Topics in Applied Physics 135

Georg Gaertner
Wolfram Knapp
Richard G. Forbes *Editors*

Modern Developments in Vacuum Electron Sources

Springer

Topics in Applied Physics

Volume 135

Topics in Applied Physics is a well-established series of review books, each of which presents a comprehensive survey of a selected topic within the domain of applied physics. Since 1973 it has served a broad readership across academia and industry, providing both newcomers and seasoned scholars easy but comprehensive access to the state of the art of a number of diverse research topics.

Edited and written by leading international scientists, each volume contains high-quality review contributions, extending from an introduction to the subject right up to the frontiers of contemporary research.

Topics in Applied Physics strives to provide its readership with a diverse and interdisciplinary collection of some of the most current topics across the full spectrum of applied physics research, including but not limited to:

- Quantum computation and information
- Photonics, optoelectronics and device physics
- Nanoscale science and technology
- Ultrafast physics
- Microscopy and advanced imaging
- Biomaterials and biophysics
- Liquids and soft matter
- Materials for energy
- Geophysics
- Computational physics and numerical methods
- Interdisciplinary physics and engineering

We welcome any suggestions for topics coming from the community of applied physicists, no matter what the field, and encourage prospective book editors to approach us with ideas. Potential authors who wish to submit a book proposal should contact Zach Evenson, Publishing Editor:

zachary.evenson@springer.com

Topics in Applied Physics is indexed by SCOPUS and books in this series are submitted for indexing to Web of Science.

2018 Impact Factor: 0.746

More information about this series at http://www.springer.com/series/560

Georg Gaertner · Wolfram Knapp ·
Richard G. Forbes
Editors

Modern Developments in Vacuum Electron Sources

Springer

Editors
Georg Gaertner
Aachen, Nordrhein-Westfalen, Germany

Richard G. Forbes
Advanced Technology Institute and
Department of Electronic Engineering
University of Surrey
Guildford, Surrey, UK

Wolfram Knapp
FMB/IFQ
Otto von Guericke University Magdeburg
Magdeburg, Germany

ISSN 0303-4216 ISSN 1437-0859 (electronic)
Topics in Applied Physics
ISBN 978-3-030-47290-0 ISBN 978-3-030-47291-7 (eBook)
https://doi.org/10.1007/978-3-030-47291-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

During the historical development of Vacuum Electronics a lot of review articles and also several more extensive textbooks have been written on vacuum electron sources, reflecting the status of knowledge of that time. One of the first textbooks was "Thermionic Emission" by Reimann (Chapman & Hall, London 1934 [1]). More famous later on became the book of Herrmann and Wagener "The Oxide Coated Cathode" (Chapman & Hall, London 1951 [2]). It appeared during the introduction of the reservoir and impregnated cathodes and suffers from this handicap. Of course also cold emission was not a topic then. There are older detailed review articles of that time by Herring and Nichols on thermionic emission [3] and by Nottingham [4]. There are also detailed chapters on thermionic emission in the context of electron tube applications by Kohl from Sylvania [5] and by Nergaard et al. from RCA [6] and also review articles, e.g., by Haas et al. [7] on thermionic emission and by others, all originating during the high noon of vacuum electron tubes before 1980. It is still instructive to read these books or reviews nowadays, since some of the knowledge of the past has been forgotten.

In the last 20 years also some reviews on cold cathodes appeared, especially the book edited by Zhu, "Vacuum microelectronics" [8] and the book "Field Emission Electronics" by Egorov and Sheshin [9], published by Springer in 2017.

More recent overviews dealing with thermionic and cold cathodes were given by Hawkes [10], Dowell et al. [11], Yamamoto [12] and Umstattd [13]. In 2008 appeared the Springer book "Vacuum Electronics—Components and Devices", containing also a chapter on vacuum electron sources by Gaertner and Koops [14], giving a wrap up on all types of historic and new cathodes used in different tubes, but not going into more detail especially needed for modern cathodes.

Hence in view of new insights, new types of cathodes and thermionic cathode improvements during the high noon of the CRTs, and due to intense research on cold cathodes in the last 20 years, it became important to compile a new textbook on vacuum electron sources. This is also intended as a comprehensive source for future cathode developments and new applications with specific demands, which is much more detailed than the last review articles. This is now realized in this book, which will mainly address the more recent types, with special emphasis on

promising new trends for future applications. Yet we have also included some already historic top results on scandate and oxide cathodes from the last phase of the CRT era as possible starting points for future developments. The requirements of high-brightness cathodes for electron beam applications are addressed in Chap. 6. Modern photocathodes are the subject of Chap. 7. In Chaps. 8 and 9 theoretical concepts of thermal, photo and mainly field emission are critically discussed. Chapters 10–12 are devoted to carbon field emitters, explosive emitters and field emitter arrays.

"Modern Developments in Vacuum Electron Sources" deals with one of the most essential components of all vacuum electron devices, namely the electron sources or cathodes, which in general are decisive for the overall performance of the respective vacuum electron device or vacuum tube. Despite the rise and the fall of once dominating types of vacuum tubes such as radio valves and cathode-ray tubes, the improvement of cathodes continues and new applications with increased demands arise, such as electron beam lithography, high-power and high-frequency microwave tubes, terahertz imaging, and electron sources for accelerators. New developments in cathodes needed for these applications are addressed by world experts in this field, wrapping up the state of the art and giving future perspectives. Let us close with an advice to the reader: if you are interested in the history of science and technology, in general, you should start with Chap. 1. If you want to know more about the basics and electron emission theory, you should start with Chaps. 8 and 9.

In case of German names with Umlaut (ä, ö, ü) always the German-English transcription (ä = ae, ö = oe, ü = ue) has been used. In case of literature searches please take both versions into account. (the same holds for German ß = ss)

General Literature: Reviews on Cathodes (Thermionic and/or Field Emission)

1. A.L. Reimann, *Thermionic Emission* (Chapman & Hall, London, 1934)
2. G. Herrmann, S. Wagener, *The Oxide Coated Cathode* (Chapman & Hall, London, 1951)
3. C. Herring, M. Nichols, Thermionic emission. Rev. Mod. Phys. **21**(2), 185–270 (1949)
4. W. Nottingham, *Thermionic Emission, Handbuch der Physik*, ed. by S. Flügge, vol. 21, Elektronen-Emission/Gasentladungen I, (Springer, Berlin, 1956), pp. 1–175
5. W.H. Kohl (Sylvania), Materials and Techniques for Electron Tubes, Reinhold Publishing 1960, revised edition; 1. Edition 1951: Materials and Technologies for Electron Tubes
6. L.S. Nergaard, R.S. Burnap et al., Electron Tube Design, RCA internal publication 1962, collection of 53 articles by RCA engineers, starting with "Fundamentals of Electron Emission" by L. S. Nergaard, pp. 1–27

7. 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
8. Ed. Wei Zhu, *Vacuum Microelectronics* (Wiley, 2001) (Topic: Field Emission)
9. N. Egorov, E.P. Sheshin, *Field Emission Electronics* (Springer, 2017)
10. P.W. Hawkes, Thermionic emission. Encycl. Appl. Phy. **21** (Wiley, 1997), 229–243
11. D.H. Dowell, J. Smedley et al., Cathode R&D for future light sources. SLAC-Pub-14002 and Nucl. Instrum. Methods Phys. Res. A **622**, 685 (2010)
12. S. Yamamoto, Fundamental physics of vacuum electron sources. Rep. Prog. Phys. **69**, 181–232 (2006)
13. R.R. Umstattd, *Advanced Electron Beam Sources, Chapter 8 in Modern Microwave and Millimeter-Wave Power Electronics*, ed. by R. Barker et al., (Wiley, 2005), pp. 393–444
14. G. Gaertner, H.W.P. Koops, Vacuum electron sources and their materials and technologies, Chap. 10 in Vacuum Electronics (Springer, Berlin, 2008), pp. 429–482

Aachen, Germany — Georg Gaertner
Magdeburg, Germany — Wolfram Knapp
Guildford, UK — Richard G. Forbes
January 2020

Contents

Contributors

Rolf Behling XTraininx—Technical Consulting, Hamburg, Germany

Nikolay V. Egorov Saint Petersburg State University, St. Petersburg, Russia

Richard G. Forbes Advanced Technology Institute & Department of Electrical and Electronic Engineering, University of Surrey, Guildford, UK

Georgiy N. Fursey The Bonch-Bruevich Saint-Petersburg, State University of Telecommunications, Saint Petersburg, Russia

Georg Gaertner Consultant, Aachen, Germany

Romain Ganter Paul Scherrer Institut, Villigen, Switzerland

Michael Green Consultant, Palo Alto, CA, USA

Kevin L. Jensen Code 6362, MSTD Naval Research Laboratory, Washington, DC, USA

Wolfram Knapp Otto von Guericke University, Magdeburg, Germany;
Knapptron GmbH Vacuum Electronics, Möser, Germany

Pieter Kruit Delft University of Technology, Delft, The Netherlands

Wei Liu Thomas Jefferson National Accelerator Facility, Newport News, VA, USA;
Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou, China;
University of Chinese Academy of Sciences, Beijing, China

Matt Poelker Thomas Jefferson National Accelerator Facility, Newport News, VA, USA

Jean-Michel Roquais Thales AVS France, Vélizy-Villacoublay, France

Evgeny P. Sheshin Moscow Institute of Physics and Technology, Dolgoprudny, Russia

John Smedley Los Alamos National Laboratory, Los Alamos, NM, USA

Bernard Vancil Ebeam Inc., Beaverton, OR, USA

Yiman Wang Beijing, China

Chapter 1
History of Vacuum Electronics and Vacuum Electron Sources and Future Development Trends

Georg Gaertner

Abstract The historical rise of vacuum electronics (VE) was enabled by the availability of electrical power and by improved vacuum techniques, but its further progress relied on improved electron sources and their control. The development of VE has been pushed by several technological waves/cycles, starting with incandescent lamps, continuing with the radio tube era and then followed by the cathode-ray tubes. Yet vacuum electronics is still alive and has specific advantages in the high-power, high-frequency domain. The improvement trends of cathodes over time, related to specific and also advanced application requirements will be addressed.

1.1 The Foundations of Vacuum Electronics

1.1.1 The Availability of Electric Power

One of the prerequisites for vacuum electronics is, of course, the availability of electric power generators and later on of a distributing grid for power supplies. Hence the advances in this field happened some decades before the rise of vacuum electronics started [1–4]. We find a similar development (limiting) growth curve as in other fields of technology.

Already in antiquity around 585 B.C., the first concepts of magnetic and electric forces were described by Thales of Miletus, electron being the Greek word for amber, wherefrom charges could be generated by rubbing, and magnetic forces being manifested in magnetic ores found in Magnesia [3, 5]. Yet they more or less remained curiosities and did not trigger any applications. An exception could be the so-called battery of Baghdad from the time of the Parthian Empire after 247 B.C., which was capable to deliver 250 μA at 0.25 V, when a saline solution was added; a possible application could have been electroplating [6].

G. Gaertner (✉)
Consultant, Aachen, Germany
e-mail: georg.f.gaertner@t-online.de

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*,
Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_1

Fig. 1.1 The 1663 electrization machine of Otto von Guericke, using a sulfur sphere and friction. The figure is based on [7], "The electrical experimenter", Sept. 1915, p.198

In modern times Otto von Guericke was not only a pioneer in vacuum technology with the Magdeburg half-spheres experiment in 1654, demonstrating that 16 horses could not draw two evacuated metal half-spheres apart [2, 7, 8]. Guericke also built an electrization machine in 1663 (see Fig. 1.1), using a sulfur sphere and friction [3, 7–11]. With this sulfur sphere, Gottfried Wilhelm Leibniz in 1672 discovered electrical sparks. In 1745 the German cleric Ewald von Kleist and the Dutch scientist Pieter van Musschenbroek from Leyden, both found that charges generated by friction could be stored and accumulated in a Kleist jar or more commonly "Leyden jar" [9]. Typical for spark experiments using Leyden jars were very high voltages and rather low discharge currents (with powers in the order of 30–50 W, see Ayrton [11]). In 1799/1800 Alessandro Volta (Italian) produced continuous electrical power for the first time (as opposed to a spark or static electricity) from a stack of 40 cells of silver and zinc plates, with felt soaked with salt solution in between the electrodes. This battery lasted for several days [3, 9, 12]. From our knowledge nowadays, it had a electromotive force of 62.8 V and should have been capable of delivering about 20 W, due to the conversion of chemical to electrical energy in a redox reaction. In a replica experiment to be followed on YouTube [13], one can see that the voltage of 22 elements' stack of Zn and Cu, which Volta also used, is about 24 V and the power in the order of 8 W. Of course, one can increase the current by using several parallel cells or larger cell sizes and one can increase the voltage by further stacking. Such an upscaling was of course realized in the following years. The introduction of the horizontal trough battery by William Cruickshank in 1802, in one version consisting of 60 Zn–Ag pairs with an estimated power of 17 W, increased battery life to several weeks. For all these galvanic cells the achievable current was limited by the electrode area, the current density being usually a factor of about 5 lower than

the maximum current density of about 50 mA/cm^2 [14, 15]. Humphrey Davy used 3 different batteries with powers ranging from about 170 to 260 W for his chemical experiments, with which he first isolated alkali elements and also demonstrated arc discharges between carbon electrodes. Based on these improved cells, William Pepys in 1808 started to construct one of the strongest batteries with 2000 plate pairs in a trough configuration and 82.6 m^2 total plate area (103 cm^2 single plate surface area) for the Royal society of London [3, 16, 17], financed by a subscription [17]. It was a plunge type battery in a nitrous and sulfuric acid solution of higher conductivity, finally installed in 1810 with an estimated power of about 7 kW at 1.7–2.2 kV (see Fig. 1.2).

A battery of 600 Zn–Cu galvanic cells with single electrode surface area of 900 cm^2 was constructed in Paris in 1813, funded by Napoleon I, with an estimated power of 4.6 kW. But both approaches had been surpassed before by Vasily Petrov in St. Petersburg in 1802/1803 with a huge, also horizontal battery of 4 troughs with 4200 Zn–Cu galvanic cells (491 cm^2 single plate area) with an estimated power of 17 kW. Yet he was still using a salt solution (as Volta) of lower conductivity. He also demonstrated the first continuous arc discharge between two carbon electrodes in 1802/3, but unfortunately published it only in Russian [18].

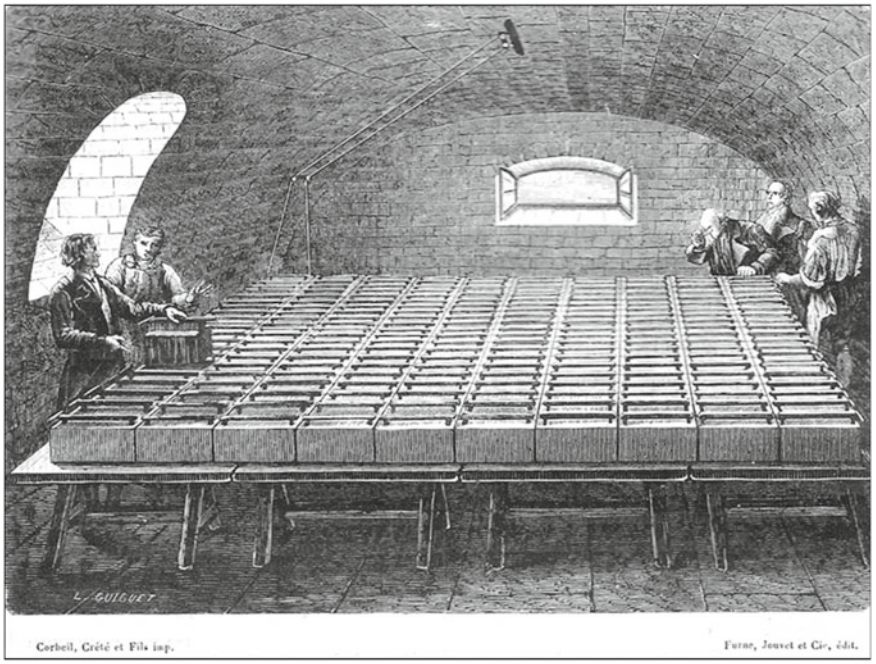


Fig. 1.2 The great Battery of London (ca. 1810), figure from [16], Louis Figuier, "Les Merveilles da la Science", Paris 1867 (Fig. 346, p. 673), reproduced by G. Gaertner

The main improvement trend was to increase the rather short life of these galvanic cells by using different and improved materials. In this context John F. Daniell (UK) in 1836 introduced a porous diaphragm between the Zn–Cu electrodes and 2 fluids in order to overcome polarization [3, 9]. In 1854 the German Wilhelm Josef Sinsteden invented the lead accumulator by replacing copper by lead and using sulfuric acid as the fluid, which was then strongly improved in 1859 by the Frenchman Gaston Planté, making the first secondary or rechargeable battery with electrodes of Pb and PbO_2 and an electrolyte of sulfuric acid technically feasible [9]. A predecessor of this accumulator was invented by J. W. Ritter in 1802 [9]. In Germany the physician Carl Gassner in 1887 developed the first dry cell.

It has to be noted that up to that time also in the later literature in most cases, no performance data on these devices were given since metrology was still in its infancy. The author estimated the performance based on material and geometrical data and our knowledge nowadays and made use of the work of Ayrton [11] and King [19]. Of course the estimated power given for different batteries up to 1825 has to be reduced, if one sets a minimum requirement for the operational time of about 1000 h.

These electrical power sources were also very valuable for establishing the laws of electricity and magnetism in the years to come, as done by the Danish man Hans Christian Oersted, by the Frenchman André- Marie Ampère in 1820 and later by the German Georg Simon Ohm (1826), the Englishman Michael Faraday (1829, 1852) and finally by James Clerk Maxwell (1865) [3].

Another approach to supply electrical power was based on the conversion of mechanical energy to electrical energy by moving magnets and inductive currents. The first usable machine was built by the Frenchman Hyppolite Pixii in 1832, by turning a permanent magnet in front of a pair of coils, producing an alternating current. In a second machine built in the same year, he introduced a commutator and obtained undulating DC current [9, 12]. A commercial application arose, when a supply for the electrical arc lamps for lighthouses was needed. In 1857–1858 Prof. Frederick Holmes constructed a magneto-electric machine for the South Foreland lighthouse in the UK. The test version had 36 permanent magnets on 6 wheels, weighed 2000 kg and gave a DC output of 1.8 kW [3] (according to [20] only of 700 W). In the final version two machines with 60 permanent stationary magnets were delivered, now with iron frames instead of wood, which weighed 5500 kg. The two wheels with coils were driven by a steam engine through a belt drive [3, 19]. Looking at the weight of these machines, they were not really efficient.

Werner Siemens in Germany showed, based on an idea of Henry Wilde, that permanent magnets were not necessary to convert mechanical to electrical energy, and built his first technically convincing dynamo-electric machine (“Elektrodynamische Maschine”) of table-size in 1866, which could deliver 26–29 W. Thus, the efficiency could be greatly improved and in the years to come more powerful machines were built, for instance, the Siemens dynamo of 1877 with a commutator delivering 20 A at 50 V (1 kW DC) [3, 5, 9, 20]. The Pearl street power station installed by T.A. Edison in 1879 generated electric power of 100 kW [3, 4]. In 1883 the company Siemens Brothers installed a dynamo machine with single-phase alternators in London with 250 kW power. In 1900 the same company showed a 1.57 MW machine at the Paris

Exhibition [3]. As an example of the worldwide activities Siemens & Halske designed and built the first public power plant of about 4 MW in the vicinity of the coal mines in Brakpan near Johannesburg in South Africa, which went into operation in 1897 [21]. Each of the 3-phase generators generated 975 kW at 700 V. The power was transmitted at 10 kV to various gold mines. A picture of the machine room is shown in Fig. 1.3.

The number of power stations strongly increased with time in the years from 1890 to 1910 and also the maximum power. In that time DC was dominant and in order to store energy for low load intervals accumulator stations were added, allowing to store about 15% of total electrical energy [4]. In the beginning of the twentieth century, the majority of new power stations supplied either alternating or 3-phase AC current. It has to be pointed out that in general the steam electric power stations need, for example, coal for steam generation and hence in total chemical energy is converted via thermal to mechanical and then to electrical energy. For a 100 MW turbo generator the amount of coal needed is enormous: 150 tons of brown coal per hour and 350 tons of steam per hour. Such a power plant was Golpa-Zschornewitz in Germany with 45 MW in 1915 (1918 180 MW), which was later topped by Boxberg (German Democratic Republic) with 3.52 GW in 1966 [9].

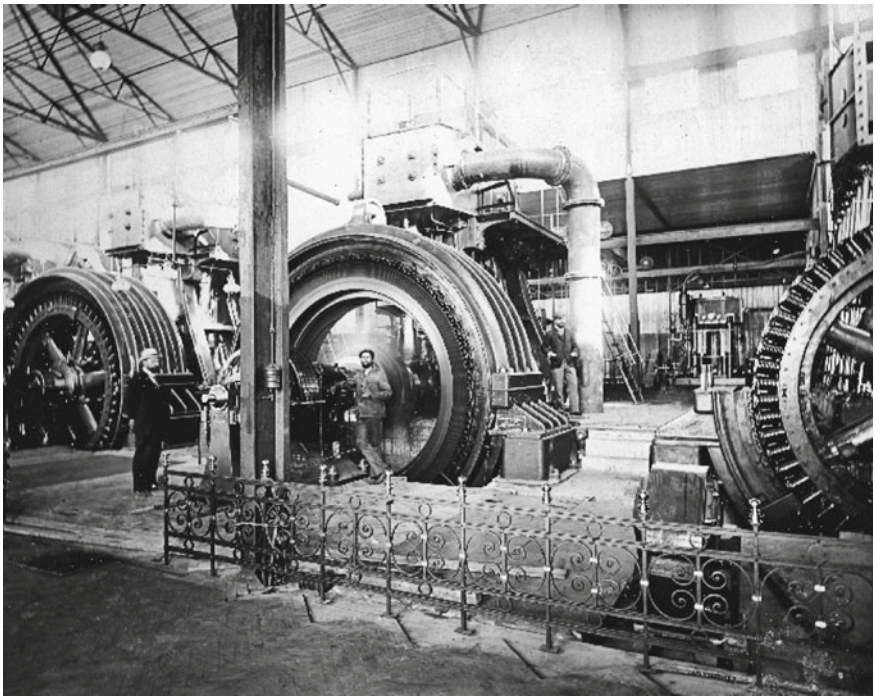


Fig. 1.3 Machine room in 1897 of the power station Brakpan in South Africa, equipped with Siemens & Halske three-phase generators (975 kW at 700 V), coupled directly to the steam engines [21]. Courtesy of Siemens Historical Institute

The first huge water power station was built by Tesla and Westinghouse in 1895 at the Niagara falls, comprising 3 aggregates of turbines and 2-phase dynamo-machines of 4 MW each [5]. In 1924/25 the Walchensee power station went into operation in Germany, with four 3-phase generators delivering 72 MW in total and two single-phase generators of 52 MW [5].

From the power stations the electrical energy was fed into a power grid, distributing it to the end-users, but DC transmission was limited to shorter distances. The first successful long-distance 3-phase transmission took place from Lauffen power station to the Frankfurt Electricity Exhibition over 175 km in 1891 and was realized by Michail von Doliwo-Dobrowolski of AEG (he was born in St. Petersburg in 1862) [4]. In 1925 long-distance transmission started, using 3-phase (rotating) current and high voltages of 110 kV or 220 kV due to lower losses [4].

The first pioneer nuclear power station Obninsk near Moscow was built in the Soviet Union in 1954 and supplied 5 MW. Calder Hall (two reactors) was built in the UK in 1956 and had an output of 69 MW (later on 180 MW); it was followed by Shipping Port in the US in 1957 with 100 MW [10]. In 1974 Biblis A in Germany generated 1.2 GW of electrical power [9], see Fig. 1.4. Nowadays the Kashiwazaki-Kariwa Nuclear Power Plant in Japan with 7 boiling water reactors and a rated power of 8.2 GW is the largest one of the world. It was completed in 1997 [22]. Of course nuclear fission is much more efficient than the burning of coal: 1 atom of U^{235} supplies 200 MeV of energy compared to 4 eV by the chemical reaction of 1 atom C^{12} with O_2, which means 1 g of U^{235} delivers about 2.5 million times more energy than 1 g of coal. In this case the nuclear reactor supplies the thermal energy for the steam–electric power plant and replaces firing of coal [4]. An intermediate drastic improvement step could be nuclear reactors using fast neutrons, with higher reactor temperature and much more efficient use of nuclear fuel, also strongly reducing the amount and decay

Fig. 1.4 Nuclear power plant Biblis in Germany 1974 [26]; Courtesy of Siemens Historical Institute

times of nuclear waste. A promising alternative also is the molten salt thorium reactor, which is currently discussed worldwide. Such a type was tested by Weinberg and his team 1965–69 in the USA (based on U^{233}) and is based on the conversion of Th^{232} to U^{233} by neutron capture. It has several advantages over the nuclear reactor types used so far: Th^{232} is 4 times more abundant than U^{238}, the reactor type is inherently safer, is more efficient, with less radioactive waste, decays faster, is not useful for nuclear weapons, but the initial radioactivity of the waste can be higher [23]. A further approach with about 60% higher efficiency compared to conventional nuclear reactors are fast breeder reactors, which have already a longer history of research and development. Here in 2016, the BN 800 fast breeder reactor in Belojarsk in Russia went into operation with a power of 800 MW. Its predecessor BN 600 started in 1980. The next reactor of this type there, the BN 1200, is under construction [24]. A still more efficient and safer approach would be nuclear fusion reactors. The first prototype of it, the International Thermonuclear Experimental Reactor (=ITER) in Cadarache (France) is planned to show the feasibility with a gain factor $Q = 10$ (fusion energy to plasma energy) around 2035 [25]. An international collaboration as in the ITER project is needed, since the problems in maintaining a high density plasma at temperatures higher than in the interior of the sun for a longer time are tremendous [25].

One should keep in mind that the majority of all power plants: coal, nuclear, geothermal, solar thermal electric power plants, waste incineration plants as well as many natural gas power plants are steam-electric (86%!). Yet it has to be mentioned that hydropower stations have surpassed the largest nuclear power stations in the meantime: the Itaipu Dam water power plant (Brazil, Paraguay) delivers 14.2 GW electric power since 1991 [5, 27], where the dam has created an artificial lake of 1350 km^2 area (total cost 20 billion US$). Besides the flooding of river valleys there are a lot more risks associated with the dam itself, as can be seen from the Sayano-Shushenskaya Dam and power plant in Russia (6.4 GW, 1985), where already several accidents have happened. By the way, since 1961 the new Niagara Falls power station delivers 2.5 GW. Yet the new world record is set by the Three Gorges Dam hydropower plant in China in 2008 with 22.5 GW and an artificial lake of 1000 km^2 area [27].

In Fig. 1.5 the maximum available electrical power per generator/power station is shown in dependence on time. After a slow increase in the seventeenth and eighteenth century, a steep increase starts in the nineteenth century with the industrial revolution. In the time before 1820, the red squares are power estimates partly based on rebuilt devices and our knowledge nowadays. The blue triangles from 1826 on (after the formulation of Ohms law) are based on measurements (data based on [3–5, 9, 20, 21, 26, 27]). The steeper slope of the blue dashed line is the limiting curve for the conversion of finally mechanical to electrical energy based on thermal, hydropower or nuclear power sources. The learning curve for a new technology such as the galvanic cells or Siemens generators may be steeper, but initially starts lower. The availability of electrical power is, of course, a prerequisite for the start of vacuum electronics, together with the development of vacuum technology. The increasing demand for electric energy at the end of the nineteenth century is driven by the demand for lighting, then followed by motors and electric transportation/tramways.

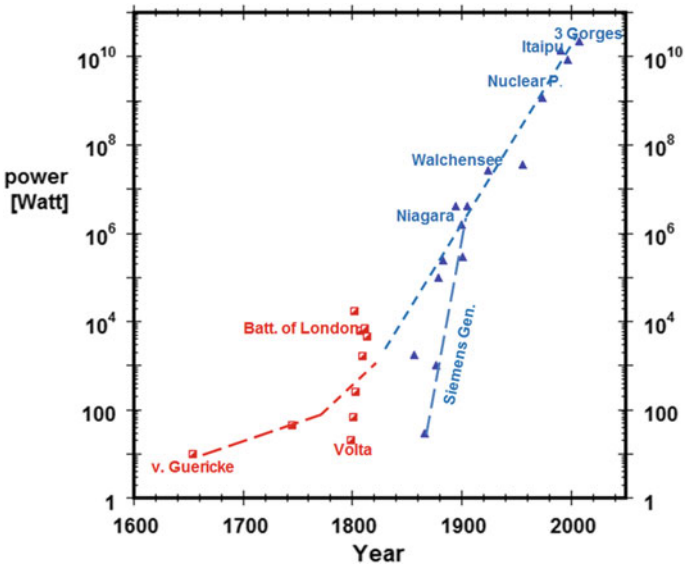


Fig. 1.5 Availability of electrical power versus time (per power generator/power plant): The red lines show the slow increase in available electrical power till 1800, when it was generated via friction or conversion of chemical to electrical energy. The steeper slope of the blue dashed line is the limiting curve for the conversion of finally mechanical to electrical energy based on thermal, hydropower or nuclear power sources. The learning curve for new technology such as the galvanic cells or Siemens generators may be steeper, but initially starts lower. Copyright Georg Gaertner, Aachen, Germany

In this context we will shortly mention alternative sources of energy, which are less risky. They did not play a role in the initial advancement of electrical energy supply, but have become important nowadays. Here the wording renewable energies is wrong, since physicists are well familiar with the conservation of energy: it at least should be renewable or sustainable energy sources, which means permanently available energy supplies, such as wind energy or light from the sun. The maximum rated power of offshore single wind turbines now reaches 8 MW [28–31], with a blade length up to 80 m, onshore values of 2–4 MW are typical. Photovoltaic power plants have been realized up to 1 GW peak (166 MW peak Solarkomplex Senftenberg in Germany; 850 MW peak solar plant near Longyangxia in China) [32]. Yet the electric energy supplied is strongly fluctuating, the average level is much lower than the peak rating, the problem of storage is not solved and their advancement is also linked to strongly increasing area consumption. The strong fluctuations still imply the need for conventional power plants for the baseload [33–35].

It is instructive to look at the following comparison. The area consumption of power plants of different kinds is not only a question of the net basement area used, but also for the required surrounding infrastructure. Let us take the Biblis nuclear power plant as an example: the total electric power available was about 2.35 GW,

assuming the planned Blocks C and D would have been realized then 4.7 GW would have been available from an area of about 0.3 km^2 at the banks of River Rhine (see Wikipedia [36]). If we compare it with wind power area requirements, we refer, for example, to the statistical evaluation of Denholm et al. from the US Department of Energy [37]. In their report they derived an average permanent direct impact area (including permanent clearing area) of 0.3 ha/MW (+ a temporary impact area of 0.7 ha/MW), but a total wind park project area of 34 ha/MW. This larger area is due to the fact that a certain distance between wind turbines is needed to avoid the turbulent flow created by other wind turbines from the initially laminar flow. From the first value for the direct impact one would calculate an area consumption of 14.1 km^2 for 4.7 GW capacity of wind power, but from the distance requirement the area consumption is 1598 km^2, which is already 62% of the federal state Saarland. If one takes into account that offshore at best 20% of the nominal power can be realized, the required area reaches about 8000 km^2, which is half of the area of Thüringen. This should not rule out wind energy as a renewable energy source in the energy mix, but its risks such as killing flying animals, reducing forest area, changing the airflow patterns and an observed drying effect on soil should be taken into account [31, 37].

1.1.2 Milestones in Vacuum Technology

The second basic condition for the rise of vacuum electronics is the availability of vacuum and hence vacuum technology. Already in antiquity Greek philosophers, especially Demokritos (460–370 B.C.), were speculating whether there might exist an absolutely empty space, in contrast to matter (filled by indivisible atoms). It was Aristotle (384–322 B.C.), who claimed that nature will not allow total emptiness and that there is a "horror vacui", which became also the dogmatic belief of the catholic church [38, 39].

Only at the beginning of modern times, with a weakening belief in dogmas, in 1641–1643 in Italy Gasparo Berti and Vincenzo Viviani could explain why suction pumps cannot pump water higher than about 10 m, namely because of atmospheric pressure. Berti first measured this pressure with a water column and demonstrated vacuum above the column. In 1643/1644 then Viviani and Evangelista Torricelli replaced water by mercury in a thin column and invented the Hg pressure manometer [5, 38, 40].

Based on his experiments with the first air pump in 1641, Otto von Guericke first could not remove water from a wooden barrel by pumping, because it was of course not airtight. He then replaced the barrel by two iron half-spheres, which exactly fitted on each other, but only when he used thicker material they withstood air pressure and he was able to evacuate them. This was eventually the first vacuum chamber. In 1654 and 1656 he showed experiments with evacuated half-spheres at the Imperial

Diets (Reichstage) in Regensburg and Würzburg. In 1657 he conducted his famous Magdeburg half-spheres experiment (now with copper) also using a solid piston pump for evacuation and reaching an estimated 13 mbar rough vacuum. 16 horses could not move the evacuated half-spheres apart. The work of Guericke was first reported in the books of Caspar Schott (Professor in Würzburg, whom Guericke had sent his results) in 1657 and 1664 and in an extended version by Guericke himself in 1672 [8, 38, 39, 41].

In 1660 Robert Boyle with the help of Robert Hooke improved the solid piston (air) pump and added a U tube Hg manometer in the vessel and thus reached 8 mbar. These Hg manometers are capable of measuring vacuum down to 1 mbar and with later improvements down to 0.1 mbar [40]. Despite a lot of new insights, for example, the Boyle–Mariotte law for ideal gases, the vacuum achieved was only slowly improved during the next 200 years, as can be seen in Fig. 1.5. Only when solid piston vacuum pumps as used by von Guericke got replaced by mercury piston pumps as first used by the glassblower Heinrich Geissler from Bonn (Germany) in 1855/56, reaching 0.13 mbar, this initiated an accelerated improvement of the ultimate vacuum [38, 40].

In 1865 Sprengel [40] devised a pump in which a train of mercury droplets trapped packets of gas in a glass tube and carried the gas away and thus reached 1.3×10^{-2} mbar. This type of pump was continuously improved in the following decades by several researchers like William Crookes and finally G. Kahlbaum, who reached 5×10^{-6} mbar in 1894 [40].

Of course this also needed improvement of pressure measurement, which was achieved by McLeod in 1874. The McLeod gauge permits pressure measurements down to 10^{-6} mbar. It is based on the compression of the gas by a mercury column to an easily measured higher pressure, and the use of Boyle's law to calculate the original pressure [40].

In the beginning of the twentieth century the German physicist Wolfgang Gaede developed several new types of vacuum pumps, which revolutionized vacuum technology [38]. His inventions were initially triggered by the need for better vacuum for metal surface investigations. It started with the development of the motor-driven rotary oil pump (fore pump) reaching 2×10^{-2} mbar, which was manufactured by the Leybold company from 1907 on. The next step was the invention of the Hg rotary vane pump also in 1905, capable of reaching about 2×10^{-6} mbar using a fore pump as above. Gaede introduced the molecular drag pump in 1913, achieving 4×10^{-7} mbar, and the mercury diffusion pump in 1915 [40, 42]. In 1916, I. Langmuir further increased the pumping speed of the Hg diffusion pump, which is needed for industrial applications, and in 1918 R. Sherwood reached 2.7×10^{-8} mbar with an improved version [42].

For such low pressures the hot-filament ionization gauge was invented by O. E. Buckley in 1916, ranging down 1.3×10^{-8} mbar [43]. In the next 30 years, only slow progress was made concerning the ultimate vacuum. In 1937, Hunt described methods to reach ultrahigh vacuum (UHV, ranging from 10^{-7} to 10^{-11} mbar), including baking and use of getters, but the sensitivity of gauges was not sufficient for pressures lower

than 10^{-8} mbar. In the years from 1935 to 1950, various getters were introduced, which after sealing and activation, helped to further pump down the tubes during operation. It was finally recognized that the limited sensitivity of the gauges was related to the creation of soft X-rays at the collector and a superimposed photoelectron current [42, 44]. A breakthrough in pressure measurement sensitivity was the Bayard–Alpert gauge, which lowered the X-ray limit drastically by reducing the surface area of the collector. It was invented by R. Bayard and D. Alpert and is able to measure down to 10^{-11} mbar [44]. Improvements in pumping soon followed. The molecular pump of Gaede was improved in the form of a multistage turbo-molecular pump by W. Becker of the company Pfeiffer Vakuum in 1958, with attainable vacua now in the range of 10^{-10} mbar [45]. Also in 1958 L. Hall of the company Varian introduced the ion getter pump, which is capable of reaching 10^{-10} mbar after first pumping down with a high vacuum pump and baking the vacuum chamber [46]. Further improvements of the Bayard–Alpert design reduced the X-ray limit further and allowed to measure pressures below 10^{-12} mbar. The years from 1950 to 1970 were very fruitful years for vacuum science and technology. Residual gas analysis was introduced by W. Paul by application of quadrupole mass filters. The ultimate vacuum was further reduced. Already Hobson [47] reported 1×10^{-14} mbar in a small glass system cooled to 4.2 K, measured with a Bayard–Alpert gauge. This was again reached by W. Thompson and S. Hanrahan in 1977. XHV results of 4×10^{-14} mbar were obtained by C. Benvenuti in 1977 and 1993 [44, 48]. In 1989, H. Ishimaru of Japan obtained 5×10^{-13} mbar by using a turbo-molecular pump for pumping down, careful baking the Al chamber and maintaining XHV by two ion getter pumps and a titanium sublimation pump. This is the lowest value for chambers at room temperature. For the measurement, he used a point collector gauge [49].

The lowest pressure reported so far, namely 6.7×10^{-17} mbar, was determined indirectly from the storage of anti-protons in a Helium cooled Penning trap by G. Gabrielse et al. in 1990 at CERN [50]. In this application of cooled Paul ion traps top results were also achieved recently by Micke et al. in 2019 [51] and by Schwarz et al. in 2012 [52], both claiming a pressure range of 1×10^{-15}–1×10^{-14} mbar (1×10^{-13}–1×10^{-12} Pa). For accelerator applications also a new pump type has been introduced, such as linear non-evaporable getter (NEG) pumps, explained in the review of Benvenuti [53].

The ultimate vacuum reached as a function of time is shown in Fig. 1.6, which is an update of G. Gaertner (see [2]) based on [2, 42, 44, 48, 51, 52]. It is a typical development function with some prominent milestones, showing a slow decrease of the ultimate vacuum reached (logarithmic scale) from 1650 to1850 (blue broken line), then followed by a much steeper decrease from then on, (red broken line, consistent with an exponential fit) partly motivated by the improvement drive of incandescent lamps and later radio tubes and CRTs. There is an indication of the progress currently slowing down, since all pressure values below 10^{-13} mbar were obtained with cryo-cooled ion traps.

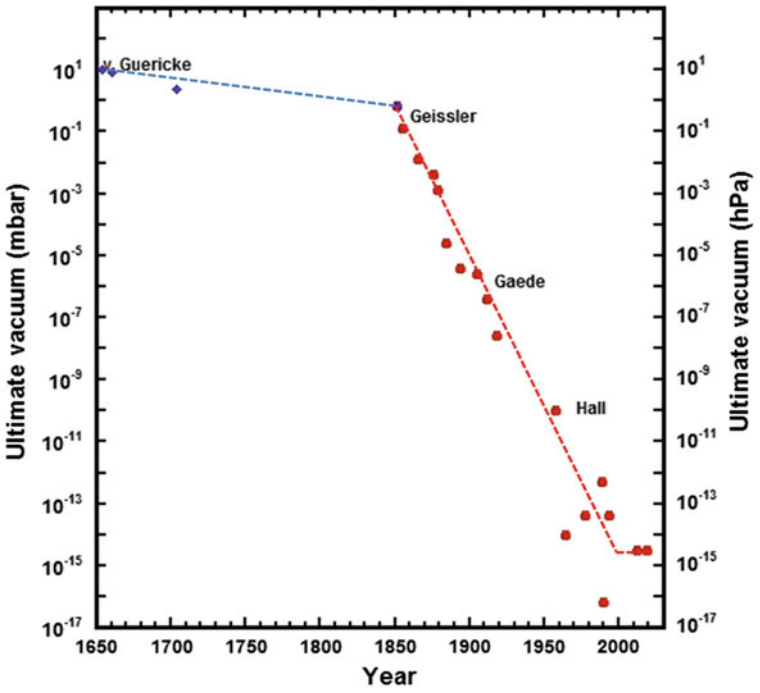


Fig. 1.6 Ultimate vacuum achieved (log scale) versus time; update of Gaertner [2] and also based on [42, 44, 48, 51, 52]

1.2 Historical Development of Vacuum Electron Tubes

1.2.1 The Rise of Incandescent Lamps

The development of vacuum tubes was first driven by lamp applications not yet suited for practical use. In 1840 Robert William Grove from the UK invented the incandescent lamp, which was further improved by J. W. Starr from the US in 1845 (US patent) and by Joseph Swan from the UK in the years 1860–1878, who got a UK patent on it in 1878 [5]. In 1854 the German mechanic Heinrich Goebel built incandescent lamps using carbonized bamboo fibers in his shop in New York with a claimed lifetime of up to 400 h. The industrial prospects of the practical application culminated in a lawsuit, where also the inventors Edison and Swan were involved. Goebel finally got a US patent on his lamp in 1893 shortly before he died [5, 9]. Yet Edison had pioneered the industrial application and sold light bulbs for 50 cents already in 1883 and the industrial production of incandescent lamps started and then

experienced a continuous rise in production numbers in the following years. The competition and the patent suits between Swan and Edison at the end lead to the foundation of the Edison and Swan Electric Light Company [9].

Figure 1.7 shows the annual production rate of incandescent lamps and electron tubes in the US from 1880 to 1940 as given by P. A. Redhead in his instructive review [54]. Since the luminous efficacy of incandescent lamps was doubled to 10 lm/W by replacing carbon filaments by tungsten filaments, they became the dominant lamp type from 1910 on. We also see that electron tube production started slowly in 1905 and made use of the existing vacuum technology for mass production of incandescent lamps, which were still produced as vacuum lamps. This will be addressed in the next paragraph. The technological cycle of incandescent lamps continued for the next hundred years and only recently started to decline due to the advent of energy-saving LED lamps [2].

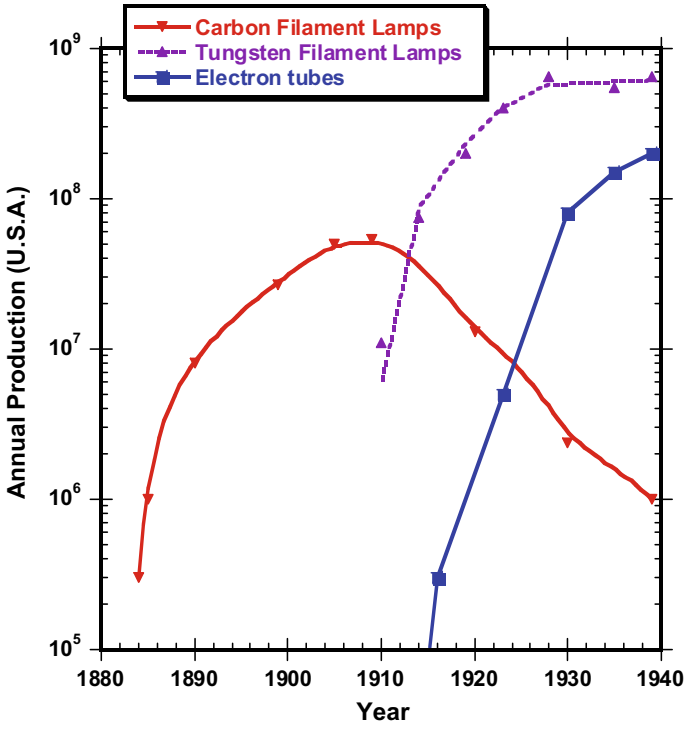


Fig. 1.7 Annual production rate of incandescent lamps and electron tubes in the US till 1940 according to Redhead [2, 54]. Reprinted with permission from J. Vac. Sci. Technol. B 30, 060801 (2012). Copyright © 2012, AVS

1.2.2 The Early History of Vacuum Tubes and the Radio Tube Era

The development of the first vacuum electron tubes started a bit later than lamp development. In 1854 the glassblower Heinrich Geissler from Bonn in Germany invented the Geissler tubes with platinum wire feedthroughs, which he evacuated with his mercury pump to some mbar residual gas pressure. By using such a Geissler tube with 2 electrodes and a Ruehmkorff inductor, the German physicist Julius Pluecker invented the gas discharge lamp [5, 9]. In1859 Pluecker experimented with invisible cathode rays, which he called "glow rays" or "Glimmstrahlen", later together with Hittorf [5]. The name "cathode rays" or "Kathodenstrahlen" was first introduced by the German physicist Eugen Goldstein in 1876. In 1878 the Englishman Sir William Crookes was the first to confirm the existence of cathode rays by displaying them, with his invention of the Crookes tube, a crude prototype for all future cathode-ray tubes [5, 55]. In 1897 J. J. Thomson proposed that the cathode rays were composed of negatively charged fragments of atoms that he called "corpuscles" and which we call electrons [56].

In 1883 Thomas Alva Edison from the USA discovered current flow to a collector plate in a light bulb and got a patent on such a diode for voltage regulation. Now the progress accelerated in view of industrial applications [54].

When experimenting with focused high-energy cathode rays, Wilhelm Conrad Röntgen discovered the X-rays in 1895. The today form of the X-ray tube was pioneered by William Coolidge (USA) in 1908, who used a thermionic tungsten wire cathode surrounded by a Wehnelt cylinder as electron source [5, 9].

In 1897 the German Karl Ferdinand Braun invented the CRT oscilloscope using a sideways anode, deflection plates, a magnetic focusing field and a phosphor screen—the Braun Tube was the forerunner of today's television and radar tubes [5, 9].

In 1904 A. Fleming first introduced diodes for detection and rectification of high-frequency electromagnetic waves in telegraph receiver stations in the UK, parallel to A. Wehnelt with rectifier tubes in Germany. This detection turned out to be much more efficient than the "coherer" used before. The next step was the triode, which was introduced by the addition of a grid in 1906 by Lee de Forest in the USA parallel to Robert von Lieben in Austria and was used for amplification of weak electrical signals by modulation of much larger electron current. The tetrode with 2 grids was first suggested by the German Walter Schottky in 1916. The pentode was patented by Bernard Tellegen of Philips in the Netherlands in 1926 [57]. In 1920 medium-wave radio transmission had started in the USA. Now after a long phase of basic inventions, the industrial applications started to take off [54]. The number of grids was further increased and tubes with up to nine grids were built. Besides amplification, they also could be used as oscillators, first shown by A. Meissner in Austria 1913 [9].

The fields of tube applications increased from valves in receivers for wireless telegraphy and telephony, from telephone repeaters to radio receivers and radio transmitters.

An interesting development was the Loewe multi-tubes ("Loewe-Mehrfachröhren"), where several (up to three) tubes were put in one bulb including passive components such as resistors and capacitors [2]. As the first integrated tubes they were a forerunner of integrated circuits, but were not successful commercially. Yet double triodes or double pentodes in one tube were quite common. In the late 1930s, several thousand different types of radio tubes were manufactured in Europe and the US by various companies. During World War II, radar applications (also using RF magnetrons) became very important.

Figure 1.7 shows the annual production rate of incandescent lamps and electron tubes in the US from 1880 to 1940 as given by P. A. Redhead in his instructive review [2, 54]. The electron tube production started slowly in 1905 and made use of the existing vacuum technology for mass production of incandescent lamps. At the beginning of the twentieth century, tube developments were dominated by radio tubes, which were used as multi-grid tubes in transmitter and receiver sets. During World War I, they got the first boost through military applications. In 1917, half a million tubes were produced in the US, mostly used in stationary amplifiers for wire telephony. In France in 1918, 300,000 of such tubes were manufactured. By 1932, the company Philips alone had produced over 100 million tubes (about 15 million in 1932) and 1 million radio receivers so far and at that time was the world's largest manufacturer of radios and Europe's largest manufacturer of radio tubes [58, 59]. In the years before, Philips had acquired Valvo in Germany (1925) and 50% of Mullard in the UK (1924) [60]. Other important tube companies were. for example, RCA, Westinghouse, GE, Western Electric in the US, GEC in the UK, AEG Telefunken and Siemens in Germany, Tungsram in Hungary and Tokyo Electric (Toshiba) in Japan. Figure 1.8 illustrates the dimensional range of Philips pentodes from 1937 (which were more efficient than tetrodes) used for transmission, ranging from 15 W (to the left) to 15 kW (PA 12/15 to the right, with a length of 61 cm) and a frequency $\leq$20 MHz. An even larger Philips transmitting tube was the TA 20/250 with 250 kW and length including water cooler of 1.4 m [61]. In World War II radar applications became decisive. Radio tubes reached their culmination in the 1940s–1960s, where about 1.2 billion tubes were produced per year around 1950 [1, 55, 57, 62]. Receiving tubes found a new application in television sets. The history of radio tubes is presented in more detail in the books of G. E. J. Tyne ("Saga of the Vacuum Tube" [62]), J. W. Stokes ("70 Years of Radio Valves and Tubes" [63]) and S. Okamura ("History of Electron Tubes", [57]). Unfortunately, the book of Tyne only covers the time till 1930, but gives a good overview of the types of tubes produced and also on tube company history with detailed references. The book of Stokes is mainly centered on the USA and the UK including major companies and covers the whole radio tube era with detailed references. The overview of Okamura mainly deals with Japan and the USA, includes electron tubes in general and gives also some commercial information, but only gives few references.

During the radio tube era also the cathodes used were improved: due to tube pressure at best in the range of pre- to high vacuum in the early years, tungsten and later thoriated tungsten were the thermionic cathodes primarily used, since they are less sensitive to emission poisoning by residual gases. With further improvements

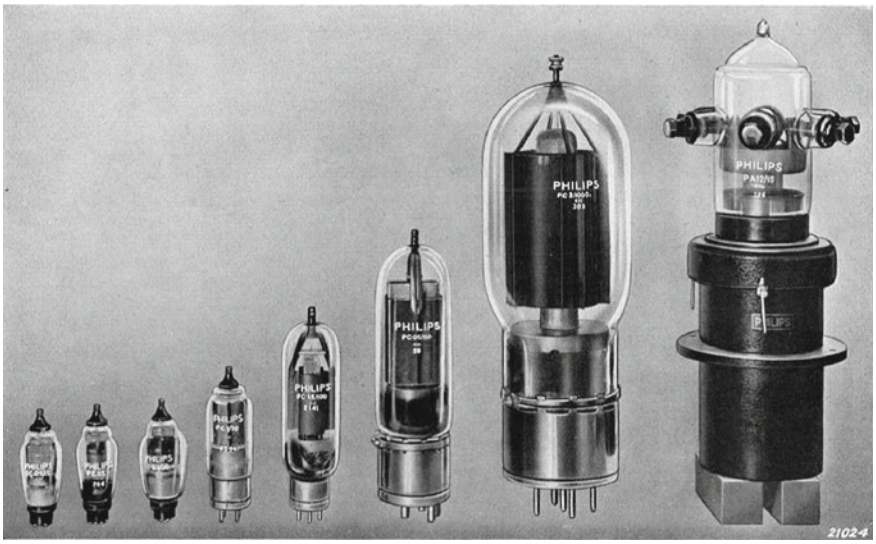


Fig. 1.8 Series of Philips pentodes from 1937 with outputs from 15 W (the smallest) to 15 kW (water-cooled PA 12/15 to the right, the length of which is about 61 cm; H. G. Boumeester, Philips Technical Review 2/4 (1937) 115–121, Fig. 6 [61]. Copyright Royal Philips N.V

of vacuum techniques by W. Gaede, I. Langmuir and others, electron tubes with oxide cathodes, first introduced by A. Wehnelt in 1904 [2, 64, 65], became feasible. From 1920 on the commercial application of oxide cathodes spread rapidly due to the advantage of operation at lower temperatures ("dull emitters" instead of W and Th/W "bright emitters"), and further enabled by the introduction of getters and the use of alkaline earth carbonates. This led to an oxide–cathode monopoly in radio-receiving valves, but they could not be used for high-voltage applications (>5 kV) due to arcing. The emission efficiency of a directly heated oxide cathode with about 40–100 mA/W was much higher than that of thoriated tungsten (6–30 mA/W) and tungsten cathodes (1.7–4 mA/W) in commercial receiver and transmitter valves in the 1930s. The efficiency values, of course, depend, for example, on rated life, emitter geometries and emission capability, especially of improved later cathode versions. Boumeester of Philips in 1937 gives the values of 3–8 mA/W for tungsten cathodes, 80 mA/W for thoriated tungsten and 200–300 mA/W for (earth alkaline) oxide cathodes [61]. Cathode and material knowledge at the end of the radio tube era was wrapped up in the books of Fred Rosebury ("Handbook of electron tube and vacuum techniques", American Institute of Physics 1993, originally MIT 1964 [60]) and of Walter H. Kohl ("Materials and techniques for vacuum devices", Reinhold Publishing Corp., New York 1967 [64]).

Then in the 1960s, electronic tubes were gradually replaced by transistors and the era of the radio valves approached its end (Fig. 1.9) with the development of Si technology and ICs at Texas Instruments and Bell and their commercialization. Yet

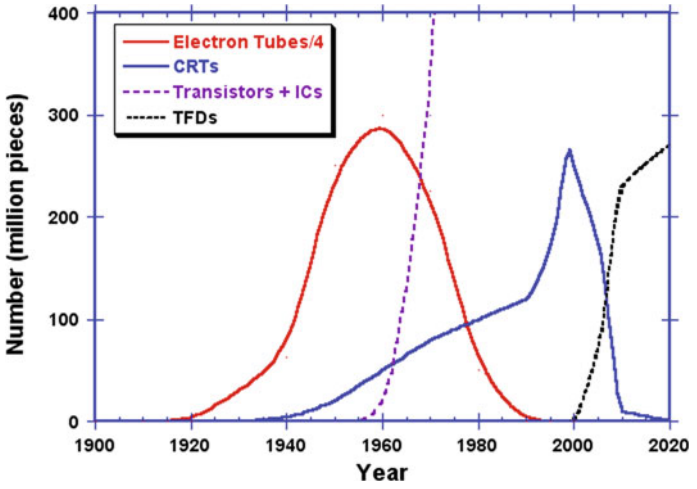


Fig. 1.9 Historical trends/technological waves in vacuum electronics and neighboring fields according to [2]. Reprinted with permission from J. Vac. Sci. Technol. B 30, 060801 (2012). Copyright © 2012, AVS

the first computers were still based on tube technology and were quite bulky, but had no chance on the way to miniaturization.

1.2.3 The Technological Cycle of Cathode Ray Tubes

During the zenith and decline of the radio tubes, the production of cathode-ray tubes (CRTs) for displays started to rise. This is illustrated in Fig. 1.9, where CRTs appear as the second technological vacuum-tube wave [2]. The picture tube was the basis of common television from the beginning around 1939 until about 2007 and experienced a continuous increase in production numbers after 1945. It was a technological extension of radio tube technology, based on the multi-gridded electron gun, where the magnetic deflection of the electron beam was added, with a fast horizontal sweep and vertical stepping, the beam spot landing on the phosphor dotted screen. Soon new features were introduced, such as color (using three guns for red, green and blue phosphor dots), increase of screen size, changes in format, and higher brightness and resolution [1, 66–70, 59]. Figure 1.10 shows a schematic cross-section of a television tube (TVT). The essential parts are the electron gun with the cathodes (three cathodes in the case of a color TVT for the colors red, green and blue), the magnetic yoke (coils) for the beam deflection, the shadow mask and the phosphor screen, where the pictures are displayed in a 50–100 Hz sequence. The cathodes used in CRTs were mainly oxide cathodes, the rest of about 10% were impregnated cathodes used for higher loads. Philips alone produced about 200 million oxide cathode units (0.65 W units) per year at the zenith of CRT production. The electron

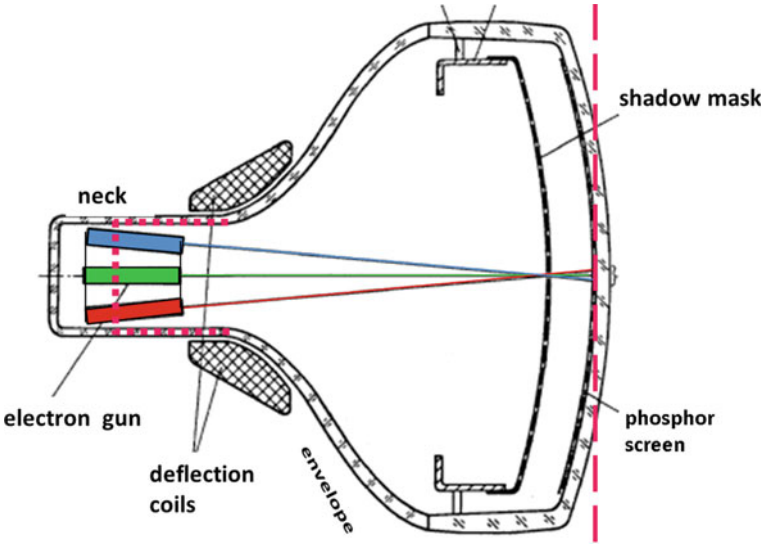


Fig. 1.10 Schematic diagram of a color picture tube. The electron gun comprises the three cathodes for red, green and blue. *Source* Handout of Philips Display Components, Eindhoven 1995, Copyright Royal Philips N. V

gun part could be separately tested in the form of dummy tubes and is essentially based on radio-tube technology. In the combat with flat displays, the last improvements were super-flat screens and super-slim tubes with a shorter electron gun, e.g. by LG. Philips Displays. High-definition television was also realized in prototypes at Philips and other companies already in 1996. The additional peak in Fig. 1.9 after 1990 was due to the application as computer monitors, culminating in 260 million CRTs sold in 1999 [2, 58]. Before the onset of color monitor tube (CMT) production, the seven largest CRT and also CTV set producers in the world in 1987 were Philips (the Netherlands, 8.6 million sets/year), Thomson (France, 6.8 million sets/year), Matsushita (Japan, 4.7 million sets/year), Sony (Japan, 3.8 million sets/year), Toshiba (3.2 million sets/year), Hitachi (Japan, 3.1 million sets/year) and Samsung (South Korea, 2.5 million sets/year) [70]. Philips had acquired Sylvania/GTE from the USA in 1980. In 1987 Thomson had taken over the radio and TV business from GE and former RCA. In 1999 Zenith (USA) was bought by LG from South Korea. Due to these concentrations and the rise of Chinese producers, the picture had somewhat changed in 1999: in television CRTs Philips (17.5% by value) and Thomson (12.8%) were still heading in market share, followed by Sony (12.5%), Matsushita (12%), Samsung (11%) and LG (7.8%). But in monitor CRTs Samsung had taken the lead (19% by value), together with Chunghwa Picture Tubes (CPT, 17.5%) and LG (15%), followed by Sony (9.5%), Philips (8.5%) and Hitachi (7.5%) [67]. The decline began in 2001, when color monitor tubes (CMTs) were phasing out, being replaced by flat and thin liquid-crystal displays (LCDs). Unlike the end of the radio tube era, where production sites could be, for example, switched to TV tube production due

to similar technology, LCD technology was disruptive and, moreover, too expensive for production in Europe or the US. Therefore, the CRT companies used several strategies for survival in a shrinking market. There already existed a lot of cross-license agreements. At that time the largest TVT producers were the newly founded LG. Philips Displays CRTs with market shares of about 24% (by number of CRTs), Samsung SDI with 21%, and Matsushita, Sony and TTD (former Thomson) with about 9% each. Concerning CMTs, Samsung SDI was leading with about 27%, followed by LG. Philips Displays with 24% and Chunghwa Picture Tubes (CPT) with 20% [68, 69]. In the years after 2000, CRTs experienced severe price erosions of up to 30% for CMTs and up to 15% for TVTs, exerting a lot of cost down pressure. Factories had to be closed down or shifted to countries with cheaper wages. This was especially a problem for factories/producers in Europe, where a lot of workers lost their jobs, since also the LCDs were much cheaper to fabricate in East Asia. Hitachi stopped its CMT production in 2001 and its TVT production in 2002. Former rivals merged their business, after LG and Philips now Matsushita and Toshiba in 2002. LG. Philips Displays CRTs went bankrupt in 2006 and factories in Europe were closed down. In this scenario of a dying industry, the post-mortem anti-trust fine of the EU commission in 2012 [71] was incommensurate, since it was not linked to the damage and hence the profits of the cartel (but instead to the turnover; the cartel already starting in 1996/1997), did not help the people in Europe who had lost their jobs years ago, and maybe helped to shrink the European company Philips further. Of course cartel formation cannot be tolerated, but with respect to CRTs after 2000 it seemed to be similar to monitoring the water in a sinking boat by the passengers. In the end the cartel was inefficient, did not stick to "agreements" and could not stop the inevitable loss of profitability. Ironically enough, one of the founders of the cartel and whistleblower was not fined, but sentenced in the US for another LCD cartel. Only companies with sizeable LCD production could remain in the TV and color display business.

Despite the lower price of television tubes, the replacement by flat-screen devices, such as LCDs or plasma panels, followed in the next years and LCDs became dominant in 2007. New features such as high and super-high definition television, much larger screen sizes and three-dimensional televisions were introduced.

1.2.4 The Continuous Progress of Other Noncyclic Vacuum Electron Tubes

Not affected by the decline of radio tubes and later of cathode-ray tubes (CRTs) were other types of vacuum tubes experiencing a continuous progress, such as X-ray tubes, microwave tubes at high powers and high frequencies, electron beam devices for materials characterization and processing, ion propulsion systems, particle accelerators and thermionic converters. Most of these are described in more detail in the

book "Vacuum electronics" edited by Eichmeier and Thumm [1] and in [2]. It is important that vacuum electron devices (VEDs) are superior to solid-state devices in the high-power and high-frequency domain, also with respect to reliability and life, as pointed out in [2, 56], and hence there is no threat of replacement. In [72, 73] J. H. Booske gives a limit of "power density" $p_{\text{avg.}} \times f^2 = p_{\text{avg.}} \times (c^2/\lambda^2)$ (used as a figure of merit; $f^2 = (c/\lambda)^2$ serves as a measure of the reverse interaction area) with respect to cost and reliability between commercial solid-state amplifiers and vacuum electronic amplifiers of

$$p_{\text{avg.}} \times f^2 = 4\left[\text{kW GHz}^2\right] \tag{1.1}$$

Above this line vacuum electronic devices (VEDs) are superior in the high-frequency, high-power region. There are three reasons for it: in solid-state devices (SSDs) the electron current generates heat via collisions in the solid, which is not the case for an electron beam in vacuum. Second, voltages in an SSD are limited to avoid breakdown. These voltage limits are much higher for VEDs. Third, VEDs can be operated at higher temperatures than SSDs. By the way, there exists also a limit in scaled "power density" for vacuum electronic devices according to J. H. Booske of

$$p_{\text{avg.}} \times f^2 = 100\left[\text{GW GHz}^2\right] \tag{1.2}$$

which is more than 7 orders of magnitude higher than the relation for SSDs. High-magnetic-field gyrotrons and free-electron lasers (FELs) approach this line, best shown in a double logarithmic plot of power versus frequency as by Booske [72, 73]. There is also a progress of this figure of merit of VEDs over time, but mainly due to the introduction of new tube concepts.

In Fig. 1.11 total annual sales (in billions of €) are depicted versus time for three important vacuum tube types, namely microwave tubes, X-ray tubes and lamps (fluorescent/CFL and incandescent; converted update of [1]). Phasing out of incandescent lamps has started due to national energy savings legislation. The rise of LEDs is also shown, data here are without fixtures, controls and without car applications [74–78]. One can also see that the world market for microwave tubes and X-ray tubes remains nearly constant over time.

Besides the boost in vacuum technology and cathode technology induced during the radio tube era and CRT tube era, also these baseline applications need and will trigger further continuous improvements. Advanced requirements to cathodes are set up for the high-power/high-frequency region of mm-wave devices, not only for terahertz imaging, but also for electron beam lithography, for particle accelerators and for thermionic converters, which we will discuss in the next paragraph.

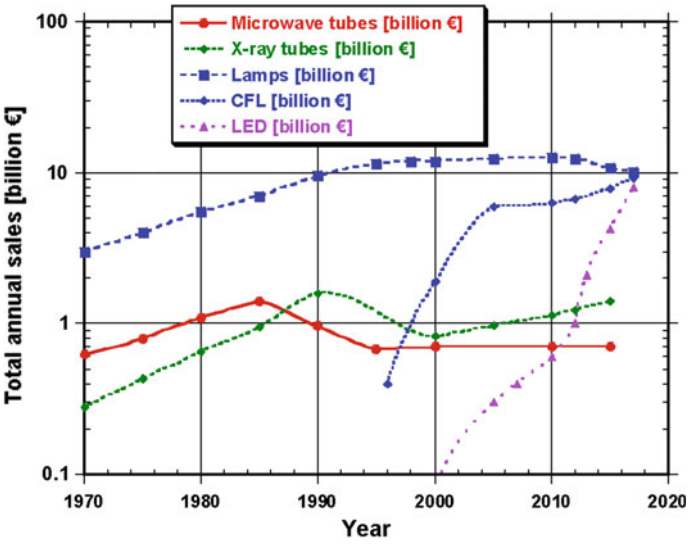


Fig. 1.11 Total annual sales (in billions of €) versus time for three important vacuum tube types, namely microwave tubes, X-ray tubes and lamps (fluorescent/CFL and incandescent, compare [1]). Phasing out of incandescent lamps has started due to national energy savings legislation. The rise of LEDs is also shown, data here are without fixtures, controls and without car applications (lighting world market see [74–78])

1.3 Historical Development and Improvement Directions of Modern Vacuum Electron Sources

During the progress of vacuum electronics, we also see continuous improvement of the emission capability of thermionic cathodes from 1910 till 2020, here, as shown in Fig. 1.12. In the last century, practically all the cathodes used have been thermionic cathodes, where reviews have been given in [1, 79–84]. The leveling-off after 2004, especially of oxide cathode improvement, is mainly due to the reduction of research efforts after the decline of the CRT era. Due to tube pressure at best in the range of pre- to high vacuum (10^{-3} to 10^{-7} mbar), tungsten and later thoriated tungsten (Th-W) were the thermionic cathodes used at the beginning of the twentieth century for more than a decade, since they are less sensitive to emission poisoning by poisonous residual gases, which especially holds for tungsten. This is the main reason why tungsten is still used in X-ray tubes (see Chap. 5 of R. Behling). But also Th-W emission could be increased by improved vacuum and controlled carburization of the tungsten base before activation [1, 80].

With further improvements of vacuum techniques by W. Gaede, I. Langmuir and others, electron tubes with oxide cathodes became feasible in the 30s [65]. They continued to be used in CRTs, since due to preparation by spray coating of the carbonates on indirectly heated Ni caps they are rather cheap in production. Although they have a very low work function of about 1.5 eV and hence much lower operating

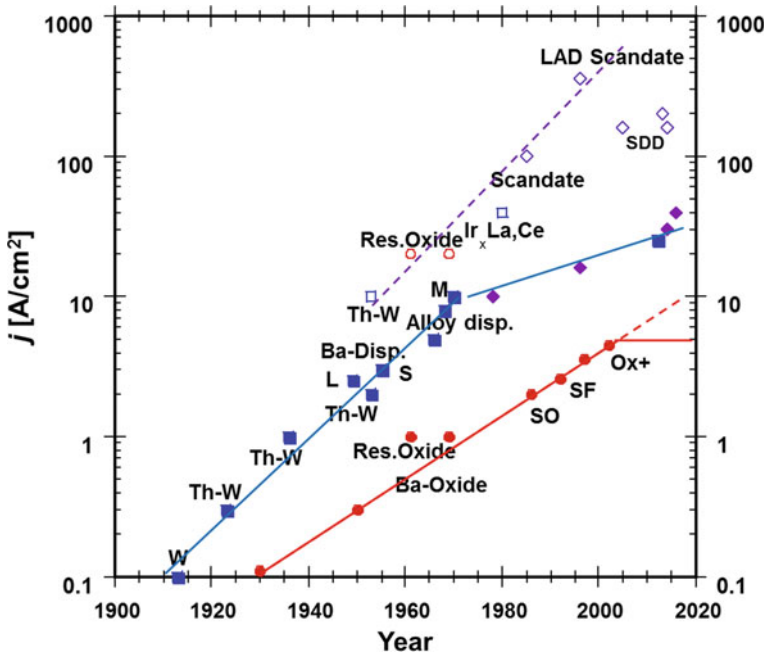


Fig. 1.12 Historical development of thermionic cathode emission capabilities—an update. The plot shows on the vertical axis the maximum current density achieved for specific cathodes versus year of achievement on the horizontal axis. The lower red line shows the evolution of DC emission of oxide cathodes (red circles). The upper blue line interpolates the DC emission of metal matrix-based cathodes including Ba dispenser cathodes (blue squares) and Ba scandate cathodes (violet diamonds). (Condition is an operational lifetime ≥20,000 h at the given DC emission current density j_{dc}). The open symbols are pulsed emission data for different cathodes including Ba scandate cathodes, where the life requirement there is ≥ 4000 h; compare [1, 2]. Explanation of the abbreviations is given in the text.

temperature than W and Th-W, they are only suitable for lower DC current densities due to their low electrical conductivity as semiconductor-based material. Yet also their DC loadability has been steadily increased in the course of time, which can be seen in the lower red line in Fig. 1.12. This improvement in performance was not only due to improved vacuum, but was achieved by different forms of doping with rare earth oxides (SO and SF cathodes) especially with Eu_2O_3, Y_2O_3 and Sc_2O_3, also by the addition of Ni particles to the porous BaO·SrO coating (oxide plus cathode of Philips). Despite increased emission and life, the main improvement direction is a further increase in the DC loadability. A further modification was reservoir oxide cathodes investigated by Lemmens and Zalm [85] and later by G. Medicus and others [1], which did not really become competitive. More details on oxide cathodes can be found in Chap. 4.

The next boost with increased emission current density came by the introduction of reservoir and impregnated Ba dispenser cathodes in the 1950s, which could be

used in tubes with ultra-high vacuum. This innovation in the cathode field was mainly achieved by the Philips company (linked to CRT research and development), as can be seen from the invention of the L-cathode in 1949 by H. Lemmens et al. (reservoir Ba dispenser cathode) and of the first impregnated (I) Ba dispenser cathodes (S, also B cathode) by H. Levi in 1955 [1]. In 1966, P. Zalm et al. introduced the so-called magic (M) cathode with an Os/Ru top layer on a tungsten base, impregnated with Ba–Ca–aluminate [1]. Similar improvements were obtained with top layers of Ir and to a lesser extend of Re, and also with alloys of the elements Os, Ir, Re, W in form of a coating or as constituents of the base matrix [80]. These Ba dispenser cathodes are the subject of Chap. 2 of this book and are mostly used in long-life microwave tubes for space applications. The lifetimes as given in Table 1.1 do depend not only on DC load and temperature, but also on the design, for Ba dispenser cathodes especially on the pellet thickness (impregnant reservoir), see [86]. The performance data of different types of thermionic cathodes applied in commercial tubes are listed in Table 1.1 in the sequence of introduction. Lifetime is specified for the DC current densities given, which are the limiting condition especially for oxide cathodes. These data are partly also the basis for Fig. 1.12.

Other variants are the so-called metal alloy dispenser cathodes such as Re_2Th, Pt_5Ba, Pd_5Ba, Ir_2La, Ir_5Ce, which were pioneered by B. Djubua and colleagues from Russia [88] and rely on monolayer films of Th, Ba, La or Ce, respectively, forming on the surface of the alloy and reducing its work function. Especially Ir_2La and Ir_5Ce with work functions of about 2.2 eV are capable of delivering 100 A/cm^2 pulsed for 1000 h and 10 A/cm^2 pulsed for 10,000 h [88].

A variant of the Ba dispenser cathodes promising still higher emission capability are the scandate cathodes, first introduced by L. Figner in 1967 in Russia in the form of

Table 1.1 Performance data of different types of thermionic cathodes applied in commercial tubes in the sequence of introduction. Lifetime is specified for the DC current densities given

Cathode type	Abbreviation	T_{op} [K]	J_{dc} [A/cm^2]	$e\Phi$ [eV]	Lifetime	References
Tungsten	W	2520	1–4	4.5	>1000 h	Chapter 5
Thoriated tungsten	Th-W	2000	2–4	2.6	10 kh	[1, 80]
Ba oxide	Ox	1050	1–4	1.4	20 kh	Chapter 4, [1, 80]
Ba dispenser, W base	W-I or S or B	1300	3	2.05	≥20 kh	Chapter 2, [1, 80]
Ba dispenser, Os/Ru coating	M or Os/Ru-I	1300	10	1.85	20–130 kh	Chapter 2, [1, 80, 86]
Impregnated Ba scandate	Imp. Scand.	1300	15	1.7	≥10 kh	Chapter 3, [1]
Ba scandate: Sc_2O_3 doped disp.	SDD	1320	40	1.45	≥4000	Chapter 3, [87]

the pressed scandate cathode, then followed by the impregnated scandate and mixed matrix and top-layer scandate cathodes investigated mainly by Philips and Hitachi [1]. Philips Research improved scandate cathodes continuously to space charge limited emission current densities higher than 360 A/cm^2 (pulsed) by introducing top-layer scandate cathodes prepared by laser ablation deposition (LAD) [89], but never introduced them in their CRT tubes for cost, lifetime and other reasons, which will be discussed in Chap. 3 of this book. Yet there is need for these thermionic cathodes with very high emission capability in advanced new applications as in high-power gyrotrons, in terahertz imaging devices and for electron beam lithography. In Beijing the group of Yiman Wang has continued research on scandate cathodes, especially nanometer-size scandia particle doped Ba dispenser cathodes (SDD) [90, 91], and reaches now top DC current densities of 40 A/cm^2 with a lifetime of larger than 10,000 h (test still continuing) [87], as we can also see in Fig. 1.12. It has to be pointed out that in this figure the quoted DC current density is always linked to a minimum lifetime >10,000 h (for CRTs ≥20,000 h) and the pulsed current density to ≥4000 h. Of course for space applications 130 kh at 10 A/cm^2 have been achieved with specially designed Ba dispenser cathodes as shown in Chap. 2. The progress in lifetime at a certain emission level is not addressed here and will need a different diagram.

Most of these improvements are due to the introduction of cathodes with lower work function and hence higher emission current density j_e, which can be seen from the Richardson Dushman equation (for derivation, see [1]), which is usually written in the form

$$j_s = A_{th} T^2 \exp(-e\Phi/kT) \tag{1.3}$$

where $A_{th} = 120.4$ Acm^{-2} K^{-2} is the thermionic constant for the ideal case, $e\Phi$ the work function and j_s the saturated emission current density, valid for zero extraction field at the surface (hence it is also called zero field thermionic emission). In reality the thermionic constant in (1.3) is replaced by the phenomenological Richardson constant A_R, which in most cases is smaller or much smaller than A_{th} and is characteristic for a certain type of cathode. For values $j_e < j_s$ the space charge field of the electron cloud in front of the surface is not compensated by the external extraction field (i.e. the Laplace field between cathode and anode) and the Child-Langmuir equation for space charge limited emission holds, which is in the first approximation independent of the temperature, but depends on the Laplace field strength U_a/D

$$j_e = 4/9\, \varepsilon_0 \sqrt{(2e/m_e)} D^{-1/2} (U_a/D)^{3/2} = K\, U_a^{3/2} \tag{1.4}$$

The geometry factor $K = 2.33 \times 10^{-6}/D^2$ is given in units of A/(cm^2 × V$^{3/2}$), where D is the cathode to anode distance in cm and U_a the anode voltage in V. There arise further corrections of space charge limited emission by taking the electron velocity distribution into account, which modify (1.4) (see [1]), but will not be discussed here.

Besides a reduction of the work function of thermionic cathodes and hence increased usable space charge limited current density, increased lifetime at the operating temperature and an increased emission uniformity over the emitting surface are further requirements. Since lower work function in most cases implies a lower operating temperature and a higher gas-poisoning sensitivity, ultra-high vacuum is needed in the electron tubes. In meeting these aims, thermionic cathode research concentrates on modification of known materials by additions, on investigation of new materials, on high-resolution characterization and modification of cathode surfaces including modern thin-film deposition methods, and design of resupply reservoirs for long life.

For cathode applications, for example, in microwave tubes, it is instructive to know that the achievable output power P_{out}, for example, of a klystron is proportional to the electron emission current I_e times of a factor dependent on anode voltage and magnetic field. Interestingly, the cost per tube then scales to a good approximation with P_{out}, the factor being in the range 1.4–3 $ per watt [2]. For space applications of microwave tubes, a long lifetime of these cathodes in the order of 10 years or more at the operating current density is needed. Cathodes used for these applications are typically Ba dispenser cathodes. Yet the higher the current density and hence the temperature, the shorter the lifetime. Here also Ba dispenser cathodes with lower work function are needed. Booske [73] has shown that for closing the terahertz gap between about 200 GHz and 3 THz, vacuum electron generators are needed. This need is most pronounced for compact mobile generators. Here in a design study based on scaling laws for 100 W CW power at 200 GHz frequency, a cathode with at least 160 A/cm^2 DC would be required under the assumption that the other parameters are optimized. J. Booske sees the only possibility in beam compression, since also long life is needed, but it has the disadvantage of nonuniformity over the beam diameter. In this respect sheet beams for further compression from scandia doped dispenser (SDD) cathodes with lateral source dimensions of 1 mm × 50 μm have been realized by Y. Wang and coworkers with up to 100 A/cm^2 current density [90, 91], which will be described in Chap. 3.

Other requirements concern the tolerable base pressure. Here poisoning sensitivity and robustness with respect to ion bombardment are the properties to be improved.

For high-resolution electron beams for microscope applications high-brightness cathodes are required [92] as discussed in Chap. 6 by P. Kruit. Here the emission current density is not decisive, but the reduced differential brightness B_r, is given as

$$B_r = dI/(dA_s\, d\Omega\ \times U)\left[\mathrm{A}/\left(\mathrm{m}^2\,\mathrm{sr\,V}\right)\right] \tag{1.5}$$

Here I is the beam current, A_s is the virtual source area, Ω the beam angle and U the beam voltage.

In the practical brightness definition, the differential values are replaced by I, the beam area at half-width and by $\boldsymbol{\Omega} = \pi \times \alpha^2$ and hence are 1.44 times higher than differential brightness. Typical emitters used are tungsten tips, LaB_6 hairpin tips and above all a specific Schottky emitter, consisting of a tungsten single crystal tip with (100) orientation coated with a Zr–O surface dipole, reducing the work function from

4.5 to 2.9 eV. This Schottky emitter has a reduced brightness of $B_r = 1–2 \times 10^8$ A/(m^2 sr V) with a current limit of 10 nA, LaB_6 exhibits $B_r = 2 \times 10^5$ A/(m^2 sr V), tungsten reaches 5×10^8 A/(m^2 sr V) and carbon nanotubes reach 3×10^9 A/(m^2 sr V).

For electron beam microscopy the current of a single beam is maybe sufficient, but for lithography arrays of thousands of parallel beams are needed with a total beam current in the order of 200 μA on the wafer and uniform beam currents of, e.g. 13 nA. Regarding the losses in the optical column in total, a thermionic emission current of 130 mA has to be supplied. Here it has been shown by the company Mapper [92], that impregnated cathodes as used for CRTs fulfill these requirements. With Os/Ru-I cathodes, $B_r = 10^6$ A/(m^2 sr V) is reached. Here scandate cathodes should exhibit a reduced brightness of $B_r = 10^7$ A/(m^2 sr V) and could be the next step in mask-less electron beam lithography. Yet here current stability, current density uniformity and sufficient life >1000 h still need to be proven for this application.

Another type of cathodes are photocathodes, which rely on Einstein's photo-electric effect of 1905. They are used as photo detectors, e.g. in night vision, in photo-multipliers and as electron injectors in particle accelerators or in free-electron masers. The classical materials used are pure metals such as Cu, but with low quantum efficiency QE, and semiconductors such as Cs_2Te, Cs_3Sb, K_2CsSb and others with higher QE. Used are also field emitters with superimposed laser photoemission [93]. Improvement directions are high peak current densities in the pulsed beam of >10^5 A/cm^2 and high beam currents of more than 1 mA, which will be discussed in detail in Chap. 7 on photocathodes.

And finally a general trend is to try to replace thermionic cathodes by cold cathodes, essentially field emitters [94–99]. Here a general trend of improvement over time cannot be shown in a plot of current density versus time, since they exhibit already current densities much higher than thermionic cathodes, which largely depend on very small emitting areas. Hence natural applications are electron beam devices. For other applications requiring higher currents, practical measures are the increase of emitting area by using bunches or arrays of them and the increase of stable DC current from field emitter arrays (=FEAs) on a larger area. A graphical overview of the most important cold cathode types, including tungsten tips, different types of field emitters and their arrays, and pn emitters (which are also known as "reverse biased junction cold cathodes" [2, 100]), is given in Fig. 1.13 in a plot of emission current density versus emitting area, in this form first given by Gaertner [2]. Here the improvement direction over time is given in trajectories more perpendicular to the lines of equal current (10 μA, 1 mA, 100 mA) in direction of higher DC currents. Zhirnov [97] in 2000 first tried to identify standardization criteria for FE measurements, since in many cases results from different experimenters are not comparable. He also stated, that the figure of merit is the total current emitted divided by the entire cathode area and also showed a plot of current density versus area partly based on Spindt emitter arrays. His diagram corresponds to the 1 mA line in the lower half of the diagram of Gaertner. Charbonnier [98] also discussed these questions, but instead of a similar diagram came up with the statement, that maximum DC current is 3 mA and maximum pulsed current is limited to about 120 mA.

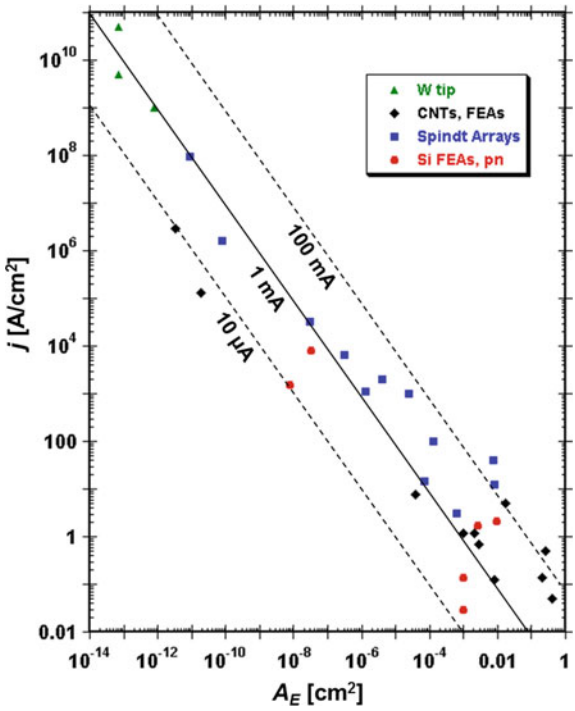


Fig. 1.13 Plot of field emission (cold emission) current density versus emitter area (including passive parts) based on the literature data for very sharp W tips, Spindt arrays, CNTs and CNT dot arrays, Si FEAs and pn emitters according to [2] and Chap. 12 (see Tables and Fig. 12.18). Lines of equal current are shown for 10 µA, 1 and 100 mA. Copyright G. Gaertner, Aachen, Germany

Wenger et al. [99] have pointed out that the field emitter arrays have not proven their usefulness in practical applications due to short life and inherent sensitivity to ion bombardment and arcing. They also showed, that when increasing current by increasing voltage of CNTs, after a Fowler Nordheim behavior at lower voltages a limited FE region follows, where the limitation is caused by space charge and by the resistance of the contact, the substrate and the emitter. At further voltage increase a normal glow discharge starts. Hence an FE-ignited glow discharge can be controlled and used as a plasma electron source. This topic will not be part of this book.

Field emission cathode features and requirements with respect to specific applications will be discussed in more detail in Chap. 10 by N. Egorov and E. Sheshin.

1.4 Future Requirements

In conclusion we have identified several improvement directions of cathodes for future requirements, depending on their application: For thermionic cathodes these are lower work function/ lower operating temperature/ higher electron emission current density/ longer life/ higher robustness versus poisoning and ion bombardment. Certain trade-offs will depend on the priorities set by the application. Currently the most advanced research concentrates on Ba scandate cathodes, since they promise

to meet the high requirements, e.g. for terahertz imaging. For high-resolution electron beams for microscope and lithography applications, high-brightness cathodes are required, but also with increased beam current. Also here Ba scandate cathodes are promising candidates. For photocathodes improvement is directed to high peak current densities in the pulsed beam of $>10^5$ A/cm^2 and to high beam currents of more than 1 mA. For laser exited photocathodes fast response times and high brightness is needed. Improvement options are higher quantum efficiency via lower work function and surface coatings on suitable metals or semiconductors.

For field emitters to become competitive in vacuum electron tubes, higher (DC) currents, higher emitting areas, higher stability and robustness are needed. Otherwise they will only be used in some niche applications. These questions will be addressed in more detail in the following chapters of this book.

