant approximation of n_ζ defined as in (6.22) and (6.27).*

Proof. Let $n_{\zeta,p}^{\Delta t}$ be the piecewise-constant approximation of $n_{\zeta,p}$ defined as in (6.22) and (6.27) where $(n_{\zeta,p})_p$ is a sequence in V having n_ζ as an accumulation point. From inclusion (6.22), there exists $h_{\zeta,p}^{\Delta t} \in B\left(n_{\zeta,p}^{\Delta t}\right)$ and $h_\zeta^{\Delta t} \in B\left(n_\zeta^{\Delta t}\right)$ such that, for any $g \in \mathcal{V}$,

$$\left(\frac{n_{\zeta,p}^l - n_{\zeta,p}^{l-1}}{\Delta t}, g\right)_{\mathcal{H}} + a(n_{\zeta,p}^l, g) + \frac{1}{\zeta}(h_{\zeta,p}^l, g)_{\mathcal{H}} = 0,$$
$$\left(\frac{n_\zeta^l - n_\zeta^{l-1}}{\Delta t}, g\right)_{\mathcal{H}} + a(n_\zeta^l, g) + \frac{1}{\zeta}(h_\zeta^l, g)_{\mathcal{H}} = 0.$$

Choosing $g = n_{\zeta,p}^l - n_\zeta^l$ with $l \geq 1$, subtracting the two equations and using the fact that B is monotone, we have

$$\left(\frac{n_{\zeta,p}^l - n_{\zeta,p}^{l-1}}{\Delta t} - \frac{n_\zeta^l - n_\zeta^{l-1}}{\Delta t}, n_{\zeta,p}^l - n_\zeta^l \right)_{\mathcal{H}} + a\left(n_{\zeta,p}^l - n_\zeta^l, n_{\zeta,p}^l - n_\zeta^l\right) \leq 0,$$

from which can be further deduced that

$$(1 - 2C_2\Delta t)|n_{\zeta,p}^l - n_\zeta^l|_{\mathcal{H}}^2 - |n_{\zeta,p}^{l-1} - n_\zeta^{l-1}|_{\mathcal{H}}^2 + 2C_1\Delta t\|n_{\zeta,p}^l - n_\zeta^l\|_{\mathcal{V}}^2 \leq 0.$$

Therefore, by proceeding as in the derivation of estimate (6.26), for a.e. $t \in (0,T)$ we have

$$|n_{\zeta,p}^{\Delta t}(t,\cdot) - n_\zeta^{\Delta t}(t,\cdot)|_{\mathcal{H}}^2 + 2C_1 \int_0^t \|n_{\zeta,p}^{\Delta t}(s,\cdot) - n_\zeta^{\Delta t}(s,\cdot)\|_{\mathcal{V}}^2 \mathrm{d}s \leq \exp(2C_2T)|n_{0,p} - n_0|_{\mathcal{H}}^2,$$

which is a discrete version of estimate (6.21). This suggests that $n_{\zeta,p}^{\Delta t} \to n_\zeta^{\Delta t}$ in V as $p \to \infty$. We already established that $n_{\zeta,p}^{\Delta t} \to n_{\zeta,p}$ in H and $n_{\zeta,p} \to n_\zeta$ in V as $\Delta t \to 0$ and the convergences are in subsequence. Therefore, we could conclude that $n_\zeta^{\Delta t} \to n_\zeta$ in subsequence in H as $\Delta t \to 0$. ■

Remark 6.3. *We point out that the weak convergence of $\widetilde{n}_\zeta^{\Delta t}$ to n_ζ in $H^1(0,T;\mathcal{V}^*) \cap V$ and subsequently the strong convergence in H do not hold for $n_0 \notin \mathcal{V}$, since $\widetilde{n}_\zeta^{\Delta t}(t,\cdot)$ does not a priori belong to $\mathcal{V}$ for $t \in [0,\Delta t)$. Hence, the proceeding from $n_0 \in \mathcal{V}$ to $n_0 \in \mathcal{H}$ is necessary in the proof of theorem 6.2.*

6.3.5 Proof of proposition 6.1

Proof. *(i)* If $n_0 \in \mathcal{K}$, we note that ψ satisfies the inclusion

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} + A\psi + \frac{1}{\zeta}B\psi \ni A\psi - \frac{1}{\zeta}.$$

Together with inequality (6.17) and the assumption that $\zeta A\psi \leq 1$, we have

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|(\psi - n_\zeta)^+|_{\mathcal{H}}^2 + a(\psi - n_\zeta, (\psi - n_\zeta)^+) \leq \left(A\psi - \frac{1}{\zeta}, (\psi - n_\zeta)^+\right)_{\mathcal{H}} \leq 0.$$

Hence, there exists a constant $C > 0$ such that $(\psi - n_\zeta)^+ \leq C(\psi - n_0)^+ = 0$ a.e. on $(0,T) \times \Omega$, which means that $n_\zeta \in K$ since $n_0 \in \mathcal{K}$.

(ii) If $n_0 \in \mathcal{H} \setminus \mathcal{K}$, let us consider the following function,

$$w(t,\boldsymbol{x}) = \frac{\beta t}{\zeta} - |(\psi - n_0)^+|_{L^\infty(\Omega)} + \psi(\boldsymbol{x}).$$

On $(0,T_0)$, with $T_0 = \dfrac{\zeta}{\beta}|(\psi - n_0)^+|_{L^\infty(\Omega)} > 0$, we have $w \leq \psi$. Therefore, w satisfies

$$\begin{cases} \dfrac{\mathrm{d}w}{\mathrm{d}t} + Aw + \dfrac{1}{\zeta}Bw \ni A\psi + \dfrac{\beta - 1}{\zeta}, \\ w(0) = -|(\psi - n_0)^+|_{L^\infty(\Omega)} + \psi. \end{cases}$$

Combining with inequality (6.17), we have the following result a.e. on $(0, T_0)$,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|(w-n_\zeta)^+|^2_{\mathcal{H}} + a(w-n_\zeta,(w-n_\zeta)^+) \leq \left(A\psi + \frac{\beta-1}{\zeta}, (w-n_\zeta)^+\right)_{\mathcal{H}}.$$

Then if ζ satisfies $\zeta A\psi \leq 1-\beta$, the right-hand-side is negative and we deduce that there exists a constant $C>0$ such that

$$(w-n_\zeta)^+ \leq C(w(0,\cdot)-n_0)^+ = C(-|(\psi-n_0)^+|_{L^\infty(\Omega)} + \psi - n_0)^+ = 0.$$

In particular, $n_\zeta(T_0,\cdot) \geq w(T_0,\cdot) = \psi$, and we arrive at the conclusion by applying the result of the case $n_0 \in \mathcal{K}$ on the penalization problem (6.17) with $t \in (T_0, T)$ and the initial datum $n_0 := n_\zeta(T_0)$. ■

6.3.6 Proof of theorem 6.3

Proof. Estimate (6.19) suggests that n_ζ is bounded in V as well as $\displaystyle\int_0^T\int_\Omega (\psi-n_\zeta)^+ \mathrm{d}\boldsymbol{x}\mathrm{d}t \to 0$ as $\zeta \to 0$. Therefore, there exists $n \in V$ such that $n_\zeta \rightharpoonup n$ in subsequence in V. Moreover, $n_\zeta \to n$ in subsequence in H according to theorem C.3 (Aubin-Lions-Simon) as $\dfrac{\mathrm{d}n_\zeta}{\mathrm{d}t} \in L^2(0,T;\mathcal{V}^*)$. We also have $n \in K$ since $(\psi-n)^+ = 0$ a.e. on Ω.

We now prove that n satisfies inequality (6.15). For each $g \in K$ such that $\dfrac{\mathrm{d}g}{\mathrm{d}t} \in V^*$, replacing $g := g - n_\zeta$ in inclusion (6.17) and integrating it on $(0,T)$ yield

$$\int_0^T \left(\left\langle \frac{\mathrm{d}n_\zeta}{\mathrm{d}t}, g-n_\zeta\right\rangle + a(n_\zeta, g-n_\zeta) + \frac{1}{\zeta}(h_\zeta, g-n_\zeta)_{\mathcal{H}}\right)\mathrm{d}t = 0,$$

with $h_\zeta \in H$ such that $h_\zeta(t,\cdot) \ni B(n_\zeta(t,\cdot))$. For the first term in the above equation, we have

$$\begin{aligned}\int_0^T \left\langle \frac{\mathrm{d}n_\zeta}{\mathrm{d}t}, g-n_\zeta\right\rangle \mathrm{d}t &= \int_0^T \left\langle \frac{\mathrm{d}g}{\mathrm{d}t}, g-n_\zeta\right\rangle \mathrm{d}t + \frac{1}{2}|g(0,\cdot)-n_0|^2_{\mathcal{H}} - \frac{1}{2}|g(T,\cdot)-n_\zeta(T,\cdot)|^2_{\mathcal{H}} \\ &\leq \int_0^T \left\langle \frac{\mathrm{d}g}{\mathrm{d}t}, g-n_\zeta\right\rangle \mathrm{d}t + \frac{1}{2}|g(0,\cdot)-n_0|^2_{\mathcal{H}}.\end{aligned}$$

For the second term, we have

$$\begin{aligned}a(n_\zeta,n_\zeta) &= a(n_\zeta,n) + a(n,n_\zeta) - a(n,n) + a(n_\zeta-n, n_\zeta-n) \\ &\geq a(n_\zeta,n) + a(n,n_\zeta) - a(n,n) + C_1\|n_\zeta-n\|^2_{\mathcal{V}} - C_2|n_\zeta-n|^2_{\mathcal{H}}.\end{aligned}$$

As $n_\zeta(t,\cdot) \rightharpoonup n(t,\cdot)$ in $\mathcal{V}$ and $n_\zeta(t,\cdot) \to n(t,\cdot)$ in $\mathcal{H}$, taking the $\liminf$ of the above inequality and using $\liminf_{\zeta\to 0} \|n_\zeta - n\|^2_{\mathcal{V}} \geq 0$ yield

$$\liminf_{\zeta\to 0} a(n_\zeta, n_\zeta) \geq a(n,n).$$

For the third term, based on the fact that $h_\zeta \leq 0$ a.e. and $g \geq \psi$ a.e., we simply have

$$(h_\zeta(t,\cdot), g(t,\cdot) - n_\zeta(t,\cdot))_{\mathcal{H}} = -G(n_\zeta(t,\cdot)) + (h_\zeta(t,\cdot), g(t,\cdot) - \psi)_{\mathcal{H}} \leq 0.$$

Consequently,

$$\int_0^T \left(\left\langle \frac{\mathrm{d}g}{\mathrm{d}t}, g - n_\zeta \right\rangle + a(n_\zeta, g) - a(n,n) \right) \mathrm{d}t \geq -\frac{1}{2}|g(0,\cdot) - n_0|^2_{\mathcal{H}},$$

and letting $\zeta \to 0$ leads to (6.19).

Lastly, let n (resp. $\widetilde{n}$) be a solution of problem (6.15) with initial datum $n_0 \in \mathcal{H}$ (resp. $\widetilde{n}_0 \in \mathcal{H}$). Let (ζ) be a sequence of small positive numbers and n_ζ (resp. $\widetilde{n}_\zeta$) be the solution of problem (6.17) with initial datum n_0 (resp. $\widetilde{n}_0$). Since $n_\zeta \rightharpoonup n$ (resp. $\widetilde{n}_\zeta \rightharpoonup \widetilde{n}$) in subsequence in V as $\zeta \to 0$, we infer from (6.21) that

$$|n(t,\cdot) - \widetilde{n}(t,\cdot)|^2_{\mathcal{H}} + 2C_1 \int_0^t \|n(s,\cdot) - \widetilde{n}(s,\cdot)\|^2_{\mathcal{V}} \mathrm{d}s \leq \exp(2C_2 T)|n_0 - \widetilde{n}_0|^2_{\mathcal{H}}, \tag{6.30}$$

which also shows the uniqueness of the solution of problem (6.15). ■

Another proof of the existence of the solution of inequality (6.15) could be inspired by [76, Theorem 5], and shown in the following result.

Corollary 6.2. *Let n be the solution of inequality (6.15). For each $\Delta t > 0$, there exists $n^{\Delta t}$ defined in the similar way as in (6.27) with the piecewise values n^l satisfying, for a.e. t,*

$$\forall g \in K, \quad \left\langle \frac{n^l - n^{l-1}}{\Delta t}, g(t,\cdot) - n^l \right\rangle + a(n^l, g(t,\cdot) - n^l) \geq 0, \tag{6.31}$$

for $l = 1, \ldots, \mathcal{L}$, such that $n^{\Delta t}$ converges to n in subsequence as $\Delta t \to 0$, weakly in V and strongly in H.

Proof. The idea is that from estimate (6.26) of piecewise-constant approximations of n_ζ, for fixed Δt and for each $l \geq 1$, n^l_ζ is bounded in $\mathcal{V}$ uniformly in ζ, so there exists $n^l \in \mathcal{V}$ such that $n^l_\zeta \rightharpoonup n^l$ in $\mathcal{V}$ and consequently, $n^l_\zeta \to n^l$ in $\mathcal{H}$ as $\zeta \to 0$ since $\mathcal{V}$ is compactly embedded in $\mathcal{H}$. Combining with the fact that $\frac{1}{\zeta}\int_\Omega (\psi - n^l_\zeta)^+ \mathrm{d}\boldsymbol{x}$ is bounded uniformly in ζ, we deduce that $n^l \in \mathcal{K}$ for $l \geq 1$. We could verify, by letting $\zeta \to 0$ in eq. (6.23), that n^l satisfies inequality (6.31) which is the discrete version of inclusion (6.14). Note that if Δt is small enough then according to the Stampacchia theorem [26, Theorem V.6], n^l is the unique solution of (6.31).

In the case $n_0 \in \mathcal{V}$, we define the approximations $n^{\Delta t}$ and $\widehat{n}^{\Delta t}$ of n in the similar way as in (6.27). There exists $\widetilde{n} \in V$ such that $n^{\Delta t}, \widehat{n}^{\Delta t}$ converge to $\widetilde{n}$ in subsequence as $\Delta t \to 0$, weakly in V since they are bounded in V uniformly in Δt and strongly in H since $\dfrac{\mathrm{d}\widehat{n}^{\Delta t}}{\mathrm{d}t} \in H \subset L^2(0,T;\mathcal{V}^*)$ (consequences of estimate (6.26) in the limit $\zeta \to 0$). Furthermore, $\widetilde{n} \in K$ since $n^{\Delta t} \in K$.

Substituting $n^{\Delta t}$ and $\widehat{n}^{\Delta t}$ in inequality (6.31) yields,

$$\left\langle \frac{\mathrm{d}\widehat{n}^{\Delta t}}{\mathrm{d}t}, g - n^{\Delta t} \right\rangle + a(n^{\Delta t}, g - n^{\Delta t}) \geq 0,$$

for a.e. t. In other words,

$$\left\langle \frac{\mathrm{d}g}{\mathrm{d}t}, g - \widehat{n}^{\Delta t} \right\rangle + \left\langle \frac{\mathrm{d}\widehat{n}^{\Delta t}}{\mathrm{d}t} - \frac{\mathrm{d}g}{\mathrm{d}t}, g - \widehat{n}^{\Delta t} \right\rangle + \left\langle \frac{\mathrm{d}\widehat{n}^{\Delta t}}{\mathrm{d}t}, \widehat{n}^{\Delta t} - n^{\Delta t} \right\rangle + a(n^{\Delta t}, g - n^{\Delta t}) \geq 0.$$

Since

$$\int_0^T \left\langle \frac{\mathrm{d}\widehat{n}^{\Delta t}}{\mathrm{d}t} - \frac{\mathrm{d}g}{\mathrm{d}t}, g - \widehat{n}^{\Delta t} \right\rangle \mathrm{d}t = \frac{1}{2}|n_0 - g(0,\cdot)|_{\mathcal{H}}^2 - \frac{1}{2}|n^{\mathcal{L}} - g(T,\cdot)|_{\mathcal{H}}^2 \leq \frac{1}{2}|n_0 - g(0,\cdot)|_{\mathcal{H}}^2$$

$$\int_0^T \left\langle \frac{\mathrm{d}\widehat{n}^{\Delta t}}{\mathrm{d}t}, \widehat{n}^{\Delta t} - n^{\Delta t} \right\rangle \mathrm{d}t = -\frac{1}{2}\sum_{l=1}^{\mathcal{L}} |n^l - n^{l-1}|_{\mathcal{H}}^2 \leq 0,$$

$$-\liminf_{\Delta t \to 0} a(n^{\Delta t}, n^{\Delta t}) \leq -a(n,n),$$

we could easily verify that inequality (6.15) holds for $\widetilde{n}$, i.e. $\widetilde{n} = n$.

Lastly, the existence of n and the convergence of $n^{\Delta t}$ to n in the case $n_0 \in \mathcal{H}$ could be demonstrated in the line of the last point of the proof of theorem 6.2 and the proof of corollary 6.1. ■

Remark 6.4. *Although* $\dfrac{\mathrm{d}n_\zeta}{\mathrm{d}t} \in L^2(0,T;\mathcal{V}^*)$*, the same conclusion could not be made for* $\dfrac{\mathrm{d}n}{\mathrm{d}t}$ *since the latter is likely not bounded in the* $L^2(0,T;\mathcal{V}^*)$*-norm. Indeed, estimate* (6.28) *is not useful since the right-hand side blows up as* $\zeta \to 0$.

Remark 6.5. *Based on the proofs of theorems 6.2 and 6.3 and corollaries 6.1 and 6.2, we have the following convergence diagram,*

$$\begin{array}{ccc} n_\zeta^{\Delta t} & \xrightarrow[\Delta t \to 0]{} & n_\zeta \\ \Big\downarrow{\scriptstyle \zeta \to 0} & & \Big\downarrow{\scriptstyle \zeta \to 0} \\ n^{\Delta t} & \xrightarrow[\Delta t \to 0]{} & n \end{array}$$

where the convergences are strong in H.

6.4 Implicit time discretization and treatment of the source terms

In light of the proposed reformulation (6.7) for the electron conservation law, the standard discharge model (2.13) is supplanted by

$$\begin{cases} \partial_t \boldsymbol{U} + \nabla \cdot \boldsymbol{F}(\boldsymbol{E}, w, \boldsymbol{U}) = \boldsymbol{S}(\boldsymbol{E}, w, \boldsymbol{U}) + \boldsymbol{S}_\psi(\boldsymbol{U}), & (t, \boldsymbol{x}) \in (0,T) \times \Omega, \\ -\nabla \cdot (\varepsilon_r(\boldsymbol{x})\varepsilon_0 \nabla \phi) = \rho, & (t, \boldsymbol{x}) \in (0,T) \times \Omega_\phi, \end{cases} \tag{6.32}$$

where $(\boldsymbol{S}_\psi(\boldsymbol{U}))_e = S_\psi(n_e)$ and $(\boldsymbol{S}_\psi)_s = 0$ for $s \neq e$ since the floor density constraint is not imposed on other species. We remark that the standard discharge system (2.13) is charge-conservative since $\sum_{s\in\mathfrak{S}} z_s S_s = 0$, so the proposed system (6.32) is not charge-conservative due to the presence of $\boldsymbol{S}_\psi$. However, since the floor density ψ that we choose in the simulations is much smaller than the charge density, the lack of charge conservation does not, in general, change the characteristics of the discharge.

In corona discharges, at each time level t^l the dielectric relaxation time Δt^l_ϕ (3.15) is usually much larger than the most severe CFL condition - that of electrons Δt^l_e (4.46) - by a factor of 10^4 to 10^7. The latter is on the order of 10^{-13}-10^{-12} s comparing to the characteristic time of ionic wind on the order of 10^{-3} s. This huge discrepancy results in unbearable CPU time which can amount to weeks or even months in two-dimensional simulations, even on multiple processors. Therefore, for simulation of corona discharge, we employ a time-stepping strategy that was put forth in [23] and described later in section 6.4.1.

6.4.1 Density-field decoupling strategy for simulation of corona discharge

For a time level $l \geq 0$, let us denote as $g^l(\boldsymbol{x})$ an approximant of the function/vector function $g(t^l, \boldsymbol{x})$ at time t^l. For the vector density $\boldsymbol{U}$, let us assume that $\partial_t \boldsymbol{U}(t^{l+1})$ is approximated by the backward Euler scheme $\dfrac{\boldsymbol{U}^{l+1} - \boldsymbol{U}^l}{\Delta t^l}$. We consider the following semi-discrete discretization (in time) of system (6.32),

$$\begin{cases} \dfrac{\boldsymbol{U}^{l+1} - \boldsymbol{U}^l}{\Delta t^l} + \nabla \cdot \boldsymbol{F}(\boldsymbol{E}^l, w^l, \boldsymbol{U}^{l+1}) \in \boldsymbol{S}(\boldsymbol{E}^l, w^l, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) + \boldsymbol{S}^{l+1}_\psi, \\ -\nabla \cdot (\varepsilon_r \varepsilon_0 \nabla \boldsymbol{\phi}^{l+1}) = \rho^{l+1}. \end{cases} \tag{6.33}$$

The superscript l of w in the time-discrete flux indicates that the drift and diffusion coefficients are evaluated at time t^l. In addition, the field are also taken at t^l in the flux, which allow the left-hand side of the first inclusion of system (6.32) to be linear with respect to $\boldsymbol{U}^{l+1}$. On the other hand, the specie density approximants $\boldsymbol{U}^l$ and $\boldsymbol{U}^{l+1}$ are mixed together in the source terms $\boldsymbol{S}$, which explains why $\boldsymbol{S}$ in system (6.33) depends on both and not only on $\boldsymbol{U}^{l+1}$ as expected in a standard implicit method. The mixing of densities is specifically chosen such that system (6.33), without the term $\boldsymbol{S}_\psi$, is linear in $\boldsymbol{U}^{l+1}$ and will be detailed in section 6.4.2.

6.4.2 Treatment of the source terms

Implicit treatment of the charge specie source terms

For charge species, a thumb rule is that if a reaction involves **one and only one** specie $s \in \mathfrak{S}$, then the density approximant is treated implicitly, i.e. taken at time t^{l+1}. Otherwise, the specie densities are taken at time t^l. For illustration, we consider the discretization of $\boldsymbol{S}$ for the two kinetic schemes in examples 2.1 and 2.2.

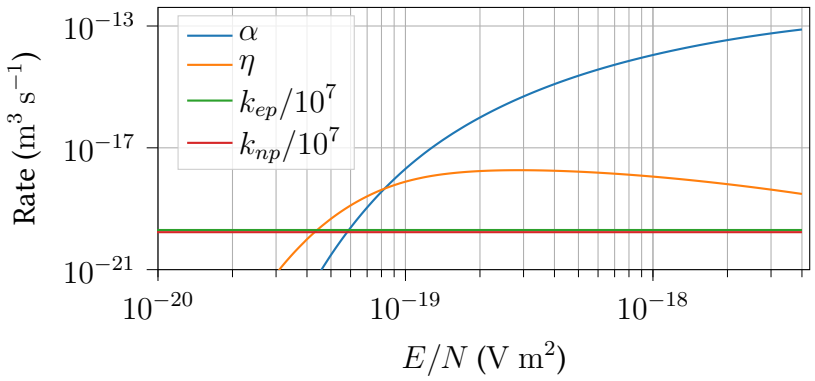


Figure 6.1: Rate coefficients of the three-specie, four-reaction kinetic model

Example (Kinetic scheme from example 2.1)**.** *The kinetic source terms of charged species read as,*

$$\begin{cases} S_e(w, \boldsymbol{U}) &= (\alpha - \eta) N n_e - k_{ep} n_e n_p, \\ S_p(w, \boldsymbol{U}) &= \alpha N n_e - k_{ep} n_e n_p - k_{np} n_n n_p, \\ S_n(w, \boldsymbol{U}) &= \eta N n_e - k_{np} n_n n_p, \end{cases}$$

and the time-discrete source terms according to the thumb rule are give in the following way,

$$\begin{cases} S_e(w^l, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) &= (\alpha^l - \eta^l) N n_e^{l+1} - k_{ep}^l n_e^l n_p^l, \\ S_p(w^l, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) &= \alpha^l N n_e^{l+1} - k_{ep}^l n_e^l n_p^l - k_{np}^l n_n^l n_p^l, \\ S_n(w^l, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) &= \eta^l N n_e^{l+1} - k_{np}^l n_n^l n_p^l. \end{cases} \tag{6.34}$$

The explicit treatment of the two-charged-body recombination reactions are somewhat justified in this case. Indeed, if we estimate that the maximum density of charged species in corona discharges is around $10^{-7} \times N$*, then the rescaled rate coefficients of the recombination reactions are well below the coefficients of the ionization and attachment reactions (see fig. 6.1) for field strength larger than the breakdown value, which is around* 10^{-19} V m^2.

Example (Kinetic scheme from example 2.2)**.** *The kinetic source terms of charged species read as,*

$$\begin{cases} S_e(w, \boldsymbol{U}) &= (\alpha - \eta - \eta_{3b} N_{O_2}) N n_e + k_{d-O2} N n_{O_2^-} + k_{d-O} N_{N_2} n_{O^-}, \\ S_p(w, \boldsymbol{U}) &= \alpha N n_e, \\ S_{O^-}(w, \boldsymbol{U}) &= \eta N n_e - \left(k_{d-O} \dfrac{N_{N_2}}{N} + k_{ct-2b} \dfrac{N_{O_2}}{N} + k_{ct-3b} N_{O_2} \right) N n_{O^-}, \\ S_{O_2^-}(w, \boldsymbol{U}) &= \eta_{3b} N_{O_2} N n_e - k_{d-O2} N n_{O_2^-} + k_{ct-2b} N_{O_2} n_{O^-}, \\ S_{O_3^-}(w, \boldsymbol{U}) &= k_{ct-3b} N_{O_2} N n_{O^-}, \end{cases}$$

and the time-discrete source terms according to the thumb rule are give in the following way,

$$\begin{cases} S_e(w^l, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) & = (\alpha^l - \eta^l - \eta_{3b}^l N_{O_2}) N n_e^{l+1} + k_{d-O2}^l N n_{O_2^-}^{l+1} + k_{d-O}^l N_{N_2} n_{O^-}^{l+1}, \\ S_p(w^l, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) & = \alpha^l N n_e^{l+1}, \\ S_{O^-}(w^l, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) & = \eta^l N n_e^{l+1} - \left(k_{d-O}^l \dfrac{N_{N_2}}{N} + k_{ct-2b}^l \dfrac{N_{O_2}}{N} + k_{ct-3b}^l N_{O_2} \right) N n_{O^-}^{l+1}, \\ S_{O_2^-}(w^l, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) & = \eta_{3b}^l N_{O_2} N n_e^{l+1} - k_{d-O2}^l N n_{O_2^-}^{l+1} + k_{ct-2b}^l N_{O_2} n_{O^-}^{l+1}, \\ S_{O_3^-}(w^l, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) & = k_{ct-3b}^l N_{O_2} N n_{O^-}^{l+1}. \end{cases}$$

In practice, we replace the densities N_{O_2} and N_{N_2} with the sole neutral density N. This requires to simply rescale the rate coefficients in the corresponding reactions with the proportion of dioxygen and dinitrogen in air.

Implicit treatment of the energy source term

For the electron mean energy, we adopt the implicit treatment as well as the linearization process of the energy source term that was put forth in [66]. The reason that the energy source term needs to be implicitly evaluated is that it is stiff in the sense that using large numerical timesteps could generate "small oscillations in the energy density which are amplified and spread rapidly throughout the whole system of equations" [63]. Let us recall that the energy source term S_ε reads as (see section 2.1.2)

$$S_\varepsilon(\boldsymbol{E}, \overline{\varepsilon}_e, \boldsymbol{U}) = -q\boldsymbol{E} \cdot \boldsymbol{F}_e - k_\varepsilon N n_e,$$

which gives rise to the following implicit time-discrete evaluation,

$$S_\varepsilon(\boldsymbol{E}^{l+1}, \overline{\varepsilon}_e^{l+1}, \boldsymbol{U}^{l+1}) = -q\boldsymbol{E}^{l+1} \cdot \boldsymbol{F}_e^{l+1} - k_\varepsilon^{l+1} N n_e^{l+1}. \tag{6.35}$$

We simplify the discrete approximant by taking the field as well as the electron density at time t^l, i.e. replacing the time-discrete source term (6.35) with the following,

$$S_\varepsilon(\boldsymbol{E}^l, \overline{\varepsilon}_e^{l+1}, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) = -q\boldsymbol{E}^l \cdot \boldsymbol{F}_e^{l+1} - k_\varepsilon^{l+1} N n_e^l. \tag{6.36}$$

Then, $\boldsymbol{F}_e^{l+1}$ as well as k_ε^{l+1} are linearized with respect to the electron mean energy $\overline{\varepsilon}_e$,

$$\begin{aligned} \boldsymbol{F}_e^{l+1} &\approx \boldsymbol{F}_e^l + \left(\frac{\partial \boldsymbol{F}_e}{\partial \mu_e}\right)^l \left(\frac{\partial \mu_e}{\partial \overline{\varepsilon}_e}\right)^l \left(\overline{\varepsilon}_e^{l+1} - \overline{\varepsilon}_e^l\right) + \left(\frac{\partial \boldsymbol{F}_e}{\partial D_e}\right)^l \left(\frac{\partial D_e}{\partial \overline{\varepsilon}_e}\right)^l \left(\overline{\varepsilon}_e^{l+1} - \overline{\varepsilon}_e^l\right) \\ &= \boldsymbol{F}_e^l - \left(\frac{\partial \mu_e}{\partial \overline{\varepsilon}_e}\right)^l \boldsymbol{E}^l n_e^l \left(\overline{\varepsilon}_e^{l+1} - \overline{\varepsilon}_e^l\right) - \left(\frac{\partial D_e}{\partial \overline{\varepsilon}_e}\right)^l \nabla n_e^l \left(\overline{\varepsilon}_e^{l+1} - \overline{\varepsilon}_e^l\right), \end{aligned} \tag{6.37}$$

$$k_\varepsilon^{l+1} \approx k_\varepsilon^l + \left(\frac{\partial k_\varepsilon}{\partial \overline{\varepsilon}_e}\right)^l \left(\overline{\varepsilon}_e^{l+1} - \overline{\varepsilon}_e^l\right), \tag{6.38}$$

where $\left(\dfrac{\partial \mu_e}{\partial \overline{\varepsilon}_e}\right)^l$, $\left(\dfrac{\partial D_e}{\partial \overline{\varepsilon}_e}\right)^l$ and $\left(\dfrac{\partial k_\varepsilon}{\partial \overline{\varepsilon}_e}\right)^l$ are resp. the approximants of $\dfrac{\partial \mu_e}{\partial \overline{\varepsilon}_e}$, $\dfrac{\partial D_e}{\partial \overline{\varepsilon}_e}$ and $\dfrac{\partial k_\varepsilon}{\partial \overline{\varepsilon}_e}$ at time t^l and can be computed from lookup tables of μ_e, D_e, and k_ε provided by the BOLSIG+ application [67] (using linear interpolation for example).

