Chemistry and Lithography

Vol. I: The Chemical History of Lithography

SECOND EDITION

Chemistry and Lithography

Vol. I: The Chemical History of Lithography

SECOND EDITION

Uzodinma Okoroanyanwu

SPIE PRESS
Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Names: Okoroanyanwu, Uzodinma, author. | Society of Photo-optical Instrumentation Engineers.
Title: Chemistry and lithography / Uzodinma Okoroanyanwu.
Description: Second edition. | Bellingham, Washington, USA : SPIE Press, 2020- | Includes bibliographical references and index. | Contents: volume. 1. The chemical history of lithography
Identifiers: LCCN 2019037747 | ISBN 9781510631557 (v. 1 ; paperback) | ISBN 9781510631564 (v. 1 ; pdf) | ISBN 9781510631571 (v. 1 ; epub) | ISBN 9781510631588 (v. 1 ; kindle edition)
Subjects: LCSH: Lithography. | Chemistry, Technical. | Semiconductors–Etching.
Classification: LCC TP156.E68 O44 2020 | DDC 686.2/315–dc23
LC record available at https://lccn.loc.gov/2019037747

Published by

SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1 360.676.3290
Fax: +1 360.647.1445
Email: books@spie.org
Web: http://spie.org

Printed in the United States of America.
First Printing.
For updates to this book, visit http://spie.org and type “PM313” in the search field.

This work as a whole is dedicated to the memory of my academic mentor
Late Professor WILLIAM C. GARDINER, JR.

This volume is dedicated to the memory of my grandmother
Late Mrs. VICTORIA ONYAH EDIMA

Contents

General Preface to the Second Edition

A second edition of a text has a number of advantages over the first edition: in addition to representing the accumulated wisdom of the readers who have suggested ideas for improving the previous edition, it also incorporates developments in the field that have taken place since the publication of the first edition. From their familiarity with the previous edition and in keeping with the general aim to make the text accessible, readers will be aware that I have provided not a slight tinkering with the text in this edition, but a wholesale rewriting of some chapters, while introducing entirely many new ones.

This present edition is a series of three volumes, namely, *The Chemical History of Lithography* (volume 1), *Chemistry in Lithography* (volume 2), and *The Practice of Lithography* (volume 3). Each volume is a unit in itself and can stand alone, which is fortunate in view of their different subject matters.

My motivation for splitting the second edition of *Chemistry and Lithography* into the above three volumes is because the subject matter of these three books is very broad and lends itself nicely to multiple angles of analysis and interpretation. Lithographically fabricated structures appear in an increasing wide range of scientific and technical fields, beyond their traditional niches in fine arts, paper printing, and electronics. In particular, the last ten years since the publication of the first edition of *Chemistry and Lithography* have witnessed a phenomenal pace of development in advanced lithography, perhaps the fastest pace of development in any decade over the 222-year history of the field. The transistor count of the leading-edge integrated circuit when the first edition was published was around 2.5 billion, and a short five years later that count rose to 10 billion. By the time the present edition is published, the transistor count of an equivalent device may well be over 20 billion, representing a nearly ten-fold increase within a decade in the ability to process, communicate, and store electronic information, made possible in no small part by advances in lithography. Such is the pace of the development in this field that the analysis I make in these three volumes is a sufficiently different and significantly expanded version of the treatment of the subject I rendered in the first edition. As such, there are a considerable number of changes within the text of the present edition.

The first edition covered some brief historical material in Part I, and in volume 1 of the present edition these historical aspects are extended. It seems reasonable to me that with the rapid advances being made in advanced lithography it is very important to realize that some of the concepts that we now take for granted in chemistry and physics that enervate lithography are by no means self-evident and often were developed only after much struggle and controversy. Improvements are still being made in all of these fields and can all the more be clearly understood by those who appreciate how our present understanding of the fields were attained.

In developing volume 1 of the present edition, I drew considerably from Part I of the first edition. Relative to Part I of the first edition, volume 1 of the present edition comprises five significantly expanded chapters: Chapters 1 and 2 of first edition with slight tinkering; an expanded version of Chapter 4 of the first edition; and two new chapters derived from expanding and splitting up Chapter 3 of the first edition. The object of volume 1 of the present edition is to weave together threads of a narrative on the history of optical and molecular physics, optical technology, chemistry, and lithography, with a view to creating a rich tapestry that gives the reader new insights into an aspect of the relationships between these fields that are often not fully appreciated: how the marriage between chemistry and optics led to the development and evolution of lithography. I show how major developments in the chemistry, physics, and technology of light influenced the invention and development of lithography, well beyond what its inventor envisioned. I also show how developments in lithography have not only influenced the development of optics and chemistry, but also played a critical role in the large-scale manufacture of integrated circuits that run the computers and machineries on which our modern electronic and information age depend.

Part of the analysis in volume 1 of the present edition is necessarily skewed toward the underlying science and technologies of advanced lithographic patterning techniques, in terms of materials, processes, and imaging, along with their unique features, strengths, and limitations. This book also provides an analysis of the emerging trends in lithographic patterning, as well as the current and potential applications of the resulting patterned structures and surfaces.

The object of volume 2 of the present edition is to deconstruct lithography into its essential chemical principles and to situate its various aspects in specific fields of chemistry. It comprises 16 chapters developed around a rewriting of Chapters 5 through 8 of Part II of the first edition and parts of Chapters 13 and 14 of Part III of the first edition. It also includes eight entirely new chapters that explore in a fundamental manner the role of chemistry in mediating specific aspects of the lithographic process.

Volume 2 of the present edition is in fact an outgrowth of the SPIE Advanced Lithography Short course Chemistry and Lithography (SC1099) that I have taught at the SPIE Advanced Lithography Symposium over the past six years. In this volume, I develop a chemistry and lithography interaction matrix, and use it as a device to illustrate how various aspects and practices of advanced lithography derive from established principles and phenomena of chemistry. For instance, lithographic unit operations involving principally the resist fall within the realm of process chemistry. Photochemistry is involved in the generation of photons from the exposure sources of optical, extreme ultraviolet, and x-ray lithographic exposure tools; the interaction of O_2, H_2O, NH_3, SO_x, and hydrocarbons with photons within the optical lithographic tool exposure chamber, as well as their roles in the oxidation, carbon deposition and growth, and formation of inorganic salt crystals (also called haze crystals) on masks and optical elements; cleaning of contaminated optics and masks with UV photons; clear defect repair of masks via photo-induced decomposition of organometallic precursors and deposition of metals on defective areas of the mask; and UV curing and photo-oxidative degradations of resists during exposure. Photochemistry is also the basis of the exposure action of photoacids and photobases in resists, as well as resist poisoning by airborne molecular bases. It is also the basis of laser-produced and discharge produced plasmas—the radiation sources for extreme-ultraviolet (EUV) lithographic patterning. Finally, photochemistry is the basis of the plasma (dry) stripping of resists and other hydrocarbon contaminants from wafers and masks.

Similarly, radiation chemistry is involved in the generation of electrons and ions from the exposure sources of electron-beam and ion-beam lithographic exposure tools, respectively; cleaning of contaminated optics and masks, and repair of defective masks with electrons and ions; the electron- and ion-mediated exposure process as well as electron-beam and ion-beam curing, and crosslinking of resists. The manufacture of lithographic exposure sources, optical elements, masks, and resist materials involves materials chemistry. Polymer chemistry is the basis of the synthesis of polymeric resist resins used in optical, charged particle, and imprint lithography, as well as block copolymers used in directed block copolymer self-assembly lithography.

Surface chemistry is the basis of priming of relevant mask and wafer substrates, and their subsequent coating with resists and associated layers used in the fabrication of masks and semiconductor device wafers. Colloid chemistry is the basis of the cleaning of masks and wafers, and development process of exposed resists. It also explains the basis for the stability of resist, developer, and wet cleaning solutions. Electrochemistry is the basis of the corrosion and electromigration processes of Cr and Mo mask or reticle absorber features, as well as electrostatic damage of the same objects. It is also the basis of electrochemical imprint lithography. Organometallic chemistry is the basis for the precursor materials used in clear defect repair of masks, as well as EUV metal-oxide resists.

The Practice of Lithography (volume 3) comprises 12 chapters, made up of Chapters 9 through 17 of the first edition, in addition to three new chapters covering full treatment of EUV lithography, as well as imprint lithography, directed block copolymer self-assembly lithography, and proximal probe lithography. The object of this volume is to present how the more important lithographic patterning techniques are used to print images on appropriately prepared flat substrate surfaces using radiations as varied as photons, electrons, and ions, as well as mechanical force, thermodynamically driven directed self-assembly of block copolymers, and even electron tunneling phenomena. In particular, I cover photolithography (or optical lithography), electron-beam lithography, ion-beam lithography, EUV lithography, imprint lithography, directed block copolymer self-assembly lithography, and proximal probe lithography.

Immersion ArF laser optical lithography is currently used in high-volume manufacture of integrated circuits at 22-, 15-, and 10-nm nodes, using double-patterning techniques to decrease feature pitch, where appropriate. EUV lithography is now entering high-volume production at the 7-nm node in some of the leading-edge semiconductor companies. Imprint lithography, especially in its roll-to-roll format, is increasingly being used in the fabrication of flexible and wearable electronic devices, diffractive optical elements, and large-area electronics. Directed block copolymer self-assembly is increasingly being used in the fabrication of functional nanostructures used in applications ranging from photonics to biomimetics, and from electrochemical energy storage to patterned electronic media. Now is an auspicious moment to provide an in-depth look into the chemistry that underpins these most advanced of lithographies.

Along with the introduction of EUV lithography into device manufacturing at single-digit-nanometer technology nodes, we enter a regime where the resist suffers from increased stochastic variation and the attendant effects of shot noise—a consequence of the discrete nature of photons, which, at very low number per exposure pixel, show increased variability in the response of the resist relative to its mean. Examples of resist response that may experience shot noise effects under very low-photon-count-per-pixel conditions within small exposure volumes, such as in EUV lithographic patterning, include photon absorption by the resist, and chemical conversion of light sensitive components in the resist, as well as the chemical changes that make a resist molecule soluble in the developer. We will examine in this volume of the present edition the role of stochastics in EUV lithography in far greater detail than we did in the first edition.

As in the first edition, I have made an attempt throughout the three books of the present edition to provide examples illustrating the diversity of chemical phenomena in lithography across the breadth of scientific spectrum, from fundamental research to technological applications. The format of this book is not necessarily chronological, but it is such that related aspects of lithography are thematically organized and presented with a view to conveying a unified view of the developments in the field over time, spanning many centuries, from the very first recorded reflections on the nature of matter to the latest developments currently at the frontiers of lithography science and technology. The emphasis is mostly placed on applications that have relevance to the semiconductor industry.

A great many of the pioneers of chemistry and lithography are not represented at all in the three books of the present edition. I can only record my immense debt to them and all who have contributed to the development of the two fields to the state in which I have reported it.

I am most grateful for suggestions from a number of experts, particularly the following: Andreas Erdman of Fraunhofer IISB, Manuel Thesen of micro resist technology GmbH, and Folarin Latinwo of Synopsys. Special thanks go to SPIE Senior Editor, Dara Burrows, for her editorial assistance in producing this book, which is much improved because of her efforts.

I am also grateful to my colleagues in the Department of Polymer Science and Engineering of University of Massachusetts at Amherst and, in particular, Prof. Jim Watkins, for lively scientific and technical discussions on polymers and flexible electronics. The opportunity to work in this department has not only helped me to broaden and deepen my scientific research interests, but also, in a direct way, has made it possible for me to write this book.

Lastly, I acknowledge the informal assistance I have received from my family members, in particular, from my wife Anett and daughter Sophie, who created a conducive atmosphere to work on this book at home.

Uzodinma Okoroanyanwu
Florence Village, Northampton, Massachusetts
January 2020

Acronyms and Abbreviations

AEE	aminoethoxy ethanol
AFM	atomic force microscope
AIBN	azobisisobutyronitrile
AMC	airborne molecular contaminant
APM	ammonia peroxide mixture
AR	antireflection
ARC	antireflection coating
att-PSM	attenuated phase-shifting mask
BARC	bottom antireflection coatings
BEOL	back end of line
BIM	binary intensity masks
BJT	bipolar junction transistor
BLR	bilayer resist
BOCST	butoxycarbonyloxystyrene
BPO	benzoyl peroxide
CaF_2	calcium fluoride
CAD	computer-aided design
CAR	chemically amplified resist
CARL	chemical amplification of resist lines
CBN	carbo-t-butoxy norbornene
CD	critical dimension
CFC	chlorofluorocarbon
CH	cyclohexanone
CMN	carbomethoxy norbornene
CMOS	complimentary metal-oxide semiconductor
CMP	chemical mechanical polishing
CMTF	critical modulation transfer function
CO	cycloolefin
COG	chrome-on-glass
COMA	cycloolefin-maleic anhydride
COP	crystal originated pit
CPU	central processing unit

CVD	chemical vapor deposition
DC	direct current
DEA	dissociative electron attachment
DEAP	diethoxyacetophenone
DMAc	dimethylacetamide
DMF	dimethylformamide
DMI	dimethyl-2-imidazolidinon
DMPA	dimethoxy phenylacetophenone
DMSDMA	dimethylsilyldimethylamine
DMSO	dimethylsulfoxide
DNQ	diazonaphthoquinone
DOF	depth of focus
DP	degree of polymerization
DPP	discharge-produced plasma
DRAM	dynamic random access memory
DRLS	development rate log slope
DRM	development rate monitor
DTBP	di-tert-butyl peroxide
DTBPIONf	di(tert-butylphenyl) iodonium perfluorobutanesulfonate (noaflate)
DUV	deep ultraviolet
EBES	electron-beam exposure system
EBL	electron-beam lithography
ECR	electron cyclotron resonance
EFM	electric-field-induced metal migration
EL	ethyl lactate
EOC	etalon output coupler
EOL	end of line
EPL	electron projection lithography
ESCAP	environmentally stable chemically amplified photoresist
ESD	electrostatic discharge
EUV	extreme ultraviolet
FEOL	front end of lithography
FET	field-effect transistor
FIB	focused ion beam
FRP	free radical polymerization
FTIR	Fourier transform infrared
FWHM	full width half maximum
HDPCVD	high-density plasma chemical vapor deposition
HEPA	high-efficiency particulate air
HF	hydrofluoric acid
HMDS	hexamethyldisilazane
HOL	hydrophilic over layer

HSQ	hydrogen silesquioxanes
HVM	high-volume manufacturing
IC	integrated circuit
IGFET	insulated gate field-effect transistor
ILD	interlayer dielectric
IMS	ion microfabrication system
IPL	ion projection lithography
ITRS	International Technology Roadmap for Semiconductors
JFET	junction field-effect transistor
KRS	ketal resist system
KTFR	Kodak thin-film resist
LBNL	Lawrence Berkeley National Laboratories
LEE	low-energy electron
LEEPL	low-energy electron projection lithography
LELE	lithography-etch-lithography-etch
LER	line edge roughness
LFLE	lithography-freeze-lithography-etch
LLD	lightly doped drain
LPP	laser-produced plasma
LPCVD	low-pressure chemical vapor deposition
L/S	line/space
LWR	line width roughness
Mac	methylacetamide
MEA	monoethanolamine
MEBES	multiple electron-beam exposure system
MEEF	mask error enhancement factor
MEMS	microelectromechanical system
MET	microexposure tool
MIBK	methylisobutyl ketone
MIF	metal-ion-free
ML	multilayer
MMA	methyl methacrylate
MOCVD	metal-organic chemical vapor deposition
MOP	methoxypropyl
MOS	metal-oxide semiconductor
MOSFET	metal-oxide semiconductor field-effect transistor
MTF	modulation transfer function
MW	molecular weight
NA	numerical aperture
NBHFA	norbornene hexafluoroisopropanol
NHA	numerical half-aperture
NH_4HF	ammonium fluoride
NILS	normalized image log-slope

nMOS	*n*-channel metal-oxide semiconductor
NMP	N-methylpyrrolidone
NVSM	nonvolatile semiconductor memory
OPC	optical proximity correction
OPD	optical path difference
OPE	optical proximity effect
PAC	photoactive compound
PAG	photoacid generator
PBOCST	poly(*tert*-butoxycarbonyl oxystyrene)
PBS	poly(butene sulfone)
PCB	printed circuit board
PCM	portable conformable mask
PDMS	polydimethylsiloxane
PE	photoelectron
PEB	postexposure bake
PECVD	plasma-enhanced chemical vapor deposition
PFOS	perfluorooctyl sulfonate
PGMA	poly(glycidyl methacrylate)
PGME	propylene glycol monomethylether
PGMEA	propyleneglycol monomethyl ether acetate
PHOST	polyhydroxystyrene
PMIPK	poly(methyl isopropenyl ketone)
PMMA	poly(methyl methacrylate)
pMOS	*p*-channel metal-oxide semiconductor
PMPS	poly(methylpentene sulfone)
ppb	parts per billion
PPDA	p-phenylenediacrylic acid
PREVAIL	projection reduction exposure with variable axis immersion lens
PROLITH	positive resist optical lithography
PSM	phase-shifting mask
PVD	physical vapor deposition
PVP	poly(vinyl pyridine)
PWB	printed wiring board
RB	rose bengal
RC	resistance capacitance
RELACS	resolution enhancement of lithography assisted by chemical shrink
RIE	reactive ion etching
ROMP	ring-opening metathesis polymerization
SADP	self-aligned double patterning
SAM	self-assembled monolayer
SCALPEL	scattering with aperture-limited projection electron lithography
S/D	source/drain

SEMC	single-electron memory cell
SLM	spatial light modulator
SPM	sulfuric peroxide mixture
STI	shallow trench isolation
STM	scanning tunneling microscope
TBEST	*tert*-butyl ester-protected 4-hydroxystyrene
TBMA	*tert*-butyl methacrylate
TBTFMA	*tert*-butyl-2-trifluoromethylacrylate
TCAD	technology computer-aided design
TE	transverse electric
TEM	transmission electron microscopy
TFE	tetrafluoroethylene
THF	tetrahydrofuran
THP	tetrahydropyranol
TM	transverse magnetic
TMAH	tetramethylammonium hydroxide
TMS	trimethylsilyl
TMSDEA	trimehylsilyldiethylamine
TPSHFA	triphenylsulfonium hexafluoroantimonate
TSI	top surface imaging
ULPA	ultralow-penetration air
UTR	ultrathin resist
UV	ultraviolet
VAP	vinyl addition polymerization
VEMA	poly(vinyl ether-*alt*-maleic anhydride)
VUV	vacuum ultraviolet
WET	wafer electrical test
XRR	x-ray reflectivity

Chapter 1

Introduction to Lithography

> "I have imposed upon myself, as a law, never to advance but from what is known to what is unknown, never to form any conclusion which is not an immediate consequence necessarily flowing from observation and experiment; and always to arrange the facts, and the conclusions which are drawn from them, in such an order as to render it most easy for beginners in the study of chemistry thoroughly to understand them."
>
> Antoine Lavoisier (1743–1794), *Traité Élémentaire de Chimie*

It is hard to think of an invention that has had a greater influence on the mass production of devices that have enabled humankind to communicate information with printed matter or photons or electrons or ions, or even with atoms, than lithography. The information superhighway and pathways of our present age—the information age—are literally paved with chips of crystalline silicon by lithography. Nearly every book, magazine, newspaper, brochure, flyer, catalog, and other print pieces that have been produced during the last 300 years and before the digital age of the last few decades have been printed with offset lithography. Nearly every integrated circuit (IC) in the chips that run the computers and telecommunication systems that power the information superhighway, as well as medical devices, electronics, home and industrial appliances, automobiles, and airplanes—to mention but a few—are made by semiconductor lithography. Without such ICs, we would have no powerful computers, no large-scale automation, no communication satellites, nor even any space exploration. There would certainly be no electronic calculators or digital watches, no transistor radios, portable tape recorders, personal digital assistants, internet, cell phones, etc. Many diagnostic procedures in medicine and dentistry rely on ICs, as do heart pacemakers and modern hearing aids. The impact of all of these things on our lives is tremendous. For instance, we can watch events on our television sets or mobile phones, or on the Internet, as they are happening thousands of miles away. We can withdraw our money from automatic teller machines almost everywhere in the world, without the aid of a bank clerk, whose function has been transformed from that of an accounts keeper to one of an intermediary between the customer and the computer. Many products are now manufactured, assembled, and shipped today by automatic machines that rely on ICs for their operation. Airplanes are guided by computer-controlled systems, and even airline seats are reserved worldwide by an instantaneous computer booking system. The list is almost endless and is growing every day. A world without lithography would be unrecognizable to anyone of us today, although we may not always be cognizant of its pervasiveness and reach in our daily lives. Lithography underpins most of the technical inventory of modern life.

More recently, lithography has reached out beyond electronics and is appearing in an increasingly wide range of scientific fields, including electrochemistry,[1] biochemistry,[2,3,4] biophysics,[5] photonics,[6,7] and medicine.[8,9] It is enabling a growing array of technical applications, including fluidic, electronic,[10,11] optoelectronic,[12] photonic, electroluminescent,[13]

---

[1] L. Rassaei, P.S. Singh, and S.G. Lemay, "Lithography-based nanoelectrochemistry," *Anal. Chem.* **83**, 3974–3980 (2011).

[2] M. Burns, B.N. Johnson, S.N. Brahmasandra, et al., "An integrated nanoliter DNA analysis device," *Science* **282**, 484–487 (1998).

[3] T. Blattler, C. Huwiler, M. Ochsner, B. Stadler, H. Solak, J. Voros, and H.M. Grandin, "Nanopatterns with biological functions," *J. Nanosci. Nanotechnol.* **6**, 2237–2264 (2006).

[4] V.N. Truskett and M.P.C. Watts, "Trends in imprint lithography for biological applications," *Trends Biotechnol.* **24**, 312–317 (2006).

[5] L. Pollack, M. W. Tate, A.C. Finnefrock, et al., "Time resolved collapse of a folding protein observed with small angle X-ray scattering," *Phys. Rev. Lett.* **86**, 4962–4965 (2001).

[6] A.S. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Paniccia, "A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor," *Nature* **427**, 615–618 (2004).

[7] V.R. Almeida, C.A. Barrios, R.R. Panepucci, and M. Lipson, "All-optical control of light on a silicon chip," *Nature* **431**, 1081–1084 (2004).

[8] M. Schena, D. Shalon, R.W. Davis, and P.O. Brown, "Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray," *Science* **270**, 467–70 (1995).

[9] A.A. Alizade, M.B. Eisen, R.E. Davis, et al., "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature* **403**, 503–511 (2000).

[10] S.R. Forrest, "The path to ubiquitous and low-cost organic electronic appliances on plastic," *Nature* **428**, 911–918 (2004).

[11] M.H.R. Lankhorst, B.W.S.M.M. Ketelaars, and R.A.M. Wolters, "Low-cost and nanoscale non-volatile memory concept for future silicon chips," *Nature Mater.* **4**, 347–52 (2005).

[12] L.M. Liz-Marzan, "Tailoring surface plasmons through the morphology and assembly of metal nanoparticles," *Langmuir* **22**, 32–41 (2006).

[13] G.H. Gelinck, H.E.A. Huitema, E. van Veenendaal, et al., "Flexible active-matrix displays and shift registers based on solution-processed organic transistors," *Nature Mater.* **3**, 106–110 (2004).

sensing,[14,15] energy-conversion, and storage devices,[16] to mention but a few. As we get ready to enter the era of "internet-of-things," whereby just about any device will be connected to the internet, the role of lithography will be even more critical because the IC of these devices will be patterned with lithographic techniques.

How did we get to where we are today? It all started with the invention of lithography in 1798, of photolithography and photography simultaneously in 1826, and subsequent developments in lithography ever since, all of which were made possible by antecedent developments in chemistry and optical physics over a period spanning more than 30 centuries, since records began to be kept. Through this unique marriage of chemistry and optics, the science and technology of lithography have evolved and made possible the mass production of printed materials, starting from the late 18th century during the industrial revolution and culminating in the microelectronics revolution that ushered in the mass production of microelectronic equipment, starting from the middle of the 20th century with the invention of the transistor in 1947 and the invention of the IC 12 years later. Innovations in lithography, new materials, and scaling to ever-smaller dimensions have led to many orders of magnitude of improvement in the capability of transistors to carry out computation, thus paving the way for the information age in which we currently live. All of these developments have radically influenced the course and trajectory of human civilization and development.

This book therefore deals with how chemistry mediates lithography, a topic that has not been previously discussed at length. In particular, it traces the arc of developments in lithography from a chemical perspective, starting from its invention and reaching back in an unmistakably continuous line to a period much earlier still. Like an arc, it has a beginning, a bow, and a tip. The beginning of this arc is the invention of lithography, the bow of the arc is the invention of photolithography, and the tip of the arc coincides with the development of semiconductor lithography and culminates in the state-of-the-art in the field today.

The objective of this book is not to lay out 30 centuries of the history of science, particularly that of physics and chemistry, like a long piece of wallpaper, and divide it into so many superficial categories after the manner of the encyclopedist and the abridger. Instead, we will focus on the lines of strategic change and carefully examine those moments in the history of

---

[14] E. Stern, J.F. Klemic, D.A. Routenberg, P.N. Wyrembak, D.B. Turner-Evans, A.D. Hamilton, D.A. LaVan, T.M. Fahmy, and M.A. Reed, "Label-free immunodetection with CMOS-compatible semiconducting nanowires," *Nature* **445**, 519–22 (2007).

[15] J. Wang, M. Musameh, and Y.H. Lin, "Solubilization of carbon nanotubes by nafion: toward the preparation of amperometric biosensors," *J. Am. Chem. Soc.* **125**, 2408–9 (2003).

[16] Z.L.Wang and J.H. Song, "Piezoelectric nanogenerators based on zinc oxide nanowire arrays," *Science* **312**, 242–6 (2006).

physics and chemistry that seem consequential to uncover the intellectual knots that had to be untied, which directly or indirectly aided the development of lithography. The treatment therefore need not follow in chronological order and linear fashion, but rather must be organized around similar coherent themes.

It is very useful to learn from the mistakes of early scientists and to examine particular intellectual hurdles associated with given periods, as well as the course of scientific developments that ran into blind alleys, but that nonetheless affected the progress of science in general and lithography in particular.

Since its invention, lithography has witnessed tremendous evolution. Many of its variants are now practiced; these range from stone plate lithography used in fine art printing, to offset lithography used in the printing of newspapers and the like, to semiconductor lithography, which utilizes a variety of exposure radiations to print ICs. While all of these variants of lithography are covered in this book, our emphasis will be on semiconductor lithography since it is the most scientifically and technically advanced form of lithography. And within semiconductor lithography, optical lithography is the most dominant technique used in fabrication of ICs. Most importantly, relative to other lithographic techniques, semiconductor lithography best exemplifies the marriage of chemistry and optics—a theme that we explore in depth in this book.

Lithography in its very essence is a series of chemical transformations—a fact recognized by its inventor, who called it "chemische druckerei" (chemical printing). In his treatise on lithography, *The Art of Drawing on Stone* (1824), Hullmandel wrote that lithography was 'an art entirely founded on chemistry.'[17] Like all chemical transformations, its currency of transaction is the electron, namely, outer-shell electrons of the atoms, molecules, and compounds of the material (also called resist) on which the image of the mask or the object to be printed is recorded. These electrons orchestrate distinct bond-breakage and bond-formation events in all of the process steps in lithography that ultimately lead to the contrast between the clear (exposed) area and the dark (unexposed) area of the image-recording medium. Preparation of the substrate, coating of the resist, the actual exposure, and subsequently the postexposure thermal and related processing are all characterized by distinct chemical processes that taken together are ultimately about chemical bond breakage and formation.

In the substrate preparation step involving priming, the surface chemistry of the substrate is modified to promote the adhesion between the substrate and the resist material during coating. In the exposure process proper,

---

[17] C. Hullmandel, *The Art of Drawing on Stone*, Rudolph Ackerman Publisher, London, p. vii (1824).

exquisite radiation chemistry takes place inside the radiation sources in order to generate the exposure radiation, but also because of the interaction of these radiations—be they photons, ions, electrons, or x-rays—with the outer-shell electrons of the radiation-sensitive compounds and molecules in the resist, leading to bond breakage and/or bond formation.

In the postexposure thermal processing steps, thermally driven diffusion and reaction of the active species that catalyze deprotection, bond scission, or crosslinking reactions in the resist are engendered. In the development step, appropriate areas of the resist film are dissolved away either through physical dissolution (involving no chemical reaction) or through acid–base neutralization reactions between the exposed areas of the resist and the developing solvent (depending on polarity). This is the basis of the contrast between the exposed and unexposed areas of the resist film.

It should be pointed out that there are some lithographic techniques where the patterning action is mediated by mesoscale and even nanoscale material transformations driven by thermodynamic forces, as in directed block copolymer self-assembly and molecular self-assembly, or by mechanical forces, as in imprint lithography, although at these length scales the roles of electrons in these processes are not apparent. However, when reduced to their quantum limits, the roles of electrons in these same processes become significant.

The main attribute of optical projection lithography that made it the manufacturing technology of choice for ICs since the beginning of the IC era is the tremendous throughput advantages it offers through its ability to reproduce an entire IC layout from a master (or reticle) in a single exposure, in contrast to other technologies that address a field point by point. In addition to the compelling throughput advantages, there were resolution and cost advantages as well. The infrastructure for light sources, lenses, reticles, photosensitive polymers, and other optical materials developed for other optical and photographic applications were appropriated and applied to IC lithography, allowing development resources to be shared.

In 1965 Gordon Moore[18] postulated that the exponential growth in the number of transistors in an IC leads to certain technical and economic advantages. Smaller transistors switch faster, allowing more operations per second. And more transistors with more interconnections enable computations of much greater complexity to be achieved. This postulate has since been codified as Moore's law,[19] which states that the complexity of ICs as measured by the number of transistors approximately doubles every two years (see Fig. 1.1). This law has led to unprecedented growth in the computer industry. Technologies that

---

[18] G.E. Moore, "Cramming more components onto integrated circuits," *Electronics* **38**(8), 114–117 (1965).

[19] G.E. Moore, "Progress in digital electronics," *IEEE Meeting 1975 Technical Digest*, 11–13 (1975); G.E. Moore, "Lithography and the future of Moore's law," *Proc. SPIE* **2437**, 2–17 (1995) [doi: 10.1117/12.209151].

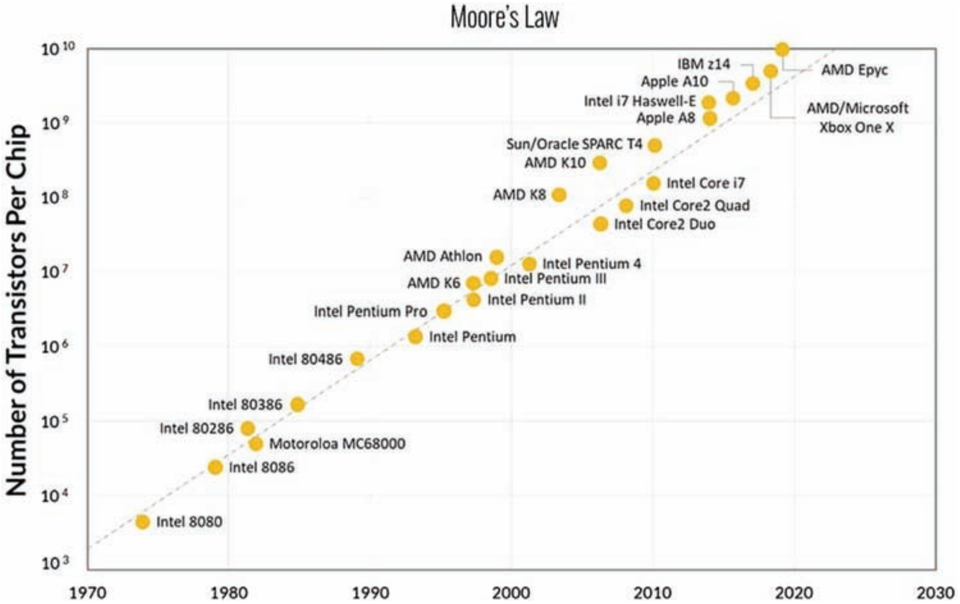


Figure 1.1 Plot of CPU transistor count versus year of introduction.[20]

were once available only in supercomputers are now commonly available in children's toys. Satellite communications networks that were once the domain of the military now help drivers find their way to their locations.

A good metric for measuring progress in IC lithography is resolution, the ability to resolve and distinguish two neighboring features on the chip. Two main approaches for improving resolution of optical projection lithography include (1) decreasing the wavelength of the optical source and (2) increasing the numerical aperture of the optical system, as derived from Rayleigh's resolution criterion[21] and shown in the following equation:

$$w = k_1 \frac{\lambda}{\mathrm{NA}}, \tag{1.1}$$

where w is the half-pitch of the feature being printed, k_1 is a process-dependent parameter, λ is the exposure wavelength, and NA is the numerical aperture of

---

[20] Wikipedia (https://en.wikipedia.org/wiki/Transistor_count); https://www.halocarbon-es.com/semiconductor-fabrication-solutions/.

[21] Lord Rayleigh, "Investigations in optics, with special reference to the spectroscope," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, Series 1–6, **18**(49), Part XXXI, pp. 261–274, Part XLVI, pp. 403–411, and Part LVI, pp. 477–486 (1879); Lord Rayleigh, "Investigations in optics, with special reference to the spectroscope," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, Series 7, **9**(53), Part V, pp. 40–55(1879); Lord Rayleigh, "On the theory of optical images, with special reference to the microscope," *London, Edinburgh, Dublin Phil. Mag. J. Sci.* **42**(255), Part XV, pp. 167–195 (1896).

the optical system, which is defined in terms of the maximum cone angle of rays ($\theta_{\max}$) subtended by the maximum pupil diameter at the image plane as $\mathrm{NA} = n \sin \theta_{\max}$ (where n is the refractive index in image space). Equation (1.1) unites chemistry with optics in optical projection lithography. For it is the interaction of the exposure photons, electrons, ions, or x-rays transmitted through the NA of the optics, with the electrons of the high-contrast recording medium—photosensitive materials (resists)—that mediate chemical phenomena on which lithographic patterning is based. The exposed part of the resist is altered relative to the unexposed part, leading to contrast between the two regions during development.

The progress of optical lithography is partly then the result of decreasing the exposure wavelength. From the initial broadband sources, the IC industry made a migration first to the mercury g-line (436 nm), then to the i-line (365 nm), and then switched to exciplex[22] laser sources—first KrF (248 nm) and now ArF (193 nm). A migration to an F_2 excimer laser source was contemplated by the IC industry but did not materialize, even after significant investments, primarily because of birefringence issues associated with calcium fluoride (CaF_2) used in the lens elements. The migration toward shorter wavelengths naturally limits the pool of available photosensitive materials that could be employed in resist formulations.

Because the NA of the optical system limits the spatial frequencies that can be transmitted to expose the resist, the NA of lens designs has migrated from 0.2 to 0.42 to 0.63 to 0.75 to 0.95. With a fundamental limit of NA = 1.0 for a conventional optical system, the introduction of immersion ArF lithography has enabled the migration to hyper-numerical-aperture (> 1.0) optical systems. It is noteworthy that the drive toward high NA is at a cost of decreased depth of focus and increased difficulty in fabricating a lens with adequate field size.[23]

---

[22] The term "exciplex" refers to a combination AB^* of two different atoms; it exists only in an electronically excited state and dissociates as soon as the excitation is lost. It differs from an "excimer," an excited-state dimer of two similar atoms AA^*. The exciplex lasers that have found widespread applications in lithography are based on KrF^* and ArF^* formed in electrical discharge in a mixture containing krypton and fluorine in KrF lasers, and argon and fluorine in ArF lasers, respectively. The only true excimer laser that has found application in lithography is based on excited-state $F_2{}^*$ dimers (lasing at 157 nm). The KrF^* and ArF^* exciplex and F_2 excimer survive for a few nanoseconds, long enough to participate in laser action. As soon as the excitation is gone, the atoms separate because the potential-energy curve of their ground state is repulsive. Unfortunately, the widely used misnomer "excimer laser" appears in the literature to describe exciplex lasers based on $XeCl^*$ (lasing at 308 nm), KrF^* (lasing at 248 nm), and ArF^* (lasing at 193 nm) when "exciplex laser" is appropriate. In this book we will use the appropriate terms. [For the photochemistry of excimers and exciplexes, see, e.g., P.W. Atkins, *Physical Chemistry*, 5th ed., W.H. Feeman and Co., New York, p. 609 (1994); P. Suppan, *Chemistry and Light*, Royal Society of Chemistry, Cambridge, England, pp. 104–110 (1994).]

[23] M.J. Bowden, "The Lithographic Process: The Physics," in L.F. Thompson, C.G. Willson, M.J. Bowden, eds., *Introduction to Microlithography*, American Chemical Society, Washington, D.C., Chapter 2, pp. 19–138 (1994).

Today, the leading-edge microelectronic devices are being made with immersion extreme-ultraviolet (EUV) lithography at 13.5 nm, which is inevitably a continuation of the progression from longer-wavelength lithographies and is dictated by the requirements for higher resolution and the drive in the IC industry toward greater packing density and higher speeds, as noted earlier.

The resolution that will be necessary for the manufacture of future generations of ICs with feature sizes below 10 nm is beyond the limits of 365-nm, 248-nm, and 193-nm UV lithographies. According to the *International Technology Roadmap for Semiconductors* (ITRS),[24] EUV lithography at 13.5 nm is a promising candidate for achieving such high resolution. With EUV lithography, sub-10-nm devices can be fabricated with conventional masks using reflective optics $\geq$0.45 NA.

Besides increasing the NA and decreasing the exposure wavelength, an approach currently in use for improving resolution in semiconductor lithography for any pitch below $0.5\lambda/\mathrm{NA}$, which is beyond the resolution of a single exposure, involves multiple patterning. Examples of the most popular multiple-patterning techniques include pitch splitting and spacer patterning. Pitch splitting involves splitting a feature into two groups, each patterned with a different mask. Spacer patterning involves patterning of a feature, followed by deposition of a spacer material (hard mask) on the patterned feature, followed by etching to remove all of the spacer material on the horizontal surfaces, leaving only materials on the sidewalls, and then removing the original feature and leaving behind only the spacer material on the two sidewalls. Since there are two spacers for each line, the line density is doubled. Another approach for improving single-exposure resolution includes the use of reticle enhancement techniques such as phase-shifting masks.

The scope of this book covers the major developments and trends in lithographic patterning since the invention of lithography, along with an overview of the patterning strategies and materials used, and the unique features, strengths, and limitations of the main lithographic techniques with resolutions in the micron- and nanometer-length scales, which are commonly referred to collectively as microlithography and nanolithography, respectively. The book also covers the emerging trends in lithographic patterning, as well as the current and potential applications of the resulting patterned structures and surfaces. Emphasis is placed on standard advanced lithographic techniques used in the manufacture of IC devices: optical, EUV, electron-beam, and ion-beam lithographies. Also emphasized are nonconventional or alternative lithographic techniques such as imprint, self-assembly, and scanning probe lithographies.

---

[24] http://www.itrs.org.

As stated above, this book attempts to systematically reappraise the main developments in chemistry and physics (particularly optical and atomic physics) that have ultimately led to lithography as practiced today, especially in semiconductor lithography—the most advanced form of lithography.[25] The task is no doubt an onerous one, but one that must be done in order to unearth the hidden connections between the various streams of thoughts that materialized as lithography and subsequently as semiconductor lithography.

---

[25] While the emphasis of this book is on semiconductor lithography, attempts will be made where necessary to highlight relevant aspects of the low-technology variants of lithography as practiced in offset lithography and fine art lithography.

Chapter 2

Invention of Lithography and Photolithography

> “God grant that it [lithography] may soon spread all over the world; that it may prove useful to mankind, and contribute to its improvement; and that it may never be abused to any dishonourable or wicked purpose; and I shall then never cease to bless the hour in which I invented it.”
>
> Alois Senefelder (1771–1834)[1]

2.1 Introduction

Before beginning this journey through how lithography and subsequently photolithography came into existence, and how the technology has evolved ever since, it is necessary to review the cultural, scientific, and technological framework out of which these planographic printing techniques emerged. A great deal has already been written about the artistic and sociological impacts of the fine art lithographers and photographers of the 19th and 20th centuries, which will not be reviewed here. The introduction of lithography at the end of the 18th century and the simultaneous introduction of photolithography and photography in the 3rd decade of the 19th century greatly influenced every facet of human endeavor, particularly in scientific and technological pursuits.

With the rise during the 17th and 18th centuries of many scientific fields based on visual observations—systematics (or the taxonomy of biological organisms), topology, and geology—the ability to make illustrative drawings of scientific observations became an increasingly important skill. A very high premium was also placed on the ability to make realistic and accurate

---

[1] A. Senefelder, *A Complete Course of Lithography*, Da Capo Press, Inc., New York, p. 85 (1997). [This book is an unabridged republication of the first English edition published in London in 1819, which was originally published in German under the title *Vollständiges Lehrbuch der Steindruckerei*.]

drawings of mechanical arts and engineering. Concomitant with these developments was the rapid industrialization brought about by the Industrial Revolution in Britain and its influence elsewhere during the late 18th century, which increased the need for draftsmen to produce drawings to aid in the production and use of the ever-expanding number of machines needed for the industries.[2]

These factors naturally increased the value of individuals who were able to record images directly as seen in nature without interpretation by an artist or a draftsman. Being able to draw provided one with upward social mobility in an era when class and social distinctions determined one's worth in society. The origins of this exact form of drawing stem from art conventions of perspective drawings set forth in the early Renaissance, which were in turn derived from Euclid's geometry. The removal of artifice from art through the codification of a system of drawing that produced what was understood to be scientifically accurate reflections of the world was the main goal of the rules of perspective. And by the 17th century, these Renaissance rules were brought to their apex by the great Dutch artists. In addition to rules of perspectives, artists and draftsmen were aided by a variety of mechanical drawing devices or drawing machines like the perspectograph (invented in 1816 by Thomas Alason[3]), the physiognotrace (invented in 1786 by Gilles-Luis Chrétien), and cameras, mostly *camera obscura* and *camera lucida.*[4]

The *camera obscura* was the most common camera in use at that time. Although it was referred to in writings dating from antiquity, the *camera obscura* was first described in detail by Johann Baptista Porta in 1533. Literally meaning "dark room," the *camera obscura* is basically a dark chamber or box with an opening through which light passes. The light entering the camera opening falls onto the wall opposite the opening to form an image. A *camera obscura* with its small opening requires no lens to be operational. However, larger openings outfitted with a lens produce very bright images.[5]

Another type of camera, the *camera lucida*, which means "light room" and may refer to the fact that the instrument did not require any enclosure, box, or lens as in the *camera obscura*, was first described in 1807 by William Hyde Wollaston, although the exact source of the name is unknown. There are basically two main types of *camerae lucidae*. The first one is a device with a reflective glass mounted on a stand, such that the glass is positioned at a

---

[2] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, pp. 4–10 (1991).

[3] T. Allason, "Papers in Polite Arts," *Transaction of the Society, Instituted at London, for the Encouragement of Arts, Manufacturers, and Commerce* **34**, 145–147 (1816).

[4] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, pp. 4–5 (1991).

[5] ibid., pp. 6–8.

45-deg angle to the paper. This allows the artist or draftsman to see the reflection of the scene on the glass and also to look through the glass to the paper. The scene is thus perceived to be superimposed on the paper and can be easily copied. Images made with this type of *camera lucida* are laterally reversed. The other type of *camera lucida* uses a four-sided prism fastened to a small stand. The use of this type of instrument calls for the artist or draftsman to position the eye so that it sees both the image of the scene reflected directly on the retina and the scene on which the paper is copied.[6]

Being more portable and convenient to use than the *camera obscura* because it was a very simple device with only a few parts, it did not require a darkened surrounding, and it was easy to use once its operation had been learned. Thus, *camera lucida* quickly found wide use among draftsmen, especially for tasks such as enlarging or reducing drawings and for copy work. Its portability made the *camera lucida* useful to both professional artists and amateurs making topological and architectural views. In particular, the *camera lucida* had special appeal to travelers because it could be used to make "correct representations"[7] of the places visited even by those who could not draw as well as artists. The renowned traveler and British naval officer Col. Basil Hall wrote a laudatory book about the *camera lucida*, in which he claimed that, once the operation of this camera was mastered, the user "could rove where he pleases, possessed of the magical secret of recording the features [of a scene] freed from the triple misery of Perspective, Proportion and Form—all responsibility respecting these taken off his hands."[8]

Equally significant was the increased demand during this period not only for unique works of art such as portraits and drawings, but also for reproductions of works of art such as prints. Demand for printed books also witnessed a phenomenal rise during this period. By sheer necessity, new printing methods and new methods of making the paper needed for the mass production of inexpensive art works and books were developed during this time. Lithography, the only new printing method discovered since the invention of the movable type some 300 years earlier, was introduced during this period. It revolutionized printmaking in every facet because it was faster and cheaper than any of the previous printing processes.[9]

---

[6] ibid., pp. 6–8.

[7] B. Hall, *Forty Etchings, from Sketches Made with the Camera Lucida in North America in 1827 and 1828*, Cadell & Co., Edinburgh (memorandum) (1829).

[8] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, pp. 6–8 (1991).

[9] ibid., pp. 8–9.

2.2 Invention of Lithography

The invention of lithography dates back to 1798, when the playwright Alois Senefelder[10] (see Fig. 2.1), writing with ink[11] (prepared with wax, soap, and lampblack on a tablet of polished Kelheim limestone,[12] the border of which he lined with wax,[13] and on the entire surface of which he applied an acid

---

[10] A. Senefelder, *Vollständiges Lehrbuch der Steindruckerei* [*A Complete Course in Lithography*, in German], Karl Thienemann, Munich and Karl Gerold, Vienna, Publishers (1818).

[11] The lithographic ink is used for writing or drawing immediately on the stone, for covering the surface of it, or for transferring drawings executed on the paper to the stone. The ink functions principally to fill the pores of the stone in those places to which it is applied with an oily greasy substance, and it also resists the action of *aqua-fortis* and other acids. Inks are generally of two kinds: one of a thicker nature for drawing on the stone, and another of a more-liquid nature for transferring drawings to the stone. Senefelder discloses the following 8 recipes as making the best ink compositions: (1) white beeswax (8 parts by weight), soap (2 parts by weight), and lampblack (1 part by weight); (2) wax (12 parts by weight), common tallow (4 parts by weight), soap (4 parts by weight), and lampblack (1 part by weight); (3) wax (12 parts by weight), shellac (4 parts by weight), soap 4 parts by weight), and lampblack (1 part by weight); (4) tallow (8 parts by weight), shellac (8 parts by weight), soap (4 parts by weight), and lampblack (1 part by weight); (5) wax (8 parts by weight), shellac (4 parts by weight), mastic (4 parts by weight), soap (4 parts by weight), and lampblack (1 part by weight); (6) wax (8 part by weight), tallow (4 parts by weight), shellac (4 parts by weight), soap (4 parts by weight), and lampblack (1 part by weight); (7) wax and gum guaicum melted together in equal quantities (12 parts by weight), tallow (4 parts by weight), soap (4 parts by weight), and lampblack (1 part by weight); (8) wax (6 parts by weight), shellac (4 parts by weight), tallow (2 parts by weight), mastic (3 parts by weight), Venetian turpentine (1 part by weight), soap (4 parts by weight), and lampblack (1 part by weight). Preparation of the ink calls for mixing together all of the different ingredients (in the manner listed above for each composition) except the soap in an iron saucepan, and exposing the mixture to a strong fire until the whole of the mass ignites. When the quantity of the mixture is reduced to one-half, the saucepan is carefully covered, or poured into a container of water to extinguish the flame and cool the substance. Colorants for the ink, besides lampblack, may include indigo, blue lake, vermillion, and red ochre, and various other colors, provided they do not alter the nature of the soap. Distilled water is recommended for dissolving the ink. [A. Senefelder, *A Complete Course of Lithography*, Da Capo Press, Inc., New York, pp. 110–118 (1997).]

[12] The best Kelheim stones for making lithographic stone plates, Senefelder reported, came from Solnhofen, a village in Bavaria, Germany. Chemically, these stones consist of calcium carbonate. In nitric and other acids, they can be almost entirely dissolved, with the carbonic acid escaping in the form of gas. The demand for these stones at the time was very high, but their availability was restricted by political and economic events such as the continental blockade and the various military campaigns of the period. Thus, early on it became a matter of national priority to find either local sources of limestone or suitable substitutes for printing stones in order to free countries like England and France from their dependence on German limestone [cited in M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 97 (1991)].

[13] A substance that is sometimes used for covering the stones, which, like the ink compositions, resists the action of *aqua-fortis*, is known under the name of etching ground, as it differs very little from that which engravers use for their copper plates. Senefelder's recipe for this calls for: wax (12 parts by weight), mastic (6 parts by weight), asphaltum (4 parts by weight), colophony (2 parts by weight), and tallow (1 part by weight). Preparation of the etching ground involves melting all of the above ingredients together on the fire in an iron saucepan, and exposing the latter to heat until the asphaltum is completely dissolved. Then, the mass is left burning until it is reduced to

Figure 2.1 Portrait of Alois Senefelder, the inventor of lithography (printed with permission from the Deutsches Museum, Munich).

solution[14] for a short period of time), led him to the serendipitous discovery that the limestone was etched away in areas on which he had not written. The wax writing solution resisted[15] the acid! Upon removing the acid, he observed that the written or image areas were raised by about 1/10 inch above the rest of the stone. By carefully rolling ink over the stone surface, he could deposit ink pigments only on the image areas of the stone, which he could easily transfer to a sheet of paper with little pressure (see Fig. 2.2 for the first lithographic press constructed by Senefelder and Fig. 2.3 for a portable hand press also constructed by Senefelder in 1818). Senefelder called his invention *steindruckerei*, which means "stone printing."[16]

---

about two-thirds, after which the fire is extinguished, and the product is cooled and formed into different shapes and preserved for use. Alternatively, a good etching ground can be prepared from plain wax if it is left boiling and burning until nearly five parts of it are consumed. [A. Senefelder, *A Complete Course of Lithography*, Da Capo Press, Inc., New York, pp. 122–123 (1997).]

[14] *Aqua-fortis*, or nitric acid, muriatic acid, vinegar, tartaric acid, malic acid, and oxalic acid are typical acids used in stone plate lithography to etch away nonimage, nonwritten areas of the stone, i.e., areas that are not properly covered with grease.

[15] Because the wax resisted the etching action of the acid, materials behaving as such within the field of lithography have since then been called resists.

[16] According to Senefelder, the invention of lithography draws its direct origin under the most trifling of circumstances in a story he thus narrated: "I had just succeeded in my little laboratory in polishing a stone plate, which I intended to cover with etching ground, in order to continue my exercises in writing backwards, when my mother entered the room, and desired me to write her a bill for the washer-woman, who was waiting for the linen; I happened not to have even the smallest slip of paper at hand, as my little stock of paper had been entirely exhausted by taking proof of impressions from the stones; nor was there even a drop of ink in the inkstand. As the matter would not admit of delay, and we had nobody in the house to send for a supply of the deficient materials, I resolved to write the list with my ink prepared with wax, soap, and lampblack, on the stone which I had just polished, and from which I could copy it at leisure.

Figure 2.2 Wooden lithographic press constructed by Alois Senefelder and used in his invention of lithography (printed with permission from the Deutsches Museum, Munich).

Some time after this I was just going to wipe this writing from the stone, when the idea all at once struck me, to try what would be the effect of such a writing with my prepared ink, if I were to bite in the stone with *aqua-fortis* [nitric acid]; and whether, perhaps, it might not be possible to apply printing ink to it, in the same way as to wood engravings, and so take impressions from it. I immediately hastened to put this idea in execution, surrounded the stone with a border of wax, and covered the surface of the stone, to the height of two inches, with a mixture of one part of *aqua-fortis*, and ten parts of water, which I left standing five minutes on it; and on examining the effect of this experiment, I found the writing elevated about a 10th part of a line (or 1/120th part of an inch.) Some of the finer, and not sufficiently distinct, lines had suffered in some measure, but the greater part of the letters had not been damaged at all in their breadth, considering their elevation; so that I confidently hoped to obtain very clear impressions, chiefly from printed characters, in which there are not many fine strokes.

I now proceeded to apply the printing ink to the stone, for which purpose I first used a common printer's ball; but, after some unsuccessful trials, I found that thin piece of board, covered with fine cloth, answered the purpose perfectly, and communicated the ink in a more equal manner, than any other material I had before used. My further trials of this method greatly encouraged my perseverance. The application of the printing ink was easier than in the other methods, and I could take impressions with a fourth part of the power that was requisite for an engraving, so that the stones were not at all liable to the danger of breaking; and, what was the greatest moment for me, this method of printing was an entirely new invention, which had occurred to nobody before me." [A. Senefelder, *A Complete Course of Lithography*, Da Capo Press, Inc., New York, pp. 9–11 (1997). This book is an unabridged republication of the first English edition published in London in 1819, which was originally published in German under the title *Vollständiges Lehrbuch der Steindruckerei* in 1818.]

Figure 2.3 Lithographic stone and small hand press constructed by Alois Senefelder in 1818 and used for fast production of lithographic prints. Note the copy made from the handwritten note on the lithographic stone. (Printed with permission from the Deutsches Museum, Munich.)

Because Senefelder was printing in relief, this method is not the same as what we call lithography today. Nevertheless, Senefelder's invention was a significant advancement over the older, copper-plate engraving printing method, which was rather laborious, slow, and costly. But the basis of all modern lithography derives also from another observation of his:[17] When a greasy (oily) image is drawn on a slab of limestone surface of appropriate size and ground flat to desired texture, perhaps with a pencil, following which a solution of gum arabic[18] (preferably acidic in the pH range between 3.5 and 5)

---

[17] According to Senefelder: "My whole process was therefore, as follows: To wash the polished stone with soap-water, to dry it well, to write or draw upon it with the composition of ink of soap and wax, then to etch it with *aqua-fortis*; and lastly, to prepare it for printing with an infusion of gum-water. I had hoped to have been able to dispense with the gum-water, but was soon convinced that it really enters into chemical affinity with the stone, and stops its pores still more effectually against the fat, and opens them to the water. In less than three days after my first idea, I produced as perfect and clear impressions as any that has since been obtained." [A. Senefelder, *A Complete Course of Lithography*, Da Capo Press, Inc., New York, pp. 31–32 (1997).]

[18] The lithographic stone, being made of limestone and therefore containing numerous pores, is equally capable of imbibing greasy as well as watery substances through these pores. These substances have the quality of adhering to the particles that make up the stone, but can be easily

is applied, and when an ink roller is subsequently rolled over the entire stone surface, ink is only deposited on the image areas on the stone (i.e., areas on the stone drawn with the greasy pencil) and is repelled in the nonimage areas (i.e., areas of the stone without the drawings from the greasy pencil). With a little pressure, he could easily transfer the image on the stone to a paper laid face down on the stone. This invention—based on the fact that oil and water repel each other—is the basis of all contemporary lithographic printing. He called this invention *chemische druckerei*, which means "chemical printing," because the process depends on the chemical interaction of grease, dilute acid, gum arabic, and water, rather than the stone from which the name lithography is derived.

The process has been variously called *steindruckerei*, polyautography, engraving or printing from stone, and *imprimerie chemique*. The French were the first to call the process by its current name, *lithographie*,[19] which means "writing with stone" and is derived from the Greek words *lithos* meaning stone and *graph* meaning to write.

Not long after Senefelder's original discovery, artists across Europe employed it to make reproductions of the works of old masters and, in time, recognized it as a valuable medium for their own original works. "Not only was it the first printing process to be introduced after the invention of the movable type 300 years earlier," Barger and White[20] assert that "it was also based on chemical, not mechanical, principles." Descriptions of the process were widely disseminated in scientific and technical journals, as well as in the popular press. By 1815, the city of Mulhouse, France, had developed into a

---

separated from them so long as the nature of the stone is not altered. This alteration is effected by sulfuric, tartaric, and phosphoric acids. Water evaporates gradually from the pores when the stone dries, but gum arabic and other greasy substances do not. Printing ink cannot adhere to the stone so long as it contains a sufficient quantity of water. In general, it adheres only faintly to the calcium carbonate surface and does not get stronger affinity to it until its pores are filled with the greasy hydrophobic etching ground (resist), to which the ink, also hydrophobic, can favorably unite. This affinity of adhesion of the printing ink with the resist on the stone takes place only when the two hydrophobic substances can directly contact each other. If this contact is interrupted (which may happen when the etching ground has penetrated too deeply into the stone and has left the surface), the etching ground will not take the printing ink until direct contact has been re-established between these two greasy substances. When applied to a well-polished stone, the solution of gum arabic, being very hydrophilic, produces the effect that the wetted spot will not accept the hydrophobic ink as long as it remains wet. However, as soon as the spot dries up, the ink adheres to it, but can be easily wiped off with a sponge and water. The gum arabic does impart to the stone the quality of rejecting the printing ink, a quality that is made all the more durable if the acid has been previously applied to this spot before the application of the gum arabic. On this principle lies lithographic printing. [A. Senefelder, *A Complete Course of Lithography*, Da Capo Press, Inc., New York, pp. 146–147 (1997).]

[19] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, pp. 8–10 (1991).

[20] ibid., p. 9.

major center for lithography in Europe, and many people came there from England and Germany to learn the process.[21] A. Hyatt Mayor asserts that by 1825 "lithography matured its classic style when Goya, as an exile in Bordeaux, drew four big stones of bullfights. At age 79, when most men are happy if they can refine what they invented years earlier, Goya created the first great works of art in the medium, the first that combine the freedom of a sketch with a black-and-white as rich and as satisfying as a painting. The road that he opened was soon followed by great artists like Delacroix, Daumier, Degas, Lautrec, Bonnard, Vuillard, Matisse, Picasso, and practically every prominent painter in France, and ultimately by those in other countries."[22]

In England, Charles Hummandel was one of the early adopters of lithography. He set up a printing shop in London to develop the technique and make it commercially viable. Subjects related to lithography appear in the Royal Institution lecture lists in the 1820s and 1830s, suggesting that the topic was of interest to the scientific community, as it was to the public. Michael Faraday and the Royal Institution during this time were engaged with research related to improving image reproduction in general and lithography in particular.

That images could lighten up dense text and make ideas more accessible, and that books could be cheaply and splendidly illustrated using lithography were facts not lost on Faraday, the Royal Institution, and artists of the era. This was quite a topical and engaging subject at the time. Faraday himself gave a discourse on the art of lithography on 3 March 1826 and pointed out how the technique was beginning to revolutionize the dissemination of imagery in books and periodical publications. He was aware of what the development of lithography could mean for modern life. Developments in lithographic printing and image reproduction techniques, coupled with the wider use of steam-driven presses, made it possible for books to be more widely available, less expensive, and better illustrated, thus making learning easier and information exchange faster.[23]

Faraday was an avid practitioner of the art of lithography. From his scrapbook in the archives of the Royal Institution, we see records of a number of lithographs of mountainous landscapes (one is dated 25 April 1825) with initials 'GB' (presumably, George Barnard) and 'MF' (presumably, Michael Faraday), showing that the two men practiced lithography together, perhaps under Hullmandel's guidance.[24] In his treatise on lithography, *The Art of*

---

[21] ibid., p. 9.

[22] A.H. Major, in A. Senefelder's *A Complete Course of Lithography*, Da Capo Press, Inc., New York, pp. V–VI (1997) (A. Hyatt Mayor provides an excellent introduction to the 1977 edition of Senefelder's classic book).

[23] J. Hamilton, *A Life of Discovery: Michael Faraday, Giant of the Scientific Revolution*, Random House, New York, pp. 114, 211–214, 272 (2002).

[24] M. Faraday's Scrapbooks, vol.1, p. 32, Royal Institution Archives, London (1825).

Drawing on Stone (1824), Hullmandel wrote that lithography was ‘an art entirely founded on chemistry,’[25] which likely was the aspect of the technique that attracted Faraday’s attention over the more mechanical reproductive arts such as engraving on metal or wood.[26]

Another notable event in the early history of lithography in England had to do with settling a dispute (as we shall see in Chapter 4) over who discovered the composition of water. There were three claimants to this discovery: James Watt and Antoine Lavoisier each claimed the discovery for himself, and the third claim was made on behalf of Henry Cavendish. In 1839, twenty years after the death of the last of the three principals in this dispute, their friends started to argue as to who was the discoverer of the composition of water. This dispute was settled in favor of Cavendish only when the President of the British Association for the Advancement of Science published a lithographed facsimile of Cavendish’s original notebook, proving that Cavendish made the discovery earlier than Watt, and certainly much earlier than Lavoisier. Here we see yet another dimension of the role lithography has played in the history of science.

From France, lithography spread westward to the United States and eastwards to Persia. Remarkably, British engravers and aquatinters fought the adoption of lithography in England, fearing that it threatened their monopoly on reproducing paintings and drawings. They persuaded Parliament to place a prohibitive import duty on high-quality German limestone, preventing it from being imported. For their part, American engravers were under the impression that lithography was just a shortcut to what engravers did; however, because they were not part of an organized body that could lobby the U.S. Congress, the successful U.S. printmakers Currier and Ives were free to exploit the process for reproducing thousands of paintings and drawings.[27]

The first steam lithographic press was invented in France in 1850 and introduced into the United States by R. Hoe around 1868.[28] Lithographic stones were used for the image, and a blanket-covered cylinder received the image from the plate and transferred it to the substrate in the steam lithographic press. Direct rotary presses for lithography using zinc and aluminum metal plates were introduced in the 1890s.[29] The first lithographic offset press was independently invented in 1903 by Ira A. Rubel (a paper manufacturer) and A.F. Harris.[30]

---

[25] C. Hullmandel, *The Art of Drawing on Stone*, Rudolph Ackerman Publisher, London, p. vii (1824).

[26] J. Hamilton, *A Life of Discovery: Michael Faraday, Giant of the Scientific Revolution*, Random House, New York, p. 214 (2002).

[27] ibid.

[28] https://dynodan.com/magazine-printing/process/printing-history/printing-history-history-of-lithography/

[29] ibid.

[30] ibid.

The lithographic offset process rose to become the most popular form of printing during the 1950s as the quality of plates, inks, paper, etc., improved tremendously. Beginning from the late 1950s, it became the dominant printing technique because it produces sharper, cleaner image reproductions over letterpress, and it was also less expensive in comparison to gravure. Today, the majority of printing, including newspapers, is done by the offset process.

It must be emphasized that lithographic printing is a planographic printing technique in that the inked imaged areas are on the same level as the uninked, nonprinting areas. It differs from relief or raised-image processes (e.g., letterpress) and intaglio or incised-image processes.[31]

As the popularity of lithography grew soon after it had been invented, some of the literature reports about the lithographic process suggested that it might be possible to transfer designs directly onto the stone if the design could be made translucent and if the lithographic stone could be varnished with a light-sensitive substance. Light passing through the design could then be transferred to the prepared stone below.[32] This idea came to fruition around 1826 at the hands of Joseph Nicéphore Niépce, who succeeded in making the first permanent photograph of nature, using photolithographic techniques.

2.3 Invention of Photolithography

Photography and photolithography share a common history. They were both invented with the same technique by the same man—Joseph Nicéphore Niépce—who succeeded in taking the first permanent photograph from nature, using the action of light on a photosensitive substrate. The invention of photography has been variously ascribed to several people: Wedgewood, who in the 1790s conceived the original idea of permanently fixing the images of the camera through the darkening action of silver salts in light, but was unsuccessful in practice; Fox Talbot, who in 1835 introduced the negative/positive process, the principle still employed in photography today; and Daguerre, who in 1837 invented the first practical process of photography (the Daguerreotype process). However, the credit belongs solely to Joseph Nicéphore Niépce, who obtained the first permanent recorded image from nature in 1826.[33]

To understand the invention of photolithography, it is necessary to examine the work of those who had been conducting experiments that could lead to permanent photographic imaging with light; these men have been

---

[31] L. Stroebel and R. Zakia, Eds., *Encyclopedia of Photography*, 3rd ed., Focal Press, Boston, p. 310 (1993).

[32] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 10 (1991).

[33] H. Gernsheim and A. Gernsheim, *The History of Photography: From the Camera Obscura to the Beginning of the Modern Era* **1**, Thames & Hudson, London (1969).

variously referred to as pioneers or fathers of photography and include Thomas Wedgewood (1771–1805), Joseph Nicéphore Niépce (1765–1833), Louis Jacques Mandé Daguerre (1789–1851), and William Henry Fox Talbot (1800–1877).

2.4 Pioneers of Photolithography

The first recorded attempts at imaging using light-sensitive materials were made by Johann Heinrich Schulze (1687–1744), who discovered the light sensitivity of silver nitrate in 1723.[34] Following Schulze's experiment, knowledge of the chemical and optical principles of photography was fairly widespread and was published in serious scientific treatises and even in popular books of amusing parlor tricks. Therefore, the circumstance that photography was not invented much earlier remains the greatest mystery of its history.[35] Influenced by Schulze's work, subsequent experimenters used light-sensitive materials as part of their light-sensitive inks. Imaging on a more sophisticated level was done sometime during the 1790s by Thomas Wedgwood (1771–1805),[36] who in his first attempt sought to use the sun to produce images on light-sensitive paper and leather. Although he was successful in producing silhouettes, he did not have a way to stop the action of light on his papers, resulting in his silhouettes gradually darkening, and eventually fading away and disappearing. The primary light-sensitive material he reported using was silver nitrate[37] as well as silver chloride, which he stated had no advantages over silver nitrate. He also noted that the images obtained in a *camera obscura* were too faint to be recorded with his sensitive paper.[38]

Although Wedgwood was the first to unite the two main ingredients of photography—the darkening of silver salts in light and the image of the *camera obscura*—it was others such as Niépce, Daguerre, Talbot (among others) who filled in the fine details in the ever-complex tapestry of ideas that

---

[34] M.S. Barger and W.B. White *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 16 (1991).

[35] H. Gernsheim and A. Gernsheim, *The History of Photography: From the Camera Obscura to the Beginning of the Modern Era* **1**, Thames & Hudson, London (1969).

[36] According to M.S. Barger and W.B. White in Ref. 34, there are conflicting reports as to when Wedgwood stopped his experimentation because of ill health. They cite John Werge, who noted that Wedgwood had given up his activities as early as 1792 [J. Werge, *Evolution of Photography with a Chronological Record of Discoveries, Inventions, Etc.*, Piper and Carter, London, p. 9 (1890)]. They also cite Wedgwood's biographer, R.B. Litchfield, who places the date later, in 1799 [R.B. Litchfield, *Tom Wedgwood, The First Photographer*, Duckworth and Co., London (1903)].

[37] T. Wedgewood and H. Davy, "An account of a method of copying painting on glass and of making profiles by the agency of light on nitrate of silver, invented by T. Wedgwood, Esq., with observations by H. Dave," *J. Royal Institution* (London) **1**, 170 (1802).

[38] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 16 (1991).

eventually materialized as photography. Apparently, it never occurred to any of the great number of artists of the 17th and 18th centuries who were using the *camera obscura* to try to fix its image permanently. It was only when scientists such as those mentioned above wanted to make pictures and turned to the *camera obscura* for assistance that the sciences of chemistry and optics were at last successfully married and united in photography.[39]

Perhaps aware that the solution to producing a permanent photographic image lay in arresting the action of light on the photochemical reactions of the silver nitrate, the pioneers of photography sought ways to accomplish just that. An account of their enormous contributions to the invention and development of photography, although a result of an evolution rather than any sudden discovery by any one of them, is presented below. For, given the knowledge of the same principles, it is only natural to expect that people working quite separately and without knowledge of each other should generate the same invention at about the same time, and this is in fact borne out by the numerous different processes for which claims of priority were made in various parts of Europe immediately after the announcement by François Arago of Daguerre's discovery on 7 January 1839.[40] An excellent account of the events surrounding the invention of photography has been provided elsewhere;[41] here, we will summarize the main events that are pertinent to the invention of photolithography.

2.4.1 Joseph Nicéphore Niépce: the inventor of photography and photolithography

The singular honor for the invention of photography as well as photolithography goes to Joseph Nicéphore Niépce (1765–1833) (see Fig. 2.4), who was born in Chalon-sur-Saône, France to a wealthy bourgeois family. Although trained for the priesthood, he, like many other members of religious orders, was forced to abandon his profession during the French Revolution. In 1792 he joined the army, but his career as a soldier was cut short because he contracted yellow fever. In 1794, he moved to Nice, where he married and became a petit bureaucrat. In 1801 he and his family returned to the family estate in Gras outside of Chalon-sur-Saône. His older brother, Claude, joined them a little later, where together, they occupied themselves with inventions.[42]

Their first major invention was the *pyréolophore*, a boat with an internal combustion engine that was patented in 1807. Their efforts to market and sell the *pyréolophore* lasted well over 20 years and took Claude to Paris in 1816

---

[39] H. Gernsheim and A. Gernsheim, *The History of Photography: From the Camera Obscura to the Beginning of the Modern Era*, Thames & Hudson, London, pp. 1–2 (1969).

[40] ibid.

[41] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, Chapter 3 (1991).

[42] ibid., pp. 17–20.

Figure 2.4 Portrait of Joseph Nicéphore Niépce, the inventor of photolithography (printed with permission from the Deutsches Museum, Munich).

and then to England in 1817. They also discovered new methods for the cultivation and extraction of dye from the woad plant, for which they received recognition from the French government. Indigo had become scarce and expensive as a result of the continental blockade during this time, while the demand for blue dye for military uniforms did not diminish in equal measure. In 1817 Joseph Nicéphore Niépce also became an inventor of a machine that is the forerunner of the bicycle.[43]

In 1813 the Niépce brothers shifted their focus to lithography. The first patent for the lithographic process had been granted in France in 1802, and the process was then primarily used for printing music. It was during Napoleon's occupation of Bavaria that the French most likely first came into contact with lithography. Although French interest in lithography did rise during the period of Napoleon's occupation of Bavaria, the growth of the technique in France was hindered because of the political situation in the country, which lasted until after the restoration of King Louis XVII to the throne in 1815. Following the restoration of the King to the throne, French lithography took off in popularity, so much so that by the end of 1817, the government, alarmed by the spread of the new technology and the relative ease of making broadsides with it, placed restrictive regulations on the process for reasons of state security. Despite these restrictions, the French became the leading practitioners of the art by 1820.[44]

---

[43] ibid., p. 17.

