THE SCHARNHORST EFFECT: SUPERLUMINALITY AND CAUSALITY IN EFFECTIVE FIELD THEORIES

by

Sybil Gertrude de Clark

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PHYSICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2016

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Sybil Gertrude de Clark, titled The Scharnhorst Effect: Superluminality and Causality in Effective Field Theories, and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Date: 8 November 2016
Sean Fleming

Date: 8 November 2016
Christian Wüthrich

Date: 8 November 2016
Bruce Barrett

Date: 8 November 2016
Dimitrios Psaltis

Date: 8 November 2016
William Toussaint

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.
I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date: 8 November 2016
Dissertation Director: Sean Fleming

Date: 8 November 2016
Dissertation Director: Christian Wüthrich

STATEMENT BY AUTHOR

SIGNED: Sybil Gertrude de Clark

ACKNOWLEDGEMENTS

Firstly, I want to express my sincere gratitude to my advisors. I want to thank my physics advisor, Sean Fleming, for his patient support through all stages of my research. It should go without saying (but I shall say it anyway) that without his readiness to accommodate my passion for foundations of physics, his guidance in delineating research projects of relevance to the latter, and then his continuous assistance over the years to bring my research to fruition, this work would not exist. I am also immensely grateful to my philosophy advisor, Christian Wüthrich, for supervising my work and providing numerous detailed and thoughtful comments. His input and advice have been invaluable to me. I could not have wished for better mentors.

In addition to my advisors, I also want to thank the other members of my doctoral committee: Bruce Barrett, Dimitrios Psaltis and Doug Toussaint. I am of course, yet truly, grateful for their willingness to serve on my committee in the first place. Furthermore, their keen and insightful feedback during the meetings that took place in the course of my research has been extremely helpful, even crucial.

I am also very grateful for the opportunities I have been given to present my work at seminars while it was still in progress. In this respect I wish to thank James Weatherall as the organizer of the "Southern California Philosophy of Physics Reading Group" at UC Irvine, as well as Chris Wüthrich for introducing me to this wonderful group. I also want to thank Hasok Chang for the opportunity to present my work to the inter-disciplinary seminar "Coffee with Scientists" which he organizes in the Department of History and Philosophy of Science at the University of Cambridge, as well as Daniel Jon Mitchell without who I would never have had the opportunity to go. The questions and observations I received during these talks really helped me to broaden my ideas, and I am very grateful to those who provided me with feedback in this way. Furthermore, these talks have had an unanticipated and wonderful outcome: they motivated me to contact several of the physicists whose fascinating work it had been my pleasure to study.

In this respect, it is with the greatest pleasure that I thank those who I have had the privilege to correspond with, and often even to meet. Being also an historian of science (notably on Ptolemy's work and 19^{th} c physics), my previous research could hardly afford me the joy of discussing the science I was studying with the very people who had discovered it. Being able to do so now has been a truly special experience, and one of the most rewarding aspects of my work. I am especially grateful to Peter Milonni, Robert Geroch, John Cramer, Shahar Ben-Menahem, Sebastiano Sonego, as well as the two pioneers who unveiled and publicized the effect that is the topic of the present dissertation: Klaus Scharnhorst and Gabriel Barton. I have been

overwhelmed by how incredibly generous they have been with their time and help, and the discussions that I have had with them over the course of this work have been incredibly gratifying.

DEDICATION

For Michael Kruse, without whom I would never have had the opportunity to accomplish this work, and for Roger Haar, who I wish could have seen it.

Table of Contents

List of Figures 12

ABSTRACT 14

Chapter 1 INTRODUCTION 16
- 1.1 Introduction 16
  - 1.1.1 Description of the Scharnhorst effect 16
  - 1.1.2 Derivations of the Scharnhorst effect: a brief history 17
- 1.2 The issue of signal velocity 19
- 1.3 Conceptual interpretation of the Scharnhorst effect 22
  - 1.3.1 The Casimir vacuum as a dielectric medium 22
  - 1.3.2 The Casimir effect 23
  - 1.3.3 Analogous conceptual model for the Scharnhorst effect 25
  - 1.3.4 Alternative conceptual models for the Casimir and Scharnhorst effects 27
- 1.4 Overview of dissertation 28

Chapter 2 SURVEY OF SUPERLUMINAL EFFECTS IN FIELD THEORIES 33
- 2.1 Introduction 33
- 2.2 The Drummond-Hathrell effect 34
  - 2.2.1 Description 34
  - 2.2.2 Historical background 37
  - 2.2.3 Interpretation 40
- 2.3 Bimetric theories 41
  - 2.3.1 Modified Newtonian dynamics (MOND) theories 41
  - 2.3.2 Varying speed of light bimetric theories 45
- 2.4 k-essence 45
  - 2.4.1 Description 45
  - 2.4.2 k-essence and superluminal propagation 49
- 2.5 Does superluminal propagation imply causal paradoxes? 51
  - 2.5.1 Causal paradox: the bilking argument and its physical implementation 52
  - 2.5.2 Causal cone 54
  - 2.5.3 The tachyonic anti-telephone thought experiment modified 57
  - 2.5.4 Stable causality 59

Table of Contents – *Continued*

2.5.5 New difficulties: propagation into the coordinate past 61
2.5.6 Initial data formulation and Cauchy problem 68
2.5.7 Robert Geroch's work 72
2.5.8 Forming CSCs within a causal cone formalism 79
2.6 Conclusion 86

Chapter 3 THE SCHARNHORST EFFECT: A HISTORICAL INTRODUCTION 89
3.1 Introduction 89
3.2 The Scharnhorst effect 90
3.3 Early derivations of the Scharnhorst effect 90
3.3.1 Scharnhorst's derivation (1990) 91
3.3.2 Barton's derivation (1990) 94
3.4 Early reactions (1990) 97
3.4.1 Milonni & Svotzil's argument 98
3.4.2 Ben-Menahem's wavefront argument 100
3.4.3 Ben-Menahem's commutator argument 102
3.5 Scharnhorst and Barton's response: implications of the Kramers-Kronig relations 107
3.6 Later work 113
3.6.1 Relation between the Scharnhorst effect and the energy density of the Casimir vacuum 113
3.6.2 Effect at oblique incidence and effective metric approaches . . 117

Chapter 4 DERIVATION OF THE SCHARNHORST EFFECT IN SOFT-COLLINEAR EFFECTIVE THEORY 122
4.1 Introduction 122
4.2 Modelling the Scharnhorst effect in the framework of SCET 122
4.2.1 The standard Euler-Heisenberg Lagrangian 123
4.2.2 Degrees of freedom in the SCET Euler-Heisenberg Lagrangian 128
4.3 Advantages of SCET *vs* the standard approach 132
4.3.1 Light-cone coordinates 133
4.3.2 SCET: scaling in the expansion parameter λ 134
4.3.3 SCET perturbative expansion and control of the errors at higher orders 141
4.3.4 Interaction between Casimir background and probe photon . . 146
4.4 Derivation of the SCET Lagrangian 147
4.5 Derivation of the SCET polarization tensor $\Pi_{\mu\nu}^{2-loop}$ 151
4.5.1 Relevant terms to model the Scharnhorst effect in $\mathcal{L}_{EH}^{SC}$ theory 152

Table of Contents – *Continued*

- 4.5.2 Expressions for the propagators 157
- 4.5.3 Evaluating the contributions to the tadpole diagram and self-energy tensor 158
- 4.5.4 Resulting polarization tensor 170

4.6 Derivation of the velocity from the SCET polarization tensor using the light cone condition 176

4.7 Derivation of the velocity from the SCET polarization tensor using the index of refraction 179

4.8 Conclusion 182

Chapter 5 FLUCTUATIONS OF THE ELECTROMAGNETIC VACUUM FIELD OR RADIATION REACTION? 183

5.1 Introduction 183
- 5.1.1 The issues 183
- 5.1.2 Vacuum fluctuations, virtual particles and zero-point energies: what are they? 185
- 5.1.3 Source field and radiation reaction: what are they? 190

5.2 Physical effects deemed evidence for vacuum fluctuations 193
- 5.2.1 The Stark effect 194
- 5.2.2 The Lamb shift 195
- 5.2.3 Spontaneous emission 201
- 5.2.4 Van der Waals force: force between two atoms 206
- 5.2.5 Casimir-Polder force 214
- 5.2.6 The Casimir effect 217
- 5.2.7 Operator ordering and ambiguity in the physical interpretation 225

5.3 The fluctuation-dissipation theorem (FDT) 228
- 5.3.1 Description 228
- 5.3.2 Examples 230
- 5.3.3 Possible interpretations 233
- 5.3.4 Interpretations of the fluctuation dissipation theorem in the physics literature 239

5.4 Consistency of the quantum theory of radiation: necessity of the vacuum field for the commutation relations to hold. 249

5.5 Discussion 251
- 5.5.1 Operator ordering 252
- 5.5.2 Are we really dealing with the vacuum field? 255

5.6 Conclusions 257

Table of Contents – *Continued*

Chapter 6 THE SCHARNHORST EFFECT IN SOURCE THEORY 260
- 6.1 Introduction 260
- 6.2 General form of the derivation 262
  - 6.2.1 The Euler-Heisenberg effective Lagrangian 263
  - 6.2.2 Derivation of the relevant equation of motion for the fields . . 267
  - 6.2.3 Derivation of the polarization and magnetization of the background electromagnetic field 275
  - 6.2.4 Derivation of the shift in the index of refraction and phase velocity 278
- 6.3 $\langle E_i E_j \rangle$ and $\langle B_i B_j \rangle$ in terms of source fields: deriving expressions for the Green's functions $\Gamma_{ij}(\mathbf{r'}, \mathbf{r'}, \omega)$ 284
  - 6.3.1 Obtaining the differential equation satisfied by the Green's function 285
  - 6.3.2 Expressing $\overleftrightarrow{\mathbf{\Gamma}}$ in terms of the "transverse electric" and "transverse magnetic" Green's functions 288
  - 6.3.3 Determining the expressions for $g^E(z, z')$ and $g^H(z, z')$ 296
- 6.4 Deriving the expression for $\langle E_i E_j \rangle$ from the Green's functions $\Gamma_{ij}(\mathbf{r'}, \mathbf{r'}, \omega)$ 317
  - 6.4.1 Derivation of the expression for $\left\langle E_i^-(r,t) E_j^+(r,t) \right\rangle$ 317
- 6.5 Conclusion 323

Chapter 7 CONCLUSION 325
- 7.1 Reminder: Debate regarding the signal velocity in the physics literature 326
  - 7.1.1 Klaus Scharnhorst and Gabriel Barton's derivations 326
  - 7.1.2 Reactions in the physics literature 328
  - 7.1.3 Scharnhorst and Barton's response 336
- 7.2 Superluminal behavior and causality 341
  - 7.2.1 "Soft-breaking" of Lorentz invariance 341
  - 7.2.2 The Casimir vacuum as a dielectric medium 344
  - 7.2.3 Concerns regarding causal paradoxes 348
- 7.3 Implications of the SCET approach 351
  - 7.3.1 Control of theoretical errors: implication 351
  - 7.3.2 Implications of the SCET approach for the signal velocity . . 353

BIBLIOGRAPHY 356

Appendix A VACUUM FIELD AND UNCERTAINTY RELATIONS 363
- A.1 Preliminary: interpretations of the uncertainty relations (UR) in the context of non-relativistic quantum mechanics (NRQM) 363
- A.2 Uncertainty relations, vacuum fluctuations and virtual particles . . . 368

Table of Contents – *Continued*

Appendix B RADIATION REACTION AND SELF-FORCE 375
B.0.1 Retardation as the origin of the self-force 378

Appendix C UNDERDETERMINATION AND ORDERING OF OPERATORS 381
C.1 First case: Q given by PE 381
C.2 Second case (two-level atom): Q given by terms of the form $E^-P^- + E^+P^+$ 383

Appendix D DERIVATION OF THE VACUUM EXPECTATION VALUES $\langle E_iE_j\rangle$ AND $\langle B_iB_j\rangle$ BETWEEN CASIMIR PLATES 385
D.1 Derivation of the vector potential $\mathbf{A}(\mathbf{r}, t)$ 385
D.2 Derivation of the fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ from the vector potential $\mathbf{A}(\mathbf{r}, t)$ 391
D.3 Derivation of $\langle E_iE_j\rangle$ and $\langle B_iB_j\rangle$ from the fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ . . 396

Appendix E DERIVATION OF THE FUNDAMENTAL SOLUTION GREEN'S FUNCTION 409

List of Figures

Figure 2.1 The “tachyonic anti-telephone” thought experiment. 53
Figure 2.2 The causal cone *vs* the light cone 55
Figure 2.3 Causal regions defined with respect to the causal cone. 56
Figure 2.4 The “tachyonic anti-telephone” thought experiment modified: propagation is constrained to take place on causal cones. . . . 58
Figure 2.5 Signal propagating into B’s coordinate past. 61
Figure 2.6 causal cone seen from the frame of an emitter moving at high speed in the same direction as the signal. 62
Figure 2.7 A signal emitted to the left propagates into B’s coordinate past. 63
Figure 2.8 A signal arriving from the right emanates from B’s coordinate future. 63
Figure 2.9 A signal arriving from the right emanates from B’s coordinate future. 64
Figure 2.10 Well-posed *vs* ill-posed initial value formulation. 69
Figure 2.11 Well-posed *vs* ill-posed initial value formulation. 70
Figure 2.12 “The hatched part shows the extended future defined by two metrics (solid and dashed lines) in the case where one metric defines a wider cone than the other one (left) and vice versa (right).” 72
Figure 2.13 The 2nd signal reaches point C after the first left point A. This holds whether or not the frame of the observer is at rest or in motion with respect to the background. It remains true no matter how fast the frame moves and the cones “tilt”, because they do so to the same degree. 80
Figure 2.14 Cones configuration needed to form a CSC. 81
Figure 2.15 Two finite bubbles moving with large opposite velocities in the x-direction and separated by a finite distance in the y-direction. 82

Figure 3.1 The two-loop processes responsible for the Scharnhorst effect. 92
Figure 3.2 The one-loop process. It is not involved in the Scharnhorst effect 92
Figure 3.3 Wavefront (or line of constant phase) for a light signal emitted at the point P. 120

Figure 4.1 Vertex in QED 125

List of Figures – *Continued*

Figure 4.2 The box diagram in QED becomes a contact term in the theory described by $\mathcal{L}_{EH}$. 126
Figure 4.3 Some of the interactions affecting a propagating photon in a theory with only photon fields. 126
Figure 4.4 Feynman diagrams in the theory described by the standard Euler-Heisenberg Lagrangian. The diagram on the left represents the leading order, kinetic term, the one on the right accounts for interactions, and is known as the "tadpole diagram." 127
Figure 4.5 Degrees of freedom in the Scharnhorst effect modeled in SCET. 129
Figure 4.6 Feynman diagrams in the theory described by the SCET Euler-Heisenberg Lagrangian. 131
Figure 4.7 Feynman diagrams used to account for the Scharnhorst effect within the theory based on the SCET Euler-Heisenberg Lagrangian. 131
Figure 4.8 The fully interacting photon propagator in the standard $\mathcal{L}_{EH}$ theory. 152
Figure 4.9 The fully interacting photon propagator in the standard $\mathcal{L}_{EH}$ theory. 154
Figure 4.10 The fully interacting photon propagator in the SCET $\mathcal{L}_{EH}^{SC}$ theory. 154
Figure 4.11 The fully interacting propagator for the Scharnhorst photon in the $\mathcal{L}_{EH}^{SC}$ theory. 156
Figure 4.12 Relevant diagram to the Scharnhorst effect in the $\mathcal{L}_{EH}^{SC}$ theory. 156

Figure 6.1 The two-loop processes responsible for the Scharnhorst effect. 260
Figure 6.2 The box diagram in QED becomes a contact term in the theory described by $\mathcal{L}_{EH}$. 261
Figure 6.3 Feynman diagrams in the theory described by the standard Euler-Heisenberg Lagrangian. The diagram on the right accounts for interactions, and is known as the "tadpole diagram." 261

Figure 7.1 The wavefront is "smeared" over a distance of $1/m_e$, but has moved measurably "beyond the light cone." 331
Figure 7.2 The "smearing" of the wavefront precludes measuring the signal speed as superluminal over the plates separation. 332
Figure 7.3 The cones themselves are blurred. 335
Figure 7.4 The two-loop processes responsible for the Scharnhorst effect. 346
Figure 7.5 Tadpole diagram responsible for the Scharnhorst effect in SCET. 346

ABSTRACT

We present two re-derivations of the Scharnhorst effect. The latter was first obtained in 1990 by Klaus Scharnhorst, soon followed by Gabriel Barton, and consists in the theoretical prediction that the phase velocity of photons propagating in a Casimir vacuum normal to the plates would be larger than c.

The first derivation given in the present work is relevant for the debates that have taken place in the physics literature regarding a possible greater-than-c value of the signal velocity. Indeed because the phase velocity result also held for the group velocity, the issue soon arose as to whether the same could be said for the signal velocity. Several arguments were presented against this notion, notably to the effect that measurement uncertainties would preclude such a measurement. These notably relied on the fact that the known phase velocity result is only valid within a certain frequency regime. Scharnhorst and Barton responded by arguing that given their previous result, the Kramers-Kronig relations imply one of two options: either the greater-than-c result holds for the signal velocity as well, or the Casimir vacuum behaves like an amplifying medium for some frequencies. Furthermore, the effect was later rederived and generalized within the framework of an effective metric approach, which has been argued to obviate the worries regarding causal paradoxes often associated with the possibility of faster-than-c signalling. However concerns related to theory errors as well as to the measurement uncertainties that had surfaced in the earlier debate have remained salient. By re-deriving the phase velocity using Soft-Collinear Effective Theory (SCET), one can address some of these concerns. Indeed, with regard to theory errors, SCET provides us with a framework where higher order corrections are known to be power-suppressed because SCET ensures that the expansion parameters are multiplied by factors of order 1. As a result, with due qualifications inherent to the nature of effective field theory, the

result obtained within the SCET approach cannot be invalidated by higher order corrections. Furthermore, the theoretical description offered by SCET provides an argument relevant to the point that measurement uncertainties would prevent measuring the signal speed to be faster-than-c. Indeed, SCET implies the interaction between the Casimir vacuum and the propagating photon to be such that the latter would have the same phase velocity irrespective of its frequency. This in turn would entail that its signal velocity would be equal to this phase velocity, which is faster-than-c.

The second calculation presented is concerned with the physical interpretation of the Scharnhorst effect, and constitutes an attempt at re-deriving it within source theory. Existing derivations imply that the Scharnhorst effect can be attributed to vacuum fluctuations. Other physical effects that share this feature have also been derived without any reference to the vacuum, but as due to source fields instead. We attempt a similar derivation for the Scharnhorst effect.

Chapter 1

INTRODUCTION

1.1 Introduction

The present work focuses on a specific instance of predicted superluminal velocity: the "Scharnhorst effect."[1] It discusses two alternative ways to rederive this velocity.[2] The first involves using the framework of Soft Collinear Effective Field Theory ("SCET"). In contrast to the standard approach used in previous calculations, with SCET higher-order corrections are well-controlled, as higher-order terms are power-suppressed. The second derivation invokes source theory. Its significance lies in its implications for the interpretation of the Scharnhorst effect.

1.1.1 Description of the Scharnhorst effect

The "Scharnhorst effect" refers to the predicted "superluminal" motion of photons in a Casimir vacuum.[3] The set-up in which the phenomenon would occur is similar

---

[1] As explained below the term "effect" is used in the sense that the phenomenon in question has been predicted from theory, but it has not (yet) been observed.

[2] What is directly calculated is the phase velocity. This represents the rate at which the phase of a specific mode (i.e. characterized by a single frequency) propagates through space. For instance, in a waveform this is the velocity of a given peak. It is defined as $v_\varphi \equiv \frac{\omega}{k}$, where ω is the frequency of the mode and k its wave number. It is also given by $v_\varphi = \frac{c}{n}$, where c is the speed of light in the usual unbounded vacuum – the so-called "trivial vacuum"–, and n is the index of refraction of the medium in which light is travelling. The group velocity is that of the envelope, and is given by $v_g \equiv \frac{d\,\omega}{d\,k}$. The velocity that has come to be considered the relevant one as far as the speed of light postulate is concerned is the signal velocity. It is defined by the motion of the wavefront. For instance in a medium, it is the speed of the front of the disturbance.

[3] By "superluminal" is meant a velocity faster than the speed of light in the usual unbounded, trivial vacuum. Although it sounds like a contradiction in terms when applied to photons themselves, the use of the term has become standard in the context of the Scharnhorst effect.

to that of the Casimir effect: two parallel, conducting but electrically neutral plates, separated by a vacuum (the Casimir effect consists in an attractive force between these plates).[4] Despite what the term "effect" may suggest, and in contrast to its Casimir counterpart, the Scharnhorst effect has not been observed: at present, it is purely a theoretical prediction. In fact, experimental confirmation is well beyond current technology. For any reasonable plate separation the predicted shift in velocity is extremely small: a correction $\Delta c/c$ of the order of 10^{-32} for a separation of a micron, which is approximately that used in experimental measurements of the Casimir effect.[5]

1.1.2 Derivations of the Scharnhorst effect: a brief history

Klaus Scharnhorst's derivation

The Scharnhorst effect was first derived in February 1990 by Klaus Scharnhorst[5]; this was followed a month later by a second derivation due to Gabriel Barton [6].[6] Both predicted that photons propagating between Casimir plates would travel faster than they do in normal, trivial vacuum for most directions, and derived the magnitude of the velocity shift for photons traveling perpendicular to the plates (the fastest). Photons propagating parallel to the plates remained unaffected by the presence of the latter and travelled at c. Like the Casimir effect, the Scharnhorst effect would decrease with plate separation [5], [6].

What inspired Scharnhorst to investigate the effect that now bears his name was his interest in the Casimir effect, together with a study due to Rolf Tarrach, which concluded that a heat bath modifies the velocity of light compared to its vacuum

---

[4] [1, 2]. In its first and best known incarnation the Casimir effect is between two flat plates; it has also been studied for other geometries [3], [4].

[5] See notably [1] for the discussion of such an experiment.

[6] Their papers were submitted to *Physics Letters B* within a week of one another (Scharnhorst's on December 4 1989 [5], Barton's on December 13 1989 [6].)

value, and analogous studies where light propagated through a background of electromagnetic fields instead.[7] Formally, Scharnhorst derived the QED effective action in the presence of two parallel, perfectly conducting plates. He modeled the plates by imposing boundary conditions, following the approach developed by Bordag *et al* in 1985 for the Casimir effect [8]. The presence of these boundary conditions modifies the photon propagator, which leads to corrections for the photon vacuum fluctuations in the effective action. From this effective action Scharnhorst obtained the permittivity and permeability tensors characterizing the vacuum, and from these the index of refraction and velocity of the photons normal to the plates.

Gabriel Barton's derivation

Gabriel Barton's derivation was meant to be simpler than Scharnhorst's.[8] Barton derived the field equations, i.e. analogous to Maxwell's equations, by varying the vector potential (in the Coulomb gauge) in the Euler-Heisenberg Lagrangian, which takes into account vacuum polarization to one loop [6]. From the resulting field equations, Barton obtained the polarization and magnetization of the vacuum, and introduced in them additional electric and magnetic fields representing the "wave whose propagation [he wished] to study", i.e. the photons. Barton then obtained the "electric and magnetic polarizabilities" (that is, the dielectric and magnetic susceptibility tensors), in which he introduced the vacuum expectation values appropriate to the Casimir set-up. These allowed him to determine the shift in the permittivity and permeability tensors, compared to the analogous quantities for the true vacuum case (i.e. in the absence of the plates). Barton then derived the shift in the index of refraction, and from it the shift in the velocity of the photons travelling normal

---

[7] [7]; [5], p.354.

[8] As Barton stated at the end of his paper, he was aware of Scharnhorst's work prior to its publication. [6], p.561.

to the plates.

Both Scharnhorst and Barton found the same result:

$$c_{\perp} = \left(1 + \frac{11}{(2^6)(45)^2}\frac{e^4}{(m_e L)^4}\right)\; c_0, \tag{1.1}$$

where c is the speed of light in free vacuum and L the distance between the Casimir plates.[9]

1.2 The issue of signal velocity

Scharnhorst and Barton's derivations shared a key feature: both obtained the velocity of the photon by deriving an index of refraction for the Casimir vacuum. Consequently, the velocity they found was a phase velocity. Because the expression they obtained for this index of refraction is independent of frequency, this phase velocity is also the group velocity – hence superluminal as well. Both Scharnhorst and Barton stated so explicitly.[10] Now, that a phase velocity, or that a group velocity, turn out to be superluminal does not *per se* conflict with the speed of light postulate: what has come to be regarded as relevant in this context is the signal velocity.[11] However, that both the phase and the group velocities be superluminal was immediately taken to beg the question: what about the signal velocity? Although he did not stress the matter, Barton asserted that it too was equal to the phase velocity – hence superluminal, but that this offered no conflict with Lorentz

---

[9] Strictly speaking this is the form given by Scharnhorst, [5], p.358. Barton expressed his result in the equivalent form:

$$v = \left(1 + \frac{11\pi^2}{2^2\,3^4\,5^2}\frac{e^2}{(m_e L)^4}\right)\; c,$$

[6], p.561.

[10] [5], p.358, [6] p.559.

[11] A well-known situation where the phase and group velocities are larger than c is that of "anomalous dispersion", i.e. when the frequency of the propagating light is close to an absorption resonance frequency of the medium; see for instance [9], chapter 7 sections 7.5B & 7.8.

invariance:

> These speeds are independent of the frequency ω (provided $\omega << m$), whence phase, group, and signal velocity coincide. Lorentz invariance is no bar to $c'_\perp > 1$, because it obtains only for boosts parallel but not normal to the mirrors.[12]

Scharnhorst and Barton's work prompted responses to the effect that their result could not constitute a signal velocity. Two papers appeared later the same year: one by Peter Milonni and Karl Zvozil in October, and another by Shahar Ben-Menahem in November [10], [11]. All in all, they presented three different arguments, notably to the effect that measurement uncertainties would preclude actually measuring the signal velocity as superluminal. Scharnhorst and Barton responded three years later, in 1993, in a joint paper. Most of their discussion relied on the Kramers-Kronig relations. They argued that the latter imply one of two options: either the signal velocity as well is superluminal, or the Casimir vacuum behaves like an amplifying medium for at least some frequencies. They stressed that they did not believe a superluminal signal velocity to be at odds with relativity in the context at stake:

> The presence of the mirrors breaks Lorentz invariance along the mirror normal (the mirrors define a preferred inertial frame), which obviates the arguments used in special relativity to prove that no signals can travel faster than light does in unbounded (Lorentz invariant) space.[13]

However they did not argue in favor of one alternative over the other.[14]

A decade later, in 2001, the effect was derived in a more general form, *i.e.* for pho-

---

[12] [6] p.559. Barton did not elaborate at this point; how exactly this statement can be interpreted will be discussed in detail in chapter 3.

[13] [12], p.2038. They did not explain how *exactly* the presence of the plates could be responsible for breaking Lorentz invariance, however as we shall discuss in chapters 2 & 3, the Casimir vacuum can be thought to behave like a medium, whose rest frame constitutes the preferred inertial frame.

[14] [6] p.559, [12].

tons traveling in arbitrary directions by Stefano Liberati, Sebastiano Sonego and Matt Visser [13, 14].

They found that in most directions the phase and group velocities were both superluminal but actually differed in general:

$$v_{\text{phase}}(\varphi) = c\left(1 + \xi\ \cos^2\varphi\right)^{1/2},$$

with

$$\xi = \frac{11\,\pi^2\,\alpha^2}{4050\,a^4\,m_{\text{e}}^4} \approx 4.36 \times 10^{-32}\left(\frac{10^{-6}\,\text{m}}{a}\right)^4,$$

and:

$$v_{\text{group}}(\varphi) = c\left(\frac{1 + (2\xi + \xi^2)cos^2\varphi}{1 + \xi\ cos^2\varphi}\right)^{1/2},$$

and become equal at first order in α^2:

$$v_{\text{phase}}(\varphi) \approx\ v_{\text{group}}(\varphi) \approx\ c\left(1 + \frac{11\pi^2}{8100}\alpha^2\frac{1}{a^4 m_e^4}\ \cos^2\varphi\right).^{15}$$

They published a second paper the following year in which they revisited these results, gave an expression for the the signal velocity which they argued to be superluminal as well, and vigorously defended the idea that this is not incompatible with special relativity. They presented several arguments to this effect. They stressed that "special relativity only requires, for its kinematical consistency, that there be an invariant speed", but does not demand the existence of a maximal one, and they showed that the Lorentz transformations can be derived without reference to such a maximal speed. They also argued that faster-than-c propagation does not lead to causal paradoxes provided that the propagation can be described by an effective metric, and they used a theorem regarding stable causality to show that

---

[15] [13], pp.6,7 [14], pp.5-7.

causal paradoxes would not occur for photons propagating superluminally between a pair of Casimir plates. The concept of effective metric was central both to the latter two arguments, as well as to their formal derivation [15, 16]. We shall discuss these issues in greater detail in chapter 2.

After discussing the relevance of an effective metric formalism at the end of chapter 2, I shall present in chapter 3 the debate on the crucial issue as to whether the signal velocity may be superluminal. As we shall see then, one concern that has prominently featured in these debates is whether the behavior of high frequency photons would deviate meaningfully from their low frequency counterparts. Indeed, all derivations of the Scharnhorst effect are only valid in the regime where the frequency of the Scharnhorst photon obeys $\omega << m_e$ with m_e the electron mass. This is unavoidable in any effective theory. However the signal velocity corresponds to the phase velocity in the infinite frequency limit, and this has naturally led to the concern that the low frequency result may therefore tell us nothing about the signal velocity. As briefly mentioned above, Scharnhorst and Barton argued that some conclusions can nevertheless be drawn by considering the Kramers Kronig relations.

1.3 Conceptual interpretation of the Scharnhorst effect

1.3.1 The Casimir vacuum as a dielectric medium

Meanwhile, the notion that photons would see their *phase* velocity undergo a shift in the Casimir vacuum, compared to its standard vacuum value of c, has been unanimously accepted. It is a result whose significance has rarely been stressed, no doubt obscured by the attention that the signal velocity has drawn. Yet it is remarkable in its own right, as it would justify viewing the "Casimir vacuum" as a dielectric medium. This in itself begs questions of interpretation: can a concrete conceptual model, an account in terms of specific processes involving explicit theoretical entities, be given for the Scharnhorst effect, as have been proposed for its Casimir counterpart? If so, which one? We would expect an account of the Scharnhorst effect to share some features at least (processes, entities...) with the conceptual model that accounts for the Casimir force. A first difficulty is that when it comes to the latter we are faced with an abundance of riches: the Casimir effect has been thought to arise from zero-point energies, which may or may not be justifiably identified with vacuum fluctuations and / or virtual particles, but it can also be accounted for without any reference to the vacuum, in terms of source fields (radiation reaction) instead. This is what motivates our attempt to derive the Scharnhorst effect as well using a source theory framework. Before discussing these matters of interpretation any further however, let us briefly introduce the Casimir effect itself.

1.3.2 The Casimir effect

As briefly mentioned above, the Casimir effect refers to the force between two electrically neutral conducting plates positioned very close to one another.[16] In its original and simplest form, the plates are flat, and attract each other with a force given by:

$$F(d) = -\frac{\pi^2 \hbar^2 c}{240\, d^4}, \tag{1.2}$$

where d is the distance between the plates.[17]

As shall be discussed in chapter 5, there are several ways to derive this result. What has become the standard, textbook derivation shows that a configuration where the

---

[16] A typical separation for the plates is of the order of microns. The Casimir effect, unlike the Scharnhorst effect, has been observed experimentally.

[17] [17], p.4247.

plates are closer to one another is less energetic than if they are further apart. It involves the concept of energy density of the vacuum.

The idea goes as follows: even in the vacuum state, the photon field has modes, whose lowest possible energy is non zero (*i.e.* their so-called "zero-point energy" $\frac{1}{2}\hbar\omega$, where ω is the frequency of the mode). In normal, unbounded vacuum these modes can be of any frequency. However in the presence of metal plates, electric and magnetic fields (hence the photon field) have to satisfy boundary conditions, so only a small fraction of the modes can exist. Many more modes can form in the regions outside the plates (where they have to vanish on only one plate, hence satisfy only one boundary condition) than in between (where they have to vanish on both).[18]

As it turns out, this implies that a configuration with the plates closer together is less energetic. At first sight, this may seem counterintuitive: the energy density between the plates is lower than outside, so bringing them closer reduces the region of lower energy. However it also modifies the value of the energy densities. Closer plates imply shorter wavelengths between them, hence higher energy modes there; meanwhile, the reverse holds true for the outer regions, i.e. the energy of the modes there is lower when the plates are closer together. Because there are many more modes outside the plates than in between, it is the change in the outer regions that dominates, and lowering the energy of the outer modes lowers the energy of the whole set-up.

This standard derivation thus relies on the concept of zero-point energies, which motivates viewing the Casimir effect as evidence for the latter. In fact, it has often been hailed as direct evidence of "vacuum fluctuations" and "virtual particles." Popular accounts, in particular, have often identified the modes with the latter and ascribed the Casimir force to their exerting more pressure outside than between

---

[18] Notably, modes of wavelength more than twice the plates separation cannot form between the plates.

the plates, pushing them together: the vacuum is described as teeming with short-lived "virtual" particle-antiparticle pairs that gain enough energy ΔE to come into existence for a brief time Δt by virtue of the Heisenberg "uncertainty" relation $\Delta E\,\Delta t \geq \frac{\hbar}{2}$. In fact the relationship between the concepts of "zero-point energies", "vacuum fluctuations" and "virtual particles" are not as straightforward as these accounts suggest, as will be discussed in chapter 5.[19]

1.3.3 Analogous conceptual model for the Scharnhorst effect

Explicit conceptual models for the Scharnhorst effect have rarely been proposed. However in a non-technical essay describing the effect for a lay-audience John Cramer did offer an explicit description, in terms of virtual particles. His description is based on interpreting the Feynman diagrams that a QFT derivation of the effect involves.[20] He ascribed the Scharnhorst effect to a decrease in the density of virtual particles in the Casimir vacuum, compared to its unbounded, trivial counterpart. Cramer portrayed photons travelling in a vacuum (whether trivial or Casimir) as slowed down by their ability to form virtual particles, which interact with virtual photons from the vacuum:

> A traveling photon may briefly be transformed into a virtual electron-positron pair, which moves forward less than one photon wavelength before annihilating to create a new photon indistinguishable from the old one. During the photon's brief existence as a pair, one of the virtual particles may initiate a "game of catch" using a virtual photon as the ball, tossing it one or more times to itself or its partner [i.e. the virtual

---

[19] For an actual derivation of the Casimir effect in terms of vacuum-field radiation pressure see notably[18].

[20] For this reason the description that follows is subject to whatever criticism one may wish to level against ascribing Feynman diagrams physical meaning, notably on the ground that the "processes" in question are not observable.

> electron and/or positron interact with virtual photons]. These QED complications of the smooth passage of photons through space have the effect of making the photon travel more slowly. In part, this is because the photon spends a fraction of its existence as an electron-positron pair which can only travel at sub-light velocity [19].

While the virtual electron-positron pair is interacting with a virtual photon, it is prevented from forming the "new photon indistinguishable from the old one." The higher the density of virtual photons, the greater the number of interactions the pair is likely to undergo before the new photon forms; this somehow results in the photon spending more time in the form of a (virtual) electron-positron pair which travels slower than the photon would. Hence the denser the vacuum is in virtual photons, the slower the observable photon travels. In normal, unbounded vacuum, this results in a velocity of c. In a Casimir vacuum where the density of virtual photons is lower, the Scharnhorst photons get slowed to a lesser extent, so travel faster:

> Because these virtual photons are absent, they cannot participate in games of catch between virtual particles. Therefore a real photon travelling between the plates spends less time as an electron-positron pair because the QED vacuum fluctuations are suppressed. For this reason, the photon travels faster across the gap [19].

This model fits with comparative studies on different types of backgrounds: electric or magnetic fields, heat baths, and gravitational backgrounds as well as Casimir vacuum – and indeed, recall that it was studies on how a heat bath and electromagnetic backgrounds modify the speed of light that had inspired Scharnhorst in the first place [7]. An interesting result has been that the change in the speed of light is indeed related to the change in the energy density of these backgrounds [20, 21]. While photons have speeds lower than c in backgrounds of higher energy density, predicted superluminal photon velocity in both the Scharnhorst effect and Hathrell Drummond effect (which we shall discuss in chapter 2) is associated with a lower energy density of the relevant background (which is gravitational in the latter case). If the Scharnhorst effect were to be confirmed experimentally, and Cramer's description is correct, it would seem to imply that in the usual, unbounded, trivial vacuum that the speed of light postulate refers to, photon speed is determined by interactions. This immediately raises the question of how this view fits with special relativity. We shall explore this crucial question at the end of chapter 2.

1.3.4 Alternative conceptual models for the Casimir and Scharnhorst effects

On the other hand, as mentioned above, the Casimir effect itself can be accounted for without reference to the concept of quantum vacuum, purely in terms of source fields, in the form of radiation reaction fields, i.e. the electromagnetic fields responsible for the energy losses experienced by accelerating charges. These are classical fields, and accounting for the Casimir effect in this way does not involve quantum effects.

In this approach, the Casimir force is derived by considering the polarization energy of the atoms in the plates, and how this energy changes as the distance between the plates is varied. This polarization can be attributed to the vacuum field (which constitutes yet an alternative way of deriving the Casimir effect), but also to the radiation reaction field.

In fact, that one can use a description in terms of radiation reaction without reference to the quantum vacuum is not only true of the Casimir effect, but of most phenomena that have been ascribed to vacuum fluctuations. This naturally leads to the question: can the Scharnhorst effect also be derived in this framework? How does the background electromagnetic field between the plates generated by radiation reaction sources within them affect the propagation speed of a photon? Obtaining a velocity of c by such a calculation would require understanding the discrepancy between the two approaches; on the other hand, finding the same superluminal results as have been obtained so far would require a different conceptual account than the one virtual particles offer.

1.4 Overview of dissertation

Before focusing on the Scharnhorst effect, we shall discuss in chapter 2 the issue of faster-than-c propagation in field theories in general. Indeed, in addition to the Scharnhorst effect, several of these theories have predicted superluminal behavior:

- the Drummond-Hathrell effect, which can be viewed as a gravitational analog to the Scharnhorst effect as photons have been predicted to move faster-than-c in some gravitational backgrounds.
- bimetric theories, which take the existence of two different metrics as a matter of hypothesis. These can be found in two research programs: the Modified Newtonian dynamics (MOND) program, and varying speed of light (VSL) theories.
- k-essence, which is a theory meant to account for the observed acceleration of cosmological expansion without introducing a cosmological constant; it is a related but alternative model to Quintessence.

That the possibility of superluminal propagation raises serious concerns is doubtless an understatement: it has often been argued to automatically imply the possibility of causal paradoxes, and to break Lorentz invariance, rendering the afflicted theories incompatible with special relativity. The remainder of chapter 2 is devoted to the exploration of these crucial issues.

We shall turn our attention to the Scharnhorst effect in chapter 3. I shall present an historical introduction, focusing on the debate that took place in the physics literature regarding the possible implications of Scharnhorst's result for the signal velocity of the photons. I will discuss the two papers in which the effect was first predicted, in 1990: in February, Klaus Scharnhorst's seminal paper was published, soon followed in March by Gabriel Barton's work which offered an alternative derivation. Both found that the phase and the group velocity of photons propagating in a Casimir vacuum normal to the plates would be larger than c. Later that year, Peter Milonni and Karl Svotzil first, followed soon afterwards by Shahar Ben-Menahem, presented several arguments to the effect that this superluminal character could not hold for the signal velocity. In particular, they emphasized that uncertainties would preclude such a measurement. Scharnhorst and Barton jointly responded in 1993. They noted that the arguments put forward largely hinged on the fact that their derivations are not valid at high frequency – but only for photons satisfying $\omega << m_e$. They countered that one can nevertheless draw partial conclusions for that frequency domain, by using the Kramers-Kronig relations, which turn out to imply one of two options: either the superluminality holds for the signal velocity as well, or the Casimir vacuum behaves like an amplifying medium at some frequencies. Although they stated that in this case, a superluminal signal velocity would not be at odds with Lorentz invariance, they did not in fact favor either alternative over the other. About a decade later, in 2001 and 2002, Stefano Liberati, Sebastiano Sonego and Matt Visser rederived and generalized the Scharnhorst effect within the framework of an effective metric approach. They defended the idea that a superluminal signal velocity in a Casimir vacuum is perfectly compatible with special relativity: their 2002 paper was almost entirely devoted to this purpose. I shall examine these discussions and point to what I believe to be unresolved issues. In particular, even the effective-metric approach of Liberati *et al* remains open to the criticisms regarding the measurability of an increase in the signal velocity, because this derivation as well is only valid for photons of frequency $\omega << m_e$, whereas the signal velocity is defined in the $\omega \to \infty$ limit. The reason why results are limited to this regime is that the theory used is based on an power expansion in $\frac{E}{m_e}$ where E is the typical energy of the photons involved.

In chapter 4, I shall present the derivation of the Scharnhorst effect within the framework of Soft Collinear Effective Theory (SCET). We shall see that it is in fact possible to address concerns related to the expansion in $\frac{E}{m_e}$ using this approach. Indeed, SCET ensures that errors associated with higher order corrections are well-controlled: in contrast to the standard theory, in SCET higher order terms are guaranteed to be power-suppressed. This is because SCET separates the physics into different energy-momentum scales. SCET involves a perturbation expansion in a dimensionless parameter ("λ"), which is the ratio of the low to the high energy-momentum scale, and particles are represented by different fields, depending on whether they have low or high energy and momentum. This guarantees that the factors that multiply the expansion parameters are of order one, which in turn ensures that higher-order terms are power suppressed with respect to their lower order counterparts, and hence that corrections are well-controlled. The first sections of chapter 4 will show how the Scharnhorst effect can be modelled within the framework of SCET. The latter part of the chapter consists in the derivation proper of the Scharnhorst effect within this framework: we shall derive the relevant SCET Lagrangian, from it the second-order polarization tensor, and from the latter we shall obtain the phase velocity of Scharnhorst photons. We shall see that, according to SCET, the result is the same, faster-than-c phase velocity as was derived by Scharnhorst and his successors.

Chapter 5 explains the rationale for a second project: the derivation of the Scharnhorst effect on the basis of source theory. Because the Scharnhorst effect is predicted to occur between Casimir plates, one may expect it to involve the same physical processes as the Casimir effect. The standard account involves "zero-point energy" and "vacuum fluctuations" — indeed, the Casimir force has often been hailed as the most direct experimental evidence we have of these. However alternative models have been proposed for the various physical effects typically accounted for by appealing to vacuum fluctuations, that is the Lamb shift, spontaneous emission, van der Waals forces, cavity modified spontaneous emission, and, most notably, the Casimir effect itself.[21] These alternative descriptions entail source theory, and the concept of radiation reaction. Accounts in terms of radiation reaction *vs* vacuum fluctuations are not necessarily incompatible: as Peter Milonni showed, the two descriptions are related to a formal choice, i.e. which ordering one selects for the quantum operators involved [17]. Whether such a choice is purely a matter of preference or there are objective reasons to prefer an ordering over another has been the object of some disagreement in the literature, and the debate has notably involved the fluctuation-dissipation theorem, which we shall also examine.

Chapter 6 relates an attempt at deriving the Scharnhorst effect in the source theory framework. The structure of the calculation follows Gabriel Barton's work. The essential difference between Barton's classic derivation and the proposed one lies in the vacuum expectation values of products of electromagnetic fields, *i.e.* $\langle E_i E_j \rangle$, $\langle B_i B_j \rangle$ and $\langle E^2 - B^2 \rangle$. In Barton's derivation, the latter represent zero-point fields. In contrast, in the present work they will stand for source fields, due to fluctuations

---

[21] Much of the material in this chapter relates work done by Peter Milonni and colleagues.

in polarization within the plates. Following Milonni and Shih, we shall ensure that these source fields alone appear in our derivation, not the vacuum field, by normal ordering the field operators [17]. However a definite result was not obtained, for reasons explained at the end of chapter (6).

The conclusion (chapter 7) will examine the implications of the result obtained by the SCET approach, *i.e.* that the phase velocity derived within the framework of SCET is the same as was previously obtained by Scharnhorst, Barton, and later researchers. After reviewing the arguments put forward in past debates regarding the possible superluminal character of the signal velocity, we shall explain what the relevance of our result is within this history.

Chapter 2

SURVEY OF SUPERLUMINAL EFFECTS IN FIELD THEORIES

2.1 Introduction

The possible occurrence of faster-than-c propagation has been suggested in several contexts described by field theories. The present chapter starts by describing these instances of "superluminal" behavior, and how they come about in the theories that involve them. The first to be discussed, *i.e.* the Drummond-Hathrell effect, like the Scharnhorst effect which is the topic of the rest of this dissertation, concerns the propagation of light itself: photons are predicted to travel faster than c in some circumstances, where c refers to their speed in free, Minkowski space vacuum. Next, I present how superluminal behavior occurs in bimetric theories, which have been proposed in the context of the MOND paradigm of gravity, and variable speed of light (VSL) theories. I then discuss the case of k-essence theories. The second half of this chapter is devoted to the various arguments that have been presented in the literature as to whether or not these instances of superluminality lead to causal paradoxes.[1]

---

[1] I mostly focus on the physics literature, but some of these ideas have also been discussed in the philosophy of physics literature. See notably [22, 23], [24], [25].
Also note that in this context what is meant by "causality" not being violated is the absence of causal paradoxes — not what physicists often mean by the term, *i.e.* the absence of superluminal signals usually thought to imply such paradoxes.

2.2 The Drummond-Hathrell effect

2.2.1 Description

The Drummond-Hathrell effect was first discussed a decade before the Scharnhorst effect. It has since sometimes been described as a gravitational version of the Scharnhorst effect.[2] And indeed both constitute theoretical predictions that photons travel at speeds faster than c. However whereas the Scharnhorst effect is ultimately due to boundary conditions, *i.e.* the presence of Casimir plates, the Drummond-Hathrell effect is due to a non-isotropic gravitational curvature. Like the Scharnhorst effect, the Drummond-Hathrell effect is tiny. It only comes to bear when the curvature length-scale L is of the order of the photon's Compton wavelength λ_C.

Unlike photons between Casimir plates, photons propagating through a gravitational background are sometimes predicted to travel *slower* than in a Minkowski spacetime. Whether their speed is smaller, the same, or larger than c depends on several factors:

- the local curvature — hence the gravitational background under consideration.
- the photons' direction of propagation.[3]
- the photons' direction of polarization.

Because of this diversity of factors, early studies of the effect focused on finding out what behavior various backgrounds generate.

In a de Sitter background, which has isotropic curvature, the propagation of the photons is not at all affected: they travel at c in all directions, irrespective of their polarization.[4]

---

[2] [15] p.168, [16] p.2.

[3] This is also true for the Scharnhorst effect since Scharnhorst photons travel at c when moving parallel to the plates. Generally speaking, their speed varies with direction, being the largest normal to the plates.

[4] [26], p.345.

For Ricci flat backgrounds (*i.e.* vacuum solutions of Einstein's equations, for which the stress energy tensor $T_{\mu\nu}$ and the Ricci tensor $R_{\mu\nu}$ vanish), two cases were studied: photon propagation through a gravitational wave background, and through a Schwarzschild one. Generally speaking, in Ricci flat spacetimes, the change in speed is equal and opposite in sign for the two transverse polarisations. This implies that the polarisation averaged velocity shift vanishes. In other non Ricci flat spacetimes, *i.e.* where the matter energy-momentum tensor is no longer zero, the polarisation averaged velocity shift is proportional to it.[5] More specifically, it was found that:
- for a background that is a plane gravitational wave, *i.e.* described by the metric $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, photons have a speed larger than c in regions where the component $b = h_{11} = -h_{22}$ is positive, when the photons travel antiparallel to the wave and are polarized in the "1" direction (in a coordinate system where the gravitational wave travels in the "3" direction). For the other polarization — *i.e.* 2, the reverse holds.[6]
- in the Schwarzschild case, radial photons travel at c. Other, transverse photons do not. When their polarization is transverse, their speed is less than c; but when it is radial, it is again larger than c.[7]
- in a Reissner-Nordstrom background, that is, one describing a charged black hole, radially directed photons travel at c, but those with an orbital velocity component see their speed affected. In this case, three contributions come into play:[8]
1) the gravitational contribution due to the mass of the black hole. This is the same as for a Schwarzschild background. It tends to decrease the speed of tangentially polarized photons, and increase that of radially polarized photons.[9]

---

[5] [27] p.387, [28] p.8.

[6] [26], pp.347-349.

[7] [26], p.349.

[8] [29] pp.644-645, [30] pp.10-11.

[9] By an amount $\frac{1}{240}\frac{\alpha}{\pi}\frac{1}{(Mm)^2}\left(\frac{2M}{r}\right)^3$ for photons moving on an orbital trajectory, where Q and M are the charge and mass of the black hole, m is the electron mass and α is the fine structure constant. [29] p.635.

2) the direct contribution of the electromagnetic field. It tends to decrease the velocity of the photons, most strongly so if they have tangential polarization (for orbital photons).[10] These first two contributions are approximately of the same order. Which one dominates depends upon the radial distance r from the black hole.[11]
3) the contribution of the gravitational field induced by the charge. It tends to increase the speed of the photons whatever their polarization, but more so when they are tangentially polarized.[12] Provided that the black hole's charge Q is of the order of Q_0, it is much weaker than the direct electromagnetic contribution however, by a power of $\frac{m^2}{\alpha}$.[13]
Therefore, generally speaking, whether radially polarized photons are faster than c hinges on the relative strength of the first two contributions, which in turn depend on r (the radial distance from the black hole). For $Q \leq Q_0$ for instance, photons moving orbitally at the horizon $r = 2M$ are superluminal.
- in a Kerr background, which describes rotating black holes, photons traveling on radial trajectories do see their speed affected — unlike in the Schwarschild and

---

[10] By an amount $\frac{1}{240}\frac{\alpha}{\pi}\frac{1}{(Mm)^2}\left(\frac{2M}{r}\right)^3\frac{7}{6}\left(\frac{Q}{Q_0}\right)^2\frac{2M}{r}$ for tangentially polarized versus $\frac{1}{240}\frac{\alpha}{\pi}\frac{1}{(Mm)^2}\left(\frac{2M}{r}\right)^3\frac{2}{3}\left(\frac{Q}{Q_0}\right)^2\frac{2M}{r}$ for radially polarized photons. $Q_0 = \frac{Mm}{\alpha}$ is the accretion limit charge.[29] p.635.

[11] [29] p.635, [30] p.1.

[12] For orbital photons: $\frac{1}{240}\frac{\alpha}{\pi}\frac{1}{(Mm)^2}\left(\frac{2M}{r}\right)^3\frac{13}{12}\left(\frac{\alpha}{m^2}\right)^{-1}\left(\frac{Q}{Q_0}\right)^2\frac{2M}{r}$ for tangentially versus $\frac{1}{240}\frac{\alpha}{\pi}\frac{1}{(Mm)^2}\left(\frac{2M}{r}\right)^3\frac{5}{12}\left(\frac{\alpha}{m^2}\right)^{-1}\left(\frac{Q}{Q_0}\right)^2\frac{2M}{r}$ for radially polarized photons.
Overall photons on an orbital trajectory (subscript refers to polarization):

$$v_r = 1 + \frac{1}{240}\frac{\alpha}{\pi}\frac{1}{(Mm)^2}\Big(\frac{2M}{r}\Big)^3\Big[1 + \frac{5}{12}\Big(\frac{\alpha}{m^2}\Big)^{-1}\Big(\frac{Q}{Q_0}\Big)^2 - \frac{2}{3}\Big(\frac{Q}{Q_0}\Big)^2\frac{2M}{r}\Big] \tag{2.1}$$

$$v_\theta = 1 + \frac{1}{240}\frac{\alpha}{\pi}\frac{1}{(Mm)^2}\Big(\frac{2M}{r}\Big)^3\Big[-1 + \frac{13}{12}\Big(\frac{\alpha}{m^2}\Big)^{-1}\Big(\frac{Q}{Q_0}\Big)^2 - \frac{7}{6}\Big(\frac{Q}{Q_0}\Big)^2\frac{2M}{r}\Big] \tag{2.2}$$

[29], p.635.

[13] [29] p.635, [30] p.1. However a Reissner-Nordstrom black hole is unlikely to have such a significant charge: it would attract oppositely charged particles more strongly than others, which would decrease its charge.

Reissner-Nordstrom cases [31], [32]. Whether their velocity increases or not depends on their polarization and how far they are from the black hole: the velocity shift vanishes at the horizon. Also, the velocity shifts are equal and opposite for the two transverse polarisations.

- in a FRW background, used to represent an expanding or contracting universe, the velocity of the photons is the same for all propagation directions and all photon polarizations. Drummond and Hathrell attribute this to the fact that the FRW metric describes a homogeneous and isotropic spacetime. Following the Friedmann cosmological model, they consider the case in which this spacetime is filled with a homogeneous and isotropic fluid. In this case, they find the photon speed to be larger than c "for all physically reasonable pressures." They also conclude that in a radiation dominated universe, the photon speed increases as one considers earlier and earlier times.[14]

2.2.2 Historical background

Drummond and Hathrell predicted the effect that bears their name in 1980 [26], but it is only in the mid-1990s that their work again arose curiosity. R.D. Daniels and G.M. Shore studied it in 1993 in the Reissner-Nordström spacetime [29], [30], which describes a charged black hole, and in 1996 in the Kerr spacetime, for a rotating black hole [31], [32].

In 2003, Shore reformulated their calculation in order to highlight pertinent physical aspects [33], [34]. Until then the equation of motion of the photons had been given in terms of the Ricci and Riemann tensors, *i.e.*:

$$k^2 \; - \; \frac{2\,b_1}{m^2} R_{\mu\lambda} k^\mu k^\lambda \; + \; \frac{8\,b_2}{m^2} R_{\mu\nu\lambda\rho} k^\mu k^\lambda a^\nu a^\rho \;\; = \;\; 0, \qquad (2.3)$$

---

[14] [26], pp.351-352.

where $k^2 = k^\mu k_\mu$, a_μ is the polarization, and b_1 and b_2 are some of the coefficients of the leading irrelevant operators in the effective action.[15]

Shore used Einstein's equation to express the equation of motion in terms of the matter energy-momentum tensor and the Weyl tensor, thereby showing the distinct contributions of matter and of gravitation *per se* to the Drummond-Hathrell effect:

$$k^2 \;-\; \frac{8\pi}{m^2}(2b+4c)T_{\mu\lambda}k^\mu k^\lambda \;+\; \frac{8c}{m^2}C_{\mu\nu\lambda\rho}k^\mu k^\lambda a^\nu a^\rho \;=\; 0, \tag{2.4}$$

where $T_{\mu\lambda}$ is the energy-momentum tensor and $C_{\mu\nu\lambda\rho}$ is the Weyl tensor.[16]

This leads to the following form for the phase velocity of the photons:[17]

$$v_{\rm ph}(0) \;=\; 1 \;+\; \frac{8\pi}{m^2}(2b+4c)T_{\mu\lambda}k^\mu k^\lambda \;+\; \frac{8c}{m^2}C_{\mu\nu\lambda\rho}k^\mu k^\lambda a^\nu a^\rho. \tag{2.5}$$

The terms containing $T_{\mu\lambda}k^\mu k^\lambda$ in Eq.(2.4) and Eq.(2.5) are not specific to the situation of interest in the Drummond-Hathrell effect: they also appear when the photons propagate through a background electromagnetic field, or a finite temperature background. When the Weyl curvature is zero ("Weyl-flat" spacetime) so that only this term modifies the photons speed, whether it is faster than c depends on the sign of $T_{\mu\lambda}k^\mu k^\lambda$ and $b+2c$. Now the weak energy condition ensures that $T_{\mu\lambda}k^\mu k^\lambda \geq 0$. The speed of the photons is therefore decreased provided that $b+2c<0$. This is

---

[15] [33] p.46, [34] p.2.

[16] [33] p.46, [34] p.2. Note that in Minkowski spacetime we would have $k^2 = k^\mu k_\mu = 0$.

[17] Or equivalently:

$$k^2 \;+\; \frac{\alpha}{360\pi}\frac{1}{m^2}\Big[-(b+2c)\;R_{\mu\nu}k^\mu k^\nu \;\pm\; 4c\;|C_{\mu\nu\lambda\rho}k^\mu m^\nu k^\lambda m^\rho|\Big] \;=\; 0$$

leads to a phase velocity:

$$v_{\rm ph}(0) \;=\; 1 - \frac{\alpha}{360\pi}\frac{1}{m^2}\Big[-(b+2c)\;R_{\mu\nu}\ell^\mu \ell^\nu \;\pm\; 4c\;|C_{\mu\nu\lambda\rho}\ell^\mu m^\mu \ell^\nu m^\rho|\Big],$$

with $R_{\mu\nu} = 8\pi T_{\mu\nu}$ (in $G=1$ units), [35] p.242, [36] p.26.

the case in an electromagnetic or thermal background. However, in the presence of gravitational curvature, $b + 2c > 0$: indeed in the low-energy effective action for QED in a background gravitational field (the analogue of the Euler-Heisenberg action), first calculated by Drummond and Hathrell in 1980, $b = 26$ and $c = -2$. This factor of $b + 2c$ also corresponds to the factor in Latorre's formula: $\frac{1}{2}(b + 2c)$ here is 11.[18]

It is the third term, the one which involves the Weyl tensor, that truly represents the phenomenon of interest in the Drummond-Hathrell effect. It depends on the polarization of the photons, so the gravitational background leads to birefringence. Shore further expressed the equation of motion in a different formalism in order to call attention to the fact that the Weyl term can increase or decrease the speed depending on this polarization. In the Newman-Penrose formalism, it takes the form:

$$k^2 \ + \ \frac{1}{2}\frac{(4b + 8c)\omega^2}{m^2} R_{\mu\nu}\ell^\mu\ell^\nu \ \ \pm \ \ \frac{4c\,\omega^2}{m^2}(\Psi_0 + \Psi_0^*) \ \ = \ \ 0, \tag{2.6}$$

with ω the frequency of the photon, $\Psi_0 = -C_{\mu\nu\lambda\rho}\ell^\mu m^\nu \ell^\lambda m^\rho$, where ℓ_μ, m_μ are tetrad vectors and ℓ_μ are in the direction of polarization.[19] G. Shore presented an

---

[18] [26] p.22, [35] pp.241-243, [36] pp.24-26. Indeed:

$$v_{\rm ph}(0) \ = \ 1 - \frac{\alpha}{360\pi}\frac{1}{m^2}\Big[-(b+2c)\ R_{\mu\nu}\ell^\mu\ell^\nu \ \ \pm \ 4c\ |C_{\mu\nu\lambda\rho}\ell^\mu m^\mu \ell^\nu m^\rho|\Big]$$

leads to (see [35] pp.242-243, [36] p.26):

$$v_{\rm ph}(0) \ = \ 1 \ + \ \frac{11}{45}\frac{\alpha}{m^2}\ (\rho + P).$$

Latorre *et al* put this (with G not set to 1) in the form of their "unified formula" by using $p = \frac{\rho}{3}$ and $\rho_G = -\rho$, obtaining (see [20], pp.61-63, [21], pp.2-4):

$$v = 1 + \frac{11}{45}\alpha G_N \frac{\rho + p}{m_e^2}.$$

[19] [33] p.46, [34] p.2.

equivalent calculation in more detail in 2007,[20] obtaining:

$$k^2 \;+\; \frac{\alpha}{360\pi}\frac{1}{m^2}\Big[-(b+2c)\;R_{\mu\nu}k^\mu k^\nu \;\pm\; 4c\;|C_{\mu\nu\lambda\rho}k^\mu m^\nu k^\lambda m^\rho|\Big] \;=\; 0, \tag{2.7}$$

which corresponds to a phase velocity:

$$v_{\mathrm{ph}}(0) \;=\; 1-\frac{\alpha}{360\pi}\frac{1}{m^2}\Big[-(b+2c)\;R_{\mu\nu}\ell^\mu \ell^\nu \;\pm\; 4c\;|C_{\mu\nu\lambda\rho}\ell^\mu m^\mu \ell^\nu m^\rho|\Big].^{21} \tag{2.8}$$

When the background is Ricci flat (*i.e.* $T_{\mu\nu} = 0$), so that only the Weyl term contributes, whether photons propagate faster than c therefore hinges upon this feature: one of the polarizations corresponds to faster than c photons and the other to "subluminal" ones.

2.2.3 Interpretation

Comparing the conceptual interpretation of the Drummond-Hathrell effect to that proposed for the Scharnhorst effect we discussed in the Introduction, one again finds both similarities and differences. Both are construed to involve photons turning into virtual electron-positron pairs as they propagate.[22] Yet while the Scharnhorst effect is described as an example of "light-by-light scattering", the Drummond-Hathrell effect is not.[23] In the paper in which they first described the effect, I. Drummond and S. Hathrell explained it as due to the photon acquiring an effective size, characterized by the electron's Compton wavelength of the electron, when it momentarily becomes a virtual electron-positron pair. The change in speed was then attributed to gravitational forces on the pair – hence the importance of

---

[20] Shore wrote this paper in response to [37, 38].

[21] [35] p.242, [36] p.26.

[22] See John Cramer's description of the Scharnhorst effect above, and notably [26], p.343, for the Drummond-Hathrell effect.

[23] [20] p.62, [21] p.3.

anisotropy. However, no conceptual attempt was made to relate relevant features of the gravitational field to explain why the speed would increase in some cases and decrease in others compared to Minkowski spacetime; nor was it made clear why and how photon polarization would affect this behavior. Reference to tidal forces appear more as an attempt to render plausible the appearance of superluminal speeds, implying that the Strong Equivalence Principle (SEP) does not hold in this context: SEP, it is argued, only allows for minimal coupling to gravity "*i.e.* through the connections only, independent of the curvature"; in the Drummond-Hathrell effect the coupling between electromagnetism and gravity does involve curvature, so the "photons" become sensitive to it, and are affected by gravitational tidal forces while they are virtual electron-positron pairs.

2.3 Bimetric theories

2.3.1 Modified Newtonian dynamics (MOND) theories

The phrase "bimetric theories" *stricto sensu* refers to theories that take the existence of two metrics as a hypothesis. Theories where several metrics come into play because they make it possible to define an effective metric (k-essence notably) are usually not regarded as such. Bimetric theories proper have been developed in mostly two contexts: within the Modified Newtonian dynamics (MOND) and the varying speed of light (VSL) research programs. Some of the bimetric theories from both traditions have involved "acausal", *i.e.* superluminal behavior.

In MOND, one of the metrics is associated with gravity, and is used to build the Einstein-Hilbert action. It determines the motion of gravitational waves. The other one is associated with matter, in particular it is used in the geodesic equations so that it defines its motion. When superluminality is predicted, it is related to a non-canonical kinetic term in the Lagrangian, as in k-essence. It is not due to the postulated bimetricity of the theories *per se*, in the sense that in the MOND context, bimetricity is a necessary condition for superluminality, but not a sufficient one. Indeed in most MOND theories the two metrics are conformal to one another, which implies that the causal cones associated with them are identical.

MOND *per se* is purely phenomenological, postulating a deviation from the Newtonian acceleration so that it is weaker far from the gravitational source, to account for the observed "flat" galactic rotation curves, without the need for dark matter. Other theories in the MOND research program consist in attempts to give this phenomenological description a theoretical underpinning. Rather than equally viable alternatives, these attempts are generally considered as a series of successors, each meant to remedy a weakness of its predecessor.

The historical progression of these theories consists in: AQUAL (non-relativistic and relativistic versions thereof), phase-coupling gravitation (PCG), Disformal relativistic AQUAL, TeVeS.[24]

AQUAL gave phenomenological MOND a Lagrangian formulation, and it turned out that this Lagrangian was "aquadratic" — and indeed AQUAL stands for "AQUAdratic Lagrangian". In the 1984 paper in which they introduced AQUAL, Bekenstein and Milgrom also proposed a relativistic version of it, and considered the propagation of gravitational waves [41]. They found that the longitudinal component of the waves propagate superluminally (except in the direction orthogonal to the acceleration background field). They argued that this "acausal propagation" was unlikely to be problematic however. Firstly, they suspected that perhaps this component of the wave could not be emitted. Secondly, they believed that these "acausal waves" could not affect the behavior of particles because the component

---

[24] [39], [40], section 6, "How TeVeS was constructed."

at stake was “only a “conformal factor” multiplying a causally propagating metric” — a Minkowski metric. Because conformally related metrics have the same causal cones, they concluded that “the acausal propagation of the conformal factor does not affect the light cone structure of the measurable metric”.[25] How this could lead to a superluminal propagation velocity for the said component is unclear. They probably did not regard this argument as strongly convincing either, for they went on to suggest that the theory could perhaps be modified to avoid the “acausal behavior” altogether.

Such modification was indeed what Bekenstein performed by proposing PCG (Phase Coupling Gravitation) in 1988 [42], [43], [44], [45]. He modified relativistic AQUAL to make the Lagrangian quadratic — at the cost of adding a second scalar field. PCG presents an interesting subtlety however. One would have expected that a quadratic Lagrangian disposed entirely of any superluminal behavior — after all, in k-essence theories as well, superluminality can be traced back to the non-canonical character of the kinetic term in the Lagrangian. Yet in fact, PCG still allowed for superluminal perturbations, but only when the background solution was unstable anyway.[26]

The history of MOND theories was far from over though. Although PCG was deemed satisfactory enough from the standpoint of causality, it failed to predict a gravitational bending of light greater than in GR, as needed to account for galactic rotation curves – which recall is what the MOND program set out to do.[27] It is in order to fix this issue that a disformal theory was proposed, that is a theory such that the cones associated with both metrics no longer coincided [50], [51]. However if the light deflection was to be *stronger* than in GR as required to match observations,

---

[25] [41], p.14.

[26] [46] p.7, [47] p.7.

[27] The problem was identified as due to the fact that the scalar field enters as a conformal factor multiplying the Einstein metric. [48] p.493, [49] p.3.

the way in which the cones came apart was deemed to imply "acausal behavior."[28] If this version of the theory was to avoid such superluminal behavior, its disformal character alone could not account for the observed light deflection. Sanders and Bekenstein further modified it in 1993-1997 [52, 53], notably by introducing an additional, vector field [48, 49]. In 2004 Bekenstein proposed further changes including a modification of the scalar field action so as to ensure that the theory be "causal" whenever the scalar field is positive. At this point the theory had become known as TeVeS.[29]

Once we are dealing with a disformal theory, the issue arises of which metric should be considered the relevant one for concerns of causality. In his 2004 paper, Bekenstein argued that the pertinent metric is the physical one, *i.e.* that representing the geometry on which matter propagates, as opposed to the gravitational metric.[30] That is, causality violations should be associated with propagation outside the cone associated with the physical metric. His motivation was that "rods and clocks are material systems with negligible self-gravity" so the coordinates to which the Lorentz transformations of special relativity refer are those of local orthonormal frames of the physical metric, not the gravitational one. He accordingly identified c, *i.e.* the invariant speed in Lorentz transformations, as the speed of light rather than that of gravitons.

---

[28] [52] p.493, [53] pp.3-4. R. H Sanders phrases the issue in this way: "if the propagation of classical gravitational waves is to be causal, then the sign of the disformal term must be such that the light bending is actually reduced over that predicted by GR", [48] p.493, [49] p.3. This corresponds to the gravitational cone being narrower than the matter cone. In order to get enhanced gravitational bending of light, it turns out that the gravitational cone must be wider than the matter one. In fact in a more recent work, Jean-Philippe Bruneton argues in favor of this option, [54] p.12, [55] p.12.

[29] [46], [47]. The metrics involved in TeVeS are still disformally related.

[30] Note that this explains the demand noted above that the gravitational cone lie within the matter cone – demand which Bruneton argued should be relaxed.

2.3.2 Varying speed of light bimetric theories

Some of the varying speed of light (VSL) theories that have been proposed to account for the accelerating expansion of the universe are also bimetric. They feature a metric for photons that differs from the gravitational metric. In the recent VSL theories of interest, the fact that the photon metric differs from the gravitational one is postulated rather than derived from an action — hence their being regarded as truly bimetric theories [56]-[57]. They do present connections with some of the other topics presented above however. The Scharnhorst and Drummond-Hathrell effects also have a different metric for photons than for gravitons, even though the existence of this metric does not have the character of a postulate. Supporters of VSL do not hesitate to point out this common feature to justify their approach. Also, in their research on the effect that bears their name, Drummond and Hathrell proposed a VSL theory.[31]

2.4 k-essence

2.4.1 Description

One class of theories known to exhibit "acausal" behavior is k-essence. It constitutes a subset of quintessence theories, with the added peculiarity that the kinetic term in their Lagrangian is not of the standard, "canonical" form:

$$X = \frac{1}{2} g^{\mu\nu} \nabla_\mu \Phi \nabla_\nu \Phi, \tag{2.9}$$

but is instead a function of X.

The context in which quintessence and k-essence arise is cosmology, so let us step

---

[31] [26] p.352.

back a moment. According to our current understanding of the Universe,[32] its history can be divided into *several* eras or epochs characterized by different dynamics. The evolution of the Universe can be described by the Friedmann equations, a set of solutions to Einstein's equations that relate the energy density and pressure p of its content to the scale factor $a(t)$, which expresses its expansion. When solving the Friedmann equations it is necessary to assume a given relation between density and pressure, known as "equation of state". The latter varies depending on the type of content which "dominates" the universe — whose energy density is greatest. Different equations of state are characterized by a parameter defined as the ratio of the pressure to the density: $\omega_q = \frac{p_q}{\rho_q}$. More specifically, a Universe dominated by non-relativistic matter is modeled by taking this matter to be pressureless — the "dust approximation." For a radiation-dominated Universe, the model used is the perfect fluid approximation with $p_q = \frac{1}{3}\rho_q$. These different scenarios correspond to the various epochs mentioned above. The early Universe, during its first 70 000 years (*i.e.* at a redshift of 3400), went through a radiation-dominated era: *i.e.* the energy density of radiation, of relativistic particles, was larger than that of matter. Expansion then led matter to become dominant — although the Universe remained opaque until "recombination" about 380,000 years after the Big Bang (redshift of 1100).

Several sets of observations all suggest that the expansion of the Universe is accelerating, and did not do so in the early stages of its history. This acceleration has been ascribed to the presence of a fluid characterized by a negative pressure — the so-called "dark energy." Dark energy is characterized by an unusual equation of state, with a negative pressure approximately the opposite of its energy density, which would give rise to this increase in expansion rate: *i.e.* $\langle p \rangle = - \langle \rho \rangle$, so that its

---

[32] "Universe" is here to be understood as what the Big Bang marked the start of.

equation of state parameter, that is the ratio of the pressure to the energy density, is: $\omega_q = \frac{p_q}{\rho_q} = -1$.

Beyond defining its dynamical features and naming it, what dark energy consists in is still very much an issue of debate. Quintessence is one of the two main proposals to account for it. Dark energy is often identified with the existence of a cosmological constant, which would have to be extremely small to match observations ($\lambda \approx \frac{3\times10^{-122}\ c^3}{\hbar G}$). Quintessence constitutes an alternative proposal, whereby dark energy is thought of as a scalar field (the term "quintessence" refers to this field itself). This field is considered a dynamic entity. Its equation of state parameter contains kinetic terms and can vary in time:

$$\omega_q = \frac{p_q}{\rho_q} = \frac{\frac{1}{2}\dot{Q}^2 - V(Q)}{\frac{1}{2}\dot{Q}^2 + V(Q)}. \tag{2.10}$$

where Q is the quintessence field and $V(Q)$ the potential energy. In particular, its value differs in the matter dominated and in the radiation dominated epochs.

This aspect of quintessence offers an advantage over the cosmological constant proposal in so far that quintessence has the ability to address the "cosmic coincidence problem": the acceleration must have begun after the formation of galaxies, otherwise it would have prevented it. The density of quintessence is made to vary in such a way that it "tracks" the density of radiation. It is thought that at first its density was lower than that of radiation, and that it kept decreasing with the latter throughout the radiation dominated era. Then its behavior would have changed abruptly when the Universe became matter dominated: at that point it would have begun behaving as dark energy which now represents about 70 % of the mass-energy of the Universe.

k-essence theories have a lot in common with quintessence. They too postulate a dynamical scalar field — the k-field or k-essence. This field also tracks radiation, with the ratio of its energy density to that of radiation remaining fixed during the radiation dominated era, so that k-essence remained subdominant during that period. k-essence cannot mimic the equation of state for matter however. When matter became dominant its energy density is believed to have rapidly decreased down to a fixed value as its pressure switched from positive to negative. However, as the density of matter decreased faster than that of k-essence with the expansion of the Universe, k-essence would have become dominant and begun driving the acceleration in this expansion. The difference between k-essence and quintessence tracker models is that quintessence energy density mimics the equation of state of the energy density of the background whether it is matter or radiation, irrespective of what this equation of state is. Only when it reaches a critical value does the energy density of quintessence freezes (and it develops negative pressure). The models therefore involve a parameter — this critical energy density value — which must be fine-tuned. In contrast the transition in k-essence behavior occurs because of the transition itself to a matter dominated Universe. In this model, that the expansion of the Universe only begun accelerating during the matter-dominated era is accounted for by k-essence's tracking of radiation prior to that, rather than having to fine-tune model parameters or initial conditions in order to account for the history of the acceleration. k-essence can therefore describe the acceleration while avoiding the fine-tuning problem. This capacity hinges on the ability of k-essence to track radiation but not matter behavior. The formal feature which allows for this is an unusual form of the kinetic term in its Lagrangian; and this non-linear character of the kinetic energy is what leads to a negative pressure in these models.

2.4.2 k-essence and superluminal propagation

k-essence theories imply that fluctuations of the field could be superluminal. Formally, the feature that leads to this is the non-canonical character of the kinetic term in the Lagrangian of the theory. k-essence theories are characterized by an action of the form:

$$S_\phi = \int d^4x \sqrt{-g}\, \mathcal{L}\left(X, \phi\right), \tag{2.11}$$

where $\mathcal{L}$ stands for the Lagrangian density and X is the canonical kinetic term:

$$X = \frac{1}{2} g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi.^{33} \tag{2.12}$$

Hence in principle $\mathcal{L}$ can contain any function of X, in contrast to most field theories where the kinetic term is X itself. This action is explicitly Lorentz invariant, yet fluctuations of the field it defines can propagate faster than c. In order to see this, let us consider the equation of motion of these fluctuations. Varying the action with respect to ϕ yields the equation of motion for ϕ:

$$-\frac{1}{\sqrt{-g}} \frac{\delta S}{\delta \phi} = \mathcal{L}_{,X}\, g^{\mu\nu} \nabla_\mu \nabla_\nu \phi + \mathcal{L}_{,XX}\, \nabla^\mu \phi \nabla^\nu \phi \nabla_\mu \nabla_\nu \phi + 2X \mathcal{L}_{,X\phi} - \mathcal{L}_{,\phi} = 0.^{34} \tag{2.13}$$

This can then be written in terms of an "effective metric", simply by grouping together the coefficients of the terms that contain $\nabla_\mu \nabla_\nu \phi$:

$$-\frac{1}{\sqrt{-g}} \frac{\delta S}{\delta \phi} = \tilde{G}^{\mu\nu} \nabla_\mu \nabla_\nu \phi + 2X \mathcal{L}_{,X\phi} - \mathcal{L}_{,\phi} = 0,^{35} \tag{2.14}$$

---

[33] [58] p.3, [59] p.3.

[34] *Idem.* A factor of $\frac{1}{\sqrt{-g}}$ is missing in [59].

[35] A factor of $\frac{1}{\sqrt{-g}}$ is missing in [59].

with

$$\tilde{G}_{\mu\nu}(\phi, \nabla\phi) \equiv \mathfrak{L}_{,X}\, g^{\mu\nu} + \mathfrak{L}_{,XX}\, \nabla^{\mu}\phi\nabla^{\nu}\phi.^{36} \tag{2.15}$$

However what we are really interested in is not the field ϕ itself, but fluctuations in it, as these are interpreted as particles. ϕ can be expressed as:

$$\phi = \phi_0 + \pi, \tag{2.16}$$

where ϕ_0 is the solution of the equation of motion for ϕ with a source and π are the fluctuations on this background.[37] The equation of motion for π can be put in the form:

$$\tilde{G}^{\mu\nu}\nabla_{\mu}\nabla_{\nu}\pi + V^{\mu}\nabla_{\mu}\pi + \tilde{M}^2\pi = \delta J, \tag{2.17}$$

with:

$$\tilde{G}_{\mu\nu}(\phi, \nabla\phi) \equiv \mathfrak{L}_{,X}\, g^{\mu\nu} + \mathfrak{L}_{,XX}\, \nabla^{\mu}\phi\nabla^{\nu}\phi, \tag{2.18}$$

where J is the source of the fluctuations, $\tilde{M}$ is an expression that represents the effective mass, and the expression which is grouped as $\tilde{G}_{\mu\nu}$ is what is referred to as "the effective metric."[38] If we expand π in plane waves, *i.e.* $\pi\ \alpha\ e^{-ik_{\sigma}x^{\sigma}}$, this equation of motion takes the form:[39]

$$(k^{\mu}k_{\mu} + \frac{\mathfrak{L}_{,XX}}{\mathfrak{L}_{,X}}(\nabla^{\mu}\phi k_{\mu})^2)\pi + ... = 0, \tag{2.19}$$

where the terms written explicitly correspond to the first term in the equation of motion, *i.e.* to $\tilde{G}^{\mu\nu}\nabla_{\mu}\nabla_{\nu}\pi + ... = 0$.

---

[36] [58] p.4, [59] p.3.

[37] [58] p.27, [59] p.22.

[38] $V^{\mu}(x) \equiv \frac{\partial\tilde{G}^{\alpha\beta}}{\partial\nabla_{\mu}\phi}\nabla_{\alpha}\nabla_{\beta}\phi_0 + \varepsilon_{,\phi X}\nabla^{\mu}\phi_0$, and $V^{\mu}(x) \equiv \frac{\partial\tilde{G}^{\alpha\beta}}{\partial\nabla_{\mu}\phi}\nabla_{\alpha}\nabla_{\beta}\phi_0 + \varepsilon_{,\phi X}\nabla^{\mu}\phi_0$.
[58] pp.3-4, 27-28, [59] pp.3-4, 21-22.

[39] $\nabla^{\mu}\nabla^{\nu}\pi$ yields $k^{\mu}k^{\nu}\pi$.

Fluctuations obeying the equation of motion:

$$(k^\mu k_\mu)\pi = 0 \tag{2.20}$$

would be propagating at the speed of light, whereas if they obeyed:

$$(k^\mu k_\mu + C)\pi = 0 \tag{2.21}$$

with C negative, they would propagate faster.[40] Looking at Eq. (2.19) this implies that a theory in which the ratio $\frac{\mathcal{L}_{,XX}}{\mathcal{L}_{,X}}$ is negative describes a field whose particles propagate faster than c.

2.5 Does superluminal propagation imply causal paradoxes?

We saw in the previous section how the concept of "effective metric" has been used in the context of k-essence. Our discussion there was purely formal, portraying $\tilde{G}^{\mu\nu}$ as a convenient way of talking about en expression that appears in the equation of motion. However, in some of the literature concerned with superluminal effects in field theory, the concept of effective metric has played an important conceptual role. Indeed it has been used to define the notion of "causal cone", which has served as an intuitive device to argue that the superluminal effects at stake would not violate causality, in the sense that they would not lead to causal paradoxes. The idea is that signals are constrained to propagate on such a causal cone, and that this would prevent Closed Signal Curves (CSCs) from forming.[41]

---

[40] Indeed $k^\mu k_\mu = \frac{\omega^2}{c^2} - \mathbf{k}^2$, so: $(k^\mu k_\mu + C)\pi = 0$ becomes $(\frac{\omega^2}{c^2} - \mathbf{k}^2 + C)\pi = 0$, hence $\frac{\omega^2}{\mathbf{k}^2} = c^2 - C(\frac{c^2}{\mathbf{k}^2})$. $\frac{\omega^2}{\mathbf{k}^2} = v^2$, so $v > c$ when $C < 0$.

[41] Because physicists usually define causality in terms of the (in)ability to send *signals* faster than c, I have opted to follow Bonvin *et al* in using the designation "CSC" for "closed signal curve" throughout the present work [60]. A CSC is defined as a "closed curve along which a signal propagates", so that it could be either a closed timelike curve or a closed null curve. An alternative phrase also used in the literature is "closed causal curve" or "CCC" ([58, 59], [15, 16]). Others

In order to see how this comes about, let us recall why it is generally believed that superluminal signaling would lead to a causal paradox.

2.5.1 Causal paradox: the bilking argument and its physical implementation

First let us clarify what is meant by "causal paradox." What is at stake is the situation described by the so-called "bilking argument", proposed in 1956 by Max Black as a *reductio ad absurdum* argument against backward causation. The famous "grand-father paradox" is an instantiation of it. The argument goes as follows:

> Imagine B to be earlier than A, and let B be the alleged effect of A. Thus we assume that A causes B even though A is later than B. The idea behind the bilking argument is that whenever B has occurred, it is possible, in principle, to intervene in the course of events and prohibit A from occurring. But if this is the case, A cannot be the cause of B; hence, we cannot have backward causation [61].

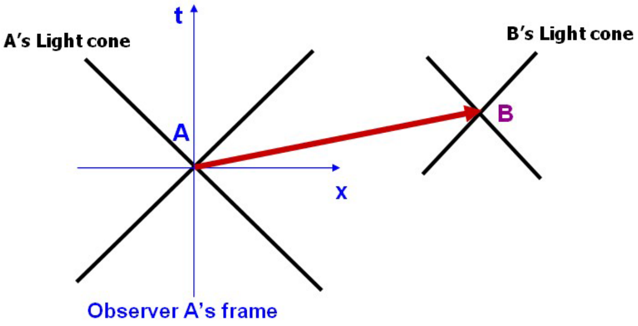
One response to this argument is to deny the possibility "to intervene in the course of events and prohibit A from occurring" by imposing constraints so that the history described be self-consistent. However, this is not what we are going to be concerned with here. Instead, we shall focus on whether the superluminal effects discussed above could lead to a situation whereby an event A (the emission of a signal) causes an event C (the reception of a second, different signal) that is earlier than A.

The way to physically instantiate this situation is based on a thought experiment illustrated below (Fig. 2.1), which has been called the "tachyonic anti-telephone."[42]

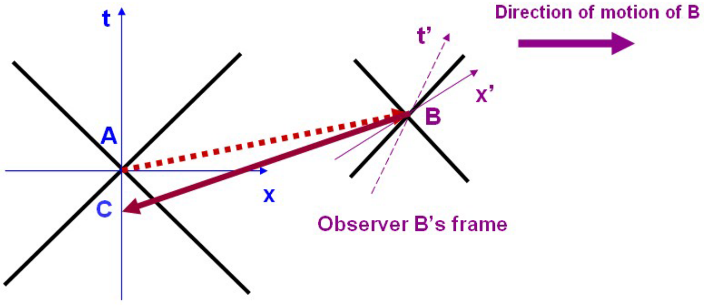
---

simply use the terminology familiar from general relativity, "closed timelike curve" or "CTC", and "closed null curve" (notably [15, 16], [37, 38], [54, 55].)

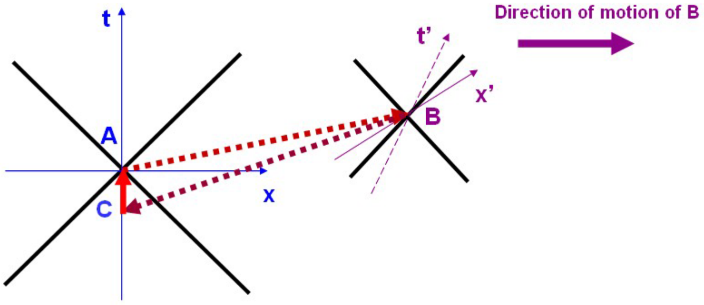
[42] Following Liberati *et al*, I use the term "tachyon" in the general sense of "particles or signals faster than" c ([15], p.176, [16] p.10, without requirement that their mass should be imaginary. Note that Babichev *et al*, whose work I discuss below, use the definition of "superluminal hypothetical particles possessing unbounded velocity $c_{tachyon} > 1$" where 1 is the speed of light ([58, 59] p.7).



(a) Signal propagates from A to B faster than the speed of light.



(b) A superluminal signal emitted by B can propagate forward in time to C. Note that it propagates into B's coordinate future.



(c) Seen from A's reference frame the second signal arrives earlier than the first one was sent.

Figure 2.1: The "tachyonic anti-telephone" thought experiment.

Consider two observers A and B who are spacelike separated, and let A send to B a signal that travels faster than c, hence propagates in A's spacelike region (Fig. 2.1a). Then let B, upon receiving this first signal, send a second signal towards A, also superluminal (so that it propagates in B's spacelike region) (Fig. 2.1b). The situation of interest occurs when B is moving away from A so fast that the x' axis in B's rest frame tilts "below" the first signal. In that scenario, although the second signal propagates towards the future in B's frame, it moves towards the past seen from A's frame. In this case the second signal reaches A before A sent the first, *i.e.* a CSC is formed (Fig. 2.1c). Clearly the reason why this can occur is because the signal is allowed to propagate towards the past in A's frame; if instead we could somehow constrain it to always move towards the future as seen by A, no CSC could form. This is how the concept of causal cone is thought to avoid CSCs.

2.5.2 Causal cone

The concept of "causal cone" can be defined from the effective metric, to represent the region of spacetime that can be reached by the propagation at stake – say, k-essence fluctuations, Drummond-Hathrell or Scharnhorst photons just to mention those we have been discussing.[43] A causal cone constitutes a well-defined hypersurface in spacetime, independent of the speed of the emitter. This hypersurface may look different (notably, tilted) to observers in different reference frames, but it constitutes an objective set of spacetime points (Fig. 2.2).

---

[43] Most authors use the phrase "causal cone" ([62], [63]), [64], [54], [37], [38]; however some refer to it as the "influence cone" ([58], [59]), the "characteristic cone" ([60]), or the "acoustic cone" ([58], [59]). The concept has especially been used in discussions relating to possible instances of superluminal propagation, but it applies to their less exotic counterparts as well (as the phrase "acoustic cone" suggests).

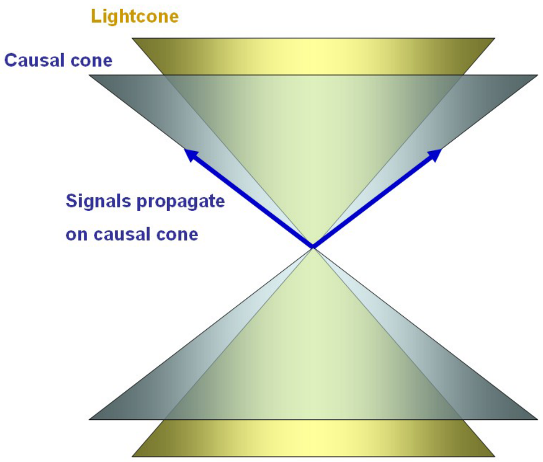


Figure 2.2: The causal cone *vs* the light cone

To see this, let us take another look at the case discussed in the previous section, and consider the equation of motion for k-essence fluctuations π, Eq. (2.17):

$$\tilde{G}^{\mu\nu}\nabla_\mu\nabla_\nu\pi + V^\mu\nabla_\mu\pi + \tilde{M}^2\pi = \delta J, \tag{2.22}$$

where, recall, $\tilde{G}^{\mu\nu}$ is the effective metric. This time, let us substitute in it an expression for π in terms of the phase of the fluctuations:

$$\pi(x) = A(x)\ e^{i\,\omega S(x)}, \tag{2.23}$$

where $S(x)$ is the hypersurface of constant phase or "characteristic surface", *i.e.* the set of all points of the same phase. This corresponds to the eikonal approximation, which is valid for high frequencies, and makes it possible to relate a description in terms of waves to a description in terms of rays. This leads to the following equation for S in the $\omega \to \infty$ limit:

$$\tilde{G}^{\mu\nu}\ \partial_\mu S\ \partial_\nu S = 0. \tag{2.24}$$

From the latter, one can derive an equation for the causal cones:

$$\tilde{G}^{-1}_{\mu\nu}\ N^{\mu}\ N^{\nu} = 0, \tag{2.25}$$

where N^{μ} and N^{ν} are the propagation vectors, and such that $N^{\mu}\ \partial_{\mu} S = 0$. Notice that Eq. (2.25) has the same form as the equation for a null (propagation) vector k^{μ}:

$$g^{-1}_{\mu\nu}\ k^{\mu} k^{\nu} = g_{\mu\nu}\ k^{\mu} k^{\nu} = 0, \tag{2.26}$$

but with the spacetime metric $g_{\mu\nu}$ replaced by the inverse effective metric $\tilde{G}^{-1}_{\mu\nu}$. By analogy, N^{μ} are deemed to be null with respect to the effective metric.

Again, the crucial point is that the causal cones are well-defined, objective hypersurfaces that divide the spacetime into observer-independent regions, *i.e.* into the same sets of events for all observers. The claim is that one can justifiably define causality with respect to these causal cones, by using them instead of (the usual) light cones to separate spacetime into different regions: the events inside a causal cone are said to form a region timelike with respect to the effective metric, while the region outside the cone is spacelike with respect to this metric (see Fig. 2.3).

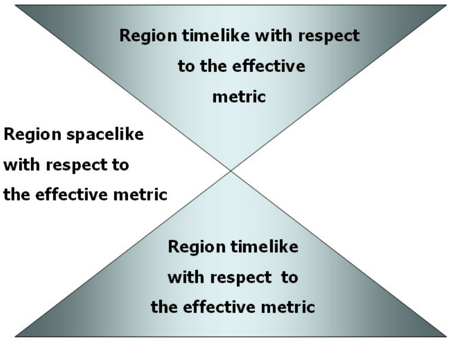


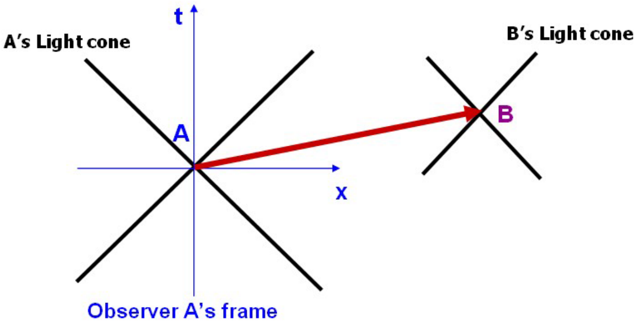
Figure 2.3: Causal regions defined with respect to the causal cone.

When the causal cone is wider than the light cone (Fig. 2.2), propagation is spacelike with respect to the light cone, with respect to the causal cone it is null-like.

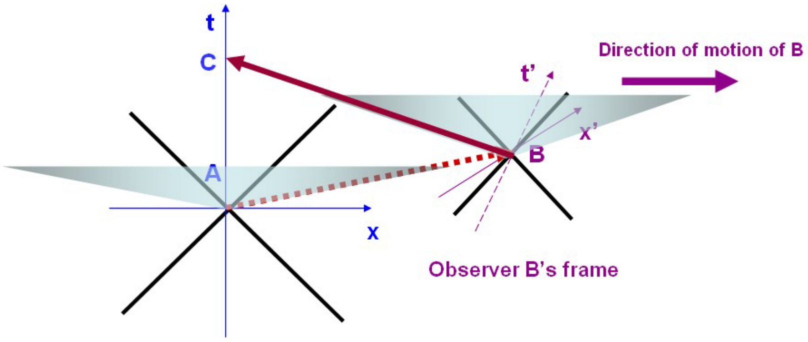
2.5.3 The tachyonic anti-telephone thought experiment modified

Because signal propagation is constrained to take place on the causal cone, the tachyonic anti-telephone thought experiment no longer leads to a closed signal curve as shown above in Fig. 2.1. What occurs instead is shown below on Fig. 2.4. Indeed the second signal as well is now constrained to move into the coordinate future of A (Fig. 2.4b), so it can only be received by A *after* the latter emitted the first signal (Fig. 2.4c). That is, a CSC cannot form – hence no bilking paradox occurs.

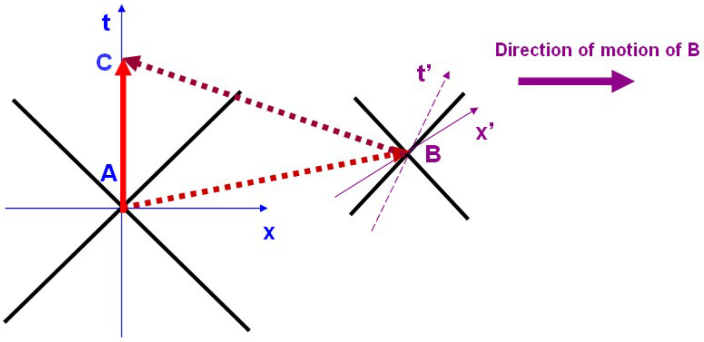
Let us contemplate what this implies for the concept of preferred reference frame. In the classic version of the thought experiment, the reason why B's signal is allowed to propagate the way it does is this: once we no longer use the light cone (which we cannot if we wish to consider superluminal signals), the only sensible concepts of past and present available are coordinate ones, *i.e.* they are frame-dependent. Yet relativity tells us our description should not be. Because the coordinate time of the emitter seems the most "natural" (not to say intuitive), choosing that one – *i.e.* the rest frame of the emitter – over all others seems as close to favoring no frame as one can hope to get. However, this means that whether or not a signal can reach an event depends only on the speed $v_{B\,wrt\,C}$ of the emitter with respect to the rest frame of the recipient – here the rest frame C hence A; and nothing prevents $v_{B\,wrt\,C}$ from being so fast that x' "swings below" a signal connecting B to C in Fig. 2.1b, because this situation can occur with $v_{B\,wrt\,C} < c$.



(a) Signal propagates from A to B faster than the speed of light. Seen from A's rest frame.



(b) A second superluminal signal propagates to C, forward by both A and B's coordinate times (when B is still moving away from A). Seen from A's rest frame.



(c) Seen from A's reference frame the second signal arrives earlier than the first one was sent.

Figure 2.4: The "tachyonic anti-telephone" thought experiment modified: propagation is constrained to take place on causal cones.

Now, in the modified version of the thought experiment, B can still move at any $v_{B\,wrt\,C} < c$, but this no longer matters because whether or not a signal can reach an event no longer depends on the speed of the emitter. However, the causal cones that constrain its motion and the speed they represent cannot be frame-invariant.[44] Indeed, this invariance is associated with the Lorentz transformations. The latter can only involve one speed, and c has been experimentally determined to be invariant, so it must be the one.[45] As a result the causal cones, unlike the light cones, look different in different reference frames. The signal looks faster to an observer moving in the direction opposite the propagation, and slower to an observer moving in the same direction, so that the causal cones look "tilted" differently to different observers.[46] They only look "straight" from one frame. So the situation described by Fig. 2.4 can only arise in theories in which a preferred reference frame can be defined. For example in k-essence, this is the rest frame of the background, and for the Scharnhorst effect, that of the Casimir plates, or equivalently of the non-trivial Casimir vacuum.

2.5.4 Stable causality

We just established that the classic thought experiment used to argue that superluminal propagation would lead to CSCs no longer leads to this conclusion when propagation is constrained on a causal cone, itself defined by an effective metric. In fact the latter provides a much more general (and formal) way to guarantee the absence of CSCs. Indeed, it allows us to avail ourselves of concepts and theorems

---

[44] At least so long as we use a description where Lorentz symmetry transformations involve c. See [15], [16], Section 2.

[45] Again, this reasoning needs to be qualified, because it has been argued that this is merely the result of using rods and clocks based on electromagnetic interactions to define our frames with. See notably [15] p.274, [16] p.9 among the works discussed in this chapter. More generally, see [65] chap.9, [66], [67].

[46] No conventional observer can move as fast as the superluminal signal of course.

first developed for that purpose in the context of general relativity. The required concept is that of stable causality:

> A spacetime is said to be stably causal if and only if it possesses a Lorentzian metric $g_{\mu\nu}$ and a globally defined scalar function t such that $\nabla_\mu t$ is everywhere nonzero and timelike with respect to $g_{\mu\nu}$.[47]

and the relevant theorem is the following:

> A stably causal spacetime possesses no closed timelike curves and no closed null curves.[48]

Note that these require that we reason in terms of spacetime and the metric that characterizes it.

Several researchers discussing theories that involve superluminal propagation have made use of this theorem beyond the realm of general relativity, applying it to a $g_{\mu\nu}$ that is not the gravitational metric of the latter, but the effective metric that defines the causal cones. Notably, Babichev *et al* did so to treat the case of k-essence, both in a Minkowski spacetime and a Friedmann universe.[49]

Of particular interest to the present research, it was also applied to the Scharnhorst effect. The relevant metric in this case is the one formed of the effective metric between the Casimir plates (notably discussed by Liberati *et al* in [14], [13]), and the Minkowski metric on each side of the plates. With this metric, one has to look no

---

[47] [68], p.198, discussed by [15] p.180.

[48] [68], p.199.

[49] [58] pp.10-12, [59] pp.9-11.

See as well Adams *et al.* ([37] p.10, [38] pp.8-9). They actually do not state the theorem used in this section, and even seem to suggest that the existence of a global time function is not a sufficient condition to exclude CSCs, in so far that they speak of the two as separate conditions: “A first requirement is that the spacetime be time orientable, meaning that there should exist a globally defined and non-degenerate timelike vector t^μ [...] The second condition for causality to hold is that there be no closed (future directed) timelike curves (CTCs).” However they continue “In the presence of CTCs, the t coordinate is not globally defined”, thereby stating that the existence of a global time function is a necessary condition for the absence of CSCs – albeit still not implying it is a sufficient one.

further than the coordinate time in the rest frame of the plates to find a global time function t: with this choice $\nabla_\mu t$ is "everywhere nonzero and timelike with respect to" the metric we just defined. The theorem then ensures that no CSCs can form in the spacetime it describes.

2.5.5 New difficulties: propagation into the coordinate past

Although in the situations involving superluminal propagation discussed so far, no CSC can be formed, another problem arises, precisely because the motion of the signal no longer depends on the speed of the emitter.

To return to the now familiar example of the tachyonic anti-telephone thought experiment, this time the difficulty comes about when B moves *towards* A at high enough speed – that is, moves in the same direction as the signal it is emitting.[50] We still do not get a CSC, however the signal looks like it is propagating towards the past seen from its emitter's (*i.e.* B's) reference frame:

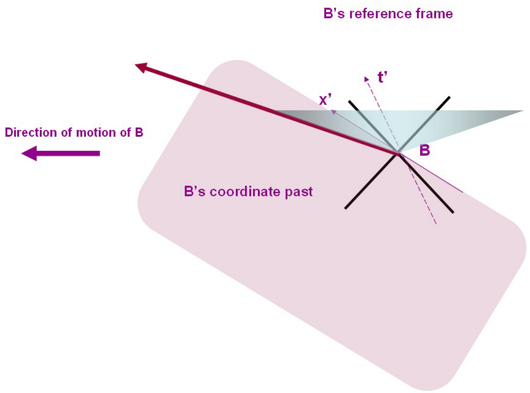


Figure 2.5: Signal propagating into B's coordinate past.

While it is null-like with respect to the causal cone, the signal is spacelike with respect to the light cone; and although we have now defined causality with respect to

---

[50] At $v_B > \frac{c}{c_s}$, where c_s is the speed of the signal with respect to the background.

the causal cone, boosting the frames of observers still involves a Lorentz transformation with the speed of light as the invariant speed, so graphically it is with respect to the light cone that the coordinate axes "scissor" when one boosts. Consequently if the boost is large enough the x' axis gets above the causal cone. As a result, one can find frames in which the signal propagates towards the coordinate past. If the emitter moves in such a way that such a frame is its reference frame, then the signal propagates towards its past in that frame.

Another way to represent this same situation, following Bonvin *et al*, is to place oneself in B's rest frame. Then the causal cone appears deformed and tilted compared to the way it looks in A's frame:[51]

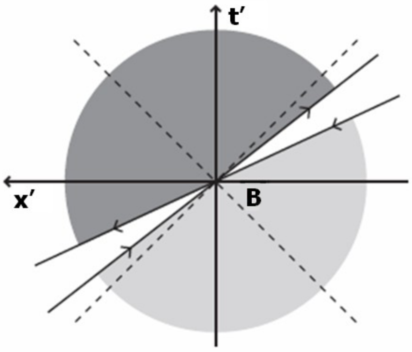


Figure 2.6: causal cone seen from the frame of an emitter moving at high speed in the same direction as the signal.

On this diagram the dotted lines represent the light cone, the light grey region is the inside of the past causal cone, and the dark grey one the future causal cone. Note that this latter region "tips" below the x' coordinate axis: some of it lies in B's past (to the left of the figure).

Again, we see that if B sends a signal to its right, opposite its direction of motion,

---

[51] Picture taken from Bonvin *et al*, *i.e.* [60], p.6.

this signal necessarily propagates into the future; but if it sends one to its left, the signal can propagate towards its coordinate past:

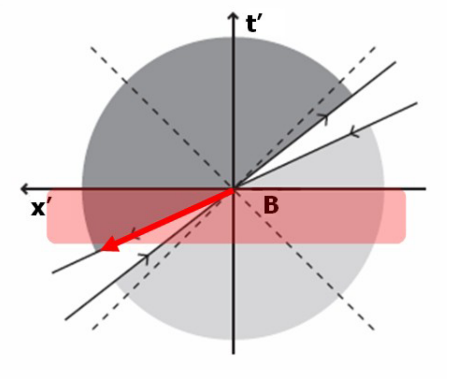


Figure 2.7: A signal emitted to the left propagates into B's coordinate past.

Similarly, part of the past causal cone lies above the x' axis, so in B's future (to the right in Fig. 2.6); so some signals can reach B from its right, that emanated in its coordinate future.

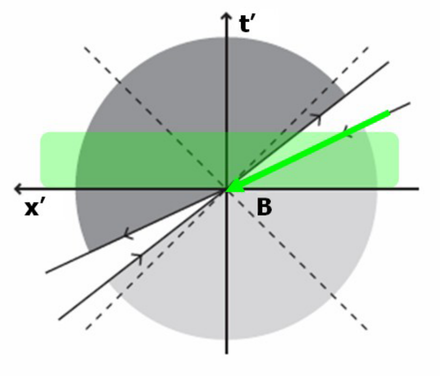


Figure 2.8: A signal arriving from the right emanates from B's coordinate future.

However, and at the risk of belabouring this point, these signals cannot lead to a grand-father paradox type predicament. In the diagrams just discussed, this can be seen from the fact that the past and future causal cones do not overlap. Another and arguably more concrete way to see one cannot find oneself in a grand-father paradox predicament is to consider the following situation:

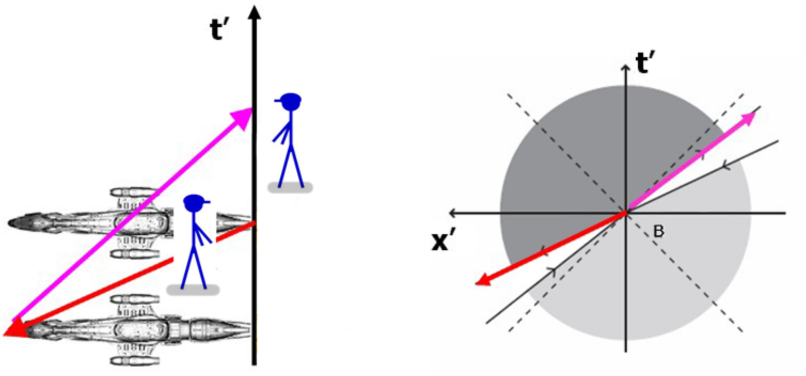


Figure 2.9: A signal arriving from the right emanates from B's coordinate future.

The spaceship depicted is travelling to the left at high enough speed for the causal cones to look as shown on the right of Fig. 2.9 in its rest frame. A superluminal signal is emitted from the back of a spaceship, and propagates towards the front of the ship. That it propagates into the coordinate past means that in its rest frame, it reaches the front before it has left the rear. However an observer at the back of the ship does not actually "see" it reach the front until later – even if the information that the signal has reached the front is transmitted to the observer at superluminal speed as well, and *a fortiori* if it were by light. According to his time, while he is at rest within the ship, he first sees the signal leave the back of the ship, and it is later that he sees it reach the front. Therefore, he can never find himself in a grand-father paradox type of predicament – and no additional consistency constraints are required to ensure that he does not prevent the departure of the signal after he sees it reach the front of the ship.

Note that these considerations are merely further ways to illustrate what the theorem discussed above guarantees: so long as we can define a global time function with

$\nabla_\mu t$, stable causality holds.

Nevertheless, it does remain that in the rest frame of the fast moving ship, the signal would travel towards the past. Considering what a signal emitted by someone at rest with respect to a k-essence background would look like to someone in such a ship, Adams *et al* explain:

> The point is that what look like perfectly natural initial conditions for a superluminal mode in one frame look like horribly fine-tuned conditions in another. [...] initial conditions that to one observer look like turning on a localized source at some unremarkable point in spacetime will appear to the other as a bewildering array of fluctuations incident from past infinity which conspire miraculously to annihilate what the original observer wanted to call the localized source.[52]

Adams *et al* and Babichev *et al* discuss what the solutions to the equations of motion would be. In terms of Green's functions, the propagation into coordinate past is expressed by the fact that what is a retarded Green's function in the rest frame of the k-essence background, and for an observed moving at $v < \frac{c}{c_s}$, becomes a sum of retarded and advanced Green's functions seen from the fast moving spacecraft – recall that advanced Green's functions represent propagation towards the future, and advanced ones towards the past.[53] This result is obtained by considering the equation of motion for the k-essence fluctuations, in the rest frame of the background:

$$\partial_t^2 \pi - c_s^2 \Delta_x \pi = \frac{c_s^2}{\mathcal{L}_{,X}} \delta J, \tag{2.27}$$

where δJ is an external source. Babichev *et al* then find its Green's function, and transform it by a Lorentz boost.[54] This yields what a retarded Green's function in

---

[52] [37] p.7, [38], p.6.

[53] Babichev *et al* have mostly k-essence in mind, and Adams *et al* simply a "theory of a single U(1) gauge field." [58, 59, 37, 38].

[54] In addition its Fourier Transform is taken for the resulting expression to be a function of frequency rather than spatial coordinates.

the rest frame of the background would look like to an observer moving with respect to it.

Babichev *et al* also compare these findings to what is obtained when first boosting Eq. (2.27), and then finding the Green's function of the result. In that case, one gets the same expression as in the former calculation for the "slow" moving observer (*i.e.* $v < \frac{c}{c_s}$).

Meanwhile, the analysis performed by Babichev *et al* brings to light a second issue, which Adams *et al* also emphasize. Babichev *et al* perform both sets of calculations in a two-dimensional and in a four-dimensional spacetimes. In the latter case, for the "fast-moving spacecraft", they find that the result of the second derivation (which boosts the equation of motion first) is unstable in the sense that it contains exponentially growing modes.[55] They dismiss this result as unphysical:

> Physically this means that we have failed to find the Green's function, which describes the propagation of the signal which the source δJ in the fast moving spacecraft tries to send in the direction of growing t'. Instead, the response to any source in the spacecraft is always driven by ... the Lorentz transformed Green's function in the rest frame.[56]

On this particular point Adams *et al* can be said to agree:

> the resulting configuration does *not* look like a small fluctuation sourced by a local source – indeed, these are explicitly stable according to the equation of motion – but rather involves turning on initial conditions at a fixed time which vary exponentially in *space*, *along* the slice. These do not represent instabilities in any usual sense; they simply represent initial conditions which we would normally rule out as unphysical.[57]

---

[55] [58], pp.16-18, [59], pp.13-14. See also [37], pp.6-8, [38], pp.6-7.

