Joseph A. Eichmeier
Manfred Thumm
Editors

Vacuum Electronics

Components and Devices

Vacuum Electronics

Joseph A. Eichmeier
Manfred K. Thumm
(Eds.)

Vacuum Electronics

Components and Devices

With 388 Figures

Professor Dr.-Ing. (i.R.) Joseph A. Eichmeier
TU München, Lehrstuhl für Technische Elektronik
Arcisstr. 21, 80290 München, Germany
E-mail: Eichmeier@tum.de; j.eichmeier@gmx.net

Professor Dr. Dr. h.c. Manfred K. Thumm
Forschungszentrum Karlsruhe, Institut für Hochleistungsimpuls- und Mikrowellentechnik
Postfach 3640, 76021 Karlsruhe, Germany
E-mail: manfred.thumm@ihm.fzk.de

Library of Congress Control Number: 2007934267

ISBN 978-3-540-71928-1 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com
© Springer-Verlag Berlin Heidelberg 2008

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Data prepared by Vtex using a Springer LaTEX macro package
Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper SPIN 10934111 57/3180/vtex 5 4 3 2 1 0

Preface

In 1989 the "Handbuch der Vakuumelektronik" (Handbook of Vacuum Electronics) was published in the R. Oldenbourg Verlag, München, Wien, edited by J. Eichmeier and H. Heynisch. During the past 18 years considerable progress has been achieved in the fields of vacuum electronic components, systems and procedures. In the present book a group of 36 well-known experts from industry, scientific institutes and universities report about the fundamentals, state-of-the-art, recent developments and their own experiences in all important areas of Vacuum Electronics.

The operation of vacuum electronic components and devices is based on the motion of electrons or ions under the influence of electric, magnetic or electromagnetic fields. Presently, research activities are concentrated on microwave tubes, especially travelling wave tubes, klystrons, gyrotrons and cross-field devices; switching tubes and vacuum relays; photomultipliers, image converter and image amplifier tubes; vacuum and plasma panel displays; electron and ion beam systems; particle accelerators; electron and ion microscopes; ion and plasma propulsion systems; light sources and gas lasers; X-ray tubes; gas discharge systems; vacuum electronic systems for semiconductor technology; and finally, vacuum technology and vacuum measuring techniques.

Vacuum electronic components, systems and procedures are applied in information, measuring and control engineering and also in high frequency, nuclear, plasma and biomedical engineering. This book presents the current state-of-the-art of vacuum electronics in research, development and production. Its reader should have a basic knowledge of electronics, especially the principles of vacuum electronics. It may be a useful source of proper information for students of electrical engineering and physics and also for engineers and physicists who develop or apply vacuum electronic components and devices in a variety of technological fields. The literature cited is the most appropriate for further reading. All the chapters begin with a historical note, and the references credit the scientists and engineers who were responsible for major breakthroughs and advancements in the various fields of vacuum electronics.

On behalf of all co-authors, the editors thank Springer-Verlag, Berlin, Heidelberg, New York, for the careful lay-out and printing of this book. The editors thank all co-authors for their excellent contributions.

Munich, Karlsruhe | *J.A. Eichmeier*
January 2008 | *M.K. Thumm*

Contents

List of Contributors xv

1 Microwave Tubes 1
G. Faillon, G. Kornfeld, E. Bosch, and M.K. Thumm 1
1.1 Introduction 1
1.1.1 Review of State-of-the-Art and Present Situation 1
1.1.2 Historical Development 1
1.1.3 Basic Operating Principles and Definitions 4
1.2 Klystrons 12
1.2.1 Klystron Amplifiers 12
1.2.2 Multibeam Klystrons 20
1.2.3 Inductive Output Tubes (IOT) 22
1.3 Traveling Wave Tube (TWT) 24
1.3.1 Introduction 24
1.3.2 TWT Design and Operation Principle 25
1.3.3 TWT Physics 29
1.3.4 TWT Applications 45
1.4 Extended Interaction Klystron EIK 50
1.4.1 Introduction 50
1.4.2 Extended Interaction Circuit Design 50
1.4.3 Typical Performance and Applications 50
1.5 Backward Wave Oscillator (BWO) 52
1.5.1 Introduction 52
1.5.2 BWO Operation Principle 52
1.5.3 BWO Applications 54

1.6 Magnetrons and Cross-Field Amplifiers ... 55
1.6.1 Magnetrons ... 55
1.6.2 Cross-Field Amplifiers (CFA) ... 58
1.6.3 Cross-Field Backward Wave Oscillator (MBWO) ... 60
1.7 Fast-Wave Devices ... 61
1.7.1 Interaction Principles ... 63
1.7.2 Dispersion Diagrams of Fast-Wave Interactions ... 65
1.7.3 Gyrotron Oscillator ... 70
1.7.4 Gyro-Amplifiers ... 74
1.7.5 Gyro-BWO ... 78
1.7.6 Free-Electron Maser (FEM) ... 78
1.8 Future Trends and Applications ... 80
References ... 82

2 Vacuum Displays ... 85
K. Blankenbach, G. Gassler, and H.W.P. Koops ... 85
2.1 Introduction ... 85
2.1.1 Historical Development ... 85
2.1.2 Overview ... 87
2.1.3 Fundamentals of Displays ... 89
2.1.4 Fundamentals of Display Parameters ... 93
2.1.5 Comparison of Various Display Technologies ... 99
2.2 Cathode Ray Tubes (CRT) ... 99
2.2.1 Historical Development ... 100
2.2.2 Electrophysical Fundamentals ... 101
2.2.3 Present State-of-the-Art ... 107
2.2.4 Future Aspects ... 111
2.3 Plasma Panel Displays (PDP) ... 111
2.3.1 Historical Development ... 112
2.3.2 Electrophysical Fundamentals ... 112
2.3.3 Present State-of-the-Art and Applications ... 115
2.3.4 Future Aspects ... 117
2.4 Vacuum Fluorescent Displays (VFD) ... 117
2.4.1 Historical Development ... 118
2.4.2 Electrophysical Fundamentals ... 118
2.4.3 Present State-of-the-Art and Applications ... 119
2.4.4 Future Aspects ... 122
2.5 Field Electron Displays (FED) ... 122
2.5.1 Historical Development ... 122
2.5.2 Electrophysical Fundamentals ... 122
2.5.3 Present State-of-the-Art and Applications ... 123
2.5.4 Future Aspects ... 124
References ... 125

3 Radiation Sensitive Vacuum Electronic Components and Devices 127
J.A. Eichmeier 127
3.1 Historical Development 127
3.2 Electrophysical Fundamentals 128
3.2.1 Photoelectron Emission 128
3.2.2 Secondary Electron Emission 130
3.2.3 Electron Optics 131
3.3 Present State-of-the-Art and Applications 132
3.3.1 Secondary Electron Multipliers 132
3.3.2 Image Converters and Image Amplifiers 139
3.3.3 Television Camera Tubes 143
3.4 Future Aspects of Radiation Sensitive Vacuum Electronic Components . . . 153
References 154

4 Electron Beam Devices for Materials Processing and Analysis 155
H. Bluhm, B. Han, A.G. Chmielewski, D. von Dobeneck, U. Gohs, J. Gstöttner, G. Mattausch, H. Morgner, H.W.P. Koops, A. Reichmann, O. Röder, S.W. Schulz, B. Wenzel, and O. Zywitzki 155
4.1 Introduction and History 155
4.1.1 Electron Optics 157
4.1.2 Electron Sources 164
4.2 Thermal Materials Processing 171
4.2.1 Welding 173
4.2.2 Thermal Surface Modification 178
4.2.3 High-rate Evaporation by Electron Beam Guns 181
4.2.4 Electron Beam Melting and Refining in Vacuum Metallurgy 184
4.3 Non-Thermal Processing Techniques 189
4.3.1 Survey of Chemical Effects Caused by Electron Beams 189
4.3.2 Electron-beam Lithography 191
4.3.3 Electron Beam-induced Processing 195
4.3.4 Curing of Organic Coatings and Surface Layers 198
4.3.5 Refinement of Polymer Materials 201
4.3.6 Control of Pathogens 202
4.3.7 Flue Gas and Wastewater Treatment 205
4.4 Materials Analysis, Imaging and Testing 206
4.4.1 Scanning Electron Microscopy 206
4.4.2 Electron Microscopy 209
4.4.3 Electron Beam Spectroscopy and Analysis 218
4.4.4 Electron Beam Testing 219
References 224

5 Ion Beam Devices for Material Processing and Analysis 231
F. Rüdenauer, H.W.P. Koops, G. Hobler, L. Palmetshofer, and H. Bluhm 231
5.1 Ion Beam Analysis 231

5.1.1 Historical Development 231
5.1.2 Electrophysical Fundamentals 232
5.1.3 Analysis Techniques 232
5.2 Ion Beam Materials Modification and Mask Repair 236
5.2.1 Historical Development 236
5.2.2 Electrophysical Fundamentals 237
5.2.3 Present State-of-the-Art and Applications 238
5.2.4 Future Aspects 245
5.3 Ion Implantation 245
5.3.1 Historical Development 245
5.3.2 Electrophysical Fundamentals 247
5.3.3 Present State-of-the-Art and Applications 248
5.3.4 Future Aspects 253
5.4 High Power Ion Beams 253
5.4.1 Historical Development 253
5.4.2 Introduction 254
5.4.3 Accelerator Physics 254
5.4.4 Beam Transport 259
References 259

6 Ion Propulsion Systems 265
H. Bassner, R. Killinger, J. Mitterauer, F. Rüdenauer, N. Koch, and G. Kornfeld 265
6.1 Electrostatic Ion Thrusters and RF-Ion Thruster RIT 265
6.1.1 Introduction 265
6.1.2 Historical Development of RIT 266
6.1.3 Electrophysical Fundamentals 266
6.1.4 Present State-of-the-Art and Applications of RIT 267
6.1.5 Applications of RIT Thrusters 268
6.1.6 Future Aspects 271
6.2 Field Emission Electric Propulsion (FEEP) 272
6.2.1 Historical Development 272
6.2.2 Electrophysical Fundamentals 273
6.2.3 Present State-of-the-Art and Applications 274
6.2.4 Future Aspects 276
6.3 Hall Effect Thrusters 278
6.3.1 Introduction 278
6.3.2 Operational Concept and Characteristics 279
6.3.3 Examples for HET Applications in Space Missions 285
6.4 High Efficiency Multistage Plasma Thruster 288
6.4.1 Introduction 288
6.4.2 Design and Physical Operation Principle 289
6.4.3 Design and Performance 293
6.4.4 Future Developments and Potential Applications 296
References 297

7 Infrared, Light, Ultraviolet, Laser- and X-ray-Tubes 303
A. Ulrich, M. Born, H.W.P. Koops, H. Bluhm, and T. Jüstel 265
7.1 General Physics of Photon Generation 303
7.1.1 Historical Development 303
7.2 Laser 307
7.2.1 Introduction 307
7.2.2 Specific Laser Devices 308
7.2.3 Ion Lasers 310
7.2.4 Excimer Lasers 311
7.2.5 Electron Beam Pumping of High Power KrF-laser 313
7.3 Smith–Purcell Effect 315
7.3.1 Historical Development 315
7.3.2 Electrophysical Fundamentals 315
7.3.3 Present State 316
7.3.4 Future Aspects for THZ Sources 317
7.4 Millimetre and Infrared Light Sources 317
7.4.1 Historical Development 317
7.4.2 Generation of IR Radiation 318
7.4.3 IR Applications 318
7.4.4 Generation of THZ Radiation 319
7.4.5 THz Applications 319
7.4.6 Further Developments 321
7.4.7 Detection of THz Radiation 322
7.5 Visible Light Sources 322
7.5.1 History and Introduction 322
7.5.2 Incandescent Lamps 323
7.5.3 Gas Discharge Lamps 324
7.5.4 Fluorescent Coatings 326
7.5.5 High-Pressure Discharge Lamps 329
7.5.6 General Light Source Survey 334
7.6 Ultraviolet Light Sources 334
7.6.1 Introduction 334
7.6.2 Thermal Tungsten Lamps for Intensity Calibration 335
7.6.3 Deuterium Lamps 335
7.6.4 UV Arc Lamps 336
7.6.5 Mercury Lamps 337
7.6.6 Hollow Cathode Lamps 338
7.6.7 Excimer Light Sources 338
7.6.8 Excimer Lamps Using Discharge Excitation Glow Discharges 340
7.6.9 Excimer Lamps Using Electron Beam Excitation 342
7.7 X-ray-Tubes 343
7.7.1 History and Generation of X-rays 343
7.7.2 Generation of X-rays 343
7.7.3 X-ray Filters 344
7.7.4 X-ray Dosimetry 345

7.7.5 X-ray Tubes 346
7.7.6 Synchrotrons 348
7.7.7 X-ray Detection 349
7.7.8 Applications 349
7.7.9 Future Developments: Miniaturized X-ray Tubes 350
References 350

8 Particle Accelerators 355
M. Seidel and K. Zapfe 355
8.1 History and Types of Accelerators 357
8.1.1 Static Accelerators 357
8.1.2 Circular Resonant Acceleration 358
8.1.3 Linear Resonant Acceleration 361
8.1.4 Acceleration by Induction (Betatron) 363
8.1.5 Particle Sources 364
8.1.6 Colliders 365
8.1.7 Synchrotron Radiation Sources 368
8.2 Essentials of Beam Dynamics 369
8.2.1 Beam Emittance and Phase Space 369
8.2.2 Equations of Motion and Strong Focusing 372
8.2.3 Synchrotron Radiation 376
8.3 Vacuum Requirements in Particle Accelerators 379
8.3.1 Beam Gas Interaction 379
8.3.2 Static and Dynamic Aspects of Accelerator Vacuum Systems 383
8.4 Layout of Accelerator Vacuum Systems 386
8.4.1 Pressure Profile Calculations for One-dimensional Systems 387
8.4.2 Pumping 389
8.4.3 Instrumentation 393
8.4.4 Material Selection and Design Implications 394
8.4.5 Cleaning and Treatment of Vacuum Chambers 397
8.5 Special Topics of Particle Accelerator Vacuum Systems 398
8.5.1 Accelerators with Cold Bore Vacuum Systems 398
8.5.2 Sychrotron Radiation Facilities 401
8.5.3 Particle Free UHV-Systems 402
References 404

9 Vacuum Interrupters 407
R. Renz 407
9.1 Historical Development 407
9.2 Physical Fundamentals 409
9.2.1 Interrupting Capability 409
9.2.2 Dielectric Properties 414
9.2.3 Current-Zero Effects 416
9.2.4 Mechanical and Thermal Aspects 416

9.3 Present State-of-the-Art and Applications 417
9.3.1 Vacuum Interrupter Design and Technology 417
9.3.2 Medium Voltage Circuit Breakers 420
9.3.3 Medium Voltage Contactors 421
9.3.4 Low Voltage Circuit Breakers 422
9.3.5 Low Voltage Contactors 422
9.3.6 High Voltage Vacuum Breakers 423
9.3.7 Load Breakers 423
9.3.8 Transformer Tap Changers 424
9.3.9 Other Applications 425
9.4 Future Aspects 426
References 427

10 Vacuum Electron Sources and their Materials and Technologies 429
G. Gaertner and H.W.P. Koops 429
10.1 Thermionic Vacuum Electron Sources 429
10.1.1 Historical Development 429
10.1.2 Fundamentals of Thermionic Emission 431
10.1.3 Types of Thermionic Cathodes, Properties and Applications 435
10.2 Non-Thermionic Electron Sources 450
10.2.1 Historic Development 450
10.2.2 Field Emission Electron Cathodes 451
10.2.3 Thermo-field Emission Electron Cathodes 453
10.2.4 Cold Field Electron Emission Cathodes 454
10.2.5 Novel Cathode Materials 456
10.2.6 Trends of Novel Developments 457
10.2.7 Metal–Insulator–Metal (MIM) Emitter Materials 459
10.2.8 Diamond Hopping Electron Emitters 459
10.2.9 Carbon Nanotube Emitters 461
10.2.10 Other Carbon Emitters and Future Development Trends 461
10.3 Other Electron Emitters 463
10.3.1 pn-emitters 463
10.3.2 Secondary Emission 466
10.3.3 Ferroelectric Electron Emission 467
10.3.4 Photo-electron Emission 467
10.4 Vacuum Electronic Construction 470
10.4.1 Spacers for FEA’s 470
10.4.2 Anodic Bonding 470
10.4.3 Vacuum Generation in On-Chip Vacuum Electronic Devices . . . 472
10.5 Materials of Vacuum Electron Sources 473
References 478

11 Vacuum Technology 483
W. Knapp 483
11.1 Introduction and History 483
11.2 Gases in Vacuum Systems 483
11.2.1 The Fundamentals 483

11.2.2 The Ideal Gas Law 484
11.2.3 The Pressure Ranges in Vacuum Technology 485
11.2.4 Surface Interactions and Outgassing 487
11.3 Vacuum Pumps 487
11.3.1 Introduction 487
11.3.2 Pumps for Rough and Medium Vacuum 488
11.3.3 Pumps for High Vacuum (HV) and Ultrahigh Vacuum (UHV) 490
11.3.4 Pump Selection 497
11.4 Vacuum Pressure Measurement 497
11.4.1 Introduction 497
11.4.2 Total Pressure Gauges 498
11.4.3 Partial Pressure Gauges 506
11.5 Vacuum Materials and Components 508
11.5.1 Material Properties for Vacuum Applications 508
11.5.2 Component Standards and Constructions 511
11.5.3 Cleaning Procedures 512
11.6 Leak and Leak Detection 514
11.6.1 Introduction 514
11.6.2 Leak Testing Methods 514
11.6.3 Helium Leak Detectors 515
11.6.4 Leak Testing of Vacuum Microelectronic Devices 517
11.7 Examples of Vacuum Systems 517
11.7.1 UHV System for Field Emission Characterisation 517
References 519

Index 521

List of Contributors

Helmut Bassner
Niemöllerallee 8
81739 München, Germany
helmut.bassner@t-online.de

Karlheinz Blankenbach
Pforzheim University
Display Laboratory
Tiefenbronnerstr. 65
75175 Pforzheim, Germany
kb@displaylabor.de

Hansjoachim Bluhm
Forschungszentrum Karlsruhe
Institut für Hochleistungsimpuls-
und Mikrowellentechnik
Postfach 3640
76021 Karlsruhe, Germany
hansjoachim.bluhm@ihm.fzk.de

Matthias Born
Philips Technology Research
Laboratories
Weisshausstr. 2
52066 Aachen, Germany
matthias.born@philips.com

Ernst Bosch
Thales Electron Devices GmbH
Söflingerstr. 100
89077 Ulm, Germany
ernst.bosch@thalesgroup.com

Andrzej G. Chmielewski
Institute of Nuclear Chemistry
and Technology
16 Dorodna Street
03-185 Warsaw, Poland
a.chmielewski@ichtj.waw.pl

Dietrich von Dobeneck
pro-beam GmbH
Behringstr. 6
82152 Planegg bei München, Germany
dietrich.dobeneck@pro-beam.de

Joseph A. Eichmeier
Technische Universität München
Lehrstuhl für Technische Elektronik
Arcisstr. 21
80290 München, Germany
eichmeier@tum.de

Georges Faillon
formerly at Thales Electron Devices
2, rue Latécoère
F-78941 Vélizy, France
georges.faillon@wanadoo.fr

Georg Gärtner
Philips Technology Research
Laboratories
Weisshausstr. 2
52066 Aachen, Germany
georg.gaertner@philips.com

Gerhard Gassler
Samtel Group Germany
Söflingerstr. 100
89075 Ulm, Germany
gerhard.gassler@samtelgroup.de

Uwe Gohs
Fraunhofer Institute for Electron Beam
and Plasma Technology
Winterbergstr. 28
01277 Dresden, Germany
uwe.gohs@fep.fraunhofer.de

Jürgen Gstöttner
Atmel Germany GmbH
Theresienstr. 2
74072 Heilbronn, Germany
juergen.gstoettner@hno.atmel.com

Bumsoo Han
EB-TECH Co., Ltd
Yongsan-dong 550, Yuseong-gu
Daejoen 305-500, Republic of Korea
bshan@eb-tech.com

Gerhard Hobler
Technische Universität Wien
Institut für Festkörperelektronik
Floragasse 7/362
A-1040 Wien, Austria
gerhard.hobler@tuwien.ac.at

Thomas Jüstel
University of Applied Sciences Münster
Stegerwaldstr. 39
48656 Steinfurt, Germany
tj@fh-muenster.de

Rainer Killinger
EADS-Astrium
Langer Grund
74239 Möckmühl, Germany
Rainer.killinger@astrium.eads.net

Wolfram Knapp
Otto von Guericke University
Magdeburg
Institute for Experimental Physics
Universitätsplatz 2
39106 Magdeburg, Germany
wolfram.knapp@physik.uni-magdeburg.de

Norbert Koch
Thales Electron Devices GmbH
Söflingerstr. 100
89077 Ulm, Germany
norbert.koch@thalesgroup.com

Hans W.P. Koops
HaWilKO PSS
Ernst Ludwig Str. 16
64372 Ober-Ramstadt, Germany
hans.koops@t-online.de

Günter Kornfeld
Thales Electron Devices GmbH
Söflingerstr. 100
89077 Ulm, Germany
guenter.kornfeld@thalesgroup.com

Gösta Mattausch
Fraunhofer Institute for Electron Beam
and Plasma Technology
Winterbergstr. 28
01277 Dresden, Germany
goesta.mattausch@fep.fraunhofer.de

Johannes Mitterauer
Technische Universität Wien
Institut für Photonik
Gusshausstr. 27–29
A-1040 Wien, Austria
joh.mitterauer@nextra.at

Henry Morgner
Fraunhofer Institute for Electron Beam
and Plasma Technology
Winterbergstr. 28
01277 Dresden, Germany
henry.morgner@fep.fraunhofer.de

Leopold Palmetshofer
Johannes Kepler Universität Linz
Abteilung für Festkörperphysik
Altenbergerstr. 69
A-4040 Linz, Austria
leopold.palmetshofer@jku.at

Axel Reichmann
Fraunhofer Institute for Electron Beam
and Plasma Technology
Winterbergstr. 28
01277 Dresden, Germany
axel.reichmann@fep.fraunhofer.de

Roman Renz
Siemens AG
Power Transmission and Distribution
Rohrdamm 88
13623 Berlin, Germany
roman.renz@siemens.com

Olaf Röder
Fraunhofer Institute for Electron Beam
and Plasma Technology
Winterbergstr. 28
01277 Dresden, Germany
olaf.roeder@fep.fraunhofer.de

Friedrich Rüdenauer
Austrian Research Centers/FM
Boltzmannstr. 1
A-2444 Seibersdorf, Austria
fritz.ruedenauer@aon.at

Siegfried W. Schulz
(decreased in 2003)
Surface Solutions, Inc.
7989 Ranchers Road
Fridley, MN 55432, USA
info@tincoat.net

Mike Seidel
Paul Scherrer Institut
CH-5232 Villigen, Switzerland
mike.seidel@psi.ch

Manfred K. Thumm
Universität und Forschungszentrum
Karlsruhe
Institut für Hochleistungsimpuls-
und Mikrowellentechnik
Postfach 3640
76021 Karlsruhe, Germany
manfred.thumm@ihm.fzk.de

Andreas Ulrich
Technische Universität München
Fakultät für Physik E 12
James-Franck-Straße
85748 Garching bei München, Germany
andreas.ulrich@ph.tum.de

Bernd Wenzel
Von Ardenne Anlagentechnik GmbH
Plattleite 19/29
01324 Dresden, Germany
wenzel.bernd-dieter@vonardenne.biz

Kirsten Zapfe
Deutsches Elektronen-Synchrotron
Notkestr. 85
22607 Hamburg, Germany
kirsten.zapfe@desy.de

Olaf Zywitzki
Fraunhofer Institute for Electron Beam
and Plasma Technology
Winterbergstr. 28
01277 Dresden, Germany
olaf.zywitzki@fep.fraunhofer.de

1

Microwave Tubes

G. Faillon, G. Kornfeld, E. Bosch, and M.K. Thumm

1.1 Introduction

1.1.1 Review of State-of-the-Art and Present Situation

For more than 60 years, microwave tubes are used in many applications as oscillators and, in particular, as powerful amplifiers of electromagnetic waves at frequencies ranging from about 300 MHz to several hundreds of GHz, with some even reaching the 1 to 2 THz range.

A number of microwave tube types exist, classified on the basis of their operating frequency and output power. At the same time, they may be divided into two general categories: pulse wave tubes and continuous wave (CW) tubes. The power/frequency relationships of the most commonly used microwave tubes are shown in Fig. 1.1.

Microwave tubes may be broadly used in three main areas of applications:

- radio, TV and telecommunications;
- radars and military systems (especially airborne);
- industrial, scientific, medical (ISM).

Even though microwave tube performance is generally described in terms of power, efficiency and gain, other factors of specific and professional interests to each user should also be taken into consideration: bandwidth, linearity, signal/noise ratio, mode of propagation, tube and transmitter weight, reliability, etc.

1.1.2 Historical Development

One century ago, in 1904, Sir J.A. Fleming discovered the valve, or the diode, immediately used to detect the electromagnetic waves recently demonstrated by H.R. Hertz. In fact, the diode was the first vacuum tube, with a poor vacuum, although vacuum. A few months later, in 1906, Lee de Forest has been inspired to

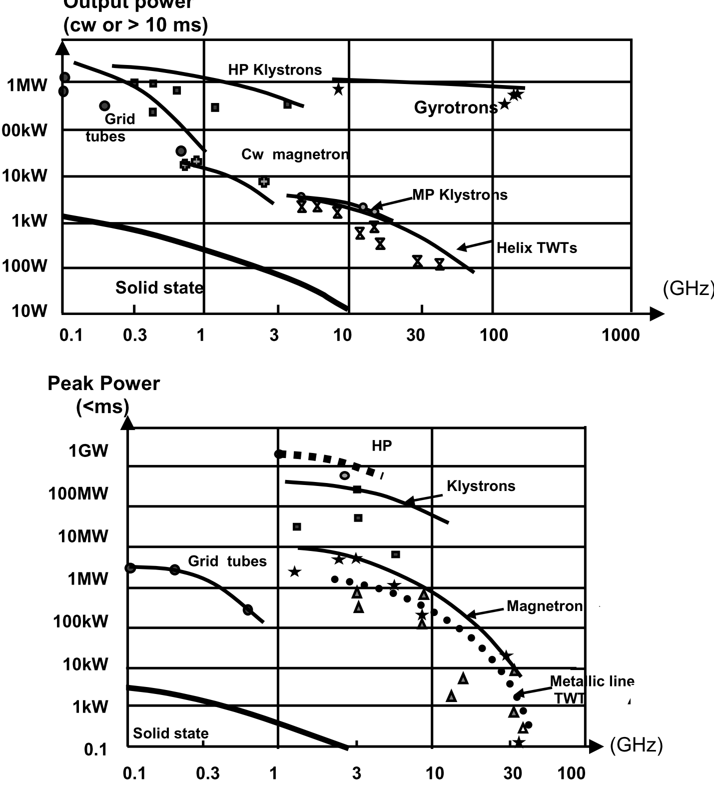


Fig. 1.1. Peak and CW power of various microwave tubes vs frequency

add a 3rd electrode, and in that way the triode was born. After about 10 years of tests and discussions, this new vacuum tube has been used as a very attractive amplifier, which gave rise, in the 1920s, to a rapid expansion of the broadcasting and the radio-links.

At the same time, research was carried not only to improve the triodes, but also to consider other vacuum devices at higher frequencies. In 1920, H. Barkhausen invented the retarding-field tube (or reflex triode), which can be regarded as the first transit time tube. The magnetron has been discovered by A.W. Hull in 1921, followed by E. Habann and A. Zacek. Twelve years later, K. Posthumous understood the operating principles of such an oscillator, and in 1939 H.A.H. Boot and J.T. Randall used "klystron" type resonators for confining the RF fields. The introduction of the oxide cathode with a high secondary emission coefficient was also a major improvement. During World War II, the urgent need for high power microwave generators for radar transmitters led to the development of the magnetron to its present state.

In 1953 W.C. Brown introduced the crossed-field amplifier (CFA), a magnetron with an interrupted RF circuit to provide input and output connections. At the same

time the backward oscillator (MBWO), or “carcinotron”, derived from the CFA but with an injected beam and a specific delay line came into light on both sides of the Atlantic Ocean. This oscillator offered the advantage to be electrically frequency tunable over a very large band.

Nevertheless, neither the triodes nor the magnetrons were high frequency and large gain amplifiers. In 1934–35, A. & O. Heil were the first scientists to imagine the use of a periodic and localized electron velocity variation in order to get bunches. After the registration of patents in 1937, Hahn & Metclaff and especially R. & H. Varian gave the exact description of the multicavity klystron. The first prototype delivered $\approx$ 50 W at 3 GHz with a gain of 30 dB, and very soon 30 to 50 dB gain 1 to 10 GHz klystrons have been widely used, for instance, on radars and particle accelerators. Born from the principles of this klystron and from previous studies on electron reflectors free of microwave fields, the “reflex klystrons” made great strides as local oscillators in radars.

The possibility of traveling wave interaction has been described in 1942 by R. Kompfner, who developed the first TWT, one year later, using a helix as the delay line. But the real starting development of the TWTs took place after 1946, when J.R. Pierce made the theory of these tubes and gave solutions to suppress many parasitic oscillations. The coupled cavity TWT appeared in 1950. The evolution of the TWTs is impressive, when we know that they are still nowadays used – in modern versions – for many purposes.

From that time, and because of important military and civilian needs, the microwave tubes entered into their industrial period. Of course the R&D was strongly going on, not only to improve the present devices and to replace them, according the new user requirements – higher frequencies, longer lifetime, greater power and bandwidth – but also to compete with the transistors and the solid state devices or, especially now, to collaborate with them.

The last 40 years period can be characterized by the following main events [1]:

- Coaxial magnetron, frequency agility magnetron.
  Mass production of magnetrons in the 1960s to provide the microwave ovens.
- Brazed and pressed helix TWTs, mainly used up to the 1980s in radar transmitters and ground radio-links.
  Variable pitch helix TWTs improving efficiency.
  Gridded electron guns and introduction of the impregnated cathodes in the 1970s.
  Then, development of TWT depressed collectors and radiating collectors.
  From 1962 space TWTs are used on satellites.
- Tunable klystrons or TV and communications. 10% instantaneous bandwidth klystrons. Inductive output tubes (IOTs).
  Multibeam medium power klystrons, mainly used in soviet countries from 1960 to 1980.
  Vapor cooled collectors, and 300 kV pulsed electron guns.
  High efficiency ($\geq$65%), pulsed power (tens of MW) and CW ($\geq$1.3 MW) klystrons. Multibeam high power klystrons (MBKs) from 1995.

- Recently born gyrotrons, which are the only tubes capable of delivering very high power (MW) at very high frequencies ($\geq$100 GHz).
  In 1959, without knowledge of the astrophysicist R.Q. Twiss's work, J. Schneider in USA and A.V. Gaponov in Russia proposed an explanation for the amplifying mechanism based on free electron gyro radiation. But the first successful experiments took place in the 1970s and became really attractive in the 1980s, thanks to the support of the thermonuclear fusion plasma community, especially in Europe. And now after having solved many severe technological problems, such as the superconducting electromagnets and the sapphire or diamond windows, the gyrotrons and the associated low loss overmoded waveguides or quasi-optical transmission lines are manufactured and their development is still progressing.

From about 20 years, the market configuration is slowly changing and the requirements are now more specific, more severe and exacting, to the detriment of mass production. Nevertheless, after the merging and the reorganization of many tube manufacturers, the applications are mainly oriented as shown in Sect 1.1.

At the same time new research and developments are conducted in the direction of better performance [1] as higher frequencies and greater power, but also towards advanced technologies, such as the field emission and cold cathodes, nanotechnologies, compactness and optimization of the whole amplifiers.

1.1.3 Basic Operating Principles and Definitions

Basic Operating Principles

A microwave tube may be defined as an evacuated envelope (vacuum), inside which an electron beam interacts with an electromagnetic wave [2, 3]. This interaction means that the electrons of the beam give up a part of their kinetic or potential energy to the electromagnetic wave, thereby generating or amplifying this wave.

Since it is relatively known how to generate and then to accelerate or, in other words, to give energy to an electron beam – especially thanks to the high voltage electron guns – the objective of this book is to go further and to present the physics behind the interaction and the transfer of the beam energy to the electromagnetic wave. Such a process involves several separate or simultaneous physical phenomena as follows.

1. Formation and acceleration of an electron beam.
2. Periodic bunching of the electrons at a frequency f. This bunching is started up by the RF input or drive power P_{d} in the case of an amplifier, or by the electromagnetic noise in the case of an oscillator.
3. Deceleration of the bunches (or reduction of their relativistic mass) in such a way that their kinetic or potential energy is converted into an electromagnetic or a microwave energy at the frequency f.
4. Forwarding of this microwave energy outside the tube, which yields the tube's output power P_{OUT}.

Characteristics and Definitions

To operate a microwave tube [4], the first step is to heat the electron-emitting cathode by connecting the attached filament to a source of electric power $P_F = V_F I_F$. Then a power supply ($P_0 = V_0 I_0$) is connected between the cathode and the anode in order to generate the required energetic electron beam, which will travel at a velocity v_0, carrying a current I_0, given by the two non-relativistic expressions (the relativistic ones will be given in the following pages)

$$mv_0^2/2 = eV_0, \tag{1.1}$$

$$I_0 = PV_0^{3/2}, \tag{1.2}$$

with P being a factor related to the geometry of the gun, called perveance.

A magnetic field is used to focus the beam [5]. This field is created either with permanent magnet(s) or just by an electromagnet with a power consumption P_{FOC}. In the case of an oscillator, there is no RF input signal, and the tube directly converts the beam energy to microwave energy at the frequency f. Let P_{OUT} denote the tube output power. Then the overall tube efficiency is given by $\eta = P_{OUT}/(P_F + V_0 I_0 + P_{FOC})$, while the interaction efficiency is $\eta_{interaction} = P_{OUT}/V_0 I_0$.

The oscillation does not begin immediately with the application of the voltage V_0, but after a fluctuating starting time τ_1, as shown in Fig. 1.2. Then, during the rise time τ_2, the output power increases up to the nominal level P_{OUT}. The frequency f changes during this time τ_2 and then stabilizes.

The power P_{OUT} is dependent on the anode current I_0 and on the load impedance Z_L, as shown in the Rieke diagram of Fig. 1.2. A pushing ratio $\Delta f/\Delta I_0$ and a pulling ratio $\Delta f/\Delta\varphi_L$, measured at constant reflection coefficient from the load (or standing wave ratio, SWR), characterize the frequency sensitivities of the oscillator tube. φ_L is the phase of the load impedance Z_L. It is sometimes useful to lock or to control the oscillation frequency f by means of a controlling signal $P_c < P_{OUT}$ at a frequency f_c (close to f), which is injected into the tube via its output. For a given P_c setting, the Adler relationship gives the total range of frequencies over which this frequency locking is possible, $\Delta f_{max} = 2f_c/Q_x\sqrt{P_c/P_{OUT}}$.

In the case of an amplifier, a microwave signal of power P_d, at frequency f, is injected at the tube input and is then amplified up to P_{OUT}. The amplifier overall efficiency is determined in practically the same way as for oscillators, $\eta = P_{OUT}/(P_F +$

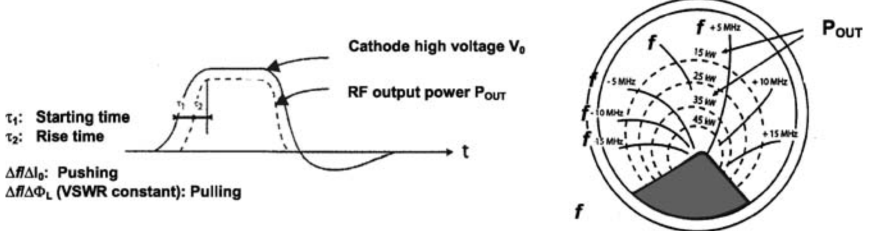


Fig. 1.2. Definitions: oscillators

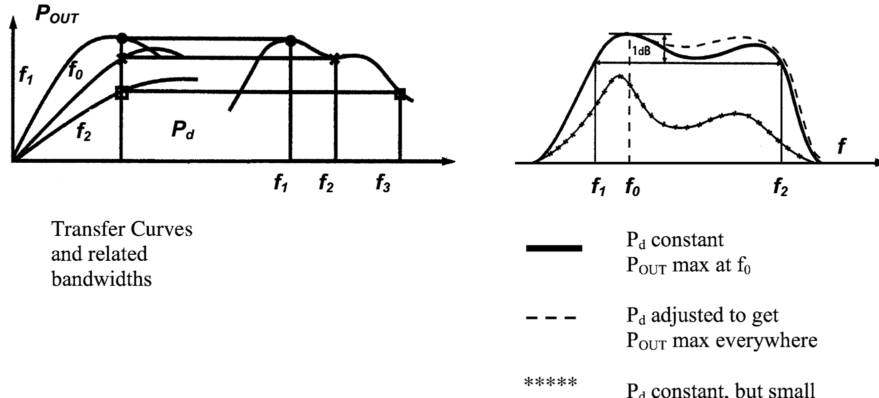


Fig. 1.3. Definitions: amplifiers

$V_0 I_0 + P_{\mathrm{FOC}} + P_{\mathrm{d}}$), and the interaction efficiency again is $\eta_{\mathrm{interaction}} = P_{\mathrm{OUT}}/V_0 \cdot I_0$. The ratio $G = P_{\mathrm{OUT}}/P_{\mathrm{d}}$ is the gain, which is often written in decibels (dB) as $G = 10\log_{10}(P_{\mathrm{OUT}}/P_{\mathrm{d}})$. Usually, the $P_{\mathrm{OUT}}/P_{\mathrm{d}}$ transfer curve (Fig. 1.3) shows an almost linear relationship at low signal levels and a saturation part at large signals. The bandwidth, generally measured at -1 dB, is defined on the basis of the variation of P_{OUT} versus the frequency, with P_{d} adjusted, at each frequency, to maximize P_{OUT}. The bandwidth can also be defined with P_{d} constant and, for instance, adjusted to the value giving maximum P_{OUT} (Fig. 1.3).

The group delay $\tau_{\mathrm{g}} = \mathrm{d}\varphi/\mathrm{d}\omega$ represents the signal delay inside the tube amplifier, $\varphi(f)$ being the phase shift between the input and the output. The influence of the load impedance Z_{L} is given by the Rieke diagram where the $P_{\mathrm{OUT}}(Z_{\mathrm{L}})$ curves are plotted.

Basic Physical Laws of E-Beams

Equation of Motion and Relativistic Corrections

The only non-negligible forces which influence electrons in microwave tubes during their interaction with the RF wave are the electromagnetic Lorentz forces. All other forces, e.g. due to its spin or the gravitation, are typically 11 or 16 orders of magnitude lower, respectively.

From the relativistic equation of motion of a single electron

$$\mathrm{d}(m_{\mathrm{e}} \cdot \boldsymbol{v}_{\mathrm{e}})/\mathrm{d}t = -e(\boldsymbol{E} + \boldsymbol{v}_{\mathrm{e}} \times \boldsymbol{B}), \tag{1.3}$$

$$m_{\mathrm{e}} = \frac{m_0}{\sqrt{1-(v_{\mathrm{e}}/c)^2}} = m_0(1 + V/V_{\mathrm{n}}), \tag{1.4}$$

$$v_{\mathrm{e}} = c\sqrt{1 - 1/(1 + (V/V_{\mathrm{n}}))^2} \tag{1.5}$$

(with the charge $e = 1.60 \times 10^{-19}$ C, the relativistic electron velocity v_e, the relativistic mass m_e, the rest mass of the electron $m_0 = 9.11 \times 10^{-31}$ kg, the velocity of light c, the acceleration voltage V and $V_n = m_0c^2/e = 511$ kV, the acceleration voltage to produce the equivalent of one electron rest mass as relativistic mass increase), we can easily conclude that the kinetic energy of an electron (or another charged particle) can only be changed by the electric field $\boldsymbol{E}$ but not by the magnetic field $\boldsymbol{B}$, because the magnetic Lorenz force vector $\boldsymbol{F}_L = -e(\boldsymbol{v} \times \boldsymbol{B})$ is perpendicular to both, the velocity $\boldsymbol{v}$ and $\boldsymbol{B}$, and because the scalar product of two orthogonal vectors is always zero. Therefore, the energy transfer from an electron beam to a RF wave or vice versa employs always only the electric field components of the RF wave and space charge bunches

$$\mathrm{d}E_{\mathrm{kin}} = \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{s} = -e(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \cdot \boldsymbol{v}\,\mathrm{d}t = -e\boldsymbol{E} \cdot \boldsymbol{v}\,\mathrm{d}t, \tag{1.6}$$

where $\mathrm{d}\boldsymbol{s}$ is the line path element along the trajectory. This does not mean that the magnetic fields $\boldsymbol{B}$ are not very important. Conversely, magnetic fields are required to maintain the focused beam properties, as we will see in Sect. 1.3.

Maxwell Equations

Since in a beam the number of electrons is far beyond the computation power of modern computers, the particle picture can be replaced for many considerations by the continuum theoretical approach of the Maxwell equations. Here the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ are self-consistently determined by the local charge and current densities ρ and $\boldsymbol{j}$, which themselves depend on a set of differential equations, boundary and start conditions on the fields $\boldsymbol{E}$ and $\boldsymbol{B}$. The Maxwell equations cannot describe localised collision processes or even the passing of particles, which occurs in RF tubes close to saturation. Nevertheless, they allow drawing helpful general conclusions and symmetry considerations and are also the platform for a lot of computer simulations and modelling.

$\mathrm{div}\,\boldsymbol{E} = \rho/\varepsilon_0;$	Charge density is a source of $\boldsymbol{E}$-fields;	(1.7)
$\mathrm{rot}\,\boldsymbol{E} = -\delta\boldsymbol{B}/\delta t;$	Induction law;	(1.8)
$\mathrm{div}\,\boldsymbol{B} = 0;$	Magnetic field is source free, no magnetic single poles;	(1.9)
$\mathrm{rot}\,\boldsymbol{B} = -\mu_0\boldsymbol{j} + \varepsilon_0\mu_0\delta\boldsymbol{E}/\delta t$	$(= -\mu_0\,\boldsymbol{j}$ for stationary beams).	(1.10)

From the stationary Maxwell equations one can conclude that the electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ can be derived from a *scalar potential* ϕ and a *vector potential* $\boldsymbol{A}$ in the form

$$\boldsymbol{E} = -\,\mathrm{grad}\,\phi, \tag{1.11}$$

$$\boldsymbol{B} = \mathrm{rot}\,\boldsymbol{A}. \tag{1.12}$$

Inserting these definitions into the Maxwell equations (1.7) and (1.10), one obtains equations for those potentials defined by the space charge density ρ and the beam current density $\boldsymbol{j}$, respectively:

$$\text{div grad}\,\phi = -\rho/\varepsilon_0; \qquad \text{Poisson's equation (the Laplace equation for } \rho = 0); \tag{1.13}$$

$$\text{rot rot}\,\boldsymbol{A} = -\mu_0 \boldsymbol{j}. \tag{1.14}$$

Since the influence of the beam current density $\boldsymbol{j}$ on the magnetic field can be in most applications completely neglected compared to the externally applied magnetic fields, the fourth Maxwell equation (1.10) reduces for the stationary case simply to

$$\text{rot}\,\boldsymbol{B} = 0; \qquad \text{For negligible current contributions to } \boldsymbol{B}\text{-field}; \tag{1.15}$$

$$\boldsymbol{B} = -\mu_0\,\text{grad}\,\Omega; \qquad \text{with } \boldsymbol{B}\text{-field defined by scalar potential } \Omega; \tag{1.16}$$

$$\text{div grad}\,\Omega = 0; \qquad \text{Magnetic equivalent to the electric Laplace equation.} \tag{1.17}$$

In this case, like for the scalar electric potential ϕ, one can define a magnetic scalar potential Ω, and (1.17) becomes the magnetic equivalent to the electric Laplace equation (1.13).

Busch's Theorem

As a direct consequence of Maxwell equations in axially symmetric (rotational symmetric) systems, the Busch theorem is very useful for the analysis of magnetically focused linear electron beams (Fig. 1.4). It states a surprising conservation law, valid along an electron trajectory, by saying that the sum of its angular momentum $m_e v_\theta r$ with respect to the z-axis and the magnetic flux Φ parallel to the symmetry axis through a circle area with the trajectory radius r and multiplied by the factor $e/2\pi$ is a constant.

$$\begin{aligned} m_e v_\theta r + e/2\pi \cdot \Phi &= \text{constant}; \\ m_e \dot{\theta} \cdot r^2 + \frac{e}{2\pi} \Phi &= \text{constant}; \end{aligned} \qquad \text{Busch's theorem, valid along trajectories of charged particles in rotational symmetric systems (Fig. 1.4).} \tag{1.18}$$

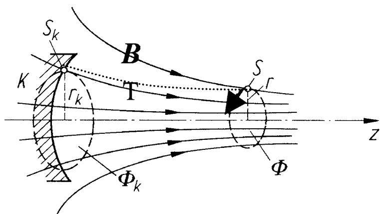


Fig. 1.4. Illustration for Busch's theorem in rotational symmetric systems

Since the angular velocity is zero at cathode emission, according to the Busch theorem (1.18) the angular velocity v_θ or $\dot{\theta}$ in S is just defined by the two flux values Φ_K, Φ and the radius r (with $\dot{\theta}$ = angular velocity around z-axis). There are a lot of other applications for the Busch theorem. Some of them will be introduced in the chapter on beam focusing.

Scaling Laws

Geometrical scaling of microwave tubes is often appropriate when an approved existing design at a given frequency f_1 has to be transferred to a new frequency f_2.

Since the tube dimensions scale with the RF wavelength λ or the reciprocal of the frequency f, the tube scaling factor γ for all linear dimensions is

$$\gamma = \lambda_2/\lambda_1 = f_1/f_2; \qquad \text{Scaling with frequency.} \tag{1.19}$$

Now the question arises how the electromagnetic properties of the tube have to be scaled in order to get identical beam trajectories in the scaled coordinate system $\boldsymbol{x}_2$?

The answer is obtained with the help of the Maxwell equations. It can be easily shown by insertion that the Maxwell equations remain invariant in the scaled, new coordinate system $\boldsymbol{x}_2$ when the following scaling laws are applied.

$$\boldsymbol{x}_2 = \gamma \cdot \boldsymbol{x}_1; \qquad \text{Linear geometrical scaling of all dimensions in system } \boldsymbol{x}_1 \text{ with the factor } \gamma \text{ in the new coordinate system } \boldsymbol{x}_2; \tag{1.20}$$

$$\boldsymbol{E}_2 = \gamma^{-1} \cdot \boldsymbol{E}_1; \qquad \text{Electric field;} \tag{1.21}$$

$$\boldsymbol{B}_2 = \gamma^{-1} \cdot \boldsymbol{B}_1; \qquad \text{Magnetic field;} \tag{1.22}$$

$$\phi_2 = \gamma^0 \cdot \phi_1; \qquad \text{Electric potential;} \tag{1.23}$$

$$\boldsymbol{A}_2 = \gamma^0 \cdot \boldsymbol{A}_1; \qquad \text{Vector potential;} \tag{1.24}$$

$$\boldsymbol{j}_2 = \gamma^{-2} \cdot \boldsymbol{j}_1; \qquad \text{Current density;} \tag{1.25}$$

$$I_2 = \gamma^0 \cdot I_1; \qquad \text{Total current (from gun);} \tag{1.26}$$

$$P_2 = \gamma^0 \cdot P_1; \qquad \text{Perveance } P = I \cdot V^{-3/2}. \tag{1.27}$$

With the above scaling laws, we can easily understand why the power handling capability P_{OUT} of a given microwave tube design scales roughly with frequency f^2 or $P_{OUT} f^2$ = constant.

Let us assume a tube scaled with $\gamma = 1/2$ for the doubled frequency $f_2 = 2f_1$. If all the electrode potentials are kept constant, also the total currents and power consumptions are kept constant. But since the current densities are scaling with γ^{-2}, we get a 4 times higher thermal loading at the critical areas. Therefore, to maintain the same material loading at the thermal material limits, the power handling capability in the doubled frequency case is roughly only ¼. A reference for the topic of scaling is given in [6].

Beam Formation in the Electron Gun

First, we consider the emission of electrons at the cathode surface. Now mainly Ba-, Ca-Aluminates impregnated dispenser cathodes are used. Since they are in detail described in Chapter 10, we limit ourselves here to reproduce the emission laws of this type of thermionic emitters described by workfunction ϕ and cathode temperature T.

Thermal Cathode Emission

In the temperature limited emission regime (all emitted electrons are drawn away from the cathode surface), the saturated emission current density j_s is given by the Richardson–Dushman equation (1911; Nobel Prize 1928)

$$j_s = A * (1 - r) \cdot T^2 \cdot \mathrm{e}^{-\phi/kT}, \tag{1.28}$$

where $A = 4\pi e m_0 k^2/h^3 = 120\,A/(\mathrm{cm}^2\,\mathrm{K}^2)$, k is the Boltzmann constant and the quantum-mechanical reflection coefficient r at the metal/vacuum surface is set to 0.

A more refined elaboration by Schottky includes the electric field E at the surface and is called the Richardson–Dushman–Schottky equation

$$\begin{aligned} j_s &= A * T^2 * \mathrm{e}^{-\phi/kT} * \mathrm{e}^{Ka*E^{1/2}} \\ \text{with } Ka &= (e^3/4\pi\varepsilon_0)^{1/2}/kT. \end{aligned} \tag{1.29}$$

Space Charge Limited Emission

Normally, due to space charge limitation the maximum emission current j_s is not obtained during operation of a linear beam microwave tube. The gyrotron cathodes only are operating in the temperature limited mode. The electron charges, emitted into a certain volume, defined by the boundary conditions for the electrostatic potential ϕ, modify the potential distribution inside this volume. As a consequence, this volume is transparent or *perveant* only for a limited current, the space charge limited emission current. The following gives the space charge limited emission for a planar diode configuration (Fig. 1.5):

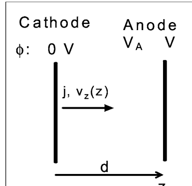


Fig. 1.5. Planar diode configuration

$$j_0 = 2.335 \times 10^{-6} \cdot V_a^{3/2}/d^2;$$ Child–Langmuir law for space charge limited current density in a diode; (1.30)

$$I_0 = F/d^2 \cdot 2.335 \times 10^{-6} \cdot V_a^{3/2};$$ Child–Langmuir law for total current;

$$P = F/d^2 \cdot 2.335 \times 10^{-6};$$ P = diode perveance; F = diode area. (1.31)

It is important to note that this space charge limited current does not depend neither on the cathode temperature nor the nature of the emitting cathode (workfunction ϕ), but only on the applied voltage and the gun geometry; here is the diode distance d and F is the diode area (see Figs. 1.6 and 1.7). The result (1.30) can be generalised as

$$I_0 = P \cdot V_a^{3/2}; \qquad \text{Child–Langmuir law for total current.} \qquad (1.32)$$

We remind of the result of the scaling laws (1.27) that the perveance P remains constant in case of a linear scaling in all dimensions.

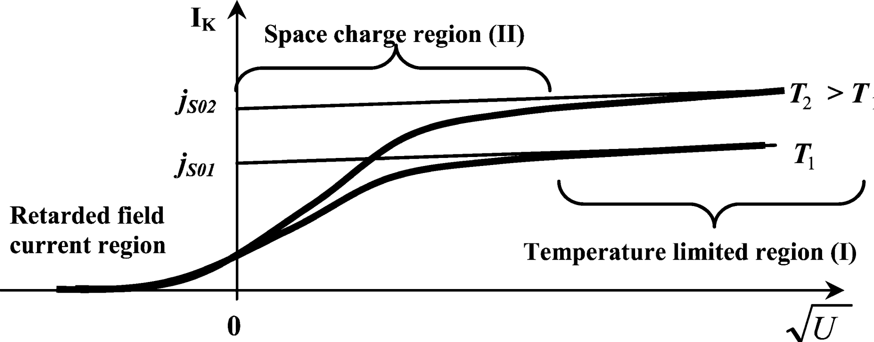


Fig. 1.6. Cathode emission vs the square root of applied anode voltage. For sufficiently high voltages the current is limited by the Richardson–Dushmann–Schottky saturation current

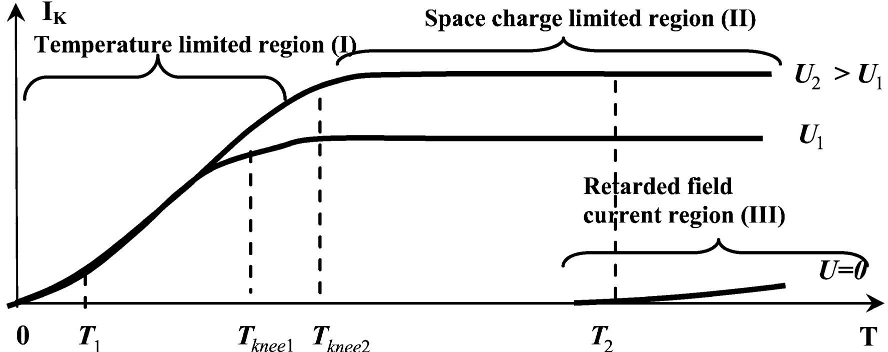


Fig. 1.7. Cathode emission vs the temperature. For low temperatures of the cathode the emission is in the temperature limited regime. At the so-called knee temperature the emission becomes space charge limited and independent of temperature

1.2 Klystrons

1.2.1 Klystron Amplifiers

Introduction

Klystrons are amplifier microwave tubes, designed mainly for use at high frequencies from about 0.3 GHz to approximately 30 GHz. They are characterized by high gains (around 50 dB) and high output powers:

- 3 kW CW tunable from 14 to 14.5 GHz with efficiency $\eta = 40\%$,
- 60 MW with 4 μs pulses at 3 GHz ($\eta = 38\%$),
- 1.3 MW CW at 352 MHz with η going up to 65%,
- 500 kW CW at 3.7 GHz with $\eta = 45\%$.

However, almost all klystron instantaneous bandwidths are fairly limited.

In a klystron (Fig. 1.8), the main functions are separated, which means that design and technology factors can be optimized for each function. These functions are the electron emission, the magnetical beam focusing, the electron bunching or the

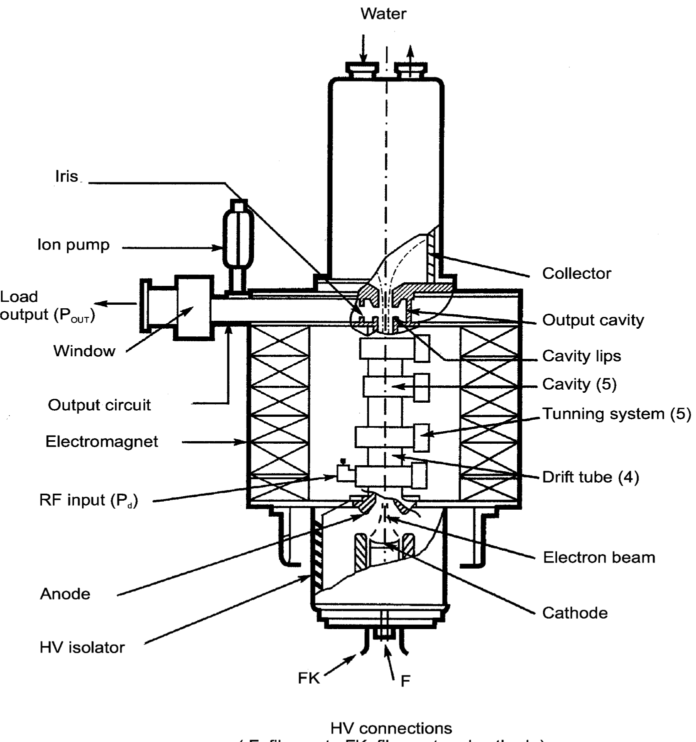


Fig. 1.8. Klystron cross-section

beam density modulation, the extraction of the microwave energy and, lastly, the dissipation of residual energy.

Interaction in a Klystron: Modulations and Energy Extraction

Thanks to a high voltage $-V_0$ applied to the cathode in regard the anode, a Pierce type electron gun generates an electron beam which carries the current $I_0 = PV_0^{3/2}$ according to (1.2) and which is maintained approximately cylindrical thanks to a magnetic field $B_z(z)$ over the entire length of the tube.

After leaving the anode, the electrons travel across the first cavity, or the input cavity. This cavity is excited by the weak input signal P_d which is to be amplified and whose frequency f is about f_0, the cavity's resonance frequency. This cavity is designed to resonate at its fundamental mode TM_{110} (rectangular) or TM_{010} (circular) with a maximum electric field E_1 at the center where the electrons pass [7]. Moreover, this electric field is increased by the fact that the drift tubes, placed on both sides of the cavity, are reentrant (Fig. 1.9).

Electrons traveling across this narrow gap of the cavity are under action of the periodic electric field $E_1 \mathrm{e}^{\mathrm{j}\omega t}$. They are accelerated for a half of period and decelerated for the other half (Fig. 1.10). Consequently, the fast moving electrons from a given half-period catch up with slow moving electrons from the preceding half-period, resulting in the creation of electron bunches. In other words, the velocity modulation produced in the first cavity creates a beam current modulation in the following drift tube, which is expressed as $I_\mathrm{b}(z,t)$ [8, 9].

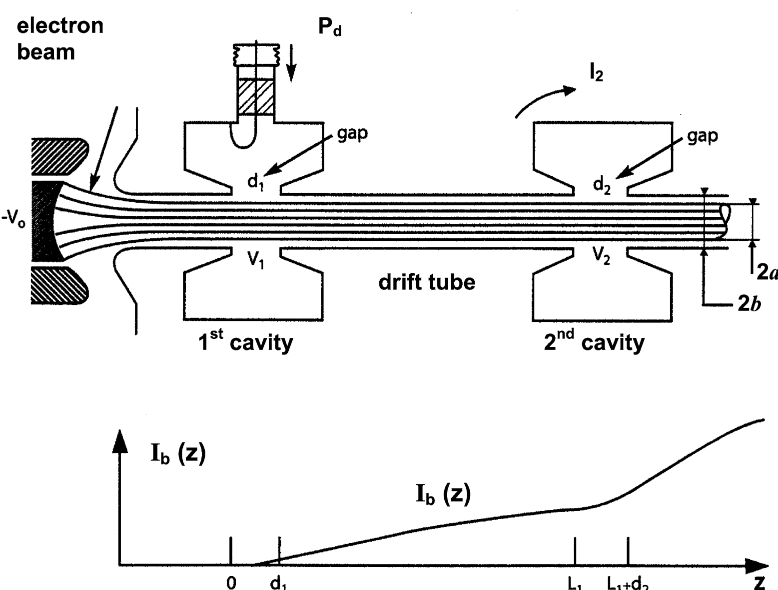


Fig. 1.9. Beam modulation along a klystron

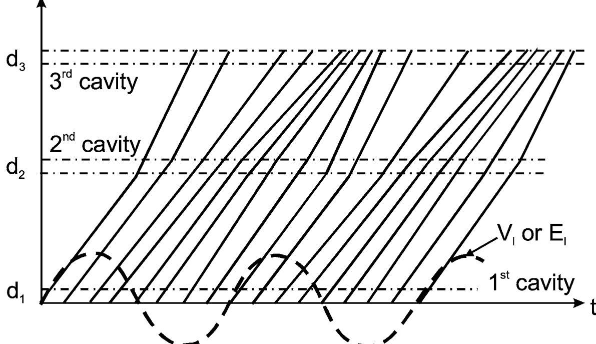


Fig. 1.10. Applegate's diagram $z(t)$

A second cavity (Fig. 1.9) is then added at the distance L_1 where the bunches form. The beam current $I_b(z,t)$ induces a current I_2 in this cavity which gives rise to a voltage $V_2 = Z_2 I_2$ (or an electric field E_2) at the gap ends, Z_2 being the cavity impedance. On its turn V_2 (or E_2) gives rise to a new velocity modulation, generally to an extent far exceeding the former modulation. Again, a third cavity is added at a distance L_2, etc.

The process is repeated at each drift tube and each cavity along the klystron, until the last, or the output, cavity is reached where the bunches are very narrow and dense. In this last cavity the very high induced current I_N gives rise to the large cavity voltage $V_N = Z_N I_N$. Then the bunches are strongly decelerated by this voltage V_N (or E_N) so that they are themselves generating. In that way, they give up a large part of their kinetic energy to the electromagnetic field which builds up in this cavity and maintains V_N (or E_N). This energy stored in the cavity yields the output power $P_{OUT} = V_N^2/2Z_L$ routed to the load, through an iris or a coupling loop and a window (Figs. 1.8 and 1.11). P_{OUT} represents the power P_d which is amplified by the klystron. Typically, klystrons have four to six cavities, gains (P_{OUT}/P_d) of about 50 dB, and interaction efficiencies ($P_{OUT}/V_0 I_0$) of about 35% to more than 65%.

From the First to the Second Cavity

Assuming several simplifications, the relationship between V_1 and P_d is given by

$$V_1^2 = 8\left(\frac{R}{Q}\right) Q(Q/Q_x) P_d \left[1 + Q^2\left(\frac{f}{f_0} - \frac{f_0}{f}\right)^2\right]^{-1} \tag{1.33}$$

with f_0, the cavity resonance frequency, and f, the operating frequency [8, 9, 11].

Solving the equation for the movement of an electron entering the gap of the cavity at a velocity v_0 (1.1) yields the velocity $v(d_1)$ at the output of the gap,

$$v(d_1) = v_0\left(1 + \frac{M_1 V_1}{2V_0} \mathrm{e}^{\mathrm{j}\omega t}\right), \tag{1.34}$$

where t is the time when the electrons pass the center of the gap and

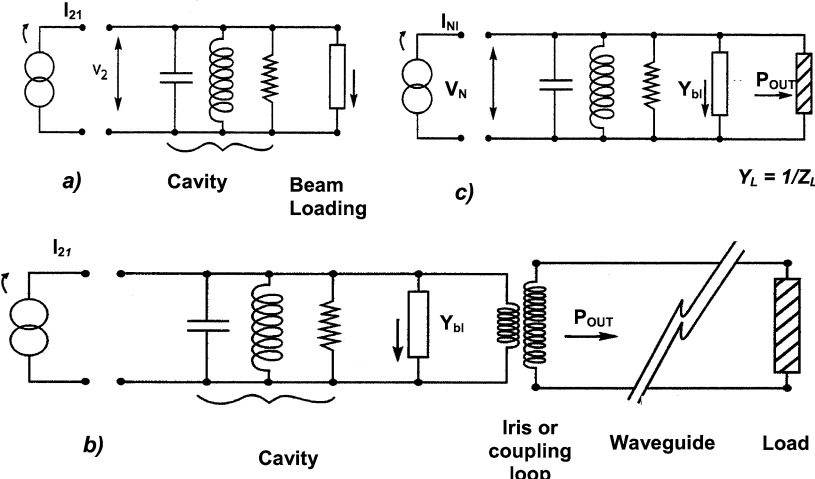


Fig. 1.11. Equivalent circuit of a klystron cavity

$$M_1 = \sin\left(\frac{\omega d_1}{2v_0}\right) \Big/ \left(\frac{\omega d_1}{2v_0}\right).$$

M_1 is called the coupling coefficient. The electrons leave the cavity and travel in the following drift tube at the velocity given by (1.34).

Due to the periodicity of the modulation (Fig. 1.10), their respective positions change over time (some electrons bunch, other separate) and bunches form, once per period. The velocity modulation caused by the first cavity therefore results in electron spatial density modulation, i.e. current modulation in the drift tube. This current along the z-axis is given by

$$I_b(z,t) = I_0 + 2I_0 \sum_{1}^{\infty} J_n(nX)\cos n(\omega t - \omega z/v_0) \tag{1.35}$$

with $X = M_1 \cdot (V_1/2V_0)(\omega z/v_0)$, J_n = Bessel function, I_0 = DC or average beam current, and M_1 given just above. Equation (1.35) shows harmonics. The first harmonic ($n = 1$) maximum is $1.16I_0$ and is located at a distance $z = L_1$ from the center of the gap of the first cavity such that $X = 1.84$. The second harmonic ($n = 2$) maximum is $0.96I_0$ and it is located at $z = L_2$ such that $X = 1.54$ [8, 9]. The second cavity is usually located near $z = L_1$ where the electron bunches are strongly formed.

Contrary to the first cavity entrance, both the electron velocity $v(L_1, t)$ from (1.34) and the current $I_b(L_1, t)$ from (1.35) at the entrance of the second cavity change over time and are time-modulated. The beam current $I_b(z, t)$ travels inside the gap (width d_2) of the second cavity. $I_b(z, t)$ is an instantaneous current along z at time t.

Each electron induces a positive image charge in the walls at the ends of the drift tubes and especially in the cavity. Thus, at time t, the current $I_2(t)$ induced inside

the cavity is the sum of all the charges induced by the electrons in the gap at time t. This can be expressed mathematically by stating that $I_2(t)$ is the integral of $I_b(z,t)$ from L_1 to $L_1 + d_2$.

After calculations, $I_2(t)$ comprises the three main terms:

- I_0, the average current, which does not interact with the cavity;
- $$I_{21} = M_2 \cdot I_{b1} e^{j\omega t} \tag{1.36}$$
  with $I_{b1} e^{j\omega t}$ = the first harmonic of the beam current $I_b(z,t)$ at $L_1 + d_2/2$ and M_2 the coupling coefficient given by (1.38) below;
- $$I_{22} = jG_0(V_2/D_2)[\exp(-jD_2/2)][\cos(D_2/2) - \sin(D_2/2)/(D_2/2)] \tag{1.37}$$
  with $G_0 = I_0/V_0$ and $D_2 = \omega d_2/v_0$. The voltage V_2 induced in this second cavity (between the ends of gap d_2) is given further in the Sect. 1.2.1.

Behaviour of the Second Cavity

The coupling coefficients, for instance M_2, are always <1 and express the fact that the field $E_2 = -V_2 e^{j\omega t}/d_2$ experienced by an electron is a function of time and changes during the traveling of this electron through the gap.

Up to now, the above calculations assumed an ideal gap with imaginary grids [9]. If we now consider a real gap with the electric field extending inside the adjacent drift tubes on the left and right and varying in both the axial and radial directions, the general expression of M_2 is given by

$$M_2 = \frac{1}{\int_0^a \int_{-\infty}^{\infty} E_2(z,0) r \, dr \, dz} \int_0^a \int_{-\infty}^{+\infty} E_2(z,r) e^{j\omega z/v} r \, dr \, dz, \tag{1.38}$$

where $\int_{-\infty}^{+\infty} E_2(z,0) \, dz = V_2$.

The narrower the gap d and the smaller the radii a or b, the more M_2 approaches 1, and only when $D_2 \geq 1.2$ rad and $\omega b/v_0 \geq 1$ rad, M_2 drops off to any large extent. These factors determine the orders of magnitude of the gap d_2, the drift tube diameter $2b$, and the electron beam width $2a$, given that the filling coefficient $a/b \approx 0.6$ and that b and d_2 are also related to the R/Q values of the cavity.

I_{22}/V_2 looks like a conductance $Y_{bl} = G_{bl} + jB_{bl}$. This is known as the beam loading conductance. A positive G_{bl} value means that an energy is taken from the modulated electron beam, whereas a negative one means that an energy is given to the electron beam. G_{bl} is at a maximum for $D_2 = \pi$ and becomes negative after $D_2 = 2\pi$. Thus, if a cavity has a gap such that $D_2 = (\omega d_2/v_0) > 2\pi$, G_{bl} is negative, $G_{bl} + G_c$ can be also negative, and an oscillation (known as "monotron" oscillation) can occur. G_c is the conductance representing the cavity losses. $\omega = 2\pi f$ is related not only to the operating fundamental mode of the cavity, but also to anyone of the numerous resonances at higher frequencies.

The expressions given above determine the equivalent schema for the second cavity as for any of the other intermediate cavities. This equivalent schema is an oscillating circuit, as represented in Fig. 1.11a, driven by the induced current I_{21} and loaded by the "losses" conductance G_c and the beam loading conductance Y_{bl}.

Space Charge

Up to now, we have neglected the space charge, i.e. the repelling force between electrons which naturally tends to separate them, thereby hindering bunching. If we disturb an electron beam in a state of equilibrium by moving an electron in relation to the other ones, this electron is firstly repelled by its neighbours and then returns to its position of equilibrium, which it overshoots, before turning back again towards the equilibrium position, like a swinging pendulum. This movement takes place within a beam moving at a velocity v_0. The frequency of this oscillation phenomenon (known as the plasma frequency $f_p = \omega_p/2\pi$) is given by $\omega_p^2 = (e/m)(\rho_0/\varepsilon_0)$ with $\rho_0 = J_0/v_0$, the beam charge density, and $J_0 \approx I_0/\pi a^2$, the current density [10]. Nevertheless, this expression is somewhat inexact because the true plasma frequency f_q must take into account the metal or conducting walls of the drift tubes.

Taking the space charge into consideration, it changes the second term of $v(d, t)$ in the expression (1.34), which has to be multiplied by $\cos(\omega_q z/v_0)$. Furthermore, the expression (1.35) also changes when the space charge is included. X becomes

$$M_1(V_1/2V_0)\frac{\omega}{\omega_q}\sin\left(\frac{\omega_q z}{v_0}\right).$$

At $\omega_q \Rightarrow 0$, this yields the original expression.

From the Second Cavity to the Output Cavity

The current I_2 creates a voltage $V_2 = Z_2(f)$. I_2 across the gap of the second cavity. This voltage in turn modulates the velocity of the electrons, whereby this modulation is not necessarily in phase with the preceding modulation but is considerably greater in magnitude. In the drift tubes between cavities 2 and 3, the new velocity modulation is transformed on its turn into a density or a current modulation (Fig. 1.10). This iterative process continues until the addition of another cavity has quite no effect on the modulation [9].

This last cavity is the output cavity and is characterized by a high degree of coupling to the external load, that is the user's device to which the tube is connected. In practical terms, a loop or an iris is used for this coupling (Fig. 1.8).

The equivalent circuit presented above applies to all cavities, including the last cavity which is coupled to the load (Fig. 1.11b and Fig. 1.11c). If we now assume that the beam loading expression has no imaginary part ($Y_{bl} = 1/Z_{bl} = G_{bl}$) and the load too ($Y_L = 1/Z_L = G_L = 1/(R/Q)Q_L$), this yields

$$Y_{\mathrm{N}} = (1 + 2jQ_{\mathrm{N}}\Delta)/(R/Q)Q_{\mathrm{N}} \tag{1.39}$$

with $\Delta \approx (\omega - \omega_0)/\omega_0 = (f - f_0)/f_0$ and $1/Q_{\mathrm{N}} = 1/Q_{\mathrm{c}} + 1/Q_{\mathrm{bl}} + 1/Q_{\mathrm{L}}$. Then

$$V_{\mathrm{N}} = I_{\mathrm{N}}/Y_{\mathrm{N}}. \tag{1.40}$$

Therefore, P_{OUT} can be written as,

$$P_{\mathrm{OUT}} = \frac{1}{2}V_{\mathrm{N}}^2 G_{\mathrm{L}} = \frac{1}{2}I_{\mathrm{N}}^2 G_{\mathrm{L}} Q_{\mathrm{N}}^2 (R/Q)^2/\left(1 + 4Q_{\mathrm{N}}^2\Delta^2\right). \tag{1.41}$$

This expression can be used to yield the gain $G = P_{\mathrm{OUT}}/P_{\mathrm{d}}$ and the interaction efficiency $\eta = P_{\mathrm{OUT}}/V_0 I_0$.

The methodology we have applied since the Sect. 1.2.1 shows how P_{OUT} can be deduced from P_{d} throughout the formulae (1.33) up to (1.41). However, it is important to bear in mind that these expressions hold only at "weak fields or signals". As soon as P_{d} increases in magnitude (i.e. modulation increases), we are no longer dealing with linear problems and the equation for P_{OUT}, as given by (1.41), becomes more complex, and in fact, even ceases to be analytical. The nonlinearities are primarily due to electron bunching and space charge forces.

This being the case, more complete and complex calculations are required, necessarily drawing on computer processing power. In fact, computer permits us to analyze considerably more than just the non-linear phenomena; it also allows us to incorporate radial movements (neglected in previous calculations), the 3D nature of certain events, relativistic effects and reflected electrons. Schematically speaking, this involves solving the electron movement equation for varying electromagnetic fields and space charge fields.

Output Power Optimization

Most of the time klystrons use 4, 5 or 6 cavities. These cavities gradually increase the RF beam current I_{b}, then I_{N} and P_{OUT}, despite the space charge forces, which become very disturbing when I_{b} is more and more large. At the same time, to optimize P_{OUT}, the velocity dispersion must be kept low, $(v - v_0)/v_0 \leq 10$ to 20%, in order to avoid reflected electrons from the output cavity. Therefore, the intermediate cavities are carefully frequency tuned to adjust the gap voltages in amplitude and in phase and, at the same time, to control the RF beam current and the velocity dispersion.

Some klystrons use a second harmonic cavity which resonates at a frequency slightly lower than $2f_0$. Such a cavity is excited by the second harmonic of the beam current and generates a voltage which reinforces the bunching and the efficiency. This cavity is usually located after the second one, where the electrons are already concentrated in a phase extension $\leq \pi$.

Due to the values of the impedance Z_{L}, the absolute value of $V_{\mathrm{N}} = Z_{\mathrm{N}} I_{\mathrm{N}}$ can exceed V_0, which is the voltage corresponding to the average electron velocity. Were V_{N} to exceed V_0, the electrons would be decelerated, but some would even be reflected back towards the cavities and the cathode, which would generate beam interception and lead to klystron malfunction or even permanent damage.

The output cavity behaves at large signals like a current generator with an internal impedance R^* around the operating point. Therefore, P_{OUT} is maximum when $Z_{\mathrm{L}} = R^*$. Since the coupling iris or the loop transform the load impedance at the user device (radar or transmitter antenna, accelerator structure) to the load impedance at the cavity, these coupling devices have to be adjusted so that $Z_{\mathrm{L}} = R^*$, in order to maximize the output power.

Klystrons Engineering and Technologies

This presentation of different steps of the interaction in a klystron allows us to understand the klystron structure which is designed around the electron beam on one hand and the cavities and the drift tubes on the other hand.

From a technological point of view, as shown in Fig. 1.8, the constitutive elements of the klystron tube are:

- The Pierce type electron gun, especially with the high voltage insulator and the cathode. The high voltage insulator is a cylinder made of alumina located between the anode and the cathode. Its dimensions are such that the DC electric fields are smaller than the breakdown limit. The electron emissions from the triple point are minimised thanks to a screen, called anticorona ring. The anode is grounded for electrical safety reasons, explaining why the gun and the cathode are usually raised to a negative voltage;
- The modulating cavities (usually frequency tunable in a factory);
- The last or output cavity coupled to the output waveguide thanks to a coupling loop or iris;
- The output window(s) which is located in the waveguide(s) and which separates the external atmosphere from the internal vacuum ($\approx 10^{-8}$ Torr) inside the tube. The drive power to be amplified is usually injected inside the first cavity through a coaxial connection which is also vacuum tight thanks to a small input window;
- The collector whose function is to collect the electrons after their interaction in the last cavity and to dissipate their remaining energy. Moreover, it must be able to dissipate the whole electron beam energy, when $P_{\mathrm{d}} = 0$ and then $P_{\mathrm{OUT}} = 0$. Therefore, it has to be efficiently cooled. The collector is usually at the same potential than the body or the anode;
- The magnetic circuit including an electromagnet, or permanent magnet(s), and the two pole pieces located near the anode and between the collector and the output cavity. These pole pieces concentrate the magnetic flux to get the required focusing B_Z parallel to the axis (B_{R}/B_Z better than several ‰).

The vacuum is usually maintained thanks to a small ion getter pump. However, this pump is not always necessary, because the beam itself has also a pumping function. The accelerated electrons ionize the residual gases. Then the generated ions are electrically attracted by the beam and slowly drained toward the cathode.

1.2.2 Multibeam Klystrons

Limitations in the Increase of P_{OUT} in a Klystron

As explained in Sect. 1.2.1, the space charge forces are opposed to a perfect electron bunching. These repulsion forces appear in the form of the plasma frequencies ω_q or ω_p, which is proportional to the square root of the perveance P.

In a conventional klystron, the perveance of the beam is in the range 0.5 to $2.5 \times 10^{-6}\,\mathrm{A\,V^{-3/2}}$. A perveance $P = 2.5 \times 10^{-6}\,\mathrm{A\,V^{-3/2}}$ is the practical limit, beyond which the electron beams are difficult to be maintained fairly cylindrical and to be focused without notable body interception. When the perveance is high, the efficiency is low. At the same time the instantaneous bandwidth is enlarged to ≈5% or more, instead of the usual 1 to 2%. This last point is explained by the high value of the beam loading conductance G_{bl} proportional to $PV_0^{1/2}$, which damps all the cavities.

On the contrary, a perveance $P = 0.5 \times 10^{-6}\,\mathrm{A\,V^{-3/2}}$ is the practical lower limit under which the smallness of the current I_0 dictates unacceptable high beam voltages V_0 and involves many electrical insulation difficulties on the tube itself and on the whole transmitter or equipment. But the low perveances are favourable for high interaction efficiencies because a strong bunching can be achieved.

In conclusion, larger output powers are obtained by increasing either the beam current I_0 or the high voltage V_0. In the first case the perveance is important; the efficiency notably decreases, but the instantaneous bandwidth is enlarged and the limitations are the control and the focusing of the beam. In the second case the perveance is kept low; the efficiency is therefore high, but the main limitations are the breakdowns and the practical use of high voltages.

The Multibeam Klystrons (MBKs)

The MBKs give the possibility to reach larger RF output powers with high efficiencies and, at the same time, acceptable high beam voltages, despite very high perveances. A conventional or one beam klystron (OBK) uses just a single electron beam which travels along the common axis of the tube and the magnet. On the contrary, the MBK, as shown in the Fig. 1.12, uses several (N) electron beams which travel all together through the same cavities, but each one has its own drift tube. Therefore, each electron beam has a low current i_0 giving rise to a high efficiency η. The whole current Ni_0 is important; then, for the same RF output power $P_{OUT} = V_0 \eta N i_0$, the cathode voltage is modest and even small. For example, the number of beams is 6 or 7. The perveance of a single beam is about 0.5 to $0.6 \times 10^{-6}\,\mathrm{A\,V^{-3/2}}$ and the total perveance $P = Ni_0 V_0^{-3/2} \approx 3$ to $4 \times 10^{-6}\,\mathrm{A\,V^{-3/2}}$. The interaction efficiency is around 65%.

In comparison with a OBK, the cathode voltage V_0 is usually reduced by ≈40%. Consequently, the electrical fields E_0 are decreased and the product $E_0 V_0$ (kV^2/mm),

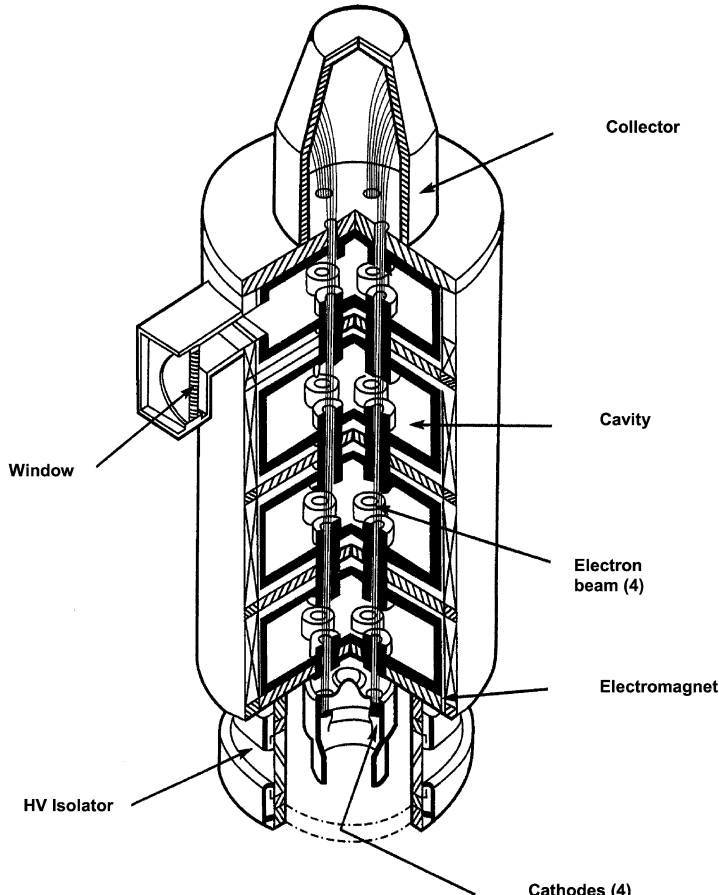
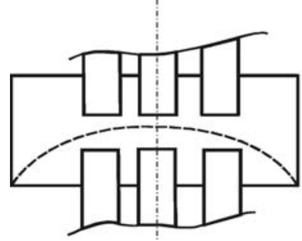


Fig. 1.12. Multibeam klystron cross-section (4 beams)

which interprets the breakdown occurrence, is also reduced by more than 50%, giving rise to the possibility of much more safe and reliable operations in longer pulses or DC voltages.

The cavities are cylindrical. The RF resonant mode is generally the conventional one, the TM^{O}_{010}, and the re-entrant drift tubes, where the beams are passing through, are concentrated around the axis [7], as shown in Fig. 1.13. The RF field pattern allows, of course, an in phase interaction of each electron beam, but also high coupling factors M between the electrons and the RF electric field E, despite a non-negligible variation of E versus the radius and the possible influence of the RF magnetic field.

The overall high perveance of the electron beams increases the conductance of the beam loading of the cavities. Consequently, the instantaneous bandwidths become larger, by a factor of 5. Finally, the cavities are designed in such a way that



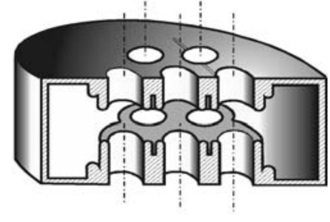


Fig. 1.13. MBK cavity (7 beams)

the adjacent modes are very far from the operating one. The reason is to prevent monotron oscillations (see Sect. 1.2.1), encouraged by the large whole beam current Ni_0.

The focusing of the electron beam is difficult because the beams axes are offset and not identical with the tube axis, generating an important radial magnetic field B_R. Therefore, the electromagnet is equipped with specific multiple pole pieces. Between those the axial field is nearly constant, and the relative radial components B_R/B_Z smaller than a few ‰.

In short, the main characteristics of the MBKs are the significantly lower cathode high voltages and the larger efficiencies and/or the broader bandwidths. The results are:

- Smaller dimensions and size and reduced electrical supplies;
- Lower breakdown or arcing risks. HV insulation in air and less and less in oil;
- Less X-rays parasitic radiations;
- Better reliability.

However, the designer has to take into account the following two difficulties:

- Focusing and transmission of the N beams, $N-1$ of which (at least) being offset in respect to the axes of the tube and the magnetic field;
- Elimination or damping of parasitic oscillations in the cavities, which can present many high order modes.

1.2.3 Inductive Output Tubes (IOT)

In the lower part of the frequency spectrum, for instance around 500 MHz in the TV bands, the ground transmitters are now equipped with Si and SiC transistors in parallel. But the RF output power is limited to several hundreds of W. To achieve much higher output power in the range of tens of kW, vacuum tubes are still used – and for a long time – thanks to their proven reliability and great efficiency. The tubes are triodes, tetrodes [11], klystrons and now mainly IOTs.

The gains of about 13 to 15 dB of triodes and tetrodes are very low. On the other hand, the large size and the unpracticable frequency tuning of the TV klystron cavities are not very attractive. That's why a hybrid tube, the IOT, has been developed about 20 years ago for the main use with transmitters operation in the 470–830 MHz

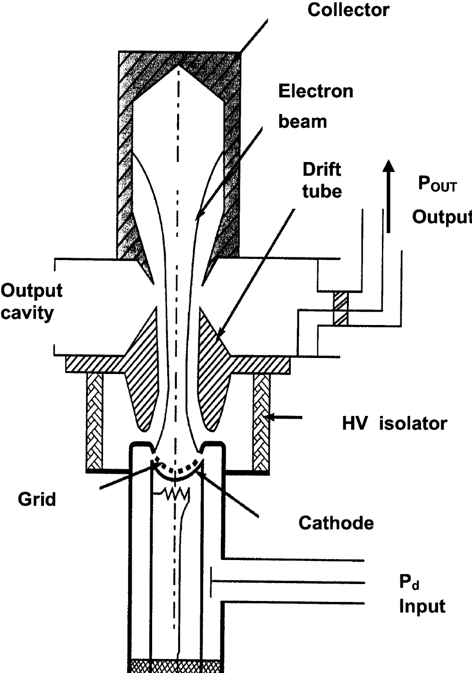


Fig. 1.14. Inductive output tube (IOT)

range, at power of several dozen kW and gains of 20 to 23 dB compatible with solid-state drivers.

IOTs represent a modification from conventional coaxial grid tubes towards a linear beam structure close to the one of the klystrons, but with just one cavity (Fig. 1.14). A Pierce type electron gun fitted with a cathode and a spherical modulation grid G_1 creates a converging electron beam, with an average current I_0, modulated by a periodic voltage V_G at the microwave frequency f.

To avoid any malfunction, no microwave interaction must take place in the large accelerating gap between the grid and the anode. This is why the gun is designed so that no microwave energy enters this gap. Such an energy could come from the tube outside or because of leakages from the coaxial cathode–grid resonator and from the output cavity through the anode. Also, no oscillation may occur in this gap; therefore, microwave resonances must be eliminated.

The electron bunches between the grid and the anode are accelerated by high voltages V_A going up to 30 kV or more, which is considerably higher that the 8–10 kV limit of triodes. After the anode, the bunches are traveling across a klystron type resonating cavity where much of their kinetic energy is transferred to a waveguide or coaxial cable, in order to yield the output microwave power P_{OUT}. The process is exactly the same as in the output cavity of a conventional klystron. It should be

noted that a too high deceleration voltage in the cavity can reflect electrons towards the cathode and damage the tube.

Finally, the remaining energy of the electrons is dissipated in the collector. Because the beam is short and don't include any cylindrical part, the focusing is simplified; just a permanent magnet around the anode is needed.

From a technological point of view, the most critical part is the grid gun, not only because the grid is located very close (≈1 mm) to the hot cathode, but also because the drive power P_d must be injected in the gun assembly which is raised to a high potential of $-V_A \approx -30$ kV. This is done thanks to a coupling loop and RF traps that are difficult to design. By the way we note that, in most of the tubes, the anode is grounded for electrical safety reasons, explaining why the gun and the cathode are raised to $-V_A$.

Before to conclude it should be point out that:

- the presence of a grid close to the hot cathode and the design of the coaxial resonator limit the operating frequencies to 1.5 GHz maximum,
- and the high voltage V_A is practically limited by the necessary use of air – and not of oil – to insulate the gun including the cathode–grid resonator.

Finally, IOTs are very attractive specific grid tubes operating below 1.5 GHz, at high voltages of about 30 kV. The gains are around 22–23 dB and efficiencies around 55%, for microwave output power of 10 to 100 kW.

1.3 Traveling Wave Tube (TWT)

1.3.1 Introduction

Looking back from the beginning of the 21st century to the history and evolution of microwave tubes (as described in Sect. 1.1.2 of this book), we might wonder why the TWT (Lindenblad, 1940; Kompfner, 1942), as the purest realization of the microwave generation principle in electron tubes, was invented after the magnetron (A.W. Hull, 1921, to K. Posthumus, 1933) and after the klystron (R.H. & S.E. Varian, 1937). The limited information exchange during the war might have been the reason why the U.S. patent of Lindenblad [12] was not known by Dr. Rudolph Kompfner, an Austrian refugee working on microwave tubes for the British Admiralty, and why it took again years from their documented inventions of the TWT principle to Kompfners publication [13] in 1946.

In fact, the TWT shows all the necessary elements for RF generation in a tube in a pure, geometrically separated and physically subsequent form. Those elements are electron beam formation and acceleration in an electron gun, power transfer from the electron beam to the amplified RF wave in an interaction section and collection of the decelerated spent electron beam in a collector. Additional elements are the RF input and output couplers at the beginning and the end of the interaction section and the magnetic beam focusing system.

In contrast to its clear physical principle, the TWT employs the most challenging manufacturing technology amongst the microwave tubes. The necessity to focus a high power density electron beam through a tiny helix with length more than 100 times the diameters ranging from 0.3 mm (60 GHz) to 4 mm (1.5 GHz) requests mastering of very small tolerances between beam forming structures and the magnetic focusing system. The fragile delay line structures limit the obtainable RF power compared to the bulk resonator structures of magnetrons or klystrons. The difference in technological maturity and power handling capability at time of TWT appearance is shown in Table 1.1.

Though the basic difference in the power handling capability was remaining over the years (see Fig. 1.1), TWTs have developed since that time, due to their excellent linearity and broadband capability, to the major microwave amplifier used in terrestrial and especially in space telecommunication. Other TWT applications are: coherent microwave sources for radar surveillance and guidance systems, satellite based Earth observation radars and as broadband Electronic Counter Measurement (ECM) amplifiers.

1.3.2 TWT Design and Operation Principle

Figure 1.15 gives a cross-section through a permanent periodic magnet (PPM) focused electrostatic 4 stage collector TWT showing the essential functional elements.

Table 1.1. Output power and application of early C-band magnetrons, klystrons and TWTs around the time of first industrial production of TWTs. Data are from [14–16] for magnetron, klystron and TWT, respectively

Parameter	Magnetron	Klystron	TWT
Year	1942	1949	1952
Manufacturer	Western Electric	Stanford University, Varian	STC
Type	WE 718	Mark III	
Application	Radar	Linear Electron Accelerator	TV, Ground Link Communication
Frequency	2.7–2.9 GHz	2.85 GHz	3.6–4.4 GHz
Output power	193 kW, pulsed	12 MW, 1 μs pulses	2 W, CW

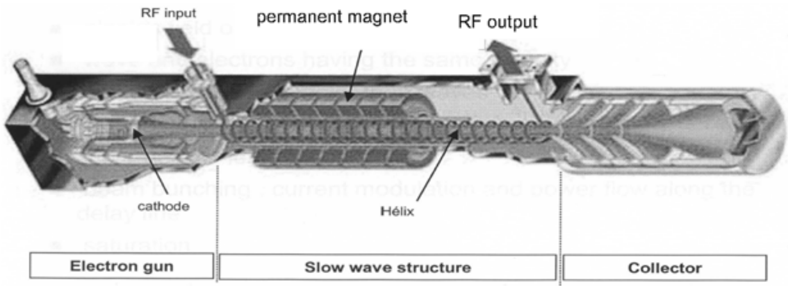


Fig. 1.15. Cut through a PPM focused, electrostatic 4 stage collector TWT

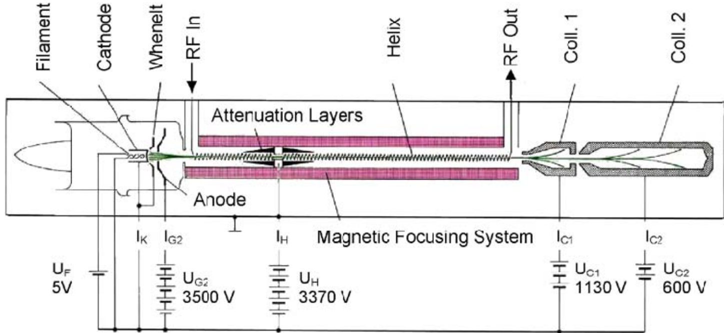


Fig. 1.16. Electric circuit of a CW-TWT with double stage collector

Figure 1.16 provides the electrical circuit needed for its operation (only 2 collector stages are shown). A vacuum envelope, comprising the electron gun, the RF interaction section and the collector is evacuated to the 10^{-9} mbar range. Metal/ceramic brazing technology is used for the electrical feed-through to the electron gun and collector and for the windows at RF in- and output. In the further description we follow the electrons from emission to collection.

Electric Circuit and Electron Gun

To emit the electron current I_K from the cathode surface, a thermionic cathode is heated to its operating temperature T_C of about 1000°C brightness by a filament power supply which provides filament voltage V_F and filament current I_F on a negative cathode potential $V_K = -V_H$. Application of positive voltages V_{G2}, V_H, V_{C1}, V_{C2} let the electrons flow through the vacuum to the respective electrodes and from there in closed circuits through their respective power supplies back to the cathode. Looking closer, they first form an accelerated electron beam which is focused by a PPM focusing system through the helix and is mainly collected by the collector electrodes. The kinetic beam power $P_{\text{beam}} = I_K \cdot V_H$ of the accelerated electron beam entering the delay line section at ground potential had to be spent as potential power by the various electrode power supplies when pumping the negative charged electrons from their electrode potentials V_{G2}, V_H, V_{C1}, V_{C2} to the negative cathode potential $V_K = -V_H$. Thus the major power supplies are the collector power supplies because anode and helix supplies have to provide only very small powers $P_{G2} = I_{G2} \cdot V_{G2}$ and $P_H = I_H \cdot V_H$ due to the small interception currents I_{G2} and I_H.

The electron beam entering the delay line is formed by the electron gun optics to a small diameter electron beam characterized by the area beam convergence factor CF as the ratio of cathode surface to beam cross-section area. For various TWT applications CF can vary in the range 1–100. For very good beam focusing properties CF should be maintained below around 30.

Focusing System and Delay Line/RF Interaction Section

The delay line, as the structure with the largest interface to the TWT environment, is kept on ground potential. This has the advantage that the interception current I_H with the electron beam, which needs to be controlled in safe operational limits, can simply be monitored on ground potential. Here the subscript "H" reminds of "helix", the RF delay line used in many TWT designs. To avoid the interception with the delay line due to the repulsive radial space charge forces in the beam, compensating magnetic forces have to be introduced by a focusing system. Constant axial fields produced by solenoids or periodic fields created by alternating permanent magnets can be used. Sufficient details of the focusing concepts will be discussed later in Sect. 1.3.2.

The RF power P_1 is fed through the RF input window to the delay line. It travels on the delay line with reduced phase velocity v_p approximately equal to the velocity of the electrons in the beam. Interacting with the electron beam, the RF wave gets amplified towards the output to the RF power P_2. Normally, the delay line is divided into an input and output section to limit the gain in each section. At the centre sever of the delay lines the RF wave is attenuated to avoid reflections and oscillations in the sections.

The maximum RF gain can be obtained when the electrons are slightly faster than the RF wave. At that condition the phase velocity is equal to the moving electron density bunch v_{bunch} which is forming in the decelerating phase of the RF wave. It is the velocity of the slow space charge wave on the electron beam moving backwards with the velocity $-v_{sc}$ or, looking from outside, with the velocity $v_e - v_{sc}$. Thus

$$v_p = v_e - v_{sc} = v_{bunch} \tag{1.42}$$

is the fundamental synchronism condition of a TWT. The understanding of this interaction and RF amplification process can be enlightened by the following everyday observation. Assume that electrons, moving across the decelerating and accelerating phases of the RF field, behave like a steady stream of cars moving over a hilly highway. On the uphill side the cars become decelerated and the car density becomes higher. The opposite occurs on the downhill side. Thus we create a bunching of cars on the uphill side, which can remain with time at the same hillside position even when the single cars move over the hill. Back in our electron picture this means that the electron bunch remains and growth for some time in the same decelerating phase with respect to the RF wave. Due to deceleration of bunch electrons, resulting in a growing of bunch space charge density and thus growing of RF current, a maximum amount of kinetic electron energy is at this synchronism condition transferred to the amplified RF wave. Thus at ideal synchronism we get the maximum power gain of the TWT. A quantitative approach to the beam/RF interaction process was given by Pierce, see Sect. 1.3.3.

At the end of the delay line we find: (a) an amplified RF wave with power P_2 coupled via the RF output coupler to a load, and (b) a decelerated spent electron beam with a wide electron velocity spectrum. Depending on the interaction efficiency, the

beam has still about 65% to 95% of its initial kinetic power left, which can be recovered up to approximately 80% in the electron collector.

To support the understanding of the interaction physics further, it is mentioned here that a linear accelerator (LINAC), which transfers power from a propagating RF wave to a stream of charged particles (e.g. electrons), can be understood as an inversed TWT. To accelerate electrons in a growing bunch, it requests for its optimal operation the coupling of the RF wave to the fast space charge wave on the particle beam, and thus the LINAC synchronism condition is given by

$$v_\mathrm{p} = v_\mathrm{e} + v_\mathrm{sc} = v_\mathrm{bunch}. \tag{1.43}$$

That means, in a LINAC the RF wave has to be faster than the charged particles (e.g. electrons) in order to maintain synchronism with the growing bunch during acceleration of single particles in the bunch. In other words, the accelerating phase of the wave front densifies the slower moving single particles in a growing bunch moving with the RF phase velocity.

Having this inversion in mind, we might wonder again why the TWT was invented only in 1942, because the LINAC principle was described already in 1924 by Gustaf Ising, a Swedish physicist, and it was built already in 1928 by the Norwegian engineer Rolf Wideröe.

Collector

The first TWT collectors built were like klystron collectors, single stage collectors on ground potential ($V_\mathrm{C} = V_\mathrm{H}$). At the grounded collector electrode the current loop is closed for the dominant portion I_C of the electron beam current I_beam. Since the spent electron beam entering the collector carries still a lot of kinetic power, this power is thermally dissipated when the electrons hit the collector walls. Now we want to know this dissipated power $P_\mathrm{C,diss}$.

Assuming that the generation of the RF power P_2 has slowed down only the electrons entering the collector, we can write, because of conservation of energy for the dissipated power in the single stage collector,

$$P_\mathrm{C,diss} = I_\mathrm{C} \cdot V_\mathrm{C} - P_2, \tag{1.44}$$

where P_2 is the RF power (fundamental and harmonics) created in the delay line circuit. The goal of TWT designers is to reduce all losses, but due to its dominance especially the collector losses. According to (1.44) this can be done for a given output power P_2 in two directions: first, by reducing the beam current and thus the collector current I_C required to produce the output power P_2, which needs improvements of the beam to RF interaction in the delay line section; second, by reducing the collector voltage V_C with respect to the cathode. This collector voltage depression is possible as long as the slowest electrons in the spent beam entering the collector have sufficient kinetic energy $E_\mathrm{kin,min}$ to arrive at the electric potential of the collector surface,

$$E_\mathrm{kin,min} > |e \cdot (V_\mathrm{H} - V_\mathrm{C})|. \tag{1.45}$$

Here e is the elementary charge of the electron. In modern space application TWT's a depression ratio up to $(V_H - V_C)/V_H = 0.55$ can be obtained. Since for a constant pitch helix the beam power efficiency η_0 is about 13%, it was a breakthrough, when H. Wolkstein [17] reached in 1958 with a single stage depressed collector TWT 30% efficiency.

Similar consideration can be made when i collector stages are introduced to further decelerate the fast electrons in subsequently depressed collector voltage stages. In the formula (1.44) the term $I_C \cdot V_C$, the electric power provided by the single stage power supply, can be replaced by the much smaller sum over the electric power supply stages i, and the dissipated power in a multi stage collector becomes

$$P_{C,\text{diss}} = \sum_i I_{C_i} \cdot V_{C_i} - P_2, \tag{1.46}$$

$$E_{i,\text{kin}} > |e \cdot (V_{C_i} - V_{C_{i+1}})|. \tag{1.47}$$

In (1.47) the entrance condition for the residual kinetic energy of electrons in stage i is related to the increase of potential energy of electrons entering the stage $i + 1$.

An exotic extreme with 10 depressed collector stages was reached by Neugebauer and Mihran [18] in 1972. It allowed increasing the efficiency of a certain klystron from 54% (single stage) to 70.9%. In modern space TWTs the best compromise between efficiency improvement and complexity is found in the range of 3 to 5 depressed collector stages depending on the width of the spent beam velocity spectrum. There, with 4 collector stages a total TWT efficiency up to 74% can be obtained.

1.3.3 TWT Physics

After the qualitative introduction to the TWT concept and its global DC-power operation, we give in here a basic description of the TWT physics and a deeper review of the TWT components. For a detailed study additional textbooks are required. Amongst others, the three English language books [6, 19, 20] are recommended.

Pierce Electron Gun for Space Charge Limited Beams

The Pierce electron gun starts a beam from a cathode surface which is larger than the final beam cross-section in order to keep the cathode current density in the order of a few A/cm^2, and thus the required operating temperature as low as possible. As Fig. 1.17 shows, this can be obtained with a spherical concave shaping of cathode surface. The equipotential lines are concentric, and the electron trajectories start perpendicular on those from cathode surface.

The perveance of such a Pierce gun can be analytically approached by the model of a conical segment of a spherical diode as calculated by Langmuir and Blodgett [21]. Here we limit ourselves to show in Fig. 1.18 the dependence of the gun perveance on the geometrical properties.

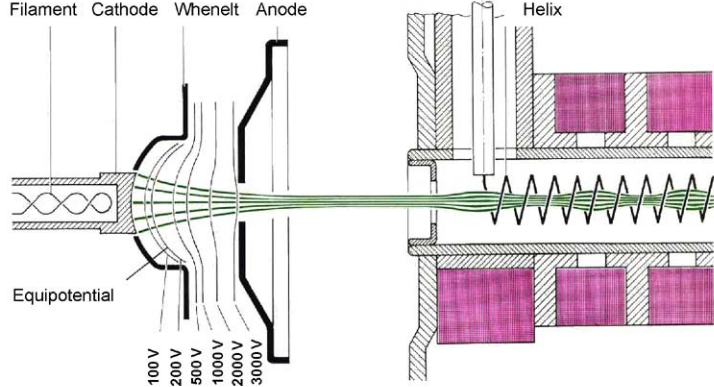


Fig. 1.17. Launch of an electron beam in a modified Pierce optic electron gun

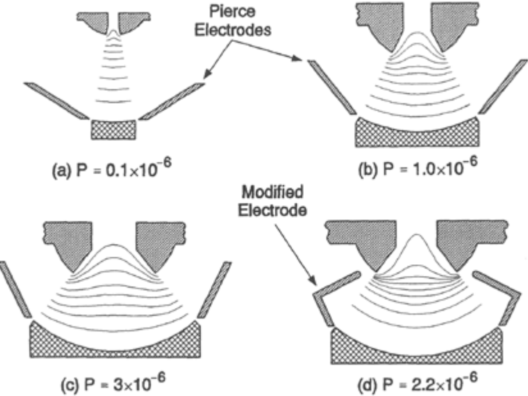


Fig. 1.18. Perveance P of different electrode configurations measured in $A/V^{3/2}$. The Wehnelt (Pierce) electrode is kept on cathode potential in all cases. The perveance increases with smaller cathode to anode distance and larger half-cone angle θ (width of cathode)

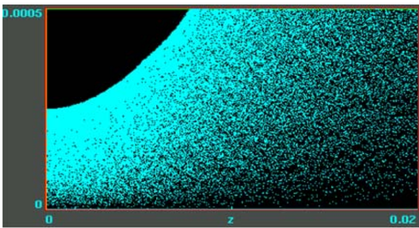
Now computer codes allow to calculate the perveance and the electron trajectories of all wanted gun geometries.

Beam Focusing Over Delay Line Length

Before dealing with the external measures to focus an electron beam, we look to the self-forces acting on the beam due to its charge and current density.

Repulsive Space Charge and Attractive Magnetic Forces in a Beam

For a uniform cylindrical beam with uniform axial velocity v, it can be shown that the radial forces at radius r on a charged particle are given by



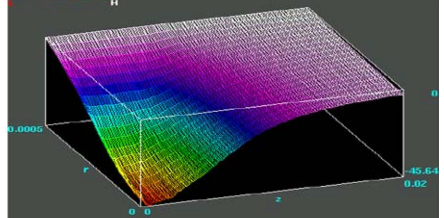


Fig. 1.19. Electron distribution and space charge potential using XOOPIC, a 2.5 dimensional particle in cell code developed by the Berkeley University in California. Simulation parameters: $I_{\text{beam}} = 110\,\text{mA}$, $V_{\text{beam}} = 7.5\,\text{kV}$, $r_{\text{beam}} = 0.25\,\text{mm}$ radius beam $r_{\text{tube}} = 0.5\,\text{mm}$ radius grounded tubing

$$F_{\text{sc}} = Qe/(2\pi r \varepsilon_0); \qquad \text{outward radial space charge force;} \tag{1.48}$$

$$F_{\text{m}} = Qe\mu_0 v^2/(2\pi r); \qquad \text{inward radial magnetic force,} \tag{1.49}$$

where Q is the total charge per unit length *inside* the radius r.

Inserting the physical identity $\varepsilon_0\mu_0 = 1/c^2$ into the ratio of both self forces one simply obtains

$$F_{\text{sc}}/F_{\text{m}} = c^2/v^2. \tag{1.50}$$

Above equations have two noteworthy consequences:

- The radial outward space charge force is dominating the magnetic pinching effect for all beam velocities, since $v < c$. But, there can be the exception that positive ions are neutralising the electron beam space charge which then can lead to a dominance of the contractive magnetic self force. (Example: vacuum switches, there this destructive effect is counteracted by an applied axial magnetic field.)
- At the inner edge of a cylindrical hollow beam the inside charge Q is zero, thus electrons on the inner radius of a hollow beam do not see a radial space charge force!

Universal Beam Spread

In Fig. 1.19, we plot the beam expansion and potential depression of an unfocused, typical beam required for a 200 W Ku-band TWT in a grounded tubing with 1 mm diameter and 20 mm length. We see that within 1 cm length the beam would expand to hit the inner diameter of a helix. The potential depression at the origin is −45.6 V and reduces due to the beam expansion.

Magnetic Focusing

Inserting the same beam into an axially constant magnet field with $B_z = 0.33\,\text{T}$, as it could be produced by a solenoid, we get a beam with a small so-called scalloping ripple, as seen in Fig. 1.20. The potential depression is now almost uniformly 45.8 V.

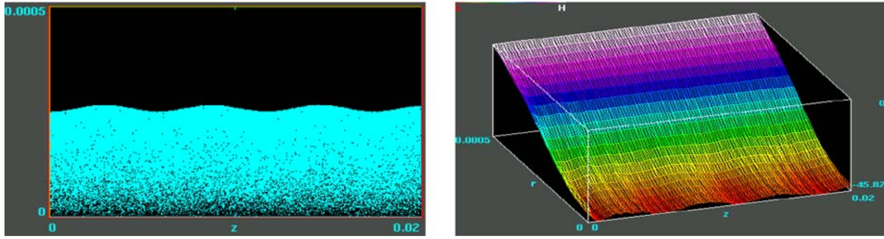


Fig. 1.20. Particle in cell simulation of a 110 mA, 7.5 kV, 0.25 mm radius beam in a 0.5 mm radius grounded tubing as simulated with XOOPIC, in a homogeneous magnetic field with $B_z = 0.33\,\text{T}$

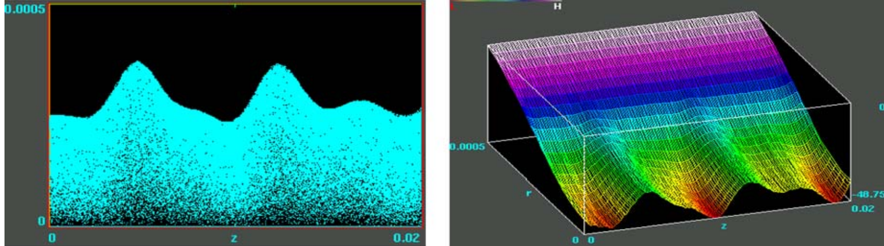


Fig. 1.21. Particle in cell simulation of a 110 mA, 7.5 kV, 0.25 mm radius beam in a 0.5 mm radius grounded tubing as simulated with XOOPIC in a PPM magnetic field with 0.33 T peak field

Figure 1.21 shows the overlay of scalloping and PPM ripple period (7 mm magnet period) for the same beam and a periodic peak field of 0.33 T. In comparison to the uniform magnetic field case, we recognise a much stronger beam ripple due to the overlay of scalloping and magnetic field periodicity with a pronounced potential variation on the axis of 16 V compared to 3 V, respectively. *Note*: The potential traps along the axis can cause ion oscillation phenomena by ionisation of residual gas in a TWT. In practical life, small disadvantages of a PPM magnet system are more than compensated by large savings with respect to mass, power, reliability and interface complexity.

Brillouin Flow

There is only one special case, known as the Brillouin flow, without any beam and potential ripples. A straightforward derivation from the Busch theorem gives the Brillouin flow field $B_{z\text{Br}}$ as function of beam current I, beam voltage V or electron velocity v_e and beam radius a:

$$B_{z\text{Br}} = \sqrt{\frac{2 \cdot I}{\varepsilon_0 \cdot \pi \cdot a^2 \cdot v_\text{e}} \cdot \frac{m_\text{e}}{\text{e}}},$$

$$\frac{B_{z\text{Br}}}{\text{Vs/cm}^2} = 8.3 \times 10^{-6} \frac{(\text{I/A})^{1/2}}{\left(\frac{a}{\text{cm}}\right)\left(\frac{\text{V}}{\text{V}}\right)^{1/4}}. \tag{1.51}$$

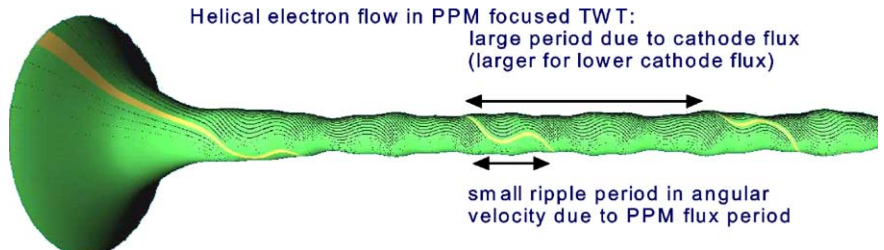


Fig. 1.22. Envelope of an electron beam with a "golden edge trajectory" from cathode surface into the PPM focused interaction region of a typical Ku-band space TWT. Parameters of the modified Pierce gun are: $I = 150\,\mathrm{mA}$, $V = 7.5\,\mathrm{kV}$ into a $\phi = 1\,\mathrm{mm}$, • grounded tunnel, $B_{\mathrm{peak}} = 0.325\,\mathrm{T}$. (Simulated with 2D-gun program and visualised with virtual reality shareware code by W. Schwertfeger, TED, Ulm. Different relative scaling, $r = 8 \times z$ direction)

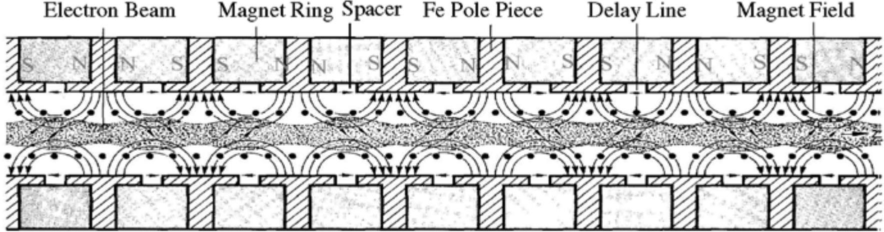


Fig. 1.23. PPM magnet stack with Fe-pole pieces and nonmagnetic spacers braced to a tight vacuum envelope

Equation (1.51) is used as a design rule for TWT magnet systems: the peak PPM magnetic field should be 1.2 to 2 times the Brillouin field to guarantee a save focusing of an electron beam over the interaction length of a TWT.

Another design rule, the focusing stability requirement, is valid for the length of the magnetic field period l_{m} in a PPM system:

$$l_{\mathrm{m}} \leq 418 \cdot \frac{V}{B_{z\mathrm{Br}}^{2}} \cdot \frac{m_{\mathrm{e}}}{e}. \tag{1.52}$$

Figure 1.22 visualises the helical movement of the electrons in the PPM focused beam. A technical realisation of a PPM focusing system is sketched in Fig. 1.23.

Delay Line System and RF Interaction

Types of Delay Lines

Depending on the application and power level, several types of delay lines are used:

- Helix; low power, very broadband (up to several octaves), needs ceramic supporting rods.

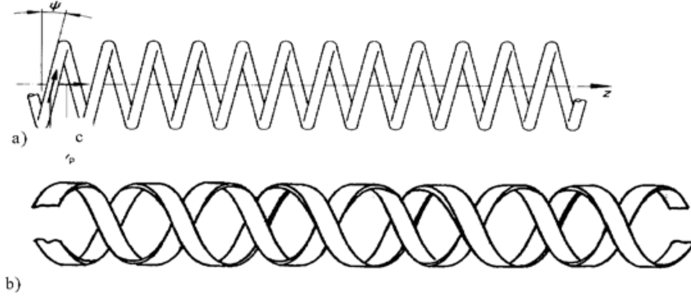


Fig. 1.24. a Helix delay line: the pitch angle ψ defines the RF phase velocity $v_\mathrm{p} = c \cdot \tan\psi = c \cdot 2\pi a/p$, $c =$ speed of light, $p =$ axial pitch; **b** Double helix delay line

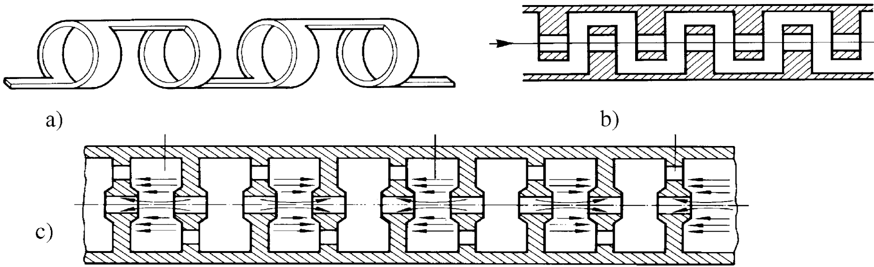


Fig. 1.25. a Ring and bar line; **b** Interdigital or comb line; **c** Coupled cavity line

- Counter wound double helix; suppresses backward wave oscillations (due to mechanical complexity scarcely used), needs ceramic supporting rods.
- Ring and bar line; suppresses backward wave oscillations, needs ceramic supporting rods.
- Interdigital line; rugged and compact delay line for high power high frequency TWTs, reduced bandwidth.
- Coupled cavity; low bandwidth, high power capability.

Figures 1.24 and 1.25 give a survey on the major types.

RF Input and Output

Depending on the delay line concept, several configurations for the RF input and output are known. We give in Fig. 1.26 only two examples for waveguide windows, as used, e.g. for a coupled cavity TWT input, and a coaxial feed through, as often used for low power helix TWT input and output.

Small Signal Theory (Pierce)

At the time Pierce developed his TWT theory, the Eulerian approach was the only success promising approach. It treats the electron beam as a charged fluid, characterised by a current I with beam velocity v_0 or beam voltage V_0 and the space charge

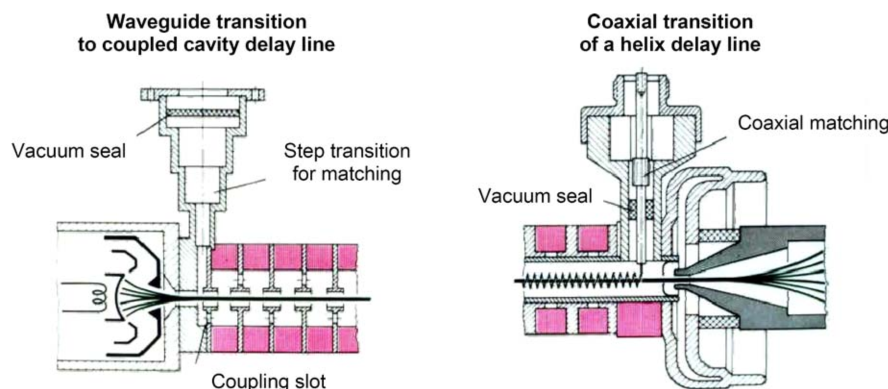


Fig. 1.26. *Left*: waveguide input transition with a Chebyshev step transformer for matching. *Right*: coaxial output transition from a helix with coaxial matching and $\lambda/4$ ceramic window for the vacuum seal

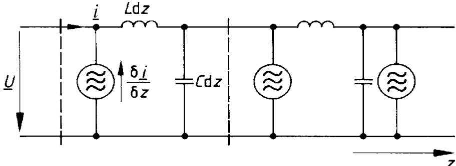


Fig. 1.27. Equivalent transmission line model for a helix delay line

bunches as an AC current i propagating with the same frequency as the RF wave. As indicated before, this approach is limited to the small signal behaviour of the TWT because saturation effects cannot be treated. The ingenious concept of Pierce combined this Eulerian view of the electron beam with the replacement of the delay line as propagation structure by an equivalent transmission line model consisting of distributed inductances L and capacitances C, as shown in Fig. 1.27.

As a result of his derivation, he obtained a fourth degree equation for the wave propagation constant β which is known as the so-called determinantal equation

$$0 = 1 + \frac{\beta_e}{(\beta - \beta_e)^2} \cdot \frac{\beta^2 \beta_c}{(\beta^2 - \beta_c^2)} 2C^3, \tag{1.53}$$

$$C^3 = \frac{I \cdot Z}{4V_0}, \tag{1.54}$$

where C is the Pierce parameter. It has four roots corresponding to (1.55)–(1.57),

$$\beta_{1,2} = \beta_e + \beta_e C/2 \pm j\sqrt{3}\beta_e C/2; \quad \text{Solutions for exponentially increasing and decreasing slow waves;} \tag{1.55}$$

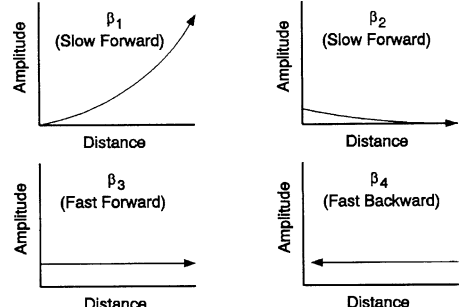


Fig. 1.28. The four waves propagating along the delay line of a TWT

$$\beta_3 = \beta_e(1 - C); \quad \text{Solution for a fast forward wave with constant amplitude;} \tag{1.56}$$

$$\beta_4 = \beta_e(C^3/4 - 1); \quad \text{Solution for a fast backward wave with constant amplitude.} \tag{1.57}$$

The meaning of the solutions becomes clear by reminding that the waves vary proportional to $e^{j(\omega t - \beta z)}$. Figure 1.28 sketches the four solutions. Considering only the exponentially increasing solution, we get from (1.55) the linear power gain G over N wavelengths of the TWT in dB

$$G = 10 \log 10 e^{\sqrt{3}\pi CN} = 47.3 \cdot CN. \tag{1.58}$$

To satisfy the boundary conditions when launching the four waves at the tube entrance, one gets an additional so-called distribution loss, and (1.58) becomes

$$G = -9.54 + 47.3 \cdot CN, \tag{1.59}$$

which is the total linear power gain of the tube with distribution loss. It should be mentioned here that above solutions are not limited to helices, but are valid for all type of delay lines when introducing their respective equivalent circuit properties.

Cutlers Beam and RF Signal Characterization up to Saturation

A great improvement in the understanding of the nonlinear behaviour of the TWT was reached when Cutler [22] published his test results on RF circuit voltage (dashed line), beam velocity (solid line) and beam current (vertical width of shaded area) at the TWT output relative to the undistorted signal phase as function of drive level (see Fig. 1.29). His findings for the 15 drive conditions from −22 dB input back off to +9 dB overdrive level are interpreted as follows. With respect to the circuit voltage, one can define an acceleration and a deceleration zone for the beam electrons.

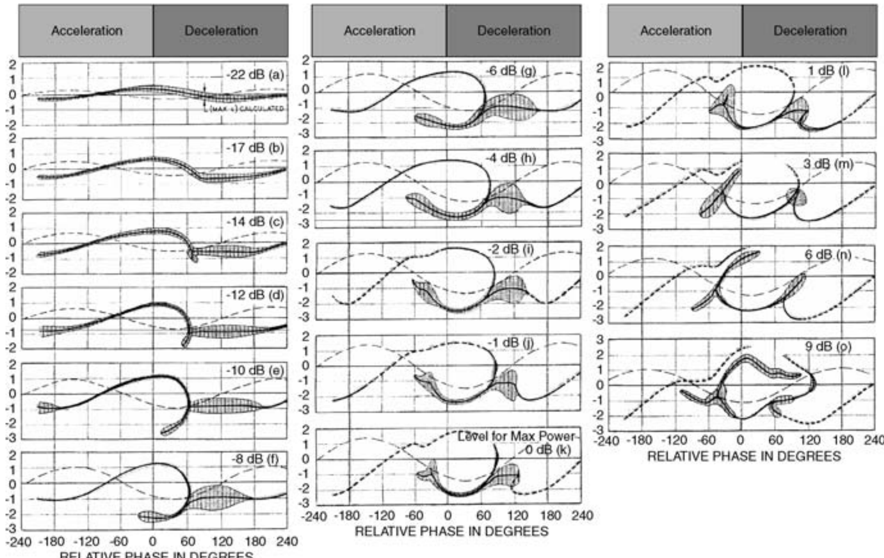


Fig. 1.29. Results of Cutler experiment in 1956

Due to the acceleration or deceleration we find the relative velocity v_r of the electrons to be increased between acceleration and deceleration zone and smaller than zero in the other regions. Also, we see that v_r starts to become double-valued for drive levels larger than 14 dB input back off because some electrons are reflected back from the deceleration zone into the acceleration zone. The amount of beam current (shaded area) is larger in the deceleration zone till saturation, where the electron bunch in acceleration zone becomes about equal due to decelerated electrons moving into this zone. Two bunches in one period means that a harmonic frequency wave has been created on the circuit. This reaches its maximum after saturation. At higher overdrive level, the electrons are smeared out again more uniformly over the relative phase. If one makes a vertical cut through the electron velocity body and integrates that over various relative phases, one gets the electron velocity spectrum at the tube exit as a function of drive. This exit velocity spectrum is entering into the collector and plays an important role for the total efficiency of the TWT as we will see in the next section. Here, it can be summarised that most of the electrons are decelerated, but few of them are accelerated compared to the start velocity v_0.

Large Signal Theory (Rowe) and Simulation Results

It is not space enough to report on the numerous theoretical approaches which deal with the nonlinear effects occurring closer to saturation drive of a TWT or even in overdrive. It should be sufficient to refer the reader to the compendium of Rowe [23] and to declare that no analytical solutions exist for the large signal operation of the TWT. The theoretical approaches are mainly based on the Lagrangian concept

which treats the electrons as particles or as representative particles and which allows multi-valued velocity distributions of electrons at a given space or phase location. Still today, fully 3-dimensional numerical simulations are too heavy for useful design work. A lot of 2.5-dimensional codes (2-dimensional axially symmetric electric and magnetic fields and 3-dimensional movement of representative electrons) (layers and shells of electrons, or statistically distributed "heavy electrons"), are providing reliable design information on gain, output power, nonlinear phase shift and even intermodulation distortions as a function of input drive and frequency as well as the trajectories of the representative electrons in the real space and phase space. As an input information to the code, they need from the delay line side only the coupling impedance $K(f, r, z)$ as a function of frequency, radius and position, the phase velocity $v_{\mathrm{p}}(f, z)$ as a function of the frequency f and the axial position, the RF attenuation $\alpha(f, z)$ as a function of the frequency f and the axial position.

For the beam, the representative charge $\rho(r, z_0)$ and current density $\boldsymbol{j}(r, z_0)$ information is required at the entrance z_0 of the delay line. Figure 1.30 gives for a modern Ku-band space TWT the simulated relative phase of the electrons, being uniformly distributed over 2π at the delay line entrance, the outer electron beam diameter and the small signal and saturation drive gain over the length of the delay line.

For Fig. 1.30a we note the crossing of electrons approaching saturation at the very last centimetres before the delay line end. From Fig. 1.30b we see again the magnetic ripple on the beam and the moderate increase of the beam radius as the electron velocity slows down close to the exit. From Fig. 1.30c we recognise the in-

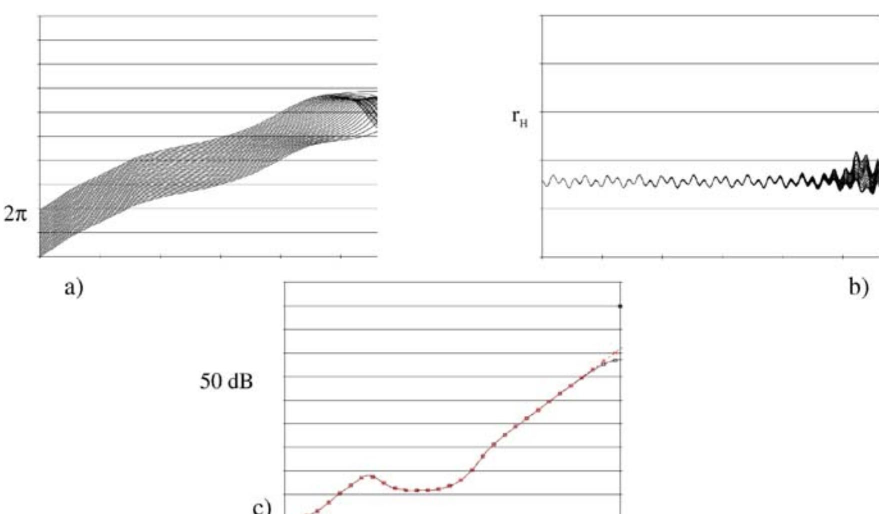


Fig. 1.30. a Applegate diagram of electrons (relative phase with respect to RF wave); **b** Outer diameter of electron beam inside delay line; **c** Small signal (crosses) and saturation (circles) gain along tube axis

sertion loss at the delay line entrance, the extended sever region where RF power is transported by the electron beam and the difference between small signal and saturation gain of less than 5 dB, which is important for excellent linearity characteristics of the tube.

Electron Beam Collector

Collection of an Electron Beam with Distributed Velocities

Figure 1.31 shows a typical velocity spectrum of a highly efficient Ku-band space TWT at saturation drive as a cumulative distribution function. It gives on the ordinate the amount of the current ratio I/I_0 which has a higher axial velocity than that corresponding to the reference acceleration voltage given on the abscissa. The original zero drive velocity distribution corresponds to the depicted rectangular distribution (all electrons have uniform velocity corresponding to helix voltage of 6250 V). The shaded areas correspond to the recovered power P_{recover} in the collector if the respective current contributions are gathered at their corresponding reduced collector potentials of a multistage collector.

Potential Depression in an Ideally Focused Cylindrical Beam

Assume an ideally focused cylindrical beam with $I = 110\,\text{mA}$ in a metallic tube at potential $\phi(b) = 0$ with the mono-energetic beam voltage corresponding to the negative cathode potential $\phi_c = -7500\,\text{V}$, as shown in Fig. 1.32. With the beam velocity

$$v_z = (2\eta(\phi_b - \phi_c))^{1/2}, \tag{1.60}$$

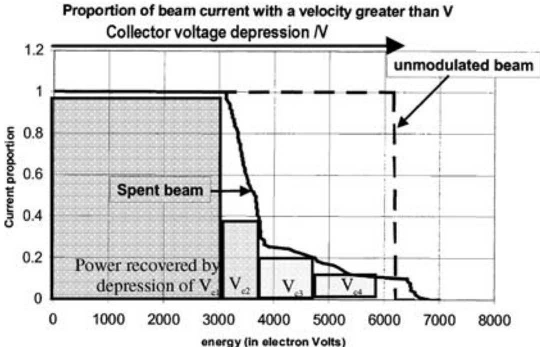


Fig. 1.31. Typical velocity/energy spectrum of the spent beam at the collector entrance and recovered power proportions (shaded area) of a Ku-band TWT with $V_H = 6250\,\text{V}$, $V_{C1} = 3250\,\text{V}$, $V_{C2} = 2500\,\text{V}$, $V_{C3} = 1750\,\text{V}$ and $V_{C4} = 450\,\text{V}$ with respect to cathode. The collector efficiency is represented by the ratio of the shaded area to the total area below the spent beam distribution curve

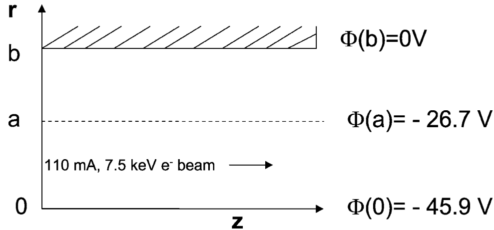


Fig. 1.32. Cylindrical electron beam of radius a in a tube with radius b and potentials $\phi(0)$ at the beam centre, $\phi(a)$ at the beam edge and $\phi(b)$ at the grounded tube

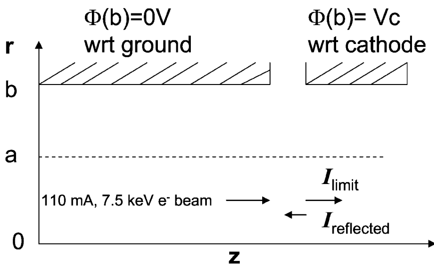


Fig. 1.33. Focused beam with voltage ϕ_{beam} entering a collector iris at a reduced collector potential V_{C}

we write the approximately valid analytical formulas for the depressed beam potential:

$$\phi(a) = \phi(b) - I \cdot \ln(b/a)/2\pi\varepsilon_0 v_z; \qquad \text{Potential at beam edge;} \tag{1.61}$$

$$\phi(0) = \phi(b) - I \cdot (\ln(b/a) + 1/2)/2\pi\varepsilon_0 v_z; \qquad \text{Potential at beam centre.} \tag{1.62}$$

It is quite satisfying that the results for the potential depression at the beam edge and the beam centre, $\phi(a) = -26.7\,\text{V}$ and $\phi(0) = -45.9\,\text{V}$, is in agreement with those from the particle in cell code XOOPIC simulation (see Figs. 1.19–1.21), where we used the same parameters $I = 110\,\text{mA}$, $\phi(b) = 0$, $\phi_{\text{c}} = -7500\,\text{V}$ and $b = 2a = 0.5\,\text{mm}$.

Collector Current Limitation by Space Charge Effects

Assume the beam of Fig. 1.32 approaching a collector entrance iris at a reduced potential, as shown in Fig. 1.33. The question arises about the limited beam current I_{limit} which can pass without reflections into the collector tunnel at a depressed collector potential V_{C}. Note: V_{C} is measured with respect to the cathode potential.

As we see from (1.63) below, the maximum current depends on the ratio of the collector tunnel radius to the beam radius b/a pressed to avoid reflection of parts of the beam for a given beam current and the ratio of the collector entrance radius b to the beam radius a (Fig. 1.34),

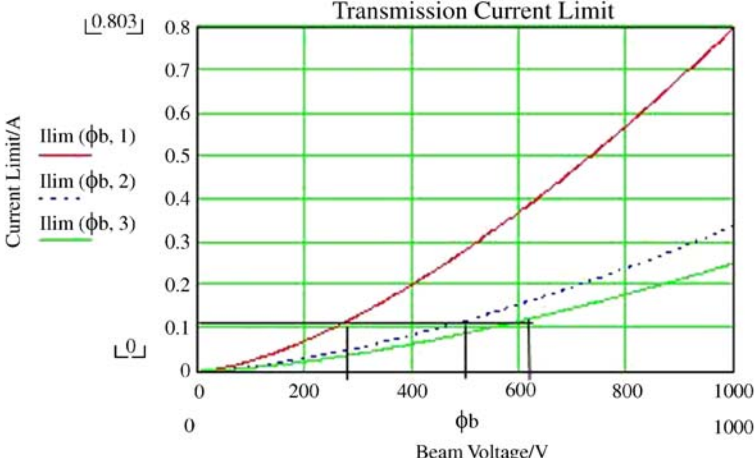


Fig. 1.34. Transmission current limit as a function of beam (collector) voltage for 3 values of tunnel to the beam radius ratio $b/a = 1$, 2, and 3. As we see for the example of a 110 mA beam, the lowest beam voltage (depressed collector voltage) can be achieved if the beam is uniformly filling the collector tunnel ($b/a = 1$)

Table 1.2. Minimum collector voltage for a 110 mA electron beam as function of the tunnel-beam radius ratio

Ratio b/a	1	2	3
Minimum collector voltage for a 110 mA beam	260 V	500 V	620 V

$$I_{\mathrm{lim}}(b/a) = 4\pi\varepsilon_0\sqrt{e/m_0}\cdot\frac{1}{1+2\ln(b/a)}\cdot\left(\frac{2}{3}(\phi_{\mathrm{beam}} - V_{\mathrm{C}})\right)^{3/2}. \qquad (1.63)$$

Equation (1.63) has again the form of a perveance or here acceptance, $I_{\mathrm{lim}} = A(b/a)\cdot(V_{\mathrm{C}})^{3/2}$, which means that the scaling laws apply (see Table 1.2).

The findings imply a theoretical limit to the collector efficiency. Depending on the width of the electron velocity spectrum, 3 to 5 collector stages are found to give an optimal compromise between the total tube efficiency and system complexity. Amongst the MDCs the two categories, electrostatic collectors and magnetically focused collectors, are distinguished.

Electrostatic Collectors

Due to the deceleration of the beam at the collector 1 entrance, the radial space charge force in the beam increases and the beam expands radially. This beam expansion affects the slower electrons more than the faster ones and is therefore used to separate the slower electrons from the faster ones which land further downstream.

Figure 1.35 sketches the electron trajectories in a 6 stage (5 stages +1 spike on cathode potential) collector as it was used in the 1980s for the 250 W space TWT TL

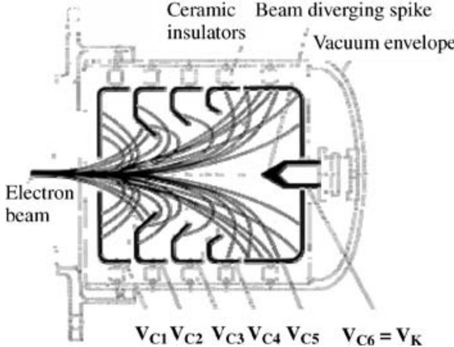


Fig. 1.35. Electrostatic 5 stage + spike collector used for radiation cooled space TWT

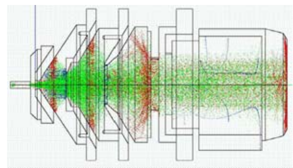


Fig. 1.36. Distribution of primary reflected and secondary electrons in an *electrostatic* 4 stage collector simulated with the Thales Electron Devices PIC program Collect 3D for a 150 W Ku-band TWT

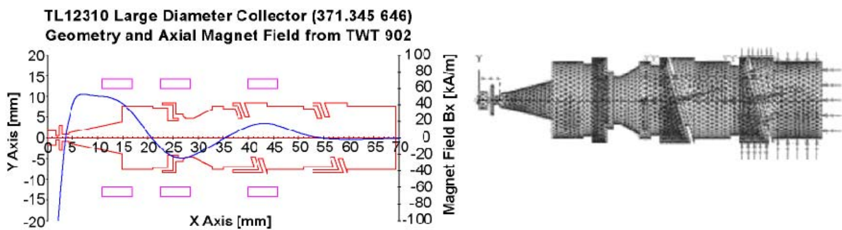


Fig. 1.37. Collector geometry, axial magnetic field and simulation mesh of a magnetically focused 4 stage collector for a 300 W Ku-band TWT (the transverse magnetic fields are not shown)

12250. It achieved typically 48% total efficiency. Figure 1.36 shows the design of a modern electrostatic 4 stage collector used in some Thales Ku-band TWTs with the simulated distribution of primary (green), reflected (blue) and secondary electrons (red) in the collector volume, which achieves about 68% total efficiency.

Magnetically Focused Collectors

These collectors minimise the radial dimensions of the collector by continuing the magnetic PPM focusing of the electron beam with an adapted periodicity and strength of the magnetic field into the various collector stages. Figure 1.37 shows the principle for a modern 4 stage collector used for a Thales 300 W Ku-band TWT.

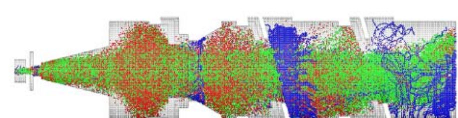


Fig. 1.38. PIC-simulation of electron distribution in the *magnetically focused* 4 stage tilted field collector at zero drive (*left*) and saturation. Primary electrons (*green*), reflected (*blue*) and secondary (*red*) simulated with the Thales Electron Devices PIC program Collect 3D

The magnetically focused collector allows an optimised beam filling of the collector at saturation drive as shown on the right side of Fig. 1.38 and thus a maximum efficiency up to 74%. To minimise backstreaming of electrons, the most advanced magnetically focused collectors are using tilted electric and magnetic fields producing compensating deflection effects in downstream direction to focus the fast electrons into the collector 4 tunnel and enhanced deflection in backstreaming direction.

Relation between TWT Total-, Beam- and Collector Efficiency

The TWT total efficiency η_{tot} and the beam power efficiency η_{beam} (or basic efficiency η_0) can be simply defined and rewritten as

$$\eta_{tot} = \frac{P_{fund}}{P_{el}} = \frac{\eta_{beam} \cdot P_{beam}}{P_{RF} + P_{losses}} = \frac{\eta_{beam} \cdot P_{beam}}{P_{beam} - (P_{beam} - P_{RF} - P_{losses})}, \tag{1.64}$$

where P_{fund} is the fundamental RF power at the TWT output, P_{RF} is the total RF power (including fundamental and harmonic RF power and RF losses), $P_{beam} = V_H \cdot I_K$ is the electron beam power and P_{losses} are the thermal losses produced in the tube. A short analysis of the thermal TWT losses P_{losses} provides

$$P_{losses} = P_{Coll,losses} + P_{H,therm} + P_{A,therm} + (P_{filament}). \tag{1.65}$$

Neglecting the small quantities (filament power, helix and anode losses), the term $P_{recover} = (P_{beam} - P_{RF} - P_{losses})$ in (1.66) becomes approximately equivalent to the recovered kinetic electron beam power by reduced collector voltages. With the definition of the collector efficiency η_C as the ratio of the recovered power to the entering beam power

$$\eta_C = \frac{P_{recover}}{P_{enter}} = \frac{P_{beam} - P_{RF} - P_{losses}}{P_{beam} - P_{RF}}, \tag{1.66}$$

we get a simplified relation between the total beam power and collector efficiency by the further neglecting the harmonic power and the RF losses in P_{RF} and by division with P_{beam}

$$\eta_{tot} = \frac{P_{fund}}{P_{el}} = \frac{\eta_{beam} \cdot P_{beam}}{P_{beam} - \eta_C (P_{beam} - P_{RF})} = \frac{\eta_{beam}}{1 - \eta_C (1 - \eta_{beam})}. \tag{1.67}$$

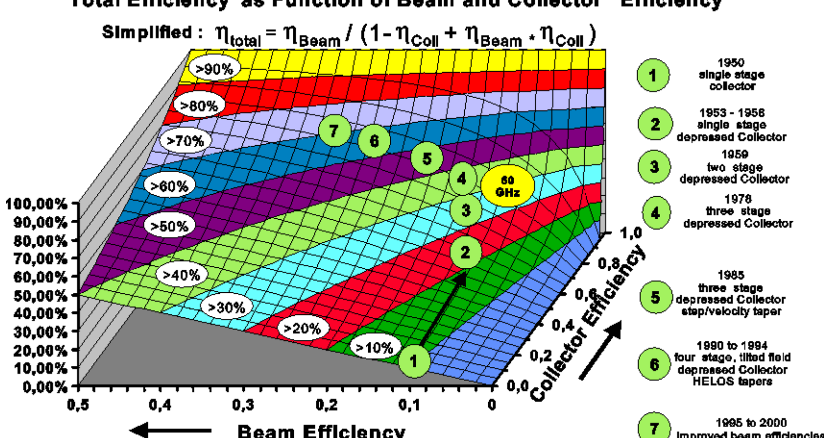


Fig. 1.39. Total efficiency as function of the two variables beam efficiency and collector efficiency. The historical path from single stage collector tubes (station 1, 10%) to 4 stage tilted field collector tubes with enhanced beam efficiency (station 7, 70%) is sketched for Ku-band TWTs. The figure is adapted from Kornfeld et al. [24]

Typical Ku Band Travelling Wave Tube of Today
Relative Contributions of Loss Mechanisms in Saturation

Cathode Heater Power 5%
Skin Losses 35%
Reflected Power 1%
Harmonic Power 8%
Helix Interception 6%
Collector Losses 45%

Cathode Heater Power
Skin Losses
Reflected Power
Harmonic Power
Helix Interception
Collector Losses

Fig. 1.40. Total distribution of thermal losses in a modern TWT

The failure in making these simplifications can be estimated from Fig. 1.39. In Fig. 1.39 the relation (1.67) is plotted. It indicates further the historical development axes for TWT efficiency. First this occurred via the collector efficiency improvement and later by essentially improving the beam efficiency with tapered helix concepts. Though Fig. 1.39 seem to imply that beam and collector efficiency are independent, this is physically not the case. The larger the beam efficiency becomes, the wider is the electron velocity spectrum entering the collector with detrimental effects on the collector efficiency.

Loss Analysis of a Typical Ku-Band Space TWT

From Fig. 1.40 we can see, that further efficiency improvements need to reduce mainly the collector losses (45%) and the skin effect losses (35%). The other losses

in a developed space TWT are well below 10%: harmonic power 8%, helix interception losses 6%, cathode heater 5%, and reflected RF power 1%.

1.3.4 TWT Applications

Table 1.3 gives a survey on the different application fields of TWTs. We note that the helix is the most used type of delay line. For communications the reason is its large bandwidth and very good linearity. For military radar and ECM, again its extreme bandwidth, which for some type of ECM tubes can achieve up to 3 octaves. In the following we provide application examples based on the Thales Electron Devices product spectrum.

Communication TWTs

TWTs for Ground Station, Airborne and Shipboard Communication

Following an internal TED communication of Francis Payen, we give a review of these TWT applications.

Depending on the system architecture, the microwave tubes used in ground-based, airborne or shipboard communications system transmit signals to a satellite (up-link) or to a ground based receiver (point to point or point to multipoint communication). Though in some communication systems, due to power requirements, also klystrons are used, TWTs are becoming predominant, because the relative bandwidth requirements are becoming more and more demanding (usually more than 10% are required). Table 1.4 presents the relative bandwidth and the output power capability, as a function of frequency, for the TWTs made by TED for up-link communications systems.

Some examples of ground-based or airborne communication systems using TED microwave tubes are DirecTV, Echostar and Astra for DBS, Iridium for commercial systems and Milstar, Syracuse, Stentor, SBIRS for military systems.

Table 1.3. Application fields for different types of TWT delay lines

Application	Communication		Radar and ECM				
Sub-application	Ground station	Space	Earth Observation	Ground & Ship Radars	Airborne Radars	Missile Seekers	ECM
Helix	•	•	•	•	•	•	•
Ring & Bar	•		•				
Coupled Cavity	•			•	•		
Interdigital Line	•			•	•	•	

Table 1.4. Survey on up-link TWTs

Frequency band GHz	C 5.85–7.1	X 7.9–8.4	Ku 12.75–14.5	Ku 17.3–18.4	Ka 25.5–31.5	Q 43.5–45.5
Relative bandwidth	Up to 20%	6%	Up to 13%	6.2%	Up to 20%	4.5%
CW output power	3 kW	2.5 kW	1 kW	500 W	350 W	250 W

Table 1.5. Comparison of historic C-band tubes

	First TWT in Use	First Space TWT	First European Space TWT	Modern Space TWT
Program	TV-Ground Link	Telstar 1	Symphonie	Measat
Manufacturer	STC	Bell Lab	AEG	TED
Year	1952	1962	1973	2003
Frequency	3.6–4.4 GHz	3.7–4.2 GHz	3.7–4.2 GHz	3.4–4.2 GHz
Output Power	2 W	2 W	13 W	70 W
Gain	25 dB	40 dB	46 dB	56 dB
Efficiency	≈1%	<10%	34%	71%
Nonlinear Phase	?	50°	50°	38°
Mass	>5000 g	>1000 g	640 g	800 g
Collector	1 stage	1 stage	1 stage depressed	4 stage depressed
Focusing Syst.	Solenoid	PPM PtCo	PPM PtCo	PPM SmCo
Cathode	Oxide	Oxide	Oxide	MM Dispenser

TWTs for Space Communication

The major requirements for microwave tubes used in satellite transponders for transmitting down-link signals are:

- long useful operating life and high reliability (>15 years, <100 FIT, respectively; more than 35 000.000 h accumulated in orbit),
- high total electrical efficiency (>60%),
- high linearity (nonlinear phase shift <48°),
- low mass (depending on output power and radiation or conduction cooling).

TWTs meet all these demanding requirements and beat by far the competition from solid state amplifiers especially with respect to life, reliability and efficiency. The driving force for the impressive improvements in all those parameters was in the past and is still today the economical pressure to save power by improving the efficiency and to reduce the mass of space amplifiers. Two figures are best characterizing this environment: the savings in launch and system costs per satellite are for DC power saving 3.000 €/W and for mass reduction 20.000 €/kg. Therefore, it might be of interest to have a short look on the history of some of the most important space TWT parameters as reviewed by Kornfeld et al. [24].

Table 1.5 shows the performance characteristics of historic C-band space tubes and compares them with the first operational TWT built as TV relay tube by STC in 1952, D.C. Rogers [25]. Though the output power was increased from 2 to 70 W, the mass was reduced.

Figure 1.41 gives the efficiency improvements vs time of the leading space tube manufacturer. The significant improvements obtained in the first 30 years after the TWT invention led in the late 1970s to the impression that the TWT technology might be completely mature and will provide only little space for further improvements. In contrast to this, a steady and for the 1990s even accelerated efficiency

Fig. 1.41. Efficiency improvement of space TWTs (courtesy of TED)

improvement was found for the space TWTs of more or less all commercial manufacturers. The main reason for this performance explosion was the availability of codes to simulate the large signal beam to RF wave interaction in improved tapered helices. It is expected that efficiency might finally approach a level between 75% and 80%. Table 1.6 presents the frequency band, the output power, efficiency and future expected trends in the respective applications.

As seen from Table 1.6, there are many different applications for space TWTs ranging from standard telecommunication systems, TV- and digital radio broadcasting to modern digital internet and multimedia services. Also Earth observation with pulsed radar TWTs becomes increasingly important. New applications where the satellite TWT technology might be used are local multimedia services from small satellites or stratospheric balloons over big cities. Also, for the low Earth orbit satellite fleet of global positioning systems (GPS, Galileo) and their continuous upgrade, powerful and highly efficient TWTs might be used.

Radar and ECM TWTs

Several types of microwave tubes are used for radars: magnetrons, crossed-field amplifiers (CFAs), klystrons and TWTs.

From a historical point of view, the magnetrons were the first microwave tubes to be used in radar transmitters, more than sixty years ago. But they are oscillators, and most of the radars, since several tens of years, require transmitters using a coherent amplification chain. Among the three types of amplifiers (CFAs, klystrons and TWTs), the TWTs are most widely used, thanks to their wide instantaneous bandwidth, high gain and noise free coherent operation.

Surface Radars

Surface radars are ground based or are used in naval systems (shipboard radars). The main types of surface radars are:

Table 1.6. Space TWT applications for communication and Earth observation

Application	Band frequency/GHz	Power efficiency	Future trend power efficiency
Direct Digital Radio Navigation/GPS & Galileo	L-band 1.1–1.5	50 to 150 W 55%	250 W 65%
Communication/TV-broadcast	S-band 2.3–2.6	70–90 W 59%	120 W 65%
Direct digital radio for automotive	S-band 2.3–2.6	200–240 W 61%	200–250 W 68%
Telecommunication & broadcasting	C-band 3.4–4.2	20–130 W 60–69%	150 W >73%
SAR, for Earth observation, Radar TWT, pulsed	C-band 5 to 6	5 kW 40%	>5 kW 45%
Scientific applications & deep space missions	X-band 7–8.5	25/120–170 W 60%	25/120–170 W 65%
Earth observation, radar TWT, pulsed	X-band 7–8	4 kW 40%	>4 kW 45%
Telecommunication and broadcasting Internet Multimedia services	Ku-band 10.7–12.75	25–200 W 62–68%	25–300 W 68–75%
Altimeter; radar application for Earth observation, pulsed	Ku-band 13–15 or 12–18	Up to 100 W 55%	150 W 60%
Telecommunication and multimedia Services	Ka-band 17–22	15–130 W 55–66%	15–220 W 55–70%
Deep Space & Scientific Mission	Ka-band 27–32	20–30 54%	20–100 58%
Multimedia Services for low orbit Satellites or Stratosphere Balloons	Q-band 40–45	40 W 40%	40–100 W 40–45%
Inter satellite links for multimedia Services	V-band 58–64	20 W 35%	20–100 W 35–40%

- long range surveillance radars for Air Traffic Control (ATC),
- air defence radars,
- tracking radars,
- fire control radars; those are most time integrated into weapon systems,
- trajectography radars.

Table 1.7 presents the relative bandwidth and the output power capability, as a function of frequency, for the TWTs (Ku- and Ka-band) and klystrons (up to and including X-band) made by TED for surface radars.

Table 1.7. Survey on surface radars

Frequency band	L	S	C	X	Ku	Ka
GHz	1.26–1.36	2.7–3.5	5.4–5.9	8.5–10.5	15–18	33–38
Relative bandwidth	3%	3 to 15%	5 to 10%	10%	10 to 20%	3 to 10%
Peak output power	4 MW	20 MW	1 MW	120 kW	2.5 kW	1 kW
Average	12 kW	20 kW	20 kW	5 kW	200 W	200 W

Table 1.8. Survey on missile radars

Frequency band	X	X	X	Ku	Ka	W
Type of tube	Helix TWT	CC TWT	Magnetron	TWT	TWT	TWT
Relative bandwidth (%)	2	3	Tunable in 600 MHz	20	3	1
Peak output power (kW)	20	120	220	2	1	0.15
Average output power (W)	800	1500	200	400	200	15

* The Ka- and W-band TWT use interdigital delay lines; W* under development

Airborne Radars

Microwave tubes (magnetrons and TWTs) are used in airborne radar transmitters in two categories:

- multimode and multifunction radars; TWTs are widely used, either with coupled cavity slow wave structure or Helix;
- Terrain following radars; generally, TWTs are used.

Missile Seekers

The requirements for microwave tubes (magnetrons, klystrons and TWTs) used in active RF missile seekers are small size and weight, high electrical efficiency, very short start-up time, capability to withstand very severe environmental conditions and high reliability after long storage periods.

Some examples are new generation MICA, ASTER and PAC3 missile seekers. The main performances are presented in Table 1.8. It shows the relative bandwidth and the output power capability for magnetrons and TWTs made by TED for airborne radars or missile seeker applications, as a function of frequency.

ECM Applications

The requirements for microwave tubes used in ECM Systems are very wide instantaneous frequency bandwidth (more than one octave), small size and low weight and high electrical efficiency.

The only microwave tube which can meet a specification with more than one octave bandwidth is the Helix TWT. Table 1.9 presents TED's helix TWTs made for ECM systems.

Table 1.9. Survey on broadband ECM tubes

	Frequency band	6 to 18 GHz	18 to 40 GHz
Pulsed TWTs	Peak output power	2 kW	–
	Average output power	80 W	–
CW TWTs	CW output power	200 W	80 W

1.4 Extended Interaction Klystron EIK

1.4.1 Introduction

Similar to klystron and TWT, the extended interaction klystron (EIK) is a linear electron beam device which tries to combine the advantages of both, the ruggedness and high power capability of a klystron and the larger bandwidth of a TWT. Therefore, the EIK is especially similar to the rugged coupled cavity and interdigital line TWT.

Since the EIK can be considered as a refinement of both microwave devices, it is not easy to mention an inventor. The main work on EIKs seems to go back to Tore Wessel-Berg [26] and his "A General Theory of Klystrons with Arbitrary, Extended Interaction Fields" issued in 1957. He was working at SLAC where he improved efficiency and power capability of klystrons for the linear accelerators. Further pioneering work was done by Chodorow and Kulke [27] in the 1960s. It was recognised by those groups that the circuit impedance is enhanced proportional to the extended interaction region and that larger gain bandwidth product and higher efficiency could be obtained relative to the conventional klystron circuit, particularly in the submillimetres/millimetres region. These characteristics suit today requirements for Ka- to W-band air and space borne radars (see [28] and [29], respectively) and also ECM applications.

1.4.2 Extended Interaction Circuit Design

Figure 1.42 compares the principle cross-sections of a conventional two cavity klystron with an extended interaction klystron. The number of cavities and interaction gaps can differ depending on the required application. A 3-dimensional impression of a multicavity multigap design of an EIK is given in Fig. 1.43. This and the previous figure are taken from [28].

The extended interaction oscillator (EIO) is a single cavity device with interaction gaps (segmented drift tube) that function like a coupled cavity TWT structure with extremely strong cavity-to-cavity coupling. At sufficiently high beam currents, oscillations are sustained. Variation of the beam voltage allows 0.4% frequency tuning.

1.4.3 Typical Performance and Applications

The today leading manufacturer of EIKs, CPI-Canada, describes the typical frequency and power range as follows (Table 1.10) [30]. The instantaneous band-

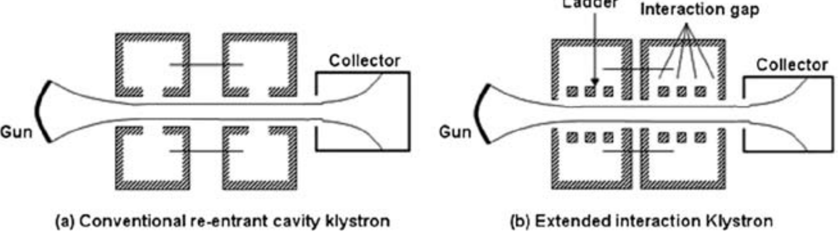


Fig. 1.42. Cross section of a conventional and extended interaction klystron

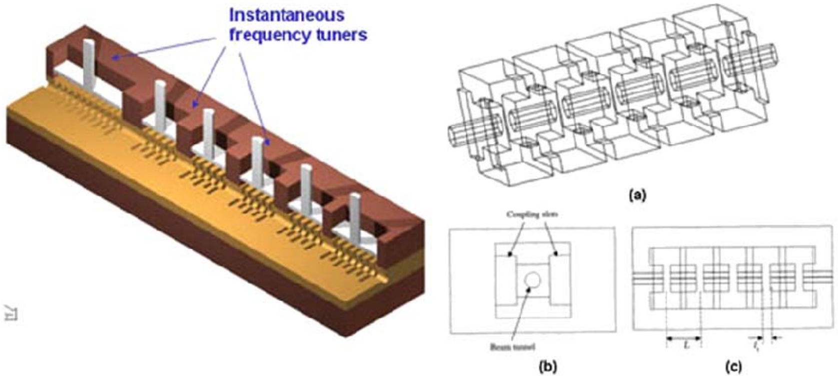


Fig. 1.43. *Left*: Rising sun type tunable EIK ladder circuit. *Right*: **a** 3-dimensional cross-section; **b** top-view; and **c** side-view of an EIK five gap structure in a cavity

Table 1.10. Typical power levels of pulsed and CW EIKs at various frequencies

Pulsed EIK		CW EIK	
3000 W	at 30 to 95 GHz	1500 W	at 30 GHz
400 W	at 140 GHz	100 W	at 95 GHz
50 W	at 220 GHz	30 W	at 140 GHz
5 W	at 280 GHz	1 W	at 220 GHz

width of those EIKs is about 1%. Because of these device characteristics, EIKs have been used for the following military, scientific and commercial applications:

- fire control radar
- seeker
- illuminator
- tracking radar
- low noise cw radar
- surveillance radar
- radio astronomy
- satellite communication
- cloud radar
- radar modelling
- fusion diagnostics
- plasma heating

1.5 Backward Wave Oscillator (BWO)

1.5.1 Introduction

To find the roots of the backward wave oscillator (BWO), one has probably to go back to Rudolf Kompfner, one of the TWT inventors (see Sect. 1.3.1) and the period around 1950 to 1952. In 1951 S. Millmann [31] published on *a spatial harmonics amplifier for 6 mm wavelength* where the energy transport in backward direction with respect to the electron beam is clearly a device feature. Kompfner, working in this period intensely on microwave sources capable of electronic frequency tuning, mentions Millmanns work in his US patent filed in 1952 [32] on *Backward Wave Tubes*. How closely the work on the oscillator type of tubes was related to the amplifier devices is indicated by the patent filed by Kompfner and Williams [33] at the same date on *Backward Wave Amplifiers*. His work published together with N.T. Williams in 1953 [34] belongs also to this period of early investigations on *Backward Wave Tubes*. The investigated devices were based on linear electron beam/slow wave circuit interaction, using periodically disturbed wave guides but also helices. H. Heffner contributed in this period with his *Analysis on Backward Wave Traveling Wave Tubes* [35].

The application, these pioneers had in mind for the oscillator type of devices, was frequency modulated signal generation. Interesting information on this period can be found in J.R. Pierces review article on R. Kompfner and his work [36]. To avoid confusion, it should be mentioned that in some literature the BWO is also called the carcinotron.

1.5.2 BWO Operation Principle

The basic idea of the BWO is to use the space charge bunches of an electron beam interacting along their path with a periodic delay line structure (the first space harmonic). The delay line can be of all types (helices, periodically corrugated waveguides, folded waveguides, interdigital lines, etc.). The preferred type depends, as in TWTs, on the frequency range and power level of interest. When passing the periodic structure, the space charge bunch initiates a wave propagating in backward direction. If the phase velocity and the bunch velocity are such that the total phase delay in the loops is $\theta_n = n \cdot 2\pi$, a backward wave oscillation can start.

In Fig. 1.44 the situation is illustrated by an electron beam passing along a tape helix. The phase delay θ_1 in a closed loop between two adjacent gaps in the delay line is given by the sum of the phase lag $\theta_e = \beta_e p$ during the time the electron bunch

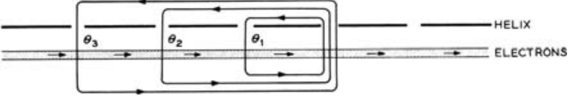


Fig. 1.44. Electrons adjacent to a tape helix interacting with the fields in the helix gaps. θ_1, θ_2, and θ_3 denote loop phase shifts for one, two and three periods, respectively

needs to travel with v_e to the next gap and the phase lag $\theta_w = \beta_w p$ during the time the wave travels back with its phase velocity v_p. Here p is the pitch of the periodic structure, β_e and β_w are the propagation constants of the electron bunch and the wave, respectively. Thus, if $\theta_1 = \beta_e p + \beta \cdot p = 2\pi$, and according to Fig. 1.44 $\theta_2 = 2 \cdot 2\pi$, $\theta_3 = 3 \cdot 2\pi$, etc., one gets a feedback loop system with total gain eventually above 1 depending on the number of sub-loops or the length of the delay line, the coupling strength between electron bunch and electric field of the wave in the gap and the current in the electron beam.

Thus, the synchronism condition for BW oscillation is given by

$$\beta_e p = 2\pi - \beta p. \tag{1.68}$$

This compares simply to the quite different synchronism condition for the forward traveling wave amplification

$$\beta_e p = \beta p. \tag{1.69}$$

It is clear that in a delay line with large dispersion, the synchronism condition (1.68) is met optimally only by one frequency, which depends strongly on the propagation constant β_e and thus the beam voltage. To oscillate, the gain loop condition must also be met. Thus, a BWO oscillates only at a current higher than a certain start current I_{st}. Above this value, the summation of all gains in the loops yields >1. Within the framework of the small signal theory, the condition for the start current is given by

$$\beta L C_{st} = 2\pi(2^{-5/3}) = 1.97 \tag{1.70}$$

with C_{st} being the Pierce parameter (see Sect. 1.3),

$$C_{st} = (K I_{st}/4V)^{1/3}, \tag{1.71}$$

at the start current for oscillation. Below this current no oscillation can occur. I_{st} is of the order of a few mA for a voltage V of a few thousand volts for conventional O-type BWOs as sketched in Fig. 1.45.

To distinguish the O-type BWO where the magnetic field is parallel to the electron beam axis (used for focusing), in Fig. 1.46 a M-type BWO (see also Sect. 1.6.3)

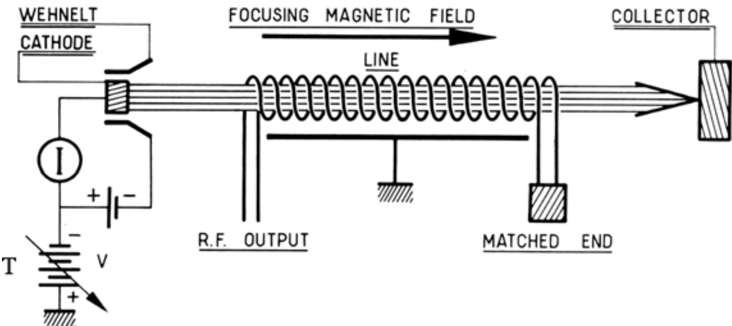


Fig. 1.45. O-type backward wave oscillator (O-carcinotron) with bifilar helix as delay line

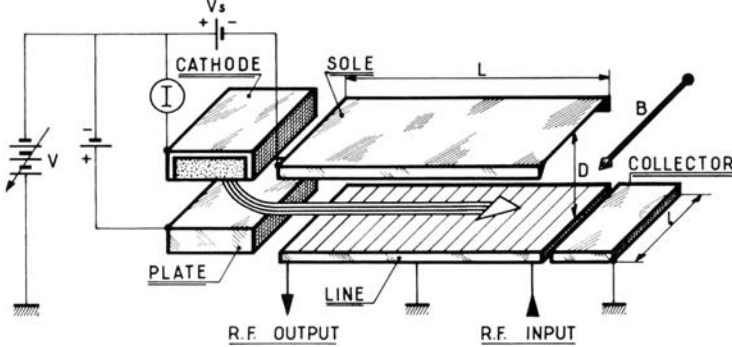


Fig. 1.46. M-type backward wave oscillator (M-carcinotron) with planar periodic delay line

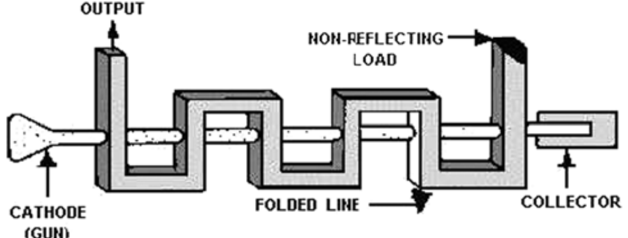


Fig. 1.47. Folded waveguide type of BWO

is sketched, which is an $\boldsymbol{E} \times \boldsymbol{B}$ device. Though the basic interaction physics and synchronism condition is the same for O-type and M-type BWOs, the M-type BWO is an $\boldsymbol{E} \times \boldsymbol{B}$ device with the magnetic field perpendicular to the beam axis and the electric field between delay line ground and sole.

A further example of an O-type BWO with a folded waveguide line is shown in Fig. 1.47. These types or interdigital lines are the most common BWOs. Further examples and a compact description of BWOs is, e.g. given by J. Arnaud in "The Encyclopedia of Electronics" issued by Charles Susskind [37].

1.5.3 BWO Applications

The typical applications for O-type BWOs are local oscillators and frequency tuneable microwave sources up to the THz range. There, 10 mW of CW power can be obtained in the frequency range from 0.1 to 1.5 THz with a tuning range of 200 GHz. The further development of THz BWO sources is a topic of increasing commercial interest.

1.6 Magnetrons and Cross-Field Amplifiers

1.6.1 Magnetrons

Introduction

Magnetrons are oscillators where all functions are grouped in one given volume between cathode and anode. Magnetron tubes are robust, compact and efficient. They are used as the final stage in ordinary transmitters (radars) but their main utilization is in industrial microwave heating applications (Table 1.11).

Microwave Circuit

Magnetrons consist of two coaxial electrodes: an inner cathode of radius r_K placed at the centre, and an outer anode of radius r_A (Figs. 1.48 and 1.49). A permanent magnet (at the origin of the designation of the magnetron) or an electromagnet is used to create a magnetic field B_Z parallel to the axis [9]. The anode is machined with an even number N of slots or resonating cavities which are coupled between themselves by their overflow or leakage field in the cathode/anode space. Consequently, we can distinguish $N/2+1$ distinct resonance frequencies corresponding to phase differences $\Delta\varphi = 0, 2\pi/N, 2\pi/(N/2), 2\pi/(N/3), \pi$ between two adjacent slots or cavities.

The modes are identified by the number n of times that the field pattern is repeated in going around the anode once [7]. In that way, $\Delta\varphi$ becomes $2\pi n/N$. In

Table 1.11. Survey on industrial magnetrons

8.5–9.6 GHz	200 kW peak/200 W average	22 kV × 27.5 A	6.5 kg
2.45 GHz	5 kW CW	7.2 kV × 0.95 A	4.3 kg

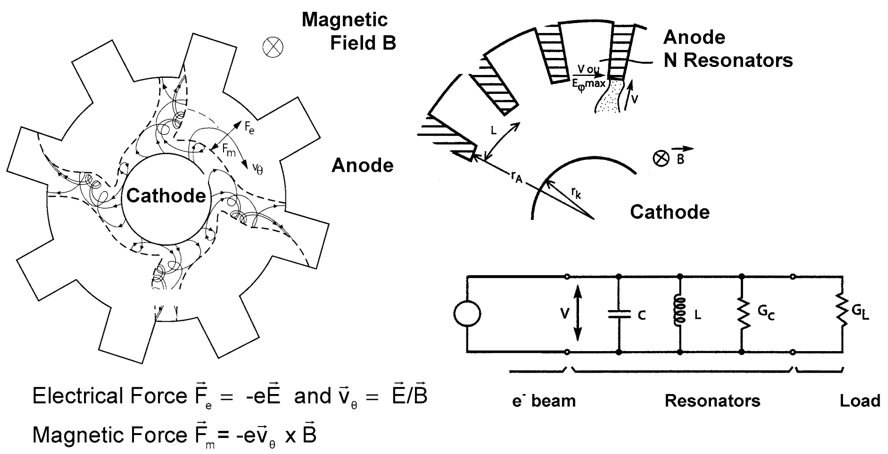


Fig. 1.48. Magnetron (interaction)

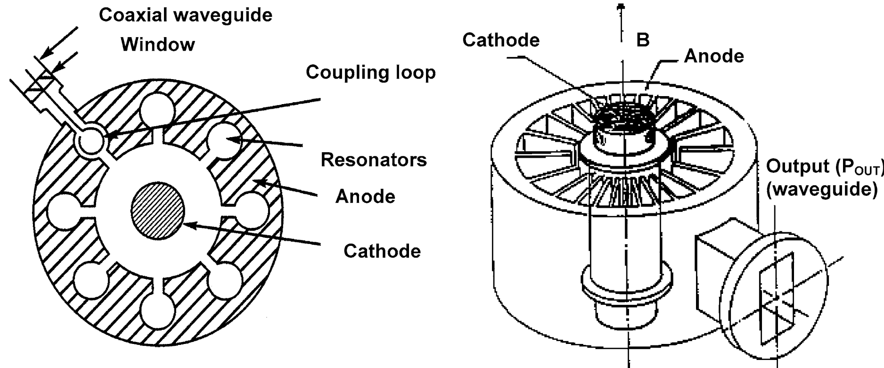


Fig. 1.49. Magnetron

practice, the most common mode is the "π" mode, with $n = N/2$, that is, a phase difference of π between adjacent cavities. The distance between two adjacent cavities is $l = \lambda n/N$, which is equal to $\lambda/2$ if π is the operating mode.

For a given resonance mode of frequency $f = \omega/2\pi$, the electromagnetic field and, for example, its E_φ component, varies along the anode (with a so-called abscissa s) as described by the usual equation of a stationary wave: $E_\varphi(s,t) = A \times \cos(2\pi s/\lambda)\cdot \cos\omega t$. But, since a stationary wave can be considered as the sum of two progressive waves moving in opposite directions, each one with an amplitude equal to half of the original wave, the stationary wave equation can be rewritten as

$$E_\varphi = (A/2)[\cos(2\pi s/\lambda + \omega t) + \cos(2\pi s/\lambda - \omega t)] \tag{1.72}$$

or

$$E_\varphi = (A/2)[\cos\omega(t + s/v_\varphi) + \cos\omega(t - s/v_\varphi)], \tag{1.73}$$

where

$$v_\varphi = \lambda f = \lambda\omega/2\pi. \tag{1.74}$$

Finally, an observer (or an electron, for example) moving along s at a velocity $v_\theta = v_\varphi$ will be in phase with one of the two waves, for example $A/2\cdot\cos\omega(t - s/v_\varphi)$, and will always see the same field.

Magnetrons are therefore multi-resonating microwave structures offering a number of resonance modes, each with the equivalent of two progressive waves, only one of which being taken into consideration for any given mode.

We will now consider the interaction process that takes place within magnetrons, i.e. the transformation of electron movement energy into electromagnetic energy.

Interaction

The electrons emitted by the cathode would move radially towards the anode, but thanks to the magnetic field B_Z, their trajectories are curved and somewhere they

become concentric ($v_r = 0$) to the cathode and the anode [3]. At that point, the electrons are subjected to an electrostatic force $\boldsymbol{F}_e = -e\boldsymbol{E}_{\text{Anode}}$ and a magnetic force $\boldsymbol{F}_m = -e(\boldsymbol{v}_\theta x \boldsymbol{B}_Z)$, acting in opposing directions (Fig. 1.48). Then a radius r exists, between r_K and r_A, for which the amplitudes of these two fields are equal and characterized by $v_r = 0$ and $v_\theta = E_A/B_Z$. Theoretically, the electron can rotate indefinitely around the magnetron axis at this radius.

The combined forces of F_e and F_m also apply – albeit very schematically – to all the electrons leaving the cathode at varying velocities. Also, all the electrons experience, to varying extents, the effects of the electromagnetic field and the space charge of the other electrons. In any case it can be said that most of the electrons rotate around in the space between cathode and anode about once, after which they terminate either on the anode or the cathode. Naturally, these electrons are replaced by those emitted non-stop by the cathode. The end result is a real cloud of electrons rotating between the anode and the cathode at a velocity approximately v_θ [9].

V_A (or E_A) and B_Z are then adjusted so that $v_\theta = v_\varphi$, with $v_\varphi = \omega(lN/n)/2\pi$ the phase velocity, deducted from the previous expressions of l and of v_φ (1.74). As explained above the rotating electrons always see the same microwave field E_φ. If E_φ is an accelerating field, they are continuously accelerated, and if E_φ is a decelerating field, they are continuously slowed down. This electron accelerating/decelerating is comparable to the TWT process, with the difference that, in the magnetron, the electrons follow a closed circular path and electron bunches are now replaced by n rotating electron arms or spokes, where $n = N/2$ if the π mode is used, that is the most common case.

The energy transfer process in the magnetron is a little bit more complicated than that in the tubes examined in the previous sections. This takes place as follows: the electrons in the spokes are attracted to the anode but, at the same time, repulsed by the RF field. In other words, they are continuously transferring their potential energy, acquired via V_A (or E_A) – to the RF field.

The whole microwave structure, including N resonators, can be regarded as a unique resonator, represented by an equivalent circuit similar to the one of the klystron cavity. In Fig. 1.48, we notice not only the 3 elements, L, C and the loss conductance G_c, but also the load G_L. The load G_L is directly experienced by the electrons and converted from the final load – for example, the antenna or the microwaves oven – through the waveguides and the coupling loop or iris located in the bottom wall of one of the resonators. The beam loading is neglected. The voltage V in Fig. 1.48 is the microwave voltage between two anode vanes and is related to the electromagnetic field $E_\varphi(r = r_A)$. Almost all the electrons are supposed to arrive at the anode, that means on the top of the vanes, with a phase such that E_φ $(r = r_A)$ is maximum. Therefore, their radial velocity is $v_r = E_\varphi(r = r_A)\max/B_Z$, while the azimuthal velocity is given by $v_\theta = v_\varphi$.

Moreover, we can write

$$E_\varphi(r = r_A) = MV/l, \tag{1.75}$$

where l is the distance between two vanes and M is the coupling factor between V and the electrons, $M \approx \sin(\omega l/(2v_\theta))/(\omega l/(2v_\theta))$.

The kinetic power – which definitively is lost – of these electrons landing on the anode, is given by

$$P_{\mathrm{A}} = I_{\mathrm{o}}(m/2e)[v_{\mathrm{r}}^2(r = r_{\mathrm{A}}) + v_\theta^2] = I_{\mathrm{o}}(m/2e)\left[\frac{E_\varphi^2(r = r_{\mathrm{A}})\max}{B_Z^2} + v_\theta^2\right]. \quad (1.76)$$

According to the equivalent circuit, the output power P_{OUT} delivered to the load G_{L} is

$$P_{\mathrm{OUT}} = \frac{1}{2}NV^2 \cdot G_{\mathrm{L}} = \frac{N}{2} \cdot \frac{\omega l^2 C}{M^2 Q_{\mathrm{L}}} E_\varphi^2(\max) \quad (1.77)$$

with $Q_{\mathrm{L}} = \omega C / G_{\mathrm{L}} = 1/(R/Q)G_{\mathrm{L}}$ and (1.75).

Thanks to the obvious expression $P_{\mathrm{OUT}} = V_0 I_0 - P_{\mathrm{A}}$, we get the approximate expression of the magnetron efficiency [9]

$$\eta = P_{\mathrm{OUT}}/V_0 I_0 = \left(1 - m v_\theta^2 / 2eV_0\right) / \left(1 + I_0 m M^2 Q_{\mathrm{L}} / B_Z^2 e N l^2 \omega C\right). \quad (1.78)$$

The efficiencies of magnetrons usually are quite large and very attractive, 30% up to 70% and even 80%.

A magnetron presents very few control or optimisation parameters: the high voltage V_{A} and the magnetic field B_Z. This is why an operating zone is defined in the plane (V_{A}; B_Z), where the magnetron correctly oscillates. Such a zone is located between two limits [3, 9]:

- the Hull parabola, below which (for B_Z increasing and/or V_{A} decreasing) the electrons are no more intercepted by the anode and begin to rotate around the cathode-anode space,
- the Hartree line that determines the synchronism condition (for a given mode). Typically, the operating point is quite close to the Hartree line.

From a technological and engineering point of view, the main parts and subassemblies look like the parts of the other microwave tubes, even if the geometries are notably different (Fig. 1.49).

However, let us notice the frequent use of cold cathodes without direct heating and based on the secondary emission. In this case, the primary electrons are the electrons, whose emission conditions are such that they are coming back and bombard the cathode (self-heating). Also quite all the magnetrons include straps, i.e. metallic wires that connect the vanes or the resonators, which must be at the same electromagnetic instantaneous potential. At the π mode, for instance, straps connect the vanes 0, 2, 4, $N-2$, N. Their objective is to minimize the risk of oscillation and the rise of unwanted oscillations or "parasitic" modes.

1.6.2 Cross-Field Amplifiers (CFA)

In the magnetrons and, generally speaking, in all the cross-field devices – also called "M-type" tubes – the electrons move under the influence of the perpendicular static electric field $E \approx V_0/(r_{\mathrm{A}} - r_{\mathrm{K}})$ and magnetic field B_Z (Fig. 1.50). As a result, the

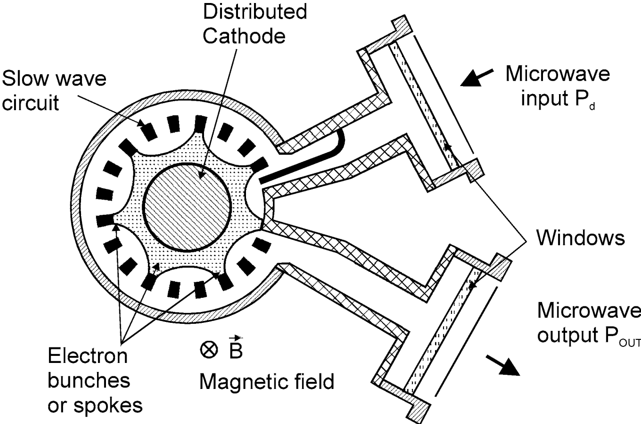


Fig. 1.50. Cross-field amplifier (CFA)

motion is perpendicular to both fields and, when the values of V_0 and B_Z are carefully adjusted, the electron beam or clouds rotate in the narrow space KA between the cathode and the anode. If now the microwave circuit – at the same time the anode – is interrupted to provide input and output connections, we deal no more with a magnetron oscillator but with a cross-field amplifier (CFA).

