

Traditions and Transformations in the History of Quantum Physics

Max Planck Research Library for the History and Development of Knowledge

Series Editors

Jürgen Renn, Robert Schlögl, Bernard F. Schutz.

Edition Open Access Development Team

Lindy Divarci, Jörg Kantel, Nina Ruge, Matthias Schemmel, Kai Surendorf.

Scientific Board

Markus Antonietti, Ian Baldwin, Antonio Becchi, Fabio Bevilacqua, William G. Boltz, Jens Braarvik, Horst Bredekamp, Jed Z. Buchwald, Olivier Darrigol, Thomas Duve, Mike Edmunds, Yehuda Elkana[†], Fynn Ole Engler, Robert K. Englund, Mordechai Feingold, Rivka Feldhay, Gideon Freudenthal, Paolo Galluzzi, Kostas Gavroglu, Mark Geller, Domenico Giulini, Günther Görz, Gerd Graßhoff, James Hough, Manfred Laubichler, Glenn Most, Klaus Müllen, Pier Daniele Napolitani, Alessandro Nova, Hermann Parzinger, Dan Potts, Sabine Schmidtke, Circe Silva da Silva, Ana Simões, Dieter Stein, Richard Stephenson, Mark Stitt, Noel M. Swerdlow, Liba Taub, Martin Vingron, Scott Walter, Norton Wise, Gerhard Wolf, Rüdiger Wolfrum, Gereon Wolters, Zhang Baichun.

Proceedings 5

**Edition Open Access
2017**

Traditions and Transformations in the History of Quantum Physics

**HQ-3: Third International Conference on the
History of Quantum Physics, Berlin,
June 28 – July 2, 2010**

Shaul Katzir, Christoph Lehner,
Jürgen Renn (eds.)

**Edition Open Access
2017**

Max Planck Research Library for the History and Development of Knowledge
Proceedings 5

Communicated by:
Olivier Darrigol

Edited by:
Shaul Katzir, Christoph Lehner, Jürgen Renn

Editorial Coordination:
Nina Ruge

Copyedited by:
Heidi Henrickson, Oksana Kuruts, Jonathan Ludwig, Marius Schneider

Cover image:
From Robert W. Wood (1923), *Physical Optics*, New York: Macmillan

ISBN 978-3-945561-22-5

Published 2017 by Edition Open Access,
Max Planck Institute for the History of Science

Reprint of the 2013 edition

Printed and distributed by

PRO BUSINESS digital printing Deutschland GmbH, Berlin

Edition Open Access

<http://www.edition-open-access.de>

Published under Creative Commons by-nc-sa 3.0 Germany Licence

<http://creativecommons.org/licenses/by-nc-sa/3.0/de/>

The Deutsche Nationalbibliothek lists this publication in the Deutsche
Nationalbibliografie; detailed bibliographic data are available in the Internet at
<http://dnb.d-nb.de>.

The Max Planck Research Library for the History and Development of Knowledge comprises four subseries, *Studies*, *Proceedings*, *Sources* and *Textbooks*. They present research results and the relevant sources in a new format, combining the advantages of traditional publications and the digital medium. The volumes are available both as printed books and as online open access publications. They present original scientific work submitted under the scholarly responsibility of members of the Scientific Board and their academic peers.

The volumes of the four subseries and their electronic counterparts are directed at scholars and students of various disciplines, as well as at a broader public interested in how science shapes our world. They provide rapid access to knowledge at low cost. Moreover, by combining print with digital publication, the four series offer a new way of publishing research in flux and of studying historical topics or current issues in relation to primary materials that are otherwise not easily available.

The initiative is supported, for the time being, by research departments of three Max Planck Institutes, the MPI for the History of Science, the Fritz Haber Institute of the MPG, and the MPI for Gravitational Physics (Albert Einstein Institute). This is in line with the *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, launched by the Max Planck Society in 2003.

Each volume of the *Studies* series is dedicated to a key subject in the history and development of knowledge, bringing together perspectives from different fields and combining source-based empirical research with theoretically guided approaches. The studies are typically working group volumes presenting integrative approaches to problems ranging from the globalization of knowledge to the nature of spatial thinking.

Each volume of the *Proceedings* series presents the results of a scientific meeting on current issues and supports, at the same time, further cooperation on these issues by offering an electronic platform with further resources and the possibility for comments and interactions.

Each volume of the *Sources* series typically presents a primary source—relevant for the history and development of knowledge—in facsimile, transcription, or translation. The original sources are complemented by an introduction and by commentaries reflecting original scholarly work. The sources reproduced in this series may be rare books, manuscripts, documents or data that are not readily accessible in libraries and archives.

Each volume of the *Textbooks* series presents concise and synthetic information on a wide range of current research topics, both introductory and advanced. They use the new publication channel to offer students affordable access to high-level scientific and scholarly overviews. The textbooks are prepared and updated by experts in the relevant fields and supplemented by additional online materials.

On the basis of scholarly expertise the publication of the four series brings together traditional books produced by print-on-demand techniques with modern information technology. Based on and extending the functionalities of the existing open access repository European Cultural Heritage Online (ECHO), this initiative aims at a model for an unprecedented, Web-based scientific working environment integrating access to information with interactive features.

Contents

Contributors	1
Introduction	3
From Classical to Quantum Physics	9
1 Theoretical Challenges by Experimental Physics: Radiation and Its Interaction with Matter	
<i>Shaul Katzir</i>	11
1.1 The Nature of X- and γ -Rays	14
1.2 Photoelectricity and the Triggering Hypothesis	16
1.3 Visible Light and X-rays	20
1.4 Later Theoretical Accounts	22
1.5 Conclusion	24
References	25
2 Challenging the Boundaries between Classical and Quantum Physics: The Case of Optical Dispersion	
<i>Marta Jordi Taltavull</i>	29
2.1 The Classical Theory of Optical Dispersion	30
2.2 Bohr's Model of the Atom and the Optical Dispersion of Molecules	34
2.3 Debye and the First Theory of Optical Dispersion According to Bohr's Model	36
2.4 Sommerfeld and Optical Dispersion in 1915	37
2.5 Reactions to Debye-Sommerfeld's Hybrid Theory	40
2.6 Sommerfeld's Counterattack	46
2.7 From the Hybrid Theory to the Light Quantum	48
2.8 A New Thread of the Story	51
2.9 Conclusion	53
Abbreviations and Archives	54
Acknowledgements	54
References	55

3	Putting the Quantum to Work: Otto Sackur's Pioneering Exploits in the Quantum Theory of Gases	
	<i>Massimiliano Badino and Bretislav Friedrich</i>	61
3.1	Biographical Overview	63
3.2	The Problem of Chemical Equilibrium	65
3.3	The Beginnings of the Quantum Theory of Gases	69
3.4	Sackur's "Elementary Regions"	73
3.5	The Generalization of Sackur's Statistical Theory	75
3.6	Conclusion	80
	References	81
	Quantum Mechanics in the Making	87
4	The Concepts of Light Atoms and Light Molecules and Their Final Interpretation	
	<i>Dieter Fick and Horst Kant</i>	89
4.1	First Corpuscular Concepts of Light	90
4.2	Light Molecules, Static Concepts	95
4.3	Dynamical Treatments of Light Multiples	98
4.4	Obituary	112
	Appendix	115
	Abbreviations and Archives	119
	Acknowledgements	119
	References	120
5	Early Interactions of Quantum Statistics and Quantum Mechanics	
	<i>Daniela Monaldi</i>	125
5.1	The Birth of Quantum Statistics	126
5.2	Heisenberg's Many-body Problem and Quantum Resonance	129
5.3	Dirac's "On the Theory of Quantum Mechanics"	137
5.4	Conclusion	144
	Abbreviations and Archives	145
	References	145
6	Pourparlers for Amalgamation: Some Early Sources of Quantum Gravity Research	
	<i>Dean Rickles</i>	149
6.1	The Torch of Unification: Mie, Hilbert, Weyl and Haas	152
6.2	Einstein on the Relationship between Gravity and Quantum	156
6.3	Quantum Meets Gravity in the Pages of <i>Nature</i>	159

6.4	On the Way to Quantum Geometry	172
6.5	Conclusion	174
	Abbreviations and Archives	175
	Acknowledgements	175
	References	175
	Extending the Framework of Quantum Physics	181
7	Superposing Dynamos and Electrons: Electrical Engineering and Quantum Physics in the Case of Nishina Yoshio	
	<i>Kenji Ito</i>	183
7.1	Engineering in Japan	185
7.2	Nishina Yoshio as an Engineering Student	187
7.3	The Klein-Nishina Formula	196
7.4	Superposition of States in the Klein-Nishina Paper	201
7.5	Conclusion	205
	Acknowledgements	205
	References	206
8	The Origins of Maria Göppert's Dissertation on Two-Photon Quantum Transitions at Göttingen's Institutes of Physics 1920-1933	
	<i>Barry R. Masters</i>	209
8.1	Physics in Germany at the Beginning of the Twentieth Century and the Development of the Institutes of Physics at the University of Göttingen	210
8.2	Göppert as a University and Doctoral Student	212
8.3	What Was the Role of Paul Dirac in Göppert's Dissertation Research?	215
8.4	Reconstruction of Göppert's Göttingen Dissertation	218
8.5	What Was Known and What Did Göppert Contribute in Her Dissertation Research?	222
8.6	Conclusion	225
	Acknowledgements	226
	References	227
9	An Act of Creation: The Meitner-Frisch Interpretation of Nuclear Fission	
	<i>Roger H. Stuewer</i>	231
9.1	The Birth of the Liquid-Drop Model	232

9.2	The Extension of the Liquid-Drop Model: Heisenberg and von Weizsäcker	235
9.3	Bohr's Theory of the Compound Nucleus	237
9.4	The Interpretation of Fission	239
9.5	Aftermath	240
9.6	Conclusion	241
	Abbreviations and Archives	242
	References	243
The Challenges of Quantum Field Theory		247
10	Tsung-Sui Chang's Contribution to the Quantization of Constrained Hamiltonian Systems	
	<i>Xiaodong Yin, Zhongyuan Zhu, and Donald C. Salisbury</i>	249
10.1	Biographical Overview	249
10.2	Studies on Constrained Hamiltonian Systems before Chang's Work	251
10.3	Chang's Contributions to Hamiltonian Systems	254
10.4	Conclusion	264
	Abbreviations and Archives	265
	Acknowledgements	265
	List of Chang's Publications	265
	References	268
11	Feynman's Struggle and Dyson's Surprise: The Development and Early Application of a New Means of Representation	
	<i>Adrian Wüthrich</i>	271
11.1	Development, Modern Application and Interpretation of Feynman Diagrams	272
11.2	Feynman's Struggle for a Physical Interpretation of the Dirac Equation	276
11.3	<i>Zitterbewegung</i>	277
11.4	Positrons and Interaction	279
11.5	Abandoning the Search for a Microscopic Analysis	281
11.6	Solution to the Problems through an Appropriate Representation of the Phenomena	282
11.7	Comparison to Other Developments of Concepts and Means of Representation	285
	Abbreviations and Archives	287
	Acknowledgements	287
	References	287

Traditions and Debates in Recent Quantum Physics	291
12 Orthodoxies on the Interpretation of Quantum Theory: The Case of the Consistent History Approach	
<i>Olival Freire</i>	293
12.1 Orthodoxies and Heterodoxies in Quantum Physics	293
12.2 The Polysemic New Orthodoxy	296
12.3 The Consistent History Approach and Its Rhetorical Resources	298
12.4 The Reception of the Consistent History Approach	302
12.5 Conclusion	304
Abbreviations and Archives	305
Acknowledgements	305
References	305
13 From Do-it-yourself Quantum Mechanics to Nanotechnology? The History of Experimental Semiconductor Physics, 1970–2000	
<i>Christian Kehrt</i>	309
13.1 New Research Technologies at the Quantum Level	311
13.2 The 1970s: A New Quantum Generation	315
13.3 Von Klitzing’s Nobel Prize: The Discovery of the Quantum Hall Effect	318
13.4 The 1980s: Founding the Walter Schottky Institute	322
13.5 Munich Nanoscience Networks	323
13.6 Nano-biotechnologies. New Forms of Interdisciplinary Cooperation?	327
13.7 Conclusion: The Reinvention of Semiconductor Physics	333
References	333

Contributors

Massimiliano Badino: Centre d'Història de la Ciència, Facultat de Ciències, Universitat Autònoma de Barcelona, Cerdanyola del Valles, 08193 Bellaterra (Barcelona), Spain
massimiliano.badino@uab.cat

Dieter Fick: Physics Department, Philipps Universität Marburg, Renthof 5, 35032 Marburg, Germany, Fritz Haber Institute of the Max Planck Society, Faradayweg 4–6, 14195 Berlin, Germany, and Max Planck Institute for the History of Science, Boltzmannstraße 22, 14195 Berlin, Germany
fick@staff.uni-marburg.de

Olival Freire Jr: Programa de Pós-Graduação em Ensino, Filosofia e História das Ciências (UFBa-UEFS), Instituto de Física, Universidade Federal da Bahia, Campus de Ondina, Salvador - BA - 40210-340, Brasil
olival.freire@gmail.com

Bretislav Friedrich: Department of Molecular Physics, Fritz Haber Institute of the Max Planck Society, Faradayweg 4–6, 14195 Berlin, Germany
bretislav.friedrich@fhi-berlin.mpg.de

Kenji Ito: The Graduate University for Advanced Studies (Sokendai), Shonan Village, Hayama, Kanagawa, 240-0193, Japan
ito_kenji@soken.ac.jp

Marta Jordi Taltavull: Max Planck Institute for the History of Science, Boltzmannstraße 22, 14195 Berlin, Germany
mjordi@mpiwg-berlin.mpg.de

Horst Kant: Max Planck Institute for the History of Science, Boltzmannstraße 22, 14195 Berlin, Germany
kant@mpiwg-berlin.mpg.de

Shaul Katzir: The Cohn Institute for the History and Philosophy of Science and Ideas, Faculty of Humanities, Tel Aviv University, Tel Aviv 69978, Israel
skatzir@tau.ac.il

Christian Kehrt: Fakultät für Geistes- und Sozialwissenschaften, Neuere Sozial-, Wirtschafts- und Technikgeschichte, Helmut Schmidt University, Holstenhofweg 85, 22043 Hamburg, Germany
kehrt@hsu-hh.de

Barry R. Masters: Department of the History of Science, Harvard University, Science Center, No. 1 Oxford Street, Cambridge, MA 02138 USA, and Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, NE47-220, Cambridge, MA 01239 USA
brmail2001@yahoo.com, bmasters@mit.edu

Daniela Monaldi: 133 Rumsey Road, Toronto, ON M4G 1P3, Canada
dmonaldi@yorku.ca

Dean Rickles: Unit for History and Philosophy of Science, The University of Sydney, NSW 2006, Australia
dean.rickles@sydney.edu.au

Donald C. Salisbury: Program in Global Science, Technology, and Society, Physics Department, Austin College, 900 North Grand Avenue, Suite 61621, Sherman, Texas 75090, USA
dsalisbury@austincollege.edu

Roger H. Stuewer: Program in History of Science and Technology, University of Minnesota, Minneapolis, MN 55455, USA
rstuewer@physics.umn.edu

Adrian Wüthrich: History and Philosophy of Science, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland, and Institut für Philosophie, Literatur-, Wissenschafts- und Technikgeschichte, Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany
awuethr@philo.unibe.ch

Xiaodong Yin: Physics Department of Capital Normal University, Beijing Xi San Huan North Road 105, Haidian District, Beijing 100048, P. R. China
xiaodongstone@139.com

Zhongyuan Zhu: Institute of Theoretical Physics, Chinese Academy of Sciences, Zhong Guan Cun East Street 55, Beijing 100190, P. R. China
zzy@itp.ac.cn

Introduction

The present volume presents results of recent research into the history of quantum physics. The thirteen papers included here show the multifaceted nature of this research. They discuss developments from the late nineteenth to the early twenty first century and go beyond the traditional focus on Europe and North America to include China and Japan. Also a wider array of subdisciplines comes into view, from optics to quantum gravity through quantum electrodynamics, from atomic and nuclear to condensed matter physics and the foundations of physics. The perspective of the papers ranges from local histories to global discussions, from conceptual changes driven by experimental practices to interactions of the new theoretical physics with social and technological forces.

Several novel aspects of the history of quantum physics emerge in these contributions. Actors who have so far played only a marginal role in the historical account, such as Otto Sackur, Maria Göppert and Chang Tsung Sui,¹ are now recognized for their roles in the development of quantum physics. Similarly, fields such as dispersion theory, physical chemistry and solid state physics receive a more prominent place in the narrative of its development. In this historical perspective, they no longer constitute just areas of applications but are seen as birth-places of important theoretical insights. Developments off the main road of the traditional narrative, such as the pursuit of the idea of light molecules or early explorations of the relations between the quantum and gravity, constitute another focal point of this volume. This collection also makes clear that recent research rightly pays increasing attention to the role of modeling and representation in the formation of quantum theories.

Despite the diversity of the themes treated, one common thread emerges: the importance of continuities in the historical development of quantum physics. The place of long established traditions can be seen, for example, in the role that traditional modes of experimental physics, associated with the nineteenth century, continued to play in developing new theoretical ideas, including those associated with the quantum hypothesis. The case of optical dispersion shows that even after the introduction of Niels Bohr's atomic model physicists continued to suggest theories of dispersion based on late nineteenth century atomistic models, and that the problem of harmonizing these theories with the developing quantum theory

¹In this introduction we use the traditional order for Chinese and Japanese names.

led to central theoretical insights. Further continuity can be discerned in the persistent role of trans-disciplinary fields such as physical chemistry or nanoscience in generating conceptual and methodological innovations, as well as in linking science and technology.

The first section of the volume deals with the transition “From Classical to Quantum Physics.” This transition is often associated with the notion that certain crucial experiments refuted tenets of classical physics and necessitated the introduction of revolutionary new theories. This transition is also often associated with the emergence of theoretical physics as an independent subdiscipline and with a new division of labor between theoreticians and experimentalists. Here it is shown, however, that the connection between empirical knowledge, experiments, and theoretical reasoning was much more complex, characterized by an overlap between classical and quantum ideas and also by a less strict division of labor than has traditionally been assumed.

In earlier periods of physics, experiments were often not tied to well-defined quantitative theoretical claims but were of a merely qualitative nature. In his contribution, Shaul Katzir shows that more qualitative, exploratory experiments of this kind, not directly guided by the intention to systematically check quantitative implications of mathematically formulated theories, did play a crucial role in the early history of quantum physics and possibly beyond.

Marta Jordi studies the crucial role of empirical knowledge, embodied in established theories of classical physics, for the emergence of the new quantum physics. She shows, in particular, how knowledge of the well-developed classical theory of dispersion was represented in the model of co-vibration of matter and light that not only survived the transition to quantum physics but also helped to shape its conceptual foundation.

The early history of quantum physics was marked by a rapid growth of the number of phenomena to which some form of quantum hypothesis was applied. Ever expanding domains of radiation and thermal physics were touched by quantum theory. How did this expansion of the quantum happen? It did not, at least initially, take place as part of a systematic research program. Rather, it occurred because existing and sometimes long-standing problems such as that of chemical equilibrium could be connected to the quantum. Moreover the transfer of the quantum hypothesis to new areas of application was not necessarily achieved by its most famous protagonists, but often by scientists simply looking for new tools to solve such long-standing problems. One such scientist was the physical chemist Otto Sackur whose contribution to the quantum theory of gases is analyzed in the paper by Massimiliano Badino and Bretislav Friedrich. They show that his pragmatic and goal-oriented attempt to address a long-standing problem of physical chemistry—how to calculate the chemical constants defining equi-

libria in chemical reactions—led to a novel and significant use of the quantum hypothesis in the theory of the ideal gas.

As the quantum hypothesis was extended to new domains, it continued to raise conceptual problems, which are the focus of the second section. Albert Einstein's introduction of the light quantum in 1905 remained controversial for at least twenty years. Debates over the nature of radiation accompanied the further development of quantum physics and often suggested innovative ideas such as the idea of considering a wave-particle duality for matter as well, or introducing a new statistics. These debates also involved ideas that, at the time, represented serious candidates for an understanding of the nature of light, but that were later dismissed and even forgotten. One such idea was the idea of light molecules introduced by Mieczysław Wolfke in 1921 and developed further by Walther Bothe, especially in an unpublished manuscript from 1925, analyzed for the first time in the contribution by Dieter Fick and Horst Kant to this volume.

The idea of light molecules was superseded by the introduction of a new statistics to radiation theory by Satyendra Nath Bose and to gases by Einstein. In her contribution, Daniela Monaldi discusses the emergence of the new statistics and its relation to early work on many-particle systems by Werner Heisenberg and Paul Dirac. She shows, in particular, that the revolutionary potential of the new statistics for developing physicist's understanding of the concept of a particle and of a physical system was not realized in their works. Instead, both Heisenberg and Dirac, in spite of the great differences between their works, stuck to a classical understanding of individual particles and their statistical independence.

Today, quantum physics and gravitation theory are two clearly separate domains whose integration is considered to be one of the most challenging conceptual problems of modern physics. In his contribution Dean Rickles shows that the need to unify them was evident to some physicists as early as the 1910s, that is, well before the formulation of either general relativity or quantum mechanics. Thus, many ideas still under discussion currently, such as the existence of more than four dimensions of space and time or the existence of a new physics at the scale of the Planck length, were broached even then. Back then, however, it was still an open question whether such ideas would actually be needed in order to complete the building of quantum theory and of general relativity, or whether this could be achieved without establishing a bridge between them.

The section "Extending the Framework of Quantum Physics" deals with examples of contributions to the expanding field of quantum mechanics after its firm establishment in the mid-1920s. This involved scientists who brought quite diverse intellectual backgrounds to the burgeoning field. The period is explored in three papers that stress the importance of these different disciplinary backgrounds and local traditions.

Ito Kenji analyzes the educational background of Nishina Yoshio, one of the leading Japanese physicists of the first half of the twentieth century. In 1928, together with Oskar Klein, Nishina developed the so-called Klein-Nishina formula describing the scattering of light quanta and electrons based on the relativistic Dirac equation for the electron. Ito makes it clear that the introduction into Japan of Western science and technology in the late nineteenth century enabled Japanese scientists of Nishina's generation to become important contributors to quantum theory, on a par with their Western colleagues. He shows, in particular, that advanced training in electrical engineering could provide an advantageous starting point for such careers.

Barry Masters takes us to one of the centers of quantum physics—Göttingen circa 1930—to explore the context and origin of Maria Göppert's dissertation on atomic transitions involving two photons (as opposed to the simple one-photon processes that Dirac had treated in his groundbreaking paper on the quantum theory of emission and absorption of radiation in 1927). Masters shows how Göppert's work is rooted in Dirac's paper and in the work of Göttingen physicists Otto Oldenberg and James Franck, who studied more complicated interactions between radiation and matter both experimentally and theoretically. The story reminds us of the importance of new applications for the establishment of a theory such as Dirac's and for its extension and corroboration despite severe internal difficulties.

Roger Stuewer describes the two distinct traditions of nuclear physics that merged when Lise Meitner and Otto Frisch puzzled out an interpretation of Fritz Straßmann and Otto Hahn's findings of unexpected elements in the decay of uranium after bombardment with neutrons. Meitner, coming from Berlin, was familiar with the work of Heisenberg and Carl Friedrich von Weizsäcker on Gamow's liquid-drop model of the nucleus. On the other hand, Frisch had been working in Copenhagen on Bohr's theory of the compound nucleus. The combination of the detailed energetic implications of the Berlin model and the dynamical emphasis of the Copenhagen approach resulted in a theoretical discovery of great consequence, nuclear fission.

The section "The Challenges of Quantum Field Theory" explores physicists' struggles to formulate a consistent quantum theory of fields by building on the early successes of Dirac's and Pascual Jordan's quantum electrodynamics. Despite these successes, it had quickly become clear that quantizing the electromagnetic field led to a set of difficulties that threatened to make the procedure meaningless. From the late 1920s until the success of the renormalization program in the late 1940s, theorists were trying to find a firmer theoretical basis for quantum electrodynamics.

The paper by Yin Xiaodong, Zhu Zhongyuan and Donald Salisbury deals with work of the Chinese physicist Chang Tsung Sui on one such theoretical difficulty of quantum electrodynamics. The problem was one encountered by Heisenberg and Wolfgang Pauli in 1928/29 in their first attempt to formulate a general theory of quantum electrodynamics. Quantizing the electromagnetic field using canonical quantization, which derives the commutation equations of a quantum theory from the Hamiltonian formulation of the corresponding classical theory, led to ambiguities. These appear in the quantization of all theories that, like electrodynamics, involve a gauge freedom, i.e. degrees of freedom in the theoretical quantities that do not have a physical meaning. Chang had visited Cambridge twice in the 1930s and 1940s. Inspired by Dirac, he wrote several papers addressing the question of how to quantize such theories. These papers anticipate results of better known works by Dirac and by James Anderson and Peter Bergmann in the 1950s.

Adrian Wüthrich's contribution is concerned with the development of Feynman diagrams, which was closely connected to another foundational difficulty of quantum electrodynamics, the infinite values predicted by the theory for most physical quantities due to its treatment of the interaction between radiation and matter. Wüthrich argues that it was Richard Feynman's attempt to find a physical interpretation of the Dirac equation in terms of the motion of a particle that led him to designing diagrams in terms of the propagation of quanta. While Feynman eventually had to abandon this interpretation, the diagrams remained as a powerful calculational tool. They now represented merely certain expressions in a calculation without presupposing that the particle actually travels along definite trajectories. It was in this sense that Freeman Dyson used Feynman diagrams to show that the infinite expressions in quantum electrodynamics could be redefined such that they only affected non-observable quantities.

In the last section "Traditions and Debates in Recent Quantum Physics," Olival Freire and Christian Kehrt examine very different aspects of recent developments within quantum physics, from its philosophical interpretation to its technological applications. Freire analyzes the view, influential in contemporary debates about the interpretation of quantum mechanics, that the consistent history approach has inherited the Copenhagen interpretation's role in the interpretation debate and has become a "new orthodoxy." Freire shows, on the basis of bibliographical data, that this view does not stand empirical scrutiny.

Based on a case study of a local research network in Munich, Kehrt argues that the emergence of the trans-disciplinary field of nanotechnology opened up new research directions for solid state physicists, in particular through the adaptation of methods from the life sciences, as well as new funding sources, entrepreneurial opportunities, and resources for public presentations. Following

Paul Forman, he sees this development as characteristic of the sciences at the turn to the twenty-first century, in the sense that technology takes primacy over science.

The present volume originated from the Third International Conference on the History of Quantum Physics (HQ-3), which took place in Berlin in summer 2010 and included speakers from five continents. The conference series was launched by the joint project on the history of quantum physics of the Max Planck Institute for the History of Science and the Fritz Haber Institute of the Max Planck Society.

We would like to thank all participants in the conference and reviewers of the papers that appear here for their helpful advice. Special thanks to Nina Ruge and her editing team: Heidi Henrickson, Oksana Kuruts, Jonathan Ludwig and Marius Schneider for ensuring that this book materialized. And thanks to our colleagues Christian Joas, Jeremiah James and Alexander Blum for their help in many ways. Lastly we are happy to acknowledge the Strategic Innovation Fund of the President of the Max Planck Society, which supported the history of quantum physics project.

From Classical to Quantum Physics

Chapter 1

Theoretical Challenges by Experimental Physics: Radiation and Its Interaction with Matter

Shaul Katzir

The development and formulation of quantum physics coincided with the consolidation of theoretical physics as a subdiscipline, the existence of which was the result of a long process whose origins go back at least to the middle of the nineteenth century. Throughout this period, it acquired characteristics, which, even if not original, distinguished it from earlier ways of theorizing about the physical world. Especially as it developed in the German-speaking world, this new theoretical physics emphasized mathematical laws and more significantly a mathematical deduction of quantitative rules from a few assumptions and the connection of these rules to empirical findings. While valuing the empirical examination of these rules, the new theoretical physics reinforced, and sometimes even created, a division of labor within physics between experimentalists and theoreticians. In this aspect, as in many others, this kind of theoretical physics became dominant in the twentieth century across linguistic and national borders.

Historians have pointed to a connection between the development of quantum mechanics on its elaborated mathematical edifice and the rise of the new theoretical physics. For example, in their classic history of the subdiscipline in Germany, Christa Jungnickel and Russell McCormmach write:

Theoretical physics experienced some of its greatest advances, and German theoretical physicists played a significant and often leading part in [the special and general theories of relativity and the early quantum theory]. (Jungnickel and McCormmach 1986, 304)

That theoretical physics played a leading role in the development of quantum physics, however, meant neither that the more traditional “experimental physics” stopped contributing to the process, nor that its participants confined their role to the testing of theories. On the contrary, as I illustrate below, they also added to the theoretical understanding of the microphysical world. The contribution of this kind of experimental physics seems to be especially significant with the de-

velopments that took place before the formulation of matrix and wave mechanics and particularly in new research areas, like photoelectricity and X-rays.

Physicists working in the more traditional style of research designed and performed experiments to explore and later understand various aspects of newly observed effects. For example, in photoelectricity many experimentalists strove to determine the source and nature of the charge carriers, leading eventually to the identification of the latter with electrons. Experimentalists took on the task of clarifying the phenomena related to the new effects and entities and to determine their characteristics (e.g., the kind of irradiated metal, the intensity, direction, polarization and wavelength of the light, and temperature). They suggested a number of hypotheses and explanations to account for the new phenomena. Yet contrary to the new theoretical physics, the kind of theoretical thought associated with experimental physics rarely involved either elaborated mathematical derivations from basic laws or exact quantitative rules for empirical test. There was no systematic theory to describe and organize the observed phenomena. Thus, experiment and theory were closely intertwined, often within the work of the same person, creating situations in which empirical investigation aimed at examining a particular hypothesis, and assumptions followed particular experimental results. Experiments were frequently explorative in character (Steinle 1998), and were used directly to supply new information and to decide between competing claims and hypotheses. Even when more encompassing hypotheses were suggested, experiments directly examined their implications, rather than a mathematical rule deduced from them.

Unlike this situation, in the ideal type of theoretical physics, experiments are confined to test theoretically derived quantitative rules. Black-body radiation provides a good example. Such experts in exact measurement as Heinrich Rubens, Ferdinand Kurlbaum, Otto Lummer and Ernst Pringsheim determined the exact mathematical curve of the wavelength distribution of black-body radiation. Their precise measurements tested the exact form of the mathematical relations but were detached from the physical assumptions made in deriving this quantitative rule. Atomic spectra following Niels Bohr's atom provides another example of measurements as a way to test an elaborate theory as a whole. The atomic and molecular emission lines became meaningful for physical claims only by comparing them with the quantitative results of the theories of Bohr and Arnold Sommerfeld. In its ideal type, an empirical deviation from a theoretical law calls for a change in the assumptions of the theory, a change that could lead to a new rule fitting the experimental results. The new rule should consequently be compared with further, more precise or extended measurements, or to measurements of other related phenomena. Sommerfeld's modification of, and addition to, Bohr's atomic

theory to improve its match with measurements showing fine structure of spectral lines is an example of this kind of study.

Thus, theoretical physics of this kind supposes a clear division between theorization and experimentation. This separation is logical, chronological and often also personal: they are different ways of doing science, they take place at different times and are often performed by different actors. However, in the early twentieth century, both traditional experimental physics and most of the physical research outside the German cultural sphere did not assume such a clear separation. Experiments contributed to theorizing by providing information about various aspects of phenomena, as well as examining the implications of specific hypotheses, and not only of one particular quantitative rule. Moreover, experimentalists suggested many hypotheses, concepts and theories while designing their experiments and reflecting on them. In Germany, the practitioners of this kind of laboratory research were called “experimental physicists,” while their colleagues who worked at precise tests of quantitative rules were coined “measuring physicists.” Measuring physics was the laboratory endeavor associated with theoretical physics (Jungnickel and McCormmach 1986, 120). In the late nineteenth century, exact measurements were the only laboratory research still practiced by German theoretical physicists. With the younger generation of theoreticians developing quantum physics, however, the separation between physicists deriving quantitative expressions and those who empirically tested them became almost complete. In other cultural realms, including the English- and the French-speaking worlds, the division between theoreticians and experimentalists emerged later. So, in the early decades of the twentieth century in these countries, physicists engaged with theory often conducted explorative and more traditional kinds of experiments. In this paper, following the German terminology of the time, this kind of research I call experimental physics.

This article illustrates how physicists working in this more traditional kind of experimental physics developed and examined new theoretical understandings of phenomena relating to ultraviolet radiation, Röntgen (X) rays and γ -rays, and their interactions with matter. Although unexpected, the discoveries of the photoelectric effect (1886) and X-rays (1895) did not seem to threaten the foundations of contemporary physics. Only further exploration of these phenomena at the beginning of the twentieth century by “experimental physicists” revealed contradictions with the common (later called “classical”) laws and concepts of physics. Conclusions, drawn from many experiments in this tradition, and not solely independent results, convinced many scientists of the need to adopt assumptions that contradicted classical physics. Moreover, new notions that emerged in this kind of research evolved into later concepts of quantum physics. In retrospects,

“quantum hypotheses” that assume discrete distribution of energy were the most important among these.

Theoretical arguments did play a crucial role in other important strands that led to quantum theory. Most famously, Max Planck needed to introduce the quantum hypothesis as a theoretical procedure in statistical thermodynamics (regardless of his own earlier “classical” interpretation of it) (Darrigol 1992). While exact measurements showed the need to change the law of black-body radiation, only a theoretical derivation of the law revealed the need to introduce the quantum hypothesis. Moreover, in 1905, Albert Einstein produced his more radical hypothesis of light quanta based on an argument that revealed a conflict between the theories of electrodynamics and thermodynamics, without evoking any novel empirical findings (Büttner, Renn, and Schemmel 2003). This kind of theoretical scrutiny, however, could not lead to similar results regarding the interaction of radiation with matter. Rather, experimental research discerned limitations of and contradictions with a conventional (classical) explanation of ultraviolet radiation, and X- and γ -rays, and their interaction with matter. Around 1911–1913, the combination of evidence from these phenomena convinced most physicists to admit discontinuity in radiation phenomena, usually in the form of a quantum hypothesis, which they connected to theoretical findings about black-body radiation.

1.1 The Nature of X- and γ -Rays

Wilhelm Röntgen’s discovery of X-rays triggered an extraordinary wave of experiments about their main properties and raised almost as many speculations about their nature. Most of these experiments studied the interaction of these rays with matter and electromagnetic radiation in the tradition of experimental physics, experiments that were mostly quantitative. Usually these experiments did not require the high precision needed in black-body measurements. Rough estimations and even an order of magnitude were often sufficient, at least in the early stages. Röntgen himself was a representative of experimental physics, and his famous discovery, like most of his research, belonged to that tradition (Katzir 2006b, 44). Within a few months, the experiments carried out in that tradition led to the general view that X-rays were some kind of electromagnetic waves. Experimentalists showed that the rays could penetrate matter (which was not characteristic of any particle), were not deflected by strong electric and magnetic fields (thus, the rays were not charged), and produced photographs on chemical paper (as did infrared, visible and ultraviolet light). Röntgen himself supposed that the rays were longitudinal or condensational ether waves, similar to acoustical waves in air, a kind of wave long sought for in continental electromagnetism, often suggested

in the experimental research on cathode rays, and thus a natural assumption for Röntgen (Darrigol 2000, 284–287). James Maxwell’s theory, however, did not allow for longitudinal waves, therefore many physicists leaned toward the view that X-rays were ordinary light, i.e., transverse waves, of very high frequency. The electromagnetic assumptions led to attempts to polarize (as only transverse waves do) and diffract (as all waves do) the rays. Yet, early failures in producing polarization and diffraction did not refute the wave assumption, since such an assumption did not provide a prediction about the magnitude of the effects. In this early phase, the non-effect constrained the assumption, i.e. the wavelength of the X-rays, but did not lead to its rejection. The rays, however, also failed to show other features of regular waves like reflection and refraction. The impulse hypothesis, according to which X-rays were transverse electromagnetic waves of very short duration, offered a way out of the failure of X-rays to present these expected features.¹

Cambridge mathematical physicist George Gabriel Stokes proposed the impulse hypothesis in 1896 as a theoretical idea within the classical electromagnetic wave theory. Still, its justification originated in qualitative findings about X-rays. It allowed for the combination of spatial continuity, characteristic of waves and fitting such an observation as neutral electricity and strong penetrating power, with the temporal discontinuity usually associated with particles, as suggested by the rays’ strength and short duration. Stokes and Joseph J. Thomson suggested a mathematical theory of the impulses and their interactions. Later in 1899, German theoretician Sommerfeld showed that the hypothesis was flexible enough to answer another direct challenge by experiment, namely X-ray diffraction (by broadening the image, without showing the interference pattern), which had been observed a few months earlier by Cornelis Wind and Hermanus Haga in Groningen. “In one form or another, the impulse hypothesis of X-rays achieved supremacy in the period 1903–1905.” (Wheaton 1983, 15–67, quotation on p. 48). Experiments on γ -rays, recently differentiated from α and β radioactive rays, pointed to a close similarity between them and X-rays. Consequently, most physicists regarded γ -rays as a stronger kind of electromagnetic impulse.

However, because it was a kind of wave assumption, the impulse hypothesis led to two major problems, which historian Bruce Wheaton in his important book on the experimental origins of particle-wave duality termed the paradoxes of quantity and quality. Since the impulse, as a kind of wave, propagates spherically, it should have ionized all atoms on its way. Yet, already in 1896 and 1897, Thom-

¹ The impulse hypothesis fitted well the corpuscular view of cathode rays: absorption of one corpuscle (electron) leads to the emission of one short electromagnetic impulse—an X-ray. Adoption of the view that cathode rays are a kind of localized concentration of electric charge, or electrons, thus encouraged the acceptance of the impulse hypothesis.

son and independently Ernst Dorn had found that only a tiny fraction of the atoms were ionized. This is the paradox of quantity. On the other hand, further experiments, especially on “secondary electrons” emitted by the absorption of X-rays, showed that X-rays communicated energy to individual electrons—of a similar magnitude to what was needed for their own generation, without any apparent weakening with distance.² How a spherical propagation, like an impulse, could deliver its whole energy to a localized place in space was the paradox of quality (Wheaton 1983, 71–79). Notice that these experiments were not carried out to examine a particular mathematical prediction but to explore the behavior and nature of these rays, mainly independently of the impulse hypothesis. Although quantitative, these experiments did not require high precision (i.e., an order of magnitude was more than enough for the quantity paradox). This kind of semi qualitative results characterized experimental rather than measuring physics.

Faced with the paradoxes of quantity and quality, William Henry Bragg suggested in 1907 that X- and γ -rays were neutral pairs of an electron and a positive charge, rather than an impulse. Working alone in Australia, Bragg was a dissident. European physicists continued to think of the rays as wave impulses, or even as simple waves. Bragg’s suggestion was far from a formal theory but led him and Charles Barkla in Liverpool to do further experimental work. Barkla, however, wanted to challenge Bragg’s conclusions. Bragg and Barkla designed their experiments to check direct qualitative consequences of the wave and the corpuscular hypotheses, without the mediation of an elaborate theory. Barkla supported the wave point of view by obtaining partial polarization of X-rays (a property of transverse waves but not of particles), as well as showing the homogeneity of secondary X-rays. In his experiments on strong X-rays (i.e. X-rays of high penetrating power) and γ -rays, Bragg, on the other hand, demonstrated their particle-like properties. For example, he showed that the direction of the emitted secondary electrons depended on the direction of the γ -rays that induced them. Their experiments and the controversy between the two physicists showed the difficulties of explaining the behavior of these rays either as particles or as waves (Wheaton 1983, 81–103; Stuewer 1971b).

1.2 Photoelectricity and the Triggering Hypothesis

Photoelectricity originated in Heinrich Hertz’s experimental research program that led him to the production and detection of electromagnetic waves. In the winter of 1886–1887, he discovered that ultraviolet light enhances electric spark-

²Since energy calculation was based on previous knowledge of electron’s mass and energy, these experiments exemplify the crucial role played by the view that there is only one kind of electron and the determination of its mass and charge.

discharge. In 1888, following Hertz, Wilhelm Hallwachs observed electric discharge from metal plates exposed to ultraviolet light in a vacuum tube. This effect became the paradigmatic example for photoelectricity. Experimentalists throughout Europe explored several aspects of this new phenomenon, examining the metals, charge carriers, gases, tubes and light that interact in these effects. Still, for more than twenty years it failed to attract much attention from theoretical physicists but remained the domain of experimental physics (Stuewer 1971a; Buchwald 1994, 243–244; Hallwachs 1916; von Schweidler 1898, 883–894).

In 1902, Phillip Lenard examined the relationship between the intensity of the incident light and the energy of the electric discharge. Exploiting recent experiments of his own and those carried out by Egon von Schweidler and Thomson that identified the carriers of the photoelectric current with electrons, Lenard measured the energy of the emitted electrons. Surprisingly, he found that their energy was independent from the intensity of light, i.e., from the power of the electromagnetic wave. He assumed, therefore, that the electrons did not leave the metal with energy received from the wave, but with energy they already possessed when they were inside the atom. Relying on Hendrik Antoon Lorentz's picture of the electron as a virtual oscillator within the atom, Lenard assumed that the electromagnetic wave only triggered the release of electrons through some undetailed resonance between their and the electrons' frequency of oscillation. Since the wave was not thought to provide energy to the electrons, this mechanism was termed the "triggering hypothesis" (von Schweidler 1898; Wheaton 1978; Heilbron 1994).

Lenard's idea was very similar to the common explanation of anomalous dispersion as a resonance between light and matter, proposed by Wolfgang Sellmeier in 1872 (Sellmeier 1872; Jordi Taltavull forthcoming). It became the most popular answer to the paradox of quality. Since, according to the hypothesis, the phenomenon is atomistic in character, the triggering hypothesis also explained gas ionization by ultraviolet light. To support its extension to other phenomena, two of Thomson's students examined empirical consequences of the hypothesis for γ - and X-rays. In 1905, John A. McClelland found a correlation between atomic weight and the intensity of secondary electrons induced by β - and γ -rays. From his point of view, this correlation indicated that the electron's energy originated in the atoms. Two years later, P. D. Innes showed that, as in the case of light, "[t]he velocity of the fastest electrons emitted from each metal [by X-rays] is completely independent of the intensity of the primary rays, but increases with the hardness of the tube ['the penetrating power' of the X-rays]" (Innes 1907, 462; Wheaton 1983, 73–75). Similar experiments in Germany and the United States led to the same conclusion. These results convinced many experimentalists that the energy of the electrons originated in the atoms, and thus that the triggering hypothesis

could be extended to X- and γ -rays. The direct experimental corroboration of theoretical assumptions displays the close connection between theoretical thinking and the laboratory. Following its extension to X-rays, physicists also saw an answer to the paradox of quantity in the triggering hypothesis, assuming that at any instance only a small portion of the atoms contained electrons oscillating at the frequency of the ray (Innes 1907). By then, the triggering hypothesis was popular among German and British physicists alike.

From the perspective of the new theoretical physics, measuring physicists should have tested the mathematical laws that followed from the triggering hypothesis, namely the relation between the frequency of light and the maximal electron velocity. This appears to be a classical case in which exact measurements should arbitrate between two alternative mathematical laws and consequently between the theories that produce them. According to Lenard's assumption, the frequency of light is proportional to the velocity of the electrons, while, according to Einstein's light quantum, the frequency is proportional to the square of the velocity. Unlike Lenard, who advanced a triggering mechanism as a generalization from and explanation of a particular experimental result, Einstein proposed the light quantum hypothesis in 1905 to answer a theoretical puzzle within statistical thermodynamics. Although he had not conceived the assumption of discrete quanta of light energy as the answer to any particular empirical findings, Einstein did employ it to explain the phenomena of fluorescence, ionization by ultraviolet light, and, more famously, photoelectricity (Einstein 1905; Klein 1963). Still, despite this apparently clear way to arbitrate between theories, experimentalists did not rush to test the predictions of the triggering hypothesis versus Einstein's law for photoelectricity. When they did examine the triggering hypothesis, they first concentrated on other implications of the hypothesis, especially those that did not require high precision to distinguish between the two mathematical laws. They found other evidence more compelling. In other words, they preferred the methods of experimental physics to those of measuring physics associated with the new kind of theoretical physics.

In 1907, Erich Ladenburg carried out a series of experiments that examined the triggering hypothesis and its implications for the electrons inside matter. If this hypothesis was true, one could learn about the energy of electrons inside the atom (or in the metal) from the emission energy of photoelectrons. Ladenburg thought that, due to thermal motion, more electrons inside the matter should have high velocities with increasing temperature, and thus he examined the changes in photoelectricity at high temperatures (up to 800°C). He found, however, no change in photoelectric current and suggested that the effect on the distribution of electron velocities was too small to be observed in the experiment. The implication of the result for the triggering hypothesis was far from decisive. Two

years earlier, A. Lienhop, a student of Lenard, viewed the independence of the effect from temperature (which he showed for low temperature) as a corroboration of the hypothesis, since he assumed that the oscillations of electrons inside the atoms are independent of the temperature (Lienhop 1906).

For Ladenburg, however, the result indicated the need to examine the triggering hypothesis further. Thus, during 1907–1908, he examined the relationship between the frequency of absorbing light and the velocity of emitted electrons. He found that “the initial velocity of the electrons is proportional to the frequency of the absorbed light” (Ladenburg 1907, 514). While this conclusion agreed with the predication of the triggering hypothesis, his finding that light releases electrons in all the many frequencies examined cast doubts on it. To keep the triggering explanation, the latter finding required the assumption of many modes of electronic oscillations in matter. Ladenburg did not worry about that, probably because he considered that electronic oscillations in the metal happened also between the atoms and not only inside the atom. Others like Thomson, however, found it problematic to explain these results, especially as Thomson assumed that metal atoms consist of dozens rather than thousands of electrons (Wheaton 1978, 136–139). By 1910, further experiments corroborated Thomson’s view about the reduced number of electrons in the atom (Heilbron 1977), and therefore made the triggering hypothesis dubious.

Furthermore, soon after the publication of Ladenburg’s results, Abraham Joffé (1907) challenged their consistency with the triggering hypothesis. Against the assertion of the experimenter, he claimed that Ladenburg’s measurements did not establish a linear relation between maximum velocity and light frequency. Joffé showed that the results agreed at least as well with Einstein’s “atomistic hypothesis of the radiation energy” (Katzir 2006a, 452–453). In the narrow spectrum of ultraviolet light that could be employed to release electrons from metals, the difference between the two mathematical relations (i.e., between a straight line and a segment of a parabola) was smaller than the experimental error. Moreover, these were delicate experiments prone to quite a few sources of error. Further attempts to measure the relations between frequency and velocity led to conflicting and ambiguous results. The experiments could not arbitrate between the quantitative relations that followed the triggering and Einstein’s hypotheses. While a few physicists trusted two 1912 measurements, one by Owen Richardson and Karl Compton and another by Arthur Hughes, which supported Einstein’s relation, other experimentalists cast doubts on some of their results, and thereby, on their conclusions. The indecisive outcome of these experiments exemplifies the limitations of the measuring approach. Only Robert Millikan’s 1916 measurements settled the issue to the satisfaction of most experimentalists. Yet, most physicists had already abandoned the triggering hypothesis at the beginning of the decade.

They had done so because of the accumulation of other evidence from different areas (Katzir 2006a, 453–454).

Lenard himself provided the strongest individual blow to the triggering hypothesis. In 1911, together with Carl Ramsauer, he looked for a correlation between the emission of electrons and the absorption of ultraviolet light, and thus of energy, in the ionization of gases. Hence, they designed the experiment to check the fundamental assumption of the triggering mechanism, i.e., that the energy of the electrons originated in matter. Against the triggering hypothesis, the experiment displayed a correlation between the absorption of light energy and electron emission, showing, in Ramsauer's words, that "the energy of the ejected electron does not come from the atoms as originally assumed by Mr. Lenard, but from the absorbed light" (Wheaton 1983, 178). In the tradition of experimental physics, the experiment directly examined a theoretical assumption—the atomic source of the photoelectrons' energy. This contradiction of the triggering hypothesis highlighted the paradoxes of radiations, paradoxes which became even more acute with the link made between ultraviolet light and X- and γ -rays.

1.3 Visible Light and X-rays

One way to avoid, or at least to lessen, the contradictory aspects of radiation was to separate X- and γ -rays from visible light and from ultraviolet light. In a sense, this was Bragg's and Thomson's strategy. Concentrating on the powerful radiation, Bragg preferred viewing X- and γ -rays as a neutral pair. After toying with the triggering hypothesis, Thomson had, by 1907, abandoned it. Instead, he looked for an answer in the structure of the electromagnetic ether. He viewed the ether as full of "Faraday tubes," which consisted of bundles of lines of electric force giving a coarse-grained appearance to the ether, rather than a continuous one, when the plane of an electromagnetic wave intersected these tubes. These apparently grained surfaces would guide the X-rays impulses only in a discrete number of specific directions. Like Röntgen's suggestion of longitudinal waves, Thomson developed the concept of tubes of force in his earlier experimental study of cathode rays and electric discharge (Smith 2001, 27–35; Navarro 2012, 60–70, 114–119). Yet, interference precluded Thomson from assuming that the apparent grainy character of the ether also influenced light.³ Instead, in 1910 and 1913 he advanced two incompatible but equally speculative atomic models that could explain the continuity in the velocity of the emitted electrons. Disregarding Lenard and Ramsauer's result, both models involved a kind of triggering hypothesis. Suggesting tubes of force inside the atom, Thomson succeeded in regaining a thresh-

³The separation of X-rays from visible light was based on the assumption that the grains of ether are small enough to be inconsequential for ordinary light.

old frequency for releasing an electron equivalent to that of Einstein's equation. Despite the deep differences in their fundamental assumptions, some physicists conflated Thomson's view of corpuscle-like radiation with Einstein's, since both ascribed particle-like properties to radiation (Wheaton 1983, 136–142; Stuewer 1970, 252–253; Millikan 1963, 221–223).

Experimental results, however, pointed to a close parallelism between X-rays and light. As mentioned previously, experiments in 1907 showed that like the velocity of electrons released by light, the velocity of electrons released by X-rays is independent of the rays' intensity, but increases with what was regarded as their hardness. In 1910, Otto Stuhlmann in the United States strengthened the other direction of the analogy. Stuhlmann prepared thin films of platinum by sputtering the material onto quartz plates, which are transparent in ultraviolet light. He showed that similar to the emission induced by X-rays, more electrons were emitted by ultraviolet light in the direction of emergent rather than incident light beam, i.e., more electrons were released in the direction in which light propagated (Stuhlmann 1910; Wheaton 1983, 234–236). In the tradition of experimental physics, these experiments were designed to examine particular claims about the behavior of these rays, rather than to check particular mathematical results. Stuhlmann, for example, tested if ultraviolet beams behaved like X-rays in producing stronger effects in the emergent direction. To that end, he produced special thin platinum surfaces in vacuum and constructed an apparatus that enabled him to compare the effects of emergent and incident beams.

Due to earlier experimental demonstrations of the similarity between different types of radiation, many physicists regarded the celebrated 1912 demonstration of X-ray diffraction merely as an additional support for an already-assumed unity between X-rays and visible light. Following Max Laue's suggestion, Walter Friedrich and Paul Knipping found interference patterns of X-ray beams passing through crystals. Interference patterns alone, however, could not form conclusive evidence for an undulatory nature. Within a few months, Lorentz showed that like continuous waves impulses also could interfere. Still, this theoretical result did not make much of a difference, especially since the subsequent experiments of William Henry Bragg with his son William Lawrence Bragg, and those of Henry Moseley and Charles G. Darwin on diffraction patterns of X-rays in crystals, showed a full analogy with light waves. "By the fall of 1913, virtually all physicists believed that crystal diffraction had indisputably demonstrated that X-rays are equivalent to ordinary light" (Wheaton 1983, 215). Experiments established the view that the rays are high frequency electromagnetic waves. To demonstrate the wave properties of radiation, William Lawrence Bragg relied on the theoretical analysis of light waves. Still, unlike experiments in the tradition

of measurement physics, the conclusion did not depend on the explicit correspondence to a specific mathematical equation (Jenkin 2008, 329–339).

On the one hand, the conclusion apparently solved the quandary about the nature of X- and γ -rays. On the other hand, it highlighted the paradoxical properties of both visible light and X-rays. The rays showed a dual character: a particle-like absorption and a wave-like diffraction. X- and γ -rays provided a strong qualitative argument against the triggering hypothesis. They led to the implausible assumption that electrons possess an exceedingly high energy inside the atom. The unification of X-rays and light also provided grounds to support Einstein's relation between the electron's energy and the frequency of light. In the relatively small range of ultraviolet frequencies, doubts about experimental precision did not allow for a clear conclusion about this relation. Yet, already by 1913, Robert Pohl and Peter Pringsheim had concluded that

extrapolation [of the Einstein relation] to the probable frequencies of the Röntgen spectrum leads to velocities for the electron liberated by Röntgen rays which agree in *order of magnitude* with those experimentally observed. (Pohl and Pringsheim 1913, 1019, emphasis added)

Thus, a rather approximate or qualitative kind of experimental research into the nature of X-rays supported even a theoretical mathematical rule concerning the exact dependence of electron velocity on light frequency.

1.4 Later Theoretical Accounts

By 1911, most physicists had abandoned the triggering hypothesis. Since they had already regarded X-rays as a kind of light before the diffraction experiments, they accepted the evidence from X-rays as supporting Einstein's linear relation between light frequency and an electron's energy. Yet they did not accept his explanation. A few of them suggested alternative theories of photoelectricity, deriving the linear relation without admitting "light quanta," but with a quantum assumption. The failure of the triggering hypothesis and the paradoxes of quantity and quality that discredited an impulse theory, and the theoretical argument from black-body radiation convinced most influential physicists of the need to introduce quantum discontinuity in their theories of the subject. In the background was the success of the quantum hypothesis in accounting for the specific heat of solids at low temperatures (Kuhn 1978, 210–220). Thus, Thomson's speculative model was the only treatment of the relationship between radiation and matter that did not require directly a quantum assumption, but some sort of structure in the ether through his Faraday tubes and a triggering mechanism.

At the same 1911 meeting in which Ramsauer abandoned the triggering hypothesis, Sommerfeld suggested a promising quantum condition for the ejection of electrons by radiation. According to his suggestion, bound electrons are emitted when the time integral on their Lagrangian (a function of their energy), which increases with the absorption of light, reaches Planck's quantum of action. In this way, part of their energy originates in the light and part in the atom. With his former assistant Peter Debye, Sommerfeld retrieved Einstein's equation (without the extraction work) for monochromatic light. Sommerfeld and Debye's theory, however, suffered from a few problems, the most severe of which turned out to be the accumulation time needed for the release of the first electron. In 1913, barely a month after they published a revised theory, Erich Marx announced his experimental results, according to which the actual time delay was at most 1% of the predicted value (Wheaton 1983, 180–189; Stuewer 1975, 55–58). In the tradition of experimental physics, Marx devised the experiment to examine a specific implication of the theory (shared by a family of electromagnetic theories), rather than to test a mathematical prediction. His experiment practically excluded all explanations that require time accumulation.

In 1913, Planck advanced another explanation of photoelectricity suggesting that a few of the electrons in the atom possess energy close to their emission energy before illumination, so that no accumulation time is needed for their release. Unfortunately, the assumption implied that the electrons' energy originates in the atoms as in the original triggering hypothesis, and against Lenard and Ramsauer's finding (Kuhn 1978, 235–254; Wheaton 1983, 178–180).⁴

The attitude of most physicists toward a theory suggested by Richardson illuminates their positions regarding the interaction between electromagnetic radiation and matter. In 1912, and in a revised version in 1914, Richardson suggested a descriptive rather than an explanatory theory of photoelectricity, "wish[ing] to avoid discussion of the vexed question of the nature of the interaction between the material parts of the system and the æthereal radiation" (quoted in Katzir 2006a, 456). Richardson accounted for the central experimental findings, including Einstein's equation, by employing thermodynamics and statistical reasoning for equilibrium between emitted and returned electrons, and a quantum assumption through the admission of Planck's distribution law of electromagnetic radiation. In particular, he did not need to assume the light-quantum hypothesis, which most physicists at the time still rejected. Nevertheless, most physicists did not adopt the new descriptive account. Even if Richardson's theory suffered from a few internal weaknesses, it seems that for most physicists a descriptive

⁴The theory shared another defect with the triggering hypothesis, namely, it required a complicated atomic structure to account for the continuous velocity spectrum of the emitted electrons, although unlike Lenard, Planck did assume a resonance mechanism.

theory, which bypasses the question of mechanism, was not enough to cope with the challenges put forward by the relations between radiation and matter. The failure of classical explanations and the accumulation of evidence from γ -rays to visible light pointed not only to an unexplained effect, but also to paradoxes. Moreover, unlike many processes of complex matter that were still unexplained, the peculiarities of radiation could not be attributed to an intricate structure whose details were still unknown. Light did not have a complex structure. Advances in the knowledge of atoms suggested that atomic structure was unlikely to account for the interaction of light with matter (Katzir 2006a).

Thus, physicists expected a radical change in the basic concepts of their discipline, and saw little value in attempts to bypass them. Millikan, a strong opponent of the light quantum, spoke for most physicists: “The new facts in the field of radiation [...] seem, then, to require in any case a very fundamental revision or extension of classical theories of absorption and emission of radiant energy,” (quoted in Katzir 2006a, 467). These facts included ionization by X- and γ -rays, frequency thresholds and black-body radiation. Most of these facts and their interpretation originated in a study characteristic of a traditional methodology of experimental physics. The challenges posed by the extensive study of different kinds of radiation and their interaction with matter made many scientists more open to ideas and concepts that departed from the classical, regarding not only these but also other physical phenomena.

1.5 Conclusion

“Experimental physicists,” in the meaning discussed in this paper, discovered photoelectricity, X- and γ -rays, revealed their surprising properties and exposed the deficiency of established notions and laws of physics to account for them. Neither elaborated theories characteristic of the emerging subdiscipline of theoretical physics nor precise measurements for testing their mathematical predictions drove the research on these phenomena. Rather, physicists working in more traditional modes of research suggested and later rejected the triggering hypothesis, displayed the similitude of visual light, ultraviolet, X- and γ -rays, revealed the paradoxes of quantity and quality and the wave-particle character of these rays. This kind of experimental physics was not detached from theory. On the contrary, theories and assumptions played a crucial role in interpreting experimental results and also in directing research. Theory and experiment were closely connected. Experimentalists explored the empirical implications of particular claims and hypotheses, designing their experiments to that end and suggesting hypotheses toward and following their results.

Thus, this kind of experimental physics, which is often associated more with the nineteenth than with the twentieth century, also played a significant role in the development of quantum physics. The contribution of this kind of research seems to depend more on the stage of the research than on the date of study. In the examples discussed here, experimental physics shaped fields that were not described by comprehensive theories. In fields described by such a theory, like black-body radiation, even when its foundations were under threat, the new theoretical physics with its associated measuring physics dominated. Theoretical and measurement physics prevailed also in more mundane fields of physics that did not experience a break with classical physics. For example, experimental research of the kind described here led piezoelectric research in the “pre-theoretical phase” of its study, i.e., from its discovery in 1880 until the formulation of a comprehensive theory a decade later. With the formulation of such a mathematical theory, the novel theoretical physics guided the study of the field (Katzir 2006b). Theoretical physics and its associated measuring physics seem to characterize research in “theoretical phases,” i.e., after the formulation of a comprehensive mathematical theory, rather than to be associated especially with a research on microphysics and the quantum. This agrees with Suman Seth’s recent claim in a study of Sommerfeld and his school that theoretical physics extended much beyond the fields that are usually associated with it like microphysics, relativity and quantum theory (Seth 2010, 4). Seth’s claim and my claim here—that quantum mechanics was far from restricted to the new theoretical physics—call for caution in linking the rise of quantum mechanics to that of theoretical physics.

References

- Buchwald, Jed (1994). *The Creation of Scientific Effects: Heinrich Hertz and Electric Waves*. Chicago: The University of Chicago Press.
- Büttner, Jochen, Jürgen Renn, and Matthias Schemmel (2003). Exploring the Limits of Classical Physics: Planck, Einstein, and the Structure of a Scientific Revolution. *Studies in History and Philosophy of Modern Physics* 34: 37–59.
- Darrigol, Olivier (1992). *From c-Numbers to q-Numbers: The Classical Analogy in the History of Quantum Theory*. Berkeley: University of California Press.
- (2000). *Electrodynamics from Ampère to Einstein*. Oxford: Oxford University Press.
- Einstein, Albert (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik* 17: 132–148.
- Hallwachs, Wilhelm (1916). Lichtelektrizität. In: *Handbuch der Radiologie*. Ed. Erich Marx. Leipzig: Akademische Verlagsgesellschaft. 245–561.

- Heilbron, John L. (1977). Lectures on the History of Atomic Physics. In: *History of Twentieth-Century Physics*. Ed. Charles Weiner. New York: Academic Press. 40–108.
- (1994). The Virtual Oscillator as a Guide to Physics Students Lost in Plato’s Cave. *Science and Education* 3: 177–188.
- Innes, P. D. (1907). On the Velocity of the Cathode Particles Emitted by Various Metals under the Influence of Röntgen Rays, and Its Bearing on the Theory of Atomic Disintegration. *Proceedings of the Royal Society A* 79: 442–462.
- Jenkin, John (2008). *William and Lawrence Bragg, Father and Son: The Most Extraordinary Collaboration in Science*. Oxford: Oxford University Press.
- Joffé, Abram (1907). Eine Bemerkung zu der Arbeit von E. Ladenburg ‘Über Anfangsgeschwindigkeit und Menge der photoelektrischen Elektronen usw.’ *Annalen der Physik* 24: 939–940.
- Jordi Taltavull, Marta (forthcoming). *Optical Dispersion at the Crossroads between Classical and Quantum Physics*. PhD Diss. Humboldt University.
- Jungnickel, Christa and Russell McCormach (1986). *Intellectual Mastery of Nature: Theoretical Physics from Ohm to Einstein, Volume 2: The New Mighty Theoretical Physics 1870–1925*. Chicago: The University of Chicago Press.
- Katzir, Shaul (2006a). Thermodynamic Deduction versus Quantum Revolution: The Failure of Richardson’s Theory of the Photoelectric Effect. *Annals of Science* 63(4): 447–469.
- (2006b). *The Beginnings of Piezoelectricity: A Study in Mundane Physics*. Dordrecht: Springer.
- Klein, Martin J. (1963). Einstein’s First Paper on Quanta. *The Natural Philosopher* 2: 57–86.
- Kuhn, Thomas S. (1978). *Black-Body Theory and the Quantum Discontinuity 1894–1912*. Oxford: Clarendon Press.
- Ladenburg, Erich (1907). Über Anfangsgeschwindigkeit und Menge der photoelektrischen Elektronen in ihrem Zusammenhange mit der Wellenlänge des auslösenden Lichtes. *Physikalische Zeitschrift* 8: 590–594.
- Lienhop, A. (1906). Über die lichtelektrische Wirkung bei tiefer Temperatur. *Annalen der Physik* 21: 281–304.
- Millikan, Robert A. (1963). *The Electron: Its Isolation and Measurement and the Determination of Some of Its Properties*. Chicago: The University of Chicago Press. Reprint of 1924 edition.
- Navarro, Jaume (2012). *A History of the Electron: J. J. and G. P. Thomson*. Cambridge: Cambridge University Press.
- Pohl, Robert and Peter Pringsheim (1913). On the Long-Wave Limits of the Normal Photoelectric Effect. *Philosophical Magazine* 26: 1017–1024.

- von Schweidler, Egon R. (1898). Über die lichtelektrischen Erscheinungen. *Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften Wien* 107(IIa): 881–909.
- Sellmeier, Wolfgang (1872). Über die durch die Aetherschwingungen erregten Mitschwingungen der Körpertheilchen und deren Rückwirkung auf die erstern, besonders zur Erklärung der Dispersion und ihrer Anomalien. *Annalen der Physik und Chemie* 145: 399–421.
- Seth, Suman (2010). *Crafting the Quantum: Arnold Sommerfeld and the Practice of Theory, 1890–1926*. Boston: MIT Press.
- Smith, George E. (2001). J. J. Thomson and the Electron, 1897–1899. In: *Histories of the Electron: the Birth of Microphysics*. Ed. Jed Buchwald and Andrew Warwick. Boston: MIT Press. 21–76.
- Steinle, Friedrich (1998). Exploratives vs. theoriebestimmtes Experimentieren: Ampères frühe Arbeiten zum Elektromagnetismus. In: *Experimental Essays – Versuche zum Experiment*. Ed. Michael Heidelberger and Friedrich Steinle. Baden-Baden: Nomos-Verlag. 272–297.
- Stuewer, Roger H. (1970). Non-Einsteinian Interpretations of the Photoelectric Effect. In: *Historical and Philosophical Perspectives of Science*. Ed. Roger H. Stuewer. Minneapolis: University of Minnesota Press. 246–263.
- (1971a). Hertz’s Discovery of the Photoelectric Effect. *Actes du XIIIe Congrès International d’Histoire des Sciences* VI: 35–43.
- (1971b). William H. Bragg’s Corpuscular Theory of X-Rays and γ -Rays. *The British Journal for the History of Science* 5: 258–281.
- (1975). *The Compton Effect: Turning Point in Physics*. New York: Science History Publications.
- Stuhlmann, Otto (1910). A Difference in the Photoelectric Effect Caused by Incident and Emergent Light. *Philosophical Magazine Series 6* 20(116): 331–339.
- Wheaton, Bruce R. (1978). Philipp Lenard and the Photoelectric Effect, 1889–1911. *Historical Studies in the Physical Sciences* 9: 299–322.
- (1983). *The Tiger and the Shark: Empirical Roots of Wave-Particle Dualism*. Cambridge: Cambridge University Press.

Chapter 2

Challenging the Boundaries between Classical and Quantum Physics: The Case of Optical Dispersion

Marta Jordi Taltavull

This paper describes one significant episode in the transition between classical and quantum theories. It analyzes the first theory of optical dispersion that ensued from the extension of Niels Bohr's quantum model of the atom to other optical phenomena outside of spectroscopy. This theory was initially developed by Peter Debye in 1915 and then was endorsed and extended by Arnold Sommerfeld in 1915 and 1917. The most interesting aspect of the Debye-Sommerfeld theory for the present paper is that it clearly typifies important features of debates concerning the boundaries between classical and quantum physics, focusing on the period from 1913 to the early 1920s.

Optical dispersion consisted of splitting white light into different colors because of its change of velocity when passing through a transparent, prismatic medium. From the 1870s onward, it was well known that light was continuously dispersed across the entire spectrum, except at those specific frequencies, characteristic of the medium, at which light was completely absorbed. In other words, dispersion and absorption of light were complementary phenomena. From 1872, this behavior was explained using one enduring theoretical representation: the *Mitschwingungen* model. This model pictured the interaction between light and matter as a continuous process of interaction between waves and particles performing induced vibrations, called *Mitschwingungen*.

In 1913, this model conflicted with certain aspects of Bohr's quantum model of the atom. Contrary to the *Mitschwingungen* model, Bohr envisioned the exchange of energy between light and matter as a discrete process, mediated by the emission or absorption of quanta of energy.

Debye and Sommerfeld's theories were the first attempts at combining optical dispersion with the new atomic model. To do so, both physicists had to come to terms with whether optical dispersion could still be considered a classical process, even in the context of Bohr's model, or had to be reinterpreted as a quantum phenomenon, in the same fashion as spectral lines.

Debye and Sommerfeld decidedly followed the first path in 1915. They borrowed various elements of the classical *Mitschwingung* theories and embedded them into Bohr's model. Most importantly, Sommerfeld saw a clear confirmation in the new theory of optical dispersion that quantum and classical physics could coexist without causing inconsistencies. Hence from 1915 to 1917, he defended it, despite skeptical and critical responses. In so doing, Sommerfeld defined a divide between two domains of physics: optical dispersion as a central example of classical physics and spectroscopy as the central phenomenon of quantum physics.

In the 1920s, Sommerfeld's theory could no longer withstand certain criticisms addressing this divide. Optical dispersion had to be regarded as a pure quantum phenomenon. This aroused the question of accounting for its continuous features in quantum terms, so well-explained by the classical model of *Mitschwingungen*. This search for a quantum explanation of optical dispersion brought about a renegotiation of quantum concepts and techniques according to different strategies from the 1921 onward, most importantly in Sommerfeld's school in Munich and in Bohr's school in Copenhagen.

In this paper, I follow the development of this intricate story from 1913 through the early 1920s, focusing particularly on Sommerfeld's intervention in the debate. In the first section, I summarize the main aspects of the classical theory of optical dispersion developed from the 1870s until 1913. I deal with Bohr's model of the atom in section 2. In sections 3 and 4, I describe at length Debye and Sommerfeld's theory. Then in section 5, I detail the extensions, comments and criticisms of the theory; and in the following section, I tackle Sommerfeld's strategies to counter them. Sommerfeld's last words on the theory are addressed in section 7. In section 8, I discuss a new direction in the debate about classical versus quantum optical dispersion. Finally, I conclude with a short introduction to the quantum theories of optical dispersion that emerged in the 1920s.

2.1 The Classical Theory of Optical Dispersion

2.1.1 Microphysics and Electromagnetic Theory

In the 1870s, some peculiar features of optical dispersion, a phenomenon recognized as early as the 17th century, were discovered. In particular, the frequency dependence of the velocity of light (and as a result, its direction) was continuous across the range of the spectrum, except at those frequencies where light was absorbed by the dispersing material.

As shown in fig. (2.1), in the neighborhood of the absorbed frequency, the index of refraction increases asymptotically as one approaches the singularity from

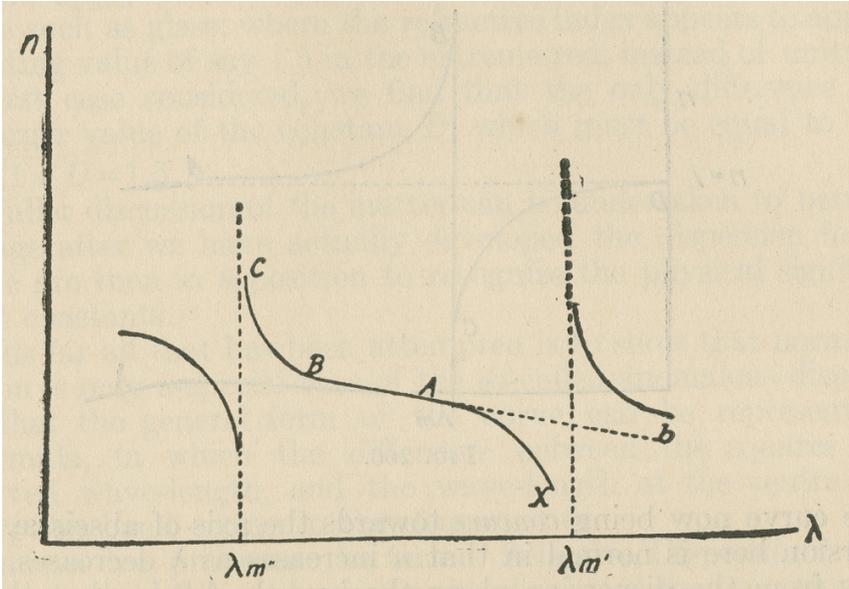


Figure 2.1: Graphic representation of the index of refraction as a function of the wavelength of light (Wood 1904, 377).

the right (region C–B), and it decreases asymptotically as one approaches from the left (region A–X). This was dubbed “anomalous dispersion.” This phenomenon was first discovered by Christian Christiansen and then thoroughly investigated for liquids by August Kundt during 1870–1872 (Christiansen 1870; Kundt 1870; 1871a; 1871b; 1872). Then Robert Wood in 1904, Rudolf Ladenburg (in collaboration with Loria) in 1908, and P. V. Bevan in 1910 provided a quantitative description for various gases (Wood 1904; Ladenburg and Loria 1908; Bevan 1910). The region B–A in fig. (2.1) corresponds to normal dispersion as observed since Newton’s time.

As previously mentioned, at least as early as 1872 physicists represented these features using a specific model of interaction between microscopical particles and light waves: the *Mitschwingungen* model (Sellmeier 1872a; 1872b; 1872c; 1872d; von Helmholtz 1875; discussed in Buchwald 1985; Whittaker 1910). According to this model, when light impinged on matter, particles and light waves oscillated together so that the dispersed light stemmed from the entangled waves induced by matter oscillations and primary waves. This

approach was radically different from all earlier explanations of optical dispersion. Previously, matter could only modify the propagation of light, without contributing to its generation.

More specifically, according to the *Mitschwingungen* model, particles were quasi-elastically bound to their equilibrium positions, which led to their having characteristic proper frequencies of vibration. When light interacted with these particles, it caused a forced oscillation, which was assumed to be the cause of the emission of a secondary set of waves, having the same frequency as the original light, but delayed by a phase factor. The primary light and the secondary radiation were presumed to interfere, and thus form new waves with different phase velocities depending on the frequency of the light ν . This yielded the following equation for the index of refraction (that is the ratio between the velocity of light in the medium and the velocity of light in vacuum):

$$n^2 - 1 = \sum_{i=1}^{\infty} \frac{K_i}{\nu^2 - \nu_i^2}, \quad (2.1)$$

where each term of the summation corresponds to one possible proper frequency of the dispersing matter. However, if the incoming light had the same frequency as the proper frequency of any of the particles ($\nu = \nu_i$), the light was entirely absorbed by the matter, without any emission of secondary waves. In such a case, the light and matter came into resonance.

The phenomenon of optical dispersion was thus defined by two parameters, the proper frequency ν_i and the constant K_i , which somehow played the role of an “intensity” of dispersion. Both parameters could be calculated a posteriori by fitting experimental data into the above formula.

The essence of this microscopic mechanism of matter interacting with light remained unchanged for over fifty years, although physicists would embed it into different frameworks. With the establishment of the electromagnetic theory of light, particles came to be considered vibrating charges. According to the electromagnetic version of the model, the vibration of these charges induced a periodic polarization of the medium, which in turn caused secondary waves to be emitted (Glazebrook 1886; von Helmholtz 1892; Drude 1894; 1900; Voigt 1899; 1901).

2.1.2 Optical Theory and the Structure of Matter

Paul Drude systematized the electromagnetic approach to optical dispersion in two editions of his *Lehrbuch der Optik* (Drude 1900) as well as in his research papers (Drude 1904a; 1904b). In these works, Drude established an extremely fruitful connection between optics and the physico-chemical properties of mat-

ter. His approach relied on two fundamental steps. First, he suggested that the charged particles involved in dispersion were in fact the recently discovered electrons, characterized by the ratio e/m , first measured by Emil Wiechert and Joseph Thomson in 1897. In 1904, Drude took up the value of this ratio e/m and, by assuming that K_i was proportional to the number of optical electrons N_i with proper frequency ν_i , obtained $K_i = 4\pi \frac{N_i e^2}{m}$.¹ This gave the following expression for the index of refraction:

$$n^2 - 1 = \frac{4\pi e^2}{m} \sum_{i=1}^{\infty} \frac{N_i}{\nu^2 - \nu_i^2}. \quad (2.2)$$

The number N_i was called the number of “dispersion electrons” thereafter. This definition of K_i led to optical dispersion becoming a very powerful tool to investigate the microphysical structure of matter.

Drude’s second step was to suggest that the electrons responsible for dispersion were the so-called valence electrons. This proposal was far from obvious. It was not until 1904 that chemical valence was connected to electron theory. Richard Abegg suggested that the valence number corresponded to the number of electrons loosely attached to an atom and having the tendency to migrate from one atom to another to form molecules (Abegg 1904). In taking up Abegg’s suggestion, Drude extended the boundaries of the concept of electronic valence from physical chemistry to the rapidly-growing field of applications of the electron theory to optics. This added a molecular dimension to the problem of optical dispersion and suggested its use as a tool for exploring the properties of the periodic table.

The identification of dispersion electrons with valence electrons was problematic. In some cases, see for example the work of Clive and Maude Cuthbertson (1910) and John Koch (1913), some slight discrepancies were found for normal dispersion in gaseous molecular hydrogen, helium, oxygen and nitrogen. More specifically, the number of dispersion electrons N calculated by fitting the experimental data using eq. (2.2) was only two-thirds of the number of valence electrons of these gases. The situation was more serious for anomalous dispersion in vapors of sodium, potassium and monoatomic hydrogen, where the discrepancies were larger by orders of magnitude (Wood 1904; Ladenburg and Loria 1908; Bevan 1910). However, the *Mitschwingung* mechanism of light-matter interaction was not affected by these inconsistencies. The relation between dispersion and valence electrons only affected issues at the border between physics and chemistry and not the more general mechanical model.

¹After Thomson, the ratio e/m was measured by other physicists, among them Walter Kaufmann. Indeed, Drude referred explicitly to Kaufmann’s measurement of e/m (Kaufmann 1902; 1906).

2.2 Bohr's Model of the Atom and the Optical Dispersion of Molecules

In 1913, a new element entered the story: Bohr's quantum model of the atom. In the series of papers he published that year (Bohr 1913a; 1913b; 1913c), Bohr assumed new laws of atomic stability derived from the introduction of a single parameter that was completely foreign to classical electrodynamics: Planck's elementary quantum of action h . Bohr's model was defined by the way matter absorbed or emitted quanta of light of energy $h\nu$.

Bohr postulated that orbiting electrons maintained constant trajectories, which were mechanically stable and on which they did not radiate. Emission and absorption of energy only took place when the electrons jumped from one orbit to another. In the new theory, the frequency ν_{12} of the radiation absorbed or emitted was related to the energies E_1 and E_2 of the orbits or levels through the relation $E_1 - E_2 = h\nu_{12}$. With this novel redefinition of the exchange of energy between matter and light, Bohr broke the classical link between the absorption/emission frequency of radiation and the mechanical frequency of the vibrating electron that was at the root of the *Mitschwingungen* model.

Bohr supported his model by applying these postulates to monoatomic hydrogen, where he found an excellent agreement between the frequencies ν_{12} and the spectral lines of the Balmer series. In addition to this simple case, in the second and third part of the trilogy, Bohr addressed at length the constitution and stability of more complex atomic systems and molecules. To fix the conditions of stability for electron orbits, Bohr introduced another postulate: the angular momentum along a stationary orbit must remain constant.

The quantized nature of the angular momentum imposed severe restrictions on the stability of the systems under the action of external forces since it allowed only two possible modifications of the orbits. Within the same plane of the orbit, changes could only occur by quantum jumps corresponding to changes in energy of value $E_1 - E_2 = h\nu_{12}$. Perpendicular to the plane of the orbit, trajectories could undergo small periodic variations, but only if the radius remained unchanged, thus preserving the constancy of angular momentum. The frequency of such small variations could be computed using the tools of ordinary mechanics.

This division into two separate mechanisms mirrored a distinction that Bohr introduced in his trilogy between the "true emission of light" and the "scattering of light." Bohr employed this dichotomy to characterize the ordinary line-spectra of atomic hydrogen ("true emission of light," following quantum rules) and the lines of the coronal spectrum of the sun discussed by John W. Nicholson in 1912 ("scattering of light," ruled by classical mechanics) (Bohr 1913a, 23–24). In addition, he relied on this distinction in the context of optical dispersion, since he

presumed that this phenomenon was also caused by mechanical perturbations of orbits:

The ordinary mechanics can be used in calculating the vibrations of the electrons perpendicular to the plane of the ring—contrary to the case of vibrations in the plane of the ring. This assumption is supported by the apparent agreement with observations obtained by Nicholson in his theory of the origin of lines in the spectra of the solar corona and stellar nebulae. In addition it will be shown later that the assumption seems to be in agreement with experiments on dispersion. (Bohr 1913b, 482)

In the second and third part of his trilogy, Bohr displayed several models of multi-electronic atoms and molecules. To test them with concrete experimental outcomes, he calculated the theoretical values of proper frequencies, using either the ordinary laws of mechanics or quantum postulates for the cases of perturbations perpendicular to the plane of the orbit and electron jumps, respectively. Molecules remained theoretically stable only when their alterations were characterized by these frequencies. Optical dispersion was one of the most effective tools to obtain empirical values of proper frequencies. Since these frequencies did not usually lie in the visible range of the spectrum, they could be derived by fitting the experimental data using eq. (2.2). The hope was that a comparison between theoretical and empirical outcomes could confirm or discard each concrete molecular model.

For example, in the case of H_2 , Bohr put forward a model in which two electrons orbited two nuclei located along the axis of symmetry, see fig. (2.2).

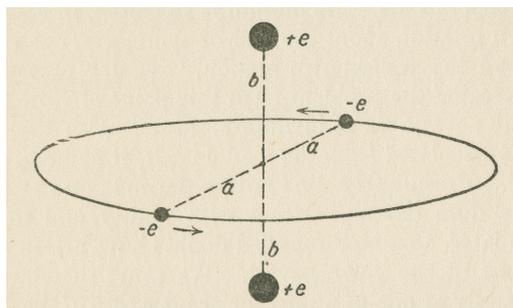


Figure 2.2: The hydrogen molecule according to Bohr (Reiche 1922, 75).

In this particular instance, the comparison seemingly confirmed Bohr's model of H_2 , since the frequency computed from the Cuthbertsons' experiments on optical dispersion (C. Cuthbertson and M. Cuthbertson 1910) turned out to be in agreement with the theoretical values. The results did not provide an unambiguous arbitration concerning the two possible theoretical mechanisms (mechanical or quantum) causing the proper frequency of optical dispersion, for the empirical value was consistent with both. However, since the model of H_2 was in any case confirmed, Bohr did not linger over this ambiguity.

2.3 Debye and the First Theory of Optical Dispersion According to Bohr's Model

The first systematic attempt to combine the classical theory of *Mitschwingungen* with the new picture of matter was carried out by Debye. He chose to address this problem because "a relation between quasi-elastically bounded electrons, which are necessary [in optical dispersion] and the rotating electrons, which are present [in Bohr's model], is missing" (Debye 1915, 1–2).

Debye took up Bohr's idea that the atom was like a planetary system, in which the quantum hypothesis came into play by delimiting the angular velocities of electrons. Furthermore, he expanded Bohr's suggestion that optical dispersion might be caused by mechanical perturbations of stationary orbits. To do so, Debye approached optical dispersion as if it were a purely classical perturbation problem, which he solved for the specific model of H_2 proposed by Bohr.

More specifically, he assumed that electromagnetic light was able to perturb molecular orbits through *Mitschwingungen* in the same way as it perturbed proper electron vibrations in Drude's theory. The essential difference between Debye and Drude was that Debye's starting point were the concrete equations of motion of the unperturbed orbits, while Drude started from the model of an ordinary oscillator. The quantum of action h was not brought into play to govern any exchanges of energy between matter and radiation but only entered the calculation of the angular momentum of the stationary orbit.

Debye's procedure boiled down to a restoration of the classical connection between the frequencies of matter oscillations (in this case, oscillations of orbits) and the frequency of the emitted light waves. Furthermore, it led to several possible modifications of the orbits instead of the unique modification suggested by Bohr. However, Debye did not dwell on the consequences of his extension of Bohr's double mechanism. Most importantly, he did not comment on fundamental questions concerning the compatibility between classical and quantum concepts.

Using the experimental data collected by Koch (1913), Debye arrived at a convincing twofold confirmation of his theory. First, the empirical number of dispersion electrons was one and a half times the number predicted by Drude. Debye obtained a new expression for the quantity K_i , which filled this gap and restored the agreement with experiments. Second, the proper frequencies s_i , ($i = 1, \dots, 5$) of the five possible perturbations turned out to be proportional to the angular frequency ω of the orbit via a parameter x_i whose value could be calculated theoretically. Using Koch's data to fix one of the frequencies s_i , Debye arrived at a value of ω which was in very good agreement with the value calculated using the expression from quantum theory $mr^2\omega = \frac{h}{2\pi}$ (r being the radius of the orbit, m the mass of the electron). Debye considered this a strong confirmation of his theory of optical dispersion from both the standpoint of the adequacy of the procedure and the validity of the specific model of molecular hydrogen.

A few months later, a doctoral student of Debye's, Paul Scherrer, extended this procedure to include an explanation of the Faraday effect for a gas of molecular hydrogen.² Further, Scherrer perceptively spotlighted the question of the compatibility between classical and quantum concepts in Debye's theory of optical dispersion. For him, its success in explaining experimental data was undoubtedly a "confirmation [...] that the classical laws of mechanics can be legitimately applied to the calculation of the perturbations in the inner atom" (Scherrer 1915, 180).

2.4 Sommerfeld and Optical Dispersion in 1915

Sommerfeld also saw a confirmation in the success of Debye's theory of optical dispersion that the quantum postulates and the classical laws of mechanics and electrodynamics could eventually coexist. Indeed, starting from Debye's approach, in 1915 Sommerfeld sought to understand the general features of this coexistence: "from [Debye's theory] one understands that the laws ruling the inner atom do not differ from classical mechanics and electrodynamics as one could presume from Bohr's postulates" (Sommerfeld 1915, 549).³

²The Faraday effect concerns the change in light polarization when light passes through a transparent medium under the action of a constant magnetic field. Indeed, the change in light polarization also depends on the frequency of the light, and this dependence is continuous for the range of the spectrum, except at the absorption frequencies of the material, as in the case of optical dispersion.

³Actually, this 1915 paper was the first in which Sommerfeld applied a form of quantization of the atom. But it was not the first time he dealt with optical dispersion. He tackled the issue as early as 1907 and returned to it in 1912 (Sommerfeld 1907; 1912). The goal of these efforts was to remove any doubts about the possibility that light could propagate faster than c in dispersive media and consequently be in contradiction with certain assumptions of relativity. To that end, he carefully examined the model of *Mitschwingungen*, applying sophisticated mathematical tools and thereby further illu-

From Sommerfeld's standpoint, Debye's theory of optical dispersion implicitly suggested the following division between classical and quantum laws: the structure and stationary features of the molecule were determined by quantum laws, while the dynamical process of light dispersion was accounted for by classical physics.

In 1915, Sommerfeld generalized Debye's approach to optical dispersion to all types of molecules with axial symmetry using a model similar to the one presented in fig. (2.2). As in Debye's case, Sommerfeld's treatment relied on the mechanical perturbations induced in the stationary orbits by electromagnetic radiation. The only step of the argument requiring quantum physics was the determination of the angular velocity of the electrons along their orbits in the molecule through the expression $mr^2\omega = \frac{h}{2\pi}$. Quantum jumps were by no means implicated in optical dispersion.

Following this procedure, Sommerfeld eventually arrived at a general expression for the index of refraction as a function of the frequency of light, which could be approximated in the regime of low frequencies by:

$$n^2 - 1 = \frac{4\pi e^2}{m} \sum_{i=1}^{\infty} \frac{N \frac{C_i}{3}}{s^2 - s_i^2}. \quad (2.3)$$

This expression showed a close analogy to Drude's formula (2.2). However, it also entailed a fundamental difference in respect to Drude's: the singularities of the index of refraction were the proper frequencies s_i at which the electronic paths were oscillating around their stationary orbits. For Drude, ν_i were the proper frequencies of the electrons around their positions of equilibrium.

Furthermore, the intensity of dispersion K_i was weighted by a factor $\frac{C_i}{3}$, not present in Drude's formula, which accounted for the possible anisotropy of the molecule. Indeed, this factor C_i eventually provided a deeper justification to the two-thirds discrepancy between Drude's expression for the number of dispersion electrons and Koch's (1913) experimental observations, which had already been accounted for by Debye. Thus, it became clear that Drude's formulation only held for the isotropic case, when the three proper frequencies of the orbits (radial, azimuthal and axial) were exactly the same.

Despite the evident advantages of this approach, Sommerfeld openly discussed some conceptual difficulties that plagued the "hybrid" theory, including "a contradiction with electrodynamics [...], as the electronic orbits are not allowed to radiate" (Sommerfeld 1915, 549). In effect, the supposition that orbits could radiate energy during optical dispersion contradicted one of Bohr's postu-

minated the process of propagation of electromagnetic waves through dispersive media. For more details on Sommerfeld's scientific activity in Munich, see (Eckert 1993).

lates: the existence of *non-radiating* stationary states. Sommerfeld tackled this issue in 1917, in a work I analyze shortly.

In the meantime, the hybrid theory involved inner contradictions that would turn out to be insuperable. Sommerfeld singled out the conflict between quantum and classical physics at exactly the proper frequencies s_i , where the dispersion was discontinuous. Two possible explanations presented themselves. On the one hand, s_i could be considered the proper frequencies at which light was resonantly absorbed. On the other hand, according to Bohr's theory, those absorption frequencies should coincide with the emission (and thus spectral) frequencies, at which the exchange of electron energy with light was governed by quantum jumps. As it was clear to Sommerfeld that both classical *Mitschwingungen* and quantum transitions had to be part of the total picture, from his point of view the elaboration of a new theory of optical dispersion meant finding a way for classical and quantum approaches to coexist peacefully and consistently. In this regard, Sommerfeld was trenchant, at the points of discontinuity Bohr won, but in the rest of the spectrum the *Mitschwingungen* held their validity:

Therefore, our dispersion formula will be correct only at a sufficient distance from the emission and absorption lines and for normal dispersion. How one has to handle anomalous dispersion and the necessary absorption by electrons lies still in the deepest obscurity. (Sommerfeld 1915, 577)

This is how Sommerfeld, by analyzing more deeply the fundamental features of Debye's theory of optical dispersion, created a divide between the classical and quantum domains.

Nevertheless, the hybrid theory did not work for the specific cases of N_2 and O_2 , at least as far as the values of ω were concerned. Theoretical values were calculated by using the quantum relation $mr^2\omega = \frac{h}{2\pi}$, assuming the axial symmetry of the molecule. For experimental values, Sommerfeld resorted to the already-published experiments on optical dispersion with these gases (C. Cuthbertson and M. Cuthbertson 1910). For the case of O_2 , while experimental data predicted $\omega = 3.39 \cdot 10^{16}$, the theoretical values according to quantum discretization of orbits predicted $\omega = 3.76 \cdot 10^{17}$. Sommerfeld was not discouraged. He explained the variance by suggesting that these molecules might not be axially symmetric, and in no way did he suggest that the conceptual grounds of the hybrid theory had been undermined. As this case illustrates, the hybrid theory also entailed a methodological commitment: to find agreement between theory and experiment, physicists should focus on exploring new molecular models.

2.5 Reactions to Debye-Sommerfeld's Hybrid Theory

Between 1915 and 1917, articles by several physicists drew attention to the hybrid theory of H_2 . For instance, within Sommerfeld's group in Munich, Paul Epstein extended it to the problem of specific heats, and Adalbert Rubinowicz generalized the Debye-Sommerfeld theory to include the mechanical perturbations in the nuclei of the H_2 molecule. However, not all reactions to the hybrid theory were so positive. The theory also received severe criticism from Hendrika Johanna van Leeuwen, a postdoctoral fellow working under the guidance of Hendrik Antoon Lorentz and Paul Ehrenfest. Her critique concerned the stability of the H_2 molecule under the influence of external electromagnetic forces. But the strongest arguments against the theory came from Carl Wilhelm Oseen and Bohr. They categorically denied the very possibility that classical and quantum physical laws could truly coexist, which was the essential cornerstone of Sommerfeld's way of dealing with optical dispersion. The next three sections spell out all these reactions to Debye-Sommerfeld's hybrid theory.

2.5.1 Extension to Specific Heats

In 1916, Epstein took up Debye-Sommerfeld's model of the hydrogen molecule H_2 , "which has been strongly supported by Debye's calculations on dispersion" (Epstein 1916, 400) and extended it to provide a quantum account of rotational specific heats.⁴

Simultaneously and independently, Frederick Krüger also saw advantages in treating the H_2 molecule as a gyroscope, see fig. (2.3). Using Felix Klein and Sommerfeld's theory of the spinning top (Klein and Sommerfeld 1898) he calculated a theoretical value of the precession frequency of Debye-Sommerfeld's H_2 molecule ($\nu = 5.11 \cdot 10^{12}$), which was in satisfactory agreement with the

⁴From the experiments on specific heats performed with H_2 by Arnold Eucken in 1912 (Eucken 1912), it was known that the dependence of the energy of the gas on the temperature was characterized by a proper frequency in the infrared region of the spectrum independent of the instantaneous temperature of the system. This frequency was $\nu = 8.85 \cdot 10^{12}$. In 1913, Albert Einstein and Otto Stern (Einstein and Stern 1913), as well as Paul Ehrenfest (Ehrenfest 1913) put forward theoretical accounts of the phenomenon by picturing the molecule as a rotator and quantizing its rotational degrees of freedom. Yet this problem turned out to be unexpectedly difficult, for such a thing as a proper frequency of a rotator does not exist (Gearhart 2010). In 1916, Karl Schwarzschild considered the possibility that rotation combines with the regular precession of the axis of symmetry, thus lifting the degrees of freedom to quantize up to two (Schwarzschild 1916). In the same year, Epstein applied Schwarzschild's theory to Debye-Sommerfeld's model of H_2 , and realized that it did not lead to a good agreement with Eucken's experiments. To retrieve Ehrenfest's formula for H_2 , Epstein had to complicate further the problem, using advanced tools of rotational dynamics, applied to three degrees of freedom, instead of two.

experimental value obtained by Arnold Eucken (1912). Differently from Epstein, however, Krüger did not quantize the motion.