[44] ibid., p. 18.

Another factor that led to the popularization of lithography in France and that was also an incentive for the Niépce brothers was the institution of premiums by the *Société d'Encouragement des Arts et Métiers* for improvements to the lithographic process. From the time of the introduction of lithography in France until well into the 1840s, yearly premiums were consistently offered for devising better methods of transferring designs to lithographic stones, finding a suitable substitute for the German limestone traditionally used as lithographic stones, and producing materials to ink lithographic stones. The *English Society of Arts* also offered premiums throughout the 1820s and 1830s.[45]

Motivated by the premiums from the *Société d'Encouragement*, during the early course of his research on lithography, Joseph Nicéphore Niépce wedded the two objectives of improving both lithography and methods of transferring designs to stone. For several years, he worked sporadically on various aspects of lithography, the first fruit of which was outlined in a letter he sent to the *Société d'Encouragement* with local stone samples suitable for lithography found in the region of Chalon-sur-Saône in September 1816. Much to his relief, these stones were judged by the *Société d'Encouragement* to be of sufficient quality to use for lithography.[46]

Also around 1816, Joseph Nicéphore Niépce was successful in capturing images using a *camera obscura* and silver chloride papers. However, like Thomas Wedgwood before him, he did not find a way to stop the action of the light, nor did he overcome a second serious problem: the images he produced were negative mirror images, reversed in both tone and geometry, of the scenes captured by his camera.[47]

Discouraged with the results he obtained on paper, Joseph Nicéphore Niépce abandoned that avenue of investigation and turned to producing images on stone, metal plates, and glass. He experimented with various resin and varnish concoctions (such as gum guaiacum, which turns color and has altered solubility properties when exposed to sun) to produce these images. He also searched for materials that would bleach rather than darken in the sun, so that the image created by the camera could be viewed directly in positive tone. Further, he reasoned that if a method that was more along the lines of printing could be devised, it would not matter that the images recorded by the camera were totally reversed because they would be reversed again in the printing process. And very significantly, with such a method, it would be possible to obtain multiple copies from an original.[48]

At last, Joseph Nicéphore Niépce made the fundamental leap when he understood that these plates could be objects in and of themselves, not just intermediaries to something else. He successfully copied engravings onto

---

[45] ibid., p. 18.

[46] ibid., p. 18.

[47] ibid., p. 18.

[48] ibid., pp. 18–19.

light-sensitive plates. He oiled these engravings to make them transparent and then laid them on lithographic stones or glass plates coated with his various concoctions of light-sensitive varnishes. After sufficient exposure, he treated the plates with various solvents that washed away the unexposed portions of the varnishes. Using acid, he was then able to etch away the clear areas of the plate that were not protected by the light-sensitive varnish. Once the plate was etched and stripped of all varnish, he could then use it for printing.

In an 1824 letter to his brother Claude, Nicéphore wrote that he had succeeded in obtaining direct images on stone and glass that then could be etched and used for printing. It turned out that he had spoken too soon about direct images, but success in this area, nevertheless, did not elude him for long.[49]

About that same time, Nicéphore began to experiment with pewter plates. Possessing a lighter and more natural color than the copper and zinc plates he had been using, pewter allowed him to attain a greater range of tones. In 1826, using a polished pewter plate to which he coated a thin film of light-sensitive substance called asphalt or bitumen of Judea that was exposed to an image of his courtyard and garden from his window with the aid of a simple *camera obscura*, he obtained the first successful direct camera image: an image of the courtyard of his country estate *Le Gras* in the village of St. Loup-de-Varennes, near Chalon-sur-Saône in France (Fig. 2.5). After dissolving the unexposed

Figure 2.5 The first photograph from nature (*Point de vue du Gras* or View from the window at Le Gras), taken in 1826 by Joseph Nicéphore Niépce of the courtyard in his country estate *Le Gras* in the village of St. Loup-de-Varenness, near Chalon-sur-Saône in France. This image was recorded on a polished pewter plate using heliography. (Reprinted with permission from the Harry Ransom Humanities Research Center, The University of Texas at Austin.)

---

[49] ibid., pp. 18–19.

parts of the photograph in turpentine oil and rinsing the plate, he obtained, without the need for any other fixing, a permanent bitumen image of the light drawing, in which the shadows were indicated by the bare pewter plate. To prevent a lateral reversal of the image, Niépce had employed a prism in front of his achromatic lens, both of which he had obtained from the Parisian optician Charles Chevalier when he purchased his first professional camera in January of 1826. He named his invention *heliography* and the direct camera image *point de vue du Gras* (view from the window at Le Gras).[50] This invention is a major milestone that marks the beginning of photography and photolithography.

This first photograph was in effect a projection printed lithographic image,[51] taken no doubt on a fine day since the image shows both sides of the courtyard to be illuminated. This is because the exposure took 8–10 hours, during which time the earth had partially rotated around its axis in front of the sun.

Bitumen of Judea is a tarlike petroleum substance that has been recovered in chunks from the Dead Sea since Egyptian times and has been used for embalming mummies, as a building material, and as a black varnish material. It has the peculiar property of softening when heated but hardening when exposed to light—essential properties of materials we call today negative-type photoresists. Earlier, the Niépce brothers had discussed using this material as part of a possible fuel mixture for the *pyréolophore*. Nicéphore had discovered that bitumen of Judea is soluble in lavender oil and that high-quality coatings could be made from the solution. He used the material to coat a variety of substrates, but his best work was done on polished pewter.[52]

As mentioned above, Nicéphore developed his image with a mixture of lavender oil and kerosene. The bitumen of Judea is rendered less soluble in the mixture of lavender oil and kerosene (it hardens) upon exposure, so during development, the unexposed parts were washed away, revealing to a varying degree the surface of the pewter. The photolithographic processes involved in

---

[50] H. Gernsheim, "The 150th anniversary of photography," *History of Photography* **1**(1), 3–8 (1977); M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 19 (1991).

[51] H. Gernsheim and A. Gernsheim, *The History of Photography: From the Camera Obscura to the Beginning of the Modern Era*, Thames & Hudson, London (1969); J.L. Mangnier, "Asphalt as the world's first photopolymer—revisiting the invention of photography in processes in photopolymer," V.V. Krongaus and A.D. Trifunac, Eds., Chapman and Hill, New York (1995) [cited in C.G. Willson, R.A. Dammel, and A. Reiser, "Photoresist materials: A historical perspective," *Proc. SPIE* **3050**, 38–51 (1997)]; H. Gernsheim and A. Gernsheim, *The Photographic Journal*, Section A, 118–120 (1952) [cited in C.G. Willson, R.A. Dammel, and A. Reiser, "Photoresist materials: A historical perspective," *Proc. SPIE* **3050**, 38–51 (1997)].

[52] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, pp. 19–20 (1991).

the imaging thus effectively created a relief structure in the remaining organic media.[53]

Two seminal events occurred in 1825 immediately after Nicéphore had made his fundamental invention in photography and photolithography that would alter the course of his work. First, he established a working relationship with the Parisian lithographer Augustin François Lemaître, to whom, for the next several years, he sent lithographic plates, which the former etched and printed. Second, that same year, a relative of Nicéphore, while visiting the opticians Charles and Vincent Chevalier of Paris to purchase a good camera to take back to Gras, mentioned that Nicéphore had succeeded in producing permanent images using a camera and showed the Chevaliers examples of his work. Quite impressed with the work, they mentioned it to another customer, Louis Jacques Mandé Daguerre, who was also interested in recording images made using a camera. Almost as soon as Daguerre heard of Nicéphore's work, he resolved to make contact with him and sent a letter in January 1826 requesting information about it. After receiving two more letters from Daguerre, Nicéphore relented and sent one of his heliographs on pewter and a print pulled from the plate to the former.[54] The two men later formed a partnership to promote and develop Nicéphore Niépce's process, which eventually led to the invention of daguerreotype process, the forerunner of modern-day photography.[55]

September 1827 saw Nicéphore take a trip to England to see his brother Claude, who was very ill and probably dying. The trip brought grave disappointments, not only because of Claude's ill health, but also because Claude had made no progress in promoting the *pyréolophore*. For more than thirty years, the brothers had devoted the family fortune to promoting this machine, and it was during this trip that it became clear to Nicéphore that there would be no return on these investments—the family fortune was lost.[56]

While in England, Nicéphore tried to promote his latest invention, *heliography*. He made the acquaintance of Francis Bauer, a prominent botanical illustrator at the Royal Botanical Gardens at Kew and a fellow of the Royal Society, to whom he showed some examples of his heliographs, including the view taken from the window at Le Gras. Bauer encouraged Nicéphore to prepare a memoir about heliography to be presented to the Royal Society, but Nicéphore was reluctant to do this because he did not want

---

[53] H. Gernsheim, "The 150th anniversary of photography," *History of Photography* **1**(1), 3–8 (1977); C.G. Willson, R.A. Dammel, and A. Reiser, "Photoresist materials: A historical perspective," *Proc. SPIE* **3050**, 38–51 (1997) [doi: 10.1117/12.275984].

[54] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science* Johns Hopkins University Press, Baltimore, p. 19 (1991).

[55] ibid.

[56] M.S. Barger and W.B. White *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, Chapter 3 (1991).

to divulge his process publicly without some financial arrangements. His descriptions of the process were intentionally vague in order to prevent others from stealing his ideas. This secretiveness, however, ran counter to the Royal Society's rules of scientific disclosure, so his memoir was not presented to this organization. He also tried to gain the royal patronage of George IV, but this plan failed as well. He finally approached the Society of Arts, hoping to find some interest in his process there, but was not successful with them, either. He was very disappointed with this lack of interest in his invention, even in England, because he thought that he had a better chance of exploiting his new process there than in France. He returned home disheartened, leaving with Bauer his memoir along with some samples of heliographs, including the view of his courtyard and garden, the first successful photograph ever taken.[57]

When he started working again, he began to use polished, silver-plated copper plates for two main reasons: (1) silver as a whiter metal than pewter would increase the number of tones in his heliographs and, (2) copper made a more suitable plate for printing. He quickly discovered that he could make another kind of image on a silver plate by treating an image made in bitumen of Judea with iodine vapor; this led to blackening of the uncovered areas of the plate. After removing the bitumen, a positive silver iodide image was left on the plate.[58]

2.4.2 Louis Jacques Mandé Daguerre

As mentioned earlier, in December of 1829, Nicéphore Niépce and Louis Jacques Mandé Daguerre (see Fig. 2.6) formed a partnership to promote and develop Nicéphore Niépce's process. Niépce subsequently presented a complete description of the process to Daguerre, who, using it as a foundation, advanced and developed the daguerreotype process. He discovered that silver iodide, which can be made by fuming silver plates with iodine just as Nicéphore Niépce had done, was sensitive to light and that, on its own, it could be used to make images. He also made the discovery that if an exposed plate with no apparent image was exposed to mercury vapor, a latent image appeared. With this method, he discovered the latent image. Unfortunately, Nicéphore Niépce did not live to see any of Daguerre's results. He died in 1833, and Daguerre came upon these improvements in 1835. At this point Daguerre still could not produce permanent images, but in 1837 he finally made the discovery that he could stop the action of light by

---

[57] These items were shown several times during the 19th century but were thought to be lost since 1898. In 1952, through the efforts of Helmut and Alison Gernsheim, who have done extensive work in chronicling the history and development of photography, the heliographs were found in a trunk that had belonged to the Bauer family and had been in storage since 1917.

[58] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, Chapter 3 (1991).

Figure 2.6 Portrait of Louis Jacques Mandé Daguerre, the inventor of daguerreotype, which was the forerunner of modern photography (printed with permission from the Deutsches Museum, Munich).

putting his plates in hot, saturated salt water. Daguerre called his new process *daguerreotype*,[59] and it is this particular process that is the forerunner of modern-day photography.

When news of the processes of the two partners reached the physicist François Dominque Arago (1786–1853)—who was at the time not only the director of l'Observatoire in Paris and the permanent secretary of the Académie des Sciences, but also a member of the Chamber of Deputies in the French Parliament—the fortunes of the two partners took a turn for the better. Being interested in light and its properties, Arago had an immediate personal interest in these processes. He particularly felt that daguerreotype could be used as a tool to unlock some of the mysteries of light, and that it would also be useful as a recording medium, not only for artists and architects, but also for scientists, archeologists, and travelers. In addition, he had the vision to foresee that this process would have an impact far beyond the mere fact of being able to record images. He thus set about to see that Daguerre and Nicéphore Niépce were duly rewarded and that this discovery brought both honor and glory to France. To this end, a brief notice about the daguerreotype process was published in the *Comptes Rendus*, the journal of the Académie des Sciences. The notice merely stated that a process had been devised and that details would be forthcoming.[60]

---

[59] ibid.

[60] ibid.

This notice generated a great deal of public interest throughout Western Europe and brought about two particular responses. First, Francis Bauer sent a letter to *The Literary Gazette* describing Nicéphore Niépce's work and offering to show the heliographs that he had given to him 11 years earlier. Bauer wanted to make sure that his friend Nicéphore Niépce was accorded the recognition due to him for his invention. Second, about this time William Henry Fox Talbot published an account of his process for making images with light in order to establish his priority as the first discoverer of such a process.[61]

2.4.3 William Henry Fox Talbot

The scientist William Henry Fox Talbot (1800–1877) (see Fig. 2.7) made his first experiments by imaging with papers prepared with silver nitrate, which he found to be not very sensitive. He then turned to silver chloride paper and soon discovered that he could stop the action of light on his sensitive papers by soaking them in a strong solution of salt water, just as Daguerre had done earlier. During the summer of 1835 Talbot was successful in making images of his house using his sensitive paper and a *camera obscura*, although with extremely long exposure time (on the order of several hours). Sometime during this period he also experimented with a solar microscope in conjunction with his sensitive paper, which resulted in the first

Figure 2.7 Portrait of William Henry Fox Talbot (printed with permission from the Deutsches Museum, Munich).

---

[61] ibid.

photomicrographs ever taken. As the publicity surrounding Arago's announcement mounted even in England, Talbot had to concede that the discoveries of Niépce and Daguerre did indeed take place prior to his own.[62]

The publicity surrounding the discoveries of Niépce, Daguerre, and Talbot also propelled others who had been thinking along the same lines to publicize their own discoveries. In May 1839, Mungo Ponton (1802–1880), a member of the Edinburgh Society of Arts, published a process based on dichromate salts, which he believed was a variant of Talbot's.[63]

Thus, the period between 1826 and 1839 saw the birth of a new era—the world witnessed the introduction of photography, which used several processes to capture permanent images using camera. None of these discoveries was the result of a concerted scientific research program based on the systematic application of knowledge of light-sensitive materials to the problem of imaging. Rather, the discoveries were the result of three men who were interested in solving problems: one was an inventor (Niépce) trying to improve lithography by competing for premiums offered by the Société d'Encouragement des Arts; one was an artist and promoter (Daguerre) trying to fulfill the need of capturing the images in his camera; and one was a gentleman scientist (Talbot) trying to make up for his inability to draw as well as an artist. These men all drew on knowledge that was widely available during their time. They applied that knowledge in new ways to devise solutions to obstacles that had earlier prevented others from solving the problem. All three were able to produce images that had relatively short exposure times (i.e., on the order of minutes); they were able to make these images permanent by stopping the action of light with some subsequent treatment.[64] For these contributions, these three men are the founding pioneers and fathers of photography. While Daguerre's and Talbot's contributions to photography are significant, the priority for the invention of photography and photolithography belongs to Niépce. Everything we do today in offset lithography as well as in optical lithography, and for that matter in other fields of semiconductor lithography, derive from Niépce's invention.

---

[62] ibid.

[63] M. Ponton, "Notice of a cheap and simple method of preparing paper for photographic drawing, in which the use of any salt of silver is dispensed with," *Edinburgh New Philosophical J.* **27**, 169–71 (1839).

[64] M.S. Barger and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 27 (1991).

Chapter 3

Physical Origins of Lithography

> "Science makes people reach selflessly for truth and objectivity; it teaches people to accept reality, with wonder and admiration, not to mention the deep awe and joy that the natural order of things brings to the true scientist."[1]
>
> Lise Meitner (1878–1968)

> "To understand a science, it is necessary to know its history."[2]
>
> Auguste Comte (1798–1857)

3.1 Introduction

As stated in Chapter 1, the so-called information superhighway is paved with chips of crystalline silicon by lithography, for it is the key technology process step with which we are able to create fine-patterned structures that ultimately form the devices and interconnects of complex functional integrated circuits (ICs) used in microelectronics. As a matter of fact, both circuit speed and integration density depend on the size of the lithographic minimum-printable feature. Today lithography represents over 35% of the chip manufacturing cost in the semiconductor industry. The increasing cost of lithographic exposure tools is recognized as one of the most critical factors for keeping up with Moore's law.[3]

---

[1] L. Meitner from a 1953 speech in Vienna to the Austrian UNESCO Commission [cited in: R.L. Sime, *Lise Meitner: A Life in Physics*, University of California Press, Berkeley p. 375, (1996)].

[2] A. Comte, *The Positive Philosophy of Auguste Comte*, 3 volumes (H. Martineau, translator), George Bell & Sons, London (1896) [reissued by Batoche Books, Kitchener vol. 1, p. 47 (2000)].

[3] G.E. Moore "Lithography and the future of Moore's law," *Proc. SPIE* **2437**, 2–17 (1995) [doi: 10.1117/12.209151].

Although lithographic patterning is not the only critical step in IC device manufacture, its specifications are by far the most stringent and also the most consequential. This arises not only from the requirements on a given lithographic patterning step for tight dimensional control and absence of defects, but also from the sheer number of lithographic patterning steps that a modern device structure undergoes. It is not uncommon for many advanced IC devices to pass through fifty or more lithographic patterning steps in the course of fabrication. Under such conditions, the need for precise, reliable, and reasonably low-cost lithographic solutions is imperative.

It is therefore only reasonable to construct the physical history of the developments that led to the emergence of lithography as the preeminent critical step in the fabrication of ICs and the mass production of printed matter like newspapers. But in order to arrive at a fuller understanding of the evolution of lithography, it is necessary to examine the work of a large number of scientists whose contributions to physics, spanning a period of over 3000 years, were instrumental in shaping the invention of lithography in 1798 by Senefelder[4] and its subsequent development and progress. This story is really not only a triumph of advances in the understanding of materials science, but is also a product of the knowledge gained in the convergence of major areas of chemistry and physics (particularly those related to optical physics, semiconductor physics, and atomic physics).

Like every other technological innovation, lithography is "like a river—its growth and development depending on tributaries and on conditions it encounters on its way. The tributaries to [such] an innovation are inventions, technologies and scientific discoveries; the conditions are the vagaries of the market place."[5]

There are several special features to the innovation that we might call lithography. First of all, it is not really a single innovation at all, but consists of a long series of linked innovations. There is a continuous chain of lithographic patterning techniques based on the simple idea that oil and water repel each other; this chain stretches from Senefelder's stone plate lithography of 1798 to Niepce's visible-light lithography of 1826, to offset lithography, to the most advanced semiconductor lithography of today. The major techniques that have evolved are extensions of photolithography. Shorter-wavelength photons and even electrons have successively been used in a quest to overcome the diffraction limitations imposed by longer-wavelength photons. Just about every conceivable radiation with wavelengths shorter than a micrometer has been investigated as a potential

---

[4] A. Senefelder, *Vollständiges Lehrbuch der Steindruckerei* [*A Complete Course in Lithography*, in German], Karl Thienemann, Munich and Karl Gerold, Vienna, Publishers (1818).

[5] E. Braun and S. MacDonald, *Revolution in Miniature: The History and Impact of Semiconductor Electronics*, Cambridge University Press, Cambridge, UK, Chapter 1 (1978).

exposure radiation. To date, only five have proven successful enough to be generally useful in advanced lithography: near-ultraviolet photons, deep-ultraviolet photons, electrons, soft x-rays, and ions. Although deep-UV lithography uses some new materials and techniques, it is more of an evolutionary development from visible-light lithography. Electron-beam lithography, x-ray lithography, and ion-beam lithography, on the other hand, employ physical concepts, techniques, and equipment that are comparatively new to microfabrication.[6]

The totality of these lithographic innovations and the innovations in semiconductor electronics that they have engendered are so massive as to make most single technological innovations pale in comparison. Apart from the scale of the innovations and the scale of its impact, lithographic technology and its associated semiconductor technology are distinguished by their great dependence on science. In more ways than one, lithography, particularly semiconductor lithography and modern electronics, which it enables, owe their very existence to science; they are both truly innovations based on science.

In the remainder of this chapter, our object is to consider how key theoretical and technological developments in physics have influenced the development of lithography. Further, we will highlight specific applications of physical theories and physical phenomena such as reflection, refraction, polarization, interference, and scattering to lithography—be they in the design and fabrication of optical elements of lithographic exposure tools, in the development of lithographic exposure sources, in metrological measurements of lithographically patterned features, in precise interferometric distance measurements in alignment of the lithographic exposure tool stage, in the interaction of radiation with lithographic materials, in lithographic image formation and process modeling, etc. A summary of these developments in physics is given in thematic order, but not necessarily in chronological order, in Fig. 3.1.

Since lithography as practiced in its many varied incarnations today involves essentially a series of chemical transformational processes mediated at some stage by some type of radiation, usually light, a narrative that traces the development and evolution of thought and ideas on the nature of light and chemical phenomena that ultimately went into the invention of and subsequent development of lithography is judged a good means for presenting this fascinating story. Along this journey we will examine fundamental theories on the nature of light and matter, spanning the earliest known human reflections on these subjects to modern quantum theory, while exploring how the marriage between chemistry (discussed in Chapter 4) and physics led to the

---

[6] N.D. Wittels, "Fundamentals of Electron and X-ray Lithography," in *Fine Line Lithography*, R. Newman, Ed., North-Holland Publishing, Amsterdam, pp. 1–104 (1980).

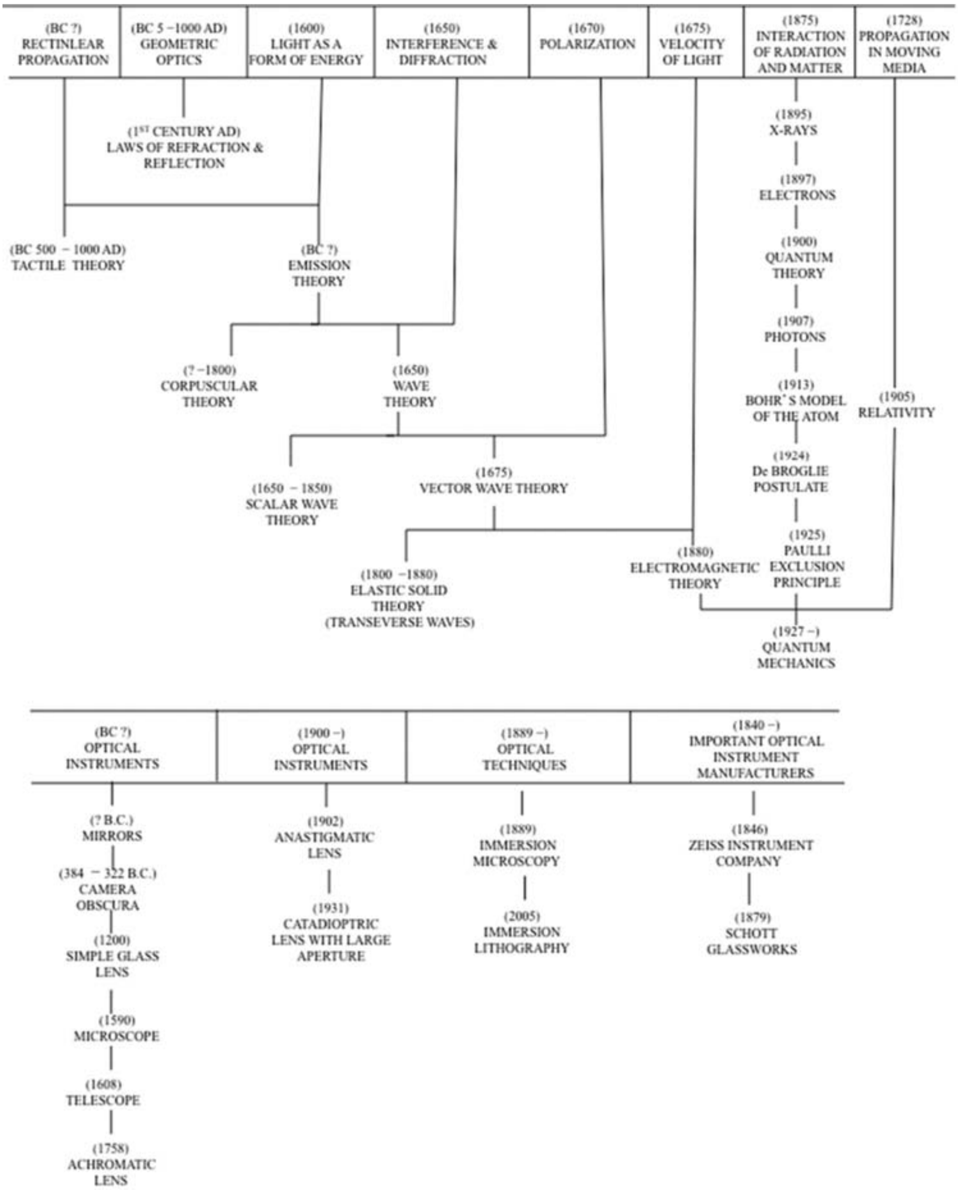


Figure 3.1 Key developments in physics and optical instruments manufacture that led to the invention and development of lithography.

invention and development of lithography. Furthermore, we will examine key developments in the optical instrument and glassmaking technologies—key critical technologies that enabled the development of instruments used in modern lithography. The whole story is one of fascinating science that has transformed human life.

3.2 Key Developments in Optical Physics that Enabled the Invention and Development of Lithography

A comprehensive treatment of the development of optical physics covering, among other topics, the theories of light has been given elsewhere;[7] here only the key developments that directly or indirectly influenced the development of lithography as practiced today are covered. In such a treatment, it is convenient to disregard the chronological order in which the developments occurred, but rather to present the materials from a historical point of view and to show how each of the developments has been incorporated into the development and practice of lithography. A summary of these developments is given in Fig. 3.1.

3.2.1 Tactile and emission theories of light

Historically, it has been known since antiquity that the earliest speculations on the nature of light were theories of vision. One school held that the eye sends out invisible antennae, or sensitive probes, or visual rays, and is thus able to feel objects that are too distant to be touched by hands or feet. This theory was called the "tactile" theory. An alternative hypothesis held that light consists of small particles that are emitted by bright objects and that, upon entering the eye, are able to affect some sensitive part of the eye and so give rise to the sensation of sight. This theory was called the "emission" theory. Both of these theories were current among Greek thinkers about 500 BC.[8]

In its inherent simplicity, the tactile theory is able to describe the unknown in terms of the known. The mystery of vision is directly related to the simpler and more obvious sense of touch. This theory does, however, experience some difficulty in explaining why things can be felt, but not seen, in the dark, and why bodies can be made visible in the dark by heating them. The fact that certain bright bodies are able to make neighboring bodies visible is also not explained by the tactile theory.[9]

Some of these difficulties of the tactile theory can be explained away by postulating that the visual probes are able to feel only certain kinds of surfaces, and then making a series of assumptions that surfaces can be modified under various conditions. But this approach only ends up making the theory intolerably complicated, as the simple sense of relation to the sense of touch—the very important attribute of the theory—is lost. These conflicts

---

[7] R.W. Ditchburn, *Light*, Dover Publications, Inc., New York (1991); P. Mason, *The Light Fantastic*, Penguin Books, New York (1981); M.I. Sobel, *Light*, University of Chicago Press, Chicago (1987).

[8] Many variations and combinations of these theories were also postulated. See, for example, R.W. Ditchburn, *Light*, Dover Publications, Inc., New York, p. 4 (1991). We need not consider them here since they are more complicated than the two theories we have described and have no significant compensating advantages.

[9] R.W. Ditchburn, *Light*, Dover Publications, Inc., New York, p. 4 (1991).

can be resolved in a simple and satisfactory way by the emission theory if it is assumed that some bodies are able to emit a radiation to which the eyes are sensitive, and that others are able to reflect or scatter this radiation so that it enters the eye.[10] For these and similar reasons, the tactile theory was gradually superseded by and eventually replaced with the emission theory. The process for this transformation was, however, very slow, and it was not until about 1000 AD that, under the influence of the Arabian astronomer Alhazen (discussed below), the tactile theory was finally abandoned.[11]

3.2.2 Early studies in optics and catoptrics

Another major notable development in the early studies of light phenomena is attributed to Euclid,[12] the great Greek mathematician of the 3rd century BC, who wrote two books on light, *Optics* and *Catoptrics* (dealing with reflection). Little is known of the latter, but a review of it, written perhaps many centuries later—and possibly spurious—discusses the properties of mirrors and provides an accurate law of reflection (see Fig. 3.2), which states that a ray is reflected from a plane surface at the same angle with which it strikes the surface. Euclid was certainly aware that light travels in straight lines, a fact that he employed in his discussions on the laws of perspective.[13]

3.2.2.1 Refraction

Another significant contribution to the early studies of light phenomena was due to Claudius Ptolemy, who made the first systematic study of refraction. In the second century AD in Greek Alexandria, he developed his cosmology in which Earth is at the center of the universe with the Sun and the planets orbiting around it. This cosmology reigned undisputed for 13 centuries. Although the Ptolemaic system was actually a compilation of works of earlier Greek astronomers, notably Hipparchus, Ptolemy recognized that to correctly determine the location of a planet he must take into account the bending of light as it enters the earth's atmosphere. He demonstrated that when light enters a dense medium from a less dense one, the ray is bent towards the

---

[10] A modern interpretation of the emission theory defines light as "visible radiation," which takes part in the following general account of the visual process: Light being emitted, reflected, or scattered enters the eye and is focused by the lens of the eye on the retina, located at a surface situated at the back of the eye. The retina contains a large number of nerve endings, which upon receiving light transforms the latter into electrical impulses through some chemical and physical action. A series of electrical impulses is thus sent from the retina along an appropriate nerve fiber to the brain [R.W. Ditchburn, *Light*, Dover Publications, New York, p. 5 (1991)].

[11] R.W. Ditchburn, *Light*, Dover Publications, New York, p. 4 (1991).

[12] Greek science reached its peak in Euclid's century. Archimedes of Syracuse was the greatest scientist of the era, and he made advances in mechanics, hydrostatics, and mathematics [M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 2 (1987)].

[13] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 2 (1987).

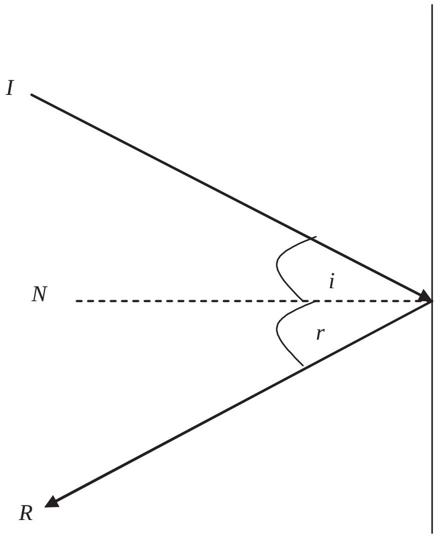


Figure 3.2 The law of reflection. An incident ray of light *I* upon hitting a mirror surface is reflected *R* such that angle of incidence $i = $ angle of reflection *r*. The dashed line *N* is the normal to the surface. The angle of incidence is the angle between the incoming ray and the normal; the angle of reflection is the angle between the reflected ray and the normal. *I, N*, and *R* are in the same plane.

normal, and conversely, it is bent away from the normal when entering a less dense medium.[14]

Ptolemy also studied air–water, air–glass, and water–glass interfaces, and made tables of the *angle of incidence* (the angle between the incoming light ray and the normal) and the corresponding *angle of refraction* (the angle between the refracted ray and the normal). He postulated, however incorrectly, that for a given interface the two angles were proportional. His *Optics*, only part of which has survived to the present day, is noteworthy not so much for its data as for the very fact of its adherence to the experimental method.[15]

The next major figure in the story of light lived more than eight centuries later, during the height of Islamic civilization.[16] Born in 965 AD, Alhazen[17] experimented and wrote extensively on optics. In his writings on optics, he used highly developed algebra and geometry, as well as an experimental

---

[14] ibid., pp. 2–3.

[15] ibid.

[16] By the year 750 AD, Islamic soldiers fanning out from Arabia had overrun many lands around their region and built an empire along the entire Mediterranean Sea and as far east as India. Among the great thinkers of this era were al-Khwarizmi the mathematician, Avicenna the physician, Averroës the philosopher, and Alhazen the physicist [M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 5 (1987)].

[17] Alhazen is actually the Latinized version of the great Arabic scholar Abu Ali al-Hasan ibn al-Hasan al-Haitham, born in 965 AD.

approach in a very modern sense, involving the use of sighting tubes, strings, and plane and curved glasses and mirrors to study the laws of reflection and refraction. He refuted Ptolemy's claim that the angle of incidence is proportional to the angle of refraction, although he himself did not obtain the correct mathematical relation. He showed that a convex lens (where each surface is a part of a sphere) can magnify an image. He carried out a rigorous mathematical treatment of reflection from spherical, cylindrical, and conical mirrors. He observed that if parallel rays of light strike a curved mirror that is a section of a sphere, the rays are not brought to a precise focus, a condition known as *spherical aberration*; but a mirror in the shape of a paraboloid (a solid figure produced by rotating a parabola about its axis) can produce a sharp focus.[18] Furthermore, he deduced that twilight, the persistence of daylight after the sun has set, was due to the refraction of sunlight from the upper layers of the earth's atmosphere. And by assuming that twilight ends when the sun's rays are refracted from the very top of the atmosphere, he deduced that the atmosphere was 20 to 30 miles high, a fairly good estimate by modern standards.[19]

Also in Alhazen's writings we find reports of very detailed early dissection of the human eye. Additionally, we find speculations on the method of propagation of light that anticipates the 17th century theory of Huygens. And we find hints, or rather the suggestion, demonstrated only in the 19th century, that light travels less easily (i.e., at lower velocity) when it enters a dense medium, and that this causes the ray to be bent toward the normal.[20]

It should be pointed out that the research and scholarship of the Muslim scholars were important not only in themselves, but also on account of the fact that they transmitted, in Latin translation, the spirit of learning to medieval philosophers like Roger Bacon and Albertus Magnus, whose work foreshadowed the age of science in the West. Bacon was certainly familiar with Alhazen's works in optics, understood how to trace rays of light through lenses and mirrors, and may very well have been the first to use a lens for spectacles. He also suggested combining two lenses to make a telescope, although it is not certain that he actually built one.[21]

3.2.2.2 Invention of the microscope and telescope

Next, our journey through the history of studies in light phenomena takes us to the age of Copernicus, specifically, the invention of the microscope by Zacharias Janssen around 1590, the invention of the telescope by Hans Lippershey in 1608, and its quick deployment by Galileo Galilei.[22] The

---

[18] M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 4–5 (1987).

[19] ibid.

[20] ibid.

[21] ibid.

[22] ibid., p. 5.

modern-day optical and electron microscopes that are widely used in the inspection of lithographically patterned wafers all trace their origins to Janssen's invention.

3.2.2.3 Laws of refraction

Around 1621, the Dutch scientist Willibrod Snell discovered the correct law of refraction, which both Ptolemy and Alhazen were unable to deduce.[23] By applying heuristic momentum conservation arguments in terms of sines, René Descartes independently derived the law in his 1637 philosophical and mathematical treatise, *Discourse on Method*. Descartes was able to solve several optical problems with the aid of this law.

Snell's law provides the relation between the angle of incidence and the angle of refraction (see Fig. 3.3) in terms of a quantity called the *index of refraction*, which is characteristic of the medium into which light travels. Specifically, the ratio (*sine of the incident angle*) / (*sine of the refracted angle*) equals the *refractive index* of the medium. According to modern wave theory, the refractive index is the ratio of the speed of light in vacuum to the speed of light in the medium.[24] Expressed another way, Snell's law states that the ratio of the sines of the angles of incidence and refraction is equivalent to the ratio of velocities in the two media, or is equivalent to the inverse ratio of the indices of refraction:

$$\frac{\sin i}{\sin t} = \frac{v_1}{v_2} = \frac{n_2}{n_1}, \tag{3.1}$$

or

$$n_1 \sin i = n_2 \sin t, \tag{3.2}$$

where sin i is the sine of the angle of incidence, sin t is the sine of the angle of refraction, v_1 and v_2 are the velocities of the light ray in medium 1 and 2, respectively, and n_1 and n_2 are the refractive indices of the light ray in medium 1 and 2, respectively.

Not long after the formulation of Snell's law, the French mathematician Pierre de Fermat unified the laws of reflection and refraction by showing that both could be deduced from the hypothesis that light travels a path of least time. In other words, given two points A and B in a region with mirrors or with different media, the path of a ray of light from point A to point B will be that for which the time of travel is least. The implication of Fermat's principle is that light travels at a finite speed.[25]

---

[23] ibid.

[24] ibid.

[25] ibid.

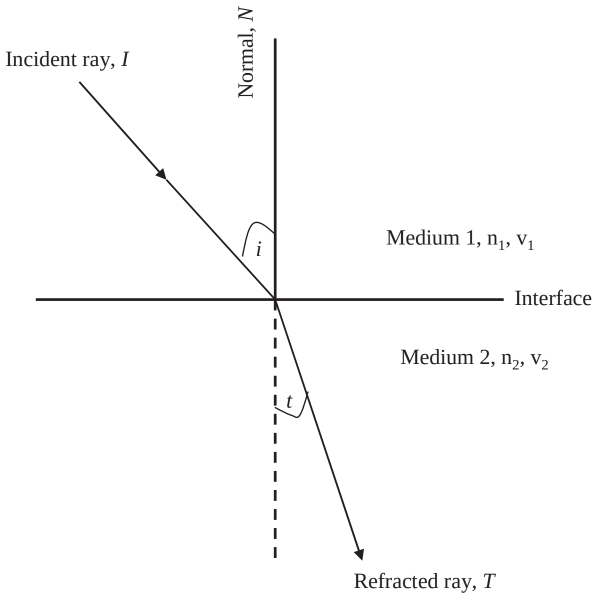


Figure 3.3 Refraction of light at the interface between two media of different refractive indices, with n_2 greater than n_1. An incident light ray *I* travels from medium 1 into medium 2, making an incident angle *i* with the normal *N*. It traverses the interface and travels into medium 2 as the refracted ray *T*, making a refracted angle *t* with the normal *N* in medium 2. Given that the velocity of light is lower in medium 2, the angle of refraction *t* is smaller than the angle of incidence *i*.

3.2.3 The nature of light

From about the mid-seventeenth century, studies aimed at elucidating the nature of light consumed the attention of most scientists. Reasoning that light provides the primary information about the experience around them, these scientists thought that understanding what it is and how it works should provide the key to understanding the diversity of natural phenomena. Such studies (see below) were the purview of not only astronomers and physicists, but also of mathematicians, chemists, biologists, and even physicians. Other workers in related fields such as opticians, glass and lens makers, and instrument makers pursued theories of light derived from specific applications and improvements of their crafts.[26] This desire to understand what light is and how it works was the impetus behind the work that eventually resulted in the invention of photography and photolithography.

---

[26] M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 12 (1991).

3.2.4 Light and color

In an attempt to elucidate why when sunlight is bent in a glass prism the color of the rainbow are produced—a fact known since antiquity—Sir Isaac Newton (1641–1727) (see Fig. 3.4) in 1666 carried out a wide variety of experiments in which he allowed a narrow pencil of sunlight to pass through a small hole in a window shade of a darkened room, producing a small circle of light on the opposite wall. Next, he placed a prism in front of the hole and observed that the beams struck a different area of the wall. The displacement of the beams is not surprising because it resulted from the refraction of the entire beam by the prism. He observed that the spot of light was now elongated and was no longer a circle. He also observed that colors were spread along its axis, red on one end, violet on the other. On measuring the angle subtended by the axes of the spot as seen from the hole, he found the long axis close to 3 deg and the short axis about one-half a degree.[27]

Next, he proceeded to analyze the beam by keeping the prism in place and introducing beyond it a barrier with a small hole. Through this hole he allowed only the red portion of the beam to pass, and later the orange, the yellow, and so on. This allowed him to experiment separately with each color, which he did by placing a second prism beyond the barrier to refract light arriving through the small hole. He monitored the angle of refraction of red light alone, orange light alone, etc., and found that the angle is different for different colors; violet light is the most bent, red the least, and the other colors fall between, leading Newton to postulate two theorems: "light which differ in color, differ also in their degrees of refrangibility [refraction]," and "The light of the sun consists of rays of differently refrangible."[28]

Figure 3.4 Sir Isaac Newton (1641–1727), who, among his many endeavors, discovered the laws of universal gravitation and was a proponent of the particle theory of light (printed with permission from the Deutsches Museum, Munich).

---

[27] M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 6–8 (1987).

[28] Sir Isaac Newton, *Optics*, Dover Publications, New York, pp. 20 and 26 (1979).

In another set of experiments, he took the beam emerging from a prism—separated into colors—and let it pass through a second prism that was inverted with respect to the first. In this arrangement, the second prism performed the opposite function of the first. The result he obtained was a combination of the separated colors, which produced a white spot on the wall. He was thus convinced that sunlight is a combination of the colors of the rainbow.[29]

In order to explain the basis of the colors of material objects, using a prism, he took light of a particular color ("uncompounded light," as he called it) and let it shine on different objects. He observed that a red object in blue light looks blue, in green light it looks green, and in red light it looks red, but in this latter case it appears brightest. This led Newton to conclude that a red object reflects all colors but reflects red more strongly, so that in the "compound" white light of the sun, the object reflects red most strongly and appears red to the viewer.[30]

3.2.5 Light as a wave or particle

Toward the end of the 17th century, two opposing theories of light were being investigated. These fall into two broad categories, namely, (1) the particle (also called the corpuscular or emission) theory and (2) the wave (or undulatory) theory.[31] Knowing that light is a form of energy that can be transferred from one place to another, many scientists of the 17th and 18th centuries sought to describe it by analogy with other methods of energy transport. They distinguished between two methods of energy transport: transport by matter or by waves. While energy transport by matter is associated with the movement of the material body or the medium in which the matter is contained, energy transport by waves is not accompanied by any bodily movement of the medium. When energy is transported, it is not easy to discern whether the mechanism of transport is matter or wave. But there is one fundamental difference—the interference phenomenon—a difference that becomes the crucial factor in the history of the study of light phenomena.[32]

Interference is the phenomenon that distinguishes waves from matter. While waves can interfere with one another in a constructive (when the superposition of the waves add up) or destructive (when the superposition of the waves cancels each other) manner, matter cannot. It was thus reasonable for many scientists of the 17th and 18th centuries to describe light in terms of either moving particles or of waves.[33]

---

[29] M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 6–8 (1987).

[30] ibid.

[31] ibid.; M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, pp. 12–13 (1991); W. Ditchburn, *Light*, Dover Publications, New York, pp. 5–6 (1991).

[32] M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 9–10 (1987); R.W. Ditchburn, *Light*, Dover Publications, New York, pp. 6–7 (1991).

[33] ibid.

3.2.5.1 Particle (or corpuscular) theory of light

The corpuscular theory of light, expounded by Sir Isaac Newton, states that light is made up of very small corpuscles or streams of particles that are emitted from one body and move at great speed through space and strike some receptive body. Founded on the premise that light travels in a straight line and therefore must cast complete shadows when it illuminates a barrier—in other words, light does not go around corners—he found it rather difficult to accept that light is a wave, even when his own experimental results should have led him to such a conclusion. For instance, his observation of the now famous Newton's rings, an interference pattern of a concentric series of alternately bright and dark colored rings created by the reflection of light between two surfaces—a spherical surface and an adjacent flat surface (see Fig. 3.5)—is a consequence of the wavelike character of light.[34]

Newton did, however, recognize that these rings indicated the presence of some kind of periodicity and that this suggested a wave theory of light. He nevertheless believed that the rectilinear propagation of light was an insuperable objection to a simple wave theory. Instead, he suggested that light is made up of corpuscles that either possess an internal vibration of their own or are to some degree controlled by waves or vibrations of the medium through which they travel.[35] During most of the 18th century, the corpuscular theory was the dominant working theory for natural philosophers concerned with questions of light, largely because of Newton's influence.[36]

It should be pointed out that the objection to a simple wave theory was removed when it was discovered that the propagation of light is not strictly linear. Light does indeed spread, although to a very small extent, from the edges of a beam defined by rays. As an example, the shadow of a straight edge

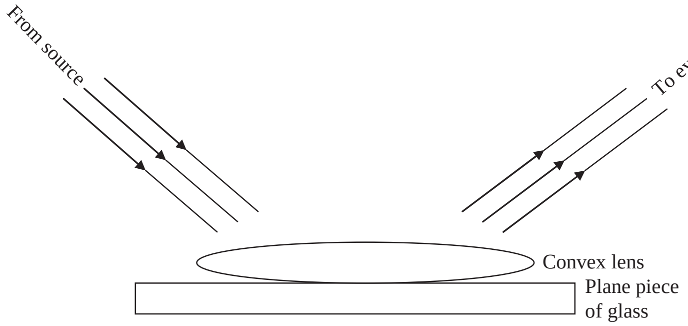


Figure 3.5 Apparatus for viewing Newton's rings.

---

[34] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 8 (1987); R.W. Ditchburn, *Light*, Dover Publications, New York , pp. 7–8 (1991).

[35] R.W. Ditchburn, *Light*, Dover Publications, New York, pp. 7–8 (1991).

[36] M.I. Sobel, *Light,* University of Chicago Press, Chicago, p. 8 (1987).

formed by a small source is not perfectly sharp when viewed under high magnification. Some light does penetrate the region that should be completely dark if light were propagated entirely in straight lines, and there exists a series of fine, light and dark bands at the edge of the region outside the shadow. The Jesuit Francesco Grimaldi who lived in Newton's lifetime made some observations of this kind, but it was not until 150 years later that the phenomenon (which is now called diffraction) was clearly elucidated. The discovery and elucidation of diffraction irrefutably showed that the propagation of light is not exactly rectilinear.[37]

Today, leading-edge lithography is all about diffraction-limited printing. The optical components of lithographic exposure tools are designed to limit the effects of diffraction, either through materials selection or the way the materials are used.

3.2.5.2 Wave (or undulatory) theory of light

The earliest systematic wave theory was proposed by the Dutch astronomer Christian Huygens (1629–1695), a contemporary of Newton. Although Huygens did not fully develop his wave theory mathematically, he showed how a train of waves striking the edge of a barrier will be bent around the edge, thus making it possible for a wave of light to penetrate regions behind a barrier. Huygens' theory was also able to explain Snell's law of refraction, if one assumes that waves travel more slowly in a dense medium than in air.[38]

Although the bending of light around an edge (otherwise called diffraction) had been observed by Grimaldi and Newton, the amount of light bent into the shadow is so extremely small that Newton did not regard the effect as evidence for the existence of light waves, but rather some property of the edge.[39] By the time the optical research of Newton (1672) and Huygens (1678) was published, it was already well established that sound was a pressure wave in air. It was not until the research of the English physician Thomas Young (1773–1829) (see Fig. 3.6) in 1801 that the wave theory of light was revived, and the similarities and differences between light and sound were explained.[40]

The first scientist to suggest the trichromatic theory of color vision, Thomas Young laid the foundation for the wave theory of light, stating that when a body becomes luminous it sends out waves of a particular amplitude and wavelength. He noted that the length of these waves determines their color in much the same way that the wavelength of sound waves determines

---

[37] R.W. Ditchburn, *Light*, Dover Publications, New York, pp. 7–8 (1991).

[38] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 10 (1987).

[39] ibid., p. 10.

[40] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 10 (1987). Young had previously published work on sound and light during the 1790s, but his primary work on light and color was published in 1801 in a lecture he delivered to the Royal Society of London, entitled "The Theory of Light and Colors."

Figure 3.6 Thomas Young (1773–1829), who laid the foundation of the wave theory of light and proposed the trichromatic theory of color vision (printed with permission from the Deutsches Museum, Munich).

their pitch.[41] Furthermore, he thought that light waves were transported in a special medium, the luminiferous ether, which vibrates and transports the waves to some receptive body at a distance.[42]

In his 1801 lecture "The Theory of Light and Colors," Young undertook to prove the wave theory of light by demonstrating interference effects; he also gave the first physical analysis of the nature of color. For instance, in explaining the basis of Newton's rings, he postulated that there exists a thin layer of air between the convex lens and the flat surface beneath it. Light being reflected from any boundary between two media, and in this case from the upper and the lower boundaries of the air layer, results in two such beams—upon arriving at the eye, having traveled paths of different lengths—that either destructively interfere to create the dark ring or constructively interfere to create the bright ring. He found that the difference between the bright and dark rays depends on the path difference of one-half the wavelength. Thus, by observing the separation between the light and dark rings, he was able to calculate the wavelength of light of different colors. He found a value of 0.00007 cm for red light, about 0.00004 cm for violet light, and values between these two wavelengths for the intermediate colors in the spectrum. With one stroke, he explained Newton's rings and a number of similar effects associated with thin films, such as the colors seen in soap bubbles or in thin layers of oil floating on water.[43]

Furthermore, he performed the simplest interference experiment, namely, allowing a light beam to fall on two parallel slits, and observing the pattern of

---

[41] M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 12 (1991).

[42] ibid.

[43] M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 10–12 (1987); M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, pp. 12–13 (1991).

bright and dark stripes, often called fringes, on a screen placed a finite distance beyond the slits. His findings show that each color in the prism's spectrum corresponds to a light wave of a particular wavelength, with the longer wavelengths appearing toward the red end of the spectrum and the shorter wavelengths toward the blue and violet end.[44]

When Young first presented his ideas, the main problem facing scientists was that the corpuscular and wave theories of light could not be resolved using available experimental knowledge—a fact made all the more difficult because the wavelength of light is so small, making its wave characteristics very difficult to detect unless the dimensions of the apparatus used are very small. For example, interference fringes can be seen only when the spacing between two reflection surfaces is very small. Diffraction can be easily seen only in the shadow of very small objects. The fact that the path of a light beam deviates from a straight line by an extremely small angle makes it extremely difficult to detect diffracted light beams—those spreading around corners.[45]

It was noted that, while the wave theory cannot give a satisfactory account of exactly rectilinear propagation of light, it is well suited to approximately describe rectilinear propagation, provided that it is assumed that the wavelength of light is small in relation to the relevant dimensions of the apparatus.[46] Young's ideas did not win immediate acceptance; however, powerful mathematical derivations based on experimental work in France and England helped sway many scientists of this era to the wave theory. Most importantly, work in the areas of color, interference, polarization, refraction, diffraction, and absorption were often better described by those who espoused Young's wave theory of light than those who espoused Newton's corpuscular theory of light.[47]

The period starting from about the late 17th century saw very important advances in the techniques of experimental physics, and the number and accuracy of experiments on light witnessed a significant increase. The observations obtained were in agreement with the wave theory, which became more exactly defined in the 19th century.[48] We describe below three important observations.

3.2.5.2.1 The wavelength of light

Many detailed experiments on interference and diffraction were made, leading to a set of determinations of the wavelength of light. It was established that, in a spectrum, the wavelength is related to the color. The wavelength is about

---

[44] M.I. Sobel, *Light,* University of Chicago Press, Chicago, p. 12 (1987).

[45] ibid.

[46] R.W. Ditchburn, *Light*, Dover Publications, New York , p. 8 (1991).

[47] M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, pp. 12–13 (1991).

[48] R.W. Ditchburn, *Light*, Dover Publications, New York, p. 8 (1991).

6.5×10^{-5} cm for red, 5.6×10^{-5} cm for green, and 4.5×10^{-5} cm for blue light. When the wavelengths were measured with different methods, consistent results were obtained.[49]

3.2.5.2.2 The velocity of light

In 1676 the Danish astronomer Ole Rømer (1644–1710) made the fundamental discovery that light does not propagate instantaneously, but travels with an enormously high, though finite, speed. He had arrived at these conclusions from a long period of observing the rotation of the moons of Jupiter and timing their eclipses as they passed around that huge planet. He observed that the moons seemed to either slow down or speed up depending on whether Jupiter was moving away from Earth or coming closer to it. This effect, he wrote,[50] was the result of the extra time that the light took to cover the distance between the planets when they were farther apart.[51]

The French physicist Armand Hippolyte Louis Fizeau (1819–1896) measured the speed of light in air to be 3×10^{10} cm/s by observing the length of time it takes light to traverse unobstructed through a pair of cogwheels set at opposite ends of a long axis and made to spin at the rate of a few thousand revolutions per minute. By interposing three mirrors between the cogwheels, with one of the mirrors placed far away, he lengthened the light path, thereby making the effect observable.[52]

Jean Foucault (1819–1868) (Fizeau's friend and collaborator) succeeded in shortening the distance by using rotating mirrors instead of cogwheels, and thus was able to measure the velocity of light in many different media—air, water, or any transparent material. He observed that the velocity of light in material bodies is less than its velocity in vacuum, thus providing a belated support to Huygens' wave theory and contradicting Newton's particle theory of light.[53]

---

[49] ibid., pp. 8–9 (1991).

[50] In describing his experiments, Roemer states: "I have been observing the first satellite of Jupiter over eight years. The satellite is eclipsed during each orbit of the planet on entering its vast shadow. I have observed that the intervals between the eclipses vary. They are the shortest when the Earth moves towards Jupiter and longest when it moves away from it. This can only mean that light takes time for transmission through space...

The speed must be so great that the light coming to us from Jupiter takes 22 minutes longer to reach us at the farthest end of our orbit round the Sun, than at the other of the orbit when we are nearest Jupiter. That is to say, light takes about 10 minutes to travel from the sun to the Earth; it does not travel instantaneously as alleged by M. Descartes" [cited in P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, p. 83 (1981)].

[51] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, p. 83 (1981).

[52] G. Gamow, *The Great Physicists from Galileo to Einstein*, Dover Publications, New York, pp. 160–162 (1961).

[53] ibid., p. 161.

Foucault's experiment was considered at the time to be very crucial, particularly in light of the fact that the corpuscular theory held that light travels faster in a dense medium because the molecules of the medium attract the particles of light, speeding up their motion as they enter; and then when the light leaves the medium, the same attractive force pulls the corpuscles back, retarding their motion as they enter the air. In contrast, the wave theory, according to arguments going back to Huygens, requires a slower speed in a denser medium than in air.[54] As predicted by the wave theory, the velocity of light in water, glass, etc., turned out to be exactly its velocity in vacuum divided by the refractive index of the material in question.[55]

In 1851 Fizeau carried out a very critical experiment, the full significance of which was, however, not realized until the publication of Einstein's theory of relativity. Fizeau was interested in determining how the velocity of light is influenced by the motion of the medium through which it propagates. For these investigations, he derived the empirical formula for the velocity V of light in a fluid moving with velocity v:

$$V = \frac{c}{n} \pm \left(1 - \frac{1}{n^2}\right) v, \tag{3.3}$$

where c is the speed of light in vacuum, and n is the refractive index of the fluid in question. Neither Fizeau nor anyone else at that time understood the implications of this result, thus the case rested until half a century later when Einstein showed that the empirical formula is a direct result of the theory of relativity.[56]

Einstein's relativity theory also predicts that mass and energy are interchangeable. His famous equation $E = mc^2$ is a relation between E, the rest energy of a particle, and m, the mass of the particle when at rest (called the rest mass).[57] When the particle is in motion, both E and m increase, but they remain connected by this fundamental equation and the universal constant c^2.[58]

---

[54] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 15 (1987).

[55] G. Gamow, *The Great Physicists from Galileo to Einstein*, Dover Publications, New York, p. 162 (1961).

[56] ibid., pp. 162–164.

[57] The Einstein mass–energy equivalence law can be derived from considerations of light pressure, first experimentally verified by the Russian physicist P.N. Lebedev and shown to be numerically equal to twice the amount of reflected energy E divided by the velocity of light c. Applying momentum arguments, one can show that a beam of light reflected from a mirror will experience a momentum equal to the product of the "mass of light" m falling on the mirror per unit time by its velocity: $P_{light} = 2\,mc$. Comparing this equation with Lebedev's empirical relation $P_{light} = 2E/c$, we obtain $m = E/c$ or $E = mc^2$.

[58] I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 205–207 (1987); G. Gamow, *The Great Physicists from Galileo to Einstein*, Dover Publications, New York, pp. 184–188 (1961); P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, p. 84 (1981).

In 1887 the American physicist A.A. Michelson (1852–1931) and his assistant E.W. Morley (1838–1932) carried out another remarkable experiment aimed at observing the effect of the earth's motion through space on the velocity of light as measured on its surface; in other words, to detect "the ether winds"—supposedly created from the motion of the earth against the luminiferous ether. The apparatus consisted of a terrestrial light source in a device Michelson had invented called an interferometer (see Fig. 3.7),[59] which measures interference—interference between two light beams that originate from the same light source (hence, are coherent). The initial beam was made to strike a glass plate positioned at a 45-deg angle; part of the beam is reflected, and part is transmitted. The two secondary beams travel at right angles and are eventually brought together by means of other mirrors, and then are viewed through a telescope. Given that the two beams have traveled different paths—perhaps one traveled parallel to the earth's motion and the other perpendicular to the earth's motion—the viewer sees an interference pattern, alternating fringes of light and dark, which itself is not surprising.

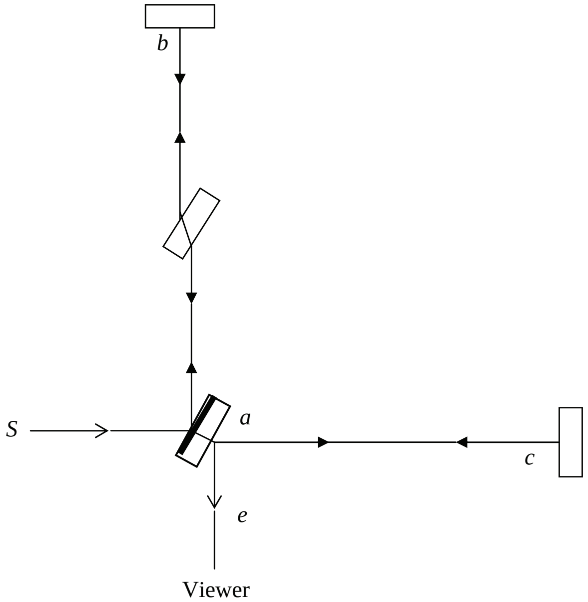


Figure 3.7 Michelson interferometer. Light from the source *S* is directed to the surface of a plane-parallel glass plate *a*, which separates it into two "coherent" pencil rays, one of which is transmitted through a glass plate en route to the surface *b*, where it is reflected back to the separating surface *a*. The other pencil ray is reflected normally from the plane mirror *c* back to the separating surface *a*, and together with the reflected beam from *b*, both return to the separating surface *a*, where they both proceed in the direction of *e*, where the resulting interference fringes may be projected on a screen or observed by the eye, with or without an observing telescope.

---

[59] A.A. Michelson, *Studies in Optics*, Dover Publications, Mineola, pp. 20–32 (1995).

However, in the final step of the experiment, the experimenters rotated the entire apparatus through 90 deg, effectively interchanging the paths of the two beams. They expected to see a shift in the fringes since the light paths through the ether had changed. Instead, the result they observed was negative, as no fringe shift was observed, implying that somehow the luminiferous ether remained stationary to the planet earth in its annual circuit around the sun. In other words, no ether wind was created, suggesting perhaps that the luminiferous ether does not exist, and even more importantly, that the speed of light in space is constant.[60]

How the speed of light is affected when the source of the light is moving became a question that attracted much theoretical speculation in the years after 1887. In 1905 Einstein found the solution lying at the heart of Maxwell's equations of electromagnetism[61] (see the section on electromagnetic theory).

Today, interferometers are used widely in the alignment system of lithographic exposure tools for coordinating the movement of the exposure stage. In particular, they have found application in high-precision measurements of extremely small distances between different objects in the exposure tool. Furthermore, interferometric lithography owes its very existence to the Michelson–Morley experiment.

3.2.5.2.3 The polarization of light

A rather curious observation was made in 1670 by Erasmus Bartholinus (1625–1698), who discovered that when a beam of ordinary light passes through certain crystals, such as natural calcite or Icelandic spar, each ray splits into two. When Bartholinus passed the two rays through a second crystal, he observed that the effect depends on the orientation of the crystal with respect to the beam. For certain orientations, the two rays each split into two. For other orientations, the two rays pass through the crystal unchanged (see Fig. 3.8). This phenomenon is known as double refraction (which is caused by birefringence). It is an indication that a beam of light that has passed through a crystal is differentiated with respect to the planes, including the direction of propagation.[62]

The simplest experiment on polarization was made much later (in 1808) by the French engineer Etienne Louis Malus (1775–1812), who was in Napoleon's army. Malus reflected a beam of light at the surfaces of two unsilvered pieces of glass (see Fig. 3.9) and observed that when the two reflections are in the same plane, a high proportion of the light incident on the second mirror (M_2) is reflected. If mirror M_2 is turned so that the second

---

[60] M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 199–202 (1987).

[61] Einstein takes Maxwell's electromagnetic theory as the starting point of his analysis. He does not make any reference to the Michelson–Morley experiment.

[62] R.W. Ditchburn, *Light*, Dover Publications, New York, p. 9 (1991).

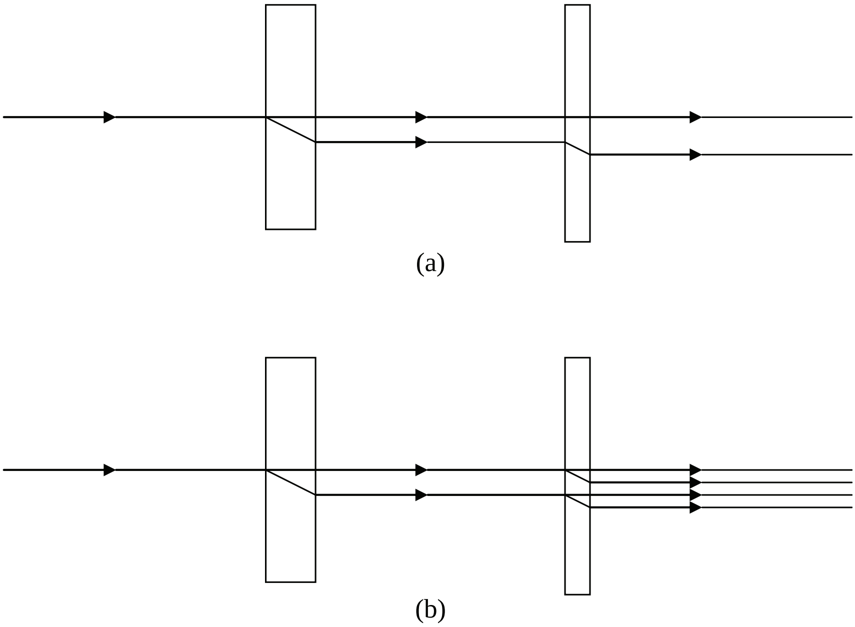


Figure 3.8 Double refraction: (a) two crystals with similar orientations and (b) different orientations of crystal axes.

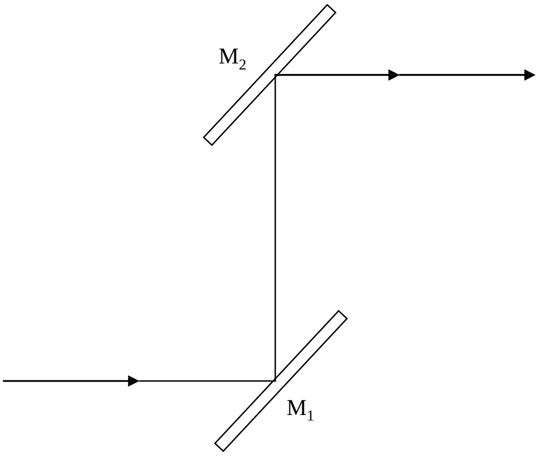


Figure 3.9 Malus' experimental setup. Note that M_1 and M_2 are unsilvered mirrors.

reflection is directed out of the plane of the paper, the reflected beam becomes weaker, approaching nearly zero brightness when the two reflections are in planes at right angles. This suggests that after the first reflection the beam of light has a special property in relation to the plane of the paper: it can be strongly reflected at a glass surface in this plane, but not at right angles. A beam of light possessing this property is said to be plane polarized.

Although this type of theory finds no place in a theory of longitudinal waves, it is adequately represented in a theory of transverse waves.[63] Thus, although polarization is a subtle property of light, its elucidation was instrumental in establishing one of the basic characteristics of light waves, along with an important difference between light and sound: light is a transverse wave, the oscillations being perpendicular to the path of propagation. This fact was realized by Huygens (1690), but it was not until the 19th century that it became possible to conduct experiments on the reflection and refraction of polarized light. Sound waves, on the other hand, were known in the 19th century to be longitudinal, consisting of oscillations of air in directions back and forth along the direction of propagation. Being longitudinal, the motion of sound can be represented as the variation of a scalar quantity, while transverse wave motion must be represented by a vector whose direction is related to the plane of polarization.[64]

The experimental and theoretical work of Thomas Young, and later of Augustin Fresnel (1788–1827), on interference and polarized light during the first few years of the 19th century yielded new evidence that supported the wave theory of light. Both scientists observed that under certain conditions light appeared to interfere with and be canceled out by other light that arrived at an eye or a lens at the same time. Young called this phenomenon *interference*. Fresnel's mathematical treatment of the interference phenomenon was far more rigorous and complete than that of Young. Fresnel developed the mathematical wave theory of light so thoroughly that the particle theory was vanquished and literally not heard of again until the 20th century.[65]

The mathematician Simeon Poisson (1781–1840) predicted (and also experimentally verified) one of the consequences of Fresnel's theory, namely, that a circular disk placed in a parallel beam of light should not only throw a circular shadow with narrow diffraction fringes around it, but should also show a bright spot ring in the middle of the shadow.[66] Fresnel also made extensive studies of diffraction by measuring fringes produced at the edges of shadows in various types of experiments; he showed that his results could be fully explained by the wave theory, with his measured values for wavelengths of light in agreement with those calculated by Young.[67]

---

[63] R.W. Ditchburn, *Light*, Dover Publications, New York, pp. 9–10 (1991).

[64] R.W. Ditchburn, *Light*, Dover Publications, New York, p. 10 (1991); M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 15–16 (1987).

[65] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, pp. 43–45 (1981); M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 13 (1991).

[66] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, p. 44 (1981).

[67] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, pp. 43–45 (1981); M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 13 (1991).

The success of the wave theory[68] greatly influenced the progress of similar work in other areas of optics, which also proceeded at a rapid pace during this period. This very success stimulated the demand for new and improved lenses and optical instruments, which had the effect of feeding the increasing demand for instruments to aid in the accurate depiction and recording of nature. Although the physics of image formation in a lens was fairly well understood during this time, new mathematics derived from the study of other light phenomena led to improved lens design and likewise improved instruments such as telescopes, microscopes, and cameras. Three notable specific examples of instruments include the *camera lucida*, the “Wollaston doublet” lens for microscopes, and a singlet landscape lens corrected to produce a flat field of view, all of which were developed by William Hyde Wollaston (1766–1828).[69] New methods for manufacturing glasses were devised during this period as well, and a more empirical understanding of the relationship between glass composition and the quality of lenses was also established.[70]

Today, in the regime of hyper-numerical-aperture imaging and immersion lithography, polarization effects have become very dominant. Understanding these effects and designing around them have relied enormously on the theories of Fresnel and Young.

3.2.6 Electromagnetic theory

The formulation of the wave theory of light predated the development of the fundamental laws of electromagnetism.[71] The notion of light waves traveling through vacuum was inconceivable to scientists of the 19 century who were schooled on the properties of water waves and sound waves (which travel through air and through liquids and solids). They therefore felt it necessary to postulate the existence of a medium that carries waves of light as transverse oscillations. This medium was the *luminiferous ether*, which pervaded all space, even out to the stars. The ether was imbued with properties that made it rarer than air since starlight is refracted when entering the atmosphere. It was incompressible, or else it would dissipate a light ray in all directions. It was perfectly elastic, such that it would not retard the passage of matter (such as the planets in their orbits), and it possessed the same kind of rigidity as a solid because only a solid is able to propagate a transverse wave. Even Lord Kelvin likened it to a mold of transparent jelly.[72]

---

[68] The wave theory resulted from efforts on the two sides of the English Channel. The success of that theory was attested to by the election of Fresnel, in 1827, as a foreign member of the British Royal Society and the election of Young, in the same year, as a foreign member of the French Academy; see also M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 15 (1987).

[69] M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 13 (1991).

[70] ibid.

[71] R.W. Ditchburn, *Light*, Dover Publications, New York, p. 10 (1991).

[72] M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 15–16 (1987).

It should be pointed out that natural philosophy in the 19th century was ripe for a bogus concept like the ether because scientists had come to accept the existence of substances other than ordinary matter. Under the shadow of Newton, science in the previous century had been dominated by a mechanistic world view in which all phenomena were due to the interaction of matter. But research involving the properties of heat, light, electricity, and magnetism made it manifestly clear that, although one might try to picture these as materials (fluids or corpuscles), they were nonetheless in some fundamental ways different from ordinary matter. Thus, these scientists thought that there were two types of matter: normal or "ponderable" matter, and the "imponderables." The ether was thus regarded as the most pervasive and subtle of the imponderables.[73]

A theory of transverse waves in such a medium presented a rather qualitative description of the fundamental phenomena of interference, diffraction, and polarization. In order to build the edifice of the wave theory, it was necessary to make special assumptions concerning the density and elasticity of this medium, and also concerning the conditions obtained at the surface separating two media such as glass and air. A close examination of these details revealed certain difficulties and appeared to indicate that there were some inconsistencies in the theory. All of these difficulties were resolved by Maxwell's electromagnetic theory of light—the great achievement of 19th century science that unified the imponderables of electricity and magnetism.[74]

Although this great theoretical synthesis was the work of James Clerk Maxwell (1831–1879) (Fig. 3.10) in Scotland, it in fact grew out of the experimental research of many scientists, including Hans Christian Ørsted (1777–1851) in Denmark, Charles Augustin de Coulomb (1736–1806), Jean-Baptiste Biot (1774–1862), and André-Marie Ampère (1775–1836) in France, and Michael Faraday (1791–1867) in England (Fig. 3.11). Ørsted was the first to demonstrate an intimate connection between electricity and magnetism when in 1820 he observed that a coil of wire carrying an electric current produces the same kind of magnetic field as a bar magnet. Faraday was the first to demonstrate the opposite phenomenon—electrical effects produced by magnets—when he observed that a magnet moved through a coil of wire induces a pulse of current in the wire (without the wire being attached to a battery or any source of electricity).[75]

Maxwell generalized a set of empirical facts that include the following:

1. Light is an electromagnetic wave with coupled electric and magnetic fields ($\vec{E}$ and $\vec{H}$, respectively) traveling through space.
2. These fields can interact with a material, resulting in four other quantities: the electric displacement $\vec{D}$, the magnetic induction $\vec{B}$, the electric current density $\vec{J}$, and the electric charge density ρ.