Now recalling that $n_\varepsilon = \overline{\varepsilon}_e n_e$ (see section 2.1.2), we have

$$\begin{aligned}\overline{\varepsilon}_e^{l+1} &= \overline{\varepsilon}_e^l + \left(\frac{\partial \overline{\varepsilon}_e}{\partial n_\varepsilon}\right)^l \left(n_\varepsilon^{l+1} - n_\varepsilon^l\right) + \left(\frac{\partial \overline{\varepsilon}_e}{\partial n_e}\right)^l \left(n_e^{l+1} - n_e^l\right) \\ &= \overline{\varepsilon}_e^l + \frac{1}{n_e^l}\left(n_\varepsilon^{l+1} - n_\varepsilon^l\right) - \frac{n_\varepsilon^l}{\left(n_e^l\right)^2}\left(n_e^{l+1} - n_e^l\right) = \overline{\varepsilon}_e^l + \frac{1}{n_e^l}\left(n_\varepsilon^{l+1} - \frac{n_\varepsilon^l}{n_e^l} n_e^{l+1}\right).\end{aligned}$$

Plugging this relation into eqs. (6.36) to (6.38), we can supplant the time-discrete source term (6.36) by a linearized energy source term in the following way,

$$\begin{aligned}S_\varepsilon(\boldsymbol{E}^l, \overline{\varepsilon}_e^l, \boldsymbol{U}^l, \boldsymbol{U}^{l+1}) &= -q\boldsymbol{E}^l \cdot \boldsymbol{F}_e^l - k_\varepsilon^l N n_e^l \\ &+ \left(q|\boldsymbol{E}^l|^2 \left(\frac{\partial \mu_e}{\partial \overline{\varepsilon}_e}\right)^l + q\boldsymbol{E}^l \cdot \frac{\nabla n_e^l}{n_e^l}\left(\frac{\partial D_e}{\partial \overline{\varepsilon}_e}\right)^l - \left(\frac{\partial k_\varepsilon}{\partial \overline{\varepsilon}_e}\right)^l N\right)\left(n_\varepsilon^{l+1} - \frac{n_\varepsilon^l}{n_e^l} n_e^{l+1}\right). \quad (6.39)\end{aligned}$$

Remark 6.6. *In [66], the linearized energy source term took into account the derivatives of the fully discrete (i.e. in time and space) electron flux that gives rise to, for example if we use the Scharfetter-Gummel scheme for flux approximation, the appearance of the terms*

$$\left(\frac{\partial \varphi(\mathfrak{p}_e)}{\partial \mathfrak{p}}\right)^l \left(\frac{\partial \mathfrak{p}_e}{\partial \mu_e}\right)^l, \quad \left(\frac{\partial \varphi(\mathfrak{p}_e)}{\partial \mathfrak{p}}\right)^l \left(\frac{\partial \mathfrak{p}_e}{\partial D_e}\right)^l,$$

before $\left(\frac{\partial \boldsymbol{F}_e}{\partial \mu_e}\right)^l$ *and* $\left(\frac{\partial \boldsymbol{F}_e}{\partial D_e}\right)^l$*, where* $\mathfrak{p}_e$ *is the numerical Péclet number of electrons and* $\varphi(\mathfrak{p}) = \frac{\mathfrak{p}}{2}\coth\left(\frac{\mathfrak{p}}{2}\right)$.

In eq. (6.39), this is not the case, which is probably less consistent comparing to the approach of [66] but much less complicated to implement since we only had the approximation of ∇n_e^l *at cell centers in COPAIER (see the end of section 3.3.3).*

Treatment of the photoionization source terms

The Helmholtz model (1.9)-(1.8) is employed to compute the photoionization source term $S_{\text{ph}}^Q(w, n_e)$, which is not stiff by numerical observations. Therefore, we evaluate S_{ph}^Q explicitly in time from the values w^l and n_e^l.

6.5 Solving the conservation laws

6.5.1 Time-discrete system of conservation laws

In this section, we consider the LFA model and the three-specie, four-reaction kinetic model in example 2.1 ($\mathfrak{S} = \{ e, p, n \}$). For the LEA as well as other kinetic schemes, such as the one in example 2.2, the discretization process in this section can be derived in the similar way.

We recall that the time-discrete system of conservation laws in (6.33) with the time-discrete

source terms (6.34) writes as, in an explicit form,

$$\begin{cases} n_e^{l+1} - n_e^l + \Delta t^l \nabla \cdot \boldsymbol{f}_e(\boldsymbol{E}^l, w^l, \boldsymbol{U}^{l+1}) = \Delta t^l(\alpha^l - \eta^l) N n_e^{l+1} - \Delta t^l k_{ep}^l n_e^l n_p^l + \Delta t^l S_\psi(n_e^{l+1}), \\ n_p^{l+1} - n_p^l + \Delta t^l \nabla \cdot \boldsymbol{f}_p(\boldsymbol{E}^l, w^l, \boldsymbol{U}^{l+1}) = \Delta t^l \alpha^l N n_e^{l+1} - \Delta t^l k_{ep}^l n_e^l n_p^l - \Delta t^l k_{np}^l n_n^l n_p^l, \\ n_n^{l+1} - n_n^l + \Delta t^l \nabla \cdot \boldsymbol{f}_n(\boldsymbol{E}^l, w^l, \boldsymbol{U}^{l+1}) = \Delta t^l \eta^l N n_e^{l+1} - \Delta t^l k_{np}^l n_n^l n_p^l \end{cases} \tag{6.40}$$

Let $\mathcal{I}$ denote the identity operator,

$$\mathcal{A}^l : \boldsymbol{U} \mapsto \begin{pmatrix} \nabla \cdot \boldsymbol{f}_e(\boldsymbol{E}^l, w^l, \boldsymbol{U}) - (\alpha^l - \eta^l) N n_e \\ \nabla \cdot \boldsymbol{f}_p(\boldsymbol{E}^l, w^l, \boldsymbol{U}) - \alpha^l N n_e \\ \nabla \cdot \boldsymbol{f}_n(\boldsymbol{E}^l, w^l, \boldsymbol{U}) - \eta^l N n_e \end{pmatrix} \equiv \begin{pmatrix} \left(\mathcal{A}^l \boldsymbol{U}\right)_e \\ \left(\mathcal{A}^l \boldsymbol{U}\right)_p \\ \left(\mathcal{A}^l \boldsymbol{U}\right)_n \end{pmatrix},$$

and $\boldsymbol{R}^l = \begin{pmatrix} n_e^l - \Delta t^l k_{ep}^l n_e^l n_p^l \\ n_p^l - \Delta t^l k_{ep}^l n_e^l n_p^l - \Delta t^l k_{np}^l n_n^l n_p^l \\ n_n^l - \Delta t^l k_{np}^l n_n^l n_p^l \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{R}_e^l \\ \boldsymbol{R}_p^l \\ \boldsymbol{R}_n^l \end{pmatrix}$. Then at t^l, system (6.40) can be recast as

$$(\mathcal{I} + \Delta t^l \mathcal{A}^l - \Delta t^l \boldsymbol{S}_\psi) \boldsymbol{U}^{l+1} \ni \boldsymbol{R}^l. \tag{6.41}$$

6.5.2 Fully discrete system of conservation laws

Let us consider a open bounded polygonal domain Ω in $\mathbb{R}^2$ and a conforming grid $\mathcal{T}$ on Ω (see section 3.1 for the grid notations). Let us consider in particular the electron conservation law in system (6.40),

$$\frac{n_e^{l+1} - n_e^l}{\Delta t^l} + \nabla \cdot \boldsymbol{f}_e(\boldsymbol{E}^l, w^l, \boldsymbol{U}^{l+1}) - (\alpha^l - \eta^l) N n_e^{l+1} + k_{ep}^l n_e^l n_p^l = h^{l+1},$$

with $h^{l+1} \in S_\psi(n_e^{l+1})$. Now averaging this equation on a cell Ω_K and dividing it by $|\Omega_K|$, we have[8]

$$\frac{\overline{n}_{e,K}^{l+1} - \overline{n}_{e,K}^l}{\Delta t^l} + \frac{1}{|\Omega_K|} \sum_{L \in \mathcal{V}_K^1} \int_{\lambda_{KL}} \boldsymbol{f}_e^{l,l+1} \cdot \boldsymbol{\nu}_{KL} \mathrm{d}\boldsymbol{l} - \left(\overline{(\alpha^l - \eta^l) N n_e^{l+1}}\right)_K + \left(\overline{k_{ep}^l n_e^l n_p^l}\right)_K = \overline{h}_K^{l+1}, \tag{6.42}$$

where, recall from section 3.1 that $\overline{v}_K \equiv \dfrac{1}{|\Omega_K|} \displaystyle\int_{\Omega_K} v(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$ for any function v, $\boldsymbol{f}_e^{l,l+1} \equiv \boldsymbol{f}_e(\boldsymbol{E}^l, w^l, \boldsymbol{U}^{l+1})$, $\mathrm{d}\boldsymbol{l}$ is the unit length element, λ_{KL} is the common edge of Ω_K and Ω_L while $\boldsymbol{\nu}_{KL}$ is the unit outward normal of Ω_K on the edge λ_{KL}.

We approximate $\alpha^l(\boldsymbol{x})$ for $\boldsymbol{x} \in \Omega_K$ (and similarly for η, k_{ep}) by $\alpha_K^l \equiv \alpha^l(\boldsymbol{x}_K^c)$, $\displaystyle\int_{\lambda_{KL}} \boldsymbol{f}_e^{l,l+1} \cdot \boldsymbol{\nu}_{KL} \mathrm{d}\boldsymbol{l}$ by $|\lambda_{KL}| \boldsymbol{f}_{e,KL}^{l,l+1} \cdot \boldsymbol{\nu}_{KL} \equiv |\lambda_{KL}| \boldsymbol{f}_e^{l,l+1}(\boldsymbol{x}_{KL}^\lambda) \cdot \boldsymbol{\nu}_{KL}$ with $\boldsymbol{x}_{KL}^\lambda$ the midpoint of λ_{KL} and then $\overline{n_e n_p}$ by $\overline{n}_e \overline{n}_p$. As such, (6.42) is replaced by

$$\frac{\overline{n}_{e,K}^{l+1} - \overline{n}_{e,K}^l}{\Delta t^l} + \frac{1}{|\Omega_K|} \sum_{L \in \mathcal{V}_K^1} |\lambda_{KL}| \boldsymbol{f}_{e,KL}^{l+1} \cdot \boldsymbol{\nu}_{KL} - (\alpha_K^l - \eta_K^l) N \overline{n}_{e,K}^{l+1} + k_{ep,K}^l \overline{n}_{e,K}^l \overline{n}_{p,K}^l = \overline{h}_K^{l+1}, \tag{6.43}$$

[8] recall that $\mathcal{V}_K^1$ is the first-level neighborhood of Ω_K (see section 4.6)

for every $\Omega_K \in \mathcal{T}$.

Now for any locally integrable function/vector function v defined on Ω, we define a piecewise constant function $\overline{v}$ on Ω such that

$$\overline{v}(\boldsymbol{x}) = \overline{v}_K, \quad \forall \boldsymbol{x} \in \Omega_K.$$

It is not evident how $\overline{h}^l$ with $h^l \in S^l_\psi$ relates to the floor density ψ and $\overline{n}^l_e$. Therefore, we make an approximation that

$$\overline{h}^l \in S_{\overline{\psi}}(\overline{n}^l_e) \equiv -\partial I_{\mathcal{K}_{\overline{\psi}}}(\overline{n}^l_e), \quad \mathcal{K}_{\overline{\psi}} = \left\{ v \;\middle|\; v \geq \overline{\psi} \right\}. \tag{6.44}$$

In order to transition to a fully discrete (both in space and time) discharge system, we use the standard Scharfetter-Gummel scheme (see section 4.2) to approximate the flux $\boldsymbol{f}^{l,l+1}_{e,KL} \cdot \boldsymbol{\nu}_{KL}$. The flux approximant writes as[9]

$$\begin{cases} \overline{f}^{l,l+1}_{e,KL} = \dfrac{D^l_{e,KL}}{d_{KL}} \left(\mathcal{B}(\mathfrak{p}^l_{e,KL}) \overline{n}^{l+1}_{e,K} - \mathcal{B}(-\mathfrak{p}^l_{e,KL}) \overline{n}^{l+1}_{e,L} \right), & \lambda_{KL} \nsubseteq \Gamma = \overline{\Omega} \setminus \Omega, \\ \overline{f}^{l,l+1}_{e,KL} = u^l_{e,KL} \overline{n}^{l+1}_{e,K}, & \lambda_{KL} \subseteq \Gamma_f, \\ \overline{f}^{l,l+1}_{e,KL} = 0, & \lambda_{KL} \subseteq \Gamma_s, \\ \overline{f}^{l,l+1}_{e,KL} = \max(u^l_{e,KL}, 0) \overline{n}^{l+1}_{e,K} + \dfrac{\overline{v}^l_e}{2} - \gamma \left(\max(u^l_{p,KL}, 0) \overline{n}^{l+1}_{p,K} + \dfrac{\overline{v}_p}{2} \right), & \lambda_{KL} \subseteq \Gamma_s, \end{cases}$$

with $d_{KL} = |\boldsymbol{x}_{KL} - \boldsymbol{x}_{LK}|$[10], $\mathfrak{p}^l_{e,KL} = -\dfrac{u^l_{e,KL} d_{KL}}{D^l_{e,KL}}$ while $u^l_{e,KL}$ and $D^l_{e,KL}$ are resp. approximants of $\boldsymbol{u}_e(t^l, x^\lambda_{KL}) \cdot \nu_{KL}$ and $D_e(t^l, x^\lambda_{KL})$ (see section 3.3.2).

Therefore, the fully discrete version of the time-discrete cell-averaged electron conservation law (6.43) reads as follows,

$$\frac{\overline{n}^{l+1}_{e,K} - \overline{n}^l_{e,K}}{\Delta t^l} + \frac{1}{|\Omega_K|} \sum_{L \in \mathcal{V}^1_K} |\lambda_{KL}| \overline{f}^{l,l+1}_{e,KL} - (\alpha^l_K - \eta^l_K) N \overline{n}^{l+1}_{e,K} + k^l_{ep,K} \overline{n}^l_{e,K} \overline{n}^l_{p,K} = \overline{h}^{l+1}_K, \tag{6.45}$$

for every $\Omega_K \in \mathcal{T}$. Similarly, the discrete flux $\overline{f}^{l,l+1}_{p,KL}$ and $\overline{f}^{l,l+1}_{n,KL}$ of positive and negative ions can be computed in the same way as $\overline{f}^{l,l+1}_{e,KL}$ and the fully discrete version of the time-discrete ion conservation laws in system eq. (6.40) are given in the following way,

$$\begin{cases} \dfrac{\overline{n}^{l+1}_{p,K} - \overline{n}^l_{p,K}}{\Delta t^l} + \dfrac{1}{|\Omega_K|} \displaystyle\sum_{L \in \mathcal{V}^1_K} |\lambda_{KL}| \overline{f}^{l,l+1}_{p,KL} = \alpha^l_K N \overline{n}^{l+1}_{e,K} - k^l_{ep,K} \overline{n}^l_{e,K} \overline{n}^l_{p,K} - k^l_{np,K} \overline{n}^l_{n,K} \overline{n}^l_{p,K}, \\ \dfrac{\overline{n}^{l+1}_{n,K} - \overline{n}^l_{n,K}}{\Delta t^l} + \dfrac{1}{|\Omega_K|} \displaystyle\sum_{L \in \mathcal{V}^1_K} |\lambda_{KL}| \overline{f}^{l,l+1}_{n,KL} = \eta^l_K N \overline{n}^{l+1}_{e,K} - k^l_{np,K} \overline{n}^l_{n,K} \overline{n}^l_{p,K}. \end{cases} \tag{6.46}$$

[9] for the flux at boundaries, see section 2.2

[10] we refer to section 4.5.1 for the definition of $\boldsymbol{x}_{KL}$ and $\boldsymbol{x}_{LK}$, noting that the subscript q is dropped since there is only one quadrature point

Let us define $\mathfrak{U}^l \equiv (\overline{n}^l_{s,K})_{s\in\mathfrak{S},\Omega_K\in\mathcal{T}} \in \mathbb{R}^{|\mathfrak{S}||\mathcal{T}|}$ the vector of unknowns at t^l, $\overline{\mathcal{I}}$ the identity matrix, $\overline{\mathcal{A}}^l$ a $|\mathfrak{S}||\mathcal{T}| \times |\mathfrak{S}||\mathcal{T}|$ matrix such that

$$\overline{\mathcal{A}}^l : \mathfrak{U} \mapsto \begin{pmatrix} \frac{1}{|\Omega_K|}\sum_{L\in\mathcal{V}^1_K} |\lambda_{KL}|\overline{f}^{l,}_{e,KL} - (\alpha^l_K - \eta^l_K)N\overline{n}_{e,K} \\ \frac{1}{|\Omega_K|}\sum_{L\in\mathcal{V}^1_K} |\lambda_{KL}|\overline{f}^{l,}_{p,KL} - \alpha^l_K N\overline{n}_{e,K} \\ \frac{1}{|\Omega_K|}\sum_{L\in\mathcal{V}^1_K} |\lambda_{KL}|\overline{f}^{l,}_{n,KL} - \eta^l_K N\overline{n}_{e,K} \end{pmatrix}_{K=1,\dots,\mathcal{N}} \equiv \begin{pmatrix} (\overline{\mathcal{A}}^l\mathfrak{U})_{e,K} \\ (\overline{\mathcal{A}}^l\mathfrak{U})_{p,K} \\ (\overline{\mathcal{A}}^l\mathfrak{U})_{n,K} \end{pmatrix}_{K=1,\dots,\mathcal{N}},$$

and $\mathfrak{R}^l = \begin{pmatrix} \overline{n}^l_{e,K} - \Delta t^l k^l_{ep,K}\overline{n}^l_{e,K}\overline{n}^l_{p,K} \\ \overline{n}^l_{p,K} - \Delta t^l k^l_{ep,K}\overline{n}^l_{e,K}\overline{n}^l_{p,K} - \Delta t^l k^l_{np,K}\overline{n}^l_{n,K}\overline{n}^l_{p,K} \\ \overline{n}^l_{n,K} - \Delta t^l k^l_{np,K}\overline{n}^l_{n,K}\overline{n}^l_{p,K} \end{pmatrix}_{K=1,\dots,\mathcal{N}} \equiv \begin{pmatrix} \mathfrak{R}^l_e \\ \mathfrak{R}^l_p \\ \mathfrak{R}^l_n \end{pmatrix}_{K=1,\dots,\mathcal{N}}.$

With these notations, the fully discrete system (6.45)-(6.46) can be recast in a more compact form,

$$(\overline{\mathcal{I}} + \Delta t^l\overline{\mathcal{A}}^l - \Delta t^l \boldsymbol{S}_{\overline{\psi}})\mathfrak{U}^{l+1} \ni \mathfrak{R}^l, \tag{6.47}$$

Here, the writing $\boldsymbol{S}_{\overline{\psi}}\mathfrak{U}^l$ is a notation abuse since $\mathfrak{U}^l$ is a discrete vector. In fact, inclusion (6.47) should be understood in the sense of (6.44) and (6.45).

We shall investigate in the next sections some algorithms to solve the discrete conservation laws (6.47). The complex structure of the operator $\overline{\mathcal{I}} + \Delta t^l\overline{\mathcal{A}}^l - \Delta t^l \boldsymbol{S}_{\overline{\psi}}$ makes the evaluation of $(\overline{\mathcal{I}} + \Delta t^l\overline{\mathcal{A}}^l - \Delta t^l \boldsymbol{S}_{\overline{\psi}})^{-1}\mathfrak{R}^l$ often hard to achieve. Therefore, we explore some splitting and iterative methods to solve this problem, namely the Lie operator splitting, the Douglas-Rachford method [94] and a Gauss-Seidel-inspired algorithm. The first one is a classical and very simple method to implement. The second one is also classical and well known in convex optimization [120]. The last method is derived from the Gauss-Seidel algorithm which is used for solving linear systems. We shall compare the performance of these methods further in section 6.6.

6.5.3 Lie splitting

Let us remark first of all that if $(\mathcal{I} + \Delta t^l\partial I_{K_\psi})g \ni h$ then from definition (6.4), we have

$$\frac{h-g}{\Delta t^l} \in \partial I_{K_\psi}(g) \iff \forall v \in K_\psi, \quad (h-g, v-g) \le 0.$$

It is shown that g is the projection of h on K_ψ, i.e. $g(t,\boldsymbol{x}) = \max(h(t,\boldsymbol{x}), \psi(\boldsymbol{x}))$, which is unique according to theorem 6.1. Therefore, if $\boldsymbol{H} = (h_e \quad h_p \quad h_n)^t$ and $J^{\Delta t}_{-\boldsymbol{S}_\psi} = (\mathcal{I} - \Delta t\boldsymbol{S}_\psi)^{-1}$ (the resolvent of $-\boldsymbol{S}_\psi$) then

$$J^{\Delta t^l}_{-\boldsymbol{S}_\psi}\boldsymbol{H} = \begin{pmatrix} \max(h_e, \psi) & h_p & h_n \end{pmatrix}^t. \tag{6.48}$$

Remark 6.7. *From inclusion (6.41), the electron density at t^{l+1} is $n^{l+1}_e = \max\left(\boldsymbol{R}^l_e - \Delta t^l(\mathcal{A}^l\boldsymbol{U}^{l+1})_e, \psi\right)$ which implies that $\overline{n}^{l+1}_e = \overline{\max\left(\boldsymbol{R}^l_e - \Delta t^l(\mathcal{A}^l\boldsymbol{U}^{l+1})_e, \psi\right)}$. By this point, making the assumption (6.44) is in fact to approximate the right-hand side of this equation with $\max\left(\overline{\boldsymbol{R}^l_e - \Delta t^l(\mathcal{A}^l\boldsymbol{U}^{l+1})_e}, \overline{\psi}\right)$. Further approximations of flux and chemical coefficients lead finally to $\overline{n}^{l+1}_e = \max\left(\mathfrak{R}^l_e - \Delta t^l(\overline{\mathcal{A}}^l\mathfrak{U}^{l+1})_e, \overline{\psi}\right)$, which is the solution of inclusion (6.47).*

Remark 6.8. *$J^{\Delta t^l}_{-\boldsymbol{S}_\psi}$ is independent of $\Delta t^l > 0$, which means that the electrons are instantaneously created to reach the floor density. Therefore, we can omit the timestep and write $J_{-\boldsymbol{S}_\psi}$ instead of $J^{\Delta t^l}_{-\boldsymbol{S}_\psi}$.*

The simplest operator splitting method is the Lie algorithm which is first-order in time. At t^l, we compute the numerical solution of (6.47), still denoted as $\mathfrak{U}^{l+1}$, in the following way,

$$\mathfrak{U}^{l+1} = J_{-\boldsymbol{S}_{\overline{\psi}}} J^{\Delta t^l}_{\overline{\mathcal{A}}^l} \mathfrak{R}^l. \tag{6.49}$$

The evaluation of $J^{\Delta t^l}_{\overline{\mathcal{A}}^l} \mathfrak{R}^l$ is classical since it is equivalent to inverting a linear system.

6.5.4 Douglas-Rachford algorithm

The second method that we consider is the Douglas-Rachford algorithm [94, 47]. The original method was introduced as a first-order-in-time operator splitting to solve the evolution problem

$$\begin{cases} \partial_\tau U + (A+B)U \ni R, \\ U(\tau = 0) = U^0 \in \mathrm{dom}(A) \cap \mathrm{dom}(B), \end{cases} \tag{6.50}$$

where A and B are maximal monotone operators. The algorithm is described as follows: we choose $V^0 \in (\mathcal{I} + \Delta\tau B)U^0$ and for $m \geq 0$, define

$$V^{m+1} = J^{\Delta\tau}_A (2J^{\Delta\tau}_B - \mathcal{I})V^m + (\mathcal{I} - J^{\Delta\tau}_B)V^m + \Delta\tau J^{\Delta\tau}_A R.$$

The approximation of U at time $m\Delta\tau$ is then $U^m = J^{\Delta\tau}_B V^m$. A crucial property of this method is that as $m \to \infty$, U^m converges to the steady-state solution U^∞ of (6.50) (i.e. $(A+B)U^\infty \ni R$) if the latter exists.

We adopt the Douglas-Rachford algorithm to solve the semi-discrete problem (6.41) by considering it as a stationary problem. More precisely, we search for the steady-state solution of the following pseudotime [77] problem,

$$\begin{cases} \dfrac{\mathrm{d}U(\tau)}{\mathrm{d}\tau} + \left(\dfrac{\mathcal{I}}{\Delta t^l} + \mathcal{A}^l - \boldsymbol{S}_\psi \right) U(\tau) \ni \dfrac{1}{\Delta t^l} \boldsymbol{R}^l, \\ U(0) = \boldsymbol{U}^l, \end{cases}$$

where τ is the pseudotime variable. We apply the Douglas-Rachford algorithm (6.50) for

$$A = \frac{\mathcal{I}}{\Delta t^l} + \mathcal{A}^l, \quad B = -\boldsymbol{S}_\psi, \quad R = \frac{1}{\Delta t^l}\boldsymbol{R}^l, \quad \Delta\tau = \Delta t^l, \tag{6.51}$$

and when it converges, we set $\boldsymbol{U}^{l+1} = U^\infty$.

Finally, the discrete version of the Douglas-Rachford method, applied to solve the discrete problem (6.47) is presented in algorithm 1.

Remark 6.9. *The requirement that A is monotone imposes a constraint on the timestep Δt^l which is related to the Townsend ionization rate α. In order to see this, let us assume for simplicity that D_s,*

Algorithm 1 Douglas-Rachford

1: Let $\epsilon > 0$, $U^0 = \mathfrak{U}^l$, $V^0 \in (\overline{\mathcal{I}} - \Delta t^l \boldsymbol{S}_{\overline{\psi}})U^0$, $\delta > \epsilon$
2: **while** $\delta > \epsilon$ **do**
3: Set $V^{m+1} = J^1_{\overline{\mathcal{I}}+\Delta t^l \overline{\mathcal{A}}^l}(2J_{-\boldsymbol{S}_{\overline{\psi}}} - \overline{\mathcal{I}})V^m + (\overline{\mathcal{I}} - J_{-\boldsymbol{S}_{\overline{\psi}}})V^m + J^1_{\overline{\mathcal{I}}+\Delta t^l \overline{\mathcal{A}}^l}\mathfrak{R}^l$
4: Set $U^{m+1} = J_{-\boldsymbol{S}_{\overline{\psi}}}V^{m+1}$ and let $\begin{pmatrix} n_e^{(m+1)} & n_p^{(m+1)} & n_n^{(m+1)} \end{pmatrix}^t = U^{m+1}$
5: Update $\delta = \max_{s\in\mathfrak{S}} \left(\dfrac{|n_s^{(m+1)} - n_s^{(m)}|_\infty}{|n_s^{(m)}|_\infty} \right)$ where $|\cdot|_\infty$ is the discrete L^∞-norm
6: **end while**
7: Set $\mathfrak{U}^{l+1} = U^m$

$\boldsymbol{u}_s$, α are constant and other reaction coefficients are zero. We define the scalar product on the space $\mathcal{H}^3 \equiv \mathcal{H} \times \mathcal{H} \times \mathcal{H}$ as

$$(\boldsymbol{U}, \boldsymbol{V})_{\mathcal{H}^3} = (n_e, w_e)_{\mathcal{H}} + (n_p, w_p)_{\mathcal{H}} + (n_n, w_n)_{\mathcal{H}},$$

with $\boldsymbol{U} = (n_e \quad n_p \quad n_n)^t$, $\boldsymbol{V} = (w_e \quad w_p \quad w_n)^t$. Since A is a linear operator it is sufficient to check the sign of $(A\boldsymbol{U}, \boldsymbol{U})^3_{\mathcal{H}}$. By some algebraic manipulation we have

$$(A\boldsymbol{U}, \boldsymbol{U})_{\mathcal{H}^3} = \sum_{s\in\mathfrak{S}} \left(\frac{1}{\Delta t^l} |n_s|^2_{\mathcal{H}} + D_s |\nabla n_s|^2_{\mathcal{H}} + \frac{1}{2} \int_\Gamma (\boldsymbol{u}_s \cdot \boldsymbol{\nu}) n_s^2 \mathrm{d}\boldsymbol{l} \right) - \alpha N |n_e|_{\mathcal{H}} (1 + |n_p|_{\mathcal{H}}).$$

If $C > 0$ with $C = \alpha N |n_e|_{\mathcal{H}}(1+|n_p|_{\mathcal{H}}) - \sum_{s\in\mathfrak{S}} D_s |\nabla n_s|^2_{\mathcal{H}} - \frac{1}{2}\sum_{s\in\mathfrak{S}} \int_\Gamma (\boldsymbol{u}_s \cdot \boldsymbol{\nu}) n_s^2 \mathrm{d}\boldsymbol{l}$, for $(A\boldsymbol{U}, \boldsymbol{U})_{\mathcal{H}^3} \geq 0$, Δt^l must satisfy

$$\Delta t^l \leq \frac{1}{C} \sum_{s\in\mathfrak{S}} |n_s|^2_{\mathcal{H}}. \tag{6.52}$$

In case of very large α such as in streamers or microdischarges, the timestep is severely restricted in spite of implicit integration.

6.5.5 Gauss-Seidel algorithm

For a discrete linear system $AU = R$ where A is a square matrix, the solution U can be obtained iteratively in the following way,

$$U^{m+1} = (M + D)^{-1} (R - NU^m), \quad m \geq 0,$$

where U^0 is an initial guess, D, M, N are resp. the diagonal, strictly lower and upper triangular components of A. By using the forward substitution, the elements of U^m can be computed sequentially as follows,

$$U_p^{m+1} = \frac{1}{1 + A_{pp}} \left(R_p - \sum_{q>p} A_{pq} U_q^m - \sum_{q<p} A_{pq} U_q^{m+1} \right), \quad p = 1, 2, \ldots,$$

where U_p^m is the p^{th} element of U^m and A_{pq} is the element of A on the p^{th} row and q^{th} column.