In CFAs, resonators are preferably replaced by a specific slow wave structure, whose ends are terminated by the input and output couplers, which must be matched at least in the expected frequency bandwidth [3]. Between these two ends, a specific short circular drift space is located in order to minimize any coupling between input and output, but also to provide a necessary electron bunching debunching process (Fig. 1.50).

As soon as the drive power P_{d} is injected, the RF fields give rise to bunches or spokes which rotate in the KA space. This process is nearly the same as in the TWTs, where the beam is coupled to a growing forward or to a backward wave. At the same time, the energy transfer inside the spokes is the same as in the magnetrons [9].

In the case of a forward wave, the electrons and the microwave energy flow in the same direction, and the spokes travel in synchronism with the circuit wave, causing it to grow. In the case of a backward wave, the electrons and the microwave energy flow in opposite directions, and the amplification is obtained when P_{d} is injected at the circuit output. The slow forward wave circuits are often helix structures, when the backward circuits use strapped bar lines or interdigital structures.

It should be remembered that in CFAs and in magnetrons, the slow wave structure acts not only as the anode but also as the collector, and in spite of a usual high efficiency, a lot of power is dissipated in the structure, which must present excellent thermal capabilities.

The CFAs, also known as “amplitrons”, are characterized by impressive output powers (for example, 700 kW peak and 10 kW average), efficiencies around 70%

and instantaneous bandwidths in the order of 5 to 10%. The cathode voltages V_0 are very acceptable ($\leq$50 kV) with perveances of $P \approx 2 \times 10^{-6}$. At the same time, the compactness and the lightness are other impressive advantages. The weak points remain not only the risk of oscillations and the possible noisy behaviour but also the low gains of $\approx$11 dB [3, 9].

As an important precaution, the RF drive pulse P_d must always be applied before V_0, for two main reasons. First, if the CFA employs a cold cathode – now the most frequent situation – the emission process may fail to build up. As a result, the modulator will be unloaded and so excessive voltage, arcing and damage. Secondly, the RF drive anterior to V_0 and strictly present throughout the voltage pulse helps to control the space charge and prevents oscillations at band edge and the generation of broadband noise.

As for the ends of the slow wave circuit, the input VSWR must be very low ($\leq$1.2:1), not only in the frequency band but also outside, in order to avoid any starting of oscillation during the voltage pulse and during the rise and the fall of V_0. The resulting spurious output is referred to as "rabbit ears" because of the way they appear on a time display of P_{OUT}.

In addition, it should be pointed out that the transmission loss of the signal through the CFA in absence of beam current is very low. Then, the CFA can be considered transparent, and this advantageous characteristic is used in many system applications.

1.6.3 Cross-Field Backward Wave Oscillator (MBWO)

The M-type backward wave oscillators, sometimes called "carcinotrons", are voltage tunable oscillators, mainly used as ECM (electronic counter measures) noise generators but also swept signal sources and drivers for high power transmitters. They are able to deliver, for example, 50 to 100 W with efficiencies of 50% from the L to the X band. They look like the former CFAs, but they use an injected thin beam, and the principle of their operation is close to the one of the backward oscillator (BWO) derived from the conventional TWTs [3].

The electron emission is not distributed over the total surface of the inner electrode, but uses a complete gun structure with a cathode of limited length, a "grid" and a controllable anode, this gun structure being at one end of the circular interaction space. The electrons are emitted approximately radially from the cathode towards the controllable anode but, because of the strong axial magnetic field B_Z, they are curved at 90° and enter the interaction region between the sole and the slow wave structure (Fig. 1.51). The sole voltage is a little bit lower than the anode potential in order to provide a radial electric field E_{r0}. Thanks to both E_{r0} and B_Z, the velocity of the electrons becomes E_{r0}/B_Z and is adjusted to get a thin electron beam following a circular path.

Usually the circuits are folded waveguides or interdigital delay lines, where the signal grows as it travels from the collector end to the electron gun end. The microwave output connector is located at this electron gun end. At the collector end,

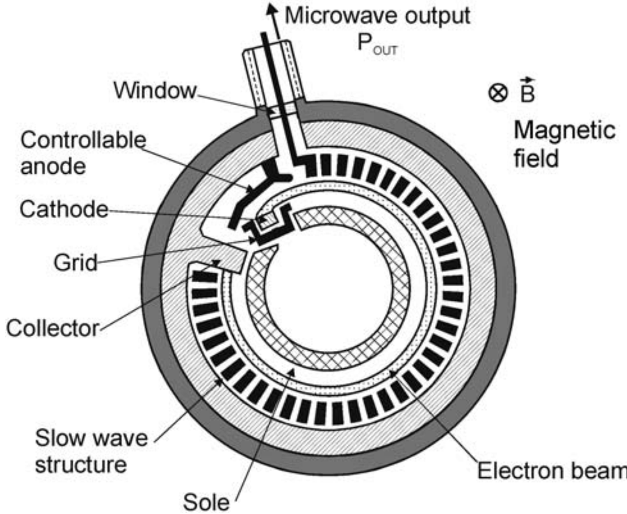


Fig. 1.51. Cross-field backward wave oscillator (MBWO)

the MBWOs require a relatively large attenuator, well bonded to the heat sink of the tube, to absorb the unavoidable microwave reflections.

The electrons interact with a backward-wave space harmonic of the circuit, the energy on the circuit flowing opposite to the direction of the electron motion. This easily provides the feedback necessary for oscillation. The circuit is terminated at the collector end, but the microwave output is removed at the gun end.

Despite this short presentation, it is understandable that the design and the technology of MBWOs are critical, but the above mentioned performances remain very attractive.

1.7 Fast-Wave Devices

Conventional microwave tubes operate with longitudinal phase bunching of the interacting electrons and are commonly referred to as slow-wave devices, since in some way, the RF structure is configured so that the phase velocity of the electromagnetic (EM) field is slowed down to be a little bit lower than the electron beam velocity and is thus less than the velocity of light. This is most evident in the helix-type TWT and BWO where the transverse RF circuit dimensions are typically a fraction of the wavelength. The dimensions of the interaction circuits of klystrons, extended interaction tubes, magnetrons and cross-field amplifiers also have dimensions in the order of the wavelength or smaller. The inherent power limitation with these conventional microwave tubes is the tremendous decrease in the dimensions of the interaction structure with increasing frequency (that is, decreasing wavelength, see Fig. 1.52). Therefore, the possibility of extracting high average power in the millimeter (mm)

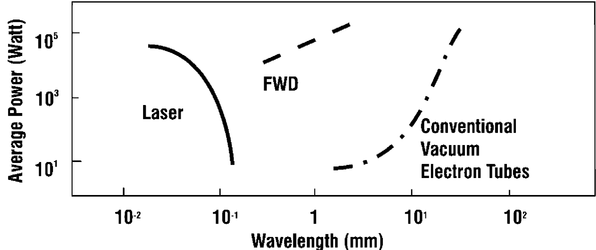


Fig. 1.52. Chart showing the limitation of conventional microwave tubes for the generation of higher power levels at elevated frequencies. Here, FWDs are fast-wave devices

wavelength range has been ruled out. In the case of two-level far-infrared LASERs, the possible average power, which can be generated, decreases with increasing wavelength since the energy between the interacting quantum levels ($hf = 0.41$ meV at 100 GHz) becomes smaller than the thermal energy quantum at room temperature ($kT = 25$ meV). To close the gap for the achievable average output power in the mm- and sub-mm wavelength range (Fig. 1.1), new interaction mechanisms had to be discovered, which operate in highly overmoded interaction circuits, i.e. in fast wave circuits where the electron beam is placed well away from the RF structure. With larger dimensions, the power-handling capability in enlarged. Since fast waves have a phase velocity larger than the velocity of light, the phase bunching mechanism in such fast-wave devices must be generated by a transverse interaction, e.g. (1) the electron–cyclotron interaction in a longitudinal magnetic field ("electron cyclotron maser": ECM) or (2) the electron undulation in a wiggler field ("free-electron laser": FEL).

The origin of the ECMs traces back to the late 1950s, when three investigators began to examine theoretically the generation of microwaves by the ECM interaction: R. Twiss in Australia [38], J. Schneider in the US [39] and A. Gaponov in Russia [40]. A short note on the possibility to use the rotational energy of a helical electron beam for microwave generation was published by German H. Kleinwächter in 1950 [41]. In early experiments with devices of this type, there was some debate about the generation mechanism and the relative roles of fast-wave interactions mainly producing azimuthal electron bunching and slow-wave interactions mainly producing axial bunching. The predominance of the fast-wave ECM resonance with its azimuthal bunching in producing microwaves was experimentally verified in the mid-1960s in the US [42] (where the term "electron cyclotron maser" was apparently coined) and in Russia [43] (where the term "gyrotron" was introduced).

FELs exploit relativistic electron beam technology to upshift the electron wiggle frequency. In this respect, perhaps a more descriptive name is that introduced by R.M. Phillips: UBITRON for an "undulated beam interaction electron tube". Long wavelength FELs ($\lambda \geq 0.5$ mm) are called free-electron masers (FEMs).

1.7.1 Interaction Principles

Fast-wave devices in which the phase velocity v_{ph} of the EM wave is greater than the speed of light c, generate or amplify coherent EM radiation by stimulated emission of bremsstrahlung from a beam of relativistic electrons. The electrons radiate because they undergo oscillations transverse to the direction of beam motion by the action of an external force (field). For such waves the electric field is mainly transverse to the propagation direction (TE or HE modes).

The condition for coherent radiation is that the contribution from the electrons reinforce the original emitted radiation in the oscillator or the incident EM wave in the amplifier. This condition is satisfied if a bunching mechanism exists to create electron density variations of a size comparable to the wavelength of the imposed EM wave. To achieve such a mechanism, a resonance condition must be satisfied between the periodic motion of the electrons and the EM wave in the interaction region [44–47]

$$\omega - k_z v_z \cong s\Omega, \quad s = 1, 2, \ldots \quad (k_z v_z = \text{Doppler term}), \tag{1.79}$$

where ω and k_z are the wave angular frequency and characteristic axial wavenumber, respectively, v_z is the translational electron drift velocity, Ω is an effective frequency, which is associated with macroscopic oscillatory motion of the electrons, and s is the harmonic number.

1. In ECMs, the EM energy is radiated by relativistic electrons gyrating in an external longitudinal magnetic field. In this case, the effective frequency Ω corresponds to the relativistic electron cyclotron frequency

$$\omega_c = \Omega_{co}/\gamma \quad \text{with } \Omega_{co} = eB_o/m_o \quad \text{and } \gamma = [1-(v/c)^2]^{-1/2} = 1 + eV_o/m_o c^2, \tag{1.80}$$

   where $-e$ and m_o are the charge and rest mass of an electron, γ is the relativistic factor, B_o is the magnitude of the guide magnetic field and V_o is the acceleration voltage. The nonrelativistic electron cyclotron frequency is $f_o/\text{GHz} = 28B_o/\text{T}$. A group of relativistic electrons gyrating in a strongmagnetic field will radiate coherently due to bunching caused by the relativistic mass dependence of their gyration frequency. Bunching is achieved because, as an electron loses energy, its relativistic mass decreases and it thus gyrates faster. The consequence is that a small amplitude wave's electric field, while extracting energy from the particles, causes them to become bunched in gyration phase and reinforces the existing wave electric field. The strength of the magnetic field determines the radiation frequency.
   The phase bunching process can be most easily understood in a reference frame in which the axial velocity vanishes. In Fig. 1.53 an annular electron beam with radius R_e is depicted in this frame. The electrons arranged around this annulus execute circular orbits with the Larmor radius $r_L = v_\perp \gamma/\Omega_0$. Typically $r_L \ll R_e$. Initially, the phase of the electrons in their cyclotron orbits is random, as shown in Fig. 1.53 (left).

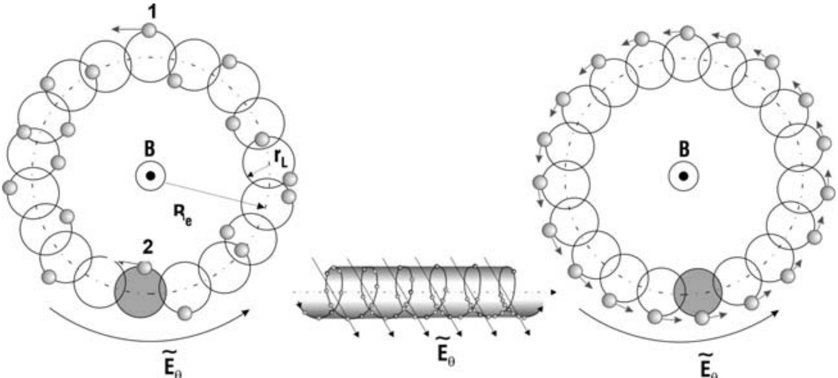


Fig. 1.53. Principle of azimuthal phase bunching in an annular electron beam with initial random phasing of electrons in their cyclotron orbits (*left*) and with electron bunched in phase in their cyclotron orbits (*right*)

In the presence of a transverse RF electric field characteristic of a microwave cavity TE_{mn} mode, the electrons will be accelerated or decelerated. As a specific example, Fig. 1.53 depicts an electric field with only an azimuthal component, as is characteristic of TE_{0n} modes in circular waveguide. With random phasing, there is no net energy exchange. Electron 1 will be decelerated by the azimuthal electric field and thus lose energy, while electron 2 will be accelerated and hence gain an equal amount of energy.
Phase bunching and net transfer can occur if the wave frequency is slightly larger than the initial value of the cyclotron frequency

$$\omega' - \frac{eB_0}{m_e \gamma_0'} = (\delta\omega)' > 0, \tag{1.81}$$

where the subscript 0 denotes the initial value and the prime denotes the reference frame in which the electron axial velocity vanishes. Then, as electron 2 gains energy, its cyclotron frequency decreases; this electron falls farther from resonance gaining less energy on each successive cycle. On the other hand, electron 1, which initially loses energy, experiences an increasing value of ω_c' and moves closer to exact resonance with the electric field, thereby using an increasing amount of energy on each successive cycle. An instability develops in which the wave energy grows in time and the electrons bunch in phase within their cyclotron orbits, as shown in Fig. 1.53 (right).

2. In the case of a spatially periodic magnetic or electric field (undulator/wiggler), the transverse oscillation frequency Ω_b (bounce frequency) of the moving charges is proportional to the ratio of the electron beam velocity v_z to the wiggler field spatial period λ_w. Thus, the operating frequency of such devices, an example of which is the FEM [48–51], is determined by the condition that an electron in its rest frame "observes" both the radiation and the periodic external force at the same frequency. If the electron beam is highly relativistic ($v_{ph} \cong v_z \cong c$),

the radiation will have a much shorter wavelength than the external force in the laboratory frame ($\lambda \cong \lambda_w/2\lambda^2$ so that $\omega \cong 2\gamma^2\Omega_b$). Therefore, FEMs are capable of generating EM radiation of very short wavelength determined by the relativistic Doppler effect. The bunching of the electrons in FEMs is due to the perturbation of the beam electrons by the ponderomotive potential well which is caused by "beating" of the EM wave with the spatially periodic wiggler field. It is this bunching that enforces the coherence of the emitted radiation

$$\Omega_b = k_w v_z, \quad k_w = 2\pi/\lambda_w. \tag{1.82}$$

In the case of the ECMs and FEMs, unlike most conventional microwave sources and lasers, the radiation wavelength is not determined by the characteristic size of the interaction region. Such fast-wave devices require no periodically rippled walls or dielectric loading and can instead use a simple hollow-pipe oversized waveguide as a circuit. These devices are capable of producing very high power radiation at cm-, mm-, and sub-mm wavelengths since the use of large waveguide or cavity cross sections reduces wall losses and breakdown restrictions, permitting the passage of larger, higher power electron beams. It also relaxes the constraint that the electron beam in a single cavity can only remain in a favourable RF phase for half of a RF period (as in klystrons and other devices employing transition radiation). In contrast with klystrons, the reference phase for the waves in fast-wave devices is the phase of the electron oscillations. Therefore, the departure from the synchronism condition, which is given by the transit angle $\theta = (\omega - k_z v_z - s\Omega)L/v_z$, can now be of order 2π or less, even in cavities or waveguides that are many wavelengths long.

1.7.2 Dispersion Diagrams of Fast-Wave Interactions

Many configurations can be used to produce coherent radiation based on the ECM instability. The departure point for designs based on a particular concept is the wave–particle interaction. Dispersion diagrams, also called ω–k_z plots or Brillouin diagrams [44–47], show the region of cyclotron interaction (maximum gain of the instability) between an EM mode and a fast electron cyclotron mode (fundamental or harmonic) as an intersection of the waveguide mode dispersion curve (hyperbola)

$$\omega^2 = k_z^2 c^2 + k_\perp^2 c^2 \tag{1.83}$$

with the beam-wave resonance line (straight) given by (1.79). In the case of a device with cylindrical resonator, the perpendicular wavenumber is given by $k_\perp = X_{mn}/R_o$, where X_{mn} is the nth root of the derivative of the corresponding Bessel function (TE_{mn} modes) and R_o is the waveguide radius. Phase velocity synchronism of the two waves is given in the intersection region. The interaction can result in a device that is either an oscillator or an amplifier. In the following subsections, different ECM devices and the FEM are classified according to their dispersion diagrams. Table 1.12 presents a review of ECM devices (also called gyro-devices) and a comparison with the corresponding conventional linear beam (O-type) devices.

Table 1.12. Review of the interaction circuits of gyro-devices and comparison with corresponding conventional linear-beam (O-type) devices

"O" TYP DEVICES	MONOTRON	KLYSTRON	TWT	TWYSTRON	BWO
TYPE OF GYRO-DEVICE	GYRO-MONOTRON	GYRO-KLYSTRON	GYRO-TWT	GYRO-TWYSTRON	GYRO BWO

Gyrotron Oscillator and Gyro-Klystron Amplifier

Gyrotron oscillators and gyro-klystrons are devices which usually utilize only weakly relativistic electron beams ($V_o < 100\,\text{kV}$, $\gamma < 1.2$) with high transverse momentum (pitch angle $\alpha = v_\perp / v_z > 1$) [46, 47]. The wavevector of the radiation in the cavity is almost transverse to the direction of the external magnetic field ($k_\perp \gg k_z$, and the Doppler shift is small) resulting, according to (1.79) and (1.80), in radiation near the electron cyclotron frequency or one of its harmonics

$$\omega \cong s\Omega_c, \quad s = 1, 2, \ldots . \tag{1.84}$$

In the case of cylindrical cavity tubes, the operating mode is close to cutoff ($v_{ph} = \omega / k_z \gg c$), and the frequency mismatch $\omega - s\Omega_c$ is small but positive in order to achieve correct phasing, i.e. keeping the electron bunches in the retarding phase. The Doppler term $k_z v_z$ is of the order of the gain width and is small compared with the radiation frequency. The dispersion diagrams of fundamental and harmonic gyrotrons are illustrated in Figs. 1.54 and 1.55, respectively. The velocity of light line is determined by $\omega = ck_z$. For given values of γ and R_o, a mode represented by X_{mn} and oscillating at frequency ω is only excited over a narrow range of B_o. Quasi-optical gyrotrons employ a Fabry–Perot mirror resonator perpendicular to the electron beam, also providing $k_\perp \gg k_z$ [46].

Cyclotron harmonic operation reduces the required magnetic field for a given frequency by the factor s. At low voltages, the number of electron orbits required for efficient bunching and deceleration of electrons can be large, which means that the resonant interaction has a narrow bandwidth and that the RF field may have moderate amplitudes. In contrast with this, at high voltages, electrons should execute only about one orbit. This requires correspondingly strong RF fields, possibly leading to RF breakdown, and greatly broadens the cyclotron resonance band, thus making possible an interaction with many parasitic modes.

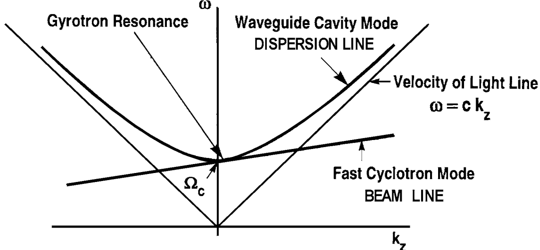


Fig. 1.54. Dispersion diagram of gyrotron oscillator (fundamental resonance)

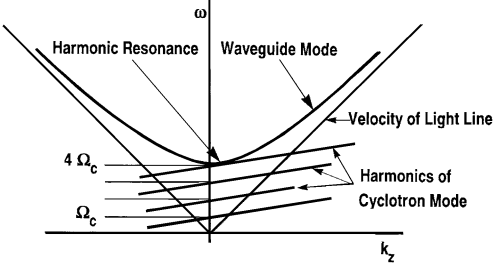


Fig. 1.55. Dispersion diagram of harmonic frequency gyrotron oscillator

Cyclotron Autoresonance Maser (CARM)

In a gyrotron with a highly relativistic beam (≥ 1 MeV), an efficient interaction will lead to an average energy loss in the order of the initial electron energy. As a result, the change in the gyrofrequency is much greater than in the mildly relativistic case. It is therefore desirable to identify the condition under which such a highly relativistic electron beam remains in synchronism with the RF field. A possibility for achieving synchronism is to utilize the interaction of electrons with EM waves propagating with a phase velocity close to the speed of light in the direction of the magnetic field. In this case, the Doppler shift term $k_z v_z$ is large, and the appropriate resonance condition is

$$\omega \cong k_z v_z + s\Omega_\mathrm{c}. \tag{1.85}$$

If $v_\mathrm{ph} \cong c$, the increase in cyclotron frequency due to extraction of beam energy (decrease of γ) nearly compensates the decrease in the Doppler shifted term. Therefore, if the resonance condition is initially fulfilled, it will continue to be satisfied during the interaction. This phenomenon is called autoresonance, and the cyclotron maser devices operating in the relativistic Doppler-shifted regime are called cyclotron autoresonance masers [52]. Figure 1.56 shows how the Brillouin diagram of the fast cyclotron wave changes during the autoresonance interaction such that the working frequency ω remains constant even though both Ω_c and v_z are based on the same instability mechanism as that of the gyrotron but changing. The CARM interaction corresponds to the upper intersection and is operated far above cutoff. The instability is convective, so a feedback, e.g. by a Bragg resonator [52], is required

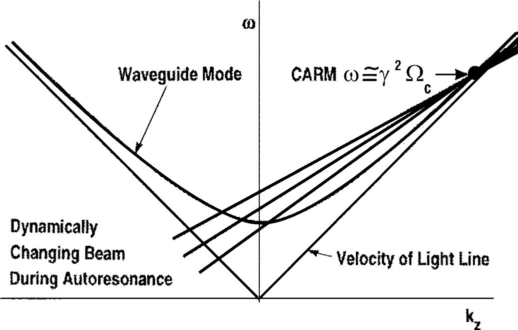


Fig. 1.56. Dispersion diagram of the cyclotron autoresonance maser (CARM)

for an oscillator; and it is necessary to carefully discriminate against the other interactions corresponding to the lower frequency intersection in the dispersion diagram (Fig. 1.56). The problem can be alleviated by employing the fundamental TE_{11} or (HE_{11} hybrid mode) and properly choosing system parameters to be within the stability limit. Compared to a gyrotron, there is a large Doppler frequency upshift of the output ($\omega \cong \gamma^2 \Omega_c$) permitting a considerably reduced magnetic field B_o. Since the axial bunching mechanism can substantially offset the azimuthal bunching, the total energy of the beam and not only the transverse component is available for RF conversion.

In contrast to the gyrotron, the CARM has an electron beam with low to moderate pitch angle ($\alpha < 0.7$). The efficiency of CARMs is extremely sensitive to spread in the parallel beam velocity. The velocity spread $\Delta v_z / v_z$ must be lower than 1% to achieve the full theoretically expected efficiency of 40% [52].

Gyro-Travelling Wave Tube (Gyro-TWT)

From the theoretical point of view, the gyro-TWT differs from the CARM only in regimes of operation. The gyro-TWT utilizes a moderately relativistic electron beam to interact with a fast waveguide mode in an oversized structure near the grazing intersection of the frequency versus wavenumber plot (see Fig. 1.57) where the resonance line is tangent to the EM mode. This produces high gain and efficiency because the phase velocities of the two modes are nearly matched and the group velocity of the waveguide mode is nearly equal to v_z. In the gyro-TWT regime ($\omega / k_z \gg c$), the axial bunching mechanism is too weak to be of any significance. To benefit from autoresonance, the cutoff frequency should be reduced relative to the cyclotron frequency.

Gyro-Backward Wave Oscillator (Gyro-BWO)

If the electron beam and/or magnetic field is adjusted so that the straight fast-wave beam line crosses the negative k_z-branch of the waveguide mode hyperbola (see

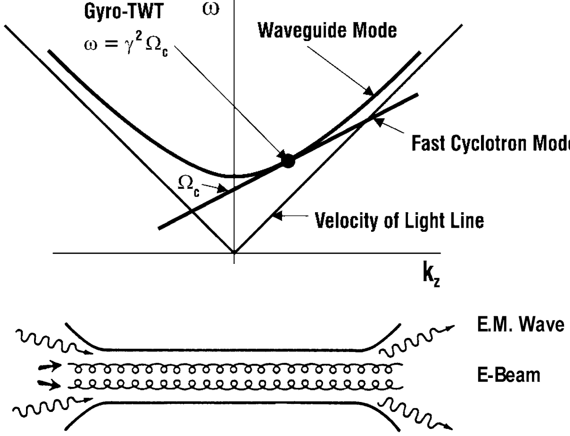


Fig. 1.57. Dispersion diagram and scheme of interaction circuit of gyro-TWT amplifier

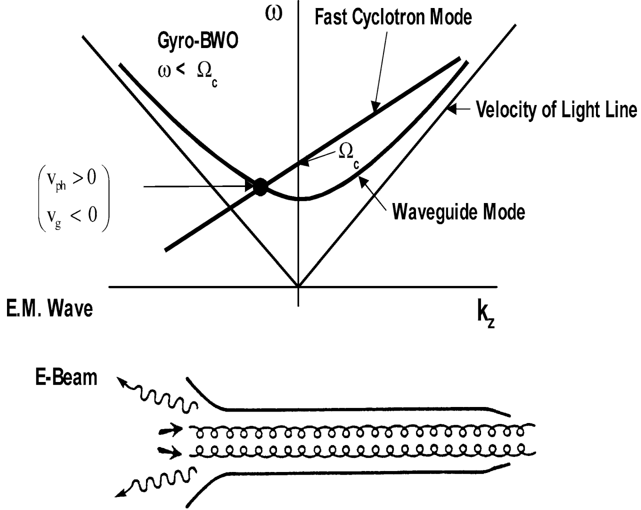


Fig. 1.58. Dispersion diagram and scheme of interaction circuit of gyro-BWO

Fig. 1.58), then an absolute instability (internal feedback) with a “backward wave” occurs. In the gyro-BWO the frequency of operation is now governed by the slope of the line, which is a function of v_z, and thus of the beam acceleration voltage V_o. Consequently, just as in the case of other BWOs (e.g. carcinotron), the frequency of oscillations can be continuously changed very fast over a broad range, using V_o in place of B_o. However, there is a Doppler down shift in frequency ($\Omega_c/2 < \omega < \Omega_c$), so that very high magnetic fields are required for high frequency operation.

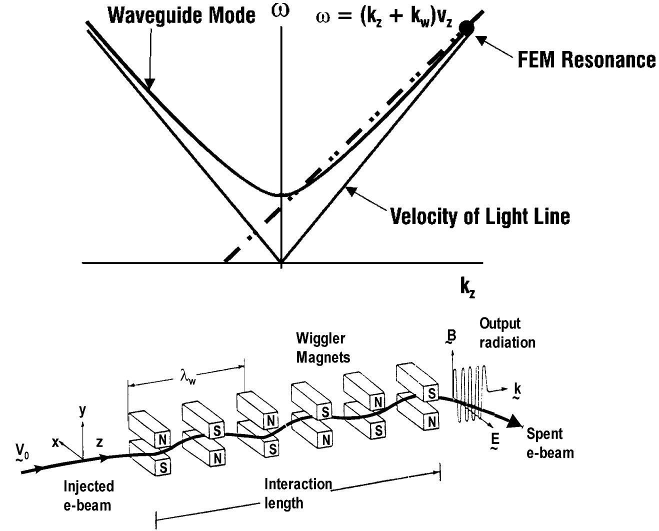


Fig. 1.59. Dispersion diagram and scheme of interaction of FEM

Free-Electron Maser (FEM)

The resonance condition in FEMs is given by

$$\omega - k_z v_z = k_\mathrm{w} v_z \quad \text{with } k_\mathrm{w} = 2\pi/\lambda_\mathrm{w}, \tag{1.86}$$

where λ_w is the spatial period of the wiggler. The generated radiation has a much shorter wavelength than the external force in the laboratory frame: $\lambda \cong \lambda_\mathrm{w}/2\gamma^2$. The basis FEM configuration and the corresponding dispersion diagram is shown in Fig. 1.59. Electrons in the injected electron beam undulate in the periodic magnetic field B_w of the wiggler.

1.7.3 Gyrotron Oscillator

Gyrotron oscillators (gyromonotrons or simply gyrotrons) were the first ECMs to undergo major development. Increases in device power were the result of Russian developments from the early 1970s in magnetron injection guns, which produce electron beams with the necessary transverse energy (while minimizing the spread in transverse energies) and in tapered, open-ended waveguide cavities that maximize efficiency by tailoring the electric field distribution in the resonator [46]. Typical conventional gyrotrons are built as shown schematically in Fig. 1.60. A gyrotron can be described as follows. The magnetron injection gun produces an annular electron beam with the desired beam parameters. The beam is transported to the interaction region, where the interaction cavity converts a fraction of the beam power to RF power. In case of axial output coupling, the spent beam will be collected on the uniform output waveguide section after the uptaper, and the RF power in the TE_{mn} mode

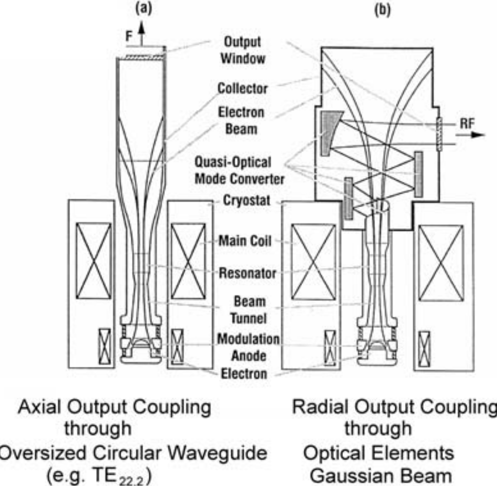


Fig. 1.60. Schematic of gyrotron with solenoids, magnetron injection gun with annular emitter ring, beam tunnel, cylindrical interaction cavity with cutoff section and output taper, cylindrical output waveguide and RF window: **a** with axial output coupling; **b** with radial output coupling [47]

is coupled through the axial output vacuum window. In the case of radial output coupling, a quasi-optical mode converter is connected to the output waveguide and it transforms the rotating TE_{mn} mode with an axial power flow to a Gaussian mode with a radial power flow. The power is then transmitted through a radially located vacuum window, and the spent beam is dissipated on the collector.

A strong externally applied magnetic field in the interaction region is chosen such that the cyclotron frequency or one of its harmonics is close to the frequency of the RF field in the beam frame of reference. The interaction region consists of an open-ended waveguide cavity, usually with a circular transverse cross-section. The electrons in the beam must have a substantial transverse velocity $v_\perp$ and the usual longitudinal velocity v_z. Most of this transverse velocity comes as a result of adiabatic compression resulting from the increasing magnetic field leading to the interaction region. The final ratio of transverse to longitudinal velocity $\alpha = v_\perp / v_z$ in the interaction region is typically between 1 and 2 for gyrotrons that use magnetron injection guns with thermionic cathodes, mostly in temperature limited operation. The electrons follow helical paths around the lines of force of the external field. At high frequencies the use of superconductive magnets is mandatory. By variation of the magnetic field, a sequence of discrete modes can be excited. The frequency scaling is determined by the value of B_o/γ. Modern high power high order volume mode gyrotron oscillators for fusion plasma applications employ an internal quasi-optical mode converter with lateral microwave output, a single-stage depressed collectors (SDC) for energy recovery and a chemical vapour deposition (CVD) diamond window.

Gyrotrons for Fusion Plasma Heating

At present, gyrotron oscillators are mainly used as high power mm-wave sources for electron cyclotron heating (ECH) applications and for diagnostics of magnetically confined plasmas in controlled thermonuclear fusion research [53]. Long-pulse (a few sec) gyrotrons utilizing open-ended cylindrical resonators which generate output powers of 100–500 kW per unit, at frequencies in the range 28–160 GHz, have been used very successfully for plasma formation, ECH and local current density profile control by noninductive electron cyclotron current drive (ECCD) in tokamaks and stellarators. Gyrotron complexes with total power of up to 4.5 MW have been installed. As experimental devices become larger and operate at higher magnetic fields ($B_o = 6$ T) and higher plasma densities ($n_{eo} = 1$–$2 \times 10^{20}/m^3$) in steady state, present and forthcoming ECRH requirements call for gyrotron output powers of at least 1 MW, CW at frequencies ranging from 100–170 GHz. Since efficient ECH&CD need axisymmetric, narrow, pencil-like mm-wave beams with well defined polarization, single-mode emission is necessary in order to generate a TEM_{00} Gaussian beam mode at the plasma torus launching antenna. Single-mode mm-wave gyrotron oscillators capable of high average power, 0.5–1 MW per tube, in long-pulse or CW operation, are currently under development in several scientific and industrial laboratories [46, 47, 53, 54].

Table 1.13 and Fig. 1.61 summarize the present status of long-pulse gyrotrons for EC H&CD applications at 110–170 GHz [54]. The maximum pulse length of commercially available 140 GHz megawatt-class gyrotrons employing synthetic diamond output windows is 30 minutes (CPI and European FZK-CRPP-CEA-TED collaboration). The world record parameters of the European 140 GHz gyrotron are: 0.92 MW output power at 30 min pulse duration, 97.5% Gaussian mode purity and 44% efficiency, employing a single-stage depressed collector for energy recovery. A maximum output power of 1.2 MW in 4.1 s pulses was generated with the JAEA-TOSHIBA 110 GHz gyrotron. The Japan 170 GHz ITER gyrotron holds the energy world record of 2.16 GJ (0.6 MW, 60 min) for tubes with an output power of more than 0.5 MW. The Russian 170 GHz ITER gyrotron achieved 0.5 MW with a pulse duration of 300 s. Figure 1.62 shows the 118 GHz and 140 GHz long-pulse gyrotrons developed by the EU team.

To achieve output powers in excess of 2 MW at frequencies around 170 GHz and long pulses, it is necessary to employ a coaxial-cavity geometry. A maximum output power of 2.2 MW (1 ms pulse length) was obtained at FZK with an efficiency of 28%. At the nominal output power of 1.5 MW the efficiency increases from 30% to 48% in operation with an SDC [47]. Two MW synthetic diamond windows are feasible. At power levels around 1 MW the coaxial-cavity gyrotron can probably operate at frequencies >300 GHz. This means that the gyrotron oscillator is the dominant mm-wave source for high power ECH and ECCD. Even in the case of local noninductive current drive for suppression of plasma instabilities in future tokamak fusion reactors, the gyrotron is a good option, since ultra broadband Brewster windows and specific magnets allow stepwise frequency tuning in the seconds time-scale in the full D-band (110–170 GHz) [47]. Diagnostic gyrotrons deliver $P_{OUT} = 40$ kW

Table 1.13. Development status of long pulse gyrotron oscillators for fusion plasma applications at 110–170 GHz [54]

Institution	Frequency (GHz)	Cavity mode	Output mode	Power (MW)	Efficiency (%)	Pulse length(s)	Fusion device
CPI,	110	$TE_{22,6}$	TEM_{00}	1.05	31	5.0	D III-D
Palo Alto				0.6	31	10.0	D III-D
	140	$TE_{28.7}$	TEM_{00}	0.9	33 (SDC)	1800	W7-X
GYCOM-M	110	$TE_{19,5}$	TEM_{00}	0.93	36	2.0	D III-D
(TORIY, IAP),				0.5	35	5.0	D III-D
Moscow,				0.35	33	10.0	D III-D
Nizhny	140	$TE_{22,6}$	TEM_{00}	0.96	36	1.2	ASDEX-U
Novgorod				0.54	36	3.0	W7-AS
	170	$TE_{25,10}$	TEM_{00}	0.9	44 (SDC)	21	ITER
				0.5	40 (SDC)	300	ITER
GYCOM-N	140	$TE_{22,6}$	TEM_{00}	0.8	32	0.8	W7-AS
(SALUT, IAP),				0.88	50.5 (SDC)	1.0	W7-AS
N. Novgorod	158.5	$TE_{24,7}$	TEM_{00}	0.5	30	0.7	T 10
JAEA,	110	$TE_{22,6}$	TEM_{00}	1.2	38 (SDC)	4.1	JT 60-U
TOSHIBA,				1.0	36 (SDC)	5.0	JT 60-U
Naka,				0.5	34 (SDC)	16.0	JT 60-U
Otawara	170	$TE_{31,8}$	TEM_{00}	1.0	43.4 (SDC)	800	ITER
				0.6	45.5 (SDC)	3600	ITER
THALES,	118	$TE_{22,6}$	TEM_{00}	0.53	32	5.0	TORE SUPRA
CEA, CRPP,				0.35	23	111	TORE SUPRA
FZK,	140	$TE_{28,8}$	TEM_{00}	1.0	49 (SDC)	12	W7-X
EUROPE				0.92	44 (SDC)	1800	W7-X

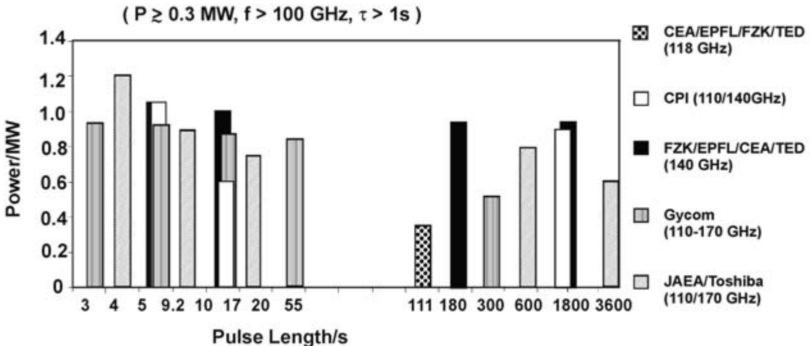


Fig. 1.61. State-of-the-art of long-pulse gyrotrons for EC H&CD applications at 110–170 GHz

with $\tau = 40\,\mu s$ at frequencies up to 650 GHz ($\eta \geq 4\%$) [53]. Operating at the fundamental or the second harmonic of the electron cyclotron frequency enables the gyrotron to act as such a medium power step tunable, mm-wave, sub-mm wave and

Fig. 1.62. CEA/CRPP/FZK/TED-gyrotrons. *Left*: 118 GHz, 0.35 MW, 11 s with liquid nitrogen Cryo-Window. *Right*: 140 GHz, 0.92 MW, 1900 s with CVD-diamond window and SDC

THz source in the frequency range from 38 GHz (fundamental) to 1005 GHz ($TE_{6,11}$ mode, second harmonic) [54].

Gyrotrons for Industrial Applications in Materials Processing

Recently, CW gyrotrons have also been successfully utilized in materials processing (e.g. sintering of advanced nanocrystalline- and piezo-ceramics, surface hardening, joining or dielectric coating of metals and alloys) and in plasma chemistry [53]. Such technological applications require sources with the following parameters: $f \geq$ 24 GHz, $P_{\mathrm{OUT}} = 10$–30 kW, $\eta \geq 30\%$. The present state-of-the-art of industrial CW gyrotrons for technological applications is summarized in Table 1.14 [54].

The use of gyrotrons appears to be of great interest if one can realize a relatively simple, low cost device which is easy to use (such as a magnetron). Gyrotrons with low magnetic field (operating at the second harmonic of the electron cyclotron frequency) which can be provided by a permanent magnet system, low anode voltage, high efficiency and long lifetime are under development (see Table 1.14).

1.7.4 Gyro-Amplifiers

Bunching of electrons in gyro-devices has much in common with that in conventional linear electron beam devices, namely, klystron, TWT, twystron and BWO. In both cases the primary energy modulation of electrons gives rise to bunching (azimuthal or longitudinal) which is inertial. The bunching continues even after the primary modulation field is switched off (at the drift sections of klystron-type and twystron-type devices). This analogy suggests the correspondence between linear-beam (O-type) devices and various types of gyro-devices (Table 1.12).

Table 1.14. Performance parameters of industrial CW gyrotron oscillators for technological applications [54]

Institution	Frequency (GHz)	Cavity mode	Output mode	Power (kW)	Effciency (%)	V_o (kV)	Magnet
CPI,	28	TE_{02}	TE_{02}	15	38	40	room temp.
Palo Alto	$28(2\Omega_c)$	TE_{02}	TE_{02}	10.8	33.6	30	room temp.
	60	TE_{02}	TE_{02}	30	38	40	cryo. mag.
CPI, NIFS	84	$TE_{15,3}$	TEM_{00}	50	14	80	cryo. mag.
GYCOM/	$24.15(2\Omega_c)$	TE_{11}	TE_{11}	3.5	23	12	room temp. PM, 116 kg
IAP,	24.15	TE_{32}	TE_{32}	36	50	33	room temp.
Nizhny	$23(2\Omega_c)$	TE_{12}	TE_{12}	13	50	25	room temp.
Novgorod				28	32	25	room temp.
	$28.3(2\Omega_c)$	TE_{12}	TE_{12}	12	20	25	PM, 68 kg
	$30(2\Omega_c)$	TE_{02}	TE_{02}	10	42	26	room temp.
				30	35	26	room temp.
	37.5	TE_{62}	TEM_{00}	20	35	30	cryo. mag.
	83	TE_{93}	TEM_{00}	10–40	30–40	25–30	cryo. mag.
MITSUBISHI, Amagasaki	$28(2\Omega_c)$	TE_{02}	TE_{02}	15	38.7	21	PM, 600 kg tapered B
UNIV. Fukui	300	$TE_{22,8}$	TEM_{00}	2.0	11	15	cryo. mag.

Figure 1.63 shows the cross-section and the photograph of a two cavity gyro-klystron amplifier. Advanced devices use several staggered cavities (up to 5) and optimized shape of magnetic field. The state-of-the-art of weakly relativistic gyro-klystrons is given in Table 1.15 [54].

Table 1.16 summarizes the status of the high-power relativistic gyro-klystron development at the University of Maryland [54]. For comparison, the experimental results of the SLAC periodic permanent magnet 11.4 GHz klystron are: 75 MW output power with 48% efficiency and 55 dB gain at a pulse duration of 1.5 μs.

The circuit employed in a gyro-TWT consists simply of a modestly oversized waveguide. Since no resonant structures are present, the gyro-TWT is potentially capable of much larger bandwidth than a gyro-klystron. Recent devices employ tapered magnetic field, interaction circuit and two partially loaded stages in order to optimize the beam-wave interaction along the waveguide [55].

The sensitivity to velocity spread can be strongly reduced by coupling between the second harmonic cyclotron mode of a gyrating electron beam and the radiation field in the region of near infinite phase velocity over a broad bandwidth by using a cylindrical waveguide with a helical corrugation on its inner surface [56]. The state-of-the-art of weakly relativistic pulse gyro-TWTs is given in Table 1.17.

The gyro-twystron, a hybrid device, is derived from the gyro-klystron by extending the length of the drift section and replacing the output cavity with a sligthly tapered waveguide section like in a gyro-TWT. The output waveguide section is excited

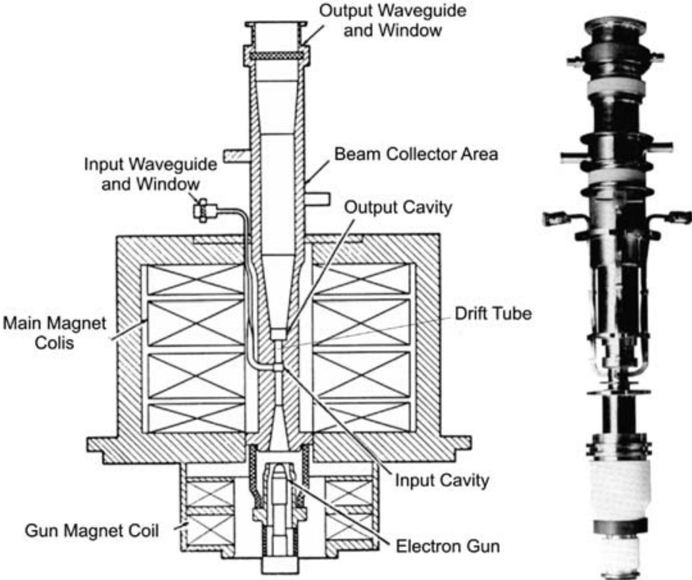


Fig. 1.63. Schematic cross-section and photograph of a pulsed gyro-klystron amplifier (CPI, formerly Varian)

Table 1.15. Weakly relativistic gyro-klystron experimental results

Institution	Frequency (GHz)	Mode	No. of cavities	Power (kW)	Effciency (%)	Gain (dB)	BW (%)
CPI, Palo Alto	10 ($2\Omega_c$)	TE_{01}	3	20	8.2	10	0.2
	28	$TE_{01/02}$	2	76	9	30	0.2
CPI, Litton, NRL,	93.8	TE_{01}	4	118 (10 av.)	29.5	24.7	0.64
UNIV. MARYLAND			5	130 (10 av.)	33	39.5	0.75
GYCOM- M	35.2	TE_{02}	2	750 (5 av.)	24	20	0.6
(TORIY),			2	350	32	19	0.9
Moscow	35.0	TE_{01}	3	250 (1.2 av)	35	40	1.4
IAP,	15.8	TE_{02}	3	160	40	30	0.5
Nizhny	32.4 ($2\Omega_c$)	TE_{02}	3	300	22	22	0.1 PM, 350 kg
Novgorod	35.1 ($2\Omega_c$)	TE_{02}	2	258	18	17	0.3
	93.5	TE_{02}	2	207	30	21	0.2
	93.2	TE_{02}	3	340	24.5	23	0.3
IAP, ISTOK	91.6	TE_{01}	4	2.5 (CW)	25	31	0.36
NRL,	4.5	TE_{10}	3	54	30	30	0.4
Washington, D.C.	34.9	TE_{01}	3	225	31	30	0.82
	93.4	TE_{01}	4	60	25	27	0.69
				84	34	42	0.37
			5	72	27	48	0.44

Table 1.16. Relativistic pulse gyro-klystron experimental results [54]

Institution	Frequency (GHz)	Mode	No. of cavities	Power (MW)	Effciency (%)	Gain (dB)	Type
IAP, Nizhny Novgorod	30	$TE_{5,3}$	2	5	25	27	$TE_{5,2}/TE_{5,3}$
UNIV. MARYLAND	8.57	TE_{01}	3	75	32	30	coaxial
	9.87		2	24	30	33	max. power
	9.87	TE_{01}	3	27	32	36	max. efficiency
			3	16	37	33	max. gain
			3	20	28	50	coaxial
	17.14 ($2\Omega_c$)	TE_{02}	3	27	13	25	coaxial
			4	18.5	7.0	23.3	
	19.76 ($2\Omega_c$)	TE_{02}	2	32	29	27	
	29.57 ($3\Omega_c$)	TE_{03}	2	1.8	2.0	14	

Table 1.17. Development status of weakly relativistic gyro-TWTs (short pulse) [54]

Institution	Frequency (GHz)	Mode	Power (kW)	Efficiency (%)	Gain (dB)	Bandwidth (%)
CPI,	5.18	TE_{11}	120	26	20	7.3
Palo Alto	93.7	TE_{11}	28	7.8	31	2
NRL, Washington,	32.3	TE_{10}	50	28	25	11
D.C.	34.0	TE_{01}	137	17	47	3.3
	35.6	TE_{11}	70	17	60	17
IAP, Nizhny Novgorod	36.3	TE_{21}/TE_{11}	180	27	27	10
MIT, Cambridge	140	$HE^{()}_{061}$	30	12	29	1.6
UNIV. HSINCHU	34.2	TE_{11}	62	21	33	12
	33.6	TE_{11}	93	26.5	70	8.6
UC LOS ANGELES	9.3	TE_{10}	55	11	27	11
	15.7 ($2\Omega_c$)	TE_{21}	207	12.9	16	2.1
	93.4	TE_{01}	75	22	60	4.5

by the beam of electrons that are bunched because of modulation in the input cavity. The gyro-twystron configuration can mitigate the problem of microwave breakdown at high power levels, since the microwave energy density in the output waveguide can be much smaller than in an output cavity.

The development of high-power gyro-amplifiers has opened up a number of possible applications to advanced mm-wave radars for high resolution ranging and imaging in atmospheric and planetary science (e.g. cloud and space-debris monitoring) as well as for deep-space and specialized satellite communications.

A 120-element phased-array 34 GHz radar system ($4'$ beam width) using two 0.5 MW gyro-klystrons with 50 MHz bandwidth, 100 μs pulse duration and 0.01 duty factor is operating in Russia. In the USA, a W-band radar system employing

a 92 kW, 94 GHz gyro-klystron with 420 MHz bandwidth and a duty factor of 0.11 (10 kW average power) is operating.

The cost of a future linear electron–positron collider with centre of mass energy in the TeV range will depend both on the number of microwave phase-coherent amplifiers required to drive the collider and on the cost of each amplifier. In the past, the microwave amplifier of choice for driving the highest energy linear electron accelerators has been the klystron (e.g. S-band). In considering microwave amplifier requirements for future supercolliders there has been a widespread perception that higher frequency and higher peak power will be required. Because of inherent limitations when simultaneously handling high power and high frequencies in klystron circuits, other types of amplifiers as relativistic gyro-klystrons and gyro-twystrons have been explored.

The klystron development programs will probably result in producing microwave amplifiers that are suitable for linear colliders in the energy range up to 1 TeV. However, for multi-TeV supercolliders, 34 GHz gyro-klystrons producing about 100 MW of peak power in 1 μs are needed [53].

Due to their larger bandwidth, gyro-TWTs can be used as output amplifiers in mm-wave communication systems.

1.7.5 Gyro-BWO

Gyro-BWOs are being developed as fast and continously frequency tunable drivers of FEMs and for materials processing applicators with extremely homogeneous EM-field distribution (by frequency sweeping). The status of weakly relativistic devices is summarized in Table 1.18.

1.7.6 Free-Electron Maser (FEM)

Free-electron lasers (FELs) differ from the other high-power microwave sources considered in this chapter in that they have demonstrated output over a range of frequencies extending far beyond the microwave spectrum, well into the visible and ultraviolet range [48–51]. To achieve this spectral versatility, FELs exploit relativistic beam technology to upshift the electron "wiggle" frequency by an amount roughly proportional to γ^2. The magnetostatic wiggler is the most common, but not the sole means, for providing electron undulation. An electrostatic wiggler or the oscillatory field of a strong electromagnetic wave can also play this role. The distinction between long wavelength free-electron maser (FEM) ($\lambda \geq 0.5$ mm) and short wavelength FELs is natural because higher current and lower energy beams are typically employed in this regime and space-charge effects are more important. In particular, the dominant interaction mechansim is often coherent Raman scattering. Also, while short wavelength FELs excite optical modes, dispersion due to the beam dielectric effects and finite transverse dimensions in the drift tubes and cavities are important effects at longer wavelengths. A low power (3 W, 2 ms pulses) FEL operating at radio frequencies (FER) employing a 420 V, 0.2 A electron beam holds the world record for long

Table 1.18. Development status of weakly relativistic pulse gyro-BWOs [54]. Only the 24.7 GHz device is operating in CW

Institution	Frequency (GHz)	Mode	Power (kW)	Efficiency (%)	Bandwidth (%)
IAP, Nizhny Novgorod, FZK, Karlsruhe	24.7	TE_{21}/TE_{11}	7	15 23 (SDC)	5
IAP, Nizhny Novgorod	35–38	TE_{21}/TE_{11}	34	7	15
MIT, Cambridge,	140	TE^{c}_{12}	2	2	9
LLNL, Livermore					
NRL, Washington, D.C.	27.8 29.2	TE^{r}_{10} TE^{r}_{10}	2 6	9 15	3 13
UNIV. Hsinchu	33.5	TE^{c}_{11}	20–67 100 164	6.5–21.7 25 41	5 8.5 1
UNIV. Strathclyde	8.5	TE_{21}/TE_{11}	60	16	17
UNIV. Utah	10	TE^{r}_{10}	0.72	10	8

r: rectangular waveguide; c: circular waveguide

wavelength ($f = 266$ MHz, $\lambda = 1.1$ m, $\lambda_w = 0.04$ m, $B_w = 0.04$ T). The highest CW power generated by a FEM is 36 W (15 GHz) whereas the IR (3.1 μm)-FEL at Thomas Jefferson National Accelerator Facility obtained a record average power of 2.13 kW at 3.5% efficiency employing beam energy recovery ([54], and references therein).

The FEM appears to be potentially capable of fulfilling all the requirements for a frequency tunable high-power mm-wave source. Coverage of the entire frequency range of 130–260 GHz presents no severe problem, and even higher frequencies are quite feasible. Rapid tunability over more than ±5% could be obtained by variation of the beam energy. The interaction occurs in a cavity operating in low-order modes, which have very good coupling to a Gaussian beam output. The relatively low RF wall loading and the use of high electron beam energy ($>$0.5 MeV) and a multi-stage depressed collector are compatible with a high unit power at high efficiencies if the electron beam interception could be maintained at an acceptable level. A survey of the development status of short pulse FEM oscillators and amplifiers is presented in Table 1.19 [54]. Electrostatic, pulseline/modulation and induction linac accelerators are employed.

The induction-linac based 140 GHz, 2 GW short-pulse (20–30 ns) LLNL-FEM was successfully employed in high-power microwave transmission measurements in fusion plasmas where the nonlinear absorption (higher transmission) at ECH frequencies was observed which is only possible at such high power levels. The UCSB-FEL was used to study nonlinear dynamics in semiconductor nanostructures and devices.

Table 1.19. Development status of pulse free-electron masers

Institution	Frequency (GHz)	Mode	Power (MW)	Efficiency (%)	Voltage (MV)	Current (kA)	Type
FOM,	206	HE^r_{11}	0.73(0.5)	5.7(3.9)	1.77	0.0072	oscillator
Nieuwegein	167	HE^r_{11}	0.36(0.26)	3.1(2.3)	1.61	0.0071	oscillator
	169	HE^r_{11}	0.1	0.9 (14 with MDC)	1.60	0.007	oscillator
IAP/INP Novosibirsk	75	TEM	100	4.2	0.8	3.0	oscillator
JINR	30.74	TM^c_{11}/TE^c_{11}	48	35(30)	0.8	0.17(0.02)	oscillator
Dubna/IAP	38.2	TM^c_{12}/TE^c_{11}	3	3(3)	0.8	0.15(0.2)	oscillator
Nizhny Novgorod	35	TE^c_{11}	30	10	1.5	0.2	amplifier
LLNL,	34.6	TE^r_{01}	1000	34(7.2)	3.5	0.85(4.0)	amplifier
Livermore	140	TE^c_{11}	2000 500–1000	13.3(10) in up to 50	6.0	2.5(3.0)	amplifier
MIT,	27.5	TE^c_{11}	1	10.3(6.3)	0.32	0.03(0.05)	oscillator
Cambridge	35.2	TE^c_{11}	0.8	8.6(5.2)	0.31	0.03(0.05)	amplifier
NRL, Washington, D.C.	13.2–16.6	TE^c_{11}	4.2	18	0.245	0.094	amplifier
RI, Moscow	6–25	TE^c_{11}/TM^c_{01}	10	1.7	0.6	1	spon. emiss.
TRW,	35	TE^r_{01}	0.1	9.2	0.3	0.004	oscillator
Redondo Beach	35	TE^r_{01}	0.002	6.9	0.29	0.0001	amplifier
UC Santa Barbara	120–880		0.027	0.5	2.6	0.002	oscillator

r: rectangular waveguide; c: circular waveguide

1.8 Future Trends and Applications

After the structural reorganization of most of the microwave tube manufacturers, as explained in Sect. 1.1.2, the market is still performing evolution and the needs of the customers change [1]. Certainly the three main applications listed in Sect. 1.1.1 will be always with us, but in different ways.

The use of TWTs in space and especially on communication satellites is important and still increases, probably because of the robustness and reliability of these tubes, long lifetime, high efficiency, and low weight of the whole associated transmitter. Also they are able to provide the large required power, 200 W/tube and several kW in total. As a matter of fact, on powerful transmitters, tubes are preferred compared to solid state devices, because the size and the cost of a tube do not at

all increase proportionally to its RF output power. On the contrary, this is also why the tubes are not seen in our everyday life, where the small power transistors are prevalent (mobile phones, computers).

The extraordinary success of microwave ovens using cheap magnetrons is also keeping with this presentation: magnetrons are robust, compact, reliable in a domestic and hostile industrial environment and capable of delivering approximately 800 W, in the S-band.

According to Fig. 1.1, we point out that the tubes become more attractive than the solid state devices when the operating frequency increases, what is the general tendency. Except some conventional ones, most of the modern radars use phased array antennas, working with a great number of low power modules and therefore solid state devices. However, in military aircrafts and especially missiles, tubes are still used (50 to 100 W, X-band, 2 to 3 octaves instantaneous bandwidths) again, because of their compactness and robustness with respect to vibrations, shocks and temperature variations.

As for the ISM applications, developments and improvements are foreseen in several directions: lower costs, fewer risks of arcing and parasitic oscillations, higher frequencies, better linearity, larger bandwidths, capabilities to deliver custom designed specific components, such as windows, waveguides and safety devices. Developments of new tubes may be extrapolated from the IOTs or the MBKs. In medical applications new low cost and compact MW magnetrons or klystrons will be requested. In the future there will be the possibility to provide a whole equipment including not only the microwave tube, the supplies and the cooling system, but also the electron linear accelerator. In that way the interface problems will disappear and the maintenance time, the overall cost, the number of failures will be reduced.

At higher frequencies, the impressive development of the present gyrotron oscillators is being followed by the one of gyroamplifiers, capable to be amplitude and frequency modulated, with the intention to be used not only on plasma fusion reactors but also for deep space communications, in some specific radars or in power transmission systems. Therefore, provided that theoretical developments, breakthroughs, and technological progress should come available on time, microwave tubes still have a confident future.

The R&D activity on microwave tubes, which requires a deep knowledge of many scientific areas and the mastery of many technologies, induces and encourages the tube physicists and technicians to be interested in the development of other electron vacuum devices, for instance:

- specific high efficiency multistage plasma (HEMP$^{@}$) thrusters, whose design is based on the TWT principle and, more precisely, on PPM focusing and depressed collectors;
- Cs atomic clock, where the RF cavity design and machining, the control of RF leakages, the screening of the magnetic fields, the quality of the vacuum and the reliability are familiar to microwave tube experts.

The presently rising micro- and nanotechnologies will also be important, especially regarding the combination of solid state and vacuum devices. For instance, 1 μm long needles (tips) with a diameter of 20 to 50 nm can be not only manufactured but also arranged in arrays of one or several mm^2, giving rise to very small cold cathodes. Associated to a grid for beam modulation at the cathode and to RF cavities, they can be the master parts of very compact microwave triodes, TWTs or IOTs, which can be used to boost the solid state amplifiers. The state-of-the-art of such vacuum microelectronic TWTs employing so-called Spindt cathodes is $P = 55\,\mathrm{W}$ at $f = 4.5\,\mathrm{GHz}$ (NEC Corporation: 50 000 tips giving 90 mA) and $P = 28\,\mathrm{W}$ at $f = 11.5\,\mathrm{GHz}$ (Northrop Grumman: 540 000 tips giving 56 mA).

In conclusion, microwave tubes are currently progressing and will progress in the future in the two main directions:

- the improvements of their basic principles and technologies;
- the evolution towards totally new devices using, at least partly, their concepts and these technologies.

Therefore, the future of microwave tubes is really encouraging and the microwave tube community is engaged in a long term program.

References

[1] R.H. Abrams, B. Levush, et al., Vacuum electronics in the 21st centuries, IEEE Microwave Mag. **2**, 61 (Sept.) (2001)
[2] M. Sedlacek, *Electron Physics of Vacuum Gaseous Devices* (Wiley, Hoboken, 1996)
[3] A.S. Gilmour, *Microwave Tubes* (Artech House, London, 1986)
[4] J.F. Gittins, *Power Travelling Wave Tube* (The English Universities Press, Ltd., 1964)
[5] A. Septier, *Focusing of Charged Particles*, vol. 1 & 2 (Academic, 1967)
[6] P.T. Kirstein, G.S. Kino, *Waters WE: Space Charge Flow* (McGraw-Hill, New York, 1967)
[7] S. Ramo, J.R. Whinnery, T. Van Duzer, *Fields and Waves in Communications Electronics* (Wiley, New York, 1984)
[8] R. Warnecke, P. Guenard, *Les Tubes Électroniques à Commande par Modulation de Vitesse* (Gauthier-Villars, 1951)
[9] J.W. Gewartowski, H.A. Watson, *Principles of Electron Tubes* (Van Nostrand, Princeton, 1965)
[10] A.H.W. Beck, *Space Charge Waves & Slow Electromagnetic Waves* (Pergamon Press, 1958)
[11] A. Leblond, *Les Tubes Hyperfréquences* (Masson, 1972)
[12] N.E. Lindenblad, U.S. Patent 2,300,052, filed May 4, 1940, issued October 27, 1942
[13] R. Kompfner, The traveling wave valve, Wirel. World **53**, 369 (Nov.) (1946)
[14] H.-T. Schmidt, Homepage. www.hts-homepage.de/WesternElectric/WesternElectric700.html
[15] http://accelconf.web.cem.ch/accelconf/p95/ARTICLES/RPA17.PDF
[16] J.R. Pierce, *Traveling Wave Tubes* (D. Van Nostrand, Princeton, 1950)
[17] H.J. Wolkstein, Effect of collector potential depression on the efficiency of traveling wave tubes, RCA Rev. **19**, 259 (June) (1958)

[18] W. Neugebauer, T.G. Mihran, A ten stage electrostatic depressed collector for improving klystron efficiency, IEEE Trans. Electron Devices **ED-19**(1) (Jan.) (1972)
[19] A.S. Gilmour Jr., *Principles of Traveling Wave Tubes* (Artech House, London, 1994 & 1999)
[20] J.R. Pierce, *Theory and Design of Electron Guns* (D. Van Nostrand, Princeton, 1954)
[21] I.L. Langmuir, K. Blodgett, Currents limited by space charge between concentric spheres, Phys. Rev. **24**, 53 (July) (1924)
[22] C.C. Cutler, The nature of power saturation in traveling wave tubes, Bell Sys. Tech. J. **35**, 841 (1956)
[23] J.E. Rowe, *Nonlinear Electron Wave Interaction Phenomena* (Academic, New York, 1965)
[24] G. Kornfeld, E. Bosch, From History to Future of Satellite TWT Amplifiers, Frequenz Zeitschrift für Telekommunikation, Band 55, 9-10/2001
[25] D.C. Rogers, Proc. IEE **100**(Pt. III), 151 (May) (1953)
[26] Tore Wessel-Berg, A General Theory of Klystrons with Arbitrary, Extended Interaction Fields, M.L. Report No. 376 (1957)
[27] M. Chodorow, B. Kulke, IEEE Trans. Electron Devices **ED-13**, 439 (1966)
[28] Y.M. Shin, G.S. Park, G.P. Scheitrum, G. Caryotakis, Circuit analysis of an extended interaction klystron, J. Korean Phys. Soc. **44**(5), 1239 (May) (2004)
[29] B. Steer, State of the art W-band EIK for Cloudsat, in *International Vacuum Electronics Conference, IVEC 2004*, Monterey, 2004
[30] B. Steer, Wide bandwidth, high average power eiks drive new radar concepts, in *International Vacuum Electronics Conference, IVEC 2000*, Monterey, 2000
[31] S. Millmann, A spatial harmonics amplifier for 6 mm wavelength, Proc. IRE **39**, 1035 (Sept.) (1951)
[32] R. Kompfner, The backward wave tube, U.S. Patent 2,985,790, filed 17 May 1952
[33] R. Kompfner, N.T. Williams, Backward-wave amplifier, U.S. Patent 2,916,657, filed 17 May 1952
[34] R. Kompfner, N.T. Williams, Backward-wave tubes, Proc. IRE **41**, 1602 (November) (1953)
[35] H. Heffner, Analysis of the backward-wave traveling-wave tube, Proc. IRE **42**, 930 (June) (1954)
[36] J.R. Pierce, R. Kompfner, Nat'l Academies Press, Biographical Memoirs, pp. 156–181 (1983). Use also: http://books.nap.edu/books/0309033918/html/156.html
[37] J. Arnaud, Backward wave tubes, in *The Encyclopedia of Electronics*, ed. by C. Susskind (Reinhold Publishing Corporation, New York, 1962)
[38] R.Q. Twiss, Radiation transfer and the possibility of negative absorption in radio astronomy, Aust. J. Phys. **11**, 424, 564 (1958)
[39] J. Schneider, Stimulated emission of radiation by relativistic electrons in a magnetic field, Phys. Rev. Lett. **2**, 504 (1959)
[40] A.V. Gaponov, Izv. VUZ Radiofiz. **2**, 450, 837 (1959)
[41] H. Kleinwächter, Zur Wanderfeldröhre, Elektrotechnik **4**, 245 (1950) (in German)
[42] J.L. Hirshfield, J.M. Wachtel, Electron cyclotron masses, Phys. Rev. Lett. **12**, 533 (1964)
[43] A.V. Gaponov, M.I. Petelin, V.K. Yulpatov, The induced radiation of excited classifical oscillators and its use in high frequency electronics, Izv. VUZ Radiofiz. **10**, 1414 (1967)
[44] V.L. Granatstein, I. Alexeff, et al., *High-power Microwave Sources* (Artech House, Boston, London, 1987)
[45] J. Benford, J. Swegle, *High-power Microwave Sources* (Artech House, Boston, London, 1992)

[46] C.J. Edgcombe (ed.), *Gyrotron Oscillators – their Principles and Practice* (Taylor & Francis, London, 1993)
[47] M.V. Kartikeyan, E. Borie, M.K.A. Thumm, *Gyrotrons – High Power Microwave and Millimeter Wave Technology* (Springer, Berlin, 2003)
[48] T.C. Marshall, *Free Electron Lasers* (MacMillan, New York, 1985)
[49] P. Sprangle, T. Coffey, New high power coherent radiation sources, Infrared Millim. Waves **13**, 19 (1985)
[50] H.P. Freund, T.M. Antonsen Jr., *Principles of Free-electron Lasers*, 2d edn. (Chapman & Hall, London, 1996)
[51] H.P. Freund, G.R. Neil, Free-electron lasers: Vacuum electronic generators of coherent radisation, Proc. IEEE **87**, 782 (1999)
[52] M.I. Petelin, One century of cyclotron radiation, IEEE Trans. Plasma Sci. **PS-27**, 294 (1999)
[53] A.V. Gaponov-Grekhov, V.L. Granatstein, *Application of High-Power microwaves* (Artech House, Boston, London, 1994)
[54] M. Thumm, State-of-the-art of high power gyro-devices and free electron masers, Update 2006, FZKA 7289 (2007)
[55] K.R. Chu, Overview of research on the gyrotron traveling-wave amplifier, IEEE Trans. Plasma Sci. **PS-30**, 903 (2002)
[56] G.G. Denisov, V.L. Bratman, A.D.R. Phelps, S.V. Samsonov, Gyro-TWT with a helical operating wavequide: new possibilities to enhance efficiency and frequency bandwidth, IEEE Trans. Plasma Sci. **PS-26**, 508 (1998)

2

Vacuum Displays

K. Blankenbach, G. Gassler, and H.W.P. Koops

The area of digital electronics and computers is not imaginable without displaying data. Vacuum based display technologies were the first being able to visualise measurements on electronic circuitry with oscilloscopes and to reproduce video on TV sets. Another strong market for electronic displays started in the mid 1980s with the invention of personal computers. Despite the advantages of flat Liquid Crystal Displays (LCD), Cathode Ray Tubes (CRT) still (2007) have a high market share. However, there are also vacuum based Flat Panel Displays (FPDs), like Plasma Panel Displays (PDP) and Vacuum Fluorescent Displays (VFD) that are widespread in different applications.

As the image quality of vacuum based displays (and others) have many common fundamentals or are based on similar principles, this chapter starts with an overview of display driving, common principles of vacuum based displays and parameters for measurement and evaluation of them.

There are many books and other references available about CRTs, but only a limited number for FPDs. Most of the publications on electronic displays are issued by the SOCIETY FOR INFORMATION DISPLAY [1] as conference proceedings, which can only be found rarely in standard libraries. Therefore, references are limited here to books or other common publications.

2.1 Introduction[1]

2.1.1 Historical Development

The story of electronic displays started in 1897 with the work of Ferdinand Braun, in Germany, by using an oscillograph to display transient phenomena [2]. Braun invented an evacuated tube (Fig. 2.1, left) with an electron source. The electrons were accelerated by a relative high voltage (anode) and deflected by magnetic fields. Light was then emitted at the location where they hit the phosphor layer on the screen. The

[1] Please refer to K. Blankenbach.

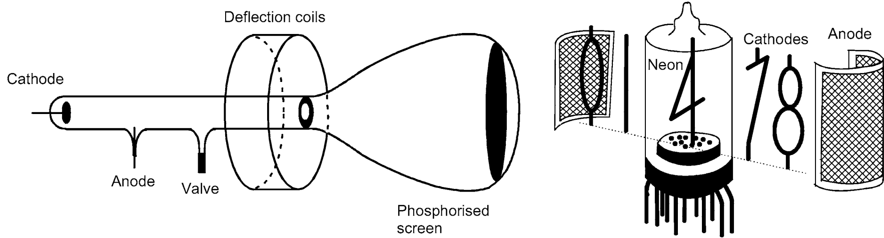


Fig. 2.1. Early displays: Braun's vacuum tube (*left*), NIXIE tube to display numbers (*right*)

name 'Cathode Ray Tube' (CRT) is not fully correct; it should be called an electron beam tube, though the latter name never became generally accepted. J. Zenneck improved Braun's tube in 1899 and adapted it for engineering measurements by using a linear sweep to display waveforms. Several improvements were made in the following years by electrostatic deflection and ZnS based phosphors for brighter screens.

However, the commercial history of oscilloscopes didn't start before 1930 when it became standard electronic laboratory equipment. As CRTs use phosphors, their data captures didn't persist long enough for examination. Several attempts to improve on this were made, including photography, but the history of analogue CRTs came to an end with the introduction of digital oscilloscopes and the replacement of the CRTs with LCDs.

Early television experiments started in the 1930s, showing it's images on CRT based systems; using video amplifying tubes and magnetic deflection signals. The story of CRT TV sets continues until today due to their competitive price. Around 1950 the first computers were built, starting with electron tubes as digital elements, which were replaced about 10 years later by semiconductor devices. Displaying the computers data was a challenge in those early years of computers. CRT based displays like TV sets couldn't be used because of the lack of fast data processing to generate continuous image streams. Consequently, initial computer displays were simple numerical indicators such as the "Numerical Indicator Experimental" NIXIE tube (Fig. 2.1, right), Numicator and Digitron (see, e.g. [3]).

A NIXIE tube consists of an evacuated glass tube containing multiple cathodes, a wired mesh anode filled with low-pressure neon. The most common form has ten cathodes in the shapes of the numerals 0 to 9, some are additionally equipped with one or two decimal points. Others could show various letters, signs and symbols. Each cathode was forced to glow (no heat is generated) by applying about 200 V DC between it and the anode. NIXIE tubes would have been ideal displays for clocks but were never used for this, as the electronics effort would have been enormous compared to a mechanical solution. The vacuum based numerical indicators were replaced in the 1970s by semiconductor light emitting diodes (LED) and Vacuum Fluorescent Displays (VFD).

Despite this, the successful story of vacuum based electronic displays continues until today with CRTs for high-resolution graphic computer terminals and TV sets. With the introduction of flat displays, like VFDs in the 1970s and Plasma Panel

Displays (PDP) in the 1990s, the relative low depth of these displays invented new products.

2.1.2 Overview

Displays are generally categorised [4, 5] as either direct view or projection devices (Fig. 2.2, left). Direct view devices are those where the observer looks directly at the image on the display, whereas projection systems make use of a screen. Emissive technologies such as vacuum based display and LEDs generate their own light. Non-emissive displays reflect or transmit ambient or artificial light, like Liquid Crystal Displays (LCD). All projection technologies, with the exception of Digital Mirror Devices (DMD, DLP) and laser displays, are based on modified direct view technologies.

The most common emissive displays are:

- Cathode ray tubes (CRT), mainly used in TV sets and computer monitors;
- Plasma display panels (PDP), as slim and large size CRT replacement;
- Vacuum fluorescent displays (VFD), for low resolutions and small screen sizes;
- Field emission displays (FED), as low power flat panel technology;
- Organic Light Emitting Diodes (OLED), as an emerging technology (2004).

The vacuum based technologies will be described in detail later in this chapter.

Typical, commercially available, screen sizes for major display technologies are shown in Fig. 2.2 (right). CRTs and LCDs can cover nearly the full range of screens sizes including projection. However, CRT based projection systems became obsolete due to their low level light output and size. For many applications, a shallow depth is a significant advantage. Therefore, Flat Panel Displays like PDPs and VFDs dominate many applications along side LCDs.

A rough comparison of CRT vs FED (Field Emission Display) is given in Fig. 2.3. A CRT has only one (monochrome) cathode for generating the electron beam. It

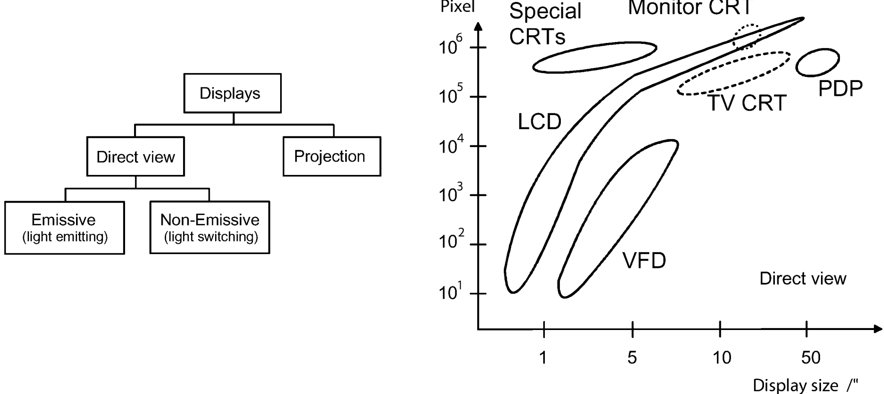


Fig. 2.2. Categorisation (*left*) of electronic displays; typical screen size and resolution for various display technologies (*right*), data from actual data sheets

is deflected by magnetic coils for TV and monitor applications and finally hits the phosphorised screen. The distance travelled to the screen varies with deflection, with greater deflections requiring greater depth. Conversely, FEDs have many cathodes per pixel hence deflection is unnecessary and subsequently they are shallower in depth.

A large screen size is, generally speaking, not suitable for all kind of devices. In many applications, like temperature meters and cashier systems, only numerical or alphanumeric representation is required, which also saves costs and space. Figure 2.4 shows how to display data depending on the basis of the image-generating element. As it can be seen, VFDs are used for low information content displays up to 40 characters. However, some low-resolution graphics are also available. CRTs, PDPs and FEDs require graphic processors to show high-resolution data (320 × 240 ... millions of pixels).

When specifying an electronic display the readability of the information is a major issue. For numbers and characters the distance between display and observer determines their height. The resolution ability of human vision is about 1′, with a tangent formula the resolvable object size can be calculated. For example, at 1 m one obtains a resolution of 0.3 mm. The height of the information element (digit, character) should be about 15 times the eye resolution – in this case around 5 mm.

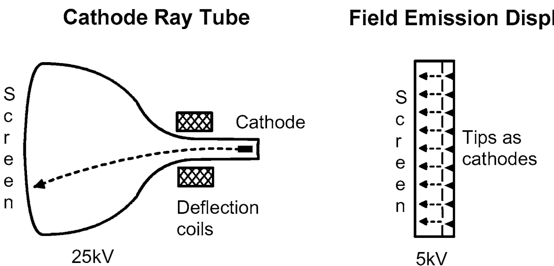


Fig. 2.3. Schematic comparison of a CRT and Flat Panel Displays (here Field Emission Display). The depth of a FPD is significantly smaller for the same screen size

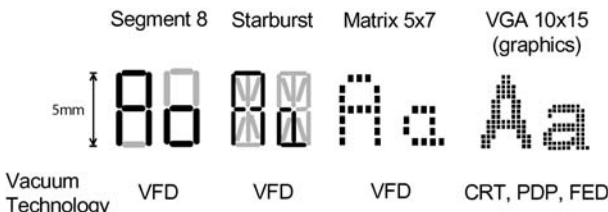


Fig. 2.4. Various representations in electronic displays for the characters ‘A’ and ‘a’ together with typical vacuum based display technologies. It can be easily seen that 8-segment (also named as Segment 8) digits are suitable for numbers and not recommended for displaying characters. A slightly better approach for cost sensitive devices is the Starburst configuration (15 segments). Sufficient readability is achieved by matrix displays. However, the best solution is a costly graphics representation. The height of 5 mm refers to an observer distance of 1 m

2.1.3 Fundamentals of Displays

Driving and Addressing

There are basically 3 methods of visualising data on electronic (vacuum based) displays (Fig. 2.5), e.g. [5, 6]:

- A sweep signal (left) drives the electron beam of analogue oscilloscopes, by electrostatic deflection, from left to right, and the measured signal modulates the vertical amplitude. No image can be displayed, only single curves; the beam intensity is constant during the sweep.
- Raster scan (centre) is used in CRTs for TV sets and computer monitors. Magnetic deflection coils move the electron beam (3 for colour CRTs) from left to right (horizontally) and from top to bottom (vertically) over the entire screen. The image is generated by intensity modulation of the electron beam representing the grey scale input signal, transferring the data serially (pixel-at-a-time addressing). During retrace (flyback) of each line or frame, the electron beam is blanked to avoid artefacts.
- Matrix drive (right) is the basis of all flat panel technologies. Each row is scanned sequentially, with the data being written in parallel to the actual addressed row

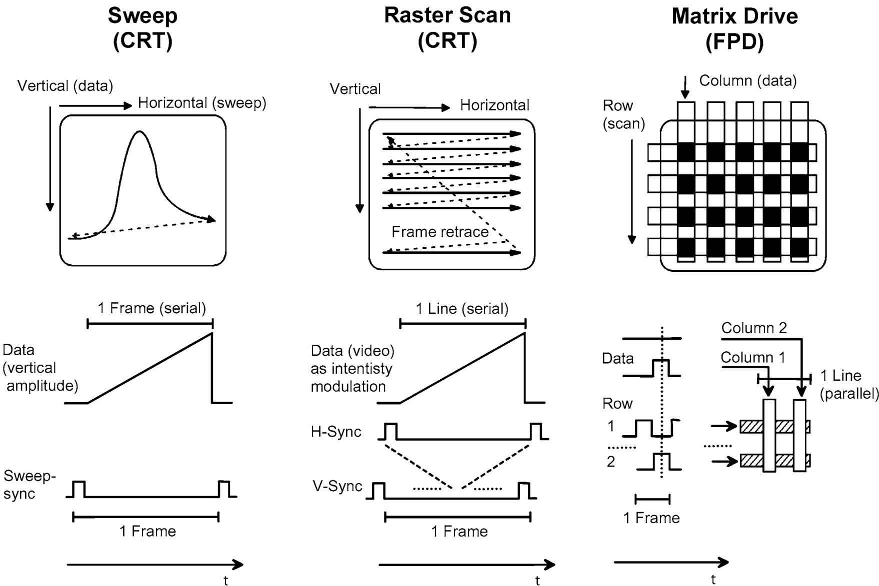


Fig. 2.5. Driving principles of CRTs (sweep and raster scan) and matrix drive for Flat Panel Displays. The drawings on the *top* show the simplified image generation, while in the *bottom* typical driving waveforms are plotted. The time for displaying a whole sweep or image is called frame time. The *dotted lines* represent the retrace (flyback) of the CRT electron beam. During this time the intensity (luminance) should be zero (blanking of video voltage)

(line-at-a-time addressing). The crossings of the two orthogonal stripe grids form the pixels. Low-resolution displays like 8-segment and Starburst VFDs have a similar driving scheme, which is called multiplexing, see also Fig. 2.38.

Sweep and raster scan are based on analogue signals for scanning. The matrix drive scheme makes use of digital data. Accordingly, some consequences can be stated:

- The frame rate and the resolution of sweep and raster scans are relatively easy to modify. Contrarily, the resolution (number of pixels) is fixed for matrix driven displays. Therefore, the resolution of the input signal has to be adapted by hardware or software to the displays resolution (see below).
- Comparing imaging principles, the time for writing the data to a dedicated pixel is a factor of the number of vertical lines, this is smaller for raster scan as opposed to matrix drive. This has some consequences for generating a certain luminance to avoid flicker (see below).

Due to limitations of the deflection, the depth of CRTs is much larger than that of matrix driven 'flat' panel displays.

Analogue and Digital Display Systems

Basically, the entire signal path of CRTs is analogue (Fig. 2.6, left); this means that all voltages and currents are amplified in the classical way. Matrix based display panels (all FPDs, Fig. 2.6, right) have a digital interface and require digital signal processing, the rows are scanned and the data are put in parallel to the columns.

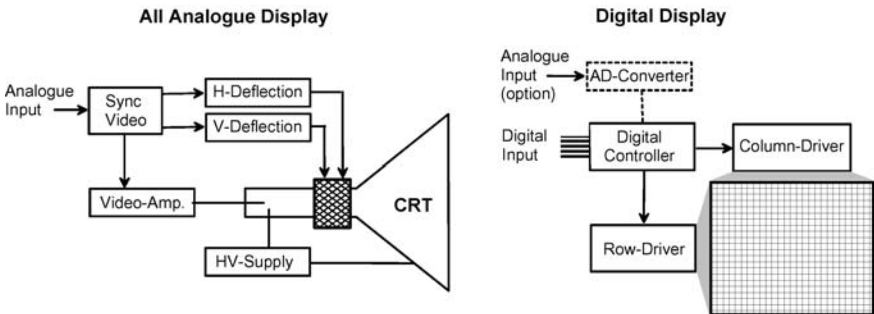


Fig. 2.6. Block diagram of a CRT based display (*left*): The input data (video) modulates the electron beams intensity by the video amplifier. The scan is done by deflection amplifiers, which are triggered by synchronisation (Sync) signals. However, oscilloscopes (sweep) have no video amplifier, the vertical deflection is driven by the input signal. Raster scan monitors use magnetic deflection because of its higher angular efficiency; electro-static deflection is applied in oscilloscopes because of their higher sweep rate and a simpler amplifier circuitry. The block diagram of a typical colour Flat Panel Display with various inputs is shown on the *right*. A digital controller provides the row scan timing and converts the data to display for the column driver. In PC based systems analogue inputs are used, some have an additional DVI input. Digital-RGB is only used for low to medium resolutions of Embedded Systems

Analogue digital display interfaces are described, e.g. in [5]. Basically all FPDs have a digital input with parallel data lines. To be compatible with analogue video signals, which were widespread until nowadays (2007), FPDs have an additional AD conversion and signal-processing unit. This compatibility is especially important for monitors and TV sets. A comprehensive review of signal processing, for displaying digital images, is provided in [7]. The advantage of all analogue displays compared to FPDs is that within certain limits various resolutions and frequencies can be displayed without costly hardware.

As previously shown (e.g. Fig. 2.2), electronic displays can have millions of pixels and, as mentioned before, they have to be addressed real time with typically 50–80 Hz. This requires high speed electronics, where the necessary pixel rate can be calculated by

$$\text{Pixel Rate} = \text{Pixel}_{\text{horizontal}} \times \text{Pixel}_{\text{vertical}} \times \text{Frame Frequency}. \tag{2.1}$$

Typical data are shown in Fig. 2.7. A review of the fundamentals of display driving and interfacing can be found in, e.g. [5]. To avoid flicker, the frame rate of CRT-based displays should be higher than that of matrix-based displays. The display driving however is limited for analogue signals by the video bandwidth (e.g. amplifier) and for flat panels by the transmission line capability. As mentioned above, the high data rate was the reason why the first computers had only numerical indicators.

As pointed out above, the resolution and signal frequencies can be easily changed for CRT monitors. The resolution of flat panels can't be changed, because the number of pixels is fixed! Therefore, the resolution of the input signal has to be scaled to the display's resolution via hardware and software. Furthermore, the frame rate of FPDs is usually 60 Hz compared to 75 Hz and higher for CRT PC monitors. For displaying interlaced video, e.g. PAL, on non-interlaced driven FPDs the two interlaced frames have to be merged to one image. All these tasks, which are described in [7], influence

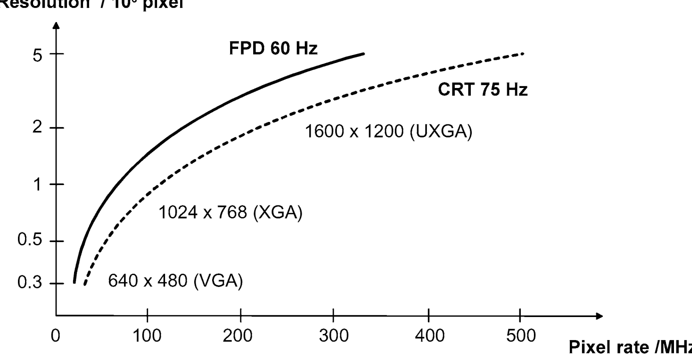


Fig. 2.7. Pixel rate and resolution for Flat Panel Displays and raster scanned CRTs. The frame rate for CRTs must be higher than that of FPDs because of flicker

Table 2.1. Tasks for displaying analogue signals on FPDs and their influence on image quality

Task	Name	Example	Image quality influence
Various resolutions	Scaling	SXGA → XGA	Reduced sharpness and uniformity
Interlaced video	De-Interlacing	interlaced → non-interlaced	Motion artefacts
Video and PC-Graphics	Frame Rate Conversion (FRC)	75 Hz → 60 Hz	Motion artefacts

the image quality. They are summarised in Table 2.1. Special processors are built in analogue input panels for displaying various resolutions and signals.

Light Generation

All vacuum based display technologies use phosphor (mostly ZnS based) to generate light. The activation is achieved for CRTs, VFDs and FEDs by an electron beam, in PDPs UV radiation is converted into the visible range. The two basic approaches for transferring the light from the activated phosphor to the observer are shown in Fig. 2.8. The light, which is produced (more or less) on the surface of the phosphor layer by the electron beam of CRT and FED based displays, has to pass through it's own (phosphor) layer and is consequently reduced in its intensity. PDPs and FEDs show typically a significantly higher luminance because the light is generated on the side of the observer. As stimulated phosphors emit light nearly in all directions (Lambertian emitter), the viewing angle is not an issue compared to LCDs.

The properties of some typical phosphors are summarised in Table 2.2, the colour co-ordinates (CIE) are described in the next chapter. A closer look on phosphors and their properties can be found in other chapters of this book. The decay time is relevant for flicker, a phenomenon, which occurs due to the pixel-at-a-time addressing of CRTs. Long persistence phosphors like amber were used in early low frame frequency monitors (slow graphics adapter) and in vector driven radar CRT screens. PDP phosphors must have a very low decay time because of the grey scale generation by subframes (see chapter about PDPs).

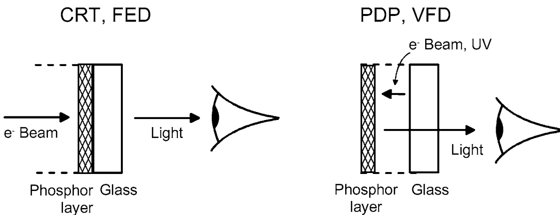


Fig. 2.8. The phosphor layer is activated by electrons on the opposite side of the observer for CRT and FED (*left*) and stimulated to the observer's side for PDP and VFD (*right*)

Table 2.2. Phosphors for vacuum based display technologies, Colour Coordinates CIE 1931, see Fig. 2.12

Application	EIA No. or type	CIE 1931		Decay range
		x	y	
CRT B/W TV	P-4	0.270	0.300	ms
CRT TV Red	P-22 R	0.640	0.340	ms
CRT TV Green	P-22 G	0.285	0.600	ms
CRT TV Blue	P-22 B	0.150	0.172	ms
VFD Bluish-Green	ZnO:Zn	0.250	0.440	ms
PDP Red YGB	$(Y,Gd)BO_3$:Eu	0.640	0.360	µs
PDP Green ZSM	Zn_2SiO_4:Mn	0.230	0.710	µs
PDP Blue BAM	$BaMgAl_2O_3$:Eu	0.140	0.050	µs

2.1.4 Fundamentals of Display Parameters

To compare and to evaluate displays, measurement parameters and procedures are necessary. The standards ISO 9241 and ISO 13406 are mainly used for determining the display quality, a summary is given in [8]. The most relevant parameters are described here in brief. Please note that all of these depend on various conditions like warm up, lifetime, temperature, vibration and ambient light.

Luminance

One of the most basic parameters of a display is its luminance, measured in units of cd/m^2. Many tasks can be done with this single parameter, like maximum luminance and uniformity of the screen. Typical application requirements are about 100 cd/m^2 for office, 300 cd/m^2 for multimedia and simulations. Avionics, military applications and presentations are in the range of 5000 cd/m^2, which is relatively small compared with 10^5 cd/m^2 of incandescent bulbs. Measuring the luminance allows calculations like contrast ratio and grey scale

Contrast Ratio

If the luminance of a bright (e.g. white) and a dark (e.g. black) pixel or area is measured and divided according to the formula,

$$C_R = \frac{L_{\text{bright}}}{L_{\text{dark}}} \underbrace{=}_{\substack{\text{with}\\\text{ambient}\\\text{light}}} \frac{L_{\text{bright}} + L_{\text{reflected}}}{L_{\text{dark}} + L_{\text{reflected}}} \tag{2.2}$$

the contrast ratio C_R is obtained. Typical values for CRTs are about 100:1, with PDPs being higher. High contrast ratios are usually regarded as quality parameter. Unfortunately, the recommended value for eyestrain-free vision is about of 10:1 the

value for paper. High contrast ratios cause dazzling in the same way as headlights of an approaching car at night do! However, when ambient light is reflected on the screen, its luminance must be added; therefore, the contrast ratio is lowered. Nearly all contrast ratios given in specifications are measured without ambient light. If the ambient light is too bright, the contrast ratio could become so low that the image can't be recognised.

Electro-Optic Curve

For displaying pictures and enhanced graphics, grey shades between dark and bright are necessary, this is called the grey scale. Due to the characteristics of the human vision, a specific non-linear relationship (electro-optic curve) between data and the luminance of the display is necessary (Fig. 2.9). The perceived brightness B (see (2.3)) shows a non-linear relationship to the luminance L of an electronic display, as the human eye is a quasi-logarithmic receiver. The parameter n lies in the region $0.3 \ldots 0.4$ (Weber–Fechner). The relationship can be also expressed by $B \sim \log(0.41L + 1)$. Both relationships have similar shapes. This is described more in detail in, e.g. [10], while the measurement procedures can be found in [8].

The perceived brightness B and data D (grey level, centre) should be proportional in computer and video systems with a display. The luminance L of a display depends on the grey scale value D (input). To compensate for the left equitation, corrections via hardware or software have to be applied to the electro-optical relationship of the display (right). The so-called gamma value (γ) has therefore to be set to about 2.5 ($n \cdot \gamma = 1$). Note that in the WINDOWS™ based PCs and TV sets a gamma value of about 2.3 is used, for APPLE MAC™ systems a value of 1.8 is chosen.

$$B \sim L^n; \quad B \sim D; \quad L \sim D^\gamma. \tag{2.3}$$

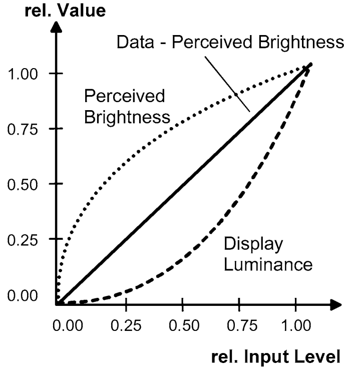


Fig. 2.9. Human vision exhibits a non-linear relationship of the perceived brightness (····) to the luminance (rel. input level). To obtain a linear relationship (—) of input signal (data, voltage) to the perception, the luminance output of the display (- - -) must compensate the perceived brightness

Due to historic reasons (few cameras – many receivers) the luminance–brightness relationship is adapted to CRTs. It seemed reasonable to have the correction circuitry in the (few) cameras and not in the large number of CRT-receivers. Fortunately, CRTs exhibit, without taking further measures, gamma values in the range 2 . . . 3, with 2.5 being typical.

Grey Scale Generation

The non-linear relationship between grey scale data D and the display's luminance L can be done in two ways (see Table 2.3): by an analogue value (e.g. voltage for most of the electronic displays) which determines the grey scale or by dividing each frame into subframes (8–12) of different length according to their specific value. The latter is applicable for fastest switching displays like PDPs (Plasma Display Panels) and DMD (Digital Mirror Devices), where the signal path from processor to the light output is strictly digital. However, this requires the shortest response time in the range of a few microseconds. The analogue and digital grey scale generation is shown in Fig. 2.10.

Colour Generation

Modern multimedia applications require the capability of displaying colour on electronic displays. For many systems a sufficient readability with monochrome (e.g. clocks) or area colour (e.g. meters with warning icons) displays can be achieved. The latter means that a specific colour can only displayed in a certain area; full colour displays show all colours independently of their location on the screen. A summary of vacuum based displays and their 'commercial' colour capability is shown in Table 2.4. Only CRTs and PDPs can be used for true colour displays. A good review of colour and displays can be found in, e.g. [8–10].

Colour in full colour displays is generated by additive colour mixture via superimposing the primary red (R), green (G) and blue (B) subpixels (Fig. 2.11). There are several possible arrangements for the subpixels, where the dot type delivers the most impressive images. If the pixel pitch and the viewing distance are geometrically in an appropriate relationship (viewing cone below 5 minutes of an arc, e.g. 0.3 mm pixel pitch at 50 cm viewing distance) the subpixels merge to one colour impression. For example, white is recognized when the three primaries are activated and the luminance of each primary has the appropriate value.

Table 2.3. Signal processing and grey scale generation for various displays; A: analogue, D: digital

Display	Signal processing	Electro-optic driving
CRT	A	A
VFD, FED, LCD, . . .	D	A
PDP, DMD, . . .	D	D

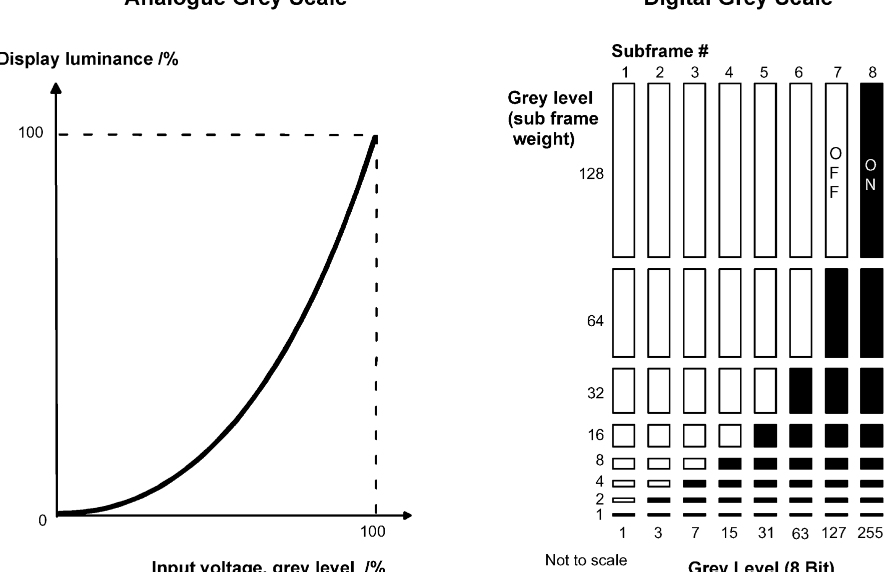


Fig. 2.10. Electro-optic curve for analogue grey scale displays (*left*), e.g. CRTs, and grey scale by subframes for all digital displays (*right*) like plasma displays

Table 2.4. Colour capabilities in terms of commercial availability (2007) for different vacuum based displays

Technology	Monochrome	Area colour	Colour	Pixel pitch/mm
CRT	X		X	0.3
PDP			X	1.0
VFD	X	X		0.5
FED	(X)			0.3

Colour Systems

Because human vision is, like all other senses, not suitable for reproducible measurements, one needs appropriate methods for evaluating and comparing displays. For colour, the first widely used system was the CIE 1931 (CIE xxxx: Commission Internationale d'Éclairage, xxxx is the year of standardization). Despite some insufficiencies it is still the current standard in 2007.

The CIE 1931 colour space is shown in Fig. 2.12 (left). All possible colours are within the boundaries of a horseshoe like curve from 380 to 770 nm, the 'purple line' connects the ends of the colour curve. In the upper area, the shades of green lie, while red is in the lower right corner and blue in the lower left one. Colours from continuous spectra, like lamps, lie within the curve, while single sharp peak spectra, like those from lasers, are on the curve. 'White' (normally 0.33/0.33) has no colour. The coordinates of the three display primaries (for typical values see Table 2.2) are

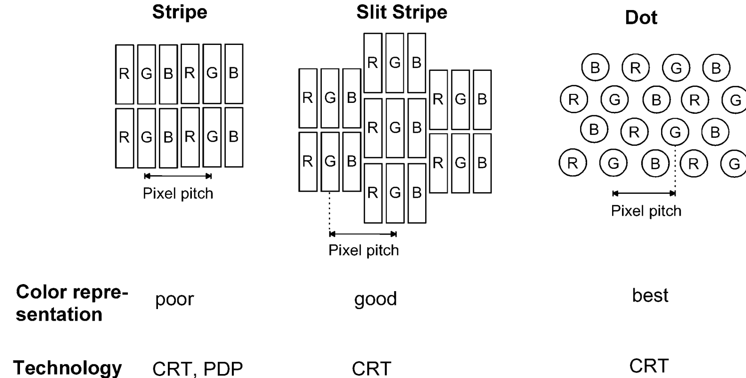


Fig. 2.11. Typical (sub) pixel arrangements for vacuum based displays. The best color mix is achieved for the dot type; however, this is only true for images and not for text. Stripe orders result in a higher luminance because the active area is higher compared to dot displays. Due to driving purposes of the matrix, all commercial flat panel displays use the regular RGB stripe, as shown on the *left*. The pixel pitch equals the sequential spatial colour period. Typical pixel pitches are 0.25–0.3 mm for CRTs and about 1 mm for PDPs which are therefore not recommended for desktop monitors with low viewing distance

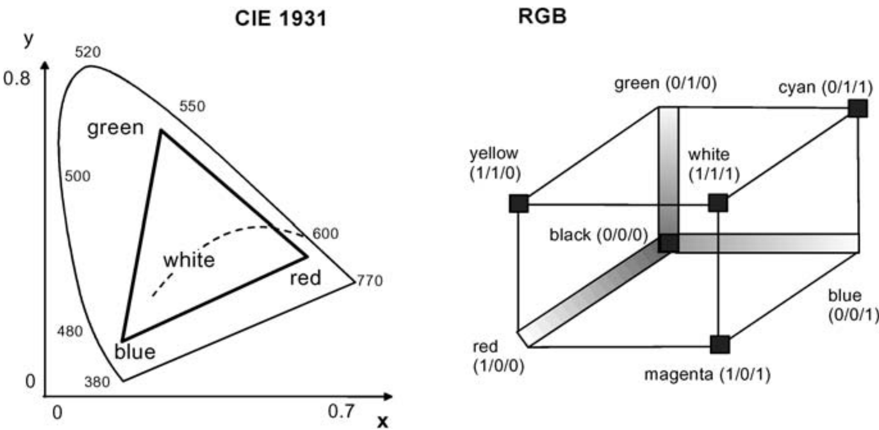


Fig. 2.12. CIE 1931 (*left*) and RGB (*right*) color system: The numbers on the horseshoe like curve equal single wavelength emitters like lasers. The black body radiation is represented by the *dotted line* and corresponds to a color temperature. For instance, white is usually defined between 5000 and 9300 K. The luminance as third axis (colour space!) is not plotted here. The RGB colour space forms a cube, it has basically no relationship to display parameters like luminance and is therefore device dependant

standardised (e.g. NTSC) and normally connected with lines to form a triangle. Only colours within this triangle can be displayed using specific grey scale values for the primaries. The area of this triangle refers to the so-called (colour) gamut to describe a potential lack of colour capability by the percentage of the standardization value.

The colour coordinates of the primaries (and those of all displayed colours) however are specified for dark room conditions. Therefore, any illumination is added to the Tristimulus values and shifts the original display colour co-ordinate to that of the light source. In case of a white illumination, we speak about ‘bleaching’ of colours because they become lightened, the gamut decreases.

From the computer programmer’s point of view, software is usually based on the RGB colour system (Fig. 2.12, right). It consists also of the three primaries RGB, but their colour co-ordinate is basically undefined. The primaries span a box whose axis is formed by the grey scale. The co-ordinates here have normalized values; therefore, white is achieved when red, green and blue are fully on. The display itself transfers, via electronics, the electro-optical curve and the properties of the phosphors to visible light; this is described in detail in, e.g. [9]. Consequently, unless precautions are taken, electronic displays are not interchangeable if colour is to be maintained. This is a task of standardization and colour management.

Image Quality

The image quality of displays is influenced by several effects (see, e.g. [8]). Some of them are described here.

- Analogue signals are limited by the bandwidth of the analogue signal path due to low pass filter characteristics. For Flat Panels the AD converter is also relevant as it affects the sharpness of the image, mainly for characters and high pixel frequencies. The effect is measured as Modulation Transfer Function (MTF). A rule of thumb formula is that the minimum analogue bandwidth should be greater than 2×horizontal resolution × horizontal frequency. For example, the video bandwidth of an XGA monitor should exceed 135 MHz.
- Deviations from the gamma curve allow different grey shades to merge together. As a consequence, a reduced number of grey shades and colours can be recognised on the display.
- Electro-magnetic interference (EMI) increases due to the sensitivity of signals to high frequency radiation, power supply artefacts and ghosting of neighbouring pixels.
- All phosphor based displays show a degradation of the emitted light related to the duration and the luminance of the corresponding pixel emissions. This limits the lifetime (e.g. time to half luminance, CRT about 10 000 h, PDP about 10 000 h); for long term fixed display content, an image will ‘burn in’. This means that the image remains ‘visible’ even if another image is shown. Another corresponding effect occurs for colour displays (RGB phosphors) if the lifetimes of the primaries are different. In most cases blue has the shortest one; therefore, the white point of the display is shifted to blue with increasing operation time.
- Effects of scaling and displaying video on Flat Panel Displays.
- The Moiré effect as colour noise occurs for colour displays if the image raster frequency and pixel pitch are nearly identical.

- Ambient light is reflected on the screen and superimposes the light that is produced by the emissive display itself. The reflected luminance affects both dark and bright pixels so that the contrast ratio is decreased. This leads, together with grey scale degradation and colour shift (to white for a white illumination resulting in a reduction of gamut due to 'bleaching' of colour), to a significant degradation of readability. Most of the emissive displays can't be read in bright sunlight. There are various methods to reduce the intensity of reflected light by louvers, anti-reflection and anti-glare methods.

2.1.5 Comparison of Various Display Technologies

When designing a new electronic device with a display, it is necessary to know which technology is the most suitable; requirements and characteristics of various applications can be found, e.g. in [11, 12]. This depends on the data to be displayed and the viewing distance. Both parameters result in an optimum screen size, pixel pitch and required resolution. Ambient light dictates the value for the luminance of the display and the need for a low reflection coefficient (in bright conditions). Power consumption, which could be expressed as efficiency, and the display depth is an issue for mobile devices. When displaying video, colour capability is required. The response time for video applications or other fast changing content should be in the region of the frame time, which is typically 16.7 ms for 60 Hz frame rate. Many industrial and public information applications operate 12 or more hours per day, so the lifetime of the display should be high (one year of continuous operation equals 8760 h). The price is compared on a relative amount per inch screen size.

Table 2.5 gives a rough review for typical features of major display technologies. CRT is the display of choice for medium size and lowest cost, while PDPs target applications with large screen size and low depth. VFDs are mainly used as low information content displays. FEDs are still under development. The situation in 2007 is as follows: Liquid Crystal Displays (LCD) are currently the most universal display technology for nearly all applications from 1″ to 50″ screen size. Standard TV sets are the dominating application for CRT because of it's low price, while larger screen sizes are the favourite application of PDPs. Turnover, production and forecast of various display technologies and applications can be obtained via [1].

2.2 Cathode Ray Tubes (CRT)[2]

There are only a very few inventions in the modern world which changed the way of life so dramatically like the Television did. Almost no other means opened the eyes of so many million people to the world. The Cathode Ray Tube as the key component is still the magic focus point of countless people every day and night.

[2] Please refer to G. Gassler.

2.2.1 Historical Development

If in the year 1897 Ferdinand Braun, the inventor of the Cathode Ray Tube, would have asked a banker for any amount of venture capital to finance the development of the "future billion pieces display", very most likely he would have exposed himself to ridicule. Hundred years later still more than hundred million CRT's are manufactured every year. Despite the upcoming of flat panels in the beginning of twenty first century with reasonable quantities, the vast majority of all displays used worldwide are still using the same basic principle of Mr. Braun's invention. Countless publications on all kind of CRT details were made; [13] gives an excellent review of CRT technologies with a lot of further references.

Table 2.5. Comparison of Vacuum Displays and Active Matrix LCDs on the basis of commercial available types with large screen size (direct view) in 2007

Feature	CRT for TV set	CRT for PC monitors	PDP	VFD	FED[a]	AM LCD[b] (PC and TV)
Screen size /″	20–30	17–19	42–70	1–5	5	15–50
Depth /cm	40–60	40	10	1	2	1–10
Resolution	PAL	SXGA	HDTV	–	QVGA	HDTV
Pixel /$\times 10^5$	4	13	4	<0.1	0.8	13
Pixel pitch /mm	1	0.3	1	0.5	0.5	0.3
Luminance /cd/m^2	100	200	500	1000	100	300
Efficiency /lm/W	2	0.5	1	1	1	2
Colour capability	++	++	+	–	o	++
Response time /ms[c]	20	10	17	10	17	25
Lifetime /h	10 000	10 000	10 000	10 000	1000	50 000
Price / €/″	3	5	20	20	–	5–20

[a] FEDs only prototypes
[b] LCD for comparison
[c] 20 ms for 50 Hz, 17 ms for 60 Hz frame rate

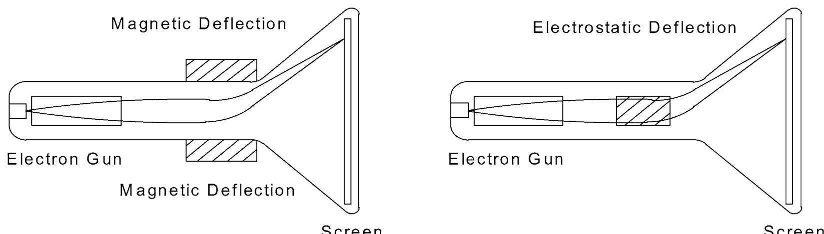


Fig. 2.13. Basic principles of magnetic and electrostatic deflected Cathode Ray Tube

2.2.2 Electrophysical Fundamentals

The basic principle of generating an image with a Cathode Ray Tube is illustrated in Fig. 2.13.

Inside an evacuated glass envelope an accelerated and deflected electron beam is focused on a screen, which converts the invisible electrons to a visible image. The deflection itself can be based either on electrostatic or on magnetic forces to the electrons.

In order to protect the electron source and the electron beam from being exposed to air, the glass envelope has to be evacuated. Consequently, bigger sized evacuated glass bulb needs mechanical protection against implosion.

Electron Beam

The vast majority of all today cathode ray tubes in TV and in professional applications are using Oxide-cathodes as a source for generating an electron beam. In cases where very long lifetime (up to 100 000 h), together with higher beam currents (2 A/cm^2), is required also more costly dispenser type cathodes can be found.

The mean thermal energy of an electron at room temperature is around 0.025 eV. To emit an electron, it is therefore necessary to heat up the emitting surface to a temperature between 600–2000°C to exceed the material dependent work function. Alkaline earth oxides like BaO with a work function between 1.2 and 1.8 eV can reach a current density of several hundred mA/cm^2 at temperatures around 700°C. The basic setup of an oxide cathode is illustrated in Fig. 2.14.

During manufacturing a layer of Barium carbonate $BaCO_3$ is sprayed on top of the cathode sleeve. This layer gets converted to BaO in a transforming process at the end of the exhausting of the cathode ray tube. The emerging CO_2 is pumped out at the same time. Thus,

$$BaCO_3 \rightarrow BaO + CO_2. \tag{2.4}$$

In order to reduce the work function furthermore, the top layer of BaO close to the surface is reduced to metallic Ba in a final activation step. The resulting oxygen and also all kinds of residual gases inside the glass envelope are absorbed by efficient getter materials.

The task of the electron gun (see, e.g. [13], p. 287ff) is to form a well defined beam of electrons, which can be modulated in its intensity and has its focal point

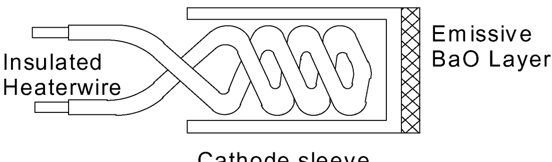


Fig. 2.14. Setup of an Oxide Cathode

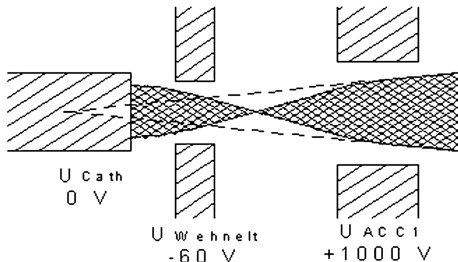


Fig. 2.15. Electron-optical setup of the "Beam-Forming-Region" of an electron gun with exemplary voltage

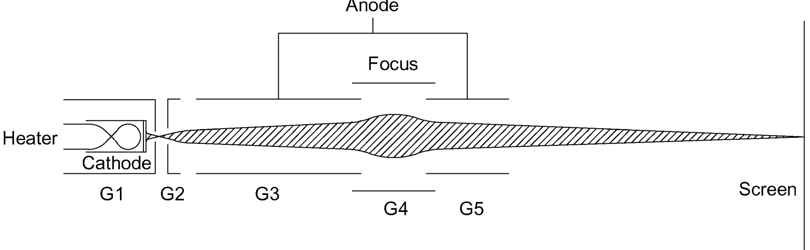


Fig. 2.16. Design of an Einzel Lens Electron Gun

on a light emitting screen (see Fig. 2.15). Electrons coming from the cathode are accelerated and focused to the so-called Crossover by an appropriate electrical field between the cathode, Wehnelt Cylinder and the first accelerating electrode. By altering the voltage at the Wehnelt Cylinder, the flux of electrons through the small aperture hole can be adjusted and therefore the resulting brightness on the screen.

The electron beam can be focused to the screen by a set of electromagnetic lenses, see [14, 15]. The focusing characteristics are depending on the speed of electrons caused by the accelerating voltages and the potential differences between neighbouring electrodes. Typical anode voltages are in the range from several kilovolts up to 40 kilovolts in special applications.

For the focusing of the electron beam basically two different types of configurations can be used, the Einzel Lens or the Bipotential Lens. An exceptional setup is formed by the Magnetic-Focus Lens. Typically the Einzel Lens design uses a low focus voltage at around 0 to 400 V and can be found in oscilloscopes, low and medium resolution monitor tubes, see Fig. 2.16.

Better resolutions at shorter overall dimensions can be achieved with the concept of a Bipotential Electron Gun design. The focus voltage varies with the specific design between 25–30% of the final anode voltage. The Bipotential Electron Gun is the standard design for Colour TV tubes and high resolution monitors, for instance for medical imaging, see Fig. 2.17. Special optimized concepts for professional applications are achieved by intensive use of computer based electron optics simulation programs.

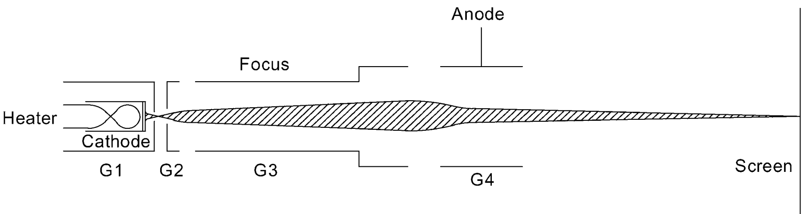


Fig. 2.17. Design of a Bipotential-Lens Electron Gun

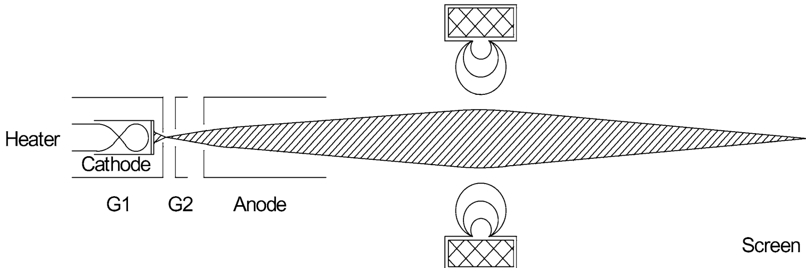


Fig. 2.18. Design of a Magnetic-Focus-Lens Electron Gun

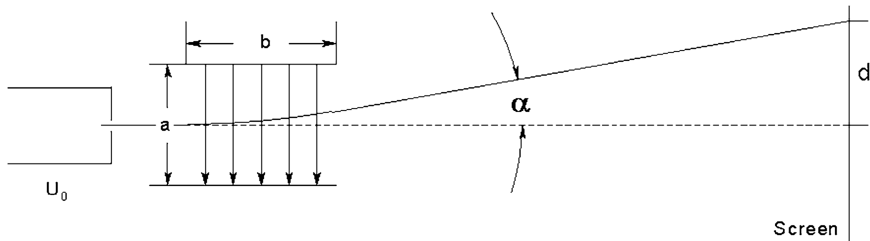


Fig. 2.19. Electrostatic deflection of an electron beam

Additional improvements up to very high resolutions are possible by using a magnetic focus lens concept. The extremely smooth characteristics of such a lens results in very low disturbances of the electron flux and therefore an optimum spot on the screen. A comparable big overall length of this tube design reduces its application to cases where resolution is of key importance. Examples are reprography systems for the film industry and high performance projection systems for flight simulators, see Fig. 2.18.

Electrostatic Deflection of Electron Beam

The principle of electrostatic deflection (see, e.g. [13], p. 200ff) is mainly used to display all kinds of electrical signals on oscilloscopes. The basic set up is shown in Fig. 2.19.

An electron beam with the initial velocity

$$v_0 = \sqrt{\frac{2eU_0}{m}} \tag{2.5}$$

is drifting through two parallel deflection plates and gets deflected by the electrical field

$$E = -\frac{U_\mathrm{p}}{d}. \tag{2.6}$$

The potential U_p is proportional to the interesting signal, which should be displayed. The deflection d or the deflection angle of the beam is given by

$$d = \frac{L \cdot b}{2a} \cdot \frac{U_\mathrm{p}}{U_0}, \tag{2.7}$$

where L is the distance of deflection plates centre to the screen, b is the length of deflection plates and a is the distance between deflection plates.

The deflection coefficient

$$k = \frac{U_\mathrm{p}}{d} \tag{2.8}$$

of an oscilloscope tube is one of the key figures. The development of fast A/D converters and low cost data storage systems brought the end to the premium class of highly sophisticated broadband and storage oscilloscopes. Except for some very specific applications, most of these technical highlights in the development of CRTs vanished in the meanwhile from the labs and are sometimes maintained only by some respectful senior technicians.

Magnetic Deflection of Electron Beam

When exposed to a magnetic field, a free flying electron beam will be deflected by the Lorentzian Force. The electron beam faces this force perpendicular to its direction of flight and perpendicular to the direction of the magnetic field. It is evident that the velocity of the electrons remains unchanged during magnetic deflection. The following calculations are described more in detail in [13], p. 233ff.

$$\vec{F} = e[\vec{v} \times \vec{B}]. \tag{2.9}$$

The deflection angle Φ arising from a homogeneous magnetic field $\boldsymbol{B}$ of length L (Fig. 2.20) is given by

$$\sin \Phi = \frac{e_0 \cdot L \cdot \boldsymbol{B}}{mv}. \tag{2.10}$$

Using the expression for the velocity of the electron beam

$$v = \sqrt{\frac{2 \cdot e_0 \cdot U_0}{m}}, \tag{2.11}$$

the resulting deflection angle will be

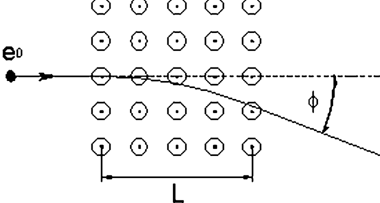


Fig. 2.20. Deflection of an electron beam in a perpendicular magnetic field

$$\sin \Phi = \sqrt{\frac{e_0 \cdot L \cdot \boldsymbol{B}}{2m \cdot \sqrt{U_0}}}, \tag{2.12}$$

where e_0 is the charge of the electron, m is the mass of the electron, v_0 is the velocity of the electron and U_0 is the accelerating voltage. This means that the deflection angle is proportional to the strength and the length of the magnetic field and inverse proportional to the square root of the accelerating voltage.

The failures created by a real magnetic deflection system can be split into three categories. Aside from the deflection angle Φ, the divergence angle α of the electron beam inside the deflection area is of importance. The failure categories are:

$$\text{coma failures} \sim \alpha^2 \Phi,$$
$$\text{astigmatism} \sim \alpha \Phi^2,$$
$$\text{distortion} \sim \Phi^3.$$

For practical use the magnetic fields are generated in a deflection coil, which is placed on the back of the tube. In order to increase the magnetic field strength, the copper wires are winded around highly efficient magnetic ferrites.

High resolution monochrome monitor tubes for medical applications are able to display images of 5 million pixels on a 21″ screen with extremely high contrast ratio. As monochrome Cathode Ray Tubes are not limited by a pixelated screen, their limitation is mostly given by the electron optical design of the electron gun, [15].

The magnification of a failure free image is described by Abbe's law (Fig. 2.21)

$$V = \frac{b}{a} \cdot \sqrt{\frac{U_K}{U_S}} = \frac{\sin \alpha_K}{\sin \alpha} \cdot \sqrt{\frac{U_K}{U_S}} \approx \frac{\alpha_K}{\alpha} \sqrt{\frac{1}{U_S}}. \tag{2.13}$$

The achievable resolution of an electron gun is influenced basically by 3 effects:

- cathode temperature,
- spherical aberration of the lenses,
- space charge repulsion.

Due to the thermal velocity distribution of the electrons coming from the cathode, the crossover will always have a certain size. Therefore, the image on the screen will show a Gaussian intensity profile, even after passing failure free lenses.

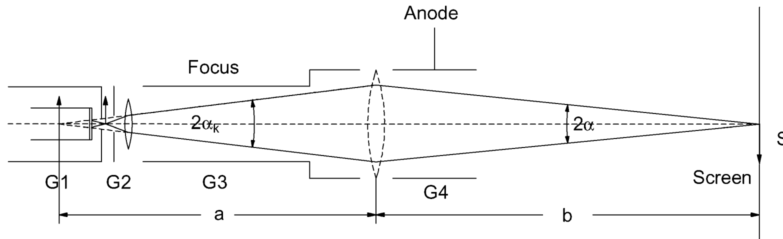


Fig. 2.21. Abbe's law; a is the object distance, α is the aperture at screen, b is the image distance, U_{K} is the potential at crossover, α_{K} is the aperture at crossover, and U_{S} is the potential at screen

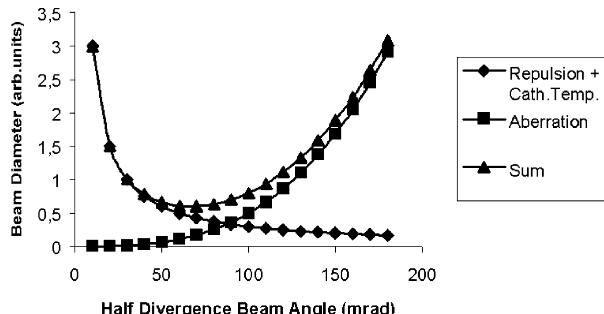


Fig. 2.22. Influence of broadening effects on the spot size

Using Abbe's law, one can determine the spot size at a given cathode temperature

$$d_{\mathrm{T}} = \frac{4r_0}{\alpha} \cdot \sqrt{\frac{3k_{\mathrm{B}}T}{eU}}, \tag{2.14}$$

where r_0 is the radius of the emitting surface on the cathode, k_{B} is the Boltzmann constant and T is the cathode temperature. This equation shows that the spot size depends on the radius of the emitting cathode surface and the beam angle. At higher brightness and therefore higher cathode current the spot size will grow consequently. The second factor for the spot size is coming from the mutual repulsion of the electrons inside the beam, and it is inverse proportional to the beam angle α.

Strictly following the analogous optical lenses also an electron optic lens has its spherical aberration failures. These failures increase with the power of three dramatically for larger beam apertures of α. All the three influences can be summed up as

$$d = \left[\left(\frac{C_1}{\alpha}\right)^2 + (C_2\alpha^3)^2\right]^{0.5}. \tag{2.15}$$

In Fig. 2.22, the graphs of the contributions to the spot size are plotted. The design of an electron gun means in practice to find an optimum solution between the different influences on the spot size. A more detailed analysis shows that the biggest efforts have to be spent first to a small beam angle α. This can be achieved by a so-called

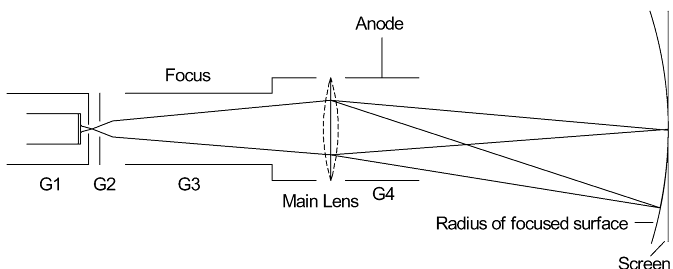


Fig. 2.23. Difference of focus surface and screen

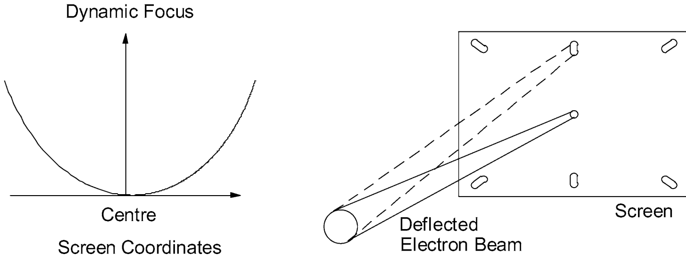


Fig. 2.24. Dynamic focus over the screen (*left*) and geometric distortion of the spot (*right*) caused by geometric projection

pre-focusing lens within the Beam Forming Region (BFR). High resolution CRTs are realized with very sophisticated, low tolerance designs of the BFR.

From Fig. 2.23 it is easily seen that a fixed focus electron gun can be brought to best focus only at a certain point. When deflected, this point will grow to a ring on a sphere, depending on the rate of deflection. The inner side of the glass faceplate will always have a different radius compared to the distance between the main lens and the screen, therefore dynamic focusing is required. This is realized by adding a dynamic parabolic term to the static focus, depending on the location of the deflected spot.

For very high resolution monitors also a dynamic astigmatism control (Fig. 2.24) can be added which corrects for distorted spot sizes due to the geometric projection of a deflected spot, see [13], p. 295ff.

2.2.3 Present State-of-the-Art

The conversion of an invisible electron beam to a viewable image gets realized when the beam hits an inorganic layer of small crystalline powder. According to the application, the type of phosphor is selected for its emission spectrum and brightness efficiency. By adding certain doping material to the basic crystal structure, the emission characteristics of the phosphor can be optimized.

The nomenclature, in Proelectron or EIA notation, emission colour, colour coordinates with reference to the Kelly-Diagram, chemical components and persistence

Table 2.6. Emission colours and chemical components of some usual phosphors

Phosphortype		Colour	CIE 1931 Coord.		Chem. comp.	Persistence
Proel.	EIA		*x*	*y*		
BE	P11	blue	0.139	0.148	ZnS:Ag	10–1000 µs
GH	P31	green	0.226	0.528	ZnS:Cu	10–1000 µs
GJ	P1	yellow/green	0.218	0.712	Zn_2SiO_4:Mn	1–100 ms
GY	P43	yellow/green	0.333	0.556	Gd_2O_2S:Tb	1–100 ms
KJ	P53	yellow/green	0.368	0.539	$Y_3(Al,Ga)_5O_{12}$:Tb	1–100 ms
RF	P56	red	0.640	0.335	Y_2O_3:Eu	1–100 ms
WB	P45	white	0.253	0.312	Y_2O_2S:Tb	1–100 ms

time of some important phosphors are listed in Table 2.6, more information is given in [13], p. 185ff.

For any process of phosphor deposition onto the glass faceplate of the CRT, the basic principle how phosphor screens are made up is very similar for all kind of applications.

- The phosphor grains of 3–10 µm particle size are glued on the glass faceplate either in a Settle-process by Potassium-Silicate or a photochemical reaction of some added components.
- On top of the phosphor a very thin acrylic layer is sprayed and dried.
- Above the acrylic layer a small amount of aluminum is evaporated under vacuum and creates a reflective mirror for all the light, which is emitted to the rear side of the tube. The acrylic layer acts only as a barrier during this aluminum evaporation process. Otherwise, the aluminum would penetrate the unprotected phosphor and destroy its emission capabilities. After completion of the screen and before closing of the CRT, the organic compounds of the acrylic material are baked out at high temperature.

Except some very special products like Penetron- and Beam Index CRTs, hundreds of millions of CRTs for consumer and professional applications are using the same concept of generating a coloured image by the combination of three electron guns for the three primary colours. The correlation of the individual electron gun and the assigned phosphor dot is made possible by a shadow mask between electron gun and phosphor matrix.

The original design for colour CRTs, where the three electron guns formed a triangle, was called a "Delta" configuration (Fig. 2.25, left). The holes in the shadow mask and the phosphor dots were round. Newer designs are using the cheaper "In-Line" gun concept (Fig. 2.25, right). In this design the two outer guns are slightly tilted against the central gun, so they can match the same hole of the shadow mask. In another modification the holes of the shadow mask became long and the phosphor is deposited in stripes.

In order to improve the contrast ratio of the screen, the area between the phosphor dots is covered with dark material and works as a "Black Matrix" in which

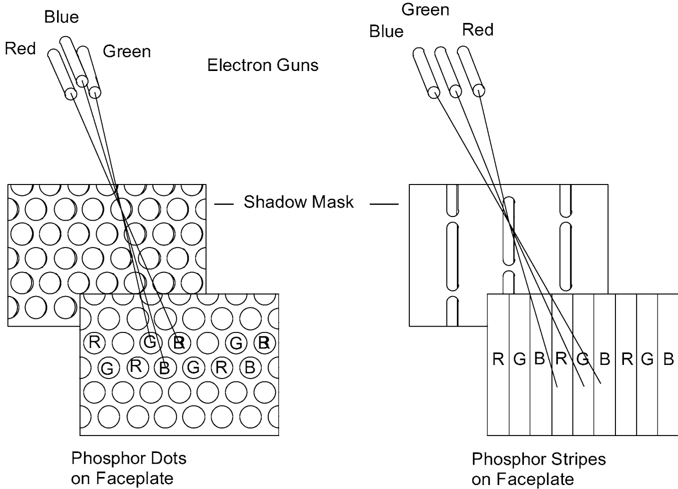


Fig. 2.25. Basic principle of a shadow mask CRT in Delta (*left*) and In-Line (*right*) configuration

the individual phosphor dots are embedded. By this, also slight misalignments of the beams do not lead to a colour shift or colour impurity when the beam does not hit perfectly the assigned phosphor dot. A comprehensive review can be found in [13], p. 263ff.

An additional increase in contrast is gained by using tinted glass for the faceplate, which has to withstand the bombardment of electrons at high accelerating voltages without browning effects. The resolution of a colour CRT is given by the distance between neighbouring RGB-triplets of phosphor dots or stripes. The so-called pitch of a standard TV CRT is approximately 0.8 mm, where very high resolution professional colour CRTs can go down to 0.15 mm.

It is obvious that the shadow mask absorbs quite a large portion of the overall electron beam. At typical anode voltages of 25 kV and beam currents of approximately 300 μA, the mask will be loaded with more than 10 Watts thermal power. The resulting extension of the metal mask leads to the "doaming effect" when the electron beam does not hit the mask holes anymore at the right position and therefore also not the right colour phosphor dots. A black coating of the mask helps here to reduce the temperature by thermal radiation. For high resolution colour CRTs it is necessary to reduce this failure further on by using a special material for the mask with a very low expansion coefficient, e.g. so-called Invar or even Super-Invar steel.

To avoid any colour fringes the three electron beams have to match each other perfectly in the mask's holes. In the centre of the screen this can be optimized by static convergence corrections. With static magnetic quadrupole and 6-pole fields the angle between the neighbouring electron beams can be fine tuned for best performance (Fig. 2.26). Outside the screen centre of a CRT equipped with an In-Line gun, convergence corrections can be implemented by design into the yoke, leading to

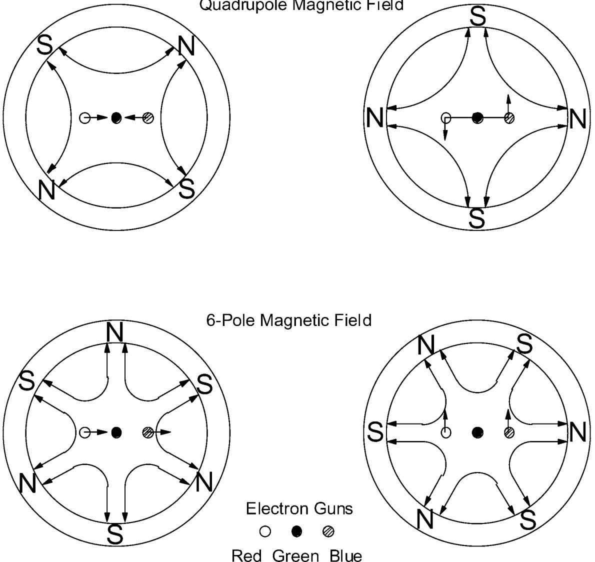


Fig. 2.26. Convergence corrections with static quadrupole- and 6-Pole magnetic fields used for an In-Line gun

a self-convergence over the entire screen. Convergence corrections for a Delta-gun need a quite complicated arrangement of small, individual magnetic tabs.

The effect of changing the magnetic environment can be easily seen on board of airplanes, when TV monitors see heavy shifts in colour purity caused by the changing Earth magnetic field. To shield the CRT against magnetic disturbances at a fixed location, a magnetic shielding is incorporated inside the glass funnel of the CRT. To achieve a good color purity all around the globe it is nevertheless necessary to adjust it to the local magnetic environment.

In addition to a vast number of CRTs for consumer applications, there is a variety of CRTs which are specifically designed for use in all kind of industrial, medical or military applications (see, e.g. [12]). The most demanding designs are required for use in civil and military aircrafts and vehicles.

The extreme optical performance requirements have to be combined with the ability to withstand the harshest environmental stress, like fast temperature changes in the range from $-54°\mathrm{C}$ to $+105°\mathrm{C}$ in operation, shock and vibration of the whole system to levels, where a standard CRT would burst immediately into piece parts, see [16, 17]. To achieve such robustness, a detailed analysis of the resonant frequencies and behaviour of all parts is necessary. With this understanding special fixtures can be designed to hold the precisely aligned electron gun inside the calibrated glass neck of the tube. The whole CRT is assembled into a massive flange and potted with the yoke mounted on into a μ-metal shielding.

Since military aircrafts are flying up to high altitudes, where the sun is shining directly on the displays of the cockpit, it is necessary to show the data symbols with

extreme brightness. Such a brightness up to 100 000 cd/m^2 can be achieved when the yoke of the CRT is driven in a vector or stroke mode. This means that the image of the CRT is not generated in a raster scan, the electron beam is moved exactly in the way the individual character is written. The consequence of this stroke mode is that no time is lost anymore for writing lines without any information. Therefore, the writing speed can be reduced to moderate levels in order to gain more efficiency from the phosphor.

Special CRTs using these technologies are Head-Up-Displays (HUD) and Helmet-Mounted-Displays (HMD). HUD-CRTs project their image onto a half-transparent glass plate, which is placed in the viewing direction of the pilot. By this, the pilot can see the information directly in front of him without moving his head down to the instruments. HMD-CRTs are mounted directly inside the helmet of a pilot and project their image onto the inner side of the helmets visor. Of course, the dimensions of these kinds of CRTs are very small, approximately $1/2''$ diameter screen size, and also very light weighted. Nevertheless, the resolution is better compared to a standard video.

There are additional numerous high performance applications for CRTs, from highest resolution medical imaging of X-ray pictures to projection CRTs for professional flight simulators or reprography systems for the film industry.

2.2.4 Future Aspects

From a pure technical point of view it is obvious that most of the development work is no more spent for improvements of Cathode Ray Tubes. Nevertheless, even during the most recent shows of the display industry CRT's have to not to hide themselves with respect of performance and image quality.

From a commercial point of view Cathode Ray Tubes are still the dominant display in the global market. Especially when one is not only looking to the most advanced societies, CRT's will be the display of choice because of cost reasons in a lot of developing countries. Asides from the cost aspect, CRTs may keep their niches in various professional applications, where small lot size production is not attractive enough for the flat panel industry to justify the development of special application displays. Therefore, Mr. Braun for sure would have made the ultimate deal if he would have secured any percentage of commission.

2.3 Plasma Panel Displays (PDP)[3]

Plasma Displays, an emissive flat panel technology combined with a relative large pixel size, are highly suitable for computer data, images and video heading public information, presentation and TV sets.

[3] Please refer to K. Blankenbach.

2.3.1 Historical Development

Electrical discharge in gases is one of the oldest known phenomenons to produce light. Around 1900 this effect became understood as electron beams. Gas discharges are the basis of devices like gaseous triode, rectifying tubes and surge arresters. They break up neutral atoms or molecules into ions and electrons, this state is called plasma – often named as fourth state of matter. An often unknown fact is that the first display application of plasma was the NIXIE tube, as briefly described in the historical development of vacuum displays (Sect. 2.1.1). The technology of plasma displays was invented in the early 1960s, being commercialised about 10 years later with monochrome displays using the orange-red emission of neon gas [18]. Plasma phosphors are based on oxides, while CRTs rely on sulphides. Colour PDPs [19] were introduced around 1995, while mass production started 5 years later. The most widespread screen size was 42″ (1.06 m) with an aspect ratio of 16:9 and WVGA (832×480) resolution. Nowadays (2007) PDPs address the HDTV market. In the mid 1990s plasma driven LCDs were developed but never became a product.

2.3.2 Electrophysical Fundamentals

The basic principle of all plasma displays is to discharge and generating a glow by exciting ions that collide with each other. This effect is also the basis of fluorescent neon tube lighting. If a voltage is applied to a capacitor within a tube filled with a low-pressure gas, several states can occur. At low voltages the gas acts as insulator. By increasing the voltage, a breakdown occurs, the gas becomes ionised and the current through the tube increases by some orders of magnitude. This current has to be limited for display applications. A display related summary of gas discharges can be found in [20, 21].

Figure 2.27 gives review of the large variety of plasma displays, details can be found, e.g. in [12, 22]. The waveform of the driving signal to generate the plasma is the main categorisation characteristic. Plasma devices can be driven either by directed (DC) or alternating (AC) current. DC electrodes interface directly with the gas, while AC electrodes, to enhance their lifetime, are shielded by dielectric films from the plasma. The electrodes form, in the simplest case, a capacitor with gas in between the plates. When they are covered by a dielectric, charges can be stored on theses films forming a capacitor – it results in the memory type. However, there is no plasma display technology available which uses the DC memory effect. Another advantage of the ‘internal’ capacitor is the limitation of the breakdown current when igniting the plasma. The refresh type uses a resistor in series to the pixel capacitor; one DC application was the NIXIE tube. For typical luminance required in applications, the dot matrix refresh types are limited to around 200 lines, which is too small for computer or video applications. Only the dot matrix AC type was in mass production when this book was written, therefore this chapter focuses mainly on the AC Memory Type Dot Matrix display, which is highlighted in Fig. 2.27.

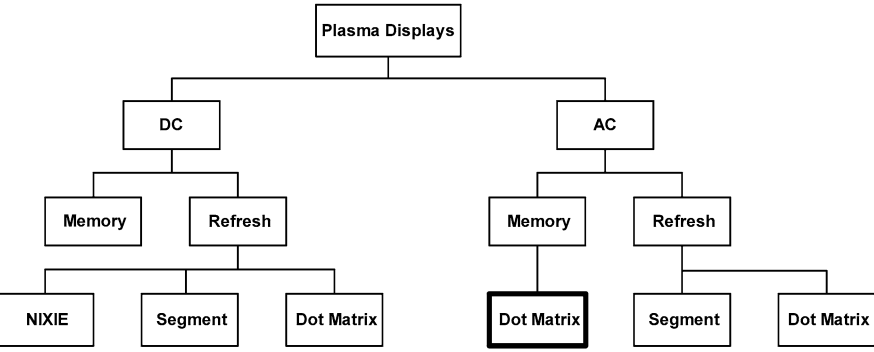


Fig. 2.27. Plasma display tree, only AC memory dot matrix displays are currently (2007) commercially available as large information panels and TV sets

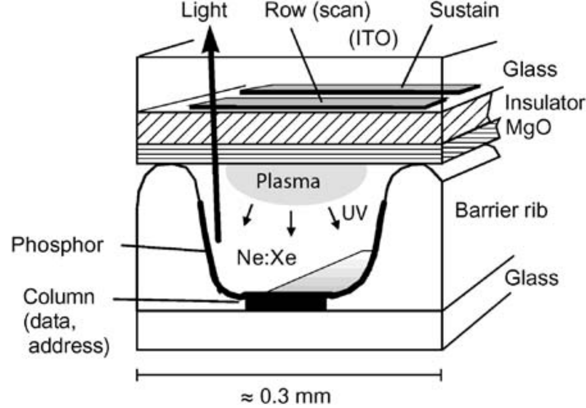


Fig. 2.28. Elements of a simplified AC Memory Plasma Display pixel

This most common plasma display is constructed as a three electrode surface discharge AC device, as shown in Fig. 2.28. The panel consists of two glass plates. The front plane with 2 parallel transparent electrodes (row and sustain, made of Indium Tin Oxide) faces the viewer. An insulation layer, covered by MgO acting as dielectric, completes this assembly. On the backplane, the column electrode is placed, often made by Ag by photographic process or screen printing. Sand blasted barrier ribs prevent crosstalk between the pixels of a column and are used as spacers keeping the two glass plates parallel. On the inner side of these ribs and on the address electrode, phosphor is placed. Both subassemblies were processed separately and then sealed together. After evacuation the panel is filled with gas, typically Neon with 10%Xe mixture at a pressure of about 500 Torr for colour plasma displays; Xe reduces the threshold voltage by the Penning effect.

If a voltage above the threshold value (typically 200–500 V) is applied between two electrodes, the electric field accelerates free electrons to high velocities to ionize the atoms of the gas. This plasma emits UV radiation at 147 nm by exited Xe atoms

for modern AC Ne:Xe displays. This invisible emission is converted to red, green and blue by exciting the corresponding phosphor. Because of the capacitive coupling an AC voltage is required for subsequent light pulses.

The sequence for activating a 3-electrode AC PDP is shown in Fig. 2.29:

- *Priming and Data-write Discharge* (*ignition, left*)
  First a voltage (about 250 V) is applied to the scan (X) and sustain (Y) electrodes causing a priming discharge. Then a voltage is applied between scan (X) and column (data) for the pixel (selected cells) to be stimulated. The gas is ignited and emits radiation. The duration of priming and data-write discharge lasts about 2 μs and is required per line. The priming discharge is sometimes made outside the visible display area in the rows at the top and bottom of the plasma panel. The surface charge on the walls of the initialised pixels last longer than the duration for scanning all the lines (memory effect).
- *Sustaining Discharge* (*centre*)
  An AC voltage (about 150 V, 50 kHz) between scan (X) and sustain (Y) electrodes is added to the voltage from the wall charge created during the ignition phase. This causes a surface discharge from which UV rays are emitted. This step is repeated as often as it is required for the grey shade to display. Cells without a wall charge (not addressed in the first step) can't light since the ignition threshold isn't reached – the sustain voltage is lower than the breakdown voltage. Because the repetition frequency is high, the light pulses merge for the observer. A mean luminance corresponding to the number of pulses is detected.

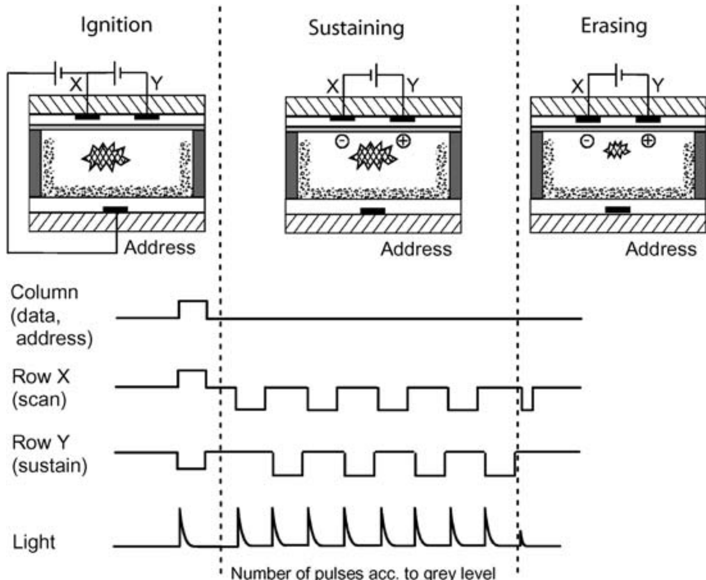


Fig. 2.29. Waveforms and light emission of an AC PDP pixel (simplified)

- *Erasing Discharge* (*right*)
  A lower voltage (and often with a shorter pulse) is applied between scan (X) and sustain (Y) electrode. This should neutralize the wall charge and prepare the pixel for the next addressing (frame or subframe).

Because the pixel information after the data-write discharge (ignition) is ‘held’ for the sustaining phase, this driving mode is called memory mode, opening the gate to high resolution plasma displays; all rows are scanned and the corresponding pixel information (ON or OFF) is written by the address electrode (column). After all rows are scanned, the panel is ‘loaded’ with the information for each pixel, and during the sustaining phase all selected pixels will light. The time needed for ignition is about 1 μs, the recovery time (gas goes from ionized to the neutral state) in the range of 10 μs. A deeper look from the systems point of view is in the chapter about driving of PDPs. An excellent review on calculations for the driving voltages can be found in [23]. The AC type cannot emit light continuously because the capacitors don’t let a continuous current pass. This is opposite to the DC type, which is limited in current by a resistor. When ions collide with the phosphor, sputtering will occur because of the relative large mass of the ions. This degrades the phosphor and reduces the lifetime of the display significantly.

Some features of gas discharges make them very attractive display devices. Passive Matrix addressing can be done as the threshold voltage dependency is very steep (only ON and OFF mode) and the efficiency is relatively high (about 1 lm/W).

2.3.3 Present State-of-the-Art and Applications

Currently (2007) the most widespread PDP has HD ready resolution (720 lines) and an aspect ratio of 16:9. This is shown on the left side of Fig. 2.30, on the right, a 42″ PDP is shown. A major issue for applications with PDPs is the different ageing of the phosphors; typically the blue one has a lower lifetime compared to red and green. Therefore, plasma panels tend to a shift of the white point to yellow with increasing operating time. One can reduce this effect by avoiding static screen content like that of check-in displays at airports.

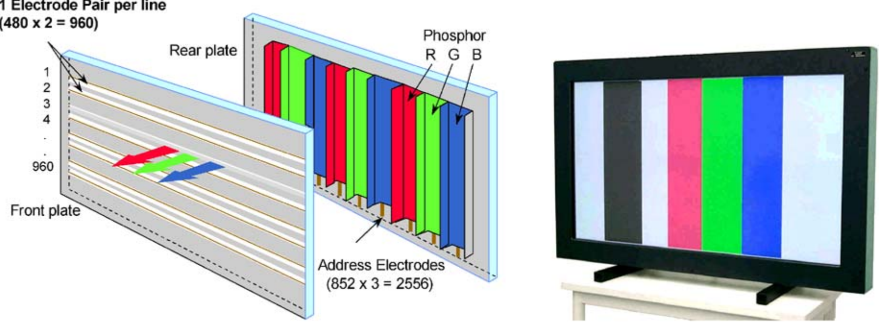


Fig. 2.30. Simplified structure of a AC colour plasma panel with WVGA resolution

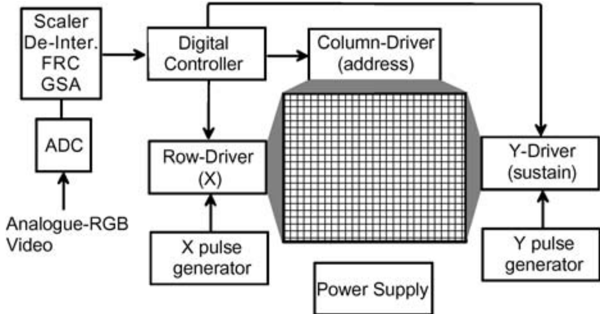


Fig. 2.31. Block diagram of AC memory plasma display electronics

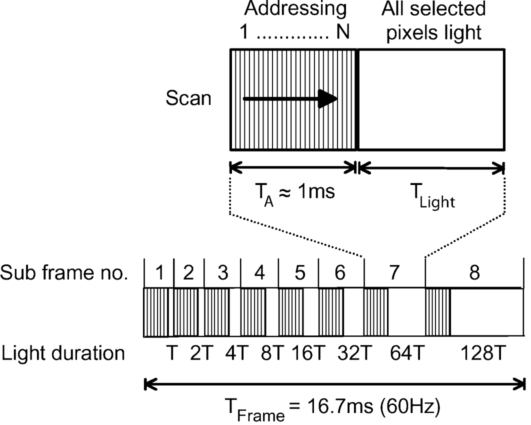


Fig. 2.32. Grey scale generation by subframes. The sequence of Fig. 2.29 has to be generated for every subframe; the value of the grey scale subframe determines the number of sustaining pulses

Figure 2.31 shows the block diagram of a PDP [11] including analogue signal processing. At first, the analogue signal is converted to digital data (ADC), and these data have to be fitted via software to PDP driving patterns; some details are described in Table 2.1. The according signals for rows and columns are shown in Fig. 2.29. The challenge in PDP electronics is the relatively high voltages needed for igniting and sustaining the plasma, which is very costly and consumes a lot of electrical power.

Due to the sharp discharge–voltage characteristic, passive matrix driving is sufficient even for high-resolution displays in comparison, e.g. to LCDs, where a costly active matrix addressing is mandatory for high pixel counts. Another consequence is that PDP pixels can only be switched ON or OFF. This allows an all-digital display, with the grey scale being generated by subframes (Fig. 2.10). As the sustaining frequency is very high, the pulses merge to a mean grey shade for the eye. The driving scheme for one frame is shown in Fig. 2.32: Each subframe consists of two parts, address and sustaining. In the address sequence (compare to Fig. 2.29, ignition, no

or nearly no light is produced) all of the pixels, which should lit in the corresponding subframe, are initialised; this lasts typically 1 ms (480 lines × 2 μs per line). All selected pixels emit light during the sustaining phase. The number of pulses represents the grey value of the subframe. The maximum luminance of the display is therefore limited from the drivers' point of view by the addressing phase which lasts about 50% of the total frame time for 60 Hz (16.7 ms) and 8 subframes. Luminance can be increased either by faster addressing or by a higher sustaining frequency, but both methods are very costly. Other optimizations are based on phosphors and gas characteristics. Furthermore, PDPs are sensitive to false contouring when displaying moving objects on the screen using subframes.

The advantages of Plasma Display Panels are their large size, while keeping the depth shallow (10 cm), video capability by adequate colour gamut and fast response time. Latest developments in 2006 target sizes large than 60″ with full HDTV resolution. Some drawbacks, like readability in bright sunlight and high power consumption (300 W for 42″ causes relative high heat generation), could be accepted in many video applications. The images to be displayed should be examined to prevent degradation of public information displays, such as timetables and advertisement, due to PDPs' relative short lifetime (about 10 000 h, burn-in) and differential RGB ageing (colour shift). These effects can be reduced by complementary colours, screen savers and by avoiding logos and icons displayed on the same area. Due to their relatively large pixel size of about 1 mm (colour pixel), PDPs are not useful for applications with low viewing distance, e.g. desktop monitors. In comparison to CRT displays, PDPs show no linearity and convergence errors and are insensitive to magnetic fields; they are lightweight (20–30 kg) compared to large-sized CRTs. Improvements in the light output are being introduced by FUJITSU™ replacing the sustain electrode by the neighbouring addressing one and driving the panel in the interlaced mode.

2.3.4 Future Aspects

In 2004, more than 2 million PDP units, most of them 42″ and 50″ types, were shipped worldwide with a total value of about 5 billion $. The sold units in 2007 are supposed to increase by a factor of 4, while the revenue rises about a factor of three. A colour pixel pitch of 1 mm makes PDPs very attractive for HDTV with 720 lines. In this case, all ergonomic requirements fit very well; typical distance of observer is about 3 m for TV displays, while the human visual resolution being there about 1 mm; the minimum diagonal for 720 lines and 16:9 aspect ratio is about 50″. This gives great opportunities for PDPs for high volume production in the future.

2.4 Vacuum Fluorescent Displays (VFD)[4]

The first flat panel display technology was created by transferring the fundamentals of CRTs to flat devices using a broad electron source and no deflection – the

[4] Please refer to K. Blankenbach.

outcome was named as Vacuum Fluorescent Displays. Their construction is like a triode. Major VFD applications target low information content displays in household appliances and automotive displays. As an emissive technology, VFDs have a large viewing angle. Due to their high luminance output, they are even applicable in bright ambient light conditions. Compared to other vacuum based displays, they have relatively low driving voltages.

2.4.1 Historical Development

The first commercial VFDs were developed by ISE and FUTABA around 1965 as 8-segment single tube displays, partially replacing NIXIE tubes. Some years later, multiple digits were integrated into one device making the construction of small numerical displays, e.g. for calculators possible. The early history of VFDs can be found in, e.g. [24]. Typically VFDs glow green. As the number of segments (pixels) increased, multiplex and matrix driving was introduced. Since 1990 graphic displays up to QVGA are available with colour capability, mainly sold as area colour devices. Because of their relatively high power consumption, VFDs were replaced in mobile devices by LCDs like in calculators around 1980.

2.4.2 Electrophysical Fundamentals

As shown in Fig. 2.33, an evacuated glass tube holds the basic elements of VFDs, see, e.g. [11, 25]. The cathode is a thin tungsten filament (diameter around 20 μm, shouldn't be visible) coated with BaO for electron emission, which is heated up to 600°C to emit thermal electrons. A metallic mesh forms the grid; the structured anodes form the pixel shape and are phosphor-coated (usually bluish-green ZnO:Zn with the peak at 505 nm and a half-width of about 100 nm). Due to these discrete elements, neither focussing nor deflection are necessary compared to CRTs. The production processes for the anode and layer structure (including insulation) are made by thin- and thick-film processing, which are reliable and relatively inexpensive. The light emission is controlled by both grid and anode voltage, typically in the range of 20–50 V at currents of about 10 mA. Therefore, low voltage phosphors are used opposite to those requiring an acceleration voltage of 20 kV as in CRTs. The electrical function is that of a triode: the grid and anode voltage, which should have both a positive polarity for light output by electrons hitting the phosphor layer, accelerates

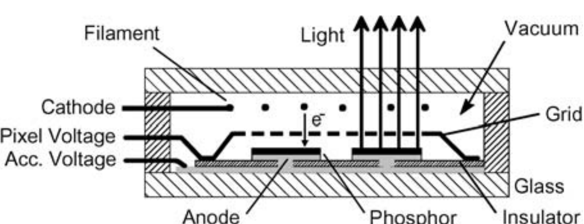


Fig. 2.33. Major elements of a Vacuum Fluorescent Display

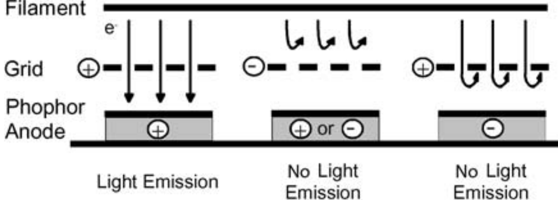


Fig. 2.34. Light emission of VFDs by modulating the number of electrons hitting the phosphor

electrons from the hot cathode. In comparison to CRTs, the electrons hit the phosphor on the same side as the observer (like PDPs), resulting in a peak luminance of up to 30 000 cd/m^2 with efficiency of 5 lm/W being achievable. Additional information on VFDs can be found in [4].

The fundamentals of driving VFDs by grid and anode voltages are demonstrated in Fig. 2.34. The cathode (filament) emits electrons, which are influenced by electric fields. In the case where both the grid and anode (left) have positive voltages, the pixel is selected, and the electrons can pass the grid and hit the anode emitting light. Because the phosphor layer would be charged, it must be conductive to avoid a negative potential. A negative potential of the grid blocks the electrons from reaching the anode despite their voltage (centre). If the grid is positive and the anode negative, the electrons are absorbed by the grid (right) without emitting any light. The grid should be as 'transparent' as possible, e.g. a fine meshed grid (typically hexagonal, see Fig. 2.35). Usually the grid and anode operate at the same potential (as mentioned 20–50 V) to avoid focussing problems. For high multiplex ratios (low duty factor) the driving voltages are raised to achieve an acceptable luminance. Examples for driving waveforms are presented in the next chapter.

2.4.3 Present State-of-the-Art and Applications

A typical realisation of a matrix (grid) driven VFD [25] for a character display is shown in Fig. 2.35. On a glass substrate (bottom) a structured anode forms the pixels of the font matrix, a dedicated grid covers this arrangements. The filament wires have to be spanned carefully in parallel without hanging through or vibrating, which leads to some non-uniformity in light emission. The last steps during production are mounting the front glass on the substrate and evacuation through the exhaust tip, which is sealed after this procedure. Vacuum fluorescent displays are available with leads or as integrated modules with dedicated drivers on a PCB or inside the glass (Chip on glass technology), some typical modules are shown in Fig. 2.36. The COG technology results in ruggedized displays which are suitable for harsh environments. A 20-character module with 2 lines (8 mm font size) is normally 4 cm high, 12 cm wide and 1 cm deep without leads. A lifetime of around 30 000 h can be achieved. Besides their low aperture ratio (usable area/total display area), relatively high power consumption, thickness and weight (vacuum based display) are major disadvantages of VFDs. The contrast ratio can be increased by using colour filters, which reduce the

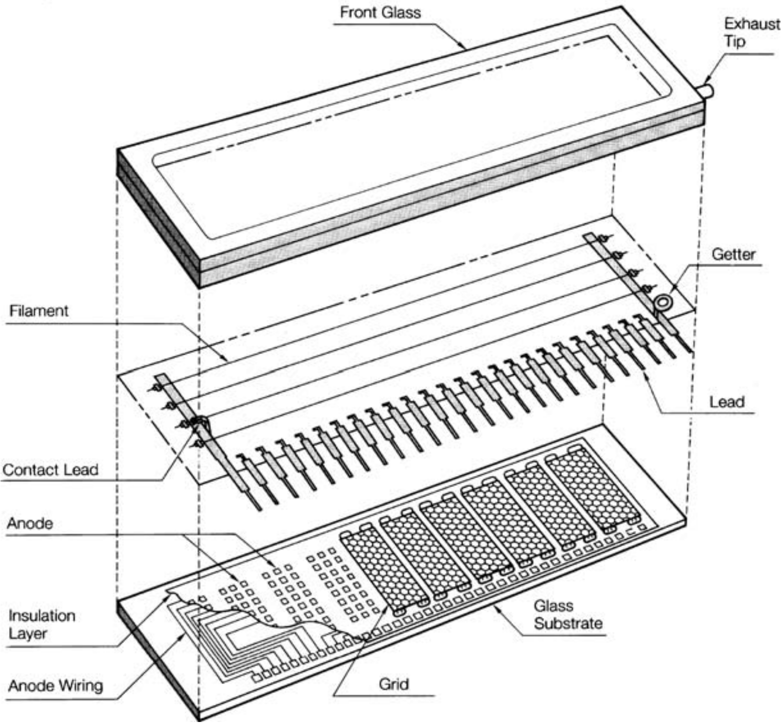


Fig. 2.35. Simplified structure of an alphanumeric VFD (with permission of FUTABA)

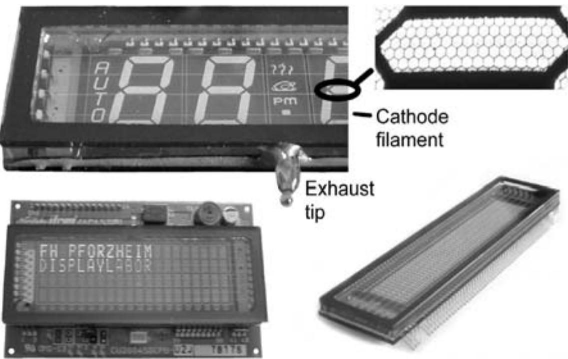


Fig. 2.36. Examples of VFDs: 8-segments with icons (*top left*, the horizontal filaments are clearly visible), honeycomb-like grid (*magnified*), matrix character display (*bottom left*) and graphics module (*bottom right*)

level of reflected ambient light due to the narrow half-width of the VFD emission. Colour filters can be also used for slight variations of the visible display colour.

As for all the flat panel technologies, direct and matrix addressing is also applicable to VFDs. Direct and low multiplex driving is used for 8-segment and character displays, while matrix addressing is the domain of graphic devices. Despite some efforts, high resolution PM and AM VFDs never became a commercial success.

A typical realisation of a microcontroller driven display system is shown in Fig. 2.37. A 4 digit display (4 grid leads) is connected via grid driver and anode driver (8 leads, 7 for digits, 1 for decimal point), so that in this case 13 connections (including one for the filaments) have to be made. All corresponding segments of the digits are connected. Hence, a multiplex driving scheme (simplified matrix principle) has to be applied. The software of the microcontroller has to provide the corresponding waveforms for the drivers to display the intended number.

The waveforms to drive such a numerical display are plotted in Fig. 2.38. The grids are scanned one by one (top) and the anodes corresponding to the number to be

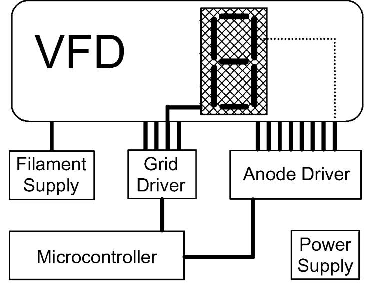


Fig. 2.37. Example of Segment VFDs driving by a microcontroller

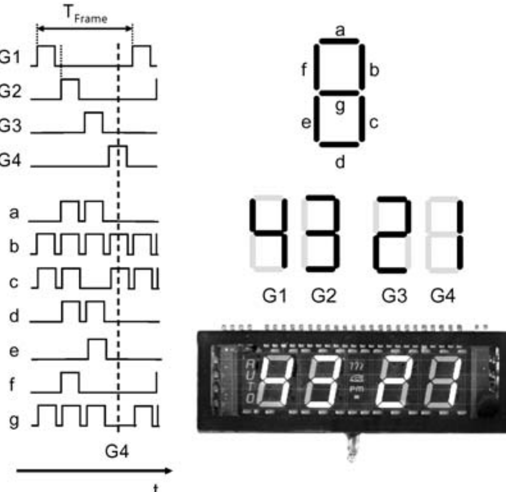


Fig. 2.38. Waveforms for an 8-segment VFD with an image of a driven display

displayed have to be activated, according to Fig. 2.34. They have a positive potential relative to the cathode. After the 4 grids are scanned, the first one (G1) is activated again; the duration (frame time T_{Frame}) of such a period should be low to prevent flicker. Between two grid signals there's a short blanking time to prevent ghosting due to decay of the phosphor. The segments (anodes) of the digits are usually named as shown in the inset. When the driving waveforms are applied, the output of the 4 digit VFD is as shown. As an example, the fourth digit (4G) is examined: if grid G4 is activated, the segments 'b' and 'c' are also positive, therefore a '1' will be displayed.

2.4.4 Future Aspects

In 2003 VFDs had a market share of Flat Panel Displays (FPDs) of about 10% on unit basis and approximately 2% on revenue basis, which equals 600 million $, which is nearly unchanged for more than a decade [1, 12]. Most of these displays show low information content as 8-segment, characters, icons and bar graphs with moderate screen sizes up to 5″. Major applications are household appliances like stationary audio and video equipment, point of sales and electronic cash registers and automotive displays. Graphic colour VFDs (passive or active matrix) never reached a broad commercial basis. The largest VFD ever made was the JUMBOTRON™ by SONY and FUTABA for the EXPO 1985 in Japan [12]. It was the largest display ever made to that time and had a screen size of about 10 m (400″). The display was capable to show colour videos even outdoors with 150 000 pixels.

2.5 Field Electron Displays (FED)[5]

2.5.1 Historical Development

Employing the micro-fabrication technology field, emitters made from molybdenum were developed as the source of electrons by Shoulders, Brodie and Spindt at Stanford Research International, CA, USA [26], and published in an extensive review [27]. Arrays of cathodes with 10 million/cm^2 could be produced, which leveled the noise of the single emitter statistically and showed lifetimes in UHV environment exceeding 10 000 hours. These arrays became attractive to develop flat panel displays [28] and array cathodes for high frequency vacuum tube devices and other applications [29, 30]. The principle of FED is briefly explained in Sect. 2.1.2, Fig. 2.3; different cathode structures are shown in Fig. 2.39.

2.5.2 Electrophysical Fundamentals

In the CRT, 3 beams compose a colour image at the screen. Each colour is served by one beam. In the FED, each pixel is served with at least 3 beams, for each colour at

[5] Please refer to H.W.P. Koops.

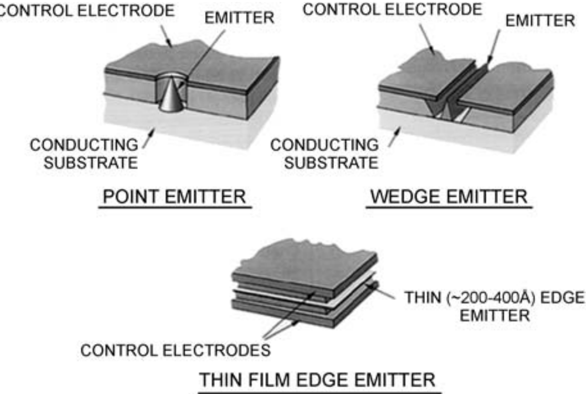


Fig. 2.39. Major types of FED cells. Point emitter (Spindt emitter (Mo) or Gray emitter (Si)). Wedge emitter and thin film edge emitter with extractor grid control electrodes

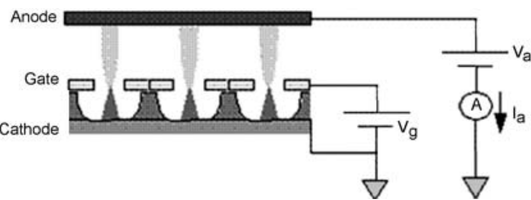


Fig. 2.40. Schematic of 1 pixel, electronic circuitry and function. Electrons are extracted via a grid voltage and accelerated to the screen at the anode

least one beam. Figure 2.40 shows the schematic of 1 pixel of an FED, its electronic circuitry and function. The electron emitter delivers the emission current. It is accelerated through a grid to the anode, which is covered with a scintillator material. The electric equivalent is a triode tube. In the FED, the source and scintillator are in close vicinity, and the image intensity is controlled switching the grid voltage. An image with 1 million pixels is composed of at least 3 million beams.

2.5.3 Present State-of-the-Art and Applications

Several novel materials applied as cathodes in FEDs are given in Fig. 2.41. Figure 2.42 shows an array of Spindt-type Mo cathodes and a cross-section of a single emitter cell. Typical performance data of field emission cathodes are emission current up to 10 µA at an extraction voltage of 40 V to 80 V. The FED anode voltage is 3 kV to 6 kV. The total current in the display is up to 10 mA. The cathodes with very low grid voltage are preferred, since it is costly to develop fast high voltage switching amplifiers.

First colour FEDs were demonstrated in 1993–1995 by Pixtech, Micron, and Futaba with Mo, Si, or diamond film emitters. In 1997, Motorola and Candescent demonstrate a VGA 15″ high voltage colour FED with Mo Spindt-type emitters.

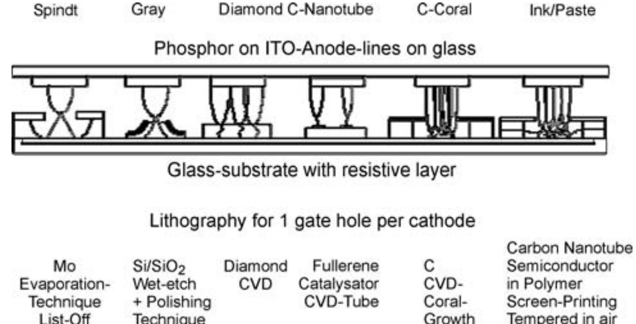


Fig. 2.41. Novel cathode materials are Molybdenum, Silicon, Diamond, corraline carbon, carbon nanotubes and particles suspended in a polymer matrix

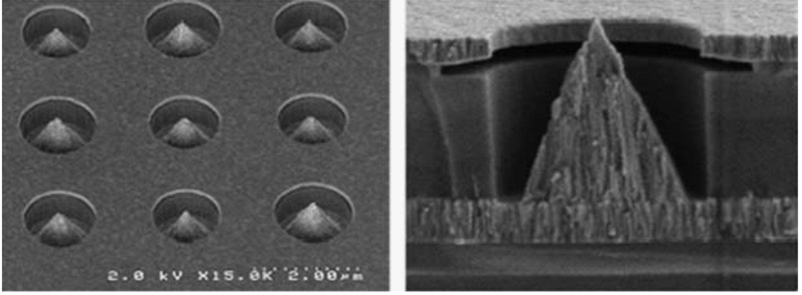


Fig. 2.42. Mo-Microtip array of Spindt type. The emission strongly depends on tip height, radius of curvature and position in the gate opening (Pictures courtesy C. Spindt SRI)

In 2003, Samsung demonstrates a 38″ HDTV colour TV with Carbon Nanotube (CNT) paste thin film edge emitters. The gate voltage is reduced. The emission current density is up to 8 mA/cm^2. The lifetime at >50% current is >10 000 h. As cathode material a binder-CNT paste is used, which is structured with photolithography. This fabrication method allows to produce 40″ displays at low cost. A screen with a gate–anode distance of 0.7 mm, an anode voltage of 3.8 kV was packaged and operated successfully. The onset voltage was 25 V. A low-cost Carbon Nanotube FED production is possible since inexpensive driver electronics can be used [31].

2.5.4 Future Aspects

The comparison of the characteristics of the competing technologies clearly shows the superiority of CNT–FED's. However, the market penetration is difficult for a new technology coming up after the times when other earlier technologies have already left their production learning curve behind, and products are sold at low prices. LCD VGA screens for computers have reached so low prices that a newcomer cannot make

it. PDP are too expensive and consume too much power. However, a good market chance is now present for the low power consuming big FED Home TV screens.

Future development trends, as visible in Japan by Futaba, Mitsubishi, Sony, and Samsung are CNT FED's to be sold as replacements of LDC screens. A market share will certainly be there for big FED 30″–50″ displays for home TV, despite of traffic and military applications. The total market for big screens is expected in 2004 to be 2.5 million units. This will be plasma displays PDP, since FED display factories still have to be built.

Start-up companies engage in using carbon CNT-emitters for lamps, especially in home appliances. Lamps having a cathode–anode gap of 0.2 mm and an anode voltage of 390 V achieve a brightness of 400 cd/m^2.

References

[1] Many articles and books about display technologies can be found at the websites of the Society for Information Display (www.sid.org). Market data are available, e.g. from DISPLAYSEARCH (www.displaysearch.com) and iSUPPLY (www.isupply.com)
[2] G. Shiers, Ferdinand Braun and the cathode ray tube, Sci. Am. **230**, 92–101 (1974)
[3] G.F. Weston, *Cold Cathode Glow Discharge Tubes* (IL-IEEE Books Ltd., London, 1968)
[4] T.J. Nelson, J.R. Wullert II, *Electronic Information Display Technologies* (Wiley SID, West Sussex, 1997)
[5] R.L. Myers, *Display Interfaces* (Wiley SID, New York, 2002)
[6] D. Bosman, *Display Engineering* (North-Holland, Amsterdam, 1989)
[7] G. Berbecel, *Digital Image Display* (Wiley SID, New York, 2003)
[8] P.A. Keller, *Electronic Display Measurement* (Wiley SID, New York, 1997)
[9] D. Travis, *Effective Color Displays* (Academic, London, 1991)
[10] H. Widdel, D.L. Post, *Color in Electronic Displays* (Plenum, New York, 1992)
[11] S. Sherr, *Applications for Electronic Displays* (Wiley, New York, 1998)
[12] J.A. Castellano, *Handbook of Display Technology* (Academic, San Diego, 1992)
[13] J. Whitaker, *Electronic Displays* (McGraw-Hill, New York, 1994)
[14] E. Harting, F.H. Read, *Electrostatic Lenses* (Elsevier Scientific Publ., New York, 1976)
[15] P.A. Sturrock, *Static and Dynamic Electron Optics* (Cambridge University Press, Cambridge, 1955)
[16] E.B. Shand, *Glass Engineering* (McGraw-Hill, New York, 1958)
[17] A. Roth, *Vacuum Sealing Techniques* (Pergamon, Oxford, 1966)
[18] H.G. Slottow, Plasma displays, IEEE Trans. Electron Devices **ED-23**(7), 760–772 (1976)
[19] H.J. Hoehn, Recent developments on three colour plasma display panels, IEEE Trans. Electron Devices **ED-20**, 1078–1081 (1973)
[20] J.P. Michel, Large area discharge displays or plasma displays, in *Display Engineering*, ed. by E. Bosman (North-Holland, Amsterdam, 1989), pp. 185–218
[21] M.M. Hirsh, H.J. Oskam, *Gaseous Electronics, Electrical Discharges,* vol. 1 (Academic, 1978)
[22] L.E. Tannas, *Flat Panel Displays and CRT*, Chapter 10 (Van Nostrand-Reinhold Company, New York, 1985)
[23] T.N. Criscimagna, P. Pleshko, AC plasma display, in *Display Devices*, ed. by J.I. Pankove, Topics in Applied Physics, vol. 40 (Springer, Berlin, 1980), pp. 91–150

[24] T. Nakamura, K. Kiyozumi, S. Mito, in *Advances in Image Pickup and Display*, vol. 5, ed. by B. Kazan (Academic, London, 1982)
[25] G.J. Hunt, Vacuum fluorescent tubes, in *Display Engineering*, Chapter II.2, ed. by E. Bosman (North-Holland, Amsterdam, 1989)
[26] E.W. Müller, Ergeb. Exakt. Naturwiss **27**, 290 (1953)
[27] K.R. Shoulders, Microelectronics using electron-beam-activated machining techniques, Adv. Comp. **2**, 135 (1961)
[28] R. Meyer, A. Ghis, P. Ramband, F. Muller, Development of matrix array of cathode emitters on a glass substrate for flat display applications, in *Proc. 1st IVMC*, Williamsburg, VA, 1988, p. 10
[29] C. Spindt, J. Appl. Phys. **39**, 3504 (1968)
[30] C. Spindt, I. Brodie, L. Humphrey, E.R. Westerberg, J. Appl. Phys. **47**(5), 248 (1976)
[31] J.E. Jung et al., Development of triode-type carbon nanotube field-emitter arrays with suppression of diode emission by forming electroplated Ni wall structure, J. Vac. Sci. Technol. B **21**(1), 375 (2003)

3

Radiation Sensitive Vacuum Electronic Components and Devices

J.A. Eichmeier

3.1 Historical Development

The fundamental phenomena, which make possible the function and application of this group of vacuum electronic components and devices, are the external photoelectric effect and secondary electron emission. Electromagnetic radiation with wavelengths in the range of infrared, visible light, ultraviolet or X-rays and particle radiation when hitting the surface of a metal, semiconductor or insulator in a vacuum tube cause the emission of electrons from this surface into the vacuum. By application of electric and/or magnetic fields within the tube, the electron current can be used:

1. for the detection of weak light signals with pulse durations down to the picosecond range and for the detection of particle-, UV- and X-radiation, respectively. In each case the input signals are transformed into small electrical current impulses with subsequent current amplification (secondary electron multipliers);
2. for the conversion and amplification of infrared, visible, ultraviolet or X-ray images into visible intensified light images (image converters, image amplifiers);
3. for the conversion of visible light pictures into electrical currents as input signals for television devices (television camera tubes).

The historical development of these types of radiation sensitive vacuum tubes is characterized by the following detections and inventions:

1887: H. Hertz observes that the ignition voltage of a spark gap decreases when the negative electrode is irradiated with UV light.
1888: W. Hallwachs shows that the effect detected by Hertz is caused by the emission of negatively charged particles from the negative electrode.
1890: J. Elster and H. Geitel, and also P. Lenard show that the charged particles are electrons. Elster and Geitel construct the first vacuum photocells for studying the photoeffect of alkali metals.

1902: L.W. Austin and H. Starke detect the secondary electron emission from metals.
1905: A. Einstein detects the energy equation of the photoelectric effect (Einstein equation).
1928: R. Suhrmann describes the photocurrent amplification by secondary electron emission and the principle of photomultipliers.
1929: V.K. Zworykin invents the first television camera tube (the Iconoscope).
1934: P. Farnsworth and G. Holst describe independently the principle of image converters and image amplifiers.
1950: P.K. Weimer et al. from RCA construct the first photoconductive camera tube (the Vidicon).

3.2 Electrophysical Fundamentals

The function of radiation sensitive vacuum electronic components described in this section is based on photoelectron emission, secondary emission, and electron optics.