---

[73] ibid.

[74] R.W. Ditchburn, *Light*, Dover Publications, New York, pp. 10–11 (1991).

[75] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 16 (1987).

Figure 3.10 James Clerk Maxwell (1831–1879), who among many other accomplishments, consolidated and extended the equations of electromagnetism (printed with permission from the Deutsches Museum, Munich).

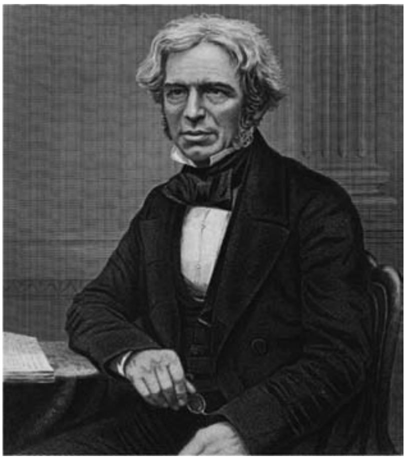


Figure 3.11 Michael Faraday (1791–1867), who among many endeavors, discovered electromagnetic induction in 1831 and established the connection between electricity, magnetism, and light; he also discovered electrolysis (printed with permission from the Deutsches Museum, Munich).

3. A changing magnetic field creates in its vicinity a changing electric field, and a changing electric field similarly creates a changing magnetic field.

Using these facts, Maxwell derived the famous equations [Eqs. (3.4) through (3.7)] that now bear his name, and which connect the rate of change of the magnetic field with the space distribution of the electric field, and vice versa:[76]

---

[76] G. Gamow, *The Great Physicists from Galileo to Einstein*, Dover Publications, New York, p. 152 (1961); P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, pp. 49–52 (1981);

$$\vec{\nabla} \times \vec{H} = \frac{4\pi}{c}\vec{J} + \frac{1}{c}\frac{\partial \vec{D}}{\partial t}, \tag{3.4}$$

$$\vec{\nabla} \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t}, \tag{3.5}$$

$$\vec{\nabla} \cdot \vec{D} = 4\pi\rho, \tag{3.6}$$

$$\vec{\nabla} \cdot \vec{B} = 0, \tag{3.7}$$

where c is the speed of light in vacuum. Equation (3.4) shows how a time-varying electric field (or displacement) causes a change in the magnetic field induction. In other words, a changing electric displacement produces a magnetic induction. Equation (3.5) shows how a time-varying magnetic induction causes a change in the electric field. In other words, a changing magnetic induction produces an electric field. Equation (3.6) shows how static charge affects the electric field, while Eq. (3.7) shows that there are no magnetic charges.[77]

It is from the properties of the materials involved that the relationships between current density and the electric field, electric displacement and the electric field, and magnetic induction and the magnetic field, are established. In the event that the material through which the electromagnetic radiation is propagating is isotropic and moving slowly relative to the speed of light, and the fields involved are time-harmonic, the three material equations become[78]

$$\vec{J} = \sigma\vec{E}, \tag{3.8}$$

$$\vec{D} = \varepsilon\vec{E}, \tag{3.9}$$

$$\vec{B} = \mu\vec{H}, \tag{3.10}$$

where σ, ε, and μ are the conductivity, dielectric constant, and magnetic permeability of the material, respectively. For transparent (or dielectric) materials, $\sigma = 0$. In general, materials used in lithography are nonmagnetic; thus $\mu = 1$.[79]

M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 16 (1987); C. Mack, *Field Guide to Optical Lithography*, SPIE Press, Bellingham, Washington, p. 11 (2006); C. Mack, *Fundamental Principles of Optical Lithography: The Science of Microfabrication*, John Wiley & Sons, Hoboken, New Jersey, pp. 30–32 (2007).

[77] ibid.

[78] ibid.

[79] C. Mack, *Field Guide to Optical Lithography*, SPIE Press, Bellingham, Washington, p. 11 (2006).

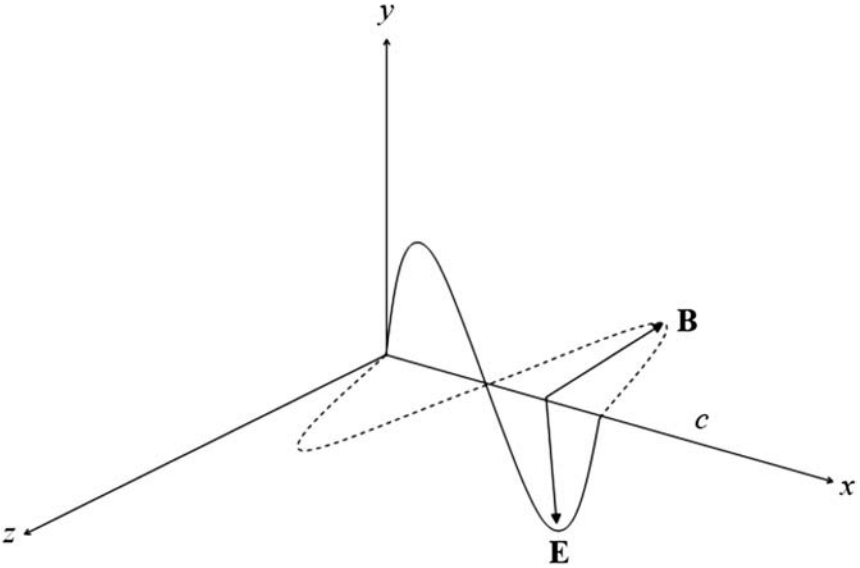


Figure 3.12 A schematic of a polarized electromagnetic wave at a fixed instant in time, showing the electric component **E** and magnetic component **B**. The light wave travels along the *x* axis with velocity *c*.

One of the great insights that derive from Maxwell's equations is the realization that both electric and magnetic fields are manifestations of the same underlying phenomenon—electromagnetism. Maxwell also deduced that these fields can exhibit wavelike properties[80] (see Fig. 3.12). As an illustration, an oscillating electric charge will create around its vicinity a region of oscillating electric and magnetic fields, and these fields in turn create oscillating electric and magnetic fields around themselves in a process that is repeated multiple times such that radiating spheres of electromagnetic fields proceed outward from the original charge. Given the periodic nature of the motion of the electric charge, the variation of the field out along a radius is also periodic, and the wave created in this way is a spherical wave.[81]

---

[80] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 17 (1987).

[81] G. Gamow, *The Great Physicists from Galileo to Einstein*, Dover Publications, New York, pp. 149–157 (1961); M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 15–21 (1987); R.W. Ditchburn, *Light*, Dover Publications, New York, pp. 10–12 (1991); E. Hecht, *Optics*, 4th ed., Pearson Education, New York (2001); R.A. Serway and J.W. Jewett, *Physics for Scientists and Engineers*, 6th ed., Brooks/Cole, Pacific Grove, California (2004); P. Tipler, *Physics for Scientists and Engineers: Electricity, Magnetism, Light, and Elementary Modern Physics*, 5th ed., W. H. Freeman and Co., New York (2004); J. Reitz, F. Milford, and R. Christy, *Foundations of Electromagnetic Theory*, 4th ed., Addison-Wesley, Boston (1992); J.D. Jackson, *Classical Electrodynamics*, 2nd ed., John Wiley & Sons, New York (1975).

In deriving his equations, Maxwell used electrostatic units for the electric field and electromagnetic units for the magnetic field. The application of these equations for describing propagating waves led him to the conclusion that the propagation velocity numerically equals the ratio of electrostatic units to electromagnetic units, which equals 3×10^{10} cm/s,[82] and which coincidentally is the velocity of light in vacuum measured by various methods long before Maxwell was born. Describing this numerical coincidence when he first published his theory of electromagnetism in the *Philosophical Magazine* in 1862, he stated, "This velocity is so nearly that of light, that it seems we have strong reason to believe that light itself (including radiant heat, and other radiations if any) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws."[83] Coincidentally, in the same year as Maxwell's publication, Jean-Bernard-Leon Foucault, using an improved method over the earlier work of Fizeau, determined a new, accurate value of the speed of light (differing by less than 1% from today's best value).[84]

In all, Maxwell formulated the equations of electromagnetism in a general form, and he showed that they are consistent with transverse electromagnetic wave propagation. He derived the velocity of propagation from constants measured in laboratory experiments on electricity and magnetism. He predicted that an oscillating electric charge will generate an electromagnetic wave and that the wave travels at the speed of light, given by $\sqrt{1/\mu_0\varepsilon_0}$, where ε_0 is electric permittivity of vacuum, and μ_0 is the permeability of vacuum. The wavelength depends on how rapidly the charge oscillates. His theory included an account of the propagation of electromagnetic waves in media such as glass. He showed that his theory gave a general account of the phenomena of reflection and refraction, including the formation of a spectrum by the dispersion of light. It is noteworthy to realize that Maxwell achieved all of this without introducing any arbitrary assumptions. The theory of light became transformed by him into a part of the theory of electricity and magnetism.[85]

It cannot be overemphasized that, because of Maxwell's electromagnetic theory, we now can understand the interaction of light and matter, including the phenomena of emission, propagation, and absorption of light, as resulting from forces acting between propagating short electromagnetic waves and

---

[82] G. Gamow, *The Great Physicists from Galileo to Einstein*, Dover Publications, New York, p. 156 (1961).

[83] J.C. Maxwell, *A Dynamical Theory of the Electromagnetic Field*; *The Scientific Papers of James Clerk Maxwell*, Dover Publications, New York, 1:535 (1965).

[84] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 17 (1987).

[85] M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 15–21 (1987); R.W. Ditchburn, *Light*, Dover Publications, New York, pp. 10–12 (1991); P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, pp. 49–52 (1981).

electrons, busily buzzing around the positively charged atomic nuclei. And, with Maxwell's equations, it is possible to explain in the most minute detail all of the phenomena and laws of optics.[86]

Today, the reach of Maxwell's electromagnetic theory extends all the way to lithography, where it guides the choice and usage of optical elements of lithographic exposure tools, as well as the interaction of these optical elements and radiation-sensitive resist materials with lithographic exposure radiations.

3.2.7 Electromagnetic spectrum

Even before the birth of Maxwell, the electromagnetic spectrum had been extended beyond the violet-to-red range revealed by the prism. As early as 1801, the German-born, British astronomer Friedrich Wilhelm Herschel (1738–1822) made the observation that, as he moved a thermometer across the beam emerging from a prism, from yellow to orange to red, heating effects increased, and they increased still further when he moved to a region beyond the red. This observation established that there is radiation beyond the red, called infrared, which, following the work of Maxwell, we recognize as electromagnetic waves with wavelengths longer than 700 nm. Herschel also observed that infrared radiation obeys the same laws of reflection and refraction that light obeys. Later workers found that infrared radiations show polarization and interference phenomena as well.[87]

At the opposite end of the spectrum, ultraviolet radiation (waves beyond the violet) was discovered on account of one of the chemical effects of light on certain salts. It had been known since early in the 18th century that light darkens silver nitrate and certain other salts. This phenomenon later became the basis of photography and was used in some of the early inorganic photoresists. Having learned of Herschel's infrared waves, the German physicist Johann Ritter (1776–1810) in 1801 discovered the darkening of silver nitrate by waves in the region beyond the violet, with wavelengths shorter than those of visible light.[88]

Maxwell's electromagnetic theory suggested the possibility of producing electromagnetic waves of other wavelengths beyond the visible. Although Herschel and Ritter demonstrated the effects of some of these waves, and by so doing, indirectly proved the existence of these waves, they were, however, not able to directly prove the existence of these waves since they did not produce them. It was not until 1887 that the required experimental evidence was provided by the German physicist Heinrich Hertz (1857–1894) (Fig. 3.13), who conclusively demonstrated the existence of electromagnetic waves. Hertz showed how to produce such waves with wavelengths much

---

[86] G. Gamow, *The Great Physicists from Galileo to Einstein*, Dover Publications, New York, pp. 156–157 (1961).

[87] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 19 (1987).

[88] ibid.

Figure 3.13 Heinrich Hertz (1857–1894), who demonstrated the existence of electromagnetic waves and thus confirmed Maxwell's electromagnetic theory (printed with permission from the Deutsches Museum, Munich).

greater than those of the visible spectrum and thereby opened the door to the world of radio and television.[89]

To generate electric sparks, Hertz used an induction coil with a vibrating spring as a contact breaker. The coil was connected to two brass balls that formed a spark gap. The balls were attached to long brass rods that acted as antennas. Given the sudden and violent surge of electricity in a spark, the motion of charges corresponds to oscillations at very high frequency. The sparks generated in the gap induce an alternating current in the rod, and the rod now acts as a transmitting radio antenna. To detect the electromagnetic waves emitted from the spark transmitter, Hertz used a secondary spark gap with its own antenna.[90] In order to enhance the received signal, he made the secondary gap small and adjustable. The experiments were an immediate success, as he observed sparks in the secondary gap at distances of several meters from the transmitter. Using parabolic mirrors made from zinc sheets enabled him to detect the waves at a much greater distance. He found that the waves could be blocked by metal sheets or by human bodies, but not by insulating materials.[91]

Hertz was able to identify these electromagnetic rays as an invisible form of light by demonstrating that they displayed the wavelike phenomena of reflection, refraction, polarization, and interference. From interference measurements, he showed that the wavelength of the waves received by the

---

[89] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, pp. 53–55 (1981).

[90] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 21 (1987).

[91] ibid.

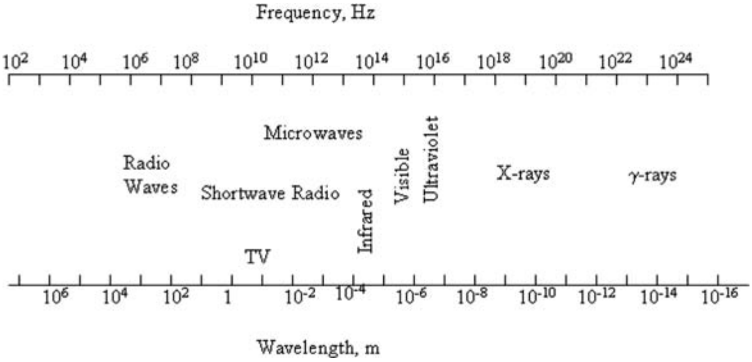


Figure 3.14 The electromagnetic spectrum, showing the frequency and wavelength of the various radiations. Visible light encompasses the 400- to 700-nm band.

detector was 66 cm, one million times longer than the wavelength of red light.[92] From both wavelength and frequency measurements of these electromagnetic waves, he was able to calculate the speed at which the waves travel, finding it to be precisely the speed of light.[93]

A great deal of the progress in experimental physics since Hertz' monumental work has consisted in the discovery of methods of producing electromagnetic waves of different wavelengths, a good number of which are now employed in lithography: near-UV (365 nm), deep-UV (248 nm and 193 nm), vacuum-UV (157 nm), extreme-UV (13.4 nm), x-ray (0.1–10 nm), etc. Some of the properties of these waves depend on their wavelength, but they all propagate with the same velocity (in vacuum) and all obey Maxwell's equations. Modern technical advances have made it possible to produce or detect waves of nearly every wavelength from above 3000 m to well below 10^{-11} cm[94] (see Fig. 3.14).

In general terms, the electromagnetic spectrum is classified by wavelengths into electrical energy, radio wave, microwave, infrared, the visible region we perceive as light, ultraviolet, x-rays, and gamma rays. The visible region occupies the band between approximately 400 nm and 700 nm, and is detected by the human eye and perceived as light. Radiation within the ultraviolet region is invisible, but it affects the skin and causes sunburn. Infrared radiation is invisible but produces the sensation of heat.

3.3 Key Developments in Optical Instruments and Glassmaking Technologies that Enabled the Development of Lithography

The discovery of glass dates back to remote antiquity. Natural glass has been in existence since the beginning of time. It is often formed when certain types

---

[92] ibid.

[93] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 22 (1987).

[94] R.W. Ditchburn, *Light*, Dover Publications, New York, pp. 10–12 (1991).

of volcanic rocks melt due to high-temperature natural phenomena (such as volcanic eruptions, lightening strikes, and even the impact of meteorites) and then rapidly cool and solidify. It is believed that stone-age people used cutting tools made of obsidian (a natural glass of volcanic origin) and tektites (naturally formed glass of extraterrestrial origin). The roots of optical technology as seen in the artifacts of the earliest manmade glass objects, consisting primarily of nontransparent glass beads, are believed to date back to about 3500 BC in areas of Egypt and Eastern Mesopotamia.[95]

The book of *Exodus*, written around 1200 BC, recounts how Bezalel, while preparing the ark and tabernacle, "made the laver [a ceremonial basin] and its base of bronze from the mirrors of ministering women who ministered at the door of the tent of the meeting."[96] In antiquity, mirrors were made of polished copper or bronze, and later on of speculum, a copper alloy rich in tin. The Greek dramatist Aristophanes alluded to the burning glass (a positive lens) in his comic play *The Clouds*, written in 424 BC. The Romans, as reported in the accounts of Pliny (23–79 AD), also made use of burning glasses. The Roman philosopher Seneca (3 BC–65 AD) wrote that a water-filled glass globe could be used for magnifying purposes.[97]

During the middle ages, the Franciscan Roger Bacon (1215–1294), considered to be the first scientist in the modern sense, was believed to have initiated the idea of using lenses for correcting vision and even hinted at the possibility of combining lenses to form a telescope. Medieval alchemists were believed to have produced an amalgam of tin and mercury that was rubbed onto the back of glass plates to make mirrors.[98] Leonardo da Vinci (1452–1519) described the *camera obscura*, which was later popularized by the work of Giovanni Battista Della Porta (1535–1615), who in 1589 discussed multiple mirrors and combinations of positive and negative lenses in his *Magia Naturalis.*[99]

The pace of optical technology accelerated tremendously in the 17th century, aided by theoretical works on the nature of light (reported above). The refracting telescope, patented by Hans Lippershey (1586–1619), a Dutch spectacle maker, in 1608 was the first major optical instrument invented in that century. Upon hearing about this invention, Galileo Galilei (1564–1642), in Padua, built his own instrument within several months, grinding the lenses by hand. The compound microscope was invented at about the same time, possibly by the Dutchman Zacharias Janssen (1588–1632), and later

---

[95] http://www.glassonline.com/infoserv/history.html (accessed 12 August 2019).

[96] *Holy Bible, The Book of Exodus*, Chapter 38, verse 8, Revised Standard Version, Catholic Edition, Camden, New Jersey, p. 82 (1966) [cited in E. Hecht, *Optics*, 2nd ed., Addison-Wesley Publishing Co., Menlo Park, p. 1 (1990)].

[97] E. Hecht, *Optics*, 2nd ed., Addison-Wesley Publishing Co., Menlo Park, p. 1 (1990).

[98] ibid., p. 2.

[99] ibid.

improvements were made on this design by Francisco Fontana (1580–1656) of Naples, who replaced the concave eyepiece in Jansen's microscope with a convex lens. A similar change was introduced in the telescope of Johannes Kepler (1571–1630). In 1668 Isaac Newton (1642–1727) invented the reflecting telescope as his solution to the problem of chromatic aberrations. John Dollond (1706–1761) in 1758 combined two elements, one of crown and the other of flint glass, and succeeded in forming a single achromatic lens used for correcting chromatic aberration.[100]

It was not until the latter stages of the Industrial Revolution that mechanical technology for mass production and scientific research into the relationship between the composition of glass and its physical qualities began to be adopted in the industry. The key figures in this development are briefly described below.

By the 19th century, optical technology had progressed to such a level that it was recognized that further development depended on a perfect marriage between science and craftsmanship—an idea first proposed by Josef von Fraunhofer (1787–1826) and implemented by the trio of Carl Zeiss (1816–1888), Ernst Abbe (1840–1905), and Otto Schott (1851–1935)—the triumvirate of Carl Zeiss company, whose story has become synonymous with that of modern optical technology, which includes the principles used in the design and construction of lithographic exposure tools. This is an interesting story,[101] and aspects of it need to be retold here, if only for its didactic lessons.

The story begins on 21 July 1801, when a glazier's workshop in a little village just outside of Munich suddenly collapsed, bringing down the whole building on top of it. The glazier's assistant, Josef von Fraunhofer, had been buried under the great rubble of bricks, beams, and glass; but as if by a dint of miracle, the debris formed an arch from which the lad had to be extricated, unhurt. Passing nearby in his carriage was none other than the Duke of Bavaria, who was attracted to the scene no doubt by the commotion. Sending for the lad, the Duke found him to be a bright, intelligent boy who, although not formally educated, had learned about glassmaking from his father. Greatly impressed by the young Fraunhofer, the Duke presented him with a bag of money, which he promptly used to buy himself a glassmaking machine, some books on optics, and a release from his apprenticeship. Throughout the remainder of his remarkably short life of thirty-nine years, Fraunhofer made outstanding advances in glassmaking, in the construction of optical instruments, in the foundation of spectroscopy, and in the wave theory of light.[102]

With a prism and a telescope he manufactured by himself, Fraunhofer discovered numerous black lines in the spectrum of sunlight thrown by a prism. He made use of these lines as wavelength markers when he was

---

[100] ibid., pp. 2–5.

[101] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, pp. 151–166 (1981).

[102] ibid., p. 151.

developing his new lenses. Subsequently, these lines would be recognized as "optical fingerprints" for identifying individual elements whose atoms absorbed only certain wavelength from the sun. Soon after, Fraunhofer became one of the world's premier scientific instrument makers.[103]

With the goal of achieving a perfect marriage between science and craftsmanship in order to improve the quality of his instruments and to distinguish them from those of his competitors, he developed the following three basic principles, which he assiduously followed: (1) the improvement of technical manufacturing methods for optical instruments and glasses, (2) research and development of basic theories of optics, and (3) the improvement of raw materials for making glass. Although each of these three principles was important in its own right, it was only by combining all three of them that he believed progress could be made beyond the old methods of trial and error used by his contemporaries. Unfortunately, when he died in 1826, the German scientific instrument industry went into decline, along with his approach.[104] Fraunhofer's illustrious approach was lost and not resurrected until 1882, when the optical team of Carl Zeiss (Fig. 3.15) and Ernst Abbe (Fig. 3.16) went into collaboration with Otto Schott (Fig. 3.17), the glassmaker.[105]

Carl Zeiss, after leaving high school in 1834 and spending an additional 12 years in apprenticeship studying and working in various scientific workshops, set up his own instrument workshop in November of 1846 and

Figure 3.15 Carl Zeiss (1816–1888), founder of the Carl Zeiss optical company in Jena in 1846 (reprinted with permission from the Deutsches Museum, Munich).

---

[103] ibid.

[104] ibid., p. 152.

[105] ibid., pp. 151–152.

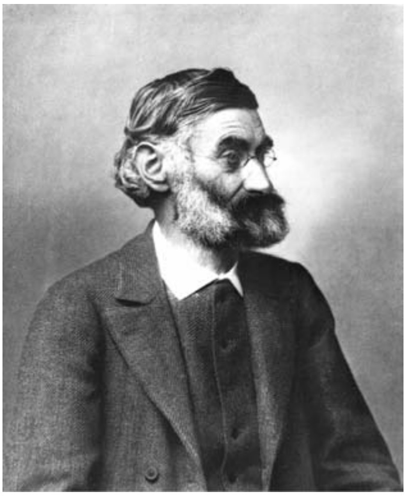


Figure 3.16 Ernst Abbe (1840–1905), who developed the Abbe sine condition—a design specification derived from wave theory for ensuring good image quality in microscopes. He also transformed the Zeiss optical company into Carl Zeiss Foundation in 1889. The greatness of this company was due not only to the quality of its optical products, but equally to the care it took of its workers at a time when exploitation and grinding poverty were the rule. (Photo courtesy of the Carl Zeiss Archives.)

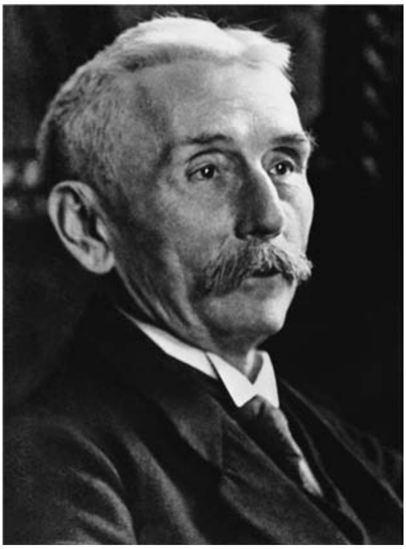


Figure 3.17 Otto Friedrich Schott (1851–1935), who along with Zeiss and Abbe established the Technical Glass Laboratory of Schott and Co. in Jena in 1884. Schott was also a partner in the Zeiss optical company. (Printed with permission from the Deutsches Museum, Munich.)

immediately went into business as an instrument maker in the little university town of Jena. Jakob Schleiden (1804–1881), a founder of modern botany and a professor at the University of Jena, offered him guidance in the manufacture of microscopes. Like many of his contemporaries in the same business, his method of manufacturing lenses relied on the trial-and-error-methods, which seemed rather wasteful and inefficient. Seeking to replace these old hit-or-miss methods with a true scientific understanding of the microscope, he enlisted the help of some of his friends at the University of Jena. First, he enlisted the help of the mathematician Friedrich Wilhelm Barfuß (1809–1854), who worked on this problem until his death, but without success. Finally, he enlisted the help of the young physics lecturer Ernst Abbe.[106]

In characteristic fashion, Abbe made a wide-ranging search of the available optical literature, but found nothing of help. After numerous visits to the Zeiss workshop, Abbe came to the conclusion that the main problem with existing microscopes of the day was that the rays of light passing through the outer parts of the lens came to a focus at a slightly different point from the rays passing through the middle, thereby degrading the sharpness of the image. He worked out theoretical ways of solving this problem and obtained an exact specification for the shapes, sizes, and positions of the lenses. He left absolutely nothing to the judgment of the instrument maker.[107]

But when a dozen microscopes were built in 1868 in accordance with Abbe's specifications, it was determined that the performance of all of them was inferior to the standard trial-and-error instruments they had been producing. Undeterred, Zeiss encouraged Abbe to try again and closely followed his progress while he spent the next two years working in the shops, studying the techniques, and designing the measuring and testing equipment that would be needed if a satisfactory theory were to be found. At long last, Abbe realized that the accurate theory for constructing microscope lenses was none other than the wave theory of light. He realized that the objects being examined under the microscope were so finely structured that the light from each tiny detail did not go simply in a straight line but was diffracted into a pattern of light and dark bands, just as in Newton's interference experiments with pinholes or Young's experiments with slits. To obtain a true image of that detail, he determined that it was necessary to collect all of the light diffracted from it and to reassemble it with the system of the lenses.[108]

Next, Abbe calculated the precise geometry of the lens system using the wave theory of light. The microscopes made to his new specification showed good improvement, but were still of a poor quality. When he next recalculated the geometry of the lens system to ensure that points a small distance from the direct axis were brought into the same sharp focus as points in the direct line

---

[106] ibid., p. 153.

[107] ibid.

[108] ibid., p. 154.

of sight, he devised what is now referred to as the "Abbe sine condition"—a specification derived from the ray theory of light that is still a key feature in modern microscope design, which has greatly influenced the design of lithographic exposure tool lenses. When the Zeiss company built "theoretical microscopes" according to the specification of the Abbe sine condition, while simultaneously meeting the specification that they collect enough diffracted rays from the object to resolve the finest possible detail, these microscopes were determined to be at least as good as the best microscopes of their day; more importantly, each was made to a reproducible specification—a decided edge over competing microscopes from other companies, which were still made by repeated trial-and-error methods. Needless to say, Abbe was greatly delighted that, upon examining a number of the best microscopes then available to find, they all fulfilled the Abbe sine condition. Thus, in a way, the old craftsmen had successfully corrected their lens systems for off-axis imaging without even realizing it.[109]

With the lens design problem solved, Abbe turned his attention to the problem of how to get the right kind of glass. At that time, a lens designer had only two distinct options when it came to the choice of glass: *crown glass*, made of silica, soda, and potash, and *flint glass*, the same mixture made heavier by the addition of oxides of lead. Flint glass characteristically refracts light more strongly, while also showing a much greater difference between its refraction of the different colors compared to crown glass. It was customary during this period to combine a crown glass with a flint glass in a way that brought light of different colors to a focus at the same point in the 'achromatic" microscope lenses being produced during this period. This helped to a certain extent in eliminating the confusing multicolored images that limited the performance of the earlier microscopes.[110] It should be pointed out that Fraunhofer had earlier shown that a complete solution to the color problem was theoretically possible, and Abbe succeeded in finding this solution only to be hampered by a lack of glass with the required optical properties, as he wrote:[111]

> "This discussion then leads to the conclusion that the deficiencies of present-day microscope objectives... are caused by the optical properties of the presently available raw materials. Therefore, perfection of the optical quality of the microscopes seems to depend primarily on the progress in the art of glassmaking."

Quite auspiciously, in 1879 the solution to the high-quality glass availability problem materialized in a large parcel Otto Schott sent to Abbe containing pieces of glass he had made by substituting lithium for the sodium

---

[109] ibid.

[110] ibid., pp. 154–155.

[111] E. Abbe, *Gesammelte Abhandlungen*, vol. 1, G. Fisher Verlag, Jena, p. 159 (1906).

or potassium in ordinary glass. Otto Schott, a young chemist whose father owned a plate glassworks in Westphalia, had written his graduation thesis on the manufacture of window glass. He was more interested in using his basic chemical knowledge of minerals to discover entirely new kinds of glass than he was in making windows or in making money for that matter. He was particularly interested in using scientific methods to study the effects of numerous chemical elements on the optical and thermal properties of glass. So it was that he requested Abbe to measure the optical characteristics of the new glass materials he had sent him.[112]

When Abbe measured the optical characteristics of Schott's new glass, he found them to differ from those of the conventional glasses, although unfortunately not in the way required for Abbe's achromatic lens design. Nevertheless, Abbe felt excited about the prospects he saw in the very existence of such an enterprising glassmaker, especially one who appreciated the importance of producing a number of small batches of glass with systematically varying composition. In no time, Schott joined the team of Zeiss and Abbe, and within three years he moved to Jena to operate a small glassmaking laboratory that the three set up jointly. In 1884 this laboratory was transformed into the Technical Glass Laboratory of Schott and Company. Two years later, it heralded a new era for optical instruments by issuing a catalog of 44 different types of glass, such as heat-resistant glasses, thermometer glass, borate, phosphate, and zinc glasses, as well as the special optical glasses that Abbe needed for his color-correction lenses.[113]

The strong bond of friendship and trust between these three men stood up to some severe testing. Having established the glassmaking company largely with the help of a large government subsidy that was to be conditioned on transfer of the works to Berlin, when faced with a decision to either abandon his colleagues and move the glassworks to Berlin to receive the full government subsidy or stay with his glassworks in Jena with his colleagues and lose the subsidy, Schott chose the latter. The government withdrew its condition.

Another severe test came when Herman von Helmholtz (1821–1894), the leading German biophysicist at the University of Berlin at the time, heard of Abbe's explanation of image formation in terms of the wave theory of light. Helmholtz had earlier worked on this problem, but without success, and so he set off eagerly for Jena to obtain a first-hand account from Abbe. Very much impressed with what he heard, he returned to Berlin and wrote to Abbe offering him a professorship at the university, but Abbe refused to accept the offer, just as Schott had refused the German government's offer. Nothing—not even money or prestige—would deter the three partners from pursuing their work together in optical science.[114]

---

[112] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, pp. 155–156 (1981).
[113] ibid., p. 157.
[114] ibid.

Shortly after the death of Carl Zeiss in 1889, Abbe developed his most advanced microscope—a microscope with a newly designed eyepiece and an oil-immersion objective, referred to as Abbe's "famous condenser," which directed the light onto the specimen. Together these two components took the instrument to the ultimate limit of resolving power. Abbe's own theory had shown that no feature finer than about one-half of a wavelength of the illuminating light could ever be resolved. Achieving the maximum resolving power in a microscope in 1889 was of the utmost importance for the study of microorganisms in the then rapidly growing field of bacteriology. Compared to the first observations of dental bacteria in 1683 by the Dutch pioneer Antonie Philips van Leeuwenhoek (1632–1723), who used a single-lens microscope to magnify the bacteria to perhaps 150 times, Abbe's new instrument could magnify up to 2000 times to resolve details in bacteria that are only a few wavelengths in size.[115]

It should be pointed out that the theoretical optical limits of resolution predicted by Abbe's theory probably saved many other microscope manufacturers from striving after the unattainable. This may also have unwittingly contributed to the belief that was common in Europe around 1890, that physical science was largely complete, and that all of the major discoveries had already been made. This belief was no doubt shattered within a few years by a succession of unexpected discoveries of the greatest importance, such as radioactivity, the electron, x-rays, etc.,[116] discussed in the following sections of this book. These unexpected discoveries were made possible in large part by advances in glassmaking technologies and optical instruments. In succeeding sections, we see the role of Crookes tubes in the discovery of electrons and x-rays, along with the role of photographic plates in the discovery of radioactivity. Crookes tubes are partially evacuated glass bulbs of various shapes with two metal electrodes on either end, and which upon application of a high voltage between the electrodes, cathode rays (electrons) are projected in straight lines from the cathode. A photographic plate is a glass plate coated with light-sensitive emulsions of salts with which a photographic image can be recorded through the interaction of light and the emulsion. Roentgen used a Crookes tube to discover x-rays. J.J. Thompson used a Crookes tube to discover electrons. Becquerel used a photographic plate to discover radioactivity.

Roentgen's discovery of x-rays opened the way to the development of x-ray diffraction, a very effective tool for exploring the atomic and molecular structure of matter. It also led to the development of x-ray lithography. The discovery of the electron by J.J. Thompson opened the way to electron microscopy, influenced no doubt by one of the great surprises of 20th-century physics—the discovery that matter, just like light, could behave as a particle

---

[115] ibid., pp. 157–158.

[116] ibid., pp. 158–159.

or as a wave, depending on the circumstances (as discussed above). An electron can behave like a particle when accelerated from the filament in a television tube to produce a point of light on the screen; it can also behave like a wave. A beam of electrons can be focused, using electric or magnetic lenses, resulting in the image of a very small object being greatly magnified and projected as a picture on the screen.[117] Similar principles are employed in exposure tools used in electron-beam lithography.

In the 1890s, Abbe turned his attention to devising ways to ensure the continued success of the Carl Zeiss enterprises, while ensuring the well-being of its workers. After having considered the limitations of the common forms of corporate structure, for example, the limited liability company or the co-operative, he found them to offer no guarantee that the social responsibility of the concern would not one day be supressed by selfish and profit-oriented motives. As a result, he created a new form of organization, the Carl Zeiss Foundation, drawing up the statues of the Foundation himself. The guiding principle of the Foundation was that the material and scientific success of the Zeiss works did not stem from the activity of a few individuals, but rather was due to the participation of the entire personnel. Consequently, he stated that all employees should share in the benefits of the firm's operations and should be entitled to life-long economic security. Noting that the Zeiss works could fulfill this objective only if its sound financial position was ensured, he included in the statues precise regulations as to the way the income was to be distributed between maintenance, replacement, production, and scientific research.

Remarkably, the regulations that attracted the most controversy were those aimed at safeguarding the rights of the workers: (1) an hourly wage or salary, once granted, cannot be reduced; (2) all employees of more than five years standing shall enjoy job protection; and (3) hiring of employees must always be irrespective of race, political affiliation, or religious belief. He also introduced a nine-hour day, which was reduced in 1900 to eight hours, sickness pay, holiday pay, retirement benefits, health insurance, and a workers' council. He made all of the provisions in the statues permanently binding on the employer, so that the personnel and economic status of the employees was not a matter of charity but an inalienable legal right.[118]

In 1889 Abbe established the Carl Zeiss Foundation, and two years later he transferred to the Foundation his ownership of the optical works and his half-share in the glassworks. By so doing, he gave up his position of owner to become one of the team of four managers who governed the company according to the statues.[119]

---

[117] ibid., pp. 158–160.

[118] ibid., pp. 160–162.

[119] ibid., pp. 162–163.

It is a lasting tribute to the inspiration of Abbe, as well as to the success of his plan, that 20 years later, at the end of the World War I, Otto Schott followed in his footsteps when he presented his share of the glassworks to the Foundation, effectively exchanging his position of owner for that of a manager.[120]

On 19 March 1945 during World War II, squadrons of British and American bombers bombed the town of Jena, destroying the Zeiss facilities. At the end the war, 80 of the leading scientists and engineers of the Zeiss optical works and a similar group of 40 from the Schott glassworks were relocated to Oberkochen in the western part of Germany, in the Allied zone, where they began with whatever they managed to salvage from Jena to build up a new optical organization that they regarded as not just a successor to the Carl Zeiss Foundation, but as the genuine article of the Foundation itself. Meanwhile in Jena, much of the damaged heavy equipment that the Russians had taken away under the reparations agreement was brought back when the factory was rebuilt, and soon after, the Carl Zeiss Company of Jena was back in business.[121]

Following the collapse of the Berlin Wall in 1989, the two Carl Zeiss Foundations were reunited and are still flourishing. Being at the forefront of modern technological developments, they are continually turning out new instruments for research and production, new microscopes, telescopes, spectroscopes, lithographic exposure tool lenses, precision measuring instruments, etc.[122]

Numerous other notable discoveries, inventions, and innovations in optical science and technology have been made since the founding of the Zeiss Company. The year 1930 witnessed the discovery of the phase-contrast phenomena by Frits Zernike (1888–1966), the principle of which underlies his invention of the phase-contrast microscope in 1938, and for which he won the Nobel Prize in Physics in 1953. Phase-contrast microscopy made possible the production of high-contrast images of transparent specimens such as living cells, subcellular organelles, lithographic patterns, etc. Also, the same Frits Zernike introduced in 1934 the "Zernike polynomials,"[123] a set of orthogonal polynomials that arise in the expansion of a wavefront function for optical systems with circular pupils. Since 1960, these polynomials have found wide application in optical design, metrology, and image analysis.

Since World War II, developments in optical technology have spanned many fronts, ranging from optical element design to the incorporation of advanced mathematics and approaches for analyzing image formation and quality, to new optical materials and processes for manufacturing them, to the use of computers in image processing and optical element design, to new

---

[120] ibid., p. 163.

[121] ibid., pp. 165–166.

[122] ibid., p. 166.

[123] F. Zernike, "Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode," *Physica* **1**, 689–704 (1934).

optical devices and optical effects, etc. A few of the notable developments, starting from about the end of World War II, are outlined below.

From about the 1950s, several workers in different countries began to infuse optics with mathematical techniques and insights of communication theory. Bound together by Fourier's mathematical analysis, the outgrowths of this approach (called Fourier optics) have been far-reaching, particularly in the areas of theory of image formation and evaluation, the transfer functions, and spatial filtering. In particular, the use of techniques like spatial frequency filters have offered a rich new way of appreciating a broad range of optical phenomena.[124]

In addition, the availability of high-speed digital computers has engendered a vast improvement in the design of complex optical systems. New practical uses were found for aspherical lens elements, and the diffraction-limited system with an appreciable field of view finally became a reality. The ion bombardment polishing technique, in which one atom is chipped away at a time, was introduced to meet the need for extreme precision in the manufacture of optical elements. Single and multilayer thin-film coatings (reflecting, anti-reflecting, etc.) became widely available. Fiber optics evolved into a practical tool, along with thin-film light waveguides. A great deal of attention was paid to the infrared region of the spectrum, with an eye to applications in surveillance systems, missile guidance, communications, etc., and these in turn have stimulated the development of infrared materials. Plastics began to be used in optics, particularly in lens elements, replica gratings, fibers, aspherics, etc. A new class of partially vitrified glass ceramics with an exceedingly low thermal expansion coefficient was developed and is now widely in use.[125]

The year 1960 saw the building of the first laser, and within a decade laser beams spanned the range from infrared to ultraviolet. The availability of high-power coherent sources led to the discovery of new optical effects such as harmonic generation, frequency mixing, etc., which in turn were incorporated into a sleuth of new devices. The technology required in the manufacture of practical optical communication systems also witnessed very fast evolution. The applications of crystals in devices such as second-harmonic generators, electro-optical and acousto-optic modulators, and the like have spurred a great deal of contemporary research in crystal optics. Holography, the wavefront reconstruction technique that produces three-dimensional (3D) images, was found to have numerous additional applications in nondestructive testing, data storage, etc.[126]

The 1980s and 1990s saw the development of lithographic exposure sources such as KrF and ArF exciplex lasers. Today, the majority of advanced IC circuits are fabricated with these laser-based exposure tools.

---

[124] E. Hecht, *Optics*, 2nd ed., Addison-Wesley Publishing Co., Menlo Park, p. 10 (1990).

[125] ibid.

[126] ibid., p. 11.

From about 1970 to the present, the applications of optics for military purposes has ranged across the spectrum from "smart bombs" and spy satellites to "death rays" and infrared gadgets that see in the dark. But profit motive coupled with the need to improve quality of life have also brought optical products into the consumer marketplace as never before. Today lasers have found widespread applications in a variety of areas, including reading videodiscs in living rooms, cutting steel in factories, setting type in newspapers, scanning labels in supermarkets, performing surgery in hospitals, exposing resists used in patterning ICs, etc.[127] The list of appliances that now rely on optics for their operations is staggeringly long.

3.4 Key Developments in Atomic and Molecular Physics that Enabled the Invention and Development of Lithography

3.4.1 The nature of matter

The belief that the atom is the ultimate particle of matter has been held since the time of Demokritus (500 BC) and was formalized by Dalton in his atomic theory in 1808 (see Chapter 4). The idea that atoms combine to form molecules was first proposed by Avogadro in 1811 and has been widely accepted since 1860 (see Chapter 4). That an atom is made up of electrons (discovered by J.J. Thomson, see below) and a nucleus (discovered by Ernest Rutherford, see below) is now a well-established scientific fact. That the nucleus is, in turn, made of up protons and neutrons—thanks to discoveries by Rutherford and coworkers—is also a well-established fact. Particle physics has since identified still smaller subatomic particles of matter. The discovery of these ultimate particles of matter, as we shall see shortly as well as in Chapter 4, was an enterprise in which physicists and chemists worked hand in hand, borrowing greatly from each other's storehouses of knowledge, and in some instances obliterating the borderland between the two disciplines. The nature of these ultimate particles of matter and the story surrounding their discovery is presented below.

3.4.2 Molecular theory of matter

In 1808 the English chemist John Dalton (1766–1844) proposed the molecular theory of matter, which states that each element is made of identical atoms, each compound is made of many identical molecules, and each molecule is made of a small number of atoms. The theory further states that a chemical reaction is the combination or recombination of several atoms to form molecules, the process being multiplied many times to become observable in nature or in the laboratory.[128] Applying the molecular theory to gases (with the

---

[127] ibid., p. 11.

[128] ibid., p. 73.

assumption that a gas consists of widely separated molecules interacting very little with each other and moving incessantly at high speeds in a random manner), James Joule (1818–1869) in 1848 explained the pressure of a gas on the walls of a container.[129] Other theorists of the mid-nineteenth century were able to deduce from this kinetic molecular theory properties of gases, including their expansion when heated and their diffusion through air.[130]

Between 1860 and 1870, the Austrian physicist Ludwig Boltzmann (1844–1906) introduced the statistical theory of heat (statistical mechanics) into the molecular theory of matter. He postulated that, although the molecules of a gas—or indeed any matter—are so numerous that one can never predict the exact path of each one, one can nevertheless predict the behavior of the gas as a whole by assigning a set of probabilities for each molecule being in a particular state. These laws of statistical thermodynamics provided a firm foundation for the kinetic theory of gases, as well as an explanation for the phenomena of temperature and heat. The laws defined temperature as a measure of the average speed of these random molecular motions and defined heat as energy associated with the random speeds of molecular motion.[131]

Today, the rational design of resist materials and optical elements used in lithography, as well as the interaction of radiation with these materials, are based on these same molecular theories of matter.

3.4.3 Electrons

By the end of the 19th century, the properties of light waves were fairly well understood. Long-wavelength radiation like infrared and radio waves were already discovered, along with short-wavelength ultraviolet radiation. Still, scientists sought to find out whether there were other shorter-wavelength radiations yet to be discovered; they also wondered which properties these types of radiation might have, along with their possible applications.

It was in this climate that the cathode rays produced by electrical discharges in gases at low pressure were discovered by Julius Plucker (1801–1868) in 1859.[132] Aided by improvements in vacuum pumps to provide a high degree of vacuum and the induction coil to provide high-voltage electricity, Sir William Crookes (1832–1919) in 1876 built the first cathode ray tube, the forerunner of the television tube. With the induction coil he discharged a high-voltage current of electricity through a highly evacuated glass tube, resulting in the occurrence of peculiar bands of fluorescent light at the negative plate, or cathode, of the glass tube. He found that he could move these bands about by moving a bar of magnet near the tube. He also observed that the cathode rays could be made to bend under the influence of a strong

---

[129] ibid., p. 74.

[130] ibid., p. 74.

[131] ibid., p. 74.

[132] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 357 (1989).

electromagnet. Crookes proposed that the movement of the bands showed that the cathode rays that formed the electrical discharge were actually streams of negatively charged particles.[133] For lack of a better name, he termed these cathode rays a fourth state of matter—for they were neither gas, nor liquid, nor solid. This was the best he could do. He could go no farther in identifying what these cathode rays were.[134] But Crook had done a valiant service. He had given humanity a new instrument of discovery. With this instrument, electrons and x-rays would be discovered, as we shall see below.

To investigate the nature of the cathode rays, J.J. Thomsom (1856–1940) constructed cathode ray tubes and evacuated the air inside them with a vacuum pump. As he passed high-voltage electricity through the evacuated glass tube, he observed the glow of cathode rays, just as Crooks had. He wondered at the cause of the bending of a beam of the cathode rays by a magnet. He observed that the beam of the cathode rays was deflected as if the rays were made up of so many iron filings attracted by a magnet.[135]

Next, he varied the conditions of his experiments. He changed the degree of evacuations of the tube. He used a different cathode and altered the intensity of the electricity he passed through the tubes. After analyzing his data, in 1897 J.J. Thomson announced that cathode rays consist of negatively charged particles of electricity torn away from the atoms of the gas in the glass tube. These cathode rays were part of the atom and were alike no matter where they originated. They were ponderable.[136] The name *electron* was coined for these negative particles of electricity by Johnston Stoney, who had calculated an approximate value for its charge in 1874.[137]

To prove the existence of electrons, Thomson set out to determine their mass. He measured the amount of bending suffered by the cathode rays in the presence of magnets of known strengths, and from the results he calculated the ratio of the electric charge of the particle to its mass—"e/m," as it is called. He announced the calculated mass of the electron as 2000 times less than that of an atom of hydrogen (H), the lightest substance hitherto known. (The currently accepted value is about 1/1850 that of the H atom). For this work, Thomson won a Nobel Prize in Physics in 1906.[138]

The central implication of Thomson's discovery of the electron was enormous. It disproved the belief that the ultimate reality of matter is the atom. Since the 1800s, the atom as conceived by Dalton had been regarded as the primordial substance from which every material in the universe was built. The atom had been generally accepted as the indivisible brick of the universe.

---

[133] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, p. 189 (1981).

[134] B. Jaffe, *Crucibles: The Story of Chemistry – from Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 199 (1930).

[135] ibid., p. 200.

[136] ibid., p. 201.

[137] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 357 (1989).

[138] ibid.

And in one fell swoop, Thomson slaughtered yet another sacred cow of chemistry and physics.

C.T.R. Wilson (1869–1959), one of Thomson's students at Cavendish Laboratory of Experimental Physics at Cambridge University, carried out another experiment designed to provide incontestable proof of the existence of electrons. The idea was to condense water vapor around electrons to slow them down long enough to capture and imprint at least one of them on a photographic plate. By themselves, electrons are too small to be photographed, but when tiny water droplets condense on them, they become visible and can be photographed. For this experiment, a high-resolution supercamera was sealed in a glass chamber in which electrons could be produced at will with the aid of a small tube of radium, which readily strips electrons from atoms. Charcoal tubes were attached to the camera and immersed in liquid air because charcoal at this very low temperature absorbs gases readily, and this made it possible to keep the chamber highly evacuated. When the desired conditions were obtained inside the chamber, the photographic plate was lowered by means of a winch into the chamber and into the field of the electrons, and a photograph was taken. Then, the vacuum in the apparatus was destroyed, and the film was removed and developed, revealing a tangled skein of thread representing the path of single electrons after expulsion from their atoms. Thus, Wilson provided incontestable proof of the existence of electrons. For this work, Wilson won the 1927 Nobel Prize in Physics.[139]

Robert Andrews Millikan (1868–1953) of the University of Chicago performed another set of experiments designed to provide incontestable proof of the existence of electrons. His apparatus consisted of two brass plates around one-third of an inch apart. The top plate had a hole the diameter of a very thin needle bored in its center, the space between the plates was illuminated by a powerful beam of light, and electrical connections were made from the brass plates to a battery that supplied 10,000 V of electricity. Using an ordinary commercial atomizer, he sprayed oil droplets of one ten-thousandth of an inch in diameter into the air above the upper plate and watched through the eyepiece of a powerful microscope as one of the neutral droplets found its way through the tiny hole and fell into the space between the plates, eventually landing on the bottom plate. When he reversed the polarity of the plates, he noticed that the motion of the droplet through the space between the plates was not affected.[140]

Next, he held a small tube of radium in the space between the plates so that the radioactive rays from the radium would strike an oil droplet to strip electrons from its atoms as it traveled from the hole in the center of the top plate to the bottom plate. At this point, he observed that the droplet, being

---

[139] ibid., pp. 202–204.

[140] ibid., pp. 204–205.

now positively charged after having lost some of its electrons, slowed down as it fell. By observing the change in speed with which the droplet traveled, he was able to determine how many electrons it had lost. He observed that the droplet always traveled at definite rates of speed, with a characteristic minimum speed. He also observed that he could double or even triple the speed of the fall by the degree of ionization to which he subjected the droplet. From this he could deduce that the slowest speed was the result of the loss of one electron, proving that the smallest invisible load he could remove from the droplet was actually one electron, and that all electrons consist of exactly the same quantity of negative electricity.[141]

From the early days of the discovery of the electron to the present time, a vast amount of scientific information has been uncovered about the electron. These discoveries have spawned the development of the giant field of electronics, which has since developed and expanded on our ever-growing knowledge of this powerful and versatile entity of matter. Thousands of new tools and machines such as the cathode-ray oscillograph, the electron microscope, television, radar equipment, and the computer were all derived from the electron and the discoveries associated with it. The electronic revolution that has defined our modern world derives directly from this discovery of the electron. Electron-beam lithography traces its very origin to this discovery. The modern theory of atomic structure derives also in substantial ways from the discovery of the electron and radioactivity,[142] as we shall see later.

3.4.4 X-rays

While investigating the nature of cathode rays produced in a Crookes tube that he had covered with a shield of black cardboard, ostensibly to prevent the cathode rays from escaping, Wilhelm Conrad Röntgen (1845–1923) (Fig. 3.18) observed in 1895 the luminescent effect on a piece of barium platinocyanide paper lying a considerable distance away. Reasoning that this luminescent effect could only be produced by some invisible ray that was different from the cathode rays but yet emanating from the Crookes' tube, while also penetrating the black cardboard shield, he gave them the "provisional" name of x-rays.[143] For this discovery, Röntgen was awarded the first Nobel Prize in Physics in 1901. Today we know that x-rays are produced when a stream of electrons falls on the metal reflector of a Crooke's tube. Also, the name x-ray is still in use everywhere in the world, except in his native Germany, where they are called Röntgen rays. Röntgen determined many properties of the x-rays and showed that they could pass through glass or wood, through aluminum (although not through lead), and even through

---

[141] ibid.

[142] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 357 (1989); P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, pp. 189–190 (1981).

[143] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, pp. 190–191 (1981).

Figure 3.18 Wilhem Conrad Röntgen (1845–1923), the discoverer of x-rays and of radiography (printed with permission from the Deutsches Museum, Munich).

human flesh. On emerging, the x-rays would make a fluorescent screen glow or a photographic plate turn black.[144]

It was George Stokes who proposed that x-rays were electromagnetic radiation with even shorter wavelengths than ultraviolet radiation. Röntgen himself erroneously thought that they were not transverse but longitudinal waves, like sound, while other scientists thought they were another kind of particle. A decade later, diffraction experiments carried out with crystals confirmed Stokes' view of x-rays as an invisible form of radiation.[145] The roots of x-ray lithography stretch all the way back to this discovery of Röntgen's.

In June 1912, Max von Laue (1879–1960) of the University of Zurich discovered a peculiar property of crystals when he exposed them to x-rays. He found that pure crystals of salt split x-rays as they split light—the minute spaces between the atoms of the crystals act as a diffraction grating and produce an x-ray spectrum. William Henry Bragg and his son William Lawrence Bragg, applying von Laue's discovery, developed a method for determining the inner structure of pure salts. In their method, they passed x-rays through very thin sections of crystals and then photographed the crystals. They found that crystals were made up of regularly spaced atoms. Using mathematical calculations, they generated a real pattern of the crystals in three dimensions, and in so doing, laid the foundation of the modern concept of crystal structure analysis by x-rays.[146]

---

[144] ibid., pp. 190–191.

[145] ibid., p. 195.

[146] B. Jaffe, *Crucibles: The Story of Chemistry – from Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 222 (1930).

While experimenting with directing a stream of electrons on the metal reflector (anticathode) of a Crookes tube and generating x-rays, Charles G. Barkla (1877–1944) determined that x-rays were principally of two kinds. The first kind was due to the stoppage of electrons as they traversed the interior of the metal reflector, and the second was sent out from the anticathode of a Crooke's tube and depended on the metal or metals of which the anticathode was composed. He also determined the length of the penetrating power of these x-rays by absorbing them in very thin sheets of metallic aluminum—an experiment for which he was awarded the Nobel Prize in Physics in 1917.[147]

3.4.5 Radioactivity

Although it had been known for a while that phosphorescent substances after exposure to sunlight become luminous or glow in the dark, very little was known as to the nature of this glow or luminescence. While trying to determine in 1896 whether the luminescence from phosphorescent substances was indeed x-rays (discovered a year earlier by Röntgen), Antoine Henri Becquerel (1852–1908) (Fig. 3.19) made the serendipitous discovery of yet another unexpected phenomenon by revealing the existence of invisible radiation with wavelengths shorter than those of x-rays. Quite by accident and without much thought, he had placed a piece of uranium ore on a sensitized photographic plate lying on a table in his darkroom. When he next checked on the plate after some hours, he observed that in the completely darkened room, without any visible light, the place that was covered with a black paper had

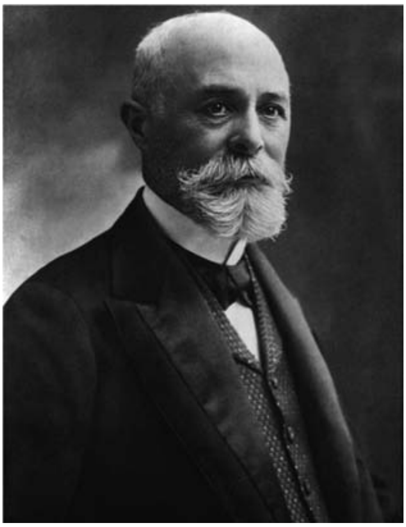


Figure 3.19 Antoine Henri Becquerel (1852–1908), who discovered natural radioactivity (printed with permission from the Deutsches Museum, Munich).

---

[147] ibid., p. 223.

been darkened under the spot on which the ore was placed. When he analyzed the ore to determine the amount of uranium it contained, he noticed that the intensity of the effect was directly proportional to the amount of uranium present in the ore.[148] Becquerel additionally found that crystals of uranium salts could also blacken a photographic plate while it was still in its black paper wrapper.[149]

Next, when he tested a certain chief ore of uranium, pitchblende, a mineral mined in northern Bohemia (present day Czech Republic), he observed that instead of the photographic effect being directly proportional to the amount of uranium present, this particular ore was much more powerful than its uranium content could account for. From this observation, he concluded that in that ore there must be another element besides uranium with power many times greater to darken photographic plates than uranium. At this point, he directed Marie Skłodowska [later known as Marie Curie (1867–1934)] to isolate this then unknown substance from the pitchblende. Working with her husband Pierre Curie (1859–1906), they extracted a small amount of bismuth salt that showed the presence of this very active element, which appeared to be about 300 times as potent as uranium. From the bismuth salt, she isolated this element, which after a battery of tests, she determined to be a new element. She announced the discovery of this new element in 1898 and named it *polonium* in honor of her beloved country, Poland. Also in 1898 and from the same pitchblende, she and Pierre Curie isolated a new element that is one million times more radioactive than uranium, which they called *radium*.[150]

The public marveled at the properties of the newly discovered radium. Its salts were self-luminous; they shone in the dark like bright stars. They continuously emitted a substantial amount of heat, about 250,000 times as much as that produced by the burning of an equal weight of coal. Scientists calculated that a ton of radium would boil 1000 tons of water for a whole year. The element was determined to be the most potent poison known to science—even acting from a distance. A small vial containing a grain the size of a pinhead and placed in the vicinity of the spinal column of a mouse paralyzed the creature in three hours; in seven hours the creature was in convulsions, and in fifteen hours it was dead. When placed near the skin, the element radium produced painful sores in humans.[151]

Subsequent work on this phenomenon of *radioactivity*, as Marie Curie called it, showed that certain elements, particularly radium, can

---

[148] B. Jaffe, *Crucibles: The Story of Chemistry – from Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 185–186 (1930).

[149] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, p. 195 (1981); J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 357 (1989).

[150] B. Jaffe, *Crucibles: The Story of Chemistry – from Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 181–190 (1930).

[151] ibid.

spontaneously release energy. These elements constantly emit positively and negatively charged particles as well as electromagnetic waves of very short wavelength, gamma rays.[152] In 1899 Madam Curie postulated that radioactive atoms are unstable and disintegrate with emission of energy.[153] Another major contributor to the early history of the discovery of radium was Friedrich Oskar Giesel, who in 1899, influenced by the discovery of Madam Curie, succeeded in preparing radiochemically pure radium in reasonable quantities. He, too, observed the emission of α and ß particles from these samples. Giesel is also credited with the first preparation of radiochemically pure actinium in 1902 and with the determination of its atomic number.

Influenced by the work of Curie and Bequerel on radioactivity and the work of Röntgen on x-rays, Ernest Rutherford (1871–1937) conducted experiments on uranium and thorium, and by 1900 he observed that thorium emitted a very minute quantity of gas that was highly radioactive. Conducting additional experiments to elucidate the nature of this gas, he discovered a hitherto unknown element, which he provisionally named thorium "emanation." With the aid of an emanation spectroscope, he made a set up to measure the electric charge on products of radioactive disintegration and was able to determine that the thorium "emanation," also called alpha rays (α rays), was indeed positively charged helium atoms.[154]

In 1903 Rutherford and Frederick Soddy (1877–1956) extended Curie's hypothesis by proposing a new theory of radioactivity: that atoms of radioactive elements are unstable and constantly disintegrating. During this disintegration process, beta particles (β particles, or electrons), gamma rays (γ rays), and positive particles (α rays) are spontaneously liberated by radioactive elements such as radium, uranium, polonium, etc. The nature of this disintegration is such that its rate, dynamics, and extent are unaffected by the application of either extreme cold or intense heat. The Greek Heraklitus was right after all: "One cannot enter the same river twice." In other words, "the only thing constant in nature is change itself." Thus, with radioactivity, we see yet another death blow to the belief in the immutability of the atom.[155]

Knowing the weight of a radioactive element (such as radium salt) that is undergoing disintegration, Rutherford could calculate that element's rate of disintegration. He could determine, for instance, that thirty-five billion atoms of radium disintegrated every second in 1 g of radium, meaning that radium lost its activity at the rate of 1% every 25 years. He determined that radium would have lost one-half of its strength at after 1700 years. The length of time it takes a radioactive element to disintegrate to one-half of its original size

---

[152] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, p. 195 (1981); J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 357 (1989).

[153] ibid.

[154] B. Jaffe, *Crucibles: The Story of Chemistry – from Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 207 (1930).

[155] ibid., pp. 207–208.

is now referred to as its half-life. For uranium, the heaviest of elements, the half-life has been determined to be about six billion years.[156] In 1907 Hans Geiger invented an instrument called a *Geiger counter* for detecting products of radioactive disintegration.

In 1938 Otto Hahn (1879–1968), who spent a year working with Rutherford, discovered atomic fission, for which he received the Nobel Prize in Chemistry in 1944. This achievement changed the world in many ways. It enabled the building of atomic bombs that set the nuclear arms race in motion. It also made it possible to harness nuclear energy for peaceful purposes, such as nuclear power plants for electrical energy generation.

When J.J. Thomson established that the negatively charged β particles emitted during radioactive decay were identical to electrons, Rutherford wondered about the nature of the positively charge α particles. He did know that α particles move at extremely fast speeds and can penetrate thin paper and pass through very thin glass, but can be stopped by the walls of ordinary glass tubes. To investigate the nature of these α particles, he decided to trap them in a glass tube and examine them by means of a highly sensitive spectroscope, which could detect one-tenth of one-millionth of a gram of a metal.[157]

To determine the nature of the α particles generated during radioactive decay, Rutherford filled the extremely thin-walled inner tube of a double-tube assembly with emanation gas from radium, before sealing the inner tube to the outer tube and exhausting the space between the tubes of gases. When after two days he examined the space between the tubes, which earlier had been carefully exhausted of all gases, he found helium gas. And knowing that only α particles could penetrate the thin walls of the inner tube and enter the space between the tubes, he could then definitely conclude that the α particles were indeed positively charged atoms of helium. Rutherford announced this discovery to the world in 1909.[158]

Thus, the identities of the α, β, and γ rays of radioactive substances were established by Friedrich Oskar Giesel, Antoine Becquerel, Pierre Curie, Marie Curie, and Ernst Rutherford between 1899 and 1903.[159] And with this, the grandeur of the whole electromagnetic spectrum could finally be appreciated (see Fig. 3.14). From the short γ rays of radioactivity to the long radio waves, the spectrum spanned more than 50 octaves.[160]

---

[156] ibid.

[157] ibid., pp. 208–209.

[158] ibid., p. 209.

[159] P. Mason, *The Light Fantastic*, Penguin Books, Middlesex, p. 195 (1981); J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 357 (1989).

[160] An octave is the span from any particular frequency to a frequency twice as great. For example, the frequency of violet light is only about 1.8 times that of the deepest red, so the visible spectrum occupies less than a single octave out of the vast span of electromagnetic radiation.

3.4.6 The beginnings of quantum theory

Although the electromagnetic theory, the high point of 19th century physics, explains how light travels through space, it tells nothing about how light is created and destroyed—processes that take place only via light interaction with matter. As a result, the theory of light could not be complete without an understanding of the fundamental nature of matter.[161]

This understanding was finally achieved in the quantum theory of 1925, which for the first time provided an adequate explanation of how matter is constructed of atoms and molecules, how atoms are constructed of nuclei and electrons, and how atoms interact with light. Each of the major developments of 19th century physical science played critical roles in leading up to quantum theory. These developments included electromagnetic theory, molecular theory of matter, and statistical thermodynamics.[162]

Influenced by the electromagnetic theoretical prediction that oscillating electric charges would emit radiation, which was demonstrated when Heinrich Hertz experimented with emitted long-wavelength radiation, prominent physicists of the late 19th and early 20th centuries pursued the idea that the oscillation of individual molecules in a solid could be taken as the source of emitted light.[163] The theoretical study of this problem marked one of the major turning points in the history of science, for it provided the final pieces of the puzzle in the construction of the modern theory of light, as well as the fundamental nature of matter; for these reasons, we will consider in detail the historical developments surrounding quantum theory.[164] Along the way, it resolved important questions dealing with blackbody radiation, photoelectricity, etc., that could not be explained by classical physics.[165]

3.4.7 Blackbody radiation

One of the difficulties encountered by physicists of the late 19th century, the resolution of which overturned the conceptual framework of physics, had to do with the radiation given off by material bodies when they are heated. It is well known that when a metal such as iron is heated in a fire to a very high temperature, say 3000 °C, and then taken out of the flame, it first glows red hot—a self-luminous source. When it is heated even further, to say 6000 °C, the radiation becomes white and even turns blue as the temperature is raised even higher. Thus, one sees that there is a continual shift of the color of a heated body from the red through the white and into the blue as it is heated to higher and higher temperatures.[166] On the basis of frequency, the emitted radiation ranges

---

[161] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 73 (1987).

[162] ibid.

[163] ibid., p. 75.

[164] ibid., p. 76.

[165] ibid., pp. 74–81.

[166] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, pp. 5–6 (1983); M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 74–75 (1987).

from a lower frequency to a higher frequency as the temperature increases. The exact frequency spectrum emitted by the body is dependent on the particular body itself. An ideal body, i.e., one that absorbs and emits all frequencies, is called a blackbody and serves as an idealization for any radiating material. The radiation emitted by a blackbody is referred to as blackbody radiation.[167]

Detailed study of this radiation in the 19th century uncovered the following four facts. (1) The spectrum is continuous like the sun's, including infrared, and it includes all colors of the visible and ultraviolet regions. (2) A plot of intensity versus wavelength (see Fig. 3.20) shows maximum intensity at one wavelength and decreasing intensity at both higher and lower wavelengths. (3) The hotter the object the shorter the wavelength of a maximum intensity. (4) The hotter the object is, the greater is the total amount of radiation emitted from a given surface area. As an example, at 3000 °C, the maximum intensity occurs in the infrared, but the tail of the spectrum stretches into the red region of the visible region of the spectrum, hence we see the object as "red hot." At a lower temperature, the maximum intensity falls within the far-infrared region, which we cannot see but can feel as radiant heat. At 6000 °C, the maximum wavelength falls within the blue region, with substantial radiation stretching into the entire visible region. The distribution (variation of intensity with wavelength) looks very much like that of the sun, and we see the object as "white hot."[168]

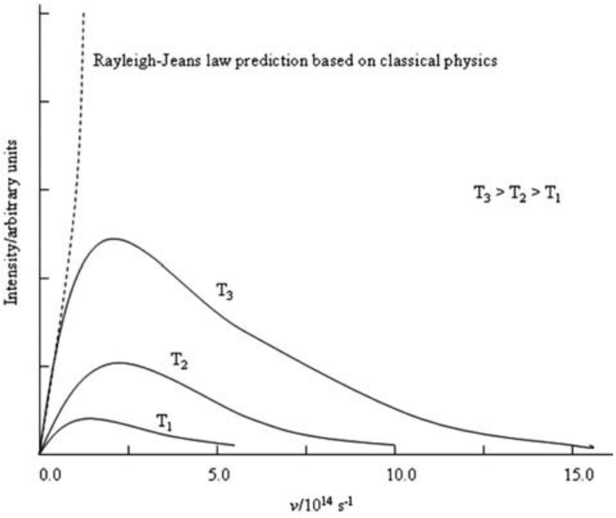


Figure 3.20 Spectral distribution of the intensity of blackbody radiation as a function of frequency for several temperatures. The dashed line is the prediction of classical physics. (Plot generated in the manner of McQuarrie.)[169]

---

[167] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, pp. 5–6 (1983).

[168] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 75 (1987).

[169] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, p. 6 (1983).

Many physicists of the late 19th century tried to derive expressions consistent with the experimental intensity-versus-frequency curves for several temperatures shown in Fig. 3.20, but without success. In fact, the Rayleigh–Jeans law, shown in Eq. (3.11), gives the expression that is derived according to the laws of 19th century physics:[170]

$$\rho(\nu,\, T)d\nu = \frac{8\pi kT}{c^3}\nu^2 d\nu, \tag{3.11}$$

where $\rho(\nu, T)d\nu$ is the density of radiative energy between the frequency ν and $\nu + d\nu$, k is the Boltzmann constant, T is the absolute temperature, and c is the speed of light. The dashed line indicated in Fig. 3.20 is the prediction of the Rayleigh–Jeans law (Fig. 3.21 shows a picture of Lord Rayleigh).

It can be seen that the Rayleigh–Jeans law reproduces the experimental data fairly well at low frequencies. However, at high frequencies, the Rayleigh–Jeans law diverges as ν^2. Since the frequency increases in the ultraviolet region of the spectrum, this divergence was called the "ultraviolet catastrophe," a phenomenon that classical physics was unable to explain theoretically. This was the first of such phenomena to be observed in physics and did in fact mark a major milestone in the annals of physics.[171]

Figure 3.21 Lord John William Strutt Rayleigh (1842–1919), a very influential physicist, who made significant contributions to optical and radiation physics. The optical resolution criterion, which is so central in lithography for judging the capability of an exposure tool, is named after him.

---

[170] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, pp. 6–7 (1983).
[171] ibid., p. 7.