References

1. G. Gaertner, H.W.P. Koops, Vacuum electron sources and their materials and technologies, in Chapter 10 of *Vacuum Electronics, Components and Devices*, ed. by J. Eichmeier, M. Thumm (Springer, 2008)
2. G. Gaertner, Historical development and future trends of vacuum electronics. J. Vac. Sci. Technol. B **30**(6), 060801 (2012)
3. P. Dunsheath, *A History of Electrical Power Engineering* (MIT Press, Cambridge, 1962)
4. Wolfram Fischer (Ed.), "Die Geschichte der Stromversorgung", Verlags- u. Wirtschaftsgesellschaft der Elektrizitätswerke m.b.H., Frankfurt a. Main, 1992
5. F.R. Paturi, *Chronik der Technik* (Chronik Verlag, Dortmund, 1988)
6. I. Buchmann, *Batteries in a Portable World*, Cadex Electronics 2011. http://www.mlelectronics.be/doc_downloads/producten/batterijen/basisinfo/BatteryKnowledgePart1.pdf
7. *The Electrical Experimenter*, Sept 1915, p. 198. https://www.americanradiohistory.com/Archive-Electrical-Experimenter/EE-1915–09.pdf
8. C. Schott, O. von Guericke, *Ottonis de Guericke Experimenta Nova Magdeburgica de Vacuo Spatio* (Amsterdam, 1672)
9. W. Conrad (ed.), *Geschichte der Technik in Schlaglichtern* (Meyers Lexikonverlag, Mannheim, 1997)
10. M. Schneider, *Elektrisiermaschinen im 18. und 19. Jahrhundert, Ein kleines Lexikon* (Universität Regensburg, 2003/2004)
11. W.E. Ayrton, *Practical Electricity; Laboratory and Lecture Course* (Cassell & Company Ltd., London, Paris & Melbourne, 1891)
12. H.W. Meyer, *A History Of Electricity And Magnetism* (MIT Press Design Department, 1971 and Burndy Library, Norwalk, Connecticut, 1972)
13. Roobert33, 2015. https://www.youtube.com/watch?v=c_0N-0lfxpE
14. B. Garg, *Introduction to Flow Batteries: Theory and Applications, coursework for PH240* (Stanford University, Fall, 2011)
15. T. Nguyen, R. F. Savinell, Flow batteries. Interface **19**(3), 54 (2010)
16. L. Figuier, *Les Merveilles de la Science*, Furne, Jouvet et Cie., Paris 1867, see Chapter "La pile de Volta", pp. 598–706
17. P. Unwin, R. Unwin, 'A devotion to the experimental sciences and arts': the subscription to the great battery at the Royal Institution 1808–91. Br. Soc. Hist. Sci. BJHS **40**(2), 181–203 (2007)
18. A. Anders, Tracking down the origin of Arc plasma science-II. Early continuous discharges. IEEE Trans. Plasma Sci. **31**, 1060–1069 (2003)

19. W. James King, *The Development of Electrical Technology in the 19th Century: l. The Electrochemical Cell and the Electromagnet*. BULLETIN 228: Contributions from the Museum of History and Technology (1962), pp. 231–418
20. V. Leiste, Siemens History Site-Im Fokus-Dynamoelektrisches Prinzip. https://www.siemens.com/history/de/aktuelles/1056_dynamoelektrisch
21. Siemens Historical Institute (ed.), *Age of Electricity* (Deutscher Kunstverlag, 2014)
22. http://www.power-technology.com/features/feature-largest-nuclear-power-plants-world/
23. S. David, E. Huffer, H. Nifenecker, High efficiency nuclear power plants using liquid fluoride thorium reactor technology. Europhys. News **38**, 24–27 (2007). https://doi.org/10.1051/EPN:2007007
24. Russian fast reactor reaches full power, 17 August 2016. http://www.world-nuclear-news.org/Articles/Russian-fast-reactor-reaches-full-power
25. M. Claessens, *ITER: The Giant Fusion Reactor*, Springer 2020
26. https://new.siemens.com/global/en/company/about/history/technology/power-generation-and-distribution/conventional-power-plants.html
27. https://en.wikipedia.org/wiki/List_of_largest_power_stations
28. https://www.energie-lexikon.info/windenergieanlage.html. RP Energie-Lexikon, Rüdiger Paschotta (2017)
29. http://www.ingenieur.de/Fachbereiche/Windenergie/Groesstes-Windrad-Welt-Herz-Nieren-getestet
30. Wind turbine, Wikipedia 2019. https://en.wikipedia.org/wiki/Wind_turbine
31. J. Peinke, D. Heinemann, M. Kühn, Windenergie, eine turbulente Sache? Physik J. **13**(7), 35f (2014)
32. https://de.wikipedia.org/wiki/Photovoltaik-Freiflächenanlage
33. http://www.ageu-die-realisten.com/archives/1473. Der Flächenbedarf von Stromerzeugungsanlagen, Günther Keil (2015)
34. A. Bachem, C. Buchal, Energiewende-quo vadis? Physik J. **12**(12), 33–39 (2013)
35. M. Bettzüge, Nationaler Hochmut oder cui bono? Physik J. **13**(5), 33–38 (2014)
36. https://de.wikipedia.org/wiki/Kernkraftwerk_Biblis
37. P. Denholm, M. Hand, M. Jackson, S. Ong, *Land-Use Requirements of Modern Wind Power Plants in the United States*. Technical Report NREL/TP-6A2-4 5834 August 2009, US Department of Energy, National Renewable Energy Laboratory
38. K. Wey, R.J. Peters, Geschichte der Vakuumtechnik. Vak. Forsch. Prax. **14**, 183 (2002)
39. N. Marquardt, *Introduction to the Principles of Vacuum Physics* (CERN Accelerator School: Vacuum Technology, 1999). http://www.iaea.org/inis/collection/NCLCollectionStore/_Public/32/011/32011645.pdf
40. P.A. Redhead, *History of Vacuum Devices* (CERN Accelerator School: Vacuum Technology, 1999), pp. 281–290. http://cdsweb.cern.ch/record/455984/files/p281.pdf
41. W. Knapp, Vacuum technology, in Chapter 11 of *Vacuum Electronics, Components and Devices*, ed. J. Eichmeier, M. Thumm (Springer, 2008)
42. P.A. Redhead, The ultimate vacuum. Vacuum **53**, 137–149 (1999)
43. A. Calcatelli, The development of vacuum measurements down to extremely high vacuum –XHV, in *Proceedings of 3rd IMEKO TC16 International Conference on Pressure Measurement, Merida 2007* (Curran Associates 2009), p. 219. http://www.imeko.org/publications/tc16-2007/IMEKO-TC16-2007-KL-034u.pdf
44. P.A. Redhead, Vacuum science and technology: 1950–2003. J. Vac. Sci. Technol. A **21**, S12–S14 (2003)
45. W. Becker (Pfeiffer Vakuumtechnik), *Turbo-molecular pump*, US Patent 3477381A, published 11 Nov 1969, priority 30 Dec 1966
46. L.D. Hall, Ionic vacuum pumps. Science **128**, 3319 (1958)
47. J.P. Hobson, Measurements with a modulated Bayard-Alpert gauge in aluminosilicate glass at pressures below 10–12 Torr. J. Vac. Sci. Technol. **1**, 1 (1964)
48. P.A. Redhead, *Extreme High Vacuum* (CERN Accelerator school: Vacuum technology, 1999), pp. 213–226

49. H. Ishimaru et al., Ultimate pressure of the order of 10^{-13} Torr in an aluminum alloy vacuum chamber. J. Vac. Sci. Technol. B **7**, 2439–2442 (1989)
50. G. Gabrielse, H. Kalinowsky et al., Thousandfold improvement in the measured Antiproton Mass. Phys. Rev. Lett. **65**, 1317–1320 (1990)
51. P. Micke, M. Schwarz et al., Closed-cycle, low-vibration 4 K cryostat for ion traps and other applications. Rev. Sci. Instr. **90**, 065104 (2019)
52. M. Schwarz et al., Cryogenic linear Paul trap for cold highly charged ion experiments. Rev. Sci. Instr. **83**, 083115 (2012)
53. C. Benvenuti, Extreme high vacuum technology for particle accelerators, in *Proceedings of the 2001 Particle Accelerator Conference (IEEE), Chicago* (2001), pp. 602–606
54. P.A. Redhead, Vacuum and the electron tube industry. J. Vac. Sci. Technol. A **23**, 1252 (2005)
55. O. Darrigol (CNRS), *Electrodynamics from Ampere to Einstein* (Oxford University Press, Oxford, 2000)
56. P.A. Redhead, Birth of electronics: thermionic emission and vacuum. J. Vac. Sci. Technol. A **16**, 1394 (1998)
57. S. Okamura, *History of Electron Tubes* (Ohmsha, Tokyo, 1994 and IOS Press, 1994)
58. G. Bekooy, *Philips Honderd–en industriele onderneming* (Europese Bibliothek, Zaltbommel, 1991)
59. D. van Delft, A. Maas, *Philips Research–100 years of Inventions that Matter* (WBooks, Zwolle, 2013)
60. F. Rosebury, *Handbook of Electron Tube and Vacuum Techniques* (American Institute of Physics, 1993, originally MIT, 1964)
61. H.G. Boumeester, Development and manufacture of modern transmitting valves. Philips Tech. Rev. **2**(4), 115–121 (1937)
62. G.E.J. Tyne, *Saga of the Vacuum Tube* (Indianapolis, 1987)
63. J.W. Stokes, *70 Years of Radio Valves and Tubes* (New York, 1982)
64. W.H. Kohl, *Materials and techniques for vacuum devices* (Reinhold Publishing Corp., New York, 1967)
65. G. Gaertner, D. den Engelsen, Appl. Surf. Sci. **251**, 24–30 (2005)
66. J.A. Castellano, in *Digest of Society for Information Display (SID) Symposium 1999* (SID, San Jose, 1999), p. 356
67. P. Combes, Display Components, Philips ppt-presentation (2000). See https://www.yumpu.com/en/document/view/217010/philippe-combes
68. J. Kitzmiller, *Industry and Trade Summary-Television Picture Tubes and other Cathode Ray Tubes*. USITC Publication 2877, May 1995
69. N. Stam, *CRT Innovations*, PCMAG.COM, 31.3.2003, based on isuppli/Stanford resources + internal information from LPD 2002
70. M. Kenney, R. Florida (eds.) *Locating Global Advantage. Industry Dynamics in the International Economy*. Stanford 2004. Chapter 4: M. Kenney, *The shifting value chain. The Television Industry in North America*, p. 182
71. European Commission, Case AT.39437, TV and computer monitor tubes. Cartel procedure, 5 Dec 2012
72. R.J. Barker, J.H. Booske, N.C. Luhmann, G.S. Nusinovic (eds.) *Modern Microwave and Millimeter-Wave Power Electronics* (Wiley-IEEE, New York, 2005)
73. J.H. Booske, Plasma physics and related challenges of millimeter-wave-to-terahertz and high power microwave generation. Phys. Plasmas **15**, 055502 (2008)
74. H. Verhaar (Philips Lighting) *Global Incandescent Phase-out; Meeting the Demand*. Presentation 12 May 2008, Shanghai. http://www.energyrating.gov.au/document/presentation-global-incandescent-phase-out-meeting-demand
75. J. Stettler, A. Leslie, A. Bell (new street Research) *The Future of Lighting: Who Will Win?*. Industry analysis, 5 May 2010. www.newstreetresearch.com, http://www.sdr.si/pdf/the%20future%20of%20lighting%20100305.pdf
76. P. Waide, *Phase Out of Incandescent Lamps*. International Energy Agency, Information paper, April 2010, https://www.oecd-ilibrary.org/docserver

77. G. Zissis, P. Bertoldi, *2014 Update on the Status of LED Market* (European Commission, Joint Research Centre, JRC Technical Report, Ispra, Italy, 2014)
78. G. Zissis, P. Bertoldi, *Status of LED-Lighting World Market in 2017*. European Commission, JRC Technical Reports, Ispra (2018)
79. G.A. Haas, Thermionic electron sources, in *Methods of Experimental Physics*, vol. **4**, part A (Academic Press, New York, 1967)
80. R.O. Jenkins, A review on thermionic cathodes. Vacuum **19**(8), 353 (1969)
81. R. Thomas, J. Gibson, G.A. Haas, R. Abrams, Thermionic sources for high brightness electron beams. IEEE Trans. ED **37**, 850–861 (1990)
82. P.W. Hawkes, Thermionic emission. Encycl. Appl. Phys. **21**, 229–243 (1997)
83. R.R. Umstattd, Advanced electron beam sources. Chapter 8 in *Modern Microwave and Millimeter-Wave Power Electronics*, ed. by R. Barker et al. (Wiley, 2005), pp. 393–444
84. S. Yamamoto, Fundamental physics of vacuum electron sources. Rep. Prog. Phys. **69**, 181–232 (2006)
85. H. Lemmens, P. Zalm, New developments in oxide-coated cathodes-oxide-coated cathodes for loads of 1 to 2 A/cm^2. Philips Techn. Rev. **23**, 21–24 (1961/62)
86. T. Aida, H. Tanuma, S. Sasaki et al., Emission life and surface analysis of barium-impregnated thermionic cathodes. J. Appl. Phys. **74**, 6482–6487 (1993)
87. W. Liang, Y. Wang, J. Wang, W. Liu, F. Yang, DC emission characteristic of nanosized-scandia-doped impregnated dispenser cathodes. IEEE Trans. ED **61/6**, 1749 (2014) and Y. Wang, Scandate cathode–what we have learned and what we expect to know, in *Proceedings of the IVESC 2016, Seoul*, p. 3
88. B. Djubua, O. Kultashev, A. Makarov, O. Polivnikova, E. Zemchikhin, Metal alloy cathodes for application in vacuum microwave devices, in Proceedings of the IVEC/IVESC 2012, Monterey, CA, paper 125
89. 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)
90. J. Wang, W. Liu, Y. Wang, M. Zhou, in *Abstract Book of 2008 International Vacuum Electron Sources Conference (IVESC), London* (2008), p. 10
91. Y. Wang, J. Wang, W. Liu, J. Li, in *Abstract Book of 2008 International Vacuum Electron Sources Conference (IVESC), London* (2008), p. 16
92. J.R. Meijer et al. (MAPPER Lithography B.V.), Electron sources for MAPPER maskless lithography, in *Proceedings of IVESC-ICEE-2014, Saint-Petersburg, Russia* (2014)
93. D.H. Dowell, J. Smedley, et al., Cathode R&D for future light sources. SLAC-Pub-14002 and Nucl. Instrum. Methods Phys. Res. A **622**, 685 (2010)
94. I. Brodie, C. Spindt, Vacuum microelectronics, in *Advances in Electronics and Electron Physics*, vol. 83, ed. by P. Hawkes (Academic Press, New York, 1992), p. 1
95. W. Zhu (ed.), *Vacuum Microelectronics* (Wiley, 2001)
96. N. Egorov, E.P. Sheshin, *Electron Field Emission, Principles and Applications* (In Russian) (Intellekt, 2011); Updated English version *Field Emission Electronics* (Springer, 2017)
97. V.V. Zhirnov, C. Lizzul-Rinne, G.J. Wojak, R.C. Sanwald, J.J. Hren, "Standardization" of field emission measurements. J. Vac. Sci. Technol. B **19**, 87 (2001)
98. F. Charbonnier, Developing and using the field emitter as a high intensity electron source. Appl. Surf. Sci. **94**(95), 26–43 (1996)
99. D. Wenger, W. Knapp, B. Hensel, S.F. Tedde, Transition of electron field emission to normal glow discharge. IEEE Trans. Electron Devices **61**(11), 3864–3870 (2014)
100. G. van Gorkom, A. Hoeberechts, "Back-biased junction cold cathodes: history and state of the art", in Vacuum Electronics 89, IOP Conf. Series **99**, 41–52 (1989)

Chapter 2
Review on Impregnated and Reservoir Ba Dispenser Cathodes

Jean-Michel Roquais, Bernard Vancil, and Michael Green

Abstract In order to overcome the conductivity limitations of Ba–oxide cathodes, Ba dispenser cathodes with a conductive metal matrix were introduced in the 1950s. Different variants of Ba dispenser cathodes are presented, either of the impregnated or the reservoir type. The optimized versions allow DC current densities of up to 10 A/cm^2 and lifetimes of more than 100,000 h. Due to their high reliability, they are the workhorse in a lot of applications, above all in different types of microwave tubes. Emission and lifetime models are presented and discussed.

2.1 Introduction: Historical Development of Different Types of Ba Dispenser Cathodes

The oxide cathode with a nickel base metal has long time been the available electron source for lamps used in radio-communication, transponders, then later CRTs for television. Used in pulsed-mode, it can deliver current densities up to 50 A/cm^2. Nevertheless, when the power through the oxide coating becomes too high, excessive sintering of the oxides takes place by overheating and the emitting layer is destroyed. This typically occurred when the oxide cathode was operated in DC mode at current densities higher than 0.5 A/cm^2 [1] (see G. Herrmann and S. Wagener, "The Oxide Coated Cathode", Chapman & Hall, London 1951). New oxide cathode types allow higher DC loadability and are treated in Chap. 4.

J.-M. Roquais (✉)
Thales AVS France, Vélizy-Villacoublay, France
e-mail: jean-michel.roquais@thalesgroup.com

B. Vancil
Ebeam inc., Beaverton, OR, USA
e-mail: Bernie@ebeaminc.com

M. Green
Consultant, Palo Alto, CA, USA
e-mail: MikeGreen747@aol.com; mverdi@aol.com

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*,
Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_2

A new generation of cathodes has emerged in the early 50s—the dispenser cathode—which overcomes this weakness by featuring a W metal matrix which can withstand high currents flowing through it. The first applicable dispenser cathode developed by Lemmens et al. [2, 3] called the L-cathode (after Lemmens) was a reservoir cathode. A similar type, the metal capillary cathode or MK cathode was introduced by Katz of Siemens [4, 5], claiming earlier priority based on a German application in 1942, but was not well known. In this L-cathode, a distinction existed functionally between the emissive surface (the substrate), and the reservoir of the material that allows it to maintain a low work function by constantly dispensing Ba (the adsorbate) to the surface. The structure of this cathode consisted in a storage chamber placed behind a porous W body, as can be seen in Fig. 2.1. This chamber contained a mixture of W powder and Ba, Ca carbonate. This cathode was prone to outgassing over long periods of time. An example of a reservoir cathode is shown in Fig. 2.1.

Levi [7–9] of North American Philips proposed a different structure in which the porous W body was impregnated by Ba, Ca aluminate. This cathode, called A- or B-type cathode depending on the impregnant composition, allowed higher average currents than the oxide cathode. The need to reduce the operating temperature to avoid rapid evaporation of the adsorbate, combined with the need for increased current densities have generated studies aiming at decreasing the work function, which is the main parameter governing the level of thermionic emission. In this scope, Zalm and van Stratum demonstrated that coating an L-cathode with a 0.5-μm-thick osmium coating substantially increased emission compared to an

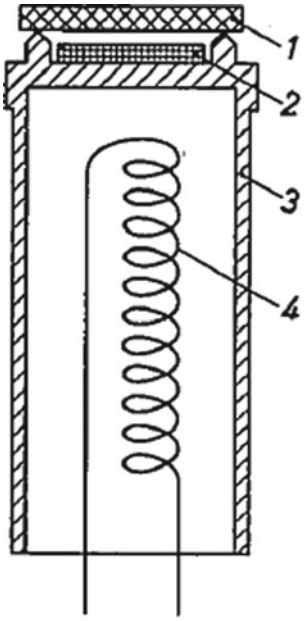


Fig. 2.1 Basic construction of an L cathode: 1 = porous tungsten body; 2 = chamber containing a reservoir of Ba–Ca Aluminate; 3 = molybdenum cylinder; 4 = heater filament; Figure from [6] P. Zalm, A. J. A. van Stratum, Philips Technical Review Vol. 27, (1966), 69–75; Copyright Royal Philips N.V

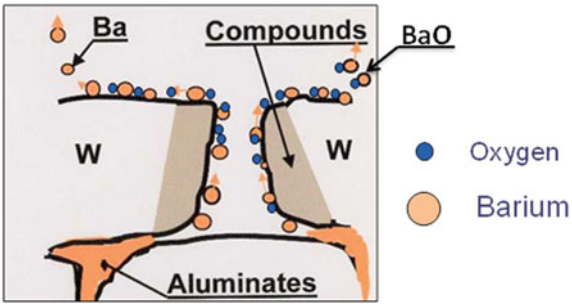


Fig. 2.2 Cross-sectional schematic view of an impregnated cathode

uncoated cathode [6]. The authors faced a paradoxical effect: the bare work function of osmium being higher than the one of bare W, the effective work function of the surface when covered with Ba becomes lower for the Os substrate than for the W surface. Theoretical explanations were provided by the authors, but more recent studies have completed their interpretation, as will be described later. Besides Os also Ir and Re top-coatings have the beneficial effect of further lowering the work function. A B-type cathode with Os coating is also called M cathode (M stands for magic). For further information about the historical development of dispenser cathodes, detailed reviews were provided by Cronin [10] and Shroff [11].

A cross section of an impregnated cathode is seen in Fig. 2.2. Barium–calcium aluminate reacts with tungsten in the porous matrix. This produces free barium which migrates to the surface of the cathode, then diffuses across the surface. The barium must be continually replenished due to evaporation at elevated temperatures.

2.2 Impregnated Ba Dispenser Cathodes (Jean-Michel Roquais)

2.2.1 Structure of Impregnated Dispenser Cathodes; The Different Pellets and Impregnants

This subchapter aims at covering the better known in-use technologies, with the exception of the reservoir cathodes and the C.P.D cathodes treated in the following subchapter. The scandate cathodes are not included either, for the same reason. The fabrication of dispenser cathodes involves the pressing and sintering of metal powders to obtain a porous W-emitting pellet with pores filled by the impregnant. Several compositions of pellets exist, as well as several compositions of Ba, Ca aluminates. The possible metallurgy of pellets of different kinds, together with the impregnant's description is summarized below:

Isostatic pressed bar: W powder of well-defined grain size distribution and purity is pressed into a bar in an isostatic press. The green porosity of the pressed bar being too

high, sintering under a protective atmosphere is then performed to decrease porosity to an average value close to 20%. In a further step, the pores of the bars are infiltrated with plastic or copper to ease the machining and to hold pores open. Pellets are machined from the W bar in the form of discs or cylinders. One face of the pellet is destined to become the surface emitting the electrons. For microwave tubes, it is common to give a concave shape to this emitting surface for easier beam focusing.

S-type and B-type: The impregnant used in Ba dispenser cathodes is a Ba, Ca aluminate. The starting materials are Ba, Ca carbonates, and aluminum oxide. Storage of the powders in the form of carbonates is preferred as they are stable in air. A conventional preparation method of the aluminates has been described [11]. Ba, Ca aluminates of different molar fractions are used:

4BaO, 1CaO, 1Al_2O_3 (4:1:1): molar fraction of "S" type cathode
5BaO, 3CaO, 2Al_2O_3 (5:3:2): molar fraction of "B" type cathode.

Other molar versions than those cited above are sometimes proposed, for example, 3:1:1 or 6:1:2. The thermochemistry of the BaO·CaO·Al_2O_3 system has been studied and can serve as a basis for the choice of the proper impregnant composition for a given application [12]. The molten barium–calcium aluminate is drawn into the porous pellets by capillary action under the protective atmosphere at high temperature. A mechanical or chemical cleaning of the surface has to be done to remove undesirable residues.

Individual die-pressed pellets: Another approach consists in die-pressing the W pellets individually instead of machining them from bars. In this case, the W powder is pressed in a cavity and is submitted to a uniaxial pressure exerted by a double-die system. Like the bars, the "green" pellets need to be sintered to obtain strength and the desired final porosity close to the typical value of 20%. Compared to bar pressing, this method is advantageous in terms of productivity as several pellets per minute can be formed by adapted automatic presses. Cutting pellets in bars takes much longer than shaping them directly inside the press cavity.

Co-pressed pellet or "Cermet": In an attempt to further increase the productivity, it is possible to realize an emitting pellet by pressing together a mixture of W and Ba, Ca aluminate. This type of pellet is often referred to as "Cermet" or "co-pressed". The process steps have been described in Shroff's patent [13]. It has an advantage over the realization of bars because giving a form in a shaped cavity is possible, for instance, a concave face for application in microwave tubes. Nevertheless, it has a clear disadvantage compared to pressing powders made of refractory metals only. As the aluminates are present in the pressed pellets, the following step of sintering cannot be performed at temperatures as high as the ones used for sintering of refractory metals. But if it is an alloy cathode, it can be sintered. Practically, the mechanical integrity is less good with co-pressed than with refractory metals matrix. Due to lower sintering temperatures, the other weakness of the co-pressed is a porosity clearly higher than the desirable 20% porosity target.

An original alternative approach for making emitting pellets has been developed by Toshiba. It consisted in pressing a 5 inches diameter disc. In a further step, pellets are laser-cut from the disc. As for the co-pressed, the method aimed at increasing the productivity in a period (the 1980s–1990s) when every manufacturer of CRTs for computers or TV sets had to ensure the production of tens of millions of units per year. Nowadays with the obsolescence of CRTs, the needs in terms of quantity have been clearly lowered and range, for a given cathode design and depending on the application, from a few units to several thousand a year.

MM cathode (Mixed metal): To obtain this type of cathode, a mixture of powders of W and of an additional refractory metal is pressed, and then sintered. Known examples are the MM-type cathodes made from a W-Os [14, 15] or W-Ir mixtures [16, 17]. These have the disadvantage that both Os and Ir inhibit the production of free barium.

M-Type cathode: A thin film of metal or of a metal alloy is deposited on the cathode surface by sputtering or evaporation. DC or RF Sputtering of the metal target is usually performed in the argon plasma. An original method of deposition by OMCVD (Organometallic Chemical Vapor Deposition) has also been developed at Naval Research Laboratory [18]. The noble metals used to form a surface film are from the platinum group metals (except platinum). Most often used are Os, Ir, Re, deposited as a single metal, or in combination with W to form surface binary alloys such as W-Os [11], W-Ir [18–21] or ternary alloys such as Os-Re-W [22, 23] and Os-Ru-W (CPI ternary alloy cathode). Some manufacturers sputter an alloy of Os/Ru rather than pure osmium.

CMM "(Coated Mixed Metal)": The thin film of metal or metal alloy is deposited on the pellet of a MM-type cathode. As with M-type cathodes, the metals are chosen among Os, Ir, or Re. Those improved cathodes have shown emission properties superior to those of M-type or MM-type [11].

2.2.2 Emission Properties of Impregnated Cathodes

2.2.2.1 Activation

When first heating the cathode in vacuum, gases are desorbed. Monitoring the evolved species by residual gas analysis is recommended. The cathode activation consists in steps of heating at high temperatures, usually in the 1000–1200 °C temperature range with a total duration of a few hours. With heating, the contaminants, mainly C and S, are progressively eliminated, and then normally only Ba and O are detected on the cathode surface together with the elements of the underlying metallic substrate, as shown by Auger Electron Spectroscopy (A.E.S) [24–29]. The evolution of the cathode surface chemistry during activation is characterized by an increasing Ba concentration together with an increase in O [26]. For an M-type cathode with Os

top-layer, the activation step at the maximum temperature of about 1200 °C induces an inter-diffusion of W and Os [26]. The formation of intermetallic phases results from this diffusion. The same inter-diffusion phenomenon is observed with Ir M-type cathode [19, 20]. This diffusion phenomenon is quite important as it determines the composition of the metallic substrate, this substrate composition being one of the factors determining the cathode work function.

2.2.2.2 Current–Voltage Characteristics, Dependence on Temperature and Work Function

The emission characteristics of a thermionic cathode can be obtained by measuring the current emitted by a cathode in a diode structure.

The Schottky plot consists in recording the current as a function of an increasing voltage V_a on the anode, at a fixed cathode temperature. On this plot, two main regions of emission appear:

- the region of emission in space charge limited regime,
- the region of emission in temperature-limited regime.

Increasing the anode voltage from zero, the cathode starts emitting in the space charge regime, and gradually enters the temperature-limited one. Plotted as a function of $V_a^{½}$, the logarithm of the current follows a straight line in the temperature-limited regime due to the so-called Schottky effect. Extrapolating the straight line to zero Volt, one determines the "zero-field current density" $J_{0,}$ which represents the practical value of cathode current density when comparing results from different sources. Schottky plots performed at two different temperatures are represented in Fig. 2.3 for an osmium-coated M-type cathode. In this plot, the current emitted by the cathode has been converted into a current density knowing the emitting area of the cathode.

The emission in the temperature-limited regime is described by the Richardson–Dushman equation:

$$J_0 = A_R T^2 e^{-\varphi/kT} \tag{2.1}$$

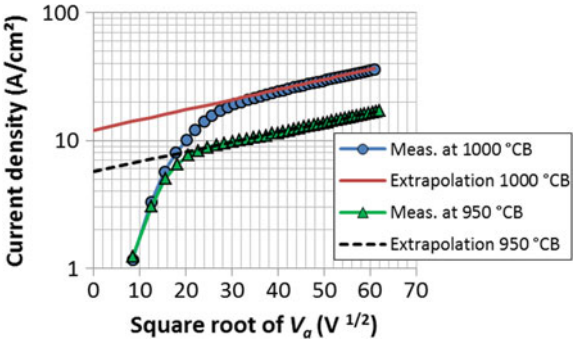


Fig. 2.3 Schottky plot on M-type cathode in a diode structure performed at two different temperatures. The extrapolation of current density at zero-field is also plotted. The temperature is expressed in $°C_B$ (brightness temperature)

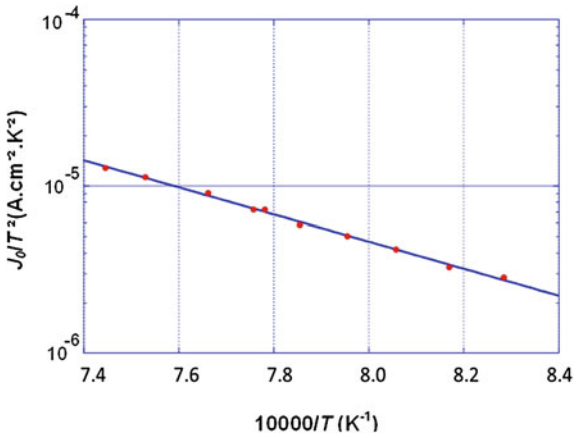


Fig. 2.4 Richardson plot of M-type cathode

where J_0 [Acm^{-2}] is the current density at zero-field, A_R [Acm^{-2} K^{-2}] is the Richardson constant, in the ideal case equal to the thermionic constant 120.4 [Acm^{-2} K^{-2}], φ [eV] is the electronic work function, k [eV/K] is the Boltzmann constant, and T[K] is the cathode true temperature.

The value of φ [eV] is directly proportional to the slope of the term $\ln(J_0/T^2)$ (Fig. 2.4) plotted as a function of the reciprocal true temperature (Richardson plot). Experimentally, for the constant in expression (2.1) the theoretical value of 120.4 Acm^{-2} K^{-2} is not found, due to the fact that the work function depends on temperature and that the cathode surface may not be uniform.

The temperature dependence of φ can be expressed as follows:

$$\varphi = \varphi_0 + \alpha T \tag{2.2}$$

With α [eV/K]: constant of proportionality.

Derived from the measured values of J_0, the work function can be calculated using (2.1), setting the Richardson constant equal to the thermionic constant. As an example, from the Richardson plot of Fig. 2.4 one obtains

$$\varphi = 1.85\,\text{eV at } 1000\,^{\circ}\text{C}_\text{B},$$

which is also called the effective work function.

Another approach based on a Richardson plot derives the work function and the Richardson constant from slope and axis intercept and regards them as phenomenological parameters. From the slope of the straight line of Fig. 2.4 based on (2.2), we derive:

$$\varphi_0 = 1.61\,\text{eV}$$

$\varphi_0/k(K)$ is the slope of the Richardson plot.

2.2.2.3 Work Function Modeling

A theory describing the work function of a refractory metal surface coated by a metallic film was proposed by Gyftopoulos and Levine [30]. In an improvement of this theory, Gyftopoulos and Steiner [31] have proposed a more complete calculation. In this second calculation, the work function φ of the bimetallic surface is, like in the first approach, expressed as a function of the coverage ratio Θ of the substrate by the adsorbate. It is related to the work function of the bare substrate disturbed by two interatomic interactions, a covalent contribution cQ and an ionic one bF.

$$\varphi(\ominus) = \varphi_s + cQ(\ominus) + bF(\ominus) \tag{2.3}$$

with

$$cQ = -\left(\varphi_s - \varphi_f\right) M(\ominus) \tag{2.4}$$

φ_s = work function of the bare metal substrate,
φ_f = bulk work function of the film of adsorbate,
θ = fraction of monolayer of adsorbate. For one monolayer $\theta = 1$.

Using the theory of Gyftopoulos and Steiner, one can calculate the work function of the surfaces constituted of alkali-earth adsorbed on refractory metals. According to these authors, the Morse function is expressed as follows:

$$M(\ominus) = 2e^{ad_0\left(1-\ominus^{-\frac{1}{2}}\right)} - e^{2ad_0\left(1-\ominus^{-\frac{1}{2}}\right)} \tag{2.5}$$

with

d_0: distance between two neighboring adsorbate atoms at $\theta = 1$ coverage,
$d_0\theta^{-\frac{1}{2}}$: distance between two neighboring adsorbate atoms at Θ coverage,
a: parameter characteristic of the interatomic bond; $ad_0 \approx 3$ [32].

Note that $M(0) = 0$ and $M(1) = 1$.

The expression of $b(\theta)$ and $F(\theta)$ are given in [31]. For $\Theta = 0$, $b = 0$ and for $\theta = 1$, $F = 0$. Thus $\varphi(\theta)$ as defined in (2.3) respects the limit conditions of the model:

$$\varphi(0) = \varphi_s \text{ and } \varphi(1) = \varphi_f$$

The theory of Gyftopoulos and Steiner has been used by Shroff and Firmain [33] to calculate the work function versus the Ba coverage, giving one piece of their lifetime prediction model for the dispenser cathode. In their calculation, Shroff and Firmain use the value of 2.97 for ad_0 corresponding to the average of many different adsorbates as calculated by Gyftopoulos and Steiner [31]. Initially, the calculation of the work function was done in the case of a film of a single species on the refractory

metal. For dispenser cathodes, it has been proved that the surface film is always comprised of a complex of barium and oxygen atoms. In the lifetime prediction model, the adsorbate [11, 33] or the film [34] stands for this barium-oxygen complex on the cathode surface.

2.2.2.4 Model of Lifetime Prediction

To calculate the cathode lifetime, Longo [35, 36] proposed an equation to express the cathode-operating current density $J_{op.}$ as a function of 2 terms, one representing the space charge regime of the cathode, the other one the temperature-limited regime. This equation featured terms without any power exponent. It was improved by Vaughan [37], who introduced the power exponent n to better fit the experimental data. Consequently, the expression of $J_{op.}$ became

$$\frac{1}{J_{op.}^n} = \frac{1}{J_{sc}^n} + \frac{1}{J_{TL}^n} \tag{2.6}$$

with

J_{op}: operating current density of the cathode in the tube;
J_{sc}: current density of the cathode in space charge regime;
J_{TL}: current density of the cathode in the temperature-limited regime. Practically, the zero-field current density J_0 is used in formulae (2.6) to represent J_{TL};
n: exponent.

Dieumegard [34] confirmed $n = 6$ as the best value to fit his experimental curves of measured current density as a function of cathode temperature on M-type cathodes. This correctly fits the transition region between space charge and temperature-limited regions in case of a well-activated cathode in early life. A value of n smaller than 6 can be used to represent a cathode at its end of life.

Both J_{sc} and J_{TL} vary as a function of the coverage θ [34]. Shroff [11] and Dieumegard [34] have developed a semi-empirical model allowing an expression of both J_{sc} and J_{TL} as a function of θ. For S-type cathodes, Palluel and Shroff [38] demonstrated experimentally that a long-term correlation exists between the decrease of the cathode emission and the consumption of the impregnant in the cathode pores, the correlation being true for both the effect of time and temperature. They demonstrated that the Ba depletion increases proportionally to the square root of time, this behavior being confirmed by other works from Higuchi et al. [39] and Roquais et al. [40]. The mechanism leading to Ba consumption has been explained [11]. The Ba aluminate reacts with W generating free Ba, which migrates to the cathode surface while reaction products remain in the pores (Fig. 2.2).

The Ba depletion measured from the cathode emitting face, called e_{Ba}, was determined on the elemental mapping of fractures of Ba pellets using either EPMA (Electron Probe Micro-analysis) [11] or EDX (Energy Dispersive X-Ray) [40]. An

example of an EDX mapping of Ba is shown in Fig. 2.5 for a cathode impregnated with 4-1-1 aluminate.

The Ba depletion has been measured on M-type cathodes aged in bulbs as a function of time for 4 different operating temperatures [40]. The cathodes selected for this experiment had a 4-1-1 impregnant, a porosity between 18 and 19% and were Os-Ru coated (~5000 Å). Those cathodes had no obstructed pores on the face opposite to the emitting face (backside of the pellet). These experimental results of depletion versus time t are presented in Fig. 2.6. The dashed curves are fitting curves with a square root variation of the depletion versus time. The equations of the fitting curves were obtained by fitting the experimental points in a log–log representation of e_{Ba} as a function of time. They have the following form:

$$e_{\mathrm{Ba}}(t, T) = A(T) \times t^{1/2} \tag{2.7}$$

with

$e_{\mathrm{Ba}}(t, T)$: barium depletion in μm and $A(T)$: coefficient in μm/h$^{1/2}$ independent of time.

The depletion e_{Ba} has been expressed by Shroff as follows [11]:

$$e_{Ba}(t, T) = A_0 \times t^{\frac{1}{2}} \times e^{\frac{-11600V_e}{2T}} \tag{2.8}$$

with

$V_e/2$ (eV): activation energy of the depletion
T (K): temperature of the cathode;
A_0: coefficient in μm/h$^{1/2}$ independent of time and temperature.

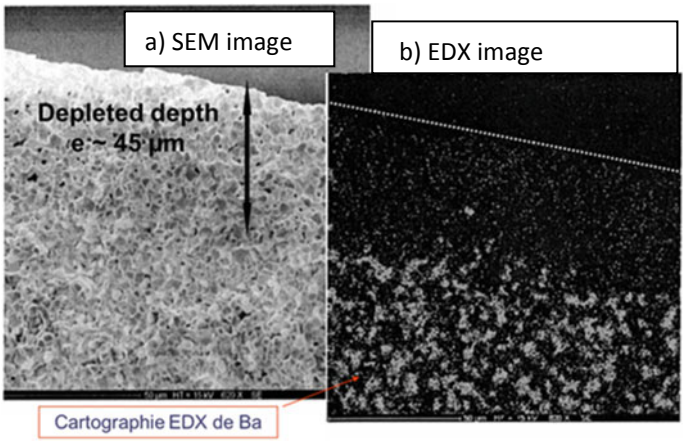


Fig. 2.5 measurement of the Ba depletion e_{Ba} in a W pellet by SEM (left image) and by EDX image of Ba (right image)

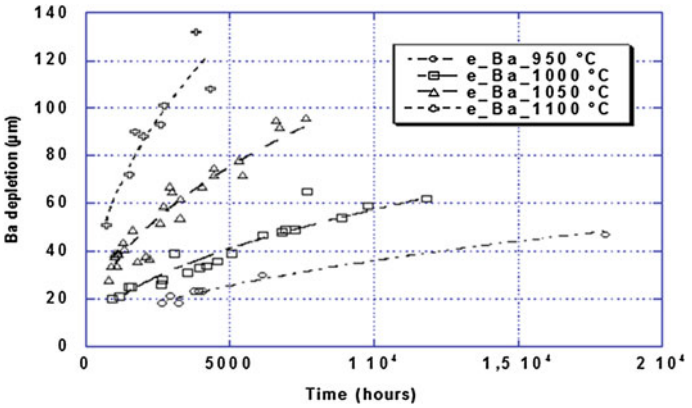


Fig. 2.6 Depletion e_{Ba} as a function of time for different cathode operating temperatures. The points as shown are experimental values determined by EDX and the dashed curves are fitting curves using a $t^{1/2}$ law of variation

From experiments of cathode aging at different temperatures, plotting the time needed for a given depletion versus the reciprocal temperature, one can extract the value of V_e. From the depletion studies, the extracted value of V_e is 2.8 eV [11]. To obtain complementary results in the study of Ba consumption, Roquais et al. [40] have conducted mass spectrometry measurements on dispenser cathodes. Such a measurement on one cathode is shown in Fig. 2.7. When heating dispenser cathodes, both Ba and BaO evaporate from the surface. They are detected as Ba^{++} ions and BaO^{+} ions as evaporated species. A continuous loss of Ba occurs during the cathode life, explaining the observed Ba depletion from the cathode surface. Because barium

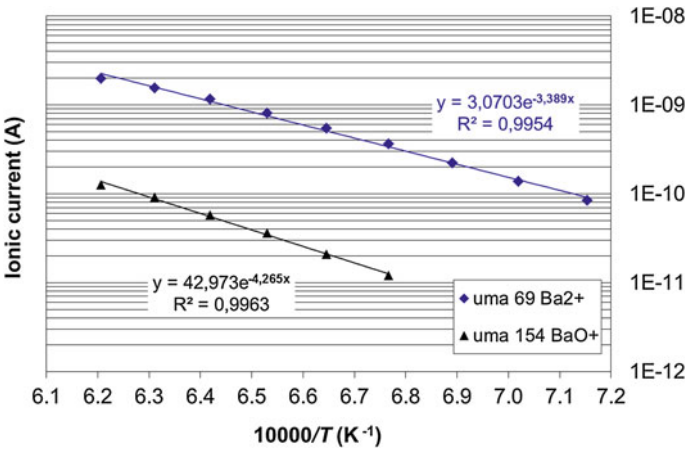


Fig. 2.7 Ionic current of the mass spectrometer for Ba^{++} and BaO^{+} species as a function of the reciprocal temperature

Table 2.1 Activation energy of evaporation of Ba and BaO species

Species	Activation energy of evaporation (eV)					
	Roquais et al. [40]	Gaertner et al. [42]	Koitabashi et al. [43]	Rittner et al. [41]	Brodie et al. [44]	Shih et al. [45]
Ba	3.23	3.2	3.1		3.23	3.38
BaO	4.02	4.1	4.0			3.65
Ba + BaO				3.00–3.25		

is doubly ionized, it produces twice the current per atom as the BaO. Thus, the top curve in Fig. 2.7 should be reduced by a factor of 2 to reflect the loss from the surface of barium versus BaO.

Fairly good agreement is obtained on the activation energies of evaporation announced by several authors [40–45] (see Table 2.1). The activation energy of the BaO evaporation is found higher than the one of Ba.

Based on the work on Ba migration from Rittner et al. [41] and Fote and Luey [46] and having found an empirical expression for depletion e_{Ba}, Shroff proposes an expression for the variation of θ versus time [11]:

$$\Delta\theta(t,T) = -\alpha \times t^{\frac{1}{2}} \times e^{\left[\frac{-11600}{T}\left(\frac{V_e}{2}+V_{L_\delta}\right)\right]} \tag{2.9}$$

with

V_{L_δ} (eV): activation energy of Ba migration length on tungsten (Rittner value: 0.7 eV).
$\Delta\theta(t,T) = \theta(t,T) - \theta(0,T)$.
t(h): time of cathode operation.
T (K): cathode temperature; α ($h^{-1/2}$): coefficient of proportionality depending on cathode.

The activation energy for Ba coverage V_θ is defined as $V_\theta = \frac{V_e}{2} + V_{L_\delta}$.

In Shroff's calculation, V_θ appears similar to the activation energy V_{t_δ} of Ba desorption from W [11]. The coefficient α is equal to 1.5×10^5 $h^{-1/2}$ for the cathodes used by Shroff to establish his model. Cathodes with different characteristics (for example, a different porosity or aluminate composition) may present a different value of α.

Equation (2.9) allows calculating the value of Θ at any interval of time. Practically, for a fixed temperature of life test, one can calculate $\Theta(t)$ and thus $\varphi(t)$ through (2.3). After having determined $\varphi(t)$, one obtains $J_{TL}(t)$ using the law of Richardson–Dushman.

Shroff has proposed the following expression for $J_{sc}(t)$:

$$J_{sc}(t,T) = J_{sc}(0,T) \times \theta(t,T)/\theta(0,T) \tag{2.10}$$

Table 2.2 Values of activation energy V_θ appearing in the models of different authors

	Shroff [11]	Shroff and Firmain [33]	Longo et al. [35]	Dieumegard et al. [34]
V_θ (eV)	2.1	2.13	2.23	2.1–2.3

A value of 0.8 has been obtained by Dieumegard [34] for θ $(0, T)$, as the existence of an initial partial monolayer appears to be a better description than a complete one [47, 48]. The calculation of $J_{TL}(t)$ and $J_{sc}(t)$ as described above gives $J_{op}(t)$ as per (2.6). As the model of Shroff is semi-empirical, it is important to have confidence in the data obtained from previous works from which coefficients are derived for the lifetime calculation. Concerning the activation energy of θ versus temperature, it is the sum of term $V_e/2$ derived from actual measurements of Shroff on his S-type or M-type cathodes and of term V_{L_δ} derived from previous work from Rittner [41]. More evaporation studies have been conducted since to determine both V_{t_δ} and V_{L_δ}. The experimental work by Free and Gibson about diffusion length on tungsten from slits [47] shows values in good agreement with those of Rittner. Thus, the value of 0.7 eV for V_{L_δ} used by Shroff finds a confirmation. The values of V_θ as proposed by different authors in their cathode life prediction model are summarized in Table 2.2.

The activation energy of the recovery time E_{rec} (Ba) for Ba coverage on cathode surface associated with the adverse Ba diffusion and Ba desorption phenomena was estimated by A.E.S by Manenschijn et al. [49] or by ion bombardment experiments by Gaertner et al. [50]. They found, respectively, 2.2 eV and 2.0 eV which appear consistent with the values of V_θ of Table 2.2. It is reasonable to think that E_{rec} (Ba) is equivalent to V_θ.

The values of V_θ are usually obtained by calibrating the life prediction model against real life tests results in special vehicles (diodes) or in real TWTs [34]. Life-test data of M-type cathodes are now available at Thales offering the possibility to compare the model with actual long-term measurements. The life test is composed of eleven diodes divided into 3 groups of cathode operating temperature: 970 °C_B, 1000 °C_B, and 1030 °C_B, with, respectively, 5, 3, and 3 diodes under test. They are operated at a nominal operating current density $J_n = 1.5$ A/cm^2. In each diode setup, the cathodes are equipped with a cathode guard ring. The anode is a gridded one to limit barium re-evaporation onto the cathode surface [11]. The cathodes have a 2 mm diameter. The zero-field temperature-limited current density of the cathode J_0 was deduced from Schottky plots (current density as a function of the square root of the anode voltage, see the paragraph about current–voltage characteristics). J_0 was measured periodically at the operating cathode temperature. The measured curves of J_0 as a function of time (averaged per temperature group) and the calculated ones using the model of Shroff are presented in Fig. 2.8.

The values of the parameters for the calculation were set as follows:

φ_s: 4.6 eV; $\theta(0, T) = 0.8$; $J_0 = 11.6$ A/cm^2 at $t = 0$ (value at 1000 °C_B).

V_θ was the parameter adjusted to fit the experimental values. Satisfactory fits were obtained with V_θ values in a 2.1–2.2 eV range, in good agreement with the values summarized in Table 2.2.

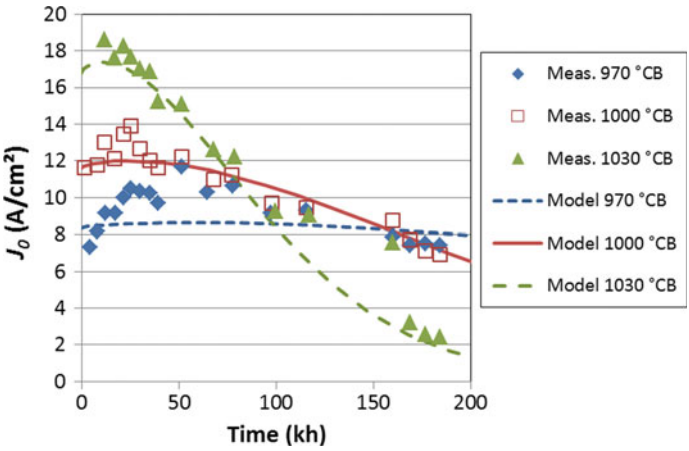


Fig. 2.8 Zero-field current density J_0 measured as a function of time for M-type cathodes (pure osmium coating) operating in diodes at different temperatures and calculated values using the model of life prediction of Shroff

Concerning the value set for φ_s, it corresponds to a W-Os intermetallic compound. As we know, it forms during cathode activation [26]. Values of φ_s for polycrystalline W-Os compounds are available in Makarov and Kultashev work [51], Mueller [52], Wall et al. [53] and Shroff [11]. The experimental points of Fig. 2.8 show a clear influence of the cathode operating temperature on the maximum value of J_0 that is reached, and also on J_0 decrease versus time. The higher the temperature, the higher the maximum reached value of J_0 but also the fastest decrease with time. One can also observe a phase of increasing emission prior to the decrease phase. The model predicts this behavior but to a lesser extent. The increase of J_0 in the calculated curves results from a decrease of the calculated work function, with the appearance of a minimum after a certain time of life test [11]. This can be understood since the minimum of the work function is estimated to occur for a value of Ba coverage less than unity [54, 55], for example, 2/3 as proposed by Dieumegard et al. as a result of calculation [34]. Adding to the evolution of the Ba coverage, some changes of the top-layer with time could participate in the evolution of the measured J_0. One evolution of the osmium top-layer is a surface micro-roughness that appears with heating, increasing the surface area, and hence the current delivered by the cathode. This positive phenomenon can contribute to the observed increase in the measured J_0. Unfortunately, it is difficult to quantify. The diffusion of W into the M-type top-layer increases the W proportion in the top-coating resulting in an expected increase of the work function and thus a decrease of the J_0. The diffusion of W into the cathode osmium–ruthenium top-layer has been determined experimentally [35, 56] and adequately fitted by Mita's model. Mita takes a value of 60% of W in the top-coating as the point when coating degradation starts provoking an increase of the work function, while its initial value lies in the typical 30–40% range [26, 35]. In Brion's work [26], the experimental values of J_0 versus the Os concentration are

reported, showing a steep slump toward values typical of an S-type cathode when Os falls down below 15%. This represents more margin of operation than the one considered by Mita, considering that the S-type situation (no osmium top-layer) is to be avoided. According to Mita's calculations [57, 58], the coating degradation becomes the dominant life-limiting factor over the consumption of the aluminates when the cathode temperature is above 1110 °C_B. Mita shows a calculation of the time when the surface coating starts to degrade as a function of temperature. Extrapolating his curve to 1030 °C_B, one can estimate that onset of degradation occurs at ~300 kh at this temperature. Thus, one can justify modeling the experimental results of Fig. 2.8 without the introduction of the calculated change of surface film composition. On the contrary, this should be done for conditions of time and temperature such that the onset of degradation is expected to occur. For instance, Longo has taken into account a change of work function versus W surface concentration in his life prediction model. As a remark, one can point out that the interpretation of life test results can become more difficult due to the combined effect of two main factors, namely Ba depletion and degradation of Os film. This characteristic of coated cathodes, in some manner, limits the possible temperature of accelerated life tests. This drawback does not exist for cathodes without surface coating (B-type, S-type, MM-type). With the possibility of life testing at high temperatures (i.e., at 1170 °CB) the MM-type cathode from Thales can be qualified for a 15 years service in about 4 months [59]. Nevertheless, Mita announces 31 as being its highest accelerating factor when running the life tests of M-type at 1100 °C_B. This offers one the possibility of observing the end of life of cathodes in a reasonable accelerated life test duration of the order of 10,000 h. From a practical point of view, one can make two important remarks about the long-life tests results presented in Fig. 2.8. Firstly, after 180,000 h in diodes, the cathodes set at 970 or 1000 °C_B still show an excellent average emission performance close to 7 A/cm^2. The cathode among these two groups with the lowest emission still exceeds slightly 5 A/cm^2 (not shown). Secondly, one observes an initial increase in emission underestimated by the model, but decreasing trends of the curves are satisfactorily described. It makes sense, in the scope of lifetime prediction, that the most important is the decreasing portion. In order to provide lifetime prediction in TWTs with maximum accuracy, Thales has calibrated the model of Shroff by fitting the average trend of the decrease of J_n on a group of seven TWTs having reached 160,000 h of life. This group was set at an initial cathode operating current of 0.6 A/cm^2 and a cathode temperature of 985 °C_B. The emission decrease ΔJ_n was 3% at the end of the test. A set of calculation of ΔJ_n versus the square root of time has been performed using these calibration curves, at two different values of initial J_n and temperatures (Figs. 2.9 and 2.10). The curves show the first phase of moderately decreasing J_n proportional to $t^{1/2}$ followed by a more abrupt decreasing, when the term featuring $1/J_{SC}$ is no longer predominant versus the term featuring $1/J_{TL}$. The end of life defined as the time necessary for a decrease of J_n of 10% occurs after 20 years for a cathode temperature $\leq$1000 °C_B. Thus, the M-type cathode modeled here is suitable for onboard TWTs of satellites (15 years of life operation is requested).

To broaden the picture on life test performance and life prediction models, it is of interest to mention works from other cathode experts. An alternative approach for

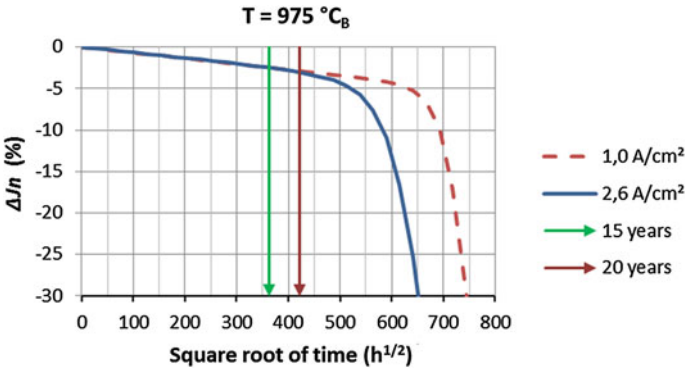


Fig. 2.9 Calculated curves of ΔJ_n versus time at 975 °C_B for M-type cathode using Shroff's model

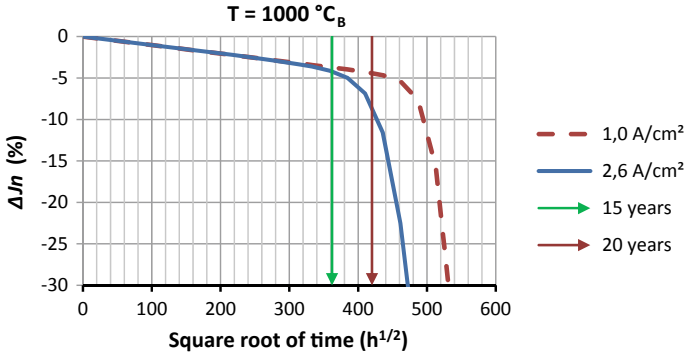


Fig. 2.10 Calculated curves of ΔJ_n versus time at 1000 °C_B for M-type cathode using Shroff's model

the prediction is to run curves of emission versus temperature at different intervals of time. Emission versus temperature has the form of the so-called Roll-Off curves [60] or Miram plots [61] or alternatively Dip-tests obtained by cutting the heater power [62]. The transition between the space charge limited regime and temperature-limited is called the "knee" and is derived from those measurements. The migration of the knee temperature versus time is plotted [60] and the projected EOL (End Of Life) corresponds to the point where the knee temperature equals the operating temperature. An approach based on the same principle has been developed by Mita [57, 58], who introduced the parameter T_{90} instead of the knee temperature. The temperature defined as T_{90} is the temperature at which the cathode current decreases to 90% of the space-charge limited current (and $\Delta T_{90} = T_{90} - T_{90\,\text{initial}}$). Mita measured the evolution of ΔT_{90} versus time on 4-1-1 M-type cathodes in simplified tubes and proposed the mathematical formulae matching the experimental points. Similar experiments were conducted by Chiba on 4-1-1 M-type cathodes in test tubes [62]. Both the authors find ΔT_{90} proportional to $t^{0.6}$. Life tests of 4-1-1 M-type

cathodes conducted in real CRTs by Toshiba [28, 29] and Thomson [40] show a dependence of the decrease of emission proportional to $t^{0.65}$ and $t^{0.5}$, respectively. This trend of emission decrease confirms the $t^{½}$ trend found by Dieumegard [34]. As a summary, the models which have proved to be useful for the prediction of the cathode lifetime are the ones from Shroff/Dieumegard [34], Tonnerre [34], Longo [35] and Mita [57, 58].

2.2.2.5 Emission Behavior Under Ion Bombardment and Gas Poisoning

Ion bombardment (=IB) and recovery was already discussed in the context of emission models and has been investigated for Os/Ru-I and Re-I cathodes by Gaertner et al. [50] and Cortenraad et al. [21]. G. Gaertner et al. observed a dynamic shielding effect in accelerated ion bombardment (life) tests caused by continuously replenished Ba, shielding O and the metal substrate below, which results in improved recovery under low ion doses and gives further experimental evidence to the model of Cortenraad et al., where Ba is sitting on top of an O bridge on the Os/Ru or Re or W surface. Gaertner et al. also derived the ion lifetime dose for an operation time of 20,000 h under typical CRT vacuum conditions as 1 mA mbar s in CRT electron guns and 10 mAs in diode configurations. The effect of locally peaked ion bombardment in a CRT gun was investigated and modeled by Higuchi et al. [63]. They found, that a fine-grained tungsten base layer with high pore density below the top-coating (here Ir) improved recovery under moderate IB [29, 63, 64].

An overview of emission poisoning of Ba dispenser cathodes by different gases is given by Cronin [65]. I cathodes are less sensitive to emission poisoning than oxide cathodes (see Chap. 4). The onset for emission poisoning by oxygen at 1000 °C is at $\geq 2.7 \times 10^{-7}$ mbar, the onset for air poisoning at 1015 °C occurs at $\geq 4 \times 10^{-6}$ mbar, the onset for CO_2 poisoning at 1020 °C happens at $\geq 1.8 \times 10^{-6}$ mbar and for water vapor at the same temperature at $\geq 1.3 \times 10^{-6}$ mbar. The trend is, the higher the operating temperature the less sensitive are I cathodes with respect to poisoning. O_2 and CO_2 poisoning of I cathodes has also been investigated by Sharma et al. [66]. This topic will be addressed again in the next chapter on scandate cathodes, where scandate cathodes and I cathodes will be compared also in this respect.

In real tube application (CRTs or microwave tubes), the environment plays a role in the emission capability of the thermionic cathodes (oxide and dispenser cathodes). In microwave tubes, metallic parts of pure metals or their alloys as well as brazing materials are possible sources of metal vapors. Pumping combined with bake-out steps of the tubes are in principle efficient to obtain a satisfactory level of vacuum to ensure the desired level of electronic emission from the cathode. Nevertheless, the outgassing of the constituting parts of the tubes can still occur during tube operation. Another situation during which cathodes can be poisoned is the outgassing of the anode in the close-spaced diode configuration. To understand and control the phenomena of poisoning, investigations by different authors to determine the effect of vapors or residual gaseous species on the emission capability of thermionic cathodes have been conducted. In their study on poisoning by vapors, Vaughn et al. [67]

have recorded the change in the temperature-limited emission in diode configuration while exposing a dispenser cathode at a fixed temperature (1100 $°C_B$) to poisoning vapors. The emission measurements were made at different brightness temperatures of the poisoning agent. The value of the saturated emission current versus time of exposure normalized to the un-poisoned starting current (I/I_0) was recorded, showing a maximum poisoning occurring after 10 min of exposure (I_0 is the saturated emission current without exposure to poisoning vapor). The authors observed that all metal vapors seem to poison the cathode to a final value I/I_0 obtained in 10 min, except copper which produced no harmful effect up to its melting point. Among the tested poisoning agents, Ag has the highest vapor pressure and poisons at the lowest temperature with Au, Cu–Ni, Ni, Ti, Fe, Pt, and Zr following in that order. It should be noted that the poisoned cathode could be re-activated by applying a hot-shot.

2.2.3 High-Resolution Surface Characterization and Correlation to Emission; Emission Model

As the emission properties of the thermionic cathodes depend mainly on the physic-ochemical composition of the cathode surface (here surface means the first atomic layers from cathode surface), Auger Electron Spectroscopy (A.E.S) has been often chosen as the most powerful technique to obtain the desired information on surface composition. In this respect, X.P.S (X-Ray Photo-electron Spectroscopy) represents a competing technique. In the 1970s, A.E.S had been preferred to X.P.S because its spatial resolution was by far more adapted to cathodes. A focused electron incident beam of the order of a few microns could be realized while the incident X-ray spot was in the cm range. Nowadays, the spatial resolution has greatly evolved for both techniques. X.P.S systems with 10 μm incident beams are available, but competing now with "nano-Auger" systems.

Several works on A.E.S have contributed to building models describing the cathode surface and the relation between its composition and the emission capability of the cathode. The studies have consisted in direct analyses on cathodes' surfaces, or on evaporation experiments of Ba and O on tungsten substrates together with A.E.S [21, 24–27, 68–70]. Most of these A.E.S characterizations were conducted at the operating temperature of the cathodes, to enable establishing a link between the surface chemistry and the emission capability of samples under tests. The link between the surface chemistry and the emission capability was best assessed when in situ measurements of cathode emission capability in the Auger system chamber could be performed [26]. These in situ emission measurements are made usually in a close-spaced diode configuration [21, 26] but other means have been employed, for instance, the shift in cut-off of low-energy electrons in a four-grid LEED system [27]. In order to study the Ba–O geometry and the Ba–O bonding, Cortenraad [21] has combined Low-Energy Ion Scattering (L.E.I.S) and A.E.S, with the possibility to relate these surface information with in situ emission measurements.

The low-energy spectrum in the range 60–85 eV brings useful information about the chemical bonding between the Ba and O atoms covering (fully or partially) the metal substrate. The low-energy region was studied in A.E.S. by the Naval Research Laboratory for BaO deposited from an external source on W, Re, and Ir substrates or Ir-W alloy substrates [68]. Their results can be compared to the low-energy spectra obtained by Cortenraad [21] on various W cathodes coated with Re, Ir, or Os/Ru, or those from Brion [26] on S-type or M-type cathodes (Os coating). A low-energy spectrum from a recent work of J.-M. Roquais presented at IVESC 2014 [71] on a pure osmium-coated 4-1-1 M-type cathode is shown in Fig. 2.11. This impregnated cathode type is similar to the one studied by D. Brion. The spectrum presents two Auger peaks of interest, one peak at 66 eV the other one at 71.2 eV. Interpretation of such peaks in the low-energy region has been provided by Haas et al. [68, 72, 73]. The peak at 66 eV corresponds to the Ba4d, Ba5p, O2p transition, the other one at ~71 eV to a Ba4d, Ba5p Ba 6s transition. The 71 eV peak accounts for Ba in the metallic state and the 66 eV for an "oxidized state", or said differently to the Ba–O electronic interaction. The relative heights between the two peaks represent a measure of the degree of charge transfer between Ba and O. This charge transfer between these two elements results from their difference in electronegativity. Haas et al. have patented the method to control the desirable surface chemistry of the cathode for optimum emission [73], consisting in measuring the ratio of the peak height of the 66 eV peak and the 71 eV peak. Derived from a set of data with statistical meaning obtained on cathodes similar to the one presented in Fig. 2.11, the average ratio is I_{pp} (66 eV)/I_{pp} (71 eV) ~ 0.56 (using peak-to-peak intensities I_{pp} on the derivative spectrum). These results are in good agreement with the values reported by Brion [26] and summarized by Cortenraad [21] for different substrates.

As a result of the deep studies by different analytical techniques of the Ba–O system on refractory metals of the past decades, the so-called "dipole model" has been adopted. Measurements by surface-extended X-ray absorption fine structure

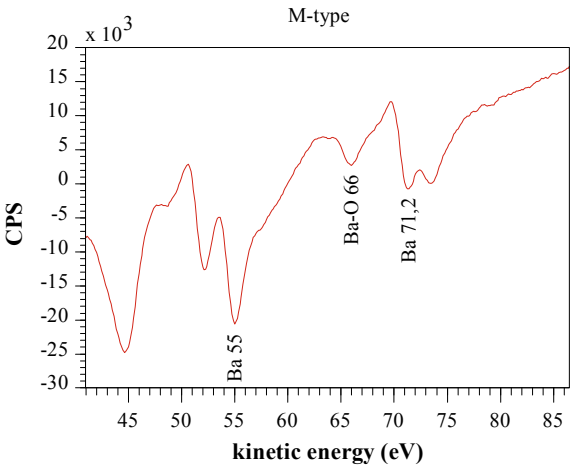


Fig. 2.11 Low-energy A.E.S spectrum of the surface of an osmium-coated M-type cathode

(SEXAFS) by Norman et al. [74] give evidence that Ba is bonded to oxygen with a well-defined short-range order. The oxygen atoms occupy the hollow sites of the substrate. The Ba–O dipoles formed on the refractory metal surface lower the "high" work function of the substrate, and therefore account for the good emission capabilities of dispenser cathodes. Nevertheless, some controversy still exists between the authors around the following points:

- What theory explains best the differences in the work function associated with the substrate composition?
- What is the geometrical arrangement of the Ba and O atoms on the surface?
- Are Ba and O atoms in a 1:1 ratio of quantities at the surface?
- What is the role of oxygen?

It is well known that for alkali or alkali-earth metals adsorbed on refractory metals, a minimum of the work function exists at an optimum sub-monolayer coverage [21, 31, 52, 70]. In his study, Cortenraad had evidence that the coverages by Ba he measured on various cathodes were below one monolayer. For his studied range of Ba coverages—i.e., below one monolayer—he showed that the work function decreased with increasing degree of coverage. The degree of coverage is dependent on the top-layer nature of the metal substrate, this dependence representing, for R. Cortenraad et al., the main reason for the lower work function on M-type compared to uncoated W type. According to these authors again, the degree of coverage of the Ba–O on the cathode surface is determined by the strength of the bonding of the O atoms in the Ba–O complex with the substrate atoms. A higher Ba coverage on M-type than on uncoated type (S-type) is a confirmed trend as it has also been found by other authors [21, 26, 75], the results being summarized in Table 2.3.

It should be noted that the minimum of work function whose value depends on the adsorbate/metal substrate couple is found by calculation for ≤ 1 [34] as depolarization is expected to result from the presence of too many dipoles.

The interpretation for the dependence of the lowering of the work function versus the nature of the substrate differs between authors. Haas et al. [68] have found similarities of variations between the amount of charge transfer between the Ba and O atoms and the variation of the lowering of the work function on an Ir-W substrate. Thus, they invoke the amount of charge transfer as being the factor that sets the lowering of the work function. In contradiction with those experimental findings, significant differences of work function between cathodes with different top-coatings were measured by Cortenraad et al. [21], while the amount of charge transfer was not different between those cathodes. As said above, these latter authors invoke more the degree of coverage by the Ba–O complex as the primary cause of the work function lowering, this degree of coverage being itself determined by the strength of the O bond to the substrate. Variable degrees of coverage by the Ba–O complex for different substrates are also introduced in the cluster approach by Mueller [52]. Other microscopic properties, i.e., the initial surface dipole μ_0, and the effective polarizability α vary with the modeled clusters. The cluster approach of

Table 2.3 Ratio of Ba coverage between M-type cathodes and uncoated cathodes, and related emission performance of characterized cathodes expressed in terms of current density or work function. The uncoated cathodes have a Ba coverage arbitrarily set at 1

Authors	Coating type	Ba coverage relative to uncoated cathode	Techniques for estimating Ba coverage	T (°C_B) during Auger and emission meas.	Cathode current density (A/cm^2)	Work function (eV)
Brion et al. [26]	Os	1.15	A.E.S	1000	M-type ~ 12 S-type ~ 4	
		1.28–1.4	X.P.S	1000		
Green [75]	Os-Ru (80-20)	1.42	A.E.S		Ratio of M-type/S-type >2	
Cortenraad et al. [21]	W (no coating)	def.: = 1	L.E.I.S	965		2.02
	Re	1.04				1.93
	Ir	1.13				1.89
	Os-Ru (80-20)	1.39				1.85

Mueller explains the influence of the substrate crystallographic nature on the value of the work function. The BaO/Os-W cluster with hcp lattice of the Os-W substrate has values of μ_0, α, and Ba coverage leading to a lower value of work function than the one of the BaO/W substrate (bcc lattice). The surface crystallography, i.e., bcc for pure W matrix and hcp for W-Os matrix was shown experimentally by X-Ray Diffraction (XRD) in a grazing-incidence geometry by Norman et al. [74]. The different interpretations for the work function lowering discussed above are summarized in Table 2.4.

As far as the surface geometrical arrangement of atoms is concerned, X.R.D combined with SEXAFS and A.E.S have allowed D. Norman et al. to propose a model of arrangement of the Ba and O atoms on the substrate surface. The O atoms are in the hollow sites of the substrate and the Ba atoms lie in a plane above the plane of the O atoms. Ion scattering results of Cortenraad et al. are consistent with this representation in the sense, that Ba is detected above O. Nevertheless, to comply with their respective experimental results, Cortenraad is in favor of a representation in which the Ba atom is directly above the O atom (upright Ba–O axis), while Norman is in favor of a tilted Ba–O axis.

The atomic ratio between O and Ba at the cathode operating temperature (965 °C Mo-brightness) has been found close to 1:1 by Cortenraad et al. [21] in their A.E.S studies, with special care given to the calibration of the Ba and O signals. The same 1:1 ratio had been found by Springer and Haas in their A.E.S studies [27]. This 1:1 ratio is not a systematic situation as Brion et al. have shown a significant variation of this ratio as a function of the cathode temperature between 800 and 1550 K [25].

Table 2.4 Models explaining the work function lowering

Authors	Samples type	Observations	Interpretation
Haas et al. [68]	BaO on Ir-W substrate (experimental)	Difference of the peak heights between 68 eV and 73 eV peaks depends on substrate nature. Lowering of the work function and difference of peak heights have the same variation versus Ir/W composition	The amount of charge transfer between Ba and O depends on the substrate composition. Its maximum value at 50%Ir/50%W accounts for the minimum of work function at this composition
Cortenraad [21]	S-type and M-type cathodes (Re, Ir, Os/Ru) (experimental)	68 eV/73 eV intensity ratios are similar for all cathodes types, i.e. independent of coating nature (Re or Ir or Os/Ru). The work function decreases with increasing degree of coverage by Ba as measured by LEIS. The Ba/O ratio is close to 1 for all cathodes	Bonding between Ba and O atoms is almost independent of the substrate type. The lowering of the work function is proportional to the number of dipoles. This number is determined by the strength of the bonding of the O atoms of the Ba–O complex with the substrate atoms
Müller [52]	S-type and M-type cathodes (Os, Pt, W/Os) (theoretical)	Calculated work functions by cluster approach match experimental values from other authors. The crystal lattice type and the dominant crystal face of the cathode surface are the characteristics of the substrate influencing the work function	The important properties, i.e., the initial surface dipole $\mu 0$, the effective polarizability α, and the surface coverage by dipoles N vary with the lattice type and the crystallographic orientation. The hcp crystal structure explains the low work function of Os-W substrates

It has been stressed by several authors that O plays a major role in the buildup of the Ba–O complex on the cathode surface. The regeneration of surface in Ba and O versus time after appropriate in situ cleaning was studied by Thomas [48] and Cortenraad et al. [21, 76]. The demonstration was made that the Ba resupply rate was faster on a pre-oxidized substrate. The diffusion length of Ba on such an oxidized surface is higher than on a non-oxidized one, resulting in a faster Ba regeneration at the surface. This larger diffusion length results from the stronger bond of Ba to the substrate in the presence of oxygen. For Cortenraad et al., the regeneration of the Ba–O complex on the cathode surface is limited by the O supply rate. Moreover, this O resupply is ensured by O atoms coming from the cathode pores [76].

2.2.4 Recent Developments on Dispenser Cathodes

To obtain high-current-density cathodes, one can be tempted to increase the operating temperature. Too high a temperature will increase the migration of W into the thin metallic top-layer (made of Os, Ir, Re, Ru, or their alloys), thus provoking the disappearance of the beneficial top-layer prematurely. Efforts to improve this situation have been made at Varian with the development of trilayer-impregnated cathodes [77]. To stabilize the surface composition, those cathodes comprise a porous buffer layer between the thin top-layer and the cathode porous matrix. More recently, a new dispenser cathode with dual-layer was described by Y. Li from the Chinese Academy of Science [78]. A thin layer of Re is deposited between the Os/W top-layer and the W pellet. During activation, Re diffuses to the surface to form a ternary alloy and acts as a barrier inhibiting diffusion of W to the surface. The same team developed a cathode coated with a thin ternary alloy (Ir/Re/W) showing 25–30% higher emission than a cathode with pure Ir coating [23]. The same type of emission improvement can be seen in the work by Fang and Maloney who found a lower work function with deposited top-layers of Os/Re/W ternary alloys compared to top-layers of Re/W or Os/W binary alloys [22]. An effective work function of 1.73 eV is reported by these authors.

2.2.5 Applications of Dispenser Cathodes

Dispenser cathodes have been produced in large numbers in the 1990s of the past century to serve as the electron source of CRTs for personal computers or television sets. Nowadays, they are still manufactured for various applications: microwave tubes, ion thrusters, E beam lithography. Different families of microwave tubes use the cathodes: TWTs, Klystrons, Gyrotrons [79], Magnetrons. Hereafter, the conditions of operation of the cathodes in different families of microwave tubes are detailed:

TWTs: In the field of satellite telecommunications, the dispenser cathodes are employed in ground-based (uplink) and space-based (downlink) TWTs. The TWTs are used as high-power amplifiers from L band to Q band, at RF power levels between few tens of watts and few kW [80]. In this case, the cathodes operate in CW mode. In recent years, the market associated with satellite systems has grown up, linked to an increasing demand on HDTV and high-speed internet. The need for higher output power and frequency has shifted the required emitting current density from ~0.7 A/cm^2 in the 1980s to 2.5 A/cm^2 nowadays. The cathodes used by Thales for space applications are 4.1.1 M-type or MM-type cathodes. The temperature of these cathodes is adjusted to an optimum value allowing delivering the required current density, with less than 10% decrease over 15 years.

For ground-based communication systems, the required lifetime is less than for space-based, the typical need being ~50,000 h.

TWTs are also widely used in radar, seeker, and ECM systems from L band to W band. For such applications, the operational missions are often short (<4 h) and repeated (every day for radars and jammers) in a harsh environment (vibrations, shocks, temperature cycles). Therefore, the dominant factor limiting the lifetime is, in general, the number of on/off cycles rather than the number of operating hours. In airborne or shipborne jammers, for example, the cathodes are operated at ~4 A/cm^2, with a minimum lifetime of ~20,000 h.

Klystrons: These tubes are used in medical, industrial, and science accelerators as the high power source (up to 100 MW pulse and up to 1.5 MW CW from UHF to X band) to feed the RF cavities which accelerate the particles (electrons, protons or ions) [80]. A well-known technology is the Multi-Beam Klystron (MBK) in which the multiple beams are obtained from several cathodes integrated into a single assembly. Such a cathode assembly made by Thales is shown in Fig. 2.12. The cathodes (4.1.1 M-type) are operated at ~1010 °C$_B$ and deliver a pulse current of 6 A/cm^2 in the MBK. In these conditions, the cathode lifetime is ~50,000 h.

Gyrotrons: These are mainly used in thermonuclear fusion research facilities (Tokamaks or Stellerators) to heat up the plasma at about 100 million °C by Electron Cyclotron Resonance Heating. Unlike the other electron guns, the Magnetron Injection Gun (MIG) features a cathode that operates in the temperature-limited regime. Space charge effects induce a spread of the transverse velocity of the electrons which degrades significantly the efficiency of the Gyrotron [79]. Hence, for this type of application, a temperature-limited regime is more appropriate than a space charge one. An example of a cathode for Gyrotron made by Thales is shown in Fig. 2.13. The cathode (4.1.1 S-type) is operated at a current density close to 2.5 A/cm^2 at a brightness temperature in the range of 900–950 °C$_B$.

Fig. 2.12 The 7 cathodes of the Multi-Beam Klystron (MBK)

Fig. 2.13 Cathode for gyrotron. The electron-emitting ring is delimited by two arrows

The requirements on electron sources for microwave tubes is always a higher and higher current density while maintaining a relatively low temperature to ensure long life and avoid high-voltage electrical insulation problems linked to Ba evaporation. Higher current densities, for instance, 5–10 A/cm^2 at operating temperatures below 950 °C$_B$ are a good target. Some recent applications, like Tera-Hertz imaging are now the most demanding in terms of narrow beams of high current density. The efforts to meet the higher current density requests have been in the past years oriented toward developing scandate dispenser cathode technologies. Compared to the M-type and MM-type cathodes (including optimum alloys) described in this subchapter, the scandate cathodes represent the real breakthrough in terms of emission capabilities.

This will be the subject of the following Chap. 3 of this book.

2.3 Reservoir Cathodes (Bernard Vancil, Michael Green)

2.3.1 Basic Structure and Historical Development

L-cathode

The first dispenser cathode or "L" cathode in 1948 was a reservoir cathode [2, 3]. It did not work very well and was supplanted in 1955 by the impregnated dispenser cathode [7–9]. An example of a reservoir cathode is shown in Fig. 2.14. The basic problem with the L-cathode was that it used barium carbonate ($BaCO_3$), the same

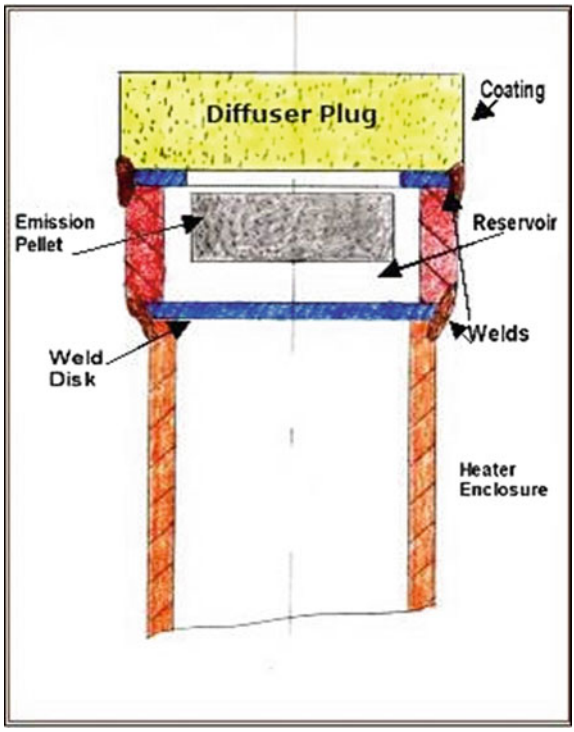


Fig. 2.14 Diagram of 2 mm diameter reservoir cathode

emissive material used in oxide cathodes. It is well known that oxide cathodes produce copious quantities of gas (CO_2 and CO) during conversion from carbonate to oxide. This is feasible in a cathode that is in an open environment, where gases can be pumped out quickly and where the carbonate is a thin layer. For a reservoir cathode, these conditions are not met. Thus, the L-cathode had a propensity to outgas for long periods. Also, hot tungsten can be oxidized by CO_2 and, in the presence of barium or barium oxides, barium tungstates can form in the pores in the diffuser, which sometimes become blocked.

MK Cathode

Later, the MK cathode was developed by Katz [4, 5]. It is shown in Fig. 2.15. It was manufactured for many years by Siemens. It avoided the problem of the L-cathode by using BaO in the reservoir along with tungsten wool to effect the conversion of BaO to free barium. It avoided the problems handling BaO hydration by coating the particles with eicosane ($C_{20}H_{42}$). This waxy, hydrophobic substance isolated the BaO from moisture in the air. The eicosane evaporates cleanly when the cathode is gently heated.

However, this cathode employed a pure tungsten matrix for the diffuser plug and consequently, it ran at higher temperatures than more modern cathodes such as the M cathode and the mixed matrix cathode. But it was tested in the Crane life test

facility [81] and achieved over 100,000 h with virtually no change in knee point. Its knee point migration is shown in Fig. 2.16. This was an important result considering the age of the design and its high operating temperature. The data is relevant to ion engines for space travel because a pure tungsten matrix may be required, due to the harsh ion environment. The MK cathode proves that reservoir cathodes can operate for very long times at very high temperatures. Incidentally, the temperatures recorded for it in Fig. 2.16 are brightness temperature. The other cathodes used true temperatures. A photograph of the Crane life test vehicle is shown in Fig. 2.17.

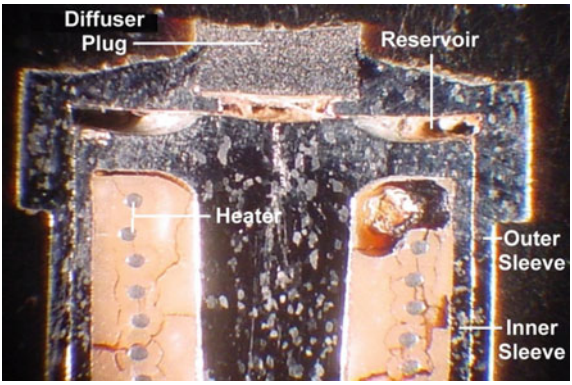


Fig. 2.15 Cross-sectional view of MK cathode

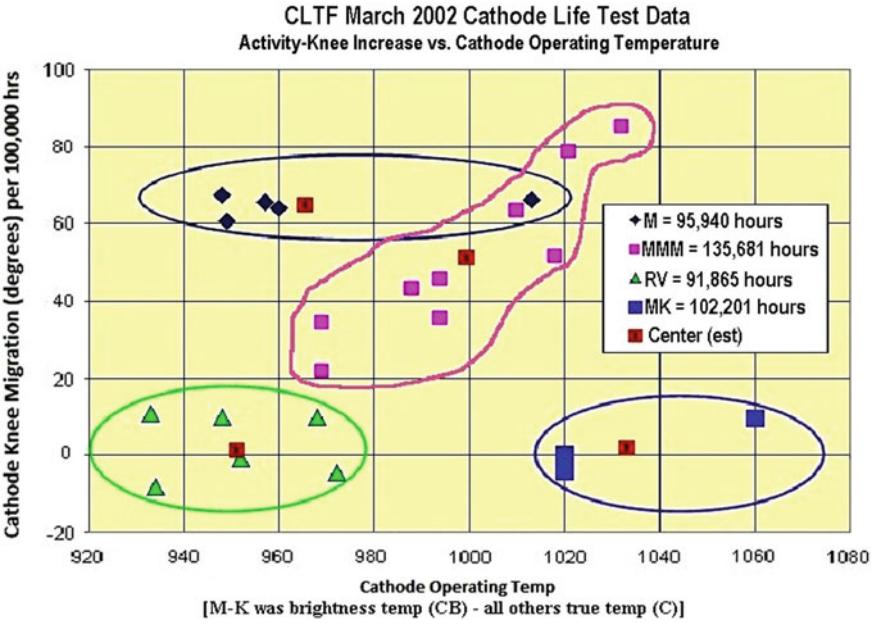


Fig. 2.16 Crane Cathode Life Test Knee Point Migration

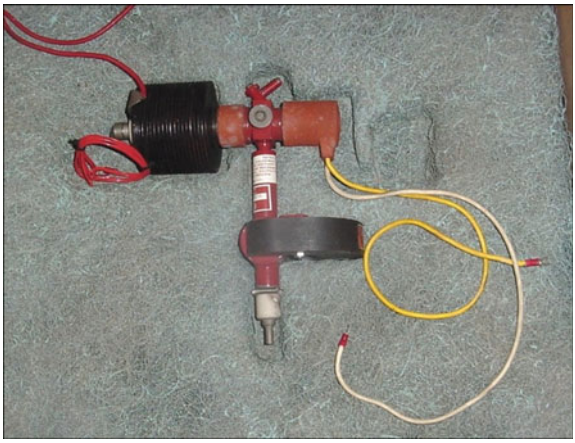


Fig. 2.17 Crane life test vehicle

The CPD Cathode

A miniature reservoir cathode was reported on in 1978 by Falce and Thomas. It was called the controlled porosity (CPD) reservoir cathode [82]. In addition to the cavity reservoir concept, it used an array of precise 5 μm laser-drilled holes in a 25-μm-thick iridium sheet over the reservoir containing barium oxide. The emission material was later changed to barium–calcium aluminate, and the sheet was changed to tungsten [83]. The problem with CPD cathodes has been their inability to control barium transport through the holes. A natural pore in a pressed tungsten matrix is on the order of one micron—whereas it is difficult to laser drill holes less than 5 μ. Also, the depth of laser-drilled holes is much less, so conductance is much higher.

One constraint is that individual laser holes have to be drilled with a single shot in order to drill the foil "on the fly" as it is rastered in X and Y on the laser work table. This is necessary to achieve a short laser drilling time and minimize the manufacturing cost. Initially, the holes were drilled at a rate of 50 per second using an off-the-shelf laser system, but the repetition rate of the laser pulses could be increased. It was possible to drill clean, tapered holes through 25-μ foil. The holes were 10 μ in diameter at the entrance end and 5 μ at the exit. The foil was oriented so that the 5-μ diameter hole mouths were positioned on the emitting surface of the CPD cathode. However, since the depth of laser-drilled holes is set by the foil thickness and the holes are much shorter than the pores in a sintered tungsten diffuser plug, as well as being larger in diameter, the conductance for barium is much higher. The barium evaporation rate from CPD cathodes was measured to be about six times higher than a conventional impregnated dispenser cathode at the 100-h life point. Excessive barium ended up creating many problems in both microwave tubes and CRTs

Considerable effort was expended to develop alternative methods of forming the pores in the CPD foil with the aim of increasing the depth of the holes and reducing their diameter. Ideally, the cross-sectional shape of holes should be a narrow slot rather than circular. This would increase the impedance to barium flow and

improve the coupling of the dispensed barium to the cathode surface. Semiconductor processing techniques such as reactive ion etching were used to drill narrow, slot-shaped holes in the foil with an increased depth-to-diameter ratio compared to the laser-drilled holes. However, this fabrication approach resulted in increased manufacturing costs.

Work on CPDs has continued with the "bundled wire" cathode recently reported on. This work is proceeding on reservoir photo-cathodes with cesium-containing compounds in the reservoir [84]. It is also being pursued as a standard thermionic cathode [85].

2.3.2 Performance of Reservoir Cathodes

2.3.2.1 Advantages and Disadvantages

Reservoir cathodes are a form of dispenser cathode in that barium is dispensed to the cathode surface, where it forms a dipole which reduces the work function of tungsten. They differ from conventional impregnated cathodes by having the barium compounds, which provide free barium, located in a reservoir behind the cathode matrix. The chemical factory which reduces the barium salts to free barium is no longer located in the pores of the matrix. An example of a reservoir cathode is shown in Fig. 2.14.

Reservoir cathodes have the advantage that the chemical factory for producing free barium is located in the reservoir, not in the matrix. Thus there are no inert reaction products, such as barium tungstates produced in the pores of the matrix. These coat the pore linings and inhibit the process whereby barium–calcium aluminate meets with tungsten to produce free barium, which then diffuses to the surface. In addition, the barium generated nearest the surface diffuses most easily to the surface. As this is depleted, the impregnant in the pores is consumed and the reaction zone recedes deeper into the porous matrix. Hence, as time passes, the barium must come from deeper within the cathode. Thus, barium has to traverse a greater length of pore with associated greater impedance to flow, and is less able to reach the surface. The result is a steady reduction in barium delivery to the surface of the cathode. The reduction in barium arrival at the cathode surface corresponds to a reduction of Ba and BaO evaporation rate, as shown for the respective quadrupole mass spectrometer ion currents (equivalent to evaporation rates) in Fig. 2.18a, b, based on data from Gaertner et al. (a) [42] and from Aida et al. (b) [86]. The result is a steady reduction in barium delivery to the surface of the cathode, until no Ba will be delivered at life end. Depending on the impregnated plug dimensions, a temperature increase above 1050 $°C_b$ will lead to a strongly reduced life. Reservoir cathodes meter barium to the cathode surface at a fixed rate.