[56] [58], pp.17-18, [59], p.14.

[57] [37] p.7, [38], pp.6-7.

However they disagree on the take home message: generally speaking, Adams *et al* argue that the difficulties encountered with theories involving superluminal fluctuations should make one very weary of taking them seriously. Babichev *et al* on the other hand seem to interpret the second type of derivation (where the equation of motion itself is boosted) as telling us something about whether someone (or something) moving at $v > \frac{c}{c_s}$ with respect to the background would be able to emit fluctuations. In contrast, boosting the Green's function obtained from the rest-frame equation of motion tells us about what a signal already propagating through the background looks like to an observer moving at such high speed. Then as far as Babichev *et al* are concerned, the fact that the former derivation yields an unphysical-looking solution implies that a fast moving entity cannot emit fluctuations described by this solution.

However the fact remains that "the Lorentz transformed Green's function in the rest frame" involves both advanced and retarded Green's functions. Again, for some researchers this serves to make the theories at stake untrustworthy.[58] Others argue that it is not really problematic.[59] Notably, Babichev *et al* think the matter reflects an unfortunate choice of formalism rather than an actual physical effect:

> One should remember [...] that the notion of past and future is determined by the past and future cones in the spacetime and has nothing to do with a particular choice of coordinates. Thus, the signals, which are future-directed in the rest-frame remain future-directed also in a fast-moving spacecraft, in spite of the fact that this would correspond to the decreasing time coordinate t'. [...] the confusion arises because of a poor choice of coordinates, when decreasing t' correspond to future-directed signals and vice versa.[60]

---

[58] In particular, see Adams *et al* and Bonvin *et al.* [37, 38, 60].

[59] See notably [54, 55, 58, 59, 62, 63].

[60] [58], p.8, [59], p.8.

and further:

> Because we cannot send a signal in the direction of growing t' one cannot associate growing t' with the arrow of time contrary to the claims in [30]. We would like to emphasize that the correctly defined causal Green functions should guarantee the propagation in the future influence cone, (which is defined geometrically) instead of propagation forward in a time coordinate.[61]

Of course, it remains that for a fast-moving observer, his/her rest-frame coordinates so happen to be such a "poor choice." Why such an argument can nevertheless be made has to do with the initial data formulation and the Cauchy problem, to which we now turn.

2.5.6 Initial data formulation and Cauchy problem

Given equations of motion for a physical system, and some initial data, the solutions are unique provided that we have a well-posed initial value formulation (or well-posed "Cauchy problem"). Then the evolution described by these solutions is self-consistent in the sense that it cannot change the initial data. A sufficient (albeit not necessary) condition for the initial value problem to be well posed is for the spacetime to be globally hyperbolic, that is to say that it must possess a spacelike slice (Σ) known as a Cauchy surface, whose domain of dependence $D(\Sigma)$ is the entire spacetime.[62] Initial data can then be specified freely on this surface. This can be represented graphically in terms of causal cones: each lobe of the cone needs to be on one side of the slice, as depicted on the left of Fig. 2.10 – recall that the evolution is constrained to take place within the causal cones. In other words, the surface needs to be spacelike with respect to the causal cones: then the lobe that constitutes the causal future lies in the future of the coordinate t (where Σ is

---

[61] [58], pp.17-18, [59], p.14.

[62] [22], [23] p.12. This also guarantees that it satisfies stable causality.

defined as the set of points satisfying $t = t_0$).[63]

Recall from Fig. 2.5 (at the start of the previous section) that the appearance of advanced Green's functions arose because the rest frame of an observer moving with $v > \frac{c}{c_s}$ has its coordinate x' axis "above" the signal. By definition, this axis lies on a hypersurface $t' = constant$. So the situation at hand corresponds to what is illustrated on the right hand side of Fig. 2.10: seen by the fast-moving observer, the causal cone has dipped below this hypersurface.

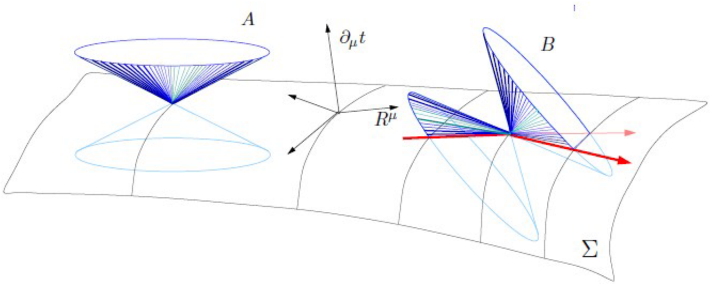
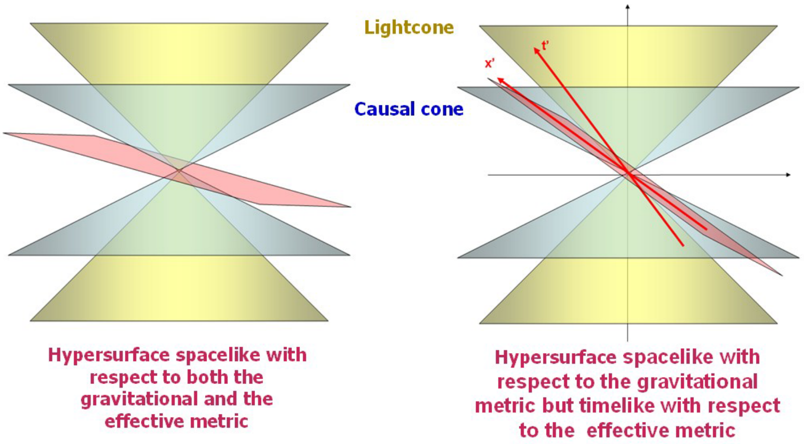


Figure 2.10: Well-posed *vs* ill-posed initial value formulation.

Seen from the rest frame of the background, the situation would look as follows (where we have included the light cone for the sake of completeness, and to illustrate that the hypersurface is at all events spacelike with respect to it):

---

[63] Fig. 2.10 is taken from [58] p.14,[59] p.12.



Hypersurface spacelike with respect to both the gravitational and the effective metric

Hypersurface spacelike with respect to the gravitational metric but timelike with respect to the effective metric

Figure 2.11: Well-posed *vs* ill-posed initial value formulation.

What the situation depicted on the right of Figs.2.10 & 2.11 implies is that if we attempt to set-up the initial configuration on that surface, the initial value problem is ill-posed. This has been interpreted to mean that someone moving at $v > \frac{c}{c_s}$ would not be able to affect the field values at will on Σ: "not everything is in the hand of the astronaut: he has no complete freedom in the choice of the initial field configuration." If he/she has any, it is within the limits that "not all possible configurations are admissible on this hypersurface but only those which could be obtained as a result of evolution of some initial configuration chosen on the hypersurface which is simultaneously spacelike with respect to both metrics $g_{\mu\nu}$ and $G^{-1}_{\mu\nu}$."[64] Babichev *et al* also integrate in this view the hitherto offending advanced Green's function:

> If the astronaut disturbs the background with some device (source function δJ) which he/she switches off at the moment of time t'_1, then the resulting configuration of the field on the hypersurface $t'_1 = const$ [...] will always satisfy the conditions needed for unambiguous prediction of the field configuration everywhere in the spacetime irrespective of the

---

[64] [58] p.18, [59] p.15.

> source δJ. The presence of the advanced mode in this Greens function plays an important role in obtaining a consistent field configuration on $t'_1 = const.$[65]

Babichev *et al* are not alone in these views. Notably, Jean-Philippe Bruneton offered a discussion of these issues that is very similar in spirit. He too is motivated by an interest in k-essence, as well as in MOND, and as the following passage illustrates, his conclusions are very much in line with those of Babichev *et al*:

> The main argument which rules out superluminal propagation is indeed that the Cauchy problem for the scalar field is not well posed for initial data that are set on surfaces which are spacelike with respect to the flat metric but timelike or null with respect to the background metric. In that case, initial data cannot be evolved because of caustics, and the Hamiltonian formalism is singular.
>
> These difficulties are however not surprising. Indeed, this only shows that initial data surfaces for the scalar field must be spacelike *with respect to the background metric*, as stated by Leray's theorem. If initial data are set on these surfaces, the theory is free of such a singular behavior. This claim therefore only arises from an unadapted choice of initial data surfaces, *i.e.*, from the postulate that the gravitational metric defines a preferred chronology. This postulate actually prevents *any* superluminal propagation. [...] In such a case, superluminal behavior is ruled out from the beginning in a rather *ad hoc* way, but not by some intrinsic (mathematical) argument.[66]

Bruneton also discusses in general terms the case of a theory involving several fields that may interact with one another, and whose propagation is constrained by different causal cones (some of which may be the light cone). He even allows for these

---

[65] *Idem.*

[66] [54] p.7, [55] pp.7-8.

cones to have any orientation with respect to one another. In such a situation causality would be defined with respect to the region illustrated below (and which could be described as an overall causal cone):[67]

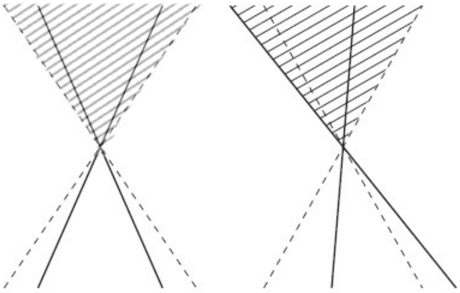


Figure 2.12: “The hatched part shows the extended future defined by two metrics (solid and dashed lines) in the case where one metric defines a wider cone than the other one (left) and vice versa (right).”

In this respect, Bruneton’s work prefigures Robert Geroch’s. Indeed the latter also offers a general discussion of interacting fields with arbitrary causal cones, and reaches the analogous conclusion that “the causal cone of the combined system is the convex hull of the causal cones of the two individual systems.”[68] In order to discuss the issues we have just treated, he develops an abstract formalism based on the notion of fiber bundles, which we shall now describe.

2.5.7 Robert Geroch’s work

Robert Geroch has offered a theoretical treatment of the implications of superluminality for causality. His approach consists in analysing the structure of physical equations using the concept of fiber bundles. The paper in which he presents this work, “Faster-than-light?”, was published in 2010.[69] However the formalism on which it relies long predated it (notably, the concept of hyperbolization), and had

---

[67] Figure taken from [54] p.4, [55] pp.7-8.

[68] [63], p.9.

[69] [62] pp.59-70, [63].

not been motivated by an interest in causal paradoxes. Geroch developed it for the purpose of studying properties of physical equations, and in the paper where he presents this work in great detail in 1996, nothing hinted at the considerations he later explored in “Faster-than-light?” [69, 70]. In 2010 he extended his work notably by defining the concept of causal cone within his formalism, which we now present.

Geroch argues that the behavior of nearly all systems in physics can be represented by systems of first order quasilinear partial differential equations. Concretely, almost any physical system can be described by equations which can be put in the form:

$$k^{Aa}{}_{\alpha}\left(\nabla_a \phi^{\alpha}\right) = j^A, \tag{2.28}$$

where ϕ^{α} are the fields of the system, α refers to the kind of field, and j^A is a current-source term. “A” indexes each equation in the system, *i.e.* there is one equation for each value of A. $k^{Aa}{}_{\alpha}$ and j^A are smooth functions of the fields ϕ^{α}, but not of their derivatives.[70] For example a physical system described by electromagnetism obeys:

$$\nabla^b F_{ab} = 0, \tag{2.29}$$

$$\nabla_a F_{bc} = 0, \tag{2.30}$$

where F_{ab} corresponds to the field ϕ^{α} in Eq. (2.28).

Second order equations can be re-expressed as first order by defining the derivative of the relevant field as a distinct field (and something analogous holds for still higher-order equations). This definition then constitutes an additional equation in the system Eq. (2.28). For instance a physical system represented by a field ψ that obeys $\nabla_a v_b = 0$ as well as the wave equation $\nabla^a \nabla_a \psi = 0$ is described in this

---

[70] [62] pp.60-61, [63] p.3.

formalism by the system of equations:

$$\begin{aligned} \nabla_a v_b &= 0 & (2.31a) \\ v_a &= \nabla_a \psi & (2.31b) \\ \nabla^a v_a &= 0. & (2.31c) \end{aligned}$$

Similarly, Einstein's equations can take the form:

$$\begin{aligned} \nabla_a\, g_{bc} &= 0 & (2.32a) \\ G_{ab} &= 0 & (2.32b) \end{aligned}$$

where the fields are the spacetime metric g_{ab} and the derivative operator ∇_a, and G_{ab} is the Einstein tensor.[71]

Therefore what Geroch means by "field" is not just what one would mean by this term in QFT. For instance he takes as fields the metric g_{ab}, and also the derivative operator ∇_a, on M.

A central notion of Geroch's formalism is that of fiber bundle. All fields ϕ^α are defined on it, that is, to each point of the 4-D manifold (M) which constitutes the base space, corresponds a fiber on which all the relevant fields are specified.[72] The actual field configuration of a system is then represented as a smooth cross-section on the fiber bundle.[73]

Geroch's formalism therefore tends to put all fields on the same footing.[74] Conceptually, this is a crucial feature, notably because it is ground for putting all causal

---

[71] [62] p.62, [63] p.4.

[72] The index a refers to this base manifold, with respect to which the derivative is taken (∇_a)....

[73] [69] pp.20, 21-25, [70] pp.3, 4-9; [62] p.21, [63] p.3.

[74] Admittedly this is not *entirely* true, in so far that in his 1996 paper, Geroch defines a special class of fields, *i.e.* background fields: "One field is a background for another if the former appears algebraically in the derivative-terms of the latter", [69] p.21, [70] p.4. However he does not make use of this concept in his work on causality, in 2010. There he notably stresses that the metric can be viewed as a field like any other.

cones on the same footing as well.

A key notion in Geroch's work on causality is that of "hyperbolization" of the system of equations.[75] Its importance is related to the issue we discussed in the previous section: it guarantees a well-posed initial value formulation:

> A hyperbolization is a casting of the system of equations (or, commonly, a subsystem of that system) into what is called symmetric, hyperbolic form. To such a form there is applicable a general theorem on existence and uniqueness of solutions. This is the initial-value formulation.[76]

A hyperbolization $h_{A\beta}$ is a tensor that characterizes an entire system of (first order quasilinear) partial differential equations. This tensor is defined by the fact that it satisfies two properties:

1) $h_{A\beta}\, k^{Aa}{}_{\alpha}$ must be symmetric in α and β.[77]

2) there must exist in the manifold M a covector n_a such that the symmetric tensor $n_a h_{A\beta} k^{Aa}{}_{\alpha}$ is positive-definite.

Geroch then defines the concept of causal cone, definition which requires the hyperbolization $h_{A\beta}$. In his formalism a causal cone is associated with the system of equations itself (Eq. (2.28)), *i.e.* it is the causal cone *of the system*, at a point p of the manifold M, and for specific field values ϕ^{α} at that point. A causal cone $\mathcal{C}$ is then defined as "the set of all tangent vectors ξ^a at $p \in M$, such that $\xi^a\, n_a > 0$ for every n_a for which $n_a\, h_{A\beta}\, k^{Aa}{}_{\alpha}$ is positive-definite." More intuitively, $\mathcal{C}$ is a "nonempty, open, convex cone of tangent vectors at the point p of M", with respect to which we can define causality:[78]

> The causal cones deserve their name. Roughly speaking, any first-order,

---

[75] But recall that "Hyperbolization" is a concept which he had also already introduced in 1996. [69] pp.20, 25-28, [70] pp.3, 9-13; [62], [63] pp.6-9.

[76] [69] pp.20, [70] p.3.

[77] Recall that the tensor k^{Aa} is defined in Eq. (2.28), *i.e.* the system of first order equations describing the physical situation: $k^{Aa}{}_{\alpha}\ (\nabla_a \phi^{\alpha}) = j^A$.

[78] [63], p.6.

> quasilinear system is capable of sending signals only within its causal cones.[79]

In many cases (he mentions Maxwell's equations, the Klein-Gordon equation, the neutrino or Dirac equations, the spin-s field equation and Einstein's equation) the causal cone is none other than the future light cone. For a (normal, classical) fluid it is the sound cone.[80] He stresses that "all of these remarks about fluids – the existence of a hyperbolization and of the causal cone – apply equally well in the case of a subluminal and a superluminal sound speed."[81] So we see that Geroch's concept of causal cone coincides with the one discussed previously.

We also recognize the condition required to have a well-posed Cauchy problem: Geroch defines the "*initial data*" for a (first-order, quasilinear) system of partial differential equations as follows:

> Fix a first-order, quasilinear system of partial differential equations, together with a hyperbolization, $h_{A\beta}$, for that system. Let S be a 3-dimensional submanifold of the manifold M, and let there be given fields ϕ_o^α on S. We call this (S, ϕ_o) *initial data* for our system provided that, at each point of S, the closure of the causal cone at that point lies entirely on one side of S. [This means, in other words, that a normal n_a to S is such that $n_a \, h_{A\beta} \, k^{Aa}{}_\alpha$ is positive-definite].[82]

In other words, the slice on which initial data is defined must be spacelike with respect to the causal cone, as also argued by Babichev *et al* and Bruneton.

Geroch stresses several key-points. One is that causal cones (and the existence of the hyperbolization needed to define them) depend on the LHS of Eq. (2.28) (more specifically, on $k^{Aa}{}_\alpha$), but not j^A. In contrast, when we "switch on" interactions

---

[79] [63], p.7.

[80] *i.e.* "the set of directions in space-time whose speed, measured with respect to the fluid 4-velocity u^a, is less than the sound speed v". [63] p.7.

[81] [63] p.7.

[82] [63], p.7.

between different fields, this formally corresponds to modifying j^A. Consequently, neither the hyperbolization nor the causal cones are affected by "turning on" interactions between two systems. When we want to describe two interacting systems, causality is then simply defined by a causal cone for the combined system which is the "convex hull of the causal cones of the two individual systems — *i.e.*, the set of all sums of the form $\xi^a + \xi'^a$, with ξ^a in $\mathcal{C}$ and ξ'^a in $\mathcal{C}'$ ", where $\mathcal{C}$ and $\mathcal{C}'$ are the causal cones of each individual system.[83] This is in agreement with Bruneton's work.

Geroch himself sums up his findings as follows:

> Each first-order, quasilinear system of partial differential equations – provided that system has a hyperbolization – carries within itself its own initial-value formulation. And, as a part of that formulation, the system carries its own causal cones for signal propagation. These cones are inherent in the structure of the equation itself, *i.e.*, they do not necessarily require that there be fixed any outside fields. We may combine such systems – and turn on interactions between systems – and when we do so the causal cones also combine, in the way we expect physically.[84]

An important conceptual aspect of Geroch's formalism is that it puts all fields and cones on the same level:

> This formulation manifests what might be called a democracy of causal cones. All systems, and their cones, are on an equal footing: No one set of fields, or one set of causal cones, has priority over any others.[85]

In particular, the light cone (more accurately, the causal cone of special relativity) has no priority over other causal cones, including wider ones. It just so happens to be the causal cone for the system of equations describing many physical systems:

---

[83] [63], p.9.

[84] [63], p.9.

[85] *Idem.*

> Special relativity is merely one more first-order, quasilinear system of partial differential equations admitting a hyperbolization. It is just one more physical theory, not dissimilar from the theory of electromagnetism or the theory of a simple fluid. Like all such systems, special relativity carries with it, by virtue of the structure of its equations, causal cones. Some systems, such as that of electromagnetism, share those cones with special relativity; while other systems, such as that of a fluid, do not. But each system – special relativity included – looks to its own causal cones – to its own system of partial differential equations – for the propagation of signals within that theory.[86]

and further:

> [The traditional view according to which the causal cones of special-relativity have a preferred role in physics arises] from the fact that a number of other systems – electromagnetism, the spin-s fields, etc – employ precisely those same cones as their own. And, indeed, it may be the case that the physical world is organized around such a commonality of cones. On the other hand, it is entirely possible that there exist any number of other systems – not yet observed (or maybe they have been!) – that employ quite different sets of causal cones. And the cones of these other systems could very well lie outside the null cones of special relativity, *i.e.*, these systems could very well manifest superluminal signals. None of this would contradict our fundamental ideas about how physics is structured: An initial-value formulation, causal cones governing signals, etc.[87]

---

[86] [63], p.10.

[87] [63], p.10.

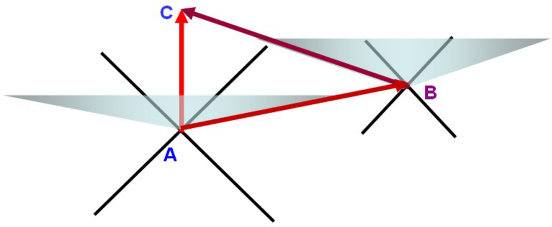
In addition, another important aspect of Geroch's approach is that it "avoids worries that the notion of field propagation he describes is coordinate or frame dependent."[88] Geroch notably applies these ideas to a type of situation that we have not yet considered, but which has been discussed in the literature as a potential source of causal paradoxes, in the form of CSCs. We now turn to these concerns before presenting Geroch's response.

2.5.8 Forming CSCs within a causal cone formalism

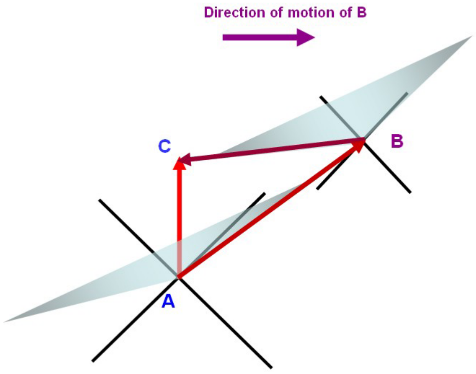
We showed above that signals constrained to move on causal cones can propagate superluminally without leading to a bilking / grand-father paradox. This is notably clear by comparing the traditional "tachyonic anti-telephone" thought experiment to the modified version thereof. Several researchers have nevertheless proposed ways to "recover" this paradox, as it were, even when propagation is so constrained. The basic idea is relatively straightforward: when discussing the modified version of the paradox, we made one assumption: that both signals (from A to B and then B to C), propagate according to causal cones that cannot tilt with respect to one another. As we then discussed, the same causal cones would look different to observers moving at different speed with respect to one another (and the background). As we showed, this, $perse$, does not permit to form a CSC. For the present discussion, let us illustrate the point in this way:

---

[88] [25], p.114.



(a) Seen from a frame at rest with respect to the background.



(b) Seen from a frame that moves with respect to the background.

Figure 2.13: The 2nd signal reaches point C after the first left point A. This holds whether or not the frame of the observer is at rest or in motion with respect to the background. It remains true no matter how fast the frame moves and the cones "tilt", because they do so to the same degree.

The 2nd signal reaches point C after the first left point A, no matter what frame we view the situation from, and, recall, no matter how A and B move with respect to one another and to the background, since the causal cones are objective regions of spacetime. The reason for this is that the boost affects the appearance of both causal cones in the same way, *i.e.* they "tilt" by the same amount. In contrast, if we recall the signal configuration that obtains in the classic form of the tachyonic anti-telephone thought experiment (see Fig. 2.1c), it becomes clear that for a CSC to form, the causal cones would need to "tilt" towards one another:

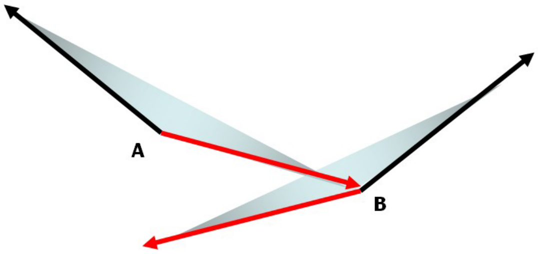


Figure 2.14: Cones configuration needed to form a CSC.

How can such a configuration be achieved? Well, how a causal cone "tilts" depends on how the frame it is viewed from is boosted with respect to the background. So for causal cones to "tilt" in different ways, the frame they are viewed from needs to have different speeds with respect to the background. Meanwhile, the causal cones of course need to be all viewed from one and the same frame. This is clearly impossible, so long as we have only one background. What is required is several backgrounds, moving at different speeds with respect to our frame – hence to one another. Then cones can appear tilted in different directions, and notably towards one another as required, just like the frame-independent light cones get tilted into forming CSCs by gravity in general relativity.[89]

How this can be implemented in practice depends on the physical context considered, and several have been discussed in the literature.[90] Notably, Adams *et al* illustrate how to form a CSC using "bubbles of non trivial vacua" moving past each other

---

[89] Note that this is essentially a way to get around the fact that signal propagation is no longer dependent on the speed of the emitter when constrained on a causal cone: the way the emitters move with respect to their respective background does not matter, but one gets the backgrounds themselves to move (with respect to one another and the observer).

[90] This has been discussed in different contexts. Dolgov & Novikov are mostly concerned with the Drummond-Hathrell effect ([71]) (as is Shore in his response to them ([33])), but they do treat Casimir plates as well. Adams *et al* use a Goldstone model which leads to "bubbles of non trivial vacua" ([37], p.10; [38], p.8). Liberati *et al* discuss Casimir plates. The conceptual issues involved as far as the possible formation of CSCs are concerned are largely similar in these different contexts.

fast enough for the cones to dip below the $t = const$ surface of the observer (that represents his/her coordinate present). Seen from the center of mass frame, the tilt of the causal cones would be symmetrical, so the situation would look as shown below:[91]

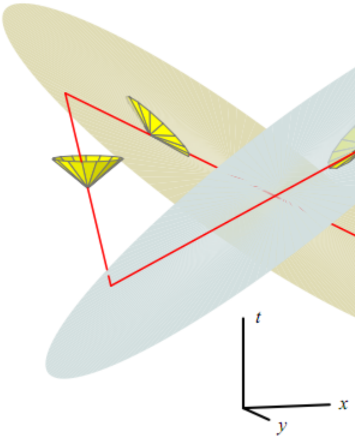


Figure 2.15: Two finite bubbles moving with large opposite velocities in the x-direction and separated by a finite distance in the y-direction.

Liberati *et al* have discussed an analogous situation created from the Scharnhorst effect, by moving two sets of Casimir plates past each other, and Geroch has considered instead two pipes filled with an hypothetical fluid of sound speed $c_s > c$. The possibility of forming CSCs in this way has inspired different responses. Adams *et al* view it as additional ground to be weary of theories that involve superluminal propagation.

Others argue that it may not be problematic either.[92] Liberati *et al* make one point in this regard which is specific to the Scharnhorst effect – of great interest to the present work.

---

[91] Figure taken from [37], p.9. The bubbles are said to be separated in the y-direction to avoid their passing through one another, which would complicate the analysis.

[92] See notably [15, 16, 62, 63, 22, 23].

Essentially, Liberati *et al* argue that one cannot have a meaningful discussion on the matter based on the actual effective metric and causal cones derived, because these results require for the Casimir vacuum to be exact. This can only be achieved by a single set of plates that need to be infinite in extent, which is incompatible with the type of thought experiments required to set up a CSC. Indeed we would have to consider either two vacua that interpenetrate as infinite set-ups pass each other, or half-infinite plates if we want to separate the vacua. In either case, the valid effective metric would not be the one they have derived, because the symmetries leading to it have been violated, and also, in addition, because of edge effect in the situation involving half-infinite plates.[93]

A more common type of argument consists in invoking chronology protection.[94] Liberati *et al* do so as well: "when the two pairs of plates are well separated, long before the two pairs approach each other, they are individually stably causal and there is no risk of closed timelike curves [...]. So if a region of closed timelike curves forms as the two pairs of plates approach each other, that region must have a boundary, and there must be a first closed null curve," and Hawking's principle of chronology protection would prevent this from happening.[95]

Considering the analogy noted above between CSCs formed in the manner described here and the CSCs found in general relativity which Hawking's chronology protection conjecture is known to address, appealing to the latter in the present case seems fairly natural.[96] Recall that according to Hawking's conjecture, the formation of a

---

[93] [15], pp.180-181.

[94] Notably, Babichev *et al* also appeal to it. [58] p.21, [59] p.17.

[95] [15], p.181.

[96] Babichev *et al* even propose the following: "It is well-known that Analogue Gravity gives more simple and intuitively clear way to investigate the properties of Hawking radiation, the effects of Lorentz symmetry breaking, transplanckian problem *etc.*, by using the small perturbations in the fluids instead of direct implication of General Relativity. In a similar way, *analogue time-machine* or *analogue Chronology Protection Conjecture* may provide one with a tool to check *Chronology Protection Conjecture* and the possibility of construction of time machines in General

first closed null curve would involve uncontrollable singularities as vacuum fluctuations "pile up on top of each other", *i.e.* a virtual (or real) particle would keep meeting up with itself along the curve an infinite number of times, coherently reinforcing itself every time. This would actually lead to a back reaction of the quantum fields that would prevent the CSC from forming. Liberati *et al* argue that "although it was developed in the context of the causal problems typically expected to arise in Lorentzian wormholes, the argument is in fact generic to any type of chronology horizon."[97]

Another line of thought is the one proposed by Geroch. As discussed above, the importance of an initial value formulation to treat the issue of propagation in the coordinate past of a fast moving observer has been stressed by a number of researchers. In addition, Geroch uses the formalism he has developed in order to apply similar arguments to the problem of CSCs. For the sake of generality, he imagines two pipes filled with a fluid whose speed of sound is superluminal. The situation where CSCs are argued to form involves these two pipes moving at high speed past one another. This situation is described by a solution to the "special-relativity-superluminal-fluid system" of equations. Geroch remarks that the very fact it has CSCs implies that "this arrangement cannot be in the domain of dependence of any surface, *i.e.*, it cannot be predicted, via the initial-value formulation, from any initial data."[98] Someone wishing to bring the situation into existence would be able to set up the initial conditions in the sense of the two pipes being at rest along one another.

> But the issue of what happens from those initial conditions must be determined by evolving, from those initial data, the differential equations for the system. [Presumably, we would include also within this system

---

Relativity."[59] pp.17-18, [58] pp.22-23.

[97] [15], p.181.

[98] [63], p.11.

> the fields describing the observer, and the initial conditions would reflect that observers resolve to get the one pipe moving.] Whatever results from these data and these equations is what results. But we know that whatever it turns out to be the result of this evolution will not consist of the two pipes moving in the prescribed manner.[99]

Note that this view encompasses the type of chronology protection mechanism envisaged above, but is more general:

> Probably, it will be difficult to include in the system interactions that will allow the observer to move the fluid around in the manner he wishes – for example, the fluid may interact back with the observer, preventing him from manipulating that fluid in the desired manner; or, because of its equations, the fluid might respond to such manipulation in an unexpected manner. It is also possible that the fluid solutions themselves might become singular when the fluid is pushed too hard.[100]

Geroch as well draws an analogy with the situation in general relativity to illustrate this point:

> There exist solutions of Einstein's equation in general relativity that manifest closed causal curves. But we do not, in light of this circumstance, allow observers to build time-machines at their pleasure. Instead, we permit observers to construct initial conditions – and then we require that they live with the consequences of those conditions. It turns out that a time-machine is never a consequence, in this sense, of the equations of general relativity, in close analogy with the situation in the

---

[99] [63], p.12.

[100] *Idem.*

> special-relativity-superluminal-fluid example above.[101]

Whatever would happen, the main point is: CSCs would not appear because given a well-posed initial value formulation (such as the two pipes at rest constitute) CSCs simply *cannot* evolve. This impossibility is guaranteed by the formalism itself, so that there is no need for a *reductio ad absurdum* argument to the effect that the fluids cannot carry superluminal signals. Provided that the system of partial differential equations that describe the physical behavior of the entities of interest have a hyperbolization, even if the signal velocities of the entities it describes is larger than c, no causal paradox can occur.

2.6 Conclusion

We have seen how the concept of effective metric and causal cones can be argued to imply that superluminal propagation would not lead to the kind of causal paradox described in the bilking argument. At the same time, it should be stressed that the theories at stake come with a preferred reference frame. This naturally begs the question of their Lorentz invariance. In fact, some of the physicists whose work we have discussed in this chapter take the view that the theories which exhibit such features are not valid precisely because they are not Lorentz invariant in all the necessary ways. Adams *et al* indeed introduce their work with these words:

> [...] we will show that some apparently perfectly sensible low-energy effective field theories governed by local, Lorentz-invariant Lagrangians, are secretly non-local [and] do not admit any Lorentz-invariant notion of causality.[102]

---

[101] *Idem.*

[102] [37] p.2, [38] p.1. What exactly they are referring to by the latter statement is unclear: it could be the very presence of superluminal propagation, the ability to define a preferred reference frame,

In contrast, Liberati *et al* argue that the type of Lorentz invariance that matters is preserved:[103]

> Light behaves in a non-Lorentz invariant way only because the ground state of the electromagnetic field is not Lorentz invariant. The Euler-Heisenberg Lagrangian, from which the existence of the effect can be deduced, as well as all the machinery of QED employed in its derivation, is still fully Lorentz invariant. For this reason, one often speaks of a *soft breaking* of Lorentz invariance in order to distinguish from a situation in which also the basic equations, and not just the ground state, are no longer Lorentz invariant.
>
> Of course, this soft breaking of Lorentz invariance has no fundamental influence on special relativity no more than being inside a material medium has.[104]

So they take the view that what needs to respect Lorentz invariance are the action and the equations of motion.[105] These remarks inspired Jeremy Butterfield the following comments:

> There are two points here.
>
> (i): Suppose a theory obeys a symmetry in the sense that a certain transformation, e.g. a spatial rotation or a boost, maps any dynamical solution to another solution. This by no means implies that every solution should be invariant, *i.e.* mapped onto itself, under the transformation:

---

or what they argue is the possibility to bring about CSCs when regions of non-trivial vacuum move past one another.

[103] They are referring to the Scharnhorst effect specifically because this is what their work focuses on, but the remarks would hold for other theories.

[104] [15] p.168, [16] pp.2-3.

[105] See also [22], [23] p.33.

> after all, not every solution of Newtonian mechanics is spherically symmetric!
>
> (ii): Agreed, the vacuum state for empty Minkowski spacetime is required to be Lorentz-invariant since it should look the same in a translated, rotated or boosted frame. But the presence of the plates breaks this symmetry, just as a pervasive inertially-moving medium would do: licensing a non-Lorentz-invariant vacuum state.[106]

Hence it can be argued that as far as the role of Lorentz symmetry *per se* is concerned, the theories in question have mundane precedents in physics. What is obviously unusual about them is that here, it finds itself associated with superluminal behavior – hence Geroch's choice to use a hypothetical fluid endowed with a superluminal sound speed. However as we have discussed this does not seem to imply causal paradoxes in the theories in question.

---

[106] [22], [23] p.33.

Chapter 3

THE SCHARNHORST EFFECT: A HISTORICAL INTRODUCTION

3.1 Introduction

In 1990, Klaus Scharnhorst, soon followed by Gabriel Barton, predicted that both the phase and the group velocity of photons propagating in a Casimir vacuum would be larger than c – which has come to be known as the Scharnhorst effect. The issue soon arose as to whether the same could hold for the signal velocity. Several arguments were presented, notably to the effect that uncertainties would preclude an actual measurement of this value. Scharnhorst and Barton responded by arguing that given their previous result, the Kramers-Kronig relations imply one of two options: either the signal velocity is superluminal as well, or the Casimir vacuum behaves like an amplifying medium at some frequencies.[1] They did not express a preference for either option, although they explained that they did not believe a superluminal signal velocity would be at odds with special relativity in this context. A decade later, Stefano Liberati, Sebastiano Sonego and Matt Visser generalized the Scharnhorst effect to arbitrary photon incidence. They used an effective metric approach, and vigorously defended the idea that a superluminal signal velocity is compatible with special relativity. In what follows, I shall examine these discussions, and point to what I believe to be unresolved issues.

---

[1] The term “superluminal” is here used to mean a “faster than the speed of light in the usual unbounded, trivial vacuum.” Although the term sounds like a contradiction in terms when applied to photons as it is here, it has become standard use in the context of the Scharnhorst effect.

3.2 The Scharnhorst effect

As explained in the Introduction, the set-up in which the Scharnhorst effect is predicted to occur is similar to that of the Casimir effect: two parallel conducting but electrically neutral plates separated by vacuum.[2] The Scharnhorst effect was first derived in February 1990 by Klaus Scharnhorst at the Humboldt University of Berlin [5]. A month later Gabriel Barton, of the University of Sussex, proposed another calculation, and obtained the same result [6]. Both predicted that photons would travel faster than they do in true vacuum when propagating in the region between the plates: photons traveling normal to them would be fastest (and it is their speed that Scharnhorst and Barton computed), while those propagating parallel to them would do so at c. The Scharnhorst effect would decrease with plates separation, just as the Casimir effect does [5], [6].

However in contrast to the latter, which has been observed experimentally, the Scharnhorst effect has not, and in fact experimental confirmation is well beyond current technology. Indeed it is extremely small: a correction $\Delta c/c$ of the order of 10^{-32} for plates a micron apart – which is typical for experimental measurements of the Casimir effect.[3]

3.3 Early derivations of the Scharnhorst effect

Both Scharnhorst and Barton derived the phase velocity by first obtaining an index of refraction for the Casimir vacuum – and both found it by first working out the permittivity and permeability tensors of this vacuum. This approach implies considering the latter as a dielectric medium. Perhaps because a similar concept has long been familiar for (real) electromagnetic backgrounds, this has raised little con-

---

[2] The Casimir effect consists in the occurrence of an attractive force between the plates.

[3] See notably [1] for the discussion of such an experiment.

cern.[4] Barton stressed that this approach required restricting the analysis to photon wavelengths much shorter than L, because implicit in the use of a local refractive index is a "semi-classical (WKB type, ray-optics) approximation untroubled by the variation of $\Delta\varepsilon$ and $\Delta\mu$, separately, on a scale of L", where L is the plate separation.[5] In terms of the frequencies ω of the photons, this means that $1/\ L << \omega$. Also, both Scharnhorst and Barton use an effective field theory approach. Notably, their derivations are based on the assumption that all the photons involved ("virtual" photons of the background and the Scharnhorst photon propagating through it) are low energy compared to the electron rest mass: $\omega << m_e$. Hence overall they worked in the regime $1/L << \omega << m_e$. This was to be a key point in most of the early responses to their work. Despite these common features, their derivations were substantially different.

3.3.1 Scharnhorst's derivation (1990)

Scharnhorst obtained the index of refraction by computing how the presence of the plates modifies the effective action for the propagating photon, using quantum electrodynamics (QED). He derived the expression for the correction to the classical action due to the most relevant polarization tensor.

$$\Gamma\left[E, B\right] = \Gamma_{cl}\left[E, B\right] - \frac{1}{2}\int d^4x\, d^4y\, A^{\mu}(x)\, \Pi_{\mu\nu}^{2-loop}(x, y)\, A^{\nu}(y) \tag{3.1}$$

---

[4] The quantum vacuum is not the only physical notion that has been treated as a refractive medium. So have a heat bath and background electric or magnetic fields. In addition to his interest in the Casimir effect, Scharnhorst was also motivated by a study due to Rolf Tarrach which concluded that a heat bath modifies the velocity of light compared to its vacuum value [7].

[5] [6], p.561.

This is the two-loop polarization tensor $\Pi^{2-loop}_{\mu\nu}$, that is the one due to these two quantum processes:[6]

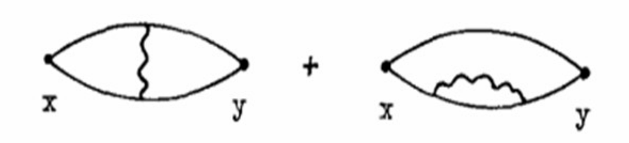


Figure 3.1: The two-loop processes responsible for the Scharnhorst effect.[7]

For comparison, the simplest correction to a photon propagating freely – what we usually mean by “vacuum polarization” – is the “one loop” diagram (where unlike in the above the external photons have been represented as well):[8]

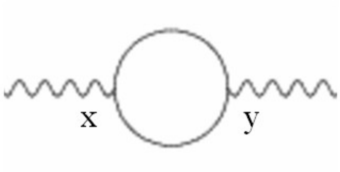


Figure 3.2: The one-loop process. It is not involved in the Scharnhorst effect

The one-loop process is not involved in the Scharnhorst effect because the off-shell

---

[6] These diagrams represent interactions between virtual particles from the vacuum, that affect photons (not shown). All the lines here refer to off-shell particles (*i.e.* “virtual particles”): as usual in QED a wiggly line stands for a photon propagator, and a straight line for an electron or positron propagator, where a propagator is a function that gives the probability amplitude of a particle (field excitation) to be found at two different spacetime points (events). x and y represent the events where the real photon interacts with the electron and positron: if the real photon was also depicted, it would be by two wiggly lines, one on each side, at x and y (so called-external lines).

The defining characteristic of off-shell (“virtual”) particles is that they do not obey the relativistic kinematical relation: $E^2 = p^2c^2 + m^2c^4$. These are theoretical entities, and taken to be impossible to observe directly.

Figure taken from [5], p.355.

[8] This corresponds to the real photon turning into a virtual electron and a virtual positron (the two solid lines) and back into a real photon.

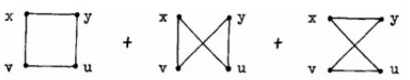
electrons and positrons are not affected by the presence of the plates.

The two loop processes responsible for the Scharnhorst effect differ from this in the additional off-shell photon they involve (in the diagram on the left this photon is exchanged between the off-shell electron and positron, and on the right it is emitted and reabsorbed by one of them). Unlike the electrons and positrons, the off-shell photon is affected by the plates, just as some of the classical electric and magnetic fields components would vanish at the surface of these conductors.

Scharnhorst derived the two-loop polarization tensor, which represents the sum of the amplitudes for the two processes in Fig.1, where the propagator for the off-shell photons is not the usual one, but has been modified by imposing boundary conditions to model the presence of the plates.[9] He then substituted his result for $\Pi_{\mu\nu}^{2-loop}(x,y)$ in Eq. (1), obtaining an expression for the effective action of the electromagnetic field in the presence of the plates. Scharnhorst then considered the expression for the effective action written in terms of the permittivity (ε_{ij}) and permeability (μ_{ij})

---

[9] Scharnhorst worked in coordinate space, and derived $\Pi_{\mu\nu}^{2-loop}$ by integrating over the one-loop four-point function $\Gamma_{\mu\nu\alpha\beta}(x,y,u,v)$ (figure taken from [5], p.355):



,

multiplied by the photon propagator modified by boundary conditions, $\bar{D}^{\alpha\beta}\,(\bar{u}-\bar{v};u_3,v_3)$. The latter quantity had been calculated five years earlier by M. Bordag and his colleagues [8]:

$$\Pi_{\mu\nu}^{2-loop}(x,y) \ = -\tfrac{i}{2}\int d^4u\; d^4v\; \Gamma_{\mu\nu\alpha\beta}(x,y,u,v)\; \bar{D}^{\alpha\beta}\,(\bar{u}-\bar{v};u_3,v_3),$$

where:

$$\bar{D}_{\mu\nu}\,(\bar{u}-\bar{v};u_3,v_3) = \tfrac{1}{2i}\int \tfrac{d^3\tilde{k}}{(2\pi)^3}\,\tfrac{1}{\Gamma(\tilde{k})}\,\left(\tilde{g}_{\mu\nu}-\tfrac{\tilde{k}_\mu\tilde{k}_\nu}{\Gamma^2(\tilde{k})}\right)\; e^{i\,\tilde{k}(\tilde{x}-\tilde{y})}\; e^{i\,\Gamma(\tilde{k})|x_3-a_i|}\;\left(h^{-1}\right)_{ij}\; e^{i\,\Gamma(\tilde{k})|y_3-a_j|}\;.$$

The subscript 3 refers to the spatial direction normal to the plates and all quantities with a tilda run only over the indices 0, 1, 2 (time and the spatial dimensions parallel to the plates). Performing the integral to obtain $\Pi_{\mu\nu}^{2-loop}$ comes down to taking the events u and v and "attaching" to them the photon propagator at x and y.

tensors:

$$\Gamma\left[E,B\right] = \frac{1}{2}\int d^4x \left(\epsilon_{ij}E^iE^j - \mu_{ij}^{-1}B^iB^j\right). \tag{3.2}$$

Comparing the two forms for the action, he obtained expressions for ε_{ij} and μ_{ij} between Casimir plates, which he used to find the index of refraction for light travelling normal to the plates:

$$n_\perp = \sqrt{\varepsilon_{11}\,\mu_{11}} = 1 - \frac{11}{2^6\cdot(45)^2}\frac{e^4}{(m_eL)^4}, \tag{3.3}$$

and the corresponding phase velocity:[10]

$$c_\perp = \left(1 + \frac{11}{2^6\cdot(45)^2}\frac{e^4}{(m_eL)^4}\right)\,c_0. \tag{3.4}$$

Scharnhorst noted that in the approximation considered there is no dispersion (the index of refraction is not frequency dependent), so the phase velocity $v_\varphi = \frac{\omega}{k}$ is also the group velocity $v_g = \frac{d\omega}{dk}$.

3.3.2 Barton's derivation (1990)

The following month Gabriel Barton proposed a different derivation of the effect, which he deemed more "elementary" [6]. As mentioned above, Barton too computed the phase velocity from the index of refraction, and the latter from the permittivity and permeability tensors. However unlike Scharnhorst he did not derive the modification in the action from the relevant polarization tensor. Instead he used the Euler-Heisenberg Lagrangian, which is an effective Lagrangian, which is meant to apply to situations where the energies of the photons are much smaller than the mass-energy of the electron. For this reason electrons have been "integrated out" of

---

[10] Indeed: $c_\perp \equiv v_{\varphi\perp} = \frac{c}{n_\perp}$, so with $n_\perp = 1 - |\Delta n_\perp|$ one gets $c_\perp = \frac{c}{1-|\Delta n_\perp|} \approx c\left(1 + |\Delta n_\perp|\right)$.

the theory that this Lagrangian embodies. As a result, unlike its QED counterpart, the Euler-Heisenberg Lagrangian does not involve fermion fields (*i.e.* electron and positron fields), but involves only the photon field A^μ. It is usually written in terms of the electric and magnetic fields E and B:[11]

$$L = \frac{E^2 - B^2}{8\pi} + g\left(\left(E^2 - B^2\right)^2 + 7\,(E.B)^2\right) \tag{3.5}$$

where:

$$g \equiv \frac{\alpha^2}{2^3.3^2.5\pi^2 m^4}. \tag{3.6}$$

The first term is the classical Lagrangian for electromagnetism, and the second term the correction to it to order α^2, that we designate by ΔL. Barton first obtained the polarization and magnetization vectors by varying this correction with respect to the electromagnetic fields:

$$P = \frac{\delta\,\Delta L}{\delta\,E} = \; 4g\;\left(E^2 - B^2\right)\;E + 14g\;(E \cdot B)\;B, \tag{3.7}$$

$$M = \frac{\delta\,\Delta L}{\delta\,B} = \; -4g\;\left(E^2 - B^2\right)\;B + 14g\;(E \cdot B)\;E. \tag{3.8}$$

He then separated the electric and magnetic fields each into a sum of two fields, one classical, the other quantized: $E \rightarrow E + \,e$, $B \rightarrow B + \,b$. The classical fields (e or b) are those of the Scharnhorst photon, and the quantized ones (still written E or B) are the background Casimir vacuum.

Barton then related the polarization and magnetization to the permittivity and

---

[11] E and B are both components of the electromagnetic field tensor $F^{\mu\nu}$ used in the covariant formulation of electrodynamics.

permeability tensors, using the corresponding susceptibility tensors $\chi^{(e)}$ and $\chi^{(m)}$:

$$\varepsilon_{ij} = \delta_{ij} + 4\pi\, \chi^{(e)}_{(ij)} \qquad\qquad \mu_{ij} = \delta_{ij} + 4\pi\, \chi^{(m)}_{(ij)}, \tag{3.9}$$

and:

$$P_i = 4g\ \left(\left\langle E^2 - B^2 \right\rangle e_i + 2\ \langle E_i E_j\rangle\ e_j\right) + 14g\ \langle B_i B_j\rangle\ e_j \equiv \chi^{(e)}_{(ij)}\, e_j, \tag{3.10}$$

$$M_i = 4g\ \left(-\left\langle E^2 - B^2 \right\rangle b_i + 2\ \langle B_i B_j\rangle\ b_j\right) + 14g\ \langle E_i E_j\rangle\ b_j \equiv \chi^{(m)}_{(ij)}\, b_j. \tag{3.11}$$

The effect of the plates comes in through the vacuum expectation values $\langle E^2 - B^2\rangle$, $\langle E_i E_j\rangle$, $\langle B_\mathbf{i} B_\mathbf{j}\rangle$.[12]

Substituting the results he obtained for these expectation values in Eqs. (3.10) and (3.11), Barton found expressions for change in the permittivity and permeability tensors due to the presence of the plates:

$$\Delta\varepsilon_{ij} = \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g\ \left[\left(-\delta^{||} + \delta^{\perp}\right)_{ij} \frac{11}{120} + \delta_{ij} 9 f\left(\xi\right)\right], \tag{3.12}$$

$$\Delta\mu_{ij} = \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g\ \left[\left(-\delta^{||} + \delta^{\perp}\right)_{ij} \frac{11}{120} - \delta_{ij} 9 f\left(\xi\right)\right]. \tag{3.13}$$

---

[12] These expectation values are obtained by solving the homogeneous wave equation for the vector potential with boundary conditions for the conducting plates ($A_{||} = 0$ and $\partial_\perp A_\perp = 0$), taking the mode expansion of $A(r,t)$, and quantizing it by imposing the usual commutation relations for the creation and annihilation operators of the photon fields. Barton then obtained E and B from A ($E = -\partial_t\, A$, $B = \nabla\times A$), and:

$$\langle E_i E_j\rangle = \left(\frac{\pi}{L}\right)^4 \frac{2}{3\pi}\ \left[\left(-\delta^{||} + \delta^{\perp}\right)_{ij} \frac{1}{120} + \delta_{ij} f\left(\xi\right)\right],$$

$$\langle B_i B_j\rangle = \left(\frac{\pi}{L}\right)^4 \frac{2}{3\pi}\ \left[\left(-\delta^{||} + \delta^{\perp}\right)_{ij} \frac{1}{120} - \delta_{ij} f\left(\xi\right)\right],$$

with: $\delta^{||}_{ij} \equiv (\delta_{i1}\delta_{j1} + \delta_{i2}\delta_{j2})$, $\delta^{\perp}_{ij} \equiv \delta_{i3}\delta_{j3}$, $f\left(\xi\right) = \frac{3-2\,\cos^2(\xi/2)}{8\,\cos^4(\xi/2)}$.

For a photon propagating normal to the plates, these take the form:

$$\Delta\varepsilon = -\left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left[\frac{11}{120} - 9f\left(\xi\right)\right], \tag{3.14}$$

$$\Delta\mu = -\left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left[\frac{11}{120} + 9f\left(\xi\right)\right]. \tag{3.15}$$

These yield an index of refraction:

$$\Delta n = -\frac{\alpha^2}{\left(m_e L\right)^4} \frac{11\pi^2}{2^2.3^4.5^2}. \tag{3.16}$$

With $\alpha = \frac{e^2}{4\pi}$, this does correspond to the result obtained by Scharnhorst, *i.e.* $\frac{11}{2^6(45)^2}\frac{e^4}{(mL)^2}$.

Both Scharnhorst and Barton stated that the group and the phase velocities of the photons were equal, so that their result represented both. In 1990 Barton also asserted that it corresponded to the signal velocity. However they did not discuss these points in any detail then. Only after responses to their papers were published did they do so, in 1993 [12].

3.4 Early reactions (1990)

Scharnhorst's and Barton's works were published in February and March 1990, respectively. They gave rise to two reactions later the same year. Both aimed to show that the velocity obtained by Scharnhorst and Barton cannot be used for sending signals. In October, a paper by Peter Milonni & Karl Svozil was published, and the following month one by Shahar Ben-Menahem, in which he gave two arguments.[13]

---

[13] For Scharnhorst and Barton's 1990 papers see respectively [5], [6], for Milonni and Svozil's response [10], and for Ben-Menahem's [11].

3.4.1 Milonni & Svotzil's argument

The argument given by Milonni and Svozil to the effect that the velocity of Scharnhorst photons cannot be used to signal is based on the time-energy uncertainty relation.

Milonni and Svozil argued that measuring the actual signal speed v requires measuring the time t that a photon takes to travel a distance L'. This travel time cannot be measured to arbitrary accuracy: there is an uncertainty Δt in its measurement. The latter arises at least from the uncertainty in the time it takes to switch on the signal. This uncertainty in the time measurement leads to an uncertainty in the velocity measurement:

$$\Delta v = c^2 \frac{\Delta t}{L'}.^{14} \tag{3.17}$$

In order to minimize Δv, they chose the distance of travel L' to be as large as possible without the photon reflecting off the plates, *i.e.* they took it to be equal to the plate separation L:

$$\Delta v = c^2 \frac{\Delta t}{L} \tag{3.18}$$

where Δt is the uncertainty in time taken to switch on a signal. They considered the case of a photon emitted by exciting an atom. Then from the uncertainty between time and energy, $\Delta t = \nabla/(2\Delta E) = 1/\omega$, where ΔE is the difference in energy of the levels between which the atomic electron transitions for the photon to be

---

[14] [10], p.438. This can be derived as follows. In general, the uncertainty in a function $R = R\,(X, Y)$ is given by $\delta R = \sqrt{\left(\frac{\partial R}{\partial X} \cdot \delta X\right)^2 + \left(\frac{\partial R}{\partial Y} \cdot \delta Y\right)^2}$. In the present case the analog of R is $v = L'\,t^{-1}$, so:

$$\delta v = \sqrt{\left(\frac{\partial\,(L'\,t^{-1})}{\partial X} \cdot \delta t\right)^2} = \sqrt{\left(-\frac{L'}{t^2} \cdot \delta t\right)^2} = \frac{L'}{t^2} \cdot \delta t = \frac{L'}{\left(\frac{L'}{v}\right)^2} \cdot \delta t = \frac{v^2}{L'} \delta t.$$

emitted, and ω is the frequency of this photon. Then:

$$\Delta v = c^2 \frac{1}{\omega L} \sim c \frac{\lambda}{L}. \tag{3.19}$$

Milonni and Svozil then considered the ratio of this uncertainty Δv in the velocity measurement to the shift in velocity predicted by Scharnhorst and Barton, and obtained:

$$\frac{\Delta v}{\Delta c} = \frac{1}{\kappa \alpha^2} \frac{\lambda}{L} \left(\frac{mcL}{\hbar} \right)^4 = \frac{\lambda}{\kappa \alpha^2} \frac{1}{\lambda_c} \left(\frac{L}{\lambda_c} \right)^3, \tag{3.20}$$

where $\kappa = \frac{11\pi^2}{2^2 3^4 5^2}$ and $\lambda_c = \frac{\hbar}{mc}$ is the Compton wavelength.[15]
To minimize this ratio, they took λ as small as possible. Scharnhorst and Barton's derivations were restricted to the regime where $\omega << m_e$, $\lambda >> \lambda_c$, so:

$$\frac{\Delta v}{\Delta c} \geq \frac{1}{\kappa \alpha^2} \frac{L^3}{\lambda_c^3} \cong 1.5 \times 10^6 \frac{L^3}{\lambda_c^3}.^{16} \tag{3.21}$$

This shows that the relative uncertainty decreases as the plate separation does. The smallest conceivable value for L, they argued, would be λ, since $\frac{1}{L} << \omega$. In this case then $L = \lambda_c$ and:

$$t = \frac{L}{c} = \frac{\lambda_c}{c} = \frac{\hbar}{mc^2} \equiv \frac{1}{\omega_{\max}} \leq \frac{1}{\omega} = \Delta t. \tag{3.22}$$

That is, the uncertainty in the time measurement would be larger than the time the photon takes to travel the distance between the plates – and they added this would

---

[15] [10], p.438. To derive this result they first convert Barton's expression to hybrid units, *i.e.* $m \to mc^2$ and $L \to \frac{L}{\hbar c}$ (fourth equality):

$$\Delta c = c \Delta n = c \frac{11\pi^2}{2^2 3^4 5^2} \frac{\alpha^2}{(mL)^4} = c\kappa \frac{\alpha^2}{(mL)^4} = c\kappa \frac{\alpha^2}{\left(mc^2 \frac{L}{\hbar c}\right)^4} = c\kappa \frac{\alpha^2}{L^4} \lambda_c^4.$$

[16] [10], p.438. Where they have used: $\kappa \cong 1.3 \times 10^{-2}$, $\alpha \cong \frac{1}{137}$. Also $\lambda_c \cong 3.9 \times 10^{-11} cm$, so $\frac{\Delta v}{\Delta c} \geq 2.5 \times 10^{37} L^3$ with L in *cm*.

imply that the uncertainty in the velocity measurement would be larger than c. As a more reasonable value for L, they considered the Bohr radius, $\mathrm{a}_0 = \lambda_c/\alpha$.[17] This yields a relative uncertainty in the velocity of:

$$\frac{\Delta v}{\Delta c} \approx \frac{1}{\kappa\alpha^5} \approx 3.7 \times 10^{12}.^{18} \tag{3.23}$$

Milonni and Svozil concluded that "the uncertainty in the measured propagation velocity will always be enormously larger than the correction to c associated with the Scharnhorst effect", and that therefore "no measurement of the faster-than-c velocity of light predicted by the Scharnhorst effect is possible."[19]

3.4.2 Ben-Menahem's wavefront argument

One of Shahar Ben-Menahem's two arguments bears some similarity to the one just discussed. Indeed it too is concerned with uncertainty and measurement, but with distance as well as time, and in particular, it is not as clearly related to the uncertainty relations.[20] Ben-Menahem argued that in order to send a signal, one would have to form a sharp-fronted wave packet. Scharnhorst and Barton's result relied on the assumption $\omega << m_e$, so such an infinitely-sharp wave packet cannot be formed using only waves of such low frequencies (hence of components whose wavelength $\lambda >> 1/m_e$, *i.e.* much larger than the Compton wavelength of the electron). Put differently, the wavefront of a wave packet comprising only such long waves, cannot

---

[17] Although they noted: "the distinction between the plates as macroscopic objects becomes blurred, of course, and repulsive forces associated with overlapping electron wavefunctions come into play to increase u [the energy density, negative in the Casimir vacuum] and weaken and remove the Scharnhorst effect." [10], p.438.

[18] In fact they wrote 13.7×10^{12} but this must have been a typo.
Note that using a plate separation of a micron, a separation typical of the measurements performed for the Casimir effect, $\Delta v \approx 2.5 \times 10^{25}$.

[19] [10], p.438.

[20] [11], pp.133-134.

be a step function (*i.e.* arbitrarily sharp): instead it is “smeared” at least on the spatiotemporal scale of the Compton wavelength $1/m_e$.

Ben-Menahem reasoned that in order to unambiguously signal faster than c, the wavefront must “move beyond the light cone”[21] by a distance larger than the uncertainty associated with the “smearing”, hence $\delta x > 1/m_e$:

$$\delta x \approx \left(\frac{1}{n_\perp} - 1\right) t \approx \zeta \frac{e^4}{(m_e L)^4}\, t > \frac{1}{m_e}, \tag{3.24}$$

where $\zeta = \frac{11}{2^6.(45)^2}$, *i.e.* just a positive constant.[22]

Ben-Menahem then showed that this inequality cannot be satisfied in the regime used to obtain the Scharnhorst effect. Like Milonni and Svozil, he required that the wave packet should not reflect off the plates. This means that t is at most $t_{max} = L/c$, *i.e.* in natural units where $c = 1$:

$$\zeta\, e^4 \frac{1}{(m_e L)^3} > 1. \tag{3.25}$$

But the effect was derived in the regime $\frac{1}{L} << \omega << m_e$, *i.e.* $m_e\, L >> 1$, so this inequality cannot hold.

Note that the measurement uncertainty Ben-Menahem discussed is not *per se* a quantum uncertainty, but expresses the impossibility to accurately define the front of the signal, even classically.

---

[21] That is, in a given time it must move an additional distance compared to what it would if it traveled at speed c.

[22] Unlike Milonni and Svozil who used Barton’s, Ben-Menahem used Scharnhorst’s form: $n_\perp = 1 - \frac{11}{2^6.(45)^2} \frac{e^4}{(m_e L)^4} = 1 - \zeta \frac{e^4}{(m_e L)^4}$. δx can be obtained from: $\Delta\, x = \Delta\, v_{\varphi\perp}\, t = \left(\frac{c}{n_\perp} - c\right) t = \left(\frac{c}{1-|\Delta n_\perp|} - c\right) t \approx c\,(1 + |\Delta n_\perp| - 1)\, t = |\Delta n_\perp|\, c t$, with $|\Delta n_\perp| = \zeta \frac{e^4}{(m_e L)^4}$ and $c = 1$ (*i.e.* natural units).

3.4.3 Ben-Menahem's commutator argument

In the same paper, Ben-Menahem presented a second argument meant to show that the velocity derived by Scharnhorst and Barton cannot be a signal velocity. Unlike those already discussed, it did not consist in an analysis of a measurement uncertainty. Instead, it appealed to a formulation of causality standard in quantum field theory: the idea that a signal cannot be transmitted faster than c provided that the commutator of the fields in question vanish "outside the light cone" – *i.e.* the commutator of $\varphi(x)$ and $\varphi(y)$, must be zero when x and y are spacelike separated.[23]

However Ben-Menahem did not use commutators in the usual form $[\varphi(x),\, \varphi(y)]$, but what he termed "commutator functions", that is "response functions [which] are essentially the vacuum expectation values of commutators of E(x) and B(x) with the fields at y":

$$\frac{\delta}{\delta\, J_\mu(y)} \left\langle E_i(x) \right\rangle , \quad \frac{\delta}{\delta\, J_\mu(y)} \left\langle B_i(x) \right\rangle . \tag{3.26}$$

It is implied that, in a Lorentz reference frame anchored to the plates, the time components of the two events are ordered with cause preceding effect: *i.e.* $x_0 > y_0$. Thus for instance, if the disturbance is a small antenna at spacetime point y with AC current in direction j, and the measured quantity at point x is the component i of the magnetic field, the corresponding commutator function is given by:

$$\frac{\delta}{\delta\, J_j(y)} \left\langle B_i^{(0)}(x) \right\rangle = \varepsilon_{ikj} \partial_k^x G_r^{(0)}(x, y) \tag{3.27}$$

where $G_r^{(0)}(x, y)$ is the retarded Green's function (a.k.a. photon propagator) for a free massless scalar field, subject to the zero-field boundary conditions imposed by

---

[23] [11], pp.134-137.
That is $[\varphi(x),\, \varphi(y)] \equiv \varphi(x)\varphi(y) - \varphi(y)\varphi(x)$ must be 0, where $\varphi(x)$ is the field evaluated at event x, and x and y are events, *i.e.* spacetime points. In general the fields at points x and y can be different fields or field components; e.g. $\mathbf{E_x}$ and $\mathbf{E_y}$, or $\mathbf{E_z}$ and $\mathbf{B_z}$, etc.

the plates. The superscript (0) refers to the fact that the field considered is free (*i.e.* no radiative corrections have as yet been included). Its free-space (*i.e.* no-plates) counterpart, Ben-Menahem denoted by $\hat{G}_r^{(0)}(x-y)$; while the fully interacting version (including both plates and radiative corrections, *i.e.* all Feynman graphs), is denoted $G_r(x,y)$.[24]

Expressed in terms of spacetime coordinates, the propagator $\hat{G}_r^{(0)}$ is:[25]

$$\widehat{\hat{G}}_r^{(0)}(x-y) = \frac{1}{4\pi\,|x|}\theta(x_0)\,\delta\,(x_0 - |x|) \tag{3.28}$$

at $y = 0$, where the $\theta(x_0)$ factor imposes "Newtonian ("time-arrow") causality" in the sense that $\hat{G}_r^{(0)}$ can only be nonzero at times later than the disturbance occurred; while the factor $\delta\,(x_0 - |x|)$ accounts for the fact that by time x_0, it had traveled a distance $|x|= c\,x_0$ (with $c = 1$), so that $\hat{G}_r^{(0)}$ is 0 everywhere else at that moment. So this relation characterizes the light cone.[26] Hence for the signal velocity to be different from c, what must differ in the fully-interacting analog for $G_r^{(0)}$ of the *RHS* of Eq. (3.28) must be $\delta\,(x_0 - |x|)$.[27] Ben-Menahem argued that for this to obtain, "to order e^4, the commutator function [must receive] a correction in which $\delta\,(x_0 - |x|)$ is replaced by a *derivative* of a delta function."[28] In such a scenario, the

---

[24] *i.e.* The hat is used when the function is *not* subject to plate boundary conditions. Note that of the three Green's functions defined here, only the second – *i.e.* the one with no plate effects and no radiative corrections – is a function of $x - y$. This property is also shared by a fourth type of retarded Green's function one may define: the fully interacting free-space $\hat{G}_r(x-y)$ *i.e.* with full radiative corrections yet no plate effects.

[25] As opposed to being expressed in momentum space; x_0 is the time coordinate.

[26] Technically a hypercone in 4D Minkowski space.

[27] Ben-Menahem also noted that from the method of images, $G_r^{(0)}$ must be related to $\hat{G}_r^{(0)}$ by: $G_r^{(0)}(x,y) = \hat{G}_r^{(0)}(x-y) + reflexions$, p.135.

[28] [11], p.135.

dispersion relation of a Scharnhorst photon:

$$\omega^2 \approx k_1^2 + k_2^2 + \frac{1}{n_\perp^2} k_3^2 \quad \text{where} \quad n_\perp < 1. \tag{3.29}$$

would cause the light cone to be "tilted" due to $n_\perp \neq 1$.[29] Ben-Menahem meant to show that in fact, "this dispersion relation receives 'dispersive' corrections that ensure that the plates cannot affect the leading light-cone singularity structure (*i.e.* $\delta\left(x_0 - |x|\right)$) and thus the wavefront of a localized disturbance moves at precisely the speed *c*."[30]

In order to demonstrate this, Ben-Menahem considered the (magnetic-magnetic) commutator function; "to all orders in the causal expansion" it obeys the recursion relation:

$$\frac{\delta}{\delta\, J_j(y)} \left\langle B_i(x) \right\rangle = \varepsilon_{ikl} \partial_k^x \left(\delta_{lj} G_r^{(0)}(x,y) + G_r^{(0)}(x,z) \Sigma_{l\mu}(z,w) \frac{\delta}{\delta\, J_j(y)} \left\langle A^\mu(w) \right\rangle \right).^{31} \tag{3.30}$$

In this expression, $\Sigma_{\mu\nu}$ is the sum of the one-particle-irreducible *causal graphs* (to all loops) just as $\Pi_{\mu\nu}$ is the sum of the one-particle-irreducible Feynman diagrams in standard QED; and the relevant ones for analyzing the Scharnhorst-Barton effect, are again the 2-loop diagrams.[32]

However the "causal graphs" introduced and analyzed by Ben-Menahem are not the usual Feynman diagrams of QED. In a Feynman diagram, the propagators are

---

[29] By "tilted" he meant that its opening hyper-angle in Minkowski space would deviate from 45 degrees in natural units.

[30] [11], p.135.

[31] Note this involves $\left\langle B_i(x) \right\rangle$, not $\left\langle B_i^{(0)}(x) \right\rangle$, *i.e.* an interacting rather than a free field. A free field would still interact with the plates via the perfect-conductor boundary conditions, but would not have radiative corrections, and hence no electron-loop contributions – and hence could not include the Scharnhorst-Barton effect. Also note that each occurrence of a free propagator (Green's function) on the *RHS*, corresponds to the sum discussed in footnote 24.

[32] Recall that virtual fermions are not affected by the plates.

Feynman, *i.e.* time-ordered propagators, and all particles (quanta) taking part in the interactions are off-shell. By contrast, Ben-Menahem's causal graphs are built of two different types of propagators: *on-shell* (which refers to the propagating particle satisfying its classical, relativistic energy-momentum dispersion relation); and *causal*, which may be retarded or advanced.[33] The retarded propagators refer to particles and the advanced ones to antiparticles, both propagating forward in time and within (for fermions) or on (for photons) the Minkowski light-cone. Ben-Menahem referred to retarded and advanced propagators as "causal propagators". Their particles are off-shell, *i.e.* not on the "mass shell" in 4-momentum space, but they are on the spacetime light cone instead. In other words, they do not obey (in natural units) $E^2 = p^2 + m^2$, but rather $|x\,| = t$ (in the case of photons) or $|x\,| \leq t$ (in the case of fermions).

The defining characteristic of a causal graph is that it "contains at least one unbroken chain of causal propagators, running forward in time from the source point y to the measurement point x."[34] A single Feynman diagram becomes, in a causal calculation, a sum of many causal graphs with the same topology, but all allowed combinations of propagators.

As with the Feynman diagrams used by Scharnhorst, the only plate-dependent internal line (free propagator) in the relevant 2-loop causal diagrams, is the propagator of the photon emitted and absorbed by the circulating fermion line. Ben-Menahem

---

[33] In the unitary Coulomb gauge; they may also be Coulomb propagators, but all these cancel thanks to the gauge invariance — enforced via the Ward identities — see [11] p.136 above Eq. (13).

[34] [11], p.136. For instance in the following graph, the lines labeled r, a, r, r and r form such a chain (figure from [11], p.137):

distinguished two types of two-loop causal graphs relevant to the Scharnhorst effect: those in which this photon propagator is *causal*, and those in which it is *on-shell*.[35] He argued that the plate-dependence of the former (*i.e.* causal) can only arise through "spatial range effects", *i.e.* when both spacetime end-points of the photon propagator are spatially near one of the plates. This is because this photon is causal and is both emitted and absorbed by virtual electrons that exist for a time of order the Compton wavelength over c, and therefore this photon can only exist for a duration of this order — and because it is on the light cone, *i.e.* it propagates at speed c, this lifetime is not long enough for it to reach (hence interact with) one of the plates, unless it is produced within a few Compton wavelengths of that plate. These arguments imply that the causal-internal-photon cannot modify the light cone singularity when the disturbance and measurement events (y and x) are at distances of order L from either plate.

Consequently, only (the plate dependent contributions to) causal graphs where the internal photon propagator is *on-shell* have a chance of modifying the light cone singularity of the (x, y) commutator function under study. Unlike for the causal photon line, the on-shell photon line causal graphs do have a plate dependence even when the events (x, y) are spatially far from either plate.[36] However, Ben-Menahem argued that even the part of $\Sigma_{\mu\nu}$ corresponding to this contribution "cannot modify the delta-function in the free commutator function [...] by a derivative of a delta-function."[37] He concluded that therefore, "wavefronts still move at speed c,

---

[35] This is the relevant difference between Ben-Menahem's diagrams c and d, not their topology. $\Sigma_{\mu\nu}$ is given by:

$$\Sigma^{(p)}_{\mu\nu}(x,y) = \int\int d^4z\, d^4w \sum_{i=1}^{2} f^{(i)}_{\alpha\beta}\ (z,w) \times \Pi^{(i)}_{\mu\nu\alpha\beta}(x,y,z,w)$$

where $f^{(1)}$ refers to the on-shell photon and $f^{(2)}$ to the causal one.

[36] Again, far relative to the Compton wavelength.

[37] [11], p.137. He referred to Appendix C of J.D. Bjorken and S.D. Drell, Relativistic quantum

even when the two-loop corrections to the commutator functions are taken into account."[38]

The import of this argument is that non-perturbative analysis is not required, but that a perturbative approach suffices to settle the issue of whether the signal speed is larger than c. Its aim is to show that in such an approach causal graphs prove that the signal speed is in fact c.[39]

3.5 Scharnhorst and Barton's response: implications of the Kramers-Kronig relations

Three years later, in 1993, Scharnhorst and Barton published a joint paper in which they responded to Milonni, Svozil and Ben-Menahem. They did so only cursorily with respect to the latter's commutator argument just discussed; instead they focused on the concerns raised in the former two. As far as the issue of measurability is concerned, although they accepted the arguments, they stressed that these rely on the photon not being permitted to reflect off the plates; indeed they summarized the issue as follows (my italics):

> *On a single traverse* the low-frequency prediction (2)[40] cannot be verified even in principle, because any wavegroup narrow enough to afford the requisite accuracy must include significant high-frequency components

---

fields (McGraw-Hill, New York, 1965) for a proof.

[38] [11], p.137.

[39] This is subject to the "usual qualifications: we do not know that the perturbative diagrams converge; there could be non-perturbative effects even at the low EM fields involved in the interacting commutator functions; etc. However, we can safely say that the claims of the previous paragraphs, are just as certain as any other rigorous prediction of perturbative QED" (personal communication).

[40] $n = 1 + \delta n$, $\delta n = -\frac{11\pi^2}{2^2.3^4.5^2}\frac{e^4}{(mL)^4}$.

> to which the effective coupling (1)[41] and therefore (2) no longer apply.[42]

They refrained from addressing the concern of measurability directly:

> Such considerations are not specific to the effect we are studying. What they show, equally for quantum and for classical waves, is that in confined geometries a single measurement on a single traverse can determine the speed of light having limited frequencies only with limited accuracy: under such conditions the operational significance of any ultimate speed is apt to remain somewhat nebulous. (As so often when applying the indeterminacy relations, one could argue that the statistically analysed average of many measurements does make it possible, in principle, to determine shifts well below the mean-square deviations, and thus to verify effects that more cursory considerations of single measurements sometimes describe as undetectable. However, this is not the place to pursue such very wide questions relating to measurement theory in general.)[43]

Instead, they meant to sidestep it and discuss signal velocity from a different perspective. The same two arguments also used the fact that the derivations of the Scharnhorst effect were done in the regime $\omega << m_e$. Scharnhorst and Barton now related this to the fact that the signal velocity[44] can be defined as:

$$v_{signal} = \frac{1}{Re\, n\,(\omega \to \infty)}, \tag{3.31}$$

so the signal velocity would be larger than c if and only if Re n $(\omega \to \infty) < 1$, a quantity which clearly cannot be found using the approximation $\omega << m_e$. In order

---

[41] $\Delta\mathcal{L} = \frac{1}{2^3.3^2.5\pi^2}\frac{e^4}{m^4}\left(\left(E^2 - B^2\right)^2 + 7\,(E.B)^2\right)$

[42] [12], p.2040.

[43] *Idem.*

[44] Understood as the velocity of a sharp wavefront advancing into an initially undisturbed medium.

to bypass this difficulty, they used an argument relating the low and high frequency regimes in order to connect their result (the index of refraction in the low energy regime, represented by $n(0)$) to the quantity of interest (its high energy counterpart $n(\omega \rightarrow \infty)$, or $n(\infty)$ for short). It relies on a theorem "based on local causality" and, most pertinently, not specific to the low energy regime.[45]

Their aim was not actually to draw a definite conclusion regarding the possible superluminal character of the signal velocity. Instead, they meant to show that the argument entailed the following dilemma:

> This connection will show that one must accept at least one of two equally unorthodox possibilities: either $n(\infty) < 1$, so that the true signal velocity, too, exceeds c; or the conventional no-photon vacuum between the (fixed!) mirrors amplifies a probe beam. In the second case the vacuum would fail to act as a passive medium.[46]

The relationship in question is the Kramers-Kronig relation, and they stressed that these conclusions:

> remain conveniently immune to non-perturbative effects, as long as these do not cause the dispersion relation [Kramers-Kronig relation] to diverge; such immunity follows simply because perturbation theory does suffice to determine $n(0)$.[47]

---

[45] [12], p.2038.

[46] [12], p.2042.

[47] [12], p.2043. In the present context the Kramers-Kronig relation becomes a dispersion relation as it expresses the relationship between index of refraction and frequency; however it is more general than that.

The Kramers-Kronig relation connects the real part of the index of refraction to its imaginary part:

$$Re\, n\,(\omega) \;=\; Re\, n\,(\infty) \;+\; \frac{2}{\pi}\int_0^{\infty} d\omega' \,\frac{\omega' Im\, n(\omega')}{\omega'^2 - \omega^2}. \tag{3.32}$$

Considering the specific case of $n(\omega) = n(0)$ and rearranging yields:

$$n\,(\infty) = n\,(0) \;-\; \frac{2}{\pi}\int_0^{\infty} d\omega' \,\frac{Im\, n(\omega')}{\omega'} \;. \tag{3.33}$$

Scharnhorst and Barton stated that this leads to two alternatives.

If the medium (*i.e.* the Casimir vacuum) is passive, Im $n(\omega) > 0$, so the integral is positive, and $n(\infty) < n(0) < 1$.[48] In this case:

$$v_{signal} = v_{\Phi}(\infty) = \frac{c}{n(\infty)} > c. \tag{3.34}$$

On the other hand, if $v_{signal} \;=\; c$, $n\;(\infty) \;=\; 1$, and since $n(0) < 1$, the integral must be strictly negative. Therefore Im $n < 0$ as well, for at least some frequencies. This entails that the Casimir vacuum fails to behave passively for at least some frequencies.[49] This would mean that it would induce the generation of real photons, which *prima facie* would seem to require energy creation out of nothing.

Now there exists a gravitational analog to the Scharnhorst effect, the "Drummond-Hathrell" effect (it predicts photon phase and group velocities larger than c in some

---

[48] Recall that $n(0) < 1$ was the main result of their 1990 derivations.

[49] That Im $n(\omega) < 0$ entails amplification can be seen from the expression for an electric field of wave group:

$$E(t,z) = \int_{-\,\infty}^{+\,\infty} d\omega\; a(\omega)\; e^{-\,i\,\omega\,t\,\left(1\;-\;\frac{n(\omega)\;z}{t}\right)}.$$

A monochromatic plane wave from the latter would be proportional to $e^{-\,i\,\omega\,t\,\left(1\;-\;\frac{n(\omega)\;z}{t}\right)}$. Hence Im n contributes $e^{-\,i\,\omega\;(-i\;Imn(\omega)\;z)} \;=\; e^{-\,\omega\;Imn(\omega)\;z}$. If Im $n(\omega) < 0$, the amplitude of the mode of frequency ω increases exponentially with z, as it propagates.

gravitational backgrounds), and it has been shown that in this context, Im n can turn negative, provided that the medium is inhomogeneous in a way that can focus the beam: there can be a local increase in amplitude without any global creation of photons. Scharnhorst and Barton argued that something analogous cannot apply to the Scharnhorst effect however, because there $n(\omega)$ is independent of position.[50] Scharnhorst and Barton ended their paper by an argument that Im n vanishes to order e^4. The dispersion relation then entails that the real part of the refractive index is constant to order e^4, that to this order we have $n(\omega) = n(0) = n(\infty)$. Then absorption and dispersion would start only at order e^6.

Yet they refrained from choosing between the two alternatives offered by the Kramers-Kronig relation. At the same time, they explained that they did not believe that a superluminal signal velocity would be at odds with special relativity:[51]

> We stress that such a speed greater than c between mirrors does not in any way contradict or pose any conceptual threats to special relativity, though admittedly it can prove eye-catching because at first sight one might think that it does. The presence of the mirrors breaks Lorentz invariance along the mirror normal (the mirrors define a preferred inertial frame), which obviates the arguments used in special relativity to prove that no signals can travel faster than light does in unbounded (Lorentz-invariant) space. By contrast, Lorentz invariance is unbroken parallel to the mirrors, and the light speed in these directions naturally must, and does, remain unchanged.[52]

---

[50] [72], [12], p.2044.

[51] In fact they had already made statements to this effect in 1990, although very briefly, merely mentioning that Lorentz invariance was broken in the direction normal to the mirrors due to the presence of the latter. [5], p.358, [6], p.559.

[52] [12], p.2038.

These remarks may seem somewhat vague in so far that they did not address how exactly the presence of the plates could be responsible for breaking Lorentz invariance; they may leave one with the impression that as far as Scharnhorst and Barton are concerned, any planar boundary conditions would break Lorentz invariance just as well. As we shall discuss below, later work by Liberati *et al* lent them more substance: the idea is that the plates lead the Casimir vacuum to behave like a medium, whose rest frame constitutes the "preferred inertial frame"[15].

One may of course wonder what assumptions hide behind the Kramers-Kronig relation. In fact it is a mathematical one more than a physical relation. It comes from applying Cauchy's residue theorem to a function of an imaginary variable. It therefore requires for that function to be analytic in the upper half plane of this variable. In the case of interest, this follows from the fact that the Fourier transform of the response function is analytic there. This in turn follows directly from the "causality condition" according to which the response must take place later than the disturbance: $\chi(t-t') < 0$ for $t < t'$.

Nevertheless the use of the Kramers-Kronig relation in the specific form Scharnhorst and Barton used has caused some concern, on the grounds that it "has been simplified, and is of the form used when a physical model for a material (like the Drude model) is used." Fearn expressed doubt whether "such a physical model applies to the vacuum."[53]

A specific worry concerns the assumption that the imaginary part of the index of refraction is odd.[54] Another has to do with the length scale associated with the scatterers:

---

[53] [73], [74], p.3.

[54] This comes in when changing the lower bound of integration from $-\infty$ (in the original relation) to 0. Fearn commented "Do we know it to be odd? This assumption comes from a physical model"; however she did not discuss which one. [74], p.4.

> Usually, when dispersion relations are employed for a material medium, some length scale is implied for the validity of the use of a refractive index to represent the material in "bulk". The wavelength of the light under observation must be much larger than the interatomic spacing for the refractive index approximation to be valid. No such "interatomic spacing" or equivalent length scale is immediately obvious in the case of the vacuum. There is no length scale specified in the original papers of Scharnhorst and Barton, and so we can assume any scale we wish.[55]

Although the latter statement is incorrect since Scharnhorst and Barton made clear that their work was only valid in the regime $1/L << \omega << m_e$, (hence $1/m_e << \lambda << L$), this does not invalidate the more general worry regarding the validity of an approach that makes use of classical concepts in order to treat the quantum vacuum. Of course, if this approach turns out to be fruitful (notably if its results were to be empirically verified, still a far away goal in the present case), the derivation of the effect from such classical concepts is precisely what would motivate describing the quantum vacuum as a dielectric medium.

3.6 Later work

3.6.1 Relation between the Scharnhorst effect and the energy density of the Casimir vacuum

Recall that Scharnhorst's work had been partly motivated by Rolf Tarrach's research on the speed of photons in a background electromagnetic field. This and other results for different physical environments motivated comparative studies, which attempted to identify a unifying principle behind the various effects. Notably,

---

[55] [74], p.5.

in 1995, Rolf Tarrach published a paper in collaboration with José Latorre and Pedro Pascual, in which they compared the speed of light in four different situations [20, 21]:

- an anisotropic vacuum due to a constant uniform background magnetic field.
- a homogeneous and isotropic thermal vacuum.[56]
- a gravitational background (Drummond-Hathrell effect). More precisely, they considered a (homogeneous and isotropic) Robertson-Walker gravitational background, with Friedmann cosmology.
- a Casimir vacuum (Scharnhorst effect).

In all four cases the speed of light is different from c, *i.e.* from what it is in an unbounded trivial vacuum in Minkowski space. In the first two this speed is less than c, but in the other two superluminality is predicted under some conditions (*i.e.* in certain directions and / or for some polarizations).

Latorre *et al*'s analysis highlighted the importance of the energy density of the vacuum. In his 1990 paper, Scharnhorst already mentioned that the space between Casimir plates was "a region of reduced vacuum energy density as compared with the usual vacuum (Casimir effect) with respect to the propagation of light," and also attributed the effect that bears his name to a change in the vacuum structure in this region.[57] Barton went further and ascribed the photons' increase in speed to this reduced vacuum energy density, also noting that an increase in temperature would have the opposite result [6]. Latorre *et al*'s work differed from these remarks by its quantitative character: the authors obtained (almost) the same equation for

---

[56] A thermal vacuum is a vacuum which, ideally, only contains electromagnetic radiation. The temperature of this vacuum is defined from the energy spectrum of this radiation: the radiation and vacuum are said to have temperature T when the energy spectrum of the radiation is that which would obtain in a vacuum chamber whose walls have temperature T.

[57] [5], p.354.

the photons speed of propagation in all four contexts.[58]

Latorre *et al* borrowed results for the speed of light previously obtained for the four different cases they meant to compare, and re-expressed them in terms of appropriate energy densities:

- for the background magnetic field:[59]

$$v_{\parallel} = 1 - \frac{8}{45}\alpha^2 \frac{\vec{B}^2}{m_e^4} \sin^2\theta \tag{3.35}$$

(polarization in the plane defined by $\vec{B}$ and the propagation direction),

$$v_{\perp} = 1 - \frac{14}{45}\alpha^2 \frac{\vec{B}^2}{m_e^4} \sin^2\theta \tag{3.36}$$

(polarization normal to the plane defined by $\vec{B}$ and the propagation direction), with an energy density of:

$$\rho_B = \frac{\vec{B}^2}{2}. \tag{3.37}$$

- for the thermal vacuum:

$$v = 1 - \frac{44\pi^2}{2025}\alpha^2 \frac{T^4}{m_e^4}, \tag{3.38}$$

with a density:

$$\rho_T = \frac{\pi^2}{15} T^4. \tag{3.39}$$

- for the FRW vacuum:

$$v = 1 + \frac{11}{45}\alpha\, G_N\, \frac{\rho + p}{m_e^2}, \tag{3.40}$$

where G_N is Newton's constant, and with a gravitational energy density $\rho_G = -\rho$, and $p = \frac{\rho}{3}$ assuming a radiation dominated universe, and $\rho + p > 0$.

---

[58] The equation for the gravitational case requires an additional assumption to be reduced to the same form, as explained below.

[59] [20], pp.61-64, [21], pp.2-5.

- for the Casimir vacuum:

$$v = 1 + \frac{11\pi^2}{8100}\alpha^2 \frac{1}{a^4 m_e^4}\cos^2\theta, \tag{3.41}$$

with a density one third that of the Casimir vacuum $\rho_a = -\frac{\pi^2}{720\,a^4}$, where a is the separation between the Casimir plates and θ the angle between their normal and the propagation direction.

Simply performing the appropriate substitutions, Latorre *et al* obtained the same expression for the background magnetic field, the thermal vacuum, and the Casimir effect:

$$v = 1 - \frac{44}{135}\,\alpha^2\,\frac{\rho}{m_e^4}. \tag{3.42}$$

For the gravitational case, it assumed the form:

$$v = 1 - \frac{44}{135}\,\alpha\,(G_N m_e^2)\,\frac{\rho_G}{m_e^4}, \tag{3.43}$$

i.e. it differed from Eq. (3.42) by a factor $\frac{G_N m_e^2}{\alpha}$, or, as Latorre *et al* remarked, would be identical to the other expressions if we were to set $G_N m_e^2$ equal to α.

Importantly, according to these results, whether v turns out to be larger or smaller than c hinges on whether the energy density of the vacuum is positive or negative.

In addition, Latorre *et al* saw further analogies between the thermal background case and the Scharnhorst effect. In the latter, the effect is ultimately due to the presence of the plates, that is, to spatial boundary conditions. Latorre *et al* noted that the velocity expression they borrowed for the Scharnhorst effect was the expression for the thermal background case with $\frac{i}{a}$ substituted for $2T$. They related this fact to the idea that "temperature corresponds to periodic boundary conditions in imaginary time."[60] They also noted that in order to recover the same pre-factors in both cases, one had to substitute a third of the Casimir vacuum density — in contrast in the thermal case the whole vacuum density is used. Far from deeming this an inconsistency, they attributed it to the Scharnhorst effect being associated with only one direction (normal to the plates), while temperature effects occur in all three dimensions.

Unlike Scharnhorst and Barton, Latorre *et al* referred to photons propagating in arbitrary directions, not just normal to the plates. However they merely assumed what form their velocity should have, by simply multiplying the correction to c found in Scharnhorst and Barton's result by the cosine of the angle between the propagation direction and the normal to the plates.[61] This velocity for photons propagating at oblique incidence was actually derived half a decade later by Stefano Liberati, Sebastiano Sonego, and Matt Visser, to whose work we now turn.[62]

3.6.2 Effect at oblique incidence and effective metric approaches

In 2001, Stefano Liberati, Sebastiano Sonego and Matt Visser rederived the Scharnhorst effect for photons travelling in arbitrary directions between the plates; the following year, they published a second paper in which they revisited their result but especially argued at length that superluminal signal velocity in the Casimir vacuum does not violate causality nor any requirement of special relativity [14, 13], [15, 16].

In their 2001 paper, Liberati *et al* meant to find out whether photons propagating in a Casimir vacuum could exhibit birefringence, *i.e.* whether photons with different polarizations could propagate at different speeds. Their interest in this question

---

[60] [20] p.76, [21] p.17.

[61] [20], p.61, [21], p.2.

[62] [14], [13]; [15] p.177-180, [16] pp.13-16.

arose in part from the gravitational analog of the Scharnhorst effect (the Hathrell-Drummond effect) where birefringence indeed occurs, and because they thought it significant in order to define an "effective metric."[63] Because birefringence could only occur for photons propagating obliquely, they studied the latter, and discovered a more general form of the Scharnhorst effect. Notably, the phase and group velocities of the photons are only equal in directions normal to (and parallel to) the plates; in arbitrary directions, they are both superluminal, but the group velocity is slightly larger.[64] Meanwhile, they aimed to determine the signal velocity from an effective metric. The idea goes as follows.[65] In covariant form, the dispersion relation they derived for a photon with 4-momentum k^{μ} is: $\gamma^{\mu\nu} k_{\mu} k_{\nu} = 0$. If it were not for the plates, the coefficients $\gamma^{\mu\nu}$ would correspond to the inverse of the Minkowski metric $\eta^{\mu\nu}$, so that $k_{\mu} k^{\mu} = 0$. In Casimir vacuum however they receive a correction and become $\gamma^{\mu\nu} = \eta^{\mu\nu} + \xi\, n^{\mu}\, n^{\nu}$, where n^{μ} is the unit vector orthogonal to the plates, and $\xi = \frac{11\,\pi^2 \alpha^2}{4050\; L^4\, m_e^4}$. Being a modified form of $\eta^{\mu\nu}$, $\gamma^{\mu\nu}$ can be viewed as the inverse

---

[63] Indeed, after establishing that birefringence does not occur in a Casimir vacuum, they commented:

"Absence of birefringence is crucial for the use of the "effective geometry" approach. Only if the propagation of light does not depend on its polarization and is thus, in a sense, universal, is it meaningful to describe it by a single effective metric. As a consequence of our analysis, photon propagation in the Casimir vacuum can indeed be phrased entirely in terms of the effective metric." [14], p.8. The reason that the Casimir vacuum is not birefringent, unlike other modified QED vacua, is that the expectation values linear in the field vanish, *i.e.* $\langle C\, |F^{\mu\nu}|\, C\rangle = 0$ where C refers to the vacuum state between Casimir plates. Only the quadratic (and higher) expectation values do not $\left\langle C\,\middle|\Omega^{\mu\nu\alpha\beta}\middle|\,C\right\rangle \neq 0$. However, Shore did use the concepts of effective light cone and effective metrics in the context of the Hathrell-Drummond effect. [75], p.26.

[64] They found the dispersion relation: $\omega^2 = \left|\vec{k}\right|^2 + \xi\, \left|\vec{k}\right|^2 \cos^2\phi$ where ϕ is the angle to the direction normal to the plates, and concluded that the phase velocity is $v_{phase}\,(\phi) = \left(1 + \xi\, \cos^2\phi\right)^{1/2}$and the group velocity $v_{group}\,(\phi) = \left(\frac{1+\left(2\,\xi+\xi^2\right)\cos^2\phi}{1+\xi\;\cos^2\phi}\right)^{1/2}$. [14], pp.5-6.

[65] [15], Section 3.2.1.

of an effective metric, itself given by:

$$g_{\mu\nu} = \eta_{\mu\nu} - \frac{\xi}{1+\xi} n_\mu n_\nu . \tag{3.44}$$ [66]

To this effective metric is associated a different light cone than to $\eta_{\mu\nu}$, a little wider (except in the direction parallel to the plates), corresponding to the fact that "c_{light}" is slightly larger than c.[67]

What they meant by "c_{light}" is the ratio $\frac{|\vec{x}|}{t}$ where $|\vec{x}|$ is the distance between two points, and t the time taken by the *wavefront* to travel this distance, both measured by an observer at rest with respect to the plates. Recall that the signal velocity is that of the wavefront, so this quantity is meant to represent the signal velocity, seen from the rest frame of the plates.[68] Noting that the wavefront obeys the equation $-ct + (g_{ij}\, x^i\, x^j)^{1/2} = 0$, Liberati *et al* obtained for the travel time:

$$ct = \left(g_{ij}\, x^i\, x^j\right)^{1/2} = \left(x^2 - \frac{\xi}{1+\xi}\,(n \cdot x)^2\right)^{1/2} = |x| \left(1 - \frac{\xi}{1+\xi}\cos^2\phi\right)^{1/2} . \tag{3.45}$$

This yields:

$$c_{light}\,(\phi) = \frac{|\vec{x}|}{t} = c \left(\frac{1+\xi}{1+\xi\,\sin^2\phi}\right)^{1/2} . \tag{3.46}$$ [69]

---

[66] This is obtained by inverting $\gamma^{\mu\nu}$. [15], pp.177-179.

[67] Recall that in Minkowski spacetime light obeys $c^2\,dt^2\,dx^2$, which results in the equation for the light cone $t = \frac{|\vec{x}|}{c}$. Clearly a larger value in the denominator gives a smaller slope so a more "open" light cone. $c^2\,dt^2 = dx^2$ itself comes from setting the interval $ds^2 = \eta_{\mu\nu}\,dx^\mu\,dx^\nu$ to zero (since we are dealing with light) so that: $\eta_{00}\,dt^2 + \eta_{11}\,dx^{(1)2} + \eta_{22}\,dx^{(2)2} + \eta_{33}\,dx^{(3)2} = -c^2\,dt^2 + dx^{(1)2} + dx^{(2)2} + dx^{(3)2} = 0$. Using $\mathrm{g}_{\mu\nu}$ instead of $\eta_{\mu\nu}$ (and considering only the $x^{(1)}$ direction for $n^\mu = (0,1,0,0)$) one would get instead of $c^2\,dt^2 = dx^2$: $c^2\,dt^2 = \frac{\xi}{1+\xi}\,dx^2$, hence $t = \sqrt{\frac{\xi}{1+\xi}}\,\frac{|\vec{x}|}{c}$. Since $\xi > 0$, $\sqrt{\frac{\xi}{1+\xi}} < 1$ and this corresponds to a wider light cone, in the $\mathrm{x}^{(1)}$ direction.

[68] However, as we discuss below, the idea that this quantity represents the actual signal velocity can be subject to criticism.

[69] Where ϕ is the angle to the normal. Note that $c_{light}\,(\phi) > c$ since $\xi > 0$ (recall $\xi = \frac{11\,\pi^2\alpha^2}{4050\;L^4\,m_e^4}$).

They illustrated the situation by the following diagram, where the closed curve centered at P represents the wavefront, hence the set of points reached after a time t by the signal travelling at speed c_{light}, for a pulse of light emitted at P at $t_0 = 0$.

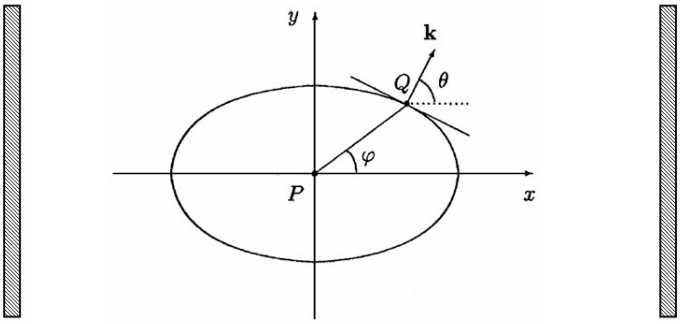


Figure 3.3: Wavefront (or line of constant phase) for a light signal emitted at the point P.

They pointed out that in general, the direction of propagation PQ is not parallel to the wave vector $\vec{k}$: except for propagation parallel and normal to the plates, the two directions form the non vanishing angle $\theta - \phi$.

Liberati *et al* also discussed another important aspect of this propagation speed: *i.e.* how c_{light} behaves in different frames. They argued that it is not frame independent, in significant contrast to c. For an observer moving with a 4-velocity u^{μ} with respect to the plates, the signal velocity of Scharnhorst photons would be:

$$c^{(u)}_{light} = c \lim_{\delta\tau \to 0} \left(1 + \frac{\xi}{1+\xi} \left(\frac{n_{\mu}\, \delta\, x^{\mu}}{n_{\nu}\, \delta\, x^{\nu}} \right)^2 \right)^{1/2} \quad . \tag{3.47}$$

They commented:

> The important points about [Eq. 3.47] are that $c^{(u)}_{light} \geq c$ always, and that $c^{(u)}_{light}$ has only one value for each observer u^{μ}. The latter feature pre-

> vents the possibility of causal paradoxes, which require signals that travel with the same speed greater than c in two different reference frames.[70]

As was discussed in chapter 2, generally speaking the importance of the concept of effective metric lies in the notion that a signal whose propagation can be described by such a metric cannot violate causality in the sense that it cannot lead to a bilking argument-type paradox, for it cannot give rise to a closed signal curve (CSC).[71] Yet it must be stressed that, despite their deriving a signal velocity from an effective metric, the treatment offered by Liberati *et al* is not immune to the criticism leveled at Scharnhorst and Barton regarding the limitations of their analysis to the low frequency regime. Liberati *et al*'s result too is based on an effective Lagrangian only valid in this range, as they themselves point out:

> Unfortunately, while it is certainly true that our treatment is valid at high frequencies (with respect to those associated to the background scales), it nevertheless also requires $\omega << m_e$, the condition under which one can use the Lagrangian, so we have no direct information about the strict $\omega \rightarrow \infty$ limit.[72]

Without access to information about what goes on in this limit, the wavefront velocity that they are deriving is not really that required to speak of signal velocity: indeed, at finite frequencies, this wavefront is not "sharp", as Ben-Menahem pointed out with reference to Scharnhorst and Barton's original derivations.

---

[70] [15], p.180.

[71] [15], Section 3.2.2.

[72] [14], p.5.

Chapter 4

DERIVATION OF THE SCHARNHORST EFFECT IN SOFT-COLLINEAR EFFECTIVE THEORY

4.1 Introduction

In the present chapter we derive the velocity of a Scharnhorst photon within the framework of Soft Collinear Effective Field Theory (“SCET”). We first introduce SCET and explain how it can be used to model the Scharnhorst effect, as well as the rationale for doing so. Taking this approach requires the use of a modified form of the Euler-Heisenberg Lagrangian, which we obtain in the following section. We then derive the polarization tensor for the Scharnhorst photon $\Pi^{\mu\nu}$ to second-order. $\Pi^{\mu\nu}$ will then be used to compute the phase velocity of Scharnhorst photons.

The advantage of the SCET approach lies in the well-controlled character of the perturbative expansion used: SCET provides a self-consistent formalism to calculate corrections to the speed of light where higher order corrections are small compared to the former. What implications this has for the interpretation of the velocity obtained will be discussed in the Conclusion.

4.2 Modelling the Scharnhorst effect in the framework of SCET

As its name suggests, SCET (*i.e.* Soft-Collinear Effective Theory) is an effective quantum field theory. The defining feature of SCET compared to standard quantum field theories is that it separates energy and momentum scales by employing more degrees of freedom to model particles, based on their momentum: for each type of particle (say, photons), those that have high momentum in a preferred direction are modelled by a different field (said to be a "collinear" field) than those that do not (represented by a "soft" field). SCET is an approach normally applied in the context of QCD, however in the present case the relevant theory is QED, since we wish to describe the interactions of photons and electrons.

Because of the concept of collinear degrees of freedom, it has notably been used to treat the properties of jets. This is also what makes it attractive to calculate the Scharnhorst effect: it is natural to describe the propagating, Scharnhorst photon (*i.e.* the probe) whose velocity we seek by a collinear field, and the photons typical of the Casimir vacuum by "soft" ones. Hence compared to previous derivations of the Scharnhorst effect, the key feature of the present approach is that it models the propagating photon on the one hand, and the photons that form the Casimir vacuum on the other, by fields that are differentiated according to their momentum-energy scale.

4.2.1 The standard Euler-Heisenberg Lagrangian

The Lagrangian we use is derived from the standard Euler-Heisenberg Lagrangian: it is, so to speak, a SCET version thereof. Recall that the Euler-Heisenberg Lagrangian is an effective Lagrangian – and indeed, it is generally regarded as the first of this kind, and is deemed to have played a foundational role in the development of the effective theories program. More specifically, it constitutes an instance of the top-down approach, whereby a low-energy theory is derived from its more general counterpart, the "full theory." The strategy to do so consists in "integrating out" the degrees of freedom that represent particles whose rest mass are much larger than the typical energy scales that the effective theory is meant to deal with: these degrees of freedoms (hence the corresponding particles/fields) no longer *explicitly* appear in the formalism of the new, effective theory. Instead, their effect appears implicitly: the coefficients of the terms in the Lagrangian are modified. These changes are what embodies the presence of the higher energy particles *i.e.* their interactions with those the effective theory explicitly describes). In the specific case of the Euler-Heisenberg Lagrangian, the fields that have been integrated out are the fermions.[1] This is why its regime of validity is limited to fields of frequencies $\omega << m_e$, where m_e is the electron rest mass (the lightest fermion), as was stressed in chapter 3. Since QED only describes fermions and photons in the first place, the Euler-Heisenberg Lagrangian only involves one remaining degree of freedom: the photon field A^μ; we will usually be dealing with it in the form of the electromagnetic field tensor $F^{\mu\nu}$.[2] The Euler-Heisenberg Lagrangian is then given by:

$$\mathcal{L}_{EH} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{\alpha^2}{m_e^4} \left\{ c_1 \left(F^{\mu\nu} F_{\mu\nu} \right)^2 + c_2 \left(F^{\mu\nu} \widetilde{F}_{\mu\nu} \right)^2 \right\}, \tag{4.1}$$

where in natural units $c_1 = \frac{1}{(2)(3^2)(5)} = \frac{1}{(2)(45)}$ and $c_2 = \frac{7}{(2^3)(3^2)(5)} = \frac{7}{(8)(45)}$.[3]
Since the low-energy effective theory has less fields than the full theory, the way to describe a given process or interaction naturally differs in the two theories, involving different Feynman diagrams, based on different vertices. In the case of interest, in

---

[1] Electrons and positrons notably.

[2] Recall $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$.
Also note that $\widetilde{F}_{\mu\nu}$ does not constitute a different field *per se*, it is merely a way to denote $\widetilde{F}_{\mu\nu} \equiv \frac{1}{2} \epsilon_{\mu\nu\alpha\beta} F^{\alpha\beta}$.

[3] The Euler-Heisenberg Lagrangian is most often given in terms of the electric and magnetic fields $\vec{E}$ and $\vec{B}$. In this case, in natural units:

$$\mathcal{L}_{EH} = \frac{1}{2}(\vec{E}^2 - \vec{B}^2) + \frac{2\,\alpha^2}{45\,m_e^4} \left\{ \left(\vec{E}^2 - \vec{B}^2 \right)^2 + 7 \left(\vec{E} \cdot \vec{B} \right)^2 \right\}.$$

In natural units $\alpha = \frac{e^2}{4\pi}$ *i.e.* $\alpha^2 = \frac{e^4}{2^4 \pi^2}$, so $\frac{2\,\alpha^2}{45\,m_e^4} = \frac{1}{(2^3)(3^2)(5)\pi^2} \frac{e^4}{m_e^4} = \frac{1}{(8)(45)\pi^2} \frac{e^4}{m_e^4}$ so the Euler Heisenberg can also be written as:

$$\mathcal{L}_{EH} = \frac{1}{2}(\vec{E}^2 - \vec{B}^2) + \frac{1}{(8)(45)\pi^2} \frac{e^4}{m_e^4} \left\{ \left(\vec{E}^2 - \vec{B}^2 \right)^2 + 7 \left(\vec{E} \cdot \vec{B} \right)^2 \right\}.$$

QED interactions are always between two fermions and one photon, represented diagrammatically by the following vertex:

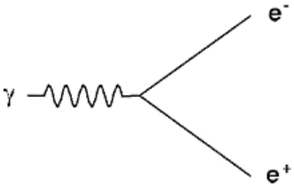


Figure 4.1: Vertex in QED

In contrast in the effective theory photons interact directly with one another. The relevant process for the Scharnhorst effect is light-by-light scattering — and it also happens to be one of the most interesting features of quantum theory compared to Maxwell's classical theory of electromagnetism, because the latter predicts no such phenomenon. In QED the lowest order process that accounts for light-by-light scattering is the box diagram, which involves four fermions. In the effective theory described by the Euler-Heisenberg Lagrangian, this diagram becomes a contact term,

---

This is the form used by Scharnhorst.

In order to express $\mathcal{L}_{EH}$ in terms of $F^{\mu\nu}$ and $\widetilde{F}_{\mu\nu}$ we use:

$$\begin{aligned} \vec{E}^2 - \vec{B}^2 &= -\frac{1}{2} F^{\mu\nu} F_{\mu\nu} \\ (\vec{E}^2 - \vec{B}^2)^2 &= \frac{1}{4} \left(F^{\mu\nu} F_{\mu\nu}\right)^2 \\ \vec{E} \cdot \vec{B} = -\frac{1}{4} F^{\mu\nu} \widetilde{F}_{\mu\nu} \qquad \text{therefore:} \qquad (\vec{E} \cdot \vec{B})^2 &= \frac{1}{16} \left(F^{\mu\nu} \widetilde{F}_{\mu\nu}\right)^2 . \end{aligned}$$

Then:

$$\begin{aligned} \mathcal{L}_{EH} &= -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{1}{(8)(45)\,\pi^2} \frac{e^4}{m_e^4} \left\{ \frac{1}{4} \left(F^{\mu\nu} F_{\mu\nu}\right)^2 + \frac{7}{16} \left(F^{\mu\nu} \widetilde{F}_{\mu\nu}\right)^2 \right\} \\ \Longleftrightarrow \mathcal{L}_{EH} &= -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{2\,\alpha^2}{45\, m_e^4} \left\{ \frac{1}{4} \left(F^{\mu\nu} F_{\mu\nu}\right)^2 + \frac{7}{16} \left(F^{\mu\nu} \widetilde{F}_{\mu\nu}\right)^2 \right\} . \end{aligned}$$

This means that $c_1 = \frac{2}{45}\frac{1}{4} = \frac{1}{(2)(45)} = \frac{1}{(2)(3^2)(5)}$ and $c_2 = \frac{2}{45}\frac{7}{16} = \frac{7}{(8)(45)} = \frac{7}{(2^3)(3^2)(5)}$.

with four photons directly interacting with one another:

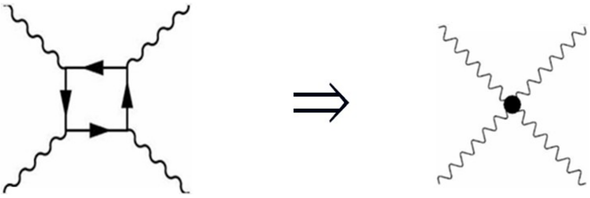


Figure 4.2: The box diagram in QED becomes a contact term in the theory described by $\mathfrak{L}_{EH}$.

Now the physical situation of interest to us is the propagation of a photon. In a theory involving only the photon field, the Feynman diagrams representing the various interactions that affect such a propagating photon look as follows:

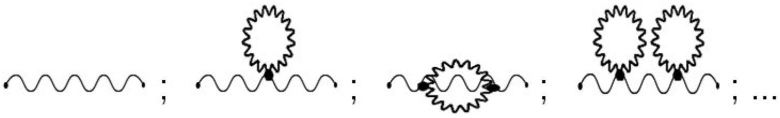


Figure 4.3: Some of the interactions affecting a propagating photon in a theory with only photon fields.