3.4.8 Planck's quantum hypothesis for the blackbody radiation law

The German physicist Max Planck (1858–1947) (Fig. 3.22) in 1900 was the first person to offer a successful explanation of blackbody radiation. He, too, like Rayleigh and Jeans earlier, assumed that the radiation emitted by the body was due to the oscillations of the electrons in the constituent particles of the material body. He pictured these electrons as oscillating in an atom much like electrons oscillate in an antenna to give off radio waves, as Hertz had demonstrated. In the case of the atomic antennae, however, the oscillations occur at much higher frequencies and hence are associated with frequencies in the visible, infrared, and ultraviolet regions rather than in the radio-wave region of the spectrum. Implicit in the Rayleigh–Jeans' derivation is the assumption of the principle of continuity—a principle so obvious and axiomatic that it lies unspoken in the heart of classical physics—namely, that physically observable variables such as the energies of the electronic oscillators responsible for the emission of radiation can take on a continuum of values. Planck's great insight was the realization that he had to move away from this mode of thinking in order to derive an expression that would fit experimental data like those shown in Fig. 3.20. He made the revolutionary assumption—one of the most daring and momentous departures in the history of physics—that the energies of the oscillators had to be proportional to an integral multiple of the frequency or, in an equation, that

$$\varepsilon = nh\nu, \tag{3.12}$$

Figure 3.22 Max Karl Ernst Ludwig Planck (1858–1947), who in 1900 proposed the quantum hypothesis, which fully explains the physics of blackbody radiation (printed with permission from the Deutsches Museum, Munich).

where n is an integer, h is a proportionality constant, now called Planck's constant, and ν is the frequency.[172] Using statistical thermodynamic arguments, he was able to derive the expression [Eq. (3.13)], now called the Planck distribution law[173] for blackbody radiation, where all of the symbols except h have the same meanings as in Eq. (3.12). The only adjustable constant is h. Planck showed that

$$\rho(\nu,\ T)d\nu = \frac{8\pi h\nu^3}{c^3}\frac{d\nu}{c^{h\nu/kT} - 1} \tag{3.13}$$

gives excellent agreement with the experimental data for all frequencies and temperatures if h is assigned the value of 6.626×10^{-34} J-s (joule-seconds).[174] Remarkably, he was initially reluctant to claim the physical reality of this quantization of energy.

3.4.9 Einstein's quantum hypothesis for the photoelectric effect

While he was carrying out the experiments that confirmed Maxwell's theory of the electromagnetic nature of light in 1886 and 1887, Heinrich Hertz first observed the photoelectric effect when he discovered that ultraviolet light causes electrons to be emitted from a metallic surface. The ejection of electrons from the surface of a metal by radiation is termed the photoelectric effect. Two experimental observations of the photoelectric effect cannot be adequately explained with the classical wave theory of light. First, the kinetic energy of the ejected electrons is independent of the intensity of the incident radiation, contrary to the prediction of classical physics. According to classical physics, electromagnetic radiation is an electric field oscillating perpendicular to its direction of propagation, whose radiation intensity is proportional to the square of the electric field. Therefore, as the intensity increases, so does the amplitude of the oscillating electric field. The surface electrons of the metal oscillate along with the field so much that, as the intensity (amplitude) increases, the electrons oscillate more violently, eventually breaking away from the surface with a kinetic energy that depends on the amplitude (intensity) of the field. Second, there is a threshold frequency $\nu_{0,}$ characteristic of the metallic surface, below which no electrons are ejected, and above which, the kinetic energy of the ejected electrons is proportional to the frequency ν—a clear contradiction of classical physics, which predicts that

---

[172] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, p. 7 (1983); M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 77 (1987).

[173] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, pp. 7–8 (1983); M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 77 (1987).

[174] ibid., pp. 7–12.

Figure 3.23 Albert Einstein (1855–1905), who among many accomplishments, explained the photoelectric effect and Brownian motion, and proposed the special and general theories of relativity (printed with permission from the Deutsches Museum, Munich).

the photoelectric effect will occur for any frequency of light as long as the intensity is sufficiently high.[175]

In order to interpret these results, the German physicist Albert Einstein (1879–1955) (see Fig. 3.23), in about 1905, applied Planck's hypothesis but extended it in an important way. In applying his energy quantization concept $\Delta\varepsilon = h\nu$ to the emission and absorption mechanism of the atomic electronic oscillators, Planck believed that once the light energy was emitted, it behaved like a classical wave. Einstein, on the other hand, suggested that the energy of a light beam is not evenly spread over the whole beam, but is concentrated in certain regions called photons.[176] In other words, the radiation itself exists as small packets of energy such that $\varepsilon = h\nu$.[177]

Einstein photons are light particles—they are localized concentrations of mass, momentum, and energy, and they are propagated like particles. There is usually a very large number of them in a light beam, the energy in any one photon being very small.[178]

---

[175] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, p. 12 (1983); R.W. Ditchburn, *Light*, Dover Publications, New York, pp. 13–14, 645–646 (1991); M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 78–81 (1987).

[176] Gilbert Luis coined the term *photon* in 1926.

[177] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, p. 12 (1983); R.W. Ditchburn, *Light*, Dover Publications, New York, pp. 13–14, 645–646 (1991); M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 78–81 (1987).

[178] R.W. Ditchburn, *Light*, Dover Publications, New York, p. 13 (1991).

Applying the conservation of energy argument, Einstein showed that the kinetic energy of the ejected electrons [(½)mu^2, where m is mass, and u is velocity] equals the energy of the incident radiation ($h\nu$) minus the minimum energy required to remove an electron from the surface of the particular metal (ϕ):

$$\frac{1}{2}mu^2 = h\nu - \phi, \tag{3.14}$$

where ϕ, called the work function of the metal, is similar to an ionization energy. The left-hand side of Eq. (3.14) cannot be negative, which implies that $h\nu \geq \phi$. The minimum frequency ν_0 that will eject an electron is just the frequency required to overcome the work function of the metal and is given by[179]

$$h\nu_0 = \phi. \tag{3.15}$$

The picture of the photoelectric effect that emerges is one in which, after the collision of the photons with the electrons of the metal, the photon energies are absorbed by the electrons, resulting in the destruction of the photons and emission of the electrons from the metal. In escaping from the metal, the electron may lose some of that energy as it overcomes the forces that bind it to the metal. The electrons ejected in the photoelectric current are thus observed to have a range of energies; those that were originally least strongly bound in the metal (i.e., those that were closer to the surface) have the maximum energy. The maximum energy of these electrons depends only on the frequency of the light, which is consistent with the hypothesis that light deposits a fixed amount of energy on each liberated electron, the quantity of energy being proportional to the frequency of the light.[180]

Further work on the photoelectric effect was done in 1916 by the American physicist Robert Mulliken (1896–1986), who succeeded in verifying Einstein's theory in detail. Specifically, he determined the value for Planck's constant h that is in agreement with the value obtained from blackbody radiation.[181] Resist imaging in EUV lithography is based on this photoelectric effect.

3.4.10 Bright and dark line spectra

The origin of spectroscopy lies in the spectral analytical work of Thomas Melville and the development of the spectroscope by William Hyde Wollaston (1766–1828) and Josef von Fraunhofer (1787–1826) (see Fig. 3.24). Melville's observation in 1752—that the spectrum of light of a flame passed through a

---

[179] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, p. 13 (1983).

[180] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 80 (1987).

[181] ibid., pp. 80–81.

Figure 3.24 Josef von Fraunhofer (1787–1826), the inventor of the diffraction grating and the spectroscope, and one of the founders of spectroscopy (printed with permission from the Deutsches Museum, Munich).

prism is not continuous like that of the sun, but rather that parts of the spectrum are bright while others are dark, with the locations of the bright spots and their colors being different when different chemicals were placed in the flame—provided yet another clue to the nature of the atomic structure.[182]

The Fraunhofer's spectroscope was constructed to receive light that had been passed through a slit and a tube to create a narrow beam, resulting in a spectrum beyond the prism that was then viewed through a telescope. This instrument allowed Fraunhofer to construct for the first time a map of the dark line spectra from the sun, the principal lines from which he named with the letters A through G—the now-familiar Fraunhofer lines. In 1814 and 1815, he published his sun and stellar studies, which gave spectral analysis a real boost.[183]

The key to understanding the dark lines, also called absorption lines, came when it was discovered that the frequencies at which they occurred were the same as the frequencies in the bright line spectrum of glowing gases. Gustav Kirchhoff (1824–1887) had shown in 1859 that he could produce dark lines in the spectra of a blackbody by passing that light through a region of cool gas, which led him to deduce that such a process was responsible for the absorption spectrum of the sun. Although the inner core of the sun (photosphere) emits a continuous spectrum, by the time the radiation traverses the outermost layer of the sun's body, known as the chromosphere, which

---

[182] ibid., p. 81.

[183] ibid.; M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 15 (1991).

contains rarefied hot gases, the elements contained in these hot gases absorb and scatter characteristic wavelengths corresponding to elements in them. This results in the dark line spectra of the sun monitored on Earth. In other words, the spectrum of the transmitted light from the sun contains dark lines at only those frequencies that are absorbed by hot gases surrounding the sun.[184]

Kirchhoff also formulated the law that now bears his name—Kirchhoff's law—which states that all substances absorb the same light frequencies that they can emit. This law has found widespread application in the study of composition of the sun and other stars many light years away,[185] as well as in the rational design of resist sensitizers and photoactive compounds used in lithography.

It has been determined that the sun contains vapors of sodium, calcium, iron, and many other elements.[186] Working together in Heidelberg, Kirchhoff and Robert Wilhelm Bunsen (1811–1899) (see Fig. 3.25) established that each kind of atom has its own signature in a characteristic array of spectral lines.[187]

Fraunhofer was also the inventor of the diffraction grating, a series of parallel wires separated by variable distances as small as 0.005 cm, which acted like an enhanced version of Young's two-slit experiment and could also

Figure 3.25 Gustav Robert Kirchhoff (1824–1887) and Robert Wilhem Bunsen (1811–1899), who, working together, made important contributions to the spectroscopic analysis of substances (printed with permission from the Deutsches Museum, Munich).

---

[184] G. Gamow, *The Great Physicists from Galileo to Einstein*, Dover Publications, New York, p. 123 (1961).

[185] ibid.

[186] The power of this law is aptly illustrated in the discovery of the element helium, which was first seen via a series of unexplained dark lines in the sun's spectrum—lines that were not related to emission lines then known on Earth. Later, helium gas was found on the earth.

[187] E. Hecht, *Optics*, 2nd ed., Addison-Wesley Publishing Co., Menlo Park, p. 10 (1990).

be used as a replacement for the prism. Light passing through the spaces between the wires forms an interference pattern beyond the grating, and the direction of constructive interference is different for each wavelength.[188]

In the 1820s, David Brewster (1781–1868), Sir John Herschel (1792–1871), and William Henry Fox Talbot (1800–1877) performed experiments on the absorption of light that were similar to those performed by Fraunhofer. Independently, these researchers made the observation that flame spectra obtained by burning compounds of different compositions varied in systematic ways. These three researchers all produced line spectra, which they tried to relate to the dark lines observed by Fraunhofer from his solar spectrum. Although there was some agreement that the spectra contained chemical information characteristic of the analyzed flame, there was no consensus on the role of light in the production of these spectra.[189] It was not until the early 20th century that it was discovered that these line spectra are not the result of light interaction with specific compounds but rather are related to their atomic structures and, therefore, to chemistry.[190]

As already stated, a common characteristic of all spectra of flames and electric discharges in gases analyzed in the spectroscope in the 19th century was the presence of bright lines located at positions characteristic of particular wavelengths. Each chemical element emits a unique set of wavelengths, which made it possible for the spectroscope to become a device that could be used to identify an element, just as a fingerprint identifies an individual.[191] The body of data on bright line spectra of gases obtained during this period presented a serious challenge to physicists and chemists that was as great as that of blackbody radiation—a challenge which in some ways was greater since the continuous emission from hot solid objects is largely independent of the nature of the object, whereas the line spectrum is different for each element.[192]

Being the simplest form of matter (since each atom or molecule is independent and not affected by forces due to the other atoms), a gas therefore made an ideal candidate for studying line spectra. The hydrogen atom being the simplest atom—as revealed from the chemical experiments and observations of John Dalton (1766–1844), Amadeo Avogadro (1775–1856), Joseph-Louis Gay-Lussac (1778–850), and others (described in Chapter 4)—was therefore chosen and its line spectrum carefully studied. Between the years 1909 and

---

[188] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 81 (1987).

[189] M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 15 (1991).

[190] Line spectra do not give any information about the physics of light; rather, the lines are related to transitions between various energy levels in the atoms making up the analyzed compound. The elucidation of what gives rise to the lines in line spectra had to wait until after the introduction of quantum mechanics and the work of Niels Bohr in the early decades of the 20th century.

[191] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 82 (1987).

[192] ibid., p. 83.

1912, the first successful theory of the hydrogen atom and its line spectrum was formulated based on the experimental work of Rutherford and the theoretical insights of Niels Bohr (1885–1962).[193]

3.4.11 Atomic structure

An accurate picture of the nature of an atom's structure began to emerge toward the end of the 19th century, aided in large part by the discovery of the electron, x-ray, radioactivity, and the nucleus and its contents, as well as the formulation of quantum mechanics. Before this time, many scientists made numerous attempts to explain the nature of the atom, which were later found to be not entirely accurate, or incomplete at best. Examples included Lord Kelvin, who pictured the atom as consisting of mobile electrons embedded in a sphere of positive electrification. J.J. Thompson had also postulated an idea similar to Lord Kelvin's, but his model also failed because it could not account for many contradictory phenomena. Rutherford's nuclear model of the atom (see next subsection) as a solar system was found to be incomplete. The main obstacle to accepting these models was that each lacked a consistent explanation of the peculiar spectra of gaseous elements when heated to incandescence.[194]

Even before the electron was discovered, Hendrik A. Lorentz (1853–1928) had postulated that spectral lines were the result of the motion of electrified particles revolving around the nucleus of the atom. He went on to predict an effect that was later observed by Pieter Zeeman (1865–1943) in 1896. Zeeman had observed that when incandescent gases are introduced into the vicinity of powerful magnetic fields, their spectral lines are split. Johannes Stark (1874–1957) showed a similar but more pronounced split 17 years later when he placed incandescent gases in strong electrical fields. All of these results point to something electrical in the atom being the cause of the spectrum of elements.[195]

In 1885 Johann Jakob Balmer (1825–1898) of Switzerland had observed that the positions of the lines in the complicated spectrum of hydrogen were not as chaotic as they seem. He showed that he could represent wavelengths by a simple mathematical formula involving a number obtained by subtracting two figures. Other scientists, for example, Johannes Rydberg (1854–1919), obtained similar remarkable numerical relationships.[196]

---

[193] ibid.

[194] B. Jaffe, *Crucibles: The Story of Chemistry – from Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 242 (1930).

[195] ibid.

[196] ibid., pp. 242–243.

3.4.12 Nuclear model of the atom

In 1911 Rutherford proposed the nuclear model of the atom, which consists of a very small, positively charged nucleus surrounded by outer electrons at relatively large distances from the nucleus. He showed that this nuclear model can explain the very large deflections sometimes suffered by α particles in passing through matter.[197] What holds the electrons in place is an attractive electrical force between each electron and the nucleus, and this force is what binds the atom together. This implies that the electrons cannot be stationary, for if they were, they would fall into the nucleus. Rather, just as the earth orbits the sun under the influence of attractive gravitational forces, the electron orbits the nucleus. This is now referred to as Rutherford's planetary model of the atom.[198]

A major prediction of this model was that stationary electrons do not exist since they would invariably be attracted by and fall into the nucleus of the atom unless their enormous speed around the center of the atom effectively counteracted the powerful pull of the nucleus of the atom. The model also predicted that electronic motion was the cause of spectral light. Contained in the very kernel of Rutherford's model was this puzzle: if electronic motion is the cause of spectral light, then Rutherford's atom ought to continuously radiate this light. However, experimentally, the spectral light was not continuous, but rather discrete. Thus a major impasse was reached that could not be explained by the known laws of classical electrodynamics.[199]

3.4.13 Bohr's model of the hydrogen atom

Realizing that Rutherford's planetary model of the atom is incompatible with Maxwell's classical theory of radiation—which stipulates that a charged electron in circular motion will continually emit radiation and thereby lose energy, and its orbit will shrink and quickly spiral into the nucleus—Niels Bohr[200] in 1913 (see Fig. 3.26) asserted that an electron in an atomic orbit simply does not radiate; in other words, Maxwell's theory does not apply at this level. Bohr's main contribution was to make two nonclassical assumptions.[201] First, he assumed the existence of stationary electron orbits, in defiance of classical physics. In other words, he postulated that, like the oscillators that produce blackbody radiation, the electron can exist in only certain quantized energy levels, and as long as it stays in a given level,

---

[197] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 360 (1989); M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 82–85 (1987).

[198] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 85 (1987).

[199] B. Jaffe, *Crucibles: The Story of Chemistry – from Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 242–243 (1930).

[200] N. Bohr, "On the constitution of atoms and molecules," *Phil. Mag.* **xxvi**, pp. 476, 857 (1913).

[201] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, p. 23 (1983).

Figure 3.26 Niels Bohr (1885–1962), who in 1913 postulated the Bohr model of the hydrogen atom, from which derives the notion of electronic shells. This notion initiated the modern atomic theory of chemical reactions, as well as the concept of energy levels that underlie the design of lasers (printed with permission from the Deutsches Museum, Munich).

it produces no light and its energy does not change. He specified these orbits by invoking a quantization condition, where he assumed that the angular momentum l of the electron must be quantized according to

$$l = mur = n\hbar \quad n = 1, 2, \ldots, \tag{3.16}$$

where m is the mass of the electron, u is the velocity of the electron, r is the radius of the electronic orbit, and $\hbar = h/2\pi$, where h is Planck's constant. The force holding the electron in a circular orbit is the coulombic force ($e^2/4\pi\varepsilon_0 r^2$) of attraction between the proton and the electron, whereas the centripetal force (mu^2/r) keeps the electron moving in the circular orbit. An electron in a stationary circular orbit therefore experiences a coulombic force that is counterbalanced by the centripetal force due to its rotation around the nucleus. Equating these two forces, we obtain

$$\frac{e^2}{4\pi\varepsilon_0 r^2} = \frac{mu^2}{r}, \tag{3.17}$$

where e is the charge on the electron, and ε_0 is the permittivity of free space, equaling $8.85419 \times 10^{-12}\ \mathrm{C^2 N^{-1} \cdot m^{-2}}$. Solving Eq. (3.17) for u and substituting into Eq. (3.20), we obtain

$$r = \frac{4\pi\varepsilon_0 \hbar^2 n^2}{me^2} \quad n = 1, 2 \ldots, \tag{3.18}$$

which shows that the radii of the allowed orbits, now called Bohr orbits, are quantized. The value of the Bohr radius corresponding to $n = 1$ is the radius of the hydrogen atom in its ground state. Under the above-described scenario, the electron can move around the nucleus only in circular orbits with radii given by Eq. (3.18).

The total energy of the electron in the hydrogen atom, being the sum of the kinetic and potential energies, is therefore given by

$$E = \frac{1}{2}mu^2 - \frac{e^2}{4\pi\varepsilon_0 r}, \tag{3.19}$$

where the negative sign in front of the potential energy term indicates that the proton and electron attract each other, their energy being less than it is when they are infinitely separated. Using Eq. (3.18) to eliminate the mu^2 in the kinetic energy term, Eq. (3.19) becomes

$$E = \frac{1}{2}\left(\frac{e^2}{4\pi\varepsilon_0 r}\right) - \frac{e^2}{4\pi\varepsilon_0 r} = -\frac{e^2}{8\pi\varepsilon_0 r}. \tag{3.20}$$

Since the only allowed values of r are those given in Eq. (3.18), substituting Eq. (3.18) into Eq. (3.20), we obtain the expression for the only allowed energies:

$$E_n = -\frac{me^4}{8\varepsilon_0^2 h^2}\frac{1}{n^2} \qquad n = 1, 2, \ldots . \tag{3.21}$$

It should be pointed out that the negative sign in this equation is an indication that the energy states are bound states; the energies given by Eq. (3.21) are less than those when the proton and electron are infinitely separated. The state of the lowest energy for the hydrogen atom corresponds to $n = 1$ and is called the ground state. States with $n > 1$ have higher energies than the ground state and are called excited states. An electron, Bohr asserted, can spontaneously fall from a higher energy level to a lower one, and when it does, it releases an amount of energy that is the difference between the two energy levels. This energy is lost by the atom and goes to create a quantum of light. In this way, light is the child of an atom in the process of change and, in another way, light is the mechanism of the atomic process; if there were no light, the atom would not be able to go from one state to another.[202]

Bohr also made the assumption that the observed spectrum of the hydrogen atom is due to transitions from one allowed energy state to another, with the energy difference between two levels given by

---

[202] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 85–86 (1987).

$$\Delta E = \frac{me^4}{8\varepsilon_0^2 h^2}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right) = h\nu. \tag{3.22}$$

The equation $\Delta E = h\nu$ is called the "Bohr frequency condition" and is the basic assumption that, as the electron falls from one level to another, the energy evolved is given off as a photon of energy $E = h\nu$. The energies and energy differences are quantized, as are the frequencies.[203] The isolated atom emits light of only certain definite frequencies, as observed in the bright line spectrum.[204]

The equations of Bohr's theory are in agreement with the observed frequencies in the hydrogen spectrum, as are the observed spectral series: Lyman series (when electrons excited to higher levels relax to the $n = 1$ state) and Balmer series (when electrons excited to higher levels relax to the $n = 2$ state, and so on). Working backwards, the observations can also be used to determine the value of Planck's constant. The value obtained in this way was found to be in agreement with the result deduced from blackbody radiation and the photoelectric effect.[205]

3.4.14 Implications of Bohr's theory

Between 1912 and 1925, Bohr's theory of the atom gave rise to a conceptual framework for the study of matter on many fronts. Applying the theory to predict the energy levels and therefore the emission frequencies of atoms more complicated than hydrogen—say helium with two electrons, lithium with three, and so on—led to the concept of electronic shells about the nucleus, the outer shell being less tightly bound, and it is these outer-shell electrons that determine the element's spectrum.[206] It should be pointed out that the theory was not fully successful in predicting the spectral lines of elements that are more complicated than hydrogen.

Furthermore, the concept of electronic shells initiated the modern atomic theory of chemical reactions. It is now well accepted that the electrons in the outer shells, transferring from one atom to another or being shared by two atoms, provide the machinery for chemical reactions and the binding of atoms to form molecules.[207]

The concept of atomic energy levels also underlies the design of lasers, which require a method to excite the electrons in atoms—a method that could

---

[203] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, p. 25 (1983)

[204] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 86–87 (1987).

[205] ibid., p. 87.

[206] ibid., pp. 89–90.

[207] ibid., p. 90.

hardly have been discovered by chance.[208] Einstein predicted that light emission could be produced if the atoms were so highly energized that the majority of the electrons at a certain energy level were excited into a higher level. In this scenario, a photon emitted from an excited atom ‘collects’ another photon from a neighboring excited atom and travels along with it, in the same phase and amplitude and in exactly the same direction. Next, these two photons in turn collect two more, and in this way a process that amplifies the number of photons is created that proceeds through the laser cavity, whether in the form of a solid crystal rod or a long tube of gas. By positioning a mirror at each end of the tube to reflect the light back and forth, an enormous intensity of these photons can be built up in a very short time. Making one of the mirrors partly reflecting allows a narrow, parallel, intense laser beam to emerge from it[209] in a new form of light that is now used in many fields—communications, surveying, range finding and welding, medical surgery, and even as light sources for lithography.

Despite its successes, Bohr’s theory did not resolve the fundamental issues that lie at the heart of the quantum theory of Planck and Einstein: What is behind the quantization of oscillators and atoms? What happens when a system changes its state, resulting in the creation of a photon? How can the generated photon behave like a particle under certain conditions and like a wave under other conditions?

3.4.15 Quantum theory of light

The answers to the above questions, not all of which need to be presented here, were formulated between 1925 and 1926 in the revolution of modern quantum theory, which shook the foundations of physics and philosophy. Remarkably, the central theme of quantum theory was the nature of light and what came to be called the “wave–particle duality.” We briefly summarize here the underpinnings of this wave–particle duality, aspects of which we have covered in the preceding sections of this chapter.

The particle theory and wave theory of light were first postulated by Isaac Newton and Christian Huygens, respectively, in the 17th century and were later expanded upon by their successors. The particle theory held that light consists of discrete particles too small and too light to be measured. This theory was generally accepted until the beginning of the 19th century when Thomas Young and later Augustin-Jean Fresnel (1788–1827), working independently, demonstrated that, under certain conditions, rays of light

---

[208] The advent of the laser is one of those rare instances in scientific research, such as the development of nylon, atomic fission, or genetic engineering, where important technologies have sprung directly from scientific theory.

[209] P. Mason, *Light Fantastic*, Penguin Books, New York, p. 232 (1981).

could be made to interfere destructively with each other to produce darkness or constructively to produce areas of brightness. This phenomenon of interference could not be adequately explained by the particle theory, but it could be explained by the wave theory of light. Thus, the wave theory of light became entrenched for a full century until it, in turn, received a jolt.[210]

Wilhelm Hallwachs (1859–1922) in 1888 observed that a beam of light falling on a negatively charged body produces a rapid loss of the charge, while a non-electrified body gradually becomes positively charged in the presence of light. No adequate explanation could be offered for this observation until two years later, after the discovery of the electron. Both J.J. Thomson and Philipp Lenard (1862–1947) observed that the emission of negatively charged particles or electrons from the surface of a metal was responsible for the phenomenon Hallwachs observed. Three years later, Lenard showed that the speed of the emitted electrons was dependent only on the color or wavelength of the light, regardless of the intensity. Subsequently, it was demonstrated that the number of electrons emitted from the metal depends on the intensity of the light. Thus, another paradox about light was revealed: How can light made up of waves in the ether release electrons from the atoms of metals? The wave theory of light had no answer to this question.[211]

In 1905 Albert Einstein theorized that the emission of electrons from metals by light could be explained if one assumes that light consists not of waves but of photons or concentrated bundles of energy traveling at the speed of light. As each photon or particle of radiation strikes the metal surface, it gives up its energy to an electron, dislodging it from its atom. This is the theory of the "photoelectric effect"[212] (discussed in Section 3.4.9).

That light possesses such contradictory properties—behaving as a wave under certain conditions and as a particle under other conditions—led some to wonder whether matter itself possesses such duality in its properties. Louis de Broglie (1892–1987) (Fig. 3.27), in his study of blackbody radiation and light quanta, was the first to behold an imperative need for a new concept of matter that would be applicable to the structure of the atom as conceived by both chemists and physicists, and at the same time would provide an explanation of the paradox of light. He gave the first inkling of this concept in 1924 in his doctoral dissertation, as well as in a paper published the same year, in which he suggested that the electron was not altogether a particle of electricity, but was composed of, possessed, or perhaps attended by, a group of waves called "matter waves" that guided its path.[213] In other words, he stated that particles

---

[210] B. Jaffe, *Crucibles: The Story of Chemistry – from Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 250–251 (1930).

[211] ibid., p. 251.

[212] ibid.

[213] ibid., p. 252.

Figure 3.27 Louis de Broglie (1892–1987), who in his 1924 doctoral thesis postulated that particles may also have wavelike properties (printed with permission from of the Deutsches Museum, Munich).

such as electrons may also possess wavelike properties[214] and that these wavelike properties would manifest themselves only in phenomena occurring on an atomic scale, as dictated by Planck's constant. He also postulated that the wavelength of these matter waves, for a given particle such as an electron or proton, would be inversely proportional to the particle's momentum p, which is a product of its mass m and speed v:[215]

$$\lambda = \frac{h}{mv} = \frac{h}{p}. \tag{3.23}$$

Experimental confirmation of de Broglie's hypothesis was soon obtained in 1927 through the observation that electron beams can be diffracted in much the same way as x-rays.[216]

---

[214] De Broglie's hypothesis sounds rather strange at first sight, but it does suggest a beautiful symmetry in nature. Surely, if light can be particlelike at times, why should matter not be wavelike at times? Realizing that Einstein had shown from relativity theory that the momentum of a photon is given by $p = h/\lambda$, de Broglie argued that both light and matter obey the equation $\lambda = h/p$.

[215] M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 92, 85–86, 375–377 (1987); D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, pp. 30–35 (1983).

[216] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 92 (1987).

It must be noted that today de Broglie's hypothesis has practical applications not only in electron microscopy, but also in electron-beam lithography, both of which use the wavelike property of electrons. The wavelength of the electrons is easily controlled with the aid of an applied voltage, and the small de Broglie wavelengths attainable offer a more precise probe (in microscopy applications) or higher resolution (in lithography applications) than systems using visible and UV photons. In addition, in contrast to electromagnetic radiation of similar wavelengths (x-rays and UV), the electron beam can be easily focused by using electric and magnetic fields.[217] Electron microscopes have found wide applications in chemistry and biology, where they are used to investigate atomic and molecular structures, as well as in semiconductor industry to inspect and review IC features.

Building on de Broglie's hypothesis,[218] theorists such as Erwin Schrödinger (1887–1961), Werner Heisenberg (1901–1976), Max Born (1882–1970), and others developed a new and fundamental theory.

In 1926, Erwin Schrödinger introduced the mathematical treatment called "wave mechanics," which treated the atom as a region permeated with waves. In specific terms, Schrödinger modified the classical wave equation describing a 3D standing wave by imposing wavelength restrictions suggested by de Broglie's hypothesis. He obtained a second-order differential equation for a wavefunction ψ with second-order differential coefficients with respect to the coordinates x, y, and z for a particle of mass m moving in a potential field U. The equation is of the form[219]

---

[217] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, p. 34 (1983).

[218] De Broglie's hypothesis today not only has practical applications in electron microscopy and electron-beam lithography, but it was also a key step in our understanding of atomic structure. It can be seen as a simple, yet physical, argument for the quantized Bohr orbit if one assumes that, as the electron revolves around the proton, it has a wavelength λ associated with it. For the orbit to be stable, the wave must be in phase as the electron makes one complete revolution. Otherwise, there will be cancelation of some amplitude upon each revolution, leading to disappearance of the wave. For the stability condition of the wave pattern around an orbit requires that an integral number of complete wavelengths fit around the circumference of the orbit. Because the circumference of a circle is $2\pi r$, we have the quantum condition: $2\pi r = n\lambda$ ($n = 1, 2, 3, \ldots$). Substituting de Broglie's relation into this equation leads to the Bohr quantization condition: $mvr = n\hbar$. [See also, D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, p. 34 (1983)].

[219] K.W. Whitten and K.D. Gailey, *General Chemistry with Quantitative Analysis*, Saunders College Publishing, Philadelphia, p. 87 (1981); J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 374 (1989); D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, pp. 77–101 (1983); This equation does not contain time and is therefore called the time-independent Schrödinger equation, the solution of which yields stationary-state wave functions. Although there is a more general Schrödinger equation that is time-dependent, a great number of problems of chemical and lithographic interest can be described in terms of stationary-state wave functions.

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} + \frac{8\pi m}{h^2}(E - U) = 0, \tag{3.24}$$

where h is Planck's constant, and E represents the total energy of the particle.

Equation (3.24) also contains a term, which, in the ordinary equation for wave motion, contains the wavelength λ. In this case, λ is taken as the wavelength defined by de Broglie's equation [Eq. (3.23)] and is contained in the fourth term of the Schrödinger equation [Eq. (3.24)]. The assumption is made that ψ is finite, continuous, and single-valued in general for all values of the coordinates, and that the square of the wavefunction ψ^2 gives the density of charge at each point, the knowledge of which gives an indication of the directions in which bond formation may occur.[220] Schrödinger's wave mechanics confirmed the work of de Broglie and even predicted some facts connected with spectral lines that were not foreseen by Bohr's model.

Werner Heisenberg proposed a new approach to quantum mechanics based on "matrix mechanics," in which the structure of the atom was represented by mathematical formulas directly connected with the frequencies and intensities of the spectral lines—phenomena that could be observed and measured.

Furthermore, in pondering the implications of the wave–particle duality, Werner Heisenberg was led in the mid-1920s to consider a problem on the measurement of the position of an electron in the following way. If we wish to locate the electron within a distance Δx, then we must use light with a wavelength at least that small since, for the electron to be "seen," a photon must interact or collide in some way with it; otherwise, the photon will just pass right by and the electron will appear transparent. The photon has a momentum $p = h/\lambda$, some of which will be transferred to the electron during the collision, implying that the very act of locating the electron necessarily leads to a change in its momentum. Furthermore, if we wish to locate the electron more accurately, we must use light with a smaller wavelength, which implies that the photons in the light beam will have greater momentum because of the relation $p = h/\lambda$. Given that some of the photon's momentum must be transferred to the electron in the process of locating it, the momentum change of the electron therefore effectively becomes greater.[221]

A careful analysis of this problem led Heisenberg to show that it is impossible to determine exactly how much momentum is transferred to the electron. This implies that the act of locating an electron within a region Δx introduces an uncertainty in the momentum of that electron. Heisenberg was

---

[220] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 374 (1989).

[221] D.A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, California, pp. 36–37 (1983).

able to show that if Δp is the uncertainty in the momentum of the electron, then

$$\Delta x \Delta p \approx h. \tag{3.25}$$

Equation (3.25) is called "Heisenberg's uncertainty principle" or "Heisenberg's principle of indeterminacy," a fundamental principle of nature, which states that "a particle may have a definite position or it may have a definite velocity, but it cannot in any sense have both." Thus, if we wish to locate any particle to within a distance Δx, then we automatically introduce an uncertainty in the momentum of the particle that is given by Eq. (3.25).[222]

It must be pointed out that this limitation to our knowledge of the physical world, embodied in Heisenberg's uncertainty principle, is not due to poor measurement or poor experimental technique, but is a fundamental property of the act of measurement itself; it is intimately associated with quantization.[223] For if light were not quantized, one could imagine making measurements with ever-smaller and smaller quantities of light and energy until we would ultimately approach a measurement that would not disturb the system being measured; or at least we could make the disturbance as small as we wished, so that the uncertainty principle could be reduced to negligible values. However, nature imposes on us a physical limit h, Planck's constant, which determines the minimum energy content of a photon; this physical limit produces uncertainties that are not negligible and that are more significant on the atomic scale.[224]

Wolfgang Pauli (1900–1958) deduced that electrons have rugged individualities, and that there are never two or more equivalent electrons in the same atom. This idea is known as Pauli's exclusion principle, which became an important addition to the knowledge of atomic structure. For their contributions to the development of quantum mechanics, Heisenberg, Pauli, and Schrödinger were awarded the Nobel Prize in Physics in 1932, 1945, and 1933, respectively.[225]

In the final analysis, quantum theory is a theory of measurements, for it tells us what we can measure within the limits of these uncertainties. If one is able to measure a quantity accurately, then one cannot make a precise

---

[222] ibid., p. 37.

[223] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 92 (1987).

[224] An another way of looking at this problem, bearing in mind that h is fixed, is to imagine using photons of very low frequency. The photon's energy $h\nu$ becomes smaller, and its disturbance of the system being measured is thus minimized. However, low frequency implies long wavelength, which for the optical instrument (e.g., a microscope) translates to poor resolving power. Effectively, we cannot obtain information about an object smaller than that wavelength; we cannot even locate the object, except with an uncertainty of the same dimension as the wavelength [also cited in M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 247 (1987)].

[225] B. Jaffe, *Crucibles: The Story of Chemistry – from Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 255–256 (1930).

measurement of a second quantity. But the possible values of the second quantity are associated with probabilities that one can calculate exactly from the equations of quantum theory. As an example, suppose that we know that the electron in the hydrogen atom is in its ground state, implying that we know its energy precisely. In this case we cannot precisely know its other quantities like the electron's position at a given time or its velocity. However, with the equations of quantum theory we can calculate the probability of the electron being at any particular point in space. Thus emerges the concept of a wave—something that is distributed through space—in contrast to a particle, which is located at a point.[226]

Today Bohr's concept of an electronic orbit is no longer tenable, but modern quantum theory substitutes for the orbit a probability distribution, for which, in the case of the hydrogen atom, the probability is concentrated in the region where the Bohr orbit was. For a free electron (i.e., one not bound to the hydrogen nucleus), the probability distribution looks like that of a wave—a confined region of oscillations, called a wave packet. These wavelike properties are extremely difficult to observe under normal conditions because typical wavelengths of electrons are extremely short—around 10^{-9} or 10^{-10} m. Wavelike properties such as diffraction and interference are only observed when the typical dimensions of the measurement system are of a dimension comparable to the wavelength.[227]

And so it was that in 1927, Clinton Davisson (1881–1958) and Lester Germer (1896–1971) were able to demonstrate wave properties in a beam of electrons by using the latticework of atoms in a crystal as the equivalent of a diffraction grating (just as had been done 15 years earlier for x-rays). Upon examining the electrons scattered by a nickel crystal, they observed a pattern of alternating minima and maxima, just like the pattern of constructive and destructive interference seen in light passed through a grating. Using the well-established equations of diffraction, these scientists calculated the wavelength of the electrons in the beam and found a value that was in agreement with the theory of de Broglie. Soon enough the existence of electron waves was confirmed by other researchers, including George Paget Thomson (1892–1975), the son of J.J. Thomson, who in 1897 had discovered the electron. George P. Thomson sent rapidly moving electrons to extremely thin sheets of gold and other metals, and took photographs of the area behind the metallic films to observe unmistakable evidence of the diffraction phenomenon of waves and thus provided even further proof of the wavelike property of electrons.[228]

Arthur H. Compton (1892–1962) announced in 1925 the phenomenon that hard x-rays, which exhibited all of the properties of light waves, showed a

---

[226] M.I. Sobel, *Light*, University of Chicago Press, Chicago, pp. 92–93 (1987).

[227] ibid., p. 93.

[228] ibid., pp. 93–94.

lowering of frequency after striking a metal plate. Remarkably, when he analyzed his photographs, he observed that the x-rays bounce like balls rather than acting like waves. For this work, he won the Nobel Prize in Physics in 1927. Soon after Compton's announcement, Chandrashekhara Venkata Raman (1888–1970), a Hindu scientist, announced the discovery of an effect similar to Compton's, observing that, when a beam of monochromatic light from a mercury vapor lamp is scattered by benzene, new lines of greater wavelength than the incident beam are revealed by spectroscopy. Both the Compton effect and the Raman effect pointed to the probability that Einstein and de Broglie were right.[229]

With the dual nature of light, x-rays, and the electron firmly established, George P. Thomson predicted that the proton would possess this duality of wave and particle behavior, as well. Soon after, Arthur J. Dempster (1886–1950) confirmed Thomson's prediction when he stripped hydrogen atoms of their electrons in a vacuum tube, directed them against a small calcite crystal, and observed from their rebound that the positively charged hydrogen atoms or protons strike the crystals as waves rather than as tiny balls of particles.[230]

Although the quantum theory of 1925 resolved the age-old dilemma of light by unifying the entities of light and matter, and by asserting that the wave–particle duality applies to both, it failed to address the issue of another fundamental distinction: the photon is created out of nothing—in the flash of an atomic jump—while the electron is permanent.[231]

It must be stated that there is a distinct case where the electron does not exist in a permanent state; in fact, it gets annihilated. This occurs at low energies when an electron (e^-) collides with its antimatter (positron, e^+) in an electron–positron annihilation process, resulting in the annihilation of the electron and the positron, and the generation of γ-ray photons: $e^- + e^+ \rightarrow \gamma + \gamma$. At high energies, other particles such as B mesons or W and Z bosons may be formed from this interaction.

All in all, the message of light since the quantum revolution of the 20th century is the message of the atom, which has given us a new understanding of how humans interact with their surroundings. It is beyond the scope of this book to dwell on the philosophical implications of quantum theory, which extend even to epistemology and logic, or on its predictive success in fields as varied as astrophysics, chemistry, lithography, metallurgy, nuclear and elementary particle physics, and materials science, as well as atomic physics and optics. It is enough, however, to say that quantum theory remains today the premier indispensable tool of the theorist in the realm of physical science.[232]

---

[229] B. Jaffe, *Crucibles: The Story of Chemistry – from Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 254–255 (1930).

[230] ibid.

[231] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 95 (1987).

[232] ibid.

Chapter 4

Chemical Origins of Lithography

> "It is the great beauty of our science, chemistry, that advancement in it, whether in a degree great or small, instead of exhausting the subjects of research, opens the doors to further and more abundant knowledge, overflowing with beauty and utility to those who will be at the easy personal pains of undertaking its experimental investigation."[1]
>
> Michael Faraday (1791–1867)

4.1 Introduction

In order to arrive at a fuller understanding of lithography, it is necessary to examine the work of a large number of scientists whose contributions to chemistry, spanning a period of over 3000 years, were instrumental in shaping the invention of lithography in 1798 by Senefelder[2] and its subsequent development, evolution, and progress. This story not only relates a triumph of advances in the understanding of materials science, but is also a product of the knowledge gained in the convergence of major areas of chemistry and physics (particularly those related to optical, atomic, and molecular physics, which are covered in Chapter 3). Our object here thus is to consider how key theoretical and technological developments in chemistry have influenced the development of lithography.

It is also imperative that we examine ways of overcoming the challenges presented by ever-more-demanding applications of sophisticated chemistry in semiconductor lithography, in terms of materials, processes, and tools.

---

[1] M. Faraday, "On electro-chemical decomposition," *LX. Experimental Researches in Electricity, 7th Ser., The Annals of Electricity, Magnetism and Chemistry and Guardian of Experimental Science*, p. 366 (1837).

[2] A. Senefelder, *Vollständiges Lehrbuch der Steindruckerei* [*A Complete Course in Lithography*, in German], Karl Thienemann, Munich and Karl Gerold, Vienna, Publishers (1818).

A modern lithography module clean room for patterning silicon chips is a masterpiece of automated chemical engineering, operating at its highest efficiency and effectiveness when it is operated by the fewest possible number of operators in order to minimize contamination and to improve both control and uniformity of lithographic processes. Moreover, the fabrication of increasingly powerful and sophisticated computers depends on advances in the design of circuits that match advances in the sophistication of IC device fabrication technologies, especially lithography.

Lithography as practiced in its many varied incarnations today involves essentially a series of chemical transformational processes mediated at some stage by some type of radiation, usually light. Therefore, a narrative that traces the development and evolution of thought and ideas on the chemical phenomena mediated by light (that ultimately led to the invention and subsequent development of lithography) is judged to be a good device for presenting this fascinating story. Along this journey, we will examine fundamental theories on the nature of matter, spanning the earliest known human reflections on these subjects to modern quantum chemistry, while exploring how these developments in chemistry have led to the invention and development of lithography. We will also examine how each key development in the our understanding of the chemical effects of light opened up many new fields. From photometry to photochemistry, a new world sprang into life after the invention of lithography, and another after the invention of photography and photolithography. Additionally, we will highlight specific applications of chemical theories, chemical phenomena, and specific chemical reactions to lithography—be they in the design and synthesis of radiation-sensitive resists, in the development of lithographic exposure sources, in the absorption and interaction of radiation with lithographic optical elements and resists, in lithographic process modeling, etc. The whole story is one of fascinating science that has transformed human life. Figure 4.1 gives a summary of these developments in thematic order, but not necessarily in chronological order.

4.2 Key Developments in Chemistry that Enabled the Invention and Development of Lithography

From very early on in human history to the present time, humans have used naturally occurring chemical compounds, or compounds they have synthesized or formulated, as medicine to alleviate and treat illnesses and diseases, as building and structural materials, or as materials for making art. They have fashioned tools from chemical elements and substances. And they have built their houses and shelters with naturally occurring polymeric materials such as wood. They made their clothing out of animal hides and plant fibers such as cotton and wool, and animal protein fibers such as silk. With the invention of lithography, they used the common lampblack made out of soot as lithographic

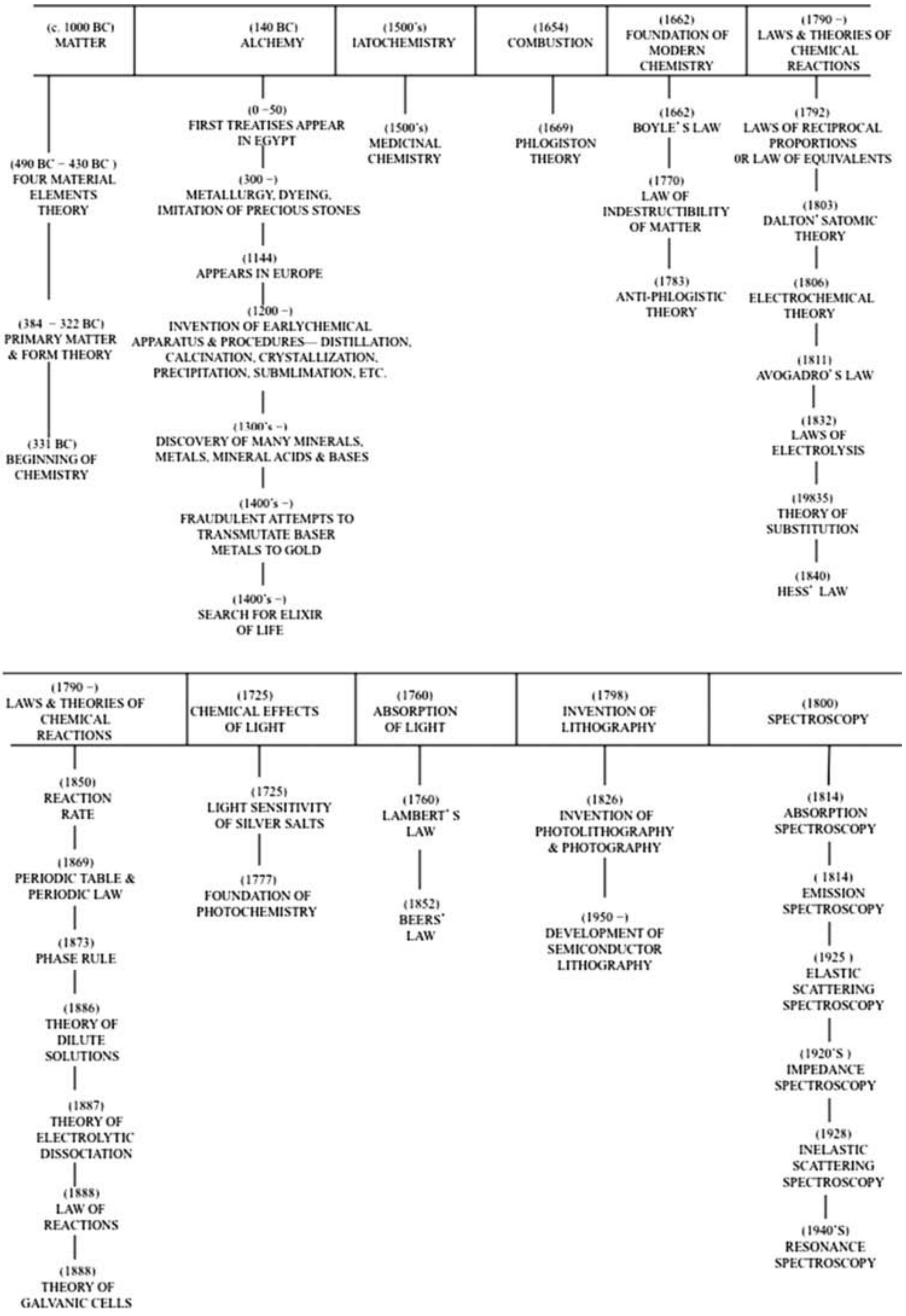


Figure 4.1 Key developments in chemistry that led to the invention and development of lithography.

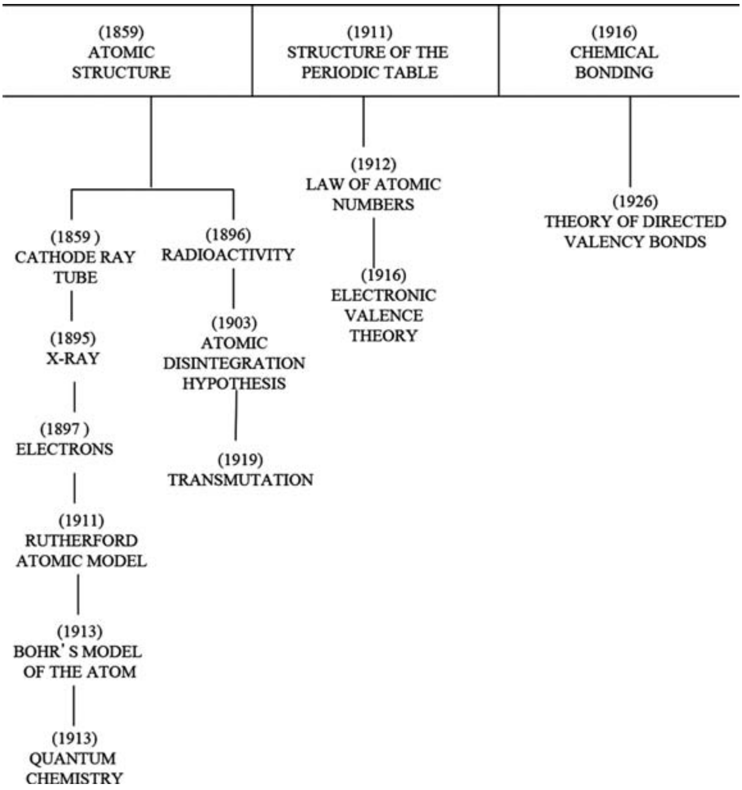


Figure 4.1 (Continued).

ink, and wax as lithographic resist. As lithography evolved into photolithography, they replaced lamp black and wax with photosensitive compounds and synthetic polymers, respectively. The process of making useful things out of naturally occurring substances involves chemical transformation. As people discovered new ways of transforming matter for specific purposes, over time, the need arose to develop a conceptual framework aimed at understanding the very basis of these transformations. Out of these conceptual frameworks evolved chemical ideas that were then passed down and modified along the way as necessary from one era to the next, until our present era. And some of these ideas influenced the course of the development of lithography.

A comprehensive treatment of the history of chemistry has been given elsewhere.[3] In this section, only the main chemical ideas that are directly or indirectly related to the invention, development, and evolution of lithography are outlined, together with the way these ideas evolved and influenced the development of other key ideas over time—stretching from the very roots of

---

[3] For a comprehensive treatment on the history of chemistry, see, e.g., J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York (1989); M.M.P. Muir, *A History of Chemical Theories and Laws*, John Wiley & Sons, New York (1907).

the field to the actual beginning of the field as a distinct discipline, and ultimately to the most advanced state of the field in the present era. Today, with our modern eyes, we might view some of these ideas as part of the false heritage of antiquity, which, instead of helping to clarify observed chemical phenomena, unwittingly served to impede the progress of chemistry.

The meaning and purpose of chemistry has undergone a lot of metamorphoses throughout recorded history. Muir writes: “The purpose of chemistry seems to have changed much from time to time. At one time chemistry might have been called a theory of life, and at another time a department of metallurgy; at one time as study of combustion, and at another time an aid to medicine; at one time an attempt to define a single word, the word element, and at another time, the quest for the unchanging basis of all phenomena. Chemistry has appeared to be sometimes a handicraft, sometimes a philosophy, sometimes a mystery, and sometimes a science.”[4] Today chemistry is viewed as the central science that enables other fields of science and technology, including lithography.

4.2.1 The four-element theory

The beginning of the story of lithography can be traced all the way back to the roots of chemistry itself, in the first clear expressions of the idea of an element in the teachings of Greek philosophers, who sought to explain the material world around them. Foremost among these philosophers were Thales (640–546 BC), who proposed that all things are made of water, Anaximenes (560–500 BC), who proposed that all things are made of air, and Herakleitos (536–470 BC), who proposed that all things are made of fire. It was Empedocles (490–430 BC) who introduced the concepts of four “roots” of things: fire, air, water, and earth; and two forces, attraction and repulsion, that joined and separated them.[5]

Plato (427–347 BC) was the first person to use the name “elements” (*stoicheia*); he assumed that things are produced from a formless primary matter, perhaps just space, taking on “forms.” The minute particle of each element, he believed, has a special shape—fire a tetrahedron, air an octahedron, water an icosahedron, and earth a cube—that are cut out of space by two kinds of right-angled triangles. To him the elements change into one another in definite ratios, by resolution into triangles and re-association of these. His dialogue *Timaeus*[6] includes a discussion of the compositions of

---

[4] M.M.P. Muir, *A History of Chemical Theories and Laws*, John Wiley & Sons, New York, p. vii (1907).

[5] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, Chapter II (1989).

[6] Plato, “Timaeus,” in *The Dialogues of Plato*, *The Great Books of the Western World* (translated by B. Jowett), M.J. Adller, Ed., Encyclopedia Britannica, Inc., Chicago, pp. 442–477 (1993).

inorganic and organic bodies and can be regarded as an elementary treatise on chemistry.

Summarizing the theories of earlier thinkers, Aristotle (384–322 BC) proposed that all substances consist of primary matter called *hule*, from which different forms could be made, just like a sculptor can make different statues from one block of marble. Aristotle preferred to think of the form as evolving from within, as in organic growth. He called the form *eidos* and believed that these forms can be removed and replaced by new ones, thus giving rise to the ideas of the transmutation of the elements. He believed that the fundamental properties of matter were hotness, coldness, moistness, and dryness. By combining these in pairs, he obtained what are called the four elements, fire, air, earth, and water, which make up in varying proportions the vast multitude of different substances in nature.[7] He also believed that a substance that burns more vigorously than another contains a higher proportion of the element fire, while one that has more fluid than another contains a higher proportion of the element water. To these four material elements, he later added a fifth, immaterial element that appears in his latter writings as the *quintessence*, which corresponds with the ether. The four material elements, he thought, were supposed to settle out naturally into earth (below), water (the oceans), air (the atmosphere), and fire and ether (the sky and heavenly bodies).[8]

He defined an element to mean "one of those bodies into which other bodies can be decomposed and which itself is not capable of being divided into others." He drew the distinction between mechanical mixture and solution, and between chemical change and chemical change with complete change of properties. He said that vapor from sea water evaporated in a vessel condenses as fresh water on the cool lid, and (wrongly) that wine would give water in a similar way, suggesting that he knew of a primitive method of distillation. He classified several chemical processes, thought that metals are mostly water (a clear influence of Plato), and listed the "homogenous" parts of animals and plants, thus constituting a primitive organic chemistry.[9]

The four-element theory endured well into the end of the 18th century, and the concept of the ether as a medium for the transmission of light endured until the end of the 19th century,[10] ending with the Michelson and Morley experiment, as described in Chapter 3. The remarkable longevity of the four-element theory for more than 20 centuries highlights the lack of chemical proof that earth, water, air, and fire were elements, or that nature's enormous variety of materials were made up of only four substances. To the ancient

---

[7] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 13–14 (1989).

[8] ibid., p. 14.

[9] ibid.

[10] ibid., pp. 14–15.

mind, the theory helped to explain many facts in ways that were easily understood. It is understandable why Thales, who worshipped the Nile, could have attributed the origin of all things to water, seeing as he would have observed that flasks of water, when heated over fires until all the water had evaporated, were left with only dull, earthy substances such as pieces of rock that appeared to have come only from the water.[11]

A phenomenon that gripped the ancient mind more than any other in its effect is fire. We will see in subsequent sections of this chapter how the explanation of fire fueled the imagination of a great many thinkers for many centuries and gave birth to theories about the nature of things and the very basis of chemistry itself. And out of attempts to understand fire came the foundation of chemistry as a distinct discipline with its own instruments, techniques, and methods.

4.2.2 Chemistry as a distinct discipline

Chemistry as a distinct discipline had its origin around the beginning of the Christian era, in the city of Alexandria in Egypt, which was founded by Alexander the Great at the mouth of the Nile in 331 BC. In Alexandria, the ancient Egyptian industrial arts of metallurgy, dying, and glassmaking were fused with the philosophical speculations of ancient Greece, which at that time had undergone many changes and had assimilated elements of Eastern mysticism. In particular, Plato's teachings were transformed into Neoplatonism. The Gnostic sects grew during the early centuries of the Christian era, also in Alexandria. Most importantly, the technical arts, i.e., the imitation and falsification of precious metals, gems and dyes, gradually was transformed during this time (influenced by the theoretical views on the nature of matter) into the "divine" or "sacred" art of making gold or silver. It was precisely with this development, which contained in its very essence the germ of chemistry, that a considerable body of positive, practical chemical knowledge came into existence.[12]

It is noteworthy that before the commencement of the Christian era we find no indication of the existence of chemistry in Europe or the Near East, although chemical operations were known to technologists of the time. The name "chemistry" first appears in an edict of the Emperor Diocletian in 296 AD, given by Suidas (10th century) from an older source, in which the books of the Egyptians (in Alexandria) on *chēmeia*—making (i.e., imitating) gold and silver—were ordered to be burned. The word also appears in the writings of Greek authors such as Plutarch, who in his treatise *On Isis and Osiris*, written about 100 AD, called Egypt *chēmeia* on account of the

---

[11] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 74 (1976).

[12] ibid., pp. 19–20.

black color of its soil.[13] Partington writes that the word probably meant Egyptian art.[14]

It is also interesting that the art of chemistry was commonly ascribed to the mysterious Hermes Trismegistus, whose name is still preserved in the expression "hermetically sealed," and who was invoked by the Gnostics as the source of all wisdom.[15]

Quite a good number of Greek treatises on the divine were in existence as early as the first century AD. These chemical treatises, which were written in Greek and found in Egypt at Alexandria, are generally considered to be the earliest known books in chemistry. Zosimos (300 AD) was the most prolific of the authors of these chemistry tomes. The books contained many examples of interesting, practical chemical information, which appeared for the first time, as well as many diagrams of chemical apparatuses. They contained information on operations such as fusion, calcinations, solution, filtration, crystallization, sublimation, and especially distillation, along with methods of heating, including open fire, lamps, and sand and water baths.[16]

4.2.3 Alchemy

It is generally accepted that alchemy began in China around 140 BC. In 1144, alchemy appeared in Europe by way of translations made in Spain from Arabic works.[17] Partington asserts that the Arabic name *alchemy* is merely the Alexandrian-Greek *chēmeia* with the Arabic definite article *al* prefixed. Both Roger Bacon and Albert the Great (Albertus Magnus) wrote about it in 1250.[18]

Alchemy was based on the concept of transmutation of substances, an idea that was derived from the marked effects of arsenic, mercury, and sulfur in changing the color of metals. It was quite a rational idea at the time, for many alchemists believed that they could produce gold from the action of the above substances on baser metals such as lead, or that they could produce silver from copper by turning the color of copper to white with arsenic. The transmuting agent was called by the Arabs *aliksir* (elixir), perhaps from the Greek *xerion*, which means a dusting powder or cosmetic; later it was

---

[13] ibid., p. 20.

[14] ibid., p. 20.

[15] ibid., p. 20.

[16] ibid., pp. 21–22.

[17] In 640 AD, Egypt was conquered by the Arabs, who, with the help of learned Greeks, Syrians, and Persians living under their rule, made translations of Egyptian books on chemistry (written in Greek) into Arabic [cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 27 (1989)].

[18] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 25 (1989).

called the philosopher's stone, or tincture (i.e., an agent for changing color of substances) by European alchemists.[19]

For many centuries, alchemy, nourished by superstition and chicanery, held the imagination of many of its adherents and adepts, who practiced their art in secrecy, often times deceiving and making claims that were impossible to realize with the instruments at their disposal. Although their claims were fraudulent, their efforts were not all entirely in vain. Francis Bacon compared alchemy "to the man who told his sons he had left them gold buried somewhere in his vineyard; where they by digging found no gold, but by turning up the mold about the roots of the vines, procured a plentiful vintage."[20] Through alchemy, many valuable discoveries were made, and many chemical facts were learned. For instance, nitric, hydrochloric, and sulfuric acids, the three most important acids used by the modern chemist, and aqua regia, a powerful solvent for gold (formed by mixing the first two of these acids), were all introduced by the alchemists of old. In their search for the seed of gold in dirt and mire, they unearthed and discovered new elements like antimony, arsenic, bismuth, and phosphorus. Many of the common chemicals in use today, such as plaster of Paris, red lead, iron, and silver salts, along with heavy barium sulfide, the first substance known to glow in the dark after exposure to sunlight, owe their discovery to the alchemist Vincenzo Cascariolo, a cobbler of Bologna.[21]

Alchemists were the first to introduce some of the apparatuses and utensils commonly used in scientific laboratories across the world today. These include the cupel, the distilling flask, the retort, the water bath, and even the balance in its crude form. Processes such as the extraction of gold by amalgamation with mercury and preparation of caustic alkali from the ashes of plants all owe their origins to alchemists. The heritage bequeathed to us by the alchemists of old is indeed a rich one. Although they were not successful in making synthetic gold, they nevertheless paved the way for the more fruitful science of chemistry.[22] However, using instruments with power beyond the capabilities of the those of alchemists, the alchemists' dream of transmuting baser metals into gold was at last realized by modern science when, in 1941, Sherr, Bainbridge, and Anderson artificially produced gold from mercury.[23]

Other notable developments in alchemy, and subsequently in chemistry, are outlined below. In the 13th century, Bacon described gunpowder, which

---

[19] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 34–40 (1989).

[20] Cited in B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York. p. 12 (1976).

[21] ibid.

[22] ibid.

[23] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 379 (1989).

became known in Europe (perhaps from China by way of Arabs) as being composed of seven parts saltpeter, five parts wood charcoal, and five parts sulfur. Theophrastus von Hohenheim, commonly called Paracelsus (1493–1541), founded iatochemistry, or chemistry applied to the service of medicine. In 1556, Georg Agricola (1494–1555), a German physician, wrote *De Re Metallica*, on mining and metallurgy. Jan Baptista van Helmont (1580–1644) clearly expressed for the first time the law of the indestructibility of matter and emphasized that metals, when dissolved in acids, are not destroyed but can be recovered again by suitable means. He invented the name ‘gas’ and was the first person to describe carbon dioxide. He regarded air and water as the fundamental elements. In 1597, Andreas Libavius (around 1540–1616), a German schoolmaster, wrote *Alchemia*, the first textbook on chemistry. Between 1646 and 1649, Johann Rudolph Glauber (1604–1670) published *New Philosophical Furnaces*, in which he describes a preparation of salts and mineral acids. In 1675, Nicholas Lemery published a textbook, *Cours de Chemie*, in which he classified substances into three “Kingdoms of Nature,” mineral, vegetable, and animal. In 1666, Otto Tachenius published *Hippocrates Chimicus*, in which he gave a clear definition of a salt: “all salts are composed of two parts, of acid and alkali.” Johann Kunckel (1630–1703) described the preparation of gold ruby glass and in 1679 wrote a treatise on glass manufacture (*Ars Vitraria*) based on an earlier work by the Italian glassmaker, Antonio Neri.[24]

4.2.4 Early theories on the nature of fire and combustion

That Herakleitos and other ancients considered fire to be the purest and most perfect of the elements of nature is not surprising. To them the flames of fire rose ever upward, reaching finally to the highest of the seven heavens—the heavens of fire and light—the empyreum, as they called it. Humankind’s greatest and earliest achievement was the discovery of the use of this fire. According to Greek mythology, the gods had not given this tool to humanity willingly. Rather, it had to be stolen by the demi-god Prometheus who, aided by Minerva, ascended to heaven and lit his torch at the chariot of the sun. Armed with this fire, humankind was able to transcend the level of the brute. So highly did the ancients regard the luminous and expansive flame that the Egyptians honored it by burning a perpetual fire in every temple, the Greeks and Persians kindled fire in every village and town, and the Romans consecrated Vestal Virgins to watch the sacred fire on their altars.[25]

The nature of fire elicited a considerable number of interpretations from the ancients. Aristotle considered it to be one of the great principles of all

---

[24] ibid., p. 65.

[25] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 27 (1976).

things. Herakleitos, as stated above, regarded it not only as the elementary principle of all things, but also as the universal force of creation. For centuries, humankind had pondered over the mystery of the flame. Initially, the fiery force that could engulf life and bring utter destruction was regarded with fear and reverence, and later it was subjected to serious investigation. Among the ancients, Plato theorized that all burnable bodies contain some inflammable property, which, however, he failed to identify. The alchemists of later centuries considered the cause of fire to be either some vague spirit of sulfur, or yellow, brittle, solid sulfur itself. They used and discarded pseudo-scientific concepts at will. Paracelsus attributed the burning of wood to the three elements possessed by the wood. He thought that wood burned because it contained sulfur, that it gave off a flame because it contained mercury, and that it left an ash because of the salt that is present in all wood. This explanation persisted for hundreds of years in western thought.[26]

4.2.5 Phlogiston theory

And then came Johann Joachim Becher (1635–1682), who first enunciated a definite, although pseudo-scientific, theory of the nature of fire. Crude though his idea was, it explained a great number of chemical phenomena to the medieval mind and gripped the scientific world for more than a century. In 1669, Becher published a book called *Physica Subterranea*,[27] in which he introduced the idea that the constituents of bodies are air, water, and three earths, one of which is inflammable (*terra pinguis*), the second mercurial, the third fusible or vitreous. These three earths correspond to the alchemists' sulfur, mercury, and salt.[28] Dispensing with the sulfur principle of his predecessors that extended all the way back to the Greek era, and finding a substance that contained no sulfur and yet could burn vigorously, he postulated a "terra pinguis"—a fatty or inflammable earth possessed by all substances that could burn. This fatty inflammable earth was later termed "phlogiston" by Georg Ernst Stahl (1660–1734) (see Fig. 4.2), which in Greek means "to set on fire."

Stahl formalized Becher's inflammability earth concept into the *Theory of Phlogiston*.[29] Like Becher, Stahl believed that all combustible and inflammable substances and metals contained a common principle of inflammability called phlogiston φ, which was contained in all substances and escaped upon combustion or calcinations in the form of fire and flame; it could be

---

[26] ibid., pp. 27–28.

[27] J.J. Becher, *Physica Subterranea*, Leipzig (1703).

[28] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 27–28 (1976).

[29] An excellent account of the history of phlogiston theory has been rendered elsewhere; see, e.g., J.H. White, *History of the Phlogiston Theory*, Arnold and Company, London (1932); K. Ullrich, *Becher, der Erfinder der Gasbeleuchtung*, Speyer, Germany (1936).

Figure 4.2 Georg Ernst Stahl (1660–1734), who formulated the phlogiston theory to explain combustion. He believed that all combustible substances and metals contain a common substance, phlogiston, which escapes upon combustion or calcinations but can be transferred from one substance to another and restored to metallic calces by heating with substances rich in phlogiston (such as charcoal, oil, etc.). (Printed with permission of the Deutsches Museum, Munich.)

transferred from one substance to another, and upon heating with substances rich in phlogiston (such as charcoal, oil, etc.),could be restored to metallic calces, reproducing the metal:[30]

$$\text{Metal} \leftrightarrow \text{calx} + \varphi$$

Scheme 4.1 Illustration of the combustion process of metallic substances according to the phlogiston theory, in which phlogiston (the common principle of inflammability) is believed to escape from the substance in the form of fire and flame during combustion, leaving behind the metallic calx (oxide of the metal). Heating the metallic calx with substances rich in phlogiston such as charcoal produces the pure metal.

To Becher and his followers, phlogiston was the concept that explained many chemical phenomena, including burning, oxidation, and calcination. The vital process of breathing could, in like manner, be explained by it when one assumes that the lungs in humans and animals constantly exhale phlogiston as a byproduct of food digestion. According to Becher, when a substance is burned, its phlogiston is violently given off in the form of the flame. He asserted that the burned body would weigh less than the original material by an amount equal to the phlogiston given off during the process. To him, phlogiston was not merely an idea, but was a chemical substance with a definite color, odor, and weight.

---

[30] J.R. Parington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 148 (1989).

A particular application of the phlogiston theory eventually led chemists to much confusion and helped to finally bring about its downfall. This confusion arose in the following manner. When a metal such as copper or lead is heated in open crucibles, it turns into a powdery substance called calx, which has no metallic properties. When weighed, the calx was observed to be heavier than the metal. The same phenomenon takes place in the rusting of iron, but in the absence of heat. Explanations typically given for this observation were that some kind of "soul" escapes from the metal, or that the matter becomes denser, or that some kind of acid is absorbed from the fire, or that fire possesses weight and is absorbed by the metal during the formation of the calx.[31] Other explanations for this observation assumed that a metal is a kind of combustible material, which upon heating, loses its phlogiston, leaving the powdery residue, calx. These phlogistic chemists knew, however, that if this calx was reheated with charcoal, it reconverted back into metal. Because charcoal would burn away almost entirely, they regarded it as being very rich in phlogiston. They explained that the heating of the calx with charcoal was responsible for restoring enough phlogiston to the calx to reconstitute the original metal. This led them to define a metal as a compound of calx and phlogiston, and calcination as the process of heating a metal to produce its calx via decomposition, which they regarded as a kind of combustion in which phlogiston escaped from the metal.

It was well known, on the other hand, as early as the 16th century, that metals increase in weight upon calcinations. But how could the weight increase since something material, namely phlogiston, had been lost from the substance of the metal? This was a serious contradiction, which some of the chemists who supported phlogiston theory disregarded, while others assumed that phlogiston had negative weight.[32] Others concocted some rather inconsistent hypotheses to explain it. Mikhail V. Lomonosov (1711–1765) in 1754 attributed the increase in weight to air being fixed by the metal.[33] Laurent Béraut (1702–1777) in his 1747 dissertation[34] concluded that the increase in weight was due to the addition of certain foreign particles mixed with the air and separated by the action of heat.[35]

Thus, the phlogiston theory went through a series of modifications throughout this period. In the old phlogiston theory, phlogiston was believed to be equivalent to *minus* oxygen; in the later theory, it was sometimes

---

[31] J.C. Gregory, *Combustion from Herakleitos to Lavoiser*, Edward Arnold, London (1934).

[32] ibid.

[33] ibid., p. 149.

[34] L. Béraut, *Dissertation sur la Cause de l'Augmentatiion de Poids, que certaines Matières acquirèrent dans leur Calcination*, Bordeaux (1747) [cited in J.R. Partington*, A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 149 (1989)].