In addition, the reservoir can be charged with a much larger amount of available barium than can be provided by the impregnant in the pores of a conventional

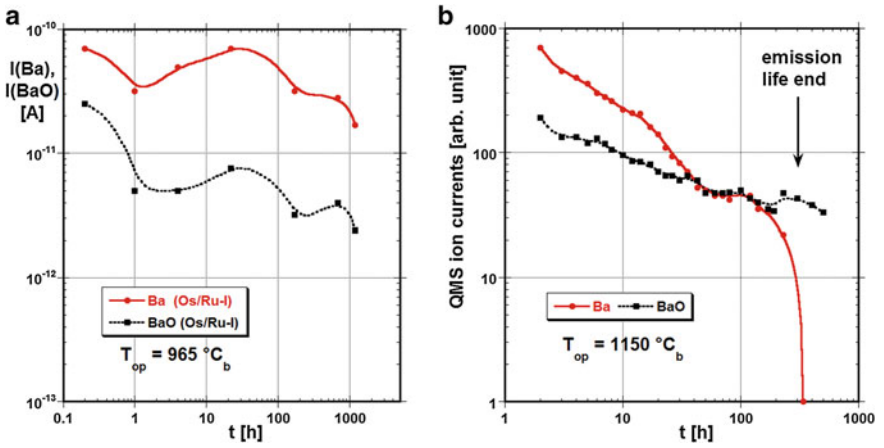


Fig. 2.18 **a** Ba and BaO ion currents [pA] for M-type cathode (Os/Ru-I) at 965 °C_b (based on [42]). **b** Ba and BaO ion currents [arbit. unit] for M-type cathode (Os/Ru-I) at 1150 °C_b (based on [86])

impregnated dispenser cathode. Thus, the operating life of a reservoir cathode can be made longer.

Another advantage of reservoir cathodes is that the cathode matrix is not restricted to materials that reduce barium compounds to elemental barium. Normally, this reducing element is tungsten. Reservoirs can use any material that is refractory and provides a low work function.

For reference, a cross section of an impregnated cathode is seen in Fig. 2.2. Barium–calcium aluminate filling the pores reacts with tungsten in the porous matrix. This produces free barium which migrates to the surface of the cathode, then diffuses across the emitting surface, forming a dipole layer that markedly lowers the surface work function. The barium in the dipole layer must be continually replenished due to evaporation from the surface at the elevated temperature at which the cathode operates.

The impregnated dispenser cathode dispenses barium at a high rate at the start of life when the impregnant in the pores is close to the cathode surface. However, as time goes on, the impregnant nearest to the pore mouths is consumed by reaction with tungsten and evaporation of barium and the reaction zone, where barium is produced recedes down the pores and deeper into the porous tungsten matrix. This increases the flow impedance for barium to reach the cathode surface (the barium is transported through the pores by a combination of surface diffusion on the pores' walls and Knudsen flow). Hence, as the cathode continues to operate, the rate of barium dispensation to the surface falls in proportion to the reciprocal of the square root of the operating time. This reduction in barium supply rate sets a limit on the lifetime of the impregnated dispenser cathode because the barium dispensation rate to the surface eventually falls to the point, where it can no longer support full coverage

of the emitting surface by the dipole layer and the surface work function increases and becomes more patchy.

In contrast to the above, a cross-sectional view of a reservoir cathode would reveal open pores, without impregnant, with barium being transported from the reservoir through the open pores of the diffuser plug by a combination of surface diffusion and Knudsen flow and finally forming a monolayer on the emitting surface of the cathode. One advantage is that a much greater quantity of dispensable barium can be stored in the reservoir than could ever be stored in the micron size pores of the tungsten matrix. Furthermore, there is no buildup of reaction products on the walls of the pores of the matrix; this buildup additionally impacts the production rate of barium with operating time in an impregnated cathode. Thus, the reservoir cathode is capable of an extremely long operating life.

In a reservoir cathode, the empty pores in the diffuser plug present a fixed impedance to the transport of barium. Hence, unlike an impregnated dispenser cathode, barium is metered to the cathode surface at a constant rate. However, one problem for the majority of reservoir dispenser cathodes is that the barium has to traverse the entire thickness of the porous diffuser plug. Hence, the transport distance along the pores is much longer than the path the barium atoms must travel in an impregnated cathode, where the barium is produced by chemical reactions within the pores from a reaction zone only a short distance below the emitting surface.

Typically, reservoir cathodes have been constructed with porous tungsten diffuser plugs fabricated using similar powder metallurgy methods and similar tungsten grain size to those employed for the porous tungsten matrices of impregnated dispenser cathodes. With comparable pore diameters and much greater pore lengths for barium flow, the conductance of the pores in the diffuser plugs of reservoir dispenser cathodes is far lower than the effective conductance of the pores in impregnated cathodes.

In the reservoir cathode, the mean free path of barium atoms is much longer than the pore dimensions, so transport through the pores is a mixture of surface diffusion of adsorbed Ba on the pore walls and molecular flow. The conductance of the pores is a combination of the conductance of the entrance orifice and the pipe conductance of the long, narrow, convoluted passage through the diffuser plug.

The mass flow of barium through the pores is set by the arrival rate of barium atoms at the entrance of the pore at the reservoir end. The low conductance of the pores means that it is essential to maintain a high partial pressure of barium vapor within the reservoir to maximize the arrival rate of barium atoms at the pore mouths. However, this also carries a risk: If the reservoir is breached, the barium flow through the pores of the diffuser plug will almost cease and the cathode will fail.

Various approaches have been employed to increase the production rate of barium within the reservoir, including different combinations and formulations of reactants. The simplest method is to increase the operating temperature of the cathode, and versions of reservoir cathodes used in microwave tubes in the 1970s had a reputation for operating hotter than equivalent impregnated dispenser cathodes.

Finally, unlike impregnated dispenser cathodes, the tungsten in the porous matrix is no longer needed to chemically convert the barium compounds to free barium. This

reduction reaction is carried out independently by tungsten powder or fine tungsten wire or wool within the reservoir.

Hence, emission-enhancing metals such as osmium, iridium, and rhenium can be incorporated in the porous matrix of the diffuser plug without any impact on the barium supply rate. In many cases, these metals (and their alloys with tungsten) have a lower work function and consequently higher emission than pure tungsten when activated by barium. However, Ru, Rh, Os, and Ir are poor reducing agents and, unlike tungsten, they do not react with barium oxide or with barium–calcium aluminate to liberate free barium.

This lack of reducing power complicates the fabrication of mixed metal matrix impregnated dispenser cathodes because it is necessary to ensure that sufficient free tungsten is still present in the matrix to act as a reducing agent. This has resulted in a lack of repeatability in the manufacture of W-Ir mixed metal matrix impregnated dispenser cathodes, note, for example the wide dispersion in operating temperatures on MMM (Ir-W) cathodes in Fig. 2.16.

But reservoir cathodes employing mixed metal matrix porous diffuser plugs can take full advantage of the work function reduction afforded by Rh, Os, and Ir, and their alloys, without suffering any impact on the production rate of barium, even if the tungsten in the mixed matrix porous diffuser is fully alloyed with the noble metal.

Mixed metal matrix reservoir cathodes are particularly suited for harsh operating environments, such as in ion thrusters, where there may be oxygen impurities in the propellant gases. Iridium and osmium are less reactive than tungsten.

2.3.2.2 Data Comparisons

The Crane life test, Fig. 2.17, provides the best data for comparing the efficacy of reservoir cathodes to other types of dispenser cathodes. Figure 2.16 shows the knee migration rate in degrees Celsius/100,000 h for 24 cathodes. Two types of reservoir cathodes and two types of impregnated cathodes were tested over a period in excess of 15 years. The knee is the transition point between temperature-limited flow and space-charge-limited flow on thermionic cathodes.

Figure 2.19 shows a family of activity curves of an e beam incorporated miniature reservoir cathode. The knee points are shown with arrows. At low temperatures and zero field, (which means space charge field compensated by the external field) these curves are temperature limited (TL) and follow the Richardson equation (2.1) with A_R in the order of the thermionic constant $A_{th} = 120.4$ [Acm^{-2} K^{-2}]

Modified by the Schottky effect, where a substantial electric field is applied, results in $J_{TL} = J_0 e^{4.4\sqrt{E}/T}$, where $E = U/d$ is given in V/cm and T in K. In the space charge limited (=SCL) regime above the knee, the emission is governed by Child's law

$$J_{SCL} = \left(2.33 \times 10^{-6}\right) \frac{U^{3/2}}{d^2} \tag{2.11}$$

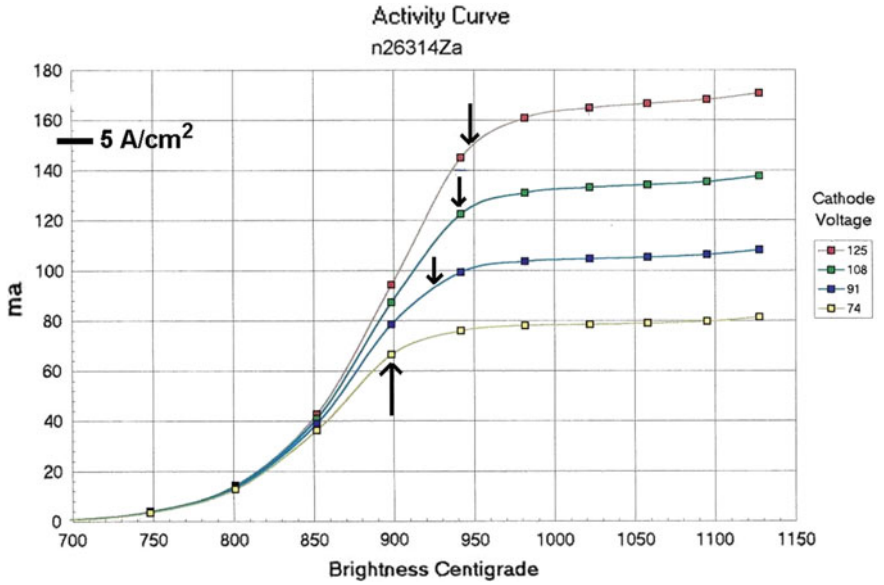


Fig. 2.19 Activity curves for e beam miniature reservoir cathode in close space diode with W-Os mixed metal matrix

where U is the positive voltage applied to the planar anode which is positioned close to the cathode (distance d) and parallel to it. The pre-factor has the dimension $[A/V^{3/2}]$.

In first-order Child's law, the emission is independent of temperature and the emission above the knee would not vary. This is seldom the case, however, on practical activity curves, as can be seen in Fig. 2.19. There is always a small increase above the knee, which can be explained by the second-order correction to Child's law, which is temperature dependent and is caused by the electron energy distribution (compare [87, p. 433]). Further contributions can come from the expansion of the support sleeve as temperature rises, which brings the cathode closer to the anode, or can be caused by high work function patches on the surface of even the best cathodes. These emit temperature-limited electrons as the temperature rises, and sometimes produce substantial slope above the knee point. The curve is, in that case, the superposition of Child's Law and the Richardson Equation. Because of this rise in emission with temperature, unrelated to work function, some workers prefer to measure current–voltage characteristics at given temperature (I-V curves) or lnI versus lnV curves to characterize their cathodes. A family of IV curves is shown in Fig. 2.20. Normally, a family of these curves is made, with each curve representing a different temperature.

The knee point in activity curves is usually sharper than for I-V curves. This allows more precise, less ambiguous assessment of cathode merit. But they do not show the full emission potential of the cathode. This is another reason activity curves are often not used by cathode researchers. The full temperature-limited emission in

Fig. 2.20 can be seen only by extrapolating the Richardson curve to the current at a given temperature that it would have achieved had there been no anode voltage limit. As a rule of thumb, the full temperature-limited emission is about twice the emission at the knee point. For example, in Fig. 2.19, the lowest curve have a current of 65 mA at the knee at 900 C_B. This current would obviously be much higher if space charge were removed. This is an important advantage in ion and hall thrusters for space propulsion, because space charge is negated by the plasma. In those devises emission current density is approximately 2 times that seen in vacuum devices at the same temperature. Furthermore, at high CW current densities, electron cooling is mitigated by ion heating of the cathode in those devices. This reduces stress on the heater which in vacuum must supply extra power to the cathode to prevent its temperature from dropping. These two benefits together make Os-W reservoir cathodes at 50A/cm^2 and scandate impregnated cathodes at 100A/cm^2 highly desirable in compact thrusters for small spacecraft such as Small Sats and Cube Sets.

In summary, we prefer activity curves because they allow precise pin-pointing of the knee temperature at a given current. This allows easy comparisons from one cathode to another. Moreover, we gain valuable insights about a cathode by viewing the sharpness of the knee and the amount of slope above the knee. The curves in Fig. 2.19 not only have low-temperature knees compared to other cathodes, but the knees are sharp and there is almost no slope at temperatures above the knees.

Returning to Fig. 2.16, the Crane life test, the two reservoir cathodes are the RV cathode from Varian and the MK cathode from Siemens. Despite their enormous

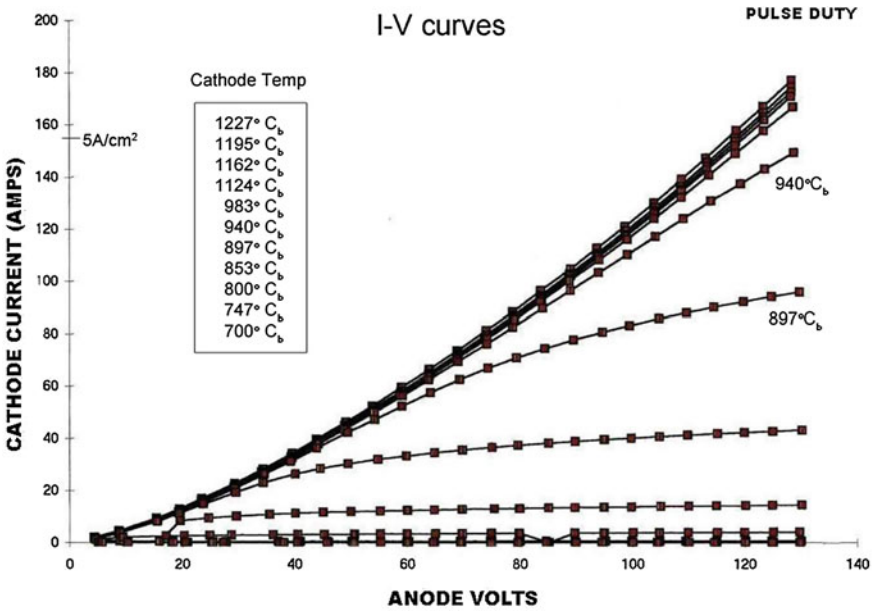


Fig. 2.20 Family of IV curves for e beam miniature reservoir cathode

difference in operating temperature, there is virtually no knee point migration over an operating lifetime of about 100,000 h. Most of the cathodes in the Crane life test were operated at 2 A/cm^2. Some in each group were operated at 4 A/cm^2. These can be seen as the points within a given perimeter that have higher operating temperatures. For example, the two rightmost cathodes among the RVs were operating at 4 A/cm^2. The one rightmost MK cathode was at 4 A/cm^2. One M cathode operated at 4 A/cm^2 and three of the mixed metal matrix iridium tungsten cathodes operated at 4 A/cm^2. All others operated at 2 A/cm^2.

It can be seen in Fig. 2.16 that both kinds of impregnated cathode had substantial knee point migration, whereas neither of the reservoir cathodes had any substantial knee point migration. This is the most important advantage that reservoir cathodes afford. The M cathodes have particularly severe knee point migration—about 70 °C per 100,000 h of life. When a knee migrates more than about 50 °C, the cathode is effectively at the end of its life. This assumes that the operating point was set 50 °C above the knee at 0 h. This was the case in the Crane life test. When a cathode's knee moves above the operating temperature, the cathode will fail unless the operating temperature is increased.

In M cathodes, there are two processes that push cathode knees higher. M cathodes have the usual problem of diminishing barium transport to the cathode surface due to the buildup of inert by-products in the pores of the matrix. They also have the problem of the inter-diffusion of the thin osmium layer at the surface into the tungsten matrix below. This reduction in osmium concentration at the surface raises the cathode's work function with temperature and time. Both the inter-diffusion problem and the barium production and transport problem are solved with reservoir cathodes. The extremely low work function, and therefore the low operating temperature of the RV cathode in Fig. 2.16 is proof that the inter-diffusion problem can be solved in reservoir cathodes. The RV cathode employed osmium in its matrix. The fourth cathode in Fig. 2.16, called MMM in the legend, also impregnated, incorporated a mixed matrix of iridium and tungsten. It is a further example of the problems one encounters in "improving" conventional impregnated cathodes. The performance of these cathodes is extremely erratic, spanning a range from 20 °C to almost 90 °C per 100,000 h in knee point migration. Meanwhile, its operating temperature is also erratic, spanning a temperature of 970–1035 °C. Incidentally, the temperatures are true temperatures. Only the MK cathode temperature is in brightness degrees Centigrade. However, the temperature adjustment factor, namely spectral emissivity, used on these cathodes is somewhat in doubt.

The reason the iridium cathodes in the Crane life test had erratic behavior was due to the fact that iridium does not convert barium salts to free barium. Consequently, these cathodes are deficient in barium. Some, for one reason or another, produced adequate barium, others did not. This is another problem that disappears when reservoir cathodes are used. Offsetting these advantages, namely flexibility on matrix composition and no pore blockage, is the fact that reservoir cathodes are more complex to build, especially very small versions of them.

2.3.3 Selected Designs and Recent Reservoir Cathode Developments

2.3.3.1 Design of the RV Reservoir Cathode

In view of its life-test performance, it is instructive to review the RV cathode design in more detail to elucidate the means by which the low work function and extended stable operating life was attained.

The stable emission characteristics, long life, and low operating temperature of the RV cathodes were achieved by a systematic approach to the design of the cathode, in which the performance constraints of both impregnated dispenser cathodes and of earlier designs of reservoir cathodes were first analyzed and assessed. Modifications were made to the RV cathode structure to remove the root causes of all time-dependent changes in the internal cathode operating parameters that could impact upon the cathode emission performance with life. The cathodes were designed with sufficient reserves of activator in the reservoir for 500,000 h of operation.

Heater

When designing for an operating lifetime of multiple hundreds of thousands of hours, the lifetime of ancillary components like the cathode heater must also be taken into account. The RV cathodes employed a conventional alumina-potted, cataphoretically coated tungsten heater. However, by changing the shape of the non-inductive heater winding, it was possible to fit a greater length of heater wire, together with a slightly larger than standard diameter wire, into the heater "can". The result was a considerably increased surface area of the heater wire. This reduced the required radiated watts per unit area of heater wire surface and enabled the heater wire in the RV cathode to run approximately 300 °C cooler than the heater of a comparably sized standard impregnated dispenser cathode at the same emitting surface temperature.

In order to ensure that there were no hot spots along the heater winding that might lead to premature failure, coaxial molybdenum cylinders were provided, both inside and outside the heater winding, to efficiently conduct heat from all parts of the wire to the reservoir and the cathode emitting surface.

Porous Diffuser Plug

It was known that previous reservoir cathode designs suffered from excessive impedance to barium flow through the pores of the diffuser to the emitting surface. This resulted in a higher required cathode operating temperature compared to a conventional impregnated dispenser cathode, and a lower resistance to poisoning by residual gases. (This was one of the factors that drove the replacement of the MK reservoir cathode by impregnated dispenser cathodes in microwave tubes in the early 1970s.)

The higher impedance to barium flow in the case of reservoir cathodes is not surprising when the porous tungsten diffusers are made from porous tungsten with properties similar to the material used for the tungsten matrices in standard impregnated dispenser cathodes. In the latter case, the tungsten grain size is nominally

4.5–5 μ average diameter, but in practice, the tungsten particle size distribution ranges from 12 μ down to submicron dust, and a histogram of the distribution of grain sizes is rather flat, not peaked. The pores are irregular in cross section and of the order of 2 μ in diameter. The porous tungsten is sintered to 80% ± 2% of solid tungsten density. It should be noted that at about 83% of full density, continuous pore-to-pore connectivity through the thickness of the tungsten begins to be lost.

In the reservoir cathode, the barium must traverse the whole thickness of the diffuser plug. With some difficulty, the diffuser plug can be machined as thin as 250 μ. Even so, the barium has to diffuse a distance of >100 pore diameters along the convoluted pores threading through the matrix.

For high manufacturing yield and for good dimensional stability and temperature uniformity of the cathode emitting surface, it is desirable that the diffuser plug is made considerably thicker than this, which further increases the impedance to barium flow. If typical impregnated cathode button thicknesses are used, the diffuser plug will be on the order of 1.5 mm thick.

By contrast, in a standard impregnated dispenser cathode, regardless of the thickness, the barium has to diffuse only a few microns from the "chemical factory" within the pores where it is produced to reach the emitting surface. It has been shown that most of the barium is dispensed from the impregnant-filled portions of the pores that are close to the cathode surface. If an impregnated dispenser cathode is, for example, 0.06" thick, most of the thickness of the impregnated matrix serves as mechanical support and heat spreader, not as a barium source.

Segregated Grain Size Tungsten Technology

The designers of the RV cathode had extensive experience with porous matrices made with segregated grain size tungsten powder and had access to a commercial vortex classifier capable of segregating GE tungsten powder into size bands approximately 0.8 microns wide. Porous tungsten made from segregated grain size powder has a narrow range of pore diameters and a very uniform distribution of pores throughout the material. As the particle size is increased, the pore diameter increases in proportion (for a given sintered density), and the impedance to barium flow drops sharply.

Segregated tungsten powder in size bands from 4 μ ± 0.8 μ up to 20 μ ± 0.8 μ was isostatically pressed at 60 kpsi into green compacts and these were sintered to 80% of full density. Diffuser plugs were machined from the resulting porous tungsten bars and the impedance of the plugs measured under molecular flow conditions, first with xenon (for experimental convenience) and then with barium. It was determined that a diffuser made of 80% dense 14-μ grain size porous tungsten would furnish an appropriate flow of barium to activate the cathode surface, while permitting the diffuser plug thickness to be considerably increased, giving improved dimensional stability and better lateral heat conduction.

Osmium–Tungsten Alloy Coating

A 5,000 Å thick layer of 55:45 osmium–tungsten alloy was deposited by co-sputtering on the tungsten diffuser to lower the work function of the activated cathode surface.

The composition of the sputter coat was chosen to give the lowest work function. This was based upon work at NRL [88] (see Fig. 2.25) and later augmented by emission testing of impregnated dispenser cathodes coated with films of either Os-W or Ir-W with multiple different elemental ratios.

The sputtering parameters were adjusted to give a stress-free Os-W alloy film with a controlled microstructure and a degree of texture in the film.

In view of the long required operating life of the RV cathodes, additional steps were taken to block inter-diffusion between the film and the underlying tungsten of the diffuser. This would otherwise change the composition of the film with time, increasing the tungsten content and moving away from the osmium: tungsten ratio associated with the lowest cathode work function.

CVD Osmium–Tungsten Diffusion Barrier Layer

In order to prevent inter-diffusion between the sputtered film and the underlying tungsten, the top-layer of the porous tungsten diffuser was converted to a mixed matrix layer of osmium–tungsten alloy prior to depositing the sputter coat.

This was carried out by first filling the pores of the diffuser plugs with copper. Next, the copper was etched with nitric acid to expose the pores to a controlled depth, followed by etching the tungsten with Murakami's etch to open the pores at the surface. A layer of osmium metal (in the form of dendritic microcrystals) was then deposited onto the surface and within the pores by CVD. The porous diffuser plugs were then heated in hydrogen to just below the melting point of copper so that the osmium sintered into the surface tungsten and the walls of the pores. Because of the large surface area of the dendritic crystals of osmium and their associated high surface energy, the sintering was rapid and a layer of osmium tungsten alloy was formed. Because the pores had previously been enlarged in diameter with Murakami's etch, the mouths of the pores were not blocked by this process. The CVD deposition and sintering process was repeated until the surface of the diffuser plug was composed of a layer of Os-W alloy 5 μ thick. The composition of the alloy layer was approximately 50% osmium at the surface. Finally, the copper was completely removed from the pores throughout the bulk of the diffuser plug. The 55:45 Os-W sputter coat was deposited on top of this alloy layer.

Since the alloy layer was approximately ten times the thickness of the sputter coating, the concentration gradient of osmium that drives diffusion into the underlying porous tungsten was greatly reduced. Assuming that the diffusion rate is proportional to the concentration gradient and obeys Fick's Law, the diffusion time scale would increase in proportion to the square of the diffusion length. Thus, the loss of osmium from the 55:45 Os-W 5,000 Å sputter coat would be slowed by two orders of magnitude by the underlying 5-μ mixed matrix alloy layer. In practice, this estimate is conservative, because it ignores the formation of intermetallic compounds between osmium and tungsten that further slows the diffusion process.

Sintered BaO Pellet

In some embodiments of the reservoir cathode, the same barium–aluminate mixtures that are melted into the pores of impregnated dispenser cathodes, for example, 4:1:1

or 6:1:2 mol ratio impregnant mixtures, are placed in the reservoir as powders to act as the source of BaO. This is done because of the difficulty of handling pure BaO powder due to its rapid reaction with atmospheric moisture. Impregnant mixtures such as 6:1:2 do not contain free BaO; instead, the available BaO is present as a solid solution in barium–calcium aluminate. This makes the 6:1:2 impregnant mix much less moisture sensitive than pure BaO. This simplifies the assembly of the cathode.

However, because the available BaO is in solid solution, its vapor pressure is reduced. This lowers the rate at which free Ba is produced in the reservoir by the reaction between BaO vapor and a reducer made of tungsten powder or wool. Some workers have partly compensated for the lower reactivity of the impregnant by mixing finely divided tungsten powder with the impregnant powder in the reservoir so that there is direct physical contact. The available BaO in the impregnant mix can then interact with tungsten by solid-state diffusion rather than vapor phase transport. This increases the rate of Ba production.

The designers of the RV cathode chose to use pure BaO in the reservoir as it provided the largest amount of available Ba for a given reservoir volume. Also, the vapor pressure of BaO in the reservoir is constant at a given temperature as long as any solid BaO remains.

They overcame the moisture sensitivity of BaO by avoiding the use of powder and placing the BaO in the reservoir in the form of a single, compact, solid pellet that could be coated with a thin layer of Eicosane to prevent any pick-up of moisture while handling it in air.

The BaO pellets were prepared by pressing barium peroxide powder into green compacts as a precursor. Barium peroxide is a pale yellow powder that does not react with water; hence it can be handled in air without elaborate precautions to exclude moisture. It loses oxygen to give BaO when heated to 500–600 °C. The BaO_2 pellets were heated with a slow initial temperature ramp to convert them to BaO, and the resulting porous BaO pellets were then raised to sintering temperature and fully densified. The diameter of the sintered, cylindrical BaO pellets was sized to be a snug fit in the inner diameter of the reservoir.

Tungsten Reducer

A large surface area of tungsten to reduce the BaO vapor in the reservoir to free barium was provided by winding tungsten wire on the order of 12 μ in diameter into a coil. Thanks to the small diameter, several hundred meters of wire could be wound into a coil that would fit inside the reservoir. For maximum stability of barium production rate, the tungsten wire reducer was loaded into the reservoir first, closest to the porous diffuser plug. This was followed by a perforated molybdenum disk to act as a spacer. Then the sintered BaO pellet was inserted and the reservoir was sealed. The intention was to prevent direct contact between the solid BaO pellet and the tungsten reducer wire while allowing BaO vapor to pass through the holes in the molybdenum spacer to react with the tungsten wire.

Summary on RV Cathode

The RV cathode was designed to operate at as low a temperature as possible by coating the emitting surface with an Os-W alloy film with a composition that would activate to a very low work function. The alloy film was stabilized against changes due to diffusion by an underlying mixed metal matrix layer. The porous diffuser plug was made using segregated grain size tungsten to increase its conductance for barium flow in order to ensure sufficient barium supply to fully activate the surface and resist poisoning. Pure barium oxide in the form of a fully dense, sintered pellet was used in the reservoir and sufficient BaO was provided for 500,000 h of operation.

2.3.3.2 Miniature Reservoir Cathode Development

Reservoir cathodes, despite their success in the Crane Life Test (Fig. 2.16) still failed to catch on commercially in the 1990s. This was partly due to their greater mechanical complexity, size, higher cost, and longer activation time. Also, there were few applications where ultra-long life was of primary importance. The CPD (Controlled Porosity Dispenser) cathode was a valiant attempt by Semicon Associates to offer a commercially viable miniature reservoir cathode. It failed for reasons noted in the history section of this chapter. Another attempt to reduce size and complexity came from the cathode ray tube industry. Tektronix developed a low-cost miniature impregnated cathode for oscilloscope CRTs [89]. This had the important feature that it used a self-supporting coated and darkened slip-in heater (see Fig. 2.21). This replaced the potted heater normally used with dispenser cathodes. The new heater dramatically lowered size, cost, and power consumption. It was also found to provide long life. The key was extending the heater sleeve backwards from the cathode. This enlarged the interface area between heater and sleeve isotherm sufficiently to lower

Fig. 2.21 Standard slip-in heater

Fig. 2.22 E beam miniature reservoir cathode (bottom) and e beam impregnated cathode (top)

heater temperature to a safe range below 1500 °C. The cathode in this geometry dissipated less than two watts, significantly lower than previous designs used in microwave tubes.

This cathode was the basis for a miniature reservoir cathode developed in the years 2000–2008 [90–92]. A comparison between the Tektronix-impregnated cathode and the new reservoir cathode is shown in Fig. 2.22. This was done under contract from NASA Glenn Research Center in Cleveland, Ohio. This device eventually achieved emission at temperatures comparable to M cathodes (Fig. 2.19). A diagram of the cathode was shown in Fig. 2.14. It dissipates under two watts. It employs a conventional powder metal pressed and sintered matrix. The sintered matrix contains one-micrometer diameter pores. Thus, it does not over-dispense barium as the CPD cathode did.

Indeed, there were issues of under-dispensing barium with tungsten four-micrometer powder mixed with 60 wt.% one-micrometer osmium powder. This problem was resolved by sealing the side of the cathode close to the reservoir. By leaving the side pores open, a low impedance route was left open for the barium from the reservoir to flow out. This reinforces the notion that the reservoir must have a positive pressure of barium vapor relative to the outside of the cathode. Once the side pores were sealed, good barium coverage occurred on the face of the cathode [91]. A photo of the laser weld more than halfway up the side of the matrix is shown in Fig. 2.23a. It is shown in cross section in Fig. 2.23b.

This weld is augmented with a Mo-Ru sealing layer above and on the weld. The various processes can be inferred from Fig. 2.14 cross section. The use of the Mo-Ru at 1900 °C means that the reservoir material cannot be installed in the reservoir until

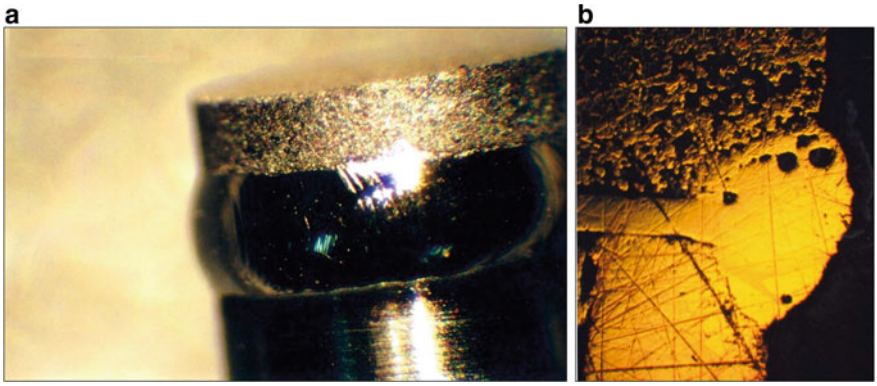


Fig. 2.23 **a** Close-up of laser weld. **b** Laser weld cross section

after this process is completed. Hence the reservoir body is left open on its lower end, away from the cathode.

The reservoir body is cut long to allow holding and turning while welding. Then it is cut off to the correct reservoir length. This design allows cathode diameters from 0.050 to 0.080 in. The Mo–Ru coating penetrates less than 0.005 inch into the matrix. It forms no brittle intermetallics with either tungsten or molybdenum. Following these processes, a bubble test is performed by attaching a tube to the open end of the reservoir, introducing nitrogen, and viewing ensuing bubbles under a microscope while the assembly is submerged in alcohol.

Next, the reservoir body is cut off to the correct length. The barium oxide-tungsten powder pressed pellet is loaded into the reservoir. This pellet is coated with Eicosane, similar to the MK and RV cathodes. The heater body is then laser welded to the cathode-reservoir assembly, seen in Fig. 2.14. The completed reservoir cathode is then mounted in the support structure seen in Fig. 2.24.

The data were encouraging. Emission data for a mixed matrix reservoir cathode employing osmium in the matrix were already shown in Figs. 2.19 and 2.20. The knee temperature at 5 A/cm^2 is comparable to M-type impregnated cathodes.

Mixed metal matrices containing varying concentrations of osmium, iridium, rhenium, and ruthenium are possible with reservoir cathodes. Figure 2.25 shows the work function reduction possible with alloys of tungsten and osmium obtained from a study of Thomas and Gibson [88]; similar results have been obtained for tungsten and iridium. A high concentration of osmium, iridium, ruthenium, or rhenium would inhibit barium conversion in an impregnated cathode. According to Fig. 2.25, the minimum work function for osmium–tungsten is near 60 wt.% osmium. This concentration will kill barium production in impregnated cathodes. At that concentration, the knee temperature at 5 A/cm^2 drops to about 910 °C$_b$. This can also be seen in Fig. 2.16 in the lower left. The RV cathode knee at 4 A/cm^2 is about 970 °C (true). Subtracting 60 °C for brightness correction yields 910 °C$_b$.

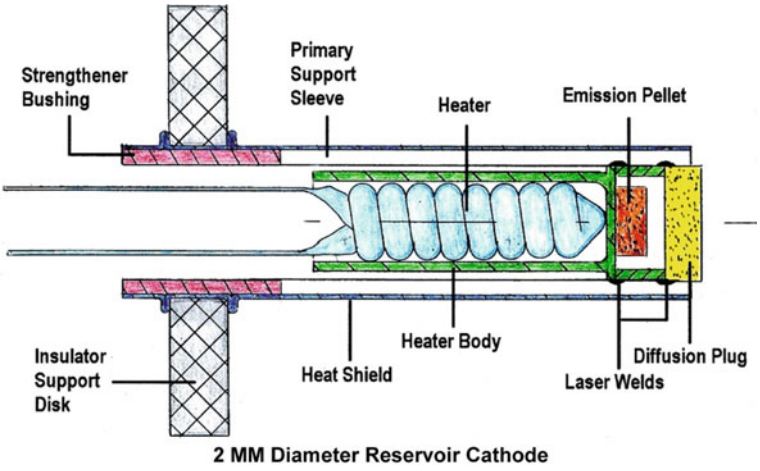


Fig. 2.24 Cross section of the complete reservoir cathode structure

However mixed metal matrix impregnated cathodes are sucessfully used by Thales in several of its devices, despite the issues with barium conversion. It is using osmium–tungsten in the ratio 70% tungsten and 30% osmium in ion neutralizers [93] for electric space propulsion. This still permits adequate barium production while significantly lowering work function and knee temperature from pure tungsten matrices.

2.3.3.3 Reservoir Cathodes for Electric Space Propulsion

The operating life of both ion thrusters and Hall Effect thrusters is to some extent limited by the cathode employed to provide ionization of the propellant gas. Current art is inadequate for long space missions, such as interplanetary flights. There are a number of advantages that reservoir cathodes offer for this purpose: (1) they allow alternatives to tungsten as the cathode emitting surface; (2) they allow far more barium than can be incorporated into the matrix of a conventional impregnated cathode; (3) they do not suffer from the barium transport problem of conventional cathodes. Barium is metered at a constant rate to the surface of the cathode. These three factors promise to give hollow cathodes for space flight an almost unlimited operating life. Furthermore, they can employ mixed Os-W matrices, which provide higher emission levels and lower temperature, and do not form barium tungstates in the matrix.

More powerful, long-lived ion and Hall thrusters are also needed for orbit-raising of heavy communications satellites from low earth to geosynchronous orbit. The orbit-raising phase may last several months. Following this, is 15 years of twice-daily ignition for station-keeping. Reservoir cathodes offer shorter orbit-raising times and more reliable operation over the subsequent 15 years.

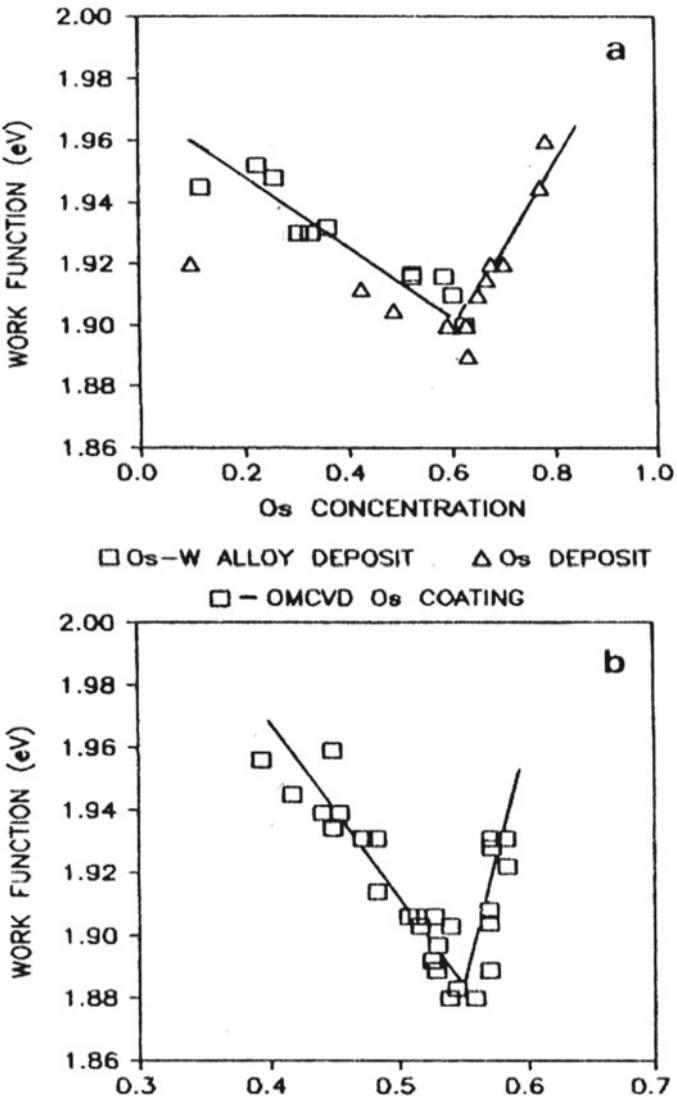


Fig. 2.25 Plot of work function versus Os concentration for three different types of surface coatings containing Os; see [88]

A project to incorporate reservoir cathodes into electric propulsion systems was funded by JPL in 2004. This project had a number of important results [94–96]. However, the cathode failed early in testing due to stress fractures in the diffuser body [97]. A new hollow reservoir cathode was developed by e beam, inc. in 2011 [98, 99]. A cross section is shown in Fig. 2.26.

A photo of the cathode insert is shown in Fig. 2.27.

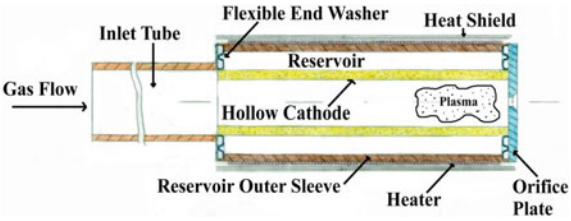


Fig. 2.26 Cross section of hollow reservoir cathode structure

Fig. 2.27 E beam hollow reservoir cathode

This cathode structure employs flexible end washers between cathode matrix and outer reservoir wall. This allows flexure and reduces stress due to thermal expansion mismatches between the outer reservoir wall and the tungsten insert. Thus, the stress fractures in the diffuser body were eliminated. This design has been extensively life tested and these tests were reported on in 2016 [99].

2.4 Conclusions and Future Aspects

A brief history of the development of the dispenser cathodes has been given, including a description of their various structures. A copious number of studies since the 1950s have allowed us to understand the chemical and physical phenomena taking place in those cathodes in such a way that their emission capabilities are well mastered [87]. At its operating temperature, Ba and O atoms forming a monolayer are always present on the dispenser cathode surface as revealed by analytical techniques like A.E.S., probing the outermost layers. The Ba–O complex on the surface of the metallic substrate accounts for the lowering of the work function, explaining the good emission capabilities of those cathodes (in the order of 10 A/cm^2 at 1000 °C). Semi-empirical models allowing estimates of cathode lifetime have been developed and have proven to be useful to choose the appropriate operating temperature adapted to the desired application. The dispenser cathodes are the source of electrons in various families of vacuum tubes ensuring reliably the requested lifetime, for instance, 15 years as is the case in TWTs on board of communication satellites. They find an application in different domains: medical, telecommunications, large instruments for science. As the trend is to develop devices requiring always higher

and higher current densities, efforts have been made in the past decades to achieve it by proposing more complex structures like the cathodes coated with a ternary alloy or coated MM-type (CMM). These approaches have led to some success in lowering the work function resulting in improved current density. In some recent works Wang et al. [100] have investigated Re-W mixed matrix impregnated cathodes and found lower evaporation rates of Ba and BaO compared to W-I cathodes, yet their emission capability lower than for M cathodes. M. Ravi et al. have studied W-Ir cathode pellets impregnated with lithium oxide doped barium–calcium aluminates [101]. They claim a work function of 1.81 eV and a life of >32,000 h with a Schottky extrapolated emission current density of 32 A/cm^2, which according to 10% deviation from space charge limit is about 25 A/cm^2, yet their test load conditions are quite unclear, which they should clarify in a further publication. Yet Li-oxide doping could be an interesting approach for improved I cathodes.

Nevertheless, the real technological breakthrough is associated with the further development of Scandate cathodes, as described in the next chapter.

In the meantime impregnated cathodes and reservoir Ba dispenser cathodes remain the workhorse of thermionic cathodes, especially for very long life space applications due to their reliability and the ability to predict their life based on reliable models and experiments and suitable design.

References

1. G. Herrmann, S. Wagener, *The Oxide Coated Cathode*. Chapman & Hall, London (1951)
2. H.J. Lemmens, M.J. Jansen, R. Loosjes, A new thermionic cathode for heavy loads. Philips Tech. Rev. **11**, 341 (1950)
3. H.J. Lemmens, U.S. Patent 2543728, published 27 Feb 1951, priority The Netherlands 26 Nov 1947
4. H. Katz, U.S. Patent 2750527, published 12 June 1956, priority Germany 19 Nov 1951 H. Katz claims earlier priority: French patent specification 903976, publ. 23 Oct 1945, priority Germany 19 Aug 1942
5. H. Katz, Metal capillary cathodes a new type of cathode. Siemens Halske Res. Bull. II, 123–126 (1951); J. Appl. Phys. **24/5**, 597–603 (1953)
6. P. Zalm, A.J.A. van Stratum, Osmium dispenser cathodes. Philips Tech. Rev. **27**, 69–75 (1966)
7. R. Levi, The impregnated cathode. Philips Tech. Rev. **19**, 186 (1957)
8. R. Levi, R. Hughes, U.S. Patent 2700000, Appl. 27 Feb 1952, granted 18 Jan 1955
9. R. Levi, New dispenser type thermionic cathodes. J. Appl. Phys. **24**, 233 (1953)
10. J.L. Cronin, Modern dispenser cathodes. IEE Proc. **128/1**, 19 (1981)
11. M. Shroff, Revue Technique Thomson-CSF **23/4**, 947 (1991)
12. D.N. Hill, Thermochemistry of dispenser cathode impregnant materials: phase equilibria in the $BaO{\cdot}CaO{\cdot}Al_2O_3$ system. Georgia Institute of Technology, School of materials engineering, Atlanta, Georgia, 30332 (1986)
13. A. Shroff, U.S. Patent 5334085, Process for the manufacture of an impregnated cathode and cathode obtained by this process, priority 9 Feb 1990 France, granted 2 Aug 1994
14. F. Bossert, R. Lotthammer, Multilayer MM-cathodes with specific additives. ITG Fachbe. **132**, 23–28 (1995). "Vakuumelektronik und Displays" conference, Garmisch-Partenkirchen
15. N. Koch, H.P. Harmann, G. Kornfeld, Status of the Thales tungsten/osmium mixed-metal hollow cathode neutralizer development, in *International Electric Propulsion Conference*, Florence, Italy, 17–20 Sep 2007

16. L. Falce, Dispenser cathodes: the current state of the technology. IEDM **83**, 448 (1983)
17. L. Falce, Electron tube with dispenser cathode. U.S. Patent 4165473, filed 27 May 1977, granted 21 Aug 1979
18. C.R.K. Marrian, A. Shih, The operation of coated tungsten-based dispenser cathodes in non-ideal vacuum. IEEE Trans. Electron Devices **36**(1), 173 (1989)
19. Y. Ouchi, S. Kimura, T. Higuchi, K. Kobayashi, M. Nikaido, K. Homma, K. Yanagibashi, Cathode characteristics and operating mechanism of Ir coated dispenser cathode. IECEJ. Technical Report, ED85-118 (1985)
20. S. Kimura, K. Homma, M. Nikaido, K. Kobayashi, T. Higuchi, Y. Ouchi, Long-life high reliability Ir-coated dispenser cathode. IEDM **87**, 689 (1987)
21. R. Cortenraad, A.W. Denier van der Gon, H.H. Brongersma, G. Gaertner, D. Raasch, A. Manenschijn, Surface analysis of thermionic dispenser cathodes. Appl. Surf. Sci. **191**, 153–165 (2002)
22. C.S. Ares Fang, C.E. Maloney, Surface studies of Os/Re/W alloy-coated impregnated tungsten cathodes. J. Vac. Sci. Technol. A **8**(3), 2329–2332 (1990)
23. H. Zhang, Y. Liu, M. Zhang, Y. Li, Emission and surface characteristic of ternary alloy Ir/Re/W-coated impregnated tungsten cathodes. Appl. Surf. Sci. **251**, 130–133 (2005)
24. G.A. Haas, C.R.K. Marrian, A. Shih, Interpretation of A.E.S data of impregnated cathodes. Appl. Surf. Sci. **24**, 430–446 (1985)
25. D. Brion, J.C. Tonnerre, A.M. Shroff, Auger spectroscopy investigations of various types of impregnated cathodes. Appl. Surf. Sci. **16**, 55–72 (1983)
26. D. Brion, J.C. Tonnerre, A. Shroff, Electron emission and surface composition of osmium and osmium-tungsten coated dispenser cathodes. Appl. Surf. Sci. **20**, 429–456 (1985)
27. R.W. Springer, T.W. Haas, Auger electron spectroscopy study of cathode surfaces during activation and poisoning. I. The barium-on-oxygen-on tungsten dispenser cathode. J. Appl. Phys. **45**, 5260 (1974)
28. S. Yamamoto, S. Taguchi, T. Aida, S. Kawase, Some fundamental properties of Ir coated impregnated cathodes. Jap. J. Appl. Phys. **23/4**, 214–217 (1984)
29. T. Higuchi, S. Matsumoto, K. Koyoma, A. Hara, H. Hamamoto, Life estimation of Ir-coated dispenser cathodes and heaters for cathode ray tubes. Appl. Surf. Sci. **146**, 109–116 (1999)
30. E.P. Gyftopoulos, J. Levine, Work function variation of metals coated by metallic film. J. Appl. Phys. **33/2**, 67–73 (1962)
31. E.P. Gyftopoulos, D. Steiner, Orbital electronegativity and physical properties of bimetallic adsorption systems, in *27th Annual Conference on Physical Electronics* (MIT, 1967), p. 169
32. J.B. Moreau, L'ionisation de contact. Revue de Physique Appliquée **3**, 286–304 (1968)
33. A.M. Shroff, G. Firmain, Long-life travelling-wave tubes-use of "M" type cathode life prediction model, in *International Electron Devices Meeting*, vol. 31 (1985), pp. 346–349
34. D. Dieumegard, J.C. Tonnerre, D. Brion, A.M. Shroff, Appl. Surf. Sci. **111**, 84 (1997)
35. R.T. Longo, E.A. Adler, L.R. Falce, IEDM Tech. Digest., 318 (1984)
36. R.T. Longo, J. Appl. Phys. **94/10**, 6966 (2003)
37. R. Vaughan, IEEE Trans. Electron Dev. **33/11**, 1925 (1986)
38. P. Palluel, A.M. Shroff, J. Appl. Phys. **51**(5) (1980)
39. T. Higuchi, S. Matsumoto, T. Yakabe, Y. Sato, S. Koshigoe, T. Yoshii, Emission life expectancy of Ir-coated dispenser cathode for CRT's, IEDM 91–935
40. J.M. Roquais, F. Poret, R. le Doze, J.L. Ricaud, A. Monterrin, A. Steinbrunn, Ba depletion study on impregnated cathodes and lifetime prediction. Appl. Surf. Sci. **215**, 5–17 (2003)
41. E.S. Rittner, R.H. Ahlert, W.C. Rutledge, Studies on the mechanism of operation of the L cathode. I. J. Appl. Phys. **28**(2), 156–166 (1957)
42. G. Gaertner, P. Geittner, D. Raasch, D.U. Wiechert, Supply and loss mechanisms of Ba dispenser cathodes. Appl. Surf. Sci. **146**, 22–30 (1999)
43. M. Koitabashi et al., SID 1994 Digest, 389, (1994)
44. I. Brodie, R.O. Jenkins, W.G. Trodden, Evaporation of Barium from Cathodes Impregnated with Barium-Calcium-Aluminate. J. Electron. Control **6**, 149–161 (1959)

45. A. Shih, J.E. Yater, R. Abrams, Thermal desorption of Ba from tungsten. Appl. Surf. Sci. **146**, 1–6 (1999)
46. A.A. Fote, K.T. Luey, Barium transport processes in impregnated dispenser cathodes, in *Tri-Service Cathode Workshop*, April 1980
47. B.A. Free, R.G. Gibson, Dependence of surface coverage on pore geometry in dispenser cathodes. Appl. Surf. Sci. **24**, 358–371 (1985)
48. R.E. Thomas, Diffusion and desorption of barium and oxygen on tungsten dispenser cathode surfaces. Appl. Surf. Sci. **24**, 538–556 (1985)
49. A. Manenschijn, P.V.D. Heide, S. Deckers, in *Tri-service/NASA Cathode Workshop, Cleveland, Conference Record* (1994), pp. 177–182
50. G. Gaertner, P. Geittner, D. Raasch, A. Ritz, D.U. Wiechert, Dynamic shielding during ion bombardment of Ba dispenser cathodes. Appl. Surf. Sci. **146**, 12–16 (1999)
51. A.P. Makarov, O.K Kultashev, A work model for barium dispenser cathodes with the surface covered by metals Os, Ir or Os(Ir) W alloy layer. Appl. Surf. Sci. **111**, 56–59 (1997)
52. W. Mueller, Computational modeling of dispenser cathode emission properties. I.E.D.M 91, 399 (1991)
53. R.N. Wall, D.L. Jacobson, D.R. Bosch, The high-temperature electron emission and vaporization of tungsten-osmium alloys. Metall. Trans. A **24A**, 951 (1993)
54. K.L. Jensen, Y.Y. Lau, B. Levush, Migration and escape of barium atoms in a thermionic cathode. IEEE Trans. Plasma. Sci. **28/3**, 772 (2000)
55. R.E. Thomas, J.W. Gibson, G.A. Haas, R.H. Abrams Jr., Thermionic sources for high-brightness electron beams. IEEE Trans. Electron Dev. **37/3**, 850 (1990)
56. N. Mita, Degradation of coated impregnated cathode's surface coatings. IEEE Trans. Electron Dev., **38/11**, 2554 (1991)
57. N. Mita, Degradation factors of a coated impregnated cathode. IEEE Trans. Electron Dev. **39/9**, 2172 (1992)
58. N. Mita, An accelerated life test method for highly reliable on-board TWT's with a coated impregnated cathode. IEEE Trans. Electron Dev. **41/7**, 1297 (1994)
59. M. Brueck, T. Hilble, S. Morel, J.M. Roquais, H. Seidel, Verification of the TED cathode models by long-term life-test data, in *IEEE International Vacuum Electronics Conference = IVEC 2009*, Rome, Book of Abstracts (2009), p. 527
60. D. Windes, J. Dutkowski, R. Kaiser, R. Justice, Triservice/NASA cathode life test facility. Appl. Surf. Sci. **146**, 75–78 (1999)
61. M. Cattelino, G. Miram, Predicting cathode life expectancy and emission quality from PWFD measurements. Appl. Surf. Sci. **111**, 90–95 (1997)
62. A. Chiba, Y. Akiyama, Life-test evaluation of 4-1-1 M cathode for highly reliable satellite TWT's. Appl. Surf. Sci. **146**, 120–125 (1999)
63. T. Higuchi, S. Yamamoto, H. Kudo, H. Murata, Modeling of life deterioration by ion bombardment of a dispenser cathode coated with an Ir/ W film. Appl. Surf. Sci. **200**, 125–137 (2002); T. Higuchi, M. Sasaki, et al., Modeling of emission slump by ion bombardment of a Ba dispenser cathode in an electron tube. IVEC/IVESC 2012, Paper 007 (2012), pp. 1–2
64. S. Kimura, T. Higuchi et al., Emission characteristics of dispenser cathodes with a fine-grained tungsten top-layer. Appl. Surf. Sci. **111**, 60–63 (1997)
65. J. Cronin, Practical aspects of modern dispenser cathodes-poisoning of impregnated dispenser cathodes. Microwave J., 4 (1979)
66. A. Sharma, A. Chopra, R. Mathew, Appl. Surf. Sci. **40**, 97–101 (1989)
67. J. Vaughn, K. Dudley, L. Lesensky, The deactivation of impregnated type cathodes due to metal vapors, in *Conference on Tube Technology* (1960), p. 140
68. G.A. Haas, A. Shih, C.R.K. Marrian, Interactions of BaO with mixed-metal substrates: W-Ir. Appl. Surf. Sci. **24**, 447–459 (1985)
69. D. Jones, D. Mc Neely, L.W. Swanson, Surface and emission characterization of the impregnated dispenser cathode. Appl. Surf. Sci. **2**, 232–257 (1979)
70. K.G. Eyink, B.C. Lamartine, W.V. Lampert, T.W. Haas, Quantification of the surface coverage of Ba and O on W substrates using Auger Electron Spectroscopy. Appl. Surf. Sci. **20**, 215–227 (1985)

71. J.M. Roquais, Surface characterization of dispenser cathodes at operating temperature and in-situ emission measurements, in *Proceedings of IVESC-ICEE-2014*, Saint-Petersburg, Russia (2014), p. 220
72. G.A. Haas, C.R.K. Marrian, A. Shih, Interatomic Auger analysis of the oxidation of thin Ba films. Appl. Surf. Sci. **16**, 125–138 (1983)
73. G.A. Haas, C.R.K. Marrian, A. Shih, Method for predicting the performance of cathode materials. US patent 4492866, Appl. 2 May 1983, granted 8 Jan 1985
74. D. Norman, R.A. Tuck, H.B. Skinner, P.J. Wadsworth, T.M. Gardiner, I.W. Owen, C.H. Richardson, G. Thornton, Surface structure of thermionic-emission cathodes. Phys. Rev. Lett. **58/5**, 519 (1987)
75. M.C. Green, Dispenser cathode physics. RADC-TR-81–211 (1981)
76. R. Cortenraad, A.W. Denier van der Gon, H.H. Brongersma, G. Gaertner, D. Raasch, A. Manenschijn, Dynamic behavior of thermionic dispenser cathodes under ion bombardment. J. Appl. Phys. **89/8**, 4354 (2001)
77. M.C. Green, Tri-layer impregnated cathode. U.S. patent 5418070, filed 28 Apr 1988, granted 23 May 1995, pp. 4113–4118
78. Y. Li, H. Zhang, P. Liu, M. Zhang, A new dispenser cathode with dual-layer. Appl. Surf. Sci. **251**, 126–129 (2005)
79. M. Thumm, Present developments and status of electron sources for high power gyrotron tubes and free electron masers. Appl. Surf. Sci. **111**, 106 (1997)
80. G. Faillon, G. Kornfeld, E. Bosch, M. Thumm, Microwave Tubes, Chapter 1 in *Vacuum Electronics–Components and Devices*, ed. by J. Eichmeier, M. Thumm (Springer, 2008), pp. 1–84
81. L. Dressman et al., *Cathode Life Test Facility Annual Report* (Crane Division Naval Surface Warfare Center, 2002)
82. L. Falce, R.E. Thomas, The controlled porosity dispenser cathode: iridium-barium oxide. *IEDM Tech. Digest*, 156–159 (1978)
83. L. Falce, The tungsten controlled porosity cathode: promise for the future, in *TriService Cathode Workshop* (1982)
84. E. Montgomery, B. Riddick, D. Feldman, A. Bolding, R. Ives, L. Falce, Lifetime and quantum efficiency advances in self-healing controlled porosity reservoir photocathodes, in *IEEE International Vacuum Electronics Conference, IVEC 2014*, pp. 131–132. https://doi.org/10.1109/ivec.2014.6857524
85. R. Ives, L. Falce, G. Collins, D. Marsden, G. Miram, S. Schwartzkopf, B. Smith, High current density long life cathodes for high frequency applications, Session 6.1, in *International Vacuum Electronics Conference*, Monterey, CA, 18–20 May 2010
86. T. Aida, H. Tanuma, S. Sasaki, T. Yaguchi, J. Appl. Phys. **74**, 6482–6487 (1993)
87. G. Gaertner, H.W.P. Koops, Vacuum electron sources and their materials and technologies", in Chapter 10 of *Vacuum Electronics, Components and Devices*, ed. by J. Eichmeier, M. Thumm (Springer, 2008)
88. R.E. Thomas, J.W. Gibson, Work function variation versus alloy concentration for dispenser cathodes. Appl. Surf. Sci. **29**, 49–66 (1987)
89. B. Vancil, Cathode Structure. U.S. Patent 4954745, filed 22 Mar 1989, granted 4 Sep 1990
90. B. Vancil, E. Wintucky, Miniature reservoir cathode–an update. Appl. Surf. Sci. **215**, 18–24 (2003)
91. B. Vancil, E. Wintucky, Weld techniques for reservoir cathodes. Appl. Surf. Sci. **251**, 101–105 (2005)
92. B. Vancil, *Long-Life High-Current Density Cathode for Low-Convergence Electron Beam Guns*. Final Technical Report, Navy Contract No. N65538-07-M-0161, Dec 2007
93. G. Kornfeld, N. Koch, H.-P. Harmann, Development & test status of the Thales W-Os mixed metal matrix hollow cathode neutralizer HKN5000, in *Presented at the 29th International* Electric Propulsion Conference, Princeton University, 31 Oct–4 Nov 2005
94. L. Schoenbeck, N. Hill, R. Shafer, W. Ohlinger, Barium source material development for reservoir hollow cathodes. *J. Am. Inst. Aeronaut. Astronaut*. (2004). 40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit. https://doi.org/10.2514/6.2004-4209

95. J. Vaughn, T. Schneider; J. Polk, D. Goebel, W. Ohlinger, N. Hill, NEXIS reservoir cathode 2000-h proof-of-concept test, in *Conference Proceedings, 40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, 11–14 July 2004. https://doi.org/10.2514/6.2004-4203
96. M. Crofton, Measurement of barium production in a hollow cathode, in *Conference Proceedings, 41st AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*, 11–13-July 2005, Tucson, Arizona. https://doi.org/10.2514/6.2005-3665
97. W. Ohlinger, Private communication, August 2010
98. B. Vancil, W. Ohlinger, J. Polk, Reservoir cathode for ion thrusters, Session 9.3, in *International Vacuum Electronics Conference*, Monterey, CA, 24–26 Apr 2012
99. B. Vancil, J. Lorr, V. Schmidt, W. Ohlinger, Reservoir hollow cathode for electric space propulsion. IEEE Trans. Electron Dev. **63**(10), 4113–4118 (2016)
100. C. Lai, J. Wang et al., Emission and evaporation properties of 75 at.% Re-25 at.% W mixed matrix impregnated cathode. Appl. Surf. Sci. **427**, 874–882 (2018)
101. M. Ravi, K. Bhat et al., Improved electron emission from W-Ir dispenser cathodes, in *IVEC/IVESC 2012*, Paper 199 (2012), p. 1

Chapter 3
State of the Art and Future Perspectives of Ba Scandate Dispenser Cathodes

Georg Gaertner and Yiman Wang

Abstract So far Ba scandate cathodes have shown the highest emission capability of all thermionic cathodes and are promising for future applications, e.g., in thermionic converters, in high-power high-frequency tubes, especially in vacuum terahertz devices. They are essentially based on Ba dispenser matrix cathodes modified with differently distributed additions of scandia. In this review, the historical development of different types of Ba scandate cathodes is addressed. Then the Philips activities on Laser ablation deposition (LAD) top-layer scandate cathodes with Re and Ba scandates are presented, which were mainly aimed at increasing ion bombardment resistivity and hence lifetime, but also led to record saturated pulsed emission of up to 400 A/cm^2. Finally, the activities on nanosized-Scandia Doped Dispenser cathodes (SDD cathode) carried out more recently at Beijing University of Technology (BJUT) and Beijing Vacuum Electronics Research Institute (BVERI) in China, which mainly focused on application in high current density, long life advanced vacuum electron devices, are introduced. The outstanding emission capability and lifetime tested at both diodes and electron guns are presented and also basic investigations on the emission model are discussed. A further motivation for this review is to derive guidelines for more robust highly emissive Ba scandate cathodes suitable for future tube applications.

G. Gaertner (✉)
Consultant, Aachen, Germany
e-mail: georg.f.gaertner@t-online.de

Y. Wang
Beijing, China
e-mail: wangym0709@yahoo.com

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*,
Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_3

3.1 Introduction: Historical Development of Different Types of Ba Scandate Dispenser Cathodes and Application Perspectives

As we have already seen in Chap. 1, vacuum electron sources or cathodes are the crucial components of a wide range of vacuum electron devices (VEDs) such as cathode-ray tubes (CRTs), X-ray tubes, or microwave tubes [1, 2]. Despite the decline of CRTs, advanced microwave and terahertz devices with applications ranging from civilian communication and scientific research to military and space-based technologies are being continuously improved [1–3]. One of the development challenges for advanced high power VEDs are electron sources with high current densities and long lifetime. VEDs operating at millimeter-wave and in THz range require cathodes that provide space charge limited (SCL) current densities up to 100 A/cm^2 with good emission uniformity and more than 10,000 h of a lifetime at reasonable operating temperatures [3–5].

Since in high power microwave devices also substantial currents are needed, thermionic cathodes are the only candidates for this task at present. Among the currently available thermionic cathodes, Ba dispenser cathodes, including mixed matrix and M-type cathodes, have long lifetimes >20,000 h with current densities of 5–10 A/cm^2 [2, 5]. However, at current densities of $\geq$20 A/cm^2 their practicality becomes problematic, due to the high temperatures and excessive evaporation of cathode materials and hence shorter lifetime. Alloy cathodes such as Ir_2La and Ir_5Ce can provide pulsed current densities of 50–100 A/cm^2, but at very high operating temperatures of 1600–1800 °C and a lifetime in the order of 1000 h [6, 7]. Ba scandate cathodes, which are Ba dispenser cathodes with scandium oxide additions, also referred to as scandate cathodes, have attracted great attention in recent years thanks to their low work function and capability of providing high current densities at operating temperatures lower than that of Ba dispenser cathodes [8–10]. In the following, we will call impregnated cathodes without scandium *Ba dispenser cathodes* (also known as I or impregnated cathodes) and cathodes with scandium *Ba scandate dispenser cathodes* (or Ba scandate cathodes for short). The development of emission capabilities versus time has already been shown in Chap. 1, Fig. 1.12.

Since the discovery of the Ba scandate cathode by Figner et al. [11] in 1967, the first rapid development of this kind of cathode was for high-resolution CRTs, which required cathodes with DC current densities of more than 10 A/cm^2 and an operating temperature lower than that of Ba dispenser cathodes. To meet the demands, several types of Ba scandate cathodes, including impregnated, mixed matrix, pressed top-layer, and thin-film top-layer types, had been developed in 1970s–1990s [9, 12–22]. However, none of the above cathodes has found applications in CRTs before the decline of this industry.

Yet it is instructive to study the different types of Ba scandate cathodes with their respective peculiarities in the sequence of their introduction, which can be found in more detail in [2, 9]. Starting with the pressed scandate cathode, which is based on a patent of Figner from 1967 [11], van Oostrom et al. [12] in 1979 realized $i_0 = 10$

A/cm^2 at 950 °C(Mo-brightness) operating temperature (1017 °C true temperature). Comparable emission was achieved with the impregnated scandate cathode invented by Philips (van Stratum et al. [13]) in 1977, where Sc_2O_3 was added to the impregnant. Mixed-matrix scandate cathodes were first introduced by S. Yamamoto (Hitachi) in 1983 [14], where the matrix consisted of a mixture of tungsten and Sc_20_3: J. Hasker (Philips) improved this in 1989 by using W + ScH_2 as a starting powder mixture [16], yielding about 100 A/cm^2 at the standard operating temperature. In 1986 Hasker had also pioneered the top-layer scandate cathode [15], where a 5 μm layer on top of the porous W body consisted of a mixed matrix of W + Sc_2O_3, but was still prepared via powder metallurgy. Further variants of top-layer scandate cathodes were then introduced by sputter coating the W-base with W + Sc_2O_3 (1986) [17, 18] and later on with W + $Sc_2W_3O_{12}$ (1989) by Yamamoto et al. (Hitachi) [19]. This approach was also continued by R. Longo et al. from Hughes and by Y. Wang using pulsed laser depositing at BVERI (1996) [20, 21], but the results of Hasker were not surpassed. G. Gaertner et al. (Philips) also investigated top-layer scandate cathode preparation by plasma-activated CVD of W and Sc_2O_3 from WF_6/H_2 and $Sc(C_5H_7O_2)_3/O_2$ starting compounds in alternate layers. They managed to reduce the inevitable C and F content of the top layer significantly by modifications of the DC glow discharge CVD process and by further thermal treatment and realized about 60 A/cm^2 pulsed emission, which also was no real improvement [22]. U. van Slooten and P. Duine from Philips reported improved ion bombardment resistivity of a ee-coated mixed matrix scandate cathode in 1996 [23]. A common feature of all these types of Ba scandate cathodes was their nonuniformity and the low robustness under ion bombardment. From all these variants only impregnated scandate cathodes based on a Philips license are still produced by Spectra-Mat and others and were used in some microwave tubes [13, 24].

In the following subchapter we will first address LAD top-layer scandate cathodes as investigated by Philips for CRT applications, which were mainly aimed at increasing ion bombardment resistivity, but also led to record saturated pulsed emission of up to 400 A/cm^2 [9]. These cathodes differ from other variants, since the top layer coating prepared by laser ablation deposition (LAD) in the form of nanoparticles consisted of Re, scandia, and (411) impregnants doped with scandia. The details given here may help to trigger new research also with reference to cheaper production methods of this type.

In the next part the activities on a new type of Ba scandate cathode, the nanosized-Scandia Doped Dispenser cathodes (in short SDD cathode) developed about two decades ago first by Beijing University of Technology (BJUT) and Beijing Vacuum Electronics Research Institute (BVERI) and later continuously carried out at BJUT in China, which mainly focused on application in high current density, long life advanced vacuum electron devices, are presented [10]. They are now the current workhorse for new device applications. More recently, in the US a new initiative on Ba scandate cathode research has been started [25].

3.2 LAD Top-Layer Ba Scandate Dispenser Cathodes with Re on W Base Matrix (G. Gaertner)

3.2.1 Motivation

The main aims for developing top-layer LAD cathodes with Re and Ba scandate at Philips Research Aachen (PRA) were to increase ion bombardment resistivity and emission uniformity with operating time and hence lifetime. In order to also overcome disadvantages of I cathodes such as high operating temperature (high Ba production!) and limited current density, Philips Research realized a top-layer scandate cathode by laser ablation deposition (LAD), which reached unprecedented pulsed emission (saturated) current density of up to 400 A/cm^2 at the same operating (true) temperature as for I cathodes of 1030 °C [9]. Thus, a reduction of the LAD scandate cathode operating temperature and also new high-end applications become feasible. The top-layer scandate cathodes were prepared by Excimer LAD of ultrafine particles of W or Re and of Sc_2O_3 or Ba, Sc containing materials from respective targets on 411-impregnated W–I cathode bases, usually already mounted in Philips 0.65 W cathode units. The top layer of typically 100–500 nm thickness had a very fine-grained structure after activation, which is favorable for fast resupply of Ba and Sc to the surface. The preparation conditions will be presented in more detail in the next paragraph. The robustness under ion bombardment [28] and also the emission uniformity of this cathode was strongly improved over previous scandate cathode types. Methods for accelerated ion bombardment are presented and the superposition model is discussed. Further topics are emission properties such as lifetime, poisoning sensitivity, emission dependence on temperature [9], and roll-off in diode and gun configuration. Cathode surface investigations by SEM/EDX and SAM and elemental depth profiles have been carried out and are correlated to the emission properties.

3.2.2 Experimental Conditions

3.2.2.1 Cathode Preparation by LAD

A common idea of nearly all concepts and structures of scandate cathodes is to provide Sc sources near the cathode or on the cathode surface, since the Sc diffusion length is much smaller than the diffusion length for Ba, the other component needed for a Ba scandate cathode. Several groups at Hitachi [17, 19] and at Hughes [20] had studied the preparation of top-layer scandate cathodes by sputter-deposition of Scandia on impregnated cathode bases. The cathodes obtained showed saturated emission current densities of 100 A/cm^2 at 1030 °C true temperature, but only limited ion bombardment resistivity and were not really an improvement over the other types discussed in the historical section. This was partly due to the fact that the composition was not really suited. At Philips Research, therefore, a more controlled approach was

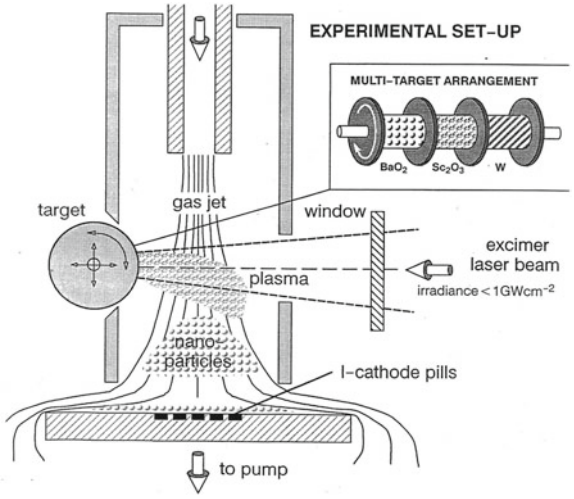


Fig. 3.1 Schematic view of the Excimer—LAD setup (first stage, static substrate); see [26]. Copyright Royal Philips N.V

used by laser ablation deposition (LAD) of thin layers of ultrafine particles and also the composition of these layers was varied [29]. This approach insured that Sc sources were in several nm distances to the emitting surface.

The LAD of top-layer scandate cathodes was carried out with a KrF Excimer laser (type EMG 202MSC, Lambda Physik; $\lambda = 248$ nm) of about 60 W average power and maximum pulse energy of 6 J, which is well suited not only for LAD of refractory metals such as W due to electronic instead of thermal excitation, but also for oxides. This setup had been used before for preparation of nanosized phosphor particles, e.g., consisting of yttria doped with Pr [26, 27] and was modified for the new task. The Excimer laser beam was guided into a stainless steel ablation chamber (UHV flanges) through a UV quartz window and hit a rotating cylindrical target. In Fig. 3.1 also a multi-target arrangement with 3 targets in a row is seen. A plasma plume with ablated nanoparticles formed above the target and the nanoparticles (also called ultrafine particles or UFPs, in the LAD case typically of 5–10 nm size) were carried by the carrier gas to the substrates. In the initial setup, the deposition was static with low yields [26, 27]. In a later stage a scale-up of deposition rates and efficiency was carried out by the addition of a nozzle above the substrate (see Fig. 3.2) and a meander scanning movement of the substrate relative to the nozzle. Also, LAD parameters such as gas flow, gas composition, and total pressure were adjusted in order to obtain improved uniformity. Thus, a UFP deposition yield of up to 60% of the ablated material was achieved. In the case of W-LAD, a major problem was the formation of W-oxide phases in the UFPs due to oxygen contaminations, despite using an inert carrier gas. During activation of the cathode, this leads to unwanted side effects such as WO_x loss by evaporation and/or formation of $BaWO_4$ (of negligible emission!) by reaction with atomic Ba. By several measures comprising the use of high purity

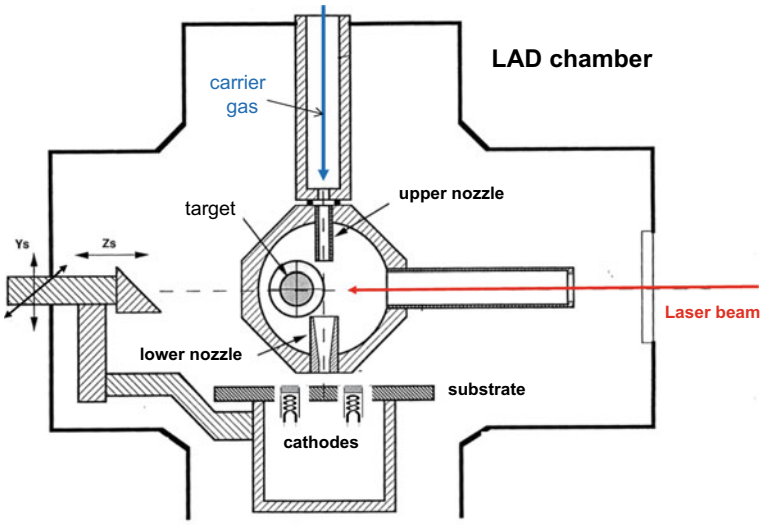


Fig. 3.2 LAD chamber (not to scale) with inner chamber and entrance nozzle for Ar carrier gas and exit (lower) nozzle (second stage of LAD setup, now dynamic; substrate with scanning movement). In holes in the substrate plate 4 I-cathode units are mounted

argon (Ar 6.0: 99.9999% purity), the addition of hydrogen (Ar/H_2 mixture), and heating of the cathode units during W deposition, the WO_x ($x \geq 2$) content could be reduced from 70% to about 20%, yet a much stronger improvement was obtained by replacing W by Re (Re-oxide content about 1%) [30].

A careful investigation of the I cathode bases (entry control) by SEM/EDX and STM [34] prior to the LAD coating revealed a varying contamination of the supplied W bases with BaO nanoparticles. These particles were remnants from the washing procedure of the pills and originated from the impregnant. After Re + Sc_2O_3 LAD it turned out, that there was a positive correlation between improved emission performance/fast ion bombardment recovery of the scandate cathodes thus obtained and the BaO nanoparticle base contamination in the range around 1.5 % Ba as determined by EDX. In conclusion G. Gaertner et al. at PRA introduced a similar artificial interface layer by LAD and investigated several Ba-oxide compounds [29–31], from targets not only such as BaO_2, $BaCO_3$, and also $Ba_3Sc_4O_9$, $Ba_2Sc_2O_5$, but also $Sc_2W_3O_{12}$, $4BaO_2{\cdot}CaAlO_4$, $8BaCO_3{\cdot}2CaCO_3{\cdot}1/2\ Sc_2O_3$. Yet the emission properties of all these layer variants were minor to the BaO-contaminated cathodes. By the way, the beneficial effect of an intermediate sputtered W + BaO layer below a sputtered W + $Sc_2W_3O_{12}$ layer on emission and its uniformity was later also shown by Sasaki et al. [35]. Therefore targets were directly prepared from the 411 impregnant melt, but with different additions of Scandia [30–33]. This approach resulted in a major breakthrough in emission capability and also in IB resistivity. It turned out that the optimum 411x target composition was about $4BaO{\cdot}CaO{\cdot}Al_2O_3{\cdot}1/3\ Sc_2O_3$ with acceptable performance also for the interval between x = 1/4 and 1/2. Yet there

Table 3.1 Mean j_{max} (1.7 kV) at operating temperature of 965 °C Mo-Br versus type of 411x target, the other layers are Re and ScOx

411x-target	Mean j_{max} [A/cm^2] and standard deviation	Number of cathodes
x = 1/4 (C)	268 ± 80	12
x = 1/3 (D)	227 ± 65	93
x = 1/2 (B)	196 ± 80	67
x = 1 (S)	208 ± 50	27
Without 411x layer	106 ± 20	16

was the tendency to get more stable targets for x ≥ ½ and better emission performance for x ≤ 1/3. As an alternative the same target compositions were prepared by sintering of pressed powder targets, but again stable targets were only obtained for x ≥ ½. Table 3.1 shows the influence of the target composition on the average emission results

Figure 3.3 shows some steps of the evolution of LAD layer structures as schematic views of the deposition sequence/layer structure of LAD top-layer scandate cathodes with alternate layers of Scandia and Re. The common porous tungsten base is impregnated with $4BaO{\cdot}CaO{\cdot}Al_2O_3$ (411). Figure 3.3b has an inverted layer structure compared to Fig. 3.3a, but here also the beneficial BaO surface contamination on W is shown schematically. A typical LAD top-layer deposition sequence (in a structural numbering #$Cd_0$122_3, in short CG1_3, one of the first cathode series with top emission) is listed below and is also schematically shown in Fig. 3.3: about 4000

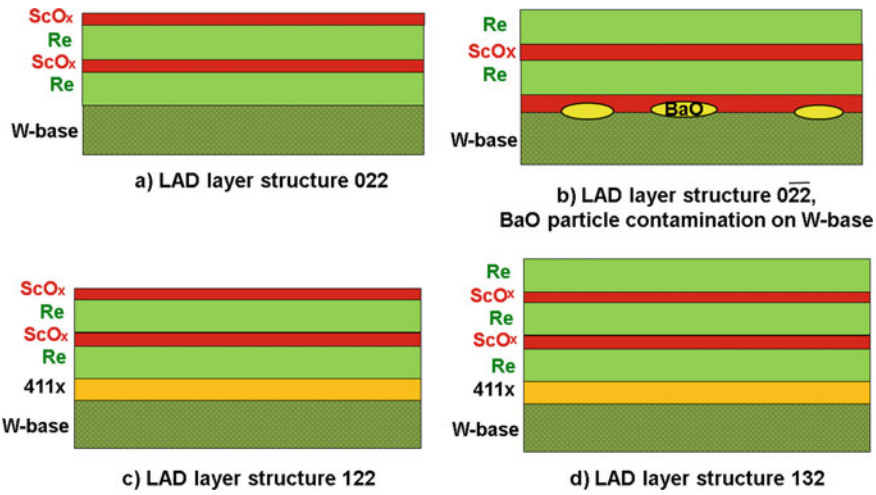


Fig. 3.3 Evolution of LAD layer structures: schematic views of the deposition sequence/layer structure of LAD top-layer scandate cathodes. The porous tungsten base is impregnated with $4BaO{\cdot}CaO{\cdot}Al_2O_3$ (411). In the following, we simply call <022>=:G0, <122>=:G1, <132>=:G2, and <132> with double thickness d_0<132>$2d_0$ = G2 + (structures see [29–33])

pulses of 411.1/4 (target C), then 6000 pulses Re, 2000 pulses Sc_2O_3, 6000 pulses Re, 2000 pulses Sc_2O_3, with the general conditions 0.4 slm Ar 6.0 carrier gas flow and a chamber pressure of 2.5–5 mbar depending on the target [29, 31]. The meander movement over the substrate was restricted to a total area of 16 mm × 16 mm. The whole process was computer controlled. The laser pulse energy was 400–460 mJ, with 4–6 Hz repetition frequency. The whole deposition then lasted about 75 min. The effective coating thickness (for 100% density) was calculated from EDX analysis referenced to the deposited mass. For CG1_3 one thus obtained about 120 nm for the 411x layer, 140 nm of Re, and 25 nm Sc_2O_3 mass equivalent layer thickness in total (assuming 100% density). Of course the layers are very porous and not 100% dense. After LAD smaller pores are covered and others are partly closed, as can be later seen in Fig. 3.21, they get open again after activation, when the structures become coarser due to grain growth. We achieved similar emission results with "132" top-layer structures (G2), where a top Re sub-layer was added, and with inverted structures, where the sequence of Sc_2O_3 and Re was inverted. The idea behind doubling the Re and Sc_2O_3 layers was to create a reservoir for ion bombardment. A doubling of the 411x layer only makes sense, if additionally a neighboring W layer is added, which implies an extension to a fourfold target arrangement.

Further improvements especially in ion bombardment resistivity were achieved by the "132" structure **and** by the so-called $2d_0$ layers, where the thickness of every sub-layer was increased by doubling the number of laser ablation pulses (e.g., #D2$d_0$132_1 = :DG2+). With the first cathode DG2+_1.2 of this series 27% of the specified ILD (ion lifetime dose) was reached!

We can see from Table 3.1, where cathodes have been classified with reference to the 411x target, disregarding their layer structure as in Fig. 3.3, that the lowest scandia content in 411x gave the best results. Yet we could not lower x further below 1/4, because the targets became mechanically unstable. The rather large scatter in the overall emission results, as judged from the j_{max}(1.7 kV) values at operating temperature shown in Table 3.1, is partly artificial, since not only variations of layer sequence and sub-layer thickness, but also changes in activation are contained in the respective groups. The sintered S targets showed less scatter, but could only be prepared for x > 0.5. For comparison in the table also results without 411x layer, only with LAD top layers of Re and ScO_x are given (G0 type), illustrating the strong improvement by introduction of this "magic" interlayer. In this reference case an average j_{max} of 106 A/cm^2 was obtained, where the saturated emission density $j_{10\%}$ is only about 40 A/cm^2. But also when all conditions were kept constant, there was still some scattering of results. Possible reasons are the target composition and the target properties (such as stability and porosity) and also changes of the target surfaces during subsequent ablations can contribute, since the same surface parts are hit several times by the laser beam in a spiraling movement, influencing the size distribution of nanoparticles. But also variations in the rest gas composition dependent on pumping down conditions can play a role or drift in the laser fluence due to internal window contamination or variation of the cathode base properties and some other not directly controlled conditions. Best conditions were achieved when the LAD setup was run continuously with 2–3 deposition runs per week and intermittent pumping-down

cycles. Technically the relevant parameters can be controlled better by improvements in the setup.

The industrialization prospects for top-layer scandate cathodes by using Excimer LAD are good in view of the deposition yield in the order of 50%. Excimer lasers are commercially available and—equipped with commercial chambers—are used for low-pressure LAD of thin compact layers. A change to a pressure range of several mbar as used here leads to the formation of UFP layers instead of compact layers. In the experiments 10–20 years ago at Philips Research Aachen, a KrF Excimer laser type EMG 202MSC of Lambda Physik with about 60 W average power was used. Yet only about 3 W were needed for the above deposition conditions for a top layer of about 300 nm thickness in total. Since the ablation rates scale with power, nowadays commercial Excimer lasers with 600 W average power and a maximum pulse repetition frequency of 600 Hz are available. Of course price predictions of LAD scandate I cathodes will strongly depend on the number of cathodes to be produced. As an alternative, the same layer structures as above could also be prepared by sputtering, since the subsequent thermal activation treatment will anyway change and roughen the surface structures as prepared.

3.2.2.2 Thermionic Emission Measurements

In the initial static stage of the LAD experiments a substrate plate with 32 circular holes was used, where impregnated cathode pills with 1 mm diameter D_c were sitting in the holes, see Fig. 3.1. This implied that these pills after LAD had to be mounted in 0.65 W Philips cathode units with heaters as shown in Fig. 3.4. In order to avoid possible damage to the LAD layer during mounting and processing, in the second stage of LAD experiments, four complete 0.65 W Philips I-cathode units were directly mounted (of course without Os/Ru layer, only W–I cathodes) under respective holes

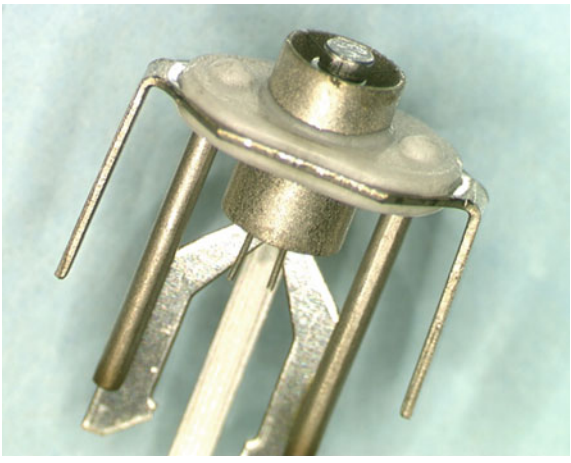


Fig. 3.4 Philips 0.65 W I-cathode unit; cathode pill diameter 1 mm; from [2], G. Gaertner, H. Koops, chapter 10 of "Vacuum Electronics, Components and Devices", Springer 2008, p. 440

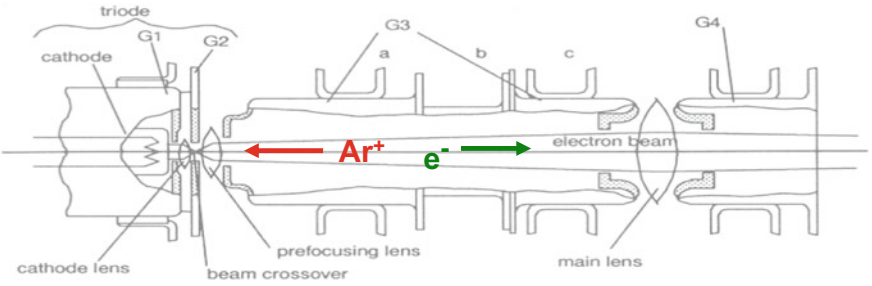


Fig. 3.5 Schematic longitudinal section of an electron gun (one electron beam). The electrons ionize the rest gas and ions (mainly Ar, N_2) are accelerated back to the cathode surface through the grid apertures: locally peaked ion beam profile! (see X. Zhang, G. Gaertner, presentation at IVESC 2002 and [44]; copyright Royal Philips N.V)

in a new substrate plate, with an additional central circular Mo reference substrate of 3 mm diameter used for layer characterization.

Electron emission measurements were carried out in either closed-spaced diode or electron gun configuration in a UHV (ultra-high vacuum) chamber, as can be seen in Fig. 3.5. In the close-spaced diode configuration the typical cathode to anode distance d_{CA} was 0.2–0.3 mm. Close-spaced diode implies that $d_{CA} \leq ¼\ D_c$ (see Hasker [38]), where D_c is the diameter of the cathode disk.

The cathode pill was mounted in a Mo cap, the side of which could be observed through a glass window for pyrometric determination of cathode temperature. The temperature could be adjusted by control of the heater voltage of 0.65 W unit. A typical setting for an operating temperature of 965 °C(Mo-brightness) or 1030 °C true temperature is a heater voltage of 6.4 V and a heater current of 103 mA. Despite heater calibration curves, the Mo-brightness temperature was determined in every measurement and corrected for window absorption.