The two terms that appear in $\mathfrak{L}_{EH}$ (Eq.(4.1)) represent the lowest orders of an expansion in the parameters α^2 and $\frac{1}{m_e^4}$. More accurately, the theory based on $\mathfrak{L}_{EH}$ is based on two separate expansions (as we shall soon discuss), and neglects all higher order terms in both of these. The only terms that remain beside the leading order, kinetic term, are $c_1 \frac{\alpha^2}{m_e^4} \left(F^{\mu\nu}F_{\mu\nu}\right)^2$ and $c_2 \frac{\alpha^2}{m_e^4} \left(F^{\mu\nu}\widetilde{F}_{\mu\nu}\right)^2$, which represent an interaction between four different fields, *i.e.* what is illustrated on the right of Fig. (4.2). This implies that when expressing the photon propagator in the language of

Feynman diagrams, the only ones relevant at the order we are working are:

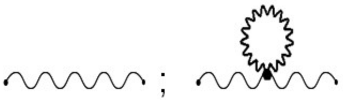


Figure 4.4: Feynman diagrams in the theory described by the standard Euler-Heisenberg Lagrangian. The diagram on the left represents the leading order, kinetic term, the one on the right accounts for interactions, and is known as the "tadpole diagram."

Neglecting in this way higher order terms in $\frac{\alpha^2}{m_e^4}$ involves neglecting higher order terms from two separate expansions: one in $\alpha = \frac{e^2}{4\pi}$ where e is the charge of the electron, and one in $\frac{E}{m_e}$, where E is the energy scale of the fields of interest and m_e the electron mass. Note that both are dimensionless with respect to energy.

The Euler-Heisenberg theory inherits the expansion in α from the perturbative nature of the coupling in QED. However it is not a feature that makes $\mathcal{L}_{EH}$ an effective theory with respect to another, more realistic one.

In contrast, the expansion in $\frac{E}{m_e}$ is the result of integrating out the electron loop from QED. It is what makes the Euler-Heisenberg theory an effective theory of QED in the sense of it being the low-energy theory derived from QED as the full theory. For this reason, neglecting terms that are higher order in $\frac{E}{m_e}$ is an approximation valid only so long as the photons being described have energies $E << m_e$ – *i.e.* frequencies $\omega << m_e$ as discussed in chapter 3. However for greater frequencies, higher order terms are not necessarily small enough compared to those in Eq. (4.1) to be negligible. This is the reason why Scharnhorst and Barton state that their result for the phase velocity of Scharnhorst photons does not give direct information about their signal velocity, which is the $\omega \to \infty$ limit of the phase velocity.[4]

---

[4] [12], see notably p.3.

The expansion in $\frac{E}{m_e}$ is related to what fields can appear in each term, and to what power they do. This is due to dimensional considerations. In QFT the dimensional basis used consists of only one base unit, that of mass-energy. The action of a theory is dimensionless – which naturally entails that every term in the action must be. When we consider the leading order, kinetic term, this implies that $F^{\mu\nu}$ must be of dimension 2.[5] When we then consider the first interaction term in the expansion, the fact that the factor of $\frac{1}{m_e}$ is to the power 4 then implies that the term must contain four factors of $F^{\mu\nu}$ (or $\widetilde{F}_{\mu\nu}$ which has the same dimensions).[6]

However the two expansions, *i.e.* the perturbative one in α and the power one in $\frac{E}{m_e}$, are not related to each other. Notably, the reasons for neglecting higher order terms are different for each.

4.2.2 Degrees of freedom in the SCET Euler-Heisenberg Lagrangian

The SCET counterpart of the standard Euler-Heisenberg Lagrangian also involves only photon fields. However it differentiates between the collinear and the soft types thereof.[7]

---

[5] This can be seen as follows (we use square brackets to denote dimensions, *i.e.* $[A] = E^n$ stands for "A has dimensions n in energy".

$S = \int d^4x \, \mathcal{L}_{EH}$ with: $[S] = E^0$ $[x] = E^{-1}$

therefore: $[\mathcal{L}_{EH}] = E^4$,

which holds for each term of $\mathcal{L}_{EH}$. Applied to the leading order term this means:

$\left[-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}\right] = E^4$ so $[F^{\mu\nu}] = E^2$ as advertised.

[6] Indeed, as discussed in the previous footnote:

[all terms in $\mathcal{L}_{EH}] = E^4$.

$\left[\frac{1}{m_e^4}\right] = E^{-4}$, so terms containing $\frac{1}{m_e^4}$ must have this factor multiplied by an expression of dimensions E^8. Since $[\alpha] = [c_i] = E^0$ it must be the operator itself that has dimensions E^8. By considering the leading order term, we found that $[F^{\mu\nu}] = E^2$, therefore there must be 4 factors of $F^{\mu\nu}$ (or of $\widetilde{F}_{\mu\nu}$) in the terms of order $\frac{1}{m_e^4}$.

[7] That is, the field $\mathcal{F}_c^{\mu\nu}$ is the collinear electromagnetic field tensor, and $\mathcal{F}_s^{\mu\nu}$ the soft one, defined with respect to the corresponding photon fields in the usual way:

$$\mathcal{F}_c^{\mu\nu} = \partial^\mu A_c^\nu - \partial^\nu A_c^\mu, \tag{4.2}$$

$$\mathcal{F}_s^{\mu\nu} = \partial^\mu A_s^\nu - \partial^\nu A_s^\mu. \tag{4.3}$$

As briefly mentioned above, the Scharnhorst photon is modelled by a collinear field, and the Casimir vacuum by soft photons.

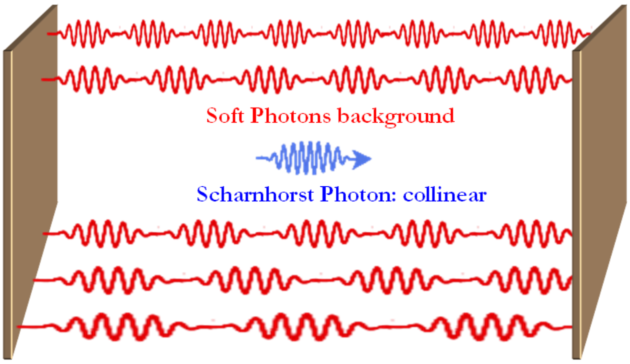


Figure 4.5: Degrees of freedom in the Scharnhorst effect modeled in SCET.

Two considerations motivate this choice: collinear fields are meant to describe particles that have high momentum in a preferred direction, so both the directionality and the magnitude of the momenta come in. More precisely, this means that the particle must have, respectively:

1) one component of its momentum much larger than the others.

2) much greater momentum than the other particles present in the physical situation one aims to describe.

The reverse must hold for soft fields, of course.

Regarding the first criterion, it is immediately obvious that the momentum of the Scharnhorst photon has preferred direction whereas the photons of the Casimir vacuum do not.

The second criterion requires more careful attention however. It states that the Scharnhorst photon must have much greater momentum than the photons of the

---

Casimir vacuum. This requirement can be reformulated in terms of energies, provided all photons concerned are only slightly off-shell, i.e. that they almost obey:

$$k^2 = E^2 - \vec{k}^{\;2} = 0, \tag{4.4}$$

for in this case their energy is of the same order as their momentum. So the propagating photon needs to be highly energetic compared to the background photons. Strictly speaking this is not the case: the region between the plates actually contains both low and high energy photons. However there are good reasons for only including low energy ones in our model. Indeed, the plates have a finite plasma frequency, above which they become transparent (the plates being metallic, the plasma frequency is in the UV range). This means that photons above that energy are not affected by them, and behave as in the usual, unbounded, trivial vacuum. Therefore these higher energy photons are not part of the *Casimir* vacuum *per se*, and do not contribute to the Scharnhorst effect. For the purpose of modelling the latter, we should therefore exclude them from our model.[8] All this requires is to consider a Scharnhorst photon of frequency much greater than the plasma one.

To sum up, the physics of the propagating photon and the Casimir vacuum belong to different energy scales (to which we shall refer as E and ϵ respectively), in such a way that:

- typical energy ϵ of the Casimir vacuum:

$$\tfrac{1}{L} << \epsilon << \omega_{plasma} \quad \rightarrow \quad \text{soft field}$$

- typical energy E of the propagating photon:

$$\omega_{plasma} << E << m_e \quad \rightarrow \quad \text{collinear field}$$

---

[8] Derivations of the Casimir force exclude them not from the start like is done here, but by computing the energy of all the modes, then subtracting the energy due to the unbounded vacuum.

On the basis of both these conditions of directionality and relative energy, collinear fields constitute a good model for the Scharnhorst photons and soft fields a good model for those that form the Casimir vacuum.[9]

In terms of Feynman diagrams, this means that the theory based on the SCET Euler-Heisenberg Lagrangian involves the following diagrams:

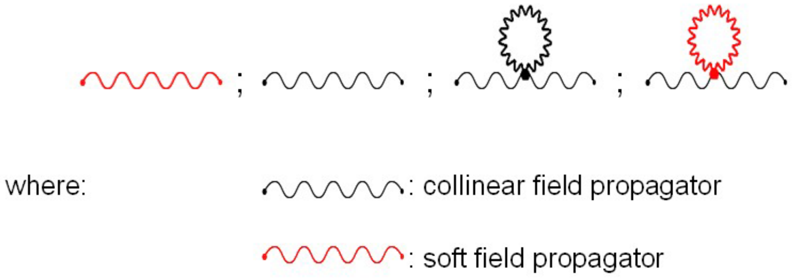


Figure 4.6: Feynman diagrams in the theory described by the SCET Euler-Heisenberg Lagrangian.

In order to model the Scharnhorst effect, we will only need two of these:

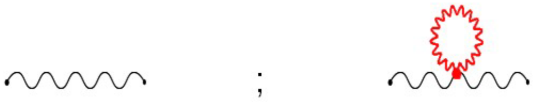


Figure 4.7: Feynman diagrams used to account for the Scharnhorst effect within the theory based on the SCET Euler-Heisenberg Lagrangian.

---

[9] Because the effect occurs in a Casimir set-up, and the Casimir force is often presented as direct evidence for "virtual" particles (see chapter 5), it may be useful to note that the distinction between "real" and "virtual" particles does not enter at the level of what degrees of freedom we choose to represent the physical entities of interest; it is not a trait that is input "by hand" in the model, as it were. This is not a specific feature of our approach or of SCET specifically. The distinction between "real" and "virtual" is something that arises naturally in quantum field theories in general: it turns out that the "particles" described by higher order, so-called "radiative" corrections do not obey the relativistic mass-energy relation; it is this trait that constitutes the actual definition of a "virtual particle", or rather, as they are more commonly called, "off-shell" particles; in contrast "real" ones are said to be "on (the mass) shell."

We do not need the soft photon propagator because the only field that propagates in our model is the collinear one; and we shall not deal with the tadpole diagram whose loop is collinear because any high energy photon from the background is not truly part of the Casimir vacuum (as it is not affected by the plates).

In these diagrams, the free propagators for the collinear fields are just the usual, trivial vacuum photon propagators. In contrast, the one that forms the soft loop is subject to the boundary conditions imposed by the plates: this is how their presence is accounted for in our approach, or, put another way, how the fact that the (Scharnhorst) photon is propagating through a Casimir vacuum instead of a normal one is modelled.

4.3 Advantages of SCET *vs* the standard approach

Our motivation for using the framework of SCET is that it allows us to systematically factor the probe physics from the Casimir background, and as a result the perturbative expansion it involves is well-controlled, *i.e.* higher order terms are power suppressed in the expansion coefficients, and their numerical coefficients are of order 1, so that the corrections they contribute are small compared to the leading order ones. Because the Lagrangian we use in our SCET calculation is derived from the Euler-Heisenberg Lagrangian, it too involves only the lowest order terms of the expansions in α and $\frac{1}{m_e}$. The crucial difference is that SCET also involves another perturbative expansion, where the power-counting parameter is a ratio of momenta. Before we can discuss this however, we must first specify our coordinate system. As often in SCET, the most convenient for our purposes are light cone coordinates.

4.3.1 Light-cone coordinates

Light-cone coordinates are coordinates in four-dimensional, Minkowski spacetime. Their defining feature is that one of the basis vectors is in the direction of a photon in (trivial *i.e.* normal, unbounded) vacuum: $n^\mu = (1, 0, 0, 1)$. Note that with this choice $n^2 = 0$, *i.e.* n^μ is indeed a light-like vector. Generally speaking, light-cone coordinates are natural coordinates for particles moving at high relativistic speeds (*i.e.* whose kinetic energy is much larger than their rest-mass), because their four-momentum is close to the direction of n^μ ("close to the light-cone"). Because collinear fields represent such particles, SCET typically uses light-cone coordinates. A second basis vector is chosen to be also light-like, and usually taken in the opposite spatial direction to n^μ: $\bar{n}^\mu = (1, 0, 0, -1)$.[10] The other two are taken normal to n^μ and $\bar{n}^\mu$ as well as to one another.

In these coordinates an arbitrary four vector can be written $p^\mu = (p^+, p^-, \vec{p}_\perp)$. This corresponds to the so-called "Sudakov" decomposition:

$$\begin{aligned} p^\mu &= \frac{n^\mu}{2}\,\bar{n}.p + \frac{\bar{n}^\mu}{2}\,n.p + p_\perp^\mu \\ &= \frac{n^\mu}{2}\,p^- + \frac{\bar{n}^\mu}{2}\,p^+ + p_\perp^\mu, \end{aligned} \tag{4.5}$$

where $p^+ \equiv \bar{n}.p$ is the component in the n^μ direction, $p^- \equiv n.p$ is the component in the $\bar{n}^\mu$ direction, and $p_\perp^\mu = (0,\, p^1,\, p^2,\, 0)$ is two-dimensional.[11]

Also, note that the relativistic invariant p^2 takes the form:

$$p^2 = p^+ p^- + (p_\perp^\mu)^2 = p^+ p^- - (\vec{p}_\perp^{\,\mu})^2, \tag{4.6}$$

---

[10] Note that $\bar{n}^2 = 0$. This is actually not the only possible choice, but it is a standard one. Also, $n.\bar{n} = 2$ constitutes the normalization convention.

[11] Note that $p_\perp^2 = -\vec{p}_\perp^{\,2}$ where $p_\perp^2$ is in Minkowski space and $\vec{p}_\perp^{\,2}$ in Euclidean space.

a result which will be useful in the next section.

4.3.2 SCET: scaling in the expansion parameter λ

In distinguishing between soft and collinear fields, SCET separates the degrees of freedom into a high momentum region and a low momentum region. Therefore, the ratio of soft to collinear momenta is very small. It is this ratio that serves as the parameter, λ, of the perturbative expansion.

In order to show the advantages of this approach, we first need to determine how the various quantities of interest (notably, those appearing in the Lagrangian) scale with λ.

Let us first consider a photon that propagates in the z-direction (*i.e.* in the x^3 direction), and call its energy E. If it is on-shell, by definition it obeys the relativistic relation:

$$p^2 = E^2 - \vec{p}^{\,2} = m^2, \tag{4.7}$$

which for photons specifically takes the form:

$$k^2 = E^2 - \vec{k}^{\,2} = 0, \tag{4.8}$$

so that $E^2 = \vec{k}^{\,2}$. Therefore in this case the momentum of the photon is E as well. This is true of a photon propagating in trivial vacuum.

Now in contrast, for a photon propagating through a Casimir vacuum, the situation is slightly more complex. Being collinear, most of its energy is associated with the p^+ component of its momentum.[12] However the photons of the Casimir vacuum

---

[12] If it were exactly on-shell and moving exactly in the z- direction, all of its momentum would be p^+.

that contribute to the Scharnhorst effect interact with it. Let us represent their typical energy scale by ϵ, which is much smaller than that of the collinear photon, E. Because most of these background photons are close to being on-shell, on average their momentum k is also of order ϵ (it does not have any preferred directionality, so this is true for all components of this momentum).[13] Interactions with the soft background photons take the propagating photon slightly off-shell, by an amount whose order of magnitude is the same as the typical energy of the background, *i.e.* ϵ. This perturbation gives the probe, Scharnhorst photon a momentum component p^- of order ϵ, as this is the smallest energy scale in the problem and p^- is the smallest component of the propagating photon. This means that the probe photon can acquire a small invariant mass on the order of ϵ. So ϵ constitutes both a measure of the "off-shellness" of our probe photon, and the energy and momentum scale of the soft photons that cause it. Because this is only a small perturbation however, the propagating photon remains very close to being on shell, and its main momentum component p^+ is of the same order as its energy, *i.e.* E. To sum up, the momentum components discussed so far are characterized by:

$$p^+ \sim E \tag{4.9a}$$

$$p^- \sim \epsilon \tag{4.9b}$$

$$p_\perp \sim \sqrt{\epsilon E} \tag{4.9c}$$

$$k \sim \epsilon. \tag{4.9d}$$

What we actually are interested in is how the components of momentum scale with λ, since this is our intended power counting parameter. By definition λ is the ratio of the typical momentum of the background soft photons, to the typical momentum

---

[13] In order to distinguish their momenta in the notation, we use p to represent the momentum of the propagating photon and k for the background photons.

of the Scharnhorst, probe photon. Because typically these photons are almost on shell, their energies are roughly of the same size as their momenta, so λ can be expressed as a ratio of energy scales as well:

$$\lambda \equiv \sqrt{\frac{\epsilon}{E}}, \tag{4.10}$$

Knowing how the momentum components scale with respect to λ tells us how they compare in size to our energy scale of reference, E.

Now given Eq. (4.9) – *i.e.* how the momentum components scale with respect to E and ϵ – together with Eq. (4.6) – the expression for p^2 in terms of these components – one can determine how they scale with respect to λ. Indeed homogeneous power counting requires that all the terms in Eq. (4.6) must be of the same order in λ. Notably, p^+p^- must be of the same order as $(\vec{p}_\perp^{\ \mu})^2$. Now we already know:[14]

$$\vec{p}_\perp^{\ \mu} \sim \ \sqrt{\epsilon E} \sim \ E\,\lambda \quad \Rightarrow \quad (\vec{p}_\perp^{\ \mu})^2 \sim \ E^2\,\lambda^2. \tag{4.11}$$

Therefore we also need to have:

$$p^+p^- \sim \ E^2\,\lambda^2. \tag{4.12}$$

We also know that the p^+ momentum component is of order E:

$$p^+ \sim \ E. \tag{4.13}$$

Therefore we get for p^-:

$$p^- \sim \ E\,\lambda^2 \tag{4.14}$$

---

[14] The second relation simply comes from the definition of λ: $\lambda \equiv \sqrt{\frac{\epsilon}{E}}$.

Incidently, this in turn means that $p^- \sim\ \epsilon$.

These result in the following behavior with respect to λ: [15]

$$p^+ \sim\ E \qquad \Rightarrow \qquad p^+\ \sim\ E\ \lambda^0\ \sim\ O(\lambda^0) \tag{4.15a}$$

$$\vec{p}_\perp^{\,\mu} \sim\ \sqrt{\epsilon E} \qquad \Rightarrow \qquad \vec{p}_\perp^{\,\mu}\ \sim\ E\ \lambda\ \ \sim\ O(\lambda^1) \tag{4.15b}$$

$$p^- \sim \epsilon \qquad \Rightarrow \qquad p^-\ \sim\ E\ \lambda^2\ \sim\ O(\lambda^2). \tag{4.15c}$$

In fact, this is taken to be the actual definition of a collinear field: a field is deemed collinear when its momentum components scale in this way with respect to the power counting parameter.

In contrast the momentum components of the background, soft photons all scale the same way with λ. Recall that because they do not preferentially propagate in a specific direction, all momentum components are of the same order in ϵ: $k^\mu \sim \epsilon$. Consequently:

$$k^\mu \sim \epsilon \qquad \Rightarrow \qquad k^\mu \sim E\ \lambda^2 \sim\ O(\lambda^2). \tag{4.16}$$

From how the different momentum components scale, we can figure out how the spatial components of the collinear fields and the soft fields scale, and from these how the corresponding measures scale. Then knowing the latter as well as the behavior of the kinetic terms, the scaling of the fields themselves can be determined.

Let us first discuss the collinear field, $F_c^{\mu\nu}$ (which includes $\widetilde{F}_c^{\mu\nu}$).[16] In light-cone coordinates, its modes take the form:

$$e^{i(p.x)} \qquad \Rightarrow \qquad e^{\,i\,(p^+x^-\,+p^-x^+\,+p_\perp x_\perp)}. \tag{4.17}$$

---

[15] Again, how a quantity scales in λ is way to express its relative size. Because $\lambda << 1$, a quantity $\sim\ O(\lambda^2)$ is *smaller* than one that obeys $\sim\ O(\lambda^1)$.

[16] Recall $\widetilde{F}^c_{\mu\nu}$ is not really a different field from $F^c_{\mu\nu}$, but a shorthand notation for $\frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F_c^{\alpha\beta}$.

The field is dominated by the modes that have terms $O(\lambda^0)$ in the exponent. Since we already know how the momentum components scale in λ, this allows us to find out how the spatial components, and ultimately how the measure, d^4x_c, scale in λ as well.

For instance let us consider the term $p^- x^+$. The modes that dominate its contribution are those for which $p^- x^+ \sim O(\lambda^0)$. We know that $p^- \sim O(\lambda^2)$, *i.e.* $p^- \sim E\,\lambda^2$. The corresponding scaling for x^+ is $x^+ \sim \frac{1}{E}\,\lambda^n$, where n is the integer needed so that $p^- x^+ \sim \lambda^0$. Clearly $n = 2$, so that for the dominant modes $x^+ \sim O(\lambda^{-2})$.

In that way we find:

$$p^+ \sim O(\lambda^0) \;\;\Rightarrow\;\; x^- \sim O(\lambda^0) \tag{4.18a}$$

$$p_\perp \sim O(\lambda^1) \;\;\Rightarrow\;\; x_\perp \sim O(\lambda^{-1}) \tag{4.18b}$$

$$p^- \sim O(\lambda^2) \;\;\Rightarrow\;\; x^+ \sim O(\lambda^{-2}). \tag{4.18c}$$

Since dx^-, dx^+, $dx_\perp$ scale the same way as x^-, x^+, $x_\perp$, this implies that for the collinear field, the measure scales as:

$$d^4x_c = dx^-\; dx^+\; d^2x_\perp \sim O(\lambda^{-4}). \tag{4.19}$$

Meanwhile, in order for the kinetic terms in the action to represent the propagation of the field as we expect them to (as opposed to accounting for perturbations to the propagation), they need to be of leading order in the perturbative expansion. Therefore they must be of order 1, *i.e.* $O(\lambda^0)$:

$$\int d^4x\; F^{\mu\nu}F_{\mu\nu} \sim O(\lambda^0). \tag{4.20}$$

This holds true for both the collinear and the soft fields.

Taken together, Eq.(4.19) and Eq.(4.20) imply:

$$\int d^4x_c \; F_c^{\mu\nu} F^c_{\mu\nu} \sim O(\lambda^0) \qquad \Rightarrow \qquad F_c^{\mu\nu} F^c_{\mu\nu} \sim O(\lambda^4). \tag{4.21}$$

Therefore the collinear field $F_c^{\mu\nu}$ scales as:

$$F_c^{\mu\nu} \sim O(\lambda^2). \tag{4.22}$$

We now follow the same reasoning for the soft field $F_s^{\mu\nu}$ (or $\widetilde{F}_s^{\mu\nu}$). It contains modes of the form $e^{i(k^\mu \, x_\mu)}$.[17] Again, the field is dominated by $O(\lambda^0)$ modes. Recall that $k^\mu \sim O(\lambda^2)$, so these modes have to satisfy:

$$k^\mu \sim O(\lambda^2) \;\; \Rightarrow \;\; x^\mu \sim O(\lambda^{-2}). \tag{4.23}$$

Consequently the soft measure scales as:

$$x^\mu \sim O(\lambda^{-2}) \qquad \Rightarrow \qquad dx^\mu \sim O(\lambda^{-2}) \qquad \Rightarrow \qquad d^4x_s \sim O(\lambda^{-8}). \tag{4.24}$$

Knowing that the kinetic term is $O(\lambda^0)$ then entails:

$$\int d^4x_s \; F_s^{\mu\nu} F^s_{\mu\nu} \sim O(\lambda^0) \qquad \Rightarrow \qquad F_s^{\mu\nu} F^s_{\mu\nu} \sim O(\lambda^8), \tag{4.25}$$

and the soft field $F_s^{\mu\nu}$ scales as:

$$F_s^{\mu\nu} \sim O(\lambda^4). \tag{4.26}$$

---

[17] Because it does not propagate in a preferred direction, there is no point in using light-cone coordinates: the scaling is the same for all directions.

As discussed above the kinetic terms in the Lagrangian are, of necessity, $O(\lambda^0)$. Now as we shall see in the next section, to order $\frac{\alpha^2}{m_e^4}$ the terms in the SCET Euler-Heisenberg Lagrangian are products of two factors of the form:

$$\mathcal{F}_c\,\mathcal{F}^c \quad \sim \quad \mathcal{F}_c\,\widetilde{\mathcal{F}}^c \quad \sim \quad O(\lambda^4) \tag{4.27a}$$

$$\mathcal{F}_s\,\mathcal{F}^s \quad \sim \quad \mathcal{F}_s\,\widetilde{\mathcal{F}}^s \quad \sim \quad \dot{O}(\lambda^8). \tag{4.27b}$$

That is, our Lagrangian contains terms of the following structure:[18]

$$(\mathcal{F}_c\,\mathcal{F}^c)^2 \quad \sim \quad (\mathcal{F}_c\,\widetilde{\mathcal{F}}^c)^2 \quad \sim \quad O(\lambda^8) \tag{4.28a}$$

$$(\mathcal{F}_c\,\mathcal{F}^c)(\mathcal{F}_s\,\mathcal{F}^s) \quad \sim \quad (\mathcal{F}_c\,\widetilde{\mathcal{F}}^c)(\mathcal{F}_s\,\widetilde{\mathcal{F}}^s) \quad \sim \quad O(\lambda^{12}) \tag{4.28b}$$

$$(\mathcal{F}_s\,\mathcal{F}^s)^2 \quad \sim \quad (\mathcal{F}_s\,\widetilde{\mathcal{F}}^s)^2 \quad \sim \quad O(\lambda^{16}). \tag{4.28c}$$

Again, these are the type of terms appearing at order $\frac{\alpha^2}{m_e^4}$, which are the only ones we consider.

In fact, because the Casimir vacuum photons are soft, and Eq. (4.28a) only involves collinear fields, this interaction is not relevant to model the Scharnhorst effect. Inversely, because our probe is a collinear photon, and Eq. (4.28c) represents interactions between soft fields only, it does not come into play either. As a result, only interactions of the form given in Eq. (4.28b) matter for the Scharnhorst effect. Higher order terms than those in Eq. (4.28) are power suppressed by additional powers of α or $\frac{1}{m_e}$.

---

[18] Several terms can have one and the same of these structures and differ from one another only in how the indices (not indicated here) contract.

4.3.3 SCET perturbative expansion and control of the errors at higher orders

Generally speaking, each terms appearing in the (power or perturbative) expansion of a theory consists of the expansion parameters (elevated to the power that characterizes the order of the term), and an expression multiplying them, made up of the numerical factors from the matching coefficients, and the operators. Concretely, if we consider for instance the first interaction term in the standard Euler-Heisenberg Lagrangian $c_1 \frac{\alpha^2}{m_e^4} \left(F^{\mu\nu}F_{\mu\nu}\right)^2$:

- expansion parameters: $\alpha^2 \frac{1}{m_e^4}$
- remaining factor: $c_1 \left(F^{\mu\nu}F_{\mu\nu}\right)^2$

While this is the most straightforward way to factor the term $c_1 \frac{\alpha^2}{m_e^4} \left(F^{\mu\nu}F_{\mu\nu}\right)^2$, because the factors chosen already appear in it as it stands, it has one major drawback: recall from the previous section that the relevant expansion parameter for the second expansion discussed is not $\frac{1}{m_e}$ *per se*, but $\frac{E}{m_e}$ where E is the typical energy of our probe photon – and indeed, it has to be if it is to be dimensionless and much smaller than 1 in the chosen regime of $\omega << m_e$. Also, recall that the term has dimension 4 in energy (once integrated by the measure of dimension -4, the result is dimensionless). The field $F^{\mu\nu}$ itself has energy dimensions E^2.

So a more meaningful way to factor $c_1 \frac{\alpha^2}{m_e^4} \left(F^{\mu\nu}F_{\mu\nu}\right)^2$ is to rewrite it in terms of a dimensionless field $\overline{F}^{\mu\nu} = \frac{F^{\mu\nu}}{E^2}$:

$$\begin{aligned}
& \int d^4x \quad c_1 \quad \alpha^2 \quad \frac{1}{m_e^4} \quad \left(F^{\mu\nu}F_{\mu\nu}\right)^2 \\
\rightarrow \quad & \int d^4(x\,E)\,\frac{1}{E^4} \quad c_1 \quad \alpha^2 \left(\frac{1}{m_e}\right)^4 \quad E^8 \left(\frac{F^{\mu\nu}}{E^2}\frac{F_{\mu\nu}}{E^2}\right)^2 \\
\rightarrow \quad & \int d^4\overline{x} \quad c_1 \quad \alpha^2 \left(\frac{E}{m_e}\right)^4 \quad \left(\overline{F}^{\mu\nu}\,\overline{F}_{\mu\nu}\right)^2 .
\end{aligned} \tag{4.29}$$

So the integrand is the product of the following expressions, which are dimensionless:

- expansion parameters: $\alpha^2 \left(\frac{E}{m_e}\right)^4$
- remaining factor: $c_1 \left(\overline{F}^{\mu\nu}\overline{F}_{\mu\nu}\right)^2$

Notice that now, not only the expansion parameters but also the remaining factor multiplying them is dimensionless.[19]

However in the case of the standard Euler-Heisenberg theory, we cannot meaningfully assess the size of the factor $c_1 \left(\overline{F}^{\mu\nu}\overline{F}_{\mu\nu}\right)^2$ because the size of $\langle\overline{F}^{\mu\nu}\rangle = \frac{\langle F^{\mu\nu}\rangle}{E^2}$ can vary greatly for different fields; and in the situation of interest to us, we know they do. We know the photons from the Casimir vacuum have energies below that corresponding to the plasma frequency, while the probe photon of interest needs to have as high an energy as permissible, for the sake of minimizing uncertainties in the position of the wavefront. Within the standard theory, both are described by the same $F^{\mu\nu}$, which can have any energy in the range $\frac{1}{L} << E << m_e$, so the size of $\langle\overline{F}^{\mu\nu}\rangle = \frac{\langle F^{\mu\nu}\rangle}{E^2}$ is not well-defined. Consequently, we cannot assess the size of $c_1 \left(\overline{F}^{\mu\nu}\overline{F}_{\mu\nu}\right)^2$.

What SCET allows us to do by differentiating fields according to their energy scales, is to ensure that the factor analogous to $c_1 \left(\overline{F}^{\mu\nu}\overline{F}_{\mu\nu}\right)^2$ is not only dimensionless, but of order 1.

As we did above for the standard Euler-Heisenberg theory, we rewrite our term of interest using dimensionless fields $\overline{\mathcal{F}}_c^{\mu\nu} = \frac{\mathcal{F}_c^{\mu\nu}}{E^2}$ and $\overline{\mathcal{F}}_s^{\mu\nu} = \frac{\mathcal{F}_s^{\mu\nu}}{E^2}$, where E refers to the

---

[19] That is, the matrix elements in these operators are dimensionless. We cannot describe $\overline{F}^{\mu\nu}$ itself as numerical since it is an operator, but $\langle\overline{F}^{\mu\nu}\rangle$ is numerical.

energy of the Scharnhorst photon:[20]

$$\begin{aligned} & \int d^4x_c \quad c_1 \quad \alpha^2 \quad \frac{1}{m_e^4} \quad (\mathcal{F}_c^{\mu\nu}\ \mathcal{F}^c_{\mu\nu}) \quad (\mathcal{F}_s^{\rho\sigma}\ \mathcal{F}^s_{\rho\sigma}) \\ \rightarrow \quad & \int d^4(x_c\, E)\ \frac{1}{E^4} \quad c_1 \quad \alpha^2 \quad \left(\frac{1}{m_e}\right)^4 \quad E^8 \left(\frac{\mathcal{F}_c^{\mu\nu}}{E^2}\frac{\mathcal{F}^c_{\mu\nu}}{E^2}\right)\left(\frac{\mathcal{F}_s^{\rho\sigma}}{E^2}\frac{\mathcal{F}^s_{\rho\sigma}}{E^2}\right) \\ \rightarrow \quad & \int d^4\overline{x}_c \quad c_1 \quad \alpha^2 \quad \left(\frac{E}{m_e}\right)^4 \quad \left(\overline{\mathcal{F}}_c^{\mu\nu}\ \overline{\mathcal{F}}^c_{\mu\nu}\right)\left(\overline{\mathcal{F}}_s^{\rho\sigma}\ \overline{\mathcal{F}}^s_{\rho\sigma}\right). \end{aligned} \tag{4.30}$$

So the integrand is the product of the following dimensionless expressions:
- expansion parameters: $\alpha^2\ \left(\frac{E}{m_e}\right)^4$
- remaining factor: $c_1\left(\overline{\mathcal{F}}_c^{\mu\nu}\ \overline{\mathcal{F}}^c_{\mu\nu}\right)\left(\overline{\mathcal{F}}_s^{\rho\sigma}\ \overline{\mathcal{F}}^s_{\rho\sigma}\right)$

So far all is quite similar to the standard theory treatment just discussed above. However now in addition, in SCET the fields and the measure have a specific scaling in $\lambda \equiv \sqrt{\frac{\epsilon}{E}}$, as we determined in the previous section.[21] The corresponding dimensionless quantities we have just defined, *i.e.* $\overline{\mathcal{F}}_c^{\mu\nu}$, $\overline{\mathcal{F}}_s^{\rho\sigma}$ and $d^4\overline{x}_c$, have the same scaling as the latter. We can make their λ dependance explicit by re-scaling these quantities: that is, we express them as the product of a quantity of order 1, and a power of λ which is a measure of their relative size:

$$\begin{aligned} \overline{\mathcal{F}}_c^{\mu\nu} &\sim O(\lambda^2) \quad \rightarrow \quad \overline{\mathcal{F}}_c^{\mu\nu} = \lambda^2\ \breve{\mathcal{F}}_c^{\mu\nu} \\ \overline{\mathcal{F}}_s^{\rho\sigma} &\sim O(\lambda^4) \quad \rightarrow \quad \overline{\mathcal{F}}_s^{\rho\sigma} = \lambda^4\ \breve{\mathcal{F}}_s^{\rho\sigma} \\ d^4\overline{x}_c &\sim O(\lambda^{-4}) \quad \rightarrow \quad d^4\overline{x}_c = \lambda^{-4}\ d^4\breve{x}_c. \end{aligned} \tag{4.31}$$

---

[20] Again, these are operators so we cannot describe them as numerical quantities, but their matrix elements as well as $\langle\overline{\mathcal{F}}_c^{\mu\nu}\rangle$ and $\langle\overline{\mathcal{F}}_s^{\rho\sigma}\rangle$ are.

[21] $\mathcal{F}_c^{\mu\nu} \sim O(\lambda^2)$, $\mathcal{F}_s^{\mu\nu} \sim O(\lambda^4)$, and $d^4x_c \sim O(\lambda^{-4})$ (respectively Eqs. (4.22), (4.26) and (4.19)).

where $\langle \breve{\mathfrak{F}}_c^{\mu\nu} \rangle$, $\langle \breve{\mathfrak{F}}_s^{\rho\sigma} \rangle$ and $\breve{x}_c$ are now of order 1, in addition to being dimensionless. In terms of the fields re-scaled in λ our term now takes the form:

$$\begin{aligned}
& \int d^4\overline{x}_c \quad c_1 \quad \alpha^2 \quad \left(\frac{E}{m_e}\right)^4 \quad \left(\overline{\mathfrak{F}}_c^{\mu\nu}\, \overline{\mathfrak{F}}^c_{\mu\nu}\right)\left(\overline{\mathfrak{F}}_s^{\rho\sigma}\, \overline{\mathfrak{F}}^s_{\rho\sigma}\right) \\
\rightarrow \quad & \int d^4\breve{x}_c\, \lambda^{-4} \quad c_1 \quad \alpha^2 \quad \left(\frac{E}{m_e}\right)^4 \quad \left(\lambda^4\, \breve{\mathfrak{F}}_c^{\mu\nu}\, \breve{\mathfrak{F}}^c{}_{\mu\nu}\right)\left(\lambda^8\, \breve{\mathfrak{F}}_s^{\rho\sigma}\, \breve{\mathfrak{F}}^s{}_{\rho\sigma}\right) \\
\rightarrow \quad & \int d^4\breve{x}_c \quad c_1 \quad \alpha^2 \quad \left(\frac{E}{m_e}\right)^4 \lambda^8 \quad \left(\breve{\mathfrak{F}}_c^{\mu\nu}\, \breve{\mathfrak{F}}^c{}_{\mu\nu}\right)\left(\breve{\mathfrak{F}}_s^{\rho\sigma}\, \breve{\mathfrak{F}}^s{}_{\rho\sigma}\right). \qquad (4.32)
\end{aligned}$$

So in SCET the integrand of this term is the product of the following dimensionless expressions:

- expansion parameters: $\alpha^2 \left(\frac{E}{m_e}\right)^4 \lambda^8$
- remaining factor: $c_1 \left(\breve{\mathfrak{F}}_c^{\mu\nu}\, \breve{\mathfrak{F}}^c_{\mu\nu}\right)\left(\breve{\mathfrak{F}}_s^{\rho\sigma}\, \breve{\mathfrak{F}}^s_{\rho\sigma}\right)$

Comparing this to the standard Euler-Heisenberg theory, we note two differences. First we now have one extra dimensionless parameter, λ which is less than 1 by definition since $\lambda \equiv \sqrt{\frac{\epsilon}{E}}$, and which appears to a positive power. Like the other two expansion parameters, it tends to power suppress higher order terms. The second difference is that the expansion parameter are now multiplied by a factor that we can be confident is of order 1, since it is a product of $\breve{\mathfrak{F}}_c^{\mu\nu}$ and $\breve{\mathfrak{F}}_s^{\rho\sigma}$ which are of order 1.[22] This is of course because we have designed them that way but the point is that in SCET we can do so, and what we get when we do is a factor of λ

---

[22] Again, more correctly $\langle \breve{\mathfrak{F}}_c^{\mu\nu} \rangle$, $\langle \breve{\mathfrak{F}}_s^{\rho\sigma} \rangle$ are.

to a positive power. That is, to summarize what was done above:

$$\begin{aligned}
\mathscr{F}_c^{\mu\nu} &= E^2 \;\; \overline{\mathscr{F}}_c^{\mu\nu} &&= E^2 \;\; \lambda^2 \;\; \breve{\mathscr{F}}_c^{\mu\nu} \\
\mathscr{F}_s^{\rho\sigma} &= E^2 \;\; \overline{\mathscr{F}}_s^{\rho\sigma} &&= E^2 \;\; \lambda^4 \;\; \breve{\mathscr{F}}_s^{\rho\sigma} \\
d^4x_c &= E^{-4}\, d^4\overline{x}_c &&= E^{-4} \;\; \lambda^{-4} \;\; d^4\breve{x}_c.
\end{aligned} \tag{4.33}$$

so that instead of simply the dimensionless $\overline{F}^{\mu\nu} = \frac{F^{\mu\nu}}{E^2}$ from the standard Euler-Heisenberg theory, whose size $\langle \overline{F}^{\mu\nu} \rangle = \frac{\langle F^{\mu\nu} \rangle}{E^2}$ is not well-defined within the theory, SCET allows us to define $\breve{\mathscr{F}}_c^{\mu\nu} = \frac{\mathscr{F}_c^{\mu\nu}}{E^2\,\lambda^2}$ and $\breve{\mathscr{F}}_s^{\mu\nu} = \frac{\mathscr{F}_s^{\mu\nu}}{E^2\,\lambda^4}$, with $\langle \breve{\mathscr{F}}_c^{\mu\nu} \rangle = \frac{\langle \mathscr{F}_c^{\mu\nu} \rangle}{E^2\,\lambda^2}$ and $\langle \breve{\mathscr{F}}_s^{\mu\nu} \rangle = \frac{\langle \mathscr{F}_s^{\mu\nu} \rangle}{E^2\,\lambda^4}$ of order of 1. In doing so one gets an extra factor of λ to a positive power, which tends to power suppress the term (over and above the power-suppression due to α and $\frac{E}{m_e}$).

A similar re-scaling of $\breve{\mathscr{F}}_c^{\mu\nu}$ and $\breve{\mathscr{F}}_s^{\rho\sigma}$ can be done for higher order terms in the α and $\frac{E}{m_e}$ expansions, and it should be clear that the structure of the resulting terms always involves an expression of order 1 containing the re-scaled fields, and a factor of λ to a positive power.

This implies that the theory errors are well-controlled as we go to higher orders in the expansion parameters. By design, the latter are all numbers much smaller than 1, so that terms of higher orders in any of them are power-suppressed in that parameter. With the standard Euler-Heisenberg theory however, the coefficients are of unknown size, so one cannot be confident that the power-suppression due to the parameters actually results in smaller and smaller corrections as we go to higher and higher order in the series expansions. In the SCET framework however, because these coefficients are of order one, we know that they cannot spoil the power suppression due to the parameters.

So SCET allows us to make a definite statement about the size of the theory errors: they would be higher order in all expansion parameters, *i.e.* in α, $\frac{E}{m_e}$ and λ.

4.3.4 Interaction between Casimir background and probe photon

A second advantage of modelling the Scharnhorst effect within the framework of SCET is that it allows us to describe the interaction between the probe photon and the Casimir vacuum in a way that suggests the phase velocity derived within this theory may in fact constitute the signal velocity as well. I shall discuss this crucial point in the Conclusion (chapter 7). For now let us examine the relevant aspects of the SCET model.

First, as discussed in the previous sections, SCET describes the interaction between the Casimir background and the Scharnhorst photon as occurring via the p^- component of the latter. Indeed, as Eq. (4.16) indicates, all momentum components of the Casimir background obey $k^\mu \sim O(\lambda^2)$. Only one of the momentum components of the Scharnhorst photon has the same scaling in λ (Eq. (4.15)): $p^- \sim O(\lambda^2)$. This expresses the fact that p^- is of the same energy scale as k^μ, in contrast to p^+ and $p_\perp$. Now it is a standard aspect of effective field theory that entities of very different energy scales hardly interact with one another. Therefore the scaling of the momentum components implies that the Casimir background, made of soft photons, only interacts with the p^- component of the propagating, collinear photon, not with $p_\perp$, and, especially, not with p^+. The latter two components of the Scharnhorst photon are decoupled from the background.

Second, to the order at which the Scharnhorst effect is calculated, SCET shows that it depends only on the background, not the energy E of the probe photon. As we saw in the previous section, the relevant interaction term is of the form (Eq. (4.32)):

$$\int d^4\breve{x}_c \quad c_1 \quad \alpha^2 \quad \left(\frac{E}{m_e}\right)^4 \ \lambda^8 \ \left(\breve{\mathfrak{F}}_c^{\mu\nu}\,\breve{\mathfrak{F}}{c}_{\mu\nu}\right)\left(\breve{\mathfrak{F}}_s^{\rho\sigma}\,\breve{\mathfrak{F}}{s}_{\rho\sigma}\right). \tag{4.34}$$

Recall that $\left(\breve{\mathfrak{F}}_c^{\mu\nu}\,\breve{\mathfrak{F}}{c}_{\mu\nu}\right)\left(\breve{\mathfrak{F}}_s^{\rho\sigma}\,\breve{\mathfrak{F}}{s}_{\rho\sigma}\right)$ is a numerical expression of order 1. Therefore the dependence of the interaction on the energies is determined by the expansion parameters. These can be written as follows:

$$\begin{aligned} &\alpha^2 \ \left(\frac{E}{m_e}\right)^4 \ \lambda^8 \\ = \ &\alpha^2 \ \left(\frac{E}{m_e}\right)^4 \ \left(\sqrt{\frac{\epsilon}{E}}\right)^8 \\ = \ &\alpha^2 \ \left(\frac{E}{m_e}\right)^4 \ \left(\frac{\epsilon}{E}\right)^4 \\ = \ &\alpha^2 \ \left(\frac{\epsilon}{m_e}\right)^4 . \end{aligned} \tag{4.35}$$

The only energy this expression depends on is that of the background, ϵ (recall that α is dimensionless). Notably, it does not depend on E, the energy of the propagating photon.

4.4 Derivation of the SCET Lagrangian

As discussed in chapter 3, the Scharnhorst effect has been derived using the standard Euler-Heisenberg Lagrangian $\mathfrak{L}_{EH}$, which defines an effective theory containing only one degree of freedom: the photon field $F^{\mu\nu}$. Computing the Scharnhorst effect within the framework of SCET means using a different form of this Lagrangian, which separates energy and momentum scales by "splitting" the photon field into two degrees of freedom: a collinear photon field $\mathfrak{F}_c^{\mu\nu}$ which represents the Scharnhorst photon and a soft photon field $\mathcal{F}_s^{\mu\nu}$ which accounts for the background. Accordingly, the SCET Euler-Heisenberg Lagrangian $\mathcal{L}_{EHSC}$ is derived by splitting $F^{\mu\nu}$ into these two fields:

$$\mathcal{L}_{EH} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{\alpha^2}{m_e^4} \left\{ c_1 \left(F^{\mu\nu} F_{\mu\nu} \right)^2 + c_2 \left(F^{\mu\nu} \widetilde{F}_{\mu\nu} \right)^2 \right\}, \tag{4.36}$$

$$\begin{aligned}
\mathcal{L}_{EH}^{SC} &= -\frac{1}{4} \left(\mathcal{F}_c^{\mu\nu} + \mathcal{F}_s^{\mu\nu} \right) \left(\mathcal{F}^c_{\mu\nu} + \mathcal{F}^s_{\mu\nu} \right) \\
&+ \frac{\alpha^2}{m_e^4} \left\{ c_1 \left(\left(\mathcal{F}_c^{\mu\nu} + \mathcal{F}_s^{\mu\nu} \right) \left(\mathcal{F}^c_{\mu\nu} + \mathcal{F}^s_{\mu\nu} \right) \right)^2 + c_2 \left(\left(\mathcal{F}_c^{\mu\nu} + \mathcal{F}_s^{\mu\nu} \right) \left(\widetilde{\mathcal{F}}^c_{\mu\nu} + \widetilde{\mathcal{F}}^s_{\mu\nu} \right) \right)^2 \right\} \\
&= -\frac{1}{4} \left(\mathcal{F}_c^{\mu\nu} + \mathcal{F}_s^{\mu\nu} \right) \left(\mathcal{F}^c_{\mu\nu} + \mathcal{F}^s_{\mu\nu} \right) \\
&+ \frac{\alpha^2}{m_e^4} \left\{ c_1 \left(\mathcal{F}_c^{\kappa\lambda} + \mathcal{F}_s^{\kappa\lambda} \right) \left(\mathcal{F}^c_{\kappa\lambda} + \mathcal{F}^s_{\kappa\lambda} \right) \left(\mathcal{F}_c^{\rho\sigma} + \mathcal{F}_s^{\rho\sigma} \right) \left(\mathcal{F}^c_{\rho\sigma} + \mathcal{F}^s_{\rho\sigma} \right) \right. \\
&+ \left. c_2 \left(\mathcal{F}_c^{\kappa\lambda} + \mathcal{F}_s^{\kappa\lambda} \right) \left(\widetilde{\mathcal{F}}^c_{\kappa\lambda} + \widetilde{\mathcal{F}}^s_{\kappa\lambda} \right) \left(\mathcal{F}_c^{\rho\sigma} + \mathcal{F}_s^{\rho\sigma} \right) \left(\widetilde{\mathcal{F}}^c_{\rho\sigma} + \widetilde{\mathcal{F}}^s_{\rho\sigma} \right) \right\},
\end{aligned} \tag{4.37}$$

where the index “c” refers to the collinear sector, and “s” to the soft one. This yields:

$$\begin{aligned}
\mathcal{L}_{EH}^{SC} &= -\frac{1}{4} \left(\mathcal{F}_c^{\mu\nu}\, \mathcal{F}^c_{\mu\nu} + 2\, \mathcal{F}_c^{\mu\nu}\, \mathcal{F}^s_{\mu\nu} + \mathcal{F}_s^{\mu\nu}\, \mathcal{F}^s_{\mu\nu} \right) \\
&+ \frac{\alpha^2}{m_e^4} \left\{ c_1 \left(\mathcal{F}_c^{\kappa\lambda}\, \mathcal{F}^c_{\kappa\lambda}\, \mathcal{F}_c^{\rho\sigma}\, \mathcal{F}^c_{\rho\sigma} + 4\, \mathcal{F}_c^{\kappa\lambda}\, \mathcal{F}^c_{\kappa\lambda}\, \mathcal{F}_c^{\rho\sigma}\, \mathcal{F}^s_{\rho\sigma} + 2\, \mathcal{F}_c^{\kappa\lambda}\, \mathcal{F}^c_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \mathcal{F}^s_{\rho\sigma} \right. \right. \\
&+ \left. 4\, \mathcal{F}_c^{\kappa\lambda}\, \mathcal{F}^s_{\kappa\lambda}\, \mathcal{F}_c^{\rho\sigma}\, \mathcal{F}^s_{\rho\sigma} + 4\, \mathcal{F}_c^{\kappa\lambda}\, \mathcal{F}^s_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \mathcal{F}^s_{\rho\sigma} + \mathcal{F}_s^{\kappa\lambda}\, \mathcal{F}^s_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \mathcal{F}^s_{\rho\sigma} \right) \\
&+ c_2 \left(\mathcal{F}_c^{\kappa\lambda}\, \widetilde{\mathcal{F}}^c_{\kappa\lambda}\, \mathcal{F}_c^{\rho\sigma}\, \widetilde{\mathcal{F}}^c_{\rho\sigma} + 2\, \mathcal{F}_c^{\kappa\lambda}\, \widetilde{\mathcal{F}}^c_{\kappa\lambda}\, \mathcal{F}_c^{\rho\sigma}\, \widetilde{\mathcal{F}}^s_{\rho\sigma} + 2\, \mathcal{F}_c^{\kappa\lambda}\, \widetilde{\mathcal{F}}^c_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \widetilde{\mathcal{F}}^c_{\rho\sigma} \right. \\
&+ 2\, \mathcal{F}_c^{\kappa\lambda}\, \widetilde{\mathcal{F}}^c_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \widetilde{\mathcal{F}}^s_{\rho\sigma} + \mathcal{F}_c^{\kappa\lambda}\, \widetilde{\mathcal{F}}^s_{\kappa\lambda}\, \mathcal{F}_c^{\rho\sigma}\, \widetilde{\mathcal{F}}^s_{\rho\sigma} + 2\, \mathcal{F}_c^{\kappa\lambda}\, \widetilde{\mathcal{F}}^s_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \widetilde{\mathcal{F}}^c_{\rho\sigma} + \mathcal{F}_s^{\kappa\lambda}\, \widetilde{\mathcal{F}}^c_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \widetilde{\mathcal{F}}^c_{\rho\sigma} \\
&+ \left. \left. 2\, \mathcal{F}_c^{\kappa\lambda}\, \widetilde{\mathcal{F}}^s_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \widetilde{\mathcal{F}}^s_{\rho\sigma} + 2\, \mathcal{F}_s^{\kappa\lambda}\, \widetilde{\mathcal{F}}^c_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \widetilde{\mathcal{F}}^s_{\rho\sigma} + \mathcal{F}_s^{\kappa\lambda}\, \widetilde{\mathcal{F}}^s_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \widetilde{\mathcal{F}}^s_{\rho\sigma} \right) \right\}.
\end{aligned} \tag{4.38}$$

In the leading order term, the product of a soft and a collinear field is unphysical in so far that these products of two fields represent propagators, and a given propagator characterizes one type of photon: a photon cannot turn from soft to collinear (or vice versa) between the two events that its propagator relates. Similarly, at next to leading order, products of a collinear field with three soft ones, or vice versa, are also unphysical, as these products of four fields represent tadpole diagrams, with two of the fields referring to the external legs, and the other two the loop. As with the free propagator, a photon cannot switch from being soft to collinear between the two events connected by the legs, nor within a loop. Our Lagrangian then simplifies to the form:

$$\begin{aligned}
\mathcal{L}^{SC}_{EH} &= -\frac{1}{4}\left(\mathcal{F}^{\mu\nu}_c\,\mathcal{F}^c_{\mu\nu} + \mathcal{F}^{\mu\nu}_s\,\mathcal{F}^s_{\mu\nu}\right) \\
&+ \frac{\alpha^2}{m_e^4}\Big\{c_1\Big(\mathcal{F}^{\kappa\lambda}_c\,\mathcal{F}^c_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_c\,\mathcal{F}^c_{\rho\sigma} + 2\,\mathcal{F}^{\kappa\lambda}_c\,\mathcal{F}^c_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\mathcal{F}^s_{\rho\sigma} + 4\,\mathcal{F}^{\kappa\lambda}_c\,\mathcal{F}^s_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_c\,\mathcal{F}^s_{\rho\sigma} \\
&+ \mathcal{F}^{\kappa\lambda}_s\,\mathcal{F}^s_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\mathcal{F}^s_{\rho\sigma}\Big) \\
&+ c_2\Big(\mathcal{F}^{\kappa\lambda}_c\,\widetilde{\mathcal{F}}^c_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_c\,\widetilde{\mathcal{F}}^c_{\rho\sigma} + 2\,\mathcal{F}^{\kappa\lambda}_c\,\widetilde{\mathcal{F}}^c_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\widetilde{\mathcal{F}}^s_{\rho\sigma} + \mathcal{F}^{\kappa\lambda}_c\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_c\,\widetilde{\mathcal{F}}^s_{\rho\sigma} \\
&+ 2\,\mathcal{F}^{\kappa\lambda}_c\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\widetilde{\mathcal{F}}^c_{\rho\sigma} + \mathcal{F}^{\kappa\lambda}_s\,\widetilde{\mathcal{F}}^c_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\widetilde{\mathcal{F}}^c_{\rho\sigma} + \mathcal{F}^{\kappa\lambda}_s\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\widetilde{\mathcal{F}}^s_{\rho\sigma}\Big)\Big\}. \qquad (4.39)
\end{aligned}$$

Now it is easily seen that:

$$\mathcal{F}^{\kappa\lambda}_c\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_c\,\widetilde{\mathcal{F}}^s_{\rho\sigma} = \mathcal{F}^{\kappa\lambda}_c\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\widetilde{\mathcal{F}}^c_{\rho\sigma} = \mathcal{F}^{\kappa\lambda}_s\,\widetilde{\mathcal{F}}^c_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\widetilde{\mathcal{F}}^c_{\rho\sigma}. \qquad (4.40)$$

Indeed:

$$\mathcal{F}^{\kappa\lambda}_c\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_c\,\widetilde{\mathcal{F}}^s_{\rho\sigma} = \mathcal{F}^{\kappa\lambda}_c\,\epsilon_{\kappa\lambda\sigma\rho}\,\mathcal{F}^{\sigma\rho}_s\,\mathcal{F}^{\rho\sigma}_c\,\epsilon_{\rho\sigma\zeta\xi}\,\mathcal{F}^{\zeta\xi}_s, \qquad (4.41)$$

and renaming dummy indices:

$$\begin{aligned}
\mathcal{F}^{\kappa\lambda}_c\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\widetilde{\mathcal{F}}^c_{\rho\sigma} &= \mathcal{F}^{\kappa\lambda}_c\,\epsilon_{\kappa\lambda\sigma\rho}\,\mathcal{F}^{\sigma\rho}_s\,\mathcal{F}^{\rho\sigma}_s\,\epsilon_{\rho\sigma\zeta\xi}\,\mathcal{F}^{\zeta\xi}_c \\
&= \mathcal{F}^{\kappa\lambda}_c\,\epsilon_{\kappa\lambda\sigma\rho}\,\mathcal{F}^{\sigma\rho}_s\,\mathcal{F}^{\zeta\xi}_s\,\epsilon_{\zeta\xi\rho\sigma}\,\mathcal{F}^{\rho\sigma}_c \\
&= \mathcal{F}^{\kappa\lambda}_c\,\epsilon_{\kappa\lambda\sigma\rho}\,\mathcal{F}^{\sigma\rho}_s\,\mathcal{F}^{\rho\sigma}_c\,\epsilon_{\zeta\xi\rho\sigma}\,\mathcal{F}^{\zeta\xi}_s \\
&= \mathcal{F}^{\kappa\lambda}_c\,\epsilon_{\kappa\lambda\sigma\rho}\,\mathcal{F}^{\sigma\rho}_s\,\mathcal{F}^{\rho\sigma}_c\,\epsilon_{\rho\sigma\zeta\xi}\,\mathcal{F}^{\zeta\xi}_s, \qquad (4.42)
\end{aligned}$$

$$\begin{aligned}
\mathcal{F}_s^{\kappa\lambda}\,\widetilde{\mathcal{F}}^c_{\kappa\lambda}\,\mathcal{F}_s^{\rho\sigma}\,\widetilde{\mathcal{F}}^c_{\rho\sigma} &= \mathcal{F}_s^{\kappa\lambda}\,\epsilon_{\kappa\lambda\sigma\rho}\,\mathcal{F}_c^{\sigma\rho}\,\mathcal{F}_s^{\rho\sigma}\,\epsilon_{\rho\sigma\zeta\xi}\,\mathcal{F}_c^{\zeta\xi} \\
&= \mathcal{F}_s^{\sigma\rho}\,\epsilon_{\sigma\rho\kappa\lambda}\,\mathcal{F}_c^{\kappa\lambda}\,\mathcal{F}_s^{\zeta\xi}\,\epsilon_{\zeta\xi\rho\sigma}\,\mathcal{F}_c^{\rho\sigma} \\
&= \mathcal{F}_c^{\kappa\lambda}\,\epsilon_{\sigma\rho\kappa\lambda}\,\mathcal{F}_s^{\sigma\rho}\,\mathcal{F}_c^{\rho\sigma}\,\epsilon_{\zeta\xi\rho\sigma}\mathcal{F}_s^{\zeta\xi} \\
&= \mathcal{F}_c^{\kappa\lambda}\,\epsilon_{\kappa\lambda\sigma\rho}\,\mathcal{F}_s^{\sigma\rho}\,\mathcal{F}_c^{\rho\sigma}\,\epsilon_{\rho\sigma\zeta\xi}\,\mathcal{F}_s^{\zeta\xi}. \qquad (4.43)
\end{aligned}$$

Q.E.D.

Therefore we can write:

$$\begin{aligned}
\mathcal{L}_{EH}^{SC} &= -\frac{1}{4}\left(\mathcal{F}_c^{\mu\nu}\,\mathcal{F}^c_{\mu\nu} + \mathcal{F}_s^{\mu\nu}\,\mathcal{F}^s_{\mu\nu}\right) \\
&+ \frac{\alpha^2}{m_e^4}\Big\{c_1\Big(\mathcal{F}_c^{\kappa\lambda}\,\mathcal{F}^c_{\kappa\lambda}\,\mathcal{F}_c^{\rho\sigma}\,\mathcal{F}^c_{\rho\sigma} + 2\,\mathcal{F}_c^{\kappa\lambda}\,\mathcal{F}^c_{\kappa\lambda}\,\mathcal{F}_s^{\rho\sigma}\,\mathcal{F}^s_{\rho\sigma} + 4\,\mathcal{F}_c^{\kappa\lambda}\,\mathcal{F}^s_{\kappa\lambda}\,\mathcal{F}_c^{\rho\sigma}\,\mathcal{F}^s_{\rho\sigma} \\
&+ \mathcal{F}_s^{\kappa\lambda}\,\mathcal{F}^s_{\kappa\lambda}\,\mathcal{F}_s^{\rho\sigma}\,\mathcal{F}^s_{\rho\sigma}\Big) \\
&+ c_2\,\Big(\mathcal{F}_c^{\kappa\lambda}\,\widetilde{\mathcal{F}}^c_{\kappa\lambda}\,\mathcal{F}_c^{\rho\sigma}\,\widetilde{\mathcal{F}}^c_{\rho\sigma} + 2\,\mathcal{F}_c^{\kappa\lambda}\,\widetilde{\mathcal{F}}^c_{\kappa\lambda}\,\mathcal{F}_s^{\rho\sigma}\,\widetilde{\mathcal{F}}^s_{\rho\sigma} + 4\,\mathcal{F}_c^{\kappa\lambda}\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}_c^{\rho\sigma}\,\widetilde{\mathcal{F}}^s_{\rho\sigma} \\
&+ \mathcal{F}_s^{\kappa\lambda}\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}_s^{\rho\sigma}\,\widetilde{\mathcal{F}}^s_{\rho\sigma}\Big)\Big\}. \qquad (4.44)
\end{aligned}$$

The first order contribution (*i.e.* the first line in Eq.(4.44)) are the kinetic terms. They are the SCET analog to the classical, Maxwell Lagrangian. Were there no other terms, the speed of photon propagation would be c: the terms that follow constitute a perturbation on the speed associated with the kinetic terms. Because in the Scharnhorst effect only the collinear field propagates, we are only concerned with the first one, $\mathcal{F}_c^{\mu\nu}\,\mathcal{F}^c_{\mu\nu}$.

The last four lines give the interaction terms of lowest order in the α and $\frac{E}{m_e}$ expansions, which are of order $\frac{\alpha^2}{m_e^4}$. Among those, products containing only collinear fields are not involved in the Scharnhorst effect: they account for high energy background photons which, as explained above, are not affected by the presence of the plates and for this reason do not contribute to the effect. Consequently, we neglect them in our derivation.

Among the remaining expressions, we find terms involving only collinear fields, *i.e.* $(\mathfrak{F}_c\ \mathfrak{F}^c)(\mathfrak{F}_c\ \mathfrak{F}^c)$ and $(\mathfrak{F}_c\ \widetilde{\mathfrak{F}}^c)(\mathfrak{F}_c\ \widetilde{\mathfrak{F}}^c)$, terms involving only soft fields, *i.e.* $(\mathfrak{F}_s\ \mathfrak{F}^s)(\mathfrak{F}_s\ \mathfrak{F}^s)$ and $(\mathfrak{F}_s\ \widetilde{\mathfrak{F}}^s)(\mathfrak{F}_s\ \widetilde{\mathfrak{F}}^s)$, and terms of the form $(\mathfrak{F}_c\ \mathfrak{F}^c)(\mathfrak{F}_s\ \mathfrak{F}^s)$ and $(\mathfrak{F}_c\ \widetilde{\mathfrak{F}}^c)(\mathfrak{F}_s\ \widetilde{\mathfrak{F}}^s)$. The first two types of terms represent interactions between collinear fields alone, and soft fields alone, respectively. The third kind represent interactions between a collinear photon and a soft photon. Because the Scharnhorst effect involves a collinear propagating photon, and the Casimir vacuum of soft photons, modelling our effect involves only this latter type of terms – *i.e.* $(\mathfrak{F}_c\ \mathfrak{F}^c)(\mathfrak{F}_s\ \mathfrak{F}^s)$ and $(\mathfrak{F}_c\ \widetilde{\mathfrak{F}}^c)(\mathfrak{F}_s\ \widetilde{\mathfrak{F}}^s)$. The other are simply not relevant to the situation of interest to us. Therefore the part of the SCET Lagrangian that is pertinent to the Scharnhorst effect is:

$$\begin{aligned}\frac{\alpha^2}{m_e^4} \quad & \Big\{c_1 \left(2\ \mathfrak{F}_c^{\kappa\lambda}\, \mathfrak{F}^c_{\kappa\lambda}\, \mathfrak{F}_s^{\rho\sigma}\, \mathfrak{F}^s_{\rho\sigma} + 4\ \mathfrak{F}_c^{\kappa\lambda}\, \mathfrak{F}^s_{\kappa\lambda}\, \mathfrak{F}_c^{\rho\sigma}\, \mathfrak{F}^s_{\rho\sigma}\right) \\ + \quad & c_2 \left(2\ \mathfrak{F}_c^{\kappa\lambda}\, \widetilde{\mathfrak{F}}^c_{\kappa\lambda}\, \mathfrak{F}_s^{\rho\sigma}\, \widetilde{\mathfrak{F}}^s_{\rho\sigma} + 4\, \mathfrak{F}_c^{\kappa\lambda}\, \widetilde{\mathfrak{F}}^s_{\kappa\lambda}\, \mathfrak{F}_c^{\rho\sigma}\, \widetilde{\mathfrak{F}}^s_{\rho\sigma}\right)\Big\}\,. \end{aligned} \tag{4.45}$$

We now determine the contribution of each of these terms to the photon self-energy, hence to the second order polarization tensor $\Pi_{\mu\nu}^{2-loop}$.

4.5 Derivation of the SCET polarization tensor $\Pi_{\mu\nu}^{2-loop}$

The quantity we shall need in order to derive the phase velocity of the propagating photon is $\Pi_{\mu\nu}^{2-loop}$. In order to obtain it – and the different contributions to it –, we first consider the fully interacting photon propagator $D^{\mu\nu}(x-y)$ in a Casimir set-up. From the latter we shall identify the relevant Feynman diagram: a tadpole diagram whose external legs involve the collinear field while the loop is formed from the soft field propagator.

4.5.1 Relevant terms to model the Scharnhorst effect in $\mathcal{L}_{EH}^{SC}$ theory

As always $D^{\mu\nu}(x-y)$ is defined by:

$$iD^{\mu\nu}(x-y) \equiv \langle\Omega| T\left[A^{\mu}(x)A^{\nu}(y)\right]|\Omega\rangle\,, \tag{4.46}$$

that is:

$$iD^{\mu\nu}(x-y) = \\ \lim_{t\to\infty(1-i\epsilon)} \frac{\left\langle 0\left| T\left\{\int DA\;\; A^{\mu}(x)A^{\nu}(y)\;\; Exp\left[i\int_{-t}^{+t}dt'\;\int_{-\infty}^{+\infty}d^3z\;\left(\mathcal{L}-\frac{1}{2\xi}\left(\partial^{\mu}A_{\mu}\right)^2\right)\right\}\right]\right|0\right\rangle}{\left\langle 0\left| T\left\{\int DA\;\; Exp\left[i\int_{-t}^{+t}dt'\;\int_{-\infty}^{+\infty}d^3z\;\left(\mathcal{L}-\frac{1}{2\xi}\left(\partial^{\mu}A_{\mu}\right)^2\right)\right]\right\}\right|0\right\rangle} \tag{4.47}$$

where t' stands for z^0. In terms of Feynman diagrams, this can be represented as:

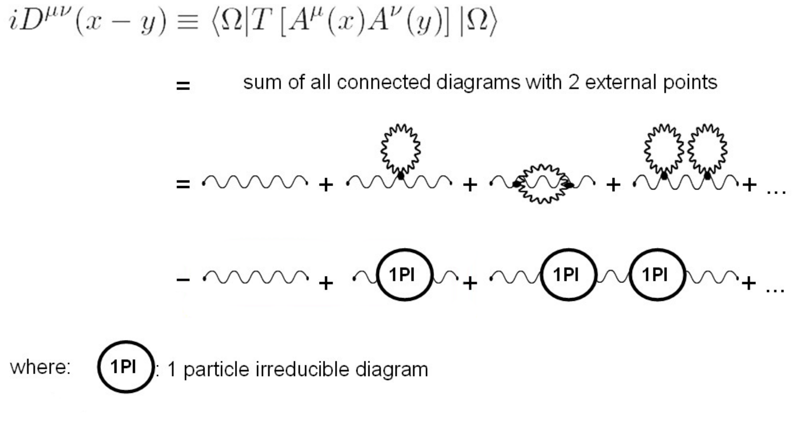


Figure 4.8: The fully interacting photon propagator in the standard $\mathcal{L}_{EH}$ theory.

In the standard theory, the propagators in the figure would represent photon propagators of arbitrary energy below the electron mass; in the SCET theory, each diagram would come in several versions that would differentiate collinear and soft fields for each propagator (say, one with collinear external legs with a soft loop, another the reverse, etc...).

Because the Euler-Heisenberg theory (be it the standard one or SCET) only involves terms to first order in α^2, if we now expand the numerator of Eq. (4.47) up to first order:[23]

$$\begin{aligned}
&\left\langle 0 \middle| T\left\{\int DA\ \ A^{\mu}(x)A^{\nu}(y)\ \ Exp\left[i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3z\ \mathfrak{L}\right]\right\} \middle| 0 \right\rangle \\
= &\int DA \left\langle 0 \middle| T\left\{A^{\mu}(x)A^{\nu}(y)\ \ Exp\left[i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3z\ \mathfrak{L}\right]\right\} \middle| 0 \right\rangle \\
\approx &\int DA \left\langle 0 \middle| T\left\{A^{\mu}(x)A^{\nu}(y)\ \ + A^{\mu}(x)A^{\nu}(y)\left[i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3z\ \mathfrak{L}\right]\right\} \middle| 0 \right\rangle .
\end{aligned} \tag{4.49}$$

So to first order Eq. (4.47) takes the form:

$$\begin{aligned}
&iD^{\mu\nu}(x-y)^{2nd-order} \\
\approx &\lim_{t\to\infty(1-i\epsilon)} \frac{\int DA \left\langle 0 \middle| T\left\{A^{\mu}(x)A^{\nu}(y)\ \ + A^{\mu}(x)A^{\nu}(y)\left[i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3z\ \mathfrak{L}\right]\right\} \middle| 0 \right\rangle}{\int DA \left\langle 0 \middle| T\left\{\ \ Exp\left[i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3z\ \mathfrak{L}\right]\right\} \middle| 0 \right\rangle} \\
\approx &\lim_{t\to\infty(1-i\epsilon)} \int DA \left\langle 0 \middle| T\left\{A^{\mu}(x)A^{\nu}(y)\ \ + A^{\mu}(x)A^{\nu}(y)\left[i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3z\ \mathfrak{L}\right]\right\} \middle| 0 \right\rangle .
\end{aligned} \tag{4.50}$$

In terms of Feynman diagrams $iD^{\mu\nu}(x-y)$ is:

---

[23] See Eq.(4.1) above for the standard Euler-Heisenberg Lagrangian:

$$\mathfrak{L}_{EH} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{\alpha^2}{m_e^4}\left\{c_1\left(F^{\mu\nu}F_{\mu\nu}\right)^2 + c_2\left(F^{\mu\nu}\widetilde{F}_{\mu\nu}\right)^2\right\}. \tag{4.48}$$

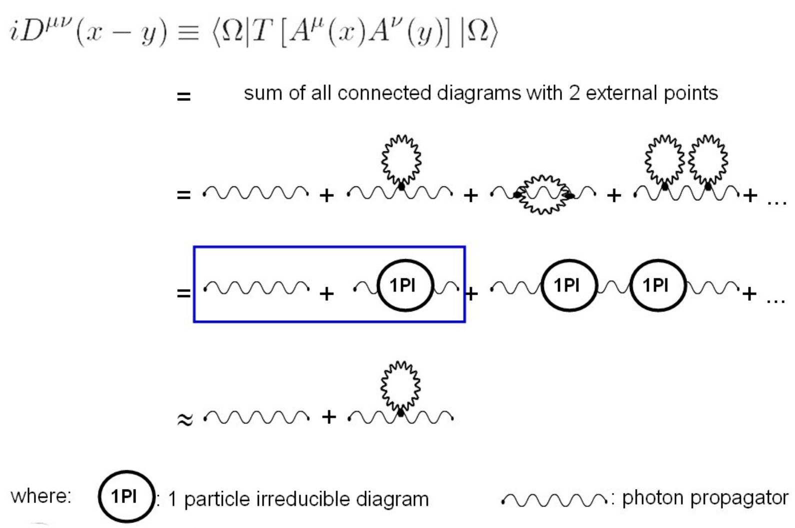


Figure 4.9: The fully interacting photon propagator in the standard $\mathcal{L}_{EH}$ theory.

In Fig. 4.9 we have used the standard Euler-Heisenberg theory.

In the SCET version of the theory, for a collinear propagating photon one gets the following instead:

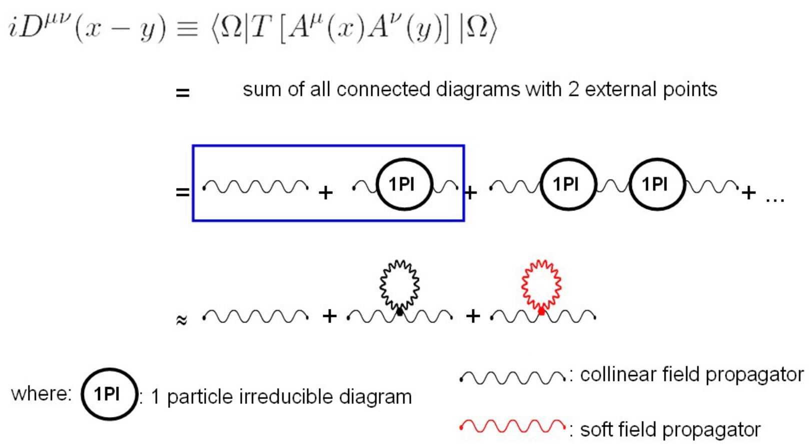


Figure 4.10: The fully interacting photon propagator in the SCET $\mathcal{L}_{EH}^{SC}$ theory.

The first term in Eq. (4.50) gives the free (*i.e.* non-interacting) result, the Feynman propagator:

$$\langle 0 \,|\, T\left\{A^{\mu}(x)A^{\nu}(y)\right\}|\, 0\rangle = D_F^{\mu\nu}(x-y). \tag{4.51}$$

Because the Scharnhorst effect is due to interactions, what is of interest to us is the second term, *i.e.*:

$$\begin{aligned} & iD^{\mu\nu}(x-y) \\ & \approx \lim_{t\to\infty(1-i\epsilon)} \frac{\int DA \left\langle 0 \middle| \, T\left\{A^{\mu}(x)A^{\nu}(y)\; i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3z\, \mathcal{L}\right\} \middle| \, 0\right\rangle}{\int DA \left\langle 0 \middle| \, T\left\{\; Exp\left[i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3z\, \mathcal{L}\right]\right\} \middle| \, 0\right\rangle} \\ & \approx \lim_{t\to\infty(1-i\epsilon)} \int DA \left\langle 0 \middle| \, T\left\{A^{\mu}(x)A^{\nu}(y)\; i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3z\, \mathcal{L}\right\} \middle| \, 0\right\rangle . \end{aligned} \tag{4.52}$$

Note that so far we have not drawn the distinction between the standard Euler-Heisenberg theory and our SCET version thereof in the algebra: up to this point the equations discussed in this section are valid for both. The difference rests in whether $\mathcal{L}$ refers to the Lagrangian of the standard Euler-Heisenberg theory $\mathcal{L}_{EH}$ or to its SCET counterpart $\mathcal{L}_{EH}^{SC}$.

For the specific purpose of modelling the Scharnhorst effect, where the background Casimir vacuum is represented by soft photons only, the second diagram in Fig.(4.10) does not come in, and we only consider:

$$iD^{\mu\nu}(x-y) \equiv \langle\Omega| T\left[A^{\mu}(x)A^{\nu}(y)\right]|\Omega\rangle \approx$$

Figure 4.11: The fully interacting propagator for the Scharnhorst photon in the $\mathcal{L}_{EH}^{SC}$ theory.

In the SCET Euler-Heisenberg theory and for the specific case of a Scharnhorst photon, to the relevant order of $\frac{\alpha^2}{m_e^4}$, $\mathcal{L}$ contains the four terms discussed in the previous section. That is for our situation of interest Eq. (4.52) contains $\mathcal{L}$ given by:

$$\mathcal{L}^{2nd\ term} = \frac{\alpha^2}{m_e^4}$$
$$\left\{2\, c_1\, \mathcal{F}_c^{\kappa\lambda}\, \mathcal{F}^c_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \mathcal{F}^s_{\rho\sigma} + 4\, c_1\, \mathcal{F}_c^{\kappa\lambda}\, \mathcal{F}^s_{\kappa\lambda}\, \mathcal{F}_c^{\rho\sigma}\, \mathcal{F}^s_{\rho\sigma} + 2\, c_2\, \mathcal{F}_c^{\kappa\lambda}\, \widetilde{\mathcal{F}}^c_{\kappa\lambda}\, \mathcal{F}_s^{\rho\sigma}\, \widetilde{\mathcal{F}}^s_{\rho\sigma} + 4\, c_2\, \mathcal{F}_c^{\kappa\lambda}\, \widetilde{\mathcal{F}}^s_{\kappa\lambda}\, \mathcal{F}_c^{\rho\sigma}\, \widetilde{\mathcal{F}}^s_{\rho\sigma}\right\} \tag{4.53}$$

Eq. (4.52) with Eq. (4.53) within corresponds to the tadpole diagram with a soft loop:

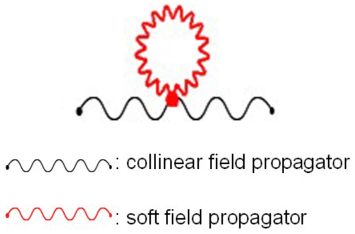


Figure 4.12: Relevant diagram to the Scharnhorst effect in the $\mathcal{L}_{EH}^{SC}$ theory.

This is what we now wish to evaluate. Before we do so however, important remarks are in order regarding the propagators involved.

4.5.2 Expressions for the propagators

Calculating the tadpole diagram of interest requires expressions for the propagators it contains. In this respect there is a crucial difference between the propagators that form the external legs and those that form the loop. The difference in question is actually not related to their collinear *vs* soft character – indeed, we find the same issue in the standard Euler-Heisenberg theory. Instead, it consists in the fact that the propagator in the loop "feels" the plates as it were: it is modified by their presence which constitutes boundary conditions on this propagator, because it is off-shell. This is the feature that is directly responsible for the Scharnhorst effect: in our model, the fact that we are dealing with a Casimir vacuum rather than trivial vacuum is instantiated by the boundary conditions imposed on the propagator that forms the loop.

In contrast, because they correspond to an on-shell photon, the propagators forming the external legs are not affected by the plates. As a result they only involve the propagator found in the usual, unbounded, trivial vacuum.

The propagator characteristic of the Casimir vacuum was derived by Bordag *et al* [8]. Following their convention I shall denote it $\overline{D}^{\mu\nu}(x, y)$ and the unbounded, trivial vacuum propagator will be represented by $D^{c\,\mu\nu}(x - y)$.

$$
\begin{aligned}
D^{c\,\mu\nu}(x-y) &= i\int \frac{d^{\,4}k}{(2\pi)^4} e^{-ik(x-y)}\tilde{D}^{\mu\nu}(k) \\
D^{c\,\mu\nu}(x-y) &= i\int \frac{d^{\,4}k}{(2\pi)^4} e^{-ik(x-y)}\frac{-i}{k^2+i\epsilon}\left(g^{\mu\nu}-(1-\xi)\frac{k^\mu k^\nu}{k^2}\right) \\
D^{c\,\mu\nu}(x-y) &= \int \frac{d^{\,4}k}{(2\pi)^4} e^{-ik(x-y)}\frac{1}{k^2+i\epsilon}\left(g^{\mu\nu}-(1-\xi)\frac{k^\mu k^\nu}{k^2}\right). \qquad (4.54)
\end{aligned}
$$

In contrast, in the Casimir vacuum propagator $\overline{D}^{\mu\nu}(x, y)$ the presence of the plates has broken the original symmetry found in $D^{c\,\mu\nu}(x-y)$ in the direction normal to the plates (the z-direction represented by the subscript 3):

$$\overline{D}^{\mu\nu}(x,y) = -\frac{1}{2}\int \frac{d^{\,3}\tilde{k}}{(2\pi)^3}\frac{i}{\Gamma}\left(\tilde{g}^{\,\mu\nu} - \frac{\tilde{k}^{\mu}\tilde{k}^{\nu}}{\Gamma^2}\right) e^{ip_\sigma(x-y)^\sigma} e^{i\Gamma|x_3-a_i|}\left(h^{-1}\right)_{ij} e^{i\Gamma|y_3-a_j|}, \tag{4.55}$$

k where $\Gamma = \Gamma(k) = (k2 - k_1^2 - k_2^2)^{\frac{1}{2}}$,

$$\tilde{k}^{\,\mu} = \begin{cases} k^{\,\mu} & \text{for } \mu \neq 3 \\ 0 & \text{for } \mu = 3 \end{cases} \qquad \tilde{g}^{\,\mu\nu} = \begin{cases} g^{\,\mu\nu} & \text{for } \mu,\nu \neq 3 \\ 0 & \text{for } \mu = 3 \text{ or } \nu = 3 \end{cases} \tag{4.56}$$

and:

$$h_{ij}^{-1} = \frac{i}{2\,Sin\left(\Gamma|a_0-a_1|\right)}\begin{pmatrix} e^{-i\Gamma|a_0-a_1|} & -1 \\ -1 & e^{-i\Gamma|a_0-a_1|} \end{pmatrix}, \tag{4.57}$$

with a_0 and a_1 the positions of the plates along the z-direction.[24]

4.5.3 Evaluating the contributions to the tadpole diagram and self-energy tensor

In order to calculate the phase velocity of the probe photon, we need to find the contributions to the vacuum polarization tensor (*i.e.* photon self-energy tensor) responsible for the effect. As discussed above, the latter is due to the tadpole diagram with collinear external legs and soft loop, as given by Eq. (4.52) for the situation at hand, *i.e.* with $\mathfrak{L}$ given by Eq. (4.53):

$$\mathfrak{L}^{2nd\ term} = \frac{\alpha^2}{m_e^4}$$
$$\left\{2\,c_1\,\mathfrak{F}_c^{\kappa\lambda}\,\mathfrak{F}^c_{\kappa\lambda}\,\mathfrak{F}_s^{\rho\sigma}\,\mathfrak{F}^s_{\rho\sigma} + 4\,c_1\,\mathfrak{F}_c^{\kappa\lambda}\,\mathfrak{F}^s_{\kappa\lambda}\,\mathfrak{F}_c^{\rho\sigma}\,\mathfrak{F}^s_{\rho\sigma} + 2\,c_2\,\mathfrak{F}_c^{\kappa\lambda}\,\widetilde{\mathfrak{F}}^c_{\kappa\lambda}\,\mathfrak{F}_s^{\rho\sigma}\,\widetilde{\mathfrak{F}}^s_{\rho\sigma} + 4\,c_2\,\mathfrak{F}_c^{\kappa\lambda}\,\widetilde{\mathfrak{F}}^s_{\kappa\lambda}\,\mathfrak{F}_c^{\rho\sigma}\,\widetilde{\mathfrak{F}}^s_{\rho\sigma}\right\}$$

---

[24] Hence h_{ij}^{-1} only depends on the distance between the plates $|a_0 - a_1|$ and $\Gamma(p)$. Also, the matrix h_{ij} itself is: $h_{ij} = e^{i\Gamma|a_i-a_j|}$.

Each of these four terms gives a contribution to the tadpole diagram, which we shall derive in the present section. This shall directly yield the relevant contributions to the vacuum polarization tensor $\Pi^{\mu\nu}$. Indeed, $i\ \Pi^{\mu\nu}$ is equal to the one-particle-irreducible diagram; and in the present case, as illustrated in Figs.(4.9) and (4.10) the 1-particle-irreducible diagram is simply the tadpole diagram. Therefore, the contributions of each term in Eq. (4.53) to the tadpole diagram also constitute their contributions to the polarization tensor.

In the fully interacting theory the photon propagator is:

$$\begin{aligned} & iD^{\mu\nu}(x-y)^{2nd\ term} \\ & \approx \lim_{t\to\infty(1-i\epsilon)} \int DA \left\langle 0 \middle| T\left\{ A^{\mu}(x)A^{\nu}(y)\ i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3z\ \mathcal{L}\right\} \middle| 0 \right\rangle . \end{aligned} \tag{4.58}$$

We shall treat each of the four terms in $\mathcal{L}$ separately. For now however, let us describe the procedure in some detail taking the first term as an example. That is, let us consider:

$$\begin{aligned} & \left\langle 0 \middle| T\left\{ A^{\mu}_c(x)A^{\nu}_c(y)\ i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3\vec{z}\ \ \left(\mathcal{F}^{\kappa\lambda}_c(z)\,\mathcal{F}^{c}_{\kappa\lambda}(z)\,\mathcal{F}^{\rho\sigma}_s(z)\,\mathcal{F}^{s}_{\rho\sigma}(z)\right)\right\} \middle| 0 \right\rangle \\ = \ \ & i\int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^3\vec{z}\ \ \left\langle 0 \,\middle|\, T\left\{A^{\mu}_c(x)A^{\nu}_c(y)\mathcal{F}^{\kappa\lambda}_c(z)\,\mathcal{F}^{c}_{\kappa\lambda}(z)\,\mathcal{F}^{\rho\sigma}_s(z)\,\mathcal{F}^{s}_{\rho\sigma}(z)\right\} \middle|\, 0\right\rangle . \end{aligned} \tag{4.59}$$

Following standard QFT procedure, the time-ordering is dealt with by performing Wick contractions. The number of terms this involves is reduced by the fact that collinear fields only contract with other collinear fields (and similarly for soft ones). Consequently the contribution of the term:

$$\begin{aligned} & \left\langle 0 \right| T\left\{ A^{\mu}_{c}(x) A^{\nu}_{c}(y) \mathcal{F}^{\kappa\lambda}_{c}(z)\, \mathcal{F}^{c}_{\kappa\lambda}(z)\, \mathcal{F}^{\rho\sigma}_{s}(z)\, \mathcal{F}^{s}_{\rho\sigma}(z) \right\} \left| 0 \right\rangle \\ = \quad & \left\langle 0 \right| T\left\{ A^{\mu}_{c}(x) A^{\nu}_{c}(y) \Big(\partial^{\kappa} A^{\lambda}_{c}\ (z) - \partial A^{\kappa}_{c}(z) \Big) \Big(\partial_{\kappa} A^{c}_{\lambda}(z) - \partial_{\lambda} A^{c}_{\kappa}(z) \Big) \Big(\partial^{\rho} A^{\sigma}_{s}(z) - \partial^{\sigma} A^{\rho}_{s}(z) \Big) \Big(\partial_{\rho} A^{s}_{\sigma}(z) - \partial_{\sigma} A^{s}_{\rho}(z) \Big) \right\} \left| 0 \right\rangle \end{aligned} \tag{4.60}$$

consists in only two series of contractions:

$$\left\langle 0 \right| \wick{ \c1{A^{\mu}_{c}(x)}\ \c2{A^{\nu}_{c}(y)}\ \Big(\c1{\partial^{\kappa} A^{\lambda}_{c}\ (z)} - \partial^{\lambda} A^{\kappa}_{c}(z) \Big) \Big(\c2{\partial_{\kappa} A^{c}_{\lambda}(z)} - \partial_{\lambda} A^{c}_{\kappa}(z) \Big) \Big(\c3{\partial^{\rho} A^{\sigma}_{s}(z)} - \partial^{\sigma} A^{\rho}_{s}(z) \Big) \Big(\c3{\partial_{\rho} A^{s}_{\sigma}(z)} - \partial_{\sigma} A^{s}_{\rho}(z) \Big) } \left| 0 \right\rangle \tag{4.61}$$

and

$$\left\langle 0 \right| \wick{ \c1{A^{\mu}_{c}(x)}\ \c2{A^{\nu}_{c}(y)}\ \Big(\c2{\partial^{\kappa} A^{\lambda}_{c}\ (z)} - \partial^{\lambda} A^{\kappa}_{c}(z) \Big) \Big(\c1{\partial_{\kappa} A^{c}_{\lambda}(z)} - \partial_{\lambda} A^{c}_{\kappa}(z) \Big) \Big(\c3{\partial^{\rho} A^{\sigma}_{s}(z)} - \partial^{\sigma} A^{\rho}_{s}(z) \Big) \Big(\c3{\partial_{\rho} A^{s}_{\sigma}(z)} - \partial_{\sigma} A^{s}_{\rho}(z) \Big) } \left| 0 \right\rangle \tag{4.62}$$

That is, by Wick's theorem:

$$\begin{aligned} & \left\langle 0 \right| T\left\{ A^{\mu}_{c}(x) A^{\nu}_{c}(y)\ i \int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^{3}\vec{z}\ \Big(\mathcal{F}^{\kappa\lambda}_{c}(z)\, \mathcal{F}^{c}_{\kappa\lambda}(z)\, \mathcal{F}^{\rho\sigma}_{s}(z)\, \mathcal{F}^{s}_{\rho\sigma}(z) \Big) \right\} \left| 0 \right\rangle = i \int_{-t}^{+t} dt' \int_{-\infty}^{+\infty} d^{3}\vec{z} \\ & \left\langle 0 \right| \Bigg(\wick{ \c1{A^{\mu}_{c}(x)}\ \c2{A^{\nu}_{c}(y)}\ \Big(\c1{\partial^{\kappa} A^{\lambda}_{c}\ (z)} - \partial^{\lambda} A^{\kappa}_{c}(z) \Big) \Big(\c2{\partial_{\kappa} A^{c}_{\lambda}(z)} - \partial_{\lambda} A^{c}_{\kappa}(z) \Big) \Big(\c3{\partial^{\rho} A^{\sigma}_{s}(z)} - \partial^{\sigma} A^{\rho}_{s}(z) \Big) \Big(\c3{\partial_{\rho} A^{s}_{\sigma}(z)} - \partial_{\sigma} A^{s}_{\rho}(z) \Big) } \\ + \quad & \wick{ \c1{A^{\mu}_{c}(x)}\ \c2{A^{\nu_{c}}(y)}\ \Big(\c2{\partial^{\kappa} A^{\lambda}_{c}\ (z)} - \partial^{\lambda} A^{\kappa}_{c}(z) \Big) \Big(\c1{\partial_{\kappa} A^{c}_{\lambda}(z)} - \partial_{\lambda} A^{c}_{\kappa}(z) \Big) \Big(\c3{\partial^{\rho} A^{\sigma}_{s}(z)} - \partial^{\sigma} A^{\rho}_{s}(z) \Big) \Big(\c3{\partial_{\rho} A^{s}_{\sigma}(z)} - \partial_{\sigma} A^{s}_{\rho}(z) \Big) } \Bigg) \left| 0 \right\rangle . \end{aligned} \tag{4.63}$$

As always the sets of contractions 4.61 and 4.62 involve propagators, and more specifically here *derivatives* of propagators, according to the following patterns:

$$\wick{\c1{A^{\mu}(x)}\ {}_{z}\partial_{\sigma}\c1{A_{\rho}(z)}} = {}_{z}\partial_{\sigma} D^{\mu}_{\rho}(y-z) \quad ; \quad \wick{\c1{A^{\mu}(x)}\ {}_{z}\partial^{\sigma}\c1{A^{\rho}\ (z)}} = {}_{z}\partial^{\sigma} D^{\mu\rho}\ (y-z) \tag{4.64}$$

where ${}_z\partial$ stands for the partial derivative operating at spacetime event z. As discussed in the previous section, whether the propagators at stake are $D^{c\,\mu\nu}(x-y)$ or $\overline{D}^{\mu\nu}(x-y)$ depends on whether they enter into external legs or loops, respectively. We can see that in both the expressions Eq. (4.61) and Eq. (4.62), the contraction between the last two factors $\left(i.e.\ \left(\partial^\rho A^\sigma_s(z) - \partial^\sigma A^\rho_s(z)\right)\left(\partial_\rho A^s_\sigma(z) - \partial_\sigma A^s_\rho(z)\right)\right)$ involves only fields at the spacetime event z. Hence they correspond to loops when expressed in the language of Feynman diagrams. The other contractions in Eq. (4.61) and Eq. (4.62) connect fields evaluated at z with fields evaluated at either x or y. They correspond to external legs. Consequently, they involve the propagator for the trivial, unbounded vacuum: it is the RHS of Eq. (4.54) that is used in the expressions of Eq. (4.64) to determine these contractions.

Contribution of the non-dual symmetric term $\mathcal{F}^{\kappa\lambda}_c\,\mathcal{F}^c_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\mathcal{F}^s_{\rho\sigma}$

The first contribution is given by the term: $2\,c_1\,\mathcal{F}^{\kappa\lambda}_c\,\mathcal{F}^c_{\kappa\lambda}\,\mathcal{F}^{\rho\sigma}_s\,\mathcal{F}^s_{\rho\sigma}$, which yields the following loop contraction:

$$\begin{aligned}
\wick{\c{\mathcal{F}}^{\rho\sigma}_s(z)\,\c{\mathcal{F}}^s_{\rho\sigma}(z)} &= \wick{\c{\mathcal{F}}^{\rho\sigma}_s(u)\,\c{\mathcal{F}}^s_{\rho\sigma}(v)}\Big|_{u=z,v=z} \\
&= \wick{\Big({}_u\partial^\rho \c{A}^\sigma_s(u) - {}_u\partial^\sigma A^\rho_s(u)\Big)\Big({}_v\partial_\rho \c{A}^s_\sigma(v) - {}_v\partial_\sigma A^s_\rho(v)\Big)}\Big|_{u=z,v=z} \\
&= 2\,\Big({}_u\partial^\rho\ {}_v\partial_\rho\ \overline{D}^\sigma{}_\sigma(u,v) - {}_u\partial_\rho\ {}_v\partial_\sigma\ \overline{D}^{\sigma\rho}(u,v)\Big)\Big|_{u=z,v=z},
\end{aligned} \tag{4.65}$$

where we have used the fact that $\overline{D}$ is symmetric.

The external legs contractions are given by:

$$
\begin{aligned}
\contraction{}{A}{^\mu(x)\,}{\mathcal{F}}
\contraction{A^\mu(x)\,\mathcal{F}_c^{\kappa\lambda}(z)\;}{\mathcal{F}}{^c_{\kappa\lambda}(z)\,}{A}
A^\mu(x)\,\mathcal{F}_c^{\kappa\lambda}(z)\;\mathcal{F}^c_{\kappa\lambda}(z)\,A^\nu(y) = \quad & {}_x\partial^\kappa\ D^{\mu\lambda}(x-z)\ {}_y\partial_\kappa\ D_\lambda{}^\nu(z-y) \\
- \quad & {}_x\partial^\kappa\ D^{\mu\lambda}(x-z)\ {}_y\partial_\lambda\ D_\kappa{}^\nu(z-y) \\
- \quad & {}_x\partial^\lambda\ D^{\mu\kappa}(x-z)\ {}_y\partial_\kappa\ D_\lambda{}^\nu(z-y) \\
+ \quad & {}_x\partial^\lambda\ D^{\mu\kappa}(x-z)\ {}_y\partial_\lambda\ D_\kappa{}^\nu(z-y).
\end{aligned}
\qquad (4.66)
$$

There is also a second set of external legs contractions:

$$
\begin{aligned}
\contraction{}{A}{^\mu(x)\,\mathcal{F}_c^{\kappa\lambda}(z)\;}{\mathcal{F}}
\contraction[2ex]{A^\mu(x)\,}{\mathcal{F}}{_c^{\kappa\lambda}(z)\;\mathcal{F}^c_{\kappa\lambda}(z)\,}{A}
A^\mu(x)\,\mathcal{F}_c^{\kappa\lambda}(z)\;\mathcal{F}^c_{\kappa\lambda}(z)\,A^\nu(y) = \quad & {}_x\partial_\kappa\ D^\mu{}_\lambda(x-z)\ {}_y\partial^\kappa\ D^{\lambda\nu}(z-y) \\
- \quad & {}_x\partial_\kappa\ D^\mu{}_\lambda(x-z)\ {}_y\partial^\lambda\ D^{\kappa\nu}(z-y) \\
- \quad & {}_x\partial_\lambda\ D^\mu{}_\kappa(x-z)\ {}_y\partial^\kappa\ D^{\lambda\nu}(z-y) \\
+ \quad & {}_x\partial_\lambda\ D^\mu{}_\kappa(x-z)\ {}_y\partial^\lambda\ D^{\kappa\nu}(z-y)
\end{aligned}
\qquad (4.67)
$$

which is the same as the expression for the first set of legs contractions obtained in Eq. (4.66).

Overall the expressions that the first term contributes to the tadpole diagram can be written:

$$
\begin{aligned}
& \contraction{}{A}{^\mu(x)\,}{\mathcal{F}}
\contraction{A^\mu(x)\,\mathcal{F}_c^{\kappa\lambda}(z)\;}{\mathcal{F}}{_s^{\rho\sigma}(z)\,}{\mathcal{F}}
\contraction{A^\mu(x)\,\mathcal{F}_c^{\kappa\lambda}(z)\;\mathcal{F}_s^{\rho\sigma}(z)\,\mathcal{F}^s_{\rho\sigma}(z)\,}{\mathcal{F}}{^c_{\kappa\lambda}(z)\,}{A}
A^\mu(x)\,\mathcal{F}_c^{\kappa\lambda}(z)\;\mathcal{F}_s^{\rho\sigma}(z)\,\mathcal{F}^s_{\rho\sigma}(z)\,\mathcal{F}^c_{\kappa\lambda}(z)\,A^\nu(y) \\
= \; & \contraction{}{A}{^\mu(x)\,}{\mathcal{F}}
\contraction{A^\mu(x)\,\mathcal{F}_c^{\kappa\lambda}(z)\;}{\mathcal{F}}{_s^{\rho\sigma}(u)\,}{\mathcal{F}}
\contraction{A^\mu(x)\,\mathcal{F}_c^{\kappa\lambda}(z)\;\mathcal{F}_s^{\rho\sigma}(u)\,\mathcal{F}^s_{\rho\sigma}(v)\Big|_{u=z,v=z}}{\mathcal{F}}{^c_{\kappa\lambda}(z)\,}{A}
A^\mu(x)\,\mathcal{F}_c^{\kappa\lambda}(z)\;\mathcal{F}_s^{\rho\sigma}(u)\,\mathcal{F}^s_{\rho\sigma}(v)\Big|_{u=z,v=z}\mathcal{F}^c_{\kappa\lambda}(z)\,A^\nu(y) \\
= \; & 4\ \Big(\ D^{\mu\lambda}(x-z)\ {}_v\partial_\rho\ {}_u\partial^\rho\ {}_x\partial_\kappa\ {}_y\partial^\kappa\ D_\lambda{}^\nu(z-y)\ \ \overline{D}^\sigma{}_\sigma(u,v) \\
& - \quad D^{\mu\lambda}(x-z)\ {}_v\partial_\rho\ {}_u\partial^\rho\ {}_x\partial_\kappa\ {}_y\partial_\lambda\ D^{\kappa\,\nu}(z-y)\ \ \overline{D}^\sigma{}_\sigma(u,v) \\
& - \quad D^{\mu\lambda}(x-z)\ {}_v\partial_\sigma\ {}_u\partial_\rho\ {}_x\partial_\kappa\ {}_y\partial^\kappa\ D_\lambda{}^\nu(z-y)\ \ \overline{D}^{\sigma\rho}(u,v) \\
& + \quad D^{\mu\lambda}(x-z)\ {}_v\partial_\sigma\ {}_u\partial_\rho\ {}_x\partial_\kappa\ {}_y\partial_\lambda\ D^{\kappa\,\nu}(z-y)\ \ \overline{D}^{\sigma\rho}(u,v)\Big)\Big|_{u=z,v=z}.
\end{aligned}
\qquad (4.68)
$$

In order to express the different contributions to the tadpole in a consistent way, and for ease of comparison with Scharnhorst's original work, let us re-write this expression with all the indices of the propagators contravariant, and let us re-label dummy indices in such a way that those on $\overline{D}$ are $\alpha\beta$:

$$\begin{aligned}
&\wick{\c1 A^{\mu}(x)\ \c1 \mathcal{F}_c^{\kappa\lambda}(z)\ \ \c2 \mathcal{F}_s^{\rho\sigma}(z)\, \c2 \mathcal{F}^s_{\rho\sigma}(z)\ \c3 \mathcal{F}^c_{\kappa\lambda}(z)\ \c3 A^{\nu}(y)} \\
= \ & 4\ \Big(D^{\mu\eta}(x-z)\ {}_v\partial_\rho\ {}_u\partial^\rho\ {}_x\partial_\kappa\ {}_y\partial^\kappa\ g_{\eta\tau}\ D^{\tau\nu}(z-y)\ g_{\beta\alpha}\ \overline{D}^{\alpha\beta}(u,v) \\
- \ & \ D^{\mu\eta}(x-z)\ {}_v\partial_\rho\ {}_u\partial^\rho\ {}_x\partial_\tau\ {}_y\partial_\eta\ D^{\tau\,\nu}(z-y)\ g_{\beta\alpha}\ \overline{D}^{\alpha\beta}(u,v) \\
- \ & \ D^{\mu\eta}(x-z)\ {}_v\partial_\alpha\ {}_u\partial_\beta\ {}_x\partial_\kappa\ {}_y\partial^\kappa\ g_{\eta\tau}\ D^{\tau\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v) \\
- \ & \ D^{\mu\eta}(x-z)\ {}_v\partial_\alpha\ {}_u\partial_\beta\ {}_x\partial_\tau\ {}_y\partial_\eta\ D^{\tau\,\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v)\Big)\Big|_{u=z,v=z}. \qquad (4.69)
\end{aligned}$$

The second contribution:

$$\begin{aligned}
&\wick{\c1 A^{\mu}(x)\ \c2 \mathcal{F}_c^{\kappa\lambda}(z)\ \ \c3 \mathcal{F}_s^{\rho\sigma}(z)\, \c3 \mathcal{F}^s_{\rho\sigma}(z)\ \ \c1 \mathcal{F}^c_{\kappa\lambda}(z)\ \c2 A^{\nu}(y)} \\
= \ & \wick{\c1 A^{\mu}(x)\ \c1 \mathcal{F}^c_{\kappa\lambda}(z)\ \ \c2 \mathcal{F}_s^{\rho\sigma}(u)\, \c2 \mathcal{F}^s_{\rho\sigma}(v)}\Big|_{u=z,v=z} \wick{\c1 \mathcal{F}_c^{\kappa\lambda}(z)\ \c1 A^{\nu}(y)}, \qquad (4.70)
\end{aligned}$$

yields the same result.

In order to evaluate the polarization tensor, we actually do not need the external legs per se, *i.e.* we do not need the propagators $D(x-z)$ and $D(z-y)$. At the same time, we wish to rewrite Eq. (4.69) and Eq. (4.70) in a form easily comparable to what can be found in Scharnhorst's original work:[25]

$$\begin{aligned}
\mathrm{C1a} \ = \ \mathrm{C1b} \ \ = & \\
4\ \Big(& \quad {}_v\partial_\rho\ {}_u\partial^\rho\ {}_y\partial^\kappa\ {}_x\partial_\kappa\ g_{\eta\tau}\ g_{\alpha\beta} - {}_v\partial_\rho\ {}_u\partial^\rho\ {}_y\partial_\eta\ {}_x\partial_\tau\ g_{\alpha\beta} \\
- & \quad {}_v\partial_\alpha\ {}_u\partial_\beta\ {}_y\partial^\kappa\ {}_x\partial_\kappa\ g_{\eta\tau} + {}_v\partial_\alpha\ {}_u\partial_\beta\ {}_y\partial_\eta\ {}_x\partial_\tau \Big)\ \overline{D}^{\alpha\beta}(u,v)\Big|_{u=z,v=z}, \qquad (4.71)
\end{aligned}$$

---

[25] *i.e.* in the standard notation where the indices on all derivatives are covariant and those of D contravariant.

where C1a refers to the contribution associated with the first set of external legs contraction, and C1b the second one.