[35] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, Chapter V (1989).

proposed to be hydrogen by Cavendish (1731–1810), Kirwan (1735–1812), and Priestly (1733–1804), or the matter of light by Pierre Macquer (1718–1784).[36]

It should be mentioned that even at its height of popularity, there were chemists who were skeptical of the phlogiston theory. Nicolas Lémery (1645–1715), for instance, did not adopt the theory of phlogiston.[37] Neither did Antoine Lavioisier, who lived 100 years later than Becher. In fact, Lavoisier is quoted to have said something to the effect of: "They have turned phlogiston into a vague principle which consequently adapts itself to all the explanations for which it may be required. Sometimes this principle has weight, and sometimes it has not; sometimes it is free and sometimes it is fire combined with earthly elements; sometimes it passes through the pores of vessels and sometimes they are impervious to it. It is a veritable Proteus changing in form at each principle."[38]

The fact cannot be overemphasized that phlogiston theory retarded the progress of chemistry and prevented a number of natural philosophers of the age from understanding the true nature of fire and burning over a span of almost a century. A number of natural philosophers during the period came very close to understanding the true nature of fire and burning, but unfortunately ended up making wrong conclusions borne out of their attempts to reconcile their results with the error-prone, prevailing phlogiston theory. So until it, along with the four-element theory, was comprehensively rejected, modern chemistry remained unfounded; for earth, air, water, and fire are not the elements, and substances do not burn because of the presence in them of a common property of inflammability.[39]

It should also be pointed out that some good came out of the phlogiston theory. This theory made it possible for a large number of chemical facts to be coordinated into a system. It helped to focus the attention of chemists on new and more fruitful fields of investigations and turned them away from the pursuit of the elixir of life and the philosopher's stone. Furthermore, the challenge of phlogiston led to a more serious and systematic study of chemical reactions, such as combustion and oxidation, and stimulated the development of the analysis of substances. Before we delve into the new vistas that opened up in chemical research when the death nail of phlogiston was struck, let us briefly survey the intellectual landscape in Europe during Becher's life, if only to contextualize the motivation that drove him and helped formulate his theory.

---

[36] ibid., p. 149.

[37] ibid., p. 149.

[38] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 30 (1976).

[39] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, Chapter V (1989).

Becher's life span (1635–1682) overlapped a period of intense intellectual activity centered in Europe. William Harvey, an English physician, demonstrated with remarkable precision the circulation of blood. Robert Hooke wrote *Micrographia*, a book dealing with microscopy. Ole Rømer, a Dane, measured the speed of light. Evangelista Torricelli, an Italian physicist and mathematician, demonstrated atmospheric pressure with his newly invented barometer. Otto von Guericke, mayor (der Burgermeister) of Magdeburg, introduced the air pump and astonished the Emperor and his court when 16 pairs of horses struggled in vain to pull apart two hollow iron hemispheres from which the air had been evacuated. Anthony van Leeuwenhoek, a Dutch scientist, gazed upon red blood cells and minute plants and animals with the microscopes he had constructed, while Christian Huygens, his countryman, invented the pendulum clock and proposed the wave theory of light, and Baruch Spinoza, a lens grinder of Amsterdam, wrote his *Ethics*. René Descartes, the French philosopher, made important contributions to the sciences of mathematics and physics. Sir Isaac Newton developed calculus and proposed his theory of universal gravitation.[40]

The field of chemistry also witnessed many developments. The French chemist Nicholas Lémery published his *Cours de Chimie*. Johann Glauber of Bavaria discovered important chemical salts and published an encyclopedia of chemical processes. Jean Rey, a French physician, wrote a curious account of the rusts of tin and lead. John Mayow of Cornwall, another physician, investigated the respiration of animals in air and its relation to the burning of metals. And in 1669, the same year that Becher published his *Physica Subterranea*, Hennig Brandt, a Hamburg merchant and alchemist, accidentally discovered fiery phosphorus, which shines in the dark, and awakened the world with this 'cold fire.' Becher made his only lasting contribution to applied chemistry with his discovery of ethylene by the action of common alcohol on sulfuric acid.[41]

And so it was that 100 years after Becher's death, history recorded that Madame Lavioiser, the wife of the founder of modern chemistry, robed as a priestess and surrounded by the scientific celebrities and elites of Paris, burned the writings of Becher along with those of his illustrious follower, Stahl, on an altar. Thus, while a solemn requiem was chanted, the theory of phlogiston was symbolically consumed by the fire, and out of its ashes, like the legendary phoenix, sprang a new chemistry.[42]

---

[40] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 35 (1976).

[41] ibid.

[42] ibid.

4.2.6 Beginnings of modern chemistry

The beginning of modern chemistry coincides with attempts to elucidate what constitutes an element; it also coincides with systematic studies on combustion, nature, and the composition of air by a group of pioneering chemists, mostly in England, France, and Germany in the 17th century. Robert Boyle (1627–1691) was one of the pioneering chemists of the modern era of chemistry. A major problem that occupied his interest was elucidating what really constitutes an element. Boyle showed through experimentation that the four elements of Aristotle (fire, air, earth, and water) and the three principles of the alchemists (mercury, sulfur, and salt) did not deserve to be called elements or principles at all since none of them could be extracted from bodies, e.g., metals. He defined an element as:

> "I mean by Elements, as those Chymists that speak plainest do by their Principles, certain Primitive and Simple, or perfectly unmingled bodies; which not being made of any other bodies, or of one another, are the Ingredients of which all those call'd perfectly mixt Bodies are immediately compounded, and into which they are ultimately resolved."[43]

Boyle believed in the atomic theory, which he used to explain chemical changes. In 1662, he discovered the law that bears his name: Boyle's law, which states that the volume of a gas in inversely proportional to the pressure.[44]

Robert Hooke (1635–1702)[45] postulated a theory of combustion involving twelve propositions in which he stated, among other things:

> "Air is the universal dissolvent of all sulfurous bodies; this action of dissolution produces a very great heat and that which we call fire; this dissolution is made by a substance inherent, and mixt with the air, that is like, if not the very same, with that which is fixt in saltpeter; in this dissolution of bodies by the air, a certain part is united and mixt, or dissolved and turn'd into the air, and made to fly up and down with it."[46]

John Mayow (1641–1679) recognized that combustion and respiration are analogous processes. He correctly stated that animal heat is developed in the muscles, and he proved that arterial blood in vacuum gives off a gas.[47]

---

[43] R. Boyle, *Sceptical Chymist: or Chymico-Physical Doubts & Paradoxes, touching the Spagyrist's Principles commonly call'd Hypostatical as they are wont to be Propos'd and Defended by the Generality of Alchymists*, London, p. 350 (1661).

[44] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 67–77 (1989).

[45] Robert Hooke claimed to have anticipated Newton in the discovery of the law of gravitation and is known for his 1660 discovery of Hooke's law.

[46] R. Hooke, *Micrographia: or some Physiological Descriptions of Minute Bodies made by Magnifying Glasses. With Observations and Inquiries thereupon*, London (1667).

[47] ibid., p. 81.

4.2.7 Discovery of simple gases in common air

The history of the rejection of the phlogiston theories begins with the discovery of simple gases in common air, particularly the discovery of oxygen, which constitutes one-fifth of common air. The story of these discoveries begins, oddly enough, in the phlogistic era and extends all the way back to the early part of the modern era in chemistry, fueled by the work of a good number of exceptional chemists.

At first it was supposed that gases were different varieties of common air, with slightly different properties. In his *Vegetable Staticks* (1727), Stephen Hales (1677–1761) describes the production of "air" from a variety of substances, and its collection and measurement, but he was not able to make any qualitative distinction between the different gases in air (such as oxygen, hydrogen, nitric oxide, carbon dioxide, coal gas, etc.), which he must have obtained.[48] In 1754, Joseph Black (1728–1799) demonstrated the chemical differences between one kind of air and common air, and gave the former the name of 'fixed air' (which is now known to be carbon dioxide). Black showed that fixed air was produced in the combustion of charcoal, in respiration when air is exhaled from the lungs, and in fermentation.[49] Carbon dioxide had been previously recognized and was called *gas sylvestre* by Jan Baptist van Helmont.[50] Black's other major contribution was the discovery of latent heat and specific heats.[51]

One of the earliest accounts of the discovery of hydrogen gas was by Paracelsus, who in the 16th century first reported bubbles of air rising from sulfuric acid when he threw pieces of iron into it. He had also discovered that this evolved gas could burn. Later, van Helmont, a Flemish physician, made a similar observation.[52] Both of these investigators neglected to study the property of this gas or the actual phenomenon that gave rise to the gas.

Then came Henry Cavendish (1731–1810), who likewise had noticed the evolution of a gas when zinc or iron was dropped into an acid. But unlike Paracelsus and Jan van Helmont, Cavendish was determined to investigate this phenomenon.[53] He introduced sulfuric acid into three separate flasks and sealed the mouths of the flasks with cork. He inserted one end of a glass tube

---

[48] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 90–93 (1989).

[49] ibid., pp. 90–93.

[50] ibid., p. 48.

[51] ibid., p. 48.

[52] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 56 (1976).

[53] Excellent accounts of the experimental research of Cavendish can be found in, e.g., H. Cavendish, *Experiments on Air*, London (1785); G. Wilson, *Life of Henry Cavendish*, Cavendish Society, London (1851); W.R. Aykroyd, *Three Philosophers* (*Cavendish, Priestley, Lavoisier*), W. Heinemann Ltd., London (1935).

into the cork and tied a bladder to the other end of the tube for each flask. Then into first, second, and third flasks he threw pieces of zinc, iron and tin, respectively, and collected the evolved colorless and odorless gas from each flask with the aid of the bladder. Next, he repeated the experiment, but this time replacing sulfuric acid with hydrochloric acid, and collected three more bladders full of gas. Next, he brought a lighted taper to his six samples of gas and watched each specimen of gas burn with the same pale blue flame, indicating that the gases were the same. What a strange result, he must have thought! He concluded, rather erroneously, that the inflammable air he had discovered was phlogiston, and that it came, not from the acids or water in the bottles, but from the metals themselves. After all, Becher had taught that metals were compounds of phlogiston and some peculiar earths.[54]

In order to convince himself that the inflammable air he had just discovered did not come from the water, he passed the gas through drying tubes to free them of all moisture, and then he weighted the pure imprisoned "phlogiston." Though extremely light, he found that the dry inflammable gas actually had weight. It was ponderable. He thought he had bottled phlogiston itself. So immersed was Cavendish in the phlogiston theory of Becher that he did not even realize that he had isolated, not the principle of fire, but pure, colorless hydrogen gas.[55]

Another important discovery Cavendish made dealt with the composition of water.[56] In this experiment, he used electrical sparks generated from a Leyden jar to fire an explosive mixture consisting of 423 measures of hydrogen gas and 1000 parts of common air in a closed, thick, glass cylinder. After the explosive reaction, all of the hydrogen and about one-fifth of the common air lost elasticity and condensed into dew, which lined the glass. When he repeated the same experiment in a globe vessel [exhausted of air with the aid of an air pump and then filled with a mixture of pure oxygen (instead of common air) and hydrogen] and fired the mixture as before, he again obtained the same liquid water, and the same gases disappeared. When he weighed the gases and their product as well as the globe, before and after the reaction, he obtained the same remarkable result—two volumes of hydrogen always combined with one volume of oxygen to form a weight of water equal to the weights of the gases. Thus, he proved conclusively the true composition of water. A corollary experiment was done a few years later by Jan Rudolph Deiman and Adriaan Paets van Troostwijk, who passed electric sparks from a frictional machine through water and decomposed it into hydrogen and oxygen.[57]

---

[54] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 56–57 (1976).

[55] ibid.

[56] ibid.

[57] ibid., p. 63.

A rather remarkable controversy[58] started two months after Cavendish had read his paper on the composition of water to the Royal Society. James Watt, the great inventor, transmitted a report to the same Royal Society, claiming the discovery of the composition of water as early as April of the preceding year. Lavoisier also laid claim to the same discovery on the basis of an oral report he submitted in conjunction with Laplace to the French Academy in June 1783. In this report, Lavoisier announced the composition of water without acknowledging any indebtedness to either Cavendish or Watt, even though he was aware of their work. This controversy came to a head around 1849, after the three principals in this matter were all dead. Their friends started a turmoil, which raged on for ten years. Watt's friends accused Cavendish of deliberate plagiarism. To vindicate Cavendish and clear his good name, the President of the British Association for the Advancement of Science published a lithographed facsimile of Cavendish's original notebook, proving that he made the discovery earlier than Watt, and certainly much earlier than Lavoisier. Today, credit for the discovery of the composition of water belongs only to Cavendish—he who first discovered it.[59] It is remarkable that it took the agency of lithography to resolve this dispute!

Cavendish made other important discoveries. He (1) introduced the concept of chemical equivalents (1766–1788), (2) investigated the properties of fixed air (CO_2), (3) determined the (nearly constant) composition of the atmosphere (1781), and (4) observed that nitric acid can be formed by repeated sparking of nitrogen of the air ("dephlogisticated air") with an excess of oxygen ("phlogisticated air") over potash solution in an enclosed vessel. A supporter of the phlogiston theory, Cavendish admitted that his results could be explained by Lavoisier's oxygen theory[60] (see Section 4.2.10).

Joseph Priestly (1733–1804) (see Fig. 4.3) was another exceptional pioneering chemist of the modern era. In 1774, he discovered oxygen, which he isolated from the decomposition of mercury oxide and lead oxide. The story of this discovery is a fascinating one, and it goes like this.[61] Using a burning lens to concentrate the rays of the sun on a sample of the red powder

---

[58] ibid., pp. 63–64.

[59] ibid.

[60] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 150–152 (1989).

[61] Many excellent accounts of the discovery of oxygen by Priestley have been provided in the following: B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 43–45 (1976); J. Priestley and J. Johnson, *Experiments and Observations on Different Kinds of Airs*, London (1776); H.R. Allenson, *Memoirs of Priestley*, written by himself, with continuation to the time of his decease by his son. (Reprint from 1809 edition), London (1904); J.R. Rutt and R.R. Hunter, *Life and Correspondence of Joseph Priestley*, London (1831); J. Priestley, *The Doctorine of Phlogiston Established and that of the Composition of Water Refuted*, Northumberland (1800); J. Priestley and J. Johnson, *Considerations on Phlogiston and Decomposition of Water*, London (1796).

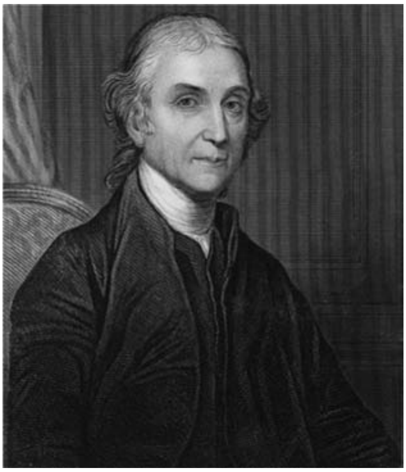


Figure 4.3 Joseph Priestly (1733–1804) (printed with permission of the Deutsches Museum, Munich).

of mercury oxide placed in a bell jar to which he attached a bottle placed over a trough of mercury, he succeeded in decomposing the latter into mercury and liberated a gas, which we now call oxygen. Many others before him had obtained the same result, including Carl Wilhelm Scheele, the great Swedish apothecary chemist, who obtained the same result three years earlier, as did Robert Boyle, one hundred years earlier. Stephen Hales, too, had liberated a gas from saltpeter but saw no connection between it and air. Eck von Salzbach, an alchemist, had performed similar experiments in Germany three centuries earlier than Priestly yet could not establish any connection between oxygen and air.

What Priestly did differently from these other researchers is that he went a step farther to investigate the nature of the oxygen gas he had liberated and collected. He observed that, when he placed a candle in the jar of this oxygen gas, the flame was not extinguished. Instead, the flame burned larger and with greater splendor. When he inserted a piece of glowing charcoal into the bottle of this oxygen gas, to his amazement, he saw that the charcoal was quickly consumed in the oxygen gas—sparkling and crackling like paper dipped in a solution of nitre. When he inserted a red-hot iron wire into the bottle of this oxygen gas, the heated metal glowed and blazed like nothing he had seen before. Next, he placed a mouse into each of two identical glass vessels inverted over water, one of which he filled with oxygen, and the other with ordinary common air, and observed them for a time. Within 15 minutes, the mouse in the glass vessel with common air died, while the mouse in the glass vessel with pure oxygen was still alive after 30 minutes. Eventually, Priestly tried to inhale a freshly made sample of this oxygen gas through a glass tube and found to his astonishment that the feeling in his lungs was not noticeably different from that when he breathed common air, although he felt that his breath was particularly light and easy some time afterward. There is a direct connection between these experiments and some of the current medical uses of pure oxygen. For instance, oxygen is administered to pneumonia patients who have reduced lung capacity and therefore are unable to breathe sufficient oxygen from the air. Also, firefighters encountering suffocating fumes, rescue parties entering mines, and aviators and mountain climbers who reach altitudes where the air is very thin all carry tanks of oxygen.

Priestly also discovered many other gases such as ammonia, hydrochloric acid gas, nitric oxide, nitrous oxide, nitrogen dioxide, nitrogen, carbon monoxide, and sulfur dioxide. He was among the first scientists to recognize that green plants make oxygen, which is responsible for restoring the goodness of air vitiated by putrefaction, or the burning of candles, or the respiration of animals through a process mediated by light. This process was elucidated in the 20th century as photosynthesis. Priestly also observed that inflammable air (H_2) is rapidly absorbed by heated metallic oxides (calces) in a process in which the latter are reduced, with water being formed if the gas is dry. He erroneously associated hydrogen with phlogiston. He also made the observation that water is formed in the explosion of inflammable air with common air or dephlogisticated air (known to us as oxygen).[62]

Carl Wilhelm Scheele (1742–1786) (Fig. 4.4) was another pioneering chemist of the modern era. Scheele was actually the first discoverer of oxygen (1772–1774) and first published this finding in 1777. He made other important discoveries, including those of (1) chlorine and manganese, (2) silicon fluoride and hydrofluoric acid from fluorspar, (3) phosphorus from bone ash and phosphoric acid from the action of nitric acid on phosphorus, (4) arsenic acid, molybydic acid, tungstic acid, arsenic anhydride, and copper arsenite (also called "Scheele's green"), and the distinction between molybdenite (MoS_2) and graphite, (5) several organic compounds, including tartaric acid, mucic acid, lactic acid, uric acid, prussic acid, oxalic acid, citric acid, malic acid, gallic acid, and pyrogallic acid, (6) glycerol and casein, (7) hydrogen sulfide, (8) the action of light on silver salts, (9) the distinction between nitrous acid, nitric acid, nitric oxide, nitrous oxide, nitrogen dioxide, and nitric acid, and (10) the formation of cyanide by the action of ammonia on a mixture of charcoal or graphite and potassium carbonate. He proposed the peculiar theory that phlogiston + fire air (oxygen) = heat.[63]

---

[62] J. Priestley, *Experiments and Observations on Different Kinds of Air* (3 vols.), London (1774–1777); J. Priestley, *Experiments and Observations Relating to Various Branches of Natural Philosophy* (3 vols., vol. i London, vols. ii-iii Birmingham) (1779–1786) [also cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 110–121 (1989)].

[63] ibid., pp. 104–109.

Figure 4.4 Statue of Karl Wilhelm Scheele (1742–1786), who is considered as the founder of the science of photochemistry (printed with permission of the Deutsches Museum, Munich).

4.2.8 Absorption of light

The way in which light is absorbed was studied by Johann Heinrich Lambert (1728–1777), who first expressed it in this differential form:

$$\frac{dI}{dz} = -\alpha I, \tag{4.1}$$

where I is the intensity of light traveling in the z direction through a medium, and α is the absorption coefficient of the medium and has units of inverse length. In a homogenous medium (i.e., where α is not a function of z), Eq. (4.1) may be integrated to yield

$$\frac{I(z)}{I_0} = e^{-\alpha z}, \tag{4.2}$$

which can be rearranged to yield

$$-\ln\frac{I(z)}{I_0} = \alpha z = A, \tag{4.3}$$

where z is the distance the light has traveled through the medium, I_0 is the intensity at $z = 0$, and A is the absorbance. The ratio of the transmitted light to the incident light [left-hand side of Eq. (4.2)] is the transmission of the light. In 1852 Augustus Beer (1825–1863) showed that for dilute solutions the absorption coefficient is proportional to the concentration of the absorbing species in the solution:

$$\alpha_{\text{solution}} = ac, \tag{4.4}$$

where a is the molar absorption coefficient (sometimes called the molar extinction coefficient) of the absorbing species, and c is the concentration.

Equations (4.2) and (4.3) are now referred to as Lambert's law, which in essence states that there is a logarithmic dependence between the transmission of light through a substance and the concentration of the substance, as well as between the transmission and the length of the material through which the light travels. Today, Lambert's law serves as the starting point of all lithographic modeling, along with the radiation chemistry/photochemistry that accompanies it.

4.2.9 Chemical effects of light

To complement the research efforts targeted at elucidating and controlling the physical characteristics of light, active research was also conducted on the chemical effects of light, dating back to the early decades of the 18th century. In fact, the roots of the science of photochemistry date back to this era, when in 1727 Johann Heinrich Schulze (1687–1744) (Fig. 4.5) discovered that if a bottle of silver nitrate crystals was placed in the sun and part of the bottle was covered with a stencil, the uncovered area would blacken. Inspired by Schulze's discovery, other experimenters discovered various photosensitive metal compounds and other silver compounds that, in time, were used in the production of silver nitrate ink and other inks used to produce permanent designs on a variety of substrates, including paper, textiles, bones and ivory, and leather. Silver nitrate was also used as a hair dye during this time.[64]

In 1777, Carl Wilhelm Scheele (1742–1786) (Fig. 4.4) published the first comprehensive book on the chemical effects of light. Considered the father of photochemistry, Scheele used his experiments to try to demonstrate that light

---

[64] M.S. Barge and W.B. White, *The Daguerreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, pp. 13–14 (1991).

Figure 4.5 Johan Heinrich Schulze (1687–1744), who discovered the darkening of silver nitrate crystals under light irradiation (printed with permission of the Deutsches Museum, Munich).

was composed of, or contained, phlogiston,[65] a fact that he deduced, rather erroneously, when he observed that when silver oxide, gold oxide, or mercury oxide was placed at the focus of a burning glass (a parabolic mirror used to focus the rays of a sun on an object to be burned), these salts changed to metal.[66]

Another major contributor to research in the chemical effects of light during this period was the Frenchman Antoine Lavoisier (1734–1794) (Fig. 4.6), who introduced into the field of chemistry accurate methods of weight measurements of materials undergoing chemical reactions. He believed that light was a true reagent, which when added to some compounds could alter their basic properties or cause a reaction. In time, testing for reactivity in light became part of the regular chemical routine of investigating new compounds and elements.[67]

The science of photometry—the measurement of the effect of light intensity, whether transmitted or absorbed, on various materials—provided

---

[65] Phlogiston was thought to be a substance that exists in all combustible materials, which upon combustion is released. It was an important chemical concept introduced at the beginning of the 18th century and maintained until the death of Joseph Priestly in 1802. Although Priestly discovered oxygen in 1774, he, like many scientists of his era, did not recognize that it was the substance responsible for combustion. Consequently, the development of chemistry was stifled due to this erroneous concept. Antoine Lavoisier made the connection that oxygen is responsible for combustion and in so doing sounded the death knell for phlogiston.

[66] M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 14 (1991).

[67] ibid., p. 14.

Figure 4.6 Antoine Laurent Lavoisier (1743–1794), the founder of modern chemistry. His method of rigid quantification, as well as his analytical definition of an element, along with his theory of combustion initiated the 'chemical revolution.' Both chemistry and physiology owe their modern forms to him. (Printed with permission of the Deutsches Museum, Munich.)

another impetus for research on the chemical effects of light during this period. These studies[68] had broad focus, spanning research on astronomy and the chemical effects of light. Very basic photometers were made with paper saturated with silver nitrate, silver chloride, or other light-sensitive solutions.[69]

Another major development in the research on the chemical effects of light was *spectral analysis*, the forerunner of modern spectroscopy. This entails the study of the physical and chemical effects of light that has been dispersed into a spectrum by being passed first through a slit and then through a prism.[70] In 1752, the Scotsman Thomas Melvill (1726–1753), when studying the light of a flame passed through a prism, discovered that the spectrum was not continuous like that of the sun; rather, parts of the spectrum were bright, and others were dark. He observed that the locations of the bright spots and their colors were different when different chemicals were placed in the flame.[71]

---

[68] Many illustrious scientists, including Thomas Young, François Arago, John Herschel, and Simeon Poisson, conducted such experiments in photometry [cited in F. Arago, *Biographies of Distinguished Scientific Men*, (transl. by W.H. Smyth, B. Powell, and R. Grant), 2nd series, Tricknor & Fields, Boston (1859); M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 14 (1991)].

[69] M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, p. 14 (1991).

[70] ibid., p. 15.

[71] M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 81 (1987).

In addition, spectral light was used to study gases and flames colored by the addition of various metal salts and transparent liquids, as well as to monitor the effects of photochemical changes. Based on his work on spectral analysis, Scheele established that there is a difference between actions of light and heat. This line of research would subsequently lead to the discovery of the infrared region of the spectrum in 1800 by Friedrich Wilhelm Herschel (discussed in Chapter 3), and of the ultraviolet region in 1801 by J.W. Ritter (discussed in Chapter 3). It also led to William Hyde Wollaston's 1802 discovery of the ultraviolet region of the spectrum, which he referred as chemical rays.[72]

In the final analysis, a great deal of both theoretical and empirical activity concerning light and its effect on a wide variety of materials was already well established by the early part of the 19th century. In particular, many chemical discoveries that showed the range of both light-sensitive materials and their properties in a variety of conditions were made. These discoveries comprised light-sensitive materials such as silver nitrate, silver chloride, silver oxalate, mercury oxalate, chromate and citrate salts, iron salts, gold salts, and platinum salts; compounds that react with light sensitive-materials, such as sodium thiosulfate; and elements that react with other materials to form light-sensitive compounds, such as chlorine, iodine, and bromine.[73] During this period, the use of optical instruments was particularly widespread among artists, artisans, and scientists.[74] It was in this milieu of abundant data on the chemical effects of light on materials, along with the widespread availability of optical instruments in the early part of the 19th century, that the invention of photolithography and its subsequent development were made possible.

4.2.10 Foundation of modern chemistry

Antoine Laurent Lavoisier (1743–1794)[75] (see Fig. 4.6) is universally acknowledged as the founder of modern chemistry. Although he discovered

---

[72] ibid., pp. 14–15; M.I. Sobel, *Light*, University of Chicago Press, Chicago, p. 81 (1987).

[73] M.S. Barge and W.B. White, *The Dagurreotype: Nineteenth-Century Technology and Modern Science*, Johns Hopkins University Press, Baltimore, pp. 15–16 (1991).

[74] ibid., p. 16.

[75] Recent biographies include: A. Donovan, *Antoine Lavoisier: Science, Administration, and Revolution*, Blackwell, Oxford (1993); J.-P. Poirier, *Lavoisier: Chemist, Biologist, Economist*, University of Pennsylvania Press, Philadelphia (1996) [originally published as *Antoine Laurent de Lavoisier, 1743-1794*, Editions Pygmalion/Gerard Watelet, Paris (1993)]. For an excellent review of these books, see E.M. Melhado, "Scientific biography and scientific revolution: Lavoisier and eighteenth-century chemistry," *Isis* **87**, 688–694 (1996). For a guide to earlier biographies, see H. Guerlac, "Lavoisier and his Biographers," *Isis* **45**, 51–62 (1954). Other notable references include: M. Beretta, Ed., *Lavoisier in Perspective*, Deutsches Museum, Munich (2005); E. Grimaux, *Lavoisier*, Paris (1888); D. Mckie, *Antoine Lavoisier: The Father of Modern Chemistry* (1936); *Notes and Records of the Royal Society* **vii**, 1 (1949); A.L. Lavoisier, *Œuvres de Lavoisier*, 6 vols., Editions Albin Michel, A. Blanchard, Paris (1864–1893);

no new substances, devised no truly novel apparatus, and worked out no improved methods of preparation, modern chemistry and physiology could not exist in their present form without the contributions of Lavoisier. It has been said that: “Lavoisier, though a great architect in the science, labored little in the quarry; his materials were chiefly shaped to his hand, and his skill was displayed in their arrangement and combination.”[76] He brought the work of Black, Priestley, and Cavendish to their logical conclusions and provided accurate interpretation of their experimental results.[77]

Although he was not the discoverer of oxygen, he was certainly the first person to recognize the consequences of its discovery and the first to realize its true nature as an element, and by carrying out ingenious quantitative experiments, he became the first to establish the true chemistry of combustion and calcinations of metals that had been foreshadowed by the work of Hooke and Mayow.[78] The following is a brief account of how Lavoisier made the above accomplishments.

History records that in October of 1774 Priestly visited Paris.[79] There he met Lavoisier, to whom he described his latest discovery—the new ‘air’ in which the flame of a candle burned much more brilliantly than in common air. He also told Lavoisier how he obtained this ‘air’ by heating the calx (oxide) of mercury or the calx (oxide) of lead.[80]

Lavoisier, it should be noted, had long been working on the related problems of combustion and calcination and, as early as 1772, had reached the conclusion that air played an important part in combustion and that the two inflammable substances, phosphorus and sulfur, combined with air when they were burned and that their weight was increased by this combination of air. He showed that tin and lead increase in weight on calcinations, and that an equal weight of air is absorbed (1774).[81] Earlier (1770), he had proved that

---

W.A. Tilden, *Famous Chemists*, G. Routledge & Sons, New York, p. 63 (1921); J.A. Cochrane, *Lavoisier*, Constable & Co., London (1931); P.J. Hartog, “The new views of Priestley and Lavoisier,” *Annals of Science* **V**, 1–56 (1941); J.R. Partington, *Nature* **clii**, 207 (1943); D.I. Duveen and H.S. Klisckstein, *A Bibliography of the Works of Antoine Laurent Lavoisier, 1745–1794*, William Dawson & Sons, London (1954) [cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 122 (1989)].

[76] W.T. Brande, *A Manual of Chemistry*, 6th ed., pp. i–xxxv (1848) [cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 122 (1989)].

[77] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 123 (1989).

[78] ibid., pp. 129–130.

[79] ibid., p. 126; B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 77 (1976); M.S. Bell, *Lavoisier in the Year One: The Birth of a New Science in an Age of Revolution*, Atlas Books, New York, p. 98 (2005).

[80] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 126 (1989).

[81] ibid., pp. 125–126.

water cannot be converted into earth,[82] as van Helmont had previously thought.[83]

Through many months of the year 1773, Lavoisier made further experiments; in November of 1774, after his meeting with Priestley, he tried Priestley's experiment of heating mercury oxide with a burning glass. He repeated the same experiment in March 1775 and concluded that the gain in weight of metals on calcination was a result of their combination with air; however, he still got no farther than supposing that it was their combination with common air in a pure state, without suspecting that only a part or constituent of the common air was involved.[84]

The year 1777 saw him in further experimentation, in which he reached the seminal conclusion that only a part of the air was involved in combustion, respiration, and calcination—the heavier part of the air—and that air itself was not a simple substance but rather consisted of two 'airs.' One of these was respirable, supported combustion, and combined with metals on calcination; this he called the 'salubrious part' of the air. The other part, he concluded, was incapable of supporting combustion or respiration, was inimical both to fire and to life, and played no part in calcination.[85]

On 3 May 1777, Lavoisier read to the Paris Academy of Sciences an account of one of the most significant experiments in the entire annals of science. On a charcoal furnace, he heated 4 ounces of mercury for 12 days in an apparatus consisting of a glass vessel (the neck of which was bent to connect with the air in a bell jar) inverted over a mercury trough. After 12 days, he observed that a red calx or oxide of mercury formed in the glass vessel and the level of water rose in the bell jar to a level corresponding to a decrease of about one-fifth of the total volume of air in the apparatus. The residual air extinguished lighted candles and asphyxiated animals. Lavoisier called this 'azote' (now called nitrogen).[86]

He then heated the red particles of mercuric oxide in a separate apparatus and collected the gas evolved from them. This gas amounted to about one-fifth of the original air and supported combustion and respiration. Lavoisier called this gas 'oxygen.' On mixing the oxygen with the air from which it had been removed, he was able to reconstitute common air once again. Thus, in this classic experiment, Lavoisier was able to show that common air consists largely of nitrogen and oxygen, both of which have strikingly different, indeed opposite, properties; and he separated these properties one from the other.[87]

---

[82] ibid., p. 124.

[83] ibid., p. 41.

[84] ibid., p. 127.

[85] ibid., pp. 127–128.

[86] ibid., p. 128.

[87] ibid., pp. 128–129.

In 1780, Lavoisier formally presented his theory of combustion as follows:[88]

1. "In every combustion there is disengagement of the matter of fire or of light."
2. "A body can burn only in pure air (oxygen gas)."
3. "There is 'destruction or decomposition of pure air,' and the increase in weight of the body burned is exactly equal to the weight of the air 'destroyed or decomposed.'"
4. "The burned body changes into an acid by addition of the substance that increases its weight."
5. "Pure air is a compound of the matter of fire or of light with a base. In combustion the burning body removes the base, which attracts more strongly than does the matter of heat, and sets free the combined matter of heat, which appears as flame, heat, and light."

In experiments designed to elucidate the composition of water (1783–1784), he decomposed water into its constituent parts: oxygen and inflammable air, which he later renamed 'hydrogen' (i.e., 'water former').[89]

Using an ice calorimeter, Lavoisier and Laplace in 1783 measured the specific heats and the heat evolved in combustion and respiration, thus laying the foundations of thermochemistry on both experimental and theoretical grounds since the authors stated the principle that "as much heat is required to decompose a compound as is liberated on its formation from the elements."[90]

He also showed that, in chemical reactions, matter is neither created nor destroyed; the total weight of the products of a chemical reaction equals the total weight of the starting reactants. All of these observations are encapsulated in the law of indestructibility of matter, which in 1789 he stated as follows:

> ... "for nothing is created in the operations either of art or nature, and it can be taken as an axiom that in every operation an equal quantity of matter exists both before and after the operation; that the quality and quantity of the principles remain the same and that only changes and modifications occur. The whole art of making experiments in chemistry is founded on this principle: we must always suppose an exact equality or equation between the principles of the bodies examined and those of the products of the analysis."[91]

---

[88] A.L. Lavoisier, *On Combustion in General* (1780) [cited in A.L. Lavoisier, *Œuvres de Lavoisier*, 6 vols., Paris, ii, 225 (1864–1893) and also in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 131 (1989)].

[89] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 142–147 (1989).

[90] A.L. Lavoisier, *Œuvres de Lavoisier*, 6 vols., Paris, ii, 283 (1864–1893); J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 132–133 (1989).

[91] A.L. Lavoisier, "*Traité Élémentaire de Chimie*," Chuchet, Paris, p. 140 (1789).

Chemistry was thus effectively established on a quantitative basis, and the first chemical balance sheet was then drawn up.

A very significant immediate result of these great advances was the reform of the language of chemistry. The old names of substances had, of course, no relation to their chemical composition since that was unknown at the time; and their chemical composition could not have been determined until it was discovered of what elements they were composed.

Lavoisier defined an "element or principle" as the "last point which analysis is capable of reaching, since we have not hitherto discovered the means of separating them, they act with regard to us as simple substances, and we ought never to suppose them compounded until experiment and observation have proved them to be so,"[92] which is essentially our modern definition of a chemical element. It must be pointed out that he adopted Robert Boyle's 1661 definition of an element as a substance that cannot be decomposed into anything simpler. He applied this idea in a very cautious way, not asserting that such substances (as they appeared to be incapable of further decomposition) were, in fact, elements, but that they should be regarded as elements until evidence to the contrary was forthcoming. He generated the first table of chemical elements, which included oxygen, hydrogen, nitrogen, sulfur, phosphorus, carbon, and a large number of metals.[93]

Having thus determined which substances were to be regarded as elements, Lavoisier and some of the French chemists who had adopted his view took it upon themselves to devise a suitable system of chemical nomenclature that emphasized giving every known substance a name that corresponded to its chemical composition. The old names that had been given before Lavoisier's chemical nomenclature came into being were often coined to indicate some physical property of a substance or its mode of preparation, or to perpetuate its discoverer's name or the place where it had been found as a mineral.

The language of science, Lavoisier maintained in his *Méthode de Nomenclature Chimique*, was itself an analytical instrument. The system of chemical nomenclature that he devised provided for every substance, in a word or two, a clean indication of its chemical nature and composition. The fact that even after more than two centuries of further discovery since he devised this system it has survived until today with very little modification is but a testament to his exceptional vision. Indeed, Lavoisier's influence on the course of modern chemistry cannot be overemphasized.

In all, Lavoisier's contributions span the entire history of modern chemistry: from the concept of chemical elements to the law of

---

[92] A.L. Lavoisier, "*Traité Élémentaire de Chimie*," Chuchet, Paris, p. 192 (1789).

[93] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 134–135 (1989).

indestructibility of matter and conservation of mass in chemical transformations, from the theories of combustion to the foundation of thermochemistry, and from chemical nomenclature to the mathematization and quantification of chemistry. The French chemist Marcellin Bertholet (1827–1907) describes Lavoisier's influence on chemistry as "*La revolution chimique*." It has been noted that a book on chemistry written before his time would not be intelligible to a student unacquainted with the history of chemistry, while Lavoisier's *Traité Élémentaire de Chimie* reads like a rather old edition of a modern textbook.[94]

The revolution in chemistry was complete; modern chemistry dates from the publication of Lavoisier's great classic, the *Traité Élémentaire de Chimie*,[95] in Paris in 1789. From this work, which ranks with Newton's *Principia* as one of the greatest of all scientific books, we may quote one sentence that particularly reveals its author's method, and that embodies his fondness for mathematical precision, his determination to give priority to experimental evidence, and his mistrust of received ideas:[96] "I have imposed upon myself, as a law, never to advance but from what is known to what is unknown, never to form any conclusion which is not an immediate consequence necessarily flowing from observation and experiments; and always to arrange the facts and the conclusions which are drawn from them, in such an order as to render it most easy for beginners in the study of chemistry to thoroughly understand them."

Equally impressive is Lavoisier's role in the discovery of the one element that is at the heart of the entire semiconductor industry and that, through the agency of lithography, is fabricated into digital and analog devices: silicon. Having generated or reduced several oxides, he had suggested in 1789 that quartz, or silica (or silicon dioxide), was probably the result of the union of oxygen with an as yet unidentified element. Since his life was cut short,[97] it was left to others to isolate and investigate the chemistry, as well as the technological applications, of silicon.

4.2.11 Post-Lavoisian evolution of chemistry

Chemistry under the influence of Lavoisier became a distinct discipline—a discipline with its own instruments, techniques, and specific technical language.[98] Even if pre-Lavoisian chemistry could be regarded as a discipline, as a set of investigative and interpretive practices, as a livelihood, and as a tradition of texts, according to Donovan, it lacked the methodological rigor to

---

[94] ibid., p. 130.

[95] A.L. Lavoisier, "*Traité Élémentaire de Chimie*," Chuchet, Paris (1789).

[96] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 130 (1989).

[97] Lavoisier was guillotined in 1794 during the reign of terror of the French Revolution.

[98] M. Crosland, "Research schools in chemistry from Lavoisier to Wurtz," *British J. for the History of Science* **36**, 333–361 (2003).

become a science.[99] In the post-Lavoisian era, chemistry developed into chemical principles such as stoichiometry, affinity, law of reaction (such as definite proportions used in determining chemical compositions), and a variety of chemical equivalents, to mention but a few. It was the successive evolutions of these doctrines that led to the atomic theory, Dalton's grand idea that changed the face of the science for good:[100]

> "... this hypothesis of atoms forms at present the foundation of all our theories, the solid base of our system of chemical knowledge. It gives a striking simplicity to the laws relating to the composition of bodies; it enables us to look into their intimate structure; it intervenes in the interpretation of their properties, reactions, and transformation; and will doubtless at some future time furnish points of support for the science of molecular mechanics."[101]

Today, the most advanced chemical principles are based on quantum mechanics and have led to remarkable developments in spectroscopy, lasers, materials synthesis, etc. All of these developments have aided the discovery and isolation of new elements, compounds, and molecules. We briefly outline a few of these developments in the following sections.

4.3 Laws and Theories of Chemical Reactions

After establishing chemistry on its modern course following the demise of the phlogistic theory of matter, chemists of the late 18th century and early 19th century turned their attention to understanding the physical basis of the chemical phenomena that underpin the formulated laws and theories governing these phenomena. Some of these theories and related developments are reviewed below.

4.3.1 Atomic theory

The origin of atomic theory dates back to the early Greek philosophers.[102] Boyle in 1661 wrote about "those theories of former philosophers, which are now with great applause revived, as discovered by latter ones."[103]

---

[99] A. Donovan, "Lavoisier as chemist and experimental physicist: A reply to Perrin," *Isis* **81**, 270–272 (1990).

[100] A. Wurtz, *A History of Chemical Theory*, Macmillan, London, Preface (1869).

[101] ibid., p. 194.

[102] J. Burnet, *Early Greek Philosophy*, 3rd ed., A. & C. Black, London (1920); I. Freund, *The Study of Chemical Composition*, Cambridge University Press, p. 226 (1904); J.R. Partington, "Origins of the atomic theory," *Annals of Science* **IV**, 245–282 (1939); J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 164 (1989).

[103] R. Boyle, *The Sceptical Chymist*, J. Caldwell, London, p. 23 (1661).

Anaxagoras of Klazomenai (near Smyrna) (498–428 BC) held that bodies are divisible without limit and retain their characteristics.[104] Zeno of Elea (born 489 BC), who was part of the Eleatic School, believed that matter is continuous, and since the universe is full of matter, there is no motion. He invented a number of ingenious paradoxes, such as that of "Achilles and the tortoise," to prove that things do not really move; we only think they do. The Eleatics held that "without vacuum there is no motion; there is no vacuum, hence there is no motion."[105]

Aristotle credits Leucippus (about 450 BC) with founding the atomic theory. Aristotle wrote that, starting from the same premise as the Eleatics, Leucippus arrives at an opposite conclusion from them, articulated as: "without vacuum there is no motion; there is motion, hence there is vacuum."[106] Leucippus had postulated that everything consists of tiny particles of various kinds, separated by space through which they travel.

The greatest exponent of atomic theory in antiquity was Demokritos of Abdera (460–370 BC), who adopted Leucippus' ideas and refined them. Demokritos believed that matter was composed of empty space and an infinite number of indivisible small particles called atoms. To him atoms were indivisible by reason of their very small size. He also believed that atoms are hard and have form and size. To Demokritos, atoms move spontaneously and ceaselessly in vacuum; they come together by necessity and form aggregates by a sort of hook-and-eye mechanism, not by attractive forces. The motion of atoms, according to Demokritos, "is like that of dust particles seen in a sunbeam in still air in a room."[107] This "necessity" of Demokritos is not unlike the statistical considerations of the kinetic theory of gases postulated by Boltzmann and Maxwell more than 2000 years later.

Epikouros (341–270 BC) adopted the atomic theory and attributed weights to the atoms. To him, the atoms fall perpendicularly through vacuum; however, at indetermined times and in indetermined places, they "swerve" and so enter into collision. The concept of "swerving" has some resemblance to the modern theory of indeterminancy.[108]

Following the early Greek era, interest in atomic theory waned and was only revived by the writings of Gassendi (1592–1655). For instance, Boyle and Lemery made extensive use of it.[109] Boyle states: "It seems not absurd to conceive that at the first Production of mixt Bodies, the Universal matter

---

[104] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 165–166 (1989).

[105] ibid., p. 164.

[106] ibid., p. 164.

[107] ibid., p. 165.

[108] ibid., p. 165.

[109] J.C. Gregory, *A Short History of Atomism*, Macmillan Co., London (1931).

whereof they among other Parts of the Universe consisted, was actually divided into little particles of several sizes and shapes variously mov'd."[110]

Another ardent atomist was Sir Isaac Newton. Newton assumed that "particles attract one another by some force, which in immediate contact is exceedingly strong, at small distances performs the chymical operations, and reaches not far from the particles with any sensible effect."[111] With the assumption that a gas is made of particles repelling one another with a force varying inversely as the distance, Newton deduced Boyle's law.[112] Furthermore, he rejected Descartes' idea that by their endless encounters some of the particles might have suffered attrition and become "old worn particles" different from the rest.[113]

The modern formulation of atomic theory is attributed to John Dalton (1766–1844), who stated it in this form:[114]

1. "The chemical elements are made of very minute indivisible particles of matter, called atoms, which preserve their individuality in all chemical changes."
2. "All atoms of the same element are identical in every respect, particularly in weight. Different elements have atoms that differ in weight. Each element is characterized by the weight of its atom."
3. "Chemical combination occurs by the union of the atoms of the elements in simple numerical ratios, e.g., 1 atom A + 1 atom B; 1 atom A + 2 atoms B; 2 atoms A + 1 atom B; 2 atoms A + 2 atoms B, etc."

Dalton's formulation of atomic theory differs from those proposed by his ancient forerunners. While Demokritos had declared the atoms to be infinite in number and infinitely various in form, Dalton postulated that the atoms of the same element are all alike, but the atoms of different elements differ in both shape and weight. Dalton also believed that the weights of the atoms of each element are always fixed and never vary.

Dalton's assumptions of atomic theory explain very well the laws of constant, multiple, and equivalent proportions (see the following sections).

---

[110] R. Boyle, *The Sceptical Chymist*, J. Caldwell, London, p. 23 (1661).

[111] I. Newton, *Opticks*, Query 31, William Innys, London, p. 364 (1730).

[112] I. Newton, *Principia*, Book 2, Prop. 23, Theorem 18, London, p. 301 (1687).

[113] I. Newton, *Opticks*, William Innys, London, p. 376 (1730).

[114] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 169 (1989). In a lecture at the Royal Institution in 1810, Dalton attributed the origin of this atomic theory to attempts to explain his law of partial pressures (1801–1802), which states that the pressure exerted by a mixture of perfect gases is the sum of the pressures exerted by the individual gases occupying the same volume. The partial pressure of a gas is the pressure a gas would exert if it alone occupied the container and if it behaved perfectly. Dalton's law of partial pressure is a more general form of Henry's law, which states that the amount of gas absorbed by a liquid is proportional to the pressure. [Cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 174 (1989).]

However, the theory could not predict the relative weights of atoms from the combining proportions unless the number of atoms in the particle of the compound is known, and this for many years remained the main point of dispute[115] in the theory.

Today, many centuries after Dalton proposed his atomic theory, an overwhelming preponderance of scientific evidence supports this theory.[116] All things considered, Dalton's atomic theory remains today one of the cardinal pillars of the edifice of chemistry, without which, chemistry would have continued to consist of nothing more than a mass of heterogeneous observations and recipes for performing experiments, or for manufacturing metals at best.

4.3.2 The law of constant or definite proportions

The first of these chemical reaction laws, the *law of constant or definite proportions* was recognized in 1797 by Joseph Louis Proust (1754–1826), who postulated that when elements combine to form chemical compounds, the elements unite in definite proportions by weight. He stated this in 1799 and 1808 as follows:

> "We must recognize an invisible hand which holds the balance in the formation of compounds. A compound is a substance to which Nature assigns fixed ratios, it is, in short, a being which Nature never creates other than balance in hand, *pondere et mensura*."[117]

Proust had been led to this law from his observation that several metals can form more than one oxide and sulfide, each of which has a definite composition.

What a remarkable order Proust had found in nature! That even "the stones and soil beneath our feet, and the ponderable mountains, are not mere confused masses of matter; they are pervaded through their innermost constitution by the harmony of numbers."[118] Kepler, Galileo, and Newton, as well as Pythagoras 2000 years before them, regarded nature as mathematical. And here was yet further proof, this time provided by Proust, that nature is

---

[115] Dalton attempted to resolve the discrepancy by making use of two general rules: (1) If only one compound of two elements is known, it is presumed to be binary (1 atom A + 1 atom B). (2) When two compounds exist, they are assumed to be binary (A + B) and ternary (A + 2B or 2A + B). When three compounds are observed, one is binary, two are ternary, and one is quaternary (3A + B or A + 3B). [Cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 169–170 (1989).]

[116] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 84–99 (1976).

[117] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 153–154 (1989).

[118] Cited in B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 92 (1976).

indeed mathematical. The composition of every true compound remains constant. This law of definite composition is a fundamental principle of chemistry.

This law, which for the first time made chemistry a mathematical science, follows directly from Dalton's atomic theory. For, if the weight of the atom of every single element is constant, as he has postulated in his atomic theory, then the composition of all compounds must be definite since all chemical reactions involve the combination of these minute immutable atoms.

4.3.3 The law of multiple proportions

First recognized by John Dalton in 1803 as a consequence of his atomic theory, this law states that when two elements combine to form more than one compound, the weights of one element that unite with identical weights of the other are in simple multiple proportions.[119] Dalton had theorized that the atoms of different elements have different weights. But being unable to determine the weight of the individual atoms (after all, mass spectroscopy and instruments for measuring atomic weight had not yet been invented), the best he could do was to find a way to compare the weights of substances he was able to measure against the weight of the lightest known substance. Thus was borne the concept of relative atomic weight, where the weight of the lightest substance known—hydrogen gas—was taken as the standard to which the weights of other substances were to be compared. In this way, the weight of hydrogen gas was assigned the value of 1 and the weights of the other substances were assigned relative to the value for hydrogen.[120]

Dalton knew that hydrogen and oxygen unite in the ratio of 1 to 7 by weight. Therefore, he assigned the number 7 to the relative weight of the atom of oxygen. In this way, he prepared the first table of relative atomic weights—a table of 14 elements, which remains a cornerstone of chemical calculations today.

While working on relative atomic weights, Dalton observed a curious mathematical relationship. Carbon reacts with oxygen in the ratio 3:4 to form carbon monoxide (CO). Carbon also reacts with oxygen to form carbon dioxide (CO_2) in the ratio 3:8. Why number 8, which is a perfect multiple of 4, he thought to himself. He found similar relationships between the oxides of nitrogen. Here the same amount of nitrogen reacts with one, two, and four parts of oxygen to form three distinct compounds: N_2O, NO, and NO_2. He also studied two other gases, methane and ethylene, and found that methane contains exactly twice as much hydrogen as ethylene. From these simple relationships, Dalton deduced the law of multiple proportions. Berzelius (see Section 4.3.5)

---

[119] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 158–159 (1989).

[120] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 93–95 (1976).

later stated this law as follows: "In a series of compounds made up of the same elements, a simple ratio exists between the weight of one and the fixed weight of the other."[121]

4.3.4 The law of reciprocal proportions or the law of equivalents

Cavendish established in 1766 that identical weights of a given acid require different weights of different bases for neutralization, and he called these weights of the bases *equivalents*.[122] His statement relates the proportions in which elements combine with each other. It was Jeremias Richter (1762–1807) who in 1795 generalized this observation into the law of reciprocal proportions: "If two elements combine or react separately with the same weight of a third element, the ratio of the masses in which they combine are either the same or a simple multiple of the mass ratio in which they combine."[123]

A simple illustration of this law is as follows. If element A reacts with element B and also with element C, then, if B and C react, the proportion by weight in which they react will be simply related to the weights of B and C, which separately react with a constant weight of A. In other words, if we know the proportion of elements in compound AB and the proportion of elements in compound BC, we can determine the proportion of elements in AC. This law is useful in understanding stoichiometry or determining the quantities of reactants and products in relation to their reactions. Take, for instance, the molecule methane (CH_4), which is made up of one atom of carbon and four atoms of hydrogen. The atomic mass of carbon is 12 atomic mass units (amu), and the atomic mass of hydrogen is 1 amu. Since we have 4 atoms of hydrogen for every atom of carbon, the atomic mass ratio of carbon to hydrogen in the molecule is 12:4, which can be simplified to 3:1. Next, let us consider the molecule water, H_2O. The atomic mass ratio of oxygen and hydrogen elements in water is 16:2 or 8:1 (since oxygen has an amu of 16 and hydrogen has an amu of 1).

Since methane and water both contain a hydrogen atom and one other element, according to this law, if we combine carbon (the other element in methane) and oxygen (the other elements in water), the proportion by weight of these elements in the product should be 3:8, or a simple multiple of that ratio. In other words, the mass ratio with which carbon reacts with oxygen should be the same as, or a simple multiple of, the mass ratio with which they separately react with a fixed mass of H. And indeed, the atomic mass ratio of these elements in CO_2 is 3:8 since the atomic mass of carbon is 12 and that of

---

[121] Cited in B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 94–95 (1976).

[122] ibid., pp. 159–160.

[123] J.B. Richter, *Uber die neurn Gegenstande der Chymie* **IV**, p. 67 (1795); This is also quoted in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 162–163 (1989).

oxygen is 16 × 2 = 32, yielding a ratio of 12:32 = 3:8; this corresponds to a ratio of the two ratios (3/1 ÷ 8/1 = 3/8) and to the fact that in methane, carbon is in a ratio of 3:1 relative to hydrogen, while in water, oxygen is in a ratio of 8:1 relative to hydrogen. Therefore, when carbon and oxygen combine, they form carbon dioxide (CO_2), which has a combination ratio of 12:32 = 3:8, exactly as predicted by the law.

Similarly, applying this law to sodium chloride (NaCl) and hydrochloric acid (HCl), we have for sodium chloride an atomic mass ratio of sodium to chloride atoms of 23:35, corresponding to the atomic mass units of sodium and chlorine atoms, respectively. For hydrochloric acid, the atomic mass ratio of hydrogen to chlorine atoms is 1:35. So when sodium combines with hydrogen, the mass ratio of these atoms in the compound should be 23:1, corresponding to sodium hydride (NaH).

4.3.5 Electrochemical theory

Luigi Aloisio Galvani (1737–1798) made a chance discovery in 1780 of active muscular contractions from the limbs of dead frogs when he observed that the leg muscles of dead frogs' legs hanging by a copper hook from an iron balcony contracted and wriggled when the carcasses touched the iron balcony. Galvani erroneously ascribed the muscular contractions to "animal electricity" that originates from within the animal. Alessandro Volta (1745–1827) thought otherwise and decided to find out the true source of the electricity. Gradually, he came to the conclusion that the produced electricity originated from the dissimilar metals completing a circuit through which electrical current could flow. The circuit was created when wind pushed the frog bodies (connected to the copper hook) against the iron balcony. To prove his hypothesis, in 1799 he connected a series of two dissimilar metals, zinc and silver, separated by a piece of cloth moistened in a solution of salt, and obtained a weak electric current. When he joined a larger series of these metals, he obtained a stronger flow of electricity. In this way, he invented the "voltaic pile," a forerunner of the modern storage electric battery.[124]

When news of this invention reached the scientific community, it did not take long for them to grasp the immense possibilities of this invention, especially with respect to using it to decompose hitherto unbreakable substances. Many scientists were soon building large voltaic piles and using them to decompose compounds such as water, acids, bases, salts, etc. Many metals were isolated in this way. Humphry Davy (1778–1829) in 1806 used a voltaic pile containing 150 cells to decompose molten potash and isolate potassium, based on the previous work of Berzelius.[125] Faraday also

---

[124] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 105–106 (1976).

[125] ibid., p. 106.

experimented with electrochemical decomposition, for example, of a solution of sodium sulfate into acid and alkali. For this experiment, he placed two current-carrying wires each onto two pieces of tin foil on a glass plate. The plate was placed over, but raised above, a piece of white paper 'so that shadows may not interfere' with the reaction. He connected one end of each of the two current-carrying wires to an electrostatic machine. Next, he joined a piece of litmus paper to a piece of turmeric paper, end to end, to test for acids and alkalis, respectively. Next, he twisted two short pieces of platinum wire to bridge the gap between the live tin foil and the piece of litmus, and between the other live thin foil and turmeric paper. Next, he dampened the litmus and turmeric papers in sodium sulfate solution. When he turned the electrostatic machine on to charge the wires, he observed the litmus paper turning red, indicating acid at the positive pole, and the turmeric turning brown, indicating alkali at the negative pole.[126] Then, he performed the next step in the experiment, described below in his own words:[127]

> "The pieces of litmus and turmeric paper were now placed each upon a separate piece of glass, and connected by an insulating string four feet long, moistened in the same solution of sulfate of soda; the terminal decomposing wire points were placed on the papers as before. On working the machine, the same evolution of acid and alkali appeared as in the former instance, and with equal readiness, nothwithstanding that the places of their appearance were four feet apart from each other."

Next, Faraday tried to determine how long the insulating string could be for him to still be able to decompose the sodium sulfate solution. He wrote:[128] "Finally a piece of string seventy feet long was used. It was insulated in the air by suspenders of silk, so that the electricity passed through its entire length: decomposition took place exactly, as in the former cases, alkali and acid appearing at the two extremities in their proper places."

Following the invention of the voltaic pile, and beyond its use in the decomposition of substances and isolation of new substances, attempts were made by major chemists of the day to elucidate the chemical effects of electricity. Out of this research came electrochemical theory, its main contributors being Jöns Jakob Berzelius (1779–1848) (Fig. 4.7) and Humphrey Davy. Other contributors to the theory included Volta, Faraday, and Lavoisier. Dualism was the basis of electrochemical theory. Oxygen, according to Lavoisier, was the central element in the system, so much so that

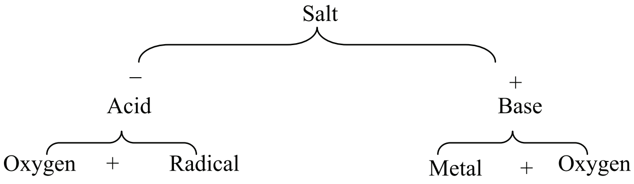
---

[126] J. Hamilton, *A Life of Discovery: Michael Faraday, Giant of the Scientific Revolution*, Random House, New York, p. 268 (2002).

[127] Cited in J. Hamilton, *A Life of Discovery: Michael Faraday, Giant of the Scientific Revolution*, Random House, New York, p. 268 (2002).

[128] ibid.

Figure 4.7 Jöns Jakob Berzelius (1779–1848), the great Swedish chemist and a major proponent of the electrochemical or dualistic theory. He produced elemental silicon in 1824. (Printed with permission of the Deutsches Museum, Munich.)



Scheme 4.2 Illustration of the dualistic or electrochemical theory of the formation of chemical compounds as consisting of two distinct species that are of opposite polarities. An acid is defined as a compound of oxygen and a radical. A base is defined as a compound of oxygen and a metal. A salt is defined as s compound of an acid and a base.

he defined an *acid* as a compound of a radical with oxygen. Davy extended this concept by showing that a *base* was a compound of a metal with oxygen, and Berzelius completed the dualistic system by assuming that, in all cases, a *salt* was a compound of an acid (actually an acid anhydride) with a base (actually a basic oxide), as illustrated in Scheme 4.2.[129]

Berzelius further noted that a similar dualism exists in electricity. This statement was influenced by his observation that salts in solution are decomposed by electric current into both bases, which travel to the negative pole and therefore seem to carry a positive charge, and acids, which travel to the positive pole and seem to carry a negative charge. Thus, he succeeded in attributing his observation to electrical polarity; bases were seen as electropositive oxides, and acids were seen as electronegative oxides—or, as Berzelius expressed it, they are electropositive and electronegative substances, respectively.[130]

---

[129] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 196 (1989).
[130] ibid., p. 197.

Davy postulated that chemical changes and electrical changes are distinct phenomena; both are produced by the same power, acting in one case (electrical) on masses, and in the other case (chemical) on particles. "The origin of the electromotive force," he says, "is contact, but when the metals are connected, the charges tend to become neutralized, and the chemical change restores the electromotive force." He observed that even apparently insoluble substances, such as glass, when moist, can be decomposed by the current, and acids may be transported through alkaline solutions, provided that they are not precipitated on the way.[131]

Based on his studies of electrolysis, Berzelius in 1803 reported the following conclusions:[132]

1. Chemical compounds are decomposed by an electric current, and their components collect at the poles.
2. Combustible bodies (hydrogen), alkalis, and earths and metals travel to the negative pole; oxygen, acids, and oxidized compounds travel to the positive pole.
3. The extent of decomposition is in compound proportion to the affinities and the surfaces of the poles; it is proportional to the quantity of electrical conductivity.
4. The chemical changes in decomposition depend firstly on the affinities of the components for the poles (metals), secondly on the affinities of the components for one another, and thirdly on the cohesion of the compounds formed.

According to Berzelius, elements may be divided into two groups based on the charges they take up on contact. The arrangement of elements according to electron affinities is called the electrochemical series. Berzelius regarded oxygen as the most electronegative element, as evidenced by the fact that in an oxygen compound, oxygen is always negative. He arranged all of the elements in a series with oxygen at the top and potassium (the most electropositive element) at the bottom, and with hydrogen separating the two classes (in an arrangement that does not differ much from modern tables of electrode potentials):[133]

O, S, N, F, Cl, Br, I, Se, P, As, Cr, Mo, W, B, C, Sb, Te, Ta, Ti, Si, $\mathbb{H}$;
Au, Os, Ir, Pt, Rh, Pd, Hg, Ag, Cu, U, Bi, Sn, Pb, Cd, Co, Ni, Fe, Zn,
Mn, Ce, Th, Zr, Al, Y, Be, Mg, Ca, Sr, Ba, Li, Na, K

The order may vary with temperature. Berzelius believed that the greater the strength of an acid or base, the more pronounced is the electrochemical

---

[131] H. Davy, *Elements of Chemical Philosophy*, J. Johnson & Co., London, p. 160 (1812).

[132] J.J. Berzelius and N. Hisinger, "Versuch, betreffend die Wirkung der elektrischen Säule auf Salze und auf einige von von ihren Basen," *Allgem. J. Chem.* **1**, 115 (1803) [cited by J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 192 (1989)].

[133] ibid., p. 200.

character of the radical. He also believed that a weak base may act as an acid toward a strong base (e.g., alumina and caustic soda), and a weak acid may act as a base toward a strong acid (e.g., boric acid and sulfuric acid).[134]

Another major contribution Berzelius made to chemistry was the introduction of a rational system of chemical shorthand, in which each element is represented by the initial letter or first two letters (in the cases where two elements have the same initial letter) of its Latin name, e.g., C for carbon, H for hydrogen, N for nitrogen, O for oxygen, P for phosphorus, and S for sulfur. He also introduced the system of joining the symbols of the elements to represent the simplest parts of compounds, e.g., copper oxide was represented as CuO. Additionally, he initiated the standard of writing algebraic exponents to designate more than one atom of an element present in a compound. The exponents were later changed by Liebig and Poggendorf to subscripts, e.g., CO_2.[135]

4.3.6 Laws of electrolysis

One of the major achievements that resulted from Faraday's discovery of electromagnetic induction (see Chapter 3) was his postulation of the laws of electrolysis, which later became known as Faraday's laws. Michael Faraday[136] (1791–1867) postulated the laws of electrolysis as follows: (1) The amount of decomposition in electrolysis is proportional to the quantity of electricity (i.e., the product of current strength and the time) that passes through the system (in 1832), and (2) the weights of substances deposited by the same current are in the same proportion as their chemical equivalent weights (i.e., electrochemical equivalents and chemical equivalents are one and the same) (in 1833).[137]

These laws rank as among the most accurate generalizations in science.[138] In honor of Faraday's contributions to electrochemistry, his name is now used

---

[134] ibid.

[135] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 108–109 (1976).

[136] A chemist-physicist, Michael Faraday made enormous contributions to physics and chemistry. His discoveries in electromagnetic induction, specific inductive capacity, and electrolysis are especially noteworthy. His papers on physics are collected in his *Experimental Researches in Electricity* (3 vols, 1839; 2nd ed., 1849); those on chemistry are collected in his *Researches in Chemistry and Physics* (1859) and include his discovery of carbon perchloride (C_2Cl_6) in 1821, of the liquefaction of gases (1823–45), of benzene in 1825, of the isomeric sulfonic acids of naphthalene in 1826, and the preparation of colloidal gold (including the description of the so-called "Tyndall phenomenon" in 1857 [cited in R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 193–194 (1989)].

[137] R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 194 (1989).

[138] J.M. Thomas, *Michael Faraday and the Royal Institution*, Taylor & Francis, New York, p. 48 (1991).

for the definition of an electrochemical equivalent unit—a faraday—the amount of electricity required to release one gram equivalent of any element from its solution. His discoveries laid the foundations of the electrochemical industry, of which electroplating, with its wide applications in manufacturing (ranging from the silver plating of utensils and jewelry to the strengthening of steel, to mention but a few), was perhaps the first and most widely visible and influential result. How ironic that, through one of Faraday's scientific discoveries, the alchemist's dream was at last being realized in our modern era: silver and gold could now, to all appearances, be made out of base metal.[139]

The electrochemical terms—electrolysis, electrolyte, anode, cathode, anion, and cation—were coined by William Whewell for Faraday.[140] Von Grotthus proposed a theory of the mechanism of current conduction in solutions in 1805.[141]

4.3.7 Law of combining volumes

While experimenting with gaseous reactions in the early 1800s, Joseph Louis Gay-Lussac (1778–1850) discovered that two volumes of hydrogen gas, when sparked with one measured volume of oxygen gas, produce exactly two volumes of water. Gay-Lussac also determined that one volume of hydrogen chloride gas reacts with one volume of ammonia gas to yield a white powder, with no residue of either gas remaining. The two gases had joined or reacted volume for volume. He tried reacting carbonic acid gas with ammonia and again found that exactly one part of odorless carbon dioxide reacts with exactly two volumes of ammonia. Another experiment carried out by his friend Bertholet had revealed that a measured volume of nitrogen gas always reacts with three times its volume of hydrogen to form exactly two measures of ammonia. From the results of the above experiments, Gay-Lussac made a remarkable observation from which he postulated that the volume of gases and the volumes of their gaseous products can always be expressed in ratios of small, whole numbers—no fractions or large numbers are involved. Thus, Gay-Lussac's discovery became known as the law of combining volumes.[142]

With the announcement of this law to the scientific world, it soon became apparent that it was not consistent with Dalton's atomic theory. The law raised the questions: Why does one volume of nitrogen unite or react with

---

[139] J. Hamilton, *A Life of Discovery: Michael Faraday, Giant of the Scientific Revolution*, Random House, New York, pp. 266–267 (2002).

[140] R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 194 (1989).

[141] ibid.; referenced also in *Annales de Chimie* **lviii**, 54 (1806) and *Ostwalds's Klassiker* No. 152.

[142] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 123–124 (1976).

one volume of oxygen to yield two volumes of nitric oxide? Why would two volumes of nitric acid (and not the one volume that Dalton's atomic theory would have predicted) result from the reaction between nitrogen and oxygen?[143]

4.3.8 Avogadro's hypothesis

Amedeo Avagadro (1776–1856) (Fig. 4.8) alone saw the error that confused the scientific world of his day. He was able to reconcile the atomic theory of Dalton with Gay-Lussac's law of combining volumes of gases. To him, the smallest part of simple elementary gases such as hydrogen, oxygen, chlorine, etc., consist not of one atom (as Dalton believed), but of two atoms chemically united. He called the pairs of hydrogen, oxygen, and chlorine atoms that are chemically united *molecules* of hydrogen, oxygen, and chlorine gases. Gaseous hydrogen, oxygen, chlorine, and nitrogen were thus not H, O, Cl, and N, but really H–H (or H_2), O–O (or O_2), Cl–Cl (or Cl_2), and N–N (or N_2), respectively. Stated another way, elementary gases such as hydrogen, oxygen, nitrogen, and chlorine are made up not of single atoms, but of groups of atoms called molecules.[144]

Avogadro arrived at his conception of pairs of atoms that combine to form molecules by postulating the hypothesis that equal volumes of all gases

Figure 4.8 Amadeo Avogadro (1776–1856), who postulated in 1811 that equal volumes of all gases at the same temperature and pressure contain the same number of molecules (printed with permission of the Deutsches Museum, Munich).

---

[143] ibid.

[144] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 124 (1976); J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 208–210 (1989).

under the same conditions of temperature and pressure contain identical numbers of molecules. He showed that his hypothesis reconciles Dalton's atomic theory with Gay-Lussac's law of combining volumes if one assumes that very simple relations also exist between volumes of gaseous substances and the numbers of simple or compound molecules that form them.[145]

Avogadro published this hypothesis in 1811 in a French scientific magazine, *Journal de Physique*, where it remained for half a century, while verbal controversies raged over dubious points about atomic weights, molecular weights, and the very essence of chemistry itself. Had the scientific world seized upon this conception and understood its full meaning, chemistry might have been advanced by decades. But the conception's formulator was little known beyond Turin and contented himself by teaching his students about his molecules and his hypothesis. In the meantime, some of the accepted theories of chemistry were found to be inconsistent with each other. Experiments yielded inaccurate results, given the accepted atomic weights of the elements. There was confusion on atomic weights, molecular weights, and equivalent weights of the elements. Theories and explanations were offered, often times conflicting with each other. There was no uniform language of chemistry.[146]

Amidst this clash of contending theories and explanations, 140 leading chemists of the era gathered in the German university town of Karlsruhe in September of 1860, determined to sort out the confusion. They arrived from all corners of the scientific world, then centered in Europe. From France came Antoine Béchamp (1816–1908), Jean-Baptist Dumas (1800–1884), and Charles Adolphe Wurtz (1817–1884). Thomas Anderson (1819–1874), Edward Frankland (1825–1899), and Henry Roscoe (1833–1915) came from England. Dmitri Mendeleev (1834–1907) represented Russia. Justus von Liebig (1803–1873), Friedrich Woehler (1800–1882), Alexander Mitscherlich (1794–1863), Hugo Erdmann (1862–1910), Emil Erlenmeyer (1825–1909), and Robert Bunsen (1811–1899) represented Germany. Stanislao Cannizaro (1826–1910) came from Italy, and Friedrich August Kekulé (1829–1896) came from Belgium.[147]

One of the issues these chemists had to settle was the distinction between the terms atom and molecule. Although the term *molecule* had been in use since the 17th century, it was used synonymously with the atoms of the ancients. Even Dalton himself had used atom and molecule interchangeably in his writings. It was in this climate of confusion that Canizzaro delivered to the illustrious audience gathered at Karlsruhe his famous speech on *molecules* and Avogadro's hypothesis. He had studied the work of his fellow countryman

---

[145] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 209 (1989).

[146] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 116 (1976).