For the nonlinear problem (6.47), by analogy to the Gauss-Seidel method we propose the following iteration for forward substitution,

$$U^{m+1} = (\overline{\mathcal{I}} + \Delta t^l \mathcal{M}^l + \Delta t^l \mathcal{D}^l - \Delta t^l \boldsymbol{S}_{\overline{\psi}})^{-1} \left(\mathfrak{R}^l - \Delta t^l \mathcal{N}^l U^m\right),$$

where $\mathcal{D}^l$, $\mathcal{M}^l$, $\mathcal{N}^l$ are resp. the diagonal, strictly lower and upper triangular components of $\overline{\mathcal{A}}^l$. Its element-wise form reads as

$$\left(\overline{\mathcal{I}} - \frac{\Delta t^l}{1 + \Delta t^l \overline{\mathcal{A}}^l_{pp}} \boldsymbol{S}_{\overline{\psi}}\right) U_p^{m+1} = \frac{1}{1 + \Delta t^l \overline{\mathcal{A}}^l_{pp}} \left(\mathfrak{R}^l_p - \Delta t^l \sum_{q>p} \overline{\mathcal{A}}^l_{pq} U_q^m - \Delta t^l \sum_{q<p} \overline{\mathcal{A}}^l_{pq} U_q^{m+1}\right).$$

As long as $1 + \Delta t^l \overline{\mathcal{A}}^l_{pp} > 0$[11], $\forall p$ so that $-\dfrac{\Delta t^l}{1 + \Delta t^l \overline{\mathcal{A}}^l_{pp}} \boldsymbol{S}_{\overline{\psi}}$ is monotone, the resolvent of $\boldsymbol{S}_{\overline{\psi}}$ is well defined as in eq. (6.48) and we have

$$U_p^{m+1} = \frac{J_{-\boldsymbol{S}_{\overline{\psi}}}}{1 + \Delta t^l \overline{\mathcal{A}}^l_{pp}} \left(\mathfrak{R}^l_p - \Delta t^l \sum_{q>p} \overline{\mathcal{A}}^l_{pq} U_q^m - \Delta t^l \sum_{q<p} \overline{\mathcal{A}}^l_{pq} U_q^{m+1}\right).$$

The complete description of the proposed algorithm is presented in algorithm 2.

Algorithm 2 Gauss-Seidel

1: Let $\epsilon > 0$, $U^0 = \mathfrak{U}^l$, $\delta > \epsilon$
2: **while** $\delta > \epsilon$ **do**
3: Forward substitution
4: $U^{m+\frac{1}{2}} = (\overline{\mathcal{I}} + \Delta t^l \mathcal{M}^l + \Delta t^l \mathcal{D}^l - \Delta t^l \boldsymbol{S}_{\overline{\psi}})^{-1} \left(\mathfrak{R}^l - \Delta t^l \mathcal{N}^l U^m\right)$
5: Backward substitution
6: $U^{m+1} = (\overline{\mathcal{I}} + \Delta t^l \mathcal{N}^l + \Delta t^l \mathcal{D}^l - \Delta t^l \boldsymbol{S}_{\overline{\psi}})^{-1} \left(\mathfrak{R}^l - \Delta t^l \mathcal{M}^l U^{m+\frac{1}{2}}\right)$
7: Let $\begin{pmatrix} n_e^{(m+1)} & n_p^{(m+1)} & n_n^{(m+1)} \end{pmatrix}^t = U^{m+1}$
8: Update $\delta = \max\limits_{s \in \mathfrak{S}} \left(\dfrac{|n_s^{(m+1)} - n_s^{(m)}|_\infty}{|n_s^{(m)}|_\infty} \right)$
9: **end while**
10: Set $\mathfrak{U}^{l+1} = U^m$

6.6 Comparison of algorithms on one-dimensional grids

6.6.1 Conservation of steady state

For the wire-to-wire test case described in section 5.2, it happens that the steady state exists (see section 5.2.2), but the capacity of the code to reproduce this steady state depends strongly on the algorithm used to solve the discrete conservation laws (6.47) as we show in the next paragraphs. In

[11] this imposes a restriction on Δt^l which is a discrete version of inequality (6.52)

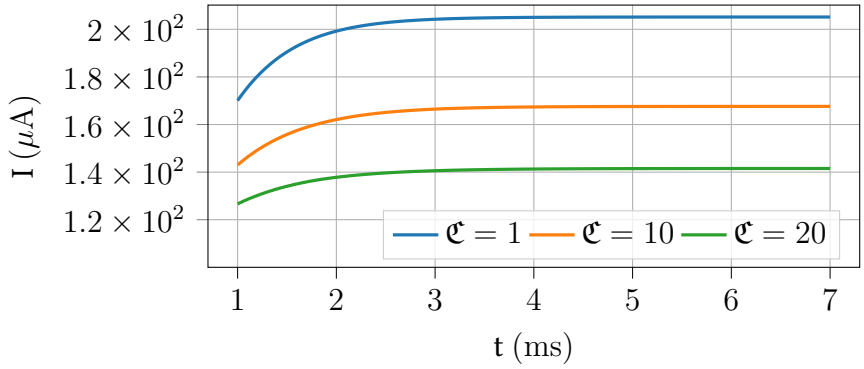


Figure 6.2: Lie splitting: electric current I with different values of $\mathfrak{C}$

this section, all the simulations are conducted on a 400-cell grid with the smallest cell size $\Delta x_{\min} = 5.5$ μm.

At each time level t^l, we compute the numerical timestep as follows,

$$\Delta t^l = \min(\mathfrak{C}\Delta t^l_{\text{ion}}, \Delta t^l_\phi),$$

where Δt^l_{ion}, Δt^l_ϕ are resp. given by eqs. (3.15) and (4.47) and $\mathfrak{C}$ is a user-defined parameter.

The simulation time is set at $T = 7$ ms for which we observe a steady-state current as defined in section 5.2.2. For $0 \leq t \leq 0.04$ ms, we fix $\mathfrak{C} = 0.01$ since at the discharge onset, there is a rapid charge multiplication and the dielectric relaxation time Δ^l_ϕ decreases quickly. So having a small timestep Δt^l ensures that the plasma dynamics are captured correctly. After $t = 0.04$ ms, we choose $\mathfrak{C}$ between 1 and 10^2. After $t = 0.4$ ms, the discharge enters the ion collection phase (slow-changing current with $\Delta^l_\phi \gg \Delta t^l_{\text{ion}}$) and we set $\mathfrak{C} = 10^3$.

Lie splitting

The computed electric currents with the algorithm (6.49) are shown in fig. 6.2. We show that the Lie splitting does not conserve the steady state: indeed, the current changes when we vary the time step.

Douglas-Rachford (DR) algorithm

We defined the stop criterion as $\epsilon = 10^{-6}$ (see algorithm 1). The electric current at steady state is conserved by the Douglas-Rachford algorithm as shown in fig. 6.3. This feature was remarked in [94]. At $T = 7$ ms, the current of $\mathfrak{C} = 10^2$ differs only 0.007% from the current of $\mathfrak{C} = 1$.

Gauss-Seidel (GS) algorithm

We defined the stop criterion as $\epsilon = 10^{-6}$ (see algorithm 2). The electric current at steady state is also conserved by the Gauss-Seidel algorithm as shown in fig. 6.4. At $T = 7$ ms, the current of $\mathfrak{C} = 10^2$ differs only $8 \times 10^{-7}\%$ from the current of $\mathfrak{C} = 1$.

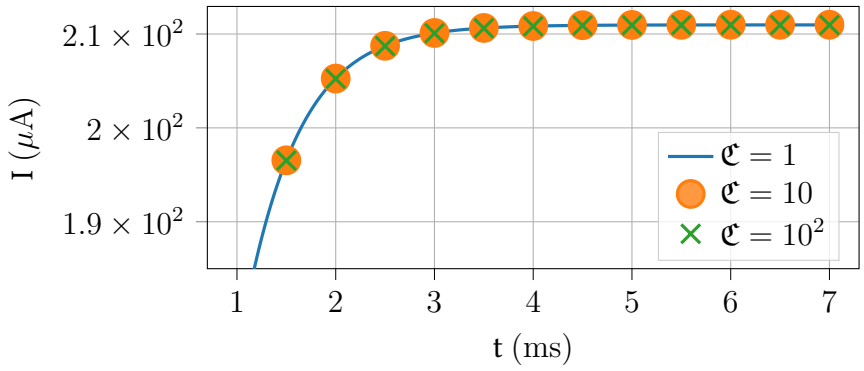


Figure 6.3: Douglas-Rachford algorithm: electric current I with different values of $\mathfrak{C}$

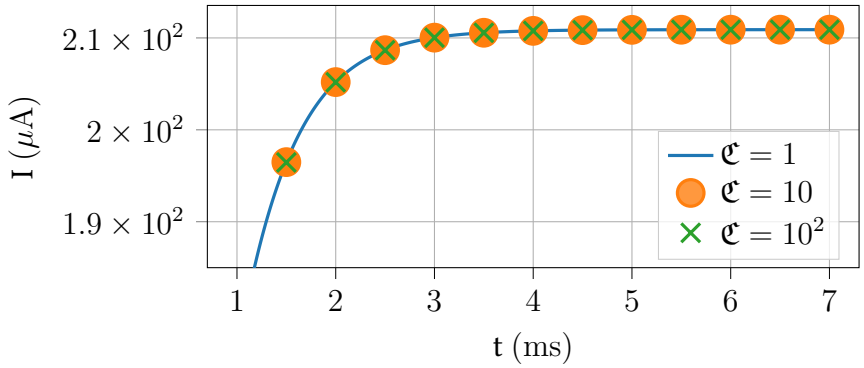


Figure 6.4: Gauss-Seidel algorithm: electric current I with different values of $\mathfrak{C}$

6.6.2 Performance comparison of the DR and GS algorithms

We have seen that the steady-state current computed with the DR and GS algorithms is virtually independent of timesteps. Therefore, these algorithms should be preferred over the Lie splitting to conduct implicit simulations of corona discharge. However, there is a huge difference between the two methods in term of computation time as shown in table 6.1.

In order to understand why, we count the number of iterations $N(t^l)$ at each time level t^l that is necessary for the algorithms to converge. Then we evaluate the averaged number of iterations $\overline{N}$ by dividing it with the number of time levels. Furthermore, we only count the iterations from $T = 1$ ms to 7 ms, since before 1 ms the dielectric time Δt^l_ϕ can fall below $\mathfrak{C}\Delta t^l_{\text{ion}}$ so $N(t^l)$ would not depend on $\mathfrak{C}$. To sum up, we have,

$$\overline{N} = \frac{\sum_{t^l=1\text{ ms}}^{7\text{ ms}} N(t^l)}{\text{Number of time levels}}.$$

We remark from table 6.1 that the averaged number of iterations of the DR algorithm scales almost linearly with $\mathfrak{C}$. This explains why the DR computation time is not shortened although we increase the timesteps. On the contrary, the averaged number of iterations of the GS algorithm is almost the same regardless of the time step and reduces the CPU time significantly. Therefore, we exclusively use the GS algorithm from now on.

$\mathfrak{C}$	DR		GS	
	$\overline{N}$	CPU time	$\overline{N}$	CPU time
1	14	45	1	17
10	84	26	1.13	4.65
100	856	24	1.33	1.68

Table 6.1: Averaged number of iterations $\overline{N}$ and CPU time (in minute) of DR and GS algorithms for simulations on a 400-cell grid

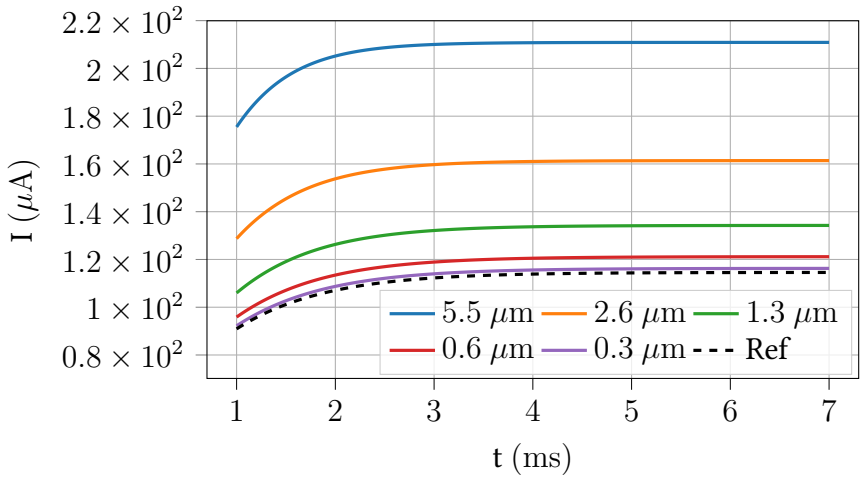


Figure 6.5: Standard SG scheme: electric current I with different mesh size $\Delta x_{\min}$

6.6.3 Mesh convergence

In this section, we conduct a mesh convergence study to ensure that the GS algorithm is able to capture correctly the discharge dynamics. We perform the simulations on five grids with 400 cells (smallest cell size $\Delta x_{\min} = 5.5$ μm), 800 cells (2.6 μm), 1600 cells (1.3 μm), 3200 cells (0.6 μm) and 6400 cells (0.3 μm). We use the standard SG scheme and the SGCC-1, SGCC-2 schemes (see section 4.3) for flux approximation. Since the latter two are nonlinear because of slope limiters, we use the fixed-point algorithm to iterate on the slopes[12].

The electric currents of each flux scheme are shown in figs. 6.5 to 6.7. They are compared to the current of the SGCC-2 scheme obtained on the 0.3 μm-grid (the black dashed curve). We can conclude that the whole numerical scheme is able to perform corona discharge simulation with confidence by observing that the electric currents converges to the black reference curve as the mesh size decreases. We also remark the high-order flux schemes (SGCC-1, SGCC-2) increases significantly the numerical precision.

Finally, we note that a mesh convergence study for the SGCC schemes with explicit time integration was only partially conducted in section 5.2 but the simulation on the 3200-cell grid was only launched up to 300 μs. The reason is that it would have taken **months** to obtain the steady-state current. In this section, the simulations are extremely rapid. For example, a computation with the

[12] see appendix E

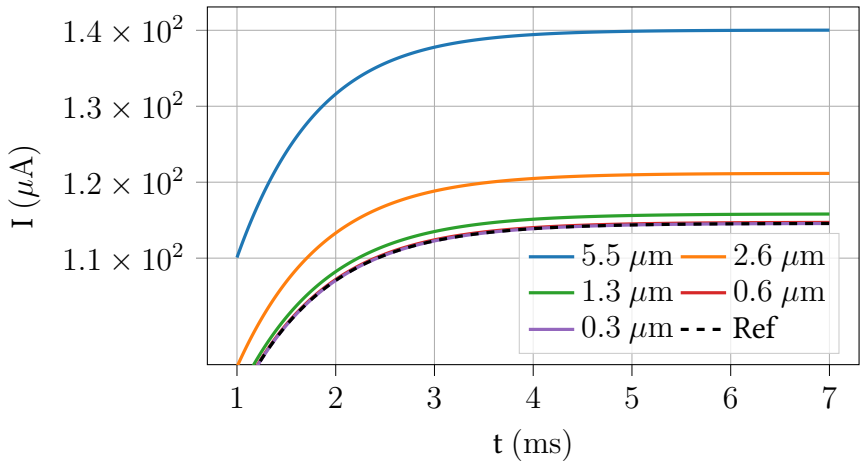


Figure 6.6: SGCC-1 scheme: electric current I with different mesh size $\Delta x_{\min}$

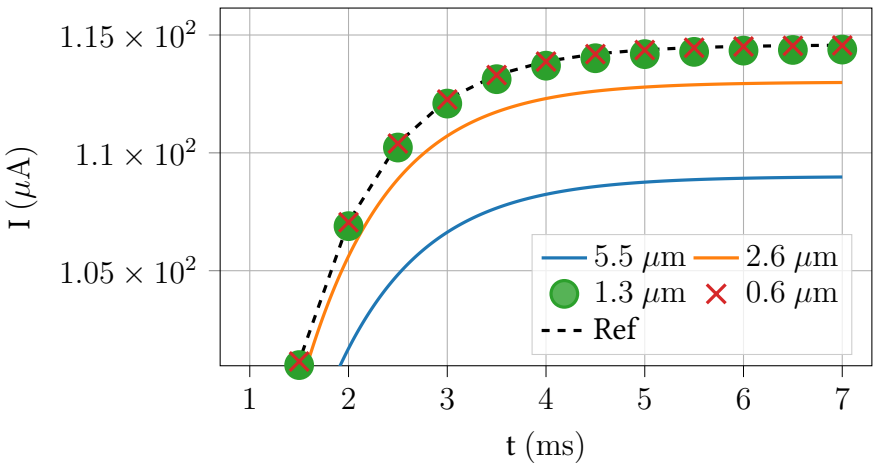


Figure 6.7: SGCC-2 scheme: electric current I with different mesh size $\Delta x_{\min}$

SGCC-2 scheme takes only **an hour** on the 3200-cell grid and **four hours** on the 6400-cell grid.

6.7 Closing remarks

The main results of this section are summarized as follows.

The conservation law of electrons were reformulated into a differential inclusion so that the floor density constraint on the electron density, $n_e \geq \psi$, appears directly in the discharge model. The accounting of this constraint manifests via a set-valued source term which is related to the subdifferential of the characteristic function of the convex set $\mathcal{K}_\psi$ of functions that satisfy the constraint.

The existence and uniqueness of the solution of the electron conservation law were demonstrated under some specific hypotheses. The proof of this result features the use of a regularization problem with a regularization parameter ζ. The smallness of ζ characterizes how fast the solution of the regularization problem moves into the convex $\mathcal{K}_\psi$ from an initial datum that is not necessarily in $\mathcal{K}_\psi$ (proposition 6.1). It has been demonstrated that the solution of the electron conservation law is the limit of the solution of the regularization problem as $\zeta \to 0$. Therefore, the solution of the electron conservation law is **instantaneously** an element of $\mathcal{K}_\psi$, meaning that electrons are created with a infinite production rate to satisfy the constraint $n_e \geq \psi$.

The conservation laws were discretized in time using the backward Euler scheme and some specific treatments of the source terms were derived to reduce the nonlinearity of the discrete discharge system. The latter is still nonlinear though because of the set-valued source term in the electron conservation law.

The mathematical reformulation of the electron conservation law was proved to be necessary as it allows to derive some iterative algorithms to correctly solve the discharge problem. The Douglas-Rachford and Gauss-Seidel algorithms, in particular, demonstrated the conservation of the steady-state current of a corona discharge, which was not the case with the widely used Lie splitting method. The Gauss-Seidel algorithm converges however much faster and thus is preferred over the Douglas-Rachford algorithm. However, we only tested the latter with the pseudo-timestep equal to the discharge timestep, i.e. $\Delta\tau = \Delta t^l$ in eq. (6.51). It would be interesting to study the Douglas-Rachford algorithm with other values of $\Delta\tau$.

The proposed implicit strategy enhances significantly the performance of the numerical solver. We were able to continue the mesh-convergence study of the SGCC-1 and SGCC-2 schemes on the one-dimensional corona discharge problem in section 5.2. Indeed, the proposed implicit strategy was able to cut short the CPU time, from **months** with an explicit time integration method, to a matter of **hours**.

6.8 Remarques finales

Les principaux résultats de cette section sont résumés comme suit.

La loi de conservation des électrons a été reformulée sous forme d'une inclusion différentielle de

sorte que la contrainte de densité de fond sur la densité des électrons, $n_e \geq \psi$, apparaît directement dans le modèle de décharge. La prise en compte de cette contrainte se manifeste par un terme source multivalué qui est lié au sous-différentiel de la fonction caractéristique de l'ensemble convexe $\mathcal{K}_\psi$ des fonctions qui satisfont la contrainte.

L'existence et l'unicité de la solution de la loi de conservation des électrons ont été démontrées sous certaines hypothèses spécifiques. La preuve de ce résultat repose sur l'utilisation d'un problème de régularisation avec un paramètre de régularisation ζ. La petitesse de ζ caractérise la vitesse à laquelle la solution du problème de régularisation se déplace dans le convexe $\mathcal{K}_\psi$ à partir d'une donnée initiale qui n'est pas nécessairement dans $\mathcal{K}_\psi$ (proposition 6.1). Il a été démontré que la solution de la loi de conservation des électrons est la limite, dans un certain espace fonctionnel, de la solution du problème de régularisation lorsque $\zeta \to 0$. Par conséquent, la solution de la loi de conservation des électrons est **instantanément** un élément de $\mathcal{K}_\psi$, ce qui signifie que les électrons sont créés avec un taux de production infini pour satisfaire la contrainte $n_e \geq \psi$.

Les lois de conservation ont été discrétisées en temps avec le schéma d'Euler implicite et certains traitements spécifiques des termes sources ont été dérivés pour réduire la nonlinéarité du système de décharge discret. Ce dernier reste cependant nonlinéaire en raison du terme source multivalué dans la loi de conservation des électrons.

La reformulation mathématique de la loi de conservation des électrons s'est avérée nécessaire car elle permet de dériver quelques algorithmes itératifs pour résoudre correctement le problème de la décharge. Les algorithmes de Douglas-Rachford et de Gauss-Seidel, en particulier, ont démontré la conservation du courant en régime stationnaire d'une décharge couronne, ce qui n'était pas le cas avec la méthode de splitting de Lie largement utilisée. L'algorithme de Gauss-Seidel converge cependant beaucoup plus rapidement et est donc préféré à l'algorithme de Douglas-Rachford. Cependant, nous n'avons testé ce dernier qu'avec un pas de pseudo-temps égal au pas de temps de la décharge, c'est-à-dire $\Delta\tau = \Delta t^l$ dans l'eq. (6.51). Il serait intéressant d'étudier l'algorithme de Douglas-Rachford avec d'autres valeurs de $\Delta\tau$.

La stratégie implicite proposée améliore considérablement les performances du solveur numérique. Nous avons pu poursuivre l'étude de convergence des maillages des schémas SGCC-1 et SGCC-2 sur le problème de décharge de couronne unidimensionnelle introduit dans la section 5.2. En effet, la stratégie implicite proposée a permis de réduire le temps de CPU des **mois** avec une méthode d'intégration temporelle explicite à seulement des **heures**.

Chapter 7

Simulations of Gas Discharge in Air with the Plasma Solver COPAIER*

7.0 Aperçu . . . 178
7.1 Overview . . . 178
7.2 Wire-to-wire corona discharge . . . 179
7.2.1 Description of the test case and numerical parameters . . . 179
7.2.2 Dynamics of the discharge . . . 181
7.2.3 Influence of the floor density on numerical solutions . . . 184
7.2.4 Comparison with experiment data . . . 184
7.3 Needle-to-ring corona discharge . . . 186
7.3.1 Description of the test case . . . 186
7.3.2 Computation of ionic wind and numerical parameters . . . 188
7.3.3 Positive corona . . . 189
7.3.4 Negative corona . . . 194
7.4 Closing remarks . . . 199
7.5 Remarques finales . . . 199

*parts of section 7.2, in altered form, has been published in [145]: *N Tuan Dung, C Besse, and F Rogier. "An implicit time integration approach for simulation of corona discharges". In: Computer Physics Communications 294 (2024), p. 108906.*

7.0 Aperçu

Ce chapitre présente la validation du modèle de densité du fond sur les électrons ainsi que la stratégie d'intégration implicite en temps qui ont été introduits dans la section 6.1 et ont été mis en œuvre dans le solveur de plasma de l'ONERA - COPAIER [46]. Les cas tests sont une décharge couronne fil-fil positive (section 7.2) et une décharge couronne aiguille-anneau (section 7.3, pour les polarités positives et négatives). Dans les deux cas, la force électromotrice V_G prend la forme d'une rampe de tension : le potentiel à l'électrode stressée est augmenté progressivement sur une courte période pour éviter le développement de micro-décharges (au moins pour les tensions positives), puis maintenu constant par la suite. Nous nous intéressons en particulier, et chaque fois qu'elle existe, au courant continu des décharges, puisque cette phase se produit sur une longue échelle de temps (de l'ordre de $1\ \mathrm{ms}$) et contribue principalement à la force EHD.

Les solutions numériques (courant de circuit, vitesse du vent ionique) sont comparées aux données expérimentales dans la mesure du possible afin de démontrer la capacité de COPAIER à reproduire des mesures réelles. Les résultats des expériences sont disponibles dans [12] pour la décharge fil-fil et dans [163] pour la décharge aiguille-anneau positive. Il est montré que l'inclusion de la contrainte de densité fond sur la densité électronique permet de calibrer les solutions numériques pour qu'elles s'adaptent aux données expérimentales. Une densité fond constante $\psi = 10^{11}\ \mathrm{m}^{-3}$ semble donner des résultats adéquats pour les décharges positives fil-fil et aiguille-anneau. Pour la décharge aiguille-anneau négative, les simulations prennent beaucoup de temps, de sorte que la comparaison avec les données expérimentales n'a pas été réalisée.

Pour rappel, les méthodes numériques utilisées dans ce chapitre sont du premier ordre en espace (schéma de Scharfetter-Gummel standard) et en temps (schéma d'Euler implicite). Le pas de temps numérique à chaque instant t^l est fixé comme suit,

$$\Delta t^l = \min\left(\mathfrak{C}\Delta t^l_{\text{ions}}, \Delta t^l_{\phi}, \frac{0.1|V_G(t^l)|}{\left|\frac{\mathrm{d}V_G(t^l)}{\mathrm{d}t}\right| + \epsilon}\right),$$

où $\mathfrak{C} = 10^3$, le pas de temps CFL des ions Δt^l_{ions} est défini dans l'eq. (4.47), le temps de relaxation diélectrique Δt^l_{ϕ} est défini dans l'eq. (3.15), $\epsilon = 10^{-6}$ pour éviter la division par zéro, et la dernière quantité du côté droit est là pour s'assurer que la simulation capture la dynamique de la décharge pendant la variation de V_G. Le nombre maximum d'itérations avant que l'algorithme de Gauss-Seidel converge est fixé à 10^4. Si l'algorithme ne converge pas, Δt^l est automatiquement réduit d'un facteur de 10 et le calcul recommence à partir de t^l.

7.1 Overview

This chapter provides the validation of the electron floor density model as well as the implicit time integration strategy that were laid out in section 6.1 and were implemented in ONERA's plasma solver COPAIER [46]. The test cases are a positive wire-to-wire corona discharge (section 7.2) and a needle-to-ring corona discharge (section 7.3, for both positive and negative polarities). In either situation, the electromotive force V_G takes the form of a ramp voltage: the potential at the stressed

electrode is raised gradually over a short time to avoid the development of microdischarges (at least for positive voltages) and then kept constant afterwards. We are interested in particular, and whenever it exists, in the direct-current component of the discharges, since this phase occurs on a long timescale (on the order of $1\ \mathrm{ms}$) and contribute mainly to the EHD force.

The numerical solutions (circuit current, ionic wind velocity) are compared to experiment data whenever possible to demonstrate the capability of COPAIER in reproducing real-life measurements. The experiment results are available in [12] for the wire-to-wire discharge and in [163] for the positive needle-to-ring discharge. It is shown that the inclusion of the floor density constraint on the electron density allows to calibrate the numerical solutions to fit the experiment data. A constant floor density $\psi = 10^{11}\ \mathrm{m}^{-3}$ seems to deliver adequate results for both positive wire-to-wire and needle-to-ring discharges. For the negative needle-to-ring discharge, the simulations take a lot of time to complete, so the comparison with experiment data has not yet been achieved.

As a reminder, the discretization methods used in this chapter are first-order in space (standard Scharfetter-Gummel scheme) and time (backward Euler scheme). The numerical timestep at each time t^l is set as

$$\Delta t^l = \min\left(\mathfrak{C}\Delta t^l_{\text{ions}}, \Delta t^l_\phi, \frac{0.1|V_G(t^l)|}{\left|\frac{\mathrm{d}V_G(t^l)}{\mathrm{d}t}\right| + \epsilon}\right),$$

where $\mathfrak{C} = 10^3$, the CFL timestep of ions Δt^l_{ions} is defined in eq. (4.47), the dielectric relaxation time Δt^l_ϕ is defined in eq. (3.15), $\epsilon = 10^{-6}$ to avoid division by zero, and the last quantity on the right-hand side is there to ensure that the simulation captures the discharge dynamics during the variation of V_G. The maximum number of iterations before the Gauss-Seidel algorithm converges is set at 10^4. If the algorithm does not converge, Δt^l is automatically reduced by a factor of 10 and the computation restarts from t^l.

7.2 Wire-to-wire corona discharge

7.2.1 Description of the test case and numerical parameters

A sketch of the actuator is shown in fig. 7.1. Two wires hanged in air are connected to a resistor R and a generator V_G. The length of each wire is L, the smallest gap between the electrode surfaces is d, the domain height is h and the smallest distance between the surface of an electrode and its closest vertical border is l. The coordinate origin $\boldsymbol{x} = (0, 0)$ is located at the center of the left (smaller) wire.

The particular geometry of the wire-to-wire actuator intensifies the electrostatic field in the electrode sheaths. Therefore, all chemical reactions are much stronger in the sheaths than in the outer region between the wires. As a result, the grids that we use are strongly refined near the electrodes as illustrated in fig. 7.2. The minimal grid size is fixed to a value $\Delta x^{\min}$.

The boundary conditions are shown on fig. 7.1 and described in section 2.2. The initial data are given in section 2.3. The electron floor density ψ is chosen between 1 and $10^{14}\ \mathrm{m}^{-3}$. The secondary emission coefficient γ is fixed to 10^{-4}. The simulation time is set at $T = 5\ \mathrm{ms}$ so that the discharge

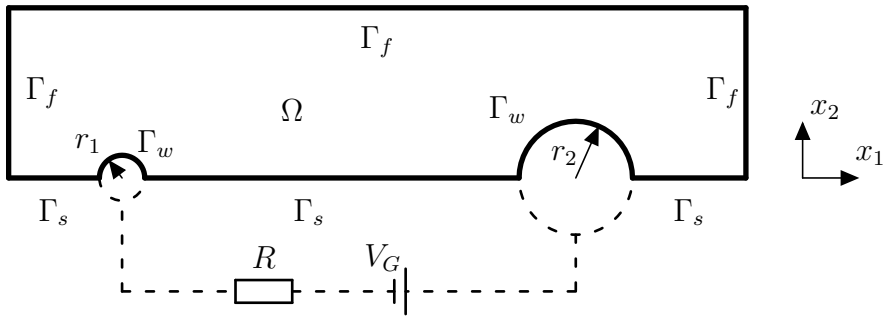


Figure 7.1: Sketch of the geometry of the actuator (not drawn to scale) and the computation domain Ω (inside the solid-line loop)

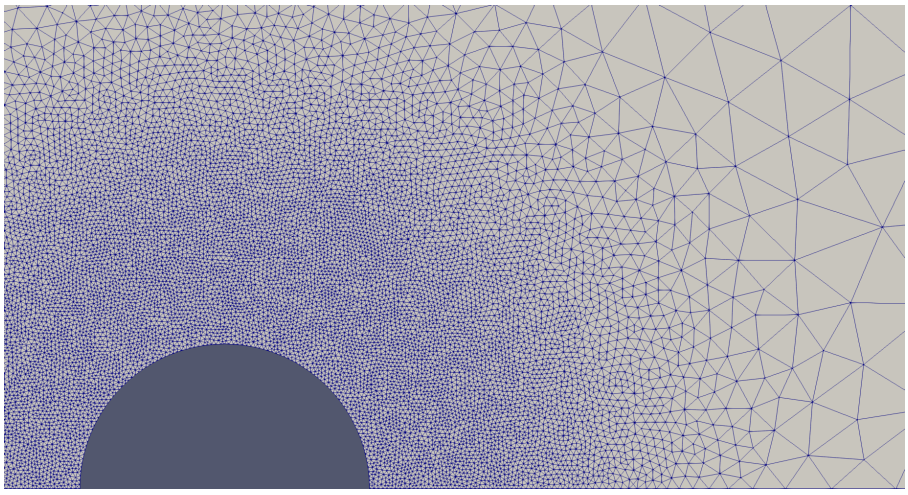


Figure 7.2: Grid refinement near the smaller wire (the left one on fig. 7.1) generated by Gmsh [58]

r_1	0.1-0.5 mm	h	8 mm	$\Delta x^{\min}$	5 μm
r_2	1 mm	l	4 mm	ψ	1-10^{14} m^{-3}
L	20 cm	V_G	9-37 kV	γ	10^{-4}
d	10-40 mm	R	10 kΩ	T	5 ms

Table 7.1: Parameters of the wire-to-wire corona discharge simulations

reaches a steady state in the sense of the definition in section 5.2.2.