The massive anode typically consisted of high melting Ta, in order to stand continuous heating by the electron current. The measurement of a current-voltage characteristic was done by starting with DC measurements from 10 to 100 V, then continuing with pulsed measurements in order to avoid excessive heating of the anode at higher voltages. Pulses were either delivered by a sawtooth pulser of up to 1.75 kV voltage and up to 3 A emission current, the pulse length being varied from 5 μs to 15, 50 or 100 μs, with a repetition frequency of 50 Hz, or by a rectangular pulser of up to 2 kV pulse height (pulse length 10 μs, repetition frequency 50 Hz, stepping up the voltage). In the original version (later on in parallel) of the sawtooth pulser the pulsed current–voltage characteristic could be observed as a stationary triggered sweep on an oscilloscope. During the whole project all emission measurements were computer controlled, the electronics being assembled in test racks. This implied an averaging of the pulsed sawtooth measurements over a chosen time interval. During the 25-year duration of several cathode projects, Philips Research Aachen (PRA) was equipped with up to 10 UHV chambers (32 diode test positions), including two chambers with Ar ion guns for ion bombardment investigations (see Fig. 3.6), additional dummy

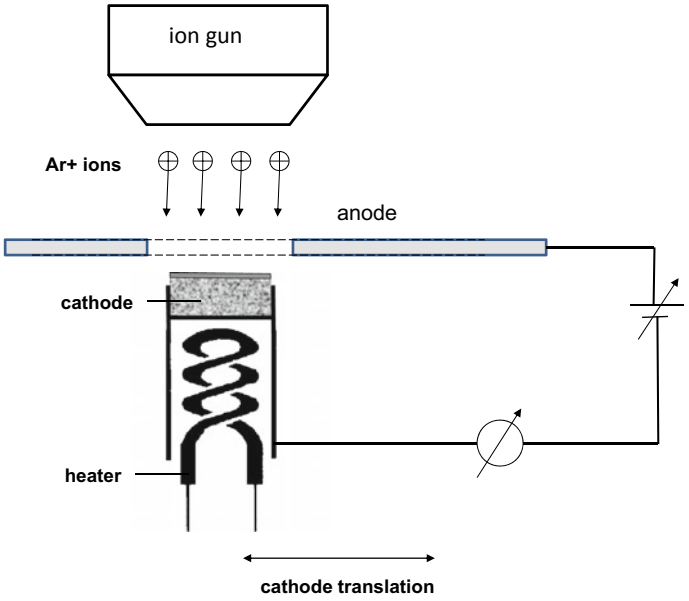


Fig. 3.6 Schematic diagram of the diode configuration with the differentially pumped Ar ion gun (Perkin-Elmer Φ 04): $d_{CA} = 0.3$ mm; the anode aperture has 2 mm diameter; the cathode can be translated to a planar diode test position (compare [44])

glass tubes for gun testing, and 4 emission test racks for measuring, e.g., current-voltage characteristics at a given temperature (status at PRA visit of IVESC 1996 participants). Cooling or heating effects under a given DC load were in most cases compensated via the heater voltage of the cathode.

Since scandate cathodes show anomalous behavior in the saturation range with only a small gradual deviation from the space charge limit (SCL) especially at higher temperatures, from the current voltage *I(U)* characteristics the 10% deviation point from SCL $I_{10\%}$ is used [9] as a good approximation of saturated emission current I_{sat} (see Manenschijn et al. [39]). For I cathodes $I_{10\%}$ is about 10% larger than I_{sat} obtained by fitting [39]. It has to be noted that the slope in a $\lg I / \lg U$ plot in the SCL range is not 1.5, but according to Child-Langmuir theory (see formula (10.9), p. 433 in [2] and the calculation of diode characteristics by Hasker [40] and Scott [37]) varying between 1.35 and 1.45. The higher values of x are obtained for current densities in the range around 100 A/cm^2 [39]. For determination of $I_{10\%,}$ it is then required that the slope x in the SCL range is above 1.4. The current density limit $j_{10\%}$ is also a reasonable measure of electron emission capability for the typical applications in the space charge limited range as in CRTs [36, 39]. Interestingly, the saturation region of scandate cathodes can also be described by a power law with slopes lower than the SCL slope. This lends support to the superposition theory of scandate patch fields,

where their contribution is reduced by prolonged sputtering as in [28]. In this context one should keep in mind that also for standard Ba dispenser cathodes a theoretical description for the complete *I(U)* characteristic is missing. It only exists for the space charge limited range and at much higher voltages for the Schottky range. In the intermediate range in between a theoretical description in closed form is missing, apart from the approximation by Hasker [40] and the numerical solution by Scott [37].

In the electron gun configuration usually 3 cathodes (for red, green, and blue) were mounted in a Philips 45AX CRT gun, as can be seen in Fig. 3.5, where the distance to grid 1 typically was 80 μm. In the diode mode of the gun, only the emission to grid 1 outside the rectangular hole of 0.5×0.7 mm^2 size was measured. This was compared to roll-off measurements in the gun mode. The gun could either be mounted in a vacuum chamber or in a dummy tube (just the neck, without phosphor screen). The path of one electron beam in gun operation is also shown in Fig. 3.5, with a beam crossover in grid 1 region, which also gives a peaked ion beam hitting back onto the cathode surface. A potential distribution for such a gun was given by T. Spanjer et al. in Fig. 8, p. 353 of Philips Technical Review 44 [42].

Since a theoretical description of the *I(U)* characteristics of Ba scandate cathodes in the saturation range is still missing and hence zero field emission cannot be derived from a curve fit, the group at BJUT of Yiman Wang uses the deviation point I_{dev}, where deviation from SCL just starts [43]. These values are lower than the $I_{10\%}$ values. In a further method, the intersection point of the straight line fits (power law slopes) in SCL range and saturation range is determined in a double logarithmic plot of the characteristic. The projected point I_{is} on the characteristic is a good approximation of zero field emission, which is more in line with the $I_{10\%}$ value. Yet this method cannot be applied, if the voltage range is not reaching far enough into the "saturation" region. In this case, a clear recipe is needed for marking the transition from SCL to saturation. This problem does not occur for *I(U)* characteristics at lower temperatures, where there are only small deviations between $I_{10\%}$ and I_{is}, as we will see later during the evaluation of temperature-dependent *I(U)* characteristics.

It should be noted that the type of equipment used for emission measurements was also used at Philips Research Eindhoven and by Philips Display Components. The 45 AX guns in the tests were applied in several commercial Philips CRTs and their design was based on electron optical calculations (see [41, 42] and other articles in Philips Research Reports or Philips Technical Review). With this equipment all types of thermionic cathodes, including thoriated tungsten, oxide cathodes, and various types of impregnated and scandate cathodes, have been tested, results have been published and are consistent with the literature. The message is that the record emission is not due to peculiarities of the test equipment, as one might speculate, since quite regularly top emission of the same cathodes was ruined by accelerated ion bombardment or by deterioration during operation (see [28, 47]) and it took some years to reach the state of the art in 1994 and then improve further.

3.2.3 Summary of Emission Results

3.2.3.1 Top-Layer Scandate Cathode Activation

It is well known that impregnated dispenser cathodes need an activation process in vacuum, in order to start the Ba supply to the surface, where Ba is generated via a reaction of the impregnant with the tungsten of the pore walls. A typical activation scheme of an I cathode consists of some preheating steps and activation by heating at 1100 $°C_{Mo\text{-}Br}$ for half an hour. It turned out, that this scheme is not really suited for LAD top-layer scandate cathodes. Of course also some preheating steps are needed for degassing of the deposited layer. A typical preactivation sequence used was 5 min at 850 $°C_{Mo\text{-}Br}$, then followed by 5 min at 965 $°C_{Mo\text{-}Br,}$ under the condition that the total pressure in the chamber is kept below 10^{-7} mbar. Yet the high-temperature activation needs to happen at 40–60 °C lower temperature than for I cathodes, otherwise lifetime and survived ion dose are strongly reduced. The optimum activation conditions are obtained under the action of an electric field (typically 75 V at $d_{CA} = 0.25$ mm) at a temperature of about 1060 $°C_{Mo\text{-}Br.}$ with a duration of 60–120 min. Figure 3.7 shows the activation curves of cathode CG1_3.2 at an average temperature of about 1060 °C(Mo-Br.) with a constant electric field of 75 V/0.25 mm. Shown is current in log scale, either DC or pulsed, versus time in log scale. The dotted line is an estimation of activation continuing in the pulsed range; see [45]. During operation

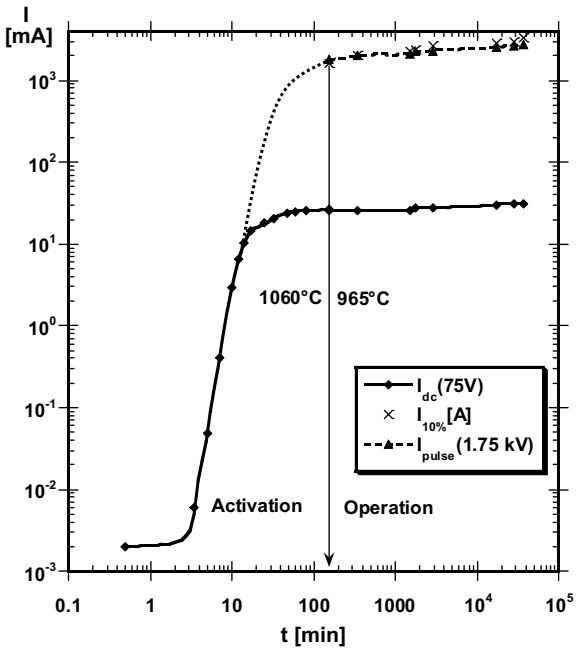


Fig. 3.7 Activation of cathode CG1_3.2 at an average temperature of about 1060 $°C_{Mo\text{-}Br.}$ with constant electric field of 75 V/0.25 mm. Shown is current in log scale, either DC or pulsed versus time in log scale; data based on [45]

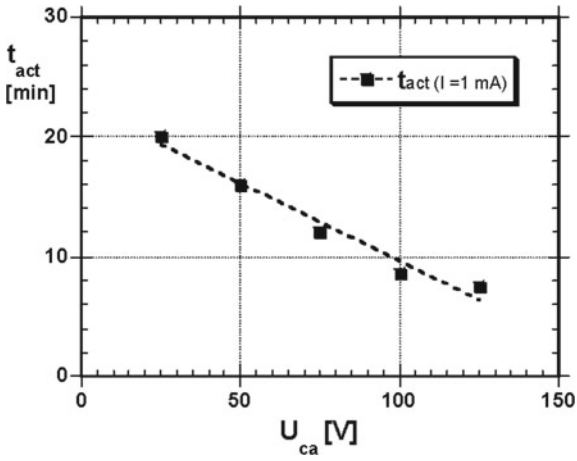


Fig. 3.8 Activation time (delay) to DC current level of 1 mA at a temperature of about 1060 °C for LAD scandate cathodes (BG1) as a function of applied voltage ($d_{CA} = 0.25$ mm); data based on [45]

of the cathode at 965 $°C_{Mo\text{-}Br.}$ following the activation interval in Fig. 3.7, one can still observe a slow further increase in the emission current.

The influence of the electric field on activation of LAD scandate cathodes of type BG1 at an activation temperature of about 1060 °C is shown in Fig. 3.8. Here we see that the lower the field strength the longer the activation time needed, but t_{act} (1 mA) has the character of a delay time, which is also dependent on the thickness of the LAD layer. It should be noted here, that for W–I, Re–I, or mixed matrix scandate cathodes, the initial delay is less than 2 min. Once a level around 0.1–1 mA has been reached, the further increase is at about the same rate as in Fig. 3.7. The influence of the electric field, especially in the higher voltage range of the *I(U)* characteristic, is consistent with the higher mobility of Sc shown in Auger experiments by J. Crombeen, where the application of a diode voltage of about 200 V would strongly increase the Sc/W signal ratio [16, 45]; he has also shown a strong influence of the oxygen concentration on activation. The advantageous action of an electric field is also known from the activation of oxide cathodes (see Chap. 4).

Of course, Ba supply during activation and on life can also be monitored by emission measurements. Here emission activation energy can be determined from the temperature dependence of activation times (e.g., time to reach half of the maximum or plateau value). Figure 3.9 shows results obtained for LAD Ba scandate cathodes with Re and uncoated impregnated (W–I) cathodes. It has to be noted that in the case of scandate cathodes two activation times can be observed, namely a fast activation—typical for Ba supply—to space charge limited emission in the low voltage range (DC-emission; Re–I characteristic!), and a slower activation to the final scandate *I(U)* characteristics in the high voltage range (pulsed emission) belonging to Ba-scandate complex formation ('Sc'). The Ba activation energy E_{act} (Ba) $= 3.12 \pm 0.1$ eV is the same for both types of Ba dispenser cathodes. The activation times to the scandate characteristic are a factor 6 prolonged, but E_{act} ('Sc') $= 3.35 \pm 0.2$ eV is

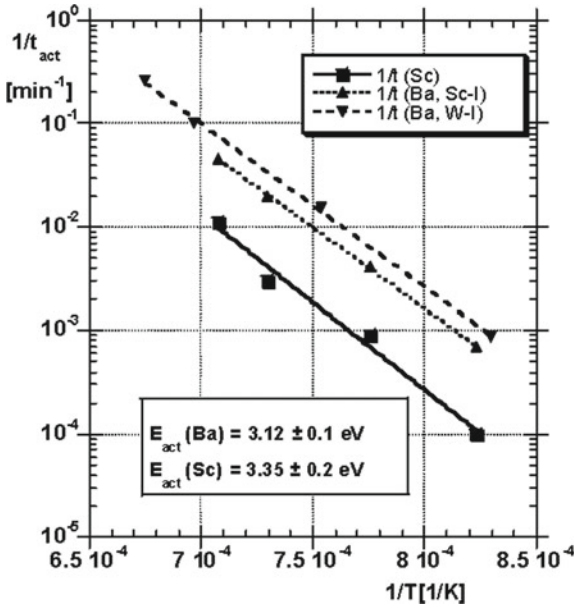


Fig. 3.9 Inverse activation times as a function of inverse true temperature for LAD Ba scandate cathodes with Re (series BG1, here denominated as Sc-I; with constant electric field), and B-type Ba dispenser cathodes (W–I); from [46] and [47], G. Gaertner et al., "Supply and Loss Mechanisms of Ba Dispenser Cathodes", Appl. Surf. Sci. 146, 22–30 (1999), with permission from Elsevier

(within the limits of accuracy) still consistent with Ba generation being the limiting factor [46, 47].

The activation times needed are dependent not only on temperature and electric field, but also on the layer thickness, the degree of oxidation of W or Re and other impurity contents, and may be prolonged by 1–2 orders of magnitude depending on the conditions, but we will not go into further details here.

3.2.3.2 Thermionic Emission Properties

The motivation for the investigation of scandate cathodes was to overcome the disadvantages of the I cathode such as high operating temperature (high Ba production!) and limited current density. Yet one of the common features of the development of different types of scandate cathodes was that they did not show a clear transition to saturation and a continuing current increase with voltage in this region instead. Since an accepted theoretical description is missing, more phenomenological practical approaches have been used, as we have already discussed in Sect. 3.2.2, namely the different transition criteria $I_{10\%}$, I_{is} and I_{dev}, where I_{dev} is marking the lower boarder of the transition range. The differences between these criteria get smaller at lower temperatures.

In Fig. 3.10 the current density versus voltage characteristics of a LAD top-layer scandate cathode type DG2, an Os/Ru–I cathode, and a W–I cathode at 965 °C Mo-brightness temperature (1030 °C_{true}), as determined in the diode mode in an electron gun configuration (45AX), are compared. The onset of saturation is given as a 10%

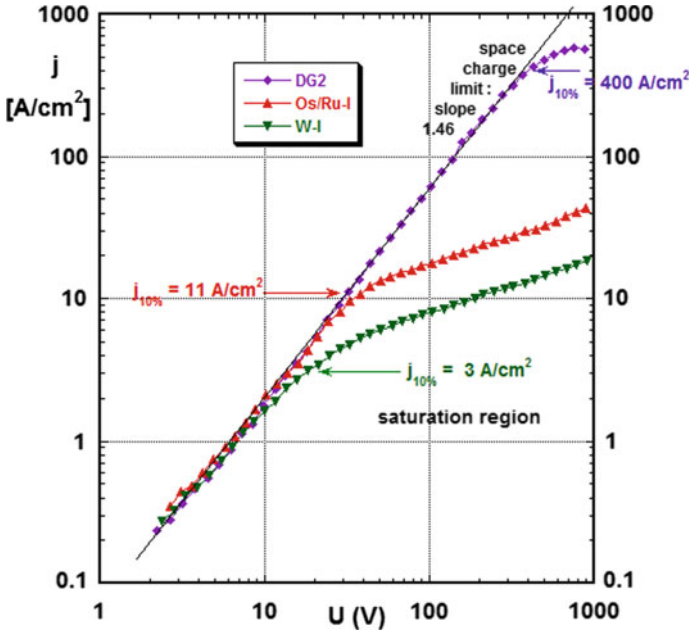


Fig. 3.10 Current density versus voltage characteristics of a LAD top-layer scandate cathode type DG2, an Os/Ru–I or M cathode and a W–I cathode at 965 °C Mo-brightness temperature, as determined in the diode mode in an electron gun configuration (d_{ca} ~ 80 μm). The onset of saturation is given as 10% deviation from the space charge limitation (slope 1.46 for DG2). For the Os/Ru–I cathode, the SCL slope is about 1.35 and the saturation slope 0.4. The graph is based on a re-evaluation of data presented in [57, 59]

deviation from the space charge limitation (slope 1.46 with second-order corrections) [57, 58]. For the Os/Ru–I cathode, the SCL slope is about 1.35 (compare [39]) and the saturation slope is 0.4. For the W–I cathode, the SCL slope is about 1.30 and the saturation slope is also 0.4. The gun data are consistent with diode measurements and the literature with reference to both types of I cathodes. In this plot also, practically no significant differences between the $j_{10\%}$ values and j_{dev} (of Wang et al. [43]) can be seen. We will discuss the correlations between gun and diode measurements later.

In Fig. 3.11 the emission current densities j of a BG1-type cathode are depicted versus diode voltage U in a double logarithmic plot at different temperatures [9]. These characteristics have been measured in a close-spaced diode configuration and are consistent with characteristics in the diode mode of a 45AX gun, where higher electric field strengths can be reached in pulsed measurements. There (Fig. 3.10) the emitting cathode area A_{cath} for determination of current density from current is A_{cath} = ($\pi \times 0.5^2$–0.35) mm^2 compared to $A_{cath} = \pi \times 0.5^2$ mm^2 = 0.7854 mm^2 in a planar diode. The power law slopes in saturation are shown at the right side of the curves. The respective $j_{10\%}$ values for the temperatures from 600 °C to 915 °C are

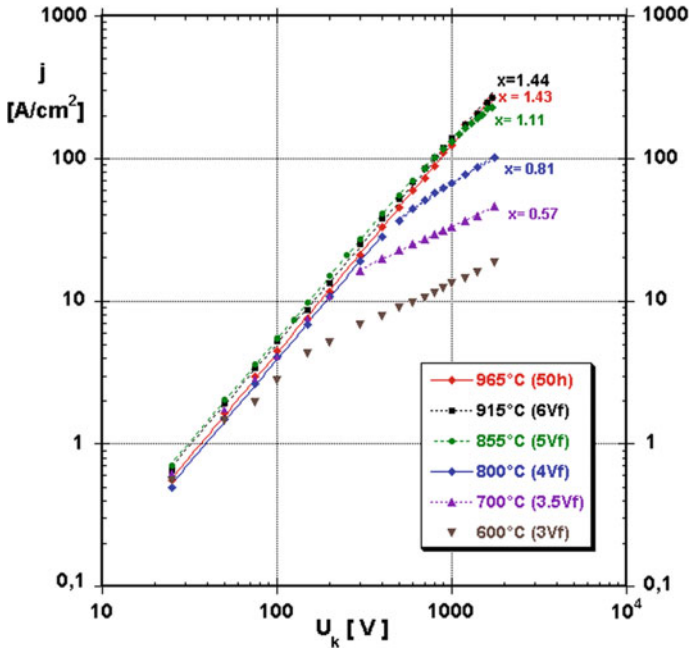


Fig. 3.11 lg j/lg U characteristics of BG1_2 cathode for different Mo-brightness temperatures in diode configuration (d_{ca} ~ 200 μm). The power law slopes x_s in the double logarithmic plot in saturation are shown at the right side of the curves. The graph is based on a re-evaluation of data presented in [9]

Table 3.2 Saturated emission of cathode type BG1_2 versus Mo-brightness temperature

T [°C $_{\text{Mo-Br.}}$]	600 °C	700 °C	800 °C	855 °C	915 °C
Heater voltage U_f [V]	3.0	3.5	4.0	5.0	6.0
$j_{10\%}$ [A/cm^2]	1.8	15	38	178	293
Slope x_s	0.57	0.57	0.81	1.11	

given in Table 3.2. Saturation cannot be determined at an operating temperature of 965 °C$_{\text{Mo-Br.}}$, since the maximum pulsed voltage is not high enough to see a deviation from space charge limit.

In a plot of saturation current density versus true temperature as in Fig. 3.12a, one can obtain a good overview of the optimization of LAD scandate cathodes. Here $j_{10\%}$ has been chosen as a good approximation for zero field emission or "saturation" current density. A general feature of all Ba scandate cathodes prepared by LAD is the work function of about 1.15 eV, as one can also derive from the Richardson plots of Fig. 3.13. The same holds also for SDD cathodes of Y. Wang and for sputtered top-layer scandate cathodes of S. Yamamoto. Only the Richardson constant A_R is

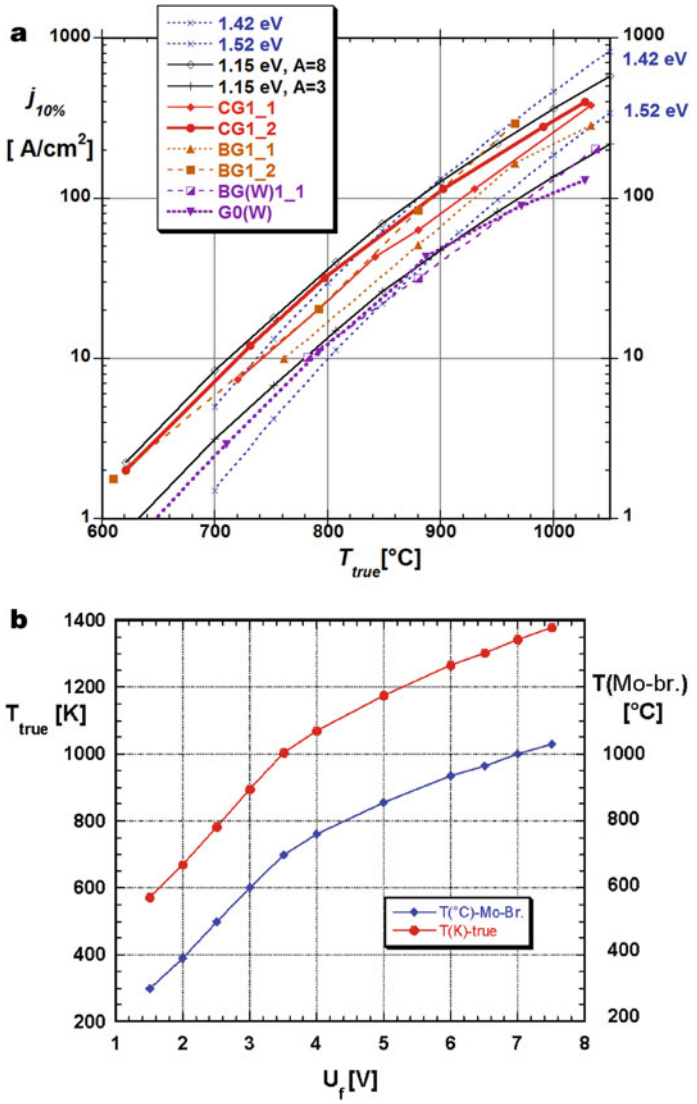


Fig. 3.12 **a** Saturation current density versus true temperature. A general feature of all Ba scandate cathodes prepared by LAD is the work function of about 1.15 eV, whereas the Richardson constant A_R is varying between 3 and 8 A/(cm^{2}K^2). For comparison, also the saturated emission density curves for work functions of 1.52 eV and 1.42 eV with the Richardson constant being the thermionic constant are shown. Results of [47, 59] are contained in this graph. **b** Temperature versus heater voltage for Philips I-cathode unit. The blue curve is the Mo-brightness temperature in °C and the red curve the true temperature in K. Measured Mo-brightness temperatures of individual units may deviate in the region of the working point 6.3 V_f by about $\pm$15 °C. During continuous DC load also, emission cooling has to be taken into account. These calibration curves are basis for the results given in [47, 59]

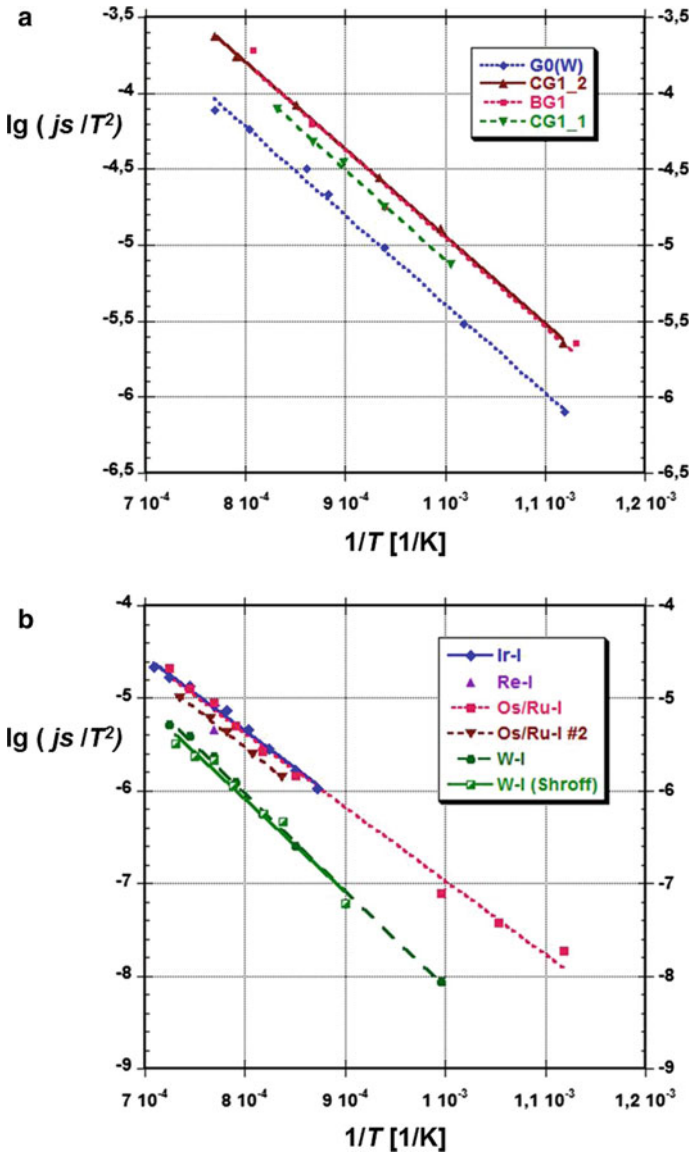


Fig. 3.13 **a** Richardson plot of LAD top-layer scandate cathodes G0 (without 411x layer), CG1_1 (x = 1/4), CG1_2 (x = 1/4), and BG1 (x = 1/2); compare [9, 58]. **b** Richardson plots of different types of Ba dispenser cathodes; see Table 3.3 and [39, 59, 61, 62]

varying between 3 and 8 A/(cm^{2}K^2), which is a measure of increasing coverage of the surface area (see Gilmour, p. 44 in [48]) and is also consistent with increasing ion bombardment resistivity when Ar = 8 A/(cm^{2}K^2) is approached. For comparison also the saturated emission density curves for work functions of 1.52 eV and 1.42 eV with the Richardson constant being the thermionic constant are shown in Fig. 3.12a,

but they do not deliver a good fit over a larger temperature interval. One can also see that only tungsten-based LAD cathodes without Re, namely BG(W)1_1 and G0(W), are located around $A_R = 3$ A/(cm^{2}K^2) at the lower boarder. In this context a Ba scandate cathode on W base has also been obtained by metal Sc-LAD on a W–I base with js in the same range (not shown here, literally Sc/W–I), where the W base is usually decorated with BaO nanoparticles as already mentioned [34]. Activation then leads to oxidation and scandate formation on a W–I base.

Very important for establishing a reliable plot as in Fig. 3.12a and also as a basis for the Richardson plots of Fig. 3.13 a + b is the determination of true temperature, which is not trivial. Such a calibration curve is given in Fig. 3.12b as a function of the heater voltage of the I-cathode unit. Of course the measured Mo-brightness temperatures of individual units may deviate, which can be about $\pm$ 15 °C in the region around the working point 6.3 V_f. During continuous DC load also electron emission cooling has to be taken into account. Thus at certain intervals during continuous runs or before/after *I(U)* characteristic measurements, the cathode temperature needs to be determined usually by optical pyrometry (visible or infrared). In regular intervals ($\leq$1 year) the pyrometer needs to be calibrated with either a tungsten ribbon lamp or a black body radiator. If the brightness temperature of the Mo cap is measured, not only the Mo emissivity as a function of temperature and observation wavelength has to be taken into account, but also the transmission of the viewing window. In case the observation spot is located on the part of the uncovered tungsten, W emissivity determines the correction.

We know from Chaps. 1 and 2 that the dependence of saturated electron emission current density on temperature is given by the Richardson–Dushman equation [2, 48], which is usually written in the form

$$j_s = A_R T^2 \exp\left(-e\phi_R / kT\right) \tag{3.1}$$

where A_R is the Richardson constant, which for real cathodes in most cases is smaller or much smaller than the thermionic constant $A_{th} = 120.4$ Acm^{-2} K^{-2} for the ideal case. $e\Phi_R$ [eV] is the Richardson work function in the form written by S. Dushman, Herring and Nichols, Nottingham, Scott [37], Hasker [40], and others [2], where within this book $e\Phi = \phi$ is used. j_s is the saturated emission current density in Acm^{-2}. From the measured dependence of j_s on true temperature T [K], one can determine work function and Richardson constant from a so-called Richardson plot [48] of log (j_s/T^2) versus *1/T*, where the physical quantities have been divided by their respective units

$$\lg\left(j_s / T^2\right) = \lg A_R - e\phi_R / (kT \ln 10) \tag{3.2}$$

It should be noted, that the Richardson constant A_R can also be expressed as a temperature dependence of work function $e\Phi = e\Phi_R + \alpha T$, with the coefficient $\alpha = k \ln (A_{th}/A_R)$. This of course only makes sense for uniform emitter surfaces.

Figure 3.13 a shows Richardson plots of LAD top-layer scandate cathodes of type G0 (without 411x layer, $\phi = 1.16$ eV, $A_{Rrel} = 2.8$), CG1_1 (x = 1/4, $\phi = 1.17$ eV, $A_{Rrel} = 6.3$), CG1_2 (x = 1/4, $\phi = 1.14$ eV, $A_{Rrel} = 6.2$), and BG1 (x = 1/2, $\phi =$ 1.15 eV, $A_{Rrel} = 7.0$). The Richardson work function of about 1.16 $\pm$ 0.02 eV and a Richardson constant of $A_R = 6$ to 7 $Acm^{-2}K^{-2}$ gives a better fit to the experimental 'Sc'-I data than an effective work function of 1.4–1.45 eV ($A_R = A_{th} = 120.4$ $Acm^{-2}K^{-2}$). This illustrates the superior emission of the LAD scandate cathodes down to temperatures of 500 °C with about constant work function.

Here in Fig. 3.13a again as in Fig. 3.12a, the only tungsten-based LAD scandate cathode without Re, namely G0(W), is characterized by A_R of about 3 $A/(cm^2K^2)$, hence in the lower range.

In order to check the validity of this interpretation of the Richardson constants of scandate-type cathodes, it should be compared with the respective results for impregnated Ba dispenser cathodes, namely the W–I or Re–I cathode bases, which in general exhibit more uniform emission than scandate cathodes. Therefore in Fig. 13b Richardson plot data of Os/Ru–I, Ir–I, and W–I cathodes from Philips and a W–I cathode from Thomson have been evaluated. The results are listed in Table 3.3.

Unfortunately respective data for Re–I have not been published; from the comparison at the operating temperature at 965 °C(Mo-Br), the Re–I line is expected in the emission range 30–40% lower than Os/Ru–I cathodes. In conclusion much higher Richardson constants for Ba dispenser cathodes are obtained from this new evaluation compared to older references. The range of A_R for Os/Ru–I and Ir–I shows a rather good agreement. Reasons for disagreement with the older literature are mainly due to the systematic errors in the determination of zero field emission current density j_0 induced by Schottky extrapolation used at that time, especially at higher temperatures, as pointed out by Hasker [38, 40]. To some extent, it may also be due to rather

Table 3.3 Fitted results of Richardson plots for LAD scandate cathodes and Ba dispenser cathodes

Ba dispenser cathode type	Reference	Richardson work function $e\Phi_R$ [eV]	Richardson constant A_R [$Acm^{-2}K^{-2}$]
LAD scandate BG1 ('Sc'/Re–I)	Gaertner et al. [59, 61]	1.15	7.0
Os/Ru–I (M)	Gaertner et al. [59, 61]	1.58	10.5
Os/Ru–I (M), #2	Manenschijn et al. [39, 36]	1.71	22.1
Ir-I (M)	Geittner et al. [61, 62]	1.60	13.0
W–I (S)	Gaertner et al. [59, 61]	2.08	171
W–I (S)	Shroff et al. [64][a]	2.01	102.6

[a]See [64] A. Shroff et al., Applic. Surf. Sci. **8**, 36–49 (1981), p. 38, Fig. 1

small temperature intervals and to problems in determining true temperature. Especially the results of Manenschijn et al. for Os/Ru–I [39] are very reliable, since they have determined j_0 by application of the theory of Hasker for the transition range between space charge limited and Schottky emission [39, 40]. The values for the W–I cathodes should be checked again, despite that a value of the Richardson constant higher than the thermionic constant is not uncommon for certain crystal faces of pure metals and corresponds in the alternate picture to a negative temperature dependence of work function. In case of the values of Shroff, the Richardson plot exhibited a slight curvature. One has to keep in mind that the results for A_R and $e\Phi_R$ will also change with operation time.

3.2.3.3 Features of an Emission Model

One of the models to describe the current–voltage characteristics in this range is the semiconductor model of Raju and Maloney [50]. We do not doubt the usefulness of this model, but it was applied to very nonuniform scandate cathodes of minor emission performance. From simulation and from continuing ion bombardment experiments [28, 49], we found that scandate cathodes can be better described by a patch or superposition model, where the ideal scandate surface portion is in first approximation only space charge-limited, but with a lot of I cathode (in our case Re–I) patches in between. Depending on the relative fractions of the Ba scandate and Re–I part, the slope in a double logarithmic plot of I versus U is increasing from about 0.4–1.46 (space charge slope or power) with increasing scandate part. This behavior at operating temperature or below is linked to the fact that the Ba-scandate surface complex has a very low work function of about 1.15 eV compared to 2.02 eV for W–I and 1.93 eV for Re–I [51–53]. This superposition model is also backed by measurements of the local nonuniform emission distribution by Hasker [16] and by measurements of the local work function of sputtered top-layer scandate cathodes by Yamamoto with the field emission retarding potential (FERP) method, yielding 1.15 eV [54]. Also, the investigation of a model system by Zagwijn et al. [55] with Sc and Ba monolayers (ML) on W(001) surface with medium energy ion scattering (MEIS) resulted in a work function of 1.18 eV for O concentrations between 1 and 1.5 ML. It is consistent with the observation that the slope is increasing during activation, which takes longer for the Ba-scandate complex than for the I cathode parts. The same happens in the reverse direction after successive ion bombardment and during emission decline at the end of life, where the I cathode area portions now increase. Yet it has to be kept in mind, that the superposition model does not imply the simple addition of contributions, since there is also a mixed term by enlargement of the Ba scandate portion by space charge expansion within the space charge coupling length, which is also called the beamlet effect by Hasker [38]. In the first approximation, this coupling length is estimated to be related to the distance d_m of the space charge maximum in front of the cathode. For a cathode temperature of 1300 K and a current density of $j = 100$ A/cm^2, one obtains $d_\mathrm{m} = 0.6$ μm and for $j = 10$ A/cm^2, $d_\mathrm{m} = 1.6$ μm [56]. Due to this space charge leakage and the low

Schottky saturation contribution of the I cathode portions, one can observe a space charge like behavior with lower power (slope in log-log plot below the space charge limit slope of about 1.45). A complete theory of this "coupled superposition" still has to be established. Surface analysis of Sc-I cathodes after activation and during life typically shows spots with Sc and Ba enrichment with an extension below the grain sizes of about 3 μm on the average. It should be noted that the superposition model is not in contradiction with a still to be established emission model of the emitting (Ba, Sc, O) complex, eventually via a semiconductor type theory of a Ba–Sc mixed oxide compound with a Ba monolayer on top, but here in the form of patches.

3.2.3.4 Low-Temperature Thermionic Emission and Field Emission

Due to the low work function, thermionic emission could be measured down to about 200 °C, where the temperature was measured with an infrared pyrometer. Field emission (FE) was observed at room temperature and under moderate heating up to $1.5V_f$ (about 300 °C) for 'Sc'/Re–I, Re–I, and W–I cathodes. The threshold field strengths (reference level 1 nA) were 3.2 V/μm for 'Sc'/Re–I, 9 V/μm for Re–I and 11 V/μm for W–I [59]. In the range from 250 to 350 °C, thermal-assisted field emission is observed and the threshold vanishes [59–61, 63] as can be seen for the respective upper curves in Fig. 3.14. These results are discussed in more detail in

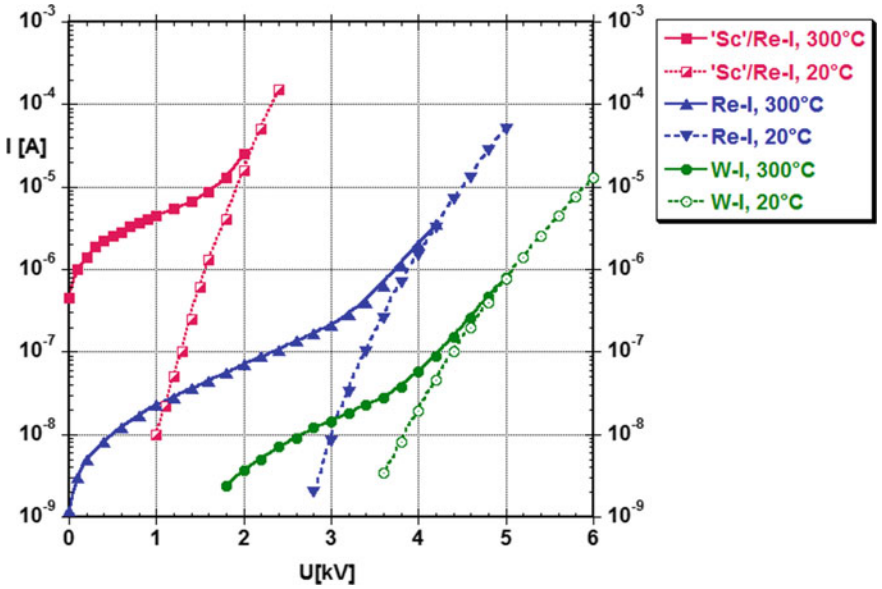


Fig. 3.14 Field emission characteristics at room temperature and low temperature *I(U)* characteristics for top-layer Re-Scandate = 'Sc'/Re–I, Re–I, and W–I cathodes; from G. Gaertner et al. [59], Copyright 2002, with permission from Elsevier

[59]. Unfortunately for stable field emission of scandate cathodes a very good UHV with chamber pressures in the order of 10^{-10} mbar is needed, including the baking of the chamber after cathode mounting. K. Jensen gives a theoretical description of the superposition of thermal and field emission in Chap. 8.

3.2.3.5 Life-Limiting Mechanisms and Experimental Results

As shown in [46, 47], the Ba evaporation rate for 'Sc'/Re–I (Scandate with Re) cathodes is lower than that of standard impregnated or I cathodes (metal matrix: W and metal surface: W or Os or Ir or Re …), at least during the initial 1000 h. Figure 2 of [47] shows that Ba evaporation from the I cathodes is higher than that of the two top-layer 'Sc'-I cathodes. An exponential fit of the data yields activation energy for Ba evaporation $E_{act} = 3.15$ eV for all 4 cathodes, which is due to the same generation mechanism. An explanation for the lower rates for 'Sc'-I cathodes could be that Ba is needed for buildup and replenishment of the Ba–Sc–O surface complex.

In conclusion, one limiting mechanism is the Ba generation and evaporation, which is mainly dependent on the design of the I cathode base. According to Aida et al. [65], the life end is above all determined by the exhaust of the Ba supply reaction of the impregnant with W (pore walls). Hence, it depends on the total amount of impregnant in the porous W matrix, i.e., for 20% porosity, and also on the thickness of the cathode pill. Aida et al. [65] have measured Ba and BaO evaporation over life (see Fig. 6 of [47]). They observed a clear correlation: end of life has been reached when the Ba evaporation (linked to Ba generation) falls below BaO evaporation. A detailed quantitative description of I cathode life-limiting mechanisms has also been given by Roquais et al. [68].

Due to the addition of Re, which is also more persistent during ion bombardment, the base emission is that of the Re–I cathode, which is about 10 A/cm^2 at T_{op} = 965 °C(Mo-Br) or about 1030 °C true operating temperature. This ensures that despite the possible loss of the Ba–Sc–O complex, the saturated emission current density or transition from space charge limited to temperature limited regime remains at this level at least for the guaranteed lifetime of the (W)-I cathode base [66, 67].

An essential mechanism of life limitation of 'Sc'-I cathodes is the loss of Sc not by evaporation, but by ion bombardment. Hence also due to the short diffusion length of Sc, a rather thick reservoir containing Sc of some 100 nm is needed below the outer surface. This is ensured by the LAD top-layer coating containing BaO and ScO_x.

In Fig. 3.15 lifetime data of LAD top-layer scandate cathode type CG1 are depicted, showing pulsed results at 1.7 kV and 200 V and also DC results for 75 V (see [9]). The time of operation was 17,328 h, with a continuous DC load of 4 A/cm^2 (75 V) applied. It has to be noted that due to electron emission cooling, the operating temperature under load is about 25 °C lower than that given above for the no-load condition. This effect will also depend on the design of the I cathode base.

A series of CG1- and BG1-type LAD scandate cathodes have been operated under UHV at operating times between 1400 to 4000 h at DC levels of 2–3 A/cm^2, with I_{dc}(100 V) remaining constant at about 5 A/cm^2. Here the general conclusion

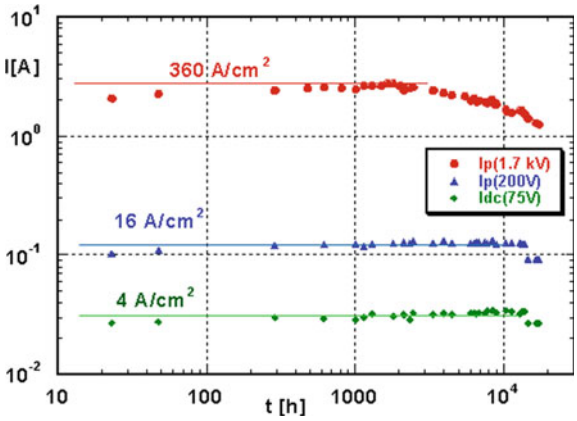


Fig. 3.15 Lifetime data of LAD top-layer scandate cathode type CG1, operated at 940 °C(Mo-Br) and 4 A/cm^2 continuous DC load in planar diode configuration. The graph is based on re-evaluated data of [9] plus additional data

for 'Sc'/Re–I cathodes is that a DC operation base level of 15 A/cm^2 should be maintained for the designed lifetime of the I cathode base, (in our case of 0.65 W Philips I-cathode units of about 20,000 h; pill diameter 1 mm, pill thickness 0.4 mm), due to the persistence of the Re surface. For operation at higher pulse test levels of, e.g., 100–200 A/cm^2, not only the surface adjacent reservoir of Sc (more specifically the amount of the Ba–Sc–O complex) will be decisive, as one can see from the ion bombardment investigations, but also a low UHV pressure level during operation is important.

3.2.3.6 Correlation Between Planar Diode and Electron Gun Emission Measurements

Since the final aim of the project was to apply the LAD scandate cathodes in CRT guns, their behavior in guns needed to be studied and correlated to planar diode results, which were of course much easier to obtain via directly mounting in UHV chambers. Moreover, the real ion bombardment happened in guns via ionization of the rest gas and had to be correlated to bombardment with Ar ions in a diode configuration [28, 44, 49], as we will discuss in the next subchapter.

The main difference between electron gun and planar diode is that in the gun for the electron beam one can only measure roll-off curves (also under-heating curves: electron current versus heater voltage) or activity curves (versus temperature) at a specific grid voltage setting for judging the emission capability of the cathode. It should be noted that there is a direct correlation between heater voltage and temperature via calibration curves (see Fig. 3.12b). In a diode, of course, it is no problem also to measure activity curves besides the *I(U)* characteristics at a set of temperatures. In the operation mode of the electron gun grid 1 usually has the same potential as the cathode (or a slight blocking voltage is applied) and the beam electrons are only extracted from the central region under the grid hole. The other grids have accelerating voltages. Yet there is the possibility to operate also the gun in a diode

mode, where grid 1 serves as the anode and at the other more distant grids higher extraction voltages are applied, so that the electrons from the central part below the hole are still extracted to the other grids. In this diode mode of the gun, the emitting area is now the cathode surface area without the area of the grid opening. Yet it is required, that the correlation is investigated just after activation at the beginning of life, before any damage in the central beam region can happen (see Fig. 3.21a, b). For well-activated cathodes from the same batch, we found good consistency between the *I(U)* characteristics in the planar diode and the diode mode in the gun and also between roll-off curves measured in the planar diode or under so-called half-drive conditions in a gun (see Fig. 3.16a, b).

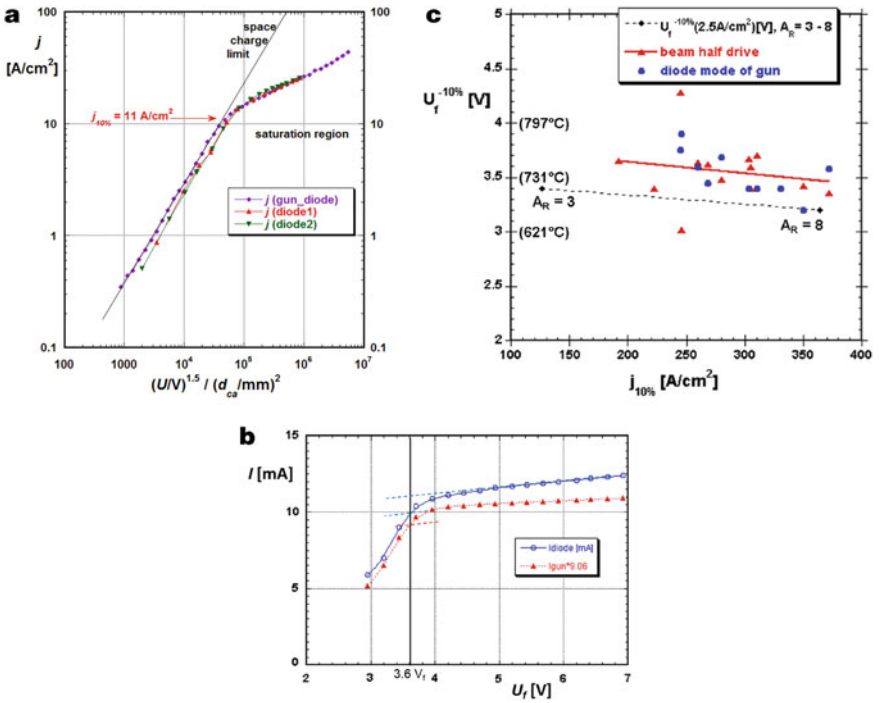


Fig. 3.16 Consistency between gun and diode measurements: **a** Electron emission current density j[A/cm^2] versus dimensionless "reduced space charge voltage coordinates" of $(U/\mathrm{V})^{1.5}/(d_{ca}/\mathrm{mm})^2$ for 3 Philips Os/Ru–I cathodes mounted in 0.65 W cathode units at an operating temperature of 965 °C$_{\text{Mo-br.}}$; based on data from [57, 59, 62]. Copyright of Fig. 3.16a: Georg Gaertner, Aachen, Germany.
b Roll-off measurements of a LAD scandate cathode type DG1 in diode mode (blue circles) and "half"-drive beam mode (red triangles) in 45AX electron gun.
c Roll-off measurements of LAD scandate cathodes type DG$_2$ in "half" drive beam mode (red triangles) and in diode mode (blue circles) in 45AX electron gun versus $j_{10\%}$ measured in the diode mode at 965 °C Mo-br. The temperatures in brackets are the true temperatures for the respective heater voltages. The theoretical values are also given for A_R varying from 3 to 8 A/(cm^{2}K^2) and a work function of 1.15 eV

The good consistency of cathode emission data obtained in a diode or in the diode mode of a gun is shown in Fig. 3.16a. A comparison of emission characteristic measurements in planar diode and in diode mode of a 45AX gun configuration was carried out for different Philips Os/Ru–I cathodes mounted in 0.65 W cathode units at the standard operating temperature of 965 $°C_{Mo\text{-}br}$. The results are shown in a double logarithmic diagram of the electron emission current density j[A/cm^2] versus dimensionless "reduced space charge voltage coordinates" of $(U/V)^{1.5}/(d_{ca}{}^2/\text{mm}^2)$. The cathode to anode distances were 70 μm in case of the diode mode of the gun, 190 μm for diode 1, and 250 μm for diode 2. The I cathodes tested were all standard Os/Ru I cathodes of Philips with 411 impregnants, mounted in 0.65 W cathode units. The agreement between the different Os/Ru I cathodes in the plot with reduced space-charge ordinate in the saturation range is remarkable. The small deviation in the space charge limited range is probably due to the fact that the lowest four data points at 25, 50, 75, and 100 V in the diode case are DC measurements of about 10–20 s duration (with some emission cooling), whereas the gun measurements and the diode saturation data are from pulsed measurements (pulse duration gun: 12 μs, pulse duration diode 5 μs). Also, a better determination of diode distance could result in still better consistency.

Since one can calculate activity curves from *I(U)* characteristics at different temperatures, one may also correlate $j_{10\%}$ at operating temperature and the knee of roll-off curves. The figure of merit is now the knee of the roll-off curve, defined as ${U_f}^{-10\%}$, i.e., the heater voltage value of the 10% deviation from the slightly inclined $I(U_f)$ line (see Fig. 3.16b). Via work function and $j_{10\%}$(T) there exists a correlation between the roll-off knee and the electron emission current density $j_{10\%}$ after IB [28, 49] (compare Fig. 3.16c). The correlation between the roll-off knee and the electron emission current density $j_{10\%}$ exists via the Richardson equation (via work function and Richardson constant), since from the half-drive condition of 1.2 mA or 2.5 A/cm^2 (a criterium used for I cathodes) the knee temperature can be determined or can be read from a plot such as in Fig. 3.12a. For example, ${U_f}^{-10\%} = 4.5$ V_f corresponds to $j_{10\%} = 40$ A/cm^2 at 6.3 V_f. Hence high pulse emission at high temperature is directly correlated with the roll-off knee at very low temperature. If ion bombardment reduces emission at high temperature in the diode, the roll-off measurement in the gun will show an increase of the knee temperature.

Figure 3.16b shows a comparison of roll-off measurements of a LAD scandate cathode type DG1 in the diode mode (blue circles) and in the "half"-drive beam mode (red triangles) in 45AX electron gun for the same ${U_f}^{-10\%} = 3.6$ V_f for both cases The ${U_f}^{-10\%}$ was determined as −10% deviation from the fitted saturation slope as a measure of the knee. The beam measurements have been carried out in the so-called half-drive mode with grid 1 at the ground, grid 2 at +1 kV, and grid 3 at +3 kV, so that the beam current is 1.2 mA at the working point of 6.3 Vf or 965 °C(Mo-br.), which corresponds to 2.5 A/cm^2 peak beam current density according to the electron optical calculations. Based on this density the equivalent current in the diode mode was calculated (I_{gun} times 9.06) and is shown as blue circles in Fig. 3.16b for a direct comparison. In the diode mode, the current density calculated from cathode area minus hole projection is 2.8 A/cm^2. The plot illustrates the good consistency

between half-drive and diode mode measurements. But it can also be used for an estimation of the maximum error, since it could mean that the relevant cathode area is about 12% larger than the geometrical projection. This would imply a diode mode correction in a gun of $j_{10\%}$ by −12%.

In Fig. 3.16c roll-off measurements of a series of LAD scandate cathodes type DG2 in "half"-drive beam mode (red triangles) and in diode mode (blue circles) in 45AX electron gun are depicted as a function of $j_{10\%}$ measured in the diode mode at 965 °C Mo-br. The temperatures in brackets are the true temperatures for the respective heater voltages. The theoretical values are also given for A_R varying from 3 to 8 A/(cm^{2}K^2) and a work function of 1.15 eV. The measurements were all carried out directly after activation ("0 h") and hence further emission increase in the following several hundred hours is not contained in this diagram.

A reliable comparison between gun and diode mode can only be carried out in the beginning, before deterioration in the central beam emitting area starts. Yet also the scatter in the beam mode may be larger, since the beam originates from a much smaller area than the diode part. In conclusion, one can see that the high emission capability of LAD top-layer scandate cathodes in guns at operating temperature is correlated with very low roll-off knee temperature.

3.2.3.7 Ion Bombardment Investigations

In cathode-ray tube applications, an important loss mechanism for Ba is ion bombardment (IB). The ions are generated by the interaction of the electron beam with the residual gas in the tube and are accelerated back to the cathode surface (see Fig. 3.5). The emerging ion beam has a strongly peaked density profile and hence sputter damage is strongly localized [28, 44, 49, 69, 70] in the central emitting area of the cathode (about 6% of the effective emitting beam area of 0.23 × 0.32 mm^2 dimensions of 45AX gun). Most electron gun experiments were done with 45AX guns in UHV chambers with a controlled gas admission valve. In this configuration, the ion life dose (ILD) is proportional to the electron current times the integral of pressure over time (average pressure times lifetime) and was specified as

$$\text{ILD (gun)} = 4.5 \times 10^{-5}\text{mA} \times \text{mbar} \times \text{h} \tag{3.3}$$

for 1.2 mA "half-drive"conditions in a Philips CRT. This corresponds to a leveling off of the initially higher pressure to nearly constant 2×10^{-9} mbar over a tube life of 20,000 h. The effect of ion bombardment in a gun is measured via the "roll-off" curve of the beam current, as we have seen above, starting at 1.2 mA (measured at grid 3) for 6.3 V_f, at given voltage settings (grid 1: 0 V; grid 2: 1 kV; grid 3: 3 kV) as a function of temperature, i.e., heater voltage U_f. The figure of merit is now the knee of the roll-off curve $U_f{}^{-10\%}$ which is correlated to the electron emission current density $j_{10\%}$ after IB [28, 49].

Accelerated ion bombardment experiments conducted in an electron gun configuration (Fig. 3.5) with controlled gas admission were then correlated with experiments in a diode configuration (Fig. 3.6) using an ion gun with a uniform ion density profile. Accelerated life testing of cathodes in the diode configuration has a number of advantageous aspects: the ions are generated externally in an ion gun and the ion current applied to the cathode is measurable, so that the number of ions hitting the surface is really known. The ion current density is constant over the cathode surface, which enables the examination of fundamental behavior and material properties of the cathodes without the influence due to the electron and ion optics in the electron gun, which in case of the gun cause a very sharply peaked ion density profile (hole burning into the emitting electron area, see Fig. 3.23a, b. Since the ion current generated by the electron beam is proportional to electron current and rest gas pressure, the effective ion dose ILD in the electron gun in units of mA*mbar*sec can be scaled to the absolute ion dose (charge) via the relation [28, 49]:

$$\text{mA} * 1\ \text{mbar} * \sec\left[\text{electron gun}\right] \cong 10^4 \mu\text{Asec [diode]} \tag{3.4}$$

This relation was mainly established via correlation of initial decay in gun and diode. In the case of the gun, the real ion dose can be determined from the specific ionization, where gas composition and electron energy need to be known, and from the path length of ionization. Most of the accelerated ion bombardment experiments were done in the diode configuration, using a moveable anode with a circular hole and a differentially pumped Ar ion gun (PerkinElmer Φ 04–303; typical Ar^+ energy 3 keV; see Fig. 3.6). In the first step, the sputter ablation rates of pure and multilayered VSP materials were determined. This gives the correlation between ion dose (ion energy 3 keV) and layer thickness removed (sputter yield Y). For W and Sc_2O_3 VSP-layers, yields of 1.1 and 0.4 atoms/ion were determined. Secondly, IB was applied to the surface of LAD cathodes and the emission recovery was investigated as a function of time, temperature, and ion dose. Qualitative main results were the following: in the W + Sc_2O_3 system emission recovery after IB is inferior and already at low doses (about 10^{17} ions/cm^2), emission falls back to a W–I characteristic. Better IB resistivity is achieved by Re + Sc_2O_3 LAD I cathodes, but the best results were obtained after further addition of the 411x intermediate layer and about 25% of the total life dose of 1700 μAs were achieved for the best variant DG2 + , where additionally the LAD scandate cathodes were not fully activated = under-activated (ua). This is depicted in Fig. 3.17, where the percentage of ILD that survived at a level of 40 A/cm^2 is related to the LAD scandate cathode type. This improvement of scandate cathode types in time scale continued from 1993 to 1997. Since these tests were accelerated tests, the shielding factor related to the dose rate has to be taken into account. With this correction the top value of 25% ILD corresponds to 63% of the reference lifetime dose. In consequence by an improvement of vacuum by a factor of 2 (reduction of average tube pressure by ½ or tube pressure 1×10^{-9} mbar), emission-induced ion bombardment in a CRT during life will be survived at a level of 40 A/cm^2. Figure 3.18 shows the degradation of the *I(U)* characteristic as a function of IB dose: the total emission can be described qualitatively as the superposition of

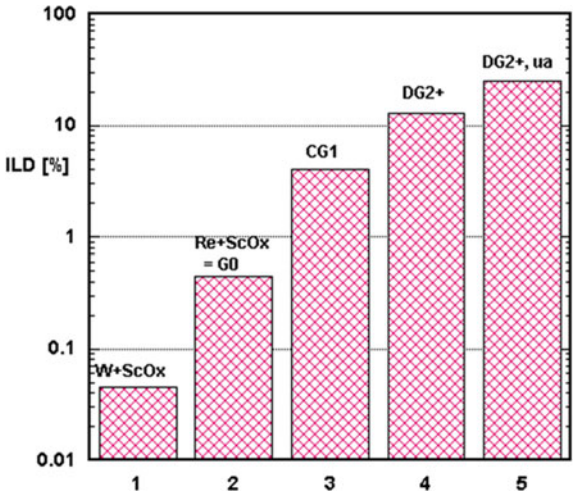


Fig. 3.17 Percentage of ILD survived at a level of 40 A/cm^2 related to the LAD scandate cathode type; compare [30, 31]

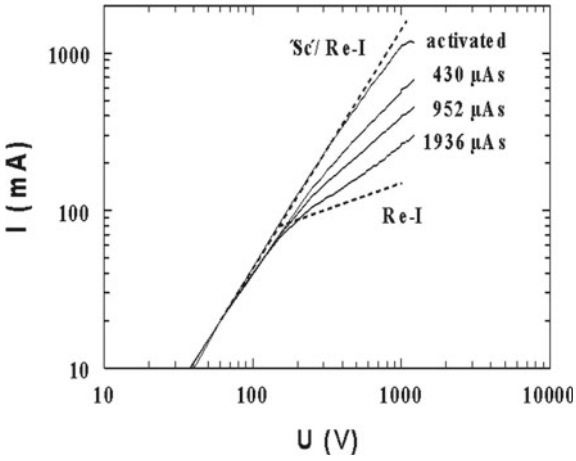


Fig. 3.18 Degradation of *I(U)* characteristic as function of IB dose in diode with Ar ion gun (ILD = 1700 µAs): Dashed lines: perfect 'Sc'/Re–I and Re–I cathodes. Total emission is superposition of area fractions of Scandate (only space charge limited) and Re–I; see [2, 28, 49], G. Gaertner, H. Koops, Chap. 10 of "Vacuum Electronics, Components and Devices", Springer 2008, p. 442

area fractions of Scandate-I (only space charge limited) and Re–I, with some extra space charge leakage into the Re fraction, where the Re–I fraction increases with increasing ion dose [28, 49].

During IB, two processes are competing, namely sputter removal of surface atoms and Ba resupply from the interior. The time change of emission current I

can be formally described [16, 28, 49] by the sum of a destruction term (τ_{IB}) and a regeneration term (τ_{rec}):

$$dI/dt = -I/\tau_{IB} + (I_0 - I)/\tau_{\mathrm{rec}} \tag{3.5}$$

With respect to emission, this initially leads to a fast decline (time constant proportional to inverse ion current) followed by a slow bending to a dynamic equilibrium which is determined by the ratio between destruction and resupply rates. This is illustrated in Fig. 3.19 showing half-drive beam currents of LAD scandate cathodes type DG2 before, during, and after ion bombardment in a 45AX gun for different levels of Ar gas admission. Duration of IB is 15 min for Ar gas admission of 4×10^{-6} mbar (blue circles) and 60 min for 6.7×10^{-6} mbar Ar (red squares). Ar was introduced via a leak valve, since it has no poisoning effect on emission and hence only acts via ionization and ion bombardment of the cathode surface. The longer poisoning interval corresponds to an ion dose of 5.4×10^{-5} mA*mbar*h or 12% of the IB life dose. Similar curves are obtained for ion bombardment in a diode configuration with an Ar ion gun, as shown in Fig. 1 in [28], Fig. 5 of [49] or Fig. 6 of [44], which are not shown here.

Compared to an ILD extrapolation in diode configuration at high IB dose rates (1 μA), the ILD extrapolation at more realistic low dose rates (10 nA) will hence yield an increase in extrapolated IB life of up to a factor of 8 due to the dynamic shielding effect [28, 49].

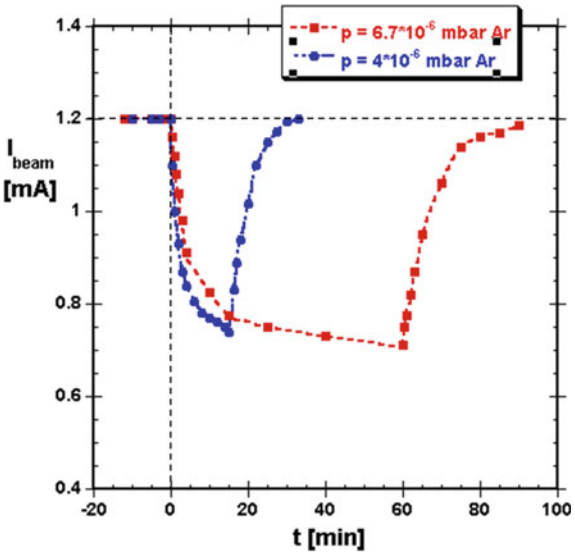


Fig. 3.19 Half-drive beam currents of LAD scandate cathodes type DG2 before, during, and after ion bombardment in a 45AX gun for different levels of Ar gas admission. Duration of IB is 15 min for the blue circles; IB duration is 60 min for the red squares; see [28, 44, 49]

A further requirement is a short recovery time after IB: as specified this should be 1 min. Best time constants achieved for Sc-recovery were 8 min at 965 °C Mo-br. Yet this requirement is only relevant in the initial phase of high tube pressures, which are continuously decreasing, equivalent to an initial dose of 4×10^{-7} mA*mbar*h.

3.2.3.8 Emission Poisoning

From emission poisoning investigations, one can conclude that Ba scandate cathodes are about an order of magnitude less sensitive to poisoning than I cathodes. The figure of merit $p\ (\Delta I = -\ 10\%) = f\ \{\text{gas, cathode type}, T, d_{CA}, i, t\}$ is defined by a 10% current decrease when increasing pressure in the UHV chamber by controlled gas admission through a UHV gas leak. The poisoning threshold $p\ (0.9\ I_0)$ for air admission of 1×10^{-5} mbar for LAD scandate cathodes is about a factor of 10 higher than that of Os/Ru–I, as can be seen from Fig. 3.20. In this case the experiments have been conducted with Re+ScO_x LAD top-layer scandate cathodes (type G0). According to the literature [71–75] and also our own experiments for scandate and I cathodes, the poisoning insensitivity for different gases can be ranked in the sequence (hierarchy): $O_2 < CO_2 <$ air $< CH_4 < CO < H_2$ (most insensitive). For I cathodes water vapor poisoning is less sensitive than O_2 poisoning. Concerning the other parameters, one finds decreasing poisoning sensitivity with increasing temperature and increasing poisoning sensitivity with increased diode spacing (less pumping effect of evaporant!) up to 1.5 mm. The mechanism of emission poisoning is based

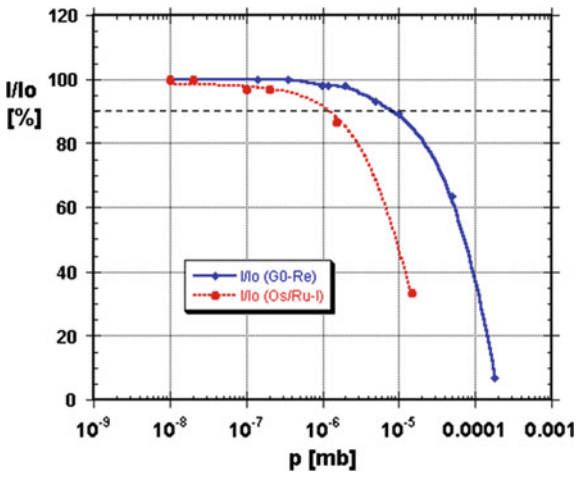


Fig. 3.20 Emission poisoning of LAD scandate cathode of type G0 (Re) and a Os/Ru—I cathode by air introduced at 965 °C Mo-Br. Depicted is the normalized emission (relative emission decline) after 1 min as a function of total pressure increased from UHV by controlled admission of air through a leak valve. The figure of merit is given by $I = 0.9\ I_0$ and is 1×10^{-6} mbar for the Os/Ru–I and 1×10^{-5} mbar for the Ba scandate cathode

on gas molecule adsorption on the cathode surface and hence the change of the work function.

A further important property of these Ba scandate cathodes is that, e.g., after air admission they can be reactivated in vacuum again to $j_{10\%} > 100$ A/cm^2, if they had reached that level before: there is no significant deterioration of emission capability.

3.2.4 High-Resolution Cathode Characterization

High-resolution characterization of LAD top-layer scandate cathodes as prepared and after a certain time of operation was carried out by SEM, EDX, and SAM. Typical SAM analysis conditions and quantification parameters are given in Table 3.4 (sample angle = 0°).

Maps were made at a primary beam voltage of 10 kV, to improve lateral resolution. All spectra were made at a primary beam voltage of 5 kV. For quantification the high energy Auger lines for AI, W, and Re were used. The low energy line for AI interferes with low Ba-lines and the low energy lines of W and Re coincide at 179 eV and cannot be used to discriminate these elements. The same holds for the W-line at 1798 eV and the Re-line at 1799 eV. For that reason, the weaker Re-line on 1858 eV was used for the quantification. The high energy lines are quite insensitive, but there was no alternative for reasonable quantification. The sensitivity data given by Physical Electronics do not give sensitivities for all these Auger lines. Sensitivities for some lines were calculated by using the ratio with other Auger lines. So sensitivities can differ from the elemental sensitivities given by Physical Electronics. So semi-quantitative values of atomic concentrations of the elements present can be obtained. For a quantitative evaluation a correction for the matrix effect and chemical binding effects needs to be carried out and also gauge samples have to be used for the specific elements present. The above conditions are similar to the conditions given by Philips Research in [23, 77].

The information depth in Auger electron spectroscopy (AES) is approximately 2 nm (0.5–5 nm), whereas the lateral resolution depends on the fine-focused electron beam and at best it is about 30 nm, but usually it is averaged over the larger area investigated. A sputter depth profile can be performed to monitor the elemental concentrations as a function of depth. Ar+ ions are used for sputtering, while the sputter conditions have been calibrated using a known SiO_2 layer thickness on Si. The measured sputter rate for SiO_2 was 9.3 nm/min (at raster 1.5 × 1.5 mm^2, 30° tilt angle, 3 keV Ar+, 15.5 nA). This sputter rate may, however, differ for the various

Table 3.4 Typical Auger analysis parameters: V-beam = 10 or 5 kV; I-beam = 20 nA

Element symbol	Sc	O	Ba	AI	W	Re
Peak energy (eV)	340	503	584	1396	1736	1858
Sensitivity	0.200	0.400	0.120	0.070	0.080	0.031

elements and elemental compositions. For Sc_2O_3, a sputter rate of 6 nm/min was calculated under these conditions.

Figure 3.21a shows an SEM micrograph of part of cathode surface SG2 + type after LAD. The surface structures of other cathodes after LAD look similar and will coarsen and shrink on the grains during operation. The result of grain growth and roughening is depicted in the SEM micrograph in Fig. 3.21b for CG1 cathode after gun operation of 750 h. It is a surface area outside the grid hole region, where practically no ion bombardment has taken place, at a magnification of 10,000. Figure 3.21c, d show elemental maps of O(c) and Sc (d) by SAM of the same surface area as in b. The composition here should be similar to the composition after activation, whereas in the sputtered region under the grid hole Sc and Ba concentrations are much lower (not shown here). It is striking that the Sc and O distributions are very similar, indicating that Sc is present as oxide. As can be judged from other elemental maps of LAD cathode surfaces, Ba is more uniformly distributed and usually covers a larger

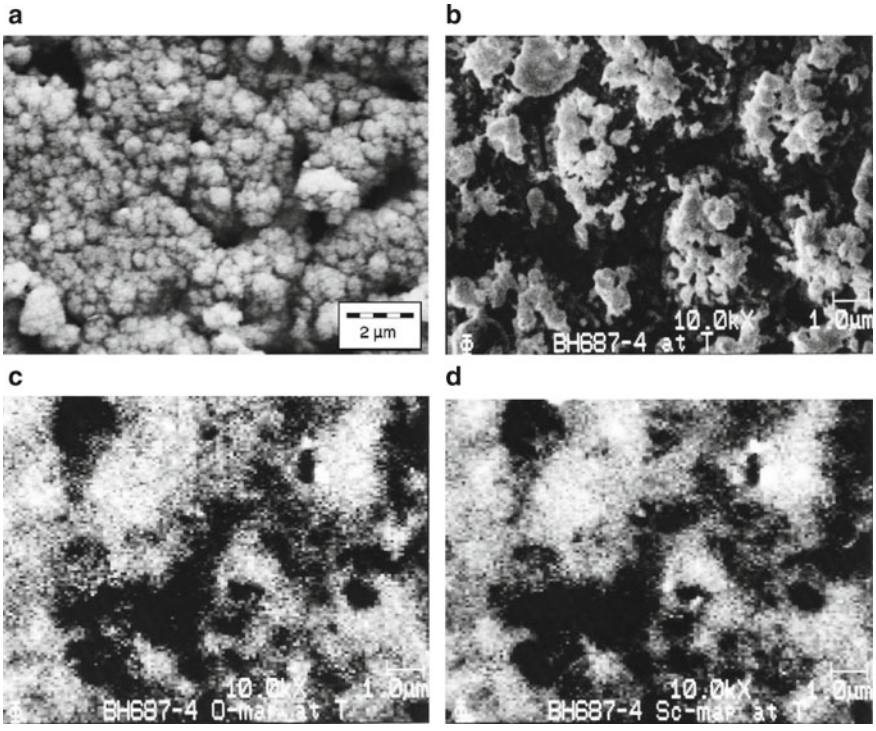


Fig. 3.21 **a** SEM micrograph of surface of SG2 + type cathode after LAD; structures will first coarsen and later on shrink on the grains during operation. **b** SEM micrograph of surface of CG1-type cathode outside grid hole region after gun operation; magnification 10,000x. **c** O elemental map by SAM of cathode surface of CG1-type after gun operation; same area as in 3.21b; magnification 10,000x. **d** Sc elemental map by SAM of cathode surface of CG1-type after gun operation; same area as in 3.21b; magnification 10,000x

surface area than Sc, which partly is concentrated, and is also found outside ScO_x patches. This is due to its resupply from the pores and its longer diffusion lengths. Hence there are two portions of Ba, the one diffusing over the whole cathode surface as a monolayer (including over the complex) and the portion bound in the Ba–Sc–O complex. The elemental composition at room temperature over the surface area at a magnification of 168x was 7% Sc, 15.9% Ba, 24% O, 11.7% Re, and 20.2% W (+4.8% Al). In the area outside the grid of Fig. 3.21b the composition was 8.5% Sc, 14.2% Ba, 20.2% O, and 7.8% W/Re. We can compare this with a similar cathode #CG1_2.1, which also reached a saturated emission current density of 293 A/cm^2 at operating temperature. There we obtain (flocky structure) 7.9% Sc, 17.2% Ba, 20.0% O, 13.5% Re, and 23.9% W (+ 3.1% Al). These results are quite similar.

A sputter depth profile of DG1 cathode (also after gun test) is given in Fig. 3.22a of an area of 200 × 200 μm^2 outside the grid hole region. For comparison the LAD preparation conditions for this cathode were 411_0.35: 8400 pulses, Re: 12,500 pulses, ScO_x: 8300 pulses ($i_{10\%}$ (under-activated) = 230 A/cm^2). Averaging the Ba:Sc:O ratio from 0 to 6 min of sputtering, which corresponds to n Sc sputter depth of 36 nm, one obtains Ba:Sc:O = 1.78: 1: 1.26. Here oxygen associated with W was not taken into account. Figure 3.22b shows another sputter depth profile now of a DG2 + cathode of a surface area of 500 × 500 μm^2. This cathode was mounted in a diode and reached a maximum saturated emission current density of 267 A/cm^2, but with an early deterioration. The LAD preparation conditions for this cathode were 411_0.35: 16,500 pulses, Re: 25,000 pulses, ScO_x: 8400 pulses. Since more Re has been ablated, a higher Re concentration of 55% was reached. Here the averaging over an Sc sputter depth of 36 nm results in Ba:Sc:O = 2.44: 1: 1.4.

The average value from the surface analysis is Ba:Sc:O = 2.04: 1: 2.78. Since Ba as a monolayer on W or Re is independent of the position of the complex, the Ba to Sc ratio of the complex is probably a bit lower.

Figure 3.23a, b show SAM analysis results of a CG1 cathode after life test in a 45AX gun. The Sc elemental map under the rectangular grid opening of 0.7 mm × 0.5 mm illustrates the final removal of Sc by ion bombardment caused by the electron beam, whereas the Ba map is more uniform due to the resupply from the pores.

The micrographs of Fig. 3.21 were first shown by G. Gaertner in a presentation at RWTH Aachen in 2000 and again at South East University Nanjing in 2004; see also [67].

We can compare these data also with the literature. van Veen by XPS [76] and Hasker by AES [77] found a Ba/Sc ratio of 1 for scandate cathodes with emission <100 A/cm^2. According to the investigations of Y. Wang and her coworkers of improved Sc–I cathodes, the emissive surface of the scandate cathodes investigated consists of an active surface multilayer containing (Ba, Sc, O) with a ratio of Ba:Sc:O around 1.9: 1: 2.2 [10]. It is interesting that J. Hasker already presented a depth profile in [15] after activation for a scandate cathode prepared with W and ScH_2 in the top layer, which looks quite similar to Fig. 3.22, Ba and Sc decreasing within the first 6 min of sputtering. The signal ratio Sc/Ba = 2 is consistent with the lower emission capability of this cathode type.

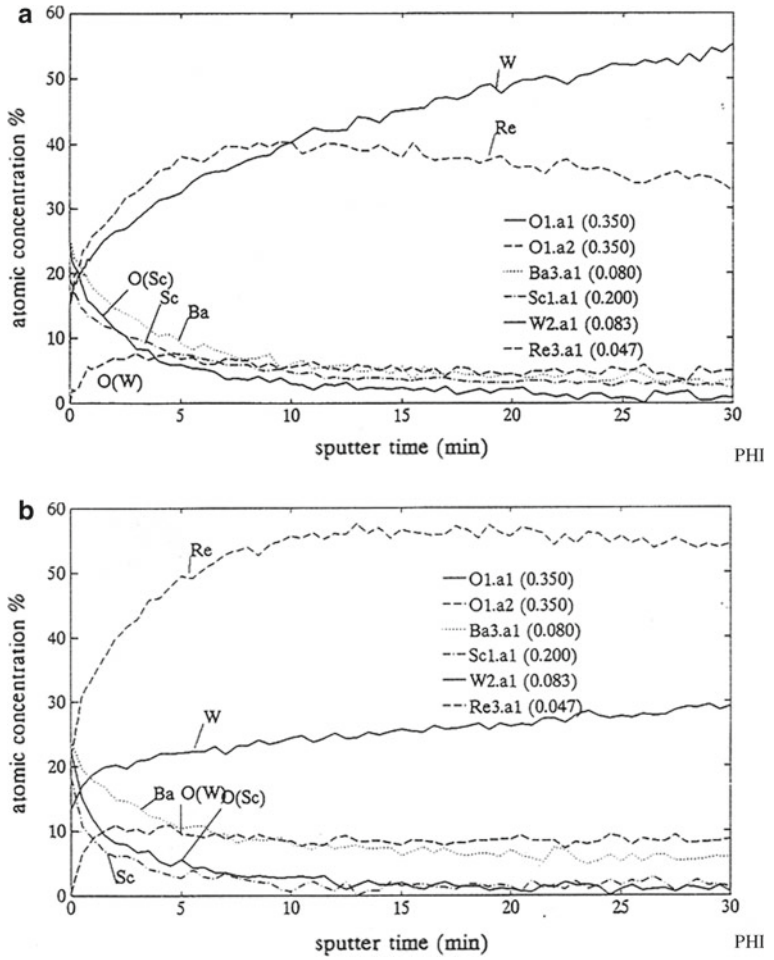


Fig. 3.22 **a** Sputter depth profile of cathode of DG2 type, area of 200 × 200 μm^2 outside grid hole region. **b** Sputter depth profile of cathode DG2 + type, surface area of 500 × 500 μm^2; 5 min sputtering of Sc corresponds to 30 nm depth

All these results would be compatible with a compound containing $Ba_3Sc_2O_6$, which is somewhat in contradiction with the stable compound $Ba_2Sc_2O_5$ in this system. It is interesting that S. Magnus and colleagues have investigated ternary compounds in the $BaO{\cdot}Sc_2O_3{\cdot}WO_3$ system [78]. Together with 10%W and BaO they found increased best emission for the compound $Ba_3Sc_2WO_9$ compared to other Sc containing oxides, yet far from optimized scandate cathode emission. This is consistent with Aachen LAD experiments, where $Ba_2Sc_2O_5$ was used as a target instead of 411x, but without success. However, it has not been discussed before, what happens if W is replaced by Re. Combining the two stable compounds $Ba_2Sc_2O_5$ and $BaReO_4$ one would obtain $Ba_3Sc_2ReO_9$. It would be an alternative route for Ba scandate

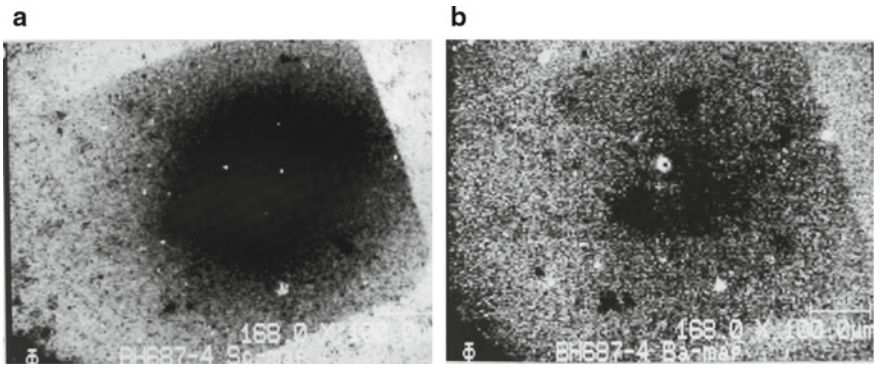


Fig. 3.23 **a** SAM analysis of CG1 cathode after life test in gun; Sc elemental map under the rectangular grid opening of 0.7 mm × 0.5 mm; magnification × 168. **b** SAM analysis CG1 cathode after life test in gun; Ba elemental map under the rectangular grid opening of 0.7 mm × 0.5 mm; magnification × 168

cathode preparation, to start and ablate directly the emissive compound/complex and not to have to rely on a high-temperature synthesis under cathode activation conditions as in our case.

3.2.5 Conclusions on LAD Top-Layer Scandate Cathodes

The outstanding emission capability of LAD top-layer scandate cathodes and their preparation has been discussed in detail. Yet there is some scatter in the emission properties, which can partly be linked to changes of the target surfaces and nanoparticle generation during repeated ablation cycles. The thermomechanical stability of the 411x target still needs to be improved. A further improvement is expected by using more fine-grained I cathode bases similar to SDD cathodes. Alternate and cheaper preparation methods such as sputtering should be investigated for the preparation of similar material layer structures, which anyway will coarsen during activation. In general the formation of the emissive Ba-scandate complex/compound via the high-temperature activation process is a critical step and it would surely help, if this compound could be identified by XRD or other methods and is, for example, prepared via a low (or high) temperature chemical route.

What is now the explanation for the very high pulsed emission capability of the LAD Ba scandate cathode? First one must notice that the high emission capability was just a side effect of achieving high robustness versus ion bombardment at a much lower emission level. In reaching this goal the surface uniformity and Sc reservoir, or better (Ba,Sc,O) reservoir below the surface was increased. The increase in uniformity and emission capability was linked to an increase of the Richardson constant A_R from 3 to 8 A/(cm^{2}K^2), whereas the Richardson work function remained

constant at the low value of about 1.15 eV. The Re base guaranteed at least a higher fall back emission level during and after ion bombardment and also reactions with W could be partly suppressed. Re and the additional 411x reservoir turned out to be essential for high IB robustness. And finally one should notice that oxide cathodes with about the same effective work function can also deliver pulsed emission current densities of 100 A/cm^2 at similar operating temperatures. Hence I am very much in favor of an oxide-semiconductor theory, where the conductivity limitation has been overcome by the conducting base with only a thin but more stable oxide layer on top and good conductivity also here provided by Re at the surface. In such a semiconductor theory, Ba will sit on top of a (Ba,Sc,O) containing oxide with an oxygen deficiency and a high density of these sites will be the aim. In the following part, this will be discussed in more detail.

3.3 Nanosized-Scandia Doped Dispenser Cathodes with W-Base Matrix (Yiman Wang)

3.3.1 Motivation for the Development of the Nanosized-Scandia Doped Dispenser (SDD) Cathodes

In the introduction, the research on different types of Ba scandate dispenser cathodes has been summarized and the LAD top-layer scandate cathode has been discussed in more detail.

In practice, improved impregnated scandate cathodes are the only Ba scandate cathodes which have been utilized in several vacuum microwave electron devices with pulsed current densities between 10 and 30 A/cm^2 at 1000–1050 °C$_b$ [79, 80]; even then they still suffered from poor emission uniformity and reproducibility.

Recently, there are growing interests in the development of the next generation of vacuum electron devices (VEDs) to operate at higher powers and frequencies and in the improvement of electron beam instruments or tubes to reduce the beam size and/or increase the beam brightness.

These make cathodes face a great challenge of providing much higher current densities at lower operating temperatures than that of presently available cathodes.

Therefore, there is strong demand and bright prospects in the application of Ba scandate dispenser cathodes for advanced vacuum electron devices and others after further improvement.

The requirements for cathodes to apply in advanced VEDs are

1. High SCL current densities in both pulsed and DC modes with excellent emission uniformity for good beam quality [48].
2. A long lifetime at reasonable operating temperatures.

3. A low evaporation rate of Ba and other volatile components for minimizing electric and vacuum breakdown.
4. Robustness in fabrication and operation.
5. Reproducibility in manufacturing.

To meet these requirements, a new type of Ba scandate dispenser cathode, the nanosized-scandia doped dispenser (simplify SDD) cathode with Scandia nanoparticles doped W matrix has been developed [10]. The characteristics of this cathode have been investigated comprehensively; high current densities have been routinely demonstrated.

3.3.2 Structure, Features, and Manufacture of Nanosized-Scandia Doped Dispenser Cathodes

3.3.2.1 Basic Ideas on Improvement of Ba Scandate Dispenser Cathodes

Historically, Ba scandate dispenser cathodes have been found capable of providing high emission current densities but suffer from the problem of extremely nonuniform current distribution. It has been commented that a Ba scandate cathode provides poor beam quality when used in electron guns due to the emission nonuniformity, and thus limits the applications of Ba scandate cathodes in VEDs [8].

A Ba scandate dispenser cathode is basically a Ba dispenser cathode with the addition of scandium oxide into the matrix, impregnant, or on its top. It is generally accepted that the electron emission capability and the emission uniformity of Ba scandate dispenser cathodes greatly depend on the distribution uniformity of scandium on the surface of cathodes [16–19]. In view of the well-known fact, that the diffusion rate of scandium is much lower than that of Ba [35, 81], two factors are favorable for the improvement of surface uniformity of Sc. The most important one is a perfect dispersity of small-sized scandium oxide in the matrix. This enables the reactions of scandium oxide with tungsten and other substances to pervade all over the matrix, leading to adequate and well-distributed release of scandium during fabrication and activation. The other is the small average size of the tungsten grains of the matrix, which favors the diffusion and coverage of Sc on the surface.

For a mixed matrix Ba scandate cathode developed by Philips, tungsten powder of 4 μm and Sc_2O_3 powder of 10 μm on average were mixed mechanically [15, 23]. The large sizes of both tungsten and Sc_2O_3 powders and the mixing method result in uneven distribution of Sc on the cathode surface. In the case of a thin-film top-layer Ba scandate cathode, as reported by Sasaki [35], the coated W + Sc-tungstate containing film could be divided into two parts after activation: film-covered impregnated areas A and the tungsten metal areas B (Fig. 3.24), where free Sc was generated in the areas A by reaction with Ba. Since surface diffusion length of scandium is too short to cover the gaps between A areas (i.e., to cover B areas), this physical nonuniformity causes

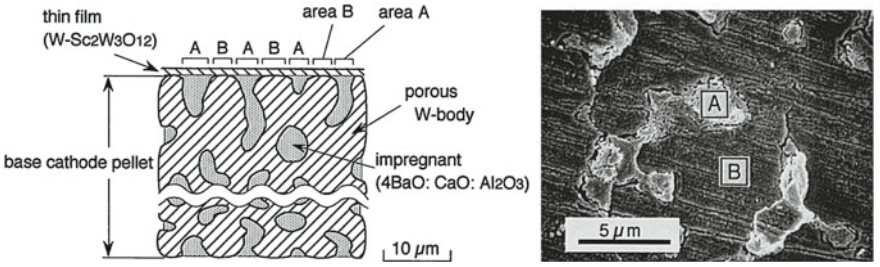


Fig. 3.24 Nonuniformity of Sc coverage on the surface of the thin-film top-layer Ba scandate cathode caused by insufficient diffusion length of scandium to cover the gap B. The graph is based on Fig. 1 of [35]

the emission nonuniformity. Hence a BaO containing layer below the Sc-containing top layer must be further introduced for improvement.