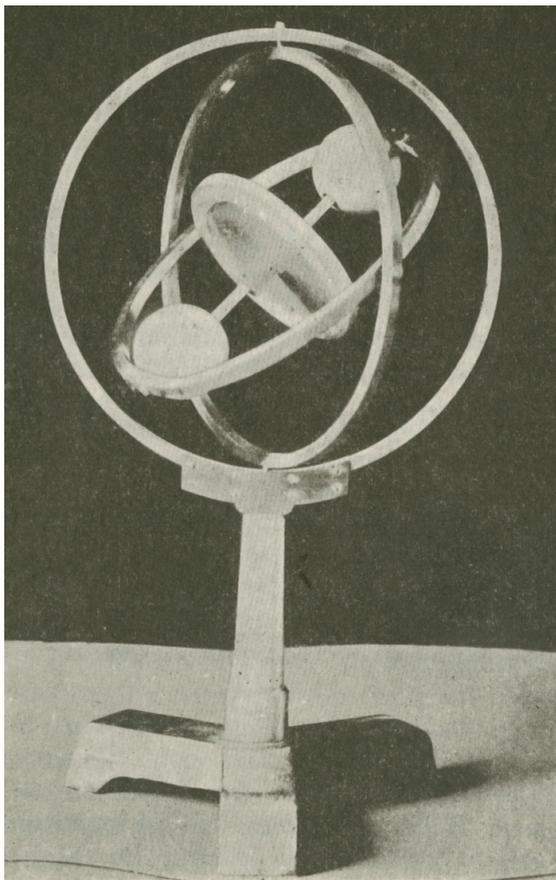


Figure 2.3: The gyrosopic model of a molecular hydrogen (Krüger 1916, 350).

In 1917, Rubinowicz imported the idea of treating Debye-Sommerfeld's model of H_2 as a gyroscope back into optics (Rubinowicz 1917). He calculated the proper motions of the molecule, taking into account both oscillations of the electronic orbits and oscillations of the symmetry axis under the influence of very small perturbations. The resulting motions are represented in the pictures below. Further examination led Rubinowicz to realize that indeed none of the precession

frequencies he obtained contributed to optical dispersion. However, the oscillation corresponding to case C_2 , see fig. (2.4), had the same proper frequency as Krüger's precession frequency in the infrared region of the spectrum. This was very important. Since this approach was a generalization of Debye's and Sommerfeld's theory, Rubinowicz's results provided a bridge between optics and research on specific heats. Optical frequencies were related to the oscillations of orbits, and the frequencies used for specific heats were related to the precession of the symmetry axis. In none of these cases, the exchange of energy took place through quantum transitions, but depended on mechanical motions of molecules, be either vibrations or precessions.

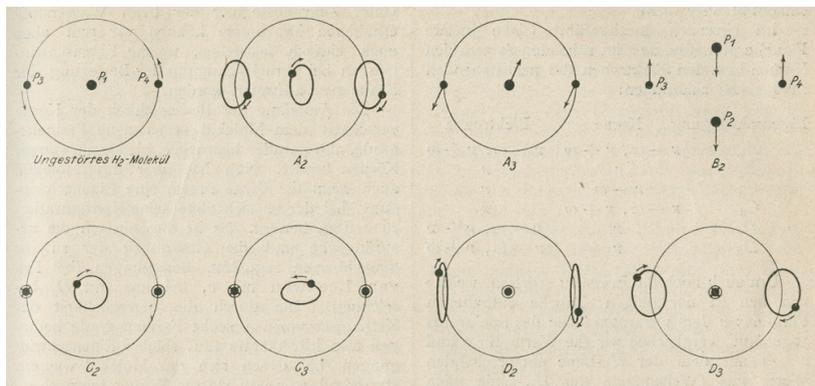


Figure 2.4: Perturbations of the diatomic molecule according to Rubinowicz (1917, 1913).

2.5.2 The Stability Problem

Despite these apparent benefits, the hybrid approach was plagued with difficulties. Ironically, hybridization led to serious problems with stability. The irony lay in that although Bohr's postulates were meant to overcome the stability problems of early atomic models (and were to some degree successful), the embedding of *Mitschwingungen* into the new framework, even as small perturbations, brought back the old problems of mechanical instability, as van Leeuwen explained in 1916 (van Leeuwen 1916).

Her detailed analysis of the hybrid theory of H₂ revealed that only three of the six possible proper motions of orbits, see fig. (2.5), had nonzero electric moments and were therefore relevant for dispersion.

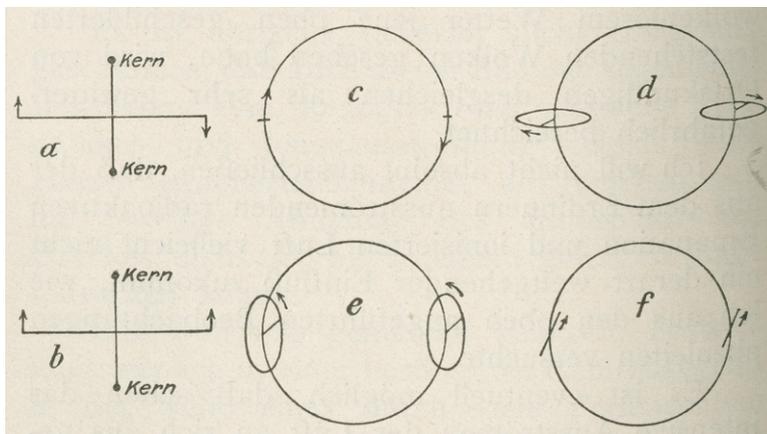


Figure 2.5: The proper motions of orbits of H_2 according to van Leeuwen (1916, 196).

Furthermore, it turned out that only one of the remaining three could actually contribute to optical dispersion. For only in the case of motion f would the orbit keep vibrating after the interruption of the primary radiation and thus be able to emit a secondary radiation. Finally, van Leeuwen determined that the case f was mechanically unstable, in such a way that the molecule would collapse after interacting with light.

To avoid this dead end, van Leeuwen analytically manipulated Debye-Sommerfeld perturbations, transforming them into mechanically stable oscillations. However, the resulting dispersion formula turned out to be in disagreement with experiments.

Under these circumstances, the only possible solution that van Leeuwen could envision was to overcome the mechanical instability of the molecule by assuming Bohr's quantum postulates. However, Bohr's model did not offer the possibility of a consistent treatment of these perturbations of orbits. Thus van Leeuwen concluded that "new assumptions on the equations of motion of [Bohr's] systems are required and as long as this is not the case, nothing can be said about the corresponding dispersion" (van Leeuwen 1916, 198).

2.5.3 Against the Foundations of the Hybrid Theory

Oseen's and Bohr's criticisms went further than van Leeuwen's. In 1915, Oseen, professor at the University of Uppsala, undertook a general analysis of the

compatibility of classical electrodynamics with Bohr's postulates. He followed a "purely logical" argument, independent of any concrete atomic or molecular model representing the problems of mechanical stability. Oseen concluded that between the quantum theory and the classical electron theory no conciliation was possible. One had to choose between them (Oseen 1915, 405).

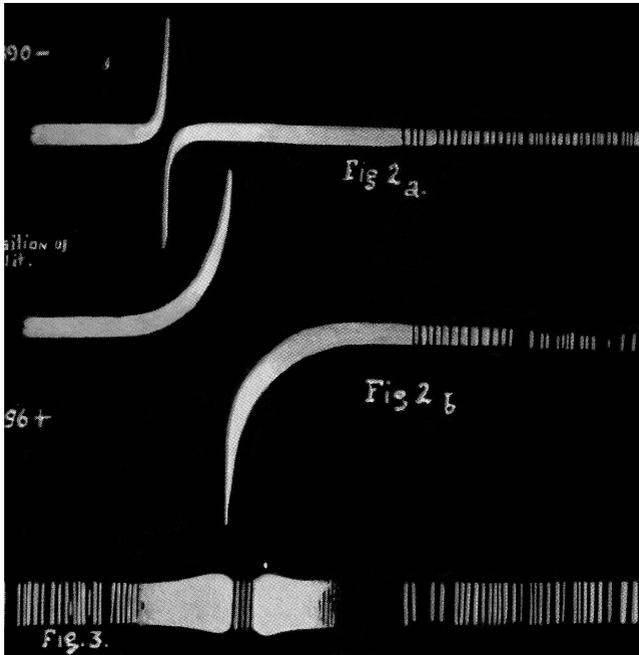


Figure 2.6: Anomalous dispersion in sodium vapor observed using a crossed-prism (Wood 1904, ii).

As mentioned earlier, Bohr had been extremely vague about the connection between the mechanical perturbations of his atomic model and the cause of optical dispersion in 1913, but in 1915 he fully agreed with Oseen. In a letter written 20 December 1915 to Oseen, Bohr expressed his agreement and very effectively summarized the problems of optical dispersion in relation to the available experimental data (*italics added by the author*):

It seems to me that if the theory of the Hydrogen atom has but the slightest connection with truth, the dispersion (at least in gases)

must be a phenomenon of quite a different nature from that assumed by Debye and Sommerfeld. In fact, it appears, e.g., from Wood's and Bevan's experiments on the dispersion in sodium and potassium vapors that the characteristic frequencies which determine the dispersion coincide with the frequencies of the principal series in the Sodium and Potassium spectra, *and one must therefore expect that the same thing holds for other gases.* (Hoyer 1981, 337–338)

Bohr's mentioning of the experiments of Wood and Bevan deserves some comment. For sodium and potassium vapors, the proper frequencies of optical dispersion were typically located in the visible range of the spectrum. This characteristic made it possible to directly observe the proper frequencies of these substances and compare them with spectral data of these substances. For both substances, it was easy to identify the dispersion frequencies with the spectral frequencies, as shown in this picture obtained by Wood in 1904, see fig. (2.6).

However, Sommerfeld and Debye employed experimental data on molecular gases, specifically H_2 . In this case, no proper frequency was identifiable in the visible spectrum; instead it was located in the ultraviolet. Hence in the hybrid theory, Sommerfeld implicitly assumed that spectral frequencies in the optical range were caused by a quantum process; whereas, the proper frequencies in the ultraviolet stemmed from a classical interaction with light. Bohr, instead, argued that all proper frequencies for the whole spectra had to be caused by the same quantum mechanism, and therefore the program of Debye and Sommerfeld was doomed to fail.

Oseen was pleased to learn that Bohr shared his opinion. Indeed, he replied to Bohr on 3 March 1916 adding a somewhat more personal evaluation of the reactions of Debye and Sommerfeld to his criticisms:

The two gentlemen have received my criticism in quite different ways. Debye admitted without reservation that there were internal contradictions in the theory but explained that he saw his mission in groping his way with attempts at hypotheses. Sommerfeld however would maintain that the theory does possess inner consistency. The hydrogen molecules should emit energy during the oscillations of the electron around the radiationless orbit, and the dispersion should originate in this emission. (Hoyer 1981, 570)

Debye seemed to have developed his theory without paying much attention to the compatibility between classical and quantum approaches; thus, he did not defend it when problems emerged. Reacting very differently, Sommerfeld took this question much more seriously in 1915, and he reiterated to Oseen his firm conviction concerning the classical aspects of the problem.

On 19 March, Bohr took up his pen and wrote to Sommerfeld. His letter in no way concealed his critical reservations regarding his German colleague's treatment of dispersion theory:

It appears to me, however, that the experiments on the dispersion in sodium and potassium vapors of Wood and Bevan indicate that the dispersion cannot be determined by means of ordinary mechanics and electrostatics from the constitution of the systems in the stationary states, but must depend essentially on the same mechanism as the transitions between the different states. (Hoyer 1981, 604)

Sommerfeld appreciated Bohr's criticisms but he insisted, in a reply from 20 August 1916, that the structure of the molecule was fixed by quantum postulates, while optical dispersion in molecules was purely classical. On that matter, he had "full scores of still unpublished calculations and results" (Eckert and Märker 2000, 565).

2.6 Sommerfeld's Counterattack

In 1917, Sommerfeld published another paper on optical dispersion, including the "unpublished calculations" that he mentioned to Bohr in the letter from the preceding August in 1916. In this paper, Sommerfeld deepened the divide between classical physics and quantum theory that he put forward in 1915 and subsequently refuted Oseen's and Bohr's claims (Sommerfeld 1917). For these purposes, Sommerfeld introduced a new ingredient into his discussion of dispersion: Ehrenfest's adiabatic principle.

The adiabatic principle allowed Sommerfeld to avoid the previously discussed conceptual contradictions between the treatment of discrete spectra and the treatment of optical dispersion. He achieved this in the following manner: an orbit was perturbed by incoming light, but as long as the ratio between its frequency and the orbital velocity of electrons was infinitesimal, all new perturbed states resulting from the initial stationary states could also be considered stationary states, on account of the adiabatic principle. Thus the continuous transition from one adiabatically modified state to another could cause continuous emission of secondary radiation without contradicting the quantum postulates.⁵

Sommerfeld also employed the adiabatic principle to address another important disparity. Quantum transitions involved in the production of spectral lines occurred only in atoms, while the adiabatic perturbation of orbits from which dispersion originated only occurred in molecules. "Is there any contradiction in this

⁵For more information about Ehrenfest's principle, see (Navarro and Pérez 2006; Pérez 2009).

different treatment of atom and molecule?" Sommerfeld asked himself in the paper. Enthusiastically he answered: "we claim: Nein!" (Sommerfeld 1917, 502).

Thus, Sommerfeld's strategy was twofold. First, he used the adiabatic principle to bridge the gaps between quantum emission and classical emission processes involving perturbed stationary states. This in turn established a new division between atomic and molecular processes, which allowed Sommerfeld to deprive Bohr's criticism of its experimental support. Wood's and Bevan's results on sodium and potassium vapors turned out to be irrelevant to the discussion of the hybrid theory.⁶

As in 1915, Sommerfeld looked for consistency between his theory and the available experimental data on H_2 , N_2 and O_2 . As I discussed in sec. (2.4), experiments implied an angular frequency ω that was at odds with the value calculated from the quantum expression $mr^2\omega = \frac{h}{2\pi}$. In 1915, Sommerfeld had tried to explain the discrepancy by appealing to possible asymmetrical configurations of the molecules. In 1917, he went so far as to invent a new quantum rule. He replaced the usual formula with $mr^2\omega = \sqrt{l} \frac{h}{2\pi}$ (l being the valence number). Admittedly, Sommerfeld could not present any "theoretical foundation for this general quantum assumption" (Sommerfeld 1917, 547), although it allowed him to maintain the validity of mechanics and electrodynamics in the inner molecule.⁷

Finally, Sommerfeld evaded the problems of stability of the H_2 molecule raised in 1915 and 1916 by van Leeuwen and Rubinowicz by alluding to the possibility of a still-unknown quantum constraint:

[F]rom the point of view of the usual mechanics, the models are unstable in a peculiar way. [...] One has to conceive a special quantum constraint on the motions in order to avoid the unstable collapse of the model. (Sommerfeld 1917, 547)

In the ensuing years, the discussion on the hybrid theory of optical dispersion reached a point of stagnation. However, Sommerfeld's division between atomic and molecular processes made an impact. In his early monograph on quantum theory, Fritz Reiche presented optical dispersion in a separate chapter devoted

⁶In 1917, the adiabatic principle also provided Sommerfeld with the conceptual tool to extend the hybrid theory to the explanation of the Faraday effect. By means of the adiabatic principle, Sommerfeld could argue that all modified states produced as a result of the presence of the magnetic field were also stationary states.

⁷Sommerfeld's new rule enjoyed very little popularity among the experts. Reiche lamented the "unaccountably strange quantum condition" which was "undoubtedly a most unsatisfactory result" (Reiche 1922, 121). Other physicists avoided it completely. For instance, the Austrian physicist Gerda Laski applied the Debye-Sommerfeld theory to different models of molecules more complex than H_2 , where valence electrons orbited in two rings, instead of one (Laski 1919). This strategy allowed her to relinquish Sommerfeld's strange quantum condition for the angular momentum.

to “phenomena of molecular models,” thus endorsing Sommerfeld’s viewpoint (Reiche 1922, 117–124).

2.7 From the Hybrid Theory to the Light Quantum

Although no alternative theory of optical dispersion threatened to overshadow the hybrid theory, from 1917 to the early 1920s, Sommerfeld’s position changed significantly.

In a letter sent to Bohr on 5 February 1919, Sommerfeld inquired about his viewpoint on the theory of optical dispersion. On this occasion, Sommerfeld did not defend his earlier theory at all costs as he had done in the 1917 paper. He even challenged Bohr to seek a better solution:

I am very excited about your position on dispersion theory. I would be very happy if you found a better approach to it. If you could replace the H_2 model, which is full of contradictions, with something better, I don’t have any objection. (Eckert and Märker 2004, 48)

Some months later, in a letter dated 19 November 1919 to Sommerfeld, Bohr attached a copy of his 1916 unpublished paper in which he unveiled the details of his argument against the hybrid theory.⁸

Once again, Bohr resorted to Wood’s and Bevan’s experiments to reiterate his point about the absence of a disparity between the quantum and classical domains of knowledge. After all, Bohr argued, Debye’s successful account of dispersion in hydrogen was due to a close coincidence, in this particular case, between the characteristic frequencies of the orbits according to ordinary electro-dynamics and those determined by the quantum relations.

Moreover, Bohr highlighted another difficulty of the hybrid theory: Wood’s and Bevan’s experiments revealed that the intensity of optical dispersion changed depending on the proper frequency considered. This behavior could not be explained by Sommerfeld’s assumed proportionality between the intensity of optical dispersion and the number N of the valence electrons, weighted by a factor C related to the anisotropy of the molecule. Bohr put forward an alternative explanation: the different values of K_i relating to different frequencies were somehow connected with the greater tendency of the quantum jumps to occur between successive stationary states than between more distant states.

⁸The reasons why the paper was withdrawn are mentioned in a letter from Bohr to Sommerfeld dated 19 March 1916. There Bohr revealed to Sommerfeld that he had “decided to postpone the publication [of this paper] and consider it all again, in view of all for which your papers [1916 Sommerfeld’s famous papers on the quantum atomic theory] have opened my eyes” (Hoyer 1981, 604).

This connection between dispersion and transition also worked in the opposite direction, suggesting a very interesting analogy to Bohr: the mechanism of transition between different stationary states resembled the mechanism of interaction of light with a classical electrodynamic vibrator: “[i]f the above view is correct [...] we must, on the other hand, assume that this mechanism [of transitions] shows a close analogy to an ordinary electrodynamic vibrator” (Hoyer 1981, 449).

Months later Sommerfeld voiced a very skeptical opinion about the hybrid theory of optical dispersion:

Debye’s apparent success with hydrogen calls us to a challenge. In the meantime, we have realized that this problem is not ripe yet, even if we exclude the case of the resonance between the exciting optical field and the proper frequencies of the atom, as I did in the past. (Sommerfeld 1921, 500)

To support this statement, Sommerfeld made use of Bohr’s own argument: he highlighted the identity between the absorption lines of optical dispersion and the spectral lines in sodium and potassium gases. Eventually, Sommerfeld accepted the quantum nature of optical dispersion.

This new attitude did not imply Sommerfeld’s subscription to Bohr’s program of bridging quantum and classical physics through the correspondence principle. Sommerfeld never concealed his skepticism about the correspondence principle. In a letter dated 11 November 1920, Sommerfeld added a comment on his treatment of the correspondence principle in the second edition of *Atombau und Spektrallinien*:

In the addenda of my book, you can see that I took the pain to formulate your correspondence principle better than I did in the first edition. [...] However I have to admit that the quantum theoretical root of your principle seems to me still awkward, although I also have to acknowledge that in this way an important relation between the quantum theory and the classical electrodynamics is revealed. (Eckert and Märker 2004, 86–87)

An alternative to the “awkward” correspondence principle was developed some years later by Gregor Wentzel and Karl Herzfeld, two of Sommerfeld’s collaborators in Munich. They relinquished the correspondence principle and the wave picture, relying instead on a purely quantum theory of light. To be sure, Sommerfeld had been advocating the quantum theory of light since 1923, shortly after he heard about Arthur Compton’s experiments. Upon his return from a lecture

trip in the United States, on 27 November 1923, Sommerfeld expressed to Max von Laue his conviction that the new theory of light could pave the way for a refutation of the correspondence principle:

Now I really have a stronger and stronger feeling that the wave theory (and the field theory) must be dismissed. Therefore Bohr's correspondence principle seems to me more and more unsatisfactory however indispensable it is. (Eckert and Märker 2004, 156)

In the same vein as these comments, Sommerfeld reiterated his doubts about the correspondence principle on the occasion of the fourth edition of *Atombau und Spektrallinien*: “personally we would like to preserve a greater hope in the magic of the quantum, rather than in considerations on correspondence or stability” (Sommerfeld 1924, 192).

These ideas also entered the daily scientific life of Sommerfeld's Theoretical Physics Institute at the University of Munich. As early as 16 November 1923, in the regularly-scheduled Wednesday colloquium, the discussion hinged on a report about “Light Quantum Hypothesis and Lattice Interference,” presented by Gregor Wentzel.⁹ Indeed, Wentzel published a paper on a similar topic shortly thereafter (Wentzel 1924). In the same year, Herzfeld took up Wentzel's ideas and applied them to optical dispersion. Both optical dispersion and the interference of light had been hitherto the most serious difficulties for the theory of light quanta.

The first step in overcoming these difficulties was to define a phase for light quanta. Wentzel (1924) had postulated that the phase of a light quantum corresponded to the $1/h$ part of its action integral along a path s between two stationary states of Bohr's atom. The phase determined a wave of probability for this light quantum to follow the path s .

Relying on the idea of the quantum phase, Herzfeld developed a theory of optical dispersion in terms of light quanta (Herzfeld 1924). He also postulated the existence of certain stationary *Zwischenbahnen* (intermediate orbits) between Bohrian stationary states. If the frequency defined by the energy difference between the states was ν_i , a light quantum of corresponding frequency was immediately absorbed by the atom or molecule. By contrast, if the frequency ν of the incoming light quantum did not coincide with ν_i , then the light quantum stayed for a duration of $1/\nu$ on the *Zwischenbahnen* before being reemitted. Drawing on formal similarities with the classical *Mitschwingungen*, Herzfeld stated that the delay of $1/\nu$ caused a phase delay, according to Wentzel's definition of quantum phase.

⁹See the register volume for “Münchener physikalisches Mittwochs-Colloquium,” 1 October 1923. See also AHQP, M/f No. 20, Sect. 001–017.

Thus, with the work of Herzfeld and Wentzel, all vestiges of the old classical mechanism disappeared. Herzfeld's approach led to a physical system in which the only mechanism of energy exchange between light and matter boiled down to transitions between states. The intensity as well as the time-dependency of optical dispersion were due to the dependence of the probability of light quanta being absorbed and reemitted on the ratio between their frequency and the frequency of transition.

2.8 A New Thread of the Story

Parallel to these theoretical developments, the quantum aspects of optical dispersion emerged from a very different perspective. Novel insights arose from experimental research by Otto Stern and Max Volmer (1919). They were dealing with another crucial optical phenomenon: fluorescence.

This phenomenon occurred when certain substances transformed the radiative energy they absorbed into secondary light, normally of lower frequency than the original radiation. In particular, Stern and Volmer were concerned with the exponential decrease of intensity of fluorescence, being parametrized with respect to the decay time τ . Classically, one expected that τ was of the same order of magnitude as the mean time between two molecular collisions. However, in a series of experiments using gaseous iodine, Stern and Volmer concluded that the decay time was indeed much shorter than expected, and it was independent of the pressure of the gas, thus the frequency of collisions.

As an alternative, Stern and Volmer attempted a quantum explanation of the fluorescence process. But, how can a time-dependent process as the exponential decrease of intensity be explained using quantum jumps? Stern and Volmer reinterpreted the decay time τ as the mean lifetime of the molecules in one state b . In doing so, they used the analogy Einstein had established in 1916–1917 between the process of spontaneous emission through quantum transitions and the decay of radioactivity governed by a statistical law (Einstein 1917). This decision implied that the only mechanism of light-matter interaction occurring in fluorescence was a quantum mechanism, namely, quantum transitions.

Eventually, to give a time-dependent explanation of the whole process, Stern and Volmer proposed the equivalence of one quantum of energy $h\nu$ with one monochromatic classical resonator of the same frequency. And they made use of the analogy in the following way: if the macroscopic decay time of a gas of n molecules could be identified with n molecular classical resonators having the same decay time, according to quantum physics, the decay time is related to the decrease in number of resonators that jump from state b to state a at each moment, n being the initial number of resonators at the state b . Thus, by applying the

analogy to the classical resonator, the time-dependent molecular process could be translated into a question of the probabilities of quantum transitions.

From this, one could draw further consequences for other phenomena, most importantly for optical dispersion. Stern and Volmer suggested that, with their analogy between a quantum of energy $h\nu$ and a classical resonator of frequency ν in hand, Drude's classical mechanism could easily be translated into a process involving only quantum transitions and rates of transitions. Thus, all proper frequencies of optical dispersion would coincide with transition frequencies, as the experiments appeared to require.

Two years later, experimentalist Rudolf Ladenburg, aware of Stern and Volmer's results, explained the experimental data on the number of "dispersion electrons" by using a similar analogy and also resorting to a probabilistic description of the elementary processes (Ladenburg 1921). Ladenburg identified the mean energy of n classical damped oscillators with the energy emitted by n molecules through quantum transitions. The latter value depended on the number M_i of molecules on the state i , statistical weights g_i and g_j of states i and j , the probability coefficient a_{ji} for spontaneous transition between these states, and the frequency of light absorbed in the jump ν_{ij} . On these grounds, Ladenburg redefined the number N_i of dispersion electrons appearing in the intensity K_i discussed in sec. (2.1.2) as:

$$N_i = M_i \frac{g_j}{g_i} a_{ji} \frac{mc^3}{8\pi^2 e^2 \nu_{ij}^2}. \quad (2.4)$$

Note that now the intensity was neither dependent on the anisotropy of the molecule, as Sommerfeld had suggested, nor proportional to the number of electrons, as Drude had assumed. Rather, the intensity depended on the probability of transition between two quantum states.

In 1923, Ladenburg and Reiche further developed the analogy with a classical resonator and formalized a kind of fictive oscillator, which they called an *Ersatzoszillator* (Ladenburg and Reiche 1923). These oscillators played the same role in the exchange of energy between light and matter as the resonating electrons in classical accounts. With this move, Ladenburg and Reiche achieved two results. First, the time-dependent features of optical dispersion—the delayed re-emission of secondary radiation by matter—were explained according to the classical model of *Mitschwingungen*. Second, resonance occurred necessarily at spectral frequencies and not at mechanical frequencies, as in the Debye-Sommerfeld theory.

It has been well documented in the secondary literature that the introduction of these *Ersatzoszillatoren* eventually led to the "virtual oscillators" of the Bohr-Kramers-Slater (BKS) theory elaborated by Bohr and two collaborators, Hendrik

Kramers and John Slater, in 1924. Although the BKS theory was soon disproved by experiments, some of its essential features remained in later accounts. Also in 1924, Kramers no longer pictured literal virtual oscillators, but resorted to the correspondence principle when elaborating the first full quantum theory of optical dispersion, based on the formal analogy between a quantum system and a system of classical oscillators perturbed by electromagnetic light (Kramers 1924).¹⁰

2.9 Conclusion

The preceding analysis has shown how optical dispersion constituted an ideal arena for discussing whether there was a divide between classical and quantum physics, which physical features had to be considered characteristically classical or quantum, and how this division influenced the creation of a consistent theoretical account.

From 1913 to 1924, the transformation of optical dispersion passed through two distinct phases. In the first phase, in the aftermath of Bohr's atomic model, it aroused a dispute between Sommerfeld and Bohr on the ultimate nature of the phenomenon, whether it was essentially classical or quantum.

In the second phase, from 1920 onward, dispersion was recognized as a quantum phenomenon, and different strategies emerged to deal with it. The two phases of the story were actually methodologically related. The way in which Bohr and Sommerfeld defined the divide between classical and quantum before 1920 had a bearing on the different strategies they chose to deal with dispersion as a quantum process after 1920.

From 1915 to 1919, Sommerfeld set the divide between the classical and quantum domains on the basis of Debye's hybrid theory. On one side lay spectroscopy and atoms, on the other, optical dispersion and molecules. Thus Sommerfeld's strategy was rooted in negotiating the divide between classical and quantum through the classification of phenomena according to the nature of their physical mechanisms.

Bohr's strategy was very different. He did not recognize any disparity between optical dispersion and spectroscopy or, more generally, between classical and quantum laws. Bohr considered optical dispersion a quantum phenomenon. In 1916, he hinted at the analogy between a classical oscillator and a quantum transition, and this analogy allegedly led to combining the continuous features of *Mitschwingungen* with quantum jumps.

After 1920, Sommerfeld and Bohr agreed on the quantum nature of optical dispersion. However, their approaches remained divergent. In Sommerfeld's

¹⁰For a thorough analysis of this development and its role in the emergence of matrix mechanics, see (Duncan and Janssen 2007).

school, the search for a physical mechanism remained very popular. Indeed, Herzfeld and Wentzel did not confine themselves to the formal analogy between the phase delay of the *Mitschwingungen* and the sojourn time of the light quantum in an atom, but they looked for a completely new mechanism of optical dispersion based on the concept of light quanta.¹¹ In contrast, in 1924, Kramers developed the first quantum theory of optical dispersion using Bohr's correspondence principle and a formal analogy between the quantum system and a system of classical electrodynamic oscillators perturbed by electromagnetic light, without providing any new explicit mechanism of interaction between light and matter.

Abbreviations and Archives

AHQP	Archive for History of Quantum Physics. American Philosophical Society, Philadelphia
------	---

Acknowledgements

This paper has been written in the framework of the project “History and Foundations of Quantum Physics” of the Max Planck Institute for the History of Science (MPIWG) and the Fritz Haber Institute of the Max Planck Society. It is also part of my ongoing PhD project, which I am carrying out under the joint supervision of Jürgen Renn (MPIWG) and Luis Navarro (University of Barcelona). My research has been financially supported with the pre-doctoral fellowships of the Generalitat of Catalunya and the Max Planck Society. I also would like to thank the editors of this volume, Shaul Katzir, Christoph Lehner, and Jürgen Renn, Alexei Kojevnikov, and my colleagues of the “History and Foundations of Quantum Physics Project” for their constructive comments on my paper's drafts. I am very much indebted to Jeremiah James for his linguistic advice. Finally, I would also like to express my gratitude to Michel Janssen for having introduced me to the fascinating research subject of optical dispersion, and very specifically to Massimiliano Badino for the innumerable discussions, and above all, for his unconditional support.

¹¹This description of the Sommerfeld school fits with the characterization in terms of “physics of problems” recently given in (Seth 2010).

References

- Abegg, Richard (1904). Die Valenz und das periodische System. Versuch einer Theorie der Molekularverbindungen. *Zeitschrift für anorganische Chemie* 39: 330–380.
- Bevan, P. V. (1910). Dispersion of Light by Potassium Vapour. *Proceedings of the Royal Society A* 84: 209–225.
- Bohr, Niels (1913a). On the Constitution of Atoms and Molecules. Part I. *Philosophical Magazine* 26: 1–25.
- (1913b). On the Constitution of Atoms and Molecules. Part II. *Philosophical Magazine* 26: 476–502.
- (1913c). On the Constitution of Atoms and Molecules. Part III. *Philosophical Magazine* 26: 857–875.
- Buchwald, Jed (1985). *From Maxwell to Microphysics*. Chicago: The University of Chicago Press.
- Christiansen, Christian (1870). Über die Brechungsverhältnisse einer weingeistigen Lösung des Fuchsin. *Annalen der Physik* 217: 479–480.
- Cuthbertson, Clive and Maude Cuthbertson (1910). On the Refraction and Dispersion of Air, Oxygen, Nitrogen, and Hydrogen, and Their Relations. *Proceedings of the Royal Society A* 83: 151–171.
- Debye, Peter (1915). Die Konstitution des Wasserstoff-Moleküls. *Sitzungsberichte der Bayerischen Akademie zu München*: 1–26.
- Drude, Paul (1894). *Physik des Aethers auf elektromagnetischer Grundlage*. Stuttgart: Ferdinand Enke.
- (1900). *Lehrbuch der Optik*. Stuttgart: Hirzel.
- (1904a). Optische Eigenschaften und Elektronentheorie. I. Teil. *Annalen der Physik* 319(9): 677–725.
- (1904b). Optische Eigenschaften und Elektronentheorie. II. Teil. *Annalen der Physik* 319(10): 936–961.
- Duncan, Anthony and Michel Janssen (2007). On the Verge of *Umdeutung* in Minnesota: Van Vleck and the Correspondence Principle. Part One and Two. *Archive for History of Exact Sciences* 61: 553–671.
- Eckert, Michael (1993). *Die Atomphysiker. Eine Geschichte der theoretischen Physik am Beispiel der Sommerfeldschule*. Braunschweig: Vieweg.
- Eckert, Michael and Karl Märker, eds. (2000). *Arnold Sommerfeld: Wissenschaftlicher Briefwechsel, Band 1: 1892–1918*. Berlin: Verlag für die Geschichte der Naturwissenschaften und der Technik.
- eds. (2004). *Arnold Sommerfeld: Wissenschaftlicher Briefwechsel, Band 2: 1919–1951*. Berlin: Verlag für die Geschichte der Naturwissenschaften und der Technik.

- Ehrenfest, Paul (1913). Bemerkung betreffs der spezifischen Wärme zweiatomiger Gase. *Verhandlungen der Deutschen Physikalischen Gesellschaft* 15: 451–457.
- Einstein, Albert (1917). Zur Quantentheorie der Strahlung. *Physikalische Zeitschrift* 18: 121–128.
- Einstein, Albert and Otto Stern (1913). Einige Argumente für die Annahme einer molekularen Agitation beim absoluten Nullpunkt. *Annalen der Physik* 345: 551–560.
- Epstein, Paul S. (1916). Über die spezifische Wärme des Wasserstoffs. *Verhandlungen der Deutschen Physikalischen Gesellschaft* 18: 398–413.
- Eucken, Arnold (1912). Die Molekularwärme des Wasserstoffs bei tiefen Temperaturen. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*: 141–151.
- Franck, James and Robert Wood (1911). Über die Überführung des Resonanzspektrums der Jodfluoreszenz in ein Bandspektrum durch Zumischung mit Helium. *Physikalische Zeitschrift* 12: 81–83.
- Gearhart, Clayton A. (2010). ‘Astonishing Successes’ and ‘Bitter Disappointment’: The Specific Heat of Hydrogen in Quantum Theory. *Archive for History of Exact Sciences* 64: 113–202.
- Glazebrook, Richard (1886). Report on Optical Theories. In: *British Association for the Advancement of Science. Report–1885*. London: Spottiswoode. 157–261.
- von Helmholtz, Hermann (1875). Zur Theorie der anomalen Dispersion. *Annalen der Physik* 230: 582–596.
- (1892). Elektromagnetische Theorie der Farbenzerstreuung. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*: 1093–1109.
- Herzfeld, Karl (1924). Versuch einer quantenhaften Deutung der Dispersion. *Zeitschrift für Physik* 23: 341–360.
- Hoyer, Ulrich, ed. (1981). *Niels Bohr Collected Works. Volume 2. Work on Atomic Physics (1912–1917)*. Amsterdam: North-Holland Publishing Company.
- Kauffmann, Hugo (1902). *Die Valenzlehre. Ein Lehr- und Handbuch für Chemiker und Physiker*. Stuttgart: Ferdinand Enke.
- Kaufmann, Walter (1902). Die elektromagnetische Masse des Elektrons. *Physikalische Zeitschrift* 4: 54–57.
- (1906). Über die Konstitution des Elektrons. *Annalen der Physik* 324: 487–553.
- Klein, Felix and Arnold Sommerfeld (1898). *Über die Theorie des Kreisels*. Stuttgart: B. G. Teubner.

- Koch, John (1913). Über die Dispersion gasförmiger Körper im ultravioletten Spektrum. *Arkiv för Matematik, Astronomi och Fysik* 8: 1–25.
- Kramers, Hendrik A. (1924). The Quantum Theory of Dispersion. *Nature* 114: 310–311.
- Krüger, Friederick (1916). Über die Eigenschwingungen freier Kreisel-moleküle. *Annalen der Physik* 355: 346–368.
- Kundt, August (1870). Über die anomale Dispersion der Körper mit Oberflächenfarben. *Annalen der Physik* 218: 346–368.
- (1871a). Über anomale Dispersion. *Annalen der Physik* 219: 259–269.
- (1871b). Über anomale Dispersion, Dritte Mittheilung. *Annalen der Physik* 220: 128–137.
- (1872). Über anomale Dispersion, Vierte Mittheilung. *Annalen der Physik* 221: 67–80.
- Ladenburg, Rudolf (1921). Die quantentheoretische Deutung der Zahl der Dispersionselektronen. *Zeitschrift für Physik* 4: 451–468.
- Ladenburg, Rudolf and Stanisław Loria (1908). Über die Dispersion des leuchtenden Wasserstoffs. *Physikalische Zeitschrift* 9: 875–878.
- Ladenburg, Rudolf and Fritz Reiche (1923). Absorption, Zerstreuung und Dispersion in der Bohrschen Atomtheorie. *Die Naturwissenschaften* 11: 584–598.
- Laski, Gerda (1919). Ultrarote Eigenfrequenzen zweiatomiger Bohrscher Gasmoleküle und die spezifische Wärme bei hohen Temperaturen. *Physikalische Zeitschrift* 20: 269–274.
- Laue, Max (1905). Die Fortpflanzung der Strahlung in dispergierenden und absorbierenden Medien. *Annalen der Physik* 323: 523–566.
- van Leeuwen, Hendrika Johanna (1916). Die Instabilität des Bohr-Debyeschen Wasserstoff-Moleküls und die Dispersion von Wasserstoff. *Physikalische Zeitschrift* 17: 196–198.
- Loria, Stanisław (1914). *Die Lichtbrechung in Gasen als physikalisches und chemisches Problem*. Braunschweig: Friedrich Vieweg und Sohn.
- Navarro, Luis and Enric Pérez (2006). Paul Ehrenfest: the Genesis of the Adiabatic Hypothesis, 1911–1914. *Archive for History of Exact Sciences* 60: 209–267.
- Oseen, Carl Wilhelm (1915). Das Bohrsche Atommodell und die Maxwellschen Gleichungen. *Physikalische Zeitschrift* 16: 395–405.
- Pérez, Enric (2009). Ehrenfest's Adiabatic Hypothesis and the Old Quantum Theory, 1916–1918. *Archive for History of Exact Sciences* 63: 81–127.
- Reiche, Fritz (1922). *The Quantum Theory*. London: Meuthen and Co.

- Rubinowicz, Adalbert (1917). Die Eigenschwingungen des Bohr-Debyeschen Wasserstoffmoleküls bei Berücksichtigung der Bewegung der Kerne. *Physikalische Zeitschrift* 18: 187–195.
- Scherrer, Paul (1915). Die Rotationsdispersion des Wasserstoffs. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen* 1: 179–185.
- Schwarzschild, Karl (1916). Zur Quantenhypothese. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*: 548–568.
- Sellmeier, Wolfgang (1872a). Über die durch die Aetherschwingungen erregten Mitschwingungen der Körpertheilchen und deren Rückwirkung auf die ersteren, besonders zur Erklärung der Dispersion und ihrer Anomalien. *Annalen der Physik* 219: 399–421.
- (1872b). Über die durch die Aetherschwingungen erregten Mitschwingungen der Körpertheilchen und deren Rückwirkung auf die ersteren, besonders zur Erklärung der Dispersion und ihrer Anomalien. *Annalen der Physik* 219: 520–549.
- (1872c). Über die durch die Aetherschwingungen erregten Mitschwingungen der Körpertheilchen und deren Rückwirkung auf die ersteren, besonders zur Erklärung der Dispersion und ihrer Anomalien. *Annalen der Physik* 221: 386–408.
- (1872d). Über die durch die Aetherschwingungen erregten Mitschwingungen der Körpertheilchen und deren Rückwirkung auf die ersteren, besonders zur Erklärung der Dispersion und ihrer Anomalien. *Annalen der Physik* 221: 525–544.
- Seth, Suman (2010). *Crafting the Quantum: Arnold Sommerfeld and the Practice of Theory, 1890–1926*. Boston: MIT Press.
- Sommerfeld, Arnold (1907). Ein Einwand gegen die Relativtheorie der Elektrodynamik und seine Beseitigung. An der Diskussion beteiligten sich W. Wien, Braun, Voigt, Des Coudres. *Physikalische Zeitschrift* 8: 841–842.
- (1912). Über die Fortpflanzung des Lichtes in dispergierenden Medien. In: *Festschrift Heinrich Weber zu seinem siebenzigsten Geburtstag am 5. März 1912*. Leipzig: B. G. Teubner. 338–374.
- (1915). Die allgemeine Dispersionsformel nach dem Bohrschen Modell. In: *Arbeiten aus den Gebieten der Physik, Mathematik, Chemie: Festschrift Julius Elster und Hans Geitel zum sechzigsten Geburtstag*. Ed. Karl Bergwitz. Braunschweig: Vieweg und Sohn. 549–584.
- (1916). Zur Quantentheorie der Spektrallinien. *Annalen der Physik* 356: 1–94, 125–167.
- (1917). Die Drudesche Dispersionstheorie vom Standpunkte des Bohrschen Modelles und die Konstitution von H_2 , O_2 und N_2 . *Annalen der Physik* 358: 497–550.

- (1921). Zur Kritik der Bohrschen Theorie der Lichtemission. *Jahrbuch der Radioaktivität und Elektronik* 17: 417–429.
- (1924). *Atombau und Spektrallinien*. 4th ed. Braunschweig: Vieweg.
- Stark, Johannes (1916). Beobachtungen über den zeitlichen Verlauf der Lichtemissionen in Spektralserien. *Annalen der Physik* 354: 731–768.
- Stern, Otto and Max Volmer (1919). Über die Abklingungszeit der Fluoreszenz. *Physikalische Zeitschrift* 20: 183–188.
- Voigt, Woldemar (1899). Über die Änderung der Schwingungsform des Lichtes beim Fortschreiten in einem dispergierenden oder absorbierenden Mittel. *Annalen der Physik* 304: 598–603.
- (1901). Weiteres zur Änderung der Schwingungsform des Lichtes beim Fortschreiten in einem dispergierenden und absorbierenden Mittel. *Annalen der Physik* 309: 209–214.
- Wentzel, Gregor (1924). Zur Quantenoptik. *Zeitschrift für Physik* 22: 193–199.
- Whittaker, Edmund (1910). *A History of the Theories of Aether and Electricity: From the Age of Descartes to the Close of the Nineteenth Century*. London: Longman, Green and Co.
- Wood, Robert (1904). A Quantitative Determination of the Anomalous Dispersion of Sodium Vapour in the Visible and Ultra-Violet Regions. *Philosophical Magazine* 8: 293–324.

Chapter 3

Putting the Quantum to Work: Otto Sackur's Pioneering Exploits in the Quantum Theory of Gases

Massimiliano Badino and Bretislav Friedrich

After its appearance in the context of radiation theory, the quantum hypothesis rapidly diffused into other fields. By 1910, the crisis of classical traditions of physics and chemistry—while taking the quantum into account—became increasingly evident. The First Solvay Conference in 1911 pushed quantum theory to the fore, and many leading physicists responded by embracing the quantum hypothesis as a way to solve outstanding problems in the theory of matter.

Until about 1910, quantum physics had drawn much of its inspiration from two sources. The first was the complex formal machinery connected with Max Planck's theory of radiation and, above all, its close relationship with probabilistic arguments and statistical mechanics. The fledgling 1900–1901 version of this theory hinged on the application of Ludwig Boltzmann's 1877 combinatorial procedure to determine the state of maximum probability for a set of oscillators. In his 1906 book on heat radiation, Planck made the connection with gas theory even tighter. To illustrate the use of the procedure Boltzmann originally developed for an ideal gas, Planck showed how to extend the analysis of the phase space, commonplace among practitioners of statistical mechanics, to electromagnetic oscillators (Planck 1906, 140–148). In doing so, Planck identified a crucial difference between the phase space of the gas molecules and that of oscillators used in quantum theory. Whereas in Boltzmann's statistical mechanics, a state corresponds to an arbitrarily small cell of the phase space, defined by a system's coordinates and momenta, the quantum hypothesis requires a partition of the space into elementary regions—or volume elements—each equal to Planck's constant, h . As we demonstrate below, phase space quantization enabled extending the quantum hypothesis to systems whose complexity exceeded that of oscillators. However, by 1906, Planck did not draw any conclusion about the phase space of gas molecules.

The other source was Albert Einstein's 1907 theory of solids, the first successful application of the quantum hypothesis outside the realm of heat radiation (Einstein 1907). By assuming that the particles of a solid behave like oscillators whose energy is given by Planck's radiation formula, Einstein was able to derive

a temperature dependence of a solid's specific heat that agreed qualitatively with Nernst's theorem (see below). Einstein's theory demonstrated that the quantum hypothesis could be used to solve outstanding problems in the theory of matter. Among its descendants are the nearly-definitive theories of the thermal properties of solids developed by Peter Debye in 1912 and by Max Born and Theodore von Kármán in 1912–1913.

From the close link between the gas and the oscillators mentioned above, one may get the impression that the leap to the quantum theory of a monoatomic gas was a fairly intuitive process. However, this was not the case, because of conceptual difficulties: its advanced formal apparatus notwithstanding, the quantum hypothesis was conceived for periodic systems characterized by a frequency. Viewed as a swarm of whirling particles, a gas was thus understood as the opposite of a paradigmatic quantum system. For instance, in 1911 Nernst resolutely denied the possibility of quantizing translation (Nernst 1911, 267), although he had advocated the extension of quantum theory to gases.¹ Furthermore, there was no experimental motivation for extending the quantum hypothesis to a monoatomic gas. As late as 1913, Arnold Eucken pointed out that no genuine quantum phenomenon had yet been observed involving such a system (Eucken 1914, 396–397).

It is no wonder that in the early twentieth century the quantum theory of a gas was more of a concern to outsiders, such as Otto Sackur, rather than to such experts as Einstein or Planck. The first attempts to develop a quantum theory of a gas thus came about in response to concerns about peripheral issues, namely chemical equilibria, instead of arising from an interest in revamping kinetic theory or thermodynamics.

Indeed, the emergent quantum framework compelled a host of young researchers to approach the problem of a gas with a pragmatic attitude: without an interest or resources to partake in the big foundational debates of the time, they ventured to squeeze an essence that could serve the specific goals of their research from the nascent quantum theory. This pragmatic attitude spurred a variety of applications of the quantum that, to the modern eye, may appear sloppy and naive, but that sometimes led to genuine progress.

Herein, we discuss one remarkable example of such a pragmatic approach, namely Sackur's exploits in the quantum theory of gases. Sackur's work was discussed in previous historical accounts;² in this paper, however, we delve deeper into his theories to uncover aspects of his style of work which were representative for part of the quantum community. A characteristic trait of this style is Sackur's use of quantum theory as a tool to tackle a problem deeply rooted in clas-

¹On this point see (Gearhart 2010).

²Mentions of Sackur's work can be found in (Darrigol 1991; Desalvo 1992).

sical physics, specifically the chemical equilibrium in gaseous systems. Sackur's goal-oriented attitude was partly responsible for some formal and conceptual inaccuracies and contributed to making his papers look tentative. However, his bold attempt to deploy the quantum hypothesis across classical statistical mechanics eventually proved instrumental in preparing Planck's path to the theory of a quantum gas.

Sackur's work developed along two non-orthogonal directions, driven on the one hand by his interest in the Nernst theorem, statistical mechanics and the problem of chemical equilibrium, and on the other, his goal to shed light on classical mechanics from a quantum vantage point. Inspired by the interplay between classical physics and quantum theory, Sackur chanced to expound his personal take on the role of the quantum in the changing landscape of physics. In this paper, we tell the story of this enthusiastic practitioner of the old quantum theory.

3.1 Biographical Overview

Otto Sackur was born in Breslau (now Wrocław), Silesia, on 29 September 1880.³ He studied chemistry first at the University of Breslau. The Chemistry Department, headed by Albert Ladenburg (1842–1911) since its foundation in 1897, was among the most prestigious in Germany. In Breslau, Sackur found an enlightened mentor, Richard Abegg (1869–1910), who introduced him to modern physical chemistry. Sackur further advanced his chemistry education at Heidelberg and Berlin before receiving his doctorate from Breslau on 31 July 1901.

Sackur's academic career at Breslau took a detour, first via the Kaiserliches Gesundheitsamt in Berlin (October 1902 – October 1904), where he worked under the direction of Theodore Paul on problems related to public health. Subsequently, he joined William Ramsey's laboratory at the University College London (October 1904 – March 1905) and then Walther Nernst's new laboratory at Berlin University (March 1905 – September 1905). During the latter stays, he became privy to the most up-to-date work in physical chemistry.

Upon his return to Breslau in October 1905, he obtained his *Habilitation* and the title of *Privatdozent*. For some years, he taught at Breslau and worked side by side with Abegg, with an eye at a more secure position. Sackur's hopes were shattered by two unfortunate events. In 1909, Ladenburg had retired because of poor health (he died two years later) and the new director of the department, the 1907 Nobel Prize winner Eduard Buchner (1860–1917), a fermentation biochemist, had little sympathy for physical chemistry. A year later, Richard Abegg

³There are only few sources available about Sackur's life. Here we have especially relied on (Kipnis 2005) and obituaries written by colleagues and friends (Auerbach 1915; Hertz 1915; Pick 1915).



Figure 3.1: Otto Sackur, 1880–1914.

died tragically in a ballooning accident.⁴ Without an academic sponsor and a laboratory, Sackur had to rely on his pedagogical skills to survive. He accepted minor teaching assignments, devised a course of chemistry for dentists and wrote textbooks on thermodynamics⁵ while trying desperately to keep abreast of the latest developments in physical chemistry. It was during this period of existential difficulties that Sackur launched his research at the intersection of physical chemistry, thermodynamics and quantum theory in the hope of a reward—a more senior academic appointment.

His hopes were fulfilled at the end of 1913, when, thanks in part to Clara Immerwahr, Fritz Haber's first wife and Abegg's former student, Sackur received a call to Haber's Kaiser-Wilhelm-Institut für Physikalische Chemie und Elektrochemie in Berlin. In 1914, he was promoted to the rank of department director. After the outbreak of the Great War, Sackur was enlisted in military research at Haber's institute, but succeeded in carrying on with his experiments on the behavior of gases at low temperatures as a side-project. On 17 December 1914, while working on a military-related project in his laboratory, Sackur was killed by an explosion at his work bench. He was only 34 years old.

3.2 The Problem of Chemical Equilibrium

In the course of the nineteenth century, the concept of a chemical equilibrium underwent several transformations. Around 1850, the old notion that a reaction is at equilibrium when all the “chemical forces” involved are balanced was gradually replaced by a kinetic view: a (reversible) chemical reaction never stops completely, but only reaches a stationary state when the reaction rates in the forward and backward directions become equal.

The ratio between the forward and backward rates was termed the equilibrium constant, K . In 1864, Cato Maximillian Guldberg (1836–1902) and Peter Waage (1833–1900) discovered that the equilibrium constant depended on the ratio of the concentrations of the reactants and products (“law of mass-action”). Their pioneering paper, published originally in Norwegian, was largely ignored until 1877, when Wilhelm Ostwald (1853–1932) adopted the law and corroborated its validity by his own experiments.⁶ Jacobus Henricus van't Hoff (1852–

⁴In Tessin, near Rostock (Arrhenius 1910; Des Coudres 1910).

⁵During this period, Otto Stern earned his doctorate under Sackur's supervision, on osmotic pressure of “generalized soda-water,” a topic of his own choice (Friedrich and Herschbach 2003). Also, it was Sackur who, using Fritz Haber's mediation, helped Stern find his post-doctoral position with Einstein in 1912 (Kuhn 1962). On Sackur's pedagogical activity see (Badino 2013).

⁶A collection of Guldberg and Waage's publications on the topic was published in the series *Ostwald Klassiker der Exakten Wissenschaften* in a translation by Abegg (1899).

1911) discovered the law independently in 1884 and two years later derived a formula that governs the temperature dependence of K (Van't Hoff 1886).

A second breakthrough in the study of chemical equilibria occurred in the 1880s, when thermodynamics was applied to chemical problems. As early as 1878, Josiah Willard Gibbs (1839–1903) used the concept of the maximum work produced by a reaction to define chemical equilibrium: a reaction has reached a stationary state when it produced all the work it was capable of producing (Gibbs 1878). Curiously, Gibbs faced a fate similar to that of Guldberg and Waage's, since the publication of his results in an obscure journal hampered their dissemination. In 1882, Hermann von Helmholtz (1821–1894) arrived at the same results independently. He distinguished between a bound and a free energy of a system. The latter is transformed into work during a chemical reaction, whose equilibrium is reached when the free energy drops to its minimum value (von Helmholtz 1882). As Arnold Eucken paraphrases it:

[T]he (maximum) work represents a quantity whose knowledge leads immediately to the solution of the [equilibrium problem]: only when a chemical process is able to perform work, it will go on spontaneously. Hence, one can also say in what direction [a reaction] develops, if it is given in which transformation [...] the maximum work has a positive sign. If in a chemical transformation no work is produced, then the system is in a state of equilibrium. (Eucken 1922, 123)

The concept of free energy amounted to a well-founded overhaul of the old intuitive notion of “affinity.” The internal energy, U , of a chemical system is made up of free energy, A , that can be used to produce work and of bound energy (also called the “latent heat of reaction”), Q . At equilibrium, the reaction work and the equilibrium constant are related by the equation

$$A = RT \ln K \quad (3.1)$$

with R the universal gas constant and T the absolute temperature.

By combining the first and second laws of thermodynamics for the case of an isothermal and isochoric gaseous reaction, it was possible to obtain the following equation, sometimes called the Helmholtz equation, for the free energy,

$$A = U + T \left(\frac{dA}{dT} \right)_v, \quad (3.2)$$

where V is the volume (Haber 1905, 18–22). The general integral of this equation is:⁷

$$A = -T \int \frac{U}{T^2} dT + Tf(V) \quad (3.3)$$

with $f(V)$ the integration constant (integration at constant volume). Thus, to evaluate the free energy—and, via eq. (3.1), the equilibrium constant—one had to determine, at a given temperature, the reaction energy U and the function $f(V)$. Experimentally, this task was far from easy and much effort was expended in the final decades of the nineteenth century at collecting the requisite data for various substances.⁸

The Heat Theorem, enunciated by Walther Nernst (1864–1941) in 1906, amounted to a third breakthrough.⁹ Nernst observed that as $T \rightarrow 0$, the slopes of the temperature dependence of the free energy and of the internal energy (heat) tend asymptotically to the same limit,¹⁰ without concluding that A and U themselves become equal at $T = 0$ as well. Nevertheless, based on his observation, Nernst was able to set the integration constant $f(V)$ in eq. (3.3) to zero (Nernst 1906a; 1906b) and thereby find at any temperature T the free energy A from U which could, in turn, be determined from thermochemical data.¹¹

However, the experimental data as well as the theoretical arguments that supported Nernst's assumptions could not be extrapolated to a gas at very low temperatures and held true only for condensed-matter systems.¹² The treatment of the equilibrium of gaseous reactions remained beyond the direct reach of the theorem. In response, Nernst devised an ingenious detour which made it possible to evaluate the equilibrium constant of gaseous reactions as well. Nernst noticed that the integration constant, C , of the Van't Hoff equation was closely related to the integration constant of the Clausius–Clapeyron equation which governs the temperature dependence of pressure of a vapor in phase equilibrium with a liquid or solid condensate (the two equations are isomorphic). Nernst dubbed this quantity the chemical constant (Nernst 1906a, 22). The chemical constant had to

⁷For the technique used to solve this equation see (Haber 1905, 22–23).

⁸For theoretical and experimental aspects of physical chemistry at the end of the nineteenth century, see (Hiebert 1971; 1983; Kormos Barkan 1999; Coffey 2006).

⁹Julius Thomsen (1826–1909) in 1852 and Marcellin Berthelot (1827–1907) in 1869 speculated on the behavior of chemical reactions in the vicinity of absolute zero. See (Kormos Barkan 1999; Bartel and Hübener 2007). For Nernst's personal take on the story, see the opening chapter of (Nernst 1917).

¹⁰Namely to zero, i.e.,

$$\lim_{T \rightarrow 0} \frac{dA}{dT} = \lim_{T \rightarrow 0} \frac{dU}{dT} = 0.$$

¹¹For a modern treatment, see (Morse 1969, 143–145).

¹²For a thorough presentation of the issue, see (Nernst 1907; 1917).

be determined experimentally for each gaseous species i involved in the reaction, which was indeed possible thanks to the absolute character of entropy, as established by Nernst's theorem (see below, section 3). Once the C_i 's are available, the equilibrium constant of a gaseous reaction follows:

$$\ln K = \frac{Q}{RT} + \frac{\ln T}{R} \Sigma(c_p)_i + \Sigma C_i, \quad (3.4)$$

where c_p is the heat capacity at constant pressure. To evaluate the chemical constant, Nernst attempted to express it as a function of the thermochemical parameters by resorting to various approximations of the Clausius-Clapeyron equation (Nernst 1907, 55–76). However, by 1911, the chemical constants were available for only a handful of substances. This was chiefly due to the unreliability of the approximations involved and to the great difficulties encountered when performing measurements at cryogenic temperatures. Sackur followed a different strategy. In 1911 he commented:

Only Nernst's theorem makes it possible to establish the constant C which determines the chemical behavior of gases from measurements on *pure* substances (vapor pressure of liquids or solids) and thereby [...] also the entropy constant [S^0]. (Sackur 1911b, 965)

The reference to the entropy constant is key: Sackur recognized that the vapor-condensate system can be dealt with by applying the second law of thermodynamics, which leads to a simple relation between the integration constant of the entropy and the chemical constant:

$$C = \frac{S^0 - c_v + R \ln R}{R}, \quad (3.5)$$

where c_v is the specific heat of a gaseous component at constant volume and S^0 its entropy constant, see eq. (3.10) below.

This relation is at the core of Sackur's 1911 advance: he set out to calculate the entropy constant S^0 —and thus C —by invoking statistical mechanics. This was a bold move not only because of the rather tentative status of the statistical approach at the time but also because entropy itself was perceived as difficult to define and, moreover, wedded to “mysterious” applications of probability. Physical chemists of the time preferred the concept of maximum work to characterize chemical equilibrium. In the most influential physical chemistry textbook of the time, Nernst's *Theoretische Chemie*, entropy comes up only to be brushed away as dispensable. However, Sackur's move was not unrewarded: he was able to derive the first quantum-statistical expression for the entropy of an ideal gas (in

the limit of high temperatures and low densities). A similar expression was discovered independently at about the same time by Hugo Tetrode (1895–1931) and is known as the Sackur-Tetrode equation. The Sackur-Tetrode equation made it possible to evaluate the entropy constant and thus the chemical constant in terms of better-known quantities.

3.3 The Beginnings of the Quantum Theory of Gases

In the first decade of the twentieth century, many proponents of the quantum, such as Nernst, Max Planck (1858–1947), and Albert Einstein (1879–1955), had come to think that kinetic theory, quantum theory, the heat theorem, statistical mechanics and (physical) chemistry were all closely related. Einstein's 1907 quantum theory of solids exemplifies the power of the quantum in treating what used to be known as the kinetic theory of matter, while Nernst, at about the same time, became convinced that the quantum implied the validity of his theorem for gases.¹³ On a more general level, it became apparent that the quantum called for new ways of applying statistical mechanics to both the theory of radiation and the theory of matter.

Presumably, Sackur learned about these developments at the scientific meetings he had attended tirelessly in those years.¹⁴ The approach to the theory of matter based on sophisticated mathematical techniques—which quantum theory entailed—was exactly his “cup of tea.” Unlike the majority of physical chemists of the time who had an aversion to the formal complications of kinetic theory (including major protagonists such as Nernst), Sackur was convinced that physical chemistry could not do without the most advanced mathematical tools available. During his academic career in Breslau, he taught classes in both kinetic theory (summer semesters 1910 and 1912) and in the “Mathematical Treatment of Chemistry” (winter semester 1906 and summer semester 1908). Sackur's lecture notes became the basis for an acclaimed book written jointly with Abegg, which was translated into English (Abegg and Sackur 1909).

¹³Although Nernst was reluctant to commit his conviction to paper, there are at least two pieces of indirect evidence: First, in about 1910 Nernst attempted to prove his theorem on the grounds of a purely thermodynamic argument independent of the nature of the substances involved (Kox 2006); second, after 1906 he put his Berlin group to work on the measurement of the specific heats of gases. Before the Solvay Conference, Nernst supervised three doctoral dissertations on the specific heats of gases—by Frank Voller (1908), Friedrich Keutel (1910), and Robert Thibaut (1910). At the same time, Mathias Pier worked on the measurement of specific heats at high temperatures (Pier 1909; 1910), while Fritz Koref and Eucken concentrated on low temperatures (Koref 1911; Eucken 1912).

¹⁴From *Physikalische Zeitschrift* and *Zeitschrift für Elektrochemie* it can be established that Sackur attended the meetings of the Bunsen Society for Physical Chemistry from 1906 until 1914, the year of his death. In 1908, he also reported about the meeting for *Physikalische Zeitschrift*.

In Sackur's usage, the term "kinetic theory" did not refer to the nineteenth-century analysis of collisions leading to the riddle of irreversibility. Instead, kinetic theory meant an attempt to trace the thermal behavior of systems, chemical or other, to the arrangements of their molecules. This is illustrated by his first paper dedicated to such problems (Sackur 1911a).¹⁵ Sackur argued that with decreasing temperature, the available molecular energy becomes increasingly restricted, limiting the number of "cells" (states) over which the molecules can be distributed. When the temperature drops close to zero, the number of cells approaches one, as a result of which the distribution becomes ordered: all the molecules end up herded in the last remaining cell. Since entropy depends on the number of molecular arrangements, W , via the Boltzmann principle

$$S = k \ln W, \quad (3.6)$$

it was apparent that $S = 0$ for $W = 1$ when all the molecules are in the same cell. Since Nernst's theorem can be expressed as $\lim_{T \rightarrow 0} S = 0$ (Planck 1911, 266–286), Sackur concluded that the third law reduces to the claim that low-temperature molecules are arranged in their energy space in the most orderly fashion.

The main sources of inspiration for Sackur's paper (Sackur 1911a) were precisely those mentioned in the introduction: Planck's 1906 *Wärmestrahlung* as well as Einstein's 1907 theory of solids (Einstein 1907). Sackur was especially intrigued by the way Einstein intertwined the classical statistico-mechanical approach with the new quantum hypothesis. In the second part of the paper, Einstein proposed a second proof of the heat theorem by showing that the quantum hypothesis, in the case of a solid, implied that the entropy constant is zero. For Sackur, the transition from the gas to the solid was justified by Einstein's general procedure:

[A] complete kinetic theory of the solid state of aggregation might perhaps be built on Einstein's presuppositions, just as the van der Waals' theory was built on the presuppositions of the classical gas theory. (Sackur 1911a, 467)

Furthermore, Sackur's paper attests to its author's internalization of Planck's concept of absolute entropy. Classical thermodynamics defines only an entropy difference and leaves the integration constant undetermined. But Planck was convinced that entropy expresses a fundamental property of nature and, therefore,

¹⁵Attempts along a similar direction were (Nernst 1911; Jüttner 1911; Polanyi 1913).

must have an absolute value. Quantum theory combined with the heat theorem affirmed that entropy's integration constant can be set to zero.¹⁶

Sackur adopted the concept of absolute entropy and used it as a building block of his new quantum theory of the ideal gas. In his second 1911 paper, he tried to construct a general application of his peculiar version of kinetic theory to chemical problems. Here "application" expresses the second direction of Sackur's work: his attempt to transform abstract procedures into potent treatments of classical problems. To Sackur, the glue that holds together quantum theory, kinetic theory and physical chemistry is the concept of probability, whose value was well proven in investigations of gases and could be applied to other problems as well:

This idea has been hitherto applied only to the variations of state of an ideal gas, that is to physically and chemically uniform substances, as well as to radiation phenomena. Now it seems that it can be applied to any spontaneously occurring processes in nature, for example to chemical reactions and to irreversible processes of any kind that are connected with an increase of the probability of a closed system. This generalization of the concept of probability to chemically distinct states of matter appears to me as a simple consequence of Boltzmann's view. (Sackur 1911b, 960)

Henceforth, Sackur set out to find a general way of comparing the probability of different states. The case of chemical equilibrium in an ideal gas was particularly suitable for this aim, as it went just a step beyond Boltzmann's gas theory.

In his attack on the problem of entropy from the kinetic side, Sackur explicitly follows Planck's statistico-mechanical treatment of an ideal gas based on the partitioning of the phase space. In classical statistical mechanics, the state of the gas is determined by the number of molecules whose coordinates lie between $q_1, q_1 + dq_1, q_2, q_2 + dq_2$ and $q_3, q_3 + dq_3$, while their velocity components lie between $v_1, v_1 + dv_1, v_2, v_2 + dv_2$, and $v_3, v_3 + dv_3$. Thus, if the state space of a single molecule is divided into "elementary regions" $dq_1 dq_2 dq_3 dv_1 dv_2 dv_3 = d\sigma$, and f is the distribution function of the number of molecules in that state space, the number of molecules in a given "elementary region" is $n = fd\sigma$. Let us assume that the elementary regions $d\sigma$ can be numbered and that $n_i = fd\sigma_i$ is the number of molecules contained in the i -th region. Thus a state distribution is the sequence n_1, n_2, \dots of the occupation numbers of each region. In 1877, Boltzmann proposed a simple recipe to calculate the probability of such a distribution, namely by counting the number of ways in which one can permute the molecules

¹⁶Planck dwelled on the concept of absolute entropy both in the third edition of his *Thermodynamik* and in the second edition of the *Wärmestrahlung* (Planck 1911; 1913).

between different regions without changing the overall distribution. From combinatorics, this number turns out to be:

$$W = \frac{N!}{n_1! n_2! \dots} = \frac{N!}{f d\sigma_1! f d\sigma_2! \dots} \quad (3.7)$$

where N is the number of molecules. By substituting for W in the Boltzmann principle, eq. (3.6), invoking the Stirling formula, setting $d\sigma_i = d\sigma$ for all i , and replacing summation by integration,¹⁷ Sackur arrived at the formula for entropy

$$S = kN \log N - k \int f \log f d\sigma - kN \log d\sigma \quad (3.8)$$

whose maximum (which corresponds to equilibrium) yields the distribution function f . Under the constraints of constant energy and number of molecules, f comes out Maxwellian, and substituted back into eq. (3.8), a formula for the translational entropy of a monoatomic gas results,

$$S = c_V \ln T + R \ln V + c_V \left[1 + \ln \left(\frac{2\pi R}{M} \right) \right] - R \ln d\sigma \quad (3.9)$$

with M the molecular weight. A comparison of eq. (3.9) with that for the entropy of an ideal gas,

$$S = c_V \ln T + R \ln V + S^0 \quad (3.10)$$

yields the entropy constant in terms of the universal gas constant, the molecular weight, and the elementary region $d\sigma$,

$$S^0 = c_V \left(1 + \ln \frac{2\pi R}{M} \right) - R \ln d\sigma. \quad (3.11)$$

So far, the argument only used well-established elements of kinetic theory and statistical mechanics. Among the ingredients contributing to the entropy constant, only the volume of the elementary region could not be specified within the framework of classical physics. Hence a determination of the entropy constant, and therefore of the chemical constant, called for a closer investigation of this notion. So what were Sackur's elementary regions?

¹⁷On one hand, the applicability of the Stirling formula relies on the assumption that the elementary regions are large enough to contain many molecules. On the other hand, replacing summations by integration requires the regions to be infinitesimal. Classical kinetic theory hinges upon the balance between these two contravening requirements, see also (Hoyer 1980; Darrigol 1988). We discuss Sackur's reflections on it in the next section.

3.4 Sackur's "Elementary Regions"

A considerable part of Sackur's paper (Sackur 1911b) deals with the properties of the elementary regions. First, he points out that the volume $d\sigma$ must have a well-defined value because the entropy constant is related to the equilibrium constant, which is indeed a constant for any given reaction. This requirement hinted to Sackur that the issue crosses beyond the boundaries of classical physics.

[S]ince pure kinetic [theory] has hitherto not been able to say anything about the magnitude of $d\sigma$, there is a gap in the sequence of inferences that lead from the kinetic hypothesis to the equation of state of a gas. (Sackur 1911b, 968)

Sackur, however, did not immediately suggest that the fixed volume of an elementary region is determined by Planck's constant, h . There might be a purely technical reason for his reluctance to do so. Following Boltzmann and Planck, Sackur worked with the position-velocity space whose volume element did not have the dimension of action. In fact, Sackur carried out the dimensional analysis of the entropy constant, but he limited himself to verifying that it depended on the volume of the gas in the expected way. Although noncommittal as to the physical meaning of $d\sigma$, the peculiarity of a fixed-magnitude elementary volume did not escape Sackur's attention. He tried to give both a visual and a conceptual interpretation of this result.

Visually, the fixed-magnitude volume seems to entail that molecules do not pass smoothly from one state to another, or equivalently, that they do not spread out into the entire space allowed.¹⁸

The gas molecules do not distribute themselves uniformly over the whole allowed space and they cannot have all possible velocities from zero to infinity, but [...] they concentrate around individual points in space, like bullets on a target, and [...] their velocity components change *in jumps*. (Sackur 1911b, 969)

Here we encounter a characteristic trait of Sackur's approach. Drawing on the "visualizability" of a discrete space, Sackur initially presumes a discontinuous behavior of the molecules. But in reality, his underlying picture remained a provisional sketch. As we show in the next section, he eventually dropped discontinuity and moved toward an operational approach which sidestepped precise commitments about the microscopic nature of gas particles. This was due not only to

¹⁸To a modern reader, Sackur's talk about a tendency of molecules to clump in phase regions may suggest some, albeit naive, anticipation of quantum statistics (Darrigol 1991). However, the main problem that quantum statistics had to deal with, namely the extensivity of entropy, was not considered. And when the problem came up, Sackur did not seem to recognize it as a fundamental issue.

the objective difficulties of framing a clear picture of the quantum phenomena, but went hand in hand with the problem-oriented character of his program. The motivation behind his research was the idea of putting the quantum in the service of classical problems—and not developing an understanding of the quantum's nature.

Conceptually, Sackur relied on Planck and traced the “atomicity” of the elementary region back to the calculation of probability. In his 1900 paper in which he enunciated the quantum hypothesis, Planck argued that an actual computation of a state's probability is possible only if the total energy is divided into elements of a finite size. The atomicity of energy then followed directly from the Boltzmann relation between entropy and probability, eq. (3.6). Sackur subscribed to Planck's view unconditionally:

[E]ither [we] postulate the physical reality of a finite elementary region (and of finite elementary quanta) or [we] give up the *unequivocal* relation between entropy and probability. (Sackur 1911b, 970)

Sackur's prime interest, however, lay in the physico-chemical properties of the elementary regions $d\sigma$. Coming to terms with what $d\sigma$ means is the theme of the second part of his analysis, whose conclusion is twofold: First, $d\sigma$ is not a universal constant but must somehow depend on the molecular mass. By combining eqs. (3.5) and (3.11), Sackur obtained

$$C \propto \frac{c_V}{R} \ln \frac{2\pi R}{M} + \ln R - \ln d\sigma \quad (3.12)$$

and argued that if $d\sigma$ were the same for all gases with the same number of molecules (as c_V is), then the chemical constant C would have to decrease with increasing molecular weight. This, however, would contradict Nernst's experiments, which had found that C in fact increases with M .

Second, Sackur concluded that $d\sigma$ depends on the number of molecules. By making use of an ingenious argument,¹⁹ Sackur contended that the probability of an ideal gas consisting of N molecules partitioned into q subsystems with N' molecules each and probability $W_{N'}$ has a total probability $W_N = (W_{N'})^q$. An application of Boltzmann's principle then led Sackur to the conclusion that the ratio $d\sigma/N$ —and not $d\sigma$ alone—is a universal constant. Sackur's result contains a grain of truth in that it renders the entropy in his formula extensive.²⁰ In his

¹⁹From this argument, one may discern another source of inspiration in Sackur's work, for it closely resembles the 1905 light-quantum paper, in which Einstein compares the probabilities for a gas to be in a given volume and in a part of that same volume.

²⁰If $S_1 = k \ln W_1$ and $S_2 = k \ln W_2$ are the entropies of two subsystems, and $W_{12} = W_1 W_2$ is the probability of the composed system, then $S_{12} = S_1 + S_2$.

Wärmestrahlung, Planck started from extensivity and the law of the composition of probabilities to derive the Boltzmann principle (Planck 1906, 136–137). Apparently Sackur did not fully realize the significance of his own argument and result, but Planck did. Three years later, while struggling with the problem of making quantum entropy extensive, he would take up Sackur's condition by stating that the elementary volume of the phase space of a gas depends on the number of molecules in the gas (Planck 1914, 9).

Sackur's paper is representative of the general concerns of his work. As stated in the title, Sackur's prime interest was the application of the methods developed by Planck in radiation theory to physical chemistry. Sackur followed Planck closely in the usage of probability and in handling the state space, but stopped short of explicitly introducing the quantum hypothesis. Instead, he attempted to clarify the properties of the elementary volume as a means to calculate theoretically the chemical constants. In the concluding section of the paper, he tested his procedure on the dissociation of iodine molecules and was able to retrieve the correct formula for the mass-action law.

3.5 The Generalization of Sackur's Statistical Theory

Sackur's focus on the burning issues of physical chemistry distracted his attention from the impact his method could have on quantum theory at large. Tetrode, whose work approached the problems of fledgling quantum statistics from a more abstract viewpoint, stressed the quantum interpretation of the elementary volume as well as the problem of extensivity right at the outset (Tetrode 1912).²¹ However, Sackur himself realized, in careful reconsideration, that his procedure suggested an intriguing perspective. He explored this viewpoint in a paper included in the Nernst *Festschrift*, published in May 1912.²² In that paper, the notion of the quantum of action and its application to the calculation of chemical constants appeared in the very title.

Sackur first tackled the problem of extensivity. He realized that the dependence of $d\sigma$ on the number of molecules was necessary to guarantee extensivity of the entropy function, but was no longer happy with the way it came about in his previous paper. At this point, Sackur was already convinced that the essence of the quantum hypothesis was probabilistic. As a consequence, he thought that a reconciliation of quantum theory with thermodynamics relied on a suitable definition of probability. To appreciate his procedure, we briefly summarize Sackur's peculiar concept of probability.

²¹For a discussion of Tetrode's paper, see (Darrigol 1991; Desalvo 1992).

²²The paper (Sackur 1912a) is dated March 1912. Since Tetrode's article was also published in March, it seems that Sackur worked out the consequences of his approach independently.

In 1911, Sackur defined the probability of a state as “the number that measures how many times a [...] state [...] is more probable than a state of the same energy and volume in which all atoms would have the same positions, directions and velocities” (Sackur 1911b, 959–960). This idiosyncratic definition, a result of Sackur’s personal reading of Planck and Boltzmann, led to the usual number of possible permutations of a state distribution. To retrieve extensivity, Sackur stated that the definition must be changed, and “we must resort [...] to the concept of molecular disorder” (Sackur 1912a, 406). Once again, Sackur obtained this concept from the tradition of Planck and Boltzmann, but reinterpreted it in a peculiar way. In an arbitrary state, it is possible that more than one molecule occupies the same energy cell (or phase-space cell). If the state is perfectly disordered, each cell contains exactly one molecule. In other words, the molecules are spread throughout the available energy (or phase) space. This concept of molecular disorder could not be more distant from Planck’s and Boltzmann’s,²³ but it accomplished its intended mission. As a result, Sackur modified the definition of probability as the ratio of the number of favorable cases to the number of possible cases, a proper fraction, and was again able to retrieve the term $d\sigma/N$ in the entropy formula, which guarantees its extensivity. This is yet another example of Sackur’s willingness to quite freely interpret the traditions of statistical mechanics to arrive at a concrete result.

The second theoretical innovation introduced in the 1911 paper is the interpretation of the elementary volume in terms of Planck’s constant. Sackur kept the position-velocity space, but relied on Arnold Sommerfeld’s authority for interpreting the volume as $d\sigma = h^3/m^3$.²⁴ This expression for the elementary volume enabled researchers “to calculate this quantity [i.e., $d\sigma$] characteristic for the determination of the vapor pressure and of the chemical behavior without any experimental determination” (Sackur 1912a, 409).

The last part of the paper deals with the calculation of the chemical constants for mono- and polyatomic gases. Sackur displays his familiarity with the experimental data and his ability to work with them. Despite an incorrect expression for the rotational energy, Sackur obtained a reassuring agreement between his theoretical values and the experimental chemical constants, which suggested that his formula was correct.

In a subsequent paper (Sackur 1912b), Sackur’s shift toward an operational approach was complete; this paper was the most ambitious of his four articles

²³Boltzmann considered molecular disorder to be the condition for applying probability in gas theory, while Planck was convinced that disorder had the power to eliminate any violation of the second law of thermodynamics. For a discussion of these two views, see (Badino 2009).

²⁴Sommerfeld proposed the interpretation of the elementary volume in terms of h as a general quantization condition for a periodic system at the First Solvay Conference and in a widely discussed paper (Sommerfeld 1911).

on gaseous equilibria and statistical mechanics. He had apparently not believed that his prior probabilistic argument was satisfactory and thus sought to present a more general analysis. The striking feature of Sackur's original approach based on the partitioning of the state space was that it resolved, with facility, the problem of the quantization of an aperiodic system. As we mention in the introduction, the quantum hypothesis was introduced to deal with periodic systems and expressed a relationship between the system's energy and characteristic frequency. But there was no obvious way of ascribing a frequency to a gas. To circumvent this issue, Sackur worked on the state space of the system rather than the system itself—quantizing elementary regions instead of the system's behavior. However, in his 1912 paper (Sackur 1912b) he changed tack. In accordance with his view that the quantum was intimately connected with probability, he endeavored to bring together the quantum hypothesis and classical probabilistic procedures used by kinetic theorists of the time. The result was a captivating argument with a nineteenth-century flavor.

The argument went as follows. Sackur first considers the case of a periodic system—a set of particles oscillating around an equilibrium point—which resembles an Einstein solid. There is no assumption about the nature of these particles except that, after a very long time, they take up all possible values of energy. Sackur then tackles the problem of calculating the probability that a particle has an energy between ϵ and $\epsilon + \Delta\epsilon$ when it passes through the equilibrium point.²⁵ Then he envisions a microscopic observer who can watch individual particles and send a signal every time a particle passes through the equilibrium point with the correct energy. As a result, the sought probability is given by the ratio of the number of signals, n , to the total number of particles, N , that is, $w = n/N$. This probability, Sackur states, is a function of energy and it is proportional to the width of the cell as well as to the span of time over which the observer watches the system, $w = f(\epsilon)\Delta\epsilon\Delta t$. Given that any value of energy is possible, Sackur insisted that it is a fundamental principle of probability calculus that if one waits long enough, all possible events will occur. This principle implies that if $\Delta\epsilon$ decreases, that is, the energy cells become smaller and more difficult to observe, the same probability can be maintained by extending the observation time Δt . In other words, $\Delta\epsilon\Delta t = \text{constant}$, a consequence that curiously resembles Heisenberg's energy-time uncertainty relation.²⁶

But if the observer is allowed to examine the system for an arbitrarily long time, he will count all the molecules infinitely many times, because they can as-

²⁵The argument could be carried out for any point of oscillation, but Sackur chooses the equilibrium point since there the energy of the particle is purely kinetic.

²⁶The energy-time uncertainty relation has a more genuine predecessor in the work of Niels Bohr (1923, 150).

sume all energy values over a long period. To ensure that each molecule is counted only once, the observation time must be set equal to the oscillation period, $\Delta t = \tau$, in which case only molecules with energies $0, \Delta\epsilon, 2\Delta\epsilon, \dots$ will be counted when $\Delta\epsilon = \frac{h}{\tau} = h\nu$. The quantum hypothesis thereby affects the energy width $\Delta\epsilon$, that is to say the “experimental error” that the microscopic observer is allowed to make. The gist of Sackur’s probabilistic argument is that the quantum hypothesis is not necessarily a new physical assumption, but rather an “extension of the old statistical methods”:

Contrary to the old picture developed by Planck for the derivation [of the equations of this theory], it [is] not necessary to assume an atomistic structure of energy or action. It suffice[s] to sharpen the (physical) concept of probability, namely by the almost obvious assumption that the verification of a result is the more probable, the longer one waits for it, and therefore even an extremely unlikely but possible result will have a finite probability to arise after an infinitely long time. (Sackur 1912b, 85)

Sackur further strived to show that the above quantum constraint emerges as soon as the probabilistic formalism is transformed into a physically workable procedure. The fictitious microscopic observer illustrates precisely this point. His attempt to demonstrate a continuity between the methods of classical statistical mechanics and quantum theory serves two related goals. First, it allows researchers to put aside some of the conundrums that came along with the quantum hypothesis. As the quotation above illustrates, Sackur was more eager to underscore the *formal similarities* between the classical and quantum approach than the weird consequences stemming from discontinuity and discretization. Second, the above similarities justified in Sackur’s eyes the application of the quantum procedure to the problems left open by classical theory. Precisely because the quantum hypothesis appeared to him as an extension of statistical mechanics, he expected it to fill the remaining gaps in the thermodynamic and statistico-mechanical treatments of chemical phenomena.

Sackur then tests his probabilistic argument by retrieving Planck’s black-body radiation law and Einstein’s formula for the heat capacity of solids at low temperatures. In this connection, Sackur makes an interesting comment, namely that in the limit of high temperatures, one obtains the Rayleigh-Jeans formula and the Dulong-Petit law for the two above cases, respectively. Then he points out for the case of the Rayleigh-Jeans formula that it is to be regarded not as a purely classical result but rather an approximate quantum result.

So much for periodic systems. For aperiodic systems, the maximization of entropy proceeds formally in the same way as for periodic ones, except that one

has to identify a proper substitute for the characteristic frequency ν . Sackur argues that the limitation imposed by the observation time must be such that the velocity of the molecule be constant during that time: “we have to choose the observation time [...] so that during $\Delta t = \tau$ the molecule experiences no change of velocity, that is no collision” (Sackur 1912b, 75–76). Sackur is apparently referring to the mean time that the molecule spends to cover the mean free path. In that span of time, the molecules maintain their velocity and the observer will count only those molecules which are in a given energy cell.²⁷

To ensure extensivity, Sackur resorted to defining the three components of velocity as the ratio between the corresponding mean free paths and the mean free times. Since the mean free path depends on the number of molecules, this procedure surreptitiously introduces the same dependence as the definition of the elementary cell in the position-velocity space. In effect, Sackur treated the gas as if it consisted of molecules distributed into V/N independent sub-volumes defined by the mean free path.

Next, Sackur wrote down his final entropy formula for a monoatomic ideal gas, after filing three other versions in his preceding papers,

$$S = c_V \ln T + R \ln V + \frac{3}{2}R \left[1 + \ln \left(\frac{2\pi R}{M} \right) \right] - 4R \ln N - 3R \ln h, \quad (3.13)$$

where

$$\frac{3}{2}R \left[1 + \ln \left(\frac{2\pi R}{M} \right) \right] - 4R \ln N - 3R \ln h \equiv S^0. \quad (3.14)$$

In the last section of his paper, Sackur extended this approach to polyatomic gases. The classical treatment of the rotational energy he presented contained an error, which Tetrode pointed out to him in a personal letter. Sackur responded by publishing a paper dedicated to the discussion of gaseous molecules with two or three atoms, in which he corrected this error (Sackur 1912c).²⁸

By the end of 1912, Sackur considered his quantum theory of gaseous equilibrium to be complete: he had calculated the chemical constants for a variety of gases, which was his initial stated goal. His contact with Fritz Haber and Haber's Kaiser Wilhelm Institut enticed him to work on applications of quantum theory to other physico-chemical problems, such as the behavior of gases at very low

²⁷Sackur's idea of using the mean free path as a parameter to quantize the gas was adopted the following year by Sommerfeld and his collaborator Wilhelm Lenz (Sommerfeld 1914).

²⁸Although Tetrode's letter to Sackur has been lost, we know about its existence from a footnote in Sackur's (1912b) paper.

temperatures (Sackur 1914).²⁹ This topic preoccupied him during the last, Berlin phase of his career.

3.6 Conclusion

The Sackur-Tetrode equation is perhaps the most lasting of Sackur's contributions to quantum theory and, indeed, to science. It is still presented in textbooks as the quantum expression for the translational entropy of a monoatomic gas at high temperatures and low densities. Like Nernst's heat theorem, it played an important part in the collective groping that led eventually to a complete theory of quantum gases. Sackur's scientific ontogenesis provides a clue about the phylogenesis of his contemporaries. Although Sackur's theoretical rendition of the chemical constants was immediately and widely accepted as a touchstone of thermochemistry, his general theoretical reflections did not have the same fate. Vacillations between phase space and microscopic observers, between discontinuity and probabilistic arguments, that is, between tradition and innovation still reveal deep uncertainties concerning the basis on which his computational results were founded.

Some of his ideas were subsequently adopted or discussed—for instance, the N dependence of the elementary cell or the division of the gas volume into V/N subvolumes to ensure extensivity. But nobody seems to have followed his suggestion for a purely probabilistic-operational foundation of the quantum hypothesis. Physicists and physical chemists favored Otto Stern's approach based on the vapor-solid equilibrium (Stern 1913; 1919), which eventually triggered Enrico Fermi's analysis of the statistics of identical particles (Fermi 1923; Rasetti 1924).

However, reducing Sackur's case to a story about which aspects of his work succeeded and which failed to have an impact would mean to miss more interesting historiographical points. We conclude with two reflections on Sackur and the complex landscape of the quantum physics in the 1910s that he helped shape.

First, Sackur was a member of the diverse and variegated scientific community that developed quantum physics. In the 1910s, the quantum hypothesis made its way into statistical mechanics, radiation theory, spectroscopy, theory of matter, physical chemistry and atomic modeling. In such a complex landscape, the sheer multitude of the specialists involved lent the contentions about the quantum a considerable polyphony but little harmony. Although formal tools and mathematical techniques were widely shared among more famous (such as Einstein,

²⁹Because of his intention to move on to the experimental determination of the gas law at very low temperatures, Sackur applied for a position in Kamerlingh Onnes's institute in Leiden. However, Onnes dismissed his application with scorn (Van Delft 2007, 477).

Planck, or Nernst) and relatively unknown (such as Tetrode, or Sackur himself) players, the quantum hypothesis meant different things to different people. For instance, Einstein's perspective, shaped mostly by his interest in radiation theory, differed greatly from Nernst's, which depended on his desire to establish the validity of the heat theorem. In our case, Sackur's background in physical chemistry was key to shaping his interpretation of the quantum. Eager to put the quantum to the service of long-standing problems of thermochemistry—as well as to impress a potential employer—Sackur stressed the continuity between classical statistical techniques and the quantum. The variety of approaches made the 1910s an interesting period both from historiographical and epistemological points of view.

Second, Sackur belonged to a part of the physics community whose research style was quite different from that of, say, Planck, Einstein or Bohr. Sackur's research was not driven by an adherence to strong principles, as he lacked a grand methodological guidance such as Planck's concept of absolute entropy. This difference echoes Suman Seth's recent distinction between "physics of principle" and "physics of problems" (Seth 2010). Sackur was, no doubt, more sympathetic to an approach of starting from and returning to concrete problems. The state of flux in which quantum physics found itself at the beginning of the 1910s facilitated and almost encouraged problem-oriented lines of attack. Thus, Sackur's was, so to speak, a "mundane" quantum physics coming "from below" as a result of his personal understanding of abstract concepts and of his pragmatic agenda. Although curious about the wider implications of his arguments, he remained largely unconcerned about the quantum formalism and, in the end, settled on a conservative position, according to which the quantum is an extension of classical statistical mechanics. For a vast majority of Sackur's peers—who eventually provided a lasting contribution—the quantum was mainly a tool. For them, as for Sackur, quantum physics made sense when viewed in the context of well-defined problems in classical physics that arose within the framework of their own disciplines.³⁰

References

- Abegg, Richard and Otto Sackur (1909). *Physikalisch-Chemische Rechenaufgaben*. Leipzig: Barth.
- Arrhenius, Svante (1910). Richard Abegg. *Zeitschrift für Elektrochemie und angewandte physikalische Chemie* 16: 554–557.
- Auerbach, Friedrich (1915). Otto Sackur – Nekrologe. *Jahresbericht der Schlesischen Gesellschaft für Vaterländische Kultur* 1: 35–37.

³⁰Massimiliano Badino and Bretislav Friedrich have contributed equally to this paper.

- Badino, Massimiliano (2009). The Odd Couple: Boltzmann, Planck and the Application of Statistics to Physics (1900–1913). *Annalen der Physik* 18: 81–101.
- (2013). Dissolving the Boundaries Between Research and Pedagogy: Otto Sackur's *Lehrbuch der Thermochemie und Thermodynamik*. In: *Research and Pedagogy. A History of Quantum Physics through Its Textbooks*. Ed. Massimiliano Badino and Jaume Navarro. Forthcoming.
- Bartel, Hans-George and Rudolf Hübener (2007). *Walther Nernst, Pioneer of Physics and of Chemistry*. Singapore: World Scientific Publishing.
- Bohr, Niels (1923). Anwendung der Quantentheorie auf den Atombau. I. Die Grundpostulate der Quantentheorie. *Zeitschrift für Physik* 13: 117–165.
- Coffey, Patrick (2006). Chemical Free Energies and the Third Law of Thermodynamics. *Historical Studies in the Physical and Biological Sciences* 36: 365–396.
- Darrigol, Olivier (1988). Statistics and Combinatorics in Early Quantum Theory. *Historical Studies in the Physical and Biological Sciences* 19: 17–80.
- (1991). Statistics and Combinatorics in Early Quantum Theory, II: Early Symptoma of Indistinguishability and Holism. *Historical Studies in the Physical and Biological Sciences* 21: 237–298.
- Desalvo, Agostino (1992). From the Chemical Constant to Quantum Statistics: A Thermodynamic Route to Quantum Mechanics. *Physica* 29: 465–537.
- Des Coudres, Theodor (1910). Richard Abegg. *Physikalische Zeitschrift* 11: 425–429.
- Einstein, Albert (1907). Die Plancksche Theorie der Strahlung und die Theorie der spezifischen Wärme. *Annalen der Physik* 22: 180–190.
- Eucken, Arnold (1912). Die Molekularwärme des Wasserstoffs bei tiefen Temperaturen. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*: 141–151.
- (1914). Die Entwicklung der Quantentheorie vom Herbst 1911 bis Sommer 1913. In: *Die Theorie der Strahlung und der Quanten*. Ed. Arnold Eucken. Halle: Knapp. 371–405.
- (1922). Der Nernstsche Wärmeersatz. *Ergebnisse der exakten Naturwissenschaften* 1: 120–162.
- Fermi, Enrico (1923). Sopra la teoria di Stern della costante assoluta dell'entropia di un gas perfetto monoatomico. *Rendiconti della Accademia dei Lincei* 32: 395–398.
- Friedrich, Bretislav and Dudley Herschbach (2003). Stern and Gerlach: How a Bad Cigar Helped Reorient Atomic Physics. *Physics Today* 56(12): 53–59.