In terms of kinetic model, we use the scheme of Bœuf et al. [15] (see example 2.1) which is available for both LFA and LEA models, or the scheme of Ferreira et al. [54] (see example 2.2) which is only available for the LFA model. The simulations of sections 7.2.2 to 7.2.4 are performed with the LFA model and the kinetic scheme of example 2.1 by **default**, unless indicated otherwise.

To avoid the formation of streamers that can reduce significantly the dielectric timestep and prolong the simulations, we increase the voltage gradually on 2 μs until it reaches the maximal value V_G.

Each simulation in this section was launched on four MPI processes[1]. The CPU time for the implicit time method introduced in chapter 6 was drastically shorten to **three to five hours** instead of a **week** for the explicit time methods that are available in COPAIER.

The characteristics of the actuator and the numerical parameters are grouped in table 7.1.

7.2.2 Dynamics of the discharge

In this section, we present the numerical solution for $V_G = 13$ kV, $r_1 = 0.1$ mm, $d = 10$ mm and $\psi = 10^{10}$ m^{-3}.

As the voltage rises, a cloud of positive ions starts to build up around the anode (the smaller wire) because of the ionization process. Figure 7.3 shows the evolution of positive ion density n_p between the electrodes on the symmetry axis. At $t = 1.17$ μs, n_p is up to 10^{14} m^{-3} in the anode sheath. We do not show the negative ion and electron densities since they are much smaller to the positive ion density. Therefore, the EHD force density $\boldsymbol{f}_{\text{EHD}} \equiv \rho \boldsymbol{E}$ depends largely on n_p and $\boldsymbol{E}$. Figure 7.5 shows that the force strength $f_{\text{EHD}} = |\boldsymbol{f}_{\text{EHD}}|$ at $t = 1.17$ μs is concentrated around the axis and on the right-hand side of the anode facing the cathode.

After the voltage reaches its maximum, the ionization process stops increasing and the positive ion cloud starts spreading out from the anode under the influence of the electrostatic field. At $t = 2.23$ μs, the cloud centers at 0.3 mm from the anode and the ion density builds up to 4×10^{17} m^{-3} which distorts the field a little bit (see Figure 7.4). We observe a "detonation" at this stage where the EHD force is released from the anode and moves away to the cathode (Figure 7.6) with an intensity up to 260 kN m^{-3}.

At $t = 3.32$ μs, the positive ion cloud diffuses progressively as they enters the low-field region

[1] see appendix F for the MPI implementation of the Gauss-Seidel algorithm

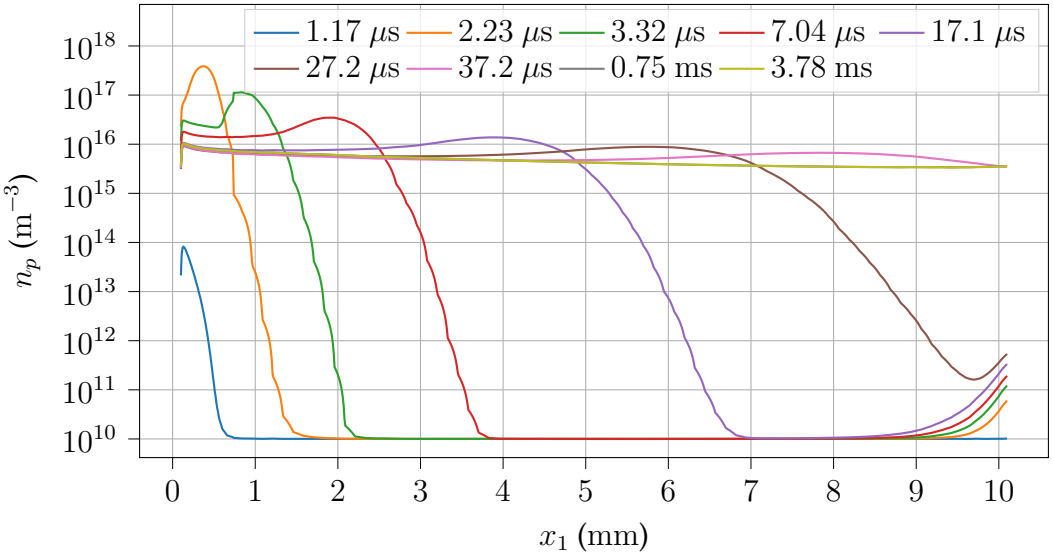


Figure 7.3: Positive ion density n_p between the electrodes on the symmetry axis

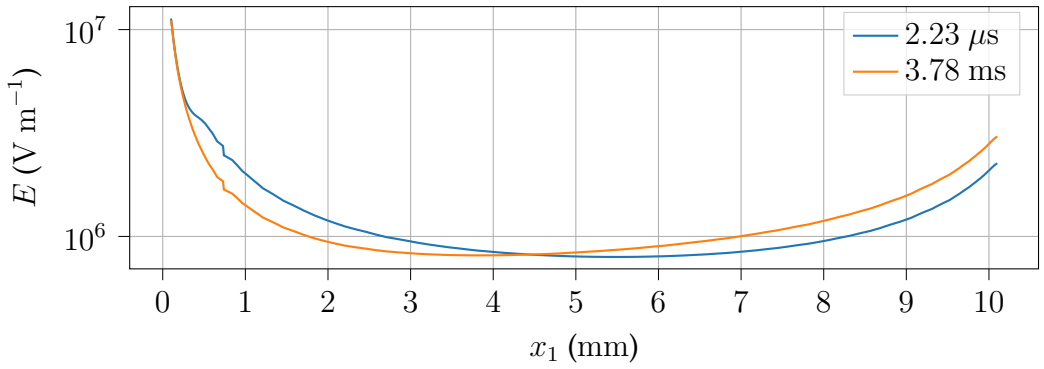


Figure 7.4: Field strength E between the electrodes on the symmetry axis

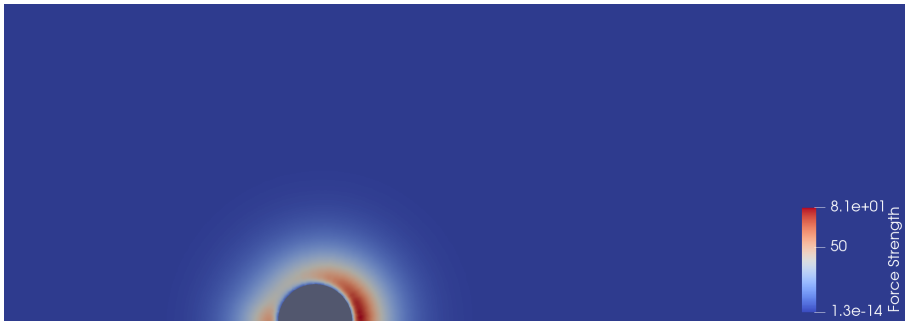


Figure 7.5: EHD force strength f_{EHD} (N m^{-3}) at $t = 1.17$ μs near the anode

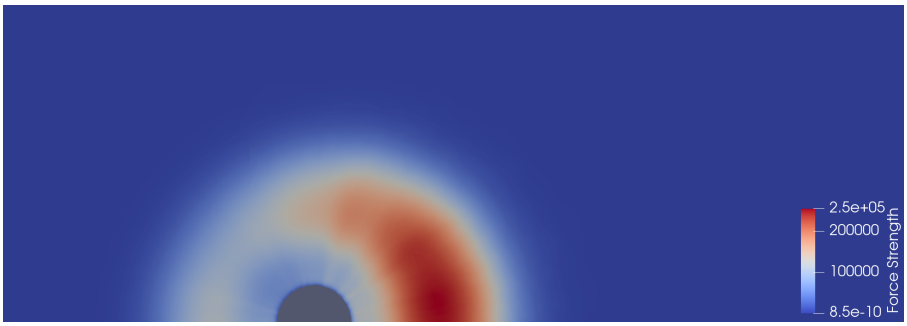


Figure 7.6: EHD force strength f_{EHD} ($\mathrm{N\ m^{-3}}$) at $t = 2.23\ \mu\mathrm{s}$

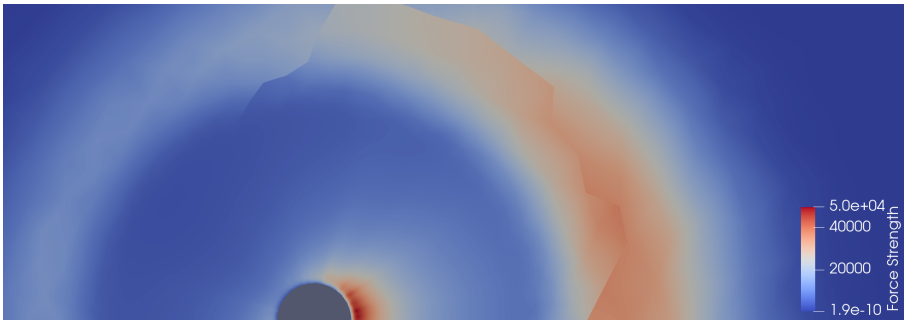


Figure 7.7: EHD force strength f_{EHD} ($\mathrm{N\ m^{-3}}$) at $t = 3.32\ \mu\mathrm{s}$

and the ion density is not high enough to create a high field as in streamers. As a result, the EHD force within the front loses its intensity gradually. Its strength now concentrates in the anode sheath (Figure 7.7) where the positive ions are continuously produced through stable ionization process.

From $t = 7.04\ \mu\mathrm{s}$ and on, the positive ions produced by ionization in the anode sheath continue to drift away, resulting in a wave traveling toward the cathode. This is the ion collecting phase of a corona discharge [109]. After some milliseconds, the ion density stabilizes at around $10^{16}\ \mathrm{m^{-3}}$ and the EHD force strength at around $16\mathrm{kN\ m^{-3}}$. Figure 7.8 shows the x_1-component of the EHD force density $f_1 \equiv \rho E_1$ at $t = 4\ \mathrm{ms}$ when the discharge reaches the steady state. The result shows that between the electrodes, there is a force density of $16\mathrm{kN\ m^{-3}}$ which points towards the cathode. On the left-hand side of the anode, the EHD force points in the reverse direction and reaches $-4.6\mathrm{kN\ m^{-3}}$.

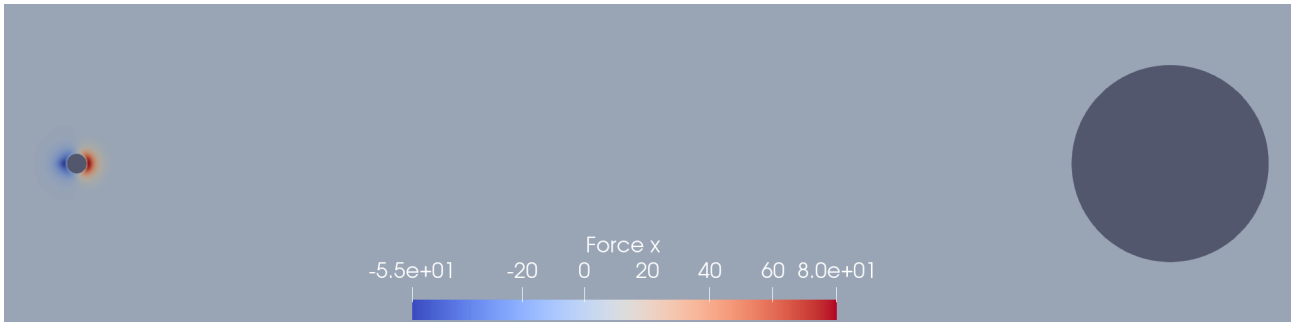


Figure 7.8: x_1-component f_1 of the EHD force density ($\mathrm{N\ m^{-3}}$) at $t = 4\ \mathrm{ms}$ (figure reflected on the symmetry axis)

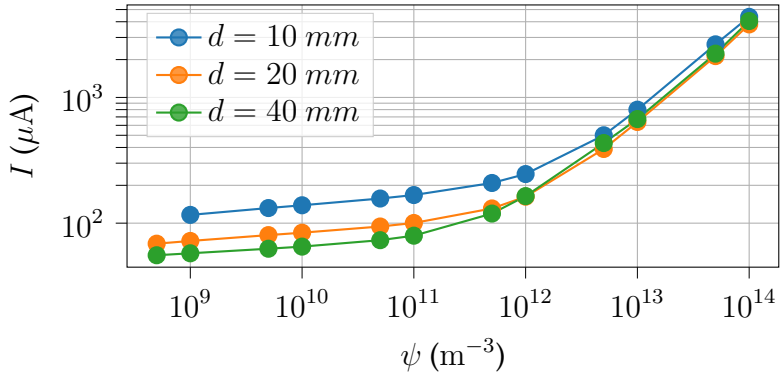


Figure 7.9: ψ-I curves for $r_1 = 0.1$ mm. The curve of $V_G = 13$ kV & $d = 10$ mm is colored in blue, $V_G = 20$ kV & $d = 20$ mm in orange, $V_G = 37$ kV & $d = 40$ mm in green.

7.2.3 Influence of the floor density on numerical solutions

We conduct a parametric study in which we compute the steady-state current for each (constant) electron floor density ψ. In fig. 7.9, we show the ψ-I curves for fixed anode radius $r_1 = 0.1$ mm while changing the wire gap d. For each d, the circuit voltage V_G is chosen closed to the maximal potential reached before the apparition of spark discharge as documented in [12]. In fig. 7.10, we do the reverse by fixing $d = 10$ mm and changing r_1.

In fig. 7.9, the minimal electron density, when there is no floor density, is more than 10^8 m^{-3} for all d, suggesting that the discharges are self-sustaining and do not need other sources of electrons other than impact ionization and secondary emission from the cathode surface. For this reason, we do not show in fig. 7.9 the values of I for ψ smaller than 5×10^8 m^{-3}. Although the discharges maintain themselves, the current is far from the experimental value if there is no additional electron source (see section 7.2.4 for $d = 10$ mm). The current I (μA) clearly depends on ψ as suggested in fig. 7.9, but the dependence seems to differ in each value interval of ψ. For ψ between 10^9-10^{11} m^{-3}, we have a (fitting) law $I \sim \psi^{0.07}$. For ψ between 10^{12}-10^{14} m^{-3}, $I \sim \psi^{0.7}$. This suggests that ψ should be between 10^9-10^{11} m^{-3} for this wire-to-wire discharge, since a too high value of ψ may overestimate the circuit current.

In fig. 7.10, we highlight the fact that in certain circumstances, the discharge is not self-sustaining unless there is an additional source of electrons other than electron impact ionization and surface secondary emission, which justifies the use of floor density. The curve corresponding to $r_1 = 0.5$ mm shows that I is close to 0 when ψ is small (1 m^{-3}), suggesting that there is no discharge at all although this is not true from experiment measurements (see fig. 7.11).

7.2.4 Comparison with experiment data

Being able to reproduce experiment data is the first step for a plasma solver to become a predictive simulation tool. In this section, we compare our numerical results to the data of Bérard et al. [12]. The simulations are conducted for $d = 10$ mm, r_1 takes values among 0.1, 0.175, 0.2 and 0.5 mm and V_G ranging from 9 to 17 kV. Figure 7.11 shows the numerical V-I curves with filled markers and the experiment data with hollow circle markers.

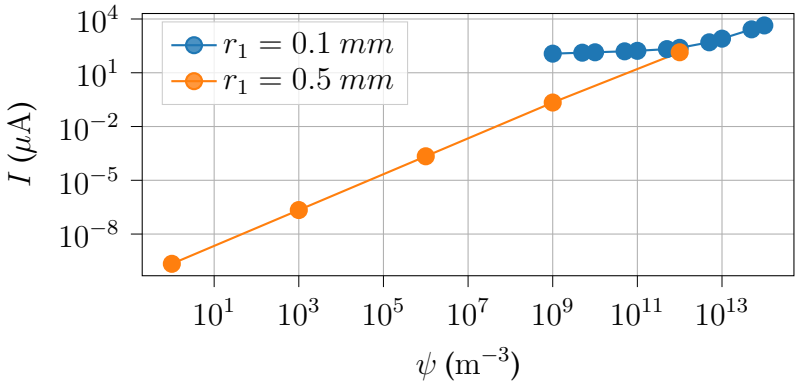


Figure 7.10: ψ-I curves for $d = 10$ mm. The curve of $r_1 = 0.1$ mm & $V_G = 13$ kV is colored in blue, $r_1 = 0.5$ mm & $V_G = 16.5$ kV in orange.

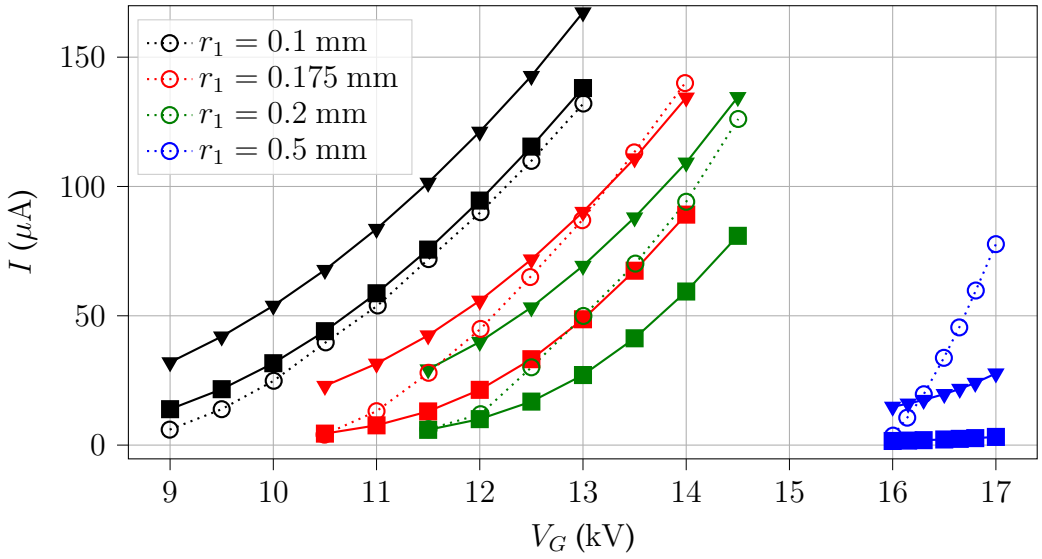


Figure 7.11: V-I curves for $d = 10$ mm and $r_1 = 0.1$ to 0.5 mm of experiment data (hollow circles) and numerical solutions (squares for $\psi = 10^{10}$, triangles for $\psi = 10^{-11}$ m^{-3})

As an attempt to fit the data, we use two values of ψ, 10^{10} m^{-3} (filled square markers) and 10^{11} m^{-3} (filled triangles). For $r_1 = 0.1$ mm, the numerical currents of $\psi = 10^{10}$ m^{-3} fit the data very well. For $r_1 = 0.175$ mm, the currents of $\psi = 10^{11}$ m^{-3} only agree with the experiments for high voltages, and as V_G decreases so does the slope of the numerical curve. The two smallest values seem to agree more with the 10^{10} m^{-3}-curve but this curve itself deviates too much for large voltages. The situation is quite similar for $r_1 = 0.2$ mm but the numerical solutions differ even more from experimental data. For $r_1 = 0.5$ mm, the simulation results are mediocre. Not only the currents are not exact, but the V-I slopes are completely wrong.

For corona discharges, the V-I characteristics is given by the law [126, Chapter 12]

$$I = k_{VI} V_G (V_G - V_c), \tag{7.1}$$

where V_c is the breakdown voltage and k_{VI} is a coefficient. The fitting of experimental and numerical currents with this quadratic law gives k_{VI} and V_c for each curve. The characteristics associated to the curves in fig. 7.11 are gathered in table 7.2.

	$r_1 = 0.1$ mm			$r_1 = 0.175$ mm			$r_1 = 0.2$ mm		
	data	$\psi = 10^{10}$	$\psi = 10^{11}$	data	$\psi = 10^{10}$	$\psi = 10^{11}$	data	$\psi = 10^{10}$	$\psi = 10^{11}$
k_{VI} (μA kV^{-2})	2.42	2.3	2.33	2.8	1.71	2.13	2.77	1.69	2.24
V_c (kV)	8.88	8.54	7.64	10.55	10.66	9.68	11.53	11.51	10.51

Table 7.2: V-I characteristics of experiment data [12] ("data") and numerical solutions for different values of ψ (m^{-3})

The results show that the ignition voltages V_c for $\psi = 10^{10}$ m^{-3} are closer to experiment data. On the contrary, the slope k_{VI} of $\psi = 10^{11}$ m^{-3} is closer to experiment data. In our opinion, the coefficient k_{VI} should be more emphasized than V_c, since V_c itself is the subject of many factors of incertitude and depends strongly on the discharge conditions (composition of gas, humidity, electrode material, etc.).

Overall, we find that the proposed numerical model provides quite good estimates of experiment data, considering the simplicity of the model, the first-order discretization in space and time and other factors in the experiments that could change the current that could not be precisely modeled, such as impurities on electrode surfaces or secondary emission.

7.3 Needle-to-ring corona discharge

7.3.1 Description of the test case

A sketch of the actuator is shown in fig. 7.12. The actuator is composed of a small metal needle which serves as the stressed electrode and a metal ring. The curvature of a the needle tip is σ and the length of the needle (excluding the tip) is L. The inner radius of the ring is r_1 and the outer radius is r_2. The thickness of the ring is h and the gap between it and the needle tip is d. A high voltage V_G is applied to the needle while the ring is grounded and the two electrodes are wired to a resistance R. The numerical computation is perform in cylindrical coordinate. In section 7.3.3, we evaluate the flow velocity generated by the discharge at a point M on the symmetry axis at a distance l downstream from the ring.

The corona ignition voltage V_c^e for different electrode gaps d were measured experimentally in [163] and are listed in table 7.3 for positive voltage polarity. For negative polarity at $d = 20$ mm, $V_c^e = -5.2$ kV. The overvoltage is defined as $\Delta V_G \equiv V_G - V_c^e$.

The grids are particularly refined near the needle tip as shown in fig. 7.13. The minimal grid size

d (mm)	20	25	30
V_c^e (kV)	7.1	7.4	8.0

Table 7.3: Experimentally measured ignition voltage V_c^e of positive corona discharge for different electrode gaps d

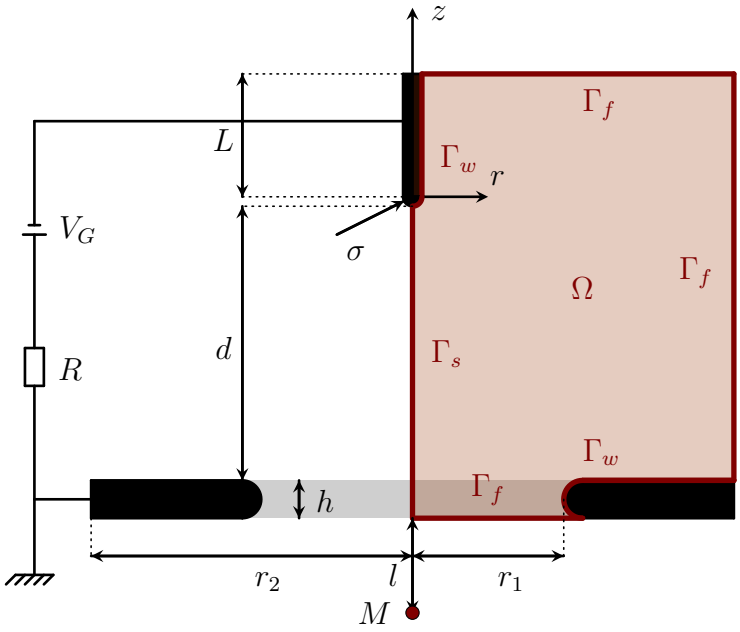


Figure 7.12: Sketch of the needle-to-ring actuator (not drawn to scale), the computation domain is colored in maroon

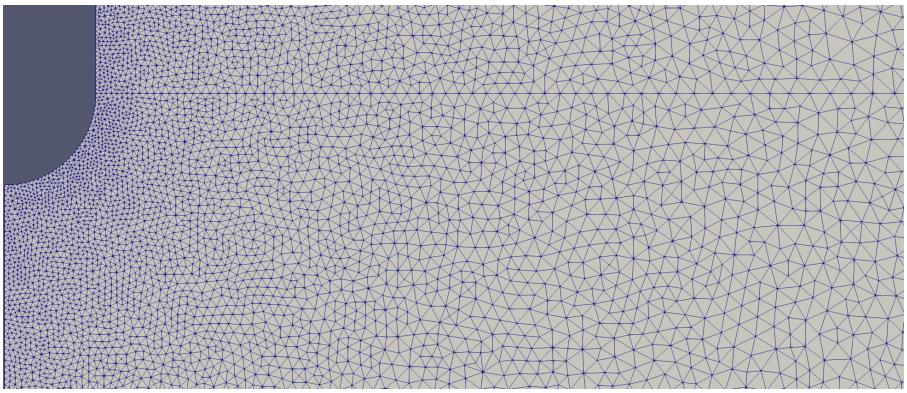


Figure 7.13: Grid refinement near the needle tip generated by Gmsh [58]

for simulations of positive corona discharge (section 7.3.3) is $\Delta x^{\min} = 4.5\ \mu\text{m}$. For negative corona simulations (section 7.3.4), the grid near the flat surface of the needle is also refined like the tip with $\Delta x^{\min} = 2.25\ \mu\text{m}$ in order to capture the microdischarges that appear near the tip in the beginning of the discharge which then propagate upwards.

The boundary conditions are shown on fig. 7.12 and described in section 2.2. The initial data are given in section 2.3. The electron floor density ψ is chosen between 10^9 and $10^{12}\ \text{m}^{-3}$. The secondary emission coefficient γ is fixed to 10^{-4} or 5×10^{-4}.

The simulation time is set at $T = 4\ \text{ms}$ for positive discharges (when steady-state solution is reached) and $T = 0.1\ \text{ms}$ for negative discharges. Furthermore, we use the kinetic scheme of Bœuf et al. in example 2.1 and the LFA model.

To avoid the formation of streamers in positive discharges, we increase the voltage gradually on $10\ \mu\text{s}$ until it reaches the maximal value V_G. On the contrary, the microdischarges always appear

during the negative corona no matter how large the voltage slope is, so we raise the voltage on $1\ \mathrm{ns}$ in this case.

Each simulation in section 7.3.3 was launched on 32 MPI processes and took **three to five hours** to complete. Each simulation in section 7.3.4, on the other hand, was on 72 processes but took about **a week** to finish.

7.3.2 Computation of ionic wind and numerical parameters

Uniquely for positive corona discharges, we use the EHD force density $\boldsymbol{f}_{\mathrm{EHD}} = \rho \boldsymbol{E}$ at steady state $T = 4\ \mathrm{ms}$ in conjunction with the incompressible Navier-Stokes equations to evaluate the induced airflow by the actuator. Since the characteristic timescales of the discharge dynamics (on the order of 10^{-5}-10^{-4} s) and the ionic wind ($1\ \mathrm{ms}$) are disparate [46], it is reasonable to solve them separately to save computation resources.

We solve for the steady-state solutions of the incompressible Navier-Stokes equations which read as follows [61, Chapter 2],

$$\begin{cases} \nabla \cdot \boldsymbol{u} = 0, \\ \nabla \cdot (\boldsymbol{u} \otimes \boldsymbol{u}) - \nabla \cdot (\nu_{\mathrm{eff}} \nabla \boldsymbol{u}) - \nabla \nu_{\mathrm{eff}} \cdot \nabla \boldsymbol{u} = -\dfrac{1}{\rho_0} \nabla p + \dfrac{1}{\rho_0} \boldsymbol{f}_{\mathrm{EHD}}, \end{cases} \tag{7.2}$$

where $\boldsymbol{u}(\boldsymbol{x})$ is the ionic wind velocity, $\nu_{\mathrm{eff}}(\boldsymbol{x})$ ($\mathrm{m^2\ s^{-1}}$) is the effective kinematic viscosity, $p(\boldsymbol{x})$ is the pressure and $\rho_0 \approx 1\ \mathrm{kg\ m^{-3}}$ is the volumetric mass of air.

The form of ν_{eff} depends on the flow model. In case of the (Newtonian) **laminar model**, $\nu_{\mathrm{eff}} = \nu$ where $\nu \approx 1.5 \times 10^{-5}\ \mathrm{m^2\ s^{-1}}$ is the kinematic viscosity of air. We also use the **realizable k-ε turbulence model** [133] with $\nu_{\mathrm{eff}} = \nu + \nu_t$ where $\nu_t(\boldsymbol{x})$ is the kinematic turbulence viscosity. ν_t is modeled by the law [104]

$$\nu_t = C_\nu \frac{k^2}{\varepsilon}, \tag{7.3}$$

with

$$\begin{gathered} C_\nu = \frac{1}{A_0 + A_S u^* \frac{k}{\varepsilon}}, \quad A_0 = 4, \quad A_S = \sqrt{6}\cos(\varphi), \quad \varphi = \frac{1}{3}\cos^{-1}(\sqrt{6}W), \\ W = \min\left(\max\left(\frac{(\boldsymbol{D}\boldsymbol{D}) \bullet \boldsymbol{D}}{\boldsymbol{D} \bullet \boldsymbol{D}}, -\frac{1}{\sqrt{6}}\right), \frac{1}{\sqrt{6}}\right), \quad \boldsymbol{D} = \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^t), \\ u^* = (\boldsymbol{D} \bullet \boldsymbol{D} + \boldsymbol{\Omega} \bullet \boldsymbol{\Omega})^{\frac{1}{2}}, \quad \boldsymbol{\Omega} = \frac{1}{2}(\nabla \boldsymbol{u} - \nabla \boldsymbol{u}^t). \end{gathered} \tag{7.4}$$

We recall that $(\boldsymbol{D}\boldsymbol{D})_{i,j} = \sum_k \boldsymbol{D}_{i,k}\boldsymbol{D}_{k,j}$ and $\boldsymbol{D} \bullet \boldsymbol{D} = \sum_{i,j} \boldsymbol{D}_{i,j}\boldsymbol{D}_{i,j}$.