3.2.1 Photoelectron Emission

Photons with wavelengths from infrared radiation down to X-rays incident on a solid surface (photocathode) cause the emission of photoelectrons according to the energy equation (Einstein equation)

$$E_k = hf - W_c, \tag{3.1}$$

where E_k is the kinetic energy of the emitted photoelectron, h is the Planck constant, f is the frequency of the photon and W_c is the work function of the photocathode. The smallest light frequency f_{min} or the largest wavelength $\lambda_{max} = c/f_{min}$ necessary for the emission of photoelectrons ($E_k = 0$) is defined by

$$f_{min} = W_c/h \quad \text{or} \quad \lambda_{max} = hc/W_c. \tag{3.2}$$

λ_{max} is the cut-off wavelength of the photocathode. A monochromatic light flux Φ (in lumen, lm) of the frequency f is equal to a power flux Φ_e (in Watt, W) through a cross-section A,

$$\Phi_e = S_{ph} hfA, \tag{3.3}$$

where S_{ph} is the number of photons incident on the surface per second and cm^2. For the wavelength $\lambda = 555$ nm (maximum eye sensitivity), the relationship between Φ (in lm) and Φ_e (in Watt) is

$$1\ \text{W} = 673\ \text{lm}\ (\lambda = 555\ \text{nm}) \quad \text{or} \quad 1\ \text{lm} = 0.001484\ \text{W}. \tag{3.4}$$

A part r of a light flux Φ_o incident on a solid surface is reflected (r is the reflection factor). The penetrating light flux Φ_a is given by

$$\Phi_a = \Phi_o(1 - r). \tag{3.5}$$

The penetrating light flux Φ_a is absorbed in the solid according to

$$\Phi_a = \Phi_o e^{-\alpha x}, \tag{3.6}$$

where α is the light absorption coefficient, x is the light pathlength in the solid and Φ_a is the light flux at x. For $x_o = 1$, we have $\Phi = \Phi_a$. x_o is called light penetration depth.

The photoelectron yield A_e of a photocathode is defined as the emitted photocurrent I_{ph} divided by the light power flux Φ_e,

$$A_e = I_{ph}/\Phi_e. \tag{3.7}$$

The lumen sensitivity s_c is defined as photocurrent per lumen

$$s_c = I_{ph}/\Phi, \tag{3.8}$$

and the quantum efficiency η_q as

$$\eta_q = S/S_{ph}, \tag{3.9}$$

i.e. the number S of photoelectrons emitted per incident photon. From (3.3) and (3.7)–(3.9) follows

$$\eta_q/\% = 0.124 A_e/(\text{mA/W})/\lambda/\mu\text{m}. \tag{3.10}$$

Semiconductors are preferred as light sensitive layers for photocathodes because of their low light reflection, high photon absorption coefficient, large emission depth and high quantum efficiency as compared with metals. The quantum efficiency, η_q, of different photocathodes as a function of photon energy, E_{ph}, and light wavelength, λ, respectively, is shown in Fig. 3.1. The points of intersection of these curves with the abscissa indicate the cut-off wavelengths of the photocathodes. Characteristic values of technically important photocathodes are presented in Table 3.1.

Table 3.1. Characteristics and applications of some photocathodes

Material	Nomenclature				Applications
	s_c (μA/lm)	A_e (mA/W)	η_q (%)	λ_{max} (μm)	
S-1/Ag-O-Cs	60	6	0.5	0.8	IR detectors
S-10/Bi-Ag-O-Cs	80	40	10	0.45	Camera tubes
S-11/Cs_3Sb on MnO	80	65	20	0.44	Multiplier, scintillation counters
S-20/(Cs)Na_2KSb	300	130	30	0.42	Image amplifiers, photometers, camera tubes

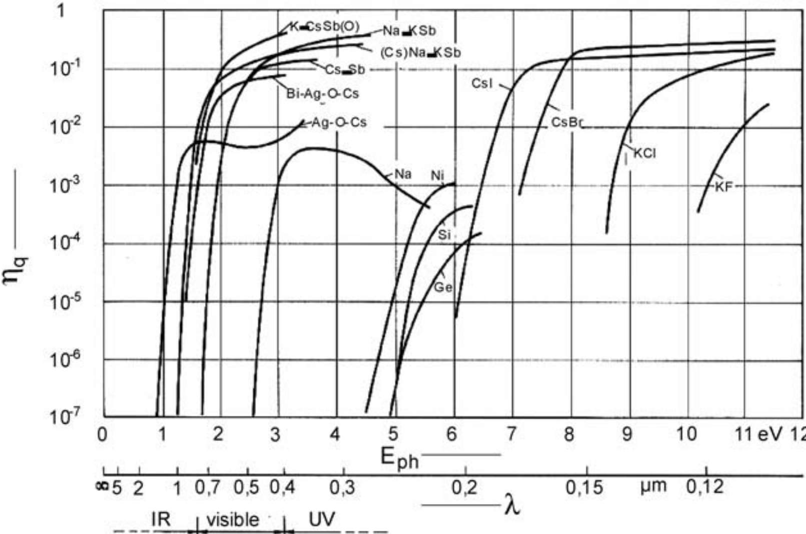


Fig. 3.1. Quantum efficiency, η_q, of different photocathodes as a function of photon energy, E_{ph}, and of the light wavelength, λ, respectively

3.2.2 Secondary Electron Emission

A primary electron beam with an energy of a few tens eV or more, incident on the surface of a solid, produces the emission of secondary electrons with a typical energy distribution, as shown in Fig. 3.2. The left maximum at low primary electron energy is due to "genuine" secondary electrons (share: 90%), the small right maximum represents elastically reflected primary electrons (about 3%) and the broad minimum is caused by backdiffused primary electrons (about 7%) which lost energy by many collisions with atoms of the solid. The thickness of the emission layer, from which most of the secondary electrons originate, is about 50 nm for metals and 500 nm for semiconductors and insulators, in which electron scattering is much smaller as compared with metals. The secondary electron current, I_s, depends on the primary current, I_{pr}, the primary electron energy, E_{pr}, and the angle of incidence of the primary ray. It depends also on the material and surface properties of the solid. The relationship between I_s and I_{pr} is given by

$$I_s = \delta I_{pr} \quad \text{or} \quad \delta = I_s/I_{pr}, \tag{3.11}$$

were δ is the secondary emission coefficient. The values of δ in dependence on the primary electron energy E_{pr} show maxima in the E_{pr}-range of some hundred eV (see Fig. 3.3). The maximum values δ_{max} are about 1 . . . 3 for metals and 3 . . . 20 for semiconductors and insulators.

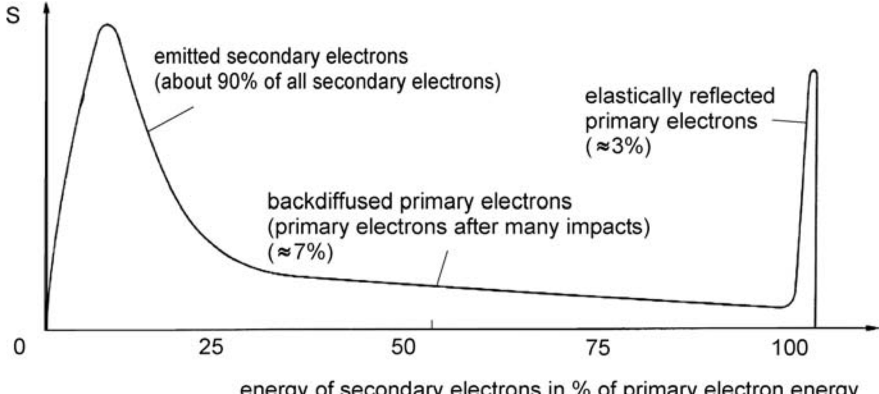


Fig. 3.2. Energy distribution of secondary electrons emitted from a solid (primary electron energy $E_{\mathrm{pr}} \approx 300 \ldots 1000$ eV)

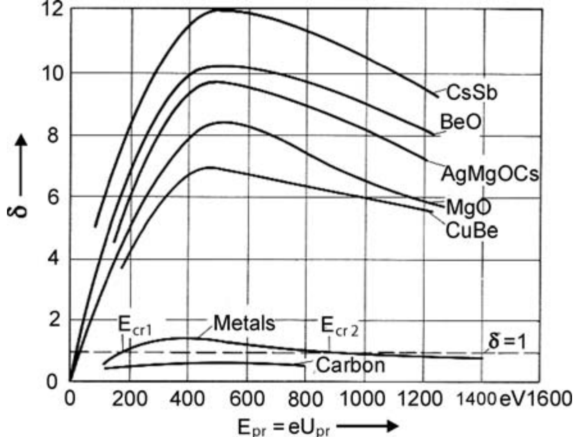


Fig. 3.3. Secondary emission coefficient, δ, as a function of the primary electron energy, E_{pr}, for different solids. $E_{\mathrm{cr}1,2}$ = crossover points for $\delta = 1$

3.2.3 Electron Optics

Electron optics deals with the quasi-optical behaviour of electron beams under the influence of electrical and magnetic fields in a vacuum system. For analogies and characteristic differences between geometrical light optics and electron optics, see [2].

Electric and magnetic electron lenses consist of electrically charged electrodes (aperture discs or cylinders) and current-carrying magnet coils, respectively, which produce electric or magnetic fields with spherically curved equipotential surfaces. Electric and magnetic deflection systems contain deflection electrodes and coils, re-

spectively, which produce flux lines perpendicular to the electron beam. Details on electron optics are discussed in [1–4].

3.3 Present State-of-the-Art and Applications

3.3.1 Secondary Electron Multipliers

Such vacuum tubes amplify a secondary electron current produced by electromagnetic or particle radiation incident on a sensitive entrance electrode. There are three types of multipliers (see Table 3.2).

Photomultipliers

Such vacuum tubes contain a photocathode and up to 12 electrodes (dynodes) for secondary electron emission. The surface of the dynodes consists of a material with high secondary emission coefficient δ. A voltage is applied to each dynode, which increases from the first to the last dynode by a constant amount. Consequently, the electron current from the photocathode is amplified at the first dynode and each of the following dynodes by a factor δ. The current amplification factor is given by

$$V_i = I_\mathrm{a}/I_\mathrm{o} = \delta^n, \tag{3.12}$$

where I_a is the anode current (current from the last electrode), I_o is the current from the photocathode, δ is the secondary emission coefficient of each dynode and n is the number of dynodes. In technical photomultiplier tubes, there is a loss of electrons between the photocathode and the first dynode (loss factor f) and between the dynodes (loss factor g). Considering these losses, (3.12) is to be replaced by

$$V_i = f(g\delta)^n. \tag{3.13}$$

Normally, the factor f amounts to about 0.9 and g to about 0.98. With

$$\delta = AU^a \tag{3.14}$$

Table 3.2. Types of secondary electron multipliers

Name	System	Type of radiation at the entrance
Photomultiplier	Photocathode + Electron multiplier system	IR-, UV- or visible light
Channel amplifier	Channel with resistive wall layer	α-, β-, γ-radiation, X- and UV-rays, fast ions and electrons
Scintillation counter	Scintillator + Photocathode + Electron multiplier system	α-, β-, γ- and X-rays

Table 3.3. Photocathode material, window material and spectral sensitivity range of modern photomultiplier tubes

Cathode material	Window material	Spectral sensitivity range/nm
Ag-O-Cs	Boron silicate glass	400 . . . 1200
CsSb	Boron silicate glass	185 . . . 650
Multialkali (Na-K-Sb-Cs)	Silica glass	185 . . . 930
CsTe	MgF_2	115 . . . 320

Equation (3.13) becomes

$$V_i = KU^{an}, \tag{3.15}$$

where A is a constant, K is a constant (containing f and g), U is the stage voltage defined as the voltage between two successive dynodes. The coefficient a is normally 0.7 . . . 0.8. According to (3.15), the current amplification factor V_i varies with the 6th to 10th power of U. The total supply voltage must therefore be highly stabilized. Its harmonic content and its temperature sensitivity must be very low.

The light entrance window of photomultipliers is either located on the side (side-on type) or on the front surface (head-on type) of the glass bulb. The side-on tubes have an opaque and the head-on ones a semitransparent photocathode, respectively. The spectral sensitivity range depends on the materials of the cathode and entrance window (see Table 3.3). The limit of sensitivity of photomultipliers is determined by the anode dark current and noise. The anode dark current increases with the supply voltage and with the temperature. The equivalent noise input is defined as the input power producing a signal-to-noise ratio equal to one at the output. It increases with the dark current, the current amplification factor and the signal band width. It decreases with a rising photosensitivity of the system.

Under the influence of a magnetic field of some mT the current amplification of photomultipliers is reduced by a factor of the order of 10 or more. The effect is especially high if the magnetic field is perpendicular to the system axis. It can be avoided by suitable shielding.

The dynamic behaviour of secondary electron multipliers, when irradiated with short light impulses, is governed by the electron transit time and the anode pulse rise time. Both values depend on the multiplier design and decrease with increasing supply voltage.

In Fig. 3.4 different commonly used types of photomultiplier tubes are shown. The tubes differ especially by the forms and positions of the dynodes. The surface of the dynodes consists of a layer of CuBe, AgMg, AlMg or $SbCs_3$ with values of $\delta \gg 1$. Characteristic data are: the voltage between each couple of dynodes is 150 to 280 V, the total voltage up to 3 kV, the current amplification $10^6 \ldots 10^8$, the sensitivity 10 . . . 60 mA/lm and the dark current 1 . . . 100 nA.

For the visualization of γ-ray pictures in Gamma cameras, position sensitive photomultipliers are applied. In such vacuum tubes the anode is, for instance, divided into 16 square fields. Each field has an area of $2.6 \times 2.6\,\text{mm}^2$ and a gap distance of 0.3 mm. An electron current, emitted from a definite point of the photocathode and

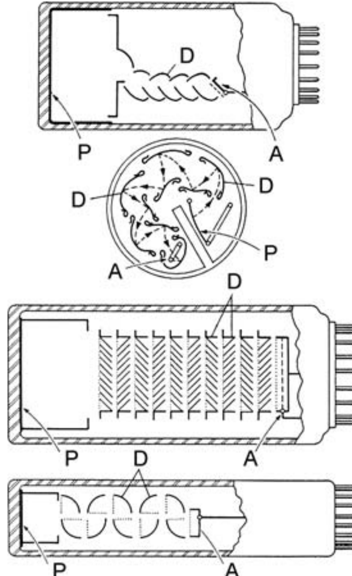


Fig. 3.4. Forms of commonly used types of photomultiplier cubes with different dynode forms. *P* is the photocathode, *A* is the anode and *D* denotes the dynodes

amplified by dynodes of the Venetian Blind type (3. tube in Fig. 3.4), hits one of the anode areas and produces an output signal. The space resolution depends primarily on the extension of the anode fields.

Channel Secondary Electron Multipliers

Channel multipliers contain – instead of single dynodes – a continuous dynode surface in the form of a resistive layer at the inner wall of a small straight, circular or spiral lead glass vacuum tube (Fig. 3.5). The length of the tube is a few centimeters and the inner diameter 1 . . . 2.5 mm. The active layer has a resistance of $10^9 \ldots 10^{11}$ Ohm and the necessary vacuum pressure within the tube is about 10^{-4} mbar. Between the two ends of the multiplier channel a direct voltage of 2 . . . 4 kV is applied (Fig. 3.6). An electromagnetic or particle radiation incident at the channel entrance causes the emission of secondary electrons from the resistive layer. The voltage drop across the layer accelerates the electrons in the direction of the outlet of the tube.

By successive impacts on the channel walls, the secondary electrons are multiplied generating an electron avalanche. The maximum current amplification is about 3×10^8. The channel tube can be used as an amplifier for currents as low as 1.6×10^{-19} A (open channel end) or as particle or quantum counter for maximum count rates of several $10^5\ \text{s}^{-1}$.

The current amplification as a function of the supply voltage of a channel multiplier is shown in Fig. 3.7. The upper limit of the amplification is caused by space charge effects. Channel multipliers show – when used as particle or quantum

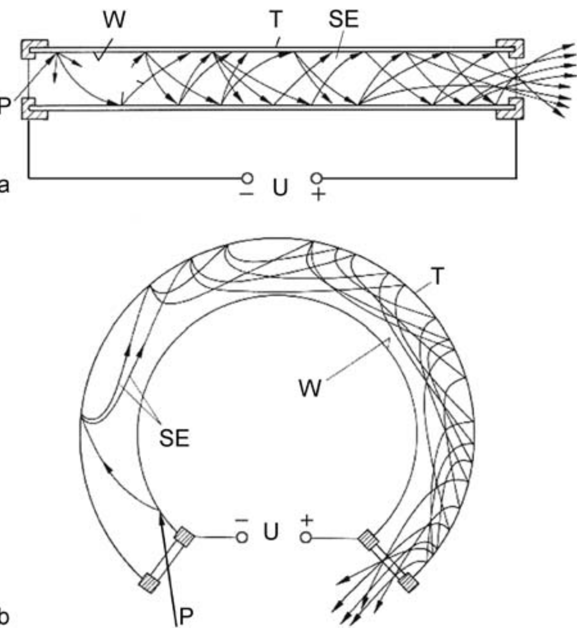


Fig. 3.5. Channel secondary electron multiplier with straight (**a**) and circular glass vacuum tube (**b**). *T* is the glass tube, *W* is the resistive wall, *P* is the incident particle or radiation, *SE* are the paths of secondary electrons and *U* is the supply voltage

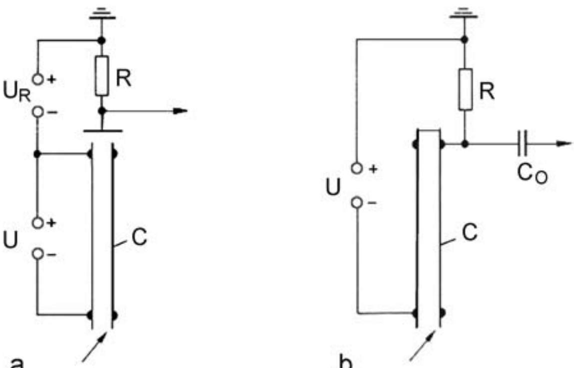


Fig. 3.6. Mode of operation of a channel multiplier with open (**a**) and closed outlet (**b**). *C* is the multiplier channel and *U* is the supply voltage

counters – a dependence of current amplification as a function of count rate. For small count rates ($1 \dots 10^3\,\mathrm{s}^{-1}$), the amplification remains constant and then decreases with rising count rate.

Channel multipliers with an inner diameter greater than 0.5 mm are bent. This prevents penetration of positive residuous gas ions to the channel entrance, where

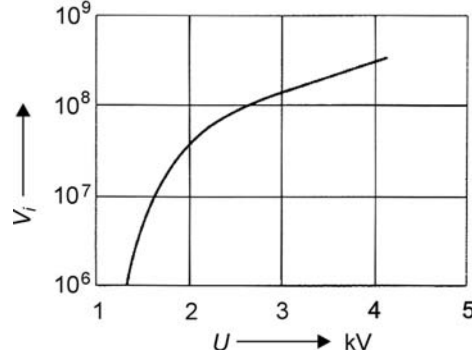


Fig. 3.7. Current amplification, V_i, as a function of the supply voltage, U, of a channel multiplier

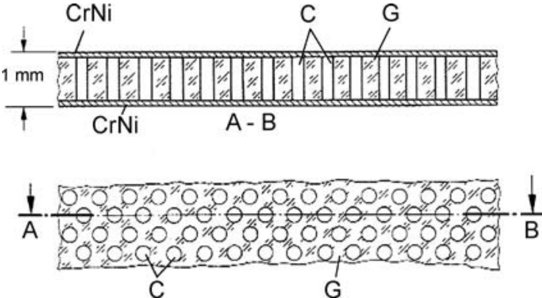


Fig. 3.8. Structure of a channel plate for the amplification of electron pictures. *G* is the glass plate, *C* is a single amplifying channel and *CrNi* is the CrNi-layer

they could release additional impulses. They are used as detectors for β-, UV- and X-rays, and also for positive and negative ions. Their impulse rise time is of the order of 5 ns, their half-amplitude pulse durations are about 10 ns and their noise count rate, for instance, $0.5\,\text{s}^{-1}$.

If a very large number of short multiplier channels (length about 1 mm) are placed in parallel and close together, a channel plate for the amplification of electron pictures is obtained (Fig. 3.8). Such secondary electron multiplier plates (diameter 2.5 . . . 10 cm) are produced by stringing, cutting and bunching of many tiny glass tubes with or without a solid core (of metal or soluble glass). The core is then chemically or electrolytically removed. Both sides of the channel plate are covered with evaporated chrome-nickel-layers. A voltage of 1 . . . 4 kV is applied between the layers. Each channel has a diameter of 25 . . . 50 µm and a wall resistance of $10^7 \ldots 10^8$ Ohm. The distance between two channel axes is 30 . . . 50 µm. The current amplification is about 10^3.

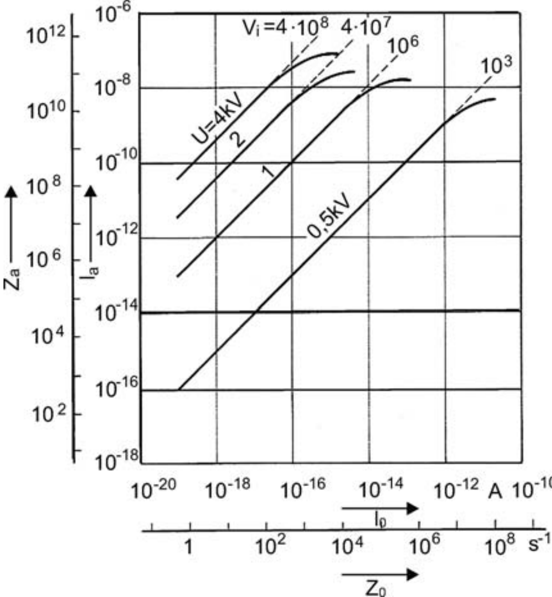


Fig. 3.9. Transfer characteristic of a channel plate (output current as a function of input current). I_0 is the input current, z_0 is the number of incident electrons per second, I_a is the output current, z_a is the number of emitted electrons per second, U is the supply voltage and V_i is the current amplification factor

The transfer characteristic of a channel plate (i.e. the output current as a function of the input current) is shown in Fig. 3.9. The structure of an ultra-fast photomultiplier with channel plate is illustrated in Fig. 3.10.

Scintillation Counters

Scintillation is defined as energy conversion of radioactive radiation into light impulses by means of a solid, liquid or gaseous medium. Commonly used inorganic scintillators are semiconductors like NaJ activated with Thallium, ZnS activated with Silver, or BGO (chemical structure: $Bi_4(GeO_4)_3$). An organic scintillator is, for instance, Anthracene. The activator atoms produce discrete energy levels in the forbidden band of the semiconductor. Particles or radiation quants of sufficient energy penetrating into the scintillation crystal cause the transition of valence electrons from the valence band to the activator energy levels. The following return of the electrons to the valence band is accompanied by the emission of photons similar to the light of a luminescent screen when irradiated with electrons. If the scintillation crystal is connected to the transparent cathode of a photomultiplier, the photons from the crystal produce photoelectron impulses from the cathode which are amplified and detected. The design of a scintillation counter with photomultiplier is shown in Fig. 3.11. The maximum pulse rate depends on the decay time of the scintillator impulses, which amounts to 1 ... some 100 ns (Table 3.4).

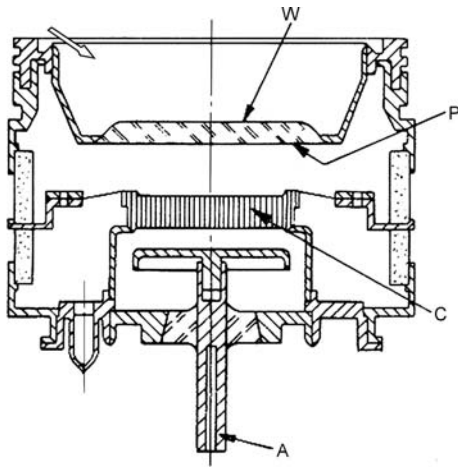


Fig. 3.10. Ultra-fast photomultiplier system with channel plate. *W* is the window, *P* is the photocathode, *C* is the channel plate and *A* is the anode

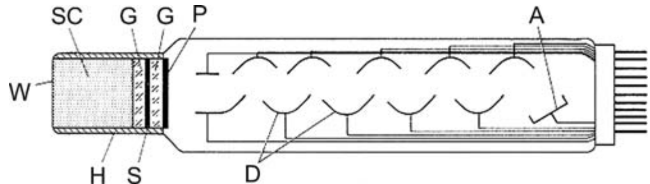


Fig. 3.11. Structure of a scintillation counter with photomultiplier. *W* is the radiation entrance window (for instance, 0.5 mm Al coated with a MgO-layer to reduce light reflection), *SC* is the scintillator crystal (for instance, NaJ(TI)), *H* is the air-tight case (because NaJ is strongly hygroscopic), *G* is the glass plate; *S* is the silicon layer, *P* is the photocathode; *D* is the dynode and *A* is the anode

Table 3.4. Parameters of different inorganic scintillation materials

	BGO $Bi_4(GeO_4)_3$	LSO $Lu_2(SiO_4)O{:}Ce$	NaJ(Tl)	BaF_2
Decay time/ ns	300	40	230	620
Absorption Length/ cm	1.12	1.14	2.56	2.06
Peak wavelength/ nm	480	420	410	310
Light yield relative to NaJ/ %	20	75	100	6

The particle or quantum energy, absorbed in the scintillator, is proportional to the amplitude of the emitted light impulse. A scintillation counter connected to a pulse height analyser can therefore be used for energy spectrometry of X-ray- and γ-quants.

3.3.2 Image Converters and Image Amplifiers

These types of vacuum tubes convert a weak light, X-ray- or γ-ray-picture into a bright visible picture. For this purpose, the light radiation containing the picture information is absorbed on the frontside of a transparent photocathode, which produces a photoelectron emission picture on its backside. The photoelectron picture is, by electron-optical means, projected to a fluorescent screen to make it visible. In the case of X- and γ-rays, the entrance of the image converter consists of a fluorescent layer (X-rays) and of a scintillator crystal (γ-rays), respectively, which convert the X- or γ-ray picture into a visible light picture on the subsequent photocathode [5].

Image Converters for Visible, Infrared or Ultraviolet Light Input Signals and for X-rays

The converters for light input signals can be divided into three different design groups (Fig. 3.12). Converter tubes of the proximity or wafer design (Fig. 3.12a) contain a plane photocathode and a plane luminescent screen at a small distance behind it (proximity structure). A short positive high voltage impulse on the screen produces a photoelectron current impulse and the corresponding short-time image of at least 1 ns duration. The projection is almost free of distortion and image inversion. The image amplification factor is about 20. Such tubes are applied as electrooptical high-speed shutters for photo cameras.

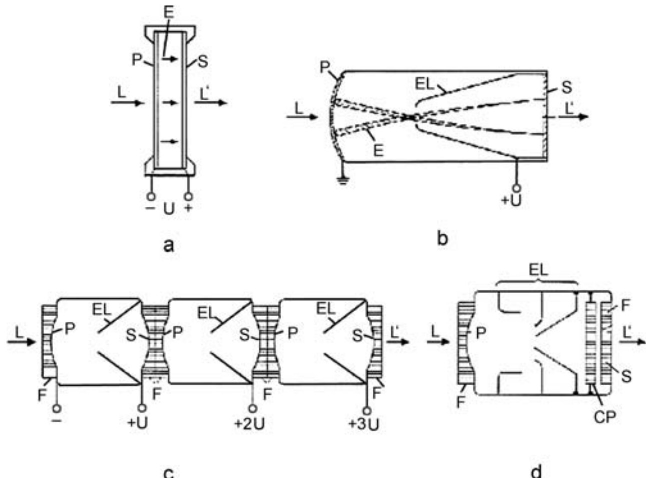


Fig. 3.12. Types of image converters for light input signals. **a** Proximity or wafer design. **b** Image amplifier with electrostatic lens. **c** Threestage image amplifier with fiberoptic coupling windows. **d** Image converter with a channel plate. *P* is the photocathode, *S* is the luminescent screen, *E* is the electron beam, *EL* is the electron lens, *F* is the fiberoptic window, *CP* is the channel plate, *L* is the incident light flux, L' is the amplified light flux and *U* is the supply voltage

In a second type of image amplifiers, an electrostatic lens is positioned between the photocathode and the screen (Fig. 3.12b) causing an image inversion on the screen. Such inverter image amplifiers have a light amplification factor of about 100 and a small image distortion. If several (normally three) image amplifier units are connected in series and coupled through fiberoptic windows to each other, a multistage image amplifier with a maximum amplification factor of 75,000 is obtained (Fig. 3.12c). A fiberoptic window consists of a very large number of fine, short and parallel positioned glass fibers which transfer the screen image of one stage to the immediately following photocathode of the next stage.

A third group of image converters belongs to the inverter type tubes containing a channel plate for additional electron multiplication between the electrostatic lens and the screen (Fig. 3.12d). The lens projects the electron image of the photocathode onto the front surface of the channel plate. The electrons leaving the channel plate are accelerated to the screen where they produce a visible bright image. The light amplification factor of such tubes depends on the channel plate voltage. Channel plates show a saturation of their transfer characteristic for increasing input currents, i.e. for stronger irradiation of the photocathode. This makes possible the amplification of images with high contrast.

The resolving power of image converters is defined as the maximum number of clearly separated line pairs per millimeter. It is limited by the fact that each point on the photocathode, when electrooptically projected on the screen, produces a tiny disc of distortion. Image converters with lens systems have a higher resolving power as compared with tubes which have no focusing lens.

Vacuum image converters are operating with a high screen voltage (maximum 15 kV) resulting in a high image brightness and small chromatic and spherical aberrations. In tubes with large diameter, the photocathode is bent to minimize scale errors. Typical parameters of converter tubes are: cathode diameter 20 ... 50 mm, cathode sensitivity $s_c = 50 \dots 200\,\mu\text{A/lm}$ at a wavelength $\lambda = 550\,\text{nm}$, light amplification factor $20 \dots 5 \times 10^4$, scale of image enlargement 0.6 ... 1.5, resolving power

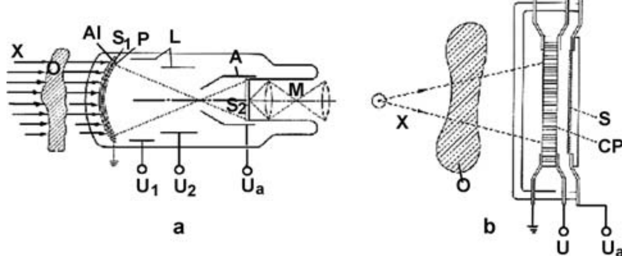


Fig. 3.13. Structures of X-ray image amplifiers. **a** Tube with one photocathode and two luminescent screens. **b** Tube with channel plate and one luminescent screen. *X* denotes X-rays, *O* is the irradiated object, *Al* is the aluminum foil, S, S_1, S_2 are the screens, *P* is the photocathode, *L* is the electron lens, *A* is the anode, *M* is the microscope, *CP* is the channel plate, U_a is the accelerating voltage and U is the voltage at the channel plate

10 ... 60 line pairs/mm and background brightness $<2 \times 10^{-7}$ Lux (caused by the dark current).

A special type of vacuum image converters are X-ray-image amplifiers (Fig. 3.13), which transform a weak X-ray image into a bright visible picture. The tube in Fig. 3.13a contains a photocathode and two luminescent screens. The first (entrance) screen S_1 is irradiated by the X-rays and produces a light picture on the adjacent photocathode P. This photocathode emits the corresponding electron picture which is projected onto the second screen S_2 by an electron lens L. The final image is transferred to a television monitor tube. In the X-ray image amplifier of Fig. 3.13b the electron-optical system consisting of the screen S_1, the photocathode P and the electron lens L is replaced by a channel plate CP in front of a luminescent screen S. The channel plate transforms the incident X-ray image into a bright visible image on the screen. Vacuum X-ray image amplifiers are to a large extent applied in medical diagnostics and in material structure analysis.

Gamma Cameras

A radioactive, γ-quants emitting isotope, when injected into the organism, takes part – like the non-radioactive isotopes of the same element – in all transport, metabolic and secretion processes of this organism. Because of its γ-radiation, the isotope can serve as an indicator or tracer, the path and concentration of which can be tracked within the organism (first application by G. von Hevešy on plants, 1913).

Within the organism the isotope traverses under the influence of metabolic processes one or more compartments (organs or part of organs), where it can at first be temporarily accumulated to form a pool before it is secreted. By detecting and measuring the corresponding variations of the organism emitted γ-radiation, the activity distribution in time and space can be analysed. Thus, information about the function and volume of the compartments and also about the flow and exchange rates between the compartments can be obtained. The pictures showing the spatial distribution of the γ-radiation within the organism are picked up with a scintillation or γ-camera (Anger camera).

The design and mode of operation of a scintillation camera is shown in Fig. 3.14. The γ-radiation emitted from the body of a patient penetrates a perforated collimator with a large number of tiny holes (for instance, diameter 39 000 holes of length 30 mm and diameter 1.6 mm; perforation pitch 0.25 mm). At the outlet of the collimator the radiation hits on a scintillation crystal with diameter of 20 ... 30 cm and thickness of 6 ... 12 mm. The scintillator can consist of a single piece of crystal (for instance NaJ) or of a large number of small crystals, each of which is shielded against the radiation of each neighbour crystal. The crystal surface opposite to the collimator is covered with a large number (about 20 ... 100) of tightly packed small photomultipliers. The locations of the scintillation points within the single crystal sensor are detected by means of a decoding matrix (resistance matrix after Anger or delay line matrix after Tanaka). In such a matrix the distance between a scintillation point and a reference point is proportional to the resistance or to the signal transit time

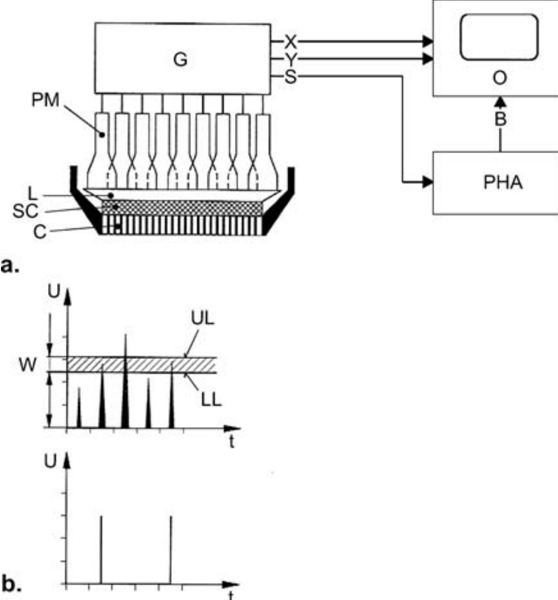


Fig. 3.14. Design and mode of operation of a scintillation camera (Gamma camera). **a** Block diagram, **b** Energy window of the pulse height analyser. *PM* is the photomultiplier array, *L* is the light conductor, *SC* is the scintillation crystal, *C* is the collimator, *G* is the generator for coordinate and sum impulses, respectively; x, y are deflection signals, *S* is the sum impulse, *O* is the oscilloscope, *B* is the brightness control, *PHA* is the pulse height analyser, *W* is the channel width, *UL*, *LL* is the upper and lower channel limit, t is the time, *U* is the signal voltage

between these points. The output signal amplitudes of the photomultipliers are proportional to the energy of the γ-quants. The influence of the γ-quantum energy can be eliminated through division of the impulse amplitudes by the amplitude of the summary impulse (Z-impulse). The Z-impulses are conveyed to the input of a pulse height analyser (two parallel analysers with different thresholds in anti-coincidence). This circuit separates within a definite energy window impulses originating from the characteristic γ-energy of a single nuclid and suppresses the Compton radiation of low energy. Therewith, the energetic separation of two γ-emitters in double nuclid studies is possible. The output impulses of the pulse height analyser modulate the brightness of the pixels on the screen. The brightness distribution corresponds to the γ-activity distribution of the body region of interest. This mapping is suitable for γ-quants within the energy range of 100 . . . 600 keV.

The scintigrams can continuously – in cardiological investigations synchronously with the electrocardiogram – be stored in a computer. In this way, for instance, changes of the function of organs in "regions of interest"can be detected and volume measurements of organs are possible.

Typical characteristics of Gamma cameras are: the sensitivity, i.e. the number of detected γ-quants divided by the number of emitted γ-quants; the resolving power, i.e. the half-amplitude pulse duration divided by the quantum energy in percent; the spatial resolution (for instance 1.8 mm); the homogeneity, i.e. the constancy of sensitivity across the scintillator surface; and the linearity and time resolution (for instance 2×10^5 impulses/s). Modern Gamma cameras contain microprocessors for homogeneity and energy correction.

3.3.3 Television Camera Tubes

This group of vacuum tubes is used for the conversion of visible light, infrared and UV images into electrical output signals (videosignals) [5].

Basic Function of Vacuum Camera Tubes

A modern camera tube consists of four principal components (see Fig. 3.15): a light sensitive converter target at the entrance of the tube, an electron beam system, a beam focusing system and a beam deflecting system.

The Converter Target

The photosensitive target of a modern vacuum camera tube is composed of a glass substrate of optical quality (window) covered with two subsequent layers: a highly transparent electrically conductive layer (signal electrode) and a photo-sensitive layer, generally consisting of several semiconducting substances. The properties of the semiconducting layers determine essentially the characteristics of the camera tube.

Three types of semiconducting targets are mainly used, namely the photoconductive resistance target, the single junction diode target and the silicon multidiode target.

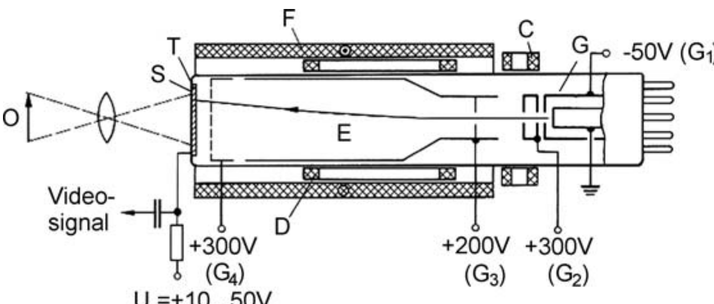


Fig. 3.15. Basic design of a Vidicon. *O* is the optical object, *S* is the signal plate, *T* is the target, *F* is the focusing coil, *D* are deflecting coils, *C* is the correction coil, *G* is the electron gun, *E* is the scanning electron beam

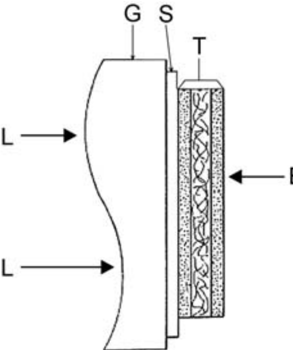


Fig. 3.16. Basic structure of the Vidicon target. *L* is the incident light flux, *G* is the glass window, *S* is the signal layer, *T* is the target layer, *E* is the scanning electron beam

The basic structure of the Vidicon target is shown in Fig. 3.16. The transparent signal electrode layer consists of indium tin oxide (ITO) or tin oxide doped with antimon ($SnO_x(Sb)$) and the target layer of three subsequent films of compact, porous and again compact semiconducting antimontrisulfide (Sb_2S_3). The films also differ in their stoichiometric composition and thickness.

In the case of single junction diode targets, the glass substrate is covered with a transparent ITO- or $SnO_x(Sb)$-signal electrode followed by several layers of amorphous or polycrytalline semiconducting materials forming a junction barrier. The semiconductor of the Plumbicon target (Fig. 3.16) is composed of subsequent n-, i- and p-conductive microcrystalline lead oxide layers with a junction barrier located in the intrinsic (i) zone. The semiconducting material of the Newvicon target is a ZnSe-film covered with In-doped ZnCdTe- and Sb_2S_3-films, forming a heterojunction within the semiconductor. The photosensitive layer of the silicon multidiode target (Fig. 3.17) is a silicon monocrystal with a light exposed and highly doped thin n^+-Si-layer followed by a normally doped n-Si-layer. On the surface of the n-Si-layer, $(6\ldots8)\times10^5$ planar diodes per cm^2 are integrated by planar diffusion. Each diode has a diameter of $6\ldots8$ μm and a mid-distance of $12\ldots15$ μm.

The different types of photosensitive targets of vacuum camera tubes convert the optical image projected onto the frontside of the target into the corresponding electrical charge distribution across the backside. In the case of the photoconductive resistance target (Vidicon target; Fig. 3.16), the photons penetrate the thin signal layer and produce a spatial conductivity distribution across the subsequent semiconducting layer. A positive bias voltage at the signal electrode causes a transport of positive charge carriers from the signal electrode through the semiconducting layer to the open surface of this layer, where they accumulate. At each point the number of flowing and accumulating charge carriers depends on the electrical conductivity at this point. Thus a definite surface charge distribution corresponding to the photon distribution of the optical image is generated on the open surface of the target. During operation of the camera tube the charge distribution is continuously scanned and thus detected by a fine electron beam produced by an electron gun within the tube.

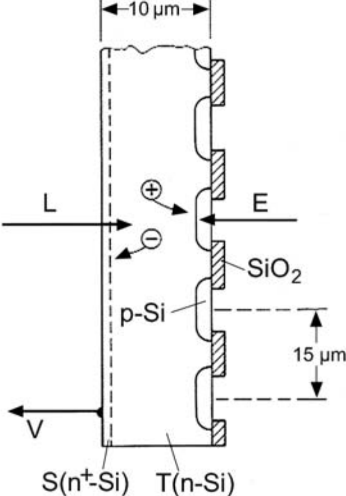


Fig. 3.17. Structure of the silicon multidiode target. *L* is the incident light flux, *S* is the signal layer, *T* is the target layer (n-Si with integrated p-Si-spots), *V* is the videosignal

In a single junction diode and multidiode target (Figs. 3.16 and 3.17) a reverse bias voltage is applied to each diode. In darkness only a very small and usually negligible reverse current is flowing over the barrier layer. Illumination of the target and consequently of the barrier layer produces electron-hole-pairs which are continuously separated by the reverse electrical field within the barrier zone. In this way the p-zone (or p-zones in the multidiode target) are positively charged. At each point of the single junction diode and in the p-zone of each diode in the multidiode target, respectively, the accumulated charge is proportional to the number of incident photons, i.e. the light intensity. This charge distribution can also be scanned by an electron beam.

The Electron Beam System

This system consists of the electron gun and the drift space. The gun can be a triode, a diode or a modified diode (Fig. 3.18).

The drift space of the electron beam system contains a cylindrical electrode connected to a grid electrode in front of and parallel to the target surface (Fig. 3.15). Both electrodes form an electrostatic lens (collimator lens) causing the electron paths to be perpendicular when the electrons hit the target surface.

The Beam Focusing and Deflecting Systems

The beam focusing and deflecting systems are magnetic coils positioned outside the vacuum tube. The focusing coil generates a homogeneous axial magnetic field along the electron beam. Electrons with additional radial velocity components are forced

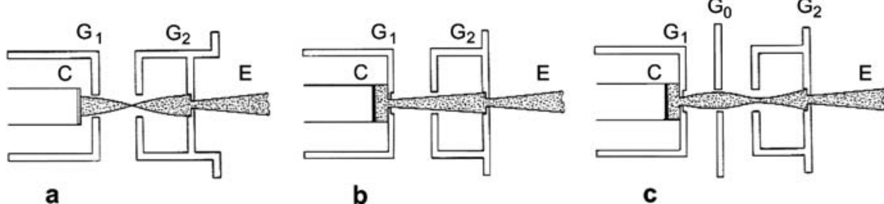


Fig. 3.18. Construction of typical electron guns for Vidicons. **a** Triode system. **b** Diode system. **c** Modified diode system. C is the thermionic cathode, G; G_1; G_2 are the beam forming electrodes, E is the electron beam

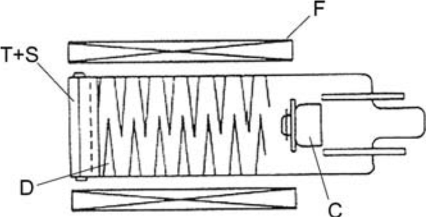


Fig. 3.19. Schematic design of a Vidicon system with magnetic beam focusing and electrostatic beam deflection. T+S is the transparent signal plate (S) with target layer (T), F is the focusing coil, D are deflecting electrodes, C is the cathode

to move on small circular paths and thus focused in one or several beam nodes. The accelerating voltage is adjusted so that the first beam node lies in the target surface.

The deflecting system consists of two pairs of coils for horizontal and vertical deflection. The coils are attached to the tube bulb and surround the bulb in the form of half-circles. The two pairs of coils have an angular displacement of 90 degrees. The application of a sweep voltage produces an alternating saw-tooth current with linear wave front and consequently two orthogonal magnetic deflecting fields. The frequencies correspond to the line and picture frequencies of the actual television standard.

The interference between the focusing and deflecting fields can be eliminated in tubes with electrostatic deflection and magnetic focusing. In this case the horizontal and vertical deflecting fields are generated by two zigzag-shaped electrode pairs (*Deflectron*, Fig. 3.19). Their forms and positions are adjusted to minimize picture defects. They are produced by evaporating a metallic layer on the inner wall of the bulb and subsequent computer-controlled laser structuring. The deflecting voltage of several hundred volts must be extremely linear and free of noise.

Mechanism of the Videosignal Production in Vacuum Camera Tubes

It has already been described above that the optical image at the entrance window of the camera tube produces an electrical charge distribution on the inner open surface of the semiconducting target layer. During operation of the tube, this charge distribution is continuously scanned with a fine and slow electron beam. The scanning

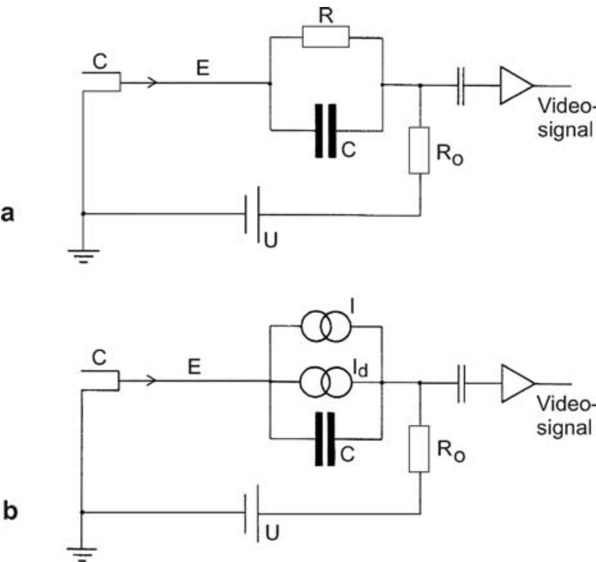


Fig. 3.20. Equivalent circuit diagram of an image point on the target. **a** Target with resistance layer. **b** Target with junction layer. C is the cathode, E is the scanning electron beam, R, C are resistance and capacitance of an image point, I, I_d are currents through an image point, U is the bias voltage across the target

process produces secondary electrons which are absorbed by the grid electrode in front of the target. Because the secondary electron coefficient is smaller than one, more primary electrons are absorbed than secondary electrons emitted. The positive charge distribution on the target surface is therefore partly or completely compensated by the scanning electron beam.

The effect of the scanning process can be described by means of the equivalent circuit of a single picture point of the target (Fig. 3.20). The diameter of each picture point is determined by the diameter of the scanning electron beam. For the photoconductive resistance target each picture point can be described as shunt connection of a high-ohm resistor R and a small capacitor C (Fig. 3.20a). In the case of the barrier junction target the equivalent circuit contains two current generators connected in parallel and a small capacitor (Fig. 3.20b). The capacitors are normally charged to the target potential (potential of the signal electrode).

During the scanning process, each picture point is hit by beam electrons within a very short time interval of about 0.1 μs. In this interval each point becomes part of the electronic circuit, in which the electron beam acts as fast switch. Without illumination the capacitors discharge to some extent between two subsequent scanning events. Through the next following scan this loss of charge is compensated. The corresponding current is the dark current of the camera tube. The voltage drop at the resistor R_o is the videosignal in darkness.

If an optical picture is projected on the (now illuminated) target, free movable electron-hole pairs are knocked off by the photons in the semiconducting target layer. Under the influence of the applied electrical field the charge carriers are separated. The free electrons move to the positive electrode (signal plate) and the holes to the negative electrode, i.e. to the open target surface which is stabilized at zero potential by the scanning electron beam. On the target surface the resulting instantaneous distribution of the positive charge density is equivalent to the light image at the tube entrance.

Between two subsequent scanning events the capacitors in Fig. 3.20 discharge. The discharge current, which is different for each image point, flows through the resistor, R, and current source, I, which are connected parallel to the capacitors, C. During the next scanning event, the charge loss of the capacitors is almost compensated. The corresponding compensation current produces a voltage drop across the resistor R_0, which is the videosignal of the tube. The series of voltage impulses with different amplitudes, generated during a frame period of the television system, contains the complete information of the television image. During the line and frame flyback, respectively, the target scanning is suppressed by negative blanking pulses on the control electrode or positive blanking pulses on the cathode of the electron gun.

Special Vacuum Electronic Camera Tubes

The *Ebsicon*: The light sensitivity of a Vidicon with silicon multidiode target can be increased by a factor of about 300 if an image amplifier is connected to the tube entrance (Fig. 3.21). In this case the target is not directly illuminated but bombarded with electrons from the image amplifier. The electron image emitted from the photocathode of the amplifier is projected on the silicon target producing a corresponding charge picture on the target surface. This type of camera tube is called Ebsicon (Ebsi = ***e***lectron ***b***ombarded ***si***licon).

The *Pyroelectric Vidicon* (Fig. 3.22): This type of camera tube is sensitive to infrared radiation. Its target is a thin dielectric disc of Triglycinsulfate (TGS) which is spontaneously polarized perpendicular to the surface. The tube entrance consists of

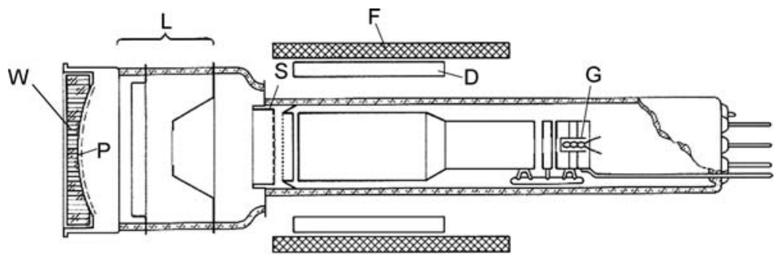


Fig. 3.21. Design of an Ebsicon (electron bombarded Si-Vidicon). *W* is the fiberoptic window, *P* is the photocathode, *L* is the electron lens system, *S* is the silicon target, *F* is the focusing coil, *D* are deflecting coils, *G* is the electron gun

a Germanium window. Incident IR-radiation, varying with time, changes the spatial temperature distribution and hence the polarization of the pyroelectric target layer. Consequently, the surface charge distribution on the target backside is also changed, whereas the illuminated frontside is kept on the constant positive potential of the adjacent signal plate. As the polarization is influenced only by variations of the IR-radiation, the radiation must be modulated through an optical chopper or by periodic movements of the camera with the result that the constant background radiation of an object has no effect and the weak contrast of a thermal scene is enhanced.

The scanned surface of the TGS-layer can have a positive or negative charge surface depending on the changes of polarization. For the electron beam scanning, it is necessary to increase the surface potential by a constant positive amount. For this purpose, the scanned surface is charged continuously with positive ions, which are generated by impacts of beam electrons on residual gas molecules near the target surface. The necessary gas pressure of Helium or Hydrogen gas is about 10^{-3} mbar. Another possibility is the production of positive surface charges by secondary emission or making the TGS-layer weakly conductive. The scanning beam diminishes the surface charge until each image point attains the cathode potential. The corresponding potential jumps yield the videosignal at the signal plate. The spectral sensitivity curve of a pyroelectric camera tube is shown in Fig. 3.23.

The optical resolution of an IR-camera tube is restricted by the thermal conductivity of the target, because the temperature profile produced by the IR-radiation is rapidly dissolving. An increasing modulation frequency causes a rising spatial resolution, but a decrease of the videosignal. For improvement of the image resolution targets with surface channel graticules are used. The channel distances are about 25 μm. The channel structure is made by ion etching with photomask technique.

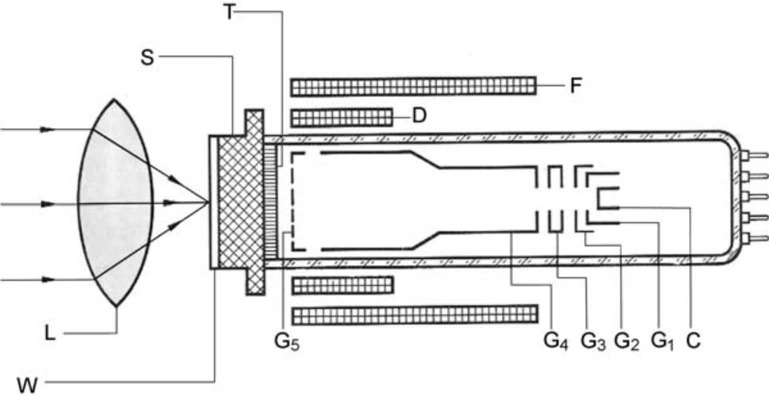


Fig. 3.22. Schematic design of a pyroelectric camera tube. *L* is the infrared lens, *W* is the IR-light window, *S* is the signal electrode, *T* is the pyroelectric target, *F* is the focusing coil, *D* are deflecting coils, *C* is the cathode, $G_{1...5}$ are the beam forming electrodes

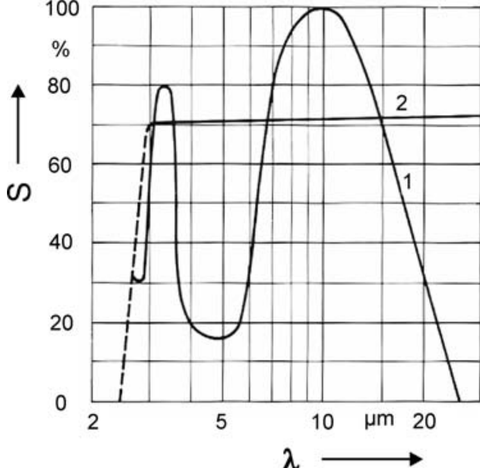


Fig. 3.23. Spectral sensitivity of pyroelectric Vidicons with different IR-entrance windows. *1* Germanium window. *2* Special IRTRAN-window. *S* is the spectral sensitivity (maximum = 100%), λ is the IR-wavelength

Geiger–Mueller Counter

Such tubes are used for intensity measurements of radioactive radiation. They consist of a glass or metal tube containing a cylindrical or spiral cathode and a straight axial metal filament as anode (Fig. 3.24). The tube is filled with gas of about 100 mbar. The gas consists of air, hydrogen or noble gases, with additional components of organic or inorganic vapours.

With increasing anode voltage U_a the following discharge processes take place in the counter tube (Fig. 3.25):

Low U_a (up to about 300 V): No formation of charge carrier avalanches (*Ionization Chamber Mode*).

Medium U_a (up to about 500 V): Formation of charge carrier avalanches on the path of incident radiation (spot of primary ionization); the discharge current is proportional to U_a and to the energy of the incident radiation (*Proportional Mode*);

High U_a (up to about 800 V): By intensive production of photons each incident particle or quant causes a “transverse ignition” of a discharge along the whole anode wire. The discharge current is not dependent on the energy of the particles or quants (*Geiger–Mueller Mode*). A further increase of U_a leads to a continuous glow discharge.

A quick extinction of each single discharge impulse can be achieved by using a RC-circuit for the tube operation (with large R and small C; see Fig. 3.24) and by addition of vapour components to the fill-gas, which cause a high absorption of photons (*self-quenching counting tube*). Because of the slow decay of the ion space charge GM-tubes show a definite *dead-time* after each discharge impulse

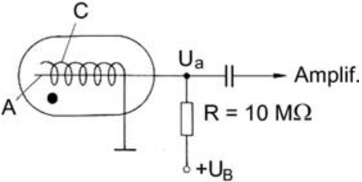


Fig. 3.24. Design and principal circuit of a Geiger–Mueller counter tube. *C* is the cathode (helix), *A* is the anode (straight wire), *U* is the supply voltage

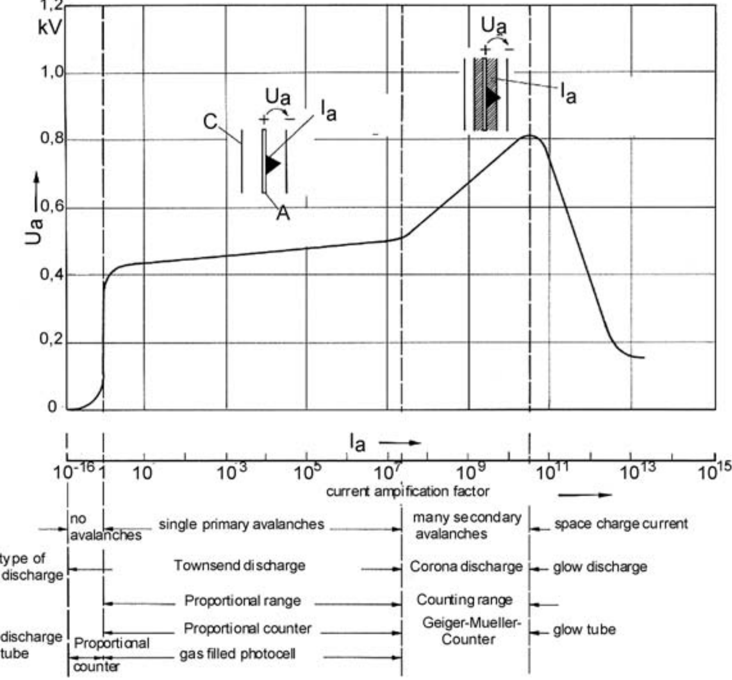


Fig. 3.25. Voltage-current-characteristic (U–I characteristic) of a Geiger–Mueller counter tube. Under the *abscissa* the different ranges of operation are shown

(0.01 ... 1 ms; no response), followed by a *recovery time* (0.1 ... 1 ms; weak response).

The pulse rate characteristic of a GM-tube, i.e. the pulse rate as a function of the anode voltage U_a, is shown in Fig. 3.26. The height of the plateau in the characteristic is proportional to the incident ion dose rate j.

Ionization Chamber

Ionization chambers are sensors for ion dose rate measurements. The glass or metal chamber contains two plane or cylindrical electrodes and is filled with pure air or

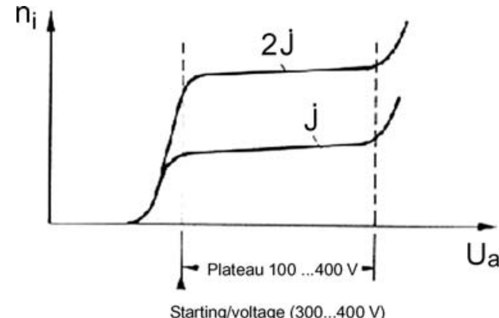


Fig. 3.26. Pulse rate characteristic of a Geiger–Mueller counter tube. n_i is the number of discharge impulses per second, U is the anode voltage, j is the ion dose rate

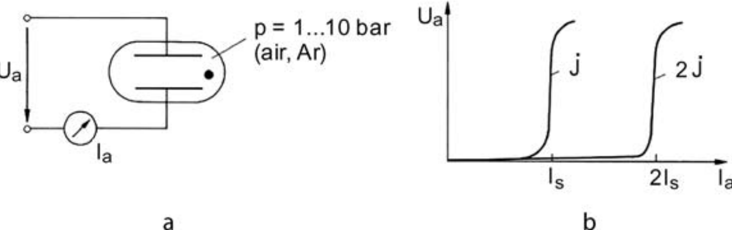


Fig. 3.27. Circuit (**a**) and U–I characteristic (**b**) of an ionization chamber. j is the ion dose rate, I is the saturation current

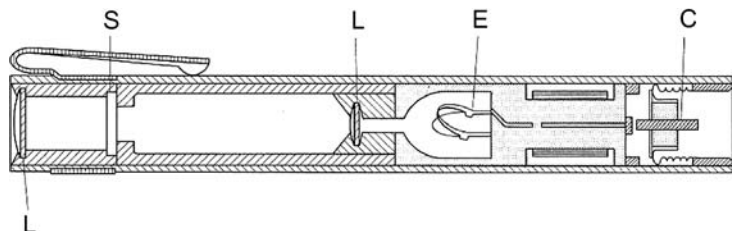


Fig. 3.28. Cross-section of a pocket dosimeter for radiology and nuclear technology. E is the quartz filament electrometer, L is the microscope lens, S is the scale, C is the contact pin for recharge

argon gas at atmospheric or a higher pressure. Therefore, strictly speaking, it does not belong to the group of vacuum tubes, but its production needs vacuum technology. The principal circuit of operation and the characteristic are shown in Fig. 3.27. The ion dose rate is defined as the charge of ions of one polarity produced in air in 1 s by the incident radiation. The saturation current (order of magnitude several pA) flowing between the electrodes is proportional to the ion dose rate. Ionization chambers are widely used as pocket dosimeters in radiology and nuclear engineering.

The dosimeter in Fig. 3.28 contains an electrically charged quartz filament, which is continuously discharged by the incident radiation. Such dosimeters have measur-

ing ranges between 5×10^{-5} and 0.15 C/kg (0.2 ... 600R; 1 R = 1 Roentgen = 2.58×10^{-4} C/kg).

3.4 Future Aspects of Radiation Sensitive Vacuum Electronic Components

Photomultiplier tubes, channel amplifiers and scintillation counters have – as compared with the corresponding semiconductor components – a very high light or radiation sensitivity, a large signal amplification factor and a fast response. Consequently, they are and will be used in a wide field of applications, e.g. for the detection of weak and short radiation impulses in physics and medicine.

A new development for future applications is the flat panel photomultiplier tube with an effective sensitive area of about 50×50 mm^2 and a thickness of only 28 mm. When used in a matrix arrangement, the dead space between the tubes is extremely small, and the effective area becomes almost 90%. Each tube contains 12 dynodes and a 8×8 multianode matrix structure for spatial resolution detection. The anode pixel size is 5.8×5.8 mm^2 and the spectral response range 300 ... 650 nm. Tubes of this type are, for example, used in Gamma cameras, Cherenkov counters and mammography units.

For photon counting in the near infrared (NIR) region (wavelength 950 ... 1400 nm), thermoelectrically cooled photomultiplier modules with integrated controller and vacuum pump have been developed. The tube has a fast time response (rise time 900 ps) and a high gain (about 10^6). Typical applications are photoluminescence, Raman spectrometry, cathodoluminescence, fluorescence and LIDAR.

Present and future applications of photo- and secondary electron multiplier tubes, described in this section, are, for example, spectrophotometers for UV-, IR- and visible light, spectrophotometers for atomic absorption, photoelectric emission, fluorescence or Raman scattering measurements for quantitative analysis of elements contained in a sample. Other spectrophotometric applications are liquid or gas chromatography, X-ray diffractometry, X-ray fluorescence analysis and electron microscopy. In semiconductor industry, narrow scanning beams of electrons, ions, light or X-rays are used to prove the quality of semiconductor surface structures, the secondary particles or rays being measured with electron multipliers or microchannel plates. For pollution monitoring, photomultipliers are used in atmospheric dust counters, turbulence meters for liquids, NO_x- and SO_x-meters. In biotechnology, electron multiplier tubes are, for instance, used in flow cytometers for biological cell counting and in DNA sequencers. In medicine, various types of scintillation counters are found in gamma cameras, positron computer tomography and liquid scintillation counting.

During the last decade image converters and television camera tubes have been partly and continuously replaced by equivalent semiconductor components. Vacuum tubes for these applications have the advantage of high sensitivity, high resolution, fast response and insensitivity against radioactive and cosmic radiation. Such properties are important in special fields like X-ray, infrared and gamma ray imaging. Consequently, it can be expected that vacuum image converter and television camera tubes will also play a considerable role in the near future.

References

[1] J. Bretting, *Technische Röhren* (Hüthig, Heidelberg, 1991)
[2] J. Eichmeier, *Moderne Vakuumelektronik* (Springer, Berlin, Heidelberg, New York, 1981)
[3] J. Eichmeier, H. Heynisch, *Handbuch der Vakuumelektronik* (Oldenbourg, München, Wien, 1989)
[4] Hamamatsu Report, *Photomultiplier Tubes* (2002)
[5] R. Steinbrecher, *Bildverarbeitung in der Praxis* (Oldenbourg, München, Wien, 1993)

4

Electron Beam Devices for Materials Processing and Analysis

H. Bluhm, B. Han, A.G. Chmielewski, D. von Dobeneck, U. Gohs, J. Gstöttner, G. Mattausch, H. Morgner, H.W.P. Koops, A. Reichmann, O. Röder, S.W. Schulz, B. Wenzel, and O. Zywitzki

4.1 Introduction and History

Electron beams have a manifold spectrum of applications in the world-wide high-tech industry. They make possible specialized work, control and production in imaging, analysis, spectroscopy, surface structurisation by lithography, deposition and beam induced processes, plasma-activated high-rate Physical Vapor Deposition PVD, welding, non-thermal materials refining, thermal surface treatment, directed vapor deposition, curing and disinfecting of goods. This all started with the practical use of the former "Kanalstrahlen" which was discovered by von Pirani [1] in 1907, when he found refractory metals at his anode molten in a gas discharge tube, see Fig. 4.1.

Imaging and materials analysis and treatment with electrons became practical by the theoretical work on electron optics of E. Riecke [2], who investigated the motion of an electrical particle in a uniform magnetic field and in an electrical glow discharge. F. Braun worked in 1897 on a method for demonstration and study of the waveform of varying currents, when developing the cathode ray oscilloscope. W. Wiechert [3] did experimental investigations of the velocity and the magnetic deflection of cathode rays in 1898. A first application of thermionic cathodes in a Braun tube was published by S. Samson in 1918 [4]. This development of the oscilloscope was a main subject in the 1920s, and Gabor used it to record traveling waves [5]. H. Busch did the calculation of the trajectory of cathode rays in axially symmetrical electromagnetic fields in 1927 and developed the formula for the focal length of a magnetic lens [6]. Methods were developed for the design of cathode ray tubes, e.g. by E. Ruska and M. Knoll [7], which led to the understanding of geometrical electron optics [8]. Using geometrical electron optical imaging of thermionic cathodes with the aid of magnetic fields by M. Knoll, F.G. Houtermans and W. Schulze in 1932

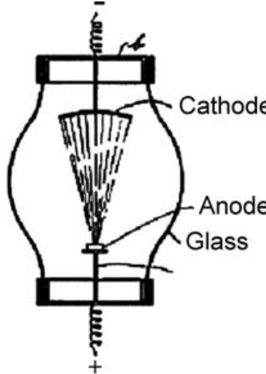


Fig. 4.1. Marcello von Pirani found molten refractory metals at the anode of his gas discharge setup and invented electron beam refining of metals

[9] led to the development of the electron microscope in 1932 [10, 11] and in subsequent patents and publications to the establishment of a new technique in materials analysis, chemistry, biology and medicine. At that time the microscope was called "Übermikroskop" for the fact that at magnifications $>1000\times$ a resolution surpassing that of the optical microscope was achieved.

Brüche [12] and Scherzer, and also Johannson [13] investigated the electrostatic electron optics and founded the theory on geometrical electron optics and the electron microscope. E.W. Müller showed imaging of atomic resolution in his field emission microscope, which he invented and theoretically explained in 1936 [14]. A new type of electron microscope, the scanning electron microscope, was invented by M. Knoll [15] in 1935. Its principle and theory [16] was described by M. von Ardenne in 1937. M. von Ardenne and R. Rühle described in 1939 the application of intense electron beams for drilling of small holes and evaporation of materials in vacuum [17]. Investigating electron sources at the AEG-Zeiss research laboratory in Mosbach, Germany, K.H. Steigerwald [18] invented in 1953 the electron-beam drilling and welding process using a long focal Wehnelt design. He was followed in this development of e-beam welding by J.A. Stohr [19] in 1958, who applied the technology in nucleonics. The use of intense electron beams was stimulated in the mid 1950s for development of new materials and new technological processes like welding, melting, evaporation, for space engineering and developments for the car industry, air plane industry and nuclear power plants. The technology development was strongly supported by the efficient metal vacuum systems which were developed after the Second World War. This allowed to develop techniques and equipment for e-beam welding, drilling, melting, materials refining and e-beam evaporation, the techniques which have today a secure position as industrial production processes. In recent years the electron beam technologies were expanded to radiation treatment of natural products, water, plastics and coatings for disinfection and materials and surface hardening. The semiconductor industry now follows the Moore Law of shrinking dimensions of transistors in electronic devices, and therefore relies very

much on the advanced technologies of scanning electron microscopy and electron-beam lithography. Those developments were described and since 1960 annually reported in international conferences, like "Symposium on Electron Beam Processes", which became later "Electron, Ion Beam Science, and Technology", first run by R. Bakish [20]. In 1973 it was renamed as "The International Conference on Electron, Ion and Photon Beam Technology". In 1995 the meeting became "The International Conference on Electron, Ion and Photon Technology and Nanofabrication", to more accurately reflect the growing importance of nanofabrication. EIPBN meets for the 50th time in 2007. Its proceedings are published annually in the Journal of Vacuum Science and Technology (JVST) and JVSTB. This series covers extensively the development of electron beam technologies in many aspects. An excellent work on the industrial use of electrons was written by Schiller Heisig and Panzer in 1977 and published in English in 1982 [21, 22]. Tables on all related questions are published by M. von Ardenne in 1961 and 1972 [23].

4.1.1 Electron Optics

History

For optical microscopy E. Abbe [24] (1873) and Helmholtz [25] (1874) calculated the resolution limit $\delta = 0.61\lambda/n\alpha > \lambda/2$, where λ is the wavelength and $n\alpha$ is the numerical aperture of the objective lens. In the visible light, $d < 200$ nm is reached. The resolution could be improved using light of shorter wavelength (ultraviolet) and quartz lenses, or immersion liquids to enlarge n, the refractive index between lens and sample.

L. de Broglie [26] (1924) postulated for the wavelength of the electron: $\lambda = h/p = h/(2em_oU(1+eU/2m_oc^2))$, which is approximated for non-relativistic electrons ($U < 60$ kV) by $\lambda = sqrt(150/U)$ in AE/V1/2 (1 AE = 1Δ = 0.1 nm), or $\lambda \approx 1.2/\sqrt{U}$ in nm/V$^{1/2}$. Here, h is the Planck constant, $p = mv + eA$ is the canonic impulse with v electron velocity and A magnetic vector potential, U is the accelerating voltage, e is the electronic charge and m is the electron mass.

Deflection of Electron Beams in Electrostatic or Magnetic Fields

The deflection angle Θ is given by $\tan\Theta = elE/mv^2 = lE/(2U_b)$; l in m, E in V/m, and U_b in V. The deflection sensitivity is given by $A/U_p = lL/(2aU_b)$, with A, the deflection amplitude at the end of distance L (distance between the middle of the deflection plates and the screen), and U_p, the voltage at the deflection plates.

Deflection of a beam in a magnetic field perpendicular to the paper plane is shown in Fig. 4.2, right. The deflection angle Θ is given by $\sin\Theta = elB/(mv) = 2.97 \times 10^5 lB/U_b^{1/2})$. The deflection sensitivity using air coils is $A/I = 0.24walL/(b^2U_b^{1/2})$. Here, w is the number of windings of the deflection coils, I is the current to the two in series connected coils, B is the diagonal distance of the conductors parallel to the axis a is the height of the coils and l is the length of the coils in axial direction. Both deflection suffer from distortion $\sim\Theta^3$, deflection coma $\sim\alpha^2\Theta$, and deflection astigmatism $\sim\alpha\Theta^2$, where α is the beam aperture.

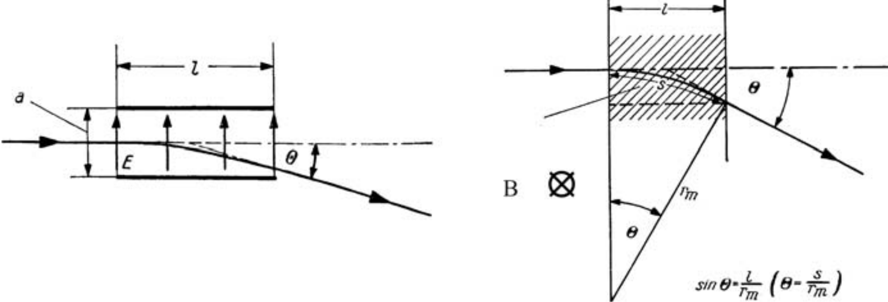


Fig. 4.2. *Left*: Deflection of a beam in a homogeneous electrostatic deflection field. *Right*: deflection of a beam in a homogeneous magnetic deflection field, perpendicular to the paper plane

Numerical Electron Optics

In the early years of electron optics, computing the electron optical properties of field and lenses was done by hand using the Laplace differential equation which gives the relation between potential distributions and the border conditions $\Delta\phi = 0$, with $\Delta\phi$ the Laplace operator of potential ϕ. The differential equation for the motion of electrons was solved using the Runge–Kutta method and introducing two fundamental rays.

Glaser describes in his book this procedure including analytic expressions for the lens aberrations [27]. More accurate results are obtained by numerical field computation and solution of the differential equation (Liebmann's mesh routine [28], Finite element method, MEBS [29] by Munroe, finite difference method or other numerical approaches like surface charge method by Rheeds [30], SPOC Software for particle optics based on 2D first order finite element method [31], see also [32]). For an estimation of the action of lenses and their aberrations, the paper of Liebmann and Grad is well suited. Mulvey describes the dimensioning of the iron circuit [33]. The paper shows that, especially when lenses for very high electron acceleration voltages are built, the lens core must be designed conically to avoid saturation and parasitic lenses in the core region.

Image Forming Properties of Magnetic Lenses

In modern electron microscopes magnetic lenses are in nearly all cases used as image forming elements. The basic configuration is as follows. A current that runs in iron-shrouded coils generates a magnetic field distribution in the gap of the shroud that can be specially formed by pole pieces. The field is rotationally symmetric about the optical axis z, and the field distribution $B(z)$ along this axis approximates very often a bell-shaped curve. The excitation of such lenses is conveniently described by the parameter $NI/\sqrt{U^*}$, where NI are the ampere turns and U^* the relativistically corrected accelerating voltage of the electron beam. In order to determine the

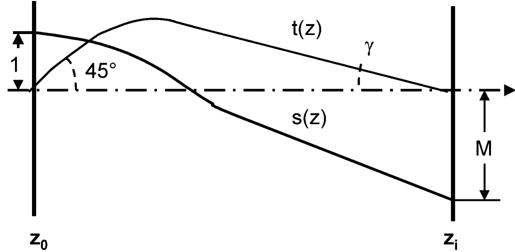


Fig. 4.3. Fundamental rays used to calculate aberration

imaging properties of a lens, one has to solve the Lorentz equation (second order differential equation) of electron motion. The ray path is determined, when, e.g. the starting point in the object plane $z = z_0$ and the slope are given. If one has two independent solutions, e.g. $s(z)$ and $t(z)$ with $s(z_0) = 1$, $s'(z_0) = 0$, and $f(z_0) = 0$, $f'(z_0) = 1$, all other solutions can be found by superposition. From these trajectories the image plane and magnification of the lens can easily be attained, as is illustrated in Fig. 4.3. Fundamental rays are used to calculate aberration. The focal length of a weak lens, where the focal length is large compared to the extent of the lens field in z-direction, is given by $1/f_\text{o} = e \int B^2(z)\,\mathrm{d}z/(8mU^*)$ with e = electron charge and m = electron mass. The image rotation which occurs in magnetic lenses is given by $\Phi_0 = \sqrt{(e/8mU^*)} \int B(z)\,\mathrm{d}z$. The rotation can be reversed by reversing the excitation of the lens. The focal length, focal position and principal planes are found, when trajectories with $s_-(-\infty) = 1$, $s'_-(-\infty) = 0$ and $s + (+\infty) = 1$, $s|(+\infty) = 0$ are chosen, as shown in Fig. 4.3.

The aberrations that occur in electron lenses can be classified as geometrical aberrations, chromatic aberrations, and diffraction error. Geometrical and chromatic aberrations are determined from the ray trajectories by calculating the deviation of the aberrated ray trajectory from an ideal one without aberrations. In the third order approximation of this calculation there are 8 geometrical aberrations: spherical aberration, isotropic and anisotropic astigmatism, field curvature, isotropic and anisotropic distortion, isotropic and anisotropic coma. If there is not good rotational symmetry of the lens field, we have in addition to these also axial astigmatism; if there is defocusing, we have to consider this, too. As to chromatic aberration, one distinguishes between axial chromatic aberration, i.e. the shift of the image plane due to the chromatic error, and the chromatic aberrations of magnification and image rotation, which occur on off-axis points.

To illustrate the design of a high resolution lens, Fig. 4.4 shows the cross-section of a top entry stage lens, as used in the ELMISKOP® 102. The specimen is introduced through the 9 mm diameter upper bore into the gap. The stage mounts on the upper pole piece, the specimen in the cartridge is coupled with the stage by cone fitting. High mechanical stability can be achieved by providing good coupling between the upper and lower pole piece and between the upper pole piece and the stage. The coils are beneath the pole piece, so that the space above the lens is kept free for additional equipment. One of the advantages of such a lens is the rotationally symmetric

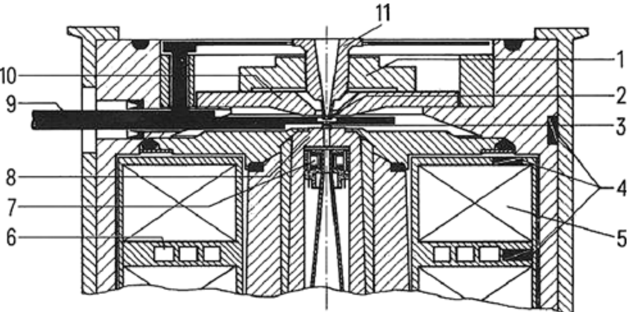


Fig. 4.4. Schematic cross-section of the objective lens of the ELMISKOP 102: *1* specimen stage, *2* specimen, *3* aperture slide, *4* thermistors, *5* objective coil, *6* cooling water chambers, *7* objective stigmator, *8* bottom objective pole piece section (pole piece), *9* cooling rod, *10* top objective pole piece section (pole plate), *11* specimen cartridge in the working position

design, providing very good mechanical and thermal stability. It is demonstrated by the fact that with this lens a line resolution of 0.7 A was attained.

Since the focal length and also the spherical and chromatic aberrations decrease with increasing excitation, it is interesting to find out what happens at very high excitations. Figure 4.5 shows the ray paths within a lens field at three different excitation values for parallel illumination. Path I is the case used in conventional objective lenses: the optical axis is crossed only once; the object is located in the first half of the field. Path II represents the telecentric mode, which is used in the condenser–objective lens according to Riecke and Ruska. Now, the object is in the centre of the field, the first half of the field acts as a condenser, the second half as objective lens, thus reducing focal length and aberrations. With this kind of lens, aberration coefficients below 1 mm can be obtained at the expense of a very small gap (=3 mm). A disadvantage of the lens for routine use in CTEM is the fact that only small areas of the specimen can be illuminated what is very inconvenient at low and medium magnifications. Path III illustrates the case of an excitation of $NI/\sqrt{U} = 28$. Now the ray path crosses the optical axis twice, and if one arranges the object in the second half of the lens field for CTEM, the focal length and aberration coefficients of the objective part are further reduced, while nearly two thirds of the lens field act as condenser system for parallel illumination (second zone mode). Thus the disadvantages of the telecentric mode are avoided. The most important point, however, is that this lens can combine low aberrations with a large gap, and therefore it is very well suited for analytical attachments and STEM.

Magnetic lenses show image rotation, which depends on the lens excitation. So, at magnification changes the final image tends to rotate. Since the direction of rotation reverses with reversed excitation, it is possible to compensate this effect, in some degree, by proper combination of excitation changes in several lenses. More important then compensating this effect is the compensation of some geometrical and chromatic aberrations that cannot be neglected at low and medium magnifications, e.g. isotropic distortion, chromatic aberration of magnification and rotation.

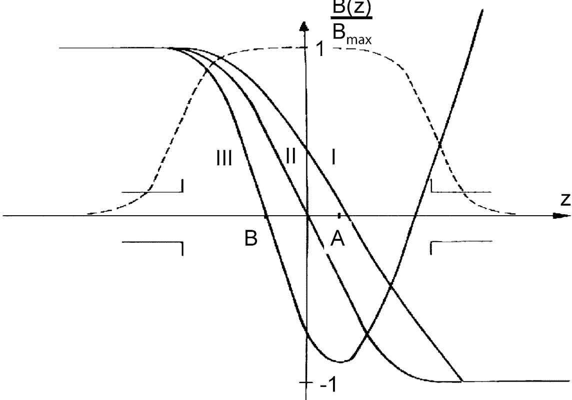


Fig. 4.5. Ray trajectories for parallel illumination for different lens excitations. Part *I*: $NI/\sqrt{U} = 14$, conventional mode; Part *II*: $NI//\sqrt{U} = 19$, telecentric mode; Part *III*: $NI/\sqrt{U} = 28$, second zone mode; *A*: Object position for CTEM; *B*: Object position for STEM

Isotropic distortion is known also as pin cushion and barrel distortion. Chromatic aberrations of magnification and rotation must be considered when examining thicker objects with energy losses at lower magnification. By proper combination of imaging lenses, distortion and other aberrations can partly be compensated because of two facts:

1. The coefficients of geometrical aberrations are excitation dependent and change sign at certain excitation values.
2. By reversing the lens polarity some aberration coefficients change sign. When knowing this, one can optimize a lens system with respect to the desired compensations by computer.

Electrostatic Refractive Index, Geometrical Optics Representation

An electron beam experiences a refraction passing through a potential jump $\Delta\phi = \phi_2 - \phi_1$. This is calculated using the power law $\sin\alpha_1/\sin\alpha_2 = v_2/v_1 = n_2/n_1 =$ sqrt$(1 + \Delta\phi/\phi_1)$, where v is the electron velocity and n is the refractive index. For $\Delta\phi > 0$ it is $\alpha_2 < \alpha_1$, which is a refraction to the normal.

This finding led Scherzer to state that in rotationally time invariant and charge free lenses the spherical aberration is positive definite [34]. Figure 4.6 shows an electrostatic “Einzel” lens having three electrodes, left and right ground and in the middle the high tension potential.

Lens Aberrations

All electron optical systems use apertures to confine the electron beam. At the edges of such apertures electron diffraction happens, and , since most fine electron sources

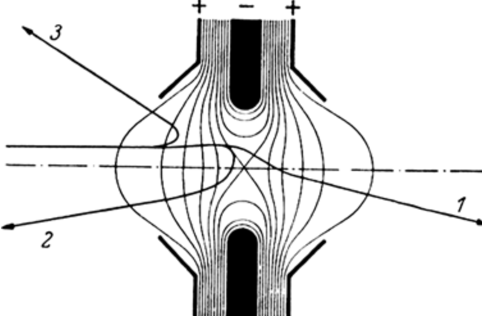


Fig. 4.6. Electrostatic “Einzel” lens having three electrodes, left and right ground and in the middle the high tension potential U_L [9]; *1* imaging ray, *2* and *3* reflected rays

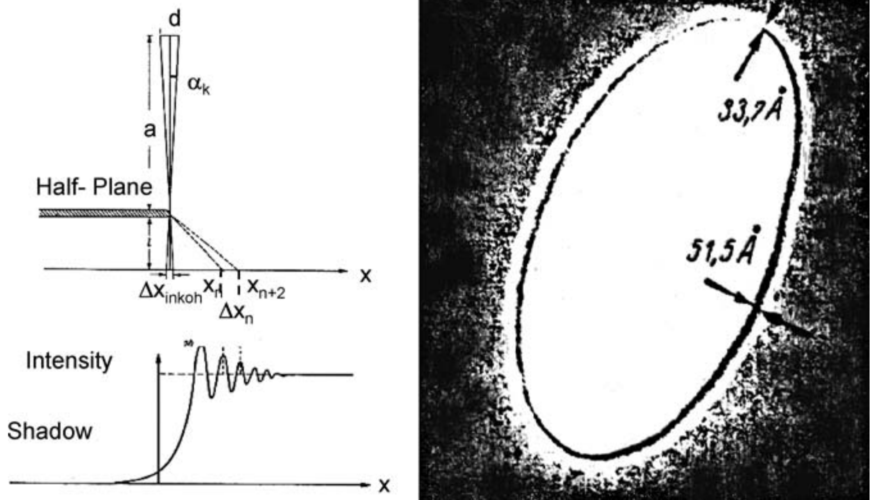


Fig. 4.7. Electron diffraction and interference at the edge of a semi-plane explains the appearance of Fresnel diffraction rings at a holy carbon foil

show a certain degree of length coherence, diffraction rings can be observed. Figure 4.7 explains the generation of the diffraction rings (left) and shows an astigmatic ring pattern observed in an TEM on a holy foil (right).

Real lenses suffer from aberrations due to principle electron optical characteristics, like the diffraction aberration, spherical and chromatic aberration. Others are generated by fabrication imperfections, like astigmatism, coma, and defocus. Figure 4.8 (left) presents such aberrations. Correction of the astigmatism is reached using two superimposed electrostatic or magnetic quadrupoles, rotated by 45° with respect to each other, to adjust for amplitude and angle of the correcting lenses. Figure 4.8 shows schematic of the ray path and an electrostatic element used to correct astigmatism. It is placed close to the lens to avoid combination aberrations

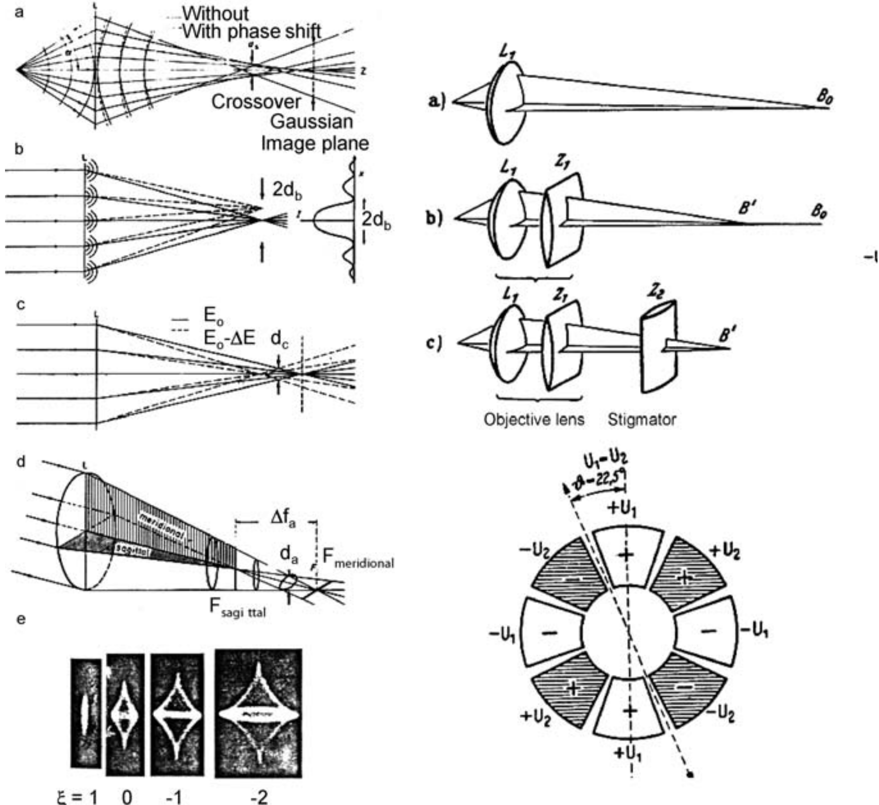


Fig. 4.8. *Left*: principal **a** (spherical), **b** (diffraction), and imperfection aberrations **c** (chromatic) and **d** (astigmatism); **e** gives caustic cross sections of an astigmatic bundle at different defocus [1]. *Right top*: schematic of the astigmatic ray path and correction of the astigmatism, and *right bottom*: an electrostatic element used to correct astigmatism

between the two lenses. The radius of least confusion is $r_{\mathrm{A}} = \Delta f_{\mathrm{a}}\alpha/2$, with Δf_{a} the astigmatic focus difference. For an unround lens bore with unroundness b, after Sturrok [4] one has $\Delta f_{\mathrm{A}}/b = 100(2 + 3(f/D)U/(NI)^2$. If $b \approx 1$ μm then is $\Delta f_{\mathrm{A}} = 2$–$3$ μm. Using an aperture $\alpha = 6 \times 10^{-3}$, the astigmatic disc of least confusion is $d = r_{\mathrm{A}} = 9$ nm. To obtain in TEM a resolution of 0.1 nm, this aberration must be corrected to $\Delta f_{\mathrm{A}} = 0.1$ μm.

Deflection aberrations are of crucial importance for high-precision deflection of electron beams. A side-field deflection system for 100 kV electrons was developed, which rendered a linear deflection for wide angles. It was employed in e-beam welding and engraving machines in the 1960s. Figure 4.9 shows the side-field deflection element which was used in Zeiss EB100 electron beam welding machines, built on an epoxy winding carrier with coils wound around a ferrite ring core, and switched to generate a deflection field.

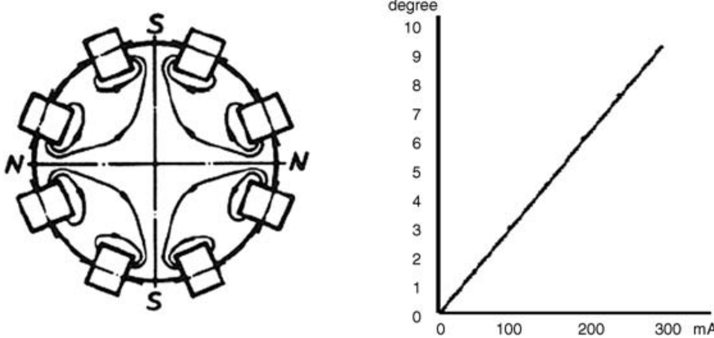


Fig. 4.9. Side field deflection element, principle, *left*, and deflection characteristic, *right*

The deflection of 9° was achieved with 0.1% deflection error. The system was 20 mm in height. An engraved field of squares shows the low deflection aberration [35]. For a detailed presentation of lens characteristics see Tables of Applied Physics by M. von Ardenne and other electron optics textbooks [36]. The immersion, 2 aperture lens and other applications of electric and magnetic fields like Wien Filters, cylinder spectrometers, magnetic and electric spectrometers, time of flight spectrometers and other accelerators are also discussed there.

Corrected Electron Optical Lenses

Using multipole correcting elements like octupoles with a refractive power $\sim r^3$, the spherical aberration can be corrected. Such aplanators are described in the microscopy section in more detail, and are now in use in high resolution scanning and transmission electron microscopes [37].

4.1.2 Electron Sources

History of Electron Sources

Gas Discharge Electron Sources

From early experiments with gas discharge tubes electron and ion sources were developed. As cathodes in vacuum tubes heated tungsten wires were used to release electrons with little requirements for local confinement. Induni [38] built the first technically employed gas discharge source. In a vacuum chamber, where the electrode distance d, the gas discharge space of length D and a gas pressure allows the mean free path Λ of the gas molecules fulfill the relation $d < \Lambda < D$, a gas discharge is maintained by the accelerating voltage. The electron beam is emitted to the apparatus through an aperture serving as anode and gas pressure determining opening nozzle of appropriate conductance. Figure 4.10 shows the gas discharge gun schematic.

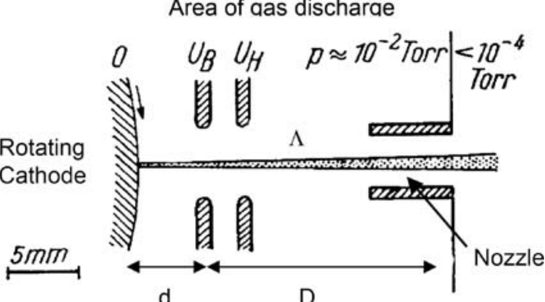


Fig. 4.10. Schematic of the gas discharge gun

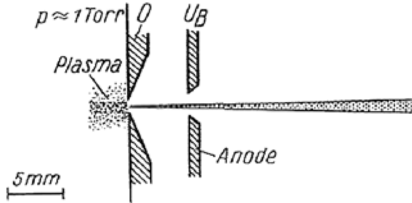


Fig. 4.11. Schematic of a two electrode system with a plasma discharge as cathode

At 50 kV a crater forms in the cathode due to ion bombardment. This crater determines the brightness [39]. To avoid this, technical applications use a rotating anode. This was measured using argon ions of 270 μA discharge current. Here no significant change of the brightness takes place over 5 hours. The following sequence of brightness was measured for different materials: $R_{\text{Mg}} = 3.5 \times 10^3$ A/cm^2 sr $>$ Rbe $>$ RAl $>$ RV2A $>$ RCuBe $= 5.7 \times 10^2$ A/cm^2 sr. For comparison of brightness values it is more convenient to give the reduced brightness R_r, which is the brightness divided by the electron energy at the measurement $R_\text{r} = R/U$b [V]. Therefore, the reduced brightness for a gas discharge source is: $R_{\text{Mg,r}} = 0.07$ A/cm^2 sr V $> R_{\text{Be}} > R_{\text{Al}} > R_{\text{V2A}} > R_{\text{CuBe}} = 0.014$ A/cm^2 sr V. Change of discharge gas causes only small variations of the brightness.

A two-electrode system with a plasma discharge as a cathode, see Fig. 4.11, is used to extract electrons from a gas discharge and to prevent a high gas flux into the beam vacuum. This type of cathode led to high-perveance electron guns of the Duoplasmatron type used to cure surfaces. The extraction aperture prevented the space charge widening of the beam [40].

Three-electrode System with W-hairpin and Long Focal Length Crossover (Fern-focus)

This system was developed to save a condenser lens in an electrostatic microscope. The emitted beam has a narrow aperture. The source image is ≈50 μm in diameter, adequate to illuminate microscopic sample for transmission electron microscopy [41].

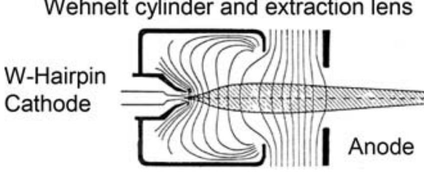


Fig. 4.12. Steigerwald's long focus electron gun combines Wehnelt shield cylinder and long-focal extraction lens

When experimenting with this system Steigerwald found, that he could weld and drill his apertures in the system with his beam, see Fig. 4.12. Now, such an electron gun is employed for electron beam welding. Using a rod cathode and 50 keV, at 2 mA emission current a crossover is achieved in 170 mm distance with a radius of 0.1 mm [42].

Modern Electron Sources

Electron sources are in common use for a long time. All the oscilloscopes, TV sets, monitors, vacuum display tube devices, and vacuum tubes use thermally released electrons from wire or dispenser cathodes, as described in Chapter 10. Only a few applications used field emitted electrons. Technical applications are high-power electron beams for metallurgy and materials recycling, waste water and gas treatment, conservation of natural products, as described in the following sections of this chapter, and also electron optical applications like microscopes and electron beam writers for lithography.

Three properties of the electron gun are important: brightness (A/cm^2 steradian), energy spread (ΔE eV), and lifetime. In addition, it is of importance which vacuum level is required for the operation of the source, and what is the technical precision required to mount the source. The brightness determines how much current can be focused into a given beam diameter from a given solid angle, and the energy spread sets the diameter of the disk of confusion due to the chromatic aberration coefficient of the final lens. The energy spread varies with gun operating conditions and has been measured for a limited number of cases. Lifetime is a customer requirement and determines the cost of ownership.

An electron source generally is composed from an emitter material of special size and form, see Fig. 4.13, a second electrode to shield or accelerate the electrons, and an anode electrode, which accelerates the electrons to the final energy. Figure 4.14 shows schematic presentations of cathodes. Left: EM7 tungsten hairpin system for early electron microscopes and beam welding and drilling systems. Here the wire is replaced by a ribbon of 1 mm width. Right: dispenser or oxide cathode for TV sets and microwave tubes [43].

Figure 4.15 gives a schematic of a high-power rod cathode heated by direct electron bombardment. Such sources are employed in e-beam melting machines. The wedge-shaped acceleration field separates outgoing electrons and back streaming ions due to their different energies [44]. The separation of electrons and ions generated by the e-beam rendered constant brightness over time, a constant high-emission

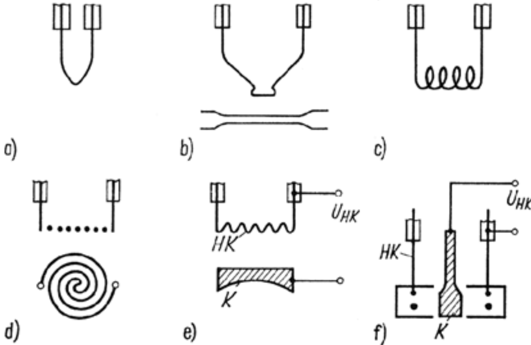


Fig. 4.13. Cathode designs: **a** hairpin cathode, directly heated, **b** strip cathode, directly heated, **c** coil cathode, directly heated large area cathode, **d** spiral cathode, directly heated large area filament cathode, **e** block cathode, large area cathode heated by electron bombardment, **f** bolt cathode, small and large area cathode, the emitting area is the front face of the bolt. *HK*: auxiliary cathode, *K*: cathode, U_{HK}: bombardment voltage between auxiliary cathode and cathode, $U_{HK} = 300$–3000 V

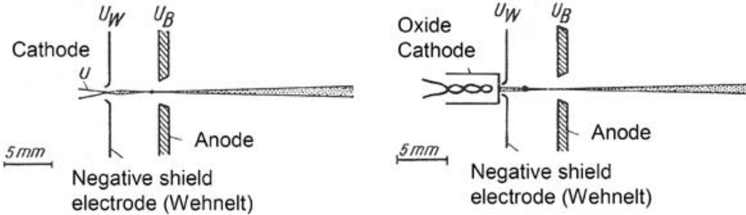


Fig. 4.14. *Left*: EM7 electron gun for microscopes. *Right*: indirectly heated dispenser or oxide cathode, e.g. as used in TV sets and vacuum tubes

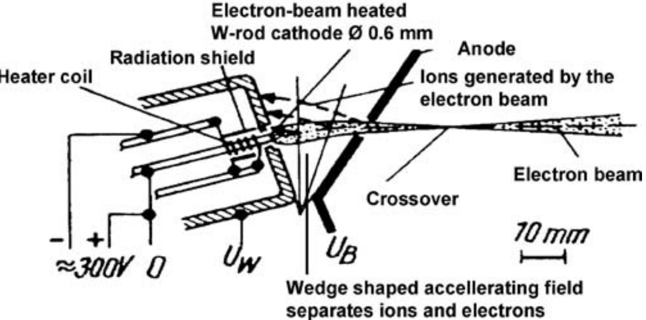


Fig. 4.15. Schematic of a high-power rod cathode heated by direct electron bombardment

current density <2 A/cm^2, and prevented a crater formation in the cathode by the ion bombardment. It was especially used in high-current systems for Eidophor TV

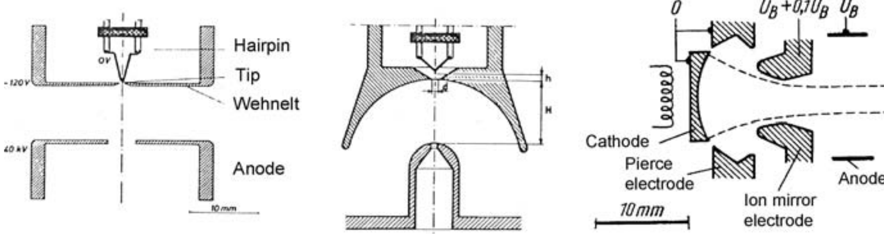


Fig. 4.16. *Left*: the EM7 flat three-electrode gun. *Middle*: the Rogowski cathode [46], as it is used in electron beam welding and drilling machines. *Right*: Pierce-cathode system of high perveance for large currents

projection systems. At 15 keV, the emission current density was $j_{em} = 0.7$ A/cm^2, and the brightness $B = 8 \times 10^3$ A/cm^2 sr corresponding to 0.53 A/cm^2 sr V.

A two-electrode system of high perveance uses an extended curved cathode. To prevent ion bombardment of the cathode, an ion mirror potential wall is generated by a positive potential with respect to the anode. This confines also the positive ion concentration, which is used in the extended beam to reduce the space charge broadening of the beam despite of having high currents in the beam [45].

Figure 4.16 shows schematic the EM7 the flat three electrode gun (left) and the Rogowski cathode (middle), as it is used in electron beam welding and drilling machines. For large beam and very high current applications the Pierce gun is used, see Fig. 4.16, right.

High-brightness Electron Sources

W-hairpins deliver at 2900 K a current density of 1 A/cm^2. Oxide and dispenser cathodes deliver up to 10 A/cm^2, but are large area and high current sources. For microscopy an increase of brightness is wanted, but not of current. Therefore, LaB_6 rod cathodes are in use, which deliver 10 A/cm^2, but those are used in a tipped configuration grinding the single crystalline material to a 4 flat pyramid, see Fig. 4.17. This delivers at 1850 K a current density of 10 A/cm^2 at a total emission current of 10 μA.

An increase of brightness is obtained for microscopy using tipped hairpin cathodes and Schottky or cold field emitters. Figure 4.18 gives schematically the used electrode configuration. For tipped thermal cathodes the electrode is a suppressor at −100 to −1000 V to confine the electron emission by the Wehnelt potential. At 50 keV electron energy and $T = 2000$°C cathode temperature tipped cathodes rendered a brightness of $B = 3 \times 10^5$ A cm^{-2} sr^{-1} or Br = 6 A/cm^2 sr V. The emission current density was $j_{00} = 5.3$ A cm^{-2} at a total beam current of $I = 10$ μA.

Using Schottky- or cold field-emitters, the extraction electrode is used to generate by current regulation the field emission conditions, high field at the tip of ≈2–4 V/nm, and the anode is used to accelerate the beam to its full potential [49]. The etched sharp tip increases the brightness, however, the lifetime of such cath-

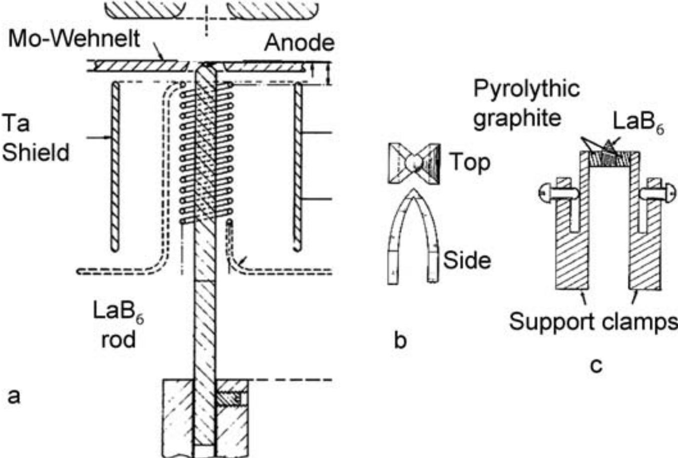


Fig. 4.17. Three types of LaB_6 cathode used with SEMs': **a** indirectly heated rod cathode [47], **b** directly heated LaB_6 hairpin [48], **c** pyrolytic sandwich cathode. Major problem with all LaB_6 cathodes remains the material itself. A sensitive balance between brightness and material evaporation at 1820°C is to be maintained. In a vacuum of 10^{-7} mb such cathodes have a lifetime of 3 months at continuous operation, e.g. in a lithography system or SEM

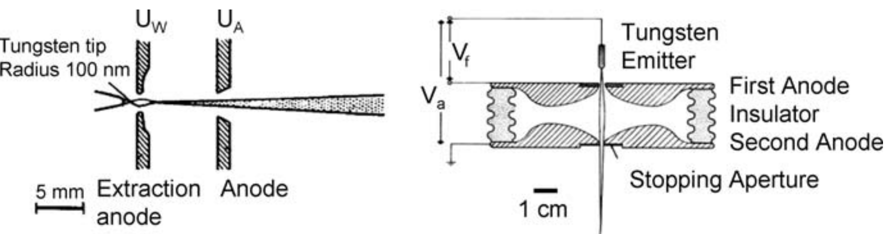


Fig. 4.18. Schematic of 3-electrode configuration. *Left*: tipped thermal cathodes or thermo-field (Schottky). *Right*: cold field emitters (Butler-gun) [51]

odes is limited in high vacuum by ion bombardment. Therefore, thermo-field and cold field-emitters are used in costly UHV gun chambers at an ambient pressure of $p < 10^{-9}$ mb. Today's most used thermo-field emitter is the Zr-O-W emitter. The zirconium layer on the top of the intermediate oxygen layer adsorbed and maintained at the <100> axis of a W single crystal tip supplies electrons with a work function of 2.9 V. The current density is <1000 A/cm^2. At 10 kV the brightness is 2×10^7 A/cm^2 sr, or $B_r = 2000$ A/cm^2 sr V [50]. Such cathodes are used in three-electrode configurations.

Electron guns with higher brightness and lower energy spread can improve resolution and reduce exposure time in SEMs, as given in Table 4.1. They will only improve resolution, however, when other limitations such as instrumental imperfec-

tions and/or electron specimen interaction phenomena do not predominate. In many cases where other limitations do predominate, greater brightness will only increase the beam current.

Table 4.1. Comparison of different electron sources for microscopy, lithography, deposition and analysis

	Field emission (cold cathode)	LaB_6	Tungsten filament
Brightness (A/cm^2 sr)	10^9	10^7	10^6
Virtual source (Å)	$<10^2$	10^5	$>10^5$
Energy spread (eV)	0.2	1.0	1.0
Service life (h)	≥ 2000	1000	40
Operating vacuum (Pa)	10^{-7}	10^{-5}	10^{-3}

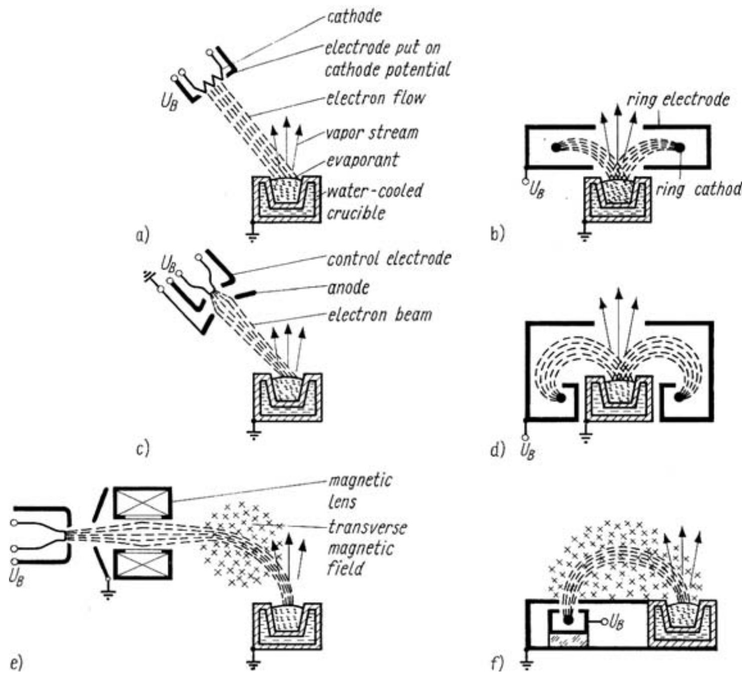


Fig. 4.19. Principles of electron sources to evaporate materials in technical processes: **a** evaporator with electron impact heating, linear cathode and electrostatic focusing; **b** evaporator with electron impact heating, ring cathode and electrostatic focusing; **c** evaporator with axial gun and electrostatic telefocusing; **d** evaporator with electron impact heating like (**b**) with electrostatic focusing and beam turning; **e** evaporator with axial gun, magnetic focusing, and magnetic beam turning by 90°; **f** evaporator with transverse gun and magnetic beam turning by 180°

Brightness

The brightness B of an electron beam is the current density per unit solid angle, with the current density $j = I/F$, the crossover area $F = \pi d^2/4$, $\Omega = \pi\alpha^2$ the unit solid angle. For a converging or diverging beam, B is given by the following expression:

$$B_r = I/F\Omega\varsigma = 4I/\pi^2 d^2 V \text{ [A/cm}^2\text{ steradian volt]}, \quad (4.1)$$

where I is the beam current (ampere), d is the beam diameter (cm), α is the beam half-angle (radian) and V electron accelerating voltage.

Experimental determination of brightness is difficult because it requires accurate measurements of beam diameter, beam angle, and beam current. Three methods are most frequently used to measure brightness: 1) the two aperture method [52], with moderate accuracy for thermal cathode electron guns where the crossover diameter is 10 µm or greater, 2) the cross-over projection method [53], a more accurate method and can be extended to smaller crossovers, and 3) measurement of the characteristics of a converging beam [54] being the most convenient for measuring the brightness of an SEM beam and also useful with field emission guns. Electron sources for materials treatments and evaporation – high power electron beams – use a high current cathode and electrostatic or magnetic beam focusing and beam stirring, see Fig. 4.19. Beam currents range from mA to hundreds of A. Accelerating voltages range from 3 to 30 kV. Applied power is from mW to 10 MW.

4.2 Thermal Materials Processing

In thermal EB machining processes the beam is used for the removal of material at the point of action by melting and vaporization, or for producing a temperature-dependent change in material properties. The two main applications in use differ in regard to the processing modes and plant design: thermal processing of solids and of thin films. For thermal processing of solids the kinetic energy of the beam electrons is converted to heat when the beam hits the surface of solid matter. This heating is used either for the removal of material or for thermally inducing structural or chemical changes at the point of beam action. In general, it will be necessary to protect the neighbourhood of the work site, at least outside a certain range around this point, from any undue thermal stresses and X-ray radiation doses that may occur. For EB machining like drilling or milling diameters, widths, depths, and profiles are limited to a processing depths T_R, which is determined by the electron range S in the corresponding material to $T_R \gtrapprox S$.

The process of energy absorption in the target material depends on the power density in a Gaussian rotationally symmetric beam, which is given by $p_F(r) = U_R j_{FO} \exp -(2r/d_F)^2$. It supplies the power distribution $p_F(r) = U_R j_{FO}(\pi d_F^2/4) = U_R I_R$, where U_R is the accelerating voltage, I_R is the beam current, d_F is the focal spot diameter, r is the distance from the spot centre, and p_{F0} and j_{FO} are the power and current densities at the centre of the spot, respectively.

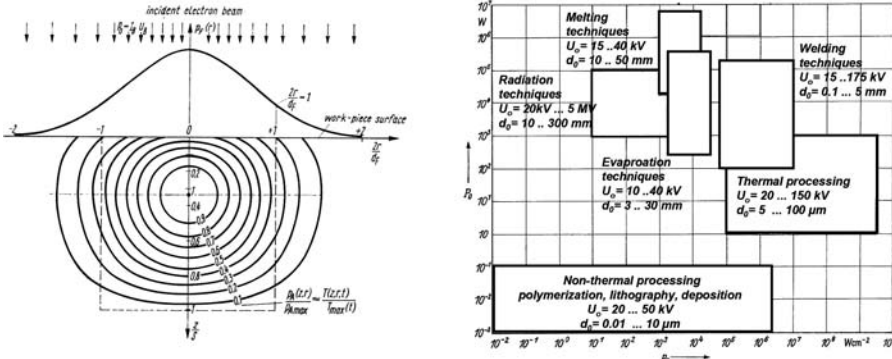


Fig. 4.20. *Left*: spatial density distribution of the relative absorbed power $p_A/p_{A\,\max}$ as a function of the reference coordinates $(2r/d_F, z/S)$ for $d_F \gg S$. *Right*: power regimes for e-beam processing

Not all of the beam power P_0 is converted to heat in the work piece, since back scattering reduces the active number of electrons within the electron range S by a fraction ηP_0, where η is the back-scattering coefficient <1. In addition, the surface topography of the work piece reduces the active power due to a topography dependent η value.

Typically, for $d_f \gg S$ one obtains $p_{A\,\max} = 1.45\eta U_R j_{FO}/S$. Since $S \sim U_R^2$, the absorbed power density is proportional to j_{FO}/U_R. This means that at similar current densities it decreases with rising accelerating voltage, despite of the increase in power density. This estimation also holds for $d_f \approx S$. As a consequence, this yields a drop in $p_{\max}$ and a wider radial extension of the absorbed power distribution as the values of d_f/S decrease, see Fig. 4.20, left. Figure 4.20, right displays the power regimes for e-beam processing applications.

In the material-removing process the power density at the hit surface is influenced by: reduction of back-scattering losses at a greater channel depth; enhanced wall admittance at reduced actions on the bottom by way of the beam aperture power density distribution at the work-piece surface and at the bottom of the hole; increased power density at the bottom of the hole due to large reflection coefficients of the marginal rays; increased wall admittance and reduced actions on the bottom by electron scattering in vapor; focusing action due to positive space charge due to metal vapor ionization; and absorption and scattering by splashes of the material.

Considering Gaussian heat sources allows to solve the inhomogeneous equation of heat conduction $((\partial/\partial t) - (\lambda/\rho c)\Delta)T = p_A/\rho c$, with $\Delta = (\partial^2/\partial r^2)$. This equation generally allows one to determine the temperature distribution $T(z, r, t)$ caused by the action of the beam with respect to time and space. Here a is the thermal conductivity, Q is the density, and c is the specific heat of the material. In the case of a flash-type point source with $p_a = E_0\,\mathrm{d}(z)\,\mathrm{d}(r)\,\mathrm{d}(t)$ a temperature distribution in a homogeneous infinite medium is produced. Here E_0 is the input energy supplied to the point $r = 0$, $z = 0$ at time $t = 0$.

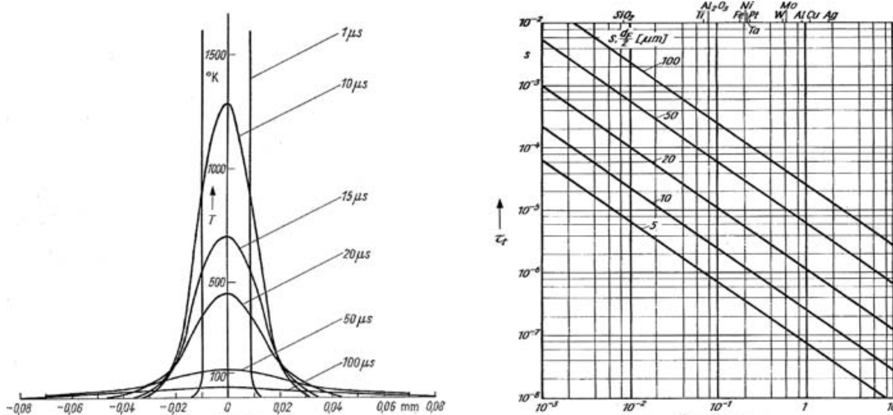


Fig. 4.21. *Left*: temperature profile produced from a point source by heat conduction for various times. *Right*: temperature propagation time τ versus the thermal diffusivity $a = \lambda(\rho c)$ for various values of S and d_F

$T(r', t) = (E_0/(\rho c(4\pi at)^{3/2}) \exp(-r'^2/4at)$, where a is the thermal diffusivity and r' is the distance of the point under consideration from the point source. The exponential term describes the temperature propagation with respect to space and time, and the previous factor causes the conservation of energy with time. Figure 4.21 plots the spatial temperature profile produced by heat conduction from a point source for various times.

By resolving the density distribution p_A of the absorbed power in a timed sequence of spatially arranged point sources and superposing the temperature distributions obtained, it is possible to calculate the temperature characteristic described above for any energy distribution with respect to time and space. If the time t approaches the order of τ_t the effect of heat conduction causes an extension of the temperature field beyond the limits of energy absorption set by the focal spot diameter d_F and the electron range S. The temperature rise in the maximum of the energy absorption falls short of the time-linear characteristic: $T_{max}(t) = P_{A\,max} t^n/\rho c$, with $n < 1, t \gtrapprox \tau_t$.

At τ_t, it finally approaches a steady-state value. The resulting temperature distribution depends solely on the input beam power and the thermal conductivity of the material. For $r > d_F$, or $r > S$ this temperature distribution is $T(r) = P_0/(2\pi\lambda r)$, to τ_t and goes far beyond the temperature distribution obtained with quasi-adiabatic heating. The temperature difference between the point of maximum energy absorption and the surface becomes almost zero.

4.2.1 Welding

First electron beam welding machines have been developed around 1950. Since then the technology of electron beam welding has found a world-wide wide-spread application [55]. The reason is the unique characteristics of this welding process, if

Table 4.2. Thermal processing effects

Thermal effect	Technological effect	Materials
1. Phase transformation in the solid state	Hardening	Hardenable steel
2. Sublimation	Material removal, drilling, milling, engraving, separation	Diamond, graphite, quartz glass
Decomposition of solids into volatile components	Etching, nanostructuring	Cellulose (paper), various plastics, green ceramics
Decomposition with at least one solid component in the cooled-down state	3-D constructive nano-structurization	Gallium arsenide, gallium phosphide EBID induced processes
3. Melting	Microwelding	Metals
	Polishing, alloying	Silicon, germanium
	Fusion scratching	Silicon, germanium, ceramics, ferrite
4. Vaporization Quasi-sublimation	Material removal (drilling, milling, engraving)	Metals, plastics, dielectrics
5. Shock waves	Crushing	Rock

compared to the commonly used electro-welding process, see Table 4.2 summarizing thermal processing effects. Very small joints of very different materials can be generated with a very high precision and with a highly automated process [56].

The main characteristics of electron beam welding are as follows. Very high power density in the focus spot of the electron beam, e.g. 10^7 W/cm^2. Up to 200 mm welding depth is possible. Very small weld joints are obtained due to an in-depth welding effect. Very low heating of the work piece. Welding in vacuum, in inert gases or in air. High reproducibility of the quality of the welding seam. High welding speed and productivity. No addition of welding material required. Automated process and weld seam positioning. Many materials and work piece combinations which cannot be welded with conventional electro-welding or arc welding can be joint with electron beam welding [57].

The schematic of the electron-beam welding system shows Fig. 4.22, left. The beam is formed with an axial electron gun, focusing magnetic lenses and a deflection system for beam stirring. The electron beam column sits on top of a work chamber with work piece stage. Further components are the vacuum system, high voltage generation, beam deflection and machine control computer. Positioning of the work piece with respect to the beam is performed using either a CCD camera peeking through an optical microscope having its objective lens inside the welding system or using backscattered electron detection with scintillator–photomultiplier amplification or semiconductor detectors. Middle shows a modern welding system. Right shows an in-depth welding seam. This is possible by the in-depth welding effect, which originates from the high power density which not only melts the materials but vaporizes it. The pressure of the metal vapor squeezes the molten material to the side and a capillary of vapor is generated, which allows the electron beam to penetrate into deeper depth. This effect produces the narrow and deep seams in electron beam

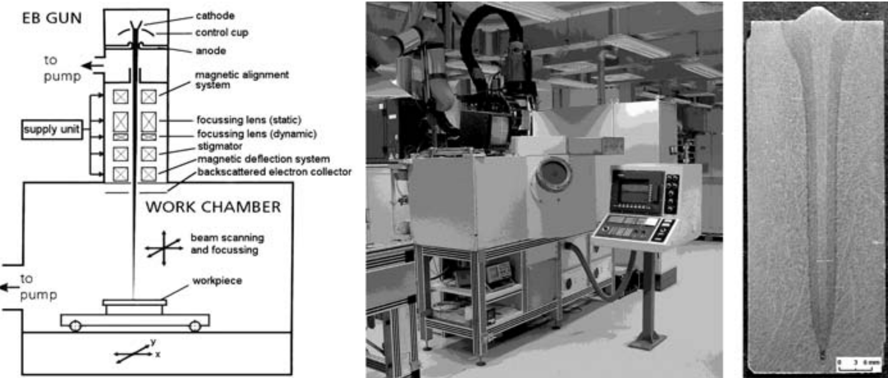


Fig. 4.22. *Left*: electron beam welding machine, schematic. *Middle*: e-beam welding system, max. beam power 10 kW, $U_{\mathrm{B}} = 60$ kV. *Right*: in-depth welding seam, 63 mm deep, 2 mm wide

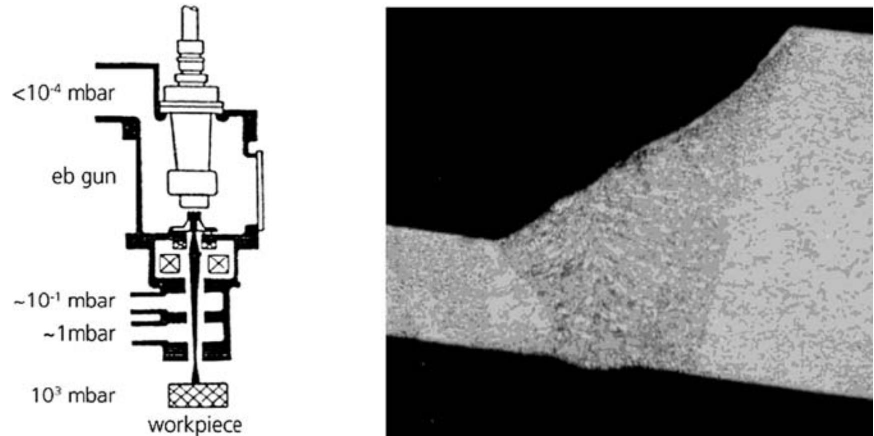


Fig. 4.23. *Left*: schematic of an e-beam welder operating in air. *Right*: weld of a tailored Al joint

welding systems, see Fig. 4.22, middle and right. Typical parameters are: an electron energy of 60 to 150 keV, a maximum power of 1 to 50 kW, and beam diameters of 0.1 mm to 1.0 mm depending on the beam power. In high vacuum chambers steel and non-iron metals like aluminium and copper are welded at 1 Pa chamber pressure with good quality, see Fig. 4.23 for the schematic of the system, left and the welding result, right. Reactive work pieces like titanium require a work chamber pressure of <0.05 Pa.

To fabricate very large steel structures, like nuclear reactor housings or pipeline tube welding, the electron beam is allowed to pass through diaphragm pressure stages out to the atmosphere [57], see also Fig. 4.23, left. Typical beam diameters are now due to the scattering in air around 1.5 to 2.5 mm. To obtain a homogeneous seam, the

beam is wobbled and guided in geometric figures inside the weld area by elaborate scanning processes.

Hole Drilling by Means of an Electron Beam

The first industrial use of holes drilled by e-beam in jewel bearings of watches was produced in 1960 at Carl Zeiss, Oberkochen, see Fig. 4.24a. Today single pulse mode or perforation and multi-pulse drilling are state of the art, with again two different techniques: percussion and trepanning, see Figs. 4.24 and 4.25. In practice, perforation is used nearly exclusively due to the high productivity compared to laser beam drilling, electro discharge machining or electro chemical machining.

The highly accelerated electrons are focused to or near to the work piece surface. At the spot of impact, they release their kinetic energy, melt and vaporize the material. For perforation at the reverse side of the work piece, a backing material is applied, which produces a high vapor pressure as soon as the beam penetrates through the work piece. This pressure of up to 100 bars blows the liquid material out of the borehole and leaves only little heat and a very thin (approx. 5 μm) recast layer at the wall of the hole. The main process parameters with a typically constant acceleration voltage of 120 kV are: mA beam current to vary the hole diameter; pulse time typically 10 μs to 10 ms depending on depth for penetration; the focus position relative to the work piece surface influences the shape (taper) of the hole. Figure 4.26 shows a modern electron beam drilling machine.

Drilling processes range from single pulse for screen production or rotogravure printing cylinders with 10 to 100 Mio holes/m^2 to multiple pulses and pulse modulation for smaller numbers or single holes, e.g. in ruby watch stone bearings to oscillating pulses for trepanning for stencils and spinnerets, and to tapered holes in special materials for jet engines. Challenges are scanning speed and beam positioning.

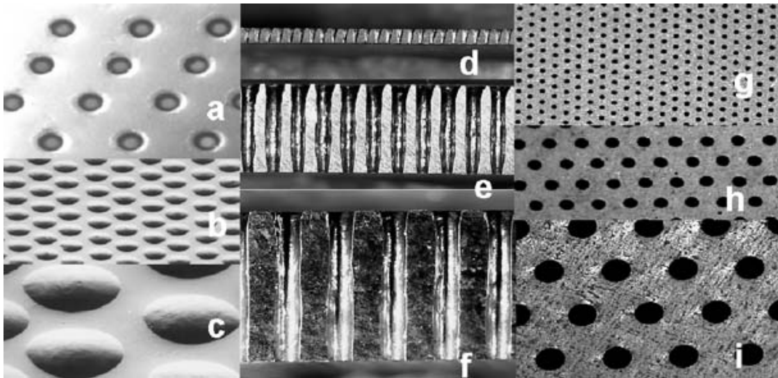


Fig. 4.24. Single pulse perforation: **a**–**c** rotogravure print cylinder, **d** holes for tofu production, **e** screen for pulp and paper centrifuges, **f** fiberizer for glass wool production in highly alloyed steel. *Right side*: top view of **d**, **e**, **f**

A high drilling rate is achieved and shrinkage problems are overcome by rolling the sheet metals around a drum and then continuously rotated and shifted on a mandrel below the beam. An incremental encoder triggers the pulses when the position of a hole is under the beam. During the time of impact, the beam follows the circumferential speed of the work piece surface (drilling on the fly). Drilling speeds between 10 holes per second for the largest possible holes and 150 000 holes per second for tiny little holes can be achieved today. The markets for drilling non-metallic materials, smaller holes and screen- or rotogravure-printing are still to develop by making use of higher speed computers. Figures 4.24 and 4.25 show different applications of the drilling process.

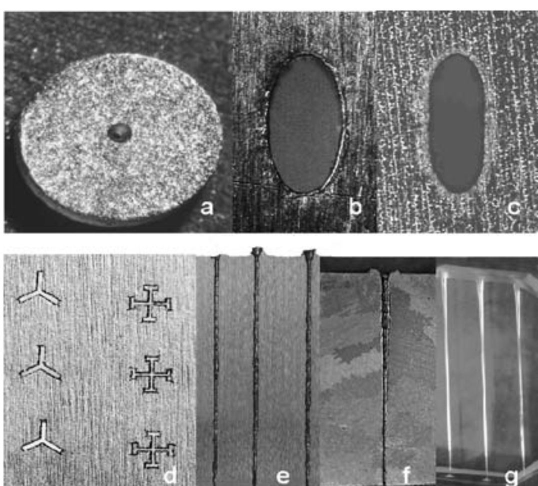


Fig. 4.25. Multi-pulse drilling: **a** rubin bearing of watches, hole diameter 0.08 mm, 0.3 mm thick; **b** trepanning of an ellipse, 3 mm long, 1.5 mm thick; **c** slotted hole with 0.6×0.25 mm; **d** trepanning of spinnerets, slot width 0.06 mm wide; **e** 60 mm depth hole with diameter 0.8 mm in steel; **f** tapered hole, 12 mm depth, exit diameter 0.12 mm; **g** quartz, 15 mm thick, hole diameter 0.1 mm

Fig. 4.26. Electron beam drilling machine

4.2.2 Thermal Surface Modification

Electron beams offer great possibilities for thermal surface modification. This is due to the high power density, small beam spot diameter and the great deflection capability of the beam. Table 4.3 presents the solid an liquid phase processes [58].

EB-hardening is employed to locally harden surface areas of steel, cast iron and can replace conventional technologies like induction or flame hardening. The e-beam is swept over the surface in the areas to be hardened and heats the material almost to the melting point. A self-quenching process by heat conduction hardens the material. Setting the duration of high surface temperature allows to set the hardening depth. Typically 10% of the wall thickness or 0.1 to 1.5 mm can be reached.

The advantages of e-beam hardening are:

- Hardening as a finishing process with exactly determined hardening depth.
- Low hardening distortions.
- Metallic shiny surfaces are obtained in the vacuum process.
- Highly reproducible as computer controlled process.
- Partial hardening of surface regions according to the technical construction.
- No tempering required.

By e-beam glowing the reshaping of the work piece is improved. EB-liquid phase processes are mainly used to improve the local corrosion and wear characteristics of work pieces.

E-beam fusing is widespread in use. In this technique the surface is shortly molten and by the high cooling velocity a fine grain phase is obtained, which results in lower wear and better corrosion resistivity. With multi-melt-zone technique typically a speed of 1 and 5 cm/sec are reached. Especially cast iron camshafts are produced by this way for automotive.

Table 4.3. Technologies for thermal surface modification with electron beam

Solid-phase processes	Liquid-phase processes
Phase transformation Hardening	Fusing
	Added material Soluble in molten base material Alloying
	Added material Insoluble in molten base material Embedding
	Added material Facing

To generate an electron beam melted alloy, an additional material is fed as wire, ribbon or rod to the melt, or the surface is coated with the alloy component before the beam application. This process improves the wear resistance of the treated surface. E-beam facing is used to reinforce highly stressed work pieces by melting the base only a little and adding the protective material.

Thermal Surface Modification by Application of High-power Pulsed Electron Beams

Intense particle beams got interest in the 1970s during the search for an effective very high power driver for inertial confinement fusion (ICF) delivering power densities of more than 100 TW/cm^2. They were not usable as the Bremsstrahlung created during interaction with the target material preheats the fuel inside the fusion pellet and thus prevents compression to high densities. Presently intense pulsed electron beams are mainly developed for material surface modification [59], see also Fig. 4.27, left and right, as pumping sources for high power excimer lasers, for intense pulsed microwave production [60], and as high-power radiation sources for flash X-ray radiography [61].

Material surface modification with pulsed electron beams efficiently transforms kinetic energy into heat. Typically a specific energy deposition of 3–15 kJ/cm^3 is required to heat a material from room temperature to above the melting point. Thus, for the achievable electron beam power densities of a few MW/cm^2 and for the penetration depth of 100–400 keV electrons ($\approx$20–100 μm) pulse durations of a few tens of μs are required. This is still short enough to neglect heat losses from the deposition zone. Using a high-current electron pulse, heating rates of up to 10^9 K/s can be obtained and melting of large surface areas (up to several 100 cm^2) to a depth of tens of μm becomes possible. Subsequent to the pulse, rapid quenching due to heat conduction into the bulk unaffected material occurs. At the resulting high cooling rates of up to 10^7 K/s amorphous, glass-like or new nano-crystalline structures are created, which improve the corrosion and wear resistance and increase the hardness

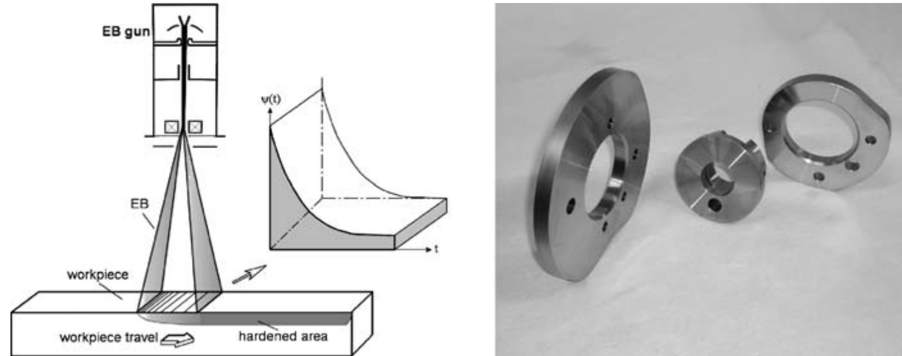


Fig. 4.27. *Left*: technology of surface hardening with e-beam. *Right*: disks from Ck45 hardened to a depth of 1 mm

of the treated surface layer, see Fig. 4.28 for a tooth hardened saw. In addition, alloys can be formed by co-melting a thin surface coating together with a thin layer of the base material.

A schematic of the pulsed electron beam facility GESA [62] at Forschungszentrum Karlsruhe is shown in Fig. 4.29, left. The electron source is a multipoint explosive emission cathode constructed of 700 graphite fiber bushels, each connected through a resistor to the high-voltage generator [63, 64]. Its emissive area amounts to 700 cm^2.

Pulsed electron beam treatment has been applied very successfully to MCrAlY superalloy coatings, which are widely used for oxidation protection on gas turbine blades at high temperatures above 900°C. Here M stands for Ni and/or Co. After melting the surface of this coating with a single 120 keV electron beam pulse of 30 μs duration, cooling rates of 2.5×10^6 K/s are obtained, and a glassy surface layer of 20 μm thickness forms. As shown in Fig. 4.29, right, the two-phase structure has changed into a single-phase structure in the melting zone. When exposing the

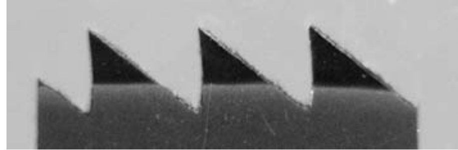


Fig. 4.28. EB-hardened teeth of a sharp industry knife. The depth of e-beam hardening is 1.2 mm. The tooth separation is 2 mm

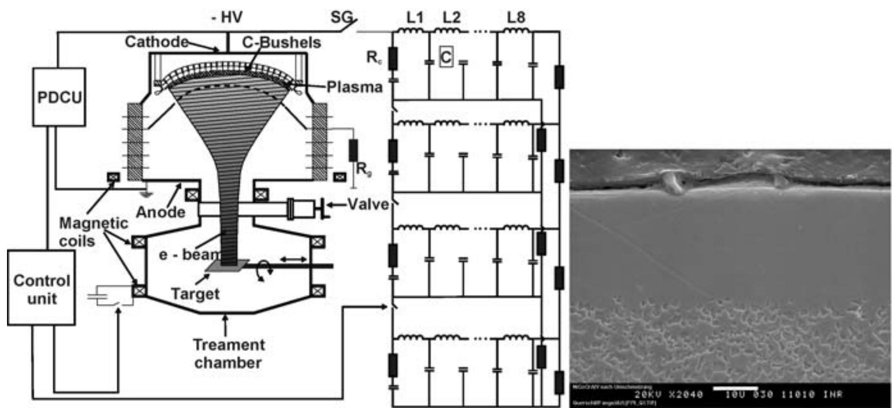


Fig. 4.29. *Left*: schematic of the pulsed electron beam facility GESA. (Electron energy: 50–150 keV, electron current: ≤700 A, energy flux on target: ≤50 J/cm^2.) PDCU stands for pulse disruption control unit which serves to control the pulse duration. *Right*: MCrAlY-oxidation protection layer after pulsed electron beam treatment. The *upper part* of the electron micrograph shows the restructured surface layer, whereas the original MCrAlY two-phase structure is still visible in the *lower part* of the micrograph. *Lower part*: original MCrAlY two-phase structure

MCrAlY layer to air at 1000°C, Al migrates to the surface and a thermally grown Al_2O_3 oxide scale (TGO) develops, which protects the underlying material against further oxidation. Much thinner and more uniform oxide scales grow on the e-beam treated surface. This TGO can also serve as a reliable bonding layer for ceramic thermal barrier coatings (TBC) like yttria stabilized zirconia (YSZO), which are used to increase the operating temperature and thereby the efficiency of future gas turbines [65].

In surface alloying thin metallic coatings (e.g. Al or Ni) are co-melted with a thin layer of the base material to form alloys which eliminate the corrosion of steel for special nuclear power reactors [66].

4.2.3 High-rate Evaporation by Electron Beam Guns

Electron-beam evaporation is used to produce thin layer coatings in a thermal process in high vacuum [67]. Different from other processes like inductive heated sources or jet evaporators the electron beam energy is converted to heat directly in the surface of the material in the crucible which is to be evaporated. Therefore, water cooled crucibles can be used, and almost all materials, even high melting point materials, can be evaporated with high purity and high evaporation rates. Also, reactive vaporization and precipitation of alloys is possible. In the 1960s high power electron guns were developed as axial or transverse evaporator guns [68].

Transverse-guns are compact electron sources constructed as a unit with the beam guiding optics and the crucible. They guide the electron beam in a 270° deflection from the electron emitting wire or coil up into the crucible, see Fig. 4.30. Typically, a power of 5 to 20 kW is used at 10 kV accelerating voltage. Such systems are mainly used in coating systems for optics and in laboratory units.

Axial guns having a dynamic beam deflection with deflection angles up to ±45° are used for evaporation in a power regime from 30 to 300 KW at 20 to 60 kV accelerating voltage [69], see also Fig. 4.31, left and right. It is required to separate the gun and the crucible area by differential pumping and pressure stages. Such guns

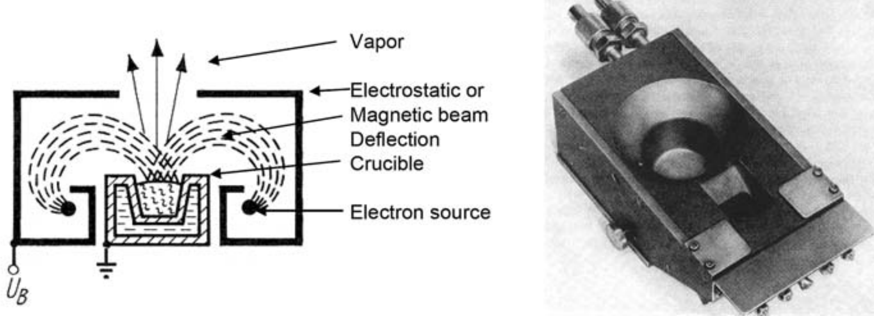


Fig. 4.30. *Left*: principle of a 270° transvers gun evaporator. *Right*: technical realization with a crucible diameter of 50 mm, 14 kW, 10 kV

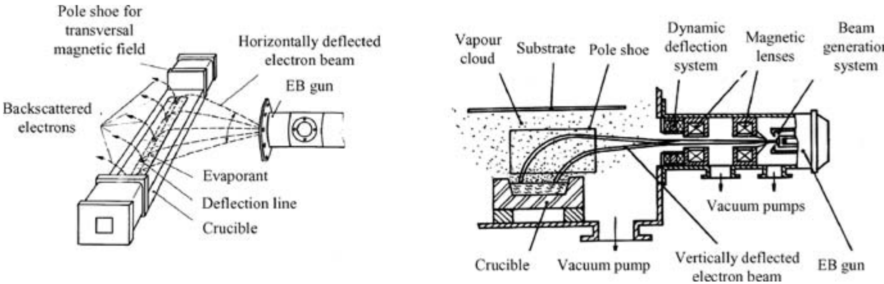


Fig. 4.31. Axial guns employed for generation of evaporation line by EB bended in a static transverse magnetic field above the crucible. *Left*: EB from an axial gun horizontally deflected. *Right*: EB vertically deflected to cover the whole bath of the melt

are especially useful for reactive evaporation and large area coaters with high vaporization rates. Crucible dimensions exceed even 1 m, and rates reach 10 μm/sec. Backscattered electrons are not contributing to heat the crucible, and special measures like magnetic traps are used to prevent them to ionize the evaporated atoms and to reach the sometimes thermally sensitive substrate.

The evaporation rate is determined by the local surface temperature and the material's vapour pressure at that temperature,

$$r_{\mathrm{V}} = \frac{\sqrt{m}\,p_{\mathrm{D}}(T_{\mathrm{V}})}{\sqrt{2\pi k T_{\mathrm{V}}}},$$

with r_{V} specific evaporation rate for the vaporized mass per area and time, T_{V} the vaporation temperature, p_{D} the vapour pressure of the vaporized material, k Boltzman's constant, m the atom, respectively, molecular mass of the vaporized particles. Integrating the Clausius–Clapeyron equation, the relation of p_{D} (T_{V}) is obtained

$$r_{\mathrm{V}} = A \cdot \sqrt{T_{\mathrm{V}}}\,\exp -\left(\frac{Q_{\mathrm{V}}}{R \cdot T_{\mathrm{V}}}\right),$$

where A denotes integration and other constants, Q_{V} is the heat of evaporation, and R is the Avogadro gas constant. The density of the vapor stream distribution $\Phi(\alpha)$ follows the law $\Phi(\alpha) = \Phi_0 \cos^n \alpha$, with α the angle between the vapour and direction perpendicular to the surface. In case of high pressure in the vessel, atom scattering of the vapour atoms widens the distribution.

Specific evaporation rates versus temperature are given for several materials in Fig. 4.32. It is necessary to use dedicated scanning techniques to keep the melt temperature within a small deviation of the required temperature in the bath. This holds especially for deposition of alloys and in deposition of materials produced in the gas reaction before the deposition.

The economically important applications of e-beam deposition with high rate of evaporation are: aluminium for barrier layers on plastic films for food packaging(100...200 nm/s), CoNi CrAlY as corrosion resistant overlay coatings on turbine blades(100... 400 nm/s) and ZrO_2-$Y2O_3$ as thermal barrier coatings on turbine

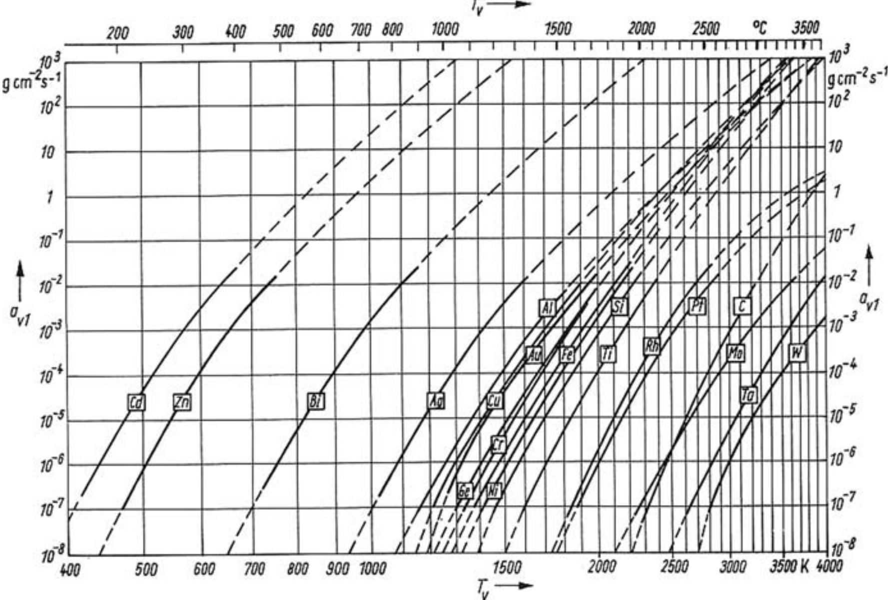


Fig. 4.32. Temperature dependence of the specific evaporation rate of some materials

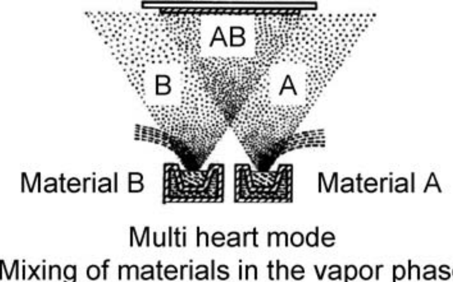


Fig. 4.33. Mixing in vapour phase by separate evaporation from two crucibles

blades. Cobalt layers as magnetic tape storage medium, and TiO_2 photocatalytic layers and high refractive layer in optical systems. SiO_x as transparent protective layer and as low refractive index layer in optical layer stacks. TiN-O as solar absorber layer on copper foils. Mixing in vapour phase by separate evaporation from two crucibles is shown in Fig. 4.33.

A dedicated compute control hardware is employed to assure constant evaporation speed and rate as well as layer thickness at the target. When depositing alloys or gas reaction generated materials, one has to consider that gases are chemisorbed, but metals are adsorbed with a stronger bond. Reactive gases therefore have to have a higher pressure to avoid non-stoichiometric deposits. Evaporated deposits are in general microcrystalline in a columnar build up.

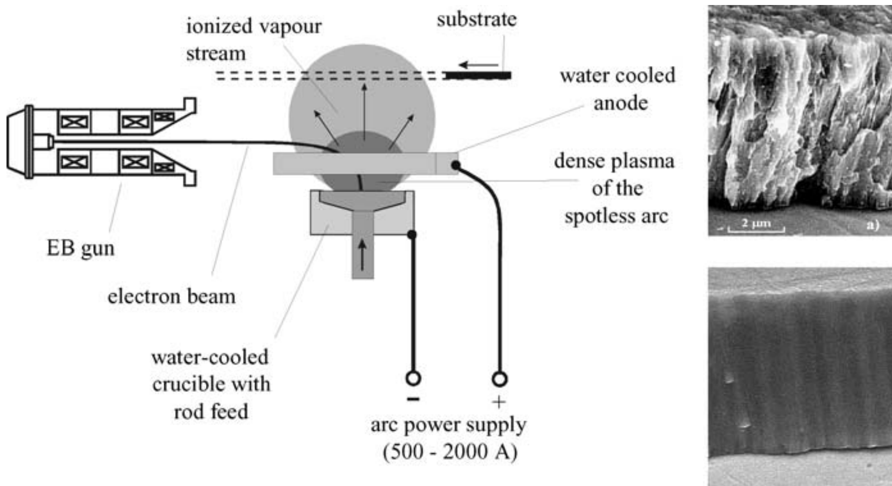


Fig. 4.34. *Left*: schematic of the spotless arc process (SAP). *Right top*: conventional deposited layer of Ti without plasma activation, substrate temperature 180–220°C, rate of evaporation 100 nm/s. *Right bottom*: SAP produced layer, substrate bias 100 V, substrate temperature 50–220°C, ion current density 50 mA/cm^2

A modern way to generate almost amorphous deposits is to generate a plasma discharge by an arc to the melt and to accelerate ionized evaporated atoms to the substrate. Figure 4.34 shows the SAP process schematic and the comparison of a conventional and a SAP or HAD produced coating, with SAD (spotless arc deposition) [70]. The HAD (hollow cathode arc activated) process delivers similar good results.

In metal, glass and plastic coating technology the material is provided in a vacuum chamber as rolled or sheet material. After a cleaning step by gas discharge, plasma or ionizing process, the material is transported into one or more evaporation and deposition chambers and finally stored in a collecting chamber. Such machines are produced, e.g. by Leybold and v. Ardenne Anlagentechnik, and can extend 10 to 30 m and coat up to 3 m^2 in one process.

4.2.4 Electron Beam Melting and Refining in Vacuum Metallurgy

A universal cathode ray furnace device with a modern concept for the beam source for vaporizing and melting was described in Rühle's patent [71] in 1939. In the 1950s industrial utilization of electron beam melting arose from the very demanding new requirements of special metals in the aircraft, aerospace, chemical and nuclear industries. The development was possible by the availability of high-performance vacuum pumps allowing reliable control of large gas loads. The first industrial electron beam melting furnaces were developed around 1955 with 30 to 250 kW power by Candidus (National Research Corporation) and Smith (Temescal Metallurgical Corporation). Subsequently erected plants employed numbers of high-power electron guns. Also,

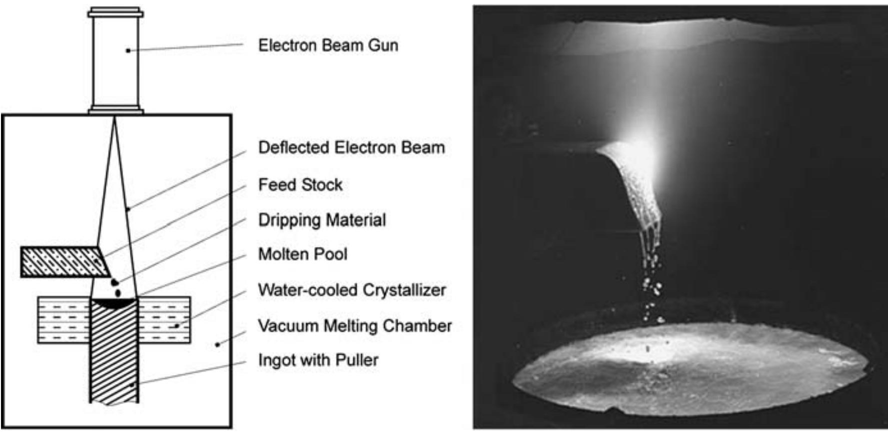


Fig. 4.35. *Left*: the principle of electron beam drip melting technology. *Right*: reality

methods for improved refining were developed [72]. Very high power furnaces of up to 6 MW were developed in the 1980s, e.g. a 6000 kW cold-hearth furnace with 5 axial guns for the melting of steel (Ingot weight 30 t) by Jumz (Russia). Other companies are: NRC (USA), Heraeus, IvA, and ESW(Germany), Airco Temescal and THT (USA) using 1200 kW drip melting furnace with 1 axial gun or 3000 kW cold-hearth furnace with 22 transverse guns, each rated 150 kW. In 1997 Timet, Oremet (USA) used a 5000 kW cold-hearth furnace for melting of titanium with an ingot weight of 25 t.

Electron beam melting (EBM) is a two-mode process. The original drip-melting mode relies on selective vacuum purification, removing impurities with vapor pressures higher than the metal being processed. In cold-hearth refining, or EBCHR, inclusions are removed as the molten metal flows through the hearth by flotation and sedimentation in addition to the evaporative purification. Those floating on the surface break down on exposure to the beam and either evaporate or dissolve in the metal. The heavy inclusions sink to the bottom of the hearth and are removed with the skull. Any light inclusions that remain floating are prevented from flowing into the ingot by a dam.

The different methods of EB drip melting are shown in Fig. 4.35. The effect of the high-power electron beam leads to melting of the rod of metal with formation of a thin liquid film on the melting zone of the rod. The liquid metal runs off the feeding rod and drops into the crucible. Molten material solidifies on the cold wall of the crucible to a shell preventing any further contact of the liquid metal with the crucible wall. In the crystallizer there is controlled cooling under the effect of the electron beam and formation of a refined ingot of metal by slow removal of the solidified metal through the base.

Multiple drip melting is used for high refining requirements, the metal is either repeatedly melted with increasingly lower chamber pressure in each melting step or not dropped directly into the crystallizer, but rather first of all collected in

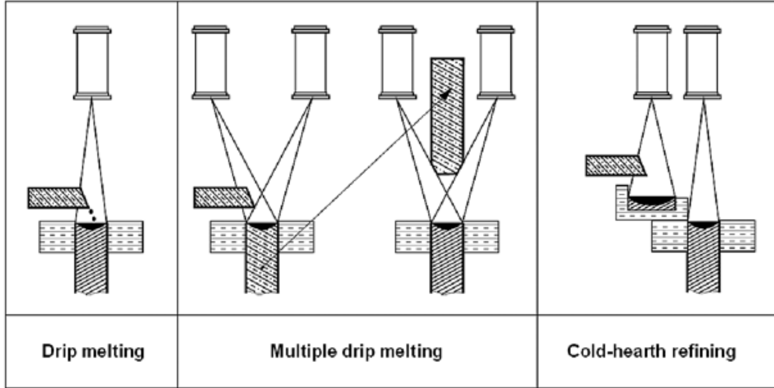


Fig. 4.36. Enhanced purification by multiple drip melting and cold-hearth melting

a cold-hearth, a flat vessel with a water-cooled copper wall (cold-hearth refining). Here, the melt is kept liquid for a longer period of time by means of electron beam heating. It is thereby refined and then flows slowly via an overflow into the crystallizer.

After coarse refining, single or multiple melting with decreasing process pressure further reduces the concentrations of metallic and non-metallic impurities. The purity from 99.5% to 99.995% of the molten metals is reached. The cold-hearth method is particularly used for Si, steel, Ti, V and Mo. The multiple drip melting method is preferred for high melting point metals (e.g. Ta, W, Nb). Figure 4.36 shows the variants of the method.

The refining effect of electron beam melting is based on several processes. Outgassing and vaporization of oxygen, nitrogen, hydrogen, carbon, oxides, and nitrides impurities, and in the case of titanium sponge also magnesium occurs due to the decreasing solubility of these substances in the liquid metal with decreasing pressure at a given temperature. Higher degrees of purity are achieved with electron beam melting compared to other vacuum-metallurgical processes. Aluminum is purified from the ore by vaporizing impurities of higher vapor pressure than the melt. Low density inclusions float on the surface of the melt and are either vaporized or dissolved on exposure to the beam, or prevented from flowing into the ingot by a dam. High density inclusions sink to the bottom of the hearth and are removed with the skull, e.g. in remelting titanium scrap and extracting tungsten carbide tool parts with this method.

Figure 4.37 shows a 100 kW electron beam furnace for research purposes and small batch production. The melting chamber is a water cooled double-walled steel vessel. Charge materials (rods or granulate) are fed through airlocks. In high productivity furnaces, rods are fed into two chambers reciprocally. Right shows the von Ardenne development EH 800V, a 800 kW gun using 60 kV and delivering a beam with a deflection angle of ±45° at 10 kHz. Figures 4.38 and 4.39 show different rod feeding ways and several procedures for special processes, e.g. to generate nano-

Fig. 4.37. Electron beam melting furnace *EMO 100* for research purposes and small batch production, beam power 100 kW. *Right*: EH 800V axial electron gun (courtesy of von Ardenne Anlagentechnik GmbH)

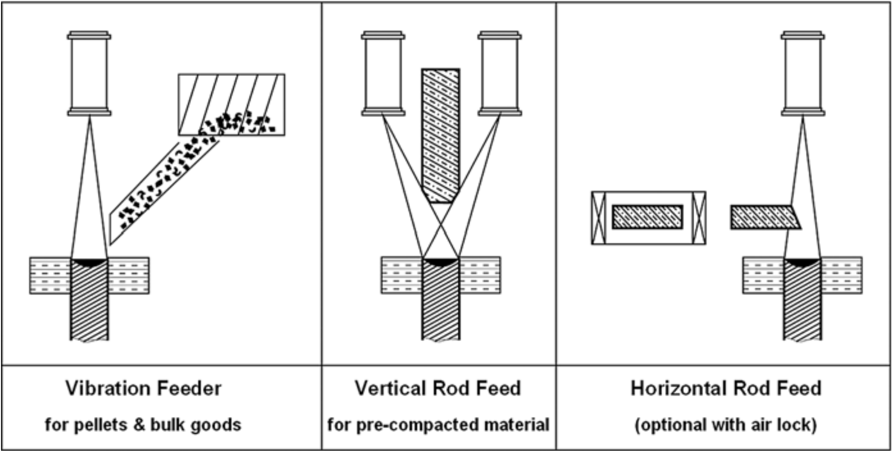


Fig. 4.38. Material feed systems in electron beam melting

powders. Non-thermal standard and electron processes are listed and compared in Table 4.4.

Current development, e.g. is aimed towards improving the metallurgical quality of niobium for superconducting, high-frequency resonators for use in advanced accelerators for nuclear and particle research. The purity of the metal is increased using up to 6-fold electron beam melting from 99.75% to 99.95%, and hence the residual resistance in the superconducting state is reduced by a factor of 300 [73].

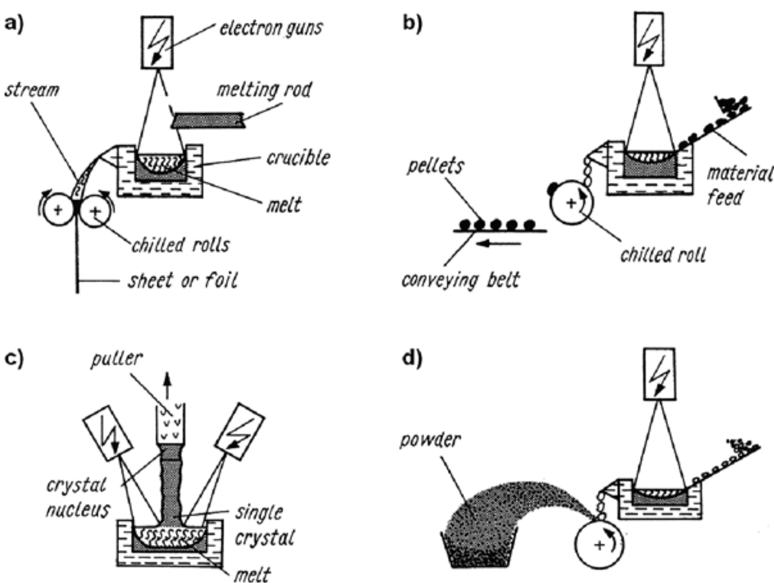


Fig. 4.39. Specialized methods of electron beam melting for fabrication of **a** metal films and strips, **b** pellets and granulates, **c** single crystals and **d** (nano) powders

Table 4.4. Non-thermal standard and electron processes

Applications	Characteristics	Additional features
Drying techniques		
hot-air drying IR and UV irradiation	efficiency, e.g. 15–20% of used energy	photosensitive additives
electron bombardment	>65% of used energy	no additives
Surface coating		
IR and UV irradiation	monodisperse resists	
electron bombardment	2 component resists give much harder surfaces	
electron-induced cross-linking of polymers, and hybrid materials	favorably-priced raw materials; characteristics are positively influenced and customized;	hardness, chemical stability and resistance, color durability, shrinkage designable
	layer properties are modified locally with a higher resolution	specialized resists for lithography
electron irradiation in presence of reaction partners	locally coating or etching of surfaces (electron induced processes EBIP)	resolution of 10 nm with beams of 1 nm diameter, 3D nanostructurization
E-beam sterilization	packaging, tools, materials and recycled goods	
e-beam disinfection	waste water, seeds	physical and chemical effects

4.3 Non-Thermal Processing Techniques

When energetic electrons penetrate into materials, they are retarded by different interactions and they transfer their energy to the atoms of the target material in a series of steps.

Energy transfer from electrons to matter:

Physical: inelastic scattering – characteristic X-ray emission, Auger electron emission by inner and outer shell ionization, secondary electron emission <50 eV, electron plasmon excitation (≈ 25 eV energy loss), phonon excitation <6 eV.

Chemical: breaking of bonds in molecules, generation of cross-linking, generation of radicals and reactive groups.

Biological: radiation damage to cells, fungus, bacteria.

4.3.1 Survey of Chemical Effects Caused by Electron Beams

The accelerating voltage U_b used in non-thermal processes for treating materials is carefully chosen such that the required electron range in the material for the particular treatment task is assured and such that there is no formation of excited nuclei as a result of inelastic collisions with the atom nuclei. The electron energies ($e \cdot U_b$) can be typically classified as follows:

- Low energy area 1 keV $< e \cdot U_b <$ 200 keV scanning microscopes, e-beam scanner systems.
- Medium energy area 200 keV $< e \cdot U_b <$ 800 keV e-beam accelerator and scanner systems.
- High energy area 800 keV $< e \cdot U_b <$ 10 MeV e-beam accelerator and scanner systems in concrete bunkers (atom dislocation in matter $>$750 keV)

Subsequent secondary reactions in the physico-chemical phase involve energy transfer via transitions to form mainly radicals. These are atomic and/or molecular groups with one of two non-bonded electrons. They are produced, for example, by cleavage processes, by decay of an excited molecule, by recombination of ions and by other subsequent reactions of the primary products that are produced.

- These reactive species cause structural changes to take place as a result of many reactions in the chemical phase (radical–molecule, ion–molecule, radical–radical and ion–ion reactions). This results in changing the properties of the material. These reactions often take place without catalysts making the manufacture of pure materials possible.

The molecular changes induced in polymers by electrons can be classified as follows:

- Scission and cross-linking of polymer molecules, with an associated decrease or increase in the molecular weight. In contrast, decomposition, cleavage and substitution reactions involve the decomposition of complex molecules and the splitting off of atoms and/or functional groups from the molecules.

Table 4.5. Temporal phases of electron treatment in organic materials

Phase	Excitation	Time (s)
Physical phase	Energy absorption	10–18
	(depends on density, atomic weight, atomic number,	10–13
	and energy),	–
	excited, ionized atoms and molecules	–
Physico-chemical phase	Secondary reactions – energy transfer/rearrangement	10–13
	Primary damages, generation of radicals	–
Chemical phase	Further complex reactions – energy transfer (depend-	–
	ing on pressure, temperature, molecular structure and	–
	state of aggregate.	–
	Radical-induced reactions. Molecule destruction,	–
	substitution, modification of molecular structure	0.1
Biological phase	Molecular changes and radiation damage in living organism can perturb or change that organism. Mutations or biochemical changes can cause cell death and death of the organism.	years
	Fight against cancer by hard radiation	

- Formation of small molecules such as H_2, CO, CO_2, SO_2, CH_4, HCl and aliphatic fragments (aldehydes, ketones), depending on the composition of the polymer.
- Changes to the molecular composition and structure of the polymer molecules, including loss or formation of unsaturated bonds. The irradiation of, e.g. transparent PVC, leads to colour development due to formation of conjugated C=C bonds. This is accompanied by the release of gaseous HCl.

In general, the four above mentioned reactions take place simultaneously. The kinetics of the reactions do however differ and hence so do their contributions to the overall reaction. Temporal phases of electron treatment in organic materials are given in Table 4.5.

The result of the induced chemical reactions depends highly on the molecular structure. For example, polymers having an aromatic ring structure have a greater resistance to electron treatment than aliphatic materials. At room temperature, the cleavage of the main chains is made more favorable by the presence of branched –C– structures in the polymer chain (e.g. CH_3 in propylene). In contrast, cross-linking reactions dominate in polymer chains containing –CH– and –CH_2– structures (e.g. polyethylene).

The G-value introduced by M. Burton quantifies the chemical reactions that are induced in materials by ionizing radiation. It indicates the number of molecules that are changed per 100 eV absorbed energy. The G-value hence allows statements to be made about the reactivity and resistance of materials to electron beam treatment, see Table 4.6.

In addition to structure dependencies, the induced chemical reactions depend on the conditions during the electron treatment like temperature and chemical environment. Inert conditions (e.g. vacuum, argon, nitrogen), the presence of reactive gases

Table 4.6. G-values for selected materials and reaction types [74]

Polymer	G_S Scission	G_C Cross-linking
Polymethylmethacrylate	1.2–2.6	
Polytetrafluoroethylene	0.1–0.2	
Polyisobutylene	1.5–2.0	
Polyethylene		2.0
Polydimethylsiloxane	0.07	2.3

(e.g. air, water vapor, sulfur dioxide) and liquids (e.g. water, alcohols) often considerably affects the final results of the treatment.

In addition to the immediate structural changes, there are also often post-treatment effects. Radicals which stay trapped in polymers may react over a longer period of time(e.g. hydrocarbons, carbon dioxide and monomers). Their storage can lead to stress and cracks in glass-like polymers (e.g. PMMA).

Molecular changes which arise during the chemical phase within a living organism can perturb or change that organism. Ongoing reactions in the biological phase generate a real damage via the development of mutations or biochemical changes which can cause cell death and death of the organism. These relationships, which are often very complex, form the basis of a separate branch of science, the molecular radiation biology. Practical applications of this discipline are radiation therapy, for example, to fight cancer, and radiation sterilization, namely the purposeful destruction of the genetic tissue of harmful microorganisms [75, 76].

4.3.2 Electron-beam Lithography

In electron-beam lithography a fine focused electron beam of a scanning type instrument is deflected and switched on and off across a wafer to expose the resist on the wafer by radiation damage according to a pre-designed pattern, and to open up windows in the resist for chemical and physical processes to construct electronic devices.

Background and Historical Development

Electron optics has been developed since 1930. Diffraction due to the wave nature of the electron having a wavelength λ and an axial aberration like spherical (C_3), and axial chromatic aberrations (C_A^F) determine the resolution in a scanning electron microscope [77, 78]. The optimum aperture is

$$\alpha_{opt} = 0.88(\lambda/C_3)^{1/4}.$$

A typical aperture ranges from 10^{-2} to 10^{-3} rad. The resolution obtainable with a probe forming lens is given by

$$d_{opt} = 0.75(C_3\lambda^3)^{1/4}.$$

An optimum current is obtained, having a 25% contribution of the aberrations to the probe diameter [79]. The probe current is limited by the Coulomb interaction [80]. The best objective lens of a scanning electron microscope for low energy electrons is presently the combined electrostatic and magnetic lens [81]. It renders a spot diameter of 1.7 nm at 1 kV, and a current density of >1000 A/cm^2 and allows very high contrast imaging at very low noise levels [82].

Throughput: A Driver for Beam Current Density and Resist Sensitivity

The obtainable current density is calculated with the numerical aperture α of the objective lens by

$$j = B/\Omega = B/\pi\alpha^2 \text{ [A/cm}^2\text{]},$$

where B is the brightness of an electron source, $B = I/F \cdot \Omega$ [A/cm^2 sr] [83].

The exposure time per area t at resist sensitivity S depends on the current in the electron probe according to S = current I · time t/area F [A s/cm^2]. The exposure time is therefore $t = S/j$ [s]. For this time interval the beam has to stop at a pixel. Most e-beam systems have a 5 to 10 MHz pixel clock frequency, a numerical aperture of 10^{-3} to 10^{-2}, and $j = 20$ A/cm^2 current density [84]. Industrial lithography systems use the Zr-O-W thermal field emission cathode, which delivers a confined beam at high brightness of 10^7 A/cm^2 sr and has a low energy broadening [85]. Typically 50 to 100 kV and power densities up to 2000 A/cm^2 are used. The throughput of electron beam lithography systems was increased from vector scan to raster scan to variable shaped beam and is highest with cell projection.

Knowing the sensitivity of the resist S [A s/cm^2], the current density j [A/cm^2], the beam current I [A], and the exposure time t [s], the productivity G [cm^2/s] is given by $G = I/S$ [cm^2/s]. To produce 30 wafer/h in a EBL projection system the required total beam current is 6.08 µA at a resist sensitivity of $S = 1$ µA s/cm^2. Mask writers use a beam current of nA or in high resolution tasks 100 pA, and are accordingly much slower.

The Electron Beam Exposure Process

To structure the surface of a silicon wafer or an already patterned wafer, this is planarized by mechanical polishing and spin-coated with an electron sensitive organic polymer from a solution in organic solvents. By electron exposure, the resist molecules suffer chain-scissoring or cross-linking of fractions of molecules to larger polymers. Both results lead to a changed solubility of the resist layer and allow either to open holes in the resist at the exposed site (positive resist) or to solve all unexposed material and leave the polymerized material covering the exposed area (negative resist). Typically PMMA (poly-methyl-metacrylate, with a molecular weight of 250 000 or up to 900 000) is a positive electron beam resist when using 50 keV electrons and a dose of 100 µC/cm^2. PMMA turns in a negative resist above 1000 µC/cm^2. PMMA is developed in MIBK (methyl-isobutyl-ketone) and IPA (isopropanol) in a 1 : 3 mixture for 90 s at room temperature [86].

Novolak based resists have similar sensitivities, but are developed with inorganic solutions. The exposure releases acids inside the resist layer and initiates the chemical amplification of the polymerization for a developable result. Such resists are called chemically amplified resists (CAR) with sensitivity of 5 μC/cm^2 [87]. All high-energy electron beam exposures suffer from proximity effect, which is an additional exposure by the forward- and backscattering of the electrons inside the resist and in the underlying layers of the substrate [88]. The dose variation by proximity exposure significantly changes the CD of the lines and patterns. Today 100 kV electron beam lithography systems are available to reduce the proximity effect.

Lithography for Chip Production

According to Moore's Law [89], the production of circuits will reduce the critical dimension (CD) or line width every 1.5 years by 0.7 to fit more devices into the same area and so maintain productivity at increased circuit speed. Optical steppers of today render a throughput of 30 wafers per hour, with wafers of 300 mm diameter having an active device area of 600 cm^2.

The production of sub-quarter micron lines became a matter of phase contrast imaging [90]. This allowed to produce 2001 patterns with 248 nm $<$ CD $<$ 124 nm, which is the 130 nm node. The phase contrast intensity distribution for proper resist exposure is obtained using optical proximity effect structures (OPC), which were introduced into the mask being much smaller than the CD [91]. This requires exposure, metrology and repair systems of higher resolution and therefore higher cost, like soft X-rays at 13.6 nm wavelength which is called extreme ultra violet lithography (EUVL) [92] and step and flash imprint lithography (SFIL) [93]. All these technologies need a stamp or mask with nanometer structures, which in turn will be produced with electron beam mask writers.

Up to 30 layers of processes and exposures are required to fabricate a SOC or Pentium computer chip, which are e.g. 10 of high resolution, 10 with middle resolution, and 10 of low resolution. All fabrication steps of today use optical or other masks.

Electron beam lithography systems have been developed since 1970 by modifying scanning electron microscopes to computer controlled electron beam systems. Most successful was the application of raster scan [94]. Mask making systems use variable shaped beam (VSB) systems as well as cell-, or character-projection [95, 96]. Mark detection and exposure control was added, which is a computer controlled beam positioning system with 16 bit D/A converters, and a pattern generator to delineate the shapes. A beam blanker switches the beam on and off in ns. Shapes are subunits which originate from the CAD geometries by subdivision of the geometries into areas, which can be addresses with an x- and y-digital coordinates system, and which are to be exposed with the same electron dose. Shapes compose a cell, several cells compose a subfield, a number of subfields compose a field and several fields finally compose a layer pattern of a device or dye. The control software has

to line up the data, the exposure conditions, and stage positions to expose the resist-coated wafer or mask blank by beam writing with the information pixel by pixel.

The CAD data files are mostly translated into GDS-II format [97] and contain the full description of the device in many layers, which are to be exposed, to open windows for the successive processing steps like metallization, doping, or etching. Depending on the electron energy used in the lithography system, the layers have to be processed with a proximity correction software [98], to assure proper shape subdivisions with equal dose to generate the required dose variations inside a shape for proper dose distributions in the resist. This is necessary to render the wanted CD after the development. Depending on the type of beam-writer, each CAD-layer is incorporated into a job description file, which takes care of the machine settings like electron energy, current, focus, spot size, writing speed, beam step, mark detection contrast, and signal evaluation. To place the pattern correctly on the wafer, or on the top of existing other layer structures, global and local marks are measured. The wafer is moved by a mechanical stage with position control by a laser interferometer measurement system and feedback to the beam position for fine corrections. Systems of today achieve $\lambda/1024 = 0.61$ nm positional accuracy of the wafer.

State of the Art Equipment: Direct Write Systems

A schematic setup of a cell projection system with all components is given in Fig. 4.40. Lines, single pixels and cells can be selected with a two-aperture method

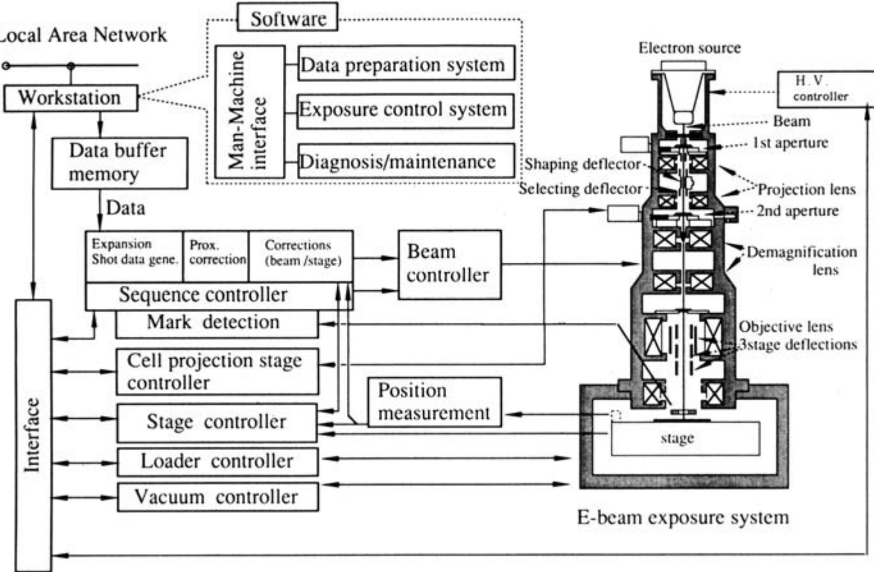


Fig. 4.40. Schematic setup of a cell projection system with all components showing the complexity of such lithography exposure systems (courtesy of Saito, Hitachi)

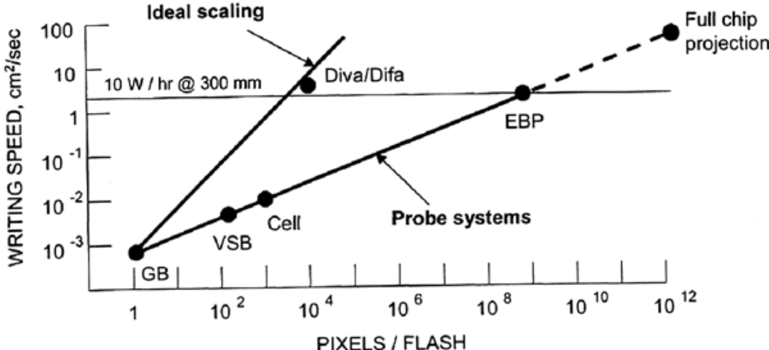


Fig. 4.41. Comparison of present and future electron beam lithography systems for direct write. Speed of lithography systems at 10 μC/cm^2 resist sensitivity: *GB*, Gaussian round beam; *VSB*, variable shaped beam; *Cell*, cell projection; *EBP*, electron projection; *DIVA*, multi-axis exposure system (Leica); and *MAPPER*, multi-beam direct-writing system

and respective mask patterns, and are then placed to the required position by the deflection systems.

To avoid the thermal bowing of the membrane mask, the scattering absorption contrast of a very thin mask is used in cell projection systems of Hitachi. System on chip dyes (SOC) have dimensions of 25×45 mm^2 $= 11.25$ cm^2, that allows to produce 50 SOC chips on one wafer. These throughput values have to be met by all the NGL lithography systems.

Future Aspects: Lithography Systems of Higher Productivity

Electron or ion beam maskless writing and production lithography schemes are now developed to avoid costly masks. Many beams are used in parallel, but being separated enough not to overlap the source images [99]. DIVA is investigated by Leica [100]. A competitive technology is MAPPER (Mapper-TU Delft).

Figure 4.41 gives the writing speed of present and future electron beam lithography systems, which are developed to meet the throughput requirements of a computer chip factory, where optical steppers and e-beam writers are working in parallel.

4.3.3 Electron Beam-induced Processing

Background and Historical Development

Electron beam-induced processing EBIP includes deposition (EBID) and etching (EBIE) of materials. It is a quickly evolving field for realization of generating, milling and patterning processing steps on the nanometre scale driven mainly by the needs of semiconductor industries and nanotechnologies.

Table 4.7. Precursors for EBIP, induced deposition, etching and reached achievements

Precursor	Product	Application and achievement
Pump-oil	Contamination	Metrology of probe astigmatism, image storage, masks for ion milling, zone plates for X-rays
WF6, carbonyls	Metal carbides	Mirrors, conductors, tips, etch masks, mask repair, STM lithography
Metal acac's	Metal carbides	3-D-AFM probe tips, 3-dimensional deposits, conductive wires, field emitters
Organic styrene	Diamond-like carbon	Carbonaceous quantum dots, carbonaceous
HMDS, HMCTS, TEOS	Siliconoxide, SiC, Silicon dioxide	AFM-tips (10 nm size, insulating, smooth) insulators, phase shifters
XeF_2, halogenes	Volatile metal fluorides	Etches MoSi, SiO_2, TaN, Si
$TMCTS/O_2$-O_3	Silicon oxide	Insulators

The roots of EBIP date back contaminations found in microscopes and on samples generated from gaseous film deposits in the systems. Depending on its chemistry, the film can get polymerized or degraded under the influence of the impinging electron beam. In the early electron beam microscopes, the observed contaminations were composed of non-volatile crack products of the vapour pump oil, i.e. carbon at the most. Nowadays, despite much better vacuum techniques, EB induced contamination is still a barrier to high-resolution microstructure analysis unless the specimen's surface is carefully cleaned (e.g. by ion etching in UHV) of residual adsorbents result from preparation at ambient atmosphere [101].

However, contamination was investigated to determine the highest possible information density obtainable by lithography methods in 1969 [102]. The scanning electron beam cracked hydrocarbon layers, and this was used as a protective mask during reactive ion etching lithography and generated 80 Å wide metal lines [103]. Next, precursors of different composition were introduced in order to evaluate the potentials of the new patterning technique and for direct writing processes SEM [104]. It was also found that with certain precursors electron-beam-induced etching was possible by the reaction of the cracked precursor with the substrate forming volatile compounds [105]. Table 4.7 gives a view of various precursors and achievements. Today, EBIP is developed to maturity for commercial use in repair of open and opaque defects in photo masks for lithography [106].