Contribution of the non-dual non-symmetric term $\mathcal{F}_c^{\kappa\lambda}\,\mathcal{F}^s_{\kappa\lambda}\,\mathcal{F}_c^{\rho\sigma}\,\mathcal{F}^s_{\rho\sigma}$

The second contribution is due to the term: $4\,c_1\,\mathcal{F}_c^{\kappa\lambda}\,\mathcal{F}^s_{\kappa\lambda}\,\mathcal{F}_c^{\rho\sigma}\,\mathcal{F}^s_{\rho\sigma}$, which yields the following loop contraction:

$$\begin{aligned}
\wick{\c{\mathcal{F}}^s_{\kappa\lambda}(z)\,\c{\mathcal{F}}^s_{\rho\sigma}(z)} &= \wick{\c{\mathcal{F}}^s_{\kappa\lambda}(u)\,\c{\mathcal{F}}^s_{\rho\sigma}(v)}\Big|_{u=z,v=z} \\
&= \wick{\Big({}_u\partial_\kappa A^s_\lambda(u) - {}_u\partial_\lambda \c A^s_\kappa(u)\Big)\Big({}_v\partial_\rho A^s_\sigma(v) - {}_v\partial_\sigma \c A^s_\rho(v)\Big)}\Big|_{u=z,v=z} \\
&= \Big(\,{}_u\partial_\kappa\ {}_v\partial_\rho\ \overline{D}_{\lambda\sigma}(u,v) - {}_u\partial_\kappa\ {}_v\partial_\sigma\ \overline{D}_{\lambda\rho}(u,v) \\
&\quad -{}_u\partial_\lambda\ {}_v\partial_\rho\ \overline{D}_{\kappa\sigma}(u,v) + {}_u\partial_\lambda\ {}_v\partial_\sigma\ \overline{D}_{\kappa\rho}(u,v)\Big)\Big|_{u=z,v=z}. \qquad (4.72)
\end{aligned}$$

The external legs contractions due to this second contribution are:

$$\begin{aligned}
\wick{\c A^\mu(x)\ \c{\mathcal{F}}_c^{\kappa\lambda}(z)\ \c{\mathcal{F}}_c^{\rho\sigma}(z)\ \c A^\nu(y)} = &\quad {}_x\partial^\kappa\ D^{\mu\lambda}(x-z)\ {}_y\partial^\rho\ D^{\sigma\nu}(z-y) \\
&- {}_x\partial^\kappa\ D^{\mu\lambda}(x-z)\ {}_y\partial^\sigma\ D^{\rho\nu}(z-y) \\
&- {}_x\partial^\lambda\ D^{\mu\kappa}(x-z)\ {}_y\partial^\rho\ D^{\sigma\nu}(z-y) \\
&+ {}_x\partial^\lambda\ D^{\mu\kappa}(x-z)\ {}_y\partial^\sigma\ D^{\rho\nu}(z-y), \qquad (4.73)
\end{aligned}$$

and:

$$\begin{aligned}
\wick{\c1 A^\mu(x)\ \c2{\mathcal{F}}_c^{\kappa\lambda}(z)\ \c1{\mathcal{F}}_c^{\rho\sigma}(z)\ \c2 A^\nu(y)} = &\quad {}_x\partial^\rho\ D^{\mu\sigma}(x-z)\ {}_y\partial^\kappa\ D^{\lambda\nu}(z-y) \\
&- {}_x\partial^\rho\ D^{\mu\sigma}(x-z)\ {}_y\partial^\lambda\ D^{\kappa\nu}(z-y) \\
&- {}_x\partial^\sigma\ D^{\mu\rho}(x-z)\ {}_y\partial^\kappa\ D^{\lambda\nu}(z-y) \\
&+ {}_x\partial^\sigma\ D^{\mu\rho}(x-z)\ {}_y\partial^\lambda\ D^{\kappa\nu}(z-y). \qquad (4.74)
\end{aligned}$$

So overall the set of contractions that represent the contribution of the second term to the tadpole diagram can be written:
For the first set of legs contraction:

$$\begin{aligned}
& \overbracket{A^\mu(x)\ \mathcal{F}_c^{\kappa\lambda}}(z)\ \ \overbracket{\mathcal{F}^s_{\kappa\lambda}(z)\,\mathcal{F}^s_{\rho\sigma}}(z)\ \overbracket{\mathcal{F}_c^{\rho\sigma}(z)\ A^\nu}(y) \\
= \ & \overbracket{A^\mu(x)\ \mathcal{F}_c^{\kappa\lambda}}(z)\ \ \overbracket{\mathcal{F}^s_{\kappa\lambda}(u)\,\mathcal{F}^s_{\rho\sigma}}(v)\Big|_{u=z,v=z} \overbracket{\mathcal{F}_c^{\rho\sigma}(z)\ A^\nu}(y) \\
= \ & 4\ \Big(\ D^\mu{}_\lambda(x-z)\ {}_v\partial_\rho\ {}_y\partial^\rho\ {}_u\partial_\kappa\ {}_x\partial^\kappa\ D_\sigma{}^\nu(z-y)\ \overline{D}^{\lambda\sigma}(u,v) \\
- \ & \quad D^{\mu\kappa}(x-z)\ {}_v\partial_\rho\ {}_y\partial^\rho\ {}_u\partial_\kappa\ {}_x\partial_\lambda\ D_\sigma{}^\nu(z-y)\ \overline{D}^{\lambda\sigma}(u,v) \\
- \ & \quad D^\mu{}_\lambda(x-z)\ {}_v\partial_\rho\ {}_y\partial_\sigma\ {}_u\partial_\kappa\ {}_x\partial^\kappa\ D^{\rho\nu}(z-y)\ \ \overline{D}^{\lambda\sigma}(u,v) \\
+ \ & \quad D^{\mu\kappa}(x-z)\ {}_v\partial_\rho\ {}_y\partial_\sigma\ {}_u\partial_\kappa\ {}_x\partial_\lambda\ D^{\rho\nu}(z-y)\ \ \overline{D}^{\lambda\sigma}(u,v)\Big)\Big|_{u=z,v=z}.
\end{aligned} \tag{4.75}$$

This can be re-written as:

$$\begin{aligned}
& \overbracket{A^\mu(x)\ \mathcal{F}_c^{\kappa\lambda}}(z)\ \ \overbracket{\mathcal{F}^s_{\kappa\lambda}(z)\,\mathcal{F}^s_{\rho\sigma}}(z)\ \overbracket{\mathcal{F}_c^{\rho\sigma}(z)\ A^\nu}(y) \\
= \ & 4\ \Big(\ D^{\mu\eta}(x-z)\ g_{\eta\alpha}\ {}_v\partial_\rho\ {}_y\partial^\rho\ {}_u\partial_\kappa\ {}_x\partial^\kappa\ g_{\beta\tau}\ D^{\tau\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v) \\
- \ & \quad D^{\mu\eta}(x-z)\ {}_v\partial_\rho\ {}_y\partial^\rho\ {}_u\partial_\eta\ {}_x\partial_\alpha\ g_{\beta\tau}\ D^{\tau\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v) \\
- \ & \quad D^{\mu\eta}(x-z)\ g_{\eta\alpha}\ {}_v\partial_\tau\ {}_y\partial_\beta\ {}_u\partial_\kappa\ {}_x\partial^\kappa\ D^{\tau\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v) \\
+ \ & \quad D^{\mu\eta}(x-z)\ {}_v\partial_\tau\ {}_y\partial_\beta\ {}_u\partial_\eta\ {}_x\partial_\alpha\ D^{\tau\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v)\Big)\Big|_{u=z,v=z}.
\end{aligned} \tag{4.76}$$

Rewriting this without the propagators of the external legs as we did to obtain (Eq. 4.71), the above takes the form:

$$\begin{aligned}
\text{C2a} \ = & \\
& 4\ \Big(g_{\eta\alpha}\ {}_u\partial_\kappa\ {}_x\partial^\kappa\ {}_v\partial_\rho\ {}_y\partial^\rho\ g_{\beta\tau}\ \overline{D}^{\alpha\beta}(u,v)\ \ -\ {}_u\partial_\eta\ {}_x\partial_\alpha\ {}_v\partial_\rho\ {}_y\partial^\rho\ g_{\beta\tau}\ \overline{D}^{\alpha\beta}(u,v) \\
- \quad & g_{\eta\alpha}\ {}_u\partial_\kappa\ {}_x\partial^\kappa\ {}_v\partial_\tau\ {}_y\partial_\beta\ \overline{D}^{\alpha\beta}(u,v)\ \ +\ {}_u\partial_\eta\ {}_x\partial_\alpha\ {}_v\partial_\tau\ {}_y\partial_\beta\ \overline{D}^{\alpha\beta}(u,v)\Big)\Big|_{u=z,v=z}.
\end{aligned} \tag{4.77}$$

For the second set of legs contraction:

$$\begin{aligned}
&\wick{\c1 A^{\mu}(x)\ \c2 \mathcal{F}_c^{\kappa\lambda}(z)\ \c3 \mathcal{F}^s_{\kappa\lambda}(z)\, \c3 \mathcal{F}^s_{\rho\sigma}(z)\ \c1 \mathcal{F}_c^{\rho\sigma}(z)\ \c2 A^{\nu}(y)} \\
= \ & \wick{\c1 A^{\mu}(x)\ \c1 \mathcal{F}_c^{\rho\sigma}(z)\ \ \c2 \mathcal{F}^s_{\kappa\lambda}(z)\, \c2 \mathcal{F}^s_{\rho\sigma}(z)\ \c3 \mathcal{F}_c^{\kappa\lambda}(z)\ \c3 A^{\nu}(y)} \\
= \ & \wick{\c1 A^{\mu}(x)\ \c1 \mathcal{F}_c^{\rho\sigma}(z)\ \ \c2 \mathcal{F}^s_{\kappa\lambda}(u)\, \c2 \mathcal{F}^s_{\rho\sigma}(v)}\Big|_{u=z,v=z} \wick{\c3 \mathcal{F}_c^{\kappa\lambda}(z)\ \c3 A^{\nu}(y)} \\
= \ & 4\,\Big(D^{\mu}{}_{\sigma}(x-z)\ {}_u\partial_{\kappa}\ {}_y\partial^{\kappa}\ {}_v\partial_{\rho}\ {}_x\partial^{\rho}\ D_{\lambda}{}^{\nu}(z-y)\ \overline{D}^{\lambda\sigma}(u,v) \\
- \ & D^{\mu\rho}(x-z)\ {}_u\partial_{\kappa}\ {}_y\partial^{\kappa}\ {}_v\partial_{\rho}\ {}_x\partial_{\sigma}\ D_{\lambda}{}^{\nu}(z-y)\ \overline{D}^{\lambda\sigma}(u,v) \\
- \ & D^{\mu}{}_{\sigma}(x-z)\ {}_u\partial_{\kappa}\ {}_y\partial_{\lambda}\ {}_v\partial_{\rho}\ {}_x\partial^{\rho}\ D^{\kappa\nu}(z-y)\ \overline{D}^{\lambda\sigma}(u,v) \\
+ \ & D^{\mu\rho}(x-z)\ {}_u\partial_{\kappa}\ {}_y\partial_{\lambda}\ {}_v\partial_{\rho}\ {}_x\partial_{\sigma}\ D^{\kappa\nu}(z-y)\ \overline{D}^{\lambda\sigma}(u,v)\Big)\Big|_{u=z,v=z}, \qquad (4.78)
\end{aligned}$$

$$\begin{aligned}
&\wick{\c1 A^{\mu}(x)\ \c2 \mathcal{F}_c^{\kappa\lambda}(z)\ \c3 \mathcal{F}^s_{\kappa\lambda}(z)\, \c3 \mathcal{F}^s_{\rho\sigma}(z)\ \c1 \mathcal{F}_c^{\rho\sigma}(z)\ \c2 A^{\nu}(y)} \\
= \ & 4\,\Big(D^{\mu\eta}(x-z)\ g_{\eta\beta}\ {}_u\partial_{\kappa}\ {}_y\partial^{\kappa}\ {}_v\partial_{\rho}\ {}_x\partial^{\rho}\ g_{\alpha\tau}\ D^{\tau\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v) \\
- \ & D^{\mu\eta}(x-z)\ {}_u\partial_{\kappa}\ {}_y\partial^{\kappa}\ {}_v\partial_{\eta}\ {}_x\partial_{\beta}\ g_{\alpha\tau}\ D^{\tau\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v) \\
- \ & D^{\mu\eta}(x-z)\ g_{\eta\beta}\ {}_u\partial_{\tau}\ {}_y\partial_{\alpha}\ {}_v\partial_{\rho}\ {}_x\partial^{\rho}\ D^{\tau\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v) \\
+ \ & D^{\mu\eta}(x-z)\ {}_u\partial_{\tau}\ {}_y\partial_{\alpha}\ {}_v\partial_{\eta}\ {}_x\partial_{\beta}\ D^{\tau\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v)\Big)\Big|_{u=z,v=z}. \qquad (4.79)
\end{aligned}$$

This yields the following contribution to the tadpole diagram:

$$\begin{aligned}
\text{C2b} \ = \ & \\
& 4\,\Big(g_{\eta\beta}\ {}_v\partial_{\rho}\ {}_x\partial^{\rho}\ {}_u\partial_{\kappa}\ {}_y\partial^{\kappa}\ g_{\alpha\tau}\ \overline{D}^{\alpha\beta}(u,v)\ -\ {}_v\partial_{\eta}\ {}_x\partial_{\beta}\ {}_u\partial_{\kappa}\ {}_y\partial^{\kappa}\ g_{\alpha\tau}\ \overline{D}^{\alpha\beta}(u,v) \\
- \ & \quad g_{\eta\beta}\ {}_v\partial_{\rho}\ {}_x\partial^{\rho}\ {}_u\partial_{\tau}\ {}_y\partial_{\alpha}\ \overline{D}^{\alpha\beta}(u,v)\ +\ {}_v\partial_{\eta}\ {}_x\partial_{\beta}\ {}_u\partial_{\tau}\ {}_y\partial_{\alpha}\ \overline{D}^{\alpha\beta}(u,v)\Big)\Big|_{u=z,v=z}.
\end{aligned}$$

(4.80)

Contribution of the dual symmetric term $\mathcal{F}_c^{\kappa\lambda}\,\widetilde{\mathcal{F}}^c_{\kappa\lambda}\,\mathcal{F}_s^{\rho\sigma}\,\widetilde{\mathcal{F}}^s_{\rho\sigma}$

The third contribution is that of the term: $2\,c_2\,\mathcal{F}_c^{\kappa\lambda}\,\widetilde{\mathcal{F}}^c_{\kappa\lambda}\,\mathcal{F}_s^{\rho\sigma}\,\widetilde{\mathcal{F}}^s_{\rho\sigma}$. It involves the following loop contraction:

$$\begin{aligned}
\wick{\c{\mathcal{F}}_s^{\rho\sigma}(z)\ \ \c{\widetilde{\mathcal{F}}}^s_{\rho\sigma}(z)} &= \wick{\c{\mathcal{F}}_s^{\rho\sigma}(u)\ \ \c{\widetilde{\mathcal{F}}}^s_{\rho\sigma}(v)}\Big|_{u=z,v=z} = \wick{\c{\mathcal{F}}_s^{\rho\sigma}(u)\ \ \epsilon_{\rho\sigma\chi\theta}\ \ \c{\mathcal{F}}_s^{\chi\theta}(v)}\Big|_{u=z,v=z} \\
&= \wick{\Big(\partial^\rho A_s^\sigma(u) - \c{\partial}^\sigma A_s^\rho(u)\Big)\ \epsilon_{\rho\sigma\chi\theta}\ \Big(\partial^\chi A_s^\theta(v) - \c{\partial}^\theta A_s^\chi(v)\Big)}\Big|_{u=z,v=z} \\
&= \Big(\ {}_v\partial^\rho\ \ \epsilon_{\rho\sigma\chi\theta}\ \ {}_u\partial^\chi\,\overline{D}^{\sigma\theta}(u,v) - {}_v\partial^\rho\ \ \epsilon_{\rho\sigma\chi\theta}\ \ {}_u\partial^\theta\,\overline{D}^{\sigma\chi}(u,v) \\
&\quad - {}_v\partial^\sigma\ \ \epsilon_{\rho\sigma\chi\theta}\ \ {}_u\partial^\chi\,\overline{D}^{\rho\theta}(u,v) + {}_v\partial^\sigma\ \ \epsilon_{\rho\sigma\chi\theta}\ \ {}_u\partial^\theta\,\overline{D}^{\rho\chi}(u,v)\Big)\Big|_{u=z,v=z}.
\end{aligned} \tag{4.81}$$

This yields a null tensor: indeed, each of the four terms in Eq. (4.81) involves contracting the indices of $\overline{D}$ with two of the indices of the Levi-Civita tensor $\epsilon_{\rho\sigma\chi\theta}$; now $\epsilon_{\rho\sigma\chi\theta}$ is antisymmetric, while $\overline{D}$ is a symmetric tensor, and contracting an antisymmetric tensor with a symmetric one yields a null result, so that:

$$\mathrm{C3a} = \mathrm{C3b} = 0. \tag{4.82}$$

Contribution of the dual non-symmetric term $\mathcal{F}_c^{\kappa\lambda}\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}_c^{\rho\sigma}\,\widetilde{\mathcal{F}}^s_{\rho\sigma}$

The fourth contribution is due to the term: $4\,c_2\,\mathcal{F}_c^{\kappa\lambda}\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}_c^{\rho\sigma}\,\widetilde{\mathcal{F}}^s_{\rho\sigma}$, which yields the following loop contraction:

$$
\begin{aligned}
\wick{\c{\widetilde{\mathfrak{F}}}^s_{\kappa\lambda}(z) \quad \c{\widetilde{\mathfrak{F}}}^s_{\rho\sigma}(z)} &= \wick{\c{\widetilde{\mathfrak{F}}}^s_{\kappa\lambda}(u) \quad \c{\widetilde{\mathfrak{F}}}^s_{\rho\sigma}(v)}\Big|_{u=z,v=z} = \epsilon_{\kappa\lambda\zeta\alpha} \; \wick{\c{\mathfrak{F}}^{\zeta\alpha}_s(u) \quad \epsilon_{\rho\sigma\chi\beta} \; \c{\mathfrak{F}}^{\chi\beta}_s(v)}\Big|_{u=z,v=z} \\
&= \quad \epsilon_{\kappa\lambda\zeta\alpha} \; \wick{\Big({}_u\partial^\zeta A^\alpha_s(u) - \c{{}_u\partial^\alpha} A^\zeta_s(u)\Big) \; \epsilon_{\rho\sigma\chi\beta} \; \Big({}_v\partial^\chi \c{A}^\beta_s(v)} - {}_v\partial^\beta A^\chi_s(v)\Big)\Big|_{u=z,v=z} \\
&= \Big(\; \epsilon_{\kappa\lambda\zeta\alpha} \;\; {}_u\partial^\zeta \;\; \epsilon_{\rho\sigma\chi\beta} \;\; {}_v\partial^\chi \;\; \overline{D}^{\alpha\beta}(u,v) \\
&\quad - \; \epsilon_{\kappa\lambda\zeta\alpha} \;\; {}_u\partial^\zeta \;\; \epsilon_{\rho\sigma\chi\beta} \;\; {}_v\partial^\beta \;\; \overline{D}^{\xi\chi}(u,v) \\
&\quad - \; \epsilon_{\kappa\lambda\zeta\alpha} \;\; {}_u\partial^\alpha \;\; \epsilon_{\rho\sigma\chi\beta} \;\; {}_v\partial^\chi \;\; \overline{D}^{\zeta\beta}(u,v) \\
&\quad + \; \epsilon_{\kappa\lambda\zeta\alpha} \;\; {}_u\partial^\alpha \;\; \epsilon_{\rho\sigma\chi\beta} \;\; {}_v\partial^\beta \;\; \overline{D}^{\zeta\chi}(u,v)\Big)\Big|_{u=z,v=z}. \qquad (4.83)
\end{aligned}
$$

Relabelling dummy indices and swapping indices on the ϵ tensors this becomes:

$$
\begin{aligned}
\wick{\c{\widetilde{\mathfrak{F}}}^s_{\kappa\lambda}(z) \quad \c{\widetilde{\mathfrak{F}}}^s_{\rho\sigma}(z)} &= \wick{\c{\widetilde{\mathfrak{F}}}^s_{\kappa\lambda}(u) \quad \c{\widetilde{\mathfrak{F}}}^s_{\rho\sigma}(v)}\Big|_{u=z,v=z} = \epsilon_{\kappa\lambda\zeta\alpha} \; \wick{\c{\mathfrak{F}}^{\zeta\alpha}_s(u) \quad \epsilon_{\rho\sigma\chi\beta} \; \c{\mathfrak{F}}^{\chi\beta}_s(v)}\Big|_{u=z,v=z} \\
&= 4 \; \epsilon_{\kappa\lambda\zeta\alpha} \;\; {}_u\partial^\zeta \;\; \epsilon_{\rho\sigma\chi\beta} \;\; {}_v\partial^\chi \;\; \overline{D}^{\alpha\beta}(u,v)\Big|_{u=z,v=z}. \qquad (4.84)
\end{aligned}
$$

The external legs contractions for the fourth contribution are the same as for the second contribution, *i.e.* Eqs. (4.73) and (4.74). The set of contractions that represent the contribution of the fourth term to the tadpole diagram take the form: For the first set of external legs contraction:

$$
\begin{aligned}
& \wick{\c{A}^\mu(x) \; \c{\mathfrak{F}}^{\kappa\lambda}_c(z) \quad \c{\widetilde{\mathfrak{F}}}^s_{\kappa\lambda}(z) \quad \c{\widetilde{\mathfrak{F}}}^s_{\rho\sigma}(z) \; \c{\mathfrak{F}}^{\rho\sigma}_c(z) \; \c{A}^\nu(y)} \\
= \;\; & \wick{\c{A}^\mu(x) \; \c{\mathfrak{F}}^{\kappa\lambda}_c(z) \quad \c{\widetilde{\mathfrak{F}}}^s_{\kappa\lambda}(u) \quad \c{\widetilde{\mathfrak{F}}}^s_{\rho\sigma}(v)\Big|_{u=z,v=z} \c{\mathfrak{F}}^{\rho\sigma}_c(z) \; \c{A}^\nu(y)} \\
= \;\; & 4\Big(\; {}_x\partial^\kappa \, D^{\mu\lambda}(x-z) \;\; \epsilon_{\kappa\lambda\zeta\alpha} \;\; {}_u\partial^\zeta \;\; {}_v\partial^\chi \;\; \overline{D}^{\alpha\beta}(u,v) \;\; \epsilon_{\rho\sigma\chi\beta} \;\; {}_y\partial^\rho \, D^{\sigma\nu}(z-y) \\
& - \;\; {}_x\partial^\kappa \, D^{\mu\lambda}(x-z) \;\; \epsilon_{\kappa\lambda\zeta\alpha} \;\; {}_u\partial^\zeta \;\; {}_v\partial^\chi \;\; \overline{D}^{\alpha\beta}(u,v) \;\; \epsilon_{\rho\sigma\chi\beta} \;\; {}_y\partial^\sigma \, D^{\rho\nu}(z-y) \\
& - \;\; {}_x\partial^\lambda \, D^{\mu\kappa}(x-z) \;\; \epsilon_{\kappa\lambda\zeta\alpha} \;\; {}_u\partial^\zeta \;\; {}_v\partial^\chi \;\; \overline{D}^{\alpha\beta}(u,v) \;\; \epsilon_{\rho\sigma\chi\beta} \;\; {}_y\partial^\rho \, D^{\sigma\nu}(z-y) \\
& + \;\; {}_x\partial^\lambda \, D^{\mu\kappa}(x-z) \;\; \epsilon_{\kappa\lambda\zeta\alpha} \;\; {}_u\partial^\zeta \;\; {}_v\partial^\chi \;\; \overline{D}^{\alpha\beta}(u,v) \;\; \epsilon_{\rho\sigma\chi\beta} \;\; {}_y\partial^\sigma \, D^{\rho\nu}(z-y)\Big)\Big|_{u=z,v=z} \\
= \;\; & 16\Big(D^{\mu\eta}(x-z) \;\; \epsilon_{\eta\alpha\kappa\zeta} \;\; {}_x\partial^\kappa \, {}_u\partial^\zeta \;\;\; \epsilon_{\tau\beta\rho\chi} \;\; {}_y\partial^\rho \;\; {}_v\partial^\chi \, D^{\sigma\nu}(z-y) \;\; \overline{D}^{\alpha\beta}(u,v)\Big)\Big|_{u=z,v=z},
\end{aligned}
$$

$$(4.85)$$

where the last line is obtained by switching indices on the epsilon tensors. Rewriting this without the propagators of the external legs as we did to obtain (Eq. 4.71), the above takes the form:

$$\mathrm{C4a} = 16 \Big(\epsilon_{\eta\alpha\kappa\zeta} \ {}_x\partial^{\kappa} \, {}_u\partial^{\zeta} \ \ \epsilon_{\tau\beta\rho\chi} \ {}_y\partial^{\rho} \ {}_v\partial^{\chi} \ \overline{D}^{\alpha\beta}(u,v) \ \Big)\Big|_{u=z,v=z}. \tag{4.86}$$

For the second set of external legs contraction:

$$\begin{aligned}
& A^{\mu}(x)\ \mathfrak{F}_c^{\kappa\lambda}(z)\ \overbrace{\widetilde{\mathfrak{F}}^{s}_{\kappa\lambda}(z)\,\widetilde{\mathfrak{F}}^{s}_{\rho\sigma}(z)}\ \mathfrak{F}_c^{\rho\sigma}(z)\ A^{\nu}(y) \\
= \ & \overbrace{A^{\mu}(x)\ \mathfrak{F}_c^{\rho\sigma}(z)}\ \ \overbrace{\widetilde{\mathfrak{F}}^{s}_{\kappa\lambda}(z)\,\widetilde{\mathfrak{F}}^{s}_{\rho\sigma}(z)}\ \overbrace{\mathfrak{F}_c^{\kappa\lambda}(z)\ A^{\nu}(y)} \\
= \ & \overbrace{A^{\mu}(x)\ \mathfrak{F}_c^{\rho\sigma}(z)}\ \ \overbrace{\widetilde{\mathfrak{F}}^{s}_{\kappa\lambda}(u)\,\widetilde{\mathfrak{F}}^{s}_{\rho\sigma}(v)}\Big|_{u=z,v=z}\ \overbrace{\mathfrak{F}_c^{\kappa\lambda}(z)\ A^{\nu}(y)} \\
= \ & 4\Big(\ {}_x\partial^{\rho}\, D^{\mu\sigma}(x-z)\ \ \epsilon_{\kappa\lambda\zeta\alpha}\ \ {}_u\partial^{\zeta}\ \ {}_v\partial^{\chi}\ \ \overline{D}^{\alpha\beta}(u,v)\ \ \epsilon_{\rho\sigma\chi\beta}\ \ {}_y\partial^{\kappa}\, D^{\lambda\nu}(z-y) \\
& -\ {}_x\partial^{\rho}\, D^{\mu\sigma}(x-z)\ \ \epsilon_{\kappa\lambda\zeta\alpha}\ \ {}_u\partial^{\zeta}\ \ {}_v\partial^{\chi}\ \ \overline{D}^{\alpha\beta}(u,v)\ \ \epsilon_{\rho\sigma\chi\beta}\ \ {}_y\partial^{\lambda}\, D^{\kappa\nu}(z-y) \\
& -\ {}_x\partial^{\sigma}\, D^{\mu\rho}(x-z)\ \ \epsilon_{\kappa\lambda\zeta\alpha}\ \ {}_u\partial^{\zeta}\ \ {}_v\partial^{\chi}\ \ \overline{D}^{\alpha\beta}(u,v)\ \ \epsilon_{\rho\sigma\chi\beta}\ \ {}_y\partial^{\kappa}\, D^{\lambda\nu}(z-y) \\
& +\ {}_x\partial^{\sigma}\, D^{\mu\rho}(x-z)\ \ \epsilon_{\kappa\lambda\zeta\alpha}\ \ {}_u\partial^{\zeta}\ \ {}_v\partial^{\chi}\ \ \overline{D}^{\alpha\beta}(u,v)\ \ \epsilon_{\rho\sigma\chi\beta}\ \ {}_y\partial^{\lambda}\, D^{\kappa\nu}(z-y)\Big)\Big|_{u=z,v=z} \\
= \ & 16\Big(D^{\mu\eta}(x-z)\ \ \epsilon_{\tau\alpha\kappa\zeta}\ \ {}_y\partial^{\kappa}\ \ {}_u\partial^{\zeta}\ \ \epsilon_{\eta\beta\rho\chi}\ \ {}_x\partial^{\rho}\ \ {}_v\partial^{\chi}\ \ \ D^{\tau\nu}(z-y)\ \overline{D}^{\alpha\beta}(u,v)\ \Big)\Big|_{u=z,v=z}.
\end{aligned} \tag{4.87}$$

This yields the following contribution to the tadpole diagram:

$$\mathrm{C4b} = 16 \Big(\ \epsilon_{\eta\beta\rho\chi} \ {}_x\partial^{\rho} \ {}_v\partial^{\chi} \ \ \epsilon_{\tau\alpha\kappa\zeta} \ {}_y\partial^{\kappa} \ {}_u\partial^{\zeta} \ \overline{D}^{\alpha\beta}(u,v) \ \Big)\Big|_{u=z,v=z}. \tag{4.88}$$

For convenience, let us summarize below, the expressions obtained for the different contributions:

$$
\begin{aligned}
\mathrm{C1a} &= \mathrm{C1b} \\
& \quad 4\Big({}_{v}\partial_{\rho}\ {}_{u}\partial^{\rho}\ {}_{y}\partial^{\kappa}\ {}_{x}\partial_{\kappa}\ g_{\eta\tau}\, g_{\alpha\beta} - {}_{v}\partial_{\rho}\ {}_{u}\partial^{\rho}\ {}_{y}\partial_{\eta}\ {}_{x}\partial_{\tau}\ g_{\alpha\beta} \\
&- \ {}_{v}\partial_{\alpha}\ {}_{u}\partial_{\beta}\ {}_{y}\partial^{\kappa}\ {}_{x}\partial_{\kappa}\ g_{\eta\tau} + {}_{v}\partial_{\alpha}\ {}_{u}\partial_{\beta}\ {}_{y}\partial_{\eta}\ {}_{x}\partial_{\tau} \Big)\ \overline{D}^{\alpha\beta}(u,v)\Big|_{u=z,v=z} \\
\mathrm{C2a} &= 4\Big(g_{\eta\alpha}\ {}_{u}\partial_{\kappa}\ {}_{x}\partial^{\kappa}\ {}_{v}\partial_{\rho}\ {}_{y}\partial^{\rho}\ g_{\beta\tau} - {}_{u}\partial_{\eta}\ {}_{x}\partial_{\alpha}\ {}_{v}\partial_{\rho}\ {}_{y}\partial^{\rho}\ g_{\beta\tau} \\
&- \ g_{\eta\alpha}\ {}_{u}\partial_{\kappa}\ {}_{x}\partial^{\kappa}\ {}_{v}\partial_{\tau}\ {}_{y}\partial_{\beta} + {}_{u}\partial_{\eta}\ {}_{x}\partial_{\alpha}\ {}_{v}\partial_{\tau}\ {}_{y}\partial_{\beta} \Big)\ \overline{D}^{\alpha\beta}(u,v)\Big|_{u=z,v=z} \\
\mathrm{C2b} &= 4\Big(g_{\eta\beta}\ {}_{v}\partial_{\rho}\ {}_{x}\partial^{\rho}\ {}_{u}\partial_{\kappa}\ {}_{y}\partial^{\kappa}\ g_{\alpha\tau} - {}_{v}\partial_{\eta}\ {}_{x}\partial_{\beta}\ {}_{u}\partial_{\kappa}\ {}_{y}\partial^{\kappa}\ g_{\alpha\tau} \\
&- \ g_{\eta\beta}\ {}_{v}\partial_{\rho}\ {}_{x}\partial^{\rho}\ {}_{u}\partial_{\tau}\ {}_{y}\partial_{\alpha} + {}_{v}\partial_{\eta}\ {}_{x}\partial_{\beta}\ {}_{u}\partial_{\tau}\ {}_{y}\partial_{\alpha} \Big)\ \overline{D}^{\alpha\beta}(u,v)\Big|_{u=z,v=z} \\
\mathrm{C4a} &= 16\Big(\epsilon_{\eta\alpha\kappa\zeta}\ {}_{x}\partial^{\kappa}\ {}_{u}\partial^{\zeta}\ \epsilon_{\tau\beta\rho\chi}\ {}_{y}\partial^{\rho}\ {}_{v}\partial^{\chi} \Big)\ \overline{D}^{\alpha\beta}(u,v)\ \Big|_{u=z,v=z} \\
\mathrm{C4b} &= 16\Big(\epsilon_{\eta\beta\rho\chi}\ {}_{x}\partial^{\rho}\ {}_{v}\partial^{\chi}\ \epsilon_{\tau\alpha\kappa\zeta}\ {}_{y}\partial^{\kappa}\ {}_{u}\partial^{\zeta} \Big)\ \overline{D}^{\alpha\beta}(u,v)\ \Big|_{u=z,v=z} \qquad (4.89)
\end{aligned}
$$

4.5.4 Resulting polarization tensor

We now add all the contributions obtained in the previous section, to get the vacuum polarization tensor. Recall from Eq. (4.53) that the terms in the Lagrangian responsible for the Scharnhorst effect are:

$$
\frac{\alpha^2}{m_e^4}\Big\{2\,c_1\,\mathcal{F}_c^{\kappa\lambda}\,\mathcal{F}^c_{\kappa\lambda}\,\mathcal{F}_s^{\rho\sigma}\,\mathcal{F}^s_{\rho\sigma} + 4\,c_1\,\mathcal{F}_c^{\kappa\lambda}\,\mathcal{F}^s_{\kappa\lambda}\,\mathcal{F}_c^{\rho\sigma}\,\mathcal{F}^s_{\rho\sigma} + 2\,c_2\,\mathcal{F}_c^{\kappa\lambda}\,\widetilde{\mathcal{F}}^c_{\kappa\lambda}\,\mathcal{F}_s^{\rho\sigma}\,\widetilde{\mathcal{F}}^s_{\rho\sigma} + 4\,c_2\,\mathcal{F}_c^{\kappa\lambda}\,\widetilde{\mathcal{F}}^s_{\kappa\lambda}\,\mathcal{F}_c^{\rho\sigma}\,\widetilde{\mathcal{F}}^s_{\rho\sigma}\Big\}.
$$

Therefore the tadpole diagram is given by:

$$
\begin{aligned}
& \text{tadpole diagram} \\
=\ & \frac{\alpha^2}{m_e^4}\big(2\,c_1\,(\mathrm{C1a} + \mathrm{C1b}) + 4\,c_1\,(\mathrm{C2a} + \mathrm{C2b}) + 2\,c_2\,(\mathrm{C3a} + \mathrm{C3b}) + 4\,c_2\,(\mathrm{C4a} + \mathrm{C4b})\big) \\
=\ & \frac{\alpha^2}{m_e^4}\big(2\,c_1\,(\mathrm{C1a} + \mathrm{C1b}) + 4\,c_1\,(\mathrm{C2a} + \mathrm{C2b}) + 4\,c_2\,(\mathrm{C4a} + \mathrm{C4b})\big) \qquad (4.90)
\end{aligned}
$$

where c_1 and c_2 are the coefficient in the Euler-Heisenberg Lagrangian. Recall:

$$c_1 = \frac{1}{(2)(3^2)(5)} = \frac{1}{(2)(45)} \tag{4.91a}$$

$$c_2 = \frac{7}{(2^3)(3^2)(5)} = \frac{7}{(8)(45)}. \tag{4.91b}$$

For ease of comparison with Scharnhorst's original result, who uses e instead of α where $\alpha = \frac{e^2}{4\pi}$, we shall express our results in terms of $\frac{\alpha^2}{m_e^4} = \frac{1}{(2^4)(\pi^2)} \frac{e^4}{m_e^4}$. Then:

$$\frac{\alpha^2}{m_e^4} c_1 = \frac{1}{2^5} \frac{1}{(45)(\pi^2)} \frac{e^4}{m_e^4} \tag{4.92a}$$

$$\frac{\alpha^2}{m_e^4} c_2 = \frac{7}{2^7} \frac{1}{(45)(\pi^2)} \frac{e^4}{m_e^4}. \tag{4.92b}$$

Note that Eq. (4.90) agrees with the expression obtained by Scharnhorst for the self-energy tensor.[26]

---

[26] Indeed Eq. (4.90) yields for our tadpole diagram the following expression:

$$
\begin{aligned}
& \frac{\alpha^2}{m_e^4} \quad \Big(2\, c_1\, (\mathrm{C1a} \; + \; \mathrm{C1b}) + 4\, c_1\, (\mathrm{C2a} \; + \; \mathrm{C2b}) + 4\, c_2\, (\mathrm{C4a} \; + \; \mathrm{C4b}) \Big) \\
= \; & \frac{1}{2^5} \frac{1}{(45)(\pi^2)} \frac{e^4}{m_e^4} \quad \Big(2\; (\mathrm{C1a} \; + \; \mathrm{C1b}) + 4\; (\mathrm{C2a} \; + \; \mathrm{C2b}) + 4\, \frac{7}{4}\, (\mathrm{C4a} \; + \; \mathrm{C4b}) \Big) \\
= \; & \frac{1}{2^5} \frac{1}{(45)(\pi^2)} \frac{e^4}{m_e^4} \quad \Big(2\; (\mathrm{C1a} \; + \; \mathrm{C1b}) + 4\; (\mathrm{C2a} \; + \; \mathrm{C2b}) + 7\, (\mathrm{C4a} \; + \; \mathrm{C4b}) \Big) \\
= \; & \frac{1}{2^4} \frac{1}{(45)(\pi^2)} \frac{e^4}{m_e^4} \; \cdot \; \Big(8\, \Big({}_v\partial_\rho \; {}_u\partial^\rho \; {}_y\partial^\kappa \; {}_x\partial_\kappa \; g_{\eta\tau}\, g_{\alpha\beta} - {}_v\partial_\rho \; {}_u\partial^\rho \; {}_y\partial_\eta \; {}_x\partial_\tau \; g_{\alpha\beta} \\
& \qquad - {}_v\partial_\alpha \; {}_u\partial_\beta \; {}_y\partial^\kappa \; {}_x\partial_\kappa \; g_{\eta\tau} + {}_v\partial_\alpha \; {}_u\partial_\beta \; {}_y\partial_\eta \; {}_x\partial_\tau \\
& \qquad + g_{\eta\alpha} \; {}_u\partial_\kappa \; {}_x\partial^\kappa \; {}_v\partial_\rho \; {}_y\partial^\rho \; g_{\beta\tau} - {}_u\partial_\eta \; {}_x\partial_\alpha \; {}_v\partial_\rho \; {}_y\partial^\rho \; g_{\beta\tau} \\
& \qquad - g_{\eta\alpha} \; {}_u\partial_\kappa \; {}_x\partial^\kappa \; {}_v\partial_\tau \; {}_y\partial_\beta + {}_u\partial_\eta \; {}_x\partial_\alpha \; {}_v\partial_\tau \; {}_y\partial_\beta \\
& \qquad + g_{\eta\beta} \; {}_v\partial_\rho \; {}_x\partial^\rho \; {}_u\partial_\kappa \; {}_y\partial^\kappa \; g_{\alpha\tau} - {}_v\partial_\eta \; {}_x\partial_\beta \; {}_u\partial_\kappa \; {}_y\partial^\kappa \; g_{\alpha\tau} \\
& \qquad - g_{\eta\beta} \; {}_v\partial_\rho \; {}_x\partial^\rho \; {}_u\partial_\tau \; {}_y\partial_\alpha + {}_v\partial_\eta \; {}_x\partial_\beta \; {}_u\partial_\tau \; {}_y\partial_\alpha \Big) \\
+ \; & 14\, \Big(\epsilon_{\eta\alpha\kappa\zeta}\, {}_x\partial^\kappa \, {}_u\partial^\zeta \, \epsilon_{\tau\beta\rho\chi} \; {}_y\partial^\rho \; {}_v\partial^\chi + \epsilon_{\eta\beta\rho\chi}\, {}_x\partial^\rho \; {}_v\partial^\chi \; \epsilon_{\tau\alpha\kappa\zeta} \; {}_y\partial^\kappa \, {}_u\partial^\zeta \; \Big) \Big) \; \overline{D}^{\alpha\beta}(u,v) \Big|_{u=z,v=z}
\end{aligned}
$$

This is half the expression obtained by Scharnhorst for $\int d^4u \; d^4v \; \Gamma^f_{\mu\nu\alpha\beta}(x,y,u,v) \; \overline{D}^{\alpha\beta}(u,v)$ in his Eqs. (12-14, 16). [5], pp.356-357. As Scharnhorst notes, his expression is i times $\Pi_{\mu\nu}$. As discussed above, our tadpole diagram is equal to i times $\Pi_{\mu\nu}$. Therefore the expression we just obtained agrees with Scharnhorst's result.

Furthermore, it turns out that not only are C1a and C1b equal to one another, but C2a = C2b and C4a = C4b as well. We can therefore write:

$$C1 \equiv \text{C1a} = \text{C1b} \qquad ; \qquad C2 \equiv \text{C2a} = \text{C2b} \qquad ; \qquad C4 \equiv \text{C4a} = \text{C4b}. \tag{4.93}$$

As a result Eq. (4.90) becomes:

$$\begin{aligned}
\text{tadpole diagram} \;&=\; \frac{\alpha^2}{m_e^4}\;\Big(\,2\,c_1\;(\text{C1a}\;+\;\text{C1b})\;+\;4\;c_1\;(\text{C2a}\;+\;\text{C2b})\;+\;4\;c_2\;(\text{C4a}\;+\;\text{C4b})\Big)\\
&=\; \frac{\alpha^2}{m_e^4}\;\Big(\,4\;c_1\;\;\text{C1}\;\;+\;\;8\;c_1\;\;\text{C2}\;\;+\;\;8\;c_2\;\;\text{C4}\,\Big)\\
&=\; \frac{1}{(2^4)(\pi^2)}\frac{e^4}{m_e^4}\;\Big(\,4\;c_1\;\;\text{C1}\;\;+\;\;8\;c_1\;\;\text{C2}\;\;+\;\;8\;c_2\;\;\text{C4}\,\Big).
\end{aligned} \tag{4.94}$$

The contributions to the polarization tensor are simply equal to corresponding contributions to the tadpole diagram multiplied by $-i$.
In this way we find, for the contribution associated with C1:

$$\Pi_{\eta\tau(1)} \;=\; \tfrac{6}{2^6}\tfrac{1}{\pi^4}\tfrac{e^4}{m_e^4\,d^4}\times$$

$$\begin{pmatrix}
\begin{array}{c} g(z_3) \\ ({}_x\partial_{1\,y}\partial_1 +_x \partial_{2\,y}\partial_2 +_x \partial_{3\,y}\partial_3) \end{array} & g(z_3)\,{}_x\partial_{1\,y}\partial_0 & g(z_3)\,{}_x\partial_{2\,y}\partial_0 & g(z_3)\,{}_x\partial_{3\,y}\partial_0 \\[2ex]
g(z_3)\,{}_x\partial_{0\,y}\partial_1 & \begin{array}{c} g(z_3) \\ ({}_x\partial_{0\,y}\partial_0 - x\partial_{2\,y}\partial_2 - x\partial_{3\,y}\partial_3) \end{array} & g(z_3)\,{}_x\partial_{2\;y}\partial_1 & g(z_3)\,{}_x\partial_{3\,y}\partial_1 \\[2ex]
g(z_3)\,{}_x\partial_{0\,y}\partial_2 & g(z_3)\,{}_x\partial_{1\,y}\partial_2 & \begin{array}{c} g(z_3) \\ ({}_x\partial_{0\,y}\partial_0 -_x \partial_{1\,y}\partial_1 -_x \partial_{3\,y}\partial_3) \end{array} & g(z_3)\,{}_x\partial_{3\,y}\partial_2 \\[2ex]
g(z_3)\,{}_x\partial_{0\,y}\partial_3 & g(z_3)\,{}_x\partial_{1\,y}\partial_3 & g(z_3)\,{}_x\partial_{2\,y}\partial_3 & \begin{array}{c} g(z_3) \\ ({}_x\partial_{0\,y}\partial_0 -_x \partial_{1\,y}\partial_1 -_x \partial_{2\,y}\partial_2) \end{array}
\end{pmatrix} \tag{4.95}$$

where in analogy to Scharnhorst's notation, we have called:

$$g(z_3) \equiv \zeta\left(4, \frac{|a_1 - z_3|}{d}\right) + \zeta\left(4, \frac{|a_2 - z_3|}{d}\right), \tag{4.96}$$

where ζ represents the Riemann Zeta function.

In the same way the other contributions are, respectively:

$$\Pi_{\eta\tau\,(2)} = \frac{1}{2^6}\frac{1}{\pi^4}\frac{e^4}{m_e^4\,d^4}\times$$

$$\begin{pmatrix}
\begin{array}{c} g(z_3) \\ ({}_x\partial_{1\,y}\partial_1 + {}_x\partial_{2\,y}\partial_2 + {}_x\partial_{3\,y}\partial_3) \\ -\frac{\pi^4}{45} \\ ({}_x\partial_{1\,y}\partial_1 + {}_x\partial_{2\,y}\partial_2 - {}_x\partial_{3\,y}\partial_3) \end{array} & \left(g(z_3) - \frac{\pi^4}{45}\right) {}_x\partial_{1\,y}\partial_0 & \left(g(z_3) - \frac{\pi^4}{45}\right) {}_x\partial_{2\,y}\partial_0 & \left(g(z_3) + \frac{\pi^4}{45}\right) {}_x\partial_{3\,y}\partial_0 \\
\left(g(z_3) - \frac{\pi^4}{45}\right) {}_x\partial_{0\,y}\partial_1 & \begin{array}{c} g(z_3) \\ (x\partial_{0\,y}\partial_0 - x\partial_{2\,y}\partial_2 -_x \partial_{3\,y}\partial_3) \\ -\frac{\pi^4}{45} \\ ({}_x\partial_{0\,y}\partial_0 - x\partial_{2\,y}\partial_2 +_x \partial_{3\,y}\partial_3) \end{array} & \left(g(z_3) - \frac{\pi^4}{45}\right) {}_x\partial_{2\,y}\partial_1 & \left(g(z_3) + \frac{\pi^4}{45}\right) {}_x\partial_{3\,y}\partial_1 \\
\left(g(z_3) - \frac{\pi^4}{45}\right) {}_x\partial_{0\,y}\partial_2 & \left(g(z_3) - \frac{\pi^4}{45}\right) {}_x\partial_{1\,y}\partial_2 & \begin{array}{c} g(z_3) \\ ({}_x\partial_{0\,y}\partial_0 - {}_x\partial_{1\,y}\partial_1 - {}_x\partial_{3\,y}\partial_3) \\ -\frac{\pi^4}{45} \\ ({}_x\partial_{0\,y}\partial_0 - {}_x\partial_{1\,y}\partial_1 + {}_x\partial_{3\,y}\partial_3) \end{array} & \left(g(z_3) + \frac{\pi^4}{45}\right) {}_x\partial_{3\,y}\partial_2 \\
\left(g(z_3) - \frac{\pi^4}{45}\right) {}_x\partial_{0\,y}\partial_3 & \left(g(z_3) - \frac{\pi^4}{45}\right) {}_x\partial_{1\,y}\partial_3 & \left(g(z_3) + \frac{\pi^4}{45}\right) {}_x\partial_{2\,y}\partial_3 & \begin{array}{c} \left(g(z_3) + \frac{\pi^4}{45}\right) \\ ({}_x\partial_{0\,y}\partial_0 -_x \partial_{1\,y}\partial_1 -_x \partial_{2\,y}\partial_2) \end{array}
\end{pmatrix} \tag{4.97}$$

where we have used the fact that between the plates:

$$\zeta\left(4, \frac{1}{2} + \frac{|a_1 - x_3| + |a_2 - x_3|}{2\,d}\right) = \zeta\,(4,1) = \frac{\pi^4}{90}, \tag{4.98}$$

where d is the plates separation.[27]

As stated above the third contribution is a null tensor: $\Pi_{\eta\tau\,(3)} = 0$.

---

[27] Indeed between the plates:
If $a_1 < x_3 < a_2$, $|a_1 - x_3| + |a_2 - x_3| = x_3 - a_1 + a_2 - x_3 = a_2 - a_1 = d$,
If $a_1 > x_3 > a_2$, $|a_1 - x_3| + |a_2 - x_3| = a_1 - x_3 + x_3 - a_2 = a_1 - a_2 = d$.

Finally, the fourth contribution to the self-energy tensor is:

$$\Pi_{\eta\tau\,(4)} = \frac{1}{2^6}\,\frac{1}{\pi^4}\,\frac{e^4}{m_e^4\,d^4}\times$$

$$\begin{pmatrix}
\begin{array}{c} -g(z_3) \\ \left({}_x\partial_{1\,y}\partial_1 +_x \partial_{2\,y}\partial_2 + {}_x\partial_{3\,y}\partial_3\right) \\ -\frac{\pi^4}{45} \\ \left({}_x\partial_{1\,y}\partial_1 +_x \partial_{2\,y}\partial_2 - {}_x\partial_{3\,y}\partial_3\right) \end{array} & \left(-g(z_3) - \frac{\pi^4}{45}\right)\,{}_x\partial_{1\,y}\partial_0 & \left(-g(z_3) - \frac{\pi^4}{45}\right)\,{}_x\partial_{2\,y}\partial_0 & \left(-g(z_3) + \frac{\pi^4}{45}\right)\,{}_x\partial_{3\,y}\partial_0 \\
\left(-g(z_3) - \frac{\pi^4}{45}\right)\,{}_x\partial_{0\,y}\partial_1 & \begin{array}{c} -g(z_3) \\ \left({}_x\partial_{0\,y}\partial_0 - {}_x\partial_{2\,y}\partial_2 - {}_x\partial_{3\,y}\partial_3\right) \\ -\frac{\pi^4}{45} \\ \left({}_x\partial_{0\,y}\partial_0 - {}_x\partial_{2\,y}\partial_2 + {}_x\partial_{3\,y}\partial_3\right) \end{array} & \left(-g(z_3) - \frac{\pi^4}{45}\right)\,{}_x\partial_{2\,y}\partial_1 & \left(-g(z_3) + \frac{\pi^4}{45}\right)\,{}_x\partial_{3\,y}\partial_1 \\
\left(-g(z_3) - \frac{\pi^4}{45}\right)\,{}_x\partial_{0\,y}\partial_2 & \left(-g(z_3) - \frac{\pi^4}{45}\right)\,{}_x\partial_{1\,y}\partial_2 & \begin{array}{c} -g(z_3) \\ \left({}_x\partial_{0\,y}\partial_0 - {}_x\partial_{1\,y}\partial_1 - {}_x\partial_{3\,y}\partial_3\right) \\ -\frac{\pi^4}{45} \\ \left({}_x\partial_{0\,y}\partial_0 - {}_x\partial_{1\,y}\partial_1 + {}_x\partial_{3\,y}\partial_3\right) \end{array} & \left(-g(z_3) + \frac{\pi^4}{45}\right)\,{}_x\partial_{3\,y}\partial_2 \\
\left(-g(z_3) + \frac{\pi^4}{45}\right)\,{}_x\partial_{0\,y}\partial_3 & \left(-g(z_3) + \frac{\pi^4}{45}\right)\,{}_x\partial_{1\,y}\partial_3 & \left(-g(z_3) + \frac{\pi^4}{45}\right)\,{}_x\partial_{2\,y}\partial_3 & \begin{array}{c} \left(-g(z_3) + \frac{\pi^4}{45}\right) \\ \left({}_x\partial_{0\,y}\partial_0 - {}_x\partial_{1\,y}\partial_1 - {}_x\partial_{2\,y}\partial_2\right) \end{array}
\end{pmatrix} \tag{4.99}$$

The required polarization tensor of the propagating Scharnhorst photon, $\Pi_{\eta\tau}^{(2-loop)}$, is given by the sum of these contributions, weighted by their respective coefficients as indicated above in Eq. (4.90):

$$\Pi_{\eta\tau}^{(2-loop)} = 4\;c_1\,\Pi_{\eta\tau\,(1)} + 8\;c_1\,\Pi_{\eta\tau\,(2)} + 8\;c_2\,\Pi_{\eta\tau\,(4)}. \tag{4.100}$$

In this way we obtain the polarization tensor needed to calculate the phase velocity of the Scharnhorst photons (see next page). In the next two sections, I present two ways to obtain this velocity: the first makes use of the light cone condition, thereby avoiding semi-classical concepts such as permittivity and permeability. For the sake of completeness, I shall also give the calculation based on these concepts and the index of refraction of the Casimir vacuum in different directions.

$$\Pi^{(2-loop)}_{\eta\tau} = \frac{1}{2^6}\,\frac{1}{(45)(\pi^4)}\,\frac{e^4}{m_e^4\,d^4}\times$$

$$\begin{pmatrix}
\begin{array}{c} 9\,g(z_3) \\ \left({}_x\partial_1{}_y\partial_1 +_x \partial_2{}_y\partial_2 +_x \partial_3{}_y\partial_3\right) \\ -\frac{11\,\pi^4}{45} \\ \left({}_x\partial_1{}_y\partial_1 +_x \partial_2{}_y\partial_2 -_x \partial_3{}_y\partial_3\right) \end{array} & \left(9\,g(z_3) - \frac{11\,\pi^4}{45}\right)\ {}_x\partial_1{}_y\partial_0 & \left(9\,g(z_3) - \frac{11\,\pi^4}{45}\right)\ {}_x\partial_2\ {}_y\partial_0 & \left(9\,g(z_3) + \frac{11\,\pi^4}{45}\right)\ {}_x\partial_3\ {}_y\partial_0 \\
\left(9\,g(z_3) - \frac{11\,\pi^4}{45}\right)\ {}_x\partial_0\ {}_y\partial_1 & \begin{array}{c} 9\,g(z_3) \\ \left({}_x\partial_0{}_y\partial_0 -_x \partial_2{}_y\partial_2 -_x \partial_3{}_y\partial_3\right) \\ +\frac{11\,\pi^4}{45} \\ \left(-_x\partial_0{}_y\partial_0 +_x \partial_2{}_y\partial_2 -_x \partial_3{}_y\partial_3\right) \end{array} & \left(9\,g(z_3) - \frac{11\,\pi^4}{45}\right)\ {}_x\partial_2\ {}_y\partial_1 & \left(9\,g(z_3) + \frac{11\,\pi^4}{45}\right)\ {}_x\partial_3\ {}_y\partial_1 \\
\left(9\,g(z_3) - \frac{11\,\pi^4}{45}\right){}_x\partial_0\ {}_y\partial_2 & \left(9\,g(z_3) - \frac{11\,\pi^4}{45}\right)\ {}_x\partial_1{}_y\partial_2 & \begin{array}{c} 9\,g(z_3) \\ \left({}_x\partial_0{}_y\partial_0 -_x \partial_1{}_y\partial_1 -_x \partial_3{}_y\partial_3\right) \\ +\frac{11\,\pi^4}{45} \\ \left(-_x\partial_0{}_y\partial_0 +_x \partial_1{}_y\partial_1 -_x \partial_3{}_y\partial_3\right) \end{array} & \left(9\,g(z_3) + \frac{11\,\pi^4}{45}\right)\ {}_x\partial_3\ {}_y\partial_2 \\
\left(9\,g(z_3) + \frac{11\,\pi^4}{45}\right)\ {}_x\partial_0\ {}_y\partial_3 & \left(9\,g(z_3) + \frac{11\,\pi^4}{45}\right)\ {}_x\partial_1{}_y\partial_3 & \left(9\,g(z_3) + \frac{11\,\pi^4}{45}\right)\ {}_x\partial_2\ {}_y\partial_3 & \begin{array}{c} \left(9\,g(z_3) + \frac{11\,\pi^4}{45}\right) \\ \left({}_x\partial_0{}_y\partial_0 -_x \partial_1{}_y\partial_1 -_x \partial_2{}_y\partial_2\right) \end{array}
\end{pmatrix} \tag{4.101}$$

4.6 Derivation of the velocity from the SCET polarization tensor using the light cone condition

Given the polarization tensor, the phase velocity of the Scharnhorst photon can be found from the light cone condition:

$$\left(p^2 g^{\mu\nu} - p^{\mu} p^{\nu} + \Pi^{\mu\nu}\right) a_{\nu} = 0, \tag{4.102}$$

where a_{ν} is the polarization vector of the propagating photon.

To this end, we need the expression for the polarization tensor in momentum space:
$\Pi^{(2-loop)}_{\mu\nu}(p)= \frac{1}{2^6}\frac{1}{(45)(\pi^4)}\frac{e^4}{m_e^4\, d^4}\times$

$$\begin{pmatrix} \begin{matrix} -9\ \tilde{g}(p_3) \\ (p_1^2+p_2^2+p_3^2) \\ +\frac{11}{45}\pi^4 \\ (p_1^2+p_2^2-p_3^2) \end{matrix} & (-9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_0\, p_1 & (-9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_0\, p_2 & -(9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_0\, p_3 \\ (-9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_0\, p_1 & \begin{matrix} 9\ \tilde{g}(p_3) \\ (-p_0^2+p_2^2+p_3^2) \\ +\frac{11}{45}\pi^4 \\ (+p_0^2-p_2^2+p_3^2) \end{matrix} & (-9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_1\, p_2 & -(9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_1\, p_3 \\ (-9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_0\, p_2 & (-9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_1\, p_2 & \begin{matrix} 9\ \tilde{g}(p_3) \\ (-p_0^2+p_1^2+p_3^2) \\ +\frac{11}{45}\pi^4 \\ (+p_0^2-p_1^2+p_3^2) \end{matrix} & -(9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_2\, p_3 \\ -(9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_0\, p_3 & -(9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_1\, p_3 & -(9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4)\ p_2\, p_3 & \begin{matrix} (9\ \tilde{g}(p_3)\ +\frac{11}{45}\pi^4) \\ (-p_0^2+p_1^2+p_2^2) \end{matrix} \end{pmatrix} \tag{4.103}$$

Now in the case of trivial vacuum, the polarization tensor can be expressed in the following form:

$$\Pi_{\mu\nu}(p) = (p^2\, g_{\mu\nu} - p_{\mu} p_{\nu})\ \Pi(p^2) \tag{4.104}$$

where $\Pi(p^2)$ is a scalar. In a non-trivial vacuum as we have it, this relation is modified by an extra term (which we shall call $\Pi^{+}_{\mu\nu}$), which is the one that modifies the propagation of the photon and causes its speed to be different from c:

$$\Pi_{\mu\nu}(p) = (p^2\, g_{\mu\nu} - p_\mu p_\nu)\ \Pi(p^2) + \Pi^{+}_{\mu\nu} \tag{4.105}$$

In the present case, we can write:

$$\Pi_{\mu\nu}(p) = (p^2\, g_{\mu\nu} - p_\mu p_\nu)\ \left(\frac{\xi}{2}\ \frac{45}{11\,\pi^4}\right)\left(9\ \tilde{g}(p_3)\ - \frac{11}{45}\pi^4\right) + \Pi^{+}_{\mu\nu} \tag{4.106}$$

where:

$(p^2\, g_{\mu\nu} - p_\mu p_\nu) =$

$$\begin{pmatrix} -p_1^2 - p_2^2 - p_3^2 & -p_0\, p_1 & -p_0\, p_2 & -p_0\, p_3 \\ -p_0\, p_1 & -p_0^2 + p_2^2 + p_3^2 & -p_1\, p_2 & -p_1\, p_3 \\ -p_0\, p_2 & -p_1\, p_2 & -p_0^2 + p_1^2 + p_3^2 & -p_2\, p_3 \\ -p_0\, p_3 & -p_1\, p_3 & -p_2\, p_3 & -p_0^2 + p_1^2 + p_2^2 \end{pmatrix} \tag{4.107}$$

$$\Pi^{(+)}_{\mu\nu} = \begin{pmatrix} -p_3^2\ \xi & 0 & 0 & -p_0\, p_3\ \xi \\ 0 & p_3^2\ \xi & 0 & -p_1\, p_3\ \xi \\ 0 & 0 & p_3^2\ \xi & -p_2\, p_3\ \xi \\ -p_0\, p_3\ \xi & -p_1\, p_3\ \xi & -p_2\, p_3\ \xi & (-p_0^2 + p_1^2 + p_2^2)\ \xi \end{pmatrix} \tag{4.108}$$

and:

$$\xi \equiv \frac{11}{(45)^2\ (2^5)}\ \frac{e^4}{m_e^4\ d^4}. \tag{4.109}$$

We then substitute Eq. (4.108) in the light cone condition Eq. (4.102). Because of the symmetry between the $x-$ and $y-$ direction, there is no loss of generality if we consider a photon propagating in the $xz-$plane.[28] With this choice:

$$\vec{p} = (|\vec{p}|\ sin\,\theta, 0, |\vec{p}|\ cos\,\theta)\,, \tag{4.110}$$

i.e. $p_1 = |\vec{p}|\ sin\,\theta$, $p_2 = 0$, $p_3 = |\vec{p}|\ cos\,\theta$, where θ is the angle between $\vec{p}$ and the $z-$direction normal to the plates. For simplicity, let us tape $a_n = (0, 0, 1, 0)$ as a compatible choice of polarization. The light cone condition then reads:

$$\big(0, 0, -p_0^2 + |\vec{p}|^2 cos^2\,\theta + |\vec{p}|^2 sin^2\,\theta + \xi\ |\vec{p}|^2 cos^2\,\theta, 0\big) = \vec{0}. \tag{4.111}$$

The only non-trivial equation that results is:

$$-\,p_0^2 + |\vec{p}|^2 + \xi\ |\vec{p}|^2 cos^2\,\theta = 0. \tag{4.112}$$

This implies that the phase velocity v_ϕ of the probe photon is given by:

$$v_\phi = \frac{p_0}{|\vec{p}|} = (1 + \xi\ cos^2\,\theta)^{\frac{1}{2}}. \tag{4.113}$$

In the specific case where the photon propagates in the direction normal to the plates, *i.e.* when $\theta = 0$ (and in natural units where $c = 1$):

$$v_{\phi\perp} = (1+\xi)^{\frac{1}{2}} \quad \sim \quad 1 + \frac{1}{2}\ \xi \quad \sim \quad 1 + \frac{11}{(45)^2\ (2^6)}\ \frac{e^4}{m_e^4\ d^4}, \tag{4.114}$$

which is equal to the result originally calculated by Klaus Scharnhorst [5].

By contrast, when the photon propagates in the direction parallel to the plates (*i.e.*

---

[28] Here I follow the same reasoning as Visser *et al* [14, 13] p.5

x in this case), $\theta = \frac{\pi}{2}$ and:

$$v_{\phi\ ||} = (1 + (\xi\,.\,0))^{\frac{1}{2}} \quad = \quad 1, \tag{4.115}$$

i.e. the phase velocity of the photon is the same as in trivial vacuum.

4.7 Derivation of the velocity from the SCET polarization tensor using the index of refraction

One can also derive the phase velocity of the Scharnhorst photon from our result for the polarization tensor by first calculating expressions for quantities that refer to the Casimir background: permittivity and permeability tensors, and from these index of refraction. This is similar to the way Scharnhorst originally derived the result [5].

In order to do so we calculate the shift in the action:

$$\Gamma' = -\frac{1}{2}\int d^4x\, d^4y \;\; A^{\mu}(x)\; \Pi^{(2-loop)}_{\mu\nu}\; A^{\nu}(y). \tag{4.116}$$

With $\Pi^{(2-loop)}_{\mu\nu}$ as obtained above at the end of section 4.5, *i.e.* in coordinate space, this yields:

$$\begin{aligned}\Gamma' \;&=\; \frac{1}{(2^7)(45)}\frac{1}{\pi^4}\frac{e^4}{m_e^4\, d^4}\Bigg(\; \left((E^1)^2 + (E^2)^2\right)\left(9\,g(z_3) - \frac{11\,\pi^4}{45}\right) + (E^3)^2\left(9\,g(z_3) + \frac{11\,\pi^4}{45}\right) \\ &-\; \left((B^1)^2 + (B^2)^2\right)\left(9\,g(z_3) + \frac{11\,\pi^4}{45}\right) - (B^3)^2\left(9\,g(z_3) - \frac{11\,\pi^4}{45}\right)\;\Bigg).\end{aligned} \tag{4.117}$$

This quantity can be expressed in terms of the permittivity and the permeability tensors:

$$\Gamma' = \frac{1}{2}\int d^4x\; \left(\epsilon_{ij}E^iE^j - \mu^{-1}_{ij}B^iB^j\right). \tag{4.118}$$

Comparing Eq.(4.117) and Eq.(4.118) we find that the expressions for the shifts in permittivity and permeability tensors compared to their trivial vacuum values are:

$$\Delta(\epsilon_{11}) = \Delta(\epsilon_{22}) = \frac{1}{(2^6)(45)}\frac{1}{\pi^4}\frac{e^4}{m_e^4\,d^4}\left(9\,g(z_3) - \frac{11\,\pi^4}{45}\right) \tag{4.119}$$

$$\Delta(\epsilon_{33}) = \frac{1}{(2^6)(45)}\frac{1}{\pi^4}\frac{e^4}{m_e^4\,d^4}\left(9\,g(z_3) + \frac{11\,\pi^4}{45}\right) \tag{4.120}$$

$$\Delta(\mu_{11}) = \Delta(\mu_{22}) = \frac{1}{(2^6)(45)}\frac{1}{\pi^4}\frac{e^4}{m_e^4\,d^4}\left(-9\,g(z_3) - \frac{11\,\pi^4}{45}\right) \tag{4.121}$$

$$\Delta(\mu_{33}) = \frac{1}{(2^6)(45)}\frac{1}{\pi^4}\frac{e^4}{m_e^4\,d^4}\left(-9\,g(z_3) + \frac{11\,\pi^4}{45}\right). \tag{4.122}$$

If we consider a Scharnhorst photon propagating in the direction normal to the plates (*i.e.* in the z-direction), and choose its polarization to be in the x-direction for instance, the relevant components of the permittivity is ϵ_{11}: indeed, it is this component of ϵ_{ij} that affects an electric field directed along the x-direction, which is precisely the case of our photon given the polarization we have chosen for it. Its magnetic field is along the 2-direction, therefore the component of ϵ_{ij} that affects it is μ_{22}. Inversely, if we chose the photon to be polarized in the y-direction, its motion would be affected by ϵ_{22} and μ_{11}. In any event, the set-up has cylindrical symmetric about the z-direction, so it is irrelevant what direction one considers in the x-y -plane – and indeed we see from Eq. (4.119) that $\epsilon_{11} = \epsilon_{22} = \epsilon_{||}$ and $\mu_{11} = \mu_{22} = \mu_{||}$. Therefore the index of refraction for a photon propagating in the z-direction, *i.e.*

normal to the Casimir plates, is given by:

$$\begin{aligned} n_\perp &= \sqrt{\epsilon_{||}\mu_{||}} = \big((1+\Delta(\epsilon_{||}))(1+\Delta(\mu_{||}))\big)^{\frac{1}{2}} \\ &= \big(1+\Delta(\epsilon_{||})+\Delta(\mu_{||})+\Delta(\epsilon_{||})\Delta(\mu_{||})\big)^{\frac{1}{2}} \\ &\approx \big(1+\Delta(\epsilon_{||})+\Delta(\mu_{||})\big)^{\frac{1}{2}} \\ &\approx 1+\frac{1}{2}\big(\Delta(\epsilon_{||})+\Delta(\mu_{||})\big)\,. \end{aligned} \tag{4.123}$$

Substituting in the expressions obtained above in Eq. (4.119) we obtain:

$$\begin{aligned} n_\perp &\approx 1-\frac{1}{(2^6)(45)}\frac{1}{\pi^4}\frac{e^4}{m_e^4\,d^4}\quad\left(\frac{11\,\pi^4}{45}\right) \\ &\approx 1-\frac{11}{(2^6)(45)^2}\frac{e^4}{m_e^4\,d^4}. \end{aligned} \tag{4.124}$$

This agrees with the result originally derived by Scharnhorst.

Meanwhile, photons propagating parallel to the plates experience an index of refraction of 1 as in trivial vacuum. Indeed, for a photon moving in the $x-y$ plane, polarized in the same plane the relevant component of ϵ_{ij} is $\epsilon_{||}$ and of μ_{ij} is μ_{33}.[29] Then:

$$\begin{aligned} n_{1||} &= \sqrt{\epsilon_{||}\,\mu_{33}} \approx 1+\frac{1}{2}\big(\Delta(\epsilon_{||})+\Delta(\mu_{33})\big) \\ &\approx 1-\frac{1}{(2^6)(45)}\frac{1}{\pi^4}\frac{e^4}{m_e^4\,d^4}\quad\left(9\,g(z_3)-\frac{11\,\pi^4}{45}-9\,g(z_3)+\frac{11\,\pi^4}{45}\right) \\ &\approx 1. \end{aligned} \tag{4.125}$$

---

[29] since its magnetic field must be normal to both its polarization and direction of propagation, it must be along the z-direction.

In the same way, for a photon propagating in the $x-y$ plane, and this time polarized in the z-direction the relevant component of ϵ_{ij} is ϵ_{33} and of μ_{ij} is $\mu_{||}$ so that:

$$\begin{aligned} n_{2||} &= \sqrt{\epsilon_{33}\,\mu_{||}} \approx 1+\frac{1}{2}\left(\Delta(\epsilon_{33})+\Delta(\mu_{||})\right) \\ &\approx 1-\frac{1}{(2^6)(45)}\frac{1}{\pi^4}\frac{e^4}{m_e^4\,d^4}\left(9\,g(z_3)+\frac{11\,\pi^4}{45}-9\,g(z_3)-\frac{11\,\pi^4}{45}\right) \\ &\approx 1. \end{aligned} \tag{4.126}$$

Consequently, the phase velocity for photons propagating parallel to the Casimir plates is c as in trivial vacuum. In contrast, normal to the plates it is given by:

$$v_\perp = \frac{c}{n_\perp} = \frac{c}{(1+\Delta n_\perp)} \approx c\,(1-\Delta n_\perp) = c\left(1+\frac{11}{(2^6)(45)^2}\frac{e^4}{m_e^4\,d^4}\right). \tag{4.127}$$

Again, this result agrees with what has been previously calculated for the phase velocity of Scharnhorst photons propagating normal to the Casimir plates.

4.8 Conclusion

We see that the SCET framework recovers the same result for the phase velocity as has been calculated by other methods. However as was explained in this chapter SCET allows us to draw solid conclusions regarding the size of the theory errors: one can feel confident that they would be higher order in all expansion parameters, *i.e.* in α, $\frac{E}{m_e}$ and λ. The crucial issue of what this may imply for the signal velocity shall be discussed in the Conclusion (chapter 7).

Chapter 5

FLUCTUATIONS OF THE ELECTROMAGNETIC VACUUM FIELD OR RADIATION REACTION?

5.1 Introduction

5.1.1 The issues

The Scharnhorst effect involves the same set-up as the Casimir force.[1] One can therefore expect the two to involve the same physical processes. It has become standard practice to attribute the Casimir force to the "zero-point energy" attributed to "vacuum fluctuations", also described as "virtual particles." In fact, the experimentally verified existence of the Casimir force has often been hailed as the most direct evidence for vacuum fluctuations.

Indeed, in past decades whether zero-point fluctuations of the electromagnetic field are physically real has been a topic of debate among physicists. Although opinion is now overwhelmingly in their favor, the strength of the evidence on which this opinion relies is still occasionally challenged,[2] and whether some of the arguments against it still have import is not clear a priori.

Evidence for the existence of vacuum fluctuations has naturally been sought in observable physical phenomena thought to involve them. Aside from the Scharnhorst effect which is our topic of interest, and the Casimir force, other physical effects that are typically accounted for by appealing to vacuum fluctuations are the Lamb

---

[1] The weak attractive force that acts between two conducting but electrically neutral plates (see the Introduction). The Scharnhorst effect is predicted to occur between such plates.

[2] See notably [76].

shift, spontaneous emission, van der Waals forces, and the Casimir-Polder force — the latter two being related to the Casimir effect. Yet it has been shown that most of these phenomena, including the famous Casimir effect, can also be accounted for without invoking vacuum fluctuations, but instead the field radiated by the charge itself — usually called either the "radiation reaction field" or the "source field" in this context. Of great interest is the fact that in many cases these two alternatives (*i.e.* vacuum *vs* source field) do not constitute completely unrelated models. One can switch from one physical picture to the other by a specific change in formalism: depending on how one chooses to order the operator relating to the system and the operator relating to the electromagnetic field, one finds the vacuum field and the radiation reaction field having different contributions to the effect. This begs the question as whether vacuum field *vs* source field constitute an instance of theory underdetermination, or whether one can find a fact of the matter regarding which one holds. We shall examine what arguments and criteria have been used in support of these alternatives. Notably, the alternative interpretations in terms of vacuum or source field have often been thought to be intimately related to the Fluctuation Dissipation Theorem (FDT). We shall discuss how the latter has been interpreted, and what can it really teach us about vacuum fluctuations. Having surveyed these debates, we shall elucidate where the matter stands at present, and what currently seems to be the best evidence in favor of vacuum fluctuations.

First however, let us present the two models in greater detail: we shall first examine the concept of vacuum fluctuations, and its relation to virtual particles and zero-point energies, then we shall turn to the notion of radiation reaction field as a source field.

5.1.2 Vacuum fluctuations, virtual particles and zero-point energies: what are they?

The concept of vacuum fluctuations arises in Quantum Field Theory (QFT), in connection to the zero-point energy of the electromagnetic field. The notion of virtual particles is often found associated to it, and both are often stated to arise courtesy of the uncertainty relations.

Quantum Field Theory (QFT) vs Classical Field Theory and Non-Relativistic Quantum Mechanics (NRQM)

The basic premise of QFT is that matter and radiation can be described in terms of quantum fields.

The difference between the quantum fields of QFT and their classical counterparts is that the former are “quantized”, in a sense analogous to the way the energy of a particle is quantized in non-relativistic quantum mechanics (NRQM).

The main difference between QFT and NRQM lies in the fact that the basic entities of interest in QFT are fields, while the particles that NRQM takes as its direct focus are derivative concepts in QFT, in the following sense. A field is by definition an entity that extends throughout space and whose value generally differs between points.[3] In QFT a field is defined for each type of particle one wishes to deal with (there is a “photon field”, a “fermion field” for electrons, one for positrons, etc...) These being the basic entities of interest, states in QFT are states *of a field.* The notion of particle arises in this context as a property of the field: the state of a field is a “multiparticle state”, which specifies how many particles have, say, a momentum of a given value when the field is in that state.[4].

---

[3] And generally between events, *i.e.* points in spacetime.

[4] Other properties of interest are energy, spin or polarization.

In this description, particles are localized energy quanta, that can be thought of as "local excitations" of the corresponding field. For instance, a photon of frequency ω is an excitation of the photon field of energy $\hbar\omega$. Now talking of particles as field excitations certainly seems to imply that the field is thought to be there anyway, even if no excitations happen to be. And indeed such a situation is thought of as a state of the field: its so-called "vacuum state." This state is the ground state of the field in the same sense as a NRQM particle has a ground state: it is its state of lowest energy.[5]

Zero-point energies and vacuum state

Recall that in NRQM a particle in a "harmonic oscillator" potential energy well can only have energies of $(n + \frac{1}{2}\hbar\omega)$.[6] This notably implies that its minimum energy is not zero, but $(\frac{1}{2}\hbar\omega)$, which is referred to as its *zero-point energy.* This is often described as a consequence of the uncertainty relations between position and

---

[5] This is ignoring many serious issues with a particle interpretation of QFTs. In reality the validity of a particle description in QFT has been vigorously debated in the philosophical literature. There are a number of concerns, associated with different requirements one would expect particles to satisfy, notably countability or localizability. When demanding that particles be countable and obey relativistic energy conditions, we do have grounds for thinking that a particle interpretation for the Fock representation of *free* fields is possible. The eigenvectors of the total number operator have properties appropriate for states containing definite numbers of particles. The no-particle, vacuum state being invariant under the Poincaré Group, it looks the same to all inertial observers, and most importantly, the energy eigenvalues of one- and multi-particle states are $n\sqrt{\mathbf{k}^2 + m^2}$ in agreement with relativity. However this no longer obtains for interacting fields. There are serious problems, both when one tries to represent the interacting field using the free field Fock representation (no state can be interpreted as containing zero quanta, for reasons related to Haag's theorem), and when seeking instead a different Hilbert space representation (attempts involve expressions that are not relativistically covariant or no longer ensure that one-particle states have the desired relativistic energy). Furthermore, even without taking interactions into account, the Unruh effect, whereby an observer accelerating in vacuum would detect Rindler quanta, also presents a challenge for countability [77], [78], [79], [80]. Localizability also faces problems of its own. The fact that under a set of reasonable conditions, the probability of finding a particle in any spatial set is zero constitutes a no-go theorem, and this conclusion has been generalized to generic spacetimes, as well as for unsharp localization [81], [82].

[6] *i.e.* whose potential energy goes as the square of the position coordinate.

momentum, which imply corresponding uncertainties for the potential and kinetic energies, respectively: the momentum and position of a particle cannot both be exactly zero simultaneously, consequently neither can its kinetic and potential energies which would require, respectively, for the particle to have precisely zero momentum, and to be exactly located at its equilibrium position.

Just as there are position and momentum commutation rules in NRQM, in QFT there are commutation relations between the quantum field and the canonical momentum.[7] The latter in turn result in commutation relations between field creation and annihilation operators.[8] As in NRQM, the non-commuting character of these operators again lead to a non-zero eigenvalue for the ground state, *i.e.* the so-called "vacuum state" — or simply "the vacuum." And again, we find an energy of $\frac{1}{2}\hbar\omega$ involved. However whereas in NRQM this value represented the energy of a particle, now in QFT it represents an energy at each point in space. Another difference from NRQM is that this "zero-point energy of the field" does not simply involve one value of ω, but is a sum over a range of values of ω, which a priori could run over arbitrarily large values. This is what leads to the energy of the vacuum state being infinite.[9] These energies of $\frac{1}{2}\hbar\omega$ are considered half quanta since in QFT as in

---

[7] Just as there is a correspondence between the Poisson brackets of classical *particle* theories and the commutators of NRQM, there is also a correspondence between Poisson brackets involving classical *field* theory variables and the canonical commutation relations of QFT.

[8] As we also find commutation relations between raising and lowering operators in NRQM.

[9] The hamiltonian for a free scalar field turns out to be given by:

$$H = \sum_k \frac{1}{2}\,\omega_k\,\left(a(\vec{k})\,a^\dagger(\vec{k}) + a^\dagger(\vec{k})\,a(\vec{k})\right),$$

for the particles, with similar terms for the anti-particles. Now if $a(\vec{k})$ and $a^\dagger(\vec{k})$ commuted, by definition that would mean that the two terms $a(\vec{k})\,a^\dagger(\vec{k})$ and $a^\dagger(\vec{k})\,a(\vec{k})$ were equal. It can be shown that $a^\dagger(\vec{k})\,a(\vec{k})$ is a number operator, $N(\vec{k})$, *i.e.* when it acts on a state it returns the number of particles of momentum $(\vec{k})$ present in that state (analogously to the NRQM number operator returning the number of energy quanta of a particle). We would then get:

$$a(\vec{k})\,a^\dagger(\vec{k}) = a^\dagger(\vec{k})\,a(\vec{k}) = N(\vec{k}),$$

NRQM energy quanta are $\hbar\omega$.[10] For a particle to be present takes a field excitation of a full quantum, so the vacuum state represents a state where no particles are present, thereby justifying the term "vacuum."

However we often find the claim that it is seething with "virtual particles", associated with the half quanta. The phrase "virtual particles" is often used interchangeably with that of "vacuum fluctuations." In the context of the vacuum state, "virtual particles" refers to the idea that the uncertainty relations between energy and time, $\Delta E \Delta t \geq \frac{\hbar}{2}$, allows the field to have an excitation of a full quantum of energy for a very brief time.[11] This time is thought to be too short *as a matter of principle* for the particles to be observed, as a consequence of the uncertainty relations. What part exactly the half quanta play in this account is often not made explicit, but they are thought to arise courtesy of the uncertainty relations by analogy to the ground

---

and when H acts on the vacuum field we would find:

$$H|0\rangle = \sum_k \omega_k \; N(\vec{k})|0\rangle = \sum_k \omega_k \; \cdot 0|0\rangle = 0,$$

i.e., the vacuum state would have an energy eigenvalue of zero.

This is not the case because $a(\vec{k})$ and $a^\dagger(\vec{k})$ do not commute, and instead $[a(\vec{k}), a^\dagger(\vec{k})] = 1$. Then what we get is:

$$a(\vec{k})\, a^\dagger(\vec{k}) = a^\dagger(\vec{k})\, a(\vec{k}) + 1 = N(\vec{k}) + 1$$

$$a(\vec{k})\, a^\dagger(\vec{k}) + a^\dagger(\vec{k})\, a(\vec{k}) = 2a^\dagger(\vec{k})\, a(\vec{k}) + 1 = 2N(\vec{k}) + 1,$$

and when H acts on the vacuum field we find:

$$H|0\rangle = \sum_k \omega_k \; \left(N(\vec{k}) + \frac{1}{2}\right) |0\rangle = \sum_k \omega_k \frac{1}{2}|0\rangle.$$

With the sum running over an infinite number of values of k, this sum is infinite, and we find that the energy of the vacuum state is infinite.

The issue is often dealt with by imposing that the operators should be "normal ordered", that is ordered in such a way that the annihilation operator is always to the right of the creation operator. Imposing this to Eq. (5.1) yields Eq. (5.1). However what we are essentially doing then is arbitrarily imposing that $a(\vec{k})$ and $a^\dagger(\vec{k})$ commute.

[10] The hamiltonian still involves $\hbar\omega N$, not half thereof.

[11] Hence the notion of "fluctuations", *i.e.* variations in time.

state energy of the NRQM harmonic oscillator.[12]

We just saw how the concept of zero-point energy can be related to the "virtual particles" often mentioned in popular accounts of the Casimir force. However it is worth stressing that the concept of "virtual particles" is not merely limited to specific contexts such as the Casimir effect, or the other physical phenomena we shall be concerned with in this chapter. The phrase also refers to "off-shell" particles, which play a staring role in QFT's description of the physical world. In this view, forces between particles are due to the latter exchanging (i.e. emitting or absorbing) other particles specific to the type of force involved. In the formalism, these other particles appear as higher order, so-called "radiative" corrections. Whenever interactions occur, the interaction particles involved are not real but "off-shell", so that according to QFT, all interactions are due to "virtual particles." In contrast to the popular description discussed above, the "off-shellness" of a particle has a very precise, quantitative physical meaning: a particle is said to be off-shell, or more precisely "off the mass shell" when it does not obey the relativistic energy-momentum relation:

$$E^2 - |\vec{p}|^2 c^2 = m^2 c^4. \tag{5.1}$$

However as Robert Klauber discusses in his textbook, the relationship between "virtual particles" understood in this way, and their more popular counterparts discussed above is not immediately obvious.[13] We shall come back to this issue in our concluding chapter (7), but is worth pointing out that in our derivation of the Scharnhorst effect in chapter 4, as in Klaus Scharnhorst's original work, the effect arises from "virtual particles" in the latter sense of "off-shell."

Finally, let us already mention an operational definition of the electromagnetic vac-

---

[12] Particle in an harmonic oscillator potential well.

[13] [83], chapter 10.

uum field that we shall encounter later. In some of the research on the topic, the concept of "free field" is used instead, by which is meant "the solution of the homogeneous field equation (without atomic source term), [which] coincides with the "vacuum field" when no [real] photons are initially present."[14]

Also, the way in which the vacuum is identified as such in derivations is often through its spectral energy density $\rho_0(\omega)$, that is its energy density not only per unit volume, but also per frequency interval $d\omega$. This quantity is the product of the energy of one mode — which in the case of the vacuum is half a quanta, $\frac{1}{2}\hbar\omega$, and the density of modes per unit volume.[15] The resulting expression is:

$$\rho_0(\omega) = \frac{\hbar\omega^3}{2\pi^2 c^3}.^{16} \tag{5.2}$$

5.1.3 Source field and radiation reaction: what are they?

Having discussed the origin of vacuum fluctuations, we now turn to that of the source fields. These are attributed to radiation reaction: the field associated with the radiation reaction force.

Strictly speaking, the radiation reaction force is the Abraham Lorentz force, the second term in the expression for the "self-force."[17] These forces emerged as classical concepts: unlike vacuum field fluctuations which find their origin in the Heisenberg uncertainty principle, the concept of a radiation reaction field exists in classical electromagnetism[18] and have historically been related to attempts to account for the mass of the electron by electromagnetism alone.[19]

---

[14] [84], p.1618. This was standard use, and can notably be found in [85].

[15] Also sometimes referred to as density of states.

[16] The density of modes is $\frac{\omega^2 d\omega}{\pi^2 c^3}$.

[17] "Strictly speaking", because the phrase "radiation reaction force" is sometimes used to refer to the self-force.

[18] Although a QFT treatment of the radiation reaction field requires quantizing it.

[19] See Appendix B.

The expression for the self-force involves all time derivatives of position except for the first:

$$F = m_{elec}\,\ddot{x} - \frac{2}{3}\frac{e^2}{c^3}\,\dddot{x} + \gamma\frac{e^2 a}{c^4}\,\ddddot{x}, \tag{5.3}$$

where m_{elec} is the "electromagnetic mass of the electron", $\frac{2}{3}\frac{e^2}{ac^2}$, a is the radius of the electron, and γ a numerical coefficient that depends on the charge distribution.[20] The "electromagnetic mass" is the factor multiplying the speed of a moving charge in the expression for the momentum of its electromagnetic field. It depends on what a is assumed to be, because it involves integrating over the surface of the charge.[21] The first term on the RHS of Eq. (5.3) *i.e.* $m_{elec}\,\ddot{x}$, corresponds to the force necessary to accelerate the electromagnetic mass. It is sometimes called the inertial force. The work done against this force is $m_{elec}\,\ddot{x}\dot{x} = \frac{1}{2}m_{elec}\,\dot{x}^2$, so it corresponds to the change in kinetic energy associated with the electromagnetic mass of the charge.[22]

The Abraham-Lorentz, radiation reaction force is the second term on the RHS of Eq. (5.3). It is the force on the charge which, *on average*, accounts for the energy it loses by radiating electromagnetic waves, and is thereby necessary for conservation of energy to hold to some extent.[23] In fact, its expression can be derived from this law.[24] Per unit time, the energy radiated by an accelerating charge (*i.e.* the power radiated) has the form:

$$P = \frac{2}{3}\frac{e^2}{c^3}\ddot{x}^2, \tag{5.4}$$

---

[20] This expression applies to an electron, but the self-force of a different charged particle would look similar except for the factor in the first term.

[21] [86], section 28-2.

[22] [86], section 28-7.

[23] Only when one performs a time-average, however. Energy does not seem conserved at a given instant of time, which is one of the several thorny issues raised by Eq. (5.3). See Appendix B for more details.

[24] [87], Appendix A.

where $\ddot{x}$ is the acceleration of the particle, e its charge and c the speed of light. We then consider the energy radiated by this particle over an interval of time, between t_1 and t_2, by integrating the expression above with respect to time:

$$\begin{aligned} E_{radiated} &= \frac{2}{3}\frac{e^2}{c^3}\int_{t_1}^{t_2}\ddot{x}^2 dt \\ &= \frac{2}{3}\frac{e^2}{c^3}\left(\ddot{x}\dot{x}\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2}\dddot{x}\dot{x}dt\right), \end{aligned} \tag{5.5}$$

where the second line is obtained through an integration by parts. If we assume that the charge is undergoing periodic oscillations, and take the time interval $t_2 - t_1$ to be a period,[25] the first term on the right-hand side in the second line of Eq.(5.5) vanishes. The energy that the radiation reaction force needs to provide to the particle in order to make up for this loss is therefore:

$$E_{reaction} = \frac{2}{3}\frac{e^2}{c^3}\int_{t_1}^{t_2}\dddot{x}\dot{x}dt. \tag{5.6}$$

Hence the reaction force is given by:

$$\begin{aligned} E_{reaction} &= \int_{t_1}^{t_2} F_{reaction}\dot{x}dt \\ F_{reaction} &= \frac{2}{3}\frac{e^2}{c^3}\dddot{x}, \end{aligned} \tag{5.7}$$

which we recognize as the Abraham-Lorentz force. Since this force involves as a factor the charge of the particle, it can be identified as due to a so-called "radiation reaction" electric field of the form:

$$\vec{E}_{RR} = \frac{2}{3}\frac{e}{c^3}\dddot{\vec{x}}. \tag{5.8}$$

---

[25] Or an integer number thereof.

The fact that the Abraham-Lorentz force involves the third derivative of x leads to several problems, as is discussed in Appendix B.[26]

Intuitively, the radiation reaction force is the recoil force on a radiating charge due to its own radiation, just as the thrust is the recoil force on a rocket due to the exhaust gases it ejects, or the recoil is the force on a gun when it shoots a bullet. In all these situations, energy and momentum conservation does not apply to the more massive system of interest (charge, rocket, gun), but to the larger system that includes what is being ejected (electromagnetic radiation, exhaust gases or bullet, respectively). As the latter carry away part of the momentum of the total system, the momentum of the system of interest also changes — so that the total momentum remains the same. This implies that a force is exerted on the system of interest, by the ejected part. So for a radiating charge, the recoil, radiation reaction force, is being exerted on the charge by the very radiation it emits. This is why $\vec{E}_{RR}$ is referred to as the "source field": it is emitted by a source, *i.e.* the charge, by opposition to the vacuum field.

5.2 Physical effects deemed evidence for vacuum fluctuations

Several physical effects have been attributed to vacuum fluctuations, and as a result constitute evidence for the existence of the latter. Aside from the Scharnhorst effect which is the topic of the present work, these effects include the Lamb shift, spontaneous emission, the van der Waals force, the Casimir-Polder force and the Casimir force. However, except for the Scharnhorst effect, they can also be accounted for by radiation reaction. As we shall see, which field is responsible can often be

---

[26] [86], chapter 28, pp.6-7. Notably, although the radiation reaction force accounts for the energy a charge loses when radiating, this only holds *on average*, over an entire period of oscillation of an emitting dipole for instance, not for the instantaneous loss.

shown to depend on the way one orders the operators for the field and the atom.

5.2.1 The Stark effect

Several of the phenomena of interest involve the Stark effect, so it is worth posing to describe it.

The Stark effect is the shift and/or splitting in the energy levels of atoms and molecules due to the presence of an external electric field. It is caused by the polarization of the atom or molecule by this field. Indeed, the latter gives rise to a force in the direction of the field on the positively charged nucleus (or nuclei in the case of molecules), and to a force in the opposite direction on the negatively charged electrons. This change in the charge configuration modifies the energy of the electrons.

For this reason, the derivations used to account for the phenomena involving the Stark effect given below are based on two expressions for the energy:

- the polarization energy of a dipole, given by the dot product of its dipole moment and the external electric field. This is the case when the atoms (or molecules) are modeled simply as point dipole oscillators.[27]
- the energy shift caused by the external field in the atomic level of interest.

The phenomena that may involve the Stark effect are the Lamb shift, van der Waals forces, the Casimir-Polder force and the Casimir force.

---

[27] Unlike a physical dipole which consists of two point charges, and whose dipole moment involves the separation between these charges (its magnitude equals the separation times the strength of one charge), a point dipole is an idealized model of the former, in the limit where this separation is taken to zero while keeping the dipole moment fixed.

5.2.2 The Lamb shift

It is fair to say that the Lamb shift is often taken as the best experimental evidence for vacuum fluctuations aside from the Casimir effect. For instance, when seeking alternative evidence for them Martin states:

> We want to investigate whether the zero-point fluctuations could manifest themselves in other physical phenomena [than the vacuum energy]. The answer to this question is usually positive and one experiment which is considered as a proof that the vacuum fluctuations are real is the measurement of the Lamb shift.[28]

The Lamb shift consists in a deviation from the prediction of non-relativistic quantum mechanics regarding the spectrum of the hydrogen atom, according to which the $2s_{1/2}$ and the $2p_{1/2}$ states[29] should have the same energy. Indeed, in the late 1930s and 1940s an energy difference was found experimentally between the two; the $2s_{1/2}$ energy level is in fact higher than the $2p_{1/2}$ by about $4.10^{-6}eV$. The first theoretical derivation of the shift was proposed by Hans Bethe in 1947.

Both derivations of the Lamb shift given below — with and without vacuum fluctuations — model the electron as a point dipole oscillator. Its energy is $W = \frac{1}{2}d.E$, where d is its dipole moment and E the electric field polarizing it. The essential difference between the two derivations lies in the origin of E.[30]

The Lamb shift from vacuum fluctuations

The effect is usually interpreted as arising from fluctuations in the position of the electron within the atomic potential. These fluctuations themselves are attributed

---

[28] [88], p.635; [89], p.61.

[29] principal quantum number $n = 2$, orbital quantum number $l = 0$ and 1 respectively.

[30] [90], p.103.

to interactions between the electron and the electromagnetic field vacuum fluctuations.[31] In essence, this makes the Lamb shift a (quadratic) Stark effect,[32] with the vacuum field in the role of the external electric field.[33]

In this framework, the effect can be derived by taking the vacuum, fluctuating field as the polarizing electric field E. Each mode of the field of frequency ω contributes a factor $E_\omega^2 = 4\pi\rho_0(\omega)d\omega$ to the energy of the dipole, where $\rho_0(\omega)$ is the spectral energy density of the vacuum field.[34]

First the expression for the energy of a point dipole oscillator, $W = \frac{1}{2}d.E$, is expressed for the case of a specific energy level, since we are dealing with an electron in an (hydrogen) atom. As the dipole moment depends on both the field and the polarizability $\alpha_j(\omega)$ through $d = \alpha(\omega)E_\omega$, and the polarizability in turn depends on both the field mode considered and the energy level, the corresponding energy takes the form:

$$W = \frac{1}{2}\alpha(\omega)E_\omega^2. \tag{5.9}$$

This is substituted in the expression for E_ω^2 due to the vacuum field, and integrating the resulting expression over all modes yields the energy W_j that an electron in energy level j experiences due to its polarization by the vacuum field.[35]

This is not quite yet the Lamb shift however: as they represent the energetic effect of being bound, atomic energies are differences between the energy of a bound electron and that of a free electron subjected to otherwise similar physics. So in order to obtain the Lamb shift, one needs to subtract from W_j the corresponding energy of a free electron, W_{free} — *i.e.* the expression for W_j evaluated for a free electron, that

---

[31] [88], pp.635-636; [89], pp.61-62.

[32] The quadratic Stark effect is the second order part of the Stark effect, so-called because it is quadratic in the external field.

[33] [90], p.103.

[34] *Idem.*

[35] [90], p.103.

is in the limit where the transition frequencies between the energy levels are small compared to the frequencies of the field modes.

The measurable level shift is then given by:

$$W'_j = W_j - W_{free} = \frac{2}{3\pi c^3} \sum_i d_{ij}^2 \omega_{ij}^3 log|\frac{\Omega}{\omega_{ij}}| \tag{5.10}$$

where Ω is a cutoff frequency; one then obtains the Lamb shift by taking $\Omega = \frac{mc^2}{\hbar}$.[36]

Alternative model of the Lamb shift: radiation reaction

If instead of substituting an expression for the vacuum field ($E_\omega^2 = 4\pi\rho_0(\omega)d\omega$) in $W = \frac{1}{2}d.E$, one uses the radiation reaction field E_{RR} due to the electron itself, still modeled as a point dipole oscillator, one can derive the same expression for the level shift W'_j.

Formally, the expression used for the radiation reaction field is its mode expansion, involving its dipole moment. The resulting expression contains the product of spatial components of the dipole moment $d_m(t)d_n(t)$, whose expectation value is considered for the state j in order to obtain the result relevant for an electron in the corresponding energy level.[37] In this case as well, the energy in the limit where the electron is free needs to be subtracted.

In addition, this time another subtraction is performed, which has been described as a mass renormalization. Indeed the radiation reaction field used involves implicitly the electromagnetic mass of the electron. When the energy ΔW_j due to its contribution is subtracted, the same result is obtained for the energy shift as when we

---

[36] *Idem.*

[37] Also, the dipole moment is time-dependent so that an integration over time is performed to get the time averaged result.

had assumed the field to be due to the vacuum:

$$W'_j = W_j - W_{free} = \frac{2}{3\pi c^3} \sum_i d_{ij}^2 \, \omega_{ij}^3 \, log|\frac{\Omega}{\omega_{ij}}|. \tag{5.11}$$

Formal origin of the difference: choice of operator ordering

It turns out that the alternative interpretations of the Lamb shift can be related to a specific difference in the formalism: the choice of operator ordering. This becomes clear if we reformulate the derivation.

For simplicity, one may consider a two-level atom model, *i.e.* an atom with only two energy levels available, $|1\rangle$ and $|2\rangle$, where $|1\rangle$ stands for the ground-state and $|2\rangle$ for the excited state.[38]

The quantity of interest to determine the energy shift is $\dot{\sigma}$, where σ is the lowering operator. Prior to choosing a specific operator ordering, this quantity is equal to:[39]

$$\dot{\sigma} = -i\omega_o \sigma + \sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[a_{\mathbf{k}\lambda} + a^{\dagger}_{\mathbf{k}\lambda}]\sigma_z, \tag{5.12}$$

where ω_o stands for the transition frequency between the two states, $a^{\dagger}_{\mathbf{k}\lambda}$ and $a_{\mathbf{k}\lambda}$ are, respectively, the creation and annihilation operators for the electromagnetic field mode of momentum $\mathbf{k}$ and polarization λ, and σ_z represents the population difference between the levels. Depending on the choice made for ordering the field and atomic operators, this expression takes the forms:[40]

---

[38] Dalibart *et al* stressed that a full treatment must take all atomic states into account, and that the two-level treatment is therefore unreliable [84], p.15. It is nevertheless useful in order to illustrate the importance of ordering, and this is why I am using it here, following Milonni's [87].

[39] [87], p.131.

[40] [87], pp.131-132.

with normal ordering:

$$\dot{\sigma} = -i\omega_o\sigma + \sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[\sigma_z \; a_{\mathbf{k}\lambda} + a^\dagger_{\mathbf{k}\lambda} \; \sigma_z]; \tag{5.13}$$

with anti-normal ordering:

$$\dot{\sigma} = -i\omega_o\sigma + \sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[a_{\mathbf{k}\lambda} \; \sigma_z + \sigma_z \; a^\dagger_{\mathbf{k}\lambda}]; \tag{5.14}$$

with symmetric ordering:

$$\dot{\sigma} = -i\omega_o\sigma + \frac{1}{2}\sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[\sigma_z(a_{\mathbf{k}\lambda} + a^\dagger_{\mathbf{k}\lambda}) + (a_{\mathbf{k}\lambda} + a^\dagger_{\mathbf{k}\lambda})\sigma_z]. \tag{5.15}$$

If we refrain from assuming that the electromagnetic field is the source field alone, or the vacuum field alone, and instead regard it as possibly containing contributions from both, the electromagnetic field creation and annihilation operators in the expressions above must take the form:

$$a_{\mathbf{k}\lambda} = a_{\mathbf{k}\lambda,0} + a_{\mathbf{k}\lambda,S}. \tag{5.16}$$

This yields:[41]

- with normal ordering:

$$\langle\dot{\sigma}(t)\rangle = -i\omega_o\langle\sigma(t)\rangle + \sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[\langle\sigma_z(t) \; a_{\mathbf{k}\lambda,S}(t)\rangle + \langle a^\dagger_{\mathbf{k}\lambda,S}(t) \; \sigma_z(t)\rangle]; \tag{5.17}$$

---

[41] [87], p.132.

- with anti-normal ordering:

$$\begin{aligned}\langle\dot{\sigma}(t)\rangle = -i\omega_o\sigma &+ \sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[\langle a_{\mathbf{k}\lambda,0}(t)\ \sigma_z(t)\rangle + \langle a_{\mathbf{k}\lambda,S}(t)\ \sigma_z(t)\rangle + \langle \sigma_z(t)\ a^{\dagger}_{\mathbf{k}\lambda,0}(t)\rangle \\ &+ \langle \sigma_z(t)\ a^{\dagger}_{\mathbf{k}\lambda,S}(t)\rangle]; \end{aligned} \tag{5.18}$$

- with symmetric ordering:

$$\langle\dot{\sigma}(t)\rangle \cong -i\omega_o\langle\sigma(t)\rangle + \frac{1}{2}\sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[\langle \sigma_z(t)\ a^{\dagger}_{\mathbf{k}\lambda,0}\rangle + \langle a_{\mathbf{k}\lambda,0}\ \sigma_z(t)\rangle].^{42} \tag{5.19}$$

As Eq. (5.17) shows, when the operators for the field and the atom are normal ordered, the contribution due to the vacuum field vanishes, and only the source field contributes. The reverse happens when one elects to use a symmetric ordering instead of Eq. (5.19), while anti-normal ordering involves contributions from both the radiation reaction field and the vacuum field Eq. (5.18).

Yet in all three cases, after further calculations, one ultimately recovers the same result:

$$\langle\dot{\sigma}(t)\rangle \cong -i[\omega_o - (\Delta_2 - \Delta_1)]\langle\sigma(t)\rangle - \beta\langle\sigma(t)\rangle.^{43} \tag{5.20}$$

When anti-normal ordering is used, the energy shift $-(\Delta_2 - \Delta_1)$ arises as the sum of a shift $(\Delta_2 - \Delta_1)$ due to the source field and a contribution $-2(\Delta_2 - \Delta_1)$ from the vacuum.[44]

This result can be generalized to a real atom to give the energy shift ΔE_n in an energy level E_n instead of simply Δ_1 or Δ_2. Doing so does lead to the correct value of the Lamb shift.[45]

---

[42] [87], p.141.
[43] [87], p.135.
[44] [87], p.140.
[45] [87], p.135.

5.2.3 Spontaneous emission

Very much the same issues as were just discussed for the Lamb shift also arise in the case of spontaneous emission. However, there is an important difference: unlike the Lamb shift, spontaneous emission cannot be interpreted as arising entirely from vacuum fluctuations.[46]

Spontaneous emission[47] is commonly interpreted as a form of *stimulated* emission, with vacuum fluctuations as the origin of the disturbance. Prior to this explanation however, the source of the disturbance used to be ascribed to the radiation reaction field of the electron.[48]

A stimulated emission rate can be calculated according to either picture, depending on the ordering of the relevant operators. However this time, in the case of a contribution limited to the vacuum field, the result comes out as only half the correct value for the spontaneous emission rate.[49]

Vacuum field contribution

The rate of stimulated emission for a transition of frequency ω_0 is the Einstein coefficient B, times the spectral density of states:[50]

$$R_{vf} = B\,\rho(\omega_0), \tag{5.21}$$

---

[46] [87], p.143.

[47] *i.e.* the emission of a photon by an atom, during which the atom transitions from a higher to a lower excited state, without the process being due to the atom having been visibly disturbed.

[48] [87], pp.142-143.

[49] [87], pp.131-132; 143.

[50] B is a measure of the amplitude (hence the probability) for an atomic state to undergo stimulated emission. The spectral density of states is the number of atomic states per unit frequency.

where $B = \frac{4\pi^2 d^2}{3\hbar^2}$, with d the transition dipole moment.
The vacuum field has a spectral energy density $\rho_0 = \frac{\hbar\omega_0^3}{2\pi^2 c^3}$.[51] This implies that if we interpret spontaneous emission as a case of stimulated emission due to the vacuum field, the transition rate should be:

$$R_{vf} = \frac{2d^2\omega_0^3}{3\hbar c^3}. \tag{5.22}$$

This however is only half the actual spontaneous emission rate.

Radiation field contribution

Let us now assume spontaneous emission to be in fact induced by the radiation reaction field that acts on the electron. We again wish to derive the emission rate, now dubbed R_{rr}, modeling the electron in its atom as a physical dipole oscillator. As discussed above, the electron experiences a radiation reaction force:

$$F = \frac{m_{elec}}{e}\,\ddot{x} - \frac{2}{3}\frac{e^2}{c^3}\,\dddot{x}. \tag{5.23}$$

This implies that it radiates energy at a relative rate:

$$\frac{dW}{W dt} = -\frac{e^2\omega_0^2}{3mc^3}. \tag{5.24}$$

---

[51] Indeed the spectral energy density $\rho(\omega)$ is equal to the number of field modes in a given frequency interval $[\omega, \omega + d\omega]$, per unit volume, times the energy of each mode. This number is given by $\frac{\omega^2 d\omega}{\pi^2 c^3}$, so since in its vacuum state each mode of the field has the 0-point energy $\frac{1}{2}\hbar\omega$:

$$\rho_0 = \frac{\omega^2 d\omega}{\pi^2 c^3}\frac{1}{2}\hbar\omega = \frac{\hbar\omega^3}{2\pi^2 c^3}.$$

The atom's rate of stimulated emission due to the radiation reaction of the electron is the product of this rate of energy loss by the oscillator strength f of the transition:

$$\begin{aligned} R_{rr} &= \frac{e^2\omega_0^2}{3mc^3}\frac{2md^2\omega_0}{\hbar e^2} \\ &= \frac{2d^2\omega_0^3}{3\hbar c^3}, \end{aligned} \tag{5.25}$$

which, again, is half the value for spontaneous emission.

This suggests that spontaneous emission can be interpreted as stimulated emission due to both the vacuum field and the radiation reaction field of the electron.

Formal origin of the difference: choice of operator ordering

As with the Lamb shift, for spontaneous emission as well, the different interpretations arise from a choice of operator ordering in the formalism. This is again most easily seen if we consider the two-level atom model.[52]

The quantity of interest for spontaneous emission is $\langle \dot{\sigma}_z(t) \rangle$, where σ_z is the population difference between the two levels. Irrespective of ordering, this is given by:[53]

$$\dot{\sigma}_z = -2\sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[a_{\mathbf{k}\lambda} + a_{\mathbf{k}\lambda}^{\dagger}][\sigma + \sigma^{\dagger}].^{54} \tag{5.26}$$

---

[52] [87], chapter 4; [91].

[53] [87], p.131.

[54] Recall that σ is the lowering operator, so $\sigma^{\dagger}$ is the raising operator, and $a_{\mathbf{k}\lambda}^{\dagger}$ and $a_{\mathbf{k}\lambda}$ are, respectively, the creation and annihilation operators for the electromagnetic field mode of momentum $\mathbf{k}$ and polarization λ.

Under different choices of operator ordering, this equation becomes:[55]

- with normal ordering:

$$\dot{\sigma}_z = -2\sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[(\sigma+\sigma^\dagger)\, a_{\mathbf{k}\lambda} + a^\dagger_{\mathbf{k}\lambda}\, (\sigma+\sigma^\dagger)]; \tag{5.27}$$

- with anti-normal ordering:

$$\dot{\sigma}_z = -2\sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[a_{\mathbf{k}\lambda}\, (\sigma+\sigma^\dagger) + (\sigma+\sigma^\dagger)\, a^\dagger_{\mathbf{k}\lambda}]; \tag{5.28}$$

- with symmetric ordering:

$$\dot{\sigma}_z = -\sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[(\sigma+\sigma^\dagger)(a_{\mathbf{k}\lambda} + a^\dagger_{\mathbf{k}\lambda}) + (a_{\mathbf{k}\lambda} + a^\dagger_{\mathbf{k}\lambda})(\sigma+\sigma^\dagger)]. \tag{5.29}$$

In each case, one then takes the expectation value of the relevant equation for a field-atom state $|0\rangle|\psi\rangle$, where the field is in its vacuum state $|0\rangle$ and $|\psi\rangle$ is an arbitrary state of the two-state atom.[56]

In all cases, carrying through the calculation yields:[57]

$$\begin{aligned}\langle\dot{\sigma}_z(t)\rangle &= -4\beta\langle\sigma^\dagger(t)\sigma(t)\rangle && (5.30)\\ &= -2\beta\,[1+\langle\sigma_z(t)\rangle] && (5.31)\end{aligned}$$

with

$$\langle\sigma_z(t)\rangle = -1 + [1+\langle\sigma_z(0)\rangle]e^{-2\beta t}.^{58} \tag{5.32}$$

---

[55] [87], pp.131-132.

[56] [87], p.132.

[57] [87], p.136.

[58] Recall that β stands for the decay rate.

This does account for the observed physical behavior: if the atom is initially in the upper state, then $\langle\sigma_z(t)(0)\rangle = |c_2(0)|^2 - |c_1(0)|^2 = 1$ and the atom decays at a rate 2β to the lower state corresponding to $\langle\sigma_z(t)\rangle = -1$. If on the other hand the atom is initially in the lower state, $\langle\sigma_z(t)(0)\rangle = -1$, then $\langle\sigma_z(t)\rangle = -1$, which means that it remains in the lower state: there is no such thing as spontaneous absorption. However, depending on the choice of operator ordering, one reaches Eq. (5.30) via different paths, which suggest different physical pictures.

In the case of normal ordering, *i.e.* taking the expectation value of Eq. (5.27) yields:

$$\langle\dot{\sigma}_z(t)\rangle = -2\sum_{\mathbf{k}\lambda} C_{\mathbf{k}\lambda}[\langle\sigma(t)\, a_{\mathbf{k}\lambda,S}(t)\rangle + \langle a^\dagger_{\mathbf{k}\lambda,S}(t)\,\sigma(t)\rangle + \langle\sigma^\dagger(t)\, a_{\mathbf{k}\lambda,S}(t)\rangle + \langle a^\dagger_{\mathbf{k}\lambda,S}(t)\,\sigma^\dagger(t)\rangle], \tag{5.33}$$

that is a result which contains only source field contributions.

Anti-normal ordering, that is taking the expectation value of Eq. (5.28) ultimately leads to:[59]

$$\langle\dot{\sigma}_z(t)\rangle \cong \langle\dot{\sigma}_z(t)\rangle_{vacuum} + \langle\dot{\sigma}_z(t)\rangle_{source} \cong -4\beta\langle\sigma_z(t)\rangle - 2\beta[1 - \langle\sigma_z(t)\rangle] \cong -2\beta[1 + \langle\sigma_z(t)\rangle], \tag{5.34}$$

while symmetric ordering by using Eq. (5.29) yields:[60]

$$\langle\dot{\sigma}_z(t)\rangle \cong \langle\dot{\sigma}_z(t)\rangle_{vacuum} + \langle\dot{\sigma}_z(t)\rangle_{source} \cong -2\beta\langle\sigma_z(t)\rangle - 2\beta \cong -2\beta[1 + \langle\sigma_z(t)\rangle]. \tag{5.35}$$

So in the case of spontaneous emission, as for the energy level shift, anti-normal ordering corresponds to contributions from both the source field and the vacuum field. However, symmetric ordering also yields a contribution from both fields to spontaneous emission, whereas in the case of the Lamb shift, only the vacuum

---

[59] [87], p.140.

[60] [87], pp.141-142.

contributed. In fact, it turns out that spontaneous emission cannot be entirely accounted for by the vacuum, whatever choice of ordering one makes.[61]

5.2.4 Van der Waals force: force between two atoms

Considering that we are particularly interested in the Casimir effect, which was first derived on the assumption that it was due to van der Waals forces, it is obviously of interest to see what models can account for the latter.

The expression “van der Waals forces” refers to forces between neutral atoms or molecules, due to their ability to behave as dipoles. Depending on the nature of the polarization, the forces are given specific names.

If the molecules have *permanent* dipole moments, the force between them is called the Keesom force. The force between a molecule with a permanent dipole (or quadrupole) moment and another molecule, whose polarization is induced by the first, is the Debye force. Even in the absence of molecules with permanent dipole moments, a van der Waals force exists between polarizable atoms and molecules, due to their induced instantaneous dipole moments; in this case it is called the London dispersion force.[62] It is the London force which is at stake in the Casimir effect.

Van der Waals forces arise because the energy level of an atom (molecule) A is modified by the presence of another atom (molecule) B. The resulting shift in A’s energy depends on the distance d between the two atoms (molecules). This gives rise to an attractive force as the energy of the system is lowered when d decreases.

---

[61] [87], p.142.

[62] It was discovered by London in 1930. The term “dispersion” refers to the fact that it depends on their polarizability, which is related to the dispersion.

van der Waals force from vacuum fluctuations

The van der Waals, London force can be accounted for on the basis of vacuum fluctuations.[63]

Recall from our discussion on the Lamb shift that generally speaking, the potential energy acquired by an atom due to its polarization by an external field is (Eq. (5.9)):

$$W = \frac{1}{2}\alpha(\omega)E_\omega^2. \tag{5.36}$$

We are now considering a situation where we have two atoms, A and B, and we are interested in the polarization energy of A.