- Gearhart, Clayton A. (2010). 'Astonishing Successes' and 'Bitter Disappointment': The Specific Heat of Hydrogen in Quantum Theory. *Archive for History of Exact Sciences* 64: 113–202.
- Gibbs, Josiah Willard (1878). On the Equilibrium of Heterogeneous Substances. *Transactions of the Connecticut Academy* 3: 108–248, 343–524.
- Guldberg, Cato M. and Peter Waage (1899). *Untersuchungen über die chemischen Affinitäten. Abhandlungen aus den Jahren 1864, 1867, 1879*. Ed. Richard Abegg. Leipzig: Englemann. Translated by Richard Abegg.
- Haber, Fritz (1905). *Thermodynamik technischer Gasreaktionen*. Munich: Oldenbourg Verlag.
- von Helmholtz, Hermann (1882). Die Thermodynamik chemischer Vorgänge. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin* 1: 22–39.
- Hertz, Walther (1915). Otto Sackur. *Physikalische Zeitschrift* 16: 113–115.
- Hiebert, Edwin N. (1971). The Energetic Controversy and the New Thermodynamics. In: *Perspectives in the History of Science and Technology*. Ed. Duhane H. D. Roller. Normal: University of Oklahoma Press. 67–86.
- (1983). Walther Nernst and the Application of Physics to Chemistry. In: *Springs of Scientific Creativity*. Ed. Rutherford Aris, Howard T. Davis, and Roger H. Stuewer. Minneapolis: University of Minnesota Press. 203–231.
- Hoyer, Ulrich (1980). Von Boltzmann zu Planck. *Archive for History of Exact Sciences* 23: 47–86.
- Jüttner, Ferenz (1911). Über die Ableitung der Nernstschen Formeln für Reaktionen in kondensierten Systemen. *Zeitschrift für Elektrochemie und angewandte physikalische Chemie* 17: 139–143.
- Kipnis, Alexander (2005). Entry on Otto Sackur. In: *Neue Deutsche Biographie*. 22. Berlin: Dunker und Humblot. 344.
- Koref, Fritz (1911). Messungen der spezifischen Wärme bei tiefen Temperaturen mit dem Kupferkalorimeter. *Annalen der Physik* 36: 49–73.
- Kormos Barkan, Diana (1999). *Walther Nernst and the Transition to Modern Physical Science*. Cambridge: Cambridge University Press.
- Kox, Anne J. (2006). Confusion and Clarification: Albert Einstein and Walther Nernst's Heat Theorem, 1911–1916. *Studies in History and Philosophy of Modern Physics* 37: 101–114.
- Kuhn, Thomas S. (1962). *Interview with Dr. Otto Stern by Thomas S. Kuhn at Stern's Berkeley Home May 29 and 30, 1962. The Oral History Interviews and Transcripts Project by the American Institute of Physics*. <http://www.aip.org/history/nbl/ohiproject.html> (accessed 03/16/2012).
- Morse, Philip M. (1969). *Thermal Physics*. 2nd ed. Reading: The Benjamin/Cummings Publishing Company.

- Nernst, Walther (1906a). Über die Berechnung chemischer Gleichgewichte aus thermischen Messungen. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen* 1: 1–39.
- (1906b). Über die Beziehungen zwischen Wärmeentwicklung und maximaler Arbeit bei kondensierten Systemen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*: 933–940.
 - (1907). *Experimental and Theoretical Applications of Thermodynamics to Chemistry*. New Haven: Yale University Press.
 - (1911). Zur Theorie der spezifischen Wärme und über die Anwendung der Lehre von den Energiequanten auf physikalisch-chemische Fragen überhaupt. *Zeitschrift für Elektrochemie und angewandte physikalische Chemie* 17: 265–275.
 - (1917). *Die theoretischen und experimentellen Grundlagen des neuen Wärmesatzes*. Halle: Knapp.
- Pick, Hans (1915). Otto Sackur. *Chemiker Zeitung* 39: 13.
- Pier, Mathias (1909). Spezifische Wärme von Gasen bei sehr hohen Temperaturen. *Zeitschrift für Elektrochemie* 15: 536–540.
- (1910). Spezifische Wärme und Gasgleichgewicht nach Explosionsversuchen. *Zeitschrift für Elektrochemie* 16: 897–903.
- Planck, Max (1906). *Vorlesungen über die Theorie der Wärmestrahlung*. 1st ed. Leipzig: Barth.
- (1911). *Vorlesungen über die Thermodynamik*. 3rd ed. Leipzig: Veit.
 - (1913). *Vorlesungen über die Theorie der Wärmestrahlung*. 2nd ed. Leipzig: Barth.
 - (1914). Die gegenwärtige Bedeutung der Quantenhypothese für die kinetische Gastheorie. In: *Vorträge über die kinetische Theorie der Materie und der Elektrizität*. Ed. David Hilbert. Göttingen: B. G. Teubner. 3–16.
- Polanyi, Michael (1913). Eine neue thermodynamische Folgerung aus der Quantenhypothese. *Verhandlungen der Deutschen Physikalischen Gesellschaft* 15: 156–161.
- Rasetti, Franco (1924). La costante assoluta dell'entropia e le sue applicazioni. *Il nuovo cimento* 3: 67–86.
- Sackur, Otto (1911a). Zur kinetischen Begründung des Nernstschen Wärmetheorems. *Annalen der Physik* 34: 455–468.
- (1911b). Die Anwendung der kinetischen Theorie der Gase auf chemische Probleme. *Annalen der Physik* 36: 958–980.
 - (1912a). *Die Bedeutung des elementaren Wirkungsquantums für die Gastheorie und die Berechnung der chemischen Konstanten*. Nernst Festschrift. Halle: Knapp.

- (1912b). Die universelle Bedeutung des sogenannten elementaren Wirkungsquantums. *Annalen der Physik* 40: 67–86.
 - (1912c). Die ‘Chemischen Konstanten’ der zwei- und dreiatomigen Gase. *Annalen der Physik* 40: 87–106.
 - (1914). Die Zustandsgleichung der Gase und die Quantentheorie. *Zeitschrift für Elektrochemie und angewandte physikalische Chemie* 20: 563–570.
- Seth, Suman (2010). *Crafting the Quantum: Arnold Sommerfeld and the Practice of Theory, 1890–1926*. Boston: MIT Press.
- Sommerfeld, Arnold (1911). Das Plancksche Wirkungswachstum und seine allgemeine Bedeutung für die Molekularphysik. *Physikalische Zeitschrift* 12: 1057–1068.
- (1914). Probleme der freien Weglänge. In: *Vorträge über die kinetische Theorie der Materie und der Elektrizität*. Ed. David Hilbert. Göttingen: B. G. Teubner. 125–166.
- Stern, Otto (1913). Zur kinetischen Theorie des Dampfdrucks einatomiger fester Stoffe und über die Entropiekonstante einatomiger Gase. *Physikalische Zeitschrift* 14: 629–632.
- (1919). Zusammenfassender Bericht über die Molekulartheorie des Dampfdruckes fester Stoffe und ihre Bedeutung für die Berechnung chemischer Konstanten. *Zeitschrift für Elektrochemie* 25: 66–80.
- Tetrode, Hugo M. (1912). Die chemische Konstante der Gase und das elementare Wirkungsquantum. *Annalen der Physik* 38: 434–442.
- Van Delft, Heike (2007). *Freezing Physics. Heike Kamerlingh Onnes and the Quest for Cold*. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen.
- Van't Hoff, Jacobus (1886). L'équilibre chimique dans les systèmes gazeux, ou dissous à l'état dilué. *Archives néerlandaises des sciences exactes et naturelles* 20: 239–302.

Quantum Mechanics in the Making

Chapter 4

The Concepts of Light Atoms and Light Molecules and Their Final Interpretation

Dieter Fick and Horst Kant

In 1900, Max Planck (1858–1947) obtained *his* famous radiation formula, see eq. (4.1) (Planck 1913, §156, eq. 275), for the energy per volume and frequency interval emitted by a black body by rather “obscure means” (Darrigol 2009):

$$u_\nu(T) = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{e^{h\nu/kT} - 1}. \quad (4.1)$$

This fitted perfectly with the data over the whole frequency-temperature range investigated at that time (Warburg 1913; Rubens 1913). The way Planck found this formula, partly by “ingenious mathematical manipulations” (Cassidy 2005), how he was driven to the assumption and, later on, to the acceptance of the discrete energy quanta of the oscillators within his model black-body radiator has been related so many times that we may disregard it here.¹ We would like to comment here that the derivation of the two factors in eq. (4.1) show quite different problems. The first factor is connected to the dynamics of the oscillators within a black-body radiator; the second derives from the combinatorial assumptions of how energy elements are distributed over resonators. Its various forms and the related controversies were discussed extensively in publications by Olivier Darrigol (1988; 1991). Satyendranath Bose (1894–1974) was the first to put both factors on equal footing (Bose 1924).

In what follows, we concentrate exclusively on the interpretation, *not the derivation*, of the second factor in eq. (4.1) in terms of light particle concepts. Albert Einstein (1879–1955) opened discussion on the particle nature of black-body radiation as early as 1905 (Einstein 1905). In discussing the entropy of a black body as function of volume, he showed that black-body radiation behaves in the Wien limit like a diluted gas consisting of light quanta. The energy of light appeared in some kind of “granular structure” (Darrigol 1988, 20). Here, we will mainly follow the historical development of the light atom and light molecule concepts, giving some emphasis to the two main actors, Mieczysław

¹For a comprehensive review, see (Kuhn 1987).

Wolfke (1883–1947) and Walther Bothe (1891–1957), as well as to their relationships with Planck, Einstein, Max von Laue (1879–1960) and Louis de Broglie (1892–1987). The early part of this story, up to the beginning of the 1920s, has already been dealt with in part by Silvio Bergia and Darrigol (Bergia, Ferrario, and Monzini 1985; Darrigol 1988; 1991).

4.1 First Corpuscular Concepts of Light

Einstein always denied interpreting light in general as being composed of independent quanta. In a letter to Hendrik A. Lorentz (1853–1928) dated 23 May 1909, Einstein writes:

[...] I am not at all of the opinion that light has to be thought of as being composed of mutually independent quanta localized in relatively small spaces. To be sure, that would be the most convenient way to explain the Wien end of the radiation formula. But the splitting of light rays on the surface of refracting media already makes this approach absolutely inadmissible. A light ray splits, but a light quantum cannot split without a change in frequency. (Klein, Kox, and Schulmann 1993, 193; for German original see: Klein, Kox, and Schulmann 1993, 123)²

In his 1916 and 1917 papers, “Zur Quantentheorie der Strahlung,” (Einstein 1916; 1917)³ Einstein showed that, analogous to particles, each light quantum in a radiation bundle carries a momentum of $h\nu/c$. However, before Einstein, a number of researchers had already understood these as real atoms. An early summary of these activities can be found in Harry Bateman’s (1882–1946) 1923 publication (Bateman 1923).

On 27 September 1910 Abram F. Ioffe (1880–1960) presented a talk, “Zur Theorie der Strahlungserscheinungen” (Ioffe 1911), at a meeting of the Physical Division of the Russian Physico-Chemical Society, the content of which he had already discussed a few weeks earlier with Planck.⁴

The headline of the second part of this publication (Ioffe 1911) “Atomistische Struktur der Strahlung” and the headline of §2 “Strahlungsquanten” both

²Einstein always maintained this point of view, as seen in a letter sent to Wolfke in 1946, see sec. (4.4).

³Since the text in both publications is identical, we will refer to the more easily accessible 1917 publication only.

⁴“I [Ioffe] tried to build [at that time] a theory of radiation energy analogous to the kinetic energy of gases” (Ioffe 1983, 63). Ioffe further aimed to discuss his “at that time heretical ideas” with Planck and to this end visited him at his resort at Lake Chiemsee in Upper Bavaria at the end of August 1910. Ioffe’s former teacher, Wilhelm Conrad Röntgen (1845–1923), arranged the meeting. Ioffe also notes:

point toward an interpretation of light in terms of an atomistic concept. Indeed, §2 in part two starts with the sentence:

Since the appearance of Einstein's article a series of facts have been discovered and discussed, which find their *simplest* explanation in an atomistic concept of radiation, or at least of its emission.⁵ (Ioffe 1911, 546–547)

A list of seven points supporting this claim followed.

Around the same time, Johannes Stark (1874–1957) tried to gain experimental insight into the description of X-rays as light quanta, or alternatively as ether waves (Stark 1910).⁶ Theoretically, he discussed the momentum conservation in electron collisions with matter in great detail.⁷ Experimentally, he analyzed the forward-backward asymmetry of X-rays emitted in electron collisions on a thin, low Z anode (charcoal). In formulating the conditions for momentum conservation under the assumption that the X-rays are light quanta, he explicitly used a vector of length $h\nu/c$ for the momentum of the emitted X-rays and predicted a pronounced forward-backward asymmetry of their momentum (energy) and intensity distributions. In contrast, if X-rays were ether-waves, he found that electrodynamics demanded an isotropic distribution (Abraham 1905). The very demanding experiment revealed pronounced forward-backward asymmetries in X-ray intensities and energies, clearly favoring the light-quantum hypothesis.⁸

In 1913, Wolfke, at the time a *Privatdozent* (private lecturer) in Zurich, was probably the first to introduce the item “light atom” (*Lichtatom*) as a center with energy ϵ , referring to Stark and to Einstein's 1905 article (Stark 1910; Einstein

[...] He [Planck] deemed my article interesting. However, he urgently insisted that I abandon the use of photons since they are inconsistent with Maxwell's electromagnetic theory of light. (Ioffe 1983, 63)

In his reminiscences, Ioffe used the term “photon,” even though it was introduced in 1926 by Lewis (G. N. Lewis 1926). Nevertheless, it was published somewhat later in *Annalen der Physik* of which Planck was an editor. During his visit, Planck had assured Ioffe:

[...] that he will not oppose the publication of the manuscript. However, he does not intend to ruin with his own hands the principles of Maxwell's construct. For him personally, the manuscript is certainly lamentable [...]. (Ioffe 1985, 425)

⁵“Seit dem Erscheinen des Aufsatzes von A. Einstein ist eine Reihe von Tatsachen entdeckt und diskutiert worden, die ihre *einfachste* Erklärung in der atomistischen Auffassung der Strahlung, oder wenigstens ihrer Emission, finden.” Unless otherwise indicated all English translations are by the authors.

⁶These are now called electromagnetic waves.

⁷At that time the German word for momentum was *Bewegungsgröße*.

⁸Even though in 1910 there was growing evidence that “X-rays and light are manifestations of the same phenomena” only a few, such as Stark, believed this was so (Wheaton 1983, 169).

1905; M. Wolfke 1913a). Light energy U was thought to be localized in a large but finite number N of these centers:

$$U = N\epsilon. \quad (4.2)$$

Wolfke chose the term “light atom” to indicate that they cannot come into existence or decay by themselves, for example, N stays constant when reflected off a moving, perfect mirror. He then used this hypothesis to derive the dispersion relation for light atoms with the following arguments: according to classical electrodynamics, the ratios of beam energies U_{in} and U_{ref} and of the frequencies ν_{in} and ν_{ref} for an incoming and reflected beam of light respectively, are identical, see (Abraham 1905, §40; Planck 1913, §77, eqs. 86/87). Thus

$$\frac{U_{in}}{U_{ref}} = \frac{\nu_{in}}{\nu_{ref}} = \frac{N\epsilon_{in}}{N\epsilon_{ref}} = \frac{\epsilon_{in}}{\epsilon_{ref}} \quad (4.3)$$

follows, since the number of light atoms N in a beam ought not to change while the beam is reflected from a perfect mirror. Therefore,

$$\frac{\epsilon_{in}}{\nu_{in}} = \frac{\epsilon_{ref}}{\nu_{ref}} = \text{const.} \quad (4.4)$$

The light atoms thus follow the dispersion relation

$$\frac{\epsilon}{\nu} = \alpha, \quad (4.5)$$

α being a universal constant.⁹ Wolfke commented on this result as follows: “The equation forms the main equation of the light atom”¹⁰ (M. Wolfke 1913a, 1125).

Using this atomistic picture of light, Wolfke interpreted radiation pressure as being caused by the hits of light atoms on a mirror (M. Wolfke 1913b). In doing so, he allocated a mass to a light atom of energy $\epsilon = h\nu$ via the relativistic relation $m = \epsilon/c^2$ for the first time. Before arriving at the heart of the problem, he related the number of light atoms per unit time and volume to the power (energy per time) of the light beam. Then he used a formula found in Planck’s book on heat radiation (Planck 1913, §60, 58, uppermost formula). It connected the radiation pressure \mathfrak{F} , with the number of light atoms hitting the mirror per unit area, with their allotted mass m and with the scattering angle Θ . This formula was obtained by Planck in a quantum picture of light assuming correctly $p = mc$ for the momentum of a light

⁹Because of a mistake in the derivation of Planck’s formula, Wolfke realized only in a subsequent publication (M. Wolfke 1913b) that α is identical to Planck’s constant.

¹⁰“Die Gleichung bildet die Hauptgleichung des Lichtatoms.”

quantum.¹¹ Wolfke finally eliminated the mass through the correct relativistic relation $m = \epsilon/c^2 = hv/c^2$ and arrived thus within a quantum picture of light at the classical relation $\mathfrak{F} = (2 \cos \Theta/c)J$ between radiation pressure \mathfrak{F} and light power J .

At this point, we pause to present Wolfke's involvement up until 1920 in the "light atom" concept in further detail, in particular his controversy with Yuriy Aleksandrovich Krutkov (1890–1952), Paul Ehrenfest's (1886–1930) pupil in St. Petersburg and at that time his visitor in Leiden (Frenkel 1971). This involvement did not lead to any considerable insight and was discussed in detail by Darrigol (1991, 254–255), as well as by Luis Navarro and Enric Perez (2004, 130–132).

First, we briefly report on Wolfke's background.¹² Wolfke was a native Pole. Under Otto R. Lummer (1860–1925), he received a doctorate in Breslau in 1910 with a dissertation on optics (M. Wolfke 1911). At the beginning of 1913, Wolfke moved to the Eidgenössische Technische Hochschule (ETH)¹³ in Zurich. He must have soon met Einstein, since Einstein served as referee in Wolfke's *Habilitationskommission* (habilitation committee) (K. Wolfke 1980). On the basis of a publication that had already appeared in 1912 in *Annalen der Physik* (M. Wolfke 1912), as well as of the positive appraisal of his personality and abilities by Einstein and others,¹⁴ he was promoted on 8 May 1913 to *Privatdozent* at the ETH.

According to a personal report of his son Karol Wolfke, Einstein often visited the Wolfkes and "played violin with father's piano accompaniment" (Sredniawa 2006, 261). It is thus rather probable that Wolfke discussed the light quantum problems with Einstein. And indeed Wolfke notes in one of his publications, at the time of his debate with Krutkov, that Einstein brought certain facts to his attention (M. Wolfke 1914b). This hypothesis is corroborated by several remarks by Darrigol, indicating that Einstein supported Wolfke in his struggle with Krutkov (Darrigol 1991, 254–259).

¹¹This remark is important: Planck assumed, within the same paragraph, that the energy of a light quantum is related to a mass through the Newton relation $\epsilon = \frac{1}{2}mc^2$, and not through the relativistically correct relation $\epsilon = mc^2$. Therefore, Planck obtained twice the classical value for the radiation pressure as Maxwell did for the first time. (For a detailed derivation of the classical radiation pressure, see (Planck 1913, §58, eq. 64)).

¹²For more details, as well as information on Wolfke's relationship with Einstein over the years, see (Kiejna 2002).

¹³Swiss Federal Institute of Technology, Zurich.

¹⁴Letter [from an unknown author, handwritten] of Section VIII from 26 May 1913 addressed to the Chair of the Swiss School Council in Zurich regarding the request for the habilitation of Wolfke: "Professors Einstein and Weiss both agree in their favorable appraisal of the submitted scientific publication and of the professional qualifications and the character of the applicant, and accept the habilitation [...]." "Die Herrn Proff. Einstein und Weiss sprachen sich übereinstimmend günstig über die eingereichten wissenschaftlichen Arbeiten, die Vorbildung und die Persönlichkeit des Gesuchstellers aus, und begrüßten die Habilitation [...].", III-71.21, 23, APAN.



Figure 4.1: Mieczysław Wolfke. Courtesy of the Archives of the Polish Academy of Sciences (APAN).

In the fall of 1922, Wolfke became a professor of physics at the Technical University in Warsaw. There, he worked experimentally on a variety of problems concerning optics, high voltages, properties of liquid helium, and obviously also on what he called “light molecules.” From a report in *Nature* on the Fifth Congress of Polish Physicists in Poznań (Anonymous 1930, 660), we learn that Wolfke presided over the congress, and we read further that “special interest was aroused by papers on association of light quanta by Wolfke.”

4.2 Light Molecules, Static Concepts

4.2.1 Wolfke 1921

At the beginning of 1914, Wolfke published a brief third note, entitled “Zur Quantentheorie” (M. Wolfke 1914a), within a series of papers released in *Verhandlungen der DPG*. Here, he speculated qualitatively on how the transition from the classical Rayleigh-Jeans limit (large radiation density) to the Wien limit (small radiation density) could proceed. A few years later while still in Zurich, he resumed this topic.¹⁵ In a publication entitled “Einsteins Lichtquanten und die räumliche Struktur der Strahlung”, he intended to demonstrate that “[...] black-body radiation [...] consists of thermodynamically-independent parts [...]”¹⁶ (M. Wolfke 1921, 378) with energy density contents of $u_{\nu,s}$, $s = 1, 2, 3, \dots$.¹⁷

His starting point was the then well-known identity for the energy density per unit frequency interval in Planck’s formula, see eq. (4.1)

$$u_{\nu}(T) = u_{\nu} = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{e^{h\nu/kT} - 1} = \sum_{s=1}^{\infty} u_{\nu,s} \quad (4.6)$$

with

$$u_{\nu,s} = \frac{8\pi\nu^2}{c^3} h\nu e^{-sh\nu/kT}, \quad (4.7)$$

the $u_{\nu,s}$ following Wien’s radiation law.^{18,19}

Using this expansion and generalizing Einstein’s method from the 1905 paper (Einstein 1905) to the full radiation spectrum, Wolfke showed that the sum of the partial entropy densities $s_{\nu,s}$ calculated for the partial energy densities $u_{\nu,s}$ adds up to the well-known expression for the entropy density of black-body ra-

¹⁵We can only guess why Wolfke took so long to publish on the light quantum problem. Einstein had already left Zurich in 1913, and during World War I, Wolfke and his family (wife and two children) faced severe economic problems with no regular income (K. Wolfke 1980).

¹⁶“[...] die schwarze Strahlung aus [...] voneinander thermodynamisch unabhängigen Teilstrahlungen besteht [...]”

¹⁷Different from the notations of the publications to be discussed, we additionally label the expansion coefficients and their associated quantities with the frequency ν to which they refer. This is necessary for Appendix 4.4, in which processes that change the frequency of light molecules, such as the Compton effect, are treated.

¹⁸Jun Ishiwara (1881–1947) discussed this expansion as early as 1912 (Ishiwara 1912).

¹⁹To follow the mathematical manipulations throughout this manuscript more easily, the following identities are useful: $\sum_{k=0}^{\infty} z^k = 1/(1-z)$; $\sum_{k=1}^{\infty} k \cdot z^k = z/(1-z)^2$; $\sum_{k=1}^{\infty} k^2 \cdot z^k = z \frac{d}{dz} (\sum_{k=1}^{\infty} k \cdot z^k) = z(1+z)/(1-z)^3$.

diation $\sum_{s=1}^{\infty} \varepsilon_{v,s} = \mathfrak{E}_v$. Thus he interpreted the partial radiation densities $u_{v,s}$ as thermodynamically independent of each other.

For a black body of volume V_0 , $u_{v,s}V_0 = \varepsilon_{v,s}$ may denote the energy per frequency interval of the s th partial radiation. Following Einstein's methodological tools, Wolfke obtained this radiation energy for the probability $W_{v,s}$ within a subvolume, V

$$W_{v,s} = \left(\frac{V}{V_0} \right)^{(\varepsilon_{v,s}/shv)}. \quad (4.8)$$

He therefore interpreted this relation such that the s th partial wave consists of $\varepsilon_{v,s}/shv$ spatially independent radiation quanta shv . He named these objects light molecules and finished his considerations with the remark:

[...] that black-body radiation, as seen from the point of view of Einstein's light-quantum hypothesis, consists of spatially independent light molecules $hv, 2hv, 3hv$ [...].²⁰ (M. Wolfke 1921, 378)

Finally, Wolfke analyzed the number density $n_{v,s}$ of light molecules per unit frequency interval:²¹

$$n_{v,s} = \frac{u_{v,s}}{shv} = \frac{8\pi v^2}{c^3} \frac{1}{s} e^{-shv/kT}, \quad (4.9)$$

yielding for the ratio of successive number densities:²²

$$\frac{n_{v,s+1}}{n_{v,s}} = \frac{s}{s+1} e^{-hv/kT}. \quad (4.10)$$

For large values of hv/kT (Wien limit), the ratio $n_{v,2}/n_{v,1}$ becomes very small in comparison to 1. Thus the radiation field consists of independent light quanta (light atoms) only, as stated by Einstein in his 1905 paper (Einstein 1905). Ap-

²⁰“[...] daß die Hohlraumstrahlung, vom Standpunkt der Einsteinschen Lichtquantenhypothese aus betrachtet, aus voneinander räumlich unabhängigen Lichtmolekülen $hv, 2hv, 3hv$, [...] zusammengesetzt ist.”

²¹Following Wolfke, for this presentation we rewrote all quantities that depend on volume and frequency such that their value per volume, v , and frequency interval, dv , is always quoted. Thus, $n_{v,s}$ is really a number density per frequency interval, dv , and not, as Wolfke mistakenly called it, the numbers of the various light molecules. In all the other publications discussed here, the n are always connected to the number of light molecules, but refer differently to the volume and the frequency interval. For example, in Bothe's early manuscript (Bothe 1923), the n denote the number of light molecules per frequency interval, even though this is not evident at first glance. In his last publication, Bothe defined n as a dimensionless number (Bothe 1927b), as did de Broglie in his brief publication (de Broglie 1922b) and Darrigol in his review article (Darrigol 1991).

²²In his paper, Wolfke discusses the inverse ratio.

proaching the classical Rayleigh-Jeans limit with decreasing values of $h\nu/kT$, the ratios $n_{\nu,2}/n_{\nu,1}$ increase. Wolfke concluded:

We see from this that with growing radiation density there occurs an association of light quanta into increasingly complex molecules, until finally—for very high radiation densities satisfying the Rayleigh-Jeans formula—the quanta agglomerate to form a continuous distribution in space. On the other hand, with decreasing radiation density the radiation continuum dissociates into simpler and simpler light molecules, until it ultimately dissolves into discrete light atoms!²³ (M. Wolfke 1921, 378)

4.2.2 Louis de Broglie 1922

About one year later, without citing Wolfke's publication, de Broglie published two short notes dealing with the particle properties of light and with Einstein's expression for the energy fluctuations, the variance $\overline{\Delta_\nu^2}$ of black-body radiation (de Broglie 1922a; 1922b). He found on purely formal grounds that the expansion of Planck's formula into $h\nu \cdot e^{-sh\nu/kT}$, $s = 1, 2, 3, \dots$ ²⁴ does not contradict Einstein's fluctuation formula (Einstein 1909):

$$\overline{\Delta_\nu^2} \cdot V d\nu = (h\nu) \cdot E_\nu + \frac{E_\nu^2}{Z_\nu}, \quad (4.11)$$

with

$$E_\nu = u_\nu \cdot V d\nu, \quad (4.12)$$

the time average of the energy at frequency ν in a volume V . According to Peter Debye (1884–1966):

$$Z_\nu = \frac{8\pi\nu^2}{c^3} V d\nu \quad (4.13)$$

denotes the number of elementary states²⁵ within a frequency interval, $d\nu$, for a black body of volume V (Debye 1910).

²³“Wir sehen daraus, wie mit zunehmender Strahlungsdichte eine Assoziation von Lichtquanten zu immer komplizierteren Lichtmolekülen stattfindet, bis schließlich die Quanten bei sehr großen Strahlungsdichten, im Gültigkeitsbereich des Jeans-Rayleighschen Strahlungsgesetzes, sich zu einem Kontinuum zusammenballen! Umgekehrt, mit abnehmender Strahlungsdichte dissoziiert das Strahlungskontinuum in immer einfachere Lichtmoleküle, bis es sich schließlich in diskrete Lichtatome auflöst!”

²⁴Equations resembling Wolfke's eqs. (4.6), (4.7).

²⁵De Broglie did not use the quantity Z_ν explicitly.

This is important historically because the two parts in Einstein's fluctuation equation at that time were often literally connected to the *particle* ($h\nu$) $\cdot E_\nu$ and *wave* aspect E_ν^2/Z_ν of light. In his results, de Broglie intuitively saw the physical picture of "coherent photons" (mono, duo, triple correlated photons) in the expansion of Planck's formula, see eqs. (4.6), (4.7).²⁶ Therefore, he was convinced that "if the theory of light-quanta ever succeeds in interpreting interference, it will require such agglomeration of quanta" (Darrigol 1991, 260).

4.3 Dynamical Treatments of Light Multiples

4.3.1 Bothe 1923

Citing Wolfke's 1921 publication only in passing, Bothe,²⁷ Planck's former student, submitted a manuscript entitled "Die räumliche Energieverteilung der Hohlraumstrahlung" to *Zeitschrift für Physik* in 1923 (Bothe 1923).²⁸ At the time, Bothe worked at the Physikalisch Technische Reichsanstalt (PTR) in Berlin-Charlottenburg. Aside from his experimental work (Fick and Kant 2009), from mid-1923 to the end of 1926, he published a few theoretical papers, all of which dealt with the "light quantum problem." In the acknowledgments and footnotes, more often than not, he thanked von Laue, who was a full professor at the Berlin University, a member of the Königlich Preußische Akademie der Wissenschaften (Prussian Academy of Sciences) and deputy director of the Kaiser Wilhelm Institut für Physik during that time (Hoffmann 2010, and references therein). In addition, Einstein was very influential. Einstein was also a member of the Prussian Academy. Beginning in 1914, he lived and worked in Berlin, and in 1916, he became a member of the *Kuratorium* (board of trustees) of the PTR. Since the light quantum problem was a primary concern of Bothe's, he certainly kept in touch with Einstein.²⁹ In his Nobel lecture, Bothe recalls this period: "Dur-

²⁶A detailed discussion of de Broglie's concerns can be found in (Bergia, Ferrario, and Monzini 1985).

²⁷Details on Bothe's vita up to the 1920s and his experimental achievements in the study of the wave-particle duality of light during the 1920s can be found in a recent article entitled "Walther Bothe's Contributions to the Understanding of the Wave-Particle Duality of Light" (Fick and Kant 2009).

²⁸"The Spatial Energy Distribution of Black-Body Radiation." The Archive of the Max Planck Society in Berlin (AMPG) retains, at the back of another manuscript, a handwritten manuscript (Bothe manuscript, III/6/104,1, AMPG) which coincides in large part with the printed version of Bothe's publication in the *Zeitschrift für Physik*. The differences in the printed version are marginal, as far as the content is concerned. On a few additional pages (microfilm pages 1211, 1213, 1224), Bothe also discusses the question of whether the structure of radiation emitted by a black body changes under processes that do not change the temperature of the radiation (reflection, refraction, absorption, etc.). He concluded that at his time such questions could not be answered experimentally.

²⁹See also the end of this section.

ing this time I had the singular good fortune of being able to discuss the [wave particle] problem constantly with Einstein” (Bothe 1964, 274).

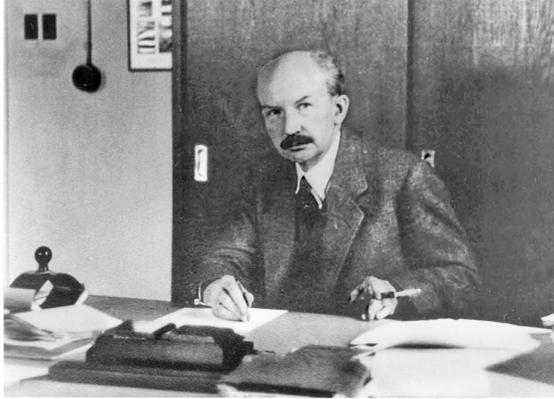


Figure 4.2: Walther Bothe. Courtesy of the Archives of the Max Planck Society, Berlin.

Beginning with Einstein’s formulation of emission and absorption processes (Einstein 1917), in his paper, Bothe studied how a two-level object (for example, an atom) achieves thermal equilibrium in a black-body radiation field. He characterized the atom by states 1 and 2 with energies ϵ_1 and ϵ_2 ($\epsilon_2 > \epsilon_1$) and occupation numbers N_1 and N_2 . According to Einstein, the atoms can undergo two types of emission processes by emitting a light quantum of energy $h\nu = \epsilon_2 - \epsilon_1$: a spontaneous one proportional to aN_2 and an induced one proportional to bN_2 . The absorption of a light quantum with energy $h\nu$, inducing transitions from state 1 to state 2 is then proportional to bN_1 .³⁰ The constants a and b are characteristic for the transitions involved.

To introduce the concept of a “quantum multiple” for a fixed frequency ν , Bothe reminds the reader that according to Einstein (1917), induced emission and absorption are, spatially, perfectly correlated processes. Moreover for stimulated emission, the inducing and stimulated quanta are perfectly correlated. They possess identical directions and energies (phases, polarizations).³¹ He noticed later that:

³⁰Without a lack of generality, we choose the statistical weights of the states 1 and 2 to be the same and equal to one, since they will not appear in the final results. This is why the b -coefficients for induced transitions and for absorption of a light quantum are equal here.

³¹We will return to this point at the end of sec. (4.3.3).

Both [quanta] are coupled seemingly; only seemingly, since in truth no forces exist between both, the dissociation energy [...] is zero. If one of both quanta is absorbed, the fate of the other is not influenced at all; this is a consequence of the assumption that the probability of an induced process is simply proportional to the spatial density of the quanta. We shall therefore better speak about *quantum pairs*. If the inducing quantum already belongs to a pair, a triple will emerge, and so on.³² (Bothe 1923, 147)

It is clear that Bothe's quantum multiples have only formally common aspects to Wolfke's and de Broglie's light molecules. Whereas both interpreted them more or less as real particles, Bothe in particular had the correlation aspect in mind. In our present understanding, we would instead call them "quasiparticles."

Denoting the number density per frequency interval dv of single quanta, pairs, triples, ... s -fold multiples, of light quanta by $n_{v,1}, n_{v,2}, n_{v,3}, \dots, n_{v,s}, \dots$, Bothe first studied the thermal equilibrium conditions.³³ Within a time interval dt , single light quanta are produced with a probability of aN_2 by spontaneous emission, and with a probability of $bN_1(2n_{v,2}h\nu)$ by absorption from a quantum pair, since $2n_{v,2}h\nu$ is the fraction of the total radiation density u_v , which belongs to quantum pairs. These are the two source terms. On the other hand, single quanta disappear with a probability $bN_1(n_{v,1}h\nu)$ through an absorption process, and with a probability of $bN_2(n_{v,1}h\nu)$ through conversion into a quantum pair in a stimulated emission process. These are the two drain terms. Since in thermal equilibrium the number of single quanta should be stationary, one obtains as a condition:

$$aN_2 + bN_1(2n_{v,2}h\nu) - bN_1(n_{v,1}h\nu) - bN_2(n_{v,1}h\nu) = 0, \quad (4.14)$$

which couples singlets and doublets of quanta with the spontaneous decay (zero quantum).

³²“Beide [Quanten] sind scheinbar gekoppelt; nur scheinbar deshalb, weil in Wahrheit keine Kräfte zwischen ihnen wirken, die Dissoziationsarbeit [...] ist Null. Wird etwa eins der beiden Quanten absorbiert, so wird das Schicksal des anderen hierdurch in keiner Weise beeinflusst; dies ist eine Konsequenz der Annahme, dass die Wahrscheinlichkeit eines Einstrahlungsprozesses einfach proportional der mittleren räumlichen Dichte der Quanten ist. Wir werden deshalb besser von Quantenpaaren sprechen. Gehört das auslösende Quant selbst schon einem Paar an, so entsteht ein Tripel, usf.”

³³In the manuscript, Bothe denotes by $n_{v,1}, n_{v,2}, n_{v,3}, \dots, n_{v,s}$ the number of light quanta $h\nu$ which form singlets, pairs, triplets, ... respectively. This definition differs from the one used by Wolfke (1921), and also from the definition used later by Bothe himself (1924). In what follows, we use Wolfke's definition (1921) of the $n_{v,s}$. It relates to the definition in the work discussed here through the relation $(n_{v,s})_{\text{Bothe}} = (s \cdot n_{v,s})_{\text{Wolfke}}$. Thus, here, the fraction of the total radiation density $u_{v,s}$, which belongs to s -fold quantum multiples, is $u_{v,s} = sn_{v,s}h\nu$.

The next higher-order rate equation, coupling singlets, doublets, and triplets with each other

$$bN_2(n_{v,1}hv) + bN_1(3n_{v,3}hv) - bN_1(2n_{v,2}hv) - bN_2(2n_{v,2}hv) = 0, \quad (4.15)$$

consists again of two source and two drain terms. The previous source term converting a doublet into a singlet now becomes a drain term for doublets, and the previous drain term for singlets now becomes a source term for doublets. This is the general structure of all higher-order rate equations. Two of the contributions always change sign in the next order equation.

Therefore, summing up these equations to the order of s , most of the terms cancel each other out. One obtains a rate equation, which couples an $(s + 1)$ - and an s -fold quantum state with a single-quantum state and the term for the spontaneous decay:

$$aN_2 + bN_1(s + 1)n_{v,s+1}hv - bN_2sn_{v,s}hv - bN_1n_{v,1}hv = 0. \quad (4.16)$$

Only the spontaneous decay process, aN_2 , cannot depend on the radiation densities for the various multiple quantum states. Since the total number of light quanta $\sum_{s=1}^{\infty} sn_{v,s}$ has to be finite, the number densities $n_{v,s}$ have to approach zero with increasing s , yielding finally:

$$-bN_1n_{v,1}hv + aN_2 = 0. \quad (4.17)$$

According to Boltzmann's law, the occupation numbers N_i ($i = 1, 2$) of the atomic states can be expressed as function of temperature³⁴

$$N_i = e^{-\epsilon_i/kT}. \quad (4.18)$$

Following Einstein (1917), the ratio a/b is then given by

$$a/b = 8\pi\nu^2/c^3. \quad (4.19)$$

Bringing everything together,³⁵ one obtains for the s th multiple:

$$n_{v,s} = \frac{8\pi\nu^2}{c^3} \frac{1}{s} e^{-sh\nu/kT} \quad s = 1, 2, 3, \dots \quad (4.20)$$

³⁴As mentioned in footnote 30, we chose without lack of generality the statistical weights of states 1 and 2 to be the same and equal to one, since they anyway will not appear in the final results.

³⁵With eq. (4.19), from eq. (4.17), one obtains $n_{v,1}$. Inserting it into eq. (4.14) yields $n_{v,2}$, and so on.

These number densities are formally identical to the ones obtained by Ishiwara and Wolfke through the expansion of Planck's formula³⁶ (Ishiwara 1912; M. Wolfke 1921). Therefore, summing up all the terms has to yield Planck's formula again. However, as pointed out above, Bothe's light multiples are correlated light quanta (quasiparticles), whereas Wolfke interpreted them as real objects.

Citing Lorentz (1916, 59), Bothe began the introduction to the publication with the remark that the energy variation (variance) of a black body holds the key to understanding radiation theory. Therefore, unlike Wolfke (1921), it was essential for him to address this topic as well. To do so, he extended Einstein's analogy between an ideal gas and the light quantum gas consisting of independent constituents to the s -fold multiples. Consequently, he assumed that the average number of s -fold multiples $n_{\nu,s}$ fluctuates in time in a Gaussian way as $\sqrt{n_{\nu,s}}$. Thus, the mean energy fluctuation squared of an s -fold multiple (energy $sh\nu$) is $\overline{\Delta_{\nu,s}^2} = (sh\nu)^2 n_{\nu,s}$. Summing over all s and using eq. (4.19), one obtains as variance

$$\overline{\Delta_{\nu}^2} = \sum_{s=1}^{\infty} \overline{\Delta_{\nu,s}^2} = \sum_{s=1}^{\infty} (sh\nu)^2 n_s = (h\nu)^2 \sum_{s=1}^{\infty} s^2 n_{\nu,s} = \frac{8\pi h^2 \nu^4}{c^3} \frac{e^{h\nu/kT}}{(e^{h\nu/kT} - 1)^2}. \quad (4.21)$$

Finally, by replacing the temperature T with u_{ν} , the energy density per frequency interval, Planck's formula, see eq. (4.1), one indeed obtains Einstein's fluctuation formula (Einstein 1917)³⁷:

$$\overline{\Delta_{\nu}^2} = h\nu \cdot u_{\nu} + \frac{u_{\nu}^2}{Z_{\nu}}. \quad (4.22)$$

Bothe demonstrated at that time that the concept of light multiples allows a correct reproduction of Einstein's fluctuation formula, in contrast to both classical theory and to simple light-quantum concepts. Even a wave concept was unnecessary to obtain the correct radiation (Planck) and fluctuation (Einstein) formulas, as long as spontaneous emission as well as phase-correlated induced emission and absorption are properly taken into account.

Nevertheless, there was at least one point that was not understood: why are there no $s = 0$ contributions in the expansion of Planck's formula, see eq. (4.6), or in other words, why must the spontaneous decay be treated separately? This

³⁶See eq. (4.9) in this contribution.

³⁷For the definition of Z_{ν} in the equation below, see eq. (4.13).

hampered an interpretation of the light quantum multiples in physical terms. We now address this question.

4.3.2 Bothe 1925a

Bothe resumed this topic a few years later in an as-yet unpublished manuscript entitled “Zur Struktur der Strahlung”³⁸ The manuscript is undated, but he worked on this until at least February 1925 when Planck, in a meeting of the Prussian Academy of Sciences, presented a paper “Zur Frage der Quantelung einatomiger Gase” (Planck 1925), which Bothe duly cited. Bose’s seminal paper had been published in the December 1924 issue of *Zeitschrift für Physik* (Bose 1924), just a couple of months earlier.

The main aim of the unpublished work was to find a bridge between the quantum picture and the wave picture of radiation. Bothe began the text with the following sentences:³⁹

Einstein’s derivation of the radiation formula [variance of energy in a black body] (Einstein 1917) stands out due to its outstanding clarity. If one attempts to develop it further [...], one obtains the strange picture that the light quanta within the stationary radiation field are in general not independent of each other, but rather bunched. Only by accounting for the bunching of the light quanta does one arrive at the correct expression for the energy fluctuations of the radiation field (Bothe 1923; M. Wolfke 1921). Unfortunately until now, no one has succeeded in building a bridge from this point to the classical wave theory, e. g., in specifying the number of degrees of freedom within a radiation volume, or even in interpreting the interferences.^{40,41}

³⁸Bothe (ca. 1925). *Zur Struktur der Strahlung*, unpublished manuscript, AMPG III/6/105.4.

³⁹It is surprising that Bothe cited Wolfke, even though Wolfke did not discuss the energy variation in his paper at all.

⁴⁰See footnote 38.

⁴¹“Die Einsteinsche Herleitung der Strahlungsformel (Einstein 1917) zeichnet sich bekanntlich durch außerordentliche Anschaulichkeit aus. Versucht man diese weiterzubilden, ohne ihr diesen Charakter zu nehmen, so gelangt man zwangsläufig zu der merkwürdigen Vorstellung, dass im stationären Strahlungsfelde die Lichtquanten im allgemeinen nicht unabhängig voneinander, sondern zu Aggregaten vereinigt sind. Durch Berücksichtigung dieser Assoziationen der Lichtquanten gelangt man erst zum richtigen Ausdruck für die Energieschwankungen im Strahlungsfeld (Bothe 1923; M. Wolfke 1921). Leider ist es bisher nicht gelungen, von diesem Punkte aus die Brücke zu der klassischen Wellentheorie der Strahlung zu schlagen, z.B: die Anzahl der Freiheitsgrade eines Strahlungsvolumens richtig anzugeben, oder gar die Interferenzen zu deuten.” (The bibliographic references in parentheses were footnotes in Bothe’s original manuscript).

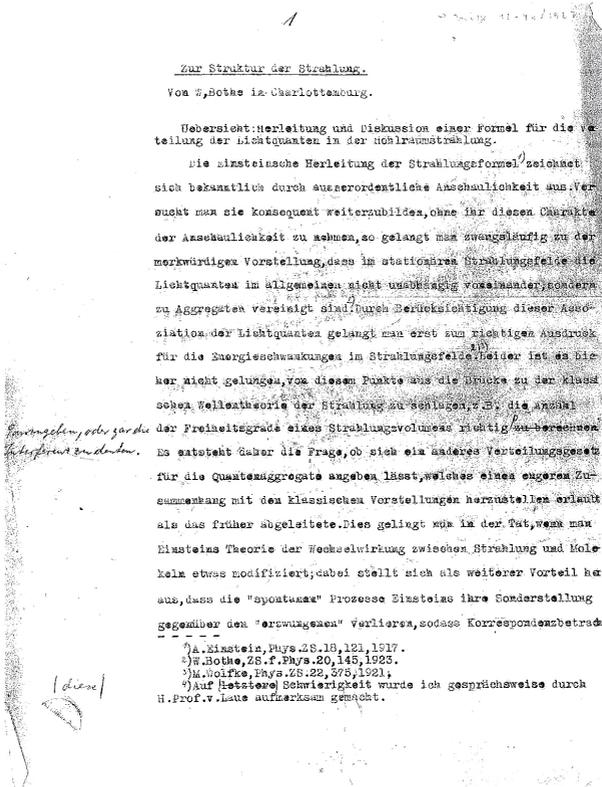


Figure 4.3: Bothe manuscript, (n.d., ca. 1925). "Zur Struktur der Strahlung." AMPG, III/6/105.4.

As we will see below, this drawback is closely connected to the missing $s = 0$ term in the expansion of Planck's formula in terms of Wienian-type s -fold light multiple energy distributions, see eqs. (4.6), (4.7), (4.20).

Bothe's main idea was to use a heuristic argument to put the spontaneous and induced emission on equal footing. With a black-body radiation field consisting of s -fold light multiples according to sec. (4.2) in mind, he started his consid-

erations again by questioning if and how thermal equilibrium is reached for his favored model system, the interaction of black-body radiation with two-level objects (atoms, molecules). He examined two processes:

1. The object is excited through interaction with an $(s + 1)$ -fold multiple from the energetically lower state with energy ϵ_1 into the energetically higher state with energy ϵ_2 , whereby the $(s + 1)$ -fold light multiple of frequency $\nu = (\epsilon_2 - \epsilon_1)/h$ is transformed into an s -fold light multiple of the same frequency ν . Using additional Boltzmann statistics for the population of the states 1 and 2, see eq. (4.18), the incidence of such a process is, as before, proportional to $(s + 1)n_{\nu,s+1}e^{-\epsilon_1/kT}$, since $(s + 1)n_{\nu,s+1}$ is the fraction of the total radiation density u_ν , which belongs to $(s + 1)$ -fold multiples.⁴²
2. The reverse process, the de-excitation of an excited molecule through interaction with an s -fold light multiple, transforms it into an $(s + 1)$ -fold one. To avoid a special role of the spontaneous transitions with respect to the induced ones, see section (4.3.1), Bothe now assumed on heuristic grounds that these transitions are proportional to $(s + 1) \cdot n_{\nu,s}$ instead of $s \cdot n_{\nu,s}$, surmising correctly, as early as 1925, a small part of modern field theory.⁴³ The incidence of these processes is then proportional to $(s + 1)n_{\nu,s}e^{-\epsilon_2/kT}$. Bothe himself commented on his ansatz very cautiously:

The addend $+1$ in the bracket replaces to some extent the spontaneous transitions, since it allows now also for emission processes which start from a zero quantum multiplet.^{44,45}

One benefit of this is that zero-fold multiples are formally allowed and would today be termed a “physical vacuum.” Thanks to Paul Dirac’s seminal work (Dirac 1927, 251/261, eq. 10), the factor $(s + 1)$ in front of terms describing the creation of light quanta (and of Bosons in general) is absolutely necessary to guarantee a symmetric wave function.⁴⁶

To achieve thermal equilibrium, both rates have to be equal, yielding

$$n_{\nu,s+1} = n_{\nu,s}e^{-h\nu/kT}, \quad (4.23)$$

⁴²Since the number density of the s -fold light multiples differs in definition from the ones of the previous section, we denote them now with a gothic n instead of a latin n .

⁴³For an early interpretation of Bothe’s choice, see Jordan’s (1928, in the very last paragraph of section I) discussion.

⁴⁴Bothe’s unpublished manuscript p. 3 (see footnote 38).

⁴⁵“Der Summand $+1$ in der Klammer ersetzt gewissermaßen die spontanen Übergänge, da er z.B. auch nullquantigen Aggregaten die Auslösung von Emissionsprozessen ermöglicht.”

⁴⁶This fact was, for example, discussed highly pedagogically in volume III of the *Feynman Lectures on Physics* (Feynman, Leighton, and Sands 1965, Chap. 4.4, Emission and absorption of photons).

for all $s \geq 0$, from which

$$n_{\nu,s} = n_{\nu,0} e^{-sh\nu/kT} \quad (4.24)$$

immediately follows. Before we report on the consequences of Bothe's ansatz, we anticipate a small detail of Bothe's 1927 publication (Bothe 1927b) discussed at the end of section (4.3.3). This detail concerns the ratio δ between the total rate of induced and spontaneous transitions in a black body. Drawing from Bothe's less precise interpretation of the two parts in the factor $(s + 1)$ (see quotation above), he obtained

$$\delta = \frac{\sum_{s=0}^{\infty} s \cdot n_{\nu,s}}{\sum_{s=0}^{\infty} 1 \cdot n_{\nu,s}} = \frac{1}{e^{h\nu/kT} - 1} \quad (4.25)$$

by means of eq. (4.24) for this ratio.

More recently, δ is called "degeneracy parameter."⁴⁷ It approaches "0" (no induced transitions) for the extreme Wien limit ($h\nu/kT \rightarrow \infty$) and "1" for the extreme Raleigh-Jeans limit ($h\nu/kT \rightarrow 0$).

In contrast to Wolfke's (1921), de Broglie's (1922b) and his own previous considerations (Bothe 1923), Bothe used his heuristic trick and found a possibility to also formally treat ($s = 0$)-multiples. Furthermore, using the last eq. (4.25), he gained an advanced interpretation of the total number density n_{ν} of light multiples, per frequency interval $d\nu$, including the zero-quantum ones. To reach this point, he identified $n_{\nu} d\nu V$, the total number of light multiples per frequency interval $d\nu$ and volume V with the number of degrees of freedom of a black body Z_{ν} , see eq. (4.13):⁴⁸

$$\sum_{s=0}^{\infty} n_{\nu,s} = \frac{Z_{\nu}}{V d\nu} = \frac{8\pi\nu^2}{c^3} = \underbrace{\sum_{s=0}^{\infty} n_{\nu,0} e^{-sh\nu/kT}}_{= n_{\nu}} \quad (4.26)$$

Performing the sum within the underlined part of the equation yields:

$$n_{\nu,0} = \frac{8\pi\nu^2}{c^3} \frac{1}{\sum_{s=0}^{\infty} e^{-sh\nu/kT}} = \frac{8\pi\nu^2}{c^3} (1 - e^{-h\nu/kT}), \quad (4.27)$$

⁴⁷One year later, Reinhold Fürth (1893–1979) discussed the same issue in other contexts (Fürth 1928a; 1928b).

⁴⁸We discuss the physical implications of this assumption in the following section, after obtaining formal results (4.3.3).

and thus

$$n_{\nu,s} = \frac{8\pi\nu^2}{c^3} (1 - e^{-h\nu/kT}) e^{-sh\nu/kT}, \quad s = 0, 1, 2, \dots \quad (4.28)$$

This equation can be rewritten as

$$n_{\nu,s} = \frac{8\pi\nu^2}{c^3} \cdot \frac{e^{-sh\nu/kT}}{\sum_{s=0}^{\infty} e^{-sh\nu/kT}}, \quad (4.29)$$

a required expression if the s -fold light multiples in a black body are distributed thermodynamically according to Boltzmann's law.⁴⁹ Thus the n_s can now be interpreted thermodynamically as occupation number densities for the s -fold multiples in a black body. Bothe was probably aware of this since he commented on his result as follows:

This is the new distribution law of the quantum multiples. From it one can easily derive the total radiation energy u_ν at frequency ν [...], that is Planck's formula.^{50,51}

$$u_\nu(T) = \sum_{s=0}^{\infty} sh\nu \cdot n_{\nu,s} = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{e^{h\nu/kT} - 1}. \quad (4.30)$$

After having obtained this result, Bothe reminds the reader that, following Debye (1910), the number of degrees of freedom of a black-body radiator at frequency ν is nothing but the number of eigenmodes of the black body at that frequency ν . Each of these eigenmodes can be replaced by a Planck oscillator of frequency ν , and the number of eigenmodes with energy $sh\nu$ can be identified according to the quantum rules with $n_{\nu,s}$. Indeed, eq. (4.28) is identical to the statistical distribution law of the energies of Z_ν Planck oscillators. He ends with the remark:

After this, we can interpret each quantum multiple in black-body radiation as the energy of an eigenmode of the black body.^{52,53}

⁴⁹Einstein used this kind of reasoning for the first time in his 1911 talk at the Solvay Conference (Einstein 1913).

⁵⁰Bothe's unpublished manuscript p. 3 (see footnote 38).

⁵¹"Dies ist das neue Verteilungsgesetz der Quantenaggregate. Aus ihm ergibt sich leicht die gesamte Strahlungsenergie E_ν von der Frequenz ν [...] d.i. die Plancksche Formel."

⁵²See footnote 38.

⁵³"Wir können hiernach also jedes Quantenaggregat in der Hohlraumstrahlung deuten als die Energie einer Eigenschwingung des Hohlraums." Sentence underlined on p. 10 in Bothe's manuscript.

In 1912, Ishiwara already had a similar interpretation in mind (Ishiwara 1912), starting from the power expansion of Planck's formula, see eq. (4.6), and relying on Debye's (1910) interpretation of phase space quantization. However, he did not get very far, since at that time in the expansion of Planck's formula ($s = 0$)-terms were only previously discussed by Einstein at the 1911 Solvay Conference (Einstein 1913).⁵⁴

Bothe finally returns to the question of whether the correct expression of the energy fluctuations in a black body follows from the distribution law in eq. (4.28). He did not doubt this since the distribution law in eq. (4.28) correctly yields Planck's formula. Nevertheless, he demonstrates this explicitly in a few lines.⁵⁵

4.3.3 Bothe 1927

One can only speculate why Bothe did not publish the manuscript at the beginning of 1925. One reason was certainly his workload at the PTR. Hans Geiger (1882–1945), the director of the laboratory for radioactivity, had left the PTR around that time, and Bothe succeeded him. In Germany, the PTR was responsible for all official calibrations of radioactive samples.

The appearance of the Bose statistics and its implications might also have caused a delay in the publication of the manuscript. Moreover, apart from his daily duties, between 1926 and 1927, Bothe performed two crucial coincidence experiments which dealt with resonance fluorescence of X-rays (Bothe 1926) and later with Compton scattering in an X-ray interference field (Bothe 1927a). Both experiments were highly complex and very time consuming.⁵⁶ It might therefore be that Bothe simply did not have enough time to finish the manuscript.

Using Compton scattering, Bothe found in the latter experiment that light quanta from an interference field still carry a momentum of $h\nu/c$. Like Einstein,⁵⁷ he considered his result to be extremely important, as we know from the handwritten notes for a talk he presented at the end of October 1926.⁵⁸ At that time, he discussed standing waves of an interference field, which definitely represents a classical situation with a vanishing energy and momentum transport in

⁵⁴See also footnote 49.

⁵⁵Quite recently one of the authors (D. Fick) presented an analysis of Bothe's 1925 heuristic assumption (Fick in print), on which sections 4.3.1 and 4.3.2 are based.

⁵⁶For details, see the extended report by Fick and Kant, "Walther Bothe's Contributions to the Understanding of the Wave-Particle Duality of Light" (Fick and Kant 2009).

⁵⁷A brief report on a talk by Einstein presented at Berlin University (renamed Humboldt University in 1948) entitled "Theoretisches und Experimentelles zur Frage der Entstehung des Lichtes" (Anonymous 1927) explicitly mentioned this experiment. However, there were also critical questions concerning its relevance (Kirchner 1930, 467–468).

⁵⁸Bothe (ca. 1926). *Lecture Notes*, unpublished, AMPG III/6/105.3.

total. Light quanta in an interference field, as well as in a directed (needle) beam, nevertheless possess a momentum of $h\nu/c$.

Gathering the main ideas of his unpublished manuscript, Bothe reacted to this situation with a short publication, again reinterpreting the concept of light multiples (Bothe 1927b). Since a monochromatic unidirectional needle beam is a non-physical object, he used von Laue's (1914) light bundles of a finite cross section and of a very small frequency and opening angle interval as "elementary cells," in which the light quanta are embedded. He considered these bundles, together with the embedded quanta, as independent entities of the radiation statistics. Each bundle could contain an arbitrary amount of light quanta (including zero), whereby the number of elementary bundles per volume V and frequency range $d\nu$ was identical to the number of degrees of freedom $Z_\nu = (8\pi\nu^2/c^3)V \cdot d\nu$, see eq. (4.26). In this way, Bothe followed the new Bose statistics, which led to a sort of coupling between the light quanta, as had already been noted by a few others (Einstein 1924; Ehrenfest 1925; Landé 1925). With this interpretation in mind, Bothe demonstrated that not only the number of light quanta for a black body in thermal equilibrium is stationary but so is their grouping.

As frequently emphasized, Bothe considered the understanding of Einstein's fluctuation formula, see eq. (4.22), to be extremely important. He therefore concluded this publication with an alternative derivation which used the dual concept of light directly. Since its formal procedure is only of minor interest here, we outline the main idea and present the final result. Following Lorentz (1916), Bothe began with the remark that in an elementary bundle, the (classical) wave energy ϵ fluctuates around its average $\epsilon_0 = (u_\nu V \cdot d\nu)/Z_\nu$ with a variance (root mean square fluctuation) δ_{wave} equal to the average wave energy itself.⁵⁹

$$\overline{\delta_{wave}^2} = \epsilon_0^2. \quad (4.31)$$

In classical physics, this would be the total fluctuation of the averaged energy. However, if one assumes that the wave energy is distributed discontinuously over "light particles" (quanta), the number of light particles i itself fluctuates around its average value $i_0 = \epsilon/h\nu$ with a variance of i_0 if a Poisson distribution is assumed. The variance of the energy due to the fluctuation of the number of light particles is therefore $i_0(h\nu)^2 = \epsilon h\nu$. Furthermore, averaging over the energy ϵ itself, one obtains for the variance (root mean square fluctuation) of this contribution:

$$\overline{\delta_{particles}^2} = \epsilon_0 h\nu. \quad (4.32)$$

⁵⁹See also (Jordan 1929), in particular, the discussion of eqs. 9 and 10.

The tricky aspect of the problem now is that both variances are coupled, since the fluctuation of the number of particles depends on the fluctuation of the (wave) energy. Nevertheless, in an appendix Bothe proved that these two variances simply add to each other:

$$\overline{\delta^2} = \overline{\delta_{particles}^2} + \overline{\delta_{wave}^2} = \epsilon_0 h\nu + \epsilon_0^2. \quad (4.33)$$

Multiplying the variance $\overline{\delta^2}$ with the number of independent elementary bundles per volume and frequency interval $(Z_\nu)/(V \cdot d\nu)$, one obtains the known result, see eq. (4.22), which we repeat here for easy comparison:

$$\overline{\Delta_\nu^2} = h\nu \cdot u_\nu + \frac{u_\nu^2}{Z_\nu}. \quad (4.34)$$

As Bothe pointed out, this equation followed from a systematic treatment of the dualistic nature of light and does not follow from a plain addition of the fluctuations of the separate results that treat light first as a wave and then as a light particle. As we now know, Poissonian distributed particle numbers, for example light quanta in a black body, always end up in energy fluctuations “expressible as the sum of contributions from the fluctuations of classical particles and the contributions of classical wave fields” (Mandel and Wolf 1965, 271, eq. 6.19 and its discussion; Born 1949, 79–82).⁶⁰

With this manuscript, Bothe completed his research on the nature of light quanta for the rest of his scientific career. He never returned to the concept of light multiples. Nevertheless, one aspect of Bothe’s manuscript stood the test of time and is still cited today, his discussion of the degeneracy parameter δ , see eq. (4.25). Due to its definition, it may also be considered the average number of light quanta that are in the same quantum state. There is a third interpretation of δ as a ratio between the multi-light quantum states ($s \geq 2$) and the one-light quantum states ($s = 1$) of thermal radiation

$$\delta = \frac{\sum_{s=2}^{\infty} s \cdot n_{\nu,s}}{n_{\nu,1}} = \frac{1}{e^{h\nu/kT} - 1}. \quad (4.35)$$

Multi-light quantum states ($s \geq 2$) are generated by various sequences of induced transitions. According to Einstein (1917), light quanta from induced transitions are perfectly correlated in direction, energy, phase and polarization. Due to the

⁶⁰To make this clear, we deliberately denoted the two contributions to eq. (4.33) with “wave” and “particle” and not as Bothe did with “wave” and “quantum.” See also “Reconstruction of and commentary on Jordan’s derivation of Einstein’s fluctuation formula” in (Duncan and Janssen 2008, sec. 3).

latter interpretation of δ , it is therefore comprehensible that correlation functions in space are proportional to the degeneracy parameter δ .⁶¹

In their seminal experiments during 1957–1958, Robert Hanbury Brown and Richard Twiss were the first to provide experimental evidence of space correlations between the outputs of two photoelectric detectors illuminated by partially-correlated light waves of a mercury lamp in a Michelson configuration. In such experiments, the magnitude of the correlation coefficient is proportional to δ .⁶²

We end this section with the remark that in 1958 Hanbury Brown and Twiss determined the correlation of light of the star Sirius A, a black body of about 10,000K surface temperature for the first time (Mandel and Wolf 1995, Chap. 9.10 and references therein). The observation of correlations is an experimental proof that the concept of light molecules à la Ishiwara (1912), Wolfke and the early Bothe, see secs. (4.2.1) and (4.3.1), is useless. It has long been known that the expansion of Planck's formula into a sum of the Wienian terms $u_{\nu,s} = (8\pi\nu^2/c^3)h\nu e^{-sh\nu/kT}$, $s = 1, 2, 3, \dots$ (eqs. (4.6), (4.7)) can be interpreted in terms of thermodynamically-independent objects (light molecules) with energy $sh\nu$, since their partial entropies ε_i add up to the total entropy \mathfrak{E} of a black body (sec. (4.2.1)). If this holds, then correlations of light from such an object should not exist; the light should be completely incoherent and thus unable to produce any interference effects at all (Laue 1906).^{63,64}

⁶¹For an extended introduction, see the lucid description of the physics involved in sec. 4.2 of (Mandel and Wolf 1965).

⁶²References to these experiments and as well its semiclassical as its field theoretical treatment can be found in (Mandel and Wolf 1995, sec. 9.9 and 14.6.1).

⁶³Because of their brevity and beauty, we repeat here von Laue's arguments, which refer partly to Planck: "The entropy \mathfrak{E} of a system is [...] connected with its probability W through the equation $\mathfrak{E} = k \cdot \log W$. For two subsystems 1 and 2, which add up to the total one, the partial entropies are accordingly: $\mathfrak{E}_1 = k \cdot \log W_1$ and $\mathfrak{E}_2 = k \cdot \log W_2$. Out of these three equations, the addition theorem for entropies $\mathfrak{E} = \mathfrak{E}_1 + \mathfrak{E}_2$ ensues if and only if $W = W_1 \cdot W_2$. A complete independence of each of the two systems would be a necessary and sufficient condition. [...] coherent light bundles are, however, not independent of each other at all. Therefore the addition theorem is not valid" (Laue 1906, 374, fn. 4). "(Die Entropie \mathfrak{E} eines Systems ist [...] mit seiner Wahrscheinlichkeit W durch die Gleichung $\mathfrak{E} = k \cdot \log W$ verknüpft. Für zwei Teilsysteme, aus denen das ganze bestehen soll, gilt entsprechend $\mathfrak{E}_1 = k \cdot \log W_1$ und $\mathfrak{E}_2 = k \cdot \log W_2$. Aus diesen drei Gleichungen folgt das Additionstheorem dann und nur dann, wenn $W = W_1 \cdot W_2$ ist; notwendige und hinreichende Bedingung dafür ist, dass die beiden Teilsysteme voneinander vollkommen unabhängig sind, [...]. Kohärente Strahlenbündel sind aber nicht voneinander unabhängig, also gilt für sie das Additionstheorem nicht.)"

⁶⁴Further work on this topic was published at around this time by von Laue (1907a; 1907b) and Ehrenfest (1925, 364, fn. 1). See also sec. 4.7 "Entropy of an Optical Field" in (Mandel and Wolf 1965) for further historical information.

4.4 Obituary

In 1930, Indian theorists R. J. Majumdar and D. S. Kothari published a thorough derivation of the ratio of Einstein's A and B coefficients, following the ideas of Bose's brief first 1924 manuscript (Bose 1924). In passing, they also mention Bothe's considerations to obtain this ratio through the requirement of achieving thermal equilibrium in a black body (Bothe 1923). Here, we will not deal with their work, but instead with their remark that "Bothe created the fairly new concept of light molecules and arrived thus at Planck's formula" (Majumdar and Kothari 1930).⁶⁵ It was probably common knowledge, at least in India, that the Indian theorist Kulesh Ch. Kar also shared the same opinion (Kar 1927). Wolfke, now a professor at the Technical University in Warsaw, reacted angrily in a short note (M. Wolfke 1930) listing all the authors who had used the concept of "light atoms" or "light molecules": Ioffe (1911), Wolfke (1921), de Broglie (1922b), Bateman (1923) and Bothe (1923).

Around this time, Wolfke still maintained his relations with Einstein. The Einstein Archive Online lists a total of nine letters exchanged between Wolfke and Einstein from 1925 to 1931. Wolfke's letter dated 13 December 1925 refers to a recent meeting with Einstein in Berlin.⁶⁶ Some of the other letters deal with Wolfke's ideas to experimentally find a signal pointing to the existence of (static) light molecules. Moreover, at Wolfke's request, Einstein presented an investigation entitled "Über die statistischen Eigenschaften der Strahlung" (Einstein 1930) at a meeting of the Prussian Academy of Sciences in which Einstein arrived at a fluctuation formula with a factor two in front of the "wave part", see eq. (4.22). The corresponding manuscript never appeared.

In two letters dated 27 and 29 December 1930,⁶⁷ Wolfke tried to corroborate Einstein's result with his own calculations. In his second letter, he was almost headed in the right direction. He started with Bose's expression for the occupation probability of a phase space cell with s light quanta (Bose 1924) and identified this probability with the number density n_s of s -fold light multiples with energy $sh\nu$.⁶⁸ By this means, he arrived at Bothe's expression for n_s , see eq. (4.28), however, with one essential difference. According to his interpretation of the s -fold multiples in terms of static light molecules, he assumed that s runs only from

⁶⁵"Bothe schuf das einigermaßen neue Konzept der Lichtmoleküle und gelangte so zur Planck'schen Formel."

⁶⁶Wolfke to Einstein, 13 December 1925, 23 507, Einstein Archive Online.

⁶⁷Wolfke to Einstein, 27 and 29 December 1930, 23 517 and 23 519, Einstein Archive Online.

⁶⁸Since this approach is similar to Bothe's unpublished manuscript, see footnote 38 in sec. (4.3.2), we denote the number densities again using fractured letters.

1 to ∞ , instead of from 0 to ∞ as Bothe interpreted. Calculating the total energy density by

$$u_\nu(T) = \sum_{s=1}^{\infty} sh\nu \cdot n_{\nu,s} \quad (4.36)$$

and its averaged energy fluctuation density squared $\overline{\Delta_\nu^2}$ through (Bothe 1923, eq. 1)

$$\overline{\Delta_\nu^2} = \frac{1}{V} kT^2 \left(\frac{du_\nu}{dT} \right) \quad (4.37)$$

he obtained the correct result for the total energy density, but a factor of two in front of the wave part of the averaged energy fluctuation density squared, see eq. (4.22), as had Einstein. One can straightforwardly identify this odd result with the missing $s = 0$ term and thus with the interpretation of the n_s . Wolfke drew the proper conclusion “[...] that on the basis of the Bose – Einstein statistics it is not permissible to interpret the radiation field as consisting of *multiple quanta*.”⁶⁹

Einstein probably made the same mistake, since we know from a brief remark at the bottom of Wolfke’s 29 December 1930 letter that he calculated the energy fluctuations assuming a mixture of locally independent multiple quanta.⁷⁰ After returning to Berlin, Einstein responded in detail in a letter dated 10 April 1931⁷¹ by saying “At first, I was very impressed by your argument. But then I found a problem.”⁷² Einstein then showed explicitly that the concept of static light molecules violates Boltzmann’s law.⁷³ No reply from Wolfke is known.

The story of “light molecules” related here ends finally in 1946. Wolfke survived the German occupation of Poland in Warsaw. Here he taught, partly at a lower level, at the Polytechnic University where he was a professor. After the war, he was eager to resume his contacts with the West, in particular to Sweden, Switzerland and, if possible, to the United States. From Stockholm, he sent a manuscript entitled “Über Multiphotone”⁷⁴ to Einstein in Princeton asking him for critical remarks. Wolfke, still drawing from his 1921 publication on black-body radiation (M. Wolfke 1921), showed that the number of phase cells that

⁶⁹ “[...] , dass es auf Grund der Bose-Einsteinschen Statistik nicht zulässig ist das Strahlungsfeld, als aus “Mehrfachquanten” bestehend, zu interpretieren” (see footnote 67).

⁷⁰ Wolfke to Einstein, 29 December 1930, 23 521, Einstein Archive Online.

⁷¹ Einstein to Wolfke, 10 April 1931, 23 522, Einstein Archive Online.

⁷² “Ihr Argument hat zu erst großen Eindruck auf mich gemacht. Aber ich finde eine Schwierigkeit.”

⁷³ This letter prompted us to explicitly show that Bothe’s interpretation of the $n_{\nu,s}$ as quantum multiples, including the $s = 0$ ones, is in accordance with Boltzmann statistics.

⁷⁴ “About Multiphotons”, APAN: III–71, 23, 5–12. According to the then general use of the word “photon,” rather than “light quantum,” he used the phrase “multiphoton” rather than “light molecule” (M. Wolfke 1946a).

contain s quanta of energy $h\nu$ equals the number density of multiphotons n_s with energy $sh\nu$, see eq. (4.9) according to Bose-Einstein statistics.⁷⁵ Subsequently, he discussed potentially feasible experiments using the light emitted by black bodies of extremely high temperatures, such as fixed stars, to enhance the ratio of $s = 2$ to $s = 1$ multiphotons, see eq. (4.10).

4 1

Stockholm. Sweden.
Odengatan 46.
Larssons-Pensionat.

Stockholm, d. 10. VII. 1946.

Herrn Prof. Dr. A. EINSTEIN,
Princeton N.Y. U.S.A.

Sehr geehrter Herr Professor!

Bezugnehmend auf unsere Korrespondenz im Jahre 1930 über die "Lichtmoleküle", erlaube ich mir Ihnen die beiliegende Abhandlung zur gefälligen Einsichtsnahme zu übersenden. Ich wäre glücklich, wenn Sie dazu Zeit und Lust finden würden mir Ihre Meinung mitteilen zu wollen.

Ich bin in Stockholm auf Durchreise, bleibe aber hier noch einige Wochen.

Mit vorzüglicher Hochachtung

Wolke

Figure 4.4: Letter from Wolfke to Einstein dated 10 July 1946, APAN, III-71/IV. 23,5.

⁷⁵Wolfke neglected—for whatever reason—Bothe's (1927b) dynamical treatment and thus the fact that Bothe obtained the same result almost two decades earlier.

In September 1946, Wolfke presented his ideas at a meeting of the Schweizerische Naturforschende Gesellschaft in Zurich.⁷⁶ Previously, on 18 January, de Broglie⁷⁷ had already presented a French version of the manuscript⁷⁸ at the meeting of the Académie des Sciences, Paris, which was published in *Comptes Rendus* (M. Wolfke 1946c).

Einstein answered Wolfke's letter one week later on 17 July 1946.⁷⁹ In this correspondence, Einstein argued against the existence of "light molecules" in discussing the passage of light molecules through a semi-reflecting plate. In doing so, he elaborated at length his argument from his 1909 letter to Lorentz⁸⁰ that "a light ray splits, but a light quantum cannot split without a change in frequency." Obviously Einstein had forgotten his much more convincing argument from his 1931 letter to Wolfke that the existence of static light molecules is not in accordance with Boltzmann's law.

Wolfke replied to Einstein's letter on 17 August 1946 without really responding to Einstein's critical remark. Wolfke passed away in Zurich on 4 May 1947 after suffering a heart attack. After Bothe's withdrawal in 1927, this incident finally brought to an end the discussion on the concept of "light molecules" as a tool for dealing with the correlations in a light quantum gas.

Appendix

Frequency Continuum and Light Molecules

Wolfgang Pauli 1923

In 1923, Wolfgang Pauli (1900–1958) tried to understand whether, and if so how, free electrons with a Maxwellian velocity distribution can achieve thermal equilibrium in a black-body radiation field (Pauli 1923). He assumed that the (relativistically-treated) Compton effect is the mechanism which drives the system into equilibrium, since it allows for a change of the energy and momentum of both light quantum and electron. Of course, the relativistically-formulated conservation laws of energy and momentum had to be fulfilled as well. He identified statistical equilibrium through the condition that each elementary process occurs

⁷⁶APAN: Wolfke manuscript, 1941, III–71, 20, 10–11. In a report to his authorities in Warsaw, Wolfke writes: "On 8. IX. [1946] I lectured at a meeting of the Schweizerische Naturforschende Gesellschaft in Zurich 'On multiple quanta in Planck's radiation,'" an abbreviated version of which was published in 1946 (M. Wolfke 1946b).

⁷⁷Wolfke probably had personal relations with de Broglie, who visited Wolfke's institute in Warsaw in 1933 (K. Wolfke 1980).

⁷⁸APAN: III–71,20, 10–11.

⁷⁹Einstein to Wolfke, 17 July 1946, APAN, III–71,24, 7.

⁸⁰See the beginning of sec. (4.1).

as often as the time-reversed process. To substantiate this requirement, Pauli devoted quite some effort to the formal description of what a time-reversed process means relativistically.

Apart from such dynamical variables as the momenta of electron and photon, and apart from the solid angle, the rates depend then on a “rate function” (*Häufigkeitsfunktion*) F which ought to depend explicitly on u_ν , the spectral density of an arbitrary radiation field at frequency ν with which the electrons are interacting. At first Pauli tested the plausible ansatz

$$F = Au_\nu, \quad (4.38)$$

whereby the function A ought to depend on the kinematical variables of electron and the light quanta, such as the momenta and energies, but be independent of temperature T . Pauli found that this ansatz leads unambiguously to Wien’s radiation law.

Guided by a qualitative analysis of this apparently limited ansatz, he analyzed the consequences of the, at first glance, rather strange ansatz

$$F = Au_\nu + Bu_\nu u_{\nu'}, \quad (4.39)$$

in which $u_{\nu'}$ denotes the radiation density at the frequency of the Compton-scattered light quantum. Pauli himself commented on this as follows: “At first glance this postulation seems odd, since an interaction of radiation bundles is assumed here, possibly with widely different frequencies”⁸¹ (Pauli 1923, 284). This ansatz causes the Compton scattering process $\nu \rightarrow \nu'$ to occur more frequently if the frequencies ν as well as ν' are present in the radiation field. And indeed, this leads to Planck’s radiation law for a black body.

Einstein and Ehrenfest reacted immediately in the next issue of *Zeitschrift für Physik* (Einstein and Ehrenfest 1923). Obviously, they had had access to Pauli’s manuscript in advance and had submitted their manuscript prior to the appearance of Pauli’s. In subsequent steps, Einstein and Ehrenfest employed a transparent formalism which allowed equilibrium conditions to be described for absorption and emission of radiation, even when the matter part has continuous energy spectra, as in Compton scattering, for example, or for moving atoms or molecules. By defining what “time reversal” in the statistical average might mean, Einstein and Ehrenfest arrived at transition probabilities that were nonlinear in the energy density of the radiation field, in general, up to any order.⁸²

⁸¹“Beim ersten Augenblick könnte diese Forderung befremdend erscheinen, weil hier eine Wechselwirkung von Strahlenbündeln mit unter Umständen weit verschiedenen Frequenzen angenommen wird.”

⁸²For further more recent discussions, see (H. R. Lewis 1973; Lanyi 2003).

Bothe 1924

Bothe also reacted to Pauli's investigations with a formulation of the problem in terms of quantum multiples. He aimed at a proof in which the "B-processes" in Pauli's sense, see eq. (4.39), are formally dispensable (Bothe 1924). We will not enter the formal details of his arguments since they were based on the light multiple concept without the $s = 0$ terms, which as we saw, is problematic, see secs. (4.2) and (4.3.1).

Bothe's final argument was based on two recollections:

1. Pauli (1923) showed that with the A-term alone, see eq. (4.38), the thermalization of a light-quantum gas in a black body through Compton scattering on a thermalized electron gas (Boltzmann distribution) will unambiguously lead to a Wienian energy distribution.
2. Referring to Wolfke (1921), see sec. (4.2.1), de Broglie (1922a), see sec. (4.2.2), and to his own work (Bothe 1923), see sec. (4.3.1), he reminded the reader that Planck's formula may be written as an infinite sum of Wienian terms, each describing an s -fold light multiple:

$$u_{\nu,s} = \frac{8\pi\nu^2}{c^3} h\nu e^{-sh\nu/kT}, \quad s = 1, 2, 3, \dots \quad (4.40)$$

Thus it will suffice to deal with Compton scattering of s -fold quantum multiples as a whole, taking into account only A-processes, see eq. (4.38). Formally, such a process might be written as $sh\nu + e \rightarrow sh\nu' + e'$. The number of s -fold multiples will be conserved and only their frequency will vary. One just has to formally replace in Pauli's calculations (Pauli 1923, 281–282) h with sh , which will result with Wienian terms in $u_{\nu',s}$, the sum of which will yield Planck's formula again and thus thermalized black-body radiation.

Bothe then added a more formal and detailed support of this brief argument. He finished the paper with the remark that in Pauli's derivation, as well as in his own, the number of light quanta is conserved. Moreover, the achievement of thermal equilibrium is independent of the prefactor $8\pi\nu^2/c^3$ in Planck's law.⁸³

Bothe 1925b

In his unpublished manuscript,⁸⁴ Bothe set about remedying this problem. He used the heuristic trick described in sec. (4.3.2). Still using the concept of s -fold

⁸³We take this opportunity to point to Pauli's article in *Handbuch der Physik*, where he discussed the heat balance between atoms and radiation in much more detail in sec. I.2. "Wärmegleichgewicht zwischen Atomen und Strahlung" (Pauli 1926, 9–22).

⁸⁴See footnote 38.

light multiples, he again assumed that the rate of processes in which an s -fold light multiple loses a light quantum is proportional to its energy density $s n_{v,s} h\nu$, but proportional to $(s + 1) n_{v,s} h\nu$ for those processes in which it gains one.⁸⁵

Apart from this assumption, he followed Pauli's strategy (Pauli 1923) and discussed the following two elementary steps in the interaction of light quanta with Maxwellian-energy-distributed electrons: in an interaction of an electron (momentum \mathfrak{G}) with an r -fold light multiple $r h\nu$ of frequency ν the latter loses one light quantum with momentum Γ which is added with momentum Γ_1 to a s -fold light multiple $s h\nu_1$ of frequency ν_1 . The electron scatters under an angle Θ_0 and thus takes care of the momentum and energy conservation ($\mathfrak{G} \rightarrow \mathfrak{G}_1$).⁸⁶ Applying once more the heuristic trick described in sec. (4.3.2), Bothe assumed that the rate of the process is proportional to the energy density $r n_{v,r} h\nu$ of the multiples which lose a light quantum, but proportional to $(s + 1) n_{v_1,s} h\nu_1$ for those which gain one. Denoting for a Maxwellian velocity distribution at temperature T the number of electrons per unit volume within the momentum interval $d\mathfrak{G}$ by $N d\mathfrak{G}$, the rate of these processes will be

$$H = B \cdot r n_{v,r} h\nu \cdot (s + 1) n_{v_1,s} h\nu_1 \cdot d\Gamma \cdot N d\mathfrak{G} \cdot d\Theta_0. \quad (4.41)$$

The strength B depends on all the relevant variables necessary to describe the process, except for the temperature T . The rate for the time reversed process can be parameterized accordingly

$$H_1 = B_1 (s + 1) n_{v_1,s+1} h\nu_1 \cdot r n_{v,r-1} h\nu \cdot d\Gamma_1 \cdot N_1 d\mathfrak{G}_1 \cdot d\Theta_0. \quad (4.42)$$

For the argument to be made, it is important only that B and B_1 do not depend on temperature T . To achieve thermal equilibrium it is sufficient that

$$H = H_1. \quad (4.43)$$

Pauli (1923, 281, eq. 25) deduced, from momentum and energy conservation for the electron-light quantum scattering, that the ratio N/N_1 obeys the relation

$$\frac{N}{N_1} = \frac{e^{h\nu/kT}}{e^{h\nu_1/kT}}. \quad (4.44)$$

⁸⁵As before, $s n_{v,s} h\nu$ denotes the fraction of the total radiation density that belongs to s -fold light multiples, see sec. (4.3.2).

⁸⁶All quantities are defined within the normal coordinate system of the individual process.

Now using the distribution for the number of multiples, see eq. (4.28)

$$n_{\nu,s} = \frac{8\pi\nu^2}{c^3} (1 - e^{-h\nu/kT}) e^{-sh\nu/kT}, \quad s = 0, 1, 2, \dots, \quad (4.45)$$

which we have already seen is in accordance with Boltzmann's law, one finds rather easily that the temperature disappears under the condition $H = H_1$, see eq. (4.43). That means thermal equilibrium will be reached independently of the very form of the functions B and B_1 .

We conclude this section of the appendix with an interpretation of the result in terms of light bundles of finite cross section and of very small frequency and opening angle interval as "elementary cells," in which the light quanta are embedded, see sec. (4.3.3). The process denoted by H , see eq. (4.41), describes a scattering process in which, out of an elementary bundle with r light quanta of frequency ν , one quantum is scattered into another bundle (called bundle 1) already hosting s light quanta with frequency ν_1 . The electron takes care of the energy and momentum conservation. H_1 describes the time-reversed process. It is now important that also ($r = 0$)-terms contribute to H , which means that in the bundles of H_1 , light quanta may appear spontaneously. For example, in the time-reversed process denoted by H_1 , a light quantum disappears completely for $r = 1$ and is created spontaneously for $s = 0$.

Abbreviations and Archives

APAN	Archive of the Polish Academy of Sciences, Warsaw
AMPG	Archive of the Max Planck Society, Berlin
Einstein Archive Online	Albert Einstein Archives and the David and Fela Shapell Digitization Project at the Jewish National & University Library, the Hebrew University of Jerusalem, http://www.alberteinstein.info/database/

Acknowledgements

We thank the staff at the following institutions for their continuous help and support: Archive of the Max Planck Society and the library of the Fritz-Haber-Institut in Berlin, the Archive of the Polish Academy of Science in Warsaw, and the library of the Physics Department of the Philipps-Universität in Marburg. Additionally, we acknowledge the indispensable help of Dr. Tilman Sauer from the

Einstein Papers Project in Pasadena (California, USA) and, last but not least, of our Polish colleagues Dr. Z. Lewandowski (Krakow), Prof. A. Kiejna (Wrocław), Prof. P. Zupranski (Warsaw) and Prof. W. Zych (Warsaw).

References

- Abraham, Max (1905). *Theorie der Elektrizität, Zweiter Band: Elektromagnetische Theorie der Strahlung*. Leipzig: B. G. Teubner.
- Afsenius, Sven-Åke (1975). Simple Approach to the Multiphoton States of a Thermal Radiation Field. *American Journal of Physics* 43: 882–884.
- Anonymous (1927). Report on talk by A. Einstein “Frage der Entstehung des Lichtes”. *Zeitschrift für Angewandte Chemie* 40: 546.
- (1930). News and Views. *Nature* 126: 660.
- Bateman, Harry (1923). On the Theory of Light Quanta. *Philosophical Magazine* 46: 977–991.
- Bergia, Silvio, Carlo Ferrario, and Vittorio Monzini (1985). Side Paths in the History of Physics: The Idea of Light Molecule from Ishiwara to de Broglie. *Rivista di storia della scienza* 2: 71–97.
- Born, Max (1949). *Natural Philosophy of Cause and Chance: The Waynflete Lectures 1948*. Oxford: Oxford University Press.
- Bose, Satyendra Nath (1924). Plancks Gesetz und Lichtquantenhypothese. *Zeitschrift für Physik* 26: 178–181.
- Bothe, Walther (1923). Die räumliche Energieverteilung der Hohlraumstrahlung. *Zeitschrift für Physik* 20: 145–152.
- (1924). Über die Wechselwirkung zwischen Strahlung und freien Elektronen. *Zeitschrift für Physik* 23: 214–224.
- (ca.1925). Zur Struktur der Strahlung. *AMPG, III/6/105.4*. Unpublished.
- (ca.1926). Lecture Notes. *AMPG, III/6/105.3*. Unpublished.
- (1926). Über die Kopplung zwischen elementaren Strahlungsvorgängen. *Zeitschrift für Physik* 37: 547–567.
- (1927a). Lichtquanten und Interferenz. *Zeitschrift für Physik* 41: 332–344.
- (1927b). Zur Statistik der Hohlraumstrahlung. *Zeitschrift für Physik* 41: 345–351.
- (1964). The Coincidence Method. In: *Nobel Lectures, Physics, 1942–1964*. Amsterdam: Elsevier. 271–276.
- Cassidy, David C. (2005). Einstein and the Quantum Hypothesis. *Annalen der Physik* 14: 15–22.
- Darrigol, Olivier (1988). Statistics and Combinatorics in Early Quantum Theory. *Historical Studies in the Physical and Biological Sciences* 19: 17–80.

- (1991). Statistics and Combinatorics in Early Quantum Theory, II: Early Symptoma of Indistinguishability and Holism. *Historical Studies in the Physical and Biological Sciences* 21: 237–298.
- (2009). A Simplified Genesis of Quantum Mechanics. *Studies in History and Philosophy of Modern Physics* 40: 151–166.
- de Broglie, Louis (1922a). Rayonnement noir et quanta de lumière. *Journal de Physique et le Radium* 3: 422–428.
- (1922b). Sur les interférences et la théorie de quanta de lumière. *Comptes Rendus de l'Académie des Sciences* 175: 811–813.
- Debye, Peter (1910). Der Wahrscheinlichkeitsbegriff in der Theorie der Strahlung. *Annalen der Physik* 33: 1427–1434.
- (1923). Zerstreung von Röntgenstrahlen und Quantentheorie. *Physikalische Zeitschrift* 24: 161–166.
- Dirac, Paul A. M. (1927). The Quantum Theory of the Emission and Absorption of Radiation. *Proceedings of the Royal Society A* 114: 243–265.
- Duncan, Anthony and Michel Janssen (2008). Pascual Jordan's Resolution of the Conundrum of the Wave-Particle Duality of Light. *Studies in History and Philosophy of Modern Physics* 39(3): 634–666.
- Ehrenfest, Paul (1925). Energieschwankungen im Strahlungsfeld oder Kristallgitter bei Superposition quantisierter Eigenschwingungen. *Zeitschrift für Physik* 34: 362–373.
- Einstein, Albert (1905). Über einen die Erzeugung und Verwendung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik* 17: 132–148.
- (1909). Über die Entwicklung unserer Anschauungen über das Wesen und die Konstitution der Strahlung. *Physikalische Zeitschrift* 10: 817–825.
- (1913). Solvay Conference 1911: Zum gegenwärtigen Stand des Problems der spezifischen Wärme. *Abhandlungen der Deutschen Bunsen-Gesellschaft* 7: 330–370.
- (1916). Zur Quantentheorie der Strahlung. *Mitteilungen der Physikalischen Gesellschaft Zürich* 18: 47–62.
- (1917). Zur Quantentheorie der Strahlung. *Physikalische Zeitschrift* 18: 121–128.
- (1924). Quantentheorie des einatomigen idealen Gases. *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*: 261–267.
- (1930). Statistische Eigenschaften der Strahlung. *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*: 543.
- Einstein, Albert and Paul Ehrenfest (1923). Zur Quantentheorie des Strahlungsgleichgewichtes. *Zeitschrift für Physik* 19: 301–306.
- Feynman, Richard P., Robert B. Leighton, and Matthew Sands (1965). *The Feynman Lectures on Physics*. III. Reading, MA: Addison-Wesely.

- Fick, Dieter (in print). Bothe's 1925 Heuristic Assumption in the Dawn of Quantum Field Theory. *The European Physical Journal H*.
- Fick, Dieter and Horst Kant (2009). Walther Bothe's Contributions to the Understanding of the Wave-Particle Duality of Light. *Studies in History and Philosophy of Modern Physics (SHPMP)* 40: 395–405.
- Frenkel, Viktor Ya. (1971). Yuriy Aleksandrovich Krutkov. *Soviet Physics Uspekhi* 13(6): 816–824.
- Fürth, Reinhold (1928a). Schwankungserscheinungen nach der neuen Quantenstatistik. *Zeitschrift für Physik* 48: 323–339.
- (1928b). Über Strahlungsschwankungen nach der Lichtquantenstatistik. *Zeitschrift für Physik* 50: 310–318.
- Hoffmann, Dieter (2010). Nicht nur ein Kopf, sondern auch ein Kerl! Zum Leben und Wirken Max von Laue (1879–1960). *Physik Journal* 9(5): 39–43.
- Ioffe, Abram F. (1911). Zur Theorie der Strahlungserscheinungen. *Annalen der Physik* 36: 534–552.
- (1983). *Vstrechi s fizikami* (Meeting physicists). Leningrad: Izd. Nauka.
- (1985). *O fizike i fizikach* (On physics and physicists). Leningrad: Izd. Nauka.
- Ishiwara, Jun (1912). Zur Theorie der Strahlungserscheinungen. *Physikalische Zeitschrift* 13: 1142–1151.
- Jordan, Pascual (1928). Das photochemische Gesetz und die molekulare Theorie der Strahlung. *Physikalische Zeitschrift* 7: 158–208.
- (1929). Der gegenwärtige Stand der Quantenelektrodynamik. *Physikalische Zeitschrift* 30: 700–712.
- Kar, Kulesh Ch. (1927). Die Quantenstatistik. *Physikalische Zeitschrift* 28: 300–302.
- Kiejna, Adam (2002). Stanislaw Loria and Mieczysław Wolfke in Wrocław — Bridging the German Past and the Polish Present. In: *Proceedings of the 17th Max Born Symposium, Wrocław, Poland, 2002*. Ed. Jerzy Lukierski and Helmut Rechenberg. Wrocław: Wydawn. Uniwersytetu Wrocławskiego. 49–69.
- Kirchner, Fritz (1930). Allgemeine Physik der Röntgenstrahlung. In: *Handbuch der Experimentalphysik*. Ed. Wilhelm Wien und Freidrich Harms. 24.1. Leipzig: Akademische Verlagsgesellschaft.
- Klein, Martin J., Anne J. Kox, and Robert Schulmann (1993). *The Collected Papers of Albert Einstein*. V. Princeton, N.J.: Princeton University Press.
- Kuhn, Thomas S. (1987). *Black-Body Theory and the Quantum Discontinuity 1894–1912*. Chicago: The University of Chicago Press.
- Landé, Alfred (1925). Lichtquanten und Kohärenz. *Zeitschrift für Physik* 33: 571–578.