Fundamentals

Deposition with nanometre resolution is employed in scanning electron microscopes, in dedicated lithography systems, in reducing image projection systems, and in scanning tunnelling microscopes. No resist is needed as generally used for lithography [107]. The uncoated sample is inserted into the microscope. Gaseous substances are brought from reservoirs where they were sublimed or vaporized through nozzles into a vacuum chamber containing the sample on which deposition or etching

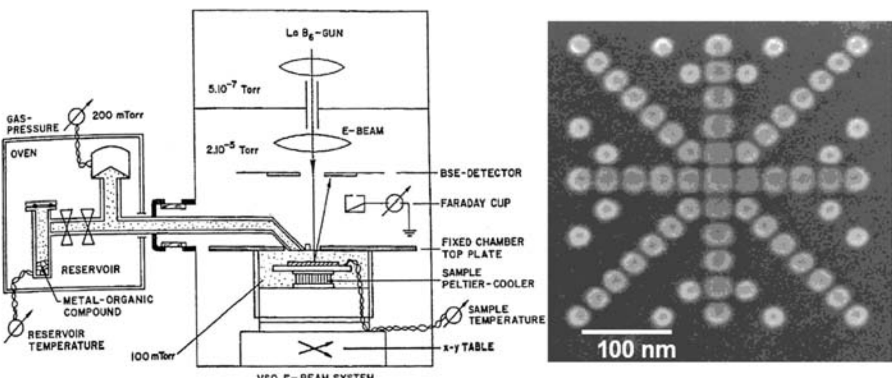


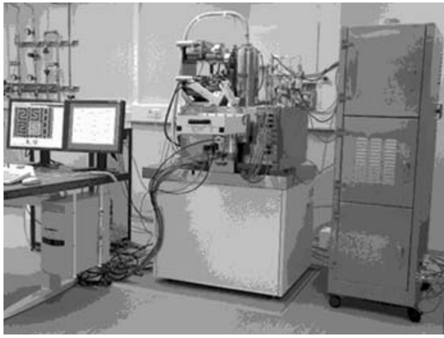
Fig. 4.42. *Left*: system for EBIP [108]. *Right*: platinum/carbon deposition on Cr reveals; 20 nm features, <4 nm edge roughness, <10 nm drift/min, highly accurate pattern generation (with MeRiT MG mask repair tool, courtesy of NaWoTec)

processes are to be performed. The nozzle forms a molecular beam of precursor molecules which is directed to the sample. Molecules of these vapours adsorb to the sample. A focused electron beam impinging on the top of the sample initiates reactions. The electron beam (current: pA to several nA, electron energy: 1 to 30 or even 100 kV) is focused to a spot with a power density of up to 60 MW/cm^2 [107]. Under such a strong radiation dose the atoms in the precursors are excited and the molecules are polymerized or even destroyed. Figure 4.42 shows such an e-beam processing system, left, and a high resolution deposition pattern, right.

Generally about 20 electrons are required to deposit 1 atom. For deposition processes this means that molecules of the vapours (precursors) dissociate into deposited and volatile fragments. The excited atoms react to permanent residues in the form of stable phases of metals (as nanocrystals) and of carbon (as fullerene parts or amorphous diamond-like deposits). For etching, reactions between substrate and gas molecules are initiated. Volatile reaction products are formed from the sample material and leave the surface into the vacuum [108].

The driving force behind these reactions is not only the electrons of the primary beam but mainly the generated secondary electrons and also backscattered electrons. Deposition and etching is faster for lower acceleration voltage with its higher secondary electron yield compared to a higher voltage. Deflecting the beam with a well-defined dwell time per pixel over the surface allows generating zero – to 3-dimensional features on the surface. Growth is always upward from the surface towards the beam here. Other experiments have shown the ability to grow material perpendicular to the beam across a hole, as it is required for the repair of stencils mask for electron-beam-projection lithography.

Controlling the exposure time per pixel with the dwell time control option of the computer control of the beam and employing beam blanking, the generation of three-dimensional structures even with arbitrarily inclined surfaces is possible. Objects with high aspect ratios up to 20 can be fabricated using this technique. The substrate



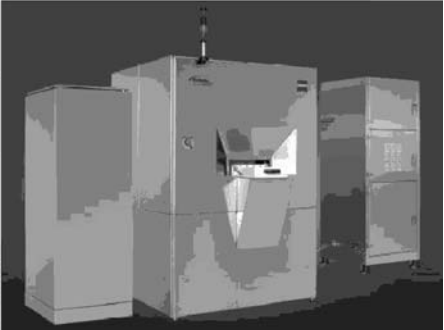


Fig. 4.43. The electron beam photo mask repair tool MeRiT MG. *Left*: proof-of-concept tool. *Right*: production tool for the 65 nm mask generation (courtesy of NaWoTec GmbH and Zeiss SMT AG)

has to have a certain minimum electrical conductivity to prevent charge-up and the resulting beam defection.

For etching, halogen compounds or other inorganic precursors are used. With this technique the sample can be etched monolayer by monolayer and an etch pit or hole can be generated. This action is used to open up opaque defects in photo masks and stencil masks.

State of the Art: Equipment and Applications

Typical tasks in mask repair are depositing bridges (for repair of stencil masks) and etching openings into absorber materials of optical lithography (as TaN for EUV lithography at 13.6 nm wavelength of soft X-rays, or MoSi for 193 nm deep UV lithography).

Figure 4.43 shows the electron beam photo mask repair tool MeRiT MG. It makes use of the described deposition and etching processes for repairing defects in 6″ photo masks. It supports the 65 nm generation photo mask manufacturing and is expected to be extendible beyond the future 45 nm node. The tool offers 3 nm spot size @ 1 kV for detection and classification of defects before repair and for inspection afterwards. A similar system for EBIP is now available from Raith called e-line.

Semiconductor industries roadmap is calling for further reductions in structure sizes. This sets the pace for development of EBIP technology beside to this the unique nanocompound materials (metallic nanocrystals immersed in a carbonaceous matrix) deposited via EBID from organometallic precursors will support research due to the fascinating capability to sustain current densities higher than 106 A/cm^2 without damage. Furthermore, an extended use of EBIP technology in nano-manufacturing and rapid prototyping is on the agenda.

4.3.4 Curing of Organic Coatings and Surface Layers

The objective of drying and curing organic materials is to convert a coating material (e.g. lacquer or printing ink) that has been applied to a substrate in a liquid state into a

solid state by curing with heat or other radiation (hot air drying, infrared irradiation). Such drying methods are slow, expensive and emit volatile organic substances.

In electron beam curing, the energy penetrates the coating material, and a portion of the energy is directly transferred to the electron shells of the molecules in the coating material and initiates polymerization and solidification of the coating material. These chemical effects take place in fractions of a second. Minimal heating of the coating material is merely a secondary effect.

In organic molecules, bond energies are typically of the order of 5 eV. This corresponds to the energy of a photon of electromagnetic radiation in the UV light region (5 eV $\leftrightarrow$ 250 nm). The first curing of organic coating materials using UV light were registered in 1938-46 [109]. In 1957 Charlesby and Wycherly reported the curing of unsaturated polyester resins with electron beams that were generated in a 2 MeV van de Graaff generator [110]. The introduction of lacquer curing using electron beams (300 kV electron generator) in the production of plastic interior trim components for cars (ashtrays, glove boxes, dashboard trim, etc.) at the Ford Motor Company represented the breakthrough for the industrial utilization of electron beam curing (abbreviation: EBC) in 1970 [111]. The advantageous processing and product properties of EBC [112–114] are: drastic reduction in energy costs, considerably lower temperature stress on products and high chemical yield. Chemical reactions have time constants considerably below 1 s. Feed rates up to 600 m/min have been realized in industry. Low space requirements for the curing station (a few meters^2 including X-ray shielding und inert gas feed). The coating materials are usually fully (100%) converted into solid materials.

No volatile organic compounds are emitted. The pot life of the coating materials is practically unlimited. Energy deposition is relatively uniform across the entire penetration depth of the electrons. Thick-film and pigmented lacquers of all shades can also be easily cured. Good adhesion of the coating material, especially on rough or porous substrates. The cured films are resistant to light, even for outdoor applications, since no photo-initiators are used. Due to the high ionization potential of the primary ($<$300 keV) and secondary electrons ($<$50 eV), the degree of cross-linking and hence the mechanical resistance (scratch resistance, hardness) is higher using EBC than what can be achieved with the photon energy ($<$10 eV) of industrially employed UV-lamps.

The EBC process is shown schematically in Fig. 4.44, left. An industrial coating unit for flat components, like wood flaps with substrate feed, printing and coating equipment and two EBC stations whose process chambers each comprise a low energy axial-type electron beam generator with a scanner and beam exit window, an inert gas system and local X-ray shielding is shown on the right.

Band-type EB generators possess a lateral dose homogeneity of $\pm$10%. This is poorer than the axial-type electron beam gun with a scanner ($\pm$3%) and compared to the latter can also not be adjusted across the width of the treatment zone. In case when the highest possible surface quality is required, it is necessary to carry out the curing process under conditions of oxygen exclusion. Oxygen will trap free radicals generated at the surface. The permissible concentration of oxygen is largely determined by the composition of the coating material. Whilst an oxygen concentration of

up to 400 ppm for lacquers and up to 200 pm for printing inks is usually tolerated in the process atmosphere, values below 50 ppm are desirable when silicone treatment is being carried out. As an protective inert gas usually pure nitrogen (produced via evaporation of liquid nitrogen), or a CO_2–N_2 gas mixture is used.

EBC is used in the following products for curing high-solid and powder lacquers, adhesives and sol-gels:

Wood products: furniture veneering, parquet, laminate, doors;
Construction parts: facade elements, window frames, car wheel rims;
Plastic films: packaging with gas barriers and tightness to aromas;
Semi-finished products: strip steel for instrument casings and wall cladding;
Printed products: paper webs, metal foils, tins;
3D-car body structures: robot arm positioned small gun and scanner unit.

Figure 4.45 shows left a low weight, low energy axial-type electron gun combined with rapid beam deflection and a beam exit window. The electron source can be

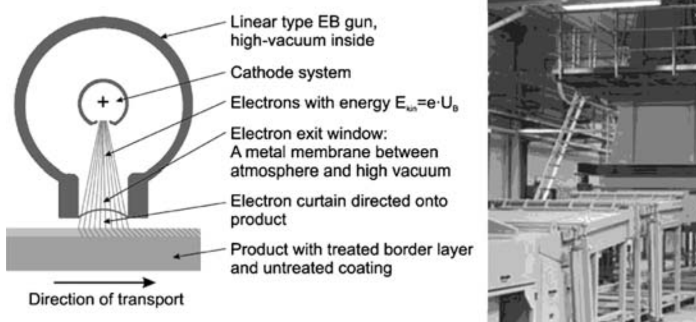


Fig. 4.44. *Left*: principle of electron beam curing. *Right*: industrial coating unit for flat components with substrate feed, printing and coating equipment and two EBC stations whose process chambers each comprise a low energy axial-type electron beam generator with scanner and beam exit window, an inert gas system and local X-ray shielding (courtesy of Scannery Holztechnik GmbH)

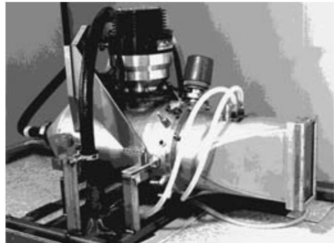


Fig. 4.45. *Left*: curing of 3D free-form surfaces with a robot-operated miniature electron beam gun with $U_B = 150$ kV, scanner and Ti-beam exit window of 200×70 mm^2. *Right*: 3-dimensional car body part (courtesy of Fraunhofer FEP)

mounted on an industrial robot and be moved like a “brush” to cure 3D-surface structures. This allows extended free-form areas to be successively treated at ambient pressure with the required dose of energy, see Fig. 4.45, right [115].

4.3.5 Refinement of Polymer Materials

Electron-induced modification of materials applies penetration depths up to several cm, and therefore employs electron beams of $e \cdot U_B = 300$ keV to 10 MeV. Linear accelerators fed by a van de Graff high tension ribbon generator were used. Today 200 keV axial type guns with a scanner and window are in commercial use to manufacture shrink films, shrink bags and for cross-linking cable insulation materials. Fundamental work on the theory of electron beam cross-linking was published by Charlesby and Chapiro [116, 117] in 1960 and 1962. Non-thermal EB refinement of polymers is utilized to enhance chemical, mechanical and thermal properties of suitable cheap raw materials at high throughput, excellent precision and high reproducibility. EB refinement of today is materials treatment in reactive atmosphere and in electron beam induced plasmas, and is gaining importance as materials engineering tool [118].

Polyethylene and rubber are two of the major polymers on which cross-linking is carried out. The cross-linking leads to covalent bonds (C–C) being formed between adjacent chains leading to a 3-dimensional network. The applied dose is in the range 20 to 200 kGy, e.g. to set the gel-fraction of polyethylene (PE). Above the gel point this results in formation of an insoluble polymer fraction. The quantity of insoluble polymer increases with the energy dose up to a limit value that depends on the ratio between scission and cross-linking (Fig. 4.46).

With suitable choice of the starting materials, the cross-linking generally requires no additives and no elevated temperatures. However, the use of additives or polymer mixtures often allows the cross-linking reaction to be favorably influenced. It must also be borne in mind that side reactions such as molecule decomposition, formation of double bonds, release of gases and discoloration can occur depending on the particular reaction conditions employed. This particularly concerns the local gas at-

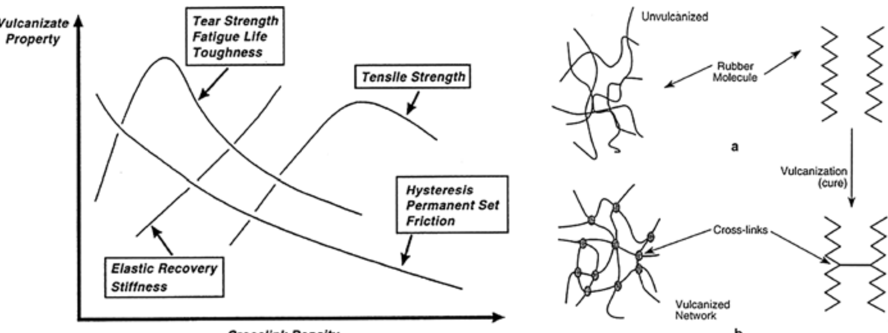


Fig. 4.46. *Left*: some material’s properties of rubber in dependence on cross-link density. *Right*: schematic of the cross-linking mechanism at molecular level (after [119])

mosphere, the chemical structure and the physical state of the polymer being cross-linked. There are therefore requirements on both the starting materials and on the management of the electron treatment process in order to achieve efficient, high-quality cross-linking.

State of the Art: Equipment and Applications

The cross-linking significantly changes the properties of the polymer material, and is currently one of the main applications of electron beam technology. Examples of the use of electron beam technology for materials refinement by industry include:

- cross-linking of polymers in the wire and cable manufacturing industry,
- cross-linking of underfloor heating pipes made of polyethylene,
- vulcanization of elastomers in the rubber processing industry,
- pre-cross-linking of adhesives in self-adhesive and release films,
- manufacture of shrink products and plastic packaging films,
- curing of laminate materials,
- curing of hydro-gels and biopolymers,
- grafting of reactive monomers/prepolymers onto substrate polymers.

In most industrial applications, the cross-linking of materials is usually carried out after the thermoplastic molding. Due to the medium energy electron and the produced higher fraction of high-energy X-rays, equipment systems must have well-shielded accelerator and treatment chambers. Meter-thick concrete walls and meander-shaped labyrinths are required to feed the products to and from the machine.

Future applications include the reactive modification of fluids, pellets and powders, tailoring of native, renewable materials, purification of gases and liquids, and modification of recycled polymers.

4.3.6 Control of Pathogens

Background and Historical Development

The term “control of pathogens” refers to the conversion of a product that is afflicted by microorganisms (e.g. bacteria, yeasts, moulds) or insects into a sterile state having an acceptable and usually legally prescribed germ load. Irradiating the products with high energy electrons effectively and safely reduces the original load. This technology is used worldwide in the area of medical technology and foods and has gained increasing importance in recent decades.

The first commercial electron beam plant with 2 MeV for sterilizing catgut sutures and treating foods was installed in 1955 by Ethicon in Scotland and in 1957 by a German spice manufacturer [120]. The design of industrial plants dedicated to sterilization or food preservation tasks is very similar to those used for curing and refining of polymers and employing high energy electrons of 2 MeV or middle energy electrons of 200 keV. Direct effects by the primary physical interaction is the direct modification of biomolecules. Indirect effects are biological effects caused by the products (e.g. H and OH radicals, peroxides) generated via radiation-chemical induced reactions in the embedding medium (e.g. water).

The harm that can be caused increases with the complexity of the organism. With single-celled organisms, the dominant effect is chemical damage to the cell components. The damage is more permanent when it involves the cell nuclei. In more complex organisms, minor damage to in particular the nucleic acids and enzymes can lead to disfunction of the material and energy metabolisms. These biochemical structural changes are not necessarily lethal immediately, but can in the longer term lead to a wide variety of adverse effects (e.g. membrane malfunction, suppression of cell division, mutations if there is damage to the chromosomes). Serious changes can cause cell death and can affect the proliferation of cells.

The test whether visible colonies are still able to develop in irradiated cell cultures is used as a criterion for the effectiveness of the radiation treatment. The biocidal effect of accelerated electrons on microorganisms is represented by deactivation curves. Here, the number of culture forming microorganisms still present on a treated product (usually indicated as the number of *culture forming units* = cfu) is plotted against the energy dose. In the simplest case these curves can be described by an exponential function $N(D) = N_0 \cdot \exp(-a \cdot D)$, where N_0 and $N(D)$ represent the number of microorganisms before and after treatment with a dose of energy D. The constant a describes the sensitivity of the microorganism. Figure 4.47 shows the deactivation curves for Bacillus subtilis and HIV virus versus the dose. Due to the bio-diversity of the different microorganisms, the profiles of the deactivation curves are generally more complicated than the examples shown here.

The profiles of the deactivation curves for electron treatment are flatter than the typical dose-effect curves for chemical poisons but indicate an effect for even the smallest doses. The deactivation curves show that there is always a certain probability of a microorganism surviving. Therefore, the sterility is defined as the probability of the presence of a surviving microorganism on or in a product (*Sterility assurance level*, SAL). For sterilized medical products in their final packaging, an SAL of 10^{-6} is required. In this case, this means the energy dose must be chosen so that for every

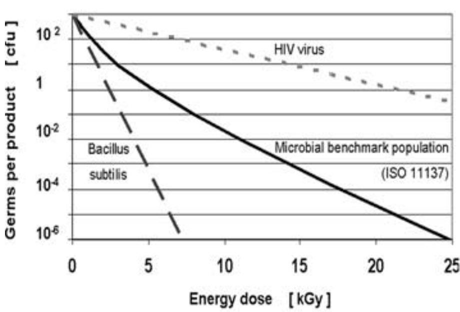


Fig. 4.47. *Left*: deactivation curves for a microbial benchmark population according to ISO 11137, for Bacillus subtilis and for the HIV virus. *Right*: pathogens control by electron treatment

Table 4.8. Effects and applications of electron treatment as a function of the energy dose applied

Dose range	Purpose of treatment
Low-dose range	Stimulation of growth/useful mutations (plant breeding)
(<1 kGy)	Insect disinfestation (grains, flour, fruits)
	Inhibition of sprouting (potatoes, onions)
Medium-dose range	Delay of ripening (vegetable seasonings)
(1 ... 10 kGy)	Inhibition of spoilage (meat, fish, fruits)
	Reduction of pathogenic bacteria/fungi (waste, spices)
High-dose range	Sterilization of medical devices and materials
(10 ... 50 kGy)	Deactivation of viruses (e.g. for vaccines)
	Preparation of sterile diets (hospital meals, space food)

1 000 000 packages there is, on average statistically, only one having a surviving germ after sterilization [121].

The sterilization of medical products and germ reduction in foods are currently important commercial applications of electron beam technology. The industrial operations meet the highest hygiene standards and efficiently produce products which are harmless to health and which can be stored for longer periods. During the treatment the products are often in their dispatch-ready packaging, so preventing renewed contamination with germs. Effects and applications of electron treatment as a function of the energy dose applied are given in Table 4.8.

In recent years the sterilization of medical products with energy-rich electrons has increasingly taken over from competing methods such as gamma-irradiation and chemical sterilization with ethylene oxide (EtO). In contrast to gamma-radiation, where radioactive preparations are used as the radiation source, electron beam plants are so-called machine sources where, for example, for maintenance work the production of ionizing radiation can be quickly and fully stopped by shutting off the power supply.

Electron beam treatment of food is permitted by the national health authorities in over 40 countries to improve the hygienic state of foods by reducing or deactivating pathogenic germs, organisms and insects and also to considerably reduce the risk of food-related illness. Series of tests over many years have demonstrated that for a maximum permitted dose of 10 kGy the nutrient and vitamin content is not adversely affected in any way and that the treated foods are harmless from toxicological and microbiological viewpoints [122]. Recently a mobile station has been developed with a dressing unit for seeds based on e-ventus® technology [123]. The process unit consists of two low-energy band-type EB generators ($U_B \leq 150$ kV) with beam exit windows arranged opposite one another. Between the two EB generators a homogeneous "electron curtain" having a working width of 1400 mm and with a total power of up to 60 kW can be realized. The seed throughput is 30 tons per hour. The seeds fall through air between the two windows, whereby the surface of each grain is reached by the electrons and is efficiently disinfected. This technology can avoid that every year, as many as 1000 tons of chemical dressings are introduced into the

ecosystem in Germany, for wheat alone. In Germany the technology has been classified as very environmentally friendly by the German Ministry for Agriculture and Forestry and its use has been specifically recommended [124].

4.3.7 Flue Gas and Wastewater Treatment

Over the last few decades, electron beam technology based techniques have been developed, demonstrated and deployed to alleviate some of environmental problems associated with gaseous and liquid effluent waste. Extended uses of electron beam technology for environmental remediation includes application of radiation technology for removal of volatile organic compounds from exhaust gases, purification of drinking water, reclamation of effluent from municipal wastewater treatment plant and hygienization of sewage sludge for use in agriculture.

Energy production generates pollutants, such as fly ash, sulphur oxides (SO_2 and SO_3), nitrogen oxides ($NO_x = NO_2 + NO$) and volatile organic compounds, into the atmosphere. High-energy electrons interact with gas creating divergent ions and radicals including oxidizing radicals and excited species. These excited species react in a various ways of neutralization reactions and dimerization. The removal of SO_2 using electrons was demonstrated in 1970–1971. Irradiation of flue gas resulted in the conversion of SO_2 to an aerosol of sulfuric acid droplets that were easily collected [125]. Ebara Co. used an electron accelerator (0.75 MeV, 45 kW) to convert SO_2 and NO_x into a dry product containing $(NH_4)_2SO_4$ and NH_4NO_3 that was usable as a fertilizer.

10 pilot and industrial scale electron beam flue gas treatment plants are in operation worldwide, with a typical flow rate of 20 000 or industrial 200 000 (Nm^3/hr) with accelerators of 800 keV delivering 160 kW and delivering a dose of 3 to 30 kGy to purify gas polluted with 1000/400 ppm SO_2/NO_x. NO is removed by oxidation and finally forms NH_4NO_3, a fertilizer [126]. In electron beam treatment the most important are radio-thermal and thermal process [127]. Radio-thermal reactions proceed through radical oxidation of SO_2 and finally result in $(NH_4)_2SO_4$ ammonium sulphate. Large-scale EB plants for flue gas cleaning have cost advantages over conventional technologies, and are constructed in China and Poland.

Electron beam processing of wastewater is non-chemical and uses fast formation of short-lived reactive radicals that can interact with a wide range of pollutants. Such reactive radicals are strong oxidizing or reducing agents that can transform the pollutants in the liquid wastes. The first full-scale application was reported for the purification of wastewater at the Voronezh synthetic rubber plant in Russia. Two accelerators (50 kW each) were used to convert the non-biodegradable emulsifier, 'nekal', present in the wastewater to a biodegradable form [128]. The installation treats up to 2000 m^3 of effluent per day. A pilot plant of 1000 m^3/d for treating textile-dyeing wastewater has been constructed in Daegu, Korea, with 1 MeV, 40 kW electron accelerator [129].

Wastewater treatment plants using high-energy irradiation produce instantaneous radiolytical transformations by energy transfer from accelerated electrons to orbital electrons of water molecules. Absorbed energy disturbs the electron system of the

molecule and results in breakage of inter-atomic bonds [130]. Hydrated electron eaq$^-$, H atom, ·OH and HO_2· radicals and hydrogen peroxide H_2O_2 and H_2 are the most important products of the primary interactions (radiolysis products) ·H_2O eaq$^-$, ·H, ·OH, HO_2·, H_2O_2, H_2, with yields (G value, μmol/J) of 0.28(eaq$^-$), 0.062(H), 0.28(OH), 0.072(H_2O_2).

High reactivity is a characteristic of water radiolysis products. Typical time of their reactions with the impurities in water is less than 1 microsecond. Hydrogen peroxide H_2O_2 and ·OH and HO_2· radicals are oxidizing species, while H atom and eaq$^-$ are chemical reducing in nature [128]. Simultaneous existence of strong oxidants and strong reductants within wastewater under treatment are remarkable and important characteristics of radiation processing. Reactions with unsaturated hydrocarbons, ·OH adds onto the double bond and, upon reaction with alkyl-compound, the ·OH captures H-atom to produce $CH_3CO{\cdot}CH_2$.

Hydrated electron eaq$^-$ and H atom are strong reductants. In reactions involving cations of metals, eaq$^-$ is able to produce neutral atoms and ions having anomalous valency. Typical reactions of eaq$^-$ consist of e^--addition to reagent. In its reactions with organic compounds, ·H shows higher reactivity than eaq$^-$. Reactions with unsaturated hydrocarbon consist of ·H addition to double bonds.

Generally, radiation processing of wastewater has maximum efficiency at pollutant concentration less than 10^{-3} mol/l (100 ppm). The treatment of such wastewater is simple, requires low dose (about 1 kGy or less) and gives almost complete elimination of odor, color, taste and turbidity. In general, pollutant transformation involves the chain oxidation, formation of insoluble compounds, coagulation of colloids, and enhancement of pollutant biodegradability.

4.4 Materials Analysis, Imaging and Testing

4.4.1 Scanning Electron Microscopy

The principle of the imaging system of a scanning electron microscope is given in Fig. 4.48, left [131]. The crossover of the electron gun is imaged at reduced scale by the condenser and objective lens to the sample. The fine focused beam is scanned by a double deflection system in a raster scan across the surface of the sample. The astigmatism of the probe is corrected with stigmator coils placed close to the objective lens. Backscattered electrons (BSE), secondary electrons (SE) or other signals like X-rays or photons from cathodo-luminescence released by the primary electron beam (PE) are generated by the electron–sample interaction and are detected with special detectors or signal amplifiers (sample current SC) and displayed on TV tubes or flat panel screens. The displayed image is also displayed in a camera tube for hardcopy or stored in an image memory of the control computer. The generally used contrast is the topography contrast. It originates when secondary electrons are generated by the primary electrons and leave the sample. The secondary electron emission is $\sim \cos\gamma$, with γ the angle between primary beam and sample surface. BSE images reveal materials contrast, see Fig. 4.49, like X-ray images of a specific spectral line, or cathodo-luminescence images, see Fig. 4.50.

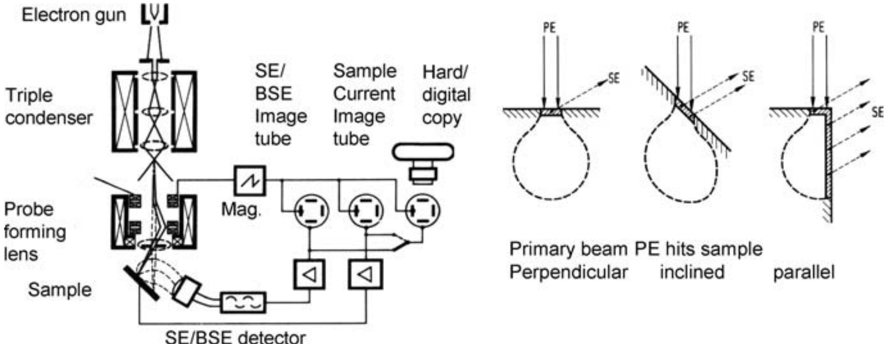


Fig. 4.48. *Left*: scanning electron microscope, schematic. *Right*: origin of topography contrast

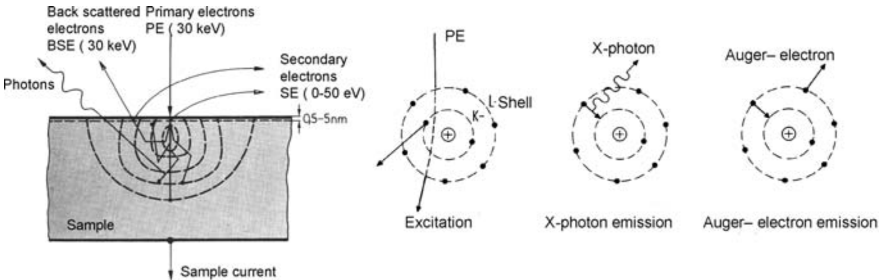


Fig. 4.49. Signal generation by primary electrons: elastic scattering gives back scattered electrons *BSE*, inelastic scattering releases secondary electrons *SE*, ionization from inner shells generates X-ray photons and by internal conversion also Auger electrons (*right*). Outer shell ionization generates cathodo-luminescence photons CL

Signal generation by primary electrons is explained in Fig. 4.49. Elastic scattering of primary electrons PE gives backscattered electrons BSE, some can leave the sample at full primary electron energy. Inelastic scattering releases secondary electrons SE. Ionization from inner shells generates X-ray photons, and by internal conversion also Auger electrons. Outer shell ionisation generates cathodo-luminescence photons CL.

Materials contrast is obtained by either BSE electron imaging or X-ray or CL-spectroscopic. Figure 4.50 gives a SEM and three X-ray images showing a defect site in a Ta thin film capacitor in 3 images taken with a wavelength dispersive spectrometer, top left. A chromium adhesion layer of the gold upper electrode (not shown) is shown in the bottom left. The Ta/Al oxide layer and the Al/Ta ground electrode are evaporated, top right, but not the Ta-oxide layer on the substrate, lower right.

Examples of surface topography (a), 3D imaging capability (b) and materials contras by BSE imaging (c) are shown in Fig. 4.51. The back-scattering coefficient is $\sim Z^2$, with Z the atomic number of the element. Therefore, high Z materials give a bright signal, whereas low Z materials appear dark in BSE imaging. The secondary

electron detector after Everhard and Thornley can be used as back-scatter electron detector, by switching off the accelerating field between scintillator entrance grid and sample. In this case only the high-energy backscattered electrons can hit the scintillator and are detected. Figure 4.51c shows phases of manganese (25), tellurium (52), and bismuth (83).

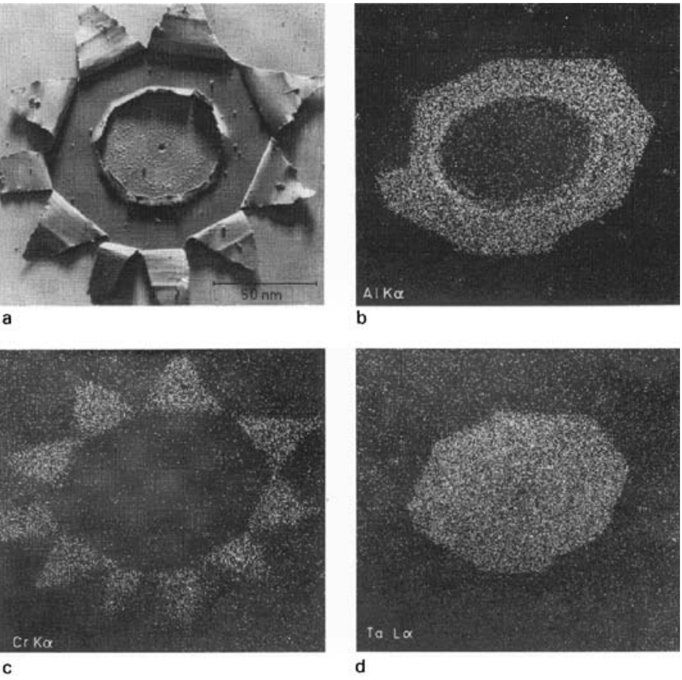


Fig. 4.50. SEM and three X-ray images showing a defect site in a Ta thin film capacitor taken with a wavelength dispersive spectrometer. *Top left*: SEM image (**a**). *Bottom left*: chromium adhesion layer under the gold upper electrode (not shown) (**c**). *Top right*: Ta/Al oxide layer and the Al/Ta ground electrode are evaporated (**b**). *Bottom right*: Ta-oxide layer on the substrate is still intact (**d**) (courtesy of Schauer, Siemens AG)

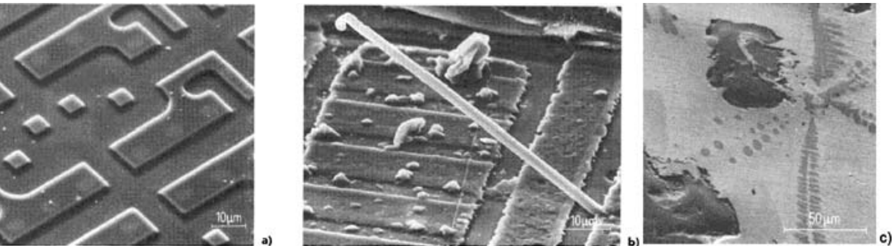


Fig. 4.51. **a** imaging with SE for surface topography, **b** SE-3D imaging capability, a whisker is shown, **c** materials contras by BSE imaging Mn, Te, Bi

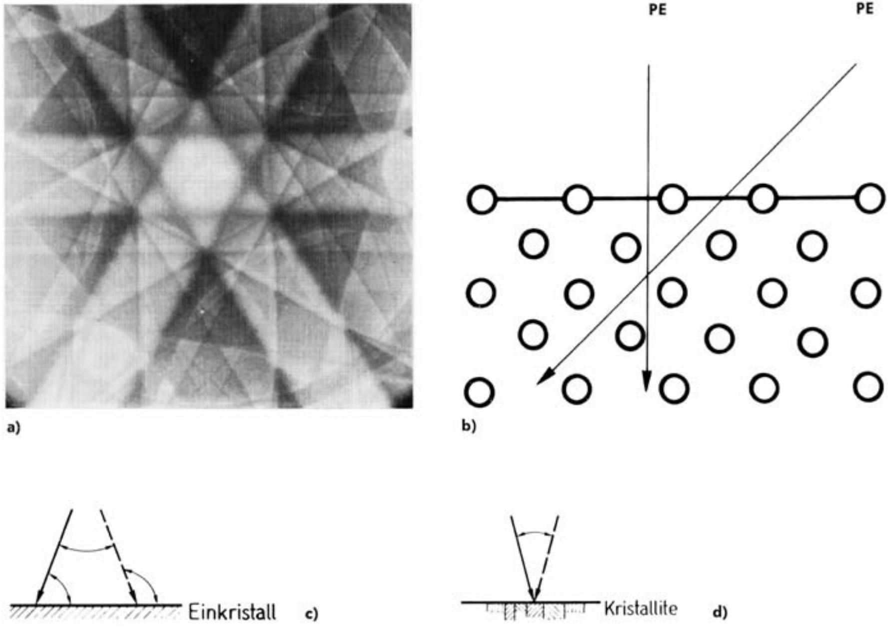


Fig. 4.52. Electron beam channelling diagram generated with a rocking beam allow to determine the orientation of single crystals, even in a polycrystalline matrix

Using a rocking beam, which is a beam focused to the sample but changing its aperture and not its position electron beam channelling diagrams can be generated, which allow to determine the orientation of single crystals, see Fig. 4.52. If electrons penetrate in between the atomic net planes, they do not scatter, but electrons hitting the atom chains do.

Electron sources have been reviewed in Sect 4.3.1. Figure 4.53 shows an expanded field emission SEM system for materials analysis with X-ray spectroscopy and image processing.

4.4.2 Electron Microscopy

Historic Development of Microscopy

Siemens built the first commercial electron microscope (1939) having more than 1000× magnification and a resolution $<0.3\ \mu m$. In 1945 a resolution of 1 nm was reached, see Fig. 4.54 where the first experimental and commercial microscopes, and also a modern spherically corrected microscope for sub-Angstrom resolution are shown.

After a series of high resolution microscopes from Siemens, Zeiss, Hitachi, JEOL, Philips operating at 100 kV and reaching a resolution of 0.3 nm, high resolution microscopes were built operating at 200 KV and 1 MeV to reduce the influence of the

Fig. 4.53. Field emission SEM system with energy dispersive X-ray spectroscopy and image processing unit (courtesy of Hitachi)

chromatic aberration. Here a resolution of 0.14 nm was reached. To overcome the limitations by spherical and chromatic aberration, corrected lenses have been developed employing multipole aplanators [133, 134]. It took 20 years until commercial high resolution microscopes with a corrector for spherical aberration were available (JEOL, FEI, Zeiss) [135], see Fig. 4.54, right.

High Resolution Electron Microscopes

To build high resolution microscopes, several engineering tasks have to be solved. Electrons are scattered in matter. To generate an electron beam, a high vacuum or even ultra high vacuum tube is needed. Today oil-free membrane pumps evacuate the tube from atmospheric pressure to 10 mb. Turbomolecular pumps with magnetic bearings and attached with a compressing "Spiro" stage evacuate the vacuum system to 10^{-7} mb and compress the gases to 15 mb. Ion getter pumps finally evacuate gun and specimen stage areas to 10^{-10} mb [136]. This vacuum is only reached after a 2 days bake out process up to 180°C during pumping. For background free electron flow the mean free path of the electrons must be $\gg$ the microscope length. The residual gas pressure therefore must be $p < 1.1 \times 10^{-5}$ U/l mm mb/V, with l the beam length.

High tension stability is required to accelerate electrons with, e.g. 200 kV. Such microscopes now have short accelerators with 4 stages. The breakdown voltage/cm, the break-down field in air is 10 kV/cm, in high vacuum 100 kV/cm over polished stainless or titanium surfaces [137].

Voltage stabilizing electronics is required to regulate the high tension to $\Delta U/U < 10^{-6}$, and the lens currents to a current stability $\Delta I/I < 10^{-6}$. However, using corrected lenses, the stability requirements for the imaging lenses reach $\Delta I/I < 10^{-7}$. The standard reference voltage diodes are stabilized to 10^{-6} only. Therefore, spe-

Fig. 4.54. Ruska's first experimental (*left*) and Siemens first commercial microscope (*middle*) [131, 132], and a modern spherically corrected microscope (Zeiss) (*right*)

cial developments are required. Currents and voltages were stabilized to 10^{-7} and better [138].

The sample is illuminated like in the optical microscope with an image from the crossover of the electron gun, placed close to the sample by demagnification of the source crossover with a first condenser lens, and by transferring the demagnified image with a second condenser lens. The condenser aperture in the second condenser determines the diameter of the illuminated area. In a microscope with twin lens or single field condenser objective, a short focal third condenser lens in close vicinity to the samples allows to realise the well known "Köhler" parallel illumination in the TEM. Here the condenser 2 aperture serves as an illumination field limiting aperture. We have $\alpha_{\mathrm{b}} = r_{\mathrm{c}}/(DM_{\mathrm{K1}}DM_{\mathrm{K2}}l) \ll \alpha_{\mathrm{Objective}}$, with α_{b} the illumination aperture, r_{c} the gun crossover radius, DM_{K1} and DM_{K2} the demagnification factors of condenser 1 and condenser 2, and l distance of the demagnified crossover image after the second condenser from the sample. The illuminated area has a few µm in diameter.

Objective lens, intermediate lens and the projector lens image the sample at a magnified scale to the final screen. The usable magnification is calculated from the resolution of the screen , which is $d_{\mathrm{S}} \approx 0.1$ mm, or the photographic plate $d_{\mathrm{P}} \approx 0.03$ mm: $V_{\mathrm{usable}} = M_{\mathrm{obj}}M_{\mathrm{Intermediate\ lens}}M_{\mathrm{Proj}} = d_{\mathrm{L}(p)}/d_{\mathrm{Sample}}$. To see 1 nm details, a necessary magnification to the screen is 100 000, and to the photographic plate 30 000 is sufficient. To correct for astigmatism to the required extent, the usable magnification should exceed 100 000! The depth of focus $T = d_{\mathrm{object}}/\alpha$, with α the objective aperture. At $a = 10^{-2}$ rad and a 0.5 nm object detail the depth of focus is 50 nm, which exceeds the transparent sample thickness. Therefore, a TEM projects all details in the thin sample into one image. Figure 4.55 presents on the left the ray path for lensless diffraction (left), magnification imaging with 2 lenses for low magnification imaging (middle), and high resolution imaging (right). The condenser

system irradiates the sample with a radius r_B, a current density j_B in A/cm^2, and an aperture α_B, as given in Fig. 4.55, right.

An image of the sample is obtained if the sample is thin enough to transmit the electrons. Figure 4.56 shows the penetration depth for imaging of several material and electron energies. Electrons interact by inelastic and elastic scattering

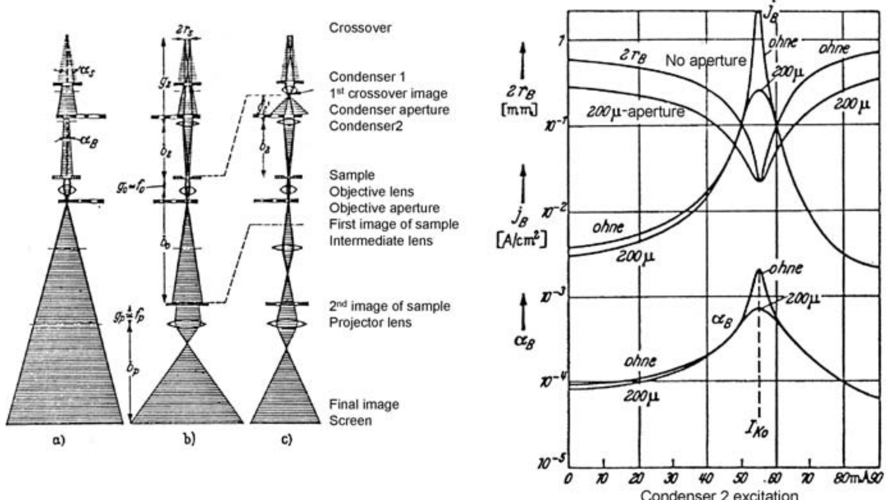


Fig. 4.55. *Left*: ray path for lensless diffraction (*left*), magnification imaging with 2 lenses for low magnification imaging (*middle*), and high resolution imaging (*right*) [139]. *Right*: characteristics of the illumination conditions; the condenser system irradiates the sample with a radius r_B, a current density j_B in A/cm^2, and an aperture α_B

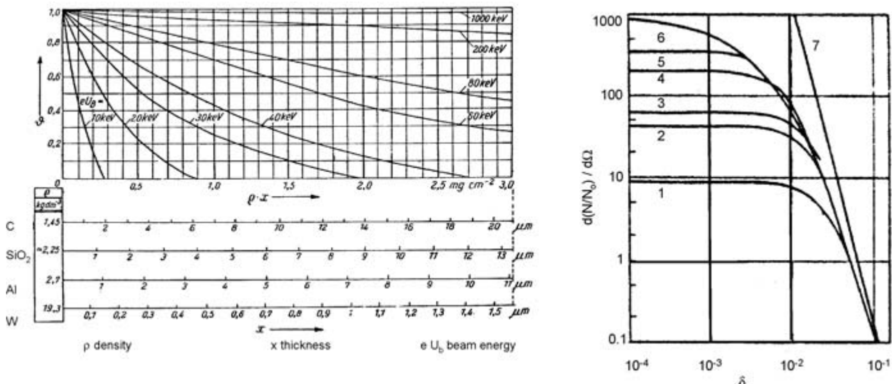


Fig. 4.56. *Left*: penetration depth for imaging of several material and electron energies. *Right*: angular distribution of the scattering intensity for elastic single event scattering of 50 keV electrons at carbon having a mass thickness of 10–6 g/cm^3 and calculated using various models (see text)

with the sample. The atomic number Z of the constituents determines scattering processes [140].

Scattering Absorption Contrast or Bright Field Amplitude Contrast

The image contrast is defined by $K = \log_{10} J_o/J$, with J_o A/cm^2 in the image without sample and J A/cm^2 in the image with sample. The contrast originates from the absorption and scattering of electrons in the sample and on the objective aperture α_o, which cuts out the electrons scattered in the angle δ. Figure 4.56 (right) gives the angular distribution of the scattering intensity for elastic single event scattering of 50 keV electrons at carbon having a mass thickness of 10^{-6} g/cm^3 and calculated using various models: 1 from the diamagnetic susceptibility (Lenz), 2 from the Hartree density distribution (Lenz), 3 after Bethe, 4 after Moliere, 5 after Bullard and Massey, 6 from the Thomas Fermi density distribution, 7 after Rutherford [141].

Figure 4.57, left displays the situation of scattering absorption contrast [142]. At constant aperture a and electron energy U the contrast is proportional to the mass thickness ρD: $K = C(\alpha, U, Z)\rho D$ with $C(\alpha, U, Z) = \log e/X_k$, where X_k is the contrast thickness, which is the mass thickness transmitting the e-part (37%) of the intensity through the aperture diaphragm. Using the atom model of Wenzel, it is $K_{\text{elastic}} \sim Z^{4/3}$, and $K_{\text{inelastic}} \sim Z^{1/3}$. With scanning transmission electron microscope and an annular detector for elastically scattered electrons, by division of the two signals an image contrast $\sim Z$ is achieved [143, 144].

Imperfections in the objective lens generate astigmatism. Using small holes in the sample, Fresnel fringes are observed. The stigmator is set in amplitude and azimuth such that the distance of the Fresnel ring is constant around the hole.

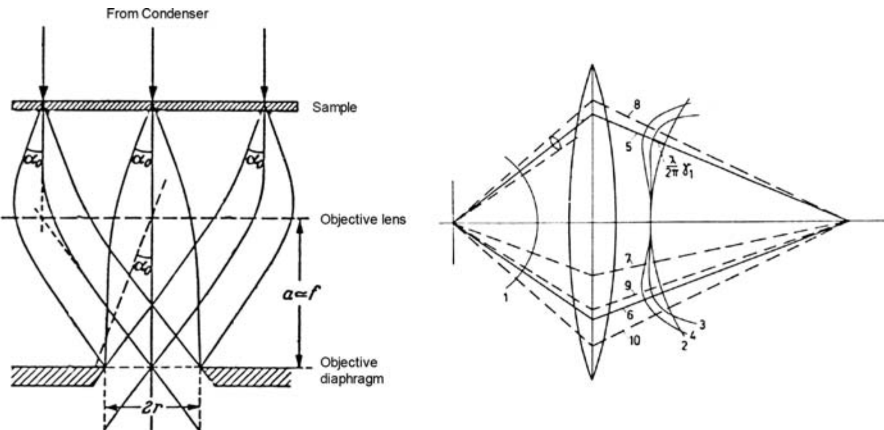


Fig. 4.57. *Left*: schematic of scattering absorption contrast. *Right*: phase shift by aberrations [145]

Phase Contrast

For higher magnifications and higher resolution phase contrast is to be considered, since a big part of the scattered electrons now transmits the aperture. However, the wave surface of the electron is no longer ball shaped like wavefront 1, but in the outer aperture regions due to the influence of spherical aberration a deviation on the wave front is observed, see Fig. 4.57, right. It is called Scherzer Focus, when the influence of the spherical aberration on the outer rays is partly compensated by a defocus setting.

For assessing the resolution of an electron lens system only spherical aberration, axial chromatic aberration, diffraction error and defocusing have to be considered. If correction of axial astigmatism is not sufficient, this has to be considered too. The theoretical resolution limit of a lens system is obtained by calculating the aberration discs of different aberrations and taking the geometric means (root of added squares) of aberration contributions. The resolution limit in electron microscopy imposed by spherical aberration and diffraction error under optimum focus conditions is $c\ Cs^{1/4}\lambda^{3/4}$ (Cs = coefficient of spherical aberration, λ = electron wavelength). The factor c lies between 0.4 and 0.6 and depends on the degree of beam coherence. In modern electron microscopes the optimum aperture to be used for the resolution limit is $\alpha \sim (\lambda/Cs)^{1/4}$.

The two main aberrations that limit the resolution of electron microscopes besides the diffraction error are spherical and axial chromatic aberration, see Fig. 4.58. A further limitation, especially in high-current optics (current >20 μA in a one- axis system), Coulombic electron–electron scattering in source images pushes the electrons and retards them along the axis (Boersch effect) or pushes them perpendicular

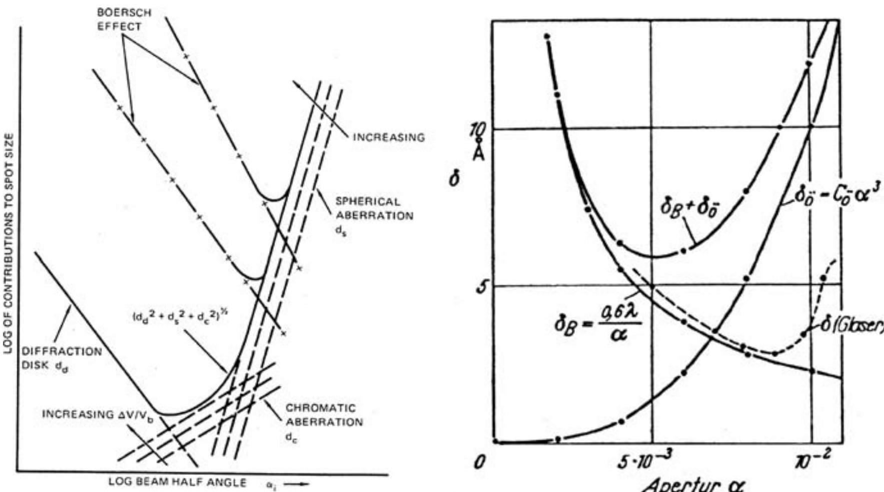


Fig. 4.58. Optimum resolution as a result of diffraction, spherical and chromatic aberration, and of electron–electron interaction (Boersch-effect)

to the axis, which leads to trajectory displacement. Using lenses without Cs and Cc correction, one has to use high lens excitations for small aberrations.

One of the fundamental problems of electron microscopy results from the fact that the specimens are three-dimensional and that the object information (i.e. the arrangement of atoms and molecules making up the specimen) has to be described by a complex function of three variables. The complete information is carried in the amplitude and phase of the electron wave field which has passed through the object. But the image can be observed only as a two-dimensional intensity distribution, represented by a real function of two variables. Therefore, several assumptions and approximations are necessary to interpret an image and to understand the imaging process [146].

In order to describe the imaging, one projects the function containing the object information into a plane shortly behind the object, so a complex function $\Psi_0(x_0, y_0)$ is obtained, where x_0, y_0 are the coordinates in this projection plane. For very thin objects it is identical with the object plane. The image forming system transforms this function into the wave function $\Psi_i(x_i, y_i)$ in the image plane, and the recording system squares it so that the image intensity distribution is $I = |\Psi_i(x_i, y_i)|^2$, where x_i, y_i are the coordinates in the image plane. In an ideal imaging system, $x_i = -Mx_0$, $y_i = -My_0$, with M as the system magnification.

In order to get a faithful image of an object property, the image forming process should be linear, i.e. the output signal has to be linearly related to the input signal, in the case of coherent illumination the image intensity $I(x_i, y_i)$ to the wave function $\Psi_0(x_0, y_0)$. In integral form this means

$$I(x_i, y_i) = \iint G(x_i, y_i, x_0, y_0)\Psi_0(x_0, y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0, \tag{4.2}$$

where the function G describes the imaging qualities of the image forming system. It is known that high-resolution coherent bright field imaging of a weak phase or amplitude object is linear, but not the coherent dark field image. In an incoherent imaging mode, as in the high-resolution bright field and dark field STEM, (4.2) is fulfilled, since then $|\Psi_0|^2$ is the input signal instead of Ψ_0.

Another requirement on a good imaging system, especially the objective lens, is that its aberrations shall have the same effect on all image points, i.e. that the aberration discs depend only on the relative distance to the corresponding Gaussian image point (the image point obtained with an aberration free system). If this is fulfilled, the image is called isoplanatic and (4.3) can be written as

$$I(x_i, y_i) = \iint G(x_i - x_0, y_i - y_0) * \Psi_0(x_0, y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0, \tag{4.3}$$

i.e. as a convolution integral. Applying the Fourier transform (denoted by F) we get

$$F[I] = F[G] * F[\Psi_0], \tag{4.4}$$

where the Fourier spectrum of the object function is directly related to the spectrum of the image intensity. $F[G]$ is called the contrast transfer function and G the pulse

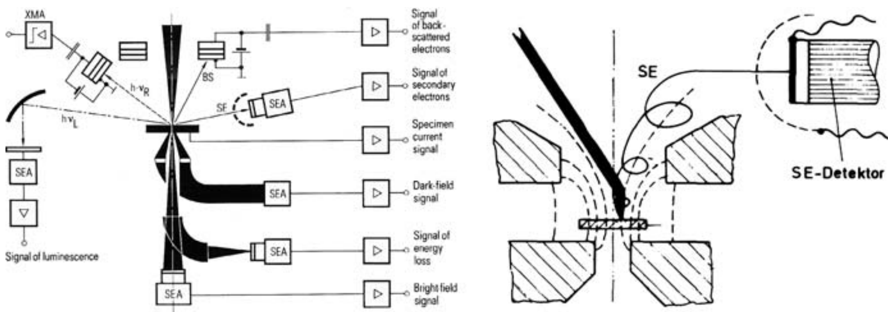


Fig. 4.59. *Left*: possibilities for obtaining signals from a specimen in a STEM. *Right*: secondary electrons detection in a high resolution and strong focusing objective lens

response function, i.e. G is the response of the system to a very steep pulse, for instance a delta function. The concept of contrast transfer functions is very important for describing the properties of optical systems, as resolution and contrast, and is nowadays widely used. Transfer functions can be determined both by theory and experiment.

The basic arrangement of a high-resolution scanning transmission electron microscope STEM imaging system, as first introduced by Crewe [144], consists of the objective lens and one or two additional condenser lenses for demagnifying the electron source. The resolution in STEM with high-brightness guns is the same as in CTEM. Change of magnification in STEM is done by deflection systems. Chromatic aberrations, as far as they originate from energy losses in the specimen, do not exist, because no imaging lenses are used behind the specimen. All geometrical aberrations, except the spherical aberration, can be compensated for by proper alignment and programming of the deflection systems, see Fig. 4.59.

A strong magnetic lens in telecentric mode is used to form a very small probe with a diameter of the order of the resolution limit of the microscope (0.2 to 0.5 nm). The most sensitive region with respect to mechanical vibrations and stray fields is in a CTEM between the objective lens and the first intermediate lens, since here the first magnification takes place. Therefore, it is necessary to couple these elements very rigidly. In a STEM the same holds between the tip and the objective lens.

The resolution obtained in a TEM is determined by the clear separation of two points. Each of them is imaged as the diffraction airy pattern. The superposition should have an intensity of 0.73 in between the two dot centers at intensity 1, see Fig. 4.60, left. On the right 2 resolution pictures from a corrected scanning electron microscope are given. Generally, gold crystals on a carbon background are in use as resolution test sample.

Figure 4.61, left shows how phase contrast originates and is transformed into amplitude contrast by a phase shifter. This is shifting the phase by γ, at best $\pi/2$. Zernicke applied an optical ring in the objective aperture to shift the scattered light with respect to the transmitted unscattered light. Scherzer realized that the action of

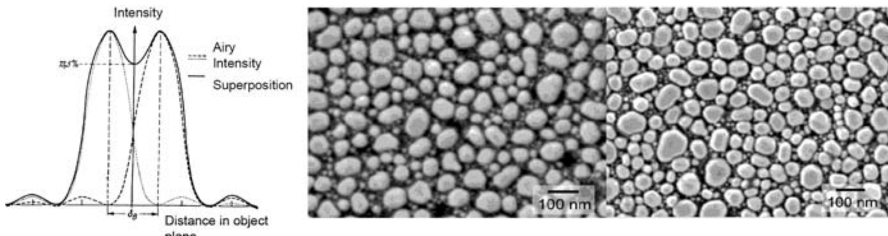


Fig. 4.60. *Left*: definition for resolution test. *Right*: 2 resolution pictures, left uncorrected, right corrected (courtesy JEOL)

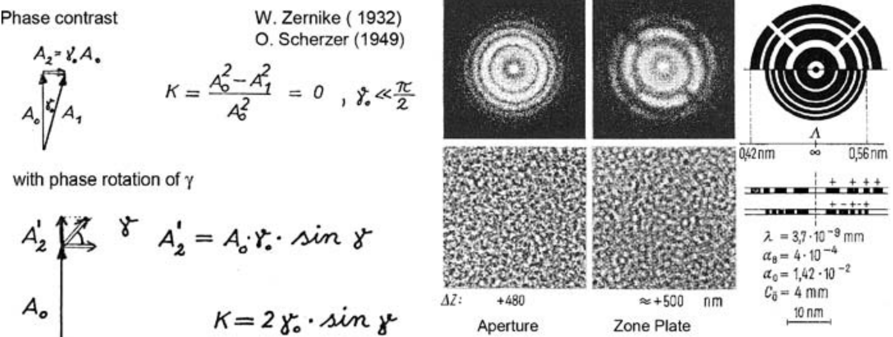


Fig. 4.61. *Left*: origin and method to convert phase contrast into amplitude contrast. *Right*: high-resolution images with diffractograms and imaging with a zone plate as an aperture to correct for negative or positive phase contrast

the phase shifter in electron optical lenses is performed by the 3rd order spherical aberration and can be balanced by the defocus.

Considering an object as a sinusoidal grating of period Λ, the diffraction says: $\Lambda = l/\Theta = f_0\lambda/X_b$, with X_b the axial distance of the first maximum of period Λ. The spherical aberration C_3 and defocus $C_1 = \Delta z$ are shifting the phase by γ in off-axial rays at the diffraction angle Θ: $\gamma = (2\pi/\lambda)(C_3\Theta^4/4 - \Delta z\Theta^2/2)$. The phase contrast is maximum for $\gamma = \pi/2(2m-1)$, where m is an integer, which means $C_3\Theta^4 - 2\Delta z\Theta^2 - (2m-1)\lambda = 0$ or $\Lambda = \lambda\{\Delta z/C_3 \pm [(\Delta z/C_3)^2 + (2m-1)\lambda/C_3]^{1/2}\}^{-1/2}$.

The phase contrast transfer function is $\Psi\Psi^* = 1 - 2a_0^* \cos\gamma - 2\phi_0^* \sin\gamma$ with $\cos\gamma$ the amplitude contrast transfer function and $-2\sin\gamma$ the phase contrast transfer function, $2a_0^*$ the Fourier transform of the amplitude attenuation, and ϕ_0^* the Fourier transform of the phase shift of the object [147].

Figure 4.61 right shows on the left the high-resolution image of a carbon foil as statistical test sample and the corresponding diffractogram (Fourier transform) showing the resolution; in the middle the same object is imaged with a zone plate as aperture blanking off the zones which image the sample with negative phase con-

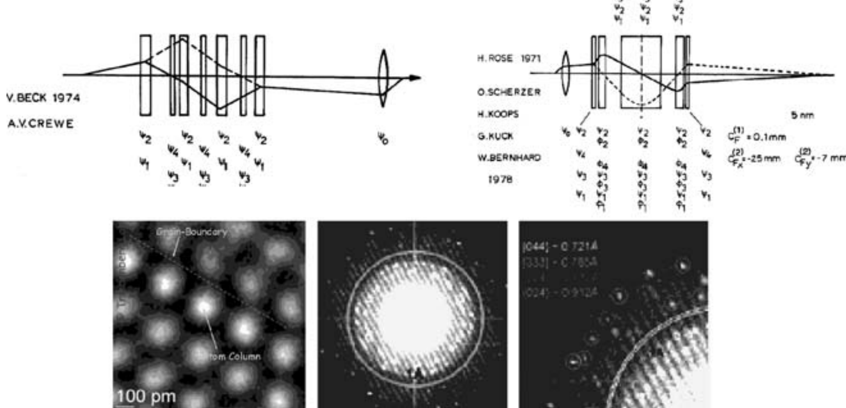


Fig. 4.62. *Top*: correctors for aberrations up to 5th order and chromatic aberrations proposed for SEM (*top left*) and TEM (*top right*) [149, 150]. *Bottom left*: gold atom chain; *middle*: Young's fringes of polycrystalline gold; *right*: fringes extend well below 1.0 Å (courtesy Lichte, Triebenberg, and FEI)

trast; on the right the zone sequence for negative and positive phase contrast and the characteristics of the imaging process [148] are given.

It is possible to correct for spherical aberration and chromatic aberration using multipole correctors, as summarized in Fig. 4.62. Such systems are now available by CEOS GmbH Heidelberg, and are incorporated in high resolution TEM from JEOL, FEI Corp. [149], and Zeiss. A corrected scanning electron microscope is presented by JEOL [150].

Using an energy filter to reduce the width of the electron energy distribution of the field emitter and therefore the chromatical error, the contrast transfer function is widened and allows higher resolution to be displayed. A hexapole spherical aberration corrector achieves sub-Angstrom resolution. Figure 4.62 shows a series of images obtained with such a microscope.

4.4.3 Electron Beam Spectroscopy and Analysis

The electron beam–chemical analysis is based on characteristic X-ray emission [151] using Moseleys law $(1/\lambda)^{1/2} = CZ$, where C is the proportionality factor, see Chapter 7. Detecting the energy loss of the primary beam in electron energy loss spectroscopy (EELS) is element specific and allows to measure the dissociation cross-sections in materials and the phonon and plasmon losses [152, 153]. The energy of the photons is measured by energy dispersive spectroscopy (EDS) or using wavelength dispersive spectrometry (WDS).

Auger electrons are detected and analyzed with Auger electron spectrometry which are high-resolution electron energy spectrometers using quadrupole optics. Electron energy loss spectrometry (EELS) is performed in STEM (scanning transmission electron microscopes) using magnetic sector field as analyzing elements,

or magnetic Omega filters in TEM to image a sample with the electrons which did undergo an energy loss at a certain element [154].

Energy dispersive detectors (EDS) are Si-Li drifted p-n detectors, facing the sample through a beryllium or mylar window, and measure the whole spectrum at once [155]. In a wavelengths dispersive spectrometer the X-ray lines are separated by Bragg diffraction at a crystal and then counted with Geiger–Müller counters [156]. To analyze all elements, it requires several crystals. WDS has a high-energy resolution of 10 eV, vs 130 eV of the EDS detector. Matrix effects have to be considered for corrections. The relative error of EDX analysis is 5%, whereas WDS has [157] an error of 1%.

Auger electrons emerge from the surface and range from 50 to 3000 eV. They are analyzed with cylinder mirror analyzers [158, 159]. In many cases an ion source is used to clean the surface and to make in depth profiling.

Electron energy loss spectroscopy (EELS) uses in transmission microscopy magnetic or electrostatic sector fields, cylinder analyzers or Wien filters [160]. Energy is lost by plasmon excitation with 0 to 50 eV [161]. The investigation of plasmon vibrations is used to analyze the structure of metals [162]. Chemical microanalysis of low Z materials show up at energy losses of 100 to 2000 eV [163].

The spectrum shows no peaks but falling wedges, which result from energy transfer to ionized inner shell electrons. Information of chemical bonding and band structures are revealed in fine structure of steep inclining wedges.

Energy filtering is also possible for whole images using magnetic energy filters in the magnifying lens system of a TEM. In addition, those filters increase the signal to noise ratio in diffraction investigation, by filtering out all inelastically scattered electrons [164]. With characteristic inelastic scattered electrons element distribution images of high resolution are obtained.

Cathodo-luminescence analysis uses an elliptical mirror and a grating spectrometer, a cooled photomultiplier and lock in techniques to display a spectral intensity of photons in the IR, visible, or UV. Impurities and damaged crystal areas in single crystals are revealed. This is of importance for investigating base materials of devices which need to have undisturbed single crystals for their function [165].

4.4.4 Electron Beam Testing

Early Testing: Mechanical Probing

In the early 1980s cluster probes for opens and shorts testing of multilayer ceramic substrates and for integrated circuits were used in manufacturing. Cluster probes used arrays of “buckling metal beams” to contact simultaneously many or all network nodes on a substrate. They touched a line at both ends and measured the flowing current. Probe development times were lengthy and highly product specific. With increasing packaging density the use of mechanical probes became impractical and expensive.

Non-contact Electron-beam Testing

To increase versatility, a non-contact test technology was developed based on the scanning electron beam microscope. Electron beam testing of electronic devices in electronics production presents several tasks:

(a) Testing of circuits on semiconductors and of multilayer wiring substrates and printed circuit boards to characterize and prove correct design and production.

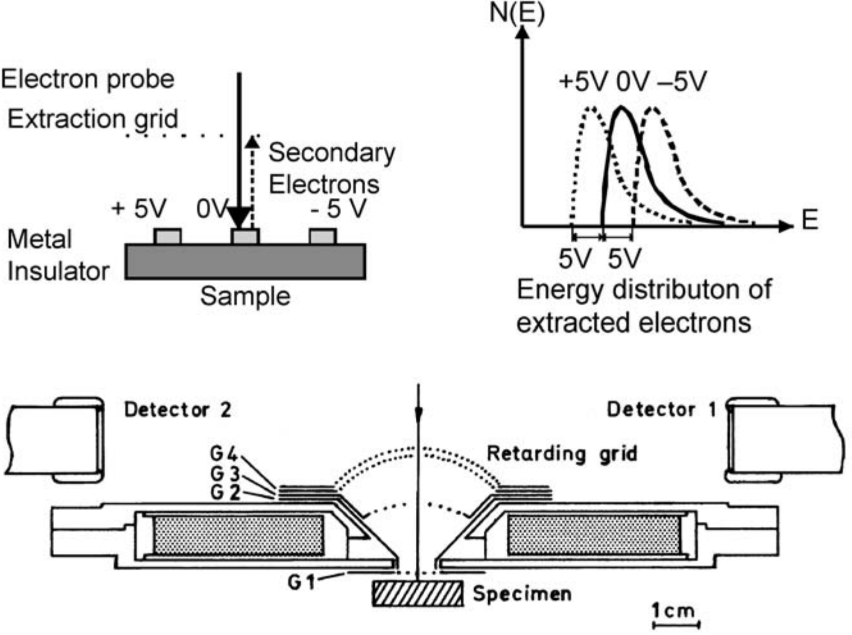


Fig. 4.63. *Top left*: principle of testing setup. *Top right*: energy distribution of extracted electrons from metal lines with different potentials. *Bottom*: through lens secondary electron spectrometer

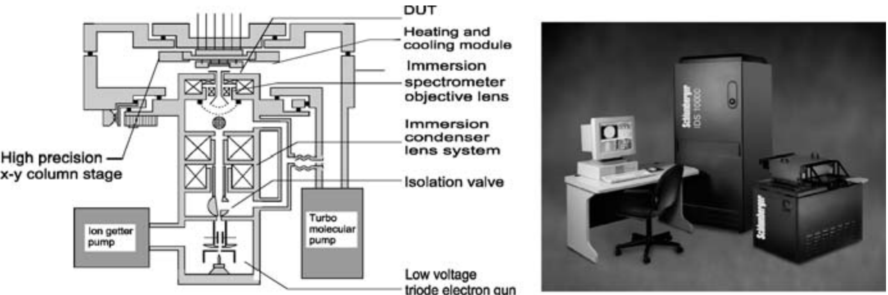


Fig. 4.64. *Left*: schematic of an e-beam probing system for VLSI circuits. *Right*: electron beam test system for chip diagnostics (Courtesy of NP Test Inc.)

Conductors, insulators, doping areas are investigated. Mechanical probing and non-contact voltage contrast method is used.

(b) Testing of circuits on semiconductors for high frequency behaviour and function: stroboscopic imaging, light emission spectroscopy of transistors.

(c) Testing of multilayer wiring metal line patterns for flip chip mounting and soldering panels: through panel charging, charged grid method, induced current method, and secondary electron spectroscopy are employed.

(d) Testing of transistors and lines in LCD production by secondary electron spectroscopy.

The investigation of devices in the scanning electron microscope with voltage contrast imaging has become inevitable necessary for the testing and debugging of processes and device structures in semiconductor device production. This speeds up the design and development of a new device and helps with circuit editing for fast development turnaround. A working and a nonworking device are easily compared and defects and errors are found, localized, and revealed from such an analysis. This technique offers the required resolution and throughput, and also the flexibility to test any package of differing geometry without problems.

Non-contact e-beam testing relies on the voltage contrast obtained when imaging metals and insulators in devices. The floating conductor can be charged positively or negatively depending on the primary electron energy, see Fig. 4.63. Applied voltages can be followed along lines.

Secondary electron energy spectrometry helps to disguise faults and to detect high frequency signals. Figure 4.63 right shows a through-lens secondary electron spectrometer as it is employed in e-beam testers. A schematic is given in Fig. 4.64 left, and a commercial tester, right.

The Dual Potential Test Method

Figure 4.65 left illustrates the "dual potential" technique [166] employed for the testing of networks with multiple nodes on the top surface of a substrate. A network was charged negatively by a relatively high-energy beam directed to one network node. The beam potential was rapidly switched to a lower reading potential and charged.

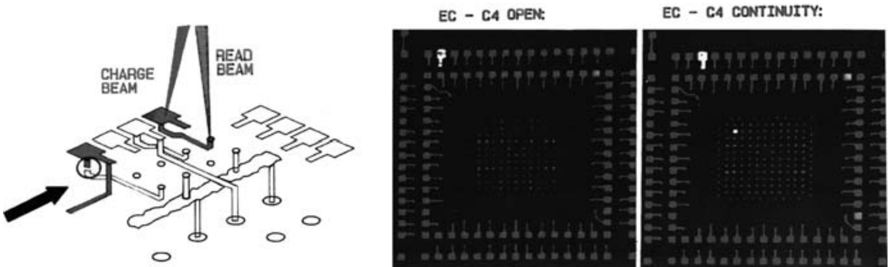


Fig. 4.65. Schematic of the dual potential test method (*left*) and voltage contrast secondary electron images showing the continuity between network nodes (*right*) or lack thereof (*middle*)

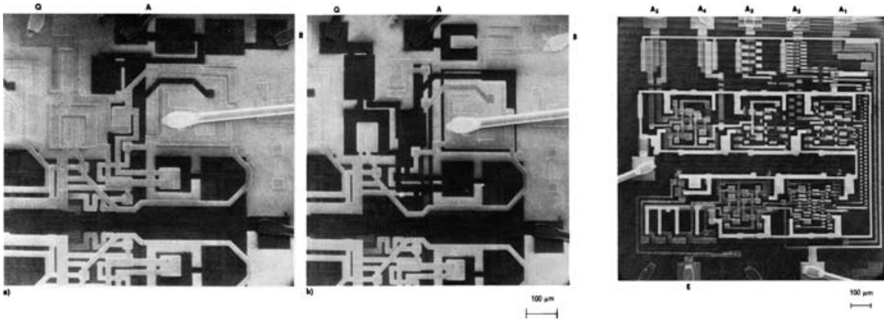


Fig. 4.66. *Left*: voltage contrast image of a gatter, one out of 2 switches: **a** the inputs A and B have +12 V (*dark*), the switching logic produces at the output Q the potential 0 V; **b** inputs: A: +12 V, B: 0 V, output Q: +12 V (*dark*). *Right*: example of high-frequency sampling imaging technique

"Reading" consisted of determining the magnitude of the signal from a voltage contrast (VC) secondary electron detector. The images in Fig. 4.65 right illustrate the VC image produced by scanning at the read potential. VC signal amplitude is represented in the image by a gray scale with white corresponding to a large signal (negative potential). The image on the right illustrates continuity between an outer pad and an inner flip-chip attachment point, whereas the image in the middle illustrates an open in the connection. The beam was directed to other network nodes to verify that they had been.

Static Voltage Contrast Imaging of "Life" Circuits

Figure 4.66, left: a and b shows the voltage contrast image of a switch for different voltage settings. When switching the input voltages, the SEM displays the function of the device and a series of images can be taken.

Slow changing states can be displayed and observed if the SEM uses TV imaging. The precondition is that the switching frequency is small compared to the image frequency. However, if the device is supplied with high frequency much larger than the imaging frequency, a striped image displaying the changing potential contrasts allows to observe the functioning of the device.

Figure 4.66, right shows the high frequency sampling technique imaging a 5 stage frequency divider. At input E a rectangular alternating voltage is supplied of frequency f. The outputs display A1: $f/2$, A2: $f/4$, A3: $f/8$, this can be seen from the doubling of the stripe width. It is also revealed that the last two stages of the device are malfunctioning.

Stroboscopic Voltage Contrast Imaging

A voltage contrast image of such devices having alternating potentials is only obtained employing a stroboscopic imaging method. This means that the imaging electron beam and the operating voltage of the device is switched on and off (pulsed

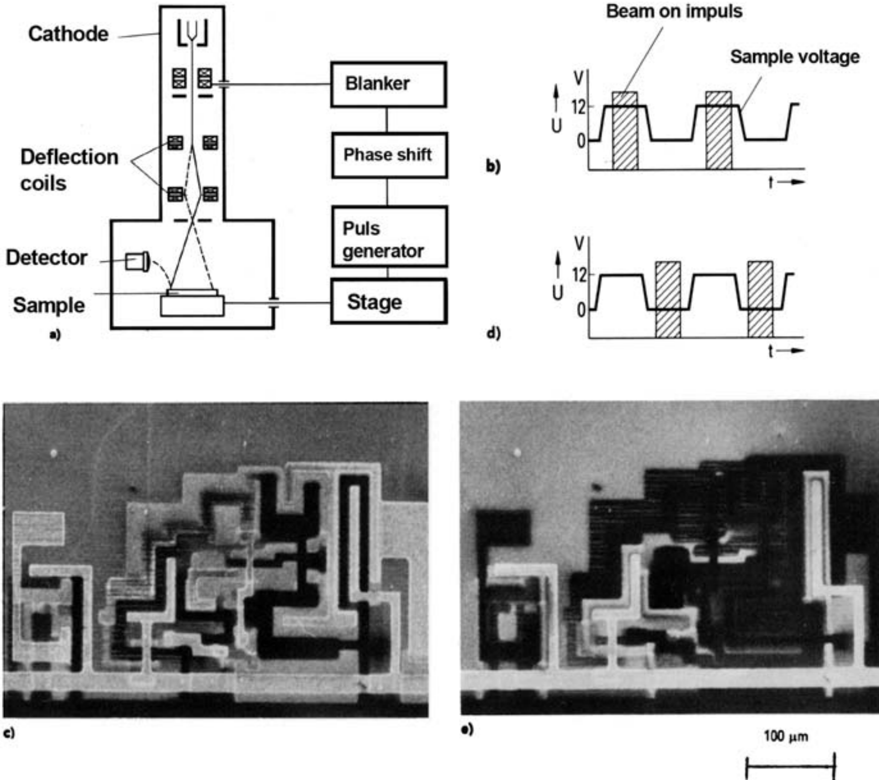


Fig. 4.67. Principle of the stroboscopic voltage contrast imaging method: **a** the stroboscopic tester system; **b** and **d** different phases of beam on impulse and device operating voltage; **c** and **e** stroboscopic voltage contrast images of a MOS device using the voltages as given in (**b**) and (**d**); frequency of operation 1 MHz

or blanked) with the same frequency. Figure 4.67a shows the schematic of the stroboscopic imaging method [167]. The device is fully wired and supplied inside the scanning microscope and operates at its frequency. A special device socket and supply voltage feed through having up to, e.g. 40 pins is used. This allows to connect the device to an external pulse generator, phase shifter and supply voltage generator. The beam blanker is triggered using the device operating frequency and voltage [168]. Using a phase shifter, the device and beam blanking can be adjusted to obtain a stroboscopic sharp respectively standing image, see Fig. 4.67b. This displays then the device potential distribution at a special phase, e.g. 12 V device voltage, see Figs. 4.67c and 4.67e. Figure 4.67d shows a state being half a period shifted, e.g. 0 V supply voltage. The operating frequency was 1 MHz.

References

[1] M. von Pirani, U.S. Patent 848,600 (1907)
[2] E. Riecke, Ann. Phys. Chem. **13**, 191–194 (1881)
[3] E. Wiechert, Nachr. d., Kgl. Ges. d. Wiss. zu Göttingen, Math.-Phys. Kl. 260–293 (1898)
[4] C. Samson, Ann. Phys. **55**, 606–632 (1918)
[5] D. Gabor, Arch. Elektrotech. **16**, 296–302 (1926)
[6] H. Busch, Arch. Elektrotech. **18**, 583–594 (1927)
[7] E. Ruska, M. Knoll, Z. Techn. Phys. **12**, 389–400, 448 (1931)
[8] M. Knoll, E. Ruska, Ann. Phys. **12**, 607–661 (1932)
[9] M. Knoll, F.G. Houtermans, W. Schulze, Verh. d. Dt. Phys. Ges. **13**, 23, 24 (1932)
[10] M. Knoll, E. Ruska, Das Elektronenmikroskop, Z. Phys. **78**, 318–339 (1932)
[11] R. Rüdenberg, Elektronenmikroskop, Naturwissenschaften **20**, 522 (1932)
[12] E. Brüche, O. Scherzer, *Geometrische Elektronenoptik* (Geometrical Electron Optics) (Springer, Berlin, 1934), 332 pp
[13] E. Brüche, H. Johannson, Elektronenoptik und Elektronenmikroskop (Electron optics and the electron microscope) Naturwissenschaften **20**, 353–358 (1932)
[14] E.W. Müller, Diss. Techn., Hochschule Berlin 1936; Z. Phys. **102**, 734–761 (1936); Z. Techn. Phys. **77**, 412416 (1936); Z. Phys. **106**, 541–550 (1937); Ergebn. Exakt. Naturwiss. **27**, 290–360 (1953)
[15] M. Knoll, Z. Techn. Phys. **16**, 467–475 (1935)
[16] M. von Ardenne, Z. Phys. **109**, 553-572 (1938); Z. Techn. Phys. **19**, 407–416 (1938)
[17] R. Rühle, DRP 764927 (1939)
[18] K.H. Steigerwald, Phys. Verhandlungen **4**, 123 (1953)
[19] J.A. Stohr, Vacuum welding by electron beam, Nucl. Power **3**(6), 272 (1958)
[20] R. Bakish, *Introduction to Electron Beam Technology* (Wiley, New York, 1962)
[21] S. Schiller, U. Heisig, S. Panzer, *Electron Beam Technology* (Wiley, New York, 1982)
[22] S. Schiller, U. Heisig, S. Panzer, *Elektronenstrahltechnologie* (Verlag Technik, Berlin, 1977)
[23] M. von Ardenne, *Tabellen zur Angewandten Physik,* vol. 1–3 (VEB Deutscher Verlag, der Wissenschaften, 1973) 2. Auflage
[24] E. Abbe, Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung (Contributions to the theory of the microscope and microscopical observation) M. Schultze's Arch. f. mikroskopische Anatomie IX, 413–468 (1873); E. Abbe, Ges. Abh. **1**, 45–100 (1904)
[25] H. von Helmholtz, Über die Grenzen der Leistungsfähigkeit der Mikroskope (On the limits of the Performance of microscopes) Monatsber. d. Akad. d. Wiss. zu Berlin, 625–626 (1873)
[26] L. De Broglie, Recherches sur la theorie des quanta (Researches on the theory of quanta) These, Masson & Cie, Paris, 1924; Ann. de Phys. **3**, 22–128 (1925)
[27] W. Glaser, *Grundlagen der Elektronenoptik* (Springer, Wien, 1952); H. Seiler, Abbildung von Oberflächen mit Elektronen, Ionen und Röntgenstrahlen, BI Hochschultaschenbücher 428/428a (1968)
[28] G. Liebmann, E.M. Grad, Imaging properties of a series of magnetic electron lenses, Proc. Phys. Soc. B **64**, 956 (1951)
[29] EO3D by E. Munroe MEBS: Munro's Electron Beam Software Ltd., London, UK
[30] http://www.electronoptics.com/
[31] SPOC by Lencova at lencova@cmail.cz

[32] E. Brüche, O. Scherzer, *Geometrische Elektronenoptik* (Springer, Berlin, 1934); B. von Borries, *Die Übermikroskopie* (Verlag Dr. W. Saenger, Berlin, 1949); P.W. Hawkes, *Image Processing and Computer Aided Design in Electron Optics* (Academic, London, 1973), p. A 68; El Kareh, *Electron Beams, Lenses and Optics* (Academic, New York, 1970), Bd. 1 and 2

[33] T.M. Mulvey, M.J. Wallingtom, Electron lenses, Rep. Prog. Phys. **36**, 347 (1973)

[34] O. Scherzer, J. Appl. Phys. **20**, 20 (1949)

[35] F. Schleich, in *Proc. 2 Electron Beam Processing Seminar*, US Trade Center Frankfurt, 3f1, June 1972

[36] M. von Ardenne, *Tabellen zur Angewandten Physik*, Bd. I (VEB-Verlag der Wiss., Berlin, 1962); P.W. Hawkes, *Electron Optics and Electron Microscopy* (Taylor & Francis, London, 1972)

[37] H. Rose, Abbildungseigenschaften sphärisch korrigierter elektronenoptischer Achromate, Optik **33**, 1–24 (1971)

[38] G. Induni, Helvet. Phys. Acta **20**, 463 (1947)

[39] P. Egelhaaf, Optik **27**, 387 (1968)

[40] J.R. Pierce, J. Appl. Phys. **11**, 548 (1940) (3.4.4.5); Von Ardenne, 1948. Tabellenzur Angewandten Physik, 1st edn (1956); Fröhlich, Nukleonik **1**, 183 (1959); Von Ardenne, Exp. Techn. Phys. **9**, 227 (1961)

[41] Steigerwald, Optik **5**, 469 (1947); Brüche, Z. Angew. Phys. **3**, 88 (1951); F.W. Braucks, Optik **15**, 242 (1958); F.W. Braucks, Optik **16**, 304 (1959)

[42] Bas-Gaydott, Z. Angew. Phys. **11**, 370 (1959)

[43] See, e.g. Von Ardenne, Z. Fernsehen **1**, 200 (1930); Maloee-Epstein, Proc. Inst. Radiat. Eng. **22**, 1386 (1934); Law, Proc. Inst. Radiat. Eng. **25**, 954 (1935); Schwaetz-Strübig-Paehe, Z. Fernseh-GmbH **1**, 5 (1938); Schwarz, Z. Techn. Phys. **19**, 454 (1938); Dosse, Z. Phys. **115**, 530 (1940); Hillier-Baker, J. Appl. Phys. **16**, 469 (1945); B. Von Borries, Optik **3**, 361, 389 (1948); Rogowski-Gbösser-Sommerfeld, Arch. Elektro-techn. **15**, 377 (1925)

[44] Bas, Optik **12**, 377 (1955); Z. Angew. Phys. **7**, 337 (1955)

[45] Samuel, Proc. Inst. Radiat. Eng. **33**, 233 (1945); Rothe-Kleen, Hochvakuum – Elektronenröhren **1**, Akad. Verlagsges., Frankfurt a. M. 219 (1955); Brewer, J. Appl. Phys. **28**, 7 (1958)

[46] Rogowski-Grösser-Sommerfeld, Arch. Elektro-techn. **15**, 377 (1925)

[47] A.N. Broers, J. Appl. Phys. **J38** (1991)

[48] W. Blair, A.N. Broers, IBM Techn. Discl. Bull. **14**(2) (1971)

[49] Hibi, J. Electronmicrosc. Japan **3**, 15 (1955); Sakaki-Möllenstedt, Optik **13**, 195 (1956); Maruse-Sakaki, Optik **15**, 485 (1958); Sakaki-Maruse, in *Verh. Int. Kongr. Elektr. Mikroskopie*, Berlin, 1958, vol. 9, p. 1 (Springer, Berlin, 1960); Drechsler-Cosslett-Nixon, ebenda, 13; Krause-Riecke, Phys. Verh. **12**, 141 (1961)

[50] L.W. Swanson, L.C. Grouser, Phys. Rev. **163**, 622 (1971); A.V. Crewe, M. Isaacson, D. Johnson, Rev. Sci. Instrum. **2**, 411 (1971)

[51] A.V. Crewe, D.N. Eggenberger, J. Wall, L.M. Welter, J. Appl. Phys. **38**, 4257 (1967); J.W. Butler, in *6th Int. Cong. Electron Micros.*, Rome, p. 191, 1966; A.V. Crewe, M. Isaacson, D. Johnson, Rev. Sci. Instrum. **40**, 241 (1969)

[52] K.J. Hanszen, K. Lauer, Zeitschr. Naturforschung **22a**, 238 (1967)

[53] B. von Borries, Optik **3**, 321 (1948)

[54] A.N. Broers, in *Proc. Scanning Electron Microscopy*, Part III, 661 (1975)

[55] H. Schulz, *Elektronenstrahlschweißen, Fachbuchreihe Schweißtechnik,* Band 93 (DVS-Verlag GmbH, Düsseldorf, 1998)

[56] D. von Dobeneck, T. Löwer, V. Adam, *Elektronenstrahlschweißen, Die Bibliothek der Technik,* Band 221 (Verlag Moderne Industrie, Landsberg/Lech, 2001)
[57] K.-R. Schulze, D.E. Powers, EBW of aluminum breaks out of the vacuum, Weld. J. (Miami) **83**(2), 32–38 (2004)
[58] R. Zenker, Electron beam surface treatment, industrial applications and prospects, Surf. Eng. **12** (Helt 4, s.), 9–12 (1996)
[59] G. Müller, V. Engelko, H. Bluhm, A. Heinzel, G. Schumacher, D. Strauss, A. Weisenburger, F. Zimmermann, V. Shulov, N. Notchovnaia, Application of pulsed electron beams for improvement of material surface properties, in *Proceedings of the 14th International Conference on High-Power Particle Beams 2002*, Albuquerque, NM, 23–28 June 2002, p. 325
[60] V.L. Granatstein, I. Alexeff, *High Power Microwave Sources* (Artech House, Boston, London, 1987)
[61] G. Cooperstein, Status of particle beam and pulsed power research in US, in *Proceedings of the 14th International Conference on High-Power Particle Beams 2002*, Albuquerque, NM, 23–28 June 2002
[62] V. Engelko, B. Yatsenko, G. Müller, H. Bluhm, Pulsed electron beam facility (GESA) for surface treatment of materials, Vacuum **62**, 211–216 (2001)
[63] M.A. Vasilievskii, I.M. Roife, V.I. Engelko, E.G. Yankin, Dynamics of plasma generated by a multitip explosive emission cathode, Sov. Phys. Tech. Phys. **31**(3), 284 (1986)
[64] M.A. Vasilievskii, I.M. Roife, V.I. Engelko, E.G. Yankin, The influence of emitter density and anode processes on the time of operation of a diode with multitip explosive emission cathodes, Sov. Phys. Tech. Phys. **33**(9), 1060 (1988)
[65] D. Strauss, G. Müller, G. Schuhmacher, V. Engelko, W. Stamm, D. Clemens, W.J. Quadakkers, Oxide scale growth on MCrAIY bond coatings after pulsed electron beam treatment, Surf. Coat. Technol. **135**, 196–201 (2001)
[66] G. Müller, A. Heinzel, G. Schumacher, A. Weisenburger, Control of oxygen concentration in liquid lead and lead–bismuth, J. Nucl. Mater. **321**, 256–262 (2003)
[67] G. Jäsch, Elektronenstrahlverdampfung von Aluminium mit Raten über 10g/s, Dissertation, TH Chemnitz, 1981, p. 21 ff
[68] S. Schiller, U. Heisig, S. Panzer, *Electron Beam Technology* (Verlag Technik GmbH, Berlin, 1995), pp. 37, 38
[69] H. Ranke, V. Bauer, A. Feuerstein, W. Dietrich, Verdampferanordnung mit einem rechteckigen Verdampfertiegel und mehreren Elektronenkanonen, German patent No DE 3639683 C2
[70] K. Goedicke, B. Scheffel, S. Schiller, Plasma activated high rate evaporation using a spotless cathodic arc, Surf. Coat. Technol. **68/69**, 799–803 (1994)
[71] R. Rühle, Verfahren zur Verdampfung im Vakuum, DRP 764927 (1939)
[72] G. Mladenov, in *Proceedings of the 7th International Conference on Electron Beam Technologies: Vacuum 77*, vol. 4 (2005), pp. 359–546
[73] Schreiter, *Sondermetalle* (Verlag für Grundstoffindustrie, Leipzig, 1961); J. Kouptsidis, F. Peters, D. Proch, W. Singer, Niob für TESLA (2001) (http://tesla.desy.de/new_pages/TESLA_Reports/2001/pdf_files/tesla2001-27.pdf)
[74] M. von Ardenne, G. Musiol, S. Reball (eds.), *Effekte der Physik* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1989)
[75] W.D. Claus (ed.), *Radiation Biology and Medicine* (Addison-Wesley, Reading, 1958)
[76] H. Dertinger, H. Jung, *Molekulare Strahlenbiologie* (Springer, Berlin, Heidelberg, New York, 1969)
[77] W. Glaser, *Elektronenoptik* (Springer, Wien, 1961)

[78] M. von Ardenne, *Tabellen zur Angewandte Physik*, vol. 1 (VEB Verlag, 1961)
[79] L. Reimer, *Transmission Electron Microscopy* (Springer, Berlin)
[80] G.H. Jansen, Coulomb interactions in particle beams in adv., in *Electronics & El. Physics*, ed. by P.W. Hawkes, Suppl. 21 (Academic, Boston, 1990)
[81] L.H. Jaksch, Field emission SEM for true surface imaging and analysis, Mater. World (October) (1996)
[82] L. Reimer, *Scanning Electron Microscopy* (Springer, Berlin)
[83] Grivet, *Electron Optics* (Elsevier, Oxford (Pergamon Print) (1965)
[84] G. Willson, Organic resist materials – theory and chemistry, Chapter 3 of *Introduction to Microlithography, Theory, Materials, Processing*, ed. by L.F. Thompson, C.G. Willson, M.J. Bowden, American Chemical Society Symposium Series 219 (1983)
[85] L.W. Swanson, L.C. Crouser, Zr-O-W cathode, J. Appl. Phys. **40**(12), 4741 (1969)
[86] M. Hatzakis, C.H. Ting, N. Viswanathan, Fundamental aspects of electron beam exposure of polymeric resist systems, in *Proc. 6th Int. Conf. On Electron and Ion Beam Sciences and Technology* (1974), p. 205
[87] CAR Resists SAL 102 from Shipley, 455 Forest St., Marlboro, MA 01752, USA
[88] T.H.P. Chang, Proximity effect in electron beam lithography, J. Vac. Sci. Technol. **12**, 1271–1275
[89] G.E. Moore, Lithography and the future of Moore's law, in *Proc. SPIE*, vol. 2437 (SPIE Press, Bellingham, 1995), pp. 2–7
[90] M. Born, E. Wolf, *Principles of Optics*, 6th edn (Pergamon, Oxford, 1980)
[91] P. Chien, M. Chen, Proximity effects in submicron optical lithography, in *Proc. SPIE,* vol. 772 (SPIE Press, Bellinham, 1987), pp. 35–49
[92] ITRS Roadmap 1997. http://public.itrs.net
[93] M. Colburn et al., Step-and-flash imprint lithography: a new approach to high resolution patterning, in *Proc. SPIE*, vol. 3676 (SPIE Press, Bellingham, 1999), p. 379; S-FIL^{TM} for sub-80 nm contact hole patterning, Solid State Technol. **27**(2) (2004)
[94] M. Gesley, F. Aboud, D. Colby, F. Raymond, A. Watson, Electron beam column for submicron – and nanolithography, Jpn. J. Appl. Phys. **32**, 5993 (1993)
[95] Y. Nakayama, S. Okazaki, N. Saitou, H. Wakabayashi, Electron-beam cell projection lithography: A new high-throughput electron-beam direct-writing technology using a specially tailored Si aperture, J. Vac. Sci. Technol. B **8**(6) (1836) (Nov./Dec.)
[96] H.C. Pfeiffer, Recent advances in electron-beam lithography for the high-volume production of VLSI devices, IEEE Trans. Electron Devices **ED-26**(4), 663–674 (1979); H.C. Pfeiffer, in *11th Symp. Electron and Ion Laser Beam Technology*, ed. by R.M.F. Thornley (San Francisco Press, 1971), p. 337
[97] M. McCord, M. Rooks, Chapter 2.6.5.1 GDS-II, in *Handbook of Microlithography, Micromachining and Microfabrication, vol. 1: Microlithography*, ed. by P. Rai-Choudhury (SPIE Press, Bellingham, 1997) 198 ff
[98] M. Hintermaier, U. Hoffman, B. Hübner, C.K. Kalus, E. Knapek, H.W.P. Koops, R. Schlager, E. Seebald, M. Weber, Proximity correction using CAPROX, evaluation and application, J. Vac. Sci. Technol. B **9**, 3043–3047 (1991)
[99] T.H.P. Chang, D. Kern, L. Murray, Arrayed miniature electron beam columns for high throughput sub 100-nm lithography, J. Vac. Sci. Technol. B **10**, 2743 (1992)
[100] T. Groves, in *Proc. 20th European Mask Conference*, GMM Fachbericht 43, Dresden, 2004, p. 209
[101] J. Hren, Barriers to AEM: contamination and etching, in *Principles of Analytical Electron Microscopy*, ed. by D. Joy, A. Romig, J. Goldstein (Plenum, New York, 1986), p. 353

[102] R. Schief, Optic **29**, 416 (1969)
[103] A. Broers, W. Molzen, Electron-beam formation of 80 Å metal structures, Appl. Phys. Lett. **29**, 596 (1976)
[104] H.W.P. Koops, J. Kretz, M. Rudolp, M. Weber, G. Dahm, K.L. Lee, Characterization and application of materials grown by electron beam induced deposition, Jpn. J. Appl. Phys. **33**, 7099–7107 (1994)
[105] S. Matsui, T. Ichihashi, Electron beam induced selective etching and deposition technology, J. Vac. Sci. Technol. B **7**, 1182 (1989)
[106] H.W.P. Koops, Charged particle beam processes and its applicability to mask repair for next generation lithographies, in *Proceedings of the 17th European Mask Conference*, GMM-Fachbericht 32 (VDE Verlag, Berlin, 2000), pp. 191–194
[107] H.W.P. Koops, R. Weiel, D.P. Kern, T.H. Baum, J. Vac. Sci. Technol. B **6**, 477–481 (1988)
[108] D. Winkler, H. Zimmermann, M. Mangerich, R. Trauner, E-beam probe station with integrated tool for electron beam induced etching, Microelectron. Eng. **31**, 141–147 (1996)
[109] Sun Chemical Corp, Preparation of photosetting coating composition, U.S. Patents 2,453,769 and 2,453,770, 1946
[110] A. Charlesby, V. Wycherley, The irradiation of unsaturated polyester resins, Int. J. Appl. Radiat. Isot. **2**, 26–34 (1957)
[111] P.G. Garratt, *Strahlenhärtung* (Curt R. Vincentz Verlag, Hannover, 1996)
[112] Chemistry & Technology of UV & EB Formulations for Coatings, Inks & Paints (Wiley/SITA Series in Surface Coatings Technology) vol. I, vol. 3 (Wiley, New York, 1998)
[113] B.J. Twomey, Radiation cured coatings – some commercial successes, Phys. Chem. **14**, 69 (1979)
[114] P. Holl, Two ideal applications for the low energy electron-beam accelerator: Vulcanisation of pressure-sensitive adhesives and controlled through-curing of coatings on parquet, Radiat. Phys. Chem. **46**, 953 (1995)
[115] G. Mattausch, M. Henry, J. Daenhard, A. Reichmann, O. Roeder, R. Bartel, Survey of Electron Beam Technologies at Fraunhofer FEP. Annual Report of the Fraunhofer Institute for Electron Beam and Plasma Technology, Dresden, 2001
[116] A. Charlesby, *Atomic Radiation and Polymers* (Pergamon, Oxford, 1960)
[117] A. Chapiro, *Radiation Chemistry of Polymeric Systems* (Wiley, New York, 1962)
[118] W. Schwarz, U. Gohs, Reactive modification of native oils by means of electrons, in *Annual Report of the Fraunhofer Institute for Electron Beam and Plasma Technology*, Dresden, 2003, pp. 20–27
[119] J.G. Drobny, *Radiation Technology for Polymers* (CRC, Boca Raton, London, New York, Washington, DC, 2003)
[120] E.S. Josephson, A historical review of food irradiation, J. Food Saf. **5**, 161 (1983)
[121] B.P. Fairand, *Radiation Sterilization for Health Care Products: X-Ray, Gamma, and Electron Beam* (CRC, Boca Raton, 2002)
[122] FOA/IAEA/WHO expert committee, Wholesomeness of irradiated food. Technical Report Series 659, WHO, Geneva (1981); S. Hackwood, The irradiation processing of foods, in *Food Irradiation*, ed. by S. Thorne (Elsevier Science, Essex, 1991)
[123] T. Schröder, O. Röder, K. Lindner, e•dressing – a unique technology for seed, ISTA News Bull. **118**, 13 (1998)
[124] M. Cutrubinis, H. Delicee, M. Stahl, O. Röder, H.J. Schaller, Detection methods for cereal grains treated with low and high energy electrons, Radiat. Phys. Chem. **72**, 639–664 (2005)

[125] S. Machi, Radiation technology for environmental conservation, Radiat. Phys. Chem. **22**(1–2), 91–97 (1983)
[126] H. Mätzing, H.R. Paur, Chemical mechanisms and process parameters of flue gas cleaning by electron beam, in *Gaseous Pollutants: Characterization and Cycling*, ed. by J.O. Nriagu (Wiley, New York, 1992), pp. 307–331
[127] O. Tokunaga, N. Suzuki, Radiation chemical reactions in NO_x and SO_2 removals from flue gas, Radiat. Phys. Chem. **24**(1), 145–165 (1984)
[128] R.J. Woods, A.K. Pikaev, *Applied Radiation Chemistry: Radiation Processing* (Wiley, New York, 1994)
[129] B. Han, J. Ko, J. Kim, Y. Kim, W. Chung, I.E. Makarov, A.V. Ponomarev, A.K. Pikaev, Combined electron-beam and biological treatment of dyeing complex wastewater. Pilot plant experiments, Radiat. Phys. Chem. **64**, 53–59 (2002)
[130] G.V. Buxton, C.L. Greenstock, W.P. Helman, A.B. Ross, Critical review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals (OH/O-) in aqueous solutions, J. Phys. Chem. Ref. Data **17**, 513–886 (1988)
[131] H. Rehme, Siemens Analysentechnische Mitteilungen 123, Sonderdruck aus Techniken der Zukunft, Heft 9 (1974), pp. 69–74; H. Rehme, Analysentechnische Mitteilungen (62) (1973)
[132] B. von Borries, E. Ruska, Die Technik des Siemens – Übermikroskops, Siemens Zeitschrift **20**(6), 217 (1940)
[133] O. Scherzer, Über einige Fehler von Elektronenlinsen, Z. Phys. **101**, 593–603 (1936); H. Rose, Abbildungseigenschaften sphärisch korrigierter elektronenoptischer Achromate, Optik **33**, 1–24 (1971)
[134] H. Koops, in *Proc. 9 Int. Congr. Electron Microscopy*, Toronto Bd. 3, 1978, p. 185
[135] H. Rose, Prospects for aberration-free electron microscopy, Ultramicrosc. **103**(1), 1–6 (2005)
[136] Berechnungsgrundlagen für die Vakuumtechnik, Leybold-Heraeus, HV 150
[137] ESPE-Knoll: Hochspannungstechnologie; G. Oberdorfer, *Lehrbuch der Elektrotechnik*, Bd. 1 (Leibnitz-Verlag, München, 1948)
[138] J. Heinzerling, J. Phys. E **9**, 131 (1976); G. Fey, Optik **55**, 55 (1980)
[139] J. Picht, I. Heydenreich, *Einführung in die Elektronenmikroskopie* (VEB-Verlag Technik, Berlin, 1966)
[140] L. Reimer, G. Pfefferkorn, *Rasterelektronenmikroskopie* (Springer, Berlin, 1977); R. Barer, V.E. Cosslett, *Advances in Optical and Electron Microscopy*, Bd. 6 (1975) S.275; M. von Ardenne, *Tabellen zur Angewandten Physik*, Bd I (VEB-Verlag der Wiss., Berlin, 1962)
[141] F. Lenz, Habilitationsschrift, TH, Aachen, 1957
[142] L. Reimer, *Elektronenmikroskopische Untersuchungs- und Präparations-methoden* (Springer, Berlin, 1967)
[143] M. Isaacson, D. Johnson, Low Z elemental analysis using energy loss electrons, in *Proc. 8 SEM Symp. IITRI*, Chicago, Bd. I, 1975, p. 157
[144] A.V. Crewe, J. Wall, L.M. Welter, J. Appl. Phys. **39**, 5861 (1968)
[145] P.W. Hawkes, *Image Processing and Computer aided Design in Electron Optics* (Academic, London, 1973)
[146] P.W. Hawkes, *Electron Optics and Electron Microscopy* (Taylor & Francis, London, 1972)
[147] K.H. Hanßen, Z. Angew. Phys. **20**, 427 (1966); F. Thon, Z. Naturforsch. **20a**, 154 (1965)

[148] F. Thon, Z. Naturforsch. **21a**, 474 (1966); F. Thon, B.M. Siegel, Berichte der Bunsen Gesellschaft für Physikalische Chemie, Bd. 74, 1116 (1970); K.-J. Hanszen, Optik **32**, 74 (1970)
[149] Titan 80–300 TEM, see http://www.feicompany.com/
[150] K. Honda, S. Takashima, JEOL News **38**, 36; S. Uno, K. Honda, N. Nakamura, M. Matsuya, J. Zach, Optik **116**, 438 (2005)
[151] P.J. Goodhew, F.J. Humphreys, *Electron Microscopy and Analysis* (McGraw-Hill, London, 1991)
[152] H. Rehme, E. Wolfgang, Siemens Zeitschrift **49**, 732 (1975)
[153] H.-J. Hunger (ed.), *Werkstoffanalytische Verfahren* (Deutscher Verlag für Grundstoffindustrie, Leipzig, 1995)
[154] C.G. Smith, *Surface Analysis by Electron Spectroscopy* (Plenum, New York, 1994)
[155] EDAX International Inc., Prairie View, Illinois 60069, USA
[156] Microspec Corporation, Sunnyvale CA94086, USA
[157] J. Walinga, X-ray elemental analysis in the electron microscope, Philips Bull. EM **108**(1, 2) (1977)
[158] W.B. Ashwell, C.J. Todd, R. Heckingbottom, Combined Auger electron spectroscopy and scanning electron microscopy, J. Phys. E: Sci. Instrum. **6**, 435–438 (1973); A. Christou, Auger-spectroscopy of solid surfaces in a dry pumped high resolution SEM, Scanning Electron Microscopy 1975, S. 149–156 IIT, Research Institute, Chic. 75
[159] S. Aksela, M. Karras, M. Pessa, E. Suominen, Study of the electron optical properties of an electron spectrograph with coaxial cylindrical electrodes, Rev. Sci. Instrum. **41**, 351–355 (1970)
[160] P. van Zuylen, Some applications of an energy analyser in electron microscopy, Philips Bull. EM **108**(1), 14 (1977)
[161] A.J.F. Metherell, *Advances in Optical and Electron Microscopy. Microscopy,* vol. 4 (Academic, London, 1971)
[162] R.P. Ferrier, R.P.T. Hills, *Advances in Analysis of Microstructural Features by Electron-Beam Techniques* (The Metals Society, London, 1974)
[163] M. Isaacson, D. Johnson, Ultramicroscopy **1**, 33 (1975); C. Colliex, V.E. Cosslett, R.D. Leapman, P. Trebbia, Contribution of electron energy loss analysis spectroscopy to development of analytical electron microscopy, Ultramicroscopy **1**, 301–315 (1976)
[164] H.T. Pearce-Percy, J.M. Cowley, Optik **44**, 273 (1976)
[165] E.M. Hörl, E. Mügschl, Scanning electron microscopy of metals using light emission, in *Proc. Vth European Congr. Electron Microscopy*, 1972, pp. 502, 503
[166] H.C. Pfeiffer, G.O. Langner, W. Stickel, R.A. Simpson, J. Vac. Sci. Technol. **19**, 1016 (1981)
[167] H. Rehme, E. Wolfgang, SIEMENS Analysentechnische Mitteilungen (166)
[168] B. Wicht, *Current Sense Amplifiers for Embedded SRAM in High-Performance System-on-a-Chip Designs* (Springer, Berlin, 2003), p. 123

5

Ion Beam Devices for Material Processing and Analysis

F. Rüdenauer, H.W.P. Koops, G. Hobler, L. Palmetshofer, and H. Bluhm

5.1 Ion Beam Analysis

5.1.1 Historical Development

The discovery in the 1930s that nuclear reactions can be induced by the impact of heavy atomic projectiles fostered research in the fields of heavy ion particle accelerators on one side and theoretical and experimental studies of atomic collisions and energy loss of fast particles in condensed matter on the other side. People like R. van de Graaff in the US and, earlier, Cockroft and Walton in Europe, designed efficient high voltage generators in the low MeV range and people like E.O. Lawrence and H. Enge designed corpuscular optics beam lines for focusing and analysis of high energy ion beams. Ion accelerators contributed heavily to the field of nuclear reactions and our knowledge on radioactive and stable nuclides. With the upcome of the age of microelectronics in the 1960s and 1970s the interest in material sciences exploded. Both, ion accelerator devices and theoretical knowledge on heavy ion collisions and propagation of ions in matter were well developed and ready to contribute to the fields of materials analysis and surface physics.

A second driver for the field of ion beam analysis was space science. When it was decided in the 1960s to send a man to the moon, it became obvious that new methods for microchemical analysis, both for material recovered from the moon and for robotic on – site analysis on planetary bodies were needed. For the latter purpose, well developed nuclear techniques (α is the particle backscattering using radioactive sources) were applied successfully, first on the SURVEYOR V Lunar mission in 1967. The former requirement gave rise to the development of a new generation of imaging microanalytical instruments with spatial resolutions down to the sub-micrometer range, e.g. the scanning electron microprobe and the ion microprobe (SIMS). Both, high and low energy ion beam analytical techniques since then

have been continuously improved and have contributed to many fields, e.g. material science, nuclear science and safety, biology, environmental science and many others.

5.1.2 Electrophysical Fundamentals

Ion beam scattering analysis is based on the fact that a high energy projectile ion is scattered in the atomic potential of a target ion (usually at rest). In most cases, such a collision can be considered with sufficient accuracy as a binary collision governed by conservation of energy and momentum. If observation is done at a fixed scattering angle, the energy of the scattered projectile depends on the ratio of target/projectile atomic mass only. If the projectile mass is known, the target mass can be inferred by measuring the energy of the scattered particle. This simple consideration holds for an isolated target atom. If the atom is embedded in a solid, the energy loss of the ingoing and of the outgoing scattered projectile must be taken into account. The energy loss $\mathrm{d}E/\mathrm{d}x$ per unit path length essentially consists of a contribution due to nuclear stopping (Coulomb interaction of screened nuclear charges of target and projectile) and a second contribution due to electronic stopping, i.e. interaction of the projectile with lattice atoms. Nuclear stopping dominates at energies up to $\approx A$ keV, where A is the atomic weight of the incident ion. The energy of the projectile determines how deep it can penetrate into a solid, make a large angle collision with a target atom and reescape from the solid again in order to yield information on the mass and the subsurface depth of the scattering target atom. The high energy (a few MeV) accelerator methods (e.g. RBS) have an accessible depth into the solid of a few micrometers. The low energy ion scattering technique (LEIS) with a projectile energy of a few 100 eV to a few keV is a purely surface sensitive method. Analysis can also be performed on the basis of electromagnetic radiation emitted by the target atoms, as a consequence of electronic and nuclear excitation by the projectiles. The particular processes are mentioned below in the description of the respective analysis techniques.

5.1.3 Analysis Techniques

"Ion beam analysis" is the generic name for a class of microchemical analysis methods for solid materials, in which ion beams in vacuum play a central role. Ion beams stimulate emission of electromagnetic or corpuscular radiation from the sample, thereby obtaining information on the kind, spatial distribution (in up to 3 coordinates) and, to a limited amount, on the chemical bonding of elements present in the sample. These ion beam methods can be classified according to the energy of the stimulating (primary) ion beam [1, 2], Table 5.1. A schematic for the experimental arrangement used in high beam energy methods is shown in Fig. 5.1.

In *Rutherford backscattering* analysis (RBS) a 0.5–2 MeV He^{++} ion beam from an accelerator penetrates into the superficial layer (0–$\approx$2 μm) of the solid sample and is backscattered from the atoms in this region. From the energy of He ions backscattered into a given direction the mass of the scattering target atom and its distance from the sample surface can be determined.

Table 5.1. Ion beam analysis methods for chemical analysis [3]

Method	Typical applications	Elements detected	Detection limits	Depth resolution	Imaging/ mapping	Lateral resolution
RBS	Quantitative thin film composition and thickness	Li–U	1–10 at% ($Z < 20$) 0.01–1 at% ($20 < Z < 70$) 0.001–0.01 at% ($Z > 70$)	2–20 nm	Yes	2 mm
PIXE	Quantitative thin film composition and thickness	Na–U	1 ppm	2–20 nm	Yes in Nuclear Microprobe	1 μm
PIGE	Quantitative thin film composition and thickness	Li–K	10–1000 ppm		–	
SIMS	Dopant and impurity depth profiling, surface, and microanalysis	H–U	1e12–1e16 at/cc (ppb-ppm)	5–30 nm	Yes	1 μm imaging, 30 μm depth profiling
FIB	Cross-sections, resist, defects, thin sections for STEM/TEM	B–U (EDS mode) H–U (SIMS mode)	0.1–1 at%		Yes	>3 nm SEM >7 nm FIB

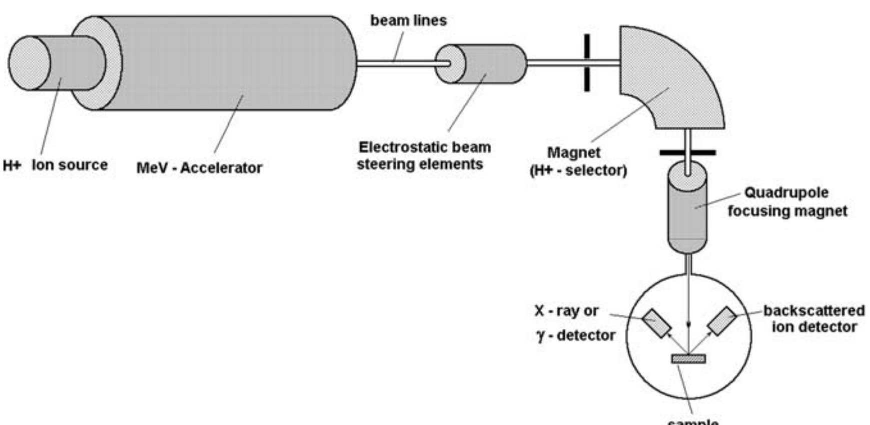


Fig. 5.1. High energy ion beam analysis methods, schematic; other beam species used are D, He

Proton Induced X-ray Emission (PIXE) makes use of the ionisation induced in the core levels of the sample atoms bombarded by a MeV proton beam [4]. Deexcitation of these deep holes gives rise to emission of characteristic X-rays, thus permitting identification of the elements present in the sample. Compared to electron based X-ray analytical techniques such as energy dispersive spectroscopy (EDS), PIXE offers better peak to noise ratios and consequently much higher trace element sensitivities, down to the ppm-level for certain elements. In a special modification of the method (*Proton microprobe*), the bombarding proton beam is focused down to a diameter of the order of 1 μm and raster-scanned across the sample surface, thus permitting to record 2D-images of the spatial distribution of elements in the surface region of the sample [5].

If the energy of the proton beam is increased to 3–4 MeV the nucleus of isotopes can be excited, leading to emission of γ-rays (*Proton Induced γ-ray Emission* [6], PIGE). Energy analysis of the emitted γ-rays allows quantitative and qualitative elemental analysis, in particular of light elements, which are often difficult to determine by other analytical techniques. The sensitivity of PIGE varies greatly from isotope to isotope.

In PIXE and PIGE the proton beam can be brought out from the high vacuum environment of the accelerator into the ambient of the laboratory (external beam). This technology makes it possible to analyse valuable artifacts or precious materials at atmospheric pressure without having to take small subsamples. The attenuation of soft X-rays in ambient air can be partially avoided by purging the area between sample and X-ray detector with He [7].

Applications of the high energy ion beam analysis methods are in mineralogy, geology, cosmochemistry, biomedical research [8], arts, microelectronics, etc. Instrument manufacturers can be found in the references [9, 10].

In *Secondary Ion Mass Spectrometry* (SIMS) a beam of ions (400 eV–20 kV) is bombarding a solid sample. This “primary” ion beam generates a collision cascade in the superficial layers of the sample, finally leading to emission of surface atoms and molecules (sputtering). A small fraction (10^{-3}% to 10%) of the sputtered particles is positively or negatively ionised (secondary ions). Mass analysis of secondary ions allows high sensitivity elemental analysis of the sample [11, 12]. Since the range of emitted secondary ions in the solid is only a few nanometers, SIMS intrinsically is a surface analytical technique. Secondary ion yields (emitted ions/primary ion) vary over a range of 3–4 orders of magnitude [13] and cannot be predicted from first principles. In all operating modes, quantification of elemental concentrations from SIMS spectra therefore is only possible using external or internal standards [11].

If the primary beam density is chosen such that during the course of an analysis, a microarea of the sample, corresponding to the extension of the collision cascade (typically 20–30 nm) in the average receives not more than a single primary ion, the original molecular structure of the sample or its adsorbed layers is reflected in the mass spectrum of the emitted particles (*Static SIMS*). This mode of analysis therefore is applied to problems in organic chemistry, medical and pharmaceutical research, structural chemistry, etc. Detection sensitivities are reaching down to 1 ppm of a single monolayer, sample damage and erosion is negligible.

If the primary current density is chosen higher (> 10^{10} ions/cm^2 s), surface molecules are destroyed by repeated ion impact at the same location and sputter erosion of the surface becomes non-negligible. This mode of *Dynamic SIMS* therefore is mainly limited to elemental analysis of solid samples. Due to the higher ion currents used, dynamic SIMS is an extremely sensitive solids analysis technique with sensitivities for many elements in the ppb range. The spatial distribution of elements at the solid surface (2D-imaging SIMS) can be determined either by focusing the primary beam and scanning this microbeam across the surface (scanning SIMS), or by making use of the imaging properties of electrostatic–magnetic mass spectrometers (direct imaging SIMS, Fig. 5.2) or by a combination of both (image dissecting SIMS) [11]. Spatial resolutions down to 20 nm have been reported [14]. By combining imaging SIMS with sputter erosion, intrinsic to primary ion bombardment, the 3rd spatial dimension (sample depth) is accessible. In this mode (3D SIMS) [15] the sample is continuously sputtered and 2D images are recorded at regular time intervals. From this 3D image stack the 3D distribution of elements in a surface–near microvolume

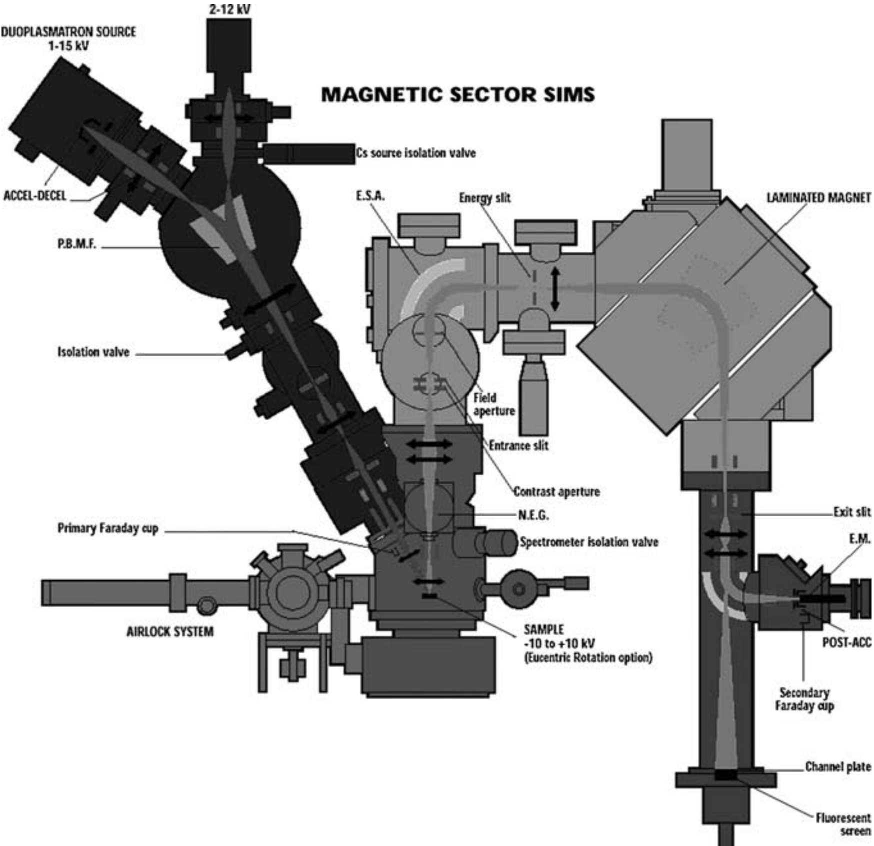


Fig. 5.2. Commercial double focusing imaging SIMS; schematic

can be reconstructed in ideal cases. A special case of 3D analysis is depth profile analysis where the concentration of elements is determined as a function of sputter depth, either directly by recording elemental signal intensity as a function of sputtering time or retrospectively by reconstruction from a 3D image stack. This analysis mode is heavily used in microelectronics device manufacture and has contributed essentially to the commercial success of SIMS [16, 17].

Mass analysis in SIMS is performed using double focusing mass spectrometers (DFMS, Fig. 5.2), time-of-flight mass spectrometers (TOF–SIMS [18–20]) or quadrupole mass spectrometers. DFMS is the standard technique for high sensitivity depth profiling and trace element and isotopic analysis. TOF–SIMS, due to simultaneous recording of all masses is of advantage for elemental and imaging analysis of extremely small sample volumina, quadrupole SIMS is an economical alternative for medium sensitivity elemental analysis and static SIMS applications.

Quadrupole dynamic SIMS often is coupled with Focused Ion Beam (FIB) micromachining equipment [21] (see Sect. 5.2). Here SIMS allows to identify defects and map elemental distributions in microelectronic circuits.

Accelerator Mass Spectrometry (AMS) is a specialised ultra-sensitive method for extreme trace analysis [22]. It basically is a medium mass resolution dynamic SIMS technique in which isobaric molecular mass interference, usually a sensitivity limit in standard SIMS instruments, is strongly reduced by accelerating the secondary ions to MeV energies and sending them through a gas or thin foil stripping cell. In the stripper, all molecules are dissociated into their atomic components so that only the selected atomic or isotopic species passes on to the detector.

SIMS is applied in microelectronic device technology, metallurgy, nuclear technology [23]; environmental analysis, biomedical research, geology, cosmochemistry [24, 25] and many others.

In *Ion Scattering Spectroscopy (ISS)* [26, 27], in particular, low energy scattering spectrometry (LEIS), mono-energetic ions are scattered by the atoms in the surface. According to the laws of conservation of energy and momentum, the energy of the backscattered ions is characteristic of the mass of the target atoms from which they are scattered. The energy spectrum obtained can thus directly be interpreted as a mass spectrum of the surface atoms. The information depth of low energy (1–5 keV) ion scattering (LEIS) is limited to one atomic layer because the low-energy noble gas ions have a high neutralization probability. This results in a negligible scattered-ion yield from target atoms below the surface layer. If the ion energy is increased, we talk of medium energy ion scattering (MEIS) or high energy ion scattering (HEIS), in their surface sensitivity. These techniques are mainly used in fundamental research or very specialised areas of industrial interest (e.g. catalysts).

5.2 Ion Beam Materials Modification and Mask Repair

5.2.1 Historical Development

Focused ion beam (FIB) systems were developed since 30 years [28], and are produced commercially for approximately ten years, primarily for large semiconductor

manufacturers. FIB systems operate in a similar fashion to a scanning electron microscope (SEM) except, rather than a beam of electrons and as the name implies, FIB systems use a finely focused beam of gallium ions that can be operated at low beam currents for imaging or high beam currents for site specific sputtering or milling.

The focused ion beam (FIB) system for nanomachining, mask repair (MRP) and circuit editing (CE) consists of an ion optical column containing an ion source, extraction electrodes, an optional mass separator, electrostatic focusing lens, and a beam blanking and deflection system. In a high vacuum chamber, a precision controlled stage with sometimes interferometric position control is used to place the sample under the beam. The beam position is controlled by a computer with special lithography software. Modifications of the sample are performed using a specialized process control software. Stability and resolution of such systems is in general better than 10 nm.

5.2.2 Electrophysical Fundamentals

The key element is the ion source. The liquid metal (LM) source consists of a sharpened tungsten needle, 0.02 to 0.002 cm in diameter, wetted with a thin film of molten metal such as gold or gallium. Ionisation of the molten metal takes place at the apex of the needle when a critical voltage is applied. The interaction of electrostatic and surface tension forces between the tip and exterior electrode causes the liquid metal to form a peaked cone (Taylor cone) of a small diameter. LM sources have a typical brightness of 1×107 A/cm^2 sr (amperes per square centimetre and solid angle) at an angular current density of 10–50 μA/sr. The energy spread is 5–21 eV at these conditions, with a virtual size of 10–50 nm. Positive ions of any metal, and from components of eutectic alloys with low melting points can be generated, although Au, Ga, Si, and Li are the most common (pure materials such as Li have such high chemical reactivity that their lifetimes as an ion source is unreliable and short-lived). The energy spread has been found to increase with both current and particle mass. It results from the Coulomb interaction of the ions at points of high density, like the cone apex of the source, and is especially large when the ions are still slow. The utilization of LM sources has been increased using eutectic alloys and separating the alloy constituents by velocity filters constructed from crossed electric and magnetic fields ($E \times B$-filter). Patterning of resist with heavy ions uses the advantage of low penetration depth and high sensitivity of polymers for ion bombardment. The lifetimes of the sources is limited to 3 months due to corrosion of the tungsten needle. Ions undergo a variety of sample interactions, see Fig. 5.3.

The Ga ion beam of typically 100 pA and 30 keV energy interacts with the sample through several mechanisms: It generates secondary electron emission, secondary ion emission, sputtering of atoms, electromagnetic radiation like X-rays, it interacts strongly by physical sputtering resulting in amorphisation of the sample top 30 nm layer, and modifies the sample by ion implantation. Ga is mixed into the deposits and the top surface layers. Ion impact produces displaced substrate atoms (damage of the crystal lattice) even as deep as 60 nm by channelling and collision cascades. Nuclear and electronic excitation, as well as heating is generated. Bulk chemical reactions

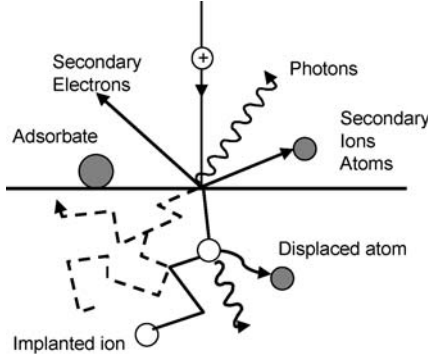


Fig. 5.3. Interaction channels for ions with the sample

and surface chemical reactions are initiated. The strongest of the above mentioned effect is the sputtering of the sample [29].

5.2.3 Present State-of-the-Art and Applications

State of the art applications are described in the literature [30]. Working with a focused ion beam is a slow serial process, but a unique combination of imaging, ion implantation, material removal and deposition. It is used for prototyping, fabrication, repair, inspection of high value items, for failure analysis of integrated circuits, re-wiring of integrated circuits, lithographic mask repair, hard disk read/write head machining, direct fabrication of integrated circuits, prototyping of MEMS devices, and micro/nano fabrication of high value objects like sensors. Table 5.2 shows characteristics of production and research FIB systems.

The most common applications of focused ion beams are, despite of micro fabrication, the refinement of sensors, like magnetic heads for disk drives, photo-mask repair (PMR), failure analysis in semiconductor production (FA), and circuit editing (CE).

Photomask Repair

There are many defects in a photomask, which are of different appearance. One finds two classes of defects, open and closed areas. Open defects require the addition of absorbing or phase shifting material, and closed defects require the removal of surplus material. Figure 5.4 shows a variety of defects. Most defect sizes are well below the CD and require therefore a repair system capable of very high resolution.

A $4\times$ Mask photo mask for a fabricated device of a certain critical dimension (CD), e.g. at 90 nm, has a critical dimension at the mask level of 360 nm. In the year 2000, a transition in optical lithography for wafer exposure systems took place. Till then binary photomasks made from chrome absorbers on quartz blanks rendering

Table 5.2. Characteristics of production and research FIB systems

Focused ion beam systems	Without mass separation	With mass separation
Application	Micromaching lithography	Implantation
Accelerating voltage	20–50 kV	5–200 kV
Current density	1–10 A/cm^2	0.1–1 A/cm^2
Beam diameter	5–500 nm	50 nm
Ion species available	Ga (elemental source)	As/Si/Be, Pd/B/As(alloys)
Systems in the field	>800	<30

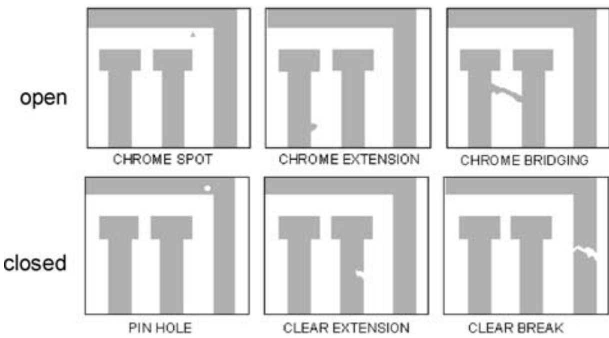


Fig. 5.4. A variety of defects in masks is classified as open and closed defects

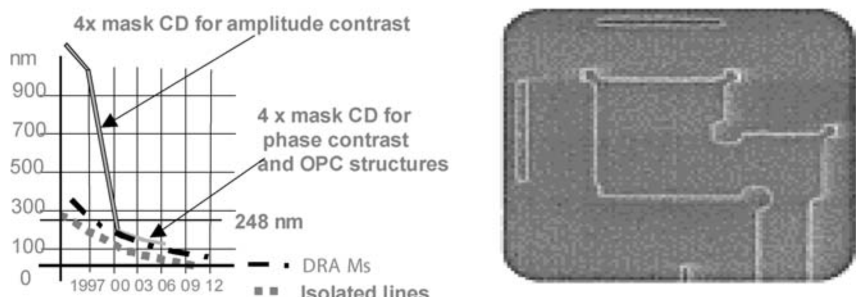


Fig. 5.5. *Left*: Due to the change from amplitude to phase contrast, 1× structures for OPC and Phase shifters have to be correct in photo- and phase shift masks. *Right*: Serifs, hammer heads and phase shifting lines of 1× size shall modify the phase contrast to deliver the correct phase contrast image

amplitude contrast device patterns were used in the 4× reducing imaging steppers. However, when the masks critical dimension reached the size of the wavelength of the UV light used in the stepper, the imaging contrast mechanism changed to phase contrast imaging, see Fig. 5.5, left. To obtain correct phase contrast intensity distributions for proper exposure of the resist, the masks patterns were corrected with optical proximity effect correcting structures (OPC). Those additional structures are as small as the CD at the wafer, which is in our example 90 nm. Such details have to be re-

paired, see Fig. 5.5, right. For mask repair of open defects an ion-induced deposition process is employed to cover the areas with an opaque carbon or platinum/carbon film, see Fig. 5.5.

The process offers fast, accurate, clear repairs that require no trimming and can be made anywhere on the mask, including isolated quartz areas. The technique consists of introducing a hydrocarbon or platinum containing organometallic precursor gas in the vicinity of the defect, while scanning it with the ion beam. The gas is broken down and deposited only where the beam has been scanned. The result is a durable opaque film consisting of carbon or Pt/C and gallium. These repairs are not affected by standard cleaning processes. The thickness of the deposition can be varied to make the area slightly transmissive when repairing attenuated phase shift masks.

Gas assisted etching repairs closed or opaque chrome defects. In order to reduce quartz damage, a reactive gas is introduced into the system near the surface of the mask. When the ion beam is scanned over the chrome, it causes the gas to react preferentially with the chrome and etches the chrome away. This produces chrome repairs with minimal quartz damage and improved transmission. An adaptive beam blanking technique can be used whereby a bit map of the defective area is constructed from image data, and based upon this map, the beam is blanked when scanning exposed quartz and unblanked when scanning chrome. This method, combined with the etching gas, minimizes undesired etching of the quartz area in the vicinity of the defect and frees the operator from having to carefully outline the defect area. This method produces repairs with "riverbeds" less than 35 nm deep and transmission better than 95% at i-line. Such repairs are usefull for 248 nm lithography and somehow for 193 nm lithography. However, the Ga implantation during the review and repair process is so severe that it limits the use of this technology.

Embedded shifter masks have become the phase shift technique of choice for many mask makers. Typical films include MoSiON and CrO_x. These can be repaired in a manner similar to binary chrome masks and the masks image quite well in the wafer stepper system. Extra shifter material is etched away by the ion beam. Missing shifter can be repaired by making a thin carbon deposition to match the transmission of the shifter material. Embedded shifter masks are often used for contact levels because they are the most critical of all levels and require the best resolution. One common defect is a missing contact hole. This can be duplicated from a good contact and placed in the proper location using a pattern copy software.

Since the ions have a low penetration depth into the material, the top surface suffers sputtering by ion impact. This allows to erase surplus material of closed defects in photo masks. Since the sputtering rate is dependent on the angle of the beam impact and is highest close to 90° with respect to the surface, all ion sputter work is characterized by "river bedding", which is the appearance of a deeper etch pit at the rim of the sputtered area. Such uncertainty leads in phase shift masks to additional phase shift. An allowed 1 degree of phase error in the 90 nm node translates to 2 nm Quartz material thickness. This is a very high accuracy requirement for the detection of the endpoint of the sputtering work.

Since Ga ions are positive and generate a large number of secondary electrons at impact (typically 20 per ion), which are extracted for imaging, the sample will

charge positively in the top layer. This charge can have a high density and generate very high potential differences at the sample during the repair work. Therefore, a low energy electron flood gun of 60 eV with a low current density is used to illuminate the mask during repair work, however intermitted with the imaging, to allow the existing charge islands to attract the negative charge required for compensation of the potential differences, see Fig. 5.6. Figure 5.7 shows: (left) the repair action schematic; and (right) a result of an open (top) and a closed effect repair (bottom).

For larger and in time lengthy repair work, especially at low currents, a drift compensation routine is employed, where a hole is drilled into an opaque area and intermittently imaged during the repair work, and after finishing the repair the hole is closed again by deposition. This procedure allows to work on 70 nm defects, but reduces the throughput of the work.

At higher resolution required for the repair, during the imaging of the defect already Ga ions are implanted into the mask, and the mask is sputtered. This reduces

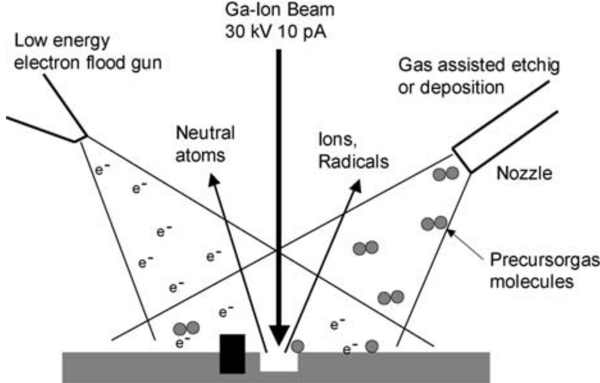


Fig. 5.6. Schematic of the focused ion beam (FIB) repair method. Molecules are issued to the mask for deposition and beam assisted etching. Charging is compensated by an electron beam of low energy flooding the sample

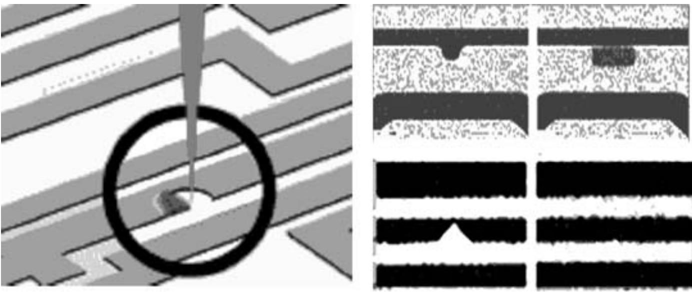


Fig. 5.7. *Left*: The deposition repair action schematic; and *right* closed defect repaired by ion milling (*top*) and repair of an open defect by deposition (*bottom*)

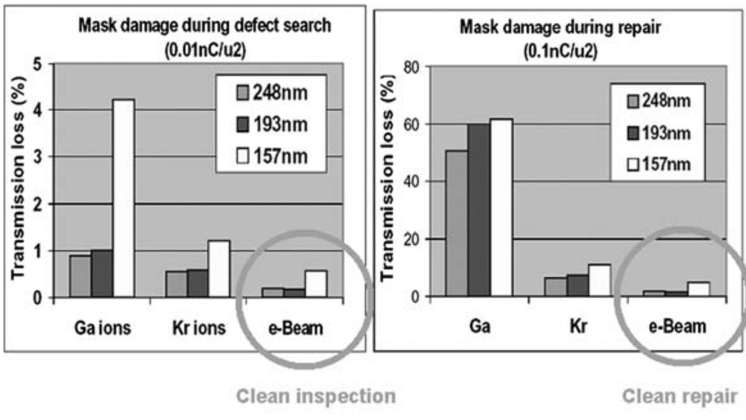


Fig. 5.8. The comparison of ion- and electron-beam mask imaging and repair shows that for 157 nm lithography it is inevitable to use electron beam mask repair

the transmission at the observed site and makes the application of FIB repair impossible for deep UV lithography masks.

A comparison of ion and electron beam mask imaging and repair is shown in Fig. 5.8. Only e-beam based mask repair tools meet the requirements of damage free defect review and repair for advanced mask technologies.

Failure Analysis by Ion Beam Milling and Electron Beam Imaging

For defect review and analysis the semiconductor industry employs dual beam systems in their production facilities. The workstation is equipped with a 30 kV ion beam, a high resolution SEM mounted under an angle to the ion beam, and a gas injection system for deposition and enhanced etching. It is used for wafer inspection, cross-sectional preparation and structural examination. The most limiting aberration is the chromatic aberration of the electrostatic ion beam probe forming system.

JEOL Ltd. has the slogan “cut and see” for this sample preparation. A wafer from the production line is inspected with a secondary electron microscope for irregularities in the pattern appearance, like bumps on a line. This indicates that a particle is located in a lower layer of the device. Using EDX analysis, a volume of 1 μm below the surface is analysed for the materials composition of this location. To image the defect, a wedge shaped hole is sputtered into the wafer with the ion beam. The wedge is required to allow the electron beam to image the side of the cut wedge. It is preferred for imaging, since scanning the sample with the ion beam sputters and changes the sample.

As an example, the specification of a commercially available crossbeam system is given in Table 5.3. This system offers a high resolution SEM and a FIB under an angle of 51° in one machine for the cost of US$ 500 000. An example for deposi-

Table 5.3. Specification of a ZEISS 1540XB CrossBeam

	SEM	FIB
Resolution	1.1 nm @ 20 kV 2.5 nm @ 1 kV	7 nm @ 30 kV guaranteed, 5 nm achievable
Magnification	20×–900k×	600×–500k×
Probe current	4 pA–10 nA	1 pA–50 nA
Acceleration voltage	0.1–30 kV	3–30 kV
Emitter	Thermal field emission type	Ga liquid metal ion source (LMIS)
Detectors	In-lens: Annular type Chamber: ET type BSD: Optional solid state or scintillator type QMS: Optional, STEM: Optional	
Operating pressure	10^{-3} mbar or lower	5×10^{-5} mbar or lower
Specimen chamber	330 mm ø, 270 mm height IR CCD-camera included for sample viewing	
Specimen stage	6-axes fully eucentric, all motorised $X = Y = 102$ mm, $Z = 43$ mm, $Z. = 10$ mm Tilt = −10–60°, Rotation = 360° continuous AWD (analytical working distance): 5 mm	
Image processing	Resolution: from 512 × 384 to 3072 × 2304 pixel Noise reduction: Pixel averaging, frame averaging, continuous averaging	
Image display	Two flicker-free XVGA monitors with image displayed at 1024 × 768 pixel	
Image hardcopy	Choice of laser, inkjet or video print media	
System control	Integrated CrossBeam®/SmartSEM™ user interface based on Windows® XP operating systems, controlled by mouse, keyboard and joystick	
Gas injection system	Up to 5 gases for selective etching, enhanced etching, material deposition, insulator deposition	

tion with e-beam and the ion beam column attached to the microscope are given in Fig. 5.9.

Circuit Modification for Device Development

The semiconductor industry needs to produce new designed faster and more powerful computer chips in shorter production cycles. To save money and time on the production of mask sets for the wafer stepper, Circuit Editing (CE) is used. For this functioning cells with critical devices areas are fabricated with variations in one mask set, and one wafer is processed in the production line. This wafer contains a variety of optimized designs for a problem in isolated islands. If necessary, an insulating top layer is milled away with the ion beam to access the metal line. Using the ion beam, electric power is connected to the device by depositing conducting lines. After testing of this part of the circuitry the lines are cut again and other parts are connected for testing. Due to bad endpointing and Ga staining of insulators the FIB CE method

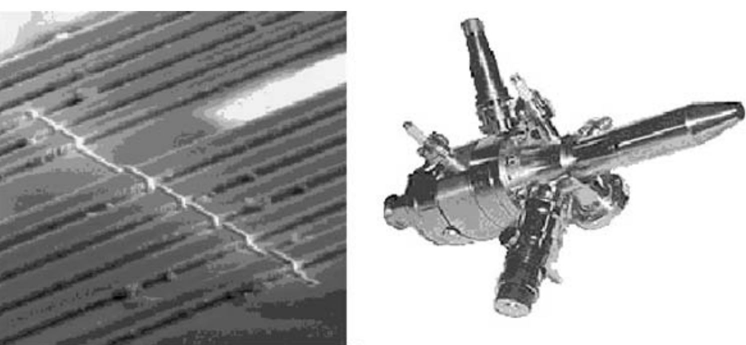


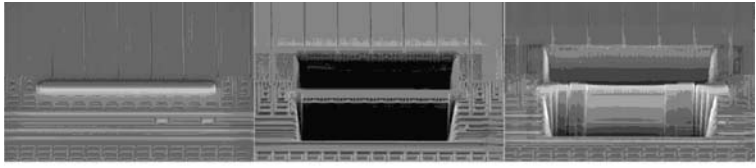
Fig. 5.9. *Left*: Zeiss Crossbeam W-deposition; and *right*: Orsay Physics Ion Beam column as it is used in the Zeiss Crossbeam

is only usable for 65 nm CD devices. Devices with smaller CD will need electron beam CE systems.

TEM-Lamella Preparation for Defect Review and Analysis in Semiconductor Manufacturing

To fabricate a cross-section, usable as a sample in a high tension Transmission Electron Microscope (TEM), the location is discriminated, generally imaging the surface with the SEM. A surface irregularity indicates that a defect is present in the depth of the semiconductor circuitry. To analyse the nature of the defect, it is necessary to cut a thin slice of the semiconductor levels using the focused ion beam and sputtering away left and right wedges of the slice at the defect location. To obtain a thin lamella, it is necessary first to deposit a line with a fine electron beam and tungsten deposit from tungsten-hexacarbonyle. Then the milling with the ion beam is performed. Generally a 30 keV Ga ions beam having a focus spot of 2 nm diameter is used. For extraction, the lamella is approached with a sharp tungsten needle using a micromanipulator. By depositing additional W on the needle and the top of the lamella this is fixed to the needle. Then the lamella is cut out at the sides and the bottom using the sputtering ion beam. Using the manipulator, the needle with attached lamella is extracted, and the lamella is then placed onto a TEM grid sample carrier coated with a holy carbon foil. It is again fixed with some W-deposition to the foil. Then the transfer of the grid with sliced lamella into the TEM is possible, and the high resolution analysis can be performed. The whole procedure takes at least 40 min. Steps of this procedure are displayed in Fig. 5.10 top and bottom rows. It is inevitable to perform such an analysis during several steps of producing a silicon chip on the wafer in the fabrication process line to obtain defect free production conditions. This is necessary to improve the yield of the fabrication lines.

The throughput for deposition and etching processes depends on current density and material supply, as discussed for ions [31] and for electrons [32]. Ions and electrons have a processing throughput limited for small defects ($<1\ \mu m^2$) by the gas supply and are therefore equal in throughput [33].



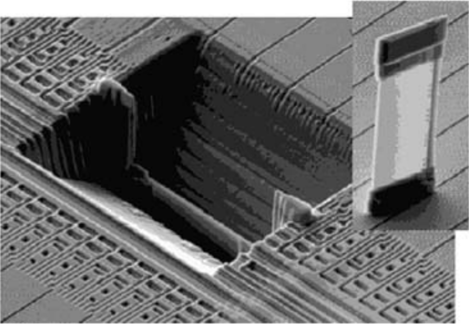


Fig. 5.10. *Top left*: the deposition of the tungsten line; *middle*: the etching of the cavities (*left* and *right*) of the lamella and *right*: the thinned lamella. *Bottom left*: the empty pit after extraction of the lamella; *right*: the inset (Courtesy ZEISS SMT AG)

5.2.4 Future Aspects

3D Rapid Prototyping and Microstructure Fabrication

For 3D structures, rapid prototyping of nanostructures, and microstructure modification a set of data for the beam control system its to be generated from a (3D) model of the device or structural change. Since deposition can only build upwards volume structures are sliced into layers, which are deposited sequentially. The schematic of data preparation and fabrication of 3D structures for rapid prototyping of nanostructures is given in Fig. 5.11 [34].

Results of 3D rapid prototyping are shown in Fig. 5.12, left: a lamella of a butterfly's wing acting as an omni directional optical mirror for visible light, right: miniature wine glass deposited from carbon polymer precursor [35, 36].

Scanning ion microscopy technology has opened a new door to forensic scientists, allowing the gun shot investigator to see inside a particle's core. Using a focused ion beam, particles can be cross-sectioned, revealing the interior morphology and character that can be utilized for identification of the ammunition manufacturer [37].

5.3 Ion Implantation

5.3.1 Historical Development

Early observations of the ability of energetic ions to modify material properties were made by nuclear scientists involved in the development of nuclear reactors during

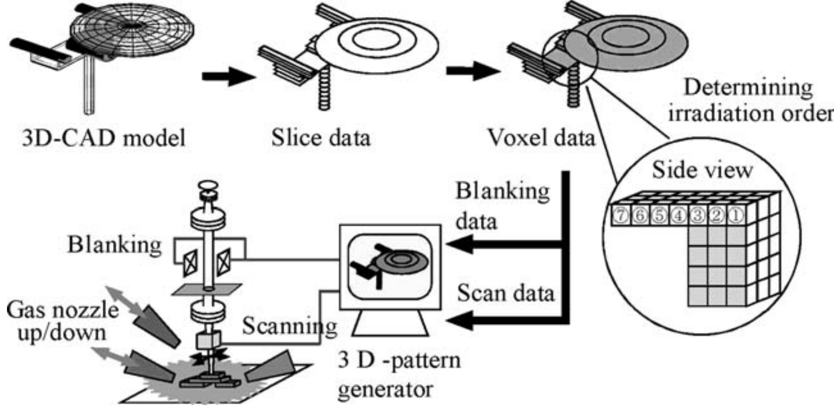


Fig. 5.11. Schematic of data preparation and fabrication of 3D structures for rapid prototyping of nanostructures

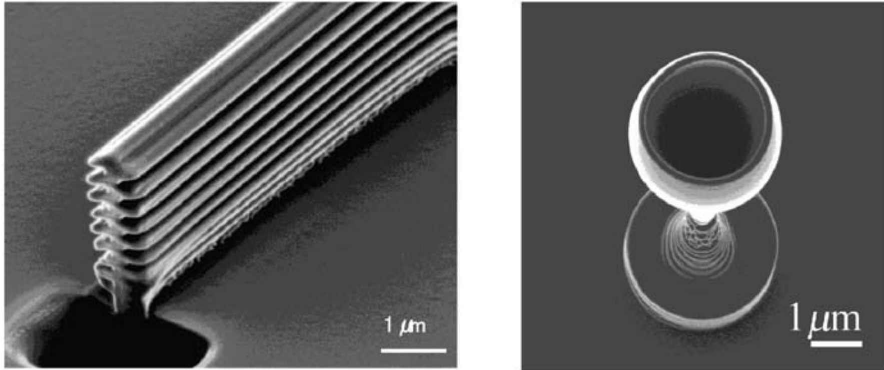


Fig. 5.12. *Left*: 3D structure fabricated after the design of the lamella on butterfly wings; *right*: nano-wineglass

World War II. These material changes, e.g. void formation, were caused by the recoiling of target atoms by fast neutrons and were regarded as deleterious only. Intentional use of ion beams for material modification was first attempted in the 1950s in the rising field of semiconductor device fabrication. In 1957 Shockley was granted a patent on ion implantation as a doping technique of semiconductor devices [38], where the ability to control the depth location of the implanted impurities by the energy of the ion beam, and the need for thermal annealing after implantation were described.

In the 1960s the theory of the stopping and the range of ions in solids was developed (LSS theory [39]), and the phenomenon of ion channeling in crystals was discovered [40]. The scientific breakthrough came in the mid 1960s when several nuclear research labs (Chalk River Nuclear Labs, Oak Ridge National Labs, AERE Harwell) started research programs on the physics and technology of ion implanta-

tion, and ion implantation became the subject of international conferences (e.g. Ion Beam Modification of Materials, Ion Implantation Technology, etc.).

Semiconductor industry was at first reserved against the new and costly technology. However, the interest grew quickly when ion implanted devices showed real advantages in performance over conventionally manufactured devices. The first commercial ion implanters were manufactured in 1973 and had a typical performance of 100 keV energy and 100 μA ion current. The following years were characterized by efforts to increase the current output (10 mA) and, rather recently, to expand the energy range to high energies (several MeV) and to very low energies (1 keV and below). Nowadays, ion implantation is indispensable for the production of, especially, CMOS devices [41]. The implanters have sophisticated process control (dose, implant angle, wafer temperature, contamination, etc.) and are capable of implanting up to 200 wafers per hour.

The use of ion implantation in other fields of materials science was studied intensively over the years. For instance, by ion implantation it is possible to change the chemical composition of the surface, to create a surface alloy or to synthesize a surface chemical compound. Such surface layers may have outstanding wear or corrosion properties. Because of the high doses needed, which result in long implantation times and high costs, the technique is only used for special niche products, e.g. medical implants.

5.3.2 Electrophysical Fundamentals

Ion implantation is a process for the modification of physical or chemical properties of solids in sub-surface layers. Atoms or molecules are ionized, accelerated in an electrostatic field, and then made to impinge on a target. The ions are slowed down inside the target by elastic collisions with screened nuclei and by inelastic interaction with electrons [42]. The range of the ions is between several nanometers and several micrometers, depending on the acceleration voltage ($\approx$1 kV to several MV) and on the mass of the ions and the target atoms. In crystalline targets, ions may have a larger range as they are steered down the open channels between the rows or planes of atoms (channeling [43]). During the stopping process, the ions recoil target atoms from their sites and thus create crystal damage. A thermal treatment is usually applied to anneal the defects.

Because of the statistical nature of nuclear scattering, the ions do not stop at a pre-determined position. Their distribution may be approximated, in the simplest case, by a Gaussian function characterized by its mean value (“projected range”) and standard deviation. In semiconductor process modelling, more sophisticated analytical functions are used nowadays [44], which allow a more accurate description of depth profiles as well as of 2- and 3-dimensional dopant distributions, e.g. at mask edges. Alternatively, ion and damage distributions may be calculated by Monte Carlo simulations [45–47].

The process of ion implantation allows the modification of almost any property of a surface layer of the target. Semiconductors, metals, insulators and superconductors have been used as targets. After ion bombardment, electrical, optical,

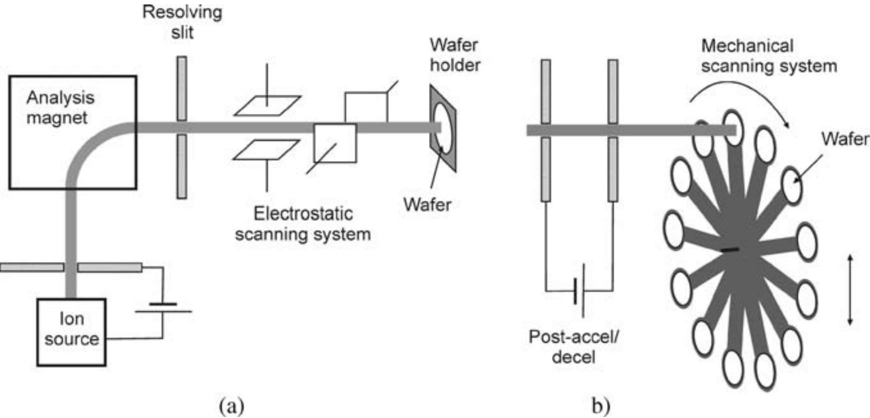


Fig. 5.13. a Traditional ion implanter with electrostatic scan system; **b** end station of a batch implanter with an additional acceleration/deceleration stage that extends the energy range of the implanter

mechanical, chemical, structural and other effects have been reported in the literature.

5.3.3 Present State-of-the-Art and Applications

The basic elements of an ion implanter are the beam generating unit consisting of an ion source and extraction optics, an analysis magnet with a resolving slit for the selection of the desired ion species, a scanning system that distributes the ions uniformly across the wafer, and a wafer handling system (Fig. 5.13a). Most commercial implanters use a Bernas source, which is essentially a hot cathode (thermionic filament) ion source with a gaseous source feed and an additional vaporizer for materials which are not conveniently obtained as gases. Purely electrostatic scanning as, shown in Fig. 5.13a, causes incidence angle variations on the wafer. In today's industrial implanters the electrostatic scan in one or both directions is substituted by mechanical scan (motion) of the wafer. A typical architecture of a single-wafer (serial) implanter comprises electrical scanning in the vertical direction with a magnetic corrector further down the beamline to compensate the angle variation, and mechanical scanning in the horizontal direction. As an alternative to electrostatic scanning, the beam may also be spread out into an unscanned "ribbon" beam that exceeds the diameter of the wafer. Batch systems, on the other hand, use purely mechanical scanning with a stationary ion beam (Fig. 5.13b). The wafers are mounted on a spinning carrousel (fast scan) whose axis performs a translational motion (slow scan).

Ion implantation is the most important method for introducing dopants into semiconductors. The main advantages are the controllability of the ion dose, the flexibility in tailoring the spatial distribution of the implanted dopants, the high reproducibility,

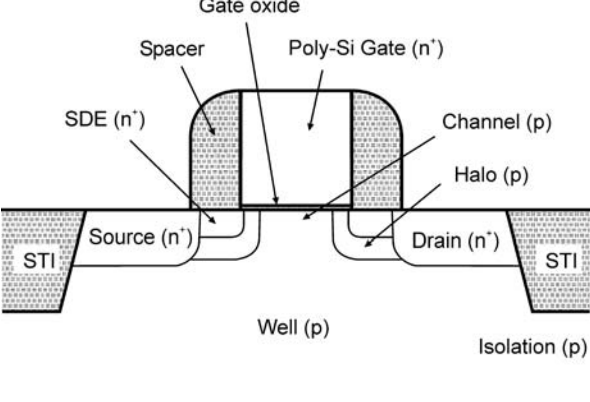


Fig. 5.14. Cross section of an n-channel MOSFET showing various doped regions. For a p-channel MOSFET n and p dopings are interchanged

and the simple masking techniques. These features are essential for semiconductor device fabrication.

In the following we demonstrate the versatility of ion implantation using a cross section through a MOSFET, the workhorse of integrated circuit technology (Fig. 5.14). In a MOSFET, a conductive path (channel) between the source and drain is generated by applying a gate voltage that exceeds a threshold which depends on the doping in the channel region. Channel implants were the first widespread application of ion implantation in semiconductor manufacturing because of the superior controllability of the resulting dopant concentration compared to the older diffusion technique or doping during epitaxial growth.

A second important application of ion implantation in MOSFET fabrication is the doping of the source and drain (SD) regions. These regions must be highly doped in order to keep the series resistance of the device low. At the same time, proper operation of the MOSFET requires the doping to be shallow compared to the channel length. Therefore, high-dose low-energy implants are required for SD doping. In practice two implants are used, a particularly shallow one reaching slightly under the gate oxide (source/drain extension, SDE) and another one further away under the contact areas. Doping of the poly-Si gate is either done together or in a similar step as the SD implant.

In order to reduce the space charge region around the p-n junction of the SDE, additional "halo" implants of the opposite type of the SDE are used. These implants are often done with a large tilt angle (20°–45°) and the gate as a mask. The well doping deeper in the bulk (cf. Fig. 5.14) has to be higher than the channel doping ("retrograde well") in order to avoid an unwanted conductive path between source and drain (punchthrough) and to isolate adjacent devices from each other. Usually multiple implants are used to optimize the well profile with respect to the electrical

characteristics of the MOSFET. Neither the flexibility in controlling the lateral extension of dopant distributions under mask edges nor the flexibility in tailoring vertical dopant profiles are provided by the diffusion technique.

Another advantage of ion implantation is the availability of almost all atoms as implant species. The most common dopants in silicon technology are B, P and As. Ge implants are commonly used to amorphize the target prior to SD implants. This allows to avoid ion channeling along low-index crystallographic axes and planes, and to activate the dopants to concentrations beyond solid solubility during solid phase epitaxial regrowth. C, F, or N ions are sometimes co-implanted with B to reduce its diffusion. The more compact profiles of Sb and In implants may be beneficial, if the lower solid solubility of these elements in Si does not matter. B, N, and Al are used as dopants in SiC, and Be, C, and Si in GaAs. Apart from doping, high-dose H implants are used for layer transfer in the production of silicon-on-insulator (SOI) material [48].

In 2004 the most advanced MOSFETs had physical channel lengths of 45 nm and SDE junction depths of 25 nm [49]. Such shallow junctions require implant energies in the keV and sub-keV range. Development of ultra-low energy implanters that deliver the required high doses in economical times was a major challenge in the 1990s. The basic problem in a high-current low-energy implanter is to reduce the space charge in the ion beam, since the space charge is proportional to the current density and inversely proportional to the ion velocity. High space charge and high associated electrostatic potentials are not tolerable in advanced MOSFET processing, since gate oxides approach thickness of 1 nm [49], and breakdown voltages exceed 1 V only slightly. Furthermore, excessive space charge leads to beam blow-up and constitutes a serious problem for beam transport. A divergent beam is also not desirable at the wafer because of its smearing effect on lateral dopant profiles at mask edges. In particular, SDE are required to be as abrupt as possible laterally to reduce the series resistance of the MOSFET.

The most important technique to reduce space charge is electron flood. Collisions with the residual gas atoms and secondary electron emission from the target provide natural sources of electrons in proportion to local beam current density. Further beam neutralization can be provided by either secondary electron flood (SEF) or plasma electron flood (PEF). In SEF, primary electrons with energies of a few hundred eV are directed to a target located close to the ion beam. Secondary electrons with about 10 eV are emitted from the target and trapped by the ion beam, thus reducing its electrostatic potential. In PEF, electrons are extracted from a discharge maintained in an arc chamber located close to the beam path. Additional neutralizing electrons come from beam-induced ionization of plasma gas bled into the beamline. It is also possible to pass the beam directly through the plasma [50]. Electrons extracted from PEF are usually less energetic than electrons from SEF, and therefore have a lower tendency of negative charging of the wafer.

In addition to electron flood, several other techniques are used to reduce space charge effects [50, 51]: Increase of the beam diameter reduces the current density for a given current and thus the space charge. It challenges, however, the design of the analysis magnet. A shorter beamline reduces the effect of space charge since the

electrostatic forces have less time to operate on the ions. Furthermore, the beam may be extracted with a higher voltage and decelerated at the end of the beamline by a voltage of opposite sign (Fig. 5.13b). The beam then travels most of the distance at higher velocity and therefore with lower space charge. Deceleration, however, has several problems that have to be dealt with. It tends to defocus the beam which results in degraded transmission to the wafer and increased beam divergence. Moreover, ions neutralized upstream of deceleration do not experience the electric field and are implanted at the higher extraction energy. They therefore cause energy contamination of the beam. They also escape the dose control system and may lead to unintended dose shifts. Possible measures to reduce the fraction of neutrals are to improve the vacuum or to add a final analysis magnet.

The effects of space charge are also reduced by the use of heavier ions, as the deflection at given speed and force decreases with ion mass. BF_2 molecular ions have traditionally been used for this reason. Also As_2, P_2, and $B_{10}H_{14}$ have been experimented with. Compared with BF_2 they have the additional advantage that 2 and 10 dopants, respectively, are implanted per (singly charged) ion. Using molecular ions, however, makes only sense if they are delivered by the ion source at rates comparable to atomic ions.

Another problem in high-current ion implantation is wafer heating [41]. Implant equipment vendors usually have to guarantee a rise in temperature of less than 80 K, since photoresist begins to flow, crack, or otherwise deteriorate around 100°C. Moreover, target amorphization becomes more difficult at elevated temperatures, so higher doses are required for amorphization, and also more residual defects remain after recrystallization. There are two ways for reducing wafer heating: First, the wafer may be cooled from the backside. The bottleneck here is the contact between the wafer and wafer pad, which normally is limited to a very small area due to surface roughness. The thermal gap between the wafer and pad may be bridged by a gas, which requires measures against bleeding into the vacuum, or by an elastomer that yields under pressure and conforms to the microscopic structure of the wafer. The pressure may be provided by clamping or, in a batch implanter, by centrifugal forces. The second way to reduce wafer heating is to distribute the power among several wafers, which is a major advantage of batch implanters (Fig. 5.13b).

Dose control is another important issue [41]. As has been mentioned, the threshold voltage of a MOSFET is very sensitive to the channel-implant dose. Excellent dose uniformity across the wafer and repeatability from wafer to wafer are necessary to achieve well-defined device characteristics. Industrial implanters usually have a closed-loop feedback system that measures the beam current and controls the scanning system in order to compensate changes in the beam current by scan speed. The current is measured by a Faraday cup that either encloses the wafer or, in a batch system, is arranged to be hit by the beam between the passage of the rotating wafers. The most common source of dose errors is charge exchange between the ions and the molecules of the residual gas, most prominently neutralization of the ions. Depending on the design of the beamline, neutral particles may be implanted into the wafer, but are not measured by the dosimetry system. In advanced dose control systems, the measured dose is corrected for charge exchanges depending on the residual gas

pressure, using semi-empirical formulae. The enhanced pressure may result from the electron flood system or from outgassing of photoresist.

The relevance of implant angle control [52] increases with device scaling because of increasing aspect ratios. The vertical dimension, e.g. of the gate stack, decreases less from generation to generation than the gate length. Shadowing effects therefore become more important, and variations in implant angle translate into larger variations in the doping distributions. Moreover, more thought has to be given to ion channeling. Ion channeling is usually considered as an adverse effect and is thus avoided, as small variations in incidence angle cause large changes in the dopant profiles. While at higher energies channeling is possible along many crystallographic axes and planes with relatively small critical angles, at low energies channeling is possible only along a few axes and planes, however, with large critical angles [43]. The traditional approach of using 7° tilt angle between the ion incidence direction and wafer normal can no longer be adopted universally since, in addition to a tendency towards smaller implant energies, there is also demand for lower implant angles to avoid shadowing and for higher implant angles to reach under the mask such as in halo implants (cf. Fig. 5.14). Major sources of implant angle variation, apart from electrostatic scanning in old implant tools, are imperfections in tuning the beamline, geometric effects in batch implanters, and crystal cut errors of the wafers.

Bombardment of solids with energetic ion beams causes sputtering which leads to wafer contamination during ion implantation [41]. Sputtered atoms originate from beamline components, uncovered areas of the process disk and materials on the surface of the implanted wafers. Since sputtered atoms have low energy, they can be found on the surface of the wafers. Elimination of most of these contaminants is possible by screen oxides grown on the wafers prior to ion implantation and removed before annealing. Contamination with energetic particles may occur with ions that have a magnetic rigidity close to that of the dopant ions and may thus be transmitted through the analysis magnet. Examples are the contamination of a $^{49}BF_2^+$ beam with $^{98}Mo^{++}$ ions from the ion source chamber and the contamination of a $^{31}P^+$ beam with $^{31}BHF^+$ ions from a previous BF_2 implant. The projected range of energetic contaminants is comparable to the range of the dopant ions and cannot effectively be suppressed by screen oxides. A careful optimization of implant conditions is necessary to keep energetic contamination below the tolerable limits. The energy contamination of ion beams in implanters with a deceleration stage mentioned above also leads to an unintended energetic implant, but with the same ion species.

The application of high-energy ion implantation (1–3 MeV) in CMOS fabrication has strongly increased in the 1990s [41, 50, 53]. The deepest implants of the vertically modulated and retrograde well structure mentioned above were done at MeV energies. The necessary energies, however, are shrinking today due to device miniaturization. A new, potential application of high-energy ion implantation is the formation of highly doped buried layers 2–4 μm beneath the silicon surface. Such buried layers can greatly enhance device performance on bulk substrates and can replace or are even superior to epitaxial silicon substrates. Highly doped silicon sub-

strates with epi layers are used because they offer latch-up suppression, low defect densities in the silicon and gate oxide, and gettering of impurity atoms. All these benefits are obtained with properly engineered implanted buried layers at a lower cost. Additionally, the end-of-range damage getters also defects induced by subsequent implants, reducing leakage currents.

Commercial high-energy implanters employ either an RF linear accelerator to obtain the high energies or use the DC tandetron principle. In a tandetron the DC acceleration voltage is used twice. First, negative ions produced in the ion source are accelerated. Then electrons are stripped away in a gas stripper resulting in positively (single or double) charged ions which are then accelerated by the same voltage again. A linear accelerator consists of an alternating series of high-voltage RF electrodes and grounded quadrupole focusing lenses. The ion beam is accelerated in phase with the RF voltage in each stage, resulting in final energies up to 20 times higher than the maximum acceleration voltage in the beamline.

5.3.4 Future Aspects

The evolution of beamline ion implantation and annealing will continue for the foreseeable future. For example, there is a need for developing ion implanters that can implant high doses at large tilt angles. In addition, several concepts such as the use of molecular ions to form ultra-shallow junctions or the application of MeV implants to produce deep gettering layers have to prove their feasibility in a production environment. The International Technology Roadmap for Semiconductors [49] also mentions plasma-immersion ion implantation (PIII) as an alternative approach for ultra-shallow junction formation. This technique makes use of ions that are accelerated by a high voltage pulse from a plasma which surrounds the workpiece to be treated [54]. The main advantage of PIII is the high current and thus short processing time achievable, its main drawback is that everything in the plasma is implanted into the wafer.

It should be mentioned that the conventional MOSFET, as shown in Fig. 5.14, will reach its scaling limit in the foreseeable future. Alternative devices, such as fully-depleted SOI MOSFETs, might be introduced for high-speed applications as soon as 2008 [49]. These devices do not need channel doping, and the SD doping may reach into the buried oxide layers, thus releasing ion implantation from some of its challenges. However, they still need lateral abruptness in the SD doping profiles. Ion implantation is therefore expected to continue to be an innovative key technology for semiconductor device manufacturing.

5.4 High Power Ion Beams

5.4.1 Historical Development

The development of intense pulsed ion beams has been stimulated to a large extend by their potential application as a driver for inertial confinement fusion

(ICF) [55, 56]. Research in this field intensified at the end 1970s when it had become obvious that intense pulsed electron beams were an inappropriate driver for ICF mainly because of their long range in matter and because the bremsstrahlung created during interaction with the target material preheats the fuel inside the fusion pellet and thus prevents compression to high densities. At the same time it was found that electron beam pinch diodes which had been developed for the production of intense electron beams were also capable of delivering intense ion beams with efficiencies around 50% [57]. At the Cornell University Humphries and Sudan showed both experimentally and theoretically that the ion beam production efficiency of vacuum diodes could be further increased by insulating the diode accelerating gap with magnetic fields supplied by external coils [58, 59]. These findings were the starting point for more than 20 years of intensive research in collective ion beam physics.

5.4.2 Introduction

Intense light ion beams are promising tools in matter research [60], inertial fusion energy (IFE) [61], and radiation (Boron Neutron Capture) therapy [62]. They can provide a technically simple, cost effective path to deliver high specific power densities (several 100 TW/g) to condensed matter. The key element of this approach is an accelerator consisting of two stages, the injector-diode and the post-accelerator diode, both operating at power densities around several 10 GW/cm^2 and at electric fields around several MV/cm, both values being two orders of magnitude higher than in conventional accelerators. In this concept the accelerator is empowered from a repetitive multi-Terawatt, multi-Megavolt pulse generator. The high energy linear induction accelerator (HELIA) approach [63] is considered to be the most attractive concept [64] for this generator.

The most extreme requirements result from driver development for IFE: Ultimately each light ion driver module should be able to deliver an ion beam with atomic masses greater than 4, particle energies greater than 30 MeV, and pulse energies of more than 500 kJ at power levels greater than 10 TW. Consequently, a beam with 1000 cm^2 cross-section must be extracted from the accelerator if the power density in the accelerating gap is limited to less than 10 GW/cm^2. To focus the beam to power densities of more than 10 TW/cm^2, a spatial convergence of more than 1000 is required. In addition, depending on the beam transport scheme, only a small divergence is tolerable.

Here we shall discuss the recent progress that has been achieved in our understanding of diode accelerator gaps, in the development of suitable ion sources, and in beam transport.

5.4.3 Accelerator Physics

To a certain extent it seems possible to break-up the light-ion accelerator development into three parts: the ion source, the injector gap and the post-accelerator gap. Although there is an interaction between the gap physics and the ion and electron source plasmas, much of the accelerator gap physics is determined by the high power densities in the gap.

Injector Gap

So far most investigations have concentrated on the injector diode gap, and here we shall summarize the results [65–71]. The injector consists of a massive anode and a hollow cathode separated by a few mm. A high power multi-megavolt pulse is applied to the diode gap for a duration of up to several 10 ns. Prior to the application of the voltage pulse, an ion source must have been prepared on the anode surface. In most concepts this is a pulsed plasma layer simultaneously produced with the accelerating pulse. Due to the large electric fields in the diode gap, electron field emission with plasma formation occurs instantaneously also on the cathode. Therefore, a strong magnetic field perpendicular to the electric field is necessary to prevent the electrons from reaching the anode and consuming most of the driver pulse energy.

Many aspects of the accelerating gap physics can be studied independently from the specific ion source if this source is homogenous and does not expand too fast across the magnetic field lines. Much of the gap physics is, however, dominated by electrons streaming into the diode along the magnetic field lines and forming an electron sheath and a virtual cathode. Initially this electron sheath is quite thin and drifts in the $E \times B$ direction. A fluid instability, known as the "diocotron" [72], spreads the electrons across the diode gap within a few nanoseconds. If the sheath comes close to the anode, the high frequency diocotron instability, whose effect on beam divergence is small, is replaced by the lower frequency ion mode instability whose frequency is roughly equal to the inverse of the ion transit time in the gap [73] and therefore has a drastic effect on beam divergence.

The occurrence of instabilities and the spreading of the electron sheath over the entire gap has a strong impact on diode operation. Since ion current density in high power diodes in general is space charge limited, space charge compensation coming along with the distributed electron sheath leads to ion current enhancement much above the Child–Langmuir value. This in turn creates a stronger diamagnetic effect of the electrons drifting in the $E \times B$ direction and results in a movement of the virtual cathode towards the anode. Since the magnetic flux is conserved, the rising magnetic pressure will finally stop it, and a new equilibrium will be achieved [74].

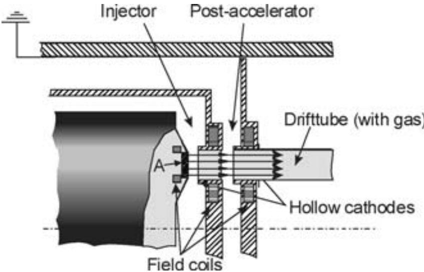


Fig. 5.15. Scheme of a two-stage ion diode. The insulating magnetic field in the accelerating gaps is established by 6 field coils supplied by capacitor banks. The beam transport sections

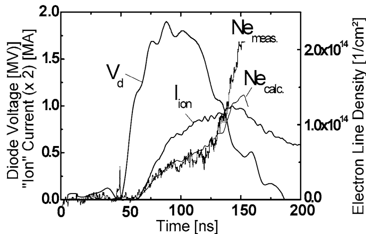


Fig. 5.16. Diode voltage V_d, ion current I_ion, and electron density Ne at a position of 3 mm from the anode surface. Also shown is the electron density Ne_calc derived from V_d and I_ion using an analytic model for the saturated (constant density) electron distribution

However, if the increasing diamagnetic field can penetrate into the electrode plasmas, the strength of magnetic insulation in the diode gap is reduced and a further degradation of the diode impedance occurs. All of these conclusions have first been drawn from analytic modelling and three-dimensional fully relativistic, electromagnetic, particle-in-cell code simulations [74, 75] for rather simplified geometries. To check these predictions experimentally, new sophisticated diagnostics had to be developed [60, 76–78]. Measurements with these tools clearly confirmed the theoretical findings. Using a very sensitive and robust dispersion interferometer, even the electron density distribution in the diode gap could be measured with high space- and time-resolution (0.5 mm, 1 ns) [79].

Figure 5.16 demonstrates the excellent agreement between the measured electron densities and those predicted by an analytic diode model [69].

By these investigations, a rather consistent understanding of the accelerating gap physics has been achieved which confirms the description of diode operation outlined above.

Post-accelerator Gap

The present experience with magnetically-insulated single-stage ion diodes suggests that the operational stability is necessary, and the required low beam divergence can only be achieved with multi-stage acceleration. Also, the goal to accelerate light ions with 30 MV can only be reached in a multi-stage configuration, since very large insulating magnetic fields would be required for single-stage acceleration. However, due to the rapidly increasing complexity, it seems unlikely that more than two stages can be realized. Because of the strong beam self-fields space charge neutralisation is necessary between stages. For that purpose the inter-stage drift section must be filled with a gas or plasma and separated from the high vacuum required in the accelerating gaps.

The expectation for reduced beam divergence is based on the observation that the frequency of the ion mode depends on the ion transit time which is different for each gap. Thus the effects of instability do not add coherently from stage to stage, and the ion divergence should decrease. Simulations with the PIC-code Quicksilver

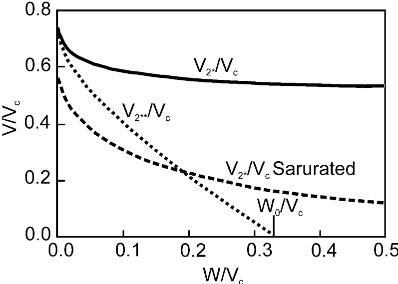


Fig. 5.17. Critical voltages V_{2*} and V_{2**} in the post-acceleration gap as a function of the injected ion energy W, normalized to the critical voltage V_c for magnetic insulation (from [80])

have confirmed that for sufficiently low beam current densities, ion divergence can be reduced significantly by post-acceleration [80].

Only very few preliminary two-stage experiments have been carried out [81–83], most of them at very low power and low voltage. Nevertheless, these experiments showed a strong reduction of beam divergence.

Some of the results that were obtained for the injector gap can certainly be extended to the post-accelerating gap. Since the electric fields in this gap will be of similar magnitude, a virtual cathode will be formed too, and it is very likely that the post-accelerating gap will also be filled entirely with electron space charge. Nevertheless, there are certain differences between the post-accelerator and the injector that need additional considerations. It has been found [84] that, like in the single-stage case, a limiting voltage V_{2*} exists – depending on the injected ion energy W and the strength of magnetic insulation B – above which unlimited current density can penetrate the stage without the formation of a virtual anode. A virtual anode can be useful, however, to suppress the acceleration of ions from plasmas that form at the injection side of the stage. It is well known from the fixed gap monopolar problem of beam injection that above a maximum current density $j_2(V_2)$, penetrating the gap, instabilities occur that prevent further transfer of current through the gap [85]. Such a limiting current also exists for the bipolar case with movable virtual cathodes. However, this limitation does only appear below a second limiting voltage V_{2**} which is unique to the multi-stage theory and depends on W and B. If the electron sheath fills the entire post-acceleration gap, then $V_{2**} = 0$. As seen from Fig. 5.17, if the suppression of unwanted ions from the injection side of the second stage would be required, saturated behaviour would limit the final beam energy to only slightly higher than the critical insulating voltage V_c determined by the magnetic insulation strength. On the other hand, if a virtual anode cannot be established, we must tolerate a certain fraction of unwanted ions extracted from the anode side of the post-accelerator. Fortunately, steady state calculations of the ion current extracted from the injection side of the second stage show that a small fraction leads already to a significant increase of V_{2*} [86].

From this discussion of the peculiarities of post-acceleration it is obvious that detailed experimental investigations of the second stage are necessary. Especially

time and space resolved electric and magnetic field measurements are needed to identify the mode of operation and to probe the existence of a virtual anode. Of course, of greatest interest is the degree of divergence reduction and the operational stability that can be attained.

Ion Source

The lack of an adequate pre-formed ion source has greatly hampered progress in the development of a light ion beam driver during the last decade. Although the pre-formed proton source used in experiments on the KALIF-facility is acceptable for studying the physical phenomena of accelerating gaps, it is inappropriate for accelerating voltages of 30 MV. The pitfalls of ion source development can be attributed to the following circumstances:

- The ion source must homogeneously emit a current density around 1 kA/cm^2 for a duration of 50 ns, i.e. supply $\approx 10^{15}$ ions/cm^2 over an area of several hundred cm^2 on the anode surface. 10^{15} ions/cm^2 are equivalent to a few atomic monolayers.
- The source should smoothly conform to the desired anode shape. This may need a conducting surface in the vicinity if the source is a plasma.
- It must not expand into the accelerating gap by more than a small fraction of the dynamic gap (5–10 mm).
- It must be prepared to emit when the voltage pulse appears in the diode.
- It should emit a single ion species.
- It should scale to repetition rates of a few Hz and to lifetimes of 10^8 pulses. For the intermediate terms of development, a single-shot source meeting the above requirements would be adequate to assess the properties of diode operation and beam transport.

All electrode surfaces, including the active source area, will be covered with a few monolayers of contaminants [87, 88]. The main adsorbates are H_2, H_2O, CO, CO_2, and various compositions of C_nH_m with binding energies on relevant electrode surfaces in the range of 0.4–1.8 eV [87]. In addition, more tightly bound oxides (1.7–3.5 eV) may exist on the surface. Various cleaning techniques (heating, RF glow-discharge) have been applied to remove the contaminants. However, at the typical base pressure conditions of 10^{-3}–10^{-2} Pa in large scale pulsed power systems the recontamination rate can be as high as a few monolayers per second. Cleaning can therefore only be successful if the base pressure in the vacuum chamber is reduced below 10^{-3} Pa.