Since we want to take vacuum fluctuations into account — a purely quantum effect — we need to treat the polarizing field as a quantum entity. So we now quantize it, which involves expanding it into modes $(\mathbf{k}, \lambda)$:

$$\mathbf{W}_A = -\frac{1}{2}\sum_{\mathbf{k}\lambda} \alpha_A(\omega_k)\mathbf{E}^2_{\mathbf{k}\lambda}(\mathbf{x}_A, t), \tag{5.37}$$

where the subscript A refers to atom A, and we are considering quantum fields, *i.e.* the $\mathbf{E}_{\mathbf{k}\lambda}$ are operators.

The external field polarizing A is assumed to contain contributions from both the dipole field of atom B, $E_{B,\mathbf{k}\lambda}$, and the electromagnetic vacuum fluctuations $E_{0,\mathbf{k}\lambda}$:

$$E_{\mathbf{k}\lambda}(\mathbf{x}_A, t) = E_{B,\mathbf{k}\lambda}(\mathbf{x}_A, t) + E_{0,\mathbf{k}\lambda}(\mathbf{x}_A, t). \tag{5.38}$$

---

[63] [87], section 3.11, pp.98-105.

Then:

$$\begin{aligned}\mathbf{W}_A &= -\frac{1}{2}\sum_{\mathbf{k}\lambda}\alpha_A(\omega_k)[\mathbf{E}^2_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)+\mathbf{E}^2_{B,\mathbf{k}\lambda}(\mathbf{x}_A,t)\\ &+ \mathbf{E}_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)\cdot\mathbf{E}_{B,\mathbf{k}\lambda}(\mathbf{x}_A,t)+\mathbf{E}_{B,\mathbf{k}\lambda}(\mathbf{x}_A,t)\cdot\mathbf{E}_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)]. \qquad (5.39)\end{aligned}$$

It turns out that the van der Waals force can be accounted for by the cross terms alone, *i.e.* the interference terms between the vacuum field and the field of atom B:

$$\begin{aligned}\mathbf{W}_{AB} &= -\frac{1}{2}\sum_{\mathbf{k}\lambda}\alpha_A(\omega_k)[\mathbf{E}_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)\cdot\mathbf{E}_{B,\mathbf{k}\lambda}(\mathbf{x}_A,t)\\ &+ \mathbf{E}_{B,\mathbf{k}\lambda}(\mathbf{x}_A,t)\cdot\mathbf{E}_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)]. \qquad (5.40)\end{aligned}$$

The vacuum field is then expressed in terms of its positive and negative frequency components, $E^+_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)$ (which contains an annihilation operator) and $E^-_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)$ (a creation operator):

$$E_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)=E^+_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)+E^-_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t). \qquad (5.41)$$

The energy W_{AB} is given by the vacuum expectation value of $\mathbf{W}_{AB}$:

$$\begin{aligned}W_{AB} &= -\frac{1}{2}\sum_{\mathbf{k}\lambda}\alpha_A(\omega_k)\left(\langle 0|(E^+_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)+E^-_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t))\cdot\mathbf{E}_{B,\mathbf{k}\lambda}(\mathbf{x}_A,t)|0\rangle\right.\\ &+ \left.\langle 0|\mathbf{E}_{B,\mathbf{k}\lambda}(\mathbf{x}_A,t)\cdot(E^+_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)+E^-_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t))|0\rangle\ \right). \qquad (5.42)\end{aligned}$$

Since $\langle 0|E^-_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)=0$ and $E^+_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t)|0\rangle=0$:

$$\begin{aligned}W_{AB} &= -\frac{1}{2}\sum_{\mathbf{k}\lambda}\alpha_A(\omega_k)\left(\langle 0|(E^+_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t))\cdot\mathbf{E}_{B,\mathbf{k}\lambda}(\mathbf{x}_A,t)|0\rangle\right.\\ &+ \left.\langle 0|\mathbf{E}_{B,\mathbf{k}\lambda}(\mathbf{x}_A,t)\cdot(E^-_{0,\mathbf{k}\lambda}(\mathbf{x}_A,t))|0\rangle)\ \right). \qquad (5.43)\end{aligned}$$

As remarked above, $E_{B,\mathbf{k}\lambda}$ is the dipole field of B. We are here dealing with the London force, so B does not have a *permanent* dipole moment, but instead *induced instantaneous* dipole moments.[64] In the present derivation, this polarization is ascribed to the effect of the fluctuating vacuum field surrounding B, so that the dipole moment of B is given by:

$$p_B(t) = \sum_{\mathbf{k}\lambda} \alpha_B(\omega_k)[\mathbf{E}^+_{0,\mathbf{k}\lambda}(\mathbf{x}_B, t) + \mathbf{E}^-_{0,\mathbf{k}\lambda}(\mathbf{x}_B, t)], \tag{5.44}$$

where α_B is the polarizability of atom B. The dipole field due to B at A, $\mathbf{E}_{B,\mathbf{k}\lambda}(\mathbf{x}_A, t)$, can then be expressed in terms of the vacuum field at B, and substituted into Eq. (5.43). After performing the necessary simplifications, the energy $W_{AB} \equiv V(r)$ can be evaluated, and in the limit of small separation r between the two atoms one then recovers the result obtained by London, *i.e.* a potential that behaves as r^{-6}:

$$V(r) \cong -\frac{3\,\hbar\,\omega_0\,\alpha^2}{4\,r^6}, \tag{5.45}$$

with α the polarizability of the atoms, where it has been assumed that one transition, at frequency ω_0, is dominant.

If instead of small separation, one considers the potential in the "retarded regime" where the two atoms are widely separated, one obtains the result derived by Casimir and Polder, with the potential decreasing as r^{-7}:

$$V(r) \cong -\frac{23\,\hbar\,c}{4\,\pi\,r^7}\,\alpha_A\alpha_B. \tag{5.46}$$

---

[64] It should be noted that for an atom in a stationary state, the *expectation value* for the dipole moment is zero. However the dipole moment is not *identically* zero; the effect depends on its *instantaneous* value, which fluctuates about, but is generally not equal to, zero.

The corresponding forces are obtained by differentiating the potentials with respect to r.

The origin of the van der Waals force is the same in both regimes (Casimir-Polder's result in the retarded limit and London's in the non-retarded one), and in the present model, it is due to the vacuum fluctuations in two ways: firstly, the relevant part of the polarization energy of the atom considered, *i.e.* A, comes from the cross terms between B's dipole field and the vacuum field at A's position, and secondly, B's dipole field arises because the dipole moment of B undergoes fluctuations about its 0-mean, due to vacuum fluctuations at B's position.

van der Waals force from radiation reaction

The van der Waals force can also be derived using a model based on radiation reaction. In this description the presence of atom B constitutes a boundary condition which modifies the radiation reaction field on atom A; hence B creates an additional contribution to the reaction field on A, compared to a free-space situation in which B would not be present. The polarization energy due to this additional field is what is responsible for the van der Waals force in this model.[65]

In order to find this field, we first seek its vector potential. Although we are not interested in vacuum fluctuations, this derivation treats the EM field as quantized. Generally speaking, the modal decomposition of a vector potential operator can be written as:

$$\hat{\mathbf{A}}(x,t) = \sum_{\alpha} \left(\frac{2\,\pi\hbar c^2}{\omega_\alpha} \right)^{\frac{1}{2}} [\hat{a}_\alpha(t)\mathbf{F}_\alpha(x) + \hat{a}^\dagger_\alpha(t)\mathbf{F}^*_\alpha(x)], \tag{5.47}$$

---

[65] [90], p.105.

where $\mathbf{F}_\alpha(x)$ and $\mathbf{F}^*_\alpha(x)$ are known as "mode functions." The mode functions are simply solutions of the classical Helmholtz equation:

$$\nabla^2 \mathbf{F}_\alpha(x) + k_\alpha^2 \mathbf{F}_\alpha(x) = 0.^{66} \tag{5.48}$$

They are required to satisfy the transversality condition:

$$\nabla \cdot \mathbf{F}_\alpha(x) = 0, \tag{5.49}$$

and, crucially, boundary conditions, which determine their form for a specific physical configuration.

The creation and annihilation operators $\hat{a}^\dagger_\alpha(t)$ and $\hat{a}_\alpha(t)$ in the vector potential are found by solving the Heisenberg equation of motion for these operators. In the case of vector potential for the radiation reaction field, $\hat{\mathbf{A}}_{RR}(x,t)$, this is done in the presence of sources. The resulting expression is:

$$\hat{\mathbf{A}}_{RR}(x,t) = 2\pi iec \sum_\alpha \omega_\alpha^{-1} \mathbf{F}_\alpha(x) \cdot \mathbf{F}^*_\alpha(x) \int_0^t dt' \ \hat{\mathbf{r}}(t') e^{i\omega_\alpha(t'-t)} + h.c., \tag{5.50}$$

where e is the charge on which the field acts.

The electric field operator for the radiation reaction field $\hat{\mathbf{E}}_{RR}(x,t)$ is then obtained by taking the time derivative of $\hat{\mathbf{A}}_{RR}(x,t)$ after it has been evaluated.

The key point for a radiation reaction account of the van der Waals force is that:

> The field of radiation reaction acting on a particle depends on the modal properties of the field, and consequently on the electromagnetic environment of the particle.[67]

---

[66] The mode functions are not operators; the operator character of the expression for the vector potential is contained in the creation and annihilation operators $\hat{a}^\dagger_\alpha(t)$ and $\hat{a}_\alpha(t)$.

[67] [90], p.105.

That is, what form Eq. (5.50) takes for a specific situation depends on the boundary conditions, which affect the expression for the mode functions.

For instance, in free-space — that is, when no other charges are present than the one experiencing the radiation reaction — they become $\mathbf{F}_\alpha(x) = V^{\frac{1}{2}}\mathbf{e}_{\mathbf{k}\lambda}e^{i\mathbf{k}\cdot x}$, which ultimately leads to the familiar radiation reaction electric field:

$$\mathbf{E}_{RR} = \frac{2e}{3c^3}\frac{d^3x}{dt^3} - \frac{\delta m}{e}\frac{d^2x}{dt^2}, \tag{5.51}$$

as in Eq. (5.3) above.

In the context of the van der Waals force, the mode functions describing the radiation reaction field at A are affected by the presence of atom B.

It turns out that these mode functions contain two terms: one corresponds to the free-space case, *i.e.* $\mathbf{e}_{\mathbf{k}\lambda}e^{i\mathbf{k}\cdot x}$, and the second one is due to the dipole field of B; *i.e.* for each mode:

$$\begin{aligned}&\mathbf{e}_{\mathbf{k}\lambda}e^{i\mathbf{k}\cdot x} + \alpha_B(\omega_k)k^3e^{i\mathbf{k}\cdot x_B}e^{ikr}\\ &\qquad\times\left[\mathbf{e}_{\mathbf{k}\lambda}\left(\frac{1}{kr}+\frac{1}{k^2r^2}-\frac{1}{k^3r^3}\right)-\frac{1}{r^2}(\mathbf{e}_{\mathbf{k}\lambda}\cdot\mathbf{r})\mathbf{r}\left(\frac{1}{kr}+\frac{3i}{k^2r^2}-\frac{3}{k^3r^3}\right)\right].\end{aligned} \tag{5.52}$$

These mode functions are then substituted in Eq. (5.50), and $\mathbf{E}_{RR}$ obtained through $\mathbf{E}_{RR} = -(1/c)\dot{\mathbf{A}}_{RR}$.

The van der Waals force is then found by calculating the shift in A's ground state energy due to E_{RR}, which again depends on the distance d between the two atoms, and differentiating it with respect to this distance, whereby one recovers the usual results (notably the energy is proportional to d^{-6} for $d \to 0$ and to d^{-7} with retardation for $d \to \infty$).

In this picture the van der Waals interaction energy is essentially a Stark effect where the external field is the contribution to the radiation reaction field on A due to the presence of B. If we recall that the Stark effect on an atom A due to A's free-space radiation reaction field results in the Lamb shift, we see that the van der Waals interaction potential can be described as the modification to A's Lamb shift due to the presence of B. Because this contribution is a function of the separation between the two atoms, it gives rise to a force between them, in contrast to the Lamb shift which only depends on the local fields.

In contrast, the description of the van der Waals interaction in terms of vacuum field fluctuations attributes it to a Stark effect in which the external field is made up of the cross terms between the vacuum field at the position of atom A and atom B's dipole field fluctuations, themselves due to the vacuum field at B's position (while the Stark effect due to the vacuum field at A *per se* gives rise to the Lamb shift).

This suggests that the Casimir effect could be interpreted as due to the same distance-dependent Stark energies, with the polarizing field coming from either of these same contributions.

When interpreted in terms of vacuum fluctuations, the atoms in, say, plate 1 attract the atoms of plate 2 (and vice versa) because the atoms in plate 1 are getting polarized (hence stark shifted) by field cross terms between the vacuum field at plate 1 and the fluctuating dipole fields emitted by the atoms of plate 2, due to their dipole moment being made to fluctuate by the vacuum field at plate 2.

When interpreted in terms of radiation reaction, the atoms in plate 1 are being polarized (hence stark shifted) by the part of their radiation reaction field affected by the presence of the atoms in plate 2.

Note that in this picture, to say that the Casimir effect is due to van der Waals forces is not to presume whether it ultimately arises from source fields or vacuum fluctuations. However, it is not immediately clear how even the vacuum fluctuations model relates to the account that the force is due to the density of vacuum fluctuations being smaller between the plates than in the outer regions.[68]

A problem which superficially at least seems closer to the Casimir effect is the origin of the Casimir-polder force, to which we now turn.

5.2.5 Casimir-Polder force

In 1948 Casimir and Polder derived the force between a (neutral) atom and a perfectly conducting wall for large separations d. For short d, the force can be obtained as a simple image problem, *i.e.* by substituting for the conducting wall an identical atom placed a distance d behind where the wall would have been, as both configurations yield the same potential configuration. What is determined in this way is the dipole-dipole interaction between the atom and its image, and it is found to be proportional to d^{-3}. However at large d, it was known that the potential must fall off as d^{-4}. This is the behavior that Casimir and Polder recovered in 1948.

The Casimir-Polder force from vacuum fluctuations

The Casimir-Polder force can be accounted for in terms of vacuum fluctuations by applying appropriate boundary conditions to the vector field for the latter.[69] The vector field inside a parallelepiped made of conducting walls takes the form:

$$A_x(\mathbf{r}) = \left(\frac{8}{V}\right)^{\frac{1}{2}} a_x \, cos(k_x x) \, sin(k_y y) \, sin(k_z z), \tag{5.53}$$

$$A_y(\mathbf{r}) = \left(\frac{8}{V}\right)^{\frac{1}{2}} a_y \, sin(k_x x) \, cos(k_y y) \, sin(k_z z), \tag{5.54}$$

---

[68] See the Introduction.

[69] [87], pp.106-7.

$$A_z(\mathbf{r}) = \left(\frac{8}{V}\right)^{\frac{1}{2}} a_z\, sin(k_x x)\, sin(k_y y)\, cos(k_z z), \tag{5.55}$$

with V the volume of the parallelepiped, and $a_x^2 + a_y^2 + a_z^2 = 1$.[70] These expressions are used to obtain the corresponding electric field at the position of the atom, a distance d from the $z = 0$ wall (*i.e.* at $(\frac{L}{2},\, \frac{L}{2},\, d)$), which is then substituted in the expression for the atomic energy level shift due to the Stark effect:

$$W_A = -\frac{1}{2}\sum_{\mathbf{k}\lambda} \alpha_A(\omega_k)\mathbf{E}^2_{\mathbf{k}\lambda}(x_A, t),^{71} \tag{5.56}$$

and the polarizability is then approximated as the static polarizability, *i.e.* $\alpha = \alpha(0)$. The interaction energy sought, *i.e.* of the atom due to a single conducting wall, is the difference between this expression, with d finite, and the limit where $d \to \infty$, *i.e.* the same configuration with the wall of interest removed. It is found to be:

$$V(r) = -\frac{3\alpha\hbar c}{8\pi d^4}, \tag{5.57}$$

which displays the desired d^{-4} dependence.[72]

As usual the corresponding force is obtained by differentiating this distance dependent energy with respect to d.

In this model the force on the atom is due to the conducting wall affecting the vacuum field at the position of the atom: it forces the vacuum field to be zero on its surface, thereby “killing off” most of its modes.

---

[70] [87], p.55.

[71] [87], p.101.

[72] [87], pp.106-107.

The Casimir-Polder force from radiation reaction

The way the Casimir-Polder force can be accounted for through radiation reaction involves a treatment entirely analogous to the way we obtained the van der Waals force above, except for the form taken by the mode functions, which must now satisfy different boundary conditions: *i.e.* a half-infinite space bounded by a perfectly conducting wall, instead of another atom. This yields for the relevant component of the radiation reaction field:

$$E_{RR,z} = \frac{2e}{3c^3}\frac{d^3z}{dt^3} - \left(\frac{4\Omega}{3\pi c^3}\right)\frac{d^2z}{dt^2} - \left(\frac{2e}{4d^2c}\right)\dot{z}(t - \frac{2d}{c}) - \left(\frac{2e}{8d^3}\right)z\left(t - \frac{2d}{c}\right), \quad (5.58)$$

where $\hat{z}$ is the direction normal to the wall. We recognize in the first term on the right hand side the free-space radiation reaction field. The additional terms correspond to the retarded dipole field from an image atom, as the presence of the conducting wall has the same effect as another, identical atom would have if placed a distance $2d$ from our atom of interest.[73]

From this field, the interaction energy can be deduced by substitution in the expression for the energy shift W_j of an atomic level j. In this way the Casimir-Polder energy is found to be:

$$W_j = -\frac{3\alpha_j \hbar c}{8\pi d^4}, \quad (5.59)$$

where α_j is the static polarizability for state j (*i.e.* $\alpha_j = \alpha_j(\omega) \to \alpha_j(0)$).

Hence the Casimir-Polder force can be interpreted, again, as due to a distance-dependent Stark effect energy; this time the field inducing the effect is the part of the radiation reaction field of the atom due to the presence of the conducting wall.

---

[73] [90], pp.104-105.

5.2.6 The Casimir effect

As stated above the Casimir effect is often said to be the most direct experimental evidence for vacuum fluctuations. Yet it too can be accounted for without appeal to zero-point energies. In fact, the original derivation by Casimir did not involve this concept, and it has been recently stressed by R.L. Jaffe that "the Casimir energy can be expressed entirely in terms of Feynman diagrams with external legs — *i.e.* in terms of S-matrix elements which make no reference to the vacuum."[74]

The Casimir effect from vacuum fluctuations

As briefly discussed in the Introduction, the Casimir force can be interpreted as arising from a difference in vacuum energies. The conducting plates act as boundaries, forcing the electric fields to vanish at their location. Consequently, for a given plate separation L, only modes of the field of wavelength λ with $\frac{n\lambda}{2} = L$ can form between the plates. Popular accounts often ascribe the effect to a difference in pressure related to the density of virtual particles, since these are viewed as equivalent to vacuum fluctuations: more modes, hence more virtual particles, can form outside than between the plates; consequently, a greater density of the latter exists outside and the associated pressure leads to the observed attractive force between the plates.

The formal derivation corresponds to a slightly different situation however. The energy density in the region between the plates is not compared to the outer region, but to what it would be in the absence of plates. This difference turns out to be dependent on the separation of the plates in such a way that closer plates imply a lower energy configuration; hence this derivation ascribes the effect to a variation in the energy associated with the vacuum fluctuations between the plates, depending

---

[74] [76], p.3; [92], p.6.

on the configuration of the latter configuration.

On these assumptions, a typical, textbook derivation of the Casimir force proceeds as follows.[75]

For simplicity, a real, massive scalar field $\Phi(t, \mathbf{x})$ is normally used instead of the fields of QED. Being a scalar field, $\Phi(t, \mathbf{x})$ obeys the Klein-Gordon equation. The latter is solved for $\Phi(t, \mathbf{x})$ by separation of variables, *i.e.* assuming that $\Phi(t, \mathbf{x})$ can be factorized into functions of its independent variables:

$$\Phi(t, \boldsymbol{x}) = X(x)Y(y)Z(z)T(t) \tag{5.60}$$

so that the Klein-Gordon equation takes the form:

$$-\frac{\ddot{T}}{T} + \frac{1}{X}\frac{\mathrm{d}^2X}{\mathrm{d}x^2} + \frac{1}{Y}\frac{\mathrm{d}^2Y}{\mathrm{d}y^2} + \frac{1}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}z^2} - m^2 = 0, \tag{5.61}$$

(where m is the mass of the scalar particle), and can be expressed as the system of four differential equations:

$$\begin{aligned} -\frac{\ddot{T}}{T} - m^2 &= C^2 && (5.62)\\ \frac{1}{X}\frac{\mathrm{d}^2X}{\mathrm{d}x^2} &= k_x^2 && (5.63)\\ \frac{1}{Y}\frac{\mathrm{d}^2Y}{\mathrm{d}y^2} &= k_y^2 && (5.64)\\ \frac{1}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}z^2} &= k_z^2. && (5.65) \end{aligned}$$

These have plane waves as solutions:

$$X(x) = A_x\mathrm{e}^{ik_x x} + B_x\mathrm{e}^{-ik_x x} \tag{5.66}$$

---

[75] [88], pp.637-638; [89], pp.63-64.

$$Y(y) = A_y \mathrm{e}^{ik_y y} + B_y \mathrm{e}^{-ik_y y} \tag{5.67}$$

$$Z(z) = A_z \mathrm{e}^{ik_z z} + B_z \mathrm{e}^{-ik_z z} \tag{5.68}$$

$$T(t) = A_t \mathrm{e}^{i\omega t} + B_t \mathrm{e}^{-i\omega t}, \tag{5.69}$$

where k_x, k_y and k_z represent the wave vectors in the corresponding spatial directions, and ω the frequency of the mode. That $C - k_x^2 - k_y^2 - k_z^2 = 0$ and $\omega^2 = C^2 + m^2$ determines the expression for ω in terms of the wave vectors:

$$\omega = \sqrt{k_x^2 + k_y^2 + k_z^2 + m^2}. \tag{5.70}$$

So far however, we have not taken the presence of the plates into account. If k_x, k_y and k_z were allowed to take continuous values, the solutions we have obtained would be those for a scalar field in free space, unconstrained. This is what the presence of conducting plates at $z = 0$ and $z = L$ prevents: as stated above the Casimir plates are treated formally by imposing boundary conditions on $Z(z)$ such that:

$$Z(0) = 0 \text{ and } Z(L) = 0. \tag{5.71}$$

This implies that k_z cannot take continuous values, but is restricted to be:

$$k_z = n\frac{\pi}{L}, \tag{5.72}$$

where n is an integer. In other words, in the z-direction only some modes can form between the plates. However, no such restriction applies to $Z(z)$ and k_z outside the plates; there is no boundary condition such that $Z(\infty) = 0$ and $Z(-\infty) = 0$, and consequently k_z can take continuous values there.

If we quantize the scalar field $\Phi(t, \mathbf{x})$, expanding it in terms of creation and annihilation operators $c_{k_x,k_y,n}$ and $c^\dagger_{k_x,k_y,n}$ that obey the usual commutation relations for bosons, we therefore find that in the presence of Casimir plates:

$$\Phi(t,\boldsymbol{x}) = \int \frac{\mathrm{d}k_x}{(2\pi)^{1/2}} \int \frac{\mathrm{d}k_y}{(2\pi)^{1/2}}$$

$$\sum_{n=1}^{\infty} \sqrt{\frac{2}{L}} \frac{1}{\sqrt{2\omega}} \sin\left(n\frac{\pi}{L}z\right) \left(c_{k_x,k_y,n}\mathrm{e}^{-i\omega t+ik_xx+ik_yy} + c^\dagger_{k_x,k_y,n}\mathrm{e}^{i\omega t-ik_xx-ik_yy}\right).^{76} \quad (5.73)$$

Knowing the expression for the field, one can then derive the energy due to the presence of the plates, and from it the Casimir force.

One first needs to obtain the energy density, which is:

$$\mathcal{H} = T_{00} = \frac{1}{2}\dot{\Phi}^2 + \frac{1}{2}\delta^{ij}\,\partial_i\Phi\,\partial_j\Phi \,+\, \frac{1}{2}m^2\Phi^2.^{77} \quad (5.74)$$

Substituting $\Phi(t,\mathbf{x})$ in this expression, and integrating over volume in order to find the total energy between the plates rather than the energy density yields:

$$\int_{\text{Cavity}} \mathrm{d}\boldsymbol{x}\,\langle 0|T_{00}|0\rangle = \left(\frac{D}{2\pi}\right)^2 \int \mathrm{d}k_x \int \mathrm{d}k_y \sum_{n=1}^{\infty} \frac{1}{2}\sqrt{k_x^2 + k_y^2 + n^2\frac{\pi^2}{L^2} + m^2},^{78} \quad (5.75)$$

where D^2 is the area of the plates, which are finite in this calculation. This expression is equivalent to:

$$\left(\frac{D}{2\pi}\right)^2 \int \mathrm{d}k_x \int \mathrm{d}k_y \sum_{n=1}^{\infty} \frac{\omega}{2} \quad (5.76)$$

with

$$\omega = \sqrt{k_x^2 + k_y^2 + n\frac{\pi}{L} + m^2}. \quad (5.77)$$

Because the formalism provides for an infinite number of modes — *i.e.* the upper bound of the integrals and sum are taken to be infinity — these expressions are

---

[77] [88], p.577; [89], p.11.

[78] [88], p.639; [89], p.65.

infinite. Consequently they need to be regularized. Using dimensional regularization results in:

$$E = \left(\frac{D}{2\pi}\right)^2 \frac{\pi}{2} \left(\frac{\pi}{L}\right)^3 \frac{\Gamma(-3/2)}{\Gamma(-1/2)} \sum_{n=1}^{\infty} n^3 = -D^2 \frac{1}{L^3} \frac{\Gamma(2)}{2^4 \pi^2} \zeta(4). \tag{5.78}$$

The energy of interest is that due to the presence of the plates, so it is the *difference* between the energy when the plates are present and when they are not, ΔE. However, after regularization the latter energy is zero: the absence of plates can be modeled by taking the distance between them (*i.e.* L) to infinity, so that ΔE is simply E.

In order to find the force itself, one differentiates the surface energy density, $\frac{\Delta E}{D^2}$, with respect to the distance between the plates. This yields:

$$\frac{\Delta E}{D^2} = -\frac{\pi^2}{1440 L^3},^{79} \tag{5.79}$$

$$F = -\frac{\partial}{\partial L}\left(\frac{\Delta E}{D^2}\right) = -\frac{\pi^2}{480\, L^4}. \tag{5.80}$$

Casimir force: Source theory and choice of operator ordering

As for the Lamb shift and spontaneous emission, the different interpretations of the Casimir force can be seen to arise from a choice of operator ordering [17]. This was notably shown by Peter Milonni in a derivation based on a more realistic set-up as the one used above for the standard calculation of the Casimir force. Instead of perfectly conducting plates and a vacuum between them, the configuration used consists in three regions of different dielectric constants:[80]

---

[79] $\zeta(4) = \pi^4/90$.

[80] This model had been used by Lifshitz [93]; [17], p.4242.

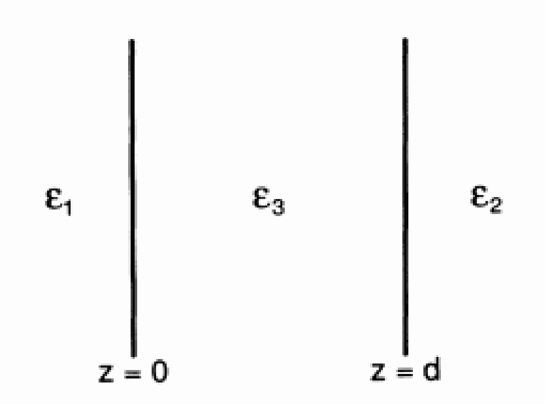


Fig. 5.1: Configuration used to derive the Casimir force.

Regions 1 and 2 represent the plates — *i.e.* whereas in the previous section they were modeled as boundary conditions imposed on the field at z = 0 and z = d, now the plates are represented as semi-infinite dielectric slabs. More specifically, they are treated as made of point dipole oscillators. What these are meant to model are current fluctuations within the plates. Note that there are no regions exterior to the plates in this model. Perhaps more surprising though is the representation of the space between the plates: it too is made of a dielectric, modeled by point dipole oscillators. The vacuum case is recovered only at the end of the calculation, by taking the relevant dielectric constant, ϵ_3, to 1. Similarly, the plates are made conducting by taking the dielectric constants ϵ_1 and ϵ_2 to infinity at the end. No change is made in the geometrical configuration however, *i.e.* regions 1 and 2 remain semi-infinite. Hence this model leaves no room for ascribing the Casimir force to a pressure difference between the Casimir vacuum and the normal vacuum outside the plates — which as discussed in the previous section does not fairly represent the standard vacuum fluctuations calculation either.[81]

The relevant quantity to derive in this case is the expectation value of the energy of the dipoles. Each dipole has potential energy $-\frac{1}{2}\mathbf{p}.\mathbf{E}$ where $\mathbf{p}$ is its dipole moment

---

[81] [17] pp.4242, 4247.

and **E** the polarizing field, hence the expectation energy per unit volume is:

$$\langle E\rangle = -\frac{1}{2}\int d^3r\langle \mathbf{P}.\mathbf{E}\rangle, \tag{5.81}$$

where $\mathbf{P} = N\mathbf{p}$ is the polarization due to N dipoles per unit volume.

The polarizing field is assumed to be composed of both the vacuum field $\mathbf{E}_0$ and the source field $\mathbf{E}_S$ due to radiation reaction:

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}_0(\mathbf{r},t) + \mathbf{E}_S(\mathbf{r},t).^{82} \tag{5.82}$$

However before substituting the two fields in the expression for the energy (Eq. (5.81)), an ordering is chosen for the atomic and field operators.[83]

- Normal ordering:

With normal ordering of **E** and **P**, Eq. (5.81) results in an expression which involves only the source field:

$$\langle E\rangle = -\frac{1}{2}\int d^3r\langle \mathbf{E}^{(-)}.\mathbf{P} + \mathbf{P}.\mathbf{E}^{(+)}\rangle,^{84} \tag{5.83}$$

i.e.:

$$\langle E\rangle = -\frac{1}{2}\int d^3r\langle \mathbf{E}_S^{(-)}.\mathbf{P} + \mathbf{P}.\mathbf{E}_S^{(+)}\rangle,^{85} \tag{5.84}$$

since $\langle 0|\mathbf{E}_0^{(-)} = 0$ and $\mathbf{E}_0^{(+)}|0\rangle = 0$. One then substitutes the mode expansion of the operators $E_S^{(+)}$ and $E_S^{(-)}$. In order to get the Casimir force, what is then derived is the change in energy $\langle \delta E\rangle$ brought about by adding a small layer of dipoles of width δd at $z = d$. When the resulting expression for the force $F(d)$ is considered

---

[82] [17], p.4243.

[83] The creation operator for the field is contained in the negative frequency component of the field, $\mathbf{E}^{(-)}$ and the annihilation operator in $\mathbf{E}^{(+)}$, so that $\mathbf{E} = \mathbf{E}^{(+)} + \mathbf{E}^{(-)}$.

[84] [17], p.4244.

[85] [17], p.4244.

for the specific case of conducting plates separated by vacuum by taking the limits $\epsilon_1 \to \infty$, $\epsilon_2 \to \infty$ and $\epsilon_3 \to 1$, one recovers the Casimir force:

$$F(d) = -\frac{\pi^2 \hbar^2 c}{240\, d^4}, \tag{5.85}$$

where d is the separation between the plates.[86]

- Symmetric ordering:

If instead of normal ordering the operators in Eq. (5.81) we choose a symmetric ordering, we obtain the same final result for the Casimir force, but only the vacuum field contributes.

This time neither the vacuum nor the source field is immediately cancelled out; Eq. (5.81)[87]

$$\langle E \rangle = -\,\frac{1}{2} \int d^3 r\; \langle \frac{1}{2}\,\mathbf{P} \cdot (\mathbf{E}^{(+)} + \mathbf{E}^{(-)}) + \frac{1}{2}\,(\mathbf{E}^{(+)} + \mathbf{E}^{(-)}) \cdot P \rangle, \tag{5.86}$$

with $\mathbf{E}^{(+)} = \mathbf{E}_0^{(+)} + \mathbf{E}_S^{(+)}$ and $\mathbf{E}^{(-)} = \mathbf{E}_0^{(-)} + \mathbf{E}_S^{(-)}$. However when one then calculates the source field contribution: $\langle E_S \rangle = -\frac{1}{4} \int d^3 r\; \langle \mathbf{P} \cdot \mathbf{E}_S + \mathbf{E}_S \cdot \mathbf{P} \rangle$, it turns out that it vanishes, so that the vacuum field alone is responsible for the effect when the operators are symmetrically ordered.[88]

---

[86] [17], p.4247.

[87] Recall it is: $\langle E \rangle = -\frac{1}{2} \int d^3 r \langle \mathbf{P}.\mathbf{E} \rangle$.

[88] [17], pp.4247-4248.

5.2.7 Operator ordering and ambiguity in the physical interpretation

Operator ordering as the source of underdetermination

As discussed above, whether a physical effect is to be attributed to the vacuum field or to the radiation reaction source field depends on the way one orders the operators for the field and the atom. In the present context, the ambiguity in the physical picture that the formalism depicts originates from the fact that the quantities of interest involve the product of two operators, and that a priori we are free to order these in any way we wish.

To see this let us represent by Q the operator for the quantity that we wish to find. In the simplest case Q is proportional to the product of an operator for the total electromagnetic field, E, with an operator that refers to some atomic variable, which we shall call P. That is:

$$Q \sim PE. \tag{5.87}$$

E and P commute at equal times, so we are free to order them any way we wish: the derivations that follow yield the same results.[89]

However different choices of ordering lead to different physical pictures because E is the sum of operators that do *not* commute with P — the 0-point vacuum field E_0 and the source field E_S:

$$E = E_0 + E_S. \tag{5.88}$$

---

[89] This is discussed by Dalibard *et al* before they treat specifically of the Lamb shift and the spin anomaly of the electron [84]. Because of the application they have in mind, they consider the time evolution of an operator G, $\frac{dG(t)}{dt}$, rather than an arbitrary operator Q as I have done here.

The quantity of interest, Q, is due to the two corresponding contributions, Q_0 and Q_S.[90] If we chose the order P E, we now find:

$$Q = PE = P(E_0 + E_S) = PE_0 + PE_S = Q_0 + Q_S, \tag{5.89}$$

and for E P:

$$Q = EP = (E_0 + E_S)P = E_0P + E_SP = Q_0 + Q_S. \tag{5.90}$$

That seems all fine, but although their sum does, E_0 and E_S do not commute with P:

$$E_0P \neq \ PE_0 \quad E_SP \neq \ PE_S.^{91} \tag{5.91}$$

So Q_0 is not the same thing in Eq.(5.89) as in Eq.(5.90) — and the same goes for Q_S of course. Therefore, depending on how we choose to order P and E, the total effect represented by Q will be due to contributions from Q_0 and Q_S in different proportions — even though one recovers the same result for Q itself whatever choice is made.

We see that the underdetermination arises because, in so far that some operators commute, we get similar final results, and in so far that the operators that make

---

[90] [84] refers to these quantities as $\left(\frac{dG(t)}{dt}\right)_{vf}$ and $\left(\frac{dG(t)}{dt}\right)_{sr}$, respectively, where the subscripts stand for "vacuum field" and "self-reaction."

[91] An explanation given for this is:

> One's freedom to change orderings in mid-calculation is limited because different operators may acquire differing time arguments and because the interaction of the field with the atom may make the separation of different degrees of freedom difficult at a time other than the initial instant. ([85], p.157)

It may seem strange that operators referring to a field and an atom, hence to different systems, can fail to commute. However, if we consider for instance the polarization of an atom, the fact that it is due to the field leads to the polarization being expressed in terms of the creation and annihilation operators of the field.

them up do not commute, we get different accounts of which field is responsible for the effects.

Reactions to the ambiguity

Different attitudes towards this underdetermination can be found in the literature. There is agreement regarding the formal equivalence of the choice of ordering. However some researchers have argued that physical interpretation itself should be ground for regarding one ordering as being the correct one. Specifically, they deem that each separate contribution (*i.e.* such as Q_0 and Q_S above) should have a clear physical meaning, and that for this to be the case, they need to be formally described by a Hermitian operator. Only symmetric ordering corresponds to this case.[92]
Therefore, at this point it would seem that imposing the condition that Q_0 and Q_S must be Hermitian implies symmetric ordering.
This is not true in all cases however, notably not when dealing with a model of relevance to our concerns: the two-level atom.[93]
Indeed in this case the operator of interest Q is given not by PE but by a sum over terms of the form $E^-P^- + E^+P^+$.[94] In this case choices of ordering that lead to different relative contributions from E_0 and E_S nevertheless correspond to Q_0 and Q_S being Hermitian. So imposing that Q_0 and Q_S be Hermitian is not enough to remove the ordering freedom and the underdetermination associated with it. In addition, one needs to demand that Q be written in a form that involves products

---

[92] [84]. See Appendix C for the derivation of this result. Normal ordering involves placing annihilation operators to the right of creation operators, anti-normal ordering places them to the left, and symmetric ordering consists in a linear combination of both.

[93] The two-level atom is often used to discuss spontaneous emission and the Lamb shift, hence its importance.

[94] E^- is the negative frequency component of the field which involves the creation operator alone, whereas the positive frequency component E^+ involves the annihilation operator. The same holds for P^- and P^+.

of only E with P,[95] not products of E^+ with P^+ (or E^- P^-) as above. That is, operators for the system and for the fields must be Hermitian operators throughout the derivation. The ordering that corresponds to this requirement is again the symmetric ordering.

Hence the symmetric one is the only ordering where:

- the vacuum field contribution and the source field contribution to the effect are represented by Hermitian operators.
- operators for the system and for the fields are Hermitian operators throughout the derivation.

For these reasons, it has been argued that the underdetermination is only superficial. The reasoning is that symmetric ordering should be the preferred choice, because observables are represented by Hermitian operators, so preferring symmetric ordering corresponds to imposing that the operators have physical meaning throughout the derivation. And symmetric ordering involves the vacuum field as well as the source field — in fact, equal contributions from both ([84], [94]). However, as will be discussed below, this line of reasoning has not met with universal agreement.

5.3 The fluctuation-dissipation theorem (FDT)

5.3.1 Description

Aside from operator Hermiticity, another issue has played an important role in the way physicists have thought of the relative roles of the vacuum and source fields: the fluctuation-dissipation theorem (FDT). The relation between the two fields has been viewed as a special instance of this theorem [95].

The FDT was proposed in 1951 by Herbert B. Callen and Theodore A. Welton. It generalizes a result derived by H. Nyquist in 1928, which relates voltage fluctuations

---

[95] or products of appropriate combinations of $E^+ - E^-$ and $P^+ - P^-$. [84], p.1625.

to resistance in electric circuits [96]. What fascinated Callen and Welton about Nyquist's relation and prompted them to generalize it was that it doesn't simply relate physical quantities for a given physical situation, but two different processes: it "relates a property of a system in *equilibrium* (*i.e.* the voltage fluctuations) with a parameter which characterizes an irreversible process (*i.e.* the electrical resistance)."[96]

Callen and Welton considered an arbitrary dissipative system, *i.e.* a system able "to absorb energy when subjected to a time-periodic perturbation."[97] Their FDT then states that, for such a system, the fluctuating force on it ($\mathscr{F}_f$) obeys:

$$\langle \mathscr{F}_f^2 \rangle = \frac{2}{\pi} \int_0^\infty R(\omega) E(\omega, T) d\omega, \tag{5.92}$$

where:

$$E(\omega, T) = \frac{1}{2}\hbar\omega + \hbar\omega[e^{(\frac{\hbar\omega}{kT})} - 1]^{-1} \tag{5.93}$$

is the energy of the system, which corresponds to the expression for the mean energy of an oscillator of natural frequency ω at temperature T, and the "generalized resistance" $R(\omega)$ characterizes the *dissipation*, given by:

$$R(\omega) = Re\left(\frac{\mathscr{F}_d}{\dot{Q}}\right),^{98} \tag{5.94}$$

---

[96] [95], p.34.

[97] *Idem.*

[98] [95], p.37.

where $\dot{Q}$ is the response of the system and $\mathscr{F}_d$ the dissipative force.[99]

Note the two terms in Eq. (5.93): the first one corresponds to zero-point energies, and in the $T = 0$ case $E(\omega, T)$ reduces to this term alone. In contrast, in the high temperature limit, $E(\omega, T)$ can be approximated by the second term which is then equal to kT.[100]

Applying the theorem requires identifying the generalized resistance, hence usually the dissipative force and the response in Eq . (5.94).

5.3.2 Examples

Nyquist's relation

By deriving the relation that led to Callen and Welton's generalization in the form of the FDT, Nyquist meant to account for the observation of an electromotive force (emf) in conductors in the absence of an external potential, due to their thermal agitation alone.

When we identify:

- the fluctuating force $\mathscr{F}_f$ with the emf,
- the generalized resistance $R(\omega)$ with the usual electric resistance,

---

[99] [95], p.35. More precisely, they considered a periodic perturbing force of the form:

$$\mathscr{F}_f = \mathscr{F}_{f0}\ sin(\omega t).$$

They found the power dissipated by the system to be proportional to the square of this perturbation, and they related the proportionality constant to generalized notions of impedance $Z(\omega)$ and resistance $R(\omega)$:

$$Power = \mathscr{F}_f^2 \frac{R}{Z}.$$

Callen and Welton actually did not distinguish between $\mathscr{F}_f$ and $\mathscr{F}_d$ as I have done here: they used the same symbol for both, suggesting that the force which fluctuates is the same force as that responsible for the dissipation. However in the examples that they discussed as applications of their theorem, the distinction is intuitively helpful, as we shall see below.

[100] This can be seen by using the binomial expansion on the second term with $kT << \hbar\omega$, which serves as the condition for the high temperature limit.

and we consider the high temperature limit where $E(\omega, T) \backsim kT$, the FDT (*i.e.* Eq. (5.92)) becomes:

$$\langle emf^2 \rangle \simeq \frac{2}{\pi} kT \int_0^\infty R(\omega)\, d\omega, \tag{5.95}$$

in agreement with Nyquist's result.[101]

Brownian motion

Applying the FDT to Brownian motion (the random movements that small particles in a fluid exhibit) one recovers the known form of the fluctuating force on the particles. In this case the following identifications need to be made:

- $\mathscr{F}_f$: force(s) due to the fluid as its molecules hit the system,
- $\mathscr{F}_d$: viscous force, responsible for dissipation, and equal to $-\eta v$, where v is the speed of the particle and $\eta = 6\pi$ viscosity $\times$ particle radius,
- response $\dot{Q}$: speed v of the particles,

consequently, from Eq. (5.94):

- the generalized resistance $R(\omega)$ is equal to $-\eta$.

Then in the high temperature limit $E(\omega, T) \backsim kT$, the FDT (*i.e.* Eq.(5.92)) becomes:[102]

$$\langle \mathscr{F}_f^2 \rangle = \frac{2}{\pi} kT\eta \int d\omega, \tag{5.96}$$

in agreement with the known force responsible for Brownian motion.

Radiating charge

The two examples given above provide an intuitive feel for the FDT, notably because they are classical, as we took the high temperature limit in both cases.

---

[101] [95], p.37.

[102] $\langle \mathscr{F}_f^2 \rangle = \frac{2}{\pi} \int_0^\infty R(\omega)\, E(\omega, T)\, d\omega$.

Of greater interest to our present concerns regarding electromagnetic fields however, Callen and Welton also applied their theorem to the radiation emitted by an accelerating charge. More specifically, they discussed a physical dipole oscillator.[103] They sought to find the force responsible for the fluctuations that their theorem implies must exist, knowing that a dissipative effect occurs due to the radiation reaction force, $-\frac{2e^2}{3c^3}\dddot{x}$, responsible for the energy loss of the charge. In order to do so, they again needed to find what the generalized resistance $R(\omega)$ is in the situation at hand.

In this case the relevant identifications are:

- $\mathscr{F}_f$: the fluctuating force on the charge,
- $\mathscr{F}_d$: the radiation reaction force, $-\frac{2e^2}{3c^3}\dddot{x}$, where $x = \frac{P_0}{e} sin(\omega t)$ with P_0 the dipole moment of the oscillator at maximum amplitude,
- $\dot{Q}$: the speed of the charge $\dot{x}$,

consequently from Eq. (5.94):

- the generalized resistance $R(\omega)$ is equal to $-\frac{2}{3}\frac{e^2\omega^2}{c^3}$.

What the FDT implies is that there must be on the charge a randomly fluctuating force $\mathscr{F}_f = e\mathscr{E}_x$, of the form:

$$\langle \mathscr{F}_f^2 \rangle = \langle e^2 \mathscr{E}_x^2 \rangle = \frac{2}{\pi} \int_0^\infty E(\omega, T) \;\; \frac{2}{3}\frac{e^2\omega^2}{c^3} \; d\omega. \tag{5.97}$$

This time Callen and Welton did not take the high-temperature limit to recover a classical result, but kept the exact form of $E(\omega, T)$ (Eq. (5.93)), including the zero-point energy term $\frac{1}{2}\hbar\omega$ that it would reduce to at $T = 0$. They noted that the energy density of an isotropic radiation field is directly related to the fluctuating

---

[103] [95], pp.37-38.

electric field by:

$$\text{energy density} = \frac{\langle \mathscr{E}^2 \rangle}{4\pi} = \frac{3}{4\pi} \langle \mathscr{E}_x^2 \rangle \tag{5.98}$$

and therefore the FDT directly leads to:

$$\text{energy density} = \frac{3}{4\pi e^2} \langle \mathscr{F}_f^2 \rangle = \frac{1}{\pi^2 c^3} \int_0^\infty \left(\frac{1}{2} \hbar\omega + \hbar\omega \left[exp\left(\frac{\hbar\omega}{kT} \right) - 1 \right]^{-1} \right) \omega^2 d\omega. \tag{5.99}$$

This energy density clearly involves two contributions:

- the Planck radiation law, given by the second term:

$$\frac{1}{\pi^2 c^3} \int_0^\infty \hbar\omega^3 [exp(\frac{\hbar\omega}{kT}) - 1]^{-1} d\omega, \tag{5.100}$$

- the zero-point energy term, to which the energy density reduces in the limit of $T \longrightarrow 0$ (but which is also present at non-zero temperatures):

$$\int_0^\infty \frac{\hbar\omega^3}{2\pi^2 c^3} d\omega. \tag{5.101}$$

Recall that the spectral energy density of the vacuum $\rho_0(\omega)$ has the form:

$$\rho_0 = \frac{\hbar\omega^3}{2\pi^2 c^3}. \tag{5.102}$$

So the FDT states that the fluctuating force due to the zero-point energy term involves an expression of the same functional form as the spectral energy density of the vacuum $\rho_0(\omega)$.

5.3.3 Possible interpretations

Looking at Eqs. (5.99) and (5.101), one would think that the implications of the FDT for the existence of vacuum fluctuations of the electromagnetic field are quite straight forward. Fluctuations in this field occur as thermal effects at non-zero temperature, but do not vanish at $T = 0$ when the field is in its ground state.

Of course in view of its derivation, using the result of the FDT for a radiating charge as evidence for the existence of vacuum fluctuations certainly does not seem free of circularity: we have used a zero-point contribution in Eq. (5.93) for $E(\omega, T)$, so we should hardly be surprised that we get one in our final result. We would get one too for the Nyquist circuit and Brownian motion if we did not choose to neglect it by taking the high temperature limit.

Yet when we consider the physics literature pertaining to debates on the issue, we do find disagreement regarding the implications of the FDT for the existence of vacuum fluctuations. Associated to those, we also find differences in the way the FDT has been interpreted. It is not even always clear from the outset whether the FDT (whatever its interpretation) is acting as a premise in the discussions. Often, it seems more like the considerations in the light of which it is being interpreted are themselves put forward as arguments for or against the idea of vacuum fluctuations. It is these interpretations that we now turn to.

The classical, high temperature limit applications of the FDT discussed above provide us with an intuitive grasp on the meaning of this theorem. Peter Milonni has described the latter as follows:

> If a system is coupled to a "bath", that can take energy from the system in an effectively irreversible way, then the bath must also cause fluctuations. The fluctuations and the dissipation go hand in hand, we cannot have one without the other. ([87], p.54)

In the case of the electric circuit discussed by Nyquist, the dissipative force and the fluctuating force are both exerted on the electrons by the ionic lattice of the metal,

	System	Bath	Generalized Resistance	Dissipative Force $\mathscr{F}_d$	Fluctuating Force $\mathscr{F}_f$
Circuit discussed by Nyquist	electrons	ionic lattice of the metal	electrical resistance	• origin: due to thermal agitation • exerted by: the metal lattice	• origin: emf fluctuations due to thermal agitation • exerted by: the metal lattice
Brownian Motion	particle	molecules of the fluid	viscosity	• origin: viscous force due to thermal agitation • exerted by: the fluid	• origin: impacts from the surrounding molecules • exerted by: the fluid

Table 5.1: The FDT applied to the Nyquist circuit and Brownian motion

so that the system is best described to be the electrons, and the bath the lattice.[104] For Brownian motion, the dissipative effect is due to the viscous force exerted on the particle by the fluid, and the fluctuating forces are also exerted by the molecules of the fluid as they hit the particle. So in this case the system is the particle, and the bath is the molecules of the surrounding fluid.

This is summarized in Table 5.1, and we note that in these cases the fluctuating and dissipative forces are both ultimately exerted by the same entity (*i.e.* metal lattice, fluid molecules).[105]

---

[104] In fact Nyquist himself considered two conductors. In his scheme, the fluctuations originate in one conductor and dissipation in the other: the dissipative effect is the heat generated in the latter, which Nyquist identified as due to *emf* fluctuations in the circuit that are ultimately ascribed to thermal agitation in the first conductor. But naturally the distinction between these two conductors is artificial, meant to associate different parts of one and the same system, *i.e.* the circuit, with the two distinct processes of fluctuation and dissipation, for the sake of conceptual clarity when discussing these processes. In reality both fluctuation and dissipation occur throughout the resistive circuit.

[105] This common source of the forces is presumably the reason why Callen and Welton did not distinguish between $\mathscr{F}_f$ and $\mathscr{F}_d$ in their derivation, representing both by $\mathscr{F}$. The rationale for distinguishing between $\mathscr{F}_f$ and $\mathscr{F}_d$ as I have done is that they play different roles, occurring in

The question of interest to us is: if we now try to give an analogous description of the FDT for a radiating charge, what can it teach us, if anything, about the radiation reaction field, the vacuum field fluctuations, and the relationship between them?

Here the dissipative effect is due to radiation reaction, so the dissipative force is exerted by the radiation reaction field. This is really only a matter of definition, since the radiation reaction field is a concept that was introduced in order to account for the (average) energy loss as a charge radiates.

Since this time we did not take the high-temperature limit, we get two contributions to the fluctuating forces: the electromagnetic force due to the Planck radiation field, and the electromagnetic force $\mathscr{F}_{f0}$. Whether the latter can be ascribed to the vacuum electromagnetic field is what is at stake.

Now for the non-zero temperature contribution, the dissipative force and the fluctuating force are exerted by the same entity: the field that exerts a reaction on the charge as the charge emits is the field it emits (see Table 5.2). Hence the bath is the field radiated by the charge, the source field.

For the $T = 0$ contribution though, things get more interesting. The relevant dissipative force is the same as for the non-zero temperature contribution, since in deriving their result Callen and Welton substituted for it only the radiation reaction force and no other contributions. So it is again the field radiated by the charge. What does this imply for $\mathscr{F}_{f0}$?

If we take the FDT to entail that the fluctuating force is exerted by the same entity as the dissipative force, then we are led to the conclusion that $\mathscr{F}_{f0}$ is due to the radiation reaction field, *i.e.* to the radiated, source field. This could be argued to be

---

the context of qualitatively different processes: $\mathscr{F}_d$, being dissipative, directly increases entropy, whereas $\mathscr{F}_f$ does not — it pertains to an equilibrium situation in so far that its expectation value is zero.

	System	Bath	Generalized Resistance	Dissipative Force $\mathscr{F}_d$	Fluctuating Force $\mathscr{F}_f$
Radiating charge	the charge	?	$-\frac{2}{3}\frac{e^2\omega^2}{c^3}$	• origin: radiation reaction • exerted by: the radiation reaction field $-\frac{2e^2}{3c^3}\dddot{x}$ *i.e.* the electromagnetic field radiated by the charge	• origin: - non-0 temperature contribution: Planck radiation - $T = 0$ contribution: zero-point effects • exerted by: - non-0 temperature contribution: the Planck electromagnetic field *i.e.* the electromagnetic field radiated by the charge - $T = 0$ contribution $\mathscr{F}_{f0}$: the zero-point electromagnetic field *i.e.* the electromagnetic field in its ground, “vacuum” state ?? Or radiation reaction field ?? In either case: $\int_0^\infty \frac{\hbar\omega^3}{2\pi^2c^3}d\omega$

Table 5.2: The FDT applied to a radiating charge

the natural interpretation of the FDT, in so far that looking at Eqs. (5.99), (5.101) and at our table, it seems clear that both thermal and zero-point effects have the same relation to the dissipative force, hence to the radiation reaction field. So if the fluctuating force associated with thermal effects is due to the radiated field, why not that associated with zero-point effects? Furthermore, one usually thinks of the FDT as involving a system and one, single bath that exerts both $\mathscr{F}_d$ and $\mathscr{F}_f$.

Then again, most applications of the FDT involve classical physics, hence the high temperature limit. So perhaps one should not take for granted that the identity of $\mathscr{F}_d$ and $\mathscr{F}_f$ necessarily holds for the zero-point energy contributions. And if one then takes the stance that the FDT does not require both types of forces to be exerted by the same entity, then one is tempted to identify $\mathscr{F}_{f0}$ as due to the vacuum field.

It is instructive to compare these two alternatives to the high-temperature examples we discussed, since being classical they are intuitive. There is an interesting difference between the bath for the Nyquist circuit and Brownian motion on the one hand, and for the Planck radiation on the other. In the former two cases, the bath is a medium that exists independently of the system — and as a result the dissipative forces exerted on the system are clearly external forces, whose appearance the system had nothing to do with. By contrast, in the case of Planck radiation, the bath has been emitted by the system itself — hence the dissipative force being a reaction force. Attributing $\mathscr{F}_{f0}$ to the vacuum field or to radiation reaction entails a similar difference, which is arguably the main, defining difference between the two fields: the vacuum field is thought to exist irrespective of whether any charge is or has ever been accelerating in it.

5.3.4 Interpretations of the fluctuation dissipation theorem in the physics literature

When we consider the physics literature on the topic, disagreements center around whether $\mathscr{F}_f$ and $\mathscr{F}_d$ can be said to be exerted by the same entity for the $T = 0$ contribution and whether or not the fluctuating forces $\mathscr{F}_f$ are due to vacuum field fluctuations.[106]

Perhaps surprisingly, the discussion has not primarily focused on whether the FDT necessarily implies that $\mathscr{F}_f$ and $\mathscr{F}_d$ are both exerted by the same entity, with the view to then draw conclusions regarding the specific case of an accelerating charge. In fact, it is often unclear whether the FDT is actually playing the part of premise in the discussions, and it seems interpreted in light of other considerations, themselves put forward as arguments for or against the idea that the vacuum field is responsible for $\mathscr{F}_f$. We now examine these arguments, and what criteria have been implicitly used to assess the origin of $\mathscr{F}_f$.

The FDT implies the existence of vacuum fluctuations

Callen and Welton Most physicists who have worked on the issue seem to have held the FDT to imply the existence of the vacuum field, understood as a distinct entity from the radiation reaction one. This certainly appears to have been true of Callen and Welton themselves. Indeed, after deriving Eq. (5.99), they commented on their result:

> The first term in this equation gives the familiar infinite "zero-point" contribution, and the second term gives the Planck radiation law. ([95], p.38)

---

[106] More precisely, whether they are partly due to vacuum fluctuations in the general case, and entirely so at $T = 0$.

Hence they identified $\mathscr{F}_f$ as being due in part to the vacuum field, and the criterion they used as the basis for this identification was the functional expression of their spectral energy density: the expression for the spectral energy density of the fluctuating field predicted by the FDT (*i.e.* the integrand of Eq. (5.101)) is identical to the known form of the spectral energy density of the vacuum field.

Peter Milonni Peter Milonni is a physicist who has devoted a lot of his research to the role of radiation reaction and the vacuum field in the various physical effects thought to involve them, and notably to the issue of operator ordering discussed in the previous section. He takes pains to make his interpretation of the FDT explicit. The full text of the excerpt quoted earlier reads:

> What we have here[107] is an example of a "fluctuation-dissipation relation." Generally speaking if a system is coupled to a "bath" that can take energy from the system in an effectively irreversible way, then the bath must also cause fluctuations. The fluctuations and the dissipation go hand in hand; we cannot have one without the other. In the present example the coupling of a dipole oscillator to the electromagnetic field has a dissipative component, in the form of radiation reaction, and *a fluctuation component, in the form of the zero-point (vacuum) field*; given the existence of radiation reaction, the vacuum field must also exist in order to preserve the canonical commutation rule and all it entails.[108]

So Milonni takes the FDT to imply the existence of the vacuum field, as a distinct physical entity from the radiation reaction field, and attributes $\mathscr{F}_f$ to the vacuum field.

---

[107] These remarks come after a discussion of the commutation relation for a dipole oscillator.

[108] [87], p.54. My emphasis.

Yet he does not base these opinions on the FDT alone. His mention of "the canonical commutation rule" refers to a derivation that precedes the remarks just quoted. We shall discuss it in detail in section 4, but it should already be said that it does not make use of the FDT. Furthermore, in addition to this derivation involving the commutation rule, Milonni must also have in mind the issue of operator ordering, which has been the focus of much of his research. Unlike the FDT *per se*, this is an approach where the radiation reaction field and the vacuum field can easily manifest themselves as compatible alternatives as much as mutually exclusive ones, since the ordering can lead to contributions from both fields as readily as from one of them.

Dennis Sciama In the chapter that he contributed to *The philosophy of vacuum* [94], Sciama discussed a number of effects involving "zero-point energy" and "zero-point noise", including of the electro-magnetic field. Of greatest interest here is his treatment of the transition of a two-level atom from the excited to the ground state, accompanied by the emission of a photon.[109] As in Callen and Welton's work, we are dealing with a radiating charge; only this time it is clearly a bound one, *i.e.* an electron in an atom.

Sciama holds that it is in this specific context that disagreement regarding the vacuum field has occurred among physicists.[110] He phrases the issue in a slightly different way: instead of asking what *field* (radiation or vacuum field) is responsible for the transition through playing the part of $\mathscr{F}_{f0}$, he asks what *process* is: stimulated emission or spontaneous emission. In fact the two approaches are similar: "stimulated emission" is "a reaction to spontaneous radiation emitted by the atom"

---

[109] [94], pp.148-150. This is what is usually called spontaneous emission, but Sciama prefers avoiding the phrase because whether it is appropriate is precisely what is at stake.

[110] This is likely due to the fact that this was the effect specifically treated by Jaynes to criticise the notion that electromagnetic vacuum fluctuations are real, as we shall discuss shortly.

whereas so-called "spontaneous emission" is thought stimulated by the vacuum field. Sciama argues that the relative contributions of stimulated and spontaneous emission to the transition rate are "precisely equal" and "each contribution is physically real."

He further argues that the processes occur even for an atom in its ground state. In this case he speaks of emission *vs absorption* rather than in terms of stimulated *vs spontaneous* emission.[111] The way he visualizes the situation in this context suggests he has the FDT or related considerations in mind: he speaks of the atom "radiating to and absorbing energy from the zero-point fluctuations of the vacuum electromagnetic field", and further states:

> The radiation rate is determined by the noise power in the zero-point fluctuations of the atomic dipole moment, and the absorption rate is determined by the noise power in the zero-point fluctuations of the electromagnetic field.[112]

What interests us are his reasons for thinking that both stimulated and spontaneous processes (hence both the vacuum and the radiation reaction fields) make "precisely equal" contributions to the effect. Although he thinks of the physical situation in terms of a system in a bath, and certainly has the FDT in mind, his argument does not *rely* on the FDT but on considerations related to operator ordering, *i.e.* that symmetric ordering is the only ordering that involves only Hermitian operators.[113]

---

[111] The switch in language naturally corresponds to the fact that what the processes now need to account for is no longer an emission, hence a net loss of energy by the atom, but the lack thereof, hence a dynamic equilibrium of loss and gain.

[112] [94], p.149. Sciama's phrasing suggests that he thinks of the bath to consist of the vacuum field alone, but this is doubtless because he is now talking about an atom in its ground state, which does not emit radiation.

[113] [94], p.148.

J. Dalibard J. Dupont-Roc and C. Cohen-Tannoudji Much of Sciama's conclusions are motivated by the work of J. Dalibard, J. Dupont-Roc and C. Cohen-Tannoudji, in which the latter argue that symmetric ordering is more physically correct than other orderings on the basis of the Hermitian character it implies for the operators. Besides this consideration however, they have other reasons for preferring symmetric ordering, which are related to the FDT. Indeed they argue that when choosing symmetric ordering, one finds physically intuitive results regarding properties of the system S and reservoir R — *i.e.* the bath.[114]

Their treatment is meant to apply even more generally than the FDT in so far that it is valid for "an arbitrary state of the reservoir", rather than requiring that it be at thermal equilibrium.[115] At the same time, the discussion is given with the Lamb shift in mind. They derive expressions for the energy shift of the system due to the reservoir, $(\delta E_a)_{rf}$, and due to self-reaction, $(\delta E_a)_{sr}$.[116] They find, respectively:

$$(\delta E_a)_{rf} = -\frac{1}{2}\sum_{ij}\int_{-\infty}^{+\infty} d\tau \;\; C_{ij}^{(R)}(\tau)\;\chi_{ij}^{(S)}(\tau) \tag{5.103}$$

$$(\delta E_a)_{sr} = -\frac{1}{2}\sum_{ij}\int_{-\infty}^{+\infty} d\tau \;\; \chi_{ij}^{(R)}(\tau)\;C_{ij}^{(S)}(\tau), \tag{5.104}$$

where:

- the C_{ij} refer to correlation functions, that describe the dynamics of fluctuations in

---

[114] [84], p.1626. They ask: "Is it possible to understand the evolution of S [the system] as being due to the effect of the reservoir fluctuations acting upon S, or should we invoke a kind of self-reaction, S perturbing R which reacts back on S?" and note that "For spontaneous emission, [...] the vacuum field, with its infinite number of modes, plays the role of R."

[115] [84], p.13.

[116] More generally, they consider the average rate of the system variable G, $\left\langle \frac{dG(t)}{dt} \right\rangle_R$ where the subscript R indicates that the average is taken over the states of the reservoir. They identify the contributions that the vacuum field and self reaction make to this rate — respectively represented by $\left\langle \left(\frac{dG(t)}{dt} \right)_{vf} \right\rangle_R$ and $\left\langle \left(\frac{dG(t)}{dt} \right)_{sr} \right\rangle_R$.[117]

either the reservoir $C_{ij}^{(R)}$ or the system $C_{ij}^{(S)}$,

- the χ_{ij} refer to linear susceptibilities, and represent the response of the system $\chi_{ij}^{(S)}$ or reservoir $\chi_{ij}^{(R)}$ to a perturbation.

They stress that these results call for a simple physical interpretation:

> One can consider that the fluctuations of R [the reservoir], characterized by $C_{ij}^{(R)}(\tau)$, polarize S [the system] which responds to this perturbation in a way characterized by $\chi_{ij}^{(S)}(\tau)$. The interaction of the fluctuations of R with the polarization to which they give rise in S has a non zero value because of the correlations which exist between the fluctuations of R and the induced polarization in S. [...] The same comments can be made [for Eq. (5.104)] as for [Eq. (5.103)], the roles of [the system] and [reservoir] being interchanged.[118]

They summarize this physical model by the diagram:

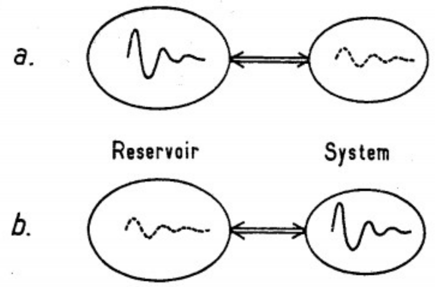


Fig. 2. — Physical pictures for the effect of reservoir fluctuations and self reaction. *a)* Reservoir fluctuations : the reservoir fluctuates and interacts with the polarization induced in the small system. *b)* Self reaction : the small system fluctuates and polarizes the reservoir which reacts back on the small system.

Fig. 5.2: Interaction between system and reservoir[119]

---

[118] [84], p.12.

[119] Figure borrowed from [84], p.12.

The FDT does not imply the existence of vacuum fluctuations

Edwin Jaynes Edwin T. Jaynes argued that effects such as the Lamb shift and spontaneous emission could be derived within what he termed “neoclassical theory”, which meant that matter was indeed quantized but the electromagnetic field was not, in obvious contrast with QED.[120] He did interpret the radiation reaction field as an operator, but as “an operator not on the “Maxwell Hilbert space” of a quantized field, but on the “Dirac Hilbert space” of the electrons”, and he saw no necessity for the “Maxwell Hilbert space” to account for the effects of interest.[121]

Unlike the physicists we just discussed, Jaynes insisted that the FDT does not imply the existence of vacuum fluctuations [97]. In a provocative talk given at the Conference on Coherence and Quantum Optics in 1977, he stated:

> This independence of the initial ordering is, then, just a very simple, general, and elegant fluctuation-dissipation theorem; but let me suggest a different physical interpretation from the usual one. This complete interchangeability of source-field effects and vacuum-fluctuation effects does not show that vacuum fluctuations are “real”. It shows that source field effects are the same as if vacuum fluctuations were present.
>
> For many years, starting with Einstein’s relation between diffusion coefficient and mobility, theoreticians have been discovering a steady stream of close mathematical connections between stochastic problems and dynamical problems. It has taken us a long time to recognize that QED was just another example of this.

---

[120] This made it a semi-classical theory. What is particularly interesting about Jaynes’ neoclassical theory is what motivated it. Peter Milonni has described it as “based on the recognition that rather few phenomena in non-relativistic radiation theory actually require field quantization for their explanation, and its purpose was to explore just how far one could get without field quantization and possibly to point the way to alternatives to QED.” [87], p.148.

[121] [97], pp.10, 18.

> But in another sense, I do have to concede that vacuum fluctuations are, after all, "very real things".[122]

Jaynes only clarified these contradictory statements by discussing an example — precisely, the case of a spontaneously radiating atom. From it he eventually concluded:

> The radiating atom is indeed interacting with an EM field of the intensity predicted by the zero-point energy; but this is just the atom's own radiation reaction field.[123]

Jaynes essentially argued that the presence in Callen and Welton's result of the expression $\frac{\hbar\omega^3}{2\pi^2 c^3}$ should not be interpreted as evidence for the existence of vacuum fluctuations (Eq.(5.101)). Central to his argument is the idea that what is physically significant is not the *spectral* energy density $\rho(\omega)$, but $\rho(\omega)$ integrated over the appropriate frequencies, *i.e.* the energy density, W_f. Note that in Callen and Welton's result, $\rho(\omega)$ appears in an integral over all frequencies in the expression for the energy density, which comes directly from $\mathscr{F}_f$. Jaynes' reasoning consists in saying that not all of these frequencies are physically present. He instead considered $\rho(\omega)$ integrated only over the small frequency bandwidth $\Delta\omega$, which he deemed "effective in causing the atom to radiate."[124]

Jaynes' argument involves deriving and comparing two expressions:

- W_{0eff}, the "effective" part of the energy density whose spectral energy is $\rho_o(\omega)$,
- W_{RReff}, the energy density of the radiation reaction field.

He found them to be given by the same expression:

$$\frac{1}{18\pi}\mu^2\left(\frac{\omega}{c}\right)^6, \tag{5.105}$$

---

[122] [97], p.12.

[123] [97], p.13.

[124] [97], p.12.

where μ stands for the electric dipole moment of the relevant atomic transition, and ω for the natural line frequency of this transition.[125]

Jaynes' derivation shows that if one takes into account only some field modes and components (deemed physically significant), the energy density due to "$\rho_o(\omega)$", W_{0eff}, is given by the same expression as the energy density of the radiation field at the position of the atom, W_{RR}. From this he concluded that what had been interpreted as a contribution of the vacuum field is in fact due to radiation reaction.[126]

At this juncture it may be useful to ask what is meant by vacuum field in the context of radiating atoms and accelerating charges. Recall that strictly speaking, the vacuum field is the electromagnetic field in its ground state, free of (real) particles. But in fact, as soon as a charge accelerates, it is bathed in its own radiation field, the real photons it is emitting. Phrases such as "vacuum fluctuations" and "zero-point (vacuum) field" can no longer refer *per se* to a vacuum surrounding the particle, since the electromagnetic quantum field is no longer in its vacuum state. So how can we make sense of the phrase "vacuum field" in this context?

It seems fair to interpret these expressions as short-hand for "fluctuations above and below the (no-longer-zero-point) energy of the electromagnetic field." What can justify this use is that these fluctuations can be thought of as "the same ones" as would be there even if there were no real photons around, in the sense that they too result from the uncertainty relations applied to the field.[127] So when one speaks of the contribution of the "vacuum field" in such a situation, one is referring to the contributions to the electromagnetic field that are due to the uncertainty relations.

---

[125] Jaynes' notation is not entirely consistent, as he first uses ω_0 to represent the natural line frequency, and simply ω in his final expressions for the two energy densities. However his derivation and reasoning require that he means the natural line frequency by this.

[126] This is most certainly the "other sense" in which he meant that vacuum fluctuations are "very real things", *i.e.* they are in fact the radiation reaction field.

[127] More precisely, they result from commutation relations between field operators.

And indeed, this is what makes the whole topic so fascinating: it involves possible implications of the uncertainty relations. Because vacuum fluctuations are the result of the latter, they would be there whether or not there is radiation present, *i.e.* whether or not a charge has ever been accelerating. Essentially, the question Jaynes could be asking is: if we first imagine the universe completely empty of real particles (especially photons and accelerating charges), and we put in it an atom in an excited state, is this atom going to emit a photon? Would "spontaneous emission" take place, even then? This is what he means when he asks whether it is the vacuum field that is responsible for the effect.[128]

Quite obviously Jaynes believes that the answer is "no, an excited atom placed in a universe free of real photons would not emit a photon." It seems clear that his reason for thinking so is that, since the modes responsible for the emission are of the same frequencies as the radiation emitted when the atom does, it is a more economical hypothesis to suppose that these modes in fact belong to the field we know for a fact exists anyway. However for Jaynes this economy goes beyond a concern with Occam's razor. It does away with a crucial difficulty, *i.e.* the infinities known to plague QED; no vacuum field, no infinities:

> The fantastic numbers noted before disappear as soon as we realize that, in order to account for spontaneous emission, there is no need for this energy density to be present in all space, at all times, in all frequency bands. It is produced automatically by the radiating atom, but in a more economical way; only the field component that is needed, where

---

[128] This also explains why the issue of what field is responsible has been described in terms of whether the emission process is spontaneous or stimulated. If the modes involved are only those radiated by the atom, they are only a consequence of the process, not a cause of it, and it is truly spontaneous. By contrast if emission is caused by the vacuum field it is of course stimulated by the latter.

> it is needed, when it is needed, and in the frequency band needed ([97], p.14).[129]

Hence Jaynes interpreted the FDT differently from the physicists discussed earlier. He thought that the bath consists of only one field, the radiation field, alone responsible for $\mathscr{F}_f$. His view was not motivated by an independent understanding of the FDT, but by the other considerations just discussed, which then led to his interpretation of the FDT.

5.4 Consistency of the quantum theory of radiation: necessity of the vacuum field for the commutation relations to hold.

In addition to arguments in favor of a preferred ordering, and considerations relating to the FDT, other evidence pertaining to the vacuum field involves the position-momentum commutation relation. Indeed a very interesting demonstration consists in showing that for $[\mathbf{X}, \mathbf{P}] = i\hbar$ to hold for an accelerating charge — which of course undergoes radiation reaction damping in the process —, the charge needs to be driven by the vacuum field. And not only are both vacuum and radiation fields needed, they also need to be related to one another in a specific way.[130]

---

[129] An interesting aspect of Jaynes' demonstration is worth noting: $\hbar$ cancels out in his derivation of W_{0eff} — as evidenced by the final result. This occurs because his expression for the bandwidth $\Delta\omega$ is proportional to $\hbar^{-1}$. This in turn comes about because he finds $\Delta\omega$ to be proportional to the Einstein A-coefficient (*i.e.* the coefficient of spontaneous emission, A_{21} is the probability per unit time that an atom in energy state 2 will spontaneously emit a photon and undergo a transition to energy state 1) and the latter is proportional to $\hbar^{-1}$. Indeed, Jaynes reasons that although the spectral density $I(\omega)$ of the radiation spontaneously emitted by the atom has a Lorentzian profile, and therefore no well-defined width, $I(\omega_0)\Delta\omega$ must be equal to the total energy radiated by the atom, $\int I(\omega)d\omega$, and the latter naturally involves the A-coefficient. This line of reasoning assumes that the bandwidth $\Delta\omega$, which refers to the frequencies of the fluctuating field causing the atom to radiate, can be identified with the width of the natural emission line of the atom. Jaynes does not strongly defend this position, merely stating that this is "presumably" the case. Perhaps he felt that his final result, *i.e.* the identity of expressions for W_{0eff} and $\mathscr{W}_{RR}$, was remarkable enough to lend credence to the assumptions he had to make in order to reach it.

[130] [90], p.106; [98], pp.1322, 1323; [99].

Peter Milonni showed this notably by considering a non-relativistic electron in free space. The Heisenberg equation of motion for the electron is:

$$\frac{d^2\mathbf{X}}{dt^2} - \frac{2e^2}{3mc^3}\frac{d^3\mathbf{X}}{dt^3} = \frac{e}{m}\mathbf{E}_0,^{131} \tag{5.106}$$

where we recognize the second term as the radiation reaction force.

If this equation is solved for $\mathbf{X}$, and we then use it to find as well $\mathbf{P} = m\dot{\mathbf{X}}$, it turns out that the commutator of these operators takes the form:

$$[\mathbf{X}, \mathbf{P}] = \frac{8\pi^2 i}{3m}\int_0^\infty d\omega \frac{\rho_0(\omega)}{[\omega^3(1+\gamma^2\omega^2)]}, \tag{5.107}$$

where $\gamma = \frac{2e^2}{3mc^3}$ and $\rho_0(\omega)$ is the spectral density of the vacuum field. When we then substitute $\rho_0(\omega) = \frac{\hbar\omega^3}{2\pi^2c^3}$ in this expression, we find the commutator equal to $i\hbar$ as it should be.

The crucial point is this: had we neglected the vacuum field in Eq. (5.106) by setting the RHS to zero, $[\mathbf{X}, \mathbf{P}]$ would not be equal to $i\hbar$. Nor would it be if the field responsible for the driving force had a spectral energy density different from $\frac{\hbar\omega^3}{2\pi^2c^3}$. Notably, the energy spectrum needs to go as the *third* power of the frequency ω because the radiation reaction force involves the *third* derivative of the position.[132]

---

[131] Note that E_0 here is the homogeneous solution of the Maxwell (Heisenberg) equation for the electric field.

[132] The reverse is also true: had we instead neglected the radiation reaction effect in Eq. (5.106) by setting the second term to zero, $[\mathbf{X}, \mathbf{P}]$ would not be equal to $i\hbar$ either.
Milonni also gives elsewhere an analogous treatment for the case of a dipole oscillator.[133] There he uses the small-damping approximation $\ddot{\mathbf{x}} \cong -\omega_0^2\mathbf{x}(t)$ in the equation for the Heisenberg-picture position operator $\mathbf{x}(t)$:

$$\ddot{\mathbf{x}} + \omega_0^2\mathbf{x} - \tau\dddot{x} = \frac{e}{m}\mathbf{E}_0(t) \tag{5.108}$$

thereby obtaining:

$$\ddot{\mathbf{x}} + \tau\omega_0^2\dot{\mathbf{x}} + \omega_0^2\mathbf{x} \cong \frac{e}{m}\mathbf{E}_0(t).^{134} \tag{5.109}$$

As he notes:

One may perhaps wonder if there isn't some circularity at play in this argument. After all, the consistency with quantum mechanics at stake here is the commutation relation that formally expresses the uncertainty relations, and it is again the uncertainty relations that lead to the prediction of vacuum fluctuations in the first place. Why should we now regard another *formal* argument based on the *same* "principle",[135] as additional and (as Milonni implies) even stronger evidence for vacuum fluctuations? However the commutation relation at stake here is not the one that gives rise to the zero point energy of the field in QFT: it is the NRQM commutation relation for the charge. Now, one can argue that we are still using the same "principle", even though it now refers to different variables — position and momentum, *vs* field and canonical momentum. However there is an important difference between these two sets of variables, which may justify regarding the NRQM uncertainty relations to be on a stronger footing than its QFT counterpart: quantum field and canonical momentum are not observable, in so far that they have zero expectation value.

5.5 Discussion

Do the arguments provided by Dalibard *et al* and Peter Milonni invalidate Jayne's view? There are at least two separate issues at stake: first, whether the ordering of operators can still be said to lead to underdetermination regarding which field is responsible for the effects of interest, and second, when we interpret an expression

---

> Without the free field $E_0(t)$ in this equation the *operator* $x(t)$ would be exponentially damped, and commutators like $[z(t), p_z(t)]$ would approach zero for $t >> (\tau\omega_o^2)^{-1}$. With the vacuum field included, however, the commutator is $i\hbar$ at all times. ([87], pp.53-54.)

In other words, the presence of $E_0(t)$ is required for the variables that describe the dipole to take on their quantum character: without this field, their commutator is no longer proportional to $\hbar$; instead it vanishes so these variables would then commute as they do classically.

[135] In so far that the uncertainty relations are often described as a principle.

as being a signature of the vacuum field, are we sure that it is indeed the vacuum field which is at stake?

5.5.1 Operator ordering

As explained above, the ambiguity regarding whether the vacuum field is definitely involved in the various relevant effects can be said to arise from issues regarding the commuting properties of operators, hence our freedom or lack thereof in ordering them. In so far that operators commute we get similar results in terms of numerical predictions, in so far that they don't we get different accounts of which field is responsible for the effects.[136]

This underdetermination does not allow for the *source field* to play no part in all effects, because in spontaneous emission all orderings involve at least some contribution from it. However, in and of itself, the ordering freedom does leave open the possibility that the *vacuum field* may play no part — hence that these effects provide no evidence for its existence.

As explained above, it has been argued that this underdetermination is only superficial, because symmetric ordering should be preferred to all other choices, and it involves equal contributions from the vacuum field as well as the source field. The proponents of symmetric ordering argue that it should be preferred because only with this ordering do we find that:

- the vacuum field contribution to the effect, and the source field contribution to it, are represented by Hermitian operators.
- throughout the derivation, operators for the system and for the fields are Hermitian operators (only products of Hermitian operators are involved, not products of,

---

[136] Recall that it is the operators for the system and the *total* field that commute, so whatever way we choose to order them, we get the same final, overall results, but that the operators for the vacuum and source field do not separately commute with operators for the system, so that different choices of ordering involve different contributions from the vacuum and source fields.

say, creation operators).

These two considerations are then elevated to the status of requirements to be imposed on derivations in order for them to make physical sense. The argument is that because in NRQM observables are represented by Hermitian operators, demanding that the operators be Hermitian throughout the calculation ensures that they have physical meaning throughout.

It should be noted that there may be some circularity in the reasoning involved, in so far that these two "requirements" do not appear as desiderata purely for their own sake, with lifting the underdetermination a mere fortunate consequence of them. In fact the latter, *i.e.* removing the ambiguity, also functions as an argument in favor of preferring one ordering over the others, and motivates seeking criteria to do so. Whether or not operators are Hermitian turns out to depend on the ordering, and has been deemed a good criterion to go by.

What can perhaps be argued to be a fortunate consequence of these requirements is the extent to which they lead to an intuitive, classical description of the relationship between system and vacuum field, whereby they can be said to polarize one another on the basis of Eq. (5.103) and Eq. (5.104).

However, not all researchers agree that these considerations justify regarding symmetric ordering as *correct*, or simply *more correct*, than the alternative choices. Ultimately, the difference between the different view points lies in the ontological status one chooses to ascribe to the positive (or negative) frequency part of a field. The physicists who argue for the superiority of symmetric orderings state:

> It would be difficult to elaborate a physical picture from an expression involving only the positive frequency part of the field which is not observable ([84], p.9).

Others, notably Peter Milonni, frame the issue in terms of the distinction between quantum and classical features, and refrain from ascribing greater import to the latter:

> Various [...] orderings give different weights to the vacuum and source fields when we try to interpret the results of a calculation. To emphasize as much as possible the classical-like aspects of the vacuum and source fields, we can choose a symmetric ordering at every stage of a calculation. ([90], p.7)

Taking this view, one is then led to describe the phenomena of interest as due, say, to only the positive frequency part of the field when normal ordering is chosen, and this physical picture should be taken as seriously as the one described by normal ordering.[137]

Also, one can contrast Dalibard *et al*'s concern with restricting the derivation to observables in the formal sense of the word, with Jaynes' operationalist motivations for doing away with the vacuum field, *i.e.* the at best indirect character of its experimental observability. Jaynes states:

> It seems to me that, if you say radiation is "real," you ought to mean by that, that it can be detected by a real detector. But an optical pyrometer sees only the Planck term, and not the zero-point term, in black body-radiation.[138]

---

[137] Normal ordering corresponds to:

$$E^{(-)}P + PE^{(+)}, \tag{5.110}$$

and $E^{(-)}$ acting on the left is equivalent to $E^{(+)}$ on the right.
It is worth noting that despite these differences of opinion, Peter Milonni actually does think that the vacuum field exists as well as the source field — simply he does not think so on the basis of the arguments that favor symmetric ordering.

[138] [97], pp.5-6. For interest's sake, Jaynes goes on: "Of course, a staunch defender of present

One gets the impression that ultimately, for Dalibard *et al* and Sciama, it is the intuitive description of the relationship between system and vacuum field that convinces them of the correctness of symmetric ordering. It is arguably in this sense that the FDT can be said to provide a strong argument in favor of the existence of vacuum fluctuations, for it provides a classical framework in which Dalibard *et al* can easily interpret their results.

5.5.2 Are we really dealing with the vacuum field?

A key point of Jaynes' discussion was that one could easily be mistaken in identifying the vacuum field. Is this point possibly relevant to the arguments put forward by Dalibard *et al* and Milonni?

As shown above, Jaynes himself used the accepted expression for the spectral energy of the vacuum, "$\rho_o(\omega)$" *i.e.* $\frac{\hbar\omega^3}{2\pi^2 c^3}$. His work stressed that the mere appearance of this expression in a result does not constitute evidence for the existence of the vacuum field. It seems fair to say that he agreed this expression represented the energy density that the vacuum field would indeed have per mode, if these modes did exist. He simply saw no reason in the appearance of this expression to assert that they do, in fact, exist. In his derivation, the piece of formalism relevant to their existence or lack thereof is the range of the frequencies one integrates $\rho_o(\omega)$ over. The vacuum field presumably involves an infinite range of frequencies, whereas the emitted radiation involves only the frequencies emitted by the atom. To account for

---

theory will say immediately that such objections reflect only naive metaphysical preconceptions of reality, not unlike pre-relativistic notions of absolute simultaneity, of just the kind that the Copenhagen interpretation of quantum theory has recognized and rightly removed from science.
...
It is a supple ontology which supposes that vacuum fluctuations are just real enough to shift the hydrogen 2s level by 4 microvolts; but not real enough to be seen by our eyes, although in the optical band they correspond to a flux of over 100 kilowatts/cm^2. Nevertheless, the dark-adapted eye, looking for example at a faint star, can see <u>real</u> radiation of the order of 10^{-15} watts/cm^2."

observable facts, Jaynes argued, the appropriate range of frequencies is the latter. This did not show that the vacuum field could not be responsible, but it did show that its existence was not required.

Now Milonni's argument involving commutation relations certainly relies on identifying $\frac{\hbar\omega^3}{2\pi^2c^3}$ as the spectral density of the vacuum field. So does Jaynes' point (*i.e.* that the field at stake may in fact not be the vacuum field) invalidate Milonni's argument?

The expression used by Milonni involves an integral over all frequencies, not simply some of them, and his result relies on performing this integral.[139] So his result does depend on the range of frequencies he considers. This suggests that Jaynes' point does not affect Milonni's result.

It does not seem to affect Dalibard *et al.*'s arguments either, although matters are less straightforward in this case: the spectral energy density $\frac{\hbar\omega^3}{2\pi^2c^3}$ does not play the role of an operational definition of the vacuum in their work. Instead, they define the vacuum field via the homogeneous solution to the field equations, *i.e.* the "free field." Now in classical electrodynamics such solutions exist and in no way correspond to what we mean by vacuum field: in that context, all fields have at some point in the past been emitted by some source. So considering the homogeneous solution alone is obviously no guarantee that we are dealing with the vacuum field. What makes it so is to add the condition that "no [real] photons are initially present",[140] which they impose through the requirement that "the radiation field is in the vacuum state at the initial time."[141] Hence it would appear that Jaynes' worry regarding the identification of the vacuum field does not apply to the work of Dalibard *et al* either.

---

[139] And not just on terms cancelling out in the integrand for instance.

[140] [84], p.1618.

[141] [84], p.1622.

5.6 Conclusions

The effects traditionally ascribed to vacuum fluctuations can also be ascribed to a source field, and we showed how this underdetermination comes about: the operator relative to the system commutes with the operator for the total field (vacuum + source), hence derivations yield the same results irrespective of the ordering chosen; however the operator for the total field does not commute with those for the source field nor for the vacuum field, so different orderings yield different relative contributions from these two fields. We saw that the ambiguity can be lifted by demanding that only Hermitian operators be used throughout the derivation, which corresponds to choosing symmetric ordering, and implies that both fields contribute — hence that the vacuum field has an observable effect. This demand is justified on the ground that symmetric ordering is more physical, in so far that it involves Hermitian operators throughout the calculations. However, it implies regarding the positive (or negative) frequency part of the field as unphysical.

At the same time, physicists have often appealed to the fluctuation-dissipation theorem when discussing the relative contributions of source and vacuum fields. Although the FDT paints a clear intuitive physical picture in the high temperature, classical limit, its interpretation in the $T = 0$ limit of zero-point contributions is not as straightforward. If we take it to mean that the fluctuating force it describes is exerted by the same entity as the dissipative force, then we should conclude that it is due to a source field. More often than not however, the FDT has been thought to suggest that the fluctuating force is due to the vacuum field. Either way the FDT requires the presence of a source field for dissipative effects to take place, and what we mean by vacuum field in such situation requires some care: in QFT the phrase refers to the electromagnetic field being in its ground state. The very presence of a source field precludes this, so what is at stake are then fluctuations in the (excited) field having the same origin as the zero-point energy of the vacuum field — *i.e.* the uncertainty relations.

Most physicists have interpreted the FDT as evidence for contributions from vacuum fluctuations. But this view has been heavily criticized at some point, notably by Edwin Jaynes, who interpreted the FDT to "show that source field effects are the same as if vacuum fluctuations were present" — *i.e.* that the fluctuating force is exerted by the same field as the dissipative one. He argued that people incorrectly assumed vacuum fluctuations to be responsible because the expression for the spectral energy density of the vacuum field appeared in their result, when in fact this expression could just as well be due to the source field.

However Jaynes' warning against unduly interpreting some expressions as representing vacuum field effects does not hold for the research involving operator ordering. Nor does it seem to undermine an argument in favor of vacuum field effects regarding the consistency of the quantum theory of radiation: taking the vacuum field into account in the Heisenberg equation of motion for an accelerating charge is necessary for the position-momentum commutation relation to hold.

Therefore, at this point it seems that the vacuum field does contribute to the physical effects in question.

If interpreting virtual particles as originating from the uncertainty relations, this would suggest that, in this context at least, the uncertainty relations should be interpreted in more than a minimal, epistemic sense, but in an ontological one as well.

Unlike the phenomena discussed in the present chapter, all accounts of the Scharnhorst effect are based on radiative corrections or the zero-point field. Because it has been related to the fact that the Casimir vacuum has lower energy density than the trivial vacuum, describing the Scharnhorst effect in terms of source fields may seem unlikely, in so far that it is a priori unclear how source fields could lead to this decrease in energy density. Yet as we have seen the two models are formally related. In the next chapter, we attempt to derive the Scharnhorst effect from source fields.

Chapter 6

THE SCHARNHORST EFFECT IN SOURCE THEORY

6.1 Introduction

As discussed in the previous chapter, the phenomena typically attributed to vacuum fluctuations (Lamb shift, spontaneous emission, Van der Waals forces, Casimir effect) can also be interpreted as due to source fields. Existing derivations of the Scharnhorst effect involve vacuum fluctuations as well. Recall that the effect that bears his name was first derived by Klaus Scharnhorst in 1990, and strictly speaking, consists in the prediction that photons propagating in a Casimir vacuum would do so with a phase and group velocities greater than their trivial vacuum value of c. In Scharnhorst's derivation the effect is due to the propagators of the off-shell (*i.e.* "virtual") photons being affect by the plates in the two processes below:

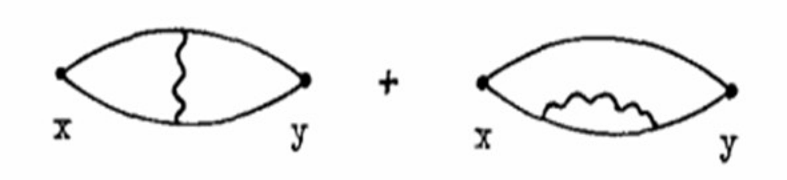


Figure 6.1: The two-loop processes responsible for the Scharnhorst effect.[1]

Barton's rederivation involves quantum expectation values of the electromagnetic field (the zero-point field), and is due to vacuum polarization. And generally, the derivations of the Scharnhorst effect have involved the Euler-Heisenberg Lagrangian. As discussed in chapter 4, the latter is the basis of an effective quantum field theory that involves only photon fields and models interactions by contact terms between them:

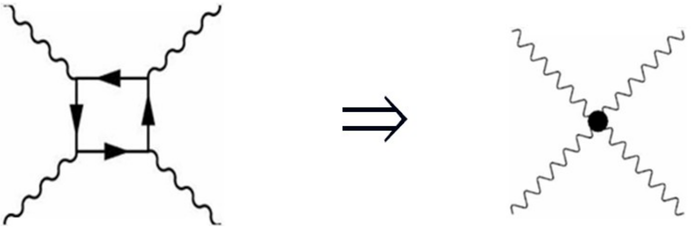


Figure 6.2: The box diagram in QED becomes a contact term in the theory described by $\mathfrak{L}_{EH}$.

In the Scharnhorst effect, the relevant processes arise from the "tadpole diagram", in which the loop represents an off-shell photon:

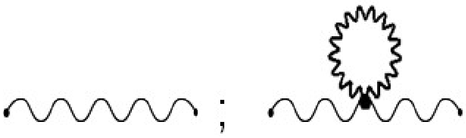


Figure 6.3: Feynman diagrams in the theory described by the standard Euler-Heisenberg Lagrangian. The diagram on the right accounts for interactions, and is known as the "tadpole diagram."

So the various derivations of the Scharnhorst effect involve off-shell, so-called "virtual" particles, which as discussed in the previous chapter are often deemed "vacuum fluctuations." Since the other physical effects involving the latter have also been accounted for by source fields, it seems reasonable to wonder whether the Scharnhorst effect too can be recovered in this way. This is what I shall attempt to do in the present chapter.

The general structure of my calculation follows Barton's quite closely; indeed source fields can easily be substituted for the zero point fields in his derivation. Furthermore, the way in which these source fields are introduced is more closely inspired by the work of Peter Milonni on the Casimir effect than by that of Julian Schwinger, Lester DeRaad and Kimball Milton [17], [100]. The approach adopted by the latter is entirely classical: they do not quantize the electromagnetic fields. In contrast, Peter Milonni does quantize them, and isolates the contribution of the source fields by normal ordering the relevant operators. This latter approach meshes more naturally with Barton's work, where the background electromagnetic fields in which the Scharnhorst photons travel are of course quantized, being zero-point fields, so this is the approach we shall follow.

6.2 General form of the derivation

Following Barton we shall model the background quantum vacuum field between the Casimir plates by quantized electromagnetic fields $\mathbf{E}$ and $\mathbf{B}$, and the Scharnhorst photons that propagate in this region by classical fields $\mathbf{e}$ and $\mathbf{b}$. The aim of the derivation is to obtain an expression for the difference in the index of refraction in the Casimir vacuum, compared to normal vacuum. This requires finding the susceptibility (ϵ_{ij}) and permeability tensors (μ_{ij}) of the Casimir vacuum, which respectively depend on its polarization (p_i) and magnetization (m_i). Using Maxwell's equations derived from an appropriate Lagrangian, expressions for p_i and m_i can be obtained in terms of the classical fields $\mathbf{e}$ and $\mathbf{b}$ that model the Scharnhorst photons and of the quantized background electric and magnetic fields E_i and B_i. More specifically these expressions involve the vacuum expectation values of products of these fields, *i.e.* $\langle E_i E_j \rangle$, $\langle B_i B_j \rangle$ and $\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle$.

It is in the way these expectation values are evaluated that the present work differs from Barton's. Whereas E_i and B_i represent the zero-point fields in Barton's derivation, here they will stand for source fields, due to fluctuations in polarization within the plates.

Before we focus on the derivation of these expectation values in the framework of source theory – which, again, is where the current work differs from Barton's – let us first present the overall calculation, that they have in common. We begin by discussing the Euler-Heisenberg effective Lagrangian, and then we shall examine how the phase velocity is derived from this Lagrangian: how the relevant equations of motion are derived, how expressions for the polarization and magnetization of the background electromagnetic field are obtained, and from them the index of refraction and phase velocity.

6.2.1 The Euler-Heisenberg effective Lagrangian

Barton's work is based on the Euler-Heisenberg effective Lagrangian, which describes virtual fermions (specifically electrons and positrons) in an electromagnetic field.

Historically, the development of this Lagrangian played a crucial role in the development of effective field theory. It is said to have "produced the paradigm for the entire field of effective Lagrangians."[2] In 1936, Werner Heisenberg and Hans Heinrich Euler on the one hand, and Victor Weisskopf on the other, considered the relevant one-loop effective action (Heisenberg and Euler used a spinor field to describe the

---

[2] [101], p.4, [102], p.448.

fermions, Weisskopf a scalar field):

$$S^{(1)}_{\text{spinor}} = -i \ln\det(i - m) = -\frac{i}{2}\ln\det({}^2+m^2) \tag{6.1a}$$

$$S^{(1)}_{\text{scalar}} = \frac{i}{2}\ \ln\det\left(D_\mu^2 + m^2\right) \tag{6.1b}$$

where the is the Dirac operator, $= \gamma^\nu\left(\partial_\nu + ieA_\nu\right)$ and m is the electron mass. $S^{(1)}$ can be expanded in powers of A_μ, the photon field that represents the background electromagnetic field. In terms of Feynman diagrams, this perturbative expansion is represented as:

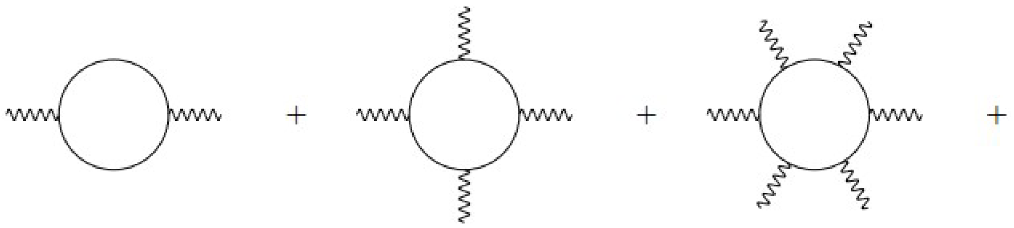


Fig. 6.1: Diagrammatic perturbative expansion of the one-loop effective action.

The wiggly lines represent the external photons — *i.e.* the real background electromagnetic field — and the loops stand for the electron/positron field they interact with. Euler, Heisenberg and Weisskopf studied this expansion in the low energy limit of the photons, which corresponds to the strength of the electromagnetic field, $F_{\mu\nu}$, being nearly constant. In this case, the effective action can be found in closed form:[3]

$$\begin{aligned}
\mathcal{L}^{(1)}_{\text{spinor}} &= \frac{1}{hc}\int_0^{\infty}\frac{d\eta}{\eta^3}e^{-\eta e\mathcal{E}_c}\Bigg\{ie^2\eta^2(\vec{E}.\vec{B})\frac{\left[\cos\left(\eta e\sqrt{\vec{E}^2-\vec{B}^2+2i(\vec{E}.\vec{B})}\right)+\text{c.c.}\right]}{\left[\cos\left(\eta e\sqrt{\vec{E}^2-\vec{B}^2+2i(\vec{E}.\vec{B})}\right)-\text{c.c.}\right]} \\
&\qquad +1+\frac{e^2\eta^2}{3}(\vec{B}^2-\vec{E}^2)\Bigg\} \\
&= \frac{1}{hc}\int_0^{\infty}\frac{d\eta}{\eta^3}\,e^{-\eta\, e\mathcal{E}_c}\left\{-i\,e^2a\,b\,\eta^2\left[\frac{\cosh[(b+ia)e\eta]+\cosh[(b-ia)e\eta]}{\cosh[(b+ia)e\eta]-\cosh[(b-ia)e\eta]}\right]\right. \\
&\qquad \left. +1+\frac{e^2\eta^2}{3}(b^2-a^2)\right\} \\
&= -\frac{1}{hc}\int_0^{\infty}\frac{d\eta}{\eta^3}\,e^{-\eta\, e\mathcal{E}_c}\left\{\frac{e^2a\,b\,\eta^2}{\tanh(eb\eta)\tan(ea\eta)}-1-\frac{e^2\eta^2}{3}(b^2-a^2)\right\}, \qquad (6.2)
\end{aligned}$$

with:[4]

$$a=\sqrt{\sqrt{\mathcal{F}^2+\mathcal{G}^2}-\mathcal{F}}\ , \quad b=\sqrt{\sqrt{\mathcal{F}^2+\mathcal{G}^2}+\mathcal{F}} \qquad (6.3)$$

$$\mathcal{F}\equiv\frac{1}{4}F_{\mu\nu}F^{\mu\nu}\ , \quad \mathcal{G}\ \equiv\frac{1}{4}F_{\mu\nu}\tilde{F}^{\mu\nu}. \qquad (6.4)$$

By expanding this expression, all the perturbative diagrams of the expansion illustrated by Fig. 6.1 can be generated.

Barton based his derivation on the expansion up to quartic order [6]:

$$\begin{aligned}
L &= \frac{1}{8\pi}(\mathbf{E}^2-\mathbf{B}^2)+g\left(\left(\mathbf{E}^2-\mathbf{B}^2\right)^2+7\left(\mathbf{E}\cdot\mathbf{B}\right)^2\right) \\
&\text{where:} \\
g &\equiv \frac{\alpha^2}{(2^3)(3^2)(5\,\pi^2\,m^4)}=\frac{e^4}{(2^3)(3^2)(5\,\pi^2\,m^4)}. \qquad (6.5)
\end{aligned}$$

---

[3] [101], p.5, [102], p.449.

[4] Following the original definitions of Heisenberg and Euler[101], p.5, [102], p.449.

The first term corresponds to the classical electromagnetic field Lagrangian $\frac{1}{8\pi}(\mathbf{E}^2 - \mathbf{B}^2)$, which we shall refer to as L_0. The other two terms represent the correction to the classical lagrangian, ΔL, to first order in g, *i.e.* $\backsim e^4$. This implies a diagram with four vertices, *i.e.* the second diagram in Fig. 6.1:

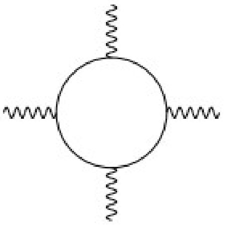


Fig. 6.2: Photon-photon scattering diagram

The physical process this corresponds to is the scattering of a photon by another: two photons come in, annihilate into a virtual electron-positron pair, which then turns into another two photons. The overall result is that two photons have seemed to scattered off one another — though the interaction has been mediated by a virtual electron-positron pair. It is perhaps worth stressing that no such phenomenon occurs in classical electrodynamics: in the context of this theory photons do not scatter off each other; they pass right through one another, ignoring each other completely. QED itself accounts for this feature of the classical theory by standing as a generalization of it: the Lagrangian for classical electrodynamics is the leading order term in the QED action, $\frac{1}{8\pi}(\mathbf{E}^2 - \mathbf{B}^2)$ in Eq. (6.5), so that the theory appears as a low order approximation of QED, unable to account for effects that only appear at higher orders. Therefore in Barton's derivation, the Scharnhorst effect is a truly quantum effect in the sense that it is a prediction done on the basis of a feature of QED absent from its classical counterpart.

Euler and Weisskopf themselves (as well as B. Köckel[5]) interpreted the effect of

---

[5] [103], English translation [104].

photon-photon scattering as conferring to the vacuum dielectric behavior. Weisskopf notably stated:

> When passing through electromagnetic fields, light will behave as if the vacuum, under the action of the fields, were to acquire a dielectric constant different from unity.[6]

This can be understood as the effect of the virtual electron-positron pairs rendering the vacuum polarizable. The latter then behaves like a dielectric. What Barton sets out to do is to calculate its index of refraction. Only the vacuum at stake in this calculation is not the usual vacuum but the Casimir vacuum, and the electromagnetic background field is the zero-point field.

6.2.2 Derivation of the relevant equation of motion for the fields

As discussed in the previous section, the relevant equation of motion is obtained from the Euler-Heisenberg Lagrangian, Eq. (6.5):

$$L \;=\; \frac{1}{8\pi}(\mathbf{E}^2-\mathbf{B}^2)+g\left(\left(\mathbf{E}^2-\mathbf{B}^2\right)^2+7\left(\mathbf{E}\cdot\mathbf{B}\right)^2\right)$$

One proceeds as one would in order to derive the inhomogeneous Maxwell's equations from L_0.[7] Recall that in that case, the electric and magnetic fields are *defined* from the scalar and vector potential fields $\mathbf{\Phi}$ and $\mathbf{A}$:

$$\mathbf{E} \equiv -\dot{A} - \nabla\Phi \tag{6.6}$$

---

[6] [105]. English translation:[106], p.1. Also quoted in [101], p.7, [102], p.451.

[7] $g \equiv \frac{\alpha^2}{2^3 \cdot 3^2 \cdot 5\ \pi^2\ m^4} = \frac{e^4}{2^3 \cdot 3^2 \cdot 5\ \pi^2\ m^4}$.
Recall that the first term corresponds to the classical electromagnetic field Lagrangian $\frac{1}{8\pi}(\mathbf{E}^2-\mathbf{B}^2)$, which we shall refer to as L_0, and the other two terms represent the correction to the classical lagrangian, ΔL.

$$\mathbf{B} \equiv \nabla \times A. \tag{6.7}$$

Because there are no charges between the plates, Eq. (6.6) reduces to:

$$\mathbf{E} \equiv -\dot{A}. \tag{6.8}$$

Φ and $\mathbf{A}$ are taken as generalized coordinates, and varying the action $\int L_0 \, dt$ with respect to them, *i.e.* taking the Euler-Lagrange equations that correspond to them, yields the two inhomogeneous Maxwell's equations.[8]

$$\nabla \cdot \mathbf{E} = 4\pi \; \rho \tag{6.9}$$

$$\nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = 0. \tag{6.10}$$

The equation of motion that we need is the modified form of Eq. (6.10), obtained from $L = L_0 + \Delta L$ instead of from L_0. It is obtained from the Euler-Lagrange equations for the field, using as the generalized coordinates the components of the 3-vector $\mathbf{A}$ (*i.e.* the spatial components of the corresponding 4-vector): [9]

$$\partial_\mu \frac{\partial L}{\partial(\partial_\mu A_\nu)} - \frac{\partial L}{\partial A_\nu} \;\; = \;\; 0. \tag{6.11}$$

L depends on $\mathbf{A}$ through *derivatives* of $\mathbf{A}$: through $\mathbf{E} \equiv -\dot{A}$, and through B_i, as $\nabla \times \mathbf{A} = B_i$, hence through spatial derivatives of $\mathbf{A}$ of the form $\partial_i A_j$ with $i \neq \; j$. L

---

[8] Following Barton, unrationalized gaussian units are used here for the electromagnetic field, together with natural units such that $\hbar = c = 1$.
Also, we are using the Coulomb gauge, so that $\nabla \cdot \mathbf{A} = 0$.