In order to overcome these disadvantages, a doping process has been developed for uniformly adding nanosized scandium oxide to tungsten powder in the manufacturing of a new type of Ba scandate dispenser cathode—nanosized-scandia doped dispenser (SDD) cathode.

3.3.2.2 Structure of SDD Cathodes

There are two types of SDD cathodes, the Scandia Doped Impregnated cathode (SDI cathode) and the Scandia Doped Pressed cathode (SDP cathode). The SDI cathode has a porous matrix fabricated by calcining, reducing, pressing, and sintering the precursor powder containing tungsten and (3–5) wt% scandium oxide. This porous matrix is then impregnated with Ba–Ca aluminate, which is traditionally used for Ba dispenser cathodes, mostly with mole ratios of BaO: CaO: Al_2O_3 of 4:1:1, at 1600–1700 °C for a short period to ensure a weight gain of (8–10) wt% is realized. After that, the surface residual impregnants are thoroughly removed by ultrasonic water cleaning [10]. The SDP cathode is made in a simpler process flow of pressing and sintering tungsten powder that already contains Sc_2O_3 and BaO, CaO, Al_2O_3 in similar mole fractions to that of 411 Ba, Ca aluminate [82].

In both cases, doping processes are used for adding scandium oxide and other compounds into tungsten.

Materials and cathodes for either type above have the following characteristics:

1. Precursor powders are tungsten, uniformly doped with nanosized scandium oxide (for SDI cathode), or compounds containing Sc, Ba, Ca, Al, and O (for SDP cathode). The powder size is several hundred nanometers to 1–2 μm.
2. Matrices of SDI cathodes, that are made from scandia doped tungsten powder and sintered at proper temperatures, have porosities of 24–28% similar to that of Ba dispenser cathodes but exhibit tungsten grains of submicron to 1–2 μ size and uniformly distributed nanosized scandia particles [43].

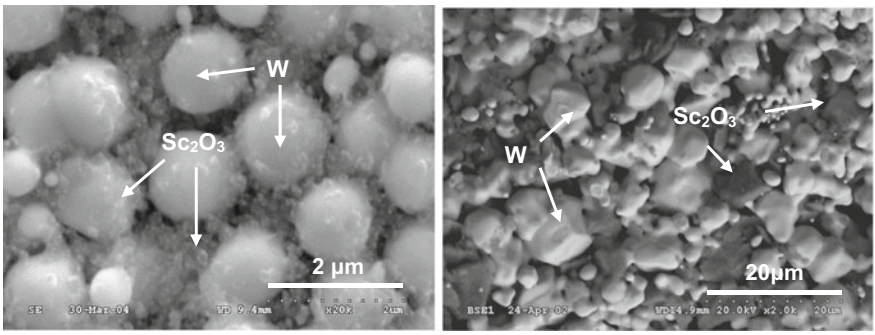


Fig. 3.25 **a** Scanning Electron Microscopy (SEM) image of a matrix composed with nanosized-scandia doped tungsten powder. **b** Back Scattered Electron (BSE) image of a matrix produced by mechanical mixing of Sc_2O_3 and W

3. Cathodes produced from the abovementioned matrices (for SDI cathodes) or the doped precursor powders (for SDP cathodes) consist of tungsten grains of submicron to 1–2 μ and uniformly distributed powder particles containing Sc, Ba, Ca, Al, and O [10, 82].

For simplicity sake, scandia-doped impregnated cathode is considered as a representative for SDD cathode in most cases below.

It has been noticed that the microstructures of matrices composed of the scandia doped tungsten powder are strongly different from that of the mechanically mixed scandium oxide with tungsten, shown in Fig. 3.25a, b, respectively. In contrast to the case of the mechanically mixed matrix where Sc_2O_3 lumps insert locally to W grains, the nanosized scandium oxide particles are evenly dispersed among and on the W grains in the scandia doped matrix. X-Ray Diffraction (XRD) analysis reveals that the Sc_2O_3 is the sole form of scandium in the W matrix [83].

Since the SDD cathode is based on above scandia-dispersed and fine-grained matrix, more efficient and extensively distributed reactions of scandium oxide with tungsten and Ba, Ca aluminates can be expected at processes of impregnation and activation as compared to other types of Ba scandate dispenser cathodes, allowing for a sufficient release and uniform spread of free scandium in the body and onto the surface of the cathode.

It has been further found that the uniformly distributed Sc_2O_3 in the matrix of Fig. 3.25a had mostly disappeared from the surface of cathode after impregnation and ultrasonic water cleaning, as shown in Fig. 3.26a. There are only small amounts of Ba, Sc, Ca, Al, and O to be observed in Fig. 3.26b within the detection depth of Energy Dispersive X-Ray Spectroscopy (EDS) analysis. This reflects that the aforementioned full reactions between Sc_2O_3 and others actually took place and the resultant reaction products were then dissolved and removed by water cleaning. The details will be discussed later.

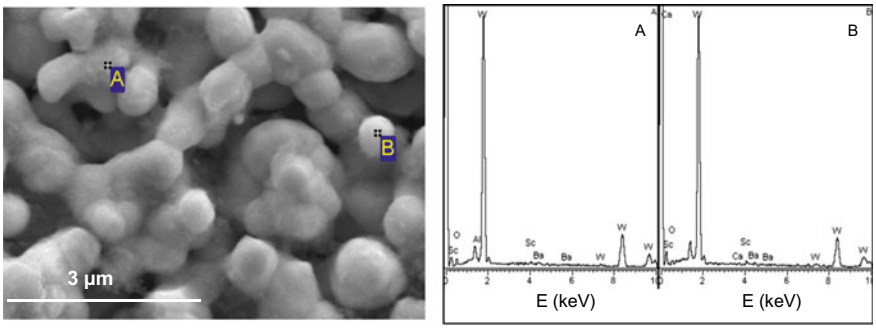


Fig. 3.26 **a** SEM image of an SDD cathode after impregnation and ultrasonic water cleaning. **b** EDS analysis at point A and B

As presented in SEM image of Fig. 3.26a, the cathode contains uniform pores with short pore distances. Such a structure is advantageous for the diffusion of active substances, including Sc, from the body of the emitter to its surface and over the tungsten grains during activation. As a result, a uniform surface layer containing Ba, Sc (and O) is formed on W substrate after activation. This has been proven by Scanning Auger Mapping (SAM) of Ba and Sc done in an activated cathode in Fig. 3.27a. The SAM maps of Sc and Ba of Fig. 3.27c, d show that Sc distributes over the surface of W grains and coincides perfectly with Ba while the latter exhibits a more even distribution.

In addition, compared with the top-layer scandate cathodes by thin-film techniques, the SDD cathodes have the advantage of being controllable in manufacturing, providing excellent resistance to possible surface abrasions in fabrication and assembling, and therefore are suited for practical applications.

3.3.2.3 Doping Processes

Several doping processes have been developed for adding scandium oxide into tungsten powder, replacing mechanical mixing technology, for scandia-doped impregnated dispenser cathodes.

The first doping technology applied to powder preparation is Liquid–Solid (L–S) doping method. In L–S doping, scandia is added to the tungsten oxide in the form of a scandium nitride aqueous solution. By reduction in a hydrogen atmosphere, the Sc-doped tungsten oxide is converted into scandia-doped metallic W powder. Compared with the powder produced by mechanical mixing, the size of scandia is obviously reduced while its distribution among W grains remains unsatisfactory.

As a major improvement, Liquid–Liquid (L–L) doping technologies were introduced. Two methods have been utilized. The technology which has been mainly adopted is a sol–gel process [84]. In the process, scandium nitride and ammonium

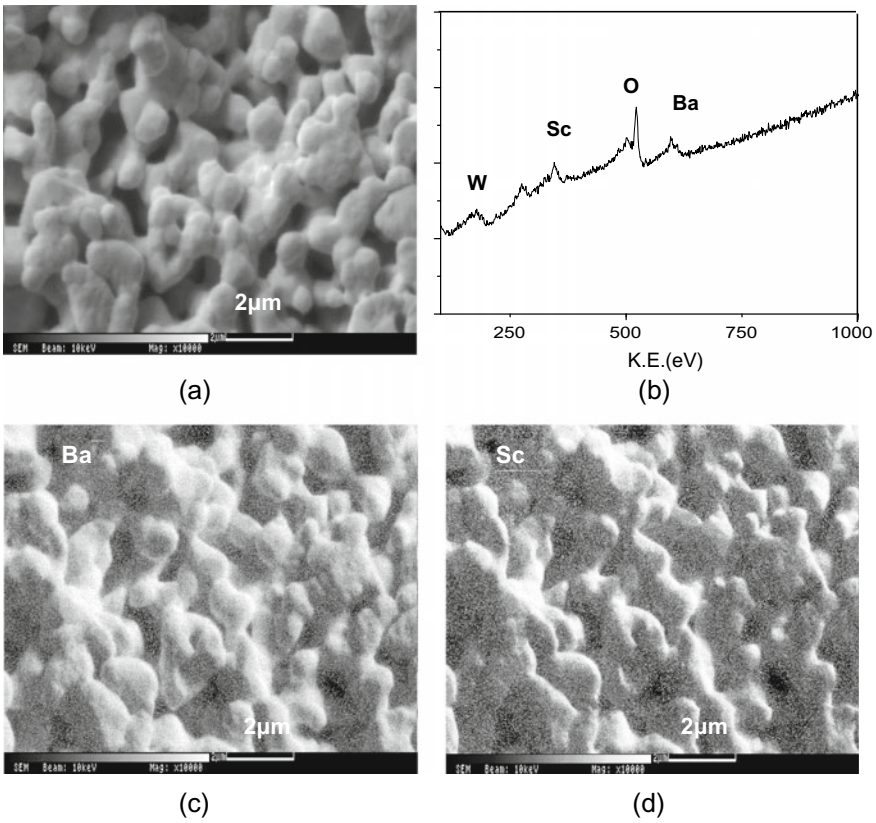


Fig. 3.27 Coverage of Ba and Sc on the surface of an SDD cathode after activation. (**a**) SEM image. (**b**) Auger Electron Spectrum taken from area shown on the SEM image. (**c**) SAM mapping of Ba. (**d**) SAM mapping of Sc

metatungstate (AMT) are used as the raw materials. The above materials were mixed in aqueous solutions to form sol and then converted into gel by bake-out. After calcining in air to remove the organics containing N and C, the Sc_2O_3 doped tungsten oxide powder was reduced into scandia-doped tungsten powder in hydrogen atmosphere. Since the raw materials are mixed in their aqueous solutions, the admixture is exceedingly uniform.

To illustrate the improvement, SEM images of the powders produced by L–S doping and sol–gel process are compared in Fig. 3.28a and the left side of Fig. 3.28b. On the right side of Fig. 3.28b a Transmission Electron Microscopy (TEM) image of a grain of the sol–gel doping powder is displayed, which indicates that tiny scandium oxide particulates are well dispersed over the surface of a tungsten grain.

By controlling the parameters of the processes, the resultant doped-tungsten grain sizes from submicron to about 2 μm can be normally produced.

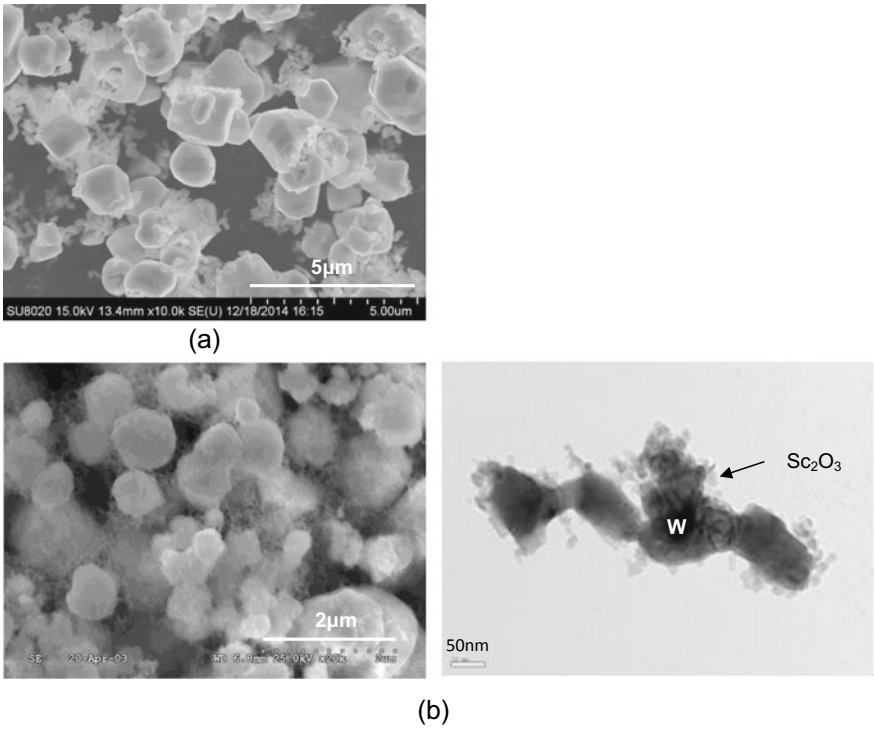


Fig. 3.28 Images of scandia doped tungsten powders. **a** SEM image of powder by L–S doping. **b** SEM image of powder by sol–gel doping process (left) and TEM image of one of the grains (right)

Another technology, spray drying, has also been developed for nanosized-scandia doped tungsten powder preparation. In the process the raw materials are mixed by spray drying method with Spray Dryer [85] and then followed by the same procedures as that of the sol–gel process above. Powders with a quasi-Gaussian distribution and mean size of around 1 or 2 μm, shown in Fig. 3.29a, can be routinely produced. The mixing uniformity of Sc_2O_3 with W, which is analyzed by Energy Dispersive X-Ray Spectroscopy (EDS) at different grains and illustrated in Fig. 3.29b, is similar to that of the sol–gel process.

In general, while the sol–gel technology results in the most homogeneous mixing of Sc_2O_3 with W, the spray drying process has the advantages that the morphology, size, and size distribution of the produced powder are more controllable based on the spraying conditions [86].

Both sol–gel and spray-drying are acceptable for the manufacture of Scandia Doped Pressed cathodes (SDP). In the doping process, ammonium metatungstate (AMT) and nitrides of scandium, barium, calcium, and aluminum are used as the raw materials.

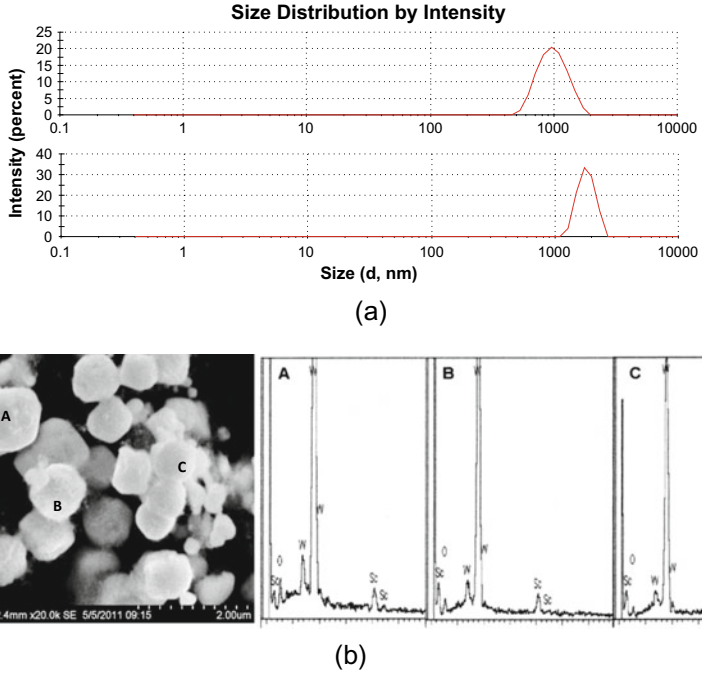


Fig. 3.29 Sizes **a** and elemental distributions **b** of scandia doped tungsten powders by spray-drying doping

3.3.3 Emission Characteristics of SDD Cathodes

3.3.3.1 Emission Evaluation Methods

The ways of emission characterization for dispenser cathodes used in VEDs have been summarized in detail by Shroff [87] in the later 90s and Gilmour [48] recently. Two types of test vehicles have been used: the close-spaced planar diode structure and the electron gun-type structure.

In close-spaced diodes, the zero-field saturated current density which is based on the Schottky equation is normally used to characterize the emission of Ba dispenser cathodes [87].

Owing to the peculiarities of Ba scandate cathodes that their saturation characteristic is not Schottky-like [8], which will be discussed later, it becomes problematic to adopt this approach in emission evaluation. Several alternatives have been proposed to approximate the saturation current density. Maloney [50] suggested the 73% slope method and Manenschijn of Philips used $I_{10\%}$ instead [39]. The latter has been discussed in detail in Sect. 3.2 of this chapter and also used for the emission evaluation of LAD scandate cathodes.

As cathode emission is required to maintain Space Charge Limited (SCL) when working in most VEDs [48], determination of SCL emission region becomes important. From the relation between current and voltage in a parallel-plane diode configuration, SCL region can be found according to the Child-Langmuir law by plotting current density as a function of voltage in log J-log U format. The plot is nearly a straight line for space charge limited operation. A critical SCL current density is then defined as the point where the deviation from the linear part of the tested log J-log U plots becomes noticeable. Though the theoretical slope of the linear part in this type of plots should be 1.5 according to the Child law, the slopes are actually affected by many factors in experiments, such as the differences between the structures of real test vehicles and the ideal parallel-plane diode, the characters of the tested cathodes themselves, and so on. It has been found that the slopes range from 1.33–1.46 for Ba scandate cathodes [9, 10, 39, 88] at operating temperatures in a diode configuration. Under comparable test conditions, the slope will reflect the emission uniformity to a certain extent.

This critical SCL current density provides a good reference for VEDs application and is adopted for the evaluation of SDD cathodes in most cases.

It should be noticed that the critical SCL current density is lower than the "saturated" current density determined by the $I_{10\%}$ for the same emission level.

Apart from testing in a diode, emission of SDD cathode is also characterized in a Pierce gun structure where a proper electrode arrangement leads to a parallel electron flow passing though the cathode–anode space. The Child-Langmuir law can be written as $I = \mathrm{P}\, U^{3/2,}$ where $I = J\mathrm{A}_{\mathrm{cath}}$ with $\mathrm{A}_{\mathrm{cath}}$ being the cathode surface area and P is known as perveance [48] (I is the emission current and U is the voltage applied between cathode and anode measured in Volt). Since the perveance is a function of geometry only, it can be designed and actualized by the electrode geometric structures. At a designed perveance, the current which meets the above law will be in the full space charge regime.

Another method, which is widely used for evaluation of Ba dispenser cathodes, is the well-known roll-off curve (or activity plot) and its normalized version—Miram curve [89]. For a certain current density, the knee temperature from SCL to Temperature Limited (TL) regions in a roll-off or Miram curve and the sharpness of the knee reflects the emission capability and emission uniformity of the cathode, respectively: the lower the knee temperature the higher the emission capability; the sharper the knee the better the emission uniformity. Among the different ways in determining the knee temperature, the intersection of linear extrapolation of the TL and SCL lines is applied.

By superposition of Miram curve against a family of normalized theoretical temperature-limited emission current curves based on Richardson–Dushman equation that have been constructed over a range of electron work function values, the Practical Work Function Distribution (PWFD) analysis method was derived [89]. Both roll-off curve and PWFD plot have been utilized as a useful tool to make emission comparison for different types of cathodes and to diagnose the characteristics of the cathodes in real applications [90]. They are applied in SDD cathode study for emission, work function evaluation, and for fundamental research.

The traditional Richardson plot is used for the estimation of the Richardson work function.

The tested data are treated and analyzed by a dedicatedly developed software.

The cathode temperatures described below are measured by an intensity comparison pyrometer with a minimum target diameter of 0.1 mm at the sleeves near the top of the cathode pellets through a quartz glass viewing window. Molybdenum brightness temperatures (°C_b) are normally utilized in emission characteristics expression. The brightness temperatures are also converted into true temperatures (°C or K), where an effective wavelength of 0.65 μm and an emissivity of 0.4 for Mo are adopted, in the determination of work functions, evaporation rates, and so on. In this conversion, a brightness temperature of 950 °C_b is related to a true temperature of 1015 °C or 1288 K, and 850 °C_b for 905 °C or 1178 K.

3.3.3.2 Cathode Emission Characteristics

(a) Cathode Activation

Cathode activation is an essential procedure for most of the thermionic cathodes. For Ba dispenser cathode, the activation is normally completed at 1200–1250 °C_b in less than one hour, because a forced diffusion of barium at 1160 °C is necessary for starting full and patchless emission by continuous barium dispensation until the entire surface is covered by barium layer at 1240 °C [91]. Unlike the Ba dispenser cathode, a relatively long time at a lower temperature of 1150 °C_b is necessary to activate Ba scandate cathodes properly. This fact has been reported for various types of Ba scandate cathodes [16, 19] as well as for SDD cathode. For the latter, the emission continuously increases with the activation time up to 2 h or longer related to different preparation processes and W grain sizes. The reason for that will be discussed in the following sections.

(b) Emission Capability

Emission capability of SDD cathode is evaluated mainly by the critical SCL current density, and tested in either parallel-plane diode configuration or Pierce-type electron gun structure.

In close-spaced diodes, cathodes mostly in diameters (D) of 2–3 mm are tested in a dynamic UHV system with a water-cooled copper anode. The distance (d) between anode and cathode is adjustable and set at about 0.3–0.4 mm, within the range of d/D = 0.1–0.2 [87], to minimize the effect of edge emission. Pulsed emission was normally measured with pulses of 5–10 μs in width and 50–100 Hz repetition frequency in an automatically graduated-raising manner.

Critical SCL current densities of over 30 A/cm^2 at 850 °C_b and over 100 A/cm^2 at 950 °C_b, Mo-brightness temperature, have been routinely measured for either SDI or SDP cathodes since their development [10, 43, 88]. This emission capability remains unchanged when operating at high duty cycles and long pulse widths. As an example, a set of experimental Log J-Log U plots measured for a cathode 2 mm

in diameter is shown in Fig. 3.30a. The relative waveforms of current and voltage with pulse widths of 50 μs–250 μs, at a same repetition frequency of 100 Hz, are displayed in Fig. 3.30b. While SCL current density of over 120 A/cm^2 is achieved at 950 °C$_b$, a total current of 3.1 A which corresponds to a current density of 100 A/cm^2 is identical for each duty cycle, indicating that the SDD cathodes are able to provide high current densities at a variety of pulse conditions.

From the I-U characteristics above, the well-known unsaturated emission behavior or abnormal Schottky effect of Ba scandate cathodes are clearly displayed, especially

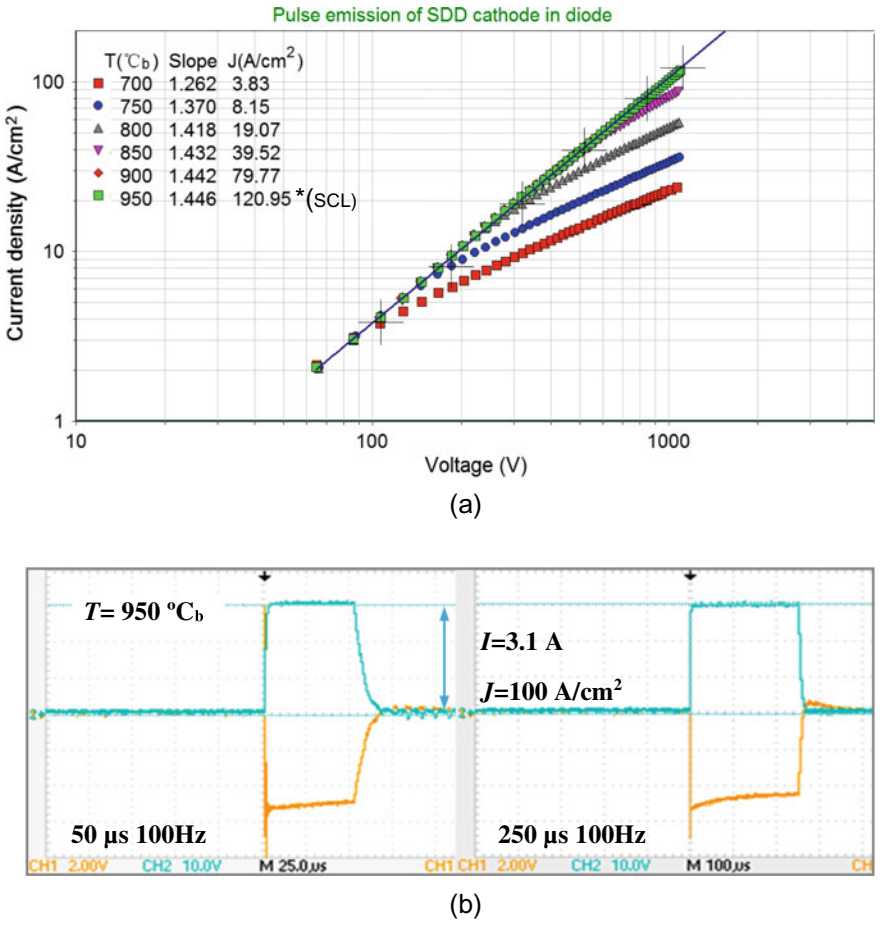


Fig. 3.30 Pulsed emission characteristics of SDD cathode tested in close-spaced diode. **a** Log *J*-Log *U* plots at different operating temperatures. Critical current densities and slopes of the linear part of the plots are shown on left corner. **b** Pulsed waveforms of I and U with pulse widths of 50 μs (left) and 250 μs (right) at a repetition frequency of 100 Hz, tested from an SDD cathode of 2 mm in diameter at cathode temperature of 950 °C$_b$. A total current of 3.1 A corresponds to 100 A/cm^2 current density

at high temperatures. This feature brings about problems in the determination of zero field emission J_0 by extrapolating a plot of ln(J) versus $\sqrt{E}$ to E = 0 at normal operating temperatures, causing inaccuracy in the Richardson work function and Richardson constant. This problem may be mitigated by only taking J_0 at relatively low temperatures. In this manner, the Richardson work function ϕ_R of around (1.13–1.15) eV and Richardson constant A_R of (2–4) $A{\cdot}cm^{-2}\cdot K^{-1}$ have been estimated and denoted in Fig. 3.31 as an example for reference. The temperature coefficient α introduced in [87], then, is about 3×10^{-4} $eV{\cdot}K^{-1}$.

In DC operating mode, the electron cooling effect begins to emerge, leading to a drop in cathode temperature. This effect becomes obvious when high DC current is emitted.

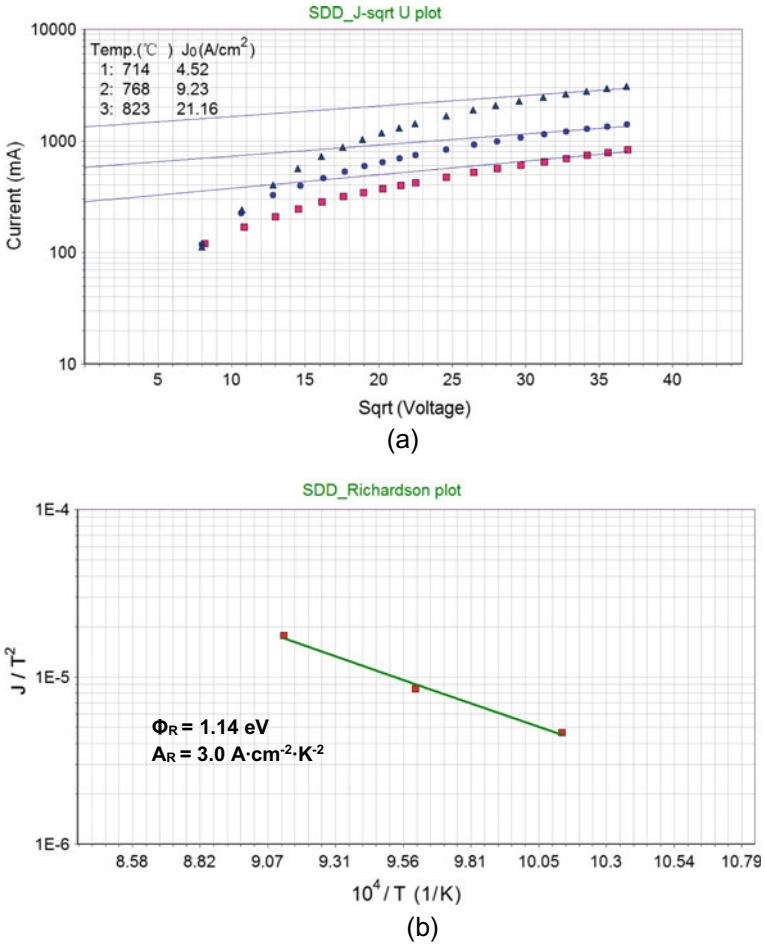


Fig. 3.31 **a** Zero field emission taken from low operating temperatures. **b** Richardson plot derived by the data from (**a**)

In order to evaluate the DC emission capability, the temperature drop needs to be corrected by heating power compensation [92]. DC characters measured at an initial temperature of 850 °C_b, with and without temperature compensation, are shown in Fig. 3.32a.

For cathodes with a reduced diameter of 1 mm in the diode configuration with d_{a-c} of 0.15–0.2 mm, DC current densities of 60–70 A/cm^2 were repeatedly measured at 950 °C_b after compensation for electron cooling. A group of Log *J*-Log *U* plots is shown in Fig. 3.32b together with a pulsed emission plot measured at 950 °C_b for comparison. The DC emission coincides with the pulsed emission until its measurement limitation is imposed by the overloading of the anode.

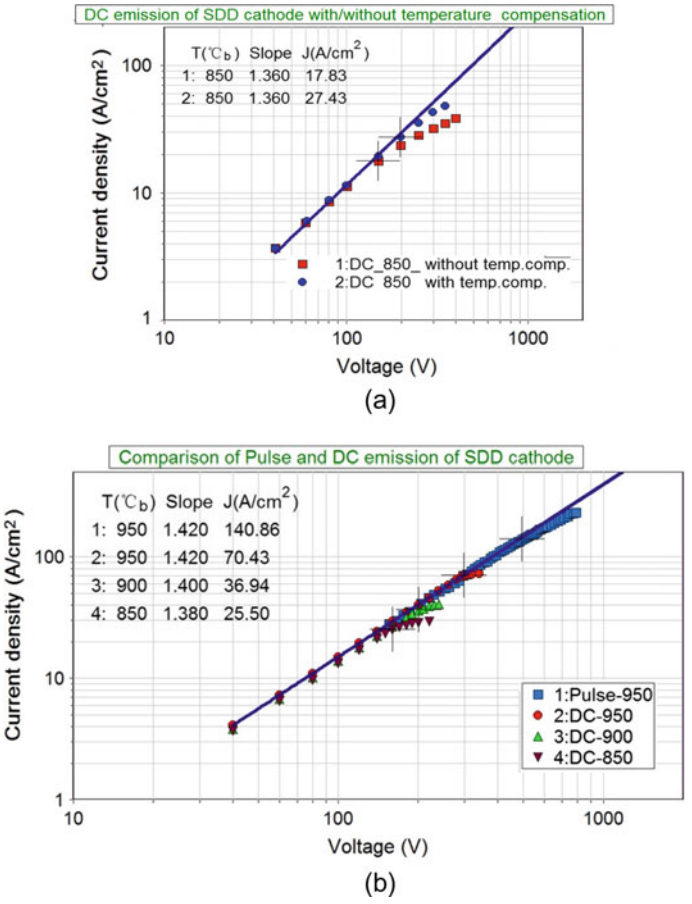
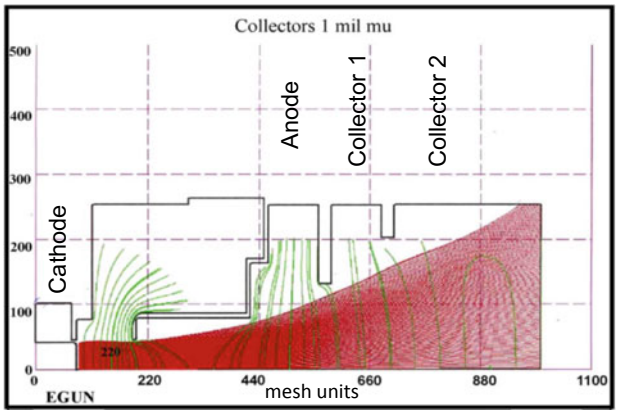


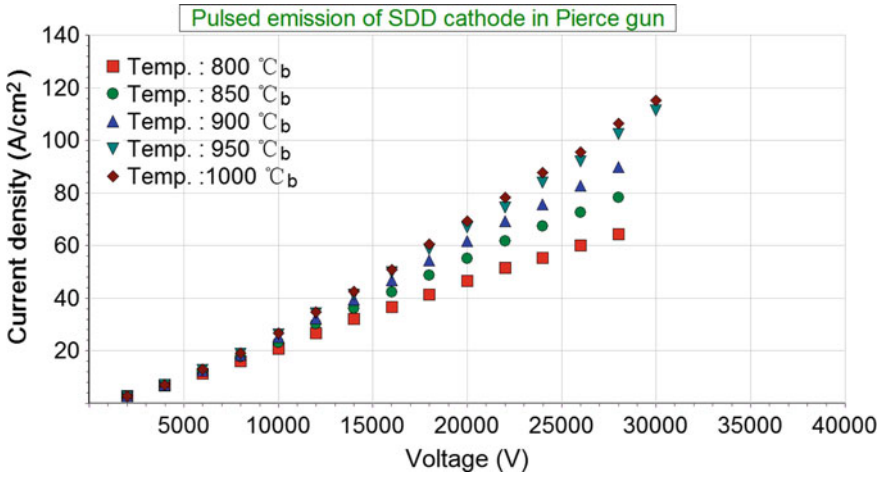
Fig. 3.32 DC emission of SDD cathode. **a** DC emission with and without temperature compensation, indicating the effect of electron cooling. **b** Comparison of pulse and DC emission with temperature compensation. The graph is based on re-evaluation of data presented in [92]

SDI cathodes have been further tested in Pierce-type electron gun tubes. This test structure allows for high current loadings without the drawbacks happening in diodes, such as cathode edge emission, barium or contaminants backscattering from the anode, and plasma formation. Furthermore, the test results are a useful reference for cathodes application in real VEDs.

A schematic diagram of a Pierce electron gun consisting of an anode and two depressed collectors with its basic electron trajectory are shown in Fig. 3.33a. The cathode disk is with a planar surface in diameter of 2 mm. The designed perveance of the electron gun is 0.6 μp [93]. Hence, a total current of 3.14 A or a current



(a)



(b)

Fig. 3.33 **a** Schematic diagram of a Pierce electron gun with its basic electron trajectory. **b** Tested pulsed emission of an SDD cathode in the Pierce gun

density of 100 A/cm^2 should be obtained at an anode voltage of 30 kV if the cathode operates in full space charge limited regime. The J-U plots tested in above Pierce electron gun structure, operating in pulse mode with a pulse width of 3 μs and a repetition frequency of 5 Hz, are shown in Fig. 3.33b. At temperatures of 950 °C$_b$ and 1000 °C$_b$, the two J-U plots are almost overlapping and nearly identical current densities of 110 A/cm^2 and 115 A/cm^2, a bit higher than the expected value of 100 A/cm^2, are reached at the nominal anode voltage of 30 kV. Given the possibly of slight errors in practical perveance, the test results distinctly demonstrate that a full space charge limited current density of 100 A/cm^2 has been reliably achieved at brightness temperature 950 °C$_b$ or true temperature 1015 °C.

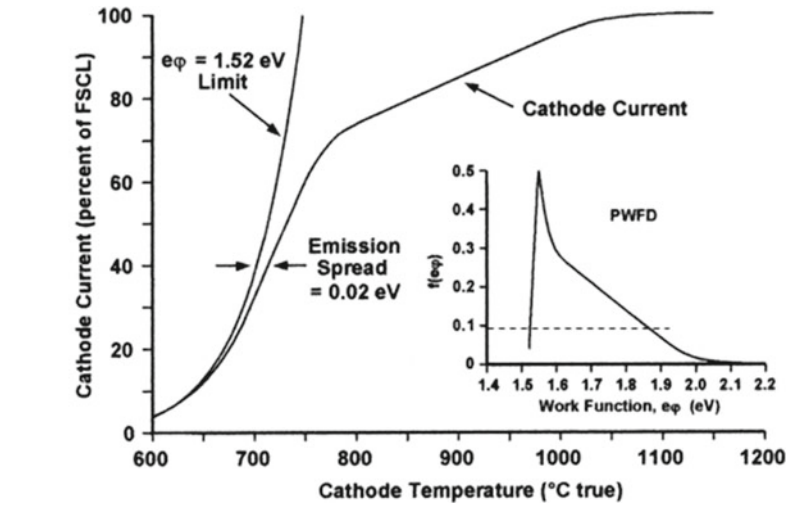
As pointed out by Gilmour [48], there is a practical limit to the current density that can be drawn from a cathode when applying in linear beam VEDs, regardless of its emission capability. This upper limit results from the electrical breakdown between cathode or focus electrode and anode in vacuum and appears to be about 100 A/cm^2 for DC currents.

Well-conducting cathodes with smooth surfaces will be a benefit for reducing the risk of local field enhancement. Hence, SDD cathodes seem most suitable for these applications.

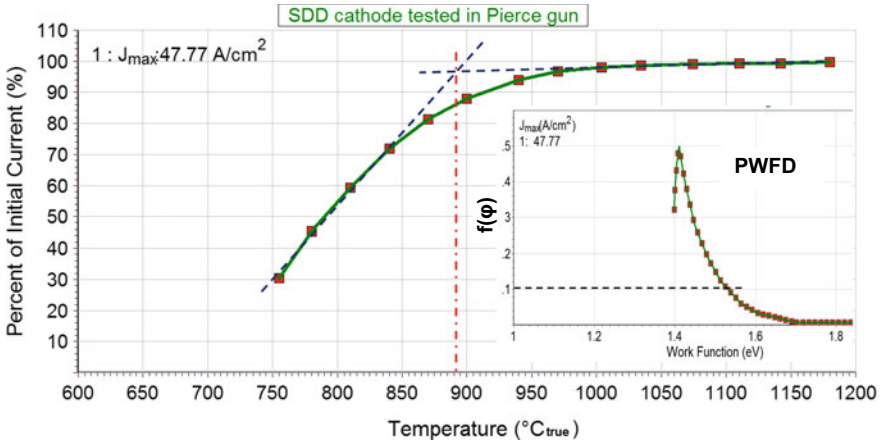
(c) Emission Uniformity

Emission nonuniformity is one of the major problems for Ba scandate cathodes. An evidence of the nonuniformity is that blurry knees on roll-off or Miram curves together with a broad PWFD appear in the early Ba scandate cathodes. Figure 3.34a shows a Miram curve and the related PWFD plot taken from an impregnated scandate cathode [48, 94]. It takes a wide range of over 300 °C, from ~750 °C to ~1100 °C, for the emission to transit from TL to SCL region. The work function of f(ϕ) $\geq$ 0.1 spreads over 0.34 eV, from 1.52 eV to 1.86 eV. This is an indication of the emission nonuniformity of the early Ba scandate cathode. With the recently developed doping technologies described above, the emission uniformity has been improved greatly. Sharp knees have been repeatedly observed on the SDD cathodes by L–L doping processes [93, 95] after a full activation. A normalized roll-off curve (Miram curve) measured in a Pierce gun structure for initial current densities of 48 A/cm^2 is exhibited in Fig. 3.34b. While the knee temperature is as low as about 880 °C true temperature, much narrower transition ranges of about 150 °C with flat SCL region appear in the curve. The peak work function is about 1.41 eV with a PWFD of only 0.14 eV, from 1.40 eV to 1.54 eV above f(ϕ) of 0.1.

It was once supposed that a Ba scandate cathode provides poor beam quality when used in an electron gun due to emission nonuniformity [8]. To inspect the actual situation, an SDD cathode with a cylindrical diameter of 2.88 mm and an emitting surface curvature radius of 3.26 mm was tested in an electron gun with perveance of 1.0–1.1 μp. After an SCL current density of 80 A/cm^2 was reached at a rated voltage of 30 kV [93], beam profiles were measured by a 3D movement apparatus at positions of 2–8 mm from the exit of the anode aperture along the beam



(a)



(b)

Fig. 3.34 Miram curves and PWFD plots of **a** an early scandate cathode. The graph is based on Fig. 5.33, [48]. **b** An SDD cathode

axis. Current distributions at related positions were also simulated by the EGUN 2 code. When an anode voltage of 10 kV is applied and at the position of 2 mm beyond the anode exit, indicated in Fig. 3.35a, the simulated current distribution is well coincident with both the tested cross-section of the electron beam and the 2D current distribution profile measured from the above electron gun at the same position, illustrated in upper, middle, and lower of Fig. 3.35b, respectively.

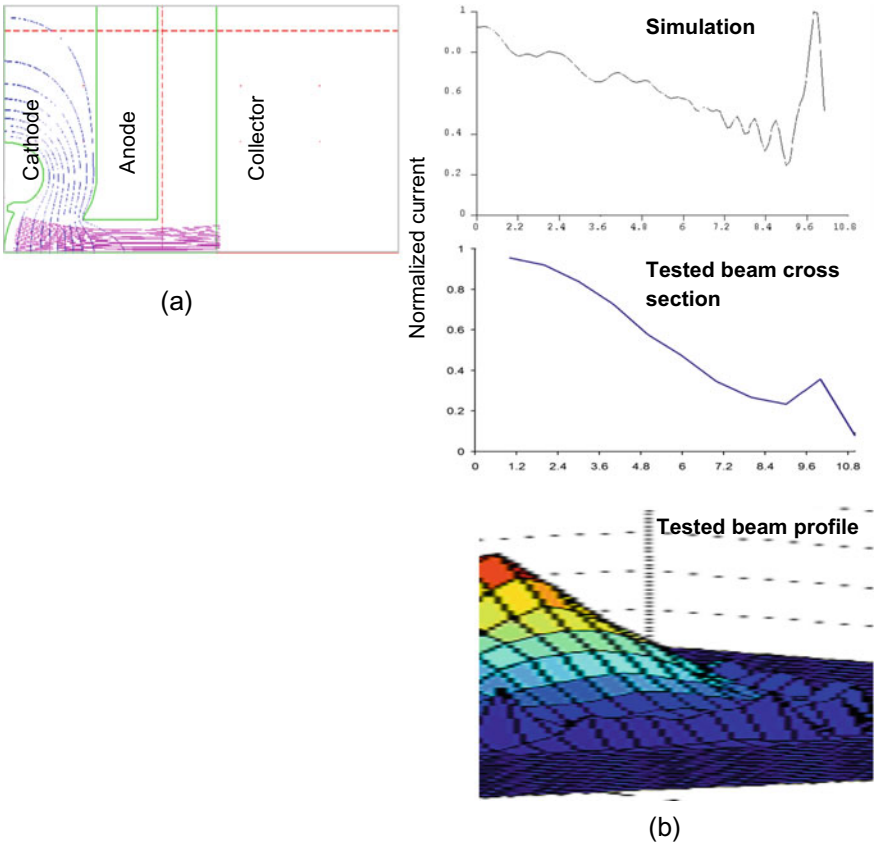


Fig. 3.35 Beam analysis in an electron gun with SDD cathode. **a** Simulated electron trajectory when collector sets at 2 mm beyond the anode exit. **b** Comparison of the simulated current distribution at collecting position (upper) with the tested cross-section of the electron beam (middle) and beam profile (lower) at the same condition. The graph is based on Fig. 8 of [93] with additional data

Since the simulation is based on uniform emission, the results prove that beam quality in an electron gun with this kind of cathode is acceptable.

The above investigations clarified that the nonuniform emission in early Ba scandate cathodes can be distinctly ameliorated in SDD cathodes with the nanosized-scandia doping technologies, optimal control of tungsten grain sizes, and proper activation procedures.

Deriving from Miram curves of SDD cathodes, peak work functions of about 1.40–1.45 eV can be established.

With the improvement for Ba Scandate cathode, the well-known "Quality of Thermionic Cathodes 'Best of Class' PWFD" [89, 96] is updated and given in Fig. 3.36.

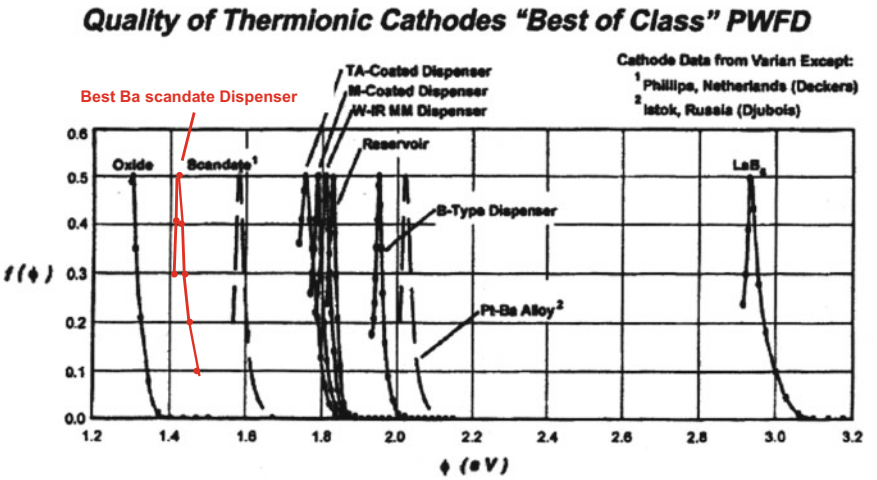


Fig. 3.36 Updated thermionic cathodes "Best of Class" PWFD, see [89, 96]

(d) Life Evaluation

Lifetimes for SDD cathodes with different doping procedures have been evaluated in both diode and Pierce-type electron gun configurations at normal operating temperatures.

In a diode with a water-cooled copper anode, a sol–gel doping SDD cathode had been operated with a continuous DC load of 2 A/cm^2 at 950 °C$_b$ while pulsed critical SCL current density at 850 °C$_b$ was measured periodically to monitor the change of the emission capability with time. The critical current density, shown in Fig. 3.37a, increased in the first two thousand hours of life and then remained stable for more than 10,000 h [10]. The stability of the current density at a critical level, at least higher than 35 A/cm^2 at 850 °C$_b$ or 1178 K, reveals that the high emission capability of SDD cathode can be well maintained for more than ten thousand hours.

With DC load of 30–40 A/cm^2, which is the highest DC load on life test for all kinds of thermionic cathodes, the SDD cathode has been working smoothly at 970 °C$_b$ for up to 5000 h, as shown in Fig. 3.37b [92]. Both the DC current load at fixed anode voltage and the critical SCL DC current densities have improved during the first ~1000 h.

In summary, by using doping processes to mix well-distributed nanosized scandium oxide into tungsten powder of proper sizes, the scandia doped dispenser cathodes provide uniform emission with critical space charge limited current density of greater than 100 A/cm^2 at 950 °C$_b$, Mo-brightness temperature, or 1015 °C true temperature. There is little difference between pulsed and DC emission. Lifetimes of $\geq$10,000 h can be achieved.

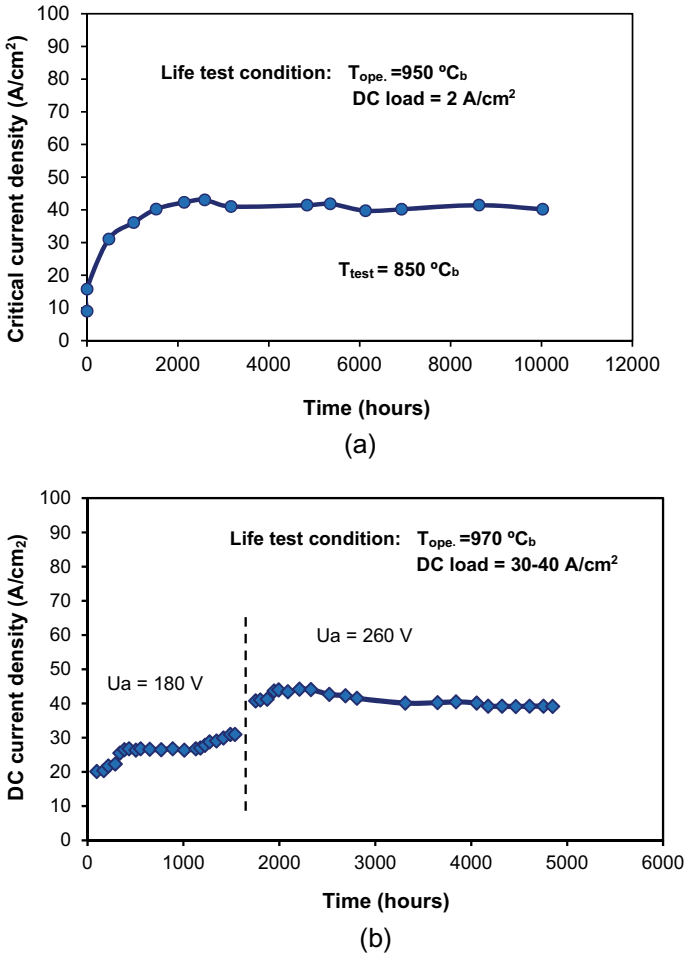


Fig. 3.37 Emission performance of SDD cathodes during life test. **a** Change of critical pulse SCL emission at 850 °C_b, checked intermittently during life test carried out at 950 °C_b with a continuous DC load of 2 A/cm^2. **b** Change of DC emission during life test at 970 °C_b with DC loads of 30–40 A/cm^2

3.3.4 Generation of Miniature Electron Beams

The advanced VEDs like millimeter wave and Terahertz vacuum electron devices demand high cathode emission capability because the required beam current density is proportional to the operating frequency squared. With the need for miniaturization, a variety of very serious challenges are posed for cathode performance. The challenges are to confine the electron beam to an increasingly small cross-section and generate and realize in the interaction space the extremely high current densities [3, 97].

Miniature electron beams can be obtained from normal cylindrical cathodes by means of specially designed compression systems [98]. However, direct generation of the miniature beams from cathodes has the advantages of simplifying the focus optics which makes the devices more compact and facilitates the formation of high-quality, high-aspect-ratio electron beams. To form the required beams, the cathodes itself must be able to provide emission current densities in the order of 100 A/cm^2 at reasonable operating temperatures. The SDD cathode described above, thus, becomes one of the most promising candidates for this application.

Before the 3D print technology could be applied to thermionic cathode manufacture, one of the approaches to obtain a miniature beam from a cathode is to cover the cylindrical cathode with an anti-emission mask, leaving a defined opening for electron emission. In this way, standard construction methods can be employed without much modification for fabricating the underlying thermionic cathode. The anti-emission mask can be a solid metal mask made from refractory metals or a thin film of emission-suppressing materials such as Ti, Hf, and Zr. An opening in the required shape and dimension is constructed on the mask by micromachining technologies [88].

A rectangular sheet beam realized by a masked cathode assembly and measured in an emission distribution testing system is shown in Fig. 3.38a. The beam has an aspect ratio of 6:1 and a current density of over 50 A/cm^2 at 950 °C$_b$ [99]. Design of a sheet-beam electron gun schematized in Fig. 3.38b is based on the above cathode. It uses one-dimensional three-fold beam cross-sectional area compression to meet the specific requirement of a beam to be formed of height 30 μm and width 600 μm at the beam–waist position with over 100 A/cm^2 uniform current density and 0.068 π-mm-mrad emittance. The electron gun is typically for 0.5 THz devices [100].

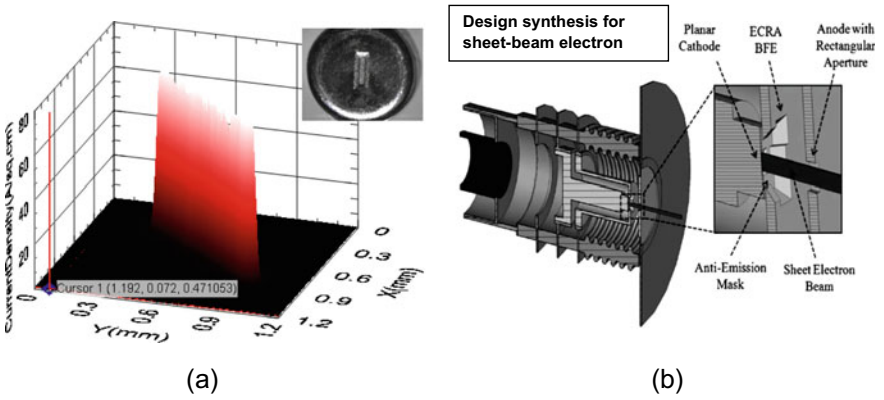


Fig. 3.38 High current density sheet-beam generated directly from an SDD cathode. **a** Beam profile of 600 μm × 100 μm with over 50 A/cm^2 SCL current density. Photo of cathode assembly is shown in the inset. **b** Schematic of half-sectioned sheet-beam electron gun. The figure is based on Fig. 1 of [100]

3.3.5 Fundamentals of Ba Scandate Dispenser Cathodes

3.3.5.1 A Brief Review of Previous Studies on Ba Scandate Dispenser Cathode Fundamentals

R. E. Thomas and his coworkers stated incisively in the later 90s [8] that "with the advent of modern surface analytical capabilities that enable one to better understand the physics and chemistry of emitter surfaces significant improvements have been made in thermionic electron sources". Based mainly on surface studies during the 70s–80s, the fundamentals of Ba dispenser cathodes have been investigated intensively and a model of Ba–O dipole monolayer on W has been deduced and widely accepted [101–103]. However, the emission mechanism of the Ba scandate cathodes is still controversial and study results tend to differ from one another.

Before a further discussion on the matter is made, the proposed emission models for Ba scandate cathode are reviewed in brief.

The different types of Ba scandate cathodes introduced in part Sect. 3.1 of this chapter can eventually be divided into three groups by the ways of adding scandium oxide: adding into the tungsten matrix or powder [10, 14, 16], adding into the impregnant [13, 79, 80], and coating on a Ba dispenser cathode as thin films [9, 17, 21, 22, 35]. Although these types of Ba scandate cathodes have different structures, they feature similar emission characteristics, i.e., (i) distinctly higher emission capability than that of Ba dispenser cathodes, though in varying degrees; (ii) obvious non-saturated emission character, though the physical distribution uniformity of scandium oxide has been increased from type to type through persistent improvements on processing technologies [95]. Therefore, it is believed that the emission mechanism for various types of Ba scandate cathode should be basically identical.

It must be pointed out that, in most of these Ba scandate cathodes, the traditional tungsten powders or matrices which are composed of Ba dispenser cathodes and even Ba dispenser cathodes themselves are constituted as part of the Ba scandate cathodes. The facts above argue strongly on a proposal presented by some researchers [25], in which it is suggested that particular crystallographic facets, like {112} facets, of the W grains may be critical to the enhanced emission performance of Ba scandate cathodes. Though facets of tungsten grains do impact the emission of thermionic cathodes, it is unlikely that this is the reason for emission enhancement of Ba scandate cathodes over Ba dispenser cathodes.

Several emission models have been suggested in the history of the development of Ba scandate cathode; they are redisplayed in Fig. 3.39a–c: a BaO layer covering on bulk Sc_2O_3 assumed by Forman [104], a Ba–Sc–O monolayer by Yamamoto [105], and a semiconductor layer by Raju and Maloney [50]. Among them special attention has been fixed on the latter two. One is the Ba–Sc–O monolayer model, which was first proposed by S. Yamamoto, based on the observation that a surface layer composed of Ba, Sc, and O on a mixed matrix Ba scandate cathode which is easily removed by Ar ion sputtering [14]. Lately, when Yamamoto investigated a top-layer Ba scandate cathode sputter-coated with a W + Sc_2O_3 thin film to improve

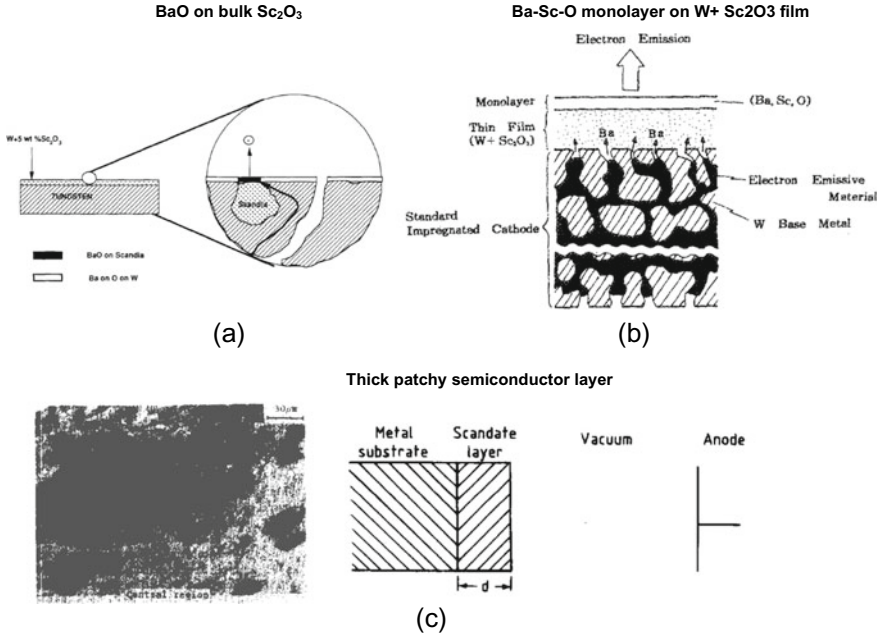


Fig. 3.39 Proposed emission models for Ba scandate cathode, based on **a** BaO on bulk Sc_2O_3, Fig. 16 of [104]; **b** Ba–Sc–O monolayer on W-Sc_2O_3 film, Fig. 1 of [105]; **c** Thick patchy semiconductor layer: Left: SEM image of the layer, Fig. 12a of [50]; Right: proposed sandate cathode in diode, Fig. 5a of [50]

the mixed matrix one, he pointed that electron emission enhancement is achieved when a surface layer consisting of Ba, Sc, and O is formed with uniform thicknesses of a monolayer on top of the W-Sc_2O_3 coating film of 50–400 nm, see Fig. 3.39b, and the work function reaches its minimum when an ordered phase of Ba, Sc, and O is formed out of a disordered phase which contains extra amounts of Ba and Sc [105]. It should be noted that the suggested monolayer above, in fact, is located on a thick Sc-containing film but not on the tungsten substrate as the conventional monolayer of Ba-O dipole on W for Ba dispenser cathode. No more details about the layer were reported. The other model is the semiconductor model presented by Raju and Maloney [50]. In their model, the cathode surface is composed of a substantially thick patchy semiconductor layer of some scandium compounds like scandium tungstate, shown in Fig. 3.39c.

Shih and Yater highlighted in their study [106] that the thick layer of semiconducting material required by Raju and Maloney's model is possible to form from the simultaneous presence of BaO, Sc_2O_3 on W. In their study, Temperature-Programmed Desorption (TPD) and Auger electron spectroscopy (AES) were used to characterize Ba and BaO on a W substrate and on a Sc_2O_3-coated W substrate to simulate the surface layer of the conventional Ba dispenser cathode and the Ba scandate cathode.

They found that a partial layer of BaO forms on W, which is similar to the Ba-O surface layer that forms on a Ba dispenser cathode. However on the Sc_2O_3-coated W substrate, there are no longer partial surface layers of BaO coverage. Instead, a stable BaO-containing compound forms and continues to accumulate to form a bulk-like layer. In addition, a high heating temperature could cause a tight binding BaO partial layer to form on the coated surface. They suggested that the layer may serve to reduce the electron affinity of the compound formed underneath, causing the Ba scandate cathode to behave like an oxide cathode.

The thick layer or multilayer proposal has been further supported by A. P. Makarov in a simulation study on jointly adsorbed atoms of barium, scandium, and oxygen on the facet (100) of a tungsten crystal [107]. The study discovered that the work function of a monolayer complex Ba–Sc–O consisting of 1 barium monolayer and 0.5 scandium monolayer with an optimal doze of adsorbed oxygen on the surface of the facet (100) W amounts to 2.1 eV, while the complex Ba–Sc–O consisting of 1 Ba monolayer and 1 Sc monolayer even comes to 2.15 eV, both are higher than the work function of 1.9–2.1 eV for Ba-O on W (100) systems. On the other hand, the work function of a multilayer film of jointly adsorbed atoms of Ba, Sc, and O on the facet (100) of tungsten crystal, consisting of 2–3 Ba monolayers and 2–3 Sc monolayers with an optimal doze of adsorbed oxygen when heated at temperature 1400–1500 K and then activated by monolayer of barium atoms, drops down to 1.70–1.75 eV, which is a bit nearer to the work function value of Ba scandate cathodes.

Based upon the above proposals, investigations have been made for understanding the fundamentals of SDD cathodes [95, 108, 109] and are summarized below.

3.3.5.2 Behavior of Surface Active Layer on Nanosized-Scandia Doped Dispenser Cathodes

To characterize the outer surface of the cathode as to explain its electron emission performance, techniques probing the first atomic layers are most adequate.

The behavior of the surface active layer on SDD cathodes has, then, been diagnosed in detail by modern surface analysis techniques, including Auger Electron Spectroscopy (AES) and X-Ray Photoelectron Spectroscopy (XPS). Atomic Force Microscopy (AFM) as a direct surface nanostructure detecting technology has also been used for topography observation.

The importance of in situ analysis has been emphasized since the surface analysis was applied in thermionic cathode research [101–103]. Thus, in situ surface analysis has been carried out during heating, activation, operation, and life test of the SDD cathodes.

(a) Surface Layer Composition

The in situ XPS studies, in which a large or even whole area of the cathode surface can be analyzed by X-ray excitation, have first been used for surface composition diagnosis. Survey spectra taken from the activated cathodes are shown in Fig. 3.40a.

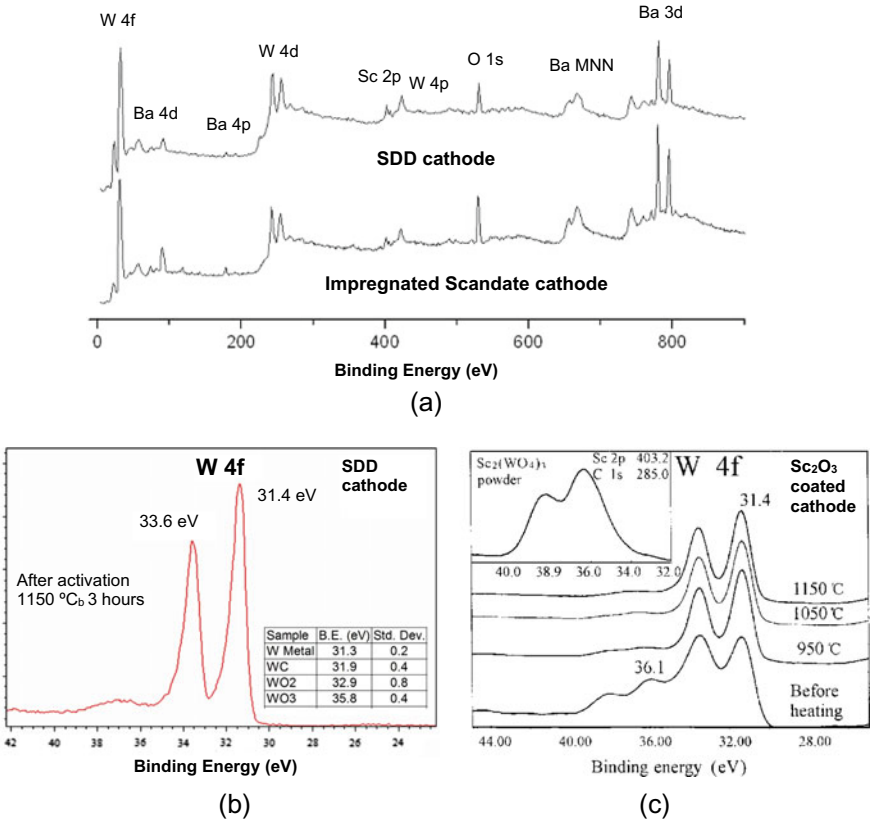


Fig. 3.40 Surface composition of a fully activated Ba scandate cathodes analyzed by in situ XPS analysis. **a** Survey spectra of SDD cathode and impregnated scandate cathode. **b** W4f spectrum of SDD cathode after activation (inset: list of BE of W4f$_{7/2}$). **c** W4f spectra of a Sc_2O_3 coated thin film cathode (insert: W4f of $Sc_2W_3O_{12}$ powder), based on Fig. 6 of [21]

Similar to that of Ba dispenser cathode, Al and Ca disappear from the surface after activation though they are the components of the impregnant [110]. There is no element other than Ba, Sc, O, and W to appear on the surface of a fully activated SDD cathode. This is also true for other Ba scandate cathode like impregnated scandate cathode, though the Sc signal is relatively weaker.

Moreover, it is noticed that the W4f$_{7/2}$ XPS peak is with binding energy (BE) of 31.4 eV for an SDD cathode after activation and it is the same for laser-deposited thin film scandate cathode [21], shown in Fig. 3.40b, c, respectively. This peak position corresponds to metallic W, denoted in the inserted list of Fig. 3.40b and much lower than the BE of a tungstate illustrated in the inset of Fig. 3.40c. The results ruled out the hypotheses that scandium tungstate such as Sc_6WO_{12} [50] or barium scandium tungstate such as $Ba_3Sc_2WO_9$ [77, 78] constitutes the surface layer of Ba scandate cathodes.

(b) Features of the Surface Layer

It is distinctly important to figure out the presentation of the suggested monolayer or the possibly thick layer described in the previous researches on the surface of Ba scandate cathodes.

In the studies of Ba dispenser cathode, the surface model was investigated through synthesizing the surface conditions existing on an impregnated tungsten cathode by depositing barium or oxidized barium on a tungsten surface with depth varied from many monolayers to partial monolayer coverage. In situ Auger analysis and effective work functions measurement were used to compare synthesized surface with the activated impregnated cathode surface [101]. Auger peak-to-peak height (APPH) ratios of surface-covering constituents to the substrate are used for surface layer thickness and coverage estimation. Based upon the above investigations, it reached a consensus that for a conventional Ba impregnated tungsten cathode there exists a monolayer or partial monolayer of barium and oxygen on a tungsten surface [101, 103, 111]. Consequently, if the monolayer assumption is suitable for Ba scandate cathodes, then similar surface features must appear. Table 3.5 summarizes the APPH ratios of surface active elements Ba, O (and Sc) to substrate W, which are taken from AES spectra of modeling systems and real cathodes. In the table, Auger peaks of Ba (584 eV), O (512 eV), Sc (334 eV), and W (179 eV) are chosen for analysis as they have similar escape depths.

Besides, Auger spectra taken from the synthesized barium and oxidized barium monolayer on a tungsten surface, an activated impregnated tungsten cathode, and an SDD cathode are compared in Fig. 3.41a, upper, middle, and lower.

From the data in the table and the figures, the APPH ratios of Ba/W of about 0.2–0.4 and O/W of 0.5–1.0 are for the Ba impregnated cathodes, agreeing with the data for monolayer. In contrast, the Auger spectra taken from the activated SDD cathodes exhibit a completely different fashion. The APPH ratio of Ba to W goes up to 1.6, almost five times higher than that of Ba dispenser cathodes. Similarly, the

Table 3.5 Auger Peak-to-Peak Height ratios for different surface configurations containing Ba, O and in some cases Sc

Surface configuration	APPH ratios			References
	Ba/W	O/W	Sc/W	
One monolayer of oxidized Ba	0.37	1.3		Figure 5 in [101]
Activated impregnated Ba tungsten cathode	0.21	0.5		Figure 2 in [112]
S-type cathode at lowest Φ	0.30	1.02		Figure 31 in [87]
Monolayer of Sc on W (100) after oxidation		1.14	0.85	Figure 6 in [113]
Ba-W dispenser cathode after activation	0.31	0.72		Under the same test condition as for SDD cathode
SDD cathode after full activation	1.6	5.2	3.3	Average of in situ analyses

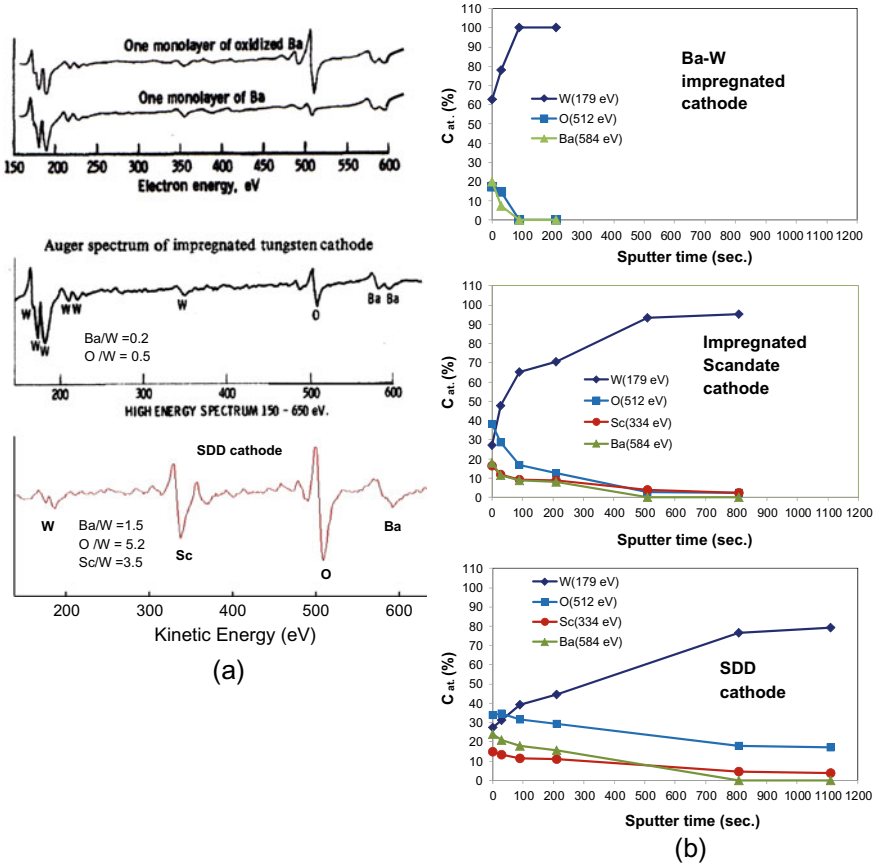


Fig. 3.41 In situ AES analysis on surface layers of thermionic cathodes. **a** Auger spectra from simulated surface layers and activated cathodes tested at operation states. Upper: Auger spectra of barium and oxidized barium monolayer on tungsten surface, based on Fig. 5 of [101]; Middle: Auger spectrum of an activated impregnated tungsten cathode, based on Fig. 2 of [112]; Bottom: Auger spectrum of an activated SDD cathode. **b** AES depth profiles of cathodes after activation, taken at room temperature. Sputtering condition: Ar^+ ions of 4 keV, 300 nA. Upper: Ba-W impregnated dispenser cathode; Middle: Impregnated scandate cathode; Bottom: SDD cathode

Sc/W of SDD cathode is about 3–4 times of that from a monolayer of oxidized Sc [113], while O/W goes even higher.

The data imply that the layer of Ba, Sc, and O on the surface of SDD cathode is far beyond a monolayer.

The above conclusion has been further verified through the in situ Auger depth profiles investigation, in a VG MICROLAB MK II surface analysis instrument, by comparing Ba-W impregnated dispenser cathode, impregnated scandate cathode, and SDD cathode. The profiles are shown in Fig. 3.41b, upper, middle, and bottom,

respectively. All the cathodes were fully activated in situ, then sputtered step by step by Ar^+ ions of 4 keV, 300 nA and analyzed at room temperature in an area of 100 μm × 100 μm to get average signal of the elements [83]. By referring to VG's technical documents on estimating sputter rates for VG ion gun, a sputter rate of about 6–8 nm/min has been estimated.

While stoichiometric Ba and O were removed simultaneously in around one minute for Ba-W impregnated cathode, a high concentration of Ba with O and Sc is slowly reduced in a period of several times longer for SDD cathode than the situation for Ba-W impregnated cathode. Moreover, O and Sc stay with little change at low levels to the end of the test.

The observations strongly suggest that Ba and O are bound with Sc, tighter than that with W, to form a Ba–Sc–O layer much thicker than a monolayer on Ba-W impregnated cathode.

A similar situation was also found for other Ba scandate cathode, such as the impregnated scandate cathode as shown in the middle of Fig. 3.41b. A Ba–Sc–O layer can be observed there, but it is thinner and with less Ba compared to that on SDD cathode.

On the other hand, the surface feature of SDD cathode significantly differs from a thick patchy layer proposed by Raju and Maloney (Fig. 3.39c).

The Ba, Sc, and O actually distribute uniformly over a single W grain and on different W grains. This has been confirmed by repeatable observation results during a decade of SDD cathode development [95, 114]. As an example, a set of AES spectra measured by an AES system (PHI700, ULVAC-PHI) with spatial resolution of 40 nm at nine different points on surface of an activated SDD cathode is presented in Fig. 3.42a. It clearly indicates the nearly identical surface composition at each point.

Furthermore, it is also demonstrated that the Ba, Sc, and O are evenly distributed along certain depth on W grains at different sites. As checked by the same AES instrument above and shown in Fig. 3.42b, where the surface active elements and substrate W versus sputter time at three points marked on the SEM image are displayed, the Ba, Sc, and O are distributed in similar depths over the whole substrate of W grain. By referring to the sputter rates calibrated with thermally oxidized SiO_2/Si, the distribution depths of up to tens of nanometers are reckoned.

The special surface feature of SDD cathode should bring visible spectacle on its topography. The obvious disparity of surface topographies between Ba dispenser cathode and SDD cathode has been observed by Atomic Force Microscopy (AFM), illustrated in Fig. 3.43. Both Ba-W impregnated cathode and SDD cathode used for diagnosis were fully activated and aged to enable stable surface structures to be established. As expected, the phase image of Ba-W cathode on the left side of Fig. 3.43a exhibits a smoothly metal-like surface. On the other hand, significantly different from the metal-like surface, a layer-like structure appears on W grain of the activated SDD cathode, shown in the middle of Fig. 3.43a. Since this structure does not present on the surface of the scandia doped matrix, as indicated on the right

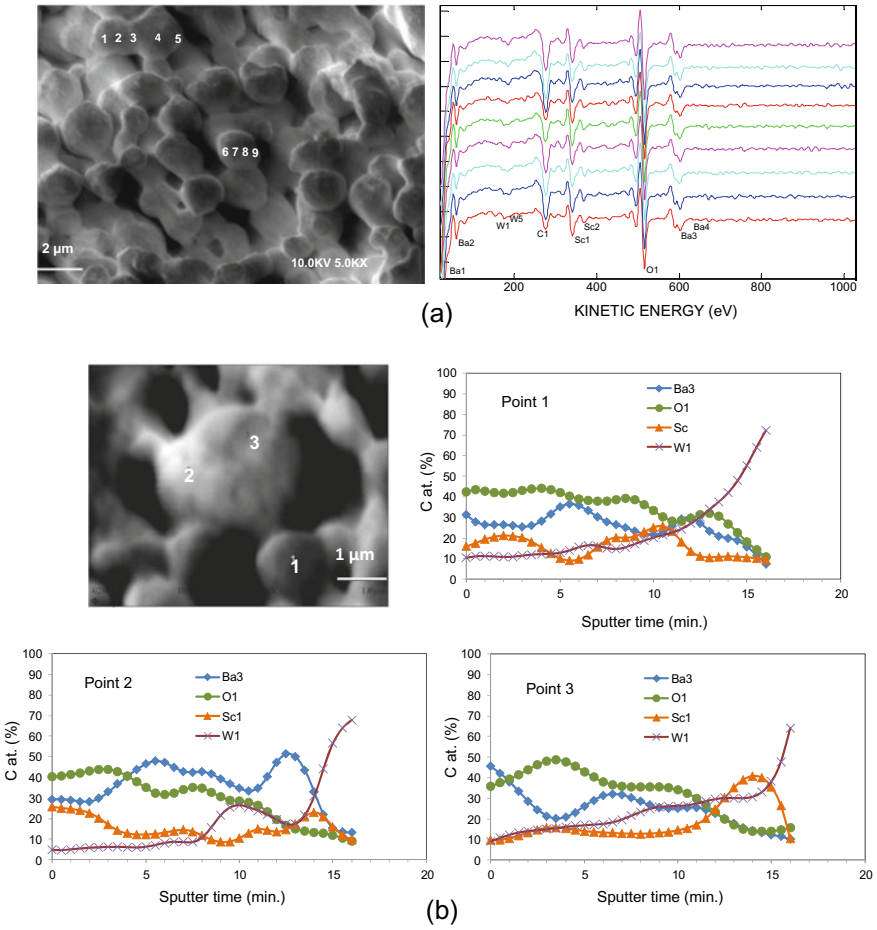


Fig. 3.42 Surface and depth distributions of Ba, Sc, and O on surface of activated SDD cathodes. **a** AES spectra (right) from nine points on the SEM image (left). **b** Depth profiles of three points on SEM image. Sputtering rate: 2 nm/min for (1–9) min and 13 nm/min. For the rest, calibrated for thermally oxidized SiO_2/Si

side of Fig. 3.43a, it is proved that the layer is formed after SDD cathode activation. From the height sensor image of an activated SDD cathode, the layer has a thickness essentially in agreement with the AES depth analysis results.

In summary, Ba, Sc, and O on the surface of SDD cathodes present neither as a monolayer as in the case of Ba dispenser cathodes, nor as a thick patchy layer. Rather, a multilayer of Sc with Ba and O, which are of higher amounts than that on Ba dispenser cathode surface, is consistent with the surface feature of a real SDD cathode. The layer is uniformly distributed on the surface of SDD cathodes in a thickness of about tens of nanometers.

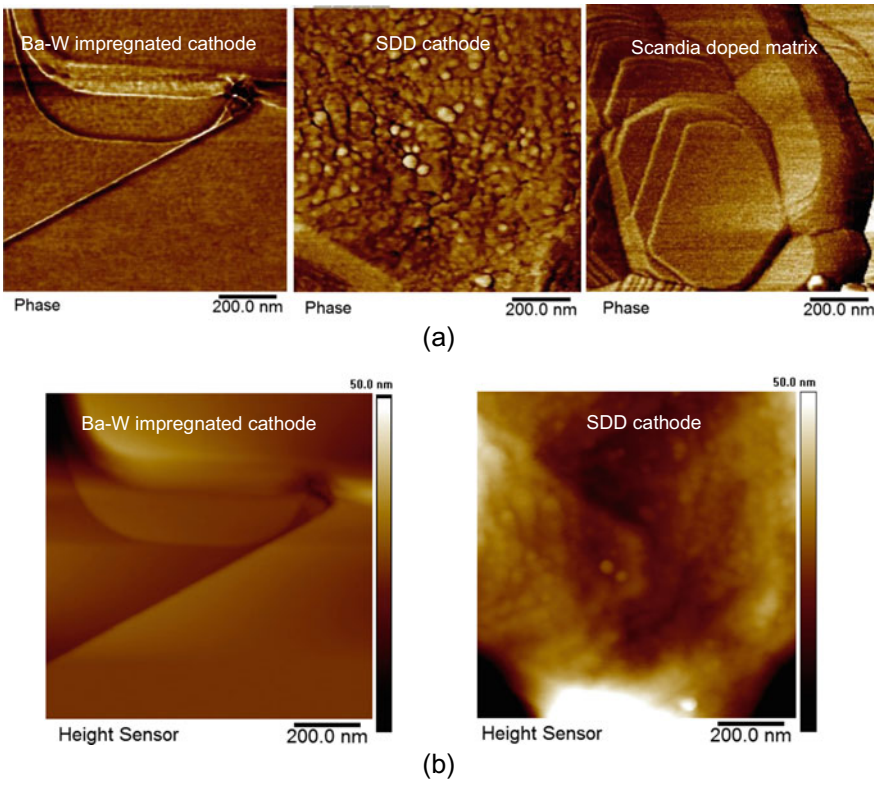


Fig. 3.43 Comparison of surface topographies of fully activated Ba-W impregnated cathode, SDD cathode and Scandia doped matrix. **a** AFM phase images. Left: activated Ba-W impregnated cathode; Middle: activated SDD cathode; Right: Scandia doped matrix. **b** AFM Height sensor images. Left: activated Ba-W cathode; Right: activated SDD cathode

(c) Correlation Between Surface Layer Parameters and Emission of SDD Cathode

A strong correlation between surface layer parameters and emission properties of SDD cathodes has been found from the summarized data of emission measurement and in situ analysis in cathode activation, operation, air exposure, and the life tests carried out in AES system.

The thickness of the layer and the atomic ratios of surface active elements are closely correlated with the emission properties of the SDD cathode [108]. A summary of the relationship between surface features and the emission is given in Table 3.6. In the table, the APPH ratio of Ba/W reflects the layer thickness, and the constituents of the active components are expressed by atomic concentration ratios of Ba:Sc and O:Sc. The layer normally has optimum atomic concentration ratios of Ba:Sc:O of about (1.5–2):1:(2–3) at operation condition after the cathodes were fully activated [95]. Improper producing processes, insufficient activation, or overheating will cause

deviations from the optimum atomic ratios and/or layer thickness, resulting in unfavorable emission. Both thickness and the surface active elements ratios remain almost the same during life test at normal operating temperature of 1000 °C_b for about ten thousand hours [115], which are demonstrated in Fig. 3.44. But all the three ratios deviate from their optimum levels when working at an elevated temperature of over 1150 °C_b for thousand hours with a coincident degradation in emission.

It was thus concluded that the copious electron emission of Ba scandate cathodes is due to a Ba–Sc–O multilayer described above and the emission capability stays unchanged as long as the multilayer remains.

This multilayer structure minimizes the influences of substrate components on the work function of the cathode, which is the case for M-type cathodes. Hence, other additives to W substrate such as Re, Os, or Ru have limited contribution to emission capability of Ba scandate cathodes.

Table 3.6 Correlation of emission of SDD cathodes with surface behavior

Cathode status	Surface behavior				Emission J(A/cm^2) at 850 °C_b
	Atomic concentration ratios			APPH ratio	
	Ba	Sc	O	Ba/W	
After full activation Average range	1.5–2.0	1	2.0–3.0	1.4–1.9	30–40
Operating at 1000 °C_b 7000 h	2.0	1	2.0	1.4	Stable
Operating at 1150 °C_b 3000 h	6.1	1	7.8	0.3	Low

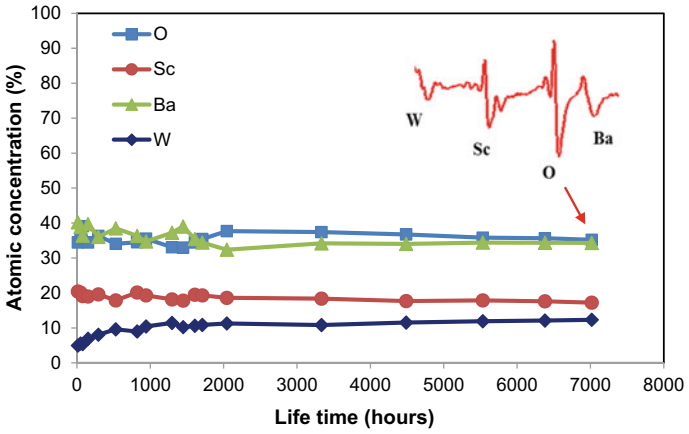


Fig. 3.44 Change of surface composition of an SDD cathode during operation at 1000 °C_b after activation, analyzed at 800 °C_b by in situ AES

3.3.5.3 Function of the Surface Layer

As a well-known fact, anomalous Schottky effect has been observed for all types of Ba scandate cathodes. It was once assumed that a patch work function is responsible for this non-saturated emission characteristic if a dipole monolayer emission model is adopted. However, as previously described, a uniform Ba–Sc–O multilayer of tens of nanometers in thickness actually exists on the surface of SDD cathodes and the emission uniformity is generally similar to that of Ba dispenser cathode. Therefore, a new interpretation is required. A. Shih has pointed out based on the thick Ba, Sc, O-containing compound layer they found in their study [106] that, if a semiconductor model is considered for Ba scandate cathodes, the emission should be basically uniform but not patchy. In the semiconductor model suggested by Maloney [50], the anomalous Schottky effect observed in the emission versus voltage measurements is thought to be the result of penetration of the external electric field into the surface layer and lowering the work function, causing a continuous increase in emission current with increased anode voltage, like that proposed by Wright for oxide-coated cathodes [116].

In a recent study, a special emission feature has been noticed in a PWFD-type analysis for SDD cathodes, and also for other kinds of Ba scandate cathodes. This feature is contrary to the typical behavior of Ba dispenser cathode, which is based on a Ba-O dipole emission model. The special feature is

1. When the actual Miram curves from Ba scandate cathodes are superimposed on a group of theoretical J_0-T curves derived only using the Richardson–Dushman equation, i.e.:

$$J_0 = 120\,T^2 \exp(-\phi/kT) \tag{3.6}$$

   to form PWFD plots, the peak work function (WF) values of the PWFD plots that result from Miram curves with different initial SCL current densities do not overlap as in the case of Ba dispenser cathodes [90].
2. The peak work function values decrease sequentially as the initial SCL current densities increase, so that the lowest peak WF always corresponds to the highest initial SCL current density. This is also the case when the Schottky effect is taken into account in determining the PWFD plots [95]. Because the initial SCL current densities are directly related to the applied electric field, the special feature implies that the lowest peak work function is subjected to the highest external electric field in the normal PWFD analysis. As an example, a comparison of PWFD plots from an SDD cathode and a Re-W base M cathode, tested in close-spaced diodes under the same conditions, is illustrated in Fig. 3.45.