If adsorbates will be present on electrode surfaces in the vicinity of a high power diode, they will eventually be released due to electrons diffusing to the anode side or due to stray ions hitting the cathode surface [89]. It has been pointed out that in the environment of a high power diode, the thermal desorption becomes much more important than the stimulated desorption [88].

Two types of ion sources must be distinguished at this point: field emission and space charge limited emission sources. Field emission threshold sources, like the

LiF-source investigated at Sandia [88], are inherently more susceptible to the release of gaseous adsorbates from the surface. If the desorbed gas layers develop into a plasma, a zero work function emitter begins to compete with the field emitter and rapidly predominates the beam composition. This behaviour has been observed on PBFA II and SABRE and was termed the parasitic load problem. While initially Li-ions are the most abundant species, later in the pulse C, O, and H prevail.

In case of a space charge limited active (plasma) source, like the TiH-source used on KALIF [90], surface contaminants are mixed into the source plasma at its creation. Since hydrogen has the largest mobility, it tends to dominate at the plasma edge. Therefore, very little C, O, Ti, etc. was observed in the extracted beam.

5.4.4 Beam Transport

Efficient transport of intense ion beams (light and heavy) is necessary for ion-driven inertial fusion energy. It has been stated earlier that the most preferable transport mode would be self-pinched transport (SPT). To realize SPT, sufficiently large beam net currents are necessary to confine the beam by its self magnetic field. For a long time ion beam SPT was considered to be impossible because of the rapid ion beam induced gas ionisation that would prevent the establishing of sufficiently high net currents. However, recent numerical simulations [91–93] and a pioneering experiment at the Naval Research Lab. (NRL) [94] suggest that ion beam SPT should be possible in a certain window of gas pressures.

References

[1] R. Bird, J. Williams, *Ion Beams for Materials Analysis* (Academic, New York, 2003)
[2] CEA, Sunnyvale/CAL. http://www.cea.com/tech.htm#rbs1 (2003)
[3] Charles Evans & Associates, Sunnyvale/CA. http://www.cea.com/tech.htm
[4] Harvard University, Materials Res. Ctr., Cambridge/MA. http://www.mrsec.harvard.edu/default.html (2003)
[5] Proton microprobe, University of Oxford, Oxford, OX1 2JD, UK
[6] CSIRO, Canberra/Australia. http://www.nmp.csiro.au/
[7] Harvard University, Cambridge/MA. http://www.mrsec.harvard.edu/cams/cams.html
[8] University of Oxford, Laboratory of Molecular Biophysics, Oxford/UK. http://biop.ox.ac.uk/www/top.html
[9] High Voltage Engineering B.V., Amersfoort/The Netherlands. http://www.highvolteng.com/ (2003)
[10] National Electrostatic Corp., Middleton/WI. http://www.pelletron.com/index.html (2003)
[11] A. Benninghoven, F.G. Rüdenauer, H.W. Werner, *Secondary Ion Mass Spectrometry* (Wiley, New York, 1987)
[12] CAMECA S.A., Courbevoie/France. http://www.cameca.fr/html/simstechnique.html (2003)
[13] H.A. Storms, K.F. Brown, J.D. Stein, Anal. Chem. **49**, 2023 (1977)
[14] R. Levi-Setti, Y.L. Wang, B. Crow, J. de Phys. **45**(C9), 197 (1984)

[15] F.G. Rüdenauer, W. Steiger, Concepts, applications and limits of 3D SIMS, in *Proc. 11th Int. Congr. on X-Ray Optics and Microanalysis*, ed. by J.D., Brown, R.H., Packwood, Univ. W. Ontario, London, 1987, pp. 210–217
[16] CAMECA S.A., Courbevoie/France. http://www.cameca.fr/index.html (2003)
[17] Physical Electronics Corp., Eden Prairie/MI. http://www.phi.com/genf.asp?ID=296 (2003)
[18] ION TOF GmbH, Münster/Germany. http://www.ion-tof.com/index.html (2003)
[19] Physical Electronics, Eden Prairie/MN. http://www.phi.com/ (2003)
[20] Surface Science Western, London/Ontario/Canada. http://www.uwo.ca/ssw/services/tofsimsfiles/tofmonel.html (2003)
[21] FEI Company, Hillsboro, OR (2003)
[22] C. Tuniz, J.R. Bird, D. Fink, G.F. Herzog, *Accelerator Mass Spectrometry: Ultrasensitive Analysis for Global Science* (CRC, Boca Raton, 1998)
[23] G. Tamborini, D.L. Donohue, F.G. Rüdenauer, M. Betti, Evaluation of practical sensitivity and useful ion yield for uranium detection by secondary ion mass spectrometry, J. Anal. At. Spectrom. **19**, 1–7 (2004)
[24] F.G. Rüdenauer, SIMS in space, in *Secondary Ion Mass Spectrometry SIMS IX*, ed. by A. Benninghoven, Y. Nihei, R. Shimizu (Wiley, Chichester 1995), pp. 901–905
[25] E. Zinner, Interstellar cloud material in meteorites, in *Meteorites and the Early Solar System*, ed. by J.F. Kerridge, M.S. Matthews (University of Arizona Press, Tucson, 1988), pp. 956–983
[26] J.W. Rabalais, *Principles and Applications of Ion Scattering Spectrometry* (Wiley, New York, 2002)
[27] B.B. Calipso, TU Eindhoven, Eindhoven, NL. http://www.calipso.nl/ (2003)
[28] J. Orloff, L.W. Swanson, Study of a field-ionization source for microprobe application, J. Vac. Sci. Tech. **12**(6) (1975)
[29] J. Melngailis, J. Vac. Sci. Technol. B **5**, 469 (1987)
[30] J. Orloff, M. Utlaut, L. Swanson, *High Resolution Focused Ion Beams: FIB and its Applications* (Kluwer Academic/Plenum, New York, 2003)
[31] L.R. Harriott, J. Vac. Sci. Technol. B **11**, 2012 (1993)
[32] H.W.P. Koops, R. Weiel, D.P. Kern, T.H. Baum, J. Vac. Sci. Technol. B **6**(1), 477 (1988)
[33] K. Edinger, T. Kraus, J. Vac. Sci. Technol. B **18**, 3190 (2000)
[34] T. Hoshino, M. Kawamori, T. Suzuki, S. Matsui, K. Mabuchi, J. Vac. Sci. Technol. B **22**, 3158 (2004)
[35] S. Matsui, et al., J. Vac. Sci. Technol. B **21**, 2732–2736 (2003)
[36] S. Matsui, et al., J. Vac. Sci. Technol. B **18**, 3181–3184 (2000)
[37] L. Niewoehner, H.W.J. Wenz, Applications of focused ion beam systems in gunshot residue investigation, J. Forensic Sci. **44**(1), 105–109 (1999)
[38] W. Shockley, Forming semiconductor devices by ionic bombardment, U.S. Patent 2,787,564 (1957)
[39] J. Lindhard, M. Scharff, H.E. Schiøtt, Range concepts and heavy ion ranges, Mat.-Fys. Medd. Dan. Vid. Selsk. **33**(14) (1963)
[40] M.T. Robinson, O.S. Oen, The channelling of energetic atoms in crystal lattices, Appl. Phys. Lett. **2**, 30–32 (1963)
[41] R.B. Simonton, W. Class, Y. Erokhin, M. Mack, L. Rubin, Ion implantation, in *Handbook of Semiconductor Manufacturing Technology*, ed. by Y. Nishi, R. Doering (Marcel Dekker, New York, Basel, 2000), pp. 105–147
[42] J.F. Ziegler, J.P. Biersack, U. Littmark, *The Stopping and Range of Ions in Solids* (Pergamon, New York, 1985)

[43] G. Hobler, Critical angles and low-energy limits to ion channeling in silicon, Radiat. Eff. Def. Sol. **139**, 21–85 (1996)
[44] K.M. Klein, C. Park, S.-H. Yang, A.F. Tasch, Accurate and efficient two-dimensional modeling of boron implantation into single-crystal silicon, in *IEDM Techn. Dig. IEEE* (1991), pp. 697–700
[45] J.F. Ziegler, http://www.srim.org
[46] G. Hobler, Monte Carlo simulation of two-dimensional dopant distributions at mask edges, Nucl. Instrum. Meth. B **96**, 155–162 (1995)
[47] M. Posselt, B. Schmidt, T. Feudel, N. Strecker, Atomistic simulation of ion implantation and its application in Si technology, Mater. Sci. Eng. B **71**, 128–136 (2000)
[48] G.K. Celler, S. Cristoloveanu, Frontiers of silicon-on-insulator, J. Appl. Phys. **93**, 4955–4978 (2003)
[49] International Technology Roadmap for Semiconductors. http://public.itrs.net/ (2003)
[50] D.W. Duff, L.M. Rubin, Ion implant equipment challenges for 0.18 μm and beyond, Sol. State Technol. (June issue) (1998)
[51] M. Graf, T. Parrill, Low-energy ion implantation meets productivity challenges, Semicond. Intl. (April issue) (2004)
[52] L. Rubin, Angle control in high-current ion implanters, Sol. State Technol. (Oct. issue) (2002)
[53] L. Rubin, W. Morris, High-energy implanters and applications take off, Semicond. Intl. (April issue) (1997)
[54] A. Anders (ed.), *Handbook of Plasma Immersion Ion Implantation and Deposition* (Wiley, New York, 2000)
[55] Papers in: *International Topical Conference on Electron Beam Research and Technology*, ed. by G. Yonas, Albuquerque, N.M., SAND76-5122 (1975)
[56] Papers in: *2nd International Topical Conference on High Power Electron and Ion Beam Research and Technology*, ed. by J.A. Nation, R.N. Sudan, Cornell University, Ithaca, N.Y. (1977)
[57] S.A. Goldstein, R. Lee, The ion induced pinch and enhancement of ion current by pinched electron and ion flow in relativistic diodes, Phys. Rev. Lett. **35**, 1079–1082 (1975)
[58] S. Humpries, Intense pulsed ion beams for fusion applications, Nucl. Fusion **20**, 1549–1612 (1980)
[59] R.N. Sudan, R.V. Lovelace, Generation of intense ion beams in pulsed diodes, Phys. Rev. Lett. **31**, 1174–1177 (1973)
[60] G.I. Kanel, K. Baumung, H. Bluhm, V.E. Fortov, Possible applications of the ion beams technique for investigations in the field of equation of state, Nucl. Instrum. Methods A **415**, 509–516 (1988)
[61] J.P. VanDevender, H. Bluhm, Light ion accelerators for ICF, in *Nuclear Fusion by Inertial Confinement*, ed. by G. Velarde, Y. Ronen, J.M. Martinez-Val (CRC, Boca Raton, Ann Arbor, London, Tokyo, 1993)
[62] H. Bluhm, P. Hoppé, Perspectives of high power pulse generators for the boron neutron capture therapy (BNCT) of tumors, in *12th IEEE Int. Pulsed Power Conf.*, ed. by C. Stallings, H. Kirbie, Monterey, USA, 1999, pp. 502–505
[63] J.J. Ramirez, The four stage HELIA experiment, in *5th IEEE Pulsed Power Conf.*, ed. by. P.J. Turchi, M.F. Rose, Arlington, Virginia, USA, pp. 143–146
[64] G.A. Moses, G.L. Kulcinski, D. Bruggink, R. Engelstad, E. Lovell, J. McFarlane, Z. Musicki, R. Petersen, I. Stanislavsky, L. Wittenberg, G. Kessler, U. Von Möllendorff, E. Stein, I. Smith, P. Corcoran, H. Nichimoto, J. Fockler, D. Cook, R. Olson, Overview of the LIBRA light ion beam fusion conceptual design, Fusion Technol. **15**, 756–765 (1989)

[65] J.E. Bailey, A.B. Filuk, A.L. Carson, D.J. Johnson, P. Lake, E.J. McGuire, T.A. Mehlhorn, T.D. Pointon, T.J. Renk, W.A. Stygar, Measurements of acceleration gap dynamics in a 20-TW applied-magnetic-field ion diode, Phys. Rev. Lett. **74**, 1771–1774 (1995)
[66] T.A. Mehlhorn, J.E. Bailey, G.A. Chandler, R.S. Coats, M.E. Cuneo, M.S. Derzon, M.P. Desjarlais, R.J. Dukart, A.B. Filuk, T.A. Haill, H.C. Ives, D.J. Johnson, R.J. Leeper, T.R. Lockner, C.W. Mendel, P.R. Menge, L.P. Mix, A.R. Moats, W.B. Moore, T.D. Pointon, J.W. Poukey, J.P. Quintenz, S.E. Rosenthal, D. Rovang, C.L. Ruiz, S.A. Slutz, W.A. Stygar, D.F. Wenger, Progress in lithium beam power, divergence, and intensity at Sandia National Laboratories, in *10th Int. Conf. on High Power Particle Beams*, San Diego, USA, ed. by W. Rix, R. White (NTIS, Springfield, 1994), pp. 53–60
[67] D.L. Hanson, M.E. Cuneo, P.F. McKay, J.E. Maenchen, R.S. Coats, J.W. Poukey, S.E. Rosenthal, W.E. Fowler, D.F. Wenger, M.A. Bernard, J.R. Chavez, W.F. Stearns, Operation of a high impedance applied-B extraction ion diode on the sabre positive polarity linear induction accelerator, in *9th Int. Conf. High Power Particle Beams*, Washington DC, USA, ed. by D. Mosher, G. Cooperstein (NTIS, Springfield, 1994), pp. 781–787
[68] J.B. Greenly, R.K. Appartaim, J.C. Olson, L. Brisette, Extraction ion diode studies for optimised performance: divergence, ion species and parasitic load, in *10th Int. Conf. on High Power Particle Beams*, San Diego, USA, ed. by W. Rix, R. White (NTIS, Springfield, 1994), pp. 398–401
[69] J.B. Greenly, R.K. Appartaim, J.C. Olson, Anode plasma dynamics in an extraction applied-B ion diode: effects on divergence, ion species and parasitic load, in *10th Int. Conf. High Power Particle Beams*, Prague, Czech Republic, ed. by K. Jungwirth, J. Ullschmied (Tiskárna "K" Ltd., Prague, 1996), pp. 111–114
[70] H. Bluhm, P. Hoppé, H. Laqua, D. Rusch, Production and investigation of TW Proton beams from a annular diode using strong radial magnetic insulation fields and a preformed anode plasma source, Proc. IEEE **80**, 995–1009 (1992)
[71] V. Licht, H. Bluhm, P. Hoppé, S.J. Yoo, Time dependent field and particle density measurements in the acceleration gap of a high power ion diode, in *12th Int. Conf. High Power Particle Beams*, Haifa, Israel, 1998, ed. by M. Markovits, J. Shiloh, pp. 203–207
[72] O. Buneman, R.H. Levy, L.M. Linson, Stability of crossed-field electron beams, J. Appl. Phys. **37**, 3203–3222 (1966)
[73] R.W. Lemke, S.A. Slutz, A full mode set electromagnetical stability analysis of magnetically insulated ion diodes, Phys. Plasmas **2**, 549 (1995)
[74] M.P. Desjarlais, Theory of applied-B ion diodes, Phys. Fluids B **1**, 1709–1720 (1989)
[75] J.P. Quintenz, D.B. Seidel, M.L. Kiefer, T.D. Pointon, R.S. Coats, S.E. Rosenthal, T.A. Mehlhorn, M.P. Desjarlais, N.A. Krall, Simulation codes for light-ion diode modeling, Laser Part. Beams **12**, 283–324 (1994)
[76] H. Bluhm, P. Hoppé, H. Bachmann, W. Bauer, K. Baumung, L. Buth, H. Laqua, A. Ludmirski, D. Rusch, O. Stoltz, S. Yoo, Stability and operating characteristics of the applied B proton extraction diode on KALIF, in *10th Int. Conf. High Power Particle Beams*, San Diego, USA, ed. by W. Rix, R. White (NTIS, Springfield, 1994), pp. 77–82
[77] Y. Maron, E. Sarid, E. Nahshoni, O. Zahavi, Time-dependent spectroscopic observation of the magnetic field in a high-power-diode plasma, Phys. Rev. A **39**, 5856–5862 (1989)
[78] S. Yoo, Spektroskopische Messungen im Anodenplasma einer fremdmagnetisch isolierten Hochleistungs-Ionendiode, Ph.D. thesis, University Karlsruhe, FZKA 5976, 1997
[79] V. Licht, H. Bluhm, A sensitive dispersion interferometer with high temporal resolution for electron density measurements, Rev. Sci. Instrum. **71**, 2710–2015 (2000)

[80] S.A. Slutz, J.W. Poukey, T.D. Pointon, Simulations of magnetically insulated multistage ion diodes, Phys. Plasmas **1**(6), 2072–2081 (1994)

[81] T. Lockner, S. Slutz, D.J. Johnson, M. Desjarlais, J. Poukey, Results of the first 2-stage diode experiments on PBFAII, in *10th Intl. Conf. on High Power Particle Beams*, San Diego, USA, ed. by W. Rix, R. White (NTIS, Springfield, 1994), pp. 61–64

[82] S. Miyamoto, K. Yasuike, K. Imasaki, C. Yamanaka, S. Nakai, Intense light ion beam divergence in single and two-stage diode, in *15th Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Darmstadt, Germany, ed. by D. König (VDE-Verlag, Berlin, Offenbach, 1992), pp. 711–715

[83] T.R. Lockner, S. Slutz, J.W. Poukey, W.A. Stygar, Theoretical and experimental studies of the 2-stage ion diode, in *9th IEEE Int. Pulsed Power Conf.*, Albuquerque, USA, 1993, ed. by K. Prestwich, W. Baker, pp. 714–717

[84] S.A. Slutz, M.P. Desjarlais, Theory of multistage intense ion beam acceleration, J. Appl. Phys. **67**(11), 6705–6717 (1990)

[85] C.K. Birdsall, W.B. Bridges, Space-charge instabilities in electron diodes and plasma converters, J. Appl. Phys. **32**(12), 2611 (1961)

[86] S.A. Slutz, Ion emission from anode foils during multistage acceleration of intense ion beams, Phys. Fluids B **5**(1), 209–215 (1993)

[87] M.E. Cuneo, The effect of electrode contamination, cleaning and conditioning on high-energy pulsed-power device performance, IEEE Trans. Dielectr. Electr. Insulat. **6**(4), 469–485 (1999)

[88] M.E. Cuneo, P.R. Menge, D.L. Hanson, W.E. Fowler, M.E. Bernard, G.R. Ziska, A.B. Filuk, T.D. Pointon, R.A. Vesey, D.R. Welch, J.E. Bailey, M.P. Desjarlais, T.R. Lockner, T.A. Mehlhorn, S.A. Slutz, M.A. Stark, Results of vacuum cleaning techniques on the performance of LiF field-threshold ion sources on extraction applied-B ion diodes at 1-10 TW, IEEE Trans. Plasma Sci. **25**(2), 229–251 (1997)

[89] R.A. Vesey, T.D. Pointon, M.E. Cuneo, T.A. Mehlhorn, J.E. Bailey, D.J. Johnson, W.A. Stygar, Electron-anode interactions in particle-in-cell simulations of applied-B ion diodes, Phys. Plasmas **6**(8), 3369–3387 (1999)

[90] H. Laqua, H. Bluhm, L. Buth, P. Hoppé, Properties of the nonquilibrium plasma from a pulsed sliding discharge in a hydrogen gas layer desorbed from a metal hydride film, J. Appl. Phys. **77**(11), 5545–5552 (1995)

[91] D.R. Welch, M.E. Cuneo, C.L. Olson, T.A. Mehlhorn, Gas breakdown effects in the generation and transport of light ion beams for fusion, Phys. Plasmas **3**, 2113–2121 (1996)

[92] D.R. Welch, C.L. Olson, Self-pinched transport for ion-driven inertial confinement fusion, Fusion Eng. Des. **32-33**, 477–483 (1996)

[93] D.V. Rose, P.F. Ottinger, D.R. Welch, B.V. Oliver, C.L. Olson, Numerical simulations of self-pinched transport of intense ion beams in low-pressure gases, Phys. Plasmas **6**(10), 4094–4103 (1999)

[94] P.F. Ottinger, F.C. Young, S.J. Stephanakis, D.V. Rose, J.M. Neri, B.V. Weber, M.C. Myers, D. Hinshelwood, D. Mosher, C.L. Olson, D.R. Welch, Self-pinched transport of an intense proton beam, Phys. Plasmas **7**(1), 346–358 (2000)

6

Ion Propulsion Systems

H. Bassner, R. Killinger, J. Mitterauer, F. Rüdenauer, N. Koch, and G. Kornfeld

6.1 Electrostatic Ion Thrusters and RF-Ion Thruster RIT[1]

6.1.1 Introduction

Electrostatic ion thrusters in general produce a plasma in the gaseous propellant enclosed in a discharge vessel by collision of electrons and neutral atoms. The energy of electrons is adjusted to produce mainly single charged ions. A part of the ions is extracted from the discharge chamber and accelerated to the beam velocity by a grid system at the end of the discharge vessel which consists of 2 or more grids, each grid on a different voltage potential. The grids contain a great number of holes, which form extraction channels. Ion beamlets are generated in each channel which result in the total ion beam.

The propellant during the first steps of the development of these thrusters was mercury, which was vaporised for ionisation. Due to erosion and degeneration problems with different spacecraft materials currently the noble gas xenon is used, but other gasses are also possible.

There exist two different methods for the ionisation of the propellant in the discharge vessel:

- *Electron Bombardment Ionisation Principle*. Electrons are generated in the discharge vessel by a hollow cathode close to the gas inlet. A part of the propellant flows through the cathode and allows the generation of a high electron current. An isolated anode on a defined potential inside the discharge vessel close to the grids forces the electrons to fly through the propellant and generate a plasma. Magnets outside the discharge vessel form a magnetic field which increase the length of the electron trajectories to the anode to a spiralic shape and increase the ionisation efficiency.
  Thrusters using this ionisation principle are developed, manufactured and used in space by USA, Japan and Europe [1].

[1] Please refer to H. Bassner, R. Killinger.

- *Radiofrequency Ionisation Principle*. A plasma is generated inside a discharge chamber by electrons, which are accelerated in an RF-field. The discharge chamber is surrounded by an RF-coil, which is connected to an RF-generator. To start the discharge, electrons from the neutraliser tip are drawn into the discharge chamber for a few seconds. As the discharge is ignited, it is self-sustaining.
  Thrusters using this ionisation principle were invented by Prof. Horst Loeb at the University of Giessen and industrially developed by EADS in Germany. A survey of the activities in Giessen and at EADS will be given below.

6.1.2 Historical Development of RIT

Radiofrequency ion thrusters (RIT), invented by Prof. Loeb at the University of Giessen in Germany, are under development at EADS Space Transportation (previously MBB, Dasa and Astrium).

Based on the work which was performed at the University of Giessen since 1962, the space section of Messerschmitt–Bölkow–Blohm (MBB) started to develop ion thrusters in 1970 in close cooperation with Giessen for the application in space.

RIT thrusters have been built and tested at discharge chamber diameters between 4 and 35 cm for thrust levels of 0.1 to 250 mN. The RIT thrusters can be operated under stable conditions in a wide range of mass flow (50 to 95% of mass efficiency) and beam voltage (800 to 2000 V).

At a constant thrust level, the operational conditions can be optimised for low input power (= low specific impulse) or for high specific impulse (= high input power). Specific impulses – defined as thrust divided by mass flow – between 5000 and 30 000 Ns/kg can be achieved at a specific power input between 24 to 40 W/mN. The highest specific impulse already demonstrated is 65 000 Ns/kg with growth margin [2].

6.1.3 Electrophysical Fundamentals

In a RIT, like in most ion thrusters, the noble gas xenon is used as propellant. The xenon atoms are ionised by collision with electrons, which are accelerated by an RF-field, inside a discharge vessel. The RF-field is connected to the inner of the electrically non-conductive discharge vessel by a coil. Figure 6.1 shows the operating principle of a RIT.

The xenon ions are transported by a pressure gradient to the acceleration grid system at the thruster's exit. In the grid system the ions are accelerated by an electrostatic field which is generated by different voltages applied to a 2- or 3-grid system.

In order to avoid a charging of the spacecraft by the positive ion current, an equal current of electrons has to be produced. These electrons are generated by a hollow-cathode neutraliser and, in addition, are necessary to ignite the RF-discharge at the beginning of each operation cycle.

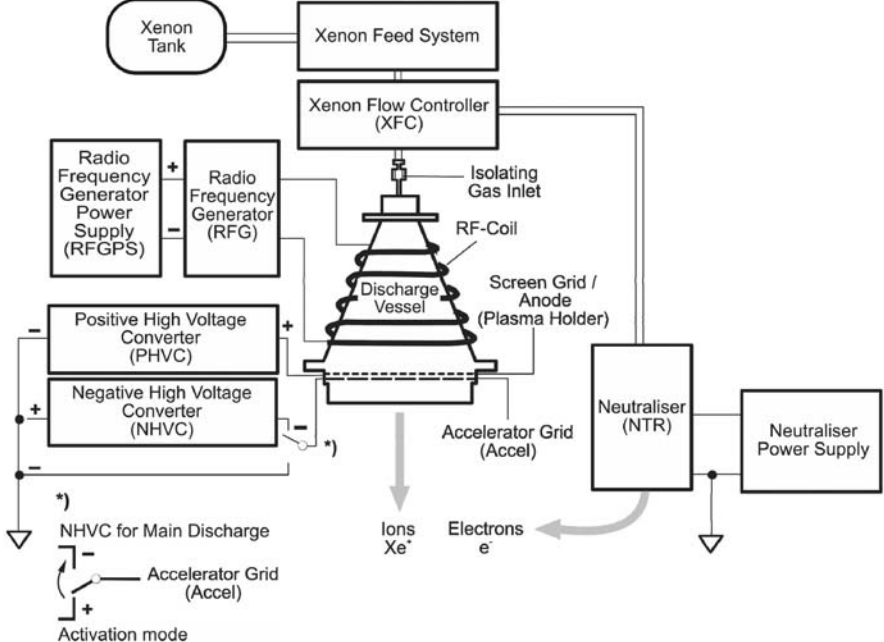


Fig. 6.1. Operating principle of an RF-ion thruster RIT

6.1.4 Present State-of-the-Art and Applications of RIT

Ion sources, using an RF-field for propellant ionisation and an electrostatic field for ion acceleration, can be used for different applications on ground and in space:

On ground fast ions are used for different applications [3]:

- surface treatment like polishing of equipment with special requirements;
- generation of thin layers of metal on surfaces by sputtering;
- implantation of ions in material surfaces;
- general material processing;
- fusion plasma heating.

For the applications on ground it is sufficient that the ion source is developed for small to medium lifetime only, compared to space applications. However, the change of life limiting components must be easy to perform. The equipment for operation of the ion source can be standard electronic and mechanical equipment with limited requirements on efficiency, lifetime, resistance against environmental conditions, mass and size.

RIT ion sources for ground applications have been mainly developed by the 1. Physical Institute of the University of Giessen in co-operation with special manufacturers for ground equipment.

A wide field of application of ion sources can be found in space, where the accelerated ions are used for the production of thrust with very high exhaust velocities.

In spite of the rather low thrust level, the thrusters can be used for nearly all tasks of orbit propulsion, like:

- attitude- and orbit control of satellites;
- orbit transfer of satellites (change of inclination angle, orbit raising, removal of satellites from their operational orbit at the end of their lifetime);
- air drag compensation of satellites flying at low attitude;
- main propulsion of interplanetary spacecraft.

The very high exhaust velocity of the ions from the ion source results in a specific impulse which is more than 10 times higher than that of chemical thrusters. This reduces the propellant mass necessary for a special mission by the same amount. The advantage is reduced by the mass of the complete electric propulsion system. Therefore, the advantages of ion propulsion pay off best for missions with high total impulse requirements, e.g. long operation time.

In the case of interplanetary spacecraft propulsion, in addition a reduction of mission time can be achieved, because time consuming fly-by maneuvres can be avoided. Moreover, this allows mission planners to be much more flexible to select launch dates. The development effort for these electric propulsion systems is rather high, because special care must be taken on the efficiency, reliability and lifetime of the system.

6.1.5 Applications of RIT Thrusters

RIT 10 Testflight on EURECA

After 20 years of detailed development of the thruster and the equipment necessary for thruster operation, a possibility for a test was offered by ESA as an experiment on the EUropean REtrievable CArrier (EURECA). As one of the 16 experiments on this platform, RITA (RIT Assembly) should be operated in an orbit of about 500 km height during one year mission time. The challenge of this project was the very limited budget and the necessity of a fully automatic operation of the experiment due to the limited ground contact time [4].

The complete hardware of this experiment except the neutraliser (developed and built by Thales, previously AEG in Ulm) has been developed and manufactured at EADS-ST in Ottobrunn. Figure 6.2 shows the complete experiment equipment mounted on one equipment support panel of the EURECA platform, which was launched in July 1992.

According to the power level available, the thruster has been operated in space at thrust levels between 5 and 10 mN for about 240 hours. The experience showed that the RITA system operated at performance values similar to those obtained during tests on ground. In addition, no effects on the satellite's electronic system or on the telecommunication were discovered.

The retrieval of the EURECA platform with all the experiments to earth by the space shuttle in July 1993 gave us the possibility to investigate the experiment in detail.

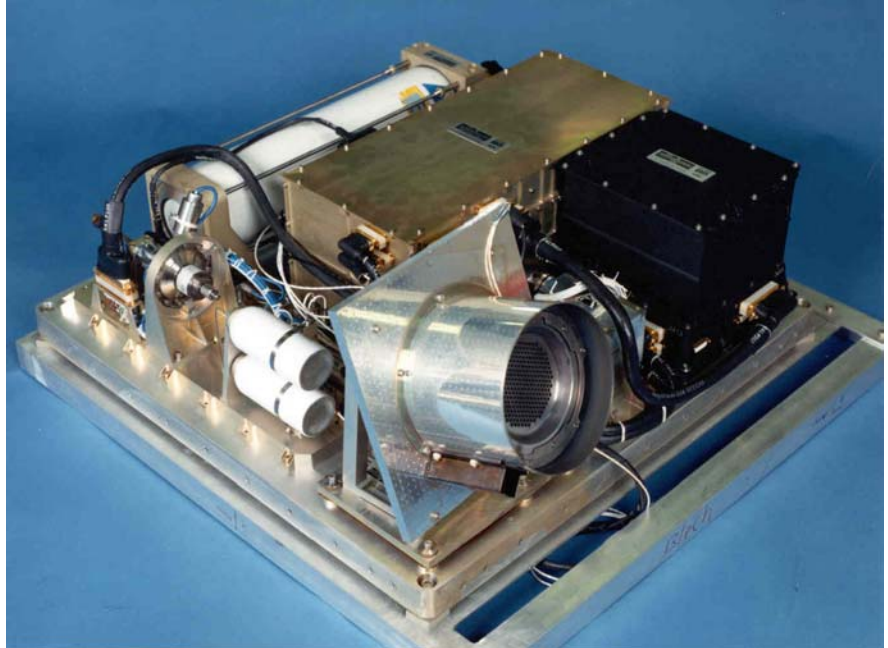


Fig. 6.2. RIT experiment equipment on EURECA

RIT 10 Application to ARTEMIS

ARTEMIS (Advanced Relay and Technology MISsion) is a geo-stationary telecommunication satellite developed and manufactured by Alenia Spazio, as the prime contractor, for the European Space Agency ESA. ARTEMIS is a three axis stabilised satellite with a launch mass of 3100 kg [5].

The main purpose of ARTEMIS is to promote advanced telecommunication technologies. In addition, the satellite was selected to demonstrate the advantages of ion propulsion for station keeping as a responsible subsystem for the total mission time of 10 years. The task of the electric propulsion system on ARTEMIS is to provide thrust for compensation of disturbances in North/South direction. This can be done by thrusting in one node only in North- or in South-direction or in both nodes. If the thrust is inclined with respect to the N/S-axis, thrusting in both nodes is necessary to compensate for the component outside the N/S-axis.

In case of ARTEMIS, 4 thrusters are mounted on the edges of the satellite, as shown in Fig. 6.3. Two RIT thrusters (RF-ionisation principle, developed by EADS-Germany) and 2 EIT (electron bombardment ion thrusters, developed by EADS-UK) are installed to take full advantage of redundancies in thruster technology and in number of thrusters. ARTEMIS was launched on July 21, 2001 from Kourou. Due to a malfunction of the ARIANE 5 upper stage, during the early phase of the 2nd stage operation, the satellite was exposed to higher vibration loads than expected, and the satellite was injected into a too low orbit (see Table 6.1).

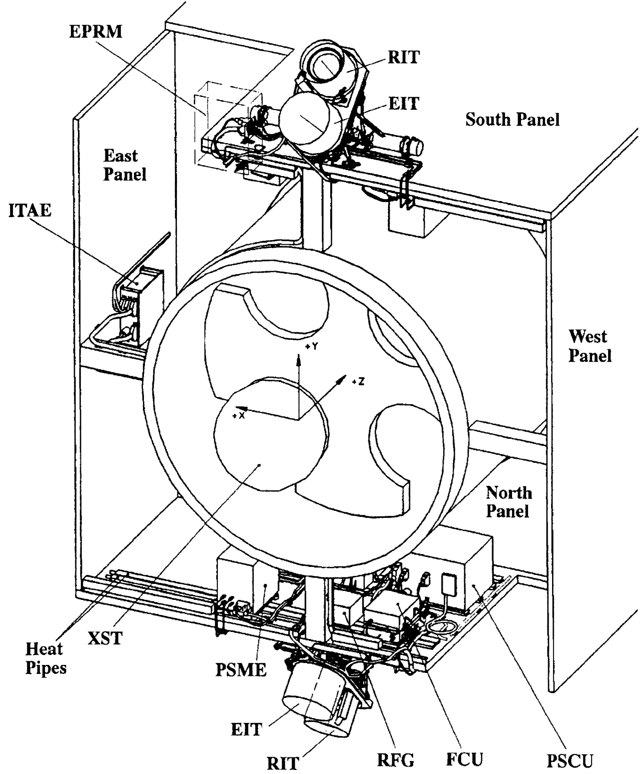


Fig. 6.3. Electric propulsion system on ARTEMIS

Table 6.1. ARTEMIS transfer orbit data

	Perigee	Apogee	Inclination
Orbit achieved	592 km	17 529 km	0°
Orbit required	858 km	35 853 km	0.8°

The satellite, although launched with a bi-propellant surplus of 200 kg, using the bipropellant system only would not be able to reach GEO and go into service. Immediately after the initial orbit was determined the ARTEMIS team decided to bring the satellite to its operational orbit by the available resources of chemical and electrical propulsion with the goal to allow 5 years mission time on station.

The rescue mission did consist of the following steps:

- Lift the perigee above the van Allen belts to avoid a decrease of the solar array power by radiation. The 400 N chemical engine has been operated during 5 successive perigee passes.
- Circularize the orbit at an altitude of 31 000 km by 3 apogee boost maneuvers using the 400 N engine.

Table 6.2. Performance of the ion thrusters during the ARTEMIS mission

	RITA 1	RITA 2	EITA 1	EITA 2
Total Ontime	1147 h	6427 h	182 h	521 h
Xenon consumption	2.1 kg	12.1 kg	0.5 kg	1.4 kg

- Raising the orbit by the ion propulsion from 31 000 km to its operational orbit of 36 000 km and compensation of the inclination failure.

The experience gained during the ARTEMIS orbit raising [6] shows the value of such a flexible propulsion architecture using bi-propellant and high specific impulse ion propulsion, which allows reacting even on nearly mission catastrophic launcher failures. After about 10 months of continuous ion thruster operation ARTEMIS has reached its final position on January 30, 2003. The following table summarizes the actual performance of the ion thrusters during the ARTEMIS mission (see Table 6.2).

RITA 2 within about one year needed for orbit raising, although operated well outside its original specifications, demonstrated an operation time comparable to more than 7 years of nominal mission for orbit control. Finally, the RITA system proved to be a well-performing and reliable candidate offering a maximum performance for future low mass and moderate power satellites.

6.1.6 Future Aspects

RIT for Commercial Applications

Ion thrusters can be used to decrease the launch mass or to increase the payload of a satellite due to the high specific impulse with respect to chemical thrusters. Different applications will have different requirements for the ion thrusters [3]:

- *Orbit Raising or Orbit Topping*. A satellite, launched into a transfer orbit by the launcher, can be brought into the final orbit by ion propulsion. This shall be done as fast as possible because commercial aspects require an early start of satellite operation. Therefore, high thrust levels should be used. As during this phase no power is needed for the payload, most of the installed solar array power can be used for it as demonstrated on the Boeing 702 satellite. Current requirements are for each ion thruster: Thrust $>$200 mN, Isp $<$35 000 Ns/kg, Power Input as low as possible.
- *Orbit Control, Mainly North-South Station-keeping*. High specific impulse will save much mass. Current requirements are for each ion thruster: Thrust $>$ 100 mN, Isp $>$ 40 000 Ns/kg.

The orbit control requirement can be satisfied by a smaller thruster at higher beam voltage and lower beam current or with a larger thruster operating in low power mode, if one type of thruster will be able to satisfy both requirements. Based on the experience gained during the development activities on RIT-EVO [7] and on ESA-XX [8], EADS-ST is developing the RF-ion thruster RIT-22 since three years.

The thruster has been operated up to now between 15 and 240 mN at beam voltages between 900 and 2300 V. This thruster meets perfectly the requirements for orbit control of geostationary satellites.

RIT for Interplanetary Missions

One of the interplanetary missions upcoming within Europe is the ESA-cornerstone BepiColombo, a mission to the planet Mercury. Two spacecrafts are currently planned to be launched by separate launchers, one will go to a Mercury orbit, the other one will carry a landing vehicle. Both spacecrafts will use a gravity assistance of the Moon and Venus to decrease the propulsion requirements. A solar electric propulsion module will be used for the interplanetary cruise on both spacecrafts to decrease the launch mass and the mission time [3].

The requirements on the electric propulsion system are:

- Thrust level per thruster 170–200 mN
- Number of thrusters installed 3
- Specific impulse >45 000 Ns/kg at the beginning of life
- Power @ 170 mN 6000 W
- Total impulse to be provided 8×10^6 Ns

The RIT-22 thruster, currently under development for commercial applications, is a candidate for this mission, which is scheduled for 2009 or later.

6.2 Field Emission Electric Propulsion (FEEP)[2]

6.2.1 Historical Development

The first field emission device which has been considered for application in space propulsion was a liquid metal droplet sprayer, developed by Krohn [9] in the 1960s at Ramo – Wooldridge Research Labs (USA), using Wood's metal as a charge material. This project, however, was abandoned, also because "unfortunately, . . . large numbers of ions are produced along with the droplets". However, it was just this ionic component with its extremely high specific impulse which later fostered development of field ion thrusters.

When the European Space Agency (ESA) started its own development program on electric space propulsion in 1968, the first activities were devoted to the exploration of a colloid thruster, and the idea of using linear emitter geometries instead of capillary or annular-shaped electrodes was advanced. This development was also based on experiments on the emission of cesium ions from capillary tubes [10].

At the European Space Research and Technology Centre (ESTEC), a field emission or field effect electric propulsion (FEEP) system based on the liquid metal ion

[2] Please refer to J. Mitterauer, F. Rüdenauer.

source (LMIS) principle with cesium as the propellant has been progressively developed and evolved from a single-pin emitter, through linear arrays of stacked needles to the presently favored slit emitter module [11].

Experimental investigations of the mechanism of ion formation in the ion emitting region were performed at the Culham Laboratory (United Kingdom) [12] and at the Vienna University of Technology (Austria) [13, 14]. The cesium development was finally transferred to Centrospazio S.A. at Pisa (Italy) [15].

However, the use of condensable liquid metal propellants such as cesium has caused reluctance on the part of potential users because of their concern about spacecraft contamination and launch-safety issues. Therefore, as early as 1978, ESTEC had been aware of these complications and awarded two research contracts to the Fulmer Research Laboratories Ltd., UK, in order to identify alternative liquid metal propellants to cesium [16, 17].

Nevertheless, for a long time cesium was considered by ESTEC as the most favored propellant for FEEP. However, in 1997 ESTEC awarded a contract to the Austrian ARC research GmbH (ARCS) for the development of an ultraprecise indium thruster for use in low thrust FEEP systems [18]. Obviously, ESTEC had been aware of the highly successful 6-year flight of the spacecraft potential control instrument on the Japanese GEOTAIL satellite, which has been developed by ARCS on the base of an indium LMIS [19], and earlier successful tests of indium LMIS on the MIR [20] space station.

6.2.2 Electrophysical Fundamentals

The ion emission mechanism in a FEEP thruster is identical to that utilised in liquid metal ion sources (LMIS). Such devices are applied as high brightness primary ion sources in SIMS (see Sect. 5.1) and in focused ion beam micromachining devices (FIB) for microelectronics device fabrication [21]. A liquid metal surface, usually adhering to a solid metal support structure, is exposed to a strong electrostatic field (Fig. 6.6). Under the combined action of this field and of surface tension the liquid metal surface deforms and, depending on the geometry of the support structure (single needle or capillary, capillary slit, pool, see Fig. 6.4), assumes an equilibrium configuration consisting of a single or of multiple conical structures, the Taylor cones (apex angle 98.6 deg). At the apices of these cones (diameter $< 10^{-8}$ m), the field is high enough (10^9–10^{10} V/m), so that ions can be field-evaporated directly from the liquid without the transitional vapour phase which is common in the technology of other ion source types. Ions emitted from the apices are replenished through hydrodynamic flow from a liquid metal reservoir, so that a continuous ion beam is emitted from each cone. Flow of the metal to the tip introduces additional hydrodynamic forces, so that the actual geometry of the emission sites is that of a Taylor cone elongated at the apex in the shape of a jet with a tip width of a few nanometers only [22]. The reaction force to the electrostatic acceleration force acting onto the beam ions constitutes the thrust acting onto the ion emitter device in the opposite beam direction.

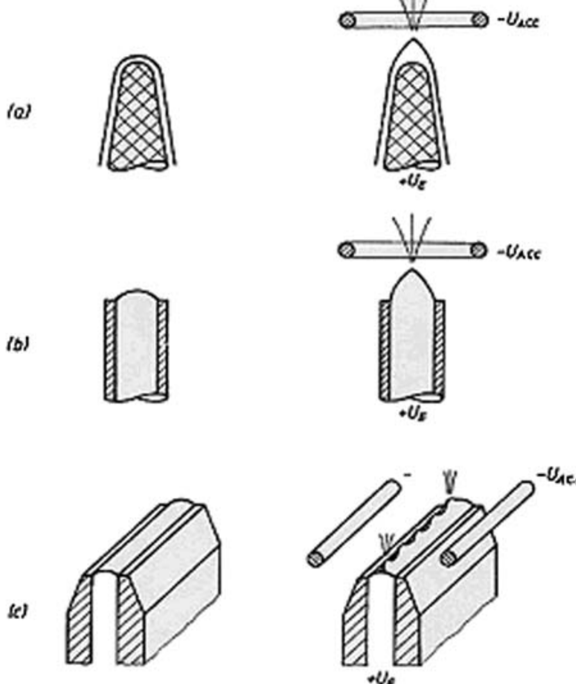


Fig. 6.4. Schematic diagram of different types of liquid metal ion sources [23]: **a** pin or needle type; **b** tube or capillary type; **c** elongated slit type. *Left*: shape of the liquid metal tip without electric field. *Right*: distortion of the liquid metal tip by an electric field due to a positive emitter potential $+U_{\mathrm{E}}$ and a negative accelerator potential $-U_{\mathrm{ACC}}$ versus common ground

6.2.3 Present State-of-the-Art and Applications

Field emission electric propulsion (FEEP) thrusters belong to a different class of ion propulsion devices, the *microthrusters* with a thrust capability ranging from the μN to the low mN range. Accordingly, these devices are not thought to be used as the main engine of a spacecraft but rather for high precision attitude control in the newly emerging microspacecraft technology with its severe constraints in mass (100 kg–1 kg), dimension (1 m–0.1 m) and power (100 W–1 W). The new class of "fundamental physics" space missions [24], in particular require ultrahigh precision attitude control. Goals of these missions are the detection of low frequency gravity waves (LISA [25]), the test of Einstein's equivalence principle with unprecedented accuracy (MICROSCOPE [26]), verification of various aspects in general relativity (HYPER [27]) or the assembly of spaceborne interference telescopes for ultraprecise astrometry (GAIA [28]) and detection of earth-like planets (DARWIN [29], TPF). In order to achieve these goals, completely dragfree experimental platforms, often on multiple spacecraft formations, must be realised with orientation and mutual distance accuracy of down to μarcsec and 50 pm, respectively. For the control and stabilisation of such platforms, thrusters with a thrust range of 1–100 μN, a control accuracy

of ≈0.1 μN and a thrust noise of <0.1 μN/Hz$^{1/2}$ (in the 0.1 Hz frequency band) are required.

At present, these requirements are thought to be achievable with FEEPs only. FEEP, however, still must be considered experimental devices at an advanced stage of development. Many of the necessary requirements have already been demonstrated in the laboratory, but essential features, in particular lifetime and spacecraft contamination, still have to be verified. The first two fundamental physics missions which already have placed contracts for supply of FEEP dragfree control are MICROSCOPE (to Centrospazio) and GOCE (to ARC research).

The actual state of FEEP thruster technology for micropropulsion is characterized by a competition between two concepts. On the one side, the slit emitter developed at the European Space Research and Technology Center ESTEC (The Netherlands) [11] and at Centrospazio (Italy) [30] is a semi-miniaturized liquid metal ion source (LMIS) with cesium as the propellant, but without any space qualification. On the other side, the space proven single needle or capillary microthrusters and charge compensating devices developed at the ARC research [31] with indium as the propellant.

The FEEP linear array slit emitter depicted in Fig. 6.5 represents an ultimate development in precision mechanics, demonstrated by a value of about 10^{-6} m for both the slit width and the round-off radius of the emitter slit edges [11]. Typical data for the emitter–accelerator geometry are as follows: $a = 0.6$ cm, $2b = 4$ mm, $d = 5$ mm, and $w = 1.2$ μm. Until now, emitter modules with a slit length of 1, 3, 5, and 8 cm have been produced. Homogeneous high current ion emission from a micron-sized slit has been demonstrated successfully, allowing the occurrence of a linear series of equally spaced emitting sites with a distance of less than 10 μm at a maximum linear current density of more than 5 mA/cm and a (calculated) thrust of about 880 μN/cm for a voltage drop of 15 kV between the emitter and accelerator electrode [23]. Ion onset linear emission current density is about 10 μA/cm for a voltage of 7.5 kV; the corresponding thrust is about 1 μN/cm. So far, the slit emitter device has only tested in the laboratory.

A different approach has been chosen by ARC research at Seibersdorf (Austria) [32]. Here, a single tungsten needle with a tip radius of ≈2 μm is wetted by liquid indium metal. When the needle is put on a potential of $\approx +5$ kV with respect to an extraction electrode, a single Taylor cone is forming at the needle tip, emitting ion currents between 1 μA (5.5 kV) and 1 mA (9 kV), corresponding to a thrust range between 0.1 μN and 100 μN (see Fig. 6.6). The emission is stable, thrust noise is of the order of 0.1 μN/Hz$^{1/2}$ at 0.1 Hz and 20 μA emission, increasing with thrust, thrust controllability is < 0.1 μN [33].

Indium has been chosen as the charge material because it is the lowest melting pure element which is not liquid at elevated ambient temperature and because of its reduced chemical reactivity and lower oxidation speed with respect to cesium. Although maximum specified thrust is lower for the In than for the Cs device, indium metal is much easier to handle than cesium. Total power consumption of the device is ≈2.5 W (secondary) at 25 μN. A complete system consisting of thrusters, electron neutraliser and miniaturised electronics has been set up, and charge neutralisation

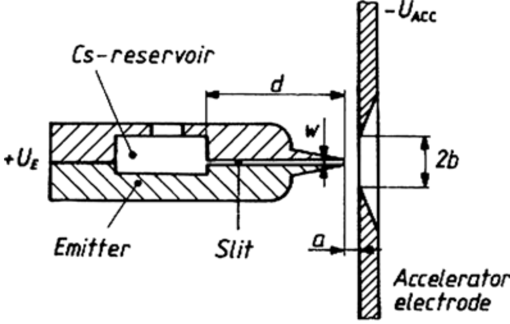


Fig. 6.5. Sectional view of the emitter-accelerator electrode configuration of a cesium slit emitter [23]: a = emitter–accelerator spacing, $2b$ = accelerator aperture, d = emitter slit depth, w = emitter slit width

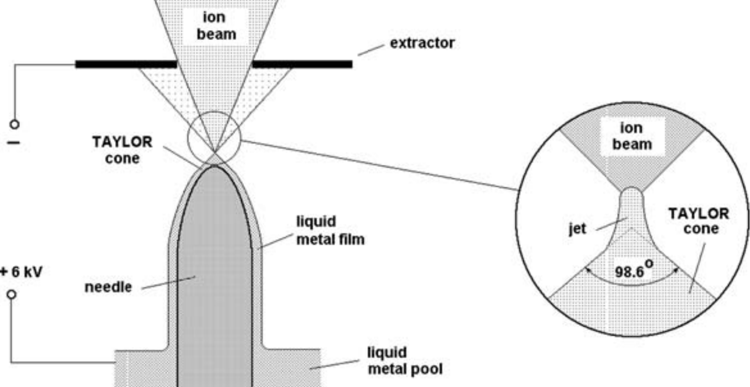


Fig. 6.6. Liquid metal ion source of the single needle type, schematic

has been demonstrated. Lifetime tests are underway, 5000 h continuous operation at a realistic thrust profile with 35 μN peak thrust has already been achieved [34] (Fig. 6.7). Long term stability and zero gravity operability of the ARC device has been proven in a number of successful space missions, where it has been flown, at reduced emission levels, as charge control device for satellites [35, 36], and as primary focused beam source for materials analysis in space [37, 38].

6.2.4 Future Aspects

The experience gained so far from long duration lifetime tests under realistic operating conditions suggest that the following issues are relevant for the development of highly reliable FEEP systems: (a) internal contamination by ion beam sputtering and deposition; (b) shielding of satellite from ion plume; (c) stability of wetting of emitter needles or slits. At least the first two problems could be greatly reduced by seriously reducing emission of metal droplets or neutral atom component. The safest

Fig. 6.7. ARC research indium FEEP cluster in endurance test

known method to achieve this is the reduction of emission current per emission site. For a given thrust and total emission current this requirement is equivalent to increasing the number of individual emission sites. This leads to the concept of an ion emitter array, which most sensibly could be produced using modern microfabrication techniques.

Within the new technical fields that have come to be known as vacuum microelectronics and microelectromechanical system (MEMS) based technology, microfabrication techniques have been developed to produce micrometre-sized cones, wedges, microvolcanos, and also linear and two-dimensional arrays of these microstructures as field emission electron and ion sources [39].

The concept of MEMS-based micropropulsion was first introduced in 1991 in the form of a proposed microfabricated FEEP thruster based on field emitter array (FEA) technology [40]. The purpose of that design study was to decrease further the mass and size of FEEP thrusters and to take advantage of the fact that critical thruster components, such as the emitter slit, were already sized in the micrometre range even for conventional designs.

In principle, any FEEP system requires an electron source as a neutralizer in order to prevent (or to control) spacecraft charging. Hollow cathodes and thermionic cathodes have traditionally been used as neutralizers; for miniaturized thrusters a more efficient technology should be compatible [41]. The miniaturization of a neutralizer requires a highly efficient electron source of low power demand, preferentially a field emitter source. Undefinable vacuum conditions in space exclude classical field emitter arrays (FEA's) to a certain extent and encourage carbon nanotube (CNT) emitters.

Just a decade ago, the miniaturization of a FEEP liquid metal ion thruster seemingly was dominated by the selection of suitable fabrication methods related either to vacuum microelectronics or to MEMS technology; as the liquid metal propellant, cesium was unquestioned.

In the meantime, the introduction of the alternative propellant indium has opened new perspectives, and the competition between the two elementary liquid metals above all is dominated by reliability issues [42].

For the future, the realization of any miniaturization will be governed by two basic aspects [43]: on one side, the optimization of the microstructure of liquid metal ion sources, and on the other side, the optimization of the liquid metal propellant itself. The first aspect relies on metal or semiconductor physics and on micro/nanotechnology; the second one embraces an exact tailoring of liquid metal alloys or compounds with prospective features superior to those of elementary metals.

6.3 Hall Effect Thrusters[3]

6.3.1 Introduction

Hall effect thrusters (HETs) are gridless plasma devices which generate thrust by emitting propellant ions [44]. The HET discharge is sustained by means of crossed electric and magnetic fields such that ions are successively produced and accelerated. Since ion extraction occurs from a space charge neutralized plasma, no extracting grids are needed, and high thrust densities are obtained. HETs offer a compact and robust design and provide specific impulses which are a factor of 5 to 10 higher than those of chemical thrusters at reasonably low electric input powers.

HET concepts were first developed in the 1960s in the United States and Germany, and in the Soviet Union [45]. Due to initially poor efficiency values, the western world focused on the development of grid ion thrusters, whereas Russian researchers and engineers continued to gradually improve the HET concept. The highest technical readiness level has been achieved with the so-called stationary plasma thrusters developed under the leadership of A. Morozov with thrust ranges from a few tens of miliNewtons to several Newtons.

From 1972 onwards, HET based electric propulsion systems have been applied in numerous satellite and space probe missions in the former USSR, and since the 1990s, HET development has been re-initiated in the US, Europe and Japan [46, 47]. Amongst European HET development activities, the most intense program has been conducted by French space agency CNES [48].

In summary, during more than 40 years of continuous development and 30 years of flight experience in space with more than 150 HETs flown or still in flight, HETs have accumulated the largest and most successful flight heritage amongst all ion propulsion devices.

In this section the operational concept and characteristics of HETs are presented and examples of HET applications in space missions are given.

[3] Please refer to N. Koch.

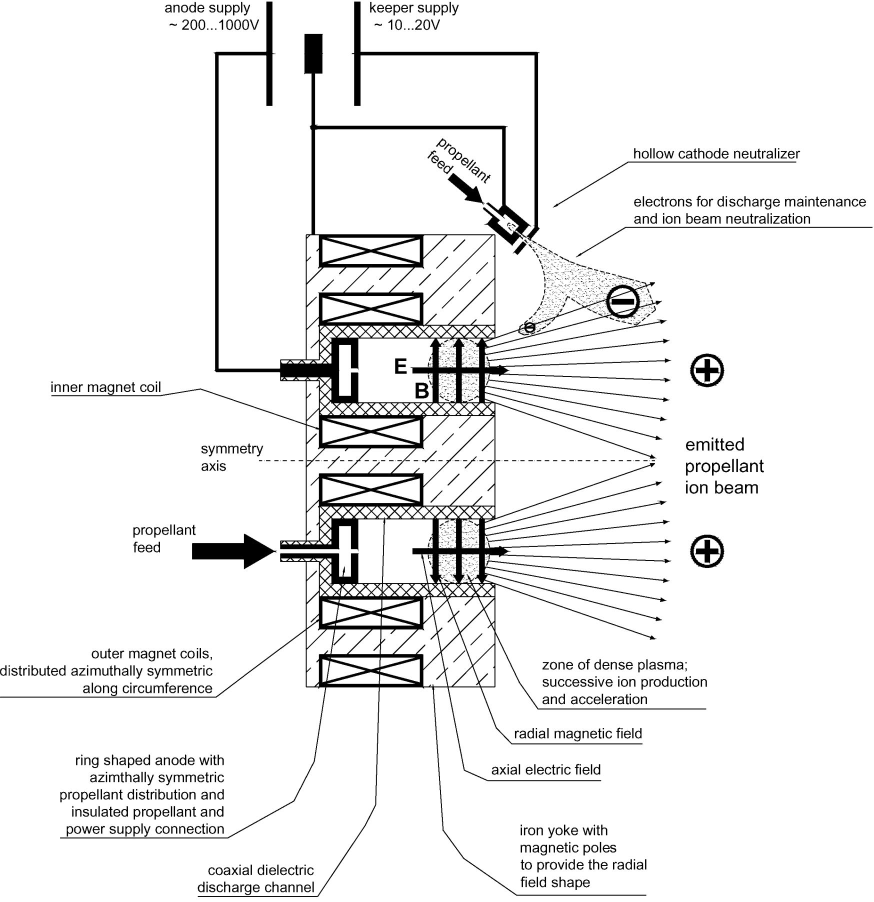


Fig. 6.8. Schematic cross-sectional view of a coaxial Hall effect thruster of SPT-type

6.3.2 Operational Concept and Characteristics

Basic Phenomena

A schematic cross-sectional view of a coaxial HET is given in Fig. 6.8. Typically, HETs employ a ring-shaped anode through which gaseous or evaporated propellant is injected into a coaxial discharge chamber. Two different types of HETs are available, which differ by the material of the discharge channel: stationary plasma thrusters (SPTs) use a dielectric channel as indicated in Fig. 6.8, whereas thrusters with anode layer (TALs) have an electrical conducting channel with the anode positioned closer to the thruster exit.[4] HETs employ a radial magnetic field across the exit-sided end of

[4] Due to their exclusive use in space missions, the description of the HET operational principle will be restricted to coaxially shaped SPTs.

the discharge channel, which is provided by means of an inner magnet coil, several outer magnet coils distributed azimuthally symmetric along the circumference. An iron yoke with magnetic poles accounts for the appropriate field shape.

Electrons, usually emitted by a hollow-cathode neutralizer placed at the thruster exit, drift towards the anode energized by the axial electrical field, which is induced by the voltage drop between anode and cathode. Electron drift is impeded by the perpendicularly oriented radial magnetic field along which the electrons fulfill a spiral motion with a Larmor radius being small compared to the width of the discharge channel.[5] In addition, the crossed electric and magnetic fields induce a strong azimuthal electron drift which results in an azimuthal Hall current. HETs are named in accordance to this type of Hall effect.[6] Electrons colliding with the discharge channel walls are either attached or reflected, or induce secondary electron emission. Wall interaction, collisions with neutral propellant atoms and ions, Bohm diffusion and local electric field fluctuations enhance electron mobility towards the anode across the perpendicular magnetic field.

The intense Hall current and the effectively long dwell time of energetic electrons in the $\mathbf{E} \times \mathbf{B}$ zone results in a high ionization rate of propellant atoms such that a dense plasma is generated. In addition, due to the high electron impedance induced by the magnetic field, a strong electric field is built up which accelerates the propellant ions created in the plasma towards the thruster exit. It is important to note, that the propellant ions are practically not affected by the magnetic field because of their high mass, which is a factor of about 10^5 higher than the electron mass.

The emitted ion beam is neutralized by an equal electron current from the hollow-cathode neutralizer to avoid charging up of the spacecraft.

Discharge Characteristics

The typical evolution of characteristic plasma parameters along the discharge channel axis of a SPT-type HET is shown in Fig. 6.9 taken from [49].[7] As a consequence of the ionization process at a given magnetic field topology, the electric potential first drops linearly and then decreases steeply towards the thruster exit. This results in a strong increase of the electric field. The charged particle density – electron density equals ion density because of the charge neutrality requirement of the discharge – rises due to continuous propellant ionization and becomes maximal towards the thruster exit. Simultaneously the neutral gas density approaches zero, i.e. most of the propellant atoms are ionized. At that point, the steep rise in electric field leads to an enhanced ion acceleration, so that the ion density decreases despite a constant ion current. Since electrons are trapped in the magnetic field, the drop in ion density

[5] At typical magnetic field levels, the electron Larmor radius is in the range of some tenths of millimeter, whereas the channel width is in the centimeter range.

[6] Due to the strong azimuthal electron drift, HETs are also referred to as closed-drift thrusters.

[7] The presented graphs stem from a numerical simulation and approach qualitatively and to a large extent also quantitatively experimental data.

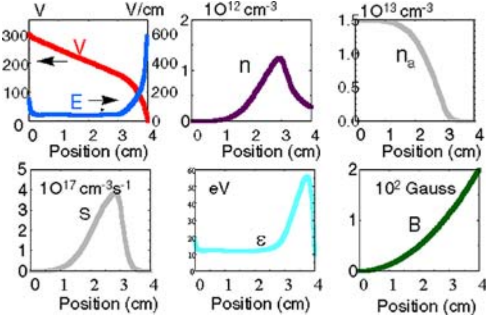


Fig. 6.9. Simulation of the axial discharge profile taken from [49]. The graphs from *up left* to *bottom right* show the axial evolution of the electric potential V and electric field E, charged particle density n, neutral gas density n_a, ionization rate density S, electron temperature ε and radial magnetic field amplitude B, respectively. Position 0 denotes the discharge sided anode plane, whilst at 4 cm the thruster exit is located. The discharge is operated at 300 V and a xenon propellant mass flow of 5 mg/s

is compensated by a rise in electron temperature to maintain space-charge neutrality. The ionization rate density evolves accordingly, since it is proportional to the product of charged particle density and neutral gas density.

Ion Beam Characteristics

HET ion beam characteristics are strongly related to the discharge process. In the photograph of Fig. 6.10 [50], a typical appearance of a HET exhaust plasma plume is shown.

The intense xenon discharge appears as a bright zone, whereas the fade light emission surrounding the plume arises from ion beam interaction with the residual gas in the vacuum test chamber. On top, the plasma bridge emitted by the hollow-cathode neutralizer is visible.

Since the subsequent processes of propellant ionization and ion acceleration are linked in a self-consistent manner, both energy and trajectory of an individual ion strongly depends on the location of its production. Typically, the polar angle distribution of the ion current density profile measured in the far field of the thruster is nearly Gaussian with a maximum on axis, and an ion beam spread that reaches up to 90° [51–54]. As a consequence, HETs exhibit divergence angles[8] of up to 45°. As an example, the polar angle distribution of the ion beam current density is shown in Fig. 6.11, taken from [54].

The typical evolution of the ion beam energy distribution is shown in Fig. 6.12 for a D-55 TAL-type HET operated at an anode voltage of 300 V [55]. A sharp peak centered at about 20 eV below anode potential (40 eV in case the thruster is operated in floating mode with neutralized ion beam) is observed with a flat shoulder in the

[8] Beam divergence denotes the polar angle range into which 95% of the total ion beam current is emitted.

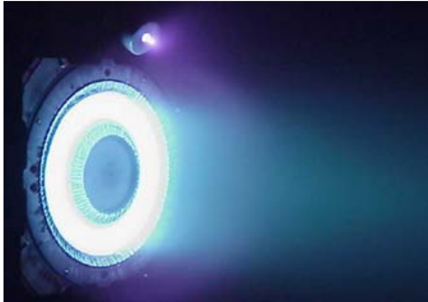


Fig. 6.10. Photograph of the exhaust plume of a PPS® 1350 HET, taken from [50]

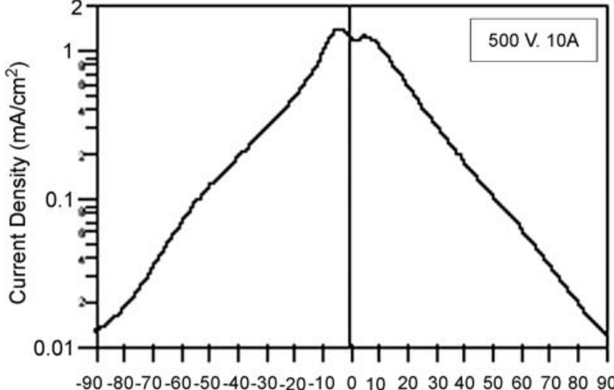


Fig. 6.11. Polar angle distribution of the ion beam current density of a P5 HET operated at a anode voltage and current of 500 V and 10 A, respectively (taken from [54])

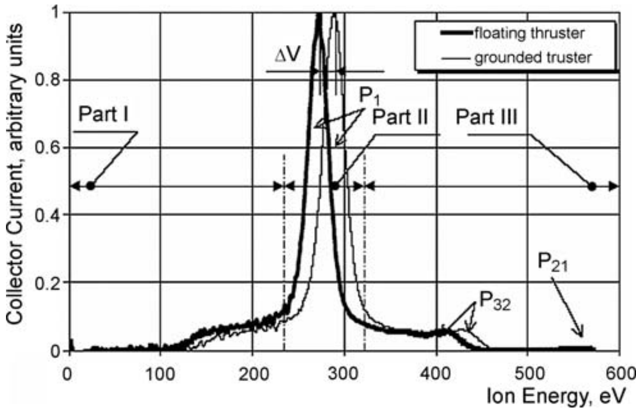


Fig. 6.12. Ion beam energy distribution of a D-55 TAL-type HET operated at an anode voltage of 300 V, taken from [55]

range of lower energies down to 130 eV. Energy levels observed above anode potential are due to multiple charged ions undergoing charge exchange collisions beyond the thruster exhaust.

Plasma Oscillations and Discharge Stability

HET discharges have complex oscillation and noise characteristics over a wide frequency spectrum [56–61]. These oscillations can result in plasma instabilities up to plasma extinction, hardware damage of thruster power supplies and periphery due to excessive discharge current fluctuations and a significant reduction in thruster efficiency. Thruster oscillations are related to space variation of the ionization front and to electron transport phenomena. Consequently, thruster stability is strongly influenced by discharge voltage and magnetic field topology, propellant inlet scheme, discharge channel geometry and the appropriate choice of the channel wall material. An example for spontaneous discharge current oscillations in a SPT-100 HET is given in Fig. 6.13, taken from [58].

Discharge Channel Erosion

HET magnetic field topology results in a electron impingement on the discharge channel wall towards the thruster exhaust. Consequently, a negative sheath potential builds up, and the impact of the resulting radial electric field on ion trajectories leads to ion impingement at the discharge channel exit. Since sputter yields of typically used channel materials are significant for ion energies in the range of several 100 eV [62], the discharge channel is eroded with time. This imposes the main life time limitation to HETs. An example of the radial erosion profile of the inner discharge channel wall as a function of the axial position is shown in Fig. 6.14, taken from [63].

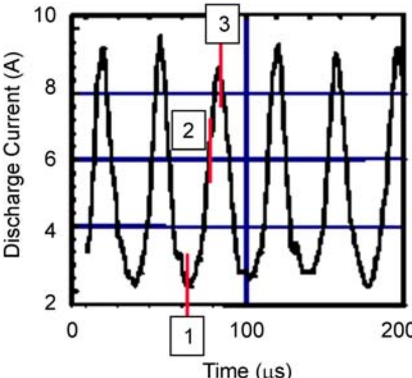


Fig. 6.13. Example for spontaneous discharge current oscillations observed at a SPT-100 HET taken from [58]

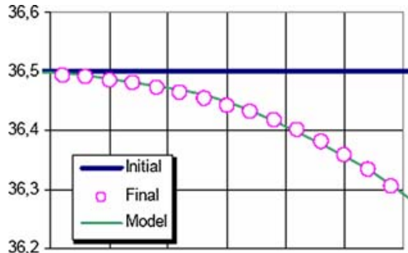


Fig. 6.14. Axial erosion profile of the discharge channel inner wall of a KM-100 SPT after 25 h of operation at 300 V discharge voltage, taken from [63]. The right side of the abscissa denotes the thruster exit, one unit equals 0.1 mm, the y-axis denotes the radial profile in units of 1 mm

After 25 hours of thruster operation at a discharge voltage of 300 V, 0.2 mm of wall material are eroded at the exit, and an erosion depth of 5.5 mm is predicted for 1000 h of operation.

Discharge channel erosion is influenced by the applied discharge voltage, magnetic field topology, discharge channel material, surface properties, temperature and plasma oscillations.

Thruster Operational Parameters and Efficiency

HETs are typically operated at moderate discharge voltages around 300 V to provide reasonable specific impulses on the order of 1500 s at a moderate power consumption of about 17 W/mN. The entire operational envelope encloses discharge voltages from 200 to 1000 V. The magnetic circuit usually is fed by coils, and the maximum of the radial magnetic field amplitude usually amounts 20 to 30 mT. In order to avoid electric power losses dissipated in the coils, efforts have been made to employ permanent magnets instead [64]. Detailed studies for SPT-type HETs have shown that discharge channels made of BN-SiO_2 ceramics ("Borosil") yield the best performance characteristic [65].

Discharge current–voltage characteristics in the typical HET operational range from 300 to 600 V are in first order flat, i.e. for a given propellant mass flow, the discharge current varies only weakly with the applied voltage. Consequently, at a given discharge voltage, discharge current and hence thrust are regulated by the propellant mass flow. In order to minimize plasma oscillations, changing of the discharge voltage requires proper tuning of the magnetic field.

Usually xenon is used as propellant due its low ionization potential and high ionization cross-section. In addition, the high atomic mass of xenon allows for low power-to-thrust ratios and negligible elastic electron energy losses. Last but not least, xenon is chemically inert and easy storable, which are significant advantages with respect to integration of a HET propulsion system on a spacecraft.

The thruster anode efficiency is given as the product of propellant ionization efficiency, beam power efficiency and angular efficiency of the ion beam.[9] Efficient propellant ionization requires sufficiently long dwell time of the neutrals in the plasma zone and minimal ion recombination. The beam power efficiency is defined as the ratio of the kinetic ion beam power to the discharge power. In order to obtain high beam power efficiency, the ratio of the discharge current to the emitted ion beam current has to approach unity, and the major part of the ions has to be produced in plasma regions at high potential close to the anode voltage. Inelastic losses, in particular due to production of multiple charged ions, reduce the beam power efficiency, electronic and ionic wall losses and strong discharge current fluctuations. The angular efficiency is defined as cosine square of the effective ion beam angle, i.e. the polar angle into which the ion beam power on average is directed. Thus an angular efficiency beyond 90% requires that ions are mainly distributed in a polar angle range of less than 20°. A high angular efficiency reduces also the beam divergence angle defined previously and is therefore a crucial parameter for HET integration on a spacecraft.

State-of-the-art HET flight models achieve anode efficiencies in the range of 35 to 50% [47]. Whereas propellant ionization and ion beam angular efficiencies are high in a range between 85% and 95%, beam power efficiencies are moderate and amount typically less than 70%. As shown in a detailed study [66], most of the losses in beam power efficiency are due to thermal losses on the discharge channel induced by plasma–wall contact. Therefore, reduction of the electronic and ionic wall contact represents the main development potential of HETs with respect to both a higher anode efficiency and lifetime.

6.3.3 Examples for HET Applications in Space Missions

Current space activities involving HET propulsion systems are mainly based on Russian experience of the past 30 to 40 years. Apart for Russian missions, qualified HET flight models have been and are being developed and employed in Europe and the United States.

Russian Missions

Due to electric power limitations on the spacecraft, SPT type HETs were initially developed for the range of low input powers and optimized for operation at low discharge voltages around 300 V, where moderate specific impulses are obtained at low power-to-thrust ratios. An overview of current Russian flight proven SPT models and their performance characteristic is given in Table 6.3 [45].

Starting from the METEOR satellite missions in 1972, more than 140 SPTs have been flown or are still in flight. HET applications in Russian space missions are listed in Table 6.4 from [45].

[9] The thruster anode efficiency neither takes into account electric power losses dissipated in the magnetic coils and in the neutralizer cathode nor the mass flow fed into the neutralizer.

European Activities

The most recent mission of the European Space Agency (ESA) employing HET technology is the SMART-1 satellite (Small Mission for Advanced Research in Technology), which has been launched in fall 2003. SMART-1 is a small lunar orbiter devoted to the demonstration of innovative key technologies for scientific deep space missions [67–69]. It involves a solar electric propulsion module based on a SNECMA PPS® 1350-G HET, which has shown excellent operational behaviour. The PPS® 1350, which has been developed in co-operation with Russian EDB Fakel, is presented in Fig. 6.15, taken from [70]. The photographs show a new PPS® 1350-G flight model and a PPS® 1350-IM engineering model after 5870 h of ground operation.

Satellite manufacturer ALCATEL, together with CNES and SNECMA, has set up and qualified HET electric propulsion platforms based on Russian SPT-100 and PPS® 1350 thrusters for the STENTOR scientific satellite and the ASTRA-1K geostationary telecommunication satellite [71]. Unfortunately, both satellites got lost due to launcher failures.

Future activities are aimed towards high power HET propulsion for the large @Bus Platform for heavy geostationary telecommunication satellites. @Bus is a

Table 6.3. Performance characteristics of Russian flight model SPTs [45]

	SPT-50	SPT-70	SPT-100
Nominal anode power	0.35 kW	0.7 kW	1.35 kW
Nominal thrust	20 mN	40 mN	80 mN
Specific impulse	1100 s	1500 s	1600 s
Lifetime (ground testes)	1500 h	3000 h	9000 h
Anode efficiency	0.35	0.45	0.5

Fig. 6.15. Photographs of space qualified SNECMA PPS® 1350 thrusters, taken from [27]. On the left, a new PPS® 1350-G flight model is shown. The picture on the right shows a PPS® 1350-MI engineering model after 5870 h of operation

Table 6.4. Russian missions with HET based electric propulsion systems [45]; 152 SPT-type HETs have been flown or are still in flight

Year	User	Satellite type	Mission	Number of thrusters
1972	VKDEM	Meteor#18 ($H = 900$ km)	Flight test, orbit correction	2 SPT-60, $\eta_t = 0.15$–0.21
1974	VNIIEM	Meteor-Priroda#1	Flight test, orbit correction	2 SPT-60M, $\eta_t = 0.26$–0.28
1976	VNIIEM	Meteor-Priroda#2	Orbit correction (S/C positioning)	2 SPT-50, $\eta_t = 0.33$
1977	VNIIEM	Meteor-Priroda#2-2	Orbit correction (S/C positioning)	2 SPT-50, $\eta_t = 0.33$
1978	VNIIEM	Astrophysics	Orbit correction (S/C positioning)	2 SPT-50, $\eta_t = 0.33$
1981	VNIIEM	Meteor-Priroda#2-4	Orbit correction (S/C positioning)	2 SPT-50, $\eta_t = 0.33$
1987	Arsenal	Plasma	Orbit correction	2×6 SPT-70
1982–2000	NPO PM	Kosmos, Loutch, et al.	S/C positioning and E-W SK	16×4 SPT-70, $\eta_t = 0.45$
1995–2001	NPO PM	GALS (2 S/C)	S/C positioning and E-W and N-S SK	2×8 SPT-100 ($\eta_t = 0.5$)
	NPO PM	Express (4 S/C)		4×8 SPT-100
	NPO PM	SFSAT		1×8 SPT-100
	PSC Energiya	Yamal-100		1×8 SPT-70

common project initiated by ESA and involves national European space agencies and satellite manufacturers ALCATEL and ASTRIUM [72]. As an example, a 5 kW class HET named PPS® 5000 manufactured by SNECMA is currently under qualifi-

cation. This device provides thrust levels and specific impulses of up to 335 mN and 3240 s at input powers of as high as 6.3 kW.

Activities in the United States

HET development activities in the US are manifold. Amongst others, they aim towards very high power devices. As an example, a 50 kW class HET has been developed which provides thrust levels of up to 3 N [73]. A large program has been initiated by LOCKHEED MARTIN Space Systems to integrate a 4.5 kW HET propulsion system on their commercial telecommunication satellite platform [74, 75]. The activities included detailed studies on thruster–spacecraft interaction and qualification of an AEROJET BPT-4000 HET.

6.4 High Efficiency Multistage Plasma Thruster[10]

6.4.1 Introduction

The idea of rocket like propulsion using charged particles (first electrons were considered, later ions), accelerated in electric fields to otherwise not achievable velocities, is about 100 years old and intimately connected with names like Konstantin E. Tsiolkovsky in Russia, Robert H. Goddard in the United States, Hermann J. Oberth and Ernst Stuhlinger in Germany. A recent review on the visionary first 50 years of electric propulsion (EP) was recently published by E.Y. Choueiri [76] and is well recommended to the interested reader. Ernst Stuhlinger is also the person who started 50 years ago in the United States to transfer the visionary ideas of EP into practical developments as can be read in the previous chapters.

Different concepts developed in the past 50 years to ionise a propellant and to accelerate the ions are certainly more numerous than different types of propellants. Only a few fundamental concepts have found their way to application in space environment, are close to the required space qualification, or have a future potential for space application. The most important of them are treated in the previous Sects. 6.1–6.3 of this book. Those are the field effect electric propulsion thruster (FEEP), the grid ion thruster (GIT) and the Hall effect thruster (HET), respectively. Their development started in the late 1950s or early 1960s.

Here, a new thruster concept is presented. It was invented only a few years ago in 1998 by the author of this article and his co-workers [77]. Between the years 2000 and 2004 the electric propulsion group at the Thales Electron Devices in Ulm, Germany, developed a device called high efficiency multistage plasma thruster (HEMP thruster). It shows an extremely wide thrust range from a few hundreds μN to more than 150 mN [78–80]. The name HEMP thruster is related to its high conversion efficiency of more than 80% of electric power into kinetic ion beam power and to its multicusp magnetic design. Though azimuthal, cusp-type magnetic field arrangements were known before in the plasma source community to reduce the plasma to

[10] Please refer to G. Kornfeld.

wall contact, the idea for using axially periodic magnetic cusp field arrangements to confine and accelerate ion beams is new and originates from the travelling wave tube (TWT) technology treated in Sect. 1.3 of this book. There, permanent periodic magnet (PPM) systems are used to focus electron beams through small diameter delay lines and to decelerate and sort electron beams effectively on collector stages with depressed potentials in magnetically focused multistage collectors. In fact, the HEMP thruster was originally thought as an inversed multistage depressed electron collector. Instead of decelerating an electron beam an ion beam, should be efficiently accelerated.

Since this is done in magnetised, quasi-neutral plasma, it introduces, in spite of the different device and magnet field topology, some similarities with the Hall effect thrusters.

The new thruster concept promises to solve unsettled problems of the standard thruster types with respect to life, reliability, flexibility, simplicity and cost. Though being developed since almost 50 years and having achieved space qualification and successful in-space flight experience, the standard technologies still suffer from design specific problems leading to life limitations. For the GITs and HETs, the limitation comes from sputter erosion of the thruster grid electrodes or ceramic channel walls, respectively. Pollution of the thruster and satellite structure with liquid metal droplets seems to be the major limitation for the FEEP thrusters. The HEMP thruster concept avoids those drawbacks by design:

- using the noble gas Xe (without being limited to it), no liquid metal pollution could impact the thruster performance as for a FEEP;
- without an interfering grid electrode system, the respective erosion as in a GIT can not occur;
- being an $\mathbf{E} \times \mathbf{B}$ device as a HET, the HEMP overcomes the HET channel wall erosion problem by an effective, multicusp magnetic confinement of the plasma in the acceleration channel and reduces their high thermal dissipation.

6.4.2 Design and Physical Operation Principle

Before describing the physical operation principle of the HEMP thruster in detail, the global physical properties of ionisation and ion beam acceleration are compared for the different thruster concepts.

In a FEEP system, production and acceleration of ions can not be separated because it is an undividable, localised field emission process at the tip of a liquid metal surface. The FEEP produces no plasma because the electrons remain in the liquid metal anode of the device. The FEEP can therefore be considered as the most elementary method to produce an ion beam. The ion beam currents must remain principally small and the applied voltages high to avoid space charge limitation effects.

In a GIT, the ionisation process of a gaseous propellant takes place in a plasma reservoir chamber either by absorption of RF power (RIT, radiofrequency ionisation thruster) or electron beam power (EIT, electron bombardment ionisation thruster).

The ionisation is completely separated from the ion acceleration. A grid mesh electrode system, adjacent to the ionisation chamber, separates the ions from the plasma electrons and accelerates them. Due to the complete separation of charges in the grid electrode system, the ion beam emission properties are fully controlled by the well-known Child–Langmuir space charge emission law. The space charge limitation explains why the GIT is principally a low ion beam current density device.

There is no space charge limitation in a HET. Instead, we find an overlap of ionisation, in form of gas glow discharge, and ion beam acceleration. To achieve a significant charge transport by the ion beam, the electron transport parallel to the thruster axis is magnetically suppressed by an $\mathbf{E} \times \mathbf{B}$ situation. This allows maintaining of quasi-neutral plasma where a dominant portion of the charge transport across the magnetic field lines is done by the accelerated ion beam.

In a HEMP thruster we find a quasi-periodic sequence of $\mathbf{E} \times \mathbf{B}$ field situations along the thruster axis at each magnetic cusp. The higher radial magnetic fields at the cusps compared to a HET reduce the overlapping of ionisation and acceleration regions along the thruster axis and result in more discrete energy velocity spectrums of the ion beam and significantly reduced thermal losses. For comparison, the thermal losses in a HET are typically 30% to 45%, while in a HEMP thruster only 10% to 20%. As a second consequence of the magnetic plasma confinement in the multicusp magnetic field structure, the plasma to ceramic wall contact is reduced and channel wall erosion is prevented, thus improving life and reliability.

Figure 6.16 sketches the cross-section of a cylinder-symmetrical HEMP thruster with 3 permanent magnet rings and three magnetic cusps located inside the thruster acceleration channel. It also shows typical, self-consistently adjusting average plasma potential along the thruster axis. The number of magnetic cusps can be chosen as appropriate for a selected application. As a general rule one can say that the higher the number of magnetic cusps, the higher the applicable maximum anode voltage and also the efficiency.

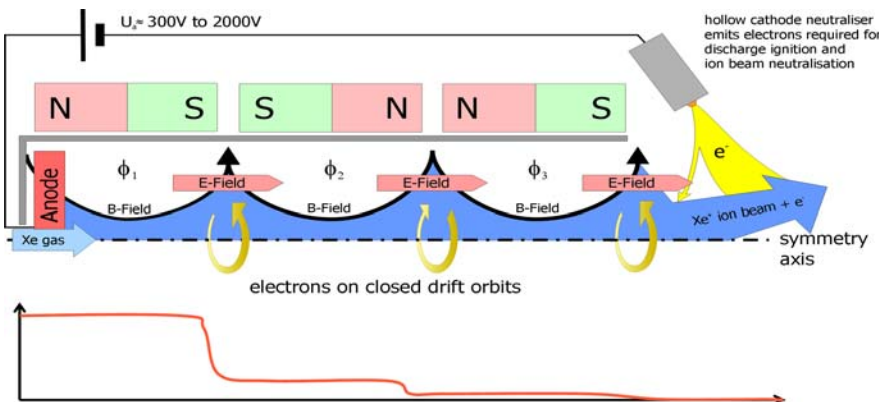


Fig. 6.16. Functional cross section of a HEMP thruster and plasma potential distribution along the thruster axis

The HEMP thruster consists of the following main elements:

- A cylindrical ceramic tube, surrounded by a
- PPM stack of magnet rings.

On the bottom of the ceramic

- A HV feedthrough to the anode electrode.
- A Xe propellant gas supply.
- A grounded hollow-cathode neutraliser located off axis at the exit side of the ceramic tube.
- An anode power supply provides the potential difference with respect to the grounded neutraliser required for ion beam acceleration.
- The neutraliser needs for its own operation an additional heater and keeper voltage supply and a Xe feed line. Since these elements are similarly required for most standard types of electric thrusters, they are not explicitly shown here.

The HEMP thruster operation principle is explained as follows. Low energy neutraliser electrons are confined by the magnetic exit cusp and maintain there a low, close to ground plasma potential at the thruster exit. Eased by collisions, a few of those primary electrons find their way from the neutraliser through the exit cusp into the discharge channel. Those electrons are trapped in spirals along their bent magnetic flux lines (a single set of those is sketched in Fig. 6.16 as bent arrows from the North to the South poles of the PPM-system). The electrons are easily movable along the thruster axis where the magnet field is parallel. There the plasma potential is almost constant. At the cusps, the electrons are effectively hindered to move across the radial magnetic field lines. Furthermore, the strong magnetic field gradient in radial direction reflects the electrons as a magnetic mirror. Thus, bouncing forth and back within the magnetic cells, the electrons can travel long ways without any wall collision and have an increased probability to ionise neutral Xe atoms at equal cell potential. Again all type of collisions in the cusp plane allow electrons to cross the cusp and to enter the next magnet cell on higher electrostatic voltage, where neutral gas ionisation takes place at an increased rate due to the higher electron and neutral gas density. Globally speaking, the electron and thus plasma density is increasing from cell to cell the closer the cell is positioned to the anode.

Considering individual electron trajectories and naming radial outward and inward parts of the electron trajectories "spokes", we find subsequent spokes rotating around the axis as a result of the $\mathbf{E} \times \mathbf{B}$ drift in a cusp. On average, the electrons moving towards the anode produce an azimuthal Hall current close to the opposite Hall current of the reflected electrons. Therefore, the net Hall current is much lower and is produced by those electrons which can cross the cusp due to elastic, inelastic or ionisation collisions. A 3-D simulation of this complex electron movement was performed with the Kobra 3D trajectory code [81] and is shown in Fig. 6.17.

As we note from Figs. 6.16 and 6.17, the plasma electrons are captured on the right, low potential side of a cusp. The small width of a gap with low electron density on the left of each cusp, is adjusting self-consistently according to the ion current to be transported across this potential drop. In other words, it seems that the device is

regulating with this gap width its perveance as required to transport and accelerate all the ions produced in the left magnetic cell, across the cusp, to the next cell on the

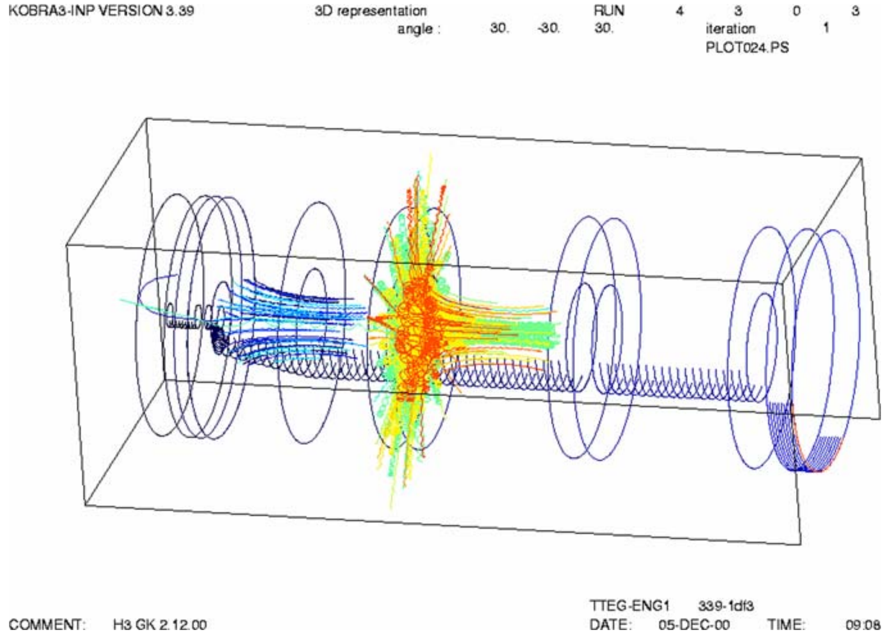


Fig. 6.17. Trajectories of electrons bounced forth and back at the radial magnetic mirror in a cusp and the axial magnetic mirror at the axial field maximum between the cusps. The trajectories are forming rotating trajectory "spokes" in the cusp plane

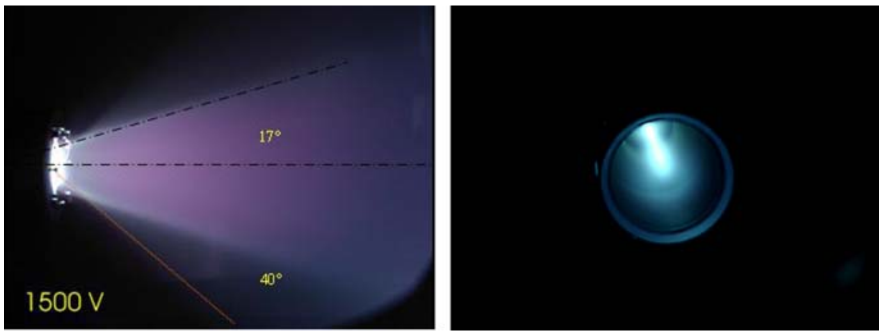


Fig. 6.18. *Left*: Ejection of two conical hollow ion beams with 17° and 40° beam angle, respectively. The ions originate from different magnetic cells at higher and lower potential inside the thruster, respectively. The thrust effective, average beam angles is 28° at 1500 V. *Right*: Front view look into the thruster acceleration channel. The confinement of the plasma and ion beam close to the axis explains the erosion free operation and the low wall losses of the HEMP thruster. The pictures are taken on the Thales HEMP 3050 DM7 thruster

right. Due to the high ionisation probability, the total ion current depends essentially, over a wide range, only on the propellant gas flow.

The residual radial kinetic energy component of the beam ions becomes apparent when the beam exits as a conical hollow beam. This and the confinement of the plasma close to the thruster axis is shown by the side and front view photos of Fig. 6.18, respectively.

6.4.3 Design and Performance

The design and performance of the cylindrical HEMP thruster, due to its rather short development time, has not yet achieved its final, flight hardware status but competes already, due to its unique characteristics, with the standard thrusters. The design of the recent BB HEMP 3050 DM7 thruster model is shown in Fig. 6.19. Its performance characterisation was obtained during a measurement campaign at the University of Giessen thrust stand. Since the only thruster supplies are anode voltage and neutral Xe flow, we report the performance parameters such as discharge current, thrust, specific impulse and power to thrust ratio over a wide range of those.

Figure 6.20 shows the discharge current. It is almost independent on the applied anode voltage and only a function of the Xe flow. If all supplied Xe atoms would be single charge ionised, a maximum discharge current of 70 mA/1 sccm Xe flow would be possible. We note that at small Xe supplies <15 sccm this limit of full ionisation is not achieved, whereas at high Xe flows the current is due to double and higher ionisation even beyond that level. On the other hand, when at high Xe flow the discharge current is due to multi-charged ions above this limit we still can have a small amount of neutrals exiting the thruster.

The ion beam current almost equals the anode current, since the amount of primary neutraliser electrons contributing to the anode current compared to the secondary electrons remains in a low of a few percent range. The thrust produced by

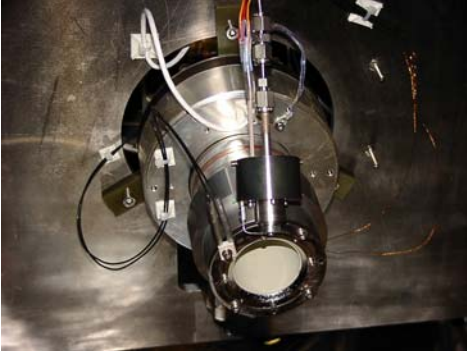


Fig. 6.19. Front view of the cylindrical HEMP 3050 DM7 thruster and its ceramic acceleration channel. The channel exit diameter is 4 cm. Around the channel the PPM – stack and above the channel the Thales hollow-cathode neutraliser is located. At the backside, the thruster is mounted on a cooling flange fixed to the University of Giessen thrust stand

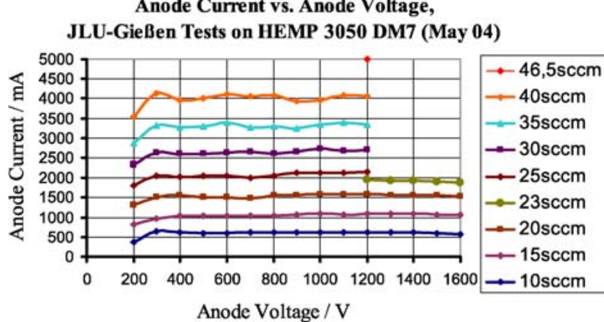


Fig. 6.20. Anode discharge current as a function of anode voltage and Xe flow (1 sccm corresponds to a Xe mass flow of approximately 0.1 mg/s

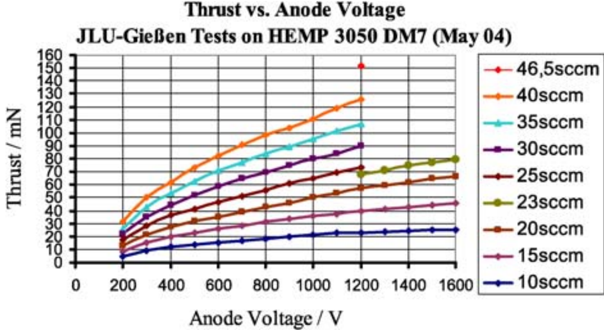


Fig. 6.21. Thrust as a function of anode voltage and Xe flow. The edge at 1200 V is caused by a power supply limitation

this amount of accelerated ions depends, as expected, square root like on the accelerating anode voltage (see Fig. 6.21). The dynamic thrust range was spanning in these tests from 152 mN down to 4.5 mN. In a separate continuous throttling test, reducing at 200 V the Xe flow below 2.8 sccm, thrust levels down to about 200 μN could be obtained. This extreme throttle-ability of the HEMP thruster is very unique for a gas driven thruster. Besides the thrust T, the specific impulse I_{sp} and the power to thrust ratio PTTR are often used to characterise the mass efficiency and the electric thrust efficiency of a thruster, respectively.

Figure 6.22 shows the specific impulse in sec (impulse per propellant weight) multiplied with the gravitational constant 9.81 m/sec^2; one obtains the impulse per propellant mass or the average axial exhaust velocity of the ions. The fact that the curves are not all on the same ideal universal curve for single ionised Xe is related to a non-ideal ionisation efficiency, non-parallel exhaust of ions (proportional to $\cos \alpha_{\mathrm{eff}}$), the effect that a portion of the ions is generated at a lower than applied anode voltage and the contribution of multiple charged ions.

Similar effects increase also the power to thrust ratio PTTR above the ideal situation. Figure 6.23 shows the respective curves.

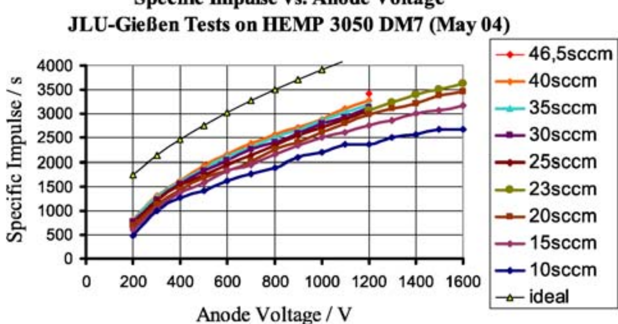


Fig. 6.22. The specific impulse Isp as function of anode voltage and Xe flow

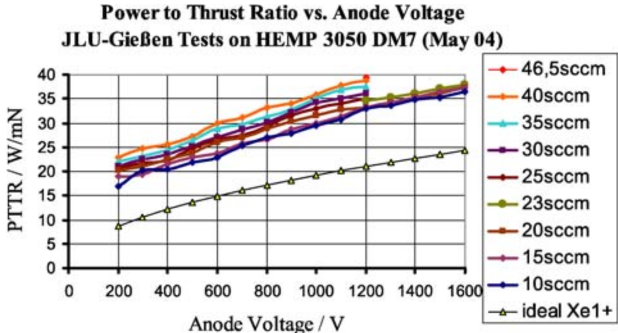


Fig. 6.23. The power to thrust ratio PTTR as a function of anode voltage and Xe flow

In the thruster community, the so-called total efficiency is often used as a figure of merit for the thruster operation. It is defined as $\eta_{\mathrm{tot}} = T^2/(2QP_{\mathrm{a}})$. Here, T is the thrust, Q is the propellant mass flow per second and P_{a} is the anode power. It can be rewritten in terms of ionisation efficiency η_{ion}, beam power efficiency η_{beam} and the effective ion beam angle α_{eff} as follows:

$$\eta_{\mathrm{tot}} = T/(2 \cdot Q \cdot P_{\mathrm{a}}) = \eta_{\mathrm{ion}} \cdot \eta_{\mathrm{beam}} \cdot \cos^2 \alpha_{\mathrm{eff}}. \tag{6.1}$$

Though not usually measured, the new parameters can easily be interpreted as the ratio of ions to neutrals exiting the thruster, the transfer ratio of electric power into kinetic beam power and the quadratic effect of the thrust effective ion beam angle. In [82] a method is described to determine these values by means of an angular resolving thermal target impact diagnostic.

Figure 6.24 shows the total efficiency. We note that it generally improves with higher anode voltage. The dominant reason above 400 V is the decreasing effective ion beam angle. Below 400 V, additional impacts from the ionisation efficiency and the beam power efficiency become apparent.

The HEMP thruster total anode efficiency can be compared with the HET values, especially if the small diameter geometry with its high surface to volume ratio is

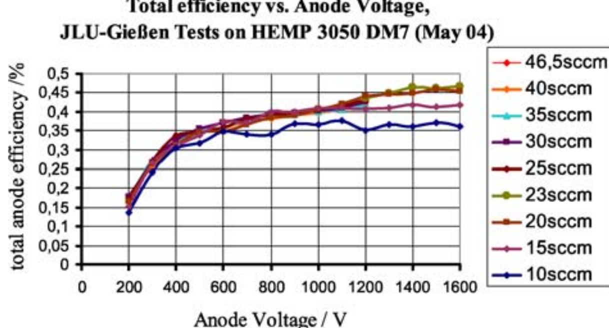


Fig. 6.24. The total efficiency as a function of anode voltage and Xe flow

Table 6.5. Comparison of typical thruster efficiencies

Type	η_{ion}	η_{beam}	$\cos^2 \alpha_{\text{eff}}$	η_{tot}
FEEP [82]	0.35	0.6	0.64	0.13
GIT (RIT 10) [83]	0.85	0.75	0.9	0.57
HET (PPS 1350/SPT100) [84]	0.85	0.66	0.83	0.47
HEMP (3050 DM7) [80]	0.75	0.82	0.75	0.46

taken into account. One of the interesting features of the HEMP thruster is its high beam power efficiency which results in a low thermal load of the satellite structure. Table 6.5 shows a comparison of the different impacts on total efficiency for the major thruster types.

The figures for the FEEP thruster are rather ambiguous since the ionisation efficiency can vary between 0.1 to 1 depending on the anode voltage and the emission currents. At higher thrust values the ionisation efficiency becomes very low due to droplet formation. Also the beam angle is given by the geometrical aperture screening the beam. For the GIT, the electrical power needed for the RF source for ionisation is included in the beam power efficiency.

Recent limited life test on a HEMP 3050 thruster has confirmed its erosion free operation at 57 mN and 2625 s over 250 h which suggests a worst case thruster life of 16 000 h corresponding to a worst case total impulse of 3.25×10^6 Ns.

6.4.4 Future Developments and Potential Applications

According to Table 6.5, the obvious development goals for the cylindrical HEMP thruster are the improvement of all efficiency contributors, especially the effective ion beam angle. Short term, for the low power HEMP 3050 (30 indicates to 3000 s, 50 to 50 mN), flight hardware total efficiency around 55% is expected. Computer-aided development using more refined plasma simulation codes should result in long term efficiency around 70%. For the new development of a high thrust, coaxial HEMP 30250 thruster short term total efficiency of about 60% to 65% is expected.

Generally, it is assumed that performance of HETs with respect to ionisation efficiency and effective beam angle can be exceeded, while maintaining the HEMP thruster superior beam power efficiency. A space qualification program for the HEMP 3050 thrusters is planned to be started by the German Space Agency DLR in 2007 with the goal of an Ion Orbit Demonstration of 4 fix mounted HEMP thrusters delivered by Thales on board the ESA ARTES 11 Small Geostationary Comunication Satellite (SGEO) to be built by OHB Systems and launched in 2011.

Since propellant and power supply requirements of the HEMP thruster are overlapping with both GITs and HETs, it fits into there respective satellite interfaces. Based on its extreme throttle-ability, all types of commercial geostationary satellites needing orbit raising, orbit topping and/or N/S station control are the target applications of the new thruster type. Also, high thrust and high specific impulse applications above 1 N, as required for future interplanetary missions are ideally suited for the coaxial HEMP thruster due to its compact and very high thrust density design.

Even μ-thrust applications down into the 1 μN range, as required for some scientific missions, are feasible with a scaled down cylindrical HEMP thruster. Due to there inherent high reliability, this might become an interesting future solution.

References

[1] H. Gray, P. Smith, D.G. Fearn, UK ion propulsion system development, J. Br. Interplanet. Soc. **49**, 163–172 (1996)

[2] H. Bassner, R. Killinger, H. Leiter, J. Mueller, Development steps of the RF-ion thruster RIT, IEPC-01-105 (2001)

[3] H.W. Loeb, K.H. Schartner, D. Feili, H.P. Harmann, H. Bassner, R. Killinger, H. Leiter, J. Mueller, Recent investigations, development and industrial applications of the RF-ion thruster in Germany, IAC-02-S.4.06 (2002)

[4] H. Bassner, H.P. Berg, R. Kukies, J. Mueller, Flight test and ground investigation results of the RITA experiment on EURECA, AIAA-94-2848 (1994)

[5] R. Killinger, H. Bassner, G. Kienlein, J. Mueller, Electric propulsion system for Artemis, AIAA-99-2271 (1999)

[6] R. Killinger, R. Kukies, M. Surauer, G. Saccoccia, H. Gray, Final report on the Artemis salvage mission using electric propulsion, AIAA-2003-4546 (2003)

[7] R. Killinger, H. Bassner, J. Mueller, Development of a high performance RF-ion thruster, AIAA-99-2445 (1999)

[8] H. Bassner, R.A. Bond, V.K. Thompson, K. Groh, The development of the ESA-XX ion thruster, IEPC-97-016 (1997)

[9] V.E. Krohn, Liquid metal droplets for heavy particle propulsion, in *ARS Electrostatic Propulsion Conference*, Monterey, CA, vol. 3–4, 1960

[10] J.F. Mahoney, et al., Electrodynamic ion source, J. Appl. Phys. **40**, 5101–5106 (1969)

[11] C. Bartoli, H. von Rohden, S.P. Thompson, J. Blommers, A liquid cesium field ion source for space propulsion, J. Phys. D: Appl. Phys. **17**, 2437–2483 (1984)

[12] K.L. Aitken, P.D. Prewett, Field emission electric propulsion: Basic investigations of the ion emitting region, Final Report on ESTEC-Contract 4462/80, 1983

[13] J. Mitterauer, Spectroscopic investigations of a field emission generated radiative zone; Mass spectroscopic measurements, Final Report on ESTEC-Contract 2734/76, 1981

[14] J. Mitterauer, Field emission electric propulsion: Spectroscopic investigations on slit emitters, Final Report on ESTEC-Contract 5051/82, 1986
[15] CENTROSPAZIO Pisa, Italy, 2003
[16] M.G. Nicholas, Propellants for field emission thrusters: A literature survey to identify alternatives to cesium, Final Report on ESTEC-Contract 3806/78, ESA Report Ref. CR(P) 1395, 1979
[17] P.D. Wilson, D. Stewart, Field emission electric propulsion: Additional propellants: Further tests on indium and alloys, Final Report on ESTEC Contract 4793/81, ESA Report Ref. CR(P) 1767, 1983
[18] M. Fehringer, F. Rüdenauer, W. Steiger, A. Genovese, M. Tajmar, FEEP microthruster components development; Ultra-precise indium thruster, Final Report on ESTEC-Contract 12376/97, 2000
[19] R. Schmidt, H. Arends, A. Petersen, F. Rüdenauer, M. Fehringer, B.T. Narheim, R. Svenes, K. Kernveen, K. Tsuruda, T. Mukai, M. Nakamura, H. Hayakawa, Results from the active spacecraft control on GEOTAIL, J. Geophys. Res. **100**, 17.253–17.259 (1995)
[20] F.G. Rüdenauer, W. Riedler, V. Berzhatyi, M. Fehringer, E. Göschl, C. Kropiunig, L. Neznamova, W. Steiger, W. Torkar, LOGION: Operation of a liquid metal ion emitter module under microgravity, ESA-ISY-4, 121–125, 1992
[21] J. Melngailis, Focused ion beam lithography, Nucl. Instrum. Methods Phys. Res. B **80/81**, 1271–1279 (1993)
[22] P.D. Prewett, G.L.R. Mair, *Focused Ion Beams from Liquid Metal Ion Sources* (Research Studies Press, Taunton, 1991)
[23] J. Mitterauer, Field emission electric propulsion: Emission site distribution of slit emitters, IEEE Trans. Plasma Sci. **PS-15**, 593–598 (1987)
[24] ESA, Paris, France, Fundamental Physics Missions, homepage. http://www.rssa.esa.int/index.php?project=sp
[25] NASA Jet Propulsion Lab, Pasadena, CA, LISA homepage. http://lisa.jpl.nasa.gov/
[26] ONERA, Palaiseau, France MICROSCOPE homepage. http://www.onera.fr/microscope/mission.htm
[27] ESA, Paris, France, HYPER homepage. http://sci.esa.int/science-e/www/area/index.cfm?fareaid=46
[28] ESA, Paris, France, GAIA homepage. http://sci.esa.int/science-e/www/area/index.cfm?fareaid=26
[29] Rutherford Appleton Lab, Chilton, Didcot, DARWIN homepage. http://ast.star.rl.ac.uk/darwin/
[30] S. Marcuccio, A. Genovese, M. Andrenucci, Experimental performance of field emission microthrusters, J. Prop. Power **14**, 774–781 (1998)
[31] M. Fehringer, F. Rüdenauer, W. Steiger, Space-proven indium liquid metal field ion emitters for ion microthruster applications, AIAA Paper 97-3057, in *33rd AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, Seattle, WA, USA, July 6–9, 1997
[32] ARC Research (2003), Seibersdorf, Austria. http://www.arcs.ac.at/W/WAS
[33] M. Fehringer, F. Rüdenauer, W. Steiger, Ultra-precise indium thruster, Report to ESA, ESTEC/12376/97/NL/PA, WP 1000-3000, 1999
[34] A. Genovese, M. Tajmar, private communication, 2003
[35] F.G. Rüdenauer, H.M. Fehringer, R. Schmidt, H. Arends, Operation of liquid metal field ion emitters under microgravity, ESA J. **17**(2), 147–153 (1993)
[36] R. Schmidt, H. Arends, A. Pedersen, F. Rüdenauer, M. Fehringer, B.T. Narheim, R. Svenes, K. Kvernsveen, K. Tsuruda, T. Mukai, M. Nakamura, H. Hayakawa, Results from the active spacecraft potential control on GEOTAIL, J. Geophys. Res. **100**, 17.253–17.259 (1995)

[37] W. Riedler, F.G. Rüdenauer, P. Beck, V. Berzhatyi, M. Fehringer, R. Finsterbusch, L. Neznamova, R. Pammer, F. Pürstl, W. Steiger, MIGMAS/A: Test of a scanning ion microscope onboard the Soviet space station MIR, ESA-ISY-4, 127–131, 1992
[38] J. Kissel, H. Höfner, G. Haerendel, P. Parigger, S. Czempiel, J. Eibl, H. Henkel, A. Glasmachers, K. Torkar, F. Rüdenauer, F.R. Krueger, E.K. Jeßberger, E. Grün, N.G. Utterback, R. Thomas, Y. Langevin, H. von Hoerner, J. Silén, K. Hornung, The COSIMA team (2003) COSIMA, A high resolution time of flight spectrometer for secondary ion mass spectroscopy of cometary dust particles (to be published)
[39] I. Brodie, C.A. Spindt, Vacuum microelectronics, Adv. Electron Phys. **83**, 1–106 (1992)
[40] J. Mitterauer, Prospects of liquid metal ion thrusters for electric propulsion, IEPC Paper 91-105, in *22nd AIDAA/AIAA/DLGR/JSASS International Electric Propulsion Conference*, Viareggio, Italy, October 14–17, 1991
[41] A. Genovese, W. Steiger, M. Tajmar, Indium FEEP microthruster: Experimental characterization in the 1–100 μN range, AIAA Paper 2001-3788, in *37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, Salt Lake City, UT, USA, July 8–11, 2001
[42] J. Mitterauer, Indium: An alternative propellant for FEEP-thrusters, AIAA Paper 2001-3792, in *37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, Salt Lake City, UT, USA, July 8–11, 2001
[43] J. Mitterauer, Micropropulsion for small spacecraft: a new challenge for field effect electric propulsion and microstructured liquid metal ion sources, Surf. Interface Anal. **36**, 380–386 (2004)
[44] H. Löb, *Nuclear Engineering for Satellites and Rockets* (Verlag Karl Thieme KG, München, 1970), p. 305
[45] V. Kim, et al., Electric propulsion activity in Russia. Paper presented at the 27th international electric propulsion conference, Pasadena, US, 2001
[46] J.D. Filliben, Electric thruster systems, Chemical Propulsion Information Agency, Report CPTR 97-65, 1997
[47] R.J. Jahn, E.Y. Choueiri, Electric propulsion, in *Encyclopedia of Physical Science and Technology*, vol. 5, 3rd edn (2002), p. 125
[48] A. Cadiou, et al., An overview of the CNES electric propulsion program. Paper presented at the 28th international electric propulsion conference, Toulouse, France, 2003
[49] L. Garrigues, J.P. Boeuf, L.C. Pitchford, Results from a quasi-neutral steady-state hybrid 1D model. http://cpat.ups-tlse.fr/operations/operation_03/POSTERS/SPT/
[50] A. Cadiou, et al., An overview of the CNES electric propulsion program. Paper presented at the 27th international electric propulsion conference, Pasadena, US, 2001
[51] R.R. Hofer, R.S. Jankovsky, The influence of current density and magnetic field topography in optimizing the performance, divergence and plasma oscillations of high specific impulse Hall thrusters. Paper presented at the 28th international electric propulsion conference, Toulouse, France, 2003
[52] J.J. Szabo, J.E. Pollard, A laboratory scale Hall thruster. Paper presented at the 31st joint propulsion conference, San Diego, US, 1995
[53] R.R. Hofer, et al., Optimization of Hall thruster magnet field topography. Paper presented at the 27th IEEE conference on plasma science, New Orleans, US, 2000
[54] R.R. Hofer, et al., A high specific impulse two-stage Hall thruster with plasma lens focusing. Paper presented at the 27th international electric propulsion conference, Pasadena, US, 2001
[55] V. Garkusha, et al., Ion energy measurement of a D-55 Hall thruster. Paper presented at the 3rd international conference on spacecraft propulsion, Cannes, France, 2000
[56] E.Y. Choueiri, Plasma oscillations in Hall thrusters, Phys. Plasmas **8**(4), 1411 (2001)

[57] J.P. Boef, L. Garrigues, Low frequency oscillations in a stationary plasma thruster, J. Appl. Phys. **84**, 3541 (1998)
[58] F. Darnon, et al., Spontaneous oscillations in Hall thrusters, IEEE Trans. Plasma Sci. **27**, 98 (1999)
[59] W.A. Hargus, B. Pote, Examination of a Hall thruster start transient. Paper presented at the 38th joint propulsion conference, Indianapolis, US, 1995
[60] N. Gascon, M.A. Capelli, Wall effects on the excitation and propagation of instabilities in Hall thrusters. Paper presented at the 28th international electric propulsion conference, Toulouse, France, 2003
[61] S. Barall, et al., Longitudinal oscillations in Hall thrusters. Paper presented at the 4th international spacecraft propulsion conference, Chia Laguna, Italy, 2004
[62] D. Pagnon, et al., Control of the ceramic erosion by optical emission spectroscopy: results of PPS1350-G measurements. Paper presented at the 4th international spacecraft propulsion conference, Sardinia, Italy, 2004
[63] M.B. Belikov, et al., Operation pecularities of Hall thruster with power 1.5–2.0 kW at high discharge voltages. Paper presented at the 28th international electric propulsion conference, Toulouse, France, 2003
[64] P. Renaudin, et al., Magnetic circuits for Hall thrusters: use of permanent magnets. Paper presented at the 28th international electric propulsion conference, Toulouse, France, 2003
[65] M. Gascon, et al., Wall material effects in stationary plasma thrusters. I. Parametric studies of an SPT-100, Phys. Plasmas **10**(10), 4123 (2003)
[66] A. Warshavsky, et al., Wall Thermal modeling and measurements of an engineering model Hall thruster. Paper presented at the 28th international electric propulsion conference, Toulouse, France, 2003
[67] G. Saccoccia, ESA spacecraft propulsion activities. Paper presented at the 4th international spacecraft propulsion conference, Sardinia, Italy, 2004
[68] G.D. Racca, Experience and lessons learnt from SMART-1, the first European electric propulsion driven space probe. Paper presented at the 4th international spacecraft propulsion conference, Sardinia, Italy, 2004
[69] D. Estublier, et al., SMART-1 electric propulsion flight experience: main working group findings. Paper presented at the 4th international spacecraft propulsion conference, Sardinia, Italy, 2004
[70] Prioul, et al., PPS-1350 qualification status and performances. Paper presented at the 4th international spacecraft propulsion conference, Sardinia, Italy, 2004
[71] P. Garnero, ASTRA-1K and STENTOR plasma propulsion subsystem experience. Paper presented at the 39th joint propulsion conference, Huntsville, US, 2003
[72] M. Lysyk, P. Garnero, Electric propulsion system on @Bus Platform. Paper presented at the 4th international spacecraft propulsion conference, Sardinia, Italy, 2004
[73] J.W. Dunning, et al., NASA's electric propulsion program. Paper presented at the 27th international electric propulsion conference, Pasadena, US, 2001
[74] V. Khayms, et al., Overview of Hall current thruster integration activities at Lockheed Martin Space Systems company. Paper presented at the 27th International electric propulsion conference, Pasadena, US, 2001
[75] J. Fisher, et al., The development and qualification of a 4.5 kW Hall thruster propulsion system for GEO satellite applications – status update. Paper presented at the 28th international electric propulsion conference, Toulouse, France, 2003
[76] E.Y. Choueiri, A critical history of electric propulsion: The first 50 years (1906–1956), J. Propul. Power **20**(2) (2004)

[77] G. Kornfeld, H. Seidel, J. Wegener, Plasma accelerator arrangement, Patent PCT/DE99/01708 filed 11 June 1999, Priority: Germany No. 198 28 704.6, filed 26 June 1998
[78] G. Kornfeld, N. Koch, G. Coustou, First test results of the HEMP-thruster concept. Paper presented at the proceedings of the 28th IEPC, Toulouse, 2003
[79] G. Kornfeld, et al., The highly efficient multistage plasma (HEMP) thruster, a new electric propulsion concept derived from tube technology. Paper presented at the proceedings of the 4th international vacuum electronic conference, Seoul, May 2003
[80] G. Kornfeld, N. Koch, H.P. Harmann, New performance and reliability results of the Thales HEMP thruster. Paper presented at the proceedings of the 4th ISPC, Cagliari, 2004
[81] P. Spädtke, KOBRA3-INP User Manual, Extended DOS Version 3.39, INP (Ingenieurbüro für Naturwissenschaft und Programmentwicklung) Wiesbaden, Germany, July 11, 2000
[82] A. Genovese, N. Buldrini, M. Tajmar, W. Steiger, 2000 H endurance test of an indium FEEP cluster. Paper presented at the IEPC 2003, Toulouse 2003
[83] H.J. Leiter, et al., in *IEPC 2001*, Paper IEPC-01-104, Cannes, 2001
[84] P. Dumazert, et al., in *IEPC 2003*, Toulouse, 2003

7

Infrared, Light, Ultraviolet, Laser- and X-ray-Tubes

A. Ulrich, M. Born, H.W.P. Koops, H. Bluhm, and T. Jüstel

7.1 General Physics of Photon Generation

7.1.1 Historical Development

Vacuum electronic devices played a key role for our understanding of light and atoms. Besides the very early work on optical spectroscopy by Kirchhoff and Bunsen in 1860 [1] using flames, it was from the great variety of gas discharge experiments that we learned about spectral lines and their assignment to the various atoms and molecules. The practical importance of vacuum electronic devices for generating light will become obvious in the following chapters. Although the quantized nature of light is much more obvious in the discrete spectral lines emitted from low pressure gas discharge devices, the concept of the "photon" as the "quantum" of electromagnetic radiation was first introduced by Max Planck in connection with the precision of measurements of the so-called black body radiation in 1900 [2]. In this context Planck's formula (7.1) can be used for comparing other light sources with this black body radiation source at temperature T (see Fig. 7.1),

$$L_\lambda \, \mathrm{d}\lambda = \frac{2hc^2}{\lambda^5} \frac{1}{\mathrm{e}^{\frac{hc}{\lambda kT}} - 1} \, \mathrm{d}\lambda. \tag{7.1}$$

Here, L_λ is the power emitted per unit area, wavelength interval and solid angle. The index λ indicates that L_λ is given versus wavelength. The constants are: Planck's constant h, velocity of light in vacuum c, and the Boltzmann constant k. Studies of emission and absorption of light performed in the years following Planck's introduction of "light quanta" led to a powerful concept for the description of these processes and finally to a completely new insight into matter and its motion on a microscopic scale – quantum mechanics. It was found that both light and matter contains aspects of waves and particles. The link between these two aspects is described by a relation between wavelength λ or wave vector $\boldsymbol{k}$ ($\boldsymbol{k} = 2\pi/\lambda$) and frequency ν on one side and momentum p and energy E on the other side,

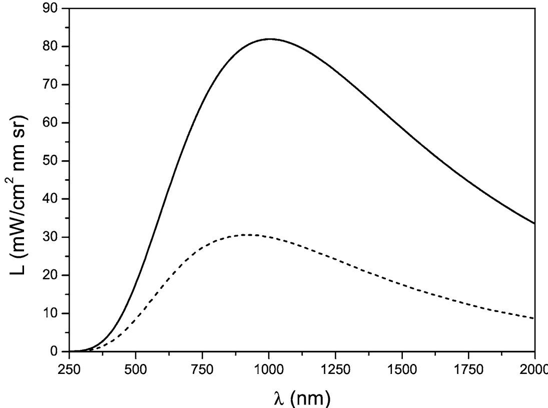


Fig. 7.1. Emission from a black body at a temperature of 2600 K (*solid line*). The emission from a tungsten strip calibration lamp (OSRAM, type Wi17g, No. 2892) with the same radiative temperature is shown for comparison. The difference is due to the finite emissivity of tungsten as measured by J.C. de Vos [3]

$$E = h\nu, \quad \vec{p} = \hbar \boldsymbol{k}. \tag{7.2}$$

The reader is asked to refer to publications about quantum mechanics for a deeper discussion of these fundamental aspects of light and matter. The important result obtained from quantum mechanics, in this context, is that bound systems, such as electrons orbiting around a nucleus, have well defined, discrete energy levels. Emission and absorption of light quanta (photons) of energy $E = h\nu$ is a possibility for the systems to make transitions from one energy level to another when the energy difference between the levels is the same as the photon energy E.

To a level of accuracy which is normally sufficient to represent all practical aspects of light sources, including lasers, the energy levels of atoms and molecules can be described by non-relativistic quantum mechanics. Coupling of these potentially light emitting or absorbing species with an electromagnetic radiation field can be described by the so-called Einstein coefficients A and B. Techniques for calculating the quantum mechanical eigenstates and energy eigenvalues of light emitting species are not discussed here. Energy levels of atoms are usually organized by main quantum numbers, quantum numbers of spin, and angular momentum. Total angular momentum is described by a quantum number J. For most practical applications energy levels of atoms can be found in tables [4, 5]. Phenomenologically, molecular levels are organized by electronic excitation (normally named X for the ground state and A, B, C, etc. for excited states), a vibrational quantum number v and a rotational quantum number I. The combination of electronic-, vibrational- and rotational energy can be calculated from parametrized descriptions of the molecular potential curves. The parameters for diatomic molecules can be found in tables [6].

Light emission from a light source is initialized by promoting atoms or molecules in the device from the ground state into an excited state, normally via collisions with

electrons. It should be noted that several eigenstates of an atom or molecule may have the same energy. An "energy level" may therefore consist of several "states". Such an energy level is said to be degenerated with a degeneracy g if g states have this particular energy. An excited species will have the tendency to return to the ground state releasing the excitation energy to its surroundings either by collisions with other species or by emitting a photon. It is the "art" in designing a light source to make photon emission in the desired wavelength region a dominant and efficient process in the device.

The process by which an undisturbed atom or molecule undergoes a transition from a level i to a level k via photon emission is called *spontaneous emission* and is governed by the Einstein coefficient A_{ik}. It has the unit 1/s and is a measure for the probability $\mathrm{d}P_{ik}$ of an atom to make the transition within a certain time interval $\mathrm{d}\tau$ ($\mathrm{d}P_{ik} = A_{ik}\,\mathrm{d}\tau$). It can therefore be used to calculate the time dependence of the number N_i of species in the upper level i starting from an initial number N_{i0},

$$N_i = N_{i0}\mathrm{e}^{-A_{ik}\tau}, \tag{7.3}$$

if photon emission is the only decay mode and the species decay independently of each other. The Einstein coefficient A_{ik} values can be calculated from basic principles in quantum mechanics. For practical applications one can refer to tabulated A_{ik} values [7]. The energy E (frequency ν, wavelength λ) of the photon emitted in this process is related to the energy difference between level i and k,

$$E = h\nu = hc/\lambda = E_i - E_k. \tag{7.4}$$

The energy level with higher energy E_i is often referred to as "upper" level and the level with lower energy E_k as "lower" level.

Resonant *absorption* is a process in which a photon is absorbed by an atom or molecule inducing a transition from a lower energy level k to a level i of higher energy. The photon energy E has to meet the requirement $E = E_i - E_k$. The probability for a given species in state k to be promoted to state i via absorption of a photon is proportional to the Einstein coefficient B_{ki} and the energy density $u(\nu)$ of the electromagnetic radiation field providing the photons which can be resonantly absorbed.

A third fundamental process is resonant interaction of a photon with an atom or molecule in the upper level i forcing it to the lower level k. It is called *stimulated emission*, and the radiation which is created is coherently superimposed on the radiation stimulating the transition. This means that the electromagnetic energy released in the transition travels in the same direction and is in phase with the incident radiation. Because of its similarity with the absorption process where radiation stimulates an "upward" (k to i) transition the process is described by a coefficient B_{ik} for the "downward" (i to k) transition. The transition probability is proportional to the coefficient B_{ik} and the energy density $u(\nu)$ of the radiation field.

In summary, the three processes for emission and absorption of photons are visualized in Fig. 7.2. To calculate the number of emission and absorption processes per unit volume and time interval for a specific transition, the density of species in the

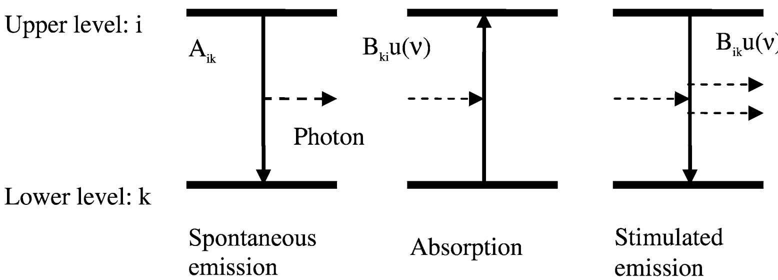


Fig. 7.2. Principle processes of light emission and absorption

upper and lower level N_i and N_k have to be known. These densities can be obtained from rate equations describing the rates at which species with energy E_i and E_k are populated and depopulated. The rates by which two selected level densities N_i and N_k change via photon emission and absorption can be obtained from the Einstein coefficients A_{ik}, B_{ik}, and B_{ki} which have been introduced above:

$$\mathrm{d}N_i/\mathrm{d}t = -A_{ik}N_i - B_{ik}N_i u(\nu) + B_{ki}N_k u(\nu)$$

and

$$\mathrm{d}N_k/\mathrm{d}t = A_{ik}N_i + B_{ik}N_i u(\nu) - B_{ki}N_k u(\nu). \tag{7.5}$$

The Einstein coefficients are related to each other in the following way:

$$g_i B_{ik} = g_k B_{ki} \text{ and } A_{ik} = (8\pi h\nu^3/c^3)B_{ik}. \tag{7.6}$$

So far a perfect resonance between the radiation field and the absorbing or emitting species has been assumed. In many cases, however, the detailed frequency dependence of the absorption-, emission-, and optical-gain in the narrow frequency region of a spectral line has to be taken into account. Various broadening effects such as Doppler- and pressure broadening influence the line shape. This can be described by a line shape function $g(\nu)$ which is normalized as

$$\int g(\nu)\,\mathrm{d}\nu = 1. \tag{7.7}$$

A beam of radiation with a specific frequency ν and intensity I (in units W/m^2) which interacts over a distance x with the atoms and molecules varies its intensity according to

$$I = I_0 \exp(-k(\nu)x). \tag{7.8}$$

The absorption coefficient $k(\nu)$ depends on the values defined above as

$$k(\nu) = \sigma(\nu)((g_i/g_k)N_k - N_i). \tag{7.9}$$

The absorption cross-section is given by

$$\sigma(\nu) = \frac{\lambda^2 A_{ik}}{8\pi\eta^2} g(\nu), \tag{7.10}$$

where η is the index of refraction of the medium. Note that relation (7.9) describes both absorption and stimulated emission. When the term $((g_i/g_k)N_k - N_i)$ becomes negative because $N_i > g_i/g_k N_k$, the intensity I will grow exponentially in the excited medium. This "**L**ight **A**mplification by **S**timulated **E**mission of **R**adiation" is the basic process of the device known as "LASER".

This brief introduction to the light emission and absorption processes allows important parameters, such as the power emitted from a lamp or the small signal gain of a gas laser tube, to be calculated.

The output characteristics of a discharge lamp could for example be estimated as follows. The emitted spectrum is given by the energy levels and corresponding transitions of the light emitting species. If absorption can be neglected, the power emitted per unit volume on a specific transition from level i to level k is

$$P/V = N_i h\nu A_{ik} \text{ with } h\nu = E_i - E_k. \tag{7.11}$$

If E_i, E_k, and A_{ik} are known from tables, the only remaining parameter which has to be determined is the density of species in the upper level i. This, however, may not always be an easy task. If N_i is not measured experimentally, a whole set of rate equations for all relevant collisional and radiative processes filling and depleting levels i and k have to be solved for obtaining a rather accurate prediction of N_i. Note that collisional processes between electrons and atoms, ions and atoms, etc. are studied as a field of its own [8]. Gain on an optical transition between two levels i and k in a gas laser tube can be predicted by a similar approach determining the population densities N_i and N_k and calculating the gain using (7.9) and (7.10).

7.2 Laser

7.2.1 Introduction

"Light Amplification by Stimulated Emission of Radiation" (LASER) was first observed by Maiman in 1960 using a flash lamp pumped ruby crystal as the laser medium [9]. Also, in that year Javan et. al. built the first gas laser with a discharge pumped low pressure He–Ne mixture as the laser medium and observed laser effect in the near infrared region [10]. Since that time one branch of laser development is based on vacuum electronic devices. The interesting history of the invention of lasers and masers was described by Charles Townes [11]. The variety of lasers which have been developed since its invention is overwhelming. Here, a short review of gas lasers which are of practical importance is given.

The general operation principle of light amplification in matter has been described in Sect. 7.1.1. The main prerequisite for getting a laser to operate is to achieve "population inversion" $(g_k/g_i)N_i > N_k$ (formula (7.9) in Sect. 7.1.1). The term inversion refers to the fact that at normal, thermal equilibrium conditions the population density N_k of the lower level is always higher than the population density N_i of

the energetically higher lying level. The "art" of laser design is to create conditions which are far away from thermal equilibrium in the sense that the N_i/N_k ratio is different from its value at thermal equilibrium, at least for two specific levels i and k. Optical feedback in the form of an optical resonator is another main ingredient for getting a gas laser operational, although some lasers like the 337 and 358 nm ultraviolet nitrogen lasers and so called soft X-ray lasers make use of light amplification in one pass through the device. Fabry–Perot type optical resonators consisting of two mirrors are normally used. Both flat and curved mirrors can form an optical cavity.

Balance between optical gain and losses in the resonator determines whether or not intensity will build up in a particular mode. Losses are due to absorption in all optical elements in the resonator, finite reflection of the laser mirrors and diffraction on the apertures of the device. Laser threshold is reached when the gain α is equal to the loss γ for light making one round trip in the resonator,

$$\alpha = \gamma.$$

Maximum small signal gain α_0 in the line center ν_0 is given by

$$\alpha_0 = \sigma(\nu_0)(N_i - (g_i/g_k)N_k).$$

The cross-section for stimulated emission $\sigma(\nu)$ is defined in (7.9) and (7.10) of Sect. 7.1.1.

Extraction of light from the laser resonator contributes to the losses of the resonator. Semitransparent laser mirrors are frequently used for extracting a certain fraction of the light intensity which builds up in the resonator during laser operation. If the amplification of light in the laser medium can be maintained over a period of time which is longer than the decay time of light in the resonator, the characteristics of the decoupled light beam are governed by the parameters of the optical resonator. A discussion of line shapes and line broadening effects can be found in the literature on lasers and atomic physics [12]. Detailed descriptions of optical gain in a laser medium and the modes of an optical resonator are, for example, given in [12] and [13].

Finding new laser media and laser transitions followed the invention of the laser. Over 6100 lines are listed [14, 15]. Compilations of original papers on gas laser technology and short wavelength lasers have been published [16, 17]. Only a few lasers are commercially successful.

7.2.2 Specific Laser Devices

A general advantage of gas lasers is that the optical properties of the laser medium are not permanently damaged by its operation. An overview over widely used gas lasers is given in Table 7.1. This allows high power levels in both pulsed and continuous wave (cw) lasers. However, a potential build up of impurities and overheating of the laser gas has to be avoided. The CO_2 laser in the infrared region at wavelengths around 10 µm is very efficient and often used in industrial processes for cutting and welding.

The *He–Ne laser* uses an energy resonance between helium atoms in the metastable state and neon atomic levels for achieving population inversion. Collisions of excited helium atoms with ground state neon atoms lead to selective excitation of these neon levels. The slight energy differences between the levels are compensated by the kinetic energy of thermal motion [12]. Note that the older Paschen notation is used in many early publications for the NeI levels. Racah notation is used here.

Helium–neon lasers use the positive column of a glow discharge as the laser medium. The tubes are made from glass with typical inner diameters of 1 to 2 mm. Large area cold cathodes are used. The cathode is often placed in an outer tube with

Table 7.1. Laser systems and their characteristics

Laser	Level transitions	Wavelength	Pumped by	Speciality/ output
He–Ne metastable states 7:1 He–Ne mixtures	$5s'[1/2]^0_1$–$3p[1/2]_1$ $5s'[1/2]^0_1$–$3p'[3/2]_2$ $4s'[1/2]^0_1$–$3p'[3/2]_2$	543.516 nm 632.816 nm 1.1529 μm	gas discharge dc-currents 10 mA to 50 mA; gas pressure 1 mbar	tube diameter 1–3 mm; output mW
Krypton ion	triply ionized	176 nm	low pressure discharge	
Argon ion	$3s^2p^5\ ^2P^0_{1/2}$ and $^2P^0_{3/2}$ 4p ($^2S^0$, $^2P^0$, $^2D^0$, $^4D^0$)–4s ($^2P_{1/2}$, $^2P_{3/2}$)	Vis to near UV 440 nm to 500 nm	two step process of forming and exciting the ions via electron collisions; discharge currents 30 A; pumps others	tens of Watts; output increases quadratically with discharge current
Excimer lasers Ar, Kr, Xe	second continua emission bands are unstructured and very broad, order of 10 nm	128 nm 150 nm 172 nm	systems pumped by intense electron beams	
ArF* KrF* XeCl		193 nm 249 nm 308 nm 100 ns up to 1 μs pulses	pulsed electrical discharges in laser tubes at 3 to 6 bar, 10 to 100 ns pulses	The three types cover 90% of market; repetition rates of 1 or 2 kHz

typically 3 cm diameter. Laser lines are selected by laser mirrors with wavelength dependent reflectivity. The total gain in a regular He–Ne laser tube is on the order of a few percent per pass. Reflectivity of one mirror is maximized (better than 99%), and the "output coupler" has a transmission of one or two percent. Typical light output power is a few milliwatts, and the overall efficiency of the device is between a per mill and a percent. Laser mirrors can be either mounted directly to the ends of the laser tube with bellows for alignment or as an external optical cavity. If the laser cell is closed by Brewster angle windows, the laser light will be linearly polarized, since Brewster windows have a low reflectivity and consequently a low loss only for one orientation of polarization. He–Ne lasers are used for alignment purposes and in metrology. Excellent spectral and spatial characteristics and stability has been achieved with helium–neon lasers, and it is still essentially the only compact, easy to use gas laser in the visible spectral range.

The great variety of collisional and radiative processes which can occur among many excited levels in neon or other rare gases leads to a great variety of laser lines [15, 16]. Some laser systems may operate continuously, others may lead to population inversion only for a short time interval following excitation by a pulsed discharge. Pulsed laser lines in neon (3p to 3s) have become important as model systems for very short wavelength lasers, often referred to as "X-ray lasers". Neon like ions are produced and excited by electron collisions in hot, laser induced plasmas [18]. With pulsed excitation there are two principle regimes for obtaining population inversion and laser effect, collisional excitation and recombination schemes. In the first case laser effect occurs during the pulse and in the latter in the afterglow by recombination processes.

Many ion laser lines can be observed in intense pulsed low pressure discharges. Triply ionized krypton at 176 nm delivers the shortest wavelength ion laser lines [14]. Soft X-ray wavelength emission was pioneered by Rocca and coworkers [19].

7.2.3 Ion Lasers

Ion lasers can also be operated in continuous mode. The best known ion laser systems use pure rare gases as the laser medium, argon in particular. In argon ion lasers singly ionized argon has the configuration . . . $3s^2p^5\ {}^2P^0_{1/2}$ and ${}^2P^0_{3/2}$. The two "ground state" levels differ in energy by only 0.18 eV. Inversion is built up because the lower 4s levels are more rapidly depleted than the upper 4p levels. The electron energy in the discharge has to be rather high for both ionization and excitation. The ionisation energy of argon is 15.759 eV and the 4p upper laser levels are again almost 20 eV above the ground state of the ion. Deviations from the quadratic increase of laser output with discharge current start above 600 A/cm^2 when formation of doubly charged ions sets in [14], p. 218.

Gas pressure and the diameter of the discharge tube of argon ion lasers are similar to those in He–Ne or other discharge pumped lasers operating on atomic transitions. Discharge currents, however, are about three orders of magnitude higher. Hot cathodes are used to provide high currents necessary for ion laser operation, and coaxial magnetic fields can improve laser performance. Ceramic tubes are used to handle

these high currents without too much sputtering. Active cooling by water or forced air has to be provided. The typical gain of ion laser lines is 1%/cm and the efficiency of the device is on the order of 0.1%. The discharge tubes are normally closed by Brewster angle windows and external optical resonators are installed. Many argon ion lasers provide the option of selecting the laser line by tuning an optical element such as a Littrow prism in the optical cavity. The argon ion laser with high output power and lines ranging from about 440 to 500 nm (near ultraviolet to green) is used to pump other lasers such as dye lasers. Finally, ion lasers are not limited to the use of rare gases as the active medium, but use metal vapors in rare gas buffer gases [15], p. 172.

7.2.4 Excimer Lasers

An interesting approach for obtaining population inversion in rare gases uses the fact that rare gas atoms can become chemically very reactive when they are not in their electronic ground state. This behaviour can be used to form molecules which exist only for a short period of time and then return to the ground state via radiative decay. The lower level of this radiative transition is instantly depleted since the species which formed the excited molecule are unbound in the ground state. Formation of the excited molecules, called “excimers” (an acronym for excited dimers), leads therefore automatically to population inversion, and lasers using this concept are known as excimer lasers. They were first demonstrated by Molchanov, Basov and coworkers in 1968 [20, 21] using electron beam excited rare gas crystals in which the excitation scheme is similar to dense gases. Excimer lasers in the gas phase were developed in the following years [22, 23]. Technical development of excimer lasers is still in progress. For basic concepts, see Charles Rhodes 1980 [24]. Excimer lasers which are widely used today are based on the formation of rare gas halide excimer molecules. From the possible combinations of He to Xe and F to I, the heavier rare gas with lighter halogen atoms show fluorescence and six combinations show laser effect [25]. The advantage of rare gas halide systems in comparison with pure rare gas excimers is their narrower emission line and the longer laser wavelength given above. Both features lead to reduced pumping power requirements.

The upper molecular level is populated via gas kinetic reactions, such as

$$\mathrm{Kr}^* + \mathrm{F}_2 \rightarrow \mathrm{KrF}^* + \mathrm{F}$$

and

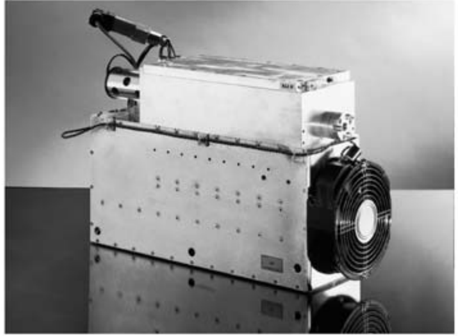
$$\mathrm{Kr}^+ + \mathrm{F}^- + \mathrm{Ar} \rightarrow \mathrm{KrF}^* + \mathrm{Ar}$$

with Ar used as a buffer gas. A general problem in excimer formation is to produce highly excited species like Kr^+ and Kr^* within a dense and rather cool gas in which molecules can form. Molecule formation can occur in the afterglow phase of a short, intense pulse of a high pressure glow discharge. Another technical difficulty in excimer laser design is to achieve a homogeneous excitation of the laser medium, since high pressure discharges tend to develop filamentation and arcing. Preionization of the laser gas with ultraviolet light is one solution of this problem. Excimer laser

transitions have lifetimes of a few nanoseconds and the electrically pumped lasers are operated using appropriate pulsed power circuits. In a short pulse regime 10 to 100 ns pulses are applied. Laser pulse duration is on the order of 10 ns in this case. In the so-called long pulse regime several 100 ns up to 1 μs pulses can be achieved. For a width of the discharge gap of 5 mm, 10 to 12 kV and, for a width of 15 mm, 25 to 30 kV are applied to drive the laser discharge. Discharge currents range from 0.1 to 1 kA. Repetition rates of 1 or 2 kHz are available from standard excimer lasers. High end devices for photolithography reach 4 kHz repetition rates with 30 to 40 W output power, time averaged.

A technical issue in the design of excimer laser tubes is to select materials which can resist the highly reactive halogen gases F_2 or Cl_2 and in particular the F and Cl radicals. Nickel and passivated aluminum are the preferred materials in excimer laser design. Other materials such as hydrocarbons have to be avoided.

Ceramic materials are used as insulators. Since the gas is chemically altered in the discharge channel, it has to be circulated in high repetition rate lasers to allow time for cooling and for the chemical components to return to their initial state, F_2 for example. A mechanical fan is used for that purpose. Its operation may contribute a significant fraction to the power consumption of the laser device. Laser tubes can provide typically 10^8 pulses before the gas has to be replaced, and 10^9 to several 10^9 pulses before the whole tube has to be at least overhauled by replacing the electrodes and optics. Optical gain per pass is very high during the short, tens of nanosecond pulses. Plane parallel resonators with a highly reflective back mirror and a highly transmissive (90%) output coupler are often used. The mirrors have to withstand an extremely high optical energy/power densities on the order of 100 mJ/cm^2 (10 ns pulses). A modern excimer laser tube is shown in Fig. 7.3. Excimer lasers have a wall plug efficiency of a few percent which is a high efficiency for short wavelength lasers.



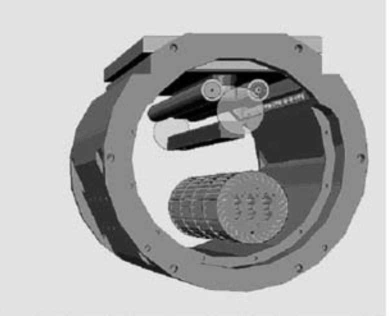


Fig. 7.3. Photograph of a compact, modern excimer laser (*left*). A cut through the laser tube is shown on the *right side*. The electrode configuration is shown in the *upper part* and the fan circulating the laser gas in the *lower part* of the drawing. Courtesy TuiLaser company, Munich, Germany

A present application which covers 70 to 80% of the market for excimer lasers is photolithography. There the lasers operate day and night with 5 to 10 mJ pulse power and 4 kHz repetition rate. The rest of the market is for micromechanics and medical applications, eye correction in particular. A number of 1000 to 10 000 shots are applied to each eye for that purpose.

7.2.5 Electron Beam Pumping of High Power KrF-laser

Because of their high efficiency, short wave length (248 nm) and high beam uniformity KrF-laser are presently the most attractive drivers for inertial fusion energy (IFE) [26]. To explore whether necessary laser beam power of several 100 TW and beam energy of several MJ can be generated with the required efficiency of more than 5% at affordable cost, programs have been started in several laboratories to develop the basic technologies [27].

High power KrF-lasers are pumped by large area high current pulsed electron beams. Because of their modularity, electron beam generators presently under development are considered to be scalable to the parameters required for a real IFE-driver. The main challenges are to develop a durable, efficient, and cost effective pulsed power system; a durable electron beam emitter; a long life, transparent pressure foil structure (that isolates the laser cell from the electron beam diode, the so-called hibachi); a recirculator to cool and quiet the laser gas between shots, and long life optical windows.

Typically, a single electron beam module must deliver 100 kA electron beam current with a particle energy around 500 keV to the laser cell for a duration of several 100 ns. Figure 7.4 shows a schematic of one half of the ELECTRA system presently constructed at NRL-Washington [27].

This system uses a capacitor/step-up transformer prime power system that pulse charges a pair of coaxial, water dielectric, pulse forming lines. The energy in the lines is then switched into the electron beam diode load using laser-triggered spark-gaps. A strong axial magnetic field is used to prevent pinching of the beam in the gas cell.

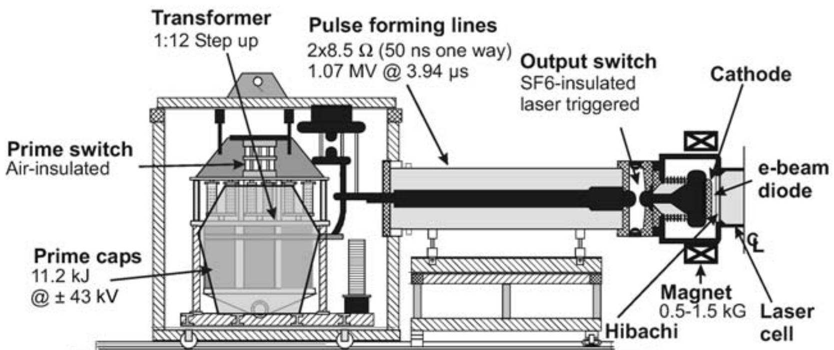


Fig. 7.4. One half of the ELECTRA pulsed power system: 500 kV, 100 kA, 100 ns flat-top electron beam, 5 Hz repetition capability

A general schematic of the diode, hibachi and gas cell is presented in Fig. 7.5. Besides the cathode, the hibachi foil support structure is the key element for a long-lived KrF-laser system. It must provide efficient and reliable electron beam injection into the gas. Plasma sources based on carbon fibres or metal dielectric structures look most promising as a durable electron source. To meet the high efficiencies required, the cathode emission must be restricted to areas between the ribs of the hibachi structure. In addition, the cathode emitter must be counter rotated by a fixed amount to compensate for the beam rotation in the axial magnetic guiding field.

Experiments have also shown that large area electron beams are subject to instabilities which can lead to energy spread and large divergence of the accelerated electrons [28]. To damp the RF oscillations, the cathode was built as a slow wave structure by loading the gaps between the segmented cathode areas with resistive elements. Long lifetimes (more than 10^8 pulses) of the hibachi foil are only possible if it is cooled efficiently. A recirculator used for that purpose does not only have to cool the foil and the laser gas but also has to quiet the gas before the next shot. A system based on louvers has been designed at NRL [29] to solve this task. It guides the gas stream towards the hibachi between shots and rapidly transfers the turbulent into a laminar flow during the pulse by opening the louvers.

Intense laser emission from electron-beam-pumped ternary mixtures of Ar, N_2, and POPOP vapor was investigated [30]. Powerful laser output was observed at 381 nm from electron-beam-pumped ternary mixtures of argon, nitrogen and 2,2′-p-phenylinebs<5-phenyloxazole (POPOP) dye vapor. The injection of the sub threshold 380.5 nm N_2-line into the electrically excited gain profile of POPOP vapor has been identified as a pumping mechanism.

The radiative energy transfer from the electron-beam-excited N_2 is apparently responsible for the excitation of the dye vapor.

The apparatus used consisted of a high-temperature high-pressure vapor cell with both electron-beam and optical access. The POPOP partial pressure in the cell was controlled by adjusting the temperature of the cell assembly with an oven enclosure. A prealigned plane-parallel optical resonator with high-reflectivity dielectric coat-

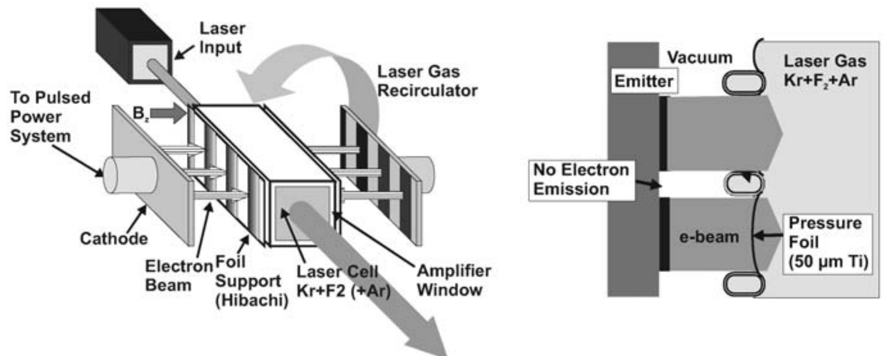


Fig. 7.5. Principal set-up of an electron beam pumped KrF-laser

ings ($R = 96\%$ at 385 ± 10 nm, $R = 70\%$ at 360 nm) was located inside the cell. The cell was isolated from the cathode of the electron-beam machine by a 0.075 mm thick titanium foil. This foil acted both as a pressure barrier and as the anode of the field-emission diode. A second 0.025 mm thick foil, located 0.3 cm away from the first one inside the cell, was used as a thermal barrier to prevent condensation of dye on the relatively cool anode foil. All experiments were performed on an S1 Apex I machine, configured for this experiment to produce a beam of 1-MeV electrons with a peak current of 28 kA in a triangular-shaped 30-nsec pulse, which corresponded to an input energy of 500 J. The output from the cell was measured with a spectrograph which had a 1 nm resolution. The output pulse was detected with a fast-vacuum photodiode and a Tektronix R7912 transient digitizer. Intense laser emission at 381 nm was observed in a 10 nsec pulse with less than 2 nm bandwidth and a 5 mrad beam divergence from an electron-beam-excited ternary mixture of 5 Torr POPOP, 2 atm Ar, and 4 atm N_2. Without optimization output power in excess of 500 kW was observed, indicating a conversion efficiency from deposited electronic energy into the active volume to optical output of at least 0.3%.

7.3 Smith–Purcell Effect

7.3.1 Historical Development

Tera-Hertz (THz) electromagnetic radiation is produced using an effect first explained by S.J. Smith and E.M. Purcell in 1953. The Smith–Purcell effect describes "Visible Light from Localized Surface Charges Moving Across a Grating". This effect occurs over a broad range of spectral wavelengths, including visible through millimetre wavelengths. The Smith–Purcell effect was the precursor of the free electron laser (FEL). Essentially, this is a form of Cherenkov radiation where the phase velocity of the light has been altered by the periodic grating [31].

7.3.2 Electrophysical Fundamentals

The electron charge of the beam induces an image charge in the metal grating. Due to the mechanical geometry of the grating bars and groves, this image charge oscillates and this generates an electromagnetic field as a standing wave above the grating. Since the electron is fast, this wave is coherent along the grating. The observed wavelength of the emitted light depends on the grating period D and the observation angle Θ with respect to the velocity vector of the electrons. This is given by

$$\lambda = D(\beta^{-1} - \cos\Theta), \tag{7.12}$$

where β is v/c, i.e. the electron velocity divided by the speed of light.

7.3.3 Present State

A strictly analytical approach explaining the Smith–Purcell effect is given by di Francia [32]. FEL systems were constructed by Walsh and his group using a conventional SEM setup [33]. The "Tabletop"-SEM FEL setup of Vermont Photonics shines an intense electron beam parallel to a mechanical grating of mm pitch. Changing the electron beam voltage allows continuous tuning of the THz output wavelength from 100 μm to 1 mm wavelength.

Present solutions for the problem are huge electron accelerators [34]. Presently worldwide 30 FEL installations using very high electron energies are installed, 10 are in development and further 10 are planned. However, the available beam time and lines number is not sufficient to exploit many applications of THz radiation. The most prominent characteristics of FEL are as follows.

Tunability by voltage and grating change. Using an undulator, a factor-of-10 tunable frequency range has already been demonstrated with the same accelerator, see Fig. 7.6. High peak power, since waste energy is carried away at nearly the speed of light and the lasing medium cannot be damaged by high optical fields. FELs can produce very high peak powers; gigawatt peak powers have been demonstrated. Flexible pulse structure using mature RF technology of linear accelerators to manipulate and control the FEL pulse structure. Picosecond pulses with sub-picosecond jitter can be produced at a rate of 10/sec. FELs have a good laser characteristics achieved by a single transverse mode, high spatial and temporal coherence, and flexible polarization properties.

Because the gain medium is transparent at all wavelengths, FELs in principle can produce radiation at any wavelength. In practice, electron beam energy, current, emittance, and energy spread requirements become more stringent as the wavelength decreases, and the cost, size, and complexity of the FEL are therefore higher at shorter

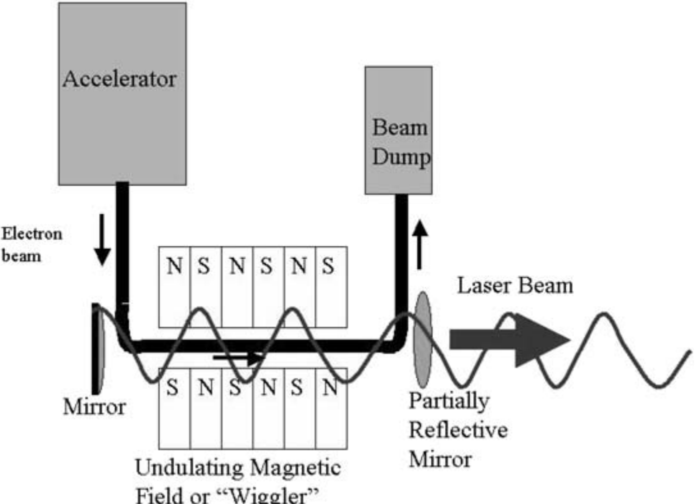


Fig. 7.6. Schematic of a FEL installation in nuclear physics or national institute

wavelengths. FELs can have significant emission at harmonics of the fundamental frequency given by the resonance condition. The shortest wavelength achieved so far in an FEL is 240 nm. FELs for scientific research have been restricted to the infrared region. Now, there are proposals to build vacuum ultraviolet and X-ray FELs. Size and cost of present FEL's used principally in central facilities are huge. Tabletop and portable powerful solutions are needed in the scientific and technical community.

7.3.4 Future Aspects for THZ Sources

Recent workshops require the development of a portable or small source for various applications. One approach is to employ a modified SEMs, as described by Walsh et al. [35, 36]. The power of the emitted light from a SP device can be calculated depending on the assumed beam shape with different formulas, which describe the linear power-gain regime [37, 38]. When achieving 1 to 218 μA currents in the beam, the exponential gain regime for the emission power is reached [39–41]. The wavelength regime from 8 μm to 1,400 mm can be covered with electron beam energies below 1000 eV. Tuning the electron energy, even X-ray generation is possible. The X-ray FEL is still under investigation since efficient S-P X-ray generation requires relativistic e-beams having a transverse momentum and dimension whose product approaches the Heisenberg uncertainty limit [42].

7.4 Millimetre and Infrared Light Sources

7.4.1 Historical Development

Millimetre and infrared light, and tera-hertz waves (THz) are electro-magnetic waves and cover a wavelength regime of 100 mm to 1 μm. Figure 7.7 gives a schematic of the dimensions of wavelength, frequency and energy for those radiations.

The propagation is governed by Maxwells equations. The interaction with matter is correlated to the weak energy of THz photons. Diffraction and/or guiding properties of devices in the THz regime are linked to the sub-mm THz wavelength, also called far infrared (FIR). Due to the period of THz waves in the picosecond regime, ultra-fast phenomena in matter can be investigated. Measurements use the time domain spectroscopy.

Infrared radiation is electromagnetic radiation having a wavelength in the range 0.000075–0.1 cm. Infrared rays thus occupy that part of the electromagnetic spectrum with a frequency less than that of visible light and greater than that of most radio waves, although there is some overlap. Infrared radiation is thermal (or heat) radiation. It was first discovered in 1800 by Sir William Herschel, who was attempting to determine the part of the visible spectrum with the minimum associated heat in connection with astronomical observations he was making. In 1847, A.H.L. Fizeau and J.B.L. Foucault showed that infrared radiation has the same properties as visible light, being reflected, refracted, and capable of forming an interference pattern. Infrared radiation is typically produced by objects whose temperature is above 10 K [43].

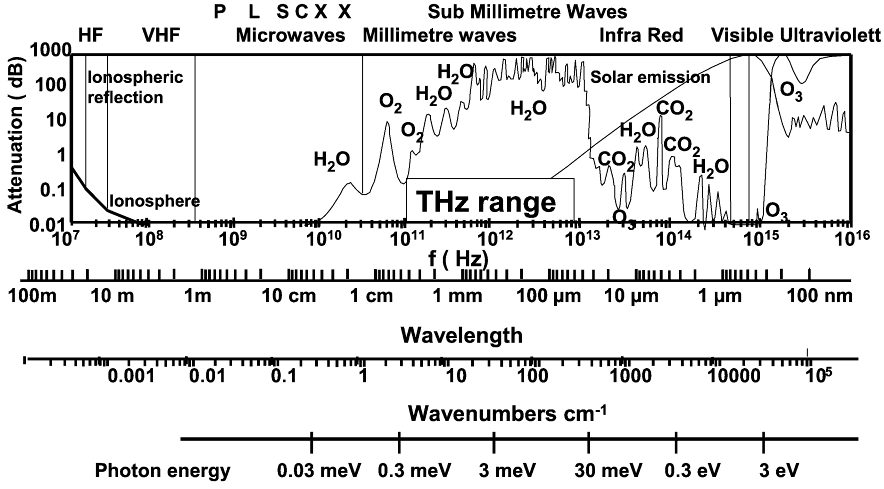


Fig. 7.7. Ranges of wavelength, frequency and energy for mm, THz, and IR radiations. Also, the attenuation in dB/km is given versus the wavelength

7.4.2 Generation of IR Radiation

Visible light is emitted from sources with a temperature of above 1900 K, as a candle does. The temperature of a human body emits radiation with a wavelength of 8 μm to 9 μm and can be observed only with an IR detecting camera or sensor. Typically, all hot surfaces emit IR radiation. Intensive IR lamps are incandescent lamps, which are generally used with a special filter to block out the visible spectrum.

7.4.3 IR Applications

There are many applications of infrared radiation. A number of these are analogous to similar uses of visible light. Thus, the spectrum of a substance in the infrared range can be used in chemical analysis. Radiation at discrete wavelengths in the infrared range is a characteristic of many molecules. The temperature of a distant object can also be determined by analysis of the infrared radiation from the object. Radiometers operating in the infrared range serve as the basis for many instruments, including heat-seeking devices in missiles and devices for spotting and photographing persons and objects in the dark or in fog [44]. Medical uses of infrared radiation range from the simple heat lamp to the technique of thermal imaging, or thermography. A thermograph of a person can show areas of the body where the temperature is much higher or lower than normal, thus indicating some medical problem. Thermography has also been used in industry and other applications. The higher sensitivity of longer-wavelength allows doctors to detect tumors using thermography, or heat analysis. It allows pilots to make better landings with improved night vision, and environmental scientists to monitor pollution and weather patterns [45]. Thermal

Fig. 7.8. *Left*: black and white image of a man's head in visible light, *left*, and the IR image, *right*. *Right*: airplane after takeoff with clearly shown jets and also hot tires of the wheels

insulation survey of buildings is performed. Other possible uses include law enforcement, search and rescue, and industrial process control [46].

In quantum detectors the THz photon excites an electron from a filled state into an empty state, for example, between an impurity state and the conduction band in a semiconductor (Ge:Ga, InSb, etc.) [47]. The sensitivity of such detectors is high: 10^{-12}–10^{-13} W/%Hz at 0.3 to 5 THz; however, they need cooling to a very low temperature (4 K or less), which is obtained by very small Stirling coolers, now commercially obtainable [48].

Figure 7.8 shows (left) the black and white image of a man's head in visible light, left, and the IR image, right. The photo on the right shows an airplane after takeoff with clearly shown jets and also hot tires of the wheels. A new infrared video camera with Infrared Focal Plane Array Technology uses quantum-well infrared photodetectors. The array is designed to detect infrared radiation of 8- to 10-μm wavelength. The camera is capable of detecting a temperature difference of 0.01°C.

7.4.4 Generation of THZ Radiation

The shortest wavelength achieved so far in an FEL is 240 nm. FELs for scientific research have been restricted to the infrared region. Now, there are proposals to build vacuum ultraviolet and X-ray FELs. Size and cost of present FEL's used principally in central facilities are huge. Tabletop and portable powerful solutions are needed in the scientific and technical community.

7.4.5 THz Applications

Applications are in the fields of spectroscopy and imaging. Radio astronomy employs remote space observatories in the FIR and THz regime with super-conducting detectors, cold mirrors and m^2 collecting aperture. Terahertz devices peer through paper and clothing, revealing concealed weapons and explosives.

Molecular vibrations in solids are used to measure spectroscopic fingerprints of molecules and compounds [49]. In biology especially unstained DNA analysis is the goal and theoretically predicted as a possibility [50]. Also, binding states can be detected and analysed. Also, single base mutations in chromosomes can be detected and help to distinguish inherited diseases. Medical applications are: tissue analysis [51], especially for cancer diagnostics in pathology [52], medical treatments without radiation damage of the surrounding tissue, imaging of human tooth for caries detection.

Other important applications are in defence and security, observation of status of materials, e.g. cracks in the foam insulation of the American space shuttles, detection of explosives and detonation chemicals, detection and analysis of drugs, screening of passengers for hidden metals, short range communications in combat, biological and chemical agent detection, toxic and chemical mapping and identification after bomb explosion.

Environment and materials applications to investigate surfaces and interfaces with THz dark field radar [53]. THZ radiation using eV to meV photons only warm up the samples a little, since the total output power of today's sources ranges in average 0.1 to 100 W DC. So far the technology lacks a portable and bright in a wide range tuneable THz-radiation source. Military and research agencies offer big money for a solution.

The analysis is simplified by a high chemical specificity. To characterize special compounds, an internationally accessible library of THz spectral signatures has to be built up. The fingerprint spectra of the compound contain rotational lines (gas); phonon bands (crystal); large amplitude collective modes (macromolecules). The presence of spectral signatures leads to non-subjective algorithms. The frequency regime for the THz fingerprint of matter is given in Fig. 7.9

For analysing the mechanical roughness of surfaces, a THz radiation source capable of emitting 300 μm radiation is sufficient and can image roughness above this wavelength, since the imaging system cannot see all the scattered noise from objects of the size of shorter wavelength.

THz screening is non-invasive, non-destructive, non-detectable, since THz is non-ionising. Very low power levels $\ll 1\ \mathrm{mW/cm^2}$ are for sensing are required. Many common, non-metallic container, shipping, and clothing materials have sufficient transmission for spectroscopy of their contents. Investigations have identified strong spectral features and fingerprints of explosives and drugs.

Preferred applications are: automated threat detection and personnel screening, molecular analysis of tissue, chemicals and fast screening of reactions without the need of a tracer or fluorescent addition to the reacting molecules.

Today one observes a rapid roll-off of the performance of electronic devices above 100 GHz. New more powerful sources are developed. For THz-radiation generation mostly negative differential conductance devices and transit time limited devices are employed, e.g. Gunn diodes, avalanche and transit time devices like Impatt diodes, and superlattice devices which use Bloch oscillations in a miniband and transit time limitation in the active region. Semiconductor heterostructure engineering with bandgap-, eigenstates-, and carrier-lifetime engineering are developed. Transis-

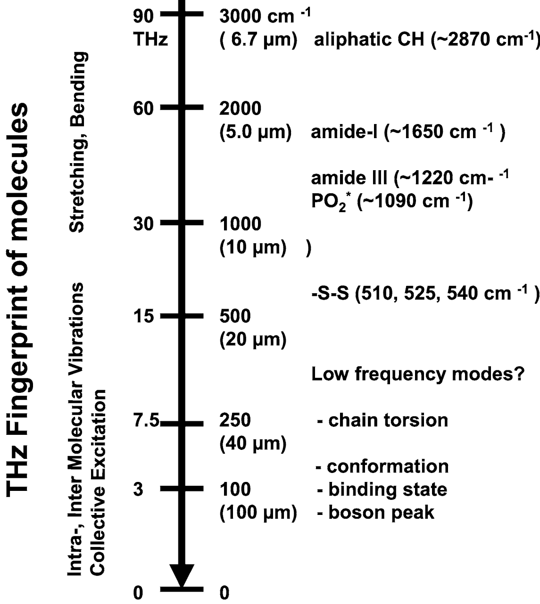


Fig. 7.9. Frequency regime for the THz fingerprint of matter with examples

tors with high mobility 2D electron gas n(HFET's) [54, 55], as well as transistors called n(HFET).

Quantum cascade lasers [56] generate THz radiation in the wavelength region 3.5 to 160 µm in pulse mode at RT and continuous mode at 77 K with tens of mW output power. Operation at room temperature is the challenge. Inter-subband transitions are used and the cascade scheme, where more phonons per electron are generated. A narrow linewidth of 4.6 THz is obtained.

Resonant tunnelling diodes are a low power source at 712 GHz, also Heterostructure Barrier Varactors.

Photomixing is employed for emitters and detectors. Powerful sources are miniaturized synchrotrons and beam transmission lines. The classic sources for THz-radiation generation are the large synchrotron and accelerator facilities which employ wigglers and Smith–Purcell grating sources and deliver mW to W output intensity. Those institutions presently offer their beam and experimental facility locations for measurements at a rate in the range of 1000 US$ per day to companies and research organisations.

7.4.6 Further Developments

To make THz spectroscopy, analysis and imaging a viable technology, however, a portable source is needed. Presently photonic sources are femto-second-laser based systems at cost 100–150 k$. Continuous wave diode laser sources cost half of this.

Wanted is a cheaper, more compact source, e.g. sub-GHz by a microelectronic approach, but will it be cheaper? Are quantum cascade lasers a solution?

7.4.7 Detection of THz Radiation

Several different detectors are in use: terahertz materials-metamaterials with negative refractive index and negative permittivity [57] and photonic bandgap materials. Others are: antenna-amplifier arrays, bolometers and optical 2 wave mixing for generation of photoconductivity in semiconductors, or using single electron transistors as detectors [58]. Those allow to detect 10^{-17} W/sqrt Hz at 500 GHz but at 70 mK.

7.5 Visible Light Sources

7.5.1 History and Introduction

The development of electrical visible light sources started with the invention and production of the incandescent lamps in 1880. Major improvements have been achieved over the past decades with respect to lamp efficiency, lifetime and colour properties, see Fig. 7.10. Further aspects like miniaturization and environmental aspects are some of the market drivers for the development of new and innovative products.

Incandescent lamps deliver thermal emission of radiation close to thermal equilibrium, low and high-pressure discharge lamps yield atomic and molecular emission from gas discharges, and LEDs emit light from solid state diodes. The latter type will not be addressed in this paper.

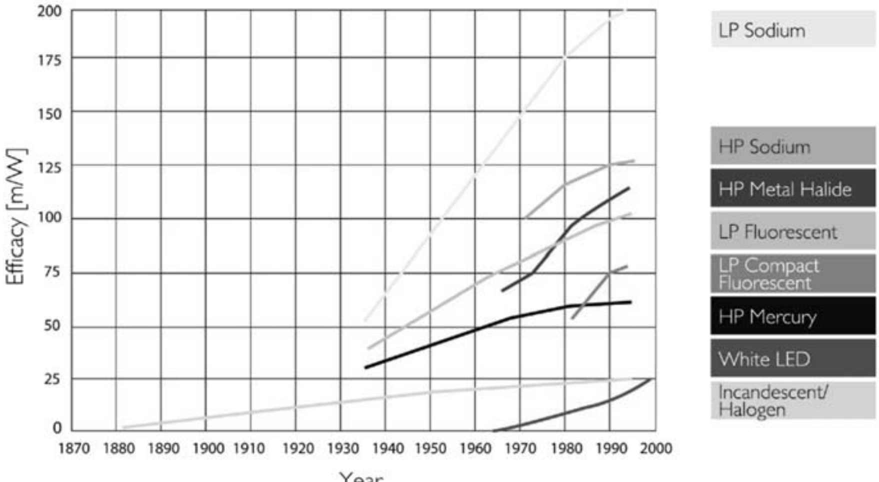


Fig. 7.10. Temporal development of luminous efficiencies of electrical light sources

7.5.2 Incandescent Lamps

The incandescent lamp is the oldest electrical light source with the largest market segment of about 11 billion pieces world-wide. It is used especially where small lumen packages are needed and where simple, cheap and compact lighting solutions are preferred [59].

Normal Incandescent Lamps

In incandescent lamps photons are produced by the electrical heating of a metal wire. Its high temperature results in radiation emission in the visible part of the spectrum.

According to Planck's law, the filament of an incandescent lamp must be heated up to at least 2400 K for production of a white emission spectrum. The metal wire is mounted in a glass bulb filled with an inert gas, as indicated in Fig. 7.11. Higher filament temperatures could render more efficient conversion of electrical energy into visible light, but would reduce lamp life due to enhanced metal evaporation. For this reason, the filaments of incandescent lamps are made of tungsten having a high melting point and a low vapor pressure. Lamp lifetime can be further improved by the reduction of the tungsten evaporation rate when adding rare gases such as krypton or xenon. Halogens are used for realizing a regenerative cycle in order to effectively transport back tungsten to the filament.

In normal incandescent lamps tungsten evaporates off the filament and condenses on the bulb wall, resulting in a reduced lumen output over lamp life. This is reduced when iodine, bromine or chlorine is added to the gas filling. The halogens are forming volatile tungsten compounds at the glass wall being transported back to the hot filament where the tungsten halides are decomposed again, see Fig. 7.11 right side. As a result of lowering the net evaporation rate of tungsten, the filament of halogen lamps can be operated at higher temperature as compared to standard incandescent

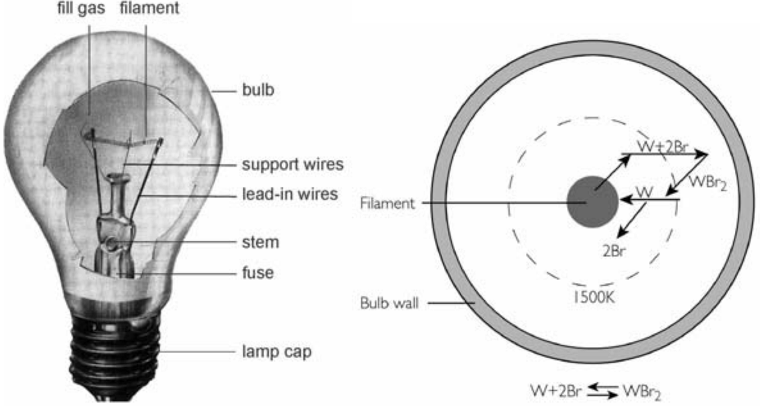


Fig. 7.11. *Left*: schematic drawing of an incandescent lamp. *Right*: schematic of chemical transport cycle in tungsten halogen lamps

lamps. Consequently, luminous efficiencies can be increased at reduced lamp size, which is relevant for applications in compact reflectors.

7.5.3 Gas Discharge Lamps

A gaseous discharge (plasma) is obtained by driving an electric current between two adjacent electrodes through a partially ionized gas enclosed in a discharge tube, typically made of quartz. Also, pulsed discharges with a dielectric barrier between the electrodes and the plasma are used for plasma display applications. Also, electrodeless microwave excited discharges can be used as light sources.

Many physical factors influence the electrical and light technical properties of a gas discharge, such as the type and pressure of the gas, electrode material, operating temperature of the electrodes, shape and surface structure of the electrodes, electrode distance, geometry of the discharge vessel, and current density. For the purpose of light generation, two basic types are distinguished: low-pressure and high-pressure discharge lamps. For lighting applications, both are operated in the arc discharge mode, which is characterized by high current densities, more than 1 A/cm^2. Electric currents are limited by using inductances or electronic ballasts [60].

In low-pressure discharge lamps, the gas pressure is typically less than 100 Pa. Consequently, the mean free path length of electrons is larger or in the order of the discharge tube diameter, e.g. a few cm. Under these conditions, they gain high energies of more than 1 eV from the applied electric field due to low collision rates with the neutral gas atoms. Finally, the cold atoms are effectively excited by inelastic collisions with the hot electrons. The temperature difference of electrons and atoms reflects the situation of non-thermal equilibrium, as shown in Fig. 7.12.

In high-pressure discharge lamps the operating pressure is typically in the range between 10 kPa and 10 MPa. Under these conditions, collisions between electrons and atoms or ions are much more frequent, and a local thermal equilibrium (LTE) with close by or equal particle temperatures is established.

In low-pressure gas discharge lamps, atomic line radiation is emitted preferably from resonance transitions of the element with the lowest excitation potential.

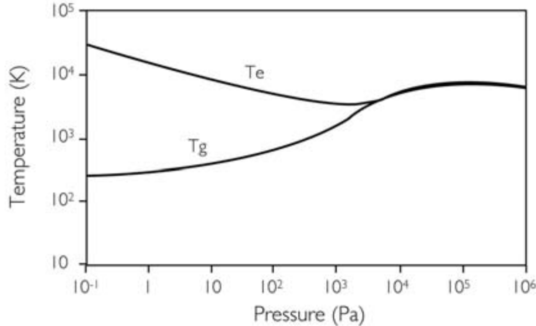


Fig. 7.12. Temperatures of electrons and neutral gas as a function of gas pressure

For mercury these transitions are in the deep UV wavelength region at 185 nm and 254 nm, whereas from sodium visible radiation at 589 nm is emitted.

In high-pressure discharge lamps various contributions to the spectrum are obtained: spectrally broadened atomic lines (resonance-, Van der Waals- and Stark-broadening), molecular radiation bands and quasi-continuous emission due to free-free (Bremsstrahlung) and free-bound (recombination of electrons with ions and atoms) transitions. As a result, the colour rendering properties of high-pressure discharge lamps are fair to excellent, depending on the type of filling [61].

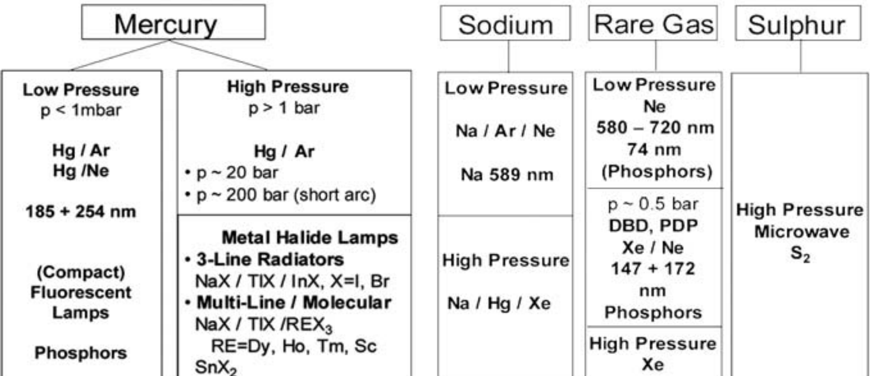
Review of Gas Discharge Lamps

Table 7.2 gives a review of the most relevant types of gas discharge lamps. The light sources differ with respect to the emission spectra and application fields. Luminescent materials are applied for conversion of UV-radiation into visible light for low-pressure mercury, sodium and xenon excimer lamps.

Low-Pressure Mercury Lamps

In Fig. 7.13 the working principle of a low-pressure mercury lamp (also known as fluorescent lamp) is shown. They are designed in the form of a linear or bended tubular bulb with the electrodes sealed in at the end parts. Also, electrodeless lamps are available, where the electrical energy is coupled inductively into the discharge vessel via externally mounted metal coils. A rare gas, e.g. argon, is filled as starting gas in addition to a few mg of mercury. At low pressures in the order of a few pascals, about 97% of the emission from Hg atoms is in the ultraviolet wavelength region. Therefore, the inner surface of the bulb is coated with a fluorescent powder (phosphor), which efficiently converts the UV radiation into visible light. The composition of the phosphor materials determines the spectral power distribution and colour of emitted wavelengths.

Table 7.2. Review of basic types of gas discharge lamps



Mercury		Sodium	Rare Gas	Sulphur
Low Pressure p < 1mbar Hg / Ar Hg /Ne 185 + 254 nm (Compact) Fluorescent Lamps Phosphors	High Pressure p > 1 bar Hg / Ar • p ~ 20 bar • p ~ 200 bar (short arc)	Low Pressure Na / Ar / Ne Na 589 nm	Low Pressure Ne 580 – 720 nm 74 nm (Phosphors)	High Pressure Microwave S_2
	Metal Halide Lamps • 3-Line Radiators NaX / TlX / InX, X=I, Br • Multi-Line / Molecular NaX / TlX /REX_3 RE=Dy, Ho, Tm, Sc SnX_2	High Pressure Na / Hg / Xe	p ~ 0.5 bar DBD, PDP Xe / Ne 147 + 172 nm Phosphors	
			High Pressure Xe	

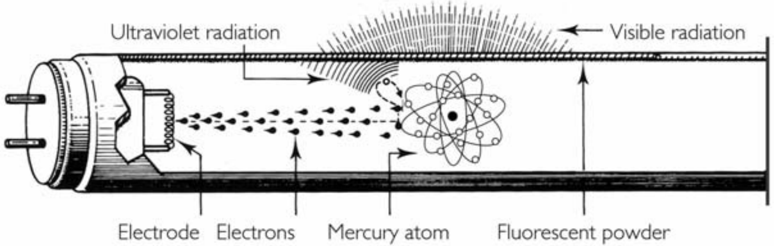


Fig. 7.13. Schematic of a low-pressure mercury discharge (fluorescent) lamp

The generation of UV photons arises from transitions of mercury atoms between the excited state levels $^1P_1$, $^3P_1$ and the ground state level $^1S_0$. About 64% of the electrical input power is converted into UV photons at a wavelength of 185 nm and 254 nm. Only 3% are directly emitted as visible light (405 nm, 436 nm, 546 nm and 589 nm). Although the quantum efficiencies of applied phosphors are close to unity, i.e. one UV photon is converted into one visible photon, the overall efficiency for visible radiation is only 28% due to the large Stokes shift. This value corresponds to a luminous efficiency of about 100 lm/W.

Typical ranges of fluorescent lamp parameters are: electrical input power up to 140 W, luminous efficiency up to 100 lm/W, colour temperature between 2700 K and 8000 K. The lamps are available in various geometries, e.g. cylindrical, circular or U-shaped tubes. The latter are also known as energy saving or compact fluorescent lamps. Depending on lamp construction, lifetime ranges from 5000 to 25 000 hours.

7.5.4 Fluorescent Coatings

The fluorescent powder or phosphor is an essential part of a fluorescent lamp since it is designed to obtain the required visible radiation. It is coated onto the inner side of the glass tube in order to effectively absorb UV photons from the discharge. The powder consists of one or several luminescent materials, which are in general inorganic compounds doped by transition metals (e.g. Mn^{2+}, Mn^{4+}) or rare-earth ions (e.g. Tb^{3+}, Eu^{3+}). The composition of the fluorescent powder is optimized for an efficient absorption of the atomic resonance lines of mercury. In addition, it efficiently emits photons in the required visible spectrum. In some cases the fluorescent composition also consists of a so-called sensitizers (e.g. Ce^{3+}). It is used if the first dopant (the activator) does not sufficiently absorb the UV radiation of mercury. The light conversion process begins with the absorption of incident photons either by the activator or the sensitiser. In the latter case, the energy is subsequently transferred to the activator. The excited activator ion, for instance Eu^{3+}, decays to the ground state by emitting a photon. Its wavelength corresponds to the energy gap between the excited and ground state. The energy difference between the emitted and the absorbed photon represents the Stokes shift. This energy loss of about 50% is dissipated as heat in the fluorescent powder.