The turbulence kinetic energy per unit mass $k(\boldsymbol{x})$ ($\mathrm{m^2\ s^{-2}}$) is given by

$$\begin{gathered} \nabla \cdot (k\boldsymbol{u}) = \nabla \cdot \left(\left(\nu + \frac{\nu_t}{\sigma_k}\right)\nabla k\right) + G_k - \varepsilon, \\ \sigma_k = 1, \quad G_k = 2\nu_t \boldsymbol{D} \bullet \boldsymbol{D}. \end{gathered} \tag{7.5}$$



Figure 7.14: Wedge-like domain for the computation of $\boldsymbol{u}$

σ	0.1 mm	L	5 mm	ΔV_G	2-14 kV (section 7.3.3)	$\Delta x^{\min}$	4.5 μm (section 7.3.3)
r_1	10 mm	l	5 mm	I_t	10%, 20%, 30%		2.25 μm (section 7.3.4)
r_2	20 mm	h	2 mm	R	20 kΩ	T	4 ms (section 7.3.3)
d	20-30 mm	γ	10^{-4}	ψ	10^9-10^{12} m^{-3}		0.1 ms (section 7.3.4)

Table 7.4: Parameters of needle-to-ring corona discharge simulations

And finally, the turbulence kinetic energy dissipation rate $\varepsilon(\boldsymbol{x})$ (m^2 s^{-3}) is given by

$$\begin{gathered}\nabla\cdot(\varepsilon\boldsymbol{u}) = \nabla\cdot\left(\left(\nu+\frac{\nu_t}{\sigma_\varepsilon}\right)\nabla\varepsilon\right) + C_{\varepsilon,1}\varepsilon(2\boldsymbol{D}\bullet\boldsymbol{D})^{\frac{1}{2}} - C_{\varepsilon,2}\frac{\varepsilon^2}{k+\sqrt{\nu\varepsilon}},\\ \sigma_\varepsilon = 1.2,\quad C_{\varepsilon,1} = \max\left(0.43,\frac{\chi}{\chi+5}\right),\quad \chi = \frac{k}{\varepsilon}(2\boldsymbol{D}\bullet\boldsymbol{D})^{\frac{1}{2}},\quad C_{\varepsilon,2} = 1.9.\end{gathered} \tag{7.6}$$

The system of eqs. (7.2) to (7.6) is entirely solved by the open-source CFD code OpenFOAM [160]. The computation of ionic wind necessitates a wedge-like domain (extrusion around the symmetry axis, see fig. 7.14) as well as a prescription of the inflow boundary conditions (on the top surface) for k and ε [61, Chapter 7] as follows,

$$k_{\text{in}} = \frac{3}{2}(I_t u^{\text{ref}})^2,\quad \varepsilon_{\text{in}} = \frac{C_\mu^{\frac{3}{4}} k_{\text{in}}^{\frac{3}{2}}}{l_m}, \tag{7.7}$$

where k_{in} and ε_{in} are resp. the inflow values of k and ε, I_t is the turbulence intensity, $u^{\text{ref}} = 3$ m s^{-1} is the reference inflow speed at the top surface, $C_\mu = 0.09$ and l_m is the Prandtl mixing length which is set at $l_m = 5\% \times 2\sigma$ (roughly 5% of the flow jet diameter).

The characteristics of the actuator and the numerical parameters are grouped in table 7.4. By **default**, $r_1 = 10$ mm, $r_2 = 20$ mm, $d = 20$ mm, $\gamma = 10^{-4}$ and $I_t = 30\%$ unless stated otherwise.

7.3.3 Positive corona

Steady-state solution

The positive ion at the steady state ($T = 4$ ms) concentrates in front of the needle tip with a density up to 2.5×10^{18} m^{-3} (see fig. 7.15). The density of other species are negligible comparing

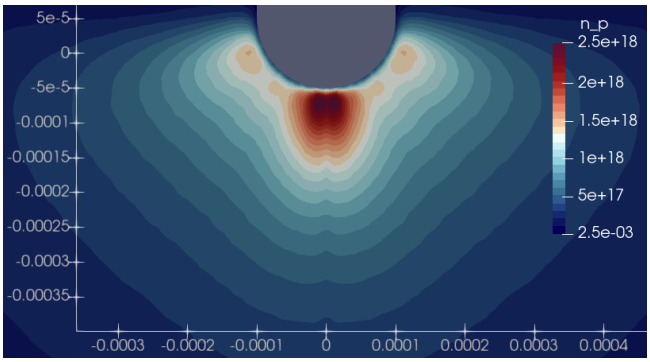
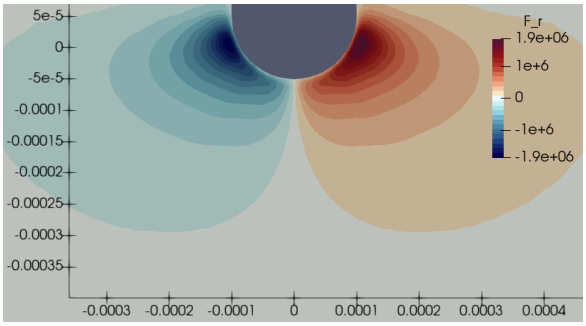
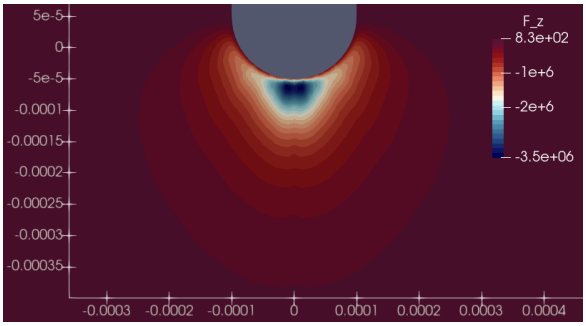


Figure 7.15: Positive ion density n_p (m^{-3}) near the needle tip for $\Delta V_G = 12\ \mathrm{kV}$ and $\psi = 10^{11}\ \mathrm{m}^{-3}$



(a) f_r



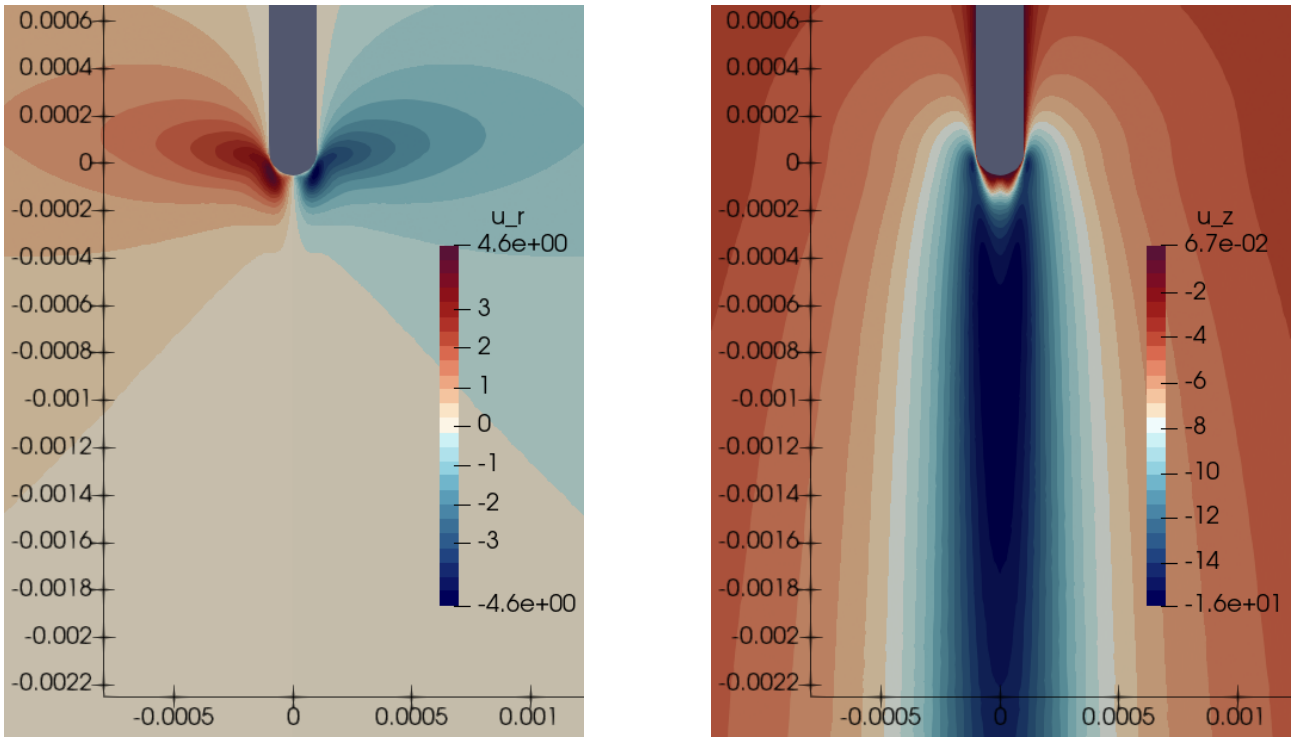
(b) f_z

Figure 7.16: Radial and axial components f_r and f_z ($\mathrm{N\ m}^{-3}$) of the EHD force density near the needle tip for $\Delta V_G = 12\ \mathrm{kV}$ and $\psi = 10^{11}\ \mathrm{m}^{-3}$

to n_p as n_e is on the order of $10^{16}\ \mathrm{m}^{-3}$ and n_n is on the order of $10^{15}\ \mathrm{m}^{-3}$. Therefore, n_p accounts for most of the EHD force. Figure 7.15 shows that the positive ions do not form a cone in front of the tip but rather a structure like the bottom of a bell pepper (the maximum position of n_p is not on but around the symmetry axis). The positive ions also forms a ring-like structure around the tip at $r \approx \sigma$ with n_p around $10^{18}\ \mathrm{m}^{-3}$. The special charge distribution around the anode seems to be a consequence of the ring geometry of the cathode.

As a result, the EHD force density has a large r-component f_r at $r \approx \sigma$ that acts on the outward direction from the needle as observed in fig. 7.16a. The influence of f_r on the induced airflow should be important in principle as the maximum strength of f_r is roughly two-third the maximum strength of the z-component force density f_z (see fig. 7.16b), but it is not clear how it effects the property of the ionic wind. Indeed, we can observe from figs. 7.17a and 7.17b that air is accelerated as if in an annular jet from the position $r \approx \sigma$ where f_r is maximal. However, despite f_r being oriented outward, the r-component of the flow velocity u_r actually points inwards to the stressed electrode.

The question on the mechanism of interaction between the plasma and the airflow will be left open for now, but it is an interesting one and should be revisited in the future. As we shall see in the rest of the discussion, the laminar Navier-Stokes equations overestimates significantly the flow speed comparing to experiment data. The use of the k-ε turbulence model with the tuning parameter I_t (turbulence intensity) gives much more coherent results otherwise, although the Reynolds number $\mathrm{Re} = u^{\max} 2\sigma/\nu$ of this jet is roughly 200, where $u^{\max} = 16\ \mathrm{m\ s}^{-1}$ is the maximum flow speed (see



(a) u_r (b) u_z

Figure 7.17: Radial and axial components u_r and u_z ($\mathrm{m\ s^{-1}}$) of the flow velocity near the needle tip obtained with the turbulence model, for $\Delta V_G = 12$ kV and $\psi = 10^{11}$ $\mathrm{m^{-3}}$

	data	$\psi = 10^9$	$\psi = 10^{11}$	$\psi = 10^{12}$
k_{VI} (μA kV^{-2})	0.174	0.12	0.121	0.119

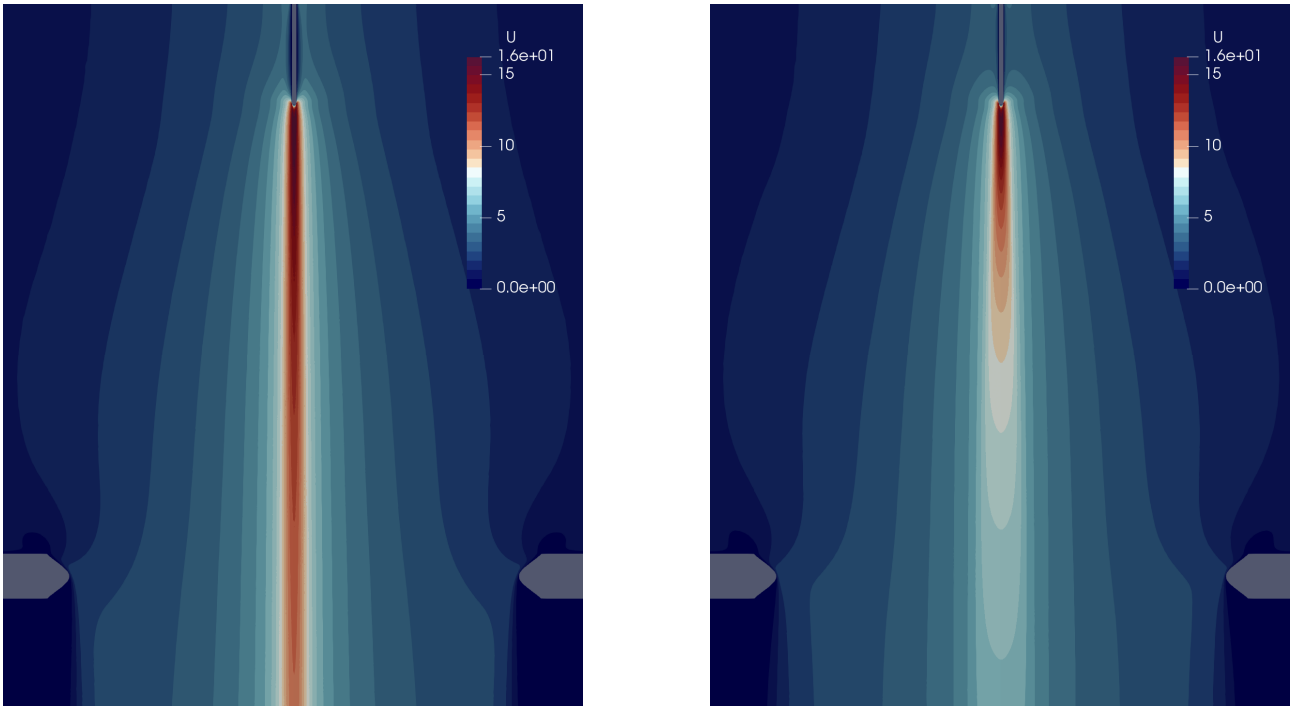
Table 7.5: V-I characteristics of experiment data [163] ("data") and numerical solutions for $d = 20$ mm and different values of ψ ($\mathrm{m^{-3}}$)

figs. 7.18a and 7.18b), 2σ is the characteristic length (diameter of the jet) and $\nu \approx 1.5 \times 10^{-5}$ $\mathrm{m^2\ s^{-1}}$ is the kinematic viscosity of air, so in principle the flow is laminar. In figs. 7.18a and 7.18b, we can already observe that the jet profiles of the laminar and the turbulence models are clearly different as the flow obtained with the turbulence model is visibly more diffusive in downstream comparing to the laminar flow.

Comparison with experiment data - circuit current

The same analysis as in section 7.2 on the V-I characteristics is carried out for the positive needle-to-ring discharge and the characteristic curves are shown on fig. 7.19 for different values of ψ. Using a quadratic fitting for the curves, we found that the V-I characteristics also satisfy the law (7.1) and the proportionality coefficients k_{VI} are listed in table 7.5.

Overall, the coefficients k_{VI} computed from the numerical solutions are practically the same but differ slightly from the experiment data from [163]. Since the curve of $\psi = 10^{11}$ $\mathrm{m^{-3}}$ seems to fit the experiment data better than $\psi = 10^9$ $\mathrm{m^{-3}}$ for high values of ΔV_G in the range of 10-12 kV, we will



(a) Laminar flow

(b) Turbulent flow

Figure 7.18: Magnitude of flow velocity $u = |\boldsymbol{u}|$ (m s^{-1}) obtained with the laminar model and the turbulence model (eqs. (7.2) to (7.6)), for $\Delta V_G = 12$ kV and $\psi = 10^{11}$ m^{-3}

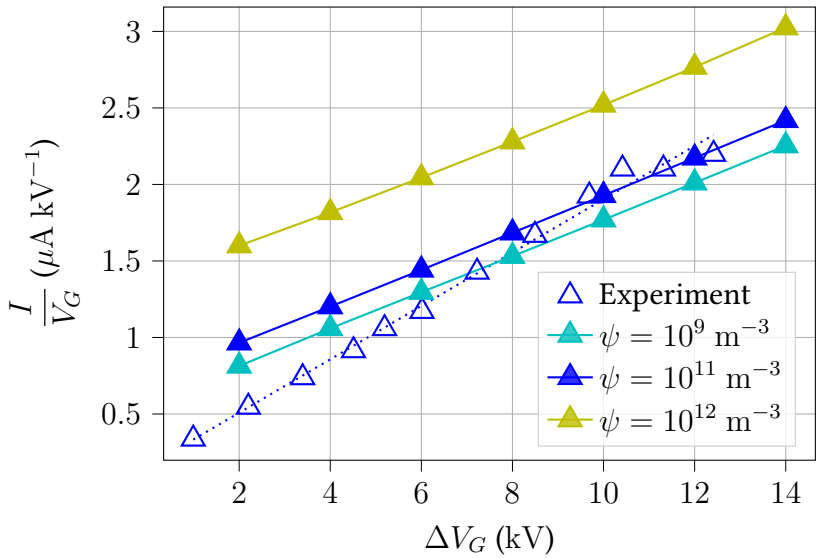


Figure 7.19: V-I curves for $d = 20$ mm. Comparison between experiment data [163] and numerical solutions for different values of ψ. The dotted line is the linear fitting of experiment data.

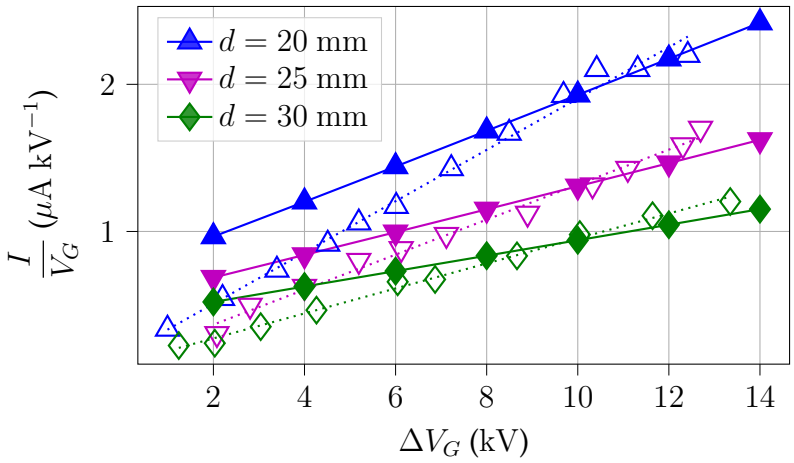


Figure 7.20: V-I curves for $\psi = 10^{11}$ m^{-3}. Comparison between experiment data [163] (hollow markers) and numerical solutions for different values of d (filled markers). The dotted lines are the linear fittings of experiment data.

	$d = 20$ mm		$d = 25$ mm		$d = 30$ mm	
	data	$\psi = 10^{11}$	data	$\psi = 10^{11}$	data	$\psi = 10^{11}$
k_{VI} (μA kV^{-2})	0.17	0.12	0.12	0.08	0.08	0.05

Table 7.6: V-I characteristics of experiment data [163] ("data") and numerical solutions for $\psi = 10^{11}$ m^{-3} and different values of d

set $\psi = 10^{11}$ m^{-3} for the rest of the positive needle-to-ring simulations.

Figure 7.20 show the V-I characteristics obtained with different values of d as well as experiment curves taken from [163]. As for $d = 20$ mm, the coefficients k_{VI} computed for $d = 25$ and 30 mm also differ slightly from experiment data (see table 7.6). Since the circuit current is quite sensitive to many factors in a discharge, such as gas composition or impurities on electrode surfaces, etc., these differences are not particularly concerning. Our conjecture is that it is more interesting to compare instead the aerodynamic effects since the timescale of ionic wind is considerably larger than that of the discharge, so it is probable that the ionic wind is less affected by discharge conditions.

Comparison with experiment data - flow velocity

After using the EHD force density corresponding to the discharge currents in fig. 7.20 to compute the steady-state flows, we find that the laminar model overestimates significantly the jet velocity comparing to experiment data even though the flow is in low-Reynolds regime. In fig. 7.21, we present the velocity profile for $d = 20$ mm and $\Delta V_G = 12$ kV on the line $z = -(\sigma + d + h + l)$, i.e. 5 mm downstream of the ring. The maximum jet speed reproduced by the laminar model is almost twice as large as the measurements in [163]. On the other hand, the maximum velocity from the experiments can be matched by numerical solutions of the turbulence model (eqs. (7.2) to (7.6)), but with a tuning parameter I_t (in eq. (7.7)) that represents turbulence intensity at inflow boundaries. We found that the computed velocity with $I_t = 30\%$ provides good agreement with data within the

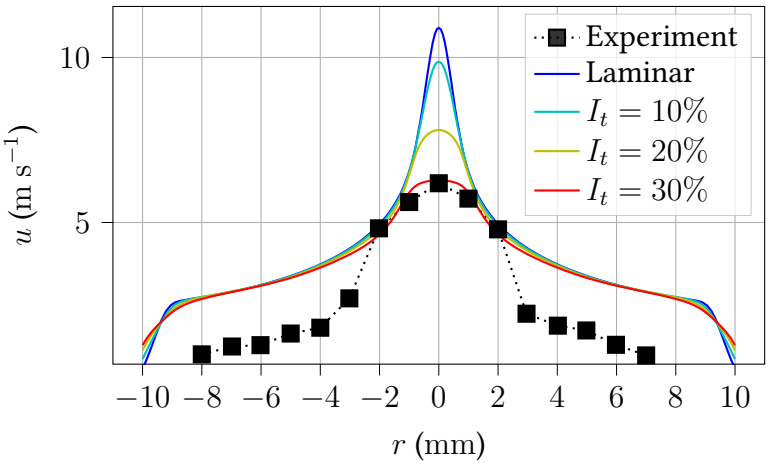


Figure 7.21: Ionic wind profile at $z = -(\sigma + d + h + l)$ (i.e. 5 mm downstream of the ring) for $d = 20$ mm and $\psi = 10^{11}$ m^{-3}. Comparison between experiment data [163] and numerical solutions obtained with laminar and turbulence models for different values of I_t.

	$d = 20$ mm		$d = 25$ mm		$d = 30$ mm	
	data	$I_t = 30\%$	data	$I_t = 30\%$	data	$I_t = 30\%$
k_{Vu} (m s^{-1} kV^{-1})	0.48	0.5	0.4	0.34	0.27	0.25

Table 7.7: V-u characteristics of experiment data [163] ("data") and numerical solutions for $I_t = 30\%$ and different values of d

jet "core" $r < 2$ mm. For this reason, we set $I_t = 30\%$ for the rest of the jet simulations.

It has been shown in [163] that jet velocity on the symmetry axis is linearly proportional to the overvoltage. In particular, we have

$$u_M = k_{\mathrm{Vu}} \Delta V_G, \tag{7.8}$$

where $k_{\mathrm{Vu}} > 0$ is the proportionality coefficient and is tabulated for each value of d in table 7.7.

The numerical coefficients k_{Vu} are actually closer to experiment data, except for $d = 25$ mm. Nevertheless, it seems that the measurements for $d = 25$ mm deviate a bit from the law (7.8) for ΔV_G in the range of 10-13 kV (see fig. 7.22). Therefore, the experimental slope k_{Vu} is probably not well estimated in this case. Otherwise, it is clear that the comparison of the voltage-velocity slope k_{Vu} is more reliable than the voltage-current slope k_{VI}.

7.3.4 Negative corona

Simulation of negative corona discharges is much more challenging as microdischarges always appear at the beginning of the simulation and the electron density can reach up to 10^{21} m^{-3}, thus heavily restricting the timesteps. The large ionization coefficient α due to strong field/electron mean energy is also a contributing factor to the stiffness of the negative corona simulations. Indeed, in section 6.5.5 we have seen the necessary condition $\mathcal{I} + \Delta t^l \overline{\mathcal{A}}^l_{pp} > 0$ for which the Gauss-Seidel algorithm could work. Let us recall that the matrix $\overline{\mathcal{A}}^l$ stems from the rearrangement of drift-diffusion-

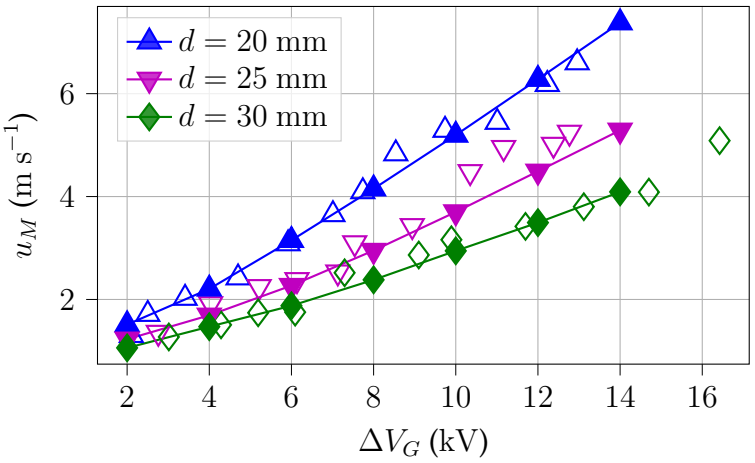


Figure 7.22: V-u curves for $I_t = 30\%$. Comparison between experiment data [163] (hollow markers) and numerical solutions for different values of d (filled markers).

reaction terms in eq. (6.43), so the condition $\mathcal{I} + \Delta t^l \overline{\mathcal{A}}^l_{pp} > 0$ can be achieved in two ways. *(i)* If the grid is sufficiently refined so that the flux-related term (from $\boldsymbol{f}_{e,K\lambda} \cdot \boldsymbol{\nu}_{K\lambda}$ in eq. (6.43)) dominates the ionization term $\alpha^l N$ (due to the factor $|\lambda|/|\Omega_K|$), then $\overline{\mathcal{A}}^l_{pp} > 0$. *(ii)* Otherwise, $\overline{\mathcal{A}}^l_{pp} < 0$ and Δt^l needs to be smaller than $(-\overline{\mathcal{A}}^l_{pp})^{-1}$. Either way, the simulation is costly in terms of computation resources. We refer to [157] for a more comprehensive reading on this constraint related to the ionization source term α.

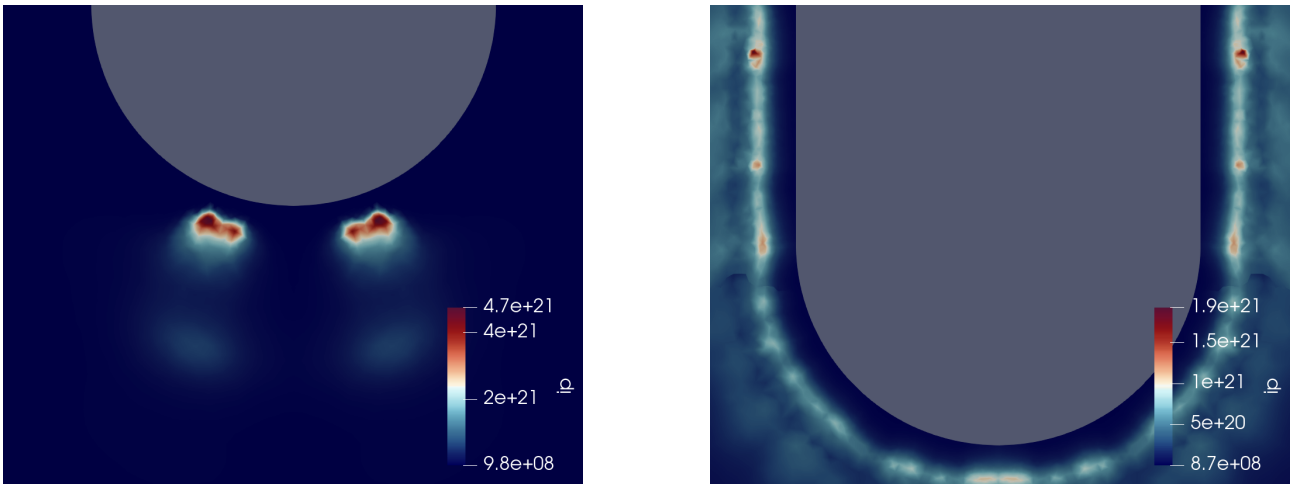
Let us denote as $S_\alpha \equiv \alpha N n_e$ the source term due to electron impact ionization reaction and $S^{l+1}_{\alpha,K}$ its discrete counterpart at t^{l+1} and on the cell Ω_K. All the simulations in sections 7.2 and 7.3.3 have been using $S^{l+1}_{\alpha,K} = \alpha^l_K N \overline{n}^{l+1}_{e,K}$ as described in eq. (6.43). In this section, simulations with this source term is not stable in the sense that the specie densities can reach nonphysical values and the code can crash. For this reason, numerical methods featuring this ionization source term are called **non-avalanche-stable** (NAS), adapting the terminology that was coined in [157]. Figure 7.23a shows the positive ion density obtained with the NAS method-LFA model at $t = 1.06$ ns for $d = 20$ mm and $V_G = -17.2$ kV. The maximum density is very high - 4.7×10^{21} m^{-3} - and keeps rising which leads to the code crashing shortly afterwards.

In order to stabilize the simulation, we use the **avalanche-stable** (AS) method proposed in [157] that features a new way to compute $S^{l+1}_{\alpha,K}$. The electrons are not created on Ω_K anymore but originates from influx electrons from neighbor cells. More precisely, $S^{l+1}_{\alpha,K}$ is replaced by

$$\widehat{S}^{l+1}_{\alpha,K} = \frac{\alpha^l_K}{G_K} \sum_{L \in \mathcal{V}^1_K} |\lambda_{LK}| \max(u^l_{e,LK}, 0) \overline{n}_{e,L}, \quad G_K = \sum_{L \in \mathcal{V}^1_K} |\lambda_{LK}| \max(u^l_{e,LK}, 0),$$

where we recall that $\mathcal{V}^1_K$ is the first-order neighborhood of Ω_K (see section 4.6) and $u^l_{e,LK}$ is an approximant of $\boldsymbol{u}(t^l, \boldsymbol{x}^\lambda_{LK}) \cdot \boldsymbol{\nu}_{LK}$.

The AS method-LFA model turns out to be not stable either. The simulation manages to go on until $t = 6$ ns but then crashes almost immediately. The last saved result shows that the positive ion density is up to 1.9×10^{21} m^{-3} but the maximum position is strangely not on the symmetry axis, as shown in fig. 7.23b.