- Lanyi, Gabor E. (2003). Thermal Equilibrium Between Radiation and Matter. *Foundations of Physics* 33: 511–527.
- Laue, Max (1906). Zur Thermodynamik der Interferenzerscheinungen. *Annalen der Physik* 20: 365–378.
- (1907a). Die Entropie von partiell kohärenten Strahlenbündeln. *Annalen der Physik* 23: 1–43.
- (1907b). Die Entropie von partiell kohärenten Strahlenbündeln: Nachtrag. *Annalen der Physik* 23: 795–797.
- von Laue, Max (1914). Die Freiheitsgrade von Strahlenbündeln. *Annalen der Physik* 44: 1197–1212.
- Lewis, Gilbert N. (1926). The Conservation of Photons. *Nature* 118: 874–875.
- Lewis, Henry R. (1973). Einstein's Derivation of Planck's Radiation Law. *American Journal of Physics* 41: 38–44.
- Lorentz, Hendrik A. (1916). *Les théories statistiques en thermodynamique*. Leipzig: B. G. Teubner.
- Majumdar, R. J. and D. S. Kothari (1930). Statistische Deutung zur Einsteinschen Ableitung des Planckschen Gesetzes. *Zeitschrift für Physik* 60(3–4): 250–254.
- Mandel, Leonard and Emil Wolf (1965). Coherence Properties of Optical Fields. *Review of Modern Physics* 37: 231–287.
- (1995). *Optical Coherence and Quantum Optics*. Cambridge: Cambridge University Press.
- Navarro, Luis and Enric Perez (2004). Paul Ehrenfest on the Necessity of Quanta (1911): Discontinuity, Quantization, Corpuscularity, and Adiabatic Invariance. *Archive for History of Exact Sciences* 58: 97–141.
- Pauli, Wolfgang (1923). Über das thermische Gleichgewicht zwischen Strahlung und freien Elektronen. *Zeitschrift für Physik* 18: 272–286.
- (1926). Quantentheorie. *Handbuch der Physik* 23: 9–23.
- Planck, Max (1913). *Vorlesungen über die Theorie der Wärmestrahlung*. 2nd ed. Leipzig: Barth.
- (1925). Zur Frage der Quantelung einatomiger Gase. *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*: 49–57.
- Rubens, Heinrich (1913). Solvay Conference 1911: Die Prüfung der Planckschen Strahlungsformel im Gebiete langer Wellen. *Abhandlungen der Deutschen Bunsen-Gesellschaft* 7: 72–76.
- Sredniawa, Bronislaw (2006). Scientific and Personal Contacts of Polish Physicists with Einstein. *Concepts of Physics* 3: 385–427.
- Stark, Johannes (1910). Zur experimentellen Entscheidung zwischen Ätherwellen- und Lichtquantenhypothese, I. Röntgenstrahlung. *Physikalische Zeitschrift* 10: 902–913.

- Warburg, Emil (1913). Solvay Conference 1911: Die experimentelle Prüfung der Planckschen Formel für Hohlraumstrahlung. *Abhandlungen der Deutschen Bunsen-Gesellschaft* 7: 65–71.
- Wheaton, Bruce R. (1983). *The Tiger and the Shark. Empirical Roots of Wave-Particle Dualism*. Cambridge, MA: Cambridge University Press.
- Wolfke, Karol (1980). Wspomnienia o ojcu, Dr. Mieczysław Wolfke (Recollections on my father, Dr. Mieczysław Wolfke). *Postepy Fizyki* 31: 551–558.
- Wolfke, Mieczysław (1911). Über die Abbildung eines Gitters bei künstlicher Begrenzung. *Annalen der Physik* 34: 277–310.
- (1912). Allgemeine Abbildungstheorie selbstleuchtender und nichtselbstleuchtender Objekte. *Annalen der Physik* 39: 569–610.
- (1913a). Zur Quantentheorie. *Verhandlungen der Deutschen Physikalischen Gesellschaft* 15: 1123–1129.
- (1913b). Zur Quantentheorie. *Verhandlungen der Deutschen Physikalischen Gesellschaft* 15: 1215–1218.
- (1914a). Zur Quantentheorie. *Verhandlungen der Deutschen Physikalischen Gesellschaft* 16: 4.
- (1914b). Welche Strahlungsformel folgt aus der Annahme der Lichtatome? *Physikalische Zeitschrift* 15: 308–379.
- (1921). Einsteinsche Lichtquanten und räumliche Struktur der Strahlung. *Physikalische Zeitschrift* 22: 375–379.
- (1930). Bemerkung zu “Statistische Deutung zur Einsteinschen Ableitung des Planckschen Gesetzes” von R.J. Majumdar und D.S. Kothari in Allahabad. *Zeitschrift für Physik* 61: 725–726.
- (1946a). Über Multiphotone. *APAN III-71*, 23: 5–12. Unpublished manuscript.
- (1946b). Über die Mehrfachquanten in der Planck’schen Strahlung. *Helvetica Physica Acta* 19: 427–429.
- (1946c). Sur la possibilité d’un effet photoélectrique des multiphotons. *Comptes Rendus* 233: 655–657.

Chapter 5

Early Interactions of Quantum Statistics and Quantum Mechanics

Daniela Monaldi

Two papers that inaugurated the quantum mechanics of multiparticle systems were published in the second half of 1926. They were “Mehrkörperproblem und Resonanz in der Quantenmechanik” by Werner Heisenberg and “On the Theory of Quantum Mechanics” by Paul Dirac (Heisenberg 1926; Dirac 1926).¹ These works are commonly credited together for having laid the foundations of the integration of quantum mechanics and quantum statistics because they introduced the quantum-mechanical expression of the symmetry of a system under exchanges of equal particles. The quantum formalism of exchange symmetry is regarded as having solved at once long-standing difficulties regarding the statistical properties of both equal particles and light quanta by clarifying and legitimizing the previously foggy notion of indistinguishable particles. Despite apparent formal similarities, however, there were significant differences between Heisenberg’s and Dirac’s approaches to multiparticle systems. Furthermore, under the surface of Heisenberg’s distrust of visualizable models and of Dirac’s ideal of abstract theorizing, the two works relied on an interpretive model of particle systems that differed from both earlier and later interpretations of quantum statistics, while remaining surprisingly close to the corpuscular model of the older statistics of James C. Maxwell and Ludwig Boltzmann. Dissonances of this kind are to be expected from two works produced in the early days of quantum mechanics, when the theory was still under construction and questions of interpretation were beginning to surface. One may recall that Max Born’s work on collisions, which opened the way to the statistical interpretation of wave functions, also dates from the summer of 1926. Wolfgang Pauli and Heisenberg formulated the principle of indeterminacy in the following fall and winter. And only in September 1927 did Niels Bohr set forth his principle of complementarity, supposedly providing conceptual unity to the so-called Copenhagen interpretation of quantum mechanics

¹Heisenberg’s paper was submitted and published in June 1926; Dirac’s paper was submitted on 26 August and published in October 1926.

and solving the puzzle of wave-particle duality, under which the interpretation of quantum statistics was filed (Jammer 1966).

While issues of interpretation in quantum mechanics have attracted much historical and philosophical scholarship, the forging of its alliance with quantum statistics remains underexamined. Recent historical analyses have uncovered a plurality of voices under the unisonant narrative of the Copenhagen interpretation, while physical and philosophical work has once again brought to the forefront questions concerning the ontology of quantum mechanics and quantum field theory that were never closed (Beller 1999; Bitbol 2007; Camilleri 2009). Therefore, an investigation of the statistical connections of the early quantum mechanics of multiparticle systems may contribute some clarification. In what follows, I set the stage by recalling the birth of quantum statistics and its first interpretation; then, I analyze Heisenberg's and Dirac's pioneering approaches to the integration of quantum statistics and quantum mechanics, with the aim of uncovering the presuppositions of their authors and their interpretations of the outcomes. The purpose is to shed light on the early phase in the historical process of understanding the statistical behavior of multiparticle systems and its connection with the wave-particle duality.

5.1 The Birth of Quantum Statistics

When Heisenberg and Dirac took up the many-particle problem in the emerging framework of quantum mechanics, quantum statistics was itself less than two years old. It had been born in the second half of 1924, when Albert Einstein applied a statistical method that Satyendra Nath Bose had just worked out for radiation to the ideal gas. This new theory, which we know as Bose-Einstein statistics, was based on the combinatorial calculation of the entropy of an assembly of “elementary entities”—light quanta or gas molecules—in a quantized phase space. As Einstein explained in the first of his quantum gas papers, the statistical-combinatorial method that Bose had adopted from Max Planck consisted of dividing the phase space of an elementary entity into “cells” of size h^3 , and defining the “microscopic state” of the assembly as the distribution of the elementary entities over the cells. Every “macroscopic state” was then assigned a quantity, W , equal to the number of different microscopic states by which the macroscopic state could be thought to be realized. The quantity W was introduced by Planck in his adaptation of Boltzmann's calculation of the equilibrium distribution of the ideal gas to thermal radiation. It was used to calculate the entropy, S , and the other thermodynamic functions through the Boltzmann principle, $S = k \log W$. Einstein called it insistently “probability (in Planck's sense)” or “probability *à la* Planck,” evidently to stress that it was no probability in his

sense (Einstein 1924, 261–262; 1925). There was nonetheless an essential difference between Bose’s calculation and the “quantum statistics” that this method had hitherto produced in the hands of Planck and others. Einstein compared the two cases in detail in the second of his quantum gas papers, labelling the former method “according to Bose” and the latter “according to the hypothesis of statistical independence of the molecules.” In both cases, the entropy of a state of the system was proportional to the logarithm of W , which Einstein characterized as the number of “possibilities of realization” of the state. Likewise, in both cases, a “macroscopically defined state,” or energy distribution, was defined by a set of numbers (n_1, n_2, n_3, \dots) representing the numbers of elementary entities in each “infinitesimal region” of energy, or “elementary region,” ΔE . Finally, within each elementary region, the molecules were to be regarded as distributed among the cells of size h^3 in the quantized phase space of a single molecule, and the ν -th elementary region contained $z_\nu = 2\pi \frac{V}{h^3} (2m)^{\frac{3}{2}} E^{\frac{1}{2}} \Delta E$ cells (Einstein 1924, 262).² Where the two methods differed was in their assumptions of what the “possibilities of realization” of a macroscopic state were. Though not explicitly defined, these represented a generalization of the configurations of molecules that Boltzmann had named “complexions,” and had to be the most specific states of equal probability by which any other state, as for example an energy distribution, could be thought to be realized. Indeed, Einstein proceeded to quietly drop the term “possibilities of realization” and to replace it with “microscopically defined states,” while also indicating the term “complexions” in parenthesis. In the old statistics, the microscopic states were defined by stating “in which cell each molecule sits,” while in the new statistics they were defined by stating “how many molecules are in each cell” (Einstein 1925, 5–6). This meant that in the old statistics, one first counted in many different ways how the n_ν molecules in the ν -th elementary region could be distributed among the z_ν phase-space cells of that region. The result was $z_\nu^{n_\nu}$. Then, one calculated the number of different ways in which the distribution (n_1, n_2, n_3, \dots) could be obtained from the n molecules, obtaining the factor $\frac{n!}{\prod n_\nu!}$. Finally, one had to multiply this factor by the product of the $z_\nu^{n_\nu}$ over all the regions, $\prod_\nu z_\nu^{n_\nu}$. In the statistics of Bose, the number of ways to

partition n_ν particles into z_ν cells was given instead by the factor, $\frac{(n_\nu + z_\nu - 1)!}{n_\nu! (z_\nu - 1)!}$, and the number of complexions corresponding to the energy distribution was given by the product of these factors over all the elementary regions.

The calculation described above was the original form of Einstein’s new statistics. For the purpose of discussing the integration of quantum statistics

²Einstein assumed that “[f]or any given $\frac{\Delta E}{E}$, however small,” one could always choose V so large that z_ν would be “a large number.”

with quantum mechanics, however, nothing will be lost if we consider a simpler form, first introduced by Erwin Schrödinger in his examination of Einstein's quantum gas theory; in this form, the elementary regions of energy are chosen to coincide with the quantum cells ($z_v = 1$) (Schrödinger 1925). For the old statistics, then, the number of complexions for a given energy distribution becomes $n! / (n_1! n_2! \dots)$, the factor that Boltzmann used, called "permutability" of the distribution, and from which he derived the Maxwell-Boltzmann distribution. For the new statistics, the number of complexions becomes equal to one. In the language of quantum atomic theory, which Heisenberg and Dirac used in their treatments of the many-body problem, the change from the Maxwell-Boltzmann statistics to that of Bose and Einstein consisted of the reduction of the statistical weight (Heisenberg's term), or a priori probability (Dirac's term), of a state from Boltzmann's permutability (different for each energy distribution) to one (equal for all the energy distributions).

Einstein realized that Bose's way of counting complexions violated an assumption that had, until then, been basic to the statistical method, namely, the statistical independence of the elementary entities. For this reason, he offered Bose's method as a new statistics alternative to the statistics of independent particles. He made this point very clearly, first in private correspondence and then in the second quantum gas paper (Einstein 1925).³ In his understanding, the new statistics expressed a mutual influence among the elementary entities, which was "for the time being of an entirely mysterious kind" (Einstein 1925, 7). The lack of statistical independence of the light quanta had already been analyzed, especially by Paul Ehrenfest, and Einstein theorized that it was responsible for the wave properties of radiation. For the light quanta, then, the new statistics simply expressed in new form something already known. Einstein surmised that the hypothetical application of the same statistics to the molecules of an ideal gas was justified by a "far reaching formal similarity between radiation and gas," which he believed to be "more than a mere analogy" (Einstein 1925, 3). He thus suggested that Louis de Broglie's theory, which attributed wave properties to material corpuscles, might provide the appropriate theoretical framing of the similarity.

Every physicist who commented in print on the new statistics in the two years after its publication adopted Einstein's interpretation of it as a statistics of non-independence. Schrödinger and Pascual Jordan, however, followed also Einstein's suggestion of a fundamental similarity between radiation and matter. Schrödinger adopted de Broglie's ideas and developed them into wave mechanics; Jordan started a treatment of the radiation field that formalized the wave-particle duality within the scheme of the new quantum mechanics (Darrigol 1986; 1992a).

³Einstein to Halpern, September 1924, Einstein Archives, Reel 12, Doc. 128; Einstein to Schrödinger, 2 February 1925, Einstein Archives, Reel 22, Doc. 2.

5.2 Heisenberg's Many-body Problem and Quantum Resonance

The first to find a connection between the quantum mechanics of particles and the new statistics was Heisenberg. Upon accepting the position of lecturer and assistant to Niels Bohr at the Institute of Theoretical Physics in Copenhagen, in April 1926, Heisenberg went to work on the problem of the helium atom. The hypothesis of electron spin that George Uhlenbeck and Samuel Goudsmit put forward in the fall of 1925 had raised hopes to find a common solution for the helium problem and for an explanation of the Pauli *Verbot*, or “Pauli exclusion rule,” the prohibition of equivalent orbits for atomic electrons postulated by Pauli at the end of 1924. Heisenberg wrote to Pauli at the beginning of May:⁴

We have found a rather decisive argument that your exclusion of equivalent orbits is connected with the singlet-triplet separation [...]. Consider the energy written as a function of the transition probabilities. Then a large difference results if one—at the energy of H atoms—has transitions to 1S, or if, according to your ban, one puts them equal to zero. That is, para- and ortho-[helium] do have different energies, independently of the energies between magnets [i.e., magnetic moments associated with spin].

Heisenberg's “decisive argument” appeared in *Zeitschrift für Physik* at the end of June. It was his first published response to wave mechanics, which Schrödinger had just set forth, presenting it as formally equivalent but physically preferable to matrix mechanics. Heisenberg had reacted positively to Schrödinger's theory, welcoming the formal connection of the two theoretical schemes, and hoping that it might be of help in reaching a physical understanding of quantum mechanics. As he wrote to Dirac:⁵

I see the real progress made by Schrödinger's theory in this: that the same mathematical equations can be interpreted as point mechanics in a non-classical kinematics and as wave theory according to Schröd[inger]. I always hope that the solution of the paradoxes in the quantum theory later could be found on this way [*sic*].

In fact, Heisenberg saw an overthrow of the classical representation of motion in space and time in quantum mechanics, hence a new kinematics, but had been concerned from the onset about the question of how the new kinematics could be understood (Camilleri 2009). Soon, however, he developed a hostility to the physical interpretation that Schrödinger proposed, namely, an open return to a physics

⁴Heisenberg to Pauli, 5 May 1926, quoted in (Mehra and Rechenberg 1987, 737).

⁵Heisenberg to Dirac, 26 May 1926, AHQP 59–2.

of continuous processes in classical space and time. Shortly before submitting his article, Heisenberg wrote to Pauli: “The great achievement of Schrödinger’s theory is the calculation of the matrix elements;” then he added: “The more I ponder the physical part of Schrödinger’s theory, the more horrible I find it.”⁶ He expressed the same view, in mitigated terms, in the opening of the paper. While acknowledging the convenience of the wave-mechanical approach and its formal connections with the matrix approach, he stressed the difficulties of a wave theory of matter. He closed the introduction stating that, despite the rising of the matter-radiation analogy, “one of the most important aspects of quantum mechanics” was that it was “based upon a corpuscular conception of matter,” even though it was not a description of corpuscles moving in ordinary space and time (Heisenberg 1926, 412). These words echoed the way in which Heisenberg, Born and Jordan had presented matrix mechanics in November 1925. They had described it as “a system of quantum-theoretical relations between observable quantities” which could not be directly interpreted in “a geometrically visualizable way,” because in it the motion of electrons could not “be described in terms of the familiar concepts of space and time” (Born, Heisenberg, and Jordan 1925, 558).

Heisenberg’s strategy for the helium atom was to extend quantum mechanics to a system composed of two electrons coupled by a potential—the simplest example of a many-body system that could model the helium atom—on the basis of an analogy with the classical effect of resonance between two coupled oscillators. He started by noting that in absence of interaction, the energy of the system was simply the sum of the energies of the two electrons and did not change under the exchange of the two electrons. He represented the total energy as $H_{nm} = H_n^a + H_m^b$, where H_n^a and H_m^b represented the energies of the two electrons, with a and b labelling the electrons, and n and m their single-electron states. Correspondingly, he also had distinct states labelled nm and mn in the matrix representing the interaction energy of the system. Being the electrons perfectly equal particles with identical energy states, the energies of the two states, rather obviously, coincided, $H_{nm} = H_{mn}$. Heisenberg expressed this property by characterizing the energy terms as “twofold” and then segued into the assertion that the states of the system exhibited what he called “the degeneracy characteristic of the resonance.” He made the same point graphically, representing the two states separately in a diagram showing the energy levels. He also admitted, both verbally and graphically, an exception to the twofoldness-degeneracy-resonance, namely, the states in which both electrons had the same energy (fig. 5.1). Heisenberg did not use too many words to justify his choice to give representation to both states, nm and mn (except for $m = n$) and his identification of twofoldness, degeneracy, and resonance; yet, the choice was neither obvious nor casual,

⁶Heisenberg to Pauli, 8 June 1926, quoted in (Mehra and Rechenberg 1987, 740).

and the identification all but trivial. As we shall see in the next section, Dirac did pause to ponder how such states should be represented in the matrix form, and made a different choice—ironically, invoking Heisenberg’s methodological principle that the new quantum mechanics should only allow the calculation of observable quantities (Dirac 1926, 666–667). Heisenberg’s priority in dealing with the helium problem was, instead, the expediency of resonance:

In other words, resonance always occurs when the two systems are not originally in the same state, for the exchange of the two systems gives the same energy. Only in the case of equal energy of the two particles the resonance (or the degeneracy) disappears. (Heisenberg 1926, 417)

In fact, if an interaction was present and acted as a perturbation on the stationary states of the non-interacting system, there resulted two corrections to the total energy. Representing with H^1 the interaction energy, in first approximation the corrections were $W_{nm}^1 = H^1(nm, nm) + H^1(nm, mn)$ and $W_{mn}^1 = H^1(nm, nm) - H^1(nm, mn)$. Each of the twofold energy terms was therefore split into two new values, while no splitting occurred for the terms in which $m = n$. It appears, therefore, that Heisenberg closely followed the analogy with the classical phenomenon of resonance, which was instrumental to the explication of the helium spectrum. At the same time, he deployed the term “degeneracy” (“*Entartung*”) from Bohr’s atomic theory, a term that carried a specific assignation of statistical weights, or a priori probabilities. Bohr’s atomic theory incorporated the rule that every stationary state of an atomic system had the same a priori probability, with the exception of the states that were “degenerate,” that is, they could be adiabatically transformed into two or more states of different quantum energies. In such cases, Bohr stated that:

[T]he probability of a given state must be determined from the number of stationary states of some non-degenerate system which will coincide in the given state, if the latter system is continuously transformed into the degenerate system under consideration. (Bohr 1918, 26)

Bohr explicitly crafted his definition of degeneracy to correspond to the classical statistical assumption that a priori probabilities were proportional to volumes in phase space. Along the same lines, Heisenberg took the unperturbed states of the two-electron system to be degenerate or non-degenerate according to whether they were transformed into two new states or into a single one by the interaction. This meant that the assignation of statistical weights was the same as that of any

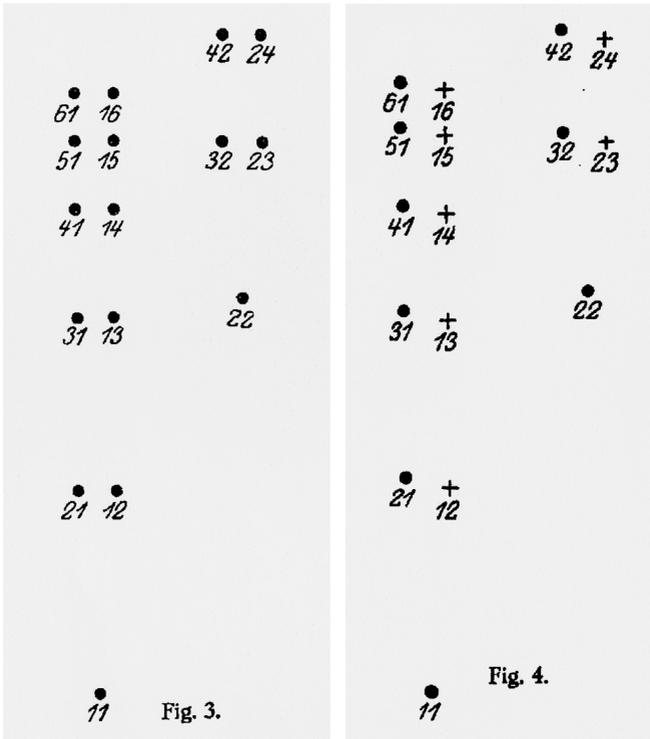
other application of statistical mechanics to quantum systems prior to Bose. Put differently, Heisenberg's formalism for non-interacting particles channeled exactly the same corpuscular model as the Maxwell-Boltzmann statistics. As we are about to see, Heisenberg was eager to connect his multiparticle quantum mechanics to the new statistics of Bose and Einstein. Nevertheless, his definition of degeneracy for the non-perturbed states reveals that he was not seeking the connection because he had come to a reexamination of the nature of particles in the light of the new kinematics.

In accord with his application of Bohr's rule, Heisenberg further asserted that the degeneracy was "eliminated in the system perturbed by the interaction" (Heisenberg 1926, 417). He also warned, however, that although he continued to label the new states nm and mn , the indices no longer referred to the states of the individual particles. He further warned that it no longer made "physical sense to talk" about the motion of single electrons (Heisenberg 1926, 418, 423).

Thus, he arrived at what he regarded as the decisive result. Because of the equality of the two electrons, the matrix representing the radiation emitted by the system had to be symmetric under electron exchanges. This requirement entailed that the new energy levels divided into two sets (which Heisenberg marked with the symbols $+$ and \bullet , see fig. 5.1), such that the amplitudes of all the transitions from the energy levels of one set to those of the other set were zero. No transition would occur from one set to the other; as Heisenberg put it, the two sets could "in no way combine with each other" (Heisenberg 1926, 418). Resonance produced, therefore, an indeterminacy in the possible solutions of the problem. Either one of the two sets of terms, as well as a mixture of the two, constituted a complete solution, and the theory alone could not decide which choice was correct.

At this point, Heisenberg examined the problem through Schrödinger's formalism. Applying the relations between matrix elements and eigenfunctions, he found that the eigenfunctions of the perturbed system were the two linear combinations of single-electron eigenfunctions. Indicating with ϕ_n^a and ϕ_m^b the eigenfunctions of the two electrons, a and b , in the unperturbed single-electron energy states of energies H_n^a and H_m^b , the eigenfunctions of the whole system in the perturbed case were $\frac{1}{\sqrt{2}}(\phi_n^a\phi_m^b + \phi_m^a\phi_n^b)$ and $\frac{1}{\sqrt{2}}(\phi_n^a\phi_m^b - \phi_m^a\phi_n^b)$ (Heisenberg 1926, 420).

Heisenberg was then able to show that the predicted phenomenon of quantum resonance could reproduce, qualitatively and in order of magnitude, the spectrum of helium. The Coulomb repulsion between the two electrons of the helium atom would cause a "large electrical resonance," and the corresponding energy separation between the two sets reproduced in order of magnitude the differences between the spectra of ortho-helium and para-helium. If then the electrons were regarded as "small magnetic spinning tops," the large resonance would be per-



(a)

(b)

Figure 5.1: Heisenberg's diagrams illustrating the states of a system of two electrons. (a) Without interaction, every energy value is twofold or degenerate, except when the electrons are in the same single-electron state. (b) With interaction, the degeneracy is broken, and the states divide into two non-combining sets, indicated by + and •. Source: (Heisenberg 1926, 417–418).

turbed. On the one hand, weak transitions between the two sets would now be permitted thanks to “the interaction between magnet and orbit.” However, each stationary state would become fourfold (there would be four possible combinations of the two spins), and a “finer exact resonance” would occur on account of the interaction between spins, which would again produce a separation into two non-combining sets (Heisenberg 1926, 421–422). Heisenberg then observed that of the two theoretically possible sets of terms, only one agreed with experiment, namely, the one in which there were no equivalent orbits and in which, therefore, the Pauli *Verbot* was satisfied. Although he could not find a reason why only one of the two sets should be selected, he was at least able to offer a formal representation of the empirical rule.

In the last section of the paper Heisenberg made the connection with Bose-Einstein statistics. In a swoop of guesswork, he argued that if his results were generalized to an arbitrary number of electrons, it then became possible to use the indeterminacy of the stationary states and the Pauli rule to deduce the statistics of the assembly. According to him, the selection of only one of the theoretically possible solutions according to the now-formalized Pauli rule would result in a “reduction of the statistical weights” of the states that corresponded precisely to “the Bose-Einstein counting” (Heisenberg 1926, 422). He claimed that this formulation of the new assignation of statistical weights surpassed Einstein’s, for it not only prescribed the choice of one specific solution out of many, it also specified the right choice, because it demanded the one set that satisfied the Pauli *Verbot*.

Heisenberg counted this outcome as a success. Even though he could not justify the natural selection of a single solution, having found a formal scheme that simultaneously reproduced the Pauli rule and the Bose-Einstein statistics was nonetheless an important result, since it indicated that they had the same cause. He recalled that an interaction was necessary among the particles for the quantum-mechanical resonance to occur, but he did not care to clarify what this meant for the molecules of the Bose-Einstein ideal gas. He also emphasized that the theoretical success came at a cost: it no longer made “physical sense to talk of” the motion of single electrons (Heisenberg 1926, 423). This startling restriction went beyond the notion that the theory described individual corpuscles in non-ordinary space and time. Heisenberg, however, defused its ontological potential by asserting that it made no sense to speak of the motion of individual electrons just as it made no sense to speak of any non-symmetrical function of the electrons, because only symmetric functions represented observable quantities. Finally, he returned to the wave interpretation to conclude that it could now be dismissed. Not only did the corpuscular interpretation alone suffice to deal with multiparticle

systems, it even afforded a smooth integration of quantum mechanics with the statistical theory that had seemed to support the undulatory conception of matter.

Despite his assurance, in the statistical part of the argument, Heisenberg was flying by the seat of his pants. Most blatantly, he did not recognize that the Bose-Einstein statistics, in which there was no restriction on the number of particles allowed in a state, was incompatible with the Pauli *Verbot*, which allowed no more than one particle in each state. A few months earlier, Enrico Fermi had published a new quantum statistics of the ideal gas that was explicitly based on the generalization of the Pauli rule (Fermi 1926, reprinted in: Amaldi et al. 1962, 181–185). Fermi, a young theorist in Rome who was not yet a member of the quantum network, had worked within the framework of the old quantum mechanics. Dirac would shortly arrive at the same quantum statistics in his treatment of many-particle systems in quantum mechanics. Heisenberg, who was evidently unaware of Fermi's work, confused the two quantum statistics, and strangely continued to confuse them even after the publication of Dirac's paper. This oversight was not the only peculiarity of Heisenberg's argument, however. In his reckless generalization from two to n electrons, Heisenberg took the Bose-Einstein statistics to amount to a "reduction of statistical weights from $n!$ to 1" (Heisenberg 1926, 423). But the statistical weight assigned by the old statistics to a macroscopic state was not $n!$. It was Boltzmann's permutability, $n! / n_1! n_2! \dots$, as we have seen in the previous section. Next, Heisenberg assumed that every energy value of the unperturbed n -particle system was $n!$ -fold degenerate, even though he had denied earlier that this was always the case. The states (1, 1) and (2, 2), for instance, were non-degenerate in his scheme (fig. 5.1). Therefore, according to his reasoning, the degeneracy factor should again have been the permutability, not the number of particle permutations, $n!$. That replacement of the permutability with the factorial of n was a standard approximation in statistical calculations, in which it was often possible to assume that the frequency of cases of particles of equal energies was negligible, may partly account for Heisenberg's slip; still, the replacement was indefensible in the case of an atomic system. Although these twin inaccuracies neutralized one another as far as Heisenberg's aim was concerned, they had the unfortunate effect of masking the actual difference between the old and the new statistics. A simple reduction of statistical weights by a common factor, such as Heisenberg implied, would have left the statistical distribution of macrostates unchanged. Furthermore, Heisenberg went on to claim that quantum resonance caused, quite conveniently, a separation into precisely $n!$ non-combining sets of states. He was unable to prove the claim, and in less than six months, Eugene Wigner dismantled it. Through a careful examination of a system of three interacting particles, Wigner showed that the separation into non-combining sets did occur, but the number of non-combining sets was only three.

To handle the problem for higher values of n , one needed the mathematical theory of groups, and the solution did not turn out to be $n!$ (Wigner 1927a; 1927b). Finally, Heisenberg oddly insisted that the reduction of statistical weights was caused by “the choice of one quantum mechanical solution out of many possible solutions,” even though according to Bohr’s rule the reduction was effected simply by the removal of degeneracy (Heisenberg 1926, 425).

Why did Heisenberg get himself into this jumble? To recover the Bose-Einstein statistics, it would have been sufficient to generalize Bohr’s rule by assuming that all the possible perturbed multiparticle states had the same statistical weight. Dirac was to take this route, as we shall see. The reason for the unnecessary complication of Heisenberg’s argument may reside in the incompleteness of his interpretation of the formalism, and in the ambiguity in which it left the ontological status of multiparticle states. Heisenberg was careful not to pronounce himself too finely on ontology beyond his advocacy of the corpuscular interpretation and his rejection of classical kinematics. Nevertheless, his apparent fondness of the number $n!$ suggests that his heuristic resource for theory construction was still the classical conception of an assembly of free particles. He regarded every energy value of the free-particle system as degenerate because it was potentially obtainable in multiple ways through exchanges of equal particles. For two particles, this classical notion accorded with Bohr’s definition of degeneracy, and Heisenberg rashly assumed that the accord would persist for any number of particles. Since he considered every multiparticle state as degenerate, he assigned it a statistical weight equal to $n!$, that is, the number of possible particle permutations, easily forgetting the anomaly presented by particles of equal energy. But what happened to the supposed degeneracy when an interaction deprived the particles of their freedom? The self-imposed prohibition to talk of individual corpuscles moving in ordinary space and time deprived Heisenberg of the possibility to interrogate the classical corpuscular model, while it did not itself provide an answer. In these circumstances, the formalism was mute. It is probably in order to resolve the ambiguity that Heisenberg reached for a mechanism capable of suppressing any potential degeneracy, and the Pauli-rule selection, albeit unjustified, seemed to him suitable to this purpose.

That Heisenberg’s treatment of the many-body problem was informed by the classical model of particle assemblies is also supported by the few explicit remarks that he made concerning the interpretation of the formalism. From the viewpoint of the corpuscular interpretation, the perturbed energies contained terms corresponding to “transitions in which the electrons exchange[d] place” (Heisenberg 1926, 417). Therefore, Heisenberg explained that in each perturbed state the electrons had “the same motion (in different phases)” (Heisenberg 1926, 418). He regarded this effect as the analog of classical resonance between

coupled oscillators. While he declared that in presence of an interaction it no longer made “physical sense to talk of” the motion of single electrons, he also advised that if you wanted to form a picture of the motions, you could imagine the electrons to “exchange place periodically in a continuous way” (Heisenberg 1926, 421 and 423).⁷

5.3 Dirac’s “On the Theory of Quantum Mechanics”

The second treatment of many-particles systems in quantum mechanics was given shortly thereafter by Dirac in a paper titled “On the Theory of Quantum Mechanics” (Dirac 1926). As for Heisenberg, this was also Dirac’s first published response to Schrödinger’s theory. He had corresponded with Heisenberg while completing his PhD thesis in Cambridge in the spring of 1926. Many years later, he wrote in his recollections that he did the work on many-particle systems after Heisenberg convinced him of the usefulness of wave mechanics. Dirac felt “at first a bit hostile” to this theory because it seemed to him that it represented a regress to “the pre-Heisenberg stage.” In a non-extant letter to Heisenberg, he criticized Schrödinger because “the wave theory of matter must be inconsistent just like the wave theory of light” (Dirac 1977, 131–132). Heisenberg agreed with this criticism but nonetheless saw Schrödinger’s theory as progress, as we have seen. Thanks to Heisenberg’s detailed explanation of the relation between the two formal schemes, Dirac could see that wave mechanics “would not require us to unlearn anything that we had learned from matrix mechanics” but rather “supplemented the matrix mechanics and provided very powerful mathematical developments which fitted perfectly with the ideas of matrix mechanics” (Dirac 1977, 133).

In Dirac’s retrospective account, it was the study of Schrödinger’s formalism that suggested to him the possibility of symmetric and antisymmetric wave functions for a system of similar particles. These “symmetry questions,” in turn, “brought in the possibility of new laws of Nature” (Dirac 1977, 133). But the inspiration to explore the symmetry of the wave functions might not have been as purely formalistic as it appeared to Dirac in hindsight. In fact, Dirac knew that Heisenberg was working on the helium atom, because Heisenberg had written to him of his idea that the explanation of the helium spectrum was “a resonance effect of a typical quantum mechanical nature.”⁸ Moreover, in an “added in proof” footnote in his paper, Dirac wrote: “Prof. Born has informed me that Heisenberg has independently obtained results equivalent to these” (Dirac 1926, 670). Dirac

⁷See (Carson 1996) for an analysis of Heisenberg’s conception of energy exchange and its offspring, the “peculiar notion of exchange forces.”

⁸Heisenberg to Dirac, 26 May 1926, AHQP 59–2.

probably met Born in late July, when the latter was in Cambridge to give a talk at the Kapitza Club (Kragh 1990, 321, note 68; Darrigol 1992b). Therefore, he knew about Born's thoughts on the superposition of wave functions, and it is also likely that he learned about Heisenberg's results before submitting his paper in late August. Nevertheless, Dirac proceeded differently from Heisenberg and he also reached significantly different results.

Instead of confronting Schrödinger's undulatory interpretation, Dirac set out to reformulate Schrödinger's formal apparatus in general terms according to his own mathematical approach. He deduced the expression of the general solution of a quantum-mechanical problem as a linear expansion with arbitrary constants in "a set of independent solutions," which he called eigenfunctions (Dirac 1926, 664). This formal milestone enabled him to develop a quantum-mechanical treatment of multiparticle systems and to reach three lasting results. He arrived at the symmetry and antisymmetry of the wave functions, formulated the statistics that we now know as Fermi-Dirac statistics, and derived a calculation of Einstein's coefficients of absorption and stimulated emission. While Heisenberg welcomed the wave theory because it showed that the mathematical apparatus could be interpreted in two ways, Dirac adopted the wave formalism as an enhanced mathematical apparatus from which it was possible to calculate the matrices of the older formulation. He believed that questions of interpretation should be broached only after a general formal scheme was developed as abstractly as possible (Darrigol 1992b). Nonetheless, since he did need an interpretive model to extend the formalism to a new kind of physical system, he openly relied on the interpretation of the matrices in terms of particles and quantum transitions, simply sidestepping the wave aspects of the theory.

As Heisenberg, Dirac adopted "an atom with two electrons" as the simplest multiparticle system. In his atom, however, all interactions between electrons could be neglected. He did not resort to the analogy with the classical phenomenon of resonance as a theoretical tool, but used only the symmetry of the two-electron system supplemented by the methodological principle for which he credited Heisenberg:

[Heisenberg's matrix mechanics] enables one to calculate just those quantities that are of physical importance, and gives no information about quantities such as orbital frequencies that one can never hope to measure experimentally. We should expect this very satisfactory characteristic to persist in all future developments of the theory. (Dirac 1926, 667)

Dirac indicated with (mn) "the state of the atom in which one electron is in an orbit labelled m and the other in the orbit n ." He then asked the question that

Heisenberg had not considered worth asking. Were the “physically indistinguishable” states (mn) and (nm) to be counted as distinct or as identical? This question was inconsequential in classical mechanics, but in the matrix formalism, it implied a choice between two different matrix representations. In one, the matrix elements corresponding to the transitions $(mn) \rightarrow (m'n')$ and $(mn) \rightarrow (n'm')$ would be represented by two separate matrix elements, in the other they would be represented by the same element. In principle, Dirac could have relied on the traditional statistics of particles to answer the question; in that case, the answer would have been that the two states were distinct except when $m = n$. In the previous section, we saw that Heisenberg had taken this course in his treatment of the helium atom. He had simply adopted the first representation and had applied symmetry considerations only to the values of the matrix elements representing the radiation emitted and absorbed by the system of interacting particles, thereby deducing the impossibility of transitions between two groups of terms. Dirac, who possibly had learned from Born of Heisenberg’s linkage of quantum mechanics and Bose-Einstein statistics, instead chose to ignore the prescription of the old statistics. He asserted that the two transitions, $(mn) \rightarrow (m'n')$ and $(mn) \rightarrow (n'm')$, were “physically indistinguishable” and that “only the sum of the intensities for the two together could be determined experimentally” (Dirac 1926, 667). From this proposition he drew the answer:

Hence, in order to keep the essential characteristic of the theory that it shall enable one to calculate only observable quantities, one must adopt the second alternative that (mn) and (nm) count as only one state. (Dirac 1926, 667)

Having so fixed the matrix formalism, Dirac applied his formula for the general solution of the two-particle model. He formed the eigenfunctions of the whole system as linear combinations of products of the eigenfunctions of the single electrons; then, he imposed the condition that they correspond to the matrices. This condition could be satisfied only by combinations that were symmetrical or antisymmetrical under exchange of the electrons. Either one of these two possibilities gave “a complete solution of the problem” and quantum mechanics did not dictate which was the correct one (Dirac 1926, 669). The choice, Dirac stated, was to be made by appealing to Pauli’s exclusion principle:

An antisymmetrical eigenfunction vanishes identically when two of the electrons are in the same orbit. This means that in the solution of the problem with antisymmetrical eigenfunctions there can be no stationary states with two or more electrons in the same orbit, which is just Pauli’s exclusion principle. (Dirac 1926, 669–670)

The symmetrical solution, however, could not be correct for “the problem of electrons in an atom” because it allowed any number of electrons in the same orbit (Dirac 1926, 670). These results could be straightforwardly extended to any system composed of similar particles, in particular, to an assembly of molecules. Dirac thus applied them to the ideal gas. He obtained the eigenfunction of the assembly by multiplying the single-molecule eigenfunctions and choosing either the symmetrical or the antisymmetrical linear combinations. At this point, he turned to statistical considerations. He implicitly made the assumption that the new states, represented by symmetrical and antisymmetrical wavefunctions, represented the energy distributions, or macrostates, of statistics. Then, he explicitly adopted as a “new assumption” the simplest extension of Bohr’s rule, namely, that “all the stationary states of the assembly (each represented by one eigenfunction) have the same a priori probability” (Dirac 1926, 671). In the case of symmetrical eigenfunctions, this rule corresponded to the Bose-Einstein statistics. In the case of the antisymmetrical eigenfunctions, whereby the number of molecules associated with each single-particle eigenfunction could only be 0 or 1, it led to the new statistics that is now known as the Fermi-Dirac statistics. Dirac concluded:

The solution with symmetrical eigenfunctions must be the correct one when applied to light quanta, since it is known that the Einstein-Bose statistical mechanics leads to Planck’s law of black-body radiation. The solution with antisymmetrical eigenfunctions, though, is probably the correct one for gas molecules, since it is known to be the correct one for electrons in an atom, and one would expect molecules to resemble electrons more closely than light quanta. (Dirac 1926, 672)

Despite having just derived the two quantum statistics from the same set of assumptions (with the difference of the Pauli principle), Dirac separated them starkly in their applicability. His integration of quantum statistics and quantum mechanics was thus sealed with an uncompromising rejection of Einstein’s analogy between light quanta and material corpuscles. Dirac, in fact, had already rejected the matter-radiation analogy two years earlier:

For the discussion of equilibrium problems, quanta of radiation cannot be regarded as very small particles moving with very nearly the speed of light. There are two important points in which this picture is inadequate. In the first place the small particles could not (according to ordinary statistical theory) have any stimulating effect on processes by which they are emitted, and they should therefore be distributed in momentum according to Maxwell’s law, which is

the same as being distributed in energy (or frequency) according to Wien's radiation law. Secondly the concentration of quanta in thermodynamical equilibrium is not arbitrary, as is the case with all kinds of material particles, but is a definite function of temperature. (Dirac 1924, 594)

Moreover, when he first studied Einstein's quantum gas theory, he followed Einstein's interpretation of Bose's method as a statistics of non-independent entities. If the gas molecules were statistically distributed as light quanta, then they would have to be "not distributed independently from one another," and hence there would have to be "some kind of interaction between them" (Dirac 1925, 7). For Dirac, therefore, the analogy of ideal gas and heat radiation was invalid because non-interacting molecules had to be statistically independent, while light quanta were known not to be so. His categorical separation of the domains of applicability of the two quantum statistics preempted the possibility to interpret his new statistics in the same way as the statistics of Einstein and Bose, that is, as a statistics of non-independence.

A retrospective comparison with the modern understanding of quantum statistics brings into relief what Dirac's argument was *not* about. He did not argue that his formal-observational symmetry signified any modification of the traditional model of particles. More specifically, he did not propose that in the new mechanics, the particles were any more indistinguishable than they already were in classical mechanics. He did not suggest that the particles lost their identity. He flatly ignored the possibility that they had wave-like properties. Finally, he did not even extend the interpretation of the Bose-Einstein statistics as a statistics of non-independent objects to his new statistics. The only implication that he drew from his implementation of the formal-observational symmetry was a confirmation of the fundamental difference between material particles (electrons, atoms and molecules) and light quanta.

Dirac's integration of quantum statistics and quantum mechanics rested squarely on the corpuscular interpretation of the matrix formalism, supplemented by a formal restriction on the solutions of the wave equation. Dirac justified the formal restriction in terms of observational symmetry. Planck had already invoked the symmetry of a gas under exchanges of equal molecules in the quantum-statistical theory of gas prior to Bose. He and others used it to rationalize the subtraction of a term depending on the number of particles from the expression of the entropy (Darrigol 1991; Desalvo 1992; Monaldi 2009). Schrödinger had compared the entropy formula obtained in this way with the formula derived by Einstein, and he had concluded that the correct implementation of Planck's exchange symmetry was the statistics of non-independence proposed by Einstein. He then floated the suggestion that the proper way to understand this

odd new statistics might be to regard the whole gas as a system endowed with a “symmetry number” equal to $n!$ rather than as an assembly of n independent individual molecules (Schrödinger 1925, 438). Therefore, a viable, if still sketchy, model based on exchange symmetry was available to Dirac to support his symmetry requirement. Nonetheless, he was clear and explicit that for him the state (mn) was not an unspecific state of the whole system endowed with symmetry, but a state “in which one electron is in an orbit labelled m , and the other in the orbit n ” (Dirac 1926, 667). In other words, Dirac remained faithful to the individuality of particles typical of the corpuscular interpretation of matrix mechanics, skirting the implications of his new statistics for the independence of electrons and gas molecules.

The fact that Dirac disregarded the undulatory interpretation of the wave function and the consequences of antisymmetry for the independence of material particles does not mean that he refrained completely from any interpretation of the general solution of the wave equation. He did put forward an interpretation in the last section of the paper, in which he outlined a perturbation theory and fruitfully put it to use. He wrote the wave equation of “an atomic system subjected to a perturbation from outside (e.g., an incident electromagnetic field),” and showed that the general solution could be written as $\Psi = \sum_n a_n \Psi_n$, where the Ψ_n were the wave functions associated with the stationary states of the unperturbed atom, and the a_n coefficients depending on time. He thus deftly switched interpretive models. He proceeded to consider the general solution as no longer representing an atom but an assembly of atoms, and to assume that the square modulus of the coefficient a_n represented “the number of atoms in the n th state” (Dirac 1926, 646–647). The general solution now was a new theoretical representation of a multiparticle system that avoided any representation of individual particles and therefore bypassed, for the time being, the question of whether two states differing only by particle exchange should be counted as a distinct or identical. Determining the time evolution of the a_n under the effect of the perturbation, Dirac was then able to derive the coefficients of absorption and stimulated emission of Einstein’s theory of radiation, under the restricting condition that the initial phases of the atoms could be averaged. In this part of the paper, Dirac made no use of the results of the previous section. Neither did he subject the assembly of atoms to the exclusion principle that he had just posited to apply to gas molecules, nor did he suppose that the atoms followed the Einstein-Bose statistics.

Dirac returned to the difference between electrons and light quanta and the emission and absorption of radiation half a year later, after having spent several months at Bohr’s institute in Copenhagen formulating the general transformation theory and a general statistical interpretation of it (Dirac 1977; Kragh 1990; Darriгол 1992b). As a result of that work, he was able to forge a link between his two

representations of multiparticle systems for the case of light quanta, and thereby launched quantum electrodynamics. He first provided a quantum-mechanical theory of the radiation field using the energies and phases of the Fourier components as dynamical variables. Then, adapting his earlier tentative interpretation of the general solution of the wave equation, and helping himself with some nimble assumptions, he built a formal equivalence between the Hamiltonian of the field and the Hamiltonian of an assembly of equal particles (Darrigol 1986; Schweber 1994). He stressed, however, that the equivalence worked only if the particles obeyed the Bose-Einstein statistics, and this meant that it could only work for light quanta. Although Dirac now enlarged the term “particles” to include these entities, it was clear that for him they still did not belong in the same category as electrons. He especially warned that the classical “light wave” did not coincide with the “de Broglie or Schrödinger wave associated with the light quanta.” Therefore, even though there was a “de Broglie or Schrödinger wave” associated with each electron, there could not be a corresponding field description for electrons with them (Dirac 1927, 241).

In addition, Dirac now had at his disposal the statistical interpretation of the theory. He was therefore able to reaffirm the connection between symmetric wave functions and Bose-Einstein statistics without the need for an additional assumption concerning the equal probabilities of the stationary states. He pointed out that the wave function of an assembly of particles could now be interpreted “in the usual manner.” By this, he meant that the wave function no longer gave “merely the probable number” of particles in any state, but gave also “the probability of any given distribution” of the particles over the various states. But the probability calculated in this way did not agree with the probability calculated “from elementary considerations” for an assembly of independent particles. It agreed with the probability of an assembly of particles that obeyed the Bose-Einstein statistics, thus confirming the latter as a statistics of non-independence. In spite of the generality of the statistical interpretation, Dirac neglected to extend the same conclusion to the Fermi-Dirac statistics (Dirac 1927, 251). For a time, he continued to regard the two statistics as revealing two fundamentally different ontologies. As he recalled in his memoir, at first he disliked Jordan and Wigner’s extension of his field quantization technique to particles obeying the Fermi-Dirac statistics. His objection was that in the case of Bose-Einstein statistics, one could form “a definite picture underlying the basic equations, namely the picture that the theory could be applied to an assembly of oscillators.” No corresponding picture was available for the Fermi-Dirac statistics, and Dirac felt that this “was a serious drawback” (Dirac 1977, 140).

5.4 Conclusion

Despite their similarities, the works of Heisenberg and Dirac on the quantum mechanics of multiparticle systems present significant differences. Heisenberg was driven by the need to solve the problem of the helium atom and by the antagonism between his corpuscular interpretation and Schrödinger's undulatory interpretation of quantum mechanics. He relied on the phenomenon of quantum resonance, for which an interaction among the particles was necessary. Although he discarded classical kinematics, he took for granted that two states of free particles that differed only by particle exchange should be represented as distinct, except for the cases in which the particles had the same energy. He mistook the Bose-Einstein statistics for a statistics compatible with Pauli's exclusion principle, and he saw the transition from the old to the new statistics as the effect of a splitting of the energy states caused by resonance plus the natural selection of the single set of states that verified the exclusion principle. Dirac, in contrast, sidestepped the wave model and was mainly motivated by his mathematical reformulation of the wave formalism. Neglecting all interactions among the particles, he relied mainly on the symmetry of a system of equal particles and on the metatheoretical precondition that the theory yield only observable quantities. On this basis, contrary to Heisenberg, he concluded that two states differing only by particle exchange should be represented as a single state. This led him to the requirement of either symmetrical or antisymmetrical stationary states. He could then relate the symmetrical states to the Bose-Einstein statistics, and the antisymmetrical states to the exclusion principle and to the Fermi-Dirac statistics. Unlike Heisenberg, Dirac drew a categorical distinction between material corpuscles, all of which in his view followed the Fermi-Dirac statistics, and light quanta, which followed instead the Bose-Einstein statistics.

The comparison also brings to light a deeper commonality. Although these two papers played a prominent role in the interaction of quantum statistics and wave-particle duality, neither of them was produced on the basis of a revision of the traditional concept of particles. On the contrary, they both reveal a firm reliance on the corpuscular interpretive model, which continued to dominate the theoretical imagination of the two theorists, notwithstanding their overthrow of the classical representation of motion and their formalistic stances. Heisenberg's traditional assignation of statistical weights within the context of quantum mechanics and Dirac's drastic separation of the two quantum statistics can be seen as symptomatic of a remarkable theoretical resilience of the corpuscular model. More specifically, they point to the relative autonomy and persistence across theory changes of two basic characteristics of the concept of individual particles, their perfect similarity and their mutual statistical independence.

Abbreviations and Archives

AHQP	Archive for History of Quantum Physics. American Philosophical Society, Philadelphia
Einstein Archives	Hebrew University of Jerusalem

References

- Amaldi, Edoardo, Enrico Persico, Franco Rasetti, and Emilio Segrè, eds. (1962). *Enrico Fermi. Collected Papers (Note e Memorie)*. Vol. 1. *Italy 1921–1938*. Chicago: The University of Chicago Press.
- Beller, Mara (1999). *Quantum Dialogue: The Making of a Revolution*. Chicago: The University of Chicago Press.
- Bitbol, Michel (2007). Schrödinger against Particles and Quantum Jumps. In: *Quantum Mechanics at the Crossroads: New Physics, Novel Historical and Philosophical Views*. Ed. James Evans and Alan S. Thorndike. Berlin: Springer. 81–100.
- Bohr, Niels (1918). On the Quantum Theory of Line Spectra. In: *Det Kongelige Danske Videnskabernes Selskab. Skrifter. Naturvidenskabelig og Matematisk Afdeling, Part 1*. Series 8, Volume 4, Number 1. 1–36.
- Born, Max, Werner Heisenberg, and Pascual Jordan (1925). Zur Quantenmechanik II. *Zeitschrift für Physik* 35: 557–615.
- Camilleri, Kristian (2009). *Heisenberg and the Interpretation of Quantum Mechanics. The Physicist as Philosopher*. Cambridge: Cambridge University Press.
- Carson, Cathryn (1996). The Peculiar Notion of Exchange Forces—I. Origins in Quantum Mechanics, 1926–1928. *Studies in History and Philosophy of Modern Physics* 27: 23–45.
- Darrigol, Olivier (1986). The Origin of Quantized Matter Waves. *Historical Studies in the Physical and Biological Sciences* 16: 197–253.
- (1991). Statistics and Combinatorics in Early Quantum Theory, II: Early Symptoma of Indistinguishability and Holism. *Historical Studies in the Physical Sciences* 21: 237–298.
- (1992a). Schrödinger’s Statistical Physics and Some Related Themes. In: *Erwin Schrödinger. Philosophy and the Birth of Quantum Mechanics*. Ed. Michel Bitbol and Olivier Darrigol. Gif-sur-Yvette: Éditions Frontières. 237–276.

- Darrigol, Olivier (1992b). *From c-Numbers to q-Numbers. The Classical Analogy in the History of Quantum Theory*. Berkeley, CA: University of California Press.
- Desalvo, Agostino (1992). From the Chemical Constant to Quantum Statistics: A Thermodynamic Route to Quantum Mechanics. *Physica* 29: 465–537.
- Dirac, Paul A. M. (1924). The Conditions for Statistical Equilibrium Between Atoms, Electrons, and Radiation. *Proceedings of the Royal Society A* 106: 581–596.
- (1925). Einstein-Bose Statistical Mechanics. *AHQP (unpublished manuscript)* M/f No. 36, Sect. 009-001: 36–39.
- (1926). On the Theory of Quantum Mechanics. *Proceedings of the Royal Society A* 112: 661–677.
- (1927). The Quantum Theory of Emission and Absorption of Radiation. *Proceedings of the Royal Society A* 114: 243–265.
- (1977). Recollections of an Exciting Era. In: *Storia della Fisica del XX Secolo. Rendiconti della Scuola Internazionale di Fisica "Enrico Fermi," LVII Corso*. Ed. Charles Weiner. New York: Academic Press. 109–146.
- Einstein, Albert (1924). Quantentheorie des einatomigen idealen Gases. *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*: 261–267.
- (1925). Quantentheorie des einatomigen idealen Gases. Zweite Abhandlung. *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*: 3–14.
- Fermi, Enrico (1926). Sulla quantizzazione del gas perfetto monoatomico. *Rendiconti della Accademia dei Lincei* 3: 145–149.
- Heisenberg, Werner (1926). Mehrkörperproblem und Resonanz in der Quantenmechanik. *Zeitschrift für Physik* 38: 411–426.
- Jammer, Max (1966). *The Conceptual Development of Quantum Mechanics*. New York: McGraw-Hill.
- Kragh, Helge (1990). *Dirac: A Scientific Biography*. Cambridge: Cambridge University Press.
- Mehra, Jagdish and Helmut Rechenberg (1987). *The Historical Development of Quantum Theory, Volume 5: Erwin Schrödinger and the Rise of Wave Mechanics. Part 2: The Creation of Wave Mechanics; Early Response and Applications 1925–1926*. Berlin: Springer.
- Monaldi, Daniela (2009). A Note on the Prehistory of Indistinguishable Particles. *Studies in History and Philosophy of Modern Physics* 40: 383–394.
- Schrödinger, Erwin (1925). Bemerkungen über die statistische Entropiedefinition beim idealen Gas. *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*: 434–441.

- Schweber, Silvan S. (1994). *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga*. Princeton: Princeton University Press.
- Wigner, Eugene (1927a). Über nicht-kombinierende Terme in der neueren Quantentheorie. I. Teil. *Zeitschrift für Physik* 40(7): 492–500.
- (1927b). Über nicht-kombinierende Terme in der neueren Quantentheorie. II. Teil. *Zeitschrift für Physik* 40(11-12): 883–892.

Chapter 6

***Pourparlers* for Amalgamation: Some Early Sources of Quantum Gravity Research**

Dean Rickles

In a lecture delivered to the British Institute of Philosophy on 15 November 1932, Sir Arthur Eddington wrote in surprisingly modern terms of the problem of merging quantum theory and the general theory of relativity:

At present theoretical physics is divided into macroscopic theory and microscopic theory, the former dealing with systems on a scale comparable with our gross senses, and the latter dealing with the minute substructure underlying the gross phenomena. Broadly speaking, relativity theory covers macroscopic phenomena and quantum theory the microscopic phenomena. The two theories must ultimately be amalgamated, but at present we have not got much beyond the *pourparlers* for amalgamation. So the gap exists—not, however, as a gap in the external world, but as a gap in our understanding of it. (Eddington 1933, 30)

This suggests that even as far back as 1932 the problem of quantum gravity already had some historical pedigree (there had been *pourparlers*, as Eddington puts it). Though in somewhat different terms, given the rapid and radical developments in physics at that time, Eddington himself had been thinking about the problem since at least 1918 (see 6.3.1 below). Yet, in the handful of historical studies of quantum gravity that exist,¹ it is claimed that quantum gravity research originated with Léon Rosenfeld’s “pioneering” pair of papers from 1930, forging, as they indeed did, both the canonical and covariant quantizations of the gravitational field—though lip service has at least been paid Albert Einstein’s own early prophetic remarks about the potential conflict between general relativity and quantum theory, in papers of his from 1916 onwards,² but this is still fol-

¹See, for example (Stachel 1998) and (Rovelli 2002).

²See (6.2) below on these studies. Also see (Stachel 1998), in which he considers an earlier and slightly more diverse group of actors. Another excellent study—though with a focus more on unified field theories than quantum gravity—is (Goldstein and Ritter 2003). This traditional view that

lowed by the claim that nothing was really done about it until Rosenfeld tackled the problem (prompted by Wolfgang Pauli).

However, though the version of quantum theory involved was significantly different (and certainly less systematic and coherent) from the present framework (or frameworks)—stemming from the pivotal 1925–1928 developments that produced quantum mechanics and quantum field theory—there was nonetheless a rich debate about the relationship between (old) quantum theory (and the atomic physics that preceded it)³ and gravity. This interaction is to be expected for two reasons:

1. the two frameworks were constructed over much the same period of time, often by the same architects;
2. the right-hand side of the Einstein field equations must naturally include (in some way) contributions from the matter best described by the quantum theory.

It is true that this work rarely, if ever, involved consideration of *quantum properties of the gravitational field*, but this could equally be said of some recent approaches that I am nonetheless perfectly content to label “quantum gravity.”⁴ When I speak of quantum gravity in this paper, I mean it very liberally to indicate any approach that involves dealing with the problem of the coexistence of quantum systems⁵ and gravitation.

In some ways, this early work closely mirrors what would come later, and indeed, many of the key notions of the later work—the importance of the Planck scale in demarcating the domain of applicability of classical general relativity, the experimental inaccessibility of quantum gravity, and the potentially radical revision of space-time concepts—were discussed even at this early stage. This must surely have contributed to later work, at least in establishing a mindset for thinking about the amalgamation of quantum and gravity. This work deserves

quantum gravity originates with Rosenfeld can be traced back at least as far as Bryce DeWitt’s brief historical review in the first of his three *Physical Review* papers on quantum gravity in 1967 (DeWitt 1967).

³Of course, “atomic” certainly does not mean “quantum”. However, many of the issues that were discussed in the context of merging atomic and gravitational physics are nonetheless relevant for later work on quantum gravity since they often involve overlapping concerns such as implementing discreteness, singular behavior, and so on, in the context of a theory like general relativity. For this reason I often discuss non-quantum, atomic physics. It will be clear from the context when this is the case. It is my contention that if one is looking for sources of the earliest examples of genuine quantum gravity research, then one cannot ignore these non-quantum examples.

⁴An example of such quantum gravity without the *quantization* of gravity are the so-called “emergent gravity” proposals. See (Hu 2009) for a nice review of these.

⁵Where, as I already indicated, this might sometimes be stretched to strictly non-quantum situations, such as early atomic physics when this physics has some properties relevant to the full quantum context.

to be discussed under the banner of “quantum gravity” just as much as any that would come later, and any attempt to force quantum gravity into the mould of the *quantization of the gravitational field* (which is implicit in the Rosenfeld story) misses out on both an important source of later ideas and an important set of ideas in their own right.

In this paper I discuss these early skirmishes into and around the problem of quantum gravity, from their prehistory (close to the creation of general relativity) to the development of wave mechanics and just prior to the Dirac equation—at which point the course of quantum gravity research, quite naturally, radically shifts, and there is an appropriate context for the kind of work carried out by Rosenfeld. These early approaches also offer a very useful probe for investigating several important agendas that were in operation at that time, including a desire to meld the cosmological and microscopic, and to unify both physics and our *knowledge* of physics via axiomatic foundations. A future task is to consider whether and how these agendas continued to play out in the later developments.

For reasons of space and convenience, I focus heavily on papers appearing in the journal *Nature* during this period. Though it has a potential to introduce an English bias, it nonetheless gives a good snapshot of the general state of research, since it was common practice to supplement more technical publications (e.g., in *Zeitschrift für Physik* or the *Proceedings of the Royal Society*) with a brief note in *Nature*⁶ describing the key findings. Indeed, it perhaps offers a closer glimpse into the state of play since a feature of *Nature*, especially at that time, is almost a direct personal correspondence between individuals that can be found in its pages, via responses in notes and letters—something that could be easily carried out on account of its weekly publication.

Though quantum gravity is today viewed as a slightly strange problem on the frontiers of physics (no doubt because of its highly theoretical and mathematical nature), in the early days surrounding their creation, a fairly natural dialogue between quantum theory and general relativity took place.⁷ Again, I am here adopting an enormously *liberal* characterization of *the problem of quantum gravity* so as to offer an account as inclusive as possible, thus minimizing the risk of sidelining what may have been important sources of later ideas, Eddington’s *pourparlers* for amalgamation. Whether any of the approaches I discuss amount to quantum gravity in themselves is of course highly debatable, but then there is *no* such theory yet.

⁶These rather non-technical accounts were often duplicated in German in *Naturwissenschaften*.

⁷Or, in somewhat less anachronistic terms, in the case of the very earliest such work, this was a dialogue between the puzzling behavior of discrete matter then observed to behave in an increasingly curious way, and gravitation.

6.1 The Torch of Unification: Mie, Hilbert, Weyl and Haas

I do not wish to cover the already well-trodden ground of the genesis of quantum mechanics and general relativity.⁸ However, a number of the initial forays into quantum gravity were in many ways extensions of debates that were conducted *before* the creation of these theoretical frameworks—hence, I am concerned with the *pre-pre-history* here. One can also find alternative accounts of gravitation—in particular those based on an electromagnetic ether—persisting well beyond the creation of general relativity. One such approach that was clearly of importance (not least to David Hilbert, Hermann Weyl and Eddington) was that of Gustav Mie. I begin with a brief account of Mie’s approach since it marks an approach to the amalgamation of the central theories of physics, as it was then, on the precipice of radical changes.

In 1912 and 1913, Mie sought to develop a (reductive) unified field theory in which both matter and force could be derived from the electromagnetic field (understood as an emergent property of the ether) alone. The core object was a “world function,” and the derivation of gravity and electromagnetism would proceed from this via the calculus of variations—he was not able to get a fully unified theory of both phenomena: Mie’s theory was based on two scalar potentials. The general thrust of Mie’s program stemmed from his belief in the significance of the new empirical facts about the behavior of atoms that had recently emerged. In 1912 he wrote:

The significance of the recently acquired empirical facts about the nature of the atoms ultimately amounts to something essentially only negative, namely that in the atoms’ interior the laws of mechanics and Maxwell’s equations cannot be valid. But regarding what should replace these equations in order to encompass from a single standpoint the profusion of remarkable facts associated with the notion of quantum of action, and in addition the laws of atomic spectra and so forth, the experimental evidence is silent. In fact, I believe that one must not expect anything like that from experiment alone. Experiment and theory must work hand in hand, and that is not possible as long as the theory has no foundation on which it can be based. Thus it seems to me absolutely necessary for further progress of our understanding to supply a new foundation for the theory of matter. With this work, I have tried in the following to make a start, but in view of the difficulty of the matter one should not right away expect

⁸For the state of the art of both quantum mechanics and general relativity, consult (Galison, Gordin, and Kaiser 2001) and (Renn 2007) respectively.

results accessible to experiment. The immediate goals that I set myself are: to explain the existence of the indivisible electron and: to view the actuality of gravitation as in a necessary connection with the existence of matter. I believe one must start with this, for electric and gravitational effects are surely the most direct expression of those forces upon which rests the very existence of matter. It would be senseless to imagine matter whose smallest parts did not possess electric charges, equally senseless however matter without gravitation. Only when the two goals I mentioned are reached will we be able to consider making the connection between the theory and the complex phenomena mentioned above. (Mie 2007, 1554)

It is quite clear here the extent to which, in Mie's mind, gravitation and electromagnetism, and matter are all manifestations of one and the same stuff—this worldview would spread to Weyl, Hilbert, Eddington and many others. Out of this mixture, Mie expected the phenomenological facts of quantum theory to emerge—this theory might somewhat perversely be viewed as an approach to quantum gravity that *predates* both general relativity and quantum mechanics. At this time in 1913, there were three other (serious) alternative theories of gravitation, those of Einstein and Marcel Grossmann (the *Entwurf* theory), Abraham Pais, and Gunnar Nordström. Abraham's theory was inadequate in several ways, chief amongst these being the incompatibility between the variable light speed adopted by Abraham and his usage of the Lorentz transformations. I will return to Nordström's theory below, for now I quickly explain how Mie's theory influenced Hilbert.

Hilbert is not a name usually associated with quantum gravity research, but he figures centrally in several early episodes, some of which played a crucial role in later work (at the time of writing, in fact).⁹ Hilbert was directly involved in aspects of the birth of general relativity, using his beloved variational principles to derive the field equations. He was led to this approach by a rather indirect route involving a modification of Mie's electromagnetic theory of matter and force: Hilbert made direct use of Mie's theory in his derivation of general relativity. This imposed a severe restriction on the form of the theory, since it depended upon a specific matter-source. As Pauli put it in his encyclopedia article on general relativity:

Hilbert's presentation [...] was not quite comfortable for the physicists, because in the first place he axiomatically defined the variational principle, and, which is more important, his equations were

⁹I am indebted in this section to the hard work of Leo Corry, David Rowe, Tilman Sauer and others in unpacking the complex relations between Mie, Hilbert and Weyl.

expressed not for an arbitrary material system, but were based on Mie's theory of matter. (Pauli 1921, 211)

Einstein¹⁰ famously found Hilbert's approach "childish" since it did not show a proper awareness of the "pitfalls of the real world"—the reason Einstein adopted a "principle theory" approach was, of course, precisely to avoid such pitfalls (in this case "risky hypotheses about the structure of the electron"). The axiomatic method was therefore not a good basis for physical theorizing, according to Einstein.

Hilbert's ideas about the foundations of mathematics (and his emerging ideas about the unity of scientific knowledge) was combined with this physical background in his celebrated work on general relativity in 1915. What is interesting about this work for this project is that, like Mie, he quite clearly believed that the physics of gravity would be able to unlock the secrets hidden in atomic processes:

As one sees, the few simple assumptions expressed in Axioms I and II suffice by sensible interpretation for the development of the theory: through them not only are our conceptions of space, time, and motion fundamentally reformulated in the Einsteinian sense, but I am convinced that the most minute, till now hidden processes within the atom will become clarified through the fundamental equations herein exhibited and that it must be possible in general to refer all physical constants back to mathematical constants—just as this leads to the approaching possibility, that out of physics in principle a science similar to geometry will arise: truly, the most glorious fame of the axiomatic method, while here, as we see, the mighty instruments of analysis, namely the calculus of variations and invariant theory, are taken into service. (Hilbert 1915, 407)

Whether it was through interactions with Einstein or Weyl, or self-realization about the magnitude of the task he had set himself, Hilbert was a little more subdued a couple of years later. In 1917, Hilbert spoke on "Axiomatic Thought" to the Swiss Mathematical Society in Zurich. At the root of his talk was a belief in the unity of scientific knowledge, with mathematics as the linchpin holding it all together:

I believe that everything which can be the subject of scientific thought, as soon as it is ripe enough to constitute a theory, falls

¹⁰Letter from Einstein to Weyl, dated 23 November 1916, cited in (Rowe 2003, 65); see also (Sauer 2002).

within the scope of the axiomatic method and thus directly to mathematics. By pursuing ever deeper-lying layers of axioms [...] we gain ever deeper insights into the essence of scientific thought itself and we become ever more conscious of the unity of our knowledge. In the name of the axiomatic method, mathematics appears called upon to assume a leading role in all of science. (Hilbert 1918, 156)

Weyl was in attendance at this talk and would adopt a formal approach to the problem of unification of gravitation and electromagnetism (though not quantum theory) that was methodologically similar to Hilbert's. Weyl generalized Riemannian geometry, adding a principle of calibration or gauge ("eich") invariance to account for the non-integrability of length transference (over non-infinitesimal distances). Einstein had similar gripes with this approach: in this case, the theory had the absurd consequence that objects taken around different paths having identical origins and termini will, at the point of termination, be found to have different sizes and rates.¹¹

The torch of "unity through axiomatization" was carried on, in a rather different way, by Arthur Haas. Haas was a strong advocate of Hilbertian axiomatization, and in his case, it led him to early speculations about matters related to the problem of quantum gravity. For example, as early as 1919, Haas writes (on the basis of "unification" ideals) that:

Arguably, one of the most important future tasks of the axiomatization of physics is the implementation of quantum theory in the system of the general theory of relativity. (Haas 1919, 749)

Though he does not explicitly name the individual constants associated with the ingredient theories (*viz.* c , \hbar , G), it is reasonable to surmise that this is what Haas had in mind in the following passage:

The main task of the axiomatization of physics will be the problem concerning the integration of the universal constants of physics. Also the solution of this question may be expected to reveal deeper knowledge of the relations, only intimated by Hilbert, holding between gravity and electricity, and of a further integration of these relations with the quantum hypothesis. (Haas 1919, 750)

This interpretation is somewhat strengthened by the fact that Haas went on to consider the various possible combinations of other constants in other contexts, investigating the way they demarcate domains (Haas 1938). In many ways, this

¹¹ See (Scholz 2001) for a fine discussion of this episode along with a translation of Weyl's text.

idea that we must consider the integration of the fundamental constants to solve the problem of the relationship between quantum theory and gravitation coincides with the modern understanding.¹²

6.2 Einstein on the Relationship between Gravity and Quantum

It is a little curious that so many great revolutionary episodes happened almost simultaneously at the beginning of the twentieth century. Perhaps one revolution made it easier for others to follow, via some kind of snowball effect. Whatever the reason, the revolution that resulted in general relativity and the revolution that resulted in quantum theory were close neighbors in time. Einstein was profoundly involved in the creation of both theoretical frameworks, though the former more so than the latter. At the time of the construction of the general theory of relativity, he firmly believed in the existence of quanta of radiation. But this only involved a belief in the property of discreteness (with no real sense of ontological substrate beyond this), rather than belief in what would become quantum mechanics (or quantum field theory—though here too his contributions on emission and absorption of radiation proved crucial). Most physicists believe *another* revolution is required to bring quantum theory and general relativity together (Rovelli 2000).

Since such quanta, with their discrete energies and other properties, would inevitably couple to the gravitational field (in however small a way), Einstein could not ignore the fact that *something* would need to be said about the nature of this interaction.¹³ Even before his article on general relativity had been published in its final form, Einstein was in correspondence with Arnold Sommerfeld about its possible relationship with quantum theory.¹⁴ Einstein heard about Sommerfeld's new theory of spectral lines first-hand while he was still working on general relativity. Sommerfeld thought that the general theory of relativity might be able to offer some help in resolving problems caused by the Stark effect (Sommerfeld 2000, 438). It is quite likely that Sommerfeld's willingness to consider the relationship between what looked at this stage like disparate fields of inquiry was

¹²Gennady Gorelik (1992) assigns the discovery that these fundamental constants might point to the limits of present physical theory to Matvei Bronstein. While I agree that Bronstein was the first to produce an explicit account of the nature of this limitation in the 1930s, I show later that Eddington also made similar claims in 1919.

¹³A little later it would also come to be understood that there is a "formal interaction" between general relativity and quantum objects stemming from the peculiar nature of fermions, including: objects with half-integer spins imposes a variety of constraints on the space-time structure, and therefore on the gravitational field (resulting in a slightly modified theory of gravitation). This was a rather slow lesson.

¹⁴Though it appears that it was Sommerfeld who led this exchange, fired up, no doubt, by the success of his application of *specialy* relativistic principles to quantum theory.

grounded in a similar belief system to that of Hilbert and Haas (and the Göttingen school), though with a far more empirical basis. Indeed, Silvan Schweber (2009, 269–278) notes that a “doctrine of pre-established harmony” (between mathematics and physics and mathematics and nature) can be found running through much of Sommerfeld’s earlier work. Given this, it is reasonable to expect some inner unity holding between so fundamental a pair of frameworks as relativity and atomic theory.¹⁵

Almost as soon as general relativity was completed, Einstein became aware of a possible conflict between it (or, more specifically, the existence of gravitational waves) and the principles of quantum theory,¹⁶ and, therefore, the need to say something about the problem of quantum gravity. Thus, he writes that:

[A]s a result of the internal-atomic movement of electrons, atoms must radiate not only electromagnetic but also gravitational energy, if only in minuscule amounts. Since this cannot be the case in nature, then it appears that the quantum theory must modify not only Maxwellian electrodynamics but also the new theory of gravitation. (Einstein 1916b, 696)

In this case, Einstein is clearly aware of the potential clash between the theoretically predicted gravitational radiation combined with the empirically observable stability of atoms: *any* moving mass (even the electrons in atoms) will radiate gravitational energy (given the right kind of motion, that is).¹⁷ In other words, something like Planck’s law of radiation would have to be found for gravitation to account for the stability. He repeated this claim again in 1918, stating that “an improved version of quantum theory would lead to changes in the gravitational theory” (Einstein 1918, 167).

This looks like a potential *empirical* motivation for pursuing quantum gravity. However, as Gorelik correctly points out, whilst atomic radiation (computed

¹⁵Norbert Wiener, who would spend much time in Göttingen, also seems to have become caught up in the general need for a harmonious structure at the foundations of physics. He writes in the second volume of his autobiography that “By 1925 [...] the world was clamouring for a theory of quantum effects which would be a unified whole and not a patchwork” (Wiener 1956, 105).

¹⁶As Helge Kragh has pointed out, the version of quantum theory that Einstein would have been thinking about at this early phase of general relativity’s development was precisely the Bohr-Sommerfeld theory (Kragh 2000, 965). Einstein would have been particularly impressed with the way the Sommerfeld theory integrated (special) relativity and quantum theory. Helmut Rechenberg claims that Sommerfeld published his results after Einstein informed him that, as one might expect, the general relativity would not modify the results in any appreciable way (Rechenberg 1995, 160).

¹⁷Though apparently not too troubled. In a letter dated 19 July 1916 (Buchwald et al. 2006, Vol. 10, 237a, 25) he writes breezily to his friend Heinrich Zangger after just completing this work (and his contemporaneous work on the quantum theory of emission and absorption of radiation), showing no signs of real concern over the fate of general relativity—though it is also very possible that his mind was preoccupied with the breakup of his marriage at this point.

along the lines of James Clerk Maxwell's theory) leads to the collapse of the atom in 10^{-10} seconds (a fact inconsistent with observations), atomic *gravitational* radiation, computed using Einstein's formula, has a collapse time of the order of 10^{37} seconds. Therefore, there would in fact be no empirical inconsistency as a result of gravitational radiation, and we should not be puzzled by the stability of atoms in this case.

Gorelik (1992, 365) argues that an "analogy with electrodynamics" lay behind this comment of Einstein's. This analogy was a persistent feature of early research on quantum gravity. One must also bear in mind that the issue of absorption and emission of radiation must have occupied a central place in his thinking at the time of writing, for his paper on the emission and absorption of radiation in quantum theory appeared very shortly afterwards—replete with the statement that "it does not seem to be doubtful that the basic idea of quantum theory must be maintained" (Einstein 1916a, 318). What is remarkable, given what we know of the certainty he professed about general relativity, is that he openly considered the possibility that the quantum theory would demand some kind of "modification" of general relativity!¹⁸

Quantum theory was invoked several times (in discussions of general relativity and unified field theories) to mark some kind of boundary of the *applicability* of a theory.¹⁹ Einstein himself expressed just this view in a lecture entitled "Ether and the Theory of Relativity" at the University of Leyden in October 1920. This address is interesting for many reasons, historical and philosophical. For our purposes, it is interesting because Einstein once again speculates on the possible restrictions that the quantum theory might place on general relativity:

Further, in contemplating the immediate future of theoretical physics we ought not unconditionally to reject the possibility that the facts comprised in the quantum theory may set bounds to the field theory beyond which it cannot pass. (Einstein 2002, 323)

Indeed, we can find several examples of Einstein expressing this kind of sentiment. Inasmuch as his comments (here and in his 1916 paper) have been investi-

¹⁸This openness of Einstein to the possibility of a quantum theoretical modification of general relativity would not last long of course, and was already beginning to sour at this stage. His taste for quantum theory soon dissolved to the extent that towards the end of his life, he was searching for ways to reproduce quantum mechanical phenomena using a purely classical field theory. Suraj Gupta (who developed a special-relativistic theory of quantum gravity in the 1950s) has a different (inverted) interpretation of Einstein's underlying reasons for distrusting quantum mechanics: "Because his theory is different from other field theories, he tried to construct unified field theories and because he could not see how his theory in the curved space could possibly be quantized, he criticized quantum mechanics" (Gupta 1962, 253).

¹⁹For example, Goldstein and Ritter (2003, 104) note how Weyl (1921) adopts this position in his *Raum, Zeit, Materie*, for which see (Scholz 2001).

gated by historians, it has tended to be in the context of the study of gravitational waves. It is true that gravitational waves are naturally involved here, but since Einstein is considering the possibility that the radiation of such waves is quantized, we ultimately have what can also be seen as heralding the beginning of research investigating the possible *quantization* of gravity.

6.3 Quantum Meets Gravity in the Pages of *Nature*

The pages of *Nature*, in the period immediately following the construction of general relativity, were littered with a variety of suggestions involving some kind of connection between gravitation and quantum phenomena.²⁰ For example, the period following the well-publicized 1919 observation to test Einstein's predicted value for the amount of deflection of starlight by the Sun resulted in a steady flow of papers probing the possible relationship between quantum theory and general relativity. This is a fairly natural line of inquiry given the context, since by this time light was, of course, understood in quantum theoretic terms and since gravity was having a direct effect on the propagation of light, it follows that there must be some relation between gravitation and quantum systems. Also the atomic theory of matter based on quantum theory was becoming established, which further deepened the need to consider the connection between gravity and matter in this form.

This episode is of wider historical interest since in many cases the articles were part of their authors' wider agenda, be it the unity of nature, a distaste for relativity, adherence to the axiomatization program, or some other underlying motivation. Indeed, what is striking about the issues of *Nature* in and around our chosen period is that there is seen to be no real division between the sciences, and certainly not between atomic physics and gravitational physics. This paper works in a largely chronological fashion, except where there are thematic links across years.

²⁰To reiterate what was said in the introduction, I am not solely focusing on proposals that aim to *unite* quantum theory and gravitation in a common framework. Rather, I am concerned with showing how quantum theory and quantum phenomena and general relativistic phenomena and general relativity did not occupy isolated conceptual schemes in the minds of physicists at this time but were very much intertwined. This often manifests itself in ways that have persisted into modern thinking on the problem of quantum gravity, such as the notion that merging quantum theory and general relativity might serve to resolve some internal problem with one or another ingredient theory. However, I also include less obviously interesting examples indicating merely that the problem of linking the two theories together was "in the air" so that the later work of Rosenfeld, for example, is seen as a fairly natural problem to focus on.

6.3.1 Eddington on Fundamental Length

Eddington discussed the relationship between gravitation and quantum theory from the period following the creation of general relativity to the end of his life.²¹ Eddington very frequently refers to quantum theory in the context of gravitation and vice versa. In fact, he began to consider the relationship between gravity and quantum at least a year *before* the deflection observation. It is highly likely that it was as a result of his (and Einstein's) work on gravitational waves that he was initially led to think about the problem for, as we have seen, Einstein had already contemplated the potential clash between quantum theory and general relativity as a result of his own work on gravitational radiation. Indeed, there are elements of Eddington's writing, in discussing the gravitational red shift, as far back as 1916 that suggest an emerging awareness of quantum theory's relevance: "The vibrations of an atom must be slower in an intense field, so that the lines of the solar spectrum should be displaced slightly to the red as compared with terrestrial spectra" (Eddington 1916, 330).

The second installment of this article (appearing in 1918) shows an even greater appreciation of the relationship. Certainly, one of the more remarkable things that emerges from Eddington's early work on general relativity is his claim that a fundamental length can be formed from the three basic universal constants, and that this length will inevitably form a piece of the future theory blending quantum and gravity:

From the combination of the fundamental constants, G , c , and h it is possible to form a new fundamental unit of length $L_{min} = 7 \times 10^{-28} cm$. It seems to be inevitable that this length must play some role in any complete interpretation of gravitation. [...] In recent years great progress has been made in knowledge of the excessively minute; but until we can appreciate details of structure down to the quadrillionth or quintillionth of a centimetre, the most sublime of all the forces of Nature remains outside the purview of the theories of physics. (Eddington 1918, 36)

This is a remarkably prescient passage; though it appears somewhat clumsily disconnected from the rest of the article appearing as the final paragraph. In it Eddington has clearly targeted what we now label "the Planck length," $\sqrt{hG/c^3}$.²²

²¹This quest has been discussed in some detail by several authors. Clive Kilmister (a student of a student of Eddington, namely George McVittie) focuses directly on Eddington's concern with the relation between gravity and quantum theory (Kilmister 1994). Ian Durham (2003) focuses more on Eddington's desire to achieve an *objective* account of physics independently of human measures.

²²Note that the value he derives is some six orders of magnitude off from the value we have today (namely $6.6 \times 10^{-34} cm$). Note also that Eddington does not mention Planck's name, so one might

This length is, of course, a fairly generic feature of all modern approaches to quantum gravity.²³ That Eddington believes this length to *inevitably* play a role in a future interpretation of gravitation plainly implies that he sees quantum theory as essentially bound up with the physics of gravitation.

In the paragraph immediately preceding the above quoted paragraph, Eddington states (again, rather presciently) that:

[W]e know that in consequence of the undulatory theory of light, a ray traversing a heterogeneous medium always takes the path of least time; and one can scarcely resist a vague impression that the course of a material particle may be the ray of an undulation in five dimensions. (Eddington 1918, 36)

Eddington clearly has in mind here a notion of the particle as a “projection” of a wave phenomenon down from five to four dimensions (as in the later more well-known Kaluza-Klein theories). One might immediately latch upon the work of Gunnar Nordström (1914) as a precedent for such five-dimensional speculations; though Nordström’s theory was based on a scalar theory of gravitation. However, given Nordström’s isolation it is highly likely that Eddington was not aware of his paper.²⁴ Although it is quite likely that Eddington’s off-hand remark might have sparked higher-dimensional thoughts in Theodore Kaluza and others.

Joseph Larmor, in discussing the possible application of quaternions to general relativity, suggests something strikingly similar (again independently, it seems, of Nordström and Kaluza),²⁵ and in a way that makes projective notions

wonder whether he came to the result (of a system of unique scales from combining these three constants) independently of Planck (1899). Of course, the idea that there is a smallest length would later become a common feature of quantum gravitational physics—Kragh (2000) provides a useful study of the earliest work (based on quantum considerations) on the notion of a minimal length. I might also add that Eddington is a missing figure from Gorelik’s otherwise superb recounting of the history of the role of the Planck units in the early history of quantum gravity research (Gorelik 1992).

²³Of course, these units derive initially from Planck’s system of “absolute units” (Planck 1899). But in that paper, Planck does not link this to any synthesis of quantum theory and gravitation, nor did he suggest that the absolute unit of length imposed any lower limit on objects and processes. He was, rather, impressed with their *independence* from the usual conventional elements involved in “terrestrial” units. That the units are just “pure numbers” encoding the laws of physics must have impressed Eddington, for just this connection would form the basis of his later (near-numerological) work on deriving the laws of physics from such pure numbers, in his last book *Fundamental Theory*, for example (Eddington 1949). I suspect (though it is not the place to discuss here) that this early realization about absolute units and their relationship with (objective descriptions of) physical laws might have played a greater role in Eddington’s later work than has previously been realized.

²⁴See (Halpern 2004) for a discussion of the (lack of) impact of Nordström’s proposal.

²⁵Daniela Wuensch points out that Larmor’s paper appeared before Kaluza’s, but argues that because Larmor had used flat space-time, it did not excite much interest (Wuensch 2003, 526). I do not think this can be the right explanation: Larmor’s approach involves flat Minkowski space only as a *projec-*

more explicit. He labels it a “hyperspatial version of the Einstein gravitational theory” (Larmor 1919b, 357). Larmor initially develops a symbolic geometrical calculus (devised by W. J. Johnston) to talk about electromagnetism in *flat* space-time (i.e., with $\delta s^2 = \delta x^2 + \delta y^2 + \delta z^2 + (ict)^2$). However, he is concerned with introducing gravitation into his scheme, and notes that this can be achieved by introducing a new dimension ξ (“preferably of space” (Larmor 1919b, 353)), such that

$$\delta s^2 = \delta x^2 + \delta y^2 + \delta z^2 + \delta \xi^2 + (ict)^2. \quad (6.1)$$

Since this includes electromagnetism too, an additional component is received by the vector potential. The idea is to have the physics of flat four-dimensional space-time “as a hypersurface within our auxiliary flat five-dimensional scheme, in which both the electrodynamic and the gravitational theory shall exist.” He develops this idea as follows:

Now any continuum of four dimensions, having a quadratic line-element, however complex, is expressible as a hypersurface in this homaloid continuum of five dimensions. If these considerations are correct, the Einstein generalisation, made with a view to include gravitation within his four dimensions, must be interpretable as the geometry of some type of hypersurface constructed in this extended homaloid of five dimensions. For the previous homaloid theory of Minkowski which ignored gravitation, this hypersurface, existing in the five dimensions, in which the world-process is represented, is flat; or more conveniently in some connections it may be taken as a closed region (hypersphere) of assigned uniform extremely small curvature, instead of the unlimited hyperplane. The problem then is to include in the scheme the influence—actually very slight in realizable cases—of gravitation; and this is to be done by recognising slight local deformations on this hypersphere in order to represent that effect. *Now in the four-dimensional Minkowski map of the historical world-process, the rays of radiation are the curves of minimum length on the locus for which the analytic element of length $\delta\sigma$ vanishes; and the paths of particles when gravitation was neglected were the curves (then straight lines in the flat) for which the length between assigned terminal points is minimum.* If the hypersurface, which is very nearly uniform of very small curvature in the actual

tive feature of the world. I find it more likely that both the cumbersome nature of the approach, the novelty and the heavily abstract formulation were more likely to be the reason that the approach did not catch on.

problem as presented in nature, can be so chosen that these two relations persist—namely, that the rays of light shall be geodesics on the locus determined by $\delta\sigma$ vanishing, and the free orbits of particles with gravitation now introduced shall be the paths of minimum length on the hypersurface—then one way of absorbing the universal phenomena of gravitation, into the mixed space-time scheme which has arisen from and has transcended and obliterated the previous idea of relativity of positions and motions, will have been accomplished.²⁶ (Larmor 1919b, 354)

Larmor associates this idea of generalizing dimensions (and dealing with the properties of one as projections in another) with Clifford. These several proposals for “dimensional expansion” indicate that when Kaluza formulated his five-dimensional formulation of gravitation and electromagnetism, he did so in a period when such speculations were not entirely rare. Of course, this idea of increasing the number of space-time dimensions is a central feature of string theoretic approaches to quantum gravity.

6.3.2 Larmor’s Paradox

In the Christmas Day edition of *Nature* in 1919, Larmor drew attention to a potential conflict (a paradox, in fact)²⁷ between the quantum theory of light and the manner in which light is treated in general relativity, to raise doubts about the veracity of the latter. As Larmor saw it (Larmor 1919a, 516), Einstein’s general theory demands on the one hand (given an undulatory description of light) that the velocity of light will be diminished as it nears the Sun, but that “the scale of time” must undergo a compensatory expansion so that, overall, there is no change in wavelength. Larmor refers to such a notion of time as “heterogeneous time” and argues that given this notion (and given that space is almost flat), the path of a ray of light will be determined “fundamentally by minimum number of wave, and not by minimum time” (Larmor 1919a, 516). In this case, claims Larmor, there ought (on kinematical grounds) to be no such deflection of light passing the Sun.

However, Larmor then considers a dynamical explanation for the deflection test, drawing in Einstein’s work on the quantum theory of light. According to this description, the velocity of light ought to increase, and according to Einstein’s theory, in just the amount observed:

²⁶The section “On Gravitational Relativity,” from which this set of quotations was taken, was added by Larmor on 20 November 1919. All emphasis added by the author.

²⁷Jose Sanchez-Ron (1999) provides an investigation of Larmor’s gripes with general relativity.

Dr. Einstein requires in another connection that light should consist of discrete bundles or *quanta* of energy. Let it also be granted that inertia and gravitation are attributes of energy. It seems to follow that each of these bundles of energy will swing round the Sun in a hyperbolic orbit, and that its velocity will be *increased* when near the Sun. It is well known that this would account for half the observed deflection. But, again, physical optics could not exist without the idea of transverse waves and their phases, which must be grafted on somehow to the bundles of energy. (Larmor 1919a, 412)

Larmor can be seen to be clearly grappling with the puzzling “wave-particle” nature of light, and seizes upon the opportunity of applying this puzzle to general relativity to render it less certain. Given this conflict, he argues that the recent deflection test conducted by Eddington should be looked upon as a “guide rather than a verification” (Larmor 1919a, 412). Of course, Einstein’s own path would involve an engagement with just such issues. His approach was to consider the quantum, particulate aspects as merely an emergent phenomenon (as special solutions) from an underlying classical field theory. He had, moreover, already (by this stage) considered the possibility of a “quantum correction” to general relativity.

6.3.3 The Cavendish Lab’s Intervention: Radioactivity and Gravitation

An experimental venture into the interaction of the gravitational field with what were slowly becoming viewed as quantum properties of particles was conducted by Ernest Rutherford and Arthur Compton at the Cavendish Laboratory. It appeared in the same Christmas Day issue as Larmor’s article discussed above. Their paper constituted a response to an article by Prof. Donnan from the previous week’s issue over the behavior of radioactive substances in strong gravitational fields.²⁸ They note that pretty much the same question was put to them by “Dr. Schuster” some years earlier. The problem considered was whether the intensity of gravitational field strength could modify the rate of transformation of various radioactive substances. Before they had a chance to put Arthur Schuster’s question to the test,²⁹ the First World War intervened.

²⁸Frederick G. Donnan deduced a relation between “the variation of mass in a physical change of state or chemical reaction and the rate of variation with gravitational potential of the corresponding change of total internal energy” (Donnan 1919, 392). Note there was a general concern around this time with the source of stellar energy. It is thus fairly natural to consider the possible dependence of the rate of emission on the star’s gravitational field.

²⁹They got as far as a method of testing radioactive decay rates (over a period of up to 100 days), and were planning on sending their various radioactive substances to parts of the world with significant differences in their gravitational field strength.

Rutherford and Compton (1919) bypass the need to use a large heterogeneous mass to generate a suitable gravitational field by using the equivalence between gravitational acceleration and centrifugal acceleration. They therefore performed the experiment by placing radioactive substances at the edge of a rapidly spinning disc (generating 20,000 times the strength of the Earth's gravitational field) and measured γ -ray rate responses (using a balance method), looking for (significant) discrepancies. However, as they noted, on the basis of Donnan's calculations, no change in rate was to be expected since, if it existed, the effect would be "very much smaller than can be detected by measurements of this character" (Rutherford and Compton 1919, 412).³⁰

Note that their method followed Einstein's own suggestion, in his "popular account" of relativity, from 1916 (Einstein 1997). He considers a setup in which a clock is situated on a spinning disc, a certain distance γ from the center. The clock's velocity, relative to a frame K at rest with respect to the moving clock, is $v = \omega\gamma$ (where ω is the angular velocity). Where v_0 represents the "number of ticks of the clock per unit time [...] relative to K when the clock is at rest," (Einstein 1997, 388) the rate of the clock when it is moving relative to K with velocity v (but at rest relative to the disc) is:

$$v = v_0 \sqrt{1 - \frac{v^2}{c^2}} = v_0 \left(1 - \frac{1}{c^2} \frac{\omega^2 \gamma^2}{2}\right). \quad (6.2)$$

He then considers the difference of potential of the centrifugal force between the clock's position and the disc's center, written ϕ :

$$\phi = -\frac{\omega^2 \gamma^2}{2}, \quad (6.3)$$

which, on substitution into eq. (6.2) gives:

$$v = v_0 \left(1 + \frac{\phi}{2}\right). \quad (6.4)$$

From which one derives the time dilation as a result of the centrifugal acceleration. The equivalence principle leads one to the result that observers rotating with the disc will find themselves in a gravitational field with potential ϕ . This is then

³⁰Sanchez-Ron (1992, 68) claims that Rutherford and Compton "did not make any effort to see whether or not their experimental results agreed with the predictions of general relativity." However, the previous remark clearly states that, for processes of such microscopic nature, it would be a practical impossibility to compare the experimental results with the theory beyond the very broad fact that no result is expected to be seen on the basis of general relativity.

applied to an atom that is emitting spectral lines, which can be viewed by analogy with the clock. The expectation will then be that:

An atom absorbs or emits light of a frequency which is dependent on the potential of the gravitational field in which it is situated. (Einstein 1997, 389, italics in the original)

But after considering a centrifugal (acceleration) example, Einstein then switches (for obvious practical reasons) to consideration of an atom on the surface of a heavenly body, noting that its frequency will be a little less than the frequency of the same element on a smaller, less massive body. One could test this with spectral lines originating on the Sun and the Earth, respectively. Though there is no Planck's constant in this example, and so this is not by any means a quantum gravitational phenomenon, at root this is about something (spectral lines) that was central in discussions of the old quantum theory. In intervening in atomic frequencies, the gravitational field was surely intervening in quantum processes.