[9] The other inhomogeneous equation is obtained by taking the time component of the 4-vector $\mathbf{A}$, *i.e.* Φ, as the generalized coordinate:

$$\partial_i \frac{\partial L}{\partial(\partial_i \Phi)} - \frac{\partial L}{\partial \Phi} \;\; = \;\; 0$$

does not depend directly on $\mathbf{A}$ itself. Consequently Eq. (6.11) reduces to:

$$\begin{aligned} \partial_\mu \, \frac{\partial L}{\partial(\partial_\mu A_\nu)} &= 0 \\ \partial_\mu \, \frac{\partial L_0}{\partial(\partial_\mu A_\nu)} + \partial_\mu \, \frac{\partial(\Delta L)}{\partial(\partial_\mu A_\nu)} &= 0. \end{aligned} \tag{6.12}$$

Let us first consider the time derivatives:

$$\partial_0 \, \frac{\partial L}{\partial(\partial_0 A_i)} = \partial_0 \, \frac{\partial L_0}{\partial(\partial_0 A_i)} + \partial_0 \, \frac{\partial(\Delta L)}{\partial(\partial_0 A_i)} \tag{6.13}$$

L only depends on Φ through E, as $E_i = -\partial_i \Phi$, so this becomes:

$$\begin{aligned} \partial_i \, \frac{\partial L}{\partial(\partial_i \Phi)} &= 0 \\ \partial_i \, \frac{\partial L_0}{\partial(-E_i)} + \partial_i \, \frac{\partial(\Delta L)}{\partial(-E_i)} &= 0 \\ \frac{1}{8\pi}\partial_i \, \frac{\partial((\mathbf{E}^2 - \mathbf{B}^2))}{\partial(E_i)} + \partial_i \, \frac{\partial(\Delta L)}{\partial(E_i)} &= 0 \\ \frac{1}{8\pi} \, \partial_i \, (2E_i) + \partial_i \, \frac{\partial(\Delta L)}{\partial(E_i)} &= 0 \\ \partial_i \, \left(\frac{1}{4\pi} \, E_i + \, \frac{\partial(\Delta L)}{\partial(E_i)} \right) &= 0 \end{aligned}$$

so that:

$$\nabla \cdot \left(\frac{1}{4\pi} \, \mathbf{E} \, + \, \mathbf{P} \right) = 0, \ \text{ with: } \mathbf{P} \equiv \frac{\delta \Delta L}{\delta \mathbf{E}} = 4g \left(\mathbf{E}^2 - \mathbf{B}^2 \right) \mathbf{E} + 14g(\mathbf{E} \cdot \mathbf{B}) \, \mathbf{B}$$

$\partial_0 A_i = E_i$ so this is:

$$\begin{aligned}
\partial_0 \, \frac{\partial L}{\partial(\partial_0 A_i)} &= \partial_0 \, \frac{\partial L_0}{\partial(E_i)} + \partial_0 \, \frac{\partial(\Delta L)}{\partial(E_i)} \\
&= \partial_0 \, \frac{1}{8\pi} \, \frac{\partial(E_i^2 - B_i^2)}{\partial(E_i)} + \partial_0 \, \frac{\partial(\Delta L)}{\partial(E_i)} \\
&= \partial_0 \, \frac{1}{8\pi} \, 2E_i + \partial_0 \, \frac{\partial(\Delta L)}{\partial(E_i)}. \qquad (6.14)
\end{aligned}$$

That is:

$$\partial_0 \, \frac{\partial L}{\partial(\partial_0 A_i)} = \partial_0 \left(\frac{1}{4\pi} \, E + \, P \right)_i \qquad (6.15)$$

with:

$$\mathbf{P} \equiv \frac{\delta \Delta L}{\delta \mathbf{E}} = 4g \left(\mathbf{E}^2 - \mathbf{B}^2 \right) \mathbf{E} + 14g(\mathbf{E} \cdot \mathbf{B}) \, \mathbf{B}. \qquad (6.16)$$

We then tackle the spatial derivatives:

$$\begin{aligned}
\partial_j \, \frac{\partial L}{\partial(\partial_j A_i)} &= 0 \\
\partial_j \, \frac{\partial L_0}{\partial(\partial_j A_i)} + \partial_j \, \frac{\partial(\Delta L)}{\partial(\partial_j A_i)} &= 0. \qquad (6.17)
\end{aligned}$$

- Let us first find the spatial derivative of the term for L_0, $\partial_j \, \frac{\partial L_0}{\partial(\partial_j A_i)}$ — which therefore would lead to the usual form of the inhomogeneous Maxwell equation Eq. (6.10). Recall:

$$L_0 = \frac{1}{8\pi}(\mathbf{E}^2 - \mathbf{B}^2) \qquad (6.18)$$

so:

$$\partial_j \, \frac{\partial L_0}{\partial(\partial_j A_i)} \;\; = \;\; -\frac{1}{8\pi} \, \partial_j \, \frac{\partial}{\partial(\partial_j A_i)} \left(\sum_{k=1}^{3} B_k^2 \right) \tag{6.19}$$

$$\frac{\partial (B_k)^2}{\partial(\partial_j A_i)} = 2 \; B_k \; \frac{\partial B_k}{\partial(\partial_j A_i)} = 2 \; B_k \; \frac{\partial \left(\epsilon_{kji} \, \partial_j A_i\right)}{\partial(\partial_j A_i)} = 2 \; \epsilon_{kji} \, B_k \frac{\partial (B_k)^2}{\partial(\partial_j A_i)} = 2 \; \epsilon_{kji} \, B_k, ^{10} \tag{6.20}$$

and taking the derivative:

$$\begin{aligned} \partial_j \frac{\partial (B_k)^2}{\partial(\partial_j A_i)} \quad &= 2 \; \epsilon_{kji} \, \partial_j B_k \\ &= -2 \; \epsilon_{ijk} \, \partial_j B_k. \end{aligned} \tag{6.21}$$

---

[10] In full details:

$$\frac{\partial (B_3)^2}{\partial(\partial_1 A_2)} = 2 \; B_3 \; \frac{\partial B_3}{\partial(\partial_1 A_2)} = 2 \; B_3 \; \frac{\partial \left(\partial_1 A_2 - \partial_2 A_1\right)}{\partial(\partial_1 A_2)} = 2 \; B_3; \text{ so } \partial_1 \frac{\partial (B_3)^2}{\partial(\partial_1 A_2)} = \; 2 \; \partial_1 B_3$$

$$\frac{\partial (B_3)^2}{\partial(\partial_2 A_1)} = 2 \; B_3 \; \frac{\partial B_3}{\partial(\partial_2 A_1)} = 2 \; B_3 \; \frac{\partial \left(\partial_1 A_2 - \partial_2 A_1\right)}{\partial(\partial_2 A_1)} = -2 \; B_3; \text{ so } \partial_2 \frac{\partial (B_3)^2}{\partial(\partial_2 A_1)} = -2 \; \partial_2 B_3$$

$$\frac{\partial (B_2)^2}{\partial(\partial_3 A_1)} = 2 \; B_2 \; \frac{\partial B_2}{\partial(\partial_3 A_1)} = 2 \; B_2 \; \frac{\partial \left(\partial_3 A_1 - \partial_1 A_3\right)}{\partial(\partial_3 A_1)} = 2 \; B_2; \text{ so } \partial_3 \frac{\partial (B_2)^2}{\partial(\partial_3 A_1)} = \; 2 \; \partial_3 B_2$$

$$\frac{\partial (B_2)^2}{\partial(\partial_1 A_3)} = 2 \; B_2 \; \frac{\partial B_2}{\partial(\partial_1 A_3)} = 2 \; B_2 \; \frac{\partial \left(\partial_3 A_1 - \partial_1 A_3\right)}{\partial(\partial_1 A_3)} = -2 \; B_2; \text{ so } \partial_1 \frac{\partial (B_2)^2}{\partial(\partial_1 A_3)} = -2 \; \partial_1 B_2$$

$$\frac{\partial (B_1)^2}{\partial(\partial_2 A_3)} = 2 \; B_1 \; \frac{\partial B_1}{\partial(\partial_2 A_3)} = 2 \; B_1 \; \frac{\partial \left(\partial_2 A_3 - \partial_3 A_2\right)}{\partial(\partial_2 A_3)} = 2 \; B_1; \text{ so } \partial_2 \frac{\partial (B_1)^2}{\partial(\partial_2 A_3)} = \; 2 \; \partial_2 B_1$$

$$\frac{\partial (B_1)^2}{\partial(\partial_3 A_2)} = 2 \; B_1 \; \frac{\partial B_1}{\partial(\partial_3 A_2)} = 2 \; B_1 \; \frac{\partial \left(\partial_2 A_3 - \partial_3 A_2\right)}{\partial(\partial_3 A_2)} = -2 \; B_1; \text{ so } \partial_3 \frac{\partial (B_1)^2}{\partial(\partial_3 A_2)} = -2 \; \partial_3 B_1$$

Therefore the spatial derivative of the term for L_0 is: [11]

$$\begin{aligned}
\partial_j \, \frac{\partial L_0}{\partial(\partial_j A_i)} &= -\frac{1}{8\pi} \, \partial_j \, \frac{\partial}{\partial(\partial_j A_i)} \left(\sum_{k=1}^{3} B_k^2 \right) \\
&= -\frac{1}{8\pi} \, \left(-2 \, \epsilon_{ijk} \, \partial_j B_k\right) \\
&= \frac{1}{4\pi} \, \epsilon_{ijk} \, \partial_j B_k \\
&= \frac{1}{4\pi} \, \left(\nabla \times B\right)_i . \qquad (6.23)
\end{aligned}$$

We note with satisfaction that combining the time and spatial derivatives of the term involving L_0 yields Eq. (6.10):

$$\begin{aligned}
\partial_\mu \, \frac{\partial L_0}{\partial(\partial_\mu A_\nu)} &= 0 \\
-\partial_0 \, \frac{\partial L_0}{\partial(\partial_0 A_i)} + \partial_j \, \frac{\partial L_0}{\partial(\partial_j A_i)} &= 0 \\
-\frac{1}{4\pi} \, \partial_0 E_i + \frac{1}{4\pi} \, \left(\nabla \times B\right)_i &= 0 \\
\nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = 0. & \qquad (6.24)
\end{aligned}$$

---

[11] In detail, for the three spatial components of $\mathbf{A}$, the term $\partial_j \, \frac{\partial L_0}{\partial(\partial_j A_i)}$ gives:

$$\begin{aligned}
\text{For } A_1 : &\quad \partial_3 \, \frac{\partial (B_2)^2}{\partial(\partial_3 A_1)} + \partial_2 \, \frac{\partial (B_3)^2}{\partial(\partial_2 A_1)} \quad \Rightarrow \quad 2 \, \partial_3 B_2 - 2 \, \partial_2 B_3 \\
\text{For } A_2 : &\quad \partial_1 \, \frac{\partial (B_3)^2}{\partial(\partial_1 A_2)} + \partial_3 \, \frac{\partial (B_1)^2}{\partial(\partial_3 A_2)} \quad \Rightarrow \quad 2 \, \partial_1 B_3 - 2 \, \partial_3 B_1 \\
\text{For } A_3 : &\quad \partial_2 \, \frac{\partial (B_1)^2}{\partial(\partial_2 A_3)} + \partial_1 \, \frac{\partial (B_2)^2}{\partial(\partial_1 A_3)} \quad \Rightarrow \quad 2 \, \partial_2 B_1 - 2 \, \partial_1 B_2
\end{aligned}$$

- We now find the spatial derivative of the term for ΔL, $\partial_j \frac{\partial(\Delta L)}{\partial(\partial_j A_i)}$:

$$\partial_j \frac{\partial(\Delta L)}{\partial(\partial_j A_i)} = \partial_j \left(\frac{\partial(\Delta L)}{\partial B_k} \frac{\partial B_k}{\partial(\partial_j A_i)} \right). \tag{6.25}$$

The last expression on the RHS is:

$$\frac{\partial B_k}{\partial(\partial_j A_i)} = \frac{\partial \left(\epsilon_{kji}\, \partial_j A_i\right)}{\partial(\partial_j A_i)} = \epsilon_{kji}. \tag{6.26}$$

So Eq. (6.25) is now:

$$\partial_j \frac{\partial(\Delta L)}{\partial(\partial_j A_i)} = \epsilon_{kji}\, \partial_j \left(\frac{\partial(\Delta L)}{\partial B_k} \right) \tag{6.27a}$$

$$= -\, \epsilon_{ijk}\, \partial_j \left(\frac{\partial(\Delta L)}{\partial B_k} \right). \tag{6.27b}$$

With:

$$\left(\frac{\partial(\Delta L)}{\partial B_k} \right) = \left(\frac{\delta(\Delta L)}{\delta B} \right)_k \tag{6.28}$$

this becomes:[12]

$$\partial_j \frac{\partial(\Delta L)}{\partial(\partial_j A_i)} = -\, \epsilon_{ijk}\, \partial_j \left(\frac{\delta(\Delta L)}{\delta B} \right)_k$$

$$= -\left(\nabla \times \left(\frac{\partial(\Delta L)}{\partial B} \right) \right)_k$$

---

[12] So in detail, for the three spatial components of **A**, the term $\partial_j \frac{\partial(\Delta L)}{\partial(\partial_j A_i)}$ gives:

$$\text{For } A_1: \ \epsilon_{kj1}\, \partial_j \left(\frac{\partial(\Delta L)}{\partial B_k} \right) \Rightarrow \partial_3 \left(\frac{\partial(\Delta L)}{\partial B_2} \right) - \partial_2 \left(\frac{\partial(\Delta L)}{\partial B_3} \right) \tag{6.29a}$$

$$\text{For } A_2: \ \epsilon_{kj2}\, \partial_j \left(\frac{\partial(\Delta L)}{\partial B_k} \right) \Rightarrow \partial_1 \left(\frac{\partial(\Delta L)}{\partial B_3} \right) - \partial_3 \left(\frac{\partial(\Delta L)}{\partial B_1} \right) \tag{6.29b}$$

$$\text{For } A_3: \ \epsilon_{kj3}\, \partial_j \left(\frac{\partial(\Delta L)}{\partial B_k} \right) \Rightarrow \partial_2 \left(\frac{\partial(\Delta L)}{\partial B_1} \right) - \partial_1 \left(\frac{\partial(\Delta L)}{\partial B_2} \right) \tag{6.29c}$$

with:

$$\mathbf{M} \equiv \frac{\delta(\Delta L)}{\delta B} = -4g\left(\mathbf{E}^2 - B_i^2\right) B_i + 14g(\mathbf{E} \cdot B_i)\,\mathbf{E}$$

$$\partial_j \, \frac{\partial(\Delta L)}{\partial(\partial_j A_i)} = -\left(\nabla \times \mathbf{M}\right)_i .$$

Combining this result with what was obtained for $\partial_j \, \frac{\partial(L_0)}{\partial(\partial_j A_i)}$ we get:

$$\begin{aligned} \partial_j \, \frac{\partial L}{\partial(\partial_j A_i)} &= \partial_j \, \frac{\partial L_0}{\partial(\partial_j A_i)} + \partial_j \, \frac{\partial(\Delta L)}{\partial(\partial_j A_i)} \\ &= \frac{1}{4\pi} \, \left(\nabla \times B\right)_i - \left(\nabla \times \mathbf{M}\right)_i \\ &= \nabla \times \left(\frac{1}{4\pi} \, B - M\right)_i . \end{aligned} \tag{6.30}$$

And our Euler-Lagrange equation finally yields:

$$\begin{aligned} \partial_\mu \, \frac{\partial L}{\partial(\partial_\mu A_\nu)} &= 0 \\ -\,\partial_0 \, \frac{\partial L}{\partial(\partial_0 A_i)} + \partial_j \, \frac{\partial L}{\partial(\partial_j A_i)} &= 0 \\ -\,\partial_t \left(\frac{1}{4\pi} \, E + \; P\right)_i + \nabla \times \left(\frac{1}{4\pi} \, B - M\right)_i &= 0. \end{aligned}$$

This gives us our result for the analog that we sought to Maxwell's equation:

$$\nabla \times \left(B_i - 4\pi \, \mathbf{M}\right) - \; \partial_t \left(\mathbf{E} + 4\pi \, \mathbf{P}\right) = 0 \tag{6.31}$$

with:

$$\mathbf{P} \equiv \frac{\delta \Delta L}{\delta \mathbf{E}} = 4g\left(\mathbf{E}^2 - \mathbf{B}^2\right) \mathbf{E} + 14g(\mathbf{E} \cdot \mathbf{B})\mathbf{B} \tag{6.32a}$$

$$\mathbf{M} \equiv \frac{\delta(\Delta L)}{\delta B} = -4g\left(\mathbf{E}^2 - \mathbf{B}^2\right) \mathbf{B} + 14g(\mathbf{E} \cdot \mathbf{B})\mathbf{E}. \tag{6.32b}$$

Or equivalently in terms of vector components:

$$P_i = 4g\left(E_j^2 - B_j^2\right)E_i + 14g\;(E_jB_j)B_i \tag{6.33a}$$

$$M_i = -4g\left(E_j^2 - B_j^2\right)B_i + 14g\;(E_jB_j)E_i. \tag{6.33b}$$

6.2.3 Derivation of the polarization and magnetization of the background electromagnetic field

So far the electromagnetic fields that we have discussed have been classical. At this point we introduce quantized fields that are to represent the background field. We do retain classical fields as well though, that describe "the wave whose propagation we wish to study", *i.e.* the Scharnhorst photons.[13] Following Barton I shall retain the symbols E_i and B_i to now stand for the quantized fields, while e_i and b_i represent the classical ones:

$$E_i \to E_i + e_i \tag{6.34a}$$

$$B_i \to B_i + b_i. \tag{6.34b}$$

In terms of these new fields, the polarization and magnetization obtained above in Eq. (6.33) become the following. We have:

$$\begin{aligned}\langle E_j^2 - B_j^2\rangle &\to \langle E_j^2 + e_j^2 + 2E_je_j - B_j^2 - b_j^2 - 2B_jb_j\rangle \\ &\to \langle E_j^2 + 2E_je_j - B_j^2 - 2B_jb_j\rangle\end{aligned} \tag{6.35}$$

---

[13] [6], p.560.

where in the second line we have dropped the terms non-linear in the classical fields e_j and b_j.

$$\begin{aligned}\langle E_j B_j\rangle &\rightarrow \langle (E_j+e_j)(B_j+b_j)\rangle \\ &\rightarrow \langle E_jB_j+E_jb_j+B_je_j+e_jb_j\rangle \\ &\rightarrow \langle E_jB_j+E_jb_j+B_je_j\rangle. \end{aligned} \tag{6.36}$$

We substitute these results in Eq. (6.33). The polarization and magnetization are the responses to the external electric and magnetic fields e and b, and we now represent them by p_i and m_i:

$$\begin{aligned} p_i &= 4g\,\langle\left(E_j^2+2E_je_j-B_j^2-2B_jb_j\right)(E_i+e_i)\rangle+14g\,\langle(E_jB_j+E_jb_j+e_jB_j)(B_i+b_i)\rangle \\ &= 4g\,\langle E_j^2E_i+2E_jE_ie_j-B_j^2E_i-2B_jE_ib_j+E_j^2e_i+2E_je_je_i-B_j^2e_i-2B_jb_je_i\rangle \\ &+ 14g\,\langle E_jB_jB_i+E_jB_ib_j+B_iB_je_j+E_jB_jb_i+E_jb_jb_i+B_je_jb_i\rangle \\ &= 4g\,\langle 2E_jE_ie_j-B_j^2E_i-2B_jE_ib_j+E_j^2e_i-B_j^2e_i\rangle \\ &+ 14g\,\langle E_jB_jB_i+E_jB_ib_j+B_iB_je_j+E_jB_jb_i\rangle \end{aligned} \tag{6.37}$$

$$\begin{aligned} m_i &= -4g\,\langle E_j^2+2E_je_j-B_j^2-2B_jb_j(B_i+b_i)\rangle+14g\,\langle E_jB_j+E_jb_j+B_je_j(E_i+e_i)\rangle \\ &= -4g\,\langle E_j^2B_i+2E_jB_ie_j-B_j^2B_i-2B_jB_ib_j+E_j^2b_i+2E_je_jb_i-B_j^2b_i-2B_jb_jb_i\rangle \\ &+ 14g\,\langle E_jB_jE_i+E_jE_ib_j+B_jE_ie_j+E_jB_je_i+E_jb_je_i+B_je_je_i\rangle \\ &= -4g\,\langle E_j^2B_i+2E_jB_ie_j-2B_jB_ib_j+E_j^2b_i-B_j^2b_i\rangle \\ &+ 14g\,\langle E_jB_jE_i+E_jE_ib_j+B_jE_ie_j+E_jB_je_i\rangle. \end{aligned} \tag{6.38}$$

Where the last equalities are obtained by discarding the terms quadratic in e and b as well as the terms $E_j^2 E_i$ and $B_j^2 B_i$. In addition, time-reversal invariance implies that the expectation values of the form $\langle E_i B_j \rangle$ are zero:

$$p_i = 4g\left(2E_jE_ie_j + E_j^2 e_i - B_j^2 e_i\right) + 14g\ \left(B_iB_je_j\right) \tag{6.39a}$$

$$m_i = -4g\left(-2B_jB_ib_j + E_j^2 b_i - B_j^2 b_i\right) + 14g\ \left(E_jE_ib_j\right). \tag{6.39b}$$

This yields:

$$\begin{aligned} p_i &= 4g\left(\langle \mathbf{E}^2 - \mathbf{B}^2\rangle e_i + 2\langle E_iE_j\rangle e_j\right) + 14g\langle B_iB_j\rangle e_j \\ &\equiv \chi_{ij}^{(e)} e_j \end{aligned} \tag{6.40a}$$

$$\begin{aligned} m_i &= 4g\left(-\langle \mathbf{E}^2 - \mathbf{B}^2\rangle b_i + 2\langle B_iB_j\rangle b_j\right) + 14g\langle E_iE_j\rangle b_j \\ &\equiv \chi_{ij}^{(m)} b_j \end{aligned} \tag{6.40b}$$

so that we have obtained the dielectric and magnetic susceptibility tensors, $\chi_{ij}^{(e)}$ and $\chi_{ij}^{(m)}$, of the vacuum in terms of the quantum electromagnetic fields $\mathbf{E}$ and $\mathbf{B}$.
In order to identify the expressions for $\chi_{ij}^{(e)}$ and $\chi_{ij}^{(m)}$ themselves we rewrite this as:

$$\begin{aligned} p_i &= 4g\left(\langle \mathbf{E}^2 - \mathbf{B}^2\rangle \delta_{ij}\ e_j + 2\langle E_iE_j\rangle e_j\right) + 14g\ \langle B_iB_j\rangle e_j \\ &\equiv \chi_{ij}^{(e)} e_j \end{aligned} \tag{6.41a}$$

$$\begin{aligned} m_i &= 4g\left(-\langle \mathbf{E}^2 - \mathbf{B}^2\rangle \delta_{ij}\ b_j + 2\langle B_iB_j\rangle b_j\right) + 14g\ \langle E_iE_j\rangle b_j \\ &\equiv \chi_{ij}^{(m)} b_j. \end{aligned} \tag{6.41b}$$

Then we find for $\chi_{ij}^{(e)}$ and $\chi_{ij}^{(m)}$:

$$\chi_{ij}^{(e)} = 4g\left(\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle \delta_{ij} + 2\langle E_i E_j \rangle\right) + 14g\ \langle B_i B_j \rangle \tag{6.42a}$$

$$\chi_{ij}^{(m)} = 4g\left(-\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle \delta_{ij} + 2\langle B_i B_j \rangle\right) + 14g\ \langle E_i E_j \rangle \tag{6.42b}$$

so that we have obtained the dielectric and magnetic susceptibility tensors, $\chi_{ij}^{(e)}$ and $\chi_{ij}^{(m)}$, of the vacuum in terms of the quantum electromagnetic fields $\mathbf{E}$ and $\mathbf{B}$.

6.2.4 Derivation of the shift in the index of refraction and phase velocity

The index of refraction itself is related to the permittivity and permeability by:

$$n = (\epsilon\mu)^{\frac{1}{2}} \tag{6.43}$$

so that:

$$n_{trivial\ vacuum} + \Delta n = (\epsilon_0 + \Delta\epsilon)^{\frac{1}{2}} (\mu_0 + \Delta\mu)^{\frac{1}{2}}\,. \tag{6.44}$$

Barton uses natural units and “unrationalized gaussian units” for the electromagnetic field, so $n_{trivial\ vacuum} = 1$, $\epsilon_0 = 1$ and $\mu_0 = 1$, and consequently:

$$\begin{aligned} 1 + \Delta n &= (1 + \Delta\epsilon)^{\frac{1}{2}} (1 + \Delta\mu)^{\frac{1}{2}} \\ 1 + \Delta n &= (1 + \Delta\epsilon + \Delta\mu + \Delta\epsilon\Delta\mu)^{\frac{1}{2}} \\ 1 + \Delta n &\simeq 1 + \frac{1}{2}(\Delta\epsilon + \Delta\mu + \Delta\epsilon\Delta\mu) \\ \Delta n &\simeq \frac{1}{2}(\Delta\epsilon + \Delta\mu) \end{aligned} \tag{6.45}$$

where we have neglected the term in $\Delta\epsilon\Delta\mu$.

Now the relative permittivity and permeability tensors ϵ_{ij} and μ_{ij} can easily be found from the electric and magnetic susceptibility tensors $\chi_{ij}^{(e)}$ and $\chi_{ij}^{(m)}$ that we defined at the end of the previous section, in Eq. (6.41):[14]

$$\epsilon_{ij} = \delta_{ij} + 4\pi\chi_{ij}^{(e)} = \delta_{ij} + \Delta\epsilon_{ij} \tag{6.46a}$$

$$\mu_{ij} = \delta_{ij} + 4\pi\chi_{ij}^{(m)} = \delta_{ij} + \Delta\mu_{ij}. \tag{6.46b}$$

In the absence of the quantized electromagnetic fields, these would have their classical values, *i.e.* $\epsilon_{ij} = \delta_{ij}$ and $\mu_{ij} = \delta_{ij}$. So we have found that the quantum fields shift the permittivity and permeability tensors by an amount:

$$\Delta\epsilon_{ij} = 4\pi\chi_{ij}^{(e)} \tag{6.47a}$$

$$\Delta\mu_{ij} = 4\pi\chi_{ij}^{(m)}. \tag{6.47b}$$

Recall our result Eq. (6.48) from the previous section:

$$\chi_{ij}^{(e)} = 4g\left(\langle \mathbf{E}^2 - \mathbf{B}^2\rangle\delta_{ij} + 2\langle E_iE_j\rangle\right) + 14g\ \langle B_iB_j\rangle \tag{6.48a}$$

$$\chi_{ij}^{(m)} = 4g\left(-\langle \mathbf{E}^2 - \mathbf{B}^2\rangle\delta_{ij} + 2\langle B_iB_j\rangle\right) + 14g\ \langle E_iE_j\rangle. \tag{6.48b}$$

The shifts $\Delta\epsilon_{ij}$ and $\Delta\epsilon_{ij}$ that we need to determine are therefore given by:

$$\Delta\epsilon_{ij} = 4\pi\left(4g\left(\langle \mathbf{E}^2 - \mathbf{B}^2\rangle\delta_{ij} + 2\langle E_iE_j\rangle\right) + 14g\ \langle B_iB_j\rangle\right) \tag{6.49a}$$

$$\Delta\mu_{ij} = 4\pi\left(4g\left(-\langle \mathbf{E}^2 - \mathbf{B}^2\rangle\delta_{ij} + 2\langle B_iB_j\rangle\right) + 14g\ \langle E_iE_j\rangle\right). \tag{6.49b}$$

---

[14] The absolute electric susceptibility tensor χ_{ij} is given by: $P_i = \chi_{ij}E_j$ with E_j an electric field; the absolute susceptibility is related to the relative susceptibility χ_r by $\chi = \epsilon_0\,\chi_r$.
Relative and absolute permittivity ϵ_r and ϵ are related by $\epsilon = \epsilon_0\epsilon_r$.

It is in the derivation of expressions for $\langle E_i E_j \rangle$ and $\langle B_i B_j \rangle$ that the present work, where the effect comes from source fields instead of vacuum fluctuations, differs from Barton's.

Using zero-point fields in these expressions, Barton obtains:[15]

$$\langle E_i E_j \rangle = \left(\frac{\pi}{L}\right)^4 \frac{2}{3\pi} \left(\frac{1}{120} \left(-\delta^{\|} + \delta^{\perp}\right)_{ij} + \delta_{ij} f\left(\xi\right) \right) \tag{6.50}$$

$$\langle B_i B_j \rangle = \left(\frac{\pi}{L}\right)^4 \frac{2}{3\pi} \left(\frac{1}{120} \left(-\delta^{\|} + \delta^{\perp}\right)_{ij} - \delta_{ij} f\left(\xi\right) \right) \tag{6.51}$$

where:

$$\delta^{\|}_{ij} \equiv \left(\delta_{i1}\delta_{j1} + \delta_{i2}\delta_{j2}\right), \qquad \delta^{\perp} \equiv \delta_{i3}\delta_{j3}, \tag{6.52}$$

$$\xi \equiv \frac{2\pi z}{L} - \pi \tag{6.53}$$

$$f\left(\xi\ \right) = \frac{3 - 2\ cos^2\left(\frac{\xi}{2}\right)}{8\ cos^4\left(\frac{\xi}{2}\right)}. \tag{6.54}$$

Substituting these results in the expressions for the shifts in the dielectric and magnetic susceptibility tensors, Eq. (6.49), Barton obtained:

$$\Delta\epsilon_{ij} = 4\pi\left(4g\left(\langle \mathbf{E}^2 - \mathbf{B}^2\rangle\delta_{ij} + 2\langle E_i E_j\rangle\right) + 14g\ \langle B_i B_j\rangle\right) \tag{6.55a}$$

$$\Delta\mu_{ij} = 4\pi\left(4g\left(-\langle \mathbf{E}^2 - \mathbf{B}^2\rangle\delta_{ij} + 2\langle B_i B_j\rangle\right) + 14g\ \langle E_i E_j\rangle\right). \tag{6.55b}$$

---

[15] The derivation of these results is presented in Appendix D.
These results, i.e. Eq. (6.50) and Eq. (6.51) yield the energy density of the Casimir effect — in the following way:

$$\begin{aligned}
\left\langle \frac{\left(\mathbf{E}^2 + \mathbf{B}^2\right)}{8\pi} \right\rangle &= \frac{\langle E_x E_x\rangle + \langle E_y E_y\rangle + \langle E_z E_z\rangle + \langle B_x B_x\rangle + \langle B_y B_y\rangle + \langle B_z B_z\rangle}{8\pi} \\
&= \frac{1}{8\pi}\left(\frac{\pi}{L}\right)^4 \frac{2}{3\pi}\left[\frac{1}{120}\left(-2+1\right) + 3f\left(\xi\right) + \frac{1}{120}\left(-2+1\right) - 3f\left(\xi\right)\right] \\
&= \frac{1}{8\pi}\left(\frac{\pi}{L}\right)^4 \frac{2}{3\pi}\left[\frac{-2}{120}\right] \\
&= \frac{-\pi^2}{720\ L^4}.
\end{aligned}$$

When we now substitute these the two equations the expressions for the vacuum expectation values, *i.e.* Eq. (6.50) and Eq. (6.51) we obtain the following results for the shifts in the permittivity and the permeability tensors, which are those given by Barton:

$$\begin{aligned}
\Delta\epsilon_{ij} &= 4\pi\left(4g\left(\langle\mathbf{E}^2-\mathbf{B}^2\rangle\delta_{ij}+2\langle E_iE_j\rangle\right)+14g\ \langle B_iB_j\rangle\right)\\
&= 4\pi\left(\frac{\pi}{L}\right)^4\frac{2}{3\pi}\Bigg(4g\left(-\frac{1}{120}+3f\left(\xi\right)+\frac{1}{120}+3f\left(\xi\right)\right)\delta_{ij}\\
&+ 8g\left(\frac{1}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{ij}+\delta_{ij}f\left(\xi\right)\right)+14g\ \left(\frac{1}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{ij}-\delta_{ij}f\left(\xi\right)\right)\Bigg)\\
&= \left(\frac{\pi}{L}\right)^4\frac{8}{3}\left(4g\left(6f\left(\xi\right)\right)\delta_{ij}+22g\ \frac{1}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{ij}-6g\ \delta_{ij}f\left(\xi\right)\right)\\
&= \left(\frac{\pi}{L}\right)^4\frac{16}{3}g\left(\frac{11}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{ij}+9\ \delta_{ij}f\left(\xi\right)\right) \qquad (6.56)
\end{aligned}$$

$$\begin{aligned}
\Delta\mu_{ij} &= 4\pi\left(4g\left(-\langle\mathbf{E}^2-\mathbf{B}^2\rangle\delta_{ij}+2\langle B_iB_j\rangle\right)+14g\ \langle E_iE_j\rangle\right)\\
&= 4\pi\left(\frac{\pi}{L}\right)^4\frac{2}{3\pi}\Bigg(4g\left(\frac{1}{120}-3f\left(\xi\right)-\frac{1}{120}-3f\left(\xi\right)\right)\delta_{ij}\\
&+ 8g\left(\frac{1}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{ij}-\delta_{ij}f\left(\xi\right)\right)+14g\ \left(\frac{1}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{ij}+\delta_{ij}f\left(\xi\right)\right)\Bigg)\\
&= \left(\frac{\pi}{L}\right)^4\frac{8}{3}\left(4g\left(-6f\left(\xi\right)\right)\delta_{ij}+22g\ \frac{1}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{ij}-6g\ \delta_{ij}f\left(\xi\right)\right)\\
&= \left(\frac{\pi}{L}\right)^4\frac{16}{3}g\left(\frac{11}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{ij}-9\ \delta_{ij}f\left(\xi\right)\right). \qquad (6.57)
\end{aligned}$$

What this implies for the shift in the index of refraction is best discussed by considering separately different directions of propagation of the Scharnhorst photons, and assuming that they can propagate as plane waves. The two limiting cases are for travel parallel to and normal to the Casimir plates.

Let us first discuss the case where they propagate parallel to the plates. Let us call the direction of propagation the x-direction (*i.e.* 1), and the direction of polarization of the electric field **e** the y-direction (*i.e.* 2).[16] $\Delta\epsilon$ is then given by $\Delta\epsilon_{22}$. Since the magnetic field **b** is normal to **e** and to the direction of propagation, it is in the z-direction, so that $\Delta\mu$ is then given by $\Delta\mu_{33}$.

$$\begin{aligned}\Delta\epsilon_{22} &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{22} - 3\,\delta_{22} f\left(\xi\right)\right) \\ &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(-1\right) - 3\,f\left(\xi\right)\right) \qquad (6.58)\end{aligned}$$

$$\begin{aligned}\Delta\mu_{33} &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{33} + 3\,\delta_{33} f\left(\xi\right)\right) \\ &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(+1\right) + 3\,f\left(\xi\right)\right). \qquad (6.59)\end{aligned}$$

The shift in the index of refraction is then:

$$\Delta n = \frac{1}{2}\left(\Delta\epsilon + \Delta\mu\right) = \frac{1}{2}\left(\Delta\epsilon_{22} + \Delta\mu_{33}\right) = 0 \qquad (6.60)$$

If we were to instead call z the direction of polarization of **e**, so that $\Delta\epsilon$ was $\Delta\epsilon_{33}$ and $\Delta\mu$ was $\Delta\mu_{22}$, we would get similarly:

$$\begin{aligned}\Delta\epsilon_{33} &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{33} - 3\,\delta_{33} f\left(\xi\right)\right) \\ &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(+1\right) - 3\,f\left(\xi\right)\right) \qquad (6.61)\end{aligned}$$

---

[16] Respectively the directions 1 and 2.

$$\begin{aligned}\Delta\mu_{22} &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{22} + 3\,\delta_{22}\, f\left(\xi\right)\right) \\ &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(-1\right) + 3\, f\left(\xi\right)\right) \qquad (6.62)\end{aligned}$$

$$\begin{aligned}\Delta n &= \frac{1}{2}\left(\Delta\epsilon + \Delta\mu\right) \\ &= \frac{1}{2}\left(\Delta\epsilon_{33} + \Delta\mu_{22}\right) = 0. \qquad (6.63)\end{aligned}$$

Therefore, the index of refraction experienced by a photon travelling parallel to the Casimir plates is the same as in normal vacuum. This makes intuitive sense from the symmetry of the set-up: in that direction, all the modes of the zero-point field present in the trivial, unbounded vacuum can form, since they are not subject to boundary conditions — other than infinity.

The other limiting case, in which we can expect the shift in the index of refraction to be most pronounced, concerns Scharnhorst photons traveling normal to the plates. Then **e** and **b** are both parallel to the plates. $\Delta\epsilon$ and $\Delta\mu$ do not contain terms parallel to $\delta^{\perp}$.

$$\begin{aligned}\Delta\epsilon_{ij} &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{ij} - 3\,\delta_{ij} f\left(\xi\right)\right) \\ &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(-\delta^{\|}\right)_{ij} - 3\,\delta_{ij} f\left(\xi\right)\right) \qquad (6.64a)\end{aligned}$$

$$\begin{aligned}\Delta\mu_{ij} &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(-\delta^{\|}+\delta^{\perp}\right)_{ij} + 3\,\delta_{ij} f\left(\xi\right)\right) \\ &= \left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\left(-\delta^{\|}\right)_{ij} + 3\,\delta_{ij} f\left(\xi\right)\right) \qquad (6.64b)\end{aligned}$$

and therefore:

$$\Delta\epsilon = -\left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120} - 3\ f\left(\xi\right)\right) \tag{6.65a}$$

$$\Delta\mu = -\left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120} + 3\ f\left(\xi\right)\right). \tag{6.65b}$$

These yield an index of refraction:

$$\begin{aligned}\Delta n = \frac{1}{2}\left(\Delta\epsilon + \Delta\mu\right) = -\left(\frac{\pi}{L}\right)^4 \frac{16}{3} g \left(\frac{11}{120}\right) \quad &= \quad -\frac{\alpha^2}{(mL)^4}\ \frac{11\pi^2}{2^2\ 3^4\ 5^2} \\ &\approx \quad -1.55 \times 10^{-48}(cm/L)^4.\end{aligned}$$

Therefore, for propagation normal to the Casimir plates the phase velocity (which Barton represents by c) acquires a value $c'_\perp$ greater than in normal vacuum:

$$c'_\perp = \frac{c}{n} \approx \left(1 - \Delta n\right) c > c. \tag{6.66}$$

6.3 $\langle E_i E_j\rangle$ and $\langle B_i B_j\rangle$ in terms of source fields: deriving expressions for the Green's functions $\Gamma_{ij}(\mathbf{r}', \mathbf{r}', \omega)$

In the derivation of the Scharnhorst effect given by Barton, the fields in $\langle E_i E_j\rangle$ and $\langle B_i B_j\rangle$ are the zero-point fields [6]. In contrast, we wish to derive the effect under the assumption that these are source fields. Following Milonni and Shih, we introduce the latter as quantized fields, and ensure that they alone appear in our derivation, not the vacuum field, by normal ordering the field operators [17]. That is, when taking vacuum expectation values, only operators containing the annihilation operator $\mathbf{a}$ act on the vacuum state, on the right and those involving the creation operator $\mathbf{a}^\dagger$ act on the left: $\langle 0|\mathbf{a}^\dagger \mathbf{a}|0\rangle$. Since the field operator that contains the annihilation operator is the positive frequency field and vice versa, this yields the following ordering for the fields themselves: $\langle 0|E^{-}E^{+}|0\rangle$, *i.e.* the positive frequency field acts on the right and the negative frequency one to the left. This is equivalent to saying that only the positive frequency field actually comes into play, since the negative frequency field only appears as its complex conjugate (as it acts on the left).

Although the physical situation of interest corresponds to the Casimir set-up, *i.e.* two conducting plates separated by vacuum, the derivation of $\langle 0|E^{-}E^{+}|0\rangle$ first deals with the more general case of three regions of arbitrary dielectric constants. Only at the end is the appropriate limit taken.

Since we wish to obtain source fields, the natural formalism to use is the Green's function formalism. In order to derive in this way expressions for the electric and magnetic fields (required to find $\langle E_i E_j\rangle$ and $\langle B_i B_j\rangle$), we first need to obtain expressions for the Green's functions $\Gamma_{ij}(\mathbf{r}', \mathbf{r}', \omega)$. We now turn to this task.

6.3.1 Obtaining the differential equation satisfied by the Green's function

What we first wish to do is to find expressions for the electric and magnetic fields by the Green's function method. In order to do so, we need to obtain the differential equation that the Green's function has to satisfy.

The Green's function $\overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r}, \mathbf{r}'; t, t')$ is defined by:

$$\mathbf{E_s}(\mathbf{r}, t) \equiv \int d^3\mathbf{r}' \int_0^t dt' \; \overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r}, \mathbf{r}'; t, t') \cdot \mathbf{P}(\mathbf{r}', t') \tag{6.67}$$

and its positive frequency by:

$$\mathbf{E}_{\mathbf{s}}^{(+)}(\mathbf{r}, t) \equiv \int d^3\mathbf{r}' \int_0^t dt' \; \overleftrightarrow{\mathbf{\Gamma}}^{(+)}(\mathbf{r}, \mathbf{r}'; t, t') \cdot \mathbf{P}(\mathbf{r}', t'). \tag{6.68}$$

The Green's function involved is a dyadic (*i.e.* tensor of rank 2), not a scalar. This is due to the fact that the components of the electromagnetic fields are not necessarily generated by the corresponding components of the sources: for instance, $\mathbf{B}_x$ is not simply generated by $\mathbf{P}_x$ but by $\mathbf{P}_y$ and $\mathbf{P}_z$ too. Each element of the dyadic Green's function corresponds not only to the field component it represents (given by the column index of the tensor), but also to a given source component (given by the row index). For example, $\mathbf{\Gamma}_{xz}$ is the x-component of the electric field generated by the z-component of the unit source (*i.e.* here, z-component of the polarization). As Eq. (6.67) expresses, in order to find $\mathbf{E_s}$ with our chosen method, we need to find the form of $\overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r}, \mathbf{r}'; t, t')$. In order to do so, we need to determine what differential equation this Green's function (or more precisely its Fourier transform) obeys. We already know what differential equations the Heisenberg picture electric field operator $\mathbf{E}$ obeys; it satisfies Maxwell's equations, notably:[17]

$$\begin{aligned} \nabla \times \mathbf{E} &= -\dot{\mathbf{B}} \\ \nabla \times \mathbf{H} &= \dot{\mathbf{D}}. \end{aligned} \tag{6.69}$$

The media filling the different regions are not magnetized, so $\mathbf{H} = \mathbf{B}$. They are linear dielectrics, so $\mathbf{D} = \epsilon\mathbf{E} + \mathbf{P}$. Therefore the equations above become:

$$\begin{aligned} \nabla \times \mathbf{E} &= -\dot{\mathbf{B}} \\ \nabla \times \mathbf{B} &= \epsilon\dot{\mathbf{E}} + \dot{\mathbf{P}}. \end{aligned} \tag{6.70}$$

---

[17] Here I am using the typical form of Maxwell's equations rather than referring to their generalized form derived earlier from the Euler-Heisenberg Lagrangian, because at the current point in the derivation the set-up consists of three dielectric regions — it is not yet the Casimir set-up specifically.

These imply that $\mathbf{E}$ obeys the following differential equation:

$$-\nabla\times\nabla\times\mathbf{E}-\epsilon\ddot{\mathbf{E}}=\ddot{\mathbf{P}}.^{18} \tag{6.71}$$

Now in order to obtain a differential equation for the Green's function $\overleftrightarrow{\mathbf{\Gamma}}$, we substitute in this the expression for $\mathbf{E}$ that defines the latter, *i.e.* Eq. (6.67). First however, we need to express the Green function in the frequency domain instead of the time domain so it can be easily differentiated, so we take the relevant Fourier Transform:

$$\overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r},\mathbf{r}';t,t')=\frac{1}{2\pi}\int_{-\infty}^{+\infty}d\omega\ \overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r},\mathbf{r}',\omega)\ e^{-i\omega(t-t')}. \tag{6.72}$$

Then we note that differentiating $\overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r},\mathbf{r}',\omega)$ with respect to time is the equivalent to multiplying it by a factor $-i\,\omega$. Inserting Eq. (6.72) in Eq. (6.67), and the resulting expression for $\mathbf{E_s}$ in Eq. (6.71) gives the following differential equation for the Green's function:

$$-\nabla\times\nabla\times\overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r},\mathbf{r}',\omega)+\omega^2\,\epsilon(\omega)\overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r},\mathbf{r}',\omega)=-\omega^2\overleftrightarrow{\mathbf{1}}\delta^3(\mathbf{r}-\mathbf{r}').^{19} \tag{6.73}$$

Note that Eq. (6.73) is simply the Helmholtz wave equation Eq. (6.71) with the point source $-\omega^2\overleftrightarrow{\mathbf{1}}\delta^3(\mathbf{r}-\mathbf{r}')$, and finding solutions for the Helmholtz equation is in fact the method of choice for solving Maxwell's equations: we start with these first order but mixed (*i.e.* involving different fields $\mathbf{B}$ and $\mathbf{E}$) equations and combine them in order to separate the fields, at the cost of ending up with the now second order wave equations. This is what we just did to get Eq. (6.71). In the present situation however, because we are dealing with dyadics, for the purpose of finding

---

[18] This is obtained by simply differentiating the second equation with respect to time, and substituting the resulting expression for $\dot{\mathbf{B}}$ into the first equation.

[19] $\mathbf{P}$ as well can be expanded in Fourier modes in frequency space, so that it gets differentiated with respect to time the same way as $\overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r},\mathbf{r}',\omega)$.

expressions for $\overleftrightarrow{\Gamma}$ we are better off dealing with mixed first order equations. So we now backtrack just a little and reconsider Eq. (6.70) and Eq. (6.71), in order to obtain "Green's functions versions" of these two equations, analogously to Eq. (6.73) being the "Green's function version" of Eq. (6.71). We get, corresponding to Eq. (6.70) and defining the magnetic Green's function $\overleftrightarrow{\mathbf{\Phi}}$:

$$\nabla \times \overleftrightarrow{\mathbf{\Gamma}} = i\,\omega\, \overleftrightarrow{\mathbf{\Phi}}, \tag{6.74}$$

and for Eq. (6.71) (which can also be obtained by substituting Eq. (6.74) into Eq. (6.73)):

$$-\nabla \times \overleftrightarrow{\mathbf{\Phi}} - i\,\omega\epsilon\, \overleftrightarrow{\mathbf{\Gamma}} = i\,\omega\, \overleftrightarrow{\mathbf{1}}\,\delta(\mathbf{r}-\mathbf{r}'). \tag{6.75}$$

6.3.2 Expressing $\overleftrightarrow{\mathbf{\Gamma}}$ in terms of the "transverse electric" and "transverse magnetic" Green's functions

We now consider the components of these two equations, where the single index refers to the vector field component, *i.e.* $\mathbf{\Gamma}_x$ represents the x-component of the electric field due to a point source of arbitrary direction — and similarly $\mathbf{\Phi}_x$ represents the x-component of the corresponding magnetic field.

Components of Eq. (6.74):

x-component:

$$\partial_y \mathbf{\Gamma}_z - \partial_z \mathbf{\Gamma}_y = i\,\omega\, \mathbf{\Phi}_x \tag{6.76a}$$

$$\textit{i.e.:} \qquad \mathbf{\Phi}_x = \frac{i}{\omega}(\partial_z \mathbf{\Gamma}_y - \partial_y \mathbf{\Gamma}_z) \tag{6.76b}$$

y-component:

$$\partial_z \mathbf{\Gamma}_x - \partial_x \mathbf{\Gamma}_z = i\,\omega\;\mathbf{\Phi}_y \tag{6.77a}$$

$$\textit{i.e.:}\qquad \mathbf{\Phi}_y = \frac{i}{\omega}(\partial_x \mathbf{\Gamma}_z - \partial_z \mathbf{\Gamma}_x) \tag{6.77b}$$

z-component:

$$\partial_x \mathbf{\Gamma}_y - \partial_y \mathbf{\Gamma}_x = i\,\omega\;\mathbf{\Phi}_z \tag{6.78a}$$

$$\textit{i.e.:}\qquad \mathbf{\Phi}_z = \frac{i}{\omega}(\partial_y \mathbf{\Gamma}_x - \partial_x \mathbf{\Gamma}_y). \tag{6.78b}$$

Components of Eq. (6.75):

x-component:

$$-\partial_y \mathbf{\Phi}_z + \partial_z \mathbf{\Phi}_y - i\,\omega\epsilon\;\mathbf{\Gamma}_x = i\,\omega\;\delta(\mathbf{r}-\mathbf{r}')\,\hat{x} \tag{6.79a}$$

$$\textit{i.e.:}\qquad \mathbf{\Gamma}_x = \frac{i}{\omega\epsilon}\partial_y \mathbf{\Phi}_z - \frac{i}{\omega\epsilon}\partial_z \mathbf{\Phi}_y - \frac{1}{\epsilon}\;\delta(\mathbf{r}-\mathbf{r}')\,\hat{x} \tag{6.79b}$$

y-component:

$$-\partial_z \mathbf{\Phi}_x + \partial_x \mathbf{\Phi}_z - i\,\omega\epsilon\;\mathbf{\Gamma}_y = i\,\omega\;\delta(\mathbf{r}-\mathbf{r}')\,\hat{y} \tag{6.80a}$$

$$\textit{i.e.:}\qquad \mathbf{\Gamma}_y = \frac{i}{\omega\epsilon}\partial_z \mathbf{\Phi}_x - \frac{i}{\omega\epsilon}\partial_x \mathbf{\Phi}_z - \frac{1}{\epsilon}\;\delta(\mathbf{r}-\mathbf{r}')\,\hat{y} \tag{6.80b}$$

z-component:

$$-\partial_x \mathbf{\Phi}_y + \partial_y \mathbf{\Phi}_x - i\,\omega\epsilon\;\mathbf{\Gamma}_z = i\,\omega\;\delta(\mathbf{r}-\mathbf{r}')\,\hat{z} \tag{6.81a}$$

$$\textit{i.e.:}\qquad \mathbf{\Gamma}_z = \frac{i}{\omega\epsilon}\partial_x \mathbf{\Phi}_y - \frac{i}{\omega\epsilon}\partial_y \mathbf{\Phi}_x - \frac{1}{\epsilon}\;\delta(\mathbf{r}-\mathbf{r}')\,\hat{z}. \tag{6.81b}$$

We are going to use this system of first order mixed equations to obtain second order equations for single components of the Green's functions just now, but in order to do so more easily we first make some simplifications.

Firstly, we take advantage of the cylindrical symmetry of our set-up, since ϵ only varies in the z-direction. We can do so by taking the Fourier transform of $\overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r}, \mathbf{r}', \omega)$ in order to obtain a Green's function whose spatial dependence involves only z and z':

$$\overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r}, \mathbf{r}', \omega) = \int \frac{(d\mathbf{k}_\perp)}{(2\pi)^2} e^{i\mathbf{k}_\perp \cdot (\mathbf{r}-\mathbf{r}')_\perp} \overleftrightarrow{\mathbf{\Gamma}}(z, z', \mathbf{k}_\perp, \omega). \tag{6.82}$$

Secondly, we make a simplifying choice of coordinates: we pick the x-axis to point in the direction of $\mathbf{k}_\perp$, the projection of the wave vector $\mathbf{k}$ onto the x-y plane. In this case $k_x = |\mathbf{k}_\perp|$, which we will designate simply by k, and $k_y = 0$. In addition, this choice also has implications for some of the partial derivatives. Indeed just as differentiating our $\overleftrightarrow{\mathbf{\Gamma}}(\mathbf{r}, \mathbf{r}', \omega)$ with respect to time boils down to multiplying them by $-i\,\omega$, Eq. (6.82) now implies that differentiating them with respect to x and y comes down to multiplying them by, respectively, ik_x and ik_y. Therefore ∂_x multiplies by ik and ∂_y by 0. This allows us to simplify the components of Eq. (6.74) and Eq. (6.75) we obtained above:

Components of Eq. (6.74):

x-component:

$$\mathbf{\Phi}_x = \frac{i}{\omega}(\partial_z \mathbf{\Gamma}_y - \partial_y \mathbf{\Gamma}_z)$$

$$\mathbf{\Phi}_x = \frac{i}{\omega}\, \partial_z \mathbf{\Gamma}_y \tag{6.83a}$$

y-component:

$$\begin{aligned} \mathbf{\Phi}_y &= \frac{i}{\omega}(\partial_x \mathbf{\Gamma}_z - \partial_z \mathbf{\Gamma}_x) \\ \mathbf{\Phi}_y &= -\frac{1}{\omega}(k\,\mathbf{\Gamma}_z + i\partial_z \mathbf{\Gamma}_x) \end{aligned} \tag{6.84a}$$

z-component:

$$\begin{aligned} \mathbf{\Phi}_z &= \frac{i}{\omega}(\partial_y \mathbf{\Gamma}_x - \partial_x \mathbf{\Gamma}_y) \\ \mathbf{\Phi}_z &= \frac{k}{\omega}\,\mathbf{\Gamma}_y. \end{aligned} \tag{6.85a}$$

Components of Eq. (6.75):

x-component:

$$\begin{aligned} \mathbf{\Gamma}_x &= \frac{i}{\omega\epsilon}\partial_y \mathbf{\Phi}_z - \frac{i}{\omega\epsilon}\partial_z \mathbf{\Phi}_y - \frac{1}{\epsilon}\,\delta(\mathbf{r}-\mathbf{r}')\,\hat{x} \\ \mathbf{\Gamma}_x &= -\frac{i}{\omega\epsilon}\partial_z \mathbf{\Phi}_y - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{x} \end{aligned} \tag{6.86a}$$

y-component:

$$\begin{aligned} \mathbf{\Gamma}_y &= \frac{i}{\omega\epsilon}\partial_z \mathbf{\Phi}_x - \frac{i}{\omega\epsilon}\partial_x \mathbf{\Phi}_z - \frac{1}{\epsilon}\,\delta(\mathbf{r}-\mathbf{r}')\,\hat{y} \\ \mathbf{\Gamma}_y &= \frac{i}{\omega\epsilon}\partial_z \mathbf{\Phi}_x + \frac{k}{\omega\epsilon}\mathbf{\Phi}_z - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{y} \end{aligned} \tag{6.87a}$$

z-component:

$$\begin{aligned} \mathbf{\Gamma}_z &= \frac{i}{\omega\epsilon}\partial_x \mathbf{\Phi}_y - \frac{i}{\omega\epsilon}\partial_y \mathbf{\Phi}_x - \frac{1}{\epsilon}\,\delta(\mathbf{r}-\mathbf{r}')\,\hat{z} \\ \mathbf{\Gamma}_z &= -\frac{k}{\omega\epsilon}\mathbf{\Phi}_y - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{z}. \end{aligned} \tag{6.88a}$$

As promised, we can now obtain second-order differential equations for specific components of the Green's functions — most easily for the y-components.
To get the differential equation satisfied by the y-component of the electric Green's function $\mathbf{\Gamma}_y$, we take the y-component of Eq. (6.75), *i.e.* Eq. (6.87a):

$$\mathbf{\Gamma}_y = \frac{i}{\omega\epsilon}\partial_z \mathbf{\Phi}_x + \frac{k}{\omega\epsilon}\mathbf{\Phi}_z - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{y} \tag{6.89}$$

and substitute in it the expressions obtained in Eq. (6.83a) and Eq. (6.85a) for $\mathbf{\Phi}_x$ and $\mathbf{\Phi}_z$ respectively, both of which only involve $\mathbf{\Gamma}_y$:

$$\begin{aligned}
&\mathbf{\Gamma}_y = \frac{i}{\omega\epsilon}\partial_z\left(\frac{i}{\omega}\,\partial_z\mathbf{\Gamma}_y\right) + \frac{k}{\omega\epsilon}\left(\frac{k}{\omega}\mathbf{\Gamma}_y\right) - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{y} \\
\Leftrightarrow\; & -\omega^2\epsilon\,\mathbf{\Gamma}_y = \partial_z^2\mathbf{\Gamma}_y - k^2\mathbf{\Gamma}_y + \omega^2\,\delta(z-z')\,\hat{y} \\
\Leftrightarrow\; & (-\partial_z^2 + k^2 - \omega^2\epsilon)\mathbf{\Gamma}_y = \omega^2\,\delta(z-z')\,\hat{y}. && (6.90\text{a})
\end{aligned}$$

To get the equation for the y-component of the magnetic Green's function $\mathbf{\Phi}_y$ we take the y-component of Eq. (6.74), *i.e.* Eq. (6.84a):

$$\mathbf{\Phi}_y = -\frac{1}{\omega}(k\,\mathbf{\Gamma}_z + i\partial_z\mathbf{\Gamma}_x),$$

and substitute in it the expressions obtained in Eq. (6.86a) and Eq. (6.88a) for $\mathbf{\Gamma}_x$ and $\mathbf{\Gamma}_z$ respectively, both of which only involve $\mathbf{\Phi}_y$:

$$\begin{aligned}
&\mathbf{\Phi}_y = -\frac{1}{\omega}\left(k\left(-\frac{k}{\omega\epsilon}\mathbf{\Phi}_y - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{z}\right) + i\,\partial_z\left(-\frac{i}{\omega\epsilon}\partial_z\mathbf{\Phi}_y - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{x}\right)\right) \\
\Leftrightarrow\; & \mathbf{\Phi}_y - \frac{k^2}{\omega^2\epsilon}\mathbf{\Phi}_y + \partial_z\frac{1}{\omega^2\epsilon}\partial_z\mathbf{\Phi}_y = \frac{k}{\omega\epsilon}\,\delta(z-z')\,\hat{z} + i\,\partial_z\frac{1}{\omega\epsilon}\,\delta(z-z')\,\hat{x} \\
\Leftrightarrow\; & \left(-\partial_z\frac{1}{\epsilon}\partial_z + \frac{k^2}{\epsilon} - \omega^2\right)\mathbf{\Phi}_y = -i\omega\,\partial_z\frac{1}{\epsilon}\,\delta(z-z')\,\hat{x} - \frac{k\omega}{\epsilon}\,\delta(z-z')\,\hat{z}. && (6.91\text{a})
\end{aligned}$$

We now have two second order and inhomogeneous but separate differential equations for $\mathbf{\Gamma}_y$ and $\mathbf{\Phi}_y$. Now recall that $\mathbf{\Gamma}_y$ and $\mathbf{\Phi}_y$ are the y-components of the electric and magnetic fields due to a point source of arbitrary direction, *i.e.* which *a priori* has components in x, y, and z. That is to say, $\mathbf{\Gamma}_y$ and $\mathbf{\Phi}_y$ are vectors, as they contain elements from each of the 3 source contributions, so Eq. (6.90a) and Eq. (6.91a) are vector equations — and indeed their RHS duly contain terms referring to different components, *i.e.* $\hat{y}$ for Eq. (6.90a), and $\hat{x}$ and $\hat{z}$ for Eq. (6.91a) So these "two" equations form six scalar equations, *i.e.*:

$$\begin{aligned}(-\partial_z^2 + k^2 - \omega^2\epsilon)\mathbf{\Gamma}_y &= \omega^2\,\delta(z-z')\,\hat{y} \\ (-\partial_z^2 + k^2 - \omega^2\epsilon)\,(\Gamma_{yx}\,\hat{x} + \Gamma_{yy}\,\hat{y} + \Gamma_{yz}\,\hat{z}) &= \omega^2\,\delta(z-z')\,\hat{y} \qquad (6.92a)\end{aligned}$$

forms:

$$(-\partial_z^2 + k^2 - \omega^2\epsilon)\,\Gamma_{yx} = 0 \qquad (6.92b)$$

$$(-\partial_z^2 + k^2 - \omega^2\epsilon)\,\Gamma_{yy} = \omega^2\,\delta(z-z') \qquad (6.92c)$$

$$(-\partial_z^2 + k^2 - \omega^2\epsilon)\,\Gamma_{yz} = 0 \qquad (6.92d)$$

and:

$$\begin{aligned}\left(-\partial_z\frac{1}{\epsilon}\partial_z + \frac{k^2}{\epsilon} - \omega^2\right)\mathbf{\Phi}_y &= -i\omega\,\partial_z\frac{1}{\epsilon}\,\delta(z-z')\,\hat{x} - \frac{k\omega}{\epsilon}\,\delta(z-z')\,\hat{z} \\ \left(-\partial_z\frac{1}{\epsilon}\partial_z + \frac{k^2}{\epsilon} - \omega^2\right)(\Phi_{yx}\,\hat{x} + \Phi_{yy}\,\hat{y} + \Phi_{yz}\,\hat{z}) &= -i\omega\,\partial_z\frac{1}{\epsilon}\,\delta(z-z')\,\hat{x} - \frac{k\omega}{\epsilon}\,\delta(z-z')\,\hat{z}\end{aligned}$$

$$(6.93a)$$

forms:

$$\left(-\partial_z \frac{1}{\epsilon} \partial_z + \frac{k^2}{\epsilon} - \omega^2\right) \Phi_{yx} = -\, i\omega\, \partial_z \frac{1}{\epsilon}\, \delta(z - z') \tag{6.93b}$$

$$\left(-\partial_z \frac{1}{\epsilon} \partial_z + \frac{k^2}{\epsilon} - \omega^2\right) \Phi_{yy} = 0 \tag{6.93c}$$

$$\left(-\partial_z \frac{1}{\epsilon} \partial_z + \frac{k^2}{\epsilon} - \omega^2\right) \Phi_{yz} = -\, \frac{k\omega}{\epsilon}\, \delta(z - z'). \tag{6.93d}$$

Now, in order to solve these equations, we define two Green's functions, in the formal sense that these functions satisfy Eq. (6.90a) and Eq. (6.91a) with a delta function on the RHS – *i.e.*:

$$(-\partial_z^2 + k^2 - \omega^2 \epsilon)\, g^E(z, z') = \delta(z - z') \tag{6.94a}$$

$$\left(-\partial_z \frac{1}{\epsilon} \partial_z + \frac{k^2}{\epsilon} - \omega^2\right) g^H(z, z') = \delta(z - z'). \tag{6.94b}$$

Indeed if we define two functions $g^E(z, z')$ and $g^H(z, z')$ such that:

$$\Gamma_y = \omega^2 g^E(z, z')\hat{y} \quad ; \quad \Phi_y = \frac{i\,\omega}{\epsilon(z')}\, \partial_{z'} g^H(z, z')\hat{x} - \frac{k\,\omega}{\epsilon(z')} g^H(z, z')\hat{z}, \tag{6.95}$$

then $g^E(z, z')$ and $g^H(z, z')$ satisfy Eq. (6.94).[20] $g^E(z, z')$ and $g^H(z, z')$ are referred to as, respectively, the "transverse electric" and the "transverse magnetic" Green's functions.

Determining the form of $g^E(z, z')$ and $g^H(z, z')$ by solving Eq. (6.94) will be the

---

[20] The equation above can also be written:

$$\Gamma_{yy} = \omega^2 g^E(z, z') \tag{6.96a}$$

$$\Phi_{yx} = \frac{i\,\omega}{\epsilon(z')}\, \partial_{z'} g^H(z, z') \tag{6.96b}$$

$$\Phi_{yz} = -\frac{k\,\omega}{\epsilon(z')} g^H(z, z') \tag{6.96c}$$

topic of the next section. First let us express the components of $\overleftrightarrow{\mathbf{\Gamma}}$ in terms of $g^E(z,z')$ and $g^H(z,z')$. From Eq. (6.86a) and Eq. (6.88a) we know that:

$$\mathbf{\Gamma}_x = -\frac{i}{\omega\epsilon}\partial_z \left(\Phi_{yx}\,\hat{x} + \Phi_{yy}\,\hat{y} + \Phi_{yz}\,\hat{z}\right) - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{x} \tag{6.97a}$$

$$\mathbf{\Gamma}_z = -\frac{k}{\omega\epsilon}\left(\Phi_{yx}\,\hat{x} + \Phi_{yy}\,\hat{y} + \Phi_{yz}\,\hat{z}\right) - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{z}. \tag{6.97b}$$

We know from Eq. (6.96b) and Eq. (6.96c) that the only non-zero components of $\mathbf{\Phi}_y$ are Φ_{yx} and Φ_{yz} so that:

$$\mathbf{\Gamma}_x = -\frac{i}{\omega\epsilon}\partial_z \left(\left(\frac{i\,\omega}{\epsilon(z')}\,\partial_{z'} g^H\right)\hat{x} + \left(-\frac{k\,\omega}{\epsilon(z')} g^H\right)\hat{z}\right) - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{x} \tag{6.98a}$$

$$\mathbf{\Gamma}_z = -\frac{k}{\omega\epsilon}\left(\left(\frac{i\,\omega}{\epsilon(z')}\,\partial_{z'} g^H\right)\hat{x} + \left(-\frac{k\,\omega}{\epsilon(z')} g^H\right)\hat{z}\right) - \frac{1}{\epsilon}\,\delta(z-z')\,\hat{z}. \tag{6.98b}$$

Then from Eq. (6.92) we see that the only non-zero component of $\mathbf{\Gamma}_y$ is Γ_{yy}, which is given by Eq. (6.96a). Therefore the non-zero components of $\overleftrightarrow{\mathbf{\Gamma}}$ are:

$$\Gamma_{xx} = \frac{1}{\epsilon}\partial_z\left(\frac{1}{\epsilon'}\,\partial_{z'} g^H\right) - \frac{1}{\epsilon}\,\delta(z-z') \tag{6.99a}$$

$$\Gamma_{yy} = \omega^2 g^E \tag{6.99b}$$

$$\Gamma_{zz} = \frac{k^2}{\epsilon\epsilon'}\,g^H - \frac{1}{\epsilon}\,\delta(z-z') \tag{6.99c}$$

$$\Gamma_{xz} = i\,\frac{k}{\epsilon\epsilon'}\partial_z\,g^H \tag{6.99d}$$

$$\Gamma_{zx} = -\,i\,\frac{k}{\epsilon\epsilon'}\,\partial_{z'} g^H \tag{6.99e}$$

where ϵ stands for $\epsilon(z)$ and ϵ' for $\epsilon(z')$ — and $g^E(z,z')$, $g^H(z,z')$ have been abbreviated to g^E and g^H.

6.3.3 Determining the expressions for $g^E(z,z')$ and $g^H(z,z')$

FORM OF $g^E(z,z')$

General form of $g^E(z,z')$ in each region

From Eq. (6.94a), $g^E(z,z')$ must obey the following in each region:

$$(-\partial_z^2 + k^2 - \omega^2\epsilon)\, g^E(z,z') = \delta(z-z'), \tag{6.100}$$

where the coordinate z refers to the points where the function is evaluated, and z' to where the sources are located. Recall that they are three regions in our set-up, characterized by different permittivity constants: region I from $z' = -\infty$ to $z' = -a$, region III from $z' = -a$ to $z' = 0$, and region II for $z' = 0$ to $z' = +\infty$. Regions I and II model the plates, and III the space between them.

From the theory of Green's functions the most general form for $g^E(z,z')$ is:[21]

$$g^E(z,z') = f^E(z,z') + h^E(z,z'), \tag{6.101}$$

where $g^E(z,z')$, $f^E(z,z')$ and $h^E(z,z')$ obey:

$$\left(-\partial_z^2 + K^2\right) g^E(z,z') = \delta(z-z') \tag{6.102}$$

$$\left(-\partial_z^2 + K^2\right) f^E(z,z') = \delta(z-z') \tag{6.103}$$

$$\left(-\partial_z^2 + K^2\right) h^E(z,z') = 0, \tag{6.104}$$

with $K^2 \equiv k^2 - \omega^2\epsilon$.

Unlike $g^E(z,z')$, $f^E(z,z')$ – called the "fundamental solution"– is not required to satisfy the same boundary conditions as the physical situation imposes on $g^E(z,z')$. However, the behavior of $h^E(z,z')$ at the boundaries must make up for that of $f^E(z,z')$, so that the sum $f^E(z,z') + h^E(z,z')$ can behave appropriately there.

---

[21] [107], p.587.

Boundary Conditions At Infinity

The physical set-up we are using, with three dielectric regions, is the one introduced by Evgeny Lifshitz in 1956 to derive "the interaction of bodies whose surfaces are brought within a small distance of one another."[22] On introducing his work Lifshitz stated:

> The interaction of the objects is regarded as occurring through the medium of the fluctuating electromagnetic field which is always present in the interior of any absorbing medium, and also extends beyond its boundaries, — partially in the form of travelling waves radiated by the body, partially in the form of standing waves which are damped exponentially as we move away from the surface of the body. It must be emphasized that this field does not vanish even at absolute zero, at which point it is associated with the zero point vibrations of the radiation field. ([93], p.73.)

Lifshitz went on to perform his calculation using the radiated traveling waves. However, what we wish to do here is to follow the approach championed by J. Schwinger, L. L. DeRaad Jr, and K. A. Milton [100] and P. Milonni and M.-L. Shih [17] because it allowed them to recover the Casimir force on the assumption that it is entirely due to source fields, not to zero-point ones. The appropriate field in this case is the one "damped exponentially as we move away from the surface of the body" — or more precisely as we move away from the sources considered, since we start by considering a space entirely filled with dielectric materials. Therefore the appropriate boundary

---

[22] [93]. Lifshitz was motivated by the need for a macroscopic model, by opposition to the "microscopic" theory calculations based on intermolecular forces. The latter relied on the assumption that these forces were additive, in the sense that the force between two molecules was unaffected by the presence of other molecules. The results of such an approach are valid in the limit of rarefied media, but are limited to this case.

conditions at $z = -\infty$ and $z = +\infty$, for a fixed value of z', are that $g^E(z, z')$ should vanish there; the fact that Green's functions are symmetric between z and z' shows that the reverse is also true, *i.e.* for a fixed value of z, $g^E(z, z')$ vanishes at $z' = -\infty$ and $z' = +\infty$.

Now the LHS of Eq. (6.103) corresponds to the "modified Helmholtz" equation in one-dimension,[23] and the textbook result for $f^E(z, z')$ in this case is:[24]

$$\begin{aligned} f^E(z,z') &= \frac{e^{-K_i|z-z'|}}{2K_i} \\ &= \frac{e^{-K_i z}\ e^{K_j z'}\ \theta(z-z') + e^{K_i z}\ e^{-K_j z'}\ \theta(z'-z)}{2K_i}, \end{aligned} \tag{6.105a}$$

where we have accounted for the possibility that z' may not be in the same region as z by ascribing to it a constant K_j different from K_i. Therefore for $f^E(z, z')$:

	General form of $f^E(z,z')$		
z z’	II	III	I
II	$\frac{e^{-K_2 z} e^{K_2 z'} \theta(z-z')}{2K_2} +$ $\frac{e^{K_2 z} e^{-K_2 z'} \theta(z'-z)}{2K_2}$	$\frac{e^{-K_3 z} e^{K_2 z'} \theta(z-z')}{2K_3} +$ $\frac{e^{K_3 z} e^{-K_2 z'} \theta(z'-z)}{2K_3}$	$\frac{e^{-K_1 z} e^{K_2 z'} \theta(z-z')}{2K_1} +$ $\frac{e^{K_1 z} e^{-K_2 z'} \theta(z'-z)}{2K_1}$
III	$\frac{e^{-K_2 z} e^{K_3 z'} \theta(z-z')}{2K_2} +$ $\frac{e^{K_2 z} e^{-K_3 z'} \theta(z'-z)}{2K_2}$	$\frac{e^{-K_3 z} e^{K_3 z'} \theta(z-z')}{2K_3} +$ $\frac{e^{K_3 z} e^{-K_3 z'} \theta(z'-z)}{2K_3}$	$\frac{e^{-K_1 z} e^{K_3 z'} \theta(z-z')}{2K_1} +$ $\frac{e^{K_1 z} e^{-K_3 z'} \theta(z'-z)}{2K_1}$
I	$\frac{e^{-K_2 z} e^{K_1 z'} \theta(z-z')}{2K_2} +$ $\frac{e^{K_2 z} e^{-K_1 z'} \theta(z'-z)}{2K_2}$	$\frac{e^{-K_3 z} e^{K_1 z'} \theta(z-z')}{2K_3} +$ $\frac{e^{K_3 z} e^{-K_1 z'} \theta(z'-z)}{2K_3}$	$\frac{e^{-K_1 z} e^{K_1 z'} \theta(z-z')}{2K_1} +$ $\frac{e^{K_1 z} e^{-K_1 z'} \theta(z'-z)}{2K_1}$

Table 6.1: General form of $f^E(z,z')$ for z and z' in different regions

---

[23] The Helmholtz equation is $(\nabla^2 + k^2)F = 0$, the "modified Helmholtz" equation has a minus sign instead: $(\nabla^2 - k^2)F = 0$.

[24] See Appendix E.

The rows in this table refer to a specific region for the location of the source, and the columns to the regions in which the function is evaluated. For instance, the expression in row “II” column “III” refers to the form of $f^E(z,z')$ in region III due to a source located in region II.

For the various scenarios regarding the position of z and z', the theta functions give:

z z’	II	III	I
II	$(-z+z')\theta(z-z')$ $+(z-z')\theta(z'-z)$ $=-\lvert z-z'\rvert$	$\theta(z-z')=0$ $\theta(z'-z)=1$	$\theta(z-z')=0$ $\theta(z'-z)=1$
III	$\theta(z-z')=1$ $\theta(z'-z)=0$	$(-z+z')\theta(z-z')$ $+(z-z')\theta(z'-z)$ $=-\lvert z-z'\rvert$	$\theta(z-z')=0$ $\theta(z'-z)=1$
I	$\theta(z-z')=1$ $\theta(z'-z)=0$	$\theta(z-z')=1$ $\theta(z'-z)=0$	$(-z+z')\theta(z-z')$ $+(z-z')\theta(z'-z)$ $=-\lvert z-z'\rvert$

Table 6.2: Theta Functions

Therefore $f^E(z,z')$ is of the form:

	General form of $f^E(z,z')$		
z z’	II	III	I
II	$\frac{e^{-K_2\lvert z-z'\rvert}}{2K_2}$	$\frac{e^{K_3 z}e^{-K_2 z'}}{2K_3}$	$\frac{e^{K_1 z}e^{-K_2 z'}}{2K_1}$
III	$\frac{e^{-K_2 z}e^{K_3 z'}}{2K_2}$	$\frac{e^{-K_3\lvert z-z'\rvert}}{2K_3}$	$\frac{e^{K_1 z}e^{-K_3 z'}}{2K_1}$
I	$\frac{e^{-K_2 z}e^{K_1 z'}}{2K_2}$	$\frac{e^{-K_3 z}e^{K_1 z'}}{2K_3}$	$\frac{e^{-K_1\lvert z-z'\rvert}}{2K_1}$

Table 6.3: General form of $f^E(z,z')$ for z and z' in different regions

As shown in Table (6.3), for a fixed value of z', $f^E(z,z')$ already satisfies the required boundary conditions, since it vanishes as $z \to +\infty$ in region II and $z \to -\infty$ in region

I. Hence the boundary conditions on $h^E(z,z')$ must simply be that it does so too. Now in order to satisfy:

$$\left(-\partial_z^2 + K^2\right) h^E(z,z') = 0, \tag{6.106}$$

$h^E(z,z')$ must have the general form:

$$h^E(z,z') \sim A_+(z')e^{Kz} + A_-(z')e^{-Kz} \tag{6.107}$$

where K usually differs in the three regions.

In order to determine the general form of $A_+(z')$ and $A_-(z')$, we note that:

- the boundary conditions require that $h^E(z,z')$ vanish as $z \to +\infty$ in region II and $z \to -\infty$ in region I. Hence, for z in region II for instance, the term $A_-(z')e^{-K_2 z}$ is present, but $A_+(z')\,e^{K_2 z}$ is not since it would "blow up" as $z \to +\infty$.

- from the theory of Green's functions, the general form of $h^E(z,z')$ must be symmetric between z and z'. This notably implies that $h^E(z,z')$ must vanish as $z' \to +\infty$ in region II and $z' \to -\infty$ in region I, since it does for $z \to +\infty$ in region II and $z \to -\infty$ in region I. This leads to the following form for $h^E(z,z')$:

	General form of $h^E(z,z')$		
z z'	II	III	I
II	$e^{-K_2 z}e^{-K_2 z'}$	$(e^{K_3 z}+e^{-K_3 z})\,e^{-K_2 z'}$	$e^{K_1 z}e^{-K_2 z'}$
III	$e^{-K_2 z}(e^{K_3 z'}+e^{-K_3 z'})$	$(e^{K_3 z}+e^{-K_3 z})(e^{K_3 z'}+e^{-K_3 z'})$	$e^{K_1 z}(e^{K_3 z'}+e^{-K_3 z'})$
I	$e^{-K_2 z}e^{K_1 z'}$	$(e^{K_3 z}+e^{-K_3 z})e^{K_1 z'}$	$e^{K_1 z}e^{K_1 z'}$

Table 6.4: General form of $h^E(z,z')$ for z and z' in different regions

Therefore the complete Green's function solution $g^E(z,z') = f^E(z,z') + h^E(z,z')$, including coefficients, must be of the form:[25]

---

[25] The terms corresponding to $h^E(z,z')$ are expressed with a factor of $\frac{1}{2K}$ for them to look similar to those of $f^E(z,z')$; the coefficients such as r_2, etc. simply "contain" a factor of $2K$ compared to

	$g^E(z,z')$		
z z'	II	III	I
II	$\frac{e^{-K_2\|z-z'\|}}{2K_2}$ $+r_2\frac{e^{-K_2 z}e^{-K_2 z'}}{2K_2}$	$\frac{e^{K_3 z}e^{-K_2 z'}}{2K_3}$ $+l_2\frac{e^{-K_3 z}e^{-K_2 z'}}{2K_3}$ $+n_2\frac{e^{K_3 z}e^{-K_2 z'}}{2K_3}$	$\frac{e^{K_1 z}e^{-K_2 z'}}{2K_1}$ $+t_2\frac{e^{K_1 z}e^{-K_2 z'}}{2K_1}$
III	$\frac{e^{-K_2 z}e^{K_3 z'}}{2K_2}$ $+r_3\frac{e^{-K_2 z}e^{-K_3 z'}}{2K_2}$ $+s_3\frac{e^{-K_2 z}e^{K_3 z'}}{2K_2}$	$\frac{e^{-K_3\|z-z'\|}}{2K_3}$ $+l_3\frac{e^{-K_3 z}e^{-K_3 z'}}{2K_3}$ $+m_3\frac{e^{-K_3 z}e^{K_3 z'}}{2K_3}$ $+n_3\frac{e^{K_3 z}e^{-K_3 z'}}{2K_3}$ $+p_3\frac{e^{K_3 z}e^{K_3 z'}}{2K_3}$	$\frac{e^{K_1 z}e^{-K_3 z'}}{2K_1}$ $+t_3\frac{e^{K_1 z}e^{-K_3 z'}}{2K_1}$ $+u_3\frac{e^{K_1 z}e^{K_3 z'}}{2K_1}$
I	$\frac{e^{-K_2 z}e^{K_1 z'}}{2K_2}$ $+s_1\frac{e^{-K_2 z}e^{K_1 z'}}{2K_2}$	$\frac{e^{-K_3 z}e^{K_1 z'}}{2K_3}$ $+m_1\frac{e^{-K_3 z}e^{K_1 z'}}{2K_3}$ $+p_1\frac{e^{K_3 z}e^{K_1 z'}}{2K_3}$	$\frac{e^{-K_1\|z-z'\|}}{2K_1}$ $+u_1\frac{e^{K_1 z}e^{K_1 z'}}{2K_1}$

Table 6.5: $g^E(z,z')$ for z and z' in different regions

Now the exact form of the coefficients is determined by how these expressions "match up" at the boundaries between the three regions — *i.e.* whether $g^E(z,z')$ is continuous or not, and if not what the discontinuity should be. The behavior of $g^E(z,z')$ alone is not all that is at stake though: so is the behavior of its derivatives with respect to z. Therefore we need the expressions for $\partial_z g^E(z,z')$ in the various regions. The expressions involving absolute values give:

$$\frac{e^{-K_i|z-z'|}}{2K_i} = \frac{e^{-K_i(z-z')}\theta(z-z') + e^{-K_i(z'-z)}\theta(z'-z)}{2K_i} \tag{6.108}$$

---

what they would be otherwise.

$$\partial_z\left(\frac{e^{-K_i|z-z'|}}{2K_i}\right)=\frac{e^{-K_i(z-z')}\left(-K_i\,\theta(z-z')+\delta(z-z')\right)}{2K_i}+\frac{e^{-K_i(z'-z)}\left(K_i\,\theta(z'-z)-\delta(z'-z)\right)}{2K_i}$$
$$=\frac{e^{-K_i(z-z')}}{2}\left(-\theta(z-z')+\frac{\delta(z-z')}{K_i}\right)+\frac{e^{K_i(z-z')}}{2}\left(\theta(z'-z)-\frac{\delta(z'-z)}{K_i}\right)$$

The derivative of $g^E(z,z')$ with respect to z therefore has the form:

	$\partial_z g^E(z,z')$		
z z’	II	III	I
II	$\frac{e^{-K_2z}e^{K_2z'}}{2}$ $\left(-\theta(z-z')+\frac{\delta(z-z')}{K_2}\right)$ $+\frac{e^{K_2z}e^{-K_2z'}}{2}$ $\left(\theta(z'-z)-\frac{\delta(z'-z)}{K_2}\right)$ $-r_2\frac{e^{-K_2z}e^{-K_2z'}}{2}$	$\frac{e^{K_3z}e^{-K_2z'}}{2}$ $-l_2\frac{e^{-K_3z}e^{-K_2z'}}{2}$ $+n_2\frac{e^{K_3z}e^{-K_2z'}}{2}$	$\frac{e^{K_1z}e^{-K_2z'}}{2}$ $+t_2\frac{e^{K_1z}e^{-K_2z'}}{2}$
III	$-\frac{e^{-K_2z}e^{K_3z'}}{2}$ $-r_3\frac{e^{-K_2z}e^{-K_3z'}}{2}$ $-s_3\frac{e^{-K_2z}e^{K_3z'}}{2}$	$\frac{e^{-K_3z}e^{K_3z'}}{2}\left(-\theta(z-z')+\frac{\delta(z-z')}{K_3}\right)$ $+\frac{e^{K_3z}e^{-K_3z'}}{2}\left(\theta(z'-z)-\frac{\delta(z'-z)}{K_3}\right)$ $-l_3\frac{e^{-K_3z}e^{-K_3z'}}{2}$ $-m_3\frac{e^{-K_3z}e^{K_3z'}}{2}$ $+n_3\frac{e^{K_3z}e^{-K_3z'}}{2}$ $+p_3\frac{e^{K_3z}e^{K_3z'}}{2}$	$\frac{e^{K_1z}e^{-K_3z'}}{2}$ $+t_3\frac{e^{K_1z}e^{-K_3z'}}{2}$ $+u_3\frac{e^{K_1z}e^{K_3z'}}{2}$
I	$-\frac{e^{-K_2z}e^{K_1z'}}{2}$ $-s_1\frac{e^{-K_2z}e^{K_1z'}}{2}$	$-\frac{e^{-K_3z}e^{K_1z'}}{2}$ $-m_1\frac{e^{-K_3z}e^{K_1z'}}{2}$ $+p_1\frac{e^{K_3z}e^{K_1z'}}{2}$	$\frac{e^{-K_1z}e^{K_1z'}}{2}$ $\left(-\theta(z-z')+\frac{\delta(z-z')}{K_1}\right)$ $+\frac{e^{K_1z}e^{-K_1z'}}{2}$ $\left(\theta(z'-z)-\frac{\delta(z'-z)}{K_1}\right)$ $+u_1\frac{e^{K_1z}e^{K_1z'}}{2}$

Table 6.6: $\partial_z g^E(z,z')$ for z and z' in different regions

Boundary Conditions At $z = 0$ and $z = -a$

As stated above, taking into account the boundary conditions for $g^E(z, z')$ and $\partial_z g^E(z, z')$ at $z = 0$ and $z = -a$ determines the form of the sixteen coefficients. The boundary conditions concern the continuity of these functions. They provide sixteen equations: eight for the continuity of $g^E(z, z')$ and another eight for that of $\partial_z g^E(z, z')$. In each case, four of the eight equations concern the boundary at $z = 0$ and the other four that at $z = -a$. As always, the components of the electric field that are parallel to these surfaces, *i.e.* E_x and E_y, are continuous:

$$\mathbf{E_{II\,x}}(0) = \mathbf{E_{III\,x}}(0) \qquad \mathbf{E_{III\,x}}(-a) = \mathbf{E_{I\,x}}(-a) \tag{6.109a}$$

$$\mathbf{E_{II\,y}}(0) = \mathbf{E_{III\,y}}(0) \qquad \mathbf{E_{III\,y}}(-a) = \mathbf{E_{I\,y}}(-a). \tag{6.109b}$$

In addition, because there is no free charge density at the boundaries between the regions, the components of the electric displacement that are normal to them are continuous as well, therefore so is ϵE_z:

$$\mathbf{D_{II\,z}}(0) = \mathbf{D_{III\,z}}(0) \qquad \mathbf{D_{III\,z}}(-a) = \mathbf{D_{I\,z}}(-a) \tag{6.110a}$$

$$\epsilon_2 \mathbf{E_{II\,z}}(0) = \epsilon_3 \mathbf{E_{III\,z}}(0) \qquad \epsilon_3 \mathbf{E_{III\,z}}(-a) = \epsilon_1 \mathbf{E_{I\,z}}(-a). \tag{6.110b}$$

That E_y is continuous across the boundaries implies that the Green's function component Γ_{yi} is as well, since it represents the y-component of the electric field (generated by the i-component of the unit source). From Eq. (6.99b) we can see that $g^E(z, z')$ is proportional to Γ_{yy}; therefore it too is continuous across these boundaries. In addition, none of the Green's function components involve $\partial_z g^E(z, z')$, which is continuous at the boundaries.

The relevant expressions for $g^E(z, z')$ at $z = 0$ (*i.e.* at the boundary between regions II and III) and $z = -a$ (*i.e.* at the boundary between regions III and I) are:

	$g^E(z,z')$			
z z'	II; $z\to0^+$	III; $z\to0^-$	III; $z\to-a^+$	I; $z\to-a^-$
II	$\frac{e^{-K_2z'}}{2K_2}$ $+r_2\frac{e^{-K_2z'}}{2K_2}$	$\frac{e^{-K_2z'}}{2K_3}$ $+l_2\frac{e^{-K_2z'}}{2K_3}$ $+n_2\frac{e^{-K_2z'}}{2K_3}$	$\frac{e^{-K_3a}e^{-K_2z'}}{2K_3}$ $+l_2\frac{e^{K_3a}e^{-K_2z'}}{2K_3}$ $+n_2\frac{e^{-K_3a}e^{-K_2z'}}{2K_3}$	$\frac{e^{-K_1a}e^{-K_2z'}}{2K_1}$ $+t_2\frac{e^{-K_1a}e^{-K_2z'}}{2K_1}$
III	$\frac{e^{K_3z'}}{2K_2}$ $+r_3\frac{e^{-K_3z'}}{2K_2}$ $+s_3\frac{e^{K_3z'}}{2K_2}$	$\frac{e^{K_3z'}}{2K_3}$ $+l_3\frac{e^{-K_3z'}}{2K_3}$ $+m_3\frac{e^{K_3z'}}{2K_3}$ $+n_3\frac{e^{-K_3z'}}{2K_3}$ $+p_3\frac{e^{K_3z'}}{2K_3}$	$\frac{e^{-K_3a}e^{-K_3z'}}{2K_3}$ $+l_3\frac{e^{K_3a}e^{-K_3z'}}{2K_3}$ $+m_3\frac{e^{K_3a}e^{K_3z'}}{2K_3}$ $+n_3\frac{e^{-K_3a}e^{-K_3z'}}{2K_3}$ $+p_3\frac{e^{-K_3a}e^{K_3z'}}{2K_3}$	$\frac{e^{-K_1a}e^{-K_3z'}}{2K_1}$ $+t_3\frac{e^{-K_1a}e^{-K_3z'}}{2K_1}$ $+u_3\frac{e^{-K_1a}e^{K_3z'}}{2K_1}$
I	$\frac{e^{K_1z'}}{2K_2}$ $+s_1\frac{e^{K_1z'}}{2K_2}$	$\frac{e^{K_1z'}}{2K_3}$ $+m_1\frac{e^{K_1z'}}{2K_3}$ $+p_1\frac{e^{K_1z'}}{2K_3}$	$\frac{e^{K_3a}e^{K_1z'}}{2K_3}$ $+m_1\frac{e^{K_3a}e^{K_1z'}}{2K_3}$ $+p_1\frac{e^{-K_3a}e^{K_1z'}}{2K_3}$	$\frac{e^{K_1a}e^{K_1z'}}{2K_1}$ $+u_1\frac{e^{-K_1a}e^{K_1z'}}{2K_1}$

Table 6.7: Expressions for $g^E(z,z')$ at the boundaries of the regions

These lead to the following boundary conditions at the interfaces:

	Boundary Conditions — $g^E(z,z')$	
z z'	at $z=0$	at $z=-a$
II	$\frac{1}{K_2}(1+r_2)=\frac{1}{K_3}(1+l_2+n_2)$	$\frac{1}{K_3}((1+n_2)e^{-K_3a}+l_2e^{K_3a})=\frac{1}{K_1}(1+t_2)e^{-K_1a}$
III	$\frac{1}{K_2}(1+s_3)=\frac{1}{K_3}(1+m_3+p_3)$ $\frac{1}{K_2}r_3=\frac{1}{K_3}(l_3+n_3)$	$\frac{1}{K_3}(p_3\,e^{-K_3a}+m_3\,e^{K_3a})=\frac{1}{K_1}u_3\,e^{-K_1a}$ $\frac{1}{K_3}((1+n_3)e^{-K_3a}+l_3\,e^{K_3a})=\frac{1}{K_1}(1+t_3)e^{-K_1a}$
I	$\frac{1}{K_2}(1+s_1)=\frac{1}{K_3}(1+m_1+p_1)$	$\frac{1}{K_3}(p_1e^{-K_3a}+(1+m_1)e^{K_3a})=\frac{1}{K_1}(u_1e^{-K_1a}+e^{K_1a})$

Table 6.8: Boundary conditions for $g^E(z,z')$ to be continuous across the regions

Note that in region III the terms involving $e^{K_3 z'}$ and those involving $e^{-K_3 z'}$ give rise to separate conditions. In order to get the boundary conditions at the $z = 0$ and $z = -a$ for $\partial_z g^E(z, z')$, we note that provided z' is not chosen right on the boundaries as well, the delta functions that appear in $\partial_z g^E(z, z')$ are equal to zero, and the theta functions become:

z z’	II; $z \to 0^+$	III; $z \to 0^-$	III; $z \to -a^+$	I; $z \to -a^-$
II	$\lim\ \theta(z-z')=0$ $\lim\ \theta(z'-z)=1$	$\lim\ \theta(z-z')=0$ $\lim\ \theta(z'-z)=1$	$\lim\ \theta(z-z')=0$ $\lim\ \theta(z'-z)=1$	$\lim\ \theta(z-z')=0$ $\lim\ \theta(z'-z)=1$
III	$\lim\ \theta(z-z')=1$ $\lim\ \theta(z'-z)=0$	$\lim\ \theta(z-z')=1$ $\lim\ \theta(z'-z)=0$	$\lim\ \theta(z-z')=0$ $\lim\ \theta(z'-z)=1$	$\lim\ \theta(z-z')=0$ $\lim\ \theta(z'-z)=1$
I	$\lim\ \theta(z-z')=1$ $\lim\ \theta(z'-z)=0$	$\lim\ \theta(z-z')=1$ $\lim\ \theta(z'-z)=0$	$\lim\ \theta(z-z')=1$ $\lim\ \theta(z'-z)=0$	$\lim\ \theta(z-z')=1$ $\lim\ \theta(z'-z)=0$

Table 6.9: Value of the theta functions at the interfaces

Then the relevant expressions for $\partial_z g^E(z, z')$ are:

	$\partial_z g^E(z,z')$			
z z’	II; $z \to 0^+$	III; $z \to 0^-$	III; $z \to -a^+$	I; $z \to -a^-$
II	$\frac{e^{-K_2 z'}}{2} - r_2 \frac{e^{-K_2 z'}}{2}$	$\frac{e^{-K_2 z'}}{2} - l_2 \frac{e^{-K_2 z'}}{2}$ $+ n_2 \frac{e^{-K_2 z'}}{2}$	$\frac{e^{-K_3 a} e^{-K_2 z'}}{2}$ $- l_2 \frac{e^{K_3 a} e^{-K_2 z'}}{2}$ $+ n_2 \frac{e^{-K_3 a} e^{-K_2 z'}}{2}$	$\frac{e^{-K_1 a} e^{-K_2 z'}}{2}$ $+ t_2 \frac{e^{-K_1 a} e^{-K_2 z'}}{2}$
III	$- \frac{e^{K_3 z'}}{2} - r_3 \frac{e^{-K_3 z'}}{2}$ $- s_3 \frac{e^{K_3 z'}}{2}$	$- \frac{e^{K_3 z'}}{2} - l_3 \frac{e^{-K_3 z'}}{2}$ $- m_3 \frac{e^{K_3 z'}}{2} \quad + n_3 \frac{e^{-K_3 z'}}{2}$ $+ p_3 \frac{e^{K_3 z'}}{2}$	$\frac{e^{-K_3 a} e^{-K_3 z'}}{2} \quad - l_3 \frac{e^{K_3 a} e^{-K_3 z'}}{2}$ $- m_3 \frac{e^{K_3 a} e^{K_3 z'}}{2} \quad + n_3 \frac{e^{-K_3 a} e^{-K_3 z'}}{2}$ $+ p_3 \frac{e^{-K_3 a} e^{K_3 z'}}{2}$	$\frac{e^{-K_1 a} e^{-K_3 z'}}{2}$ $+ t_3 \frac{e^{-K_1 a} e^{-K_3 z'}}{2}$ $+ u_3 \frac{e^{-K_1 a} e^{K_3 z'}}{2}$
I	$- \frac{e^{K_1 z'}}{2} - s_1 \frac{e^{K_1 z'}}{2}$	$- \frac{e^{K_1 z'}}{2} - m_1 \frac{e^{K_1 z'}}{2}$ $+ p_1 \frac{e^{K_1 z'}}{2}$	$- \frac{e^{K_3 a} e^{K_1 z'}}{2} - m_1 \frac{e^{K_3 a} e^{K_1 z'}}{2}$ $+ p_1 \frac{e^{-K_3 a} e^{K_1 z'}}{2}$	$- \frac{e^{K_1 a} e^{K_1 z'}}{2}$ $+ u_1 \frac{e^{-K_1 a} e^{K_1 z'}}{2}$

Table 6.10: Expressions for $\partial_z g^E(z, z')$ at the boundaries of the regions

And they give rise to the following boundary conditions for $\partial_z g^E(z,z')$ at the interfaces between the regions:

	Boundary Conditions — $\partial_z g^E(z,z')$	
z z’	at $z=0$	at $z=-a$
II	$r_2=l_2-n_2$	$(1+n_2)\,e^{-K_3a}-l_2e^{K_3a}=(1+t_2)e^{-K_1a}$
III	$s_3=m_3-p_3$ $r_3=l_3-n_3$	$p_3\,e^{-K_3a}-m_3\,e^{K_3a}=\,u_3\,e^{-K_1a}$ $e^{-K_3a}(1+n_3)-l_3e^{K_3a}=e^{-K_1a}(1+t_3)$
I	$s_1=m_1-p_1$	$p_1e^{-K_3a}-(1+m_1)\,e^{K_3a}=u_1e^{-K_1a}-e^{K_1a}$

Table 6.11: Boundary conditions for $\partial_z g^E(z,z')$ to be continuous across the regions

So that all in all, the conditions required at the interface between the regions are:

	Boundary Conditions — $g^E(z,z')$ and $\partial_z g^E(z,z')$	
z z’	$z=0$	$z=-a$
II	$\frac{1}{K_2}(1+r_2)=\frac{1}{K_3}(1+l_2+n_2)$ $r_2=l_2-n_2$	$\frac{1}{K_3}((1+n_2)e^{-K_3a}+l_2e^{K_3a})=\frac{1}{K_1}(1+t_2)e^{-K_1a}$ $(1+n_2)e^{-K_3a}-l_2e^{K_3a}=(1+t_2)e^{-K_1a}$
III	$\frac{1}{K_2}(1+s_3)=\frac{1}{K_3}(1+m_3+p_3)$ $\frac{1}{K_2}r_3=\frac{1}{K_3}(l_3+n_3)$ $s_3=m_3-p_3$ $r_3=l_3-n_3$	$\frac{1}{K_3}(p_3\,e^{-K_3a}+m_3\,e^{K_3a})=\frac{1}{K_1}u_3\,e^{-K_1a}$ $\frac{1}{K_3}((1+n_3)e^{-K_3a}+l_3\,e^{K_3a})=\frac{1}{K_1}(1+t_3)e^{-K_1a}$ $p_3\,e^{-K_3a}-m_3\,e^{K_3a}=u_3\,e^{-K_1a}$ $e^{-K_3a}(1+n_3)-l_3e^{K_3a}=e^{-K_1a}(1+t_3)$
I	$\frac{1}{K_2}(1+s_1)=\frac{1}{K_3}(1+m_1+p_1)$ $s_1=m_1-p_1$	$\frac{1}{K_3}(p_1e^{-K_3a}+(1+m_1)e^{K_3a})=\frac{1}{K_1}(u_1e^{-K_1a}+e^{K_1a})$ $p_1e^{-K_3a}-(1+m_1)e^{K_3a}=u_1e^{-K_1a}-e^{K_1a}$

Table 6.12: Boundary conditions needed to find the coefficients of $g^E(z,z')$

Coefficients of $g^E(z,z')$

From the system of sixteen equations in Table (6.12) we can find the coefficients:

z' in II	$r_2=\frac{e^{2aK_3}(K_1+K_3)(K_2-K_3)-(K_1-K_3)(K_2+K_3)}{e^{2aK_3}(K_1+K_3)(K_2+K_3)+(K_1-K_3)(K_3-K_2)}$ $l_2=-\frac{2K_3(K_1-K_3)}{e^{2aK_3}(K_1+K_3)(K_2+K_3)+(K_1-K_3)(K_3-K_2)}$ $n_2=\frac{(K_3-K_2)\left(e^{2aK_3}(K_1+K_3)-K_1+K_3\right)}{e^{2aK_3}(K_1+K_3)(K_2+K_3)+(K_1-K_3)(K_3-K_2)}$ $t_2=\frac{4K_1K_3e^{a(K_1+K_3)}}{(K_1+K_3)(K_2+K_3)e^{2aK_3}+(K_1-K_3)(K_3-K_2)}-1$
z' in III	$r_3=-\frac{2K_2(K_1-K_3)}{e^{2aK_3}(K_1+K_3)(K_2+K_3)+(K_1-K_3)(K_3-K_2)}$ $l_3=\frac{1}{\frac{e^{2aK_3}(K_1+K_3)}{K_3-K_1}+\frac{2K_2}{K_2+K_3}-1}$ $n_3=\frac{1}{\frac{e^{2aK_3}(K_1+K_3)(K_2+K_3)}{(K_3-K_1)(K_3-K_2)}-1}$ $s_3=\frac{(K_2-K_3)\left(e^{2aK_3}(K_1+K_3)+K_1-K_3\right)}{e^{2aK_3}(K_1+K_3)(K_2+K_3)+(K_1-K_3)(K_3-K_2)}$ $m_3=\frac{1}{\frac{e^{2aK_3}(K_1+K_3)(K_2+K_3)}{(K_3-K_1)(K_3-K_2)}-1}$ $p_3=\frac{1}{\frac{e^{-2aK_3}(K_1-K_3)}{K_1+K_3}+\frac{K_2+K_3}{K_3-K_2}}$ $t_3=\frac{2K_1(K_2+K_3)e^{a(K_1+K_3)}}{(K_1+K_3)(K_2+K_3)e^{2aK_3}+(K_1-K_3)(K_3-K_2)}-1$ $u_3=\frac{2K_1(K_3-K_2)e^{a(K_1+K_3)}}{e^{2aK_3}(K_1+K_3)(K_2+K_3)+(K_1-K_3)(K_3-K_2)}$
z' in I	$s_1=\frac{4K_2K_3e^{a(K_1+K_3)}}{(K_1+K_3)(K_2+K_3)e^{2aK_3}+(K_1-K_3)(K_3-K_2)}-1$ $m_1=\frac{2K_3(K_2+K_3)e^{a(K_1+K_3)}}{(K_1+K_3)(K_2+K_3)e^{2aK_3}+(K_1-K_3)(K_3-K_2)}-1$ $p_1=\frac{2K_3(K_3-K_2)e^{a(K_1+K_3)}}{(K_1+K_3)(K_2+K_3)e^{2aK_3}+(K_1-K_3)(K_3-K_2)}$ $u_1=\frac{e^{2aK_1}\left((K_1-K_3)(K_2+K_3)e^{2aK_3}-(K_1+K_3)(K_2-K_3)\right)}{(K_1+K_3)(K_2+K_3)e^{2aK_3}+(K_1-K_3)(K_3-K_2)}$

Table 6.13: Coefficients of $g^E(z,z')$

Note that because $g^E(z,z')$ is symmetrical with respect to z and z', these same coefficients can also be obtained by using as boundary conditions the continuity of $g^E(z,z')$ and of its derivative with respect to z', $\partial_{z'} g^E(z,z')$, at $z' = 0$ and $z' = -a$. Doing so also yields the results in Table (6.13).

FORM OF $g^H(z,z')$

General form of $g^H(z,z')$ in each region

Recall that the "transverse electric Green's function" $g^E(z,z')$ is the solution of Eq. (6.94a):

$$(-\partial_z^2 + k^2 - \omega^2 \epsilon)\, g^E(z,z') = \delta(z - z'). \tag{6.111}$$

By contrast, the "transverse magnetic Green's function" $g^H(z,z')$ has to satisfy Eq. (6.94b):

$$\left(-\partial_z \frac{1}{\epsilon} \partial_z + \frac{k^2}{\epsilon} - \omega^2\right) g^H(z,z') = \delta(z - z'). \tag{6.112}$$

For the same reasons as the full solution to Eq. (6.94a) is of the form Eq. (6.101):

$$g^E(z,z') = f^E(z,z') + h^E(z,z'), \tag{6.113}$$

we now have:

$$g^H(z,z') = f^H(z,z') + h^H(z,z'), \tag{6.114}$$

where $g^H(z,z')$, $f^H(z,z')$ and $h^H(z,z')$ obey:

$$\left(-\partial_z^2 + K^2\right) g^H(z,z') = \epsilon\, \delta(z - z') \tag{6.115}$$

$$\left(-\partial_z^2 + K^2\right) f^H(z,z') = \epsilon\, \delta(z - z') \tag{6.116}$$

$$\left(-\partial_z^2 + K^2\right) h^H(z,z') = 0, \tag{6.117}$$

with $K^2 \equiv k^2 - \omega^2\epsilon$.

Recall that $f^E(z,z')$ was the textbook Green's function fundamental solution Eq. (6.105a) for the modified Helmholtz equation Eq. (6.103):

$$\left(-\partial_z^2 + K^2\right) f^E(z,z') = \delta(z-z')$$

$$f^E(z,z') = \frac{e^{-K_i|z-z'|}}{2K_i} = \frac{e^{-K_i z}\ e^{K_j z'}\ \theta(z-z') + e^{K_i z}\ e^{-K_j z'}\ \theta(z'-z)}{2K_i}.$$

Because the equation that $f^H(z,z')$ satisfies only differs in that the RHS is multiplied by ϵ, $f^H(z,z')$ is:

$$f^H(z,z') = \epsilon\,\frac{e^{-K_i|z-z'|}}{2K_i} = \frac{e^{-K_i z}\ e^{K_j z'}\ \theta(z-z') + e^{K_i z}\ e^{-K_j z'}\ \theta(z'-z)}{2K_i'},$$

where $K_i' \equiv \frac{K_i}{\epsilon}$.[26]

As for $h^H(z,z')$, it satisfies the same equation as $h^E(z,z')$ did since Eq. (6.104) was:

$$\left(-\partial_z^2 + K^2\right) h^E(z,z') = 0 \tag{6.118}$$

and Eq. (6.117) is entirely analogous. Therefore $h^H(z,z')$ has the same general form as $h^E(z,z')$ (Table (6.4)):

	General form of $h^H(z,z')$		
z z’	II	III	I
II	$e^{-K_2 z} e^{-K_2 z'}$	$(e^{K_3 z} + e^{-K_3 z})\ e^{-K_2 z'}$	$e^{K_1 z} e^{-K_2 z'}$
III	$e^{-K_2 z}(e^{K_3 z'} + e^{-K_3 z'})$	$(e^{K_3 z} + e^{-K_3 z})(e^{K_3 z'} + e^{-K_3 z'})$	$e^{K_1 z}(e^{K_3 z'} + e^{-K_3 z'})$
I	$e^{-K_2 z} e^{K_1 z'}$	$(e^{K_3 z} + e^{-K_3 z}) e^{K_1 z'}$	$e^{K_1 z} e^{K_1 z'}$

Table 6.14: General form of $h^H(z,z')$ for z and z' in different regions

---

[26] The rationale for defining $K_i' \equiv \frac{K_i}{\epsilon}$ is that the coefficients in $g^H(z,z')$ then look analogous to those in $g^E(z,z')$, as we shall see below.

Hence the complete Green's function solution $g^H(z,z') = f^H(z,z') + h^H(z,z')$ is:[27]

	$g^H(z,z')$		
z z’	II	III	I
II	$\frac{e^{-K_2\|z-z'\|}}{2K'_2}$ $+r'_2\,\frac{e^{-K_2 z}e^{-K_2 z'}}{2K'_2}$	$\frac{e^{K_3 z}e^{-K_2 z'}}{2K'_3}$ $+l'_2\,\frac{e^{-K_3 z}e^{-K_2 z'}}{2K'_3}$ $+n'_2\,\frac{e^{K_3 z}e^{-K_2 z'}}{2K'_3}$	$\frac{e^{K_1 z}e^{-K_2 z'}}{2K'_1}$ $+t'_2\,\frac{e^{K_1 z}e^{-K_2 z'}}{2K'_1}$
III	$\frac{e^{-K_2 z}e^{K_3 z'}}{2K'_2}$ $+r'_3\,\frac{e^{-K_2 z}e^{-K_3 z'}}{2K'_2}$ $+s'_3\,\frac{e^{-K_2 z}e^{K_3 z'}}{2K'_2}$	$\frac{e^{-K_3\|z-z'\|}}{2K'_3}$ $+l'_3\,\frac{e^{-K_3 z}e^{-K_3 z'}}{2K'_3}$ $+m'_3\,\frac{e^{-K_3 z}e^{K_3 z'}}{2K'_3}$ $+n'_3\,\frac{e^{K_3 z}e^{-K_3 z'}}{2K'_3}$ $+p'_3\,\frac{e^{K_3 z}e^{K_3 z'}}{2K'_3}$	$\frac{e^{K_1 z}e^{-K_3 z'}}{2K'_1}$ $+t'_3\,\frac{e^{K_1 z}e^{-K_3 z'}}{2K'_1}$ $+u'_3\,\frac{e^{K_1 z}e^{K_3 z'}}{2K'_1}$
I	$\frac{e^{-K_2 z}e^{K_1 z'}}{2K'_2}$ $+s'_1\,\frac{e^{-K_2 z}e^{K_1 z'}}{2K'_2}$	$\frac{e^{-K_3 z}e^{K_1 z'}}{2K'_3}$ $+m'_1\,\frac{e^{-K_3 z}e^{K_1 z'}}{2K'_3}$ $+p'_1\,\frac{e^{K_3 z}e^{K_1 z'}}{2K'_3}$	$\frac{e^{-K_1\|z-z'\|}}{2K'_1}$ $+u'_1\,\frac{e^{K_1 z}e^{K_1 z'}}{2K'_1}$

Table 6.15: $g^H(z,z')$ for z and z' in different regions

In other words, $g^H(z,z')$ is similar to $g^E(z,z')$ with K_i replaced by:

$$K_i \rightarrow K'_i = \frac{K_i}{\epsilon_i} \tag{6.119}$$

except in the exponentials.

The derivative of $g^H(z,z')$ with respect to z therefore has the form:

---

[27] The terms corresponding to $h^H(z,z')$ are now expressed with a factor of $\frac{1}{2K'}$, rather than $\frac{1}{2K}$ as we had in $h^E(z,z')$. The aim is again to make these coefficients look more akin to those of $f^H(z,z')$. As a result, coefficients such as r'_2, etc. now "contain" a factor of $2K'$ compared to what they would otherwise be.