To explain this special emission feature quantitatively, the influence of external electric fields on the emission properties when semiconductor models for thermionic emission are taken into account, which was proposed by Wright [116], has been

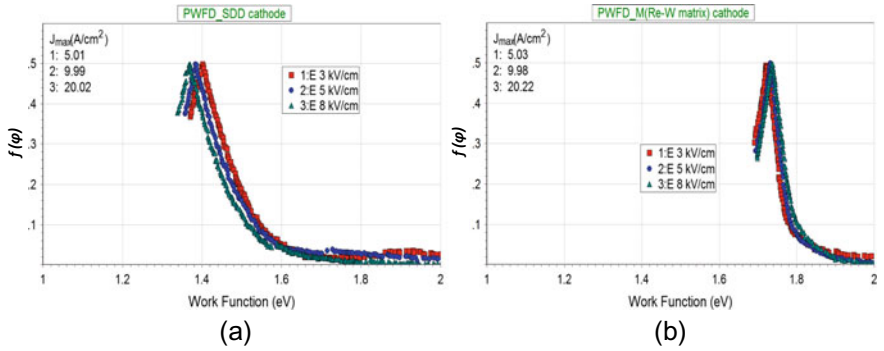


Fig. 3.45 PWFD plots at different initial SCL current densities, corresponding to different external electric fields, for **a** SDD cathode and **b** M (Re-W matrix) cathode. The peak work functions shift with the increase of the electric fields only for SDD cathode

introduced. According to Wright's model, the applied external field penetrates the surface of the semiconductor due to the low concentration of free electrons in its conduction band, forming a layer of space charge. The energy levels are then tilted in the semiconductor layer, producing a reduction $\delta\chi$ of the work function ϕ, so that

$$\phi' = \phi - \delta\chi \tag{3.7}$$

$\delta\chi$ is determined as

$$\delta\chi = 2kT\sinh^{-1}\frac{E}{4(2K\pi n_0 kT)^{1/2}} \tag{3.8}$$

where E is the electric field at the surface of the cathode produced by the applied anode potential, n_0 is the concentration of electrons in the conduction band of the semiconductor at T, and K is related to the dielectric constant of the semiconductor. By combining Schottky effect and above semiconductor modification, the Temperature Limited (TL) emission $J_{TL\text{-}S}$ is then described as

$$J_{TL-S} = J_0 e^{\frac{\delta\chi}{kT}} e^{\frac{4.4\sqrt{E}}{T}} \tag{3.9}$$

While the tested Miram curves of an SDD cathode with different initial SCL current densities, which are related to different anode voltages at a fixed anode–cathode distance, i.e., electric fields E, are superimposed to (i) a family of J_0-T curves calculated according to the Richardson–Dushman equation and (ii) the $J_{TL-S} - T$ curves calculated by taking both Wright's semiconductor model and the Schottky effect into account, two groups of PWFD plots are derived and illustrated in Fig. 3.46a, b, respectively.

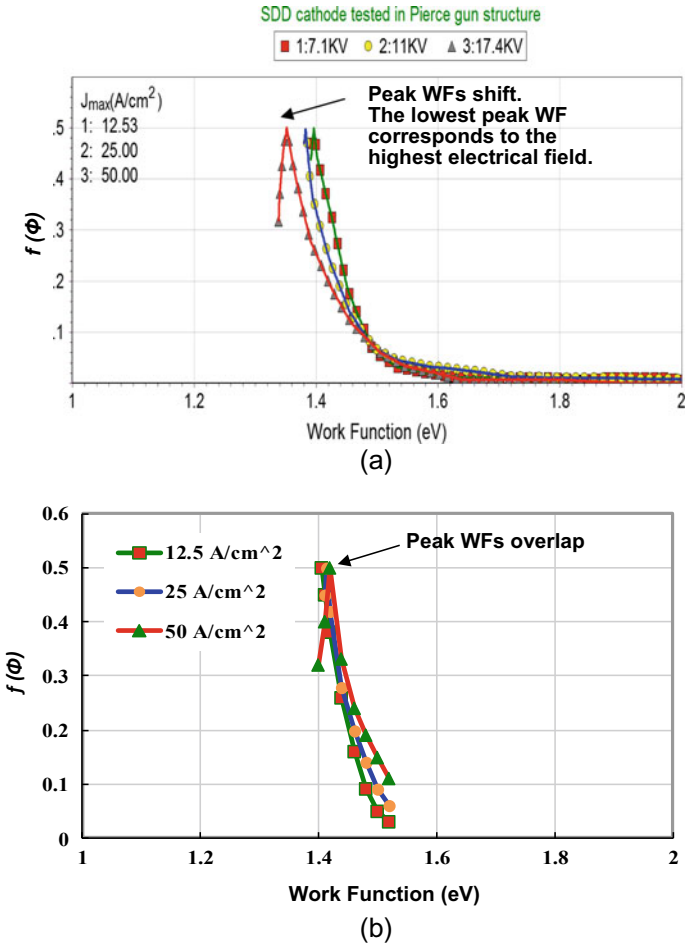


Fig. 3.46 PWFD curves of an SDD cathode resulted from superimposing of tested Miram curves at different initial SCL current densities (external electric fields) with **a** a family of J_0-T curves calculated according to the Richardson–Dushman equation. **b** The $J_{TL\text{-}S}$ -T curves calculated by taking both the Schottky effect and the Wright's semiconductor model into account

The peak work function values for the three initial SCL current densities shift obviously in the case of (i) and the lowest WF is related to the highest anode voltage. However, in the case (ii), where the influence of extra electric fields on emission from the semiconductor model has been involved, the peak WFs tend to normally overlap with an average WF of around 1.41 eV. The analyses results denote that the emission from SDD cathode (and other Ba scandate cathodes) is more related to a semiconductor structure rather than a dipole monolayer model.

This proposal could be substantiated after applying the Longo–Vaughan (L–V) relations of thermionic cathodes to the SDD cathodes. In L–V relation, the cathode current density J can be expressed with the following equation [117],

$$1/J^{\alpha} = 1/J_{SCL}^{\alpha} + 1/J_{TL}^{\alpha} \tag{3.10}$$

where J_{SCL} is given by Child-Langmuir's law, J_{TL} is determined by Richardson's equation with Schottky correction, and α is called the shape factor. By fitting the tested J with the equation, the work function can be estimated. This approach has been successfully applied for Ba dispenser cathode performance evaluation [118].

However, similar to the case of PWFD analysis, when fitting the experimental data from SDD cathodes to the equation, the obtained WFs from different initial SCL current densities are inconsistent with each other and the lowest WF-value is always corresponding to the highest initial SCL loading.

Nevertheless, by taking into account Wright's modification on the emission and choosing the concentration of electrons, n_0, from 10^{14} to 10^{16} electrons per cm^3, it is found that WFs are almost identical with values of around 1.41 eV if n_0 is in the region of 10^{15} electrons per cm^3, indicated in Fig. 3.47.

From the above analyses, it can be concluded that the extraordinary emission properties of Ba scandate cathodes are mainly attributed to the influence of external electric field on the semiconductor layer on the cathode surface rather than the patch effect. In other words, the Ba–Sc–O multilayer on the surface of Ba scandate cathode functions as a semiconductor layer.

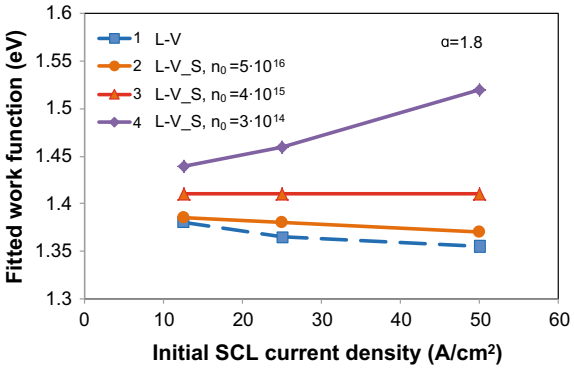


Fig. 3.47 Variation trend of the fitted work functions related to initial SCL current densities. The dashed line noted as "L–V" indicates the resultant work functions by fitting the tested data with Longo–Vaughan relation. The solid lines noted as "L–V_S" indicate the fitted work functions when Wright's modification for semiconductor is included and with different concentrations of electrons per cm^3 (n_0) in conduction band of the semiconductor layer. A shape factor of 1.8 is adopted for fitting

3.3.5.4 Formation of the Surface Layer

Owing to the essentiality of the surface layer to emission performance of Ba scandate cathodes, a deep understanding of the formation process of the layer is propitious for operation mechanism investigation and manufacturing technology development.

Though the Ba scandate cathode is basically constituted by adding scandium oxide to Ba dispenser cathode, whatever to matrix, impregnants, or on top of the cathode, the original added scandium oxide itself does not, in fact, join the surface layer directly.

In the case of SDD cathode, as pointed out in cathode structures description, the scandium oxide dispersed in the matrix no longer exists after impregnating with Ba, Ca aluminate. Compounds such as scandium tungstates, barium scandates, Ba, Sc-tungstates and Ba, Sc-aluminates, as well as Ba, Ca-aluminates like that in conventional Ba dispenser cathodes are produced and have been detected by X-Ray Diffraction (XRD) in thermochemical experiments [108]. The generation of these compounds is coincident with the previously reported possible reactions by the thermochemistry investigation for $BaO{\cdot}Sc_2O_3{\cdot}WO_3$ ternary system [78] and real cathodes [119]. These water-soluble compounds are then removed from the surface area of cathodes after ultrasonic water cleaning as displayed on the left side of Fig. 3.48a.

The formation of the surface Ba–Sc–O layer is, therefore, a re-accumulating of Sc with Ba and O, by diffusing from the body to the surface with increase of temperatures, during activation in vacuum. It is observed in an in situ AES analysis exhibited on the right side of Fig. 3.48a, that on the surface of an SDD cathode which was firstly ion etched to remove the residual impregnants, the surface atomic concentration ratios of Sc, Ba, and O to substrate W, i.e., Sc/W, Ba/W, and O/W, increase obviously with temperature and time, indicating the diffusion process of Sc with Ba and O during heating and activation. Furthermore, the identical migration patterns of Sc/W and Ba/W from the pole area to W grains on a 2 μm thick W film deposited on SDD cathode surface, shown in Fig. 3.48b, reveal that a lateral co-diffusion of Sc with Ba, O occurs over the surface.

In contrast, when heating the Sc_2O_3 doped W matrix up to activation temperature, the concentrations of Sc and O remain unchanged [83]. This implies that little diffusion of scandium oxide takes place in the mentioned temperature region. The same conclusion has also been drawn from observations made on the thin film Ba/BaO on Sc/Sc_2O_3 model cathodes by in situ Photoelectron Emission Microscopy (PEEM)/Thermionic Emission Microscopy (ThEEM). The observation confirms that scandium as a metal diffuses on tungsten but Sc_2O_3 does not diffuse at operating temperatures [120].

To sum up, free or ionic Sc is released by reactions of Sc-containing compounds, which are produced from originally added Sc_2O_3, with other compounds or oxides during processes of impregnation and activation. The free Sc diffuses to the W surface with Ba and O to form the Ba–Sc–O layer during activation. The above processes are time consuming; that could explain the aforementioned longtime activation for Ba scandate cathodes.

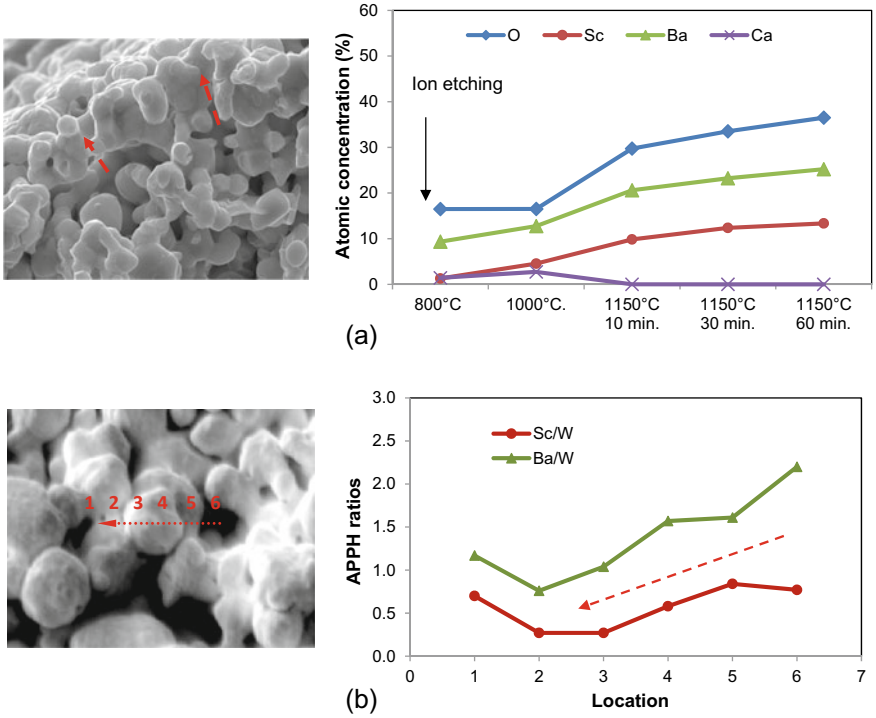
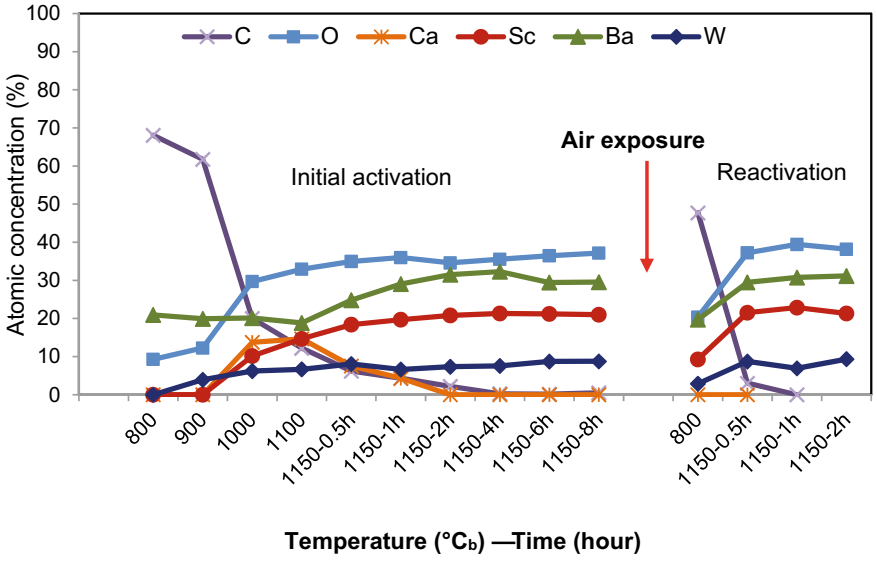
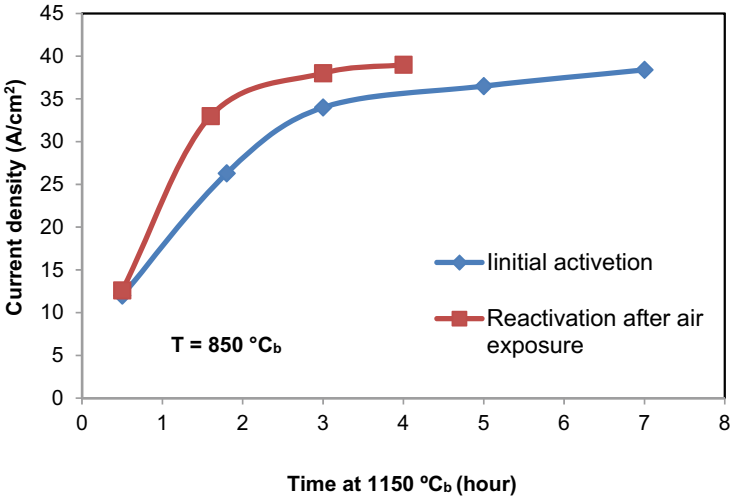


Fig. 3.48 Diffusion of Sc with Ba, O during activation and surface migration of Ba, Sc along a W grain. **a** Changes of AES atomic concentration ratios of Sc/W, Ba/W, and O/W during activation with temperature (°C_b) and time (right); analyzed at the surface of cathode after impregnating and surface cleaning shown in SEM image (left). **b** Variation of Auger Peak–Peak Height APPH ratios of Ba/W and Sc/W (right); from pore to a W grain shown in SEM image (left)

On the other hand, a different picture is valid for the cathode after air exposure. The Ba–Sc–O layer formed during a long initial activation, by removing residual impregnants such as Ca-containing compounds and gradually diffusing Sc, Ba, and O, is soon established after air exposure, just when the surface adsorbed carbon-containing contamination is eliminated, see Fig. 3.49a. Accordantly, the emission rises up to the original or even better level in a period much shorter than that at initial activation, illustrated in Fig. 3.49b. The simultaneous change of all surface elements at the beginning of re-activation indicates the surface layer is already in existence and the Ba, Sc, and O has a close chemical association. Once the layer forms at initial activation, it is basically stable during air exposure.



(a)



(b)

Fig. 3.49 Performances of an SDD cathode at initial activation and reactivation after air exposure. **a** Change of surface elements at different status by in situ Auger analysis. **b** Critical SCL current density at 850 °C_b with activation time

3.3.5.5 Properties Related to the Surface Layer

The Ba–Sc–O multilayer surface feature of Ba scandate cathode, which differs from a monolayer Ba-O dipole on Ba dispenser cathode, brings about distinctions in several properties.

Lower Evaporation Rates

Evaporation properties of SDD cathode have been investigated and compared with that of traditional Ba dispenser cathodes by using of a vapor-collection apparatus equipped with a quartz crystal oscillator and a Time-of-Flight Mass Spectrometer (ToFMS) [108].

The total evaporation rate of SDD cathodes measured by the quartz crystal oscillator is listed in Table 3.7. At an operating temperature of 1050 °C_b or a true temperature of 1400 K, the average evaporation rate is about 1.15×10^{-9} g·cm^{-2}·s^{-1} for SDD cathode. It is almost one magnitude lower than the evaporation rate of 1.48×10^{-8} g·cm^{-2}·s^{-1} for the ordinary Ba dispenser cathode at its normal operating temperature of 1100 °C_b or 1457 K. Moreover, the intensities of ionic Ba peaks on ToFMS mass spectra for SDD cathodes are much lower than those of the Ba dispenser cathodes and with higher activation energy. A comparison of ionic current for Ba^{++} (69), from 1000 °C_b to 1150 °C_b, measured from traditional Ba-W impregnated cathode, M-type cathode, and SDD cathode, is shown in Fig. 3.50.

The lower Ba evaporation rate is consistent with the surface analysis results and the inference described before: barium as well as oxygen binds much stronger to the surface of SDD cathode, by forming a Ba–Sc–O layer, when compared to the Ba on W for Ba dispenser cathode.

In practical applications, the low evaporation of SDD cathodes is beneficial to reduce the possibilities of vacuum electrical breakdown or arcing, secondary emission, and fogging of the output windows when applying the cathodes into VEDs.

Higher Robustness to Residual Gas Poisoning

The Ba–Sc–O semiconductor layer on the surface of Ba scandate cathode also exhibits less sensitivity to residual gas poisoning [121].

Poisoning by oxygen and air for SDD cathodes and M cathode has been compared under the same test conditions and the results are listed in Table 3.8, where partial pressure corresponding to the emission drop to 90% of its initial value is defined as the critical poisoning pressure. The comparison reveals that the critical pressures for SDD cathodes are about half to one order of magnitude higher than that of M

Table 3.7 Average evaporation rates of M-type and SDD cathodes

Cathode type	Average evaporation rate (g·cm^{-2}·s^{-1}) at true temperature (K)		
	1400	1457	1513
M	6.23×10^{-9}	1.48×10^{-8}	5.68×10^{-8}
SDD	1.15×10^{-9}	3.38×10^{-9}	1.02×10^{-8}

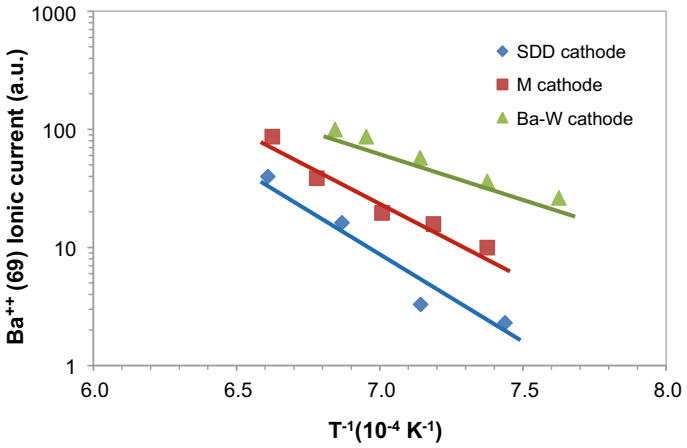


Fig. 3.50 Comparison of Ba evaporation rates of the SDD cathode and Ba dispenser cathodes

cathode at the same temperature, at which the SDD cathode is capable of providing much higher current density than M cathode.

In addition, a full emission recovery is visible for SDD cathode after the poisoning gases are pumped off. The recovery character of SDD cathode is also better than that of M cathode, as indicated in Fig. 3.51. For either cathode, the poisoning-recover test started with an initial current density of 20 A/cm^2 at an operational temperature of 1000 °C$_b$. However, owing to the difference in their poisoning critical pressures, the O_2 partial pressure was 5×10^{-5} Pa for M cathode but 5×10^{-4} Pa for SDD cathode in the testing. A faster emission drop with time and a relatively long recovery was observed for M cathode while the SDD cathode exhibits moderate emission degradation and completes recovery after the termination of the oxygen admission.

The high poisoning pressures and the reversible character for Ba scandate cathodes imply that the Ba–Sc–O layer is more tolerant of a gas poisoning than the Ba-O dipole layer. A physical adsorption of gases rather than a chemical reaction may dominate the poisoning process for this kind of cathode.

Table 3.8 Critical poisoning pressures of air and oxygen for SDD and M cathode

Operating temperature (°C$_b$)	Critical pressure (Pa)			
	Air		O_2	
	SDD	M	SDD	M
900	6.5×10^{-4}	1.1×10^{-4}	2.1×10^{-4}	3.6×10^{-5}
1000	1.4×10^{-3}	2.1×10^{-4}	5.4×10^{-4}	4.9×10^{-5}
1100	6.5×10^{-3}	5.8×10^{-4}	1.1×10^{-3}	2.2×10^{-4}

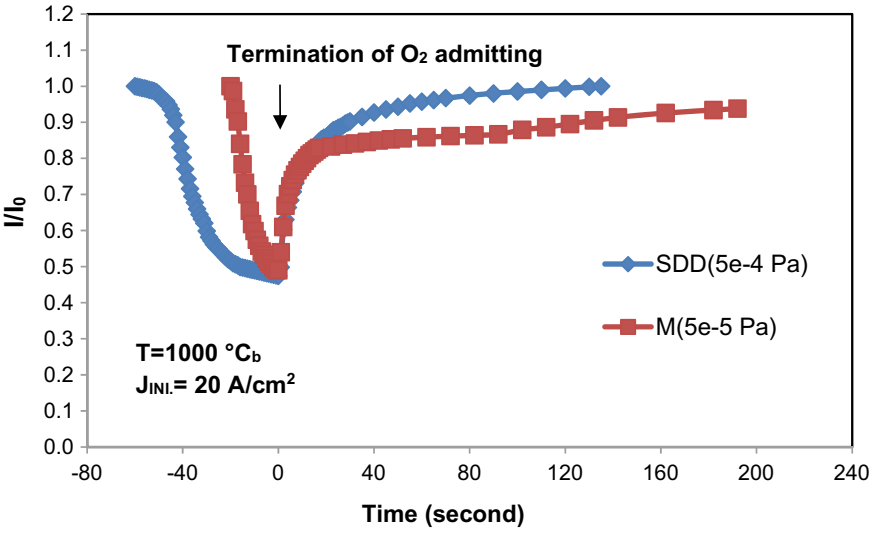


Fig. 3.51 Emission poisoning by O_2 and recovering on SDD cathode and M cathode

Uniqueness of the Contribution of Scandium Oxide to Emission Improvement
Since the addition of scandium oxide to Ba dispenser cathode leads to great enhancement in thermionic emission, other rare-earth oxides such as yttrium and europium oxides were once assumed to evoke equal effects as scandia. To verify the assumption, the W powders have been doped with single Y_2O_3, and with 50/50 mixtures of Sc_2O_3–Eu_2O_3 or Sc_2O_3–Y_2O_3 in a similar manner as doping of Sc_2O_3. These precursor powders were then fabricated into matrices and cathodes in the standard ways. The added contents (in wt%) of Sc_2O_3, Eu_2O_3, and Y_2O_3 in precursor powders and the atomic percentage concentrations (at. %) of elements on the surface of matrices and cathodes after heating or activating are listed, respectively, in Table 3.9. The emission performance of each sample is also indicated for comparison.

It is interesting to find that among the added rare-earth elements, only Sc has been detected on the surface of these cathodes though other elements do exist on the surface of their relative matrices. As a result, for the single Y_2O_3 added cathode (Y6), no active elements other than Ba and O present on W substrate and its emission is just similar to that of the normal Ba dispenser cathode. For the cathodes co-doped of Sc_2O_3–Eu_2O_3 (E3) or Sc_2O_3–Y_2O_3 (Y3), the emission is close to, but not excess, the level of a single Sc_2O_3 doped cathode (S5), resulting only from the appearance of Sc (with Ba and O) on the surface of these cathodes.

Consequently, it is believed that rare-earth oxides other than scandium oxide have little contribution to electron emission though they may have similar chemical and physical properties. The innate character of this interesting phenomenon is expected to trigger/stimulate further investigations.

Table 3.9 Original doped contents of rare-earth oxides in precursor powders and atomic concentrations of Sc/other rare-earth elements on surface of matrices and cathodes with the relevant emission properties of the cathodes

Sample	Doped content (wt%) in precursor powder			Content on matrix surface at 1150 °C_b (at.%)				Content on cathode surface after activation at 1150 °C_b (at.%)					Emission J (A/cm^2)	
	Sc_2O_3	Eu_2O_3	Y_2O_3	Sc	Eu	Y	O	Sc	Eu	Y	Ba	O	850 (°C_b)	1000 (°C_b)
S5	5			33			39	17			25	41	35.7	
E3	3	3		20	18		36	9	0		20	20	20.6	
Y3	3		3	15		7	41	15		0	30	34	34.7	
Y6			6			20	40			0	28	22		3.4

3.3.6 Summary and Discussion

3.3.6.1 Summary

The nanosized-scandia doped dispenser cathode, as a new kind of Ba scandate cathode, exhibits the characteristics as follows:

(1) The cathode has the ability to provide a critical Space Charge Limited (SCL) current density of over 100 A/cm^2 at Mo-brightness temperature of 950 °C$_{B\text{-}Mo}$ or true temperature of 1015 °C. There is little difference between pulsed and DC emission with temperature compensation for electron cooling in DC operation.
(2) The lifetime, during which the high critical SCL emission level can be stably maintained, is longer than ten thousand hours.
(3) The peak work functions determined from Practical Work Function Distribution(PWFD) plots are about 1.40–1.45 eV. The Richardson work functions estimated from zero field emission at low temperatures are around 1.13–1.15 eV with Richardson constants of (2–4) A·cm^{-2}·K^{-2}.
(4) The emission uniformity is greatly improved and becomes similar to that of Ba dispenser cathode. The beam quality is acceptable for application in VEDs.
(5) A Ba–Sc–O semiconductor layer of tens of nanometers thick with certain atomic concentration ratios on W substrate is the source for the copious electron emission.
(6) The high emission property of the cathode remains stable as long as the Ba–Sc–O layer maintains.
(7) The layer has lower evaporation rate, higher tolerance to gas poisoning than that of Ba-O dipole monolayer.
(8) Nanosized Scandia particles doped W matrix in grain size of submicron to 1–2 micron promotes a uniform distribution of Ba–Sc–O layer on W, leading to improvements on both emission capability and uniformity. The more uniform the layer on W substrate, the higher the emission and the better the emission uniformity.

3.3.6.2 Discussion

(a) The Constituents of the Ba–Sc–O Layer

Owing to the difficulty in determining the exact composition and structure of the surface layer on a real Ba scandate cathode by experimental approaches, hypotheses to describe the physics behind the enhanced emission of Ba–Sc–O systems presented recently are mostly based on the investigation of model systems or theoretical analysis [122–124]. These investigations provided interesting results and promoted an understanding of the possible emission mechanism. However, so far no consensus is achieved.

As described before, the surface layer is formed by diffusion of free Sc (with Ba and O) during activation for a real Ba scandate cathode like SDD cathode. But Sc_2O_3

may possibly form on the surface again from the out-diffused free Sc, by a reaction proposed by Hasker [15]:

$$3BaO + 2Sc = Sc_2O_3 + 3Ba \tag{3.11}$$

$$\Delta G^0 = -26 \text{ kcal/(mol. Ba) at } 1200\,K \tag{3.12}$$

In addition, it has been verified by investigations of thermochemistry on the ternary system [78] and practical cathode materials [108, 119] that a reaction of BaO with Sc_2O_3 occurs at 950 °C to produce ternary compound $Ba_2Sc_2O_5$ and decompose into $Ba_3Sc_4O_9$ above 1100 °C.

$$2BaO + Sc_2O_3 = Ba_2Sc_2O_5 \text{ at } 950\,^{\circ}C \tag{3.13}$$

$$2Ba_2Sc_2O_5 = Ba_3Sc_4O_9 + BaO \text{ at } 1100\,^{\circ}C \tag{3.14}$$

As these reactions are possible to take place during cathode activation, a proposal is then drawn up. That is a Ba adsorbed or doped barium scandate, such as $Ba_3Sc_4O_9$, thin layer may represent the Ba–Sc–O layer on the surface of Ba scandate cathode. The proposal is supported by the atomic ratios measured on real cathodes, shown in Table 3.10.

Further investigation is expected to explore the exact constitution of the surface layer. That is important not only for emission mechanism understanding, but also for guiding the improvement of cathode technologies.

(b) The Essential Processes in Forming of Free or Ionic Scandium

Emission degradation by ion bombardment in practical applications is sometimes unavoidable. This problem is especially serious while cathodes work in a relatively poor vacuum environment, such as in CRTs, and consequently caused a particular concern in the early development stage of the Ba scandate cathode [9, 16, 19, 22]. The VEDs usually operate with excellent vacuum conditions and in SCL regime, reducing the influence of ion bombardment. However, this is still a great challenge for all kinds of Ba scandate cathodes.

Ultimately, the recovery of emission after ion bombardment is related to the restoration of the surface layer. It has been noticed that when the surface elements are removed by strong ion beam etching, reactivating arouses a full recovery of Ba

Table 3.10 Atomic ratios of barium scandates and Ba–Sc–O layer on SDD cathode

	Atomic ratio		
	Ba	Sc	O
$Ba_2Sc_2O_5$	1	1	2.5
$Ba_3Sc_4O_9$	0.75	1	2.25
Ba–Sc–O layer	1.5–2	1	2–3

in a short time while the re-diffusion of Sc only provides a partial recovery in a time depending on the W grain size. This process brings about a big confusion that why the obvious liberation of free or ionic Scandium taken place at initial activation can only be partially performed during reactivation after ion etching.

Till now the reactions to release the free or ionic Sc are still unclear. Two possible reactions have been suggested.

The one is proposed by Yamamoto [18]:

$$Sc_2W_3O_{12} + 3Ba = 2Sc + 3BaWO_4 \tag{3.15}$$

and the other is by Magnus and Hill [78]:

$$Ba_3Sc_4O_9 + W = Ba_3Sc_2WO_9 + 2Sc \tag{3.16}$$

However, no thermodynamic data were ascertained.

Although the microstructure of SDD cathode is beneficial to liberation and diffusion of free Sc, it, nevertheless, needs an essential improvement on the recovery property of scandium. To achieve this, a full understanding of the fundamental generation mechanism of free Sc is one of the major expectations.

(c) The Nature of the Surface Nanoparticles

In the context of the above discussion, it is interesting to observe that, apart from uniformly spread Ba–Sc–O multilayer, a multitude of nanoparticles have been observed on the surface of the layer, especially along the visible growth boundaries of the underlying W grains [10].

The particles, which can be clearly seen from the AFM amplitude image shown on the left-hand side of Fig. 3.52a, appear only for SDD cathodes after full activation and operation [10]. Neither such particles have been found for the activated Ba dispenser cathodes as shown on the right-hand side of Fig. 3.52a, nor are they present in SDD cathodes before activation.

No other elements except Ba, Sc, and O have been detected at the nanoparticle areas by AES with 40 nm spatial resolution. Because the sizes of the particles are below the resolution, it is not clear if the analyzed results correspond well to the actual constituent of the particles.

The potential effects of these nanoparticles on emission have been evaluated by assessing the local electric field enhancement around the particles. Referring to Edgcombe and Valdre's report [125], for a cylindrical shank in parallel-plane electrode geometry with tip-to anode distance of 0.1 mm or more, the field enhancement factor γ can be expressed as

$$\gamma \approx 1.2(L/R + 2.15)^{0.9} \tag{3.17}$$

where L is the tip length and R is the tip radius.

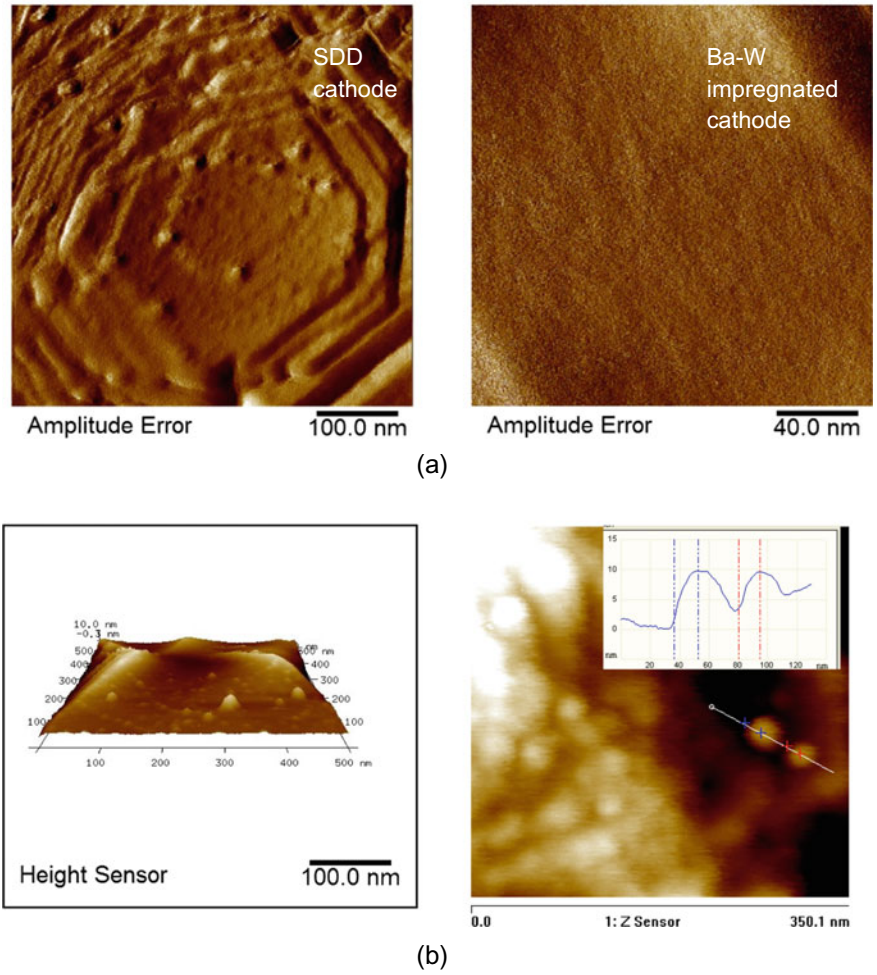


Fig. 3.52 **a** AFM amplitude images of activated SDD cathode (left) and activated Ba-W impregnated cathode (right). **b** Shapes (left) and dimension estimation (right) of the nanoparticles in AFM height sensor images

By considering a nanoparticle as a projection with a length of 10 nm and radius of 5 nm, like that shown by AFM in Fig. 3.52b, the field enhancement factor will roughly equal to 4.3.

Thus, at an average field strength of about 40 kV/cm between cathode and anode, which is related to the field that an SCL critical current density of 100 A/cm^2 can be reached under normal experimental conditions such as the cases of Figs. 3.30 and 3.32, the electric field near the nanoparticles will be around 170 kV/cm, i.e., about 1.7×10^7 V/m. This field strength is still much lower than 10^8–10^9 V/m for the

starting of field emission unless the L/R of the particle reaches an unrealistic value of about 30.

However, the influences of this electric field on electron emission can be exerted by Schottky effect on thermionic emission and, especially, by field penetration when a semiconductor model is considered for Ba scandate dispenser cathodes to heighten the emission. If that is the case, the appearance of nanoparticles on SDD or other Ba scandate cathodes will not only strengthen the emission, but also cause emission nonuniformity while the cathodes operate at low temperatures in TL regime, but will have limited effects in SCL regime.

It must be pointed out that the SEM and AFM observations were not obtained at cathode operating temperatures. The activated or operated cathodes were cooled down to room temperature before analysis. The exact shapes and dimensions of these nanoparticles at operation conditions are not clearly figured out yet.

The nature of these surface particles, their formation process and existence status, behavior, and effects on emission are interesting subjects for a deeper study.

Ba scandate cathodes with their proven outstanding emission capability have attracted continuous attention for almost half a century. However, as described above, many phenomena are still unclear and/or controversial. The comment for the old oxide-coated cathode made by Gilmour [48] should be also suitable for this kind of cathodes: "It is clear that as far as science goes, the state of the oxide coated cathode is confused indeed. Not only does theory fail to help us make good cathodes; it seems not to fit all the evidence. Maybe cathodes work sometimes one way and sometimes another, which isn't as ridiculous as it may sound. At any rate, while some physicists are uninterested and others confused, cathode making continues to be an art practiced, it seems, by artists with a good deal of witch doctor in their make-up. When something goes wrong, this or that demon is exorcised, and eventually things go right again. And, unlikely as it seems, cathodes continue to improve".

3.4 General Conclusions and Outlook

In the preceding sections, two successful versions of Ba scandate-type cathodes have been described in detail, namely the LAD top-layer Ba scandate + Re dispenser cathodes on W–I base as pioneered by Philips Research [9] and nanosized-scandia doped dispenser cathodes with W-base matrix (SDD) as introduced first by Beijing University of Technology (BJUT) and Beijing Vacuum Electronics Research Institute (BVERI) and continued by BJUT in China [10]. A common feature of all Ba scandate-type cathodes is that they exhibit a low Richardson work function of about 1.15 eV and Richardson constants varying from 2 to 8 A/(cm^2K^2), with increasing uniformity and electron emission capability. This feature is connected with a second characteristic property that Ba scandate cathodes show anomalous behavior in the saturation range which does not follow the Schottky effect as for Ba dispenser cathodes, but is similar to oxide cathodes. This results in only a small gradual deviation from the space charge limit in the current-voltage characteristics at higher temperatures. From emission

characterization of LAD and SDD scandate cathodes, it was found that one can describe saturation by a power law with power x_s (slope x_s in double logarithmic *I(U)* plot), which is smaller than the space charge limit (slope x_{scl}). In LAD and SDD scandate cathode experiments, it was observed as a rule that higher uniformity (higher A_R in the case of LAD cathodes) caused steeper slopes x_s. A convincing model and a theory of the saturation behavior is still needed, but there are a lot of indications that basically an oxide-type model can explain most of the properties, but eventually linked to an increasing area fraction of a surface compound containing (Ba, Sc, O) and, possibly, a Ba monolayer on top. In this context it would be helpful, to prepare very flat polished surfaces of Ba scandate cathodes for high-resolution work function mappings as, for example, described by Eng [126] or Back et al. [127] and correlate it with Auger elemental mappings.

On average, the AES depth profile results for LAD scandate cathodes gave a ratio of Ba:Sc:O = 2.04: 1: 2.78 to a depth of about 36 nm. This is in line with the results for SDD cathodes of Ba:Sc:O = (1.5–2): 1: (2–3). W or Re do not play a direct role and just serve as the conducting base. The Ba:Sc:O ratios are consistent with a surface compound composition of $2BaO{\cdot}Sc_2O_3$ with a surface coverage of 2 Ba (about a monolayer). It is not clear if there exist 2 phases or maybe an unknown high-temperature phase only formed after activation.

The success of the LAD scandate cathode and SDD cathode is linked to higher emission capability and higher uniformity of emission. This is due to finer surface structures and more homogeneous distribution of Sc, in case of LAD cathodes due to surface coating with nanoparticles of initially uniform distribution (but on a coarser base) and in case of SDD cathodes due to the fine-grained structuring of porous tungsten base uniformly doped with scandia nanoparticles. The advantage of the SDD cathode is that it is prepared by a much cheaper and practical method than the top-layer preparation by LAD, which only can get competitive for high volume production of cathodes as in the high noon of CRT production. Nowadays scandate cathodes are still needed for high-end applications such as THz imaging devices, but in smaller numbers. Therefore SDD cathodes especially with high DC emission are the cathodes of choice. Yet there are good prospects that one can also use the less-expensive sputtering instead of LAD, but with targets of the same composition as used for LAD to realize the high-performance cathode. In view of the great demands on applications of high current density cathodes, the doors for new ideas and technologies are always open for further improvement of the currently existing Ba scandate cathodes.

References

1. G. Gaertner, Historical development and future trends of vacuum electronics, J. Vac. Sci. Technol. B **30/6**, 060801 (2012)
2. G. Gaertner, H.W.P. Koops, Vacuum electron sources and their materials and technologies, chapter 10 of *Vacuum Electronics, Components and Devices*, Ed. J. Eichmeier, M. Thumm, Springer (2008)

3. R. Barker, Miniature Electron Sources for Tomorrow's Vacuum THz Devices, Report No. AFRL SR-AR-TR-06-0330, Institute for Research in Electronics and Applied Physics, Univ. of Maryland, July 2006
4. M.J. Rosker, Vacuum electronics and the world above 100 GHz, in Proceedings IVEC 2008, Monterey, CA, p. 5–7 (2008)
5. M.C. Green, Cathode technology overview—Current status and future Directions, in Proceedings IVEC 2008, Monterey, CA, p. 3–4 (2008)
6. B.C. Djubua, A.P. Makarov, A.A. Negirev, S.E. Rozhkov, E.M. Zemchikhin, Metal alloy cathodes for very high emission- density applications, in Proceedings IVEC 2010, Monterey, CA, p. 155 (2010)
7. B.C. Djubua, Research and development of dispenser, metal alloy and oxide thermionic Cathodes, 1994 TRI-Service/NASA Cathode Workshop, Cleveland/Ohio, Conference Record, p. 3–6 (1994); based on older results from 1967/1979
8. R.E. Thomas, J.W. Gibson, G.A. Haas, R.H. Abrams, Thermionic sources for high-brightness electron beams. IEEE Trans. Electron. Dev. **37**(3), 850–861 (1990)
9. 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)
10. Y. Wang, J. Wang, W. Liu, K. Zhang, J. Li, Development of high current-density cathodes with scandia-doped tungsten powders. IEEE Transact. ED **54**(5), 1061–1070 (2007)
11. A. Figner, A. Soloveichik, I. Judinskaya, Metal Porous Body Having Pores Filled with Barium Scandate, US Patent 3 358 178, filed 27.10.1964, granted 12.12.1967
12. A. van Oostrom, L. Augustus, Activation and early life of a pressed barium scandate cathode. Appl. Surf. Sci. **2**(2), 173–186 (1979)
13. A. van Stratum, J. van Os, J.R. Blatter, P. Zalm, Barium-Aluminum-Scandate Dispenser Cathode, US-Patent 4007393, granted 8-2-1977, priority 21-2-1975
14. S. Taguchi, T. Aida, S. Yamamoto, I.E.E.E. Trans, Electron Dev. **31**(7), 900–903 (1984)
15. J. Hasker, J. van Esdonk, J.E. Crombeen, Properties and manufacture of top-layer scandate cathodes. Appl. Surf. Sci. **26**, 173–195 (1986)
16. J.E. Crombeen, J. Hasker, Some experiments on the role of oxygen and surface reactions for tungsten and scandate thermionic emitters. IEEE Trans. Electron Dev. **37**(12), 2589–2594 (1990)
17. S. Yamamoto, S. Sasaki, S. Taguchi, I. Watanabe, N. Koganezawa, Application of an impregnated cathode coated with W-Sc_2O_3 to a high current density electron gun. Appl. Surf. **33–34**, 1200–1207 (1988)
18. S. Yamamoto, S. Taguchi, T. Aida, S. Kawase, Study of metal film coating on Sc_2O_3 mixed matrix impregnated cathodes. Appl. Surf. Sci. **17**, 517–529 (1984)
19. S. Yamamoto, I. Watanabe, S. Taguchi, S. Sasaki, T. Yaguchi, Formation mechanism of a monoatomic order surface layer on a Sc-type cathode. Jpn. J. Appl. Phys. **28**, 490–494 (1989)
20. R. Longo, D. Dibb, Dispenser scandate cathode: a progress report, Conf. Record, 1992, TRI Cathode Workshop, Greenbelt/Md, 51–56
21. Y. Wang, T. Pan, Investigation of pulsed laser depositing Sc-coated cathode, Appl. Surf. Sci. **146**, 62–68 (1999) and Vac. Sci. and Technology (in Chinese), vol. **16/1**, p. 10, 1996
22. G. Gaertner, P. Janiel, J.E. Crombeen, J. Hasker, Top-layer scandate cathodes by plasma-activated CVD, *Vacuum Microelectronics 1989*, Institute of Physics Conf. Series No. **99**, Bristol and New York, Ed. R. E. Turner, 25–28 (1989)
23. U. van Slooten, P. Duine, Scanning Auger measurements of activated and sputter cleaned Re-coated scandate cathodes. Appl. Surf. Sci. **111**, 24–29 (1997)
24. SMI Spectra-Mat Inc., Dispenser Cathode Products, PB-100-A- 04/14 (2014); Emission Characteristics for Scandium Type Dispenser Cathodes, TB 119 (2001) and "Notes on Dispenser and Oxide Cathodes", TB 128 (1999), Watsonville (USA) and b) S. Fukuda et al.,Appl. Surf. Sci. **146**, 84–88 (1999)
25. D. Kirkwood, J. Balk et al., Frontiers in thermionic research. IEEE Trans. ED **65**, 2061–2071 (2018)

26. G. Gaertner, P. Janiel, H. Lydtin, L. Rehder, Generation of very small particles by laser ablation in a carrier-gas, deposition and characterization, in Synthesis and Measurement of Ultrafine Particles, Delft University Press 1993, pp. 41–50
27. G. Gaertner, H. Lydtin, Review of ultrafine particle generation by laser ablation from solid targets in gas flows. NanoStruct. Mater. **4/5**, 559–568 (1994)
28. G. Gaertner, P. Geittner, D. Raasch, A. Ritz, D. Wiechert, Dynamic Shielding during ion bombardment of Ba dispenser cathodes. Appl. Surf. Sci. **146**, 12–16 (1999)
29. G. Gaertner, P. Geittner, E. Klein, "Elektrische Entladungsröhre oder Entladungslampe und Scandat-Vorratskathode", priority 31.7.1995 DE, Europ. Patent EP 0757370 granted 5.7.2000 and US-Patent 6 348 756 B1, "Electric discharge tube or discharge lamp and Scandate dispenser cathode", granted 19.2.2002
30. G. Gaertner, P. Geittner, H. Lydtin, "Scandat-Vorratskathode mit Re und Barium-Calcium-Aluminat-Beschichtung", German Patent application 19828729, filed 29.6.1998
31. G. Gaertner, P. Geittner, D. Raasch, D.U. Wiechert, "Scandat-Vorratskathode", German Patent 19961672B4, filed 20.12.1999, granted 9.4.2009
32. G. Gaertner, E. Klein, Scandate Dispenser Cathode, European patent application EP 04106783.6, filed 21.12.2004, and WO 2006/067670 A2, publ. 29.6.2006
33. G. Gaertner, W. Keur, "Target material for Barium Scandate Dispenser Cathode or Material", US-Patent application 3.8.2011, 2011P00729US_110803, 31.7.2012 patent filed WO2013/018027A1, "Target for Barium Scandate Dispenser Cathode", published 7.2.2013
34. C. Makovicka, G. Gaertner, A. Hardt, W. Hermann, D.U. Wiechert, Impregnated cathode investigations by SFM/STM and SEM/EDX. Appl. Surf. Sci. **111**, 70–75 (1997)
35. S. Sasaki, T. Yaguchi, N. Mori, S. Taguchi, M. Shibata, Non-uniform emission distribution on a scandate impregnated cathode. Appl. Surf. Sci. **146**, 17–21 (1999)
36. P. van der Heide, "Os/Ru coated Impregnated Cathode for Use in Cathode Ray Tubes", 1992 TRI-Service/NASA Cathode Workshop, Conf. Record, 155–159 (1992)
37. J. B. Scott, "Extension of Langmuir space-charge theory into the accelerating field range". J. Appl. Phys. **52**(7), 4406–4410 (1981)
38. J. Hasker, P. van Dorst, Pitfalls in the evaluation of cathode properties from I-V characteristics. IEEE Trans. ED **36**, 201–208 (1989) and J. Hasker, J.E. Crombeen, P. van Dorst, Comment on progress in Scandate cathodes. IEEE Transact. Electr. Dev. **36**, 215–219 (1989)
39. A. Manenschijn, S. Deckers, T. Weekers, P. van der Heide, *"Emission characterization of impregnated cathodes and Scandate cathodes", Conference record of 1992 TRI-Service/NASA Cathode Workshop* (Greenbelt/Md, USA, 1992), pp. 67–71
40. J. Hasker, Calculation of diode characteristics and proposed characterization of cathode emission probability. Appl. Surf. Sci. **16**, 220–237 (1983)
41. J. Hasker, Beam current characteristic and cathode loading of electron guns with rotational symmetry: some important properties and method of calculation. Philips Res. Reports **27**, 513–538 (1972)
42. T.G. Spanjer, A.A. van Gorkum, W.M. van Alphen, Electron guns for projection television. Philips Tech. Rev. **44**, 348–356 (1989)
43. H. Yuan, X. Gu, K. Pan, Y. Wang, W. Liu, K. Zhang, J. Wang, M. Zhou, J. Li, Characteristics of scandate-impregnated cathodes with sub-micron scandia-doped matrices. Appl. Surf. Sci. **251**, 106–113 (2005)
44. X. Zhang, G. Gaertner, Ion bombardment investigations of impregnated cathodes. Appl. Surf. Sci. **215**, 25–32 (2003)
45. G. Gaertner, P. Geittner, J. Crombeen, "Aktivierung von thermionischen Kathoden", German patent application P 199 32 564.2, filed 20.12.1999, disclosed 28.6.2001
46. G. Gaertner, S. Ording, T. Bisschops, "Comparison of Ba evaporation from different Ba-dispenser cathode types", ITG-Fachber. **150**, *Displays and Vacuum Electronics*, VDE-Verl., 489–494 (1998)
47. G. Gaertner, P. Geittner, D. Raasch, D.U. Wiechert, Supply and loss mechanisms of Ba dispenser cathodes. Appl. Surf. Sci. **146**, 22–30 (1999)

48. A.S. Gilmour, Jr., Chapter 5, "Thermionic Cathodes" in *Klystrons, Traveling Wave Tubes, Magnetrons, Crossed-Field Amplifiers and Gyrotrons*, ARTECH HOUSE (2011)
49. G. Gaertner, P. Geittner, D. Raasch, A. Ritz, D.U. Wiechert, Investigation of the shielding-effect during ion bombardment of ba-dispenser-cathodes, ITG-Fachbericht **150**, *Displays and Vacuum Electronic*s, VDE-Verlag, 483–487 (1998)
50. R.S. Raju, C.E. Maloney, Characterization of an impregnated scandate cathode using a semiconductor model. IEEE Trans. Electron Dev. ED **41**(12), 2460–2467 (1994)
51. R. Cortenraad, A.W. Denier van der Gon, H.H. Brongersma, G. Gaertner, A. Manenschijn, Thermionic cathodes studied by low-energy ion scattering, ITG-Fachbericht **150**, *Displays and Vacuum Electronics*, VDE-Verlag, 477–482 (1998)
52. R. Cortenraad, A. van der Gon, H. Brongersma, G. Gaertner, A. Manenschijn, Quantitative LEIS analysis of thermionic dispenser cathodes. Appl. Surf. Sci. **146**, 69–74 (1999)
53. R. Cortenraad, A. van der Gon, H. Brongersma, G. Gaertner, A. Manenschijn, Surface analysis of thermionic dispenser cathodes. Appl. Surf. Sci. **191**, 153–165 (2002)
54. S. Yamamoto, T. Yaguchi, S. Sasaki, I. Watanabe, Work function measurement of (W-$Sc_2W_3O_{12}$)-coated impregnated cathode by Retarding Potential method utilizing titaniated W(100) field emitter. Jpn. J. Appl. Phys. **28**, L865f (1989)
55. P. Zagwijn, J. Frenken, U. van Slooten, P.A. Duine, A model system for scandate cathodes. Appl. Surf. Sci. **111**, 35–41 (1997)
56. J. Eichmeyer, *Moderne Vakuumelektronik*, Springer 1981, pp. 208–210
57. G. Gaertner, D. Barratt, New developments and life aspects of oxide and Ba dispenser cathodes, ITG-Fachbericht **183**, *Displays and Vacuum Electronics*, VDE-Verlag, 111–116 (2004)
58. G. Gaertner, P. Geittner, H. Lydtin, Emissionseigenschaften von Top -Layer Scandat Kathoden, ITG Fachbericht **132**, "Vakuumelektronik und Displays", VDE Verlag, 35–40 (1995)
59. G. Gaertner, P. Geittner, D. Raasch, Low temperature and cold emission of Scandate cathodes. Appl. Surf. Sci. **201**, 61–68 (2002)
60. G. Gaertner, P. Geittner, H. Lydtin, Low-temperature cathode having an emissive nanostructure, US-Patent 5 866 975, granted 2.2.1999, priority 5.5.1995 DE
61. P. Geittner, G. Gaertner, D. Raasch, "Low temperature properties of Ba-dispenser cathodes", Technical Digest 05.19p, 12th Int. Vacuum Microelectronics Conf. 1999, Darmstadt, 190–191 and J. Vac. Sci. Techn. B **18**, 997–999 (2000)
62. D. Raasch, P. Geittner, G. Gaertner, "Ba losses due to oxygen adsorption on Ba-dispenser cathodes", Technical Digest 05.20p, 12th Int. Vacuum Microelectronics Conf. 1999, Darmstadt, 192–193, and J. Vac. Sci. Techn. B **18**, 1000–1002 (2000)
63. G. Gaertner, P. Geittner, D. Raasch, U. Schiebel, Low Temperature Thermionic Cathodes for CRTs and FPDs, Eurodisplay '99, 283–287 (1999)
64. A. Shroff et al., Performance and life tests of various types of impregnated cathodes. Applic. Surf. Sci. **8**, 36–49 (1981)
65. T. Aida, H. Tanuma, S. Sasaki, T. Yaguchi, Emission life and surface analysis of barium impregnated thermionic cathodes. J. Appl. Phys. **74**, 6482–6487 (1993)
66. G. Gaertner, P.A.M. van der Heide, New developments in CRT cathodes, in Proceedings of IDW 2000, Kobe, Japan, CRT4-1 (invited), 513–516 (2000)
67. G. Gaertner, D. Barratt, Life-limiting mechanisms in Ba-oxide, Ba dispenser and Ba Scandate cathodes. Appl. Surf. Sci. **251**, 73–79 (2005)
68. J.-M. Roquais, F. Poret, R. le Doze, J.L. Ricaud, A. Monterrin, A. Steinbrunn, Appl. Surf. Sci. **215**, 5–17 (2003)
69. T. Higuchi et al., Modeling of life deterioration by ion bombardment of a dispenser cathode coated with W/Ir film. Appl. Surf. Sci. **200**, 125–137 (2002)
70. T. Higuchi et al., Modeling of emission slump by ion bombardment of a dispenser cathode in an electron tube, Ext. Abstracts IVEC-IVESC 2012, Paper 07
71. C. Marrian, A. Shih, The operation of coated tungsten based dispenser cathodes in nonideal vacuum. IEEE Transact. Electr. Dev. **36**(1), 173–179 (1989)
72. A. Sharma, A. Chopra, R. Mathew, Emission poisoning studies on impregnated tungsten dispenser cathodes under O_2 and CO_2 environment, Appl. Surf. Sci. **40**, 97–101 (1989)

73. S. Kimura, D. Miyazaki, M. Hara, M. Fujiwara, Emission characteristics and surface composition of dispenser cathodes in realistic vacuum. Vacuum **41**, 1763–1765 (1990)
74. H. Gallagher, Gas poisoning of cathodes, IEEE Conf. Record, 9th Conf. on Tube Technology New York 1968, 15–22
75. R.O. Jenkins, W. Trodden, The poisoning of impregnated cathodes. J. Electron. Control **7**, 393–415 (1959)
76. N. van Veen, XPS on impregnated cathodes: surface concentrations and thermal stability. Appl. Surf. Sci. **29**, 113–126 (1987)
77. J. Hasker, H. van Stoffelen, "Alternative" Auger analysis reveals important properties of M type and scandate cathodes. Appl. Surf. Sci. **24**, 330–339 (1985)
78. S. Magnus, D. Hill, W. Ohlinger, Emission properties of compounds in the $BaO{\cdot}Sc_2O_3{\cdot}WO_3$ ternary system, Appl. Surf. Sci**. 111**, 42–49 (1997); Thermochemistry in the $BaO{\cdot}Sc_2O_3{\cdot}WO_3$ ternary system, Appl. Surf. Sci. **111**, 50–55 (1997)
79. J. Li, S. Yua, W. Shao, Q. Chen, M. Zhu, Investigation and application of impregnated scandate cathodes. Appl. Surf. Sci. **215**, 49–53 (2003)
80. T. Shintake, N. Akasaka, H. Matsumoto, Y. Ohkubo, H. Yonezawa, Development of C-band 50 MW pulse klystron for e^+ e^- linear collider, Proc. Particle Accelerator Conference, 1997, Vancouver, Canada, p. 533 (1997), https://doi.org/10.1109/pac.1997.749748
81. E. Uda, O. Nakamura, S. Matsumoto et al., Emission and life characteristics of thin film top-scandate cathode and diffusion of Sc_2O_3 and W. Appl. Surf. Sci. **146**, 31–38 (1999)
82. J. Wang, W. Liu, L. Li, Y.C. Wang, Y. Wang, M. Zhou, A study of scandia-doped pressed cathodes. IEEE Trans. Electron Dev. ED **56**, 799 (2009)
83. W. Liu, K. Zhang, Y. Wang, K. Pen, X. Gu, J. Wang, J. Li, M. Zhou, Operating model for scandia doped matrix scandate cathodes. Appl. Surf. Sci. **251**, 80–88 (2005)
84. J. Wang, M. Zhou, Y. Wang, W. Liu, Manufacturing method of precursor powder for Sc-containing dispenser cathode, China Patent ZL 200510053831.6 (2005)
85. J. Wang, Y. Cui, W. Liu, Y. Wang, F. Yang, F. Zhou, M. Zhou, A study of scandia-doped impregnated cathode fabricated by spray drying method. IEEE Trans. Electron Dev. ED **62**, 1635–1640 (2015)
86. Y. Cui, J. Wang, W. Liu, Preparation and characterization of scandia-doped tungsten powders prepared by a spray-drying method, Res. Chem. Intermed., published on line, https://doi.org/10.1007/s11164-011-0248-4
87. A.M. Shroff, Review of dispenser cathodes. Revue Technique, Thomson-CSF **23**, 948–1026 (1991)
88. L. Li, J. Wang, Y. Wang, Generation of high-current-density sheet electron beams. IEEE Electron Device Lett. **30**, 228–230 (2009)
89. M.J. Cattelino, G. Miram, Predicting cathode life expectancy and emission quality from PWFD measurements. Appl. Surf. Sci. **111**, 90–95 (1997)
90. T. J. Grant, A powerful quality assurance technique for dispenser cathodes and electron guns, in IEDM Tech. Dig., 1984, p. 334–337
91. A. Sandor, Activation Process of Impregnated Dispenser Cathode Viewed in the Large-screen Emission Microscope. Int. J. Electron. **13**(5), 401–416 (1962). https://doi.org/10.1080/00207216208937447
92. W. Liang, Y. Wang, J. Wang, W. Liu, F. Yang, DC emission characteristic of nanosized-scandia-doped impregnated dispenser cathodes. IEEE Trans. Electron Dev. ED **61**, 1749–1753 (2014)
93. W. Liu, Y. Wang, J. Wang, Y.C. Wang, B. Vancil, Emission characteristics of nanosized scandia-doped dispenser cathodes in open electron-gun structures. IEEE Trans. Electron Dev. ED **58**, 1241–1246 (2011)
94. G. Miram, L. Ives, M. Read, R. Wilcox, *Emission Spread in Thermionic Cathodes* (Technical Digest, Fifth IEEE IVEC, 2004), pp. 303–304
95. Y. Wang, J. Wang, W. Liu, L. Li, Y.C. Wang, X. Zhang, Correlation Between Emission Behavior and Surface Features of Scandate Cathodes. IEEE Trans. Electron Dev. ED **56**, 776 (2009)

96. G. Miram, M. Cattelino, Life Test Facility for Thermionic Cathodes, Conference Record 1994, Tri-Service/NASA Cathode Workshop, Cleveland, U.S, March 1994, p. 29–31
97. R. L. Ives, G. Collins, M. Read, G. Miram, D. Marsden, Electron Guns for Terahertz Vacuum Electron Sources, Terahertz Science and Technology vol. **4/4**, p. 230 (2011), ISSN 1941–7411
98. M.A. Basten, J.H. Booske, Two-plane focusing of high-space-charge sheet electron beams using periodically cusped magnetic fields. J. Appl. Phys. **85**, 6313 (1999)
99. L. Li, Y. Wang, W. Liu, Y.C Wang, J. Wang, A. Srivastava, J.-K. So, G.-S. Park, Development of high-current sheet beam cathodes for terahertz sources, IEEE Trans. Electron Dev. **ED 56**, p. 762 (2009)
100. A. Srivastava, J.-K. So, Y. Wang, J. Wang, R.S. Raju, S.-T. Han, G.-S. Park, Design of sheet-beam electron gun with planar cathode for terahertz devices. J. Infrared Millim. & Terahertz Waves **30**, 670–678 (2009)
101. R. Forman, Surface studies of barium and barium oxide on tungsten and its application to understanding the mechanism of operation of an impregnated tungsten cathode. J. Appl. Phys. **47**, 5272–5279 (1977)
102. G.A. Haas, A. Shih, C.R.K. Marrian, Interatomic Auger analysis of oxidation of thin Ba film II. Applications to impregnated cathodes. Applications Surf. Sci. **16**, 139–162 (1983)
103. K.G. Eyink, B.C. Lamartine, W.V. Lampert, T.W. Haas, Quantification of the surface coverage of Ba on W substrates using Auger electron spectroscopy. Applications Surf. Sci. **20**, 215–227 (1985)
104. G. Lesny, R. Forman, Surface studies on Scandate cathodes and synthesized scandates. IEEE Trans. Electron Dev. ED **37**, 2595–2604 (1990)
105. S. Yamamoto, S. Taguchi, l. Watanabe, S. Kawase, Impregnated cathode coated with tungsten thin film containing Sc_2O_3. J. Vac. Sci. Technol. **A5,** 1299 (1987)
106. A. Shih, J.E. Yater, C. Hor, Ba and BaO on W and on Sc_2O_3 coated W. Appl. Surf. Sci. **242**, 35–54 (2005)
107. A. P. Makarov, The change of a multilayer film work function of jointly adsorbed atoms of barium, scandium and oxygen on the facet (100) of a tungsten crystal at adsorption and heating, in Proceedings IVEC 2015, Beijing, China, April. 27–29, (2015), p. 41. DOI:10.1109/IVEC.2015.7223742
108. Y. Wang, J. Wang, W. Liu, X. Zhang, L. Li, Emission mechanism of high current density scandia-doped dispenser cathodes, J. Vac. Sci. Technol. **B 29**, p. 04E106-1–9 (2011)
109. Y. Wang, J. Wang, W. Liu, Development of scandate cathode and its prospect_ a review, in Proceedings IVESC, 2012, Monterey, USA, p. 41 (2012)
110. R.W. Speringer, T.A. Haas, Auger electron spectroscopy study of cathode surface during activation and poisoning the Barium on Oxygen on Tungsten dispenser cathode. J. Appl. Phys. **45**, 5260 (1974)
111. A. Shih, G.A. Haas,.R.K. Marrian, "Preparation and oxidation of a thin Ba film", Applications Surf. Sci. **16**, 93–105 (1983)
112. R. Forman, A proposed physical model for the impregnated tungsten cathode based on Auger surface studies of the Ba-O-W system. Appl. Surf. Sci. **2**, 258–274 (1979)
113. T. Kawano, Y. Takai, R. Shimizu, Characterization of Sc-O/W(100) surface as Schottky emitter: work function change for activation processing. Jpn. J. Appl. Phys. **39**, 577–580 (2000)
114. F. Yang, J. Wang, Y. Wang, W. Liu, X. Zhu, Investigation of nanosized-scandia-doped dispenser cathodes with machined surfaces. IEEE Trans. Electron Dev. ED **63**, 1728–1733 (2016)
115. J. Wang, Y. Wang, X. Wang, X. Zhang, F. Yang, W. Liu, M. Zhou, Surface characteristics of scandate dispenser cathodes during life, in Proceedings IVEC, 2013, Paris, France (2013). https://doi.org/10.1109/ivec.2013.6571168
116. D.A. Wright, J. Woods, The emission from oxide-coated cathodes in an accelerating field. Proc. Phys. Soc. **65**, 134–148 (1952)
117. R. Vaughan, A Synthesis of the Longo and Eng Cathode Emission Models. IEEE Trans. Electron Dev. ED **33**, 1925 (1986)

118. R.T. Longo, Physics of thermionic dispenser cathode aging. J. Appl. Phys. **94**, 6966 (2003)
119. Z. Zeng, Q. Su, M. Li, L. Cai, Structure analysis study in emissive materials of the Barium Scandate dispenser cathode. J. Electron. (in Chinese), vol.**12/.2**, 179 (1990)
120. J.M. Vaughn, K.D. Jamison, M.E. Kordesch, In situ emission microscopy of scandium/scandium-oxide and barium/barium-oxide thin films on tungsten, IEEE Trans. Electron Dev. **ED 56**, 794 (2009)
121. Y. Yang, Y. Wang, W. Liu, Z. Pan, J. Li, J. Wang, Robustness investigation on nanosized-scandia-doped dispenser cathodes. IEEE Trans. Electron Dev. ED **65**, 2072–2076 (2018)
122. R.M. Jacobs, J.H. Booske, D. Morgan, Electron emission energy barriers and stability of Sc_2O_3 with Adsorbed Ba and Ba–O. J. Physical Chem. C **118**, 19742–19758 (2014)
123. M.E. Kordesch, J.M. Vaughn, C. Wan, Model scandate cathodes investigated by thermionic-emission microscopy, J. Vac. Sci. Technol. **B 29(4)**, 04E102-1, Jul/Aug 2011
124. V.I. Kapustina, I.P. Lia, A.V. Shumanov, YuYu. Lebedinskii, A.V. Zablotskii, Physical operating principles of scandate cathodes for microwave devices. Phys. Electron. **62**, 116–126 (2017)
125. C.J. Edgcombe, U. Valdre, Microscopy and computational modelling to elucidate the enhancement of factor for field electron emitters. J. Microscopy **203**, 188 (2001)
126. C. Eng, H. Kan, Scanning Auger and work function measurements applied to dispenser cathodes, Aerospace corporation report SD-TR-81-35, (1981)
127. T.C. Back, S. Fairchild et al., Work Function Characterization of Directionally Solidified LaB_6 –VB_2 Eutectic, Ultramicroscopy **183**, May 2017

Chapter 4
Modern Developments in Ba Oxide Cathodes

Georg Gaertner

Abstract The oxide cathode was the predominantly used cathode in radio electron tubes, electronic tubes, cathode ray tubes and in gas discharge lamps from about 1940 to 2010. A new rise of oxide cathode research and development occurred in the nineties, partly driven by price pressure on CRTs, which gave a number of new insights and improvement directions wrapped up in this article. These new results concern accelerated life tests of oxide cathodes and life increase, their poisoning behavior and measures to increase electrical conductivity. This led to a more refined model of the oxide cathode. Oxide cathodes are nowadays still applied in RF and microwave tubes. Since thinner oxide layers allow higher dc loadability, also alternate preparation methods will be discussed, which could eventually open new application areas. Some of the concepts are also discussed in the context of scandate cathodes.

4.1 Introduction

The oxide cathode was one of the most used cathodes in radio electron tubes, electronic tubes, cathode ray tubes and in gas discharge lamps. At the end of the 20th century about 700 million oxide cathodes were produced for CRTs and more than 7 billion oxide cathodes for gas discharge lamps. These numbers have strongly decreased since then, due to the replacement of CRTs by LCDs and of CFL lamps by LEDs. The major part of research and development on oxide cathodes was carried from 1930 to 1970, with a new rise in the nineties, partly driven by price pressure on CRTs. The knowledge at that time is wrapped up in the book of Herrmann and Wagener, "The Oxide Coated Cathode" [1] and in several longer articles by Nergaard and his colleagues from RCA [2–4] and Zalm [5], which partly put more emphasis on semiconductor emission models [2, 3, 5]. The more recent activities gave a number of new insights, which are compiled in this article and in shorter reviews by Barratt and Jenkins [6] and Gaertner [7]. One of the insights was, that the oxide cathode has a memory. F. i. when an oxide cathode gets poisoned and recovers completely, this will

G. Gaertner (✉)
Consultant, Aachen, Germany
e-mail: georg.f.gaertner@t-online.de

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*,
Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_4

eventually lead to a shorter operating life. It was characteristic for oxide cathodes, that one only gets the same results when the conditions over time are completely identical. This sets new limits on the experiments. The oxide cathode nowadays is still a. o. applied in microwave tubes of lower power. Some of the concepts are also discussed in the context of scandate cathodes.

4.2 Historical Development and Model Development

More than 100 years ago in 1903 Arthur Wehnelt discovered the high emission of alkaline earth oxides on heated platinum wire bases, especially of BaO and CaO [8–10]. These systematic investigations of various oxides had been triggered by the observation, that contaminated Pt wires showed lower cathode fall. When varying temperature Wehnelt found, that the Richardson equation was not only valid for metals but also for oxides. There were a lot of investigations from 1920 on, which profited from improved vacuum conditions, and proved, that the emission of the cathode is due to an excess of metallic Ba in the oxide, which is produced during cathode activation. This excess Ba was thought to be present in the form of a monoatomic Ba film, which also exhibits a much lower Ba evaporation rate than bulk Ba [10]. Since electrical conductivity of oxide cathodes is a limiting factor, different models of conduction through the bulk semi-conductor, through the pores and also surface conduction have been proposed, which in the end all contribute to a different degree. Due to further improvements of the base vacuum in tubes f.i. by the introduction of getters and by the use of alkaline earth carbonates for preparing the oxide coating, in the 1920s and 1930s the commercial application of oxide cathodes in tubes spread rapidly. In the 1950s, the role of the activators in the cathode Ni base was investigated in more detail and it was shown, that Ba was generated via an exchange reaction of the activator with BaO and that the process was controlled by activator diffusion and supply [2–4, 11–15]. With the advent of CRTs, oxide cathodes were the first choice because of their high efficiency also as indirectly heated cathode. The emission efficiency of directly heated oxide cathodes with about 40–100 mA/W was much higher than that of thoriated tungsten (6–30 mA/W) and tungsten cathodes (1.7–4 mA/W) in commercial valves in the 1950s. The efficiency values of course depend on cathode and heater design, on rated life and on emission capability especially of improved later cathode versions. H. Boumeester of Philips in 1937 quotes higher values of 3–8 mA/W for tungsten cathodes, 80 mA/W for thoriated tungsten and 200–300 mA/W for (earth alkaline) oxide cathodes [16], but also here the relative improvement for oxide cathodes is important.

In their review in 1951 Herrmann and Wagener [1] give a range for the dc emission current density of 0.1–0.3 A/cm^2 and peak values for pulsed emission of 10–30 A/cm^2 for state of the art oxide cathodes at that time. A lot of basic investigations and compilation of knowledge was carried out at Philips [5–7, 11] and at RCA by

Nergaard [2, 3] and Chin et al. [17]. In the 1960s the dc emission current density was still limited to 0.5–1 A/cm^2 in order to avoid excessive heating and destruction of the coating.

Due to their low cost there was a strong drive in research and development for higher dc loads in CRTs. In 1986 Saito et al. from Mitsubishi [18] introduced scandia particle doping in triple BaSrCa-oxide coatings in order to increase the limiting dc current density to 2 A/cm^2. Derks from Philips in 1987/1992 replaced scandia finally by rare earth atomic doping of the Ba·Sr·oxides, especially using co-precipitated europia or yttria, which were called SF (super fine) cathodes [19]. It was later shown by Gaertner et al., that the electrical conductivity and hence the dc loading capability is increased in proportion to the dopant concentration, yet at the expense of increased oxygen poisoning sensitivity of the oxide cathode [20]. Mitsubishi have presented work on a fine-grained tungsten coating of nickel caps, increasing activator diffusion that enabled a current density of 3.6 A/cm^2 [21]. As a further measure to overcome the conductivity limitation, Philips introduced nickel filament additions to the coating, which as a further beneficial effect reduce the cut-off drift in CRT applications [22]. This Ni addition together with increased activator concentration in the Ni base was one of the distinctive features of the "Oxide Plus" cathode of LG. Philips displays [23] and enabled a dc-loading of up to 4.5 A/cm^2 during life. The increase of the dc loadability of oxide cathodes as a function of time is shown in Fig. 4.1.

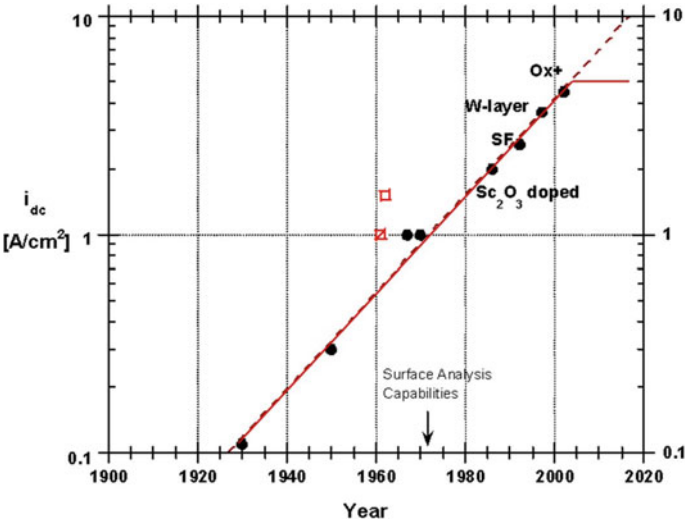


Fig. 4.1 Improvement of dc loadability of Ba oxide cathodes versus time (typical cathode life t_L > 10 kh, the red squares [80] have $t_L \geq 5$ kh); based on Gaertner et al. [10]

4.3 Peculiarities of the Oxide Cathode

In contrast to thermionic cathodes with a metal bulk, oxide cathodes have peculiar properties. These are above all related to their limited conductivity, since the oxide grains are semiconducting, and their porous structure. Including this porous oxide layer, the oxide cathode consists of the principal regions: base metal—interface layer—oxide layer—surface dipole layer/surface (according to Nergaard [2]).

1. There is a field or voltage limitation when operating oxide cathodes in a given diode or gun configuration, since otherwise sparking and crater formation occurs and the oxide cathode is destroyed.
2. In dc operation mode a slow decay of the dc emission current level at a given voltage happens, till some equilibrium value after e.g. 1 min is reached. In general the stable dc emission levels are in the order of 0.5–4 A/cm^2. Pulsed emission current densities are much higher than stable dc current densities, which in first order is not caused by anode heating. The pulsed levels decrease to dc level when the pulse length is increased [7].
3. Free adsorbed Ba is needed for the oxide cathode to function. This is provided by the so-called activators (Si, Mg, Al, W, etc.) embedded in the cathode Nickel base. After conversion of the carbonate layer to the oxide, the activators diffuse to the surface and to the pore surfaces of the oxide layer and form free Ba via an exchange reaction of the activators with BaO, most of this happens on top of the Ni interface. Ba is both needed for the low work function at the cathode surface and also for the conductivity of the oxide, since it can provide a conductive Ba monolayer at the grain surfaces along the pores, can eventually modify the semiconducting properties of the inner surfaces and can also be helpful for generation of an electron gas within the pores. In this context also a relation between emission and conductivity of the oxide layer was observed, yielding increased emission with increasing conductivity according to Hannay [12] and others. The conductivity of the oxide cathode in the activated state is much higher than of the oxide in the non-activated state.
4. High to ultrahigh vacuum is needed to avoid emission poisoning of the oxide cathode. If poisoning and recovery happens, the oxide cathode shows a memory effect limiting cathode life. Hence it is very important, that the cathodes tested are all treated in the same way to get reproducible results.
5. Addition of rare earth dopants such as Sc_2O_3, Y_2O_3, Eu_2O_3 or others can improve oxide cathode properties, especially conductivity, but may also increase sensitivity to emission poisoning.
6. In the different semiconductor models the nature of either mobile donors or mobile acceptors (e.g. Ba vacancies) has not been proven and has the status of hypotheses. We will discuss it in context of the proposed models.

4.4 Oxide Cathode Preparation and Activation

A comprehensive overview of oxide cathode preparation and activation was given by Kohl ([24], 504–509) and by Herrmann and Wagener [1]. After assembly of the cathode base unit with Ni cap and heater, the coating is applied on the Ni cap. A

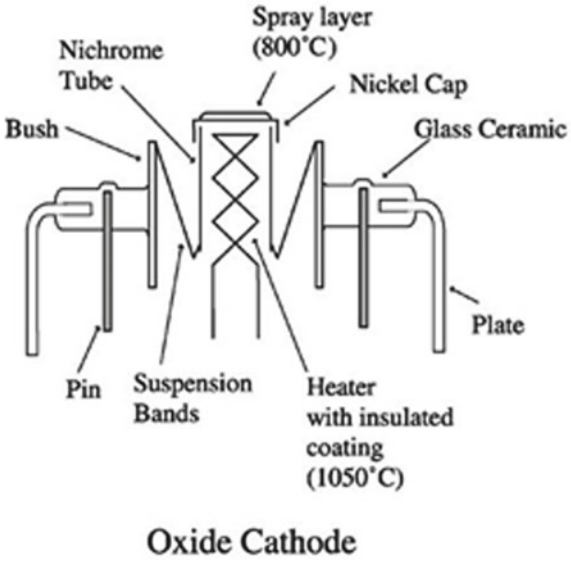


Fig. 4.2 Schematic cross section of a Philips oxide cathode unit according to D. Barratt et al. [6]. Courtesy of Springer

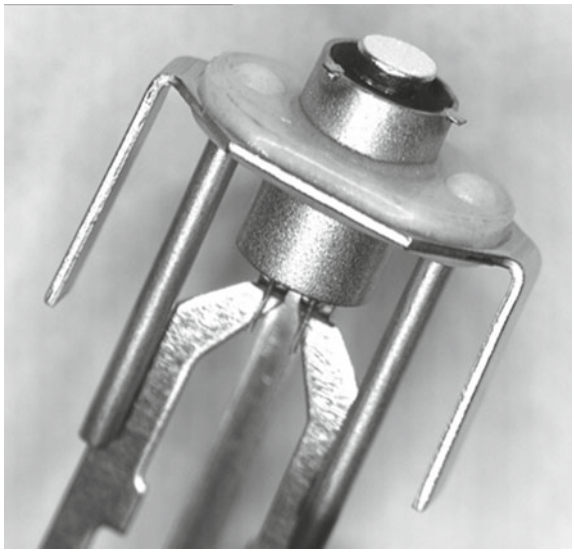


Fig. 4.3 Philips 0.65 W oxide cathode unit. The cathode surface diameter is 1.2 mm; from [7], Springer 2008, p. 440; see also [38, 103]

schematic view of an oxide cathode is given in Fig. 4.2. The main elements are the oxide coating on top of the Ni cap and the heater in a Nichrome tube sitting below. A photo of a 0.65 W oxide cathode unit is shown in Fig. 4.3 (see [7, 38, 103]). The coating is usually applied via a suspension fluid, which typically contains nitrocellulose, ethyl alcohol, amyl- or butyl-acetate with a total solids content of around 12%. The Ba-, Ca-, Sr-carbonate particle sizes of the coating are in the order of 1–10 μm. The mixture is ground in a ball mill for about 20 h and the viscosity of the milled mixture is quoted to be 38 ± 1 centipoise [24]. Most of the commercial tube manufacturers produced their own coating and suspension composition. The coating mixture is usually applied to the nickel cathode base by spraying with a spray gun while a number of cathodes are mounted in a spray frame. Usually this is done in several passes of the spray gun, with intermittent drying of the cathodes in a tunnel (e.g. procedure of Thomson, communicated by J.-M. Roquais). The type of nozzle used for the gun, the air pressure applied to the gun, the distance of the gun from the cathodes and the relative amount of humidity in the spray chamber are all important parameters that determine the quality of the resulting coating. Typical coating densities (density of the carbonates about 4 g/cm^3) are in the order of 0.6–1.4 g/cm^3 [1, 2, 24] and the coating weight may range from 2 to 8 mg/cm^3. Dependent on cathode geometry, other coating techniques used are e.g. cataphoresis, dip coating, film casting or use of transfer tapes.

After deposition it is necessary to remove the nitrocellulose binder. Here one of the recipes is to dry the carbonate (+oxides) for 12 h at 110 °C before use and allow to cool it in a dry atmosphere. McNair has shown, that complete elimination of the binder requires firing of the coated cathode at 200–250 °C in a stream of flowing air or in an oxygen atmosphere for a period from 5 to 15 min [24]. After 1944 it has been practice in Europe to bake the coated cathodes in air at 400 °C to break down the binder, as it was found that baking at 200 °C left some carbon in the coating. The temperature of the binder combustion has also influence on the mechanical strength of the layer [24]: Burn-off at 160 °C < T < 400 °C strongly reduces mechanical strength, whereas below 160 °C and above 400 °C it is preserved. To remove any oxide film formed during such a treatment on the Ni base, a second firing at 500–550 °C in hydrogen is sometimes used. After this treatment the color of the cathode should have changed from gray via a patchy blackish gray to a pure white, which also is a measure of success.

After removal of the binder (1) the next two steps of cathode activation are (2) conversion of the carbonates to the oxides and (3) partial reduction of the oxides at the metal coating interface to produce free Ba throughout the coating by diffusion. The conversion of the carbonates (2) takes place at brightness temperatures between 850 and 900 °C: onset of breakdown may be observed above 750 °C. During conversion the emissivity of the cathode decreases and the heater wattage eventually has to be adjusted. Conversion should take place within a few minutes and is accompanied by a fall of the system pressure from about 10^{-3} mbar peak to 10^{-6} mbar at the end of conversion. If the cathode behaves differently, this is generally linked with poor emission. It has to be noted, that in general mixed crystals of the oxides are obtained [1, 26, 27], in case of Philips and Matsushita (Ba, Sr)O.

After completion of the conversion, activation (3) is achieved by raising the temperature of the cathode for a few minutes to 1000–1200 °C, depending on type and base metal. This may be accompanied by a transient pressure rise to about 3×10^{-6} mbar. A dc diode voltage may be applied until a current density of ≥ 0.1 A/cm^2 is drawn, while keeping the pressure below 10^{-6} mbar. Subsequently the temperature is reduced to the operating range of 770–830 °C for stabilizing emission. In general, aging is done after seal-off. Some manufacturers do conversion and activation after seal-off.

Of course there is the possibility of choice to draw current from the cathode on the pump or afterwards. It is feasible to do "thermal" activation on the pump and "electrolytic" activation after seal-off and getter activation. An exposure to air subsequent to activation usually leads to severe poisoning by formation of Ba/Sr/Ca-hydrates. Here maintaining a cathode temperature of 150 °C prevents conversion to the hydrates and only hydroxides are formed. Flaking of the coating is thus prevented [24].

With respect to Philips procedures typical values for a 0.65 W oxide cathode activation and aging are given below in Table 4.1 both for diode and gun (45AX) configuration [54]. The first 3 steps are the conversion steps, then followed by voltage assisted activation:

Since the activation in the diode configuration occurs at a typical cathode to anode distance of 0.25 mm, the diode values are about a factor 3.3 higher than the grid 1 values (cathode to grid 1 distance about 80 μm). In some variants in case of ac voltages rectified ac voltages have been used.

The application of low dc voltages already strongly reduces the risk of detrimental flashover. Yet the ac currents during the application of ac-voltages have to be limited and hence load resistors (or load lamps) have to be used between voltage sources and grid 1 and grid 2, e.g. 820 Ω or 1000 W (11 W). Also during operation sparking can happen and can lead to partly destruction of the oxide coating, especially to

Table 4.1 Steps of a typical conversion and aging process for oxide cathodes (see M. Pralle, thesis for diploma, Aachen [54])

Duration t [min]	Heater voltage [V]	Cathode temperature (Mo–Br.)	In diode: voltage U [V]	In gun: grid 1 voltage	In gun: grid 2 voltage
ca. 2	3.0	450 °C	–	–	–
ca. 2	4.0	600 °C	–	–	–
ca. 8	6.0	725 °C	–	–	–
3	7.3	800 °C	0	0	0
5	9.5	960 °C	0	0	0
5	9.5	960 °C	50 V_{dc}	15 V_{dc}	15 V_{dc}
5	9.5	–	70 V_{dc}	21 V_{dc}	21 V_{dc}
5	9.5	–	90 V_{dc}	27 V_{dc}	27 V_{dc}
30	9.5	–	150 V_{ac}	70 V_{ac}	90 V_{ac}
35	9.5	–	15 V_{dc}	5 V_{dc}	90 V_{ac}

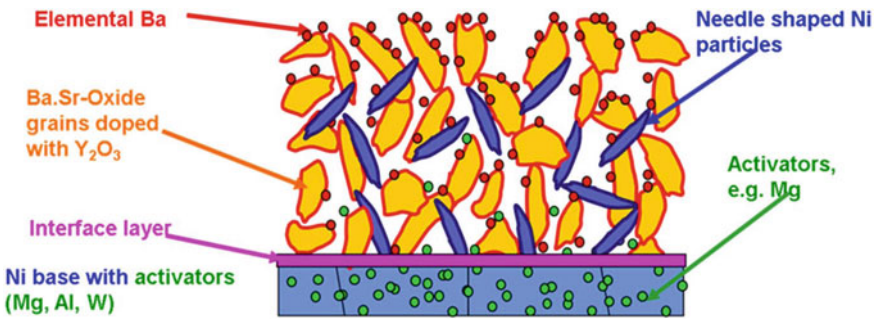


Fig. 4.4 Schematic view of the cross section of an oxide plus (cermet) cathode of Philips, (original by the author in 1999, presented at ITG 2004, compare [28])

crater formation and exposure of the bare Nickel with negligible emission (own investigations). Eisenstein [27] has determined sparking limits (typically > 2–3 kV) for Ba/Sr oxide cathodes—here with Si doped Ni base and a synthetic Ba-Silicate interface layer of 6 μm-as a function of time. At low temperatures (low current densities) the sparking limits are much higher than at high temperatures and higher current densities. Cathodes without Si in the Ni base show higher limits. For different oxide cathode types these limits have to be newly determined.