There were, then, some early experimental suggestions concerning the influence of gravity on elementary processes, but these quickly died out. It is possible that this was due to Eddington's theoretical calculations and these null results from the Cavendish Lab.³¹ However, it was, of course, already known that the gravitational effects on single atoms would be miniscule simply by inspecting the size of the gravitational constant. In this sense, the Cavendish Lab's results merely confirmed what was already believed.³²

Before I leave this section, I note that it seems that the debate discussed here was in many ways a direct continuation of an earlier one on the relationship between gravitation and temperature following experiment work by Philip E. Shaw (1916). This work stretches back to a period before general relativity was established and that remained largely independent of general relativity even when it did become better established. Shaw had conducted experiments in 1915, with a Cavendish torsion balance. These pointed to a positive temperature coefficient for the gravitational constant. Shaw's theoretical position was roundly criticized,

³¹ However, the same question was tackled in 1942, with the benefit of new theoretical knowledge and improved experimental techniques (then able to produce centrifugal fields of $1000000g$), by Freed, Jaffey and Schultz (1942)—they explicitly cite the Cavendish results in their work. Even at these centrifugal field strengths, no effect was seen that could be distinguished from experimental error. For a discussion on how this early experimental research developed into the modern era of quantum gravity research, see (Gillies and Unnikrishnan 2002a; 2002b), especially p.127 of the latter.

³² Curiously, John Joly returned to the general issue of the connection between rates of radioactive decay and the principle of relativity posing the question of whether radioactive clocks might offer an invariant way to measure absolute time, or whether "radio-activity [is] also 'in the conspiracy'" (Joly 1920, 468). Clearly the centrifugal experiments were not sufficient to determine the answer one way or the other. Of course, it was only fairly recently that the question was answered and the effect of gravitational time dilation on atomic clocks was confirmed (Hafele and Keating 1972).

not least by Oliver Lodge (1916; 1917), who pointed to several problems with momentum non-conservation and potential empirical inadequacies. However, Shaw's experimental work was positively received, and it was this that filtered through into the later work on the relationship between radioactivity and gravitation.

6.3.4 Einstein on the Development of Relativity

In 1921, following an outline of how he arrived at the form of general relativity, Einstein concluded his account with a list of:

[...] important questions which are awaiting solution [...]. Are electrical and gravitational fields really so different in character that there is no formal unit to which they can be reduced? Do gravitational fields play a part in the constitution of matter, and is the continuum within the atomic nucleus to be regarded as appreciably non-Euclidean? (Einstein 1921, 784)

As I understand it, here Einstein is, firstly, hinting at a unified field theory, through which both gravitational and electrical forces are described. On the basis of this, the question is begged as to whether gravitational force plays any role in holding atoms together.³³ Finally, and most interesting, it is natural to consider what kind of gravitational field would exist in the interior of an atom—though Einstein thinks directly in terms of what space-time would look like inside atoms. This has a bearing on the other questions since, unless the space-time is appreciably curved, there will be no work for gravitation to do in the structure of matter. What can be reasonably inferred from this is that Einstein was considering the possibility that general relativity might have something to say about the constitution of matter, and a fortiori the nature of quantum theory. This is backed up by remarks that Einstein is reported to have made following a lecture at King's College London in 1921:

After the public lecture Prof. Einstein was the guest of the Principal [...]. In responding to his health, Prof. Einstein made an interesting revelation of his attitude to the quantum theory. This theory was, he said, presenting a difficult problem to physics, but the very nature of the difficulty served to bring into relief the attractiveness and satisfaction of the principle of relativity. That principle had served to give a simple and complete explanation of experimental facts which under

³³Recall that at this time electromagnetism alone was thought to be responsible for the structure of matter.

any other aspect were discordant. In the quantum theory as it stood at present we were faced with discordant experimental facts, and were searching for the principle on which to interpret them. (Anonymous 1921, 504)

One inference to make here is that Einstein expected that the general theory of relativity itself might be able to supply such a missing principle on which to found a satisfactory theory of matter, quantum or not. However, Sauer, in a private email to the author on 15 September 2011, interpreted the passage as more likely indicating that Einstein had in mind, not that the principle of relativity would itself serve as a *guide* for quantum theory, but rather that something *analogous* to the principle of relativity, though of a different sort, more relevant for quantum theory (such as the correspondence principle or the adiabatic principle), might be required to interpret the experimental data then available. However, given that, at this time, Einstein considered the *direct* role of gravitation in the constitution of matter one of his most important questions in need of resolution, it is perfectly possible that he intended what I suggest.

6.3.5 Jeans on Indeterminism in GR and QM

In 1926, James Jeans presented a curious argument suggesting that the “unpalatable determinism” brought forth by relativity (in the form of the “block universe”) might be somehow *cured* by developments in quantum theory. To modern ears, this might sound the wrong way around, since it is the indeterminism of quantum theory that is unpalatable to many.

It is clear that Jeans has in mind *fatalism* here, since he writes that “Einstein’s work on relativity changed the universe from a drama into a picture drama” (Jeans 1926, 311). Clearly he supposed that the random nature of atomic processes could inject some much-needed randomness into all processes:

[R]elativity is not the whole of natural science; it is not even the whole of Einstein’s work. His contributions to science fall into two columns which, unhappily, are parallel and show no signs of meeting. The first column contains his contributions to the theory of relativity, [...] the second column contains his contributions to the theory of quanta [...]. It is not yet altogether clear which of these columns will figure most prominently in the history of present-day science when this is finally written in its proper perspective. But it already seems possible that the second column of Einstein’s work may contain the needed antidote to the determinism and automatism to which the first column, if it stood by itself, would seem to condemn us. (Jeans 1926, 311)

Jeans is clearly well aware of the *problem* of bringing together quantum and gravity—that is, of the task of saying something about the “meeting” between the “two columns”—however, the probabilistic nature of quantum theory was still not fully appreciated by the wider scientific community,³⁴ nor were the dynamics of general relativity (*qua* theory describing the evolution of geometrical data) understood. Indeed, Jeans apparently viewed the four-dimensional nature of Einstein’s theory as its core distinguishing feature. Or, as he puts it, “the dynamical explanation of a gravitational force crumbled in the hands of Einstein” (Jeans 1926, 311). However dated Jeans’s specific comments may strike us now, they mark a very clear expression of the problem of quantum gravity as a potential *conceptual* (rather than “merely technical”) clash. It is also an early example of a proposal to *utilize* one of the ingredient theories of quantum gravity to resolve some supposed problem with the other. In this case, it was the conceptual problem of the block-like nature of the universe according to general relativity, and Jeans proposed that the theory of quanta might offer some assistance in changing the worldview “back into a drama.”

6.3.6 Klein on Five-dimensional Quantum Relativity

Oskar Klein came up with the idea of a five-dimensional approach while visiting Ann Arbor as an instructor in theoretical physics at the University of Michigan. Klein began working on the approach in 1924, then returned to this initial foray when he returned to Copenhagen in 1925. He published a paper in *Nature* a little after the more well-known paper from *Zeitschrift für Physik*, though both appeared in 1926. It was Pauli who, early in 1926, informed Klein that Kaluza had already published on a similar idea (Pais 2000, 131). However, there is genuine novelty in Klein’s approach in that Planck’s constant emerges as a consequence of topological structure. Or, as Klein puts it, his result “suggests that the origin of Planck’s quantum may be sought just in this periodicity in the fifth dimension” (Klein 1926b, 516).

This was a new development of the much older idea that geometry and topology could be used as a “resource” in world-building. In fact, in his later recollections of how he came to the five-dimensional idea, Klein describes an approach broadly similar to that envisaged by Larmor and, earlier, Eddington (as described above). He notes how he was searching for “a wave background to the quantization rules” and had been playing with:

³⁴Jeans’s examples are based on radiation, involving the disintegration of uranium and Einstein’s own work on the emission and absorption of light quanta (described by Jeans as “the statistics of their jumping about”); with such phenomena, says Jeans, “we seem to be beyond the domain of [...] natural laws” (Jeans 1926, 311).

[T]he idea that waves representing the motion of a free particle had to be propagated with a constant velocity, in analogy to light waves—but in a space of four dimensions—so that the motion we observe is a projection on our ordinary three-dimensional space of what is really taking place in four-dimensional space. (Klein 1991, 108–109)³⁵

Klein was in discussions with Paul Ehrenfest and George Eugene Uhlenbeck in the spring of 1926, during a visit to Leiden at Ehrenfest's invitation, and the discussions were enough to lead to a paper on five-dimensional relativity by Ehrenfest and Uhlenbeck (1926), appearing around the same time as Klein's own note in *Nature* (Ford 2009, 9–10).³⁶ Ehrenfest had a long fascination with the concept of dimensionality. Long before Klein began thinking about his dimensional expansion, Ehrenfest (1917) had written on the possible reasons why space is three-dimensional, showing how various processes and the stability of orbits depend on it. This might well have been behind Klein's own suggestion, given that physical quantities would be periodic functions of the compact dimension and observables would be given as averages over the small circumference, that ordinary space must be three-dimensional (Klein 1991, 110).

As with several other approaches mentioned in this paper, Klein's work on five-dimensional relativity might be seen to fall somewhat outside of the category of quantum gravity. However, as with many of the other approaches discussed, the influence of the work on later quantum gravity research cannot be underestimated. Further, it shows how, in some sense, the shape of space (a feature dynamically determined within general relativity) can determine what would be otherwise inexplicable features of the world (in this case the existence of a quantum of action). Of course, it was already known following general relativity that geometry offers a potentially exceptional explanatory resource, but Klein's work showed that this resource was more widely applicable than previously supposed.

As mentioned above, Klein's approach was closely related to the earlier efforts of Kaluza (1997), and Klein begins by outlining Kaluza's approach. Kaluza

³⁵Klein claims that Bohr had earlier made similar suggestions (Klein 1991, 109). However, he recalls his later discussions with Bohr and Heisenberg as being received with "kind skepticism" (Klein 1991, 112).

³⁶Klein (1991, 112) notes that Ehrenfest had asked Lorentz to invite Klein (on a Lorentz Fellowship) after having read a copy of Klein's paper that was given to him by Llewellyn Hilleth Thomas (himself passing through Leiden on his way from Copenhagen to Cambridge). Uhlenbeck refers to an unpublished paper between himself, Ehrenfest, and Klein (see interview of Uhlenbeck by Thomas S. Kuhn on 9 December 1963, Niels Bohr Library & Archives). Uhlenbeck recalls that at the time he believed Klein had something like a theory of everything: "It seemed then that one was very close to a world formula—one equation containing everything, you see. I remember that I had the feeling that 'Golly, we now perhaps know everything'" (Klein 1991, 112)—though he notes that the same was not true of Ehrenfest.

himself was inspired by Weyl's earlier modification of the metric of general relativity, so as to have a total metric that could account for (what was then) "all physical phenomena" (Kaluza 1997, 53). The introduction of a space-time fifth dimension was necessary in such a theory since in four dimensions the only Christoffel symbols that are available are those of the gravitational field. Kaluza imposed a "cylinder condition," effectively eliminating variations with respect to the fifth dimension by regarding x^0 derivatives as having very small or zero magnitudes.

Though Kaluza's approach was purely classical, he does conclude his 1921 article with a consideration of its *microscopic* significance:

[M]atter, in its fundamental constituents at least, is not weakly charged; in the words of H. Weyl its 'macroscopic placidity' stands in sharp contrast to its 'microscopic turbulence', and this is true in particular for the new coordinate x^0 : for the electron or H-nucleus the quantity $\frac{\rho_0}{\mu_0}$ [the ratio of charge density to rest-mass density, or the "specific charge" of matter] and with it the "velocity"-component is anything but small! In the form demanded by Approximation II [very small specific charge] the theory can describe at most macroscopic phenomena and the key question is whether it can be used for the above elementary particles.

If one tries to describe the motion of electrons by geodesics in R_5 one encounters immediately a difficulty that threatens to destroy the whole structure. The problem is that, if one takes the earlier assumptions literally, the fact that $\frac{e}{m} = 1.77 \times 10^7$ (in lightseconds) means that the quantity u^0 is so large that the last term in $[\bar{v}^\lambda = \frac{dv^\lambda}{d\sigma} = \Gamma_{\rho\sigma}^\lambda v^\rho v^\sigma + 2\alpha F_\kappa^\lambda u^0 v^\kappa - \mathfrak{h}_{,\lambda}(u^0)^2]$ ³⁷, instead of disappearing, takes a value much greater than is observed experimentally and becomes the leading term. [...] [I]t would seem to be impossible to proceed in the old manner without some new hypotheses. (Kaluza 1997, 57)

Kaluza's solution to the problem was to throw out the gravitational constant "so that gravitation would appear as a sort of difference-effect" (Kaluza 1997, 57). This, Kaluza argues, would have as an "attractive feature," the fact that a "statistical role" could be attributed to the gravitational constant. He finishes by remarking that "for the moment the consequences of this hypothesis cannot be foreseen;

³⁷The \mathfrak{h} term here is Kaluza's expression for the g_{00} th component (i.e., the "corner potential") of the metric tensor; $d\sigma$ is a five-dimensional Riemannian line-element (given by $d\sigma = \sqrt{\sum \gamma_{ik} dx^i dx^k}$); $\alpha = \sqrt{\frac{\kappa}{2}}$, with κ the gravitational constant.

and of course there are other possibilities to consider. And threatening all universal hypotheses is the Sphinx of modern physics, the quantum theory” (Kaluza 1997, 58).

Klein focuses directly on Kaluza’s “Sphinx”, on the microscopic description. He also diverges from Kaluza in assigning a definite scale to his x^0 and treating it realistically. The approach involves the establishment of a link between Kaluza’s unified theory and the (then) brand new work on quantum mechanics of Louis de Broglie and Erwin Schrödinger. He characterizes Kaluza’s approach as a unified theory in which the unification is achieved via the coefficients γ_{ik} of the five-dimensional Riemannian line-element mentioned in footnote 37. Klein shows how one can view the equations of motion for charged particles propagating in an electromagnetic field (constructed by Kaluza) as radiation equations (that is, according to which matter is a wave phenomenon). When this viewpoint is adopted, a generalization of the wave equation follows. Restricting to a class of solutions in which the fifth dimension has a period related to Planck’s constant, then, Klein argues, the standard quantum mechanical laws drop out. Hence, one has a unified theory of electricity and magnetism, *and* one has an elementary notion of quantum theory that appears *as a consequence* of the theory. In a sense we find in Klein’s approach geometry being used as a *resource* in the construction, deduction, or explanation of other puzzling phenomena.

6.4 On the Way to Quantum Geometry

The relationship of gravity (and indeed general relativity) to the phenomena revealed by quantum theory was used strategically by those who opposed relativity around the 1920s. Lodge, for example, thought that “if posterity is forced to accept and employ devices [...] for dispensing with the ether I fear that a damaging blow will have been dealt at physics” (Lodge 1919, 62).

However, Lodge signaled an early warning for those who might wish to link the discontinuity of matter with space-time:

May I parenthetically urge philosophers to be on their guard against any system which introduces discontinuity into space or time, or even energy? Matter is discontinuous, electricity is discontinuous, I venture to say that real number is discontinuous; but space and time and ether are continuous. Energy may acquire a discontinuous aspect in its relation with matter, and the quantum is an important metrical fact, but it is explicable in terms of the atom or electron, and is not a feature in energy itself. Time is absolutely continuous, however it be measured and expressed numerically. (Lodge 1919, 62)

Developments from as early as the 1930s—that have persisted to the present day, packaged in the concept of “quantum geometry”—would follow just the path attacked by Lodge, himself too strongly committed to classical physics and the ether theory to budge.

An even earlier statement of potential short-distance gravitational distortions was given in an editorial of 1919:

If the distortion of space were very great, the customary methods of dynamics might lose their significance; and the question arises: Will, on Einstein’s theory, the space inside an atom be so far from Euclidean that ordinary dynamical methods are unjustifiable? The answer to this question is “No.” There are two lengths which have special significance in connection with the atom; one of these is what we call the radius of the atom, and is of the order 10^{-8} cm; the other we call the radius of the electron, and is about 10^{-13} cm. Even at the smaller of these distances the gravitational potential due to the mass of the atom, and therefore the distortion from Euclidean space, would be exceedingly small compared with the corresponding quantities due to Earth at its surface, so that there is no special distortion inside the atom, except at distances from the centre which are infinitesimal even when compared with the radius of an electron. (Lodge 1921a, 354)

Not unrelated is the suggestion made by Norman Campbell in 1921, arguing that better sense could be made of the interior of atoms if the distinction between stationary and moving electrons were abolished by arguing that time ceased to make sense in the interior of atoms:³⁸

The suggestion that I made is that, by means of a generalized principle of correspondence, the distinction between moving and fixed electrons might be abolished and the conceptions that have proved so fruitful in explaining spectra made available immediately for explaining also such things (if there are such things) as are only explicable by fixed electrons. Thus the distinction would be abolished if ‘time’ had no meaning inside the atom. For the difference between electrons following an orbit and electrons fixed at points on that orbit can only be expressed in terms of temporal conceptions; if all such conceptions are totally invalid in dealing with problems of atomic structure the distinction vanishes. (Campbell 1921, 170)

³⁸Hence, this is far more radical than Einstein intimated in his question over whether the interior of atoms is non-Euclidean.

One wonders whether such suggestions could have been conceptualized in this way were it not for the parallel debates in the context of general relativity.

The gravitational field of the electron (though classically conceived) received an interesting early speculative treatment by George Jeffrey (1921). As with many other papers already discussed, Lodge (1921b) thought fit to comment on this paper of Jeffrey's in *Nature*. The aim of Jeffrey's paper was to show how the gravitational field might be involved in the structure of the electron, with the conclusion that the electrical and mass potentials offer some kind of stabilizing effect by opposing each other. Again, Lodge translates the "ether-free" discussion into one concerning the state of the ether close to an electron.

The introduction by Uhlenbeck and Samuel Goudsmit (1926) of the hypothesis of quantized angular momentum of electrons (to explain several puzzling results in quantum theory and experiment) radically altered the landscape, both for quantum theory and for the kinds of models needed in general relativity. Eddington (1926) discusses a potential conflict between the spinning electron hypothesis and relativity theory. He notes that some have been perplexed by what seems like a straightforward conflict between relativity's prohibition of superluminal velocities and the fact that the electron's periphery apparently moves at just such velocities. Eddington dissolves the perplexity in two ways: firstly, the prohibition applies to the propagation of *signals*, but clearly no such signaling is possible by utilizing the electron's angular velocity. Secondly, the spin is a quantum number: it represents, as Eddington says, "a state of the world" (Eddington 1926, 652). Finally, he notes that the idea that the electron has a space-like (superluminal) J^μ vector was already postulated by Weyl in connection with his investigation into the relationship between gravitational and electrical fields, and was deduced purely from his action-principle. Again, this clearly points to the fact that the domains of the large and small, gravitational and atomic, were not seen to be disconnected. The view that the world of the quantum and of gravity is a schizophrenic one came with later (failed) attempts to directly *quantize* general relativity.³⁹

6.5 Conclusion

In this paper I have examined the very earliest work on the problem of quantum gravity (understood very liberally). There was a very lively debate in this early stage, and no suggestion that such a theory would not be forthcoming. Indeed, there are, rather, many suggestions explicitly advocating that an integration of quantum theory and general relativity (or gravitation, at least) is *essential* for future physics, to construct a satisfactory foundation. I have also demonstrated how

³⁹See e.g., (Ashtekar 2005).

this belief was guided by a diverse family of underlying agendas and constraints, often of a highly philosophical nature. A subsequent paper will trace the fate of these agendas as quantum theory was put on a firmer footing.

Abbreviations and Archives

Niels Bohr Library & Archives	American Institute of Physics, College Park, MD, USA, www.aip.org/history/ohilist/4922_5.html
----------------------------------	---

Acknowledgements

I would like to thank Shaul Katzir and Tilman Sauer for their helpful comments on an earlier draft. I would also like to thank the Australian Research Council for their support via the Discovery Project Grant DP0984930.

References

- Anonymous (1921). Prof. Einstein's Lectures at King's College London, and the University of Manchester. *Nature* 107(2694): 504.
- Ashtekar, Abhay (2005). The Winding Road to Quantum Gravity. *Current Science* 89: 2064–2075.
- Buchwald, Diana, Tilman Sauer, Ze'ev Rosenkranz, Josef Imes Illy, and Virginia Iris Holmes (2006). *The Collected Papers of Albert Einstein. Volume 10. The Berlin Years*. Princeton: Princeton University Press.
- Campbell, Norman (1921). Atomic Structure. *Nature* 107(2684): 170.
- Corry, Leo (1997). David Hilbert and the Axiomatization of Physics (1894–1905). *Archive for History of Exact Sciences* 51: 83–198.
- (1999). From Mie's Electromagnetic Theory of Matter to Hilbert's Unified Foundations of Physics. *Studies in History and Philosophy of Modern Physics* 30(2): 159–183.
- DeWitt, Bryce S. (1967). Quantum Theory of Gravity I. The Canonical Theory. *Physical Review* 160(5): 1113–1148.
- Donnan, Frederick G. (1919). Heat of Reaction and Gravitational Field. *Nature* 104(2616): 392–393.
- Durham, Ian T. (2003). Eddington and Uncertainty. *Physics in Perspective* 5: 398–418.

- Earman, John and Clark Glymour (1978). Einstein and Hilbert: Two Months in the History of General Relativity. *Archive for History of Exact Sciences* 19: 291–308.
- Eddington, Arthur S. (1916). Gravitation and the Principle of Relativity I. *Nature* 98(2461): 328–330.
- (1918). Gravitation and the Principle of Relativity II. *Nature* 101(2524): 34–36.
- (1926). Spinning Electrons. *Nature* 117(2949): 652.
- (1933). Physics and Philosophy. *Philosophy* 8: 30–43.
- (1949). *Fundamental Theory*. Cambridge: Cambridge University Press.
- Ehrenfest, Paul (1917). In What Way Does It Become Manifest in the Fundamental Laws of Physics that Space Has Three Dimensions? *Proceedings of the Amsterdam Academy* 20: 200–209.
- Ehrenfest, Paul and George Uhlenbeck (1926). Graphische Veranschaulichung der De Broglieschen Phasenwellen in der fünfdimensionalen Welt von O. Klein. *Zeitschrift für Physik* 37: 895–906.
- Einstein, Albert (1916a). Strahlungs-Emission und Absorption nach der Quantentheorie. *Verhandlungen der Deutschen Physikalischen Gesellschaft* 18: 318–323.
- (1916b). Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin* XXXII: 688–696.
- (1918). Über Gravitationswellen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*: 154–167.
- (1921). A Brief Outline of the Development of the Theory of Relativity. *Nature* 106(2677): 782–784.
- (1997). On the Special and General Theory of Relativity (A Popular Account). In: *The Collected Papers of Albert Einstein. Volume 6. The Berlin Years: Writings, 1914–1917*. Ed. Anne J. Kox, Martin J. Klein, and Robert Schulmann. Princeton: Princeton University Press. 247–420. Translation of 1916 edition. [Doc. 42: authorized translation by Alfred Engel, 1997].
- (2002). Ether and the Theory of Relativity. In: *The Collected Papers of Albert Einstein. Volume 7. The Berlin Years: Writings, 1918–1921*. Ed. Michel Janssen, Robert Schulmann, József Illy, Christoph Lehner, and Diana Kormos Buchwald. Princeton: Princeton University Press. 305–323. Translation of 1920 edition. [Doc. 38: authorized translation by Alfred Engel, 2002].
- Ford, George (2009). *George Eugene Uhlenbeck, 1900–1988: A Biographical Memoir*. Washington, D.C.: National Academy of Sciences.
- Freed, Simon, Arthur H. Jaffey, and Melvin L. Schultz (1942). High Centrifugal Field and Radioactive Decay. *Physical Review* 63: 12–17.

- Galison, Peter L., Michael Gordin, and David Kaiser, eds. (2001). *Quantum Mechanics: History of Modern Physical Science*. 4. New York: Routledge.
- Gillies, George T. and C. S. Unnikrishnan (2002a). Experimental Perspectives on the Interplay of Quantum and Gravity Physics. In: *Advances in the Interplay between Quantum and Gravity Physics*. Ed. Peter G. Bergmann and Venzo de Sabbata. Dordrecht: Kluwer. 103–122.
- (2002b). Quantum Physics-Motivated Measurements and Interpretation of Newtonian Gravitational Constant. In: *Advances in the Interplay between Quantum and Gravity Physics*. Ed. Peter G. Bergmann and Venzo de Sabbata. Dordrecht: Springer. 123–132.
- Goldstein, Catherine and Jim Ritter (2003). The Varieties of Unity: Sounding Unified Theories 1920–1930. In: *Revisiting the Foundations of Relativistic Physics*. Ed. Abhay Ashtekar, Robert S. Cohen, Don Howard, Jürgen Renn, Sahotra Sarkar, and Abner Shimony. Dordrecht: Springer. 93–149.
- Gorelik, Gennady (1992). The First Steps of Quantum Gravity and the Planck Values. In: *Studies in the History of General Relativity*. Ed. Jean Eisenstaedt and Anne J. Kox. Boston: Birkhäuser. 364–379.
- Gupta, Suraj (1962). Quantum Theory of Gravitation. In: *Recent Developments in General Relativity*. Ed. Leopold Infeld. Oxford: Pergamon Press. 251–258.
- Haas, Arthur (1919). Die Axiomatik der Modernen Physik. *Naturwissenschaften* 7(41): 744–750.
- (1938). The Dimensionless Constants of Physics. *Proceedings of the National Academy of Sciences of the United States of America* 24(7): 274–276.
- Hafele, Joseph C. and Richard E. Keating (1972). Around-the-World Atomic Clocks: Predicted Relativistic Time Gains. *Science* 177(4044): 166–168.
- Halpern, Paul (2004). Nördstrom, Ehrenfest, and the Role of Dimensionality in Physics. *Physics in Perspective* 6: 390–400.
- Hilbert, David (1915). The Foundations of Physics (First Communication). In: *The Genesis of General Relativity, Volume 4. Gravitation in the Twilight of Classical Physics: The Promise of Mathematics*. Ed. Jürgen Renn. Dordrecht: Springer 2007. 1003–1015.
- (1918). Axiomatisches Denken. *Mathematische Annalen* 79: 405–415.
- Hu, Bei-Lok (2009). Emergent/ Quantum Gravity: Macro/ Micro Structures of Spacetime. *Journal of Physics: Conference Series* 174: 1–16.
- Jeans, James (1926). Space, Time, and Universe. *Nature* 2939(117): 308–311.
- Jeffrey, George B. (1921). The Field of an Electron on Einstein's Theory of Gravitation. *Proceedings of the Royal Society A* 99(697): 123–134.
- Joly, John (1920). Relativity and Radio-Activity. *Nature* 104(2619): 468.

- Kaluza, Theodore (1997). Zum Unitätsproblem der Physik. In: *The Dawning of Gauge Theory*. Ed. Lochlainn O’Raifeartaigh. Princeton: Princeton University Press. 53–58. Reprint of 1921 edition.
- Kilmister, Clive (1994). *Eddington’s Search for a Fundamental Theory*. Cambridge: Cambridge University Press.
- Klein, Oskar (1926a). Quantentheorie und fünfdimensionale Relativitätstheorie. *Zeitschrift für Physik A* 37(12): 895–906.
- (1926b). The Atomicity of Electricity as a Quantum Theory Law. *Nature* 118: 516.
- (1991). From My Life of Physics. In: *The Oskar Klein Memorial Lectures*. Ed. Gösta Ekspång. Singapore: World Scientific. 103–117.
- Kragh, Helge (2000). Relativity and Quantum Theory from Sommerfeld to Dirac. *Annalen der Physik (Leipzig)* 9(11–12): 961–974.
- Larmor, Joseph (1919a). Gravitation and Light. *Nature* 104(2617): 412.
- (1919b). On Generalized Relativity in Connection with Mr. W. J. Johnston’s Symbolic Calculus. *Proceedings of the Royal Society A* 96(678): 334–363.
- Lodge, Oliver (1916). Gravitation and Temperature. *Nature* 97(2433): 321.
- (1917). Gravitation and Thermodynamics. *Nature* 99(2475): 104.
- (1919). Time, Space, and Material, II. Symposium: Time, Space, and Material: Are They, and If so in What Sense, the Ultimate Data of Science? *Proceedings of the Aristotelian Society, Supplementary Volumes* 2: 58–66.
- (1921a). Gravitation and Light. *Nature* 104(2614): 354.
- (1921b). The Gravitational Field of an Electron. *Nature* 107(2691): 392.
- Mie, Gustav (2007). Foundations of a Theory of Matter (Excerpts). In: *The Genesis of General Relativity*. Ed. Michel Janssen, John D. Norton, Jürgen Renn, Tilman Sauer, and John Stachel. 4. Boston Studies in the Philosophy of Science, Volume 250. Boston: Springer. 1554–1619. Reprint of source text 1912–1913.
- Nordström, Gunnar (1914). Über die Möglichkeit, das elektromagnetische Feld und das Gravitationsfeld zu vereinigen. *Physikalische Zeitschrift* 15: 504–506.
- Pais, Abraham (2000). *The Genius of Science*. Oxford: Oxford University Press.
- Pauli, Wolfgang (1921). Relativitätstheorie. In: *Enzyklopädie der Mathematischen Wissenschaften*. V. Berlin: B. G. Teubner. 539–775.
- (1958). *Theory of Relativity*. New York: Pergamon. English translation of 1921 edition.
- Planck, Max (1899). Über irreversible Strahlungsvorgänge. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin* 5: 440–480.

- Rechenberg, Helmut (1995). Quanta and Quantum Mechanics. In: *Twentieth Century Physics*. Ed. Laurie M. Brown, Abraham Pais, and Brian Pippard. I. New York: Institute of Physics Publishing. 143–248.
- Renn, Jürgen, ed. (2007). *The Genesis of General Relativity*. 4 Volumes. Dordrecht: Springer.
- Rovelli, Carlo (2000). The Century of the Incomplete Revolution: Searching for General Relativistic Quantum Field Theory. *Journal of Mathematical Physics* 4(6): 3776–8801.
- (2002). Notes for a Brief History of Quantum Gravity. In: *The Ninth Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories*. Ed. Vahe G. Gurzadyan, Robert T. Jantzen, and Remo Ruffini. Proceedings of the MGIXMM Meeting. Held 2–8 July 2000 in The University of Rome "La Sapienza," Italy. Singapore: World Scientific. 742–768.
- Rowe, David E. (2003). Hermann Weyl, the Reluctant Revolutionary. *Mathematical Intelligencer* 25(1): 61–70.
- Rutherford, Ernest F. and Arthur Compton (1919). Radio-Activity and Gravitation. *Nature* 104(2617): 412.
- Sanchez-Ron, Jose M. (1992). General Relativity Among the British. In: *Studies in the History of General Relativity*. Ed. Jean Eisenstaedt and Anne J. Kox. Boston: Birkhäuser. 57–88.
- (1999). Larmor versus General Relativity. In: *The Expanding Worlds of General Relativity*. Ed. Hubert Goenner, Jürgen Renn, Jim Ritter, and Tilman Sauer. Boston: Birkhäuser. 405–430.
- Sauer, Tilman (2002). Hope and Disappointments in Hilbert's Axiomatic 'Foundations of Physics'. In: *History of Philosophy of Science*. Ed. Michael Heidelberger and Friedrich Stadler. Dordrecht: Kluwer. 225–237.
- Scholz, Erhard (2001). *Hermann Weyl's Raum-Zeit-Materie and a General Introduction to His Scientific Work*. Basel: Birkhäuser.
- Schweber, Silvan S. (2009). Weimar Physics: Sommerfeld's Seminar and the Causality Principle. *Physics in Perspective* 11: 261–301.
- Shaw, Philip E. (1916). The Newtonian Constant of Gravitation as Affected by Temperature. *Philosophical Transactions of the Royal Society of London, Series A* 216: 349–392.
- Sommerfeld, Arnold (2000). *Wissenschaftlicher Briefwechsel. Band 1. 1892–1918*. Ed. Michael Eckert and Karl Märker. Berlin: Deutsches Museum, Verlag für Geschichte der Naturwissenschaften und der Technik.
- Stachel, John (1998). The Early History of Quantum Gravity (1916–1940). In: *Black Holes, Gravitational Radiation and the Universe*. Ed. Bala R. Iyer and Biplab Bhawal. Dordrecht: Kluwer Academic Publishers. 525–534.

- Uhlenbeck, George and Samuel Goudsmit (1926). Spinning Electrons and the Structure of Spectra. *Nature* 117(2946): 264–265.
- Weyl, Hermann (1921). *Raum-Zeit-Materie*. Berlin: Julius Springer.
- Wiener, Norbert (1956). *I Am a Mathematician: The Later Life of a Prodigy*. Cambridge, MA: MIT Press.
- Wuensch, Daniela (2003). The Fifth Dimension: Theodor Kaluza's Ground-Breaking Idea. *Annalen der Physik (Leipzig)* 12(9): 519–542.

Extending the Framework of Quantum Physics

Chapter 7

Superposing Dynamos and Electrons: Electrical Engineering and Quantum Physics in the Case of Nishina Yoshio

Kenji Ito

Research in quantum physics began in Europe and then spread to many parts of the world. Outside Europe and North America, Japan was one of the places where quantum physics research successfully took off before World War II. Although Japan started absorbing European scientific knowledge in the 18th century,¹ modernized educational and research institutions for the sciences appeared only in the late 19th century, after the so-called Meiji Restoration. Nevertheless, in the late 1920s, Japanese physicists began producing first-rate theoretical research,² and in 1935, Yukawa Hideki published his Nobel Prize-winning work on meson theory (Yukawa 1935).

How could this happen? Answering this would be of some importance for understanding the spread of quantum physics worldwide, or more generally for understanding the global dissemination of modern science in the 20th century. Unless we believe in the teleologically driven dissemination of science, it requires an explanation why other, vastly different cultures came to adopt European science. Instead of giving a full answer to the question of dissemination of quantum physics, which I will attempt in a separate and much larger work, in this paper I will show a possible link between preexisting conditions in Japan and the introduction of quantum physics. The main goal is to identify one possible aspect of these conditions that eased the introduction of quantum physical research into Japan: electrical engineering.

Nishina Yoshio³ (1890–1951) was among the first generation of Japanese quantum physicists who produced noteworthy theoretical research in quantum physics. He established himself as a physicist while staying in Europe, in particular Copenhagen, and in collaboration with Oskar Klein, carried out signifi-

¹ The earliest known book on Newtonian physics in Japan is Shizuki Tadao's liberal translation in 1798 of John Keill's *Introductiones ad Veram Physicam et Veram Astronomiam* (1725).

² One example, other than Nishina Yoshio's work with Oskar Klein described below, is Sugiura Yoshikatsu's work on the Heitler-London method, see (Sugiura 1927).

³ Following the common academic convention, I write Japanese names in the traditional order, the family name first, the given name second.

cant theoretical work on quantum physics, resulting in the Klein-Nishina formula (Klein and Y. Nishina 1929).⁴ After his return to Japan in late 1928, he introduced quantum mechanics there. He became a leading figure in this field in Japan, paving the way for other Japanese physicists.⁵

Curiously, Nishina was originally trained as an electrical engineer, not a physicist. When he entered the Sixth Higher School in Okayama at the age of 20 in 1910, he chose engineering as his major.⁶ Later he advanced to the Department of Electrical Engineering at Tokyo Imperial University in 1914. It was only after his graduation and employment at the Institute for Physical and Chemical Research (RIKEN) that he changed his career to physics. RIKEN allowed him to study abroad, from 1921 to December 1928. His stay in Europe during this period definitively made Nishina a quantum physicist.

Becoming a quantum physicist from Japan must have been difficult enough. How could Nishina move across, not only the cultural boundary between Japan and Europe, but also the disciplinary boundary between electrical engineering and physics? In an attempt to develop a partial answer to this question, I ask: Did the electrical engineering training that Nishina received prepare him for research in quantum mechanics?

A late-comer to the industrialized world, Japanese society prioritized practical subjects such as electrical engineering. These training fields may have provided intellectual resources and institutional bases that helped motivate, legitimize and sustain quantum physical research. The goal of this paper is to explore such possible links between electrical engineering training and theoretical practices of quantum mechanics in the local Japanese context to understand how new theoretical scientific practices traveled across cultures.

The role of an engineering background in the development of physics practices has been studied in several cases, including Albert Einstein, Henri Poincaré and Julian Schwinger (Pyenson 1982; 1985; Galison 1997; 2003). It is not within the scope of this paper to fully study physicists who were involved in engineering research during their training, as a hobby, or in the context of commissioned military research. However, given the topic of this paper, an obvious example for such an investigation is Paul Dirac, who received electrical engineering training in Bristol. In studying Dirac's engineering training and his wartime engineering work on isotope separation, Richard Dalitz stresses Dirac's engineering and practical sides and discusses how these could have been integrated into his atti-

⁴For the history of the Compton effect, see (Stuewer 1975; Brown 2002). For the historical background and significance of the Klein-Nishina formula in relation to Dirac's relativistic quantum mechanics, see, for example (Kragh 1990).

⁵For a biographical account of Nishina, see (Ito 2002; Kim 2007).

⁶"Higher School" *kōtō gakkō* in pre-World War II Japan was a liberal arts institution of higher education, and is not the same as a high school in the United States, for example.

tude to physics (Dalitz 1990). Peter Galison (2000) shows how Dirac expressed his ideas in drawings, using a form of projective geometry, which was part of his engineering training, and how these drawings were suppressed in his published papers.

These studies indicate that the relationship between engineering and physics was not a simple deterministic one. In a similar vein, this paper is not an attempt to provide a causal explanation of how socio-cultural or intellectual contexts shaped or influenced the way quantum mechanics was introduced into Japan and how Japanese physicists came to practice quantum physics in a different way. Nor do I claim that Nishina turned to quantum mechanics or carried out specific theoretical research (such as his collaboration with Klein on the Klein-Nishina formula) *because of* his electrical engineering background. In particular, I need to emphasize that it is not my intention to show that Nishina did theoretical physics *differently* from other physicists because of his earlier training in electrical engineering. Since the main focus of this paper is work that Nishina and Klein coauthored, such an analysis is out of the question anyway. My goal here is different and more modest. While one could point out various characteristics shared by quantum mechanics and electrical engineering in general terms, this paper aims to point out particular characteristics relevant to the case of Nishina's engineering training and his specific research in quantum mechanics, and to locate them in the historical context surrounding him. Rather than showing differences, I seek to find out how Nishina came to do things similar to the work of European physicists.

I pay particular attention to details of the mathematical practices in engineering to which Nishina was exposed and accustomed. I explore technical details of electrical engineering and physics to indicate some resemblances between the theoretical practices of the engineering and the quantum physics to which Nishina was exposed and upon which he drew.

7.1 Engineering in Japan

Scrutinizing socio-cultural contexts of physics in Japan is not the purpose of this paper, but a brief description might be helpful. Engineering had higher priority and more prestige than physics in early 20th century Japan. In the process of modernization after the Meiji Restoration, Japan included engineering in its higher education from the very beginning, motivated by the need for engineers to build modern infrastructure, such as telegraphs and railroads. Science was mostly, if not entirely, meant to be a basis for engineering and was taught in schools mainly to prepare engineers (Hirosige 1973; Bartholomew 1989). Hence, it was natu-

ral that atomic physics was often identified as a basis of electrical engineering. Physics provided a basic understanding of electricity.

At the same time, the electron was one of the foci of atomic physics as it was introduced into Japan in the early 20th century.⁷ The result was that there was an emphasis on the theory of the electron in textbook physics. Simultaneously, physicists themselves took advantage of the notoriety of electrons in their popularizing activities. Here, I give three examples for this. The first, Nagaoka Hantarō (1865–1950), known for his Saturnian model of an atom,⁸ was one of the early Japanese professors of physics at Tokyo Imperial University and the leading Japanese atomic physicist before the introduction of quantum mechanics. His popular book, *Genkon no denki-gaku* (Studies of electricity today) (Nagaoka 1912), is an account of the physics related to electricity, including not just classical electromagnetic theory but also atomic physics, with an obvious focus on the electron.

The second, Mizuno Toshinojō (1862–1944), was a contemporary but less-known physicist at Kyoto Imperial University. While Nagaoka's activities extended into various areas, Mizuno focused on studies of the electron. His 1912 book, *Denshiron* (The electron theory), was basically a textbook of atomic physics (Mizuno 1912). More interestingly, his 1918 book, *Densi no katsudō* (Activities of the electron), was a more popular book, aimed at explaining the recent physics of the electron to a lay audience (Mizuno 1918).

The third example is Aichi Keiichi (1880–1923), another theoretical physicist who worked at Tohoku Imperial University. Known for many highly mathematical works on various topics, he wrote a book entitled *Denshi no jijoden: Tsūzoku kagaku kōgi* (The autobiography of an electron: A popular science account), a popular account of atomic physics (Aichi 1923). As in Charles Gibson's *The Autobiography of an Electron* (1911) published over ten years earlier, the electron takes the role of narrator and describes its properties and roles in nature and society.

Apparently, these physicists tried to gain popular support and interest by using the electron to connect their field to electricity. At the same time, as professors at Japan's key higher education institutions, they were in a position to teach students, who might seek jobs at electric companies, as well as to teach electromagnetism to future engineers at secondary and tertiary schools.

In light of this context, Nishina's career appears more cohesive; beginning in his undergraduate years, he mostly studied topics related to the electron. In fact,

⁷Other important pre-quantum mechanical topics in physics that had practical implications were spectroscopy and X-ray physics, which provided training grounds and institutional bases for experimental atomic physicists. Nishina was familiar with these traditions, but I do not discuss this issue here.

⁸For a biography of Nagaoka in English, see (Yagi 1974).

his elder brother, Nishina Teisaku, described the purpose of Nishina's long stay in Europe as a study of electricity and assumed that his study would have practical implications for electrical engineering. This was probably the way Nishina justified his study abroad to his family and relatives.⁹

7.2 Nishina Yoshio as an Engineering Student

The connection between electrical engineering and physics went farther than the fact that both fields dealt with electricity. The theoretical and mathematical aspects of these research areas reveal additional similarities. As a student of electrical engineering, Nishina was immersed in a highly theoretical and mathematical school of electrical engineering under Hō Hidetarō (1872–1931), who was Nishina's academic advisor at Tokyo Imperial University.

Hō is known as the author of many electrical engineering textbooks, in particular works on alternating current circuits and transition phenomena. He is also known for introducing Charles Proteus Steinmetz's theory of alternating current (Steinmetz 1893) to Japan.¹⁰ He is best known for his work in Japan on the Hō-Thévenin theorem.

Nishina's disciplinary identity was deeply lodged in the alternating current theory of electrical engineering under the influence of Hō and Steinmetz. In his later years, when asked to discuss the books that inspired him in his youth, Nishina listed four: *The Mathematical Theory of Electricity and Magnetism* by James Jeans published in 1908, *Theory and Calculation of Alternating Current Phenomena* by Steinmetz published in 1897, *Wechselstromtechnik* by Engelbert Arnold published in 1902, and *Kōryū riron* (Alternating current theory) by Hō published in 1912 (Y. Nishina 1946).¹¹

To illustrate the kinds of physical phenomena Hō treated in his textbook, let me summarize one example from his *Hadō, shindō oyobi hirai* (Wave, vibration and lightning arrester), first published in 1915.¹² When electric current flows along an ideal wire (whose resistance and inductance can be ignored), nothing happens. Hō constructed a theory that predicts what happens when there are various kinds of electromagnetic "barriers" along its way. Here, as an example, I

⁹Nishina Akira, a nephew of Yoshio's, remembered that when Nishina's elder brother Teisaku mentioned that Yoshio was studying electricity at a party to celebrate Yoshio's return to Japan in 1929, Teisaku emphasized how Japan was backward as far as the study of electricity was concerned and asked the other guests for support for Yoshio (A. Nishina 1975).

¹⁰For a biography of Steinmetz, see (Kline 1992).

¹¹Steinmetz's innovation in alternating current theory was the use of imaginary numbers in understanding alternating current circuits. Steinmetz replaced vector diagrams with imaginary numbers.

¹²In the following discussion, I use the 1923 version of the book (Hō 1923), which probably better represents the instruction that Nishina received in his student years.

take the case of electric current in a sinusoidal (wave-like) form and a coil acting as a “barrier.”

Suppose there is a wire whose inductance and capacitance per length are L_1 and C_1 respectively, and there is a coil at point b , whose inductance is L . The coil is infinitesimally short and can be treated as a point. There are currents in the form of incoming, reflecting and penetrating waves at point b . Let us say functions $f_1(t)$, $f_2(t)$, and $f_3(t)$ are the currents of the incoming, reflecting and penetrating waves. From classical electromagnetic theory, it follows:

$$\begin{aligned} f_2(t) &= e^{-nt} n \int e^{nt} f_1(t) dt + Ae^{-nt}, \\ f_3(t) &= f_1(t) - e^{-nt} n \int e^{nt} f_1(t) dt - Ae^{-nt}, \end{aligned} \quad (7.1)$$

where $n = \frac{2}{L} \sqrt{\frac{L_1}{C_1}}$ and A is determined by initial conditions.

For example, if the incoming wave is a half wavelength of a sinusoidal wave with a certain angular frequency ω :

$$f_1(t) = I_1 \sin \omega t \quad (0 \leq t \leq \frac{\pi}{\omega}), \quad (7.2)$$

the penetrating and reflecting waves would be:

$$\begin{aligned} f_2(t) &= I_1 \sqrt{\frac{\omega^2}{n^2 + \omega^2}} \sin(\omega t + \phi) - \frac{n\omega_1}{n^2 + \omega^2} e^{-nt} \quad (0 \leq t \leq \frac{\pi}{\omega}) \\ f_3(t) &= \frac{nI_1}{\sqrt{n^2 + \omega^2}} \sin(\omega t - \phi) + \frac{n\omega_1}{n^2 + \omega^2} e^{-nt} \quad (0 \leq t \leq \frac{\pi}{\omega}) \\ f_3(t) &= -f_2(t) = \frac{n\omega_1}{n^2 + \omega^2} (1 + e^{-n(t - \frac{\pi}{\omega})}) \quad (\frac{\pi}{\omega} < t) \\ \phi &= \arctan \frac{\omega}{n}. \end{aligned} \quad (7.3)$$

A case like the one above resembles a one-dimensional scattering problem: a wave collides with a particle, making a certain interaction with it. In a sense, Hō solved it to the first order.¹³ While it would be absurd to see a direct connection between such work and the problem that Nishina would later deal with in Compton scattering, there nevertheless seems to be a mathematical affinity between the problems in physics Nishina worked on starting in the 1920s and the kind of problems that Hō dealt with and taught in the electrical engineering department.

A possibly more interesting link between Hō’s alternating current theory and quantum physics is found in the notion of linearity that lies at the heart of both

¹³The assumption that the coil is infinitely small is equivalent to substituting for the actual curve a flat line within that length, that is equivalent to the first-order approximation.

theories. In his textbook on alternating current theory, after describing notations and fundamental notions, Hō (1912) started the main part of his textbook on alternating current theory with a discussion of the principle of superposition, just as Dirac (1930) started his textbook on quantum mechanics. Suppose there are three configurations: A , B , and C . They have the same circuit elements, except:

- Configuration A : There is a voltage source E_1 at point A , but none at B ;
- Configuration B : There is a voltage source E_2 at point B , but none at A ;
- Configuration C : There is a voltage source E_1 at point A , and E_2 at B .

Then one obtains the solution to Configuration C , where there are sources at both points A and B , by adding up the solutions to Configurations A and B .

The principle of superposition turned out to be a key component of Hō's fame in Japan. In 1922, in seeking to apply this principle to the problem of power lines, he rediscovered what is now known as Thévenin's equivalent circuit theorem, a cornerstone of circuit theory, which every electrical engineering student learns today. The theorem makes the calculations of complicated electrical circuits much easier than they would be if one applied Kirchhoff's laws directly. Thévenin's theorem resembles what Edwin Layton called "engineering sciences" (Layton 1971, 567). While the value of this theorem mostly lies in its practicality, it is a mid-level theorem, derived rigorously through theoretical considerations from fundamental principles, namely the principle of superposition and Kirchhoff's laws. This theorem had been derived previously and sometimes independently by several scientists and engineers, including Hermann von Helmholtz, Léon Charles Thévenin, Hans Ferdinand Mayer and Edward Lawry Norton. Although presented in different formulations, the theorem is essentially a way to substitute a part of a complex circuit with a simpler equivalent circuit consisting of a certain voltage source and a resistance. Thévenin formulated the theorem as follows:

Assuming any system of linear conductors connected in such a manner that to the extremities of each one of them there is connected at least one other, a system having some electromotive forces, E_1, E_2, \dots, E_n , no matter how distributed, we consider two points A and A' belonging to the system and having actually the potentials V and V' . If the points A and A' are connected by a wire ABA' , which has a resistance r , with no electromotive forces, the potentials of points A and A' assume different values of V and V' , but the current I flowing through this wire is given by the equation $I = \frac{V-V'}{R+r}$, in which R represents the resistance of the original system, this resistance being measured between the points A and A' , which are considered to be electrodes. (Suchet 1949, 843–844)

The proof can be stated simply. Here is a textbook presentation of Thévenin's proof without much change from the original. Let us define the following four configurations:

1. Configuration I is defined as the original system where A and A' are not connected as in fig. (7.1).¹⁴
2. Configuration I' is defined as the one where A and A' are connected, and there is a voltage source of $V - V'$ at point B . The voltage source is connected in the opposite direction to A and A' , so that there is no current between them as in fig. (7.2).
3. Configuration II is defined as the system having the same resistance as Configuration I' but no voltage source, except the one at B in the opposite direction to the one in Configuration I' as in fig. (7.3).
4. Configuration III is the system where A and A' are connected by a wire of resistance r as in fig. (7.4).

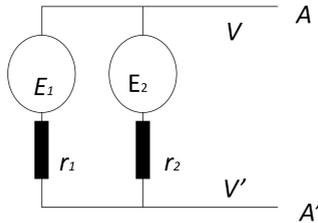


Figure 7.1: Configuration I: Unconnected circuit.

¹⁴Figures (7.1)–(7.4) produced by the author.

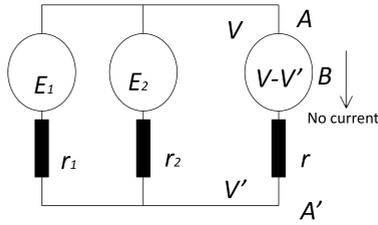


Figure 7.2: Configuration I': Circuit with an added voltage source.

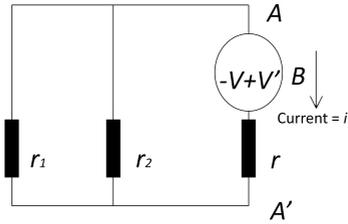


Figure 7.3: Configuration II: Thévenin's equivalent circuit.

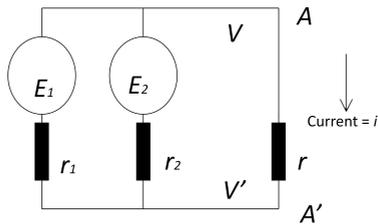


Figure 7.4: Configuration III: Connected circuit.

Since there is no current between A and A' in Configuration I', Configuration I' gives the same voltages and currents at each point as in Configuration I.

According to the principle of superposition, Configuration III can be obtained by adding Configuration I' to Configuration II. Since there is no current at B , the current at B in Configuration III comes only from Configuration II, which is what the theorem states.

Hō reached a form of this theorem without knowing that others had already found it. In his 1922 paper on power transmission, Hō devised a way to calculate the effects of an accidental grounding of a transmission line, by ingeniously using the principle of superposition. The result was the same as Thévenin's theorem, except that Hō discussed an alternating current circuit instead of a direct current circuit and grounding instead of shorting. Hō's proof was equivalent to Thévenin's. Hō considered the transmission line grounded by a wire with impedance R as in fig. (7.5). Suppose the voltage at point a is given by:

$$v_a = V_m \sin(\omega t - \theta_0), \quad (7.4)$$

where V_m , ω , θ_0 are the amplitude, angular frequency, and initial phase of the voltage. There will be no current through R if there is an electromotive force with the same strength but in the opposite direction as in fig. (7.6). If there is an electromotive force with the same strength but in the opposite direction at the same point as in fig. (7.7), then those two electromotive forces cancel each other and the result should be the same as fig. (7.5). Since fig. (7.7) can be obtained by superposing fig. (7.6) and fig. (7.8), the current through R can be calculated by fig. (7.8) (Hō 1922).

Although this was a special case of what we today call Thévenin's theorem, Hō's proof was the same as the proof for the general case. Because of this work, the theorem is called the Hō-Thévenin theorem in Japan, with Hō's name firmly attached. Whether Hō is entitled to be named one of the discoverers of this theorem is not the issue here. What is of interest is that the principle of superposition was so central to Hō's work.

Nishina was situated deeply in this tradition of alternating current theory, in which the principle of superposition dominated. Not only did Nishina read Hō's textbooks and attend his classes, he wrote his bachelor's thesis along the line manifested by Hō's derivation of the Hō-Thévenin theorem. The main question in Nishina's thesis was how unbalanced loads would affect an alternator, a motor, or a rotary transformer in a polyphase system. It relied heavily on Hō's and Steinmetz's work (Y. Nishina 1918).

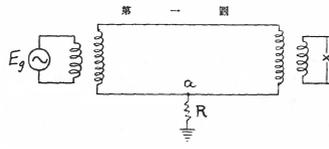


Figure 7.5: Grounded circuit (Source: Hō 1922, 193).

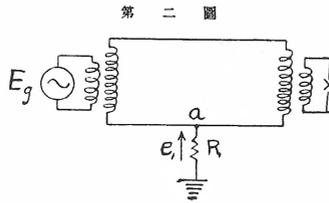


Figure 7.6: Circuit equivalent to non-grounded circuit (Source: Hō 1922, 194).

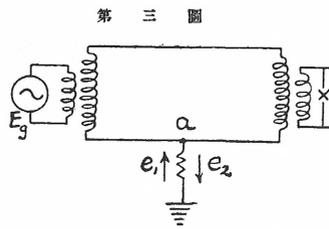


Figure 7.7: Circuit equivalent to grounded circuit (Source: Hō 1922, 194).

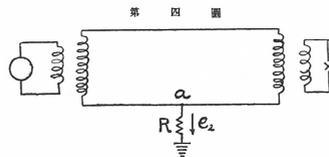


Figure 7.8: Hō's equivalent circuit (Source: Hō 1922, 194).

Nishina started his bachelor's thesis with definitions of a few main concepts. In an N -phase system, if the voltage is equal in all branches and the phase difference between the branches is one N th, the system is called symmetrical. If not, it is asymmetrical. If the sum of the power in all N branches is constant, it is called balanced, if not, unbalanced. A symmetrical system, for example, can be unbalanced when loaded unequally.

According to Nishina, the problem of imbalances in three-phase systems was very practical. Nishina thought that, as the centralization of the electrical power supply continued, the three-phase system would be the most efficient for generating and transmitting electric power. However, after the introduction of the single-phase commutator motor, there arose a demand for single-phase electrical power supplies. If a single-phase load was supplied with electricity directly from a three-phase system, the voltage would become unbalanced. Hence, the problem of an unbalanced load would ensue. With such motivating factors in mind, Nishina proceeded to the main part of his thesis, which discussed how unbalanced loads would affect a few types of alternating current device, such as an alternator, a motor and a rotary transformer.

In the case of the alternator, Nishina examined what would happen when loads were connected in an unbalanced fashion to a three-phase alternator (that is when loads were connected to only one or two of the three phases). Treating the problem theoretically, Nishina argued that unfavorable effects would result. Terminal voltage would become "unsymmetrical" both in phase and in magnitude (Y. Nishina 1918, 94–95). This would increase both iron and copper loss, reducing efficiency and producing more heat. The unbalanced load would also cause odd higher harmonics, which would result in an "uncomfortable" humming noise.

In analyzing the unbalanced system, Nishina applied reasoning similar to Hō's "principle of superposition." In his discussion of the unbalanced three-phase system, he claimed that it could be considered a superposition of two balanced three-phase systems rotating in opposite directions, or in his words: "An unbalanced polyphase system can be resolved into two balanced components with opposite phase rotations, one positive and the other negative." Nishina cited R. E. Gilman and Charles LeGeyt Fortescue, who originally "discovered" and "proved" this theorem (Y. Nishina 1918, 20). In his thesis, Nishina reproduced their proof.

The proof goes as follows. Define ϵ as:

$$\epsilon = \exp(2\pi j/n), \quad (7.5)$$

where j is the imaginary unit and n is the number of the phase. E_1, E_2, \dots, E_n , and E'_1, E'_2, \dots, E'_n , are terminal voltages of two symmetrical n -phase systems, rotating in opposite directions. Since a factor of ϵ rotates the phase of a complex

number by $2\pi/n$, these terminal voltages of the symmetrical n -phase systems can be written as: $E_1 = E_1, E_2 = \epsilon E_1, E_3 = \epsilon^2 E_1, \dots, E_n = \epsilon^{n-1} E_1$ and $E'_1 = E'_1, E'_2 = \epsilon E'_1, E'_3 = \epsilon^{-2} E'_1, \dots, E'_n = \epsilon^{-n+1} E'_1$. Nishina's claim above states that for any n phase system, of which the terminal voltages are V_1, V_2, \dots, V_n , there are E_1, E'_1 , that satisfy $V_1 = E_1 + E'_1, V_2 = E_2 + E'_2, \dots, V_n = E_n + E'_n$. Nishina's proof goes as follows. If the equations above are multiplied by $\epsilon^n, \epsilon^{n-1}, \dots, \epsilon^1$, then:

$$\begin{aligned} \epsilon^n V_1 &= \epsilon^n E_1 + \epsilon^n E'_1, \\ \epsilon^{n-1} V_2 &= \epsilon^n E_1 + \epsilon^{n-2} E'_1 \\ &\dots \\ \epsilon V_n &= \epsilon^n E_1 + \epsilon^{-n+2} E'_1. \end{aligned} \tag{7.6}$$

By summing both sides of the equations, and using the definition of ϵ , the result is:

$$\epsilon^n V_1 + \epsilon^{(n-1)} V_2 + \dots + \epsilon V_n = n \epsilon^2 E_1. \tag{7.7}$$

This determines E_1 . E'_1 can be derived similarly. This derivation indicates Nishina's familiarity with the idea of analyzing the physical system by separating it into superposed components, as well as with the ways of exploiting the linearity of alternating current circuits, just as Hō had done.

The most interesting aspect of this proof is, however, that it was wrong. Eq. (7.7) is a necessary condition for the original equations for E_1 and E'_1 , but it is not guaranteed that the derived forms of E_1 and E'_1 and the other terminal voltages satisfy the original equations. In fact, they do not satisfy these equations in general, which one can confirm by simple substitution. As famously shown by Fortescue (1918), decomposing an N -phase unbalanced system requires N -balanced components. However, since the actual problems Nishina dealt with were three-phase systems, this theoretical mistake was not catastrophic. In short, Nishina's thesis arguments were mathematically inaccurate but, probably helped by his physical intuition,¹⁵ his conclusions were physically correct.

Nishina's thesis shows his commitment to Hō's theoretical tradition of electrical engineering, his close ties to Steinmetz's tradition,¹⁶ and his ability in the-

¹⁵Nishina could have decomposed an arbitrary unbalanced three-phase system into three, rather than two, symmetrical systems: two rotating in opposite directions and one not rotating at all. Nishina used the reverse component to show the production of higher harmonics and other undesirable effects. These qualitative conclusions did not change significantly whether or not one took the stationary component into consideration.

¹⁶As for Steinmetz's tradition, see (Kline 1992).

oretical and physical reasoning. Nishina's work was theoretical in the sense that he derived fairly general characteristics of the three-phase system. Although the thesis reveals Nishina's relative mathematical weakness compared to some of the mathematical wizards entering physics at the time, it nonetheless demonstrates Nishina's ability to draw physically correct conclusions and highlights his immersion in the Hō tradition, especially his familiarity with Hō's strategy for exploiting linearity and superposition to represent physical phenomena.

7.3 The Klein-Nishina Formula

The principle of superposition occupies a central place in quantum mechanics. In particular, the idea plays a crucial role in Dirac's formulation of quantum mechanics, as manifested by Dirac's textbook first published in 1930 (Dirac 1930). This book was soon translated into Japanese by Nishina and his students, including Tomonaga Sin-Itiro (Dirac 1936). Hence, it is reasonable to suppose that Nishina's familiarity with the principle of superposition through electrical engineering was useful to him when learning quantum mechanics and when carrying out quantum theoretical research.

This section and the next closely examine Nishina's earliest and most important work in theoretical physics, performed in collaboration with Klein, and resulting in the so-called Klein-Nishina formula (Klein and Y. Nishina 1929). I explore how Nishina, along with Klein, actually employed the idea of superposition in his theoretical research in quantum mechanics.

Historically, the Klein-Nishina formula was one of the earliest applications of Dirac's theory. It was very quickly confirmed experimentally giving strong empirical evidence to support Dirac's theory, which had various conceptual problems, including the issue of negative energies. Klein and Nishina derived this formula through a semi-classical treatment of Compton scattering. Following Walter Gordon and Dirac, they used Dirac's relativistic theory of the electron instead of the non-relativistic theory. Such a semi-classical approach for the Compton effect was first explored by Gordon (1926). In his 1926 paper, Gordon proceeded by comparing classical and quantum mechanical calculations. He had a relatively simple picture behind his calculation. Incoming radiation disturbs and imparts motion to an electron through electromagnetic interaction. Gordon first calculated how the incoming radiation would interact with the electron, both in classical mechanics and quantum mechanics. When moving, the electron, a charged particle, emits radiation, which Gordon calculated using a classical electromagnetic formula. He assumed the emitted radiation to correspond to the outgoing X-ray observed in the experiment. What Gordon did can be written mathemati-

cally as follows. He assumed that the incoming radiation was a monochromatic plane wave, setting its (four-)vector potential Φ_α as:

$$\begin{aligned}\Phi_\alpha &= a_\alpha \cos\phi, \quad a_4 = ia_0 \\ \phi &= \frac{2\pi\nu}{c}(\sum n_k x_k - ct),\end{aligned}\tag{7.8}$$

where c is the speed of light, n_k the vector that gives the direction of the radiation, ν the frequency of the wave, a_α the amplitude of the wave, and α the index of a four-vector taking the values 1 through 4 (alternatively 0 through 3), whereas k is the index of a three-vector, taking the values 1 through 3. The first task was to solve the equation of motion for the electron. For the quantum mechanical treatment of this problem, Gordon chose to use the Klein-Gordon equation:

$$\left(\sum\left(\frac{\hbar}{2\pi i}\frac{\partial}{\partial x_\alpha} - \frac{e}{c}\Phi_\alpha\right)^2 + m^2 c^2\right)\Psi = 0,\tag{7.9}$$

where Ψ is the wave function. In the presence of the above-mentioned incoming radiation, this equation can be solved to first order, giving:

$$W = px + \frac{pb}{pl}\sin\phi, \quad \Psi = e^{\frac{2\pi i}{\hbar}W},\tag{7.10}$$

where $b_\alpha = \frac{e}{c}a_\alpha$, $l_k = \frac{2\pi\nu}{c}n_k$, $l_0 = i2\frac{p\nu}{c}$, and px, pb, pl are all inner products of four vectors.

In classical mechanics, where the electron can be considered a point mass, the electromagnetic wave resulting from its motion is easily calculated. In particular, the frequency of this wave is trivially the same as the frequency of the moving electron. The quantum mechanical treatment required a more complicated procedure, since the electron needed to be considered not as a point mass but as a spatially distributed wave.

From the specific solutions of the equation, Gordon wrote up the general form of the solution as an arbitrary superposition of them:

$$\Psi = \int z(\mathbf{p})C(\mathbf{p})e^{\frac{2\pi i}{\hbar}W}d\mathbf{p} \quad (d\mathbf{p} = dp_1 dp_2 dp_3),\tag{7.11}$$

where $z(\mathbf{p})$ is a weight and $C(\mathbf{p})$ is a normalization factor (Gordon 1926, 125). Gordon assumed that the electric current in quantum mechanics should take the following form:

$$s_\alpha = \frac{1}{i}(\bar{\Psi}\frac{\partial\Psi}{\partial x_\alpha} - \Psi\frac{\partial\bar{\Psi}}{\partial x_\alpha} - \frac{4\pi i}{h}\frac{e}{c}\Phi_\alpha\Psi\bar{\Psi}).\tag{7.12}$$

Then he plugged this expression into the classical electromagnetic formula for the retarded potential,

$$\Phi'_\alpha = \frac{1}{c} \int \frac{[s_\alpha]}{R} dx \quad (dx = dx_1 dx_2 dx_3), \quad (7.13)$$

which gives the electromagnetic field caused by the current. Here, R is the spatial distance between the volume element dx of the integral and the point in question. The brackets [] indicate that t in s_α should be substituted by $(t - R/c)$. Then, Gordon calculated the frequency and intensity of the induced radiation. The result, according to Gordon, agreed with the one obtained by Dirac in his 1926 paper:

$$I = \frac{e^4}{m^2 c^4 r^2} I_0 \frac{\sin^2 \phi}{(1 + \alpha(1 - \cos \theta))^3}, \quad (7.14)$$

where ϕ is the angle between the electric field and the observed direction, θ the angle between the direction of the incoming radiation and the observed direction, I_0 the intensity of the incoming radiation, and r the distance between the point of scattering and the point of observation. It was approximately identical to the formula obtained by Arthur Compton.

Klein and Nishina adopted Gordon's approach, but used Dirac's relativistic theory of electrons instead of the Klein-Gordon equation and the formula for the electron current given by Gordon. The first step is to solve the Dirac equation for a free electron, whose spin is zero on average, and for an electron in a monochromatic unpolarized radiation. The Klein-Gordon equation is replaced by the Dirac equation:

$$\left(\frac{E + cV}{c} + \rho_1(\boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c}\boldsymbol{\mathfrak{A}}) + \rho_3 mc \right) \Psi = 0, \quad (7.15)$$

in which $\boldsymbol{\sigma}, \boldsymbol{\rho}$ are (three-)vectors of 4 by 4 matrices given by Dirac, whose elements satisfy the following relations:

$$\begin{aligned} \sigma_r^2 &= 1, & \sigma_r \sigma_s + \sigma_s \sigma_r &= 0 \quad (r \neq s) \\ \rho_r^2 &= 1, & \rho_r \rho_s + \rho_s \rho_r &= 0 \quad (r \neq s) \\ \rho_r \sigma_t &= \sigma_t \rho_r &= 0. \end{aligned} \quad (7.16)$$

Suppose the monochromatic radiation is given by a vector potential of the following form:

$$\boldsymbol{\mathfrak{A}} = \mathbf{a} e^{i\nu(t - \frac{r}{c})} + \bar{\mathbf{a}} e^{-i\nu(t - \frac{r}{c})}, \quad (7.17)$$

where \mathbf{n} is the unit vector in the direction of the incoming wave. Nishina and Klein calculated the solution up to first order:

$$\begin{aligned}\phi(\mathbf{p}) &= \phi_0(\mathbf{p})\{1 + f(\mathbf{p})e^{iv(t-\frac{n\mathbf{r}}{c})} + \bar{f}(\mathbf{p})e^{-iv(t-\frac{n\mathbf{r}}{c})}\}, \\ \psi(\mathbf{p}) &= \{1 + g(\mathbf{p})e^{iv(t-\frac{n\mathbf{r}}{c})} + \bar{g}(\mathbf{p})e^{-iv(t-\frac{n\mathbf{r}}{c})}\}\psi_0(\mathbf{p}),\end{aligned}\quad (7.18)$$

in which $\phi_0(\mathbf{p})$ and $\psi_0(\mathbf{p})$ are eigenfunctions of a free electron:

$$\phi_0(\mathbf{p}) = u(\mathbf{p})e^{\frac{i}{\hbar}[Et-(\mathbf{p}\mathbf{r})]}, \quad \psi_0(\mathbf{p}) = v(\mathbf{p})e^{\frac{-i}{\hbar}[Et-(\mathbf{p}\mathbf{r})]}, \quad (7.19)$$

and f, \bar{f} and g, \bar{g} are constant matrices, determined by:

$$\left. \begin{aligned}f(\mathbf{p}) &= \frac{e}{2\hbar v(E/c-(\mathbf{n}\mathbf{p}))}\{2(\mathbf{a}\mathbf{p}) - h(\boldsymbol{\sigma}\boldsymbol{\eta}) - ih\rho_1(\boldsymbol{\sigma}\boldsymbol{\epsilon})\}, \\ \bar{f}(\mathbf{p}) &= -\frac{e}{2\hbar v(E/c-(\mathbf{n}\mathbf{p}))}\{2(\bar{\mathbf{a}}\mathbf{p}) + h(\boldsymbol{\sigma}\bar{\boldsymbol{\eta}}) - ih\rho_1(\boldsymbol{\sigma}\bar{\boldsymbol{\epsilon}})\}, \\ g(\mathbf{p}) &= -\frac{e}{2\hbar v(E/c-(\mathbf{n}\mathbf{p}))}\{2(\mathbf{a}\mathbf{p}) + h(\boldsymbol{\sigma}\boldsymbol{\eta}) + ih\rho_1(\boldsymbol{\sigma}\boldsymbol{\epsilon})\}, \\ \bar{g}(\mathbf{p}) &= \frac{e}{2\hbar v(E/c-(\mathbf{n}\mathbf{p}))}\{2(\bar{\mathbf{a}}\mathbf{p}) + h(\boldsymbol{\sigma}\bar{\boldsymbol{\eta}}) + ih\rho_1(\boldsymbol{\sigma}\bar{\boldsymbol{\epsilon}})\}.\end{aligned}\right\} \quad (7.20)$$

Here, $\boldsymbol{\epsilon}$ and $\bar{\boldsymbol{\epsilon}}$ or $\boldsymbol{\eta}$ and $\bar{\boldsymbol{\eta}}$ respectively depend on the electric and magnetic fields of the radiation, $\boldsymbol{\mathfrak{S}}$ and \mathfrak{D} :

$$\left. \begin{aligned}\boldsymbol{\epsilon} &= -\frac{iv}{c}\mathbf{a}, & \bar{\boldsymbol{\epsilon}} &= \frac{iv}{c}\bar{\mathbf{a}}, \\ \boldsymbol{\eta} &= -\frac{iv}{c}[\mathbf{n}\mathbf{a}], & \bar{\boldsymbol{\eta}} &= \frac{iv}{c}[\mathbf{n}\bar{\mathbf{a}}].\end{aligned}\right\} \quad (7.21)$$

The general solution of the wave equations arises from a superposition of all possible solutions of the form (7.18) up to the considered approximation.

$$\Phi = \int \phi(\mathbf{p})d\mathbf{p}, \quad \Psi = \int \psi(\mathbf{p})d\mathbf{p}. \quad (7.22)$$

The second step is to calculate the current density using this solution. The following formula for the current density is given by Dirac in his 1928 paper (Dirac 1928):

$$\boldsymbol{\mathfrak{S}} = ec\Psi\rho_1\boldsymbol{\sigma}\Psi = ec \int \int \phi(\mathbf{p})\rho_1\boldsymbol{\sigma}\psi(\mathbf{p}')d\mathbf{p}d\mathbf{p}'. \quad (7.23)$$

Inserting the solutions above, we have:

$$\begin{aligned} \mathfrak{J} = \mathfrak{J}_0 + ce \int \int dp dp' \{ & u(\mathbf{p}) [\rho_1 \boldsymbol{\sigma} g(\mathbf{p}')] \\ + f(\mathbf{p}) \rho_1 \boldsymbol{\sigma} v(\mathbf{p}') e^{\frac{i}{\hbar} [(E+hv-E')t - (\mathbf{p}+n\frac{hv}{c} - \mathbf{p}')r]} & \\ + c.c. \}. \end{aligned} \quad (7.24)$$

Using the formula for the retarded potential, this electric current causes radiation \mathfrak{A}' :

$$\begin{aligned} \mathfrak{A}' = \frac{c}{r} \int \int dp dp' \{ & e^{\frac{i}{\hbar} (E+hr-E')(t-\frac{r}{c})} \int dx u(\mathbf{p}) [\rho_1 \boldsymbol{\sigma} g(\mathbf{p}')] \\ + f(\mathbf{p}) \rho_1 \boldsymbol{\sigma} v(\mathbf{p}') e^{-\frac{i}{\hbar} [\mathbf{p}-\mathbf{p}'+n\frac{hv}{c} - n'\frac{E+hv-E'}{c}]r} & \\ + c.c. \}, \end{aligned} \quad (7.25)$$

where $\int dx$ means the integration $\int dx_1 \int dx_2 \int dx_3$ over the entire region available to the electron.

By further calculation, and using Maxwell's equations and various relations that the Dirac matrices satisfy, we have for the magnetic field \mathfrak{D}_0 :

$$\begin{aligned} \mathfrak{D}_0 = \frac{(2\pi\hbar)^3 e^2 v'}{2mc^2 r (v-v' + \frac{2mc^2}{\hbar})} \sqrt{\frac{E' v'}{mc^2 v}} \{ & d(\frac{1}{v'} (\mathbf{n}' \boldsymbol{\epsilon}) (v' - v) [[\mathbf{n}' \boldsymbol{\pi}] \\ - v' (\frac{1}{v} + \frac{1}{v'})^2 \frac{mc^2}{\hbar} [\mathbf{n}' \boldsymbol{\epsilon}]] - i [(\frac{1}{v'} - \frac{1}{v}) ((\boldsymbol{s}, \mathbf{n}v - \mathbf{n}' v') ((\mathbf{n}' \boldsymbol{\epsilon}) \boldsymbol{\pi} & \\ - (\mathbf{n} \boldsymbol{\pi}') \boldsymbol{\epsilon}) + (v - v' + \frac{2mc^2}{\hbar}) ((\boldsymbol{s} \mathbf{n}') \boldsymbol{\epsilon} - (\mathbf{n}' \boldsymbol{\epsilon}) \boldsymbol{s})) & \\ + \frac{2}{v} (\boldsymbol{\epsilon} \mathbf{n}') ((\mathbf{n}' \boldsymbol{s}) (\mathbf{n}v - \mathbf{n}' v') + (v' - (\mathbf{n} \boldsymbol{\pi}') v) \boldsymbol{s}) & \\ - (\frac{1}{v} + \frac{1}{v'}) ((\mathbf{n} [\boldsymbol{\epsilon} \boldsymbol{s}]) v [\mathbf{n}' \boldsymbol{\pi}] + v' (\mathbf{n}' [\mathbf{n} \boldsymbol{\epsilon}]) [\mathbf{n}' \boldsymbol{s}])] \} e^{iv'(t-\frac{r}{c})} & \\ + c.c. \end{aligned} \quad (7.26)$$

Here, \mathbf{n}' is a unit vector in the direction of observation, v' is the frequency of the outgoing radiation, and the following abbreviations are used:

$$\left. \begin{aligned} u(\mathbf{p}) \boldsymbol{\sigma} v(\mathbf{p}') = \boldsymbol{s}, \quad u(\mathbf{p}') \boldsymbol{\sigma} v(\mathbf{p}) = \bar{\boldsymbol{s}} \\ u(\mathbf{p}') v(\mathbf{p}') = d, \quad u(\mathbf{p}') v(\mathbf{p}) = \bar{d}. \end{aligned} \right\} \quad (7.27)$$

The fourth and final step is to calculate an observable physical quantity, such as the intensity of scattered radiation. For this, we need to calculate the expectation value of \mathfrak{D}_0^2 . Beside obviously meaning a long calculation, including manipulations of Dirac matrices, this also involves the question of how to evaluate

expectation values, which I discuss closely in the following section. The result is:

$$\bar{\mathfrak{D}}_0^- = \frac{e^4}{m^2 c^2 r^2} \left(\frac{v'}{v}\right)^3 \left\{ \left(\frac{v}{v'} + \frac{v'}{v}\right) \epsilon^2 - 2(n'\epsilon)^2 \right\}. \quad (7.28)$$

From this, if we write the angle between the observation direction and the direction of the incoming wave as Θ , and the angle between the observation direction and the electric force of the incoming wave as θ , the intensity of the outgoing radiation is:

$$I = I_0 \frac{e^4}{m^2 c^4 r^2} \frac{\sin^2 \theta}{(1 + \alpha(1 - \cos \Theta))^3} \left(1 + \alpha^3 \frac{(1 - \cos \Theta)^2}{2 \sin^2 \theta (1 + \alpha(1 - \cos \Theta))} \right), \quad (7.29)$$

where $\alpha = \frac{h\nu}{mc^2}$, I_0 is the intensity of the incoming radiation.

One could point out that this particular approach to this problem is reminiscent of electrical engineering in various ways. After all, calculating electromagnetic waves emitted by electric current is one of the foremost topics of electrical engineering.

7.4 Superposition of States in the Klein-Nishina Paper

Although the fourth step above might appear to be a very lengthy and complex, but straightforward calculation, in reality, it was more than that. Yazaki Yūji (1992) has clarified Klein and Nishina's thought process by closely studying the Klein-Nishina paper and the extensive archival materials that Nishina left at RIKEN. According to Yazaki, the main physical problem Nishina and Klein faced was to determine the initial and final states of the electron and to calculate the average of the magnetic field. The procedure for calculating an average of a physical quantity was neither standardized nor clear at this point. It was even less apparent because of the semi-classical approach. Today, we know that we need to calculate contributions from orthogonal states, but in their time, the notion of orthogonality and its relevance to the calculation of a quantum statistical average was not clear. Klein and Nishina thought that they would take the average of two "independent states," such as the two states having magnetic moments (spin) in opposite directions. However, the states for spin up and down along the x_3 -axis satisfy the Dirac equation only when $\mathfrak{p} = 0$. Eventually, they solved this problem using the method of Lorentz transformation that Klein developed (see below on page

203). As Yazaki points out, physical considerations were Klein and Nishina's main concern (Yazaki 1992).¹⁷

My aim here, however, is not to revisit their physical considerations, which Yazaki has already studied. Instead I intend to make explicit and confirm that the idea of superposition of states did indeed play an important role in the calculation of expectation values Klein and Nishina carried out.

Since \mathfrak{D}_0^2 is a sum of bilinear terms such as $u(\mathbf{p}')\alpha v(\mathbf{p}) \cdot u(\mathbf{p})\beta v(\mathbf{p}')$, where α and β are certain matrices, Klein and Nishina had to calculate these terms. The authors reduced all the terms to the calculation of $u(\mathbf{p}')v(\mathbf{p}')$ and $u(\mathbf{p}')\sigma v(\mathbf{p}')$. For these terms, Klein and Nishina derived the following values:

$$u(\mathbf{p}')\bar{\sigma}v(\mathbf{p}') = 0, \quad (7.30)$$

$$u(\mathbf{p}')v(\mathbf{p}') = u^*(\mathbf{p}')v^*(\mathbf{p}') = 2(2\pi h)^{-3}, \quad (7.31)$$

where the overbar means average.

The factor of 2 in this equation is remarkable, especially because eq. 20 in the Klein-Nishina paper (Klein and Y. Nishina 1929, 858) states:

$$u(\mathbf{p})v(\mathbf{p}) = (2\pi h)^{-3}. \quad (7.32)$$

How did they reach these values? This part of their calculation hinged on their choice for the initial and final states of the electron.

For the initial state, it seems appropriate to choose spin up and down states as two independent states. To explain this choice, it is necessary to explicitly fix spinor and matrix elements. Klein and Nishina chose ρ_3 and σ_3 to be diagonal. Hence:

$$\rho_3 = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (7.33)$$

From the Dirac equation for free electrons, it follows that in the case of $\mathbf{p} = 0$,

$$u(0)(1 + \rho_3) = 0, \quad (1 + \rho_3)v(0) = 0. \quad (7.34)$$

Thus, one can choose two independent solutions, with either u_1, v_1 or u_2, v_2 being nonzero. Hence, the solutions are:

¹⁷Yazaki's articles are unfortunately not in English. For a brief English account of his articles, see (Brown 2002).

$$u(\mathbf{p}) = \begin{pmatrix} a_1 e^{i\delta_1(\mathbf{p})} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad v(\mathbf{p}) = \begin{pmatrix} a_1 e^{-i\delta_1(\mathbf{p})} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad (7.35)$$

and

$$u(\mathbf{p}) = \begin{pmatrix} 0 \\ a_2 e^{i\delta_2(\mathbf{p})} \\ 0 \\ 0 \end{pmatrix}, \quad v(\mathbf{p}) = \begin{pmatrix} 0 \\ a_2 e^{-i\delta_2(\mathbf{p})} \\ 0 \\ 0 \end{pmatrix}, \quad (7.36)$$

where $a_1^2 = a_2^2 = (2\pi h)^{-3}$, and δ_1, δ_2 are phases that can be chosen freely.

For the final state, since the electron is not at rest, the equations are not so simple. As Yazaki (1992) shows, they solved this by introducing a contact transformation:

$$S = \alpha + i\beta\rho_2(\boldsymbol{\sigma}\mathbf{p}), \quad S^{-1} = \alpha - i\beta\rho_2(\boldsymbol{\sigma}\mathbf{p}), \quad (7.37)$$

$$\alpha^2 + \beta^2\mathbf{p}^2 = 1,$$

where, α, β are given by the following equations:

$$\alpha = \sqrt{\frac{m^* + m}{2m^*}}, \quad \beta = \sqrt{\frac{m^* - m}{2m^*\mathbf{p}^2}}. \quad (7.38)$$

They define $u^*(\mathbf{p})$ and $v^*(\mathbf{p})$ by:

$$u(\mathbf{p}) = u^*(\mathbf{p})S(\mathbf{p}), \quad v(\mathbf{p}) = S^{-1}(\mathbf{p})v^*(\mathbf{p}). \quad (7.39)$$

Then the Dirac equations take the same form as for $\mathbf{p} = 0$:

$$u^*(1 + \rho_3) = 0, \quad (1 + \rho_3)v^* = 0. \quad (7.40)$$

In the case of the final state of the Compton effect, Klein and Nishina claim both u_1^*, v_1^* and u_2^*, v_2^* must be considered finite (Klein and Y. Nishina 1929, 864).

Hence, it appears the wave functions for the final state have both components, namely:

$$\begin{aligned}
 u^*(\mathbf{p}') &= \begin{pmatrix} a_1 e^{i\delta_1(\mathbf{p}')} \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 e^{i\delta_2(\mathbf{p}')} \\ 0 \\ 0 \end{pmatrix}, \\
 v^*(\mathbf{p}') &= \begin{pmatrix} a_1 e^{-i\delta_1(\mathbf{p}')} \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 e^{-i\delta_2(\mathbf{p}')} \\ 0 \\ 0 \end{pmatrix}.
 \end{aligned} \tag{7.41}$$

In other words, for the final state of the electron, they combined (or superposed) the two independent solutions u^* , v^* with variable phase factors. Thus, behind the scenes, Nishina, with Klein, went back to his old friend from his electrical engineering student years—the principle of superposition—to carry out his first and most important work in quantum mechanics.

For the value of $u(\mathbf{p}')\boldsymbol{\sigma}v(\mathbf{p}')$, Klein and Nishina employed physical considerations and direct calculation. They claim the final state “should contain the two independent solutions with equal strength,” the average of $u(\mathbf{p}')\boldsymbol{\sigma}v(\mathbf{p}')$ over phases must vanish (Klein and Y. Nishina 1929, 866). Hence the expectation value of $\boldsymbol{\sigma}$, the magnetic moment of the electron, is zero. The same value can be derived through straightforward calculation by plugging in the expression in eq. (7.41), and averaging over δ_1 and δ_2 . As for $u^*(\mathbf{p}')v^*(\mathbf{p}')$, one can get $2(2\pi\hbar)^3$ by inserting the expression in eq. (7.41).

Klein and Nishina did not clearly show how they justified this procedure for calculating a statistical average. This procedure of summing over phases did not appear in the previous theory of the Compton effect, such as Gordon’s (1926). Although Klein and Nishina did not state it explicitly, they probably took into consideration that different phases give different directions of spin, not unlike Steinmetz’s alternating current theory, where different imaginary numbers give different directions of vectors in vector diagrams. Summation over phases meant summation over the direction of the magnetic moment. At least for the initial state, the physical meaning was then clear. As for the final state, the situation was somewhat different, because the physical meanings of $u^*(\mathbf{p})$ and $v^*(\mathbf{p})$ were not transparent. They probably justified their assumptions by showing that $u(\mathbf{p}')\bar{\boldsymbol{\sigma}}v(\mathbf{p}')$ vanishes by direct calculation. Since they were considering unpo-

larized light as the incoming radiation, the average of the magnetic moment in the final state should also be zero.

Since summation over phases eliminates non-diagonal elements and sums diagonal elements, this procedure can be considered equivalent to taking a trace. Because the density matrix in this case is proportional to the unit matrix, this procedure agrees with a quantum statistical calculation for a mixed state. A mathematical theory about this procedure of quantum statistics had already been presented by John von Neumann in 1927, see (von Neumann 1927). Klein and Nishina were likely unaware of the relevance of Neumann's paper but managed to do a mathematically-equivalent calculation relying on physical considerations (if they had noticed it, their calculation would have been much different; taking a trace from the beginning would have made the calculation much shorter and easier).

7.5 Conclusion

As I wrote at the beginning of this paper, I am not claiming any deterministic, causal connections. It would be ridiculous to claim that Nishina worked on quantum mechanics because he studied electrical engineering. Nor do I claim that Nishina took a certain research style in quantum mechanics different from others because of his electrical engineering background. What I claim is that quantum mechanics, not only its experiments but also its theoretical research, might not be as disconnected from other fields of investigation, such as engineering, as we might assume. In terms of theoretical practices, there are some similarities between alternating current theory and quantum physics, at least in the way they were experienced by a person like Nishina.

Therefore, the fact that Nishina was originally trained in electrical engineering does not mean that Nishina started learning quantum mechanics "from scratch." Nishina's engineering training probably prepared him for theoretical research in quantum mechanics to some extent. At least in this limited sense, it seems the institutional and pedagogical developments in engineering helped introduce theoretical research in quantum mechanics into Japan.

Acknowledgements

This work is partially supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Grant-in-Aid for Scientific Research C (Kakuhhi) # 21500976.

References

- Aichi, Keiichi (1923). *Denshi no jijoden: Tsūzoku kagaku kōgi* (The autobiography of an electron: A popular science account). Tokyo: Shōkabō.
- Bartholomew, James R. (1989). *The Formation of Science in Japan: Building a Research Tradition*. New Haven and London: Yale University Press.
- Brown, Laurie M. (2002). The Compton Effect as One Path to QED. *Studies in History and Philosophy of Modern Physics* 33: 211–249.
- Dalitz, Richard H. (1990). Another Side to Paul Dirac. In: *Reminiscences about a Great Physicist: Paul Adrien Maurice Dirac*. Ed. Behram N. Kursunoglu and Eugene P. Wigner. Cambridge: Cambridge University Press. 69–92.
- Dirac, Paul A. M. (1926). Relativistic Quantum Mechanics with an Application to Compton Scattering. *Proceedings of the Royal Society A* 111: 405–423.
- (1928). The Quantum Theory of the Electron. *Proceedings of the Royal Society A* 117: 610–624.
- (1930). *The Principles of Quantum Mechanics*. Oxford: The Clarendon Press.
- (1936). *Ryōshi rikigaku* (Quantum mechanics). Tokyo: Iwanami Shoten.
- Fortescue, Charles L. (1918). Method of Symmetrical Co-ordinates Applied to the Solution of Polyphase Networks. *AIEE Transactions* 37, part II: 1027–1140.
- Galison, Peter L. (1997). *Image and Logic: A Material Culture of Microphysics*. Chicago: The University of Chicago Press.
- (2000). The Suppressed Drawing: Paul Dirac's Hidden Geometry. *Representation* 72: 145–166.
- (2003). *Einstein's Clocks, Poincaré's Maps: Empires of Time*. New York: W. W. Norton & Company.
- Gibson, Charles R. (1911). *The Autobiography of an Electron: Wherein the Scientific Ideas of the Present Time Are Explained in an Interesting and Novel Fashion*. Philadelphia, London: J.B. Lippincott Company, Seeley & Co., Limited.
- Gordon, Walter (1926). Der Comptoneffekt nach der Schrödingerschen Theorie. *Zeitschrift für Physik* 40: 117–133.
- Hirosige, Tetu (1973). *Kagaku no shakaishi: Kindai Nihon no kagaku taisei* (Social history of science: Scientific regime in modern Japan). Tokyo: Chūō Kōron Sha.
- Hō, Hidetarō (1912). *Kōryū riron* (Alternating current theory). Tokyo: Maruzen Kabushiki Kaisha.

- (1922). Sōdensen no secchi to chōjō no ri (Grounding of power line and the principle of superposition). *Denki Gakkai zasshi* (Journal of the Institute of Electrical Engineers of Japan) 42: 193–198.
- (1923). *Hadō shindō oyobi hirai* (Wave, vibration, and arrestoring). Tokyo: Maruzen.
- Ito, Kenji (2002). *Making Sense of Ryōshiron* (Quantum Theory): Introduction of Quantum Physics into Japan, 1920–1940. PhD thesis. Harvard University.
- Kim, Dong-Won (2007). *Yoshio Nishina: Father of Modern Physics in Japan*. New York and London: Taylor & Francis.
- Klein, Oskar and Yoshio Nishina (1929). Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac. *Zeitschrift für Physik* 52: 853–868.
- Kline, Ronald R. (1992). *Steinmetz: Engineer and Socialist*. Baltimore: Johns Hopkins University Press.
- Kragh, Helge (1990). *Dirac: A Scientific Biography*. Cambridge: Cambridge University Press.
- Layton, Edwin (1971). Mirror Image Twins: The Communities of Science and Technology in Nineteenth Century America. *Technology and Culture* 12: 562–580.
- Mizuno, Toshinojō (1912). *Denshiron* (The electron theory). Tokyo: Maruzen Kabushiki Kaisha.
- (1918). *Densi no katsudō* (Activities of the electron). Tokyo: Maruzen Kabushiki Kaisha.
- Nagaoka, Hantarō (1912). *Genkon no denkgaku* (Studies of electricity today). Tokyo: Kōdōkan.
- von Neumann, John (1927). Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen* 1: 245–272.
- Nishina, Akira (1975). Memories of Dr. Nishina Yoshio. *Takahashigawa* 32: 4–19.
- Nishina, Yoshio (1918). *Effects of Unbalanced Single-Phase Loads on Poly-Phase Machinery and Phase Balancing* (B. A. Thesis). Tokyo Imperial University.
- (1946). Watashi wa nani o yondaka (What I have read). In: *Genshiryoku to watashi: Nishina Yoshio hakase ikōshū* (Atomic power and I: Posthumous collection of Dr. Nishina Yoshio's writings). Tokyo: Gakufū Shoten. 224–230.
- Pyenson, Lewis (1982). Audacious Enterprise: The Einsteins and Electrotechnology in Late Nineteenth-Century Munich. *Historical Studies in the Physical Sciences* 12: 373–392.

- Pyenson, Lewis (1985). *The Young Einstein: The Advent of Relativity*. Bristol: Hilger.
- Steinmetz, Charles Proteus (1893). Complex Quantities and Their Use in Electrical Engineering. *Proceedings of the International Electrical Congress*: 37–53.
- Stuewer, Roger H. (1975). *The Compton Effect: Turning Point in Physics*. New York: Science History Publications.
- Suchet, Charles (1949). Léon Charles Thévenin (1857–1926). *Electrical Engineering* 68: 843–844.
- Sugiura, Yoshikatsu (1927). Über die Eigenschaften des Wasserstoffmoleküls im Grundzustande. *Zeitschrift für Physik* 45: 484–494.
- Yagi, Eri (1974). Hantaro NAGAOKA. In: *Dictionary of Scientific Biography*. 9. New York: Charles Scribner's Sons. 606–607.
- Yazaki, Y. (1992). Klein-Nishina kōshiki dōshutsu no katei: Riken no Nishina shiryō o chūshin ni (How was the Klein-Nishina formula derived?: Based mainly on the source materials of Y. Nishina in RIKEN). *Kagakushi Kenkyu* 31. Part I: 81–91. Part II: 129–137.
- Yukawa, Hideki (1935). On the Interaction of Elementary Particles I. *Proceedings of the Physico-Mathematical Society of Japan* 17(3): 48–57.