Table 7.3. Review of most commonly used luminescent materials in fluorescent lamps

Composition of luminescent materials	Emitted peak wavelength (nm)	Color point x, y[a] (after CIE 1931)	Example of application area
LaB_3O_6:Bi,Gd	311	–	medical lamps
$LaPO_4$:Ce	320	–	tanning lamps
$BaSi_2O_5$:Pb	350	–	tanning lamps
SrB_4O_7:Eu	368	–	black light lamps
$Sr_2P_2O_7$:Eu	420	0.167, 0.014	reprography lamps
$BaMgAl_{10}O_{17}$:Eu	453	0.150, 0.070	colour 80 lamps[b]
Zn_2SiO_4:Mn	530	0.256, 0.700	decoration lamps
$LaPO_4$:Ce,Tb	543	0.343, 0.585	colour 80 lamps[a]
$CeMgAl_{11}O_{19}$:Tb	543	0.350, 0.582	colour 80 lamps[a]
$GdMgB_5O_{10}$:Ce,Tb	543	0.346, 0.531	colour 80 lamps[a]
$Y_3Al_5O_{12}$:Ce	560	0.453, 0.523	colour 90 lamps[a]
$Ca_5(PO_4)_3$(F,Cl):Sb, Mn	575	0.356, 0.377	halophosphate lamps
Y_2O_3:Eu	611	0.643, 0.344	colour 80 lamps[a]
$GdMgB_5O_{10}$:Ce,Tb, Mn	630	0.602, 0.382	colour 90 lamps[a]
$Mg_4GeO_{5.5}$F:Mn	660	0.700, 0.287	decoration lamps

[a] At 254 nm excitation
[b] See text

Today, fluorescent materials compositions with quantum efficiency close to unity are commercially available. Table 7.3 gives a review of most commonly used phosphors, their characteristics and examples of application areas. For illumination purposes, halophosphate phosphors are still favoured. They yield a white emission spectrum with a reasonable efficiency and colour rendering. However, a trichromatic phosphor blend results in improved lamp efficiency and better colour rendition. An ordinary phosphor blend comprises $BaMgAl_{10}O_{17}$:Eu, $LaPO_4$:CeTb, and Y_2O_3:Eu. In Table 7.3 such so-called colour 80 lamps are indicated with an asterisk. They exhibit a spectral power distribution as shown in Fig. 7.14.

A further improvement of color rendition is obtained by using a tetra- or pentachromatic phosphor blend. Colour 90 lamps additionally consist of phosphors with an emission band position where the trichromatic phosphor blend does not radiate, viz. in the yellow ($Y_3Al_5O_{12}$:Ce) and deep red ($GdMgB_5O_{10}$:Ce,Tb,Mn) part of the spectrum.

Today, the tuneability of fluorescent lamp spectra results in many other application areas beyond illumination, such as medical and cosmetic skin treatment, decoration, reprography and horticultural lighting. The latter, for instance, comprises a dichromatic phosphor blend consisting of $BaMgAl_{10}O_{17}$:Eu and Y_2O_3:Eu. Its emission spectrum is optimised with respect to the action spectrum for photosynthesis of green plants.

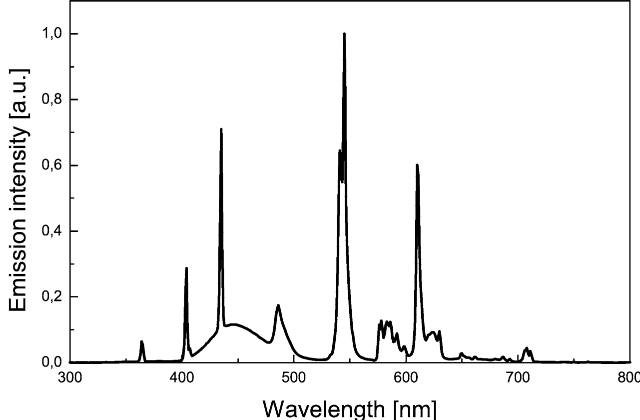


Fig. 7.14. Spectrum of colour 80 fluorescent lamp comprising a blend out of $BaMgAl_{10}O_{17}$:Eu, $LaPO_4$:Ce,Tb, and Y_2O_3:Eu

Other Low-Pressure Discharge Lamps

The main difference of low-pressure sodium lamps to mercury lamps is given by the fact that sodium significantly emits radiation directly in the visible spectrum. The Na D-lines deliver almost monochromatic yellow light from the doublet at 589.0 nm and 589.6 nm. Furthermore, the melting point of Na is higher compared to Hg resulting in an optimal operating temperature of about 530 K. At these temperatures the Na metal is highly reactive. Therefore, a chemically stable wall material like quartz (SiO_2) or alumina (Al_2O_3) is required. As indicated in Fig. 7.15, thermal losses are reduced by mounting the discharge tube inside an evacuated outer bulb.

Due to the low excitation energies and convenient spectral position of the Na D-lines, these lamps exhibit the highest ever achieved luminous efficiency. At a discharge power of 200 W values of 200 lm/W at an overall system efficiency of 172 lm/W are realized. Unfortunately, a low rendition of colours is obtained due to missing emission at other wavelengths, e.g. in the blue, green and red part of the spectrum. For this reason the application of low-pressure sodium lamps is limited mainly to street and outdoor lighting where high efficiencies and long lifetime of about 20 000 hours are required. The poor colour rendition can be further improved by increasing the sodium vapor pressure to atmospheric values. In high-pressure sodium lamps, the Na D-lines are spectrally broadened, as described in the section below.

Besides low-pressure mercury and sodium lamps also other types of radiators such as neon and xenon are known. For example, low-pressure neon lamps are applied in automotive brake lights and advertisement lighting. Excimer UV-radiation from xenon is used in plasma displays and flat backlights, photocopier lamps and UV-purification devices. Here, atomic and molecular radiation of xenon at 147 nm and 172 nm is emitted, respectively. A major advantage of these discharges is their

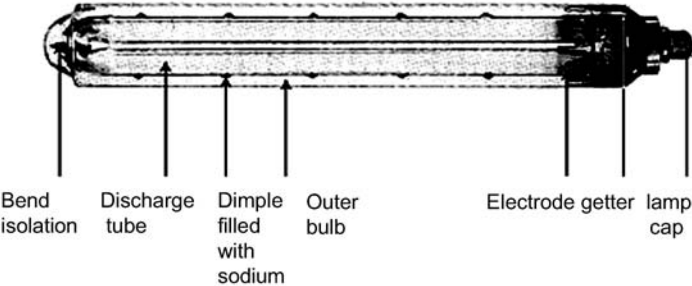


Fig. 7.15. Build up of a low-pressure sodium lamp

fast switching behaviour since in contrast to mercury and sodium, xenon is already in the vapor state.

7.5.5 High-Pressure Discharge Lamps

When raising the pressure in a low-pressure discharge, the rate of elastic collisions increases. Although each collision transfers only little energy to heavy particles, the plasma is heated effectively due to large particle densities. As a result, a plasma temperature gradient establishes in order to sustain central plasma temperatures in the range between 4000 and 10 000 K. The heat flux along the temperature gradient to the discharge wall represents a thermal loss of energy limiting the radiation efficiency to around 60%. Selective emission of the plasma arc is used for efficient visible radiation. High-pressure discharge lamps contain metals like mercury or sodium.

In the so-called metal halide lamps also other efficient radiators are added to the lamp filling. Different combinations of metals generate visible radiation with various colours and radiant efficiencies. Theoretically, at least, no fewer than 50 different metals, typically dosed in the form of metal halide compounds, can be used as additives. Lamp manufacturers have introduced various combinations on the market with a wide range of applications.

In high-pressure sodium and metal halide lamps a major technical breakthrough has been realized by the introduction of polycrystalline alumina (PCA) – a sintered Al_2O_3 ceramic – as a wall material. Compared to quartz, it is much more chemical resistant against hot sodium vapor. Therefore, PCA is used as a wall material in high-pressure sodium lamps. Generally, ceramic materials allow lamp operation at high wall temperatures resulting in an effective evaporation of the salt fillings. In contrast, lamps with quartz envelopes are limited to maximum wall temperatures of about 1370 K due to re-crystallization effects. However, the design of PCA lamps and vacuum tight sealing of electrodes is more complicated as compared to quartz due to inconvenient thermal expansion behaviour and brittleness.

High-Pressure Mercury Lamps

In Fig. 7.16 (left) the construction of a high-pressure mercury lamp is illustrated. The discharge tube is filled with mercury and argon as a starting gas. The discharge vessel is mounted in an evacuated outer bulb in order to reduce thermal losses. Ignition of the plasma is realized by applying a high voltage pulse in the order of several kV. After ignition, the plasma enters the arc regime which is characterized by a low operating voltage (about 20 V) and high discharge current (>1 A). At this stage, the lamp is mainly emitting UV radiation from mercury at 254 nm and some light from the starting gas. As a result of thermal losses from the plasma, wall temperature increases with time causing enhanced evaporation of the liquid mercury. With a further increase in mercury vapor pressure, the radiated energy is concentrated progressively towards spectral lines of longer wavelengths since higher lying energy levels are excited. The run-up of the lamp enters a steady-state phase when a mercury pressure of typically 200–1000 kPa is established. Even at these pressures, part of the emission is in the UV region which can be converted into visible light by coating the inner surface of the outer bulb with a fluorescent powder, e.g. YVO_4:Eu. Besides efficiency, also colour rendering is improved significantly.

In high-pressure discharge lamps, mercury is introduced also as a so-called buffer gas resulting in elastically scattering of electrons by mercury atoms. By this mechanism, a low electrical conductivity of the plasma is adjusted. According to Ohm's law, electrical field strengths on the order of several 10 V/mm are present in the discharge column. Typically, lamp voltages in the range of 60–90 V are needed since at a given electrical input power only moderate discharge currents can be tolerated over lamp life. Electrode temperatures are in the range between 2400 and 3500 K, the

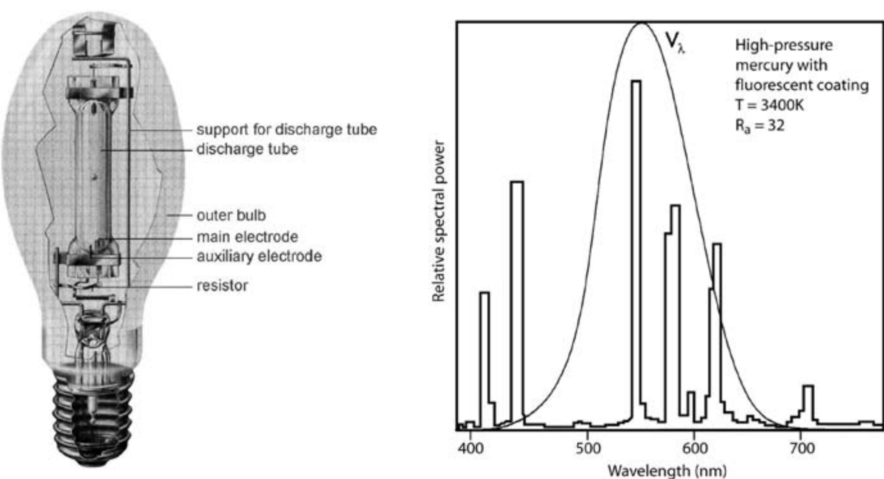


Fig. 7.16. *Left*: construction of a high-pressure mercury lamp. The inner surface of the outer bulb may be coated with a phosphor additionally converting UV radiation into visible light. *Right*: spectrum of a high-pressure mercury lamp

electrons being generated thermionically. For mercury both properties – its volatile character enabling large atom densities and its large cross-section for elastic electron scattering – are unique. Unfortunately, mercury is an environmental hazard and the lighting industry aims for the realization of mercury-free products in the near future [59].

High-pressure mercury lamps are used in various indoor and outdoor applications. The power ranges from 50 up to 1000 W with luminous efficiencies around 50–60 lm/W. Colour rendition is only moderate to poor. Colour temperatures range from 3400 to 6000 K.

A recent development of a high-pressure mercury discharge is the UHP lamp, which was invented by PHILIPS in the 1990s. This lamp has a very high operating mercury pressure (about 2×10^7 Pa) and a short arc length of about 1 mm only. This point-like white light source is applied in TV projection systems and beamers. In contrast to conventional high-pressure mercury lamps, a significant part of the visible radiation is generated not only by pressure broadened atomic mercury lines but also by Hg_2 molecules. UHP lamps are available on the market in the power range between 100 and 250 W. One of the key issues of UHP lamps is to maintain stable properties of the plasma and electrodes for several thousands of hours. As lined out in the next section, this can be achieved by adjusting a so-called regenerative chemical cycle: evaporated tungsten is transported back onto the electrodes via compounds including oxygen and bromine. Furthermore, UHP lamps have to be operated with special electronic devices in order to control electrode erosion and arc jumping.

Metal Halide Discharge Lamps

The construction of metal halide lamps is quite similar to high-pressure mercury lamps. Besides mercury and argon, metal halides are added to the lamp fill. Large particle densities are needed for efficient radiation emission in the visible spectrum. Metal halides are also less aggressive with respect to wall corrosion as compared to the pure metals.

Three main groups of metal halide lamps are distinguished: multi-line radiators, molecular radiators and three-band colour radiators. In the first type rare-earth halides like DyX_3, HoX_3, TmX_3 or ScX_3 ($X = I$, Br) are added to the lamp filling. These species have a large number of transitions at low mean excitation energy. Such plasmas efficiently emit multi-line radiation in the visible spectrum. Typically, NaI and TlI are added to such lamp fillings to improve luminous efficiency and colour properties. The second type is based on molecular radiators like SnI_2 or $SnCl_2$, which produce a quasi-continuous spectrum. The third class comprises mixtures of NaI, TlI and InI_3. The resulting spectra consist of three colour bands in the yellow, green and blue, respectively.

In Fig. 7.17 the spectra of these basic lamp types are shown. Application areas are found for various indoor and outdoor lighting purposes. The electrical input power ranges from 35 to more than 1000 W. Metal halide lamps deliver high luminous efficiencies of up to 100 lm/W and good to excellent colour rendition. Lifetime is up to 20 000 h, depending on the type of filling.

A specialty of the lamp types of Fig. 7.17 (right) and Fig. 7.18 (left) is the introduction of a metal halide cycle. This principle is illustrated in Fig. 7.18 (right). When a metal halide lamp first starts, the spectrum is initially that of mercury and starting gas vapor, since the halides remain unevaporated at the cold discharge tube wall. As the wall temperature increases, the metal halides melt and begin to vaporize. Diffusion and convection cause a transport of the vapor species into the hot region of the arc. Here, the halide compounds dissociate into halogen and metal atoms. As shown in Fig. 7.18 (right), different halides dissociate at indicated temperatures. The metal atoms are excited at the hot plasma centre having a temperature of about 5000 K where the main contribution of atomic radiation is emitted. The metal and halogen atoms then move towards colder regions near the discharge wall where they recombine and once again form the halide compound. This cycle repeats itself continu-

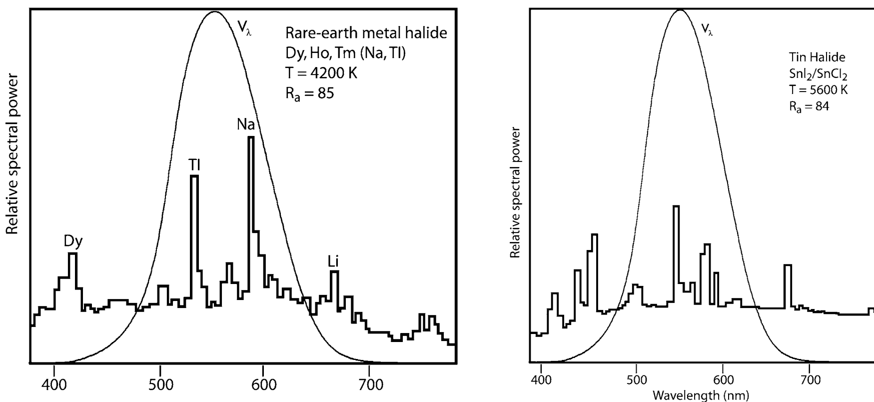


Fig. 7.17. *Left*: spectrum of a $NaI/TlI/DyI_3/HoI_3/TmI_3$ lamp for sports and flood lighting. *Right*: spectrum of a $SnI_2/SnCl_2$ lamp for studio and theatre lighting

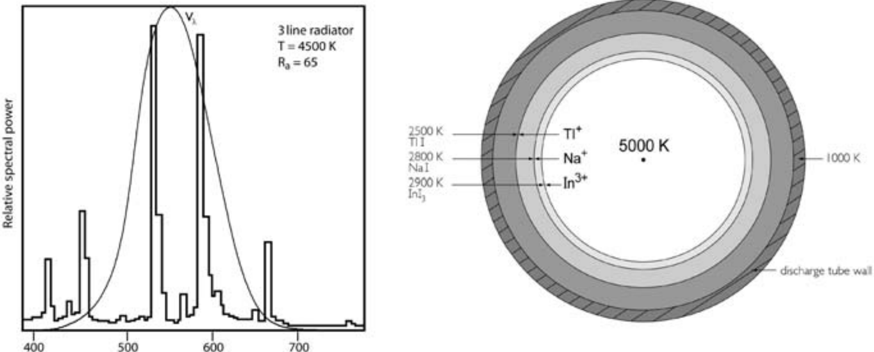


Fig. 7.18. *Left*: spectrum of a three-band colour (NaI, Tl, InI_3) metal halide lamp. The lithium line is due to the presence of impurities in the wall material (quartz). *Right*: schematic diagram of the halide cycle of a three-band colour metal halide lamp (NaI, TlI, InI_3). The values indicate temperatures of wall, arc centre and dissociation of molecules

ously. When adding oxygen and bromine, for example, also a regenerative tungsten cycle can be realized which significantly prevents the lamps from wall blackening.

As already mentioned, a recent improvement of metal halide lamps was achieved by the introduction of polycrystalline alumina (PCA) as a wall material instead of quartz. Such lamps exhibit excellent colour stability and long lamp life. Unfortunately, they are not suitable for optical applications, e.g. in projection systems. The reason is the translucent property of PCA resulting from multiple scattering of photons at grains and pores. Nevertheless, the total transmission for visible light is nearly 95%. Also application of PCA to high power metal halide lamps is limited: cracking of the discharge tube induced by thermal shock and expansion, especially during run-up or when switching off the lamps, is observed. In the lighting industry, the applicability of ceramic wall materials is still an actual field in research and development of modern light sources. Furthermore, upcoming legislation initiated several activities on the replacement of mercury in this type of discharge by non-toxic materials [59]. The first mercury-free metal halide lamp for automotive headlight application was introduced to the market by Philips in 2004.

Other High-Pressure Discharge Lamps

Besides high-pressure mercury and metal halide lamps, the so-called blended-light-lamp is a derivative of the conventional high-pressure mercury lamp. It has a built-in ballast in the form of a tungsten filament connected in series with the discharge. Therefore, radiation from plasma and filament combine (or blend). However, the improved colour property is realized at the expense of reduced system efficiency.

In high-pressure sodium lamps, broadening of the D-lines results in improved colour quality compared to the low-pressure version. Luminous efficiencies of up to 130 lm/W are obtained. Using PCA instead of quartz dramatically improves lamp life and stability of the spectra since it withstands the aggressive Na vapor.

Another class of high-pressure discharge lamps is represented by the sulphur lamp, invented by Turner et al. at FUSION SYSTEMS in 1994. The advantage of this lamp is a long life at high luminous efficiency with electrodeless operation. Sulphur lamps are excited by a microwave generator at a wavelength of 2.45 GHz. The electrical input power is in the order of several kW. The spherical bulb having a diameter of about 30 mm is filled with sulphur powder. At continuous operation, a sun-like white light is emitted by S_2 molecules having a partial pressure of about 600 kPa. High efficiencies of the discharge of about 170 lm/W can be realized since at these pressures UV radiation of the sulphur molecules is reabsorbed by the plasma. Thus, radiation emission mainly occurs in the visible part of the spectrum. A technical drawback of microwave lamps is the quite low energy efficiency of the generator of about 65%. Therefore, the overall efficiency of the system is reduced to about 100 lm/W. Finally, sulphur lamps turned out not to be competitive with conventional metal halide lamps. However, electrodeless discharge lamps are still a hot topic in the lighting industry. One advantage of these systems is their potentially long life due to the lack of plasma interaction with electrodes. Light may be distributed by so-called light engines which are coupled to light guiding pipes. This technique could

Table 7.4. Electrical and light technical data of basic types of visible light sources

Light source	Electrical input power (W)	Luminous flux (lm)	Luminous efficacy (lm/W)	Colour rendering quality
Incandescent	10–1000	80–15 000	8–15	excellent
Halogen	20–2000	300–60 000	15–30	excellent
Low-pressure Hg discharge	7–150	350–15 000	50–100	good
High-pressure Hg discharge	50–1000	2000–60 000	40–60	good
Metal-halide discharge	20–2000	1600–24 000	80–120	good to excellent
Low-pressure Na discharge	20–200	2000–40 000	100–200	poor
High-pressure Na discharge	40–1000	1600–14 000	40–140	moderate to good
Sulphur microwave discharge	up to 5000	up to 450 000	80–90 (system)	good
White dichromatic inorganic LED	1–5	20–150	20–30	good
White trichromatic inorganic LED	1	20–25	20–25	excellent
Organic LED (at 1,000 cd/m^2)	15 mW (per cm^2)	0.25 lm (per cm^2)	15	good

be advantageous for industrial and general lighting applications, e.g. where central light sources are preferred as compared to conventional distributed lighting systems.

7.5.6 General Light Source Survey

Table 7.4 gives a summary over main types of visible light sources together with selected electrical and light technical properties.

7.6 Ultraviolet Light Sources

7.6.1 Introduction

Ultraviolet (UV) light extends from the visible to the X-ray region of the electromagnetic spectrum and is divided in spectral regions UVA, UVB and UVC. This is relevant, e.g. for medical skin treatment and sun tanning studios. The CIE definition is 315 to 400 nm for UVA, 280 to 315 nm for UVB, and 100 to 280 nm for UVC. The region from 200 to 280 nm may be called far ultraviolet (FUV). The term vacuum ultraviolet (VUV) is used for the region 100 to 200 nm. Wavelengths of 105 and 110 nm are the shortest cut-off wavelengths for optical windows (LiF or MgF_2,

respectively) and can be used to distinguish between “extreme ultraviolet” (XUV) with $\lambda < 110$ nm and VUV. Note that the term vacuum ultraviolet may be misleading in the sense that absorption of air sets in at about 200 nm only because of the Schumann–Runge bands (B to X transition) of molecular oxygen and if pure nitrogen or rare gases are used to purge the optical system, electromagnetic radiation may propagate through dense gases without significant absorption down to wavelengths as short as 60 nm, corresponding to the first excitation energy of helium atoms. The expressions deep UV and extreme UV (EUV) are used in photolithography for the VUV and soft X-rays ($\lambda = 13.6$ nm), respectively.

From the various types of UV light sources only portable, sealed-off devices are described here. A broader description of vacuum ultraviolet technology, including e.g. synchrotron sources, can be found in [62] and [63]. Laser induced plasmas are incoherent VUV and XUV sources but will also not be discussed here.

7.6.2 Thermal Tungsten Lamps for Intensity Calibration

Although the intensity of thermal radiation decreases rapidly towards short wavelengths according to Planck’s law (see formula (7.1) and Fig. 7.1 in Sect. 7.1.1), there is an important wavelength region around 300 nm in which thermal emitters overlap with other sources such as deuterium lamps, which will be discussed below. This allows intensity calibrations in the UV and VUV regions if an absolutely calibrated thermal emitter is used in combination with a UV light source for which the relative intensities in the emission spectrum is known. The emission of a tungsten strip lamp is shown in Fig. 7.1 together with black body radiation. This lamp is calibrated down to 250 nm, and the emission from deuterium lamps has been calibrated up to 350 nm (see next paragraph). This allows intensity calibration of optical systems from 115 to 2600 nm using these two light sources.

7.6.3 Deuterium Lamps

Deuterium lamps are the most widely used broad-band ultraviolet light sources. Their emission spectrum consists of a combination of several hydrogen molecular bands and continua which are emitted between 120 and 500 nm [63], p. 96. The intensity at wavelengths longer than 350 nm, however, is very low. The Lyman-α line at 121.57 nm is the only pronounced atomic line radiation in the spectrum. The so-called Lyman- and Werner bands which are due to B to X and C to X transitions, respectively, cover the wavelength range from 120 to 130 and 140 to 170 nm, respectively. The structureless continuum at wavelengths longer than 170 nm is due to the radiative decay of the molecular state a ($^3\Sigma_g^+$) to the repulsive b ($^3\Sigma_u^+$) state. A spectrum emitted from a deuterium lamp is shown in Fig. 7.19.

Deuterium lamps are commercially available in various forms as sealed off discharge lamps. Their short wavelength cut-off is determined by the material of the output window. Quartz cut-off is at about 160 nm and MgF_2 at about 110 nm (see Fig. 7.19). The lamps are normally dc arc lamps which are started by use of a hot filament and 300 to 600 V starting voltage. Typical operation conditions are 300 mA

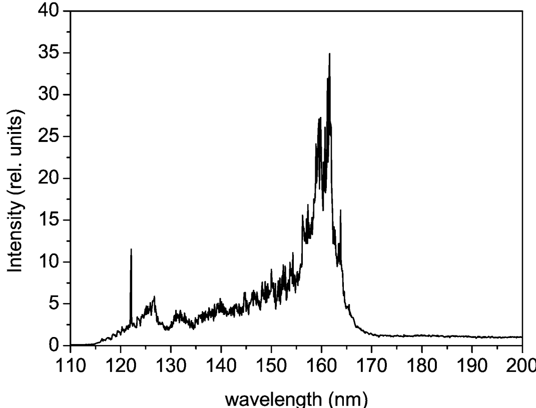


Fig. 7.19. Emission spectrum from a deuterium lamp in the VUV spectral region. The spectrum has not been corrected by the response function of the monochromator detector system. An assignment of the emission bands is given in the text. The line at 121 nm is the Lyman-α line of atomic hydrogen

current and 80 V voltage drop (24 W), and wall plug power is about 100 W. The arc burns between an anode plate and a box-shaped cathode with two openings which define an optical axis. One opening constricts the arc to about 1 mm diameter, and the light is emitted through the second opening of the cathode box. The filament is mounted inside the cathode box, displaced from the optical axis. Besides the standard lamps, water cooled high power devices are available. The shape of the emission spectrum of deuterium lamps has been studied in detail and the spectral radiance of specific deuterium lamps has been tabulated, so that they can be used as transfer standards in the spectral range from 115 to 350 nm [64].

7.6.4 UV Arc Lamps

High-current wall stabilized arc lamps produce an emission spectrum which is a combination of a continuum extending over a wide range in the UV and VUV and line radiation depending on the atoms which are present in the plasma. A light source using a flow of argon at atmospheric pressure has been described in detail in the literature as a transfer standard for spectral radiance [65]. It uses a wall stabilized arc of typically 30 V operating voltage and 50 A current for excitation. The arc is constricted by a 6.3 mm thick water cooled copper plate with a 4 mm diameter hole. The authors provide a table of radiance values between 330 and 114.4 nm and report radiances of, for example, 70 mW/(cm^2 nm sr) at 330 nm and still 6.5 mW/(cm^2 nm sr) at 151 nm for these operating parameters. A more detailed description of the geometry, operating conditions and ways to add different light emitting species to the buffer gas is given in [66]. A wall stabilized arc with a stable radiance on the order of 1 W/(cm^2 nm sr) is described in [67] when xenon was used as the light emitting species.

Arc discharges in sealed off quartz- or ceramic bulbs are usually used for lighting applications. The gas filling is often xenon and mercury with high operating pressures up to 100 bar. They emit a continuous spectrum which resembles a black body spectrum of typically 5000 K radiation temperature with some narrow lines from Hg and Xe on top. The technology of such lamps is described in Sect. 7.5. The UV part of the emission spectrum can be used for simulating the solar spectrum, for example, for testing paints and dyes. Light sources for simulating irradiation by sunlight are commercially available for that purpose.

7.6.5 Mercury Lamps

Although presently the semiconductor industry introduces the ArF laser with a wavelength of 193 nm as light source for photolithography, high-power mercury arc lamps emitting a line dominated spectrum are still of great importance as UV sources in this field, in devices named DUV and i-line stepper. The wavelengths used are 248 and 365 nm, respectively.

Low-power mercury lamps provide the line spectrum shown in Fig. 7.20. The mercury atoms are excited in a low pressure glow discharge with rare gas used as the buffer gas. The concentration of mercury atoms in all mercury lamps is defined by Hg-vapor pressure and therefore by the temperature of the discharge tube. The most widespread application of low-pressure mercury lamps is to provide the ultraviolet light which is converted into visible light in fluorescent lamps. Such light sources are described in detail in Sect. 7.2.2. Some mercury lamps are used as “black lamps” emitting UVA light for fluorescence applications. Fluorescent coatings may be used to shift light emitted in the deep UV to the desired UVA wavelength region, and filters protect the user from the hard UV emitted from mercury.

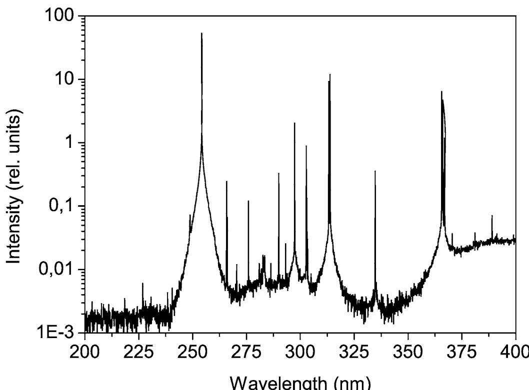


Fig. 7.20. Spectrum of the emission of mercury in the wavelength region 200 to 400 nm. The spectrum was recorded using a small Hg glow discharge lamp for wavelength calibration of monochromator detector systems

Small low-pressure mercury glow discharge lamps with the discharge burning in U-shaped, narrow quartz capillaries are used for wavelength calibration of UV spectrometers (see Fig. 7.20).

7.6.6 Hollow Cathode Lamps

Low-pressure hollow cathode lamps are widely used spectral line sources for atomic absorption spectroscopy in analytical chemistry. They emit a spectrum which is characteristic of the cathode material due to deliberately sputtering the cathode material. Many lines are in the UV part of the electromagnetic spectrum. The lamps consist of a cylindrical cathode and a ring shaped anode inside a glass or quartz tube. A cathode material is introduced to a gas filling of typically 10 mbar neon or argon by sputtering via ions which are accelerated in the cathode fall of the discharge. Line radiation from atoms of about 70 elements, either individually or in combinations, are commercially available via excitation in hollow cathode discharge lamps. The excitation process and technology of hollow cathode lamps is described in more detail in Chap. 4 of [62].

7.6.7 Excimer Light Sources

General Aspects

Excimer light sources are very efficient sources. They cover a wide range of the ultraviolet spectral region and have found their way into practical applications when dielectric barrier discharges were introduced as the excitation method. The excimer light emission process is also of interest for plasma flat panel displays when, for example, dense xenon gas is used as the light emitting species.

Formation and decay of rare gas excimer molecules has been studied for a long time [68, 69]. A schematic energy level diagram for two rare gas atoms versus internuclear distance is plotted in Fig. 7.21. It shows that excited rare gas atoms have an attractive potential and thus can form molecules at elevated gas density via three body collisions.

Rare gas ultraviolet excimer light sources may reach the efficiency of more than 50%. This is due to the favourable energy level scheme and gas kinetics [70]. Briefly, collisional excitation of rare gas atoms leads in a first step to the formation of excited atoms and ions. Ions rapidly form ionic molecules in a dense gas. Molecular ions undergo a so-called dissociative recombination leaving one of the atoms in the ground state and the other in a highly excited state. Highly excited atoms which are either produced by this process or by direct excitation will cascade down to the lowest lying excited states. These states are either metastable or have a long effective lifetime due to radiative trapping of the light in the dense gas. Therefore, essentially all the atoms in the first excited states will form excimer molecules via three body collisions. Excimer molecules decay to the repulsive ground state emitting broad-band continua which cover a wide spectral range from about 60 to 200 nm, as shown in Fig. 7.22. Note that all photons which are emitted from molecular states have a lower energy

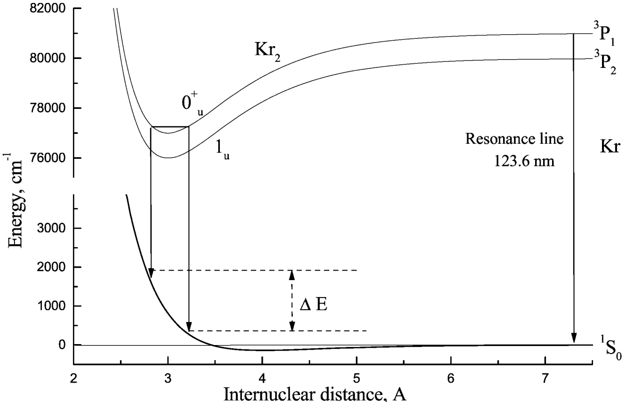


Fig. 7.21. Schematic drawing of a potential energy diagram of the lowest lying levels of two rare gas atoms versus internuclear distance (Kr as an example)

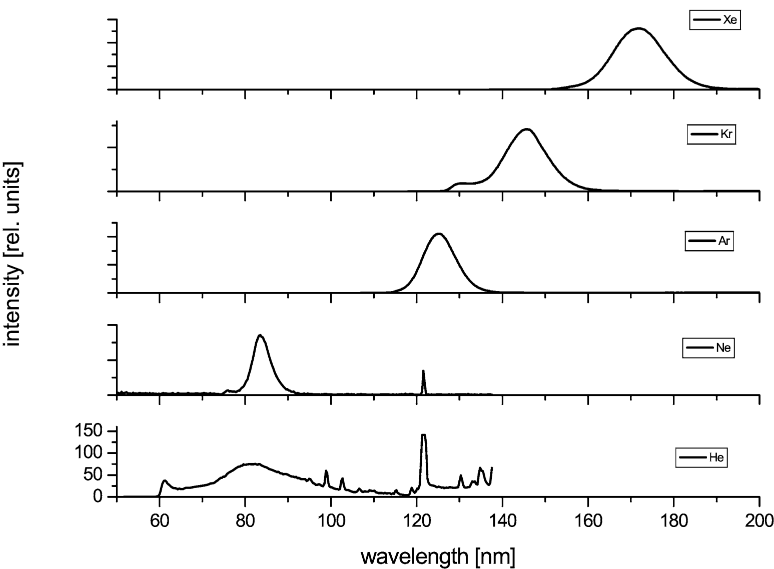


Fig. 7.22. Review of the rare gas emission spectra recorded at a relatively high gas density (gas pressure 1 bar, 1000 hPa)

than photons emitted on the resonance lines of rare gas atoms and are therefore not radiatively trapped in the dense gas.

Emission spectra of rare gas excimer light sources can be strongly modified by using rare gas mixtures and by adding other gases. The reason is that the energeti-

cally high lying levels in rare gas atoms can transfer energy to other species leading to their excitation and ionisation with subsequent emission of light due to transitions in the species which have been added. This effect, however, leads to the problem of pure rare gas excimer lamps that already very low impurity levels can drastically reduce their spectral performance and efficiency. For example, by adding molecular hydrogen on a level of 1%, the neon excimer disappears completely, and the spectrum is dominated by the atomic Lyman-α line at 121.57 nm [71–73]. This example shows that excimer lamps can be easily converted from broad band into narrow band ultraviolet sources and that the emission wavelength can be shifted to longer wavelengths via energy transfer processes. Other examples of ultraviolet sources in which the emission spectrum of rare gases is strongly modified by small admixtures of other gases have been described in the literature: Ar–Xe [74], Kr–Xe [75], Ar–N_2 [76]. The great variety of UV emitting rare gas halogen mixtures used for excimer lasers systems can also be used for incoherent UV and VUV light sources by applying appropriate excitation methods.

Specific Excimer Light Sources

Incoherent excimer light sources have attracted great attention since about 1990. A practical difficulty in designing an excimer lamp is to excite rare gas atoms to energetically high lying levels in a rather cold and dense gas so that the excited species can efficiently form molecules with atoms in the ground state. Pioneering work in the field of excimer lamps was performed by B. Eliasson and U. Kogelschatz using so called dielectric-barrier discharges for exciting dense gases [77]. Earlier work used condensed discharge devices for forming excimer molecules [68]. This approach has been extended very successfully to higher power and the rare gas halide excimers by V. Tarasenko and coworkers [78]. The micro-hollow-cathode discharge was introduced into the field by K.H. Schoenbach et al. [79]. It was also shown that corona discharges can efficiently emit excimer light [80]. Electron beam induced excimer emission can also be produced in compact, portable UV and VUV light sources by using very thin entrance foils for the beam [81].

7.6.8 Excimer Lamps Using Discharge Excitation Glow Discharges

Since glow discharge devices are rugged and easy to use, some powerful incoherent excimer light sources were built using this excitation mechanism. In a particular case a light source was built in the form of two 40 cm long coaxial quartz tubes [82]. The outer tube had an inner diameter of 60 mm, and discharge currents up to 0.5 A were sent through the gap between the outer and inner tube which was either 2.5 or 8 mm wide. The inner tube was used for water cooling of the device. Operating regimes were either high voltages (5 to 7 kV) and low currents (2 to 3 mA) or lower voltages (2 to 3 kV) and higher currents (up to 500 mA). An UV output power up to 130 W was produced with efficiencies on the order of 10%, and was improved to 200 W output and 15% efficiency. Such devices are limited by constriction of the discharge into an arc, a condition under which excimers cannot form.

Dielectric Barrier Discharges

Dielectric barrier discharges (DBD), also called silent discharges, are an excitation method for incoherent excimer light sources [83]. The concept is schematically shown in Fig. 7.23. At least one of two electrodes, which are attached to a gas filled cell, is covered by an insulating material, the dielectric barrier, and an alternating voltage with typically tens of kHz repetition rate and tens of kV amplitude is applied to the electrodes. Capacitive coupling leads to ignition in the gas. Hot streamer discharges develop when the cell is filled with dense gas. But these streamers are stopped already after a few nanoseconds since the dielectric wall charges up due to the current flow in the streamer. In the afterglow phase of the streamer, excimer molecules form from the atoms which were excited and ionized in the streamer phase. Dielectric barrier discharges can be designed to produce only one localized streamer at a time, but normally they make use of a concept where the self-terminating streamer discharges are formed in large quantities randomly in space and time, so that a quasi-homogeneous and quasi-continuous discharge pattern develops. Many parameters such as gas pressure, frequency, and pulse shape of the applied voltage are relevant for dielectric barrier discharge operation. Geometry and size of the devices can be varied over wide ranges and adapted for specific applications. DBD development is an active field of research [84–86].

An advantage of DBDs is that the gas can be kept clean, for example in quartz tubes where it is not in direct contact with the electrodes. Dielectric barrier discharges reach conversion efficiencies of up to 80% light output from power deposited in the discharge volume. The overall efficiency, however, is significantly lower due to the limited coupling efficiency of electrical power into the discharge. This coupling has to be optimized by the electrical design of the device (repetition rate and pulse shape). Small, portable and also large stationary ultraviolet light sources have been built. These devices are the most advanced excimer light sources so far, and are already commercially available from major lighting companies, as UV and VUV sources and also large area, mercury-free visible light sources by using phosphors as wavelength shifters.

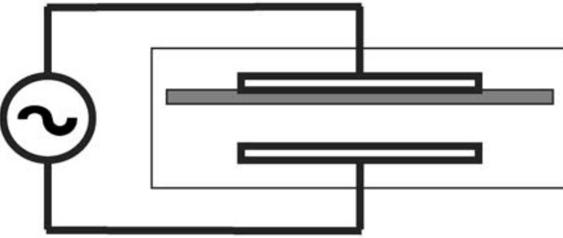


Fig. 7.23. Basic geometry for DBD discharges. Streamer discharges develop between two electrodes. At least one of the electrodes is covered by an insulating material schematically shown as the gray plate next to the top electrode

Micro-Hollow-Cathode Discharges

The use of micro-hollow-cathode discharges (MHC) for excimer light sources has been pioneered by K.H. Schoenbach and his group based on a study of the electrical characteristics of a hollow cathode discharge with small hole diameter (0.7 mm) [87]. The product of gas pressure p and hole diameter D determines the operation regime of the discharge. Successful operation is observed for values which are typically between 0.03 and 13 mbar cm. A light source was demonstrated with argon and xenon at pressures up to one bar and hollow cathodes of 100 μm diameter. Micro-hollow-cathode discharges are appropriate for excimer formation, since the electron energy distribution function reaches up to 100 eV in the cathode fall of this type of discharge, and effective molecule formation is supported by the relatively high gas pressures p which are possible due to the pD scaling of the discharge and small D values. The current–voltage characteristic of the DC- MHC-discharges has a positive slope, so that many MHC discharges can be operated in parallel from the same power supply without electrical interference. The perspective is to build large area, flat UV emitting devices. MHC discharges have also been used with pulsed operation and high instantaneous discharge currents. Due to the high current densities in the narrow, constricted discharge there is a high rate of electrode erosion and spectral lines of the electrode material (e.g. molybdenum) have been observed in the emitted light.

Corona Discharges

A corona discharge can also be used for excimer formation (St. Elmo's fire) [80, 88]. Like a micro-hollow-cathode discharge it is not electrode-less, but the discharge burns more freely between a sharp metal tip and a metal mesh. The operating voltage is of the order of 1 kV and an output power on the second continuum of xenon of 35 mW/cm^2 has been observed with a prototype lamp of 50% efficiency. The needle electrodes used in this setup can also be operated in parallel for designing large area lamps. Needle electrodes have also been used by the same research group with RF-excitation [89]. It could be shown that this reduces electrode erosion significantly, even with very corrosive gases such as ArF and F_2.

7.6.9 Excimer Lamps Using Electron Beam Excitation

Electron beam excitation can provide the non-equilibrium conditions required to populate the rather high lying precursor levels of excimer molecules [90]. All early excimer lasers were pumped by high energy, high current electron beam devices. An advantage of the concept is that there is no feedback between the gas and the excitation process. For example, no ignition conditions depending on gas pressure or gas composition have to be met, and the excitation can have any time structure from short pulses to fully continuous.

A practical issue in all electron beam systems is the separation of the vacuum part where electrons are accelerated from the dense target gas in which the light is produced. Thin metal foils have been used since the early experiments performed

by Lenard in 1894 [91]. Electron beam excitation with very low particle energy has become possible by using extremely thin ceramic membranes as the entrance foils for electron beams [81]. Energy loss of 15 keV electrons in a 300 nm thick SiN_x entrance foil is on the order of 10% [92] and pressures up to several bar can be applied. Time averaged beam current densities of 100 μA/mm^2 can be sent through the membranes into dense gas targets. The range of the electrons in the heavier rare gases is typically 1 mm. Very compact and brilliant excimer light sources can be built when applying this technique [93, 94]. Efficiencies of excimer light sources of 35% were reached and applications using He and Ne excimer radiation extend to wavelengths below 100 nm [95].

7.7 X-ray-Tubes

7.7.1 History and Generation of X-rays

In October of 1895, Wilhelm Conrad Röntgen (1845–1923), who was professor of physics and the director of the Physical Institute of the University of Wurburg, became interested in the work of Hillorf, Crookes, Hertz, and Lenard. In June 1895 he had obtained a Lenard tube from Muller and had already repeated some of the original experiments that Lenard had created. He reproduced Lenards work generating cathode rays in free air and continued investigating of cathode rays. He modified a Crookes tube fitted with an anode and cathode, separated from each other by a few centimeters in the tube. He used a Rhumkoff induction coil to produce a cathode–anode potential of a few thousand volts knowing that a stream of charged particles would originate in the cathode and would be attracted to the anode.

His first report "On a New Kind of Rays" was published in the Proceedings of the Physical Medical Society of Wuerzburg on December 28, 1895. Not knowing what these emanations were, he uses the term X-ray to describe the rays he was producing. Later, in 1896, he accepted the Rumford gold medal of the Royal Society, and in 1901 he was the first winner of the Nobel Prize in physics.

The characteristics of X-radiation are: invisible, can excite substances to emit light, penetrates matter, cannot be deflected by electric or magnetic fields.

7.7.2 Generation of X-rays

Electron Interaction with Matter. There are 4 basic processes electrons undergo when impinging to matter. Inelastic scattering by photoeffect with ionisation, which lead to emission of characteristic X-radiation and secondary electrons. Auger effect where by internal conversion emitted characteristic X-radiation is absorbed and ionizes an electron from a higher shell. Furthermore, electrons are swinging around the atoms and delivering "Bremsstrahlung" like synchrotron radiation. Elastic scattering generates no X-rays, but changes the direction of the primary electrons.

Fig. 7.24. *Right*: Röntgen. *Left*: the Nobel prize winning photo with the invented X-rays of his wife's hand

Primary electron of high energy suffer inelastic losses by exciting electrons from the shells of the atoms of the sample into higher shells with unoccupied states. Filling the empty level from the next shell above is performed dissipating the energy as electromagnetic radiation. The energy difference determines the frequency and wavelength by

$$\Delta E = h\nu = hc/\lambda. \tag{7.13}$$

Ionizing the K-shell delivers characteristic Kα- and Kβ-radiation.

The generation of "Bremsstrahlung" follow the Duane–Hunts law

$$h\nu_{\text{max}} = hc/\lambda_{\text{min}}. \tag{7.14}$$

Here, $\nu_{\text{max}} = 242 \times 10^{12}$ (Hz). U_{a} (V), with U_{a} the electron acceleration or anode voltage, or $\lambda_{\text{min}} U_{\text{a}} = 12.4 \times 10^{-10}$ (kV m) the lower wavelength limit.

The X-ray lines can be calculated using

$$E_{\text{n}} = -0.5(e^2/4\pi\varepsilon_0)(1/r_H)(Z^2/n^2) = -13.6\ (\text{eV}) \cdot (Z/n)^2. \tag{7.15}$$

A tungsten anode $Z = 74$, k-shell $n = 1$, $eU_{\text{a}} = 745.5$ (keV), delivers by k-shell ionisation and filling of the hole: $h\nu = E_{\text{L}} - E_{\text{K}} = 55.9$ (keV), or $\lambda_{\text{K}\alpha} = 0.022$ nm.

7.7.3 X-ray Filters

Using metal foils of special atomic weight Z, and making use of the fact that X-rays are absorbed in atoms with an effectivity depending on the shell energies. At X-ray energies close to the shell energies, the photoeffect by X-rays changes with a jump, which allows lower energies to pass the filter, but higher energies to be blocked.

X-rays are absorbed by matter of thickness x according to the exponential law of mass absorption

$$I(x) = I_0 \exp(-\mu x). \tag{7.16}$$

For technical applications the mass absorption coefficient μ/ρ and the mass thickness ρx are used, with ρ being the density (kg/cm^3).

According to the mass absorption coefficient μ (cm^{-1}), lead and concrete are used to absorb – shield X-rays in all applications. Here the half-dose thickness $d_{1/2} = \ln 2/\mu = 0.693 x_e$, with $x_e = 1/\mu$ the penetration depth. For example, lead: density $\rho = 11.3$ (g/cm^3), linear attenuation coefficient $\mu = 56.5$ (cm^{-1}), penetration depth $x_e = 0.18$ (mm).

Water has at the same energy the mass attenuation coefficient of $\mu/\rho = 0.2$ (cm^3/g), and the penetration depth $x_e = 5$ (cm). A water layer thickness of $4x_e = 20$ cm absorbs 98.2% of the energy used. The attenuation coefficient is $\approx Z5$.

7.7.4 X-ray Dosimetry

The biological effect of X-rays depends on the number of radiation damages like ionisation events. The dose is proportional to the energy loss divided by the mass. The unit is Gy (Gray).

$$D_e = \Delta E/m \text{ (J/kg)}. \tag{7.17}$$

The ionisation dose $D_i = \Delta Q/m$ (A s/kg).

Formerly the units rad and Roentgen were used:

$$1 \text{ rad} = 10 \text{ mGy} = 10^{-2} \text{ (J/kg) and } 1 \text{ Röntgen} = 2.578 \times 10^{-4} \text{ (A s/kg)}.$$

The mean energy for forming an ion pair is in the range of $w = 27$–34 eV.

The absorption of the energy ΔE produces $\Delta N = \Delta E/w$ ion pairs. With the ionisation constant $k_j = w/e$, one obtains

$$Df = \Delta E/m = w\Delta N/m = w\Delta Q/em = wD_j/m = k_j D_j.$$

In air: D_e (air) = 33.7 V D_j (air).

There is no heating effect in the human body by X-rays. For example, a dose of 10 Gy delivers 10 J to each kg of mass, which converted to heat results in a temperature rise of the body by 2.4 mK, which is negligible. However, 10 Gy is a lethal dose. The natural radiation dose/year is 1 mGy. Tumors are treated in radiation therapy using 30 to 100 Gy, with the radiation applied locally. One X-ray image needs 1–100 mGy.

For radiation – therapy the biological effect depends on the nature of the radiation. Those are classified with factors q describing the relative biological effectivity compared to photons ($q = 1$) and electrons ($q = 1$).

The dose-equivalent $D_q = qD_e$ (J/kg = Sv); Sv = Sievert.

A dose-equivalent $D_\mathrm{q} = 10$ Sv corresponds to a radiation of $q = 10$ radiating to tissue with an energy dose $D_\mathrm{e} = 1$ Gy, the biological effect, however, corresponds to an X-ray dose in tissue $D_\mathrm{e} = 10$ Gy.

The q-factor for differing radiations are: X-rays, gamma-rays, electrons have $q = 1$; thermal neutrons $q = 3$; fast neutrons (1 MeV) $q = 10$; α-particles (10 MeV); protons (10 MeV) $q = 10$; heavy ions $q = 20$.

7.7.5 X-ray Tubes

Conventional X-ray Tubes

The schematic of an X-ray diffractometre for crystal structure analysis is shown in Fig. 7.25, showing also the schematic of the X-ray tube. The electrons from a thermal electron emitter are accelerated directly to the anode. The beam spot emits then characteristic and Bremsstrahlung. The characteristic radiation of narrow wavelength distribution is used to irradiate the sample. The filter blocks out the Bremsstrahlung to use the characteristic radiation for spectroscopy. A detector, generally a Geiger–Müller detector, is used to count the X-ray photons and amplify the signal by the gas discharge in the counter tube.

The angular position of crystal reflections are governed by the Laue equations and allow to determine the electron density distribution of the elementary cell of the crystal.

X-rays interact with matter by 4 effects. The photo-effect with ionisation; the photo-absorption coefficient is proportional to $Z^4\ E_{-3}$. The Compton effect is inelastic scattering of a photon at a weakly bound outer shell electron of an atom and results in a changed wavelength and changed direction. Pair formation at energy above 1.022 MeV, a photon gets materialized to one electron and one positron. With

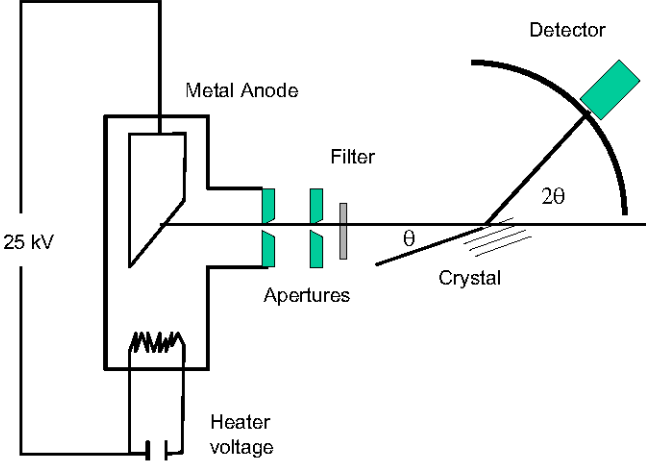


Fig. 7.25. Schematic of an X-ray diffractometer for crystal structure analysis

the Rayleigh scattering it generates radiation perpendicular to the beam which generates dose exposure outside the beam direction.

Figure 7.26 shows a technical tube as supplied by Oxford, showing the electron emitter thermal housing, the anode, and the beryllium window to pass the radiation generated at the fixed or rotating anode to air.

Typical data for a microfocus X-ray tube for analytics are: electron energy 90 kV, beam current 2 mA, total power 80 W, 4 mm anode to object distance, true round spot, grounded target = high power, 80 W, air-cooled. The tube is ideal for X-ray imaging. The tube is designed for those applications where flux density is important. With a true round focal spot and high power loading, it is ideally suited for use with X-ray optics.

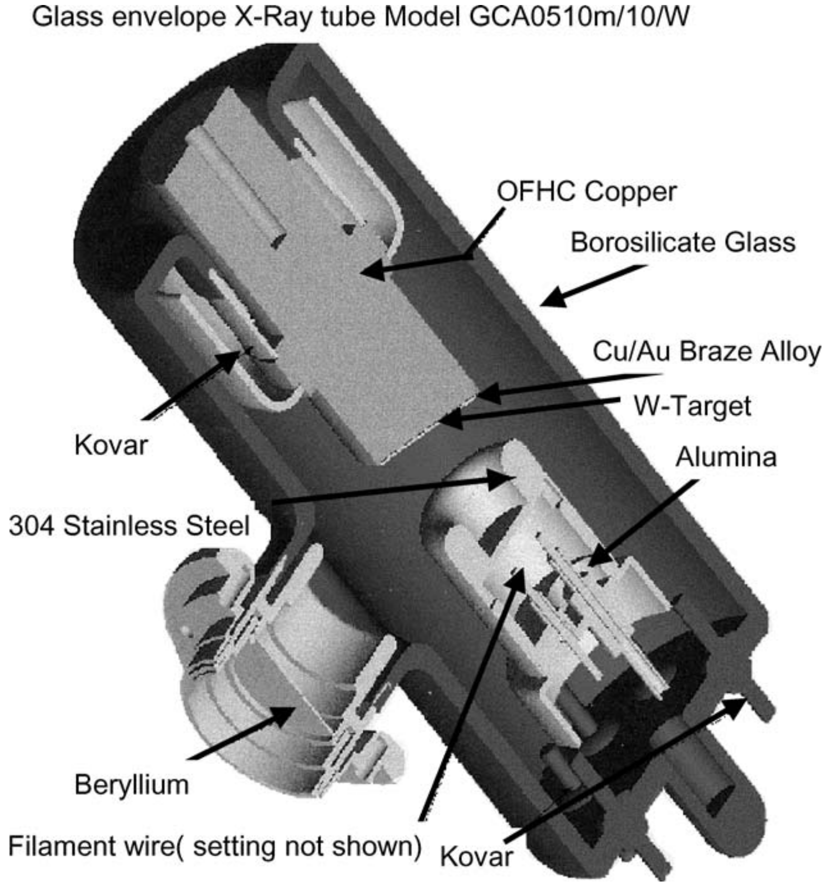


Fig. 7.26. Technical tube as supplied by Oxford, showing the electron emitter thermal housing, the anode, and the beryllium window to pass the radiation generated at the fixed or rotating anode to air

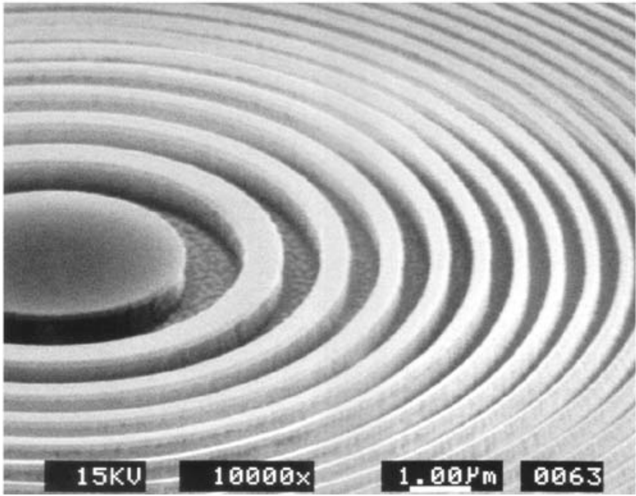


Fig. 7.27. Fresnel zone plate, which is a diffractive grating with variable grating constant, used for focusing

X-ray Optics. Fresnel zone plates, Wolter telescopes, and aspherically polished X-ray mirrors, coated with multiplayer mirrors for high reflection, are the only possibilities to focus X-ray radiation and to generate X-ray images. The semiconductor industry hopes to build chips with the novel Extreme Ultraviolet (EUV) radiation at 13.6 nm. This will help to produce 35 nm CD computer-chip structures in the next 5 years.

Figure 7.27 shows a Fresnel zone plate, which is a diffractive grating with variable grating constant. This has the effect that the first order of the diffracted beam hits in one spot. Therefore, imaging with such X-ray optical elements is possible [96].

7.7.6 Synchrotrons

In physics, the synchrotron radiation, emit electromagnetic radiation by high-speed electrons spiralling along the lines of force of a magnetic field. Depending on the electron's energy and the strength of the magnetic field, the maximum intensity will occur as radio waves, visible light, or X-rays. The emission is a consequence of the constant acceleration experienced by the electrons as they move in nearly circular orbits; according to Maxwell's equations, all accelerated charged particles emit electromagnetic radiation. Although predicted much earlier, the synchrotron radiation was first observed as a glow associated with protons orbiting in high-energy particle accelerators, such as the synchrotron. In astronomy, the synchrotron radiation has been suggested as the mechanism for producing strong celestial radio sources like the Crab Nebula. The synchrotron radiation is employed in a host of applications ranging from solid state physics, lithography for semiconductor industry, LIGA a method to build micromechanical high aspect ratio structures by dell X-ray lithography and

electroforming in the resist mould, and last but not least in medicine. As excellent producers of X-rays, synchrotron sources offer unique probes of the semiconductors that lie at the heart of the electronics industry. Both ultraviolet radiation and X-rays generated by synchrotrons are also employed in the treatment of diseases, especially certain forms of skin cancer. Soft and hard X-rays are used in medical device sterilization, mail sanitation, food pasteurisation and sterilization. Since investment and operation costs for the same quality of irradiation and production capacity are equal for gamma- and X-ray sterilization, the industry has to answer to the challenges, not any longer to rely on cobalt-hard X-ray emitters from nuclear decay in materials, but employ a switchable X-ray source like miniaturized sources in brachiotherapy and other medical internal and external applications, e.g. cancer radiation therapy.

7.7.7 X-ray Detection

X-ray film being specially sensitised for X-rays, is produced in large sheets and is exposed through the object, tooth, body, mechanical structure, etc.

More modern is digital X-ray detection using arrayed X-ray detectors in charged particle devices. These detectors are preferably used in technology, medicine, and astronomy. CXDI-11, an X-ray imaging system, uses the Canon 43 cm $\times$ 43 cm LANMIT amorphous Silicon Sensor System and reaches excellent image quality with more than 7 million detector elements built from amorphous silicon, each of them delivering 4096 grey levels. This allows to compensate for over- and underexposure, and reduces the number of X-ray exposures and repeated exposures.

Immediate image presentation is an especially time saving feature. No costly silver containing plates are required. Images can be transferred and stored, archived, communicated to others electronically, etc.

7.7.8 Applications

Medical: radiography, bone densitometry, X-ray imaging; on-line elemental analysis, particle size analysis; density measurements, process control, inspection, on-line process control; X-ray spectroscopy for materials analysis and composition control, e.g. factories producing concrete are controlled with automated analysis X-ray diffractometers, which determine the concentrations of different materials to be filled into the burning oven.

Electronics/semiconductor: solder joints quality control, MEMS structure control after assembly by bonding. Life sciences: protein and large molecule crystallography, drug discovery and research, “burning” computer chip designs into metal wafers, studying molecule shapes and protein crystals, analyzing chemicals to determine their composition, watching living cells as they react to drugs, fluorescence studies, semiconductor material analysis and structural studies, geological material analysis. Detecting breast cancer tumors is refined by a novel technique that creates three-dimensional X-rays that make tumors much easier to see. Full-field digital tomosynthesis, or TOMO, takes multiple pictures at different angles as the X-ray tube rotates in an arc.

7.7.9 Future Developments: Miniaturized X-ray Tubes

To employ X-ray in medical applications miniaturized tubes are investigated and produced [97, 98]. Micropinches generated in a low inductance vacuum spark (micropinch discharge) are intensive sources of soft X-rays [99]. Brachyo-therapy needs a source of 1 mm diameter and 3 mm in length. Such a source with a thermal emitter and 25 kV delivering 30 μA of current is developed by Xoft [100, 101].

References

[1] G. Kirchhoff, R. Bunsen, Chemische Analyse durch Spectralbeobachtungen, Poggend. Annal. Bd. **110**, S161–169 (1860)
[2] M. Planck, Über das Gesetz der Energieverteilung im Normalspektrum, Drudes Annalen 553 (1901)
[3] J.C. de Vos, A new determination of the emissivity of tungsten ribbon, Physics **XX**, 690–714 (1954)
[4] A.R. Striganov, N.S. Sventitskii, *Tables of Spectral Lines of Neutral and Ionized Atoms* (Plenum, New York, Washington, 1968) (translated from Russian)
[5] Atomic data can also be found on the internet, e.g. at web sites of national bureaus of standard. http://physics.nist.gov/cgi-bin/AtData/main_asd?XXR6000q3000qHgqI
[6] K.P. Huber, G. Herzberg, *Molecular Spectra and Molecular Structure IV Constants of Diatomic Molecules* (Van Nordstrand Reinhold Company, New York, 1979)
[7] W.L. Wiese, *Atomic Transition Probabilities*, vol. I: H through Ne. vol. II: Na through Ca (National Burea of Standards, National Standard Reference Data Series NSRDS-NBS 4, Washington, 1966)
[8] H.S.W. Massey, E.H.S. Burhop, H.B. Gilbody, *Electronic and Ionic Impact Phenomena*, vol. I–V (Clarendon Press, Oxford, 1969–1974)
[9] T.H. Maiman, Stimulated optical radiation in ruby, Nature **187**, 493–494 (1960)
[10] A. Javan, W.R. Bennett Jr., D.R Herriott, Population inversion and continuous optical maser oscillation in a gas discharge containing a He–Ne mixture, Phys. Rev. Lett. **6**, 106 (1961)
[11] C.H. Townes, *How the Laser Happened, Adventures of a Scientist* (Oxford University Press, New York, Oxford, 1999)
[12] D. Eastham, *Atomic Physics of Lasers* (Taylor & Francis, London, Philadelphia, 1989)
[13] B.A. Lengyel, *Lasers*, 2nd edn (Wiley-Interscience, New York, 1971)
[14] R. Beck, W. Englisch, K. Gürs, *Table of Laser Lines in Gases and Vapors* (Springer, Berlin, Heidelberg, New York, 1980)
[15] M.J. Weber, *CRC Handbook of Laser Science and Technology*, vol. II (Gas Lasers) (CRC Press, Boca Raton, 1982)
[16] J.G. Eden (ed.), Selected papers on *Gas Laser Technology*, SPIE Milestone Series, volume MS 159 (SPIE Press, Bellingham, 2000)
[17] R.W. Waynant, M.N. Edinger (ed.), Selected papers on *UV, VUV, and X-Ray Lasers*, SPIE Milestone Series, volume MS 71 (SPIE Press, Bellingham, 1993)
[18] D.L. Matthews, P.L. Hagelstein, M.D. Rosen, M.J. Eckart, N.M. Ceglio, A.U. Hazi, H. Medecki, B.J. MacGowan, J.E. Trebes, B.L. Whitten, E.M. Campbell, C.W. Hatcher, A.M. Hawryluk, R.L. Kauffmann, L.D. Pleasance, G. Rambach, J.H. Scofield, G. Stone, T.A. Weaver, Demonstration of a soft X-ray amplifier, Phys. Rev. Lett. **54**, 110–113 (1985)

[19] J.J. Rocca, V. Shlyaptsev, F.G. Tomasel, O.D. Cortazar, D. Hartshorn, J.L.A. Chilla, Demonstration of a discharge pumped table-top soft-x-ray laser, Phys. Rev. Lett. **73**, 2192–2195 (1994)

[20] A.G. Molchanov, I.A. Poluektov, Y.M. Popov, The possibility of the generation of vacuum ultraviolet radiation by electron excitation of inert-gas crystals, Sov. Phys. Solid State **9**, 2655 (1968)

[21] N.G. Basov, O.V. Bogdankevich, V.A. Danilychev, A.G. Devyatkov, G.N. Kashnikov, N.P. Lantsov, Cathodoluminsecence of solid xenon in the ultraviolet region of the spectrum, JETP Lett. **7**, 317–318 (1968)

[22] P.W. Hoff, J.C. Swingle, Ch.K. Rodes, Demonstration of temporal coherence, spatial coherence, and threshold effects in the molecular xenon laser, Opt. Commun. **8**, 128–131 (1973)

[23] J.B. Gerardo, A.W. Johnson, High-pressure xenon laser at 1730 Å, IEEE J. Quantum Electron. **Q-9**, 748–755 (1973)

[24] C.K. Rhodes (ed.), *Excimer Lasers, Topics in Applied Physics*, vol. 30, 2nd enlarged edn (Springer, Heidelberg, New York, Tokyo, 1984)

[25] C.A. Brau, page 87 in [24]

[26] S.E. Bodner, et al., High gain direct drive target design for laser fusion, Phys. Plasmas **7**, 2298 (2000)

[27] F. Hegeler, M.C. Myers, M. Friedman, J.D. Sehtian, S.B. Swanekamp, D.V. Rose, D.R. Welch, Efficient electron beam deposition for repetitively pulsed krypton fluoride lasers, in *Proceedings of the 14th International Conference on High-Power Particle Beams 2002*, Albuquerque, NM, 23–28 June 2002, p. 357

[28] M. Friedman, S. Swanekamp, S. Obenschain, Y. Chan, L. Ludeking, D. Smithe, Stability of large area electron beam diodes, Appl. Phys. Lett. **77**, 1053 (2000)

[29] F. Hegeler, private communication

[30] G. Marowsky, R. Cordray, F.K. Tittel, W.L. Wilson, C.B. Collins, Appl. Phys. Lett. **33**(1), 60 (1978)

[31] S.J. Smith, E.M. Purcell, Phys. Rev. **92**, 1069 (1953)

[32] T. di Francia, Il Nuovo Cimento **16**, 1085 (1960)

[33] Vermont Electronics homepage. www.vermontphotonics.com

[34] National Research Council, *The Free Electron Laser* (National Academy Press, Washington, 1982); C. Pellegrini, Encyclopedia of Applied Physics **8**, 353 (1994)

[35] M. Goldstein, J.E. Walsh, M.F. Kimmit, J. Urata, C.L. Platt, Appl. Phys. Lett. **71**, 452 (1997)

[36] J.H. Brownell, J. Walsh, G. Doucas, Phys. Rev. E **57**, 1075 (1998)

[37] J. Urata, Spontaneous and stimulated Smith–Purcell radiation experiments in the far infrared, Thesis, Dartmouth College, Hanover, USA, 1997

[38] O. Haeberle, P. Rullhusen, J. Salome, N. Maene, Phys. Rev. E **49**(4), 3340 (1994)

[39] J. Urata, M. Goldstein, M.F. Kimmit, C. Platt, J.E. Walsh, Phys. Rev. Lett. 516 (1998)

[40] J.E. Walsh, J.H. Brownell, J.C. Swartz, J. Urata, M.F. Kimmit, Nucl. Instr. Methods A **429**, 457 (1999)

[41] F. Floreani, H.W. Koops, W. Elsässer, Concept of a miniaturised free-electron laser with field emission source, Nucl. Instrum. Methods Phys. Res. A **483**, 488–492 (2002)

[42] M.J. Moran, X-ray generation by the Smith–Purcell effect, Jpn. J. Appl. Phys. **27**, 408–412 (1988)

[43] M. Schlessinger, *Infrared Technology Fundamentals*, 2 edn. rev. and expanded (Dekker, New York, 1995)

[44] B. Bhanu, I. Pavlidis, *Computer Vision Beyond the Visible Spectrum* (Springer, London, 2005)

[45] K. Stahl, G. Miosga, *Infrarottechnik: Grundlagen, Strah lungssender und Detektoren, Infrarotbildaufnahmen und -wiedergabe, Fernmeßverfahren*, 2 edn (Hüthig, Heidelberg, 1986)
[46] R. Borchert, W. Jubitz, *Infrarotstrahler: zur Erwärmung, Trocknung und Aushärtung, für die Lack-, Textil- und andere Industrien, für die Strahlentherapie*, vol. 5 (Schriftenreihe des Verlages Technik, Berlin, 1951)
[47] E. Martinet, F. Luc, E. Rosencher, P. Bois, E. Costard, S. Delaitre, E. Bockenhoff, in *Intersubband Transitions in Quantum Wells*, ed. by E. Rosencher, B. Vinter, B. Levine (Plenum, New York, 1992), p. 299 (in France)
[48] S.D. Gunapala, S.V. Bandara, *Quantum Well Infrared Photodetector (QWIP) Focal Plane Arrays*. Semiconductors and Semimetals series, vol. 62 (1999)
[49] Shen, et al., Appl. Phys. Lett. **83**, 3118 (2003)
[50] M. Bykhovskaia, et al., Theor. Chem. Acc. **106**, 22 (2001)
[51] P. Han, et al., Opt. Lett. **25**, 242 (2000)
[52] T. Löffler, et al., Opt. Express **9**(12), 616–621 (2001)
[53] Hasegawa, et al., Remote identification of surface texture, Appl. Phys. Lett. **83**, 3996 (2003)
[54] S. Bollaert, et al., Fmax of 490 GHz metamorphic In Al As/In Ga As HEMT's, Electron. Lett. **38**, 389 (2002)
[55] W. Knap, et al., Terahertz emission by plasma waves in 60 nm gate High Electron Mobility Transistor, Appl. Phys. Lett. **48**, 2331 (2004)
[56] R. Köhler, Quantum cascade laser, Nature **417**, 156 (2002)
[57] M. Perrin, et al., JOPA Special Issue on metamaterials
[58] Astafiev, et al., Appl. Phys. Lett. **79**, 1199 (2001)
[59] J.R. Coaton, A.M. Marsden, *Lamps and Lighting* (Arnold and Contributors, 1997)
[60] W. Elenbaas, *Light Sources* (Crane, Russek & Company, Inc., 1972)
[61] M. Born, T. Jüstel, Phys. J. (February), 43 (2003)
[62] J.A. Samson, D.L. Ederer (eds.), in *Vacuum Ultraviolet Spectroscopy*, vol. 1 and vol. 2, Experimental Methods in the Physical Science, vol. 31 (Academic, San Diego, London, Boston, New York, 1998)
[63] J.A.R. Samson, *Techniques of Vacuum Ultraviolet Spectroscopy* (Wiley, New York, London, Sydney, 1967)
[64] Deuterium lamps as transfer standards for spectral radiance in the spectral range 115 nm–350 nm, Physikalisch Technische Bundesanstalt, Inst. Berlin, Abbestr. 2-12, Berlin, Germany, Final report for contract 1405/1/0/002/83/11–BCR–D(30), Sept. 29 (1987); P.J. Key, D.H. Nettleton, Deuterium lamps as transfer standards for spectral radiance measurements, National Physical Laboratory, Teddington, Middlesex TW 11 OLW, UK, Commission of the European Communities BCR information, Applied Metrology Contract No. 615/1/0/002/80/1-BCR-UK(30)
[65] J.M. Bridges, W.R. Ott, Vacuum ultraviolet radiometry. 3: The argon mini-arc as a new secondary standard of spectral radiance, Appl. Opt. **16**, 367–375 (1977)
[66] J.A. Samson, D.L. Ederer (eds.), in *Vacuum Ultraviolet Spectroscopy*, vol. I, chap. 3.4, Experimental Methods in the Physical Science, vol. 31 (Academic, San Diego, London, Boston, New York, 1998)
[67] R.A.B. Zijlmans, J.H. van Helden, D.C. Schram, R. Engeln, The cascaded arc – a novel bright light source for sensitive broadband absorption spectroscopy, P-85, page 355 in Light sources, 2004; G. Zissis, in *Proceedings of the 10th International Symposium on the Science and Technology of Light Sources*, Toulouse, France, 18–22 July 2004, IOP Conference Series Number 182 (Institute of Physics Publishing, Bristol and Philadelphia, 2004)

[68] W. Weizel, Chr. Füchtbauer, Kernschwingungen im Bandenspektrum des Heliums, Z. Phys. **44**, 431–434 (1927)
[69] Y. Tanaka, A.S. Jursa, F.J. LeBlanc, Continuous emission spectra of rare gases in the vacuum ultraviolet region. II. Neon and helium, J. Opt. Soc. Am. **48**, 304–308 (1958)
[70] M.V. McCusker, The rare gas excimers, Chapter 3, in *Excimer Lasers*, 2nd edn., ed. by Ch. Rhodes. Topics in Applied Physics, vol. 30 (Springer, Berlin, Heidelberg, New York, Tokyo, 1984)
[71] J. Wieser, M. Salvermoser, L.H. Shaw, A. Ulrich, D.E. Murnick, H. Dahi, Lyman-alpha emission via resonant energy transfer, J. Phys. B **31**, 4589–4597 (1998)
[72] P. Kurunczi, H. Shah, K. Becker, Hydrogen Lyman-α and Lyman-β emissions from high-pressure microhollow cathode discharges in Ne-H_2 mixtures, J. Phys. B **32**, L651–L658 (1999)
[73] A. El-Dakrouri, J. Yan, M.C. Gupta, M. Laroussi, Y. Badr, VUV emission from a novel DBD-based radiation source, J. Phys. D **35**, L109–L114 (2002)
[74] T. Efthimiopoulos, D. Zouridis, A. Ulrich, Excimer emission spectra of rare gas mixtures using either a supersonic expansion or a heavy-ion-beam excitation, J. Phys. D **30**, 1746–1754 (1997)
[75] A. Morozov, B. Krylov, G. Gerasimov, A. Arnesen, R. Hallin, A study of atomic and molecular energy transfer channels in Kr-Xe gas mixtures excited with radio frequency discharges, J. Phys. B **35**, 1929–1940 (2002)
[76] A. Schmitt, K. Wendt, A. Ulrich, Electron beam pumped UV light source for resonant spectroscopy of atoms, Maier-Leibnitz-Laboratorium für Kern- und Teilchenphysik, LMU and TU Munich, Annual Report 2002, p. 62
[77] B. Eliasson, U. Kogelschatz, UV excimer radiation from dielectric-barrier discharges, Appl. Phys. B **46**, 299–303 (1988)
[78] A.N. Panchenko, E.A. Sosnin, V.F. Tarasenko, Improvement of output parameters of glow discharge UV excilamps, Opt. Commun. **161**, 249–252 (1999)
[79] A. El-Habachi, K.H. Schoenbach, Emission of excimer radiation from direct current, high-pressure hollow cathode discharges, Appl. Phys. Lett. **72**, 22–24 (1998)
[80] M. Salvermoser, D.E. Murnick, High-efficiency, high-power, stable 172 nm xenon excimer light source, Appl. Phys. Lett. **83**, 1932–1934 (2003)
[81] J. Wieser, D.E. Murnick, A. Ulrich, H.A. Huggins, A. Liddle, W.L. Brown, Vacuum ultraviolet rare gas excimer light source, Rev. Sci. Instrum. **68**, 1360–1364 (1997)
[82] M.I. Lomaev, A.N. Panchenko, V.S. Skakun, E.A. Sosnin, V.F. Tarasenko, M.G. Adamson, B.R. Myers, F.T. Wang, Excilamp producing up to 130 W of output power and possibility of its application, Laser Part. Beams **15**, 339–345 (1997)
[83] U. Kogelschatz, J. Salge, High-pressure plasmas: dielectric-barrier and corona discharges – properties and technical applications, in *Low Temperature Plasma Physics*, ed. by R. Hippler, S. Pfau, M. Schmidt, K.H. Schoenbach (Wiley-VCH, Berlin, Weinheim, New York, Chichester, Brisbane, Singapore, Toronto, 2001)
[84] G. Zissis (ed.), Light sources, 2004, in *Proceedings of the 10th International Symposium on the Science and Technology of Light Sources*, Toulouse, France, 18–22 July 2004, IOP Conference Series Number 182 (Institute of Physics Publishing, Bristol, Philadelphia, 2004)
[85] J. Meichsner, D. Loffhagen, H.-E. Wagner (eds.), in *XXVI International Conference on Phenomena in Ionized Gases (ICPIG) 2003*, Greifwald, Germany, July 15–20, 2003 (4 volumes), ISBN 3-00-011689-3
[86] V.F. Tarasenko, G.V. Mayer, G.G. Petrash (eds.), in *Int. Conf. On Atomic and Molecular Pulsed Lasers*, vol. V, Tomsk, Russia, Sept. 15–19, 2003; Proc. SPIE, vol. 5483, SPIE Press, Bellingham, and earlier proceedings of this conference series

[87] K.H. Schoenbach, R. Verhappen, T. Tessnow, F.E. Peterkin, W.W. Byszewski, Appl. Phys. Lett. **68**, 13–15 (1996)
[88] M. Salvermoser, D.E. Murnick, Efficient, stable, corona discharge 172 nm xenon excimer light source, J. Appl. Phys. **94**, 3722 (2003)
[89] M. Salvermoser, D.E. Murnick, Stable high brightness radio frequency driven micro-discharge lamps at 193 (ArF^*) and 157 nm (F_2^*), J. Phys. D **37**, 180–184 (2004)
[90] C.A. Brau, Rare gas halogen excimers, Chapter 4 (4.3), in *Excimer Lasers*, 2nd edn., ed. by Ch.K. Rhodes. Topics in Applied Physics, vol. 30 (Springer, Berlin, Heidelberg, New York, Tokyo, 1984)
[91] P. Lenard, Ueber Kathodenstrahlen in Gasen von atmosphärischem Druck und im äussersten Vacuum, Annal. Phys. Neue Folge Band **51**, 225–267 (1994)
[92] R. Gauvin, P. Hovongton, D. Drouin, P. Horny, H. Demers, CASINO program (public domain) Université de Sherbrooke, Québec, Canada. http://www.gel.usherb.ca/casino/
[93] M. Salvermoser, D.E. Murnick, J. Wieser, A. Ulrich, Energy flow and excimer yields in continuous wave rare gas-halogen systems, J. Appl. Phys. **88**, 453–459 (2000)
[94] F. Mühlberger, J. Wieser, A. Ulrich, R. Zimmermann, Single photon ionization (SPI) via incoherent VUV-excimer light: robust and compact time-of-flight mass spectrometer for on-line, real-time process gas analysis, Anal. Chem. **74**, 3790–3801 (2002)
[95] A. Fedenev, A. Morozov, R. Krücken, S. Schoop, J. Wieser, A. Ulrich, Applications of a broadband electron-beam pumped XUV radiation source, J. Phys. D **37**, 1586–1591 (2004)
[96] G. Schmahl, D. Rudolf (eds.), *X-ray Optics and Spectroscopy* (Plenum, New York, 1987)
[97] H.H. Busta, J.M. Chen, Z. Shen, K. Jansen, S. Rizkowski, J. Matey, A. Lanzillotto, Characterization of electron emitters for miniature X-ray sources, J. Vac. Sci. Technol. B **21**(1), 344 (2003)
[98] H.H. Busta, in *Vacuum Microelectronics,* Chap. 7, ed. by W. Zhu (Wiley, New York, 2001), p. 315
[99] O.G. Semyonov', A.E. Gurey, A.P. Kanavin, A.A. Tikhomirov, Applications of micropinch x-ray source, J. Vac. Sci. Technol. B: Microelectron. Nanometer Struct. **19**(4), 1235–1240 (2001)
[100] Xoft microTube, Inc., homepage
[101] R. Nath, L.L. Anderson, G. Luxton, et al., Dosimetry of interstitial brachytherapy sources: Recommendations of the AAPM Radiation Therapy Committee Task Group 43, Med. Phys. **22**, 209–234 (1995)

8

Particle Accelerators

M. Seidel and K. Zapfe

Particle accelerators are facilities for the generation and acceleration of charged particle beams to high energies. Accelerators are used since 1920 to investigate the structure of matter. With a particle beam of energy E_b it is possible to probe structures with typical dimensions of the *de Broglie* wavelength $\lambda \approx hc/E_\mathrm{b}$ in scattering experiments. The higher the provided beam energy the smaller are the experimentally accessible dimensions. Another successful concept is the production of new particles by colliding two opposing particle beams with well-defined centre of mass energy. With increasing beam energy heavier particles can be produced. Many important discoveries in particle physics where achieved in the twentieth century using accelerators. As an originally disturbing side effect, intense UV- and X-ray radiation is produced in the bending magnets of electron accelerators. This effect was predicted already in 1898 and observed in an accelerator in 1947. Since the 1960s this so-called synchrotron radiation from accelerators is used for experimental research with increasing popularity.

Nevertheless, the research in the field of elementary particle physics has been the driving force behind the development of particle accelerators over the past 80 years. With the discoveries of new particles at higher and higher masses the energies required by the particle physics experiments were raised enormously. While the first accelerators generated energies of a few thousand electron-volts and had enough space on a laboratory table, present-day accelerators reach hundreds of billion electron-volts (GeV) with dimensions of several kilometers. But not only the beam energies were raised. The diminishing probabilities for interesting collision events at higher energies required also a steady increase of the beam densities to obtain reasonable event rates. This was achieved by raising the beam currents and decreasing the beam cross-sections at the interaction points. Naturally, these developments resulted in higher investment and operation costs and also outstanding operational complexity of modern accelerator systems (see, e.g. Fig. 8.1). A large facility based on two linear electron/positron accelerators with beam energies of 250 GeV and beyond, and a length exceeding 30 km is presently being discussed as the next

Fig. 8.1. View of the HERA accelerators, an electron/proton collider facility with a circumference of 6300 m and a bending radius of 800 m. The superconducting magnets for the 920 GeV proton beam are installed on *top* of the normal conducting magnets for the 27.5 GeV electron beam

step in high energy accelerator development. The construction of such a facility is only possible as an international project.

Besides the described development at the energy and interaction rate frontier, there is also an increasing demand on medium and small size accelerators with very specific requirements. Accelerators are used for medical applications, e.g. for the production of radioactive isotopes and cancer therapy using proton and ion beams. High power proton accelerators are utilized, for example, to drive neutron spallation sources or pion/muon production targets. Another application with a dynamic development over the recent years is the generation of synchrotron radiation for research purposes. A number of dedicated synchrotron radiation sources have been built worldwide with circumferences of several hundred meters and energies of a few GeV. The applications of such light sources are widely spread and cover, for example, material, biological and medical research. The technical challenges of a synchrotron light source are different from a particle physics research accelerator. Synchrotron radiation sources have optimized magnet arrangements that allow for the generation of beams with small dimensions (emittance) to produce the required high intensity light fans and very stable beam orbits. A relatively new development are free electron lasers (FEL's) that provide even higher brilliance and coherence

of the generated UV- and X-ray beams. Free electron lasers produce coherent radiation from high density beam particle packets, also called bunches, that develop a micro-structure due to intense interaction with the produced laser light. This micro-structure exhibits the periodicity of the radiation wavelength. The radiation intensity generated by micro-modulated bunches scales quadratically with their charge and is therefore by orders of magnitude stronger than incoherent radiation that is linearly proportional to the charge.

The vacuum system is an essential component of an accelerator. In storage rings with circulating particle beams the residual gas density determines the beam lifetime. For particle physics and synchrotron light applications the gas density is a crucial parameter for the beam induced background at the experiments. Typically residual gas pressures of the order of 10^{-9} mbar or below are required and routinely established in ultra-high vacuum systems (UHV). This means that the required gas densities are at least 12 orders of magnitude lower than under normal atmospheric conditions. In such low gas densities a proton undergoes an inelastic collision every 7×10^{11} km, or 30 days travel at the speed of light! A broad range of advanced techniques, procedures, production methods and instruments has been developed in the field of vacuum technology for particle accelerators to meet these requirements.

This chapter will provide the reader with a basic review of the different types of particle accelerators, the mechanisms of acceleration and beam transport, the definition of essential particle beam properties and the physics of synchrotron radiation. In the context of this book we will concentrate then on the requirements and the technical realization of accelerator vacuum systems.

8.1 History and Types of Accelerators

The aim of this section is to give a brief review of the different types of accelerators, their working principles and their historical development. Readers interested in technical details are referred to the classical textbooks [1–3] and those with an interest in the historical development will find more information in [4].

8.1.1 Static Accelerators

Acceleration of charged particles using static electrical fields was the first applied acceleration principle. It also represents the simplest possibility from the technical point of view. Particles with charge q that traverse an electrical field of total voltage U gain an energy of $\Delta E = qU$, commonly expressed in electron volts (eV). The technical challenge lies in the generation and reliable operation of high electrical voltages. One of the first high voltage accelerators was the *Cockroft–Walton* generator, named after its inventors. It was used in 1932 to split Li atoms with 400 keV protons [5]. For this experiment Cockroft and Walton received the Nobel Prize in 1951. The base unit of this high voltage generator is a voltage doubling circuit that employs transformers, diode tube rectifiers and capacitors. Cascading of the base

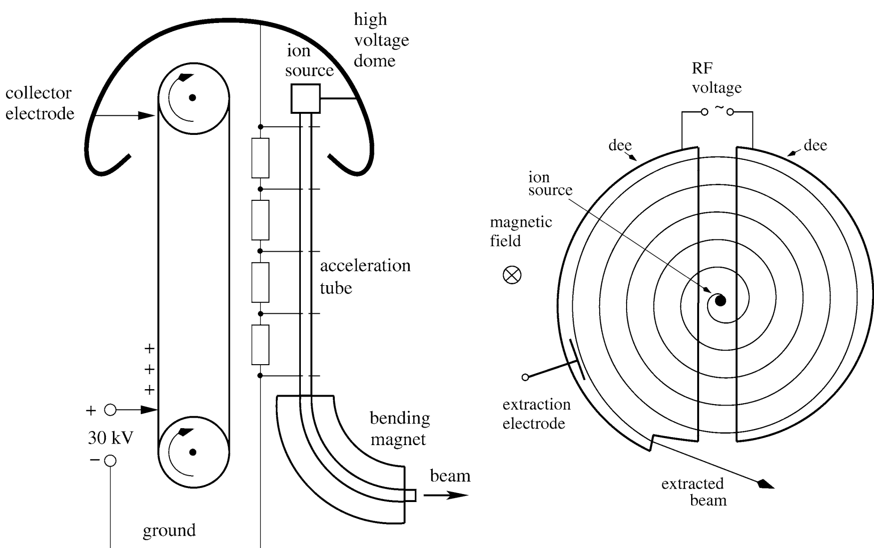


Fig. 8.2. Examples of static and resonant acceleration. *Left*: van de Graaff accelerator. *Right*: cyclotron (a magnetic dipole field points *into the plane*)

unit allows the generation of voltages in the range of 1–2 MV. The acceleration of the charged particles is done in straight drift tubes at different voltage potentials.

Another classical apparatus for the production of high static voltages is the *van de Graaff generator* (Fig. 8.2). A circulating rubber band continuously transports charge into a dome-shaped high voltage terminal where it is accumulated. The particle source is installed inside this dome. The first devices of this type reached accelerating voltages of 1.5 MV, later more than 20 MV were achieved. A modification of the van de Graaff generator is the *tandem accelerator* which allows to double the beam energy. In the first section negatively charged ions are accelerated, then the electrons are stripped from the ions and the same voltage is used to accelerate the particles, which are now positively charged, a second time. Further energy increase was achieved by installing the critical components in a pressurized freon atmosphere which raised the sparking threshold. The most important advantages of the van de Graaff type accelerators are the sharply defined energy of the beam and the possibility of wide energy variations. Accelerators of this type are still in use. An example is the *Vivitron* in Strasbourg [6] that provides energies up to 20 MeV.

8.1.2 Circular Resonant Acceleration

Static high voltage accelerators are limited by the technical feasibility of high voltage generation. It was quickly realized that resonant acceleration by applying an accelerating field many times is more efficient and powerful. This is achieved, for example, by recirculating the beam and alternating the polarity of a single electrical field with the circulation frequency.

The *cyclotron* (see Fig. 8.2) is a classical example for this principle. First cyclotrons were built in the 1930s by Livingston and by Lawrence [7]. Protons with an energy of 1.25 MeV were produced by Lawrence's cyclotron in 1932 to split atoms as well. For the purpose of acceleration an alternating high voltage at radio frequency (RF) is applied to two D-shaped hollow electrodes, the "dees". Ions from a central ion source are repetitively accelerated from one dee to the other on a spiral path. The ions are kept on a piecewise circular path by the application of a vertically oriented magnetic field of magnitude B. In practice the magnetic field must decrease slightly with increasing radius to provide stability for the beam in axial direction. The frequency of the accelerating voltage is kept constant and has to match the cyclotron frequency

$$f_c = \frac{qB}{2\pi\gamma m_0}. \tag{8.1}$$

Here, q is the charge of the ion, m_0 its rest mass and the relativistic factor γ represents the ion energy in units of rest energy $\gamma = E/m_0c^2$. As soon as relativistic energies are reached, i.e. $\gamma > 1$, the required frequency drops down, and the ions slip out of the accelerating phase. This effect limits the achievable energy of the cyclotron acceleration principle. In practice, γ factors of 1.02 can be reached.

This limitation can be overcome with the *synchro cyclotron*, where the RF-frequency is ramped down with an appropriate function of the energy to follow the varying circulation time of the ions. Unfortunately, this scheme admits only acceleration of one ion packet for each ramp cycle of the frequency, whereas the classical scheme allows to fill every RF cycle with particles. Consequently, the average accelerated current typically decreases by a factor of 1000 as compared to the classical cyclotron, which limits the practical usefulness of this scheme.

Another solution to overcome the relativistic mass increase was provided by the *isochronous cyclotron* [8]. Magnets with alternating strong and weak azimuthal regions provide an additional axial focusing. In contrast to the classical cyclotron, this allows to raise the average magnetic field with increasing radius as it is required to match the increasing mass of the ions.

Cyclotrons with much improved and sophisticated layouts are today in widespread use for different applications. The Paul Scherrer Institute in Switzerland operates a facility containing a chain of two cyclotrons which generates a 590 MeV proton beam with a continuous power of more than a megawatt [9]. Today this presents the highest average beam power available, although the linac-based project SNS, which is being commissioned in the US, has the potential to produce even higher beam powers.

The *synchrotron* was developed in the 1940s independently by E. McMillan [10] and V. Veksler [11]. The layout of a small storage ring that contains all essential components of a synchrotron is shown in Fig. 8.3. In the synchrotron the guide field increases with particle energy in order to keep the orbit stationary. The acceleration is provided with an RF voltage using a cavity at a localized position in the ring. Working principles of RF cavities are described in the next section. The important new mechanism in the synchrotron is the principle of phase stability where a focusing effect is employed for the longitudinal coordinates. For example, a particle with

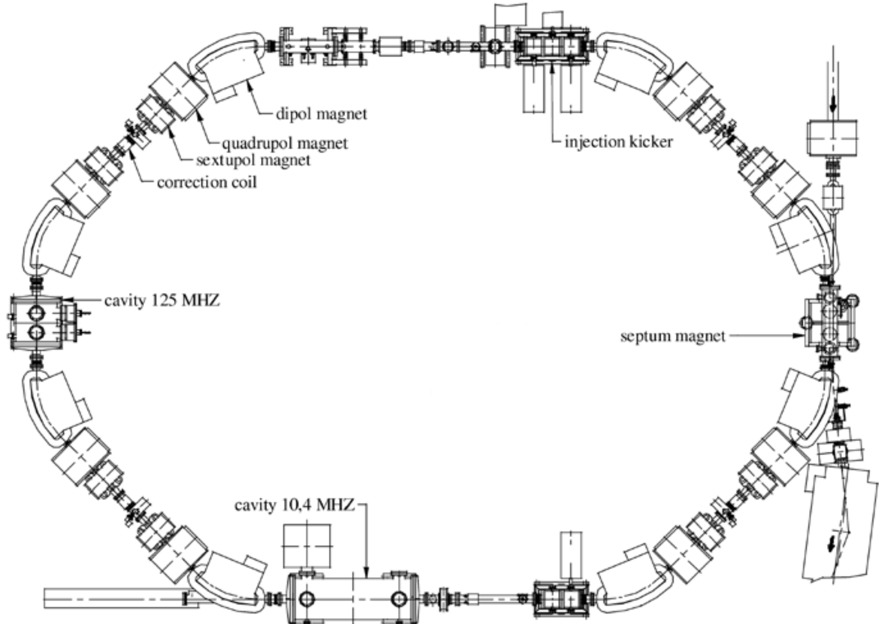


Fig. 8.3. Example of a small synchrotron used as positron intensity accumulator (PIA) at DESY, Hamburg. The circumference is 27.5 m

higher than nominal energy will travel with a higher speed on a different orbit and will arrive at a different time at the accelerating cavity. Either the velocity change or the change of the path length can be more important depending on the rest mass, the energy and the specific magnet lattice of the accelerator. The different arrival time leads to less energy gain on the next turns. In this way the periodic accelerating field collects the particles into circulating bunches, within which the particles perform so-called *synchrotron oscillations* relative to an ideal reference particle at the centre of the bunch.

A mechanism of transverse focusing for the beam is essential as well. The first synchrotrons used the so-called *weak focusing* principle where the guide field is slightly decreased with increasing radius. The disadvantages of weak focusing are the large beam size requiring large diameter vacuum pipes, costly magnets and also the necessity to provide a very high accuracy for the focusing gradient. The energy limit for weak focusing accelerators due to technical limitations lies around 10 GeV.

In 1952 E. Courant, M. Livingston and H. Snyder [12] proposed *strong focusing* employing consecutive quadrupoles with alternating gradients. This new scheme provided a dramatic improvement for the design of synchrotrons in view of achievable energy and cost reduction. Applying strong focusing, the beam size and the dimensions of the accelerator become independent of each other. The basic ingredients of the synchrotron, dipole magnets for bending, alternating quadrupole magnets for focusing and localized RF cavities for acceleration are arranged in a loop with prac-

tically arbitrary length. The function of these components and the concept of strong focusing is described more detailed in Sect. 8.2.2. Practically, all modern circular accelerators for particle and synchrotron radiation research employ the synchrotron principle.

A further major improvement of the accelerator technology was the application of the superconducting magnet technology. Today the largest proton accelerators operate with superconducting bending magnets that achieve field strengths of up to 8 T, which is to be compared with 1.5 T regularly achieved in iron yoke magnets.

8.1.3 Linear Resonant Acceleration

Another branch of the resonant acceleration principle was established with the *linear accelerator*. In 1924 Ising [13] proposed to apply time varying fields produced by a single RF generator using several drift tubes with alternating polarity. As shown in Fig. 8.4, the particles experience an accelerating field in-between the tubes, and the field reverses when they move inside the tubes. The length of the individual drift tubes is adapted to the increasing velocity of the accelerated particles. In 1928 Wideröe built the first accelerator that followed this principle [14].

The rest mass of electrons is 1836 times smaller than that of protons. Consequently, electrons become relativistic already at moderate energies and then they move practically at the speed of light, independent on their energy. Linear accelerators for electrons employ therefore a different working principle using a constant length of the accelerating cells. Such accelerating structures were first developed at the Stanford University and the Massachusetts Institute of Technology in 1946.

In the following the basic working principle of a linear accelerator structure will be explained. An electromagnetic wave can be transported with small losses in a conducting hollow cylinder. Different electromagnetic field configurations, so-called modes, are possible. The modes are solutions of the Maxwell equations with appropriate boundary conditions. One of the lowest order modes is called TM_{010} and exhibits the following electrical field component E_z along the axis of the tube:

$$E_z = \hat{E} J_0(k_c r) \cos(\omega t - k_z z). \tag{8.2}$$

Here, $\hat{E}$ is the peak field strength, J_0 the zeroth order Bessel function and ω the circular frequency of the wave. The wave number k_z is related to the free space wave number $k = \omega/c$, the so-called cutoff wavelength λ_c and corresponding wavenumber

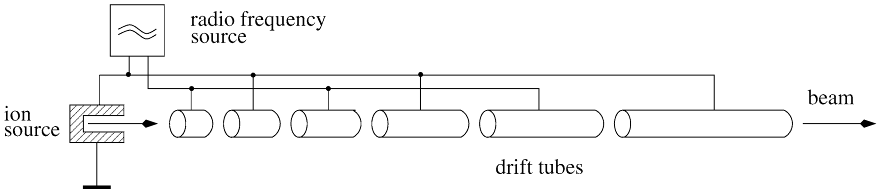


Fig. 8.4. Drift-tube accelerator after Wideröe

$k_c = 2\pi/\lambda_c$ via $k_z = \sqrt{k^2 - k_c^2}$. Only waves with $\lambda < \lambda_c$ can travel along the tube. Since the field must vanish on the metallic surface at the radius of the tube $r = b$, the relation $k_c = \nu_0/b$ must be satisfied, where $\nu_0 = 2.405$ is the first zero of the Bessel function J_0.

In principle, such a wave would be well suited to accelerate electrons if they would "ride on the wave" or, more exactly, if they had the same speed as the phase velocity of the wave. Unfortunately, the phase velocity v_{ph} of all modes is larger than c. For the discussed mode it turns out to be

$$v_{ph} = \frac{\omega}{k_z} = \frac{c}{\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}}} > c. \tag{8.3}$$

An electron at the speed of light would be accelerated only over a short distance, then the field orientation would reverse, and the particle would be decelerated again. On average there would be no energy gain. However, there are ways to overcome this problem. One solution to obtain effective acceleration in a long structure is the installation of periodic apertures in the round pipe. Thereby it is possible to slow the phase velocity down until it matches exactly the speed of light at a certain frequency. Such an accelerating structure is called disc-loaded waveguide or multi-cell cavity (Fig. 8.5).

Contrary to the smooth waveguide, there will always be reflections at the irises of the disc-loaded waveguide, which will produce diffraction like patterns in the longitudinal field distribution. A regular field pattern that suits the purpose of acceleration is only obtained if the wavelength is chosen as a multiple of the disc distance, $\lambda = nd$ (Fig. 8.5). The possible modes are named according to the angular phase advance per cell: $2\pi/n = k_z d$. The most commonly used ones are the π-mode ($n = 2$) and the $2\pi/3$-mode ($n = 3$). In storage rings often single cell cavities are employed.

Classical room temperature accelerator structures are usually made from copper for reasons of high electrical and thermal conductivity. A relatively new development to improve the performance of the accelerating structures is the employment of superconducting cavities. These structures are mostly made from ultra-pure niobium, operated at liquid helium temperatures. With superconducting cavity walls resistive RF losses are reduced to a minimum. The quality factor of such cavities, which is the number of cycles for an e-folding decay of the field amplitude, can reach 10^{10}. This number is to be compared with several 10^4 for conventional copper structures. The advantage of the superconducting structures is the absence of noticeable heat

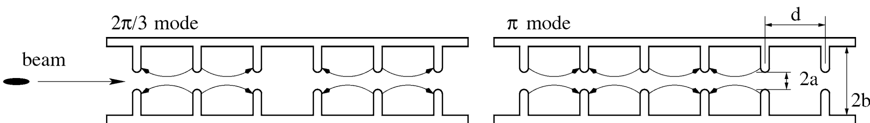


Fig. 8.5. Side view of circular disc-loaded waveguide for resonant acceleration. The figure shows schematically the electrical field lines for the two most commonly applied acceleration modes and relevant dimensions of the structure

Fig. 8.6. Superconducting TESLA shaped 9-cell cavity for operation at 1.3 GHz

loads which permits continuous wave operation or at least operation with millisecond pulses, whereas copper structures can withstand only sub-microsecond pulses at the highest gradients. The first project where superconducting cavities have been used on a larger scale was CEBAF [15]. The proposed International Linear Collider (ILC) [17] uses superconducting 9-cell cavities at 1.3 GHz with anticipated gradients up to 35 MV/m (Fig. 8.6).

8.1.4 Acceleration by Induction (Betatron)

This principle of acceleration by induction was first discussed, but not published, by Wideröe. The particle beam is circulated in an upright magnetic field. When the field strength is dynamically increased with time, a circular electrical field is induced which accelerates the beam along its circular path (Fig. 8.7). Basically, the beam is the equivalent to a secondary winding in a transformer. The relationship between magnetic field **B** and electric field **E** is given by the Maxwell equation

$$\operatorname{rot}\mathbf{E} = -\dot{\mathbf{B}}. \tag{8.4}$$

In this configuration the dynamic combination of magnetic and electric field provides bending, i.e. spacial confinement of the beam and acceleration at the same time.

In order to keep the trajectory radius constant during acceleration, a specific radial variation of the magnetic field is required. This field profile is usually achieved by shaping the pole contour. In the following, a relation for the necessary magnetic field characteristics will be derived. If one requires equality of centrifugal and magnetic bending force on the particle, it holds: $pv/\rho = evB_z$, where p is the particle momentum, v the velocity, ρ the bending radius and B_z the magnitude of the vertically oriented magnetic field. On the other hand, the application of Stokes integral law on (8.4) delivers $2\pi\rho E_\phi = -\pi\rho^2 \frac{\mathrm{d}}{\mathrm{d}t}\overline{B}_z$, where $\overline{B}_z$ is the magnetic field averaged over the circle enclosed by the particle trajectory. The change of particle momentum due to the circular oriented electric field E_ϕ is given by $\frac{\mathrm{d}p}{\mathrm{d}t} = -eE_\phi$. If

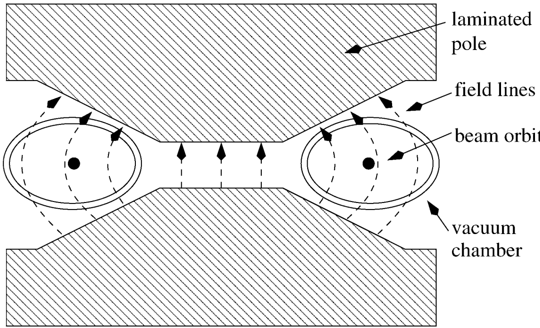


Fig. 8.7. Schematic side view of a betatron

these conditions are combined, we obtain a relation for the B-field on the circle and the average B-field inside the circle: $\frac{\mathrm{d}}{\mathrm{d}t}B_z = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\overline{B}_z$. Integration over time finally leads to the well-known Wideröe condition for stable orbits in a betatron accelerator

$$B_z(t,\rho) = \frac{1}{2}\overline{B}_z(t) + B_0. \tag{8.5}$$

In words: for stable orbits the field on the particle trajectory has to be raised half as much as the average enclosed field. The constant offset B_0 can be chosen at will to adjust the orbit diameter according to the available apparatus.

8.1.5 Particle Sources

Particle sources provide the particles for the accelerator, usually at moderate energies. Often the performance of the whole accelerator depends critically on the beam parameters the particle source delivers. This refers to the time structure of the produced beam, the current and most importantly the beam density which relates to the transverse and longitudinal dimensions of the beam.

The classical principle for the generation of electron beams employs a heated cathode for the emission of electrons. A modern and more advanced example of an electron source is shown in Fig. 8.8 [18]. The electrons are emitted from a special photo-cathode, made of Cs_2Te, which is illuminated by laser light with a wavelength of 263 nm. The cathode material is optimized such as to obtain a maximum yield of electrons per number of photons. The material is very sensitive and has to be handled with extreme care. The high efficiency is lost when exposed to air or even poor vacuum conditions. The laser beam is pulsed in time according to the beam pattern required by the downstream accelerator. In the described electron gun the laser produces 5000 pulses with a length of 15 ps each in a train of totally 800 μs length. This pattern is repeated 10 times per second.

After the emission from the cathode the cloud of electrons tends to blow up because of the repulsive Coulomb forces between the equally charged particles. This force is largely compensated by an attractive magnetic force with increasing beam

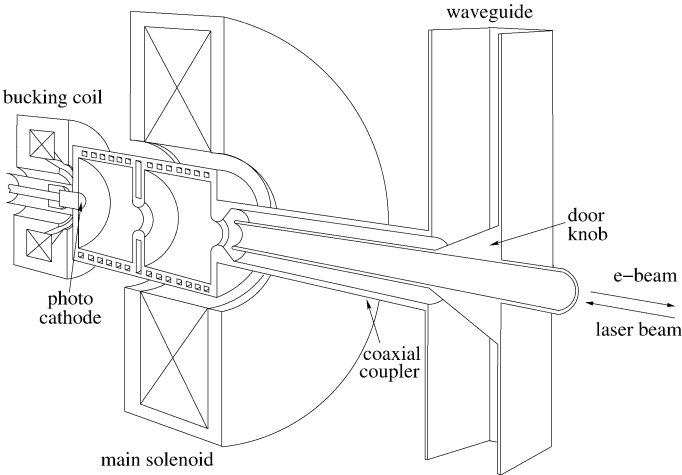


Fig. 8.8. RF gun built for the Tesla Test Facility (DESY) [18]

energy. Therefore, it is important to accelerate the electrons as quickly as possible to relativistic energies in order to obtain small transverse beam dimensions which are related to the so-called emittance (see Sect. 8.2.1).

In the discussed electron source the acceleration is achieved with a short accelerating structure that consists of one and a half cavity cells. The cathode is placed in the end plate of the half-cell, where the maximum gradient occurs. The RF wave is fed to the cells through a coaxial coupler. At a frequency of 1.3 GHz and an RF power of 4.5 MW over a pulse duration of 1 ms one achieves an accelerating gradient of 50 MV/m. With this electron source it is possible to produce an average pulse current of 8 mA and a normalized emittance of $\varepsilon_\mathrm{N} \approx 1$ mm mrad.

For the generation of high energy proton beams the accelerator chain often starts with an H^- source. This provides the advantageous possibility to combine a circulating proton beam and an injected H^- beam in the same phase space. The beams are bend by a magnetic field in opposite direction. At a location where the orbits coincide the electrons are stripped from the H^- ions using a thin foil. If positively charged protons were injected, such an accumulation would not be possible. A description of a pulsed H^- source can be found in [19]. This source produces pulses with a length of 150 µs, a current of 40 mA and a repetition rate of 8 Hz. The achieved normalized emittance amounts to $\varepsilon_\mathrm{N} = 0.06$ mm mrad.

Electron and proton sources are the most commonly used ones. In addition there exists a variety of ion sources and sources for the production of antiparticles, in particular positrons and antiprotons. The interested reader is referred to [2].

8.1.6 Colliders

Already the first accelerators ever built were used to investigate the structure of matter using fixed target experiments. A material sample was irradiated with a parti-

cle beam, and scattered secondary particles were observed behind the target with appropriate detectors. By deep inelastic scattering of antiparticles, e.g. electrons and positrons, it is possible to produce new particles if the collision energy is high enough. However, in case one of the collision partners is at rest, as for the fixed target setup, most of the beam energy is lost in the recoil process. The available *centre of mass energy* E_{cm} scales only with the square root of the beam energy E_{b}: $E_{\mathrm{cm}} \approx \sqrt{2E_0 E_{\mathrm{b}}}$, where $E_0 = m_0 c^2$ is the energy of the particle at rest. This approximation holds for beam energies much higher than the rest energy of the beam particles.

Very soon it became clear that higher collision energies can be achieved if the target is not fixed but two opposing beams are collided. If again the energies E_1 and E_2 of the two beams are much higher than the rest energies of the particles, the center of mass energy of the collision is given by $E_{\mathrm{cm}} \approx 2\sqrt{E_1 E_2}$. For most colliders the two beam energies are equal and $E_{\mathrm{cm}} \approx 2E_{\mathrm{b}}$.

Besides the energy, the frequency of interesting collision events is the other important parameter for each collider. Obviously, this rate is proportional to the number of beam particles and inversely proportional to the cross-sections of the beam spots at the interaction points. As a measure for the interaction rate the *luminosity* $\mathcal{L}$ is introduced

$$\mathcal{L} = \frac{f_{\mathrm{b}} N_1 N_2}{4\pi \sigma_x \sigma_y}. \tag{8.6}$$

Here, f_{b} is the collision frequency of particle packets, N_1 and N_2 are the particle numbers in the packets of the two beams and σ_x and σ_y the rms width and height of the two beams assuming the beams have equal dimensions. The interaction rate for a specific process is obtained by multiplying the luminosity with the interaction cross-section for this process. Practically, luminosities in the range $10^{30} \ldots 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ have been achieved in colliders. To maximize the luminosity, all colliders use special insertions, arrangements of quadrupoles, to achieve very small beam spots at the interaction points.

The concept of high-energy particle colliders provided outstanding possibilities for the research on elementary particles. Over the past 40 years many colliders were built and the energies were raised several orders of magnitude, see Fig. 8.9. Most facilities utilize electron/positron collisions, but some of them are proton/antiproton colliders. Electron/positron storage rings are limited by the particles energy loss from synchrotron radiation (see Sect. 8.2.3). Much higher energies can be reached with protons/antiprotons since they produce nearly no radiation losses. However, since these particles possess a complicated inner structure, their reactions are more difficult to interpret.

In some cases it is advantageous to collide beams with different energies. Examples are the presently operated B-factories PEP-II and KEK-B where it is necessary for the detection of decay lengths to provide a moving centre of mass for the experimental detector. The energies chosen are approximately 3 and 9 GeV. Another example is the HERA accelerator where electrons and protons are collided. Because of the mentioned reason, it is relatively easy to achieve a high proton en-

ergy of 920 GeV, whereas the electron energy is 27.5 GeV. A new facility with the highest energy of $E_{cm} = 14$ TeV and a length of 27 km is the *Large Hadron Collider* (LHC), presently under construction at CERN. This accelerator complex will collide two proton beams.

The synchrotron radiation problem in storage rings can be overcome by the linear collider concept where the beams are accelerated along a straight line, basically without deflections. Linear colliders employ long chains of accelerating structures (see Sect. 8.1.3) to reach high beam energies within a single beam passage. One large scale linear accelerator, the SLC, was successfully operated with energies up to 50 GeV at the Stanford Linear Accelerator Center (SLAC) in the US. In recent years several proposals for linear colliders with beam energies up to 500 GeV have been discussed. The operating frequencies of the accelerator structures in the proposed facilities reach from 1.3 GHz up to 30 GHz. With higher frequency also higher accelerating gradients can be achieved. However, high frequency structures have small dimensions and the interaction of the electromagnetic fields generated by the beam with the cavity walls is stronger. Among other complications this leads to tighter alignment tolerances.

After careful judgement of different aspects the technology of superconducting accelerator structures at 1.3 GHz has been chosen as the basis for the International Linear Collider (ILC) [17].

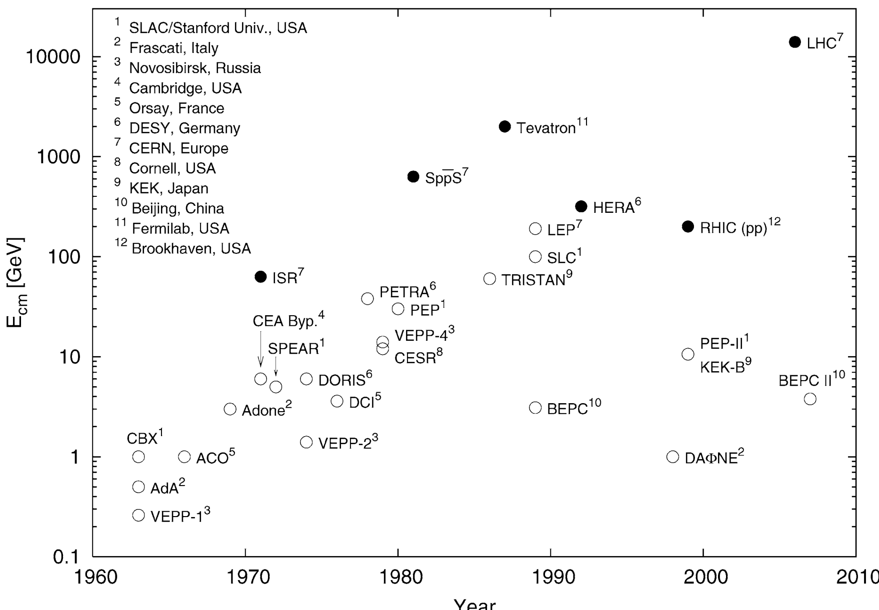


Fig. 8.9. Beam energy versus year of completion for various collider facilities around the world. The hollow points represent e^+e^--colliders, the filled ones $p\bar{p}$- or ep-facilities

8.1.7 Synchrotron Radiation Sources

Synchrotron light, typically X-ray radiation in the keV region, is emitted when a charged particle beam is bend in a magnetic field. Details of this mechanism and the properties of the radiation are discussed in Sect. 8.2.3. Synchrotron radiation occurs as a side effect in circular electron accelerators and storage rings. It was quickly realized that the intense radiation allows a broad range of experimental research in different areas. In the early 1960s the synchrotron DESY in Hamburg was the first accelerator regularly used for synchrotron radiation research in a parasitic mode. A storage ring that was exclusively planned for synchrotron radiation production, named TANTALUS [20], was built in 1968. The mentioned facilities did not take any special measures to maximize the intensity of the produced radiation and are called first generation light sources. Later the magnet lattice of such storage rings was optimized in order to obtain very small equilibrium beam dimensions. In this way these second generation light sources were able to generate much intenser photon beams from dipole radiation. The performance was further improved in third generation light sources that employ specific insertion devices for the deflection of the beam. Insertion devices are *undulator* magnets that bend the beam back and forth (Fig. 8.10). Similar devices with stronger deflection are called *wiggler* magnets. Because the overall direction of the beam is not changed, the radiation of several dipole magnets is superimposed which leads to enhanced radiation at a specific wavelength.

The latest developments for the generation of photon radiation with accelerators are finally represented by *free electron lasers*. Here the beam is sent through undulator magnets as well, with very high charge density. If the interaction between radiation and particle beam is intense and long enough, a self-bunching effect sets on. The forces from the generated electromagnetic waves lead to a concentration of the beam in small packets, micro-bunches, which radiate coherently. This process leads to an exponential growth of the radiation intensity until the particle beam is fully bunched, i.e. modulated with the wavelength of the coherent radiation. The FEL process has been studied and demonstrated at several accelerator facilities worldwide. Figure 8.11 shows the observation of the mentioned exponential growth with

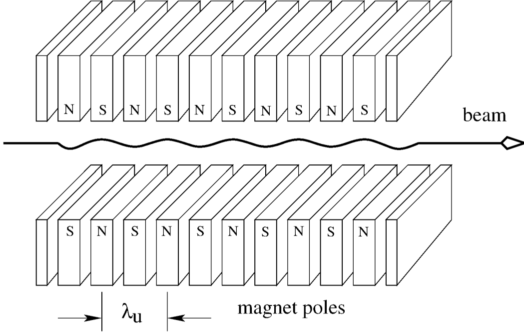


Fig. 8.10. Principle of an undulator magnet for the generation of coherent radiation

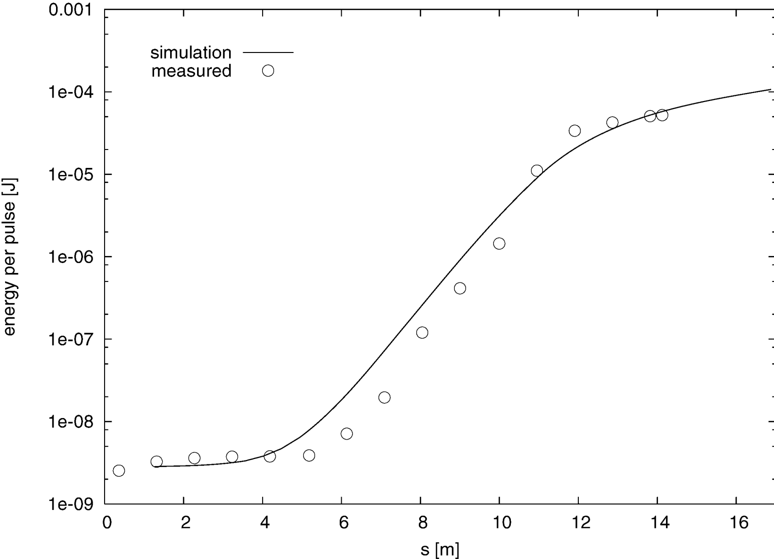


Fig. 8.11. Energy per laser pulse as a function of the undulator length in the TTF FEL. The dots represent measured values, the line a numerical prediction [22]

increasing undulator length and saturation of the process in the TESLA Test Facility (TTF) [21].

A measure for the photon intensity of synchrotron radiation sources is the *brilliance*, which is basically the density of the generated photons in the six-dimensional phase space. The density is expressed in number of photons radiated per area of beam cross-section, per space angle, per unit time and per photon energy interval. As a practical convention, the brilliance is often normalized on 0.1% of the bandwidth in terms of photon frequency. Figure 8.12 shows the historical development of typical brilliances achieved in synchrotron radiation facilities of different generations.

8.2 Essentials of Beam Dynamics

The purpose of this section is to discuss three selected topics of accelerator theory with practical importance and to derive some fundamental relations in somewhat more depth. Nevertheless, the reader should note that the study of additional literature is required for a thorough understanding of the beam dynamics in accelerators.

8.2.1 Beam Emittance and Phase Space

In an accelerator the particles are guided and focused by magnetic fields. The magnets are installed along an *ideal orbit* of the particle beam. It is convenient to describe the motion of an individual particle in terms of coordinates relative to this ideal orbit.

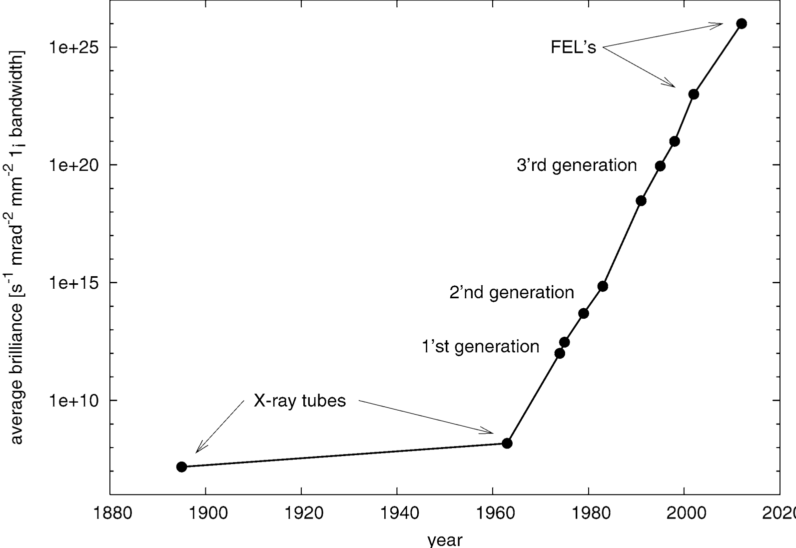


Fig. 8.12. Average brilliance values for accelerator-based synchrotron radiation facilities and X-ray tubes over time

The motion path of an individual particle is called trajectory, whereas the term orbit denotes the sample average of particles in a beam. The instantaneous position of the particle will be specified by $[x, y, s]$, where x, y are the horizontal and vertical deviations from the ideal orbit and s is the distance along the orbit from a reference point (Fig. 8.13). For the complete definition of a particles state three more coordinates, which are connected to the instantaneous momentum, are needed. Commonly used are $[x', y', E]$, where $x' = p_x/p$ and $y' = p_y/p$ are the transverse angular deviations from the ideal orbit and $E \approx cp$ is the particle energy which is proportional to the momentum for highly relativistic particles. Instead of the absolute energy it is often more convenient to use the energy deviation δE to the reference particle.

For many applications of particle accelerators the minimally achievable beam dimensions, or more generally the phase space density in the beam, are very important. The brightness of X-ray beams generated in synchrotron light sources or the interaction rate achieved in colliders for purposes of high energy physics research depend on the beam dimensions. In the first case it is desirable to obtain a small divergence or angular spread of the beam particles at the location where the light is emitted since this will result in a well focused, high brightness X-ray beam. At the interaction point of a collider, where two opposing beams penetrate each other, the goal is to generate small beam cross-sections to maximize the rate at which interesting collisions occur. Both mentioned properties of the beam, the divergence and the dimension are coupled together and are related to the transverse *emittance* parameter.

The beam particles are distributed in a certain volume of the six-dimensional phase space. A volume covered by a surface of constant particle density remains

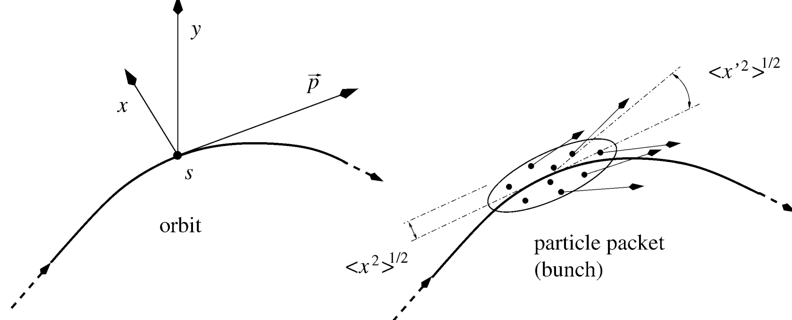


Fig. 8.13. Coordinate system related to the ideal orbit (*left*) and the statistical properties rms width and rms angular spread of particles in a bunch

constant under Hamiltonian transformations. It can be rotated, stretched or arbitrarily deformed, but the enclosed volume is conserved. This property of the phase space is related to the so-called *Liouville theorem*. It has important consequences for the beam dynamics. For example, it allows to compute the propagation of statistical beam properties in a transport channel. The conservation of phase space density is not fulfilled for non-Hamiltonian forces that act on individual particles, e.g. Coulomb scattering or the emission of photons in an external magnetic field.

For practical purposes one considers projections of the full six-dimensional phase space, for example, onto the (x, x')-plane of the transverse coordinates. The horizontal emittance ε_x of the beam is a measure of the average phase space area the beam covers in these coordinates. The rms emittance can be computed statistically from the coordinates of the individual particles. We assume particle coordinates relative to the orbit here, i.e. $\langle x \rangle = 0$, $\langle x' \rangle = 0$:

$$\varepsilon_x = \sqrt{\langle x^2 \rangle \langle x'^2 \rangle - \langle x x' \rangle^2}. \tag{8.7}$$

As an important consequence from the Liouville theorem this value is conserved as long as no coupling to the other four degrees of freedom occurs, which is commonly avoided in accelerator designs. The same emittance value can be defined for the vertical coordinates and also for the longitudinal ones. The so-called *normalized emittance* $\varepsilon_{\mathrm{N}} = \beta\gamma\varepsilon$ is an adiabatic invariant in the absence of radiation and remains constant when $\beta\gamma$ is raised by acceleration. Here, $\beta = v/c$ is the velocity of the particle in units of the speed of light. As a consequence, for proton and ion beams or electrons in a linear accelerator the transverse beam size shrinks during acceleration as $(\beta\gamma)^{-1/2}$.

Besides the beam energy and current, the transverse and longitudinal emittance values belong to the major parameters characterizing an accelerator. In the next sections we discuss how the transverse beam dimensions are related to the emittance in a magnet lattice and the existence of an equilibrium emittance in storage rings with synchrotron radiation.

8.2.2 Equations of Motion and Strong Focusing

This section gives a short introduction into the linear beam dynamics in a strong focusing (compare with Sect. 8.1.2) magnet lattice. This will enable the reader, for example, to compute basic beam properties such as transverse beam size or divergence with the help of an optics table.

In an electromagnetic field the *Lorentz force* $\mathbf{F}$ acts on a particle with charge e and velocity vector $\mathbf{v}$

$$\mathbf{F} = e\mathbf{v} \times \mathbf{B} + e\mathbf{E}. \tag{8.8}$$

In accelerators the electric force (the second term on the right) is commonly used for acceleration and the magnetic force for deflection and focusing. The magnetic force acts always perpendicular to the direction of motion and consequently it is incapable of changing the kinetic energy of a particle. In a dipole magnet with field strength B and length l a particle with momentum p is deflected on a circular path with a bending radius ρ and a deflection angle θ

$$\frac{1}{\rho} = \frac{eB}{p} \approx 0.3\,[\mathrm{m}^{-1}]\,\frac{B\,[\mathrm{T}]}{p\,[\mathrm{GeV/c}]}, \quad \theta \approx \frac{l}{\rho}. \tag{8.9}$$

Besides dipole fields, modern accelerators and beam transport lines use quadrupole fieldsv which apply a bending force that is proportional to the particles deviation from the magnetic axis, thereby providing transverse focusing for the beam. In fact, a particle beam is focused by a quadrupole in a similar way as a light ray by an optical lens. The typical shapes of these two most elementary magnets are shown in Fig. 8.14. Another magnet type of importance is the sextupole. It acts like a position dependent quadrupole and is used to compensate for the momentum dependence of the quadrupole focusing.

Contrary to the optical lens, a magnetic field in a current free region cannot apply a focusing effect to the beam in the horizontal and vertical plane simultaneously. This can be seen by applying the Maxwell equation and expanding the Lorentz force (8.8) around the beam orbit up to linear terms

$$\begin{aligned} \mathrm{rot}\,\mathbf{B} = 0 \rightarrow & \frac{\partial}{\partial x} B_y - \frac{\partial}{\partial y} B_x = 0, \\ & F_x \approx -gx + F_x(0), \quad F_y \approx gy + F_y(0), \\ & g \equiv e\beta c \frac{\partial}{\partial x} B_y \bigg|_{x=0} = e\beta c \frac{\partial}{\partial y} B_x \bigg|_{y=0}. \end{aligned} \tag{8.10}$$

A focusing force is a back-driving force that pushes a particle with any deviation back to the ideal axis. In the above notation, F_x is focusing and F_y defocusing for $g > 0$. An ideal quadrupole field exhibits only the linear terms.

Though focusing for both planes is impossible in a single quadrupole, an effective focusing for both planes is attainable by installing quadrupoles with alternating gradients in a sequence. In-between the quadrupoles one can either install dipoles to form an arc or leave the space empty in case of a transport channel. This scheme is

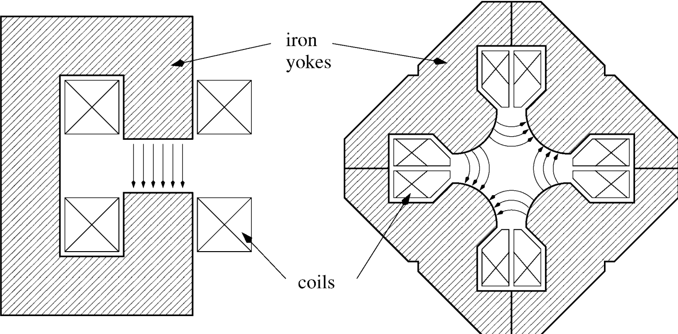


Fig. 8.14. Dipole magnet (*left*) and quadrupole magnet (*right*), both with iron yokes

called alternating gradient focusing or FODO structure since focusing quadrupoles (F), drift spaces (O) and defocusing quadrupoles (D) are employed. The full theory of particle dynamics for the alternating gradient scheme is beyond the scope of this book. The interested reader is referred to [1] or the excellent original publication of Courant and Snyder [23]. We discuss here the basic relations which are important from the practical point of view. If we neglect bending fields and consider the motion in quadrupole fields with respect to the ideal orbit, the equations of motion are:

$$x'' + k(s)x = 0, \quad y'' - k(s)y = 0. \tag{8.11}$$

The prime denotes derivation with respect to s, $k = g/cp$ in quadrupoles and $k = 0$ elsewhere. For a given arrangement of magnets an amplitude function $\beta(s)$ can be computed. The solutions of (8.11) are conveniently expressed with the help of this β-function. We discuss here only the solution for x since that for y is equivalent.

$$x(s) = \sqrt{2J_x\beta_x(s)}\cos(\varphi_x(s) + \varphi_{x0}), \quad \varphi_x(s) = \int_{u=0}^{s} \frac{\mathrm{d}u}{\beta_x(u)}. \tag{8.12}$$

The particles oscillate around an ideal beam orbit with an amplitude that is modulated with the square root of the β-function. At the same time the β-function presents the local wavelength of the oscillation due to its relation with the phase advance $\varphi_x(s)$. The β-function plays a central role for many beam dynamics calculations. The choice of the symbol β is a bit unfortunate since it can be mixed up with the relativistic β that was introduced earlier. To discriminate the two meanings, we always index the function β with x or y, corresponding to the transverse planes of motion.

Trajectories with equal amplitudes and different phases are shown together with the envelope function $\sqrt{\beta_x(s)}$ in Fig. 8.15 for a FODO lattice. The β-function is usually computed with optics codes and tabulated in lists for each accelerator. For completeness we give here also the differential equation for the β-function, which can be deduced from (8.11) and (8.12),

$$2\beta_x\beta_x'' - \beta_x'^2 + 4\beta_x^2 k(s) = 0. \tag{8.13}$$

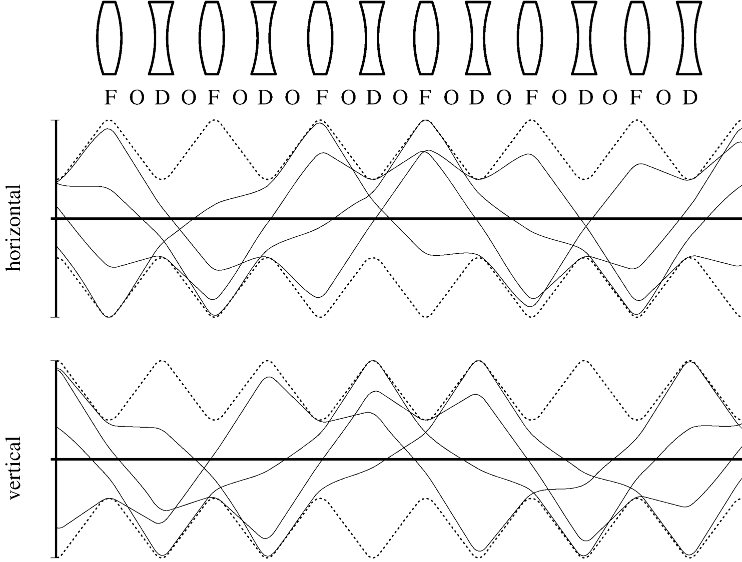


Fig. 8.15. Particle beam (*dashed*) and trajectories for both transverse planes in a FODO structure

Equation (8.12) is the solution of a second order equation of motion and contains two integration constants that define the initial conditions of motion. J_x is the particle action, a measure for the amplitude of the oscillation around the ideal axis, and φ_{x0} the initial phase. The angular deviation of an individual particles motion $p_x/p \approx x'(s)$ is derived by computing the derivative of (8.12) with respect to s

$$x'(s) = -\sqrt{\frac{2J_x}{\beta_x(s)}}(\alpha_x(s)\cos(\varphi_x(s) + \varphi_{x0}) + \sin(\varphi_x(s) + \varphi_{x0})). \tag{8.14}$$

Here, another optical function $\alpha_x(s) = -1/2\, d\beta(s)/ds$ is introduced, which is related to the derivative of the β-function. If the particle coordinates x and x' are measured at a specific location s in an accelerator, the action J_x is computed by

$$J_x = \frac{1}{2}\left(\left(\frac{\alpha_x(s)}{\sqrt{\beta_x(s)}}x + \sqrt{\beta_x(s)}x'\right)^2 + \frac{x^2}{\beta_x(s)}\right). \tag{8.15}$$

Lines with constant action form ellipses in the phase space (x, x'). An example of a phase space distribution is shown in Fig. 8.16. If we observe the coordinates x, x' of an individual particle at a specific location in a storage ring turn by turn, it also samples an ellipse. While the amplitude of the oscillation remains constant, if we neglect synchrotron radiation for a moment, the phase of the oscillation advances every turn by a value $\Delta\varphi_x = 2\pi Q_x$, where Q_x is called the horizontal *betatron tune* of the accelerator. The tune is an important parameter of the magnet lattice that should not approach an integer. In case the tune equals an integer, trajectories in

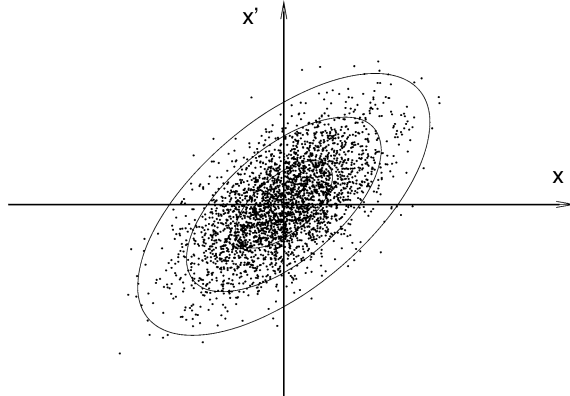


Fig. 8.16. Simulated Gaussian phase space distribution in the coordinates x, x'. The ellipses for one, two and three standard deviations are indicated

consecutive turns would have the same phases, small distortions would add up turn by turn and would finally lead to particle losses. Also, higher resonances where the tune approaches rational numbers with small integer denominators are dangerous.

In electron storage rings the emission of synchrotron radiation (Sect. 8.2.3) leads to an equilibrium Gaussian distribution of the beam particles in the transverse coordinates. Often also the distribution of a proton beam is well approximated by a Gaussian one, although a proton beam is never in an equilibrium state due to the practical absence of radiation. For this case the distribution functions of a beam containing N particles are given below in terms of the phase space coordinates x, x', and the above introduced action angle variables J_x, φ_x. For the sake of simplicity, an accelerator position with $\alpha_x = 0$ has been chosen here.

$$\begin{aligned}\rho(x, x')\,\mathrm{d}x\,\mathrm{d}x' &= \frac{N}{2\pi\varepsilon_x}\exp\left(-\frac{\beta_x(s)x'^2 + \dfrac{x^2}{\beta_x(s)}}{2\varepsilon_x}\right)\mathrm{d}x\,\mathrm{d}x',\\ \rho(J_x, \varphi_x)\,\mathrm{d}J_x\,\mathrm{d}\varphi_x &= \frac{N}{2\pi\varepsilon_x}\exp\left(-\frac{J_x}{\varepsilon_x}\right)\mathrm{d}J_x\,\mathrm{d}\varphi_x.\end{aligned} \tag{8.16}$$

The distribution for y is equivalent. By computing

$$\langle J_x\rangle = \int \mathrm{d}J_x\,\mathrm{d}\varphi_x\,J_x\,\rho(J_x, \varphi_x)/N,$$

we find that the expectation value of the action is the emittance $\varepsilon_x \equiv \langle J_x\rangle$. In the Gaussian phase space distribution shown in Fig. 8.16, the ellipses at one and two standard deviations are indicated. The beam contains 39%, 86% and 99% of particles inside one, two and three standard deviations, respectively. These numbers are valid for the two-dimensional Gaussian distribution of (8.17) and should not be confused

with the ones of the often used one-dimensional distribution. As discussed above, the essential parameter is the emittance ε_x. The rms beam width σ_x and divergence σ_x' are given by

$$\sigma_x = \sqrt{\beta_x \varepsilon_x}, \quad \sigma_x' = \sqrt{(1+\alpha_x^2)\varepsilon_x/\beta_x}. \tag{8.17}$$

Here we see the importance of the tabulated optical functions α_x, β_x. Together with the emittance value, they allow to compute beam width and divergence for any position in the accelerator.

8.2.3 Synchrotron Radiation

Any accelerated charge radiates electromagnetic power. This fact has many important consequences for particle accelerators. They range from radiation induced heat-load on vacuum components over radiation damping in a storage ring to such subtle effects as unavoidable energy uncertainty during the collisions in a linear collider at very high beam energies.

In principle radiation is emitted for any acceleration, however, the strongest radiation occurs when the acceleration is applied perpendicular to the direction of motion of the particle, e.g. in a magnetic field [24]. The radiation power of a single particle on a circular path with radius ρ, rest mass m_0 and energy E was computed by Liénard already in 1898 [25],

$$P_\gamma = \frac{e^2 c}{6\pi\varepsilon_0} \frac{1}{(m_0 c^2)^4} \frac{E^4}{\rho^2}. \tag{8.18}$$

The rest mass m_0 enters (8.18) with the fourth power. With a rest mass ratio of 1836 a proton radiates $1836^4 \approx 10^{13}$ times less power than an electron when moving on the same bending radius with the same energy. At the beam energies available today radiation is only important in electron or positron accelerators. However, in the LHC collider, presently under construction, at 7 TeV beam energy also the radiation from protons becomes relevant.

In more practical units, the power radiated by an electron beam with current I_b over a length l is

$$P_{\text{beam}}\,[\text{kW}] \approx 14.1 \frac{E^4[\text{GeV}^4]\, l\,[\text{m}]\, I_b\,[\text{A}]}{\rho^2\,[\text{m}^2]}. \tag{8.19}$$

The electron/positron collider LEP at CERN was the storage ring with the highest lepton energy ever built. With a bending radius of 3100 m, a top energy of 104 GeV per beam and a beam current of 6 mA the total radiation power was 2×20 MW. This power is deposited as heat in the beam pipe around the ring. The power loss must be continuously compensated by the RF system of the accelerator. The example of LEP demonstrates the limitations of the storage ring concept. Even with the large circumference of 27 km it is not possible to extend the beam energy substantially beyond 100 GeV. The strong scaling of the required power with the fourth power of the beam energy, see (8.18), would require enormous RF powers. The way out are linear colliders as described in Sect. 8.1.3.

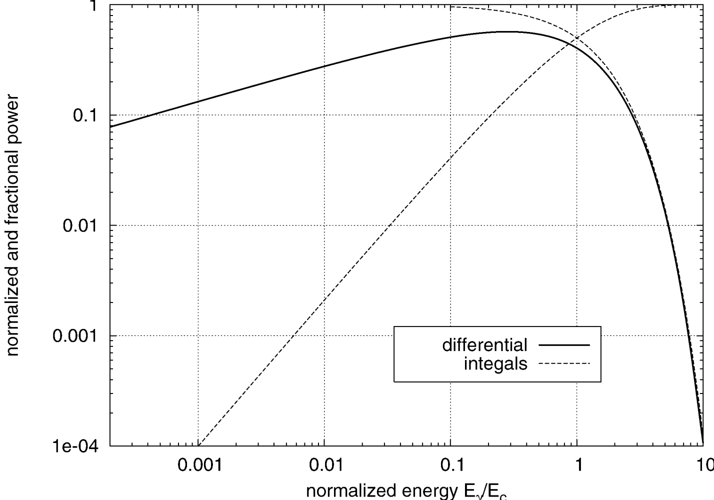


Fig. 8.17. Normalized differential synchrotron radiation spectrum and integrals of the spectrum from both sides. The photon energy is given in units of the critical energy E_c

Besides the power, another important characteristics of the radiation is their spectral distribution, as shown in Fig. 8.17. The exact distribution function was first computed by Schwinger [26] in 1946. For most applications it is enough to know a typical photon energy, the so-called *critical energy* E_c. The power content of the spectrum is exactly 50% below and above the critical energy.

$$E_c = \frac{3}{2}\hbar c \frac{\gamma^3}{\rho}, \qquad E_c\,[\text{keV}] \approx 2.218 \frac{E^3[\text{GeV}^3]}{\rho[\text{m}]}. \tag{8.20}$$

The derivation of the spectral function is described in [24]. For practical purposes one is often interested in the integral of the spectrum, for example, to determine how much power is radiated above a certain photon energy. Figure 8.17 also contains the integrals of the spectrum from both sides.

The photons are emitted in a narrow cone around the instantaneous direction of motion of the particle. The typical opening angle is γ^{-1}. However, the average emission angle is correlated with the energy of the photons. Photons with higher energy are more sharply peaked around the moving direction. The rough scaling as a function of energy follows the relation $\theta_{\text{rms}} = (E_c/E_\gamma)^{1/3}/\gamma$.

The emission of synchrotron light in a storage ring results in an important effect called *radiation damping*. Generally, beam particles are not moving exactly on the nominal beam orbit but carry some small transverse momentum. The transverse momentum is statistically distributed among the beam particles and leads to the finite beam dimensions. As discussed above, the emission of photons is peaked in the direction of the instantaneous momentum vector of the particle. Due to momentum conservation, the recoil of the emission process leads to loss of longitudinal and

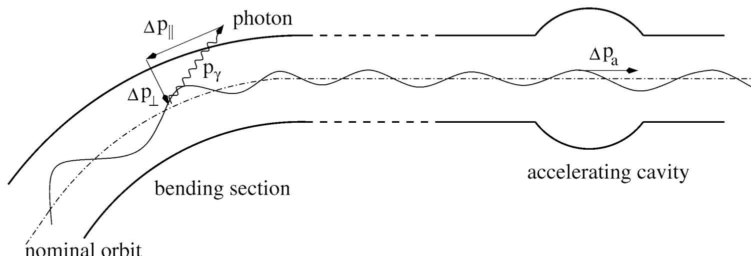


Fig. 8.18. Principle of transverse radiation damping; the emission of a photon reduces longitudinal and transverse momentum of a particle, but only longitudinal momentum is gained back in the accelerating cavity

transverse momentum for the beam particle. In the accelerating field of the storage ring the momentum loss is compensated. However, the particle gains only longitudinal momentum back, and the transverse momentum shrinks to some small equilibrium value. This mechanism is sketched in Fig. 8.18. In a similar way also oscillations of the particle energy around the nominal energy are damped. Robinson [27] formulated a theorem that relates the damping times of the energy oscillations (τ_E) and the two transverse planes (τ_x, τ_y) of motion. The *Robinson Theorem* is based on rather general assumptions and will be given here without proof. The essential content of the theorem states that the sum of the damping rates, i.e. the inverse damping times, is a constant:

$$\frac{1}{\tau_x} + \frac{1}{\tau_y} + \frac{1}{\tau_\mathrm{E}} = \frac{4}{\tau_0}, \qquad \tau_0 = 2\frac{R}{\rho}\,\frac{E_\mathrm{b}}{P_\gamma}. \tag{8.21}$$

Here, R is the effective ring radius, i.e. the circumference divided by 2π, and ρ the bending radius in the dipole magnets. If the ring contains no vertical deflection, it holds $\tau_y = \tau_0$. In addition, for most designs one has $\tau_x \approx \tau_0$ and $\tau_\mathrm{E} \approx \tau_0/2$. Horizontal and longitudinal damping rates can be traded against each other by modifying the accelerator optics. This “shifting” of the damping distribution also results in different horizontal and longitudinal equilibrium emittances. Examples for practical damping rates in existing accelerators are given in Table 8.1.

If the emission of radiation would occur continuously without granularity, the beam size would shrink to infinitely small dimensions. However, the energy is radiated in the form of individual photons that carry certain portions of momentum. This introduces fluctuations for the particles transverse momenta around zero and also around the nominal momentum in the longitudinal direction. The rate $\mathrm{d}N_\gamma/\mathrm{d}t$ of photons radiated in a dipole magnet with bending radius ρ is given by [28]

$$\frac{\mathrm{d}N_\gamma}{\mathrm{d}t} = \frac{5\alpha_\mathrm{f}}{2\sqrt{3}}\frac{\gamma}{\rho}c. \tag{8.22}$$

Here, $\alpha_\mathrm{f} \approx 1/137$ is the fine structure constant. Note that the rate in (8.22) is about 3 times higher than the naive estimate P_γ/E_c from (8.18) and (8.20). The reason lies in the asymmetric shape of the spectral distribution.

Table 8.1. Some parameters of electron storage rings related to synchrotron radiation and radiation damping. The listed parameters in the given order are: beam energy, ring length, bending radius, critical energy, horizontal emittance, damping time, radiated photons per particle and turn

	E_b (GeV)	l (m)	ρ (m)	E_c (keV)	ε_x (nm)	τ_0 (ms)	N_γ/turn	Location
DAFNE[a]	0.51	98	1.4	0.21	1000	36	66	Frascati/Italy
SLS[b]	2.4	288	5.7	5.4	5	9	310	PSI/Switzerland
SPRING 8[b]	8.0	1436	39	29	6	8.2	1040	JAERI/Japan
PEP II/HER[a]	9.0	2199	165	9.8	49	38	1200	SLAC/USA
LEP II[a]	104	26 660	3100	810	44	5.5	13 500	CERN/Europe

[a] particle physics facilities
[b] synchrotron light sources

After several damping times the particle distribution assumes a Gaussian shape with an equilibrium emittance. It is possible to optimize the optics of a storage ring with the goal to minimize the emittance. This is commonly done in rings that work as dedicated synchrotron radiation sources. The principles applied for magnet lattice optimization are described, for example, in [1].

8.3 Vacuum Requirements in Particle Accelerators

The required level of vacuum for an accelerator depends mainly on the degree of degradation the beam could tolerate while passing through the beam pipe. Further restrictions are imposed by the users of the accelerator. Even lower pressure may be requested close to the experimental detector to provide acceptable background conditions. The vacuum pressure achieved in an accelerator often is not only given by the thermal outgassing from the components installed, but can be affected in a number of ways by the interaction of the beam with the vacuum system. More details of the topics discussed in this section can be found, for example, in [2, 29, 30].

8.3.1 Beam Gas Interaction

One very important figure of a storage ring is the beam lifetime. There are always effects that lead to losses of particles from the beam resulting in a usually slow decay of the beam current. For the processes discussed here, the particle losses are proportional to the number N of particles in the beam, $\mathrm{d}N \propto N\,\mathrm{d}t$, which implies an exponential decay of the beam current. The e-folding time of this decay, mathematically expressed as $\tau = -N/(\mathrm{d}N/\mathrm{d}t)$, is called the *beam lifetime*. In storage rings lifetimes above 10 hours are usually desirable. Often the most important parameter in this context is the average residual gas pressure in the accelerator beam pipe together with the gas composition. Scattering of beam particles off the residual gas molecules leads to beam blowup, in the extreme case to particle losses. Also in

linear accelerators, where the beam passes only once, one is interested in good vacuum conditions. Although there is no such figure as a beam lifetime, it is required to keep beam losses at a small level to avoid damage of accelerator components and to obtain clean background conditions for the users of the beam.

When a particle within the stored beam is scattered off a residual gas molecule, it receives a scattering angle and it loses some fraction of its energy. Both mechanisms can cause particle losses from the beam. The rate at which beam particles are scattered is basically given by the product of beam velocity βc, particle number N and effective cross-section $\sigma_{i,l}$ of the residual gas atoms which is related to the interaction probability by

$$-\left(\frac{\mathrm{d}N}{\mathrm{d}t}\right)_l = N\beta c \sum_i n_i \sigma_{i,l} = \frac{N}{\tau_l}. \tag{8.23}$$

The sum has to be carried out over the different species of gas atoms contained in the residual gas composition, and n_i is the volume density of the atoms. The interaction cross-section between beam particles and residual gas atoms of type i is denoted by $\sigma_{i,l}$. Different types of scattering processes are possible and are indexed with l. The total beam lifetime is obtained from the inverse addition of the lifetimes related to the individual loss mechanisms, $1/\tau_{\mathrm{eff}} = \sum_l 1/\tau_l$. We will discuss next the relation between the beam lifetime and pressure in an electron storage ring and describe the problems specific to proton/ion storage rings later.

The elastic or Coulomb scattering of an electron is described by the well-known formula for *Rutherford Scattering* which gives the differential cross-section for the occurrence of a scattering angle θ,

$$\frac{\mathrm{d}\sigma_i}{\mathrm{d}\Omega} = \frac{Z_i^2 r_\mathrm{e}^2}{4\gamma^2} \frac{1}{\sin^4(\theta/2)}. \tag{8.24}$$

Here, the charge of the residual gas atom is Z_i and the classical electron radius is $r_\mathrm{e} = 2.8\,\mathrm{fm}$. If we integrate this differential cross-section from the angle θ_0 above which particles are lost to π and make use of the approximation $\theta_0 \ll 1$, the total elastic scattering cross-section is

$$\sigma_{i,\mathrm{el}} = \frac{2\pi Z_i^2 r_\mathrm{e}^2}{\gamma^2} \frac{1}{\theta_0^2}. \tag{8.25}$$

The limiting angle θ_0 can be estimated using a typical value for the β-function $\overline{\beta_y}$ and the minimum aperture A_y of the accelerator, $\theta_0 = A_y/\overline{\beta_y}$. We use the values for the vertical plane here since the vertical aperture is usually smaller than the horizontal one in electron storage rings. In principle, the geometric mean value of the β-functions at the locations of particle scattering and particle loss have to be used. However, these positions are not known for most practical situations. Combining the above relations and carrying out the sum over different atom species, we obtain the following formula for the beam lifetime due to elastic scattering:

$$\tau_{\mathrm{el}}^{-1} = \frac{2\pi r_\mathrm{e}^2 c}{\gamma^2} \frac{\overline{\beta_y}^2}{A_y^2} \sum_i n_i \sum_j k_{ij} Z_j^2. \tag{8.26}$$

Here, k_{ij} represents the number of atoms of type j within the molecule of type i. By inserting numbers for the fundamental constants and expressing the gas density in terms of pressure at room temperature, we obtain the following formula for the beam lifetime in electron rings due to elastic scattering:

$$\tau_{\mathrm{el}}\,[\mathrm{h}] = 2839 \frac{E^2\,[\mathrm{GeV}^2]\; A_y^2\,[\mathrm{mm}^2]}{\overline{\beta}_y^2\,[\mathrm{m}^2]} \left(\sum_i P_i\,[\mathrm{pbar}] \sum_j k_{ij} Z_j^2 \right)^{-1} . \tag{8.27}$$

Because of the quadratic dependence on Z, the beam lifetime is crucially affected if heavy gas species are present in the gas composition.

As an example, we consider the 27 GeV storage ring HERA-e at DESY. The typical gas composition of a synchrotron radiation dominated storage ring consists of 25% CO and 75% H_2, see Sect. 8.3.2. The effective nuclear charge for this composition amounts to $Z_{\mathrm{eff}} = \sqrt{\langle Z_i^2 \rangle} \approx 3.6$. For HERA we assume an average total pressure of 10^{-8} mbar, a minimum aperture of $A_y = 20\,\mathrm{mm}$ and a β-function of $\overline{\beta_y} \approx 25\,\mathrm{m}$. The resulting beam lifetime from elastic scattering is $\tau_{\mathrm{el}} = 5200\,\mathrm{h}$. In fact elastic scattering is not an important contribution for the beam lifetime and practically observed lifetimes are much lower due to different mechanisms. However, in storage rings with small apertures, e.g. with special wiggler magnet insertions, elastic scattering is potentially more important.

The other major loss mechanism is inelastic scattering caused by *Bremsstrahlung* interaction of the beam particles with residual gas molecules. This process involves the deceleration of a beam particle in the Coulomb field of a residual gas atom and the emission of a high energy photon. The energy loss from the interaction can cause the particle to leave the energy acceptance of the accelerator, and then it will eventually be lost by further energy loss from synchrotron radiation. The important parameter in this context is the largest allowed relative energy deviation for the particles to stay confined within the beam, $\delta_{\mathrm{E}} = \Delta E / E_0$. The cross-section for the inelastic process is

$$\sigma_{\mathrm{inel}} \approx -\frac{4}{3} \frac{V_n}{N_{\mathrm{A}}} \frac{1}{X_0} \ln(\delta_{\mathrm{E}}), \tag{8.28}$$

where N_{A} is the Avogadro constant and $V_n = 22.4$ l/mol is the molar volume under standard conditions. The *radiation length* X_0 is the length over which a particles energy has dropped by a factor $1/e$. X_0 scales roughly inversely proportional to the square of the nuclear charge of the residual gas and also inversely proportional to its density. Radiation length values for common gases under normal conditions are tabulated, for example, in [31]. In Table 8.2 we list X_0 for the important gases of accelerator vacuum systems. To compute the beam lifetime, the effective radiation length is obtained by inverse averaging over the different gas species:

$$\frac{1}{\tau_{\mathrm{brems}}} = -\frac{4}{3} \frac{c}{P_n} \ln(\delta_{\mathrm{E}}) \sum_i \frac{P_i}{X_{0,i}}, \tag{8.29}$$

where P_n is the atmospheric pressure for which tabulated values $X_{0,i}$ of different gas species are usually given and P_i are their partial pressures in the vacuum system. This expression can be converted into the following formula for practical applications:

Table 8.2. Radiation length X_0 and inelastic interaction length λ_i for different gases under atmospheric pressure and 20°C

	H_2	He	CH_4	H_2O	CO	N_2	C_2H_6	Ar	CO_2	air
A	2	4	16	18	28	28	30	40	44	
X_0 (m)	7530	5670	696	477	321	326	364	117	196	304
λ_{inel} (m)	6107	3912	1103	1116	763	753	606	704	490	747

$$\tau_{\text{brems}}\,[\text{h}] = \frac{-0.695}{\ln(\delta_\text{E})}\left(\sum_i \frac{P_i\,[\text{pbar}]}{X_{0,i}\,[\text{m}]}\right)^{-1}. \tag{8.30}$$

In the HERA-e storage ring lifetimes of $\tau_\text{b} > 10\,\text{h}$ are required. In practice the energy acceptance is $\delta_\text{E} = 8\times10^{-3}$ and gas composition and pressure are assumed as discussed in the example on Coulomb scattering. This results in a theoretical lifetime of $\tau_\text{b} = 16\,\text{h}$, whereas the lifetime observed in practice is in the range of 12–17 h.

For a proton or ion beam already the degradation of the beam emittance from elastic gas scattering at very small angles is harmful. Due to the absence of radiation damping any decrease of the beam density over the storage time cannot be recovered. This accumulation of small scattering effects over time leads generally to tighter vacuum requirements than for electrons. In a collider facility the slow increase of the beam dimensions is translated into a finite *luminosity lifetime*. It is possible to define a growth- or lifetime for the beam emittance:

$$\frac{1}{\tau_\varepsilon} = \frac{1}{\varepsilon}\frac{\text{d}\varepsilon}{\text{d}t}. \tag{8.31}$$

The scattering of the beam particles at the residual gas causes a diffusive growth of the mean squared angular deviation of the particles momentum vector which is linear in time. The emittance growth is related to this angle as follows (θ_p is the rms scattering angle projected on a transverse plane):

$$\frac{\text{d}\varepsilon_y}{\text{d}t} = \frac{1}{2}\overline{\beta_y}\frac{\text{d}(\theta_\text{p}^2)}{\text{d}t} = \frac{1}{2}\overline{\beta_y}\frac{(13.6)^2}{(\beta cp)^2\,[\text{MeV}^2]}\frac{\beta c}{P_n}\sum_i \frac{P_i}{X_{0,i}}. \tag{8.32}$$

For highly relativistic beams the growth time is then given by

$$\tau_\varepsilon\,[\text{h}] \approx 34.2\frac{\varepsilon_y\,[\text{m rad}]\;E^2\,[\text{GeV}^2]\;T\,[\text{K}]}{\overline{\beta}_y\,[\text{m}]}\left(\sum_i \frac{P_i\,[\text{pbar}]}{X_{0,i}\,[\text{m}]}\right)^{-1}. \tag{8.33}$$

The temperature T has been included in the given formula since many proton accelerators usesuperconducting magnets and cold beam pipes. The described mechanism takes only Coulomb scattering into account.

As a numerical example, we consider here the proton ring of HERA with an operating energy of 920 GeV, an average β-function of 50 m, a gas temperature of 4.5 K for the major part of the ring and a gas pressure of 5×10^{-11} mbar. The gas composition should be exclusively hydrogen at this temperature and the beam emittance

amounts to 5×10^{-9} m rad. With these numbers we obtain a rather slow rise time of 1960 h.

Another process is the complete removal of particles from the beam by *inelastic scattering*. The beam lifetime for this effect can be computed using the inelastic interaction length λ_{inel}, which is also tabulated in Table 8.2,

$$\frac{1}{\tau_{\text{inel}}} = \frac{\beta c}{P_0} \sum_i \frac{P_i}{\lambda_{\text{inel},i}}. \tag{8.34}$$

The inelastic interaction length is related to the corresponding nuclear cross-section via $\lambda_{\text{inel}} = A/\rho\, N_{\text{A}}\, \sigma_{\text{inel}}$, where A is the molar mass and ρ the density. If we again include the gas temperature and take out all constants, we obtain the following formula for the inelastic beam-gas lifetime:

$$\tau_{\text{inel}}\,[\text{h}] = 3.2 \times 10^{-3} T\,[\text{K}] \left(\sum_i \frac{P_i\,[\text{pbar}]}{\lambda_{\text{inel},i}\,[\text{m}]} \right)^{-1}. \tag{8.35}$$

Once more we use HERA-p with the above parameters for a numerical example and read the interaction length from Table 8.2. The lifetime is computed with 1740 h, and indeed lifetimes of 1000 h or more are regularly achieved in the proton ring when no electron beam is present. With beam–beam collisions other effects come into play that cause smaller lifetimes.

8.3.2 Static and Dynamic Aspects of Accelerator Vacuum Systems

For a given geometry and pump speed the lowest achievable pressure in a vacuum system is determined by the thermal outgassing of the vacuum chambers and components installed. This outgassing is initially dominated by the desorption of adsorbed gas molecules from the surfaces, e.g. water. After sufficient pumping time, the outgassing is mainly hydrogen originating from diffusion out of the bulk materials. Hence the *static vacuum* is given by the desorption rate of the used materials and the pumping speed installed into the system. However, the vacuum pressure of particle accelerators can be affected in a number of ways by the interaction of the beam with the vacuum system. This is usually referred to as *dynamic vacuum*.

Intense synchrotron radiation, focused into a forward-directed narrow cone, is produced in all high-energy accelerators for electrons and positrons. For the first time a significant level of synchrotron radiation will also occur at the high-energy proton beams of the LHC. The energetic photon flux produces strong outgassing from the vacuum system leading to large dynamic pressure increases, which could limit the beam lifetime in a storage ring or cause increased background in the area of an experiment.

The *synchrotron radiation induced desorption* of gas molecules is considered to occur in a two stage process. The photons primarily produce photoelectrons. Subsequently, tightly bound gas molecules are desorbed by electron stimulated desorption from the surfaces [32, 33]. For very high photon energies the electrons are produced

by Compton scattering instead of the photo-effect. Due to scattered and reflected photons and also secondary electrons the molecules are practically desorbed from the whole internal vacuum chamber surface. Following (8.22), the total number of emitted photons per second around the accelerator circumference increases linearly with beam intensity I and particle energy E,

$$\frac{\mathrm{d}N_\gamma}{\mathrm{d}t} = 8.08 \times 10^{17} I\,[\mathrm{mA}]\; E\,[\mathrm{GeV}]. \tag{8.36}$$

The photon flux per unit length (s^{-1} m^{-1}) is given by

$$\frac{\mathrm{d}N_\gamma}{\mathrm{d}t\,\mathrm{d}s} = 1.28 \times 10^{17} \frac{I\,[\mathrm{mA}]\; E\,[\mathrm{GeV}]}{\rho\,[\mathrm{m}]}, \tag{8.37}$$

with ρ being the bending radius. The resulting flux of desorbed gas is related to the photon flux by the desorption yield η_γ, the number of desorbed gas molecules per incident photon. Instead of quoting the molecular desorption yield, the vacuum performance of an accelerator is usually expressed by the dynamic pressure rise $\Delta P/\Delta I$ which includes implicitly also the installed pumping speed of the system.

The desorption yield η_γ depends on the chosen material and its preparation, e.g. cleaning. It needs to be determined experimentally. Figure 8.19 shows an example of η_γ as a function of accumulated photon flux, more figures and references are given in [29]. Variations in η_γ by a factor of 10 for various gas species are not unusual. The

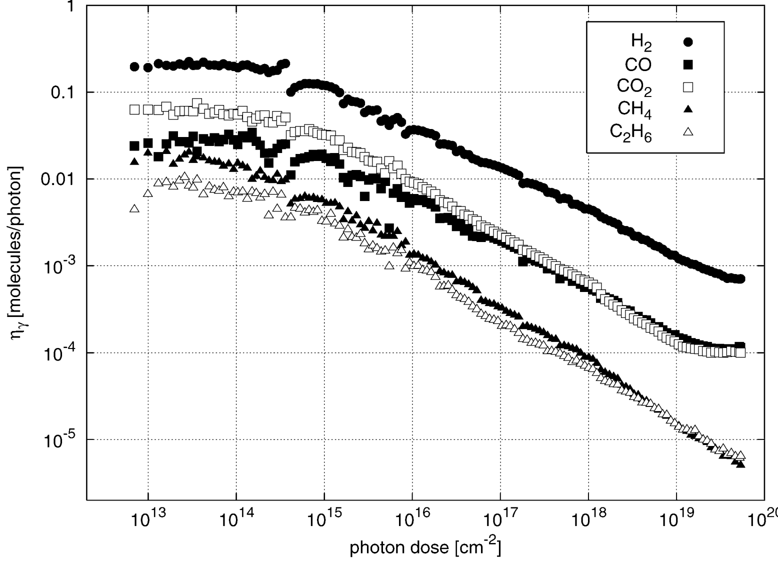


Fig. 8.19. Desorption yield η_γ as a function of accumulated photon flux for unbaked copper [34]

desorbed gas is dominated by hydrogen, followed by carbon monoxide and carbon dioxide. In addition, methane is released at a rate of about a factor 10 less, while water when present is not much desorbed.

During operation of the accelerator, η_γ decreases initially very fast and later on more gradually by several orders of magnitude with increasing accumulated photon flux. As the initial pressure rise normally exceeds the static pressure by several orders of magnitude, the resulting initial beam lifetime in storage rings may be very short. However, in most facilities it is an accepted design concept that the required vacuum performance will be achieved after an initial conditioning time only. Figure 8.20 shows the evolution of the dynamic pressure of LEP over a period of more than 10 years [35]. The overall dynamic pressure rise decreases inversely proportional to the integrated beam current.

An important aspect for the operation of accelerators is the need of reconditioning of the vacuum system following any exposure to atmospheric pressure, e.g. in case of a repair or other intervention into the system. Fortunately, formerly conditioned vacuum chambers keep a memory, at least partially. Usually the training starts at a lower initial desorption level. Venting with dry and clean nitrogen helps to preserve the previous conditioning as much as possible.

For storage rings with intense, positively charged beams of protons or positrons the pressure can increase due to *ion-induced desorption* from the walls of the vacuum chambers [36]. The circulating beam particles ionize residual gas molecules. The positive space-charge potential of the beam accelerates these ions towards the vacuum chamber wall. The ions gain an energy of up to several keV per ampere of

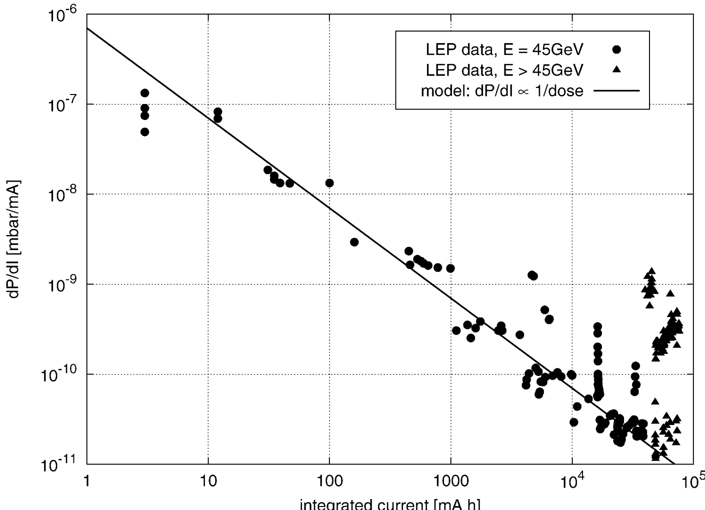


Fig. 8.20. Evolution of the dynamic pressure of LEP over a period of more than 10 years [35]. The beam energy was increased gradually from an initial value of 45 GeV up to 100 GeV at the end of the operation time of LEP

circulating beam. Therefore, the ions are very efficient in desorbing molecules from the walls. Since the rate of ionization is proportional to the gas density, an avalanche process may occur resulting in a continuously increasing pressure.

Similar to the synchrotron radiation-induced desorption, the flux q of desorbed molecules per unit length can be expressed in terms of a desorption yield η_{ion} (here the molecules per incident ion) by

$$q = \eta_{\mathrm{ion}} \sigma P \frac{I}{e}, \tag{8.38}$$

where σ is the ionization cross-section of the residual gas molecules and P the vacuum pressure. Adding the thermal outgassing rate q_0 to (8.38), the balance between the total gas flow into the system and the molecules pumped with an effective pumping speed $\mathcal{S}$ per unit length is given by

$$P\mathcal{S} = \eta_{\mathrm{ion}} \sigma P \frac{I}{e} + q_0, \tag{8.39}$$

thus

$$P = \frac{q_0}{\mathcal{S} - \eta_{\mathrm{ion}} \sigma \frac{I}{e}}. \tag{8.40}$$

The pressure increases with beam current, and no equilibrium pressure exists for conditions above a critical value

$$\eta_{\mathrm{ion}} I > \frac{\mathcal{S} e}{\sigma}. \tag{8.41}$$

Up to now, an ion-induced pressure instability has been observed in one proton facility only [37], setting a severe limitation for the maximum circulating beam current. The designs of subsequently built accelerators have been worked out carefully to avoid this effect. As the stability depends on the local values of pumping speed and desorption coefficient, i.e. on the local degree of cleanliness of the vacuum chamber, the pressure run-away will always develop at the weakest place in the accelerator. Contrary to conditioning with synchrotron radiation it is not possible to obtain any significant improvement for the ion-induced desorption by merely running the accelerator with beam for long periods. This can be attributed to the much lower integrated ion dose that is obtained under practical conditions.

8.4 Layout of Accelerator Vacuum Systems

Todays accelerators range from small facilities of some 10 meters in length up to complex accelerator systems comprising tens of kilometers at large national and international research institutes. Most of these accelerators serve as user facilities offering reliable beam conditions to be booked well in advance. High availability and reliability of the whole accelerator complex and all its subsystems are mandatory, often for 24 hours a day, 7 days a week.

Apart from the huge size of todays accelerators, which makes both access and control complicated, the large number of components puts high demands on the reliability. For example, the vacuum systems of the largest accelerators like HERA, LHC or RHIC contain several 10 000 components.

Already when planning and designing a new facility the automation, self-safe operation, accessibility and simplicity have to be taken into account, while keeping the cost reasonable. Another important aspect is to include sufficient redundancy to minimize the number of events where the failure of a single component would stop beam operation. For larger projects, the planning and installation phase easily sums up to a decade. In addition, most accelerators are not fixed systems over their lifetimeof typically 10 years and more. Modifications and substantial upgrades of the accelerator must be possible to improve the beam quality and the experimental conditions.

Following the pressure demands, as described in Sect. 8.3.1, accelerator vacuum systems are ultra-high vacuum systems. The required pressure conditions should be reached soon after initial startup with beam. Thus, short conditioning times in case of dynamic vacuum effects are desirable. Also, in case of venting to atmospheric pressure, the system must be capable of quick recovery.

The vacuum beam pipe of accelerators is closely linked to the accelerating, bending and focussing elements, i.e. resonators and magnets. The magnet costs increase rapidly with the width of the gap; typical gap sizes are a few centimeters. On the other hand the size of the vacuum chamber is closely fitted to the aperture required by the beam. As a consequence, the gas conductance of the beam vacuum chamber is often quite low, e.g. $<100\,\mathrm{l\,s^{-1}\,m}$ (N_2), requiring a large number of relatively small pumps. However, compact magnet designs often permit only limited access for pump connections. In such cases quite sophisticated mechanical solutions have to be designed which are well adapted to the system, e.g. sufficient pumping speed inside long magnets with significant dynamic vacuum load. In this section we will discuss various aspects for designing and constructing accelerator vacuum systems, pumping equipment and frequently used instrumentation.

8.4.1 Pressure Profile Calculations for One-dimensional Systems

For complex geometries of accelerator vacuum systems, modeling of the gas load profile, pumping geometries and gas flow often is quite useful. During the design phase, not only the expected pressure profile could be simulated, but the layout of the vacuum system be optimized. This is particularly true for accelerators with significant amounts of synchrotron radiation, where the profile of the synchrotron radiation flux and the induced gas load has to be integrated into the calculations. Pressure profiles are also crucial in estimating the beam gas-induced background in a detector installed at a certain location of the accelerator. Various methods have been described in literature and several computer programs do exist, e.g. Monte Carlo calculations for complex geometries [38] or commercial programs [39] to simulate and design complex vacuum systems.

Accelerator vacuum systems usually consist of beam pipes with a longitudinal dimension much larger than their transverse size. Therefore, it is a common practice

to compute pressure profiles as a function of the longitudinal position along the beam path. In general, the particle density distribution in a vacuum system is described by a time dependent diffusion equation. However, with practical parameters equilibrium conditions are reached very fast and the time dependence can be neglected. The pressure distribution which is finally reached results from a balance between outgassing from the chamber wall and gas flow into the vacuum pumps. In the one-dimensional case the gas flow $Q(s)$ is proportional to the derivative of the pressure $P(s)$ with respect to the position, and the constant of proportionality is the specific conductance of the vacuum chamber $\mathcal{C}$ [$\mathrm{l\,s^{-1}\,m}$]: $Q(s) = -\mathcal{C} \cdot \partial P(s)/\partial s$. Several formulae for the calculation of conductances for pipes with simple cross-sections are given elsewhere in this book.

Another relation is the continuity equation $\partial Q(s)/\partial s + \mathcal{S} P(s) = q$, where $\mathcal{S}$ [$\mathrm{l\,s^{-1}\,m^{-1}}$] is the specific linear pumping speed and q [$\mathrm{mbar\,l\,s^{-1}\,m^{-1}}$] is the outgassing rate per unit length. These two equations can be combined and deliver the desired diffusion equation

$$\frac{\partial}{\partial s}\mathcal{C}\frac{\partial}{\partial s}P(s) - \mathcal{S}P(s) + q = 0. \tag{8.42}$$

An elegant solution of (8.42), suited for numerical computations, is the piecewise transformation of pressure and flow over sections of vacuum chambers using matrix multiplications with $\alpha = \sqrt{\mathcal{S}/\mathcal{C}}$ [40]:

$$\begin{pmatrix} P(l) \\ Q(l) \end{pmatrix} = \begin{pmatrix} \cosh(\alpha l) & -\frac{1}{c\alpha}\sinh(\alpha l) \\ -\alpha c \sinh(\alpha l) & \cosh(\alpha l) \end{pmatrix} \begin{pmatrix} P(0) \\ Q(0) \end{pmatrix} + \frac{q}{\alpha} \begin{pmatrix} \frac{1-\cosh(\alpha l)}{\alpha c} \\ \sinh(\alpha l) \end{pmatrix}. \tag{8.43}$$

For sections without pumping, $\mathcal{S} = 0$, the elements containing α in (8.43) have to be taken in the limit $\alpha \to 0$. A realistic pressure profile containing 7 sections and computed with the described method is shown in Fig. 8.21.

In practice two principle types of pressure profiles occur. In sections without pumping ($\mathcal{S} = 0$) one obtains a quadratic solution. For example, a beam pipe with two lumped pumps on both ends will develop a quadratic pressure bump in-between the pumps. The quadratic solution is also obtained from (8.43) in the limit $\alpha \to 0$. The case of two equal pumps, each having a total pumping speed S [l/s], separated by a distance l and connected with a pipe of conductance $\mathcal{C}$ and outgassing rate q is an important one since it occurs often in practice. If we center the origin of the coordinate system in-between the pumps, the pressure profile is given by

$$P(s) = \frac{ql}{S} + \frac{q}{8\mathcal{C}}(l^2 - 4s^2). \tag{8.44}$$

For the average and maximum pressure we obtain

$$P_{\mathrm{avg}} = \frac{1}{l}\int_{-l/2}^{l/2} \mathrm{d}s\, P(s) = ql\left(\frac{1}{S} + \frac{l}{12\mathcal{C}}\right), \quad P_{\mathrm{max}} = ql\left(\frac{1}{S} + \frac{l}{8\mathcal{C}}\right). \tag{8.45}$$

These expressions have two terms, the left one is related to the applied pumping speed and the right one to the conductance of the vacuum chamber. It is obvious

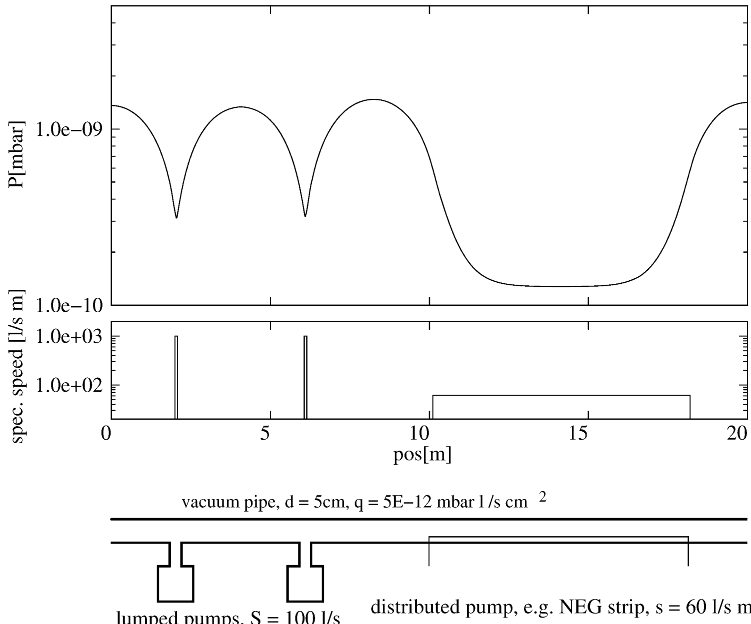


Fig. 8.21. Pressure simulation of a fictitions one-dimensional vacuum system with lumped and distributed pumps. The upper plot shows the computed pressure distribution and the lower one the pumping speed per unit length

that the pressure cannot be improved arbitrarily by installing larger pumps. Even for reasonably large vacuum chamber cross-sections, the average pressure within the beam pipe quickly becomes determined by the chamber conductance. Instead of using a few huge pumps at larger distance, it is rather more efficient to increase the number of small pumps by decreasing their distance. This solution, however, could become quite expensive due to the large number of pumps and connecting ports to the beam pipe.

In any case the conductance must be matched to the speed of the pumps. A good balance is achieved if the right term of the average pressure in (8.45) is in the range of 1–2 times the left one. Whenever the required pumping speed per meter of vacuum system exceeds about 100 l/s the use of distributed linear pumping becomes more efficient than discrete pumping. The other mentioned type of pressure profile occurs in a section with a distributed pump ($\mathcal{S} \neq 0$). Here the pressure follows an exponential function and reaches the value $q/\mathcal{S}$ far from the ends of the section. Both cases are shown in the numerical example in Fig. 8.21.

8.4.2 Pumping

Todays accelerators aim for average pressures ranging from 10^{-7} mbar down to 10^{-11} mbar. For facilities of large length, thus containing a huge number of com-

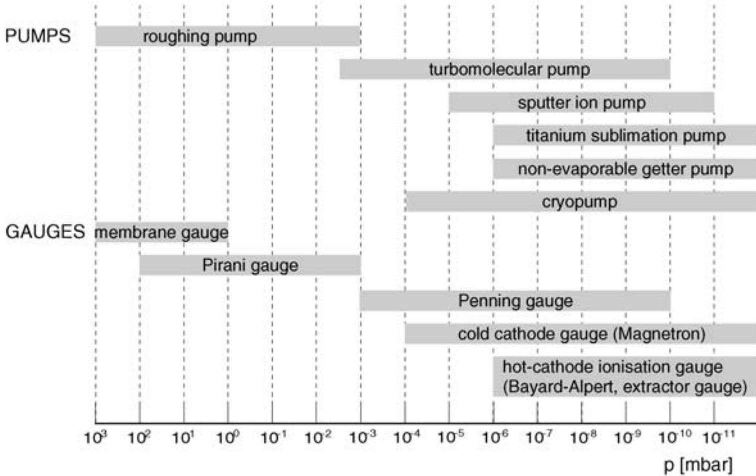


Fig. 8.22. Review of the most commonly used types of pumps and gauges in large accelerator vacuum systems including their operating range

ponents, reliability of the overall system becomes an issue. Pumps without moving mechanical parts are preferred from this point of view.