It should be noted, that different types of oxide cathodes such as the oxide cathode of Mitsubishi doped with fine scandia particles, or the SF oxide cathode, which is doped with superfine (SF) additions of rare earth oxides, or the oxide plus cathode of Philips with Ni particle additions, are all prepared with no real change in the rather cheap manufacturing procedures. The main changes are introduced in the materials used as for oxide plus [22, 23, 28, 65, 66, 101, 103]: rare earth oxide additions to the double carbonate, use of cathode Ni with increased activator content and addition of acicular Ni particles to the Ba·Sr-carbonate. A schematic cross section of an oxide plus (cermet) cathode of Philips is shown in Fig. 4.4, illustrating the high porosity of about 70% of the oxide layer and the positive effects of Ba surface diffusion on the semiconducting grains and of the Ni particles increasing conductivity.

4.5 Emission Characteristics and Work Function

4.5.1 Pulse Emission Decay

It has already been noted, that pulsed emission of oxide cathodes reaches much higher values than dc emission due to the limited conductivity of the porous oxide layer. The transition region from short pulsed emission to dc was investigated by Eisenstein [29], Sproull [30] and by Nergaard [2, 31] and others. Nergaard e.g. showed, that the pulsed levels decrease to dc level when the pulse length is increased up to 1 s

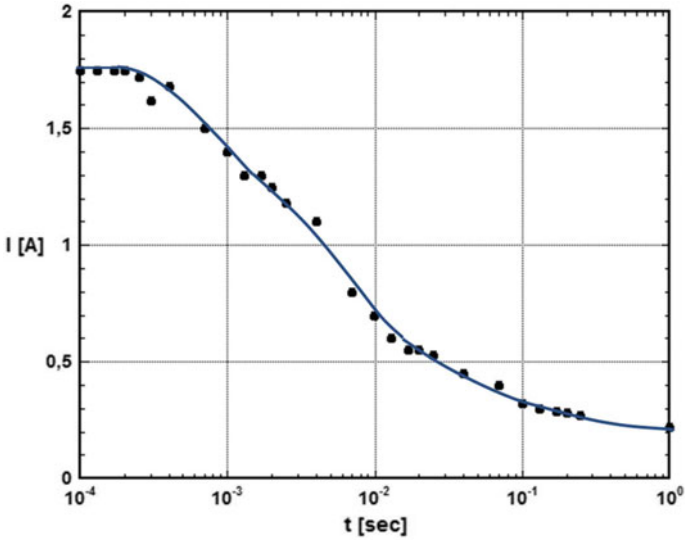


Fig. 4.5 Current decay of an oxide cathode in a diode on application of a rectangular pulse (237 V) of t sec duration (max. t = 1 s), repetition rate 2/min, T = 1190 K; estimated cathode area 2 cm^2; according to Nergaard [2], RCA Review

as can be seen in Fig. 4.5 [2]. Of course the pulse decay rate depends on the type of oxide cathode and on other test conditions. It increases with higher temperature and higher pulse emission current density [29]. Yet there are also other interfering effects, especially the vacuum conditions and degassing of the anode. According to Nergaard the decay is caused by mobile donor depletion, which will be discussed later.

4.5.2 Typical Experimental Procedures to Determine the Emission Parameters

For a comparison of oxide cathode results of Philips with other groups a short description of Philips measurement procedures and of the evaluation steps is given and methods to evaluate the emission parameters are described.

The oxide cathodes were usually mounted as 0.65 W units (Philips) or other units with heaters in a close spaced diode configuration with distance d = 250–300 μm in an ultrahigh-vacuum system with a massive Ta anode. The UHV system was usually baked at about 140 °C before activation and start of the emission measurements (pressure < 5 × 10^{-8} mbar). The measurements were computer-controlled. The standard measurements comprised the determination of pulsed *I/U* characteristics, with *U* up to 800 V and pulse lengths of 12–20 μs (rectangular pulser) or 15 μs

or 50 μs (saw-tooth pulser), sampling frequency 50 Hz, which were averaged over several sweeps. The pulsed current-voltage (*I*-*U*) characteristics were measured at different cathode temperatures between 650 and 930 °C (Heater voltages in the range ~4.8–8.5 V). DC *I*-*U* characteristics were determined by stepping dc voltage from 20 to 100 V (this limit set to avoid sparking). Eventually this range can also be measured with the voltage pulsers, but since dc loadability needs to be known, dc measurements are preferred and a further continuous dc measurement at 100 V for 1 min is added.

Examples of temperature dependent *I*-*U* characteristics are given in Fig. 4.6 [54]. The results of a pulsed diode characteristic emission measurement are usually plotted in a double logarithmic diagram of lg *I* versus lg *U*. In this diagram the space charge limit (SCL) gives a straight line with slope 1.46 (instead of 1.5, taking into account second order correction to the space charge law). In the lower part the characteristic is identical with the space charge limit, at sufficiently higher voltages it starts to deviate, which marks the onset of saturation. Within Philips and also in the respective publications the 10% deviation from SCL is used as a practical measure of the saturated emission current (zero surface field emission current). This is a very good approximation for practical use of the cathode in the SCL regime, the error to theory being only 5–10%, see Manenschijn et al. [32]. It can be easily determined as the crossover of 0.9 × SCL line with the *I*-*U* characteristic (see Fig. 4.7a), and thus one obtains $j_{10\%} = I_{10\%}/A_{\text{cath}}$ (where A_{cath} is the emitting surface area of the cathode). Other

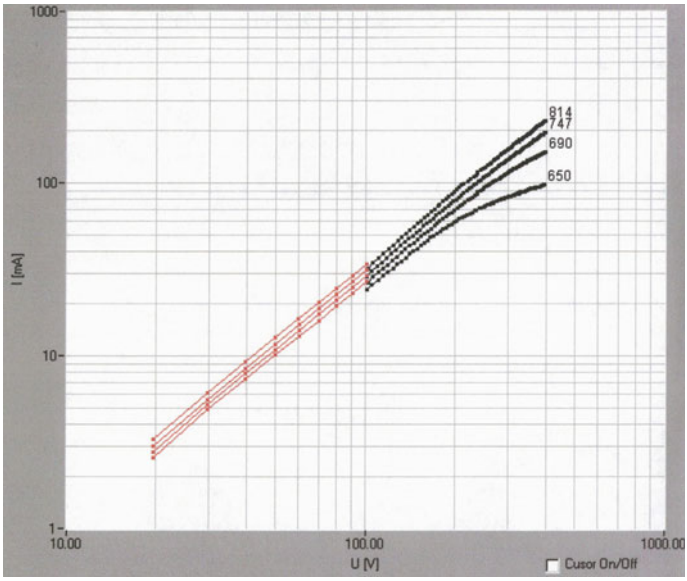


Fig. 4.6 Current voltage characteristics of Philips SF oxide cathode at different temperatures (650–814 °$C_{\text{Ni–Br.}}$) in double logarithmic scale. The red points are dc measurements, the black dots are pulsed measurements (Pralle [54]). The slope of dc and pulsed measurements below 150 V is space charge limited

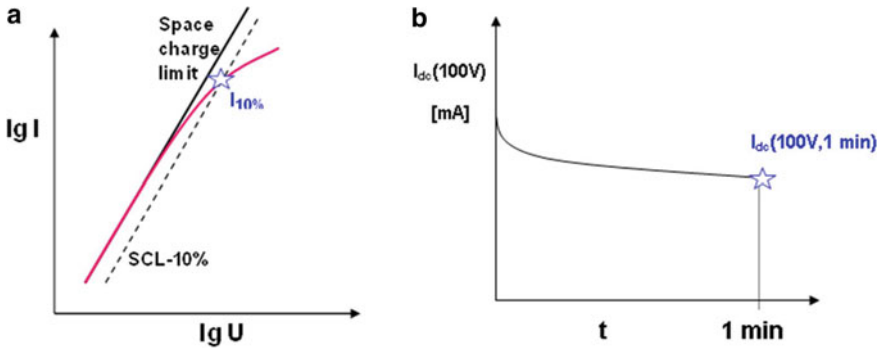


Fig. 4.7 **a** Schematic illustration of the determination of $I_{10\%}$. **b** Continuous I_{dc} measurement for 1 min at 100 V to judge critical j_{dc} current density

possibilities e.g. are to determine the point of inflection in the characteristic or, if the saturation behavior is known e.g. according to Schottky, determine it from a Schottky plot (which is not a good option for oxide cathodes due to their anomalous Schottky behavior). The temperatures are usually measured as °C Ni-brightness with optical pyrometers or via a heater current versus temperature calibration and are corrected for true temperature. The emitting area A_{cath} was determined from the diameter D_c of the oxide coated circular area on the cathode surface: for Philips oxide cathode it is 1.2 mm for 0.65 W units and 1.4 mm for 1.2 W units.

Since it is well known from oxide cathode theory (see Sect. 4.5.2. and Fig. 4.5) and also from measurements, that due to conductivity limitations of the oxide coating (semiconductor) the initially high pulsed or sub-second dc emission decays to some equilibrium value, we monitor the dc emission current at 100 V over 1 min in order to detect a current decay. If we observe a decay reaching equilibrium within <1 min (see Fig. 4.7b), the I_{dc}(100 V, 1 min) value usually is identical with the actual critical dc emission current as described by Nergaard [2]. For a well activated cathode during early life I_{dc}(100 V) stays constant over 1 min, which means that the critical dc current is at least higher than I_{dc}(100 V). Approaching the end of life typically a significant decay is seen. At Philips research the life end specification for I_{dc}(100 V, 1 min)/A_{cath} was set to 2.0 Acm^{-2} w.r.t. the application in CRTs and was in line with other CRT manufacturing companies.

Examples of continuous dc measurements up to 220 V for a SF cathode and an oxide plus cathode of Philips [28, 65, 66, 101] are given in Fig. 4.8, which are record data for oxide plus above 100 V.

It has to be noted, that dependent on the oxide layer conductivity the cathode temperature may rise by ohmic heating during the 1 min time interval.

From the dependence of $j_{10\%} = j_{\text{sat}}$ on true temperature the (Richardson) work function $e\Phi_R$ of the oxide cathode and the Richardson constant A_R can be determined via a Richardson plot, derived from the Richardson equation [7]:

$$j_{\text{sat}} = A_R T^2 \times \exp(-e\Phi_R/kT) \tag{4.1}$$

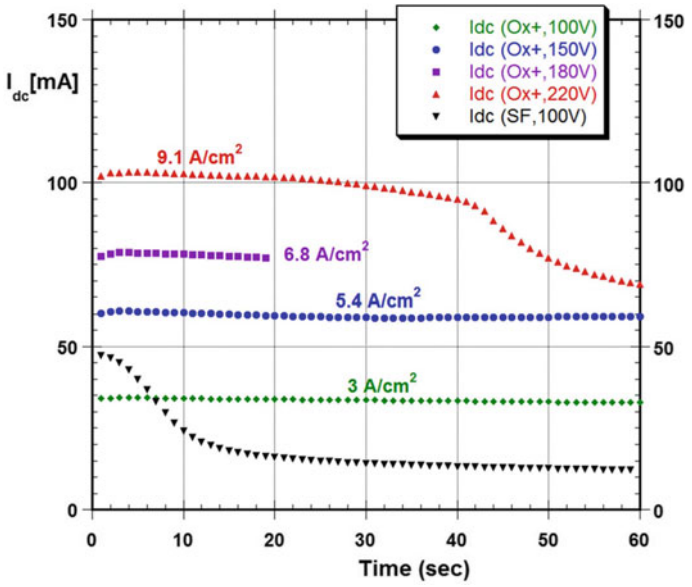


Fig. 4.8 Top example of continuous dc measurements of 1 min duration for a SF cathode of Philips doped with yttria (black triangles) and an oxide plus cathode of Philips [28, 65, 66, 101] doped with yttria, double activator concentration and acicular Ni particles at 750 °C. For oxide plus dc emission remains stable up to about 6 A/cm^2

In this plot ln (j_{sat}/T^2) is plotted versus $1/T$, corresponding to (4.1) written as

$$\ln\left(j_{sat}/\,T^2\right) \;=\; \ln A_R \;-\; e\Phi_R/\,kT \tag{4.2}$$

Thus from the slope the work function and from the axis interception the Richardson constant can be calculated. It should be noted, that for the oxide cathode as a semiconductor the Richardson-Fowler equation is valid (4.3). We will discuss it in more detail in the paragraph on models, but since the $T^{5/4}$ factor is only a weak contribution to the total temperature dependence, we use (4.1) for comparison with other cathode types.

For a one donor level semiconductor (Wilson model) [33], the Richardson equation has to be modified according to Fowler:

$$j_0 = B(1-r)n_0^{1/2}T^{5/4} \times \exp(-(\mathrm{X} + \Delta E/2)/\,kT) \tag{4.3}$$

Here B is a constant, r is the reflection coefficient, n_0 is the donor density, X is the electron affinity and ΔE is the donor level with respect to the conducting band edge. A further difference to standard semiconductor theory is, that the charge distribution in the oxide coating is not generated by immobile donors and acceptors, but by mobile

electrons and donors [3, 17, 34, 35] and hence is influenced by field penetration and fields induced by the surface states. One consequence also is, that a critical current can be derived under the assumption that the constant number of donors is redistributed under the action of current flow. Increasing average current then causes the the electron density to vanish at the positive surface of the semiconductor at a critical (dc) current density j_{crit} [34, 35]:

$$j_{\text{crit}} = (2kT/e) \times \left(n_0 e \mu_- / d\right) \tag{4.4}$$

where n_0 is the average density of the donors, μ_- is the mobility of the electrons and d is the thickness of the oxide semiconductor. This can quantitatively explain the observation of limited dc–loadability of the oxide cathode.

For pure metals the Richardson constant is in the order of magnitude of the Thermionic constant = 120.4 A × cm^{-2} K^{-2}. In a small temperature range, by setting A_R = thermionic constant, the so called effective work function $e\Phi$ is obtained ($e\Phi = e\Phi_R + \alpha T$, α the temperature coefficient of WF, compare Chap. 3). A typical starting value for oxide cathodes is $e\Phi = 1.45$ eV, after life it increases to 1.6 eV. For real cathodes, A_R and $e\Phi$ can differ from expected values, which may be caused e.g. by patchy surfaces, incomplete activation or deterioration on life, hence A_R and $e\Phi_R$ are used as fit parameters.

It is interesting to see, that w.r.t. short pulsed emission current, where the pulse length is in the order of 10 µs, pulsed emission of well-activated Ba-oxide cathodes of top performance is nearly identical with pulsed emission of LAD top-layer Scandate cathodes of Philips [36], as can be seen in Fig. 4.9. It surely points to the fact, that Ba monolayers on metals, eventually as Ba-O dipoles, cannot further reduce the work function below 1.8 eV, and hence similar to oxide cathodes also in the case of top performance Scandate cathodes the Ba monolayer on the surface spreads on a semiconductor containing BaO below. We will discuss this further in context of emission model development.

4.6 Life Limiting Effects and Accelerated Life Tests of Oxide Cathodes

In view of the limited electrical conductivity and dc loadability of oxide cathodes, the cathode life in a vacuum tube under operating conditions is a very important parameter for any application. There are several life limiting mechanisms, but we will address here the three most important ones in the context of accelerated life tests, which are needed to evaluate for example improvements in material composition and structure for application in CRTs or other vacuum tubes. In this chapter we will discuss accelerated life by increased operating temperature, accelerated life by increased continuous dc-load, and finally life acceleration by intermittent

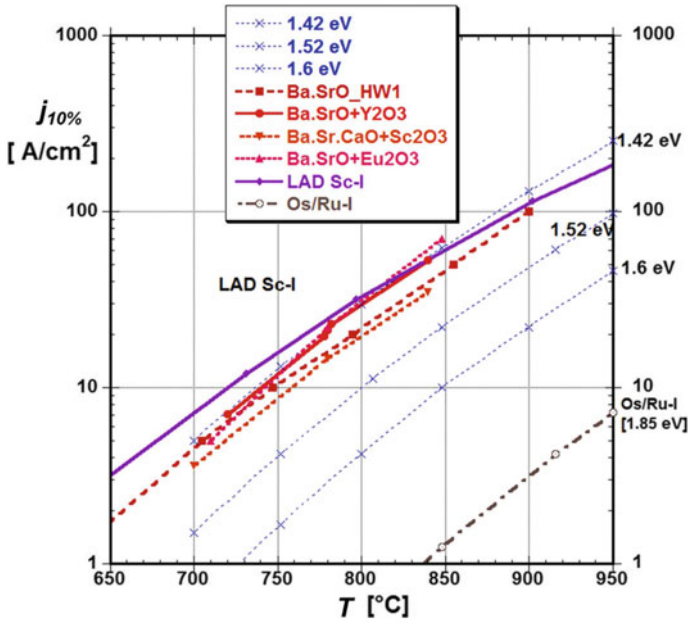


Fig. 4.9 Saturated pulsed emission current density versus true temperature for several oxide cathode types (Ba·SrO + Y_2O_3: Philips SF, Ba·SrO + Eu_2O_3: Philips and MEC; Ba·Sr·CaO + Sc_2O_3: Hitachi; Ba·SrO_HW1 according to Herrmann and Wagener [1]) compared with LAD top-layer Sc/Re-I and Os/Ru-I cathodes of Philips [36]. The theoretical curves for work functions of 1.42, 1.52 and 1.6 eV with A_R = thermionic constant are also shown for reference

or continuous poisoning and quantify them for Ba·Sr-oxide cathodes. We will use a dc loadability of 2 A/cm^2 as the general criterium for end of life in the experimental evaluation, which also means, that we will start at higher dc current density level. We will see later, that this also implies pulsed emission decay to the same level.

4.6.1 Accelerated Life by Increased Operating Temperature

Of course in this context also the cathode model and the operation mechanisms play a decisive role, especially the role of the activators in the cathode Ni base, and this and the temperature dependence was already well studied in the 1950's and 1960's.

Ba generation is mainly caused by reaction of the activators-contained in the cathode nickel base-with BaO after diffusion to the surface. Typical activators used are Mg, Al, Si, Zr and W. The free energy change ΔF of Ni for oxidation in a stream of hydrogen and water vapour as given by Allison and Samelson [14] is positive and hence Ni is less prone to oxidation. In Table 4.2 the corresponding ΔFs are compared [14].

Table 4.2 Free energy change for oxidation ΔF according to [14]

Oxide	Al_2O_3	MgO	NiO	SiO_2	ZrO_2
ΔF [kcal/mole]	−178.7	−68.7	+12.6	−71.8	−120.4

To be more general, the formation (reaction) enthalpy of the respective activator oxides (the most stable oxide phase in the temperature range of interest) has to be compared with BaO, SrO, CaO formation enthalpies ΔH (−131.0; −141.3 and −151.8 kcal/mole respectively) and needs to be more negative for the activator oxides in order to reduce the earth alkaline oxide. Eventually between Ba and Mg there is only a small difference in ΔH, the difference between Ba–Si and Ba–Al is much larger! We do not go into details here, since not the chemical reaction, but following Peterson et al. [13] diffusional supply of the activators from the base Ni is the limiting step, which was also proven by comparison of the experimental results of cathode life versus temperature with the theoretical predictions.

Since activator diffusional supply to the Ni surface and subsequent loss and hence also Ba generation and Ba loss e.g. by evaporation increase with operating temperature, cathode life will be shortened. Hence instead of running the cathodes in the UHV chamber at the usual operating temperature of 780 °C (true temperature), accelerated life tests are carried out at 880 °C true temperature, which typically will reduce life (e.g. referenced to a dc-load of 2 A/cm^2) from 20,000 h to about 2000 h, which is a reasonable testing time (see [37, 38]).

Via a theoretical model and experimental results of accelerated life-tests at elevated temperature life predictions for lower and for the operating temperature can be derived. The theoretical model is based on the temperature dependence of the diffusion coefficients of the activators in cathode Ni. The diffusion constants at 1050 K and the pre-exponential factors and diffusion activation energies according to Peterson [13] and Allison et al. [14] are listed in Table 4.3. From these parameters the temperature dependence of the diffusion coefficients has been calculated by $D = D_0 \times \exp(-Q_d/kT)$ and is depicted in Fig. 4.10. Essentially it can be seen, that the

Table 4.3 Diffusion parameters and diffusion coefficients of activators in cathode Ni according to Peterson [13]

Activator	Zr	Mg	Al	Si	Ti	Mn	Mo	W
Diffusion activation energy Q_d [kcal/mole]	48.3	56.6	64.0	61.7	61.4	67.1	68.9	76.8
Pre-exponential factor D_0 [cm^2/s]	0.034	0.44	1.87	1.50	0.86	7.5	3.0	11.1
D(1050 K) [10^{14} cm^2/s]	300	71.5	8.74	21.6	14.0	7.9	1.34	0.112
D(1050 K)/D(1143 K)	0.154	0.110	0.082	0.092	0.091	0.073	0.068	0.05

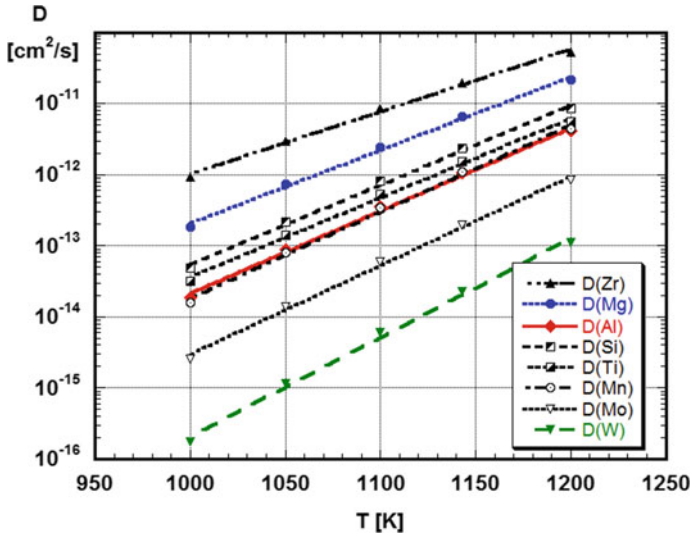


Fig. 4.10 Temperature dependence of diffusion coefficients of activators in Ni based on Peterson et al. [13] and Allison [14]; Mg, Al and W are contained in cathode Ni of Philips oxide cathodes; compare [37]

increase of diffusion coefficients is comparable for different activators for a temperature increase of about 100 K, which can be seen from the list of the ratios between the diffusion coefficients at 1050 K to $D(1143\ \mathrm{K})$ for the most important activators in Table 4.3.

This would correspond to a reduction in cathode life when increasing the temperature from 747 $°C_{Ni–Br}$ to 840 $°C_{Ni–Br}$ to a range between 15% and 5%, or between 11% and 5% for Mg, Al and W as activators in Philips oxide cathodes. Now the average life of Philips SF oxide cathodes at 840 $°C_{Ni–Br}$ is about 1600 h compared to 20,000 h at the standard operating temperature, which amounts to 8%, consistent with Al as the life-determining activator. Another calculation may illustrate this and can be used for estimation: the diffusion length x after 20,000 h is $x(Mg) = 72\ \mu m$, $x(Al) = 25\ \mu m$ and $x(W) = 2.8\ \mu m$. This means, that for a Ni base thickness of 100 μm, the Mg concentration has already drastically diminished, the Al concentration is about half of its initial value and the W diffusion is only efficient in a very thin layer at the Ni base top. If one now assumes cathode lifetime to be correlated to an activator loss of about 50%, lifetimes for the different activators may be predicted from $t_L(T) = (25\ \mu m)^2/D(T)$. This is shown in Fig. 4.11, where the straight (Al) and dotted (Mg) lines depict oxide cathode lifetime estimated on basis of about 50% activator loss from cathode Ni. It can be seen, that the experimental points (black circles) are consistent with Al diffusion in Ni as the life limiting factor. Hence possibilities to increase life are either an increase of Al concentration in Ni, or use of an activator with a diffusion constant a bit smaller than Al, or both. A

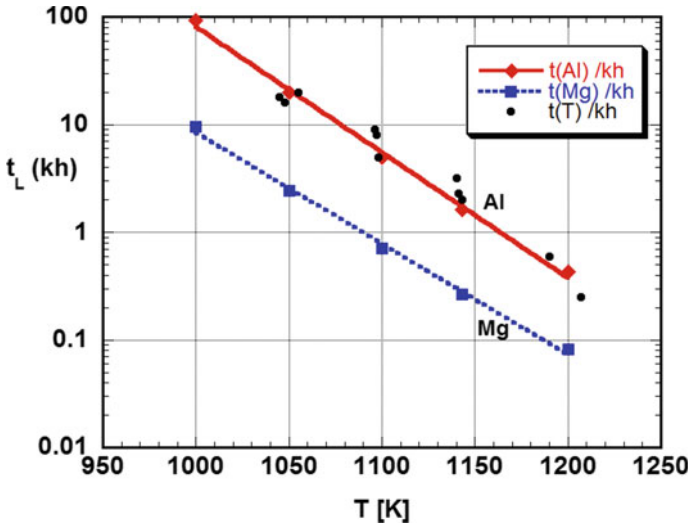


Fig. 4.11 Straight (Al) and dotted (Mg) lines: Oxide cathode lifetimes estimated on basis of about 50% activator loss from cathode Ni via $x^2 = D(T) \times t_{life}$, with $x = 25$ μm. The solid circles are experimental data points, which are consistent with life limitation by Al, according to [37, 38]. From G. Gaertner et al. [38]: Copyright 2003, with permission from Elsevier

much faster diffusor in Ni than the activators listed above, but also an oxide cathode activator is Carbon, which is mentioned here for completeness. C may always be present as an impurity, which can play an important role in initial life. The diffusion coefficient of C in Ni at 1050 K is nearly 3 orders of magnitude larger than for Zr.

Itazu [39] computed cathode life on basis of (Mg, Si) activator diffusion and of the Wilson n-type semiconductor model for emission dependence on donor concentration. The result is shown in Fig. 4.12 for the Ba concentration at the two temperatures 750 and 800 °C versus time. A plot with similar behaviour is obtained for pulsed

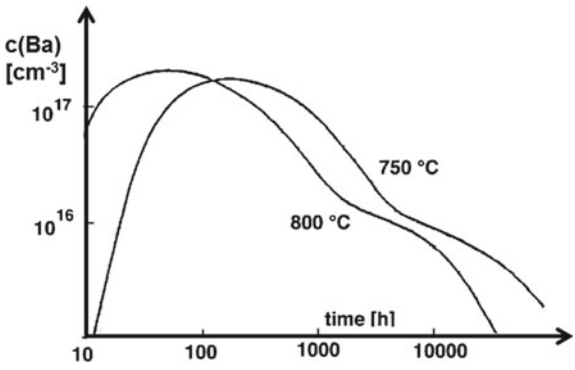


Fig. 4.12 Chronological change of Ba content c(Ba) according to model calculations of Itazu [39] in triple oxide cathode with 0.01 weight % Mg and 0.01 weight % Si activator concentration in cathode Ni

emission. Of course lifetime at 750 °C is longer. From the structure of the curve the first maximum is due to the fast activator Zr and the following declining plateau due to Si, which is hence determining overall life. Investigations with Zr single additive Ni alloys have also been conducted by Kern [15] . Here also a clear correlation could be seen between pulsed emission (at a continuous dc base level of 0.2 A/cm^2) and Ba production versus operation time at 1015 K.

Investigations of the oxide compounds formed at the Ni interface of course will depend on the activators used. For Mg and Si the compounds MgO and $BaSiO_4$ were observed at the Ni surface, where unfortunately $BaSiO_4$ exhibits a high interface resistance and can also lead to delaminations, a disadvantage of the Si activator. In case of Mg and Al activators and Ba·Sr·oxide Jenkins et al. [40] besides MgO also observed $MgAl_2O_4$ (also by Roquais et al., additionally Al_2O_3, see [41, 42]) and (Ba, Sr)Al_2O_4, whereas in literature so far only $BaAl_2O_4$ was claimed. These compounds give further evidence for the exchange reactions, where free Ba is produced.

As an instructive example an accelerated life measurement of a Philips SF oxide cathode (Ba·Sr·oxide doped with Y_2O_3) at increased temperature is presented. The operating temperature was set to 840 °C Ni–Br. instead of the standard 750 °C

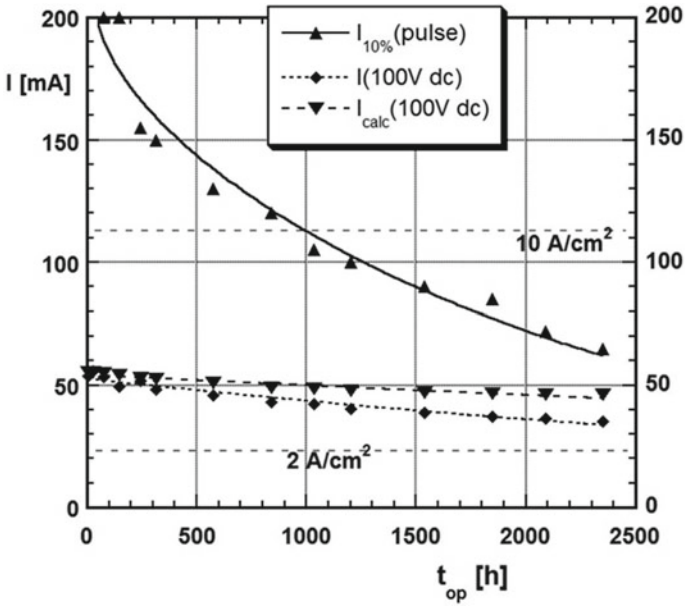


Fig. 4.13 Early life prediction in accelerated life tests: I(100 Vdc) and saturated emission $I_{10\%}$ (pulse) measured at 750 °C versus operation time (at 840 °C) for a Y_2O_3 doped Ba·Sr·oxide cathode (SF cathode of Philips). The calculated dc curve is based on diode distance enlargement by sintering of the oxide layer [37, 38]. From G. Gaertner et al. [38]: Copyright 2003, with permission from Elsevier

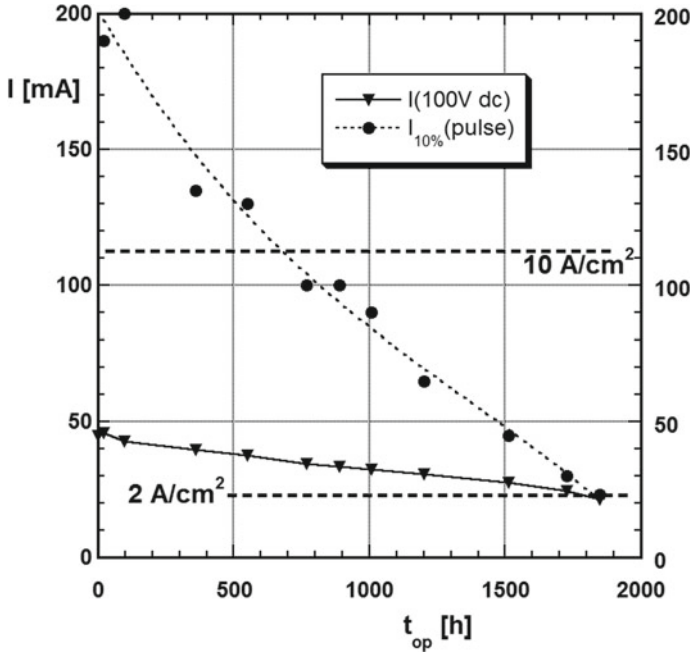


Fig. 4.14 I(100 V dc) and saturated emission $I_{10\%}$ (pulse) measured at 750 °C Ni–Br. versus operation time at 840 °C Ni–Br. (no load) of a Philips oxide plus cathode. From G. Gaertner et al. [38]: Copyright 2003, with permission from Elsevier

Ni–Br. During operation no load was applied. Measurements were usually done in 1 week distance at three temperatures (about 815 °C, 750 °C and 700 °C). The results obtained at 750 °C for $I(U_k = 100$ V dc; t meas. $= 1$ s), and $I_{10\%}$ (pulse) are depicted versus t_{op} in the diagram Fig. 4.13. The difference between $I(U_k = 100$ V dc; t meas. $= 1$ s) and $I(U_k = 100$ V dc; t meas. $= 1$ min) (not shown here) is less than 8%. $I_{calc}(U_k = 100$ V dc) is also shown in this diagram and was calculated on basis of the shrinking of the oxide layer by sintering. It is at least one contribution to the decrease of the dc current. The extrapolated cathode life is larger than 2500 h at 840 °C Ni–Br., when a dc level of 2A/cm^2 is reached.

For comparison an accelerated life measurement of a Philips oxide plus cathode (Y_2O_3 doped Ba·Sr·oxide with Ni particles and double activators in Ni) is presented in Fig. 4.14, also with 840 °C Ni–Br as operating temperature. The results obtained at 750 °C reference temperature (temperature decrease before measurement) for $I(U_k = 100$ V dc) and $I_{10\%}$ (pulse) are depicted versus t_{op}. The measurements of $I(U_k = 100$ V dc) after $t = 1$ min are nearly identical with the measurements at $t = 1$ s and are not shown in the diagram. Here we do not see an improvement of lifetime of the oxide plus cathode compared with the SF cathode. Yet we will see

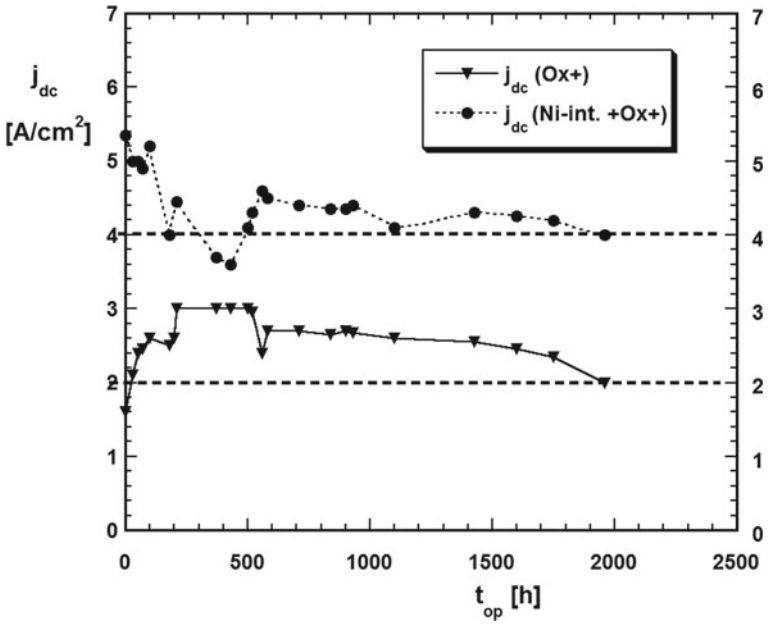


Fig. 4.15 Dc emission current density j_{dc} at 750 °C (Ni–Br.) versus operating time t_{op} at 840 °C (Ni–Br.) in an accelerated life test of standard oxide plus cathodes and oxide plus cathodes with a sputtered fine grained interlayer on cathode Ni (averaged results)

later, that oxide plus has the advantage of increased life after poisoning, which means it is more robust if vacuum pressures are a bit higher than 10^{-9} mbar. A further improvement of life is seen, if a fine grained Ni interlayer is coated on top of the cathode Ni, as done by Samsung [43] via sputtering or by Philips by laser ablation deposition of Ni or also by sputtering of Ni. This is depicted in Fig. 4.15 for standard oxide plus cathodes and cathodes with a sputtered Ni interlayer on top of cathode Ni. The ultra-fine nickel interlayer obtained by sputtering Ni on top of cathode Ni improves zero hour and life performance compared to oxide plus cathodes without this layer. The dc emission current density data are averaged results and have been obtained by accelerated life tests at 840 °C (Ni–Br.) without dc load, whereas the depicted measurements have been carried out intermittently at 750 °C (Ni–Br.).

A general question arising from these diagrams is, what are the reasons not only for pulsed but also for dc emission degradation? It may best be explained when looking qualitatively at the change of the *I/U* characteristics over time in a double logarithmic plot at a given temperature in a diode configuration as shown in Fig. 4.16 [44]. There are three effects superimposed:

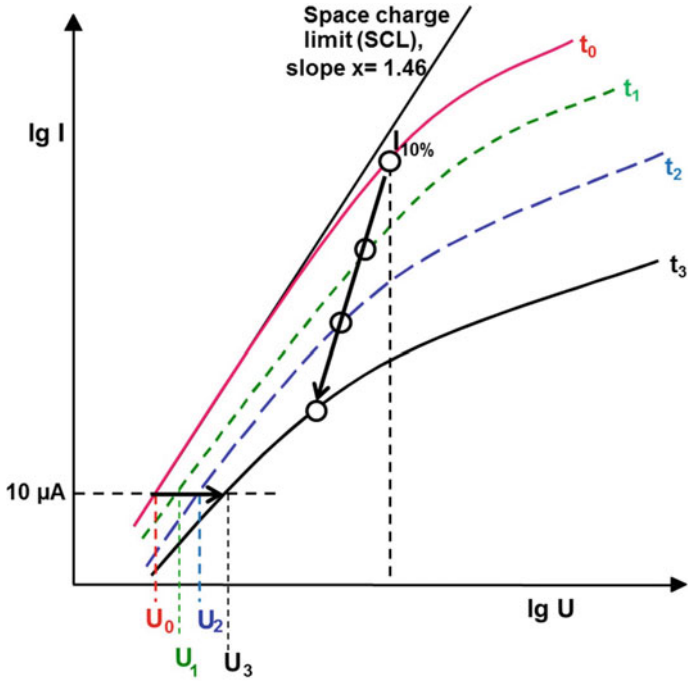


Fig. 4.16 Schematic change of current-voltage characteristics of an oxide cathode with increasing operation times $t_3 > t_2 > t_1 > t_0$ (t_0 after activation, t_3 near end of life) at the operation temperature or a higher constant temperature during an accelerated test [28, 44]; modified from G. Gaertner et al. [44], Copyright 2005, with permission from Elsevier

(a) the space-charge limit (SCL) is shifting to lower current values
(b) the slope of the "space-charge limited *I/U* line" in a log-log plot is decreasing, which means increasing non-uniformity via superposition of SCL sites and increasing portion of low-saturating sites
(c) saturated (pulsed) emission current is decreasing

Case (c) can be explained mainly by activator consumption as discussed at the beginning of this chapter, whereas (a) is dominated by sintering shrinkage of the oxide layer and coarsening of the grain structures, leading to a decrease of emitting area.

These changes by sintering are also the reason not to exceed the +100 °C limit above the operating temperature during accelerated life tests.

From Fig. 4.16 one can also explain features of the oxide cathode such as cutoff drift, which implies the increase of voltage $U_0 \rightarrow U_3$ needed in order to maintain a certain (minimum) current level, here e.g. 10 µA. One can also observe the decrease of saturation current $I_{10\%}$ with increasing time.

4.6.2 Accelerated Life by Increased Continuous DC-Load

It is of course well known, that voltage/load dependent effects play an important role for the oxide cathode, especially for the activation procedure (see Hermann and Wagener, "The Oxide-Coated Cathode, 64–67 [1]). Yet they claim, that voltage assisted activation is only necessary for "combined" cathodes, where a severe oxidation of the coating and the metal has been carried out before, but not for the uncombined cathodes, which are activated via reduction of the oxide by the activators. Yet drawing current is also necessary there for degassing of the anode. Apart from these masking effects, we have carried out an experiment, where we compared oxide cathodes with and without voltage assisted activation, but with the same temperature sequence. The only thermally activated cathode showed minor performance after activation and only reached the standard emission level after a delay of 100 h of operation.

In order to study the dependence on continuous load in detail, in accelerated life tests Philips SF oxide cathodes were operated at higher temperature with different continuous dc-loads in a UHV chamber (baked at 150 °C before activation) and compared with an oxide plus cathode, in order to quantify the influence of dc-load on life. The results are shown Fig. 4.17, where accelerated life of BaSr-oxide cathodes at 805, 840 and 870 °C (Ni–Br) operating temperatures is depicted as a function of continuous dc load. The Ni brightness temperatures given above correspond to 840,

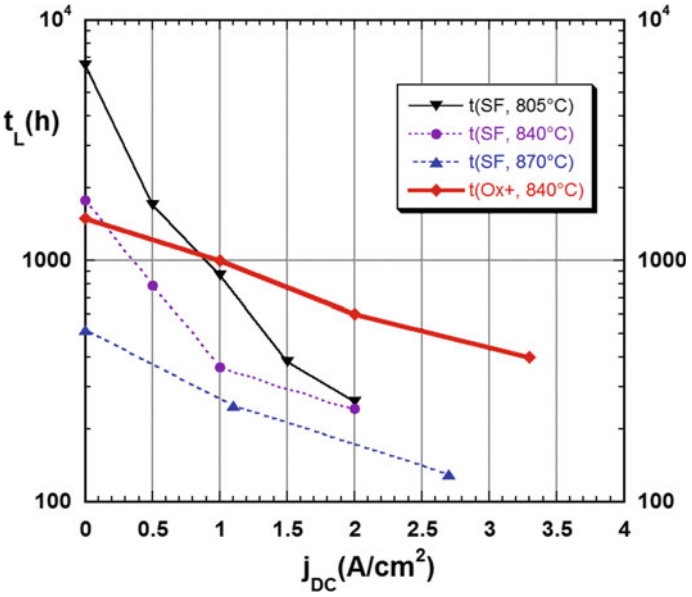


Fig. 4.17 Accelerated life test results as a function of continuous dc load at different (Ni–Br.) temperatures for SF and oxide plus cathodes of Philips (840 °C); from G. Gaertner et al. [44], Copyright 2005, with permission from Elsevier

880 and 920 °C true temperature, respectively. The oxide plus test was conducted at 840 °C Ni–Br. Of course the respective temperature rise induced by the continuous dc current was compensated by heater voltage adjustment. Despite the strong decrease of life with dc load (the dc life end criterium again at 2 A/cm^2) this is not representative for CRT conditions, where the average duty factor is about 10%. One can see, that the influence of continuous dc becomes dominating over the temperature dependence at higher dc load, but is less pronounced for oxide plus cathodes, which also exhibit higher electrical conductivity, as we will see later.

As can be seen from our comparative measurements on activation and operation with and without load, there is clearly an acceleration of transport phenomena over time due to current flow. This effect is directly linked to the electrical conductivity σ_{el} of the type of oxide cathode in operation and its decrease over time, being less pronounced for oxide plus cathodes with higher electrical conductivity σ_{el}. Since σ_{el} according to Chin et al. [17] is mainly based on the grain surface layers including Ba, it is quite evident that current flow along the grain surface can be most detrimental for oxide cathodes with low conductivity and hence cause accelerated deterioration, which is faster for the SF and slower for the oxide plus cathode of Philips, where more Ba is generated. Since Ba (monolayer) accumulation occurs on the cathode surfaces and on inner pore surfaces, this can also explain the relation between electron emission and electrical conductivity observed in several experiments. Yet compared with oxide cathodes of other manufacturers one should not forget, that despite of the same activator concentration in the Ni base, also the thickness of the Ni base will play a role and higher conductivities and longer life will be reached the thicker the cathode Ni base, if the activator concentrations are comparable.

This effect of current induced deterioration is not yet sufficiently understood, but probably is caused by a combination of several effects [1], including current-induced sintering and additional coarsening of the grain structure as observed by Gaertner et al. [28], by activator diffusion enhancement via electro-migration, by increased electrolytical Ba production at the interface layer to the Ni base and by field assisted Ba desorption also known from other references. There is also some evidence for an increased loss of Sr under current load. Less efficient electrolytic conduction may increase during operation, including Ba ion movement to the base, and electro-migration effects can play a role. According to Sproull [30], the high emission under load should lead to an increased loss of ionized Ba ad-atoms into the bulk. A systematic investigation of these current-induced effects is needed for oxide cathodes, in order to find a better understanding.

From analysis of the (Ba, Sr)O cathode composition over life, usually there is a loss of BaO in the unloaded case, f.e. by Ba evaporation or by Ba ion diffusion to the interface. Also an increase of Mg, Al and W can be found in the oxide coating; Mg is especially increasing at higher continuous dc current loads, also in the oxide plus cathode.

Despite of the fact, that bulk and ionic conductivity doesn't play a strong role in oxide cathodes, as can be seen from [17], some of the material present in oxide cathodes as possible ingredients in form of dopants or interface compounds can also belong to the group of solid electrolytes, e.g. ZrO_2, or γ-Al_2O_3 (Na+), or a mixture

of $ZrO_2 + Y_2O_3 + Sr_2O_3$ (O^-) or (Ba, Sr)Al_2O_4 or mixtures of BaO/SrO, MgO and Al_2O_3, as present at the interface (see Lexikon der Physik [45] and G. He et al. [46, 47]. This also implies that Ba++ is the ionic conductor moving to the Ni base.

The effect of electro-migration is more accessible. A good description on electro-migration in thin metal films is given by F. d'Heurle and P. Ho [48]. We may directly apply this work to the electro-migration of the activators in cathode Ni, which could explain the life acceleration with higher dc current densities. Unfortunately the typical electro-migration investigations are carried out at lower temperatures (range 140–300 °C) and high current densities (5×10^{-4} to 2×10^{-6} A/cm^2), in order that the electro-migration effects are not masked by increased grain boundary diffusion at higher temperatures. Yet if we extrapolate the temperature and current density dependence given in ([46], p. 283, Fig. 8.12) for apparent failure energies to the region 30 A/cm^2 (in Ni base) and 1000 K, comparative times to failure are obtained. This at least indicates, that there is probably a measurable effect in addition to thermal grain boundary diffusion. In electro-migration the driving force is usually considered to be the sum of two effects: the electrostatic interaction between the electric field and the ionic core of the atoms stripped of their valence electrons, and a friction force between these ions and the flowing charge carriers, which is often called the "electron wind" force. In metals, which are good electrical conductors, the electron wind force is usually dominant. Mass transport in form of the atomic flux J_l due to electro-migration is then proportional to the atomic density of the lattice N_l, to the diffusion coefficient D_l, the electric field E and the effective charge $Z_l \times e$ (T = temperature), as given in the following equation [48]:

$$J_l = N_l \times D_l / kT \times Z_l \times eE \tag{4.5}$$

The question now is, which quantities can we measure to assess compositional and structural changes including electro-migration effects under current flow? We can have two contributions: first, the activator diffusion can be enhanced, which will lead to increased Ba production and to an acceleration of activator consumption during operation, and second extra Ba and Sr could be generated e.g. at the interface via an electrolytically/current stimulated reaction.

In both cases we should observe increased Ba evaporation via increased Ba generation. In the second case only we would not have an extra activator consumption, and life should not be shortened too much, if no other effects like enhanced sintering get involved. Of course a superposition of both effects may be expected. An experimental approach could be based on mass spectroscopic Ba evaporation measurements as a function of continuous dc load, on analysis of Ba/BaO deposits on the anode and on monitoring of activator consumption via analysis of the cathode base Nickel.

In general, the application of this effect to life predictions is more on a heuristic basis and is also strongly dependent on the type of oxide cathode investigated, but is useful for judging the dc loadability in practical applications.

4.6.3 *Emission Life Limitation by Continuous or Intermittent Gas Poisoning*

Poisoning investigations of oxide cathodes, especially w.r.t. oxygen poisoning, usually yielded no consistent and really reproducible results [49, 50]. Typical was a continuing degradation of poisoning resistivity observed with every subsequent poisoning interval as can be seen from the investigations of Jenkins et al. [49] or Itoh et al. [51], where they number the sequence of the measurements. This can be explained for oxidizing gases from our current model, since free Ba or interstitial oxygen vacancies are oxidized. Pumping down again e.g. after oxygen admission does not lead to a permanent recovery, since changes in the bulk of the grains and in the interface layer remain, in contrast to the poisoning of *I* cathodes, where an oxygen overlayer is only loosely bound and is desorbed again. In order to recover, extra Ba has to be generated, which needs a certain accumulation time competing with losses; the interface layer is also growing. In the end there are only few reliable publications on oxide cathode poisoning, some of the results for different types are compiled in Fig. 4.18, in a plot of current I(p) over unpoisoned current I_0 versus partial pressure of poisoning gas introduced into the chamber. Among these results

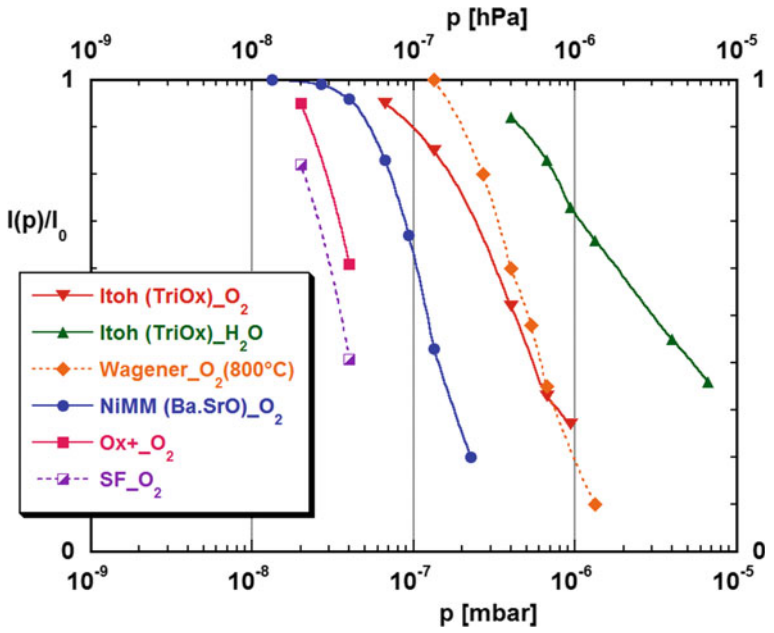


Fig. 4.18 Overview of O_2 poisoning results for different types of oxide cathodes: SF and oxide plus (Ox+, Philips), Ni matrix oxide cathode (O_2, see No. 6 in Fig. 9 of [50]) Jenkins et al. [49, 50], Wagener [52] and Itoh et al. [51] (triple oxide). I_o is unpoisoned, and $I(p)$ poisoned emission current at pressure p of the poisoness gas. The green curve is for H_2O poisoning.

are also the results of Jenkins et al. [49] for oxygen poisoning of a mixed matrix cathode consisting of 50% Ni particles and 50% Ba·SrO particles at 900 °C, which is of course not representative, but very well documented. Some results of Itoh et al. [49] for oxygen and hydrogen poisoning are also given, but they do not give any information on the temperature and one can only speculate, that it is in the usual region of operating temperatures between 750 and 780°C. Their investigations also suffer from the handicap, that they switched to the next pressure after 1 min, and this time is too short to wait for the equilibrium between poisoning and resupply, which in case of our investigations is typically reached after 40–60 min. Thus their cathodes seem to be more insensitive to poisoning than they really are. Yet the value of their publication lies in the comparison of different gases, which gives increasing poisoning sensitivity in the sequence $Ar < N_2 < H_2 < CO < CO_2 < H_2O < O_2$, $NO_2 < Cl_2 < SO_2$, where the last two lead to irreversible decay without any recovery. In the practical case in tubes O_2 is the most critical gas and therefore the results for Philips SF and oxide plus cathodes are added. We see that a higher concentration/reservoir of activators helps to increase poisoning robustness. The results of Wagener [52] given for O_2 and probably triple oxide cathodes at 800 °C (brightness or true temperature?) in the article of Gallagher [53] are also included. The higher temperature could partly explain the lower O_2 poisoning sensitivity compared to SF and oxide plus (Ox+).

The usual poisoning formalism [53] relies on surface poisoning and hence equates the unpoisoned current density j_0 in relation to the poisoned $j(p)$ at a pressure p. From the Richardson-Fowler (4.3) the ratio j/j_0 can be derived to be:

$$j/j_0 = \exp(-e \cdot \Delta\Phi/kT) = \exp(-\sigma_m \times \Theta \times M/\varepsilon) \tag{4.6}$$

with σ_m being the density of possible adsorption sites on the surface, Θ = the fraction of those sites covered by gas, M = the dipole moment formed by the adsorbed gas and ε = permittivity.

This of course influences pulsed emission, yet with respect to dc current limitation, the critical current density from the semiconductor picture is given by—as we recall from (4.4)

$$j_{\text{crit}} = (2kT/e) \times n_D e \mu_- / d$$

where n_D is the average density of the donors in the oxide bulk, μ–is the mobility of the electrons and d is the thickness of the oxide semiconductor. Here the average donor density n_D is either identical or proportional to the Ba density. Then according to [53] poisoning is following one of the 5 adsorption isotherms, here especially type III: $n = c(p/p_0)^2$, p_0 being the gas vapour pressure at temperature T, as one can derive from Fig. 4.19 from the quadratic dependence on pressure of the curve fits.

In Fig. 4.19 cathode life according to the criterion $j_{dc} \geq 2$ A/cm^2 is depicted versus total pressure in the test chamber. It can be seen, that a decrease of chamber pressure by 1 order of magnitude results in a lifetime increase by 2 orders of magnitude. The lifetime results for Philips SF cathodes can be fitted by the relationship

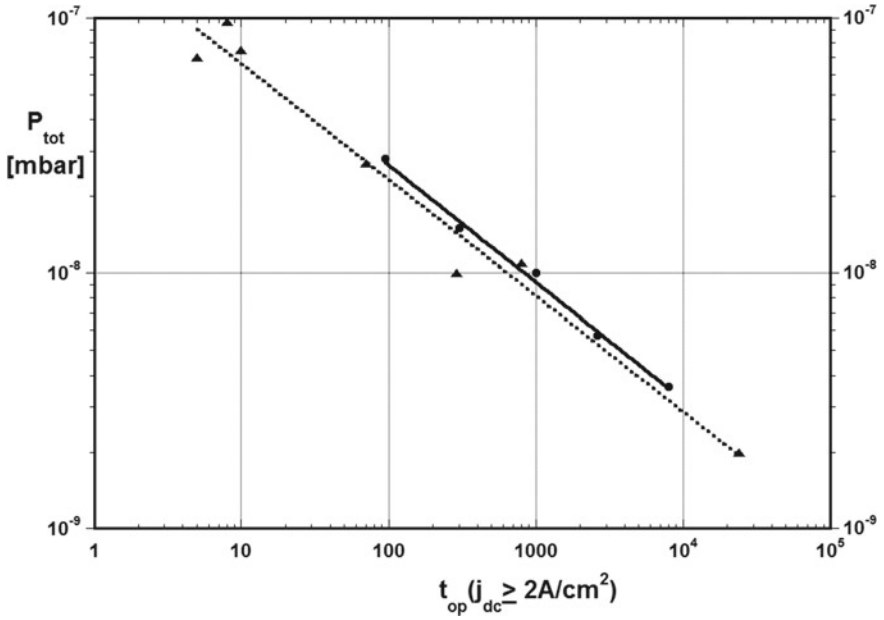


Fig. 4.19 Relationship of oxide cathode life w.r.t. $j_{dc} \geq 2$ A/cm^2 for Philips SF oxide cathode (black triangles) and for triple oxide cathode doped with Scandia (black dots) at $T_{op} = 780$ °C true temperature (750 °C Ni–Br) versus total pressure in the test chamber; compare fig./Abb. 42 of Pralle [54]

$$\tau_{dc}(780\,^{\circ}\mathrm{C},\, j_{dc} \geq 2\,A/\mathrm{cm}^2)\,[h] \;=\; \{10^{-6.8}/\,(p/mb)\}^2[h] \tag{4.7}$$

A similar relation holds for triple oxide cathodes. It is very interesting, that an improvement in chamber pressure during prolonged pumping can also improve j_{dc}, e.g. from 1.8 to 2.4 A/cm^2 and hence also τ_{dc}. The constant in the numerator has the dimension of a pressure (p_0) and is just the axis value at a lifetime of 1 h. Of course not the total pressure per se, but the poisonous gas partial pressures, mainly H_2O and oxygen (compare Fig. 4.18) are responsible for oxide cathode emission poisoning. Typical rest gas analysis results obtained in one of the test chambers show H_2O partial pressure of about 80% and of O_2 of 3% of total pressure in the 1×10^{-7} mbar range compared to about 90% H_2, 6% H_2O and 1% O_2 in the region of 5×10^{-9} mbar [54]. It is consistent with mainly initial H_2O poisoning and its memory effect on cathode life and also stresses the importance of baking and degassing the UHV chamber before cathode conversion and activation in order to reduce the partial pressure of water vapor. Since the relationship depends on the type of cathode and also on the rest gas composition, it cannot really be used for lifetime prediction, if composition and structure of oxide cathodes are varied. Yet it is a guideline, that for best emission performance of any oxide cathode the total pressure in the chamber should be at least below 3×10^{-9} mbar [55].

Typical emission poisoning tests are carried out by intermittent gas poisoning, where usually the poisoning gas was introduced via a gas leak at a certain pressure for 60 min, as shown in Fig. 4.20 for oxygen admission. Here again we see, that oxide plus cathodes (Y-doped W1) are superior to SF cathodes, where higher Y_2O_3 doping level is more critical.

In Fig. 4.21 the effect of 1 h O_2 gas poisoning in contrast to short time recovery of Fig. 4.20 is shown for the whole operational life w.r.t. the life end criterium of 2A/cm^2 dc for several types of Philips oxide cathodes doped with 60 ppm yttria (compare [38]). It can clearly be seen, that increasing the activator concentration in cathode Ni (2 fold or 4 fold compared to standard SF) has the strongest effect on longer life after poisoning, or in other words: the lower the activator concentration, the more detrimental the poisoning effect on cathode life! Also the addition of Ni particles in oxide plus results in a further improvement over SF with double activator content in cathode Ni. Comparing the two Figs. 4.20 and 4.21 one can further state, that the oxide cathode has a memory. Despite initial recovery it may die earlier, if the generation of Ba is less efficient. One should note, that 1 h O_2 poisoning after 140 h was conducted in the time domain of the Mg activator, which mainly spreads by diffusion; the overall behavior may be different, if poisoning happens e.g. after 1000 h in the domain of Al, which also evaporates onto the grains.

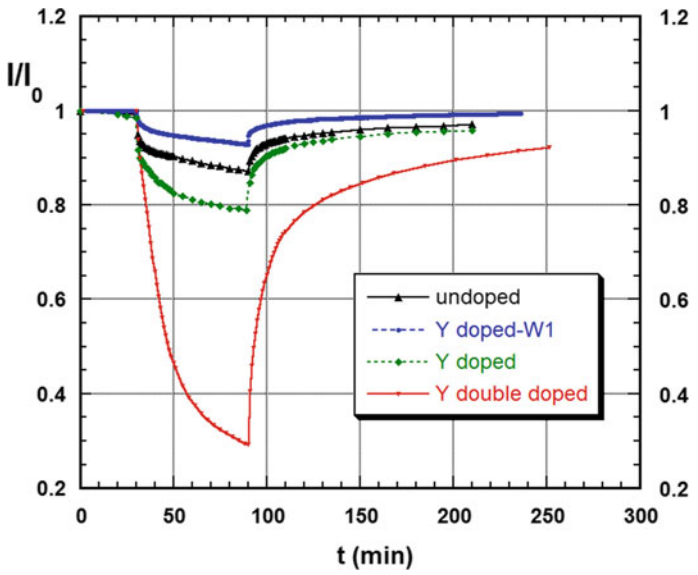


Fig. 4.20 Effect of oxygen admission of about 2×10^{-8} mbar into the chamber for 60 min at T_{op}= 750 °C(Ni–Br.). Depicted are SF type cathodes without yttria and with 60 ppm (SF) and with 120 ppm yttria and a SF type cathode with double activator concentration in the Ni base (W1), similar to oxide plus, which is more robust w.r.t. emission poisoning. The graph is partly based on [38], Fig. 6

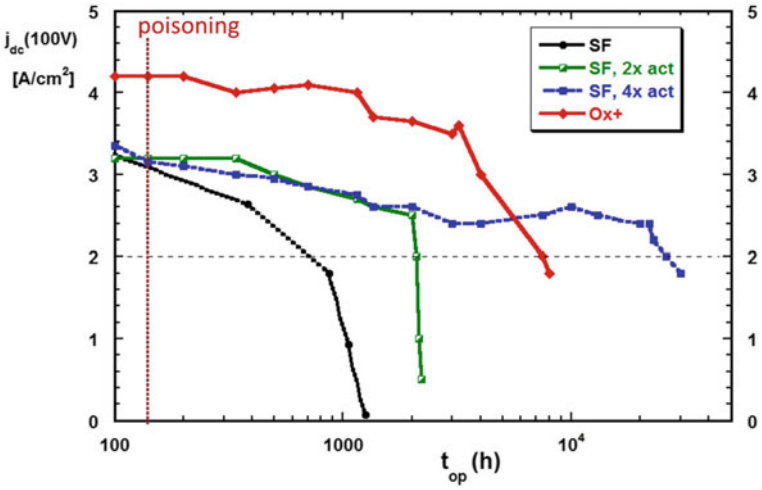


Fig. 4.21 Effect of 60 min oxygen poisoning after 140 h at 750 °C (Ni–Br.) on cathode life for different types of oxide cathodes doped with 60 ppm yttria and different activator content in the Ni base, including SF and Oxide+; (compare [38])

By the way, the doping of the oxide with rare earth oxides was first introduced by doping with Scandia particles by Saito et al. of Mitsubishi [18], with the positive effect, that the $BaSiO_3$ (+MgO) interface layer now got kind of 'dissolved', thus avoiding reduced activator supply and high interface resistivity. P. Derks of Philips first introduced molecular doping with Eu_2O_3 or Y_2O_3 [19], which leads to higher conductivity. Hayashida et al. of Matsushita (MEC) [56] did an extensive investigation of doping effects of the group III, IV, or V elements on CRT oxide cathodes, leading to increased conductivity. Unfortunately they did not study poisoning, but similar sensitivity should be expected.

Figure 4.22 shows a further improvement by introducing a fine-grained Ni interlayer on the Ni base, either done by sputter coating as by LG. Philips Displays Blackburn (results shown in this Fig., communication by D. Barratt) or by Laser ablation deposition of ultrafine Ni particles on the cathode Ni. This feature also allows higher dc loads at zero hour and also improves life performance after poisoning. Improved cathode performance with a W interlayer on Ni was also shown by Saito et al. [21] and with a fine-grained Ni interlayer on Ni by Kim et al. of Samsung [43, 57].

4.7 Electrical Conductivity of Oxide Cathodes

We have seen in the preceding paragraphs, that one of the handicaps of the oxide cathode is the limited electrical conductivity σ_{el}, which is essentially due to (Ba, Sr)O or the triple oxide being semiconductors. We have also noted, that electron emission

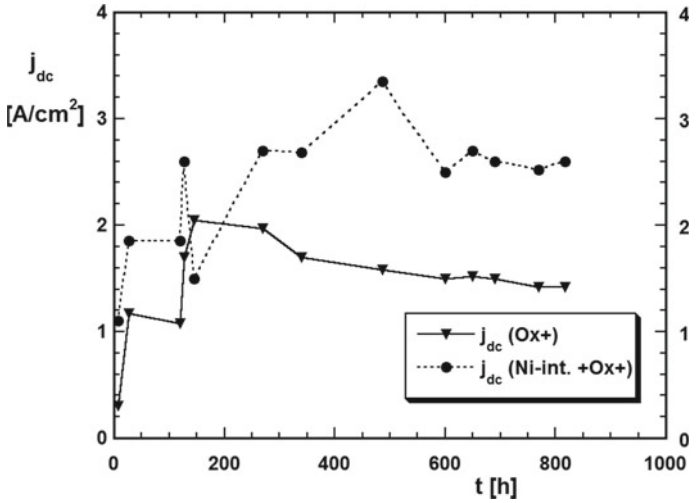


Fig. 4.22 Influence of poisoning on emission: Improved performance of oxide plus cathode with additional sputtered fine-grained Ni interlayer on top of cathode Ni is maintained after oxygen poisoning (at $t = 0$) in comparison with standard oxide plus cathode; *Top* = 750 °C (Ni–Br.), $t =$ 0 after about 100 h of operation. Evaluation of results of LG. Philips Blackburn; communicated by D. Barratt

capability and electrical conductivity are strongly linked for activated oxide cathodes. In this paragraph we want to address some basic conductivity models and present experimental results for oxide cathode conductivity. Then a new direct method of σ_{el} determination is presented and is applied to advanced new oxide cathodes. One should keep in mind, that oxide cathodes have a high porosity of 50–75%, depending on preparation. High porosity is also linked to high emission, whereas lower porosity oxide cathodes yield lower emission according to Loosjes and Vink [11]. This is also seen over life, since the cathodes sinter, become denser and the pore portion decreases, also accompanied by an emission decrease. This can also reflect a change in the conduction mechanisms as illustrated in Fig. 4.23 (compare [59]).

For electron conduction from the base Ni through the porous oxide to the emitting surface essentially three pathways exist: namely through the bulk of the grains and their contacts, along the surface of the grains and third via an electron gas through the pores [11, 58–60]. Of course all three can contribute and the question is, which is the dominant contribution. Later we will also discuss the contribution of percolation by addition of Ni particles to the porous oxide. It had already become clear before 1945, that generation of free Ba via reaction of Ba with the activators is essential for oxide cathode functioning, and that it is counteracted by oxygen poisoning, which led to a Ba monolayer model on BaO. After 1945 several researchers tried to explain oxide cathode behavior by semiconductor models of the bulk grains. Essentially two models had been promoted for solid state conduction through the grains, namely the "Mobile Donor Model" originally formulated by Nergaard [2, 3] and refined and promoted

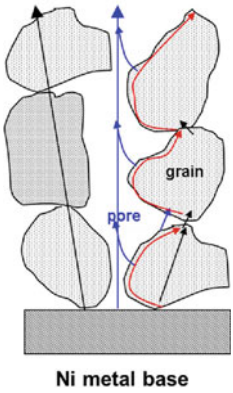


Fig. 4.23 Schematic view of porous oxide conductor; Contribution of different conduction mechanisms: grain conduction (black arrows), surface conduction (red) and pore conduction (blue); simplified model (compare [59])

by Zalm [5] and Chin et al. [17], and the "Mobile Acceptor Model" postulated by Okumura and Hensley [61, 62]. Of course there is a lot of experimental evidence that BaO and also (Ba, Sr, Ca)O are excess semiconductors. The experiments of Doloff [63] has shown, that the electrical conductivity of BaO single crystals is increased by excess Ba. The way how this can interact with the crystal lattice is illustrated by Herrmann and Wagener [1, Fig. 88] with a $Ba^{++}O^{--}$ lattice with some O vacancies coupled to Ba. The vacancies in the BaO lattice can then be compared with F-centers e.g. in NaCl in the same crystal lattice type. As for NaCl with excess Na, BaO is also coloured by excess Ba uptake (here blue). From measurements of the Hall-effect Wright [64] found the sign of the Hall effect being that of an excess electronic semiconductor, the Hall constant H being an exponential function of $1/T$ as required by theory (It has to be noted that $\sigma \times \mathrm{H} = 3\pi/8 \times \mu$ is nearly equal to the mobility μ). A simple model first presented by Wilson [33] can be used for giving a quantitative description of the $\boldsymbol{\sigma}$ versus T dependence, by just taking 1 donor level at a depth $\mathrm{E_D}$ below the bottom of the conduction band E_C ($E_C - E_D = \Delta E$). It is assumed, that the band-gap is so wide that the contribution of electrons excited from the valency band to the conduction band is negligible. The density of free electrons in the conduction band can then be determined from the donor density n_D following the Wilson model [33]:

$$\sigma_{el} = n_e e\, \mu_e \text{ with } n_e = n_D^{1/2} \times N_C^{1/2} \exp(-\Delta E/2kT) \tag{4.8}$$

Here N_C is a number density dependent on temperature and is given by

$$N_C = 2 \times (2\pi\, \mathrm{m} kT)^{3/2} / \mathrm{h}^3 \tag{4.9}$$

In contrast the mobile acceptor model-promoted by Okumura and Hensley [61, 62] on basis of diffusion experiments- is not convincing due to not well controlled experimental conditions and is also in contradiction to the Hall-effect measurements

of Wright. They had suggested, that the principle acceptors in e.g. BaO consist of Ba-vacancies which are highly mobile and that the principle donors are oxygen vacancies which are much less mobile.

A difference to standard semiconductor theory is, that the charge distribution in the oxide coating is not generated by immobile donors and acceptors, but by mobile electrons and donors [2, 3, 5, 17] and hence is influenced by field penetration and fields induced by the surface states. Then a critical current density j_{crit} can be derived under the assumption, that the constant number of donors is redistributed under the action of current flow and the electron density vanishes at the positive surface of the semiconductor at $\boldsymbol{j_{\mathrm{crit}}}$ as given in (4.4). This can quantitatively explain the observation of limited dc-loadability of the oxide cathode. In consequence, the critical dc-current density can be increased by reduction of the oxide coating thickness and/or by increase of the average donor density. The higher the electrical conductivity, the smaller is the Debye length λ (which is derived for small band-bending, $\Delta E << kT$):

$$\lambda = (\varepsilon kT/4\pi\, e^2 n_i)^{1/2} \tag{4.10}$$

The Debye length λ is introduced here, since it is also needed for the discussion of the model of Chin et al., which explains σ mainly with surface conduction of the oxide cathode grains. Taking into account band bending and surface conduction as in [17], the interpretation of the parameters given above has to be changed, but the basic relations still hold.

According Chin et al. [17] the electrical conductivity of a BaO single crystal at 1100 K is 1–2 × 10^{-4} (Ohm × cm)$^{-1}$; SrO and CaO have much lower values. $\boldsymbol{\sigma_{el}}$ of an un-activated BaSr.O coating lies in the range 3 × 10^{-6} to 1 × 10^{-4} (Ohm × cm)$^{-1}$ which is at least 2 orders of magnitude lower than for the activated coating: 2 × 10^{-3} to 1 × 10^{-2} (Ohm × cm)$^{-1}$. According to Doloff [63] a BaO crystal with excess Ba can reach conductivity values of up to 1(Ohm × cm)$^{-1}$ at 1000 K. These Ba excess densities (between 0.3 and 6 × 10^{17}/cm^3) are obtained by heating in a Ba atmosphere, the crystals turning blue. But excess Ba is only transient and not stable over longer times and the coloured crystals bleach again; it is only instructive for what could be achieved.

Based on the facts, that Ba is supplied to the surfaces, where the oxide structure is modified, Chin et al. promoted the surface conduction model based on the mobile donor model and point out, that at the grain contacts surface conduction is the main current transport mechanism. They claim electron space charge accumulation at the surface of the grains with band bending (at the cathode surface) and calculate an average electron density $n_a = 2.5 \times 10^{15}$/cm^3 for an activated coating with $\boldsymbol{\sigma_{el}} = 3 \times 10^{-3}$ (Ohm × cm)$^{-1}$, with a value of 5 cm^2 × (Vsec)$^{-1}$ of the electron mobility μ_e. At the surface the electron density is higher and they approximate the thickness of the surface conduction layer with twice the Debye length, which means 50–100 nm.

It should be noted, that the vapour pressure of BaO at 1100 K is about 10^{-9} mbar, whereas the vapour pressures of SrO and CaO are at least 4 orders of magnitude lower [1, 7]. For the respective metals this is different, the vapour pressures are much

higher (0.3 mbar for Ba at 1100 K) and increase in the sequence Ba < Ca < Sr < Mg (see page 154, vol 2 of [1]). This can also explain, that more than 1 Ba monolayer is not stable under UHV conditions.

In Fig. 4.24 besides more recent results also the temperature dependence of σel for a well activated non doped oxide cathode as measured by Loosjes and Vink [11] are depicted, which are representing a lot of experimental data from the 1950s. In the intermediate range the slope gives a conductivity activation energy of about 1 eV, which is the same as the Richardson work function. From that they derive, that in this region an electron gas emitted into the pores is determining the conductivity, whereas at lower temperatures solid state conductivity becomes dominant. Hensley [62] has derived a general expression for pore conductivity of:

$$\sigma_p = j_0 e d_p / kT \tag{4.11}$$

Here d_p is some average pore diameter weighted in favour of the narrower dimensions. The dependence of σ on temperature is of course determined by the electron density in the pores and hence by electron emission into them and is then proportional to j_0 as given by the Richardson-Fowler equation.

We also see in Fig. 4.24, that Ni particle additions as introduced by Philips and continued by LG-Philips Displays in Blackburn, improve conductivity drastically also in the lower temperature range [6, 22] due to the percolation effect [66, 67].

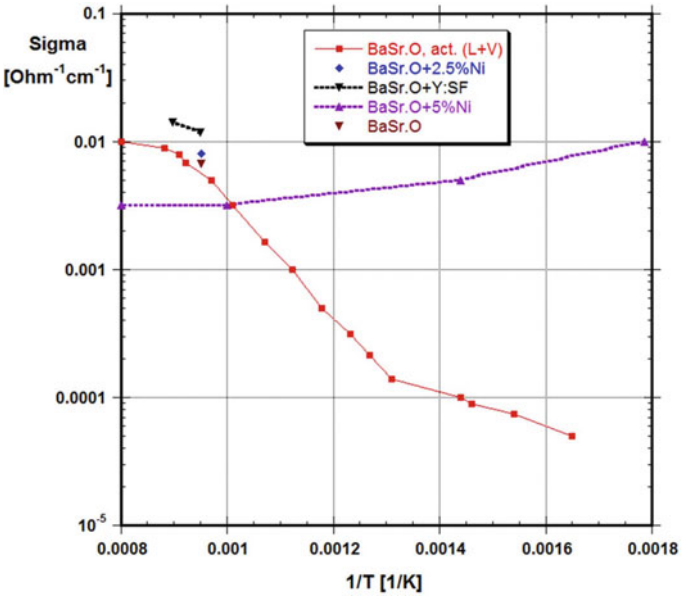


Fig. 4.24 Electrical conductivity of oxide cathodes; comparison with some literature results: L + V are results of an activated BaSr-oxide cathode of Loosjes and Vink [11]; BaSr.O + 5% Ni particles are results of Barratt and Jenkins [6]

The change of slope in the lower temperature range could also be explained by a conductivity contribution of Ba monolayers on the grain surfaces. From the known temperature dependence of bulk Ba one can calculate σ_{el} (773 K, bulk) = 176 (Ohm × cm)$^{-1}$ (see Gmelin [68]). It is well known, that conductivity decreases with decreasing film thickness according to a formula from the Fuchs-Sondheimer theory given by Yamauchi ([69], p. 377), which we can use for estimating the conductivity of a Ba monolayer via

$$\rho(d) = \rho_{\text{bulk}} \times (1 + 3/8 \times (1 - p) \times \lambda_e/d) \tag{4.12}$$

where d is the thickness of the thin film and p is a parameter having values $0.5 < p < 1$ for epitaxial and single-crystalline thin films, and is nearly 0 for polycrystalline films. With the known mean free path λ_e of electrons in the bulk λ_e(Ba) = 6.5 nm (compare e.g. [70] C. Kittel, Chap. 6, Tables 1–3) we can then calculate σ_{el} (773 K, 1 ML) = 14.4 (Ohm × cm)$^{-1}$. With a simple model of rectangular grains of 3 μm × 3 μm cross section and equal pore dimensions with 75% porosity and using the atomic radius of Ba = 0.217 nm as monolayer thickness one can calculate the monolayer portion of the cross section, resulting in σ_{el} (773 K, oxide) = 1 × 10^{-3} (Ohm × cm)$^{-1}$ which is in the same order of magnitude as the values of Loosjes and Vink [11] in Fig. 4.24.

The two conductivity curves in Fig. 4.24 for undoped cathodes [11, 22, 23] have both been measured with the two cathode method, where two cathodes are mounted face to face and are then moved till they touch, so that the resistivity can directly be determined. A disadvantage of this method is, that e.g. directional effects are averaged, as for instance the electrolytic Ba^+ ion flow is reversed in one cathode. Another method was first used by Coomes [25], where the pulsed emission with two different duty cycles is used for reaching the same temperature. From the difference the cathode resistance can be determined, but by neglecting the effect of emission cooling. Hence these values are only approximate and refer to ac conductivity. A further modification of the approach of Coomes had to be introduced by Hayashida et al. [56] and Choi et al. [71] for indirectly heated cathodes. These problems were overcome by Gaertner et al. [20] by a new direct measurement of dc conductivity by a compensation method.

In this procedure cathode temperature at a given heater power is monitored as a function of increasing continuous dc-load (dc current after 1 min) with an optical pyrometer in a close-spaced planar diode configuration in a UHV chamber. The temperature observed is the result of a superposition of electron emission cooling - proportional to the dc current I_e—and ohmic heating proportional to the square of the dc-current drawn = $R \times I_e^2$ [20, 72, 73] Of course the emission cooling contribution is not strictly linear in I_e, since the work function threshold is modified either by space charge or by the mirror image charge, which is again dependent on I_e, as one can see in (4.12) [20]. In Fig. 4.25 the resulting temperature is shown as a function of dc current I_{dc}. (1 min). It can be seen, that the temperatures without load and at the compensation point, where emission cooling compensates ohmic heating, are equal.

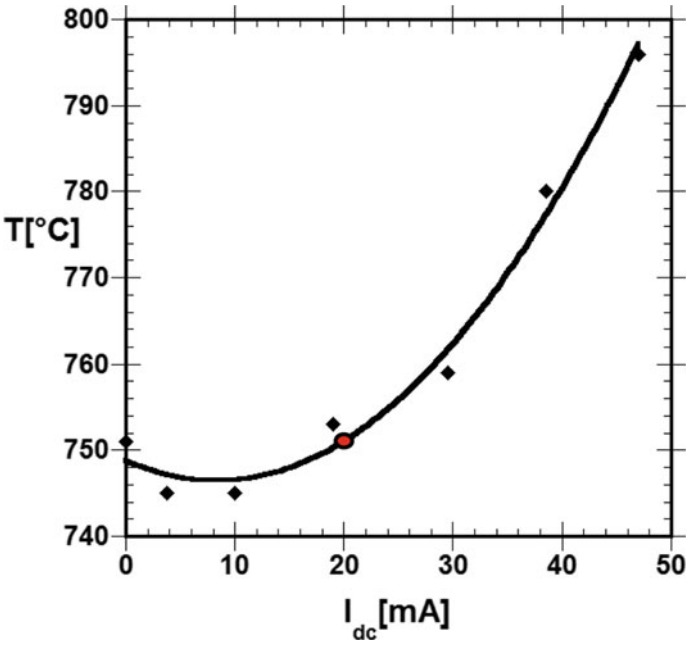


Fig. 4.25 Superposition of Ohmic heating and electron emission cooling of oxide cathode coating: Ni-brightness temperature as a function of dc current (1 min) for oxide cathode doped with Y_2O_3 (after 1500 h). The compensation point is shown in red; from G. Gaertner et al. [20], Copyright 2002, with permission from Elsevier

Hence at the compensation point the resistance and hence the conductivity can be determined from the (4.13)

$$R\,I_e^2 = I_e \times [\Phi_e + kT/e \times \ln(I_s/I_e) + 2kT/e] \tag{4.13}$$

inserting the respective work function Φ_e and saturated emission *current* I_s values. Of course pulsed and dc emission characteristics are also measured as a function of cathode temperature in order to determine the work function.

Using this method the results depicted in Fig. 4.26 have been obtained as a function of operation time. Hence Y_2O_3 or Eu_2O_3 molecular doping increases the electrical conductivity of the un-doped mixed oxide by introducing stable donors. Quantitatively Philips Research found ([20] and information by G. Gaertner), that the increase of σ_{el} is additive to the conductivity of the un-doped oxide and amounts to about $+1 \times 10^{-3}\ \Omega^{-1}\text{cm}^{-1}$ per 40 ppm yttria at about 300 h. Unfortunately also the poisoning sensitivity increases, probably since the donors are less mobile after oxidation, and also sintering is enhanced the higher the dopant concentration. We can see, that σ_{el} is decreasing with operation time, which can be explained by diminishing Ba supply, by sintering and coarsening of the structure, leading to lower porosity and less inner surface area. This effect can be partly overcome by addition of Ni particles

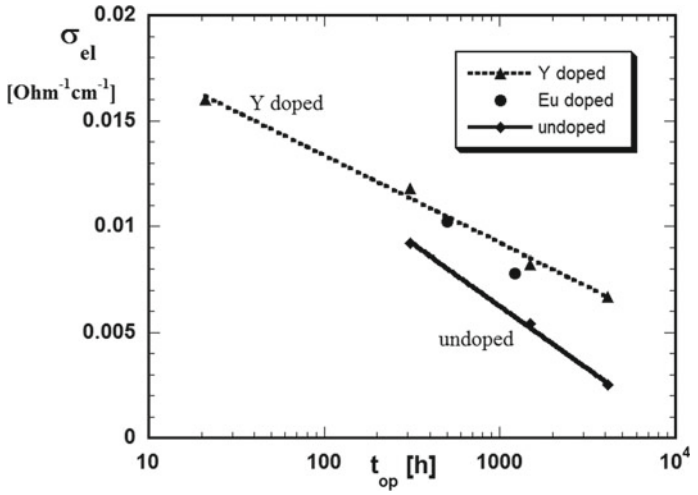


Fig. 4.26 Electrical conductivity σ_{el} of undoped and yttria or europia doped BaO.SrO oxide cathodes versus operation time at 780 °C true temperature (750 °C Ni brightness) according to Gaertner et al. [20], Gaertner and Koops [7], Springer 2008, Fig. 10.12

into the porous oxide matrix, as can be seen from Fig. 4.24 and the improved lifetime performance of Philips Oxide Plus cathodes.

4.8 Current Model of the Oxide Cathode

In the preceding subchapter on electrical conductivity of oxide cathodes we have already discussed the 3 (or 4) main conduction mechanisms. Eventually we do not have to make a choice between the three, but have to evaluate their contributions in a well-activated oxide cathode and find out, which is the dominant one, depending on the emitter temperature and the structural and compositional changes experienced over life. Based on the work of Zalm [5] and Chin et al. [17], semiconductor surface conduction with some contribution of pore conduction by the electron gas, which is also denser near the grain surfaces and needs the conducting inner surfaces as electron emitters, is responsible for the high conductivity in early life (compare [6]), whereas conduction through the grains will become more pronounced near the end of life. This statement especially holds for a temperature range from 1000 to 1100 K with a preferred oxide cathode operating temperature of 1050 K, and can also be derived from the slope of conductivity versus reciprocal temperature relation in Fig. 4.25. It should be mentioned, that Dearnaley [74, 75] introduced a further phenomenological conductivity model of formation of filamentary conduction paths through the grains or along the surface, but without any quantitative results. One consequence is, that the Richardson equation holds and high local emission spots should exist, consistent

with the brighter spots on the surface detected in some experiments. Their theory could also help to explain flicker noise in oxide cathodes, as observed by Lindemann and van der Ziel [76].

That free Ba generation and resupply is needed in a well activated cathode has been proven by numerous experiments, and also the role of the activators in the Ni base for this purpose is well established. The oxide cathode performance is improved by increasing the amount of activators, as we have seen in the preceding subchapters, especially by investigations of H. Kern [15] and by the oxide plus cathode variants of Philips and LG. Philips Displays [6, 38, 44]. The Ba monolayer or sub-monolayer on top of the cathode and on the surface of the grains is essential for a well emitting oxide cathode, together with the semiconducting base below it. One should not forget, that a complete Ba monolayer on all grain surfaces alone can explain the conductivity to a large extent. This leads to a modified coexistence of several conduction mechanisms within the same system and implies accommodating the bulk semiconductor and surface theories within the same overall model, although there is still some way to go to reach a complete unified model [6]. When looking at the sticking probability of Ba on a mixed BaO/SrO surface, it is already known from low energy ion scattering investigations of Cortenraad et al. [77] of Ba on Ba dispenser cathodes, that sticking via an oxygen bridge is much more likely than direct sticking on a metal atom. From these findings it is likely, that extra Ba will sit on top of O at the surface and hence the maximum Ba coverage should only be half a monolayer.

Based on the conductivity measurements of doped cathodes of Philips [20] D. den Engelsen and G. Gaertner have also updated the solid semiconductor model by incorporating the role of the (molecular) dopants [58–60], which of course has also consequences for the surface states. In the model of Zalm the electron donors are oxygen vacancies according to

$$V_O = V_O^{2+} + 2\,e^- \tag{4.14}$$

By the way, the formalism would not change for ionized interstitial Ba. From typical values of electric conductivity after activation in Sect. 4.7 we have already calculated an average electron density $n_a = 2.5 \times 10^{15}/\mathrm{cm}^3$ for an activated coating with $\boldsymbol{\sigma}_{el} = 3 \times 10^{-3}\ (\mathrm{Ohm} \times \mathrm{cm})^{-1}$, with a value of $5\ \mathrm{cm}^2 \times (\mathrm{Vsec})^{-1}$ of the electron mobility μ_e [17]. For a well activated SF cathode with $\sigma_{el} = 1.2 \times 10^{-2}\ (\mathrm{Ohm} \times \mathrm{cm})^{-1}$ one obtains $n_a = 1 \times 10^{16}/\mathrm{cm}^3$. Near the end of life electrical conductivity diminishes by two orders of magnitude same as n_a, the voltage drop over the oxide gets larger than the decomposition voltage of BaO of 2.3 V at 1050 K and Ba^{2+} will move to the interface. Now the role of the rare earth ions is, that part of them can become donors according to:

$$R^{2+} = R^{3+} + e^- \tag{4.15}$$

and if only 20 ppm are dissolved in BaO, and 1% become donors, they can additionally supply about $+n_a = 5 \times 10^{15}/\mathrm{cm}^3$ electrons, with a low effect at the beginning but a pronounced effect near the end and hence increase life compared to the un-doped

case [58]. But it becomes also clear, that oxygen poisoning and oxidation of these donors has to be avoided.

It is interesting, that also in other applications the replacement of Ba ions by rare earth ions oxygen vacancies are created, e.g. in the case of $BaAl_2O_4$:Eu,Dy or $SrAl_2O_4$:Eu,Dy prepared as nano powders for phosphor applications [78].

Besides conductivity one of the important features of an oxide cathode is the low work function and a possible explanation by the model. Already Wright and Woods [81] have set up a simple semiconductor surface model with one donor level (at distance R below the conduction band edge), as shown in Fig. 4.27. In absence of a field, the work function is $\phi = R/2 + \chi$. The emitting surface can be characterized by the electron affinity χ of the semiconductor (compare [102]), which may be altered (by $\delta\chi_0$) in the presence of a dipole layer. Following the more refined model of Zalm [5] and Chin et al. [17], a positive Ba ion layer on top of the grains, especially on SrO and CaO rich surface, induces band-bending and a high electron density at the surface already without an external field (see also Venema [104]), which is exhausted during short high voltage pulses (pulsed emission). The short pulse vacuum emission is hence dominated by the surface properties, which are of course also partly dependent on the bulk properties (e.g. raising of the Fermi level towards the conduction band edge during activation).