Chapter 8

The Origins of Maria Göppert's Dissertation on Two-Photon Quantum Transitions at Göttingen's Institutes of Physics 1920-1933

Barry R. Masters

In the 1920s, the University of Göttingen was a nexus of theoretical and experimental physics, as well as mathematics (Hund 1983; 1987; Jungnickel and McCormmach 1986; Rupke 2002). In this case study of Maria Göppert, a doctoral student under the tutelage of the theoretical physicist Max Born, we see the influence of the experimental groups in James Franck's physics institute and the role of Paul Dirac's scientific papers on her dissertation research (Kamp et al. 1983).

Göppert's dissertation work on the theory of two-photon¹ transitions of atoms is significant in the history of quantum mechanics. It not only provided a theoretical foundation for the experimental findings that were the origin of her research, but more importantly, it served as the basis of nonlinear optics (Boyd 2008; Masters and So 2004).

This paper seeks to answer the following historical questions. Why did Göppert choose to study at the University of Göttingen? How did she become a student of Born? What influenced her selection of a research problem for her dissertation work? What theoretical techniques did she use in her research? And finally, why did Göppert, and not others, calculate the probability for two-photon transitions?²

¹Göppert wrote "two light quanta" (*zwei Lichtquanten*) in her publications. In 1926, Gilbert N. Lewis coined the term "photon." The modern usage is "two-photon" or "multi-photon" processes (Masters and So 2008).

²Göppert's theory predicted two-photon absorption and emission processes of atoms in her 1931 Göttingen dissertation. Since double or two-photon transitions are related to the square of the intensity of light, they are extremely improbable with the light sources available prior to the development of the laser (Maiman 1960). In honor of her important discovery, the two-photon absorption cross-section unit, GM, is given the name Göppert-Mayer.

This paper examines the sources and reconstructs Göppert's 1931 Göttingen dissertation to answer these questions and improve our understanding of the history of quantum mechanics.³

In addition, the methods used in my research include an analysis of the following sources: Dirac's 1927 paper on the emission and absorption of radiation written during his visit to Bohr in Copenhagen, and Dirac's 1927 dispersion paper that he wrote in Göttingen while he was a visitor in Born's Institute of Theoretical Physics.⁴

Other sources include experimental studies described in the 1928 papers of Otto Oldenberg and those of Franck from Göttingen's physics institute, and Göppert's 1929 paper and her 1931 Göttingen dissertation, as well as her contributed chapter on dispersion theory for Born and Pascual Jordan's *Elementare Quantenmechanik* (M. Born and Jordan 1930). Part of my methodology was a comparison of all of these sources and an analysis of which theoretical techniques Göppert took from Dirac's publications and which were her original contributions.

The case study of Göppert's dissertation illustrates the synergistic interaction between Franck's experimental group and Born's theoretical group at the Göttingen physics institutes in the 1920s. Furthermore, this paper examines the role of visitors to the physics institutes. In particular, I compare the influence of Dirac on two of Born's contemporary doctoral students: Göppert, the focus of this paper, and Victor Weisskopf. I posit that the combination of her mathematical expertise, physical insight, and the selection of a research topic associated with two-photon processes, together with a deep understanding of the theoretical techniques used in Dirac's dispersion paper, all contributed to Göppert's successful theoretical prediction and calculation of the probabilities of two-photon processes.

8.1 Physics in Germany at the Beginning of the Twentieth Century and the Development of the Institutes of Physics at the University of Göttingen

Theoretical physics began a strong tradition in Göttingen, beginning with Wolde-
mar Voigt. In 1883, Voigt became a full professor for theoretical (mathematical)

³I use her maiden name, Göppert, as she did prior to 1931. The authorship of her 28 October 1929 paper in *Die Naturwissenschaften* is listed as Göppert. On 18 January 1930, she married Joseph Mayer, an American Rockefeller Fellow who was an assistant to James Franck. In March, she completed her final examination and the Göttingen dissertation. Afterwards, she signed her Göttingen dissertation of 7 December 1930, which was published in *Annalen der Physik* (Leipzig) in 1931, with the name Göppert-Mayer. After 1931, I refer to her married name as she did in her publications.

⁴Dirac visited Bohr in Copenhagen from September 1926 through February 1927, when he wrote his paper on transformation theory as well as his paper on the emission and the absorption of radiation by matter (Dirac 1927a; 1927b). Following that visit to Bohr, Dirac remained in Göttingen from February through the end of June 1927 (Bacciagaluppi and Valentini 2009, 84; Kragh 1990, 43). During this time he wrote his paper on dispersion theory (Dirac 1927c).

physics, as well as the director of the mathematical physics institute; in addition, he was made the co-director of the physical department of the mathematical-physical seminar (Jungnickel and McCormmach 1986, 115; Hund 1987, 30).

At the beginning of the twentieth century, many physics departments in German universities were dedicated to experimental work, however, there were also institutes of theoretical physics and full (ordinary) professors of theoretical physics (Eckert 2001; Heilbron 1967; Hund 1987; Jungnickel and McCormmach 1986; Rupke 2002; Seth 2010). In 1914, Peter Debye joined the university and became director of the institute for theoretical physics.⁵

Debye's lectures, during 1917–1918 for example, included the following topics: new research in quantum theory, optics for physicists and mathematicians. Also, Debye and David Hilbert initiated the joint seminar "On the Structure of Matter" (Schirmmacher 2003).

When Debye left Göttingen in 1920, his replacement was the theoretician, Born. Prior to his departure, Debye collaborated with Hilbert, and with the arrival of Born in Göttingen, he was offered the directorship of the mathematical department of the physics institute, formerly held by Debye. But Born exploited the confusion of the postwar ministry of culture and convinced them to divide Debye's former department into two new departments, one for theory and another for experimental research (Hund 1987; Jungnickel and McCormmach 1986, 357). Born arranged for his friend Franck, an experimentalist, to become an ordinary professor in the adjacent institute. Franck's research was centered on experimental atomic physics.

Shortly after, in 1921, three institutes were created: Robert Pohl directed the First Institute of Physics, Franck directed the Second Institute of Physics, and Born directed the Institute for Theoretical Physics. Pohl, who was made an ordinary professor of physics in 1920, was one of the founders of experimental solid state physics (Hund 1987; Jungnickel and McCormmach 1986).⁶

The scientific collaboration between Born and Franck extended beyond friendship; it was bilaterally synergistic. Born and Franck were friends and colleagues who closely interacted, and their physics institutes were located in the same building (Hund 1983; 1987).⁷

⁵As described by Jungnickel and McCormmach (1986, 301), after Hilbert heard Debye's lecture, he decided to have Debye join the faculty in Göttingen. In order to have Debye head an institute, Voigt agreed to turn the directorship of the institute over to Debye with the agreement that Voigt would still share the institute and the teaching of theoretical physics.

⁶These institutes were in the main physics building on Bunsenstrasse, which was built in 1905. The Mathematics Institute was next door to the Physics Institutes. Ludwig Prandtl headed the Institute for aerodynamics research that was on the opposite side of Bunsenstrasse (Hentschel 1999; Hund 1987).

⁷The life-long friendship between Born and Franck began when they were both students at the University of Heidelberg and met in a mathematics class (Greenspan 2005, 24–25; Lemmerich 2007, 24).

The productive synergism between Born and Franck's groups is further described in a recent biography of Franck (Lemmerich 2007), and is expressed by Gyeong-Soon Im:

After Born moved from Frankfurt to Göttingen in 1921, he conducted a research program in quantum theory with a distinct style: he selected as simple physical problems as possible for which there already existed extensive empirical evidence. He then sought general solutions to these problems with the help of rigorous mathematical techniques. Since Franck systematically performed experiments associated with the quantum theory, he accumulated *inter alia* many observational results on quantum excitation during collision processes, including ionization energies of atoms and molecules. Born's close collaboration with Franck was well suited to his research style: a formal and mathematical description of nature based upon plentiful observational data. (Im 1995, 74)

8.2 Göppert as a University and Doctoral Student

What influences impact the development of a scientist? Is it family, friends, neighbors, teachers and mentors? Is it primary education and university education? In Göppert's scholarly development, we can trace multiple examples of these influences (Johnson 1999; 2004; Masters 2000; 2008a; McGrayne 1993).

Göppert was born in 1906, the only child of Friedrich Göppert and his wife, Maria. In 1910, the family moved to Göttingen, where Friedrich Göppert obtained a position as professor. Göppert was proximate to this center of intellectual activity and her family was physically and socially connected to many of Göttingen's great intellectuals. For example, the Göpperts lived next door to Hilbert and they were personal friends. In 1921, Göttingen brought two new physicists to the university, first Born and then Franck (H. Born and M. Born 1962). The Göpperts became and remained their good friends. Other family friends included Richard Courant, Edmund Landau, and Hermann Weyl, who were members of the mathematics faculty. Göppert's own close friends included Born, Max Delbrück, Franck, Linus Pauling, Hertha Spöner, Leo Szilard, and Victor Weisskopf (McGrayne 1993; Sachs 1982).

In this section, I explore some of the plausible reasons why the young Göppert chose to study and then to perform her doctoral research in Göttingen, and after she earned her doctorate, why she chose to fulfill her professional life as a physicist outside of Germany.

There are three reasons why Göppert chose to study at Göttingen. First, Göppert fostered a strong interest in mathematics. Göttingen was home to

Germany's leading mathematics department (Jungnickel and McCormach 1986; McGrayne 1993; Rowe 1989). Another important attraction for her was that Hilbert, Richard Courant and Carl Runge were interested in both physics and mathematical physics, which coincided with her joint interests (Hund 1987; Schirmmacher 2003).⁸

Second, Göttingen and its university had a long and famous standing in liberalism and freedom from censorship (Georg-August Universität 2011a; 2011b).⁹

Third, the University of Göttingen was home to some outstanding women, and that set a precedent and provided role models for Göppert to pursue her graduate work at that institution. For example, Emmy Noether came to Göttingen in 1916 at the invitation of both Hilbert and Felix Klein, remaining there until 1934. It was the efforts of Hilbert, a strong proponent of women's educational rights, that helped Noether undergo her habilitation¹⁰ and thus gave her the right to lecture at a university. Finally in 1922, with a doctorate earned thirteen years previously, Noether was made a *Privatdozent*; now she could legally teach in the university under her own name. In the course of her second habilitation lecture, she presented her work on invariant forms in mathematics, or what is now known as "Noether's Theorem." In 1922, she did not receive the status of a civil servant (*Beamtin*) and she had the following title: *nichtbeamteter außerordentlicher Professor*. This was basically a volunteer professorship that had no university salary, although she received student's fees as a *Privatdozent*; in Göttingen she was never made a full professor (McGrayne 1993, 175–200).

Hertha Sponer, who was a friend of Göppert's, worked on molecular spectroscopy and photochemistry in Franck's laboratory during the time that Göppert was a graduate student with Born (Lemmerich 2007). Sponer had studied at the University of Tübingen, but after one year in Tübingen she moved to the University of Göttingen where she was a doctoral student with her supervisor, Debye. She graduated in 1921 with a doctorate; this was a very significant achievement since she was part of a small group of women who obtained a doctorate in physics at a German university in addition to her habilitation (obtained under Franck's supervision in 1925).

⁸When Born studied for both his doctorate and his habilitation in Göttingen, he was influenced by famous scientists and mathematicians including Klein, Hilbert, Hermann Minkowski, Runge, Karl Schwarzschild (full professor of astronomy and director of the observatory), and Voigt.

⁹Consistent with this liberal spirit is the story of the Göttingen Seven (Lampe 2002; Marchand 1996). In 1837, Dahlmann and the other six protesters demonstrated against any alteration of the constitution of the Kingdom of Hanover. They were all dismissed from the university.

¹⁰In Germany and other European countries, before a person with a research doctorate could teach in the university, they had to obtain a habilitation which gave them this right. Habilitation research differs from the research doctorate; while the research doctorate is performed under the supervision of a guiding professor, habilitation research is based on independent scholarly work. In general, the level of scholarship for the habilitation is significantly higher as compared to the research doctorate.

Göppert's first plan was to study mathematics, which was her strongest interest. Therefore, in 1924 Göppert began her studies in mathematics at the University of Göttingen. Shortly after beginning her studies, Born asked her to join the physics seminar, and her interest in the newly-evolving area of quantum mechanics—coupled with her training and interest in mathematics—influenced her move from mathematics to physics (Greenspan 2005). By the time she became a graduate student under Born, she was already adept in mathematics and that helped her with the new quantum mechanics. Nevertheless, Franck's strong experimental approach remained with her during her doctoral research, as well as in her later works in the field of nuclear physics (Masters 2000, 38–41).

Göppert worked in Born's institute from 1924 until she graduated in 1931. After a period in the United States, she returned to Göttingen from Baltimore in the summers to continue to work with Born, at least until 1933. Together they published a major review on the dynamic lattice theory of crystals, which appeared in the 1933 edition of the prestigious *Handbuch der Physik* (M. Born and Göppert-Mayer 1933, 623–794).

Göppert wanted a career in science as a full professor. She recognized that such an aspiration had a very low probability if she remained in Germany. This followed from her knowledge that neither Noether, nor Lise Meitner, nor her good friend Sponer ever achieved a full university professorship in Germany (McGrayne 1993, 184–191).¹¹

On 1 April 1930, Göppert-Mayer and her husband moved to Baltimore, Maryland, where Mayer held an assistant professorship at Johns Hopkins University. Her summer visits to Born in Göttingen ceased in 1933. Social and political realities in Germany resulted in forced migrations of many academics. Following the 7 April 1933 enactment of the Law for the Restoration of the Professional Civil Service, almost all non-Aryan civil servants (including tenured university professors) were removed from their positions in Nazi Germany. Born left Göttingen to take a position in the United Kingdom (H. Born and M. Born 1962). Franck, the recipient of the 1925 Nobel Prize in physics, quit his university professorship in protest against Nazi racial policies and emigrated to the United States (Lemmerich 2007).

¹¹That plan did not materialize in the United States for many decades. In the United States, Göppert-Mayer spent many years working as a volunteer in the physics departments of Johns Hopkins University, Columbia University, and as a voluntary associate professor and later as a voluntary professor at the University of Chicago (McGrayne 1993). At the same time, her husband working at the same institutions moved up the academic ranks to full professor. In 1956, she was elected to the National Academy of Sciences. Finally in 1960, she accepted a full professorship with pay at the University of California, San Diego. In 1963, Eugene Wigner, Göppert-Mayer and Johannes H. D. Jensen shared the Nobel Prize in Physics (Göppert-Mayer 1948; 1955; McGrayne 1993).

8.3 What Was the Role of Paul Dirac in Göppert's Dissertation Research?

8.3.1 Dirac's 1927 Publications

Three of Dirac's 1927 publications had a great influence on Göppert-Mayer's dissertation work (Dirac 1927a; 1927b; 1927c). She directly acknowledged Dirac's contribution to her research in her 1931 Göttingen dissertation and in her chapter on dispersion theory in *Elementare Quantenmechanik* (M. Born and Jordan 1930, 404–408; Dirac 1927b; 1927c). In the latter book chapter, a footnote states that the considerations in her section follow from Dirac's paper on the quantum theory of the emission and absorption of radiation and from Dirac's paper on the quantum theory of dispersion (Dirac 1927b; 1927c; Göppert-Mayer 1930). Furthermore, in their preface to *Elementare Quantenmechanik*, Born and Jordan state that Göppert contributed sections on Dirac's theory of emission, absorption and dispersion (M. Born and Jordan 1930, VII–VIII). In this section, I examine Dirac's contributions and evidence of his influence on Göppert's dissertation research.

8.3.2 Dirac's Paper on the Emission and Absorption of Radiation (Dirac 1927b)

Dirac states that the mathematical development in this paper on emission and absorption of radiation follows from his previous paper on the general transformation theory of quantum matrices (Dirac 1927a). In this paper, Dirac proceeds as follows: he considers an atom interacting with a field of radiation, which is confined to a cavity, to have a discrete set of degrees of freedom (Dirac 1927b). Dirac considers a finite cavity to enclose the radiation to establish a relationship between the number of light quanta per stationary state and the intensity of the radiation. He restricts the treatment to the non-relativistic case. In the absence of interaction between the atom and the radiation, the Hamiltonian consists of two terms: the field and the atom. In the presence of interaction, a third term from classical theory would be added to the Hamiltonian. From this formulation, Dirac derives the "correct" results for the action of the radiation and the atom on each other. Thus he derives the "correct" laws for the emission and the absorption of radiation and the "correct" values for Einstein's A and B coefficients (Einstein 1916).

8.3.3 Dirac's Paper on the Quantum Theory of Dispersion (Dirac 1927c)

Initially, Dirac explains that while the new quantum mechanics uses analogies found in classical theory, it cannot be applied to a class of problems where the analogies are obscure, for example, the problems of resonance radiation and the widths of spectral lines. Dirac proposes that the radiation field can be treated as a dynamical system composed of harmonic components with energies and phases, where each one is a harmonic oscillator. The interaction of this field with an atom can be described by a Hamiltonian function. Dirac then requires the use of perturbation methods to solve the Schrödinger equation. Dirac shows through the use of second-order perturbation theory that a double process can occur: first a transition from the initial state to an intermediate state, and then a transition from the intermediate state to the final state. Each of these processes does not conserve energy, but energy is conserved in the total process consisting of the two transitions, for example, from the initial to the final state in a double process. Dirac resolves the electromagnetic field into its components of plane-polarized, propagating waves, with each component of a definite frequency, direction, state of polarization. He confines the radiation to a cavity to discretize the number of components. Then, he formulates the Hamiltonian function in terms of a vector potential that describes the interaction of the field with the atom, which he considers a single electron in an electrostatic field with a potential. For the case of resonance, Dirac assumes a range of frequencies of the incident radiation, and he calculates the equations for the probability of the emission and the absorption of light quanta.

I now elucidate some of the details of his paper on dispersion theory (Dirac 1927c).

1. The basic idea of Dirac's theory of radiation is to describe the total system of radiation and the atom as the sum of three terms: the first term represents the energy of the atom, the second is the electromagnetic energy of the radiation field, and the third term is the interaction energy of the atom and the radiation field. In the absence of the third term, the atom could neither absorb nor emit radiation. Initially, Dirac decided not to consider the radiation in infinite space, but to represent the radiation as confined to a cavity, of finite volume (V) and with perfectly reflecting walls. Later, the cavity would expand to become infinite, and that would represent the radiation in free space. Then, the oscillations of the confined electromagnetic field are represented as the superposition of a finite number of fundamental vibrations; each one corresponds to a system of standing waves. The electromagnetic field of a monochromatic, plane standing wave in the cavity can be described by a vector potential. Next, the Hamiltonian of the atom

and the radiation field are described. The electromagnetic energy of the radiation field can be shown to have the same Hamiltonian as a system of uncoupled harmonic oscillators. The Hamiltonian for the total system of atom and radiation field is the sum of three terms: for the radiation field, for the atom, and the term of the interaction of the radiation and the atom. The last interaction term is the coupling term for the atom and the radiation field. Then, Dirac develops his time-dependent perturbation theory to calculate the probabilities of transitions of energy for the atom and for the radiation field. This is studied for a variety of cases: absorption, emission and induced emission.

2. Dirac uses a semiclassical treatment; the electromagnetic field is treated classically and the atoms with which the field interacts are treated quantum mechanically. The semiclassical approach “correctly” describes absorption and induced emission, but it fails to “correctly” describe the influence of the atoms on the electromagnetic field.
3. In the mathematical description of a plane linear-polarized monochromatic wave that is resolved into its Fourier components, there appears the frequency of the wave, an amplitude which is a complex vector, and two complex components of the wave amplitude; they are each multiplied by a unit polarization vector, which represents the two independent states of linear polarization.
4. To make the number of degrees of freedom discrete, Dirac assumed that the radiation field is confined to a cavity. According to Dirac's theory, radiation in a cavity can be described by giving the amplitude of each standing wave at a particular time; therefore, the amplitude can be considered a coordinate that follows the laws of quantum mechanics. In his theory of the interaction of atoms and radiation, he calculated the probabilities of both induced emission and spontaneous emission (no radiation present). In addition, it provided a new theory for dispersion and light scattering.
5. In the treatment of an atom and its interaction with a radiation field, the process of the absorption of a photon by an atom involves the increase in the energy of the atom by a quantum of energy, and the decrease of the harmonic oscillators comprising the radiation field by a quantum of energy. The combined energies of the electron and the radiation oscillators follow the law of conservation of energy.
6. Dirac's perturbation theory included two cases: time-dependent and time-independent perturbations. An example of the former case is the calculation of absorption of light or the induced emission of light by an atom in a radiation field.

7. Dirac's time-dependent perturbation theory can be used to calculate transitions between discrete energy levels, as well as in physical systems with continuous energy levels. For example, in particle collisions, the eigenfunctions of the free particles, that is, electrons colliding with atoms, are described as plane waves, and the energy of the particles is not quantized, but can take different positive values. If the particles are now confined to a box, the eigenvalue or the energy of the particle is now quantized. As the size of the box increases to infinity, the free particle eigenfunctions and energy eigenvalues approach those of the free particle. For a free particle in a box, the quantized energy eigenvalues can be calculated by perturbation theory for discrete energy levels. Then the size of the box is increased to infinity, and the result obtained is valid for continuous energy levels.
8. Raman scattering is another example of a two-photon process. A photon is absorbed and another photon is emitted; the atom makes a transition from the initial to the final state. The energy difference between the initial and the final states is equal to the energy difference of the two photons. Second-order perturbation can be used to calculate the Raman transition probabilities, which are the square of the transition amplitudes for the process. Time-dependent perturbation theory is required to calculate the rates of the transitions.
9. Dirac states that the exact interaction energy of the field and the atom is too complicated, therefore he uses the dipole energy. That approximation results in a divergent series that appears in the calculation. In his calculation of dispersion and resonance radiation, there is no divergent series, but when he attempts to calculate the breadth of a spectral line, a divergent series appears.

As we shall see in the following section, many of these aspects of Dirac's 1927 publications were directly incorporated into Göppert's Göttingen dissertation.

8.4 Reconstruction of Göppert's Göttingen Dissertation

The origin of Göppert's dissertation research were two publications by Oldenberg and Franck on electronic excitation of atoms due to inelastic collisions with electrons and the subsequent luminescence (Oldenberg 1928; Franck 1928). The significance of these experiments is that they demonstrated the discrete energy levels of atoms. The inelastic collisions of electrons and atoms can result in the transfer of energy to the atoms and can excite the atoms without ionizing them. These experiments were conducted at the Second Institute of Physics, and they

provided Göppert with an opportunity to seek a theoretical explanation for these purported two-photon findings.

Next, I review Oldenberg's and Franck's 1928 papers (Franck 1928; Oldenberg 1928). The basis of Oldenberg's experiments was the question: could an atom become excited (its electrons are raised to higher energy states than the ground electronic state) through a single act of collision between electrons and an incident light field? He also discussed the concept that two light quanta can work together in one elementary act to excite an atom or molecule, for example, the Smekal-Raman effect (Smekal 1923).

Oldenberg produced experimental evidence on the broadening of resonance lines of mercury atoms when the excited atoms collide with slow particles multiple times. He showed that the excitation energy of the mercury atoms can be transferred as kinetic energy to the particles, and the difference frequency is radiated as light. The publication contains an equation that shows how two light quanta, with two different frequencies, can work together in a single elementary act to excite an atom (double absorption or two-photon absorption). In the second section of Göppert's Göttingen dissertation, she constructed the theory of "the working together of light and collisions [electrons] in one elementary act" (Göppert-Mayer 1931, 288). Her theoretical analysis agrees with Oldenberg's previous experimental results (Oldenberg 1928).

Franck focused his research program on atomic physics and spectroscopy.¹² In Göttingen, Franck continued to experiment with collisions of fast electrons and atoms. He explored the effect of the velocity of colliding electrons on the spectral lines of atoms. He studied the ionization of atoms due to collisions with slow and fast electrons and the subsequent luminescence that was observed. According to Franck (1928), this process is due to the recombination of ions and electrons.

Göppert worked on the theory of atom-photon interactions. Building on Hans Kramers and Werner Heisenberg's dispersion theory, and Dirac's time-dependent perturbation theory, she developed analytical expressions of the transition probability for multi-photon absorption and stimulated emission, as well as Raman scattering processes (Kramers and Heisenberg 1925).¹³ Note that in her 1929 paper, she stated that Dirac's dispersion theory described not only the Ra-

¹²In 1926, Franck and Gustav Hertz received their Nobel prize in physics. Franck was awarded the Nobel for his work during the 1912–1914 period, specifically the Franck-Hertz experiment based on the inelastic scattering of electrons by mercury atoms in the gas phase. Franck and Hertz demonstrated that a collision between an electron and an atom can result in a transition of the atom from its ground state to a stationary state of higher energy; in the process, the electron loses an equal amount of energy (Franck and Hertz 1914). Their experiment provided an important confirmation of the quantization of an atom's energy levels.

¹³To obtain a sense of the physical theories and techniques that were in use at the time of Göppert's graduate research, I recommend that the reader examine *Elementare Quantenmechanik* (M. Born and Jordan 1930).

man effect but also the reverse process in which two photons act together in a single elementary event to promote an atom from the ground state to an excited state (Dirac 1927c; Göppert 1929).

What theoretical and mathematical techniques did Göppert use in her dissertation research that followed Dirac's previous publication (1927c)? To address this question, I surveyed physics and mathematics books published in the 1920s. In particular, the book series edited by Born and Franck entitled *Struktur der Materie in Einzeldarstellungen* (1925), and a book by Franck and Jordan (1926) on collisions. Although it is likely that she read these volumes as a student, I refer to Göppert's publications in which she explicitly cites experimental works from the Franck group and Dirac's theoretical papers as major influences on her dissertation research. In particular, a careful analysis of her dissertation reveals four similarities with Dirac's dispersion paper (Dirac 1927c). The first section of her dissertation is concerned with two light quanta working together in one elementary act (Göppert-Mayer 1931, 273–284). The four similarities are listed below in extracts from Göppert's dissertation:

1. With the help of Paul Dirac's dispersion theory, the probability of an analogous Raman effect process is calculated, namely the simultaneous emission of two light quanta. It is shown; that a probability exists for an excited atom to divide its excitation energy into two light quanta [...]. If an atom is irradiated with light of a lower frequency than the frequency associated with an eigenfrequency of the atom, there additionally occurs a stimulated double emission [...]. Kramers and Heisenberg (1925) calculated the probability of this last process in a corresponding manner [273].
2. The reverse process is also considered, namely the case that two light quanta, whose sum of frequencies is equal to the excitation frequency of the atom, work together to excite the atom. It is further investigated how an atom responds to colliding particles, when at the same time it has the possibility of spontaneously emitting light. Oldenberg (1928) experimentally found a broadening of the resonance lines of mercury, when he allowed the excited atoms to collide many times with slow particles [273]. For this process, an equation is derived here that is analogous to the Raman effect or double emission [274]. Finally, in relation to a study by James Franck (1928), an attempt is made to explain the behavior of the intensity of excitation of spectral lines, induced by collision [of atoms] with fast electrons in such a double process [274]. The calculation shows a probability for such a process, the nature of which will be discussed [275].
3. The following calculation is closely associated with the work of Dirac on emission, absorption, and dispersion [275].

4. Let us consider the interaction of an atom with a [electromagnetic] radiation field. To make the number of degrees of freedom countable, think of the radiation contained in a cubic box of volume V , which constrains the light waves to the condition of periodic repetition [standing waves]. Later this box will be assumed to be infinitely large. Such a radiation field is equivalent to a system of uncoupled harmonic oscillators. The radiation can be decomposed in plane, linear polarized waves, let A be the vector potential [...] [276].¹⁴

Perusal of her Göttingen dissertation indicates that Göppert made use of the following assumptions and techniques:¹⁵

1. the confinement of the radiation field in a cavity so that the number of the degrees of freedom can be discrete,
2. the use of the vector potential [277],
3. the description of the total Hamiltonian function consisting of three components: the Hamiltonian of the radiation field (the uncoupled harmonic oscillators), the Hamiltonian of the atom, and the Hamiltonian of the interaction between the atom and the radiation field [277], the electric dipole approximation in which it is assumed that the wavelength of the light is much larger than the atom's diameter, that is the assumption that the electromagnetic field is constant over the atom's diameter [277–278],
4. the use of second-order, time-dependent perturbation theory [278–284],
5. the use of two-photon transitions via virtual intermediate states [278–284], and
6. the “method of variation of constants,” mentioned by Göppert [280].

The state of an atom is represented by an expansion in terms of the unperturbed energy eigenfunctions. The Hamiltonian operator is different from the true Hamiltonian by a very small term, which is the perturbation. The method of variation of constants derives its name from the fact that the constant coefficients used in the expansion of the wave function, in terms of the true energy eigenfunctions, vary with time.

The second part of Göppert's Göttingen dissertation addresses the way light and collisions (electrons) work together in one elementary act (Göppert-Mayer 1931, 284–294). First, she defines the Hamiltonian function of the total system in which the interaction energy is separated into two parts: one term is the interaction of the atom and the radiation, and the second term is the interaction between

¹⁴Page numbers in square brackets refer to Göppert's dissertation, published in *Annalen der Physik* (Göppert-Mayer 1931). The quotations 1–4 here are translated from German by the author (Masters 2010).

¹⁵Page numbers in square brackets in this list of six assumptions and techniques refer to Göppert-Mayer (1931, 277–284).

the atom (nucleus) and the electron, which is approximated by the Coulomb field. The electron waves are enclosed in a cavity with the same conditions as for the radiation: periodic standing waves. In the first case, she assumed only one atom and one electron in the cavity and no radiation; thus, there are only emission processes. She calculated the probabilities for transitions in the state of the atom due to light alone, and performed a similar calculation for the transitions due to electron collisions alone. Then, she used second-order perturbation theory to study how light and collisions work together. The second part of her Göttingen dissertation was stimulated by the experimental results of Franck's research group, and it confirmed many of their findings (Göppert-Mayer 1931, 284–294).

The significance of this careful reconstruction of her Göttingen dissertation, together with a thorough comparison of the two papers that Dirac published in 1927, demonstrates that Göppert not only used and cited Dirac's papers, but the extent to which she incorporated theoretical techniques from those two papers is significant. Previously, this incorporation of Dirac's work into her Göttingen dissertation has either not been described or has been ignored in the literature on the history of quantum mechanics.

8.5 What Was Known and What Did Göppert Contribute in Her Dissertation Research?

Göppert and Weisskopf were contemporary doctoral students under Born. In this section, I compare the influence of Dirac on Göppert's research and the influence of Wigner on Weisskopf's research. My studies of both Göppert's and Weisskopf's Göttingen dissertations raised the question of the level of originality required at that time for a doctoral dissertation. Both dissertations are at approximately the same level of originality. It is important to understand the role of the dissertation and habilitation to put this question of originality into perspective.

Göppert's Göttingen dissertation relied on second-order, time-dependent perturbation theory. Since perturbation theory was a major mathematical technique in her doctoral theoretical research, it is necessary to look into its antecedents. What are the sources of this theory and how did approximation methods from celestial mechanics find a place in quantum mechanics?

The early development of these perturbation techniques derived from problems in astronomy (Masters 2008b, 36–41). To solve three-body problems or n-body problems, a number of techniques were developed. When the Hamiltonian for the exact problem is known, and it differs slightly from the Hamiltonian for the less complex soluble problem, then approximation or perturbation techniques were derived. The fundamental basis of all the perturbation methods is that the solutions of the perturbed system are only slightly different from the solutions

(the integrated form of the equations of motion) of the equations of motion of the unperturbed system that are already integrated. The main mathematical problem to overcome is that when series expansions were used as approximations of a function, they did not always converge or sum to a finite term; in many cases they diverged to infinity (M. Born 1924; 1925).

In the winter semester of 1922–1923, Born arranged a course on perturbation theory at his institute in Göttingen. In 1922, Paul Epstein independently developed his form of perturbation theory with applications to quantum mechanics (Epstein 1922a; 1922b; 1922c). Earlier in 1916, Epstein developed a perturbation method to treat the helium atom (Epstein 1916). His method was based on similar work by the French astronomer Charles Eugène Delaunay. Born recognized that the perturbations in his theory were similar to the degenerate perturbations in celestial mechanics called “secular perturbations” (M. Born 1924; 1925). The word “secular” was first used in classical mechanics to describe a perturbation that has a very slow and cumulative effect on the orbit of a planet.

Much of the later progress on perturbation theory stems from the works of Born, Schrödinger, Epstein, and Dirac; these methods built on the earlier work of Henri Poincaré. The early formulations of perturbation theory were modified for their application in both old and new quantum theories (Masters 2008b, 36–41). In 1926, Schrödinger published five papers on his newly-derived wave mechanics and some applications to the “Stark effect” of the “Balmer lines.” He developed his time-dependent wave equation and was able to calculate the intensities and polarization of the “Stark effect” on the “Balmer series” of electronic transitions in the hydrogen atom. His expression for the energy shifts is equivalent to that derived by Epstein. In 1926 and 1927, Dirac developed his time-dependent perturbation theory (Dirac 1926; 1927b). Dirac's time-dependent theory was the basis of Göppert's dissertation.

In the second part of her Göttingen dissertation, she calculated the probabilities of the combined action of light and electron collisions in the electronic transitions of atoms. Göppert's dissertation contained the theoretical basis for two-photon absorption and emission processes; she called the effects “double absorption” and “double emission.”¹⁶

The probability of the two-photon process is proportional to the square of the light intensity, and the rate constant for the two-photon process is very low

¹⁶It is significant that in Born and Jordan's *Elementare Quantenmechanik* (1930), section 74 on the absorption and emission of radiation by atoms, they cite in the footnote on page 400 Dirac's 1927b and 1927c papers, and they state that the theoretical development in the section follows Dirac's work. In section 75 on scattering and dispersion, a footnote states the text is analogous to Göppert's 1929 “Die Naturwissenschaften” paper, but in fact it is largely taken from Göppert's 1931 Göttingen dissertation. A careful comparison of section 75, her 1929 publication, and her 1931 Göttingen dissertation clearly indicates that an early draft of her Göttingen dissertation is the basis of section 75.

compared to a single-photon process that has a rate constant that is proportional to the light intensity. Göppert predicted nonlinear interactions between light and matter mediated by multi-photon processes. Furthermore, she showed that in a double transition or a two-photon transition via intermediate states or a virtual state, each part of the transition does not obey the conservation of energy law; however, the total transition from the initial state to the final state follows the law of conservation of energy. This is exactly what Dirac showed in his dispersion paper.

Weisskopf and Göppert were contemporary doctoral students of Born in Göttingen. It is interesting to explore Dirac's influence on Weisskopf's research and to compare Dirac's influence on the two doctoral students. Perusal of Weisskopf's Göttingen dissertation and his biography provide additional support for the mutual interaction of the experimental and theory groups in Göttingen's physics institutes, as well as the effect of the visitors on both Göppert's and Weisskopf's research programs (Weisskopf 1931; 1991).

Although Weisskopf arrived in 1928, both published their Göttingen dissertation in the same 1931 volume of *Annalen der Physik*. In his 1991 biography, Weisskopf cites the people who had seminal influences on his research in Göttingen: Franck, the experimental physicist who could accurately predict the results of an experiment or a theoretical calculation; Hilbert and especially Courant, who taught Weisskopf advanced mathematics; the three young teachers, Walter Heitler, Lothar Nordheim and especially Herzberg who taught the course "Introduction to Quantum Mechanics," which included the latest developments in the field (Weisskopf 1991).

According to Weisskopf, it was Dirac's 1927 paper, "The Quantum Theory of Emission and Absorption of Radiation" (Dirac 1927b), which was published prior to Weisskopf's arrival in Göttingen, that influenced Weisskopf's choice of a thesis problem (Weisskopf 1991). Dirac's paper demonstrated how to calculate the rates of the emission and absorption of light from an atom, but not how to calculate the line width of the transitions; Weisskopf decided to investigate the line width shapes for the transition from the first excited state to the ground state.

Born had a stroke shortly before Weisskopf arrived in Göttingen in 1928; therefore, Weisskopf turned to Wigner for mentorship. Wigner often visited Göttingen from Berlin. Together, they started with Dirac's 1927 paper on radiation (Dirac 1927b) and developed a novel theory that was published in two papers in 1930 (Weisskopf and Wigner 1930; 1930). Their first paper was entitled "Calculation of the natural line width due to the Dirac theory of light" (Weisskopf and Wigner 1930) in which the authors twice credit Dirac for previously publishing the techniques used in their present calculations for the interaction of light and matter. These include standing waves of radiation in a cavity and the ma-

trix methods to calculate transitions. The authors also wrote a footnote crediting Göppert for a similar calculation published in her 1929 paper (Göppert 1929).

In their second paper, Weisskopf and Wigner extended their calculation of the natural line width due to the Dirac theory of light interacting with an atom. The authors found that their quantum mechanically calculated line width of a harmonic oscillator coincides perfectly with the line width as calculated by classical theory (Weisskopf and Wigner 1930).

Weisskopf and Wigner's two 1930 publications incorporated the assumption "that all the atomic states that were not directly involved in the emission of radiation could be neglected" (Weisskopf 1991, 43). According to Weisskopf, this technique differed from the perturbation techniques, "which assumed that the interaction between the atom and the light is very small" (Weisskopf 1991, 43).

This so-called Weisskopf-Wigner theory was later used to solve other problems in quantum field theory (Weisskopf 1991). Because this joint research could not be submitted as his dissertation work, Weisskopf used the same theoretical approach to solve the problem of the re-emission of light absorbed from atoms. The title of his Göttingen dissertation is "Zur Theorie der Resonanzfluoreszenz" (Weisskopf 1931). Weisskopf's selection of this topic was also influenced by the work of Robert Wood, an experimental physicist who worked at Johns Hopkins University and published spectroscopic data on resonance fluorescence (Wood and Ellett 1924). Weisskopf discussed Wood's spectroscopic studies with Franck, whose spectroscopic group was involved with measurements of line widths (Weisskopf 1991). At the end of his Göttingen dissertation, Weisskopf thanked Born, Franck and Wigner for many supportive suggestions and discussions; these thanks provide further evidence of the interactions between the experimental group headed by Franck and the theoreticians Born and Wigner (Weisskopf 1991).

In summary, a study of Göppert's and Weisskopf's Göttingen dissertations indicates the fundamental influences of Dirac's prior publications. They also illustrate the communication between Franck's experimental groups and these two graduate students in Born's theory group.

8.6 Conclusion

From the previous discussion, I conclude that Dirac's 1927 publications had a substantial influence on Göppert's and Weisskopf's doctoral research, on their 1931 Göttingen dissertations, and on their publications of 1929 and 1930, respectively. Although it was not previously described, I suggest that Göppert's research borrowed more heavily from Dirac's theoretical techniques (with appropriate citations to Dirac) than did Weisskopf. In fact, my comparison of the dissertations

and Dirac's previously published papers indicate that Göppert borrowed Dirac's theoretical techniques to an extraordinary extent.

The question remains: why was Göppert able to predict two-photon transitions and calculate their probabilities for several cases? I propose the following answer. Oldenberg suggested that the experimental findings are indicative of a two-photon process. In addition, Dirac had provided the theoretical techniques to calculate the probabilities for two-photon transitions via virtual states in his dispersion paper of 1927. Göppert possessed superb mathematical skills, as well as a deep insight into experimental physics, was able to perform a synthesis of the works of Oldenberg and Dirac, and was able to work through the detailed quantum mechanical calculations that resulted in a theoretical understanding of Oldenberg's results.

She calculated the transition probabilities for two-photon absorption, two-photon emission and two-photon Raman processes for the Stokes and the anti-Stokes cases. With the invention of the laser, her theoretical predictions of two-photon processes of light absorption and emission would later be verified (Boyd 2008; Maiman 1960; Masters and So 2008).

Acknowledgements

I appreciate the cogent comments and suggestions from the participants of the *Third International Conference on the History of Quantum Physics*, Max Planck Institute for the History of Science, Berlin, held 28 June to 2 July 2010. I thank Shaul Katzir for helpful comments and for editing the paper. I thank Lisa Crystal and Myrna Lynn Perez of the Department of the History of Science at Harvard University for helpful suggestions. I thank Katie Deuschl (formerly Lagoni-Masters) for checking my translation of Göppert's Göttingen dissertation. I thank Arkajit Dey and Peter Musial for their assistance with the LaTeX software. Finally, I thank the anonymous reviewer for helpful comments.

References

- Bacciagaluppi, Guido and Antony Valentini (2009). *Quantum Theory at the Crossroads, Reconsidering the 1927 Solvay Conference*. Cambridge: Cambridge University Press.
- Born, Hedwig and Max Born (1962). *Der Luxus des Gewissens*. München: Nymphenburger Verlagshandlung GmbH.
- Born, Max (1924). Über Quantenmechanik. *Zeitschrift für Physik* 26: 379–395.
- (1925). *Vorlesungen über Atommechanik*. Berlin: Julius Springer.
- Born, Max and James Franck (1925). *Struktur der Materie in Einzeldarstellungen*. Berlin: Julius Springer.
- Born, Max and Maria Göppert-Mayer (1933). Dynamische Gittertheorie der Kristalle. In: *Handbuch der Physik, Zweite Auflage*. Ed. Hans Geiger and Karl Scheel. Berlin: Julius Springer. 623–794.
- Born, Max, Werner Heisenberg, and Pascual Jordan (1926). Zur Quantenmechanik II. *Zeitschrift für Physik* 35: 557–615.
- Born, Max and Pascual Jordan (1930). *Elementare Quantenmechanik, Zweiter Band der Vorlesungen über Atommechanik*. Berlin: Julius Springer.
- Boyd, Robert (2008). *Nonlinear Optics, Third Edition*. San Diego: Academic Press.
- Dirac, Paul A. M. (1926). On the Theory of Quantum Mechanics. *Proceedings of the Royal Society A* 112: 661–677.
- (1927a). The Physical Interpretation of the Quantum Dynamics. *Proceedings of the Royal Society A* 113: 621–641.
- (1927b). The Quantum Theory of Emission and Absorption of Radiation. *Proceedings of the Royal Society A* 114: 243–265.
- (1927c). The Quantum Theory of Dispersion. *Proceedings of the Royal Society A* 114: 710–728.
- Eckert, Michael (2001). The Emergence of Quantum Schools: Munich, Göttingen and Copenhagen as New Centers of Atomic Theory. *Annalen der Physik (Leipzig)* 10: 151–162.
- Einstein, Albert (1916). Strahlungs-Emission und -Absorption nach der Quantentheorie. *Verhandlungen der Deutschen Physikalischen Gesellschaft* 18: 318–323.
- Epstein, Paul S. (1916). Zur Theorie des Starkeffektes. *Annalen der Physik* 50: 489–520.
- (1922a). Die Störungsrechnung im Dienste der Quantentheorie I. Eine Methode der Störungsrechnung. *Zeitschrift für Physik* 8: 211–228.
- (1922b). Die Störungsrechnung im Dienste der Quantentheorie II. Die numerische Durchführung der Methode. *Zeitschrift für Physik* 8: 305–320.

- Epstein, Paul S. (1922c). Die Störungsrechnung im Dienste der Quantentheorie III. Kritische Bemerkungen zur Dispersionstheorie. *Zeitschrift für Physik* 9: 92–110.
- Franck, James (1928). Beitrag zum Problem der Wiedervereinigung von Ionen und Elektronen. *Zeitschrift für Physik* 47: 509–516.
- Franck, James and Gustav Hertz (1914). Über Zusammenstöße zwischen Elektronen und Molekülen des Quecksilberdampfes und die Ionisierungsspannung desselben. *Verhandlungen der Deutschen Physikalischen Gesellschaft* 16: 457–467.
- Franck, James and Pascual Jordan (1926). *Anregung von Quantensprüngen durch Stöße*. Berlin: Julius Springer.
- Georg-August Universität (2011a). *The History of the University*. <http://www.uni-goettingen.de/en/52652.html> (accessed 02/27/2012).
- (2011b). *Freedom from Censorship*. <http://www.uni-goettingen.de/en/30768.html> (accessed 02/27/2012).
- Göppert, Maria (1929). Über die Wahrscheinlichkeit des Zusammenwirkens zweier Lichtquanten in einem Elementarakt. *Die Naturwissenschaften* 17: 932.
- Göppert-Mayer, Maria (1930). Absorption und Emission von Strahlung durch Atome. In: *Elementare Quantenmechanik, Zweiter Band der Vorlesungen über Atommechanik*. Ed. Max Born and Pascual Jordan. Berlin: Julius Springer. 400–429.
- (1931). Über Elementarakte mit zwei Quantensprüngen. *Annalen der Physik (Leipzig)* 9: 273–294.
- (1948). On Closed Shells in Nuclei. *The Physical Review, Second Series* 74(3): 235–239.
- Göppert-Mayer, Maria and Johannes Hans D. Jensen (1955). *Elementary Theory of Nuclear Shell Structure*. New York: John Wiley and Sons, Inc.
- Greenspan, Nancy T. (2005). *The End of the Certain World: The Life and Science of Max Born*. New York: Basic Books.
- Heilbron, John L. (1967). The Kossel-Sommerfeld Theory and the Ring Atom. *Isis* 58: 450–485.
- Hentschel, Klaus (1999). The Physical Tourist. Some Historical Points of Interest in Göttingen. *Physics in Perspective* 1: 110–117.
- Hund, Friedrich (1983). Born, Göttingen und die Quantenmechanik. In: *James Franck und Max Born in Göttingen: Reden zur akademischen Feier aus Anlass der 100. Wiederkehr ihres Geburtsjahres*. Göttingen: Vandenhoeck und Ruprecht. With contributions by Norbert Kamp, Peter Haasen, Gerhart W. Rathenau and Friedrich Hund.

- (1987). *Die Geschichte der Göttinger Physik*. Göttingen: Vandenhoeck und Ruprecht.
- Im, Gyeong-Soon (1995). Experimental Constraints on Formal Quantum Mechanics: The Emergence of Born's Quantum Theory of Collision Processes in Göttingen, 1924–1927. *Archive for History of Exact Sciences* 50: 73–101.
- Johnson, Karen E. (1999). Science at the Breakfast Table. *Physics in Perspective* 1: 22–34.
- (2004). From Natural History to the Nuclear Shell Model: Chemical Thinking in the Work of Mayer, Haxel, Jensen, and Suess. *Physics in Perspective* 6: 295–309.
- Jungnickel, Christa and Russell McCormmach (1986). *Intellectual Mastery of Nature: Theoretical Physics from Ohm to Einstein. Volume 2. The Now Mighty Theoretical Physics 1870–1925*. Chicago: The University of Chicago Press.
- Kamp, Norbert, Peter Haasen, Gerhart W. Rathenau, and Friedrich Hund (1983). *James Franck und Max Born in Göttingen: Reden zur akademischen Feier aus Anlass der 100. Wiederkehr ihres Geburtsjahres*. Göttingen: Vandenhoeck und Ruprecht.
- Kragh, Helge (1990). *Dirac: A Scientific Biography*. Cambridge: Cambridge University Press.
- Kramers, Hendrik A. and Werner Heisenberg (1925). Über die Streuung von Strahlen durch Atome. *Zeitschrift für Physik* 31: 681–708.
- Lampe, Jörg H. (2002). Politische Entwicklungen in Göttingen vom Beginn des 19. Jahrhunderts bis zum Vormärz. In: *Göttingen: Geschichte einer Universitätsstadt*. Ed. Dietrich Denecke and Ernst Böhme. Göttingen: Vandenhoeck und Ruprecht. 43–102.
- Lemmerich, Jost (2007). *Aufrecht im Sturm der Zeit. Der Physiker James Franck, 1882–1964*. Berlin: GNT-Verlag.
- Maiman, Theodore H. (1960). Stimulated Optical Radiation in Ruby. *Nature* 187: 493–494.
- Marchand, Suzanne L. (1996). *Down from Olympus: Archaeology and Philhellenism in Germany, 1750–1970*. Princeton: Princeton University Press.
- Masters, Barry R. (2000). The Scientific Life of Maria Göppert-Mayer. *Optics and Photonics News* 11(9): 38–41.
- (2008a). The Scientific Life of Maria Göppert-Mayer. In: *Handbook of Biomedical Nonlinear Optical Microscopy*. Ed. Barry R. Masters and Peter T. C. So. New York: Oxford University Press. 30–35.
- (2008b). The History of Perturbation Theory from Astronomy to Quantum Mechanics. In: *Handbook of Biomedical Nonlinear Optical Microscopy*. Ed.

- Barry R. Masters and Peter T. C. So. New York: Oxford University Press. 36–41.
- Masters, Barry R. (2010). *English Translation of Göppert (1929) and Göppert-Mayer (1931)*. http://quantum-history.mpiwg-berlin.mpg.de/news/workshops/hq3/hq3_talks/21_masters_appendix.pdf (accessed 06/12/2012).
- Masters, Barry R. and Peter T. C. So (2004). Antecedents of Two-Photon Excitation Laser Scanning Microscopy. *Microscopy Research and Technique* 63: 3–11.
- eds. (2008). *Handbook of Biomedical Nonlinear Optical Microscopy*. New York: Oxford University Press.
- McGrayne, Sharon B. (1993). Maria Göppert-Mayer. In: *Nobel Prize Women in Science: Their Lives, Struggles, and Momentous Discoveries*. New York: Birch Lane Press. 175–200.
- Oldenberg, Otto (1928). Über ein Zusammenwirken von Stoss und Strahlung. *Zeitschrift für Physik* 51: 605–612.
- Rowe, David E. (1989). Klein, Hilbert, and the Göttingen Mathematical Tradition. *Osiris, 2nd Series* 5: 186–213.
- Rupke, Nicolaas (2002). *Göttingen and the Development of the Natural Sciences*. Göttingen: Wallstein Verlag.
- Sachs, Robert G. (1982). Maria Göppert-Mayer—Two-Fold Pioneer. *Physics Today* 35(2): 46–51.
- Schirmacher, Arne (2003). Planting in His Neighbor's Garden: David Hilbert and Early Göttingen Quantum Physics. *Physics in Perspective* 5: 4–20.
- Seth, Suman (2010). *Crafting the Quantum: Arnold Sommerfeld and the Practice of Theory, 1890–1926*. Boston: MIT Press.
- Smekal, Adolf G. (1923). Zur Quantentheorie der Dispersion. *Die Naturwissenschaften* 11: 873–875.
- Weisskopf, Victor (1931). Zur Theorie der Resonanzfluoreszenz. *Annalen der Physik (Leipzig)* 9: 23–66.
- (1991). *The Joy of Insight, Passions of a Physicist*. New York: Basic Books.
- Weisskopf, Victor and Eugene Wigner (1930). Berechnung der natürlichen Linienbreite auf Grund der Diracschen Lichttheorie. *Zeitschrift für Physik* 63: 54–73.
- Weisskopf, Victor and Eugene Wigner (1930). Über die natürliche Linienbreite in der Strahlung des harmonischen Oszillators. *Zeitschrift für Physik* 65: 18–29.
- Wood, Robert and Alexander Ellett (1924). Polarized Resonance Radiation in Weak Magnetic Fields. *Physical Review* 24: 243–254.

Chapter 9

An Act of Creation: The Meitner-Frisch Interpretation of Nuclear Fission

Roger H. Stuewer

In late December 1938 Lise Meitner met her nephew Otto Robert Frisch in Kungälv, north of Göteborg, Sweden, to spend the Christmas holidays with Swedish friends. They had not planned to meet in that unlikely place; both were victims of Nazi persecution: Frisch had been dismissed from his position in Hamburg after the promulgation of the Nazi Civil Service Law on 7 April 1933, and after a year in London he found a haven in the fall of 1934 in Niels Bohr's Institute for Theoretical Physics in Copenhagen (Frisch 1979a; 1979b, 41–108; Peierls 1981, 285–290). Meitner, protected by her Austrian citizenship, had remained in Berlin until mid-July 1938, four months after the Anschluss of Austria, when she was spirited out of Berlin and into The Netherlands, from whence she flew to Copenhagen and then went on to Stockholm (Sime 1990; 1996, 184–209). Thus, Meitner traveled to Kungälv from Stockholm, and Frisch from Copenhagen.

Both Meitner and Frisch had deep personal worries on their minds: Meitner was fifty-nine years old and had a distinguished scientific reputation, but Manne Siegbahn gave her no financial or personal support in Stockholm to continue her researches. Frisch's father, a lawyer employed by the publishing firm of Bermann-Fischer in Vienna, had been dismissed from his position in June 1938, and now in December was incarcerated in Dachau, while his mother was threatened and worried sick in Vienna.¹

Yet, when aunt and nephew first met at breakfast in their hotel in Kungälv, probably on the morning of 24 December 1938, it was not these deep personal worries that commanded their attention, but a letter that Meitner had just received from Otto Hahn in Berlin in which he reported that he and Fritz Strassmann, in continuing the experiments they had begun with Meitner, had found barium as one of the products of the reaction when uranium was bombarded with neutrons.² Pondering Hahn's letter, first in their hotel, then on a hike outdoors in the snow, it suddenly occurred to Meitner and Frisch that the liquid-drop model of the

¹Frisch to Meitner, 1 October 1938, Frisch Papers. See also (Frisch 1973).

²Hahn to Meitner, 19 December 1938, Meitner Papers.

nucleus might provide the key to understanding Hahn and Strassmann's experiments. Rough calculations substantiated their conjecture, and on 1 January 1939, filled with excitement, Meitner returned to Stockholm and Frisch to Copenhagen.

Two days later, Frisch caught Bohr and told him of his and Meitner's interpretation. Bohr was greatly surprised. Much later, Frisch said that Bohr burst out with the words, "Oh, what fools we have been! We ought to have seen that before" (Frisch 1967b, 47).

Now, Bohr's surprise must be taken seriously, because it is precisely such a moment of surprise that reveals that a great transformation in thought has taken place. We recall Arthur Koestler's analysis of the act of creation in which he argued that the essence of the creative act is the synthesis of two previously unrelated "matrices of thought" (Koestler 1964, 207). Koestler also noted that:

[The] more original a discovery the more obvious it seems afterwards. The creative act is not an act of creation in the sense of the Old Testament. It does not create something out of nothing; it uncovers, selects, re-shuffles, combines, synthesizes already existing facts, ideas, faculties, skills. (Koestler 1964, 120)

Koestler, of course, was mainly concerned with the creative acts of individuals; I think, however, that we shall see that his analysis also seems applicable to the creative act of the two physicists, Meitner and Frisch, at the end of 1938.³

We therefore are presented with several historical questions: Why was Meitner and Frisch's interpretation so obvious to Bohr as soon as he learned about it? Why was Bohr so astonished that he, or for that matter, other contemporary physicists had failed to see what Meitner and Frisch had seen? Conversely, why were Meitner and Frisch able to achieve their insight, their creative act? To answer these questions, we must examine the history of the liquid-drop model of the nucleus in some detail. We will see that by December 1938 that model was exactly ten years old, and that it had been developed in two distinct phases, from 1928 to 1935 and from 1936 to 1937. Moreover, as we will see immediately, the inventor of the liquid-drop model was not Niels Bohr, as many physicists believe today, but rather George Gamow.

9.1 The Birth of the Liquid-Drop Model

In June 1928 twenty-four-year-old George Gamow left Leningrad to spend the summer in Max Born's institute in Göttingen where immediately after his arrival he made his first major contribution to nuclear physics, his quantum-mechanical

³For a full account, see (Stuewer 1994).

theory of alpha decay (Gamow 1928; Stuewer 1986).⁴ The following month he wrote to Niels Bohr, making arrangements to visit Copenhagen before returning to Leningrad in the fall.⁵ He arrived in Copenhagen in September, and when he told Bohr about his theory Bohr was so impressed with it, and with Gamow personally, that he offered Gamow a fellowship to enable him to spend the entire academic year 1928–1929 in his institute. Bohr also arranged for Gamow to visit the Cavendish Laboratory in Cambridge, England, for around five weeks, from early January to mid-February 1929. Gamow thrived in both places. In particular, just before leaving Copenhagen for Cambridge, he invented the liquid-drop model of the nucleus, which he presented for the first time on 7 February 1929, at a meeting of the Royal Society in London to which Ernest Rutherford had invited him.⁶

Gamow's basic idea was that the nucleus consists of a collection of N alpha particles with short-range attractive forces between them that balance their Coulomb repulsion. All are in a state of lowest energy since they obey Bose-Einstein statistics. The alpha particles exert an outward pressure owing to their kinetic and potential energy and are held inside the nucleus by its "surface tension," so one can calculate the total "drop energy" E of the nucleus in terms of the number N of alpha particles inside it, that is, in terms of the atomic weight of the nucleus.

Gamow returned to Russia that summer and in the fall was back in Cambridge on a Rockefeller Foundation fellowship where he pursued his liquid-drop model further during the academic year 1929–1930. The result was a paper, "Mass Defect Curve and Nuclear Constitution" (Gamow 1930), which Rutherford communicated to the *Proceedings of the Royal Society* on 28 January 1930. In it Gamow developed his liquid-drop model quantitatively.

Gamow's basic idea was the same as before. He assumed that the nucleus consists of N alpha particles, each of mass m , charge e , kinetic energy T , and potential energy V . He estimated T from Heisenberg's uncertainty principle and determined V from the virial theorem, finding that $T + V \approx -h^2/(4mr^2)$, where h is Planck's constant and r is the radius of the nucleus. Then, in a separate calculation, he equated the surface tension $S(r, N)$ to the internal pressure $P(r, N)$, both of which are functions of r and N , and found that $r = RN^{1/3}$, where $R = 2 \times 10^{-13}$ cm. The total internal energy E_i of the alpha particles thus is $E_i = N(T + V) \approx -(h^2/4mR^2)N^{1/3}$. The Coulomb repulsive energy E_c of the alpha particles at the nuclear surface is $E_c \approx (2eN)^2/r = (4e^2/R)N^{5/3}$. The

⁴The same theory was published independently and virtually simultaneously by Ronald W. Gurney and Edward U. Condon; see (Gurney and Condon 1928; 1929).

⁵Gamow to Bohr, 21 July 1928, BSC.

⁶"Discussion on the Structure of Atomic Nuclei" (Rutherford et al. 1929).

total energy E of the nucleus therefore is $E = E_i + E_c = -(h^2/4mR^2)N^{1/3} + (4e^2/R)N^{5/3}$, or the sum of an attractive term varying as $N^{1/3}$ and a repulsive term varying as $N^{5/3}$. Fig. (9.1) is his plot of E versus N , where the shading of the curve indicates the approximate nature of his calculation. Since it was in poor agreement with Francis W. Aston's mass-defect measurements (dots in the picture), he then took account of the electrons that everyone assumed were present in nuclei and found that the minimum in the mass-defect curve moved out to higher atomic weights as Aston's measurements required.

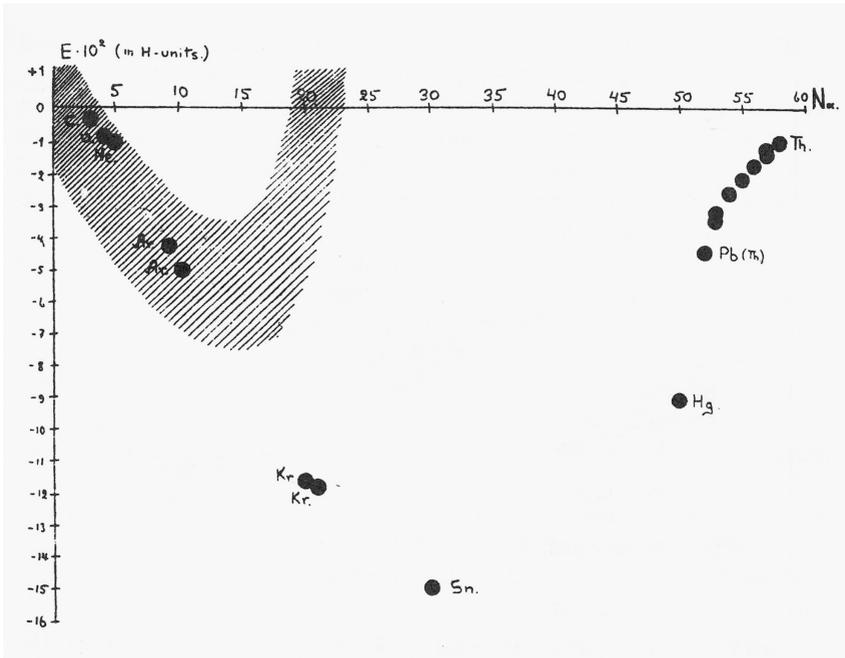


Figure 9.1: Gamow's plot of the total nuclear energy E ($\times 10^2$ in hydrogen-mass units) versus the number of alpha particles N (N_α in Gamow's notation) in nuclei. Source: (Gamow 1930, 637).

During the following academic year 1930–1931 Gamow was back in Copenhagen, where he made no further progress on his mass-defect calculations. He again returned to Russia in the summer to renew his passport so that he could attend an international conference on nuclear physics—the first of its kind—that

Enrico Fermi was organizing in Rome for October 1931.⁷ The Soviet authorities, however, refused to renew his passport, so he remained in Leningrad where he taught physics, married, and made plans to escape. He thus was in Leningrad when the field of nuclear physics was fundamentally transformed between the end of 1931 and the end of 1932 by the discoveries of deuterium, the neutron, and the positron, and the construction of the Cockcroft-Walton accelerator and the cyclotron.

9.2 The Extension of the Liquid-Drop Model: Heisenberg and von Weizsäcker

The most important of these developments for our purpose was James Chadwick's discovery of the neutron in February 1932 (Chadwick 1932a; 1932b), because a few months later, between June and December 1932, Werner Heisenberg submitted for publication his seminal theory of nuclear structure in which he assumed that a neutron and a proton in the nucleus are bound together by exchanging an electron between them (Heisenberg 1932a; 1932b; 1933a). Then, in early 1933, Ettore Majorana, while visiting Heisenberg's institute in Leipzig, published, on Heisenberg's urging, a new neutron-proton force involving the exchange of both charge and spin (Majorana 1933). A compelling argument in its favor was that it saturated at the alpha particle, a highly stable particle, whereas Heisenberg's force saturated at the deuteron, a much less stable particle. Heisenberg therefore adopted Majorana's exchange force and recalculated the total energy E of a nucleus in terms of the total number of neutrons n_1 and protons n_2 in it. He presented his calculation and a plot of the resulting mass-defect curve, see fig. (9.2), at the seventh Solvay Congress in Brussels in October 1933 (Heisenberg 1933b). He found, just as Gamow had, that the mass-defect curve has a minimum in it at nuclei of intermediate atomic weights. That must have been particularly gratifying to Gamow, who was present at the Solvay Congress: Quite mysteriously, both he and his wife were given passports and visas to attend it—an opportunity they took to leave the Soviet Union for good (Gamow 1970, 118–133).

Two years later, in 1935, Heisenberg's student Carl Friedrich von Weizsäcker built on his mentor's work and extended Gamow's liquid-drop model still further by introducing his famous semi-empirical mass formula (von Weizsäcker 1935). The result was the same, namely, a minimum in the mass-defect curve at nuclei of intermediate atomic weights,⁸ see fig. (9.3). Von Weizsäcker's semi-empirical mass formula thus represented, historically speaking, the culmination of the line

⁷(Reale Accademia d'Italia 1932).

⁸To anticipate: If a heavy nucleus like uranium ($N + Z \sim 200$) were to split up into two nuclei in the middle of the periodic table ($N + Z \sim 75$), then the difference in their energies per nucleon $E/(N + Z)$,

of development that Gamow inaugurated when he invented the liquid-drop model in 1928.

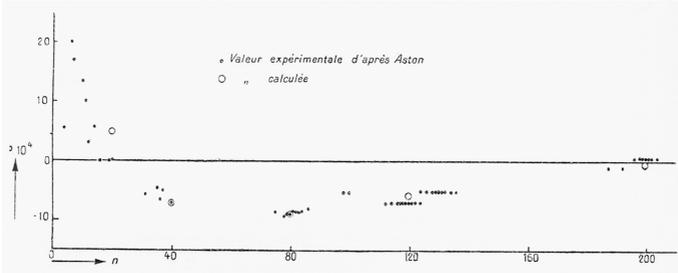


Figure 9.2: Heisenberg’s plot of the total nuclear energy E ($\times 10^4$ in units of mc^2) versus the total number of neutrons and protons $n = n_1 + n_2$ in nuclei. The dots represent Aston’s experimental mass-defect data, the circles Heisenberg’s theoretically calculated values. Source: (Heisenberg 1933b, 318).

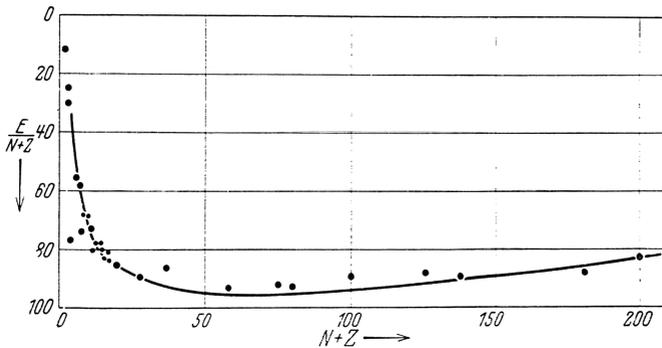


Figure 9.3: Von Weizsäcker’s plot of the mass defect or energy per nucleon $E/(N + Z)$ versus the total number of nucleons (N neutrons + Z protons) in nuclei. Source: (von Weizsäcker 1935, 457).

as represented by the difference in their ordinates on the mass-defect curve, would be released in the process.

9.3 Bohr's Theory of the Compound Nucleus

A second phase in the history of liquid-drop model began in February 1936 when Niels Bohr published his theory of the compound nucleus (Bohr 1936). The greatest influence on Bohr's thought was Enrico Fermi's discovery in October 1934 that slow neutrons are much more efficacious in producing nuclear reactions than fast ones (Fermi et al. 1934), contrary to what might be expected on simple energy considerations. The basic idea that Bohr presented was that an incident neutron, for example, interacts with many neutrons and protons in a target nucleus, producing an excited long-lived compound nucleus, which then decays by the emission of a proton, neutron, gamma ray, or by any process that is consistent with conservation of energy.

For our purposes, the principal point of interest is Bohr's picture of what would happen if the energy of the incident neutron increased more and more:

Even if we could experiment with neutrons or protons of energies of more than a hundred million volts, we should still expect that the excess energy of such particles [...] would in the first place be divided among the nuclear particles [...]. [We] may, however, in such cases expect that in general not one but several charged or uncharged particles will eventually leave the nucleus as a result of the encounter. For still more violent impacts, with particles of energies of about a thousand million volts, we must even be prepared for the collision to lead to an explosion of the whole nucleus. (Bohr 1936, 348)

Thus, first one, then two, and eventually all of the particles would be exploded out of the target nucleus.

Bohr also concluded that nuclear excitations, in general, were due to oscillations of the surface of a nucleus, not to oscillations of its volume, much like those of an elastic solid. He discussed these ideas widely in lectures on a trip around the world in the first half of 1937 with the help of a picture, see fig. (9.4), drawn for him by Otto Robert Frisch. We see that an incident neutron transfers energy to the target nucleus, causing the surface of the compound nucleus to oscillate and the temperature of the nucleus to rise, after which a single particle is emitted and the nucleus cools down, and finally returns to its initial state and temperature by emitting a gamma ray. He published these ideas in a paper of October 1937 that he wrote with his young assistant, Fritz Kalckar (1937).

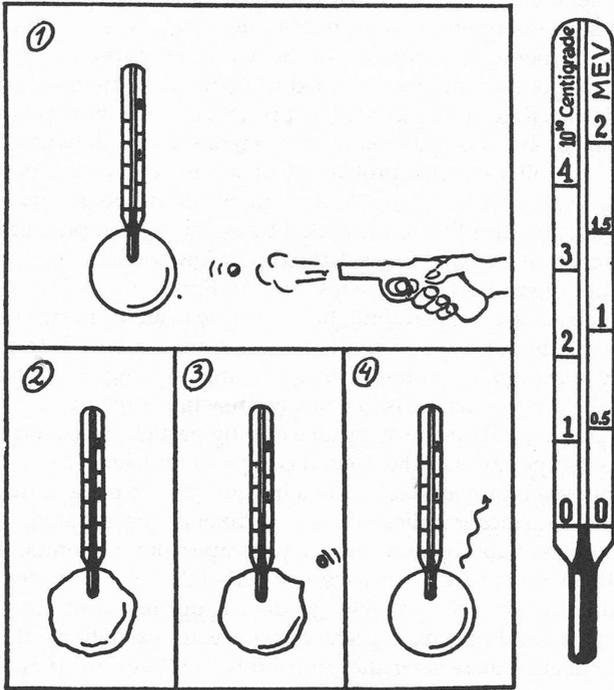


Figure 9.4: Bohr's illustration (as drawn by Frisch) of the excitation and deexcitation of a heavy nucleus as absorption and evaporation processes. Source: (Bohr 1937, 163).

9.4 The Interpretation of Fission

We have seen that there were two stages in the development of the liquid-drop model: Stage I, from 1928 to 1935, was delineated by the work of Gamow (1928–1931), Heisenberg (1933), and von Weizsäcker (1935), who applied the liquid-drop model to a calculation of the nuclear mass-defect curve. Their focus, in other words, was on static features of the model. Stage II, from 1936 to 1937, was delineated by the work of Bohr (1936) and Bohr and Kalckar (1937), who applied the liquid-drop model to a calculation of nuclear excitations. Their focus, in other words, was on dynamic features of the model.

These two stages formed what I might call the Berlin and Copenhagen traditions of research on the liquid-drop model, both of which persisted into 1938. Thus, in April 1938 von Weizsäcker published a paper that was largely devoted to a discussion of the liquid-drop model and its application to the calculation of nuclear mass defects (von Weizsäcker 1938). And in August 1938 Bohr presented a report at a meeting of the British Association for the Advancement of Science in Cambridge that concentrated on the problem of nuclear excitations,⁹ analyzing them as he had in his paper with Kalckar.

Now, just between these two dates, on 13 July 1938, Lise Meitner was spirited out of Berlin and into The Netherlands and then went on to Copenhagen and Stockholm. She was thoroughly embedded in the Berlin tradition, being completely familiar with both Heisenberg's and von Weizsäcker's work. Thus, she was present at the seventh Solvay Congress in October 1933 where she heard Heisenberg outline his calculation of the nuclear mass-defect curve and saw its agreement with Aston's experimental data, and she came into close contact with von Weizsäcker when, after completing his *Habilitationsschrift* under Heisenberg in Leipzig (von Weizsäcker 1936), he moved to Berlin in July 1936.

Similarly, Otto Robert Frisch was thoroughly embedded in the Copenhagen tradition. He was present in Bohr's institute when Bohr conceived his theory of the compound nucleus, and he immediately carried out experiments related to Bohr's theory (Frisch 1937). Moreover, as I noted above, he even drew the pictures that Bohr used to illustrate his theory in his lectures in Europe and on his trip around the world during the first half of 1937.

It was these two research traditions, then, that came together for the first time in the minds of Meitner and Frisch, probably on the morning of 24 December 1938, to produce an entirely new application of the liquid-drop model, their momentous interpretation of nuclear fission. I have reconstructed their conversation based upon Frisch's several recollections of their memorable walk in the

⁹(Nuclear Physics 1938).

snow in Kungälv, Sweden (Frisch 1967a, 144–145; 1967b, 47; 1973; 1979a, 71; 1979b, 114–116).¹⁰

- Meitner rejected the idea that Hahn had made a mistake; he was too good a chemist to have done so: When he said that he and Strassmann had found barium when neutrons bombarded uranium, they *had* found barium.
- Both Meitner and Frisch realized that this could not result from a chipping off or cracking up of the uranium nucleus.
- Meitner then seems to have thought of the liquid-drop model of the nucleus in this connection, and she made a sketch of a large circle with a smaller circle inside it.
- Frisch immediately interpreted Meitner’s sketch as an end-on view of a dumbbell—as an elongated liquid drop with a constriction between its two halves.
- Meitner then estimated the amount of energy that would be produced if the uranium nucleus would split into two nuclei at the middle of the periodic table, finding it to be about 200 MeV—an enormous amount. As Frisch remarked, she “had the mass-defect curve pretty well in her head” (Frisch 1967b, 47).
- Meanwhile, Frisch had estimated that the electric charge of a nucleus would diminish its surface tension to around zero for a nucleus of nuclear charge $Z = 100$, and he also calculated that two nuclei in the middle of the periodic table, if initially in contact, would fly apart under their mutual Coulomb repulsion with an energy of about 200 MeV—in agreement with Meitner’s figure.
- As Frisch said, “We put our different kinds of knowledge together” (Frisch 1973, 833). In other words, as Koestler would have claimed, a synthesis of Meitner’s and Frisch’s different “matrices of thought” occurred.

9.5 Aftermath

We now can understand, I think, Bohr’s great surprise when Frisch told Bohr of Meitner’s and his interpretation of Hahn and Strassmann’s experiments in Copenhagen on 3 January 1939, and why Bohr had not seen what they now saw. Thus, when Bohr presented his theory of the compound nucleus in February 1936 his picture of the disintegration of a heavy nucleus when bombarded with neutrons

¹⁰See also: Interview of Otto Robert Frisch by Charles Weiner on 3 May 1967, Niels Bohr Library & Archives; and (Frisch 1973). Nuclear Pioneer Lecture Honouring Lise Meitner. Presented to the Society of Nuclear Medicine, Miami Beach, Florida, June 1973. File C74, p. 9–10, Frisch Papers; and Frisch, Otto R. (no date). The Origin of Nuclear Fission. File C74, Frisch Papers.

was a completely different one: With ever increasing neutron energy first one, then two, then an increasing number of neutrons and protons would be expelled from the compound nucleus until eventually, at very high neutron energies, it would explode, sending its neutrons and protons out in all directions. Moreover, in his and Kalckar's paper of October 1937, Bohr was inclined to view the nucleus as an elastic solid, not as a liquid drop, when considering its surface oscillations and excitations. Perhaps most fundamentally, however, since Bohr was primarily interested in nuclear reactions and excitations, he never seems to have incorporated Gamow's, Heisenberg's, and von Weizsäcker's mass-defect calculations into his thinking. He never seems to have appreciated the compensatory roles played by the repulsive surface charge and attractive surface tension in delicately maintaining the stability of a heavy nucleus like uranium.

Now, however, given Meitner and Frisch's interpretation, Bohr immediately understood it and discussed it with Frisch over a period of four days before he boarded a ship for the United States on 7 January 1939, where immediately after his arrival in New York on 16 January his traveling companion, Léon Rosenfeld, let the cat out of the bag in Princeton, and where Bohr and Fermi announced the discovery and interpretation of nuclear fission at the fifth Conference on Theoretical Physics at George Washington University in Washington, D.C., on 26 January 1939 (Stuewer 1985), after which the news spread rapidly from coast to coast in the United States. Bohr then went on to Princeton where he became completely absorbed in the fission process, and where he and John Archibald Wheeler produced their classic paper, "The Mechanism of Nuclear Fission" (Bohr and Wheeler 1939).

Meanwhile, in Copenhagen, Frisch began to carry out experiments and around 3:00 A.M. on the morning of Friday, 13 January 1939, he first detected the fission fragments from uranium (Frisch 1939). He recalled that four hours later the postman woke him up and handed him a telegram from his mother saying that his father had been released from Dachau, and that both of his parents now could emigrate to Sweden (Frisch 1973, 833).¹¹

9.6 Conclusion

By early 1939 the history of the liquid-drop model had become obscured, because Bohr, in his and Kalckar's paper of October 1937, had failed to cite Gamow as its creator, even though Gamow had conceived it in Bohr's institute in December 1928, perhaps because Bohr saw his application of it as being so different

¹¹See also: Interview of Otto Robert Frisch by Charles Weiner on 3 May 1967, Niels Bohr Library & Archives; and (Frisch 1973). Nuclear Pioneer Lecture Honouring Lise Meitner. Presented to the Society of Nuclear Medicine, Miami Beach, Florida, June 1973. File C74, p. 13, Frisch Papers.

from Gamow's. And Bohr's omission was immediately propagated in the literature. Hans A. Bethe, in the second part of his famous three-part article in the *Reviews of Modern Physics* (the Bethe Bible) of April 1937, based his discussion of the liquid-drop model on Bohr and Kalckar's paper, which he read in its pre-publication form when he met Bohr in the United States during his trip around the world. Consequently, Bethe unwittingly and ironically stated in his first sentence that: "It was pointed out by Bohr and Kalckar (B33) that a nucleus should be considered as a drop of liquid [...]" (Bethe 1937, 86, 170).¹² Moreover, Bohr did not correct his omission of Gamow, or Bethe's mistake, in his and Wheeler's paper of 1939.

Gamow himself, however, never forgot his contribution. In January 1968 he prepared what no doubt was his last curriculum vitae (he died on 20 August 1968), and under the section he entitled "Theory of Nuclear Fluid" he noted that in Bohr's institute he had "proposed a hypothesis that atomic nuclei can be treated as little droplets of so-called 'nuclear fluid'. These views led ultimately to the present theory of nuclear fission and fusion."¹³

Abbreviations and Archives

AHQP	Archive for History of Quantum Physics. American Philosophical Society, Philadelphia
BSC	Bohr Scientific Correspondence in the AHQP
Niels Bohr Library & Archives	American Institute of Physics, College Park, MD, USA, www.aip.org/history/ohilist/4616.html
Frisch Papers	Trinity College Archive, University of Cambridge
Gamow Papers	Manuscript Division, Library of Congress, Washington, D.C.
Meitner Papers	Churchill College Archive, University of Cambridge

¹²Bethe revealed his subtle sense of humor in his reference B33, because, knowing Bohr's propensity for writing many drafts of his papers before releasing them for publication, Bethe gave 1939 as the year of publication of Bohr and Kalckar's paper.

¹³Curriculum Vitae dated January 1968 and entitled "GAMOW, GEORGE, American Physicist Born March 4, 1904, Odessa, Russia.", Gamow Papers.

References

- Amaldi, Edoardo, Enrico Persico, Franco Rasetti, and Emilio Segrè, eds. (1962). *Enrico Fermi. Collected Papers (Note e Memorie)*. Vol. 1. *Italy 1921–1938*. Chicago: The University of Chicago Press.
- Bethe, Hans A. (1937). Nuclear Physics. B. Nuclear Dynamics, Theoretical. *Reviews of Modern Physics* 9: 69–244.
- Bethe, Hans A., Robert F. Bacher, and Stanley M. Livingston (1986). *Basic Bethe. Seminal Articles on Nuclear Physics, 1936–1937*. New York: Tomash Publishers and American Institute of Physics.
- Beyer, Robert T., ed. (1949). *Foundations of Nuclear Physics. Facsimiles of Thirteen Fundamental Studies as They Were Originally Reported in the Scientific Journals*. New York: Dover.
- Bohr, Niels (1936). Neutron Capture and Nuclear Constitution. *Nature* 137: 344–388.
- (1937). Transmutation of Atomic Nuclei. *Science* 86: 161–165.
- Bohr, Niels and Fritz Kalckar (1937). On the Transmutation of Atomic Nuclei by Impact of Material Particles. I. General Theoretical Remarks. *Det Kongelige Danske Videnskabernes Selskab. Matematisk-fysiske Meddelelser* 14(10): 1–40.
- Bohr, Niels and John A. Wheeler (1939). The Mechanism of Nuclear Fission. *Physical Review* 56: 426–450.
- Chadwick, James (1932a). Possible Existence of a Neutron. *Nature* 129: 312.
- (1932b). The Existence of a Neutron. *Proceedings of the Royal Society A* 136: 692–708.
- Fermi, Enrico, Edoardo Amaldi, Bruno Pontecorvo, Franco Rasetti, and Emilio Segrè (1934). Azione di sostanze idrogenate sulla radioattività provocata da neutroni. I. *La Ricerca Scientifica* 5: 282–283.
- Frisch, Otto R. (1937). On the Selective Capture of Slow Neutrons. *Det Kongelige Danske Videnskabernes Selskab. Matematisk-fysiske Meddelelser* 14(12): 1–31.
- (1939). Physical Evidence for the Division of Heavy Nuclei under Neutron Bombardment. *Nature* 143: 276.
- (1967a). The Interest is Focussing on the Atomic Nucleus. In: *Niels Bohr: His Life and Work as Seen by His Friends and Colleagues*. Ed. Stefan Rozen-tal. Amsterdam: Wiley. 137–148.
- (1967b). How It All Began. *Physics Today* 20(11): 43–48.
- (1973). Walk in the Snow. *New Scientist* 60: 833.
- (1979a). Experimental Work with Nuclei: Hamburg, London, Copenhagen. In: *Nuclear Physics in Retrospect: Proceedings of a Symposium on the*

- 1930s. Ed. Roger H. Stuewer. Minneapolis: University of Minnesota Press. 65–75.
- Frisch, Otto R. (1979b). *What Little I Remember*. Cambridge: Cambridge University Press.
- Gamow, George (1928). Zur Quantentheorie des Atomkernes. *Zeitschrift für Physik* 51: 204–212.
- (1930). Mass Defect Curve and Nuclear Constitution. *Proceedings of the Royal Society A* 126: 632–644.
- (1970). *My World Line: An Informal Autobiography*. New York: Viking.
- Gurney, Ronald W. and Edward U. Condon (1928). Wave Mechanics and Radioactive Disintegration. *Nature* 122: 439.
- (1929). Quantum Mechanics and Radioactive Disintegration. *Physical Review* 33: 127–140.
- Heisenberg, Werner (1932a). Über den Bau der Atomkerne I. *Zeitschrift für Physik* 77: 1–11.
- (1932b). Über den Bau der Atomkerne II. *Zeitschrift für Physik* 78: 156–164.
- (1933a). Über den Bau der Atomkerne III. *Zeitschrift für Physik* 80: 587–596.
- (1933b). Considérations théoriques générales sur la structure du noyau. In: *Structure et propriétés des noyaux atomiques. Rapports et discussions du septième conseil de physique tenu à Bruxelles du 22 au 29 octobre 1933*. Ed. Institut International de Physique Solvay. Paris: Gauthier-Villars. 289–323.
- Institut International de Physique Solvay (1934). *Structure et propriétés des noyaux atomiques. Rapports et discussions du septième conseil de physique tenu à Bruxelles du 22 au 29 octobre 1933*. Paris: Gauthier-Villars.
- Koestler, Arthur (1964). *The Act of Creation*. New York: Macmillan.
- Majorana, Ettore (1933). Über die Kerntheorie. *Zeitschrift für Physik* 82: 137–145.
- Meitner, Lise and Otto R. Frisch (1939). Disintegration of Uranium by Neutrons: A New Type of Nuclear Reaction. *Nature* 143: 239–240.
- Nuclear Physics (1938). *Nature Supplement*. 142. (September 17). 520–522.
- Peierls, Rudolf (1981). Otto Robert Frisch. 1 October 1904–22 September 1979. *Biographical Memoirs of Fellows of the Royal Society of London* 27: 283–306.
- ed. (1986). *Niels Bohr Collected Works*. 9. *Nuclear Physics (1929–1952)*. Amsterdam: Elsevier.
- Reale Accademia d’Italia (1932). *Convegno di Fisica Nucleare, Ottobre 1931–XI*. Roma: Reale Accademia d’Italia.

- Rozental, Stefan, ed. (1967). *Niels Bohr: His Life and Work as Seen by His Friends and Colleagues*. Amsterdam: Wiley.
- Rutherford, Ernest F., Francis W. Aston, James Chadwick, Charles D. Ellis, George Gamow, Ralph H. Fowler, Owen W. Richardson, and Douglas R. Hartree (1929). Discussion on the Structure of Atomic Nuclei. *Proceedings of the Royal Society A* 123: 262–267.
- Sime, Ruth Lewin (1990). Lise Meitner's Escape from Germany. *American Journal of Physics* 58: 262–267.
- (1996). *Lise Meitner: A Life in Physics*. Berkeley: University of California Press.
- Stuewer, Roger H., ed. (1979). *Nuclear Physics in Retrospect: Proceedings of a Symposium on the 1930s*. Minneapolis: University of Minnesota Press.
- (1985). Bringing the News of Fission to America. *Physics Today* 38(10): 2–10.
- (1986). Gamow's Theory of Alpha-Decay. In: *The Kaleidoscope of Science*. Ed. Edna Ullmann-Margalit. 1. *The Israel Colloquium: Studies in History, Philosophy, and Sociology of Science*. Dordrecht: D. Reidel. 147–186.
- (1994). The Origin of the Liquid-Drop Model and the Interpretation of Nuclear Fission. *Perspectives on Science* 2: 39–92.
- Ullmann-Margalit, Edna, ed. (1986). *The Kaleidoscope of Science*. 1. *The Israel Colloquium: Studies in History, Philosophy, and Sociology of Science*. Dordrecht: D. Reidel.
- von Weizsäcker, Carl Friedrich (1935). Zur Theorie der Kernmassen. *Zeitschrift für Physik* 96: 431–458.
- (1936). Über die Spinabhängigkeit der Kernkräfte. *Zeitschrift für Physik* 102: 572–602.
- (1938). Neuere Modellvorstellungen über den Bau der Atomkerne. *Die Naturwissenschaften* 26: 209–217, 225–230.

The Challenges of Quantum Field Theory

Chapter 10

Tsung-Sui Chang's Contribution to the Quantization of Constrained Hamiltonian Systems

Xiaodong Yin, Zhongyuan Zhu, and Donald C. Salisbury

The quantization of constrained systems is one of the cornerstones of modern elementary dynamical theories. Important fundamental physical theories, such as quantum electrodynamics, quantum chromodynamics, electro-weak unified theory and string theories make use of it. The most widely used, currently canonical formulation for the quantization of constrained Hamiltonian systems was proposed by Paul Dirac in 1950 (Dirac 1950; 1951; 1964), and independently by Peter Bergmann and his collaborators (Bergmann and Brunings 1949; Bergmann, Penfield, et al. 1950; Anderson and Bergmann 1951). Later, in 1967, Ludvig Fadeev and Victor Popov made important progress in the path-integral quantization of the Yang-Mills field (Fadeev and Popov 1967).

In this paper, we emphasize the contribution of a Chinese theoretical physicist, Professor Tsung-Sui Chang, to this topic. In 1946, Chang pointed out that the previous canonical formulations of constrained systems could not be applied to quantum theory because they did not provide a method for dealing with one of the key features of the analogous classical theories—the appearance of undetermined multipliers. Chang worked out a feasible quantization procedure for such systems.

In the following section, we present a summary of Chang's education, training, and professional development from the 1930s to his death in 1969. In subsequent sections, we outline theoretical developments in the field leading up to Chang's own advances.

10.1 Biographical Overview

Chang was born in Hangzhou, Zhejiang Province, on 12 July 1915. He studied physics at Yenching University in 1930, then in 1931 he joined the Physics Department of Tsinghua University, headed by Wu Youxun. It was one of the most prestigious universities in China. In 1934, Chang began a masters degree pro-

gram under Wu's supervision. Wu recommended that he continue his training at Cambridge University.

In August of 1936, Chang entered the Mathematics Department of Cambridge University, supported by the Boxer Indemnity. He studied statistical physics as a doctoral student under Ralph Howard Fowler. Chang completed several important works on cooperative phenomena (solid solution, adsorption). The well-known text "Statistical Thermodynamics" by Fowler and Edward Guggenheim (1939) includes a section "The Combinatory Formulae of Chang."

After receiving his doctorate at Cambridge in 1938, Chang decided to extend his research field beyond statistical physics to quantum field theory. In 1938, Fowler endorsed Chang's application to Bohr:¹

I think I can whole-heartedly recommend him to you. He has done very well in his two years in Cambridge showing very considerable initiative and skill in developing the formal consequences such as order and disorder in alloys. I think you would find him very pleasant to deal with, and thoroughly industrious and able.

In 1939, Chang went to the Theoretical Physics Institute at the University of Copenhagen as a postdoctoral fellow and commenced his research on quantum field theory. His academic career began in earnest with stays at different locations in Europe, including: Copenhagen (September 1938 to February 1939), Zurich (February to June 1939), and Paris (June to October 1939). His acquaintances included Wolfgang Pauli, Niels Bohr and Aage Bohr. In Copenhagen, Chang lived in Bohr's home and established a very good relationship with Bohr's family. He completed two articles, "The Azimuthal Dependence of Processes Involving Mesons" (Chang 1940) was published in 1940, then another article on the nature of pseudo-scalar mesons. The latter publication was delayed until 1942 due to Chang's 1939 return to China and wartime communication difficulties.

Once back in China, Chang became the youngest professor of physics at the National Central University in Chongqing. For the next six years, during the Sino-Japanese War, Chang continued his research on statistical physics and quantum field theory, including the quantization of constrained systems. He published about ten articles during the difficult war period. Meanwhile, he was eager to pursue an international academic exchange. Chang's hope was fulfilled at the end of 1945 when he had the opportunity to return to Cambridge, thanks to Joseph Needham, who was the head of the Sino-British Science Cooperation Bureau in China, and Dirac. The Bureau provided information and aid for institutions and universities in China in wartime by sending papers to Western journals, offering scientific instruments and sending Chinese scholars to the United Kingdom.

¹Correspondence by Fowler to Bohr on 8 June 1938, Niels Bohr Archive.

During 1944–1946, the Sino-British Science Cooperation Bureau assisted eight Chinese professors in going abroad, including Chang.

In addition to his connection to Bohr, Chang had a close relationship with Dirac, one of Fowler's previous students. Dirac had already become a well-known professor during Chang's first visit to Cambridge. Dirac's research style served as a model for Chang. It was under Dirac's influence that Chang undertook the study of quantum field theory in 1939. Further, Dirac recommended that Chang teach a Cambridge course on quantum field theory,² and they had many discussions on this and other topics.

Chang's first two papers on constrained systems published in Britain were communicated by Dirac, and in the third paper, published in 1947, he thanked Dirac for his interest and discussions.

In the autumn of 1947, Dirac went to the Princeton Institute for Advanced Studies for a short-term visit. He suggested that Chang join him there. Chang spent six months with Dirac in New Jersey. He was then invited to work at the Carnegie Institute of Technology in Pittsburgh, Pennsylvania, for five months.

In the autumn of 1949, Chang left the United States and returned again to China. He successively became a professor in the Physics Department of Peking University, Beijing Normal University and the Institute of Mathematics of the Chinese Academy of Sciences. He also became an academician of the Chinese Academy of Sciences in 1957. He is recognized as one of the founders of quantum field research in China. Despite his successes, he suffered during the Cultural Revolution period and committed suicide on 30 June 1969 at 54 years of age.

10.2 Studies on Constrained Hamiltonian Systems before Chang's Work

10.2.1 Initial Studies

To better understand Chang's contributions, we now turn to initial studies of constrained systems. The challenge of quantizing a classical constrained dynamical model was present at the birth of quantum field theory. Classical electromagnetism is such a theory, so efforts to describe the quantum interaction of the field with charged particles needed to address the problem of constraints directly. Historically, the first such successful theory, with full quantum electromagnetic interaction, was introduced by Dirac in 1927 (Dirac 1927). With regard to the electromagnetic field itself, Dirac followed a method advanced by Pascual Jordan in 1926 in a joint publication with Born and Heisenberg (Born, Jordan, and

²Cambridge University Reporter, Reporter issues for the academic year 1936–1937, Vol. 67, 682; 1937–1938, Vol. 68, 633.



Figure 10.1: Tsung-Sui Chang, 1915–1969, Source: Prof. Yi Ci Chang.

Heisenberg 1926). The basic method is to decompose the radiation field into harmonic oscillators, so the quantization of the electromagnetic field was reduced to the quantization of these oscillators. The Dirac scheme dealt exclusively with transverse components of the field and was therefore not obviously relativistically covariant. Subsequently, in 1929, Heisenberg and Pauli established a canonical quantization procedure for general quantum fields (Heisenberg and Pauli 1929; 1930). However, when applying their method to the electromagnetic field, they encountered a stubborn difficulty that was eventually overcome by Heisenberg: the classical momentum conjugate to the scalar potential of electromagnetic fields vanishes identically. This means that the related canonical degrees of freedom were not independent. The canonical variables were subject to constraints. The immediate consequence was the contradictory conclusion that the commutator of this vanishing momentum with the scalar potential would not vanish. Thus, a procedure was needed to avoid this contradiction.

10.2.2 Earlier Approaches to Constrained Systems by Rosenfeld and Dirac

Obvious problems existed in early approaches to the quantization of the electromagnetic field.³ The first Heisenberg-Pauli method added a new term to the

³See (Salisbury 2009a) for a detailed analysis.