[147] ibid., p. 117.

and became convinced of the validity of Avogadro's hypothesis. He pleaded for the admission of the molecule into its honored place as a worthy peer of the atom. At the close of the conference, he distributed to audience members reprints of a pamphlet he had written two years earlier called *Sunto di un Corso di Filsofia Chimica* (*Outline of a Course in the Philosophy of Chemistry*). The conference ended without any resolution to its main subject matter. However, as some of the attendees read the pamphlet and understood its content, Avogadro's hypothesis was at last accepted, albeit slowly, by the scientific world. Chemists soon had in their hands a tool with which to unlock many of the mysteries of atomic weights and formulas.[148]

Following the acceptance of Avogadro's hypothesis, scientists sought ways and developed ingenious schemes to verify it. First among these scientists was James Clerk Maxwell (1831–1879), who, in developing a kinetic theory of gases, made the first estimate of the number of particles in one cubic centimeter of a gas, and in so doing, mathematically verified the theory.[149]

Next was Jean Baptiste Perrin (1870–1942),[150] who prepared fine powder from gamboge gum, and placed a weighed amount of this gum in water, creating a suspension of fine particles. For three whole weeks he violently agitated this suspension until the heavier particles of gum settled to the bottom and were removed, leaving only the tiniest grains of the powder suspended in the water. Then, he placed a drop of the water containing the fine particles on a slide and examined the drop with a microscope that was equipped with a powerful beam of light such that each time a speck of gamboge particle passed across the field of the microscope, there was a momentary flash of light. In this way, he counted the number of specks in the gamboge and, knowing the weight of the gum he used, was able to determine the number of molecules in a cubic centimeter of gas as 3.15×10^{18}.

Another attempt at counting molecules in a given volume of gas was made by Robert Andrews Millikan (1868–1953).[151] Millikan used a different method from Perrin and yet arrived at the number 2.85×10^{18}, which is very close to the number obtained by Perrin. Another verification of Avogadro's hypothesis was provided by Irvin Langmuir (1881–1957),[152] who successfully dissociated a hydrogen molecule at a very high temperature and found that it actually contains two atoms.

The theory was found to be almost universally applicable. Later Jacobus Henricus van't Hoff (1852–1911) extended its application to the problem of solutions[153] (as we shall see later).

---

[148] ibid., pp. 117–121.

[149] ibid., pp. 125–126.

[150] ibid., p. 126.

[151] ibid., pp. 126–127.

[152] ibid., p. 127.

[153] ibid., p. 127.

The importance of Avogadro's hypothesis when it was finally accepted by the scientific world cannot be overemphasized. It allowed chemistry to advance by clearing away confusion, inconsistencies, conflicting opinions, and impossible conventions about atomic weight, molecular weight, equivalent weight, atoms, and molecules. It made it possible to determine accurate atomic weights of elements as well as the molecular weights and densities of a large number of gases. It showed that the elementary gases such as oxygen, hydrogen, chlorine, and nitrogen contain two atoms in each molecule, and that, therefore, their molecular weights are twice their atomic weights; ignorance of this fact in earlier times had led to many difficulties and inaccurate results.[154]

4.3.9 Law of mass action

The law of mass action was first proposed[155] with full generality by Cato Maximilian Guldberg (1836–1902) and Peter Waage (1833–1900) in 1864.[156] These Norwegian chemists postulated that the "force" responsible for the equilibrium reaction/transformation is proportional to the product of the masses involved in the equilibrium reaction:

$$A + B = A' + B', \tag{4.5}$$

which is given by kpq (for the forward reaction) and $k'p'q'$ for the reverse reaction, where k and k' are the activity coefficients for the forward and backward reaction, respectively; p and q are the active masses (defined as number of molecules per unit volume) for substances A and B, respectively, and p' and q' are active masses for substances A' and B', respectively. The value of the active masses can be determined experimentally, from which the ratio k/k' can be easily determined for any original condition of the four substances.[157]

---

[154] ibid., p. 127.

[155] Earlier in 1850, Ludwig Wilhelmy was the first person to investigate the rate of progress of a chemical reaction: the inversion of cane sugar by water in the presence of an acid. This reaction is particularly suited to experimental investigation because its progress can be followed by a polarimeter without disturbing the conditions of the reacting system. Wilhemy found that, in the presence of a large and practically constant mass of water, the amount of sugar changed in a small interval of time dt is proportional to the amount M actually present: $-dM/dt = kM$, where k is a constant, and, after integration, leads to the following equation: $-\log M = kt + constant$. Also in 1850, Williamson pointed out the dynamical character of a chemical equilibrium, viz., that it is the result of two opposite changes taking place at equal rates (cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 324 (1989).

[156] ibid., pp. 325–327.

[157] ibid., pp. 326–327.

In 1877 van't Hoff replaced the indefinite "force" by velocity of reaction. The velocity with which A and B react to form A' and B' he expressed as $v' = kpq$, and the velocity with which A' and B' react to form A and B he expressed as $v' = k' p' q'$. The actual velocity of the complete reaction is therefore given by $V = v - v' = kpq - k'\ p'\ q'$. At equilibrium, $V = 0$, and $p'q'/pq = k/k'$.[158]

Kinetics, the branch of physical chemistry dealing with rates of chemical reactions, therefore had its origins in the above-mentioned works of Guldberg and Waage as well as that of van't Hoff. Today, the techniques developed by these scientists are used to study rates of various reactions involved in lithographic patterning. A few examples of these lithographic reactions include deprotection kinetics of chemical amplification resists, development kinetics of resists, kinetics of photolysis of photoacid generators in chemical amplification resists, and color center formation in quartz glass of lithographic exposure tools, to mentioned but a few.

4.3.10 Thermochemical laws

The foundations of the branch of physical chemistry dealing with thermochemistry was laid by Lavoisier and Laplace, who through their collaborative efforts showed that the heat evolved in a reaction equals the heat absorbed in the reverse reaction. They investigated the specific and latent heats of a number of substances, as well as the amount of heat generated in combustion. It was Germain Henri Hess (1802–1850) who in 1840 enunciated the law that now bears his name—Hess' law:[159] The evolution of heat in a reaction is the same whether the process is accomplished in one step or in a number of stages. As an illustration, the heat from the formation of CO_2 is the sum of the heat from the formation of CO and the heat from the oxidation of CO to CO_2.[160]

When the mechanical theory of heat was enunciated, Hess' law was soon realized to be a consequence of the law of conservation of energy, which states that energy cannot be created nor destroyed, but can be transformed from one form to another. A consideration of the second law of thermodynamics, enunciated by Nicolas Léonard Sadi Carnot (1796–1832) in 1824, influenced Hermann von Helmholtz (1821–1894) in 1882 to estimate the work done by the chemical forces not as the heat evolved in the reaction, but as the largest quantity of work that can be gained when the reaction is carried out in a reversible manner, e.g., electrical work in a reversible cell. This maximum work is regarded as the diminution of the free or available energy of the

---

[158] ibid., p. 327.

[159] H. Hess, "Thermochemische Untersuchungen," *Ann. Physik*. **50**, 385–404 (1840); *Ostwald's Klassiker* No. 9 [cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 327–328 (1989)].

[160] ibid., p. 327.

system, while the heat evolved is regarded as a measure of the diminution of the total energy of the system.[161]

The application of the second law of thermodynamics to chemistry, first occurring notably in the study of dissociation phenomena in solids, influenced Henry John Horstman (1854–1929) in 1873 to point out that such changes are similar to physical changes of state, and that the thermodynamic equation derived by Benoît Paul Émile Clapeyron (1799–1864) and Rudolf Clausius (1822–1888) for changes of state are also applicable here, i.e., $dp/dT = Q/T(v' - v)$, where p is the dissociation pressure, Q is the heat of dissociation, T is the absolute temperature, and v' and v are the volumes of the system after and before dissociation, respectively. In the case in which a gas or vapor formed is supposed to behave as an ideal gas, and the volume of the solids is neglected in comparison with that of the gas, the equation becomes $d \log p/dT = Q/RT^2$, where R is the universal gas constant.[162]

Between 1884 and 1886, van't Hoff generalized this equation to make it applicable to all cases of chemical equilibrium between gases or substances in dilute solution. To this end, he replaced p with the equilibrium constant K in the above equation.[163] However, although the integration of this equation gives a means of finding Q from measurements of the effect of temperature on K, it does not solve the problem of calculating K from the heat of reaction Q since the equation contains an unknown constant of integration: $\log K = -Q/RT +$ constant. It was in 1906 that Walther Nernst (1864–1941) showed that by making another assumption it was possible to calculate the value of this constant, and the problem of finding K and hence the affinity from the heat of reaction was solved.[164]

4.3.11 The phase rule

The phase rule enunciated by Josial Willard Gibbs (1839–1903) in 1876 is the most useful guide for predicting equilibrium in systems that are not homogeneous. The rule deals with systems containing more than two different bodies, called *phases*, which are separated by boundaries, e.g., liquid and gas, or liquid and solid. Such a system can be built up from a certain minimum number of chemical substances called *components*. In order for a system to be in equilibrium, Gibbs asserts, certain conditions of temperature, pressure, and the concentrations in the various phases must be satisified. He called the number of these variables, which must be fixed before there is equilibrium,

---

[161] *Ostwald'd Klassiker* No. 124 [cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 328 (1989)].

[162] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 328–329 (1989).

[163] ibid., p. 329.

[164] ibid., p. 329.

the number of *degrees of freedom* of the system. If P is the number of phases, C is the number of components, and F is the number of degrees of freedom, then the Gibbs phase rule states that $P + F = C + 2$.[165]

4.3.12 Theory of solutions

The modern theory of solutions is primarily based on the research of François-Marie Raoult (1830–1901) and Wilhelm Pfeffer (1845–1920). Raoult showed that the depression of the freezing point or of the vapor pressure of a solvent by the dissolved substance could be used to calculate the molecular weight of the latter in the following manner. If C represents the depression due to P grams in 100 grams of solvent, then $CM/P = K$, where M is the molecular weight, and K a constant representing the value of C when $P = M$, i.e., the molecular depression of the freezing point, which is the same for a given solvent for all dissolved substances.[166]

Pfeffer's research was on osmotic pressure. It had been well established from early in the 18th century that when a solution is separated from pure water by a semipermeable membrane such as a bladder, the water permeates the membrane and dilutes the solution, setting up a pressure if the volume of the latter is prevented from increasing. This pressure was termed osmotic pressure. Pfeffer in 1877 was able, by depositing copper ferrocyanide on the walls of a porous pot, to prepare semipermeable membranes that were permeable to water, but were not permeable to dissolved sugar. Using these semipermeable membranes, he was able to measure the osmotic pressures of solutions.[167]

Interpreting Pfeffer's results, van't Hoff showed in 1886 that the osmotic pressure exerted by a dissolved substance equals the pressure it would exert as a gas confined in a space equal to the volume of the solution. He remarked: "It gradually appeared that there is a fundamental analogy, nay almost an identity, with gases, more especially in their physical aspect, if only in solutions we consider the so-called osmotic pressure instead of the ordinary gaseous pressure… we are not here dealing with a fanciful analogy, but with one which is fundamental."[168] Van't Hoff also postulated in 1887 a theory of solutions, which suggested that dissolved substances obey the same laws as gases.[169]

By applying the principles of thermodynamics, van't Hoff was able to deduce a quantitative connection between osmotic pressure and lowering of the freezing point or lowering of the vapor pressure.[170] However, like Raoult

---

[165] ibid., pp. 330–331.

[166] ibid., pp. 331–332.

[167] ibid., p. 332.

[168] ibid., p. 333.

[169] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 172 (1976).

[170] ibid.

before him, he was not able to explain why solutions of salts in water exhibit higher osmotic pressures than would be expected based on van't Hoff's theory. The full explanation of this phenomenon was offered in the theory of electrolytic dissociation[171] of Svante Arrhenius, which is based on the fact that solutions that show abnormally high osmotic pressures are conductors of electricity.[172]

4.3.13 Theory of electrolytic dissociation

One of the most puzzling problems in chemistry around the end of the 19th century was how to gain a rational understanding of what occurs in a solution when an electric current is passed through it. Even before this period, in fact nearly a century earlier, when the first experimenters in electrochemistry arranged the two poles of a galvanic battery so that an electric current might pass through it, this problem had puzzled and perplexed the leading chemists of the day. Particularly baffling aspects of this phenomenon included the fact that an electric current could not be transmitted through distilled water or solid salts such as sodium chloride (table salt). Yet when salt and water were mixed, their solution became a liquid through which electricity would pass with ease, the more dilute the solution. And as electric current passed through the aqueous solution, decomposition of the salt was observed, sometimes accompanied by evolution of gases. What was the basis for this strange behavior of solutions?[173]

Both Humphrey Davy and Theodor von Grothuss (1785–1822) had attempted to explain this phenomenon. Faraday, who discovered electromagnetic induction, had also investigated this subject and even created its terminology. Yet none of these investigators could deduce a correct explanation of electrolysis until Arrhenius came onto the scene.[174]

Svante Arrhenius (1859–1927) (Fig. 4.9) was the first scientist to offer an accurate explanation for this strange behavior. In his classic paper "On the dissociation of substances in aqueous solutions," published in the first volume of *Zeitschrift für Physikalische Chemie* in 1887, in which he postulated the theory of electrolytic dissociation, he wrote that the dissolution of a solid salt like sodium chloride in water is accompanied by a tremendous invisible change. The molecules of sodium chloride are instantly dissociated or ionized into particles, which years before, Faraday had labeled *ions* at the suggestion of William Whewell (1794–1866). Faraday thought that the ions were produced by the electric current, whereas Arrhenius said that they were

---

[171] The full statement of the theory of electrolytic dissociation was first published in *Zeitschrift für Physikalische Chemie* **1**, p. 631 (1887).

[172] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 332–333 (1989).

[173] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 164–165 (1976).

[174] ibid., p. 164.

Figure 4.9 Svante Arrhenius (1859–1927) proposed the theory of electrolytic dissociations (1887) and investigated the viscosity of solutions, the effect of temperature on reaction rates (1889), etc. (Printed with permission of the Deutsches Museum, Munich.)

already present in the solution even before the electric current was passed through the solution. Arrhenius represented the dissociation as follows:

$$\begin{array}{ccccc} \text{sodium chloride solution} & \rightarrow & \text{sodium ions} & + & \text{chloride ions} \\ \mathrm{NaCl} & \rightarrow & \mathrm{Na}^{+} & + & \mathrm{Cl}^{-} \end{array}$$

Arrhenius thought that only sodium and chloride ions were present in solution once the salt was fully dissociated, and that there were no longer any sodium or chloride atoms or molecules of sodium chloride present in the solution. The dissociated ions were absolutely free to move anywhere in solution. The dissociated ions differed from their parent atoms because they were electrically charged, while atoms were not. Dissociation in essence had changed the atoms into ions, and the charge of electricity had changed the ions to such an extent that they differed fundamentally from their parent atoms.[175] The revolutionary idea of this theory was that only ions take part in chemical reactions in solutions[176] and by extension in other phases.

With this theory, Arrhenius opened doors to a new chemistry: the chemistry of ions—strange, infinitesimal particles of matter bearing infinitely small electric charges that carry current through solutions, and then, as they touch electrodes, give up their electric charges and return once again to the atomic state. This dissociation phenomenon is observed to take place every time an inorganic acid, alkali, or salt is dissolved in water. Electrolysis,

---

[175] ibid., pp. 164–175.

[176] ibid., p. 169.

electroplating, and other applications of electrochemistry have their very foundations firmly rooted in this theory. Other scientific fields like physiology, medicine, and bacteriology, too, have found the theory to be very useful.[177] Today, in lithography, the theory of electrolytic dissociation applies in the analysis of the development step of exposed resists, in lithographic mask degradation due to corrosion, in electromigration of chromium ions, etc.

Between 1853 and 1859, Johann Wilhelm Hittorf (1824–1914) showed by numerous experiments (to which Arrhenius later drew attention) that the current in electrolysis is carried unequally by the two ions that move at different speeds; he called the fraction carried by the anion the transport number n.[178] Friedrich G.W. Friedrich Kohlrausch (1840–1910) had discovered that all ions of the same element, regardless of the compound from which they were formed, behave in exactly the same way.[179] Extending Hittorf's findings, Kohlrausch showed in 1874 that the equivalent conductivity (i.e., the specific conductivity κ divided by the concentration of the solution in equivalents per cubic centimeter, $\kappa/c = \Lambda$) of a very dilute solution of a salt is the sum of two terms, one depending only on the cation and the other only on the anion: $\Lambda = l_c + l_a$. He called l_c and l_a the mobilities of the ions. If n is Hittorf's transport number, then $l_a = n\Lambda$.[180]

When Raoult's results were examined in light of Hittort's and Kohlrausch's results, it became apparent that in solution salt dissociates into its two ions in a process called ionization. The electric current serves to drive the positively and negatively charged ions toward the electrodes of opposite polarity. Given that the number of particles in a given salt solution is increased by ionization, the abnormally high osmotic pressures and the related abnormal depressions of the freezing point were thus explained.[181]

Arrhenius showed that the degree of ionization calculated from the electrical conductivity Λ/ Λ_∞, where Λ_∞ is the equivalent conductivity at infinite dilution (when there is complete ionization), was very approximately the same as that calculated from the deviations from Roults' law in the case of the freezing point measurement. Obtaining nearly identical results with two independent methods substantiated Arrhenius' theory. Today, we use a slightly modified version of the original formulation of Arrhenius' theory that primarily involves strong acids, bases, and salts, which are now viewed as being practically completely ionized in fairly dilute solutions. The deviations from the results for complete ionization are ascribed to interaction between the ions of

---

[177] ibid., pp. 164–175.

[178] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 324 (1989).

[179] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 171 (1976).

[180] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 324 (1989).

[181] ibid., p. 335.

the opposite charge, which causes a reduction in the mobilities of the ions and hence a diminution of Λ.[182]

Important extensions of Arrhenius' theory were made by Walther Nernst (1864–1941) (see Fig. 4.10), who in 1889 put forward the theory of galvanic cells, which states that the production of electromotive force in galvanic cells can be explained in terms of a "solution pressure" of the metal electrodes that tends to throw off charged ions into the solution, this tendency being balanced by the osmotic pressure of the dissolved ions. Nernst also introduced in 1889 the theory of solubility product, which explains precipitation reactions. In addition, he investigated diffusion in solutions, including liquid contact potentials (1888–1889), and he postulated the so-called heat theorem (third law of thermodynamics) (1906), which states that the entropy change is zero at absolute zero.[183]

Another important extension of Arrhenius theory is attributed to Alfred Werner (1866–1919), who developed a theory of complex ion formation, which shows that metals combine chemically in two ways. The first method of combination is through formation of ionic or primary bonding. The second is

Figure 4.10 Walther Nernst (1864–1941) showed in 1889 that the production of electromotive force in galvanic cells can be explained in terms of a "solution pressure" of the metal electrodes. He made substantial contributions to chemical thermodynamics. (Printed with permission of the Deutsches Museum, Munich.)

---

[182] P. Debye and E. Huckel, "Zur Theorie der Elektrolyte. 1. Gefrierpunktserniedrigung und Verwandte Erschinungen" [The theory of electrolytes. 1. Lowering of freezing point and related phenomena], *Physikalische Zeitschrift* **24**, pp. 185–206 (1923).
[183] ibid., pp. 336–338.

through formation of nonionic or secondary bonding. For example, in the compound potassium ferrocyanide, $K_4Fe(CN)_6$, the primary bond is between K^{+1} and $Fe(CN)_6^{-4}$. The secondary bond is between Fe^{+2} and $(CN)^{-6}$. This type of chemical bonding has specific orientations about the central metal ion, and these orientations influence the behavior of this complex ion.[184]

4.3.14 Le Chatelier's law of reactions

Henry Louis Le Chatelier (1850–1936) put forward the law of reaction, which governs the effects of pressure, temperature, concentration, and catalyst on reaction equilibrium. This law states that "if a change of conditions (stress) is applied to a system at equilibrium, the system will respond in the way that best reduces the stress in reaching a new state of equilibrium."[185]

4.3.15 The periodic table and the periodic law

From the ancient time to about 1869, 63 chemical elements were known to science. The stories behind their discoveries are as fascinating as the properties of the elements. Gold, silver, copper, iron, mercury, lead, tin, sulfur, and carbon were all known to the ancients, who fashioned instruments out of them. The alchemists discovered new elements in their futile search for the seed of gold and the elixir of life. The German physician Basil Valentine (Basilius Valentinus) (1394–1450) fancifully described antimony in 1492, the year Christopher Columbus discovered America. In 1530, Georgius Agricola (1494–1555), another German, described bismuth in his *De Re Metallica*, a book on mining. Paracelsus, born Theophrastus von Hohenheim (1493–1591), was the first person to describe zinc to the Western World. Hennig Brandt (1630–1710), a German alchemist, discovered phosphorus in 1669 in urine that he had boiled down to a tar-like substance. Arsenic and cobalt were later added to the list of elements known to medieval scientists.[186]

Before the end of the 18th century, 14 more elements were added to the list of known elements. In 1735, Don Antonio de Ulloa (1716–1795), a Spanish naval officer, while on an astronomical mission to Choco, Columbia, found a heavy nugget and almost discarded it as worthless before the valuable properties of the metal platinum were recognized. Then came the discovery of lustrous nickel, inflammable hydrogen, inactive nitrogen, life-sustaining oxygen, death-causing chlorine, manganese, which is used among other things for burglar-proof safes, tungsten, which is used in incandescent lamps, chromium, which is used in stainless steel, molybdenum and titanium, both useful in steel

---

[184] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 173 (1976).

[185] K.W. Whitten and K.D. Gailey, *General Chemistry with Qualitative Analysis*, Saunders College Publishing, Philadelphia, p. 427 (1981).

[186] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 154 (1976).

alloys, and tellurium, zirconium, and uranium, which is the heaviest of all the elements. And around beginning of the 19th century, the Englishman Hatchet discovered niobium (also called columbium) in a black mineral that had found its way from the Connecticut Valley to the British Museum.[187]

By 1869, all of the 63 known chemical elements had been isolated and described in the chemical journals in England, Germany, France, and Sweden. All of these elements had different Daltonian atoms, as well as different atomic weights, which ranged from 1 (hydrogen) to 238 (uranium). Their chemical and physical properties differed to greatly varying degrees. Some of them were gases, including oxygen, hydrogen, chlorine, and nitrogen. Others were liquid at normal temperature and pressure conditions, and these included mercury and bromine. The rest of the elements were solids. Some of them were very hard metals such as platinum and iridium, and some were soft metals like sodium and potassium. Lithium was a metal so light that it floated on water. Osmium, on the other hand, was 22.5 times heavier than water. And there was mercury, a metal, which was not a solid at all, but a liquid. These elements also had different colors. Copper was red, gold was yellow, iodine was steel gray, phosphorus was white, and bromine was red. Some of the metals, for example, nickel and chromium, could be polished to a very high degree, while others, like lead and aluminum, were dull. Gold, on exposure to air, never tarnished, but iron rusted readily; iodine was sublimed and changed directly into vapor from the solid phase. Some elements reacted with one atom of oxygen, and others with two, three, or even four atoms. A few, like potassium and fluorine, were reactive when contacted with other elements and substances; others were highly nonreactive toward other elements and substances.[188]

It was in this context of massively varying and dissimilar properties of the then known elements that scientists of the era sought to find order, connections, and relationships between the elements, with a view to proposing a law that governs their properties. The result of this effort was the periodic table and the periodic law governing the position of the elements in the table.

The periodic law, which first disclosed a chemical and material relationship among chemical elements, represents one of the highest achievements of the revolution in chemistry that Lavoisier started. Its roots stretch all the way back to Johann Wolfgang Döbereiner (1780–1849), who in 1817 and 1829 reported that in certain groups of three elements, e.g., calcium, strontium, and barium, the atomic weight of the middle element was approximately the mean of those of the first and the third. These groups of elements were called Döbereiner's triads.[189]

---

[187] ibid., pp. 154–155.

[188] ibid., p. 155.

[189] For the Dobreiner triads, see "Auszug eines Briefes vom Hofrath Wurzer, Prof. der Chemic zu Marburg," *Ann. Physik* **lvi**, 331 (1817) (Ca, Sr. Ba); "An attempt to group elementary substances according to their analogies," *Ann. Phys.* **xv**, 301 (1829) (other triads); *Ostwald*

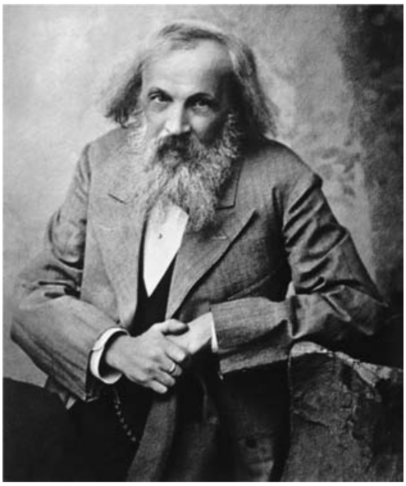


Figure 4.11 Dmitri Ivanovich Mendeleev (1834–1907) independently formulated the periodic law of the elements and constructed the periodic table (printed with permission of the Deutsches Museum, Munich).

The periodic law was proposed almost simultaneously and quite independently by Dmitri Ivanovich Mendeleev (1834–1907)[190] (see Fig. 4.11) in Russia and Julius Lothar Meyer (1830–1895)[191] (see Fig. 4.12) in Germany. Mendeleev published his periodic law in 1869, which states that "the properties of the elements are in periodic dependence upon their atomic weights." Lothar Meyer had previously reported (also in 1869) that certain triads of elements, such as lithium, sodium, and potassium, as well as magnesium, calcium, and strontium, exhibit similar chemical properties. In 1870, he published his well-known atomic volume curve, in which the periodic dependence of a quantitative property was shown as a function of the atomic weight, revealing that, as the atomic weight steadily increases, the property alternately rises and falls over definite periods of the elements.[192] Following leads of the above type, Mendeleev gave each known element a position in a single grand ensemble now called the periodic table and identified gaps that he predicted would be filled by yet undiscovered elements that he called *eka-aluminum* (with properties similar to aluminum), *eka-boron* (with properties similar to boron) and *eka-silicon* (with properties similar to silicon).[193]

Klassiker No. 66; J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 343 (1989).

[190] D.I. Mendeleyev, "Essai d'une system des elements," *J. Russ. Phys. Chem. Soc.* **i**, 60 (1869); see also J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 348 (1989).

[191] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 348–349 (1989).

[192] ibid.

[193] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 158 (1976).

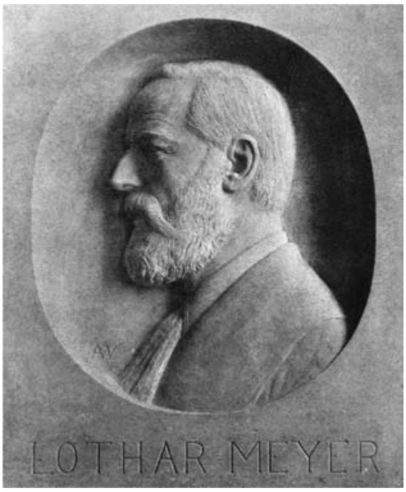


Figure 4.12 Julius Lothar Meyer (1830–1895), who pointed out that triads of elements, such as lithium, sodium, and potassium, and magnesium, calcium, and strontium, had similar chemical properties. He independently formulated the periodic law of the elements at about the same time as Mendeleev. (Printed with permission of the Deutsches Museum, Munich.)

In constructing his table, Mendeleev arranged the elements into seven groups, starting with lithium (atomic weight 7), and followed by beryllium (atomic weight 9), boron (atomic weight 11), carbon (atomic weight 12), nitrogen (atomic weight 14), oxygen (atomic weight 16), and fluorine (atomic weight 19). Because the next element in increasing order of atomic weight (sodium, atomic weight 23) resembled lithium very closely in both physical and chemical properties, he placed it below lithium in his table. After placing five more elements, he came to chlorine, which had properties very similar to fluorine, so he placed it under fluorine. With this method, he arranged all of the known 63 elements. When he completed his list, he noticed a most remarkable order. All of the elements fit into their places beautifully. The very active metals, lithium, sodium, potassium, rubidium, and cesium, all fell into one group (Group I). The extremely active nonmetals, fluorine, chlorine, bromine and iodine, all fell into the seventh group (Group VII).[194]

When he examined his table, he discovered that the properties of the elements were periodic functions of their atomic weights; in other words, their properties repeated themselves periodically after each seventh element. Another astonishing fact that he observed was that all of the Group I elements reacted with oxygen, two atoms to one. All of the Group II atoms reacted with oxygen, atom for atom. The Group III elements reacted with oxygen, two atoms to three. He observed similar trends in the remaining

---

[194] ibid., pp. 156–157.

groups of elements. What a remarkable observation! Knowing the properties of one element of a certain group meant knowing, in a general sense, the properties of all of the elements in that group. In 1869, he presented his paper *On the Relation of the Properties to the Atomic Weights of the Elements* to the Russian Chemical Society, in which he postulates that the properties of the elements show periodic dependence on their atomic weights.[195]

In 1875, the eka-aluminum predicted by Mendeleev was discovered by the Frenchman Lecoq de Boisbaudran (1838–1912) in a zinc ore mined in the Pyrenées. When the ore was analyzed, it was found to have properties similar to aluminum: it yielded two new lines on the spectroscope, was easily fusible, could form alums, and its chloride was volatile. Lecoq named it *gallium* after the ancient name of his native country.[196]

A short while later in 1877, in Germany, Clemens Winkler (1838–1904) discovered an element that matched the properties of Mendeleev's eka-silicon. From the silver ore argyodite he isolated eka-silicon and found it to have an atomic weight of 72.3 and a density of 5.5. When he heated it in air, he found its oxide to be exactly as heavy as Mendeleev had predicted. He synthesized its ethide and found it to boil at exactly the temperature Mendeleev had predicted. Its spectral lines were also in agreement with Mendeleev's prediction. Winkler named the new element *germanium* in honor of his fatherland.[197] Two years later in 1879, out of Scandinavia came the report that Lars Fredrik Nilson (1840–1899) had isolated eka-boron from the ore of euxenite. He named it scandium after his native Scandinavia.[198]

Influenced by the x-ray spectrometric method developed by William Henry Bragg (1862–1942) and his son William Lawrence Bragg (1890–1971) (discussed in Chapter 3), Henry Moseley (1887–1915) set out to compare the photographs of the x-ray spectra of different elements (38 in number—from aluminum to gold), from which he sought to determine the nature of the electric charge of their nuclei. The motivation for this experiment was the idea, proposed by Rutherford around 1911, that the main mass of the atom is concentrated in a tiny nucleus of positive electricity surrounded by enough electrons to make the atom electrically neutral.

To determine the nuclear charge, Moseley found that when metallic plates made of individual elements are bombarded with cathode rays (electrons), they emit x-rays of characteristic frequency, dependent on their atomic number. In other words, each element has its own characteristic x-rays, which

---

[195] ibid.

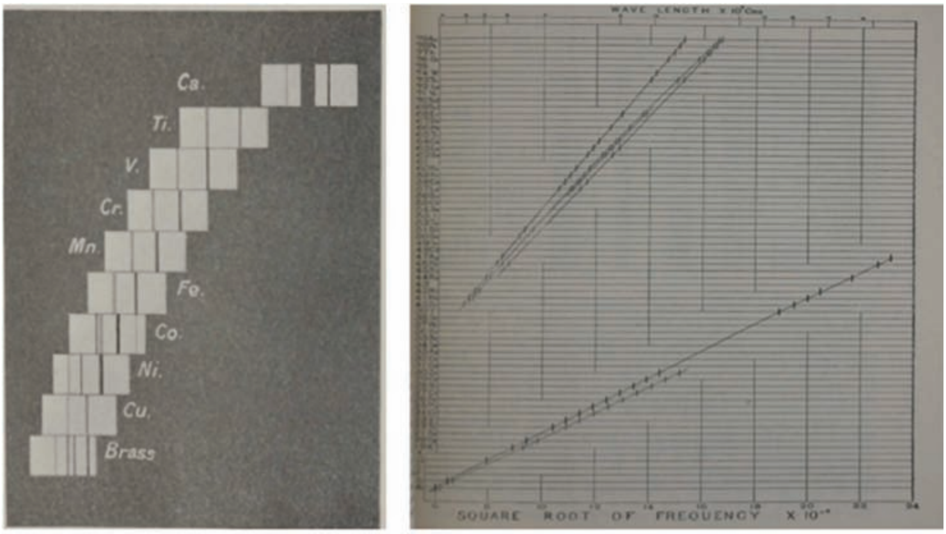
[196] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 151–152 (1976).

[197] ibid.; F. Seitz and N.G. Einspruch, *Electronic Genie: The Tangled History of Silicon*, University of Illinois Press, Urbana, p. 3 (1999); J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 346–349 (1989).

[198] ibid.

are almost like a fingerprint of the element.[199] He also observed a definite relationship: The heavier the element, the shorter and more penetrating are the x-rays it produces. He arranged the figures of his x-ray spectra for the elements on graph paper, obtaining a staircase formation, now called Moseley's "staircase," in which the line pattern is shifted to the shorter wavelength as the charge of the nucleus increases, as illustrated in Fig. 4.13(a).[200]

When he plotted the numbers of the elements (representing their position in Mendeleev's table) against the square root of the vibrational *frequencies* of a characteristic line of their x-rays, he made a remarkable discovery: The elements fall on a straight line in the exact order of their atomic weights, indicating that in the atom there is a fundamental quantity that increases by one unit from each element to the next in the periodic table [see Fig. 4.13(b)].[201] This quantity, Moseley surmised, can only be the charge of the central positive nucleus. This was the essence of his fundamental discovery, which he postulated in the law of atomic numbers.[202] With this



(a) (b)

Figure 4.13 (a) Moseley's photographic recording of Kα and Kβ x-ray emission lines for a range of elements showing the "ladder staircase" of the elements. (b) Moseley's plot of the atomic numbers of elements versus the square root of the vibrational frequencies of a characteristic line of their x-rays.

---

[199] H.G.J. Moseley, "High frequency spectra of the elements, Part 1," *Philos. Mag*. **6**(26), 102 (1913).

[200] ibid.

[201] H.G.J. Moseley, "High frequency spectra of the elements," Part II, *Philos. Mag*. **6**(27), 703 (1914).

[202] ibid.

law, he prepared a new periodic table of the elements that was far more fundamental than that of his predecessor Mendeleev. With the new periodic table, he bequeathed[203] to our world an infallible road map of all of the elements in the universe—a chart based not on *atomic weights* but on *atomic numbers.*[204]

It is thus now established that the nucleus of the atom is made up of protons and neutrons, a neutron being a particle of approximately the same mass as the proton but without the charge. The mass of the atom has now been determined to approximately equal the sum of the masses of the protons and neutrons. The nuclear charge equals the atomic number, given the position of an element in the periodic table. An element is now defined by this atomic number, and different isotopes of an element have the same atomic number.[205]

Table 4.1 shows the modern periodic table, in which the elements are arranged according to their atomic numbers in columns (groups) and rows (periods) and presented to emphasize their periodic properties,[206] from hydrogen 1 to lawrencium 103.[207] While Dmitri Mendeleev is usually credited with the form of the modern periodic table, many other excellent chemists made profound contributions to its development, including Johann Döbereiner, John Newlands, Lothar Meyer, Sir William Ramsay, (of course) Moseley, and others.[208]

Twenty-five years after the publication of Mendeleev's table, two Englishmen,William Ramsay (1852–1916) and Morris Travers (1872–1961), following a clue that Henry Cavendish had uncovered almost 150 years

---

[203] My use of the word *bequeath* is not accidental. Three years after publishing his monumental discovery, and at only 26 years of age, Henry Moseley was killed at the battle of Gallipoli, Turkey, during World War I. What a great loss to science and to humanity! In expressing the tragic death of this young man, Millikan wrote: "In a research which is destined to rank as one of the dozen most brilliant in conception, skillful in execution, and illuminating in results in the history of science, a young man twenty-six years old threw open the windows through which we can glimpse the subatomic world with a definiteness and certainty never dreamed of before. Had the European War had no other result than the snuffing out of this young life, that alone would make it one of the most hideous and most irreparable crimes in history." [Quotation cited in B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 219–220 (1976).]

[204] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 222–226 (1976).

[205] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 360–361 (1989).

[206] The periodic law unearthed a number of previously unsuspected analogies among the elements and, as a result, stimulated the study of inorganic chemistry, which had been rather neglected in the second half of the 19th century owing to the popularity of organic chemistry at that time.

[207] A. Stwertka, *A Guide to the Elements*, Oxford University Press, Oxford, p. 6 (2002).

[208] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 342–349 (1989).

Table 4.1 The modern periodic table.[209]

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Period																		
1	${}_{1}$H																	^{2}He
2	${}_{3}$Li	${}_{4}$Be											^{5}B	^{6}C	^{7}N	^{8}O	^{9}F	^{10}Ne
3	${}_{11}$Na	^{12}Mg											^{13}Al	^{14}Si	^{15}P	^{16}S	^{17}Cl	^{18}Ar
4	${}_{19}$K	^{20}Ca	^{21}Sc	^{22}Ti	^{23}V	^{24}Cr	^{25}Mn	^{26}Fe	^{27}Co	^{28}Ni	^{29}Cu	^{30}Zn	^{31}Ga	^{32}Ge	^{33}As	^{34}Se	^{35}Br	^{36}Kr
5	${}_{37}$Rb	^{38}Sr	^{39}Y	^{41}Zr	^{41}Nb	^{42}Mo	^{43}Tc	^{44}Ru	^{45}Rh	^{46}Pd	^{47}Ag	^{48}Cd	^{49}In	^{50}Sn	^{51}Sb	^{52}Te	^{53}I	^{54}Xe
6	${}_{55}$Cs	^{56}Ba	${}^{*}{}_{57}$La	^{72}Hf	^{73}Ta	^{74}W	^{75}Re	760s	^{77}Ir	^{78}Pt	^{79}Hg	^{80}Hg	^{81}Ti	^{82}Pb	^{83}Bi	^{84}Po	^{85}At	^{86}Rn
7	${}_{87}$Fr	^{88}Ra	^{**89}Ac	^{104}Rf	^{105}Db	^{106}Sg	${}_{107}$Bh	^{108}Hs	^{109}Mt	^{110}UUn	^{111}Uuu	^{112}Uub		^{114}Uuq				
*Lanthanides			^{58}Ce	^{59}Pr	^{60}Nd	^{61}Pm	^{62}Sm	^{63}Eu	^{64}Gd	^{65}Tb	^{66}Dy	^{67}Ho	^{68}Er	^{69}Tm	^{70}Tb	^{71}Lu		
*Actinides			^{90}Th	^{91}Pa	^{92}U	^{93}Np	^{94}Pu	^{95}Am	^{96}Cm	^{97}Bk	^{98}Cf	^{99}Es	^{100}Fm	^{101}Md	^{102}No	^{103}Lr		

---

[209] ibid., p. 6.

earlier, discovered a new group of elements, of which Mendeleev was not even aware. These elements constituted a new group—Group Zero (0)—as they were later named. The members of Group 0 comprise the noble gases, which show limited or no reactivity to other elements under normal conditions.[210] The first member of this group was for the first time observed in the sun's photosphere during a solar eclipse in August 1868 by Norman Lockyer (1836–1920), when he looked through a new instrument—the spectroscope invented by Bunsen and Kirchhoff—and saw the distinctive, bright, orange–yellow spectral line associated with the element. Since it was present in the photosphere of the sun, he called it helium. Twenty years later, William Hillebrand (1853–1925) of the United States Geological Survey observed this gas being expelled from the rare mineral cleveite.[211] Ramsay and Travers isolated the other five members of the group: neon, argon, krypton, xenon, and radon.[212] Xenon tetrafluoride and several other noble compounds were successfully prepared in 1962; for this reason, these gases are no longer regarded as being nonreactive.[213]

The modern statement of the periodic law is that the chemical and physical properties of the elements vary in a periodic manner with their atomic numbers.[214] The periodic table is now an indispensable reference tool for scientists because of its ability to predict the chemical behavior of elements. The high frequency of its use by scientists highlights the central role it plays in making sense out of what otherwise might be a chaotic jumble of facts about the elements and their molecular combinations,[215] bearing in mind the fact that, even with the variety and complexity of our world, everything in it—air, water, rocks, living tissue, and the almost infinite number of other objects and materials around us—is actually made up of only a limited number of chemical elements. By the time Mendeleev died in 1907, 86 elements were listed in the periodic table, a four-fold increase since the days of Lavoisier.[216] We know today that only 92 of such elements exist naturally on the earth, although several new artificial elements have been created under laboratory conditions.

What is then responsible for the periodic behavior of the elements? And why do elements within a particular group have similar chemical behavior?

---

[210] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 159–150 (1976).

[211] ibid., pp. 150–151 (1976).

[212] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, pp. 159–160 (1976).

[213] ibid., p. 159.

[214] A. Stwertka, *A Guide to the Elements*, Oxford University Press, Oxford, pp. 6–15 (2002).

[215] ibid., p. 15.

[216] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 160 (1976).

The reason stems from the fact that atoms are attracted to each other by electric forces. The atomic number, i.e., the number of positively charged protons in the nucleus, determines the number of negatively charged electrons contained in the atoms of a particular element, and it is the outer-shell electrons, also called valence electrons, that determine the reactivity, or how strongly and in what way the atoms will bind with other atoms. The chemical behavior of an element is thus ultimately determined by the way in which the electrons orbiting its nucleus are structured.[217]

Elements in the same group of the periodic table possess the same number of electrons in their outer shells and are therefore said to have the same valence electronic configuration and, consequently, similar chemical and physical properties. As electrons fill into the inner shells of an atom, the atom's outer shell takes on a specific valence configuration that is determined by the rules that govern how many electrons can occupy a particular shell, as described in Section 3.4.11 on quantum mechanics. It is this very regularity in the number of outer-shell electrons that explains the periodic behavior shown by the elements as the atomic number increases. Similarly, properties such as atomic size are determined by the number of shells an atom contains. For example, the radius of the atoms of the elements within a particular group in the periodic table increases from the top of the group to the bottom.[218]

Today, in lithography, knowledge of the periodic table is employed in the rational design of lithographic chemicals and materials. Specifically, H, Si, C, O, N, F, and S are the most commonly used elements in lithography. Si is primarily used in the fabrication of wafer substrates and organo-silicon resists, as well as in the synthesis of optical glass/quartz elements of the lithographic exposure tool. C, O, F, and S are widely used in the synthesis of resists and resist sensitizers. Lithographic exposure light sources operating in the deep-UV region are based on KrF and ArF exciplex lasers, lasing at 248 nm and 193 nm, respectively. Extreme-UV lithography relies on soft x-ray radiations of 13.5-nm wavelength generated from plasma derived from the interaction of CO_2-laser-produced and elemental Sn droplets.

4.3.16 Electronic theory of valency

Influenced by Bohr's model of the atom and developments in quantum physics, in 1916 Gilbert N. Lewis (1875–1946) proposed a simple model of the atom in which the number of electrons in the outer layer is eight in the atoms of inert gases, and tends to become eight in other atoms either (1) by gain or loss of outer electrons, forming negative or positive ions, or (2) by sharing

---

[217] A. Stwertka, *A Guide to the Elements*, Oxford University Press, Oxford, p. 13 (2002).

[218] ibid., pp. 14–15.

electrons in pairs with other atoms, forming ordinary valency bonds.[219] This theory was extended to heavier atoms by I. Langmuir in 1919.[220]

Two types of bonds[221]—ionic and nonionic (covalent) bonds—were recognized early on in the formulation of the electronic theory of valency. Lewis made clear distinctions between ionic and covalent bonds. Ionic bonds are formed by transfer of electrons and production of separate charged ions, as explained by Albrecht Kossel (1853–1927) in 1916. Covalent bonds are formed, according to Lewis, by sharing of electrons in pairs, a single bond consisting of one shared pair, a double bond of two, and a triple bond of three.

A type of bond produced when both electrons of a pair come from the same atom was first described by Langmuir (1919).[222] G.A. Perkins in 1921 drew attention to this bond,[223] as did Martin Lowry (1874–1936), who in 1923 called it a "mixed double bond" or "dative bond,"[224] and Nevil Sidgwick, who in 1923 called it a "coordinated link," since such bonds are present in the so-called coordination compounds.[225] In 1931, Linus Pauling (1901–1994)[226] proposed that different types of bonds in molecules can become equalized via rapid shifting of electrons, through a process called "resonance." The resulting

---

[219] G.N. Lewis, "The atom and the molecule," *J. Amer. Chem. Soc.* **xxxviii**, 762 (1916); W. Kossel, "Über Molekülbildung als Frage des Atombaues," *Ann. Physik* **xlix**, 229–362 (1916).

[220] I. Langmuir, "The arrangement of electrons in atoms and molecules," *J. Amer. Chem. Soc.* **xli**, 868–934 (1919); **xlii**, 274 (1920); J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 360–362 (1989).

[221] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 361–362 (1989).

[222] I. Langmuir, "The arrangement of electrons in atoms and molecules," *J. Amer. Chem. Soc.* **xli**, 868–934 (1919); **xlii**, 274 (1920); J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 360–362 (1989).

[223] G.A. Perkins, "Expression of octet theory of valence in structural formulas," *Philippine J. Sci.* **19**(1), 1–22 (1921); also cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 366–367 (1989).

[224] M. Lowry, "The electronic theory of valency. Part I: Intramolecular ionization," *Trans. Faraday Soc.* **xviii**, 285 (1923); "Studies of electrovalency. Part I: The polarity of double bonds," *J. Chem. Soc.* **cxxiii**, 822 (1923); "The uniqueness of hydrogen," *J. Soc. Chem. Ind.* **xlii**, 43–47 (1923) [cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 368 (1989)].

[225] N.V. Sidgwick, "Coordination compounds and Bohr atom," *J. Chem. Soc.* **cxiii**, 725 (1923); *The Electronic Theory of Valency*, Oxford University Press (1927); W.G. Palmer, *Valency, Classical and Modern*, Cambridge (1944) [cited in J.R. Partington, A *Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 368 (1989)].

[226] L. Pauling, "The nature of the chemical bond. Part II: The one-electron bond and the three-electron bond," *J. Amer. Chem. Soc.* **53**(9), 3225–3237 (1931); "The nature of the chemical bond. Part III: The transition from one extreme bond type to another," **54**(3), 988–1003 (1932); *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, New York (1939) [also cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 368–369 (1989)]. This phenomenon was recognized earlier in some theoretical discussions by other researchers, but its application in chemistry is largely due to Pauling.

"resonance hybrid" contains contributions from the different possible electronic configurations.[227]

Coordination compounds, also called "molecular compounds," formed by the addition of molecules in which the valencies of the atoms are apparently satisfied, were first investigated in 1893 by Werner, who showed that the number of atoms or groups associated with a central atom is often 4 or 6, which are the "coordination numbers" (other coordination numbers, up to a maximum of 8, occur, but less often).[228] Various types of isomerism, including optical isomerism, are fairly well explained by Werner's theory. Sidgwick showed in 1923 that groups that coordinate with metals do so by contributing electron pairs to the valency shell of the metal; hence, the name "coordinate link" was given to this type of bond.[229]

4.3.17 Theory of the periodic table

The structure of the periodic table was explained by Bohr and others based on four quantum numbers and Pauli's exclusion principle, described below. These four quantum numbers represent the solution of the Schrödinger equation (described in Chapter 3) for hydrogen atoms, and they describe the various states available to hydrogen's single electron. The electrons are described as point charges revolving around the central positive nucleus of the atom in circular orbits, just like the planets revolve around the sun. The energy state of the electron is described by a principal quantum number n, with each orbit having a particular value of n, which may take any integral value: 1, 2, 3, 4, ... By analogy with the planetary orbits (which obey the inverse square law of attraction), the orbits would be expected to be ellipses, and for a given value of n, ellipses of varying eccentricity are therefore possible.[230]

Arnold Sommerfeld (1868–1951) introduced the subsidiary (or azimuthal) quantum number l in 1925 to designate the geometric shape of the region in space that an electron occupies. Within each energy level (defined by the value of n), l may take integral values from 0 up to and including $(n - 1)$, i.e., $l = 0, 1, 2, \ldots, (n - 1)$. The maximum value of l is thus $(n - 1)$. The subsidiary

---

[227] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 367–369 (1989).

[228] A. Werner, "New ideas in inorganic chemistry," *Z. Anorg. Chem.* **iii**(1), 267–330 (1893); **xl**, 15 (1905); also cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 369–373 (1989).

[229] N.V. Sidgwick, "Coordination compounds and the Bohr atom," *J. Chem. Soc.* **cxiii**, 725 (1923); *The Electronic Theory of Valency*, Oxford University Press (1927); W.G. Palmer, *Valency, Classical and Modern*, Cambridge University Press (1944); also cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 373 (1989).

[230] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 362–365 (1989); K.W. Whitten and K.D. Gailey, *General Chemistry with Qualitative Analysis*, Saunders College Publishing, Philadelphia, pp. 69–101 (1981); P.W. Atkins, *Physical Chemistry*, 3rd ed., W.H. Freeman and Company, New York, pp. 340–370 (1986).

quantum number designates a sublevel, or a specified kind of atomic orbital, that an electron may occupy. With each value of l, a letter has become associated, each corresponding to a different kind of atomic orbital: $l = 0$ (s), 1(p), 2(d), 3(f), . . . , (n – 1). The s, p, d, and f designations arise from the characteristics of spectral emission lines produced by electrons occupying the orbitals: s (sharp), p (principal), d (diffuse), and f (fine structure).[231]

If the plane of the ellipse is defined by a weak magnetic field, it can assume a number of discrete positions that are defined by the magnetic quantum number m, which designates the spatial orientation of an atomic orbital. Within each sublevel, m takes all integral values from $-l$ through 0 up to and including $+l$, i.e., $m = (-l) \ldots, 0, \ldots, (+l)$. The maximum value of m depends on the value of l. For example, when $l = 1$, which designates the p sublevel, there are three permissible values of m, namely, –1, 0, and +1. There are thus three distinct regions of space, called atomic orbitals, associated with a p sublevel. Atomic orbitals are defined as regions of space where there is a high probability of finding electrons.[232]

A fourth quantum number s is used to define the spin of the electron about its own axis and the orientation of the magnetic field produced by the motion of an electron. It has two possible values, $\pm\frac{1}{2}$, corresponding to two opposite directions of spin.[233]

The values of n, l, m, and s describe a particular atomic orbital, and each atomic orbital can accommodate no more than two electrons, one with $s = -\frac{1}{2}$ and a second with $s = +\frac{1}{2}$. Pauli's exclusion principle (1925) states that no two electrons in an atom may have identical sets of the four quantum numbers n, m, l, and s. No exceptions to this principle have been uncovered to date.[234]

By taking the values 1, 2, 3, and 4 for n and making use of the above relations, it can be shown that the maximum numbers of electrons with these values of n are 2, 8, 18, and 32, and these follow a general formula, $2n^2$. These numbers correspond to the numbers in the periods of the periodic table (see Table 4.1):[235]

$n = 1$, hence $l = 0$ and $m = 0$; $s = \pm\frac{1}{2}$, corresponding with two electrons.
$n = 2$, hence $l = 1$ or 0. For $l = 0$, $m = 0$; for $1 = 1$, m is –1, 0, or 1. There are four values of m, and for each, s is $\pm\frac{1}{2}$, making a total of 8.
$n = 3$, hence $l = 2$, 1, or 0; for these three cases, $m = (2, 1, 0, -1, -2)$, $(1, 0, -1)$, and 0; there are 9 cases, and for each, $s = \pm\frac{1}{2}$, making 18 in all.
$n = 4$, hence $l = 3$, 2, 1, or 0, giving 16 values of m, making 32 in all.

---

[231] ibid.
[232] ibid.
[233] ibid.
[234] ibid.
[235] ibid.

The electronic structures of the atoms of all of the elements have now been elucidated with a reasonable degree of certainty from spectroscopic data. Although Bohr's atomic model, with electrons revolving in orbits, is no longer accepted, the significance of the quantum numbers, with rather different interpretations, remains valid even today. Today, electrons are believed to be distributed in "shells" corresponding to special values of n, the maximum numbers of electrons in these shells, starting with the one next to the nucleus, being 2, 8, 18, and 32. The end of a period is occupied by an inert gas, and the number of electrons in the outer shell of an inert gas is 2 for helium and 8 for all the others.[236]

It is now well-established that the atomic nucleus is made up of protons and neutrons, with the number of protons giving the positive charge of the nucleus or the atomic number. The electrons outside the nucleus, with an equal number of negative charges and positive charges in the protons of the nucleus, make the atom neutral. These electrons have principal quantum numbers n of 1, 2, 3, etc. There exists a well-defined scheme[237] for building up the periodic table with elements, from the simplest element, hydrogen, to the heaviest elements.

4.3.18 Theory of directed valencies

The theory of directed valencies derives from the Schrödinger equation, the solution of which mathematically represents the distribution of charge of an electron regarded as distributed through space with a certain density of charge at a particular point. The solution to this equation is the wavefunction ψ, which is assumed to be finite, continuous, and single valued, in general, for all values of the coordinates, and whose square ψ^2 gives the density of charge at each point. It thus follows that, since the electron is in an atom, knowledge of the charge distribution gives an indication of the directions in which bond formation may occur.[238]

Thus, the Schrödinger wave equation led to the theory of directed valency, which states that the charge density of an electron in an atom may be either spherically symmetrical (as in the hydrogen atom), or more prominent in certain directions, in other words, have decided directions in space, corresponding to the directions of the valency bonds. The formation of a valency bond is thus interpreted as the pairing of an electron in one atom with another electron having an opposite spin in another atom. Chemists had earlier assumed, based on the facts of stereochemistry, that valency bonds have definite directions in space, a fact now borne out also by Schrödinger's

---

[236] ibid., p. 364.

[237] The scheme discussed here is excerpted from the descriptions rendered in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 364–365 (1989).

[238] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 374 (1989).

wave theory. This theory also correctly predicts the energy levels in the hydrogen atom.[239]

Application of the modern theory of valency in organic chemistry has led to new remarkable preparations of many compounds, such as ethylene and benzene, two of the main building blocks of materials used as resist resins in lithography. Today, valence bonds are regarded as examples of bonds produced by hybridization rather than examples of conventional double bonds.[240]

4.3.19 Theory of substitution

Jean Baptiste Dumas (1800–1884) formulated the theory of substitution as follows:[241]

1. "When a substance containing hydrogen is submitted to the dehydrogenating action of chlorine, of bromine, of iodine, of oxygen, etc., for each atom of hydrogen which it loses it gains an atom of chlorine, of bromine, or half an atom of oxygen."
2. "When the compound contains oxygen the same rule holds good without modification."
3. "When the hydrogenised body contains water, this loses its hydrogen without replacement, and then, if a further quantity of hydrogen is replaced, substitution takes place, as in the former cases."

4.4 Rational Synthesis of Chemical Substances

Although chemists had been synthesizing chemical substances since ancient times, the rational design and synthesis of chemical substances started when chemists—now equipped with the laws and periodic table discussed above, coupled with an accurate understanding of atomic weight, molecular weight, equivalent weight, and the definition of a molecule and an atom, following the controversy of the Karlsruhe congress of 1860—could write a chemical reaction on a piece of paper and predict *a priori* what the product would be. Before this time, synthetic chemistry was essentially an art, a series of recipes that people followed, often on a trial-and-error basis. The above-mentioned chemical laws made chemistry a quantitative and predictive science. Armed with these laws and the periodic table, chemists set out to synthesize new chemical substances.

It's true that many metals had been extracted from their ores since time immemorial, as was just mentioned. It's also true that many naturally occurring metals had been mined since ancient times, as was also mentioned.

---

[239] ibid.

[240] ibid., p. 376.

[241] ibid., p. 241.

And many elements had been isolated via electrolysis, as was described above. However, for the longest time, all of the synthesized substances were inorganic. An erroneous belief existed, which stated that inside the living body of plants and animals is an invisible flame that burns steadily, and through this flame a mysterious vital force builds up the sugars, starches, proteins, and hundreds of other very complex organic compounds that keep plants and animals alive. The belief held that this vague creative force exists in the animal and plant kingdoms, but not in the mineral world. As such, according to the belief, substances made from plants and animals differ from those of the mineral world and therefore cannot be synthesized in the laboratory. Some of the early chemists even doubted whether organic compounds obey the laws of chemistry. Even some of the celebrated chemists of the era spoke of the impassable gulf that separates organic compounds from inorganic substances.[242] This very belief, or rather opinion, guided synthetic chemistry until 1828, as we shall see later.

The first marginally successful synthesis of elemental silicon was carried out in 1811 by Joseph Louis Gay-Lussac (1778–1850) and Louis J. Thénard (1777–1857), who reacted silicon fluoride with heated potassium. However, a much more successful method was used by the Swedish chemist Jöns Jacob Berzelius (1779–1848) in 1824, which involved the reduction of potassium fluorosilicate with metallic potassium. The element silicon was eventually recognized to be one of the most abundant in the lithosphere, that is, the crust of the earth. Remarkably, even at that early stage, a debate developed as to whether silicon was a metal or an insulator, with Berzelius taking the view that it is a metal and Humphry Davy (1778–1829) that it is an insulator. This issue, aided by a deeper understanding of crystalline solids, was finally fully resolved during World War II.[243]

The scientific study of silicon continued to advance during the 19th century. Notable contributions were made by Friedrich Wöhler (1800–1882) (Fig. 4.14), who developed crystalline forms of silicon between 1856 and 1857. Although the chemistry of silicon compounds occupied a central position in the field of inorganic chemistry during much of the 19th century, precious little seems to have been reported in the literature about this elemental crystalline form until the last quarter of the century, when it began to interest metallurgists.[244]

Wöhler was also the first person to isolate the element aluminum. His method involved treating a solid salt of aluminum with potassium. His process

---

[242] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 129 (1976).

[243] F. Seitz and N.G. Einspruch, *Electronic Genie: The Tangled History of Silicon*, University of Illinois Press, Urbana, p. 3 (1999).

[244] ibid., p. 3; J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, pp. 230–232, 238 (1989).

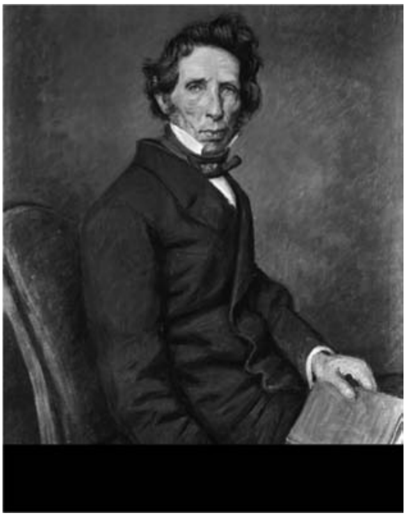


Figure 4.14 Friedrich Wöhler (1800–1882), the founder of modern organic chemistry, who in 1828 succeeded in producing urea (an organic substance) by evaporating a solution of isomeric ammonium cyanate (an inorganic substance). He also succeeded in producing crystalline silicon and boron between 1856 and 1857. (Printed with permission of the Deutsches Museum, Munich.)

was not simple nor economical. A much simpler and more-economical process for the production of aluminum was patented by Charles Martin Hall (1863–1914) in 1886; this marked the beginning of the Aluminum Company of America (ALCOA), which in the years since, has been producing millions of tons of pure aluminum.[245] ALCOA is still in operation today.

Wöhler also isolated beryllium and yttrium, and because of illness, *almost* discovered the fourth metal vanadium, which was later isolated by Nils Gabriel Sefström (1787–1845).[246] The story surrounding the discovery of this metal is a very interesting one and needs to be retold here, if only for didactic reasons. Wöhler had sent a specimen of a lead ore containing this unknown metal to his friend and mentor Berzelius, asking him to analyze it. Berzelius gave it to his student Sefström, who analyzed the mineral and discovered that it was indeed vanadium oxide, from which he isolated vanadium. Berzelius wrote back to Wöhler about the discovery with the following story:[247]

> "In the remote regions of the north there dwells the goddess Vanadis, beautiful and lovely. One day there was a knock at her door. The goddess was weary and thought she would wait to see if the knock

---

[245] B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 139 (1976).

[246] Cited in B. Jaffe, *Crucibles: The Story of Chemistry – From Ancient Alchemy to Nuclear Fission*, 4th ed., Dover Publications, New York, p. 141 (1976).

[247] ibid.

> would be repeated, but there was no repetition. The goddess ran to the window to look at the retreating figure. "Ah" she said to herself, "it is that fellow Wöhler." A short time afterward there was another knock, but this time so persistent and energetic that the goddess went herself to open it. It was Sefström, and thus it was that he discovered *vanadium*. Your specimen is, in fact, oxide of vanadium. But, the chemist who has invented a way for the artificial production of an organic body can well afford to forgo all claims to the discovery of a new metal, for it would be possible to discover ten unknown elements with the expenditure of so much genius."

The first successful synthesis of an organic compound dates back to 1828, when Wöhler showed that urea, a typical organic compound, could be produced by merely heating ammonium cyanate, an inorganic compound; both compounds are isomeric. The prevailing view at the time, as stated above, was that in the organic world, the existence of a mysterious "vital force" was responsible for the formation of compounds, and their preparation in the laboratory could hardly be expected.[248] Influenced by Wöhler, many organic chemists went on to prepare new chemicals. Marcellin Bertholet (1827–1907) was the first person to prepare formic acid, the liquid responsible for the sting of insects. Hermann Kolbe (1818–1884) prepared acetic acid without the use of sweet cider or the mother-of-vinegar bacteria. William Perkin (1838–1907) discovered a method for synthesizing mauve—the first in a long series of coal tar dyes, which rival the colors of nature. Adolf von Baeyer (1835–1917) discovered the formula for indigo and made possible the manufacture of synthetic indigo 15 years later by the Badische Company of Germany, the parent company of BASF (Badische Anilin und Soda Fabrik).[249]

Other notable discoveries in organic chemical synthesis that directly influenced the development of lithography include the discovery by Peter Greiss (1829–1888) in 1860 of azo compounds, which are obtained by the action of nitrous acid at low temperatures on a solution of salts of aromatic amines such as aniline. These azo compounds form the starting point in the preparation of numerous organic compounds, including azo dyes, the first member of which (aminoazobenzene, or "aniline yellow") was obtained by C. Mène in 1861, and even diazonaphthoquinone (DNQ) sensitizer used in resist applications for g-line lithography, h-line lithography, and i-line lithography, and lithographic printing plates.

Interest in synthetic polymer chemistry received a tremendous boost in the 1920s from the success of the industrial-scale production of polystyerene under the direction of Herman Mark (1895–1992) at I.G. Farben in Germany.

---

[248] ibid., p. 220.

[249] ibid., pp. 139–140.

In the 1930s, high-pressure or low-density polyethylene was developed by ICI in England. By the 1930s, Wallace Carothers (1896–1937) at DuPont started synthesizing polymers by applying well-established reactions of organic chemistry such as esterification and amidation. His products—polyesters and polyamides—contained many ester and amide linkages, in contrast to the single ester or amide linkages in products formed from simple organic molecules undergoing similar reactions. From this research came the development of the first commercially successful synthetic rubber, *neoprene*, in 1931, followed by *nylon* at DuPont. World War II saw the development of a large industrial program to develop styrene-butadiene rubber (SBR), a substitute for natural rubber. This program received important contributions from many scientists of the era, including the prominent work of William J. Sparks at Esso Research. Karl Ziegler (1898–1973) and Giulio Natta (1903–1979) also made major contributions to synthetic polymer chemistry. Paul J. Flory made significant contributions to polymer physics.

Industrial centers for fundamental research on polymers were established as early as the 1920s and 1930s at I.G. Farben in Germany and DuPont in the U.S., respectively. Some universities in the U.S., Europe, and Japan established centers for polymer research, starting in the 1930s. A few examples include the Institute of Rubber Chemistry at Akron, Ohio, which was established in the 1930s; Polymer Research Institute at Brooklyn Polytechnic Institute, which was established in the 1940s; and the Department of Polymer Science and Engineering at the University of Massachusetts at Amherst, which was established in the 1960s.

The majority of resists currently used in lithography are made of organic polymeric materials. Improvements in the manufacture and processing of these polymeric resins are used in all forms of lithography today and derive directly from advances in polymer science.

4.5 The Structure of Chemical Compounds and Phenomena Deriving from Them

The spatial orientation of chemical compounds can have a significant influence on their chemical properties. Isomers—compounds with the same composition and molecular formula but different arrangement of atoms and structural formula and properties—have been known since 1827 when Friedrich Wöhler prepared silver cyanate and discovered that, although its elemental composition was identical to that of silver fulminate, prepared by Justus von Liebig (1803–1873) in 1826, its properties were quite different.[250]

---

[250] F. Kurzer, "Fulminic acid in the history of organic chemistry," *J. Chem. Educ.* **77**(7), 851–857 (2000).

4.5.1 The structural formula of benzene

Benzene was one of the synthetic compounds whose chemical structure posed a significant challenge for chemists to elucidate in the late 19th century. August Kekulé (1829–1896), falling asleep in front of his fire in Ghent, had a dream of wiggling snakes and awoke knowing the solution of a difficult problem on the structure of benzene.[251] He postulated the hexagonal formula for benzene in 1866 (see Fig. 4.15), in which the six atoms of carbon in the benzene ring are regarded as being joined by alternating single and double linkages to preserve the tetravalency of carbon.[252]

The importance of Kekulé's benzene formula in organic chemistry cannot be overemphasized. It explains isomerism in the derivatives of benzene and has proved to be sufficiently useful during the years of intensive work since it was proposed. X-ray studies of the structure of aromatic compounds have also confirmed the existence of the flat hexagonal ring of carbon atoms.[253] Today, benzene is a major feedstock used in the manufacture of thousands of compounds, including a great majority of lithographic resists.

Early attempts, at least in the modern era, to characterize polymeric substances generally indicated that polymers possess high molecular weights, even if the results were not too accurate. These early investigators found it easier to interpret the results in terms of colligative properties than to simply accept the possibility that these substances were high-molecular-weight compounds, composed from smaller units that are held together by covalent bonds.

As early as 1826, Faraday (1791–1867) had already arrived at C_5H_8 as the empirical formula of rubber, and by 1860, isoprene was identified as a product of the destructive distillation of rubber. However, the idea that a natural polymer such as rubber somehow "contained" isoprene gradually emerged,

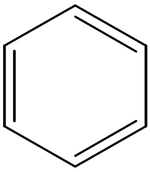


Figure 4.15 Structural formula of benzene proposed by Kekulé.

---

[251] Kekulé reports that the origin of the discovery of the structure of benzene appeared to him while he was asleep in his study in Ghent, whereby he saw the long chains of carbon atoms like snakes twisting and curling until one "gripped its own tail and the picture whirled scornfully before my eyes" [cited in *Berichte der Deutschen Chemische Gesellschaft* **xxiii**, 1302 (1890)].

[252] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 290 (1989).

[253] ibid., p. 292.

but the nature of its involvement was more elusive and sometimes outright contentious.[254]

The late and slow progress in the elucidation of the basic structure of polymers is generally agreed to be due primarily to a lack of interest on the part of organic and physical chemists. On one hand, organic chemists were more interested in determining the structure of ordinary-sized organic molecules during the early years of the 20th century, and this probably contributed to their reluctance to look beyond structures of convenient size. Physical chemists, on the other hand, were more interested in intermolecular forces during this period, and the idea that polymers were the result of some sort of association between low-molecular-weight constituent molecules prevailed much longer than it should have.[255]

Hermann Staudinger (1881–1965) (Fig. 4.16) is generally credited with being the father of modern polymer chemistry, for proving that polymers are long-chain molecules made up of repeating units and not loose colloidal associations as had been previously thought. In 1920, he proposed the chain formulas we use today, maintaining that structures are held together by covalent bonds that are equivalent in every way to those in low-molecular-weight compounds. These were rather controversial ideas at the time.

Figure 4.16 Hermann Staudinger, the father of polymer chemistry (printed with permission of the Deutsches Museum, Munich).

---

[254] For excellent books on polymer chemistry, see, e.g., P.C. Himenez, *Polymer Chemistry: The Basic Concepts*, Marcel Dekker, New York (1984); H.R. Allcock and F.W. Lampe, *Contemporary Polymer Chemistry*, 2nd ed., Prentice Hall, Englewood Cliffs, New Jersey (1990); P.J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, New York (1953).

[255] ibid.

Nevertheless, over time, Staudinger's ideas were gradually accepted, culminating in him being awarded the Nobel Prize in Physics in 1953.

In addition to the intramolecular covalent bonds that keep the polymer molecules intact, we now know that intermolecular forces—hydrogen bonds, dipole–dipole interactions, and London forces—hold assemblies of these molecules together in the bulk state. Physical chemists like Paul J. Flory (1910–1985), Werner Kuhn (1899–1963), Hermann Mark (1895–1992), and a host of others eventually turned their attention to polymers, and soon they were applying thermodynamics, statistics, and crystallography to describe the multitude of forms that long-chain molecules could assume.[256] Today, polymer studies span the entire gamut of physical sciences and engineering, e.g., in polymer chemistry, polymer physics, polymer engineering, and so on.

4.5.2 Optical activity

Optical activity involving the ability to rotate the plane of polarized light was first observed by Jean-Baptiste Biot (1774–1862) in 1815–1835 in a number of naturally occurring organic compounds, such as turpentine, camphor, sugars, and tartaric acid. Since optically active compounds exhibited this property both in their crystalline form as well as in solution, it was reasoned that this property is inherent in the molecules. Eilhard Mitscherlich (1794–1836) in 1844 made the observation that, although tartaric and racemic acids are isomeric, the former and its salt are optically active, while racemic acid is inactive.[257]

In an attempt to elucidate the observation of Mitscherlich, in 1848 Louis Pasteur (1822–1895) observed that the crystals of tartaric acid are characterized by small facets that had unaccountably been overlooked by Mitscherlich and were similar to the hemihedral facets observed by René Just Haüy (1743–1822) in optically active quartz. Quartz crystals are in two forms, i.e., right and left handed, depending on the position of the facets. The two forms are related to each other by the same relation that an object has with its own mirror image. John Herschel (1792–1871) pointed out in 1821 that this was probably related to Biot's observation that some quartz crystals rotate the plane of polarized light to the right and some to the left.[258]

Pasteur's other main important discovery, among many others in this field, was the resolution of sodium ammonium racemate, an optically inactive compound, by crystallization (1848–1850). In describing his discovery, he remarks:[259]

---

[256] ibid.

[257] ibid., p. 300.

[258] ibid., pp. 300–301.

[259] Alembic Club Reprint No. 14; *Delephine, Bull. Soc. Chim.* **xxxvii** (1925) [cited in J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 302 (1989)].

> "I carefully separated the crystals which were hemihedral to the right from those which were hemihedral to the left, and examined their solutions separately in the polarizing apparatus. I then saw with no less surprise than pleasure that the crystals hemihedral to the right deviated the plane of polarization to the right, and that those hemihedral to the left deviated it to the left; and when I took an equal weight of each of the two kinds of crystals, the mixed solution was indifferent towards the light in consequence of the neutralization of the two equal and opposite individual deviations."