	$\partial_z g^H(z,z')$		
z z’	II	III	I
II	$\frac{e^{-K_2 z}e^{K_2 z'}}{2}$ $\left(-\frac{K_2}{K_2'}\theta(z-z')+\frac{\delta(z-z')}{K_2'}\right)$ $+\frac{e^{K_2 z}e^{-K_2 z'}}{2}$ $\left(\frac{K_2}{K_2'}\theta(z'-z)-\frac{\delta(z'-z)}{K_2'}\right)$ $-r_2'\frac{K_2}{K_2'}\frac{e^{-K_2 z}e^{-K_2 z'}}{2}$	$\frac{K_3}{K_3'}\Big(\frac{e^{K_3 z}e^{-K_2 z'}}{2}$ $-l_2'\frac{e^{-K_3 z}e^{-K_2 z'}}{2}$ $+n_2'\frac{e^{K_3 z}e^{-K_2 z'}}{2}\Big)$	$\frac{K_1}{K_1'}\Big(\frac{e^{K_1 z}e^{-K_2 z'}}{2}$ $+t_2'\frac{e^{K_1 z}e^{-K_2 z'}}{2}\Big)$
III	$\frac{K_2}{K_2'}\Big(-\frac{e^{-K_2 z}e^{K_3 z'}}{2}$ $-r_3'\frac{e^{-K_2 z}e^{-K_3 z'}}{2}$ $-s_3'\frac{e^{-K_2 z}e^{K_3 z'}}{2}\Big)$	$\frac{e^{-K_3 z}e^{K_3 z'}}{2}\left(-\frac{K_3}{K_3'}\theta(z-z')+\frac{\delta(z-z')}{K_3'}\right)$ $+\frac{e^{K_3 z}e^{-K_3 z'}}{2}\left(\frac{K_3}{K_3'}\theta(z'-z)-\frac{\delta(z'-z)}{K_3'}\right)$ $+\frac{K_3}{K_3'}\Big(-l_3'\frac{e^{-K_3 z}e^{-K_3 z'}}{2}$ $-m_3'\frac{e^{-K_3 z}e^{K_3 z'}}{2}$ $+n_3'\frac{e^{K_3 z}e^{-K_3 z'}}{2}$ $+p_3'\frac{e^{K_3 z}e^{K_3 z'}}{2}\Big)$	$\frac{K_1}{K_1'}\Big(\frac{e^{K_1 z}e^{-K_3 z'}}{2}$ $+t_3'\frac{e^{K_1 z}e^{-K_3 z'}}{2}$ $+u_3'\frac{e^{K_1 z}e^{K_3 z'}}{2}\Big)$
I	$\frac{K_2}{K_2'}\Big(-\frac{e^{-K_2 z}e^{K_1 z'}}{2}$ $-s_1'\frac{e^{-K_2 z}e^{K_1 z'}}{2}\Big)$	$\frac{K_3}{K_3'}\Big(-\frac{e^{-K_3 z}e^{K_1 z'}}{2}$ $-m_1'\frac{e^{-K_3 z}e^{K_1 z'}}{2}$ $+p_1'\frac{e^{K_3 z}e^{K_1 z'}}{2}\Big)$	$\frac{e^{-K_1 z}e^{K_1 z'}}{2}$ $\left(-\frac{K_1}{K_1'}\theta(z-z')+\frac{\delta(z-z')}{K_1'}\right)$ $+\frac{e^{K_1 z}e^{-K_1 z'}}{2}$ $\left(\frac{K_1}{K_1'}\theta(z'-z)-\frac{\delta(z'-z)}{K_1'}\right)$ $+u_1'\frac{K_1}{K_1'}\frac{e^{K_1 z}e^{K_1 z'}}{2}$

Table 6.16: $\partial_z g^H(z,z')$ for z and z' in different regions

Boundary Conditions At $z = 0$ and $z = -a$

As for $g^E(z,z')$, the form of the sixteen coefficients is determined from the boundary conditions at $z = 0$ and $z = -a$. Recall from Eq. (6.109a) and Eq. (6.110a) that E_x and ϵE_z are continuous across the interfaces between the regions. That ϵE_z is continuous means that the $\epsilon\,\Gamma_{zi}$ are, hence from Eq. (6.99c) and Eq. (6.99e) that $g^H(z,z')$ too. Continuity of E_x implies that of Γ_{xi}, which in turn, from Eq. (6.99a) and Eq. (6.99d), implies that of $\frac{1}{\epsilon}\partial_z g^H(z,z')$. The relevant expressions for $g^H(z,z')$ at $z=0$ and $z=-a$ are:

	$g^H(z{,}z')$			
z z’	II; $z{\to}0^+$	III; $z{\to}0^-$	III; $z{\to}{-a}^+$	I; $z{\to}{-a}^-$
II	$\frac{e^{-K_2z'}}{2K_2'}$ $+r_2'\frac{e^{-K_2z'}}{2K_2'}$	$\frac{e^{-K_2z'}}{2K_3'}$ $+l_2'\frac{e^{-K_2z'}}{2K_3'}$ $+n_2'\frac{e^{-K_2z'}}{2K_3'}$	$\frac{e^{-K_3a}e^{-K_2z'}}{2K_3'}$ $+l_2'\frac{e^{K_3a}e^{-K_2z'}}{2K_3'}$ $+n_2'\frac{e^{-K_3a}e^{-K_2z'}}{2K_3'}$	$\frac{e^{-K_1a}e^{-K_2z'}}{2K_1'}$ $+t_2'\frac{e^{-K_1a}e^{-K_2z'}}{2K_1'}$
III	$\frac{e^{K_3z'}}{2K_2'}$ $+r_3'\frac{e^{-K_3z'}}{2K_2'}$ $+s_3'\frac{e^{K_3z'}}{2K_2'}$	$\frac{e^{K_3z'}}{2K_3'}$ $+l_3'\frac{e^{-K_3z'}}{2K_3'}$ $+m_3'\frac{e^{K_3z'}}{2K_3'}$ $+n_3'\frac{e^{-K_3z'}}{2K_3'}$ $+p_3'\frac{e^{K_3z'}}{2K_3'}$	$\frac{e^{-K_3a}e^{-K_3z'}}{2K_3'}$ $+l_3'\frac{e^{K_3a}e^{-K_3z'}}{2K_3'}$ $+m_3'\frac{e^{K_3a}e^{K_3z'}}{2K_3'}$ $+n_3'\frac{e^{-K_3a}e^{-K_3z'}}{2K_3'}$ $+p_3'\frac{e^{-K_3a}e^{K_3z'}}{2K_3'}$	$\frac{e^{-K_1a}e^{-K_3z'}}{2K_1'}$ $+t_3'\frac{e^{-K_1a}e^{-K_3z'}}{2K_1'}$ $+u_3'\frac{e^{-K_1a}e^{K_3z'}}{2K_1'}$
I	$\frac{e^{K_1z'}}{2K_2'}$ $+s_1'\frac{e^{K_1z'}}{2K_2'}$	$\frac{e^{K_1z'}}{2K_3'}$ $+m_1'\frac{e^{K_1z'}}{2K_3'}$ $+p_1'\frac{e^{K_1z'}}{2K_3'}$	$\frac{e^{K_3a}e^{K_1z'}}{2K_3'}$ $+m_1'\frac{e^{K_3a}e^{K_1z'}}{2K_3'}$ $+p_1'\frac{e^{-K_3a}e^{K_1z'}}{2K_3'}$	$\frac{e^{K_1a}e^{K_1z'}}{2K_1'}$ $+u_1'\frac{e^{-K_1a}e^{K_1z'}}{2K_1'}$

Table 6.17: Expressions for $g^H(z,z')$ at the boundaries of the regions

These lead to the following boundary conditions at the interfaces, where as before the terms involving $e^{K_3z'}$ and $e^{-K_3z'}$ give rise to separate conditions in region III:

	Boundary Conditions — $g^H(z,z')$	
z z’	at $z=0$	at $z=-a$
II	$\frac{1}{K_2'}(1+r_2') = \frac{1}{K_3'}(1+l_2'+n_2')$	$\frac{1}{K_3'}((1+n_2')e^{-K_3a}+l_2'e^{K_3a}) = \frac{1}{K_1'}(1+t_2')e^{-K_1a}$
III	$\frac{1}{K_2'}(1+s_3') = \frac{1}{K_3'}(1+m_3'+p_3')$ $\frac{1}{K_2'}r_3' = \frac{1}{K_3'}(l_3'+n_3')$	$\frac{1}{K_3'}(p_3'\,e^{-K_3a}+m_3'\,e^{K_3a}) = \frac{1}{K_1'}u_3'\,e^{-K_1a}$ $\frac{1}{K_3'}((1+n_3')e^{-K_3a}+l_3'\,e^{K_3a}) = \frac{1}{K_1'}(1+t_3')e^{-K_1a}$
I	$\frac{1}{K_2'}(1+s_1') = \frac{1}{K_3'}(1+m_1'+p_1')$	$\frac{1}{K_3'}(p_1'e^{-K_3a}+(1+m_1')e^{K_3a}) = \frac{1}{K_1'}(u_1'e^{-K_1a}+e^{K_1a})$

Table 6.18: Boundary conditions for g$^H(z,z')$ to be continuous across the regions

Using table (6.9) we find the relevant expressions for $\partial_z g^H(z,z')$ at $z=0$ and $z=-a$ are:

	$\partial_z g^H(z,z')$			
z z’	II; $z\to0^+$	III; $z\to0^-$	III; $z\to-a^+$	I; $z\to-a^-$
II	$\frac{K_2}{K_2'}\Big(\frac{e^{K_2z}e^{-K_2z'}}{2}$ $-r_2'\frac{e^{-K_2z}e^{-K_2z'}}{2}\Big)$	$\frac{K_3}{K_3'}\Big(\frac{e^{K_3z}e^{-K_2z'}}{2}$ $-l_2'\frac{e^{-K_3z}e^{-K_2z'}}{2}$ $+n_2'\frac{e^{K_3z}e^{-K_2z'}}{2}\Big)$	$\frac{K_3}{K_3'}\Big(\frac{e^{K_3z}e^{-K_2z'}}{2}$ $-l_2'\frac{e^{-K_3z}e^{-K_2z'}}{2}$ $+n_2'\frac{e^{K_3z}e^{-K_2z'}}{2}\Big)$	$\frac{K_1}{K_1'}\Big(\frac{e^{K_1z}e^{-K_2z'}}{2}$ $+t_2'\frac{e^{K_1z}e^{-K_2z'}}{2}\Big)$
III	$\frac{K_2}{K_2'}\Big(-\frac{e^{-K_2z}e^{K_3z'}}{2}$ $-r_3'\frac{e^{-K_2z}e^{-K_3z'}}{2}$ $-s_3'\frac{e^{-K_2z}e^{K_3z'}}{2}\Big)$	$-\frac{K_3}{K_3'}\frac{e^{-K_3z}e^{K_3z'}}{2}$ $+\frac{K_3}{K_3'}\Big(-l_3'\frac{e^{-K_3z}e^{-K_3z'}}{2}$ $-m_3'\frac{e^{-K_3z}e^{K_3z'}}{2}$ $+n_3'\frac{e^{K_3z}e^{-K_3z'}}{2}$ $+p_3'\frac{e^{K_3z}e^{K_3z'}}{2}\Big)$	$\frac{K_3}{K_3'}\frac{e^{K_3z}e^{-K_3z'}}{2}$ $+\frac{K_3}{K_3'}\Big(-l_3'\frac{e^{-K_3z}e^{-K_3z'}}{2}$ $-m_3'\frac{e^{-K_3z}e^{K_3z'}}{2}$ $+n_3'\frac{e^{K_3z}e^{-K_3z'}}{2}$ $+p_3'\frac{e^{K_3z}e^{K_3z'}}{2}\Big)$	$\frac{K_1}{K_1'}\Big(\frac{e^{K_1z}e^{-K_3z'}}{2}$ $+t_3'\frac{e^{K_1z}e^{-K_3z'}}{2}$ $+u_3'\frac{e^{K_1z}e^{K_3z'}}{2}\Big)$
I	$\frac{K_2}{K_2'}\Big(-\frac{e^{-K_2z}e^{K_1z'}}{2}$ $-s_1'\frac{e^{-K_2z}e^{K_1z'}}{2}\Big)$	$\frac{K_3}{K_3'}\Big(-\frac{e^{-K_3z}e^{K_1z'}}{2}$ $-m_1'\frac{e^{-K_3z}e^{K_1z'}}{2}$ $+p_1'\frac{e^{K_3z}e^{K_1z'}}{2}\Big)$	$\frac{K_3}{K_3'}\Big(-\frac{e^{-K_3z}e^{K_1z'}}{2}$ $-m_1'\frac{e^{-K_3z}e^{K_1z'}}{2}$ $+p_1'\frac{e^{K_3z}e^{K_1z'}}{2}\Big)$	$-\frac{K_1}{K_1'}\frac{e^{-K_1z}e^{K_1z'}}{2}$ $+u_1'\frac{K_1}{K_1'}\frac{e^{K_1z}e^{K_1z'}}{2}$

Table 6.19: Expressions for $\partial_z g^H(z,z')$ at the boundaries of the regions

Now in each region:

$$\frac{1}{\epsilon_i}\frac{K_i}{K_i'} = \frac{1}{\epsilon_i}\frac{K_i}{\frac{K_i}{\epsilon_i}} = 1. \tag{6.120}$$

Hence the boundary conditions that $\frac{1}{\epsilon}\partial_z g^H(z,z')$ has to satisfy at the interfaces are:

	Boundary Conditions — $\frac{1}{\epsilon}\partial_z g^H(z,z')$	
z z’	at $z=0$	at $z=-a$
II	$-r_2' = -l_2' + n_2'$	$(1+n_2')\,e^{-K_3 a} - l_2'\,e^{K_3 a} = (1+t_2')\,e^{-K_1 a}$
III	$s_3' = m_3' - p_3'$ $r_3' = l_3' - n_3'$	$p_3'\,e^{-K_3 a} - m_3'\,e^{K_3 a} = u_3'\,e^{-K_1 a}$ $(1+n_3')\,e^{-K_3 a} - l_3'\,e^{K_3 a} = (1+t_3')\,e^{-K_1 a}$
I	$1 + s_1' = 1 + m_1' - p_1'$	$-(1+m_1')\,e^{K_3 a} + p_1' e^{-K_3 a} = -e^{K_1 a} + u_1'\,e^{-K_1 a}$

Table 6.20: Boundary conditions for $\frac{1}{\epsilon}\partial_z\, g^H(z,z')$ to be continuous across the regions

In order to determine the coefficients, we use the following relations:

	Boundary Conditions — $g^H(z,z')$ and $\frac{1}{\epsilon}\partial_z g^H(z,z')$	
z z’	$z=0$	$z=-a$
II	$\frac{1}{K_2'}(1+r_2') = \frac{1}{K_3'}(1+l_2'+n_2')$ $r_2' = l_2' - n_2'$	$\frac{1}{K_3'}((1+n_2')e^{-K_3 a}+l_2' e^{K_3 a}) = \frac{1}{K_1'}(1+t_2')e^{-K_1 a}$ $(1+n_2')e^{-K_3 a} - l_2' e^{K_3 a} = (1+t_2')e^{-K_1 a}$
III	$\frac{1}{K_2'}(1+s_3') = \frac{1}{K_3'}(1+m_3'+p_3')$ $\frac{1}{K_2'}r_3' = \frac{1}{K_3'}(l_3'+n_3')$ $s_3' = m_3' - p_3'$ $r_3' = l_3' - n_3'$	$\frac{1}{K_3'}(p_3'\,e^{-K_3 a}+m_3'\,e^{K_3 a}) = \frac{1}{K_1'}u_3'\,e^{-K_1 a}$ $\frac{1}{K_3'}((1+n_3')e^{-K_3 a}+l_3'\,e^{K_3 a}) = \frac{1}{K_1'}(1+t_3')e^{-K_1 a}$ $p_3'\,e^{-K_3 a} - m_3'\,e^{K_3 a} = u_3'\,e^{-K_1 a}$ $(1+n_3')e^{-K_3 a} - l_3' e^{K_3 a} = (1+t_3')e^{-K_1 a}$
I	$\frac{1}{K_2'}(1+s_1') = \frac{1}{K_3'}(1+m_1'+p_1')$ $s_1' = m_1' - p_1'$	$\frac{1}{K_3'}(p_1' e^{-K_3 a}+(1+m_1')e^{K_3 a}) = \frac{1}{K_1'}(u_1' e^{-K_1 a}+e^{K_1 a})$ $p_1' e^{-K_3 a} - (1+m_1')e^{K_3 a} = u_1' e^{-K_1 a} - e^{K_1 a}$

Table 6.21: Boundary conditions needed to find the coefficients of $g^H(z,z')$

These are similar those for for $g^E(z,z')$, but with K_i replaced by K_i' except in the exponents.

Coefficients for $g^H(z, z')$

The sixteen equations in Table (6.21) can be used to find the sixteen coefficients for $g^H(z, z')$. Because the boundary conditions themselves only differ from those on $g^E(z, z')$ by the replacement just noted, the resulting coefficients for $g^H(z, z')$ also only differ from those of $g^E(z, z')$ by this substitution, as can be seen in the table below:

z' in II	$r'_2 = \frac{e^{2aK_3}(K'_1+K'_3)(K'_2-K'_3)-(K'_1-K'_3)(K'_2+K'_3)}{e^{2aK_3}(K'_1+K'_3)(K'_2+K'_3)+(K'_1-K'_3)(K'_3-K'_2)}$; $l'_2 = -\frac{2K'_3(K'_1-K'_3)}{e^{2aK_3}(K'_1+K'_3)(K'_2+K'_3)+(K'_1-K'_3)(K'_3-K'_2)}$ $n'_2 = \frac{(K'_3-K'_2)\left(e^{2aK_3}(K'_1+K'_3)-K'_1+K'_3\right)}{e^{2aK_3}(K'_1+K'_3)(K'_2+K'_3)+(K'_1-K'_3)(K'_3-K'_2)}$; $t'_2 = \frac{4K'_1K'_3e^{a(K_1+K_3)}}{(K'_1+K'_3)(K'_2+K'_3)e^{2aK_3}+(K'_1-K'_3)(K'_3-K'_2)} - 1$
z' in III	$r'_3 = -\frac{2K'_2(K'_1-K'_3)}{e^{2aK_3}(K'_1+K'_3)(K'_2+K'_3)+(K'_1-K'_3)(K'_3-K'_2)}$; $l'_3 = \frac{1}{\frac{e^{2aK_3}(K'_1+K'_3)}{K'_3-K'_1}+\frac{2K'_2}{K'_2+K'_3}-1}$ $n'_3 = \frac{1}{\frac{e^{2aK_3}(K'_1+K'_3)(K'_2+K'_3)}{(K'_3-K'_1)(K'_3-K'_2)}-1}$; $s'_3 = \frac{(K'_2-K'_3)\left(e^{2aK_3}(K'_1+K'_3)+K'_1-K'_3\right)}{e^{2aK_3}(K'_1+K'_3)(K'_2+K'_3)+(K'_1-K'_3)(K'_3-K'_2)}$ $m'_3 = \frac{1}{\frac{e^{2aK_3}(K'_1+K'_3)(K'_2+K'_3)}{(K'_3-K'_1)(K'_3-K'_2)}-1}$; $p'_3 = \frac{1}{\frac{e^{-2aK_3}(K'_1-K'_3)}{K'_1+K'_3}+\frac{K'_2+K'_3}{K'_3-K'_2}}$ $t'_3 = \frac{2K'_1(K'_2+K'_3)e^{a(K_1+K_3)}}{(K'_1+K'_3)(K'_2+K'_3)e^{2aK_3}+(K'_1-K'_3)(K'_3-K'_2)} - 1$; $u'_3 = \frac{2K'_1(K'_3-K'_2)e^{a(K_1+K_3)}}{e^{2aK_3}(K'_1+K'_3)(K'_2+K'_3)+(K'_1-K'_3)(K'_3-K'_2)}$
z' in I	$s'_1 = \frac{4K'_2K'_3e^{a(K_1+K_3)}}{(K'_1+K'_3)(K'_2+K'_3)e^{2aK_3}+(K'_1-K'_3)(K'_3-K'_2)} - 1$; $m'_1 = \frac{2K'_3(K'_2+K'_3)e^{a(K_1+K_3)}}{(K'_1+K'_3)(K'_2+K'_3)e^{2aK_3}+(K'_1-K'_3)(K'_3-K'_2)} - 1$ $p'_1 = \frac{2K'_3(K'_3-K'_2)e^{a(K_1+K_3)}}{(K'_1+K'_3)(K'_2+K'_3)e^{2aK_3}+(K'_1-K'_3)(K'_3-K'_2)}$; $u'_1 = \frac{e^{2aK_1}\left((K'_1-K'_3)(K'_2+K'_3)e^{2aK_3}-(K'_1+K'_3)(K'_2-K'_3)\right)}{(K'_1+K'_3)(K'_2+K'_3)e^{2aK_3}+(K'_1-K'_3)(K'_3-K'_2)}$

Table 6.22: Coefficients of $g^H(z, z')$

As was true for $g^E(z, z')$, there is an alternative way of obtaining the coefficients for $g^H(z, z')$, since it too is symmetrical with respect to z and z'. Instead of using the continuity with respect to z of $g^H(z, z')$ and $\frac{1}{\epsilon}\partial_z g^H(z, z')$ as was done here, one can rely on the continuity with respect to z' of $g^H(z, z')$ and $\frac{1}{\epsilon'}\partial_{z'} g^H(z, z')$, matching these two functions at $z' = 0$ and $z' = -a$. Again, this method yields the results in Table (6.22).

Complete forms of $g^E(z,z')$ and $g^H(z,z')$ in the regions of interest

To summarize the results of the previous two sections:

- the form of $g^E(z,z')$ is given in Table (6.5), with the coefficients of Table (6.13).
- the form of $g^H(z,z')$ is given in Table (6.15), with the coefficients of Table (6.22).

What we are interested in however are fields in region III due to sources in regions I and II (with z in region III and z' in regions I and II). Therefore the functions of interest to us are:

z z’	III	
	$g^E(z,z')$	Coefficients
II	$\frac{e^{K_3 z} e^{-K_2 z'}}{2K_3}$ $+l_2 \frac{e^{-K_3 z} e^{-K_2 z'}}{2K_3}$ $+n_2 \frac{e^{K_3 z} e^{-K_2 z'}}{2K_3}$	$l_2 = -\frac{2K_3(K_1-K_3)}{e^{2aK_3}(K_1+K_3)(K_2+K_3)+(K_1-K_3)(K_3-K_2)}$ $= \frac{2K_3}{(K_2-K_3)} \left(\frac{1}{1-\frac{(K_3+K_1)(K_3+K_2)e^{2aK_3}}{(K_3-K_1)(K_3-K_2)}} \right)$ $n_2 = \frac{(K_3-K_2)\left(e^{2aK_3}(K_1+K_3)-K_1+K_3\right)}{e^{2aK_3}(K_1+K_3)(K_2+K_3)+(K_1-K_3)(K_3-K_2)}$ $= \frac{K_3-K_2}{K_3+K_2} - \frac{2K_3}{(K_2+K_3)} \left(\frac{1}{1-\frac{(K_3+K_1)(K_3+K_2)e^{2aK_3}}{(K_3-K_1)(K_3-K_2)}} \right)$
I	$\frac{e^{-K_3 z} e^{K_1 z'}}{2K_3}$ $+m_1 \frac{e^{-K_3 z} e^{K_1 z'}}{2K_3}$ $+p_1 \frac{e^{K_3 z} e^{K_1 z'}}{2K_3}$	$m_1 = \frac{2K_3(K_2+K_3)e^{a(K_1+K_3)}}{(K_1+K_3)(K_2+K_3)e^{2aK_3}+(K_1-K_3)(K_3-K_2)} - 1$ $= \frac{2K_3 e^{aK_1}}{(K_1-K_3)\left(\frac{K_3-K_2}{K_3+K_2}\right)e^{-aK_3}+(K_1+K_3)e^{aK_3}} - 1$ $p_1 = \frac{2K_3(K_3-K_2)e^{a(K_1+K_3)}}{(K_1+K_3)(K_2+K_3)e^{2aK_3}+(K_1-K_3)(K_3-K_2)}$ $= \frac{2K_3 e^{aK_1}}{(K_1+K_3)\left(\frac{K_3+K_2}{K_3-K_2}\right)e^{aK_3}+(K_1-K_3)e^{-aK_3}}$

Table 6.23: $g^E(z,z')$ in region III due to sources in regions I and II

Note that:

$$n_2 = \frac{K_3 - K_2}{K_3 + K_2}(1 + l_2) \tag{6.121a}$$

$$m_1 = -1 + \frac{K_3 + K_2}{K_3 - K_2}\, p_1 \tag{6.121b}$$

As for $g^H(z, z')$, it has the same form with K_i' substituted for K_i everywhere except in the exponents.

6.4 Deriving the expression for $\langle E_i E_j \rangle$ from the Green's functions $\Gamma_{ij}(\mathbf{r}', \mathbf{r}', \omega)$

Having derived expressions for the Green's functions, we are now in a position to calculate the expressions for $\langle E_i E_j \rangle$ and $\langle B_i B_j \rangle$ in source theory. We will then be able to calculate the index of refraction as indicated in the first section of the present chapter, just as Barton obtained it from his $\langle E_i E_j \rangle$ and $\langle B_i B_j \rangle$ which involved the zero-point fields instead of source fields [6]. Following Milonni and Shih, we ensure that our quantized source fields alone appear in the derivation by normal ordering the field operators [17]. Consequently, the relevant expressions are of the form: $\langle 0|E^- E^+|0\rangle$.

6.4.1 Derivation of the expression for $\left\langle E_i^-(r,t) E_j^+(r,t) \right\rangle$

We first seek what form $\left\langle E_i^-(r,t) E_j^+(r,t) \right\rangle$ takes. Because the electric field is due to sources in our model, it can be expressed in terms of Green's functions. The positive and negative frequency parts of the quantized electric field can then in turn be expressed in terms of positive and negative frequency parts of Green's functions:

$$\begin{aligned} &\langle E_i^-(r,t) \;\; E_j^+(r,t) \rangle \\ = \;\; &\left\langle 8\pi \int_{-\infty}^{+\infty} d^3r'' \int_0^t dt'' G_{km}^-(r, r''; t, t'').P(r'', t'') \;\; 8\pi \int_{-\infty}^{+\infty} d^3r' \int_0^t dt' G_{nl}^+(r, r'; t, t').P(r', t') \right\rangle \end{aligned} \tag{6.122}$$

where:

$$
\begin{aligned}
G^{-}_{km}(r,r'';t,t'') &= \frac{1}{2\pi}\int_0^{+\infty} d\omega_b\,\Gamma_{km}(r,r'',\omega_b)\; e^{+i\omega_b(t-t'')} && (6.123)\\
G^{+}_{nl}\,(r,r'\,;t,\;t') &= \frac{1}{2\pi}\int_0^{+\infty} d\omega_a\,\Gamma_{nl}(r,r',\omega_a)\; e^{-i\omega_a(t-t')}. && (6.124)
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
&\left\langle E_i^-(r,t)\;\; E_j^+(r,t)\right\rangle\\
&= \Bigg\langle 8\pi\int_{-\infty}^{+\infty} d^3r''\int_0^t dt''\left(\frac{1}{2\pi}\int_0^{+\infty} d\omega_b\,\Gamma_{km}(r,r'',\omega_b)\; e^{+i\omega_b(t-t'')}\right).P(r'',t'')\\
&\quad 8\pi\int_{-\infty}^{+\infty} d^3r'\int_0^t dt'\left(\frac{1}{2\pi}\int_0^{+\infty} d\omega_a\,\Gamma_{nl}(r,r',\omega_a)\; e^{-i\omega_a(t-t')}\right).P(r',t')\Bigg\rangle\\
&= 16\Bigg\langle\int_{-\infty}^{+\infty} d^3r''\int_0^t dt''\int_0^{+\infty} d\omega_b\,\Gamma_{km}(r,r'',\omega_b)\; e^{+i\omega_b(t-t'')}.P(r'',t'')\\
&\quad\int_{-\infty}^{+\infty} d^3r'\int_0^t dt'\int_0^{+\infty} d\omega_a\,\Gamma_{nl}(r,r',\omega_a)\; e^{-i\omega_a(t-t')}.P(r',t')\Bigg\rangle\\
&= 16\int_{-\infty}^{+\infty} d^3r''\int_0^t dt''\int_0^{+\infty} d\omega_b\,\Gamma_{km}(r,r'',\omega_b)\; e^{+i\omega_b(t-t'')}\\
&\quad\int_{-\infty}^{+\infty} d^3r'\int_0^t dt'\int_0^{+\infty} d\omega_a\,\Gamma_{nl}(r,r',\omega_a)\; e^{-i\omega_a(t-t')}\,\langle P(r'',t'')P(r',t')\,\rangle \quad (6.125)
\end{aligned}
$$

The quantity $\langle P(r'',t'')P(r',t')\,\rangle$ represents the effect of fluctuations in the polarization. Physically these are microscopic fluctuations in the charge distribution within the metal plates. In the model, they are represented by microscopic dipoles, that see their dipole moments fluctuate randomly. The expectation values of the resulting polarizations, $\langle P(r',t')\,\rangle$ or $\langle P(r'',t'')\,\rangle$, vanish: the dipole moment of a given dipole points sometimes this way, as often the other way, so overall its expectation value is zero. Therefore so is the expectation value of the corresponding macroscopic variable, the polarization $\langle P(r',t')\,\rangle$. However, $\langle P(r'',t'')P(r',t')\,\rangle$ is not necessarily zero. It only vanishes when the polarizations involved are not at the same place and time, or in the same direction:

$$\langle P(r'',t'')P(r',t')\,\rangle \quad = \quad \delta_{mn} f(r')\delta^3(r'-r'')\delta(t'-t'') \tag{6.126}$$

where the function $f(r')$ depends on the intensity of the polarization at r'. It is of the form:

$$f(r') = i\,\frac{\hbar}{2}\,\alpha\,\langle\, d_m(t'')\, d_n(t')\,\rangle\, N(r'). \tag{6.127}$$

Substituting Eq. (6.126) in Eq. (6.125) for our quantity of interest, one obtains:

$$\begin{aligned} & \left\langle E_i^-(r,t)\ \ E_j^+(r,t)\right\rangle \\ = \;& 16 \int_{-\infty}^{+\infty} d^3r'' \int_0^t dt'' \int_0^{+\infty} d\omega_b\, \Gamma_{km}(r,r'',\omega_b)\ e^{+i\omega_b(t-t'')} \\ & \int_{-\infty}^{+\infty} d^3r' \int_0^t dt' \int_0^{+\infty} d\omega_a\, \Gamma_{nl}(r,r',\omega_a)\ e^{-i\omega_a(t-t')}\ \delta_{mn}\, f(r')\delta^3(r'-r'')\delta(t'-t''). \end{aligned}$$

Integrating with respect to r'' is unproblematic:

$$\int_{-\infty}^{+\infty} d^3r''\, \Gamma_{km}(r,r'',\omega_b)\, \delta^3(r'-r'') = \Gamma_{km}(r,r',\omega_b). \tag{6.128}$$

However one has to be careful when integrating over time, because the integral over a delta function is defined with limits at infinity, whereas our time integrals have finite limits, at 0 and t. This means that these integrals have to be replaced by equivalent expressions involving theta functions instead of those finite limits.

$$\int_0^t dt'' \int_0^t dt'\ e^{+i\omega_b(t-t'')}\ e^{-i\omega_a(t-t')}\ \delta(t'-t'')\ : \tag{6.129}$$

$$\begin{aligned}\int_0^t dt'' \; e^{+i\omega_b(t-t'')} \; \delta(t'-t'') &= \int_{-\infty}^{+\infty} dt'' \; e^{+i\omega_b(t-t'')} \; \delta(t'-t'') \; \Theta(t'')\Theta(-t''+t) \\ &= e^{+i\omega_b(t-t')} \; \Theta(t')\Theta(-t'+t). \qquad (6.130)\end{aligned}$$

So:

$$\begin{aligned}&\int_0^t dt'' \; \int_0^t dt' \; e^{+i\omega_b(t-t'')} \; e^{-i\omega_a(t-t')} \; \delta(t'-t'') \\ =\; &\int_0^t dt' \; e^{-i\omega_a(t-t')} \; e^{+i\omega_b(t-t')} \; \Theta(t')\Theta(-t'+t) \\ =\; &\int_0^t dt' \; e^{-i\omega_a(t-t')} \; e^{+i\omega_b(t-t')}. \qquad (6.131)\end{aligned}$$

Getting back to the expression we want:

$$\begin{aligned}&\left\langle E_i^-(r,t) \;\; E_j^+(r,t)\right\rangle \\ =\;& 16 \int_{-\infty}^{+\infty} d^3r' \int_0^t dt' \int_0^{+\infty} d\omega_a \, \Gamma_{im}(r,r',\omega_a) \; e^{-i\omega_a(t-t')} \\ &\qquad \int_0^{+\infty} d\omega_b \, \Gamma_{mj}(r,r',\omega_b) \; e^{+i\omega_b(t-t')} \; f(r') \\ =\;& 16 \int_{-\infty}^{+\infty} d^3r' \int_0^t dt' \int_0^{+\infty} d\omega_a \int_0^{+\infty} d\omega_b \\ &\qquad \Gamma_{im}(r,r',\omega_a) \; \Gamma_{mj}(r,r',\omega_b) \; e^{-i\omega_a(t-t')} \; e^{+i\omega_b(t-t')} \; f(r') \\ =\;& 16 \int_{-\infty}^{+\infty} d^3r' \int_0^t dt' \int_0^{+\infty} d\omega_a \int_0^{+\infty} d\omega_b \\ &\qquad \Gamma_{im}(r,r',\omega_a) \; \Gamma_{mj}(r,r',\omega_b) \; e^{-i(\omega_a-\omega_b)t} \; e^{+i(\omega_a-\omega_b)t'} \; f(r'). \\ &\qquad (6.132)\end{aligned}$$

Now:

$$\int_0^t dt' e^{+i(\omega_a-\omega_b)t'} \quad = \quad \left.\frac{e^{+i(\omega_a-\omega_b)t'}}{i(\omega_a-\omega_b)}\right|_{t'=0}^{t'=t} \quad = \quad \frac{e^{+i(\omega_a-\omega_b)t}-1}{i(\omega_a-\omega_b)} \tag{6.133}$$

so:

$$\begin{aligned} e^{-i(\omega_a-\omega_b)t}\int_0^t dt' e^{+i(\omega_a-\omega_b)t'} &= e^{-i(\omega_a-\omega_b)t} \; \frac{e^{+i(\omega_a-\omega_b)t}-1}{i(\omega_a-\omega_b)} \\ &= \frac{1-e^{-i(\omega_a-\omega_b)t}}{i(\omega_a-\omega_b)} \end{aligned} \tag{6.134}$$

and therefore our expression becomes:

$$\begin{aligned} &\left\langle E_i^-(r,t) \;\; E_j^+(r,t)\right\rangle \\ &= \; 16\int_{-\infty}^{+\infty} d^3r' \int_0^{+\infty} d\omega_a \int_0^{+\infty} d\omega_b \; \Gamma_{im}(r,r',\omega_a)\Gamma_{mj}(r,r',\omega_b) \; \frac{1-e^{-i(\omega_a-\omega_b)t}}{i(\omega_a-\omega_b)} \; f(r'). \end{aligned} \tag{6.135}$$

Now from the Riemann Lebesgue lemma, if *f(x)* is L^1 integrable on $\mathbb{R}^d$, then: $\int_{\mathbb{R}^d} f(x)e^{iz.x}dx \to 0$ as $|z| \to \infty$. So in our case:

$$\int_0^{+\infty} d\omega_a \, \frac{e^{-i\omega_a t}}{i(\omega_a-\omega_b)} = \int_0^{+\infty} d\omega_a \; \Gamma_{im}(r,r',\omega_a) \; \frac{e^{i\omega_a(-t)}}{i(\omega_a-\omega_b)} \to 0 \text{ as } t = |-t| \to \infty \tag{6.136a}$$

$$\int_0^{+\infty} d\omega_b \, \frac{e^{-i(-\omega_b)t}}{i(\omega_a-\omega_b)} = \int_0^{+\infty} d\omega_b \; \Gamma_{mj}(r,r',\omega_b) \; \frac{e^{i\omega_b t}}{i(\omega_a-\omega_b)} \to 0 \text{ as } t = |t| \to \infty \tag{6.136b}$$

and therefore:

$$\int_0^{+\infty} d\omega_a \int_0^{+\infty} d\omega_b \;\Gamma_{im}(r,r',\omega_a)\;\Gamma_{mj}(r,r',\omega_b)\,\frac{e^{-i(\omega_a-\omega_b)t}}{i(\omega_a-\omega_b)} \;=\; 0. \quad (6.137)$$

So we get for our expression:

$$\begin{aligned} &\langle E_i^-(r,t)\;\;E_j^+(r,t)\rangle \\ =\;& -i16\int_{-\infty}^{+\infty} d^3r' \int_0^{+\infty} d\omega_a \int_0^{+\infty} d\omega_b\, \Gamma_{im}(r,r',\omega_a)\Gamma_{mj}(r,r',\omega_b)\,\frac{1}{\omega_a-\omega_b}\,f(r') \\ =\;& -i16\int_{-\infty}^{+\infty} dz' \int_0^{+\infty} d^2r'_\perp \int_0^{+\infty} d\omega_a \int_0^{+\infty} d\omega_b\, \Gamma_{im}(z,z',r'_\perp,w_a)\Gamma_{mj}(z,z',r'_\perp,\omega_b)\,\frac{1}{\omega_a-\omega_b}\,f(r'). \end{aligned} \quad (6.138)$$

In order to have expressions in momentum space, we take the Fourier transforms:

$$\Gamma_{im}(z,z',r'_\perp,w_a) \;=\int_{-\infty}^{+\infty} \frac{d^2k_{1\perp}}{(2\pi)^2}\,e^{ik_{1\perp}(r_\perp-r'_\perp)}\,\Gamma_{im}(z,z',k_{1\perp},w_a) \quad (6.139a)$$

$$\Gamma_{mj}(z,z',r'_\perp,\omega_b) \;=\int_{-\infty}^{+\infty} \frac{d^2k_{2\perp}}{(2\pi)^2}\,e^{ik_{2\perp}(r_\perp-r'_\perp)}\,\Gamma_{mj}(z,z',k_{2\perp},\omega_b). \quad (6.139b)$$

Substituting these:

$$\begin{aligned} &\langle E_i^-(r,t)\;\;E_j^+(r,t)\rangle \\ =\;& -i16\int_{-\infty}^{+\infty} dz' \int_0^{+\infty} d^2r'_\perp \int_0^{+\infty} d\omega_a \int_0^{+\infty} d\omega_b \int_{-\infty}^{+\infty} \frac{d^2k_{1\perp}}{(2\pi)^2} \int_{-\infty}^{+\infty} \frac{d^2k_{2\perp}}{(2\pi)^2} \\ & e^{ik_{1\perp}(r_\perp-r'_\perp)}\;e^{ik_{2\perp}(r_\perp-r'_\perp)}\;\Gamma_{im}(z,z',k_{1\perp},w_a)\Gamma_{mj}(z,z',k_{2\perp},\omega_b)\,\frac{1}{\omega_a-\omega_b}\,f(z'). \end{aligned} \quad (6.140)$$

We then integrate over $r'_\perp$, which as we see below sets $k_{1\perp} = -k_{2\perp}$:

$$\int_0^{+\infty} d^2r'_\perp e^{ik_{1\perp}(r_\perp - r'_\perp)}\ e^{ik_{2\perp}(r_\perp - r'_\perp)} \ = \ e^{i(k_{1\perp}+k_{2\perp})r_\perp} \int_0^{+\infty} d^2r'_\perp e^{-i(k_{1\perp}+k_{2\perp})r'_\perp}. \tag{6.141}$$

Now:

$$\int_0^{+\infty} d^2r'_\perp e^{-i(k_{1\perp}+k_{2\perp})r'_\perp} \ = \ \delta(k_{1\perp}+k_{2\perp}). \tag{6.142}$$

So we set $k_{1\perp} = -k_{2\perp}$, *i.e.* $k_{1\perp} = k_\perp$, $k_{2\perp} = -k_\perp$. Then:

$$\begin{aligned}
& \int_0^{+\infty} d^2r'_\perp \int_{-\infty}^{+\infty} \frac{d^2k_{1\perp}}{(2\pi)^2} \int_{-\infty}^{+\infty} \frac{d^2k_{2\perp}}{(2\pi)^2}\ e^{ik_{1\perp}(r_\perp - r'_\perp)}\ e^{ik_{2\perp}(r_\perp - r'_\perp)}\ \Gamma_{im}(z,z',k_{1\perp},w_a)\Gamma_{mj}(z,z',k_{2\perp},\omega_b) \\
= \ & \int_{-\infty}^{+\infty} \frac{d^2k_{1\perp}}{(2\pi)^2} \int_{-\infty}^{+\infty} \frac{d^2k_{2\perp}}{(2\pi)^2}\ e^{i(k_{1\perp}+\,k_{2\perp})\,r_\perp} \int_0^{+\infty} d^2r'_\perp\ e^{-i(k_{1\perp}+\,k_{2\perp})\,r'_\perp}\ \Gamma_{im}(z,z',k_{1\perp},w_a)\Gamma_{mj}(z,z',k_{2\perp},\omega_b) \\
= \ & \int_{-\infty}^{+\infty} \frac{d^2k_{1\perp}}{(2\pi)^2} \int_{-\infty}^{+\infty} \frac{d^2k_{2\perp}}{(2\pi)^2}\ e^{i(k_{1\perp}+\,k_{2\perp})\,r_\perp}\ \delta(k_{1\perp}+k_{2\perp})\,\Gamma_{im}(z,z',k_{1\perp},w_a)\Gamma_{mj}(z,z',k_{2\perp},\omega_b) \\
= \ & \int_{-\infty}^{+\infty} \frac{d^2k_\perp}{(2\pi)^2}\ e^{i(k_\perp -\, k_\perp)\,r_\perp}\ \ \Gamma_{im}(z,z',k_\perp,w_a)\,\Gamma_{mj}(z,z',-k_\perp,\omega_b) \\
= \ & \int_{-\infty}^{+\infty} \frac{d^2k_\perp}{(2\pi)^2}\,\Gamma_{im}(z,z',k_\perp,w_a)\,\Gamma_{mj}(z,z',-k_\perp,\omega_b)
\end{aligned} \tag{6.143}$$

and we now get for our expression:

$$\begin{aligned}
& \left\langle E_i^-(r,t)\ \ E_j^+(r,t)\right\rangle \\
= \ & -i16 \int_{-\infty}^{+\infty} dz' \int_0^{+\infty} d\omega_a \int_0^{+\infty} d\omega_b \int_{-\infty}^{+\infty} \frac{d^2k_\perp}{(2\pi)^2}\,\Gamma_{im}(z,z',k_\perp,w_a)\Gamma_{mj}(z,z',k_\perp,\omega_b)\ \frac{1}{\omega_a-\omega_b}\,f(z').
\end{aligned} \tag{6.144}$$

6.5 Conclusion

Completing the derivation requires making a choice for the function $f(z')$ in Eq.(6.144). Concretely, this means making choose the functional form of $N(r')$ in

Eq. (6.127), which represents the spatial distribution of the sources of polarization. The simplest choice would be:

$$N(r') = N(z') = \frac{\epsilon(z') - 1}{4\,\pi\alpha}, \tag{6.145}$$

with ϵ taking different values in the three different regions (i.e. the two semi-infinite regions representing the plates, and the space between them).

However it turns out that with this choice for $N(r')$, Eq. (6.144) yields a divergent result when we take the conducting limit in the regions representing the plates. Hopefully, a different choice for $N(r')$ could yield a sensible result, but at present it is not possible to draw any conclusion regarding whether the Scharnhorst effect can be recovered within the framework of source theory.

Chapter 7

CONCLUSION

The Scharnhorst effect predicts that photons propagating through a Casimir vacuum would do so faster than c.[1] Strictly speaking, it concerns the phase velocity of these photons, in so far that this is the quantity that was first derived by Klaus Scharnhorst [5], and soon thereafter Gabriel Barton [6]. That a phase velocity should be superluminal is not especially remarkable;[2] however, as we discussed in the Introduction (chapter 1) and in greater details in chapter 3, whether the same result holds for the signal velocity has been the object of debate in the literature. Arguments to the effect that it did not were presented within months of the original result [10], [11]. When Scharnhorst and Barton responded, they did not present conclusive evidence that the signal velocity was also superluminal, but argued that this was one of two unusual possibilities. About a decade after Klaus Scharnhorst and Gabriel Barton's work, Liberati *et al* presented re-derivations of the effect, and notably calculated the velocity at which the wavefront of Scharnhorst photons propagates from an effective metric approach [14, 13, 15, 16]. For a classical propagating wave, the signal velocity is usually identified with the wavefront velocity. However their result was subject to the same caveat as Scharnhorst's and Barton's. Let us

---

[1] Klaus Scharnhorst was the first to derive the effect that bears his name in February 1990; a month later, Gabriel Barton offered an alternative derivation. They found that photons travelling in a Casimir vacuum would do so faster than c (unless their motion was parallel to the plates). Scharnhorst and Barton calculated the magnitude of the shift in velocity perpendicular to the plates – for which the photons are fastest [5], [6].

[2] As explained earlier, the term "superluminal" is here used in the sense of "faster than the speed of light in trivial vacuum", i.e. c. Although talking of superluminal photons may seem unfortunate in so far that it smacks of being a contradiction in terms, this use has already become standard in the relevant literature.

briefly review the specific issues, on which the debate focused.

7.1 Reminder: Debate regarding the signal velocity in the physics literature

7.1.1 Klaus Scharnhorst and Gabriel Barton's derivations

Both Klaus Scharnhorst and Gabriel Barton first calculated the index of refraction for the Casimir vacuum (thereby treating it like a dielectric medium), and obtained the velocity of the photon from it. The velocity one finds in this way is the phase velocity $v_\Phi = \frac{c}{n}$.

Recall from chapter 3 that in order to get the index of refraction, they derived expressions for the permittivity and the permeability of the Casimir vacuum. However the details of how Scharnhorst and Barton did so differed.

Scharnhorst obtained an expression for the effective action of the photon, and then compared it to the known expression for the action given in terms of the permittivity (ε_{ij}) and permeability (μ_{ij}) tensors. Deriving the former expression for the action meant taking into account how it is modified by the presence of the plates, compared to its trivial vacuum counterpart. This notably involved calculating the two-loop polarization tensor, which represents the sum of the amplitudes for the two quantum processes responsible for the effect. Hence Scharnhorst's calculation was mostly based on quantities and concepts from quantum field theory.[3]

In contrast, Barton obtained the permittivity and permeability tensors from the corresponding susceptibility tensors $\chi^{(e)}$ and $\chi^{(m)}$, which he in turn calculated from the polarization P and magnetization M of the Casimir vacuum. He derived expressions for the latter by varying the correction to the Lagrangian with respect to the electric and magnetic field, i.e. $P = \frac{\delta \Delta L}{\delta E}$, $M = \frac{\delta \Delta L}{\delta B}$, where L is the Euler-Heisenberg

---

[3] Again, these amplitudes differ from their trivial vacuum counterpart. This is because the presence of the plates affects the propagator of the virtual photons.

effective Lagrangian. The resulting polarization and magnetization involve vacuum expectation values (VEVs) of products of E and B; it is these VEVs that are affected by the presence of the Casimir plates. The derivation made no explicit use of QFT concepts such as Feynman diagrams, photon propagators, or polarization tensor.

Hence Scharnhorst and Barton's derivations differ significantly in their detail. Notably, the physical quantities directly affected by the plates differ. In Scharnhorst's work, it is the the propagators of the virtual photons which are subject to boundary conditions, whereas in Barton's the latter are imposed on the background electromagnetic field whose VEVs are then used to find the index of refraction.

However the two calculations share an important feature, which the later derivation by Liberati *et al* also has: they are based on the Euler-Heisenberg effective Lagrangian. The latter involves the leading order terms of a power expansion in $\frac{E}{m_e}$ (where E is the typical energy of the field and m_e is the electron mass), so it is only valid when $E << m_e$. This implies that their result is only guaranteed to hold in the regime $\omega << m_e$ where ω is the frequency of the photons, including the Scharnhorst, propagating photon. Hence it cannot be guaranteed to hold for a Scharnhorst photon of arbitrarily large energy. Furthermore, on the low energy end, the eikonal approximation (needed to describe the dynamics of the photon field in terms of propagation along a trajectory) requires that $1/L << \omega$.[4]

Since the velocity derived by both Scharnhorst and Barton is simply a phase velocity, one may wonder what justified interest in whether it could also be the signal velocity. The reason for this is that the expression obtained for the index of refraction implied that it is also the group velocity – as both Scharnhorst and Barton pointed out. Indeed this expression is not frequency dependent, indicating that there is no

---

[4] where L is the plate separation, ω the frequency of any of the photons (be they propagating Scharnhorst photons or virtual ones), which gives their energy, and m_e is the electron rest mass.

dispersion “in the approximation considered” [5].

This motivated Barton to state that the signal velocity was the same as well, hence superluminal, which he argued was not in conflict with Lorentz invariance – essentially because the presence of the plates results in a soft-breaking of this symmetry in the direction normal to them.

7.1.2 Reactions in the physics literature

Reactions to the effect that the phase velocity result could not constitute a signal velocity appeared a few months later [10], [11].[5] They offered three different arguments, two of which were based on the idea that uncertainties would make it impossible to measure the signal velocity as superluminal. Both of these proofs relied on the fact that Scharnhorst and Barton's calculations were only valid in the photon frequency regime $1/L << \omega << m_e$.

The first one, given by Peter Milonni and Karl Svozil, is based on the time-energy uncertainty relation. They argued that the fundamental quantum uncertainty the relation implies for our ability to measure time leads to a corresponding uncertainty in the measurement of the signal velocity of a Scharnhorst photon, which is larger than the velocity difference from c predicted by Scharnhorst and Barton's calculations for all reasonable values of plate separation.

The uncertainty in the velocity measurement is minimal, when the travel distance is greatest. However, allowing for reflection off the plates complicates the analysis too much to draw reliable conclusions. So Milonni and Svozil considered this distance to be bounded from above by the plate separation, L. This assumption made, a given uncertainty in the time measurement dictates a specific uncertainty in the signal velocity measurement. Wishing to minimize the time uncertainty, they chose

---

[5] Scharnhorst's work was published in February 1990, Barton's in March, the paper by Peter Milonni and Karl Zvozil in October, and Shahar Ben-Menahem's in November.

it to be only associated with the emission event of a photon by an atom, as the latter transitions from an excited state to a lower energy level. The time-energy uncertainty relations dictate that the energy difference ΔE between the two levels imply an uncertainty Δt in the time of emission, which translates into a velocity measurement uncertainty Δv. Milonni and Svozil derived an expression for the ratio $\frac{\Delta v}{\Delta c}$, where Δc is the velocity difference from c predicted by Scharnhorst and Barton.[6] For the signal velocity to be measurably greater than c, this ratio needs to be smaller than 1, otherwise the Scharnhorst velocity shift Δc will get "washed out" by the uncertainty in the measurement Δv. They found that $\frac{\Delta v}{\Delta c}$ is proportional to λL^3 (where λ is the wavelength of the Scharnhorst photon and recall that L is the separation between the two plates), all other factors in the expression being constants. To minimize λ they appealed to the fact that Scharnhorst and Barton's derivations are restricted to the frequency regime $\omega << m_e$. This entails that the photon wavelength is bounded from below: $\lambda >> \lambda_c$, so they considered λ_c.[7] For L they took λ itself to be the smallest distance one might wish to consider, because as we saw the frequency is bounded from below $1/L << \omega$.[8] So the smallest plate separation they considered was λ_c. They showed that taken together, even these extreme assumptions regarding λ and L imply that the ratio $\frac{\Delta v}{\Delta c}$ would be much too large to allow measuring the small shift in velocity predicted by Scharnhorst and Barton ($\frac{\Delta v}{\Delta c} \geq 1.5 \times 10^6$) – in fact it would mean an uncertainty in the velocity measurement greater than c itself, or put another way an uncertainty in the time measurement

---

[6] $\Delta v \sim c\frac{\lambda}{L}$, where λ is the wavelength of the Scharnhorst photon; Milonni and Svozil then derived the ratio $\frac{\Delta v}{\Delta c}$to be:

$$\frac{\Delta v}{\Delta c} = \frac{\lambda}{\kappa\alpha^2}\frac{1}{\lambda_c}\left(\frac{L}{\lambda_c}\right)^3,$$

where $\lambda_c = \frac{\hbar}{mc}$ is the Compton wavelength, and κ is the constant $\frac{11\pi^2}{2^2 3^4 5^2}$. [10], p.438.

[7] This entails $\frac{\Delta v}{\Delta c} \geq 1.5 \times 10^6 \frac{L^3}{\lambda_c^3}$.

[8] They also discussed the more reasonable option of using the Bohr radius instead, and noted that even in this case the plates no longer really constitute distinct macroscopic objects [10], p.438.

greater than the travel time from one plate to the other.

Shahar Ben-Menahem too argued that measuring a superluminal signal velocity with Scharnhorst photons would be impossible due to issues relating to uncertainty and measurement – albeit not the quantum uncertainty relations [11]. His argument relies as well on the frequency regime of the photons, and that the travel distance of the photon can at most be L.

Ben-Menahem used the fact that the signal velocity is the velocity of a *sharp* wavefront, i.e. essentially a step-function. He argued that the frequency regime $\omega << m_e$ for which the effect was calculated precludes forming such a wavefront, hence using it to signal superluminally: this would require forming a wave packet with an infinitely-sharp front, which in turn necessitates using component modes of arbitrarily low frequencies. The wavefront of any wave packet we can form using photon modes in the allowed regime is "smeared" – at best on the order of the scale $1/m_e$. Therefore, in order to signal faster than c, the wavefront must "move beyond the light cone"[9] by a distance $\delta x > 1/m_e$. The idea can be illustrated as in Fig. 7.1 below, where the blurred region at the top right represents the wavefront.

Ben-Menahem as well restricted his analysis to photons travelling from one plate to the other without reflecting, so that the time of travel, t, can at most be equal to L (in natural units where $c = 1$).

He found that under these circumstances, the superluminal signalling condition $\delta x > 1/m_e$ entailed the requirement $\zeta\, e^4 \frac{1}{(m_e L)^3} > 1$ (where ζ is a constant).[10]

---

[9] That is, in a given time it must move an additional distance compared to what it would if it traveled at speed c.

[10] $\zeta = \frac{11}{(2^6)(45)^2}$.

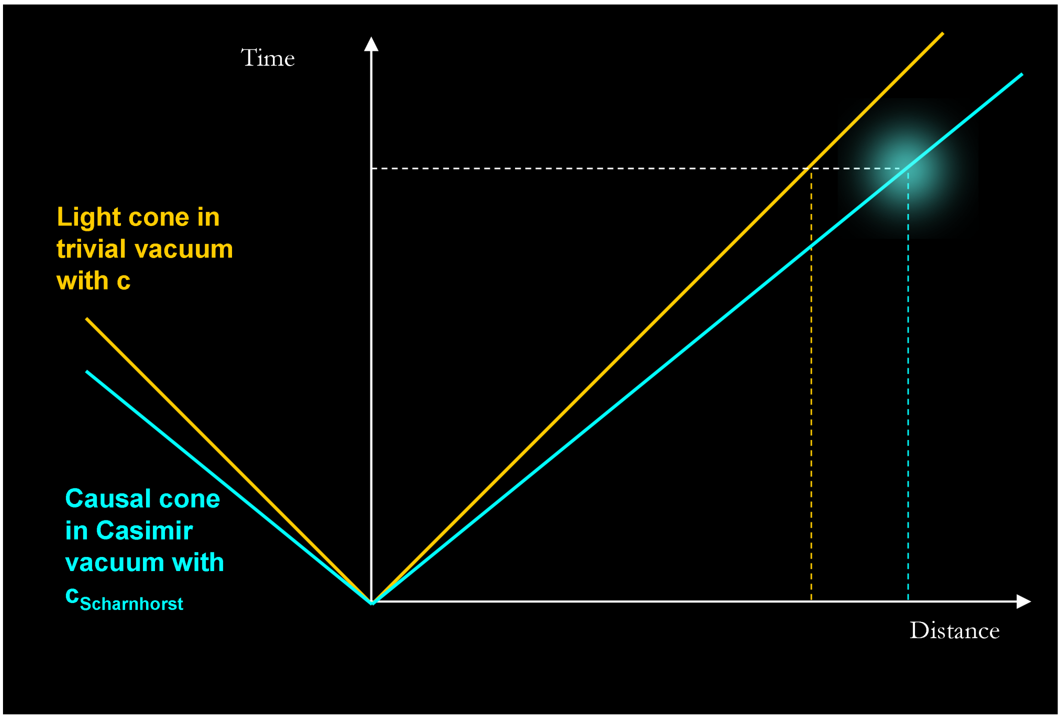


Figure 7.1: The wavefront is “smeared” over a distance of $1/m_e$, but has moved measurably “beyond the light cone.”

The reason why this inequality cannot be satisfied is, again, that the Scharnhorst velocity was derived for the regime $\frac{1}{L} << \omega << m_e$. This implies $m_e L >> 1$, which is at odds with the requirement for superluminal signalling derived by Ben-Menahem. The conclusion is that the “smearing” of the wavefront precludes measuring the signal speed as superluminal over L, which can be illustrated as follows (note that over a travel distance equal to the plates separation the blurring of the wavefront overlaps with the light cone):[11]

---

[11] The smearing is represented as being in both space and time so the illustration may either Milonni and Svozil or Ben-Menahem’s arguments; recall that, respectively, they discuss uncertainties in time and distance measurements.

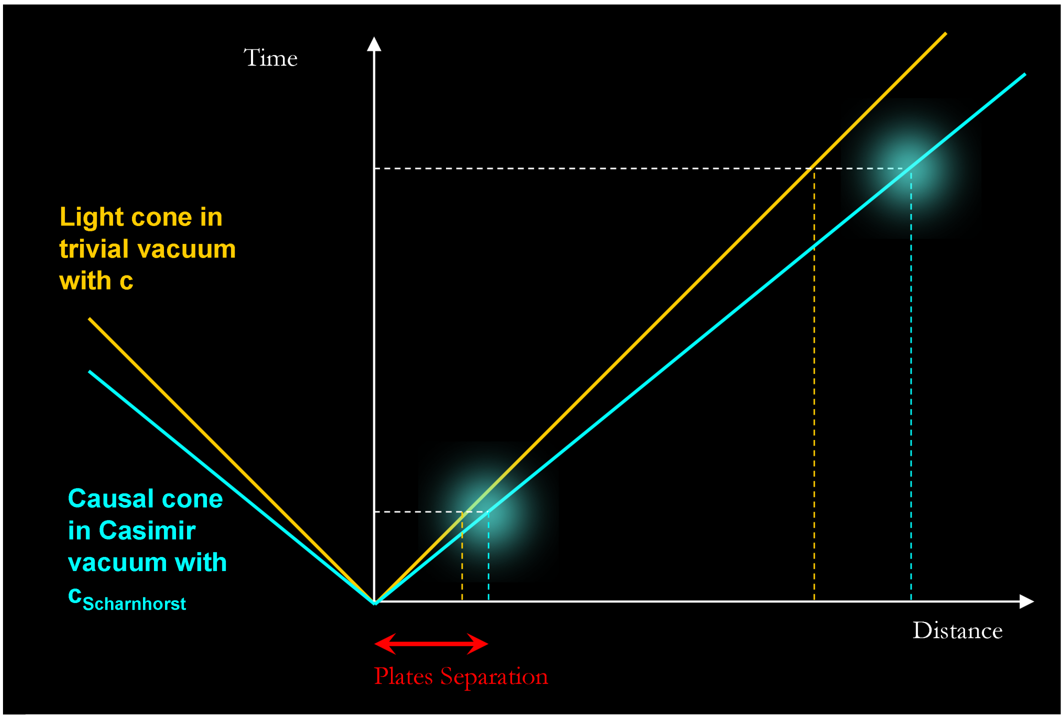


Figure 7.2: The "smearing" of the wavefront precludes measuring the signal speed as superluminal over the plates separation.

The kind of measurement uncertainty discussed by Milonni and Svozil on the one hand, and Ben-Menahem on the other differ profoundly: the former is related to the uncertainty relations, the latter reflects the impossibility to define the front of the signal to arbitrary accuracy, even classically. However these proofs share two essential features, which constitute necessary conditions for the conclusion to hold. Firstly, both restrict their analysis to photons that travel at most the distance between the plates. Again, this is required for the arguments to hold: if the motion of the photons could somehow be measured over an arbitrarily long travel distance, uncertainties would not be able to "blur" the velocity measurement to the point of precluding a superluminal signal velocity, as Fig. 7.2 illustrate. Restricting the propagation distance to be at most L is partly motivated by the fact that taking reflections into account would complicate the analysis considerably. This is also true for other options one may wish to consider, such as piercing holes through the plates, which may involve edge effects, or shining a photon right through the plates (since they are transparent for photons whose energy is larger than the plasma frequency) where one would have to take the dielectric property of the plates into account. However increased difficulty in the theoretical analysis does not imply that such measurements could not be performed.[12] On the other hand, it may be that all schemes one would use to increase the travel distance would slow down the light along parts of its trip in such a way that it would not be possible to measure the signal to propagate between the emission event and the measuring event at an average speed faster than c.[13]

Secondly, both arguments also hinge on the fact that Scharnhorst and Barton's result is only guaranteed to hold for photons of frequency $\frac{1}{L} << \omega << m_e$. Without this requirement, neither can conclude that the measurement uncertainties at stake are too large to measure the signal velocity to be superluminal. Notably, if ω is not bounded from above, then in the relation $\frac{\Delta v}{\Delta c} = \frac{\lambda}{\kappa\alpha^2}\frac{1}{\lambda_c}\left(\frac{L}{\lambda_c}\right)^3$ derived by Milonni and Svozil, λ can be taken to be arbitrarily small, implying that $\frac{\Delta v}{\Delta c}$ can be correspondingly minimized. Likewise in Ben-Menahem's work: without $\omega << m_e$, $\frac{1}{L} << \omega$ no longer implies that $\frac{1}{L} << m_e$, *i.e.* that $m_e L >> 1$ as required to show that the inequality $\zeta\, e^4 \frac{1}{(m_e L)^3} > 1$ cannot be satisfied. From a more intuitive standpoint, being able to include photon modes of arbitrarily large frequency would

---

[12] One would of course have to contend with the fact that the effect is far too tiny to be within experimental reach at present. Then again one could also imagine stacking many Casimir set-ups against one another to increase the effect. One could imagine interference experiments between a photon having travelled through such a set-up and one through trivial vacuum.

[13] The dielectric behavior of the plates would definitely cause such a slowing down; reflections certainly would as well as the photon would have to be absorbed and re-emitted by the atoms in the plates. Whether edge effects associated with holes pierced in the plates would cause such a slowing down seems less clear.

make it possible to form a wave packet with an arbitrarily sharp wavefront – *i.e.* to arbitrarily reduce its "smearing." This would make it possible to measure the signal speed as superluminal even for a travel distance no longer than L (and even shorter).

That the result derived is only valid within the regime $\omega << m_e$ also affects the work done a decade later by Stefano Liberati, Sebastiano Sonego and Matt Visser [13, 14, 15, 16]. Importantly, they showed that the motion of Scharnhorst's photons can be described in terms of an effective metric, which as discussed in chapter 2, entails that a superluminal signal speed would not violate causality in the sense of allowing for the Bilking argument paradox – often illustrated by the grand-father paradox. Liberati *et al* derived an expression for the velocity of the wavefront, hence what should be the signal velocity, and found it to be superluminal. Yet because the regime of validity of their calculations is restricted to the frequency range $\omega << m_e$ (which they point out), Ben-Menahem's remarks regarding the impossibility to define the wavefront accurately enough to measure its motion as superluminal still applies, and one could argue that the effective causal cone is not *wider* by an amount δx but *blurred* by that amount. Alternatively, one could argue that the effective cone is not truly the causal cone, in the sense of the cone corresponding to the behavior of measurable signals.

Indeed there are two ways to think of the arguments presented by Milonni, Svozil and Ben-Menahem. In Figs. (7.1) and (7.2), I have represented both the light cone and the causal cone as arbitrarily sharp. This is a natural way to think of the situation when one has relativity in mind. In this context photons track the light cone, but the latter has meaning independently of their motion: the light cone is associated with c understood as a fundamental, natural constant, whose value is well-defined independently of our ability to measure what photons do. One is then led to think of the light cone as arbitrarily "sharp", as illustrated above. Having this view in mind, it is natural to think of the causal cone in the same way, hence as perfectly well-defined. This is clearly the view one would be led to if one were to think of $c_{Scharnhorst}$ in the same way as c, *i.e.* as the value taken by c in the Casimir vacuum, $c_{Casimir\ vacuum}$ *vs* $c_{trivial\ vacuum}$. This would correspond to a single-metric as opposed to a bimetric description of the effect – which is not to say that this conception of the cones is *necessarily* incompatible with bimetricity.

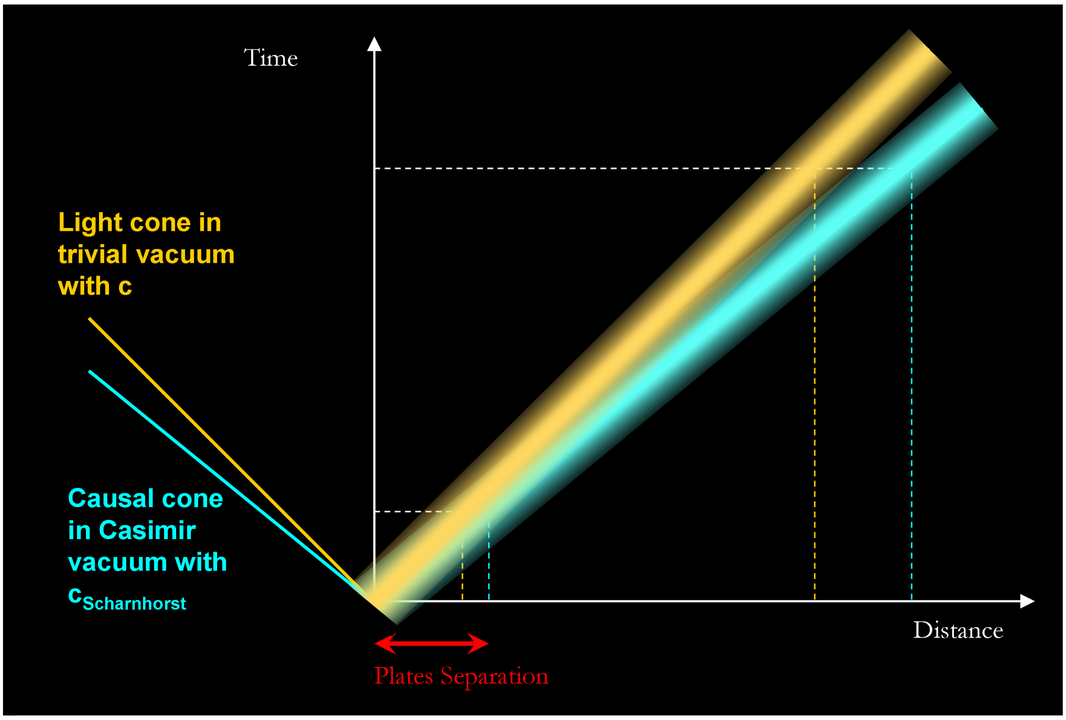


Figure 7.3: The cones themselves are blurred.

However, as illustrated above (Fig. 7.3), one can also have a different picture in mind, and consider that the cones themselves are blurred (or at least the effective one). If one thinks of the causal cone as *defined by* the motion of the wavefront, and the latter as smeared out, then one is led to view the cone itself as blurred. In this case the signal velocity is never a perfectly well-defined concept; close to the point of emission of the light, it is practically meaningless, and becomes better and better defined as one considers propagation over greater and greater distances (recall that the speed is represented by the opening angle of the cone).

These two different conceptions of the cones have implications regarding the type of questions one can ask. If the cone is arbitrarily well-defined, one can argue that the notion of signal velocity (or more precisely, of it having a definite value) still has meaning even between the plates. One would then hold that the impossibility of measuring it to be superluminal over L is simply a measurement problem. In contrast, if the cone itself is blurry, one cannot take this view. Conceivably, one could still want to resolve the difference between the light cone and the effective cone by taking many measurements.[14] However one cannot speak of the signal speed in the same way as in the former case, notably as having a specific value.

7.1.3 Scharnhorst and Barton's response

As discussed in chapter 3, Scharnhorst and Barton did address the concerns we just described in a joint paper published in 1993 [12].

They agreed that taken together, the requirement to limit the travel distance to be at most the plate separation L and the exclusion of photon frequencies higher than $\omega << m_e$ would prevent the measurement of a faster than c signal velocity.[15]

---

[14] Recall that in their joint paper, Scharnhorst and Barton remarked:

> As so often when applying the indeterminacy relations, one could argue that the statistically analysed average of many measurements does make it possible, in principle, to determine shifts well below the mean-square deviations, and thus to verify effects that more cursory considerations of single measurements sometimes describe as undetectable. [12], p.2040.

[15] Recall that they stated:

Limit of the travel distance

One of their comments suggests that they did not view this as having fundamental implications for the value of the signal velocity itself:

> Such considerations are not specific to the effect we are studying. What they show, equally for quantum and for classical waves, is that in confined geometries a single measurement on a single traverse can determine the speed of light having limited frequencies only with limited accuracy: under such conditions the operational significance of any ultimate speed is apt to remain somewhat nebulous.

It seems fair to say that for their critics, this very nebulosity is precisely the whole point: from an operationalist standpoint, being unable to measure that the signal velocity is superluminal means that it is not. The remark just quoted might suggest that Scharnhorst and Barton are taking a realist stance on the issue, and hold that the concept of signal velocity in fact has meaning beyond the possibility of measuring it.

However this is not their point, for they continue:

> (As so often when applying the indeterminacy relations, one could argue that the statistically analysed average of many measurements does make it possible, in principle, to determine shifts well below the mean-square deviations, and thus to verify effects that more cursory considerations of

---

On a single traverse the low-frequency prediction (2) [$n = 1 + \delta n$, $\delta n = -\frac{11\pi^2}{(2^2)(3^4)(5^2)}\frac{e^4}{(mL)^4}$] cannot be verified even in principle, because any wave-group narrow enough to afford the requisite accuracy must include significant high-frequency components to which the effective coupling (1) [$\Delta\mathcal{L} = \frac{1}{(2^3)(3^2)(5\pi^2)}\frac{e^4}{m^4}\left(\left(E^2 - B^2\right)^2 + 7\left(E.B\right)^2\right)$] and therefore (2) no longer apply" [12], p.2040. My italics.

> single measurements sometimes describe as undetectable.)[16]

This is certainly a very interesting point that would require further attention.

As a matter of principle, it cannot be excluded that thought experiments based on a different (presumably more complex) set-up may allow for high enough accuracy for a shift in the signal velocity to be measured. This of course is a critique which can be levelled at any thought experiment meant to demonstrate an impossibility, and is by no means specific to those we just described. In the present case, because the results hinge on the measurement being done over a maximum distance of L, alternative thought experiments would presumably seek to relax this requirement. Recall that it was made because allowing the photon to reflect off the plates would greatly complicate the analysis.[17]

However Scharnhorst and Barton chose to focus instead on another response to the two proofs described above, aimed specifically at the implications associated with the $\omega << m_e$ regime of validity of their result.

Photon frequency regime bounded from above

Scharnhorst and Barton could not draw a definite conclusion regarding the faster-than-c character of the signal velocity. In agreement with the point that defining a signal velocity requires arbitrarily high frequencies, they considered the following

---

[16] [12], p.2040.

[17] So would piercing the plates and allowing the photon to propagate through the gap, as one would then have to contend with edge effects that may affect the Casimir vacuum structure, and diffraction effects on the photon's part. Now a high energy photon (such as is of interest) would propagate right through the plates anyway since the latter have a finite plasma frequency. It would be considerably slowed down by the experience, so that it could never be measured to connect two points outside the plates faster than c. However it would perhaps be interesting to analyse a thought experiment in which such a photon is made to interfere with another which would not have gone through the plates – and of course one could "stack" many Casimir set-ups behind one another for the first photon to go through.

definition:

$$v_{signal} = \frac{1}{Re\, n\,(\omega \to \infty)}.^{18} \tag{7.1}$$

Recall that the phase velocity is given by $v_\Phi = \frac{c}{n}$, so Eq. (7.1) is the formal expression of the idea that the signal velocity is the phase velocity in the high frequency, $\omega \to \infty$ limit.

The definition Eq. (7.1) entails that $v_{signal} > c$ if and only iff Re n $(\omega \to \infty) < 1$; but obviously the quantity Re n $(\omega \to \infty) < 1$ cannot be directly determined by an analysis restricted to a regime where ω is bounded from above.

In order to bypass this difficulty, Scharnhorst and Barton needed to relate the low frequency regime $\omega << m_e$ in which their phase velocity result was derived, to the high frequency regime $\omega \to \infty$ in which the phase velocity eventually coincides with the signal velocity, as given by Eq. (7.1). More specifically, in order to use Eq. (7.1) they needed to relate the index of refraction in the two regimes – which they designated respectively by $n(0)$ and $n(\infty)$.

In order to do so, they used what is known as the Kramers-Kronig relation. Strictly speaking, the latter is a *mathematical* relation between the real and imaginary parts of a given quantity.[19] Applied to the index of refraction, it directly leads to the result:

$$n\,(\infty) = n\,(0) \;-\; \frac{2}{\pi}\int_0^\infty d\omega'\,\frac{Im\, n(\omega')}{\omega'}\;.^{20} \tag{7.3}$$

---

[18] where they took $c = 1$.

[19] Taking this quantity to be $n(\omega)$ it takes the form:

$$Re\, n\,(\omega) \;=\; Re\, n\,(\infty) \;+\; \frac{2}{\pi}\int_0^\infty d\omega'\,\frac{\omega' Im\, n(\omega')}{\omega'^2 - \omega^2}. \tag{7.2}$$

[20] This is obtained by taking $n(\omega) = n(0)$ as a specific case of the general expression given in the previous footnote, and rearranging.

This, Scharnhorst and Barton argued, leaves only "two equally unorthodox possibilities":

> Either $n(\infty) < 1$, so that the true signal velocity, too, exceeds c; or the conventional no-photon vacuum between the (fixed!) mirrors amplifies a probe beam. In the second case the vacuum would fail to act as a passive medium.[21]

Indeed if Im $n(\omega) > 0$, the integral in Eq. (7.3) is positive as well, and therefore $n(\infty) < n(0) < 1$.[22] This means that the signal velocity itself is larger than c.[23] The only way to escape this conclusion is to suppose Im $n < 0$, which corresponds to the Casimir vacuum failing to behave passively for at least some frequencies. What this concretely means is that the Casimir vacuum would induce the generation of real photons, which Scharnhorst and Barton noted seems to require energy creation out of nothing.

Now for the "Drummond-Hathrell" effect, which can be deemed a gravitational analog to the Scharnhorst effect in so far that it predicts photon phase and group velocities would be larger than c in some gravitational backgrounds, it has been shown that Im n can indeed be negative so that signal velocities remain smaller than c. However for this to happen, photons need to be focused: in that way the situation allows for a local increase in amplitude, while avoiding the creation of photons overall. This requires that the "medium" (*i.e.* in that case the gravitational background) be inhomogeneous in the appropriate way. Scharnhorst and Barton pointed out that something analogous to this cannot apply to the Scharnhorst effect, because $n(\omega)$ is not position-dependent, *i.e.* the medium (Casimir vacuum) is

---

[21] [12], p.2042.

[22] Recall that $n(0) < 1$ was the main result of Scharnhorst and Barton's 1990 derivations.

[23] Since $v_{signal} = v_{\Phi}(\infty) = \frac{c}{n(\infty)}$, $n(\infty) < 1$ entails $v_{signal} > c$.

homogeneous.[24]

Scharnhorst and Barton stressed that the superluminal option "does not in any way contradict or pose any conceptual threats to special relativity."[25] However these remarks were not meant to imply they favored this option – simply that this possibility should not to be outright discounted. Scharnhorst and Barton did not believe that their argument made it possible to draw a definite conclusion on the matter:

> We believe that in strict logic the choice between these two alternatives remains open; indeed, subjectively the two present writers probably incline to opposite choices.[26]

This statement would still hold today: subsequent derivations are in the $\omega << m_e$ regime as well so that Milonni, Svozil and Ben-Menahem's criticisms are still pertinent, and on the other hand it has not be shown that the Casimir vacuum fails to behave passively.

7.2 Superluminal behavior and causality

The prospect of a faster-than-c signal velocity evidently raises concerns about a possible conflict with Lorentz invariance – hence special relativity – and causality in the sense of Bilking argument type paradoxes. However as was discussed in chapter 2, in the context of the Scharnhorst effect such fears are likely unwarranted.

7.2.1 "Soft-breaking" of Lorentz invariance

With respect to Lorentz invariance specifically, it is important to differentiate what we can rightfully expect to be invariant or not, and what turns out to be (or fails

---

[24] [72], [12], p.2044.

[25] [12], p.2038.

[26] [12] p.2039.

to be) in the situation of interest.

What is required by relativity is for *theories* to be Lorentz invariant. Concretely, this means that the Lagrangian and the equations of motion derived from it must be Lorentz invariant. Both the standard Euler-Heisenberg Lagrangian and its SCET counterpart duly are (as are also the other theories that might involve superluminal propagation discussed in chapter 2).

The solutions of these equations, however, are a different matter. As Jeremy Butterfield puts it:

> Suppose a theory obeys a symmetry in the sense that a certain transformation, e.g. a spatial rotation or a boost, maps any dynamical solution to another solution. This by no means implies that every solution should be invariant, i.e. mapped onto itself, under the transformation: after all, not every solution of Newtonian mechanics is spherically symmetric![27]

Similarly, in the case of the Scharnhorst effect, the vacuum state – that is, a specific solution to the Lorentz invariant equations of motion – does not obey the symmetry. However this, of course, threatens in no way the Lorentz invariance of the theory itself, as Liberati *et al* emphasized:

> Light behaves in a non-Lorentz invariant way only because the ground state of the electromagnetic field is not Lorentz invariant. The Euler-Heisenberg Lagrangian, from which the existence of the effect can be deduced, as well as all the machinery of QED employed in its derivation, is still fully Lorentz invariant. For this reason, one often speaks of a *soft breaking* of Lorentz invariance in order to distinguish from a situation in which also the basic equations, and not just the ground state, are no

---

[27] [22], [23] p.33.

> longer Lorentz invariant.
>
> Of course, this soft breaking of Lorentz invariance has no fundamental influence on special relativity no more than being inside a material medium has.[28]

The reason for this "soft-breaking" of Lorentz invariance is clearly to be attributed to the presence of the Casimir plates, as Scharnhorst and Barton already stressed in their joint paper:

> The presence of the mirrors breaks Lorentz invariance along the mirror normal (the mirrors define a preferred inertial frame), which obviates the arguments used in special relativity to prove that no signals can travel faster than light does in unbounded (Lorentz-invariant) space. By contrast, Lorentz invariance is unbroken parallel to the mirrors, and the light speed in these directions naturally must, and does, remain unchanged.[29]

How the plates achieve this is somewhat less obvious. Somehow, they define a preferred reference frame – their rest frame.[30] A natural way to think of this for anyone accustomed to macroscopic physics is to view the Casimir vacuum as a medium. Jeremy Butterfield uses the concept as a useful analogy:

> Agreed, the vacuum state for empty Minkowski spacetime is required to be Lorentz-invariant since it should look the same in a translated, rotated or boosted frame. But the presence of the plates breaks this symmetry,

---

[28] [15] p.168, [16] pp.2-3.

[29] [12] p.2038.

[30] That the description involves a preferred reference frame is also true of the other theories discussed in chapter 2.

> just as a pervasive inertially-moving medium would do: licensing a non-Lorentz-invariant vacuum state.[31]

But this idea has also appeared as more than an analogy in some of the research on the Scharnhorst effect.

7.2.2 The Casimir vacuum as a dielectric medium

The work due to Latorre *et al* in 1995 has explicitly portrayed the Casimir vacuum as a dielectric medium, as it compared the phase velocity of light propagating through different backgrounds, the Casimir vacuum being but one of these. Their study consisted in relating the phase velocities of the propagating photons to the energy densities characteristic of these backgrounds [20, 21], so it is clear that in this framework, the Casimir vacuum appears as a medium. The superluminal character of the phase velocity is shown to arise from the energy density of the Casimir vacuum being lower than in trivial vacuum.

More generally, thinking of the Casimir vacuum as a medium is tempting in view of the history of the Euler-Heisenberg Lagrangian: the photon-photon scattering process it was developed to account for inspired physicists to view the background electromagnetic field involved as a dielectric medium. This is likely why both Scharnhorst and Barton found it natural to use the concept of index of refraction to derive the photon speed. However there is a difference between the Scharnhorst effect and the previous studies of photon-photon scattering: one is now dealing with a vacuum instead of a real (as opposed to virtual) electromagnetic field. One may think of the Casimir vacuum as a modified zero-point electromagnetic field – indeed, this is the idea used by Barton in his derivation. Furthermore, Barton clearly distinguishes between the propagating photon (which he models as

---

[31] [22], [23] p.33.

a classical field), and the background (which he represents by quantum VEVs).

Both Scharnhorst's and the SCET derivation presented in chapter 4 also lend themselves to viewing the Casimir vacuum as a medium – and indeed, I have described the SCET model in these terms, referring to the Casimir vacuum as a separate (*i.e.* soft) field that forms a background through which the collinear Scharnhorst photon propagates.

However, the presence of the plates comes into the problem by affecting the propagators of the off-shell (*i.e.* "virtual") photons (again, in both Scharnhorst's work and the SCET calculation); yet if one is to give an interpretation to the Feynman diagrams responsible for the Scharnhorst effect, the idea that these off-shell photons constitute an external medium seems less natural.[32] Essentially, one finds oneself confronted to the type of issues discussed in chapter 5, notably those pointed out by Jaynes who argued that the modes responsible for the Lamb shift (attributed to "vacuum fluctuations") did not exist in the absence of the radiating atom, but were generated by it. In the present case, this concern becomes whether the photons that "form" the Casimir vacuum exist in the absence of the probe photon.

Recall from the discussion in chapter 3 that when we consider the full-theory (as opposed to the effective one) the Scharnhorst effect arises from the following two-loop processes (where the external photon propagators are not represented, but would connect at x and y):

---

[32] Admittedly my description of the model used in chapter 4 seems to take this idea for granted. This was partly for convenience, but also because the model refers to the effective theory, and in the latter this description is *arguably* more natural, as we shall soon see.

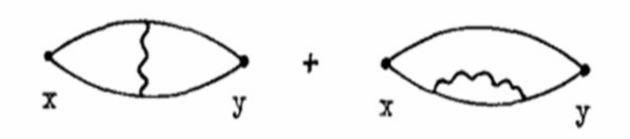


Figure 7.4: The two-loop processes responsible for the Scharnhorst effect.[33]

When one wishes to give a conceptual interpretation of these diagrams as processes, the usual description is that they represent the exchange of an off-shell photon either between two fermions (diagram on the left) or its emission and reabsorption by the same fermion (on the right). In both cases, the off-shell photon is thought to be emitted by one of the fermions, not to pre-exist in a vacuum conceived as an external bath of virtual particles – which could easily be thought of as a medium.

Now one could conceivably think of these off-shell photons in the latter manner, and say that the diagrams represent the interaction of photons already "present" in the vacuum that now interact with the fermion(s). However such a description would have to account for the fact that in this case, momentum conservation at the vertices would not take into account the momentum of incoming or outgoing photons. In any event, thinking of the off-shell photons as forming an independently-existing bath of virtual particles is certainly not the only option.

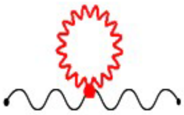


Figure 7.5: Tadpole diagram responsible for the Scharnhorst effect in SCET.

Now both Scharnhorst's calculation and the SCET one use the effective (Euler-Heisenberg) theory, not QED *per se.* In the effective theory, the diagram responsible for the effect is a tadpole, so what is at stake in this case is the status of the photon loop. Because the interaction takes place at a single event, viewing the off-shell photon as emitted and reabsorbed is perhaps less natural than for the fermions of the diagrams above. Interacting at a single event with a virtual photon from a separate background might appear somewhat more natural in this case.

Therefore, when it comes to the external medium-or-not dichotomy, intuition may seem to pull in different directions whether we consider the full theory or its effective counterpart. One could argue that on the basis of its status as full theory, the description afforded by QED should be taken more seriously – although rejecting off-hand ontological claims of effective descriptions should not be done lightly. In any case we see that describing the Casimir vacuum as a medium is not an obvious choice devoid of difficulties when we take a closer look at the interactions involved. However, it is certainly worth recalling that one *can* give a description of the Scharnhorst effect based on the concept of refractive index, understood as a property of a medium – and one then obtains the same result as with later calculations that avoid this concept. So it seems fair to say that the space between Casimir plates *behaves* as a dielectric – at least, it is a simpler description than saying the propagating photons behave as if they propagated through a dielectric.

Another aspect of the issue is that if the Casimir vacuum can be said to be an external medium, this certainly implies that the trivial vacuum is not "empty" either – as popular works are so fond of stressing. Latorre *et al*'s work certainly suggests this view: if photons propagate faster in the Casimir vacuum than in trivial vacuum because the energy density is lower there, it stands to reason that there is more of *something* in the latter than in the former.

7.2.3 Concerns regarding causal paradoxes

However, whether or not the Casimir vacuum can rightfully be viewed as an external medium, the presence of the Casimir plates alone arguably provides a preferred reference frame, which from a formal standpoint at least accounts for the non-Lorentz invariant character of the vacuum solution.

Yet not everyone has agreed that this soft-Lorentz breaking is unproblematic, so long as the invariance of the theory itself is preserved. On the contrary, Adams *et al* have cautioned that such "apparently perfectly sensible low-energy effective field theories governed by local, Lorentz-invariant Lagrangians" are in fact not acceptable, precisely because they are secretly non-local [and] do not admit any Lorentz-invariant notion of causality."[34] Although what precisely they mean by this latter statement is not entirely clear, what follows throughout their paper shows their main concern to be the very possibility of superluminal propagation in the theories in question because, they argue, this would directly lead to the possibility of Closed Timelike Curves. And this, indeed, in the form of the so-called "tachyonic anti-telephone" thought experiment, has long been the main worry associated with the possibility of faster-than-c signalling. This thought experiment indeed shows how in the absence of a preferred reference frame, faster-than-c propagation can result in a Bilking argument causal paradox – popularly exemplified by the grand-father paradox. However, as discussed in chapter 2, no such paradox can occur if the propagation is constrained to take place on a "causal cone", associated with an effective metric. And indeed, Liberati *et al* have derived such an effective metric for the Scharnhorst effect, thereby addressing this concern.[35]

---

[34] [37] p.2, [38] p.1.

[35] Recall that, as explained above, the fact that the SCET derivation presented in chapter 4 recovers the same phase velocity for the Scharnhorst photons as was hitherto obtained with on the basis of the standard Euler-Heisenberg Lagrangian implies that the causal cone described by Liberati *et al* is indeed sharply wider than the light-cone, not simply blurred by measurement

Furthermore – and, again, as Liberati *et al* pointed out – the existence of an effective metric makes it possible to use the concept of stable causality, introduced in general relativity, and the theorem that applies to it:

> A stably causal spacetime possesses no closed timelike curves and no closed null curves.[36]

Recall from chapter 2 that: "A spacetime is said to be stably causal if and only if it possesses a Lorentzian metric $g_{\mu\nu}$ and a globally defined scalar function t such that $\nabla_\mu t$ is everywhere nonzero and timelike with respect to $g_{\mu\nu}$."[37]

This theorem has been applied beyond the realm of general relativity, by interpreting $g_{\mu\nu}$ as the effective metric that determines the causal cones.[38] In the context of the Scharnhorst effect, one can easily find a global time function satisfying the above requirement on $\nabla_\mu t$: the coordinate time in the rest frame of the Casimir plates.[39] Having shown in this way that the spacetime is stably causal, the theorem then guarantees that no closed timelike curves nor closed null curves can form in the Casimir vacuum despite the superluminal behavior of photons propagating through it.

This holds at any rate for a single pair of Casimir plates. Adams *et al* imagined configurations analogous to taking two set-ups and moving them with respect to one another in such a way that a CTC could form.[40] In keeping with the analogy with general relativity, Liberati *et al* have argued essentially that something akin to the

---

uncertainties.

[36] [68], p.199. [68], p.198, discussed by [15] p.180.

[37] [68], p.198, discussed by [15] p.180.

[38] Notably in the context of k-essence, both in a Minkowski spacetime and a Friedmann universe.[58] pp.10-12, [59] pp.9-11.

[39] Also recall that for the Scharnhorst effect, the relevant $g_{\mu\nu}$ is formed of the effective metric between the Casimir plates, and the Minkowski metric on either side of them.

[40] They actually considered "bubbles of non trivial vacua" moving past one another. See chapter 2 for details.

Chronology Protection Conjecture. would prevent CTCs from forming.

Robert Geroch has drawn essentially the same conclusion. He did not specifically consider the Scharnhorst effect however: his work is meant to be quite general, and can be viewed as justifying the Chronology Protection Conjecture.

Geroch imagined a thought experiment where two pipes are filled with some hypothetical fluid, the sound speed of which is superluminal. The issue is then whether this superluminal character can be thought to imply that a CTC could be formed – again, by moving the pipes with respect to each other while sending signals through them. Using his formal work on the structure of differential equations, Geroch has argued that no such causal paradox would occur: as discussed in chapter 2, provided that the system of equations describing the system possesses a hyperbolization (which of course would have to be verified in the case of the Scharnhorst effect for his reasoning to apply there), whatever would follow from well-posed initial conditions could not be any CTC, simply because "this arrangement [i.e. the pipes moving in the manner required for the formation of a CTC] cannot be in the domain of dependence of any surface, *i.e.*, it cannot be predicted, via the initial-value formulation, from any initial data."[41] More concretely, given well-posed initial conditions, "what happens from those initial conditions must be determined by evolving, from those initial data, the differential equations for the system.[...] Whatever results from these data and these equations is what results. But we know that whatever it turns out to be the result of this evolution will not consist of the two pipes moving in the prescribed manner."[42] In this case, the initial conditions in question could be the two pipes lying at rest with respect to one another while superluminal signals propagate through them. The analog situation for the Scharnhorst effect would consist in Scharnhorst photons propagating superluminally in Casimir set-ups at

---

[41] [62], p.68, [63], p.11.

[42] [62], p.68, [63], p.12.

rest with respect to one another. Provided that Geroch's formalism does apply to the Scharnhorst effect, there should be no cause for worry that this might result in causal paradoxes.

7.3 Implications of the SCET approach

In chapter 4, we used the SCET framework to derive the phase velocity of a photon propagating in a Casimir vacuum normal to the plates. We found the same result as Scharnhorst, Barton, and later researchers had, *i.e.* the phase velocity is larger than its trivial vacuum value of c:

$$v_\perp = c\left(1 + \frac{11}{(2^6)(45)^2}\frac{e^4}{m_e^4\, d^4}\right). \tag{7.4}$$

Previous derivations were based on the standard Euler-Heisenberg theory.

7.3.1 Control of theoretical errors: implication

As explained in chapter 4, the advantage of the SCET approach consists in the fact that theory errors are well-controlled, so that higher order terms are power-suppressed compared to lower order ones.

Both the standard and the SCET frameworks involve a power expansion in $\frac{E}{m_e}$ as well as a perturbative expansion in α. Also in both cases (and as is usual in effective theory), each term can be expressed as a product of dimensionless expressions. These expressions consist in the expansion parameters (to the power characteristic of the order of the term), and a factor containing the coefficients and the operators. By design, the expansion parameters are numbers smaller than 1, so that each of them tends to power-suppress the terms of higher order in it.

In the standard theory however, that the expansion parameters are much smaller than 1 is not sufficient to ensure that, as we go to higher orders in the series expansions, higher order terms represent smaller and smaller corrections. This is because in any of these terms, the size of the factor multiplying the parameters is not known: it contains expressions of the form $\frac{\overline{F}^{\mu\nu}}{E^2}$, that is fields that have been re-scaled in energy, but because in the standard theory the fields can have any energy, the size of these factors (or more precisely of their expectation value) is not well-defined within the theory. In other words, in the standard theory the field can be re-scaled to be dimensionless, but one cannot go any further and re-scale it to be of order 1 as well. In contrast, SCET differentiates fields of different energy scales. It introduces a dimensionless parameter λ which involves the ratio of the energy-momentum scale of the soft field ϵ to that of the collinear, Scharnhorst photon E ($i.e.$ $\lambda \equiv \sqrt{\frac{\epsilon}{E}}$). As explained in chapter 4, this makes it possible to re-scale the fields into new fields that are not only dimensionless, but also of order 1. The price for this re-scaling process is that it generates in the terms an additional expansion parameter: a factor of λ to a positive power appears. Since λ, like the other parameters in the theory (α and $\frac{E}{m_e}$), is much smaller than 1, this additional factor contributes to the power suppression of the higher order terms.

The key point is that, through this re-scaling, SCET makes it possible to write each term as a product of the expansion parameters multiplied by a numerical expression of order 1 (involving the re-scaled fields). Since these numerical expressions are now of order 1, it is unlikely that they would be able to spoil the power suppression due to the parameters, and higher order terms in the various expansions are most likely smaller than lower order ones – although as a matter of principle one would need to calculate higher order terms in order to be sure of this.

Subject to this qualification, the fact that the phase velocity calculated to order $\frac{\alpha^2}{m_e^4}$ within the SCET framework is equal to $c + \Delta v$ as found by Scharnhorst, Barton and other researchers implies the following: although higher order corrections change the value of Δv somewhat, they are unlikely to do so to the point of totally cancelling out the superluminal character of the phase velocity. To use the concept of causal cone discussed above, higher order corrections to the opening of the causal cone would not be large enough for it to coincide with the light cone.

7.3.2 Implications of the SCET approach for the signal velocity

As we just saw, control of theoretical errors ensures that the opening of the causal cone is not significantly different from what is derived at the order the Scharnhorst effect is calculated, and is in fact wider than the light cone (see Fig. 7.3). However, as discussed above in section 7.1.2, this does not necessarily tell us about the signal velocity *per se*. Since this is arguably what has been the crucial issue in the literature, one is led to ask: does SCET give us any reason to think that the phase velocity we have derived may also constitutes the signal velocity? In fact, it would seem that it does. More specifically, SCET provides an argument to the effect that the blurring of the wavefront as illustrated in Fig. 7.2 (or of the causal cone itself, see Fig. 7.3), can be arbitrarily reduced.

The reason for this is that in the SCET approach, at the order the effect is calculated the phase velocity of the Scharnhorst photon depends only on the energy of the background, not on the energy of the propagating photon itself. This is due to the manner in which background and probe are modelled to interact. Recall that the defining feature of SCET is its separation of energy scales. Using light-cone coordinates, this leads to a description where the propagating photon has typical energy E associated with its p^+ momentum component, while the Casimir background is made of soft photons whose momentum k^μ has the same typical size as the p^- component of the Scharnhorst photon – *i.e.* ϵ, or $O(\lambda^2)$. As explained in section 4.3, this means that the Casimir vacuum only interacts with the probe photon via its p^- momentum component; the background (*i.e.* k^μ) is decoupled from both $p_\perp$ and p^+. Furthermore, SCET shows that at the order it is calculated, the interaction term responsible for the effect depends only on the energy of the background, ϵ, not on the energy E of the probe photon. Interactions between the background and the other momentum components of the Scharnhorst photons do occur at higher order in λ; but they are power suppressed by λ, so they should only contribute small corrections to the effect, modifying the lowest order speed only slightly – although as a matter of principle one would need to calculate all higher order terms in order to be sure.

This means that the Scharnhorst effect depends on properties of the background – the quantity ξ in the vacuum polarization tensor, or the index of refraction if we want to use semi-classical concepts – which are ultimately fixed by ϵ. Via the p^- momentum component of the probe photon, these properties affect the behavior of this photon, determining its phase velocity; however these background properties are not in turn affected by the energy E of the propagating photon, because E is associated with p^+ which is not coupled to k^μ.

This entails that a photon of arbitrarily large E experiences the same background as one with $\omega << m_e$. Hence, a photon of arbitrarily large frequency would have the same phase velocity as the photons described by the Euler-Heisenberg theory, *i.e.* the phase velocity we derived in chapter 4.

Regarding the arguments that uncertainties would preclude measuring the signal speed to be superluminal, recall that a well-defined signal velocity requires forming an arbitrarily sharp wavefront, which in turns involves using photons of arbitrarily high frequency (arbitrarily short wavelength); hence being limited to photons with $\omega << m_e$ leads to a smeared, blurred wavefront. Again, as always in effective theory, only if all higher order terms could be calculated could one make definitive statements. With this warning in mind however, the way SCET models the interaction between the Casimir vacuum and the propagating photon entails that photons of arbitrarily high frequency would have the same, greater-than-c, phase velocity as their lower frequency counterparts, and therefore could be used to form wave packets with a sharp wavefront travelling at this speed. Phrased another way, in the debates that have taken place in the physics literature the signal velocity has been defined as the infinite-frequency limit of the phase velocity. Therefore, that the phase velocity of a photon of arbitrarily large frequency has the value first derived by Scharnhorst suggests that the signal velocity has this same value: *i.e.* that the signal velocity of Scharnhorst photons is larger than c.

Bibliography

[1] G. Bressi, G. Carugno, R. Onofrio, and G. Ruoso, Physical review letters **88**, 041804 (2002).

[2] G. Bressi, G. Carugno, R. Onofrio, and G. Ruoso, arXiv preprint quant-ph/0203002 (2002).

[3] V. Bezerra, G. Klimchitskaya, and C. Romero, Modern Physics Letters A **12**, 2613 (1997).

[4] A. Canaguier-Durand, P. A. M. Neto, I. Cavero-Pelaez, A. Lambrecht, and S. Reynaud, Physical review letters **102**, 230404 (2009).

[5] K. Scharnhorst, Physics Letters B **236**, 354 (1990).

[6] G. Barton, Physics Letters B **237**, 559 (1990).

[7] R. Tarrach, Physics Letters B **133**, 259 (1983).

[8] M. Bordag, D. Robaschik, and E. Wieczorek, Annals of Physics **165**, 192 (1985).

[9] J. D. Jackson, *Classical Electrodynamics* (Wiley, 1998), 3rd ed., ISBN 047130932X.

[10] P. W. Milonni and K. Svozil, Physics Letters B **248**, 437 (1990).

[11] S. Ben-Menahem, Physics Letters B **250**, 133 (1990).

[12] G. Barton and K. Scharnhorst, Journal of physics. A, mathematical and general **26**, 2037 (1993).

[13] S. Liberati, S. Sonego, and M. Visser, arXiv preprint quant-ph/0010055 (2000).

[14] S. Liberati, M. Visser, and S. Sonego, Phys. Rev. D **63**, 085003 (2001).

[15] S. Liberati, S. Sonego, and M. Visser, Annals of Physics **298**, 167 (2002).

[16] S. Liberati, S. Sonego, and M. Visser, arXiv preprint gr-qc/0107091 (2001).

[17] P. Milonni and M.-L. Shih, Physical Review A **45**, 4241 (1992).

[18] P. Milonni, R. Cook, and M. Goggin, Physical Review A **38**, 1621 (1988).

[19] J. G. Cramer (1990), URL `http://www.npl.washington.edu/av/altvw43`.

[20] J. I. Latorre, P. Pascual, and R. Tarrach, Nuclear Physics B **437**, 60 (1995).

[21] J. Latorre, P. Pascual, and R. Tarrach, arXiv preprint hep-th/9408016 (1994).

[22] J. Butterfield, International Studies in the Philosophy of Science **21** (2007).

[23] J. Butterfield, Arxiv preprint arXiv:0708.2189 (2007).

[24] S. Weinstein, Synthese **148**, 381 (2006).

[25] J. O. Weatherall, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics **48**, 109 (2014).

[26] I. T. Drummond and S. J. Hathrell, Phys. Rev. D **22**, 343 (1980), URL `http://link.aps.org/doi/10.1103/PhysRevD.22.343`.

[27] G. Shore, Nuclear Physics B **460**, 379 (1996), ISSN 0550-3213, URL `http://www.sciencedirect.com/science/article/pii/055032139500646X`.

[28] G. Shore, arXiv preprint gr-qc/9504041 (1995).

[29] R. Daniels and G. Shore, Nuclear Physics B **425**, 634 (1994).

[30] R. Daniels and G. Shore, arXiv preprint hep-th/9310114 (1993).

[31] R. Daniels and G. Shore, Physics Letters B **367**, 75 (1996).

[32] R. Daniels and G. Shore, arXiv preprint gr-qc/9508048 (1995).

[33] G. Shore, in *Time and Matter* (2006), vol. 1, pp. 45–66.

[34] G. Shore, pp. 45–66 (2003), `gr-qc/0302116`.

[35] G. Shore, Nuclear Physics B **778**, 219 (2007).

[36] G. Shore, arXiv preprint hep-th/0701185 (2007).

[37] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, Journal of High Energy Physics **2006**, 014 (2006).

[38] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, arXiv preprint hep-th/0602178 (2006).

[39] C. Skordis, Classical and Quantum Gravity **26**, 143001 (2009).

[40] C. Skordis, arXiv preprint arXiv:0903.3602 (2009).

[41] J. Bekenstein and M. Milgrom, The Astrophysical Journal **286**, 7 (1984).

[42] J. Bekenstein, Physics Letters B **202**, 497 (1988).

[43] J. Bekenstein, J. Bond, B. Carter, T. Damour, S. Deser, and M. Duncan, General Relativity and Gravitation **19**, 68 (1987).

[44] J. Bekenstein, in *Developments in General Relativity, Astrophysics and Quantum Theory: A Jubilee in Honour of Nathan Rosen*, edited by J. R. FI Cooperstock, LP Horwitz (Bristol: IOP Publishing, 1990), p. 156.

[45] J. Bekenstein, p. 905 (1992).

[46] J. Bekenstein, Physical Review D **70**, 083509 (2004).

[47] J. Bekenstein, arXiv preprint astro-ph/0403694 (2004).

[48] R. Sanders, The Astrophysical Journal **480**, 492 (1997).

[49] R. Sanders, arXiv preprint astro-ph/9612099 (1996).

[50] J. Bekenstein, Physical Review D **48**, 3641 (1993).

[51] J. Bekenstein, arXiv preprint gr-qc/9211017 (1992).

[52] J. Bekenstein and R. Sanders, The Astrophysical Journal **429**, 480 (1994).

[53] J. Bekenstein and R. Sanders, arXiv preprint astro-ph/9311062 (1993).

[54] J. Bruneton, Physical Review D **75**, 085013 (2007).

[55] J. Bruneton, arXiv preprint gr-qc/0607055 (2006).

[56] M. Clayton and J. Moffat, Physics Letters B **460**, 263 (1999).

[57] M. Clayton and J. Moffat, arXiv preprint gr-qc/9910112 (1999).

[58] E. Babichev, V. Mukhanov, and A. Vikman, Journal of High Energy Physics **2008**, 101 (2008).

[59] E. Babichev, V. Mukhanov, and A. Vikman, arXiv preprint arXiv:0708.0561 (2007).

[60] C. Bonvin, C. Caprini, and R. Durrer, arXiv preprint arXiv:0706.1538 (2007).

[61] J. Faye, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta (2015), winter 2015 ed.

[62] R. Geroch, *Advances in Lorentzian Geometry: Proceedings of the Lorentzian Geometry Conference in Berlin* (Amer Mathematical Society, 2011), chap. Faster Than Light?, Ams/Ip Studies in Advanced Mathematics Series, ISBN 9780821853528, URL `http://books.google.ca/books?id=uyRb3c8PcykC`.

[63] R. Geroch, arXiv preprint arXiv:1005.1614 (2010).

[64] J. Earman, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics **48**, 102 (2014).

[65] J. S. Bell, *Speakable and unspeakable in quantum mechanics: Collected papers on quantum philosophy* (Cambridge university press, 2004).

[66] H. R. Brown, *Physical relativity: Space-time structure from a dynamical perspective* (Oxford University Press on Demand, 2005).

[67] N. Huggett and C. Hoefer, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta (2015), spring 2015 ed.

[68] S. W. Hawking and G. F. R. Ellis, *The large scale structure of space-time*, vol. 1 (Cambridge university press, 1973).

[69] R. Geroch, *General Relativity: Proceedings of the Forty Sixth Scottish Universities Summer School in Physics, Aberdeen, July 1995* (Taylor & Francis Group, 1996), chap. Partial differential equations of physics, SUSSP Proceedings, ISBN 9780750304191, URL `http://books.google.com/books?id=qHrqYH6kprsC`.

[70] R. Geroch, arXiv preprint gr-qc/9602055 (1996).

[71] A. Dolgov and I. Novikov, Physics Letters B **442**, 82 (1998).

[72] A. Dolgov and I. Khriplovich, Zhurnal Eksperimental noi i Teoreticheskoi Fiziki **85**, 1153 (1983).

[73] H. Fearn, Laser physics **17**, 695 (2007).

[74] H. Fearn, arXiv preprint arXiv:0706.0553 (2007).

[75] G. M. Shore, International Journal of Modern Physics A **18**, 4169 (2003).

[76] R. Jaffe, Physical Review D **72**, 021301 (2005).

[77] J. Bain, Erkenntnis **53**, 375 (2000).

[78] J. Earman, A. Arageorgis, and L. Ruetsche, Philosophy of Science **70**, 164 (2003).

[79] D. Fraser, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics **39**, 841 (2008).

[80] J. Bain, *Relativity and Quantum Field Theory* (Springer, 2010).

[81] D. B. Malament, in *Perspectives on quantum reality* (Springer, 1996), pp. 1–10.

[82] H. Halvorson and R. Clifton (2002).

[83] R. D. Klauber, *Student friendly quantum field theory* (Sandtrove Press, 2013).

[84] J. Dalibard, J. Dupont-Roc, and C. Cohen-Tannoudji, Journal de Physique **43**, 1617 (1982).

[85] L. Allen, *Optical Resonance and Two-level Atoms*, vol. 28 (Courier Corporation, 1975).

[86] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics, vol. 2: Mainly electromagnetism and matter* (Addison-Wesley, 1979).

[87] P. Milonni, *The quantum vacuum: an introduction to quantum electrodynamics* (Academic Press, 1994), ISBN 9780124980808, URL `http://books.google.com/books?id=P83vAAAAMAAJ`.

[88] J. Martin, Comptes Rendus Physique (2012).

[89] J. Martin, arXiv preprint arXiv:1205.3365 (2012).

[90] P. Milonni, Physica Scripta **1988**, 102 (1988).

[91] P. W. Milonni and W. A. Smith, Physical Review A **11**, 814 (1975).

[92] R. Jaffe, arXiv preprint hep-th/0503158 (2005).

[93] E. Lifshitz, Soviet Phys. JETP **2**, 73 (1956).

[94] D. W. Sciama, in *The philosophy of vacuum*, edited by S. Saunders and H. R. Brown (Clarendon Press, Oxford, 1991).

[95] H. B. Callen and T. A. Welton, Physical Review **83**, 34 (1951).

[96] H. Nyquist, Physical review **32**, 110 (1928).

[97] E. Jaynes, in *Coherence and Quantum Optics IV*, edited by L. Mandel and E. Wolf (Plenum Press, New York, 1978), URL `http://bayes.wustl.edu/etj/articles/electrodynamics.today.pdf`.

[98] P. Milonni, Physical Review A **25**, 1315 (1982).

[99] P. Milonni, Physics Letters A **82**, 225 (1981).

[100] J. Schwinger, L. L. DeRaad Jr, and K. A. Milton, Annals of Physics **115**, 1 (1978).

[101] G. V. Dunne, arXiv preprint hep-th/0406216 (2004).

[102] G. V. Dunne (2005).

[103] H. Euler and B. Kockel, Naturwissenschaften **23**, 246 (1935).

[104] H. Euler and B. Köckel, Annalen der Physik **26**, 398 (1936).

[105] V. F. Weisskopf, Dan. Mat. Fys. Medd. **14**, 1 (1936).

[106] V. Weisskopf (1936).

[107] K. Riley, M. Hobson, and S. Bence, *Mathematical methods for physics and engineering* (1998).

[108] J. Hilgevoord and J. Uffink, *The uncertainty principle-stanford encyclopedia of philosophy* (2006).

[109] H. R. Pagels, *The cosmic code: Quantum physics as the language of nature* (Courier Corporation, 1982).

[110] C. Callender, *Nothing to see here: Demoting the uncertainty principle*, `http://opinionator.blogs.nytimes.com/2013/07/21/nothing-to-see-here-demoting-the-uncertainty-principle/?_r=0` (July 21st 2013).

[111] W. Heisenberg, W. Heisenberg, W. Heisenberg, and W. Heisenberg, *Die physikalischen prinzipien der quantentheorie*, vol. 1 (S. Hirzel Leipzig, 1930).

[112] W. Heisenberg, U. Chicago Press, Chicago p. 21 (1930).

[113] W. Heisenberg, Zeitschrift für Physik **43**, 172 (1927).

[114] J. A. Wheeler and W. H. Zurek, *Quantum theory and measurement* (Princeton University Press, 1983).

[115] D. J. Griffiths and E. G. Harris, *Introduction to quantum mechanics*, vol. 2 (Prentice Hall New Jersey, 1995).

[116] T. C. Quinn and R. M. Wald, Physical Review D **60**, 064009 (1999).

[117] T. C. Quinn and R. M. Wald, Physical Review D **60**, 064009 (1999).

[118] S. E. Gralla, A. I. Harte, and R. M. Wald, Physical Review D **80**, 024031 (2009).

[119] S. E. Gralla, A. I. Harte, and R. M. Wald, arXiv preprint arXiv:0905.2391 (2009).