Figure 8.22 shows a review of the most commonly used types of pumps in large accelerator systems. The functionality of pumps is described more detailed elsewhere in this book. Therefore, we will concentrate here on the pumping aspects relevant to accelerator vacuum systems.

Turbomolecular pumps in combination with rotary pumps, usually installed on mobile pump carts, are used for pump down of the beam vacuum. They are usually disconnected thereafter for reasons of cost saving and to exclude the potential risk of contamination in case of accidental venting and back-streaming of oil from the roughing pump. Sputter ion pumps, titanium sublimation pumps (TSP) and non-evaporable getter (NEG) pumps are almost exclusively used for routine operation in storage rings and large linear accelerators. While sputter ion pumps could be operated already at quite high pressure, their pumping speed is limited. Typical values range from 60 to 250 l/s. Besides the high price, sputter ion pumps also have the disadvantage of heavy weight, e.g. a typical pump with 250 l/s weights about 70 kg. If the operating pressure is low enough, both TSPs and NEG pumps offer significant higher pumping speed, while being much more compact. As the pumping process of both TSPs and NEG pumps is based on chemisorption of gas molecules, they do not pump noble gases and chemically inactive gases like methane. Therefore, they are always used in combination with small sputter ion pumps, which are capable to pump all gases. The application of huge turbomolecular or cryopumps for continuous pumping of the beam vacuum usually is restricted to special applications, e.g. experiments with increased gas load. Cryopumps must be isolated by large aperture gate-valves to allow periodic warm-up.

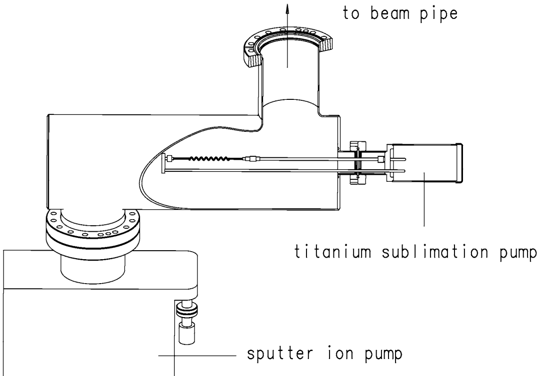


Fig. 8.23. Combination of titanium sublimation and sputter ion pump

Sputter ion pumps are capable of pumping all gases. Hydrogen, the dominating component of the residual gas molecules of a typical UHV system, is pumped by diffusion into the bulk of the cathodes. All other reactive gases are chemisorbed by the cathode material. A typically used material is titanium, which is sputtered onto the walls and anodes by the ions upon incident on the cathodes. Noble gases are physically buried by the sputtered cathode atoms, which may be subsequently uncovered and released into the vacuum. This will induce pressure bursts and thus could disturb the beam significantly. This is one important reason to fix leaks in accelerator vacuum systems as soon as possible to avoid significant amounts of argon or helium to enter into the vacuum. As the pumping speed for noble gases usually is small, it might be increased to values of 25–30% of that for N_2 replacing the cathode plate by heavy material such as tantalum. Such *noble diodes* are often used in systems with enhanced risk of helium leaks as in accelerator sections using superconducting magnets or resonators cooled with liquid helium.

Titanium sublimation pumps are sorption pumps in which titanium is evaporated and deposited on a room temperature inner wall as the getter surface. Even though the TSPs have extraordinarily high pumping speed for active gases, they have limited pumping capacity which is directly proportional to the surface area covered by the titanium film. The saturated pumping surface is renewed by deposition of fresh titanium from a heated filament by sublimation. As a rule of thumb, the saturation time of a pump with 1000 l/s pumping speed is one hour at a pressure of 10^{-7} mbar. A typical arrangement of a TSP in combination with a sputter ion pump is shown in Fig. 8.23 offering a pumping speed of more than 1000 l/s for hydrogen. Usually it is very important that a line-of-sight shield is incorporated such that titanium atoms do not enter the beam chamber or contaminate surfaces, e.g. mirrors, ceramic insulators or instrumentation.

Non-evaporable getter pumps are sorption pumps as well. The NEG material is made of special alloys which form stable chemical compounds with the majority of active gas molecules, while the sorption of hydrogen is thermally reversible. As for

TSPs, NEG pumps have a limited pumping capacity. The NEG material is activated by heating to temperatures of 400–750°C for about 30–60 min, reactivation is done at somewhat lower temperatures for commercially available pumps. During heating the NEG material is not evaporated, but the adsorbed molecules diffuse into the bulk material. The only exception is hydrogen, which is released again into the gas phase, thus requiring other pumps during the process of activation and reactivation. The heating produces empty surface sites for further adsorption of active gases. The NEG material usually is sintered onto flexible strips. In this way the pump geometry can be easily adopted to the space available, and quite compact units can be build. Heating of NEG pumps is mostly done by resistive heating, conductive or radiative heating are other options.

In synchrotron radiation facilities and in high energy electron or positron storage rings the dynamic outgassing induced by synchrotron radiation exceeds the thermal degassing by several orders of magnitude. Most of this radiation is produced within the bending magnets, which are often several meters long. In order to cope with this large degassing, much more pumping speed than in accelerators without synchrotron radiation needs to be installed. One option is to design the vacuum chambers such that the synchrotron radiation will be absorbed at discrete absorbers between magnets only. Massive local pumping must be provided right at the absorbers with as large a conductance as possible between source and pump. Alternatively, the radiation is absorbed continuously and pumps must be distributed evenly along the beam vacuum chamber.

An elegant and cost-effective solution are so-called integrated sputter ion pumps inserted linearly into a channel parallel to and in good vacuum contact with the beam channel [41, 42]. When installed in the arcs, they make use of the magnetic field of the bending magnets of the accelerator. This solution has been adopted at the electron ring of HERA (see Fig. 8.24), PETRA, PEP and TRISTAN reaching typical pump speeds of 25 l/s/m. However, as long as the magnets are not powered, e.g. during initial pump-down, maintenance or even in-between two fillings of the storage ring, these pumps are not active. Thus additional pumps are required to obtain and

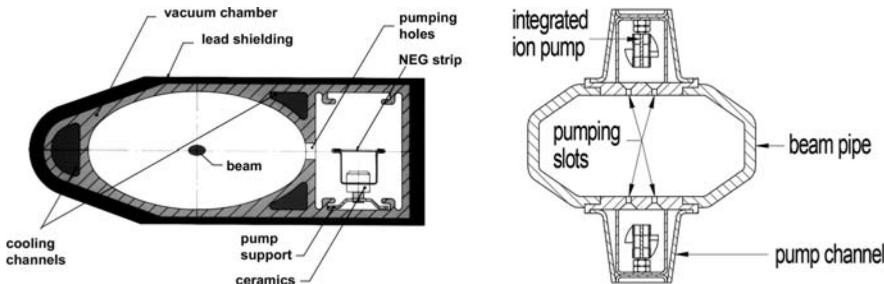


Fig. 8.24. Cross-section of the LEP vacuum chamber (*left*) equipped with NEG pumps [43] and cooling channels. The whole chamber is wrapped into a lead shield to absorb as much synchrotron radiation as possible. Cross-section of the HERA-e vacuum chamber (*right*) with integrated sputter-ion pumps as used for the quadrupole magnets

maintain the vacuum. Furthermore, the magnetic field intensity is decreased when the particles are injected at lower energies, resulting in a reduced pumping speed. This becomes crucial for accelerators where the injection energy and thus the corresponding magnetic field are much lower than the nominal beam energy, such that the discharge in the sputter ion pumps would extinguish and the available pump speed would be insufficient.

An alternative approach without the disadvantages described above is the application of linear strips with NEG material providing initial pumping speeds of more than $1000\,\mathrm{l\,s^{-1}\,m^{-1}}$ next to the strips. This technique has been developed for LEP [43] and is now applied in many accelerators. At LEP about 24 km of the beam pipe have been equipped with 30 mm wide and 0.2 mm thick constantan ribbons coated with 0.1 mm thin NEG material on both sides. As shown in Fig. 8.24, the strips are installed onto a rigid stainless steel carriage via insulating ceramics inside a separate pump channel. For heating the pumps are connected to electric current feedthroughs.

More recently thin film coatings of TiZrV sputtered onto a vacuum chamber have been found to allow even much lower activation and reactivation temperatures down to values of 200°C [44, 45]. These coatings can be passively activated during a vacuum chamber bake-out. If the coating covers the complete inner surface of the beam pipe, not only the pump surface is much larger than using strips, but also the outgassing of the beam pipe itself is drastically reduced as the gas is directly pumped before entering into the gas volume. First tests of coated vacuum chambers in an accelerator have been performed at the European Synchrotron Radiation Facility (ESRF) [46] showing that not only the thermal outgassing, but also the dynamic pressure increase produced by particle bombardment is drastically reduced. Based on the positive results this technique will be used for the room temperature vacuum chambers of the LHC. Also at synchrotron light sources coated chambers are increasingly used, especially in areas with limited access for other pumps. However, one has to keep in mind that the vacuum chambers need to be baked in-situ and the magnets be protected during this process. Detailed studies of these coatings have shown that with increasing number of ventings the activation temperature increases rapidly to values of 300°C. Thus, the advantage of low activation temperatures is restricted to applications where the vacuum system will be rarely vented to atmospheric pressure.

More and more accelerators make use of superconducting magnets or accelerating structures. With the beam pipe being integrated into the cryostat this leads to a *cold bore* vacuum system which takes on the characteristics of a huge cryopump. More details of such systems are discussed in Sect. 8.5.1.

8.4.3 Instrumentation

The beam vacuum system of accelerators usually is divided into several sections using gate valves. Another aspect for segmentation is the accessibility of components, which need more frequent access to the vacuum parts and thus should be easily separated from the remaining vacuum system for venting. Specially designed gate

valves with apertures adopted to the actual beam pipe cross-section and shielding of the gaps in the mechanics to carry the image currents of the beam without major disturbances are available. Fast shutters, which could stop a shock wave in case of unwanted venting within less than 15 ms, are frequently used in sensitive areas, e.g. to separate the beam lines to the experiments from the accelerator at synchrotron light sources.

While the absolute precision of the pressure measurement is of less importance, it is mandatory to monitor the pressure of each vacuum sector of the accelerator continuously. By using an appropriate control system, this allows to immediately diagnose unexpected pressure rises during operation of the facility. Necessary further actions like closing valves are taken to avoid damage to large parts of the accelerator in case of unacceptable pressure rises. In addition, variations of the vacuum pressure are often produced from material being hit by particles originating from changes of the beam parameters and thus can be used for beam diagnostic purposes.

Typical total pressure gauges including their operating range are shown in Fig. 8.22. Each vacuum sector should be equipped with at least one high pressure gauge. At low pressures cold cathode gauges or Bayard–Alpert gauges are frequently used. As the current drawn by a sputter ion pump is proportional to the residual gas pressure down to pressures of 10^{-10} mbar, these pumps supply valuable pressure indications. To save on costly measuring equipment, most accelerators use the current readings which are particularly attractive for large facilities.

Residual gas analysers (RGA) are an essential tool in analysing the residual gas composition of an accelerator vacuum system with respect to both thermal and dynamic outgassing. Due to its high price permanent installations of RGAs are usually restricted to very few applications. As the readout electronics in most cases is very sensitive to radiation damage, special attention must be given to proper shielding.

8.4.4 Material Selection and Design Implications

The choice of materials for the beam pipe and other beam vacuum components strongly depends on the specific requirements of each accelerator. Saving of costs by a careful detail design usually is important, especially for large facilities. The application of industrial processes and methods is favored including series production for most parts of the systems. Nevertheless, specific solutions often need to be worked out for critical components and special areas like experiments. In the following some general requirements and selection criteria for materials and design implications illustrated by specific examples from various accelerators are given. More details might be found in [2, 29], and references therein.

The low pressure required for beam operation in combination with resistance to radiation and corrosive atmospheres usually requires all metal solutions for the beam vacuum system. The selection of the most appropriate beam pipe material is an important decision for starting the design work. The material should be inexpensive, while easy to be produced and machined in large quantities with well established manufacturing processes. Acting as a vacuum vessel, it has to withstand safely the air

Table 8.3. Thermal and electrical conductivity and yield strength $\sigma_{0.2}$ for several commonly used materials of beam vacuum chambers

Material	Thermal conductivity W/K/m	Electrical conductivity $10^6 / \Omega$/m	Yield strength $\sigma_{0.2}$ N/mm^2
SS(316LN)	14	1.4	280–300
Al	229	29	35
AlMgSi0.5	190	29	70
Cu	384	58	40–80
CuSn2	184	25	150

pressure from outside; usually a non-magnetic material is required in order not to disturb the external guide field of the beam. The beam pipe must provide an electrically conducting and smooth wall to carry the image currents of the beam without major disturbances. Good thermal conductivity is required for facilities with high load of synchrotron radiation on the beam pipe wall. The quest for very low pressures, bake-out temperatures of up to 250°C and applications at very low temperatures give further restrictions on the possible choice of materials.

Most accelerators are built out of stainless steel, copper or aluminum vacuum chambers. With respect to both electrical and thermal conductivity, copper and aluminum surfaces are by factors better than stainless steel, see Table 8.3. However, the formation of thick oxide layers, e.g. during cleaning processes, could result in unfavorable electrical properties of the surface as well.

For proton accelerators austenitic stainless steel has become the most widely used material. Beam pipes are often fabricated from seamless tubes with discrete pumps attached every few meters. More complicated parts are machined and welded together. For sealing usually welded stainless steel flanges and copper Conflat^{TM} gaskets are used.

Aluminum is favored for electron facilities due to its high thermal conductivity. The synchrotron radiation can be absorbed directly by the vacuum chamber if the power density is not exceeding values of 100 W/m. With appropriate water cooling of the chamber a power load of several 10 KW/m can be accepted, depending on the height of the synchrotron radiation fan. Often quite elaborate beam pipe cross-sections in combination with pumping and cooling ducts can be economically produced by continuous extrusion, as shown in the example of Fig. 8.24. More complex chambers are produced from solid blocks. Different chambers are often connected by aluminum flanges in combination with aluminum conflat, diamond formed aluminum or Helicoflex^{TM} gaskets. Alternatively, aluminum/stainless steel transitions allow the usage of standard stainless steel Conflat^{TM} flanges. However, the sealing of aluminum systems is not as reliable as the standard stainless steel system with copper gaskets.

In some cases copper chambers instead of aluminum ones are used like at HERA-e, KEK-B and PEP. Here brazing techniques are mostly applied to fabricate the chamber profiles and to connect stainless steel flanges. Due to its even higher thermal conductivity in comparison to aluminum, water-cooled copper blocks are

often used as absorbers for synchrotron radiation at locations of high power density.

When using superconducting magnets and thus cooling the beam pipe to very low temperatures of 4.5 K and below as at HERA-p, RHIC and LHC, further restrictions apply to the possible choice of material due to the thermal stresses applied. Usually the beam pipes are made out of a certain stainless steel type with the inner surface copper coated to enlarge the thermal conductivity. In accelerators using superconducting accelerating structures the cavities and thus the beam pipe are fabricated out of niobium.

The use of other metals like titanium, etc. is limited to special cases. Beam pipes within an experiment often require a high transparency to secondary particles, as the detector arrays for the collision products are mostly located outside the vacuum. Thus extremely thin (0.1 mm stainless steel or aluminum) and low atomic number materials like beryllium, aluminum beryllium or composite materials (carbon fibers) are used. Similarly, exit windows are often made out of beryllium, kapton or mylar foils.

Ceramics, most commonly alumina, are used as insulators, e.g. in electrical feedthroughs. In some cases ceramics are even used for vacuum chambers like inside rapidly changing magnetic fields of kickers (see [47]), or diagnostic elements. Mostly ceramic components are connected to metal by brazing. Glass windows are available in various qualities with respect to absorption and transmission of light. Plastic materials are usually not adequate due to their high outgassing rates and applications are restricted to small quantities.

Inert gas shielded arc welding, electron-beam welding, laser beam welding as well as brazing in a furnace are standard techniques to make all-metal UHV connections. Soldering and glueing are not acceptable for UHV applications.

In most cases the beam pipe must carry the image currents of the beam without major disturbances. Thus variations of the vacuum chamber cross-section must be made gradual and tapered, bellows and pump ports must be shielded. An example of a shielded pump port and bellows is shown in Fig. 8.25. Bellows are often electrically shielded by inserting a set of flexible, sliding spring contacts of copper–beryllium alloys. Perforated electric screens with circular holes of some millimeters in diameter or slits of some centimeters length are used for pump ports forming a compromise in reducing the pumping speed, while keeping the disturbance of the beam at an acceptable level. On the other hand, unproper shielding could lead to untolerable heating of some areas. Consequences are flange connections developing leaks or even melting of material, e.g. in a bellows.

Depending on the degree of radiation produced by the accelerator, all components of the vacuum system installed next to the beam must be resistant to radiation and often also to corrosive atmospheres produced by the primary radiation. This includes also cables and electronics, which must either be properly chosen or well protected. For high energy electron/positron storage rings and synchrotron radiation facilities often the whole chamber is wrapped into a lead shield to absorb as much radiation as possible. Figure 8.26 shows the attenuation of X-rays in various materials as a function of the photon energy.

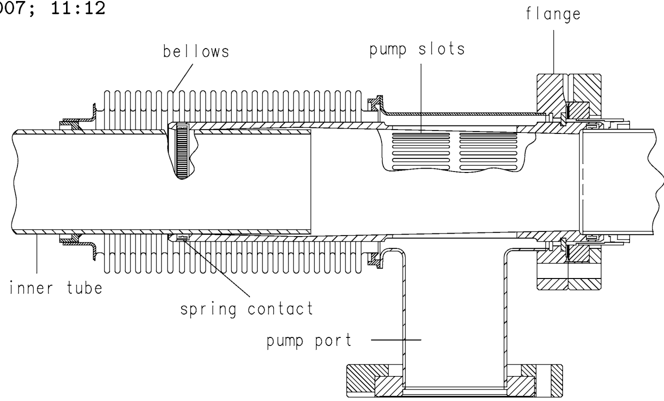


Fig. 8.25. Shielding of a bellows and pump port to minimize the disturbance of the beam as used for a connection in the 4 K cold proton beamline in HERA-p

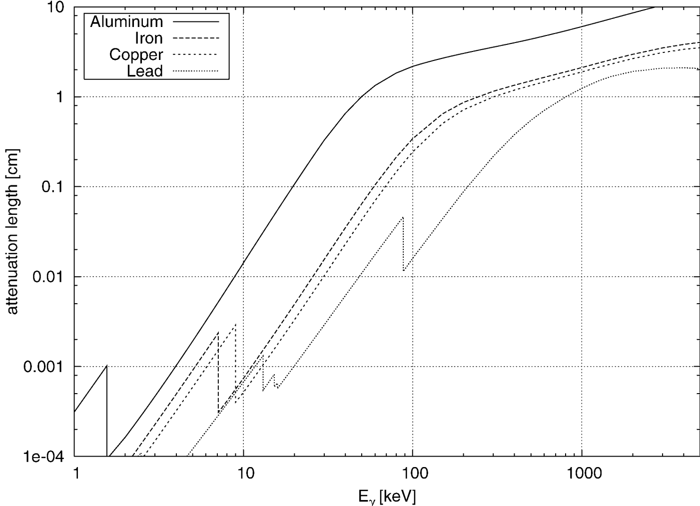


Fig. 8.26. Attenuation of X-rays in aluminum, copper, iron and lead as a function of the photon energy

8.4.5 Cleaning and Treatment of Vacuum Chambers

A proper cleaning and a careful further treatment of all vacuum chambers is crucial for the performance of the beam vacuum system.

Removal of machining lubricants and desorbable gases are necessary before installation of the chambers. Cleaning of all-metal pieces usually starts with a chemical treatment to remove any kind of grease and oxide layers, followed by thorough rinsing with distilled water and drying. More details could be found elsewhere in this book.

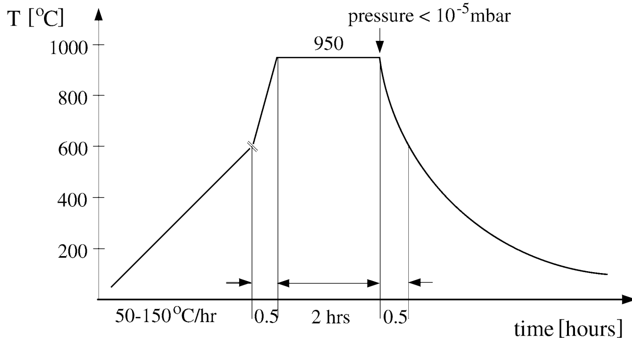


Fig. 8.27. Temperature cycle for the vacuum firing of stainless steel components as developed at CERN

In addition, various heat treatments either during the fabrication process or as an in-situ bake of the complete system after installation often is used to reduce the outgassing. The diffusion of hydrogen out of stainless steel is reduced significantly by firing at 950°C in a vacuum furnace [48]. Figure 8.27 shows the temperature cycle as developed at CERN. This procedure is often applied when aiming for very low pressures like in proton or ion accelerators. A bake-out at temperatures as high as 400°C before installation or in-situ at 250°C for stainless steel or 180°C for aluminum and copper effectively removes adsorbed water and other lightly bound molecules. As the technical effort for an in-situ bake is quite significant compared to the gain in time to start up a facility, this method is less favored for new accelerators except for heavy-ion accelerators or experimental areas requiring the lowest pressure achievable.

In electron or positron rings the most effective treatment after installation is the clean-up/conditioning of the chamber walls by the intense synchrotron radiation itself. Some hundreds of mA-hours of high-energy (multi-GeV) beam operation usually results in acceptable low desorption rates, see Sect. 8.3.2. For some older facilities (ISR, PETRA and CESR) in-situ cleaning by glow discharge in an argon atmosphere has been performed prior to beam operation.

Venting to atmospheric pressure should be done using ultra-pure nitrogen to passivate the surface, thereby avoiding the collection of water molecules onto the surfaces. During short interventions flowing nitrogen through the open chamber during repair or installation work might reduce the necessary pump down time to restart beam operation.

8.5 Special Topics of Particle Accelerator Vacuum Systems

8.5.1 Accelerators with Cold Bore Vacuum Systems

On the way to increase the particle energies to the technical limits more and more accelerators use superconducting elements cooled with liquid helium to temperatures

as low as 2 K. Most of todays high energy proton storage rings use superconducting magnets instead of standard iron yoke magnets. Thus, higher field strengths and smaller bending radii are reached. For these designs the vacuum chamber is integrated into the magnet cryostat. In most cases the liquid helium is not only cooling the magnet itself, but also the beam pipe to temperatures between 4.5–2 K. In a few cases the beam pipe is operated at somewhat higher temperatures in the range 20–70 K. For high energy electron or positron storage rings and electron, positron or proton linear accelerators the use of helium-cooled superconducting accelerating structures has become quite popular, as the technology is nowadays competitive to normal conducting structures. Here the accelerating structures itself represent the beam pipe. This leads to *cold-bore vacuum systems* which take the characteristics of a huge cryopump once the system is cooled down.

In order to cool and operate magnets or accelerating structures at such low temperatures, they are inserted into a steel tank, which will be pumped for reasons of thermal insulation. As can be seen from the example shown in Fig. 8.28, a couple of helium lines with operating pressures of several bars and heat shields are integrated into the insulating vacuum vessel.

When operating the beam pipe at temperatures $\leq$4.5 K, all gases except helium are effectively pumped by the vacuum chamber itself. However, additional pumps are necessary to supply some pumping speed for helium. Pumping is also necessary when the system is not cooled down and during warm up, when the gas trapped by the cold beam pipe is released. Often sputter ion pumps with enhanced helium pumping speed (noble diodes, see Sect. 8.4.2) are connected from outside the insulating vessel

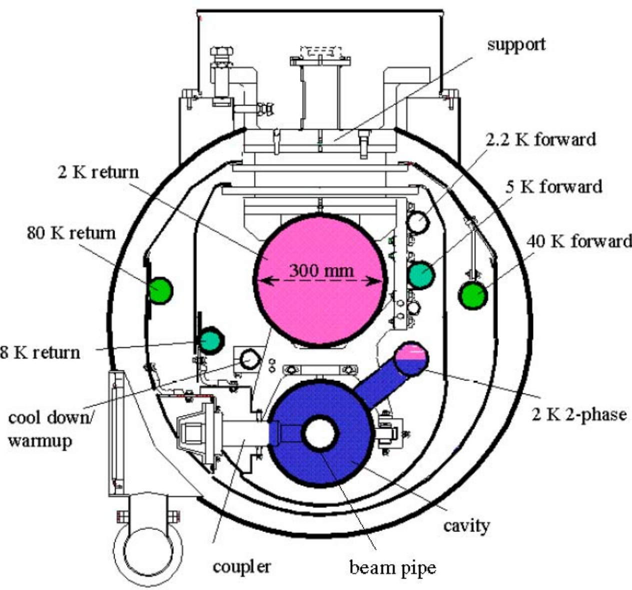


Fig. 8.28. Cross-section of the insulating vessel for the TESLA cavities [16]

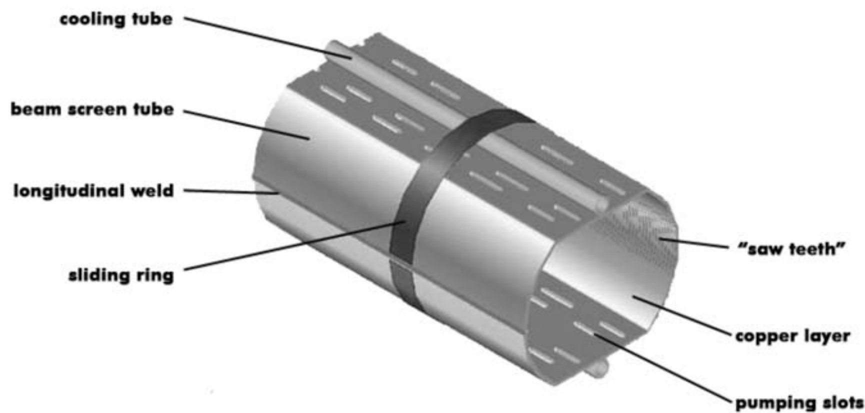


Fig. 8.29. Beam screen for the cold bore beam pipe of the LHC [49]

through conducts to the cold beam pipe at large distances. These pumps could also be used as pressure monitors, as described in Sect. 8.4.3. However, one has to note that the pressure P_w and gas density ρ_w at the room temperature pump are different from P_c and ρ_c inside the cold bore:

$$P_\mathrm{c} = \sqrt{\frac{T_\mathrm{c}}{T_\mathrm{w}}} \cdot P_\mathrm{w}, \quad \rho_\mathrm{c} = \sqrt{\frac{T_\mathrm{w}}{T_\mathrm{c}}} \cdot \rho_\mathrm{w}. \tag{8.46}$$

The pumping of the surfaces of a cold bore beam pipe could become problematic when large amounts of energetic particles hit the cold surfaces. In case of the 7 TeV proton storage ring LHC a significant flux of synchrotron radiation is produced. This could cause the desorption of the condensed gas leading to a positive feedback and pressure runaway. In addition, the incident radiation is equivalent to a heat load which needs to be cooled away by the cryogenic system. As a counteract, so-called *beam screens*, as shown in Fig. 8.29, will be inserted into the cold bore apertures of the magnets [49]. The screens will be kept at somewhat enhanced temperatures between 5 K and 20 K. In order not to loose the benefit of the cryopumping on the cold bore, the screens have narrow pumping slots which allow the gas molecules to leave the beam channel aperture and condense on the 1.9 K cold beam pipe. In this way both the synchrotron radiation induced desorption and the heat load will be kept at a tolerable level.

To minimize the heat transfer through gas convection from the ambient cryostat wall to the magnet cold mass, accelerating structure and helium lines, an insulating vacuum of $<10^{-5}$ mbar is required. The system must be pumped to about $<10^{-3}$ mbar before cool down. The further reduction in pressure will occur by cryopumping on the cold surfaces inside the insulating vacuum vessel. For this purpose mobile pump stations with roughing and turbomolecular pumps are used. Depending on the amount of material installed for insulation, e.g. multi-layer insulating foil usu-

ally containing significant amounts of water, the pump out could last several days. For routine operation the number of pump stations usually is reduced.

For large systems, many insulating vessels are installed consecutively to build up large cryogenic units. This requires sliding sleeves in-between the vessels as well as in-situ welds of the cryogenic lines at each interconnect. The sleeves are either welded or O-ring sealed. For the latter solution one has to take into account the permeation of air into the insulating vacuum, which may result in a substantial additional gas load. A careful testing procedure of all welds is mandatory to avoid serious problems with leaks during operation. Serious leaks need to be located and fixed, while small leaks could be compensated by additional pump stations installed to the insulation vacuum vessel if a repair is not possible. However, in this case one has to take the risk that helium might penetrate into the beam vacuum if there is also a leak between beam and insulating vacuum.

8.5.2 Sychrotron Radiation Facilities

Facilities dedicated to the production of intense synchrotron radiation for experiments have to face various technical challenges specific to these accelerators. The radiation produced inside the bending magnets (see Sect. 8.1.7) cannot completely be guided to the experiments, but huge amounts rather hit the vacuum chamber enclosures. Usually most of this radiation is absorbed by special absorbers made out of pure or coated copper blocks or GlidCop^{TM} (dispersion strenghened copper). A careful cooling of these absorbers is necessary to avoid overheating or large temperature gradients. The remaining radiation hits the beam chamber walls.

Another approach to handle such high loads of synchrotron radiation is the concept of the *chamber–antechamber* design. Here the beam chamber is open to one side by a narrow channel leading to an antechamber, as shown schematically in Fig. 8.30. The synchrotron radiation can pass through the channel to the antechamber where discrete absorbers intercept the photons that are not used. In addition, the antechamber is used for distributed pumping. This way the absorber function is strongly decoupled from the beam vacuum chamber resulting in an increase of thermal stability of

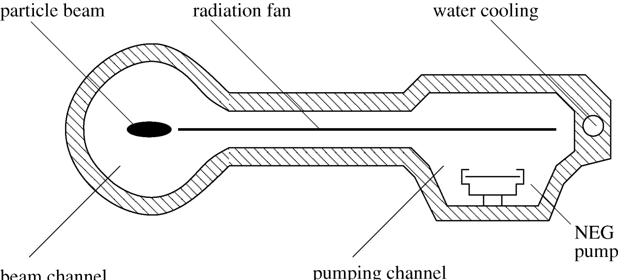


Fig. 8.30. Schematic cross-section of a beam vacuum chamber with antechamber, which is equipped with NEG-strips and water cooling

the beam chamber. This is of importance for sensitive diagnostics like beam position monitors and the position stability of the magnets itself. The chamber–antechamber concept is widely used for third generation light sources.

One important parameter on the way to increase the quality of the produced photon beams is the strength of the magnetic field acting on the beam particles. Strong fields require small gaps between the magnetic poles of the insertion devices. Presently vacuum chambers with less than 10 mm total height are in use, typically several meters long. Although the horizontal dimensions are more relaxed (up to 100 mm), effective pumping often is difficult due to the small conductance. Consequently, such chambers develop pressure bumps which could be quite significant due to synchrotron radiation induced desorption (see Sect. 8.3.2). These pressure bumps could not only lead to significant amounts of radiation by bremsstrahlung, but also influence the beam life time in storage rings. Here conditioning of the chambers with beam is extremely important to reduce the gas load. A recent development to counteract these problems are NEG coatings, as described in Sect. 8.4.2. However, one has to keep in mind that such coatings have a quite rough surface, and the vacuum chambers need to be baked in-situ to temperatures of at least 180°C.

In order to reduce the gap of the insertion devices to its absolute minimum, the whole array of permanent magnets is placed inside the vacuum chamber [50]. Using these *in-vacuum undulators*, a minimal gap size of 4 mm has been achieved.

For the next generation of synchrotron radiation facilities using free electron lasers the beam quality of the electron bunches passing through the undulators is of extreme importance. Ultrashort bunches with very high peak currents, low emittance and energy spread are required. Transporting such intense beams, the impedance of the beam pipe has to be as small as possible. While the resistivity is a property of the chosen material, the surface roughness and discontinuities can be minimized, e.g. for bunches of 25 μm length a surface roughness below 200 nm is required, a challenging goal from the technical point of view.

8.5.3 Particle Free UHV-Systems

For facilities using superconducting accelerating structures of high gradients, the requirements to the cleanliness of the UHV components has become more severe with increasing performance. Similar demands hold for synchrotron radiation beamlines connected to an accelerator using optical components like mirrors for the photon beam transport.

The maximum achievable gradient, i.e. acceleration per length, for superconducting cavities has been improved substantially during the past decade, see Sect. 8.1.3. One major contribution to this improvement is the consistent treatment and preparation of the cavities in clean rooms with procedures similar to standards in semiconductor industry. Dust particles can act as field emitters and thus limit the performance of the superconducting cavities. Therefore, particles on the inner surface of the cavities need to be absolutely avoided. The cold cavities with a wall temperature of a few Kelvin are an integral part of the accelerator beam pipe. Thus, the risk to contami-

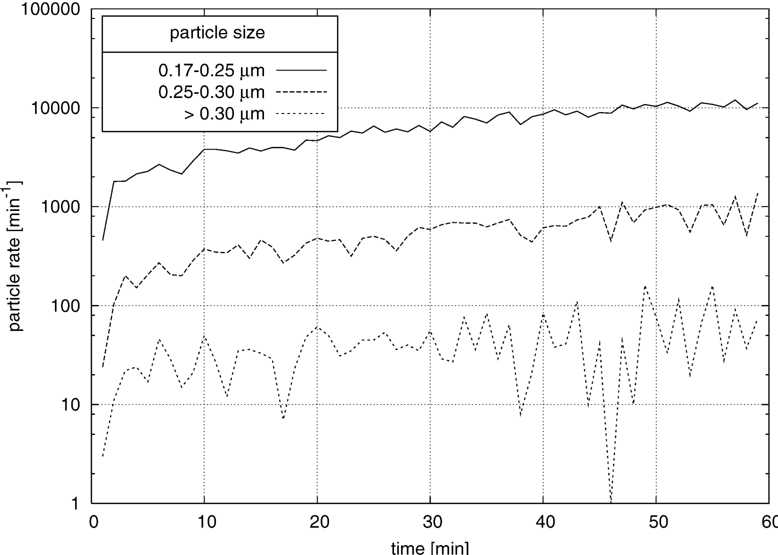


Fig. 8.31. Measured rates of particles introduced into a UHV chamber during venting with air. Inserting a particle filter into the ventline reduces the total number of particles to the required level of a few counts per minute

nate the superconducting cavities with particles from other vacuum sections during assembly and operation must be minimized.

Synchrotron radiation beamlines on free electron lasers or third generation synchrotron radiation sources like storage rings have a large output of coherent photons. Key components in the synchrotron radiation beamlines are mirrors which are used to deflect and tailor the photon beam for experimental needs. Especially in the X-ray regime dust on the mirrors will destroy partially the coherence properties of the beam. To which extent particles on reflecting surfaces have to be avoided is still under investigation.

As a consequence, the preparation of all vacuum components should include procedures similar to those applied to the superconducting cavities. As described in [51], the standard UHV-cleaning processes for degreasing should be followed by further treatment in a clean room of at least class 100 (less than 100 particles >0.5 µm per ft^3). Here particles are effectively removed using an ultra-sonic bath followed by rinsing, both processes using ultra-pure water. Up to 110°C hot filtered air according to class 100 specification or ultra pure nitrogen can be blown through the cleaned components longitudinally to shorten the drying process.

During assembly of the accelerator vacuum system the particle cleanliness is preserved by using local clean rooms. Special care must be given to the pump down and venting procedures, e.g. the usage of small apertures to reduce the gas flow therefore avoiding strong turbulences. Filters prevent particles to enter the vacuum

section during venting. Figure 8.31 shows an example for the particles introduced into a vacuum chamber during venting with air without using any particle filter. Oil-free pump stations are routinely used at such systems.

References

[1] H. Wiedemann, *Particle Accelerator Physics II* (Springer, Berlin, 1994)
[2] A. Chao, M. Tigner, *Handbook of Particle Accelerator Physics and Engineering* (World Scientific, 1999)
[3] K. Wille, *Physik der Teilchenbeschleuniger und Synchrotronstrahlungsquellen* (Teubner, Stuttgart, 1992)
[4] P.J. Bryant, CERN Accelerator School 94-1, CERN (1994)
[5] J.D. Cockroft, E.T.S. Walton, Proc. R. Soc., Ser. A **136**, 619–630 (1932)
[6] M. Letournel, Vivitron Group, Nucl. Instrum. Methods A **268**, 295 (1986)
[7] E.O. Lawrence, N.E. Edlefsen, Science **72**, 376 (1930)
[8] L.H. Thomas, Phys. Rev. **54**, 580 (1938)
[9] T. Stammbach, et al., in *Proc. Cyclotrons and their Applications*, East Lansing, MI, 2001, p. 423
[10] E.M. McMillan, Phys. Rev. **68**, 1434 (1945)
[11] V. Veksler, J. Phys. USSR **9**, 153 (1945)
[12] E.D. Courant, M.S. Livingston, H.S. Snyder, Phys. Rev. **88**, 1190 (1952)
[13] G. Ising, Arkiv för Matematik, Astronomi och Fysik **18**, 1 (1924)
[14] R. Wiederöe, Arch. Elektrotech. **21**, 387 (1928)
[15] C. Reece, et al., in *Proc. Part. Acc. Conf.*, Dallas, 1995, p. 1512
[16] R. Brinkmann, et al. (eds.), TESLA Tech. Design Report, DESY 2001-11 (2001)
[17] International Linear Collider Reference Design Report. www.linearcollider.org (2007)
[18] B. Dwersteg, K. Flöttmann, J. Sekutowicz, C. Stolzenburg, Nucl. Instrum. Methods A **393**, 93 (1997)
[19] J. Peters, in *Proc. of the XXI Linear Accelerator Conference*, Geyongji, Korea. http://linac2002.postech.ac.kr/html/proceedings.html (2002)
[20] E.M. Rowe, F.E. Mills, Part. Accel. **4**, 211 (1973)
[21] V. Ayvazyan, et al., Phys. Rev. Lett. **88**, 104802 (2002)
[22] M. Dohlus, et al., Nucl. Instrum. Methods A **528**, 448 (2004)
[23] E.D. Courant, H.S. Snyder, Ann. Phys. **3**, 1 (1958)
[24] J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975)
[25] A. Liénard, L'Eclairage Elect. **16**, 5 (1898)
[26] J. Schwinger, Phys. Rev. **17**, 798 (1946)
[27] K. Robinson, Phys. Rev. **111**, 373 (1958)
[28] M. Sens, The physics of electron storage rings. An introduction, in *Proc. of the Int. School of Physics Enrico Fermi*, 1971, ed. by B. Touschek
[29] CERN Accel. School on Vacuum Technology, ed. by S. Turner, CERN 99-05 (1999)
[30] C. Benvenuti, R. Calder, O. Gröbner, Vac. **37**, 699 (1987)
[31] R.M. Barnett, et al., Review of particle physics, Phys. Rev. D **54**, 1 (1996)
[32] G.E. Fischer, R.A. Mack, J. Vac. Sci. Technol. **2**, 123 (1965)
[33] M. Bernardini, L. Malter, J. Vac. Sci. Technol. **2**, 130 (1965)
[34] G. Vorlaufer, Ph.D. Thesis, Technische Universität Wien, 2002, p. 200
[35] J.C. Billy, et al., Vac. **60**, 183 (2001)
[36] O. Gröbner, R.S. Calder, IEEE Trans. Nucl. Sci. **NS-20**, 760 (1973)

[37] R. Calder, et al., in *9th Int. Conf. on High Energy Accel.*, Stanford, 1974, p. 70
[38] A. Pace, A. Poncet, Vac. **41**, 1910 (1990)
[39] VacSim^{TM}, http://www.softsim.com
[40] V. Ziemann, SLAC/Pub/5962 (1992)
[41] U. Cummings, et al., J. Vac. Sci. Technol. **8**, 348 (1971)
[42] H. Hartwig, J.S. Kouptsidis, J. Vac. Sci. Technol. **11**, 1154 (1974)
[43] C. Benvenuti, Nucl. Instrum. Methods **205**, 391 (1983)
[44] C. Benvenuti, et al., J. Vac. Sci. Technol. A **16**(1), 148 (1998)
[45] C. Benvenuti, et al., Vac. **53**, 219 (1999)
[46] P. Chiggiato, R. Kersevan, Vac. **60**, 67 (2001)
[47] M. Kinsho, et al., Vac. **73**, 187 (2004)
[48] R. Calder, G. Lewin, Br. J. Appl. Phys. **18**, 1459 (1967)
[49] O. Gröbner, Vac. **60**, 25 (2001)
[50] W. Gudat, J. Pflüger, Z. Phys. B, Condens. Matter **61**, 483 (1985)
[51] U. Hahn, et al., Vac. **73**, 231 (2004)

9

Vacuum Interrupters

R. Renz

9.1 Historical Development

In the medium voltage range the vacuum switching principle is well-established. Today vacuum circuit breakers are available up to 52 kV voltage and 72 kA short-circuit current. Main advantages are:

- high number of operations,
- maintenance-free operations,
- environmental compatibility.

Most of the applications are in power transmission and distribution. But the vacuum interrupting principle is entering new domains which have been dominated by other quenching media in the past.

Different AC switching devices have in common the principle of arc interruption during current zero. The differences are in the extinguishing media. During middle of the twentieth century air- and oil-circuit breakers dominated the medium voltage applications. Since the 1970s SF_6- and vacuum-interrupters have gained attention. Today more than 60% of switching devices applied in the medium voltage range from 7.2 kV up to 50 kV are vacuum interrupters. In the past, the vacuum switching technology was first developed by American (General Electric) and English (Vacuum Interrupter Ltd.) manufacturers, followed by Japanese and Germans (Toshiba, Siemens, ABB).

The main reasons for such a long development time of the vacuum interrupters have been shortcomings in industrial vacuum technology and a lack of fundamental knowledge about the vacuum arc behaviour. Finally, long term basic research and development made the introduction of a new modern switching principle possible. Convincing the customer to move from proved switching principles to vacuum interrupters was not easy in the beginning. It was possible by giving information concerning the function principle, properties and reliability of vacuum interrupters.

At first the number of sold vacuum circuit breakers was not high, but since 1980 there has been a rapid increase in world wide demand. The global production (excluding China) in 2004 was approximately 1.2 million vacuum interrupters for medium voltage circuit breakers. Within Europe (including Russia) the estimated unit quantity was about 400 000 in 2004, e.g. ABB, Eaton Electrical, AREVA and Siemens together sold more than 800 000 units in 2005. In more than 35 years of experience a measure for reliability was obtained in the form of mean time to failure (MTTF) of 40 000 VI-years. This knowledge is an important base for further development of VI's.

The manufacturing technology and design of vacuum interrupters changed parallel with the increased output. In the beginning the tubes were produced in so-called many-fold technology. After pre-assembling and brazing of contact- and housing-parts the tubes were mounted, welded and evacuated via an exhausting pipe. Today the one-shot brazing technology is well established (Fig. 9.1). Here the complete vacuum interrupters are assembled in a dust-protected environment. In a batch of several tens or hundreds the bottles will then be evacuated and brazed simultaneously (one-shot) at 800–900°C in an ultrahigh-vacuum furnace in one process. This technology is an important premise for economical mass-production of vacuum interrupters.

To produce the vacuum in the range of 10^{-6} hPa or less, turbo-molecular pump systems are the best choice. During heating up in the vacuum furnace the tubes are out-baked and water vapour desorbs from the inner surfaces. After cooling down the

Fig. 9.1. One-shot-brazing technique (Siemens). In a batch of several tens or hundreds the bottles will be evacuated and brazed at 800–900°C in an ultrahigh-vacuum furnace in one process (one-shot)

tightness of the bottles has to be controlled. Normally, the PENNING-principle is used, in which the residual gas in the tube is 100% ionized by an impulse voltage in a magnetic field. The ion current then is a direct measure for the internal pressure. Multiple measurements within a certain time give precise information concerning the vacuum tightness.

In the middle of 1970s the first samples for the high voltage range (84 kV) had been developed in Japan. Since 1984 vacuum technology has been applied in the low voltage range (<1 kV) as well. But even in 2004 the vacuum circuit breakers are mainly used in the medium voltage range. New challenges have arisen from the higher ratings of voltages and currents. Even for power plants an increasing trend towards vacuum circuit-breakers is emerging in order to achieve a better system management. In low voltage applications high reliability is necessary too. Maintenance-free operations and high number of operations are to the best advantage to the users. Hermetically sealed housings guarantee the prevention of an external extinction arc and thus a high environmental compatibility.

9.2 Physical Fundamentals

9.2.1 Interrupting Capability

When contacts carrying current are separated, a metal vapour arc will be created. This arc, which consists exclusively of the vaporising contact material, is fed by the external supply of energy until the next current zero. At the instant this current zero takes place, the arc is finally extinguished, and the vacuum interrupter regains its insulating capability and is able to withstand the transient recovery voltage.

Charge carriers in the vacuum arc are metal ions from the anode and electrons emitted from cathode-spots. After current zero the free charge carriers recombine in a few microseconds resulting in a very fast dielectric recovery [1]. On the other hand, the concentration of the neutral metal vapour in the arc decreases in hundreds of microseconds.

At currents around 10 kA the vacuum arc begins to contract, being initially noticeable in the form of anode spots. One way in which the switching capability can be improved is to create a contact geometry with a self-generated magnetic field. Until today, spiral contacts and contrary slotted pot-shaped systems are used [2, 3]. These contacts generate a radial magnetic field (RMF), which causes an azimuthal electromagnetic force acting on the contracted arc. The contracted arc moves over the contact surface at a speed of up to 150 m/s. This high velocity ensures that there is less contact erosion significantly improving the current interrupting capability up to 50 kA.

The movement of the constricted burning arc (Fig. 9.2) is caused by a radial magnetic field component B_r due to the Lorentz force [4]. Because these contact systems also produce an azimuthal field B_φ, a radial force has an effect on the arc movement and arc stability. Hence, both magnetic field components B_r and B_φ have to be optimised in order to ensure a proper switching behaviour.

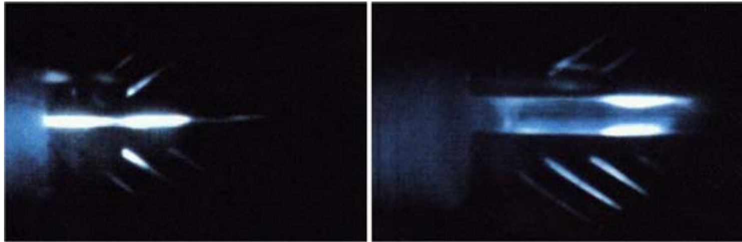


Fig. 9.2. Arc rotation. The movement of the constricted burning arc is caused by a radial magnetic field component B_r due to Lorentz force

For pot-shaped contacts we have some parameters like slot angle, slot number, pot depth and pot diameter which can be varied at a given contact diameter. Considering also manufacturing aspects, one set of optimised parameters results from the calculations. Typical values for the specific magnetic field B/I are:

$$\frac{B_r}{I} = 6 \ldots 12 \frac{\mu T}{A} \tag{9.1}$$

and

$$-5\frac{\mu T}{A} < \frac{B_\varphi}{I} < 0\frac{\mu T}{A}. \tag{9.1$'$}$$

Due to the negative sign of B_φ/I the radial component of the Lorentz force is directed from the contact centre to outwards. In case of a non-optimised system the rotating arc forms jets which can interact with the surrounding interrupter housing. This tendency increases depending on the contact stroke d. A limiting condition is given by

$$I \cdot d < C_L, \tag{9.2}$$

where C_L is a constant for a given contact material and I is the interrupting current. The arc movement is correlated with the radial magnetic field. For a given contact diameter D the number of arc rotations N is approximately

$$N = C_r \cdot I \cdot \frac{B_r}{I} \cdot \frac{d}{D}, \tag{9.3}$$

where C_r is a constant depending on the contact system. The arc movement in (9.3) results from simple thermodynamic considerations of heat dissipation on the contact surface due to the vacuum arc. For maximum ac-arcing times the interrupting current is limited due to overheating of the contact surface. In this case N_{max} is nearly constant, independent of contact diameter D and maximum current I_{max} [5]. This corresponds to the experience that maximum interrupting current I_{max} scales rather well with the contact diameter D for RMF-systems.

A physical measure for the interrupting capability is the maximum amount of electrical charge Q which can be carried by the arc, divided by an effective contact surface A_{eff} (molten area). In this case the experience is summarized by the following relation:

Fig. 9.3. Diffuse burning arc mode. Due to an inverse pinch effect (see below) the arc constriction is shifted towards higher currents

$$\frac{Q_{\mathrm{RMF}}}{A_{\mathrm{eff}}} = \frac{C_{\mathrm{RMF}}}{1 + \Gamma_{\mathrm{RMF}} \cdot d}, \tag{9.4}$$

where C_{RMF} is a constant depending on the contact material and Γ_{RMF} is approximately a constant term, which causes a strong dependence on the contact stroke, as shown in Fig. 9.4. However, in case of a small gap (i.e. less than 5 mm), relation (9.4) becomes independent on the contact stroke, respectively constant. A_{eff} is determined by the arc radius r_{arc} which is weak proportional to $I^{1/3}$ and the rotating orbit length $\pi \cdot D$. Therefore, (9.4) also indicates a linear dependence of the interrupting capability on D.

For a further increase of the switching capability (from 63 kA to 80 kA), e.g. for applications close to generator terminals, the contacts generating an axial magnetic field (AMF) are used. Thereby the contraction of the arc is shifted towards higher currents. The arc is diffuse (Fig. 9.3) and the supplies of arc-energy and the contact erosion are reduced drastically. The main parameters which have to be considered when dimensioning AMF contacts are the size, distribution and phase relation of the axial magnetic field [4]. The goal is to minimise the phase displacement between the high current and the magnetic field it generates.

Also in the case of AMF-contacts different geometries are known. Multiple armed coil systems [6] and identical slotted pot-shaped AMF-contacts [7] are well-established. Based on simple and cost effective manufacturing techniques, an axial field contact with plane slits running in the same direction was developed by Siemens [8]. Accurate magnetic field calculations were performed in order to optimise the contact system. The contacts were designed in such a manner that the axial component of the specific magnetic induction is in the range [4]

$$\frac{B_z}{I} = 2.5 \ldots 5 \frac{\mu\mathrm{T}}{\mathrm{A}}. \tag{9.5}$$

In case of pot-shaped AMF-contacts there are also some geometrical parameters like slot angle, slot number, pot depth and pot diameter which can be varied at a given contact diameter. Hence, it is advantageous to optimise both the axial magnetic field and the ohmic resistance of the contact system.

Considering the interrupting capability, the specific electrical charge is also limited by a relation similar to (9.4)

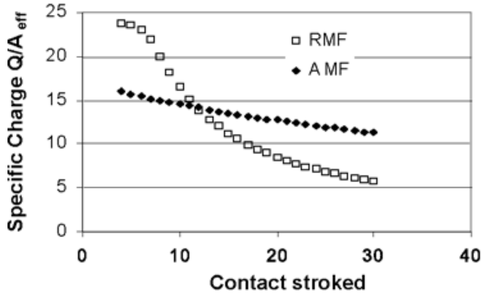


Fig. 9.4. Interrupting capability of RMF- and AMF-contacts. Experimental data in accordance to (9.4) and (9.6)

$$\frac{Q_{\mathrm{AMF}}}{A_{\mathrm{eff}}} = \frac{C_{\mathrm{AMF}}}{1 + \Gamma_{\mathrm{AMF}} \cdot d}. \tag{9.6}$$

Contrary to the strong dependence on the contact stroke in case of RMF, Γ_{AMF} is a weak constant, as shown in Fig. 9.4. C_{AMF} is a constant depending on the contact material.

The effective contact surface (molten area) A_{eff} for the AMF-systems is related with the axial magnetic induction. If we define a yield-factor η_{AMF} by

$$A_{\mathrm{eff}} = \eta_{\mathrm{AMF}} \cdot D^2 \cdot \frac{\pi}{4} \tag{9.7}$$

with D as the contact diameter, our experience gives a very good approximation for η_{AMF} in terms of B_z/I

$$\eta_{\mathrm{AMF}} \approx C_\gamma \cdot \sqrt{\frac{B_z}{I}}. \tag{9.8}$$

If we assume that $B_z/I \propto 1/D$ for this Helmholtz-like coil system, from (9.6)–(9.8) a well-known approximation results

$$Q_{\mathrm{AMF}} = C_{\mathrm{q}}(d) \cdot D^{1.5}. \tag{9.9}$$

Relation (9.9) fits Yanabu's experience [6] for the interrupting capability of AMF contacts very well. $C_{\mathrm{q}}(d)$ is weakly dependent on the contact stroke.

In a first order approximation, (9.4) and (9.6) reflect the energy balance in the vacuum arc. If we assume that the electrical arc energy

$$W_{\mathrm{el}} = \int u_{\mathrm{arc}} \cdot i_{\mathrm{arc}} \cdot \mathrm{d}t \approx u_{\mathrm{arc}} \cdot Q \tag{9.10}$$

is equal to the thermal energy loss at the contact surface

$$W_{\mathrm{th}} = \iint \lambda \cdot \nabla T_{\mathrm{S}} \cdot \mathrm{d}A \cdot \mathrm{d}t \approx \lambda \cdot \frac{\Delta T_{\mathrm{S}}}{\delta} \cdot A_{\mathrm{eff}} \cdot t_{\mathrm{arc}} \tag{9.11}$$

with the thermal conductivity λ, melting temperature T_{S} and penetration depth δ of the contact material, we obtain

$$\frac{Q}{A_{\text{eff}}} \approx \lambda \cdot \frac{\Delta T_{\text{S}}}{\delta} \cdot \frac{t_{\text{arc}}}{u_{\text{arc}}}. \tag{9.12}$$

Further, if the arcing voltage u_{arc} is a linear function in terms of the contact stroke d, (9.12) can be estimated by

$$\frac{Q}{A_{\text{eff}}} \approx \frac{C}{1 + \Gamma \cdot d} \tag{9.13}$$

according to (9.4) and (9.6). Equation (9.8) is only a simple empirical relation fitting the experimental data. The diffuse burning mode of high-current arcs is correlated with the distribution and magnitude of the axial magnetic field [9]. A minimum value of B_z/I is obviously necessary as an existing criteria for the diffuse arc mode.

The arcing voltage also shows different behaviours, dependent on its state. In the diffuse mode, e.g. at low currents (<10 kA) or AMF-arc, the voltage is rather low (20–60 V) and smooth during burning time. At higher currents in case of the constricted mode, the voltage increases up to several 100 V and high-frequency oscillations appear. The arcing voltage is a combination of three parts: the cathode-fall, the anode-fall and the voltage at the plasma-column. Cathode- and anode-fall are in the range of 20 V (CuCr), while the plasma-column voltage increases with increasing current (ohmic behaviour).

In the high-current region (>10 kA) the diffuse AMF-arc is spread over the whole anode surface. Exceeding a certain current density a central arc-constriction appears accompanied with bright anode-spots. The arc mode changes into the diffuse columnar arc. This is regarded as a transition phase to the completely constricted arc at higher current densities [10]. Now the interrupting limit is reached.

A better knowledge about the physics of AMF-arcs is important to increase the interrupting capability. The magneto-hydrodynamic theory leads to the known generalised Ohm's law

$$\mathbf{E} = \frac{\mathbf{j}}{\sigma} + \frac{1}{n \cdot e} \cdot [(\mathbf{j} \times \mathbf{B}) - \nabla p] \tag{9.14}$$

with electrical field vector $\mathbf{E}$, magnetic field vector $\mathbf{B}$, current density $\mathbf{j}$, electrical conductivity σ, plasma density n, elementary charge e and plasma pressure p. If we presume

$$\mathbf{E} = (0, 0, E_z) \quad \text{and} \quad \mathbf{B} = (0, B_\varphi, B_z) \tag{9.15}$$

the Maxwell's equations require in general case non-cylindrical symmetry, which is in accordance to experimental experiences (Fig. 9.3).

Considering rotational symmetry, one obtains

$$\mathbf{j} = (j_{\text{r}}, j_\varphi, j_z) \tag{9.16}$$

with current density components j_{r}, j_φ, j_z depending on B_φ, B_z and E_z. In order to minimise the radial current density component j_{r} and to obtain an optimised inverse pinch-effect, the following boundary condition should be realised:

$$\frac{\partial^2 (B_z)^2}{\partial r^2} < 0. \tag{9.17}$$

Obviously, the radial distribution of the axial magnetic flux between the contacts is also very important for AMF-systems [11].

9.2.2 Dielectric Properties

Approaching current zero in any case, RMF or AMF, the arc mode changes to a diffuse form determined by cathodic spots. Due to a starvation effect in carrier production the arc will chop just before current zero. Natural in case of vacuum arcs this chopping current is related with a steep $\mathrm{d}i/\mathrm{d}t$ which may cause over-voltages in inductive loads. The chopping current depends on the contact material. Therefore, this property has to be optimised during contact material development.

One of the best contact materials also taking this fact into account is a composition of Cu and Cr (Fig. 9.5). For example, CuCr50 has a chopping current below 5 A. Some other very important properties of CuCr are:

- high resistance to arc corrosion,
- high breaking capacity,
- high electrical conductivity,
- low tendency to contact welding,
- high dielectric strength.

After current zero the transient recovery voltage (TRV) imposed by the circuit appears at the contacts within some tens of microseconds. The dielectric strength of the contact gap, however, recovers in a few microseconds. This is correlated to the neutralization of the metal plasma in the arc. The contact gap is now insulating, and the current is successfully interrupted.

In principle, the vacuum stroke is an ideal insulator because there are no free electrical carriers. So the interrupter can withstand very high voltages, even light-

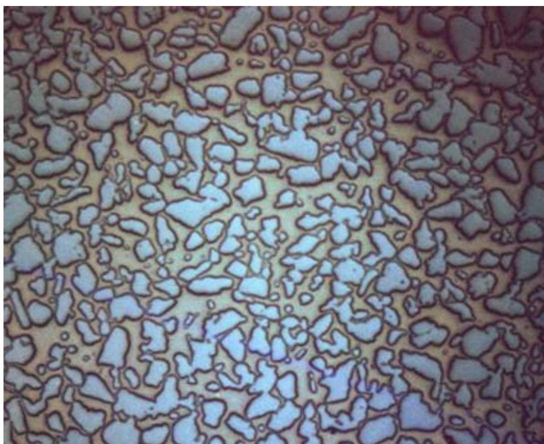


Fig. 9.5. CuCr contact material

ning impulse voltages. Limits are given by current emitting mechanism or particle discharges [12].

Micro-protrusions on the metallic surfaces will enhance the electrical field strength E locally by a factor β of the order of several hundreds, depending on the surface conditions. Due to the quantum mechanical tunnel-effect electrons will be emitted. According to the theoretical description given by Fowler and Nordheim [13], the field-emission current density j_{FE} is

$$j_{FE} = C_1 \cdot (\beta \cdot E)^2 \cdot e^{-\frac{C_2}{(\beta \cdot E)}}, \tag{9.18}$$

where C_1 and C_2 are coefficients in terms of the work function of the material and weakly dependent on the field strength. In practice they may be considered as constants. The field enhancement factor β can be determined experimentally by the current voltage analysis from the plot slope of $\ln(I_{FE}/U^2)$ versus $1/U$. This Fowler–Nordheim analysis (FNA) gives important information regarding the surface condition and dielectric quality.

At a critical electrical field strength E_{crit} the field emission current density increases rapidly producing a discharge,

$$\beta \cdot E_{crit} \approx 10^{10} \, \frac{\mathrm{V}}{\mathrm{m}}. \tag{9.19}$$

Field emission mechanisms limit the dielectric strength in vacuum depending on the local field enhancement, caused by micro-protrusions (surface structure). But there is also a FN-like electron emission from semi-conducting layers on metallic surfaces in vacuum (Shottky emission). These layers may be caused by metal oxides or other surface contaminations. Even the joint of ceramic and metal in the interrupter (triple-junction) produces Shottky emission under the influence of an electrical field. Hence, clean surfaces and field protected triple-junction areas are essential for high dielectrical properties.

The dielectric limits in vacuum due to emission effects are linearly correlated to the electrode distances, hence to the electrical field strength. With increasing contact gap d, however, non-linear relations between the discharge voltage U_d and d appear. This experience is of most importance for vacuum interrupters at rated contact gap. The dielectric strength is nearly square-root dependent from the gap [14]. This square-root law can be described with a particle discharge mechanism [15]. Microparticles are charged under the influence of the local electrical field and accelerated. At a certain particle energy a discharge is initiated due to a secondary process, e.g. particle vaporization at impact. This model leads to the following relationship:

$$U_d = C_p \cdot \sqrt{d}, \tag{9.20}$$

where the constant C_p depends on the material, particle size and surface quality. Because of this it is obvious that a dust-protected environment in the interrupter production line is necessary.

In spite of the extremely clean production conditions a final high-voltage treatment of the tubes is needed. Between the contacts a voltage is applied which is higher

than the internal dielectric strength. This causes powerful internal breakdowns just at the position of protrusions or particles where the electrical field is enhanced. Due to these breakdowns the dielectric disturbances vaporize (micro-explosion) and decrease the local field enhancement (β). With increasing applied voltage step by step and/or decreasing the contact stroke the final desired dielectric strength will be 'produced'. This process is called high-voltage conditioning. The voltage may be AC and/or impulse.

9.2.3 Current-Zero Effects

Approaching current-zero, the plasma burns in a diffuse mode rooting from plenty of cathode spots. Each cathode spot carries a limited amount of current in the range of 100 A, depending on the contact material. Correlated to the extinguishing of the last spot, arc-instabilities appear and the arc 'starves' just before current-zero. The current at this moment is called chopping-current i_{chop}. This chopping-current shows a statistical distribution around a main value, also depending on the contact material. Due to very fast dielectric recovery in vacuum the current chopping results in a steep $\mathrm{d}i/\mathrm{d}t$ in the range of 10^8 A/s.

In the case of an inductive network, these high-frequency phenomena may lead to over-voltages u_{surge} at the load-side of the circuit:

$$u_{\mathrm{surge}} \approx \frac{\mathrm{d}i}{\mathrm{d}t} \cdot L \cdot 2 \cdot \sin\left(\frac{i_{\mathrm{chop}}}{2 \cdot \frac{\mathrm{d}i}{\mathrm{d}t} \cdot \sqrt{L \cdot C}}\right), \tag{9.21}$$

where L and C are the circuit inductivity and capacity. In practice the sine-argument is very small, so the following approximation is sufficient:

$$u_{\mathrm{surge}} \approx i_{\mathrm{chop}} \cdot \sqrt{\frac{L}{C}}. \tag{9.22}$$

For medium voltage circuit-breakers CuCr is well-established with a maximum i_{chop} at 5 A. In the case of contactors WCu-based materials with 3 A and WCAg with 1 A are used as well [16].

9.2.4 Mechanical and Thermal Aspects

In order to realize the switching function in a vacuum-sealed interrupter a moveable contact is needed. A usual number of operations is between 10 000 and 30 000 during an interrupter's lifetime. The most reliable solution is moving via stainless steel bellows. The mechanical lifetime of these bellows is only a matter of dimensioning. If D_{a} and D_{i} are the outer and inner diameter, z the number of waves and s the material thickness, the mechanical life N_{mech} for a certain stroke d is approximately given by

$$N_{\text{mech}} \propto \left[\frac{(D_{\text{a}} - D_{\text{i}})^2 \cdot z}{s \cdot d} \right]^4. \tag{9.23}$$

This strong dependence on geometric parameters allows the designer to optimize the tube's dimensions.

Another demanding application is the making operation on a short-circuit current. Depending on the circuit conditions, a peak current I_{max} in the range of three times short-circuit current will appear, which may be 100 kA or more. Due to current loops in the contact path, repulsive forces F_{c} can lead to bouncing and at least to contact welding. To avoid this, a contact pressure is necessary which is higher than the repulsive force. Usually this is realized by the switching mechanism. The repulsive force F_{c} is approximately given by

$$F_{\text{c}} = K \cdot I_{\text{max}}^2, \tag{9.24}$$

where the constant K depends on the contact system. Due to attractive forces produced in an AMF-system, K is approximately 40% smaller for AMF than for RMF.

In the closed position, the interrupter must also be able to carry the short-circuit current for a few seconds without inadmissible warming or contact welding, according to the standards. The thermal energy loss is determined by the Ohm losses in the contact resistance R_{c} and narrowness resistance R_{n} on the contact surface. R_{c} depends on the surface cleanness and is in the range of a few $\mu\Omega$ for vacuum interrupters, while the narrowness resistance is of the order of

$$R_{\text{n}} = \frac{\rho_{\text{el}}}{D_{\text{c}}}, \tag{9.25}$$

where D_{c} is the diameter of the thermal contact area and ρ_{el} the specific Ohm-resistance of the contact material. In order to minimize the Ohm-losses $I^2 \cdot (R_{\text{c}} + R_{\text{n}})$, ultra-clean surfaces are needed and the thermal contact area has to be maximized.

The time-dependent temperature $T(t)$ related to the initial temperature T_0 on the contact surface can be estimated by

$$T(t) = T_0 + \frac{2}{\sqrt{\lambda \cdot \rho_{\text{m}} \cdot c \cdot \pi}} \cdot \frac{I^2}{D_{\text{c}}^2 \cdot \frac{\pi}{4}} \cdot \left(R_{\text{c}} + \frac{\rho_{\text{el}}}{D_{\text{c}}} \right) \cdot \sqrt{t}, \tag{9.26}$$

where λ is the thermal conductivity, ρ_{m} the mass density and c the specific heat of the contact material (Table 9.1). If $T(t)$ exceeds the melting temperature, welding occurs.

9.3 Present State-of-the-Art and Applications

9.3.1 Vacuum Interrupter Design and Technology

The vacuum interrupter is a switching component for ac-networks. But even in dc-circuits an interruption is possible due to forced current-zero by taking external measures. The principle properties of an interrupter are:

Table 9.1. Material values

Material		$Cu_{300\,K}$	$Cu_{1000\,K}$	$Cr_{300\,K}$	$Cr_{1000\,K}$	$CuCr50_{300\,K}$	$CuCr50_{1000\,K}$
Mass density	ρ_m (10^3 kg/m^3)	8.9	≈8,5	6.9	≈6,7	8	≈7,6
Spec. resistance	ρ_{el} (10^{-8} Ω m)	1.7	≈7	12.9	≈50	5.5	≈20
Therm. conductivity	λ (W/m/K)	400	≈350	100	≈60	200	≈150
Spec. heat	c (J/kg/K)	380	≈500	450	≈680	400	≈600
Melt. temperature	T_m (K)	1383		2203		1383	
Boil. temperature	T_b (K)	2895		2942		2895	

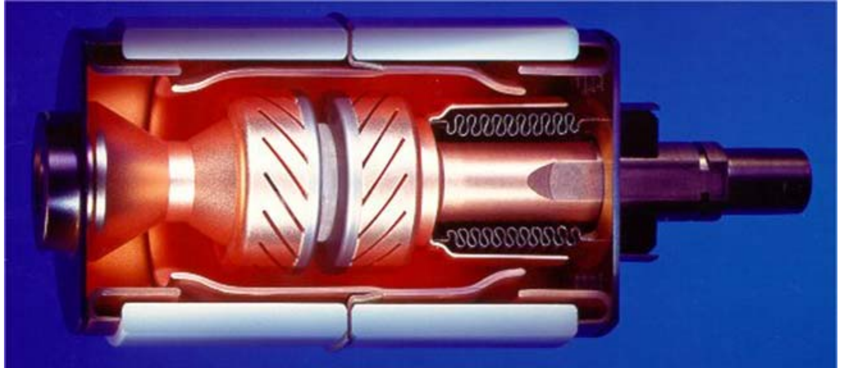


Fig. 9.6. Vacuum interrupter

- conduction of rated (permanent) and short-circuit currents (few seconds),
- interruption of rated and short-circuit currents,
- insulation of rated, power frequency and impulse voltages,
- making (closing) operations at rated and short-circuit conditions,
- mechanical switching operations.

To realize these functions, copper bolts, contact parts, ceramics, internal shields, bellows and a vacuum sealed envelope are necessary. There are many different tube designs depending on the application and "manufacturer's philosophy". For the lower ratings tubes with only one ceramic and one internal shield are sufficient. High-ranged interrupters have two ceramics and sophisticated shield systems (Fig. 9.6). Tubes with an arcing chamber in the middle are also advantageous.

These shield systems protect the inner ceramic surfaces against metal vapor and improve the dielectrical strength. Electrical field calculations and optimizations are state-of-the-art today. The insulators are made of Al_2O_3-ceramic which is metallized

at the joint surfaces to the metal parts. Stainless steel bellows allow up to several million of mechanical operations without any loss of vacuum. All joints are absolutely vacuum-sealed for life cycle by brazing or welding technology. The product life is over 20 years. The brazing materials are based on copper–silver alloys with additions. Copper parts are oxygen-free (OFCu) and all internal materials have only low gas content. Especially the contact material, e.g. CuCr, is extremely outgassed and predominantly free of impurities. Thus, there are only a few suppliers for these high-tech materials word wide.

Knowing the parameters which determine the contact system properties, one is able to take into account the advantages and disadvantages of both worlds, RMF and AMF. One has to compare the switching behavior for different voltage ratings, the ohmic losses during operation with rated current, mechanical forces to keep the contact closed and, of course, the costs.

Looking at the switching behavior, there is a higher interrupting capability at smaller gaps for RMF than for AMF contacts (Fig. 9.4). In practice it is more useful to consider the interrupting capability in dependence on the contact diameter D. From (9.4) resp. (9.9) we get another graph in Fig. 9.7. Due to the condition according to (9.2), a saturation effect occurs for RMF and the interrupting capability is limited in contrast to AMF. Hence, AMF is used at higher voltages and higher interrupting currents.

The RMF arc-mode tends to form jets which may interact with the internal parts of the tube, while the AMF arc-mode has an effect like a magnetic cage. Thus, AMF systems make smaller tube designs possible. Considering the ohmic losses RMF is more advantageous than AMF. This is because of the rather long length of the current path in the AMF contact design. Of course, due to several geometrical parameters AMF systems can be optimized concerning the electrical resistance and the axial magnetic field. For making operations the force to keep the contacts closed is also an important parameter. In accord to (9.24) this force is approximately 40% smaller for AMF than for RMF.

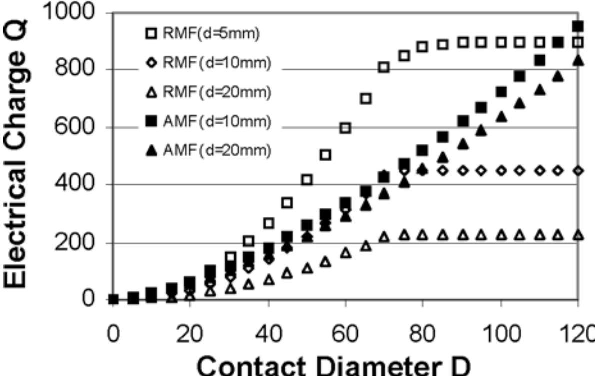


Fig. 9.7. Interrupting capability in dependence on contact diameter and stroke. Saturation effect for RMF due to the limiting rule (9.2)

Table 9.2. Contact system portfolio

	<20 kA	25 kA	31.5 kA	>31.5 kA
≤1 kV	RMF	RMF	RMF	RMF
12 kV	RMF	RMF	RMF	AMF
24 kV	RMF	RMF/AMF	AMF	AMF
36 kV	RMF	AMF	AMF	AMF
>36 kV	AMF	AMF	AMF	AMF

Finally, a more complicated AMF design often results in higher production costs compared to RMF. In view of the above mentioned facts it is obvious that the application of RMF is preferred at lower voltage ratings and AMF is preferred at higher interrupting currents. For the low voltage range, i.e. less than 1 kV, there is no alternative to RMF in order to interrupt proper short-circuit currents. On the other hand, at higher voltages (>24 kV) and higher currents (>40 kA) AMF is the best choice (Table 9.2). Of course, if there is a need for smallest interrupter design at lower ratings, e.g. at 12 kV level, sometimes AMF is used as well. In these cases it depends on customer's requirements.

9.3.2 Medium Voltage Circuit Breakers

Today the main applications of vacuum interrupters are in circuit breakers for the medium voltage range 7.2–50 kV. The rated currents are several hundreds or thousands of amperes. VCB's are able to interrupt the rated current some tens of thousand times without any significant reduction in dielectric strength. Also short-circuit current interruptions up to tens of kiloamperes will be managed properly. Even in the case of power generators more than 72 kA short-circuit currents with a high dc-component will be interrupted more than 30 times.

High dc-components are caused by an asymmetrical shift of the sinusoidal ac-current due to the network behavior after the short-circuit fault. Even the absence of current-zeros is possible during the first power-frequency cycles. After some tens of milliseconds – depending on the time-constant τ of the network – the first current-zero appears, which is essential for current interruption. Modern VCB's have a rather short response time t_o for opening after short-circuit occurs, thus rather high dc-components appear,

$$\mathrm{dc} = \mathrm{e}^{-\frac{t_\mathrm{o}}{\tau}} \approx 0.3 \ldots 0.6. \tag{9.27}$$

Therefore, the inevitable interrupting capability for tubes is given by the electrical arc-charge carrying the asymmetrical current during the arcing-time t_arc,

$$Q_\mathrm{dc} \approx Q_\mathrm{sym} \cdot \left(1 + \frac{\pi}{2} \cdot dc\right). \tag{9.28}$$

The user likes to have rapid interrupters in order to shorten the fault time. Thus, at given ratings highest dc-components are demanded. On the other hand, this leads to larger contact and tube diameter. A balanced design will result in an optimized cost position for the interrupter.

9.3.3 Medium Voltage Contactors

In the medium voltage range vacuum contactors are in operation all around the world: switching transformers, capacitors, reactors, resistive loads or motor starters up to 12 kV. Recently designed vacuum contactors extend the supply range to 24 kV.

The switching frequency is much higher than for circuit-breakers; therefore, contactors and their vacuum interrupters have to exhibit much longer mechanical and electrical lifetimes. As a rule, the minimum required number of switching cycles over the life is one million. Usually the contact material consists of WCAg instead of CuCr commonly used in circuit-breakers. The hard-metal component, tungsten carbide (WC), ensures small erosion caused by arcing [17]. Also with WCAg the chopping current is below 0.5 A. This helps considerably in reducing possible over-voltages caused by the $\mathrm{d}i/\mathrm{d}t$ in high-inductance circuits. But contact materials based on CuCr or WCu are well-established in medium voltage contactors too.

The first use of contactors for the 24 kV level was in 1998 operating the drivers of the TRANSRAPID magnetic transportation system (Fig. 9.8). Here switching devices have to handle higher rated currents and voltages, and also up to one million operating cycles. This large number of operations in the 24 kV range is possible by innovative contactor designs with a special contact material used in the vacuum interrupter.

Fig. 9.8. 24 kV contactor. Switching device for the magnetic transport system TRANSRAPID

Fig. 9.9. Low voltage vacuum interrupter. First vacuum interrupter for 50 kA at 690 V; 80 mm tube diameter

9.3.4 Low Voltage Circuit Breakers

In 1990 the first vacuum interrupter for a low voltage range was introduced into the market (Fig. 9.9). With a lifetime of several 10 000 operations at 2500 A and a few tens of short-circuit current interruptions at more than 50 kA new trends were set in low voltage power engineering.

References are given in facilities with high demands like cranes for open-cast mining. Minimum standstill-times and maximum availability reduce the production costs considerably. Circuit breakers on ships should be maintenance-free and without exhausting systems. Furthermore, in the chemical industry hermetically sealed switching devices are desirable. Vacuum technology has improved itself in plenty of low voltage applications. There is no doubt that the vacuum technology will also be available for ratings higher than 100 kA in the future.

9.3.5 Low Voltage Contactors

The use of vacuum interrupters for low voltage is well-established for contactors. Starting with 630 A contactors in the early 1980s, the development now tends towards smaller currents.

Main problems are the production costs compared to air contactors. New designs and modern vacuum mass-production technologies are imperative in order to introduce economical products (Fig. 9.10). Contactors are available in vacuum technology with a significantly higher electrical lifetime (up to 3 million switching operations) than conventional devices. Besides the compact design and high mechanical and electrical lifetime, the hermetic capsulation of vacuum contactors is a further advan-

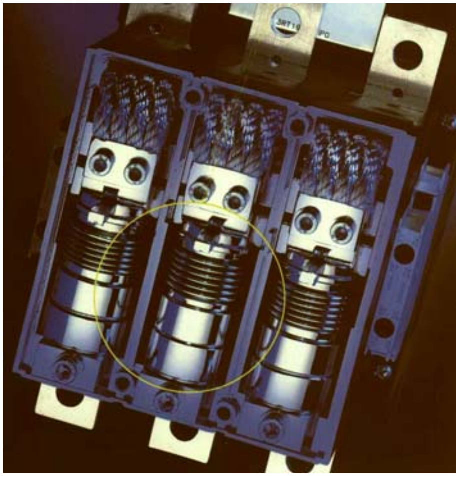


Fig. 9.10. Low voltage contactor: for 300 A at 690 V; 30 mm tube diameter

tage. Because of no interaction between the arc and the environment, applications in aggressive or explosive atmospheres are possible.

9.3.6 High Voltage Vacuum Breakers

By using series arrangements of two or more vacuum interrupters, it is basically possible to double or multiply the dielectric strength without increasing the operation energy substantially. Applications with two 24 kV or 36 kV standard vacuum interrupters in series for rated voltages of 52 kV or 72 kV are well known.

In addition, the rise of the dielectric recovery after current zero is steeper than for a single device. By this the interrupter is able to quench arcs successfully after shorter arcing times compared with a single interrupter [18]. In principle, there is no restriction on the medium voltage range for single vacuum interrupters. Tubes for 72 kV and even more than 125 kV applications are described by Japanese manufacturers [19]. However, due to the nearly square-root dependence of the dielectric strength on the contact stroke [see (9.20)] it is not trivial to find acceptable solutions. Optimized shield systems, excellent surface technologies and mellow contact systems for long-stroke interruption capability (e.g. AMF-system) are essential for high voltage interrupters.

Due to basic research and development high voltage interrupters in 'medium voltage sizes' are possible today (Fig. 9.11). The electrical field strength is minimized and is in the rage of medium voltage bottles. Even the X-ray emission under rated voltage (72 kV) seems to be no problem.

9.3.7 Load Breakers

Unlike the interrupters used in contactors, an interrupter in a load-break switch only has to interrupt currents in order of magnitude of the rated current. The maximum

Fig. 9.11. High voltage vacuum interrupter (prototype): for 72.5 kV and 31.5 kA; 150 mm tube diameter; 500 mm length

mechanical and electrical lifetimes are about 10 000 switching cycles. Switches must also be capable of disconnecting capacitor banks without re-striking. Capacitive switching, for which high dielectrically strength is essential, is also required for frequent disconnections of overhead lines and cables under no-load or low-load conditions. The highest demand on dielectric strength exists when the rated cable-charging breaking current has to be interrupted under earth-fault conditions.

Based on the considerable differences between the contactor and the switch applications, the challenge is to design an interrupter that is economically priced, compact and which can be used for different applications with the same external dimensions and the same internal construction. The difference is only in the contact material.

9.3.8 Transformer Tap Changers

A special application of vacuum interrupters is the use for transformer tap changers. Requirements are large numbers of operations like contactors with current-ratings as circuit breakers. Nevertheless, a very high reliability is essential. This was a great challenge for the tube developers and producers.

Fig. 9.12. Transformer Tap Changer (MR Germany). A special application of vacuum interrupters for transformer tap changers

In the early 1990s the world-wide leading manufacturer of transformer tap changers introduced vacuum switching interrupters successfully (Fig. 9.12). The tap changer operation is divided into three major functions:

- arc interruption and re-closing by use of the vacuum interrupters in conjunction with the associated by-pass switches;
- selection of the next position by a selector switch assemblies in proper sequence with the operation of a vacuum interrupter and by-pass switch;
- operation of reversing or coarse/fine switches in order to double the number of tap positions.

9.3.9 Other Applications

There are many other applications for vacuum switching devices, for example, circuit breakers for railroads in low-cycle networks (e.g. 16 Hz) and power-breakers for nuclear fusion reactors (ITER). In these cases rather long arcing-times compared to normal power frequency (50/60 Hz) and high interrupting currents (up to 80 kA) may appear. This requires tubes with highest interrupting capability.

Vacuum interrupters are not able to interrupt dc-currents. But with the help of current commutation an artificial current zero can be generated in order to extinguish the arc. A typical application is in Tokamak-devices for nuclear fusion experiments [20]. Best experiences were made with the use of vacuum interrupters, and maybe in future the 'sun on earth' will be switched on with vacuum bottles.

For special applications some types of interrupters must be able to operate properly at more severe stress, e.g.

- switches for arc furnaces operate frequently with high currents,
- switches for back-to-back capacitor banks that operate rather frequent inrush currents with high frequency and amplitude,
- switches for small inductive currents which may cause over-voltages at the load-side.

Highly sophisticated mechanical designs, excellent dielectrical performances and suitable contact materials are essential in order to realize these demands.

9.4 Future Aspects

Up to now most of the applications of vacuum interrupters are in the medium voltage range 7.2–52 kV with maximum short current interruption capability of 72 kA. World-wide the manufactures are working on a further increase of the switching capability, e.g. for power generators. The next level to reach is 80 kA according to ANSI standards. But also more than 100 kA seems to be possible. For this high-current interruption optimized AMF-contacts are necessary. A better knowledge about the physics of plasma-arcs is important to increase the interrupting capability. Calculated field distributions combined with plasma models have to reflect the experimental results. Plenty of investigations are focused on this [21].

Looking at the low voltage range a beginning was made with 690 V contactors for 300 up to 800 A. However, the lower the ratings the higher are the production costs. Hence economical limits are given. For the low voltage circuit breakers the situation is quite similar. Low cost solutions for ratings higher than 50 kA are needed. The question is, whether 200 kA interruptions are possible with common contact systems [22].

On the other side also the high voltage regime is ready for vacuum technology. Assuming vacuum interrupters for 125 kV rated voltage with a single contact gap are available in the future, the fiction is a 750 kV arrangement [25]. New technologies like magnetically driven switchgears open a wide field for innovations, e.g. phase-controlled switching just before current zero in order to minimize the arc energy. Intelligent electronics and vacuum electronics will melt together, leading to computer-controlled vacuum interrupters (CCVI).

Further tubes in cast-resin insulated poles allow compact switchgear designs and outdoor applications. Small dimensions and compatibility to older breakers allow a trouble-free replacement [23]. There is also a certain pool for vacuum interrupters in special cases mentioned before, like reclosers, nuclear fusion research (ITER) and high-speed transfer switches.

Increasing production quantities are reflecting the progressive trend of vacuum technique. For example, in China there is a clear decision for vacuum switching-technology [24]. Maintenance-free and large numbers of operations, high reliability

of the vacuum chamber with mean time to failure (MTTF) of several tens of thousands of tube-years, proper switching performance and environmental compatibility predestine the vacuum switching principle for this millennium.

References

[1] G. Lins, The influence of electrode separation on the ion density in the vacuum arc, in *XIVth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Santa Fe, USA, Sep. 1990

[2] H.N. Schneider, Contact structure for an electrical circuit interrupter, U.S. Patent 2,949,520, Aug. 1960

[3] M.P. Reece, A review of the development of the vacuum interrupter, Phil. Trans. R. Soc. London, A **273**, 121–129 (1973)

[4] R. Renz, On criteria of optimized application of AMF- and RMF-contact systems in vacuum interrupters, in *XIXth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Xi'an, China, 2000

[5] E. Dullni, E. Schade, W. Shang, Vacuum arcs driven by cross-magnetic fields (RMF), in *XXth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Tours, France, 2002

[6] S. Yanabu, E. Kaneko, H. Okumura, T. Aiyoshi, IEEE Trans. Power App. Syst. **PAS-100**(4), 1966–1974 (1973)

[7] R. Renz, Axialfeldkontakte für Hochleistungs-Vakuumschaltroehren, ITG-Fachbericht **108**, 262 (1989)

[8] B. Paul, R. Renz, Europaeische Patentschrift 0 155 376, U.S. Patent 4,620,074, 1987

[9] H. Fink, R. Renz, Future trends in vacuum technology, in *XXth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Tours, France, 2002

[10] E. Schade, D. Shmelev, Numerical modeling of plasma behavior and heat flux to contacts of vacuum arcs with and without external magnetic field (AMF), in *XXth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Tours, France, 2002

[11] A. Henon, T. Altimani, P. Picot, H. Schellekens, 3D finite element simulation and synthetic tests of vacuum interrupters with axial magnetic field contacts, in *XXth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Tours, France, 2002

[12] R. Renz, in *Hochspannungstechnik in der Vakuumelektronik*, ed. by Eichmeier/Heynisch, Handbuch der Vakuumelektronik (Oldenbourg Verlag, Muenchen, Wien, 1989), pp. 124–131

[13] R.H. Fowler, L. Nordheim, R. Soc. Proc. **A119**, 173 (1928)

[14] R. Renz, et al., Archiv Bd. **9**, 123 (1987)

[15] L. Cranberg, The initiation of electrical breakdown in vacuum, J. Appl. Phys. **23**(5), 518 (1952)

[16] P.G. Slade, Advances in material development for high power vacuum interrupter contacts, IEEE Trans. Comp. Pack. Manuf. Tech. Part A **17**(1) (1994)

[17] W. Hartmann, L. Kellmann, R. Renz, K.-D. Rohde, N. Wenzel, in *Kontaktwerkstoffe und Pruefverfahren in der Vakuumschalttechnik*, VDE-Fachbericht 55 (VDE Verlag GmbH, Berlin, Offenbach, 1999), pp. 191–196

[18] T. Betz, H. Blechschmitt, H. Fink, M. Heimbach, *Switching Performance of Two Vacuum Interrupters in Series* (Cigre, Paris, 2002)

[19] H. Okubo, S. Yanabu, Feasibility study on application of high voltage and high power vacuum circuit breaker, in *XXth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Tours, France, 2002

[20] I. Benfatto, et al., Life test on vacuum breaking 50 kA unidirectional current, in *IEEE 1990 Summer Meeting*, Minneapolis, MN, 1990, p. 1
[21] *XXth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Tours, France, 2002
[22] R. Renz, On criteria of optimized application of AMF- and RMF-contact systems in vacuum interrupters, in *XIXth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Xi'an, China, 2000
[23] E. Dullni, H. Fink, M. Heimbach, C. Reuber, A family of vacuum circuit breakers with worldwide application using common components, in *CIRED* 2001, Paper 1_8.pdf
[24] W. Jianhua, W. Jimei, Review of theoretical research in vacuum arc and their application in China, in *XIXth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Xi'an, China, 2000
[25] D.-H. Liu, J.-M. Wang, S.-X. Xiu, Z.-Y. Liu, Z.-Y. Wang, R. Yang, Research on 750 kV vacuum circuit breaker composed of several vacuum interrupters in series, in *XXIth Int. Symp. on Discharges and Electrical Insulation in Vacuum*, Yalta, Crimea, 2004

10

Vacuum Electron Sources and their Materials and Technologies

G. Gaertner and H.W.P. Koops

10.1 Thermionic Vacuum Electron Sources

10.1.1 Historical Development

Since the discovery of cathode rays in 1859 by Prof. Julius Plücker in Bonn [1] in the so-called Geißler tubes, manufactured by the skilled German mechanic Heinrich Geißler in Bonn since 1854, who had also developed an improved mercury vacuum pump, cathode ray based tubes have contributed a major part to the technological progress of the western world in the next 150 years to come. Cathode rays are nothing else than electron beams, and it lasted another 32 years since George J. Stoney from the UK introduced the term electron [1]. Usually only the name of T.A. Edison from the USA is known to the broader public, who started to sell incandescent light bulbs for 50 cents in 1883. Yet he had only developed this bulb in parallel to Joseph W. Swan from the UK, who introduced an incandescent lamp with a carbonized bamboo fiber as heated filament and an extraordinary long life of 40 hours in 1878. Edison had already experimented with a lot of different electrode materials, besides carbon-based materials also with Pt, Ir, Ba, Rh, Ru, Ti and Zr, but with limited success. But in the year 1883 T.A. Edison also showed that a current is flowing from the glowing cathode to the anode in a vacuum tube [1].

Typically vacuum electron tubes consist of an electron source, a vacuum or low pressure region where the extracted electrons interact with electric fields of either dc or ac/rf nature (e.g. defined by several grids or metal boundaries at certain applied potentials) and the collector or anode where the electrons are collected and continue to flow into a solid conductor. Our main interest in this chapter will be focused on the emission mechanisms and the different types of electron sources.

Regarding electron emission capabilities as a function of time (Fig. 10.1), there has been a continuous improvement [2, 3]. This was by part triggered by material systems change, partly by improved structural design and in the last two decades by

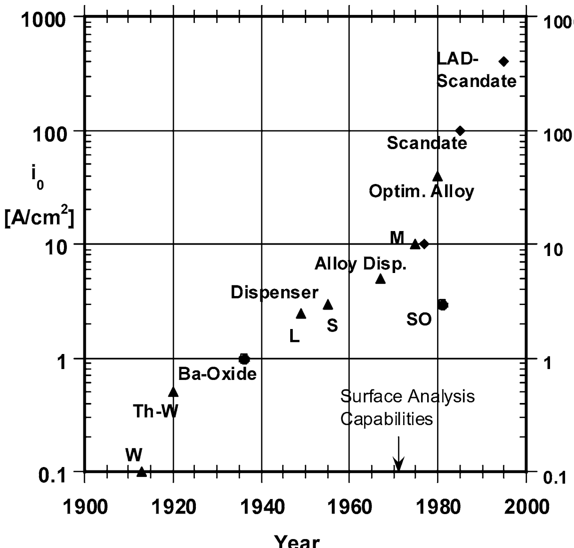


Fig. 10.1. Historical development of thermionic cathodes emission capabilities (life t_{op} at saturated emission current density ≥4000 h) [7]

increased contribution of sub-μm scale technologies, accompanied by high resolution characterization. The emission current densities given in Fig. 10.1 are related to a cathode life ≥4000 h. Of course, the introduction of new cathodes was also linked to certain tube applications.

In 1898 the Austrian Carl Auer von Welsbach introduced osmium wire as electrode in an incandescent light bulb, thus improving the usable life. In 1908 William David Coolidge from the USA found a method to draw thin tungsten wires and equipped incandescent light bulbs with heated tungsten wire coils [1]. Such tungsten filaments are used up till nowadays in X-ray tubes as cathodes due to their rather high insensitivity against gas poisoning and ion bombardment. Yet their emission current density at an operating temperature of about 2600°C is rather low, about 1 A/cm^2.

The oxide cathode was discovered by Arthur Rudolf Wehnelt in 1903 [4], who found that Geißler tubes equipped with such a cathode consisting of alkaline earth oxides could be used as current rectifiers. Since then the oxide cathode has been subject of extensive scientific studies and has become of great technological importance. Curiously enough, this stems from the art of making and utilizing it and not from a thorough understanding of the cathode. Development and improvement was in part guided by the contemporary physical models, which failed in other aspects later on or turned out to be incomplete [5]. In the years before World War II, the predominant view was (e.g. Reimann 1934 [6]) that Ba on the oxide surface, which was supplied by the interface reaction via diffusion, determined electron emission.

Part of the past and present improvements of vacuum electron tubes were made possible by improved cathodes [3]. Especially in the last 50 years Philips, as one

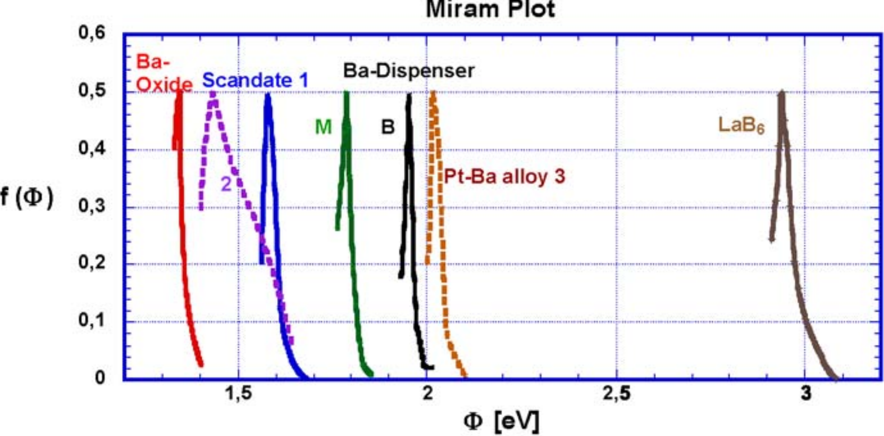


Fig. 10.2. Practical work function distributions (PWFD) or Miram plots, based on underheating or roll-off curves, of "best of class" thermionic cathodes [8]. Data are from Varian except *1* Deckers (mixed matrix scandate: Philips, The Netherlands), *2* Gaertner (LAD top-layer scandate: Philips, Germany) and *3* Djubua (Istok, Russia)

of the major vacuum tube manufacturing companies, was always at the top of innovation in the cathode field, as can be seen from the invention of the L-cathode in 1949 by Lemmens et al. and of the first impregnated Ba-dispenser (or I-) cathodes by Levi in 1955. In 1966 then Zalm et al. introduced the so-called M (magic) cathode with an Os/Ru top-layer on a 411 impregnated tungsten base. In 1995/97 G. Gaertner et al. improved the emission capability of Scandate cathodes from 120 A/cm^2 to 400 A/cm^2 at a true temperature of 1030°C (Os/Ru-I: 15 A/cm^2), which still represents world record in thermionic emission [7]. In a historical review Fig. 10.1 shows the development of the emission capabilities of thermionic cathodes in the last century (life end at saturated emission current density $\geq$4000 h). In Fig. 10.2 practical work function distributions (PWFD) or Miram plots based on underheating or roll-off curves are shown for "best of class" thermionic cathodes [8]. The LAD top-layer Scandate cathode peak (2) is situated at 1.43 eV near the oxide cathode peak.

In the following contribution a review of the basics and of present research and development activities on thermionic cathodes for vacuum tubes is given and also the progress in other emitters and the specific application advantages of the different cathode types are commented. Besides basic investigations of I cathodes and Scandate cathodes also improvements of oxide cathodes are addressed in more detail.

10.1.2 Fundamentals of Thermionic Emission

The basic equations governing thermionic emission or thermal electron emission can be derived straightforward for metals, where the energy levels are occupied up to the Fermi level E_F, which in this case lies in the conduction band. If the metal is heated to a temperature T, some of the electrons in the conduction band acquire sufficient

energy $> E_{\text{min}}$ in order to reach the vacuum level and escape from the metal,

$$E_{\text{min}} = E_{\text{F}} + e\Phi, \tag{10.1}$$

where $e\Phi$ is the work function (Austrittsarbeit).

We can now calculate the density of electron gas in a short distance in front of the metal surface when in equilibrium with the emitter and also the number of electrons crossing a plane (cathode surface A) in this distance at time t_0 in both directions. Under saturation conditions all electrons of sufficient energy leave the cathode, with exception of a small fraction reflected at the surface represented by the reflection coefficient. A detailed derivation can be found in "Moderne Vacuumelektronik" by J. Eichmeier, pages 59–63 [9] and also in some other standard textbooks and review articles on thermionic emission [10, 11, 14, 15].

The differential density dN of electrons in the velocity interval between $v_{x,y,z}$ and $v_{x,y,z} + \mathrm{d}v_{x,y,z}$ is given by

$$\mathrm{d}N = 2v_x A t_0 m_{\mathrm{e}}^3/h^3 * (1/(1+\exp((E - e\Phi)/kT))\,\mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z, \tag{10.2}$$

where $E = m_{\mathrm{e}}(v_x^2 + v_y^2 + v_z^2)/2$. The differential emission current density belonging to dN then is given by

$$\mathrm{d}j_{\mathrm{s}} = e\mathrm{d}N/(At_0). \tag{10.3}$$

By integration of dj_{s} from $v_{x,\text{min}}$ to ∞, j_{s} can be obtained, also taking $E \gg E_{\mathrm{F}}$ into account for emitted electrons

$$j_{\mathrm{s}} = \int \mathrm{d}j_{\mathrm{s}} = \int \frac{e\,\mathrm{d}N}{t_0 A} = \frac{2em^3}{h^3}\mathrm{e}^{E_{\mathrm{F}}/kT} \int_{v_{x,\text{min}}}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} v_x \mathrm{e}^{-\frac{m}{2kT}(v_x^2+v_y^2+v_z^2)}\,\mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z. \tag{10.4}$$

Integration then yields the equation

$$j_{\mathrm{s}} = 4\pi emk^2T^2/h^3 * \exp(E_{\mathrm{F}}/kT) * \exp\left(-mv_{x,\text{min}}^2/2kT\right), \tag{10.5}$$

which is known as the Richardson–Dushman equation. It is usually written in the form

$$j_{\mathrm{s}} = A_{\mathrm{R}}T^2 \exp(-e\Phi/kT), \tag{10.6}$$

where A_{R} is the Richardson (or better thermionic) constant, since the term Richardson constant is often used also for phenomenological values different from 120.4 A cm^{-2} K^{-2}, and $e\Phi$ is the work function. Both constants can be determined from the I/U emission characteristic, where j_{s} is the point of deviation from the space charge limit, via a Richardson plot of $\ln(j_{\mathrm{s}}/T^2)$ versus $1/T$.

Of course, the validity of this equation is limited to ideal cases of metals with uniform surface and zero extraction field at the surface, which in reality is only obtained if the space charge field of the electron cloud in front of the surface is compensated by the external extraction field, i.e. the Laplace field between cathode and anode. Hence saturated thermionic emission is also called zero field thermionic emission. If the fixed thermionic constant is used in a Richardson plot instead of a fitted constant A_{R}, a temperature dependent term in the work function is obtained. Usually, due to the variations in A_{R}, this term is much larger than the theoretical estimation, which is of the order of k/e [11].

The equation can also be applied at very low current densities (e.g. at very low temperatures below the usual operating temperatures), but here in the generally called retarding field region with retarding potential U_a also the contact potential between cathode and anode (= difference between cathode and anode work function $= e\Phi_K - e\Phi_A$; order of magnitude 1 eV) has to be taken into account. The emission current density j_a is then given by

$$j_a = j_s \exp\{(-eU_a + (e\Phi_K - e\Phi_A))/kT\}. \quad (10.7)$$

Using the thermionic cathode at intermediate current densities up to saturation, a space charge limited emission is obtained (virtual cathode at space charge maximum at distance d_m). Here the space charge limited current I_{SCL} is practically independent of the temperature, but depends on the field strength U_a/D in front of the cathode,

$$I_{SCL} = (4/9)\varepsilon_0\sqrt{2e/m_e}(A_k/D^2)U_a^{3/2} = KU_a^{3/2}. \quad (10.8)$$

The geometry factor $K = 2.33 \times 10^{-6} A_k/D^2$ is given in units of A/V$^{3/2}$, where A_k is the emitting cathode surface area, D the cathode to anode distance and U_a the anode voltage in V. Further corrections of space-charge limited emission arise by taking the electron velocity distribution into account; U_a then has to be replaced by the difference of $U_a - U_m$, where U_m is the threshold potential and D_m the distance of the space charge maximum. The equation for I_{SCL} is then modified to [9]

$$I_{SCL} = (4/9)\varepsilon_0\sqrt{2e/m_e}A_k(U_a - U_m)^{3/2}/(D - D_m)^2 * \{1 + 2.66\sqrt{(kT/(eU_a - eU_m))}\}. \quad (10.9)$$

Also in the saturation range a further slight increase of current density with increasing field strength can be observed. This is due to the mirror image charge and is called Schottky effect in honour of Walter Schottky, who first derived it. Extracting the saturation current using an accordingly high anode voltage U_a, e.g. in a diode configuration, reduces the work function slightly by the high field strength E_k at the cathode. The Schottky effect can result in an increase of the saturation current of up to 10% or more. The resulting saturation current density in most cases can be measured in pulsed operation only, due to the high power load on the anode.

The superposition of the $1/r$ potential of the mirror image charge with the Laplace electric field as shown in Fig. 10.3 leads to a reduction of the potential maximum in front of the cathode by

$$\Delta U_k = \sqrt{((eE)/(4\pi\varepsilon_0))}. \quad (10.10)$$

This then implies further modification of the Richardson–Dushman equation by inclusion of this field dependent decrease of the work function [9], where E is given in V/cm and T in K,

$$j_s = A_R T^2 \exp(-e\Phi/kT)\exp\left(4.4E^{1/2}/T\right). \quad (10.11)$$

It has to be pointed out that this equation has to be modified further, since space charge, of course, is also present in the accelerating field range and hence the Laplace

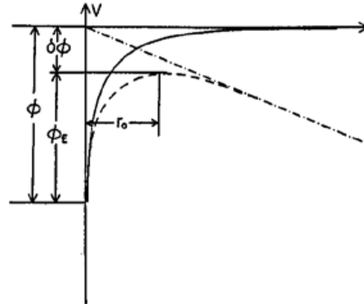


Fig. 10.3. The effect of an accelerating electric field on the work function of a metal: image potential (—); potential due to the applied electric (Laplace) field (-·-·-); and total potential (- - - -) [12]

field is modified by the space charge. A complete theoretical description of this range is given by Scott (numerical approach) and Hasker [13], where Hasker deduced an approximate theory in closed form.

Another reduction of a clean metal work function can be caused by adsorption of an electropositive monolayer. This subtractive dipole moment of the adsorbed atoms reduces the work function by $4\pi ned$, where n is the density of adsorbed atoms per cm^2, e is the electron charge and d is the dipole distance, typically of the order of 0.1 nm. Of course, adsorption of an electronegative monolayer will then lead to a similar increase of the work function [16].

When an electron leaves the cathode surface, it takes with it an average kinetic energy of the amount $2kT$. Also, since the electrons that must replace it are at the Fermi level on average, an additional amount of energy $e\Phi$ is required to raise an electron to the vacuum level. This electron emission cooling effect was first derived theoretically by O.W. Richardson in 1903 and is used for a calorimetric determination of the work function [14].

For high emission currents the cooling effect partly compensates cathode heating (e.g. via resistive or radiative heating) and leads to a current dependent temperature decrease of the cathode. This cooling can become noticeable at operating temperatures and sufficiently high dc-emission currents. For pulsed currents the duty cycle has to be taken into account. In a planar configuration the cooling power ΔP_{cool} is proportional to the the emission current I_{e} and to the average energy uptake of the electrons leaving the cathode according to [15, 19], being valid for the space charge and for the Schottky region:

$$\Delta P_{\text{cool}} = I_{\text{e}}(\Phi + kT/e * \ln(I_{\text{s}}/I_{\text{e}}) + 2kT/e), \tag{10.12}$$

where T is the true temperature. For a cylindrical configuration $2kT/e$ has to be replaced by $5kT/2e$. In case of $I_{\text{s}} = I_{\text{e}}$ the case of onset of saturation or zero field emission I_{s} is obtained:

$$\Delta P_{\text{cool}}(\text{sat}) = I_{\text{s}}(\Phi + 2kT/e). \tag{10.13}$$

The contribution of the Thomson effect usually is negligible. The term with $\ln(I_s/I_e)$ for $I_e > I_s$ also takes the work function decrease due to the Schottky effect in the accelerating field range into account. The cooling of the cathode can be evaluated from the $T(P)$ diagram, where temperature T[K] versus heating power P[W] is plotted. The change in temperature T by $-\Delta T$ (in the ordinate) by emission cooling corresponds to a change in power P by $-\Delta P_{cool}$ (in the abscissa).

10.1.3 Types of Thermionic Cathodes, Properties and Applications

Metal Cathodes

In order to be useful as a cathode, a clean metal must give sufficient emission current density at a temperature where it does not evaporate too rapidly. As a rough guide vapour pressures in the range 10^{-7} to 10^{-9} mbar can be regarded tolerable for most purposes. This limits the choice to a few metals such as W, Re or Ta with a high melting point, which usually belong to the refractory metals and are not so easily machineable, especially tungsten [6].

This choice is substantiated by the following Table 10.1, where melting points and work functions of metals/elements used in vacuum electronics are listed. It is taken from Jenkins [16], with some additions. It has to be noted that to a good approximation the metal work function can be estimated by the image force Coulomb potential at the metal atomic radius: $e\Phi = e^2/(8\pi\varepsilon_0 r)$.

W, Ta and Re have the advantages that due to their high operating temperature contaminating electronegative gases are rapidly evaporated, and thus they are able to emit satisfactorily in poor vacua needing only a small temperature increase to suppress poisoning. For this reason tungsten is generally used in ionization gauges, high power electron guns for electron beam welding and in unbaked demountable systems. The disadvantages are that it needs a high amount of heating power (70 W cm^{-2} at 2500 K), is also subject to a reversible reaction with water vapour and can form

Table 10.1. Usable electron emission from clean metals/elements

Metal	MP (K)	T (K) for $p_{vap} = 1.3 \times 10^{-7}$ mb	A_R (A cm^{-2} K^{-2})	$e\Phi$ (eV)	j_s (A cm^{-2})
W	3640	2520	80	4.54	0.4
Ta	3270	2370	60	4.10	0.6
Re	3440	2330	700	4.7	0.26
Mo	2890	1970	55	4.15	5×10^{-3}
Os	2973	2310	120.4	5.5	6.4×10^{-4}
Ir	2727	1770	120.4	5.4	1.6×10^{-7}
C	4400	2030	48	4.35	2×10^{-3}
Pt	2050	1650	170	5.40	2×10^{-8}
Ni	1730	1270	60	4.1	5×10^{-9}
Ba	1120	580	60	2.11	1×10^{-11}

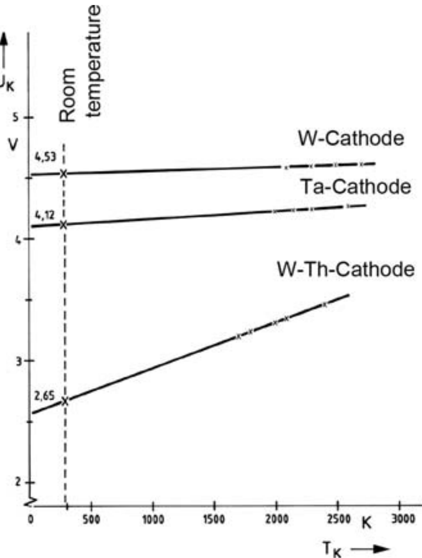


Fig. 10.4. Work function dependence on operation temperature for tungsten (W), tantalum (Ta), and thoriated tungsten (Th–W) wire cathodes [15]

tungsten oxides with oxygen in the rest gas, which may also react to CO with carbon containing contaminants. This is the reason why the more expensive Re or Ir are preferred in mass spectrometers for residual gas analysis.

For geometries other than wire usually Ta is preferred due to its high ductility and hence good machineability. It has the disadvantage that it is embrittled by hydrogen. Os, which was used at the beginning of the last century in incandescent lamps, has the disadvantage of forming a very volatile poisonous osmiumtetroxide with oxygen. The dependence of the work function on the operation temperature is shown in Fig. 10.4 for tungsten (W), tantalum (Ta), and thoriated tungsten (Th–W) wire cathodes.

Monolayer Thin Film on Metal Cathodes

The work function of pure metals can be lowered by a monolayer coating with a suitable element or dipole layer of two elements , as we have seen in the preceding paragraph. Of course, this layer can be provided from an external evaporation source of these elements, but usually this is impractical for long-life cathodes, who need a continuous resupply for compensation of the evaporation or ion bombardment induced loss of this film. Therefore, a reservoir, usually in the interior of the cathode, is needed where the monolayer elements are generated via a chemical reaction and migrate to the cathode surface, e.g. by grain boundary or pore diffusion. Such film cathodes are therefore also called dispenser cathodes [6].

Thoriated Tungsten Cathodes

By provision of an electropositive monolayer of Th on W, the work function of such a system $e\Phi = 2.7\,\text{eV}$ is by far lower than the work function of tungsten of 4.54 eV and even lower than that of Th of 3.5 eV [17].

The supply of Th is achieved by doping W with ThO_2 in form of small embedded grains and by carburization of a top layer of tungsten, e.g. to one third of the W wire diameter. After activation and during life, free Th is then generated via the reaction between thoria and tungsten carbide:

$$2\,W_2C + ThO_2 \rightarrow 4W + 2\,CO + Th. \tag{10.14}$$

Th is migrating to the surface via grain boundary diffusion and spreads on the surface via surface diffusion [16].

Thus, much higher current densities at much lower wire temperatures can be provided by thoriated tungsten cathodes (see Fig. 10.5). The typical operating temperature is 2000 K, where the Th–W cathodes can provide a saturated emission of

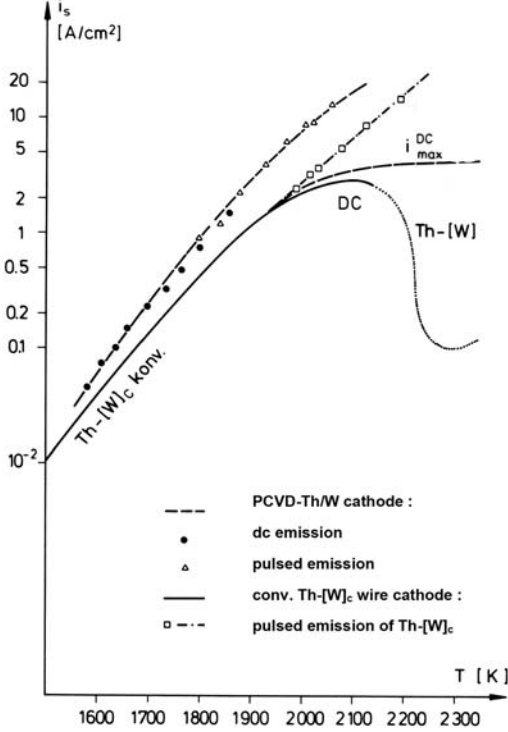


Fig. 10.5. Saturated emission current density as a function of temperature for thoriated tungsten: conventional carburized thoriated tungsten wires were measured in a cylindrical diode. Samples of Th/W cathodes manufactured by plasma-activated CVD were measured in a planar diode configuration [20, 21]

4 A cm^{-2} with a life in excess of 10 000 h. The Th–W cathodes are typically applied in rf tubes such as tetrodes or magnetrons used for rf/radio transmission and rf heating. The thoriated tungsten wires are either assembled into a "squirrel cage" or with an open basket weave, and are directly heated. The valves usually contain Zr heated by radiation from the cathode to act as a getter for the oxidizing gases.

In 1987 G. Gaertner et al. succeeded in providing unipotential cylinder cathodes of thoriated tungsten by plasma-activated CVD from a reactive gas phase containing WF_6/H_2 and Th-β-diketonate starting compounds [21]. There exists a lot of variants of thoriated cathodes, e.g. characterized by replacement of W by Re, Os, Ta at least at the surface or by surface layers with preferential crystal orientation, as pioneered by Ira Weisman [18] for W. Carburization can also be replaced by boriding, e.g. with the aid of diborane, then obtaining a borided tungsten cathode. Also improvements in emission can be achieved by addition of the so-called mobilizers such as Pt, Zr, Hf, Al , Ta enhancing grain boundary diffusion [22].

Lanthanated Molybdenum Cathodes

Replacing thorium by lanthanum and tungsten by molybdenum the lanthanated Mo cathode or LM cathode is obtained, which was introduced and investigated by C. Buxbaum et al. of BBC in the years 1976–1980 [22].

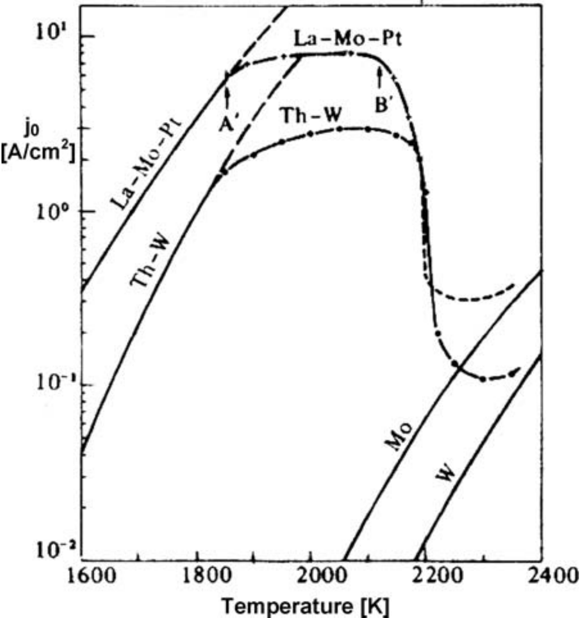


Fig. 10.6. Emission current density of La–Mo–Pt sintered cathodes as a function of temperature. At point A′La evaporation becomes dominating over diffusional supply. Thoriated tungsten, Mo and W are depicted for comparison [22]

The advantage over Th–W cathodes are lower operating temperatures for comparable emission current densities. Thus, in the range 1780–1900 K the LM cathode can deliver 3.5–8 A/cm^2 (see Fig. 10.6). These cathodes are rather insensitive to gas poisoning and can be completely reactivated after venting the chamber or tube. Mo is also easily machineable. The typical dopant amount of La_2O_3 in Mo is 1–4% by weight, the optimum being about 2%. Mo is carburized similar to tungsten. It is also advantageous to coat the surface with a 0.2–10 μm thin Pt layer, where La has a higher sticking coefficient, i.e. the cathode can be stably operated at somewhat higher temperature.

Ba-dispenser or I-cathodes

For high end television tube applications Philips in 1990 introduced Os/Ru coated impregnated (=I) cathodes, which show stable emission performance at 5–10 A/cm^2 up to 25 000 h [23, 27]. Here the cathode base consists of porous tungsten with about 20% porosity, impregnated with $4BaO{\cdot}CaO{\cdot}Al_2O_3$ (=411). The use of osmium

coated and uncoated Ba-dispenser cathodes was a breakthrough with respect to loadability and low temperature film cathode operation. A predecessor of the I cathode was the metal capillary cathode (Siemens designation MK cathode) or reservoir cathode (Philips designation L-cathode), where a reservoir containing a.o. BaO is situated behind a porous tungsten plug [16]. A recent version of it is described by B. Vancil [25]. Figure 10.7 presents the work function dependence on the operation temperature for osmium coated and uncoated Ba-dispenser cathodes, especilly BaO reservoir cathodes. Despite their operating temperature being about 250°C higher than for oxide cathodes related to their higher work function (about 0.4 eV higher)

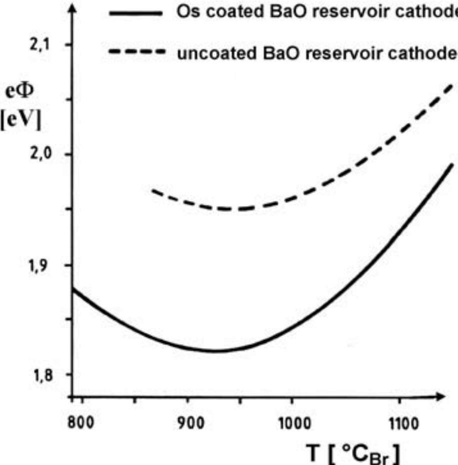


Fig. 10.7. Work function dependence on operation temperature for Osmium coated and uncoated Ba-dispenser cathodes [15]

and their higher production cost, they allow improved brightness and resolution of high end CRTs. The design of the 0.65 watt I-cathode units (see Figs. 10.8 and 10.9) is compatible with 0.65 watt oxide cathode units (see Fig. 10.14), so that they are exchangeable in the same tube type.

One of the important application issues is ion bombardment (IB) resistivity, since ions are created by electron impact on the rest gas. The IB behaviour of several Ba dispenser cathodes such as W-I, Re-I, Ir-I, Os/Ru-I and also Scandate-I was investigated in gun and diode configurations. In accelerated IB life tests at lower ion dose

Fig. 10.8. Philips 0.65 watt I cathode unit for CRTs

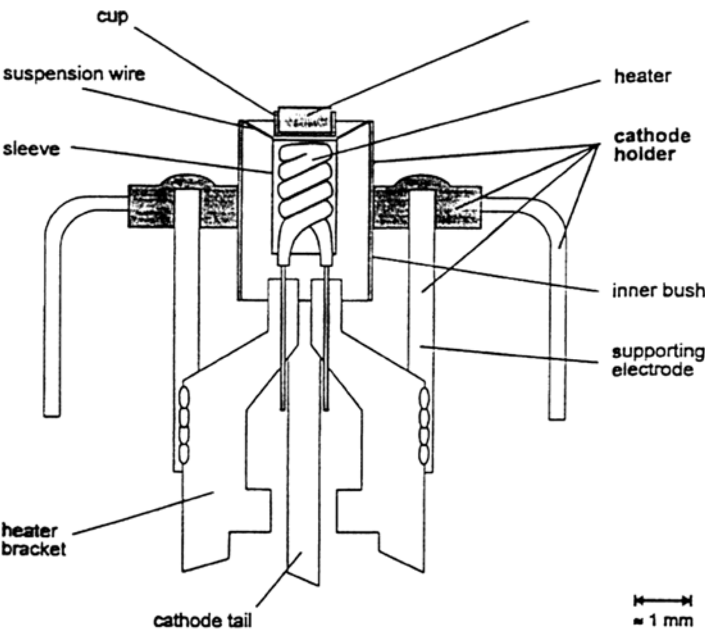


Fig. 10.9. Schematic cross-section of a Philips 0.65 watt I cathode unit

rates (see Fig. 10.11) a "dynamical shielding effect" [36] was observed due to continuous Ba resupply during IB, which improves extrapolation to realistic dose rates. In collaboration with TU Eindhoven, IB characterization was recently extended to more basic I-cathode surface investigations using low energy ion scattering (LEIS) and also Auger parallel to emission measurements [27]. Quantification of the results showed increased neutralization of the scattered Ar ions at I cathode surfaces the lower their work function. It was also observed that Ba will mainly reside on surface sites where oxygen is present in order to form a relatively stable Ba-O dipole layer. Hence IB recovery is also limited by presence and resupply of oxygen.

The highest Ba-O density is found for Os/Ru-I, which also has the lowest effective work function of all I-cathode types of about 1.8 eV. The concept of a fine-grained intermediate layer, in this case W, was also applied to Ir-I-cathodes by Toshiba [28], showing improved ion bombardment resistivity.

Ba-Scandate Dispenser Cathodes

Currently the most prominent thermionic cathode with respect to highest zero field emission current density i_0 is the Scandate cathode, despite the fact that high end applications are still dominated by advanced I-cathode types. The reason for this will be discussed later. Also, in Scandate cathode development a steady improvement of i_0 with time can be observed, linked to different types of Scandate cathodes. Starting with the pressed Scandate cathode, which is based on a patent of Figner in 1967, van Oostrom et al. [29] in 1979 realized $i_0 = 10\,\mathrm{A/cm^2}$ at $950°\mathrm{C}_{\text{Mo-Br}}$ operating temperature (1017°C true temperature). Comparable emission was achieved with the impregnated Scandate cathode invented by Philips (P. Zalm, A. van Stratum) in 1977, where Sc_2O_3 was added to the impregnant. Mixed matrix Scandate cathodes were first introduced by S. Yamamoto (Hitachi) in 1983 [30], where the matrix consisted of a mixture of tungsten and Sc_2O_3; J. Hasker (Philips) improved this in 1989 by using $W + ScH_3$ as a starting powder mixture [31, 32], yielding about $100\,\mathrm{A/cm^2}$ at the standard operating temperature. In 1984 J. Hasker had also pioneered the top-layer Scandate cathode, where a 5 μm layer on top of the porous W body, impregnated with BaCaAluminate, consisted of a mixed matrix of $W + Sc_2O_3$. Further variants of top-layer Scandate cathodes (in short, 'Sc'-I) were then introduced by sputter coating the W-base with $W + Sc_2O_3$(1986). Further improvement was found using $W + Sc_2W_3O_{12}$ (1989) [33] by S. Yamamoto et al. (Hitachi). In 1984 they also published results obtained with a combination of mixed matrix Scandate cathode with Ir, Os, Pt and Mo surface coating. U. van Slooten and P. Duine from Philips report increased ion bombardment resistivity of a Re-coated mixed matrix Scandate cathode in 1996 [34]. Top-layer (=T-L) Scandate cathodes prepared by plasma-activated CVD were investigated by G. Gaertner et al. [35], yielding $60\,\mathrm{A/cm^2}$ at $950°\mathrm{C}_{\text{Mo-Br}}$ in the best version. Currently besides the research activities of Philips and Hitachi on Scandate cathodes, also Hughes (sputtered T-L), Thomson, AEG, Samsung and other companies are active in this field.

To overcome disadvantages of the I cathode like high operating temperature (high Ba production!) and limited current density, Philips research realized a top-layer

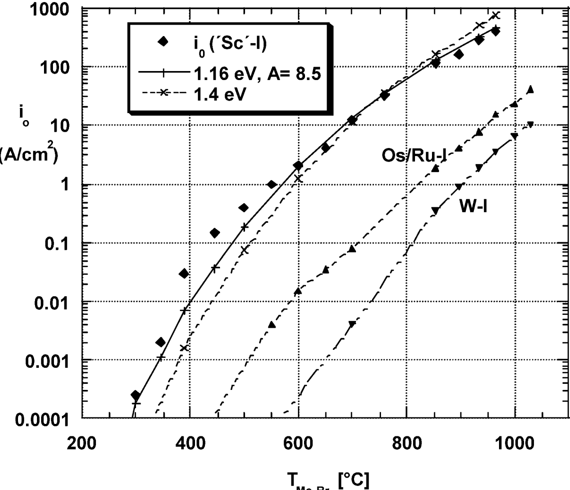


Fig. 10.10. Saturated emission current density i_0 as a function of temperature (Mo-brightness) for LAD top-layer Scandate cathodes ='Sc'/Re-I, Os/Ru-I and W-I cathodes. The *solid line* is the theoretical line for a Richardson work function of 1.16 eV and a Richardson constant of $A_R = 8.5\,\text{A/cm}^2\,\text{K}^2$ [36]

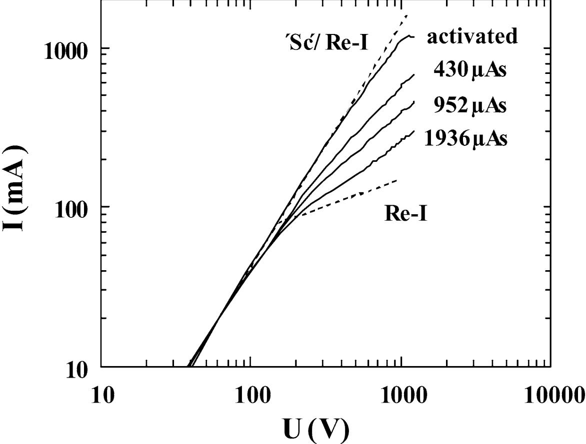


Fig. 10.11. Degradation of I/U characteristic as a function of IB dose: 10^4 µAs in diode @1 mA∗1 mbar∗s in electron gun; *dashed lines*: perfect 'Sc'/Re-I and Re-I cathodes; total emission is superposition of area fractions of scandate (only space charge limited) and Re-I [36]

Scandate cathode by laser ablation deposition (LAD), which is capable of unprecedented 400 A/cm^2 at the same operating true temperature as the I cathode of 1030°C [7, 36]. This allows a reduction of the 'Sc'/Re-I cathode operating temperature and also makes possible new high end applications. The top-layer Scandate cathodes

were prepared by Excimer-LAD of W/Re + Sc_2O_3 on 411 impregnated W-I cathode bases, usually already mounted in 0.65 W cathode units. The top layer of usually 100–500 nm thickness has a very fine grained structure prepared via nanoparticles [24], which is favorable for grain boundary and pore diffusion, a concept which has also been applied for other cathode types. The ion bombardment resistivity and also the emission uniformity of this cathode is strongly improved over previous scandate cathode types, which were either prepared by powder metallurgy or by sputter coating the top-layer. In order to illustrate also the exceptionally high emission of 'Sc'-I at low temperatures, in Fig. 10.10 the saturated emission current density i_0 is shown as a function of temperature (Mo-brightness) over the whole range from 200 to 1000°C for 'Sc'/Re-I, for Os/Ru-I and W-I cathodes [36]. The solid line is the theoretical line for a Richardson work function of 1.16 eV and a Richardson constant of $A_R = 8.5\ \mathrm{A\,cm^{-2}\,K^{-2}}$, as determined from Richardson plot in the interval 600–1000°C. It gives a better fit to the experimental 'Sc'/Re-I data than an effective work function of 1.4 eV (with $A_R = 120.4\ \mathrm{A\,cm^{-2}\,K^{-2}}$).

The measurements underline that the work function remains nearly constant over a wide temperature range. Also field emission with a low threshold of 3.2 V/μm has been observed for 'Sc'/Re-I. In the range from 250 to 350°C thermal assisted field emission is observed and the threshold vanishes [37]. Finally sputtered top-layer Scandate cathodes are investigated by Hitachi, Toshiba and by Y. Wang in China (see [23]) trying to improve them by variation of layer composition and structure. In one version Y. Wang and her group prepared a fine-grained mixed matrix Scandate cathode type with submicron matrix structure and improved emission behaviour [26].

Oxide Cathodes

Alkaline Earth Oxide Cathodes

Oxide cathodes are still lowest in cost. This is due to the rather cheap preparation method of spray coating (Ba,Sr,Ca)- or (Ba,Sr)-carbonate particles on a cathode Ni base. As a result a rather porous structure of about 75% porosity is obtained. Yet they are limited in dc current density to less than 3 A/cm^2. In 1986 Saito et al. from Mitsubishi [38] introduced scandia particle doping of the triple Ba·Sr·Ca-oxide coating in order to increase the critical dc current density $j_{crit,dc}$. P. Derks from Philips then in 1987/1992 replaced scandia finally by rare earth atomic doping of the oxides, especially using coprecipitated europia or yttria [22]. It was later shown by G. Gaertner et al. that the electrical conductivity and hence $j_{crit,dc}$ is increased proportionally to the coprecipitated rare earth addition, yet at the expense of increased oxygen poisoning sensitivity of the oxide cathode [40], if no additional measures are taken.

In Fig. 10.12 the effect of atomic doping with yttrium on conductivity is shown versus operation time. The comparison with the undoped cathode shows, that σ_{el} is increased over life. This was one of the reasons that now oxide cathodes doped with Ni particles are developed. The basic idea is to overcome the conductivity limitation by addition of acicular Ni particles, where 5% are already at the percolation threshold

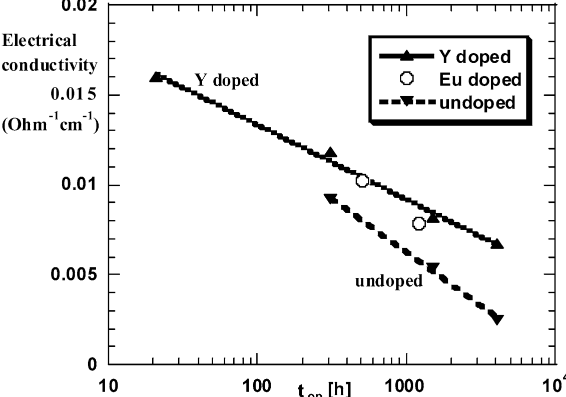


Fig. 10.12. Electrical conductivity σ_{el} of yttria doped and Eu doped Ba·Sr-oxide cathode compared to non-doped cathode versus operation time [40]

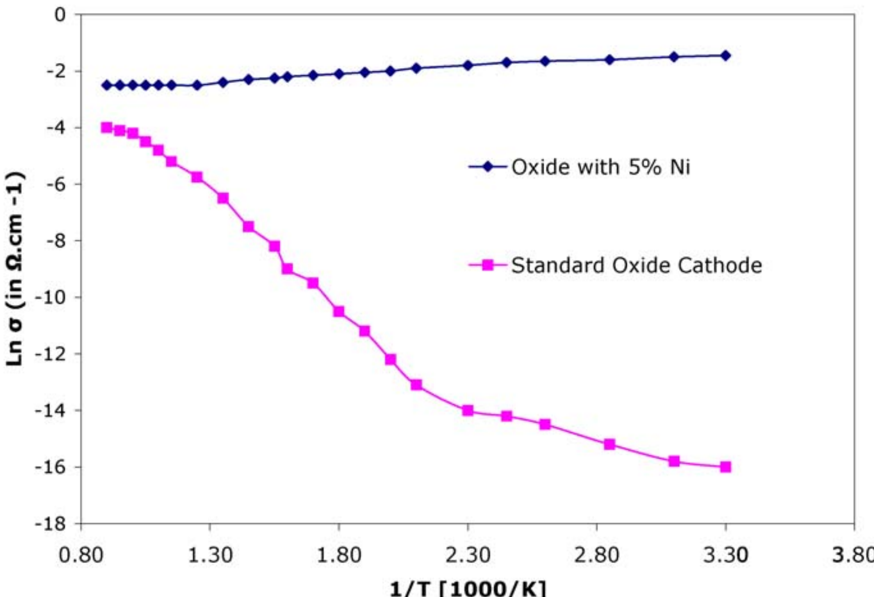


Fig. 10.13. Electrical conductivity versus reciprocal absolute temperature for Philips standard oxide and with 5% Ni particles added [41]

for metallic conduction. Conductivity measurements at the Sheffield–Hallam University for a 5% Ni doped cathode showed about 1×10^{-2} $(\Omega\,\text{cm})^{-1}$ from 1000 K down to room temperature (see Fig. 10.13) [41], whereas oxide cathodes without Ni exhibit a strong conductivity decline with decreasing temperature. For 2.5% Ni addition $\sigma_{el} = 8 \times 10^{-3}$ $(\Omega\,\text{cm})^{-1}$ was directly determined under operating conditions

Fig. 10.14. Philips 0.65 W oxide cathode unit for CRTs

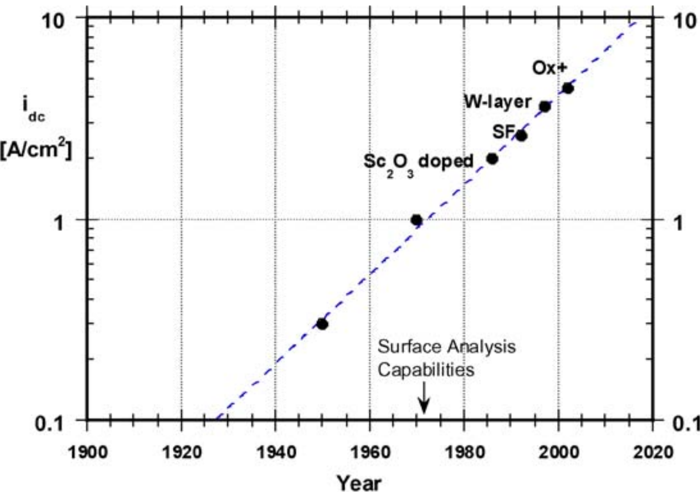


Fig. 10.15. Ba-Oxide cathode dc loadability improvement versus time (typical cathode life > 10 000 h)

at the (true) operating temperature of 1050 K in a planar diode, as described in [40]. Increased dc loadability at lower temperatures can be derived from Fig. 10.13.

As a further beneficial effect, the embedded nickel filaments have also been found to reduce the cut-off drift in CRT applications. This Ni addition is one of the distinctive features of the cermet Oxide Plus cathode of LG-Philips Displays (see Fig. 10.14) besides other improvements, which is able to deliver 3.5 A/cm^2 over life, with peak (dc) values up to 10 A/cm^2 [39]. The increase of the dc loadability of oxide cathodes as a function of time is shown in Fig. 10.15. Of course also a sufficient effort will be needed to continue with this progress in the future.

Improvements of Ba-oxide and Ba dispenser cathodes are also investigated by several other CRT companies and have been reviewed recently by e.g. Higuchi [42]. Therefore, we only want to give a short review of some promising developments. Besides numerous variations of the oxide additives started by Mitsubishi [38], one improvement option in oxide cathodes is the introduction of an additional layer at the interface. Here Samsung [43] introduced a fine-grained Ni top layer on the Ni base of about 1–2 μm thickness, which favors the activator diffusion along the grain boundaries to the oxide coating and implies increased Ba generation. These cathodes, additionally doped with Mg–La compound, can still be operated at 2 A/cm^2 after 12 000 h. Mitsubishi investigated a tungsten intermediate layer, but unfortunately delamination problems had to be overcome.

A further trend are much denser oxide cathodes, like the HIP cathode of NEC [23], where a mixture of Ni, Scandia and (Ba,Sr,Ca)-carbonate powders is pressed into a Ni–Cr container. Cr from the walls is playing the role of the activator. An emission life of 22 000 h at 3 A/cm^2 and 1153 K is claimed. Also, improvement is obtained by a much denser arc plasma deposited oxide cathode with respect to lower BaO evaporation and improved emission.

Other Oxide Cathodes: ThO_2

Thoria coated W, Ta or Ir filaments have been used in rf tubes and especially ThO_2 on Ir is still in use. They deliver current densities of about 5 A/cm^2 at 2000 K due to a low work function of 2.6 eV [93:502-503].

Boride cathodes: LaB_6, CeB_6

The thermionic emission properties of the borides of the alkaline-earth and rare-earth metals and thorium have been investigated in the 1950s [45]. These compounds all have the same formula MB_6 and the same crystal structure consisting of a three-dimensional boron framework in whose interlattice spaces the metal atoms are embedded. The valence electrons of the metal atoms are not accepted by the B_6 complex, thus giving rise to the presence of free electrons which impart a metallic character to these compounds. This, together with the strong bonds between the boron atoms in the framework, produces a series of compounds which have high electrical conductivities and high thermal and chemical stabilities, which are ideal properties for a cathode material. When this structure is heated to a sufficiently high temperature, the metal atoms at the surface evaporate. They are, however, immediately replaced by diffusion of metal atoms from the underlying cells. The boron framework does not evaporate but remains intact.

This process gives a mechanism for constantly maintaining an active cathode surface. Thermionic emission measurements made on these materials show the rare-earth metal borides to be superior to the others. The highest emission was obtained from lanthanum hexaboride, namely 6 A/cm^2 at 1600°C. Its emission constants for the Dushman equation were $e\Phi = 2.66$ eV and $A_R = 29$ A/(cm^2 K^2) [45]. This is higher than the emission normally obtained from thoria H. Ahmed and A. Broers in 1971 obtain [46] $e\Phi = 2.4$ eV and $A_R = 40$ A/(cm^2 K^2). Lanthanum boride has a

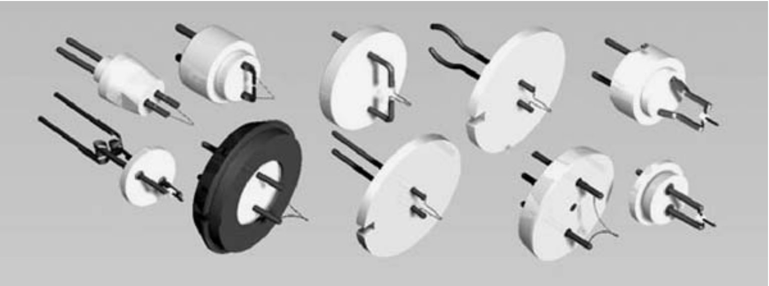


Fig. 10.16. Tungsten and boride cathode mounts for electron guns of many suppliers (courtesy of Kimball Physics [44])

relatively low evaporation rate corresponding to a latent heat of evaporation of 169 kilocalories per mole. If the hexaborides are operated at high temperature in contact with the refractory metals, boron diffuses into their metal lattices forming interstitial boron alloys with them. When this occurs, the boron framework which holds the alkaline-earth or rare-earth metal atoms collapses, permitting the latter to evaporate. However, the hexaboride cathodes may be operated at high temperatures in contact with tantalum carbide or graphite. Typical cathode mounts are shown in Fig. 10.16.

Lanthanum boride cathodes are especially useful in applications where high current densities and electron beams of high brightness are required. Despite these advantages they have a much higher operating temperature than Ba dispenser or Ba oxide cathodes. They are also suitable for high voltage applications because they stand up well under positive ion bombardment. Since they are atmospherically stable and activate easily, they have found wide use in experimental demountable systems.

LaB_6 cathodes use single crystal lanthanum hexaboride as emitter material, having a work function of 2.69 eV. The crystal with <100> orientation is mounted on a carbon heater rod, and held in place by a precision carbon ferrule. Special accuracy is employed to fabricate a round and smooth micro-flat at the top of the crystal of $1.7 \times 10^{-6}\,\text{cm}^2$ area. The heating current of 1.7 A to 2.1 A is delivered through a precision-machined, single-piece carbon rod with central slot and mounting strips. The ceramic sub-base provides rigidity and easier mounting. The Kimball Physics [44] ES-423E (extended life) lanthanum hexaboride cathode is a high performance, resistively heated, thermionic electron source, see Fig. 10.17. Guaranteed is a lifetime of thousands of hours in a clean vacuum system (UHV quality) at a pressure $<10^{-7}$ mbar. Operation at 20–30 A/cm^2 current density is recommended. The reactive lanthanum hexaboride crystal is clamped between or mounted in carbon. The temperature of the crystal is balanced to have high brightness of $10^7\,\text{A/(cm}^2\,\text{sr)}$, at a source size of 5 μm, and to have a low chemical evaporation rate of the crystal. This equilibrium is reached at a crystal temperature of 1700–1900 K. Due to the space charge limited emission from a surface at 1700 K the electron beam energy distribution has a low width of approximately 0.4 eV.

The small area of the heating current loop keeps the unwanted heater current magnetic field low. Because the rod is one single piece, no heating current passes

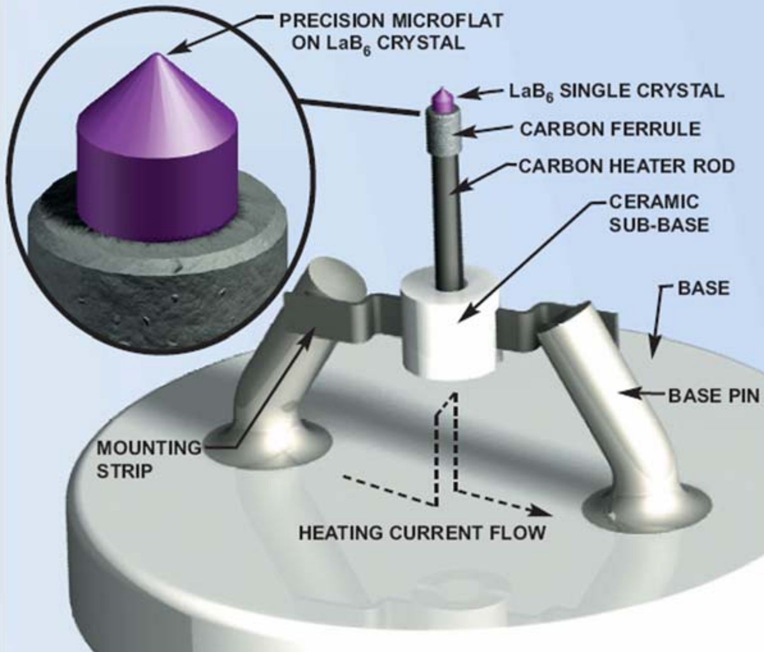


Fig. 10.17. Kimball Physics ES-423E (extended life) lanthanum hexaboride cathode [44]

through the crystal; there are no high temperature current-carrying joints. A high degree of axial symmetry keeps mechanical motions small. The small physical size fits most Wehnelts with ease. In the ES-423E, the crystal can be completely evaporated away without affecting the heating circuit. The very tight tolerances and the enclosed structure prevent the loss of LaB_6 in the mounting region throughout the entire crystal life. Reduced material loss also means less Wehnelt contamination.

In SEM type instruments, lifetimes up to 3000 to 4000 hours may be achieved at operating temperatures of 1850 K (corresponding to material surface loss rates in the 0.025 micron/hour range), with full brightness and excellent stability. With somewhat reduced brightness, as required by typical TEM instruments, lifetimes can be even longer. The ES-423E mounting structure will last more than 10 000 hours. Neither the electrical heating circuit drive impedance nor the thermal properties will drift perceptibly over that period. Chemical reactivity and mechanical drift problems have been eliminated. The real figure of merit of a thermionic electron emitter is the number of coulombs of electrons which may be boiled off per kilogram of cathode surface evaporated away. LaB_6 is an order of magnitude superior to the refractory metals in this key parameter. Any failure of a cathode mounting structure, before the LaB_6 cathode itself has been used up, represents a waste of cathode life. The ES-423E single-piece stress-free ultra-stable carbon mount is unique. Unlike other designs which operate near the temperatures where chemical instabilities will set in, the ES-423E carbon mount is almost impossible to destroy by accidental over-temperature.

The melting point of lanthanum hexaboride itself is somewhat over 2800 K; there have been examples of crystals being melted (extreme over-temperature), in which the Kimball Physics ES-423E carbon mount survived.

Applications

Electron sources with LaB_6 and CeB_6 cathodes are employed in many brightness-limited electron optical systems like scanning electron microscopes, transmission electron microscopes, electron beam lithography systems, electron accelerators, X-ray sources, free electron lasers and other customer applications.

Lifetimes in excess of 6 months of continuous operation are regularly achieved in commercial SEM's and TEM's with suitable gun vacuum. It is based on a well-proven heater structure. Continuous operation at the full operating temperature improves the thermal stability of the gun and hence beam current stability. Stable beam conditions are achieved immediately, e.g. for performing quantitative EEL or EDX measurements.

Alkali Based Emitters

Whereas Ba based thermionic cathode systems are widely used in vacuum tubes and dominate applications in the lower temperature range from 700–1000°C, there only exist academic investigations on alkali based systems due to their high chemical reactivity with oxygen, water and CO_2. Nevertheless, still lower work functions than for the Ba based systems have been achieved with alkali based systems, so that emission current densities in the range of 2–10 A/cm^2 can be reached in the range 400–600°C. This implies a significant reduction of the heating power.

The first investigations of Cs coated W or Cs on oxidized W substrates were conducted by I. Langmuir and K. Kingdon in the years 1924–1930. For this purpose they used a Cs oven heatable to temperatures of about 100°C coating the substrate with the evaporated Cs. Depending on the equilibrium partial pressure and hence the degree of Cs coverage, emission maxima were observed between 400°C and 700°C [47] (see Fig. 10.18). But such a solution with a Cs oven is impractical for vacuum tubes, is risky and too expensive for commercial applications. The dwell-time of Cs on the surfaces is short and is decreasing with increasing temperature. According to Fomenko [48] the work function of Cs on W (optimum coverage for about a monolayer) is about 1.5 eV, for Cs-O-W the work function F is about 1.44 eV and for Cs-O_2-W ca. 1.2 eV. Lower work functions are obtained on Ni, Ti, ZrC and TaC bases ($F\{\text{Cs-Ni}\} = 1.4$ eV; $F\{\text{Cs-Ti}\} = 1.2$ eV; $F\{\text{Cs-ZrC}\} = 1.3$ eV; $F\{\text{Cs-TaC}\} = 0.85$ eV).

For application in thermionic converters A. Makarov et al. [49] investigated Cs films on pyrolytic graphite in an arrangement similar to the one of Langmuir. They obtained a current density of 75 A/cm^2 at 580°C operating temperature and a Cs reservoir temperature of 220°C. They had to cope with the same disadvantages of alkali based systems: a saturation of the tube rest gas with Cs vapour is of course not acceptable.

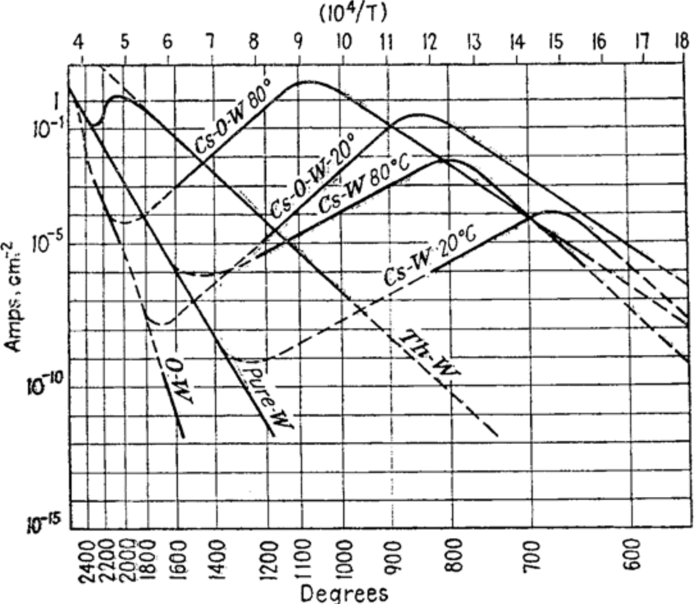


Fig. 10.18. The emission of monoatomic films on tungsten, including Cs on W and Cs-O-W (after Dushman [47])

Only with the introduction of a local supply mechanism of Cs from a moderately heated CsAu compound, the application of Cs based systems with low work function became feasible. This was achieved by Philips in the context of replenishing Cs on semiconductor cold cathodes, which increases the efficiency of ACC cathodes (avalanche cold cathode) by orders of magnitude [50].

Well known is also the application of alkali based photo cathodes with low work functions of 1.5 to 1 eV, where impinging light induces electron emission. Yet this is only possible up to 50°C, since then Cs losses become too excessive. Typical compounds used are K_2CsSb or AgCsO, RbCsSb; in general, multiantimonide-compounds, as given in a review by P. Coates [51]. Low work functions have also been found for conductive Cs suboxides and Cs-Rb binary suboxides, but have not been used so far for any applications.

10.2 Non-Thermionic Electron Sources

10.2.1 Historic Development

Field Emitters

R.W. Wood in 1897 first described field emission as "fireworks" originating from sharp electrode points in his discharge tube. In 1922 J.E. Lilienfeld for the first time

used a pointed cold cathode in an X-ray tube, drawing erratically several mA. He also observed a further activation of the cathode by Cs. In the following years, also with improved vacuum, the field emission (FE) became more stable. It could be shown experimentally that the fields necessary for FE are a factor of hundred lower than predicted by Schottky in 1923 [52]. In 1928 R.H. Fowler and L. Nordheim published their paper with a theoretical description of FE based on electron tunnelling [53]. In 1936 E.W. Müller introduced the field emission microscope using, e.g. a tungsten point source and a phosphor screen [54]. He showed that the emission originated from submicron protrusions on the tip. Later thermo-field emitters were investigated, and recently diamond emitters and carbon nanotubes (CNT) are used for field emission sources in vacuum microelectronic devices and displays.

From Vacuum Microelectronics to Field Emitter Displays

Shoulders proposed 1956 in [55], based on the knowledge how to build manually machined and assembled tubes which could be operated at frequencies as high as 4 GHz [56], that employing the microfabrication technology developed for on-chip large scale integration of micron-sized solid-state devices will allow a revolution to obtain much faster vacuum microelectronic devices having dimensions which are three orders of magnitude smaller. This was thought possible using field emitters as the source of electrons instead of the commonly used thermionic emission from hot filaments. Based on the theory and experiments of field electron emission [57], at Stanford Research International, CA, USA, Shoulders, Brodie and Spindt developed field emission cathodes made from molybdenum [58] and published an extensive review on their development work [59]. Arrays of cathodes with 10 Mio/cm^2 could be produced, which levelled the noise of the single emitter statistically and showed life times in UHV environment exceeding 10 000 hours. These arrays became attractive to develop flat panel displays [60] and cathodes for high frequency vacuum tube devices [61]. More applications are described in a review by Brodie and Spindt [62], see also Table 10.2.

10.2.2 Field Emission Electron Cathodes

Electrons are released from a material surface or from an atom using very different mechanisms. Figure 10.19 gives schematic the mechanism for various electron emission sources. The Brightness, which is the current per unit area emitted into a steradian, is also given. Field emitters represent the class with the highest brightness of all electron sources.

Field emission appears, if at an metallic surface an electric field (E) of the strength of 10^9 V/m is applied. Such field strengths are generated by low voltages at existing sharp edges and very fine tips $E = U/d \sim U/5r$ with d the electrode distance and r the tip radius. A common form of field emitter is an electrolytically etched tungsten wire of 150 μm diameter, which renders a fine tip of 100 nm radius. To obtain this, the wire must be dipped a defined length of, e.g. 2 mm into the NaOH base. When monitoring the resistance of the etch cell, a sharp rise of the resistance

Table 10.2. History: milestones in the development of FED's

1958	Buck and Shoulders (SRI): Proposal of vacuum microelectronic devices using field emitters
1968	Spindt (SRI): Development and production of molybdenum-based field emission cathodes with integrated control grid
1981	Henry Gray (NRL): Lateral Vacuum Triode with Silicon Field Emitters
1985	Robert Meyer: Demonstration of a matrix-addressable monochrome FED
1988	Meyer, Baptist (LETI): develop Mo-emitters on a resistive layer for FED
1993	Pixtech: demonstrates a colour FED; Micron: demonstrates a 1″ colour FED with Si emitters
1993	Kumar (MCC): demonstrates a FED with Carbon Diamond film emitters
1995	Pixtech/Futaba: demonstrate a low voltage FED
1997	Motorola and Candescent-demonstrate a VGA high-voltage colour FED with Mo Spindt-type Emitters
2003	Samsung: demonstrates a 38″ HDTV colour TV with carbon nanotube (CNT) paste emitters

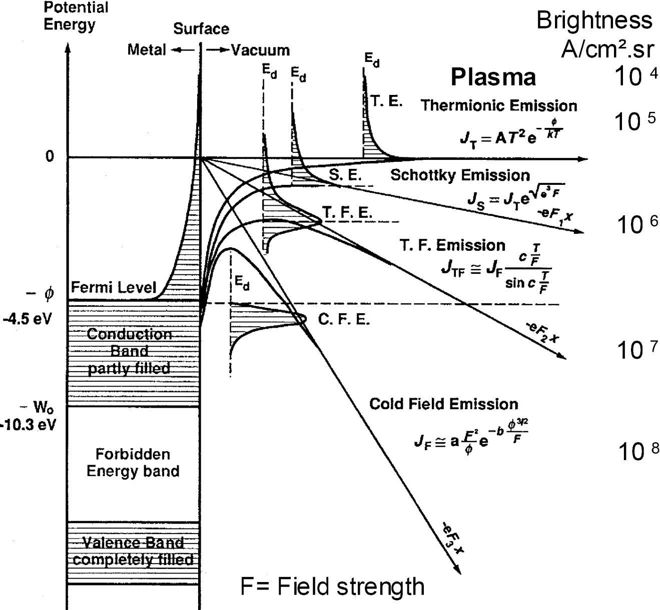


Fig. 10.19. Electron emission mechanisms and brightness of electron sources at 10 kV

will be observed, when the immersed piece of the wire is dropped off due to gravity, and the tip is formed. At this moment it is necessary to cut the voltage off, otherwise enhanced etching will make the sharp tip round. Radii observed are 100 nm with electronic switch-off.

The current density emitted from such tips and edges can reach very high values (e.g. 10^8 A/cm^2). However, the emission currents are generally small (μA, due to the

fact that metals segregate at current densities >250 kA/cm^2 (aluminium)). (Segregation is the disruption of conductors due to ion and atom migration in fields.) In addition, at such current densities an extreme heating of the emitter tip appears and vaporizes this.

10.2.3 Thermo-field Emission Electron Cathodes

To release electrons from a metal surface, the extracting field E_x must be sufficiently strong that electrons can tunnel with a high probability through the potential wall in front of the tip. This wall is the thinnest one close to the tip. At a temperature of 0 K electrons fill the bands in the metal up to the Fermi energy E_F. To have them tunnel from there, a very high field strength of >0.4 V/nm is required. Electrons can however reach energy levels above the Fermi level, by operating the field emitter at elevated temperatures.

Thermionically emitted electrons are not monochromatic, but have a energy distribution with a full width at half-maximum of

$$\Delta E[\mathrm{eV}] = 2.45T[\mathrm{K}]/11\,600. \tag{10.15}$$

Field emission at elevated temperatures is called thermo-field (T-F) emission or Schottky emission. The current obtainable is given by [63]

$$J_\mathrm{es} = J_s(\pi q)/(\sin(\pi q)), \tag{10.16}$$

where

$$q = h^{1/4}F^{3/4}/\pi(2m)^{1/2}kT = \hbar e^{1/4}F^{3/4}/\pi(2m)^{1/2}kT \tag{10.17}$$

and

$$J_\mathrm{s} = 120T^2 \exp\left\{\left(-\Phi + (e^3F)^{1/2}\right)/kT\right\}[\mathrm{A/cm}^2], \tag{10.18}$$

with T the absolute temperature in Kelvin, Φ the work function and F the electric field.

Figure 10.20 shows the potential distribution and the electron distribution in metal at elevated temperature. Since electrons at $T = 0$ fill the metal potential well up to the Fermi level, a high field is required to release them by tunnelling through the potential into the vacuum. Since in a heated metal the electrons above the Fermi level have a Maxwellian energy distribution, they can tunnel at lower field strength and deliver an emission current in thermo-field emission [64].

To increase the emission current, not only heating of a metal tip is used, like with a W <100> thermal field build up tip, but also adsorbed layers are supplied to lower the work function of the material. It has been shown that zirconium, having a work function of 3.9 eV, adsorbs on W <100> planes selectively, by an intermediate layer of oxygen adsorbed to the W-tip. This is schematically shown in Fig. 10.21.

This adsorbed double layer also lowers the work function from 4.6 eV for W down to 2.7 eV for Zr-O-W tip coatings. This reduces the required field strength and extraction voltage, and confines the emission site to the area of low work-function,

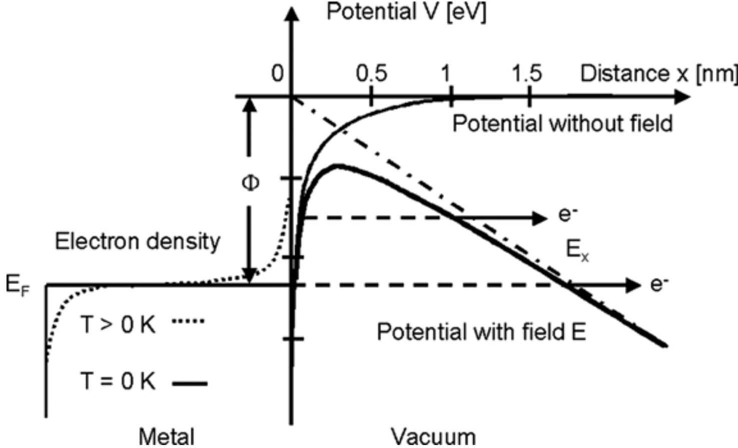


Fig. 10.20. Potential distribution and the electron distribution in metal at elevated temperature, and potentials for tunnelling of electron in T-F emission

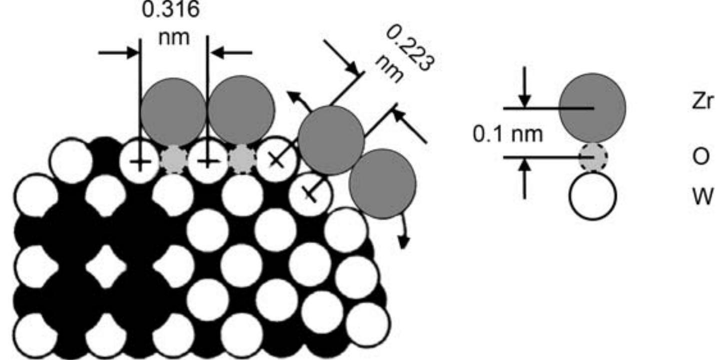


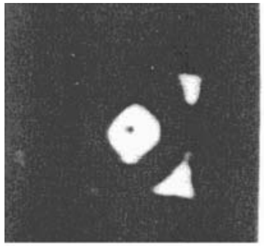
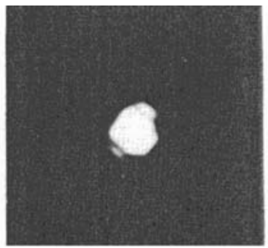
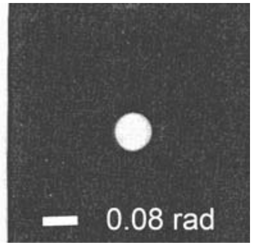
Fig. 10.21. Schematic representation how zirconium adsorbs on W <100> planes of the W-tip selectively by an intermediate layer of oxygen

see Fig. 10.22 [65]. The reduction of emission sites at the cathode renders in a reduced emission current and gives a better beam stability. In addition, it reduces the energy broadening of the emitted electrons due to Coulomb interaction effects.

10.2.4 Cold Field Electron Emission Cathodes

For applications in microscopy and materials analysis the brightest possible cathodes are required. Those are found with cold field emission cathodes. Very fine etched metal tips emit up to 10^8 A/cm^2 current in narrow emission angles. The field emission current density (J_{F0}) at room temperature (300 K) is given by

$$J_{F0} = C_1 E^2 \exp(-C_2/E), \tag{10.19}$$

Zr-O-W <100> reactivation of Zr-0- monolayer at 1850 ± 50 °K

Fig. 10.22. Field emission microscope observation of emission confinement by absorption of Zr on O on a W field emitter tip

where E is the electric field strength and C_1, C_2 are material constants depending on the work function. At higher temperatures (<1500 K) the emission current density is given by the analytic expression

$$J_{\mathrm{FT}} = J_{\mathrm{F0}}(TR/\sin \pi R) \tag{10.20}$$

with $R = 9.3 \times 10^5 \Phi^{1/2} T/E$, Φ the work function (eV) and T the temperature [K].

The Fowler–Nordheim equation describes the current density obtainable in field emission, taking the image force and geometry into account (see Fig. 10.23 top), by

$$J = (AE^2)/(\Phi t^2(y)) \exp((B\Phi^{3/2} v(y))/E). \tag{10.21}$$

Here, Φ is the work function, E is the field strength, A, and B are factors, $y(\Phi, E)$ is a function of field strength and work function, $y(\Phi, E) = 3.79 \times 10^{-4} \Phi^{1/2}/F$, with $v(y) \approx 0.95 - y^2$, and $t^2 \approx 1.1$ representing the influence of the image force.

For better evaluation of the emission characteristics a logarithmic display of the current density divided by the square of extraction voltage over the reciprocal value of the extraction voltage, generally called the Fowler–Nordheim plot is used (see Fig. 10.23 bottom)

$$\log_{10}(J/U^2) \sim 1/U. \tag{10.22}$$

The extrapolated ordinate value of the FN line is a measure for the obtained reduced current density. The inclination of the FN line corresponds to the work function and the field in front of the emission site.

This field strength is difficult to measure and to determine. Therefore, these data only give relative values. The field strength at the tip depends also on the materials structure, or structure composition, and on the semiconductor characteristics of the material. Fowler–Nordheim equation and logarithmic representation are used to characterize field emission sources. Characteristic data are taken from the inclination and the ordinate section value and are used in SG-plots, showing these data in a graph for different sources.

The major types of field emission cathodes and emitter cells for arrayed field emission cathodes FEA-cells like point emitter (Spindt-emitters (Mo) or Gray-emitters (Si)), wedge emitter, and thin film edge emitter are used. Figure 10.24 shows Spindt-type Mo-emitters.

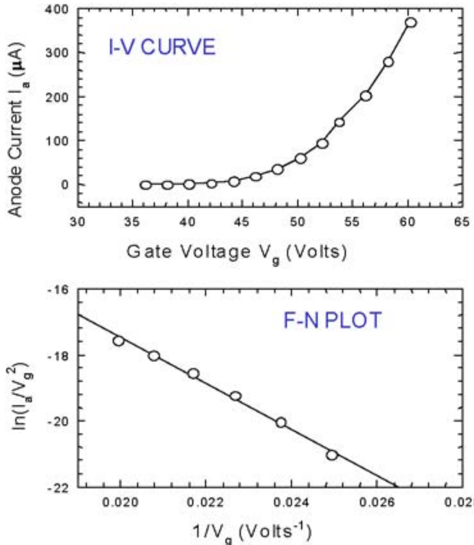


Fig. 10.23. *Top*: current–voltage curve shows the exponential emission characteristic; *bottom*: the logarithmic representation according to Fowler–Nordheim gives the straight line, which allows to judge on the performance

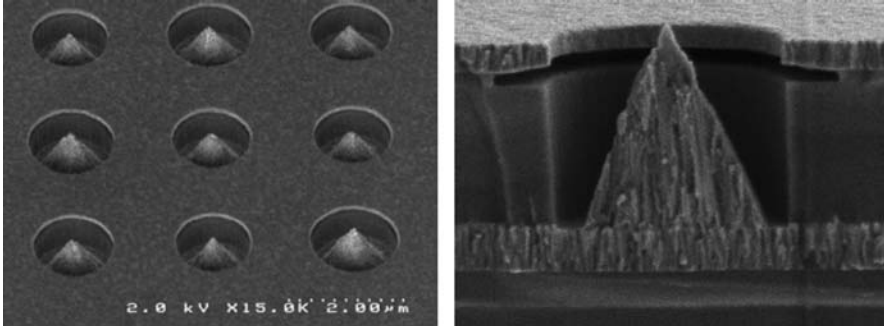


Fig. 10.24. Mo-microtip array of Spindt type (courtesy of C. Spindt SRI). The emission strongly depends on tip height, radius of curvature, position in the gate opening

10.2.5 Novel Cathode Materials

Several novel materials are developed and applied in technical electron sources and FED's, and are given in Sect. 10.2.5, Table 10.3. Novel cathode materials are molybdenum, silicon, diamond, carbon in corraline carbon, carbon nanotubes and semiconductor particles suspended in an oxide matrix (not shown). Also novel materials produced by electron beam induced deposition (EBID) with a nanocrystalline composition are used, e.g. Au/C and Pt/C nanocomposits, see Table 10.3.

Table 10.3. Table of typical performance data of field emission cathodes

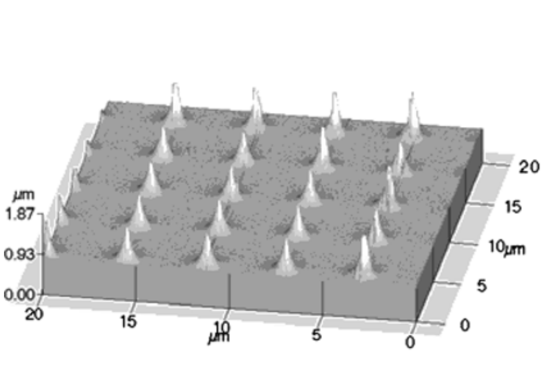
Electron sources	Year of market entry	Brightness A/cm^2 sr V	Efficiency	Maximum current
Gas discharge	1930	1	0.05	100 µA
Thermal emitter W	1932	10	0.35	200 µA–15 mA
BaO	1940	100	0.35	50 µA
LaB_6	1960			
Thermofield Zr-O-W	1970	1000	0.40	100 nA
Cold FE	1964	10 000	0.95	5 µA
Spindt-Mo	1970			500 µA
Gray-Si	1985			50 µA
Diamond	1995			200 µA
Carbon nanotubes	2003	10 000	0.95	50 µA
EBID Au/C Pt/C	1992	10 000	0.95	1 mA

10.2.6 Trends of Novel Developments

Field emitter arrays (FEAs) are one of the most important electron sources for vacuum microelectronics. One of the issues arising from FEA processing is incomplete homogeneity in tip shape over the array. Such inhomogeneity gives rise to different emission behaviour among FEAs. Tip surface modifications have been performed for the purpose of obtaining a sharp tip and a clean surface using ion milling [66], reactive ion etching [67] and plasma etching [68], see Fig 10.24. The modification of Si tip surfaces by anodization [69] has been used to improve the emission behaviour of Si FEAs, see Fig 10.25. Tip shape modification using a focused ion beam (FIB) has been performed in order to obtain a sharp tip [70].

Electron-beam-induced deposition (EBID) is caused by the dissociation of adsorbed molecules induced by highly energetic electrons [71]. The dissociation process has not been clarified yet due to the variety of potential excitation channels available even for small molecules. The EBID process facilitates a maskless process down to 5 nm depending on electron beam spot size and reactions. EBID can yield three-dimensional structures on any kind of surface as long as the precursor molecules adsorb [72]. A field emission tip achieved by EBID has been reported, though the structure was that of the two-tip arrangement [73], see Fig. 10.26. Pt-FEA tips with a gate electrode were fabricated by EBID emitter generation in Spindt-type cells having a blunt emitter tip, which showed no emission. The emission characteristics were measured before and after Pt deposition. Nanometer-sized field emitters have also been fabricated using FIB-etching and EBID [74].

Smart FEA structures incorporate a current stabilizing MOSFET transistor and are fabricated in MOSFET Technology, see Fig. 10.27. Transistor-controlled cells with single, dual and triple gate have been developed and employed to stabilize the emission. This leads to a stable intensity [75]. Characteristics obtained were: at gate diameters of 3 µm the onset voltage was 75 V, the maximum current was 10 µA at



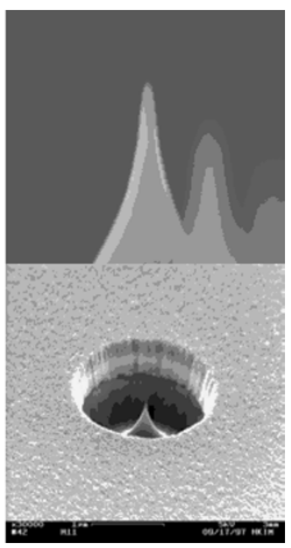


Fig. 10.25. Silicon field emitter arrays (courtesy of T. Akinwande, MIT (1997))

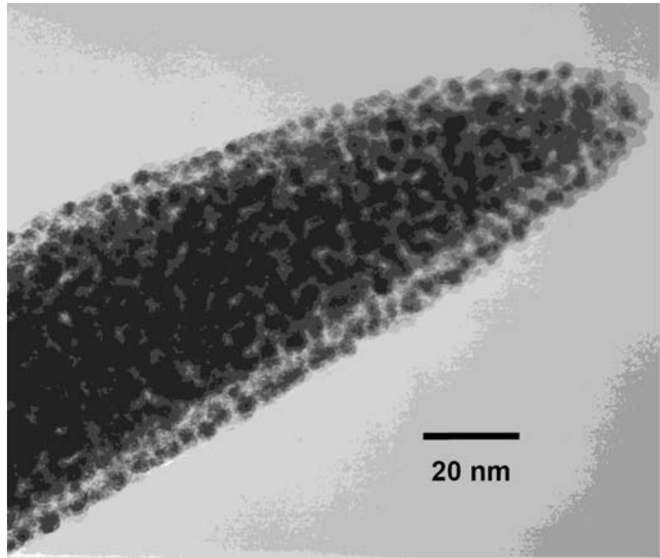


Fig. 10.26. Top of a tip deposited from dimethyl-gold-trifluoro-acetylacetonate (Me2-Au-tfac) at 300 pA and 25 kV shows nanocrystalline composition imaged in a 200 kV TEM

100 V gate voltage and the current stability over 24 hours in a vacuum of 10^{-8} Pa was 25% in a 10×10 tip array.

Candescent emitters produced with ion beam technology uses 10^8 to 10^9 ions/cm^2 (random pattern) multiply ionized Xe or Ar at 2 to 13 MeV energy, see Fig. 10.28.

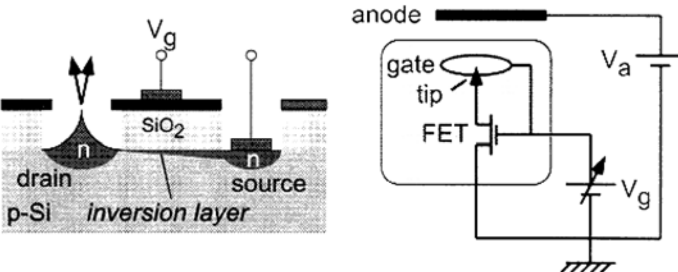


Fig. 10.27. Smart FEA structures incorporate a current stabilizing MOSFET transistor and are fabricated in MOSFET Technology

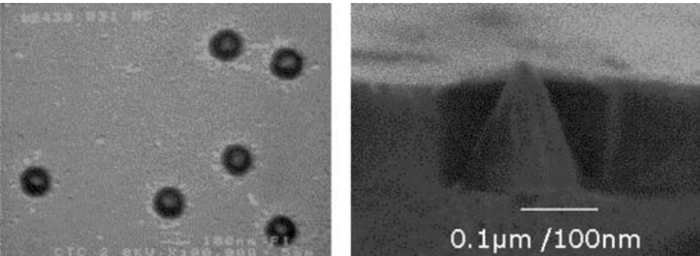


Fig. 10.28. Candescent FED emitters are produced by ion beam irradiation to avoid patterning of extractor holes by electron beam lithography or other methods, like optical holography or ball lens exposure. *Left*: 100 nm scale gate holes. *Right*: 150 nm size emitter cones

A 1 cm^2 diameter beam is rastered over the active area of the display, with a uniformity of $<5\%$. 320×340 mm panels are irradiated at the Lawrence Livermore National Lab (LLNL); 60 seconds are needed to iradiate the panel. The successful technology was stopped by financial considerations within the US-Asia powerplay.

10.2.7 Metal–Insulator–Metal (MIM) Emitter Materials

PFE's screen printable cold cathode contains conducting or semi-conducting particles of ≈ 3 μm in size suspended in an inorganic insulating matrix, generating a metal–insulator–metal–insulator–vacuum (MIMIV) device, which emits electrons by tunnelling, see Fig. 10.29.

Screen printed emitters are used in a 32×32 device, as shown below, and have higher particle and emission site density (300 k sites per cm^2), flatter surface, are easier to fabricate, and render $\approx 30\%$ higher emission uniformity. They require 50 volts swing drive electronics. To make FED's economically competitive, it is inevitable to lower the gate switching voltage to 25 V or below.

10.2.8 Diamond Hopping Electron Emitters

Diamond-like layers have insulating diamond grains in sp3 carbon configuration and grain boundaries in sp2 carbon configuration. The latter are conducting and deliver

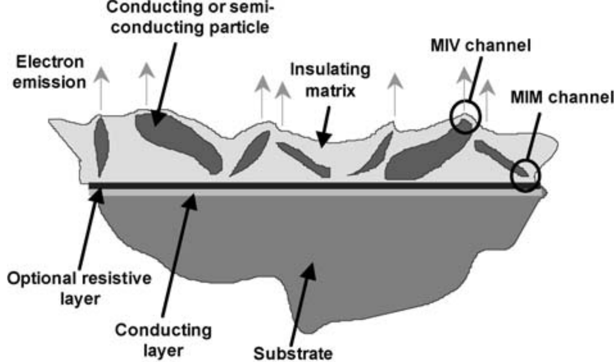


Fig. 10.29. Printable inks made from semiconductor material embedded in insulator matrix present metal insulator vacuum channels for field electron emission after firing (courtesy of PFE Inc.)

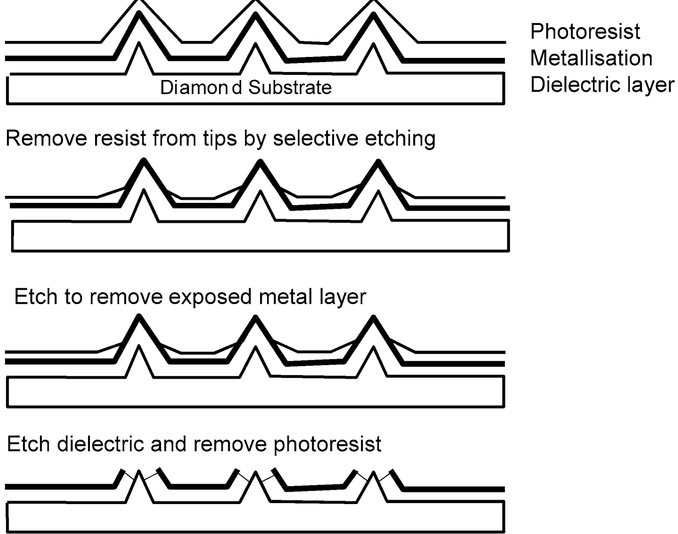
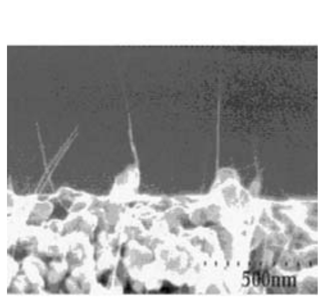


Fig. 10.30. Schematic of the fabrication method for gated diamond emitters

current by hopping. Typically 5–20 V/μm field strength is applied to a diode configuration to characterize such emitters.

Extreme Devices did fabricate CVD gated diamond microtip arrays for high power, high temperature and high radiation environments, see Fig. 10.30. These diamond cathodes shall replace thermionic cathodes in: microwave amplifiers, space applications, sensors, TV-tubes. The gate is self-aligned. Electron-extraction field: 3 V/μm, high-current-density emission: 10 A/cm^2, low onset voltage 22 V. Stable emission over long time, also in high vacuum (courtesy of Extreme Devices Inc.)



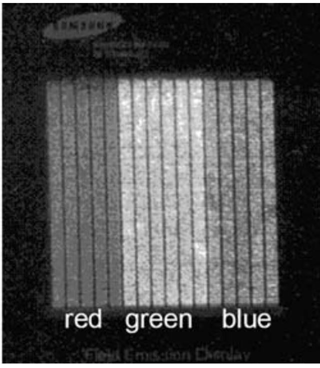


Fig. 10.31. Display with carbon nanotubes from Samsung (courtesy of Choi et al., SAIT Samsung, Korea)

However, flat diamond layers render measured maximum emission current densities for N doped, P doped and B doped diamond all below 4×10^{-5} A/cm^2 [76]. Increased emission is obtained with B doped diamond tips with increased sp^2 content. Emission increases from 1 μA to 20 μA per tip. The turn-on field is 9 V/μm [77].

10.2.9 Carbon Nanotube Emitters

Carbon nanotubes (CNT) grow in gas discharges from butene and other low molecular weight carbon gases and form fullerene-type hoses, which are single to multiwalled and conduct current. Seeding a surface with Fe or Co through lithographicaly generated holes and lift-off allows to select the location of growth of single CNT's. It is possible to fabricate distributed arrays with 5–10 nanotubes/μm^2. They render a bright FED display at 0.2 mm cathode–anode distance and 600 V anode voltage. The maximum current density is 0.1 A/cm^2 [78], see Fig. 10.31.

A much cheaper method to distribute the CNT's than lithography is ink jet printing or sieve printing using CNT filled polymers and optical lithography. Figure 10.32 shows electron emitter sites generated by such a technique. HDTV images with 5″ full colour and moving picture (320 × 240× R, G, B) is shown in Fig. 10.33. Specifications of the Samsung 38″ CNT FED are: screen size 38″ diagonal, pixel number 1280 × RGB × 768, brightness 180–360 Cd/m^2 (F.W.), dark contrast 530:1, power consumption 90 W. Samsung 38″ CNT FED was presented at IVMC 2003 [79], see Fig. 10.33.