The interpretation of Pasteur's results in terms of the relation of optical activity to the presence of an asymmetric carbon atom was given independently in 1874 by Jacobus Hencricus van't Hoff (1852–1911) and Joseph Achille Le Bel (1847–1930); in so doing, they laid the foundation of stereochemistry. According to van't Hoff: [260]

> "In the case where the four affinities of an atom of carbon are saturated by four different univalent groups, two and only two different tetrhedra can be obtained, of which one is the mirror image of the other…, two structural formulae of isomers in space."

Earlier in 1867, Kekulé had described the tetrahedral model of the carbon atom, with a central sphere as the atom and four wires pointing to the corners of the tetrahedron as the valencies. Van't Hoff also stated that "Every carbon compound which in solution impresses a deviation on the plane of polarization possesses an asymmetric carbon atom," i.e., one in which the four groups that are attached to the carbon are all different.[261]

It has since been established that optical activity can exist in compounds with no asymmetric carbon atoms, as long as they can form two non-superimposable molecular configurations. For instance, optical activity has been established in the case of compounds of selenium, tin, sulfur, chromium, etc., some of the compounds being quite free of carbon.[262] Optical activity in modern chemistry is believed to originate in the enantiomorphous molecular configurations of compounds that possess it.

---

[260] J.R. Partington, *A Short History of Chemistry*, 3rd ed., Dover Publications, New York, p. 304 (1989).

[261] ibid., p. 305.

[262] ibid., pp. 306–307.

Chapter 5

Evolution of Lithography

> "The old order changeth, yielding place to new."
> Alfred Tennyson (1809–1892), *The Idylls of the King*

5.1 Introduction

From the time of Senefelder to the present, lithography has undergone tremendous evolution, but overall, its basic principle remains intact. It remains a planographic printing process in which the image and nonimage areas are on the same plane of the printing substrate.[1] Just as in Senefelder's time, today, the contrast between the image and nonimage areas depends on the interfacial tension of oily inks and water-retaining surfaces. This is in fact the very same principle that governs the immiscibility of oil and water—hydrophobic–hydrophilic interactions between the image and nonimage areas. The inks (or resist) are naturally hydrophobic and can be made resistant to the action of etchants when dry.

Lithography has evolved into a broad family of techniques that are distinctly different in many respects, yet share similar attributes as to their ultimate goal—a method to transfer information from one substrate to another with a high degree of fidelity and throughput relative to other comparable printing techniques. Senefelder's original direct stone plate lithography of the late 18th century evolved into offset lithography in the 19th century and is now primarily used in the printing industry. Offset lithography in turn evolved into semiconductor lithography in the 20th century, where it is now used in the fabrication of ICs. Many variants of lithography are currently practiced, ranging from stone plate lithography used in fine art printing, to offset lithography used in the printing of newspapers and the like, to semiconductor lithography that utilizes a variety of exposure radiations to print ICs, to emerging lithographies based on molecular self-assembly, imprinting or

---

[1] Lithography is one of the three basic printing methods, the other two being intaglio (gravure or engraving) and typography (relief printing).

embossing, scanning probes, etc. Although very interesting in their own right, stone plate and offset lithographies will not be discussed in detail in this chapter.

Another notable evolution in lithography relates to the manner in which the image is created. While Senefelder drew his image by hand using ink pencils, this mode of lithographic printing is currently practiced only by fine-art lithographers. Most lithographic images today are made through radiation-induced chemical reactions on appropriate sensitive substrates, using photons, electrons, ions, and even atoms.

Still another aspect of the evolution of lithography relates to the speed with which it is carried out. While Senefelder's stone plate lithography took hours, if not an entire day, to complete just one print, today, an IC device layer on a silicon wafer substrate containing hundreds of devices, each of which may have nearly one billion transistors, can be patterned in just under one minute.

Despite the tremendous evolutionary changes and developments that lithography has undergone, at its fundamental core, its printing and patterning action is ultimately about chemical transformations brought about by the interactions of electrons within the ink or resist material, or any lithographic patterning media, as the case may be. Such electronic interactions may be based on the affinity of oil for oil and the repulsion of oil and water, as in stone lithography (solid plate printing), or on the repulsion between incompatible blocks of copolymers, leading to phase separations (as in block copolymer self-assembly lithography); it could also be mediated by photons (as in photolithography), by electrons (as in electron-beam lithography), and by ions (as in ion-beam lithography). When photoradiations (such as visible light photons, UV photons, x-ray photons, etc.), electrons, and ions are employed in the imaging process, they function primarily to excite the electrons of the sensitized lithographic substrates or radiation-sensitive materials in resists to appropriate energy levels, altering their physical properties such as solubility in an appropriate solvent, and thereby providing the basis of the contrast between the exposed and unexposed areas of the substrate or the resist. When such interactions are done properly, the result is a successful marriage between the exciting agent (electrons, photons, ions, etc.,) and the electrons of the lithographic substrate, whose offspring is a well-resolved and well-patterned planographic feature. The tone of this feature could be either positive (when the exposed area is washed or dissolved away from the substrate) or negative (when the exposed area remains on the substrate).

As mentioned Chapter 1, lithography is venturing beyond its traditional application field of electronics to other fields of science and technology. This trend is driven largely by a combination of two factors. First, advanced lithographic fabrication is becoming more accessible to a widening pool of new users because investment costs have decreased substantially and clean room facilities have become standard fixtures in many research laboratories and institutes. Second, advances in advanced lithographic fabrication that are driven mostly by the semiconductor industry have made it possible to fabricate increasingly complex and small structures, continually creating opportunities for new fundamental experiments and technologies.[2]

The motivation for reducing the device feature size in, for example, ICs is premised on the fact that smaller features result in higher device speeds and decreasing energy consumed per computing function. The patterning of most IC devices is currently done with photolithography, where the minimum feature scales with the exposure wavelength. Initial device miniaturization was realized by reducing the exposure wavelength. However, the lack of high-throughput laser sources, resist materials, and optically transparent materials for lenses at energies beyond 193 nm have slowed down continued wavelength reduction as a device-feature-scaling strategy. As such, a combination of immersion (in water) optics and resolution enhancement techniques is the strategy being pursued today to push the resolution of optical lithography beyond its current limits.[3,4,5,6,7,8,9,10,11] This approach is not without issues: the exponential increases in manufacturing costs offset, in significant part, the benefits of decreased feature sizes. Should this trend continue, it will become economically very expensive to implement new manufacturing technologies that rely solely on the conventional semiconductor manufacturing processes in the nanometer-scale regime.[12,13] To circumvent the above limitations, new

---

[2] L. Rassaei, P.S. Singh, and S.G. Lemay, "Lithography-based nanoelectrochemistry," *Anal. Chem.* **83**, 3974–3980 (2011).

[3] R.F. Pease and S.Y. Chou, "Lithography and other patterning techniques for future electronics," *Proc. IEEE* **96**, 248–270 (2008).

[4] S. Owa and H. Nagasaka, "Advantage and feasibility of immersion lithography," *J. Microlithogr. Microfabr. Microsyst.* **3**(1), 97–103 (2004) [doi: 10.1117/1.1637593].

[5] S. Owa, H. Nagasaka, K. Nakano, and Y. Ohmura, "Current status and future prospect of immersion lithography," *Proc. SPIE* **6154**, 5408–54011 (2006) [doi: 10.1117/12.656887].

[6] H. Sewell, D. Mulkens, D. McCafferty, L. Markoya, B. Streefkerk, and P. Graeupner, "The next phase for immersion lithography," *Proc. SPIE* **6154**, 615406 (2006) [doi: 10.1117/12.657574].

[7] R. Menon and H.I. Smith, "Absorbance-modulation optical lithography," *J. Opt. Soc. Am. A* **23**, 2290–2294 (2006).

[8] D. VanDenBroeke, K. Wampler, X. Shi, R. Socha, and K. Gronlund, "Model-based RET using interference maps, algorithms for random contacts at 65 nm," *Solid State Technol.* **47**, 63–6 (2004).

[9] Y.C. Pati, A.A. Ghazanfarian, and R.F. Pease, "Exploiting structure in fast aerial image computation for integrated circuits," *IEEE Trans. Semiconductor Manuf.* **10**, 62–75 (1995).

[10] T.A. Brunner, "Why optical lithography will live forever," *J. Vac. Sci. Technol. B* **21**, 2632–2637 (2003).

[11] C.G. Willson and B.J. Roman, "The future of lithography: Sematech litho forum," *2008 ACS Nano* **2**, 1323–1328 (2008).

[12] S.E. Thompson and S. Parthasarathy, "Moore's law: the future of Si microelectronics," *Mater. Today* **9**, 20–25 (2006).

[13] M. Lundstrom, "Moore's law forever?" *Science* **299**, 210–211 (2003).

alternative lithographic patterning techniques that do not rely solely on focused energetic beams and are thus not wavelength-limited are now being actively researched and developed.

These new alternative lithographic patterning techniques include imprint lithography, molecular self-assembly lithography based on the self-assembly of colloidal particles and block copolymers, and scanning probe lithography. These techniques can be applied to a broad range of substrates such as, for example, biological substrates,[14,15,16] some of which are not patternable with radiation-based lithography. In addition, they can be combined with conventional semiconductor lithography in a sort of hybrid mode. Furthermore, they are simple enough to be used in a typical university laboratory setting and do not require extremely expensive equipment.[17]

Imprint lithography is based on the pressure-driven flow of soft materials to transfer patterns onto surfaces. It can be used with a variety of adsorbate and substrate systems and enables rapid, and in some cases parallel, printing of features with sizes down to single-digit nanometers, and most importantly, at a fraction of the cost of standard radiation-based lithographic strategies. Molecular self-assembly is based on the thermodynamics of the assembling molecules' or particles' drive to minimize their surface energies. Scanning probe lithography (SPL), on the other hand, is based on the interaction between the substrate and the probe in a scanning probe microscope, such that this interaction induces physical changes or chemical reactions on the substrate. Such interactions include applied forces,[18,19,20] potentials,[21,22,23]

---

[14] D.G. Castner and B.D. Ratner, "Biomedical surface science: foundations to frontiers," *Surf. Sci.* **500**, 28–60 (2002).

[15] C.S. Chen, M. Mrksich, S. Huang, G.M. Whitesides, and D.E. Ingber, "Micropatterned surfaces for control of cell shape, position, and function," *Biotechnol. Prog.* **14**, 356–363 (1998).

[16] A. Bernard, J.P. Renault, B. Michel, H.R. Bosshard, and E. Delamarche, "Microcontact printing of proteins," *Adv. Mater.* **12**, 1067–1070 (2000).

[17] H.M. Saavedra, T.J. Mullen, P. Zhang, D.C. Dewey, S.A. Claridge, and P.S. Weiss, "Hybrid strategies in nanolithography," *Rep. Prog. Phys.* **73**, 036501(2010).

[18] B. Cappella and H. Sturm, "Comparison between dynamic plowing lithography and nanoindentation methods," *J. Appl. Phys.* **91**, 506–512 (2002).

[19] Y. Sugimoto, M. Abe, S. Hirayama, N. Oyabu, O. Custance, and S. Morita, "Atom inlays performed at room temperature using atomic force microscopy," *Nature Mater.* **4**, 156–159 (2005).

[20] D.W. Abraham, H.J.G. Mamin, and J. Clarke, "Surface modification with the scanning tunneling microscope," *IBM J. Res. Dev.* **30**, 492–499 (1986).

[21] E.S. Snow and P.M. Campbell, "Fabrication of Si nanostructures with an atomic-force microscope," *Appl. Phys. Lett.* **64**, 1932–1934 (1994).

[22] Y. Li, B.W. Maynor, and J. Liu, "Electrochemical AFM 'dip-pen' nanolithography," *J. Am. Chem. Soc.* **123**, 2105–2106 (2001).

[23] J.A. Dagata, J. Schneir, H.H. Harary, C.J. Evans, M.T. Postek, and, J. Bennett, "Modification of hydrogen-passivated silicon by a scanning tunneling microscope operating in air," *Appl. Phys. Lett.* **56**, 2001–2003 (1990).

and/or electrons.[24,25,26,27,28,29] It is possible to fabricate localized and site-specific functional structures that can be simultaneously characterized *in situ* with SPL because of the nanoscale precision and atomic resolution of the technique. Just like imprint lithography, SPL techniques can be utilized in a wide range of materials, including metals, semiconductors, polymers, and biological molecules, while operating in liquid, air, or vacuum.[30]

It should be pointed out that the resolution of these alternative lithographic techniques are set by physical and chemical interactions, such as diffusion, electrostatic forces, and intermolecular interactions.[31,32] As an example, patterned clusters and small molecules often suffer from pattern degradation or dissolution arising from molecular diffusion.[33,34] These techniques suffer from a lack of registration, addressability, interconnectability, and scalability to make functional devices. The issue of registration and addressability of the alternative lithographic techniques can be mitigated by combining the latter technique with one or more of the conventional, top-down advanced lithographic techniques in a hybrid mode. As opposed to conventional top-down lithography, which is subtractive, with hybrid approaches between conventional and alternative lithographies, it is also

---

[24] M.A. McCord and R.F.W. Pease, "Lithography with the scanning tunneling microscope," *J. Vac. Sci. Technol. B* **4**, 86–88 (1986).

[25] M.A. McCord, D.P. Kern, and T.H.P. Chang, "Direct deposition of 10-nm metallic features with the scanning tunneling microscope," *J. Vac. Sci. Technol. B* **6**, 1877–1880 (1988).

[26] M.A. McCord and R.F.W. Pease, "Exposure of calcium-fluoride resist with the scanning tunneling microscope *J. Vac. Sci. Technol. B* **5**, 430–433 (1987).

[27] J.R. Sheats and B.W. Smith, Eds., *Microlithography: Science and Technology*, Marcel Dekker, New York (1998).

[28] H. Iwasaki, T. Yoshinobu, and K. Sudoh, "Nanolithography on SiO_2/Si with a scanning tunneling microscope," *Nanotechnology* **14**, R55–R62 (2003).

[29] E.C.H. Sykes, L.C. Fernandez-Torres, S.U. Nanayakkara, B.A. Mantooth, R.M. Nevin, and P.S. Weiss, "Observation and manipulation of subsurface hydride in Pd(111) and its effect on surface chemical, physical, and electronic properties," *Proc. Natl Acad. Sci. USA* **102**, 17907–17911 (2005).

[30] H.M. Saavedra, T.J. Mullen, P. Zhang, D.C. Dewey, S.A. Claridge, and P.S. Weiss, "Hybrid strategies in nanolithography," *Rep. Prog. Phys.* **73**, 036501 (2010).

[31] E. Delamarche, H. Schmid, A. Bietsch, N.B. Larsen, H. Rothuizen, B. Michel, and H. Biebuyck, "Transport mechanisms of alkanethiols during microcontact printing on gold," *J. Phys. Chem. B* **102**, 3324–3334 (1998).

[32] A.A. Dameron, J.R. Hampton, R.K. Smith, T.J. Mullen, A.D. Gillmor, and P.A. Weiss, "Microdisplacement printing," *Nano Lett.* **5**, 1834–1837 (2005).

[33] A.A. Dameron, J.R. Hampton, S.D. Gillmor, J.N. Hohman, and P.S. Weiss, "Enhanced molecular patterning via microdisplacement printing," *J. Vac. Sci. Technol. B* **23**, 2929–2932 (2005).

[34] J.R. Hampton, A.A. Dameron, and P.S. Weiss, "Double-ink dip-pen nanolithography studies elucidate molecular transport," *J. Am. Chem. Soc.* **128**, 1648–1653 (2006).

possible to build on surface features, while also providing control over structure and function.[35,36,37,38,39,40]

Relative to other patterning techniques, how and why did lithography become the dominant printing method for the fabrication of ICs? An adequate answer to this question must trace the evolution of lithography from its invention through all of its myriad incarnations that have been practiced ever since then, through the invention of the transistor and the IC, as well as the appropriation of its techniques in the large-scale fabrication of semiconductor devices. This journey must perforce highlight the unique properties of this patterning technique, with particular emphasis on semiconductor lithography, which has made it the dominant patterning technique used in the fabrication of ICs.

Figure 5.1 is a schematic of the major evolutionary trends and developments in lithography, starting from its very invention in 1798 and extending all the way to its role in the development of microelectronics—first the printed circuit board, next the transistor, and then the integration of the transistors into microprocessors, memory chips, and controllers—that have ushered in a cornucopia of machines that manipulate information by streaming electrons through silicon. Today, the entire microelectronics industry rests on lithographic techniques that routinely fabricate structures smaller than 10 nm across (that is, 10 billionths of a meter). This size is tiny by the standards of everyday experience—about five hundred and fifty-two thousandths the width of a human hair—but is comparatively large on the scale of atoms and molecules.

The major consequential transitions in the evolution of lithography as highlighted in Fig. 5.1 include the following: the transition (1) from engraving and relief printing to stone plate lithography, (2) from fleeting images to permanent photographs, (3) from photolithography to offset lithography and semiconductor lithography, (4) from arc lamps to exciplex and excimer lasers

---

[35] A. Hatzor and P.S. Weiss, "Molecular rulers for scaling down nanostructures," *Science* **291**, 1019–1020 (2001).

[36] Y. Huang, X.F. Duan, Y. Cui, L.J. Lauhon, K.H. Kim, and C.M. Lieber, "Logic gates and computation from assembled nanowire building blocks," *Science* **294**, 1313–1317 (2001).

[37] M.E. Anderson, M. Mihok, H. Tanaka, L.P. Tan, M.W. Horn, G.S. McCarty, and P.S. Weiss, "Hybrid approaches to nanolithography: photolithographic structures with precise, controllable nanometer-scale spacings created by molecular rulers," *Adv. Mater.* **18**, 1020–1022 (2006).

[38] J. Henzie, J.E. Barton, C.L. Stender, and T.W. Odom, "Large-area nanoscale patterning: chemistry meets fabrication," *Acc. Chem. Res.* **39**, 249–257 (2006).

[39] M.W. Li, R.B. Bhiladvala, T.J. Morrow, J.A. Sioss, K.K. Lew, J.M. Redwing, C.D. Keating, and T.S. Mayer, "Bottom-up assembly of large-area nanowire resonator arrays," *Nature Nanotechnol.* **3**, 88–92 (2008).

[40] T.J. Morrow, M.W. Li, J. Kim, T.S. Mayer, and C.D. Keating, "Programmed assembly of DNA-coated nanowire devices," *Science* **323**, 352 (2009).

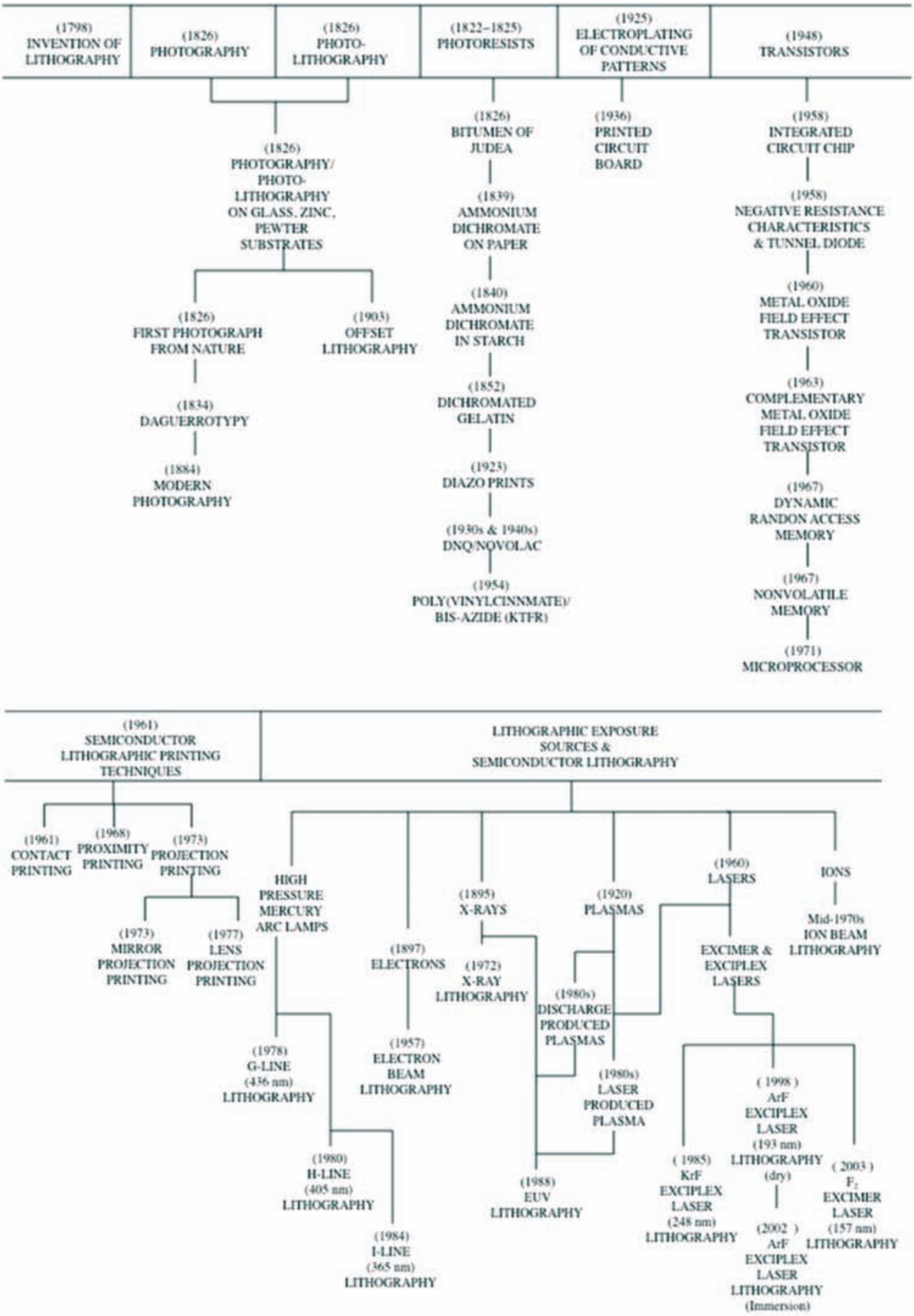


Figure 5.1 Evolution of lithography.

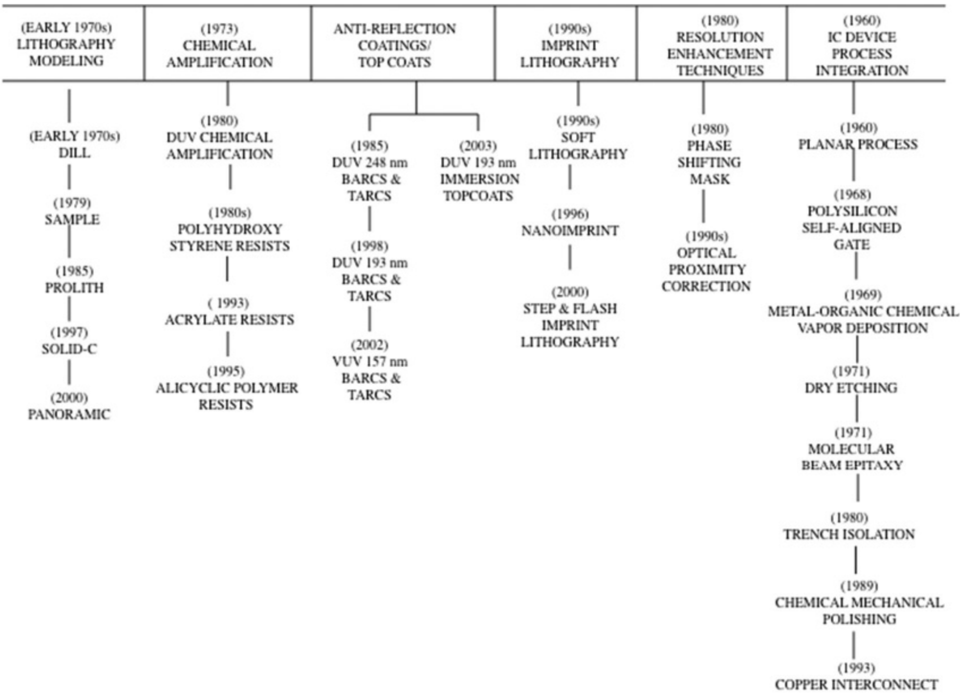


Figure 5.1 (Continued).

to discharge- and laser-produced plasma, (5) from DNQ/novolak to chemical amplification resists, (6) from single patterning to double patterning, (7) from binary masks to phase-shifting masks, (8) from reticle enhancement technologies to optical proximity correction schemes, (9) from vacuum tubes to transistors, (10) from discrete transistors to ICs and to memories and microprocessors, and (11) from device-speed scaling to number-of-cores scaling.

The early history of the invention of lithography, photography, and photolithography has already been covered in Chapter 2 and will not be repeated here. Instead, our starting point shall be the developments that took place after the invention of photolithography in 1822 by Niépce. While the developments in fine-art lithography[41] and offset lithography[42] are interesting subjects in their own rights, here our emphasis will be on how developments in these lithographies were appropriated and employed in the development of

---

[41] The most notable exponent of fine-art lithography was Toulouse Lautrec. For literature on key developments in fine-art lithography, see J. Renard, *Histoire Naturelles* (*Natural History*), (English Translation by R. Howard), Horizon Press, New York (1966); M. Joyant, *Henri de Toulouse-Lautrec*, Arno Press, New York (1968).

[42] See, e.g., H. Gernsheim and A. Gernsheim, *The History of Photography: From the Camera Obscura to the Beginning of the Modern Era*, Thames & Hudson, London (1969).

printing plate technology, offset lithography, and, subsequently, semiconductor lithography, the most advanced form of lithography.

5.2 Offset Lithography

We have come a long way from the stone plate lithography that Senefelder invented. As you can very well imagine, it would be quite difficult to mount a piece of stone on a modern high-speed lithographic press. While Senefelder's press was flat, making the use of stone plates rather easy, modern lithographic printing presses use curved cylinders to hold the offset plates—image carriers such as thin paper, plastic, or a metal sheet—which, once exposed and processed, can be wrapped around a cylinder of a press for printing. Offset lithography therefore refers to a printing method in which the inked lithographic plate prints the image on a rubber blanket or roller (the offset), which in turn prints the image on the object (for example, a newspaper or a circuit board).

The image areas of a modern offset plate accept ink and are hydrophobic, while the nonimage areas accept water and are hydrophilic. Although this basic principle is common to all forms of offset lithography, there are many differences between offset plates and the method they use to separate the image from the nonimage areas.[43]

Today, most lithographic plates are made of aluminum sheets of varying thickness with a granular finish. The grains serve to imbue the plate with water-carrying properties and anchorage to the image-forming material. The plates are typically pre-sensitized, i.e., coated with light-sensitive diazo compounds and photopolymer resins that make them able to be responsive to light exposure.[44] This photopolymer coating reacts with ultraviolet radiation, much like photographic film. Upon exposure to UV radiation through a film negative, the exposed part of the photopolymer on the plate (i.e., the part under the transparent section of the film negative) is cross-linked and "hardened," while the unexposed part (i.e., the part under the dark areas of the film negative) remain unaffected by the UV radiation and thus can be easily removed in an appropriate solvent, resulting in a printing plate that can be used for offset lithography.

The seven major types of lithographic plates that are commonly used in the printing industry today include diazo plates, photopolymer plates, silver halide plates, electrophotographic plates, bimetallic plates, waterless plates, spark discharge plates, laser (digital waterless) plates, and plates made from

---

[43] P.B. Meggs, *A History of Graphic Design*, 3rd ed., John Wiley & Sons, Hoboken, New Jersey, pp. 146–150 (1998); "Offset Printing," Encyclopedia Britannica, https://www.britannica.com/technology/offset-printing, retrieved 4th November 2019; "Printing Process, https://dynodan.com/magazine-printing/process/uncategorized/printing-process/, retrieved 4th November 2019.

[44] ibid.

various materials using computer-to-plate technology. In addition to digital information from computers, film positives and negatives are used in making or exposing most types of offset plates today. While there are several types of offset plates, they are all generally classified as either positive or negative working plates, depending on the tone of their printing.[45]

The ink used in offset lithography has properties similar to those used in a letterpress in that both inks are very viscous. The water solution, called a "fountain solution," is used to keep the ink from adhering to the non-image areas of the offset plate. The fountain solution is formulated from materials such as desensitizing gums, cleaners, buffering agents, etc.[46] Offset lithographic printing is the dominant printing method used in the printing of newspapers, textbooks, and the like.

5.3 The Printed Circuit Board and the Development of the Electronics Industry

Printed circuit boards are in almost all electronics and computerized devices that make our modern world possible. They are the essential framework on which electronic components are assembled and integrated to make functional devices. The printed circuit board (PCB)[47] and modern resist technologies are both outgrowths of the printing industry, whereas resists had been known for more than a century before the invention of the PCB. In early PCB processes, etch resists were applied by offset lithographic printing, resulting inevitably in the transfer of printing technologies to the PCB industry.[48] To better understand the role that lithography, specifically offset lithography, played in the development of PCB, a brief discussion on the invention and development of PCBs is presented.

The roots of PCBs stretch all the way back to 1903, when the German scientist Albert Hanson filed a British patent[49] for a device meant to improve telephone exchange boards. His simple device was made using through-hole construction, which allowed conductor wires to be passed through the holes in the substrate to be soldered or bonded to copper or brass traces on the flat surface of the substrate. The traces were adhesively bonded to the substrate (parrafined paper and the like). Although Hanson's device was not a truly *printed* circuit board, as the conductor metal patterns were either cut or stamped out, it did produce conductive metal patterns on a dielectric

---

[45] ibid.

[46] ibid.

[47] The account rendered here is adapted from a similar one given in W.S. DeForest, *Photoresist: Materials and Processes*, McGraw-Hill, New York, Chapter 1 (1975).

[48] ibid., p. 6.

[49] A. Hanson, "Improvements in or connected with electrical cables and jointing of the same," British Patent 4,681 (1903).

substrate. And, most importantly, it was the first documented circuit invention that anticipated many aspects of the modern PCB.

The next major contributor to the development of PCBs was Charles Ducas, who in 1925 used a stencil to print a conductive paste on a plastic base and electroplated it with metals such as copper, gold, and silver. Ducas was granted a patent for a method of forming electroplated conductor patterns in copper, gold, or silver on a nonconductive base material. Although he formed the conductor patterns by stenciling rather than photoetching, these patterns are of interest due to their remarkable resemblance to modern-day circuit patterns.[50]

Ducas designed these conductors to interconnect circuit elements in place of wires, while using the base as an insulator in a method he described in one of the claims of the invention for "constructing electrical circuits arranged ... upon the panel surfaces ... for interconnecting terminals of electrical instruments secured to [the] surfaces."[51] Remarkably, there exists no evidence that extensive commercial use was made of this invention, or of any subsequent circuit-board invention, until World War II.[52]

The year 1936 saw Paul Eisler,[53] an Austrian inventor living in England, conceive what is probably the earliest PCB as we know it today. Using his background in the printing industry, he arrived at the idea of printing electronic circuits on boards. In 1948, he was granted a patent[54] for a process to manufacture electric circuit components. The process involved printing the circuit pattern on a foil bonded to an impervious insulating backing using a medium (ink) that adhered to the foil and was capable of protecting the covered metal surfaces from chemical attack. He formed the pattern where the unprotected foil was etched away to the base. Today, this process is generally referred to as "print and etch." Eisler subsequently filed other patents pertaining to circuit boards.[55]

Following the end of World War II, it was disclosed in 1947 that large numbers of circuit boards had been manufactured for use in ordnance devices in the United States during the war. This revelation spurred a lot of interest in PCBs, resulting in the development of numerous processes for making them by the early 1950s. The most popular techniques for fabricating both PCB

---

[50] C. Ducas, "Electrical apparatus and method of manufacturing the same," U.S. Patent 1,563,731 (1925).

[51] ibid.

[52] W.S. DeForest, *Photoresist: Materials and Processes*, McGraw-Hill, New York, pp. 6–7 (1975).

[53] P. Eisler, *The Technology of Printed Circuits*, Heywood and Company, London, pp. 3–6 (1959).

[54] P. Eisler, "Manufacture of electric circuit components," U.S. Patent 2,441,960 (1948).

[55] P. Eisler, "Electric capacitor and method of making it, U.S. Patent 2,607,825 (1952); P. Eisler, "Apparatus for regeneration of etching media," U.S. Patent 2,748,071 (1956); P. Eisler, "Method of making printed circuits," U.S. Patent 2,747,977 (1958).

circuitry and interconnections during this early phase were mechanical fabrication techniques;[56] these gave way to plated-foil processes in the latter part of the decade.[57] The use of eyelets for interconnection was the preferred method until electroless processes for metallizing through holes were introduced.[58] Conductive ink patterns[59] applied by stenciling or screening also found reasonable acceptance and were quite often used in both plated and unplated conditions well until the late 1950s.

Plating as a means of forming a conductive path in a circuit pattern, from one side of the board to the other, evolved rather slowly during the decade despite the fact that methods for metallizing the through holes were available. As early as 1948, the two-step catalyzing process for electroless copper deposition for copper metallization of nonconductors was already in use.[60] Such a process involved first sensitizing glass or other nonconductive surfaces with a dilute stannous chloride solution that acted as a mordant, followed by a reduction of a precious metal such as gold from a subsequent acid-activating solution. These early electroless copper deposition processes were plagued by instability, cost, and reliability problems that mitigated deploying them in high-volume production. A reliable activating and plating process was eventually developed by Ralph Atkinson[61] in the mid-1950s, which was followed by a few others.[62] The year 1960 saw the introduction of the colloidal

---

[56] J.T. Beck, "Contacts for electrical circuits and methods for making same," U.S. Patent 2,823,286 (1958); J.B. Little and B.E. Phelps, "Printed circuit connection and method of making the same," U.S. Patent 2,902,629 (1959); O.I. Steigerwalt and H.I. Ashry, "Method of making a printed circuit," U.S. Patent 2,912,745 (1959); H.I. Ashry, O.I. Steigerwalt, and J.D. Heibel, "Method of making printed circuit panels," U.S. Patent 2,912,746 (1959); H.I. Ashry, O.I. Steigerwalt, and J.D. Heibel, "Connecting means for etched circuitry," U.S. Patent 2,912,747 (1959); W.D. Frazier, "Electrical mounting devices," U.S. Patent 2,915,678 (1959); R.B. Gray, "Method of making printed circuit panels," U.S. Patent 2,912,748 (1959); N.N. Berger, "Connecting means for etched circuitry," U.S. Patent 2,889,393 (1959).

[57] H. Shapiro, *Plating*, pp. 607–611 (June 1957).

[58] J.T. Beck, "Contacts for electrical circuits and methods for making same," U.S. Patent 2,823,286 (1958); W.D. Frazier, "Electrical mounting devices," U.S. Patent 2,915,678 (1959); N.N. Berger, "Connecting means for etched circuitry," U.S. Patent 2,889,393 (1959).

[59] D.K. Cardy, "Method of forming a printed circuit," U.S. Patent 2,864,156 (1958); R.C. Sanders, "Electrical components," U.S. Patent 2,876,392 (1959); M. Martin and C.S. Bojanowski, "Electrically conductive body and method of making same," U.S. Patent 2,861,911 (1958); S.K. Tally and V.F. Dahlgren, "Printed circuit baseboard," U.S. Patent 2,876,393 (1959); W.G. Lee, "Printed electric circuits," U.S. Patent 2,940,018 (1960).

[60] S. Wein, *Metal Finishing* **46**(8), 58 (August 1948).

[61] R.B. Atkinson, "Material and method for electroless deposition of metal," U.S. Patent 3,119,709 (1964).

[62] M.C. Agens, "Process of stabilizing autocatalytic copper plating solutions," U.S. Patent 2,938,805 (1960); R.M. Lukes, "Copper plating process and solution," U.S. Patent 2,996,408 (1961).

catalyst[63] by the Shipley Company. This catalyst quickly became the standard sensitizing-activating process for the PCB industry, and it remains so today.[64]

In the early and mid-1960s, multilayer boards arrived on the scene, particularly for military use, where cost was not an issue; the premium was on compactness, weight, and reliability. These boards evolved with the advances that had occurred in metallizing and photoresist processes.[65] Standards and specifications were prepared for conventional boards[66] in the early 1960s, and in the latter part of the decade for multilayers.[67] Due to migration problems, silver was permanently eliminated as a competitor to copper in metallizing circuit board holes.[68]

As was already implied, the early selection of offset lithographic printing as a means of applying etching resists was quite natural considering the nature of the inks that must be used for lithography and other available methods. The rubber image more easily conforms to uneven surfaces than to the metal lithographic plate, and very little moisture is transferred with the ink compared to direct lithographic printing. Copper is receptive to most printing inks and is often used in lithographic plates. In spite of all the good attributes of offset lithographic printing, several problems are associated with this technique: (1) The process results in a thin coating, which usually necessitates two or more passes. (2) Coatings are generally nonuniform, particularly when used for making large areas. (3) In some instances, misalignment and poor registration do occur.[69]

Another very important method for fabricating PCBs is screen printing. More PCBs have been made using screen printing than perhaps any other method. Screen printing can be used to form thick, inexpensive, chemically resistant images with reasonable accuracy and at an exceptionally high production rate. Once a screen is prepared, it can be used to print many boards in rapid succession, with the aid of automatic equipment. Images are formed on tensioned screens with naturally occurring colloid resists such as gelatin and albumin, or with synthetic polymers such as polyvinyl alcohol, polyacrylics, etc., which are rendered photosensitive by dichromate or other means. Silk, stainless steel, nylon, and polyester screens are typically used.

---

[63] "How to Bond to Copper while Metallizing Plastics," Shipley Company Data Sheet DS-124, January (1960); C.R. Shipley, Jr., "Method of electroless deposition on a substrate and catalyst solution therefor," U.S. Patent 3,011,920 (1961).

[64] W.S. DeForest, *Photoresist: Materials and Processes*, McGraw-Hill, New York, p. 9 (1975).

[65] ibid.

[66] "Military Specifications for Printed Wiring Boards," MIL-P-55110A; "Printed Wiring for Electronic Equipment," MIL-STD-273A.

[67] "Plated Through Hole Multilayer Circuit Boards," MIL-P-55640.

[68] D.E. Yost, in *Symposium on Printed Circuits*, University of Pennsylvania, Jan. 20, pp. 53–56 (1955); S.W. Chaikin, J. Janney, F.M. Church, and C.W. McClelland, "Silver migration and printed wiring," *Ind. Eng. Chem.* **51**(3), 299–304 (1959).

[69] W.S. DeForest, *Photoresist: Materials and Processes*, McGraw-Hill, New York, p. 12 (1975).

The processing that follows image formation can take the form of baking to harden the features before they are subjected to etching.[70]

In screen printing, the image is formed when a supply of ink is moved across the screen surface by the squeegee and forced through the screen opening onto the board. The screen is depressed against the board at a point of contact by the squeegee, and as the squeegee passes a given point, the screen lifts, leaving behind ink that is forced through the openings. The squeegee blade forms a seal with the binder/resin and forces the ink ahead of it, while wiping clean the smooth surface. Screen-printed resist images are compatible with most processes for either pattern plating or etching and are applicable to most single- and double-sided board processes.[71] As a general rule, screen printing of PCBs is done with negative resists.[72]

Resists as different as naturally occurring colloids (such as gelatin, albumin, and shellac) or synthetic polymers (such as polyvinyl alcohol, polyacrylics, polyvinyl pyrrolidone, polyvinyl butyral) that have been rendered photosensitive by dichromate or other means, as well as synthetic photosensitive polymers such as polyvinyl cinnamates (Kodak positive resists), diazonaphthoquinone/novolac, and solid-film resists such as Riston® (introduced by Dupont in 1968) have all been used in PCB fabrication. Some of these resists are still in use today.

5.4 The Transistor and Microelectronics Revolution

5.4.1 The invention of the transistor

In 1947, John Bardeen (1908–1991) and Walter Brattain (1902–1987),[73] both research scientists at Bell Telephone Laboratories, made a startling discovery while trying to understand the nature of the electrons at the interface between a metal and a semiconductor. They found that by making two point contacts very close to one another, they could make a three-terminal device: the first "point-contact" transistor[74] (Fig. 5.2). They made the two point contacts at the bottom of the triangular quartz crystal from two strips of gold foil separated by about 50 μm, and pressed them onto a semiconductor surface of germanium. With one gold contact biased forward (i.e., positive voltage with respect to the third terminal) and the other reverse biased, they observed transistor action. In other words, they observed the amplification of the input signal.[75]

---

[70] ibid.

[71] ibid.

[72] ibid, pp. 13–14.

[73] J. Bardeen and W.H. Brattain, "The transistor, a semiconductor triode," *Phys. Rev*. **74**, 230–231 (1948).

[74] Nobel Foundation 2007: http://nobelprize.org/educational_games/physics/transistor/history/index.html.

[75] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, Chapter 1 (2004).

Figure 5.2 The first point-contact transistor, which was invented at Bell Laboratories on 23 December 1947 (reprinted with permission from Alcatel-Lucent).

The researchers quickly made a few of these transistors and connected them to some other components to make an audio amplifier. After showing this audio amplifier to executives at Bell Telephone Company, who were very impressed that it didn't need time to "warm up" (like the heaters in vacuum tube circuits), they quickly realized the power of this new technology.[76] It should be mentioned that the original motivation for this invention was to find a way to amplify voices in telephones, a key product sold by the Bell Telephone Company—a goal that the inventors ably met.

This invention ignited a huge research effort in solid state electronics across the world. Bardeen and Brattain received the Nobel Prize in Physics in 1956,[77] together with William Shockley (1910–1989), "for their researches on semiconductors and their discovery of the transistor effect." Shockley had developed a so-called junction transistor in 1949,[78] which was built on thin slices of different types of semiconductor materials pressed together. The junction transistor was easier to understand theoretically, could be manufactured more reliably, and became the preferred semiconductor device, which ushered in the modern electronic era.[79]

---

[76] Nobel Foundation 2007: http://nobelprize.org/nobel_prizes/physics/laureates/1956/index.html.

[77] ibid.

[78] Nobel Foundation 2007: http://nobelprize.org/educational_games/physics/transistor/history/index.html.

[79] ibid.

Transistors eventually made their way into portable radios and other electronic devices, and are most prominently used today as building blocks of ICs. Remarkably, the invention of the point contact and *p–n* junction transistors, followed by Shockley's classic paper[80] on *p–n* junctions and bipolar transistors (also in 1949), spawned the microelectronics revolution, which has had a profound impact on our technological society over the latter half of the 20th century and into the 21st century in such diverse fields as automobiles, computers and communications, medicine, energy, and home entertainment, to mention but a few.

What, then, really is a transistor? The transistor is a three-terminal, solid state electronic device. It consists of a three-layer structure that comprises an *n*-type semiconductor layer sandwiched between *p*-type layers (*p–n–p* configuration) or a *p*-type layer between *n*-type layers (*n–p–n* configuration) (see *Chemistry and Lithography*, Second Edition, Vol. 3).

In such a three-terminal device, electric current or voltage between two of the terminals can be controlled by applying an electric current or voltage to the third terminal. In this way, a transistor regulates current and voltage flow, and acts as a switch or gate for electronic signals. A small change in the current or voltage of the inner semiconductor layer (which acts as the control electrode) can produce a large, rapid change in the current passing through the entire component.[81]

The three-terminal character of the transistor allows one to use it as an amplifier for electrical signals, such as the one in a radio. With the three-terminal transistor, one can also make an electric switch that can be controlled by another electrical switch. By cascading these series of switches (i.e., switches that control switches, which in turn control other switches, etc.), one can build up very complicated logic circuits,[82] resulting in switching times that are extremely short, on the order of nanoseconds today. Such logic chips are at the heart of the personal computer and many other electronic gadgets in use today.[83]

5.4.2 Limits of discrete transistors

Following the invention of the transistor, for many years, transistors were made as individual, discrete electronic components and were connected to other electronic components (resistors, capacitors, inductors, diodes, etc.) on printed circuit boards to make an electronic circuit. The transistor's small size and low power consumption made it an ideal candidate to replace the bulky

---

[80] W. Shockley, "The theory of *p–n* junction in semiconductors and *p–n* junction transistors," *Bell Syst. Tech. J.* **28**(3), 435–489 (1949).

[81] Nobel Foundation 2007: http://nobelprize.org/educational_games/physics/transistor/history/index.html.

[82] ibid.

[83] ibid.

vacuum tubes then used to amplify electrical signals and switch electrical currents. These beneficial attributes of transistors made it possible for them to be used in making ever-more-complex electronic circuits in place of vacuum tubes. However, it was not long before the limits of this approach of building circuits were reached. Circuits based on many individual, discrete transistors became too large, too cumbersome, and too difficult to assemble and produce for the simple reason that they contained too many electronic components. In addition, because transistor circuits were faster than vacuum tube circuits, noticeable problems due to time delays for electric signals to propagate a long distance in these large circuits were observed. To make the circuits even faster, one needed to pack the transistors closer and closer together, a task that became all the more daunting, if not outright impossible, to accomplish.[84]

5.4.3 The integrated circuit

In 1958 Jack Kilby (1923–2005)[85] at Texas Instruments and in 1959 Robert Noyce (1927–1990)[86] at Fairchild Camera independently invented the IC as a solution to the problem of building electronic circuits with large numbers of components. Instead of making transistors one-by-one, several transistors could be made at the same time with this technique, on the same piece of semiconductor. In addition to transistors, other electric components such as resistors, capacitors, and diodes could be made by the same process with the same set of materials.[87]

The rudimentary IC that Kilby made contained one bipolar transistor, three resistors, and one capacitor, all made in germanium and connected by wire bonding. For his invention of the IC, Jack Kilby was awarded the Nobel Prize in Physics in 2000.[88] The monolithic (*monolith* means "single stone") IC that Noyce proposed in 1959 was a flip-flop circuit containing six devices, in which the aluminum interconnection lines were obtained by etching an evaporated aluminum layer over the entire oxide surface using a lithographic technique.[89]

Since the invention of the IC, the pace of innovation has been scorching. The IC is now the foundation of microprocessors, memory chips, and myriad other kinds of semiconductor devices.

---

[84] Nobel Foundation 2007: http://nobelprize.org/educational_games/physics/transistor/history/index.html.

[85] J.S. Kilby, "Invention of the integrated circuit," *IEEE Trans. Electron Devices* **ED-23**, 648 (1976); J.S. Kilby, "Miniaturized electronic circuits," U.S. Patent 3,138,743 (1964).

[86] R.N. Noyce, "Semiconductor device-and-lead structure," U.S. Patent 2,981,877 (1961).

[87] Nobel Foundation 2007: http://nobelprize.org/educational_games/physics/transistor/history/index.html.

[88] ibid.

[89] R.N. Noyce, "Semiconductor device-and-lead structure," U.S. Patent 2,981,877 (1961).

5.4.4 Other notable developments in transistor technology

Since the invention of the point-contact transistor in 1947, the number and variety of semiconductor devices have witnessed a tremendous increase as advanced technology, new materials, and broadened comprehension have been applied to the creation of new devices. Without exception, all of these new devices[90] operate on the same basic principles of transistor action discovered in 1947. Some of the most notable of these are outlined below.

The year 1952 saw Ebers[91] develop the basic model for the thyristor, an extremely versatile switching device. Chapin and co-workers[92] developed the solar cell in 1954, using a silicon *p–n* junction. The solar cell is a device for harnessing energy from the sun because it can convert sunlight directly into electricity in a very environmentally benign manner. In 1957 Herbert Kroemer[93] proposed the heterojunction bipolar transistor (potentially one of the fastest semiconductor devices) to improve transistor performance. In 1958 Leo Esaki[94] observed negative-resistance characteristics in a heavily doped *p–n* junction, an observation that ultimately led to the discovery of the tunnel diode, which is important for ohmic contacts and carrier transport through thin layers.

The year 1960 saw the invention of the planar process by Jean Hoerni.[95] This process involves the formation of an oxide layer over a semiconductor surface. With the aid of lithography, windows are defined on the oxide surface, from which portions of the oxide layer are subsequently removed. Impurity atoms are diffused only through the exposed semiconductor surface, and *p–n* junctions are formed in the oxide window areas.

The MOSFET (metal-oxide semiconductor field-effect transistor), first reported by Dawon Kahng and John Atalla[96] in 1960, is the most important device for advanced ICs.[97] With the increase in complexity of ICs, device fabrication technology transitioned from *n*-channel MOSFET (nMOS) to complimentary MOSFET (CMOS), which employs two complementary transistors per gate—nMOS and pMOS (*p*-channel MOSFET)—to form

---

[90] A short review on the invention of these devices is provided in G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, Chapter 1 (2004). A good number of the classic papers on devices are collected in S.M. Sze, Ed., *Semiconductor Devices: Pioneering Papers*, World Science, Singapore (1991).

[91] J. Ebers, "Four terminal *p–n–p–n* transistors," *Proc. IRE* **40**, 1361 (1952).

[92] M.D. Chapin, C.S. Fuller, and G.L. Pearson, "A new silicon *p–n* junction photocell for converting solar radiation into electrical power," *J. Appl. Phys.* **25**, 676 (1954).

[93] H. Kroemer, "Theory of a wide-gap emitter for transistors," *Proc. IRE* **45**, 1535 (1957).

[94] L. Esaki, "New phenomenon in narrow germanium *p–n* junctions," *Phys. Rev.* **109**, 603 (1958).

[95] J.A. Hoerni, "Planar silicon transistors and diodes," *IRF Int. Electron Devices Meet.*, Washington, D.C. (1960).

[96] D. Kahng and M.M. Atalla, "Silicon-silicon dioxide surface device," presented at IRE Device Research Conference, Pittsburgh (1960).

[97] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, p. 4 (2004).

logic elements. The CMOS device was invented by Wanlass and Sah[98] in 1963. The CMOS technology offers the advantage that its logic elements draw significant current only during the transition from one state to another (e.g., from 0 to 1) and draw very little current between transitions, thus making it possible for power consumption to be minimized. CMOS technology is the dominant technology for advanced ICs in the world today.[99]

An extremely small MOSFET with a channel length of 15 nm was demonstrated by Yu and co-workers[100] at Advanced Micro Devices, Inc. (AMD) in 2001. This device may serve as the basis for the most advanced IC chips containing well over one trillion devices.[101]

The dynamic random access memory (DRAM), a two-element circuit, was invented by Dennard[102] in 1967. The DRAM cell contains one MOSFET and one charge-storage capacitor. The MOSFET functions as a switch to charge or discharge the capacitor. Although a DRAM is volatile and consumes relatively high power, it is expected that DRAMs will continue to be the first semiconductor memory of choice for nonportable electronic systems in the foreseeable future.[103]

The nonvolatile semiconductor memory (NVSM) was invented by Kahng and Sze[104] in 1967. The NVSM can retain its stored information when the power supply is switched off. Although very similar to the conventional MOSFET, it differs from the latter by the use of a "floating gate," which makes possible semipermanent charge storage. Attributes such as nonvolatility, high device density, low power consumption, and electrical rewritability (i.e., the stored charge can be removed by applying voltage to the control gate) have made the NVSM the dominant memory for portable electronic systems such as cellular phones, notebook computers, digital cameras, and smart cards.[105]

The ultimate floating-gate nonvolatile memory is the single-electron memory cell (SEMC), which requires only one electron for information storage. SEMCs can be fabricated by reducing the length of the floating gate

---

[98] F.M. Wanlass and C.T. Sah, "Nanowatt logics using field-effect metal oxide semiconductor triodes," *Tech. Dig. IEEE Int. Solid-State Circuit Conf.*, p. 32 (1963).

[99] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, p. 8 (2004).

[100] B. Yu, H. Wang, A. Joshi, et al., "15 nm gate length planar CMOS transistor," *IEEE IEDM Technical Digest*, Washington, D.C., p. 937 (2001).

[101] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, p. 4 (2004).

[102] R.M. Dennard, "Field effect transistor memory," U.S. Patent 3,387,286 (1968).

[103] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, p. 8 (2004).

[104] D. Kahng and S.M. Sze, "A floating gate and its application to memory devices," *Bell Syst. Tech. J.* **46**, 1283 (1967).

[105] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, p. 5 (2004).

to ultrasmall dimensions on the order of 10 nm. At this dimension, when an electron enters the floating gate, the potential of the gate is altered such that it prevents the entry of another electron. In 1954 Yano et al.[106] were the first to demonstrate the operation of a SEMC at room temperature. The operation of a SEMC may one day serve as the basis for the most advanced semiconductor memories, which may contain well over one trillion bits.[107]

With a view to improving device performance, Kerwin and co-workers[108] proposed the polysilicon self-aligned gate process in 1969. This process not only improved device reliability, it also reduced parasitic capacitances. Furthermore, in 1969 Manasevit and Simpson[109] developed the metal-organic chemical vapor deposition (MOCVD) process, which found widespread adoption in the fabrication of compound semiconductors such as GaAs.

In order to support the ever-shrinking device dimensions, a dry etching technique was developed to replace wet chemical etching for high-fidelity pattern transfer by Irving and co-workers[110] in 1971 in a process that involved the use of a CF_4/O_2 gas mixture to etch silicon wafers. Molecular beam epitaxy, developed also in 1971 by Cho,[111] offers the advantage of near-perfect vertical control of composition and doping down to atomic dimensions, and is now applied extensively in the fabrication of photonic devices and quantum effect devices.[112]

The year 1971 saw the invention of the microprocessor by Hoff and co-workers,[113] who put the entire central processing unit (CPU) of a simple computer on one chip. A four-bit microprocessor (Intel 4004) with a chip size of 3 mm × 4 mm and containing 2300 MOSFETs, it was fabricated with a *p*-channel polysilicon gate process using an 8-μm design rule. Remarkably, this microprocessor performed as well as its much larger counterparts, the IBM mainframe computers of the early 1960s, each of which needed a CPU

---

[106] K. Yano, T. Ishii, T. Hashimoto, et al., "Room-temperature single-electron memory," *IEEE Trans. Electron. Devices* **41**(9), 1628–1638 (1994).

[107] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, p. 5 (2004).

[108] R.E. Kerwin, D.L. Klein, and J.C. Sarace, "Method for making MIS structure," U.S. Patent 3,475,234 (1969).

[109] H.M. Manasevit and W.I. Simpson, "The use of metal-organic in the preparation of semiconductor materials. I. Epitaxial gallium-V compounds," *J. Electrochem. Soc.* **116**, 1725 (1969).

[110] S.M. Irving, K.E. Lemons, and G.E. Bobos, "Gas plasma vapor etching process," U.S. Patent 3,615,956 (1971).

[111] A.Y. Cho, "Film deposition by molecular beam technique," *J. Vac. Sci. Technol.* **8**, S31 (1971).

[112] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, Chapter 1 (2004).

[113] The inventors of the microprocessor include M.E. Hoff, F. Fagin, S. Mazor, and M. Shima. For a profile of M.E. Hoff, see R. Slater, *Portraits in Silicon*, MIT Press, Cambridge, p. 175 (1987).

the size of a large desk. The invention of the microprocessor was a major breakthrough for the semiconductor industry. Currently, microprocessors constitute the most advanced segment of the IC industry.[114] Intel Corporation and Advanced Micro Devices, Inc. are the two companies that make more than 95% of all of the microprocessors in the world.

Since the early 1980s, many new technologies were developed to support the requirements of ever-shrinking minimum feature lengths. The three most significant technologies to be developed during this period include trench isolation, chemical mechanical polishing (CMP), and copper interconnect. Trench isolation was developed by Rung and co-workers[115] in 1982 to isolate CMOS devices. This technique eventually replaced all other prior isolation methods for fine-featured ICs. In 1989 Davarik and co-workers[116] developed the CMP technique for global planarization of interlayer dielectrics, a key enabling process technique for multilevel metallization. Electromigration, the transport of metallic ions through a conductor due to passage of an electrical current, is a widely known device failure mechanism, particularly at submicron dimensions. Although aluminum has been in use since the early 1960s as an interconnect material, it suffers from electromigration at high electrical current; it is also too resistive. Paraszczak and co-workers[117] introduced copper interconnects in 1993 to replace aluminum for minimum feature lengths approaching 100 nm and smaller.[118]

5.4.5 Overall device technology trends

Device technology trends over the last 70-odd years are illustrated in Fig. 5.3, which shows the growth curves for different IC devices. The bipolar transistor drove the IC device technology at the beginning of the modern electronic era (1950–1970). Between 1970 and 1990, the DRAM and the microprocessor based on MOS devices were the technology drivers, largely due to the rapid growth of personal computers and advanced electronic circuits. The nonvolatile semiconductor memory has been the technology driver since 2000, due mainly to the rapid growth of portable electronic systems.[119]

---

[114] G.S. May and S.M. Sze, *Fundamentals of Semiconductor* Fabrication, John Wiley & Sons, Hoboken, New Jersey, p. 8 (2004).

[115] R. Rung, H. Momose, and Y. Nagakubo, "Deep trench isolated CMOS devices," *Tech. Dig. IEEE Int. Electron Devices Meet.*, p. 237 (1982).

[116] B. Davarik, C.W. Koburger, R. Schulz, et al., "A new planarization technique using a combination of RIE and chemical mechanical polish (CMP)," *Tech. Dig. IEEE Int. Electron Devices Meet.*, Washington, D.C., p. 61 (1989).

[117] J. Paraszczak, D. Edelstein, S. Cohen, et al., "High-performance dielectrics and processes for ULSI interconnection technologies," *Tech Dig. IEEE Int. Electron Devices Meet.*, Washington, D.C., p. 261 (1993).

[118] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, p. 8 (2004).

[119] ibid.

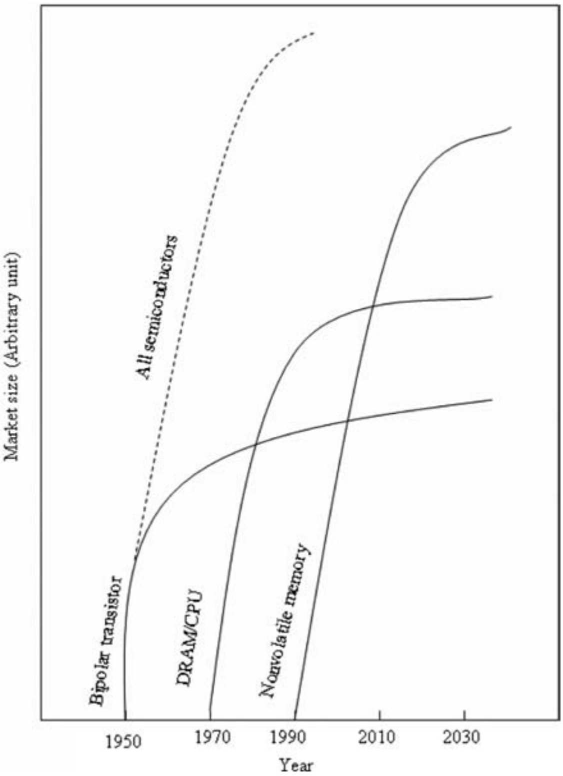


Figure 5.3 Growth curves for different IC devices that have been technology drivers over the years (plot generated in the manner of Masuoka[120]).

The fact that demand for memory has continued to be an important driving force within the IC industry over the last 50 years and is expected to remain so for the foreseeable future is a testament to the heightened sensitivity of these devices to yield factors. Yield depends significantly on the surface area of the device and is often negatively impacted by contaminants—either from particulates outside the process equipment or generated by the process equipment, or redeposits of process materials during processing. For this reason, yield is heavily affected by packing density, which in turn is a most apparent restriction in memory devices, making memory devices the technology drivers in the IC industry.[121] Memory devices lead the industry in both device density and linewidth.[122] Packing density is determined by linewidth, so any increases in memory size require smaller linewidths or larger

---

[120] F. Masuoka, "Flash memory technology," *Proc. Int. Electron Devices Mater. Symp.* **83**, Hsinchu, Taiwan (1996).

[121] VLSI Research Inc., "Wafer Fab Fabrication Equipment," Report 2233-11P (1989).

[122] ibid.

die size, or both.[123] These requirements call for lithographic exposure tools such as high-NA scanners and steppers that can reliably expose larger die sizes with a high degree of fidelity and without significant yield losses.[124]

Since the beginning of the microelectronics era, the smallest linewidth (or minimum feature length) of an IC has shrunk at a rate of about 13% per year, motivated no doubt by the inherent advantages of device miniaturization. Device miniaturization leads to increased reduced unit cost per circuit function. For example, the cost per bit of memory chips has decreased by one-half every two years for successive generations of DRAM circuits. Also, as device dimensions shrink, the intrinsic switching time also decreases. Device speed has in fact improved by four orders of magnitude since 1959. Higher speeds lead to increased performance. Furthermore, as devices become smaller, they consume less power. The energy dissipated per logic gate has decreased by a factor of more than one million since 1959.[125]

In the 46 years since the microprocessor was invented, microprocessor computational power has increased by a factor of 2 every 18 months since 1970,[126] so much so that computer clock speeds have increased by a factor of one thousand (from a few megahertz to a few gigahertz) and memory capacity has grown even more (from kilobytes to gigabytes). These facts are illustrated in Fig. 1.1 for microprocessors and graphic chips made by companies such as Intel Corporation, Advanced Micro Devices, Inc., International Business Machines (IBM), Apple Inc., etc. Even in the face of this relentless drive toward miniaturization of computer chips over the last three decades, one aspect of computer hardware has remained largely unchanged. Until recently, most general-purpose computers were equipped with only one CPU, where programs are executed and calculations are performed. Now a major shift has been implemented. Many of the latest general-purpose computers from AMD and Intel are equipped with multiple processor chips that are fabricated on a single slab of silicon. These multiple processors are meant to share the work of computation and multiply the machine's power.[127] Examples of such processor systems include AMD's Zen$^+$ microprocessor chip with up to 32 cores and containing 4.8 billion transistors (see Fig. 5.4).

A natural question to ask is this: Why build chips with multiple processors? Why not just keep increasing the speed of the single CPU? If the latter option were feasible, the chipmakers would have adopted it. Instead,

---

[123] ibid.

[124] ibid.

[125] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, p. 9 (2004).

[126] See Fig. 1.1 in Chapter 1.

[127] B. Hayes, "Computing in a parallel universe," *American Scientist*, pp. 476–480, November–December (2007).

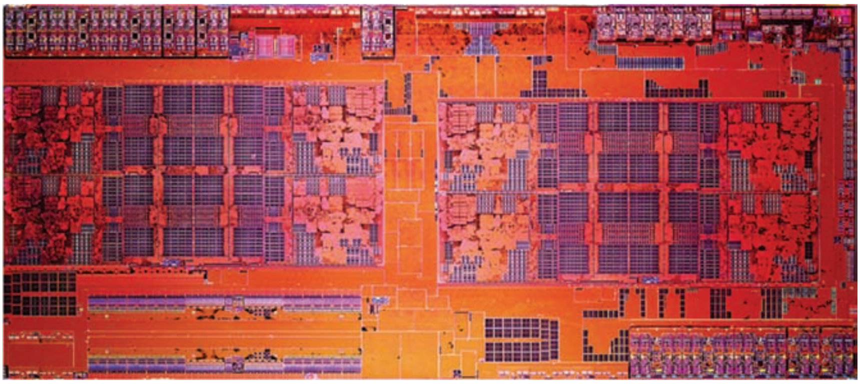


Figure 5.4 AMD's Zen+™ microprocessor chip with up to 32 cores. Manufactured with a 12-nm technology node process, it has critical dimension of 12 nm, 12 metal layers, 4.8 billion transistors, and a silicon die size of 212.97 mm^2. The 32 cores are separate processors that act in parallel.[128]

they are turning to multicore systems only because the path to higher clock speeds seems to be very difficult, at least for now.[129]

The causes of this impasse derive from the peculiar physical and economic laws that govern the design of ICs.[130] The most celebrated of these laws stipulates that, as transistors or other components are made smaller and packed more densely on the surface of a silicon chip, the cost of producing the chip remains nearly constant, so much so that the number of transistors on a state-of-the-art chip doubles every year or two, as already alluded to. In effect, the cost per transistor steadily declines over time. This extraordinary fact is the basis of Moore's law,[131] formulated in 1965 by Gordon E. Moore, one of the founders of Intel Corporation. The law reflects the fantastic progression of circuit fabrication, which has unleashed previously unimagined computational power.[132]

Less famous than Moore's law but equally influential are several "scaling laws," first formulated in 1974 by Robert H. Dennard and co-workers[133] at IBM, who asked: When the size of a transistor is reduced, how should the other factors such as voltages and currents that control its operation be adjusted? And what effect will the changes have on performance? Dennard and co-workers found that voltage and current should be proportional to the

---

[128] https://en.wikichip.org/wiki/amd/microarchitectures/zen%2B.

[129] ibid.

[130] ibid.

[131] See, e.g., D.C. Brock, Ed., *Understanding Moore's Law: Four Decades of Innovation*, Chemical Heritage Press, Philadelphia (2006).

[132] See Fig. 1.1 in Chapter 1.

[133] R. Dennard, F. Gaensslen, H.-N. Yu, V.L. Rideout, E. Bassous, and M. LeBlanc, "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE J. Solid State Circuits* **SC-9**(5), 256–268 (1974).

linear dimensions of the device, implying that power consumption (the product of voltage and current) will be proportional to the area of the transistor. This was a very significant discovery because it means that, even as the number of devices on a chip increases, the total power density remains constant.[134]

The conclusion drawn by Dennard and co-workers about performance was even more startling, but in a good way. Their scaling laws show that switching delay—the time it takes a transistor to go from the conducting to the nonconducting state, or vice versa—is proportional to size. This implies that as circuits shrink, they can be operated at higher speed. The switching delay is a crucial measure of transistor performance. [135]

Taken together, these observations suggest that the laws of nature seem to work in our favor in microelectronics. Moore's law and Dennard's scaling rules promise circuits that gain in both speed and capability, while cost and power consumption remain constant. From this favorable circumstance arises the entire bonanza of modern microelectronics.[136] However, as with every bonanza, there are limits and tradeoffs.

Over the last decade, chip designers have struggled with such limits and tradeoffs, best exemplified by the twin problems of memory bottleneck and power management. Although CPUs are 1000 times faster today than they were 30 years ago, memory speed has increased by a factor of only 10 or so within the same period. As far back as the 1980s, reading a bit from main memory took a few hundred nanoseconds, which corresponded to the time needed to execute a single instruction in a CPU. Then the memory and the processor cycles were well matched. Currently, a processor could execute one hundred instructions in the time it takes to obtain data from memory.[137]

A strategy employed by chip designers for tackling the memory bottleneck is to transfer data in large blocks rather than single bits of bytes. This strategy improves throughput (bits per second), but not latency (the delay before the first bit arrives). To tackle the latency problem, computers are now equipped with an elaborate hierarchy of cache memories, which surround the processor core. Data instructions that will be executed immediately are held in the innermost, first-level cache, which has only a small capacity but is built for very high speed. For slightly less urgent information, there is a second-level cache, which is larger but a little slower than the first-level cache. Some processor systems even have a third-level cache.[138]

---

[134] B. Hayes, "Computing in a parallel universe," *American Scientist*, pp. 476–480, November–December (2007).

[135] ibid.

[136] ibid.

[137] ibid.

[138] ibid.

Over-reliance on cache memory presents some difficulties in predicting which data instructions a program will call for next, a wrong prediction being rather catastrophic. Devoting a large fraction of the silicon of processor chips to caches and the logic circuits that control them means that fewer processors will be made from a given silicon wafer on a unit basis. In addition, the lithographic patterning of such processors is fraught with many difficult challenges in terms of defectivity, overlay control, and critical dimension control. As the disparity between memory and CPU speed widens even farther, at some point, all of the benefits of any further increase in processor speed will be used up by the demand for more cache.[139]

The second problem that now plagues chip designers is a power crisis. The scaling laws of Dennard and co-workers predict that power density should remain constant, even as the number of transistors and their switching speed increase. However, for that prediction to hold, voltages have to be reduced in proportion to the linear dimensions of the transistor. In reality, chip manufacturers have not been able to lower operating voltages that steeply.[140]

Historically, for each successive generation of processor chips, the linear dimensions (such as gate width of a transistor) has been scaled by a factor of 0.7, which yields an area reduction of 0.5. In other words, chip density has doubled for every successive generation. Moreover, the switching delay (the reciprocal of processing speed) is proportional to the linear size. If the operating voltage could also be lowered by a factor of 0.7, a transistor's power consumption should be proportional to this surface area, and the power density of the entire chip should therefore remain constant. However, in reality, the scaling factor for voltages has been 0.85 rather than 0.7, with the result that power density has been rising steadily with each new generation of chips, so that power and heat have now become limiting factors, so much so that computer machines are now equipped with fans to cool them. [141]

Going forward, even the 0.85 voltage reduction strategy is problematic because, as voltage is lowered, transistors become leaky, like valves that cannot completely shut off. Presently, the leakage current accounts for roughly one-third of total power consumption; with further reductions in voltage, leakage could become even more unmanageable. On the other hand, without reducing the voltage, the clock rate cannot be increased.[142]

Some have taken to viewing these problems with memory latency and power density as signaling the end of Moore's law—a stance that might not be entirely accurate. After all, chip manufacturers can still pack more transistors onto a chip and manufacture it for a roughly constant cost. In fact, the

---

[139] ibid.

[140] ibid.

[141] ibid.

[142] ibid.

Semiconductor Industry Association's "Roadmap"[143] calls for increasing the number of transistors on a microprocessor from a few hundred million today to more than 12 billion by 2020. What happens to be ending, or at least slowing down dramatically, is the scaling law that allows processor speeds to keep climbing. The chip industry can still make smaller circuits, thanks to lithography, but not faster ones. And therein lies the reason behind the strategy of multicore chips, in which lots of little processors work in parallel to perform ever-more-complex computational tasks.[144] According to Mark T. Bohr of Intel Corporation, "achieving very high operating frequencies is no longer the prime target for microprocessors. Instead, the goal has shifted to delivering higher performance combined with lower power. Power efficiency is the main scaling goal for chips used in small handheld devices and in large data centers."[145]

An illustration of this new strategy of having multiple processor on the same slab of silicon working in parallel is provided by AMD's Zen^{+}™ microprocessor chip with up to 32 cores (see Fig. 5.4). Released in 2018, this 32-core processor has 4.8 billion transistors, each one of which has three electrical contacts, reaching a total of 14.4 billion contacts per chip. In AMD's fabrication facilities, these electrical contact holes are made at the rate of roughly 35 billion per second with the aid of lithography. Imagine drilling a hole every three seconds, perhaps with the aid of an electrical drill. It would take more than 3000 years to drill 35 billion electrical contact holes, at the end of which one would not have produced the contacts for even one chip.[146] Again, this illustrates the speed and efficiency of semiconductor lithography. But how did the ancient art of lithography make the transition from the printing industry, where it was used for reproducing artistic impressions, blueprints, newspapers, and the like, into a technique for patterning of ICs?