[120] J. F. Woodward, *Radiation Reaction, year = 1999, url = http://physics.fullerton.edu/ jimw/general/radreact/, urldate = 1998.*

Appendix A

VACUUM FIELD AND UNCERTAINTY RELATIONS

A.1 Preliminary: interpretations of the uncertainty relations (UR) in the context of non-relativistic quantum mechanics (NRQM)

The uncertainty relations have been interpreted in different ways. As Jan Hilgevoord and Jos Uffink note:

> The interpretation of these relations has often been debated. Do Heisenberg's relations express restrictions on the experiments we can perform on quantum systems, and, therefore, restrictions on the information we can gather about such systems; or do they express restrictions on the meaning of the concepts we use to describe quantum systems? Or else, are they restrictions of an ontological nature, i.e., do they assert that a quantum system simply does not possess a definite value for its position and momentum at the same time? The difference between these interpretations is partly reflected in the various names by which the relations are known, e.g. as 'inaccuracy relations', or: 'uncertainty', 'indeterminacy' or 'unsharpness relations'.[1]

Which interpretation of the uncertainty relations (UR) one tends to adopt is generally related to which interpretation of quantum mechanics as a whole one favors. For this reason, Jan Hilgevoord and Jos Uffink deem as the "minimal interpretation" that which "seems to be shared by both the adherents of the Copenhagen interpretation and of other views."[2]

---

[1] [108], p.13.

[2] [108], p.34. See also notably: [109], pp.82-83, [110].

The minimal interpretation is a statistical one, with the uncertainties identified with spreads of probability distributions: one considers a very large number of copies of the system all formally described by the same state vector,[3] and look at the outcomes of, say, position (or momentum) measurements on all these copies. The UR are then taken to mean that "the position and momentum distributions cannot both be arbitrarily narrow."[4]

What makes it possible for this description to serve as a common-ground interpretation is the fact that it is purely epistemic in the sense that it only restricts what we can know about the systems, while remaining agnostic regarding the source of this restriction. In addition to forming such common-ground, it is also minimal in a second sense: it does "little more than filling in the empirical meaning" of an inequality which is widely considered as the formal expression of the UR:

$$\Delta_\Psi p \, \Delta_\Psi q \geq \frac{\hbar}{2} \tag{A.1}$$

where $\Delta_\Psi p$ and $\Delta_\Psi q$ are *standard deviations* of momentum and position, respectively:

$$(\Delta_\Psi p)^2 = \langle\Psi|p^2|\Psi\rangle - \langle\Psi|p|\Psi\rangle^2, \tag{A.2}$$

and similarly for position.[5]

This is a particular case of a more general theorem about Hermitian operators (**A** and **B** here):

$$\Delta_\Psi \mathbf{A} \, \Delta_\Psi \mathbf{B} \geq \frac{1}{2}|\langle\Psi| \, [\mathbf{A}, \mathbf{B}] \, |\Psi\rangle|.^6 \tag{A.3}$$

---

[3] Which from an experimental standpoint implies that they must all be prepared in the same state.

[4] [108], pp.34-35.

[5] This was derived by Kennard in 1927, and Heisenberg presented it in his own work in 1930. [108], pp.19-20, [111], [112].

[6] This generalization was obtained by Robertson in 1929. [108], p.20.

Modern textbooks typically identify the UR with this inequality, which is the result of a formal derivation based on the quantum formalism itself, by contrast with Heisenberg's original formulation to which we shall soon turn. As can be seen from Eq.(A.3), the uncertainty involved is due to the commutation relations: when **A** and **B** commute, the RHS vanishes and the standard deviations can be simultaneously arbitrarily small; uncertainty only obtains when this is not the case.

The way Heisenberg originally discussed the UR had little to do with the spread of statistical distributions however.[7] Instead it was related to the inaccuracy of instruments. Both Heisenberg and Bohr saw in the UR more than a restriction on our ability to obtain information about quantum systems (of course in their view it entailed that too, but not only). This was the result of an operationalist standpoint: Heisenberg held that physical quantities (position, momentum) only have meaning if we can specify an experiment by which to measure them (what Hilgevoord and Uffink term the "measurement=meaning principle"). Restrictions in our ability to measure accurately (which *per se* is an epistemic issue) then imply corresponding inaccuracies in the *meaning* of physical quantities:

> If there are, as Heisenberg claims, no experiments that allow a simultaneous precise measurement of two conjugate quantities, then these quantities are also not simultaneously well-defined.[8]

---

[7] Even though he viewed in Kennard's result an exact proof of his UR, [108], p.20.

[8] [108], pp.6-8. In Heisenberg's view inaccuracies were related to the occurrence of discontinuous changes, i.e. in the thought experiment where he imagines trying to measure the position of an electron using light:

> At the instant of time when the position is determined, that is, at the instant when the photon is scattered by the electron, the electron undergoes a discontinuous change in momentum. This change is the greater the smaller the wavelength of the light employed, i.e., the more exact the determination of the position. At the instant at which the position of the electron is known, its momentum therefore can be known only up to magnitudes which correspond to that discontinuous change; thus, the more precisely the position is determined, the less precisely the momentum is known, and conversely.[113], pp.174-5, English translation [114], pp.62-84. See as well [108], p.7.

But Heisenberg occasionally expressed views that could be described as more extreme, if denying the *existence* of something can be deemed more extreme than denying the possibility of *defining* said thing. He famously wrote:

> I believe that one can formulate the emergence of the classical path of a particle pregnantly as follows: the path comes into being only because we observe it.[9]

on which Hilgevoord and Uffink comment:

> Apparently, in his view, a measurement does not only serve to give meaning to a quantity, it creates a particular value for this quantity. This may be called the measurement=creation principle. It is an ontological principle, for it states what is physically real.[10]

This in turn suggests an interpretation of the UR that can also be deemed ontological:

> Before the final measurement, the best we can attribute to the electron is some unsharp, or fuzzy momentum. These terms are meant here in an ontological sense, characterizing a real attribute of the electron.[11]

---

[9] [113], p.185.

[10] [108], p.12.

[11] [108], p.12. Hilgevoord and Uffink state this after summing up Heisenberg's views:

> First we measure the momentum of the electron very accurately. By measurement= meaning, this entails that the term "the momentum of the particle" is now well-defined. Moreover, by the measurement=creation principle, we may say that this momentum is physically real. Next, the position is measured with inaccuracy δq. At this instant, the position of the particle becomes well-defined and, again, one can regard this as a physically real attribute of the particle. However, the momentum has now changed by an amount that is unpredictable by an order of magnitude $|p_f - p_i| \sim \frac{\hbar}{\delta q}$. The meaning and validity of this claim can be verified by a subsequent momentum measurement. The question is then what status we shall assign to the momentum of the electron just before its final measurement. Is it real? According to Heisenberg it is not. Before the final measurement, the best we can attribute to the electron is some unsharp, or fuzzy momentum. These terms are meant here in an ontological sense, characterizing a real attribute of the electron.

Bohr's understanding of the UR further differed from Heisenberg's. Notably, Bohr did not ascribe the origin of the UR to discontinuous changes in momentum, but to his Principle of Complementarity:

> It is not so much the unknown disturbance which renders the momentum of the electron uncertain but rather the fact that the position and the momentum of the electron cannot be simultaneously defined in this experiment.[12]

Especially interesting is the derivation of the UR Bohr offered in 1928, in which uncertainties were defined in terms of the features of a wave packet.[13] Δx and Δt represented the spatial and temporal extensions of the wave packet, $\Delta\sigma$ and $\Delta\nu$ its range of inverse wave numbers ($\sigma = \frac{1}{k}$) and frequencies.

What is remarkable about this derivation is that it makes no use of commutation relations, in contrast to the now standard one discussed above. It relies instead on Fourier analysis to obtain the product of uncertainties, responsible for the restriction in defining quantities *simultaneously* (since as one factor increases the other must decrease):[14]

$$\Delta x \, \Delta\sigma \approx \Delta t \, \Delta\nu \approx 1. \tag{A.4}$$

Then Bohr obtained the UR by substituting for frequency and wavenumber in these relations, using, respectively, the relation between energy and frequency $E = h\nu$, and between momentum and wavelength $p = \frac{h}{\lambda}$:[15]

$$\Delta x \, \Delta p \approx \Delta t \, \Delta E \approx h. \tag{A.5}$$

---

[12] "Addition in Proof" to [113]; [108], p.24. The thought experiment in question involves determining the position of an electron by scattering a photon off it, thereby leading to a discontinuous change in the momentum of the electron.

[13] A wave packet is a superposition of waves of different frequencies, hence of different wave numbers.

[14] [108], p.29.

[15] *Idem.*

Two remarks are in order regarding the use of $E = h\nu$ and $p = \frac{h}{\lambda}$. First, it is through them that the quantum character of the relations appears: $\Delta x\,\Delta\sigma \approx \Delta t\,\Delta\nu \approx 1$ are obtained from purely classical considerations. This is also obvious from the fact that it is the substitution of $E = h\nu$ and $p = \frac{h}{\lambda}$ which introduces h in the final result, i.e. in the UR. Second, Eq. (A.4) can be thought to express the principle of complementarity so dear to Bohr, with E and p regarded as having to do with the particle aspect of entities and ν and σ the wave one.

A.2 Uncertainty relations, vacuum fluctuations and virtual particles

As briefly discussed in introduction, in NRQM a particle in a "harmonic oscillator" energy well has a ground state, "zero-point" energy of $\frac{1}{2}\hbar\omega$, and this is generally seen as a consequence of the UR for position and momentum, which implies corresponding URs for potential and kinetic energies respectively.

One derivation relating the UR to the "zero-point" energy goes as follows. If one considers the mechanical energy of the particle and:

- one substitutes the "momentum uncertainty" Δp for p in the kinetic energy term and the "position uncertainty" Δx for x in the potential energy specific to a harmonic oscillator,[16]
- one makes use of the minimal form of the UR $\Delta x \Delta p = \frac{\hbar}{2}$, substituting for Δp,
- one minimizes the resulting expression for the mechanical energy with respect to Δx,

one finds that this energy has a global minimum (at $\Delta x = \sqrt{\frac{\hbar}{2m\omega}}$), and that its value there is $\frac{1}{2}\hbar\omega$.[17]

Note that no assumption has been made regarding the exact meaning of Δx and Δp, i.e. whether they should be viewed as representing experimental inaccuracies due

---

[16] Taking the ground state energy to be: $E_0 = \frac{1}{2m}(\Delta p)^2 + \frac{1}{2}m\omega^2(\Delta x)^2$.

[17] Note that this derivation relies on using equality in the UR, and that this is possible here because we are dealing with an harmonic oscillator potential: the wave function of the particle in such a potential is a gaussian, which constitutes a case of minimum uncertainty.

to the limitations of instruments, statistical spreads such as standard deviations, or are associated with wave packets.

This derivation would seem to motivate an interpretation of the UR that goes beyond an epistemic one, whereby the relations would be understood as limiting only our knowledge of the position and momentum.[18] That is, if we interpret the zero-point energy result to mean that a particle in its ground state *actually has* an energy of $\frac{1}{2}\hbar\omega$, this would imply that the UR tell us something about the value a variable really has. How exactly this would come about seems less clear. One account would be to say that the uncertainty in the UR arises because of an intrinsic indeterminacy in the variables at stake, i.e. that there is truly no fact of the matter as to how much potential(/kinetic) energy the particle has exactly; that therefore the potential and kinetic energies cannot be both actually and exactly zero, hence their sum, the total energy, cannot either. In this story, ironically, the degree of indeterminacy in the component energies determines the exact value of their sum. This would correspond to an ontological interpretation of the UR as discussed by Hilgevoord and Uffink in connection to Heisenberg's ideas, since the potential and kinetic energies would have to be unsharp in such a sense.

If instead one views (Δx) and (Δp) as standard deviations, one can restrict oneself to the minimal interpretation of the UR, which is only epistemic. This would imply that the $\frac{1}{2}\hbar\omega$ value for the zero-point energy would have to be interpreted in a similar way.

We just considered zero-point energies in NRQM, we now turn back to QFT, and the half-quanta of its vacuum field. As stated previously, these are what are termed "vacuum fluctuations", and they are often described as evanescent "virtual particles" that borrow from the vacuum enough energy to become real for an amount of time too short for them to be observed, courtesy of the UR between time and

---

[18] Hence of the potential and kinetic energies.

energy.[19]

Now whether we are justified to talk about zero-point energies in terms of "fluctuations", which suggests a dynamical process, and whether we have ground to accept the "virtual particles" account are two different issues. Yet they are related: the latter requires the former, since "virtual particles" require energy to vary with time by small amounts, in turn increasing and decreasing — i.e. they require energy to fluctuate.

What makes the issue of whether this account is correct a fascinating one are its implications for the meaning of the UR. These play a double role in the scheme, because the uncertainty in energy and the uncertainty in time play different ones. For the energy uncertainty, the UR have to concern a matter of fact about the world, and to tell us something about the properties of entities — as could already be said in the context of zero-point energies in NRQM. So it requires the UR to be ontological in the sense used by Hilgevoord and Uffink. But it could be argued that now, in addition, it is ontological in a further sense: this "uncertainty" is responsible for the very existence of some entities. Indeed the scheme describes virtual particles to actually *exist* because they acquire enough energy to become full quanta courtesy of the UR. The reason the implications of zero-point energies in NRQM are not as extreme is because the latter cannot describe particle creation, so the quanta concern the energy of a particle that NRQM has to take as pre-existing. In QFT by contrast, the very *existence* of an entity (a particle) depends on the *value of a property* of another (the field): the energy quanta pertain to the field, and not to a particle from the outset; whether one is there or not depends on the value of a property of this field, i.e. its energy. While the energy is a full quantum, a particle actually exists. Its virtual character consists in its inability to maintain this existence long enough

---

[19] It may be worth stressing that this use of the phrase "virtual particles" differs from the standard, formal use of it in the context of Feynman diagrams notably: there "virtual particles" are particles that are "off shell", meaning that they are not obliged to satisfy the relativistic energy-momentum relation. What the two definitions have in common is that in both cases, the particles are said to be unobservable as a matter of principle.

to be observed, which brings us to the second role played by the UR.

While the uncertainty in energy is responsible for the "effect" itself (i.e. the existence of the particles), the uncertainty in time provides an explanation for what could otherwise be deemed a contradiction with experimental facts, i.e. our inability to observe it. Hence in this account the uncertainty in time is both about a physical fact (the particles exist for a short time), and a limit on our knowledge (virtual particles are unobservable as a matter of principle).

However, there are serious issues with the "virtual particles" account.

Indeed, what part the energy half quanta are thought to play in this description is not clear. Their relation to the concepts of vacuum fluctuations and virtual particles certainly seems to be mediated by the UR: virtual particles are explicitly said to be the direct result of time-energy UR, fluctuations are certainly ascribed to uncertainty (although exactly how is again unclear), and field half quanta are thought to be due to the UR by analogy to the zero-point energy of the NRQM harmonic oscillator.[20] This analogy is suggested by the fact that both the NRQM particle and the QFT electromagnetic field have non-zero ground state energies as a result of operators not commuting. In this context the field and canonical momentum of QFT are the analogs of the NRQM position and momentum respectively. However there are issues with this analogy.

A first issue is that what is needed for the "virtual particles" account to make sense is an analog of the NRQM energy-time UR, not position-momentum. To what extent this is problematic is questionable: being relativistic, QFT puts position and time on the one hand, and energy and momentum on the other on the same footing. Hence quantum field and canonical momentum are both functions of time and position. Since their commutation relation involves time as much as it does position, one could argue that the field-canonical momentum commutation relation is the relevant one. Furthermore, it has been argued that "in the context of special relativity the

---

[20] Particle in an harmonic oscillator potential well.

energy-time form might be thought of as a consequence of the position-momentum version, because x and t (or rather, ct) go together in the position-time four vector."[21] So the field-canonical momentum relation could be said to stand in the same relation to both the energy-time and the position-momentum URs.[22] However it is unclear how the "virtual particles" account should be phrased to reflect the field-canonical momentum relation.

A second problem with the "virtual particles" account, and with even merely talking of fluctuations is that they implicitly invoke a dynamical process. Although the field-canonical momentum commutation relation indirectly involves time, it can hardly be said to express dynamical evolution. The operators may be time dependent, but these commutation relations are usually considered at equal times, so it is difficult to see how an uncertainty in time would arise — never mind how that would then yield a dynamical account. On the other side of the correspondence, as noted by Robert Klauber, being energy eigenstates half quanta do not sit well with a dynamical story either:

> Those half quanta appear to simply be steadily "sitting" in the vacuum, and not "popping in and out" of it. There is no apparent mechanism whereby they exist part of the time, but not all of the time.[23]

Partly related to the latter is a third, and perhaps more worrisome issue — which too would undermine not only an account in terms of virtual particles, but even the soundness of speaking of fluctuations. There exists an important disanalogy between the NRQM commutation relation and its QFT counterpart: unlike position and momentum in NRQM, quantum fields and their conjugate momenta are not observable, in that their expectation value is zero. For this reason, Klauber has

---

[21] [115], p.112.

[22] even though only position-momentum is based on commutation relations.

[23] [83], p.270. Klauber's work is primarily a textbook on QFT but some parts of it, notably his chapter 10 entitled "The vacuum revisited" is a critical reflection on the uses of the phrase "virtual particles".

argued that the UR one can associate with this commutation relation is in fact "meaningless."[24]

So all in all, it is hard to see how the commutation relations responsible for the half quanta in QFT are related to the time-energy UR invoked in the "virtual particles" account — and perhaps required to speak of fluctuations too.

Can the time-energy UR nevertheless be used to justify this account, independently of any reference to the field-canonical momentum commutation relations? After all, both Heisenberg's and Bohr's views on the UR did not *require* commutation relations.[25] As we saw, Bohr even provided a derivation of the position-momentum as well as the time-energy UR without using commutation relations. And above all, the time-energy UR are not associated with a commutation relation anyway.

This approach may be more hopeful, but even in this case it is not clear how a dynamical account could emerge. One may seek inspiration in the fact that the time-energy UR has been used to account for the period of oscillation of a system between its stationary states: the energy difference between two eigenstates is related to the period of oscillation by this UR. However this requires the system to be in a superposition of the two energy eigenstates, which a field in the vacuum state by definition is not. Another suggestive account is the relationship between the lifetime of a particle type and the spread found in its mass (i.e. rest energy) through repeated measurements, which is also given by the time-energy UR. However in this account what the UR controls is the time it takes the particles to cease to exist; by contrast what we need as a matter of priority for the "virtual particles" account is for it to provide them with the

---

[24] [83], p.273. The same remarks apply to the commutation relation between creation and annihilation operators.

[25] The relation between UR and commutation relations were introduced by Kennard, as discussed above.

required energy to become full quanta in the first place. Still, it is a suggestive story.

Yet a different approach would be to seek dynamics without invoking the UR for that purpose. This would not provide much ground for the virtual particles account, but would presumably justify the use of the term "fluctuations". In this respect the FDT could be helpful, since it has been interpreted to show that in the $T = 0$ limit, the fluctuating force is due to the vacuum field.[26] This certainly would seem to require that the latter fluctuates, in the same, dynamical sense as the Planck radiation field does in the high temperature limit. Generally speaking, in applications of the FDT, zero-point effects and high-temperature limit, classical effects stand in the same relation to the fluctuating force, each contributing a term to it. It therefore seems reasonable that if thermal effects give rise to a fluctuating, dynamical force, so do the zero-point effects.

In so far that the associated zero-point field can be said to originate from the UR (which in itself is questionable), this would indeed seem to justify interpreting the latter in more than an epistemic sense but in an ontological one as well, i.e. as actually affecting the value of variables. This conclusion would seem to hold irrespective of whether these fluctuations are thought to require the pre-existence of the radiation field or not, since the UR does imply a zero-point term in both cases.

---

[26] i.e. the zero-point term in Eq. (5.93) for $E(\omega, T)$ manifests itself as fluctuations.

Appendix B

RADIATION REACTION AND SELF-FORCE

Historically, discussions regarding this self-force have been related to issues pertaining to the mass of the electron.[1] In the late 19th and early 20th centuries, it was hoped that this mass could perhaps be accounted for by electromagnetism alone. An "electromagnetic mass" of the electron, m_{elec}, was defined through the momentum not of the electron itself, but of its electromagnetic field.[2] The expression for this momentum involves the speed of the electron, and the coefficient multiplying it was defined to be m_{elec}:

$$p_{field} = \frac{2}{3}\frac{e^2}{ac^2}v, \tag{B.1}$$

where a is the radius of the electron, which was not assumed necessarily point-like, e is its charge and v its speed. Hence:

$$m_{elec} = \frac{2}{3}\frac{e^2}{ac^2}. \tag{B.2}$$

$$F = m_{elec}\,\ddot{x} - \frac{2}{3}\frac{e^2}{c^3}\,\dddot{x} + \gamma\frac{e^2 a}{c^4}\,\ddddot{x}, \tag{B.3}$$

Equation B.3 involves a number of issues.

First it implies that for a point charge, i.e. as $a \to 0$, terms higher than the radiation reaction one vanish. What is problematic in this limit is that because $m_{elec} = \frac{2}{3}\frac{e^2}{ac^2}$, the first term on the RHS goes to infinity. When discussing the issue, Richard

---

[1] [86], chapter 28.

[2] Indeed this momentum is derived using the Poynting vector, which implies that it is physically located in the electromagnetic field due to the charge.

Feynman hints that the assumption of a point-like distribution is likely unphysical.[3] Yet he did discuss several attempts to get rid of the offending term while retaining the radiation reaction one — necessary to preserve energy conservation in the context of radiating charges.[4] Dirac notably proposed that the self-force should be given by half the difference of the forces due to the retarded and advanced fields[5] - rather than the force due to the retarded fields only. This indeed turns out to do the trick, giving:

$$F = -\frac{2}{3}\frac{e^2}{c^3}\,\dddot{x} + \ \ldots\ (higher\ terms) \tag{B.4}$$

Feynman himself, together with Wheeler, preferred to do away with the concept of self-force, and interpret the needed radiation resistance force as due to the advanced fields of all the other charges in the universe.[6]

Both attempts therefore made use of the advanced solutions to Maxwell's equations, which are usually discarded as they represent waves travelling backward in time.

Another issue with eq.(B.3) has to do with the fact that the Abraham-Lorentz force term involves the third derivative of x, which implies that the force vanishes when the charge accelerates uniformly. Yet, a uniformly accelerating charge emits electromagnetic radiation, too. How can this force possibly ensure that energy is conserved in this case? Recall that in order to derive the expression for this force, we integrated over a period. It can be shown that more generally, if we consider a charge that starts and ends with no acceleration, the energy it radiates is accounted for by the work done by the Abraham Lorentz force on this charge *over the entire motion*; i.e. energy is conserved *on average*, not *instantaneoulsy*. How to account for this

---

[3] "The infinity arises because of the force of one part of the electron on another - because we have allowed what is perhaps a silly thing, the possibility of the "point" electron acting on itself".[86], section 28-4.

[4] [86] section 28-5

[5] i.e. advanced wave solutions to Maxwell's equations.

[6] [86], section 28-5

apparent violation of conservation of energy has been the object of some debate.[7] Third, the Abraham Lorentz force in equation B.3 implies that the acceleration of the charge would grow exponentially in the absence of an external force.[8] For a charge of mass m subjected to this force alone, Newton's second law takes the form:

$$\tau \dddot{x} = m\ddot{x}, \tag{B.5}$$

where $\tau = \frac{2}{3}\frac{e^2}{c^3}$, i.e.:

$$\tau \dot{a} = ma, \tag{B.6}$$

where $a = \ddot{x}$. The solution of this equation is:

$$a(t) = a(0)e^{\frac{t}{\tau}}. \tag{B.7}$$

This shows firstly, that $a(t)$ hence also the trajectory of the charge, depends on the value of the acceleration at the initial time $a(0)$, which is unusual. Second and most importantly, it indicates that the acceleration increases exponential in time, as advertised.

Admittedly, considering a charge subject to no other force besides the Abraham Lorentz one is not a satisfactory physical scenario. However the issue remains in the presence of external forces; the acceleration then takes the form:

$$a(t) = \left(a(0) - \frac{1}{m\tau} \int_0^t dt' F_{ext}(t')e^{-\frac{t'}{\tau}} \right) e^{\frac{t}{\tau}}. \tag{B.8}$$

This expression shows that even in the presence of additional forces, the acceleration increases exponentially — a behavior charges do not seem to display in nature —

---

[7] [116], [117], [118], [119], [120]; curiously Feynman does not stress this issue when he discusses radiation reaction in [86], Chapter 28. Instead he only insists that the Abraham Lorentz term is necessary in order to account for conservation of energy.

[8] [87], pp.156-157.

unless its initial value is:

$$a(0) = \frac{1}{m\tau} \int_0^\infty dt' F_{ext}(t') e^{-\frac{t'}{\tau}}, \tag{B.9}$$

which would then lead to the acceleration taking the form:

$$a(t) = \frac{1}{m\tau} \int_0^\infty dt' F_{ext}(t' + t) e^{-\frac{t'}{\tau}}. \tag{B.10}$$

Dirac notes this in 1938, and took it seriously as the solution to the runaway solution predicament. As one might expect with a gesture that consists in imposing a certain form to an “initial” quantity, the description implied has peculiar causal features. Specifically, the acceleration at time t, $a(t)$ depends on the force at time $t + t'$, that is *at a later time*. This feature of Dirac's solution is called “pre-acceleration”.

B.0.1 Retardation as the origin of the self-force

When discussing the self-force, Feynman ascribes a specific process to account for it. The latter requires that the electron be modelled as an extended object — to begin with at least. For simplicity, its charge is taken to be uniformly distributed on the surface of a sphere. Each region of the sphere exerts an electric repulsion on all the others. In the rest frame of the particle, these forces cancel out so that no net force arises on the particle as a whole (Fig. B.1).[9]

---

[9] [86], section 28-4, Fig 28-3 a.

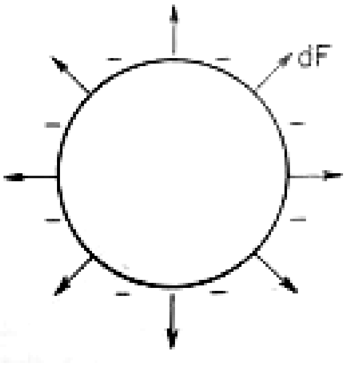


Fig. B.1: Self-force on an electron seen from its rest frame.

If the particle is accelerating however, this is no longer true. In Fig. B.2 below (where the dashed circle represents the electron at an earlier time than that indicated by the solid line), the influence of region β needs to propagate further to reach region α than if the electron was standing still. The force experienced by region α due to region β is therefore weaker in this case than for a stationary electron. By the same reasoning, the force experienced by region β due to region α is stronger.[10]

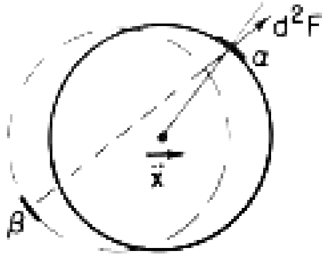


Fig. B.2: Self-force on an accelerating electron: retardation effect.

Hence in general the forces exerted by two parts of the electron on one another do not cancel out. Consequently, overall, the accelerating charged particle exerts

---

[10] [86], section 28-4, Fig. 28-3 b.

a net force on itself, in the direction opposite its acceleration: the “self-force”, or “self-reaction force” (Fig.B.3).[11]

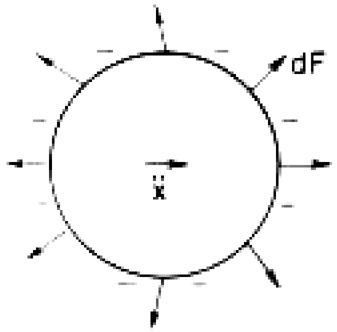


Fig. B.3: Self-force on an accelerating electron: net force.

As discussed above, what is called the radiation reaction force is a part of this self-reaction force.

As Feynman notes, one would expect these retardation effects to also affect a charge moving at constant velocity. He states that this is not the case, as evidenced by the expression he then gives, i.e. Eq. (B.3). And indeed were it so, it would mean that a self-force would act on a particle undergoing inertial motion, but not on a stationary one. Hence, provided that this force be measurable, it could be used to ascribe an absolute velocity to the charge. This would clearly violate both Galilean and Lorentz invariance. Unfortunately Feynman does not discuss how the retardation model is modified in order for the $\dot{x}$ contribution not to appear in Eq. (B.3).

---

[11] [86], section 28-4, Fig. 28-3 b.

Appendix C

UNDERDETERMINATION AND ORDERING OF OPERATORS

We examine the relationship between the hermitian character of the operator for a variable of interest, Q and the symmetric character of the ordering of the field operators E that enter in the definition of Q. We distinguish two cases: $Q = PE$, and Q given by terms of the form $E^-P^- + E^+P^+$, where P refers to an atomic variable.

For $Q = PE$, it is found that requiring Q to be hermitian implies that the field operators are symmetrically ordered.

When Q is given by terms of the form $E^-P^- + E^+P^+$, the hermiticity of Q is a necessary but not sufficient condition for the ordering of the field operators to be symmetric. The latter requires in addition that it be possible to rewrite the expression for Q in a form that involves products of only E with P, but not products of either E^+ or E^- with P^+ or P^-.

C.1 First case: Q given by PE

We denote by Q the operator for the quantity of interest, the field operator by E and the operator for the quantity that describes the atomic system by P.

If:

$$Q = PE, \tag{C.1}$$

and we refrain from making any assumption regarding the ordering of E and P within Q, the most general form the latter can take can be written:

$$Q = \lambda PE + (1 - \lambda)EP, \tag{C.2}$$

where λ is arbitrary. This implies that the contributions to Q due to the vacuum and source fields, i.e. respectively , Q_0 and Q_S can in turn be written:

$$Q_0 = \lambda P E_0 + (1-\lambda) E_0 P; \quad Q_S = \lambda P E_S + (1-\lambda) E_S P. \tag{C.3}$$

P, E_S and E_0 are Hermitian, so their conjugates are:

$$P = P^*; \quad E_0 = E_0^*; \quad E_S = E_S^*, \tag{C.4}$$

and therefore the Hermitian conjugates of Q^s and Q_0^* are:

$$Q_0^* = \lambda E_0^* P^* + (1-\lambda) P^* E_0^* = \lambda E_0 P + (1-\lambda) P E_0, \tag{C.5}$$

$$Q_S^* = \lambda E_S^* P^* + (1-\lambda) P^* E_S^* = \lambda E_S P + (1-\lambda) P E_S. \tag{C.6}$$

In order for Q_0 and Q_S to be hermitian, the following needs to be satisfied:

$$Q_0 = Q_0^*; \quad Q_S = Q_S^* \tag{C.7}$$

$$\lambda P E_0 + (1-\lambda) E_0 P = \lambda E_0 P + (1-\lambda) P E_0; \quad \lambda P E_S + (1-\lambda) E_S P = \lambda E_S P + (1-\lambda) P E_S \tag{C.8}$$

$$\lambda = 1 - \lambda \Rightarrow \lambda = \frac{1}{2} \tag{C.9}$$

That is, Q_0 and Q_S hermitian implies that $\lambda = \frac{1}{2}$. In this case the two operators take the form:

$$Q_0 = \frac{1}{2} P E_0 + \frac{1}{2} E_0 P; \quad Q_S = \frac{1}{2} P E_S + \frac{1}{2} E_S P \tag{C.10}$$

This corresponds to the following expression for Q:

$$\begin{aligned} Q &= Q_0 + Q_S \\ &= \frac{1}{2}(PE_0 + PE_S) + \frac{1}{2}(E_0P + E_SP) \\ &= \frac{1}{2}PE + \frac{1}{2}EP. \end{aligned} \tag{C.11}$$

If we now expand the field operators in terms of their positive and negative frequency components:

$$Q = \frac{1}{2}PE + \frac{1}{2}EP = \frac{1}{2}P(E^+ + E^-) + \frac{1}{2}(E^+ + E^-)P, \tag{C.12}$$

which corresponds to a choice of symmetric ordering of the field operators, as announced.[1]

C.2 Second case (two-level atom): Q given by terms of the form $E^-P^- + E^+P^+$

Now Q is no longer by PE but by a sum over terms of the form:

$$E^-P^- + E^+P^+, \tag{C.13}$$

where E^- and P^- can be ordered arbitrarily since they commute, and the same holds for E^+ and P^+. If we then choose to order them differently in both terms, i.e.:

$$E^-P^- + P^+E^+ \qquad \text{or} \qquad P^-E^- + E^+P^+ \tag{C.14}$$

---

[1] Recall that normal ordering consists in placing annihilation operators to the right of creation operators, anti-normal ordering the is the reverse, and symmetric ordering the linear combination of both in equal proportions.

we notice that we get back these two expressions in the same order when taking their hermitian conjugates:

$$(E^-P^- + P^+E^+)^\dagger = P^+E^+ + E^-P^- = E^-P^- + P^+E^+ \tag{C.15}$$

$$(P^-E^- + E^+P^+)^\dagger = E^+P^+ + P^-E^- = P^-E^- + E^+P^+ \tag{C.16}$$

If we now want to express this in terms of the separate contributions Q_0 and Q_S of the vacuum and source fields, the same structure is preserved as we substitute $E^+ = E_0^+ + E_S^+$ and $E^- = E_0^- + E_S^-$. That is, for instance:

$$Q_0^\dagger = (E_0^-P^- + P^+E_0^+)^\dagger = P^+E_0^+ + E_0^-P^- = E_0^-P^- + P^+E_0^+ = Q_0. \tag{C.17}$$

So we see that with both ordering choices in Eq.(C.14), Q_0 and Q_S are hermitian. Yet these two choices are not equivalent to one another: they constitute different choices of ordering that lead to different relative contributions from the vacuum and source fields.

In such a situation, i.e. when the variable of interest Q is given by a sum over terms of the form Eq.(C.13), removing the ordering freedom requires imposing an additional requirement, in addition to demanding that Q_0 and Q_S be hermitian. One needs to impose also that the expression for Q be rewritten in a form that involves products of only E with P,[2] not products of either E^+ or E^- with P^+ or P^- as above. The ordering that corresponds to this requirement is again the symmetric ordering.[3]

---

[2] Or products of appropriate combinations of $E^+ - E^-$ and $P^+ - P^-$.

[3] These derivations are clarified versions of what can be found in [84].

Appendix D

DERIVATION OF THE VACUUM EXPECTATION VALUES $\langle E_iE_j\rangle$ AND $\langle B_iB_j\rangle$ BETWEEN CASIMIR PLATES

Here we rederive the expressions for the vacuum expectation values of $\langle E_iE_j\rangle$ and $\langle B_iB_j\rangle$ between Casimir plates. The result corresponds to Eq. (6) in Gabriel Barton's 1990 paper.

D.1 Derivation of the vector potential $\mathbf{A}(\mathbf{r},t)$

In a region free of charges and currents, the three-vector potential $\mathbf{A}(\mathbf{r},t)$ must satisfy the homogeneous wave equation:

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial^2 t}\right)\mathbf{A}(\mathbf{r},t) \quad = \quad 0 \tag{D.1}$$

with the boundary conditions that it must vanish at the plates, i.e. at $z = 0$ and $z = L$.

In component notation this gives us the three equations:

$$\left(\partial_i^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)A_i(\mathbf{r},t) \quad = \quad 0, \qquad i = x, y, z. \tag{D.2}$$

We obtain the solutions by separation of variables.

We assume a solution of the form:

$$A_i = X_i\ Y_i\ Z_i\ T_i \tag{D.3a}$$

where:

$$A_i = A_i(x,y,z,t), \quad X_i = X_i(x), \quad Y_i = Y_i(y), \quad Z_i = Z_i(z), \quad T_i = T_i(t) \tag{D.3b}$$

After substituting this assumed form of the solution into Eq. (D.2), and dividing through by $X_i\ Y_i\ Z_i\ T_i$ one obtains:

$$\begin{aligned} \frac{1}{X_i}\frac{d^2}{dx^2}X_i + \frac{1}{Y_i}\frac{d^2}{dy^2}Y_i + \frac{1}{Z_i}\frac{d^2}{dz^2}Z_i - \frac{1}{c^2}\frac{1}{T_i}\frac{d^2}{dt^2}T_i &= 0 \\ \frac{1}{X_i}\frac{d^2}{dx^2}X_i + \frac{1}{Y_i}\frac{d^2}{dy^2}Y_i + \frac{1}{Z_i}\frac{d^2}{dz^2}Z_i = \frac{1}{c^2}\frac{1}{T_i}\frac{d^2}{dt^2}T_i &= -k^2 \qquad \text{(D.4)} \end{aligned}$$

where we have chosen the separation constant negative in anticipation of getting oscillatory solutions.

Let us first consider the space-dependent functions, X_i, Y_i, Z_i.

$$\frac{1}{X_i}\frac{d^2}{dx^2}X_i + \frac{1}{Y_i}\frac{d^2}{dy^2}Y_i + \frac{1}{Z_i}\frac{d^2}{dz^2}Z_i = -k^2 \qquad \text{(D.5)}$$

k^2 can be expressed as the sum:

$$k^2 \;=\; k_x^2 + k_y^2 + k_z^2 \qquad \text{(D.6)}$$

so that one can write:

$$\begin{aligned} \frac{1}{X_i}\frac{d^2}{dx^2}X_i &= -k_x^2 \qquad \text{(D.7)} \\ \frac{1}{Y_i}\frac{d^2}{dy^2}Y_i &= -k_y^2 \qquad \text{(D.8)} \end{aligned}$$

There is no boundary in the $x-$ nor in the $y-$direction, so the spatial components of X_i and Y_i, $i = x, y, z$ have the general form:

$$\begin{aligned} X_i &= e^{\mp i\ k_x x} \qquad \text{(D.9)} \\ Y_i &= e^{\mp i k_y y} \qquad \text{(D.10)} \end{aligned}$$

or more compactly:

$$X_i Y_i \quad = \quad e^{\mp i\mathbf{k}.\mathbf{r}_{||}} \tag{D.11}$$

Physically, the constant k_i represent the wave numbers of modes propagating in different directions.

The boundary conditions at $z = 0$ and $z = L$ determine Z_i since this is the part of A_i that varies with z. Z_i obeys:

$$\frac{1}{Z_i}\frac{d^2}{dz^2}Z_i \quad = \quad -k_z^2 \tag{D.12}$$

Therefore it has the general form:

$$Z_i \quad \alpha \quad e^{\pm i\, k_z\, z}$$

or equivalently:

$$Z_i \quad \alpha \quad \cos\left(k_z\, z\right) + \sin\left(k_z\, z\right)$$

Because $\mathbf{A}(\mathbf{r}, t) = \frac{\partial}{\partial t}\mathbf{E}(\mathbf{r}, t)$, $A_{||}$ obeys the same boundary conditions as $E_{||}$, and $\mathbf{E}(\mathbf{r}, t) = \nabla \times \mathbf{A}(\mathbf{r}, t)$ implies that $A_{\perp}$ obeys the same as $B_{\perp}$. Therefore:

$$A_{||} = A_x = A_y = 0 \qquad \text{at } z = 0, z = L$$

$$\Rightarrow Z_x(0) = Z_y(0) = 0 \tag{D.13a}$$

$$\Rightarrow Z_x(L) = Z_y(L) = 0 \tag{D.13b}$$

$$\partial_{\perp} A_{\perp} = \partial_z A_z = 0 \qquad \text{at } z = 0, z = L$$

$$\Rightarrow \partial_z Z_z(0) = 0 \tag{D.13c}$$

$$\Rightarrow \partial_z Z_z(L) = 0 \tag{D.13d}$$

The boundary condition Eq. (D.13a) implies that Z_x and Z_y involve $\sin\left(k_z\, z\right)$, since $\sin\left(k_z\, 0\right) = 0$ satisfies $Z_x = Z_y = 0$, whereas $\cos\left(\mathrm{k}_z\, 0\right) = 1$ does not.

The boundary condition Eq. (D.13b) at $z = L$ then means that $\sin(k_z\, L) = 0$, and since $\sin(n\pi) = 0$, n integer, this in turns implies that $k_z = \frac{n\pi}{L}$. Hence Z_x and Z_y are of the form:

$$Z_{x,y} \quad \alpha \quad \sin\left(\frac{n\pi}{L}\, z\right) \tag{D.14}$$

The boundary condition Eq. (D.13c) implies that Z_z involves $\cos(k_z\, z)$, since $\partial_z \cos(k_z\, z)|_{z=0} = -\sin(k_z\, 0) = 0$ satisfies $\partial_z Z_z(0) = 0$ whereas $\partial_z \sin(k_z\, z)|_{z=0} = cos\,(k_z\, 0) = 1$ does not. We have already established that $k_z = \frac{n\pi}{L}$, so Z_z is of the form:

$$Z_z \quad \alpha \quad \cos\left(\frac{n\pi}{L}\, z\right) \tag{D.15}$$

This indeed satisfies the fourth boundary condition, Eq. (D.13d), since $\partial_z \cos(\frac{n\pi}{L}\, z)|_{z=L} = \sin(n\pi) = 0$.

Finally, let us find the form of the time-dependent function, T_i. The second equality in Eq. (D.5) yields a solution for T_i of the form:

$$\begin{aligned} \frac{1}{c^2}\frac{1}{T_i}\frac{d^2}{dt^2}T_i &= -k^2 \\ \frac{d^2}{dt^2}T_i &= -\omega^2\, T_i \\ T_i &\quad \alpha \quad e^{\pm i\,\omega\, t} \end{aligned} \tag{D.16}$$

as $\omega = ck$. The latter also implies $\omega^2 = k_{||}^2 + k_z^2 = k_{||}^2 + (\frac{n\pi}{L})^2$ which will come useful later. Taking all these results into account, $A_i = X_i\ Y_i\ Z_i\ T_i$ has the form:

$$A_x = \sum_n c_x(\mathbf{k},\mathbf{n})\ \mathbf{e}^{\mp\mathbf{ik.r}_{||}}\ \sin\left(\frac{\mathbf{n}\pi}{\mathbf{L}}\,\mathbf{z}\right)\mathbf{e}^{\pm\mathbf{i}\,\omega\,\mathbf{t}} \tag{D.17a}$$

$$A_y = \sum_n c_y(\mathbf{k},\mathbf{n})\ \mathbf{e}^{\mp\mathbf{ik.r}_{||}}\ \sin\left(\frac{\mathbf{n}\pi}{\mathbf{L}}\,\mathbf{z}\right)\mathbf{e}^{\pm\mathbf{i}\,\omega\,\mathbf{t}} \tag{D.17b}$$

$$A_z = \sum_n c_z(\mathbf{k},\mathbf{n})\ \mathbf{e}^{\mp\mathbf{ik.r}_{||}}\ \cos\left(\frac{\mathbf{n}\pi}{\mathbf{L}}\,\mathbf{z}\right)\mathbf{e}^{\pm\mathbf{i}\,\omega\,\mathbf{t}} \tag{D.17c}$$

which can be written:

$$\begin{aligned}\mathbf{A} = \sum_{\mathbf{n}} \quad & \left[(c_x(\mathbf{k},\mathbf{n}) + \mathbf{c_y}(\mathbf{k},\mathbf{n}))\ \sin\left(\frac{\mathbf{n}\pi}{\mathbf{L}}\,\mathbf{z}\right)\hat{\mathbf{r}} + \mathbf{c_z}(\mathbf{k},\mathbf{n})\ \cos\left(\frac{\mathbf{n}\pi}{\mathbf{L}}\,\mathbf{z}\right)\hat{\mathbf{z}}\right]\ \mathbf{e}^{\mathbf{ik.r}_{||}-\mathbf{i}\,\omega\,\mathbf{t}} \\ + \quad & \left[(c_x(\mathbf{k},\mathbf{n}) + \mathbf{c_y}(\mathbf{k},\mathbf{n}))\ \sin\left(\frac{\mathbf{n}\pi}{\mathbf{L}}\,\mathbf{z}\right)\hat{\mathbf{r}} + \mathbf{c_z}(\mathbf{k},\mathbf{n})\ \cos\left(\frac{\mathbf{n}\pi}{\mathbf{L}}\,\mathbf{z}\right)\hat{\mathbf{z}}\right]\ \mathbf{e}^{-\mathbf{ik.r}_{||}+\mathbf{i}\,\omega\,\mathbf{t}}\end{aligned} \tag{D.18}$$

This result can be re-written in terms of the modes' direction of propagation $\hat{k}$:

$$c_i(\mathbf{k},\mathbf{n}) \quad = \quad \sum_\lambda a_\lambda\ \hat{\varepsilon}^i_\lambda = a_1\ \hat{\varepsilon}^i_1 + a_2\ \hat{\varepsilon}^i_2 \tag{D.19}$$

where $i = x, y, z$, and $\hat{\varepsilon}^i_\lambda$ is the polarization vector in the λ direction.[1]

The form of the $c_i(\mathbf{k},\mathbf{n})$ can be derived by imposing the orthogonality of the polarization

---

[1] i.e. $\mathbf{A}$ becomes:

$$\begin{aligned}\mathbf{A} = \sum_{\mathbf{n}} \quad & \left[(a_1^+\ \hat{\varepsilon}^x_1 + a_2^+\ \hat{\varepsilon}^x_2 + a_1^+\ \hat{\varepsilon}^y_1 + a_2^+\ \hat{\varepsilon}^y_2)\ \sin\left(\frac{n\pi}{L}\,z\right)\hat{r}\right. \\ + \quad & \left.(a_1^+\ \hat{\varepsilon}^z_1 + a_2^+\ \hat{\varepsilon}^z_2)\ \cos\left(\frac{n\pi}{L}\,z\right)\hat{z}\right]\ e^{i\mathbf{k.r}_{||}-i\,\omega\,t} \\ + \quad & \left[(a_1^-\ \hat{\varepsilon}^x_1 + a_2^-\ \hat{\varepsilon}^x_2 + a_1^-\ \hat{\varepsilon}^y_1 + a_2^-\ \hat{\varepsilon}^y_2)\ \sin\left(\frac{n\pi}{L}\,z\right)\hat{r}\right. \\ + \quad & \left.(a_1^-\ \hat{\varepsilon}^z_1 + a_2^-\ \hat{\varepsilon}^z_2)\ \cos\left(\frac{n\pi}{L}\,z\right)\hat{z}\right]\ e^{-i\mathbf{k.r}_{||}+i\,\omega\,t}\end{aligned}$$

where the a are still function of $\mathbf{k},\mathbf{n}$.

vectors $\hat{\varepsilon}_1$, $\hat{\varepsilon}_2$ and the wave vector $\hat{k}$:

$$\hat{\varepsilon}_1 \cdot \hat{\varepsilon}_2 = 0 \tag{D.20a}$$

$$\hat{\varepsilon}_1 \cdot \hat{k} = 0 \tag{D.20b}$$

$$\hat{\varepsilon}_1 \cdot \hat{k} = 0 \tag{D.20c}$$

as well as imposing the Coulomb gauge condition $\nabla \cdot \mathbf{A} = 0$.
This results in:

$$\mathbf{A}(\mathbf{r},t) = \frac{1}{\pi}\left(\frac{\pi}{L}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k \frac{1}{\omega^{\frac{1}{2}}}$$
$$\Bigg(e^{i\mathbf{k}.\mathbf{r}_{||}-i\omega t} \left\{ a_1^+ \sin\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} \times \hat{\mathbf{z}} + a_2^+ \left[\frac{i}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} - \frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right\}$$
$$+ e^{-i\mathbf{k}.\mathbf{r}_{||}+i\omega t} \left\{ a_1^- \sin\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} \times \hat{\mathbf{z}} + a_2^- \left[\frac{-i}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} - \frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right\}\Bigg) \tag{D.21}$$

Quantizing the field the functional coefficients are promoted to annihilation (a_1, a_2) and creation operators ($a_1^\dagger$, $a_2^\dagger$):

$$\begin{aligned} a_1^+ &\to a_1 \quad ; \quad a_2^+ \to a_2 \\ a_1^- &\to a_1^\dagger \quad ; \quad a_2^- \to a_2^\dagger \end{aligned} \tag{D.22}$$

Then:

$$\mathbf{A}(\mathbf{r},t) = \frac{1}{\pi}\left(\frac{\pi}{L}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k \frac{1}{\omega^{\frac{1}{2}}}$$
$$\Bigg(e^{i\mathbf{k}.\mathbf{r}_{||}-i\omega t} \left\{ a_1 \sin\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} \times \hat{\mathbf{z}} + a_2 \left[\frac{i}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} - \frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right\}$$
$$+ e^{-i\mathbf{k}.\mathbf{r}_{||}+i\omega t} \left\{ a_1^\dagger \sin\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} \times \hat{\mathbf{z}} + a_2^\dagger \left[\frac{-i}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} - \frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right\}\Bigg) \tag{D.23}$$

D.2 Derivation of the fields $\mathbf{E}(\mathbf{r},t)$ and $\mathbf{B}(\mathbf{r},t)$ from the vector potential $\mathbf{A}(\mathbf{r},t)$

$$
\begin{aligned}
&\mathbf{A}(\mathbf{r},t) = \sum_{\lambda} a_\lambda \mathbf{A}_\lambda(\mathbf{r}) e^{-i\omega_\lambda t} + \mathbf{A}_\lambda^\dagger a_\lambda^\dagger(\mathbf{r}) e^{+i\omega_\lambda t} \\
&= \frac{1}{\pi}\left(\frac{\pi}{L}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k \frac{1}{\omega^{\frac{1}{2}}} \\
&\Bigg(e^{i\mathbf{k}.\mathbf{r}_{||}-i\omega t} \left\{ a_1 \sin\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}}\times\hat{\mathbf{z}} + a_2 \left[\frac{i}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} - \frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right\} \\
&+ e^{-i\mathbf{k}.\mathbf{r}_{||}+i\omega t} \left\{ a_1^\dagger \sin\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}}\times\hat{\mathbf{z}} + a_2^\dagger \left[\frac{-i}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} - \frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right\}\Bigg)
\end{aligned}
\tag{D.24}
$$

where $a_i = a_i(\mathbf{k}, n)$, $\omega \equiv \left(k^2 + \frac{n^2\pi^2}{L^2}\right)^{\frac{1}{2}}$.

The electric field $\mathbf{E}(\mathbf{r},t)$ can be found from the vector potential $\mathbf{A}(\mathbf{r},t)$ through:

$$
\mathbf{E}(\mathbf{r},t) \;=\; -\frac{\partial}{\partial t}\mathbf{A}(\mathbf{r},t) \tag{D.25}
$$

$$
\begin{aligned}
&\mathbf{E}(\mathbf{r},t) = \sum_{\lambda} i\,\omega_\lambda\, a_\lambda \mathbf{A}_\lambda(\mathbf{r}) e^{-i\omega_\lambda t} - i\,\omega_\lambda\, \mathbf{A}_\lambda^\dagger a_\lambda^\dagger(\mathbf{r}) e^{+i\omega_\lambda t} \\
&= \frac{1}{\pi}\left(\frac{\pi}{L}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k\, \omega^{\frac{1}{2}} \\
&\Bigg(e^{i\mathbf{k}.\mathbf{r}_{||}-i\omega t} \left\{ i a_1 \sin\left(\frac{n\pi z}{L}\right) (\hat{\mathbf{k}}\times\hat{\mathbf{z}}) + a_2 \left[\frac{-1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} - i\frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right\} \\
&+ e^{-i\mathbf{k}.\mathbf{r}_{||}+i\omega t} \left\{ - i a_1^\dagger \sin\left(\frac{n\pi z}{L}\right) (\hat{\mathbf{k}}\times\hat{\mathbf{z}}) + a_2^\dagger \left[\frac{-1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} + i\frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right\}\Bigg)
\end{aligned}
\tag{D.26}
$$

The magnetic field $\mathbf{B}(\mathbf{r},t)$ is:

$$
\mathbf{B}(\mathbf{r},t) \;=\; \nabla \times \mathbf{A}(\mathbf{r},t) \tag{D.27}
$$

In order to find what this yield we first note that $\mathbf{A}(\mathbf{r},t)$ contains terms whose spatial part is of the form:

$$\sin\left(\frac{n\pi z}{L}\right)(\hat{\mathbf{k}}\times\hat{\mathbf{z}})\,e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} \tag{D.28a}$$

$$\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}\,e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} \tag{D.28b}$$

$$\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\,e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} \tag{D.28c}$$

These expressions are of the form $\mathbf{V}\Psi$,[2] with $\mathbf{V}$ equal to $\sin\left(\frac{n\pi z}{L}\right)(\hat{\mathbf{k}}\times\hat{\mathbf{z}})$, $\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}$ and $\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}$ respectively, and $\Psi = e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}$ Now:

$$\nabla\times(\mathbf{V}\;\Psi) \;=\; \Psi\,\nabla\times\mathbf{V}+\nabla\Psi\times\mathbf{V} \tag{D.29}$$

So these three expressions are going to contribute:

$$\begin{aligned}
\nabla\times\left(\sin\left(\frac{n\pi z}{L}\right)(\hat{\mathbf{k}}\times\hat{\mathbf{z}})\,e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}\right) &= e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}\,\nabla\times\left(\sin\left(\frac{n\pi z}{L}\right)(\hat{\mathbf{k}}\times\hat{\mathbf{z}})\right)+\nabla(e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}})\times\sin\left(\frac{n\pi z}{L}\right)(\hat{\mathbf{k}}\times\hat{\mathbf{z}})\\
\nabla\times\left(\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}\,e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}\right) &= e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}\,\nabla\times\left(\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}\right)+\nabla(e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}})\times\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}\\
\nabla\times\left(\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\,e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}\right) &= e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}\,\nabla\times\left(\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right)+\nabla(e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}})\times\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}
\end{aligned} \tag{D.30}$$

Now in the present situation we have:

$$\hat{\mathbf{k}} = \frac{k_x}{k}\hat{\mathbf{x}}+\frac{k_y}{k}\hat{\mathbf{y}} \tag{D.31a}$$

$$\hat{\mathbf{k}}\times\hat{\mathbf{z}} = \frac{k_y}{k}\hat{\mathbf{x}}-\frac{k_x}{k}\hat{\mathbf{y}} \tag{D.31b}$$

where $k=\sqrt{||\mathbf{k}||^2}=\sqrt{k_x^2+k_y^2}$.
And we know that for an arbitrary vector $\mathbf{C}$ lying in the x-y plane, as $\hat{\mathbf{k}}$ and $\hat{\mathbf{k}}\times\hat{\mathbf{z}}$ do:

$$\nabla\times\mathbf{C} \;=\; -\partial_z C_y\,\hat{\mathbf{x}}+\partial_z C_x\,\hat{\mathbf{y}}+(\partial_x C_y-\partial_y C_x)\,\hat{\mathbf{z}} \tag{D.32}$$

---

[2] i.e. where $\vec{V}$ is a vector field and Ψ a scalar field.

Therefore:

$$\begin{aligned}
\nabla \times \left(\sin\left(\frac{n\pi z}{L}\right)(\hat{\mathbf{k}} \times \hat{\mathbf{z}})\right) &= -\partial_z \left(\sin\left(\frac{n\pi z}{L}\right)\right)\left(-\frac{k_x}{k}\right)\hat{\mathbf{x}} + \partial_z \left(\sin\left(\frac{n\pi z}{L}\right)\right)\left(\frac{k_y}{k}\right)\hat{\mathbf{y}} + (0)\,\hat{\mathbf{z}} \\
&= \frac{k_x}{k}\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{x}} + \frac{k_y}{k}\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{y}} \\
&= \left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\left(\frac{k_x}{k}\hat{\mathbf{x}} + \frac{k_y}{k}\hat{\mathbf{y}}\right) \\
&= \left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} \qquad \text{(D.33)}
\end{aligned}$$

$$\begin{aligned}
\nabla \times \left(\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}\right) &= -\partial_z \left(\sin\left(\frac{n\pi z}{L}\right)\right)\left(\frac{k_y}{k}\right)\hat{\mathbf{x}} + \partial_z \left(\sin\left(\frac{n\pi z}{L}\right)\right)\left(\frac{k_x}{k}\right)\hat{\mathbf{y}} + (0)\,\hat{\mathbf{z}} \\
&= -\frac{k_y}{k}\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{x}} + \frac{k_x}{k}\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{y}} \\
&= \left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\left(-\frac{k_y}{k}\hat{\mathbf{x}} + \frac{k_x}{k}\hat{\mathbf{y}}\right) \\
&= \left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\left(-\hat{\mathbf{k}} \times \hat{\mathbf{z}}\right) \qquad \text{(D.34)}
\end{aligned}$$

and $\nabla \times \hat{\mathbf{z}} = \hat{\mathbf{x}}\,\partial_y - \hat{\mathbf{y}}\,\partial_x$:

$$\nabla \times \left(\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right) = \partial_y \cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{x}} - \partial_x \cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{y}} = 0 \qquad \text{(D.35)}$$

These results are going to allow us to find $\nabla \times \mathbf{V}$.

For $\nabla\Psi$:

$$\nabla e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} = \nabla e^{i\,(k_x x + k_y y)} = i k_x \hat{\mathbf{x}} + k_y \hat{\mathbf{y}})e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} = i\mathbf{k}\,e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} \qquad \text{(D.36)}$$

Hence for $\nabla\Psi \times \mathbf{V}$:

$$i\mathbf{k} \times \sin\left(\frac{n\pi z}{L}\right) (\hat{\mathbf{k}} \times \hat{\mathbf{z}})\, e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} = i\,\sin\left(\frac{n\pi z}{L}\right) \left(-\frac{k_x^2}{k} - \frac{k_y^2}{k}\right) \hat{\mathbf{z}}\, e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}$$

$$= -i\,\sin\left(\frac{n\pi z}{L}\right)\, k\,\hat{\mathbf{z}}\, e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} \tag{D.37a}$$

$$i\mathbf{k} \times \sin\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}}\, e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} = 0 \tag{D.37b}$$

$$i\mathbf{k} \times \cos\left(\frac{n\pi z}{L}\right) \hat{\mathbf{z}}\, e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} = i\,\cos\left(\frac{n\pi z}{L}\right)\, (k_y\hat{\mathbf{x}} - k_x\hat{\mathbf{y}})\, e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}$$

$$= i\,\cos\left(\frac{n\pi z}{L}\right)\, k\,(\hat{\mathbf{k}} \times \hat{\mathbf{z}})\, e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} \tag{D.37c}$$

where $k = \sqrt{k_x^2 + k_y^2}$.

So with Eq. (D.29):

$$\nabla \times (\mathbf{V}\ \Psi) \quad = \quad \Psi\ \nabla \times \mathbf{V} + \nabla\Psi \times \mathbf{V} \tag{D.38}$$

The three expressions Eq. (D.30) become:

$$\nabla \times \left(\sin\left(\frac{n\pi z}{L}\right) (\hat{\mathbf{k}} \times \hat{\mathbf{z}})\, e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}\right) = \left(\left(\frac{n\pi}{L}\right) \cos\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} - ik\,\sin\left(\frac{n\pi z}{L}\right)\ \hat{\mathbf{z}}\right) e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} \tag{D.39a}$$

$$\nabla \times \left(\sin\left(\frac{n\pi z}{L}\right)\ \hat{\mathbf{k}}\, e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}\right) = \left(\left(\frac{n\pi}{L}\right) \cos\left(\frac{n\pi z}{L}\right)\ \left(-\hat{\mathbf{k}} \times \hat{\mathbf{z}}\right)\right) e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} \tag{D.39b}$$

$$\nabla \times \left(\cos\left(\frac{n\pi z}{L}\right)\ \hat{\mathbf{z}}\, e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}}\right) = \left(ik\,\cos\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}} \times \hat{\mathbf{z}})\right) e^{i\,\mathbf{k}\cdot\mathbf{r}_{||}} \tag{D.39c}$$

Recalling the expression we obtained for **A** Eq. (D.24)

$$
\begin{aligned}
&\mathbf{A}(\mathbf{r},t) = \frac{1}{\pi}\left(\frac{\pi}{L}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k \frac{1}{\omega^{\frac{1}{2}}} \\
&\Bigg(\; e^{i\mathbf{k}.\mathbf{r}_{||}-i\omega t} \; \Bigg\{ a_1 \sin\left(\frac{n\pi z}{L}\right) \; \hat{\mathbf{k}} \times \hat{\mathbf{z}} \; + a_2 \left[\frac{i}{\omega}\frac{n\pi}{L} \sin\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} - \frac{k}{\omega} \cos\left(\frac{n\pi z}{L}\right) \hat{\mathbf{z}} \right] \Bigg\} \\
&+ e^{-i\mathbf{k}.\mathbf{r}_{||}+i\omega t} \Bigg\{ a_1^{\dagger} \sin\left(\frac{n\pi z}{L}\right) \; \hat{\mathbf{k}} \times \hat{\mathbf{z}} \; + a_2^{\dagger} \left[\frac{-i}{\omega}\frac{n\pi}{L} \sin\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} - \frac{k}{\omega} \cos\left(\frac{n\pi z}{L}\right) \hat{\mathbf{z}} \right] \Bigg\} \Bigg)
\end{aligned}
$$

B $= \nabla \times$ **A** becomes:

$$
\begin{aligned}
&\mathbf{B}(\mathbf{r},t) = \frac{1}{\pi}\left(\frac{\pi}{L}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k \frac{1}{\omega^{\frac{1}{2}}} \\
&\Bigg(\; e^{i\mathbf{k}.\mathbf{r}_{||}-i\omega t} \; \Bigg\{ a_1 \left(\left(\frac{n\pi}{L}\right) \cos\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} - ik \sin\left(\frac{n\pi z}{L}\right) \; \hat{\mathbf{z}} \right) \\
&+ a_2 \left[-\frac{i}{\omega}\left(\frac{n\pi}{L}\right)^2 \cos\left(\frac{n\pi z}{L}\right) \left(\hat{\mathbf{k}} \times \hat{\mathbf{z}}\right) \; - i\frac{k^2}{\omega} \cos\left(\frac{n\pi z}{L}\right) \; (\hat{\mathbf{k}} \times \hat{\mathbf{z}}) \right] \Bigg\} \\
&+ e^{-i\mathbf{k}.\mathbf{r}_{||}+i\omega t} \Bigg\{ a_1^{\dagger} \left(\left(\frac{n\pi}{L}\right) \cos\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} + ik \; \sin\left(\frac{n\pi z}{L}\right) \; \hat{\mathbf{z}} \right) \\
&+ a_2^{\dagger} \left[\frac{i}{\omega}\left(\frac{n\pi}{L}\right)^2 \cos\left(\frac{n\pi z}{L}\right) \left(\hat{\mathbf{k}} \times \hat{\mathbf{z}}\right) \; + i\frac{k^2}{\omega} \cos\left(\frac{n\pi z}{L}\right) \; (\hat{\mathbf{k}} \times \hat{\mathbf{z}}) \right] \Bigg\} \Bigg)
\end{aligned}
\tag{D.40}
$$

$$
\begin{aligned}
&\mathbf{B}(\mathbf{r},t) = \frac{1}{\pi}\left(\frac{\pi}{L}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k \frac{1}{\omega^{\frac{1}{2}}} \\
&\Bigg(\; e^{i\mathbf{k}.\mathbf{r}_{||}-i\omega t} \; \Bigg\{ a_1 \left(\left(\frac{n\pi}{L}\right) \cos\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} - ik \sin\left(\frac{n\pi z}{L}\right) \; \hat{\mathbf{z}} \right) \\
&+ a_2 \left[\left(-\frac{i}{\omega}\left(\left(\frac{n\pi}{L}\right)^2 + k^2 \right) \right) \cos\left(\frac{n\pi z}{L}\right) \; (\hat{\mathbf{k}} \times \hat{\mathbf{z}}) \right] \Bigg\} \\
&+ e^{-i\mathbf{k}.\mathbf{r}_{||}+i\omega t} \Bigg\{ a_1^{\dagger} \left(\left(\frac{n\pi}{L}\right) \cos\left(\frac{n\pi z}{L}\right) \hat{\mathbf{k}} + ik \; \sin\left(\frac{n\pi z}{L}\right) \; \hat{\mathbf{z}} \right) \\
&+ a_2^{\dagger} \left[\left(\frac{i}{\omega}\left(\left(\frac{n\pi}{L}\right)^2 + k^2 \right) \right) \cos\left(\frac{n\pi z}{L}\right) \; (\hat{\mathbf{k}} \times \hat{\mathbf{z}}) \right] \Bigg\} \Bigg)
\end{aligned}
\tag{D.41}
$$

Taking into account that $\omega^2 = \left(\frac{n\pi}{L}\right)^2 + k^2$:

$$\mathbf{B}(\mathbf{r},t) = \frac{1}{\pi}\left(\frac{\pi}{L}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k \frac{1}{\omega^{\frac{1}{2}}}$$
$$\Bigg(\quad e^{i\mathbf{k}.\mathbf{r}_{||}-i\omega t} \quad \left\{ a_1 \ \left[\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\ \hat{\mathbf{k}} - ik\ \sin\left(\frac{n\pi z}{L}\right)\ \hat{\mathbf{z}}\right] - ia_2\ \omega\ \cos\left(\frac{n\pi z}{L}\right)\ \left(\hat{\mathbf{k}}\times\hat{\mathbf{z}}\right) \right\}$$
$$+ e^{-i\mathbf{k}.\mathbf{r}_{||}+i\omega t}\left\{ a_1^\dagger \ \left[\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\ \hat{\mathbf{k}} + ik\ \ \sin\left(\frac{n\pi z}{L}\right)\ \hat{\mathbf{z}}\right] + ia_2^\dagger\ \omega\ \cos\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}}) \right\}\Bigg) \tag{D.42}$$

D.3 Derivation of $\langle E_i E_j\rangle$ and $\langle B_i B_j\rangle$ from the fields $\mathbf{E}(\mathbf{r},t)$ and $\mathbf{B}(\mathbf{r},t)$

From Eq. (D.43) for $\mathbf{E}(\mathbf{r},t)$ and Eq. (D.42) $\mathbf{B}(\mathbf{r},t)$ we want the expression for $\langle 0|E_iE_j|0\rangle$ and $\langle 0|B_iB_j|0\rangle$.
Since $a|0\rangle = 0$ and $\langle 0|a^\dagger = 0$, the only non-zero terms are of the form $aa^\dagger$:

$$\mathbf{E}(\mathbf{r},t) = \sum_\lambda i\ \omega_\lambda\ a_\lambda \mathbf{A}_\lambda(\mathbf{r})e^{-i\omega_\lambda t} - i\ \omega_\lambda\ \mathbf{A}_\lambda^\dagger a_\lambda^\dagger(\mathbf{r})e^{+i\omega_\lambda t}$$
$$= \frac{1}{\pi}\left(\frac{\pi}{L}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k\ \omega^{\frac{1}{2}}$$
$$\Bigg(\quad e^{i\mathbf{k}.\mathbf{r}_{||}-i\omega t} \quad \left\{ ia_1\ \sin\left(\frac{n\pi z}{L}\right)\ \ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})\ + a_2\left[\frac{-1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} - i\frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right] \right\}$$
$$+ e^{-i\mathbf{k}.\mathbf{r}_{||}+i\omega t}\left\{ - ia_1^\dagger\ \sin\left(\frac{n\pi z}{L}\right)\ \ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})\ + a_2^\dagger\left[\frac{-1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}} + i\frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right] \right\}\Bigg) \tag{D.43}$$

$$\langle 0|E_iE_j|0\rangle =$$

$$= \left\langle 0 \left| \frac{1}{\pi^2}\left(\frac{\pi}{L}\right) \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k\, \omega^{\frac{1}{2}} \right.\right.$$

$$\left(ia_1 \sin\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\ + a_2 \left[\frac{-1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}_i - i\frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right)$$

$$\sum_{n'=0}^{\infty} \frac{1}{(1+\delta_{n'0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k'\ \omega'^{\frac{1}{2}}$$

$$\left.\left(-\, ia_1^\dagger \sin\left(\frac{n'\pi z}{L}\right)\ (\hat{\mathbf{k}}'\times\hat{\mathbf{z}})_j\ + a_2^\dagger \left[\frac{-1}{\omega}\frac{n'\pi}{L}\sin\left(\frac{n'\pi z}{L}\right)\hat{\mathbf{k}}'_j + i\frac{k'}{\omega'}\cos\left(\frac{n'\pi z}{L}\right)\hat{\mathbf{z}}\right]\right) \right| 0 \right\rangle \quad \text{(D.44)}$$

$$\langle 0|B_iB_j|0\rangle =$$

$$= \left\langle 0 \left| \frac{1}{\pi^2}\left(\frac{\pi}{L}\right) \sum_{n=0}^{\infty} \frac{1}{(1+\delta_{n0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k\, \frac{1}{\omega^{\frac{1}{2}}} \right.\right.$$

$$\left(a_1\ \left[\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}_i - ik\,\sin\left(\frac{n\pi z}{L}\right)\ \hat{\mathbf{z}}\right] - ia_2\ \omega\ \cos\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\right)$$

$$\sum_{n'=0}^{\infty} \frac{1}{(1+\delta_{n'0})^{\frac{1}{2}}} \int_{-\infty}^{+\infty} d^2k'\ \frac{1}{\omega'^{\frac{1}{2}}}$$

$$\left.\left(a_1^\dagger\ \left[\left(\frac{n'\pi}{L}\right)\cos\left(\frac{n'\pi z}{L}\right)\hat{\mathbf{k}}'_j + ik'\ \ \sin\left(\frac{n'\pi z}{L}\right)\ \hat{\mathbf{z}}\right] + ia_2^\dagger\ \omega'\ \cos\left(\frac{n'\pi z}{L}\right)\ (\hat{\mathbf{k}}'\times\hat{\mathbf{z}})_j\right) \right| 0 \right\rangle \quad \text{(D.45)}$$

Now we know that the commutators of a and $a^\dagger$ obey:

$$[a_p(\mathbf{k},n), a^\dagger_{p'}(\mathbf{k}',n')] \quad = \quad \delta_{pp'}\ \delta(\mathbf{k},\mathbf{k}')\ \delta_{nn'}, \qquad \text{(D.46)}$$

and since:

$$\begin{aligned}\langle 0|[a_p(\mathbf{k},n), a^\dagger_{p'}(\mathbf{k}',n')]|0\rangle &= \langle 0|a_p(\mathbf{k},n)a^\dagger_{p'}(\mathbf{k}',n')|0\rangle - \langle 0|a^\dagger_{p'}(\mathbf{k}',n')a_p(\mathbf{k},n)|0\rangle \\ &= \langle 0|a_p(\mathbf{k},n)a^\dagger_{p'}(\mathbf{k}',n')|0\rangle, \end{aligned} \qquad \text{(D.47)}$$

each expression of the form $a_p(\mathbf{k},n)a^\dagger_{p'}(\mathbf{k}',n')$ in our vacuum expectation value is equivalent to $\delta_{pp'}\ \delta(\mathbf{k},\mathbf{k}')\ \delta_{nn'}$:

$$\begin{aligned}\langle 0|a_p(\mathbf{k},n)a^\dagger_{p'}(\mathbf{k}',n')|0\rangle &= \langle 0|\delta_{pp'}\ \delta(\mathbf{k},\mathbf{k}')\ \delta_{nn'}|0\rangle \\ &= \langle 0|0\rangle\delta_{pp'}\ \delta(\mathbf{k},\mathbf{k}')\ \delta_{nn'} \\ &= \delta_{pp'}\ \delta(\mathbf{k},\mathbf{k}')\ \delta_{nn'}\end{aligned}$$

So we can now write:

$$\begin{aligned}&\langle 0|E_iE_j|0\rangle = \\ &= \frac{1}{\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty} d^2k\ \omega \\ &\left(\ i\sin\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\ +\left[\frac{-1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}_i - i\frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right) \\ &\left(-i\sin\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j\ +\left[\frac{-1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}_j + i\frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right]\right)\end{aligned} \tag{D.48}$$

$$\begin{aligned}&\langle 0|B_iB_j|0\rangle = \\ &= \frac{1}{\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})^{\frac{1}{2}}}\int_{-\infty}^{+\infty} d^2k\ \frac{1}{\omega} \\ &\left(\left[\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}_i - ik\ \sin\left(\frac{n\pi z}{L}\right)\ \hat{\mathbf{z}}\right] - i\omega\ \cos\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\right) \\ &\left(\left[\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}_j + ik\ \ \sin\left(\frac{n\pi z}{L}\right)\ \hat{\mathbf{z}}\right] + i\omega\ \cos\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j\right)\end{aligned} \tag{D.49}$$

Expanding gives rise to terms that contain one of these expressions: $(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j$, $(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\ \hat{\mathbf{k}}_j$ or $\hat{\mathbf{k}}_i(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j$, $(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\ \hat{\mathbf{z}}$ or $\hat{\mathbf{z}}\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j$, $\hat{\mathbf{k}}_i\hat{\mathbf{k}}_j$, $\hat{\mathbf{k}}_i\hat{\mathbf{z}}$ or $\hat{\mathbf{z}}\hat{\mathbf{k}}_j$, $\hat{\mathbf{z}}\hat{\mathbf{z}}$.
$\hat{\mathbf{k}}$ lies in the x-y plane. So do $(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i$ and $(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j$: since one of the vectors in the cross products $\hat{\mathbf{k}}\times\hat{\mathbf{z}}$ is in the $\hat{\mathbf{z}}$ direction, $(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i$ and $(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j$ are normal to $\hat{\mathbf{z}}$.

Now for any unit vectors $\hat{\mathbf{C}}$ and $\hat{\mathbf{D}}$ that lies in the x-y plane:

$$\hat{\mathbf{C}}_i\hat{\mathbf{D}}_j = \begin{cases} 1 & i = j = 1 \text{ or } i = j = 2 \\ 0 & \text{otherwise} \end{cases}$$

so:

$$\hat{\mathbf{C}}_i\hat{\mathbf{D}}_j = \frac{1}{2}(\delta_{i1}\delta_{j1} + \delta_{i2}\delta_{j2}) \equiv \frac{1}{2}\delta^{||}_{ij} \tag{D.50}$$

This leads to the following results for the expressions of interest:

- $(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_i\ (\hat{\mathbf{k}} \times \hat{\mathbf{z}})_j$:

Both $(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_i$ and $(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_j$ lie in the x-y plane so from Eq. (D.50):

$$(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_i(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_j = \frac{1}{2}\delta^{||}_{ij} \tag{D.51}$$

- $(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_i\ \hat{\mathbf{k}}_j$ or $\hat{\mathbf{k}}_i(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_j$:

Both $(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_i$ and $\hat{\mathbf{k}}_j$ (or $\hat{\mathbf{k}}_i$ and $(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_j$) lie in the x-y plane so from Eq. (D.50):

$$(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_i\ \hat{\mathbf{k}}_j = \hat{\mathbf{k}}_i(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_j = \frac{1}{2}\delta^{||}_{ij} \tag{D.52}$$

- $\hat{\mathbf{k}}_i\hat{\mathbf{k}}_j$:

Both $\hat{\mathbf{k}}_i$ and $\hat{\mathbf{k}}_j$ lie in the x-y plane so from Eq. (D.50):

$$\hat{\mathbf{k}}_i\hat{\mathbf{k}}_j = \frac{1}{2}\delta^{||}_{ij} \tag{D.53}$$

- $(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_i\ \hat{\mathbf{z}}$ or $\hat{\mathbf{z}}\ (\hat{\mathbf{k}} \times \hat{\mathbf{z}})_j$:

Since $(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_i$ and $(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_j$ are normal to $\hat{\mathbf{z}}$:

$$(\hat{\mathbf{k}} \times \hat{\mathbf{z}})_i\ \hat{\mathbf{z}} = \hat{\mathbf{z}}\ (\hat{\mathbf{k}} \times \hat{\mathbf{z}})_j = 0 \tag{D.54}$$

- $\hat{\mathbf{k}}_i\hat{\mathbf{z}}$ or $\hat{\mathbf{z}}\hat{\mathbf{k}}_j$:

Since $\hat{\mathbf{k}}_i\hat{\mathbf{z}}$ and $\hat{\mathbf{z}}\hat{\mathbf{k}}_j$ are normal to $\hat{\mathbf{z}}$:

$$\hat{\mathbf{k}}_i\hat{\mathbf{z}} = \hat{\mathbf{z}}\hat{\mathbf{k}}_j = 0 \tag{D.55}$$

- $\hat{\mathbf{z}}\hat{\mathbf{z}}$:

$\hat{\mathbf{z}}\hat{\mathbf{z}} = 1$. This can be written in terms of i and j as well:

$$\hat{\mathbf{x}}_i\hat{\mathbf{x}}_j = \begin{cases} \hat{\mathbf{z}}\hat{\mathbf{z}} = 1 & i = j = 3 \\ 0 & \text{otherwise} \end{cases}$$

So:

$$\hat{\mathbf{z}}\hat{\mathbf{z}} = \delta_{i3}\delta_{j3} \equiv \delta_{ij}^{\perp} \tag{D.56}$$

Taking these results into account, Eq. (D.48) and Eq. (D.49) become:
- with Eq. (D.54) and Eq. (D.55):

$$\begin{aligned}
&\langle 0|E_iE_j|0\rangle = \\
&= \frac{1}{\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty} d^2k\ \omega \\
&\Bigg(\left(i\,\sin\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\right)\left(-\,i\,\sin\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j\right) \\
&+\left(i\,\sin\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\right)\left(-\,\frac{1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}_j\right) \\
&+\left(-\,\frac{1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}_i\right)\left(-\,i\,\sin\left(\frac{n\pi z}{L}\right)\ (\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j\right) \\
&+\left(-\,\frac{1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}_i\right)\left(-\frac{1}{\omega}\frac{n\pi}{L}\sin\left(\frac{n\pi z}{L}\right)\hat{\mathbf{k}}_j\right) \\
&+\left(-\,i\,\frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right)\left(i\frac{k}{\omega}\cos\left(\frac{n\pi z}{L}\right)\hat{\mathbf{z}}\right)\Bigg)
\end{aligned} \tag{D.57}$$

$$\langle 0|B_iB_j|0\rangle =$$

$$= \frac{1}{\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty}d^2k\,\frac{1}{\omega}$$

$$\Bigg(\left(\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\,\hat{\mathbf{k}}_i\right)\left(\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\,\hat{\mathbf{k}}_j\right)$$

$$+\left(\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\,\hat{\mathbf{k}}_i\right)\left(i\omega\,\cos\left(\frac{n\pi z}{L}\right)\,(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j\right)$$

$$+\left(-\,ik\,\sin\left(\frac{n\pi z}{L}\right)\,\hat{\mathbf{z}}\right)\left(ik\ \ \sin\left(\frac{n\pi z}{L}\right)\,\hat{\mathbf{z}}\right)$$

$$+\left(-\,i\omega\,\cos\left(\frac{n\pi z}{L}\right)\,(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\right)\left(\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi z}{L}\right)\,\hat{\mathbf{k}}_j\right)$$

$$+\left(-\,i\omega\,\cos\left(\frac{n\pi z}{L}\right)\,(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_i\right)\left(i\omega\,\cos\left(\frac{n\pi z}{L}\right)\,(\hat{\mathbf{k}}\times\hat{\mathbf{z}})_j\right)\Bigg) \tag{D.58}$$

- now using Eq. (D.51), Eq. (D.52), Eq. (D.53), and Eq. (D.56):

$$\langle 0|E_iE_j|0\rangle =$$

$$= \frac{1}{\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty}d^2k\,\omega$$

$$\Bigg(\sin^2\left(\frac{n\pi z}{L}\right)\,\frac{1}{2}\delta^{||}_{ij}\;-\;i\frac{1}{\omega}\frac{n\pi}{L}\sin^2\left(\frac{n\pi z}{L}\right)\,\frac{1}{2}\delta^{||}_{ij}\;+\;i\,\frac{1}{\omega}\frac{n\pi}{L}\sin^2\left(\frac{n\pi z}{L}\right)\,\frac{1}{2}\delta^{||}_{ij}$$

$$+\left(\frac{1}{\omega}\frac{n\pi}{L}\right)^2\sin^2\left(\frac{n\pi z}{L}\right)\frac{1}{2}\delta^{||}_{ij}\;+\;\left(\frac{k}{\omega}\right)^2\cos^2\left(\frac{n\pi z}{L}\right)\,\delta^{\perp}_{ij}\Bigg) \tag{D.59}$$

$$\langle 0|E_iE_j|0\rangle = \frac{1}{\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty}d^2k$$

$$\left(\frac{1}{2}\,\omega\,\sin^2\left(\frac{n\pi z}{L}\right)\,\delta^{||}_{ij}\;+\frac{1}{2\omega}\left(\frac{n\pi}{L}\right)^2\sin^2\left(\frac{n\pi z}{L}\right)\delta^{||}_{ij}\;+\;\frac{k^2}{\omega}\cos^2\left(\frac{n\pi z}{L}\right)\,\delta^{\perp}_{ij}\right) \tag{D.60}$$

$$\langle 0|B_iB_j|0\rangle = \frac{1}{\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty} d^2k\,\frac{1}{\omega}$$

$$\Bigg(\left(\frac{n\pi}{L}\right)^2\cos^2\left(\frac{n\pi z}{L}\right)\,\frac{1}{2}\delta_{ij}^{||} + i\omega\left(\frac{n\pi}{L}\right)\cos^2\left(\frac{n\pi z}{L}\right)\,\frac{1}{2}\delta_{ij}^{||} + k^2\,\sin^2\left(\frac{n\pi z}{L}\right)\,\delta_{ij}^{\perp}$$

$$-\,i\omega\left(\frac{n\pi}{L}\right)\,\cos^2\left(\frac{n\pi z}{L}\right)\,\frac{1}{2}\delta_{ij}^{||} + \omega^2\cos^2\left(\frac{n\pi z}{L}\right)\,\frac{1}{2}\delta_{ij}^{||}\Bigg) \tag{D.61}$$

$$\langle 0|B_iB_j|0\rangle = \frac{1}{\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty} d^2k$$

$$\left(\frac{1}{2}\frac{1}{\omega}\left(\frac{n\pi}{L}\right)^2\cos^2\left(\frac{n\pi z}{L}\right)\,\delta_{ij}^{||} + \frac{k^2}{\omega}\,\sin^2\left(\frac{n\pi z}{L}\right)\,\delta_{ij}^{\perp} + \frac{1}{2}\omega\,\cos^2\left(\frac{n\pi z}{L}\right)\delta_{ij}^{||}\right) \tag{D.62}$$

Now:

$$\sin^2\left(\frac{n\pi z}{L}\right) = \frac{1}{2} - \frac{1}{2}\cos\left(\frac{2n\pi z}{L}\right) \tag{D.63}$$

$$\cos^2\left(\frac{n\pi z}{L}\right) = \frac{1}{2} + \frac{1}{2}\cos\left(\frac{2n\pi z}{L}\right) \tag{D.64}$$

So $\langle 0|E_iE_j|0\rangle$ and $\langle 0|B_iB_j|0\rangle$ can be written in the form:

$$\langle 0|E_iE_j|0\rangle =$$

$$= \frac{1}{4\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty} d^2k$$

$$\left\{\left(2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp} + \left(\omega\,+\frac{1}{\omega}\left(\frac{n\pi}{L}\right)^2\right)\delta_{ij}^{||}\right) + \left(2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp} - \left(\omega\,+\frac{1}{\omega}\left(\frac{n\pi}{L}\right)^2\right)\delta_{ij}^{||}\right)\cos\left(\frac{2n\pi z}{L}\right)\right\} \tag{D.65}$$

$$\langle 0|B_iB_j|0\rangle =$$

$$= \frac{1}{4\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty} d^2k$$

$$\left\{\left(2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp}+\left(\omega\;+\frac{1}{\omega}\left(\frac{n\pi}{L}\right)^2\right)\delta_{ij}^{||}\right)\;+\;\left(-\,2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp}+\left(\omega\;+\frac{1}{\omega}\left(\frac{n\pi}{L}\right)^2\right)\delta_{ij}^{||}\right)\cos\left(\frac{2n\pi z}{L}\right)\;\right\} \tag{D.66}$$

$$\begin{aligned} \omega^2 &= k^2+\left(\frac{n\pi}{L}\right)^2 \\ \Rightarrow \omega+\frac{1}{\omega}\left(\frac{n\pi}{L}\right)^2 &= 2\,\omega-\frac{k^2}{\omega} \end{aligned} \tag{D.67}$$

$$\langle 0|E_iE_j|0\rangle = \frac{1}{4\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty} d^2k$$

$$\left\{\left(2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp}+\left(2\,\omega-\frac{k^2}{\omega}\right)\delta_{ij}^{||}\right)\;+\;\left(2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp}-\left(2\,\omega-\frac{k^2}{\omega}\right)\delta_{ij}^{||}\right)\cos\left(\frac{2n\pi z}{L}\right)\;\right\} \tag{D.68}$$

$$\langle 0|B_iB_j|0\rangle = \frac{1}{4\pi^2}\left(\frac{\pi}{L}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_{-\infty}^{+\infty} d^2k$$

$$\left\{\left(2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp}+\left(2\,\omega-\frac{k^2}{\omega}\right)\delta_{ij}^{||}\right)\;+\;\left(-\,2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp}+\left(2\,\omega-\frac{k^2}{\omega}\right)\delta_{ij}^{||}\right)\cos\left(\frac{2n\pi z}{L}\right)\;\right\} \tag{D.69}$$

Now we know that:

$$\int d^dk\; f(k) \;=\; \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}\int_o^{+\infty} dk\; k^{d-1}f(k)$$

with $d = 2$:

$$= \; 2\pi\int_o^{+\infty} dk\; kf(k) \tag{D.70}$$

so $\langle 0|E_iE_j|0\rangle$ and $\langle 0|B_iB_j|0\rangle$ become:

$$\langle 0|E_iE_j|0\rangle = \frac{1}{2L}\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_0^{+\infty} dk\ k$$

$$\left\{\left(2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp} + \left(2\,\omega - \frac{k^2}{\omega}\right)\delta_{ij}^{||}\right)\ +\ \left(2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp} - \left(2\,\omega - \frac{k^2}{\omega}\right)\delta_{ij}^{||}\right)\cos\left(\frac{2n\pi z}{L}\right)\ \right\} \tag{D.71}$$

$$\langle 0|B_iB_j|0\rangle = \frac{1}{2L}\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_0^{+\infty} dk\ k$$

$$\left\{\left(2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp} + \left(2\,\omega - \frac{k^2}{\omega}\right)\delta_{ij}^{||}\right)\ +\ \left(-\,2\,\frac{k^2}{\omega}\,\delta_{ij}^{\perp} + \left(2\,\omega - \frac{k^2}{\omega}\right)\delta_{ij}^{||}\right)\cos\left(\frac{2n\pi z}{L}\right)\ \right\} \tag{D.72}$$

The following integrals now need to be performed:

$$\int_0^{+\infty} dk\ k\ \frac{k^2}{\omega} \quad ; \quad \int_0^{+\infty} dk\ k\ \omega \tag{D.73}$$

With $\omega = \left(k^2 + \left(\frac{n\pi}{L}\right)^2\right)^{\frac{1}{2}}$ this is:

$$\int_0^{+\infty} dk\ \frac{k^3}{\left(k^2+\left(\frac{n\pi}{L}\right)^2\right)^{\frac{1}{2}}} \quad ; \quad \int_0^{+\infty} dk\ k\ \left(k^2+\left(\frac{n\pi}{L}\right)^2\right)^{\frac{1}{2}}$$

$$\int_0^{+\infty} dk\ \frac{k^3}{(k^2+l^2)^{\frac{1}{2}}} \quad ; \quad \int_0^{+\infty} dk\ \frac{k}{(k^2+l^2)^{-\frac{1}{2}}} \tag{D.74}$$

with $\frac{n\pi}{L} \equiv l$. With the change of variable $k \rightarrow lt$:

$$\int_0^{+\infty} dt\ \frac{t^3\,l^4}{l\,(t^2+1)^{\frac{1}{2}}} \quad ; \quad \int_0^{+\infty} dt\ \frac{t\,l^2}{l^{-1}\,(t^2+1)^{-\frac{1}{2}}}$$

$$l^3\int_0^{+\infty} dt\ \frac{t^3}{(t^2+1)^{\frac{1}{2}}} \quad ; \quad l^3\int_0^{+\infty} dt\ \frac{t}{(t^2+1)^{-\frac{1}{2}}} \tag{D.75}$$

These integrals are of the form $\int_0^{+\infty} dt\ \frac{t^a}{(t^2+1)^b}$. They can be evaluated by making use of the relation between the Beta function and Gamma functions.

$$B(x,y) \;=\; \int_0^{+\infty} dq \frac{q^{x-1}}{(q+1)^{x+y}}, \quad Re(x) > 0, Re(y) > 0 \tag{D.76}$$

With the change of variable $q \rightarrow t^2$ this becomes:

$$B(x,y) \;=\; 2\int_0^{+\infty} dt \frac{t^{2x-1}}{(t^2+1)^{x+y}}, \quad Re(x) > 0, Re(y) > 0 \tag{D.77}$$

Now the Beta function is related to Gamma functions by:

$$B(x,y) \;=\; \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{D.78}$$

Therefore our integrals can be evaluated using the relation:

$$\int_0^{+\infty} dt \frac{t^{2x-1}}{(t^2+1)^{x+y}} \;=\; \frac{1}{2}\ \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{D.79}$$

$l^3 \int_0^{+\infty} dt\ \frac{t^3}{(t^2+1)^{\frac{1}{2}}}$ corresponds to $x = 2$, $y = -\frac{3}{2}$ so:

$$l^3 \int_0^{+\infty} dt\ \frac{t^3}{(t^2+1)^{\frac{1}{2}}} \;=\; l^3\ \frac{1}{2}\ \frac{\Gamma(2)\Gamma(-\frac{3}{2})}{\Gamma(\frac{1}{2})} = l^3\ \frac{2}{3} \tag{D.80}$$

$l^3 \int_0^{+\infty} dt\ \frac{t}{(t^2+1)^{-\frac{1}{2}}}$ corresponds to $x = 1$, $y = -\frac{3}{2}$ so:

$$l^3 \int_0^{+\infty} dt\ \frac{t}{(t^2+1)^{-\frac{1}{2}}} \;=\; l^3\ \frac{1}{2}\ \frac{\Gamma(1)\Gamma(-\frac{3}{2})}{\Gamma(-\frac{1}{2})} = l^3 \left(-\frac{1}{3}\right) \tag{D.81}$$

Summarising these results:

$$\int_0^{+\infty} dk\ k\ \frac{k^2}{\omega} = \frac{2}{3}\left(\frac{n\pi}{L}\right)^3 \quad ; \quad \int_0^{+\infty} dk\ k\ \omega\ = -\frac{1}{3}\left(\frac{n\pi}{L}\right)^3 \tag{D.82}$$

This allows us to evaluate the expressions:

$$\begin{aligned}
\frac{1}{2L}\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_0^{+\infty} dk\ k\left(2\,\frac{k^2}{\omega}\right) \quad &; \quad \frac{1}{2L}\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\int_0^{+\infty} dk\ k\left(2\,\omega-\frac{k^2}{\omega}\right)\\
\frac{1}{2L}\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\,\frac{4}{3}\left(\frac{n\pi}{L}\right)^3 \quad &; \quad \frac{1}{2L}\sum_{n=0}^{\infty}\frac{1}{(1+\delta_{n0})}\left(-\frac{2}{3}-\frac{2}{3}\right)\left(\frac{n\pi}{L}\right)^3\\
\frac{2}{3}\,\frac{\pi^3}{L^4}\sum_{n=1}^{\infty}n^3 \quad &; \quad -\frac{2}{3}\,\frac{\pi^3}{L^4}\sum_{n=1}^{\infty}n^3
\end{aligned} \tag{D.83}$$

And the relevant expressions involving $\sum_{n=1}^{\infty} n^3$ can be determined as follows. $\sum_{n=1}^{\infty} n^3$ itself can be evaluated from its relation to the Zeta function:

$$\begin{aligned}
\zeta(s) &= \sum_{n=1}^{\infty}\frac{1}{n^s}\\
&\text{so:}\\
\sum_{n=1}^{\infty}n^3 &= \zeta(-3)=\frac{1}{120}
\end{aligned} \tag{D.84}$$

and $\sum_{n=1}^{\infty}\frac{1}{n^s}\cos\left(\frac{2n\pi z}{L}\right)$ gives:

$$\sum_{n=1}^{\infty}n^3\cos\left(\frac{2n\pi z}{L}\right) = \frac{2+\cos\left(\frac{2\pi z}{L}\right)}{8\,\sin^4\left(\frac{\pi z}{L}\right)} \tag{D.85}$$

Using the trigonometric identities:

$$\cos\,(2\theta) = 1-2\,\sin^2\theta \tag{D.86a}$$

$$\sin\theta = \cos\left(\theta-\frac{\pi}{2}\right) \tag{D.86b}$$

$$\sum_{n=1}^{\infty}n^3\cos\left(\frac{2n\pi z}{L}\right) = \frac{3-2\,\cos^2\left(\frac{\pi z}{L}-\frac{\pi}{2}\right)}{8\,\cos^4\left(\frac{\pi z}{L}-\frac{\pi}{2}\right)} \tag{D.87}$$

with $\xi \equiv \frac{2\pi z}{L} - \pi$:

$$\sum_{n=1}^{\infty} n^3 \cos\left(\frac{2n\pi z}{L}\right) \quad = \quad \frac{3 - 2\cos^2\left(\frac{\xi}{2}\right)}{8\cos^4\left(\frac{\xi}{2}\right)} \equiv f(\xi) \tag{D.88}$$

Substituting Eq. (D.83), Eq. (D.84) and Eq. (D.88) in Eq. (D.71) for $\langle 0|E_iE_j|0\rangle$ and Eq. (D.72) for $\langle 0|B_iB_j|0\rangle$:

$$\begin{aligned}\langle 0|E_iE_j|0\rangle = \Bigg| & \left\{\left(\delta_{ij}^{\perp}\left(\frac{2}{3}\,\frac{\pi^3}{L^4}\,\sum_{n=1}^{\infty} n^3\right) - \delta_{ij}^{||}\left(\frac{2}{3}\,\frac{\pi^3}{L^4}\,\sum_{n=1}^{\infty} n^3\right)\right)\right. \\ + & \left.\left(\delta_{ij}^{\perp}\left(\frac{2}{3}\,\frac{\pi^3}{L^4}\,\sum_{n=1}^{\infty} n^3\right) + \delta_{ij}^{||}\left(\frac{2}{3}\,\frac{\pi^3}{L^4}\,\sum_{n=1}^{\infty} n^3\right)\right)\cos\left(\frac{2n\pi z}{L}\right)\right\} \\ = & \left(\frac{2}{3}\,\frac{\pi^3}{L^4}\right)\left\{\left(\delta_{ij}^{\perp}\frac{1}{120} - \delta_{ij}^{||}\frac{1}{120}\right) + \left(\delta_{ij}^{\perp} f(\xi) + \delta_{ij}^{||} f(\xi)\right)\right\}\end{aligned} \tag{D.89}$$

$$\begin{aligned}\langle 0|B_iB_j|0\rangle = & \left\{\left(\delta_{ij}^{\perp}\left(\frac{2}{3}\,\frac{\pi^3}{L^4}\,\sum_{n=1}^{\infty} n^3\right) - \delta_{ij}^{||}\left(\frac{2}{3}\,\frac{\pi^3}{L^4}\,\sum_{n=1}^{\infty} n^3\right)\right)\right. \\ + & \left.\left(-\delta_{ij}^{\perp}\left(\frac{2}{3}\,\frac{\pi^3}{L^4}\,\sum_{n=1}^{\infty} n^3\right) - \delta_{ij}^{||}\left(\frac{2}{3}\,\frac{\pi^3}{L^4}\,\sum_{n=1}^{\infty} n^3\right)\right)\cos\left(\frac{2n\pi z}{L}\right)\right\} \\ = & \left(\frac{2}{3}\,\frac{\pi^3}{L^4}\right)\left\{\left(\delta_{ij}^{\perp}\frac{1}{120} - \delta_{ij}^{||}\frac{1}{120}\right) + \left(-\delta_{ij}^{\perp} f(\xi) - \delta_{ij}^{||} f(\xi)\right)\right\}\end{aligned} \tag{D.90}$$

Also, by the definitions in Eq. (D.50) and Eq. (D.56):

$$\delta_{ij}^{||} + \delta_{ij}^{\perp} \quad \equiv \quad \delta_{i1}\delta_{j1} + \delta_{i2}\delta_{j2} + \delta_{i3}\delta_{j3} \tag{D.91}$$

This is equal to 1 provided that $i = j$ and to 0 otherwise, so it is δ_{ij}.

$$\langle 0|E_iE_j|0\rangle = \left(\frac{\pi}{L}\right)^4 \frac{2}{3\pi}\left\{\left(\delta_{ij}^{\perp} - \delta_{ij}^{||}\right)\frac{1}{120} + \delta_{ij} f(\xi)\right\} \tag{D.92}$$

$$\langle 0|B_iB_j|0\rangle = \left(\frac{\pi}{L}\right)^4 \frac{2}{3\pi} \left\{ \left(\delta^{\perp}_{ij} - \delta^{||}_{ij} \right) \frac{1}{120} - \delta_{ij} f(\xi) \right\} \tag{D.93}$$

Appendix E

DERIVATION OF THE FUNDAMENTAL SOLUTION GREEN'S FUNCTION

Here we derive the expression for the textbook fundamental Green's function $f(z,z')$, solution of the differential equation:

$$\left(-\partial_z^2 + K^2\right) f(z,z') = \delta(z-z'). \tag{E.1}$$

The derivative is with respect to z, so having an explicit expression for the z-dependance would be nice. We can get one by using a trick, i.e. by taking a Fourier Transform of $f(z,z')$, like so:

$$f(z,z') = \int \frac{dk}{2\pi}\, e^{ikz}\, \tilde{f}(k,z') \tag{E.2}$$

Substituting this expression for f(z, z') into our differential equation:

$$\left(-\partial_z^2 + K^2\right) \int \frac{dk}{2\pi}\, e^{ikz}\, \tilde{f}(k,z') = \delta(z-z') \tag{E.3}$$

We note with satisfaction that the z-dependance of the integrand is now entirely contained in e^{ikz}.

Nothing forbids us from swapping the order in which $\left(-\partial_z^2 + K^2\right)$ and $\int \frac{dk}{2\pi}$ act so:

$$\int \frac{dk}{2\pi}\, \left(-\partial_z^2 + K^2\right)\, e^{ikz}\, \tilde{f}(k,z') = \delta(z-z'), \tag{E.4}$$

and we can now perform the derivations with respect to z:

$$\int \frac{dk}{2\pi}\, \left(-(ik)^2 + K^2\right)\, e^{ikz}\, \tilde{f}(k,z') = \delta(z-z') \tag{E.5}$$

i.e.:

$$\int \frac{dk}{2\pi}\, \left(k^2 + K^2\right)\, e^{ikz}\, \tilde{f}(k,z') = \delta(z-z') \tag{E.6}$$

At this point we use another a trick, this time to deal with the RHS of the equation. We realize that we can put it in a form that somewhat resembles the LHS if we recall that a delta function can be expressed as an integral with respect to an arbitrary, dummy variable. We choose to call it k, as indeed from a formal standpoint the variable k on the LHS of our equation is also a dummy variable. Thus we take:

$$\delta(z - z') = \int \frac{dk}{2\pi} \, e^{ik(z-z')} \tag{E.7}$$

and substitute this in our previous result eq.(E.6):

$$\int \frac{dk}{2\pi} \, \left(k^2 + K^2\right) \, e^{ikz} \, \tilde{f}(k, z') = \int \frac{dk}{2\pi} \, e^{ik(z-z')} \tag{E.8}$$

In order for this equality to hold generally, whatever our dummy variable k may be, the integrands themselves must be equal:

$$\left(k^2 + K^2\right) \, e^{ikz} \, \tilde{f}(k, z') = e^{ik(z-z')} \tag{E.9}$$

i.e.:

$$\left(k^2 + K^2\right) \, \tilde{f}(k, z') = e^{-ikz'} \tag{E.10}$$

We now have obtained an explicit expression for $\tilde{f}(k, z')$:

$$\tilde{f}(k, z') = \frac{e^{-ikz'}}{(k^2 + K^2)} \tag{E.11}$$

What we want however is an explicit expression for $f(k, z')$. But we already have an equation relating $f(k, z')$ to $\tilde{f}(k, z')$ — none other than our Fourier Transform of $f(k, z')$, eq.(E.2). So we get:

$$f(z, z') = \int \frac{dk}{2\pi} \, e^{ikz} \, \frac{e^{-ikz'}}{(k^2 + K^2)} \tag{E.12}$$

i.e.:

$$f(z, z') = \int \frac{dk}{2\pi} \, \frac{e^{-ik(z'-z)}}{(k^2 + K^2)} \tag{E.13}$$

This integral can be solved by the method of contour integration, taking the integration constant k to be complex for this purpose. The value of the integral along a closed contour in the complex plane, i.e. parameterized by k, is equal to the value of the integral along the real axis plus its value along a line in the upper (or lower) part of the complex plane:

$$\oint = \int_{\text{along real axis}} + \int_{\text{along line in upper/lower half-plane}} \tag{E.14}$$

The integral we actually want, i.e. Eq. (E.13), corresponds to the integral along the real axis:

$$\int_{\text{along real axis}} = \oint - \int_{\text{along line in upper/lower half-plane}} \tag{E.15}$$

If we are lucky enough for the integrand to vanish as we take the limit of the imaginary part of k to either $+i\infty$ or $-i\infty$ (as indeed is the case here) then we can take advantage of such blessing by choosing our contour of integration wisely. The last integral in eq.(E.15) is then zero and the result we want (the integral along the real axis) is simply equal to the contour integral in the complex plane.

In order to evaluate said contour integral, one uses the "residue theorem", which states that when integrating clockwise along a closed contour, the contour integral is:

$$\oint = 2\pi i \times \text{the sum of the residues inside the contour,} \tag{E.16}$$

where each residue is given by:

$$R(k_0) = (k - k_0).I(k)\Big|_{k=k_0} \tag{E.17}$$

with k_0 a pole,[1] and $I(k)$ the integrand. When integrating counterclockwise contour eq.(E.16) picks up a minus sign on the RHS.

---

[1] values of k for which the integrand would "blow up" as its denominator would be 0.

Our integral of interest has two poles, at $k = +iK$ and $k = -iK$ — i.e. on the imaginary axis.

Now we need to choose our contour of integration wisely in order to simplify eq.(E.15). In order to do so, we need to distinguish two situations, which correspond to two choices for the integration contour.

Integration in the upper-half plane:

When $z' - z < 0$, the line integral in the upper half-plane is the one that vanishes when the imaginary part of k is taken to infinity. Indeed, with $z' - z < 0$, for the exponent of $e^{-ik(z'-z)}$ to tend to $-\infty$ we need $-ik \to +\infty$, hence $k \to +i\infty$, which corresponds to k in the upper plane.

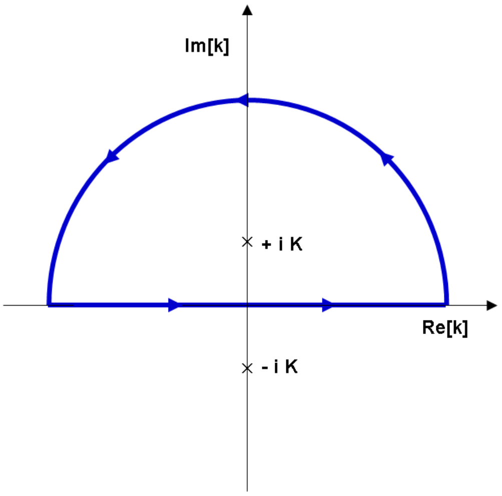


Fig. E.1: Integration contour in the upper half-plane

Therefore in this first case, we choose to integrate along a closed contour that runs, clockwise, from $-\infty$ to $+\infty$ along k's real axis, and to infinity in the upper, positive region of the complex plane.

There is one pole enclosed by this contour: the positive one, $k_0 = +iK$. From eq.(E.17) its residue is:

$$\begin{aligned} R(+iK) &= (k-iK).\frac{1}{2\pi}\frac{e^{-ik(z'-z)}}{(k^2+K^2)}\bigg|_{k=+iK} \\ &= (k-iK).\frac{1}{2\pi}\frac{e^{-ik(z'-z)}}{(k+iK)(k-iK)}\bigg|_{k=+iK} \\ &= \frac{1}{2\pi}\frac{e^{-ik(z'-z)}}{(k+iK)}\bigg|_{k=+iK} \\ &= \frac{1}{2\pi}\frac{e^{K(z'-z)}}{2iK} \end{aligned} \tag{E.18}$$

So from eq.(E.16):

$$\begin{aligned} \oint &= 2\pi i \sum \text{residues inside the closed contour} \\ &= 2\pi i R\,(+iK) \\ &= 2\pi i \frac{1}{2\pi}\frac{e^{K(z'-z)}}{2iK} \\ &= \frac{e^{K(z'-z)}}{2K} \end{aligned} \tag{E.19}$$

Integration in the lower-half plane:

If instead $z' - z > 0$, the line integral that vanishes as we take the imaginary part of k to (minus) infinity runs in the lower half-plane. Indeed, if $z' - z > 0$, for the exponent of $e^{-ik(z'-z)}$ to tend to $-\infty$ what is now required is $-ik \to -\infty$, hence $k \to -i\infty$, so that k needs to be in the lower plane.

Therefore in this situation we need to integrate along a closed contour that runs from $-\infty$ to $+\infty$ along k's real axis, and to infinity in the negative region of the complex plane — counterclockwise this time.

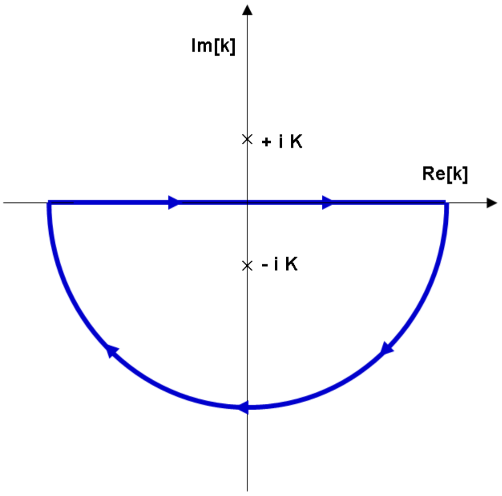


Fig. E.2: Integration contour in the lower half-plane

In this case, it is the other pole that is enclosed: the negative one, $k_0 = -iK$. From Eq. (E.17) its residue is:

$$\begin{aligned} R(-iK) &= (k+iK).\frac{1}{2\pi}\frac{e^{-ik(z'-z)}}{(k^2+K^2)}\bigg|_{k=-iK} \\ &= (k+iK).\frac{1}{2\pi}\frac{e^{-ik(z'-z)}}{(k+iK)(k-iK)}\bigg|_{k=-iK} \\ &= \frac{1}{2\pi}\frac{e^{-ik(z'-z)}}{(k-iK)}\bigg|_{k=-iK} \\ &= -\frac{1}{2\pi}\frac{e^{-K(z'-z)}}{2iK} \end{aligned} \tag{E.20}$$

So from eq.(E.16), with a minus sign because the integration contour now runs counterclockwise:

$$\begin{aligned}\oint &= -2\pi i \sum \text{residues inside the closed contour} \\ &= -2\pi i R\,(-iK) \\ &= -2\pi i\,(-)\frac{1}{2\pi}\frac{e^{-K(z'-z)}}{2iK} \\ &= \frac{e^{-K(z'-z)}}{2K} \qquad \text{(E.21)}\end{aligned}$$

Overall result for the integral:

Putting these two results together we have:

- when $z' - z < 0$, i.e. $\theta(z - z') = 1$:

$$f(z,z') = \int \frac{dk}{2\pi}\,\frac{e^{-ik(z'-z)}}{(k^2+K^2)} = \frac{e^{K(z'-z)}}{2K} \qquad \text{(E.22)}$$

- when $z' - z > 0$, i.e. $\theta(z' - z) = 1$:

$$f(z,z') = \int \frac{dk}{2\pi}\,\frac{e^{-ik(z'-z)}}{(k^2+K^2)} = \frac{e^{-K(z'-z)}}{2K} \qquad \text{(E.23)}$$

So:

$$\begin{aligned}f(z,z') &= \theta(z-z')\,\frac{e^{-K(z-z')}}{2K} + \theta(z'-z)\,\frac{e^{-K(z'-z)}}{2K} \\ &= \frac{e^{-K|z-z'|}}{2K} \qquad \text{(E.24)}\end{aligned}$$

Q.E.D.