Lagrange function, multiplied by a small parameter ϵ . This had the effect that the momentum no longer vanished. However, the ϵ term destroyed the manifest gauge invariance of the Lagrangian. The second Heisenberg-Pauli method set the scalar potential A_0 equal to zero, thus destroying manifest Lorentz covariance. Gauss's law was then imposed as an initial condition on quantum states. In this same paper, they showed that a method that had—in the meantime—been put forth by Fermi was equivalent to adding a Lorenz gauge-fixing term to the Lagrangian, and this gauge condition also needed to be imposed as an initial quantum condition (Fermi 1929; Heisenberg and Pauli 1930).

Pauli invited the young Rosenfeld to join him in Zurich in 1929 to establish a firmer theoretical foundation for the methods that he, Heisenberg and Fermi had employed in their treatment of quantum electromagnetic field theory. Rosenfeld set himself the task of formulating a Hamiltonian procedure for dealing with local gauge symmetries in the two fundamental interactions that were known at the time, electromagnetism and Einstein's curved space-time gravitation. From the start he focused on the problem of implementing the full gauge symmetry group as a canonical transformation group acting on the phase space field variables. He made enormous progress in this effort, although his results were largely unknown (or in the case of Dirac, perhaps forgotten) by subsequent researchers. There is no indication that Chang was acquainted with Rosenfeld's work, although we do know that Dirac was aware of it in 1932 (Salisbury 2009a).

Rosenfeld showed that, as a consequence of local Lagrangian symmetries, identities arise that relate the canonical momenta and configuration variables (when the former are understood as functions of the configuration variables and their time derivatives). Following Rosenfeld, we represent the ensuing constraining relations as $F^r(p, q) = 0$, where the index r ranges over the total number of such so-called primary constraints. It followed that the time development for given initial conditions was not unique. Furthermore, Rosenfeld showed that this time development could be represented in first-order Hamiltonian form, with a Hamiltonian

$$H = H_0(p, q) + \lambda^r F_r(p, q), \quad (10.1)$$

where the λ^r are arbitrary space-time functions.⁴ The Hamilton equations are then

$$\frac{dq^\alpha}{dt} = \frac{\partial H_0}{\partial p_\alpha} + \lambda^r \frac{\partial F_r}{\partial p_\alpha} \quad (10.2)$$

⁴We assume here for the sake of simplicity that the number of physical variables q is finite. The extension to field theory is straightforward.

and

$$\frac{dp_\alpha}{dt} = -\frac{\partial H_0}{\partial q^\alpha} - \lambda^r \frac{\partial F_r}{\partial q^\alpha}. \quad (10.3)$$

Even though Rosenfeld presented an explicitly q -number version of his formalism in his 1930 article, he did not address the question of how one would or could incorporate the arbitrary functions λ^r into the quantum theory. However, as pointed out elsewhere (Salisbury 2009a), he had all the tools required to construct gauge invariant objects using his symmetry group generators. In a review of quantum electromagnetism published two years later (Rosenfeld 1932), he simply reverted to the Fermi scheme, after having convinced himself that his prior analysis justified the procedure. Curiously, however, he did not express this conviction in writing.

In 1933, Dirac published a paper entitled “Homogenous Variables in Classical Dynamics” in which he considered a far narrower class of models than Rosenfeld had examined, not citing him even though it is clear from an exchange of letters in 1932 that Dirac was familiar with Rosenfeld’s work. Dirac wrote in this paper:

The well-known methods of classical mechanics, based on the use of a Lagrangian or Hamiltonian function, are adequate for the treatment of nearly all dynamical systems met with in practice. There are, however, a few exceptional cases to which the ordinary methods are not immediately applicable. For example, the ordinary Hamiltonian method cannot be used when the momenta p_r , defined in terms of the Lagrangian function L by the usual formulae $p_r = \partial L / \partial \dot{q}_r$, are not independent functions of the velocities. (Dirac 1933)

Dirac used the electromagnetic field and the massless relativistic particle as examples to illustrate his program. The outcome is the same Hamilton equations exhibited by Rosenfeld, with the same arbitrary functions (though designated by Dirac by ρ rather than λ). Referring to Dirac’s paper, in 1946 Chang observed that the appearance of these arbitrary functions seemed to preclude passage to a quantum theory.

10.3 Chang’s Contributions to Hamiltonian Systems

From July 1944 to June 1946, Chang published three papers on the quantization of constrained systems: “A Note on the Hamiltonian Theory of Quantization” (Chang 1945), “A Note on the Hamiltonian Equations of Motion” (Chang 1946), and “A Note on the Hamiltonian Theory of Quantization (II)” (Chang 1947).

The first and second papers were completed under very difficult conditions in Chongqing, a southwestern city in China, during the Sino-Japanese War. They were published in the Proceedings of the Royal Society of London and Proceedings of the Cambridge Philosophical Society, respectively, and were communicated by Dirac. In the second and third papers, Chang expressed his thanks to Dirac for discussions. The third paper, completed at Cambridge University, was the most extensive, summarizing some results from the previous two, and is the principal subject of our analysis. This 1947 paper is divided into three sections in which Chang discusses the need for dealing with the arbitrary functions λ , a proposal for quantizing models in which constraints are imposed at the Lagrangian level through the use of Lagrange multipliers, and a proposal for quantizing systems with primary constraints.

First, Chang observed, in referring to Dirac's 1933 paper,

The Lagrangian equations for cases with missing momenta have been studied some time ago by Dirac by making use of homogeneous variables. It was shown that the equations of motion can always be put in canonical form. However, the final equations still contain quantities of the nature of unknown Lagrange multipliers, and are thus not suitable for passing to a quantum theory. (Chang 1947)

This may be the first published observation of the challenge that the undetermined functions posed for the canonical quantization program.

10.3.1 Models with Lagrange Multipliers

The first models that Chang considered were models in which constraints were imposed "by hand" through the use of Lagrange multipliers. He considered systems for which the Lagrangian contained only first derivatives $q_{,\mu}^{\alpha} := \frac{\partial q^{\alpha}}{\partial x^{\mu}}$. He supposed that variations of the action

$$I := \int L(q^{\alpha}, q_{,\mu}^{\alpha}, x) d^4x, \quad (10.4)$$

were subject to f auxiliary conditions

$$f^{(\xi)}(q^{\alpha}, q_{,\mu}^{\alpha}, x) = 0 \quad (\xi = 1, 2, \dots, f). \quad (10.5)$$

Then it followed that

$$\frac{\partial L}{\partial q^{\alpha}} + \sum \mu^{(\xi)} \frac{\partial f^{(\xi)}}{\partial q^{\alpha}} - \frac{\partial}{\partial x^{\mu}} \left(\frac{\partial L}{\partial q_{,\mu}^{\alpha}} + \sum \mu^{(\xi)} \frac{\partial f^{(\xi)}}{\partial q_{,\mu}^{\alpha}} \right) = 0. \quad (10.6)$$

Eqs. (10.5) and (10.6) are the field equations. The Lagrange multipliers μ_ξ are understood to depend on the space-time coordinates, represented collectively by the symbol x , where we will take them to be real, with $x^\mu = \{ct, \vec{x}\}$.⁵ Superscripts μ, ν, \dots run from 1 to 4, while superscripts r, s, \dots go from 1 to 3. The q^α are the dynamical field variables.

Chang claimed to have achieved a first-order canonical form for his field equations by first implicitly defining the functions $b^\alpha(p_\alpha, q^\alpha, q_{,r}^\alpha, x)$ and $\eta_\xi(p_\alpha, q^\alpha, q_{,r}^\alpha, x)$ through the relations

$$f^{(\xi)}(q^\alpha, q_{,r}^\alpha, b^\alpha, x) = 0 \quad (10.7)$$

and

$$p_\alpha - \frac{\partial}{\partial b^\alpha} L(q^\alpha, q_{,r}^\alpha, b^\alpha, x) - \eta_\xi \frac{\partial}{\partial b^\alpha} f^{(\xi)}(q^\alpha, q_{,r}^\alpha, b^\alpha, x) = 0, \quad (10.8)$$

where a sum over ξ is understood. He then defined the Hamiltonian to be

$$H(p_\alpha, q^\alpha, q_{,r}^\alpha, x) := -L(q^\alpha, q_{,r}^\alpha, b^\alpha, x) + p_\alpha b^\alpha, \quad (10.9)$$

where a sum over α is understood.

Using this Hamiltonian function, the canonical equations became

$$\frac{\partial q^\alpha}{\partial t} = \frac{\delta \tilde{H}}{\delta p_\alpha(\vec{x})} = \frac{\partial H}{\partial p_\alpha} \quad (10.10)$$

and

$$\frac{\partial p_\alpha}{\partial t} = -\frac{\delta \tilde{H}}{\delta q^\alpha(\vec{x})} = -\left(\frac{\partial H}{\partial q^\alpha} - \frac{\partial}{\partial x^r} \frac{\partial H}{\partial q_{,r}^\alpha} \right), \quad (10.11)$$

where $\tilde{H} := \int H d^3x$.

Substituting for H we obtain

$$\frac{\partial q^\alpha}{\partial t} = b^\alpha + \eta_\xi \frac{\partial f^{(\xi)}}{\partial b^\beta} \frac{\partial b^\beta}{\partial p_\alpha} \quad (10.12)$$

⁵Chang used $x_4 := ict$.

and

$$\begin{aligned} \frac{\partial p_\alpha}{\partial t} &= \frac{dL(q^\alpha, b^\alpha, x)}{dq^\alpha} + \eta_\xi \frac{df^{(\xi)}(q^\alpha, b^\alpha, x)}{dq^\alpha} \\ - \frac{\partial}{\partial x^r} \left\{ \frac{dL(q^\alpha, b^\alpha, x)}{dq_{,r}^\alpha} + \eta_\xi \frac{df^{(\xi)}(q^\alpha, b^\alpha, x)}{dq_{,r}^\alpha} \right\}, \end{aligned} \quad (10.13)$$

where we define the total derivative with respect to q^α as

$$\frac{dL(q^\alpha, b^\alpha, x)}{dq^\alpha} := \frac{\partial L(q^\alpha, b^\alpha, x)}{\partial q^\alpha} + \frac{\partial L(q^\alpha, b^\alpha, x)}{\partial b^\beta} \frac{\partial b^\beta}{\partial q^\alpha}. \quad (10.14)$$

Chang claimed that eq. (10.12) yielded

$$\frac{\partial q^\alpha}{\partial x_4} = b^\alpha,$$

failing to mention that this form is achievable only by requiring that

$$\eta_\xi \frac{\partial f^{(\xi)}}{\partial b^\beta} \frac{\partial b^\beta}{\partial p_\alpha} = 0.$$

He also did not observe that additional restrictions regarding the q^α -dependence of the constraining relations (10.5) arise by requiring that the dynamical eq. (10.13) be equivalent to the Euler-Lagrange field eq. (10.6).

Generally this means that there are severe limitations to the applicability of Chang's method. We can, however, give a simple illustrative example in which the procedure may be implemented. We consider a system with

$$L = \frac{1}{2} \left(\frac{dq}{dt} \right)^2$$

and an auxiliary condition

$$f = \frac{dq}{dt} = 0.$$

Then, employing Chang's symbols, we have

$$\begin{aligned}
 f &= b = 0, \\
 p &= \frac{\partial L}{\partial b} + \eta \frac{\partial f}{\partial b} = b + \eta = \eta, \\
 H &= pb - \frac{1}{2}b^2 = 0, \\
 \{q, p\}_{P.B.} &= 1, \\
 \dot{q} &= \frac{\partial H}{\partial p} = 0, \\
 \dot{p} &= -\frac{\partial H}{\partial q} = 0.
 \end{aligned}$$

We note also that the relevant additional consistency condition that is required to achieve the correct Euler-Lagrange equations is

$$\eta \frac{df(q, b)}{dq} = \eta \left(\frac{\partial f}{\partial q} + \frac{\partial f}{\partial b} \frac{\partial b}{\partial q} \right) = 0. \quad (10.15)$$

This does vanish since f is independent of q , and b vanishes identically. So although we have a non-trivial model with a Lagrange multiplier that leads to a self-consistent quantum theory here, it is dubious that the procedure could enjoy widespread applicability.

10.3.2 Models with Missing Momenta

Chang next discussed models which possessed missing momenta, i.e., models that in our current terminology have primary constraints. Chang identified those momenta as “missing” that vanished due to the absence of the time derivative of the conjugate configuration variable in the Lagrangian. He assumed that a transformation of these variables had been undertaken so that all primary constraints would take this simple form, thus dividing the configuration variables into a set q^α such that the q_0^α could be written in terms of the conjugate momenta p_α , and another set Q^l with momenta $P_l = 0$. Chang appears to have been the first to recognize in print that this decomposition was achievable at the Lagrangian level, although he did not note that to implement it, it might be necessary to add total derivative terms to the Lagrangian. Dirac added such terms for general relativity in 1958 (Dirac 1958).

In terms of these variables, the Lagrangian field equations take the form

$$\frac{\partial L}{\partial q^\alpha} - \frac{\partial}{\partial x^\mu} \left(\frac{\partial L}{\partial q_{,\mu}^\alpha} \right) = \frac{\delta \tilde{L}}{\delta q^\alpha} - \frac{\partial p_\alpha}{\partial t} = 0 \tag{10.16}$$

and

$$f_l(q^\alpha, Q^l, q_{,0}^\alpha, x) := \delta \tilde{L} / \delta Q^l = 0, \tag{10.17}$$

where $\tilde{L} := \int L d^3x$. Thus Chang recognized that the relations (10.17) were themselves constraints, in addition to the primary constraints.⁶

He then presented a procedure about which he explicitly recognized that it was applicable only when the constraints could be employed to eliminate the Q^l as independent variables. That is, he assumed that the relations

$$p_\alpha - \frac{\partial L(q^\alpha, B^l, b^\alpha, x)}{\partial b^\alpha} = 0 \tag{10.18}$$

and

$$f_l(q^\alpha, B^l, b^\alpha, x) = 0 \tag{10.19}$$

could be solved for

$$Q^l =: B^l(q^\alpha, q_{,r}^\alpha, p_\alpha, x). \tag{10.20}$$

Then under these circumstances, the Hamiltonian

$$H(p_\alpha, q^\alpha, x) = -L(q^\alpha, B^l, b^\alpha, x) + p_\alpha b^\alpha$$

delivers

$$\frac{\partial q^\alpha}{\partial t} = \frac{\partial H}{\partial p_\alpha} = b^\alpha(q, q_{,r}, p, x) \tag{10.21}$$

and

$$\frac{\partial p_\alpha}{\partial t} = -\frac{\delta \tilde{H}}{\delta p_\alpha(\vec{x})} \tag{10.22}$$

and is equivalent to the original Lagrangian field equations.

Thus Chang recognized correctly that systems may exist where the constraints can be solved. And in this case, upon passing to the quantum theory, one can have canonical Hamiltonian equations and commutation relations just for the

⁶In the following section, Chang showed that in gauge covariant models these constraints arise due to the demand that primary constraints be preserved under time evolution.

independent variables p_α and q^α with the Q^l becoming functions of p_α, q^α . He did not observe that this circumstance arises when the constraints have non-vanishing Poisson brackets among themselves, or in the language that was introduced by Dirac in 1950, when the constraints are second class (Dirac 1950).

The models for which the constraints cannot be solved belong to another type. Chang also discussed a limited version of this case, which we address in the next subsection. This is the situation with gauge theories for which the constraints do have vanishing Poisson brackets among themselves. Dirac called such constraints first class. So although Chang did not characterize them in this way, he was certainly aware that there existed two kinds of constrained systems, and he presented a preliminary procedure for quantizing them.

Dirac was one of the first to propose a quantization procedure that dealt with both types. In gauge theories, he noted that gauge conditions need to be invoked, and he introduced modified brackets that respected these conditions (Dirac 1950). It seems likely that Chang's work would have influenced Dirac's movement in this direction.

10.3.3 Gauge Covariant Models

Having realized that his program for second class constraints did not work for systems like electromagnetism, Chang then embarked on an alternate approach. He considered a system in which the Lagrangian L is such that some of the p_α are missing, but with the assumption that L is increased by an amount $F^L(x)$ under the transformation

$$q^\alpha(x) \rightarrow q^\alpha(x) + F^\alpha(x).$$

We represent the resulting variations with the symbol $\bar{\delta}$, writing $\bar{\delta}q^\alpha = F^\alpha$. The resulting variation of the Lagrangian is

$$\bar{\delta}L = \frac{\partial L}{\partial q^\alpha} F^\alpha + \frac{\partial L}{\partial q_{,\mu}^\alpha} F_{,\mu}^\alpha =: F^L. \quad (10.23)$$

Chang remarks that the variation he is considering "is not the same as an ordinary gauge transformation," but he calls it a gauge transformation "for lack of a better name." Yet he does assume that his Euler-Lagrange equations are covariant under this transformation. This result follows immediately from the assumption that F^L is a function only of x (and not of the dynamical variables). He is perhaps assuming, at least initially, that the F^α do not depend on arbitrary space-time functions. Also, general gauge transformations have a dependence on the dynamical variable, i.e., $\bar{\delta}q^\alpha = F^\alpha(q^\beta, q_{,\mu}^\beta, x)$. Such is the case, for example,

with general relativity and also with the homogeneous models that were treated by Dirac (1933). In any case, the electromagnetic model that he cites specifically as susceptible to his analysis is clearly gauge covariant in the current sense.

Next, Chang points out that when the field equations are satisfied, one obtains, after performing a spatial integration by parts and letting the gauge variations vanish at spatial infinity,

$$\frac{\partial}{\partial t} \int d^3x \left(F^\alpha \frac{\partial L}{\partial q_{,0}^\alpha} \right) - \int d^3x F^L = 0. \quad (10.24)$$

At this point, Chang assumes that the F^α involve arbitrary functions Φ of the time up to order u , and this leads him to a remarkable result: it follows from eq. (10.24) that the coefficients of each order of time derivative must separately vanish. This result was already known, to be sure, to Rosenfeld in 1930—but it was independently rediscovered in 1951 by James L. Anderson and Bergmann and is generally attributed to them (Anderson and Bergmann 1951). Indeed, these authors introduced the terminology that is still in use today. The requirement that primary constraints be preserved under time evolution may lead to secondary constraints. These in turn may lead to tertiary constraints, and so on.

Chang notes that the highest derivative term, $\frac{\partial^u \Phi}{\partial t^u}$, that appears in eq. (10.24) will not involve a time derivative of momenta. This term is

$$\int d^3x \left(\frac{\partial F^\alpha}{\partial t} p_\alpha \right), \quad (10.25)$$

and isolating the coefficient of $\frac{\partial^u \Phi}{\partial t^u}$ in the integrand, we deduce the existence of primary constraints that Chang represents as $g_u(q, p) = 0$. But then the coefficient of $\frac{\partial^{u-1} \Phi}{\partial t^{u-1}}$ will be in the form

$$g_{u-1}(p, p) := \frac{\partial g_u(q, p)}{\partial t} - C_{u-1}(q, p) = 0. \quad (10.26)$$

In other words, the preservation of the primary constraint requires the existence of a secondary constraint $C_{u-1}(q, p) = 0$, and so on. Thus eq. (10.24) may be rewritten as

$$\int d^3x \left\{ g_0(p, p, 0) \Phi + g_1(p, p, 0) \frac{\partial \Phi}{\partial t} + \dots + g_u(p, p, 0) \frac{\partial^u \Phi}{\partial t^u} \right\} = 0 \quad (10.27)$$

with

$$g_0 = g_1 = \dots g_u = 0. \quad (10.28)$$

Actually, Chang did not state explicitly that constraints C would arise, but it is an immediate consequence of the relation (10.24).

Although Chang did not provide a concrete example, it is instructive to take the electromagnetic field as an example to illustrate his formulation—and identify one shortcoming. As mentioned above, Chang did note that the electromagnetic model satisfied his assumptions regarding the gauge invariance of the Lagrangian. This is true, provided that the charged source current is understood to be a non-dynamical external field, as we highlight below.

The Lagrangian for this model is

$$L = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + j^\mu A_\mu, \quad (10.29)$$

where

$$F_{\mu\nu} := A_{\nu,\mu} - A_{\mu,\nu} \quad (10.30)$$

and j^μ is an external current. Then the conjugate momenta are

$$p^a = \frac{\partial L}{\partial \dot{A}_a} = F^{a0} = -\dot{A}_a - A_{0,a}, \quad (10.31)$$

while $p^0 =: P$ vanishes identically.

The Lagrangian equations of motion are

$$\partial_0 F^{0a} + \partial_b F^{ba} + j^a = 0 \quad (10.32)$$

and

$$\partial_a F^{a0} + j^0 = 0. \quad (10.33)$$

The gauge variation is

$$\bar{\delta}A_\mu = \Phi_{,\mu}. \quad (10.34)$$

The variation of the Lagrangian under this transformation is

$$\bar{\delta}L = F^{\mu\nu}\Phi_{,\mu\nu} + j^\mu\Phi_{,\mu} = j^\mu\Phi_{,\mu} =: F^L. \quad (10.35)$$

Thus, provided that j^μ is a prescribed function of x , the variation may be written as a total time derivative, and the variation is a true gauge transformation.

Performing integrations by parts of the action, and letting the variations vanish at spatial infinity, we find that if the equations of motion are satisfied, then

$$\frac{\partial}{\partial t} \int d^3x (P\dot{\Phi} - p_{,a}^a \Phi) = - \int d^3x \partial_{aj}^a \Phi + \int d^3x j^0 \dot{\Phi} \quad (10.36)$$

or

$$\begin{aligned} \int d^3x [P\dot{\Phi} + (\dot{P} - p_{,a}^a - j^0) \dot{\Phi} - (\dot{p}_{,a}^a + j_{,a}^a) \Phi] &= 0 \\ &=: \int d^3x (g_2 \ddot{\Phi} + g_1 \dot{\Phi} + g_0 \Phi). \end{aligned} \quad (10.37)$$

Thus, the Gauss's law constraint $C := p_{,a}^a + j^0 = 0$ may be a consequence of the required vanishing of the time derivative of the primary constraint $P = 0$.

Chang showed generally that the consistency conditions (10.37) guarantee that if the equation of motion

$$\frac{\delta L}{\delta Q} = 0 \quad (10.38)$$

is fulfilled at the initial time, then it is fulfilled for all time. For the electromagnetic example, this means that if

$$\frac{\delta L}{\delta Q} = \vec{\nabla} \cdot \dot{\vec{A}} - \nabla^2 Q + j^0 = 0 \quad (10.39)$$

is fulfilled at an initial time, then it will be fulfilled at all future times. Chang therefore proposes that the variables $Q := A_0$ and P can be eliminated entirely by setting $Q = 0$, and therefore setting as an initial condition

$$\vec{\nabla} \cdot \dot{\vec{A}} = -j^0. \quad (10.40)$$

Finally in passing to the quantum theory, he imposes this classical initial condition as a condition on the quantum state Ψ ,

$$(\vec{\nabla} \cdot \vec{p} - j^0)\Psi = 0. \quad (10.41)$$

So the ultimate practical outcome of this analysis for quantum electrodynamics is that Chang provided another, somewhat more general proof of the legitimacy of the second quantization procedure proposed by Heisenberg and Pauli (1930).

10.4 Conclusion

In summary, we have identified the following five original contributions of Chang to the quantization of constrained systems:

1. He was the first to recognize in print, and to offer a preliminary resolution of, a problem for quantization posed by the appearance of arbitrary space-time functions in classical gauge theories.
2. Chang proposed a procedure for imposing quantum constraints, using Lagrangian multipliers, for a limited class of non-singular classical theories.
3. He recognized that constrained systems could exist in which the constraints could be solved, thereby entirely eliminating some canonical degrees of freedom. Canonical Poisson bracket relations of the remaining phase space variables could then be replaced by canonical quantum commutation relations. Furthermore, he showed that in these models, the functions that multiply the constraints become functions of the dynamical variables, and that they therefore become non-trivial operators in quantum theory. Dirac later discovered that these models possessed the property that the Poisson brackets of the constraints among themselves did not vanish.
4. He discovered for a limited class of classical gauge theories that the preservation of primary constraints under time evolution leads to additional constraints. This discovery has until recently been attributed to Anderson and Bergmann in 1951, even though the general proof was already demonstrated by Rosenfeld in 1930.
5. Chang developed a technique for quantizing a class of gauge covariant models that could be viewed in the case of electromagnetism as a somewhat more general proof of the legitimacy of the quantization procedure proposed by Heisenberg and Pauli in 1930. This method formulates the gauge condition as a restriction on initial quantum states.

There is a clear link between Chang's activities during 1945–1947 and Dirac's subsequent publications beginning three years later on constrained Hamiltonian systems. In future work, we intend to further investigate the detailed manner in which they influenced each other.

Abbreviations and Archives

Niels Bohr Archive	Bohr Scientific Correspondence, Fowler to Bohr. Folder 19. No. 380608
Cambridge University Reporter	Reporter issues for the academic year 1936–1937, Vol. 67; 1937–1938, Vol. 68

Acknowledgements

Xiaodong Yin was supported by NSFC under No.11075109. We would like to express our gratitude to the academician of the Chinese Academy of Science, Professor Yuanben Dai, for his detailed discussions.

List of Chang's Publications

Papers in English in International Journals:

- (1937) An Extension of Bethe's Theory of Order-Disorder Transitions in Metallic Alloys, *Proceedings of the Royal Society A*, 161, 546–563.
- (1937) Specific Heats of Solids due to Molecular Rotation, *Proceedings of the Cambridge Philosophical Society*, 33, 524–533.
- (1938) Superlattice Formations of the Type AB in Alloys with Applications to the Theory of Adsorption, *Proceedings of the Cambridge Philosophical Society*, 34, 224–237.
- (1939) Statistical Theory of the Adsorption of Double Molecules, *Proceedings of the Royal Society A*, 169, 512–531.
- (1939) The Number of Configurations of an Assembly with Long-Distance Order, *Proceedings of the Royal Society A*, 173, 48–58.
- (1939) Superlattice Formation of the Type AB in an Adsorbed Layer, *Proceedings of the Cambridge Philosophical Society*, 35, 70–83.
- (1939) The Number of Configurations in an Assembly and Cooperative Phenomena, *Proceedings of the Cambridge Philosophical Society*, 35, 265–292.
- (1940) The Azimuthal Dependence of Processes Involving Mesons, *Proceedings of the Cambridge Philosophical Society*, 36, 34–42.
- (1941) A Note on Bethe-Kirkwood's Partition Function of a Binary Solid Solution, *Journal of Chemical Physics*, 9, 169–174.

- (1941) Second Neighbour Interactions and the Critical Behaviours of Binary Solid Solution, *Journal of Chemical Physics*, 9, 174–176.
- (1942) Arrangements with Given Number of Neighbours (and C. C. Ho), *Proceedings of the Royal Society A*, 180, 345–365.
- (1942) Properties of Mesons Described by a Pseudo-Scalar Wave Function, *Det Kongelige Danske Videnskabernes Selskab Mathematisk-Fysiske Meddelelser*, 19, 17–19.
- (1943) Arrangements of Double Molecules on a Lattice, (and S. K. Kao), *Transactions of the Faraday Society*, 39, 288.
- (1944) The Impulse-Energy Tensor of Material Particles, I. Mesons and Electrons, *Proceedings of the Royal Society A*, 182, 302–318.
- (1945) A Note on the Hamiltonian Theory of Quantization, *Proceedings of the Royal Society A*, 183, 316–328.
- (1946) A Note on the Hamiltonian Equations of Motion, *Proceedings of the Cambridge Philosophical Society*, 42, 132–138.
- (1946) Quantum Electrodynamics with $\frac{\delta A_{\mu}}{\delta x_{\mu}} = 0$, *Proceedings of the Royal Society A*, 185, 192–206.
- (1947) A Note on the Hamiltonian Theory of Quantization (II), *Proceedings of the Cambridge Philosophical Society*, 43, 196–204.
- (1947) A Note on Relativistic Second Quantization, *Proceedings of the Cambridge Philosophical Society*, 43, 183–195.
- (1948) Field Theories with High Derivatives, *Proceedings of the Cambridge Philosophical Society*, 44, 76–86.
- (1948) Surface Effects in Cooperative Phenomena, *Proceedings of the Cambridge Philosophical Society*, 44, 598–600.
- (1949) Relativistic Field Theories, *Physical Review*, 75, 967.
- (1950) The Quantum Mechanics of Localizable Dynamical Systems, *Physical Review*, 78, 592.

Papers in Chinese, English and Russian in Chinese Journals:

- (1948) Remarks on the Theory of S Matrices, *National Peking University Semi-Centennial Papers*, 81–88.
- (1949) On Weiss' Theory of Fields, *Chinese Journal of Physics*, 7, 265–27 (in English).
- (1951) Relativistic Nature of Coulombian Interaction, *Chinese Journal of Physics*, 8, 123–130.
- (1953) Calculations of Some Operators in Relativistic Quantum Mechanics, Part I, *Acta Mathematica Sinica*, 3, 59–72.

- (1953) Calculations of Some Operators in Relativistic Quantum Mechanics, Part II, *Acta Mathematica Sinica*, 3, 73–85.
- (1954) A Remark on Boltzmann H Theorem, *Acta Physica Sinica*, 10, 1–12.
- (1956) A Remark on the Application of Mayer's Theory of Imperfect Gases to Regular Solid Solutions, *Acta Physica Sinica*, 12, 379–398.
- (1956) Classical Motion of Material Particles, *Journal of Beijing Normal University*, 1, 1–13.
- (1956) Direct Calculations of Spinor Matrices, *Journal of Beijing Normal University*, 1, 15–18.
- (1956) Reduction of Canonical Equations, *Acta Physica Sinica*, 12, 393–401.
- (1957) Free Energy of a Solid Solution on a Face-Centered Cubic Lattice (and H. C. Ting, T. K. Tsiao), *Acta Physica Sinica*, 13, 515–524.
- (1957) Note on Quantum Electrodynamics with High Derivatives, *Journal of Beijing Normal University*, 2, 53–60.
- (1957) Configurational Partition Function of Binary Solid Solutions, (and T. M. Wang, H. C. Hsu), *Acta Physica Sinica*, 13, 525–542.
- (1958) A Note on Interactions Representation, *Science Record*, 2, 250 (in English).
- (1958) A New Approximate Method of Constructing Configurational Partition Functions of Cooperative Assembles, *Scientia Sinica*, 7, 313.
- (1958) Some Account of Works of H. H. Bogoliubov and E. Lifshitz, *Acta Physica Sinica*, 14, 274–279.
- (1958) High Order Equations in Quantum Theory, *Acta Physica Sinica*, 14, 307.
- (1958) Normal Dependent Terms in Wave Equations in Interaction Representation, *Acta Physica Sinica*, 14, 308.
- (1958) A Note on the Ergodic Theorem for Quantized Systems, *Acta Physica Sinica*, 14, 400–404.
- (1958) Some Remarks on the Theory of Expansors, *Acta Physica Sinica*, 14, 405–422.
- (1959) The Configurational Partition Functions of Solid Solutions, *Acta Physica Sinica*, 15, 42–54.
- (1959) A Simplified Proof of Dispersive Relations, *Acta Physica Sinica*, 15, 609–615.
- (1959) Remarks on Chew-Low Equations, *Acta Physica Sinica*, 15, 616–623.
- (1959) Regular Solutions With Molecules Occupying Several Sites, *Acta Physica Sinica*, 15, 652–663.

- (1959) Configurational Partition Function for Regular Solid Solutions, *Scientia Sinica*, 8, 401.
- (1961) Analyticity of Perturbation Expansions, *Acta Physica Sinica*, 17, 1–17.
- (1961) Integral Representations of Commutators, *Acta Physica Sinica*, 17, 163–169.
- (1962) Analyticity of Perturbation Expansions, *Acta Physica Sinica*, 18, 91–117.
- (1964) Asymptotic Behavior of a Ladder, *Acta Physica Sinica*, 20, 378–380.
- (1964) Inelastic Amplitudes in the Lee Model Possessing Crossing Symmetries, *Acta Physica Sinica*, 20, 1090–1100.
- (1965) The Regge Cuts in a Perturbation Theory, *Acta Physica Sinica*, 21, 1381.
- (1965) On Singularities of Amplitudes in a Perturbation Theory (and F. T. Wu, T. C. Jiu), *Acta Physica Sinica*, 21, 1544–1551.
- (1965) Motion of Poles in the Lee Model, *Acta Physica Sinica*, 21, 1882–1888.

Dissertation (at Cambridge University):

- (1938) *Applications of Bethe's Method in Statistical Mechanics*, Cambridge University

References

- Anderson, James L. and Peter G. Bergmann (1951). Constraints in Covariant Field Theories. *Physical Review* 83: 1018–1025.
- Bergmann, Peter G. and Johanna H. M. Brunings (1949). Non-linear Field Theories II. Canonical Equations and Quantization. *Reviews of Modern Physics* 21: 480–487.
- Bergmann, Peter G., Robert Penfield, Ralph Schiller, and Henry Zatkis (1950). The Hamiltonian of the General Theory of Relativity with Electromagnetic Field. *Physical Review* 80: 81–88.
- Born, Max, Pascual Jordan, and Werner Heisenberg (1926). Zur Quantenmechanik II. *Zeitschrift für Physik* 35: 557–615.
- Chang, Tsung-Sui (1940). The Azimuthal Dependence of Processes Involving Mesons. *Proceedings of the Cambridge Philosophical Society* 36: 34–42.

- (1945). A Note on the Hamiltonian Theory of Quantization. *Proceedings of the Royal Society A* 183(994): 316–328.
- (1946). A Note on the Hamilton Equations of Motion. *Proceedings of the Cambridge Philosophical Society* 42: 132–138.
- (1947). A Note on the Hamiltonian Theory of Quantization II. *Proceedings of the Cambridge Philosophical Society* 43: 196–204.
- Dirac, Paul A. M. (1927). The Quantum Theory of the Emission and Absorption of Radiation. *Proceedings of the Royal Society A* 114: 243–265.
- (1933). Homogeneous Variables in Classical Dynamics. *Proceedings of the Cambridge Philosophical Society* 29: 389–401.
- (1950). Generalized Hamiltonian Dynamics. *Canadian Journal of Mathematics* 2: 129–148.
- (1951). The Hamiltonian Form of Field Dynamics. *Canadian Journal of Mathematics* 3: 1–23.
- (1958). The Theory of Gravitation in Hamiltonian Form. *Proceedings of the Royal Society A* 246: 333–343.
- (1964). *Lectures on Quantum Mechanics*. New York: Yeshiva University Press.
- Faddeev, Ludvig and Victor Popov (1967). Feynman Diagrams for the Yang-Mills Field. *Physics Letters B* 25: 29–30.
- Fermi, Enrico (1929). Sopra l'elettrodinamica quantistica. *Rendiconti Lincei* 9: 881–887.
- Fowler, Ralph H. and Edward A. Guggenheim (1939). *Statistical Thermodynamics*. Cambridge: Cambridge University Press.
- Heisenberg, Werner and Wolfgang Pauli (1929). Zur Quantendynamik der Wellenfelder. *Zeitschrift für Physik* 56: 1–61.
- (1930). Zur Quantentheorie der Wellenfelder II. *Zeitschrift für Physik* 59: 168–190.
- Rosenfeld, Léon (1930). Zur Quantelung der Wellenfelder. *Annalen der Physik* 5: 113–152.
- (1932). La théorie quantique des champs. *Annales de l'Institut Henri Poincaré* 2: 25–91.
- Salisbury, Donald (2009a). Léon Rosenfeld and the Challenge of the Vanishing Momentum in Quantum Electrodynamics. *Studies in History and Philosophy of Modern Physics* 40: 363–373.
- (2009b). Translation and Commentary of Léon Rosenfeld's "Zur Quantelung der Wellenfelder", *Annalen der Physik* 397, 113 (1930). *Max Planck Institute for the History of Science* Preprint 381, November 2009. <http://www.mpiwg-berlin.mpg.de/en/resources/preprints.html>.

Chapter 11

Feynman's Struggle and Dyson's Surprise: The Development and Early Application of a New Means of Representation

Adrian Wüthrich

Around the year 1948, Richard Phillips Feynman (1918–1988) began to use a particular kind of diagram for the theoretical treatment of recalcitrant problems in the theory of quantum electrodynamics (QED), that is, the calculation of the self-energy of the electron. Soon thereafter, these so-called Feynman diagrams became a ubiquitous tool in theoretical elementary particle physics.

In this contribution, I first briefly sketch how Feynman diagrams are used today, how they are most often interpreted and how their genesis is usually described, see sec. (11.1). In the second part I present my reconstruction of how Feynman, starting from a search for an appropriate interpretation of the “Dirac equation” (Dirac 1928), arrived at an innovative representation of quantum electrodynamic phenomena, see sects. (11.2)–(11.5).

My reconstruction of Feynman's struggle is based on manuscript pages which the Archives of the California Institute of Technology (Caltech) kindly made accessible to me. Some of these manuscript pages can also be found in Silvan Schweber's book on QED (Schweber 1994) and in some of his articles, for example (Schweber 1986). I hope, however, that my reconstruction of the material reveals more clearly how Feynman's diagrammatic representations were the means by which he defined and further developed physical models to interpret theoretical equations.

The third and last part is concerned with Freeman Dyson's systematization and theoretical updating of Feynman's framework, see sec. (11.6).¹ I end with a comparison with two other case studies concerning developments of concepts and means of representation, see sec. (11.7).

¹A more detailed account of what is described in sects. (11.1) to (11.6) can be found in (Wüthrich 2010). Some passages and figures from that publication are reproduced here with kind permission of Springer Science and Business Media.

11.1 Development, Modern Application and Interpretation of Feynman Diagrams

The modern application of Feynman diagrams goes something like this: To calculate, for instance, the probability amplitude for the scattering of an electron and a positron, we first draw a line for each incoming and outgoing particle, see fig. (11.1(a)). We read the diagram from the bottom up. For electrons (e^-), we indicate on the lines whether the particle is coming in or going out using arrows. For positrons (e^+), which are the antiparticles of the electrons, we use lines which have their arrow pointing in the opposite direction as for electrons.

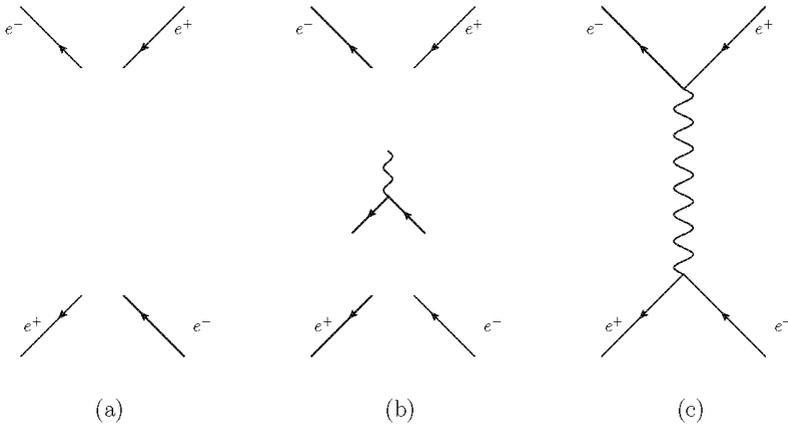


Figure 11.1: The modern application of Feynman diagrams.

The fundamental element of a quantum electrodynamic interaction is the absorption and emission of a light-quantum by an electron or a positron, which is represented by a point in which two solid lines and a wavy line end or start, see fig. (11.1(b)). Two such elementary interactions suffice to bring about the interaction in this example, see fig. (11.1(c)). Other diagrams, most of them more complex, would complete the representation of the elastic scattering of an electron and a positron using Feynman diagrams.

The numerical value for the reaction rate is obtained by translating the Feynman diagrams into mathematical expressions according to the so-called Feynman rules. In fig. (11.2), we see how Michael E. Peskin and Daniel V. Schroeder show, in their 1995 textbook, what graphical element corresponds to what mathematical expression (Peskin and Schroeder 1995).

While most of the authors on Feynman diagrams warn us of incorrectly interpreting the diagrams in terms of, for instance, particle trajectories, some of them go further and deny any physical interpretation whatsoever. James Robert Brown, for example, maintains such a position and claims: “We see the lines in the diagram; we do not visualize the physical process itself, nor any sort of abstract version of it” (Brown 1996, 267).

$$\begin{array}{l}
 \begin{array}{c} \text{---} \leftarrow \\ p \end{array} \\
 \begin{array}{c} \text{~~~~~} \\ \leftarrow p \end{array} \\
 \begin{array}{c} \mu \\ \text{~~~~~} \\ \swarrow \quad \searrow \end{array}
 \end{array}
 \quad = \quad
 \begin{array}{l}
 \frac{i(\not{p} + m)}{p^2 - m^2 + i\epsilon} \\
 \frac{-ig_{\mu\nu}}{p^2 + i\epsilon} \\
 iQe\gamma^\mu
 \end{array}$$

Figure 11.2: The Feynman rules as presented in Peskin and Schroeder (1995, 801–802).

The view that Feynman diagrams are simply a tool for organizing calculations and the many warnings against making incorrect interpretations also have a bearing on accounts of their origin. Several authors claim that the diagrams were developed to find abbreviations for complicated mathematical expressions. Silvan Schweber says that “[Feynman] diagrams evolved as a shorthand to help Feynman translate his integral-over-path perturbative expansions into the expressions for transition matrix elements being calculated” (Schweber 1994, 434). Brown also suggests that the diagrams are the result of Feynman’s attempts to simplify the task of finding and organizing terms in complicated perturbative calculations: “When Richard Feynman was working on quantum electrodynamics in the late 1940s, he created a set of diagrams to keep track of the monster calculations that were required” (Brown 1996, 265).

Most authors reconstruct the route that led Feynman to devise his new method of diagrams according to the premise that the physical content of the theory remained the same throughout the development. Also, when it comes to evaluat-

ing what was achieved, most authors maintain that no changes have occurred, as far as the physics is concerned, either in Feynman's or in related work by Julian Schwinger and Sin-Itiro Tomonaga. Dyson, who was one of the main actors in their development, would say in 1965:

Tomonaga, Schwinger, and Feynman rescued the theory without making any radical innovations. Their victory was a victory of conservatism. They kept the physical basis of the theory precisely as it had been laid down by Dirac, and only changed the mathematical superstructure. (Dyson 1965, 589)

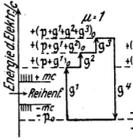
From a closer study of Feynman's unpublished manuscripts and early publications, this view of the development and interpretation of Feynman diagrams does not quite fit what emerges as Feynman's main concern. Feynman almost always used diagrams as a calculational tool and also as a means to represent the physical model by which he interpreted the theoretical equations.

If one is willing to accept that diagrams can have the two functions of calculating and representing at the same time, one could gain further insight into the development of quantum electrodynamics—not only in Feynman's work, but also beyond it. The differences in the means of representation in use at different stages of the development reflect differences in calculational techniques; in addition, they reflect profound differences in the way quantum electrodynamic phenomena were conceptualized.

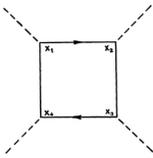
In fig. (11.3), we see three examples of processes which are, on the one hand represented by variations of atomic term schemes (left column) and, on the other hand, by a Feynman diagram (right column). The three processes are light-by-light scattering, the Compton effect and the scattering of an electron off an external electromagnetic potential. The use of Feynman diagrams is associated with different, often much more effective ways of calculating. With Feynman diagrams, one often does not need to distinguish between mathematical terms; without Feynman diagrams, those terms have to be treated separately. Without Feynman diagrams, the procedure was, therefore, more complicated and error-prone (Weinberg 1995, 37; Halzen and Martin 1984, 99).

However, the way of calculating was not the only difference that came with the use of Feynman diagrams. Without Feynman diagrams, the phenomena were represented as transitions between energy levels, almost like in traditional atomic physics, as we can see in the top row of fig. (11.3). With Feynman diagrams, see bottom row of fig. (11.3), the phenomena were analyzed into a succession of free propagation of initial quanta which are annihilated when intermediate quanta are created, and those then also get annihilated when the final state quanta are created.

Light-by-light scattering

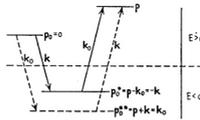


Euler (1936, 419)
(a)

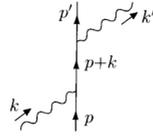


Karplus & Neuman
(1950, 381)
(b)

Compton scattering

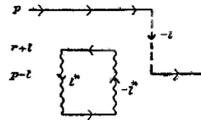


Heitler (1944, 190)
(c)

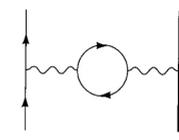


Peskin & Schroeder
(1995, 158)
(d)

Electron scattering off a potential



Koba & Takeda (1949, 69)
(e)



Peskin & Schroeder
(1995, 244)
(f)

Figure 11.3: Comparison of representations of elementary particle interactions with Feynman diagrams (lower row) and without Feynman diagrams (upper row). The two diagrams in the first column represent light-by-light scattering; the two diagrams in the second column represent Compton scattering; and the two diagrams in the third column represent the scattering of an electron off a potential. Diagram (a) is a detail from (Euler 1936, 419); (b) is from (Karplus and Neuman 1950, 381); (c) from (Heitler 1944, 190); (d) from (Peskin and Schroeder 1995, 158); (e) from (Koba and Takeda 1949, 69); and (f) from (Peskin and Schroeder 1995, 244). In the third edition of Heitler's *Quantum Theory of Radiation* from 1954, the diagram for Compton scattering reproduced here is absent.

11.2 Feynman's Struggle for a Physical Interpretation of the Dirac Equation

In my account of the origin of Feynman diagrams, Feynman's preoccupation with complicated mathematical expressions recedes, while his efforts to develop an appropriate way of representing quantum electrodynamic phenomena comes to the fore. Other accounts, such as Brown's and Schweber's, tend to neglect the latter in favor of the former. According to my reconstruction, Feynman diagrams are the final result of Feynman's quest for an informative physical interpretation of Dirac's well-known equation. Unpublished manuscripts indicate that Feynman began to seriously "struggle" with the Dirac equation in 1947.

In a letter to his student and friend Theodore Welton, Feynman announced a private research project:²

I am engaged now in a general program of study—I want to understand (not just in a mathematical way) the ideas in all branches of theor. physics. As you know I am now struggling with the Dirac Eqn.

Dirac's equation had been well known and long-used. What was Feynman looking for? It was not a new version of the equation nor a new method of solution. It was, rather, a deeper understanding of the equation:

The reason I am so slow is not that I do not know what the correct equations, in integral or differential form are (Dirac tells me) but rather that I would like to *understand* these equations from as many points of view as possible. So I do it in 1, 2, 3 & 4 dimensions with different assumptions etc.³

Clearly, for Feynman, the search for a deeper understanding should lead to an appropriate representation of the phenomena which are described by the equation. The physical interpretation which comes along with the representation should then provide the grounds for circumventing problematic consequences of the equation:

Of course, the hope is that a slight modification of one of the pictures will straighten out some of the present troubles.

²The following three quotations are taken from Richard Feynman's 1947 letter to Theodore Welton, in "Papers of Richard Phillips Feynman," California Institute of Technology (Caltech) Archives, Box 11, Folder 2. Partially reproduced in (Wüthrich 2010, 83–95). See also (Schweber 1994, 406–408).

³The italicized word is underlined in the original letter.

The “troubles” were that the Dirac equation yielded empirically well-confirmed predictions when it was applied in a rough approximation (perturbative expansion to the first order). However, when physicists attempted to obtain yet more accurate results by applying the equation in a more rigorous approximation (perturbative expansion to higher orders), the predictions became less precise and completely unusable and uninterpretable: they turned out to be infinite.

11.3 *Zitterbewegung*

Although part of the material on which my reconstruction of the development of Feynman diagrams is based has previously been analyzed by other scholars, I hope to be able to emphasize more clearly that Feynman's visualizations were not only a source of inspiration, and what he represented was far more than “toy models” for academic exercises.

In 1947, Feynman began to elaborate on the physical interpretation of Dirac's equation as describing a *Zitterbewegung*, a quivering motion, of the electron, which both Gregory Breit (1928) and Erwin Schrödinger (1930) had proposed.⁴ Breit considered an electron in an electromagnetic field while Schrödinger only considered the case of no forces. In either the one or the other form, the interpretation of the Dirac equation as describing a *Zitterbewegung* must have been well-known since it had been published in influential journals and was discussed in Dirac's book *The Principle of Quantum Mechanics* (Dirac 1935) and in his Nobel lecture (Dirac 1933).

In his investigations, Feynman restricted himself, most of the time, to the one-dimensional version of the equation. From the Hamiltonian which corresponded to the one-dimensional Dirac equation for an electron in an electromagnetic field

$$H\psi = \phi\psi + \alpha(p - A)\psi - \beta\mu\psi, \quad (11.1)$$

Feynman deduced that, according to Dirac's equation, the instantaneous velocity of the electron always equaled the speed of light. This was no novel result, however, since it was discussed in Breit's, Schrödinger's and other aforementioned publications.

In the above eq. (11.1), the electron's wave function ψ has two components; α and β are constant matrices: $\alpha = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and the momentum p is defined as the operator of partial differentiation with respect to the spatial variable

⁴See also (Fock 1929).

x , that is, $p = \frac{1}{i} \frac{\partial}{\partial x}$. ϕ and A are, respectively, the scalar and the vector potential of the electromagnetic field; μ is the mass of the electron.⁵

The result that the electron's speed was always identical to that of light meant, in Feynman's notation and units, that the velocity operator \dot{x} equaled the two-by-two Dirac matrix α . To derive this result, Feynman used the familiar relationship between the total time derivative (\dot{F}) of any operator (F) and its commutator with the Hamiltonian operator (H) and the partial temporal derivative ($\frac{\partial F}{\partial t}$) of the operator: $\dot{F} = i(HF - FH) + \frac{\partial F}{\partial t}$.

The result is puzzling at first since it seems to contradict the fact that no massive particle is known which moves at the speed of light. Also, the theory of special relativity precludes the existence of such a particle because, according to relativistic laws, the particle would contain an infinite amount of energy.

In Breit's and Schroedinger's publications, as well as in Dirac's book and Nobel lecture, the apparent contradiction to the empirical observations was resolved: we are dealing with a *Zitterbewegung*, a quivering motion, of the electron of which we can only observe the displacements *on average*. Thus, although the instantaneous velocity is the speed of light, the observable velocity is finite.

Feynman tried to incorporate this interpretation into his earlier work on an alternative formulation of non-relativistic quantum mechanics. In that work, Feynman had attempted to eliminate the divergences associated with self-interactions of the particles by representing the time development of the wave-function in terms of the *action*. The most important results of that work are published in an article in the *Reviews of Modern Physics* (Feynman 1948a). However, to a large extent, he had already obtained these results in 1942 in his doctoral thesis, which has been published only recently (Feynman 2005). In these two publications, Feynman showed explicitly that the time evolution of the wave-function was given by the sum of contributions of all paths that led a particle from its initial to its final position. The contribution of any one path was given by the action evaluated along that path.

In his unpublished notes from 1947, the first problem he tackled was that no justifiable expression for the action was known at the time. Therefore, Feynman considered a lattice of one space and one time dimension and interpreted the two components of Dirac's wave function as describing a particle that traveled either to the right or to the left. Feynman graphically represented a special case of the situation in a diagram. In this special case, the initial wave function of the particle has only a "right" component, and Feynman wants to determine the "left" component at the lattice point P which was N lattice spacings ϵ away in one diagonal direction, and M lattice spacings in the other diagonal direction.

⁵For details and a facsimile of the manuscript page (Caltech Archives, Box 11, Folder 2), see (Wüthrich 2010, 66–68).

Through a change of variables and an iterative solution procedure, Feynman deduced a factor $i\epsilon$ by which to count each reversal in the direction of the particle. The problem of finding the action thus came to working out how many paths there were that had a given number of changes in direction.⁶

He could then count each change by a factor $i\epsilon$ and sum over all paths. The obtained expression thus determined the time evolution of the wave-function exactly as the action would do.

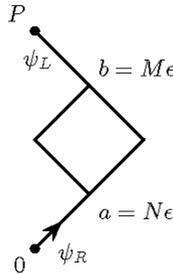


Figure 11.4: Abstract graphical representation of the quivering electron, which Feynman used for counting the number of possible zigzag paths. Reproduced by the author from a manuscript page by Feynman probably dating from 1947, see (Wüthrich 2010, 69).

Feynman thus solved the Dirac equation by “path counting,” as he wrote in his notes. Actually, by counting the zigzag paths he obtained the Green’s function associated with Dirac’s equation, which Feynman mentioned in passing on the same manuscript page from which fig. (11.4) is reproduced.

11.4 Positrons and Interaction

One of the striking features of the Dirac equation was its implication, or at least suggestion, of the existence of antiparticles. Feynman had not yet taken this into account, as far his “struggle” with the Dirac equation was concerned. Maybe he had an uneasy sense of the difficulties which would arise in the application of his method of path-counting to those positrons.

Taking an idea that John A. Wheeler, his PhD supervisor, had communicated to him in the autumn of 1940, Feynman conceived of the positron as an electron

⁶For details and a facsimile of the manuscript page (Caltech Archives, Box 11, Folder 2), see (Wüthrich 2010, 68–75).

moving backward in time.⁷ In my sketch of the modern application of Feynman diagrams, I briefly mentioned that such a conception is still in use with Feynman diagrams. For Feynman's model of the quivering electron and its possible paths, this meant that he now also had to account for loops of a path. The presence of electrons moving backward in time opened up the possibility of paths going through the same point twice, forming loops. The undesirable consequence of this was that now an infinite number of paths were possible between a given initial and a given final position, even though Feynman considered a space-time lattice and not a continuum. Therefore, the path-counting amounted to an infinite series which did not seem to converge, and the method seemed inapplicable.

However, in his notes, Feynman recognized that the possibility of a path that contained a loop implied the possibility of a path that contained the same loop but went through it in the opposite direction. Moreover, the contributions of the two paths canceled each other out, and Feynman concluded that "any completely closed loop cancel[ed]."⁸ Therefore, paths containing loops could be dismissed, and the method of path-counting was saved.

After having successfully dealt with positrons, Feynman moved to the next problematic issue. He attempted to incorporate the interaction between two or more particles into his model system of the Dirac equation. To this end, he tried to construct a Hamiltonian operator out of the action function which he had used, together with Wheeler, in his alternative formulation of classical electrodynamics (Wheeler and Feynman 1949).⁹

However, unlike with the incorporation of positrons, the difficulties were insurmountable. He was only able to treat the special case of an interaction which vanished after a certain time. In his notes, he expressed his dissatisfaction with this state of affairs. Of his attempts to describe a system of two particles by a joint wave function Φ he says:

It is a bit hard to see how to define Φ for path pair AB and CD , since there are some terms from interaction at x from y which is unspecified. However if the interaction is zero beyond P we are OK. Hence, at present, I can only specify Φ for paths which are long enough that they go beyond the time of interaction (this stinks).¹⁰

⁷For details of how Feynman learned about Wheeler's idea, see (Schweber 1986, 460; 1994, 387–388; Feynman 1966, 702, 705–706). See also Charles Weiner's Interviews with Dr. Richard Feynman, 4 March to 28 June 1966, vol. 2, p. 32, in Niels Bohr Library & Archives.

⁸For a facsimile of the page in question, (Caltech Archives, Box 12, Folder 9), see (Wüthrich 2010, 97; Schweber 1986, 482).

⁹For details, see (Wüthrich 2010, 104–108).

¹⁰Caltech Archives, Box 11, Folder 2. For a facsimile, see (Wüthrich 2010, 110).

11.5 Abandoning the Search for a Microscopic Analysis

Because he was not able to satisfactorily incorporate the interaction of particles into the model of the quivering electron, Feynman abandoned it. To him, it seemed impossible to analyze interacting electrons and positrons in terms of the microscopic *Zitterbewegung* implied by Dirac's equation. However, a closer inspection of his published papers (Feynman 1949a; 1949b) shows that he did not give up the attempt to construct a model system of the Dirac equation altogether. Only the level of specification of his explanatory model was about to change. Up to that moment, he had analyzed the propagation of an electron by a superposition of microscopic zigzag paths. This had led to a description of the propagation by a Green's function. Afterward, Feynman would work directly on this less-specific level of Green's functions and leave the propagation of the electron from one point to another without further analysis.

In this way, Feynman eventually succeeded in adequately describing the interaction between two electric particles. To obtain such a description, he took the classical mathematical expression for the potential as a basis. However, unlike in his previous unsuccessful attempt to incorporate interactions, he did not attempt to construct a quantum description by a Hamiltonian operator out of the classical expressions. In the previous attempt, he had needed the Hamiltonian to describe the quivering motion of the interacting particles. This time, he left aside the quivering motion and tried to fit the classical expressions into his more coarse-grained model of propagating particles described by Green's functions.

Feynman was probably more prepared than other physicists to interpret the electromagnetic interaction as being brought about by emission and absorption of light quanta because, in his "cut-off" paper (Feynman 1948b, 1431), he saw a way to avoid the difficulties, related to the polarization of the light quanta, which such an interpretation usually had to face. After some modifications of the classical potential, which put the expression for the potential side by side with the Green's functions that described the propagation of the electrons and positrons, the interpretation of the potential as describing the propagation of a light quantum must, therefore, have seemed natural to him.¹¹

This reinterpretation is clearly visible in the graphical representation Feynman used at that time, compared to the one he had used in the earlier work with Wheeler, see fig. (11.5), upon which his previous attempt to adequately describe interactions had been based.

Feynman thus achieved a description of the time evolution of a quantum electrodynamic system, including interaction, by Green's functions describing the free propagation of electrons, positrons and photons. This marks the endpoint

¹¹For details, see (Wüthrich 2010, 133–136).

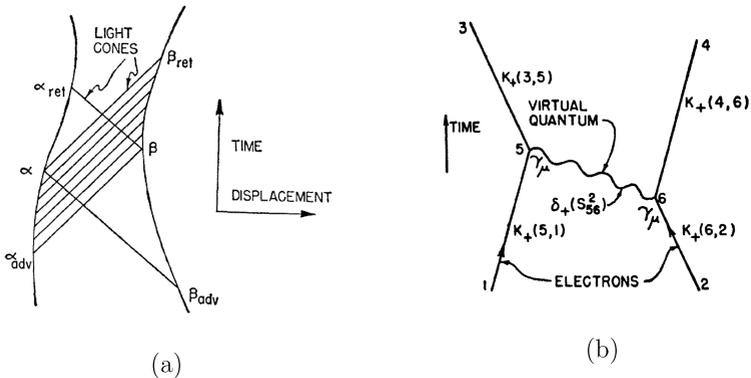


Figure 11.5: Graphical representation of the electromagnetic interaction between two particles (a) in Wheeler's and Feynman's alternative classical electrodynamics (Wheeler and Feynman 1949, 431) and (b) in Feynman's approach to quantum electrodynamics (Feynman 1949b, 772).

of Feynman's contributions in the period covered by this reconstruction of the genesis of Feynman diagrams. However, to understand how the modern form and use of Feynman diagrams came about, we cannot stop here. The modern form and use of Feynman diagrams goes back not only to the diagrams Feynman left us in his "space-time approach" (Feynman 1949b), rather, to a considerable extent, it goes back to what Dyson made of them in the context of the quantum field theory of the time. I address Dyson's contributions (Dyson 1949a; 1949b) in the next section.

11.6 Solution to the Problems through an Appropriate Representation of the Phenomena

With the innovative interpretation of the classical interaction as a propagation of a light quantum, Feynman fulfilled his aspiration, which he had announced in his letter to Welton (see footnote 2), to find a "picture," the "slight modification" of which would remove the theory's problematic infinities. He had now reduced all QED processes to the free propagation of initial quanta which are annihilated when intermediate quanta are created, these also get annihilated when the final state quanta are created. All QED processes were seen to be composed of a fundamental process: the emission or absorption of a light quantum by an electron or a positron described as a sequence of propagation of the particles and the quan-

tum. Feynman had thus found what he had been looking for ever since he began his "struggle" to fully understand the Dirac equation: a means of representation that made explicit which features of the theory could be held responsible for its divergence problems. If all phenomena the theory of QED described were essentially made up of one fundamental process, an appropriate modification of the representation of this process should suffice to eliminate the problems.

The community of physicists working at the time skeptically received Feynman's modified theory of QED, with which Feynman intended to solve the problems of the then-current theory, and they had a hard time understanding what it was all about. However, the main reason for the skeptical attitude of most physicists was not Feynman's extensive use of diagrams but rather the obsolete theoretical principles on which Feynman had based his theory. For instance, Feynman understood that, as in quantum mechanics, wave functions are probability amplitudes for the position of a particle. However, the theory of QED of the time was a theory in which the wave function was replaced with an operator-valued field, the quanta of which are electrons and positrons.

It was Dyson who rescued Feynman's diagrams from their obsolete theoretical setting. He began to interpret them in the context of 1940s state-of-the-art quantum field theory and eventually brought them to fruition. Dyson showed, for instance, that Feynman's Green's functions corresponded to quantum-field-theoretical vacuum expectation values (Dyson 1949a, 494). Thanks to the theoretical updating, Dyson could eliminate the problematic divergences in a more systematic manner than Feynman, and to arbitrary high orders in perturbation theory (Dyson 1949b).

Dyson showed that problematic divergences arose from two types of basic processes, which he represented graphically as shown in fig. (11.6(a)) and fig. (11.6(b)). To precisely determine observable quantities like cross sections and reaction rates, one should take all combinations of these processes into account, such as the one shown in fig. (11.6(c)).

However, Dyson was able to show how one could dispense with all of these problematic processes. He modified the *Green's functions* (or vacuum expectation values) which describe the free propagation of the field quanta and the operator which occurs in the description of the interaction, such that only "irreducible graphs" (Dyson 1949b, 1743), like the one which is shown in fig. (11.6(d)), have to be taken into account.

Dyson (1949b, 1754–1755) argued that the infinities arose from an over-idealized description of the interaction. The problematic diagrams represented effects that are entirely unobservable since they represented inevitable fluctuation processes. In the quantitative evaluation of the irreducible graphs, the infinite factors appeared only in combinations which could be interpreted as an *effective* mass

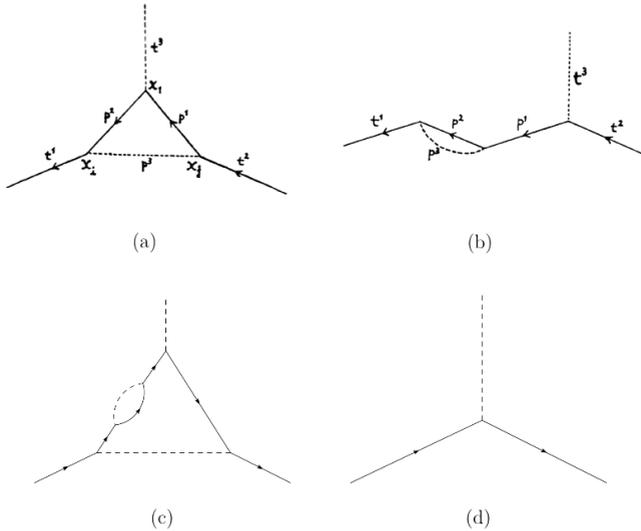


Figure 11.6: (a) and (b) show the two basic types of processes which are responsible for the problematic divergences according to Dyson's analysis (Dyson 1949b, 1741–1742); (c) shows a combination of the two; (d) the corresponding “irreducible graph” (drawings (c) and (d) by the author).

and charge. The different types of divergences could thus be calibrated against each other such that only the observed values for the mass and charge appeared in the observable quantities. This method came to be known as *renormalization*.

Dyson was surprised by the ease with which the recalcitrant problems concerning the divergences could be eliminated. For him, the cancellation of the infinities to yield the finite observable quantities was a physical fact:

The surprising feature of the [theory] as outlined in this paper, is its success in avoiding difficulties. Starting from the methods of Tomonaga, Schwinger and Feynman, and using no new ideas or techniques, one arrives at an S matrix from which the well-known divergences seem to have eliminated themselves. This automatic disappearance of the divergences is an empirical fact, which must be given due weight in considering the future prospects of electrodynamics. (Dyson 1949b, 1754)

The apparent simplicity of the solution can, at least partially, be explained by Dyson's use of an appropriate representation suggested by Feynman. Feynman's representation identified exactly the right elements to eliminate the divergences in a systematic and physically interpretable manner. Once Feynman had provided the appropriate representation, the successful modification of the theory, as performed by Dyson, no longer involved fundamental revisions but certainly deep insights as to the consequences of the new way of representing the phenomena. The development of an appropriate means of representation by Feynman was the fundamental, though not yet complete, revision Dyson successfully put to use.

11.7 Comparison to Other Developments of Concepts and Means of Representation

The early history of Feynman diagrams lends itself to a comparison with other developments of concepts and means of representation, in particular to James Clark Maxwell's abandonment of the mechanical model (Siegel 1991) and to medieval representations of change (Schemmel 2010).¹² I close my contribution with a short sketch of what seem to be the most interesting similarities and differences, with respect to the aforementioned cases, worthwhile to pursue further.

11.7.1 Maxwell's Abandonment of the Mechanical Model

According to Daniel Siegel, Maxwell interpreted his mechanical vortex model for electromagnetic phenomena more realistically than most of the scholarship on Maxwell would acknowledge. In the years following 1862, however, Maxwell partially abandoned the model and aimed to formulate his theory on the displacement current and light without a full commitment to the model which initially provided much of the necessary insights to the construction of the theory. Similarly, Feynman's move from the representation of the quivering electron to the representation of the electron's and positron's propagation by Green's functions is an abstraction of a theoretical description from some of the features the object under consideration was, up to then, assumed to have.

I would emphasize, however, that Feynman did not thereby abandon the physical interpretation of the diagrams; only the level of specification of the representation changed. To what extent this would also be true of Maxwell's case I am not able to assess, and it is not clear to me how much Siegel would endorse such a view. But it seems plausible that Maxwell's progress was also a process of changing the level of specification of the representation without giving up the physical interpretation of his equations.

¹²I thank Shaul Katzir and Jürgen Renn for having suggested these comparisons.

11.7.2 Medieval Representations of Change

According to Matthias Schemmel (2010), the medieval representations of change were an essential ingredient for the development of an appropriate concept of velocity when early modern scientists, like Thomas Harriott, employed them in their investigations. The form of the diagrams remained nearly unchanged while their interpretation was significantly revised.

Two phases of the development of Feynman diagrams show similar characteristics. Hans Euler, Walter Heitler, and Zirô Koba and Gyô Takeda used traditional means to represent phenomena, which they described using novel concepts. In traditional term schemes, the horizontal lines represented stable energy levels which manifested themselves as lines in spectroscopic analyses. In the hands of Euler and others, these horizontal lines indicated, instead, energy levels of electrons and positrons which were never observed. They represented intermediate states in a process in which only the initial and final states were observable.

The other phase in which the diagrams hardly changed but their interpretation did was the systematization of Feynman's theory by Dyson. With Feynman, the diagrams operated in the context of quantum mechanics, which describes particles by a wave-function. With Dyson's systematization, the context is quantum field theory, which describes particles as quanta of fields. With Feynman, the lines in the diagrams represented particle propagators, while with Dyson, the lines represented vacuum expectation values of field operators.

The most significant conceptual change, however, occurred with the invention of Feynman's diagrams during his search for an adequate interpretation of Dirac's equation. Compared to the traditional means of representing QED phenomena by adaptations of term schemes, Feynman's diagrams differ significantly, both in their form and in their interpretation. Instead of horizontal lines as main graphical elements, we have lines and vertices. Instead of transitions between energy levels, we have propagation of particles.

Rather than catalysts which stimulate a conceptual development but remain unchanged, as in the case Schemmel describes, the Feynman diagrams are a product of a process in which both the graphical means of representation and the conceptual framework were developed at the same time. It is hard to distinguish the process of conceptual development and the development of appropriate means of representation. The genesis of Feynman diagrams is a case in point to show that conceptual developments are carried out through a concrete manipulation of graphical means of representation.

Abbreviations and Archives

Caltech Archives	Archives of the California Institute of Technology, Pasadena, CA
Niels Bohr Library & Archives	American Institute of Physics, College Park, MD, USA

Acknowledgements

I am grateful to the Max Planck Institute for the History of Science, Berlin for the generous travel stipend for junior scholars for participation in the conference, and to Shaul Katzir and an anonymous referee for valuable comments. The research for my PhD thesis on the genesis and interpretation of Feynman diagrams has been funded by the Swiss National Science Foundation (grant no. 100011–113589), supervised by Gerd Graßhoff and co-refereed by Tilman Sauer.

References

- Breit, Gregory (1928). An Interpretation of Dirac's Theory of the Electron. *Proceedings of the National Academy of Sciences of the United States of America* 14: 553–559.
- Brown, James R. (1996). Illustration and Interference. In: *Picturing Knowledge: Historical and Philosophical Problems Concerning the Use of Art in Science*. Ed. Brian S. Baigrie. Toronto: University of Toronto Press. 250–268.
- Dirac, Paul A. M. (1928). The Quantum Theory of the Electron. *Proceedings of the Royal Society A* 117: 610–624.
- (1933). *Theory of Electrons and Positrons*. Nobel Lecture, December 12. http://www.nobelprize.org/nobel_prizes/physics/laureates/1933/dirac-lecture.pdf (accessed 11/29/2011).
- (1935). *The Principles of Quantum Mechanics*. 2nd ed. Oxford: Clarendon Press.
- Dyson, Freeman J. (1949a). The Radiation Theories of Tomonaga, Schwinger, and Feynman. *Physical Review* 75: 486–502.
- (1949b). The S-Matrix in Quantum Electrodynamics. *Physical Review* 75: 1736–1755.
- (1965). Tomonaga, Schwinger, and Feynman Awarded Nobel Prize for Physics. *Science* 150: 588–589.

- Euler, Hans (1936). Über die Streuung von Licht nach der Diracschen Theorie. *Annalen der Physik* 418: 398–448.
- Feynman, Richard P. (1948a). Space-Time Approach to Non-Relativistic Quantum Mechanics. *Reviews of Modern Physics* 20: 367–387.
- (1948b). Relativistic Cut-Off for Quantum Electrodynamics. *Physical Review* 74: 1430–1438.
- (1949a). The Theory of Positrons. *Physical Review* 76: 749–759.
- (1949b). Space-Time Approach to Quantum Electrodynamics. *Physical Review* 76: 769–789.
- (1966). The Development of the Space-Time View of Quantum Electrodynamics. *Science* 153: 699–708.
- (2005). The Principle of Least Action in Quantum Mechanics. In: *Feynman's Thesis: A New Approach to Quantum Theory*. Ed. Laurie M. Brown. Singapore: World Scientific Publishing. 1–69.
- Fock, Vladimir (1929). Über den Begriff der Geschwindigkeit in der Diracschen Theorie des Elektrons. *Zeitschrift für Physik* 55: 127–140.
- Halzen, Francis and Alan D. Martin (1984). *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. New York: John Wiley and Sons.
- Heitler, Walter (1944). *Quantum Theory of Radiation*. London: Oxford University Press.
- Karplus, Robert and Maurice Neuman (1950). Non-Linear Interactions between Electromagnetic Fields. *Physical Review* 80: 380–385.
- Koba, Zirô and Gyô Takeda (1949). Radiation Reaction in Collision Process III: First Radiative Correction for an Arbitrary Process Including Electrons, Positrons, and Light Quanta. *Progress of Theoretical Physics* 4: 60–70.
- Peskin, Michael E. and Daniel V. Schroeder (1995). *An Introduction to Quantum Field Theory*. Boulder (Colorado): Westview Press.
- Schemmel, Matthias (2010). *Medieval Representations of Change and Their Early Modern Application*. Preprint 402. Berlin: Max Planck Institute for the History of Science.
- Schrödinger, Erwin (1930). Über die kräftefreie Bewegung in der relativistischen Quantenmechanik. *Sonderausgabe aus den Sitzungsberichten der Preußischen Akademie der Wissenschaften Phys.-Math. Klasse* 24: 418–428.
- Schweber, Silvan S. (1986). Feynman and the Visualization of Space-Time Processes. *Reviews of Modern Physics* 58: 449–508.
- (1994). *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga*. Princeton (NJ): Princeton University Press.
- Siegel, Daniel M. (1991). *Innovation in Maxwell's Electromagnetic Theory: Molecular Vortices, Displacement Current, and Light*. Cambridge: Cambridge University Press.

- Weinberg, Steven (1995). *The Quantum Theory of Fields*. Cambridge: Cambridge University Press.
- Wheeler, John A. and Richard P. Feynman (1949). Classical Electrodynamics in Terms of Direct Interparticle Action. *Reviews of Modern Physics* 21: 425–433.
- Wüthrich, Adrian (2010). *The Genesis of Feynman Diagrams*. Archimedes Series. Ed. by Jed Buchwald. Dordrecht: Springer.

Traditions and Debates in Recent Quantum Physics

Chapter 12

Orthodoxies on the Interpretation of Quantum Theory: The Case of the Consistent History Approach

Olival Freire

Most of the historical narratives about the foundations of quantum theory center on the themes of orthodoxies and heterodoxies. Niels Bohr's and John von Neumann's early approaches were considered the orthodox views on the issue. In the 1950s, this research was marked by David Bohm's and Hugh Everett's heterodoxies, and according to such physicists who led the field in subsequent years as John Bell and John Clauser, its development faced the stigmas associated with this research. Since the blossoming of this research in the late twentieth century, warnings against the revival of old orthodoxies have been heard. A poignant alert was launched by Jeffrey Bub in *Interpreting the Quantum World*, published in 1997, when he dubbed the weaving of strands including decoherence, Everett's interpretation, and the consistent history approach the "new orthodoxy." Bub pointed to Roland Omnès writings as examples of this new orthodoxy.

Here, I analyze these claims, particularly the consistent history approach. I consider not only the rhetorical strategies adopted by its proponents and critics, such as Bub himself, but also the effective influence achieved by this approach. Bub's claim that the consistent history approach is a new orthodoxy is an overstatement. This paper presents a summary of the use of terms such as "orthodoxy" and "heterodoxy" in reference to quantum mechanics. In addition, it deals with the polysemic manner in which the concept of orthodoxy appears in Bub's book; and I present a synopsis of the consistent history approach, of its claims and rhetorical strategies. The final part is dedicated to the analysis of the effective influence of this approach on physicists. Further, I draw some conclusions from this history about the uses of the terms orthodoxy and heterodoxy in debates on the foundations of quantum mechanics.

12.1 Orthodoxies and Heterodoxies in Quantum Physics

Between 1925 and 1927, a polyphony of interpretations of the newly-born quantum theory emerged. This concurrence was narrowed in October 1927 when Bohr

presented his complementarity principle at the Solvay Conference. Bohr's interpretation was not accepted by such physicists as Albert Einstein and Erwin Schrödinger. However, it was supported by a number of others, including Werner Heisenberg, Wolfgang Pauli, and Max Born. While the term orthodoxy was not commonly used at the time, its meaning hung in the air. Louis de Broglie, who arrived at the conference suggesting a causal interpretation of quantum mechanics which was at variance with the notion of complementarity, left disillusioned with his own proposal. When faced with the duties of teaching quantum mechanics in Paris, he "joined the ranks of the adherents to the orthodox interpretation which was accepted by the overwhelming majority of the participants at the Solvay meeting" (Jammer 1974, 114). In 1928, Einstein wrote to Schrödinger, both men in a clear-cut minority among the founding fathers of this physical theory, on complementarity: "The Heisenberg-Bohr tranquilizing philosophy—or religion?—is so delicately contrived that, for the time being, it provides a gentle pillow for the true believer from which he cannot very easily be aroused. So let him lie there."¹

In the early 1930s, the mathematician von Neumann presented a fully consistent treatment of quantum theory in terms of Hilbert spaces. Together with complementarity, von Neumann's treatment conveyed the feeling that both the philosophical implications and the mathematical formalism of the theory were settled forever. Moreover, in the 1930s, physicists failed to exploit the differences between Bohr's and von Neumann's views regarding completeness and measurement problems.

In the 1950s, the manner in which physicists referred to the dominant view of the interpretation of quantum mechanics began to change. Critics of complementarity referred to it as the "usual" interpretation, as Bohm (1952), or "Copenhagen interpretation," as Everett (Osnaghi, Freitas, and Freire 2009, 105, footnote 111). Later, the historian of physics Max Jammer (1974, 250) dubbed the orthodoxy "the monocacy of the Copenhagen school." The term "Copenhagen interpretation," apparently created by Heisenberg, was not consensually accepted by adepts of Bohr's complementarity. Most importantly, it was used by critics of Bohr's views in general (Osnaghi, Freitas, and Freire 2009, 99). In the early 1960s, Eugene Wigner conspicuously called von Neumann's mathematical presentation of the measurement problem "the orthodox view" in quantum mechanics, only to say that either quantum mechanics was incomplete and could be complemented by a nonlinear modification or one should accept the mind's role during measurement processes (Wigner 1963). If Bohr were alive, it is unlikely that he would accept either of Wigner's choices. As I have argued elsewhere, Wigner indeed became a heterodox in the foundations of quantum mechanics and supported most

¹Einstein to Schrödinger, 31 May 1928 (Jammer 1974, 130).

of the research in this field during the late 1960s (Freire 2007). From 1970 on, the term “Princeton school” was used to distinguish Wigner and von Neumann’s views from Bohr’s as well as to signal that the monolithic support behind what was once considered the orthodox view had waned or had been split (Freire 2007).

In the 1960s, a new meaning for orthodoxy was emerging among the new generations of physicists interested in the foundations of quantum mechanics. Bell, who would play a key role in subsequent years in this research, co-authored a paper with Michael Nauenberg in 1966 saying:

[W]e emphasize not only that our view [that quantum mechanics is, at best, incomplete] is that of a minority but also that current interest in such questions is small. *The typical physicist feels that they* [issues on foundations of quantum mechanics] *have long been answered*, and that he will fully understand just how if ever he can spare twenty minutes to think about it. (Freire 2006, 583, emphasis added by OFJ)

The same sentiment was conveyed by Abner Shimony, in a later recollection:

[T]he preponderance of the physics community at that time accepted some variant of the Copenhagen interpretation of quantum mechanics and believed that satisfactory solutions had already been given to the measurement problem, the problem of Einstein-Podolsky-Rosen, and other conceptual difficulties. (Shimony 1993, XII)

Thus, when research on the foundations of quantum mechanics began to appeal to a larger number of physicists in around 1970 (Freire 2004; 2009), orthodoxy was a polysemic term meaning Bohr’s complementarity, von Neumann’s mathematical presentation, and the vague but influential idea that problems in the foundations of quantum mechanics had already been solved by the founding fathers of the discipline.

A conclusion may be drawn from this short review. Orthodoxy is a term that was never used by the supporters of the complementarity view to refer to themselves. Often it is currently used without implicit assumptions, but mostly orthodoxy is used by critics of the complementarity view or Bohr’s legacy. Such assessments suggest that Bohr and adepts of the complementarity view were closed-minded to the diversity of possible interpretations of quantum mechanics, and their authority helped suffocate debate on the subject. Heinz-Dieter Zeh sharply criticized the inappropriateness of authority’s role: “I have always felt bitter about the way how Bohr’s authority together with Pauli’s sarcasm killed any discussion about the fundamental problems of the quantum.”² The term orthodoxy has been

²Zeh to Wheeler, 30 October 1980, Wheeler Papers, Series II, Box Wo-Ze, folder Zeh. Cited in (Freire 2009, 282).

used in the controversy over the foundations of this physical theory as a rhetorical strategy, either by the critics of Bohr's views or von Neumann's mathematical formulation of this theory. It is a strategy used to open or keep open the debate about alternative interpretations or approaches to issues important to the foundations of quantum mechanics. Rhetorical strategies were also used by defenders of complementarity, a process the philosopher Mara Beller called "The Copenhagen Dogma: The Rhetoric of Finality and Inevitability" (Beller 1999). Léon Rosenfeld, for instance, criticized Heisenberg's use of the term Copenhagen interpretation because it conveys the idea that complementarity is just one among other possible interpretations (Osnaghi, Freitas, and Freire 2009, 99).

12.2 The Polysemic New Orthodoxy

Almost fifteen years ago, physicist and philosopher Bub warned about the appearance of a new orthodoxy as regards the interpretation of quantum mechanics. The term has since gained some currency among physicists and philosophers, as well as among historians of physics.³ Bub opened his chapter dedicated to the new orthodoxy recalling a long-lasting attitude among physicists. According to Bub:

For most physicists, the measurement problem of quantum mechanics would hardly rate as even a 'small cloud' on the horizon. The standard view is that Bohr had it more or less right, and that anyone willing to waste a little time on the subject could easily straighten out the sort of muddle philosophers might get themselves into. (Bub 1997, 212)

I have argued elsewhere that such an attitude has been blamed for hampering our understanding of foundational issues of quantum mechanics, as far as the hidden-variable issue and its related Bell's theorem are concerned. In fact, it was this kind of orthodoxy that Bell and Clauser referred to as the stigma against research on hidden variables (Freire 2006; 2009). However, those obstacles were eventually overcome, and the field is today generally considered a regular field of research, even reaping some of the fruits of the quantum information boom. Therefore, Bub's warning seems to alert us to past obstacles created by a prevailing orthodoxy encountered by physicists and philosophers who dealt with the foundations of quantum mechanics. Yet Bub was not only speaking of an already existing orthodoxy. According to him:

³See (Schlosshauer 2004; Hagar 2007; Ghirardi 2008; Camilleri 2009).

There seems to be a growing consensus that a modern, definitive version of the Copenhagen interpretation has emerged, in terms of which the Bohr-Einstein debate can be seen as a rather old-fashioned way of dealing with issues that are now much more clearly understood. (Bub 1997, 212)

It is reasonable to question if this new consensus is producing a new orthodoxy, that is, creating new intellectual and professional obstacles, as in the past, thus hampering the development of quantum physics. To assess the reach of Bub's statement, we have to first examine what he meant by old and new orthodoxies. However, Bub used the term orthodoxy in a polysemic manner, which is not unusual in the literature on quantum physics, as already explored in the previous section. Indeed, Bub introduced the concept of orthodoxy in four different ways:

1. The first meaning is that transcribed at the beginning of this paper, which hinges on the founders of the discipline. In sum, the shared view that foundational issues were already solved by the founding fathers of the discipline and do not deserve attention from younger practitioners (Freire 2006, 583; Shimony 1993, XII).
2. Bub (1997, 221) also used orthodoxy with a second meaning, which he called "the orthodox (Dirac-von Neumann) interpretation principle (the 'eigenvalue-eigenstate link')."
3. Citing Zurek, Bub also included "Bohr's 'Copenhagen Interpretation'" as a third meaning for orthodoxy (Bub 1997, 223).
4. Lastly, in the boldest statement, Bub presented the new orthodoxy as a mix of several strands, such as the physical phenomenon of environment-induced decoherence, elements of Everett's relative state formulation and the notion of "consistent histories." And he singled out the French physicist Omnès as the spokesman of the new orthodoxy.

Bub recalled that:

Omnès refers to '*the* interpretation of quantum mechanics, not *an* interpretation,' and characterizes the view as 'simply a modernized version of the interpretation first proposed by Bohr in the early days of quantum mechanics.'⁴ (Bub 1997, 212, emphasis added by OFJ)

While the last three of Bub's concepts of orthodoxy, the orthodox interpretation principle, the Copenhagen interpretation, and Omnès's new orthodoxy, can be found in texts by various authors, the orthodoxy, the founder's orthodoxy, cannot be attributed to anyone in particular, as it is simply an unwritten belief held by

⁴Bub cites Omnès (1994, XIII and 498).

a professional group, physicists in this case. This founders' orthodoxy has no evident or single authorship, but it is a very effective, professionally-grounded attitude. It is not completely independent of Bub's second and third meanings, because the intellectual authority of some of the founding fathers contributed to its wide acceptance. However, hypothetically at least, the second and third meanings of orthodoxy could have existed independent of the founders' orthodoxy, and thus Bub's second and third meanings would not have hindered research on the foundations of quantum physics.

The belief that foundational issues had already been solved survived the founding fathers of the discipline; however, it was challenged by new generations, and eventually research on the foundations of quantum mechanics blossomed. The days of the supremacy of the authority of the founding fathers of the discipline are gone. Therefore, the issue of historical and practical interest seems to be: is the new orthodoxy, Bub's fourth meaning, resuscitating the founders' orthodoxy, his first meaning? If this is the case, such a symbiosis may be harmful to the development of research on the foundations of quantum mechanics. The issue deserves close scrutiny. However, instead of investigating Bub's new orthodoxy as a whole, which is of uncertain authorship, I choose to focus my analysis on the approach represented by Omnès, whom Bub singled out in his fourth meaning, the new orthodoxy. The analysis that follows is thus focused on the consistent history approach and whether it represents the new orthodoxy in quantum theory.

12.3 The Consistent History Approach and Its Rhetorical Resources

The consistent history approach developed between 1984 and 1990, and its founding fathers include Robert Griffiths, Omnès, Murray Gell-Mann and James Hartle. Griffiths is a prominent statistical physicist working at the Carnegie-Mellon University in Pittsburgh, who in the early 1980s turned his attention to research on the interpretation of quantum mechanics. In his seminal paper, "Consistent Histories and the Interpretation of Quantum Mechanics," published in 1984, Griffiths suggested mathematical criteria to use classical rules of probability to produce conditional probabilities for sequences of events at different times. He showed that such criteria could be applied to systems described by the usual quantum mechanical formalism (Griffiths 1984). He called these criteria a consistent history approach because they were able to identify sequences of events, now called consistent histories, which were meaningful in a quantum mechanical treatment. These criteria constitute a regulatory principle to adopt in quantum theory. For Griffiths (1984, 219), the main advantage of his approach was that it could be applied to closed (isolated) quantum systems between successive measurements, thus without taking measurement as a central process for quantum theory. There-

fore, one could speak about the physical meaning of a quantum state even in the absence of measurement processes, which is an advantage for a philosophical approach to quantum physics in terms of realism. By the same token, the new approach solved the conceptual difficulties associated with measurement in other interpretations of standard quantum mechanics. Among them, Griffiths pointed to two interpretations. The first requires conscious observers, a reference to von Neumann, Fritz London and Edmond Bauer, and Wigner. The second approach includes classical apparatuses, an indirect reference to Bohr.

While his approach differed from these traditional interpretations, Griffiths did not see it as an alternative interpretation. Rather, he saw it “as an extension and [we hope] clarification of what is, by now, a ‘standard’ approach to quantum probabilities” (Griffiths 1984, 221). However, Griffiths did not present his paper as a reinforcement of any orthodoxy. He saw it as part of “an extended controversy which is far from being resolved” about the “physical interpretation to the solutions (including boundary and initial conditions)” of the Schrödinger equation (Griffiths 1984, 221). From the immense literature on quantum interpretation, he singled out papers by Kurt Gottfried, Marcelo Cini, Peter Moldauer and Everett for comment and criticism, in addition to the “orthodox views” by von Neumann and Wigner. It is remarkable that he did not reveal any special influence from Everett’s interpretation (Griffiths 1984, 257–265).

Omnès is a theoretical physicist from the Université de Paris XI in Orsay. Before changing his focus to the foundations of quantum mechanics, he worked on particle and field physics. In his answer to a referee of the first major publication of his proposal, Omnès highlighted his own contribution to the consistent history approach. Asked about “what is common and what is different in [his] approach with Griffith’s [sic] history description,” he replied that “as far as mathematical techniques are concerned, Griffith’s [sic] construction is used,” and added “the conceptual foundations are different because what is proposed here is a revision of the logical foundation of quantum mechanics” (Omnès 1987, 172). Omnès revealed in this answer his intellectual heritage, that of the modern axiomatization which comes from the mathematician David Hilbert. Omnès acknowledges this influence through his debts to Henri Cartan’s teachings (Omnès 1988a, 931). In three-paper follow up, he developed the logical and theoretical machinery that allowed him “to construct consistent Boolean logics describing the history of a system, following essentially Griffiths’s proposal” (Omnès 1988a, 893).

While Omnès recognized discussions with other physicists interested in the foundations of quantum physics, such as Bell, Jean-Pierre Vigièr, and mainly Bernard d’Espagnat, he did not relate his work to the ongoing controversy over quantum physics, except for Griffiths’s contributions. In particular, he did not cite Everett’s interpretation, a distance he would keep. Furthermore, while he admit-

ted Bohr's shortcomings—"when singling out strict classical physics for expressing experimental data, Bohr was creating new, deep problems" (Omnès 1999, 52)—he tended to present the entire consistent history approach as "significant progress [...] towards a consistent and complete reformulation of the Copenhagen interpretation" (Omnès 1992, 339).

In contrast, Gell-Mann and Hartle came from very different backgrounds; it was the quantization of gravitation which led them to foundations of quantum physics, as they acknowledged in their first joint paper: "we will discuss the implications of quantum cosmology for one of the subjects of this conference—the interpretation of quantum mechanics" (Gell-Mann and Hartle 1989, 322). Previously, in 1983, Hartle, in collaboration with Stephen Hawking, had worked out what is now known as the Hartle-Hawking wave function of the universe, a solution of the Wheeler-DeWitt equation for quantizing gravitation. In the late 1980s, Hartle from the University of California, Santa Barbara, and his former PhD supervisor, the particle physics 1969 Nobel Prize winner Gell-Mann from the California Institute of Technology, approached the issue of interpreting quantum mechanics. The main merit of their contribution was to associate the attribution of classical probabilities in quantum systems as preached by Griffiths and Omnès with decoherence, a quantum feature whose understanding was just emerging. The connection was that "decoherence requires a sufficiently coarse-grained description of alternative histories of the universe" (Gell-Mann and Hartle 1989, 321; 1990, 425). According to Gell-Mann, "coarse graining typically means following only certain things at certain times and only to a certain level of detail" (Gell-Mann 1994, 144). While the first papers they jointly published were more programmatic, they eventually published a more technical work in which "the connections among decoherence, noise, dissipation, and the amount of coarse graining necessary to achieve classical predictability are investigated quantitatively" (Gell-Mann and Hartle 1993, 3345).

As for affiliations, Gell-Mann and Hartle departed from the point of view that all standard interpretations, Copenhagen included, which presuppose a classical domain or an external observer, are inadequate for cosmology because "measurements and observers cannot be fundamental notions in a theory that seeks to discuss the early universe when neither existed." They acknowledged Everett as the first to suggest "how to generalize the Copenhagen framework so as to apply quantum mechanics to cosmology." However, they considered Everett's work incomplete as Everett was not able to "adequately explain the origin of the classical domain or the meaning of the 'branching' that replaced the notion of measurement" (Gell-Mann and Hartle 1990, 429–430). Thus, Gell-Mann and Hartle considered the works of Wojciech Zurek, Erich Joos and Zeh with regard to decoherence as a "post-Everett" stage, and included this trend into their own proposal,

along with Griffiths's and Omnès's (Gell-Mann and Hartle 1990). Later, in a book for a wider audience, Gell-Mann made a distinction between the interpretation of quantum mechanics by the founding fathers and the modern one. The former he considered marked by a "curiously restrictive and anthropocentric fashion," as it was based on the existence of observers and classical domains, while the latter was presented as an approach still under construction (Gell-Mann 1994, 136 and ch. 11–12).

As for the rhetoric of orthodoxy, these authors cannot be put in the same category. Certainly Griffiths did not frame his proposal in terms of a new orthodoxy. Instead, he explicitly considered it part of the ongoing quantum controversy, "which is far from being resolved" (Griffiths 1984, 220). Unlike Griffiths, Gell-Mann was seduced by the idea of a new orthodoxy. He presented his own approach as the "modern" interpretation, contrasting it with that of the founding fathers. Adopting the rhetorical strategy of presenting two interpretations, one old-fashioned and the other modern, Gell-Mann (1994, 136–173) implicitly conveyed the idea of a new orthodoxy. He christened the former "the approximate quantum mechanics of measured systems" and introduced it saying that "when first formulated by its discoverers, quantum mechanics was often presented in a curiously restrictive and anthropocentric fashion." He presented the latter saying "for describing the universe, a more general interpretation of quantum mechanics is clearly necessary, since no external experimenter or apparatus exists and there is no opportunity for repetition, for observing many copies of the universe," and added "that is one reason why what I call the modern interpretation of quantum mechanics has been developed over the last few decades" (Gell-Mann 1994, 136–173). By presenting his own approach as the "modern" interpretation emerging as part of the "post-Everett" stage, Gell-Mann is, indeed, excluding other possible interpretations and thus playing the game of the "new orthodoxy." The irony of history, a new orthodoxy barely fits with the idea of a "post-Everett" stage, as Everett's own ideas in their time were considered supreme heresy (Osnaghi, Freitas, and Freire 2009).