10.2.10 Other Carbon Emitters and Future Development Trends

The gate voltage amplifier electronics makes PDP displays expensive. With CVD carbon nanotube gated cathodes the gate voltage is reduced. Emission current density: up to 8 mA/cm^2, Life time at $>50\%$ current $>10\,000$ h, cathode material with binder-CNT paste with photolithography structured. This fabrication method allows

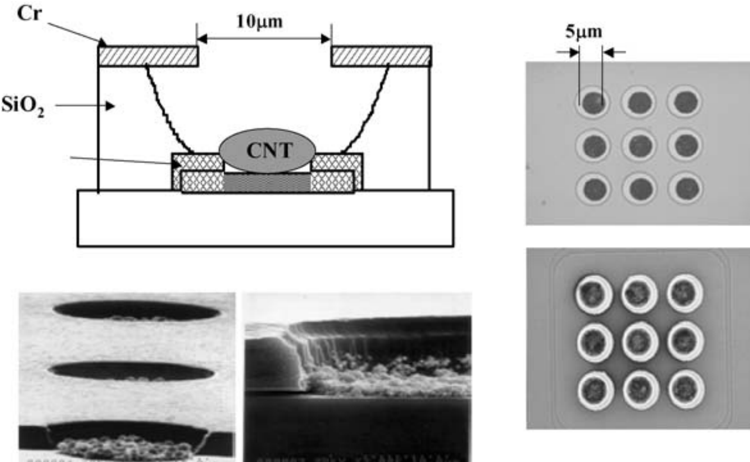


Fig. 10.32. Samsung carbon nanotube FED with a normal thin film gate triode structure. This is possible due to the application of a photosensitive CNT containing cathode paste applied with screen printing methods and UV curing for the definition of the cathode (courtesy of J.M. Kim, SAIT Samsung)

Fig. 10.33. HDTV images with 5″ full colour and moving picture (320 × 240× R, G, B) (courtesy J.M. Kim, SAIT Samsung)

to produce 40″ displays at low cost development at Motorola; gate–anode distance 0.7 mm, anode voltage 3.8 kV, packaged successfully. The onset voltage was 25 V. A major challenge to the display industry is to find low-cost carbon nanotube FED emitters using inexpensive driver electronics [80]. Future possible commercialization will take place in the fields of small displays as replacements of LDC small screens and VGA computer displays. The big FED 30″–50″ display for home TV is a possible major driver for the development. Lamps as picture tube elements, and

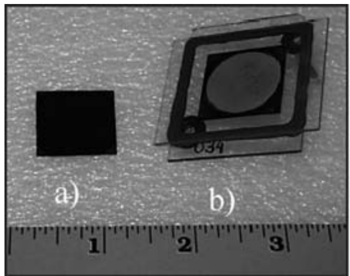


Fig. 10.34. Flat diode lamps, *left*: **a** 1″ × 1″ nanotube cathode on silicon substrate, **b** flat panel diode package. Cathode–anode gap is 0.2 mm. Right: blue and red pictured light sources under test. Anode voltage 390 V, brightness 400 cd/m^2 (courtesy of C. Hunt, UC Davis, CA) [81]

also lamps for room lighting are developed and will be mass-produced in the near future.

Flat diode lamps are investigated at UC Davis, CA, in collaboration with Sandia National Laboratories Livermore, CA, and are shown in Fig. 10.34. Table 10.4 gives a comparision of FED with other flat panel displays.

10.3 Other Electron Emitters

10.3.1 pn-emitters

Avalanche Cold Cathodes

Apart from improved thermionic cathodes also low drive cold cathodes are developed. The principle of operation is the avalanche cold cathode [83, 84]. The cold emitter based on silicon IC technology consists of a shallow p$^+$n^{++}-junction, where the n^{++}-layer has a thickness of 10–20 nm (Fig. 10.35). The internal field of the pn-junction heats the electrons to a much larger temperature than the lattice temperature, and with a Cs covered surface several percent of the diode current are emitted into vacuum at current densities up to 8000 A/cm^2 over an area of about 1 μm lateral extension. One of ACCs main advantages over thermionic cathodes in CRTs is that only a few volts of video-drive are necessary, which leads to a power reduction of the set. Also, excellent picture quality has been demonstrated. Yet sustaining continuous Cs coverage on the ACC surface has led to new vacuum-technological concepts like the addition of a Cs-Au compound thin film for Cs resupply.

As a cathode with low operation voltage for the vacuum microelectronic devices, a new silicon tip avalanche cathode (STAC); which is a silicon field emitter with a shallow n1–p1 junction formed on the tip, is studied [85]. The measured emission characteristics of STAC show that the gate voltage at which an appreciable emission current begins to flow is lowered, and the gate voltage required to produce a desired emission current is reduced. These emission characteristics of STAC result from the hot electrons generated and accelerated by the avalanche breakdown at the n1–p1 junction on the tip, see Fig. 10.36, and Fig. 10.37.

Table 10.4. Comparison of FED with other flat panel displays [82]

Feature	Thin film transistor LCD	Electro-luminescent display	FED	Plasma display panels	OLED display
Brightness (cd/m^2)	200	100	150 (low-V) >600 (high-V)	300	300
Viewing angle (degrees)	±40	±80	±80	±80	±80
Emission efficacy (lm/W)	3–4	0.5–2	1.5–3 (low-V) 10–15 (high-V)	1.0	10–15
Response time (ms)	30–60	<1	0.01–0.03	1–10	<0.001
Contrast ratio (intrinsic)	>100:1	50:1	300:1	100:1	100:1
Number of colors (Mio)	16	16	16	16	16
Number of pixels	1024 × 768	640 × 480	800 × 600	852 × 480	640 × 480
Resolution mm in pitch	0.31	0.31	0.27	1.08	0.012
Power consumption (W),	3	6	2	200	6
Size (cm)	25.4	25.4	25.4	106.7	15.2
Maximum screen size in diagonal (cm)	55.9 (22″)	25.4 (10″)	35.6 (14″)	106.7 (42″)	15.2 (6″)
Panel thickness	8	10	10	75–100	3
Operating temperature range (°C)	0–50	−5–±85	−5–+85	−20–±55	−25–±65

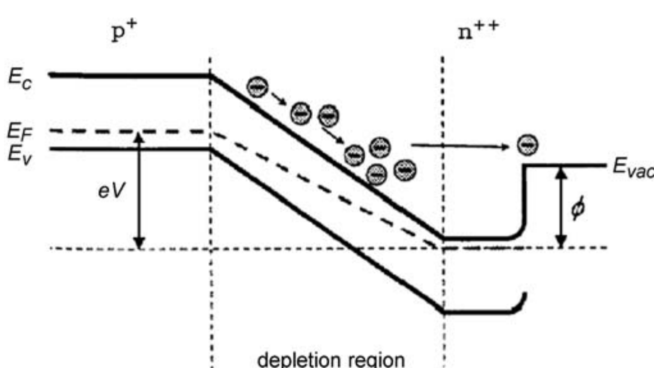


Fig. 10.35. Electron energy band diagram of an avalanche cold cathode driven in reverse bias. Electrons generated in the depletion layer are accelerated towards the surface

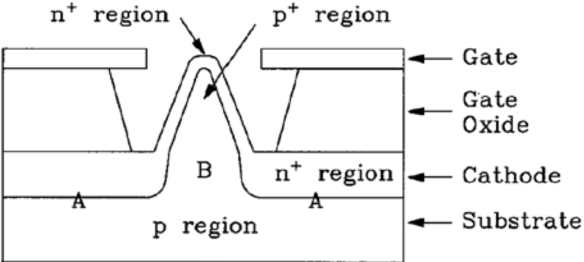


Fig. 10.36. Practical design of an ACC-field emission source (after Lee [85]), a shallow n-p junction is formed at the tip by lithographic steps and ion implantation to $10^{15}/\mathrm{cm}^2$

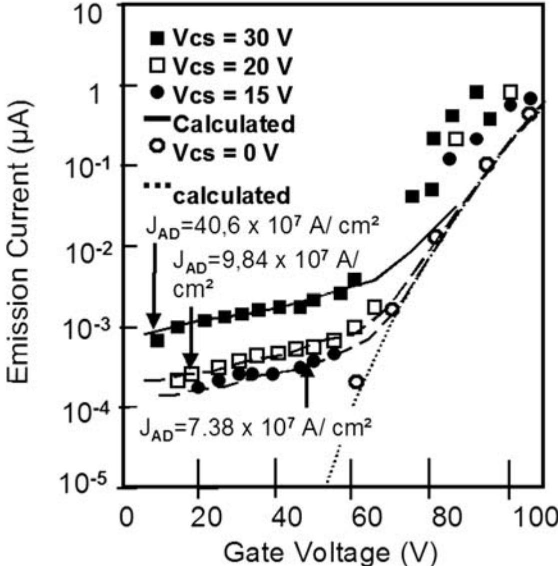


Fig. 10.37. Comparison between the calculated emission current curves (*solid line*) for $J_{\mathrm{AB}} =$ 7.38, 9.84, and 40.63 × 10^7 A/cm^2 and the measured emission current data for $V_{\mathrm{CS}} = 15$, 20, and 30 V. Also, the calculated emission current curve (*dashed line*) and the measured emission current data for $V_{\mathrm{CS}} = 50$ V are shown [86]

The low gate voltage characteristics of STAC is well explained by a simple analysis which takes into consideration the emission of hot electrons. However, the measured emission current of STAC in the high gate voltage region is not well resolved by the analysis. Also a large current is needed for the avalanche breakdown, and this results in low emission efficiency and large power consumption.

Ballistic Electron Emitters

The mechanism of ballistic electron emission from nanocrystalline silicon diodes is obtained from nanocrystallized polysilicon (NPS) based devices. The electron

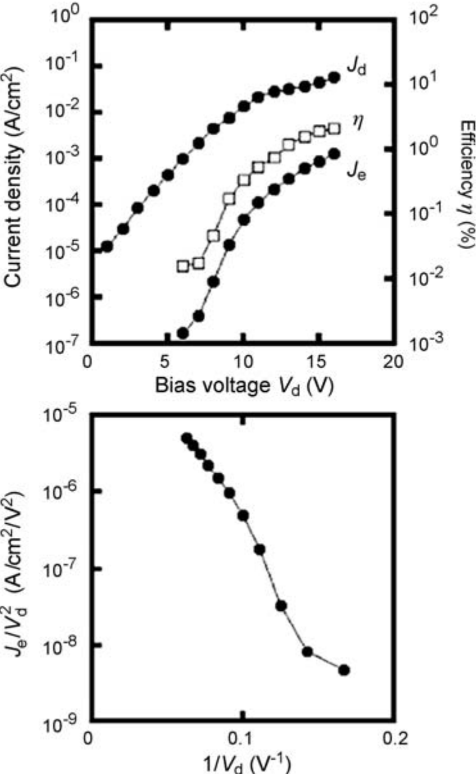


Fig. 10.38. The emission current curve and the diode current curve. Both determine the efficiency h, which is as low as 2.8%

emission characteristics performs due to of chainlike nanocrystalline silicon ∼nc-Si structure, which is produced along columnar poly-Si grains. A sufficient electron emission current density of 3.0 mA/cm^2 was observed with a high emission efficiency ≈2.8% and stability, see Fig. 10.38. The surface and interfacial oxidation of nc-Si particles is an important factor for efficient emission. The results indicate that control of interconnected nc-Si structures is a key issue for the efficient ballistic electron emission [86]. It is found that nc-Si chains interconnected via thin oxides play a key role for efficient electron emission. This supports the ballistic electron transport model based on multiple tunneling in the nc-Si layer [87, 88].

10.3.2 Secondary Emission

For a detailed description see Sect. 3.2.2.

10.3.3 Ferroelectric Electron Emission

Weak electron emission during polarization reversal of ferroelectric material was first reported by Rosenman et al. in 1984 [89]. The spontaneous electrical polarization of ferroelectric materials can be changed either by reversal or phase transition between ferroelectric and non-ferroelectric state. If spontaneous polarization changes are induced by fast temperature rise, by a mechanical pressure pulse, by a laser pulse or a high voltage pulse applied to the sample, strong uncompensated surface charge densities and related polarization fields are generated, which may lead to intense self-emission of electrons from negatively charged surface areas. Thus, in electron guns the extraction-field free emission can be separated from high field regions of accelerating gap or gate structures. In 1988 at CERN strong ferroelectric pulsed electron emission of up to 100 A/cm^2 was achieved. Inherently dc-emission is not possible with this principle. The first application was the use of ferroelectric emission as a trigger to ignite the discharges of high power gas switches. Another application was the alternating gradient space charge focussing of low energy heavy-ion beams, i.e. for an accelerator technology.

10.3.4 Photo-electron Emission

Photoelectric emission is the emission of electrons caused by the irradiation of matter by light. Fundamentals are described in Sect. 3.2.1. It is investigated in dependence of the existence and properties of solid surfaces. Photoemission occurs in the fundamental optical range. Thus, it is related to other optical experiments and is used for solid state spectroscopy, i.e. the determination of band structure. In theories of electron emission one generally assumes that all electrons whose energy associated with the movement normal to the surface is larger than the potential in vacuo can escape. Thus, the probability of escape P is taken to be

$$P = 1 \quad \text{if } E - n^2 k||/2m > U_{\text{vac}}, \tag{10.23a}$$

$$P = 0 \quad \text{otherwise}, \tag{10.23b}$$

where $k||$ is the wave vector parallel to the surface, m is the free electron mass, and U_{vac} is the asymptotic value of the potential in vacuo. In reality, electrons satisfying (10.23a) are partially reflected because of diffraction by the periodic potential of the solid. It is not generally possible to match a Bloch wave in the solid with a single plane wave in vacuo without the existence of reflected Bloch waves. Considerable reflection occurs at low electron energies and therefore $P(E)$ exhibits a pronounced structure. Since diffraction is highly directional, its effect on energy distributions of photoelectrons will be larger in measurements accepting electrons in a small solid angle of directions. In hemispherical retarding field analysers diffraction effects will be approximately averaged out.

The energy diagram of the surface of a semiconductor shows Fig. 10.39. It is characterized by a space charge near the surface that is compensated by a surface charge, S, composed of electrons occupying surface states. These charges produce a

macroscopic potential $V(x)$. Electrons are retained in the material by the potential $U(x)$ which has the shape of the image force at large distances in vacuo. It is characterized by the work function Φ, the electron affinity χ and the ionization energy ξ, as shown in Fig. 10.39. All these quantities can be measured by photoemission. The relationship between $\mathrm{d}V/\mathrm{d}x$ at $x = 0$ and $(E_F - E_V)_s$ is particularly interesting, because it yields information on surface states. The surface of a metal is much simpler. It is fully defined by the work function $\Phi = U_{vac} - E_F$.

Band bending near the surface of semiconductors is attributed to charge carriers residing in surface states. In principle, these electrons can be excited by the incoming light and contribute to photoemission. Their observation is facilitated in n-type semiconductors by the fact that electrons in surface states have higher energies than valence electrons. They are isolated at the high-energy end of energy distributions. Experiments with cesiated InAs have shown that, indeed, emission from a few as 0.01 electrons per surface atom could be observed [90]. The Schottky effect in photo emission is the reduction in work function and resulting increase in emission caused by the application of an electric field normal to the surface. This increase in current is a monotonic function of the applied field.

The vectorial photoeffect intrigued early investigators of photo emission and has been relatively neglected in the more recent years. When linearly polarized light falls on a surface at an oblique angle and the components of the electric field parallel and normal to the surface transmitted into the solid are calculated according to Fresnel's equations, it is found that the component of the electric field normal to the surface is much more efficient in producing photoemission than the tangential component. In thin films of simple metals, the enhanced photoemission for E_s is accompanied by an equal enhancement of optical absorption. Both can be accounted for by excitation

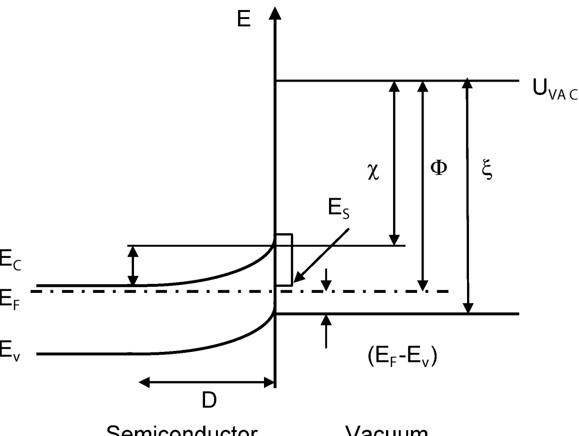


Fig. 10.39. Energy diagram of a semiconductor surface; V_B is the band bending, E_F is the Fermi level, E_c and E_V are the edges of conduction and valence band, E_S is surface states, χ is the electron affinity, Φ is the work function, ξ is the ionization potential and U_{Vac} is the asymptotic value of image force potential

Table 10.5. Characteristics of standard photocathodes in use

Photo cathode	Sensitivity efficiency μA/lm	Dark current density A/cm^2	Quantum detection efficiency	Wavelength nm
Ag-O-Cs	30 (75)	$10^{-13}..10^{-11}$	1	1200
Alkaline-antimony	50–250	$10^{-18}..10^{-14}$		600–870
Multi-alc-antimony	250–600	10^{-15}	20–25	950
NEA	500–2000	10^{-15}	10–50	1200

of plasma oscillations by the light. The vectorial photoeffect decreases drastically as the film thickness increases beyond 100 nm.

Angular distribution of emitted electrons yields considerable information on the band structure of crystals from which the electrons were emitted without appreciable scattering. It was shown for clean silicon and germanium surfaces, that electrons are preferentially emitted in directions that depend on the crystallographic orientation of the plane of polarization of normally incident, linearly polarized light. This effect shows the presence of a substantial fraction of electrons that were excited by direct transitions and emitted without scattering. These measurements do not rule out the presence of some scattered electrons.

It was observed from the investigation of the angular distribution of electrons emitted from a GaAs crystal, whose surface was covered with Cs and CsO to produce a negative electron affinity, that the electrons were emitted preferentially in a very narrow cone normal to the surface. The angular distribution was found to be Gaussian, with a half-width that depends on wavelength of UV-light and surface treatment but is of the order of 10 degrees. The explanation of this effect is that the majority of the electrons that have been excited at great depth ($\approx$1 μm) have lost their kinetic energy and are "thermalized" in the bottom of the conduction band ($k = 0$). They are accelerated in a direction normal to the surface by the space charge field and emitted after negligible scattering at the surface. For more details and literature, see Fischer 1972 [91]. Table 10.5 summarizes the characteristics of standard photo cathodes in common use.

Inclusion of getter materials in a cavity allows to operate handheld vacuum photosensitive electronic devices on a chip. Getter materials are used to disperse alkali coatings of surfaces for photo-sensible structures. Alkali metal dispensers (AMDs) are compact sources of alkali metals. They may be used to prepare photosensitive surfaces of photo cathodes or wherever an ultrapure alkali metal film is needed. The alkali metal generating source is a mixture of an alkali metal chromate with a reducing agent. A research program at SAES Getters S.p.A. into methods of controlling the release of pure alkali metals led to the development of SAES AMDs. SAES alkali metal dispensers are especially suitable when very pure alkali metal films are required, evaporation rate must be controllable and reproducible and loose particles cannot be tolerated. The device must be free of harmful gases during deposition of

the alkali metal. The reducing agent in SAES AMDs is SAES' St101 (Zr 84%-Al 16%) getter alloy. In addition to reducing the alkali back to its metallic state, St101 also removes chemically reactive gases from the device preventing them from contaminating the alkali metal vapor. SAES AMDs are available to deposit Cs, K, Na, Rb and Li. They are available in precut sizes with terminals or in bulk wire. Curved shapes are available [92].

10.4 Vacuum Electronic Construction

10.4.1 Spacers for FEA's

To separate the cathode plate and the screen plate in a FE-display, spacers are required. It is not sufficient to have insulators, but semiconductors or resistive supports are required to assure the proper function of the display. Researchers from Saint-Gobain Display Glass, France, could solve the problem by making a semi-conducting glass. Figure 10.40 shows a cross-spacer which fits in between pixels and does not attenuate the electron beam at his site. Several spacers are to be distributed over the area of a large display at least in 1 to 5 cm distance depending on the thickness of the glass sheets. Earlier solutions were made from ionic conducting glass [93].

10.4.2 Anodic Bonding

Anodic Bonding

The substrates are bonded at elevated temperature (≈400°C) by placing and clamping the substrates between two metal electrodes. A high DC potential (up to >1 kV) is applied between the electrodes creating an electrical field, which penetrates the substrates. One substrate is a glass that contains sodium ions, which at the elevated temperature are displaced from the bonding surface of the glass by the applied electrical field. The depletion of sodium ions near the surface of the glass makes the

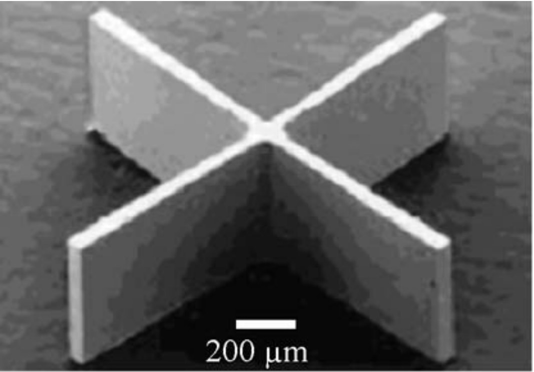


Fig. 10.40. Semiconducting glass cross spacer, which fits in between pixels and does not attenuate the electron beam at his site (curtesy of Saint-Gobain Display Glass, France)

surface highly reactive with the silicon surface of the other substrate forming a solid chemical bond.

Fusion Bonding

The substrates are first forced into intimate contact by applying a high contact force. Once in contact the substrates hold together due to atomic attraction forces (Van der Waals), which are strong enough to allow the bonded substrates to be handled. The substrates are then placed in a furnace and annealed at high temperature, after which a solid bond is formed between the substrates.

Glass Frit Bonding

The alignment tolerance for registration of CAD data to features on the wafer is typically 10 μm. A batch process is possible. The second substrate material may be e.g. Pyrex (Corning 7740) having a substrate diameter of 50 mm and a substrate thickness of 300 to 1000 μm. Bonding takes place at a temperature of 400°C. Silicon wafers may have up to 100 mm wafer size.

Solder Bonding

A gold metal layer is employed on both sides of the parts to be bonded. The alignment tolerance is 5 μm to 200 μm if needed. Alignment is performed manual optical. The substrates are exposed to a vacuum of 0.1 mb during processing. The gold is bonded at a temperature of 350 to 450°C.

Epoxy Bonding

Epoxy bonding is performed in air, with alignment. It is necessary to use Piranha (H_2SO_4/H_2O_2)clean (non-MOS clean) on both surfaces. If there are metal parts on any of the wafers, select an organic NMP clean. Epoxy coat the front to a thickness of 5 μm to 8 μm. Three methods for alignment are possible: optical implies front-front align, backside IR implies front-back align, manual implies no fine alignment.

Bonded materials are preferred pairs of materials bonded by this process glass (Hoya), Pyrex (Corning 7740), silicon, silicon on insulator. The pressure in the process chamber during processing is 1 atm. Wafers of 75 to 100 mm in size can be bonded.

Resist Bonding

This process requires several steps: (1) dehydration bake both surfaces, (2) prime front with HMDS, (3) photoresist coat front with resist material Shipley 1827 of 2.7 μm thickness, (4) resist bond both, (5) perform photoresist hardbake (110°C) both parts. This process is applicable to wafers of 25 mm to 100 mm diameter.

Other Bonding Processes

They are: aluminum microwave bonding, copper microwave bonding, gold microwave bonding, nickel microwave bonding, thermocompression bonding, adhesive bonding, eutectic bonding, low-temperature glass bonding, microwave bonding [94].

10.4.3 Vacuum Generation in On-Chip Vacuum Electronic Devices

Inclusion of getter materials in a cavity allows to operate handheld vacuum electronic devices on a chip. Getter materials chemically bond residual gases and generate vacuum for a long time. SAES Getter supplies a St707 getter alloy for vacuum systems. The St707 wide operating temperature range down to room temperature and its availability in different formats make this alloy the best solution for several applications, from particle accelerators to plasma fusion machines, from laboratory vacuum systems to batteries. It is a getter material composed as a ternary alloy with the following nominal composition: zirconium 70%, vanadium 24.6%, iron 54%. To pump the St707 getter needs activation, since it has a protective passivation layer that must be eliminated to start the gettering action. This process, called activation, is carried out by heating the getter under vacuum or in an inert gas atmosphere. The optimum activation conditions are at 450–500°C for 10 minutes. Once the alloy is activated, reactive molecules such as O_2, H_2O, N_2, CO, CO_2 and H_2 are adsorbed via a three step adsorption mechanism: surface dissociation, surface sorption and bulk diffusion. Hydrocarbons are adsorbed at lower pumping speed at temperatures above 200°C. Once adsorbed, oxygen, nitrogen and carbon atoms cannot be released by the St707 due to the formation of strong chemical bonds with the alloy atoms. Hydrogen reacts differently: it diffuses into the St707 getter bulk even more quickly than the other species and it distributes almost uniformly within the bulk even at low temperatures. However, since the bonds hydrogen-alloy are weak, some of the hydrogen sorbed at low temperatures can be released at higher temperature. The sorption speed for various gases at different temperatures for ST707 getter from SAES Getter shows Fig. 10.41 [94].

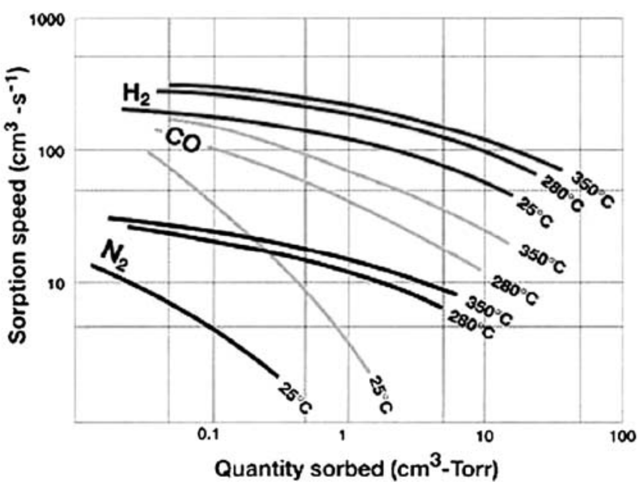


Fig. 10.41. The sorption speed for various gases at different temperatures for ST707 getter from SAES Getter

Nanogetter Packaging

The process is complex and needs the following steps: (1) silicon cap wafer preparation, (2) nanogetter deposition and patterning; the nanogetter deposition is performed using a specially fabricated shadow mask; the nanogetter is deposited by sputter-deposition, (3) fusion bonding with alignment requires a pre-fusion-bond clean of both surfaces to be bonded, a fusion pre-bonding of both surfaces and finally a nitrogen anneal of both parts.

As bonding method can be used anodic bonding (Si to Pyrex), eutectic bonding, fusion bonding (Si to Si direct) or glass-frit bonding. As capping material Kovar Pyrex (Corning 7740) on silicon is recommended. To incorporate the getter or other device, a cavity etch produced by MEMS silicon micromechanical etching is needed [95]. Typically the depth of the etch in the capping wafer is 10 to 550 μm depending on wafer thickness. The thickness of the nanogetter to be deposited must be 5 to 500 nm. A through wafer etch is sometimes performed for electrical connections through the capping substrate.

10.5 Materials of Vacuum Electron Sources

Suitable materials for vacuum technology [96–98] have to meet special requirements with respect to sustaining high pressure differences of up to 15 orders of magnitude between vacuum and outside atmospheric pressure and obtaining constant operating conditions. Dependent on the applications and on the pressure range different groups of materials are used:

1. materials for the vacuum chambers (stainless steel, glass, oxide ceramics);
2. materials for components such as pumps, connecting tubes, valves, flanges and traps (stainless steel, Al, glass);
3. materials for seals (Cu, Au, Fe-Co-Ni alloys, Ag containing solders, Mo, Konstantan, etc.);
4. materials for inner system parts (stainless steel, W, Ta, Mo, Ni, Pt, Ag, Au, Pd, Al, Be, Zr, Ti, diamond, etc.);
5. emitter materials for electron sources (BaO, CaO, SrO, W, Mo, ThO_2, La_2O_3, LaB_6 , Cs, Si, C, etc.).

The choice of a material depends on properties such as high mechanical stability, low gas permeation, low vapour pressure, good heating properties for degassing and good matching of thermal expansion coefficients for different system parts or on special properties such as high electron emission. Further classes of materials are getter materials for pumping and luminescent materials for display screens. A much more detailed description is given in the book of Walter H. Kohl: "Materials and Techniques for Vacuum Devices" [98] and also in the following Chapter 11, "Vacuum Technology". A large part of the materials given above is also used for vacuum electron source construction, and in the following we will specifically stress this application.

Metals

The following table (Table 10.6) gives a review of metals usually applied in vacuum technological applications. In general, a metal consists of many small crystallites. The form and size of these crystallites determine the mechanical properties. Fine

Table 10.6. Materials for vacuum technology applications

Metal	MP (K)	Density ρ (g cm^{-3})	Application examples
Stainless steel	1,803	7.86	Vacuum vessels, flanges, valves, pumps
Fe-Co-Ni and Fe-Ni alloys	1,723	8.2–8.3	Feedthroughs in glass and ceramics (sealing–in process)
W	3,640	19.3	Components for high temperatures, cathodes
Ta	3,270	16.8	Anodes, grids, cathodes
Mo	2,890	10.3	Cathodes, anodes, arc electrodes
Pt	2,046	21.4	Feedthroughs in soft glass, coatings of glass and ceramics
Pd	1,828	12.0	Hydrogen leaks, leakage detectors
Au	1,336	19.3	Vacuum solder, coating for grids and sockets
Ag	1,233	10.5	Solder component, coating of electrical conductors, sockets and copper electrodes
Ni	1,730	8.9	Anodes, grids, shields, oxide cathode bases
Cu	1,356	8.9	Vacuum seals, cooling walls, conductors
Al	932	2.7	Deflection plates, electron windows
Be	1,553	1.85	X-ray windows
Ti	1,963	4.52	Anodes, grids, getter material
Zr	2,403	13.1	Getter material
Ba	987	3.5	Cathodes, getter materials
Cs	301.7	1.9	Cathodes

and elongated crystals (hard-rolled) make the metal hard and brittle. During longer heating above the so-called recrystallization temperature T_R, larger crystallites are formed, whereby the metal becomes softer and ductile (soft anneal). As a rule, the recrystallization temperature T_R can be derived from the melting temperature by

$$T_R = 0.4T_S.$$

Ductility is one of the criteria for good machineability/formability of the respective metal. At heating, at insufficient deformation or at too high temperature very large crystals are formed, and the metal gets brittle. The strength of metals and other materials is usually determined by applying a load to the material under study and by observing the resulting deformation.

Drawn, rolled or pressed metal forms are useable in vacuum technology. Alloys can only be used if the components still have a sufficiently low vapour pressure at the highest operating temperature. Only forming/machining methods are allowed, which do not lead to micro-leaks (e.g. deep drawing, arc erosion, laser cutting). One of the most important materials for UHV chambers and electron guns is wrought stainless steel. Stainless steels are defined as iron–base alloys containing at least 12% of chromium. The type of the stainless steel used (there are more than 37 compositions) will depend on the specific application.

The refractory metals W, Mo, Ta and Nb are of primary interest for electron device construction. Their principal merit is their mechanical strength at elevated temperature which is an approximate function of their (high) melting points. Among these tungsten exhibits the highest transition temperature from ductile to brittle behaviour of 350°C. Hence Mo with a transition temperature of about 0°C is still machineable, whereas W needs powder metallurgical methods for forming, e.g. porous tungsten only becomes machineable when the pores are filled with copper. In general, the choice of metals becomes more and more limited the lower the pressure in the vacuum system is required to be. For UHV chambers only stainless steel, copper and gold are appropriate, but for cathode bases and anodes also W, Mo and Ta are used.

Figure 10.42 shows an example of a UHV diode emission test flange as used by Philips Research. Here the material of the so-called ConFlat flange, the feedthroughs and the mounting posts are made of is stainless steel, the insulating part of the current feedthroughs consists of glass ceramics, whereas the anode is made of Ta and the cathode unit is a combination of a $4BaO \cdot CaO \cdot Al_2O_3$ impregnated tungsten pill in a Mo cap. The cathode to anode distance is 0.3 mm. The insulation here is obtained by quartz cylinders. This flange is then mounted via a copper gasket on the corresponding flange (inner width 35 mm) of the UHV chamber (made of stainless steel) and is tightend by six screws with controlled rotary torques.

Glass, Glass Ceramics and Ceramics

Although glass is a solid in the sense that it has the rigidity that is generally associated with solid bodies, it lacks the continuous crystalline structure that is characteristic for metals and ceramics. On cooling down of the melt, the liquid state is “frozen

Fig. 10.42. UHV diode emission test flange with a 0.65 W I cathode unit below the Ta anode on top, as used by Philips Research (courtesy of Philips Research Aachen)

in", which means that on cooling down no crystals are formed because the viscosity increases sufficiently fast. At a viscosity of more than 10^{12} Pa s (transformation point) glass transforms from a plastic tension free state to an elastic brittle state. The transformation temperature is essential for glass machining, because this is the lower boundary for decrease in tension in glass in a reasonable time. Glass does not possess a sharply defined melting point where liquid and solid phases exist in equilibrium, but looses its solid-like character gradually on heating. The ease of forming and the ease of sealing to metals are two of the advantages of glass. Further properties of glass important for vacuum technology are a reasonable mechanical strength, but fragility and low shock resistance, chemical and thermal robustness, low vapour pressure, low electrical conductivity, transparency to radiation/good visible light transmission and reasonable cost. The main glass constituents besides SiO_2 are B_2O_3, Al_2O_3, CaO, PbO, K_2O, Na_2O and others.

Glass ceramics consist of a mixture of crystals and a residual glass phase. It is prepared by adding nuclei for crystal growth such as TiO_2 or ZrO_2 to the glass melt, which after cooling down and heating up again grow to larger crystals. Glass ceramics have a higher density, higher viscosity and higher mechanical strength than glass. The thermal expansion can be nearly zero, since the thermal expansions of the crystals and the residual glass phase nearly compensate each other.

Modern ceramics are crystalline sintered materials of pure Al_2O_3 or BeO, which are produced at temperatures far below the melting point. They exhibit high values of thermo-mechanical stability, of chemical robustness and of the specific electrical resistivity and are not permeable for gases also without glass coating. They are used for vacuum chambers of particle accelerators, for corrosion resistant vapour beams and for the insulation of current feed-throughs.

Emitter Materials

The emitter materials used in electron sources again can be divided into metals/ elements (see above, including Si, C) and compounds such as oxides, especially of the alkaline earths and of the scandium group, and of the rare earths. Among these we will only address the last three groups of oxides, since they are the most important for low temperature emitters and are not treated elsewhere. Usually, the alkaline earth oxides are prepared in the form of carbonates and then have to be transformed into oxides at elevated temperature, typically under UHV. In nearly all cases the oxides are then used as sources of the respective metals via a suitable chemical reaction in order to form, e.g. a Ba monolayer on the cathode surface either on a metal or an oxide base with a very low work function. Due to the low melting temperature and high vapour pressure of Ba, Ca and Sr this indirect way of Ba, Ca or Sr supply has to be chosen, and hence these cathodes are either oxide or dispenser cathodes. The vapour pressure p of BaO, CaO and SrO are shown in Fig. 10.43 as a function of temperature, the vapour pressure of the respective metals are depicted in Fig. 10.44. Even more critical for UHV is the use of Cs. Here a unique new way is the use of controlled Cs resupply from a CsAu compound.

The oxides of the scandium group and of the rare earths are in most cases only used in small quantities as dopants, e.g. of barium-based oxide cathodes, or as mixed oxides, e.g. Ba-Scandate in the case of Scandate cathodes.

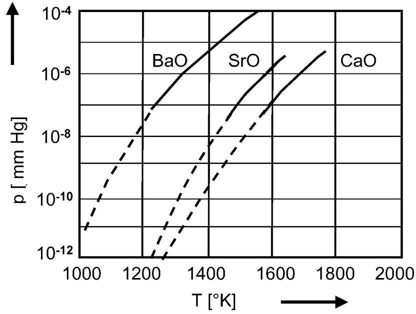


Fig. 10.43. Vapour pressure p (in mmHg = Torr) of alkaline earth oxides as a function of temperature (according to Herrmann and Wagener [97])

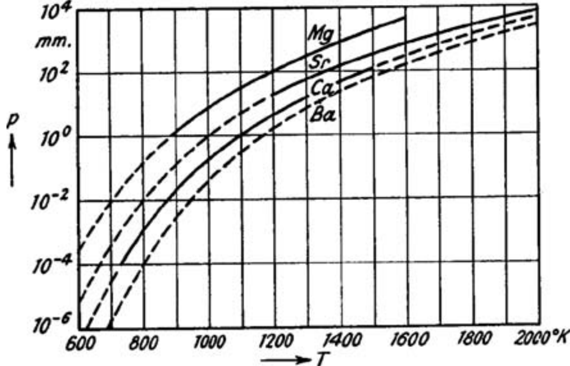


Fig. 10.44. Vapour pressure p (in mmHg = Torr) of alkaline earth metals as a function of temperature [2]

References

[1] F.R. Paturi, *Chronik der Technik* (Chronik-Verlag in der Harenberg Kommunikation, Dortmund, 1989)

[2] G. Herrmann, S. Wagener, *The Oxide Coated Cathode* (Chapman & Hall, London, 1951)

[3] G. Haas, R. Thomas, Thermionic sources for high brightness electron beams, IEEE Trans. Electron Devices **37**(3), 850–861 (1990)

[4] A. Wehnelt, Über den Austritt negativer Ionen aus glühenden Metallen, Annal. Phys. Ser. 4 **14**, 425–468 (1904)

[5] L.S. Nergaard, Electron and ion motion in oxide cathodes, in *Halbleiter-probleme,* Bd. 3 (F. Vieweg-Verlag, Braunschweig, 1956), pp. 154–206

[6] A.L. Reimann, *Thermionic Emission* (Chapman & Hall, London, 1934)

[7] G. Gaertner, P. Geittner, H. Lydtin, A. Ritz, Emission properties of top-layer Scandate cathodes prepared by LAD, Appl. Surf. Sci. **111**, 11–17 (1997)

[8] G. Miram, A. Cattelino, Life test facility for thermionic cathodes, in *TRI Service/NASA Cathode Workshop*, Cleveland, Conf. Record, 1994, pp. 233–236

[9] J. Eichmeier, *Moderne Vakuumelektronik* (Springer, Berlin, 1981)

[10] W. Nottingham, Thermionic emission, in *Handbuch der Physik*, ed. by S. Flügge. Elektronen-Emission/Gasentladungen I, vol. 21 (Springer, Berlin, 1956), pp. 1–175

[11] G. Haas, R. Thomas, Thermionic emission and work function, in *Techniques of Metal Research*, Vol. 6/1, ed. by E. Passaglia. (Interscience Publ., 1972), pp. 94–262

[12] A.G. Knapp, Surface potentials and their measurement by the diode method, Surf. Sci. **34**, 289–316 (1973)

[13] J. Hasker, Calculation of diode characteristics and proposed characterization of cathode emission capability, Appl. Surf. Sci **16**, 220–237 (1983)

[14] C. Herring, M. Nichols, Thermionic emission, Rev. Mod. Phys. **21**(2), 196 (1949)

[15] J. Eichmeier, H. Heynisch (ed.), *Handbuch der Vakuumelektronik* (R. Oldenbourg, München, Wien, 1989)

[16] R.O. Jenkins, A review of thermionic cathodes, Vacuum **19**(8), 353 (1969)

[17] I. Langmuir, The electron emission from thoriated tungsten filaments, Phys. Rev. **22**, 357–398 (1923)

[18] I. Weissman, J. Appl. Phys. **36**(2), 406 (1965)

[19] G. Gaertner, D. van Houwelingen, Electron emission cooling of thermionic thoriated tungsten cathodes under high dc-loads, Elektronenröhren u. Vakuumelektronik, NTG-Fachbericht 95, VDE-Verlag (1986), pp. 224–229
[20] G. Gaertner, NTG Fachberichte **85**, 228–232 (1983)
[21] G. Gaertner, P. Janiel, H. Lydtin, ITG Fachbericht **108**, 297–302 (1989)
[22] G. Gessinger, C. Buxbaum, High Temp.-High Press. **10**, 325–328 (1978)
[23] G. Gaertner, P. van der Heide, New Developments in CRT Cathodes, IDW 2000, Technical Digest CRT4-1, pp. 513–517
[24] G. Gaertner, D. Barratt, Appl. Surf. Sci. **251**, 73–79 (2005)
[25] B. Vancil, E. Wintucky, Appl. Surf. Sci. **251**, 101–105 (2005)
[26] Y. Wang, et al., Appl. Surf. Sci. **251**, 80–88 (2005)
[27] R. Cortenraad, A. van der Gon, H. Brongersma, G. Gaertner, A. Manenschijn, Appl. Surf. Sci. **146**, 69–74 (1999)
[28] S. Kimura, T. Higuchi, et al., Appl. Surf. Sci. **111**, 60–63 (1997)
[29] A. van Oostrom, L. Augustus, Appl. Surf. Sci. **2**(2), 173–186 (1979)
[30] S. Taguchi, T. Aida, S. Yamamoto, IEEE Trans. Electron. Devices **31**(7), 900–903 (1984)
[31] J. Hasker, J. van Esdonk, J.E. Crombeen, Appl. Surf. Sci. **26**, 173–195 (1986)
[32] J. Hasker, J.E. Crombeen, Trans. Electron. Devices **37**(12), 2589–2594 (1990)
[33] S. Yamamoto, I. Watanabe, S. Taguchi, S. Sasaki, T. Yaguchi, Jap. J. Appl. Phys. **28**, 490–494 (1989)
[34] U. van Slooten, P. Duine, Appl. Surf. Sci. **111**, 24–29 (1997)
[35] G. Gaertner, P. Janiel, J.E. Crombeen, J. Hasker, Vacuum microelectronics 1989, in IOP Conf. Ser. **99**, 25–28 (1989)
[36] G. Gaertner, P. Geittner, D. Raasch, A. Ritz, D. Wiechert, Appl. Surf. Sci. **146**, 12–16, 22–30 (1999)
[37] G. Gaertner, P. Geittner, D. Raasch, Low temperature and cold emission of Scandate cathodes, Appl. Surf. Sci. **201**, 61–68 (2002)
[38] M. Saito, et al., Higher current density oxide cathode for CRT, NTG Fachberichte **95**, 165–170 (1986)
[39] G. Gaertner, D. Raasch, D. Barratt, S. Jenkins, Accelerated life tests of CRT oxide cathodes, Appl. Surf. Sci. **215**, 72–77 (2003)
[40] G. Gaertner, P. Janiel, D. Raasch, Direct determination of electrical conductivity of oxide cathodes, Appl. Surf. Sci. **201**, 35–40 (2002)
[41] S. Hodgson, C. Goodhand, P. van der Heide, et al., Processing and performance of a novel cathode material, Appl. Surf. Sci. **146**, 79–83 (1999)
[42] T. Higuchi, Recent trends in thermionic cathodes, in *IDW'98*, CRT3-1, p. 393
[43] Y.C. Kim, K. Joo, J. Choi, H. Yang, in *IVESC 2000 Techn. Digest* D-4
[44] www.kimballphysics.com
[45] J.M. Lafferty, Boride cathodes, J. Appl. Phys. **22**(3), 299–309 (1951)
[46] H. Ahmed, A. Broers, J. Appl. Phys. **43**, 2185–2192 (1972)
[47] S. Dushman, Thermionic emission, Rev. Mod. Phys. **2**, 405–414 (1930)
[48] V. Fomenko, Emission properties of materials, NTIS, JPRS-56579 (1972)
[49] A. Makarov, et al., Cesium coated graphite emitter, Sov. Phys.-Tech. Phys. **22**(12), 1463–1465 (1978)
[50] P.A. Duine, in *IVMC*, Darmstadt, 1999, pp. 368–369
[51] P. Coates, Thermionic emission from photocathodes, J. Phys. D: Appl. Phys. **5**, 1489–1498 (1972)
[52] R.H. Good Jr., E.W. Müller, Flügge, Elektronen-Emission, in *Handbuch der Physik*, Bd. XXI (Springer, Berlin, 1956), S. 176–231

[53] R.H. Fowler, L. Nordheim, Proc. R. Soc. London Ser. A **119**, 173 (1928)
[54] E.W. Müller, Ergeb. Exakt. Naturwiss **27**, 290 (1953)
[55] K.R. Shoulders, Microelectronics using electron-beam-activated machining techniques, Adv. Comp. **2**, 135 (1961)
[56] J.W. Gewartowski, H.A. Watson, *Principles of Electron Tubes* (D. Van Norstrand, Princeton, 1965), p. 229
[57] W.P. Dyke, W.W. Dolan, *Advances in Electronics and Electron Physics*, vol. 8 (Academic, New York, 1956), p. 89
[58] C.J. Spindt, Appl. Phys. **39**, 3504 (1968)
[59] C. Spindt, I. Brodie, L. Humphrey, E.R. Westerberg, J. Appl. Phys. **47**, 5248 (1976)
[60] R. Meyer, A. Ghis, P. Ramband, F. Muller, Development of matrix array of cathode emitters on a glass substrate for flat display applications, in *Proc. 1st IVMC*, Williamsburg, VA, 1988, p. 10
[61] I. Brodie, C.A. Spindt, Appl. Surf. Sci. **2**, 149 (1979)
[62] I. Brodie, C.A. Spindt, Vacuum microelectronics, in *Advances in Electronics and Electron Physics*, ed. by P.W. Hawkes, vol. 83 (Academic, New York, 1992), p. 1
[63] D.W. Tuggle, L.W. Swanson, Emission Characteristics of the Zr-O-W thermal field electron source, J. Vac. Sci. Technol. B **3**, 220 (1985)
[64] L.W. Swanson, A comparative study of the zirconiated and build-up W cathode, J. Vac. Sci. Technol. **12**(6), 1228 (1975)
[65] H.G. König, H. Koops, A Study of Zr-O-W- and W-field emitters in an electron source at high vacuum conditions, in *Proc. Int. Conf. Microcircuit Engineering,* Berlin, 1984, ed. by A. Heuberger, H. Beneking (Academic, New York, 1985), pp. 195–202
[66] T. Asano, T. Tamon, Tech. Dig. IVMC91, 1991, p. 88
[67] M. Komuro, H. Hiroshima, J. Vac. Sci. Technol. B **9**, 2656 (1991)
[68] P.R. Schwoebel, C.A. Spindt, Appl. Phys. Lett. **63**, 33 (1993)
[69] P.R. Wilshaw, E.C. Boswell, J. Vac. Sci. Technol. B **12**, 662 (1994); M. Takai, M. Yamashita, H. Wille, S. Yura, S. Horibata, M. Ototake, Appl. Phys. Lett. **66**, 422 (1995)
[70] M. Takai, M. Yamashita, H. Wille, S. Yura, S. Horibata, M. Ototake, J. Vac. Sci. Technol. B **13**, 441 (1995)
[71] S. Matsui, K. Mori, J. Vac. Sci. Technol. B **4**, 299 (1986)
[72] H.W.P. Koops, J. Kretz, M. Rudolph, M. Weber, G. Dahm, K.L. Lee, Jpn. J. Appl. Phys. **33**, 1345 (1994)
[73] J. Kretz, M. Rudolph, M. Weber, H.W.P. Koops, Three dimensional structurization by additive lithography, analysis of deposits using TEM and EDX, and application for field emitter tips, Microelectron. Eng. **23**, 477–481 (1994)
[74] M. Takai, T. Kishimoto, M. Yamashita, H. Morimoto, S. Yura, A. Hosono, S. Okuda, S. Lipp, L. Frey, H. Ryssel, J. Vac. Sci. Technol. **14**, 1973 (1996)
[75] T. Hirano, S. Kanemaru, H. Tanoue, J. Itoh, Jap. J. Appl. Phys. **35**, 6637 (1996)
[76] K. Okano, et al., Appl. Surf. Sci. **146**, 274–279 (1999)
[77] J.F. Xu, et al., Appl. Surf. Sci. **146**, 280–286 (1999)
[78] W.B. Choi, D.S. Chung, J.H. Kang, H.Y. Kim, Y.W. Jin, I.T. Han, Y.H. Lee, J.E. Jung, N.S. Lee, G.S. Park, J.M. Kim, Appl. Phys. Lett. **75**, 3129 (1999)
[79] S.K. Kang, J.H. Choi, J.H. Park, J.-H. Han, J.-B. Yoo, J.-W. Nam, C.K. Lee, J.M. Kim, Relationship between field emission property and composition of carbon nanotube paste for large area cold cathode, J. Vac. Sci. Technol. B **22**(3), 1345 (2004)
[80] K.A. Dean, B.F. Coll, Y.W. Xie, A.A. Talin, A. Howard, C.D. Moyer, J. Trujillo, J. Jaskie, Motorola Inc. Physical Science Labs, Tempe, AZ

[81] A.G. Chakhovskoi, C.E. Hunt, G. Forsberg, T. Nilsson, P. Persson, Reticulated vitreous carbon field emission cathodes for light source applications, J. Vac. Sci. Technol. B **21**, 571 (2003)
[82] H. Busta, Field emission flat panel displays, Chapter 7, in *Vacuum Microelectronics*, ed. by W. Zhu (Wiley, New York, 2001)
[83] E.A. Hijzen, The avalanche cold cathode for CRTs, in *IDW'98*, 1998, pp. 405–408
[84] A.M.E. Hoebrechts, G.G.P. van Gorkum, Design, technology, and behavior of a silicon avalanche cathode, J. Vac. Sci. Technol. B **4**, 105 (1986)
[85] S.H. Jo, J.D. Lee, Fabrication and analysis of a silicon tip avalanche cathode, J. Vac. Sci. Technol. B **13**(2), 469 (1995)
[86] W. Fuhs, P. Kanschat, K. Lips, Bandtails and defects in microcrystalline silicon (mu c-Si:H), J. Vac. Sci. Technol. B **18**, 1792 (2000)
[87] J.-S. Kim, T. Hoshi, K. Sawada, M. Ishida, Planar metal–insulator–semiconductor type field emitter fabricated on an epitaxial Al/Al2O3/Si (111) structure, J. Vac. Sci. Technol. B **22**, 1358–1361 (2004)
[88] H.A. Baba, T. Yoshida, T. Asano, Field emission characteristics of defect-controlled polyimide tunnelling cathode, J. Vac. Sci. Technol. B **22**, 1353–1357 (2004)
[89] Riege, Ferroelectric electron emission: Principles and technology, Appl. Surf. Sci. **111**, 318–324 (1997)
[90] J.D. Clewley, A.D. Crowell, D.W. Juenker, Changes in photoelectron emission from molybdenum due to exposure to gases, J. Vac. Sci. Technol. **9**, 877 (1972)
[91] T.E. Fischer, Photoemission and surfaces, J. Vac. Sci. Technol. **9**, 860 (1972)
[92] http://www.saesgetters.com
[93] D. Martin, et al., Design of glass substrates and spacers for FEDs, Saint-Gobain Display Glass, France, Technical digest EURO FE, 2002
[94] http://www.mems-exchange.org/catalog/
[95] J. Dziuban, R. Walczak, Microwave enhanced wet anisotropic etching of silicon utilizing a memory effect of KOH activation – a remote E2MSi process, in *The 16th European Conference on Solid-State Transducers Eurosensors XVI*, Prague, Czech Republic, T1A3, September 15–18, 2002
[96] J. Eichmeier, *Moderne Vakuumelektronik* (Springer, Berlin, 1981), chapter 9.4
[97] J. Eichmeier, H. Heynisch (ed.), *Handbuch der Vakuum-elektronik* (R. Oldenbourg, München, Wien, 1989), chapter 2
[98] W. Kohl, *Handbook of Materials and Techniques for Vacuum Devices* (Reinhold Publishing corporation, 1967)

11

Vacuum Technology

W. Knapp

11.1 Introduction and History

Vacuum is defined as a space that is entirely devoid of matter, i.e. an enclosed volume that is not filled with air or any other gases. Galileo, at the beginning of the seventeenth century, generated the first partial vacuum with a simple piston. Then he showed that bodies with different density fall with the same velocity. Torricelli invented the mercury barometer in 1643. Some years later the first vacuum pumps were designed by von Guericke and described by Casper Schott in 1657. The most famous historical vacuum experiment in Germany is that of Otto von Guericke with the Magdeburg halfspheres, which dates as early as 1654 (Fig. 11.1). The interest in low pressure was slight until in the second half of the nineteenth century discharge phenomena hastened the development. In 1874 McLeod made the compression manometer and measured pressures down to 10^{-3} Pa. The first major use of vacuum technology in industry occurred about 1900 in the manufacture of electrical light bulbs. In 1905 Gaede constructed the rotary vane pump, which was sealed by mercury. In 1915 he invented the mercury diffusion pump, the first pump able to achieve high vacuum [1]. These inventions provided art which made possible the vacuum technology from light bulbs to high-tech vacuum systems, such as coating plants, surface analytical instruments, space simulators and so on. This chapter gives a review of fundamentals, components and devices for vacuum measurement, production and other key aspects of vacuum technology.

11.2 Gases in Vacuum Systems

11.2.1 The Fundamentals

The macroscopic state of a gas is described by the so-called variables of state: volume, pressure and temperature. Volume V is the content of a space, which is filled completely by the gas. The gas exerts pressure p on the walls of the vacuum chamber

Fig. 11.1. Historical vacuum experiment of Otto-von-Guericke with the Magdeburg half-spheres [1]

in which it is enclosed. Pressure p is defined as the quotient of standard force exerted on a surface (unit of force per unit of surface area)

$$p = \frac{\mathrm{d}F}{\mathrm{d}A}. \tag{11.1}$$

Taking (11.1) as a basis, the result is the unit for pressure, a derived unit in the International System (SI),

$$[p] = \text{Newton/meter}^2 = \text{Pascal};\ \text{N/m}^2 = \text{Pa}. \tag{11.2}$$

Other permissible units of measure for pressure include:

- bar (standard abbreviation bar); 1 bar $= 10^5\ \mathrm{N\,m^{-2}} = 10^5$ Pa,
- millibar (standard abbreviation mbar): 1 mbar $= 10^{-3}$ bar.

Other, less commonly used, units for pressure are:

- Torr: 1 Torr $= 133.3224$ Pa ≈ 1.33 mbar,
- technical atmosphere (standard abbreviation at):
  1 at $= 1\ \mathrm{kp\,cm^{-2}} = 0.981$ bar (1 kp $= 9.81$ N),
- physical atmosphere (standard abbreviation atm);
  1 atm $= 760$ Torr $= 1.013$ bar $= 1.033$ at.

11.2.2 The Ideal Gas Law

Equation (11.3), the equation of a state for the ideal gas, shows that at a given temperature T and mass m of a substance, the product of pressure p and volume V will remain constant.

$$p \cdot V = \frac{m}{M} \cdot R \cdot T. \qquad (11.3)$$

Here the quotient m/M corresponds to the number n of the moles (M = molar mass) found within the volume V, and R is the ideal gas constant.

The primary function of vacuum technology is to reduce the number density of the molecules inside a given volume. At constant temperature and volume, this will always be equivalent to reducing the gas pressure. At this juncture it is necessary to point out that it is possible to lower the pressure (V = const) not only by reducing the number of molecules present, but also lowering the temperature of the same number of molecules. This fact must always be taken into account when the temperature within the evacuated space is not uniform throughout.

11.2.3 The Pressure Ranges in Vacuum Technology

It is common in vacuum technology to subdivide the wide range of pressures, which can be technically achieved, today covering more than 16 powers of ten, into individual sectors, which are listed in Table 11.1.

Depending on the prevailing vacuum pressure range, there are three types of gas flow:

1. The first is viscous or *continuous flow*, which is found almost exclusively in the low vacuum range. The interplay of the molecules one with another determines the characteristics of this type of flow. The internal friction and the viscosity of the flowing substance have a major effect. If eddies occur, then this is a turbulent flow; if various layers of the flowing medium slide one over the other, then the flow is characterized as laminar. Laminar flow in cylindrical tubes with parabolic velocity distribution is known as Hagen–Poiseuille flow. This special case is encountered particularly often in vacuum technology. Viscous flow is normally present when the mean free path l_{free} of the molecules is much shorter than the diameter of the vacuum chamber or tube,

$$l_{\text{free}} \ll d. \qquad (11.4)$$

   A characteristic magnitude for the state known as viscous flow is the Reynolds number Re. Re is a dimensionless value. At $Re > 2{,}300$, the flow becomes increasingly turbulent. Where $Re < 2{,}300$, the flow is more laminar.

Table 11.1. Vacuum regions

Name of the vacuum	Short symbol	Pressure range
Low (or rough) vacuum	LV	300 to 1 mbar[a]
Medium vacuum	MV	1 to 10^{-3} mbar
High vacuum	HV	10^{-3} to 10^{-7} mbar
Ultrahigh vacuum	UHV	10^{-7} to 10^{-12} mbar
Extremely ultrahigh vacuum	XHV	$\leq 10^{-12}$ mbar

[a] 1 mbar = 100 Pa = 1 hPa

2. *Molecular flow* prevails in the high, ultrahigh, and extremely ultrahigh vacuum ranges. In these ranges the molecules can move freely, without interfering one with another. Molecular flow is present when the mean free path l_{free} for a module is much larger than the diameter of the vacuum chamber or tube,

$$l_{\text{free}} \gg d. \tag{11.5}$$

3. *Knudsen flow* is found at the transition between viscous and molecular flow. It is present above all in the medium vacuum range

$$l_{\text{free}} \approx d. \tag{11.6}$$

The product of pressure p and tube parameter d for a certain gas at a certain temperature can serve as a characteristic magnitude for the various types of flow. For air at 20°C the following equivalent relationships apply:
low vacuum (viscous flow)

$$p \cdot d > 6 \times 10^{-1}\ \text{mbar cm}, \quad l_{\text{free}} < \frac{d}{100}; \tag{11.7}$$

medium vacuum (Knudsen flow)

$$6 \times 10^{-1}\ \text{mbar cm} > p \cdot d > 1.3 \times 10^{-2}\ \text{mbar cm}, \quad \frac{d}{100} < l_{\text{free}} < \frac{d}{2}; \tag{11.8}$$

high, ultrahigh and extremely ultrahigh vacuum (molecular flow)

$$p \cdot d < 1.3 \times 10^{-2}\ \text{mbar cm}, \quad l_{\text{free}} > \frac{d}{2}. \tag{11.9}$$

In the viscous flow range, the preferred vectors for the velocity of all the gas molecules is equivalent to the macroscopic direction of flow for the gas. Such an orientation is enforced by the fact that the gas molecules are densely packed and collide far more frequently one with another than with the limiting envelope represented by the equipment walls. In the molecular flow range, on the other hand, collisions with the envelope predominate. As a result of reflection but also due to desorption following a certain dwell time on the envelope walls, a molecule of gas in the high vacuum range can assume any arbitrary direction of movement. It is no longer possible to refer to flow in the macroscopic sense. In the high, ultrahigh and extremely ultrahigh vacuum ranges the properties of the vacuum container's wall are of decisive significance. Below 10^{-3} mbar there are more molecules on the surface than in the enclosed space itself.

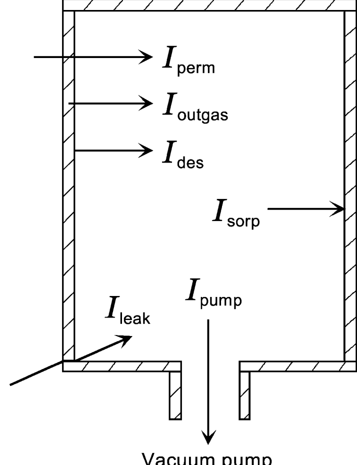


Fig. 11.2. Sources of gas which present a load to the vacuum pump

11.2.4 Surface Interactions and Outgassing

The interactions between gas molecules and the walls of the vacuum chamber in which the gas is retained by the solid received the name *sorption*. This includes two mechanisms: *adsorption* and *absorption*. The term *adsorption* refers to the process whereby molecules are attracted to and become attached to the surface of a solid, the resulting layer of adsorbed gas being one molecule (*monolayer*) or a few molecules (*multilayer*) thick. The attracting forces of the solid may be physical (*physisorption*) or chemical (*chemisorption*). The term *absorption* refers to gas which enters into solid in much the same manner as gas dissolving in a liquid. The solid which takes up the gas is known as adsorbent or absorbent, the gas removed is known as *adsorbate* or *absorbate*. Terms as *adatom* or *admolecule* are also used to refer to the specific particles involved in the process. When a material is placed in a vacuum, the gas which was previously ad- or absorbed begins to *desorb*, i.e. to leave the material. The *desorption* is influenced by the pressure, the temperature, the shape of material, and a kind of its surface. Gas from the *desorption* is known as *outgassing*. The *specific outgassing rate* is defined as the rate at which gas appears to emanate from unit area of surface (geometrical surface) and is usually measured in units of mbar s^{-1} cm^{-2}. The gas flows are shown schematically in Fig. 11.2. More information about sorption and desorption phenomena are given in [2, 3].

11.3 Vacuum Pumps

11.3.1 Introduction

Various types of vacuum pumps are used to produce vacuum in the laboratory or industrial environments (Fig. 11.3). This section describes examples of pumps in

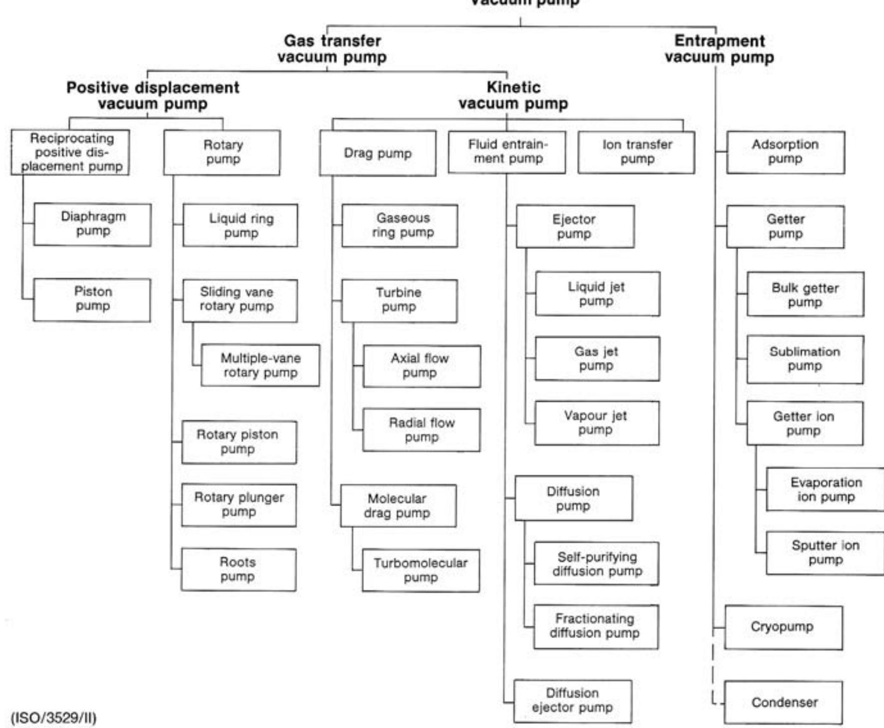


Fig. 11.3. Classification of vacuum pumps

common use, chosen to show principles of operation in various groups. It is realised that it is by no means exhaustive and does not cover innumerable variations and manufactures special features. For further information, the special edition of vacuum pumps and in particular the catalogues of vacuum companies are recommended [4, 5].

11.3.2 Pumps for Rough and Medium Vacuum

Rotary Vane Vacuum Pumps

The rotary vane vacuum pump is a typical example of an oil-immersed positive displacement pump. The central component of a rotary vane vacuum pump is the pumping system. It consists of a cylinder with ports leading to the outside. These ports are used to take in and exhaust the gases to be pumped. Inside the cylinder, there is an eccentrically-arranged rotor. The vanes are fitted into slots on the rotor. The vanes, which glide along the cylinder wall, divide the available inner space into working chambers. During one full rotation of the rotor, the chamber volume increases from zero to the maximum volume and then decreases continually until it reaches the minimum value. Actual pumping is effected by the increase and decrease in size

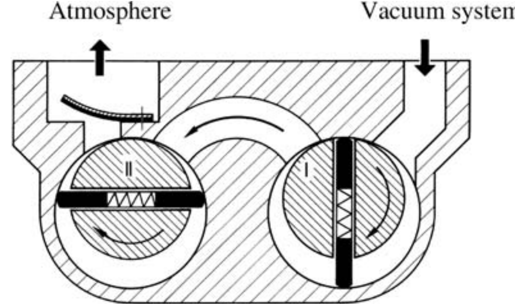


Fig. 11.4. A cross-sectional diagram of a two-stage rotary vane vacuum pump

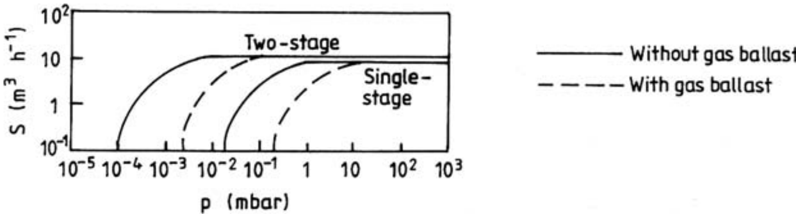


Fig. 11.5. Typical pumping speeds for rotary vane vacuum pumps [2]

of the sickl-shaped chambers of the working space. The decrease in chamber volume compresses the enclosed gases. The compression at approximately 200 mbar above atmospheric pressure allows for the gas pressure to be higher than the opening pressure of the exhaust valve. Rotary vane vacuum pumps are available in single- and two-stage models. Two-stage pumps have a lower ultimate pressure than single-stage pumps (Fig. 11.4). Rotary vane vacuum pumps can be used without problem whenever the medium to be pumped is a gas that cannot condense at the operating temperature of the pump (Fig. 11.5) and at atmospheric pressure.

Diaphragm Pumps

The diaphragm pump operates by the flexing of an elastomer diaphragm moved by a motor driven piston. Principle of the diaphragm pump is as follows. The elastomer diaphragm (see Fig. 11.6) is moved up and down by an eccentric. On the down-stroke it draws the air or gas being handled through the inlet valve. On the up-stroke the diaphragm forces the medium through the exhaust valve and out of the head. The compression chamber is hermetically sealed from the drive mechanism by the diaphragm. The pumps transfer, evacuate, and compress completely oil-free.

Features of diaphragm pumps:

- uncontaminated transfer, no pollution of the media,
- 100% maintenance-free,
- compact design,

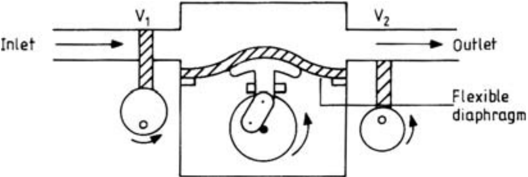


Fig. 11.6. Scheme and view of diaphragm pump

- smooth running,
- easy installation,
- excellent gas tightness,
- with very few exceptions, the pumps can operate in any installed position.

These pumps will operate to an ultimate pressure of 1 mbar and are used for oil-free roughing of UHV systems to allow other pumps to be started, especially turbomolecular pumps (see Sect. 11.3.3).

11.3.3 Pumps for High Vacuum (HV) and Ultrahigh Vacuum (UHV)

Vapour Diffusion Pumps

This pump is mainly used in equipment for the study of clean surfaces and in radio frequency sputtering. Pumping speeds are available up to 5,000 cubic meters per minute with an operating pressure range of 10^{-2} to less than 10^{-9} mbar when water-cooled baffles are used and less than 10^{-11} mbar when refrigerated baffles are employed. The pumping speed for a vapour pump remains constant from about 10^{-3} mbar to well below the ultimate pressure limitations of the pump fluid. The best fluids allow pressures of better than 10^{-9} mbar. The diffusion pump is initially evacuated by an oil-sealed rotary pump to a pressure of about 10^{-1} mbar or less. When the pump fluid in the boiler is heated, it generates a boiler pressure of a few mbar within the jet assembly. High-velocity vapour streams emerge from the jet assembly, impinge and condense on the water or air-cooled pump walls, and return to the boiler. In normal operation part of any gas arriving at the inlet jet is entrained, compressed, and transferred to the next stage. This process is repeated until the gas is removed by the mechanical forepump (Fig. 11.7).

The oil-vapour booster pump works on the same principles as the diffusion pump, but it employs a higher boiler pressure. Normal operating pressure range is 1 to 10^{-4} mbar. When backed by an oil-sealed rotary pump, this pump is widely used for achieving high vacuum in thin-film evaporation units, accelerators, and in TV tube pumping.

Cryopumps

Operation is achieved by the condensation, freezing and/or sorption of gas at surfaces maintained at extremely low temperatures, thus removing them from the gas phase in

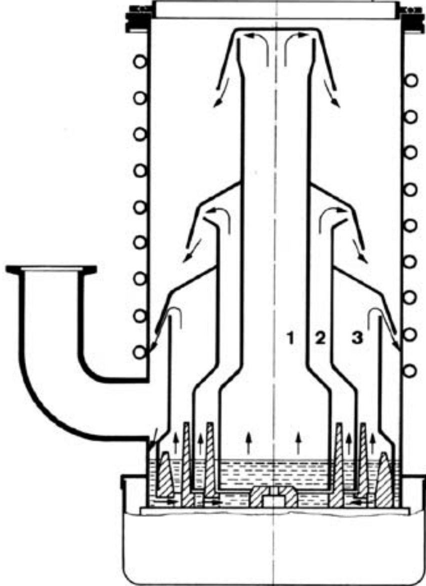


Fig. 11.7. Cross-sectional diagram of a diffusion pump

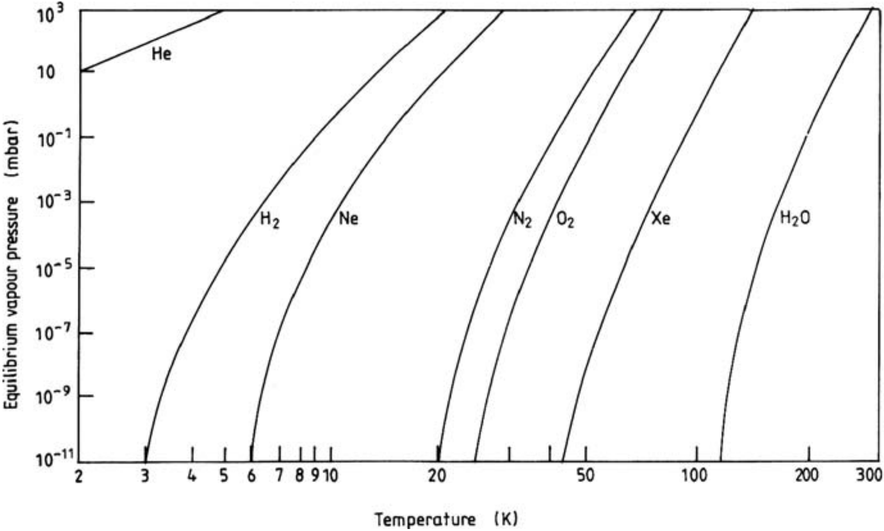


Fig. 11.8. Diagram of temperature and vapour pressure for various gases in a cryopump

the vacuum. A cryopump use the fact that van der Waals forces keep gas molecules on a container surface cooled to a sufficiently low temperature [6]. Figure 11.8 shows the temperatures required to maintain an equilibrium vapour pressure equal to or less than the required low pressure in the chamber.

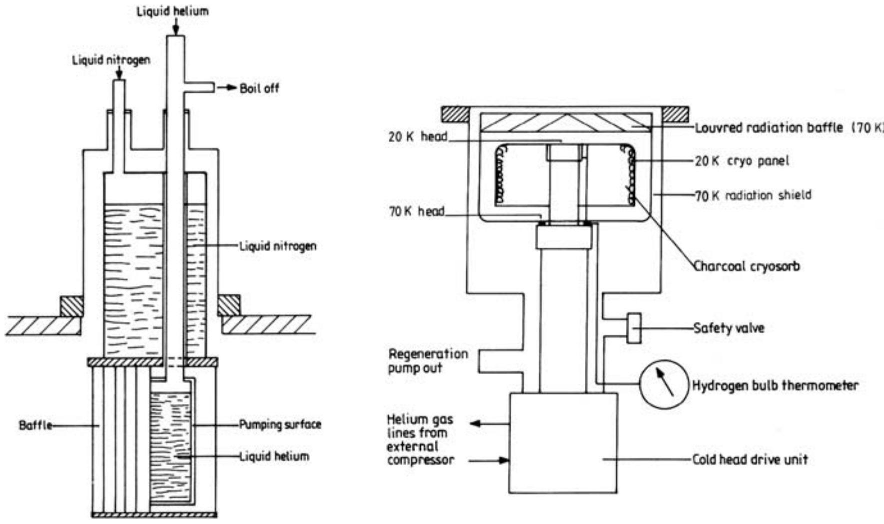


Fig. 11.9. Cross-section diagrams of cryogenic pumps: a liquid helium cryopump (*left*) and a self-contained cryopump with refrigerator (*right*)

Cryopumps, most of which use helium as working gas, are clean UHV or XHV pumps and, unlike ion pumps, do retain no condensed gases after heating and must be vented periodically. Two helium cryopumps have evolved: a liquid helium (bath cryopump) and a self-contained cryopump with refrigerator (Fig. 11.9).

Evaporable Getter (EG) Pumps

This pump uses the principle of "gettering" where gas is retained by chemical combination with a getter in the form of metal or metal alloy, either in bulk or in the form of a deposited thin film. The vacuum is produced during the baking and exhaust step through a thin tubulation that is pinched off at the end of the process. The Ba getter, usually in the shape of a small ring, is then flashed using an external RF coil, as in the case of cathode ray tubes (CRTs). A few mm of barium film are produced in one corner of the display, in a region which does not affect the emission of light.

Non-Evaporable Getter (NEG) Pumps

The use of NEG-based pumping devices in these vacuum systems has become an established approach. When the pumping performance needs a boost, and when a lower ultimate pressure and a cleaner vacuum environment are essential, non-evaporable getter pumps are coupled with other UHV pumps to make the difference.

The unique pumping mechanism of NEG pumps with the highest hydrogen pumping speed, the possibility to operate without a power source, their compactness and flexibility in configuration make them the ideal solution for a variety of

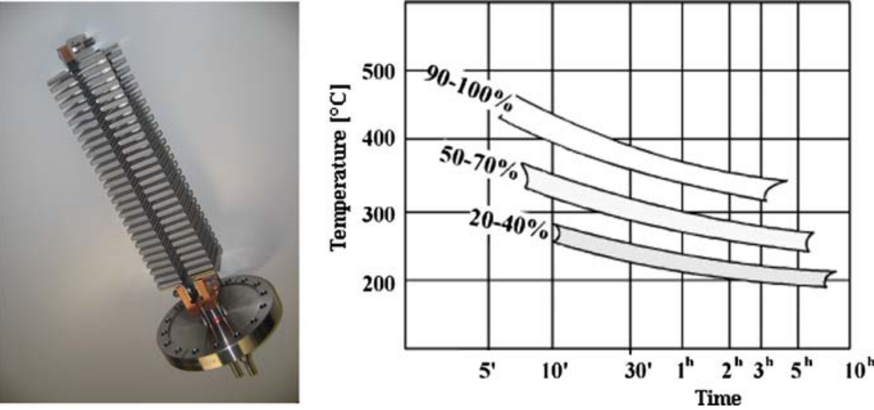


Fig. 11.10. Getter pump with two NEG strips (*left*) and diagram of activation of St 707^{TM} (*right*)

applications in the R&D and industrial fields, e.g. important applications in vacuum electronic devices are X-ray tubes, travelling wave tubes, and vacuum interrupters.

The NEG-activation process consists of heating the getter under vacuum, cf. Fig. 11.10 (right). During the activation process, the oxides present on the getter's surface decompose and diffuse towards the bulk of the getter material. In general, the higher the temperature of the getter material the higher the diffusion rate of atoms of the molecules absorbed on the surface, providing the potential to fully exploit a large portion of getter material. In principle, when the surface of getter material is saturated by gas molecules at room temperature, a process of re-heating under vacuum will re-activate the getter material to further improve its sorption capacity. In practice, the reactivation process of the getter material in vacuum-electronic devices is possible only during the last fabrication steps, so it is not a viable solution during the life of vacuum-electronic devices.

Ion Pumps

This makes use of the gettering principle, in which a cathode material, typically titanium, is vaporised or sputtered by bombardment with high-velocity ions. The active gases are pumped by chemical combination with the sputtered titanium, the inert gases by ionisation and burial in the cathode, and the light gases by diffusion into the cathode. Capacities are available up to 400 cubic feets per minute, with an operating pressure range of 10^{-11} mbar. The full speed of the pump is developed in the pressure range from about 10^{-6} to 10^{-8} mbar, although the characteristic at the lower pressure is dependent on the pump design.

A typical pump consists of two flat rectangular cathodes with a stainless steel anode between them made up of many open-ended boxes (cf. Fig. 11.11). This assem-

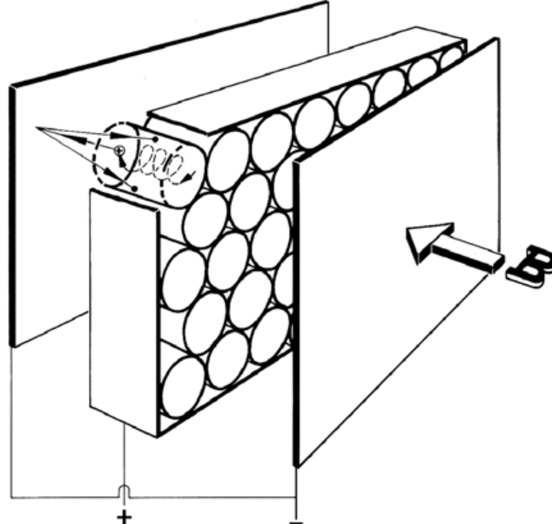


Fig. 11.11. Ion pump: scheme of ionization (diode type)

bly, mounted inside a narrow box attached to the vacuum system, is surrounded by a permanent magnet. The anode is operated at a potential of about 7 kV, whereas the cathodes are at ground potential. The sputter ion pump has low speed and sometimes is unstable when pumping inert gases. To improve its characteristics, other types of sputter ion pumps have been developed: the slotted cathode, triode (cf. Fig. 11.12), differential, and magnetron pumps. To start up a sputter ion pump, it is necessary to reduce the pressure to at least 2×10^{-2} mbar, and preferably much lower, by means of a roughing pump. Sputter ion pumps can operate in any position and do not need water or liquid nitrogen supplies. They have a long life and can provide very clean, ultrahigh vacuum, free of organic contamination and vibration. They are employed mainly for the clean-surface studies and in those applications where any organic contamination will give unsatisfactory results.

Sublimation Pumps

A sublimation pump is a UHV pump which pumps chemically active gases (H_2, O_2, CO, CO_2, H_2O, etc.), and therefore is especially useful pumping, together with a turbopump, hydrogen. The cross-section of a sublimation pump is shown schematically in Fig. 11.13. Titanium atoms evaporate from a directly heated Ti-Mo-alloy wire and sublime on the cold (often LN_2) chamber walls where they make a thin clean layer which binds active gases by adsorption, diffusion, and formation of chemical compounds. The Ti-Mo-alloy wire evaporates at a rate of $\approx$100 mg per hour, so that it can be used for a few hours. The active pumping speed is 3 … 15 l s^{-1} cm^{-2}, in dependence of temperature. Those are the reasons why a sublima-

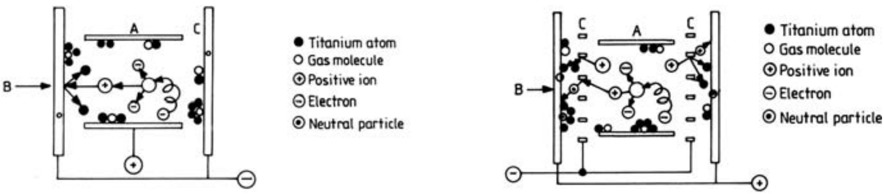


Fig. 11.12. Two basic principles of an ion pump: diode type (*left*) and triode type (*right*)

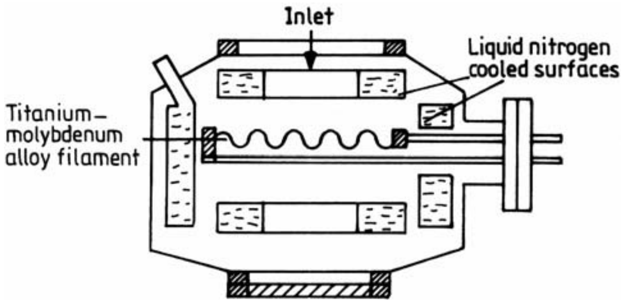


Fig. 11.13. Cross-section of a sublimation pump

tion pump can be used only in a UHV system to complete other pumps, especially with ion pumps and turbomolecular pumps.

Turbomolecular Pumps

A turbomolecular pump (in short, turbopump) essentially comprises a casing with a rotor and a stator. Rotating and fixed disks (or blades) are arranged alternately. All disks contain oblique channels whereby the rotor disk channels are arranged in mirror image to the channels on the stator disks (see Fig. 11.14). A rotor disk together with the stator disk forms a pump stage that generates a particular compression ratio (that for air is approximately 30). The compression effect is multiplied by the sequential switching of several pump stages and very high ratios are attained (for example, for air $\geq 10^{12}$). The important vacuum data of a turbomolecular pump, the pumping speed S and the compression ratio K, depend on the version of the pump. Pumping speed S is proportional to the intake flange A and approximately proportional to the rotation speed of the blades v,

$$S \approx 1/4A \cdot v. \tag{11.10}$$

In the molecular flow range, the pumping speed is independent of the pressure and reduces in the transition range to the laminar flow for the reason given. A typical pumping speed in dependence on the intake pressure can be seen in Fig. 11.15.

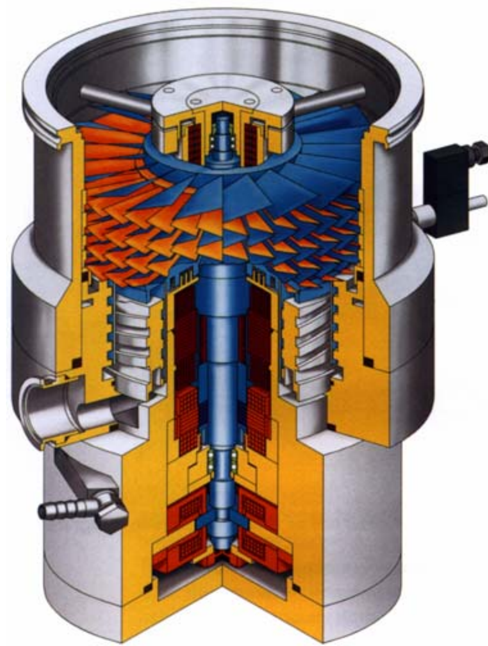


Fig. 11.14. Cross-section of a turbopump with magnetic bearing and Holweck stage

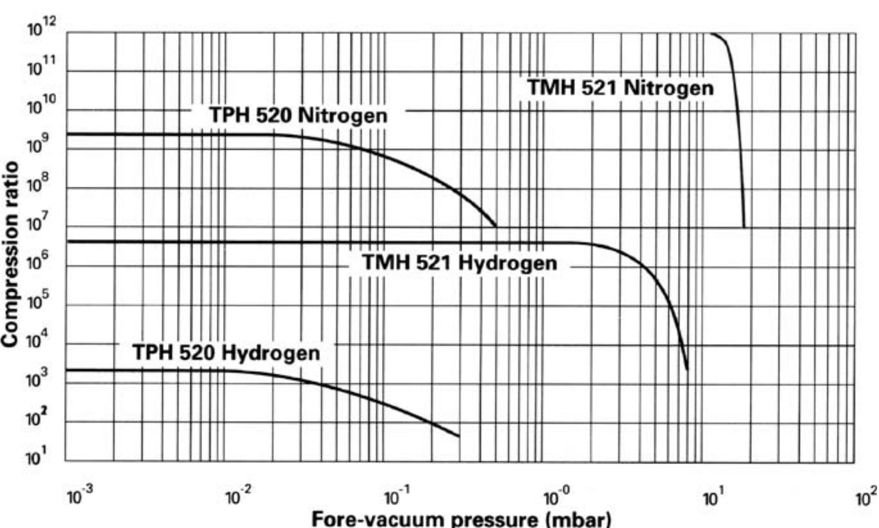


Fig. 11.15. Comparison of the compression ratios of the classical turbopump (TPH 520) and turbo drag pump (TMH 521), and as a function of various gases [6]

The compression ratio K depends on the mean molecular speed, the proportion of the root of the molecular mass M and the blade speed v. In this case the dependence

is exponential. For the maximum compression rate $K_{\max}$ concerned the following is valid:

$$K_{\max} \approx e^{\sqrt{M \cdot v}}. \tag{11.11}$$

Very high compression for heavy gases arises from the dependence of the compression ratio on the molecular mass. The compression breaks down very quickly on the transition from the molecular into the laminar flow range. Therefore, a fore-vacuum pressure of lower 0.1 mbar must be ensured in order to operate the classical turbopump. If the critical backing pressure is exceeded, back streaming can cause shock penetration of the fore vacuum in the intake area of the turbopump. Different technical approaches are used to improve the exhaust pressure of a turbopump. A fully developed construction is to add a Holweck drag stage downstream of the turbo stage. This stage consists of a rotating cylinder and a stationary part with a thread. As with the turbomolecular stages, the Holweck stages are most effective in the molecular flow range. The typical dimension of a pump channel on the Holweck stages is smaller than on the turbo disks by a factor of 10 to 50. Molecule collisions attain significance only at higher pressures. The molecular flow is therefore retained up to a few mbars.

Fully magnetic bearing turbopumps have been designed for special applications. Electromagnets are used to hold the rotor electronically suspended in axial direction. The radial position can be regulated either via the permanent magnetic bearing or also electronically. Because there is no direct contact of the rotor with the casing, the vibration level is extremely low. Additionally, out of balance compensation takes place, and this reduces the vibration even further. It is less than for comparable ball bearing turbopumps by approximately an order of magnitude of one. A further advantage besides the absence of oil on the fore-vacuum side is the reduced wear and tear and the freedom of maintenance. In the event of a power failure, the magnetic bearing is supplied with the power by the energy from the rotor. This ensures continuation of power supply for a few minutes. Any mounting position can be selected for magnetic bearing turbopumps in turbo drag and normal turbo-technology.

The rotor of every turbopump is driven and controlled. This is done using electronic drives that are either integrated into the turbopump or formed as a separate unit depending on the design of turbopump.

11.3.4 Pump Selection

The correct selection of a pumping system is an important condition for vacuum-system design. The correct selection will depend on many factors. A overview is given in Fig. 11.16.

11.4 Vacuum Pressure Measurement

11.4.1 Introduction

The range of vacuum technologies extends about 18 orders of magnitude of pressure below atmospheric. There is no single gauge which is able to measure the full pres-

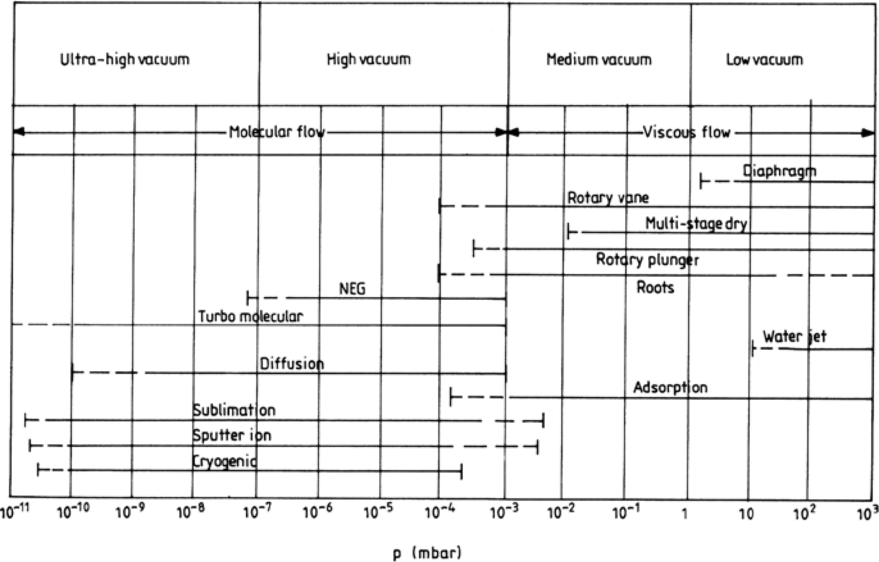


Fig. 11.16. The working pressure range of vacuum pumps

sure range. Each type and kind of gauge is sensitive to variations of pressure in a specific range. Most of the gauges measure total pressure [4, 5, 7].

11.4.2 Total Pressure Gauges

Thermal Conductivity Gauges

For measuring pressures in the range 10^{-2} to 100 mbar, thermal conductivity gauges are commonly used. The thermal conductivity gauge is a vacuum gauge (also known as a *Pirani gauge*) containing two surfaces at different temperatures between which heat can be transported by gas molecules. Changes in the temperatures, or in the heating power required to maintain the temperature of one of the surfaces constant, can be correlated with the gas pressure. Thermal conductivity gauges differ in their method of indicating the temperature change.

Pirani gauges detect the cooling effect of residual gas molecules on a heated filament. The rate of heat transfer to the gas is related to pressure and causes a change in the electrical resistance of the filament or the amount of power required to maintain it at constant temperature. The filament is normally connected in a bridge circuit, see Fig. 11.17. In this design, a sensor wire is heated electrically, and the pressure of the gas is determined by measuring the current needed to keep the wire at a constant temperature. The thermal conductivity of each gas is different, so the gauge has to be calibrated for the individual gas being measured. A Pirani gauge will not work to detect pressures above 1 mbar, because, above these pressures, the thermal conductivity of the gases no longer changes with pressure. The Pirani gauge is linear in the 10^{-2}

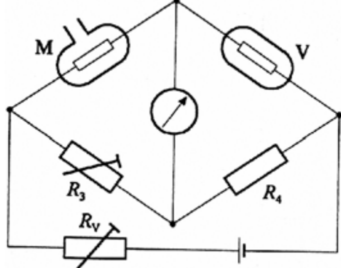


Fig. 11.17. Measurement circuit (Wheatstone bridge circuit) of a Pirani gauge

Fig. 11.18. Comparison of MicroPirani (*left*) and conventional Pirani gauges

to 10^{-4} mbar range. Above these pressures, the output is roughly logarithmic. Pirani gauges are inexpensive, convenient, and reasonably accurate. They are 2% accurate at the calibration point and 10% accurate over the operating range.

Thermistor gauge is a form of *thermal conductivity gauge* in which the temperature-sensitive elements are made of semiconducting material instead of metal. *Thermocouple gauge* is a *thermal conductivity gauge* which contains a heated filament and a thermocouple for measuring the filament temperature as a function of gas pressure.

The introduction of the solid state MicroPirani sensor (cf. Fig. 11.18) has established new performance standards for thermal conductivity gauges. The microprint is based on a 1×1 mm^2 silicon chip construction where thermal conductivity is measured in a very small cavity where gas is passed by diffusion only and not by flow. Compared to the well-known wire-based Pirani gauge, the solid state MicroPirani sensor has the advantages of higher accuracy, wider measuring range, insensitivity to mounting position, mechanical robustness, small dead volume and very fast response to pressure changes. The measuring range of the Microprint is extended 1–2 decades compared to traditionally Pirani gauges, and thereby can be used from atmospheric pressure down in the high vacuum range of 10^{-5} mbar. Measuring accuracy of 10%

is not unusual for a traditionally wire-based Pirani gauges. The high reproduction of the solid state MicroPirani sensor element provides a measuring accuracy of less than 5% of reading.

Capacitance Manometer Gauges

A capacitance manometer has a thin metal diaphragm separating an unknown pressure from a vacuum or known pressure (see Fig. 11.19). The electrical capacitance between the diaphragm and the fixed electrodes gives an accurate measurement of the distance between them. Since this distance is determined by the deflection of the diaphragm, which in turn is determined by the pressure of gas, the capacitance can be converted into a reading of the unknown pressure. Capacitance manometers measure pressure independently of gas or vapour composition. Because the pressure-deflection-capacitance is directly related, this type of gauge may be classified as direct reading.

Capacitance manometers are the most accurate devices for measuring the absolute pressure (with very high vacuum on the reference side; the tube 4 is closed) and differential pressure (with a known pressure on the reference side; tube 4 is open). They measure all gases and materials that are vapours at the gauge's operating temperature. Gauge heads are specified by their maximum pressure (atmospheric pressure down to 1×10^{-1} mbar) and have a dynamic range approximately of 10^4. Thin diaphragms can measure down to 10^{-5} mbar, while thicker diaphragms can measure in the low vacuum to atmospheric range. To cover a wide vacuum range, one can connect two or more capacitance sensing heads into a multi-range package.

The capacitance diaphragm gauge is widely used in the semiconductor industry, because its Inconel body and diaphragm are suitable for the corrosive services of this industry. They are also favored because of their high accuracy and immunity to

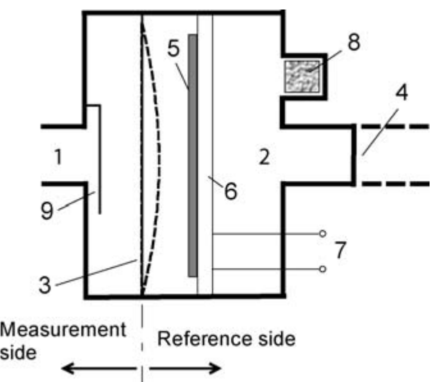


Fig. 11.19. Scheme and view of a capacitance manometer: *1* measured vacuum, *2* reference vacuum, *3* diaphragm, *4* tube, *5* electrodes, *6* electrodes holder, *7* electrical feed through, *8* chemical getter (NEG), *9* shield

contamination. Capacitance manometers commonly offer accuracies of 0.25%, while high-accuracy products can offer 0.08%.

Spinning Rotor Gauges (SRG)

In 1968 Fremerey [8] developed a gauge with a freely rotating steel ball for the measurement of the viscosity of the gas. According to Fig. 11.20, this steel ball was suspended by permanent magnets and magnetic coils whose excitation current was electronically adjusted. The ball was made to rotate by a special coil. After reaching a desired frequency the excitation was stopped, and the remanent magnetism of the ball induced an electric voltage in a coil, so that the frequency could be determined electronically at intervals of time.

The viscosity and the pressure could be calculated from the decrease of the frequency ν with time from

$$p = \frac{\pi \rho_K r \bar{v}}{10 \alpha t} \ln \frac{\nu(t)}{\nu(0)}, \tag{11.12}$$

where r is the radius of the ball, ρ_K its density, $\nu(t)$ the frequency of the ball at time t, $\nu(0)$ the frequency at the beginning of the measurement, α the accommodation coefficient and $\bar{v}$ is the arithmetic average velocity of gas particles. The advantage of this gauge is a simple construction of the sensor. It consists only of a flange connected with a cylindrical tube closed at the end opposite to the flange and containing the ball. The arrangement of the permanent magnets and the coils is a separate structure fixed on the tube. The lower pressure limit is 10^{-7} mbar or lower, the upper pressure limit is about 10^{-1} mbar. Due to the quantities α and ν being specific to the gas, the pressure reading depends on the composition of the gas. The precision of this gauge is so excellent that it can be used as a secondary standard for the calibration of other gauges.

Crossed-field Ion Gauges (Penning Gauges)

These gauges are based on the gas discharge. In the usual mechanism, an atom (in a gas molecule) is ionised in an electric field, the extracted electron and the ion travel towards opposite electrodes and their kinetic energy serves to ionise more atoms creating an avalanche discharge. Therefore, the gauge needs a high voltage of up to

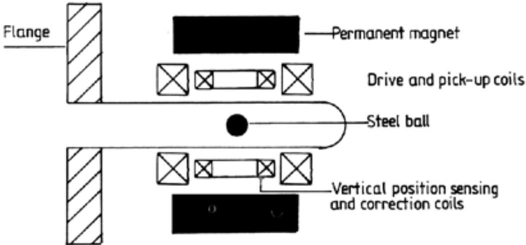


Fig. 11.20. Schematic diagram of a spinning rotor gauge (SRG)

2 kV to generate a discharge between cathode and anode (Fig. 11.21). If the mean free path of molecules is large, the chain reaction is not possible because of the lack of partners. In the Penning gauge, the flight path of electrons is increased artificially by a strong magnetic field perpendicular to the electric field. Heavier ions are deflected to a much lesser extent than the electrons and reach the cathode. The ion current across the electric field is proportional to the pressure (Fig. 11.22). The limits are 10^{-2} mbar where the normal gas discharge mechanism causes a much larger electron current superimposed on the ion current and 10^{-7} mbar where the mean free path becomes too long even for electrons forced on the magnetic detour.

In crossed-field gauges (Penning gauges, cf. Fig. 11.21 and Fig. 11.22), the ion current I_+ and pressure p relationship follows the equation

$$I_+ = K \cdot p^m, \tag{11.13}$$

where m depends on the type of gauge and varies between 1…1.4. However, m also depends on the pressure and this makes the situation rather complicated for reliable measurements. Generally, m is higher for lower pressures and may reach values up to 2. Therefore, if in gauge controllers the relation for higher p is extrapolated to

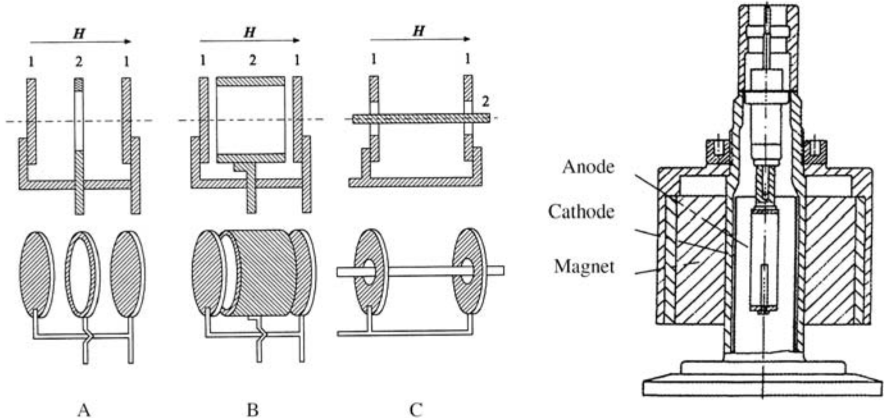


Fig. 11.21. Schematic draw of cathode–anode arrangements (*left*) and cross-section of a Penning gauge (*right*) [5]

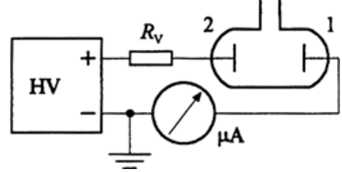


Fig. 11.22. Schema of Penning gauge (*1* cathode, *2* anode) in measurement circuit

very low pressures ($<10^{-7}$ Pa), the gauge will indicate lower pressures than actually present.

Emitting-Cathode Ionisation Gauges (Hot-Cathode Gauges)

The earliest ionisation vacuum gauges were simple triode vacuum tubes adapted from early radio vacuum tubes. In ionisation gauges, electrons emitted from a hot filament ionise the gases present in the tube, and these ions are collected on a collector to generate an electrical current proportional to pressure.

In the 1950s, Bayard and Alpert invented the inverted triode gauge, which minimized the effect of X-rays on the measurement and extended the measuring range down into the ultrahigh vacuum range ($<10^{-7}$ mbar). Driven by the needs of the aerospace market in the 1950s and 1960s, the Bayard–Alpert gauge became the workhorse and virtual standard for high vacuum measurements. By the 1980s, limitations of these traditional gauges were becoming an impediment to the new and rapidly growing semiconductor market (Fig. 11.23).

In emitting-cathode ionisation gauges, the ion current and pressure relationship follows the simplified equation

$$I_+ = C \cdot p \cdot I_-, \tag{11.14}$$

where I_- is the electron emission current. The gauge sensitivity C depends on the gas species. Attempts to correlate this gas specific sensitivity accurately with ionisation cross-sections fail due to other gas specific effects like ion capture probability, dissociation effects and secondary electron generation. Values for relative ionisation

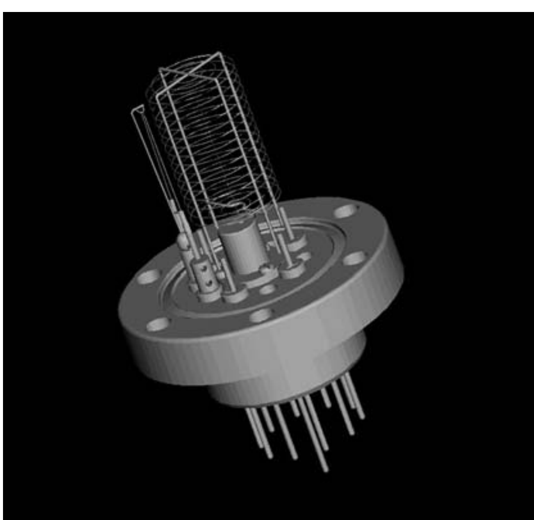


Fig. 11.23. Constructional drawing of hot-cathode ionization gauge, type IE414 (INFICON AG)

sensitivities (normalised for nitrogen = 1) presented in tables [9] can be applied with some confidence while going from one gas to another, but the level of accuracy is only 10 to 20%. If higher precision is required, the gauges must be calibrated individually and for the gas in use.

Combination of Vacuum Pressure Gauges

It is a long standing desire of vacuum system designers to be able to measure vacuum pressure over the full range which can be generated with today's vacuum technology with one sensor. Unfortunately, physical principles employed in today's vacuum gauges make this goal unattainable. To measure a pressure from atmosphere down to 10^{-10} mbar, one has to employ at least two different measurement principles. For example, one can employ a Pirani gauge from atmosphere down to 10^{-3} mbar and then switch over to an ionization gauge such as a Bayard–Alpert hot ionization gauge. Traditionally this required two different flanges on the vacuum chamber and two different vacuum gauge controllers. Recently, controllers have been integrated with gauges, so that the gauges could be directly connected to a data acquisition system. These integrated systems are called active gauges. Further, combination gauges have appeared on the market, which combine different measurement technologies in one flange attachment. A combination gauge saves cost for the original equipment manufacturer since it reduces system requirements, such as the number of flanges

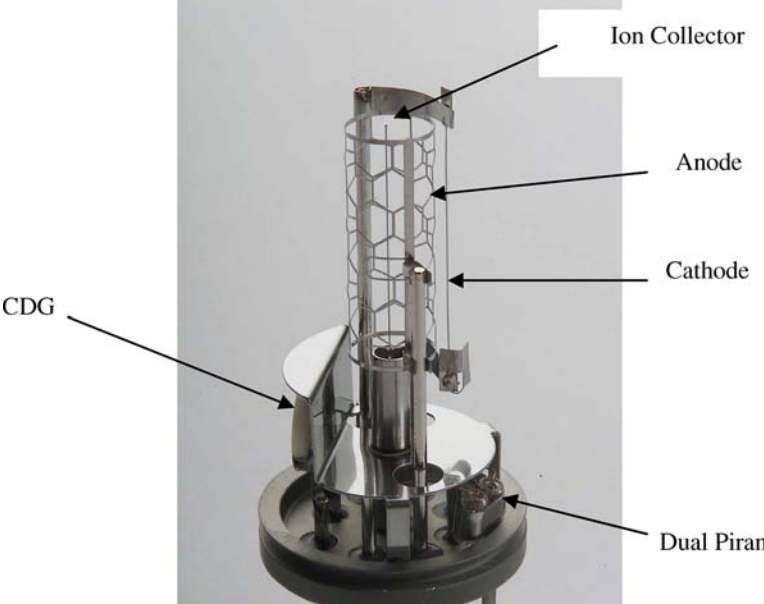


Fig. 11.24. The TripleGauge™ sensor is a combination of three different sensors: a Bayard–Alpert, a Pirani and a capacitance diaphragm gauge (CDG)

and cables required together with data acquisition system coding and test time of the system. Previous wide range combination gauges have limitations. One limitation is that Pirani gauges above 10 mbar measure a pressure which depends on the gas composition. This is particularly annoying during venting. If a heavy gas, e.g. argon, is used to vent to atmosphere, the Pirani gauge will never display atmospheric pressure despite having already reached a pressure above atmospheric pressure. This can lead to dangerous situations during venting process and is therefore a safety hazard. If venting is done with a light gas, e.g. helium, the gauge will indicate a pressure above atmosphere, even though the actual pressure is still much below atmospheric pressure. As a consequence, the vacuum chamber door can not be opened.

For example, the TripleGauge™ sensor presented here is an active combination gauge eliminating the limitations of previous generations of combination gauges. The TripleGauge™ sensor has a large measurement range from 5×10^{-10} to 1500 mbar (13 decades). It combines three different measurement technologies in one housing:

1. gas-type independent capacitance diaphragm measurement (CDG) for low vacuum (1 mbar to 1500 mbar),
2. heat loss measurement according to Pirani for medium vacuum (5×10^{-3}–10 mbar), and
3. Bayard–Alpert hot cathode ionisation measurement for high vacuum (5×10^{-10}–3×10^{-2} mbar). The TripleGauge™ sensor head is depicted in Fig. 11.24.

The CDG provides gas type independent pressure measurement above 10 mbar, ideal for load lock applications. The Pirani protects the hot filament of the Bayard–Alpert gauge from premature burnout and bridges the pressure range between the Bayard–Alpert and the CDG measurement range.

In the range between 1 and 10 mbar, the output is a mixed signal made from CDG and Pirani signals. Based on the Pirani values, the algorithm decides when it is save to turn on the ionization gauge. In the range from 5×10^{-3} to 5×10^{-2} mbar, the output is a mixed signal made from Bayard–Alpert and Pirani (see Fig. 11.25).

The TripleGauge™ sensor can be used for pressure measurement in semiconductor process, transfer and loadlock chambers, in industrial coating, and also in general vacuum measurement and control systems in the low to ultra-high vacuum range. Its particular advantages are:

- TripleGauge™ sensor (e.g. Typ BCG450) saves cost and tool space and reduces the complexity of vacuum measurement installation and setup,
- Gas-type-independent pressure measurement above 10 mbar provides more reliable loadlock control for any gas-mixture,
- Pirani interlock protects the hot filament from premature burnout,
- Automatic high vacuum and atmospheric Pirani adjustment reduces operator interventions,
- Differential pressure measurement at atmosphere eliminates uncertainty related to atmospheric pressure changes,

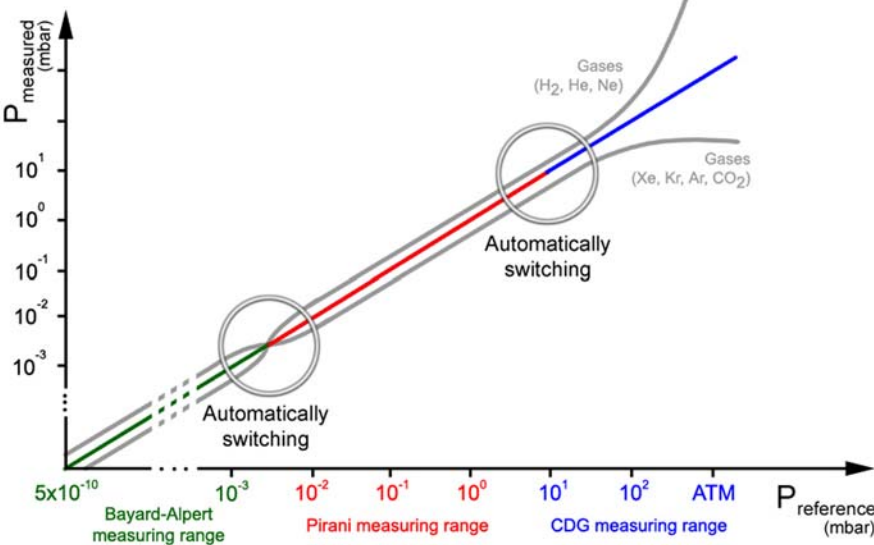


Fig. 11.25. Switching ranges of a TripleGauge^{TM} sensor (INFICON AG)

- Easy-to-exchange sensing element with on-board calibration data guarantees reproducibility,
- Optional graphic display and Fieldbus interfaces available.

So far, the TripleGauge^{TM} sensor comes closest to the ideal vacuum gauge which can measure any vacuum pressure with a single gauge independently of the gas type.

11.4.3 Partial Pressure Gauges

There are a number of different types of mass filter – mass spectrometer – that have been considered for the analysis of the gases found in vacuum systems. All are based upon the hot-cathode ionisation gauge, with the addition of some form of mass filter placed between the ionisation chamber and the ion collector. The pressure range to be considered is from approximately 10^{-4} mbar down to ultra-high vacuum. The four instruments that have been responsible for the major share of the development are:

- the magnetic deflection mass spectrometer,
- the Omegatron,
- the time-of-flight mass spectrometer,
- the quadrupole mass filter.

All have the advantages and disadvantages of hot-cathode devices (e.g. problems due to filament outgassing) and all operate over roughly the same pressure range. The quadrupole mass spectrometer is now well established as the instrument used almost universally for partial pressure measurement (cf. Fig. 11.26). This spectrometer can be considered as a hot-cathode extractor ionisation gauge with the addition

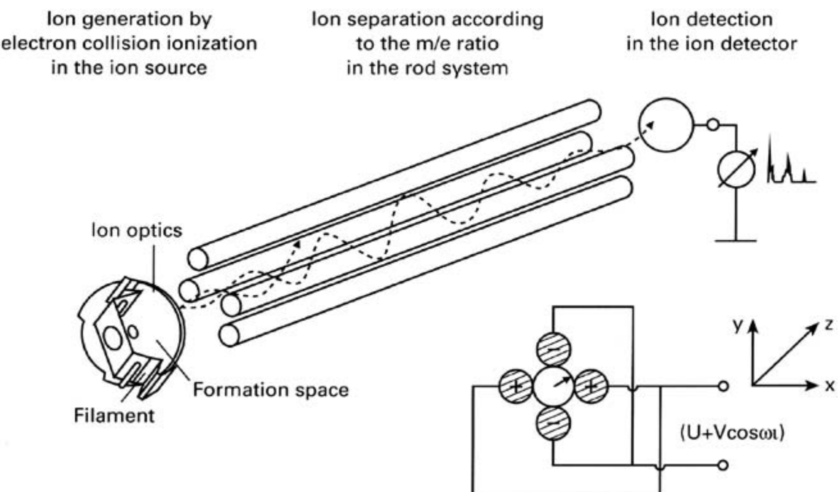


Fig. 11.26. Overall scheme of the quadrupole mass spectrometer

of an electrostatic mass filter between the ion source and the ion collector. This filter consists of a square array of circular rods approximately 6 mm diameter and between 100 to 150 mm length. Positive ions are injected parallel to the central axis towards the ion collector. A filtering action is achieved by applying combined radio frequency and dc potentials between opposite pairs of rods. The characteristics of the filter can be controlled by small changes in the relative values of the radio frequency and dc voltages and the mass range by changes in the frequency of the ac signal. Care must always be taken when operating this mass spectrometer in an analytical mode because of non-linearities caused by space charge and other effects which are important, particularly in the injection region between the ion source and filter. Also the filter is subject to loss of sensitivity over the long term, generally caused by contamination from hydrocarbon impurities in the vacuum system. This is not normally a problem in ultrahigh vacuum operation. Although almost all filters have been designed to operate with relatively low voltages applied to the rods, alternative modes are being considered which require distinctly high voltages or lower frequencies.

The determination of gas composition is also very important, and is typically done with a compact mass spectrometer known as a residual gas analyser (RGA). This produces a characteristic mass spectrum (see, e.g. Fig. 11.27). Most of the less expensive RGA's are based on a Quadrupole Mass Spectrometer (QMS), whose principle is explained in Fig. 11.26. Higher mass resolution is obtained in more specialized Magnetic Sector instruments, which are typically attached to specialist facilities for, e.g. cluster research or atom probe microanalysis.

Display and Interpretation of Spectra

Analog mode (cf. Fig. 11.27) provides a line graph representation of the acquired mass spectrum (partial pressure vs mass number). The span, resolution and noise

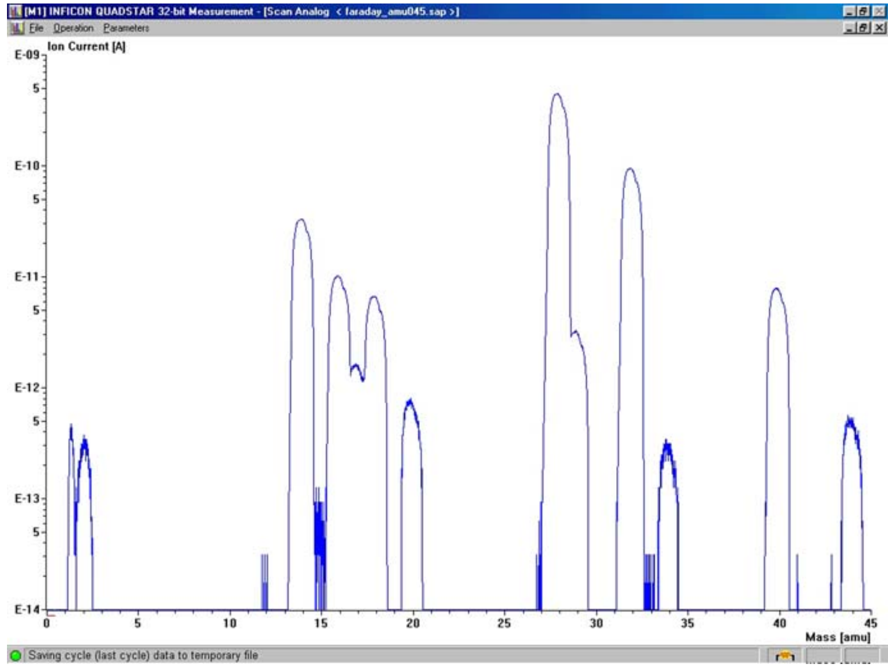


Fig. 11.27. Display view of a QMS: analog mode

floor can each be set. Scans can be single-shot, timed or taken continuously. Total pressure is available in analogue and most other modes of operation.

Pressure vs time (cf. Fig. 11.28) presents a strip chart of partial pressures for selected masses and provides a complete time history of data. Complete scrolling and zoom control is available even while data is being acquired. This mode is most often used for monitoring process trends.

Library mode contains a comprehensive list of gases that can be used to compare against the current spectrum. A search mode allows the operator to select up to some important masses and identify and display (numerically and graphically) the intensity of all gases that contain these masses.

11.5 Vacuum Materials and Components

11.5.1 Material Properties for Vacuum Applications

Selection Criteria

The selection of materials for use in vacuum systems is a very important part of the design and should be considered in view of application. Not only must the material be capable of being fabricated into the required components but it must stand up

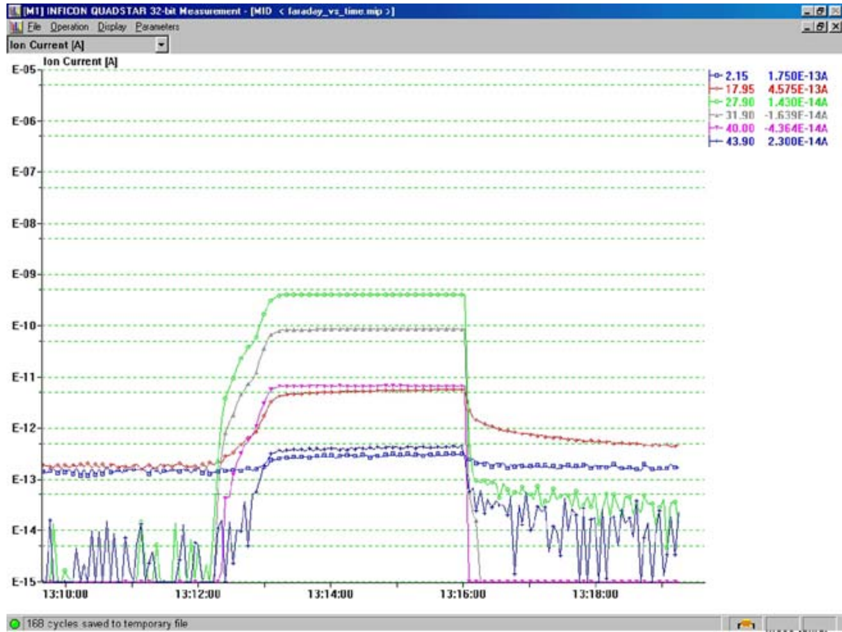


Fig. 11.28. Display view of a QMS: partial pressures of different gases vs time

to the environmental conditions of temperature, pressure, and chemical attack, etc. imposed on it by the vacuum processes, without limiting the attainable pressure that is required.

Material Property Requirements

1. Mechanical Properties: The material must be capable of being machined and fabricated. It must have adequate strength at maximum and minimum temperatures to be encountered, and must retain it's elastic, plastic, and/or fluid properties over the expected temperature range.
2. Thermal Properties: The material's vapour pressure must remain low at the highest temperature. Thermal expansion of adjacent materials must be taken into account, especially at joints.
3. Gas Loading: Materials must not be porous. Materials must be free of cracks and crevices which can trap cleaning solvents and become a source of virtual leaks later on. Surface and bulk desorption rates must be acceptable at extremes of temperature and radiation.

Commonly Used Materials

1. Metals for the vacuum components:
   Austenitic Stainless Steel is the most commonly used metal for high and ultrahigh vacuum systems, since it fulfills all of the requirements above. Stainless

steel material 1.4306 (U.S. 304) are chosen most frequently for satisfactory argon-arc welding. Stainless steel is relatively economical, has acceptable outgassing rates, and can be fabricated easily.

Aluminum and Aluminum Alloys are very cheap, easy to machine, and have a low outgassing rate as long as the alloy does not have a high zinc content. They have the disadvantage of low strength at high temperatures and high distortion when welding. Alloys with copper content also present welding problems. Aluminum that will be exposed to vacuum should never be anodized due to serious outgassing problems. There are also some potentially violent chemical reactions that can develop when cleaning freshly machined Aluminum with Trichloroethane or Trichloroethylene based vapour degreasers.

Mild Steel may be used down to about 10^{-3} mbar or lower if plated. High permeability to hydrogen and possible rusting make this material unsuitable for ultra-high vacuum applications.

Oxygen Free High Conductivity (OFHC) Copper is easily machined with good corrosion resistance and is widely used for vacuum applications. It is not generally used for vacuum envelopes that require baking due to possible heavy oxidation, scaling, and the difficulty of brazing in a hydrogen atmosphere.

2. Metals used in demountable seals:

   Copper Rings are commonly used for high and ultra-high vacuum applications. Conflat flanges (CF) use a copper ring compressed between two knife edges, are bakeable to 450°C, and are widely used.

   Aluminum Wire Rings are very cheap and bakeable to 200°C.

   Indium Wire can be used between flat flanges. It is very soft and continues to flow after initial tightening.

   Gold wire is often used for ultra-high vacuum seals between flat surfaces and can be baked to 450°C. Gold is somewhat easier to recycle than Indium, offsetting it's high initial cost.

   Swagelok, *Tylok*, *and other Proprietary Brand Seals* are now widely used. The manufacturer's directions for each brand of seal must be followed to the letter. Never mix one brand of seal with another – mating surfaces will most likely be cut at different angles and they will not be compatible. Mating components of the various seals should also be of the same metal, e.g. don't use brass ferrules on stainless steel fittings and vice versa.
3. Plastics: Generally, the use of plastics should be kept to a minimum due to their high gas permeability and high desorption rates compared with metals, glass and ceramics. In spite of this, plastics are often used in low vacuum systems because of their insulating properties, elasticity, and price.
4. Ceramics: Fully vitrified electrical porcelain and vitrified alumina are excellent insulator that have a low outgassing rate, low gas permeability, and can be used to 1500°C. There are also some machinable ceramics available. All ceramics are brittle and must be handled with care.
5. Glass: Borosilicate Glass is often used for small systems and viewing windows. Glass can be obtained as components from stock, is easy to fabricate into components, and has high corrosion resistance.

Table 11.2. Outgassing rates of different materials [9]

Material: Various steels	Outgassing rate[a]	Material: Nonferrous metals	Outgassing rate[a]
Mild (ingot) steel	540	Aluminium[b]	6.3
Mild steel, slightly rusty	600	Aluminium, treated	
Steel, de scaled	307	in various ways	4.1 … 6.6
Steel, chromium-plated[b]	7.1	Duraluminium	170
Steel, chromium-plated[c]	9.1	Gold, wire[b]	15.8
Steel, nickel-plated[b]	4.2	Copper[b]	40
Steel, nickel-plated	2.8	Copper[c]	3.5
Steel, niclad[b]	8.3	Copper, OFHC	18.8
Stainless steel	90 … 175	Copper, OFHC[c]	1.9
Stainless steel[b]	13.5	Brass	400
Stainless steel, blast-cleaned	8.3	Molybdenum	5.2
Stainless steel[c]	1.7	Titanium	4 … 11.3
Stainless steel, electropolished	4.3	Zinc	220

[a] All data in 10^{-9} mbar s^{-1} cm^{-2}
[b] As produced
[c] Mechanically polished

6. Glass to Metal and Glass to Ceramic Seals: None of the above are suitable for glass to metal seals, which are usually made of *Inconel* or *Kovar*. These materials are very difficult to machine, very expensive, and best bought as parts of components from vacuum equipment suppliers.

Materials that should not be used in vacuum: *Cadmium Plating*, often used for small screws, must be avoided for vacuum applications due to some really nasty outgassing problems. Small screws used in the chamber can be nickel-plated brass or plain copper, and should be drilled or relieved (i.e. file off one side of the threaded area) to prevent virtual leaks. *PVC Insulated wire* generally has a high outgassing rate and should be replaced with Teflon insulated wire. *Paint* is a common source of outgassing and virtual leaks. In general, the choice of metals becomes more and more limited the lower the pressure in the vacuum system is required to be. For UHV chambers only stainless steel, copper and gold are appropriate. Table 11.2 shows the outgassing rates of different materials related to the surface area (desorbed molecule flow rate density) after a pumping time of one hour. The preceding material overview is by no means complete, but it's a basic orientation. More details are in [2–6].

11.5.2 Component Standards and Constructions

Pumps, chambers, and gauges are joined by flanges, tubes, bellows, and valves, all of which must be vacuum tight, made with low partial pressure materials, and have

Table 11.3. Selected vacuum components with different flanges and corresponding photos, drawings and symbols

Vacuum component	Photo view	Constructional drawing	Graphic symbol[a]
Adapter Nipple CF/ISO-K			
Adapter Nipple CF/KF			
T-Pieces, DN 40 CF			
4-Way Cross, DN 40 CF, 2 flanges rotable			

[a] All standard graphic symbols can be found in [10]

clean, smooth surfaces. Each is tightened with seals made of elastic and deformable material, whose choice is governed by partial pressure and permeation.

In medium- to high-vacuum systems the O-ring is the common seal, which has elastic properties and fills small irregularities. O-rings are used in systems whose ultimate pressure is higher than 10^{-7} mbar. O-ring permeation is larger ($\approx 10^{-8}$ mbar l/s) because of voids between the polymer chains; and thus for UHV systems aluminium, indium, or gold ring seals, or standardized flanges ("conflat"; CF flanges) with copper rings are used (cf. Table 11.3).

11.5.3 Cleaning Procedures

Experience shows that in order to achieve all but the most modest levels of vacuum, it is necessary to clean vacuum vessels and components in some way. This is because

Table 11.4. Simple cleaning procedures

Pressure range	Procedures
1 bar	Hot water wash
	Swab or wash with detergent
1 mbar	Wash in hot aqueous detergent
10^{-6} mbar	Ultrasonic wash in aqueous detergent
	Rinse with hot pure water
10^{-9} mbar	Ultrasonic wash in hot solvent
	Rinse with hot pure water
	Vacuum bake at 250°C

as far as vacuum is concerned, the world is a dirty place! In general, vessels and components will have been machined, worked in some way or another, or handled. Such processes may use greases or oils which have high outgassing rates or vapour pressures and will remain on or in the surface. Marking pens and adhesive tapes leave residues on surfaces that can also enhance outgassing. Water, solvents and other liquids can remain embedded in cracks or pores in a surface and can outgas over long periods. All such contamination can limit the base pressures attainable in a system.

In general terms as far as vacuum is concerned, we define contamination as anything which

- prevents the vacuum system from reaching the desired base pressure,
- introduces an unwanted or detrimental species into the residual gas, and
- modifies the surface properties of all or part of the vacuum system in an undesired way.

For example, at room temperature, a pool of liquid mercury in a vacuum system will limit the base pressure to about 2×10^{-3} mbar, its vapour pressure. If the system were required to operate at, say 10^{-5} mbar, the mercury would be a contaminant.

There is no one cleaning process which is "right" for all vacuum systems, vessels or components. Some of the things which will need to be taken into account are as follows:

- the level of vacuum required (rough, high, UHV, etc.),
- if there is a particular performance requirement (e.g. low desorption),
- whether there is a particular contaminant (e.g. hydrocarbons) whose partial pressure must be minimized,
- what materials the items are made from,
- how the items are constructed,
- safety,
- cost.

Table 11.4 illustrates in a very schematic fashion some typical common cleaning processes which would be used in sequence. To use this table, select the approximate pressure required and then apply all the processes in order from the top to the level corresponding to the pressure required. It should be noted that there is a lot of different "recipes" in the literature and in the folklore of vacuum.

11.6 Leak and Leak Detection

11.6.1 Introduction

Leak detection is one of the most important but sometimes one of most tedious and frustrating aspects of vacuum technology. No vacuum apparatus or plant is absolutely vacuum-tight and, in principle, does not need to be. Important is that the leak rate is comparatively small and so does not influence the required working pressure, gas content and ultimate pressure in the evacuated vacuum system. Hence it follows that the requirement with regard to the tigthness of a vacuum system is stricter the lower the required pressure.

In order to comprehend leaks or leakages quantitatively, the physical quantity of leak rate has been introducted. Its symbol is q_{L} and its unit in vacuum technology mbar l s^{-1}. This means that the leak rate q_{L} amounts to $q_{\mathrm{L}} = 1$ mbar l s^{-1} if in a closed vacuum system of 1 l volume the pressure increased during one second by 1 mbar.

11.6.2 Leak Testing Methods

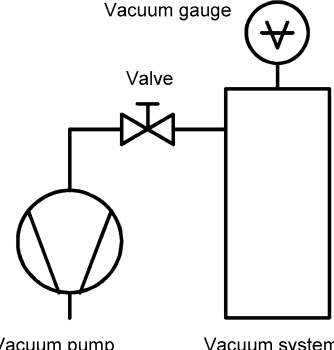
There are several leak detection techniques to choose from, the following being especially suitable for use in vacuum systems for production of vacuum electronic devices:

- measuring pressure rise in vacuum systems with vacuum gauge,
- measuring rest gas composition in vacuum system with mass spectrometer,
- specific gas detection testing with helium leak detectors (see next section).

The method of testing a vacuum system depends on the size of the leakage, on the tightness aimed at and also on whether the apparatus is of metal or glass.

Before beginning the actual leak searching, information on the size of the leakage present should be obtained. This is given by the *pressure rise method*. For this, the valve on the pump side of the vacuum system is closed (Fig. 11.29, left). Then the time is measured during which the pressure rises by a given amount (about ten times). The valve is open again and pumping continued for some time, then the procedure is repeated. If the time for the pressure rise remains constant, a real leak is present, provided that the waiting time between the separate pressure-rise experiments was sufficiently long. If the pressure rise becomes smaller, it is probably due to gas evolution from the inner surface. From the course of the pressure-rise curve one can attempt to distinguish between leakage and contamination. On a linear scale, the pressure-rise curve remains linear, even for high pressures. If the pressure rise is due to gas evolution from walls, the pressure rise becomes gradually smaller and tends to a limiting value. Usually, both cases arise simultaneously, so that a separation between the two sources is often difficult. The relations are sketched in Fig. 11.29 (right).

If the rise in pressure is due to a genuine leak and if system volume is known, then the leak rate q_{L} can be determined quantitatively on the basis of the pressure rise through time using (11.15).



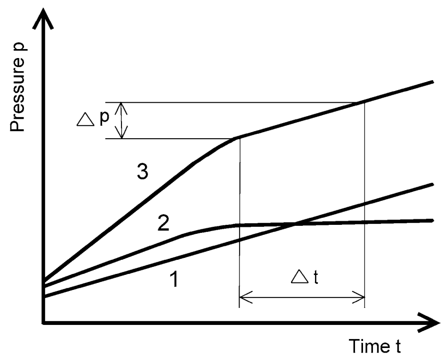


Fig. 11.29. Vacuum scheme (*left*) and pressure rise through time in a vacuum system without the pump connection (valve closed): *1* pressure rise due to a leak; *2* pressure rise due to other phenomena (evaporation, outgassing, desorption); *3* superposition of *1* and *2*

$$q_{\mathrm{L}} = \frac{\Delta p}{\Delta t} V. \tag{11.15}$$

Generally, in high vacuum systems the following values apply:

- total leakage $q_{\mathrm{L}} \leq 10^{-6}$ mbar l s^{-1}: vacuum system is very leak-tight,
- total leakage $q_{\mathrm{L}} \approx 10^{-5}$ mbar l s^{-1}: vacuum system is adequately leak-tight,
- total leakage $q_{\mathrm{L}} \geq 10^{-5}$ mbar l s^{-1}: vacuum system not leak-tight.

Hence, from (11.15) the end pressure p_{end} of vacuum system is given by

$$p_{\mathrm{end}} = q_{\mathrm{L}} \cdot S_{\mathrm{eff}}^{-1}. \tag{11.16}$$

Naturally, an improvement of this pressure, in case it is unsatisfactory, can be achieved by the use of a larger pump with higher effective pumping speed S_{eff}. A desired increase in S_{eff} depends in practice on economic and constructional considerations. It is easier to find and then to delete the real leak.

11.6.3 Helium Leak Detectors

The most sensitive and reliable leak detectors are mass spectrometer helium leak detectors. The mass spectrometer (see Fig. 11.30) consists of the ion source (1 to 5) and the deflection system (6 to 11). In the ion source neutral gas molecules become ionized through electron impact. The ion beam passes through the slit (5) and enters the magnetic field (7) with a definite energy. In the magnetic field the ions are deflected in circular path, the radius being small for small masses ($M < 4$) and greater for larger mass ($M > 4$). By correct adjustment of accelerating voltage and magnetic field strength, the ions of mass 4 (helium) describe a quarter circle of defined radius so that they pass through the diaphragm (6) and thus arrive at the aperture of the electronic lens (8).

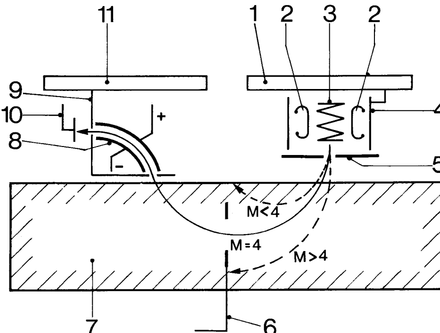


Fig. 11.30. Schematic diagram of a double focusing mass spectrometer as used in helium leak detector. Legend: *1* flange to the ion source unit, *2* cathodes, *3* anode (heated), *4* shield tube, *5* diaphragm (also used as ion collector for measuring total pressure), *6* intermediate diaphragm , *7* magnetic field, *8* cylindrical condenser, *9* diaphragm, *10* ion collector, *11* flange of the deflection unit

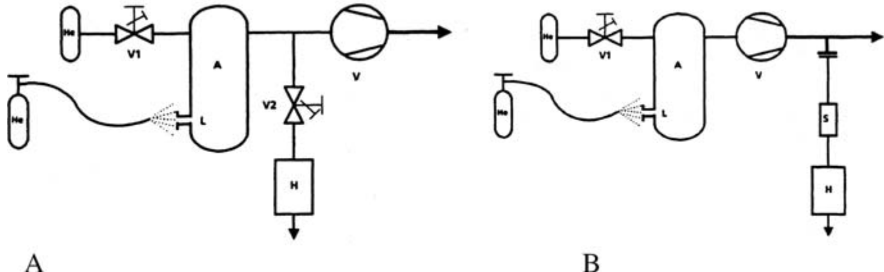


Fig. 11.31. Schematic diagram of vacuum leak testing using a helium leak detector in partial-flow operation (**A**) and using a helium sniffer (**B**)

Vacuum Leak Testing Using Helium Leak Detectors

Figure 11.31(A) shows the setup for a vacuum system to which a helium leak detector is connected for testing in partial-flow operation. A metering valve is used to limit the sub-stream directed to the leak detector to the maximum permissible flow of test gas for the leak detection process. Leak testing is undertaken in two phases:

1. *Calibration*: A "test leak" with a known leakage rate is used to determine both the detector's response threshold and the maximum response time for the overall system – the helium leak detector and the vacuum system itself. The test leak should be installed as far as possible from the detector.
2. *Measurement*: Helium is sprayed from the outside around any suspicious points on the vacuum container. If a leak is present, helium will enter the system and will be sensed by the detector.

Vacuum Leak Testing Using a Helium Sniffer

When operating in the low vacuum range, it is possible to conduct leak tests to locate larger leaks without having to make any connection to the vacuum circuit and without having to interrupt production. In this case a helium leak detector with a helium "sniffer" is used, the probe tip of which submerged in the stream of discharged gas. The "sniffer" used in leak detection is a probe which draws in small quantities of gas and forwards it to the mass spectrometer of a helium leak detector to check for traces of helium. As in partial flow vacuum testing, helium is sprayed around those points which are suspected to be leaking, cf. Fig. 11.31(B). If helium enters the system it will be moved to the process vacuum pump and will appear in the discharge line. A helium concentration of just 5 ppm in the exhaust gas will trigger a clear leak indication. The presence of a liquid product poses no problem; it will be penetrated by the helium.

11.6.4 Leak Testing of Vacuum Microelectronic Devices

A pressure/vacuum method of leak detection is used for testing hermetically sealed devices containing a small cavity which evacuated (or may be gas-filled). The devices to be tested (e.g. X-ray tubes, field emission displays and other vacuum microelectronic devices) get in a pressure chamber exposed to the search gas, preferably helium. During an exposure time of up to several hours at a high search-gas pressure (5 to 10 bar), search gas will penetrate through any leaks present into the test object. This part of the process is referred to as "bombing". After such pressurizing or "bombing" the objects are tested for search gas emission in a vacuum chamber, following the same procedure as in the vacuum hood test described above. Test objects having a coarse leak may loose their inner search gas concentration during pump down of the vacuum chamber. This is why in many cases a bubble leak test (leaks are indicated by the appearance of bubbles in the liquid) is made prior to the test in the vacuum chamber. This method permits the detection on the smallest leak rates and is preferably used for industrial scale automatic leak testing, especially of components, which cannot be gas-filled by other means.

11.7 Examples of Vacuum Systems

11.7.1 UHV System for Field Emission Characterisation

For field emission investigations a UHV setup was developed [11]. The vacuum components of the main system (field-emission measurement chamber) are stainless steel components with standard DN40 CF (Conflat) flanges. In the vacuum chamber a base pressure in the 10^{-8} mbar range (without baking) is reached us-

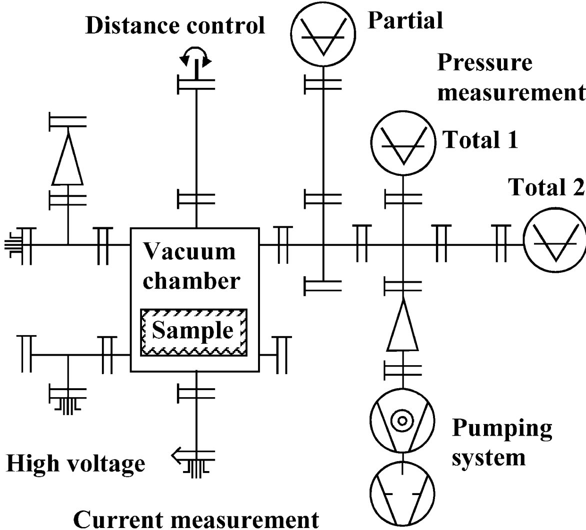


Fig. 11.32. Experimental ultrahigh vacuum (UHV) setup for field-emission investigations

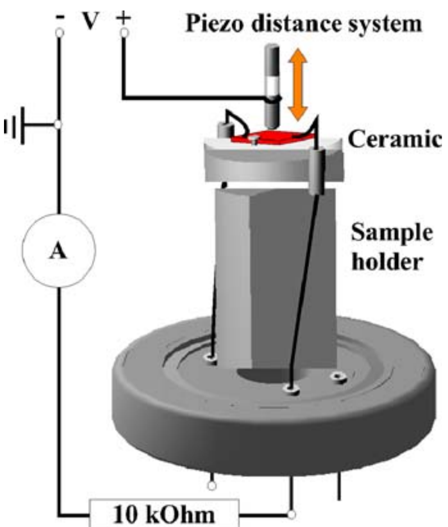


Fig. 11.33. Experimental set-up for cathode–anode distance control (diode system)

ing a turbomolecular pumping system consisting of a turbomolecular drag pump and diaphragm pump (Fig. 11.32). The total pressure in the main chamber is determined by a Bayard–Alpert ionization gauge, while the partial pressures were determined by a quadrupole mass spectrometer (QMS) of the residual gas analyzer type.

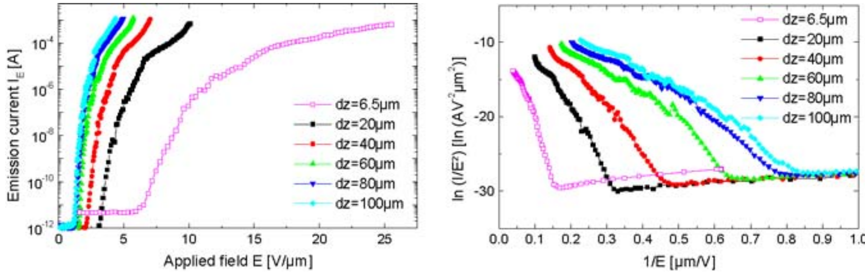


Fig. 11.34. Current–voltage curves and Fowler–Nordheim plots for electron field emission of CNT buckypaper showing dependence on emitter–anode distance (hemispheric anode of stainless steel) in diode structure

A second total pressure measuring system makes possible a full range vacuum pressure control. The gas inlet system used for the investigations of field emission and residual gas interactions is differentially pumped by a rotary vane pump. With the help of a piezo-element system with a distance controller the anode–cathode (emitter) distance can be adjusted in the range of 1…8 μm. The setup is shown schematically in Fig. 11.33. Furthermore, a precise UHV manipulator or special spacer are used to realize distances over 8 μm (e.g. 20, 40, 100 μm).

To investigate the emitter properties of Carbon Buckypaper, the emission current is measured between emitter (cathode) and ground potential. The measuring procedures are carried out as follows. The amplitude of the applied voltage (between anode and ground potential) is increased incrementally in 1 V steps (every second), starting from ground potential up to a programmed maximal value. The field emission measurement is controlled by a real-time graphical display with the opportunity of operator interventions. Nevertheless, a constant anode voltage is applied for long-term field emission stability measurement (Fig. 11.34).

Automated data acquisition and control are accomplished with LabView data acquisition software using the IEEE-488 interface bus. Important parameters are vacuum pressure, gas flow and composition, and emission parameters. Additional a digital video camera with a framing time of 40 ms is used to record all optical activity (luminescence and discharge) in the emitter–anode distance. An oscilloscope is used to monitor the time-profile of the current and voltage during micro-discharges.

References

[1] C. Schott, *Technica Curiosa* (Archiv Otto-von-Guericke-Gesellschaft Magdeburg e.V., Nürnberg, 1664)

[2] A. Chambers, R.K. Fitch, B.S. Halliday, *Basic Vacuum Technology* (Adam Hilger, Bristol, Philadelphia, New York, 1991)

[3] A. Roth, *Vacuum Technology* (Elsevier, Amsterdam, 1990)

[4] A. Wutz, *Handbuch Vakuumtechnik* (Vieweg, Wiesbaden, 2004) (in German)

[5] C. Edelmann, *Vakuumphysik* (Spektrum, Heidelberg, 1998) (in German)
[6] R.A. Haefer, *Kryo-Vakuumtechnik* (Springer, Berlin, Heidelberg, New York, 1981) (in German)
[7] J.H. Leck, *Total and Partial Pressure Measurement in Vacuum Systems* (Blackie, Glasgow London, 1989)
[8] J.K. Fremerey, High vacuum friction manometer, J. Vac. Sci. Technol. **9**, 108–119 (1972)
[9] R.L. Summers, NASA Tech. Note. NASA TN D-5285 (1969)
[10] Standard DIN 28401 (in German)
[11] W. Knapp, D. Schleussner, Field-emission characteristics of carbon buckypaper, J. Vac. Sci. Technol. B **21**, 557–561 (2003)

Index

3D analysis, *see also* three-dimensional analysis 235
3D rapid prototyping 245
β-function 373
π-mode 56, 362

A
Abbe's law 106
aberration 159, 214
  coefficient 160
absolute pressure 500
absorber 401
absorption 308, 345, 487
  coefficient 306, 346
  cross-section 306
AC memory type dot matrix display 112
accelerating voltage 239
accelerator 232
  collider 365
  structure 362
  vacuum system 357, 387, 394, 403
accelerator mass spectrometry (AMS) 236
acrylic layer 108
active cathode surface 446
active gauge 504
Adler relationship 5
adsorption 487, 494
alkali based emitter 449
alkali based photo cathode 450
alkali metal dispensers (AMD) 469
alkaline earth metals 478
alkaline earth oxides 477
alternating gradient focusing 373
amplifier 1, 6, 46
  channel 132, 153
  harmonics 52
  large gain 3
  three stage image 139
amplitron 59
analogue display 91
analogue grey scale generation 95
anisotropic astigmatism 159
anisotropic coma 159
anisotropic distortion 159
anode 5, 11, 43, 55, 57, 60, 118, 156, 164, 409, 429, 433, 475, 493, 502, 516, 518
  current 5
  dark current 133
  efficiency 286
  potential 60
  surface 413
  voltage 11, 102, 118, 433, 461, 463
anodic bonding 470, 473
anticorona ring 19
aperture 157, 217
  lens 164
Applegate diagram 14, 38
arc
  constriction 411
  movement 409, 410
  stability 409
  welding 174
aspherically polished X-ray mirror 348
astigmatism 157, 162, 206
atomic absorption 338
atomic absorption spectroscopy 338
Auger
  effect 343
  electron 207
  electron emission 189
avalanche cold cathode (ACC) 450, 463
axial aberration 191
axial astigmatism 159, 214
axial chromatic aberration 191, 214
axial electron gun 174
axial gun 181

axial magnetic field (AMF) 411, 417, 419, 420
  arc 413
  contact 411, 412, 419, 426
axial output coupling 71
azimuthal electron bunching 62
azimuthal phase bunching 64

B
Ba based thermionic cathode 449
Ba getter 492
Ba-dispenser cathode, *see also* I cathode 439
back-scattering coefficient 172
backscattered electron detection 174
backscattered electrons (BSE) 206
backward wave amplifier 52
backward wave oscillator (BWO), *see also* carcinotron 2, 52
  M-type BWO (MBWO) 2, 53, 54, 60
  O-type BWO 53, 54
ballistic electron emission 465
bandwidth 1, 6, 12, 22
BaO evaporation 446
barrier junction target 147
Bayard–Alpert gauge 394
beam
  aperture 172
  deflecting system 143, 145
  deflection 174, 181, 200
  density 364
  diameter 239
  dynamics 369
  efficiency 43, 44
  emittance 369
  focusing system 143, 145
  lifetime 379, 380
  loading conductance 16
  power efficiency 29, 43, 295
  screen 400
  transport 250, 254
    self-pinched 259
  vacuum chamber 401
  vacuum system 393, 397
beam forming region (BFR) 107
beam-gas lifetime 383
bending magnet 401
Bernas source 248
betatron 363
  accelerator 364
  tune 374
bipotential electron gun 102
bipotential lens 102
black body radiation 303, 335
blended-light-lamp 333
bone densitometry 349
borosilicate glass 510
breaking capacity 414
Bremsstrahlung 179, 344
  interaction 381
Brewster angle windows 72, 310, 311
brightness 168, 170
  efficiency 107
brilliance 369
Brillouin flow 32
bulk desorption rate 509
buried layers 252
Busch theorem 8

C
candescent emitter 458
capacitance manometer 500
  gauge 500
carbon nanotube (CNT) 125, 451, 456, 461
  emitter 277
  FED 462
carcinotron 3, 60
cathode 20, 30, 55, 57, 118, 122, 155, 164, 430, 493, 502, 506, 516, 519
  auxiliary 167
  boride 446
  borided tungsten 438
  cold 4, 58, 60, 82, 451
  cold field emission 454
  emission 9
  hot 24, 503, 506
  lanthanum hexaboride 448
  long-life 436
  material 456, 461, 493
  nanotube 463
  ray 184, 429
    oscilloscope 155
  rod 166
  surface 10, 26, 29, 33, 436, 477
  thermal 168, 171
  virtual 433
cathode ray tube (CRT), *see also* electron beam tube 85, 86, 100, 492
  monitor 91

ceramic window 35
cesium 272, 278
  ions 272
CF flange 512
channel implants 249
channel multiplier 136
channel plate 136, 137, 140
channel secondary electron multiplier 134, 135
channeling 246, 247, 250, 252
character display 88
Chebyshev step transformer 35
chemically amplified resists (CAR) 193
Cherenkov radiation 315
Child–Langmuir law 11
chopping current 414, 416
chromatic aberration 159, 160, 210
chromatic aberration coefficient 166
circuit editing (CE) 237, 243
clean metal work function 434
coaxial electrode 55
coaxial magnetron 3
coaxial output transition 35
Cockroft–Walton generator 357
coefficient of spherical aberration 214
coherent radiation 63, 357
cold bore vacuum system 393, 399
cold field emitters 168
collector 3, 19, 24, 60, 506
  efficiency 39, 41, 43, 44
collider 370
colloid thruster 272
colour 96
  capability 95
  rendition 327, 331
coma 162
combination gauge 504
compression ratio 496
Compton effect 346
Compton scattering 384
computer-controlled vacuum interrupters (CCVI) 426
condenser aperture 211
condenser lens 165, 211
conflat flange 512
conical hollow beam 293
contact stroke 410, 412, 416, 423
contact welding 414
continuous flow 485
continuous wave (CW) tubes 1
contrast ratio 93, 94
  of the screen 108
convergence factor 26
corona discharge 342
correcting lenses 162
coupling coefficients 16
coupling factor 57
critical dimension (CD) 193, 238
critical energy 377
cross-field amplifier (CFA) 2, 47, 55, 58, 61
cross-linking of polymers 202
cross-over projection method 171
crossed-field ion gauge, *see also* Penning gauge 501
crossover 102, 166, 171, 206, 211
cryopump 491
crystal damage 247
Cs on W 449
CsAu compound 450
CTEM 160, 216
curing of laminate materials 202
curing organic materials 198
current 407
  amplification 136
  density 9, 239, 324, 413
  limitation 40
Cutlers beam 36
cutoff wavelength 361
CVD carbon nanotube gated cathode 461
CVD-diamond window 71
cyclotron 359
  classical 359
  frequency 359
  harmonic operation 66
  interaction 65
cyclotron autoresonance maser (CARM) 67, 68

D
damping time 378
dark current 147
de Broglie wavelength 355
deactivation curves 203
deceleration 4, 27, 37, 251
  zone 37
decoding matrix 141
deep space communications 81

deflection 157
aberration 163
amplifier 90
angle 104, 157, 186
astigmatism 157
coils 157
coma 157
error 164
field 163
sensitivity 157
system 195, 216
delay line 27
comb line 34
counter wound double helix 34
coupled cavity 34
helix 33
interdigital line 34
ring and bar line 34
system 33
Delta gun 108
depressed collector 29, 71
depth profiles 247
depth profiling 236
depth resolution 233
desorption, *see also* outgassing 486, 487, 509, 515
yield 384
detection of THz radiation 322
deuterium lamp 335, 336
diamond cathode 460
diamond hopping electron emitter 459
diaphragm pump 489
dielectric barrier discharge (DBD) 341
dielectric strength 414, 415, 420, 423
differential pressure 500
differential pump 494
diffraction 308
aberration 162
error 159
diffuse AMF arc 413
diffuse burning arc mode 411
diffusion 493, 494, 499
pump 483, 490
rate 493
digital display 91
digital grey scale generation 95
digital mirror devices (DMD) 95
dipole magnet 372
directed vapor deposition 155
discharge 265, 266, 307, 324, 450
chamber 265
channel 279, 283, 291, 312
dispenser cathode, *see also* film cathode 10, 166, 436, 477
dispersion diagram 65, 67
display quality 93
dissociative recombination 338
distortion 157
divergence 376
doping 246, 249
Doppler effect 65
Doppler shift term 63, 67
double deflection system 206
double focusing mass spectrometer 516
dragfree satellite 274
drilling 171
rate 177
systems 166
droplet sprayer 272
dual potential test method 221
Duane–Hunts law 344
ductility 436, 475
duoplasmatron 165
dynamic astigmatism 107
dynamic pressure 385
dynamic SIMS 235
dynamic vacuum 383
dynode 132, 134

E
e-beam 156, 172, 179, 192
accelerator 189
disinfection 188
evaporation 156, 181
fusing 178
glowing 178
hardening 178
high rate evaporation 182
liquid phase processes 178
machining 171
melting 166, 188
probing system 220, 221
scanner systems 189
sterilization 188
treated surface 181
welding 156, 163, 166, 168, 174, 175
writers 195
Ebsicon 148

effective contact surface 410, 412
effective ion beam angle 295
effective work function 443
efficiency 1, 12, 18, 46, 70, 72, 99
Einstein coefficients 304, 306
Einstein equation 128
Einstein's equivalence principle 274
Einzel lens, *see also* bipotential lens 102
elastic scattering 343, 380
electric deflection system 131
electric electron lens 131
electric propulsion (EP) 288
electrical charge distribution 146
electrical lifetime 421, 422, 424
electrical visible light sources 322
electrically charged quartz filament 152
electro chemical machining 176
electro discharge machining 176
electro welding, *see also* arc welding 174
electro-magnetic interference (EMI) 98
electron beam 4, 24, 59, 79, 92, 101, 130, 131, 139, 145, 146, 149, 241, 313
electron beam cold-hearth refining (EBCHR) 185
electron beam collector 39
electron beam curing (EBC) 199, 200
electron beam drilling machine 177
electron beam drip melting technology 185
electron beam (EB), *see also* e-beam 155, 156
  energy recovery 79
  melting (EBM) 185
  photo mask 198
  projection (EBP) 195
  spectroscopy 218
  treatment 205
electron beam-induced deposition (EBID) 195, 198, 456, 457
electron beam-induced etching (EBIE) 195
electron beam-induced processes (EBIP) 188, 195, 198
electron-beam drilling 156
electron-beam lithography 157, 191, 192, 195
electron bombardment ion thruster (EIT) 269, 289
electron bombardment ionisation principle 265
electron bunching 4, 13
electron cyclotron frequency 63
electron cyclotron heating (ECH) 72
electron cyclotron maser (ECM), *see also* gyrotron 62
  applications 49
electron emission 429, 450, 452, 467, 473
  cooling effect 434
  current 503
electron energy loss spectroscopy (EELS) 218
electron gun 4, 24, 60, 102, 144, 165, 166, 171, 211
electron lens 159
electron linear accelerator 81, 399
electron microscope 156, 209, 214
electron optical lens 164, 217
electron optical system 161
electron optics 128, 131, 155, 157, 164, 191
electron plasmon excitation 189
electron source 429, 449, 451, 456, 473, 477
electron spatial density modulation 15
electron storage ring 379
electron–positron collider 78
electronic counter measurement (ECM) 25
electrooptical high-speed shutter 139
electrostatic collector 41
electrostatic deflection 86, 103, 146
  field 157
electrostatic Einzel lens 162
electrostatic focusing 170
electrostatic lens 140, 192
electrostatic microscope 165
electrostatic scanning 248
emission current 430, 434, 452, 453, 463, 466, 519
  density 10, 168, 430, 437, 438, 441, 442, 454, 461, 466
emission display 517
emission of electrons 9, 10
emission spectrum 107, 336
emissive displays 87
emittance 369
emitter materials 477
emitting-cathode ionisation gauge, *see also* hot-cathode gauge 503
energy contamination 251
energy dispersive spectroscopy (EDS) 218

epoxy bonding 471
equilibrium emittance 371
equivalent noise input 133
erasing discharge 115
European Space Agency (ESA) 286
European Space Research and Technology Centre (ESTEC) 272
eutectic bonding 473
evaporable getter (EG) pump 492
evaporation 156, 169, 171, 200, 436, 490, 515
  rate 469
excimer laser 311, 313, 340, 342
  tube 312
excimer light source 338
excitation energy 305, 328, 335
excited dimer 311
extended interaction klystron (EIK) 50
extended interaction oscillator (EIO) 50
external photoelectric effect 127
extraction aperture 165
extreme ultra violet lithography (EUVL) 193
extreme UV (EUV) 335, 348
extremely ultrahigh vacuum (XHV) 485

F
far infrared (FIR) 317
far ultraviolet (FUV) 334
Faraday cup 251
fast wave circuit 62
fast-wave device 61
ferroelectric electron emission 467
field electron emission 454, 460
field emission cathode 451, 457
field emission current density 415
field emission display (FED) 87, 122, 451, 452, 464
field emission electric propulsion (FEEP) 272, 274, 278, 288, 296
field emission microscope 156
field emission SEM system 209
field emitter array (FEA) 457
field evaporation 273
field-emission measurement chamber 517
filament temperature 499
film cathode 436
finite element method 158
fixed target 365
flat diode lamp 463
flat panel display (FPD) 85, 89, 91
flicker 90, 122
fluorescent coating 326
fluorescent lamp 325, 326
fluorescent screen 139
focal length 155, 159, 165
focused ion beam (FIB) 236, 237, 245, 273, 457
focusing coil 145
focusing magnetic lenses 174
focusing system 27
FODO lattice 373
folded waveguide 54
frame rate 90
frame time 89, 99
free electron laser (FEL) 62, 315, 317, 356, 368, 403
  pulse 316
free-electron maser (FEM) 62, 70, 78
  amplifier 79
  oscillator 79
Fresnel zone plate 348
fusion bonding 471, 473

G
gain 1, 12, 18, 27
Gamma camera 141, 153
gas
  discharge 164, 501
  laser 307
  loading 509
  poisoning 430, 439
  pressure 485, 499
gaseous discharge (plasma) 324
gaseous propellant 289
gauge 497
  sensitivity 503
Gaussian mode 72
Gaussian phase space distribution 375
Geiger–Mueller counter 150, 219, 346
Geiger–Mueller mode 150
Geißler tubes 429
geometrical aberration 159, 161
geometrical electron optics 155
getter 438, 472, 492
  material 469, 472, 474, 493
glass frit bonding 471, 473
glow discharge device 340

grey scale 94
grid ion thruster (GIT) 288
grid voltage 118
gridded electron gun 3
gyro-amplifier 74
gyro-backward wave oscillator (gyro-BWO) 66, 68, 78
gyro-klystron 66, 75, 76
gyro-travelling wave tube (gyro-TWT) 66, 68, 69, 75
gyro-twystron 66, 75, 77
gyrotron 4, 62, 66, 70, 72
  cathode 10
  oscillator 66, 67, 72, 75
  for industrial applications 74

H
hairpin cathode 167
Hall effect thruster (HET) 278, 288
halo implants 252
harmonic frequency gyrotron 67
Hartree line 58
head-up-displays (HUD) 111
helium leak detector 516
helium sniffer 517
helix 27
helmet-mounted-displays (HMD) 111
HERA accelerator 356, 366
hibachi foil 314
high current cathode 171
high efficiency multistage plasma thruster (HEMP-T) 288, 293
high energy ion scattering (HEIS) 236
high energy linear induction accelerator (HELIA) 254
high pressure gauge 394
high resolution 102, 105, 121
  electron microscope 210
  lens 159
high voltage vacuum breakers 423
high voltage vacuum interrupter 424
high-power electron gun 184
high-pressure discharge lamp 324, 329
high-pressure mercury lamp 330
high-pressure sodium lamp 329
highly relativistic beam 67
HIP cathode 446
hollow cathode arc activated (HAD) 184
hollow cathode lamp 338
hollow ion beams 292
hollow-cathode neutralizer 266, 280, 281, 291, 293
Holweck stage 497
homogeneous excitation 311
hopping electron cathode 465
hot-cathode gauge 503
Hull parabola 58

I
I cathode, *see also* Os/Ru coated impregnated (I) cathode 431, 439, 476
ideal gas law 484
ideal quadrupole field 372
image amplification factor 139
image converter 139
imaging 233
  lens 210, 216
  SIMS 235
imperfection aberration 163
impuls per propellant weight 294
impulse rise time 136
In-Line gun 108
in-vacuum undulator 402
incandescent lamp 322, 323, 436
indium thruster 273
indium tin oxide (ITO) 144
induction linac 79
inductive output tube (IOT) 3, 22
inelastic scattering 343, 381, 383
inertial confinement fusion (ICF) 179
inertial fusion energy (IFE) 313
  requirements 254
infrared focal plane array technology 319
inorganic scintillator 137
integrated sputter ion pump 392
interaction efficiency 5, 20, 27
interaction of beam energy 4
interaction process within magnetron 56
interference telescope 274
International Linear Collider (ILC) 363, 367
interrupting capability 409, 410, 413, 419, 425
interrupting current 410, 419, 425
ion beam 242
  analysis 232
  energy distribution 281

ion bombardment (IB) 430, 440
  resistivity 441, 443
ion current 409, 502, 503
ion emitter array 277
ion implantation 238, 247
ion implanters
  batch 248
  high-current 251
  ultra-low energy 250
ion laser 310
ion scattering spectroscopy (ISS) 236
ion source 237, 248
  contaminants 258
  field emission 258
  space charge limited 258
ion-induced deposition process 240
ion-induced desorption 385
ionisation
  chamber 150, 151, 506
  constant 345
  dose 345
  efficiency 295
iron yoke magnet 361, 399
ISM applications 81
isochronous cyclotron 359
isotropic astigmatism 159
isotropic coma 159
isotropic distortion 160

K
klystron 2, 12, 14, 47, 48, 74
  amplifier 12
  cavity 15
  reflex 3
  tunable 3
  two cavity 50
Knudsen flow 486

L
L-cathode 431
LaB_6 evaporation rate 447
LaB_6 rod cathode 168
ladder circuit 51
laminar flow 485
lanthanated molybdenum (LM) cathode 438
Laplace equation 8
large electron-positron collider (LEP) 376, 385
  vacuum chamber 392
Large Hadron Collider (LHC) 367
large signal operation of the TWT 37
light amplification by stimulated emission of radiation (laser) 307
  ArF 337
  argon ion 310
  ablation deposition (LAD) 442
  beam drilling 176
  cell 313
  CO_2 308
  continuous wave (cw) 308
  diode 321
  discharge pumped 310
  He–Ne 309
  KrF 313
  medium 307, 308, 310
  mirror 308
  pulse duration 312
  resonator 308
  quantum cascade 321
  ultraviolet nitrogen 308
  X-ray 308, 310
leak detection 514
lens 157, 159
  aberration 158, 161
  field 160
  system 214
lethal dose 345
lifetime 3, 93, 98, 99, 112, 117, 119, 166, 168, 169, 237, 286, 314, 323, 326, 328, 338, 416, 422, 447, 449
  tests 276
light amplification factor 140
light emission 114
light emission spectroscopy 221
light penetration depth 129
light sensitive converter target 143
light wavelength 130
light-ion generator
  current enhancement 255
  diocotron instability 255
  injector gap 254
  ion mode instability 255
  virtual cathode 255
LINAC 28
linear accelerator 361, 390
linearity 1, 25, 39, 46, 143
Liouville theorem 371

liquid effluent waste 205
liquid helium cryopump 492
liquid metal droplet 289
liquid metal ion source (LMIS) 243, 273
  capillary type 274
  needle type 274, 276
  slit type 274
  sources 237
  propellant 273, 278
lithography 170
Littrow prism 311
local oscillator 54
local thermal equilibrium (LTE) 324
Lorentz force 372
low energy electron flood gun 241
low energy ion scattering (LEIS) 236, 441
low vacuum (LV) 485, 500, 505, 510, 517
low voltage circuit breakers 422
low-pressure discharge lamp 324
low-pressure mercury lamp, *see also* fluorescent lamp 325
low-pressure neon lamp 328
low-pressure sodium lamp 328
low-resolution display 90
lumen sensitivity 129
luminance 93, 97, 98, 119
luminosity 366
luminosity lifetime 382

M
M cathode 431
M-carcinotron 54
magnet cryostat 399
magnetic bearing turbopump 497
magnetic deflection 90, 155
  mass spectrometer 506
  of electron beam 104
  system 105, 132
magnetic electron lens 131
magnetic focusing 31, 146
magnetic lens 155, 158, 192, 216
magnetic-focus-lens electron gun 103
magnetically focused collector 42
magnetron 2, 24, 47, 55, 81
  efficiency 58
  injection gun 70
  oscillator 59
  pump 494
maser 62, 307
mask repair (MRP) 237
mass absorption 345
mass spectrometer helium leak detector 506, 515
matrix drive 89, 90
maximum amplification factor 140
maximum current amplification 134
Maxwell equations 7
mean time to failure (MTTF) 408, 427
mechanical lifetime 421, 424
medium energy ion scattering (MEIS) 236
medium vacuum (MV) 485, 505
medium voltage range 407, 409, 420, 426
melting 156, 171
  point 178, 181, 435, 449, 475
  temperature 412
  zone 185
MEMS 277
mercury arc lamp 337
mercury diffusion pump 483
mercury-free metal halide lamp 333
metal halide lamp 329, 331
metal–insulator–metal–insulator–vacuum (MIMIV) device 459
micro-hollow-cathode (MHC) discharge 342
microfabrication 277
MicroPirani gauge 499
micropropulsion 275, 277
microscope 156
microspacecraft 274
microthruster 274, 275
microwave ovens 81
microwave tube 1, 10, 24, 45, 62
microwelding 174
mixed matrix Scandate cathode 441
Mo-brightness 442
mode of propagation 1
modulating cavities 19
modulation transfer function (MTF) 98
modulations and energy extraction 13
MoirT effect 98
molecular flow 486
molecular radiator 331
monotron oscillation 16, 22
Monte Carlo simulations 247
MOSFET 249
multi-cell cavity 362
multi-line radiator 331

multi-pulse drilling 176
multibeam high power klystrons (MBK) 3, 20
multicavity klystron 3
multicavity multigap design 50
multiplier channel 134, 136
multipoint explosive emission cathode 180
multistage collector 39

N
nanocrystallized polysilicon (NPS) 465
nanogetter 473
natural radiation dose 345
negative phase contrast 218
neutral propellant atoms 280
neutralizer 277, 280, 285
Newvicon target 144
Ni doped cathode 444
nominal thrust 286
non-contact e-beam testing 221
non-evaporable getter (NEG) pump 390, 391, 492
  coating 393
non-thermal EB refinement 201
non-thermal materials refining 155
non-thermal processes for treating materials 189
normalized emittance 371
numerical aperture 192
numerical indicator experimental (NIXIE) tube 86

O
O-ring 512
objective aperture 211, 216
objective lens 160, 174, 192, 206, 213
Omegatron 506
on-line elemental analysis 349
one beam klystron (OBK) 20
one-dimensional vacuum system 389
one-shot brazing technology 408
onset ion emission 275
optical lithography 238
optical proximity effect correcting structures (OPC) 239
optical resonator 308
organic scintillator 137
Os/Ru-I cathode 443
oscillator 1, 3, 47
outgassing 487, 506, 515
  rate 510, 511, 513
output power 58, 312
  optimization 18
oxide cathode 2, 101, 167, 430, 439, 474, 477
  alkaline earth oxide cathode 443
  ThO_2 cathode 446
oxygen free high conductivity (OFHC) copper 510

P
parasitic lens 158
partial pressure 507
partial vacuum 483
particle
  accelerator 355, 398
  size analysis 349
  source 364
peak current 417
penetrating light flux 128
penetration depth 212, 412
Penning gauge 502
permanent magnet 19, 55, 74, 402
permanent periodic magnet (PPM) 25, 42, 289, 291
perveance 5, 9, 20, 165, 168
PFE's screen printable cold cathode 459
phase contrast 214
phase lag 52
phase space 369
phased array antenna 81
phased-array radar 77
phosphor 92, 98, 122
  grain 108
photo cathode 128, 132, 139, 140, 148, 364, 469
photo emission 468
photo-effect with ionisation 346
photo-electron emission 467
photoconductive resistance target 143, 147
photoelectron 128
  current 139
  emission 128
    picture 139
photomask repair (PMR) 238
photomultiplier 132, 153
photon emission 305
photon energy 130, 304, 305

photon flux 384, 385
photosensitive target 143
physical vapor deposition (PVD) 155
Pierce electron gun 29
Pierce type electron gun 13, 19, 23
Pierce-cathode 168
Pirani gauge 499
pixel pitch 95, 96, 98, 117
Planck's formula 303, 335
plane luminescent screen 139
plane photocathode 139
plasma discharge 165, 184
plasma display 112
plasma display panel (PDP) 85, 87, 95
  grey scale 116
plasma frequency 17
plasma-activated CVD 438, 441
plasma-immersion ion implantation (PIII) 253
Plumbicon target 144
pocket dosimeter 152
Poisson equation 8
positive electron beam resist 192
positive ion bombardment 447
positron linear accelerator 399
post-accelerator gap
  limiting voltage 257
  multi-stage acceleration 256
  virtual anode 257
pot depth 410, 411
pot diameter 410, 411
pot-shaped AMF contact 410, 411
potential depression 40
power 1
  density distribution 172
  supply 5, 98
  transmission 81
practical work function distribution (PWFD) 431
pre-focusing lens 107
pressure 484
  instability 386
  profile 389
  rise method 514
primary electron 344
primary electron energy 130
priming and data-write discharge 114
probe astigmatism 196
projected range 247
propellant 265, 278, 288, 297
  for ion thrusters 265
  gas flow 293
  inlet scheme 283
  ionization 280
    efficiency 285
  ions 280
  mass 281
    flow 284, 295
  weight 294
proportional mode 150
proton induced γ-ray emission (PIGE) 234
proton induced X-ray emission (PIXE) 234
proton linear accelerator 399
proton microprobe 234
pulse wave tubes 1
pump 483, 487
pumping speed 386, 490, 495
purification of drinking water 205
Pyroelectric Vidicon 148
pyrolytic sandwich cathode 169

Q
quadrupole magnet 372, 373
quadrupole mass filter 506
quadrupole mass spectrometer (QMS) 236, 507
quantum counter 135
quantum detector 319
quantum efficiency 129, 130, 327
quasi-optical mode converter 71

R
radar 81
radar and ECM TWTs 47
radial current density 413
radial magnetic field (RMF) 409, 417, 419, 420
  contact 412
  system 410
radial output coupling 71
radiation damping 377, 379, 382
radiation emission 323
radiation length 381
radiation power 376
radiation sensitive vacuum electronic components 128
radiation sensitive vacuum tube 127
radiation sensitivity 153

radical-induced reactions 190
radioactive radiation 150
radiofrequency ionisation principle 266
radiofrequency ionisation thruster (RIT) 266, 289
radiography 349
radiometer 318
rapid prototyping 245
rare gas emission spectra 339
rare gas excimer light source 339
raster scan 89, 90
Rayleigh scattering 347
reactive evaporation 182
recirculator 313
reentrant cavity 13
reflection coefficient 5, 10, 99
relativistic equation of motion of a single electron 6
reliability 1, 22, 49, 289
repetition rates 312
residual gas analyser (RGA) 394, 507
residual gas pressure 379
resist bonding 471
resistive layer 134
resolution 86, 88, 91, 117
resolving power 143
resonant absorption 305
response time 99
retarding-field tube 2
RF gun 365
RGB colour system 98
ribbon beam 248
Richardson work function 442
Richardson–Dushman equation 10
Richardson–Dushman–Schottky equation 10
Rieke diagram 5
RIT 10 testflight on EURECA 268
RIT for commercial applications 271
RIT for interplanetary missions 272
Robinson theorem 378
rotary pump 390
rotary vane pump 483, 488
rotating anode 165
Rutherford backscattering analysis (RBS) 232
Rutherford scattering 380

S
saturation current 152, 433
‘Sc’-I cathode 441
‘Sc’/Re-I cathode 442
scalar potential 7
scaling 92
  laws 9
Scandate cathode 431, 441, 477
scanning beam 149
scanning electron microscope (SEM) 156, 169, 196, 206, 210, 218, 221, 222, 316, 317, 449
scanning process 147
scanning system 248
scanning transmission electron microscope (STEM) 161, 216, 218
scanning tunnelling microscope 196
scattering absorption contrast 213
Schottky effect 435, 468
scintigrams 142
scintillation camera 141
scintillation counter 132, 137, 138, 153
scintillation crystal 137, 139, 141
screen oxides 252
second harmonic cavity 18
secondary electrons (SE) 206
  current 130, 132
  emission 127, 130, 132, 189, 237
  energy spectrometry 221
  multiplier 132
  spectroscopy 221
secondary emission 128, 149, 466
  coefficient 131, 132
secondary ion emission 237
secondary ion mass spectrometry (SIMS) 234
  static 234
self-contained cryopump 492
semiconductor cold cathode 450
semiconductor device fabrication 249
sensitive entrance electrode 132
sensitivity 143
separate evaporation 183
sextupole magnet 372
signal amplification factor 153
signal-to-noise ratio 1, 133
silicon multidiode target 143–145
silicon tip avalanche cathode (STAC) 463
silicon-on-insulator (SOI) 250

single contact gap 426
single crystal sensor 141
single junction diode target 143–145
single stage collector 28
single-stage rotary vane vacuum pump 489
slot angle 410, 411
slot number 410, 411
slotted cathode pump 494
small geostationary communication satellite (SGEO) 297
small mission for advanced research in technology (SMART) 286
small signal theory (Pierce) 34
Smith–Purcell effect 315
solder bonding 471
source/drain extension (SDE) 249
space charge 7, 10, 17, 18, 27, 29, 30, 134, 250, 433, 434, 442, 447, 467
  density 8
  field 469
  limited current 11
  limited emission 10
space materials analysis 276
space resolution 134
spacecraft charging 277
spacecraft contamination 273, 275
spatial conductivity distribution 144
spatial resolution 143
specific excimer light sources 340
specific impulse 266, 268, 271, 272, 278, 284, 288, 293, 297
specific outgassing 487
specific pumping speed 389
spectral sensitivity 133, 150
spectroscopy 155
spherical aberration 159, 214, 218
Spindt cathode 82
Spindt-type Mo cathode 123
spinning rotor gauge 501
spiral cathode 167
spiral contact 409
spontaneous emission 305
spot size 106
spotless arc deposition (SAD) 184
spotless arc process (SAP) 184
sputter ion pump 390, 391, 494
sputtering 234, 252
sputtering of atoms 237
  anode 493
standing wave ratio (SWR) 5
Stanford Linear Accelerator Center (SLAC) 367
starvation effect 414
stationary plasma thruster (SPT) 279
stencils mask 197
step and flash imprint lithography (SFIL) 193
sterility assurance level (SAL) 203
stimulated emission 305, 308
Stirling cooler 319
streamer 341
strip cathode 167
stroboscopic imaging method 222
strong focusing 360, 372
sublimation pump 494
superconducting cavity 362, 402
superconducting magnet 361, 393
surface analysis 234
surface desorption rate 509
sweep signal 89
switching capability 409, 411, 426
synchro cyclotron 359
synchronism condition 53, 65
synchrotron 359
  oscillations 360
  radiation 348, 355, 376, 379, 395, 401–403
    induced desorption 383
    production 368
    research 361, 368
    source 356
    spectrum 377

T
tandem accelerator 358
Taylor cone 237, 273
telecentric mode 160
television camera tube 143
television tube 439
temperature limited emission 10
tera-hertz (THz) 315
  applications 319
  photon 317
  radiation 320
    generation 320
  spectroscopy 321
TESLA Test Facility (TTF) 369
thermal annealing 246

thermal barrier coatings (TBC) 181
thermal conductivity, *see also* thermal diffusivity 172, 173, 395, 396, 412
  gauge, *see also* Pirani gauge 498
thermal diffusivity 173
thermal EB machining processes 171
thermal electron emission 431
thermal field emission cathode 192
thermal processing of solids 171
thermal processing of thin films 171
thermal surface treatment 155
thermally grown oxide (TGO) 181
thermionic cathode 26, 71, 155, 431, 441, 460
thermionic constant 432
thermionic electron source 447
thermionic emission 431
thermistor gauge 499
thermo-field emission cathode 453
thermo-field (T-F)/emission, *see also* Schottky emission 453
thermocouple 499
  gauge 499
thermography 318
thin film 492
thin tungsten filament 118
thin-film evaporation 490
Thomson effect 435
thoriated tungsten cathode 437
thoriated tungsten (Th–W) wire cathode 436
three-band colour radiator 331
three-dimensional analysis 236
thrust 267, 271, 273, 278
  level 266, 268, 271, 288, 294
thruster with anode layer (TAL) 279
time-of-flight mass spectrometer 506
titanium sublimation pump (TSP) 390, 391
TOF–SIMS 236
top-layer (T-L) Scandate cathode 441
total efficiency 37, 42–44, 46
total gain 53
transfer curves 6
transfer of beam energy 4
transformer tap changers 424
transient recovery voltage (TRV) 414
transit angle 65
transparent photocathode 139
transverse emittance 370
transverse gun 181
traveling wave tube (TWT) 3, 24, 29, 45, 47, 49, 289, 493
  applications 45
  collector 28
  communication 45
  coupled cavity 3
  efficiency 29, 44
  helix 3
  total efficiency 43
Triglycinsulfate (TGS) 148
triode 2, 22, 118, 145
  compact 82
  pump 494
tungsten (W) wire cathode 436
turbo-molecular pump system 408
turbomolecular pump, *see also* turbopump 390, 494, 495
TV klystron 22
TV tube 490
two aperture method 171
two-stage rotary vane vacuum pump 489
twystron 74

U
UBITRON 62
UHP lamp 331
UHV diode 475
UHV pump 492
ultra-fast photomultiplier 137, 138
ultrahigh vacuum (UHV) 485, 490, 494, 503, 507, 518
ultraviolet (UV) light 334
  UVA 334
  UVB 334
  UVC 334
undoped cathode 443
undulator 64
  magnet 368
universal beam spread 31
UV arc lamp 336
UV lithography 242
UV photon 326

V
vacuum based display (VFD) 85, 92, 118
  waveform 121
vacuum chamber 473
vacuum circuit breakers 407, 409

vacuum conductance 389
vacuum fluorescent display (VFD) 85, 87, 118
vacuum interrupter 407, 415, 421, 426, 493
design 417
vacuum microelectronic 82, 278
vacuum pressure 485, 497, 504, 519
vacuum pump 473, 483, 487, 517
vacuum tube 132, 143, 449
vacuum ultraviolet (VUV) 334
van de Graaff generator 358
vapour diffusion pump 490
vapour pressure 477, 491
variable shaped beam (VSB) 193
vector potential 7, 9
vectorial photoeffect 468
video bandwidth 91
videosignal 149
production 146
Vidicon 143
target 144
viscosity 476
visible photon 326
Vivitron 358
vulcanization of elastomers 202

W
W-I cathode 443
wafer 242, 244
contamination 252
handling system 248
heating 251
wall plug efficiency 312
wavelength dispersive spectrometry (WDS) 218
weak focusing principle 360
Wehnelt electrode 30
welding 155, 173
depth 174
process 156, 173
speed 174
system 166, 174
Wideröe condition 364
wiggler 64, 70, 78
magnet 368
Wolter telescopes 348
work function (Austrittsarbeit) 10, 101, 128, 259, 432, 433, 437, 455, 477

X
X-ray 343
diffractometer 346
dosimetry 345
FEL 317
filter 344
image amplifier 140
imaging 347, 349
lines 344
spectroscopy 209
tube 430, 451, 493, 517
Xe propellant gas 291
XHV pump 492

Z
Zeiss Crossbeam 243, 244
zero field thermionic emission 432