5.5 Semiconductor Lithography

The beginning of semiconductor lithography dates back to 1957, when Jules Andruss, using etch-resistant photoresist, successfully transferred photomask patterns of a device into semiconducting substrate.[147] Since then, lithography has evolved to become a key technology in the semiconductor industry. The phenomenal growth of the semiconductor industry derives, to a significant extent, from improvements in lithographic technology. Lithography also

---

[143] 2015 International Technology Roadmap for Semiconductors (ITRS): http://www.itrs.net/Links/2005ITRS/Home2005.htm.

[144] B. Hayes, "Computing in a parallel universe," *American Scientist*, pp. 476–480, November–December (2007).

[145] M.T. Bohr, Senior Fellow, Intel Corporation, personal communication (2017).

[146] U. Okoroanyanwu, Thought experiment, unpublished results (2019).

[147] J. Andrus, "Fabrication of semiconductor devices," U.S. Patent 3,122,817 (1964).

constitutes a significant economic factor, currently representing well over 35% of IC manufacturing costs.[148]

Borrowed from the printing industry, semiconductor lithography is ideally suited for the fabrication of ICs, which requires the patterning of many layers of thin-film materials used in building up the device on a semiconductor substrate. Viewed another way, semiconductor lithography is basically an extension of photography. In semiconductor lithography, one first fabricates the equivalent of a photographic negative/positive containing the pattern required for some part of the microchip's circuitry. This negative/positive, called the mask, is then used to copy the pattern onto the semiconductor substrate of an IC. The process comprises two parts: the preparation of the mask (a one-time event that can be slow and expensive) and the transfer of the mask pattern to the semiconductor substrate. In effect, the semiconductor lithographic process involves transferring a circuit pattern—as might, for example, be contained in a mask—into a radiation-sensitive film (resist) and subsequently replicating that pattern in an underlying thin conductor or dielectric film of the semiconductor substrate.[149]

To make a mask for a part of an IC, a manufacturer first designs the pattern of the circuitry on a conveniently large scale and converts it into a pattern of opaque metallic film (usually chromium) on a transparent plate (usually glass or quartz). Semiconductor lithography is then used to reduce the size of the pattern in a process that resembles that used in photography. This process is shown schematically in Fig. 5.5 and comprises a number of steps. First, an ultrahigh-purity wafer substrate, usually silicon wafer (the "stone"), is coated with radiation-sensitive resist (typically made of organic polymer or inorganic compounds), baked to evaporate the casting solvent, and exposed to radiation through a mask. The exposure process involves directing a beam of radiation (typically UV light from a mecury arc lamp or deep-UV light from exciplex and excimer lasers or electrons, or even ions) to a mask, creating a transmitted or reflected image, which then is made to pass through a set of lenses or mirrors, as the case may be, which in turn focuses the image onto the resist on the surface of a semiconductor substrate such as a silicon wafer.[150]

Following exposure, the exposed resist film may be baked again, causing the catalytic photoacid[151] generated from the photoactive compound contained in the photoresist to act on the polymers. Next, the exposed resist

---

[148] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, Chapter 1 (2004).

[149] G.M. Whitesides and J.C. Love, "The art of building small," *Scientific American Reports* **17**(3), 13–21 (2007); W.M. Moreau, *Semiconductor Lithography: Principles, Practices, and Materials*, Plenum, New York (1988); L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C. (1994).

[150] ibid.

[151] Applicable only to chemically amplified resists.

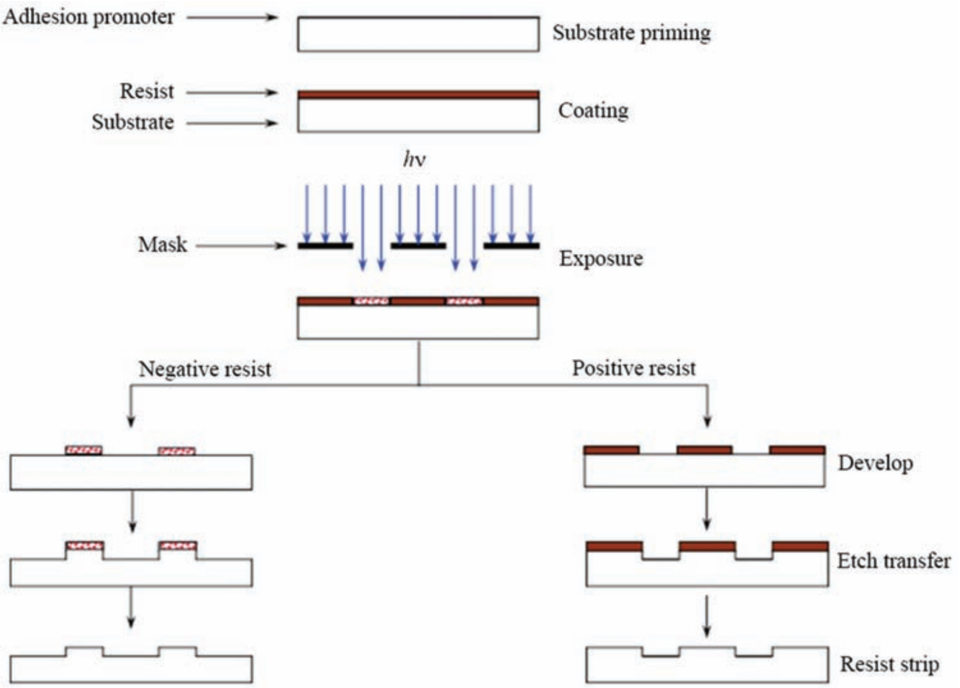


Figure 5.5 Schematic representation of the photolithographic process sequence.

film is developed in a solvent to produce a 3D relief replica of the mask pattern. Finally, the image in the resist is etched into the underlying substrate by liquid-based substrate etching systems or a variety of plasma (dry) etch processes.[152] And at the end of the process, the remaining resist is stripped off, leaving the patterned features on the semiconductor substrate, which then undergoes subsequent processing to transform it into an actual IC device.

The two fundamental relationships describing optical lithographic resolution R and depth of focus (DOF) are given by

$$R = k_1 \frac{\lambda}{\mathrm{NA}}, \tag{5.1}$$

$$\mathrm{DOF} = k_2 \frac{\lambda}{(\mathrm{NA})^2}, \tag{5.2}$$

where λ is the wavelength of the imaging radiation, and NA is the numerical aperture of the imaging system. The parameters k_1 and k_2 are empirically

---

[152] G.M. Whitesides and J.C. Love, "The art of building small," *Scientific American Reports* **17**(3), 13–21 (2007); W.M. Moreau, *Semiconductor Lithography: Principles, Practices, and Materials*, Plenum, New York (1988); L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C. (1994).

determined and take on values that depend on the critical dimension (CD) tolerance and the size of an acceptable IC manufacturing process window. Equation (5.1) demonstrates that improvements in resolution can be achieved by decreasing wavelength and/or by increasing NA. Equation (5.2), which describes the depth of focus of a feature of resolution R, shows that smaller features have less depth of focus.

Photoresists function by radiation-induced alteration of the materials' solubility. There are two basic classes of resist materials: negative and positive resists (see Fig. 5.5). Negative resists become less soluble upon exposure to radiation; i.e., the unexposed areas can be selectively removed by treatment with an appropriate developer solvent. Positive resists selectively undergo an increase in solubility upon exposure, enabling the exposed regions to be selectively removed in the developer. Both types of resists are formulated from polymers designed to have physical and chemical properties consistent with semiconductor processing. Negative resists that undergo radiation-induced crosslinking often suffer from solvent-induced swelling, which imposes a limit on their resolution. Positive resists do not usually suffer from swelling problems and therefore have dominated the fabrication of leading-edge devices. Details of these processes are described by Moreau[153] and Thompson et al.[154] In the following section, we briefly review the various semiconductor lithographic techniques that are either currently being used to manufacture IC devices or are under development for potential future lithographic applications.

5.6 Advanced Lithographic Patterning Techniques and Imaging Mechanisms

Figure 5.6 shows various advanced lithographic patterning techniques that are either currently being used to fabricate IC devices, or are currently being actively researched. We will briefly review the main attributes of these techniques as a way of providing background information that will be required later in our analysis.

5.6.1 Optical lithography

Optical lithography (also called photolithography), which uses visible and ultraviolet radiation (436 nm to 157 nm) to transfer the pattern from the mask to a photosensitive resist, is the dominant technology in use today for the fabrication of ICs. The technology was adopted during the early days of planar transistor development and is still the main technology of choice.[155]

---

[153] W.M. Moreau, *Semiconductor Lithography: Principles, Practices, and Materials*, Plenum, New York (1988).

[154] L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C. (1994).

[155] W.S. DeForest, *Photoresist Materials and Processes*, McGraw-Hill, New York (1975).

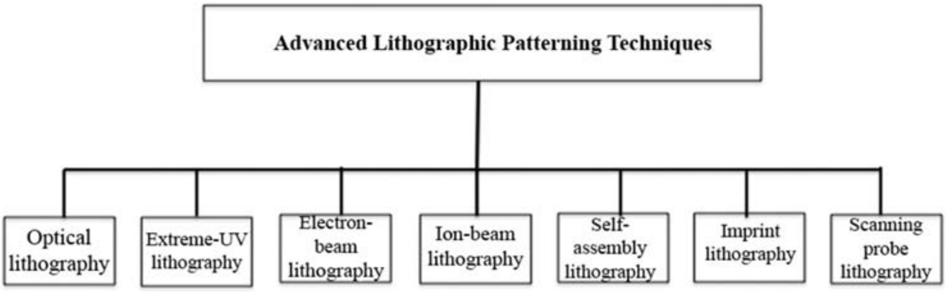


Figure 5.6 Advanced lithographic patterning techniques.

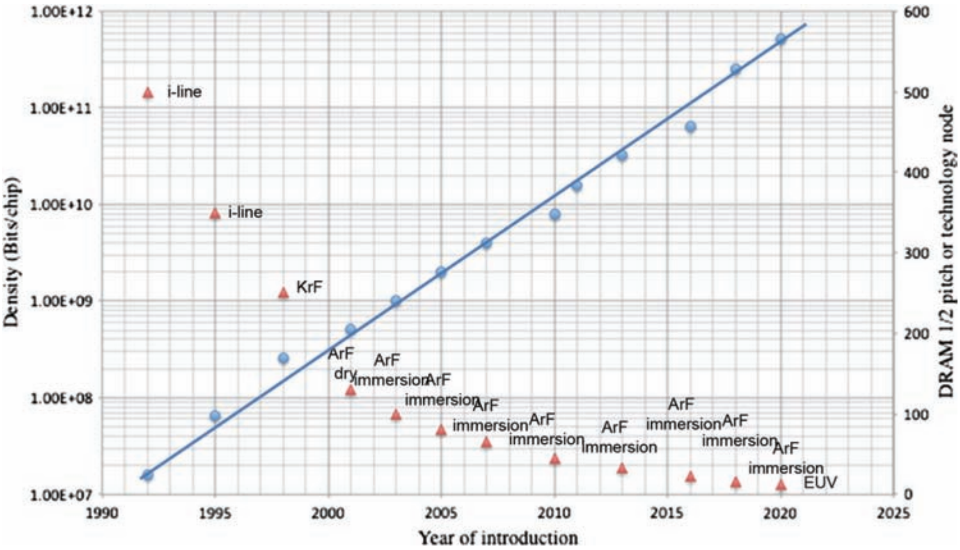


Figure 5.7 Evolution of semiconductor lithography as reflected in the exponential increase of dynamic random access memory (DRAM) density versus year of introduction, based on the 1999, 2011, and 2012 Semiconductor Industry Association Roadmap for Semiconductors.

Since its introduction, photolithography has undergone many innovations and evolutions,[156] as shown in Fig. 5.7 and Table 5.1.

The oldest photolithographic technique is contact printing,[157] whereby the mask is placed close to the wafer surface and is visually aligned to the previous pattern on the wafer in a process called registration and alignment. The mask is then pressed into hard contact with the resist-coated wafer, following which

---

[156] An excellent account on developments in optical lithography between 1967 and 2007 is provided in J.H. Bruning, "Optical lithography: 40 years and holding," *Proc. SPIE* **6520**, 652004 (2007) [doi: 10.1117/12.720631].

[157] W.M. Moreau, *Semiconductor Lithography: Principles, Practices, and Materials, Plenum*, New York (1988); W.S. DeForest, *Photoresist Materials and Processes*, McGraw-Hill, New York (1975).

Table 5.1 Evolution of lithography from 1967 to 2015 (S&R: step and repeat; DNQ: diazonaphthoquinone; PBOCST: poly(butoxycarbonate styrene).

Year of Introduction	Lithography Type	Resist	Device	Minimum Feature (μm)
1967	Contact printing	Cyclized rubber	256-bit DRAM	15–20
1971	Near-contact printing	Cyclized rubber	1-Kbit DRAM	8–12
1971	Near-contact printing	DNQ/Novolac	1-Kbit DRAM	8–12
1974	Near-contact printing	DNQ/Novolac	4-Kbit DRAM	6
1977	436–360 nm, 1:1 projection printing	DNQ/Novolac	16-Kbit DRAM	4
1980	436–360 nm, 5–10× S&R	DNQ/Novolac	64-Kbit DRAM	2
1982	436 nm (g-line), 5× S&R	DNQ/Novloac	256-Kbit DRAM	1.4
1985	g-line, 5× S&R	DNQ/Novolac	1-Mbit DRAM	1
1988	365 nm (i-line), 5× S&R	DNQ/Novolac	4-Mbit DRAM	0.7
1992	365 nm (i-line), 5× S&R	DNQ/Novolac	16-Mbit DRAM	0.5
1995	365 nm (i-line), 5× S&R	DNQ/Novolac	64-Mbit DRAM	0.35
1998	DUV (248 nm, KrF laser), 4×; i-line), 5× S&R;	PBOCST	256-Mbit DRAM	0.25
2001	DUV (248 nm/KrF laser), 4× S&R	PBOCST	1-Gbit DRAM	0.18
2005	DUV (193 nm/ArF laser), 4× scanner	Acrylate & Alicyclics	4-Gbit DRAM	0.13
2007	DUV (193 nm/ArF laser), 4× scanner	Acrylate & Alicyclics	16-Gbit DRAM	0.1
2010	DUV (immersion 193 nm/ArF laser), 4× scanner	Acrylate & Alicyclics	64-Gbit DRAM	0.07
2015	DUV (immersion 193 nm/ArF laser), 4× scanner	Acrylate & Alicyclics	256-Gbit DRAM	0.05

the resist is flood exposed through the mask with ultraviolet light.[158] The theoretical resolution of contact printing is given by[159]

$$\mathrm{CD} = \frac{3}{2}\sqrt{\frac{\lambda z}{2}}, \tag{5.3}$$

---

[158] L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C., p. 9 (1994).

[159] M.J. Bowden, "The lithographic process: The physics," in L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C., Section 2.1.1, pp. 22–23 (1994).

where CD is the critical dimension (provided that the feature is a grating with a period of twice the CD), λ is the exposure wavelength, and z is the thickness of the photoresist.

Unfortunately, contact printing was plagued by a number of practical and technical difficulties. Bringing the mask and the wafer into direct contact causes damage to the mask and creates defects that are reproduced in subsequent exposures. In addition, the nonflatness of the wafer and mask results in alignment error, which further limits the resolution capability of the technique.[160]

Practical and technical difficulties associated with contact printing led to a modification of the technique, such that the mask and the wafer were separated by a small, accurately controlled gap of 10–25 μm during exposure. Introduced in the early 1970s, this technique, known as proximity printing,[161] minimizes mask and wafer damage caused by contact, but at the expense of resolution. This gap, however, limits resolution because of diffraction. The theoretical resolution of proximity printing is given by[162]

$$\mathrm{CD} = \frac{3}{2}\sqrt{\lambda\left(g + \frac{z}{2}\right)}, \tag{5.4}$$

where g is the separation distance between the mask and the resist surface.

Around 1973, a technique known as projection printing, in which the mask image is projected onto the wafer by means of a 1:1 reflective or refractive optical system, was developed.[163] Given that the mask was no longer in direct contact with the wafer, the technique offered greatly increased mask lifetimes and marked reduction in defect density. The early projection systems were designed to image the full-wafer exposure without demagnification. Being 1× projection systems, they were amenable to simple optical designs that permitted the use of a relatively broad light spectrum. With increasing wafer size, projection printing was incapable of imaging all features onto a wafer simultaneously. The image field therefore had to be partitioned.[164]

---

[160] A.K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Bellingham, Washington, p. 14 (2001) [doi: 10.1117/3.401208].

[161] D.O. Massetti, M.A. Hockey, D.L. McFarland, "Evaluation of deep-UV proximity mode printing," *Proc. SPIE* **221**, 32–38 (1980) [doi: 10.1117/12.958621].

[162] M.J. Bowden, "The lithographic process: The physics," in L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C., Chapter 2, p. 24 (1994).

[163] D.A. Markle, "The future and potential of optical scanning system," *Solid State Technol.* **27**(9), 159–166 (1984); M.C. King, "Future development for 1:1 projection photolithography," *IEEE Trans. Electron Devices* **ED-26**, 711 (1979).

[164] A.K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Bellingham, Washington, p. 14 (2001) [doi: 10.1117/3.401208]; M.J. Bowden, "An introduction to microlithography," in L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C., Chapter 1, pp. 9–10 (1994).

One of the first partitioning approaches was based on scanning projection printing,[165] in which the reticle was projected onto the wafer by scanning a ring-shaped image field, consisting of two concentric spheres with a radius ratio such that the image field was well corrected for aberrations.[166] The exclusive use of mirrors during this time permitted the use of relatively broadband illumination since chromatic aberration was not a concern. This technique was introduced in the 1970s.

Due to the continuous increase in wafer size and reduction of feature dimension, full-wafer printing without demagnification became increasingly difficult to implement because of mask dimension control and pattern placement difficulties. These difficulties were alleviated by the introduction of reduction projection printing toward the end of the 1970s.[167] Known as step-and-repeat photolithography, this reduction projection printing technique involves exposing a smaller area (on the order of 1 to 2 cm^2) and stepping this pattern over a large-diameter wafer. The technique improves resolution and alignment accuracy compared to full-wafer printing.[168] These step-and-repeat exposure systems, or steppers, were designed with either 1:1 or reduction optics (10×, 5×, and 4×).[169]

Despite the strategy of image field reduction, design of reduction projection optics still presented difficult technical challenges, exacerbated by the drive toward higher-numerical-aperture lenses for improved resolution and the need for aberration correction over the (still) substantial field area. For exposure wavelengths below 300 nm, fused silica and calcium fluoride were and still remain the only practical lens-making material, making chromatic aberration correction impossible in refractive exposure systems. Although reflective or catadioptric[170] systems could be used, the required surface finish of the aspheric mirrors in such systems is extremely difficult to achieve, but is easily achieved in lenses used in immersion ArF (193 nm) lithography.[171]

---

[165] R.M. Scott, "Annular field optical imaging system," U.S. Patent 3821763 (1974); A. Offner and C.T. Darien, "Unit power imaging catoptric anastigmat," U.S. Patent 3748015 (1973).

[166] D.A. Markle, "The future and potential of optical scanning system," *Solid State Technol.* **27**(9), 159–166 (1984); M.C. King, "Future development for 1:1 projection photolithography," *IEEE Trans. Electron Devices* **ED-26**, 711 (1979).

[167] V. Pol, "High resolution optical lithography: A deep ultraviolet laser-based wafer stepper," *Solid State Technol.* **30**(1), 71–76 (1987).

[168] L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C., p. 10 (1994).

[169] ibid.; A.K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Bellingham, Washington, p. 15 (2001) [doi: 10.1117/3.401208]; M.J. Bowden, "An introduction to microlithography," in L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Chapter 1, pp. 9–10 (1994).

[170] Using both reflective and refractive elements.

[171] A.K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Bellingham, Washington, p. 15 (2001) [doi: 10.1117/3.401208].

Line narrowing of the incident radiation was also adopted by the semiconductor industry as a solution for the design problems associated with reduction projection printing. The reasoning here involves the fact that narrowing the spectral width of the illumination light effectively minimizes chromatic aberration to the point where color correction is not needed in refractive systems. The degree of spectral purity requirement increases with increasing NA, so much so that 248-nm exposure systems with a 0.7 NA have a spectral bandwidth on the order of a picometer. Such a narrow spectrum requires high spectral power, which can be obtained with high-repetition-rate laser sources.[172]

Furthermore, both the trend toward increased chip size and the ever-more-stringent critical-dimensional control requirements presented additional technical challenges that had to be overcome, even as stepper technology was being stretched to its limits. The imaging issues associated with this trend were again mitigated by reducing the exposure area with step-and-scan exposure systems,[173] which are essentially a hybrid of the early full-wafer scanning systems and step-and-repeat systems. In the step-and-scan exposure systems, the reticle is scanned rather than exposed all at once. Through repeated stepping and scanning, an entire wafer can be exposed.[174]

The size of the feature that can be printed with optical lithography is ultimately limited by the wavelength of the exposing radiation, which explains the continuous shift to shorter wavelengths that the semiconductor industry has witnessed over the last forty years (Table 5.1). During this period, exposure tools that operate in the near-UV (405 nm–360 nm), deep-UV (250 nm–190 nm), and vacuum-UV (157 nm) region of the spectrum were commercialized. Because the materials used in conventional refractive optical systems are too opaque at these wavelengths, quartz and CaF_2 reduction lenses were introduced for the deep-UV and vacuum-UV exposure tools, respectively. The near-UV systems use filtered Hg arc lamps at 405 nm and a 365-nm light

---

[172] A.I. Ershov, T. Hoffmann, W.N. Partlo, I.V. Fomenkov, G. Everage, P.P. Das, and D. Myers, "Feasibility studies of operating KrF lasers at ultra-narrow spectral bandwidths for 0.18 μm linewidths," *Proc. SPIE* **3334**, 1021–1030 (1998) [doi: 10.1117/12.310732]; U. Stamm, R. Pätzel, K. Kleinschmidt, K. Vogler, W. Zschocke, I. Bragin, and D. Basting, "ArF excimer laser for 193 nm lithography," *Proc. SPIE* **3334**, 1010–1013 (1998) [doi: 10.1117/12.310730].

[173] S. Hirukawa, K. Matsumoto, K. Takemasa, "New projection optical system for beyond 150 nm patterning with KrF and ArF sources," *Proc. SPIE* **3334**, 414–422 (1998) [doi: 10.1117/12.310769]; D. Cote, K. Andersen, D. Cronin, H. Harrold, M. Himmel, J. Kane, J. Lyons, L. Markoya, C. Mason, D. McCafferty, M. McCarthy, G. O'Connor, H. Sewell, and D. Williamson, "Microscan III: Performance of a third generation, catadioptric step and scan lithographic tool," *Proc. SPIE* **3051**, 806–816 (1997) [doi: 10.1117/12.276001]; J. van Schoot, F. Bornebroek, M. Suddendorf, M. Mulder, J. Van der Spek, J. Stoeten, and A. Hunter, "0.7-NA DUV step-and-scan system for 150-nm imaging with improved overlay," *Proc. SPIE* **3679**, 448–463 (1999) [doi: 10.1117/12.354357]; N. Deguchi and S. Uzawa, "150-nm generation lithography equipment," *Proc. SPIE* **3679**, 464–472 (1999) [doi: 10.1117/12.354359].

[174] A.K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Bellingham, Washington, pp. 15–16 (2001) [doi: 10.1117/3.401208].

source, while the deep-UV systems use line-narrowed KrF and ArF exciplex lasers at 248 nm and 193 nm, respectively, as light sources. The vacuum-UV systems use a line-narrowed F_2 excimer laser at 157 nm as a light source.

Currently, DUV 193-nm (ArF) lithography dominates the lithographic technology in terms of market share, closely followed by DUV 248-nm (KrF) lithography. Dry ArF lithography was used to fabricate devices at 90-nm and 65-nm device technology nodes. Vacuum-UV 157-nm (F_2) lithography was abandoned midway during development and was not used in production because of significant birefringence issues with the CaF_2 lens. Introduced in its place was water immersion ArF lithography, which is currently being used to fabricate devices at the 45-nm, 32-nm, and 22-nm technology nodes, and is also being extended to both the 15-nm and possibly 10-nm technology nodes. EUV lithography is currently being inserted into high-volume manufacturing at the 7-nm technology node in leading-edge semiconductor companies and is expected to be the dominant technology that will be scaled to the 3-nm node, and possibly to the end of CMOS technology.

5.6.1.1 Challenges of decreasing exposure wavelength in optical lithography

The strategy of shifting to ever-shorter wavelengths as a means of improving resolution in photolithography cannot continue indefinitely because of a number of factors. The first factor concerns the lack of exposure radiation sources with adequate output in the vacuum-UV and extreme-UV regions of the spectrum. The second factor limiting resolution, particularly as feature sizes become comparable to the exposure wavelength of the coherent light source (laser), is diffraction, which degrades the quality of the printed features. Various resolution enhancement techniques such as off-axis illumination, phase-shifting masks, and the use of immersion optics that have enabled hyper-numerical-aperture imaging have made it possible to push the limits of photolithography to the realm of 15-nm and 10-nm technology nodes. But these technical improvements may not be sufficient to push the limits of photolithography to the sub-3-nm technology nodes.

The third limitation factor is cost, which follows from the second factor in that, because it is technically difficult to pattern features using short-wavelength photons, it is also very expensive to do so. It is true that the use of phase-shifting masks (alternating, attenuated, chromeless, etc.), optical proximity correction (OPC) schemes, and off-axis illumination schemes have enabled the extension of the useful span of photolithography; however, the complexity of mask production and mask inspection presents difficult challenges for which solutions have not been found. In addition, mask cost is extremely high. The photolithographic tools that will be used to make chips well below the 5-nm technology node will invariably be extremely expensive, which may preclude some device manufacturers from buying them.

These inherent limitations of photolithography have thus necessitated the search for alternative semiconductor lithographic technologies. Electron beams, x-rays, and ion beams have been employed in conventional semiconductor lithography, leading to the development of electron-beam, x-ray, and ion-beam lithographies. By subjecting soft, rubbery materials like polydimethylsiloxane to mechanical processes of printing such as stamping, molding, embossing, and cutting, a new form of semiconductor lithography called soft lithography has also been developed within the last 25 years. And lastly, semiconductor lithography has been further extended into the realm of proximal probe techniques based on atomic force microscopy and scanning tunneling microscopy, collectively called scanning probe lithography. In the remainder of Section 5.6, we briefly review each of these new extensions of semiconductor lithography.

5.6.2 X-ray lithography

Lithography using x-rays with wavelengths between 0.1 and 10 nm was first described by Spears and Smith[175] in 1972, and since then, significant research efforts have been devoted to it in many research institutes and IC companies across the globe, with the motivation of using it as a possible replacement for optical lithography. There are quite a few reasons why x-ray lithography is theoretically more attractive than optical lithography. Because x-rays have much shorter wavelengths than the UV radiation currently used in optical lithography, they reduce wavelength-related diffraction problems, which limit resolution in optical lithography. Also, because resist films are transparent to x-rays, high-aspect-ratio patterns can be fabricated with x-ray lithography. Furthermore, there is practically no field size limitation associated with x-ray lithography. And because the index of reflection is practically the same for all materials (close to unity) in the x-ray spectral region, standing-wave effects and other reflection-related problems that plague optical lithography are eliminated in x-ray lithography.

Early x-ray systems employed conventional impact sources, which were relatively low-brightness sources. These systems were designed to expose an entire wafer in a single exposure, but the low brightness of the source necessitated the use of very sensitive resists. High-brightness sources such as synchrotron storage rings or pulsed laser-induced plasmas[176] became available over the last three decades, and several companies even attempted to incorporate these sources into step-and-repeat systems, which improved

---

[175] E. Spears and H.I. Smith, "High-resolution pattern replication using soft X rays," *Electron Lett.* **8**, 102 (1972); E. Spears and H.I. Smith, "X-ray lithography: A new high resolution replication process," *Solid State Technol.* **15**(7), 21 (1972).

[176] D.W. Peters, J.P. Drumheller, R.D. Frankel, A.S. Kaplan, S.M. Preston, and D.N. Tomes, "Application and analysis of production suitability of a laser-based plasma x-ray stepper," *Proc. SPIE* **923**, 28 (1988) [doi: 10.1117/12.945630]; D.W. Peters and R.D. Frankel, "X-ray lithography: The promise of the past and reality of the present," *Solid State Technol.* **32**(3), 77 (1989); A. Heuberger, "X-ray lithography," *J. Vac. Sci. Technol. B* **6**, 107 (1988); R. Hollman, "X-ray lithography using broadband sources," *J. Vac. Sci. B* **6**, 186 (1988).

registration and relaxed the resist sensitivity requirements. Many problems still exist, however, including the fabrication of stable x-ray masks, alignment, and the commercial availability of suitable resist systems.[177]

In spite of all the efforts that have been expended in the development of x-ray lithography, it has not yet been employed in any meaningful manner in IC device production. The difficulty of fabricating defect-free and stable x-ray masks have presented nearly insurmountable challenges. The extension of optical lithography (photolithography) to technology nodes that were once deemed beyond its resolution limits, barely three decades ago, have also helped to consign x-ray lithography to the sidelines of IC device manufacture. In addition, conventional lenses are not capable of focusing x-rays. Furthermore, many of the materials used in the fabrication of masks and lenses are rapidly degraded by the highly energetic x-ray radiation. Consequently, there are no current efforts to develop a production-worthy x-ray lithography capability within the semiconductor industry.

5.6.3 Electron-beam lithography

Lithography using beams of electrons to expose resist is currently a mature technology, having evolved from the scanning electron microscope developed in the mid-1950s.[178] In fact, it was one of the earliest processes used for fabricating ICs, dating back to 1957.[179] Almost from its very beginning, sub-100-nm resolution was reported with the technique. Broers reported

---

[177] M.J. Bowden, "An introduction to microlithography," in L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C., pp. 11–12 (1994).

[178] K.C.A. Smith and C.W. Oatley, Jr., "The scanning electron microscope and its fields of application," *Br. J. Apply Phys.* **6**, 391 (1955); T.E. Everhart, "Persistence pays off: Sir Charles Oatley and the scanning electron microscope," presented at the 40th International Conference on Electron, Ion, and Photon Beam Technology and NanoFabrication, Atlanta, Georgia, May 28–31 (1996); A.N. Broers, in *Proc. Symp. Electron Ion Beam Sci. Technol.*, pp. 3–25 (1972); A.N. Broers, "A new high-resolution electron probe," *J. Vac. Sci. Technol.* **10**, 979 (1973); D.S. Alles, F.R. Ashley, A.M. Johnson, and R.O. Townsend, "Control system design and alignment methods for electron lithography," *J. Vac. Sci., Technol.* **12**, 1252 (1975); D.R. Herriot, R.J. Collier, D.S. Alles, and J.W. Stafford, "EBES: A practical electron lithographic system," *IEEE Trans. Electron Devices* **ED-22**, 385 (1975); H.C. Pfiffer, T.R. Groves, and T.H. Newmann, "High-throughput, high resolution electron beam lithography," *IBM J. Res. Dev.* **32**(4), 494 (1988); J.T. Clemens, "Electron-beam lithography: A gating item," *Solid State Technol.* **32**(3), 69 (1989); A.N. Broers, "Resolution limits for electron-beam lithography," *IBM J. Res. Dev.* **32**(4), 502 (1988); H.C. Pfeiffer, "Variable spot shaping for electron beam lithography," *J. Vac. Sci. Technol.* **15**, 887 (1978); D.R. Herriot and G.R. Brewer, "Electron-beam lithography machines," in *Electron-Beam Technology in Microelectronic Circuit Fabrication*, Academic Press, Orlando, Florida, pp. 141–216 (1981).

[179] D.A. Buck and K. Shoulders, "An approach to micro/mini structure printed systems," *Proc. of Eastern Joint Computer Conference*, American Institute of Electrical Engineers, New York, p. 55 (1957).

50-nm lines ion milled into metal films using a contamination resist patterned with a 10-nm-wide electron beam as early as 1964.[180] The year 1976 saw the fabrication of 8-nm lines on Au-Pd with a 0.5-nm probe and a 10-nm-thick carbon membrane substrate using electron-beam lithography.[181] In 1985, Murray et al. reported 1- to 2-nm features patterned on metal halide resists using this technique.[182] Except for more recent reports on atomic scale resolution with proximal probe techniques, the resolution of e-beam lithography has been unsurpassed by any other lithographic technique.[183]

Although all high-volume production of ICs is currently done with optical lithography, electron-beam lithography nevertheless continues to dominate two small but vital segments of the overall semiconductor lithography field. First, it is used to generate the masks used in optical lithography. Second, it is used in low-volume manufacture of specialty, high-resolution devices, custom circuits, and prototype devices. These two vital functions performed by electron-beam lithography will in all probability not diminish in importance in the foreseeable future.[184]

The main drawback of electron-beam lithography is its low-throughput and high capital cost. The major advantages of electron-beam lithography include its ability to register accurately over small areas of a wafer, low defect densities, direct generation of patterns from circuit design data, and lack of diffraction issues of the type that limit resolution in optical lithography because electrons do not diffract at atomic scales.[185]

The two principal modes of operation of electron-beam exposure systems include maskless direct-write electron-beam machines and mask-based electron-beam machines. Direct-write electron-beam machines operate directly from design data and are capable of extremely high resolution.[186] Mask-based electron-beam systems utilize masks in their imaging process. The implementation of electron-beam lithography in mix-and-match mode

---

[180] A.N. Broers, "Micromachining by sputtering through a mask of contamination laid down by an electron beam," *Proc. of the First International Conference on Electron and Ion Beam Technology*, R. Bakish, Ed., John Wiley & Sons, New York, p. 181 (1964).

[181] A.N. Broers, W.W. Molzen, J.J. Cuomo, and N.D. Wittels, "Electron-beam fabrication of 80-Å metal structures," *Appl. Phys. Lett.* **29**, 596 (1976).

[182] A. Murray, M. Scheinfein, M. Isaacson, and I. Adesida, "Radiolysis and resolution limits of inorganic halide resists," *J. Vac. Sci. Technol. B* **3**, 367 (1985).

[183] E.A. Dobisz, F.K. Perkins, and M.C. Peckerar, "E-beam and proximal probe processes for nanolithography," in *Microlithography: Science and Technology*, J.R. Sheats and B.W. Smith, Eds., Marcel Dekker, New York, pp. 715–775 (1998).

[184] M.J. Bowden, "An introduction to microlithography," in L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C., Chapter 1, p. 11 (1994).

[185] ibid.

[186] ibid.

with optical lithography in a manufacturing environment has been demonstrated.[187]

Several innovations in electron-beam reduction projection systems aimed at improving the low throughput of electron-beam writing machines have been reported.[188] For instance, workers at Hitachi developed the cell projection electron-beam system, which projects entire transistor "cells" in a single exposure, which are then sequentially stepped over the entire device area.[189] This technique appears to be best suited for devices that are composed primarily of identical transistor elements, such as those as found in DRAMs. In addition, workers at AT&T developed a reduction projection method known as SCALPEL (scattering with aperture-limited projection electron-beam lithography), which uses a mask that scatters high-energy (70–200 kV) electrons through a thin patterned mask. The SCALPEL mask is made of a thin and low-atomic-number membrane and a high-atomic-number metal scattering element that is patterned. The scattered electrons are removed from the imaging systems through a back focal plane aperture.[190] The SCALPEL technology suffers from space-charge effects, which make it nearly impossible to achieve both high resolution and high throughput simultaneously. As a result, the SCALPEL technology is no longer being pursued by the semiconductor industry. It should be pointed out, however, that when compared with conventional serial scanning systems, both the Hitachi and AT&T innovations allow simultaneous printing of a much larger number of pixels.[191]

Another major innovation in maskless electron-beam technology is multiple-electron-beam direct-writing systems, which use thousands of electron beams operating in parallel to write circuit patterns on wafers or

---

[187] W.W. Flack, D.H. Cameron, V.J. Alameda, and G.C. Malek, "Mix-and-match lithography in a manufacturing environment," *Proc. SPIE* **1671**, 126–136 (1992) [doi: 10.1117/12.136008].

[188] Y. Sohda, Y. Makayama, N. Saitou, H. Itoh, and H. Todokoro, "Electron optics for high throughput electron beam lithography system," *J. Vac. Sci. Technol. B* **9**(6), 2940 (1991); H.W.P. Koops, "Capacity of electron beam reducing image projection systems with dynamically compensated field aberrations," *Microelectronic Eng*. **9**(1–4), 217–220 (1989); S.D. Berger and J.M. Gibson, "New approach to projection-electron lithography with demonstrated 0.1 μm linewidth," *Appl. Phys. Lett*. **57**, 153 (1990); S.D. Berger, J.M. Gibson, R.M. Camarada, R.C. Farrow, H.R. Huggins, J.S. Kraus, and J.A. Liddle, "Projection electron-beam lithography: A new approach," *J. Vac. Sci. Technol. B* **5**, 2996 (1991).

[189] H. Itoh, H. Tadokoro, Y. Sohda, Y. Nakayama and N. Saitou, "Cell projection column for high speed electron-beam lithography system," *J. Vac. Sci. Technol. B* **10**, 2799 (1992).

[190] S.D. Berger and J.M. Gibson, "New approach to projection-electron lithography with demonstrated 0.1 μm linewidth," *Appl. Phys. Lett*. **57**, 153 (1990); S.D. Berger, J.M. Gibson, R.M. Camarada, R.C. Farrow, H.R. Huggins, J.S. Kraus, and J.A. Liddle, "Projection electron-beam lithography: A new approach," *J. Vac. Sci. Technol. B* **5**, 2996 (1991).

[191] M.J. Bowden, "An introduction to microlithography," in L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C., Chapter 1, pp. 11–12 (1994).

mask substrates, eliminating the need for the costly photomasks used in optical lithography. Companies such as Mapper Lithography BV (now part of ASML) of the Netherlands and IMS Nanofabrication GmbH of Vienna are the leaders in this field.

5.6.4 Ion-beam lithography

Lithography using beams of ions has been investigated by several researchers[192] who were motivated by the fact that ion-beam lithography can achieve higher resolution than electron-beam lithographic techniques because ions have a higher mass and therefore scatter less than electrons. The technique also allows the use of resists with lower sensitivities than the other lithographic approaches. One major drawback of ion-beam lithography is that it may suffer from random (or stochastic) space-charge effects, causing broadening of the ion beam. The most important application of ion-beam lithography to date is the repair of optical lithographic masks.[193]

There are two types of ion-beam lithography systems: a scanning focused-beam system and a large-area projection system. The former is similar to the electron-beam machine, in which the ion source can be Ga^+ or H^+, the major differences being in the source and deflection lenses.[194] The latter system is similar to an optical 5× reduction projection step-and-repeat system, which projects 100-keV light ions such as $H_2{}^+$ through a stencil mask.[195] It must, however, be stated that, before these ion-beam lithographic techniques can become competitive with electron-beam lithography, there must be significant advances in high-brightness sources, high-speed deflection systems for the scanning system, and stable masks and large lenses for the projection tools.[196]

---

[192] T.M. Hall, A. Wagner, and L.F. Thompson, "Ion beam exposure characteristics of resists," *J. Vac. Sci. Technol.* **16**, 1889 (1979); T. Kaneko, T. Umegaki and Y. Kawakomi, *Proc. Kodak Interface '80*, p. 25 (1980).

[193] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, pp. 79–80 (2004).

[194] J. Melngailis, "Focused ion beam technology and applications," *J. Vac. Sci. Technol. B* **5**, 469 (1987); K. Hosond, H. Morimoto, Y. Watanabe, and T. Kato, "Practical technologies for FIB direct-writing applications," *Proc. SPIE* **923**, 84–91 (1988) [doi: 10.1117/12.945635].

[195] G. Stengl, W. Loschner, and P. Wolf, "Ion projection lithography machine IPLM-01: A new tool for sub-0.5-μm modification of materials," *J. Vac. Sci. Technol. B* **6**, 194 (1986); G. Stengl, G. Bosch, A. Chalupka, J. Fegerl, R. Fischer, G. Lammer, H. Loschner, L. Malek, R. Nowak, C. Traher, P. Wolfe, P. Mauger, A. Shimkunas, S. Sen, and H.C. Wolfe, "Ion projector wafer exposure results at 5× ion-optical reduction obtained with nickel and silicon stencil masks," *J. Vac. Sci. Technol. B* **10**(6), 2824 (1992).

[196] M.J. Bowden, "An introduction to microlithography," in L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., *Introduction to Microlithography*, 2nd ed., American Chemical Society, Washington, D.C., Chapter 1, pp. 11–12 (1994).

5.6.5 Extreme-ultraviolet lithography

Extreme-ultraviolet (EUV) lithography is a promising next-generation lithographic technique for replacing optical lithography, perhaps at the sub-10-nm technology nodes. Invented independently in the mid-1980s by researchers at NTT and AT&T Bell Laboratories,[197] who were inspired by the availability of mirrors of reasonable efficiency at soft x-ray wavelengths ($\lambda \approx 10$–$15\,\text{nm}$), it offers distinct advantages of wide process margins and extendibility to multiple lithographic nodes. Its main drawback is the enormous cost of bringing up its infrastructure.

The soft x-rays with wavelengths of 10–14 nm used in EUV lithography are produced from a laser-produced plasma or synchrotron radiation. A mask that is produced by patterning an absorber material deposited on multilayer-coated flat silicon is used to reflect the EUV radiation from the nonabsorbing regions of the mask through a 4× reduction camera and imaged onto a thin layer of resist on the wafer. Due to the strong absorptivity of all materials at EUV wavelengths, the lithography process is performed in vacuum. In particular, the mirrors are coated with multilayer coatings that produce distributed quarter-wave Bragg reflectors. The mask blank is also multilayer-coated to maximize its reflectivity at wavelengths of 10–14 nm.[198]

The major problems that plagued EUV lithography and delayed its insertion into high-volume manufacturing in previous years have now been largely resolved, paving the way for insertion of the technology into high-volume production at the 7-nm technology node at leading-edge fabs such as Samsung, TSMC,[199] and Intel. Problems such as low source power, high mask defects (exacerbated by difficulty of producing defect-free mask blanks), and protecting the fabricated masks from contaminants (given the lack of pellicles to protect the masks), as well as a lack of resists that simultaneously met sensitivity, resolution, and line edge roughness requirements, were all major challenges that had to be overcome. The source power of the latest exposure tools is now sufficiently high,[200] mask defectivity has been significantly reduced,[201] and resists with sufficiently

---

[197] D.M. Tennant, J.E. Bjorkholm, R.M D'Souza, L. Eichner, R.R. Freeman, J.Z. Pastalan, L.H. Szeto, O.R. Wood, T.E. Jewell, W.M. Mansfield, W.K. Waskiewica, D.L. White, D.L. Windt, and A.A. MacDowell, "Reflective mask technologies and imaging results in soft x-ray projection lithography," *J. Vac. Sci. Technol. B* **9**(6), 3176 (1991).

[198] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, John Wiley & Sons, Hoboken, New Jersey, pp. 79–80 (2004).

[199] Optics.org, "TSMC orders production EUV systems," (2014) https://optics.org/news/5/11/35.

[200] H. Mizoguchi, H. Nakarai, T. Abe, et al. "High power LPP-EUV source with long collector mirror lifetime for high volume semiconductor manufacturing," *2108 China Semiconductor Technology International Conference (CSTIC)* (IEEE, publisher), Shanghai, China (2018).

[201] M. Lapedus, "EUV mask bland battle brewing," *Semiconductor Engineering*, Nov. 15 (2018).

high sensitivity and resolution and low line edge roughness have been developed.[202,203,204]

5.6.6 Soft lithography

Soft lithography, developed in the 1990s in George Whitesides' Laboratories at Harvard University, involves printing, molding, and other mechanical processes carried out using an elastic stamp to fabricate patterns with very fine resolutions. Photolithography or electron-beam lithography is used to fabricate the stamp in a resist layer on the surface of silicon wafer; this results in a relief master in which islands of resist stand out from the silicon. Patterning of features with this stamp is accomplished by pouring dimethylsiloxane—a free-flowing liquid—over the bas-relief master and curing it into the rubbery solid polymer called polydimethylsiloxane (PDMS). With this stamp, several techniques, including microcontact printing, micromolding in capillaries, and nanoskiving, can be used to fabricate nanostructures. Related techniques based on step-and-flash imprint lithography and nanoimprint lithography that use hard masters instead of elastic PDMS have also been developed and have shown excellent resolution.[205]

Soft lithography, it must be pointed out, is not well suited for fabricating structures with multiple layers that stack precisely on top of one another, as is the case in conventional ICs. Deformations and distortions of the soft PDMS stamp can result in errors in the replicated pattern and a misalignment of the pattern with any underlying pattern previously fabricated. The use of rigid stamps in imprint lithography has proved reasonably successful in correcting this shortcoming. However, it remains extremely difficult to produce rigid masters without defects, which given that this is a 1× printing technique, are reproduced on the wafer with a high degree of fidelity.[206]

One significant advantage of imprint lithography over standard photolithography is the ability to pattern 3D topographies in a single step, thus reducing the cost of fabricating such a device. Such topographies are critical for connecting different layers on an IC.[207]

---

[202] N. Mojarad, J. Gobrecht, and Y. Ekinci, "Beyond EUV lithography: a comparative study of efficient photoresists' performance," *Scientific Reports* **5**, 9235 (2015).

[203] N. Mojarad. M. Hojeij, L. Wang, J. Gobrecht, and Y. Ekinci, "Single-digit-resolution nanopatterning with extreme ultraviolet light for the 2.5 nm technology node and beyond," *Nanoscale* **7**, 4031–4037 (2015).

[204] S.T. Meyers and A. Grenville, "Metal oxide EUV photoresists for N7 relevant patterns," presented at the 2016 International Symposium on EUV Lithography, Berkeley (2016).

[205] G.M. Whitesides and J.C. Love, "The art of building small," *Scientific American Reports* **17**(3), 13–21 (2007); B.D. Gates, Q.B. Xu, M. Stewart, D. Ryan, C.G. Willson, and G.M. Whitesides, "New approaches to nanofabrication: molding, printing, and other techniques," *Chemical Reviews* **105**(4), 1171–1196 (2005).

[206] ibid.

[207] ibid.

5.6.6.1 Microcontact printing

In the microcontact printing technique, the PDMS stamp is inked with a reagent solution consisting of organic molecules called thiols. Next, the stamp is brought into contact with a thin film of gold on a glass, silicon, or polymer plate, resulting in a reaction between the thiols and the gold surface that forms a highly ordered film (a self-assembled monolayer) that replicates the stamp's pattern. Given that the thiol ink does spread somewhat beyond the pattern area after it contacts the surface, the resolution of the monolayer cannot be quite as high as that of the PDMS stamp. However, when performed correctly, microcontact printing can produce patterns with features as small as 50 nm.[208]

5.6.6.2 Micromolding in capillaries

The PDMS stamp in this technique is placed on a hard surface containing liquid polymer such that, through capillary action, the liquid polymer flows into the recesses between the surface and the stamp. Upon solidification, the polymer takes the form of the desired pattern. Patterning is complete when the stamp is released from the solidified polymer. This technique can replicate features smaller than 10 nm.[209]

5.6.6.3 Nanoskiving

In nanoskiving, a PDMS stamp is first used to mold a hard plastic such as epoxy with, for example, cylindrical posts. Coating the molded epoxy with a thin (about 50 nm) film of metal followed by coating more epoxy and cutting parallel to the plane of the sandwich structure creates a thin sheet of plastic containing nanostructures shaped like the cross-section of the original molded structures; for cylindrical posts, the resulting shapes are rings. The thickness of the slice determines their height, and the deposited film determines their thickness. These techniques are well suited for fabricating subwavelength optical devices, waveguides, and optical polarizers on the one hand, and on the other hand, for creating biochips based on nanofluidics and microfluidics that are used in biochemical research.[210]

5.6.7 Nano-imprint lithography

Developed by Stephen Chou of Princeton University in the 1990s, nano-imprint lithography uses a rigid master placed on a film of polymer that has

---

[208] G.M. Whitesides and J.C. Love, "The art of building small," *Scientific American Reports* **17**(3), 13–21 (2007).

[209] ibid.

[210] ibid.

been heated to a temperature near its melting point to facilitate the embossing process. This technique has shown resolution down to 20 nm with good fidelity.[211]

5.6.7.1 UV imprint lithography

In this technique, UV illumination is applied to cure photocurable resist film once it has been imprinted. A version of this technique, developed by C. Grant Willson of the University of Texas at Austin around 2000, differs from the soft lithographic techniques developed by George Whitesides in that, in step-and-flash imprint lithography, lithography is used to pattern a quartz plate, yielding a rigid bas-relief master, which is directly pressed against a thin film of liquid polymer (thus eliminating the step of making the elastic PDMS stamp from the master as in Whitesides' approach), which fills the master's recesses. Then the master is exposed to UV radiation, which cures and solidifies the polymer to create the desired replica. Releasing the quartz plate from the hardened polymer provides the patterned structures with resolutions down to 20 nm.[212]

5.6.8 Proximal probe lithography

One of the newest additions to lithographic patterning techniques is proximal probe lithography.[213] Not long after the invention of the scanning tunneling microscope (STM) by Heinrich Rohrer and Gerd K. Binnig of IBM Zurich Research Laboratories in 1981,[214] researchers began to investigate the application of proximal probes to lithography.[215] Proximal probes are attractive from a number of standpoints because they offer the potential for atomic-scale resolution. They have been used to define lithographic patterns in commercial resists, surface oxide, surface modification, electrochemical etching, and chemical vapor deposition. Furthermore, simple electronic

---

[211] G.M. Whitesides and J.C. Love, "The art of building small," *Scientific American Reports* **17**(3), 13–21 (2007).

[212] B.D. Gates, Q.B. Xu, M. Stewart, D. Ryan, C.G. Willson, and G.M. Whitesides, "New approaches to nanofabrication: molding, printing, and other techniques," *Chemical Reviews* **105**(4), 1171–1196 (2005).

[213] C.R.K. Marrian, Ed., *The Technology of Proximal Probe Lithography*, *Proc. SPIE* **10310** (1993) [doi: 10.1117/12.2283725].

[214] G. Binnig and H. Rohrer, "Scanning tunneling microscopy," *Helv. Phys. Acta* **55**, 726 (1982); G. Binnig, H. Rohrer, Ch. Gerber, and W. Weibel, "Tunneling through a controllable vacuum gap," *Appl. Phys. Lett.* **40**, 178 (1982); G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, "Surface studies by scanning tunneling microscopy," *Phys. Rev. Lett.* **49**, 57 (1982).

[215] M.A. McCord and R.F.W. Pease, "High resolution, low-voltage probes from a field emission source close to the target plane," *J. Vac. Sci. Technol. B* **3**, 198 (1985).

structures have been made and their transport properties have been measured with proximal probe lithography.[216]

Proximal probes are defined as tools in which a tip is held close to or in intimate contact with a surface so that there is an interaction between the atoms on the tip and the substrate.[217] The most widely used proximal probes include the STM, in which current is induced to flow across a potential difference between a sharp tip and the substrate, and the atomic force microscope (AFM), in which an electric field potential is established between the tip and substrate, resulting in overlapping atomic repulsive potentials. In both STM and AFM, a conducting tip is held in close proximity to a substrate and potential is applied between the tip and the sample.[218]

In STM lithography, the tip-to-sample separation is adjusted to maintain a constant current of approximately 0.003–3 nA between the tip and the sample. The tip-to-sample current is fed into the circuitry that drives the piezoelectric transducer, which in turn controls the tip height. Lateral piezoelectric transducers scan the tip across the surface.[219]

In AFM lithography, the tip rests on the end of a cantilever, with the force (1–100 nN) between the tip and the sample being determined by measurement of the deflection of the cantilever. This measurement can be optical, capacitive, or piezoresistive. Here, the tip-to-sample separation is adjusted to maintain a constant force between the tip and the sample.[220]

---

[216] E.E. Ehrichs, W.F. Smith, and A.L. de Lozanne, "Four-probe resistance measurements of nickel wires written with a scanning tunneling microscope/scanning electron microscope system," *J. Ultramicrosc*. **42–44**, 1438 (1992); L. Stockman, I. Heyvaert, C. van Haesendonck, and Y. Bruynseraede, "Submicrometer lithographic patterning of thin gold films with a scanning tunneling microscope," *Appl. Phys. Lett*. **62**, 2935 (1993); K. Matsumoto, M. Ishii, K. Segawa, Y. Oka, B.J. Vartanian, and J.S. Harris, "Room temperature operation of a single electron transistor made by the scanning tunneling microscope nanooxidation process for the TiO_x/Ti system," *Appl. Phys. Lett*. **68**, 34 (1996); T. Fayfield and T.K. Higman, "Fabrication and transport measurements of atomic force microscope modified silicon metal-oxide semiconductor field-effect transistors," *J. Vac. Sci. Technol. B* **13**, 1285 (1995); P.M. Campbell, E.S. Snow, and P.J. McMarr, "Fabrication of nanometer-scale side-gated silicon field effect transistors with an atomic force microscope," *Appl. Phys. Lett*. **66**, 1388 (1995); S.C. Minne, H.T. Soh, Ph. Flueckiger, and C.F. Quate, "Atomic force microscope lithography for fabricating a 0.1 μm metal-oxide field-effect transistor," *Appl. Phys. Lett*. **66**, 703 (1995); K. Matsumoto, S. Takahashi, M. Ishii, M. Hoshi, A. Kurokawa, S. Ichimura, and A. Ando, "Application of STM nanometer-size oxidation process to planar-type MIM diode," *Jpn. J. Appl. Phys*. **34**, 1387 (1996).

[217] C.J. Chen, *Introduction to Scanning Tunneling Microscopy*, Oxford University Press, New York (1993).

[218] E.A. Dobisz, F.K. Perkins, and M.C. Peckerar, "E-beam and proximal probe processes for nanolithography," in *Microlithography: Science and Technology*, J.R. Sheats and B.W. Smith, Eds., Marcel Dekker, New York, p. 737 (1998).

[219] ibid., pp. 737–738.

[220] ibid., p. 738.

Several physical approaches to writing patterns by STM and AFM have been reported. For AFM, these include moving nanoparticles around and arranging them in patterns by AFM tips, and scratching the surface. For STM, by increasing the current flowing from the tip of the STM, the microscope becomes a very small source for an electron beam, which can be used to write nanometer-scale patterns by probe-induced chemical vapor deposition,[221] lithography in resists,[222] and surface modification.[223] The STM tip can also push individual atoms around on a surface to build patterns.[224]

Several challenges must be addressed before proximal probe lithography can make the transition into mainstream lithographies. The most important issues concern the availability of high-quality resists or imaging layers, control and maintenance of the integrity of the tip, and throughput.[225]

A new scanning probe method, called dip-pen lithography, was developed by Chad Manning of Northwestern University during the first decade of the 21st century. In this technique, the tip of an AFM is coated with a thin film of thiol molecules that are insoluble in water but react with a gold surface. When the device is placed in an atmosphere containing an appreciable concentration of water vapor, a minute drop of water condenses between the gold surface and the microscope's tip. Surface tension pulls the tip to a fixed distance from the gold, and this distance remains fairly fixed as the tip moves across the surface. The drop of water effectively acts as a bridge over which the thiol molecules migrate from the tip to the gold surface, where they form a self-assembled monolayer. Lines that are a few nanometers across have been successfully written with this procedure.[226]

Although dip-pen lithography is a relatively slow technique, it has great versatility with respect to the many varieties of molecules it can use as "inks," which makes it an attractive candidate for introducing chemical flexibility to

---

[221] F.K. Perkins, M. Onellion, S. Lee, and P.A. Dowben, "Demonstrating the utility of boron based precursor molecules for selective area deposition in a scanning tunneling microscope," *Mater. Res. Soc. Proc.* **236**, 153 (1992); D.S. Scully, A. Ermakov, E.L. Garfunkel, and P.A. Dowben, "Electron-beam induced patterned deposition of allylcyclopentadienyl palladium using scanning tunneling microscopy," *J. Appl. Phys.* **76**, 7639 (1994).

[222] E.A. Dobisz, F.K. Perkins, and M.C. Peckerar, "E-beam and proximal probe processes for nanolithography," in *Microlithography: Science and Technology*, J.R. Sheats and B.W. Smith, Eds., Marcel Dekker, New York, pp. 715–775 (1998).

[223] G.M. Shedd and P. Russel, "The scanning tunneling microscope as a tool for nanofabrication," *Nanotechnology* **1**, 67 (1990).

[224] J.A. Stroscio and D.M. Eigler, "Atomic and molecular manipulation with the scanning tunneling microscope," *Science* **254**, 1319–1326 (1991); For pictures of atom manipulation, see http://www.almaden.ibm.com:80/vis/stm/gallery.html.

[225] E.A. Dobisz, F.K. Perkins, and M.C. Peckerar, "E-beam and proximal probe processes for nanolithography," in *Microlithography: Science and Technology*, J.R. Sheats and B.W. Smith, Eds., Marcel Dekker, New York, p. 739 (1998).

[226] G.M. Whitesides and J.C. Love, "The art of building small," *Scientific American Reports* **17**(3), 13–21 (2007).

nanometer-scale writing. Parallel arrays of independent scanning probes have significantly improved the throughput of dip-pen lithography in recent times.[227]

5.6.9 Atom lithography

A variety of direct atom manipulations have been demonstrated with proximal probes.[228] The electric field strength in the vicinity of a probe tip is very strong and inhomogeneous, so much so that it can be used to manipulate atoms, including sliding atoms over surfaces and transferring atoms by pick (erase) and place (write). These atomic manipulation processes can be classified as parallel processes and perpendicular processes. In parallel processes, an adsorbed atom or molecule is induced to move along the surface (sliding); in perpendicular processes, the atom or molecule is transferred from the surface to the tip of a proximal probe, or vice versa. While the theoretical resolution of this technique is a single atom, its major drawback is the extremely long time it takes to generate even the simplest of features.[229] The use of the atomic processing microscope (APM) by Kamerski et al.[230] in 1993 has opened up new methods for considerably shortening the time. These researchers succeeded in imaging selected atoms on a surface, stripping them off, and replacing them with other atoms, all in a matter of minutes. The APM relies on some important technical innovations such as photon biasing, whereby a precise voltage and photon pulse are tuned" to a specific atom to strip it from the surface. Optical biasing (called "optical tweezers") has been used to manipulate a variety of individual particles on a variety of surfaces.[231]

5.6.10 Self-assembly lithography

Unlike top-down lithographic techniques in which patterns are formed by reductive processes involving removal of matter to define the pattern, self-assembly lithography is based on bottom-up hierarchical and spontaneous self-assembly of molecular materials (monolayers and block copolymers) and particulate materials (colloidal particles) to create larger patterns that can be arranged in a random or ordered manner.

---

[227] ibid.

[228] J.A. Stroscio and D.M. Eigler, "Atomic and molecular manipulation with the scanning tunneling microscope," *Science* **254**, 1319–1326 (1991).

[229] M.J. Madou, *Fundamentals of Microfabrication: The Science of Miniaturization*, 2nd ed., CRC Press, New York, p. 61 (2002).

[230] Editorial, "Ingenious STM puts atoms right where you want them," *Res. & Dev.*, p. 71, April (1993).

[231] M.J. Madou, *Fundamentals of Microfabrication: The Science of Miniaturization*, 2nd ed., CRC Press, New York, p. 61 (2002).

5.6.10.1 Molecular self-assembly lithography

Molecular self-assembly, a chemical process in which molecules spontaneously organize to form larger ordered structures, represents one of the newest emerging lithographic techniques. In one form of this technique, self-assembling precursor molecules from solutions or vapor phase react at interfaces to produce layers of monomolecular thickness that are chemically bonded to solid surfaces through thermodynamically favorable bond formation involving chemisorption. Such layers belong to a class of materials called self-assembled monolayers (SAMs) and make good ultrathin film resist candidates. These SAM films can be molecularly engineered to be patterned by various types of energetic radiation, including deep UV, soft x-rays, ion beams, and low-energy electrons.[232] Scratching (Section 5.6.7) and microcontact printing (Section 5.6.6.1) have been used to pattern SAMs. Device features smaller than 20 nm have been successfully patterned on SAMs with STMs and conventional electron-beam lithography systems.[233]

Resists based on SAMs have also been successfully deposited on metals such as gold, aluminum, titanium, zirconium, silver, copper, and platinum, as well as SiO_2, GaAs, and other surfaces, [234] and were no doubt inspired by the first demonstration of the formation of long-chain ω-substituted dialkyldisulfides on gold in 1983.[235] Alkane-thiols and dialkyldisulfides, which are lipidlike organic molecules having the general formula $HS–(CH_2)_n–X$ and $X–(CH_2)_n–SS–(CH_2)_m–Y$, respectively, where n and m indicate alkyl chain length, and X and Y indicate the end groups [$–CH_3$, –azobenzene, –OH, etc.],[236] are well-known precursor materials for SAMs on gold, the most extensively studied substrate. Organosilanes are often used to form SAMs on Si surfaces through a Si–O-substrate siloxane bond; the alkyl group R is responsible for the ordered nature of the film.[237]

5.6.10.2 Colloidal particle self-assembly lithography

When solutions of monodisperse spherical colloidal particles are coated on a surface to a thickness of a monolayer or double layers, they tend to

---

[232] C.R.K. Marrian and E.S. Snow, "Proximal probe lithography and surface modification," *Microelectron. Eng*. **32**, 173–189 (1996).

[233] H.U. Muller, C. David, B. Volkel and M. Gunze, "Nanostructuring of alkanethiols with ultrasharp field emitters," *J. Vac. Sci. Technol. B* **13**, 2846–2849 (1995).

[234] H.A. Biebuyck, N.B. Larsen, E. Delamarche, and B. Michel, "Lithography beyond light: microcontact printing with monolayer resists," *IBM J. Res. & Dev*. **41**(1.2), 159–170 (1997).

[235] R.G. Nuzzo and D.L. Allara, "Adsorption of bifunctional organic disulfides on gold surfaces," *J. Am. Chem. Soc*. **105**, 4481 (1983).

[236] E. Delamarche, B. Michel, H.A. Biebuyck, and C. Gerber, "Golden interfaces: surface self-assembled monolayers," *Adv. Materials* **8**, 719–729 (1996).

[237] M.J. Madou, *Fundamentals of Microfabrication: The Science of Miniaturization*, 2nd ed., CRC Press, New York, pp. 62–63 (2002).

spontaneously self-assemble into long-range-ordered periodic arrays with hexagonal close packing, driven simply by entropic depletion and gravity, as well as by capillary forces arising from the evaporation of the solvent.[238,239,240] The exploitation of this phenomenon has resulted in periodic spherical particle arrays, which are subsequently used as masks to deposit materials in the interstices or spaces between the assembled particles constituting colloidal lithography. Fisher and co-workers used such arrays of spherical nanoparticles as shadow masks to fabricate platinum structures in the early 1980s.[241] Since then, a variety of lithographic methods based on this principle, in which spherical nanoparticles are used as masks for depositing nanomaterials and as scaffolds for templating 2D and 3D patterns, have been developed.[242,243,244,245]

5.6.10.3 Directed block copolymer self-assembly lithography

A new variant of the self-assembly technique, called directed block copolymer self-assembly lithography, pioneered in Tom Russel's group at the University of Massachusetts at Amherst and Paul Nealy's group at the University of Wisconsin, Madison, involves the use of external fields (such as interfacial interactions, electric fields, thermal fields, capillary force, solvent evaporation, flow fields), internal fields, gradient fields, and coupled fields to direct the self-assembly of appropriate precursors, typically block copolymers, on surfaces into desired patterns that can be imaged with various types of energetic radiations, including deep UV, soft x-rays, and electron beams, and transferred into underlying substrates. A key advantage of directed self-assembly is the interplay between thermodynamic driving forces and kinetic/transport forces that mediate the phase separation between the polymer

---

[238] N.D. Denkov, O.D. Velev, P.A. Kralchevsky, I.B. Ivanov, H. Yoshimura, and K. Nagayama, "Two-dimensional crystallization," *Nature* **361**, 26 (1993).

[239] Y.N. Xia, B. Gates, Y. Yin, and Y. Lu, "Monodisperse colloidal particles: old materials with new applications," *Adv. Mater*. **12**, 693–713 (2000).

[240] S.M. Yang, S.G. Jang, D.G. Choi, S. Kim, and H.K. Yu, "Nanomachining by colloidal lithography," *Small* **2**, 458–475 (2006).

[241] U.C. Fischer and H.P. Zingsheim, "Submicroscopic pattern replication with visible light," *J. Vac. Sci. Technol*. **19**, 881–885 (1981).

[242] C.L. Haynes and R.P. Van Duyne, "Nanosphere lithography: a versatile nanofabrication tool for the study of size-dependent nanoparticle optics," *J. Phys. Chem. B* **105**, 5599–5611 (2001).

[243] J.G.C. Veinot, H.Y. Veinot, S.M. Smith, J. Cui, Q. Huang, and T.J. Marks, "Fabrication and properties of organic light-emitting 'nanodiode' arrays," *Nano Lett*. **2**(4), 333–335 (2002).

[244] Y.N. Xia, B. Gates, Y. Yin, and Y.Lu, "Monodispersed colloidal spheres: Old materials with new applications," *Adv. Mater*. **12**, 693–713 (2000).

[245] J.M. McLellan, M. Geissler, and Y.N. Xia, "Edge spreading lithography and its applications to the fabrication of mesoscopic gold and silver rings," *J. Am. Chem. Soc*. **126**, 10830–10831 (2004).

blocks into a rich variety of lithographically desirable morphological patterns ranging from spheres to cylinders to lamellae, etc.

5.6.11 Stereolithography

Stereolithography, a technique in which light exposure is used to solidify a special liquid resin into a desired 3D shape, has been developed over the last few decades.[246] This is accomplished by polymerizing monomers (such as acrylates) contained in a mold of the desired pattern in a step-wise fashion with UV radiations or electron-beam radiations until the desired pattern is realized. Following polymerization and solidification, the formed pattern is released from the mold. Complex objects with a resolution better than 5 μm in the x, y, and z directions have been fabricated in this manner.[247] The key to this 3D imaging is the use of photoresists that are capable of undergoing two-photon photochemistry. Such resists can respond to the square of the radiation intensity. This necessitates the use of extremely high-power radiation. Remarkably, this approach enables lithographic patterning on nonplanar substrates.

New variants of this technique, called 3D printing, involve maskless printing or depositions of appropriate precursor solutions (hot molten liquid polymers, metals, etc.) into appropriate molds in a layer-by-layer fashion, followed by thermally cooling, hardening, and solidifying the resulting structures, and then releasing the mold.

5.7 Outlook on the Evolution of Lithography

Since its invention, lithography has witnessed phenomenal evolution, exemplified by the myriad techniques and forms in which it is now practiced, and most importantly, by the sheer ubiquity of the devices that it enables, most of which make up the technical inventory of modern life. The dawning of the nanotechnology age has presented new patterning challenges and requirements, which are fueling the outward march of lithography into new scientific and technological areas, beyond its traditional fields of fine art, paper printing, and electronics. The cost, throughput, and resolution of the lithographic patterning process are the main imperatives driving this evolution. Some advanced lithographic techniques currently have resolutions approaching the atomic scale, but the throughputs of their processes make

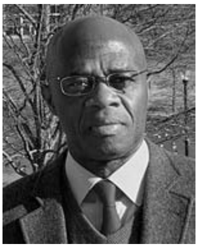
---

[246] K. Ikuta and K. Hirowatari, "Real three-dimensional micro fabrication using sterolithography," in *Proc. IEEE Micro Electro Mechanical Systems (MEMS '93)*, Ft. Lauderdale, Florida, pp. 42–47 (1993); K. Ikuta, K. Hirowatari, and T. Ogata, "Three-dimensional integrated fluid systems (MIFS) fabricated by stereo lithography," in *IEEE International Workshop on Micro Electro Mechanical Systems (MEMS '94)*, Oiso, Japan, pp. 1–6 (1994).

[247] M.J. Madou, *Fundamentals of Microfabrication: The Science of Miniaturization*, 2nd ed., CRC Press, New York, pp. 66–67 (2002).

them unacceptable in a production environment. A good illustration of this is single-atom manipulation in STM lithography. The combination of conventional top-down lithographic patterning and bottom-up hierarchical molecular self-assembly is an approach currently under active exploration and offers the possibility of a path forward to overcome some of the critical limitations associated with advanced lithographic patterning.

Index



Uzodinma Okoroanyanwu is a research associate professor in the department of polymer science and engineering of the University of Massachusetts at Amherst. His research interests span the disciplines of electrochemistry, materials science, and lithography, and are broadly geared toward developing thin film, solid state batteries and microbatteries; and printed, flexible, flexible hybrid, and wearable electronic sensor platforms with applications in environmental monitoring and point-of-care diagnostics. He is also the founder of Enx Labs, a company that translates his research results into devices and instruments that help to improve the human condition and sustain the environment. He previously worked at Advanced Micro Devices, where he spent twelve years conducting research on advanced lithography and organic polymer memories, and at GLOBALFOUNDRIES, where he spent four years conducting research on advanced lithography. He has published extensively on lithography science and technology and on electronic applications of polymers. His books include *Molecular Theory of Lithography* (2015) and *Chemistry and Lithography* (2010), both published by SPIE Press. A holder of 37 U.S. patents, he was educated at The University of Texas at Austin, where he earned the following degrees: Ph.D. in physical chemistry (1997), M.S. in chemical engineering (1995), M.A. in physical chemistry (1994), and B.S. in chemistry and chemical engineering (1991).