(a) $t = 1.06$ ns (b) $t = 6$ ns

Figure 7.23: Positive ion density n_p (m^{-3}) near the needle tip, obtained with the LFA-NAS model (left) and the LFA-AS method (right) right before they crash

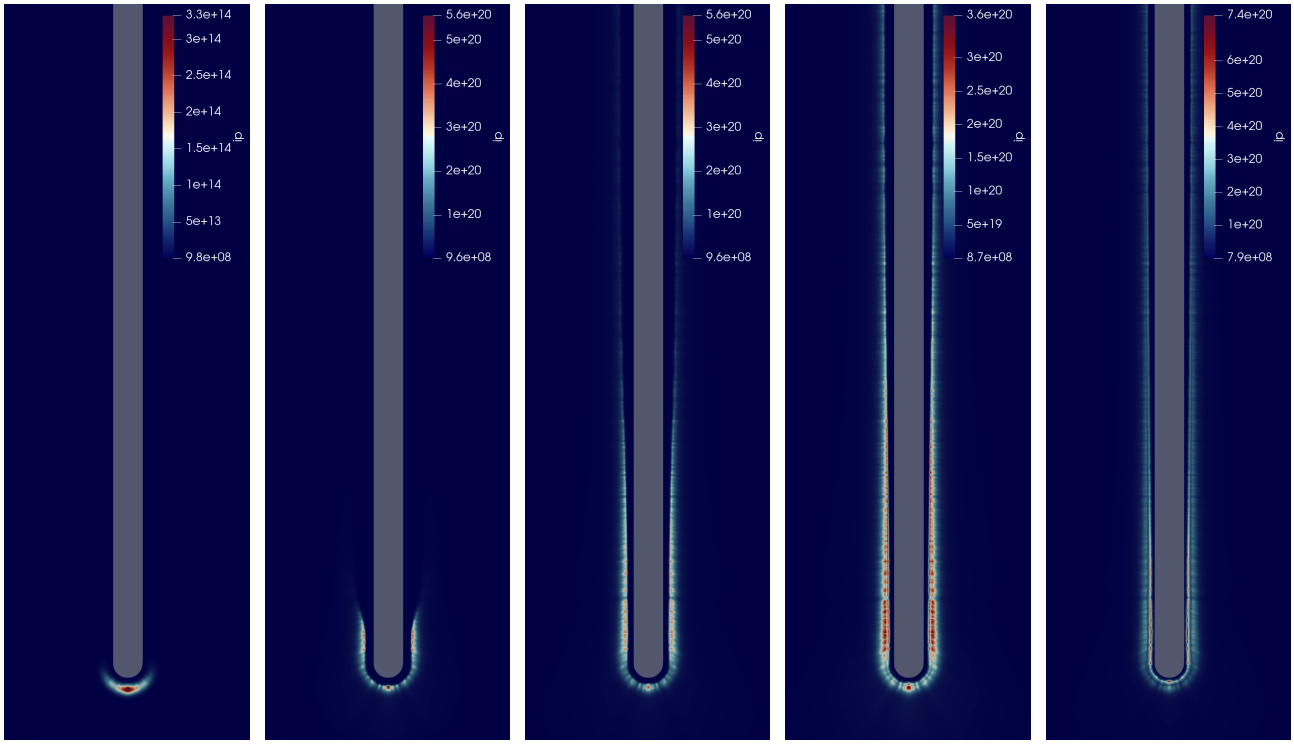
The problem lies in the limit of the LFA model that was discussed in section 2.5. The high concentration of positive charges close to the stressed electrode (cathode) generates an extremely strong field that is ready to ionize electrons that are diffused into the region near the cathode, since the ionization coefficient α depends on field strength in the LFA model. To remedy this problem, we use eq. (2.19) to modify α. This approach is dubbed as the **full-flux scheme** (FFS) in [138] and allow the simulation to complete at $T = 0.1$ ms.

The evolution of positive ion density, obtained with the LFA-AS-FFS method, during the first ten nanoseconds of the discharge is shown in fig. 7.24. The first avalanches occur in front of the tip and then extend upwards along the needle surface until the charges completely surround the electrode surface. The maximum positive ion density as well as electron density (not shown) well exceed 3×10^{20} m^{-3}, which explain the immense circuit current, on the order of -0.1 to -1 A on 0-10 ns (see fig. 7.25a).

From 10 ns to 60 μs, the charge density slowly decreases to around 7×10^{18} m^{-3} and so does the circuit current. But from $t = 60$ μs, successive current pulses begin to form and persist until the end of the simulation (see fig. 7.25b). The pulse peaks are on the order of -1 mA with the frequency of repetition around 0.7-1 MHz.

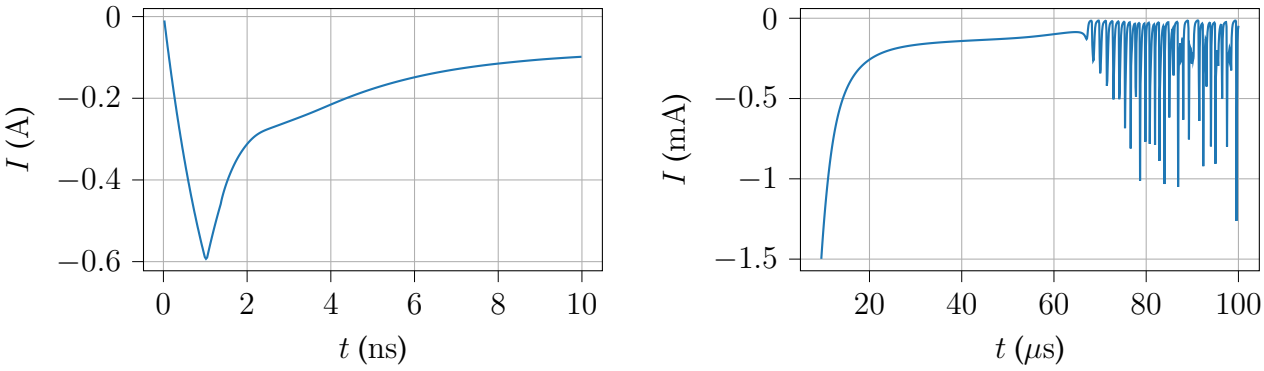
We also tested the LEA model combining with the AS method. The LEA-AS method persists until $t = 1.487$ ns but then crashes because the condition $\mathcal{I} + \Delta t^l \overline{\mathcal{A}}^l_{pp} > 0$ was violated for the electron mean energy. This suggest that the AS method should be extended to the (positive part of the) energy source term. Figure 7.26 compares the positive ions density obtained with the two methods LFA-AS-FFS and LEA-AS around $t = 1.5$ ns. It reveals that the maximum density of the first model is substantially smaller than the latter one, but the morphology of the particle distribution is the same. This suggests that the LEA-AS method is on the right track of development and only needs some improvements to be totally avalanche-stable.

Finally, we switched the explicit resolution method of the Poisson equation (see section 3.3.1) to



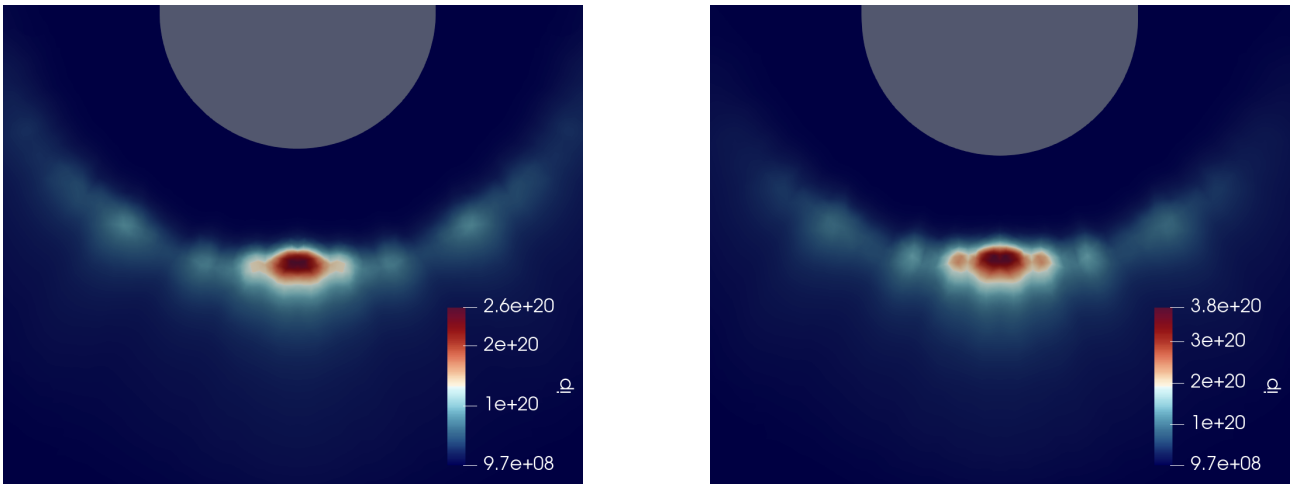
(a) $t = 1$ ns (b) $t = 2$ ns (c) $t = 4$ ns (d) $t = 6$ ns (e) $t = 10$ ns

Figure 7.24: Evolution of the positive ion density n_p (m^{-3}) near the needle tip during the first ten nanoseconds, obtained with the LFA-AS-FFS method



(a) 0-10 ns (b) 0-100 μs

Figure 7.25: Circuit current I obtained with the LFA-AS-FFS method



(a) $t = 1.508$ ns (b) $t = 1.487$ ns

Figure 7.26: Positive ion density n_p (m^{-3}) near the needle tip, obtained with the LFA-AS-FFS method (left) and the LEA-AS method (right) right before the latter crashes

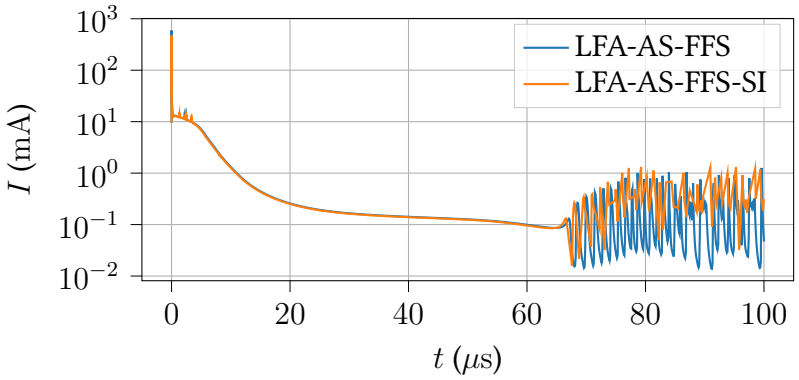


Figure 7.27: Comparison between the circuit currents I (log-scale of absolute values) obtained with the LFA-AS-FFS method and the LFA-AS-FFS-SI method

the semi-implicit method (SI) introduced in section 3.3.4. The comparison between the circuit currents I obtained with the LFA-AS-FFS method (explicit resolution of the field) and the LFA-AS-FFS-SI method in fig. 7.27 shows that the two methods yield very similar results. The timesteps of the LFA-AS-FFS-SI method is computed in the following way,

$$\Delta t^l = \min\left(\mathfrak{C}\Delta t^l_{\text{ions}}, \Delta t_{\phi,0}, \frac{0.1|V_G(t^l)|}{\left|\frac{\mathrm{d}V_G(t^l)}{\mathrm{d}t}\right| + \epsilon}\right),$$

with $\mathfrak{C} = 10^3$ and $\Delta t_{\phi,0} = 10^{-11}$ s. The LFA-AS-FFS-SI method took about **one hour** to compute the first ten nanoseconds of the discharge, where the dielectric relaxation time is on the order of 10^{-13}-10^{-12} s, as opposed to **three hours** with the LFA-AS-FFS method on the same computer cluster and with the same number of MPI processes.

7.4 Closing remarks

The implicit time method developed in chapter 6 has been implemented in COPAIER and tested against a wire-to-wire and a needle-to-ring corona discharges with direct-current voltages and constant floor densities ψ.

For the positive wire-to-wire discharge, a parametric study on the value of ψ shows that the numerical electric currents obtained with $\psi = 10^{11}\ \mathrm{m}^{-3}$ are coherent with experiment data for small anode radii. The numerical V-I curve, however, deviates more from the experimental curve as the larger the anode radius is. This highlights the drawback of the proposed discharge model in application to less-energetic discharges. However, since large-voltage discharges are more interesting as they generate stronger ionic wind, we find that the numerical results in this test cases are quite satisfactory, especially with the fact that they were obtained with a very simple kinetic model. It was also shown that the integration of ψ in the model is necessary to maintain the discharge in some cases.

For the positive needle-to-ring discharge, the numerical V-I curves are not coherent with experiment data, but the voltage-flow speed characteristics agree well with experiment measurements. This suggests that V-u curves are better comparison indicators than V-I curves since the circuit current I is more sensitive to discharge conditions than the flow speed u. It was also shown that the flow is turbulent and a boundary parameter of the flow - the turbulence intensity I_t - needs to be calibrated to obtain good values of flow speed. The cause of turbulent flows, as well as the correlation between I_t and u, are not yet determined.

Overall, the CPU time of positive discharges in this chapter was cut short to **a few hours** instead of **weeks** with explicit methods. Therefore, the performance of COPAIER is significantly enhanced with the proposed implicit method.

For the negative needle-to-ring discharge, the advantage of the proposed implicit method is hindered due to the appearance of microdischarges, especially during the first $10\ \mathrm{ns}$ of discharge where the numerical current can reach the order of $-1\ \mathrm{A}$. However, it still only took **five to seven days** to reach the simulation time $T = 0.1\ \mathrm{ms}$ as opposed to more than a **month** with explicit methods. Moreover, it has been shown that negative discharges are more challenging to compute than positive ones. Indeed, an avalanche-stable scheme as well as a modified evaluation of the ionization coefficient were employed lest more refined grids need to be used. Even so, a series of low-intensity microdischarges, exhibited by intermittent peaks of current, appear from $t = 70\ \mu\mathrm{s}$ and on but their nature is not yet to be understood. They could be Trichel pulses or numerical artifacts perhaps due to insufficient refinement of the grid.

7.5 Remarques finales

La méthode implicite en temps développée dans le chapitre 6 a été mise en œuvre dans COPAIER et testée sur les décharges couronne fil-fil et aiguille-anneau avec les tensions de courant continu et les densités de fond ψ constantes.

Pour la décharge positive fil-fil, une étude paramétrique sur la valeur de ψ montre que les courants électriques numériques obtenus avec $\psi = 10^{11}\ \mathrm{m}^{-3}$ sont cohérents avec les données expérimentales pour des petits rayons d'anode. La courbe numérique V-I, cependant, s'écarte davantage de la courbe expérimentale à mesure que le rayon de l'anode est plus grand. Ceci met en évidence l'inconvénient du modèle de décharge proposé pour l'application aux décharges moins énergétiques. Cependant, étant donné que les décharges à haute tension sont plus intéressantes car elles génèrent un vent ionique plus fort, nous trouvons que les résultats numériques dans ce cas test sont tout à fait satisfaisants, d'autant plus qu'ils ont été obtenus avec un modèle cinétique très simple. Il a également été montré que l'intégration de ψ dans le modèle est nécessaire pour maintenir la décharge dans certains cas.

Pour la décharge positive aiguille-anneau, les courbes numériques V-I ne sont pas cohérentes avec les données expérimentales, mais les caractéristiques tension-vitesse d'écoulement s'accordent bien avec les mesures expérimentales. Cela suggère que les courbes V-u sont des meilleurs indicateurs de comparaison que les courbes V-I puisque le courant du circuit I est plus sensible aux conditions de décharge que la vitesse d'écoulement u. Il a également été démontré que l'écoulement est turbulent et qu'un paramètre aux limites de l'écoulement - l'intensité de turbulence I_t - doit être calibré pour obtenir de bonnes valeurs de la vitesse d'écoulement. La cause des écoulements turbulents, ainsi que la corrélation entre I_t et u, ne sont pas encore déterminées.

Dans l'ensemble, le temps de calcul des décharges positives dans ce chapitre a été réduit à **quelques heures** au lieu de **semaines** avec les méthodes explicites. Par conséquent, la performance de COPAIER est considérablement améliorée avec la méthode implicite proposée.

Pour la décharge négative aiguille-anneau, l'avantage de la méthode implicite proposée est entravé par l'apparition de micro-décharges, en particulier pendant les 10 premières nanosecondes de la décharge où le courant numérique peut atteindre de l'ordre de -1 A. Cependant, il n'a fallu que **cinq à sept jours** pour atteindre le temps de simulation $T = 0.1$ ms, contre plus d'un **mois** avec les méthodes explicites. En outre, il a été démontré que les décharges négatives sont plus difficiles à calculer que les décharges positives. En effet, un schéma stable-aux-avalanches ainsi qu'une évaluation modifiée du coefficient d'ionisation ont été utilisés afin d'éviter l'utilisation de maillages plus raffinées. Malgré cela, une série de microdécharges de faible intensité, caractérisée par des pics de courant intermittents, apparaît à partir de $t = 70\ \mu$s, mais leur nature n'a pas encore été élucidée. Il pourrait s'agir d'impulsions de Trichel ou d'artefacts numériques dus à un raffinement insuffisant du maillage.

Part III

Conclusion and Prospects

Conclusion et Perspectives

Conclusion

La modélisation numérique des décharges de plasma dans l'air n'a jamais été une tâche facile en raison de sa nature multi-échelle. En effet, une décharge électrique est le siège de nombreux phénomènes physiques qui se produisent à des échelles de temps très différentes. Du point de vue de la modélisation, les modèles mathématiques et numériques doivent être suffisamment riches pour décrire correctement la dynamique du plasma, mais aussi suffisamment simples pour limiter la complexité numérique. Cette thèse travaille avec des descriptions hydrodynamiques du plasma, où les lois de conservation des espèces prennent la forme d'équations de dérive-diffusion, et aborde deux objectifs dans la modélisation numérique des décharges de gaz dans l'air : *(i)* amélioration de la qualité de la précision des solutions numériques et *(ii)* réduction du temps CPU des simulations.

Pour le premier objectif (abordé dans les chapitres 4 et 5), nous avons conçu de schémas de flux d'ordre élevé connus collectivement sous le nom de méthodes de Scharfetter-Gummel avec correction de courant (SGCC) pour résoudre les équations de dérive-diffusion, ce qui sont la généralisation du schéma standard de Scharfetter-Gummel (SG) qui est largement utilisé dans la simulation de plasma. La dérivation d'un flux SGCC est basée sur une approximation polynomiale de degré $p \in \mathbb{N}$ du flux de particules (le flux exact) dans un voisinage de chaque interface de cellule. Il a été démontré théoriquement que les flux SGCC-p sont consistants avec le flux exact et validés numériquement pour $p \leq 4$ avec des cas tests simples de transport-diffusion unidimensionnels et bidimensionnels. Ils ont également été appliqués à des simulations de décharges de gaz sur des maillages cartésiens, à savoir une décharge couronne fil-fil et une décharge de streamer positif. Sur le même maillage, les schémas SGCC d'ordre élevé fournissent des solutions numériques beaucoup plus précises que le schéma SG standard, et permettent donc d'économiser plus de ressources informatiques que le schéma SG pour obtenir la même qualité de résultats numériques.

Pour le deuxième objectif (abordé dans les chapitres 6 et 7), nous avons développé une stratégie d'intégration implicite en temps pour la simulation de la décharge couronne qui est basée sur une nouvelle formulation du modèle de plasma. Ce travail se concentre sur la réduction du temps de CPU tout en garantissant que les résultats numériques soient comparables avec les mesures expérimentales. Afin de calibrer les solutions numériques, une contrainte qui exige que la densité électronique soit toujours plus grande qu'une densité de fond ψ a été implémentée auparavant dans le solveur plasma de l'ONERA - COPAIER. Le système de décharge avec cette contrainte peut être résolu avec des schémas temporels explicites en utilisant une simple méthode de splitting puisque les pas de temps numériques sont petits, mais cette stratégie n'est pas pratique pour les schémas implicites puisque

les solutions de régime stationnaire dépendraient des pas de temps numériques. Par conséquent, la loi de conservation des électrons a été reformulée sous forme d'une inclusion différentielle afin que la contrainte de densité de fond puisse être directement intégrée dans le modèle de décharge. Certains algorithmes ont ensuite été explorés pour résoudre le système nonlinéaire résultant de la discrétisation implicite du modèle de décharge et un algorithme nonlinéaire de Gauss-Seidel s'émerge comme le candidat le plus solide. Cette approche a ensuite été utilisée dans des simulations de décharges couronnes fil-fil et aiguille-anneau avec COPAIER (sur des maillages triangulaires). Des études paramétriques ont ensuite montré que les solutions numériques avec une densité de fond de l'ordre de 10^{10}-10^{11} m^{-3} sont en bon accord avec les données expérimentales, et que le temps de CPU a été réduit avec succès, des **jours** ou **semaines** avec des méthodes explicites, à seulement des **heures** avec la méthode implicite proposée.

Perspectives

Il y a certainement des possibilités d'amélioration des méthodes numériques que nous avons introduites dans cette thèse, ainsi que des idées de travaux futurs.

Les méthodes SGCC (chapitres 4 et 5) n'ont été mises en œuvre que sur des maillages cartésiens. Il y a donc encore beaucoup de travail à faire pour que les schémas puissent être testés sur des maillages nonstructurés. De plus, le solveur de plasma que nous avons développé dans cette thèse pour valider les schémas SGCC n'a pas été parallélisé comme dans COPAIER. L'implémentation du parallélisme des schémas d'ordre élevé n'est pas nécessairement un travail facile puisque la reconstruction de la densité nécessite l'échange de données à travers les sous-domaines, et le volume de l'échange est déterminé par le stencil du schéma de flux qui est plus grand à mesure que l'ordre de précision augmente. Un algorithme de reconstruction spécifique adapté à MPI pourrait être nécessité, par exemple celui proposé dans [69].

En outre, le schéma standard de Scharfetter-Gummel est connu pour être uniformément convergent en fonction de la taille du maillage [127], ce qui signifie que la constante d'erreur du schéma est indépendante du gradient de la solution. Le schéma décentré classique du premier ordre, par exemple, ne possède pas cette propriété car son erreur de discrétisation augmente en réalité dès que le maillage est raffiné au point où les points de maillage commencent à entrer dans la couche de gradient fort de la solution, et ne diminue comme il se doit que lorsque le maillage est suffisamment raffiné et qu'il y a suffisamment de points de maillage à l'intérieur de la couche de gradient fort pour l'approcher de manière efficace. Il serait intéressant de savoir si les schémas SGCC héritent de la propriété de convergence uniforme du schéma SG.

En ce qui concerne la méthode implicite développée pour la simulation de la décharge couronne (chapitres 6 et 7), nous n'avons jusqu'à présent utilisé que des schémas du premier ordre pour la discrétisation en espace et en temps. Par conséquent, les solutions numériques pourraient ne pas être précises, ce qui entraîne une certaine incertitude quant à l'estimation de la densité de fond ψ qui a été obtenue par comparaison avec les données expérimentales. En effet, notre étude dans cette thèse a conclu que ψ est de l'ordre de 10^{10}-10^{11} m^{-3}, mais les résultats dans [99] (simulations par

COPAIER avec des schémas explicites du second ordre en espace et en temps) ont montré que ψ est de l'ordre de 10^{11}-10^{12} m^{-3}. La valeur de ψ pourrait bien sûr varier en fonction de chaque cas test, mais une étude complète utilisant des méthodes d'ordre élevé, par exemple les schémas SGCC, devrait être anticipée avant de tirer des conclusions.

En outre, la vitesse de convergence de l'algorithme nonlinéaire de Gauss-Seidel - le moteur qui a été construit pour résoudre le système discret de décharge - est d'autant plus lente que le maillage est fin. Il s'agit d'une observation à partir des simulations de décharge effectuées dans le cadre de cette thèse. Sur un maillage de taille minimale de 4.5 $\mu\mathrm{m}$, l'algorithme GS peut prendre jusqu'à 100 itérations pour converger; avec une taille minimale de 2.25 $\mu\mathrm{m}$, il peut prendre jusqu'à 10^3 itérations et avec une taille minimale de 1 $\mu\mathrm{m}$, le nombre maximum d'itérations peut atteindre 10^4. Par conséquent, il convient de concevoir un préconditionneur pour l'algorithme GS.

Enfin, le développement d'une technique implicite pour réduire le temps de CPU de la simulation des microdécharges en est encore à son début, comme le montre le dernier chapitre. Le point de départ est le schéma semi-implicite pour résoudre l'équation de Poisson. L'étude de cette question est toujours en cours et nécessitera plus d'efforts, certainement en dehors de la durée de cette thèse, pour traiter ce régime de décharge complexe.

Conclusion and Prospects

Conclusion

Numerical modeling of electric discharge in air has never been an easy task owing to its multiscale nature. An electric discharge is the siege of numerous intertwined physical phenomena that occur on disparate time scales. From the modeling point of view, mathematical and numerical models must be rich enough to correctly describe the plasma dynamics, but also sufficiently simple to limit numerical complexity. This thesis works with hydrodynamic descriptions of plasma, where the conservation laws of species take the form of drift-diffusion equations, and addresses two objectives in numerical modeling of gas discharge in air: *(i)* improving the precision quality of numerical solutions and *(ii)* reducing the CPU time of simulations.

For the first objective (addressed in chapters 4 and 5), we have conceived a set of high-order flux schemes collectively known as the Scharfetter-Gummel methods with correction of current (SGCC) to solve the drift-diffusion equations, which is a generalization of the standard Scharfetter-Gummel (SG) scheme that is widely used in plasma simulation. The derivation of a SGCC flux is based on a polynomial matching of degree $p \in \mathbb{N}$ of the particle flux (the continuous flux) in the neighborhood of each cell interface. The SGCC-p fluxes have been theoretically demonstrated to be consistent with the continuous flux and numerically validated for $p \leq 4$ with simple one-dimensional and two-dimensional transport-diffusion test cases. They have also been applied to simulations of gas discharge on cartesian grids, namely a wire-to-wire corona discharge and a positive streamer discharge. On the same grid, the high-order SGCC schemes provide much more accurate numerical solutions than the standard SG scheme, therefore they allow to save more computation resources than the SG scheme to obtain the same quality of numerical results.

For the second objective (addressed in chapters 6 and 7), we have developed an implicit time integration strategy for simulation of corona discharge that is based on a new formulation of the plasma model. This work focused on shortening the CPU time while simultaneously ensuring that the numerical results match the experiment measurements. In order to calibrate the numerical solutions, a constraint which requires that the electron density must be always larger than a floor density ψ had been implemented before in ONERA's plasma solver COPAIER. The discharge system with this constraint can be solved with explicit time schemes by using a simple splitting method since the numerical timesteps are small, but this strategy is impractical for implicit schemes since the computed steady-state solutions would depend on timesteps. Therefore, the conservation law of electrons were reformulated into a differential inclusion so that the floor density constraint could be directly integrated into the discharge model. Some algorithms were subsequently explored to

solve the nonlinear system that resulted from the implicit discretization of the discharge model and a nonlinear Gauss-Seidel algorithm stood out as the strongest candidate. This approach was later used in simulations of wire-to-wire and needle-to-ring discharges with COPAIER (on triangular grids). Parametric studies then showed that numerical solutions with a floor density on the order of 10^{10}-10^{11} m^{-3} agreed well with experiment data, and the CPU time was successfully cut short, from **days** or **weeks** with explicit methods, to a matter of **hours** with the proposed implicit method.

Prospects

There certainly rooms for improvements of the numerical methods that we have introduced in this thesis as well as for some ideas on future works.

The SGCC methods (chapters 4 and 5) were only implemented on cartesian grids. Hence, there is still a lot of work to do be done so that the schemes can be tested on unstructured grids. Furthermore, the plasma solver that we have developed in this thesis to validate the SGCC schemes was not parallelized as in COPAIER. The parallelism implementation of high-order schemes is not necessarily a light work since the density reconstruction requires exchange of data across subdomains, and the volume of exchange is determined by the stencil of the flux scheme which is larger as the precision order increases. A specific MPI-friendly reconstruction algorithm might be required, for example the one proposed in [69].

In addition, the standard Scharfetter-Gummel scheme is known to be uniformly convergent with respect to grid size [127], which means that the error constant of the scheme is independent of the gradient of the solution. The classical first-order upwind scheme, for example, lacks this property because its discretization error actually increases as soon as the grid is refined to the point where grid points begin to enter the large-gradient layer of the solution, and only decreases as it should have to when the grid is sufficiently refined and there are enough grid points inside the gradient layer to effectively approximate it. It would be interesting to know if the SGCC schemes inherit the uniform convergence property of the standard SG scheme.

On the implicit method developed for the simulation of corona discharge (chapters 6 and 7), so far we have only employed first-order schemes for discretization in space and time. As a result, the numerical solutions could be not accurate, causing some incertitude on the estimation of the floor density ψ which was obtained from comparison with experiment data. Indeed, our study in this thesis concluded that ψ is on the order of 10^{10}-10^{11} m^{-3}, but the results in [99] (simulations by COPAIER with explicit second-order schemes in space and time) showed that ψ is on the order of 10^{11}-10^{12} m^{-3}. The value of ψ could of course vary depending on each test case, but a comprehensive investigation using high-order methods, for example the SGCC schemes, should be expected before making any conclusions.

Furthermore, the convergence speed of the nonlinear Gauss-Seidel algorithm - the engine that was built to solve the discrete discharge system - is slower the more the grid is refined. This is an observation from the discharge simulations in this thesis. On a grid with a minimum size of 4.5 $\mu\mathrm{m}$, the GS algorithm can take up to 100 iterations to converge; with a minimum grid size of 2.25 $\mu\mathrm{m}$,

it can take up to 10^3 iterations and with a minimum grid size of $1\ \mu\mathrm{m}$, the maximum number of iterations can be as high as 10^4. As a result, a preconditioner for the GS algorithm should be devised.

Finally, the development of an implicit technique to reduce the CPU time of microdischarge simulations is still in its early stage as shown in the last chapter. The starting point is the semi-implicit scheme for solving the Poisson equation. The study on this issue is still ongoing and will take more efforts, certainly outside the duration of this thesis, to address this complex discharge regime.

Appendices

A Notations on multi-dimensional arrays - tensors

For an integer $m > 0$, let $\mathcal{S}^m\left(\mathbb{R}^2\right)$ be the space of symmetric m-dimensional arrays, or symmetric tensors of rank m, with 2 components on each dimension. For example, a vector is an element of $\mathcal{S}^1\left(\mathbb{R}^2\right)$, while a symmetric 2×2 matrix is an element of $\mathcal{S}^2\left(\mathbb{R}^2\right)$. For two elements $\boldsymbol{a}$ and $\boldsymbol{b}$ of $\mathcal{S}^m\left(\mathbb{R}^2\right)$, we define the **contraction** of $\boldsymbol{a}$ and $\boldsymbol{b}$ as

$$\boldsymbol{a} \bullet \boldsymbol{b} \equiv \sum_{k_1=1}^{2} \cdots \sum_{k_m=1}^{2} (\boldsymbol{a})_{k_1 \ldots k_m} (\boldsymbol{b})_{k_1 \ldots k_m}.$$

where $(\boldsymbol{a})_{k_1 \ldots k_m}$, $(\boldsymbol{b})_{k_1 \ldots k_m}$ are the elements of $\boldsymbol{a}$, $\boldsymbol{b}$.