Omnès's rhetoric is rather close to the idea of a new orthodoxy. It may indeed have raised concerns about the claim to be a new definitive solution to the problems in the foundations of quantum theory. Omnès presented the new approach as bringing together three different achievements ("the decoherence effect," "the emergence of classical physics from quantum theory," and the "constitution of a universal language of interpretation by means of consistent histories") and went on to conclude that the consistent history approach is a method which "provides a logical structure for quantum mechanics and classical physics as well" (Omnès 1999, 69). He further states that:

[W]hen these three ideas are put together, they provide a genuine theory of interpretation in which *everything is derived directly from the basic principles alone and the rules of measurement theory become so many theorems*. (Omnès 1999, 70, emphasis added by OFJ)

Thus, according to Omnès, the consistent history approach settles the main issues in the foundations of quantum theory. It is not by chance that Bub singled out Omnès's discourse as the target for his criticism.

12.4 The Reception of the Consistent History Approach

Rhetorical strategies, however, are not enough to explain the existence of a new orthodoxy. After all, heterodox interpretations, such as Bohm's hidden variables and Everett's relative states, were also presented with rhetorical strategies that promised to solve all the problems in the foundations of quantum mechanics (Freire 2005; Osnaghi, Freitas, and Freire 2009). As my colleague Joan Bromberg once remarked, the existence of an orthodoxy requires the existence of followers. Thus, answers to such questions as who are the followers of the consistent history approach and how influential are they, seem to be the litmus test for the existence of a new orthodoxy. In light of this, we now deal with more questions. How was the consistent history approach received by most physicists? Is there, indeed, a growing adhesion to the consistent history approach as the solution to the problems in the foundations of quantum theory?

To gain an insight into the reception of the consistent history approach, I bring into play scientometrics. We know how misleading this source may be if considered independently from other analytical resources (Freitas and Freire 2003). Given this, I have not only considered raw figures concerning citations but also some qualitative cues. I use as my example Griffiths's 1984 paper, not only because it was the first in this approach and introduced the term "consistent history approach," but also because it is the most cited among papers by Omnès, Gell-Mann, Hartle and Griffiths on this topic. According to the Web of Science, see fig. (12.1), it amassed 450 citations, which is very good for citations of a paper in physics. This first positive impression is slightly marred by the data in the following chart, which registers the number of citations per year from 1985 to 2009. Citations of Griffiths's paper took off after 1990, probably due to the connection made by Gell-Mann and Hartle between consistent histories and decoherence. After a decade of rising numbers of citations, however, citations began to decline then remained steady before declining again. These fluctuations can hardly be said to be evidence of a new orthodoxy.

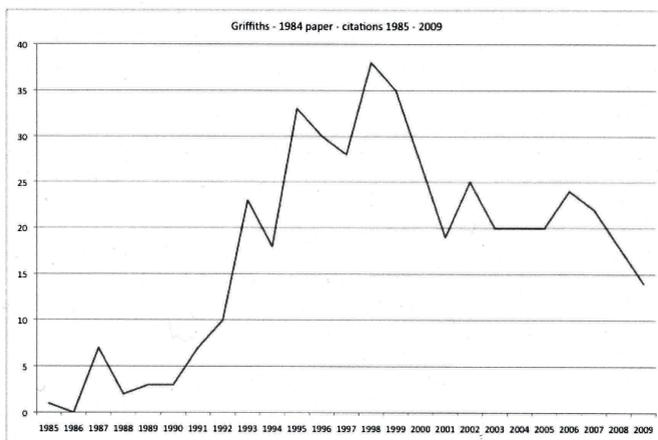


Figure 12.1: Citations of (Griffiths 1984). Data source: Web of Science, query, 4 June 2010, image by the author.

In addition, I checked publications citing Griffiths's paper. I was interested in particular in discovering any connections between this approach and the ever-growing experimental activities on the foundations of quantum mechanics. I sorted all the physicists who cite Griffiths's paper more than ten times. Their names and the number of times they cited Griffiths's paper are listed in the table below:

J. J. Haliwell – 29	R. Omnès – 21	B. L. Hu – 15
C. Anastopoulos – 22	R. B. Griffiths – 18	J. B. Hartle – 13

In addition to the expected cross-citations among the team who suggested such an approach, the other three physicists are also theoretical physicists, which suggests a weak connection between the consistent history approach and the blossoming experimental physics in this field. I also assessed the number of times Griffiths's paper was referred to by physicists working on experiments concerned with the foundations of quantum mechanics. I did not find any citations of Griffiths's paper by prominent experimentalist researchers in the foundations of quantum physics, such as Serge Haroche, Anton Zeilinger and Herbert Walther. It appears that the consistent history approach cannot be considered a relevant theoretical framework for the flourishing experimental research in this domain.

Leaving aside quantitative data, let me make a few qualitative comments. That the proponents of the consistent history approach have suggested a new approach to the foundations and interpretations of quantum physics is, in itself, evidence that they did not consider “the measurement problem of quantum [...] a ‘small cloud’ on the horizon,” or that “Bohr had it more or less right,” as indicated in the founders’ orthodoxy described by Bub (1997, 212). In addition, it is doubtful whether Griffiths, Omnès, or Gell-Mann and Hartle possess the kind of professional and intellectual authority that Bohr, Pauli and Heisenberg had in the golden days after the creation of quantum mechanics. In the introduction to his new book in 2002, Griffiths mentioned his “fellow consistent historians” naming Gell-Mann, Hartle and Omnès. He also identified some of the critics such as d’Espagnat, Giancarlo Ghirardi, Basil Hiley, Adrian Kent, Euan Squires, Angelo Bassi and Fay Dowker. This suggests that after more than a decade, the number of adherents remained the same.

12.5 Conclusion

Bub’s fear of a new orthodoxy was not unfounded, as this field was handicapped in the 1950s and 1960s by the widespread idea that foundational issues had already been solved by Bohr and other founding fathers of quantum physics. Bub himself, as a graduate student of Bohm in the 1960s, probably experienced these adversities. However, in spite of having good reasons for fearing a new orthodoxy, my conclusion is that Bub, as far as the consistent history approach is concerned, overstated the existence of a new orthodoxy, at least in the first and fourth meanings described in this paper, namely, the founder’s orthodoxy and the combined meaning represented by Omnès’s approach. The consistent history approach seems to be simply one more candidate, albeit a strong one, in the plethora of possible interpretations for quantum theory. Thus, while following the steps of previous researchers in this field who criticized the rhetoric of orthodoxy in quantum mechanics, Bub appears to have used the same resource in a new and different context. The efficacy of using the same resources for different contexts is therefore doubtful.

It is possible that Bub missed the target while singling out the consistent history approach for the new orthodoxy and Omnès as its representative. Research on the foundations of quantum mechanics may face other kinds of obstacles at the time of this writing. Perhaps Bub was worried about the widespread feeling that decoherence is the solution to the quantum measurement problem. However, the relationship between decoherence and foundational issues is better addressed in terms of an ongoing controversy than in terms of orthodoxies, as Maximilian Schlosshauer’s (2004) review on the subject may evidence. Others, such as Ghi-

rardi (2006, 2913), are concerned with the uncertain idea that quantum physics is ultimately a physical theory of information, an idea that he called the “quantum information interpretation.” Amit Hagar and Meir Hemmo (2006, 1295), along the same lines, state that “quantum information theory has by now become to a large extent a new orthodoxy in the foundations of quantum mechanics,” but, as evidence that orthodoxy is a polysemantic word in quantum mechanics, Bub’s approach, in terms of information, is itself the target of Hagar and Memmo’s criticisms. However relevant these views may be to the unfolding research on quantum information, analyzing them is another and quite different story from that of the consistent history approach. At any rate, as the term orthodoxy has become so polysemous, with an increasing number of different orthodoxies, it has lost its rhetorical efficacy: it is pointless to speak either of many orthodoxies or of many heterodoxies.

Abbreviations and Archives

Wheeler Papers	American Philosophical Society, Philadelphia
----------------	--

Acknowledgements

I am thankful to the National Council of Technological and Scientific Development (CNPq) and the Bahia State Research Support Foundation (FAPESB) of Brazil for their support and to Christoph Lehner, Shaul Katzir, Kent Staley, Michael Stöltzner and John Stachel for their comments.

References

- Beller, Mara (1999). *Quantum Dialogue: The Making of a Revolution*. Chicago: The University of Chicago Press.
- Bohm, David (1952). A Suggested Interpretation of the Quantum Theory in Terms of “Hidden” Variables—I and II. *Physical Review* 85(2): 166–179, 180–193.
- Bub, Jeffrey (1997). *Interpreting the Quantum World*. Cambridge: Cambridge University Press.
- Camilleri, Kristian (2009). A History of Entanglement: Decoherence and the Interpretation Problem. *Studies in History and Philosophy of Modern Physics* 40: 290–302.

- Freire, Olival (2004). The Historical Roots of “Foundations of Quantum Physics” as a Field of Research (1950–1970). *Foundations of Physics* 34(11): 1741–1760.
- (2005). Science and Exile—David Bohm, the Cold War, and a New Interpretation of Quantum Mechanics. *Historical Studies in the Physical and Biological Sciences* 36(1): 1–34.
 - (2006). Philosophy Enters the Optics Laboratory: Bell’s Theorem and Its First Experimental Tests (1965–1982). *Studies in History and Philosophy of Modern Physics* 37: 577–616.
 - (2007). Orthodoxy and Heterodoxy in the Research on the Foundations of Quantum Physics: E.P. Wigner’s Case. In: *Cognitive Justice in a Global World: Prudent Knowledges for a Decent Life*. Ed. Boaventura de Sousa Santos. Lanham, MD: Lexington Books. 203–224.
 - (2009). Quantum Dissidents: Research on the Foundations of Quantum Mechanics circa 1970. *Studies in History and Philosophy of Modern Physics* 40: 280–289.
- Freitas, Fabio and Olival Freire (2003). Sobre o uso da Web of Science como fonte para a história da ciência. *Revista da sociedade brasileira de história da ciência* 1(2): 129–147.
- Gell-Mann, Murray (1994). *The Quark and the Jaguar—Adventures in the Simple and the Complex*. New York: W. H. Freeman.
- Gell-Mann, Murray and James Hartle (1989). Quantum Mechanics in the Light of Quantum Cosmology. In: *Proceedings of the 3rd International Symposium on the Foundations of Quantum Mechanics*. Tokyo: Physical Society of Japan. 321–343.
- (1990). Quantum Mechanics in the Light of Quantum Cosmology. In: *Complexity, Entropy, and the Physics of Information*. Ed. Wojciech H. Zurek. Redwood City [CA]: Addison-Wesley. 425–458.
 - (1993). Classical Equations for Quantum Systems. *Physical Review D* 47(8): 3345–3382.
- Ghirardi, Giancarlo (2006). Some Reflections Inspired by My Research Activity in Quantum Mechanics. *Journal of Physics A: Mathematical and Theoretical* 40: 2891–2917.
- (2008). Collapse Theories. In: *The Stanford Encyclopedia of Philosophy*. Ed. Edward N. Zalta. Fall 2008. Stanford: Stanford University. <http://plato.stanford.edu/archives/fall2008/entries/qm-collapse/> (accessed 06/04/2010).
- Griffiths, Robert (1984). Consistent Histories and the Interpretation of Quantum Mechanics. *Journal of Statistical Physics* 3(1–2): 219–272.
- (2002). *Consistent Quantum Theory*. Cambridge: Cambridge University Press.

- Hagar, Amit (2007). Experimental Metaphysics₂: The Double Standard in the Quantum-Information Approach to the Foundations of Quantum Theory. *Studies in History and Philosophy of Modern Physics* 38: 906–919.
- Hagar, Amit and Meir Hemmo (2006). Explaining the Unobserved—Why Quantum Mechanics Ain't Only about Information. *Foundations of Physics* 36(9): 1295–1324.
- Jammer, Max (1974). *The Philosophy of Quantum Mechanics—The Interpretations of Quantum Mechanics in Historical Perspective*. New York: Wiley.
- Omnès, Roland (1987). Interpretation of Quantum Mechanics. *Physics Letters A* 125(4): 169–172.
- (1988a). Logical Reformulation of Quantum Mechanics. I. Foundations. *Journal of Statistical Physics* 53(3–4): 893–932.
- (1988b). Logical Reformulation of Quantum Mechanics. II. Interferences and the Einstein-Podolsky-Rosen Experiment. *Journal of Statistical Physics* 53(3–4): 933–955.
- (1988c). Logical Reformulation of Quantum Mechanics. III. Classical Limit and Irreversibility. *Journal of Statistical Physics* 53(3–4): 957–975.
- (1992). Consistent Interpretations of Quantum Mechanics. *Reviews of Modern Physics* 64(2): 339–382.
- (1994). *The Interpretation of Quantum Mechanics*. Princeton: Princeton University Press.
- (1999). *Understanding Quantum Mechanics*. Princeton: Princeton University Press.
- Osnaghi, Stefano, Fabio Freitas, and Olival Freire (2009). The Origin of the Everettian Heresy. *Studies in History and Philosophy of Modern Physics* 40(2): 97–123.
- Schlosshauer, Maximilian (2004). Decoherence, the Measurement Problem, and Interpretations of Quantum Mechanics. *Reviews of Modern Physics* 76(4): 1267–1305.
- Shimony, Abner (1993). *Search for a Naturalistic World View: Volume 1: Scientific Method and Epistemology*. New York: Cambridge University Press.
- Wigner, Eugene (1963). The Problem of Measurement. *American Journal of Physics* 31: 6–15.

Chapter 13

From Do-it-yourself Quantum Mechanics to Nanotechnology? The History of Experimental Semiconductor Physics, 1970–2000

Christian Kehrt

Given the hype surrounding nano-technology (NT), few people realize that some of us have been practicing NT for over 30 years—we just didn't call it NT. (Kroemer 2005, 959)

Herbert Kroemer has been influencing the field of semiconductor physics, surface science and quantum electronics from the 1960s to the present day.¹ From his perspective, nanotechnology is mainly a re-labeling of the well-established and highly dynamic field of experimental semiconductor physics that traditionally stands between science and technology. He denies the claim of novelty by arguing that recent developments of so-called nanotechnology are rooted in the experimental practices of semiconductor physics from the early 1970s, when do-it-yourself quantum mechanics was made possible by new research technologies, such as Molecular Beam Epitaxy (MBE).² Nevertheless, the aim of this paper is not to follow Kroemer's defensive and rather skeptical argument and to reduce nanotechnology to the traditions of semiconductor physics or surface science (Kehrt and Schüßler 2010). Instead, I propose to carefully contextualize the discourse of nanotechnology in the 1990s from a historical perspective and to look for continuities and changes in specific scientific practices within a wider societal and political framework.³

¹In 2000, Kroemer received the Nobel Prize together with Zhores I. Alferov and Jack Kilby in physics for his work on semiconductor heterostructures and optoelectronics.

²A similar observation is made by the pioneer of MBE, John Arthur, who heard a radio broadcast on nanotechnology that left him "a bit impatient because of the heavy emphasis on the more flamboyant future possibilities that research may provide" (Arthur 2002, 190). Nevertheless, he was fascinated by the discussion since obviously he himself had been practicing nanotechnology for over thirty years: "It struck me that for more than thirty years, some of us have been doing this, in one dimension at least, by the process known as molecular beam epitaxy" (Arthur 2002, 190).

³These are results of an interdisciplinary case study, funded by the Volkswagen Stiftung, that I conducted together with Peter Schüßler on the practices and knowledge production of nanotechnology in Munich at the Deutsches Museum. It was based on oral history interviews and bibliometrics of local

My approach within the recent or contemporary history of science starts in the present, but tries to avoid the pitfall of constructing *ex post* a linear genealogy of nanotechnology (Rheinberger 2006; Söderqvist 1997). From my perspective, a “history of nanotechnology” is not possible in the sense of inventing milestones or traditions of nanotechnology, as its proponents intend to do, or by critically showing that scientific work at the nanoscale had been practiced in many fields already in the course of the twentieth century. Nanotechnology is a boundary object that has different social relevance for different groups of actors (Gieryn 1999, 5–6; Star and Griesemer 1989, 70; Kehrt and Schüßler 2010). Therefore, a closer look at specific scientific communities, their strategies, and their research traditions is necessary to explain why the rather vague and often stereotypical—but highly popular—futuristic discourse has been actively shaped by semiconductor physicists. These scientists have been working, as Kroemer noted, since the early 1970s at the nanoscale but only identified themselves as “nanoscientists” at the turn of the twenty-first century. So basically, I will tell a story of experimental or “do-it-yourself quantum mechanics” (Esaki 1992) that starts in the 1970s, culminates in the 1980s and looks for new orientations in the 1990s.

How did this dynamic field between science and technology evolve in this period? Is experimental semiconductor physics at the quantum level a case of so-called technoscience (Latour 1998; Nordmann 2006), mode II science (Gibbons, Limoges, and Nowotny 1994), finalized science (Böhme 1978), or—to cite Paul Forman—postmodern science (Forman 2007; Carson, Kojevnikov, and Trischler 2008)? As philosopher of science Joachim Schummer and many others have pointed out, nanotechnology is an umbrella term that encompasses almost every branch of science and thus is not helpful in specifying new fields of research (Schummer 2009; Decker 2006, 42). However, the wide use and active participation of scientists in the visionary nanotechnology discourse has real impact on the formation of local networks, research agendas, and careers. In the case of solid-state science, the reference to the rather vague idea of “nanotechnology” helps scientists to cross disciplinary boundaries and work with new experimental systems and methods from the life sciences. Besides this intra-scientific, trans-disciplinary dimension, the participation in the public nanodiscourse highlights the extra-scientific, social and technological significance of this research that appears to be related to a future key technology. This seems to cohere with Forman’s claim that the downgrading of science and the upgrading of technology indicates an epochal change that took place in the 1980s:

nanotech networks. All interviews and translations in this paper are done by the author. I also want to thank Michael Eckert and Paul Forman for their helpful comments and critical remarks and Fred Koch for his careful reading of the Klaus von Klitzing story.

Indeed, the transition from modernity to postmodernity, whatever else it involves, involves an upward revaluation of technology and a downward revaluation of science, leading to a far-reaching change in the culturally presupposed relationship between science and technology. (Forman 2010, 160)

More particularly, my argument is that nanotechnology is a funding and media strategy scientists use to pursue undirected free research at universities with the intent of emphasizing the technological relevance of their research and to still be able to freely play with molecules (Kehrt 2011). Therefore, the reorientation of semiconductor physics after the end of the Cold War reflects a general ideological shift from science to technology, without necessarily abandoning basic research or aiming merely to realize technological goals (McCray 2005; Johnson 2004).

This paper is based on a case study of local nanotech networks in the city of Munich, Germany. The high-tech region of Munich—with Siemens as a major employer for physics students, two high-ranking physics departments at the Ludwig Maximilians University (LMU) and the Technical University Munich (TUM), the resulting Center for Nanoscience (CeNS), and the national excellence network Nanoinitiative Munich (NIM)—is a good place to study nanoscience networks. While much work has been done on the discourses and futuristic background of nanotechnology, there are few studies that explicitly deal with the scientific networks, practices, and historical dimensions involved. The 1970s recently gained attention in the general history community (Jarausch 2008; Doering-Manteuffel and Raphael 2010; Trischler 1999; 2001). Unfortunately, there are almost no studies in the history of science about developments in microelectronics, semiconductor physics or experimental quantum mechanics in Germany that deal with developments in the period from the 1970s to the present.

13.1 New Research Technologies at the Quantum Level

A closer look at the research practices of nanoscientists at TUM and LMU shows that scientists in the field of semiconductor physics conduct experiments with quantum effects in semiconducting materials, such as quantum wires and dots, that confine the movement of electrons in two, one and zero dimensions. These nanostructures provide the opportunity to investigate new physical phenomena and promise new technological possibilities. Quantum phenomena of electron transport in two dimensional electron gases were first predicted theoretically by John Robert Schrieffer, who “pointed out that for high electric fields in surfaces of high perfection it would be necessary to consider quantum effects” (Landwehr

1975, 50). In 1966, researchers at Bell Labs proved quantum behavior of electrons in two-dimensional electron gases (Fowler et al. 1966, 901). This was the beginning of a new research field that dealt with electron transport at the quantum level.⁴ Quantum states were realized in experimental systems with ultra-pure silicon samples or heterostructures of molecular thin layers of gallium and arsenide with a high electron mobility in vacuum conditions at low temperatures, but also at room temperatures.

The intense contemporary interest in the physics and technology of thin films, surfaces and ultra-thin multilayer heterostructures has been motivated, at least in part, by the remarkable development of the solid-state electronics industry in the past thirty years. These areas are intriguing because, apart from their obvious technological importance, they offer the possibility of new effects that are not present in the bulk of a solid. (Dingle 1975, 21)

These new experimental possibilities at the quantum level were based on advances in materials (Mönch 1973, 242). The production of high-quality silicon wafers demanded new and extremely costly silicon growth and production techniques that were only realizable in large-scale industries. The aim was to build electronic devices with better qualities and performance. Especially the metal oxide semiconductor field effect transistors (MOSFETs) stimulated new research about electron transport in semiconductor surfaces and interfaces (Eckert and Schubert 1986, 200).⁵ MOSFETs were developed in the 1960s and allowed an increase in performance through the ability to integrate more transistors and connections on a chip (Bassett 2002; Eckert and Schubert 1986, 200).

With the beginning of the seventies, a new era became visible that has been directly connected with extreme demands concerning high packing densities (in integrated circuits) or homogeneity (in high power devices). (H. Hermann, Herzer, and Sirtl 1975, 281)

In contrast to the invention of the transistor, the development of the silicon MOSFET is based on technological advances and the control of surface phenomena and not so much on theory (Ernest Braun and MacDonald 1978, 101; Morris 1990, 85;

⁴The fact that these quantum effects depend on the size of the devices and materials involved corresponds to the formal but vague definition of nanotechnology that assumes new effects at the nanoscale.

⁵The MOSFET is a sandwich-device built with layered materials of semiconducting silicon, conducting metal and non-conducting silicon dioxide. It can control the flow of electrons in the surfaces of silicon layers with the help of a thin metal film (or more recently, polysilicon). This gate electrode steers the flow of electrons by inducing a conducting channel between two electrodes called "source" and "drain." This principle allows for amplifying signals and constructing electronic switches.

Handel 1999, 208). Nevertheless, practical questions about the purity of the materials involved and the resulting “considerable renaissance in materials research and particularly diagnostic techniques” (H. Hermann, Herzer, and Sirtl 1975, 281) also led to research that went far beyond the daily business of industrial research labs:

Quite apart from its technological importance in the form of the metal oxide semiconductor field effect transistor (MOSFET), the space charge layer on a semiconductor surface is a fascinating physical system. Under the influence of a surface electrical field, electrical charge is accumulated on the surface in a narrow channel typically 10 lattice constants in depth [...]. The electrons in the surface space charge layer are bound in their motion normal to the surface in discrete quantum mechanical states. They are free with respect to their motion parallel to the surface, and electron states thus form a two-dimensional band—the electric subband, as it is called. (Koch 1975, 79)

The early 1970s can be seen as a period with an experimental breakthrough in quantum mechanics and the beginning of a new and highly dynamic field of research. At that time, IBM researcher Leo Esaki proposed the so-called artificial superlattice where electron tunneling determines electron transport:

It should be possible to obtain a novel class of man-made semiconductor materials, at least as far as electronic properties are concerned, and one expects the properties to depend not only on band parameters of the host crystal, but also on the characteristics of the superlattice. (Esaki and Tsu 1979, 61)

According to Esaki, new instrumental practices and technological equipment were crucial for this kind of experimental work at the quantum level that allowed one to operate with theoretical assumptions formulated in the early 1930s:

A general tendency in those early days of quantum mechanics existed to try to explain any unusual effects in terms of tunneling. In many cases, however, conclusive experimental evidence of tunneling was lacking, primarily because of the rudimentary stage of material science. (Esaki 1974, 1149)

With the development of new research technologies such as MBE, it became possible for theories, models, and concepts of quantum mechanics from the 1930s to be realized in experimental physics:

Weimann: I mean, quantum wells, quantum films [...] or potential pots as we called them back then. They were calculated in the 1930s, when quantization was introduced and we had gotten used to it. There were models, but only conceptual models. Now [in the 1970s], for the first time, we really could use and create it in components and in semiconductors and really see that the qualities in the components improved.⁶

Besides silicon, which was favored by industry due to its stable surface properties and cleanliness, gallium arsenide compounds were also of interest since they promised future devices with superior performance in comparison to silicon MOSFETs. Especially scientists at universities, those with a greater interest in basic physical processes, moved to experimental systems with III–V element semiconductors (Ernest Braun and MacDonald 1978, 138).

One key research technology that made quantum experiments possible was MBE. It enables the precise tailoring of material structures at the nanometer or Angstrom level, so that quantum phenomena determine the transport of electrons. This research technology was developed at AT&T Bell Laboratories by Alfred Y. Cho and John R. Arthur in 1970 (Cho 2004, 199); both were interested in surface phenomena.⁷ At the center of MBE is an ultra-high vacuum chamber with several heating pots that contain semiconducting materials, such as gallium and arsenide, that evaporate and finally condensate in ultra-thin “nanolayers”:

Thus, it has been possible to produce a large range of unique structures including quantum well devices, superlattices, lasers etc., all of which benefit from the precise control of composition during growth. Because of the cleanliness of the growth environment and because of the precise control over composition, MBE structures closely approximate the idealized models used in solid state theory. (Arthur 2002, 189)

Scientists spoke of “band gap engineering” and “artificial atoms” that are created by new research technologies and simultaneously promise new high-speed electronic devices as well as new, rather fundamental scientific discoveries and principles (Esaki 1985, 27; Capasso 1987). According to Terry Shinn and Bernward

⁶“Weimann: Ich meine, Quantenfilm oder Quantenbrunnen oder Quantentröge oder Potenzialtöpfe haben wir es eigentlich früher genannt. Die hat man in den 30er-Jahren schon gerechnet. Das kam auf, nachdem man die Quantisierung eingeführt und sich an die gewöhnt hatte. Das gab ja die Modelle, aber immer nur als Gedankenmodelle. Hier konnten wir das jetzt wirklich im Bauelement, im Halbleiter ausnutzen, herstellen und tatsächlich auch sehen, dass man sehr viel bessere Bauelementeigenschaften bekommen hat.” (Interview with Weimann, 20 February 2008).

⁷First attempts to grow III–V element heterostructures go back to Siemens laboratories in the 1950s (Günther 1958).

Joerges, a research technology brings different actors from science and industry, electronics, semiconductor physics and also theoretical physics together (Joerges and Shinn 2001). In Germany, it was Klaus Ploog who pioneered MBE in the 1970s at the Max Planck Institute for Solid State Research in Stuttgart.⁸ Kroemer also worked with MBE, as well as Gerhard Abstreiter and Günter Weimann at the Walter Schottky Institute (WSI) in Munich. In contrast to the tunneling microscope or electron microscope, this widespread research technology did not gain much attention beyond the realm of involved experts. Only with the new interest in the origins of nanotechnology was it identified as a precursor of today's nanotechnology (McCray 2007).

13.2 The 1970s: A New Quantum Generation

The biographies of leading Munich scientists like Jörg Kotthaus, who founded Munich's CeNS in 1998, or Abstreiter, director of the WSI, point at the origins of today's nanotechnology research projects in the early 1970s. Both belong to the generation that studied and worked in Munich in the 1970s and then actively shaped local nanotechnology networks and research projects in the 1990s. In an interview, Kotthaus stated:

Esaki [...] started to work with Molecular Beam Epitaxy at IBM in the early 1970s. And that is what fascinated me completely. I have to say, for me, [...] the beginning was when people started to build artificial semiconductors by layering materials. That was essentially the beginning of experimental nanoscience, if you leave Feynman out.⁹

In this passage, the Munich scientist distances himself from the official storyline of nanotechnology that starts with a thought experiment by Richard Feynman. In an after-dinner speech in 1959, the pioneer of quantum electrodynamics came up with the idea that it should be possible to build electronic structures with single atoms and electrons. The ex post reference to Feynman's long-unnoticed talk is an invention of traditions by which nanoscientists try to emphasize the credibility of their research. Feynman's slogan "there is plenty of room at the bottom" then became the official headline of the US nanotechnology initiative at the turn of the

⁸Interview with Klaus Ploog, 1 July 2008.

⁹"Kotthaus: Ich meine Esaki hat bei IBM die Molekularstrahlepitaxie angefangen Anfang der 70er-Jahre. Und das ist auch das, was mich völlig fasziniert hat. Da muss ich sagen, für mich fing es an, als Leute künstliche Festkörper gebaut haben durch Schichtung von Materialien. Und das war auch im Grunde genommen der Beginn der Nanowissenschaften im experimentellen Bereich, also wenn man Feynman mal weglässt." (Interview with Kotthaus, 19 January 2006).

millennium (Nordmann 2006; Junk and Riess 2006). Nevertheless, Kotthaus does not locate the origins of today's nanotechnology endeavors in Feynman's speech but in the research traditions and experimental practices of semiconductor physics in the early 1970s.

This new experimental work at the quantum level was enthusiastically pursued at TUM, especially in Koch's research group.¹⁰ At the University of Maryland, Koch was already experimenting successfully with electrons that were bound in metal surfaces by magnetic fields (Doezema and Koch 1972). He was fascinated by the idea of applying this approach to semiconductors to study quantum behavior. In an interview on 15 June 2009, Koch explained:

Koch: Epitaxy. When you build layered structures. That goes back to important things that Esaki had done. I was there, in the USA. Leo Esaki was one of the first who dreamed of growing semiconductors in such dimensions that something would happen [if you built in electrons], because he also took the slow electrons into account. And if you build such electrons into nanostructures, [...] if they are confined to certain dimensions so to say, then their properties will change.

Kehrt: And that's exciting?

Koch: That's absolutely exciting. That is completely fundamental physics. That is the wave mechanics of the 1930s; that is where it was recognized. That is Heisenberg and Sommerfeld. Sommerfeld not so much, but Heisenberg and Max Planck and so forth. So the whole quantum physics of the electron is involved.¹¹

Then in 1972, Koch took the chance to start a new branch of semiconductor physics at TUM. The appointment of distinguished American scientists was meant to help close the knowledge gap between the United States and Germany; the latter had lost ground in cutting-edge fields like semiconductor physics and

¹⁰Kotthaus was Koch's assistant in 1973. Abstreiter was his first PhD candidate.

¹¹"Koch: Epitaxie. Dass man eine Schichtstruktur aufbaute. Und das geht jetzt auch einher mit wichtigen Dingen, die Esaki gemacht hat. Ich war in den USA dabei. Also Leo Esaki war einer der ersten, der davon träumte, Halbleiter in solchen Dimensionen zu wachsen, dass sich was tun würde, weil er auch die langwelligen Elektronen erkannte. Und wenn ich solche Elektronen in Nanostrukturen einbaue oder habe oder die Elektronen erscheinen dadurch, dass man sie injiziert oder irgendwas macht, dass ein Elektron da ist [...] und wenn dann ein Elektron in solchen Dimensionen sozusagen beherbergt ist, eingesperrt ist, dann ändern sich seine Eigenschaften.

Kehrt: Und das ist spannend?

Koch: Das ist absolut spannend. Das ist ganz grundlegende Physik. Das ist die Wellenmechanik der 30er-Jahre, da hat man das erkannt. Das sind Heisenberg und Sommerfeld, Sommerfeld nicht so richtig, aber Heisenberg und Max Planck usw. Also die ganze Quantenphysik der Elektronen kommt da zum Tragen." (Interview with Koch, 15 June 2009).

electronics after World War II. TUM introduced the American department structure to create better and supposedly more successful learning and research conditions (W. Hermann 2006, 505). But so far, the research focus at TUM's physics department was on nuclear physics due to the strong influence of Heinz Maier-Leibnitz (1911–2000) and Rudolf Mößbauer.¹² In the early 1970s, Koch had the chance to start a completely new direction of experimental physics. At TUM he had a distinguished position where he could basically build everything up from the very beginning and cooperate closely with Siemens:

Well, that really opened up my eyes, when I saw that here [in Munich] you have Siemens just around the corner. I met the Siemens people. [They said:] we will provide you with samples. In the US I couldn't compete with IBM or Bell Labs. They had their own research teams. And here in Munich I saw the chance, since there was nothing going on in semiconductor physics at all. The whole physics department was based on nuclear physics, nuclear methods, Mößbauer, Maier-Leibnitz and the research reactor on campus over there [directly opposite Koch's office].¹³

Abstreiter, one of the five most-cited authors in semiconductor physics (Tsay, Jou, and Ma 2000, 505), also identifies Koch's group at TUM as the starting point of nanoscience in Munich: "I was the first doctoral student at the SFB (*Sonderforschungsbereich*, collaborative research center), also the first doctoral students of Prof. Koch's professorship that was newly established in 1973. You could roughly say that it was a kind of precursor to nanoscience, this special research field."¹⁴ This so-called *Sonderforschungsbereich* investigated quantum phenom-

¹²Maier-Leibnitz had a strong influence in the realm of nuclear physics in Munich as well as on the German nuclear research. At TUM, he held a chair in Technical Physics. He founded the first research reactor, the so-called atomic-egg (*Atomei*) that was the nucleus of the Garching research campus. He also motivated Mößbauer to return to Munich from CalTech and was a key figure in establishing the physics department structure at TUM (Eckert 1988). Mößbauer studied physics at TUM under Maier-Leibnitz. In 1961, he received the Nobel Prize for the discovery of the so-called Mößbauer Effekt—based on the investigation of recoil-free emission and absorption of gamma ray photons by atoms bound in solids.

¹³"Also mir sind die Augen aufgegangen. Dass ich sah, Siemens hier vor der Haustür. Ich habe die Siemens-Leute getroffen gehabt. Wir beschaffen euch die Proben. In den USA konnte ich ja nicht mit IBM und Bell Labs konkurrieren. Die hatten ihre eigenen Forscherteams. Und ich sah diese Chance hier in München, wo es Null Komma Nichts an Halbleiterphysik gab. Das ganze Department war auf Kernphysik, kernphysikalische Methoden, Mößbauer, Maier-Leibnitz, der Reaktor da drüben." (Interview with Koch, 15 June 2009).

¹⁴"Ich war da in dem SFB der erste Doktorand, auch der erste Doktorand im Lehrstuhl von Professor Koch, der '73 da neu aufgebaut wurde und man könnte grob sagen, das war so eine Art Vorläufer für Nanowissenschaften, dieser Sonderforschungsbereich." (Interview with Abstreiter, 22 November 2007).

ena in the surfaces of semiconductors. It was a highly successful research effort started in 1978, in which two future Nobel laureates, von Klitzing and chemist Gerhard Ertl, were working and also where the first scanning tunneling microscope (STM) was introduced in Munich.¹⁵ In the 1990s, these approaches would probably not have been called surface chemistry or surface science but nanotechnology (Mody 2004, 364).

13.3 Von Klitzing's Nobel Prize: The Discovery of the Quantum Hall Effect

The discovery of the Quantized Hall Effect (QHE) was the result of systematic measurements on silicon field effect transistors—the most important device in microelectronics. Such devices are not only important for application but also for basic research. (von Klitzing 1985, 316)

The discovery of the QHE by von Klitzing in 1980 is a milestone in the field of experimental semiconductor physics. Its origins go back to the early 1970s with the intensifying experimental work on quantum effects in two-dimensional electron gases: “The first indications for the QHE were already obtained by von Klitzing in 1974, when he measured the magnetoresistance of a MOS Hall bar between the current contacts and observed a plateau” (Landwehr 2003, 2). Von Klitzing was appointed to be professor at TUM while he still was at the high magnetic field facility of the Max Planck Institute for Solid State Research in Grenoble, doing the decisive Hall measurements.¹⁶ But the discovery of the QHE was not just a Munich or Bavarian story that then resulted in later discoveries in Grenoble. At that time, many research groups worldwide, especially in Japan, were interested in localization phenomena and conducted Hall resistance and magneto transport measurements.¹⁷ In 1977, Japanese scientist and theoretician Tsuneya Ando from the department of physics of the Tokyo Institute of Technology was a

¹⁵Interview with Behm, 16 December 2008.

¹⁶Koch strongly supported the appointment of von Klitzing as professor to be able to conduct experiments like Gottfried Landwehr in Würzburg.

¹⁷In 1879, Edwin Hall discovered that if an electric current in a conductor flows through a magnetic field, that field exerts a separating force on the charge carriers so that an electrical field builds up perpendicular to the magnetic field and to the current's direction. If the Hall Effect is produced in a two-dimensional semiconductor at low temperatures, a series of steps appear in the Hall resistance as a function of magnetic field instead of a monotonic increase. Von Klitzing realized that the Hall conductivity of a two-dimensional electron system is quantized in whole fractions of e^2/h (Thouless 1984, 147; Landwehr 2003, 9). This Quantized Hall Effect is taken as a natural constant to define the Ohm resistance with an uncertainty better than 10^{-6} ; it does not depend on the material of the samples.

visiting scholar at TUM. In Munich, he introduced the possibility of constructing MOS Si probes with a multicontact geometry as they were produced in Japan. He also had data from Shinji Kawaji, a semiconductor physicist from Gakushuin University Tokyo, who measured—besides the normal longitudinal resistance—the so-called transversal Hall resistance with the now-famous von Klitzing steps. According to Koch, Tsuneya Ando pointed to the relations between these steps and the phenomenon of localization.¹⁸ Koch remembers clearly that his Japanese colleague interpreted these Hall resistance steps as a mathematical artifact that was founded in the phenomenon of localization of electrons and thus saw these only as approximate quantum measurements. This was the general tendency of the early discussions about the Hall steps before von Klitzing's discovery. Research groups in Japan that conducted Hall measurements in semiconductors could also show plateaus in the Hall resistance values. But it was von Klitzing who realized in the early 1980s in Grenoble that these energy plateaus are quantized stepwise with very high precision.

The silicon MOS-structure that was later used by von Klitzing in the high magnetic field facility of the Max Planck Institute for Solid State Research in Grenoble was conceived and designed at TUM and then produced by Siemens. This so-called MOS Hall bar, a high-quality MOSFET with high electron mobility could only be provided by industrial research labs.¹⁹ Koch's research group designed the masks for the lithography process of such multicontact probes at Siemens. The probes that then were used for the measurements in Grenoble resulted from these. In an interview, Koch stated:

That structure was built for us, the way Hitachi did it for their researchers and neighboring universities. And with that we gained a basic insight. Von Klitzing's true merit was not in the steps in the diagram—the Japanese scientists had them already and I had Japanese visitors here who showed me this data and so on. It wasn't the insight that there were steps in it. One of our theoreticians was sitting in the room next door. Back then he said: forget about the steps, that is a mathematical artifact. But von Klitzing realized: wait a second, there is a natural constant in there. And the real meaning of his discovery was to point that out to an infinite number of places behind the decimal point—no one has shown yet, how many places it is. Or

¹⁸Personal communication between Koch and the author, 6 June 2011.

¹⁹The other samples and control measurements were conducted by Michael Pepper at Cambridge's Cavendish laboratories with samples produced by the Plessey company (the Munich group worked with Siemens). The Hall bar structure is a sample configuration that measures the different components of the conductivity tensor. So in a MOS probe, there are four additional contacts besides the usual source and drain contacts to measure electrical resistance and the Hall voltage.

it's so many places that you don't even have to ask. It's a natural constant that emerges from this experiment. That's von Klitzing's true merit.²⁰

It is the merit of von Klitzing to have realized that:

[T]he Hall resistance at particular, experimentally well-defined surface carrier concentrations has fixed values which depend only on the fine-structure constant and speed of light, and is insensitive to the geometry of the device.²¹ (von Klitzing, Dorda, and Pepper 1980, 494)

Gerhard Dorda from Siemens, who wrote the decisive paper together with von Klitzing, did not just provide the samples; he himself conducted research on quantized phenomena in the early 1970s. Dorda was confronted with measurements in MOSFETs that could not be understood with the band structure model. He had to assume quantum states to explain the transport behavior in inversion layers underneath the surface of semiconductors:

The rapid development of MOS devices with traceable surface characteristics has led to measurements of the physical properties of semiconductor inversion layers. In almost all considerations it was supposed that the band structure of the bulk is also applicable to the surface. Schrieffer has pointed out that in the interpretation of transport properties of inversion layers a quantization of carrier motion perpendicular to the surface has to be considered. (Dorda 1971, 2053)

In 1972, Dorda first presented his results at an international conference in Hawaii, where he also met Koch, who went to Munich within a year.²² In contrast to Dorda, Koch and his team at TUM, as well as von Klitzing, had more freedom to

²⁰“Koch: Diese Struktur wurde für uns geschaffen, genau so wie Hitachi für ihre Forscher das machten und benachbarte Hochschulen. Und daraus wurde dann die Grunderkenntnis gewonnen. Der wahre Verdienst von von Klitzing sind nicht die Stufen, die haben die Japaner vorher gehabt und ich hatte japanische Besucher hier, die mir diese Daten zeigten und so was. Es war nicht die Erkenntnis, dass da Stufen drin sind. Wir hatten einen Theoretiker, der saß im Nebenraum. Der sagte mir damals: Vergiss die Stufen, das ist mathematisches Artefakt. Von Klitzing hat da erkannt: Augenblick mal, eine Naturkonstante steckt darin. Und die Bedeutung seiner Entdeckung war wirklich darauf hinzuweisen, dass bis auf unendlich viele Stellen hinterm Komma, noch niemand gezeigt hat, wie viele Stellen es sind. Oder es sind so viele Stellen, dass man gar nicht danach fragen muss, ist das eine Naturkonstante, die aus diesem Experiment raus kommt. Das ist der wahre Verdienst von von Klitzing.” (Interview with Koch, 15 June 2009).

²¹Koch explained these details in written form to the Nobel committee and also requested the inclusion of the Japanese colleagues.

²²Koch was appointed to TUM on 1 December 1972.

deal with quantum phenomena that were not the focus at Siemens. Dorda himself said this kind of research was tolerated but not really motivated by Siemens, a kind of “submarine science” (“*U-Boot-Tätigkeit*”) that takes place unobserved and then suddenly pops up to the surface with a new discovery. Dorda explains:

That was always my motivation: to deal with fundamental questions. I pursued that along the way and also within the universities. And they [at Siemens] acknowledged that in so far as they said, he is a typical scientist/researcher. We called that submarine work; it remained underground, nobody knew that this was happening, because it was not condoned. And then, when I was successful, I resurfaced, so to say [...]. They [the managers at Siemens] said I was a typical scientist of this kind. After the Nobel Prize, I of course got absolute freedom to do whatever I wanted. I was the last one. They said I was the last Mohican at Siemens, because before this time, before they started working with semiconductors at Siemens, they had also discovered the III-V semiconductor at Siemens. That was in Erlangen.²³

Von Klitzing’s success was seen as a triumph of experimental physics (Landwehr 2003, 11; Thouless 1984, 147) and also as a result of basic research: “Not applied, but basic research led to a very substantial improvement of the accuracy of the resistance standard” (Landwehr 2003, 12). Interestingly, the first jury member for *Physical Review Letters* initially refused von Klitzing’s decisive paper for publication since it did not contain enough theory (Landwehr 2003, 15). In fact, von Klitzing’s discovery was possible without direct theoretical prediction and was based on experimental laboratory work with refined methods and measurement techniques. Nevertheless, this kind of experiment with quantized phenomena is based on quantum theory and a creative interaction between experiment and quantum theories about the behavior of electrons in semiconductors.

²³“Dorda: Das ist mein inneres Bestreben, immer so Grundlagenfragen zu erörtern, ich habe das nebenbei weiter getrieben und eben über die Universitäten. Und sie haben das dann anerkannt, insofern dass sie sagten, na ja, ich bin der typische Forscher. Man nennt das U-Boot-Tätigkeit, also im Untergrund, ohne dass jemand was, weil es nicht gebilligt war, wusste. Und wenn ich dann erfolgreich war, bin ich wieder aufgetaucht sozusagen. Und die sagen, wenn man das so macht, ist es auch okay. Es ist also tolerierbar. Und ich bin ein typischer Forscher dieser Art, haben sie gesagt. Und dann, also nach dem Nobelpreis selbstverständlich, habe ich dann absolute Freiheit bekommen. Ich konnte dann quasi machen, was ich wollte. Ich war der Letzte. Sie sagen mir, ich bin der letzte Mohikaner bei Siemens, weil noch vor dieser Zeit, also als sie angefangen haben bei Siemens, mit Halbleitern zu arbeiten, da haben sie bei Siemens ja auch die III–V-Halbleiter entdeckt. Das war doch in Erlangen.” (Interview with Dorda, 17 June 2008).

13.4 The 1980s: Founding the Walter Schottky Institute

The founding of the WSI in 1988 is closely related to the prestigious event of von Klitzing's Nobel Prize in 1985. It has to be seen in the context of an increasing competition for the best scientists in a global microelectronics race. The idea of such an interdisciplinary center to facilitate knowledge flow between universities and industry was formulated by Abstreiter and Ploog after a visit to Japanese research facilities that were equipped with MBE systems.²⁴ In comparison to industrial research labs, the WSI pursues rather basic and long-term perspectives:

Kehrt: But what you do here is science, not engineering science?

Abstreiter: That's in-between. We also have engineering, but not in the sense of classical engineering, we rather look for new principles.²⁵

The WSI holds a strategic middle position between basic science and technology development that did not exist previously. However, in the 1990s, following growing competition in the globalized semiconductor industry, Siemens—like many other big companies—cut down its research department and focused on shorter innovation cycles. That was the time when basic research in semiconductor physics lost contact with industry, and nanotechnology was entering the focus of such scientists as Kotthaus, Koch or Abstreiter, who had been working at the quantum level with semiconductor “nanostructures” since the early 1970s:

Kehrt: There was a move away from microelectronics as a key technology?

Koch: Yes.

Kehrt: In these research fields that were previously closer to microelectronics?

Koch: That's it. Right. That's what Abstreiter and I and Kotthaus did in the early 1970s until the 1980s, but in the middle of the 1980s that began to diverge. And then in the 1990s, when the companies also withdrew; that's when such nano-institutes did things that were far from real applications.²⁶

²⁴Interview with Abstreiter, 22 November 2007.

²⁵Kehrt: Aber was sie hier machen, das ist Naturwissenschaft, keine “engineering science”?

Abstreiter: Das ist zwischendrin. Wir haben auch “engineering”, aber im Sinn nicht das klassische “engineering”, sondern wirklich neue Prinzipien.” (Interview with Abstreiter, 22 November 2007).

²⁶Kehrt: D.h. es gibt so eine Wegorientierung von der Mikroelektronik als Schlüsseltechnologie?

Koch: Ja.

In the 1990s, undirected basic research in semiconductor physics lost ground and made new orientations and strategies necessary (Angel 1994, 3; Gerybadze, Meyer-Krahmer, and Reger 1997, 20; Hack 1998, 102). Obviously, the general consensus for basic research as it was practiced in the Cold War—especially in fields related to the military, such as semiconductor physics—vanished, and the rise of the life sciences forced semiconductor physics to reorient its research strategies. Furthermore, there was a general crisis in the German innovation system after reunification (Caspar 2007, 76; Nusser 2006, 66–67; Cuhls, Uhlborn, and Grupp 1996, 53; Bundesbericht Forschung). The need for a new visionary technology seemed to be fulfilled by the promises of nanotechnology (Bachmann 1998). The German Ministry for Science and Education (BMBF) was well aware that the Clinton presidential administration in the United States was creating a new nanotechnology strategy and started its own German initiative. The main reason was not to fall behind at the beginning of radical new technological developments but to support the possibility of future key innovations.²⁷ Von Klitzing predicted a blossoming of nanoelectronics based on future quantum devices before the ultimate physical limits of miniaturization were reached. He criticized the reduction of basic research and the dominance of economic restraints, and he argued for long-term perspectives in—and basic research on—quantized phenomena in semiconductors (von Klitzing 1995, 26).

13.5 Munich Nanoscience Networks

The perception of nanotechnology as a new scientific trend began in the late 1980s and early 1990s when new developments in the field of semiconductor physics allowed for designing nanostructures for basic science as well as for future technologies:

It is anticipated that the independent technologies will be married in the next decade, with consequent production of structures that are atomically engineered in all three dimensions to nanometer design rules. (Kelly 1987, 264)

Due to the advancement of research and materials processing technologies, it became possible to artificially design structures that confined the movement of elec-

Kehrt: In diesen Forschungsfeldern, die vorher näher an der Mikroelektronik dran waren.

Koch: So ist es. Ganz richtig. Also das, was Abstreiter und ich und Kotthaus in den frühen 70er-Jahren machten und in die 80er hinein [...]. Mitte der 80er fing sich an, das zu divergieren. Und dann in den 90ern, als die Firmen sich zurückziehen, dann haben solche Nanoforschungsinstitute ganz andere Dinge getan, die weit weg sind von der wirklichen Anwendung.” (Interview with Koch, 15 June).

²⁷Interview with Secretary of BMBF, Wolf-Michael Catenhusen, January 2007.

trons in one and zero dimensions (Kuchar, Heinrich, and Bauer 1990; Reed 1993, 118). “Top-down” techniques, such as electron beam lithography, were coupled with “bottom-up” approaches from chemistry and the life sciences to artificially design new materials, such as nanotubes and quantum dots, that do not exist in nature: “The study of quantum dots is the result of tremendous advances in molecular beam epitaxy, dry processing, and advanced lithography” (Smith 1990, 10). Also, the STM (invented in the early 1980s by Heinrich Rohrer and Gerd Binnig at IBM) was identified as a key instrument of nanotechnology. It allowed experiments with self-organizing processes of molecular clusters in very different fields of research (Hennig 2011). Already in 1988, an article in *Nature* assumed the possibility of atomic engineering with the help of the STM (Pethica 1988, 301). In a special section of *Science* entitled “Engineering a Small World” in 1991 (Science. Special Section 1991), all topoi that constitute the future nanodiscourse were formulated: the idea of engineering atoms and molecules, the visions of Feynman and Eric Drexler, the processes of self-assembly, the use of biological materials, as well as the key role of the STM. More specifically in the field of semiconductor physics, nanotechnology was associated with the possibility to conduct experiments with quantum dots, nanowires and nanotubes (Corcoran 1991, 78). However, the relabeling of these well-established research fields under the heading and hype of nanotechnology was motivated primarily by science policy considerations, when the Clinton administration started its National Nanoinitiative in 1998, and thus research at the nanoscale became very attractive because of its association with a future key technology.

The founding of the Munich CeNS in 1998 was related to the emerging public nanohype. At the time when the national nanotechnology strategies were formulated, Kotthaus, together with colleagues from the experimental physics department of LMU, quickly realized the potentials of research at the nanoscale and came up with the idea of a center for nanoscience. This local nanoscience network tries to meet the new transdisciplinary, media and economic challenges of science at the turn of the twenty-first century. Obviously, the freedom of scientists to play with molecules beyond established disciplinary boundaries requires other, more flexible forms of interaction and strategies.

Publication statistics show that, at an international and a national level, Munich has a leading position and is a good example for studying general trends in nanotechnology. Research in nanotechnology is mainly taking place at universities (Kostoff, Koytcheff, and Lau 2007, 576) and basic research dominates (Heinze 2006, 113). Also in Munich, the two major universities—LMU and TUM—dominate nanopublications, while only 6% can be located in industrial research labs.²⁸ The fact that local nanotech endeavors are rooted in semicon-

²⁸Result of a bibliometric study of Munich nanotech networks (Kehrt 2011).

ductor physics can be explained by the high density of microelectronics in the Munich area. Siemens especially was the main employer for physics students and had a strong influence on the field and career patterns of semiconductor physics. The main protagonists of the new nanotechnology networks stemmed from the semiconductor physics community. The twenty founding members of CeNS belong to the field of experimental physics, and 65% of the involved professors are located in semiconductor physics or biophysics. This orientation of semiconductor physics toward nanotechnology at the turn of the century is confirmed by an analysis of the leading German journal *Advances in Solid State Physics*.²⁹

In the context of quantum mechanics, it is interesting that the idea for such a bottom-up nanoscience network is related to quantum electronics and its organizational structures in the United States. The founding father of CeNS, experimental physicist Kotthaus, refers to the US Center for Quantized Electronic Structures (QUEST) that he knew through his long contacts at the University of California in Santa Barbara, where he had studied in the 1970s and where Kroemer has been working since the late 1970s. In an interview with the author on 19 January 2006, Kotthaus remembers:

The idea for such a center, frankly speaking, is something that had moved me since the beginning of the 1980s, when I saw how such centers were created in the USA. QUEST was certainly a role model. QUEST meant “Quantum Electronic Structures” and was a close cooperation among scientists at UC Santa Barbara that was truly based on common interests. Back then, I was in Santa Barbara almost every summer for a month or two.³⁰

Such problem-oriented, interdisciplinary centers were pushed in the 1980s to facilitate cooperation between disciplines (and between universities and industry) (Thompson Klein 1992, 36). Nevertheless, the strategies of successful and influential scientists like Kotthaus and Abstreiter changed in the 1990s. Semiconductor physics lost its immediate relationship to industry and had to look for new alliances and visions. Now basic research—even at universities—needs a stronger legitimacy in utility. The university itself has turned into a place for

²⁹A database search of all nano composites in titles or abstracts shows that 35% were written in the years 1990–1999 and 65% during 2000–2008. The word “nanotechnology” appears only since the year 1999, while word composita with “nano” appear earlier.

³⁰“Kotthaus: [...] das heißt die Idee so was zu machen, ehrlich gesagt, hat mich an sich bewegt seit Anfang der 80er-Jahre, als ich gesehen habe, wie in den USA solche Zentren entstanden; Vorbildfunktion hat für mich das QUEST gehabt. Das QUEST hieß eben ‘quantum-electronic structures’ und war eine Zusammenarbeit von Wissenschaftlern in Santa Barbara, die eben wirklich auf gemeinsamen Interessen beruhte und ich war damals praktisch, ja, jeden Sommer ein bis zwei Monate in Santa Barbara [...]” (Interview with Kotthaus, 19 January 2006).

entrepreneurial science. This strategy began with Wolfgang Hermann's appointment as president of TUM in 1995. According to Hermann's new entrepreneurial philosophy, such traditional disciplines as physics, biology or medicine should have a closer relationship to technology (W. Hermann 2006, 931). This new orientation can be observed within Munich nanotech networks. Doctoral students learn to address the media, defend the usefulness of their endeavors and are encouraged to found spin-off companies, such as Attocube, Nanion or Nanotools. These Munich nanotech firms often directly result from PhD work in semiconductor physics.

Despite this new entrepreneurial spirit promoted by nanotech spin-off companies, the research conducted at universities has no direct link to the market, is far away from direct application and follows rather long-term perspectives. Indeed, there are few chances for direct technological development stemming from nanoscience research. This current state of affairs was already realized when nanotechnology was identified as a new research field in the early 1990s. Nanotechnology provides "wonderful tools for science," but it does not offer clear economic or technological perspectives (Ball and Garwin 1992, 766). For example, Don Eigler, who gained public attention through his first manipulation of single atoms by writing "IBM" with xenon atoms, is rather critical of overrated hopes of utility and application:

However, on the nanometer scale, we simply do not have a robust, practical method for mass production. [...] Nanotechnology is now in the single device invention stage, and there is no clear vision about how one could practically integrate devices in a second stage. (Brus and Eigler 1994, 273–274)

This situation, that nanotechnology was rather in the stage of basic research, did not change, although in Munich a dozen small university spin-off companies such as Attocube, Nanion or Nanotools were founded. A closer look reveals that these "nanotechnology" enterprises still do produce high precision scientific instruments and analytic tools to enable basic research at the nanoscale.³¹

³¹There are a dozen firms founded by students of LMU that are conducting research in experimental semiconductor physics and biophysics. These firms use the "nano" label to promote their equipment for scientific research. Nanotools was founded in 1997 by students of Kotthaus. Using the atomic force microscope, they realized that scanning required much stronger tips and thus constructed these special tips to improve research with the instrument (Interview with Bernd Imer, Founder of Nanotools, 10 March 2009). Attocube was founded in 2001, also by scientists of the Kotthaus group. Attocube produces high precision piezo-engines for scientific instruments working in high magnetic fields or in ultra-high vacuum conditions (Interview with Attocube—Prof. Karrai and Dr. Haft—28 January 2009). Nanion, founded in 2002 by Niels Fertig, also a former doctoral student of Kotthaus, uses the patch clamp method to develop labs-on-a-chip. In 1991, Bert Sakmann received the Nobel

Therefore, the reference to university spin-offs and future key technologies is a sign of a new scientific culture that now already positions economic thinking within the realm of the university. But this does not imply that basic research and traditional modes of knowledge production are completely abandoned. In regards to Forman, I argue that the promotion of spin-off companies and the reference to entrepreneurial PhD students is a sign of an ideological shift. While in former times students of semiconductor physics went directly to Siemens, these big companies no longer offer career opportunities. In this context, the formation of nanotech-networks has real effects on career patterns and the strengthening of local science clusters. The nanohype allocates money from the government and supports projects that identify themselves as being related to nanotechnology. In the beginning, CeNS was an informal network to bring scientists together and exchange ideas.³² Then, with the resulting success of the excellence initiative NIM, money from the government was turned into new careers, professorships, and infrastructures.³³

13.6 Nano-biotechnologies. New Forms of Interdisciplinary Cooperation?

In previous sections, Munich nanoscientists were located in the field of experimental semiconductor physics. There are clear continuities from the 1970s to the 1990s related to quantum phenomena of electron transport in low dimensional physical systems. Yet, there are also significant new transdisciplinary developments and changes that cannot be explained by these research traditions. In this regard, Munich nanoscientists emphasize their close cooperation between semi-

Prize for this method to measure electric currents in ion channels between cells to understand the communication between cells.

³²At this point, one could critically ask why or if this rather normal science communication and exchange of new ideas was not possible within the traditional disciplinary and institutional setting of the university.

³³Three professorships, those of Alexander Holleitner, Thorsten Hugel and Christina Scheu, were fully sponsored by money from the NIM. Also the following research projects received funds from NIM: Prof. Philip: Tinnefeld Biophysics (LMU); Prof. Lukas Schmidt-Mende: Hybrid/ Colloidal Nanosystems (LMU); Prof. Dieter Braun: Physical Aspects of Hybrid Nano-Bio Systems (LMU); Prof. Don Lamb: Live Cell Imaging (LMU); Prof. Scheu: Transmission Electron Microscopy of Nanostructures (LMU); Prof. Alexander Högele: Nanophysics (LMU); Prof. Ulrich Gerland: Theoretical Nanophysics (LMU); Prof. Bettina Lotsch: Synthetic Chemistry (LMU); Prof. Ulrich Schollwöck: Theoretical Physics (LMU); Prof. Holleitner: Nano-technology and -materials (TUM); Prof. Friedrich Simmel: Bioelectronics (TUM); Prof. Hugel: Molecular Machines (TUM); Prof. Tim Liedl: Bio Interfaces (LMU). Also the WSI was able to enlarge its research facilities and build a new “nanoscience building” with money from the NIM (Peter Sonntag, general manager of NIM, email communication, 23 October 2009).

conductor physics and the life sciences. In an interview on 20 November 2006, Heckl stated:

In fact that would not have been possible twenty years ago, that someone like Kotthaus—a semiconductor physicist working in clean rooms—suddenly started to touch DNA. He would have never done that in the past. He would have said: “My lab will get contaminated by that kind of organic stuff.” But a lot happened back then, and especially here in Munich with its research environment, because obviously [...] or maybe I’ll put it the other way around [...] that is certainly a reason why we have now, for example, become an “Exzellenzuniversität” (Excellence University).³⁴ Because in many fields things have changed, moved forward.³⁵

According to Munich nanoscientists, a characteristic trait of the Munich nanoscience landscape seems to be the close cooperation between life sciences and semiconductor physics. In fact, at TUM, Erich Sackmann established a school of biophysics and his pupils introduced the STM and atomic force microscope (AFM) to study processes of molecular self-assembly that were then identified as being an integral part of so-called nanotechnology (Mody 2004; Hennig 2011). In the 1990s in Munich, a large biotechnology cluster also emerged near the village of Martinsried (Heßler 2007, 167–187). However, it is not clear in what sense there are direct interdisciplinary cooperations between biotechnology, biochemistry and genetics on the one hand and semiconductor physics and surface science on the other. Do semiconductor physicists really cooperate closely with scientists from the life sciences in concrete interdisciplinary nanoscience research projects?

While early bibliometric studies (Meyer and Person 1998, 203) seem to confirm the interdisciplinary nature of nanotechnology, others rather doubt this claim (Heinze 2006, 111; Schummer 2004, 461). Indeed, a closer look at the Munich

³⁴The German University Excellence Initiative was a national competition between universities for the prestigious title “Excellence University.” This official campaign started in 2005 and aimed at funding cutting-edge research. LMU received the title Excellence University, and the local nanoscience network became the excellence cluster known as Nanosystems Initiative Munich (NIM). LMU Presseinformation 13 October 2006, Entscheidung im Exzellenz-Wettbewerb. “LMU ist Spitzenuniversität”, <http://www.nano-initiative-munich.de>, accessed 15 October 2007.

³⁵“Heckl: Also das hat es eben vor 20 Jahren nicht gegeben, dass jemand wie der Kotthaus, der also ein Halbleiterphysiker mit Reinraumlabor ist, plötzlich eine DNA anlangt, ja. Das hätte der nie gemacht früher. Der hätte gesagt: “Meine Kammer wird verunreinigt durch so ein organisches Gezeugse.” Also, da ist schon viel passiert auch, aber gerade auch bei uns natürlich auch in München in dem Umfeld, weil natürlich, oder ich sage es jetzt mal andersrum und das ist mit Sicherheit auch ein Grund, warum wir jetzt, zum Beispiel, eine Exzellenzuniversität geworden sind. Weil in vielen Feldern sich etwas bewegt hat, was vorwärts gegangen ist.” (Interview with Heckl, 20 November 2006).

nanoscience landscape and their transdisciplinary cooperations show that experimental physics is open to life sciences methods and approaches, but there is no real interdisciplinary cooperation between different disciplines.³⁶ Therefore, I argue that the boundary object of nanotechnology—with its rather vague and indefinite character—opens up new venues and spaces for research beyond disciplinary boundaries, but it does not necessarily lead to strong interdisciplinary interaction or the emergence of nanotechnology as a distinct scientific discipline.

Simmel's experimental work with DNA is an example for new transdisciplinary approaches in the direction of nanobiotechnology or synthetic biology. In his PhD, Simmel analyzed quantum dots in Kotthaus's research group (Simmel 1999). Then as a postdoctoral researcher, he went to Bell Labs in New York, in a period when cutting-edge basic research was still promoted there. In Bernard Yurke's research group, they developed a so-called nanotweezer, based on DNA strings, that can open and close and thus possibly lead to the foundation of new principles for future molecular scale devices (Yurke et al. 2000). In the beginning of this new research, the hopes were high to be able to "construct simple, machine-like nanomechanical devices" (Simmel and Dittmer 2005, 285). They used DNA to create new artificial nanosystems that do not exist in nature. Characteristic of Simmel's work is the radical change of experimental systems. Simmel explains:

I have to say, frankly speaking, that the production of semiconductor chips was no longer any fun after a couple of years. I think I don't like that clean room work very much. And then in 1998, for example, new work was published by Uri Sivan³⁷ who proposed for the first time to use radically new methods of production based on the principle of molecular self-organization and biological material. And that fascinated me somehow and I thought, if I want to stay in this field at all, then I want to work in this biological self-organization direction.³⁸

³⁶I distinguish interdisciplinarity from transdisciplinarity. While transdisciplinarity implies the transcending of disciplinary boundaries, interdisciplinarity involves a stronger form of cooperation, where scientists from different disciplines work together on the basis that each partner has to learn the pre-supposition of the other's discipline to come up with a new project, idea or technological device (Thompson Klein 2001; Schummer 2004, 11; Kehrt and Schübler 2010, 38).

³⁷In 1998, Sivan and his colleagues from the University of Haifa used DNA as a template to attach a silver wire to construct an electric circuit (Erez Braun et al. 1998).

³⁸"Simmel: Und da muss ich aber sagen, dass mir ehrlich gesagt diese ganze Produktion der Halbleiterchips nach ein paar Jahren keinen Spaß mehr gemacht hat. Ich mag diese Reinraumarbeit nicht besonders, glaube ich. Und da kamen dann im Jahr 98 Arbeiten raus von Uri Sivan z.B., wo die Idee vorgebracht wurde, dass man vielleicht ganz neue Produktionsmethoden nutzen könnte, die auf Selbstorganisation und biologischem Material basieren. Und das hat mich irgendwie fasziniert

Simmel, who holds a chair of bioelectronics at TUM, no longer works with quantum dots and computer chips, but with DNA and methods from the life sciences. That is a radical step beyond his original field of research. He does not operate in the clean room any more. His laboratory, which moved into the new nanoscience building at the WSI looks more like a biotechnology lab. The aim is to find new ways of handling and using DNA as a building block for future “DNA machines,” DNA computers as a template for materials synthesis or intelligent drug delivery systems. For Simmel, DNA is not just a carrier of information and a basic unit of life that scientists try to understand; it also has interesting physical, electrical and mechanical properties, it is something to “play around” with and to see how artificial molecular machines behave with their abilities to host other molecules or to act as semiconductors.

Biochemist Nadrian Seeman has influenced this nanobiotechnological research field since the 1980s (Seeman 1999, 11; 2002, 53–84; 2003, 33–37). He is interested in the functional properties of DNA to create radically new systems and DNA structures, so-called Nano-Origami, with potential technological applications:

For the past half-century, DNA has been almost exclusively the province of biologists and biologically-oriented physical scientists, who have studied its biological impact and molecular properties. During the next 50 years, it is likely they will be joined by materials scientists, nanotechnologists, and computer engineers, who will exploit DNA’s chemical properties in a non-biological context. (Seeman 2003, 431)

But experts doubt that DNA will ever be able to compete directly with silicon-based technology. Therefore, such far-reaching technological visions of DNA as a building block for future computers has no direct meaning for technology development and is more a question of basic research practiced by university-based scientists. As Simmel points out, these questions are rather basic and a DNA computer is not realistic so far:

And now they want to bring these two worlds together. That is incredibly difficult in a technological sense, and maybe even unrealistic. So we have to ask in what direction that should go. On the other hand, I still think that for some kinds of things this is useful, if you want to solve some basic questions. But basically when you say you want to combine semiconductor technology and biotechnology

und ich habe mir gedacht, also wenn ich überhaupt in dem Feld bleibe, dann möchte ich in diese Bioselbstorganisations-Richtung.” (Interview with Simmel, 30 September 2008).

for example with a lab on a chip or biosensors or such things. I see perspectives there, because you are interested exactly in the interface of the two worlds, so to speak. But if the question is, whether it is realistic to build a Pentium Processor out of DNA, then I would say this isn't realistic.³⁹

In the early 1990s, researchers hoped that DNA would one day replace silicon as the basis for a new generation of computers, "scientists have realized that there are numerous problems inherent in DNA computing and that they would have to live with their silicon-based computers for quite a while yet" (Parker 2003, 7).

Despite these new experimental practices that combine new methods from the life sciences with approaches and research questions from experimental physics, there are only few signs of close interdisciplinary cooperation. In most of the cases, scientists from life science departments are not really interested in what their physics colleagues try to do with DNA:

Simmel: I think that the influence of biophysics was very important in Munich because biophysics is interdisciplinary in its roots. And that was also an important influence in CeNS and then later NIM concerning research topics that were chosen. Because ultimately biophysics works at the border to biochemistry. But in contrast, there were almost no direct influence from biochemistry or biology on the nano-developments here in Munich, as far as I can see.

Kehrt: So strongly oriented towards physics?

Simmel: Yes.

Kehrt: Physics is opening up, while chemistry remains within its classical structures?

Simmel: Exactly. Here with CeNS and NIM there is almost no participation with chemistry and almost none with biology [...]. Sometimes they [the biochemists] say we are really dealing with the important biological questions while what you are doing is simply playing around. So in the end, in their view, what I do is of

³⁹"Simmel: Und jetzt will man diese zwei Welten zusammen bringen. Das ist an manchen Punkten einfach technisch wahnsinnig schwierig und vielleicht auch unrealistisch. Also da muss man sich fragen, in welche Richtung das gehen soll. Umgekehrt glaube ich aber schon, dass man es für manche Dinge brauchen kann, zumindest als einerseits um Grundlagenfragen zu beantworten. Aber eben dann wenn man sagt, die Verknüpfung aus Halbleitertechnologie und DNA oder Biotechnologie findet meinetwegen lab-on-a-chip oder im Biosensorikbereich oder solche Sachen. Da sehe ich durchaus Perspektiven, weil da ist man ja genau an diesem Interface sozusagen interessiert zwischen den beiden Welten. Wenn es aber jetzt darum geht, ist es realistisch, mit DNA einen Pentium-Prozessor zu bauen, dann würde ich sagen, dass es nicht realistisch ist." (Interview with Simmel, 30 September 2008).

course purely playing around. In this sense nanoscience in Munich remained within physics, perhaps because in physics it is more easily accepted that scientists play around without any clear goals. But I have to say that is different in the US. There people like me almost always work in interdisciplinary centers with a strong participation of biochemistry and chemistry, which is quite remarkable.⁴⁰

Only a few experimental physicists like Simmel adopt methods from biochemistry and leave their discipline far behind without really closely cooperating with their neighboring disciplines from the life sciences. There is no direct cooperation or interdisciplinary exchange with scientists from the life sciences. Doctoral students from the life sciences also hesitate to work in physics departments because of their strict career patterns. So if we want to identify the trading zones between physics and the life sciences, it is the laboratories of experimental physicists like Simmel in which knowledge is transferred from the life sciences in order to use DNA as an experimental system to build artificial devices and lay the foundations of future DNA computing. This is one of the rather seldom cases in the history of physics where physicists adopt and incorporate approaches from other disciplines (Kragh 1999, 445). In this instance, the boundary object of nanotechnology facilitates knowledge transfer and the sometimes radically new methods beyond disciplinary boundaries that obviously would have been difficult to pursue within the framework of semiconductor physics.

⁴⁰“Simmel: Ich glaube, dass in München der Einfluss der Biophysik sehr wichtig war, weil die Biophysik von der Anlage her interdisziplinär ist und das war ja auch bei CeNS und dann später NIM ein sehr wichtiger Einfluss bei Themen, die dann letztlich auch gewählt wurden. Weil die Biophysik zwangsläufig an der Grenze zur Biochemie arbeitet. Im Gegensatz dazu ist aber aus der Biochemie selber oder auch der Biologie kaum ein Einfluss auf die ganze Nano-Entwicklung hier in München gewesen, soweit ich das sehen kann.

Kehrt: Stark physikorientiert?

Simmel: Ja.

Kehrt: Die Physik öffnet sich, während die Chemie in ihren klassischen Strukturen drin bleibt?

Simmel: Genau. Also man hat auch im NIM und im CeNS und was es da alles gibt fast keine Beteiligung von Seiten der Chemie und so gut wie gar keine von der Biologie. [...] Manchmal bekommt man auch mitgeteilt, mehr oder weniger, wir kümmern uns um die wirklich wichtigen biologischen Fragestellungen und das andere ist halt Spielerei. Also letztlich, das was ich mache ist auch, aus deren Sicht natürlich, pure Spielerei. Insofern ist es gerade hier in München relativ physiklastig geblieben, vielleicht weil das eben in der Physik eher akzeptiert wird, dass man so ein bisschen rumspielt, ohne ganz klare Zielrichtung. Ich muss aber sagen, das es im Gegensatz dazu in den USA anders ist. Also da ist meine Konkurrenz fast immer in interdisziplinären Zentren, in denen die Biochemie und die Chemie sehr stark beteiligt ist, was ganz kurios ist.” (Interview with Simmel, 30 September 2008).

13.7 Conclusion: The Reinvention of Semiconductor Physics

The Munich case study shows that nanotechnology is deeply embedded in the history of semiconductor physics (Choi and Mody 2009; McCray 2007). The story of experimental physicists dealing with the confinement of electrons in two, one and zero dimensional systems—so-called quantum wells, wires and dots—started in the early 1970s with new instrumental practices at the quantum level. This can be shown by looking at the careers of that generation of physicists who finished their doctorates in the 1970s, were of political interest in the 1980s chip war, and then reoriented their research efforts in the direction of nanotechnology at the end of the 1990s. The relabeling of semiconductor physics' research traditions was mainly stimulated by science policy and motivated by extra-scientific interests, such as the necessity to present research in the media, emphasize its economic potential and receive funding from partners outside academia. Therefore, nanotechnology is more a rhetorical tool and ideologically motivated science policy strategy which has emerged to cope with new challenges that university-based research had to face at the turn of the twenty-first century. Scientists have to legitimate their research by referring to the potential utility and innovations that might result from that research without necessarily being directly involved in innovation processes. In this sense, the Munich case study confirms Forman's thesis that there is a primacy of technology in so-called postmodernity. But the story of Munich nanotechnology networks differs from Forman's diagnosis, which dates the changes and shifts toward postmodern science to the 1980s. While the research practices of this field started in the early 1970s, the new and explicit orientation toward nanotechnology appeared in the 1990s—exactly at a point when that field lost its crucial contact to the semiconductor industry. At that time, new developments within the life sciences stimulated new approaches in experimental physics. The boundary object of nanotechnology helped physicists leave the traditions of semiconductor physics behind and adopt new methods and experimental systems from the life sciences. Therefore, nanotechnology—with its strong rhetoric of innovation, its dizzying transgressions and redefinitions of existing institutional frameworks, and its presence in the public sphere—is rather typical for science at the end of the 20th century.

References

- Angel, David P. (1994). *Restructuring for Innovation. The Remaking of the U.S. Semiconductor Industry*. New York and London: The Guildford Press.
- Arthur, John (2002). Molecular Beam Epitaxy. *Surface Science* 500: 189–217.

- Bachmann, Gerd (1998). *Innovationsschub aus dem Nanokosmos – Technologieanalyse*. Düsseldorf: Bericht des VDI-Technologiezentrum, Abteilung Zukünftige Technologie des Vereins Deutscher Ingenieure (VDI).
- Baird, Davies, Alfred Nordmann, and Joachim Schummer, eds. (2004). *Discovering the Nanoscale*. Amsterdam: IOS Press.
- Ball, Philipp and Laura Garwin (1992). Science at the Atomic Scale. *Nature* 355(6363): 761–766.
- Bassett, Ross Knox (2002). *To the Digital Age. Research Labs, Start Up Companies and the Rise of MOS Technology*. Baltimore: Johns Hopkins University.
- Böhme, Gernot, ed. (1978). *Die gesellschaftliche Orientierung des wissenschaftlichen Fortschritts*. Frankfurt am Main: Suhrkamp.
- Braun, Erez, Yoav Eichen, Uri Sivan, and Gdalyahu Ben-Yoseph (1998). DNA-Templated Assembly and Electrode Attachment of a Conducting Silver Wire. *Nature* 391: 775–778.
- Braun, Ernest and Stuart MacDonald (1978). *Revolution in Miniature. The History and Impact of Semiconductor Electronics*. New York: Cambridge University Press.
- Brus, Louis and Don Eigler (1994). Devices and Desires. *Nature* 369: 273–274.
- Bundesbericht Forschung 1996. Drucksache 13/4554 vom 08.05.1996. <http://dip21.bundestag.de/dip21/btd/13/045/1304554.asc> (accessed 10/06/2011).
- Capasso, Frederico (1987). Band-Gap Engineering. From Physics and Materials to New Semiconductor Devices. *Science* 235(4785): 172–176.
- ed. (1990). *Physics of Quantum Electron Devices*. Berlin: Springer.
- Carson, Cathryn, Alexei Kojevnikov, and Helmuth Trischler (2008). Beyond “Weimar Culture” – Die Bedeutung der “Forman Thesen” für eine Wissenschaftsgeschichte in kulturhistorischer Perspektive. *Berichte zur Wissenschaftsgeschichte* 31: 301–306.
- Caspar, Steven (2007). *Creating Silicon Valley in Europe. Public Policy towards New Technology Industries*. Oxford: Oxford University Press.
- Cho, Alfred Y. (1999). How Molecular Beam Epitaxy (MBE) Began and Its Projection into the Future. *Journal of Crystal Growth* 201/202: 1–9.
- (2004). Recollections about the Early Developments of Molecular Beam Epitaxy (MBE). In: *50 Years Progress in Crystal Growth*. Ed. Robert Feigelson. A Reprint Collection: Elsevier. 199–201.
- Choi, Hungsoi and Cyrus Mody (2009). The Long History of Molecular Electronics. The Microelectronics Origins of Nanotechnology. *Social Studies of Science* 39(1): 11–50.
- Corcoran, Elisabeth (1991). Nanotechnik. *Spektrum der Wissenschaften* 91(1): 76–86.

- Cuhls, Kerstin, Christian Uhlborn, and Hariolf Grupp (1996). Foresight in Science and Technology—Future Challenges of the German S&T System. In: *Science, Technology, and Innovation in Germany—Changes and Challenges in the 1990s*. Ed. Wilhelm Krull and Frieder Meyer-Krahmer. London: Cartermill. 63–81.
- Decker, Michael (2006). Eine Definition von Nanotechnologie: Erster Schritt für ein interdisziplinäres Nanotechnology Assessment. In: *Nanotechnologien im Kontext. Philosophische, ethische und gesellschaftliche Perspektiven*. Ed. Alfred Nordmann, Joachim Schummer, and Astrid Schwarz. Berlin: Akademische Verlagsgesellschaft. 33–48.
- Dingle, Raymond (1975). Confined Carrier Quantum States in Ultrathin Semiconductor Heterostructures. *Festkörperprobleme* 15: 21–48.
- Doering-Manteuffel, Anselm and Lutz Raphael (2010). *Das Ende der Zuversicht. Perspektiven auf die Zeitgeschichte seit 1970*. II. Göttingen: Vandenhoeck & Ruprecht.
- Doezema, Ryan E. and Frederick J. Koch (1972). Magnetic Surface Levels in Cu—Observation and Analysis of Microwave Resonances and Determination of Fermi Velocities. *Physical Review B* 5: 3866–3882.
- Dorda, Gerhard (1971). Piezoresistance in Quantized Conduction Bands in Silicon Inversion Layers. *Journal of Applied Physics* 42(5): 2053.
- Eckert, Michael (1988). Neutrons and Politics: Maier-Leibnitz and the Emergence of Pile Neutron Research in the FRG. *Historical Studies in the Physical and Biological Sciences* 19(1): 81–113.
- Eckert, Michael and Maria Osietzki (1989). *Wissenschaft für Macht und Markt. Kernforschung und Mikroelektronik in der Bundesrepublik Deutschland*. München: Beck.
- Eckert, Michael and Helmut Schubert (1986). *Kristalle, Elektronen, Transistoren. Von der Gelehrtenstube zur Industrieforschung*. Hamburg: Deutsches Museum.
- Esaki, Leo (1974). Long Journey into Tunneling (Nobel Lecture). *Science* 183(413): 1149–1155. http://nobelprize.org/nobel_prizes/physics/laureates/1973/esaki-lecture.pdf (accessed 10/06/2011).
- (1985). Semiconductor Superlattice and Quantum Wells through Development of Molecular Beam Epitaxy. In: *Molecular Beam Epitaxy and Heterostructures*. Ed. Klaus Ploog and Leroy L. Chang. Dordrecht, Boston, Lancaster: Martinus Nijhoff Publishers. 2–26.
- (1992). Do-It-Yourself Quantum Mechanics in Low-Dimensional Structures. *Physica Scripta* T42: 102.

- Esaki, Leo and Ray Tsu (1979). Superlattice and Negative Differential Conductivity in Semiconductors. *IBM Journal of Research and Development* 14(1): 61–65.
- Forman, Paul (2007). The Primacy of Science in Modernity, of Technology in Postmodernity, and of Ideology in the History of Technology. *History and Technology* 23(1): 1–152.
- (2010). (Re)cognizing Postmodernity: Helps for Historians—of Science Especially. *Berichte zur Wissenschaftsgeschichte* 33: 157–175.
- Fowler, Alan B., Frank F. Fang, William E. Howard, and Philip J. Stiles (1966). Magneto-Oscillatory Conductance in Silicon Surfaces. *Physical Review Letters* 16(20): 901–903.
- Gerybadze, Alexander, Frieder Meyer-Krahmer, and Guido Reger, eds. (1997). *Globales Management von Forschung und Innovation*. Stuttgart: Schäffer-Poeschel.
- Gibbons, Michael, Camille Limoges, and Helga Nowotny (1994). *The New Production of Knowledge*. London: Sage.
- Gieryn, Thomas F. (1999). *Cultural Boundaries of Science. Credibility on the Line*. Chicago: The University of Chicago Press.
- Götzberger, Adolf and Max Schulz (1973). Fundamentals of MOS Technology. *Festkörperprobleme* 13: 309–336.
- Günther, Karl-Georg (1958). Evaporated Layers of Semiconducting III–V Compounds. *Naturwissenschaften* 45: 415.
- Hack, Lothar (1998). *Technologietransfer und Wissenstransformation. Zur Globalisierung der Forschungsorganisation von Siemens*. Münster: Westfälisches Dampfboot.
- Handel, Kai (1999). *Anfänge der Halbleiterforschung und -entwicklung. Dargestellt an den Biographien von vier deutschen Halbleiterpionieren*. PhD thesis. Aachen: RWTH.
- Heinze, Thomas (2006). *Die Kopplung von Wissenschaft und Wirtschaft. Das Beispiel der Nanotechnologie*. Frankfurt am Main: Campus.
- Hennig, Jochen (2011). *Bildpraxis. Visuelle Strategien in der frühen Nanotechnologie*. Bielefeld: Transcript.
- Hermann, Hans, Heinz Herzer, and Erhard Sirtl (1975). Modern Silicon Technology. *Festkörperprobleme* 15: 279–316.
- Hermann, Wolfgang, ed. (2005). *Unternehmen Universität – Universität Unternehmen. Deutschland im Paradigmenwechsel des Hochschulsystems*. München: Technische Universität.
- ed. (2006). *Technische Universität München. Die Geschichte eines Wissenschaftsunternehmens*. Berlin: Metropol.

- Heßler, Martina (2007). *Die kreative Stadt. Zur Neuerfindung eines Topos*. Bielefeld: Transcript (Urban Studies).
- Jaraus, Konrad H., ed. (2008). *Das Ende der Zuversicht? Die siebziger Jahre als Geschichte*. Göttingen: Vandenhoeck & Ruprecht.
- Joerges, Bernward and Terry Shinn (2001). *Research Technologies. Instrumentation between Science, State and Industry*. Dordrecht: Kluwer Academic Publishers.
- Johnson, Ann (2004). The End of Pure Science: Science Policy from Bayh-Dole to the NNI. In: *Nanoscale*. Ed. Davis Baird, Alfred Nordmann, and Joachim Schummer. 217–230.
- Junk, Andreas and Falk Riess (2006). From an Idea to a Vision: There's Plenty of Room at the Bottom. *American Journal of Physics* 74(9): 825–830.
- Kehrt, Christian (2011). Mit Molekülen spielen. Die Nanotechnologie als politische Strategie der universitären Grundlagenforschung. In: *Neue Technologien in der Gesellschaft: Akteure, Erwartungen, Kontroversen und Konjunkturen*. Ed. Christian Kehrt, Peter Schübler, and Marc-Dennis Weitze. Bielefeld: Transcript. 315–332.
- Kehrt, Christian and Peter Schübler (2010). Nanoscience is 100 Years Old. The Defensive Appropriation of the Nanotechnology Discourse within the Disciplinary Boundaries of Crystallography. In: *Governing Future Technologies. Nanotechnology and the Rise of an Assessment Regime*. Ed. Mario Kaiser, Monika Kurath, Sabine Maasen, and Christoph Rehmann-Sutter. Dordrecht, Heidelberg, London and New York: Springer. 37–54.
- Kelly, Michael J. (1987). Nanometer Physics and Microelectronics. In: *Solid State Science, Past, Present and Predicted*. Ed. Denis L. Weaire and Colin G. Windsor. Bristol: Adam Hill Publishing. 263–273.
- von Klitzing, Klaus (1985). *The Quantized Hall Effect. Nobel lecture*. http://nobelprize.org/nobel_prizes/physics/laureates/1985/Klitzing-lecture.pdf (accessed 10/06/2011).
- (1995). *Grenzen der Mikroelektronik: Quantenphänomene in mikrostrukturierten Halbleitern*. Jena: Universitätsverlag (Schriftenreihe Ernst-Abbe-Kolloquium Jena 11).
- (1997). Physics and Applications of the Quantum Hall Effect. In: *Physics and Applications of Quantum Wells and Superlattices*. Ed. Emilio E. Mendez and Klaus von Klitzing. 170. New York and London: NATO ASI Series, Series B, Physics. 229–248.
- von Klitzing, Klaus, Gerhard Dorda, and Michael Pepper (1980). New Method for High Accuracy Determination of the Fine Structure Constant Based on Quantized Hall Resistance. *Physical Review Letters* 45(6): 494–497.

- Koch, Frederick J. (1975). The Dynamics of Conduction Electrons in Surface Space Charge Layers. *Festkörperprobleme* 15: 79–112.
- Kostoff, Ronald N., Raymond G. Koytcheff, and Clifford G. Y. Lau (2007). Global Nanotechnology Research Metrics. *Scientometrics* 70(3): 565–601.
- Kotthaus, Jörg, Gerhard Abstreiter, Frederick J. Koch, and Ronald Ranvaud (1975). Cyclotron Resonance of Localized Electrons on a Si Surface. *Physical Review Letters* 34: 151–154.
- Kragh, Helge (1999). *Quantum Generations. A History of Physics in the Twentieth Century*. Princeton: Princeton University Press.
- Kroemer, Herbert (2005). Nano-Whatever. Do We Really Know Where We Are Heading? *Physica Status Solidi (a)* 202(6): 957–964.
- Kuchar, Friedemar, Helmut Heinrich, and Günther Bauer, eds. (1990). *Localization and Confinement of Electrons in Semiconductors*. Proceedings of the Sixth International Winter School, Mauterndorf, Austria, February 19–23. Berlin: Springer Series in Solid State Sciences.
- Landwehr, Gottfried (1975). Quantum Transport in Silicon Inversion Layers. *Festkörperprobleme* 15: 49–77.
- (1986). The Discovery of the Quantum Hall Effect. *Festkörperprobleme* 26: 17–39.
- (2003). 25 Years Quantum Hall Effect. How It All Came About. *Physica E* 20: 1–13.
- Latour, Bruno (1998). *Wir sind nie modern gewesen. Versuch einer symmetrischen Anthropologie*. Frankfurt am Main: Suhrkamp.
- McCray, Patrick W. (2005). Will Small Be Beautiful? Making Policies for Our Nanotechnology Future. *History and Technology* 21(2): 177–203.
- (2007). MBE Deserves a Place in the History Books. *Nature Nanotechnology* 2(5): 2–4.
- Meyer, Martin and Olle Person (1998). Nanotechnology—Interdisciplinarity, Patterns of Collaboration and Differences in Application. *Scientometrics* 42(2): 195–205.
- Mody, Cyrus (2004). *Crafting the Tools of Knowledge. The Invention, Spread and Commercialization of Probe Microscopy, 1960–2000*. PhD thesis. Cornell University.
- Mönch, Winfried (1973). On the Physics of Clean Silicon Surfaces. *Festkörperprobleme* 13: 241–274.
- Morris, Peter R. (1990). *History of the World. Semiconductor Industry*. London: Peter Peregrinus Ltd.
- Nordmann, Alfred (2004). Was ist Technowissenschaft? – Zum Wandel der Wissenschaftskultur am Beispiel von Nanoforschung und Bionik. In: *Bionik. Aktuelle Forschungsergebnisse in Natur-, Ingenieur- und Geisteswissenschaften*

- ten. Ed. Thorsten Rossmann and Cameron Tropea. Berlin: Springer. 209–218.
- (2006). Denkmuster hinter der Nanotechnologie. Die Welt als Baukastensystem. *Politische Ökologie* 101: 20–23.
- Nusser, Michael (2006). Wissenschaftliche Bedeutung und Wettbewerbsfähigkeit forschungs- und wissensintensiver Branchen. Innovationsreport. Handlungsoptionen zur Sicherung der internationalen Wettbewerbsfähigkeit. *TAB-Brief* 30: 65–67.
- Parker, Jack (2003). Computing with DNA. *EMBO [European Molecular Biology Organization]* 4(1): 7–10.
- Pethica, John Bernard (1988). Atomic-Scale Engineering. *Nature* 331: 301.
- Ploog, Klaus (1978). Molekularstrahlepitaxie – Grundlagen und Anwendung für die Bauelementherstellung. *Nachrichtentechnische Zeitschrift* 31(6): 435–441.
- Ploog, Klaus and Leroy L. Chang, eds. (1985). *Molecular Beam Epitaxy and Heterostructures*. Dordrecht, Boston, Lancaster: Martinus Nijhoff Publishers.
- Reed, Marc (1993). Quantum Dots. *Scientific American* 268(1): 118–123.
- Rheinberger, Hans-Jörg (2006). Rezente Wissenschaft und ihre Erforschung. Das Beispiel Molekularbiologie. *Medizinhistorisches Journal* 41: 187–199.
- Schummer, Joachim (2004). Multidisciplinarity, Interdisciplinarity, and Patterns of Research Collaboration in Nanoscience and Nanotechnology. *Scientometrics* 59(3): 425–465.
- (2009). *Nanotechnologie. Spiele mit Grenzen*. Edition Unseld Nr. 23. Frankfurt am Main: Suhrkamp.
- Science. Special Section (1991). Engineering a Small World: From Atomic Manipulation to Microfabrication. *Science* 254: 1300–1335.
- Seeman, Nadrian C. (1999). DNA Nanotechnology. *Nature Biotechnology* 17: 11.
- (2002). It Started with Watson and Crick, but It Sure Didn't End There. Pitfalls and Possibilities beyond the Classic Double Helix. *Natural Computing* 1: 53–84.
- (2003). DNA in a Material World. *Nature* 421: 33–37.
- Simmel, Friedrich C. (1999). *Transporteigenschaften von Quantenpunkten: Statistik der Leitwertresonanzen im Coulomb-Blockade-Regime; nicht-linearer Transport durch einen kleinen Quantenpunkt; Transport durch Quantenpunkte unter dem Einfluß zeitveränderlicher Felder*. PhD thesis. Berlin, Logos Verlag: LMU München.
- Simmel, Friedrich C. and Wendy U. Dittmer (2005). DNA Nanodevices. *Small* 3(1): 284–299.
- Smith, T. B. (1990). Electron Confinement in Quantum Dots. In: *Localization and Confinement of Electrons in Semiconductors*. Ed. Friedemar Kuchar, Helmut

- Heinrich, and Günther Bauer. Proceedings of the Sixth International Winter School, Mauterndorf, Austria, February 19–23. Berlin: Springer Series in Solid State Sciences. 10–19.
- Söderqvist, Thomas, ed. (1997). *The Historiography of Contemporary Science and Technology*. Amsterdam: Harwood Academic Publisher.
- Star, Susan Leigh and James R. Griesemer (1989). Institutional Ecology. “Translations” and Boundary Objects: Amateurs and Professionals in Berkeley’s Museum of Vertebrate Zoology 1907–1939. *Social Studies of Science* 19: 387–420.
- Thompson Klein, Julie (1992). *Crossing Boundaries, Knowledge, Disciplinarity and Interdisciplinarity*. Charlottesville and London: University Press of Virginia.
- (2001). *Transdisciplinarity. History, Theory, and Practice*. Detroit: Wayne University Press.
- Thouless, David J. (1984). Theory of Quantized Hall Effect. *Surface Science* 142: 147–154.
- Trischler, Helmuth (1999). Hoffnungsträger oder Sorgenkind der Forschungspolitik? Die bundesdeutsche Großforschung in den “langen” siebziger Jahren. In: *Wissenschaftsfördernde Institutionen im Deutschland des 20. Jahrhunderts*. Ed. Rüdiger vom Bruch and Eckart Henning. Dahlemer Archivgespräche 5. Berlin: MPG-Archiv. 200–214.
- (2001). Das bundesdeutsche Innovationssystem in den “langen 70er Jahren”: Antworten auf die “amerikanische Herausforderung”. In: *Innovationskulturen und Fortschrittserwartungen im geteilten Deutschland*. Ed. Johannes Abele, Gerhard Barkleit, and Thomas Hänseroth. Köln: Böhlau. 47–70.
- Tsay, Ming-Yueh, Shio-Jen Jou, and Sheau-Shin Ma (2000). A Bibliometric Study of Semiconductor Literature, 1978–1997. *Scientometrics* 49(3): 491–509.
- Yurke, Bernard, Andrew J. Turberfield, Allen P. Mills, Friedrich C. Simmel, and Jennifer Neumann (2000). A DNA-Fuelled Molecular Machine Made of DNA. *Nature* 406: 605–608.