For two tensors $\boldsymbol{a} \in \mathcal{S}^{m_1}\left(\mathbb{R}^2\right)$ and $\boldsymbol{b} \in \mathcal{S}^{m_2}\left(\mathbb{R}^2\right)$, the **tensor product** of $\boldsymbol{a}$ and $\boldsymbol{b}$ is an element of $\mathcal{S}^{m_1+m_2}\left(\mathbb{R}^2\right)$ and denoted as $\boldsymbol{a} \otimes \boldsymbol{b}$ such that

$$(\boldsymbol{a} \otimes \boldsymbol{b})_{k_1 \ldots k_{m_1} l_1 \ldots l_{m_2}} = (\boldsymbol{a})_{k_1 \ldots k_{m_1}} (\boldsymbol{b})_{l_1 \ldots l_{m_2}}.$$

For a tensor $\boldsymbol{a} \in \mathcal{S}^m\left(\mathbb{R}^2\right)$ and an integer $k > 0$, the power k of $\boldsymbol{a}$ is an element of $\mathcal{S}^{km}\left(\mathbb{R}^2\right)$ and is defined as

$$\begin{cases} \boldsymbol{a}^{\otimes 1} = \boldsymbol{a}, \\ \boldsymbol{a}^{\otimes k} = \boldsymbol{a}^{\otimes k-1} \otimes \boldsymbol{a}, \quad \text{for } k > 1. \end{cases}$$

For an integer $m > 0$, let $\mathfrak{O}^m$ be the group of permutations of the set $\{ 1, \ldots, m \}$ and $\overline{\mathcal{J}}^m \equiv \{ 1, 2 \}^m$ be the set of indices of a $\mathcal{S}^m\left(\mathbb{R}^2\right)$-tensor. The group $\mathfrak{O}^m$ induces a natural equivalence relation on $\overline{\mathcal{J}}^m$ [69] which is defined in the following way,

$$(i_1, \ldots, i_m) \sim (k_1, \ldots, k_m) \iff \exists \pi \in \mathfrak{O}^m, (i_1, \ldots, i_m) = \left(k_{\pi(1)}, \ldots, k_{\pi(m)}\right).$$

Hence, for a symmetric tensor $\boldsymbol{a}$ of $\mathcal{S}^m\left(\mathbb{R}^2\right)$ we have $(\boldsymbol{a})_{i_1,\ldots,i_m} = (\boldsymbol{a})_{k_1,\ldots,k_m}$ if $(i_1, \ldots, i_m) \sim (k_1, \ldots, k_m)$.

Let $\boldsymbol{i} = (i_1, \ldots, i_m) \in \overline{\mathcal{J}}^m$, then we could find a permutation $\pi \in \mathfrak{O}^m$ such that $i_{\pi(1)} \leq \cdots \leq i_{\pi(m)}$. Each equivalence class of $\overline{\mathcal{J}}^m$ is then represented by a unique element $\boldsymbol{i}$ such that $i_1 \leq \cdots \leq i_m$. Therefore, we define the quotient set $\mathcal{J}^m \equiv \overline{\mathcal{J}}^m / \sim$ that contains these representatives of all equivalence classes of $\overline{\mathcal{J}}^m$. We define also an order on $\mathcal{J}^m$ as follows,

$$(i_1, \ldots, i_m) \preceq (k_1, \ldots, k_m) \iff \begin{cases} i_1 < k_1, \\ (i_2, \ldots, i_m) \preceq (k_2, \ldots, k_m), \quad \text{if } i_1 = k_1. \end{cases}$$

with $(i_1, \ldots, i_m), (k_1, \ldots, k_m) \in \mathcal{J}^m$. Furthermore, let $\boldsymbol{i} \in \mathcal{J}^m$ and $\mathtt{1}(\boldsymbol{i})$ be the number of indices in $\boldsymbol{i}$ that equals 1, then the equivalence class of $\overline{\mathcal{J}}^m$ represented by $\boldsymbol{i}$ contains $\dfrac{m!}{\mathtt{1}(\boldsymbol{i})!(m - \mathtt{1}(\boldsymbol{i}))!}$ elements.

Example A.1. *Let $\boldsymbol{a} \in \mathcal{S}^3\left(\mathbb{R}^2\right)$. Then its components corresponding to each equivalence class of $\overline{\mathcal{J}}^3$ are* $\begin{cases} a_{111} \\ a_{112} = a_{121} = a_{211} \\ a_{122} = a_{212} = a_{221} \\ a_{222} \end{cases}$ *. There is only need to store the components corresponding to the elements of $\mathcal{J}^3$ which are sorted as $a_{111} \preceq a_{112} \preceq a_{122} \preceq a_{222}$.*

For any tensor $\boldsymbol{a} \in \mathcal{S}^m(\mathbb{R}^2)$, we define the tensor $\widetilde{\boldsymbol{a}}$ such that for any $\boldsymbol{i} = (i_1, \ldots, i_m) \in \mathcal{J}^m$,

$$\widetilde{\boldsymbol{a}}_{\boldsymbol{i}} = \frac{1}{\mathtt{1}(\boldsymbol{i})!(m - \mathtt{1}(\boldsymbol{i}))!} \left(\boldsymbol{a}\right)_{\boldsymbol{i}}.$$

B A brief presentation on total variation diminishing schemes

Definition B.1 ([90, 91]). *A function $n \in L^1_{\text{loc}}((0,T) \times \mathbb{R})$ belongs to the set of bounded-variation functions, denoted as $\text{BV}((0,T) \times \mathbb{R})$, if $|n|_{\text{BV}((0,T)\times\mathbb{R})} < +\infty$, where*

$$|n|_{\text{BV}((0,T)\times\mathbb{R})} = \limsup_{\epsilon \to 0} \frac{1}{\epsilon} \int_0^T \int_{\mathbb{R}} |n(t, x+\epsilon) - n(t,x)| \, \mathrm{d}x\mathrm{d}t + \limsup_{\epsilon \to 0} \frac{1}{\epsilon} \int_0^T \int_{\mathbb{R}} |n(t+\epsilon, x) - n(t,x)| \, \mathrm{d}x\mathrm{d}t.$$

For a piece-wise constant function n, for example $n(t,x) = n_i^l$ on $[t^l, t^{l+1}) \times \Omega_i$ for $l = 0, \ldots, \mathcal{L}$ and $i = 1, \ldots, \mathcal{N}$, the semi-norm $|n|_{\text{BV}((0,T)\times\mathbb{R})}$ simply reduces to

$$|n|_{\text{BV}((0,T)\times\mathbb{R})} = \sum_{l=0}^{\mathcal{L}} \sum_{i=-\infty}^{+\infty} \left(\Delta t^l \left|n_{i+1}^l - n_i^l\right| + \Delta x_i \left|n_i^{l+1} - n_i^l\right|\right),$$

where we have simply extended the function $n(t,\cdot)$ by 0 outside the domain Ω. The BV spaces play an very important role in demonstrating the convergence[2] of numerical solutions to the solution of eq. (4.1) when Δt, Δx tend to 0, since they are endowed with a compactness property. In fact, the following examples from resp. [53] and [90] are compact sets in $L^1_{\text{loc}}((0,T) \times \mathbb{R})$,

$$\mathcal{S}^1 = \text{BV}((0,T) \times \mathbb{R}) \cap L^\infty((0,T) \times \mathbb{R}),$$
$$\mathcal{S}^2 = \text{BV}((0,T) \times \mathbb{R}) \cap \{\, n \mid \exists M > 0, \text{Supp}(n(t,\cdot)) \subset [-M, M] \text{ a.e. } t \in [0,T] \,\}.$$

In [53], the $\mathcal{S}^1$ space was used to prove the convergence of numerical solutions, obtained by monotone flux schemes, for a general nonlinear hyperbolic equation. [90] extended the issue for a class of numerical methods known as **total variation diminishing** (TVD) schemes. Both used the Lax-Wendroff theorem (documented in these references) to attain the convergence proof.

[2] in the scalar case

Let us recall the time explicit discretization of (4.1),

$$\frac{\overline{n}_i^{l+1} - \overline{n}_i^l}{\Delta t^l} + \frac{\overline{f}_{i+\frac{1}{2}}^l - \overline{f}_{i-\frac{1}{2}}^l}{\Delta x_i} = 0, \tag{9}$$

where the numerical flux $\overline{f}_{i+\frac{1}{2}}^l$ is the function of $\overline{n}_k^l$ on some neighborhood cells Ω_k of Ω_i. Consider a function $\overline{n}(t,x)$ in $L^1_{\mathrm{loc}}((0,T)\times\mathbb{R})$ such that

$$\overline{n}(t,x) = \begin{cases} \overline{n}_i^l & \text{on } [t^l, t^{l+1}) \times \Omega_i, \\ 0 & \text{on } [0,T) \times (\mathbb{R} \setminus \Omega). \end{cases}$$

In this case, we have a useful lemma which says that we simply need to evaluate the total variation with respect to the space discretization.

Lemma B.1 ([90]). *Assume that $\overline{f}_{i+\frac{1}{2}}^l$ is Lipschitz-continuous with respect to all its arguments and there exists some $C > 0$ such that*

$$\left|\overline{n}(t^l,\cdot)\right|_{\mathrm{BV}(\mathbb{R})} \leq C, \quad l = 0, \dots, \mathcal{L}, \tag{10}$$

then $\overline{n} \in \mathrm{BV}((0,T)\times\mathbb{R})$. Here $\mathrm{BV}(\mathbb{R})$ is a set of functions $v \in L^1_{\mathrm{loc}}(\mathbb{R})$ such that $|v|_{\mathrm{BV}(\mathbb{R})} < +\infty$ with

$$|v|_{\mathrm{BV}(\mathbb{R})} = \limsup_{\epsilon\to 0} \frac{1}{\epsilon} \int_{\mathbb{R}} |v(x+\epsilon) - v(x)| \, \mathrm{d}x.$$

If v is a piece-wise function such that $v(x) = v_i$ on Ω_i, then we simply have $|v|_{\mathrm{BV}(\mathbb{R})} = \sum_{i=-\infty}^{+\infty} |v_{i+1} - v_i|$.

The numerical methods that satisfy eq. (10) are also known as BV schemes. The aforementioned TVD schemes, i.e. numerical methods such that $\left|\overline{n}(t^{l+1},\cdot)\right|_{\mathrm{BV}(\mathbb{R})} \leq \left|\overline{n}(t^l,\cdot)\right|_{\mathrm{BV}(\mathbb{R})}$ for $l = 0, \dots, \mathcal{L}-1$, constitute in fact a subset of BV schemes[3].

C Monotone operators and some useful results from functional analysis

Definition C.1 (Monotone operators [5, 8]). *A set-valued map M from $\mathcal{H}$ to $\mathcal{H}$ is called monotone, or accretive, if and only if $\forall g, h \in \mathrm{dom}(M), \forall v \in Mg, \forall w \in Mh$,*

$$(v - w, g - h)_{\mathcal{H}} \geq 0.$$

Furthermore, M is maximal monotone, or m-accretive, or maximal accretive[4] if there is no other monotone set-valued map $\widetilde{M}$ whose graph strictly contains the graph of M. Here, the graph of M is the set $\mathrm{graph}(M) = \{\, (g,v) \in \mathcal{H}\times\mathcal{H} \mid v \in Mg \,\}$.

[3] with the assumption that $\overline{n}(t^0,\cdot) \in \mathrm{BV}(\mathbb{R})$

[4] the three definitions coincide on the Hilbert space $\mathcal{H}$ and its dual $\mathcal{H}^*$ identified by the scalar product $(\cdot,\cdot)_{\mathcal{H}}$

An important property to identify maximal monotone operators is introduced in the following theorem.

Theorem C.1 ([5, 8]). *Let M be a monotone set-valued operator on $\mathcal{H}$. Then M is maximal monotone if and only if, for any $\lambda > 0$, $\mathrm{R}(\mathrm{Id} + \lambda M) = \mathcal{H}$, where $\mathrm{R}(M)$ denotes the range of M and Id denotes the identity operator.*

A class of operators "nearly" monotone/accretive is introduced as follows.

Definition C.2 (ω-accretive operators [8]). *A set-valued map M from $\mathcal{H}$ to $\mathcal{H}$ is called ω-accretive (resp. ω-m-accretive, with $\omega \in \mathbb{R}$, if $M + \omega I$ is accretive/monotone (resp. m-accretive/maximal monotone).*

Finally, we cite some results from functional analysis that are useful for the proofs in chapter 6 and appendix D.

Lemma C.1 ([25, Corollary 2.7]). *Let M and N be two maximal monotone operators. If $\mathring{\mathrm{dom}}(M) \cap \mathrm{dom}(N) \neq \emptyset$ then $M + N$ is maximal monotone and $\overline{\mathrm{dom}(M) \cap \mathrm{dom}(N)} = \overline{\mathrm{dom}(M)} \cap \overline{\mathrm{dom}(N)}$. Here $\mathring{S}$ and $\overline{S}$ are resp. the interior and closure of the set $S \subset \mathcal{H}$ with respect to the norm $|\cdot|_{\mathcal{H}}$.*

Theorem C.2 ([8, Chapter 3]). *Let M be an ω-m-accretive operator. Then, for each $n_0 \in \mathrm{dom}(M)$, there exists a unique function $n \in W^{1,1}([0,T];\mathcal{H})$ such that*

$$\begin{cases} \dfrac{\mathrm{d}n}{\mathrm{d}t}(t,\cdot) + Mu(t,\cdot) \ni 0, \quad \text{for a.e.} \quad t \in (0,T), \\ n(0,\cdot) = n_0, \end{cases}$$

and satisfies

$$\left|\frac{\mathrm{d}n}{\mathrm{d}t}(t,\cdot)\right|_{\mathcal{H}} \leq \exp(\omega t)\left|M^0 n_0\right|$$

a.e. on $(0,T)$. Here $W^{1,1}([0,T];\mathcal{H})$ is the space $\left\{ g \in L^1([0,T];\mathcal{H}) \mid \dfrac{\mathrm{d}g}{\mathrm{d}t} \in L^1([0,T];\mathcal{H}) \right\}$ and the map $u \mapsto M^0 u$ maps $u \in \mathrm{dom}(M)$ to the element of Mu having the smallest norm, i.e. $M^0 u$ is the projection of 0 on Mu, which exists and is unique since M is maximal monotone and thus Mu is a closed convex subset of $\mathcal{H}$ [5, Chapter 3].

Theorem C.3 (Aubin-Lions-Simon [22, Theorem II.5.16]). *Let B_0, B_1, B_2 be three Banach spaces, the embedding of B_1 in B_2 be continuous and the embedding of B_0 in B_1 be compact. Let p, r be such that $1 \leq p, r \leq +\infty$. For $T > 0$, let us define*

$$E_{p,r} = \left\{ g \in L^p(0,T;B_0) \mid \frac{\mathrm{d}g}{\mathrm{d}t} \in L^r(0,T;B_2) \right\}.$$

Then,

1. *if $p < +\infty$, the embedding of $E_{p,r}$ in $L^p(0,T;B_1)$ is compact;*
2. *if $p = +\infty$ and $r > 1$, the embedding of $E_{p,r}$ in $\mathcal{C}^0(0,T;B_1)$ is compact.*

D Regularity results for problems (6.2) and (6.3)

Lemma D.1. *Let $n_0 \in \mathrm{dom}(A)$ and $\zeta > 0$. Then the penalization problem (6.17) has a unique solution $n_\zeta \in W^{1,1}([0,T];\mathcal{H}) \cap V$ such that*

$$\left|\frac{\mathrm{d}n_\zeta}{\mathrm{d}t}(t,\cdot)\right|_{\mathcal{H}} \leq \exp(C_2 t)\left|A^0 n_0\right| \tag{11}$$

a.e. on $(0,T)$, where C_2 is the constant from eq. (6.13).

Proof. This result is a corollary of theorem C.2. Let us verify that the operator $A + \frac{1}{\zeta}B$ is ω-m-accretive. Indeed, for any $\lambda > 0$, the bilinear form $(\cdot,\cdot)_{\mathcal{H}} + \lambda\widetilde{a}(\cdot,\cdot)$ is continuous on $\mathcal{V}\times\mathcal{V}$, with $\widetilde{a}(\cdot,\cdot) \equiv a(\cdot,\cdot) + C_2(\cdot,\cdot)_{\mathcal{H}}$, and $|g|^2_{\mathcal{H}} + \lambda\widetilde{a}(g,g) \geq \lambda C_1\|g\|^2_{\mathcal{V}}$ for any $g \in \mathcal{V}$ (see eq. (6.13)). Therefore, according to the Lax-Milgram theorem [26, Chapter V], for any $h \in \mathcal{V}^*$, there exists a unique $g \in \mathcal{V}$ such that $(g,v)_{\mathcal{H}} + \lambda\widetilde{a}(g,v) = \langle h, v\rangle$ for all $v \in \mathcal{V}$, or in other words, $\langle g, v\rangle + \lambda\left\langle \widetilde{A}g, v\right\rangle = \langle h, v\rangle$ with $\widetilde{A} \equiv A + C_2\mathrm{Id}$. In particular, the range of the restriction of $\mathrm{Id} + \lambda\widetilde{A}$ to $\mathcal{H}$, still denoted as $\mathrm{Id} + \lambda\widetilde{A}$, is $\mathcal{H}$, i.e. $\mathrm{R}(\mathrm{Id} + \lambda\widetilde{A}) = \mathcal{H}$. Hence, $\widetilde{A}$ is a maximal monotone operator on $\mathcal{H}$ according to theorem C.1.

By the virtue of lemma C.1 with $M := \frac{1}{\zeta}B$ and $N := \widetilde{A}$, the operator $\widetilde{A} + \frac{1}{\zeta}B$ is maximal monotone on $\mathcal{H}$ (and $\mathrm{dom}\left(\widetilde{A} + \frac{1}{\zeta}B\right) = \mathrm{dom}(A)$ since $\mathrm{dom}(B) = \mathcal{H}$). Consequently, $A + \frac{1}{\zeta}B$ is ω-m-accretive with $\omega = C_2$.

From theorems C.2 and 6.2, $n_\zeta \in W^{1,1}([0,T];\mathcal{H}) \cap V$ and satisfies inequality (11). ∎

This lemma provides a bound on $\frac{\mathrm{d}n_\zeta}{\mathrm{d}t}(t,\cdot)$ which is usable to prove the existence of a strong solution of the variational inequality (6.14). Indeed, if $n_0 \notin \mathrm{dom}(A)$, then there is no viable estimate of $\frac{\mathrm{d}n_\zeta}{\mathrm{d}t}(t,\cdot)$ (see remark 6.4).

Theorem D.1. *Given $n_0 \in \mathrm{dom}(A) \cap \mathcal{K}$, then the problem (6.14) has a unique strong solution $n \in K$.*

Proof. It has been established that n is the unique weak solution of inequality (6.14). For $\zeta > 0$, let n_ζ be the solution of the penalization problem (6.17). From the proof of theorem 6.2, we have $n_\zeta \rightharpoonup n$ in V and $n_\zeta \to n$ in H. As a result, $\frac{\mathrm{d}n_\zeta}{\mathrm{d}t} \to \frac{\mathrm{d}n}{\mathrm{d}t}$ as $\zeta \to 0$ in the distribution sense.

We now show that[5] $|A^0_\zeta n_0|_{\mathcal{H}}$, with $A_\zeta \equiv A + \frac{1}{\zeta}B$, does not depend on ζ provided that $n_0 \in \mathrm{dom}(A) \cap \mathcal{K}$. Indeed, if $v_0 = A^0_\zeta n_0$ then $(v_0, v - v_0)_{\mathcal{H}} \geq 0$ and in other words, $|v_0|_{\mathcal{H}} \leq |v|_{\mathcal{H}}$ for any $v \in A_\zeta n_0$, since v_0 is the projection of 0 on the set $A_\zeta n_0$. Since $n_0 \in \mathcal{K}$ which means that $Bn_0 \ni 0$, we have $An_0 \in A_\zeta n_0$. Therefore, $|A^0_\zeta n_0|_{\mathcal{H}} \leq |An_0|_{\mathcal{H}}$ for any $\zeta > 0$.

[5] recall that the definition of the operator A^0 for an operator A is given in theorem C.2

From eq. (11), we have in particular that

$$\left\| \frac{\mathrm{d}n_\zeta}{\mathrm{d}t} \right\|_{L^\infty(0,T;\mathcal{H})} \leq \exp(C_2 T)|A_\zeta^0 n_0|_{\mathcal{H}} \leq \exp(C_2 T)|An_0|_{\mathcal{H}}.$$

As a consequence, there exists $v \in L^\infty(0,T;\mathcal{H})$ and a subsequence, still denoted as (ζ), such that $\dfrac{\mathrm{d}n_\zeta}{\mathrm{d}t}$ converges to v weakly-* in $L^\infty(0,T;\mathcal{H})$ and in particular weakly in $H = L^2(0,T;\mathcal{H})$ since Ω is bounded. We can identify next that $\dfrac{\mathrm{d}n}{\mathrm{d}t} = v$ and hence, $n \in H^1(0,T;\mathcal{H})$.

Multiplying the first equation of (6.16) with $g - n_\zeta$ with any $g \in K$ and using the fact that $(Bn_\zeta, g - n_\zeta)_{\mathcal{H}} \leq 0$ yield

$$\left\langle \frac{\mathrm{d}n_\zeta}{\mathrm{d}t}, g - n_\zeta \right\rangle + a(n_\zeta, g - n_\zeta) \geq 0.$$

Now letting $\zeta \to 0$ and using the property $\liminf_{\zeta \to 0} a(n_\zeta, n_\zeta) \geq a(n,n)$, we conclude that n is the strong solution of problem (6.14) in the sense of definition 6.3. ■

E Implicit integration of the SGCC schemes

Recall that the SGCC-p flux for a specie $s \in \mathfrak{S}$, at time t^l, reads as follows (see definition 4.2 and remark 4.6),

$$f^{l|p}_{s,i+\frac{1}{2}} = \frac{D^l_{s,i+\frac{1}{2}}}{\Delta x_{i+\frac{1}{2}}} \left(\mathcal{B}(\mathfrak{p}^l_{s,i+\frac{1}{2}}) Q^{l|p}_{s,i+\frac{1}{2},i} - \mathcal{B}(-\mathfrak{p}^l_{s,i+\frac{1}{2}}) Q^{l|p}_{s,i+\frac{1}{2},i+1} \right).$$

It can be decomposed as

$$\overline{f}^{l|p}_{s,i+\frac{1}{2}} = \overline{f}^{l|0}_{s,i+\frac{1}{2}} + \widetilde{f}^{l|p}_{s,i+\frac{1}{2}},$$

which is the sum of the standard Scharfetter-Gummel flux and a high-order correction which writes as

$$\widetilde{f}^{l|p}_{s,i+\frac{1}{2}} = \frac{D^l_{s,i+\frac{1}{2}}}{\Delta x_{i+\frac{1}{2}}} \Big(\mathcal{B}(\mathfrak{p}^l_{s,i+\frac{1}{2}}) \big(Q^{l|p}_{s,i+\frac{1}{2},i} - \overline{n}^l_{s,i} \big) - \mathcal{B}(-\mathfrak{p}^l_{s,i+\frac{1}{2}}) \big(Q^{l|p}_{s,i+\frac{1}{2},i+1} - \overline{n}^l_{s,i+1} \big) \Big),$$

In the above equation, $\big(Q^{l|p}_{s,i+\frac{1}{2},i} - \overline{n}^l_{s,i} \big)$ is usually a nonlinear term due to the presence of slope limiters. Following the steps in section 6.5, we can write the discretization of the conservation laws with the SGCC scheme in the following way,

$$(\overline{\mathcal{I}} + \Delta t^l \overline{\mathcal{A}}^l + \Delta t^l \widetilde{\mathcal{A}}^{l|p} - \Delta t^l \boldsymbol{S}_{\overline{\psi}}) \mathfrak{U}^{l+1} \ni \mathfrak{R}^l,$$

where $\widetilde{\mathcal{A}}^{l|p}$ is the nonlinear operator ensued from $\widetilde{f}^{l|p}$. We adapt a fixed-point method to solve this problem (see algorithm 3).

Algorithm 3 Fixed-point

1: Let $\epsilon > 0$, $U^0 = \mathfrak{U}^l$, $\delta > \epsilon$
2: **while** $\delta > \epsilon$ **do**
3: Set $V^q = \widetilde{\mathcal{A}}^{l|p} U^q$
4: Use algorithm 1 or algorithm 2 to solve $(\overline{\mathcal{I}} + \Delta t^l \overline{\mathcal{A}}^l - \Delta t^l \boldsymbol{S}_{\overline{\psi}}) U^{q+1} \ni \mathfrak{R}^l - \Delta t V^q$
5: Let $\left(n_e^{(q+1)} \quad n_p^{(q+1)} \quad n_n^{(q+1)}\right)^t = U^{q+1}$
6: Update $\delta = \max_{s \in \mathfrak{S}} \left(\frac{|n_s^{(q+1)} - n_s^{(q)}|_\infty}{|n_s^{(q)}|_\infty} \right)$
7: **end while**
8: Set $\mathfrak{U}^{l+1} = U^q$

F MPI implementation of the Gauss-Seidel algorithm

The sequential nature of the Gauss-Seidel algorithm makes the parallel-computing implementation not evident. In fact, if we imagine a simple partition of the simulation domain (see fig. 28) and the cells on the second subdomain all have ordering numbers larger than those on the first subdomain, then the density update on the second subdomain would have to wait for the update on the first subdomain to finish before getting started. For this reason, we rather send the values of U^m than those of U^{m+1} (U^m, U^{m+1} defined in algorithm 2) on the border cells (yellow triangles) from the first subdomain to the second and vice versa. Meanwhile, the variables on interior cells (red triangles) are updated normally by the Gauss-Seidel algorithm. The overall algorithm resembles a block-Jacobi method where each block is affiliated to a subdomain.

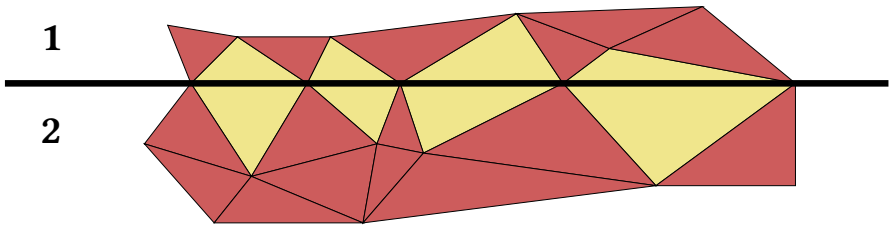


Figure 28: In this figure the computation domain is partitioned into two subdomains, sharing the thick black line as interface. The triangles which share an edge with the interface are called border cells and are painted in yellow. The other triangles are called interior cells and are painted in red.

Bibliography

[1] M Abramowitz, IA Stegun, and RH Romer. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. 1988.

[2] K Adamiak and P Atten. "Simulation of corona discharge in point–plane configuration". In: *Journal of electrostatics* 61.2 (2004), pp. 85–98.

[3] G Allaire. *Numerical analysis and optimization: an introduction to mathematical modelling and numerical simulation*. OUP Oxford, 2007.

[4] E Arcese, F Rogier, and JP Boeuf. "Plasma fluid modeling of microwave streamers: Approximations and accuracy". In: *Physics of Plasmas* 24.11 (2017).

[5] JP Aubin and A Cellina. "Differential Inclusions". In: *Grundlehren der mathematischen Wissenschaften* (1984).

[6] JLF Azevedo, LF Figueira da Silva, and D Strauss. "Order of accuracy study of unstructured grid finite volume upwind schemes". In: *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 32 (2010), pp. 78–93.

[7] B Bagheri et al. "Comparison of six simulation codes for positive streamers in air". In: *Plasma Sources Science and Technology* 27.9 (2018), p. 095002.

[8] V Barbu. *Nonlinear differential equations of monotone types in Banach spaces*. Springer Science & Business Media, 2010.

[9] TJ Barth. "Aspects of unstructured grids and finite-volume solvers for the Euler and Navier-Stokes equations". In: *AGARD, Special Course on Unstructured Grid Methods for Advection Dominated Flows* (1992).

[10] TJ Barth and D Jespersen. "The design and application of upwind schemes on unstructured meshes". In: *27th Aerospace sciences meeting*. 1989, p. 366.

[11] Ph Belenguer and JP Boeuf. "Transition between different regimes of rf glow discharges". In: *Physical Review A* 41.8 (1990), p. 4447.

[12] Ph Bérard, D Lacoste, and C Laux. "Measurements and simulations of the ionic wind produced by a DC corona discharge in air, helium and argon". In: *38th AIAA Plasmadynamics and Lasers Conference In conjunction with the 16th International Conference on MHD Energy Conversion*. 2007, p. 4611.

[13] M Bessemoulin-Chatard. "A finite volume scheme for convection–diffusion equations with nonlinear diffusion derived from the Scharfetter–Gummel scheme". In: *Numerische Mathematik* 121.4 (2012), pp. 637–670.

[14] M Bessemoulin-Chatard, C Chainais-Hillairet, and MH Vignal. "Study of a finite volume scheme for the drift-diffusion system. Asymptotic behavior in the quasi-neutral limit". In: *SIAM Journal on Numerical Analysis* 52.4 (2014), pp. 1666–1691.

[15] JP Boeuf, Y Lagmich, and LC Pitchford. “Contribution of positive and negative ions to the electrohydrodynamic force in a dielectric barrier discharge plasma actuator operating in air”. In: *Journal of applied physics* 106.2 (2009), p. 023115.

[16] JP Boeuf and LC Pitchford. “Two-dimensional model of a capacitively coupled rf discharge and comparisons with experiments in the Gaseous Electronics Conference reference reactor”. In: *Physical Review E* 51.2 (1995), p. 1376.

[17] JP Boeuf and LC Pitchford. “Electrohydrodynamic force and aerodynamic flow acceleration in surface dielectric barrier discharge”. In: *Journal of Applied Physics* 97.10 (2005), p. 103307.

[18] JP Boeuf, T Unfer, and LC Pitchford. *Studies of the Electrohydrodynamic Force Produced in a Dielectric Barrier Discharge for Flow Control. Report no. 2, Phase 3*. Tech. rep. Université Paul Sabatier Toulouse (France), 2010.

[19] JP Boeuf et al. “Electrohydrodynamic force in dielectric barrier discharge plasma actuators”. In: *Journal of Physics D: Applied Physics* 40.3 (2007), p. 652.

[20] H Borradaile et al. “Flow reversal in millimetric annular DBD plasma actuator”. In: *Journal of Physics D: Applied Physics* 54.34 (2021), p. 345202.

[21] A Bourdon et al. “Efficient models for photoionization produced by non-thermal gas discharges in air based on radiative transfer and the Helmholtz equations”. In: *Plasma Sources Science and Technology* 16.3 (2007), p. 656.

[22] F Boyer and P Fabrie. *Mathematical Tools for the Study of the Incompressible Navier-Stokes Equations andRelated Models*. Vol. 183. Springer Science & Business Media, 2012.

[23] D Breden, K Miki, and LL Raja. “Self-consistent two-dimensional modeling of cold atmospheric-pressure plasma jets/bullets”. In: *Plasma Sources Science and Technology* 21.3 (2012), p. 034011.

[24] H Brezis. “Inéquations variationnelles paraboliques”. In: *Séminaire Jean Leray* (1971), pp. 1–10.

[25] H Brezis. *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. Elsevier, 1973.

[26] H Brezis. “Analyse fonctionnelle, théorie et applications, dunod, Paris, 1999”. In: *Nouvelle présentation* (2005).

[27] H Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Vol. 2. 3. Springer, 2011.

[28] TMP Briels et al. “Positive and negative streamers in ambient air: measuring diameter, velocity and dissipated energy”. In: *Journal of Physics D: Applied Physics* 41.23 (2008), p. 234004.

[29] S Celestin. “Study of the dynamics of streamers in air at atmospheric pressure”. PhD thesis. Ecole Centrale Paris, 2008.

[30] C Chainais-Hillairet, JG Liu, and YJ Peng. “Finite volume scheme for multi-dimensional drift-diffusion equations and convergence analysis”. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 37.2 (2003), pp. 319–338.

[31] C Chainais-Hillairet and YJ Peng. “Convergence of a finite-volume scheme for the drift–diffusion equations in 1D”. In: *IMA journal of numerical analysis* 23.1 (2003), pp. 81–108.

[32] C Chainais-Hillairet and YJ Peng. “Finite volume approximation for degenerate drift-diffusion system in several space dimensions”. In: *Mathematical Models and Methods in Applied Sciences* 14.03 (2004), pp. 461–481.

[33] M Chatard. “Asymptotic behavior of the Scharfetter–Gummel scheme for the drift-diffusion model”. In: *Finite Volumes for Complex Applications VI Problems & Perspectives*. Springer, 2011, pp. 235–243.

[34] FF Chen et al. *Introduction to plasma physics and controlled fusion*. Vol. 1. Springer, 1984.

[35] I Choquet, P Degond, and B Lucquin-Desreux. “A hierarchy of diffusion models for partially ionized plasmas.” In: (2007).

[36] S Clain, S Diot, and R Loubère. “A high-order finite volume method for systems of conservation laws—Multi-dimensional Optimal Order Detection (MOOD)”. In: *Journal of computational Physics* 230.10 (2011), pp. 4028–4050.

[37] P Colella and PR Woodward. “The piecewise parabolic method (PPM) for gas-dynamical simulations”. In: *Journal of computational physics* 54.1 (1984), pp. 174–201.

[38] T Corke, B Mertz, and M Patel. “Plasma flow control optimized airfoil”. In: *44th AIAA Aerospace Sciences Meeting and Exhibit*. 2006, p. 1208.

[39] TC Corke, CL Enloe, and SP Wilkinson. “Dielectric barrier discharge plasma actuators for flow control”. In: *Annual review of fluid mechanics* 42 (2010), pp. 505–529.

[40] A Cristofolini, CA Borghi, and G Neretti. “Charge distribution on the surface of a dielectric barrier discharge actuator for the fluid-dynamic control”. In: *Journal of Applied Physics* 113.14 (2013), p. 143307.

[41] JA Cross and GN Haddad. “Negative point-plane corona in oxygen”. In: *Journal of Physics D: Applied Physics* 19.6 (1986), p. 1007.

[42] A Debien et al. “Unsteady aspect of the electrohydrodynamic force produced by surface dielectric barrier discharge actuators”. In: *Applied Physics Letters* 100.1 (2012), p. 013901.

[43] P Degond and B Lucquin-Desreux. “The asymptotics of collision operators for two species of particles of disparate masses”. In: *Mathematical Models and Methods in Applied Sciences* 6.03 (1996), pp. 405–436.

[44] P Degond and B Lucquin-Desreux. “Mathematical models of electrical discharges in air at atmospheric pressure: a derivation from asymptotic analysis”. In: *International Journal of Computing Science and Mathematics* 1.1 (2007), pp. 58–97.

[45] AA Dubinova. “Modeling of streamer discharges near dielectrics”. In: (2016).

[46] G Dufour and F Rogier. “Numerical modeling of dielectric barrier discharge based plasma actuators for flow control: the COPAIER/CEDRE example”. In: *Aerospace Lab* 10 (2015).

[47] J Eckstein and DP Bertsekas. “On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators”. In: *Mathematical Programming* 55.1 (1992), pp. 293–318.

[48] O Eichwald et al. “Effects of numerical and physical anisotropic diffusion on branching phenomena of negative-streamer dynamics”. In: *Journal of Physics D: Applied Physics* 45.38 (2012), p. 385203.

[49] CL Enloe, MG McHarg, and ThE McLaughlin. "Time-correlated force production measurements of the dielectric barrier discharge plasma aerodynamic actuator". In: *Journal of applied physics* 103.7 (2008), p. 073302.

[50] CL Enloe et al. "Mechanisms and responses of a dielectric barrier plasma actuator: Geometric effects". In: *AIAA journal* 42.3 (2004), pp. 595–604.

[51] CL Enloe et al. "Mechanisms and responses of a single dielectric barrier plasma actuator: plasma morphology". In: *AIAA journal* 42.3 (2004), pp. 589–594.

[52] R Eymard, J Fuhrmann, and K Gärtner. "A finite volume scheme for nonlinear parabolic equations derived from one-dimensional local Dirichlet problems". In: *Numerische Mathematik* 102.3 (2006), pp. 463–495.

[53] R Eymard, Th Gallouët, and R Herbin. "Finite volume methods". In: *Handbook of numerical analysis* 7 (2000), pp. 713–1018.

[54] NGC Ferreira et al. "Simulation of pre-breakdown discharges in high-pressure air. I: The model and its application to corona inception". In: *Journal of Physics D: Applied Physics* 52.35 (2019), p. 355206.

[55] K Fujii. "Three flow features behind the flow control authority of DBD plasma actuator: Result of high-fidelity simulations and the related experiments". In: *Applied Sciences* 8.4 (2018), p. 546.

[56] DN de G. Allen and RV Southwell. "Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder". In: *The Quarterly Journal of Mechanics and Applied Mathematics* 8.2 (1955), pp. 129–145.

[57] EC Gartland. "Uniform high-order difference schemes for a singularly perturbed two-point boundary value problem". In: *Mathematics of computation* 48.178 (1987), pp. 551–564.

[58] C Geuzaine and JF Remacle. "Gmsh: A 3-D finite element mesh generator with built-in pre-and post-processing facilities". In: *International journal for numerical methods in engineering* 79.11 (2009), pp. 1309–1331.

[59] VV Gorin et al. "Boundary conditions for drift-diffusion equations in gas-discharge plasmas". In: *Physics of Plasmas* 27.1 (2020), p. 013505.

[60] S Gottlieb, CW Shu, and E Tadmor. "Strong stability-preserving high-order time discretization methods". In: *SIAM review* 43.1 (2001), pp. 89–112.

[61] CJ Greenshields and HG Weller. "Notes on computational fluid dynamics: General principles". In: (2022).

[62] Y Guan et al. "Experimental and numerical investigation of electrohydrodynamic flow in a point-to-ring corona discharge". In: *Physical Review Fluids* 3.4 (2018), p. 043701.

[63] GJM Hagelaar. *Modeling of microdischarges for display technology*. Vol. 109. Technische Universiteit Eindhoven Eindhoven, 2000.

[64] GJM Hagelaar. "Modelling methods for low-temperature plasmas". PhD thesis. Université Toulouse III Paul Sabatier (UT3 Paul Sabatier), 2008.

[65] GJM Hagelaar, FJ De Hoog, and GMW Kroesen. "Boundary conditions in fluid models of gas discharges". In: *Physical Review E* 62.1 (2000), p. 1452.

[66] GJM Hagelaar and GMW Kroesen. "Speeding up fluid models for gas discharges by implicit treatment of the electron energy source term". In: *Journal of Computational Physics* 159.1 (2000), pp. 1–12.

[67] GJM Hagelaar and LC Pitchford. "Solving the Boltzmann equation to obtain electron transport coefficients and rate coefficients for fluid models". In: *Plasma sources science and technology* 14.4 (2005), p. 722.

[68] F Haider, JP Croisille, and B Courbet. "Stability analysis of the cell centered finite-volume m uscl method on unstructured grids". In: *Numerische Mathematik* 113 (2009), pp. 555–600.

[69] F Haider et al. "High order approximation on unstructured grids: Theory and implementation". In: *Preprint* (2012).

[70] A Harten. "High resolution schemes for hyperbolic conservation laws". In: *Journal of computational physics* 135.2 (1997), pp. 260–278.

[71] A Harten and S Osher. "Uniformly High-Order Accurate Nonoscillatory Schemes. I". In: *SIAM Journal on Numerical Analysis* (1987), pp. 279–309.

[72] A Harten et al. "Uniformly high order accurate essentially non-oscillatory schemes, III". In: *Upwind and high-resolution schemes*. Springer, 1987, pp. 218–290.

[73] D Hong et al. "Measurement of the surface charging of a plasma actuator using surface DBD". In: *Journal of Electrostatics* 71.3 (2013), pp. 547–550.

[74] S Ichimaru. *Statistical plasma physics, volume I: basic principles*. CRC Press, 2018.

[75] AM Il'in. "Differencing scheme for a differential equation with a small parameter affecting the highest derivative". In: *Mathematical Notes of the Academy of Sciences of the USSR* 6.2 (1969), pp. 596–602.

[76] K Ito and K Kunisch. "Parabolic variational inequalities: The Lagrange multiplier approach". In: *Journal de mathématiques pures et appliquées* 85.3 (2006), pp. 415–449.

[77] A Jameson. "Time dependent calculations using multigrid, with applications to unsteady flows past airfoils and wings". In: *10th Computational fluid dynamics conference*. 1991, p. 1596.

[78] B Koren. *A robust upwind discretization method for advection, diffusion and source terms*. Vol. 45. Centrum voor Wiskunde en Informatica Amsterdam, 1993.

[79] IA Kossyi et al. "Kinetic scheme of the non-equilibrium discharge in nitrogen-oxygen mixtures". In: *Plasma Sources Science and Technology* 1.3 (1992), p. 207.

[80] K Kourtzanidis. "Modélisation numérique d'actionneurs plasma pour le contrôle d'écoulement". PhD thesis. Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), 2014.

[81] K Kourtzanidis, G Dufour, and F Rogier. "Self-consistent modeling of a surface AC dielectric barrier discharge actuator: in-depth analysis of positive and negative phases". In: *Journal of Physics D: Applied Physics* 54.4 (2020), p. 045203.

[82] K Kourtzanidis, G Dufour, and F Rogier. "The electrohydrodynamic force distribution in surface AC dielectric barrier discharge actuators: do streamers dictate the ionic wind profiles?" In: *Journal of Physics D: Applied Physics* 54.26 (2021), 26LT01.

[83] AA Kulikovsky. “A more accurate Scharfetter-Gummel algorithm of electron transport for semiconductor and gas discharge simulation”. In: *Journal of computational physics* 119.1 (1995), pp. 149–155.

[84] AA Kulikovsky. “Two-dimensional simulation of the positive streamer in N2 between parallel-plate electrodes”. In: *Journal of Physics D: Applied Physics* 28.12 (1995), p. 2483.

[85] AA Kulikovsky. “Positive streamer between parallel plate electrodes in atmospheric pressure air”. In: *Journal of physics D: Applied physics* 30.3 (1997), p. 441.

[86] AA Kulikovsky. “Positive streamer in a weak field in air: A moving avalanche-to-streamer transition”. In: *Physical Review E* 57.6 (1998), p. 7066.

[87] Y Lagmich et al. “Model description of surface dielectric barrier discharges for flow control”. In: *Journal of Physics D: Applied Physics* 41.9 (2008), p. 095205.

[88] I Langmuir. “The interaction of electron and positive ion space charges in cathode sheaths”. In: *Physical review* 33.6 (1929), p. 954.

[89] DS Lee et al. “The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018”. In: *Atmospheric Environment* 244 (2021), p. 117834.

[90] RJ LeVeque. *Numerical methods for conservation laws*. Vol. 132. Springer, 1992.

[91] RJ LeVeque. *Finite volume methods for hyperbolic problems*. Vol. 31. Cambridge university press, 2002.

[92] SZ Li and HS Uhm. “Investigation of electrical breakdown characteristics in the electrodes of cylindrical geometry”. In: *Physics of Plasmas* 11.6 (2004), pp. 3088–3095.

[93] JL Lions. *Quelques méthodes de résolution des problemes aux limites non linéaires*. Dunod, 1969.

[94] PL Lions and B Mercier. “Splitting algorithms for the sum of two nonlinear operators”. In: *SIAM Journal on Numerical Analysis* 16.6 (1979), pp. 964–979.

[95] VA Lisovskiy et al. “Validating the Goldstein–Wehner law for the stratified positive column of dc discharge in an undergraduate laboratory”. In: *European journal of physics* 33.6 (2012), p. 1537.

[96] L Liu et al. “The complete flux scheme—error analysis and application to plasma simulation”. In: *Journal of Computational and Applied Mathematics* 250 (2013), pp. 229–243.

[97] N Liu and VP Pasko. “Effects of photoionization on propagation and branching of positive and negative streamers in sprites”. In: *Journal of Geophysical Research: Space Physics* 109.A4 (2004).

[98] LB Loeb. “Ionizing waves of potential gradient: luminous pulses in electrical breakdown, with velocities a third that of light, have a common basis”. In: *Science* 148.3676 (1965), pp. 1417–1426.

[99] A Loehrmann et al. “Numerical and Experimental correlation on Electro-Hydro Dynamic system for Small Aircraft propulsion”. In: *2021 AIAA/IEEE Electric Aircraft Technologies Symposium (EATS)*. IEEE. 2021, pp. 1–12.

[100] A Luque et al. “Photoionization in negative streamers: Fast computations and two propagation modes”. In: *Applied physics letters* 90.8 (2007).

[101] PA Markowich et al. “A singular perturbation approach for the analysis of the fundamental semiconductor equations”. In: *IEEE Transactions on Electron Devices* 30.9 (1983), pp. 1165–1180.

[102] R Marskar. “3D fluid modeling of positive streamer discharges in air with stochastic photoionization”. In: *Plasma Sources Science and Technology* 29.5 (2020), p. 055007.

[103] AA Martins. “Simulation of a wire-cylinder-plate positive corona discharge in nitrogen gas at atmospheric pressure”. In: *Physics of Plasmas* 19.6 (2012), p. 063502.

[104] OA Marzouk and ED Huckaby. “Simulation of a Swirling Gas-Particle Flow Using Different k-epsilon Models and Particle-Parcel Relationships.” In: *Engineering Letters* 18.1 (2010).

[105] JC Matéo-Vélez. “Modélisation et simulation numérique de la génération de plasma dans les décharges couronnes et de son interaction avec l’aérodynamique”. PhD thesis. École nationale supérieure de l’aéronautique et de l’espace (Toulouse; 1972 …, 2006.

[106] JC Matéo-Vélez et al. “Modelling wire-to-wire corona discharge action on aerodynamics and comparison with experiment”. In: *Journal of Physics D: Applied Physics* 41.3 (2008), p. 035205.

[107] MA Medvedev et al. “Internal microstructure of current channel of the long spark discharge”. In: *Bulletin of the Lebedev Physics Institute* 48.12 (2021), pp. 373–377.

[108] M Moisan and J Pelletier. “Physique des plasmas collisionnels”. In: *Physique des plasmas collisionnels*. EDP sciences, 2021.

[109] E Moreau. “Airflow control by non-thermal plasma actuators”. In: *Journal of physics D: applied physics* 40.3 (2007), p. 605.

[110] R Morrow. “Theory of negative corona in oxygen”. In: *Physical Review A* 32.3 (1985), p. 1799.

[111] R Morrow. “The theory of positive glow corona”. In: *Journal of Physics D: Applied Physics* 30.22 (1997), p. 3099.

[112] GV Naidis. “On photoionization produced by discharges in air”. In: *Plasma Sources Science and Technology* 15.2 (2006), p. 253.

[113] V Namias. “Application of the Dirac delta function to electric charge and multipole distributions”. In: *American Journal of Physics* 45.7 (1977), pp. 624–630.

[114] S Nijdam, J Teunissen, and U Ebert. “The physics of streamer discharge phenomena”. In: *Plasma Sources Science and Technology* 29.10 (2020), p. 103001.

[115] C Ollivier-Gooch and M Van Altena. “A high-order-accurate unstructured mesh finite-volume scheme for the advection–diffusion equation”. In: *Journal of Computational Physics* 181.2 (2002), pp. 729–752.

[116] DF Opaits et al. “Surface charge in dielectric barrier discharge plasma actuators”. In: *Physics of Plasmas* 15.7 (2008), p. 073505.

[117] S Pancheshnyi. “Role of electronegative gas admixtures in streamer start, propagation and branching phenomena”. In: *Plasma Sources Science and Technology* 14.4 (2005), p. 645.

[118] S Pancheshnyi. “Photoionization produced by low-current discharges in O2, air, N2 and CO2”. In: *Plasma Sources Science and Technology* 24.1 (2014), p. 015023.

[119] B Parent, MN Shneider, and SO Macheret. “Sheath governing equations in computational weakly-ionized plasmadynamics”. In: *Journal of Computational Physics* 232.1 (2013), pp. 234–251.

[120] N Parikh, S Boyd, et al. “Proximal algorithms”. In: *Foundations and trends® in Optimization* 1.3 (2014), pp. 127–239.

[121] VP Pasko. “Theoretical modeling of sprites and jets”. In: *Sprites, elves and intense lightning discharges* 225 (2006), pp. 253–311.

[122] M Patriarca et al. “Highly accurate quadrature-based Scharfetter–Gummel schemes for charge transport in degenerate semiconductors”. In: *Computer Physics Communications* 235 (2019), pp. 40–49.

[123] CO Porter et al. “Plasma actuator force measurements”. In: *AIAA journal* 45.7 (2007), pp. 1562–1570.

[124] J Qu et al. “A review on recent advances and challenges of ionic wind produced by corona discharges with practical applications”. In: *Journal of Physics D: Applied Physics* 55.15 (2021), p. 153002.

[125] A Quarteroni, R Sacco, and F Saleri. *Numerical mathematics*. Vol. 37. Springer Science & Business Media, 2010.

[126] YP Raizer and JE Allen. *Gas discharge physics*. Vol. 1. Springer, 1991.

[127] HG Roos, M Stynes, and L Tobiska. *Robust numerical methods for singularly perturbed differential equations: convection-diffusion-reaction and flow problems*. Vol. 24. Springer Science & Business Media, 2008.

[128] JR Roth, DM Sherman, and SP Wilkinson. “Electrohydrodynamic flow control with a glow-discharge surface plasma”. In: *AIAA journal* 38.7 (2000), pp. 1166–1172.

[129] N Sato. “Discharge current induced by the motion of charged particles”. In: *Journal of Physics D: Applied Physics* 13.1 (1980), p. L3.

[130] DL Scharfetter and HK Gummel. “Large-signal analysis of a silicon read diode oscillator”. In: *IEEE Transactions on electron devices* 16.1 (1969), pp. 64–77.

[131] P Ségur et al. “The use of an improved Eddington approximation to facilitate the calculation of photoionization in streamer discharges”. In: *Plasma Sources Science and Technology* 15.4 (2006), p. 648.

[132] P Seimandi, G Dufour, and F Rogier. “An asymptotic model for steady wire-to-wire corona discharges”. In: *Mathematical and computer modelling* 50.3-4 (2009), pp. 574–583.

[133] TH Shih et al. *A new k-epsilon eddy viscosity model for high Reynolds number turbulent flows: Model development and validation*. Tech. rep. 1994.

[134] CW Shu. “Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws”. In: *Advanced numerical approximation of nonlinear hyperbolic equations*. Springer, 1998, pp. 325–432.

[135] VR Soloviev and VM Krivtsov. “Surface barrier discharge modelling for aerodynamic applications”. In: *Journal of Physics D: Applied Physics* 42.12 (2009), p. 125208.

[136] G Strang. “On the construction and comparison of difference schemes”. In: *SIAM journal on numerical analysis* 5.3 (1968), pp. 506–517.

[137] PK Sweby. “High resolution schemes using flux limiters for hyperbolic conservation laws”. In: *SIAM journal on numerical analysis* 21.5 (1984), pp. 995–1011.

[138] J Teunissen. “Improvements for drift-diffusion plasma fluid models with explicit time integration”. In: *Plasma Sources Science and Technology* 29.1 (2020), p. 015010.

[139] J Teunissen and U Ebert. “3D PIC-MCC simulations of discharge inception around a sharp anode in nitrogen/oxygen mixtures”. In: *Plasma Sources Science and Technology* 25.4 (2016), p. 044005.

[140] J Teunissen and U Ebert. “Simulating streamer discharges in 3D with the parallel adaptive Afivo framework”. In: *Journal of Physics D: Applied Physics* 50.47 (2017), p. 474001.

[141] J Teunissen and U Ebert. “Afivo: A framework for quadtree/octree AMR with shared-memory parallelization and geometric multigrid methods”. In: *Computer Physics Communications* 233 (2018), pp. 156–166.

[142] JHM ten Thije Boonkkamp and MJH Anthonissen. “The finite volume-complete flux scheme for advection-diffusion-reaction equations”. In: *Journal of Scientific Computing* 46.1 (2011), pp. 47–70.

[143] P Tsoutsanis. “Stencil selection algorithms for WENO schemes on unstructured meshes”. In: *Journal of Computational Physics: X* 4 (2019), p. 100037.

[144] N Tuan Dung, C Besse, and F Rogier. “High-order Scharfetter-Gummel-based schemes and applications to gas discharge modeling”. In: *Journal of Computational Physics* 461 (2022), p. 111196.

[145] N Tuan Dung, C Besse, and F Rogier. “An implicit time integration approach for simulation of corona discharges”. In: *Computer Physics Communications* 294 (2024), p. 108906.

[146] Th Unfer and JP Boeuf. “Modelling of a nanosecond surface discharge actuator”. In: *Journal of physics D: applied physics* 42.19 (2009), p. 194017.

[147] Th Unfer et al. “An asynchronous scheme with local time stepping for multi-scale transport problems: Application to gas discharges”. In: *Journal of Computational Physics* 227.2 (2007), pp. 898–918.

[148] B Van Leer. “Towards the ultimate conservative difference scheme I. The quest of monotonicity”. In: *Proceedings of the Third International Conference on Numerical Methods in Fluid Mechanics*. Springer. 1973, pp. 163–168.

[149] B Van Leer. “Towards the ultimate conservative difference scheme. II. Monotonicity and conservation combined in a second-order scheme”. In: *Journal of computational physics* 14.4 (1974), pp. 361–370.

[150] B Van Leer. “Towards the ultimate conservative difference scheme III. Upstream-centered finite-difference schemes for ideal compressible flow”. In: *Journal of Computational Physics* 23.3 (1977), pp. 263–275.

[151] B Van Leer. “Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection”. In: *Journal of computational physics* 23.3 (1977), pp. 276–299.

[152] B Van Leer. “Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov’s method”. In: *Journal of computational Physics* 32.1 (1979), pp. 101–136.

[153] AGR Vasconcelos, DMS Albuquerque, and JCF Pereira. “A very high-order finite volume method based on weighted least squares for elliptic operators on polyhedral unstructured grids”. In: *Computers & Fluids* 181 (2019), pp. 383–402.

[154] J Velechovsky, R Liska, and M Shashkov. “High-order remapping with piece-wise parabolic reconstruction”. In: *Computers & Fluids* 83 (2013), pp. 164–169.

[155] PLG Ventzek et al. “Two-dimensional hybrid model of inductively coupled plasma sources for etching”. In: *Applied physics letters* 63.5 (1993), pp. 605–607.

[156] A Villa et al. “An asymptotic preserving scheme for the streamer simulation”. In: *Journal of Computational Physics* 242 (2013), pp. 86–102.

[157] A Villa et al. “Mesh dependent stability of discretization of the streamer equations for very high electric fields”. In: *Computers & Fluids* 105 (2014), pp. 1–7.

[158] A Von Engle, R Seeliger, and M Steenback. “On the glow discharge at high pressure”. In: *Zeit. fur Physik* 85.144 (1933), pp. 144–160.

[159] S Wang et al. “Numerical and experimental study of heat-transfer characteristics of needle-to-ring-type ionic wind generator for heated-plate cooling”. In: *International Journal of Thermal Sciences* 139 (2019), pp. 176–185.

[160] HG Weller et al. “A tensorial approach to computational continuum mechanics using object-oriented techniques”. In: *Computers in physics* 12.6 (1998), pp. 620–631.

[161] H Xu et al. “Flight of an aeroplane with solid-state propulsion”. In: *Nature* 563.7732 (2018), pp. 532–535.

[162] M Yang and ZJ Wang. “A parameter-free generalized moment limiter for high-order methods on unstructured grids”. In: *47th AIAA Aerospace Sciences Meeting Including The New Horizons Forum and Aerospace Exposition*. 2009, p. 605.

[163] Y Zhang et al. “Characteristics of ionic wind in needle-to-ring corona discharge”. In: *Journal of Electrostatics* 74 (2015), pp. 15–20.

[164] Y Zhang et al. “Trichel pulse in various gases and the key factor for its formation”. In: *Scientific reports* 7.1 (2017), p. 10135.

[165] L Zhao and K Adamiak. “EHD flow in air produced by electric corona discharge in pin–plate configuration”. In: *Journal of electrostatics* 63.3-4 (2005), pp. 337–350.

[166] MB Zhelezniak, AKh Mnatsakanian, and SV Sizykh. “Photoionization of nitrogen and oxygen mixtures by radiation from a gas discharge”. In: *High Temperature Science* 20.3 (1982), pp. 357–362.

[167] Y Zhu et al. “Nanosecond-pulsed dielectric barrier discharge-based plasma-assisted anti-icing: modeling and mechanism analysis”. In: *Journal of Physics D: Applied Physics* 53.14 (2020), p. 145205.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Université de Toulouse

Titre : Développement d'une méthode numérique multiéchelle pour les plasmas atmosphériques et application au contrôle d'écoulements

Mots clés : méthode multiéchelle, schéma implicite, méthode de volumes finis, plasma non-équilibre, décharge électrique, modèle hydrodynamique

Résumé : Les récents événements météorologiques extrêmes qui se sont produits dans le monde entier récemment suscitent de grandes inquiétudes concernant le changement climatique et renforcent la nécessité de développer des technologies écologiques permettant d'enrayer la tendance croissante à la destruction de l'écosystème. Selon une étude publiée en 2021, l'industrie aéronautique est responsable de 3.5% des facteurs de changement climatique liés aux activités humaines entre 2010 et 2018, dont les émissions de dioxyde de carbone et d'oxydes d'azote. Les études menées au cours des deux dernières décennies ont montré que les dispositifs à décharge électrique appelés actionneurs de plasma sont capables de contrôler l'écoulement autour d'un profil aérodynamique et de réduire la traînée sur sa surface, ce qui est prometteur pour réduire la consommation de carburant des aéronefs. Cette thèse contribue à la modélisation numérique des décharges électriques dans l'air qui n'a jamais été une tâche facile en raison de sa nature multi-échelle à la fois en espace et en temps. Nos travaux se focalisent sur deux axes : l'amélioration de la qualité de précision des solutions numériques et la réduction du temps CPU des simulations. Pour le premier objectif, nous avons concevons des schémas d'ordre élevé connus collectivement sous le nom de méthodes de Scharfetter-Gummel avec correction de courant (SGCC) pour résoudre les équations de dérive-diffusion qui apparaissent dans le modèle mathématique de la décharge. Les schémas SGCC sont une généralisation du schéma standard Scharfetter-Gummel (SG) qui est largement utilisé dans la simulation des plasmas. Nous allons montrer, à l'aide de nombreux cas tests, qu'ils fournissent des solutions numériques beaucoup plus précises que celles du schéma SG standard pour les mêmes conditions de simulation. Pour le deuxième objectif, nous développons une stratégie d'intégration temporelle implicite pour la simulation de la décharge couronne qui est basée sur une nouvelle formulation du modèle de plasma. Ce travail se concentre sur la réduction du temps CPU tout en garantissant que les résultats numériques sont cohérents avec les données expérimentales disponibles dans la littérature, en imposant une contrainte minimale sur la densité d'électrons. La méthode implicite proposée a permis de réduire le temps de calcul à quelques heures. Au contraire, le temps de calcul des méthodes explicites utilisées précédemment peut atteindre plusieurs semaines. En conséquence, nous bénéficions d'une amélioration significative de la performance du solveur de plasma, ce qui pourrait potentiellement ouvrir la porte à des simulations de plasma plus réalistes dans l'avenir.

Title: Development of a multiscale numerical method for gas discharge in air and application to flow control

Key words: multiscale method, implicit scheme, finite-volume method, non-equilibrium plasma, gas discharge, hydrodynamic model

Abstract: Recent extreme weather events that have been occurring more frequently around the globe have raised further concerns about climate change and fuels the need of developing environment-friendly technologies that could halt the increasing trend of ecosystem destruction. According to a study in 2021, the aviation industry is responsible for 3.5% of all drivers of climate change from human activities from 2010 to 2018, including emissions of harmful gases such as carbon dioxide and nitrogen oxides. Studies carried out over the last two decades have shown that electric discharge devices called plasma actuators are capable of controling the flow around an airfoil and reducing the drag force on its surface, thus are promising to decrease the fuel consumption of aircrafts. This thesis contributes to the numerical modeling of electric discharge in air which has never been an easy task owing to its multiscale nature both in space and time. Our works focus on two axes : improvement of the precision quality of numerical solutions and reduction of the CPU time of simulations. For the first objective, we conceive a set of high-order flux schemes collectively known as the Scharfetter-Gummel methods with correction of current (SGCC) to solve the drift-diffusion equations that appear in the discharge mathematical model. The SGCC schemes are a generalization of the standard Scharfetter-Gummel (SG) scheme that is widely used in plasma simulation. It will be shown, through numerous test cases, that they provide much more accurate numerical solutions than those of the standard SG scheme for the same simulation conditions. For the second objective, we develop an implicit time integration strategy for simulation of corona discharge that is based on a new formulation of the plasma model. This work focuses on shortening the CPU time while simultaneously ensuring that the numerical results are coherent with experiment data available in the literature, by imposing a minimum constraint on the electron density. The CPU time was successfully reduced to a matter of hours with the proposed implicit method. On the contrary, the CPU time of the previously used explicit methods can be weeks long. As a result, we obtain a significant enhancement in the performance of the plasma solver, which could potentially open the door to more realistic plasma simulations in the future.