The contributions to the work function by band bending $\delta\chi$ and mirror image charge $\delta\phi_s$ for an external electric field X are depicted in Fig. 4.27. It should be noted, that $\delta\chi$ reduces the work function w.r.t. the unchanged Fermi energy level (which was wrongly depicted in the original figure in [81]. Here also the theoretical contribution of field penetration by Morgulis [79] has to be taken into account. Wright and Woods [81] give the following values ½ R = 0.7 eV, ϕ' = 1 eV and f.e. $\delta\chi$ = 0.07 eV at 1000 K and an electric field X = 15 kV/cm (we will call it F in the following) for a coating with low resistance and $n_0 = 10^{14}/\text{cm}^3$, whereas $\delta\phi s$ = 0.465 eV, yielding χ = 0.835 eV (In these definitions of Wright and Woods e is contained in ϕ' and ϕs).

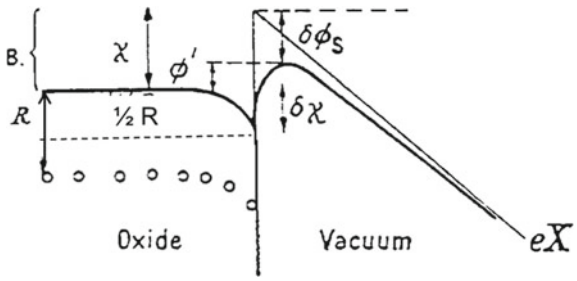


Fig. 4.27 Schematic energy levels in an oxide coating according to Wright and Woods [81], but corrected for the Fermi level, when space-charge and Schottky effects are both taken into account, and the voltage drop in the coating is neglected. With applied field eX, the work function is modified to $\phi' = \phi - \delta\chi - \delta\phi_s$; the open circles are the donors

In their investigations of 5 nm BaO on Ir by low energy electron reflections (LEER] Morgulis [79] obtained an electron affinity of 0.6 eV and a work function $\phi = 1.3$ eV, which implies $R/2 = 0.7$ eV.

Morgulis, Wright and Woods, Raju and Maloney (for finite thickness [80]) have derived implicit formulas for $\delta\chi$ of the type [81]:

$$eF/\varepsilon_{ox} - I/\sigma = -4 \times (2\pi\, n_0 kT/\varepsilon_{ox})^{1/2} \times \sinh(\delta\chi/2kT) \tag{4.16}$$

where $\boldsymbol{I}$ is the emission current and $\boldsymbol{\sigma}$ the conductivity of the oxide. For saturation the electric field at the surface vanishes and hence $F = 0$. One can easily see from these derivations, that the saturation behavior is different from pure Schottky theory, which is only part of it.

Furthermore, due to ionized donors (ionized Ba at the surface) there is still a higher surface electron density and an additional reduction of work function at $F = 0$ at the surface. $F = 0$ at the surface of course does not mean, that the Laplace field strength *U/d* is zero, but just compensates the space charge at the surface.

This picture neglects, that the cathode surface is not uniform, due to the pore openings, eventually filled with electron gas, and due to the surface morphology. Also the time dependent effects such as charge carrier depletion during pulsed emission and delayed resupply are not contained in it. Time dependent effects have been studied by Sproull [30], who proposes an electrolytic conduction mechanism, and by Krusemeyer and Pursley [82]. They investigated step-like changes in dc voltage and measurements of dc current over time and determined time constants. They explain their results with redistributions of donors and a change of their total number. These are also facets of a more complete theory.

4.9 Types and Variants of Oxide Cathodes

Since a large portion of earth alkaline oxide cathodes used in vacuum tubes or in gas discharge lamps consist of the triple oxides BaO, SrO and CaO, Grey [83] has set up a contour ternary plot of saturated emission current density j_0—obtained from short-pulsed emission of undoped cathodes—versus the respective concentration in a triangular diagram, the so-called Grey diagram. He obtained the maximum of $j_0 = 8$ A/cm^2 at a composition of 47% BaO, 43% SrO and 10% CaO, probably at a true temperature of 1000 K. Whereas Philips and Matsushita (MEC) were only using the Ba·Sr double oxide with $j_0 = 2.7$ A/cm^2, the majority of other manufacturers use triple oxide with admixtures of CaO of 4–8%, which is based on the former RCA-design. One should mention that a much wider variety of triple oxide electrode compositions is used by manufactures of gas discharge tubes, in which case tungsten is the base, but that is not our subject here. The applicability of the Grey diagram of course assumes equal activator amounts in the Ni base and would have to be modified for different thicknesses and porosities of the oxide layers. Before the introduction

of dopants further differences in cathode performance can arise by using different activators with different concentrations in the Ni base and different thicknesses of cathode Ni.

Due to problems with Si as activator (high interface resistivity and delamination), Philips and Matsushita used Mg, Al and W in the Ni base. Mitsubishi [18], Sony and Thomson [41, 42] applied Ni with Mg, Si and W as activators similar to RCA, whereas Hitachi, NEC [39], Samsung [43] and Toshiba only used Mg and Si. Depending on the manufacturer the Ni base thickness varied between 100 μm (Philips) and 200 μm. Also the dopants were different, as we have seen in the subchapters on poisoning and on electrical conductivity. In 1961 Lemmens and Zalm of Philips [84] had also investigated a double layer oxide cathode, where the first layer on the Ni contained Ti and Zr as activators, which had been sprayed as hydrides on Ni together with the triple carbonates, and the top layer was sprayed without the activators. The cathode reached 2500 h life at a dc load of 1 A/cm^2, which was an improvement compared to single layer oxide cathodes at that time, which had the handicap of increasing resistivity due $BaSiO_4$ interface formation (with Si as one of the activators). Of course omitting Si as activator and using e.g. Al instead is another possibility, which was used by Philips in the 1970s.

We already know, that Saito et al. of Mitsubishi [18] introduced doping of the oxide layer with scandia particles, followed P. Derks of Philips with molecular doping with Eu_2O_3 and Y_2O_3 [19]; Matsushita was also using Eu_2O_3. Hayashida et al. of Matsushita (MEC) [56] did an extensive investigation of doping effects of the group III, IV, or V elements. The highest doping effect, i.e. lowest emitter resistance showed Eu, but still lower was co-precipitated ZrO_2. In their patents Hitachi claim a ZrO_2 doping, but later also a Ba-Scandate can be found. Samsung was using La- and Mo-oxide doping. The other CRT oxide cathode manufacturers did not use any additional dopants.

Also the interface on the Ni base was modified by several companies to improve cathode life. Y. Kim et al. [43] of Samsung in 2000 introduced a sputtered fine-grained Ni interlayer on the Ni base, where best results were obtained with the largest thickness of 1.5 μm. Philips used laser ablation deposition of ultrafine Ni particles on the cathode Ni. This feature together with Ni particles in the oxide layer also allows higher dc loads. Improved cathode performance with a W interlayer on Ni was also shown by Ohira et al. of Mitsubishi [85, 86], where W of course also acts as a slow activator. Sawada of Mitsubishi in 2004 presented a further improvement of their Sc_2O_3 doped oxide cathode by sandblasting the Ni base before coating and forming W islands, allowing a dc load of 4 A/cm^2 [87].

It was already mentioned, that Philips with LG. Philips Displays as successor introduced nickel particle additions to the coating of their CRT oxide cathodes (cermet cathodes) based on the percolation effect [66, 67], which increased electrical conductivity over life and also reduced the cut-off drift in CRTs [22, 67, 88, 89]. Together with increased activator concentration in the Ni base this was one of the distinctive features of the “Oxide Plus” cathode of LG. Philips displays [23] and enabled a dc-loading of up to 4.5 A/cm^2 during life (see Fig. 4.1).

A different form was the reservoir oxide cathode introduced by Lemmens and Zalm of Philips in 1961 [84], which also tried to overcome the problem of increasing interface resistivity. Here the emitting (Ba, Sr)O layer is spray-coated via the carbonates on a Ni wire gauze. Below it a half-open Ni container is filled with (Ba, Sr)O eventually with 3% (by weight) Ti, the Ni of the walls also containing activators. In a variant, the reservoir consists of $BaAl_2O_4$ and 25% Ti. This cathode was capable of 5000 h life at a continuous dc load of 1.5 A/cm^2.

A variant of the "oxide" cathode in the direction to dispenser cathodes was a special Ni matrix cathode introduced by Sugimura et al. of NEC in 1999 [90], the so-called HIP cathode, where a mixture of Ni, Scandia and Ba·Sr·Ca-carbonate powders was hot isostatically pressed (HIP) into a Ni-Cr container. Cr from the walls plays the role of the activator. With a thin (initial) oxide layer on top of the oxide-Ni mix, the cathode reached an emission life of 22000 h at 3 A/cm^2 and 1153 K. At this temperature the HIP is superior to a standard oxide cathode, but at 1053 K their properties become similar. The concept of the HIP cathode is not new, but corresponds to the Ni matrix cathode, where the coating was soaked in (impregnated), see e.g. Jenkins [50]. In more detail it was described by Pantano and Wittberg [91]. This cathode was prepared from 50% Ni and 50% triple carbonate on 220 Nickel (rolled Ni matrix cathode) and is named after G. Medicus, who got a US patent for it in 1971 [92]. The earliest approach to an oxide-Impregnated Nickel-matrix cathode was by Balas et al. [93]. Its work function/emission lies between L cathode and sprayed oxide cathode.

A modification of the oxide cathode with Ni particles was the coated powder cathode of Maurer and Pleass [94], where 1–3% Ni was coated on the carbonate particles. This was done via a wet fluid bed coating with $Ni(CO)_4$ on the (Ba, Sr, Ca)carbonate particles. At a continuous dc load of 1 A/cm^2 the CPC could reach lifetimes of 40–50 kh. The advantage of this cathode is not only increased conductivity via the Ni coating, but also the inhibition of excessive grain growth and coarsening.

In Beijing new variants of reservoir oxide cathodes were introduced by Wang et al. [95, 96], in one version consisting of a Ni sponge and a Ba·Sr·Ca(CO_3) reservoir below, later a Ni-Sc sponge, a Ni-Re–Ir sponge and finally a Ni sponge impregnated with ammonium-perrhenate. Also the top oxide coating was prepared by plasma spraying. Reservoir and plasma sprayed oxide cathodes exhibit a much stronger emission cooling effect than conventional oxide cathodes, due to higher electrical conductivity. At 780 °C true temperature the Ni-Sc sponge cathode delivered 2.3 A/cm^2 dc current. The Ni-Re–Ir sponge cathode provided 56 A/cm^2 at 800 °C brightness temperature at a duty cycle of 10% [97].

4.10 Alternate Preparation Methods

Based on measurements by Tischer [35] of the critical current density versus oxide thickness, an increase with reduced thickness was observed, reaching from 0.5 A/cm^2 for a thickness of 100 μm to 3 A/cm^2 for 20 μm. Using also semiconductor theory,

this would extrapolate to 50 A/cm^2 for an oxide thickness of 1 μm, which would be consistent with the mechanism of conductivity limitation. This leads to the idea to improve oxide cathode performance by preparation of a thin oxide layer, which of course should be more stable to meet lifetime requirements.

Such an approach to further improve the performance of oxide cathodes w.r.t. higher dc current densities was by vacuum arc plasma deposition of the BaO/SrO coating (thickness in the order of 1 μm) directly on the cathode base by Maslennikov et al. [98] and Umstattd et al. [99, 100], instead of using the traditional process by spraying or painting as described above. Umstattd avoided the carbonate conversion process and reached pulsed current densities of 30 A/cm^2, but unfortunately these cathodes were very sensitive to water-vapor poisoning and needed H_2O partial pressures in the low 10^{-9} mbar range. He used different types of activator doped cathode Ni as a base. Unfortunately in his thesis Umstattd [100] did not comment on any load assisted activation process (only mentioning temperature remaining below 1000 °C), which probably could have increased the robustness.

Maslennikov et al. [98] prepared oxide cathodes by vacuum arc plasma deposition with a density of the oxide layer of 4–4.5 g/cm^3. Yet the emission performance of these cathodes did not really show an improvement with 0.5–0.7 A/cm^2 mean current density. Also here other conditions such as cathode Ni activator content or activation treatment have not been commented and could give space for improvement. This approach of a compact layer has also two further disadvantages: first, an advantage of the high porosity of the conventional oxide cathodes provides a large inner surface area for a monolayer Ba coating (also increasing conductivity) and show pore conduction, which is not the case for a compact layer; second, a μm thin layer has no reservoir for the inevitable evaporation over time and a thermally more stable oxide than BaO would be needed.

There are two possible ways to save this concept, namely to prepare nano porous cathodes, eventually with a nano porous interlayer on a suitable activator doped cathode Ni base, and second to use a more stable (Ba, Sc-oxide) compound together with BaO. Similar concepts have been followed in the case of LAD top layer Scandate cathodes [36, 44].

4.11 Conclusions and Outlook

With improved vacuum techniques in the 1930s, oxide cathodes became the predominantly used cathodes in radio electron tubes, electronic tubes, and later also in cathode ray tubes (CRTs) and in gas discharge lamps. After a first culmination of oxide cathode research in the 1940–1960s, a new rise of oxide cathode research and development occurred in the nineties due to price pressure on CRTs, which resulted in new insights and major improvements of oxide cathodes. At the end of the 20th century about 700 million oxide cathodes were produced for CRTs and more than 7 billion oxide cathodes for gas discharge lamps. These numbers have strongly decreased since then, due to the replacement of CRTs by LCDs and of CFL lamps by

LEDs. Oxide cathodes are nowadays still applied e.g. in microwave tubes of lower power and in RF tubes, but with much lower production numbers than in 2000 and there are practically no research efforts any longer, compared to the time 15 years ago. That is also the reason, why the performance progress curve of Fig. 4.1 does not continue, but levels off after 2005. This happens despite of the fact that a lot of further improvement options tested and proven in research before 2005 are not implemented in oxide cathode manufacturing and would have continued the progress line of Fig. 4.1 till nowadays. For example a "super oxide plus" cathode comprising additional options besides higher activator concentration in the Ni base and rare earth doping, such as fine grained interface layer and addition of **activator doped Ni particles**, was not realized and rolled out into production. The motivation for this contribution therefore was to give a review of these options for the future including the functional and theoretical insights based on improved oxide cathode types. New insights have especially been obtained w.r.t. conductivity, rare earth oxide doping, oxygen poisoning and dc loadability in the context of accelerated life tests. Looking at the list of peculiarities in the second subchapter, for nearly all the topics there exist now theoretical or phenomenological models and explanations, maybe with the exception of the accelerated life tests with continuous dc load, which need more investigations. Yet one may still call it a patchwork theory of the oxide cathode.

From Chap. 3 we know, that Ba scandate cathodes with their high emission capability can be regarded as a combination of Ba dispenser cathodes with a thin and rather stable oxide layer on top, which can be described by a semiconductor model explaining the low work function similar to Ba oxide cathodes. Hence Ba oxide cathodes may also profit from further progress in Ba scandate cathode models.

References

1. G. Herrmann, S. Wagener, *The Oxide Coated Cathode* (Chapman & Hall, London, 1951)
2. L.S. Nergaard (RCA), Studies of the oxide cathode. RCA Rev. **12**, 464–545 (1952)
3. L.S. Nergaard (RCA), Electron and ion motion in oxide cathodes. Halbleiter-probleme, Bd. **3**, F. (Vieweg Verlag, Braunschweig,1956), pp. 154–206
4. T. Cunningham et al. (eds.), Electron tube design (RCA (Radio Corporation of America), Harrison (NJ), 1962). (containing: L.S. Nergaard, Fundamentals of electron emission, 1–27; C. Hadley, Oxide-coated Emitters 28–36; C. Meltzer, E. Widell, Electron-emission coating for the oxide cathode 37–63; C. Meltzer, E. Widell, Nickel base metal fort the oxide cathode, 64–76; E. Hanig, The interface and other oxide cathode impedances, 77–89)
5. P. Zalm, Thermionic cathodes. Adv. Electron. Electron Phys. **25**, 211–275 (1968)
6. D.S. Barratt, S.N. Jenkins, The evolution of oxide cathodes for cathode ray tube applications. J. Mat. Sci. Mater. Electron. **17**, 735–743 (2006), https://doi.org/10.1007/s10854-006-0020-5
7. G. Gaertner, H.W.P. Koops, Vacuum electron sources and their materials and technologies, in *Vacuum Electronics, Components and Devices (chapter 10)*, ed. by J. Eichmeier, M. Thumm (Springer, 2008), ISBN 978-3-540-71928-1
8. A. Wehnelt, Über den Austritt negativer Ionen aus glühenden Metallverbindungen und damit zusammenhängende Erscheinungen, Annalen der Physik, Serie **4**, vol. 14, 425–468 (1904). (A. Wehnelt, Sitzungsberichte der physikalisch-medizinischen Sozietät Erlangen, 150–158, 1903)

9. F. Jentzsch, Über die Elektronenemission glühender Metalloxyde. Ann. Phys. **27**, 129–156 (1908)
10. G. Gaertner, D. den Engelsen, Hundred years anniversary of the oxide cathode-a historical review. Appl. Surf. Sci. **251**, 24–30 (2005)
11. R. Loosjes, H.J. Vink, The conduction mechanism in oxide-coated cathodes. Philips Res. Rep. **4**, 449–475 (1949)
12. N.B. Hannay, D. McNair, A. White (Bell), Semiconducting properties in oxide cathodes. J. Appl. Phys. **20**(7), 669–681 (1949),
13. R. Peterson, D. Anderson, W. Shepherd, Influence of the cathode base on the chemical activation of oxide cathodes. J. Appl. Phys. **28**, 22–33 (1957)
14. H. Allison, H. Samelson (Bell), Diffusion of Al, Mg, Si and Zr in Nickel. J. Appl. Phys. **30**(9), 1419–1424 (1959)
15. H.E. Kern, Bell. Lab. Rec. **38**, 451–456 (1960)
16. H.G. Boumeester, Development and manufacture of modern transmitting valves. Philips Technical Rev. **2**(4), 115–121 (1937)
17. T.N. Chin, R.W. Cohen, M.D. Coutts, Electronic processes in oxide cathodes. RCA Rev. **35**, 520–538 (1974)
18. M. Saito et al., Higher current density oxide cathode for CRT. NTG Fachber. **95**, 165–170 (1986)
19. P. Derks (Philips), European Patent Application 300568, Appl. 14.7.88, Priority 23.7.87, NL + EPS 482704, 29.4.92, Prior. 20.10.90 NL
20. G. Gaertner, P. Janiel, D. Raasch, Direct determination of electrical conductivity of oxide cathodes. Appl. Surf. Sci. **201**, 35–40 (2002)
21. M. Saito, H. Teramoto, T. Ohira, Recent trends in CRT cathodes. SID 99 Digest 25.2, 360–363 (1999)
22. G. Gaertner, P.A.M. van der Heide, New developments in CRT cathodes, IDW 2000, Kobe, Japan, Technical Digest CRT4-1, 513–517 (2000)
23. D. Barratt, N. Filkin, I. Bakker, Oxide plus-a newly developed cathode for CRT applications. SID 2003 Digest, 12.3, 162–165 (2003)
24. W.H. Kohl, Handbook of materials and techniques for vacuum devices, in *Cathodes and Heaters (chapter 16)* (Reinhold Publishing Corporation, New York 1967), pp. 475–528
25. E. Coomes (MIT), The pulsed properties of oxide cathodes. J. Appl. Phys. **17**, 647–654 (1946)
26. A. Eisenstein (MIT), A study of oxide cathodes by X-ray diffraction methods. Part I, J. Appl. Phys. **17**(6), 434–443 (1946), Part II, J. Appl. Phys. **17**(8), 654–663 (1946)
27. A. Eisenstein (MIT), Some properties of the Ba_2SiO_4 oxide cathode interface. J. Appl. Phys. **20**(8), 776–790 (1949). (A. Fineman, A. Eisenstein, Studies of the interface of oxide-coated cathodes. J. Appl. Phys. **17**(8), 663–668 (1946)
28. G. Gaertner, D. Barratt, New developments and life aspects of oxide and barium dispenser cathodes, in *ITG-Conference Displays and Vacuum Electronics*, *ITG Fachbericht*, vol. 183 (Garmisch-Partenkirchen, 2004), pp. 111–116
29. A. Eisenstein (MIT), Oxide coated cathodes. Adv. Electron. **1**, 1–64 (1948)
30. R. Sproull (Cornell University/ RCA), An investigation of short-time thermionic emission from oxide-coated cathodes. Phys. Rev. **67**, 166–178 (1945)
31. R. Matheson, L.S. Nergaard (RCA), The decay and the recovery of the pulsed emission of the oxide cathode. J. Appl. Phys. **23**(8), 346–349 (1952)
32. A. Manenschijn, S. Deckers, T. Weekers, P. van der Heide, Emission characterization of impregnated cathodes and Scandate cathodes, in *Conference record of 1992 TRI-Service/NASA Cathode Workshop* (Greenbelt/Md, USA, 1992), pp. 67–71
33. A.H. Wilson, Proc. Roy. Soc. **A133**, 458 (1931)
34. L.S. Nergaard (RCA), The physics of the oxide cathode. RCA Rev. **18**, 486–511 (1957)
35. K.M. Tischer, Halbleitertheoretische Behandlung der Oxidkathodenemission, Teil I, Frequenz **32**(7), 208–212 (1978), Teil II: Frequenz **32**(8), 220–226 (1978)
36. 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)

37. G. Gaertner, P. Janiel, D. Raasch, K. Schlageter, P.A.M. van der Heide, S. Jenkins, Accelerated life tests of oxide cathodes, ITG Tagung, in *Displays and Vacuum Electronics 2001, Proceedings*, ITG Fachbericht, vol. 165 (2001), pp. 219–224
38. G. Gaertner, D. Raasch, D. Barratt, S. Jenkins, Accelerated life tests of CRT oxide cathodes. Appl. Surf. Sci. **215**, 72–77 (2003)
39. M. Itazu, Life time estimation of cathode ray tubes, in *Japan Display* (1988), pp. 156–159
40. S. Jenkins, W. Johnstone, P.A.M. van der Heide, P. Rommers, M. van der Straaten, G. Gärtner, Investigations of the interface layer between cathode Nickel and oxide coating, ITG Tagung, in *Displays and Vacuum Electronics 2001, Proceedings*, ITG Fachbericht, vol. 165 (2001), pp. 231–236
41. J.-M. Roquais, F. Poret, R. Le Doze, P. Dufour, A. Steinbrunn, Initial chemical transport of reducing elements and chemical reactions in oxide cathode base metal. Appl. Surf. Sci. **201**, 85–95 (2002)
42. F. Poret, J.-M. Roquais, The base metal of the oxide-coated cathode. Appl. Surf. Sci. **251**, 31–41 (2005)
43. Y.C. Kim, K. Joo, J. Choi, H. Yang (Samsung): Improvement in oxide cathode life through Ni coating on cathode base metal. IVESC 2000, Technical Digest D-4
44. G. Gaertner, D. Barratt, Life-limiting mechanisms in Ba-oxide, Ba-dispenser and Ba-Scandate cathodes. Appl. Surf. Sci. **251**, 73–79 (2005)
45. W. Greulich (ed.), *Lexikon der Physik*, vol. 2 (Spektrum Akademischer Verlag, Berlin, Heidelberg, 1999)
46. G. He, T. Narushima, Y. Iguchi, T. Goto, T. Hirai, Electrical conductivity of Alkaline-earth metal-Aluminas and their application to a CO_2 gas sensor. Solid State Ion. **156**(3), 29–336 (2003)
47. G. He, T. Narushima, T. Goto, T. Hirai, Y. Iguchi, Synthesis and electrical property measurements of Ba-Alumina in a BaO-MgO-Al_2O_3 System. J. Ceram. Soc. Jpn. **106**(11), 1048–1050 (1998)
48. J. Poate, K. Tu, J. Mayer (eds.), *Thin Films-Interdiffusion and Reactions* (Wiley, New York, 1978). (especially: F. d'Heurle, P. Ho: Electro-migration in thin films (1978), pp. 243–304)
49. R.O. Jenkins, W. Trodden, The poisoning of barium-nickel matrix cathodes. J. Electr. Control **9**, 81–96 (1961)
50. R.O. Jenkins, A review of thermionic cathodes. Vacuum **19**(8) 353–359 (1969)
51. S. Itoh, M. Yokoyama, K. Morimoto (Futaba), Poisonous gas effects on the emission of oxide-coated cathodes. J. Vac. Sci. Technol. **A5**(6), 3430–3435 (1987)
52. S. Wagener, Proc. Phys. Soc. B **67**, 369 (1954)
53. H. Gallagher, Gas poisoning of cathodes, in *IEEE Conference rec., 9th Conference on Tube Technology* (New York, 1968), pp. 15–22
54. M. Pralle, *Massenspektrometrische Untersuchungen der Restgaszusammensetzung bei Aktivierung und Betrieb von Oxidkathoden, sowie der Emissionsvergiftung durch sauerstoffhaltige Gase* (Diplomarbeit, Fachhochschule Aachen-Jülich, 1999)
55. J. van den Berg, P. Boelens (Philips), Poisoning of thermionic cathodes by oxygen gas and argon ions, residual gases in electron tubes, in *Proceedings of 4th International Conference, Florenz 1971* (Academic Press, 1971, 1972), pp. 277–285
56. Y. Hayashida, T. Ozawa, H. Sakurai (MEC), Doping effects of the group III, IV, or V elements on the emitter of CRT oxide cathodes, IVESC'98, Ext. Abstracts p-9,137/8 and Appl. Surf. Sci. **146**, 7–11 (1999)
57. D. Barratt, G. Gaertner, Recent developments in oxide cathode research for CRT applications. Appl. Surf. Sci. **215**, 65–71 (2003)
58. D. den Engelsen, The oxide cathode revisited. IDW 2002, CRT5.1, pp. 627–630
59. D. den Engelsen*, G. Gaertner, Centennial of the oxide cathode, IDW 2004, CRT4–1
60. D. den Engelsen, G. Gaertner, Model of dopant action in oxide cathodes. Appl. Surf. Sci. **251**, 50–58 (2005)
61. K. Okumura, E.B. Hensley (Univ. Missouri), Drift and diffusion of activation in oxide-coated cathodes, J. Appl. Phys. **34**(3), 519–531 (1963)

62. E.B. Hensley (Univ. Missouri), Conduction mechanism in oxide-coated cathodes. J. Appl. Phys. **27**(3), 286–290 (1956)
63. R.T. Doloff, Electrical conductivity of barium oxide single crystals as a function of temperature and excess barium density, J. Appl. Phys. **27**(12), 1418–1426 (1956)
64. D.A. Wright, Nature **164**, 714 (1949) and Brit. J. Appl. Phys **1**, 150–153 (1950)
65. G. Gaertner, P. van der Heide, W. Johnstone, Kathodenstrahlröhre mit einer Verbundwerkstoffkathode. European Patent application 00308163.5, 19.9.2000, European Patent EP 1 232 512 B1, Cathode ray tube comprising a cathode of composite material, granted 7.1.2009; US-Patent 6833659, granted 21.12.2004, CRT with composite cathode
66. G. Gaertner, D. Raasch, C. Goodhand, P. van der Heide, T. Lee, D. Barratt, Oxidkathode. European Patent EP 1 232 511 B1, granted 15.8.2007, priority 19.09.2000 EP 00308164 and 02.05.2001 EP 01201836; US-Patent 7019450, granted 28.3.2006, Oxide Cathodes for Vacuum Electron Tubes
67. G. Gärtner, H. Lydtin (Philips), European Patent 0560436, Kathode mit einem Festkörperelement, granted 26.7.1995, priority German patent application DE-P 4207220.4 on 7.3.1992, US-Patent 5592043, Cathode including a solid body (Percolation Cathode), granted 7.1.1997, application 23.2.1993
68. L. Gmelin, Gmelins Handbuch der anorganischen Chemie: Ba, Barium, Verlag Chemie 1937; now Springer Gmelins Handbook of Inorganic Chemistry
69. N. Yamauchi, Thin films, electrical properties, in *Encyclopedia of Applied Physics* ed. by G. Trigg, vol. 21 (Wiley-VCH Verlag GmbH, 1997), p. 377
70. C. Kittel, *Einführung in die Festkörperphysik* (R. Oldenbourg Verlag, München, Wien, 1983)
71. J. Choi, K. Choi, K. Joo, S. Lee, Long-life oxide cathodes, in *IDW 1996 Proceedings* (1996), pp. 225–228
72. C. Herring, M. Nichols, Thermionic emission. Rev. Mod. Phys. **21**(2), 196 (1949)
73. G. Gaertner, D. van Houwelingen, Electron emission cooling of thermionic thoriated tungsten cathodes under high dc-loads, Elektronenröhren und Vakuumelektronik, NTG-Fachbericht, vol. 95,(VDE-Verlag, 1986), pp. 224–229
74. G. Dearnaley, A theory of the oxide-coated cathode. Thin Solid Films **3**, 161–174 (1969)
75. G. Dearnaley, Electronic conduction through thin unsaturated oxide layers. Phys. Lett. **25A**(10), 760–761 (1967)
76. W. Lindemann, A. van der Ziel (Univ. Minnesota), New mechanisms for the generation of Flicker noise in tubes with oxide-coated cathodes. J. Appl. Phys. **27**(10), 1179–1183 (1956)
77. R. Cortenraad, A.W. Denier van der Gon, H.H. Brongersmaa, G. Gaertner, D. Raasch, A. Manenschijn, Dynamic behavior of thermionic dispenser cathodes under ion bombardment. J. Appl. Phys. **89**(8), 4354–4364 (2001)
78. R.G. Ianoș, *Temperature and atmosphere influence during combustion synthesis of metal oxide (nano) powders* (Politechnical University of Timișoara, Thesis, 2015)
79. N. Morgulis, The Schottky effect for composite semiconductor electron emitters. J. Phys. **11**(1), 67–71 (1957)
80. R. Raju, C. Maloney, Characterization of an impregnated Scandate cathode using a semiconductor model. IEEE Trans. ED **41**(12), 2460–2467 (1994)
81. D.A. Wright, J. Woods (GEC), The emission from oxide-coated cathodes in an electric field. Proc. Phys. Soc. **65B**, 134–145 (1952)
82. H. Krusemeyer, M. Pursley (Bell), Donor concentration changes in oxide-coated cathodes resulting from changes in the electric field. J. Appl. Phys. **27**(12), 1537–1545 (1956)
83. L. Grey (GEC), Pulsed emission from the BaO-SrO-CaO system. Nature **167**, 522 (1951)
84. H. Lemmens, P. Zalm, Neue Entwicklungen auf dem Gebiet der Oxydkathoden - Oxydkathoden für Belastungen von 1 bis 2 A/cm^2, Philips Technische Rundschau Jahrgang **23**, 21–24 (1961/62)
85. T. Ohira, H. Teramoto, M. Saito, T. Shinjo (Mitsubishi), An analysis on the surface of the Ni-W layer of tungsten-film coating cathode. Appl. Surf. Sci. **146**, 47–50 (1999)
86. H. Teramoto, T. Ohira, M. Saito, T. Shinjo (Mitsubishi), An oxide cathode with tungsten -film coating. IDW'98, CRT-p 2, 453–456 (1998)

87. T. Sawada (Mitsubishi), High current density oxide cathodes for CRTs, in *Proceedings of IVESC 2004, Beijing*, pp. 104–106
88. S. Hodgson, A. Baker, A. Ray, A. Al-Ajili, J. Travis, C. Goodhand, T. Lee, P. van der Heide, Progress on the percolation cathode. IDW'99, CRT6-4, 1111+1112 (1999)
89. S. Hodgson, A. Baker, C. Goodhand, P. van der Heide, T. Lee, A. Ray, A. Al-Ajili, Processing and performance of a novel cathode material. Appl. Surf. Sci. **146**, 79–83 (1999)
90. T. Sugimura, M. Narita, R. Takeda, T. Hirai (NEC), A newly developed HIP cathode sintered under high temperature and high pressure. IDW'99, CRT6-1, 533–536 (1999)
91. C. Pantano, N. Wittberg, Surface studies of the medicus Ni matrix cathode. Appl. Surf. Sci. **4**, 385–400 (1980)
92. G. Medicus, US Patent 3615901 (1971)
93. W. Balas, J. Dempsey, E. Rexer (Honeywell), Oxide impregnated Ni matrix cathode. J. Appl. Phys. **26**(8), 943–948 (1955)
94. D. Maurer, C. Pleass, The CPC: a medium current density high reliability cathode. Bell Syst. Tech. J. **46**, 2375–2404 (1967)
95. X. Liao, X. Wang, Q. Zhao, M. Meng, Development of new types of oxide cathodes. Appl. Surf. Sci. **251**, 64–68 (2005)
96. X. Wang, X. Liao, J. Luo, Q. Zhao, An improved reservoir oxide cathode. Appl. Surf. Sci. **251**, 69–72 (2005)
97. X. Wang, S. Qi, Y. Liu et al., Influence of new material and technology on oxide cathode performance, in *Proceedings of IVESC-ICEE* (Saint Petersburg, Russia, 2014), pp. 272–273
98. O. Maslennikov, B. Katsnelson, A. Tyomin (Toriy), Advanced cathodes for use in cathode ray devices, IVESC 2000, Orlando (Florida), Technical Digest E5, and http://inspirehep.net/record/920872/files/HEACC98_336-338.pdf
99. R.J. Umstattd, Advanced electron beam sources, in *Modern Microwave and Millimeter-Wave Power Electronics (chapter 8)* ed. by R. Barker et al. (Wiley, 2005), pp. 393–444
100. R.J. Umstattd, *Plasma deposition of oxide-coated cathodes* (University of California at Davis, Thesis, 1998)
101. G. Gaertner, D. Raasch, Kathodenstrahlröhre mit dotierter Oxid-Kathode, German Patent application P 100 45 406.2.0, priority 14.9.2000; US- patent 6600257, granted 29.7.2003: Cathode Ray Tube Comprising a Doped Oxide Cathode
102. G.A. Haas, R. Shih, R. Thomas, Determination of conducting band edge and electron affinity in surface potential measurements of BaO. J. Appl. Phys. **47**, 5400–5404 (1976)
103. P. van der Heide, G. Gaertner, D. Barratt, Lifetime performance of cathodes in CRTs, Information Display Workshop IDW´03, Fukuoka, Japan, Technical Digest CRT3.4 (2003)
104. A. Venema, Thermionic emission, Handbook of Vacuum Physics Part 2, 181–298 (1964)

Chapter 5
Cathodes of Medical X-Ray Tubes

Rolf Behling

Abstract Cathodes of typical medical diagnostic X-ray tubes operate under adverse conditions. The X-ray focal spot on the anode in view of the cathode may generate metallic and other vapors. The residual gas pressure is comparatively high, as clinical tubes are not subject to active evacuation during operation. Subcomponents like the anode may release significant amounts of gas. Electron currents of up to 1 A or more produce ions under high electric fields, which may bombard the cathode. Ion energies may reach 150 keV. Occasional vacuum plasma discharges may appear close to or even invoking the electron emitter. Tubes for computed tomography may experience centrifugal acceleration of more than 30 times the gravitational acceleration. There exists an indisputable quest for utmost reliability. All these requirements conserved the classical thermionic tungsten emitter as the workhorse of medical diagnostic X-ray tubes, which will be the primary focus of this article.

5.1 Introduction

5.1.1 X-Ray Tubes for Medical Diagnostics

A modern treatment of diagnostic X-ray sources can be found in Behling [1]. The text covers history, actual status, and future developments of the technology in a comprehensive manner and was the initial basis for the following treatment. This section will concentrate on the subcomponent cathode which releases and accelerates electrons and focuses and directs electron beams for the purpose of generating braking-radiation upon impact on atomic nuclei of the target material. X-ray tubes are not only the current workhorses for medical X-ray imaging. These vacuum electronic devices are also expected to remain without practical mid- or long-term alternatives for the generation of the required photon flux, see [24].

R. Behling (✉)
XTraininx—Technical Consulting, Hamburg, Germany
e-mail: admin@XtrainingX.com

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*,
Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_5

5.1.2 The Basics of the Generation of Bremsstrahlung

Since the early days of X-ray production in ion tubes of the type shown in Fig. 5.1, the bremsstrahlung process has been dominating X-ray generation for medical diagnostics and will continue doing so. In spite of the seeming energetic inefficiency of the conversion of electrical power to X-ray intensity, no better ways of X-ray generation have been introduced into the breadth of the clinical routine. As an example, an advanced spectral detection system for computed tomography (CT) is depicted in Fig. 5.2. Its technical concept is based on the polychromatic nature of bremsstrahlung. This Philips system shown operates a modern premium performance X-ray tube. A cut model of this advanced vacuum electronics device is depicted in Fig. 5.3. As all bremsstrahlung sources do, it generates electromagnetic radiation of sub-nanometer wavelength by acceleration and deceleration of electrons during interaction with the nuclei of the target matter. For a nonrelativistic electron, the total radiated power P is, according to the classical Larmor formula and stated in SI units, $P = \frac{e^2}{6\pi\varepsilon_0 c^3}\left|\ddot{X}\right|^2$. The character e denotes the electron charge, **X** the position vector, c the speed of light, ε_0 the vacuum dielectric constant. During this process, only a small fraction of the kinetic energy is converted to X-ray energy. Losses in the anode are caused, e.g., by exciting plasma oscillations in the electron cloud. The gross energy conversion factor for medical X-ray tubes amounts to about 1% only, even for the high-Z-target

Fig. 5.1 Ion tube with a water-cooled target, manufactured by C.H.F. Müller, Hamburg, ca. 1910
Picture courtesy of Philips

Fig. 5.2 Philips computed tomography system iQon® with spectral detection ("color X-ray"). Polychromatic bremsstrahlung from the X-ray source is attenuated by the patient. In addition to the X-ray intensity, this unique system analyzes the spectrum at the detector downstream of the patient. It enables spectral differentiation between bones, tissue, and contrast agent (iodine). The electron density per volume element of the patient is measured in addition to the attenuation per volume element. In comparison with other concepts, the spectral detection concept delivers the best achievable temporal registration and clinical workflow Picture courtesy of Philips. Adapted from Behling [1]

material tungsten ($Z = 74$). The benefit of materials with a large atomic number is their three-dimensional matrix of nuclei and electrons which generates large electric fields and strong acceleration and deceleration of traversing electrons.

In contrast to visual light, bremsstrahlung for medical diagnostics of humans emerges in an isotropic manner from the target and cannot be efficiently focused. Due to the absence of suitable X-ray lenses for the energy range of human X-ray imaging of between about 16 keV (mammography) and up to 150 keV (chest exposures and CT), the size of the focal spot determines on one hand the spatial resolution which a "shadow imaging" system may deliver. It also limits the electrical input power which the anode can withstand. The term medical X-ray "beam" may be misleading. Fans of X-rays are formed by annihilating unwanted or even hazardous radiation with apertures which delimit the illuminated volume. Typically, 99% of the generated photons end up in these radiation shields. Thus, the total output power of usable X-rays is about four orders of magnitude lower than the electric input. Heat management of the target and other subcomponents of an X-ray tube is essential.

Kramers' law associates the X-ray spectrum with the tube voltage. Other than, e.g., for laser-plasma-generated fast electrons, the thermal energy of the electrons

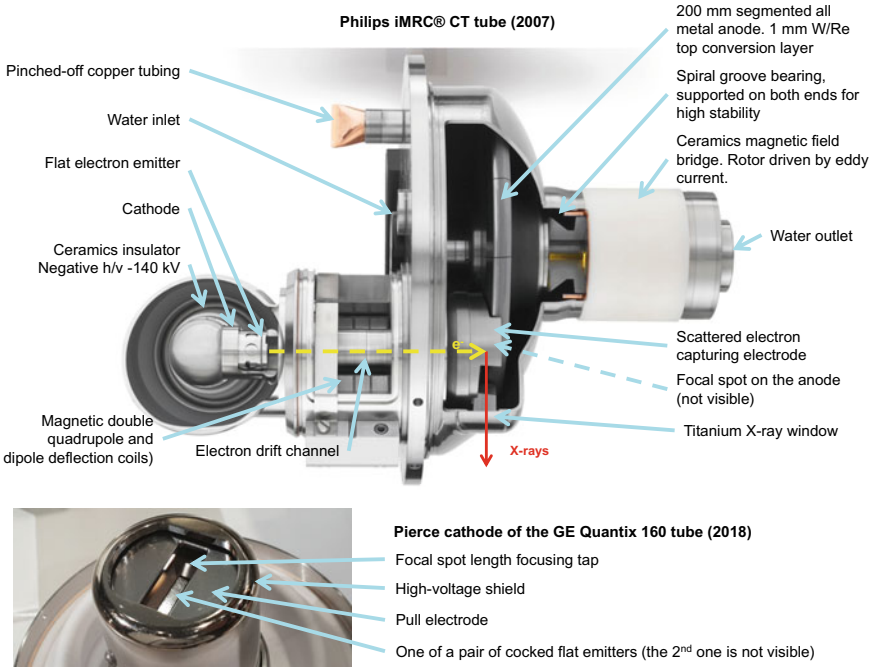


Fig. 5.3 Top: Cut view of the X-ray tube Philips iMRC® for the computed tomography system iQon, shown in Fig. 5.2 (input power 120 kW for 4 s, focal spot dimension 1.1 mm width, and 1.2 mm X-ray optical length, FWHM). The tube revolves about 4 times per second about the patient with a centrifugal acceleration of more than 30 g. A similar tube was launched by Siemens (Vectron®) in 2013 and GE (Quantix® 160) in 2018. Bottom: Pierce-type cathode of the GE Quantix® tube. While Philips' and Siemens' cathodes operate in saturation mode (low electron return rate), the Quantix is space charge controlled by a pull electrode Adapted from Behling [1], Fig. 6.13. Picture courtesy of Philips

released from the thermionic emitter of a medical X-ray tube (a fraction of an electron volt) is usually four to six orders of magnitude smaller than the energy which the electrons gain in the electric pull field before they impinge on the target. Thus, the tube voltage determines the kinetic energy of the electrons at impact and with it the X-ray spectrum, see Behling [1], Fig. 2.8. The tube voltage also defines the Duane–Hunt limit, that is, the maximal achievable photon energy. The X-ray spectrum determines the image contrast and the degree of attenuation in the object which the X-ray camera is to image. Thus, the spectrum has to be optimized for best visibility of object features at minimal X-ray dose.

The cathode of a medical diagnostic X-ray tube has to supply electron currents of between a few milliamperes and about one ampere. Given the available means for focusing the electrons into the source area of the X-rays on the anode, the so-called focal spot, current densities at the electron-emitting surfaces reach a few A/cm^2. Space charge effects have to be carefully considered.

5.2 Conditions of Operation of Medical Diagnostic X-Ray Tubes

It has always been important to control the vacuum conditions inside of X-ray tubes. Only Conrad Röntgen's very first tubes operated attached to an active mercury vacuum pump. But, the active exhaust is neither acceptable nor necessary in current clinical practice. As shown in Fig. 5.4, diagnostic X-ray tubes operate pinched-off from any active vacuum pump. As a consequence, the vacuum conditions in these tubes are usually rather poor. The basic residual gas pressure is in the order of 10^{-7} Pa when the tube in the idle state. It may reach up to 10^{-2} Pa when the anode is at maximal temperature or in the event of a high-energy vacuum discharge. These latter events usually occur in a stochastic manner. Their frequency may range from one per year for excellent tubes up to several per second for poor or damaged tubes. Vacuum discharges are caused by electronic instabilities of the electrode surfaces, see, e.g., Fursey [2]. The gas released is typically mainly comprised of CO, H_2, and to a minor extent water (in particular in glass tubes) and hydrocarbons. Chemical

Fig. 5.4 A glass tube for general radiography is pinched-off from the mechanical vacuum pump after the exhaust process is completed. From this point onwards, the vacuum inside the tube is maintained by ion implantation into electrodes and to a smaller extent, by a chemical getter (see Behling [1], Fig. 6.92)

getters help stabilizing the vacuum quality, see Behling [1], Fig. 6.92. Inert gasses, primarily argon, may be present when internal leaks from hidden voids may release residual air into the volume. Metal vapor adds to the blend of atoms and molecules impinging on the cathode.

Thus, summarized, cathodes of typical medical diagnostic X-ray tubes operate under adverse conditions:

- metallic vapor from the anode, in particular from the focal spot, see Fig. 5.5;
- residual gas pressure above the UHV level during desorption of contaminants form hot anodes and other subcomponents, thermally, electron or ion stimulated, see Fig. 5.6;
- ion bombardment, see Fig. 5.7a;
- emitter temperatures between 1800 and 2600 °C, recrystallization, see Fig. 5.7b and d;
- tube voltages of up to 150 kV across distances of about 2 cm, see Fig. 5.3;
- tube currents of up to 1,5 A, see Fig. 5.8;

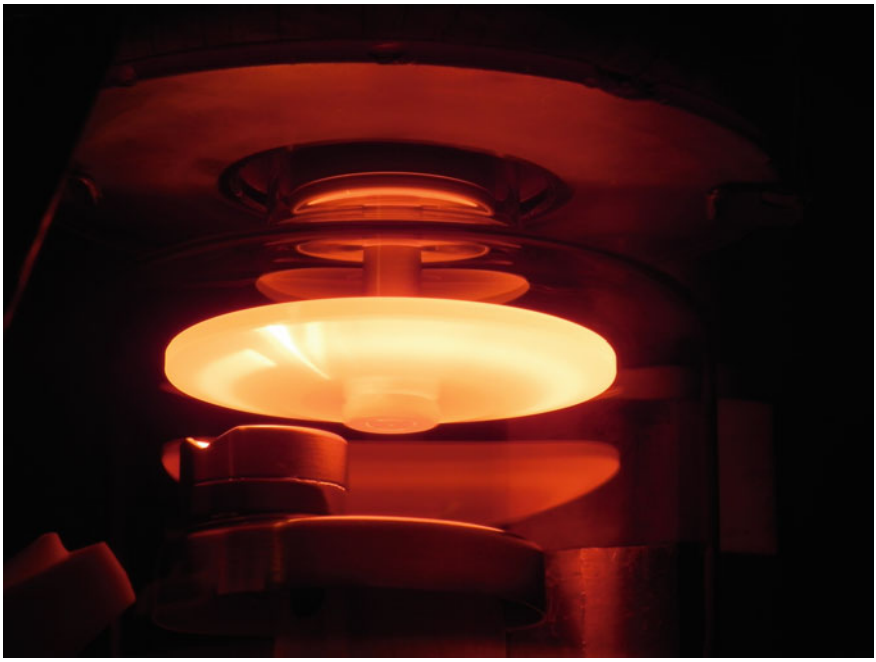


Fig. 5.5 Degassing of a glass tube for general radiography. Electrical power is converted into heat and to a small extent X-ray intensity on the surface of the rectangular focal spot, which leaves a hot comet tail on the focal track. The anode of this sample tube has a diameter of 90 mm and rotates for this picture with 49 Hz. Some visual light from the hot tungsten electron emitter is reflected from the anode as well and visible as a large radial streak parallel to the focal spot

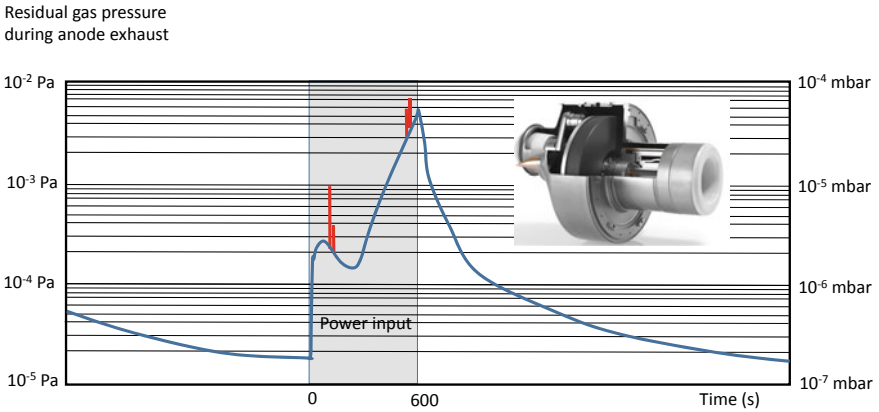


Fig. 5.6 Typical residual gas pressure during degassing of a typical X-ray tube for computed tomography. The pressure was measured at the vacuum pump of the exhaust station. The reading of the gauge was multiplied by a conversion factor such that the chart reflects the real pressure inside the tube. Spikes on top of the continuous curves illustrate pressure bursts from vacuum discharges. Peak values at the cathode inside the tube are typically much higher than indicated. The spikes are smoothened out due to the limited gas conductivity of the vacuum tubing and the volume of the measurement chamber

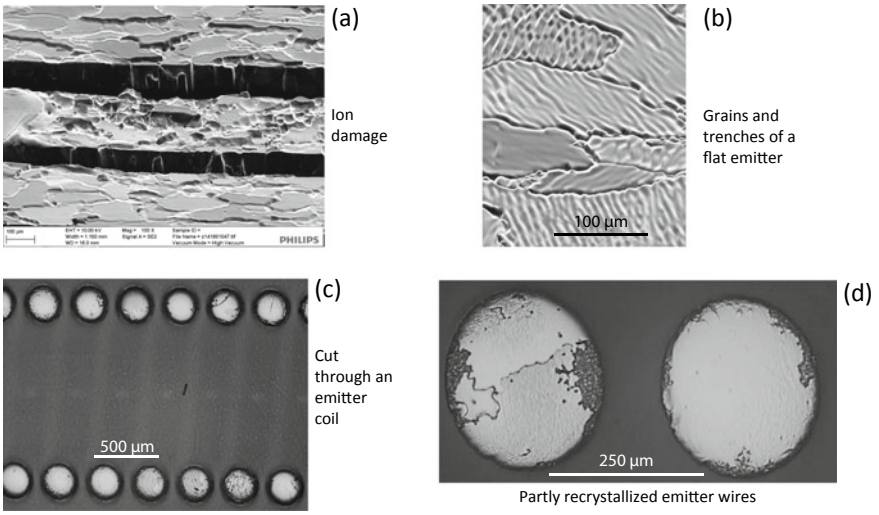


Fig. 5.7 Close up of thermionic tungsten electron emitters for medical X-ray tubes. **a** Flat emitter surface eroded by ion bombardment. Ions of high kinetic energy may originate from zones close to the focal spot on the anode, where the primary electron density and the density of backscattered electrons as well as the vapor pressure of anode material are all highest. **b** Grain structure at the surface of a flat electron emitter after a period of operation. A heterogeneous structure as well as trenches between grains are visible. **c** is a cut view through a coil emitter as used in the cathode in Fig. 5.12. The cut **d** presents more details of the grain and surface structure of a partly recrystallized coil emitter. Adjusting the optimal state of recrystallization is crucial for the alignment and the long-term stability of the emitter within the focusing structure

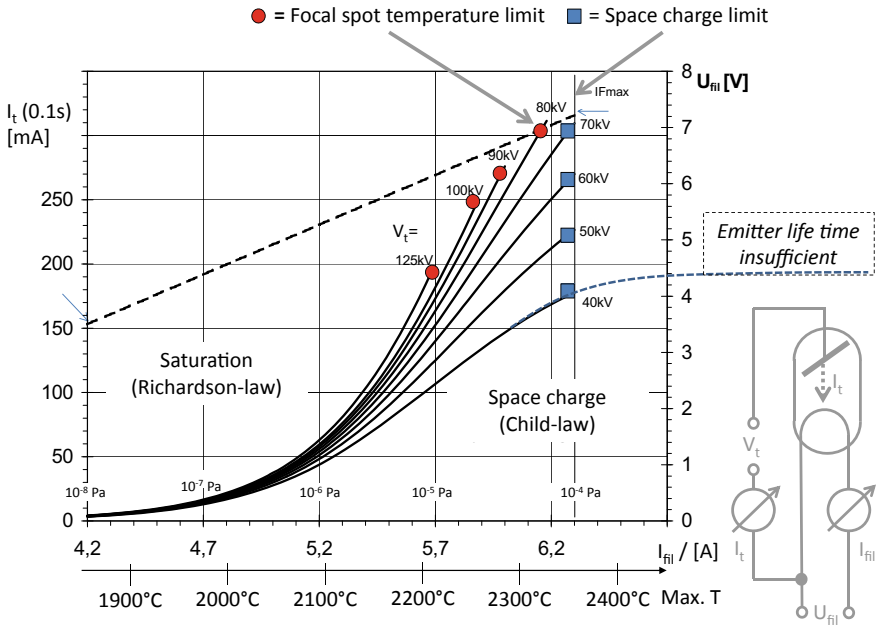


Fig. 5.8 Maximal tube current I_t as a function of heating current I_{fil} and wire temperature T for the small focal spot of a sample tube manufactured by Philips. The maximal tube current permitted for 0.1 s exposure time is shown, which is limited by the absolute temperature and the temperature gradient in the focal spot. The dotted curve indicates the voltage U_{fil} for driving the AC heating current. The tube has a tungsten coil emitter, delivers a focal spot size of 0.6 mm width and 0.9 mm projected X-ray optical length in the central beam, an anode angle of 11°. The anode runs with a focal spot track speed of 23 ms^{-1}. Philips recommends a maximal heating current of 6.3 A. The dotted line extending to the right illustrates the space charge limitation of the cathode for 40 kV tube voltage. Emitter temperatures in excess of the recommended maximum should not be applied for the sake of reliability. The gain of tube current would be minor Adapted from Behling [1], Fig. 6.18

- current densities of about 3 A/cm^2 at the emitter and up to 50 A/cm^2 at the target, see Fig. 5.9;
- sudden vacuum plasma discharges with pressure bursts in the order of kilobars in their center between electrodes, near the electron emitter, see Fig. 5.10;
- centrifugal acceleration of more than 30 g in a computed tomography system, see Fig. 5.2;
- 10^6 s of operation until the tube may be replaced, see Behling [1], Sect. 9.5.

In a large portion of examinations, patients undergo an interventional procedure during X-ray imaging. Consequently, there exists an indisputable quest for utmost reliability.

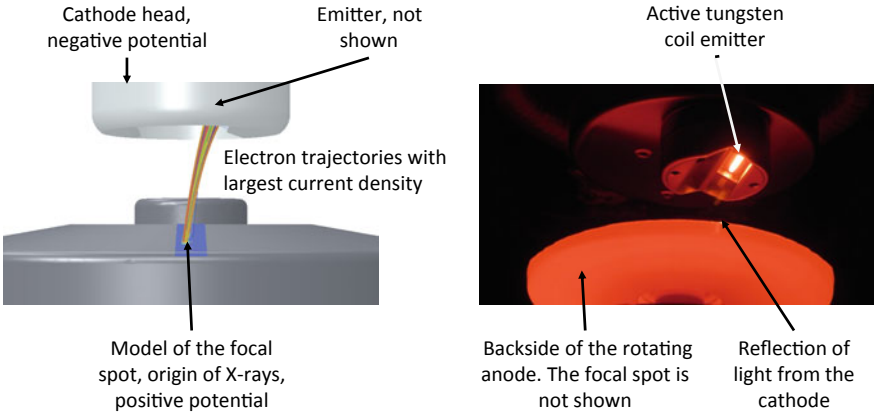


Fig. 5.9 Electron ray tracing (left), which delivers the core of the focal spot and renders the nominal focal spot size. The real situation of an operating thermionic cathode in a rotating anode glass tube during exhaust is shown to the right Adapted from Behling [1], Fig. 6.23

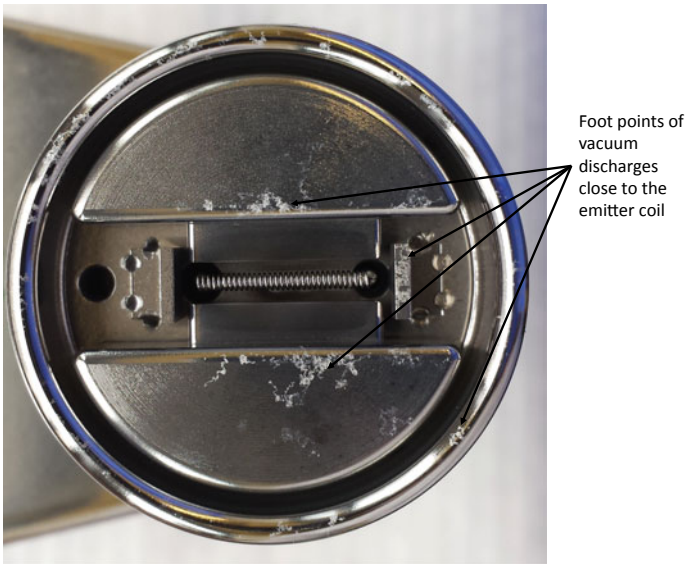


Fig. 5.10 Craters left on the cathode after multiple vacuum discharges in a bipolar metal center section tube. Micro-craters represent foot points of discharge events, see Fursey [2]. Craters concentrate in the zones of high electric field, but may also appear on plain surfaces

5.3 Release of Electrons into the Vacuum Space—Historical Development

The earliest electron sources for intentional X-ray generation were electrodes bombarded with ions from gas discharges. For more than 20 years since the discovery of X-ray in 1895, ion tubes with stationary targets were used and optimized in several ways, as the one depicted in Fig. 5.1. "Soft" tubes which allowed for high tube currents even at low tube voltages delivered high soft-tissue contrast. But, the image quality suffered from reduced penetrating power of the produced X-rays. Over time, the gas content diminished by ion implantation. Soft tubes "hardened". Reduced gas pressure required enhancing the tube voltage. At some point, hard tubes failed totally, could not maintain a suitable gas discharge, and had to be replaced. Already in 1882 Goldstein had prevented these gas discharges from extinguishing at low pressure by heating the cathode, see Dörfel [3], page 5. This was early evidence that the thermionic Edison or Richardson effect of thermal ionization of the cathode was a suitable supplement or even a replacement for the gas discharge. Around the year 1913, Coolidge, [4] and Lilienfeld, see Lilienfeld [5], proposed thermionic emission of electrons from hot tungsten. A tube of the Lilienfeld type, depicted in Fig. 5.11, is on display in the German Röntgen Museum, Remscheid-Lennep, Germany. Since

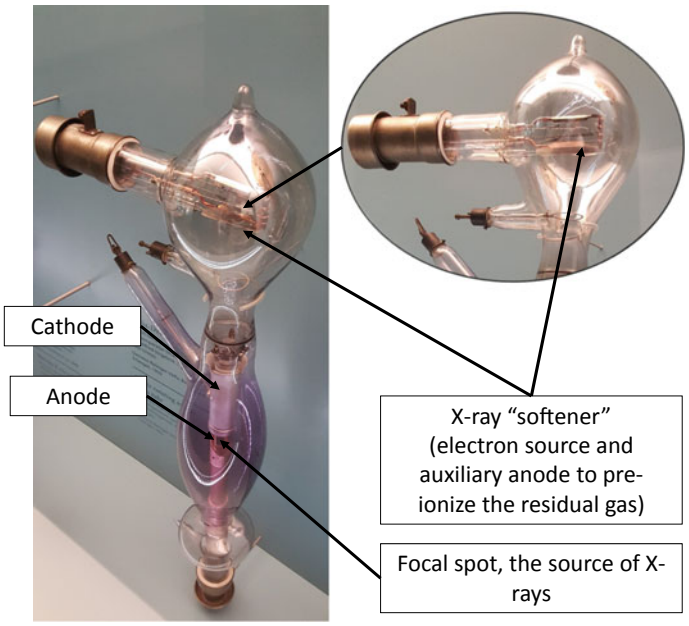


Fig. 5.11 Photo of a Lilienfeld tube, the predecessor of tubes with thermionic electron release. The tube is on display at the German Röntgen Museum, Remscheid-Lennep, Germany

Coolidge and Lilienfeld, tube current and tube voltage could be controlled independently. This feature is essential. X-ray intensity and X-ray spectrum could from then on be optimized separately to minimize the patient dose of ionizing radiation and to employ the maximum of the thermal capacity of the anode. The Lilienfeld-tube marks the transition from ion tubes to tubes with pure thermionic generation of electrons, commercialized by Coolidge.

Figure 5.3 illustrates the function of a high-performance rotating anode X-ray tube for a Philips high-performance computed tomography system IQon in Fig. 5.2. A simpler glass tube is shown in Fig. 5.4. For details of the combination of source and detection technology see Shefer [6]. In contrast to other vendors' systems, which alter the source spectrum for spectral differentiation of an object, the Philips concept allows for spectral differentiation of the polychromatic X-rays at the detector end instead. This concept is highly beneficial in view of the temporal registration of the projections and for the clinical workflow.

Cathode and anode of the high-performance tube used are enclosed in high vacuum inside a metal envelope. X-rays emerge from a rectangular area hit by impinging electrons on the truncated cone of the rotating anode. For electron release, the cathode comprises a flat tungsten electron emitter, which is directly heated by an auxiliary current of some amperes. A cathode "cup" and magnetic quadrupoles are focusing the beam. The anode disk is a segmented sintered compound of tungsten with some percentage of rhenium at the top and an underlying bulk of alloy of molybdenum, zirconium, and titanium. The exemplary anode primarily cools by heat conduction through a liquid metal lubricated spiral groove bearing and may reach bulk temperatures of 1500 °C and focal spot temperatures of ca. 3000 °C.

For standard applications in radiology, simpler glass tubes are used. They allow looking into the interior during operation, see Fig. 5.5: The rotating anode was heated by electrons from thermionic emission in the cathode at the bottom. The glowing emitter at the bottom of Fig. 5.5 is an isolated coil of tungsten wire of a quarter millimeter diameter. Close-ups are shown in Figs. 5.10, 5.12, and 5.13.

5.4 Tungsten Emitters

5.4.1 Why Still Tungsten Emitters?

Various alternative concepts of electron emission are discussed in Behling [1], Figs. 6.12, 6.13, 6.14, 6.15, 6.16. Thermal ionization of a tungsten wire or a flat sheet of tungsten has proven to be the most robust electron emission concept, given the adverse conditions in a medical X-ray tube. Usually, the temperature of the tungsten wire or flat sheet emitter is adjusted to between 1900 and 2500 °C. Inevitably, this causes evaporation of tungsten and limits the emitter lifetime, see Karyugin et al. [7]. Emitter tests in ultra-high-vacuum (UHV) of the order of 10^{-7} Pa often result in larger life times than in the real practice of an operated X-ray tube. When an

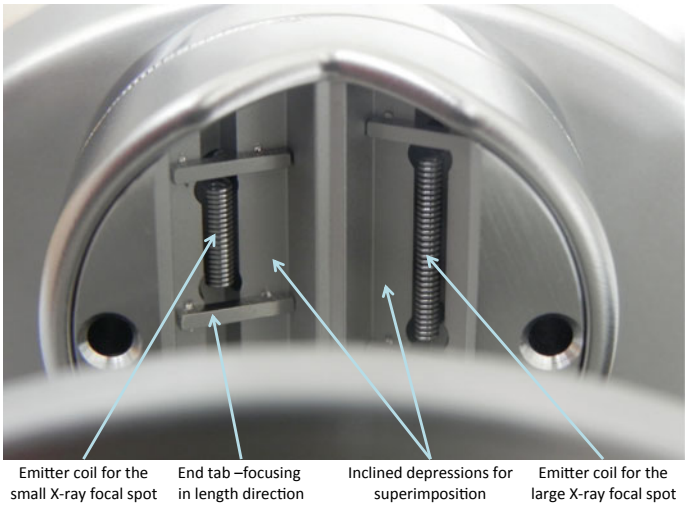


Fig. 5.12 Close-up of a dual emitter cathode for a tube with superimposed focal spots on the anode. Adapted from Behling [1], Fig. 6.21

anode is present, residual gas is released and partly ionized by the electron beam. If not shielded or deflected by special means, at least a portion of the produced ions may bombard the origin of the electrons, i.e., the electron emitter, and may destroy it, see Fig. 5.7a. Rotating anodes constitute a further challenge. During rest, the cool anode adsorbs residual gas. Thermal desorption upon application of high voltage immediately releases the material in a gas burst during the first few anode rotations of some milliseconds duration. The local gas and vapor density around the cathode may, therefore, be much higher than measured with a remote gas analyzer. Figure 5.6 shows the gas pressure at such a remote gauge with temporal damping at a time constant of ca. 100 ms.

Tungsten emitters are capable to withstand nonoxidizing and non-carbonizing vapors with pressures of up to more 10^{-2} Pa and ion bombardments from tube currents of an order of an ampere in the same gas atmosphere. High-voltage discharges and high-pressure plasma bursts, which have their foot points close to the emitter, will also usually not lead to destruction.

Failure of the cathode in a medical device may be hazardous. Robustness and reliability are key. Other concepts of electron release, like dispenser cathodes or carbon nanotube field emitters are lacking the degree of robustness of the tungsten emitter. Tungsten wire and flat emitter technology have matured. “Non-sag” tungsten wires with stabilizing additives, e.g., 30–70 ppm potassium, also withstand the necessary temperatures of up to 2500 °C while maintaining sufficient mechanical strength. As potassium does not dissolve in pure tungsten due to lattice mismatch, this additive preferably accumulates in voids and crossings of grain boundaries. Islands of potassium prevent grains from growth and guarantee mechanical stability. Maintaining the mechanical integrity of flat electron emitters is more challenging. The sheet material

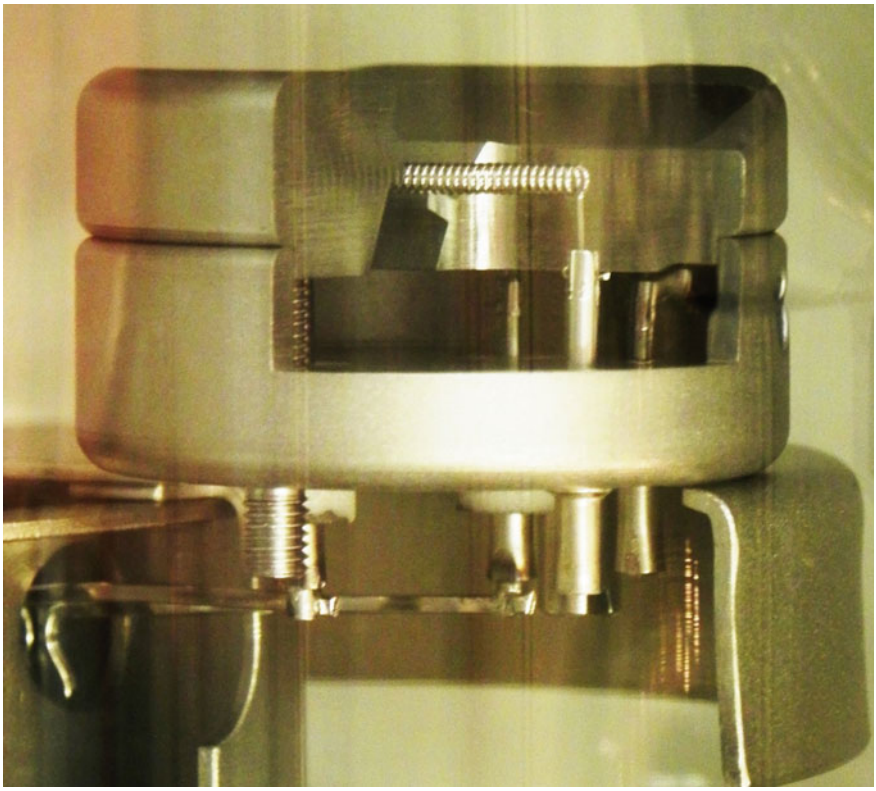


Fig. 5.13 Wiring of the cathode of a glass tube for general radiology

is rolled and annealed in multiple steps. The proper state of recrystallization, orientation of grains relative to the meander structure, centrifugal forces during operation, thermomechanical stress relief in the supporting structure, and more aspects have to be kept well under control.

5.4.2 Work Function of Tungsten

Among the metallic elements, tungsten has a comparatively large work function W_W between 4.2 and 5.3 eV, depending on the crystalline orientation of the emitting surface, as presented in Table 5.1. In practice, tungsten emitters present a mix of orientations at their surface. The average work function equals about 4.5 eV. Figure 5.7a and b gives an impression of the polycrystalline structure and the mix of grain orientations at the surface of a real flat tungsten electron emitter. Figure 5.7a illustrates the result of ion bombardment of an un-screened surface. Figure 5.7b shows a close-up of the grain structure of a used flat emitter. Cut views of coil emitters are in Fig. 5.7c

Table 5.1 Work function of tungsten as a function of the orientation of the emitting crystal surface, after Skotnicová et al. [8]

Miller index	(310)	(111)	(100)	(211)	(110)
Work function	4.2 eV	4.4 eV	4.6 eV	4.9 eV	5.3 eV

and d. It is clearly visible that the emitting surface has a heterogeneous topology for flat as well as coil emitters. The orientation of grains and the effective average work function depend on the history of the emitter. The importance of the micro-topology of the emitter surface has been demonstrated by Barmina et al. [9].

5.4.3 Some Basics of Thermionic Emission

Figure 5.14 illustrates the energetic conditions of electrons during thermionic emission in a high electric field in which most emitters in medical X-ray tubes operate. The emission scheme has been described by O. W. Richardson and was named after him. In contrast to space charge limited emission, the electric field is large enough to reduce the share of electrons $\tilde{r}$ returning to the emitter after their primary departure from the metal to between 10% and about 50%. As the shape of emitters in medical X-ray tubes is usually far from planar the Richardson equation is not fully applicable. Most emitters are coil shaped or resemble meandered flat bars in modern flat emitters. Gaps, trenches, and interleaved electrodes disturb the electric field. Although the field which acts on a flat emitter is usually much more homogeneous, even these generally operate in a mix of space charge limited and thermionic emission.

5.4.4 Characteristics of Cathodes—Emission Chart

Emission charts are an important metric of the performance of a medical X-ray tube. They indicate not only the performance of the cathode, but also the match with the thermal capacity of the anode. Figure 5.8 shows an exemplary emission chart for an interventional angiography tube with a tungsten coil emitter. Its characteristic is best described by both, space-charge-limited emission and thermionic emission. Space-charge-governed emission follows the Child–Langmuir $V_t^{3/2}/d^2$ law, V_t being the tube voltage and d the cathode–anode distance and the Richardson equation copied in Fig. 5.14 with its exponential temperature dependency. The low current area of Fig. 5.8 is characterized by thermionic emission. Space charge effects are minor in this regime.

Squares terminate the lines of tube currents in Fig. 5.8 at the maximum permitted emitter temperature. This temperature is a function of the required tube life and the assumed tube application. At tube voltages V_t below the iso-watt point, that is, the

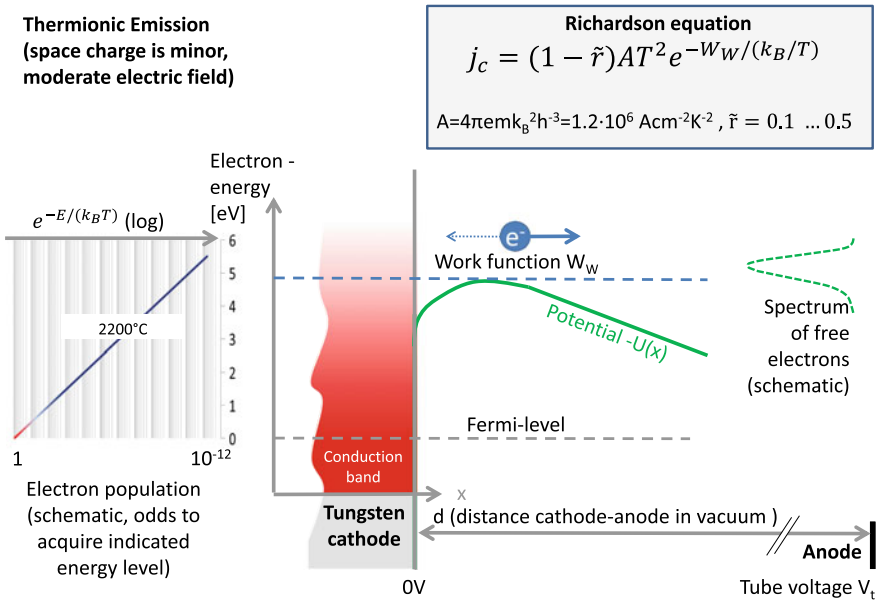


Fig. 5.14 Scheme of thermionic emission, quantified by the Richardson equation. Space charge and with it the electron reflection factor r˜ are assumed to be small. The electric pull field is generated by the tube voltage V_t between cathode and anode along the distance d. j_c denotes the current density at the emitter surface, ε_0 the vacuum dielectric constant, e the absolute value of the electron charge. m stands for its rest mass. The metal is characterized by its work function W_W. A is the Richardson–Dushman proportionality factor, T the emitter surface temperature, k_B Boltzmann's constant, h Planck's constant. The graph on the left indicates the probability of an electron in the partly empty conduction band to thermally acquire kinetic energy above the baseline Fermi level. The spectrum of energies of the emitted electrons relative to the Fermi energy is shown at the very right. Other emission mechanisms are treated in Behling [1], Sect. 6.2.1.2 Adapted from Behling [1], Fig. 6.13

tube voltage where cathode and anode limitations meet, coil emitters produce an approximately linear relationship between maximal tube current I_t and tube voltage. For higher tube voltages beyond the iso-watt point, the temperature of the focal spot on the anode is limiting, indicated by the circles.

5.4.5 *Heating of the Emitter*

The emitter, which is on cathode potential, is heated by a direct heating current of several amperes, which is supplied by a current source of the generator through the high-voltage cable.

The surface temperature of an emitter varies along its emitting area. Tube current adds to heating current.

During departure, emitted electrons extract power from the emitter at the amount of the work function (typically 4.5 eV for tungsten) times the emitted tube current (typically between a few milliamperes and an ampere). When not compensated by raising the heating current (which is usually done automatically by the high-voltage generator), the emitter inevitably cools down. This causes a significant drop in the tube current as soon as the tube voltage is switched on. The time constant depends on the thermal capacity and heat conduction of the tungsten emitter. It corresponds to the time for the initial heating of the emitter prior to X-ray exposure. The cooling power loss is minor, usually in the order of one watt, in relation to the up to several dozen watts for emitter heating. But, when operating in saturation emission (high tube voltage, low tube current, minor, or no space charge) the high sensitivity of the current density with respect to the temperature of the emitting surface results in a well noticeable drop of the tube current. Usually, the heating control circuitry of the high-voltage generator is programmed to compensate for this. Figure 5.15 illustrates the typical response of the tube current of a sample tube for cardiovascular application on emitter cooling. The curves were measured with a deactivated generator heating

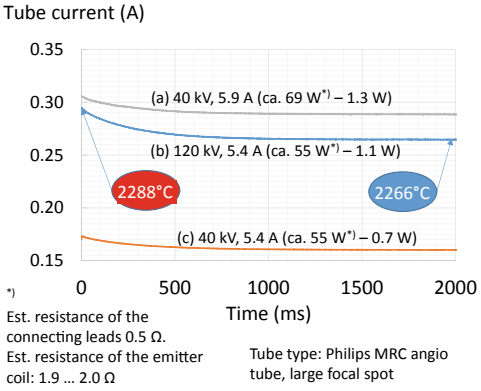


Fig. 5.15 Left chart: tube currents of the large focal spot of an exemplary angiographic X-ray tube. The emitter is preheated by the indicated heating current before the exposure starts and brought to a steady-state temperature. After switching on the indicated tube voltage at time zero, the electron emission starts and the tube current drops after an initial peak. The curves are labeled with tube voltage, heating current, and (in brackets) coil heating power and emission cooling power. **a** Space charge limits the tube current with high emitter temperature and low tube voltage. Changing temperature hardly impacts the emitted electron current. The emission current is dominated by space charge limitation and basically follows the Child–Langmuir law. This results in nearly constant tube current, irrespective of emission cooling. Further, the emitter temperature is stabilized by the high share of heat radiation at high temperatures compared with the heat dissipated by emission cooling. **b** The other extreme is shown. Space charge is nearly absent in this case. The emitter temperature rules the emission current according to the Richardson law, see Fig. 5.14. The visible 10% drop of the tube current over a time period of 500 ms is indicative of a reduction of the emitter temperature by 22 K. The time pattern reflects the thermal time constant of the emitter coil. **c** Mixed mode. Part of the emitter coil operates space charge limited, while the center part, which is subject to the highest electric field, is in saturation mode. Right: A Picture of the situation Adapted from Behling [1], Fig. 6.9

control loop. After the wire has reached a steady-state temperature upon application of a defined heating current, high voltage is switched on at time zero. The tungsten temperature drops with a typical time constant of several hundred microseconds. This time constant depends on the thermal capacity of the emitter and the share of heat dissipation by radiation cooling and heat conduction. Emission cooling is most clearly visible when space charge is absent and saturation emission dominates at high tube voltage and small heating and tube current. The other extreme occurs at low tube voltage and high emitter temperature. Usually, the cathode operates in a mixed mode.

The time constant measured during cooling equals the constant for initial heating. In clinical practice, the preparation time prior to exposure will be shortened by stand-by heating. In the idle phase of a medical diagnostic system, the temperature of the emitter is adjusted such that no measurable electron current will be emitted (below ca. 1800 °C) but the electrical resistance of the tungsten emitter is at an elevated level. Heating from the cold would take additional time, as the resistance and with it, the heating power at constant current would be smaller, as for most metals.

Thermionic emitters deteriorate in several ways. The primary cause of failure is partial evaporation. Vapor pressure data for tungsten sheet metal are indicated in Fig. 5.8. Hot spots, trenches between grains and voids build up. When operated with constant heating current, the local heating power grows with the electric resistivity, and causes run-away destruction. Surface diffusion of tungsten from hot to cold zones, thermomechanical stress, ion bombardment, and reactions with residual gas contribute as well. Carbonization of tungsten emitters is often experienced during the initial exhaust process of tubes with carbon anodes as soon as hydrocarbons of elevated molecular mass are desorbing. During this exhaust procedure, residual gas is intentionally mobilized from subcomponents of a tube to be pumped. A typical pressure curve measured at the depicted X-ray tube type is shown in Fig. 5.6. The emission capacity of the tungsten surface may drop below half of the initial value in cases of such chemical “emitter poisoning”, while tube voltage and heating currents are kept constant. This apparent increase of the work function can be recovered to a large extent by continued heating of the tungsten when the tube is processed properly and the hydrocarbon attack is kept under control. In clinical practice, deterioration of the electron emissivity is a strong indicator of oil leaking into the vacuum, and often accompanied by high-voltage instabilities.

5.5 The Electron Beam

5.5.1 Beam Focusing and Focal Spot Metric

Most tube types with focal spots of different lengths and widths are equipped with multiple emitter coils. The focal spot size is adapted to the clinical question and selected by heating the respective emitter. Figure 5.12 is a close-up, seen from the

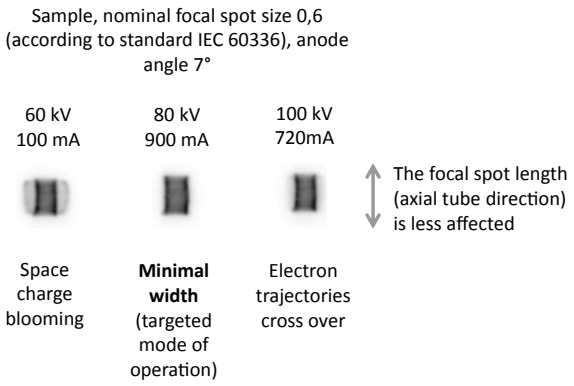


Fig. 5.16 X-ray focal spot photographs, rendered by a pinhole camera as depicted in Fig. 5.17. The focal spot size is depending on the amount of space charge in the focusing space Adapted from Behling [1], Fig. 6.9

anode. Figure 5.9 shows the result of the ray tracing modeling of a single operating tungsten emitter coil. As long as space charge can be ignored, purely electrostatic focusing by the local shape of the cathode head is invariant with respect to changing the tube voltage. Neither focal spot size nor the current density distribution pattern varies with tube voltage. Thus, controlling the size of the focal spot is simple. However, as aforementioned, usually, a significant amount of electronic space charge is generated around coil emitters using standard operating conditions of tubes with cathodes as shown in Fig. 5.12. Figure 5.16 illustrates the impact of space charge on the focal spots of such tubes, the so-called blooming effect, measured with a focal spot pinhole camera as shown in Fig. 5.17.

5.5.2 Advanced Electron Optics

Recent developments of the focusing technology for medical X-ray tubes have significantly reduced space charge limitations which appear in legacy X-ray tubes with thermionic emitters, see Behling [10]. Flat directly heated tungsten sheet emitters were commercialized on a broader scale by Siemens for mammography tubes, later for the rotating frame CT tube Straton®, and for angiography tubes. Philips enhanced the emitter size further and added a magnetic double quadrupole focusing system. Unlike Siemens' Straton® tube from 2003 with its single quadrupole system, see below, the double quadrupole concept of the Philips iMRC® tube from 2007, see the picture at the top of Fig. 5.3, allows for a high electron trajectory compression factor. Siemens adopted this technology also in the Vectron® tube for CT, launched in 2013. In 2018, GE presented a similar tube, the Quantix 160. It serves the premium tier GE Revolution® CT system, which allows for tube voltage switching for spectral X-ray imaging. Extending the basic design of Philips and later Siemens, it comprises a Pierce-type cathode; see the picture at the bottom of Fig. 5.3. An

Fig. 5.17 Digital focal spot camera complying with the IEC 60336 standard. X-rays from the focal spot of the X-ray tube pass an X-ray opaque slab through a 30 μm wide pinhole at a distance of min. 100 mm from the source and are detected by a pixelated spatial resolving detector. For a virgin tube, the image on the detector is a (typically threefold) magnified representative of the electron current density in the focal spot. Samples are shown in Fig. 5.16. Non-isotropic roughening of the focal spot track along the radial direction in aged rotating anode tubes may cause distortions

additional system of pull and focusing electrodes enables tube current modulation in a matter of microseconds by modulating the relatively high space charge in front of two tungsten flat emitters. Upon request prior to scanning, tube power and focal spot size are kept constant during modulation of the tube voltage between 80 and 140 kV. Siemens produces dual-source-dual-detector systems for this purpose. Instead, Philips employs detector-based spectral CT imaging, which is always active. Spectral material decomposition can be retrieved retrospectively, a feature that improves the clinical workflow.

A single quadrupole magnet would leave the cross-sectional area of the beam untouched. As realized in Siemens' Straton® tube, it converts a circular beam from the cathode in the center of the rotating frame tube into an elliptic which in combination with grazing impact of 45° to normal defines a set of well suitable focal spots for this tube. But, beam current density and minimal focal spot sizes are interrelated and restricted in an undesired manner. This disadvantage can be overcome with a

double quadrupole system. Even considering mechanical tolerances, this system is capable of producing unprecedented focal spot sizes down to the order of 0.3 mm width on the anode and a length of 3 mm, without compromising the beam current. As the X-rays are taken off in the central beam at a small angle of 8° from the anode, the large physical length of 3 mm translates to an X-ray optical length in the order of the width. The cathodic space charge limitation is practically gone, even at low tube voltages. Figure 5.18 is a schematic representation of the concept. First, the beam from the flat emitter is accelerated while the cross section hardly changes. Then, it is widened by the first magnetic quadrupole and finally compressed by the second. In addition, the figure illustrates the capability for the so-called z-deflection in focal spot length direction along the tube rotor axis, which is oriented parallel to the patient axis in a CT system. Also x- (width) deflection is possible by energizing extra dipole coils. Z-deflection helps reducing image artifacts, x-deflection enhancing the in-plane spatial resolution in a computed tomography system. The price paid is the higher effort for active control of the magnetic quadrupole focusing field.

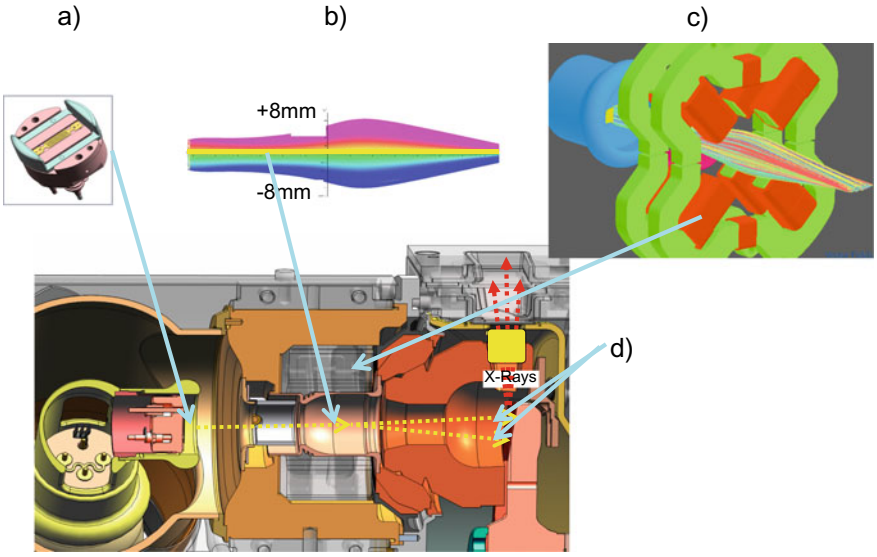


Fig. 5.18 Modern magnetic focusing as an alternative to the simpler electrostatic focusing concept as sketched, for example, in Fig. 5.9 and depicted in Fig. 5.12. Electrons (**d**) are released by a flat electron emitter (**a**), which is connected to the negative terminal of the high-voltage source. The electrons gain speed between the cathode and the funnel of a grounded drift tube, which is on ground potential, pass through the focusing quadrupole and a deflecting dipole magnet system and impinge on the anode. Electron trajectories are shown in (**b**). **c** The magnet system is depicted. As shown in (**d**), the electron beam can be deflected to render a deflection of the X-ray source area. Unlike electrostatic focusing and using coil emitters, space charge limitations are greatly reduced. Adapted from Behling [1], Fig. 6.36

5.5.3 *Electrostatic Beam Compression and Current Switching*

The starting conditions of electrons depend on the point of emission. In particular at coil emitters, the electric field is strongly angulated at large portions of the surface. Electrons starting off normal with a fraction of an electron volt of thermal energy gain up to several tens of electron volts kinetic energy in the lateral component of their velocity before they turn to directions close to normal to the cathode. The emittance of the electron beam is low. Under these circumstances, small focal spots can only be designed by sacrificing emitter size and emitting area, i.e., the capacity to generate large tube currents. In turn, the necessary enhancement of the emitter temperature results in short tube life. Given a small emitting surface, the emitter current density must be elevated through the application of enhanced temperatures, limited by space charge effects.

A cathode of the type shown in Fig. 5.19 helps narrowing focal spots. A single tungsten emitter coil is surrounded by isolated control electrodes to the left and right. Negative bias voltage with respect to the emitter narrows the emitting area under a pulling electric field such that the lateral components of the initial velocity of emitted electrons are minimized. Side rays are suppressed. Tabs in the length direction of the tungsten coil shorten the focal spot and homogenize the electron intensity distribution on the anode. High bias voltages of about 3 kV and higher allow cutting off the electron emission altogether. This type of switching is used to minimize unwanted X-ray emission when pulsed X-ray emission of millisecond duration is required, e.g., for angiography. The electric field at the emitter is repelling thermally released electrons on the entire surface. This mode is called grid switching in remembrance of legacy radio tubes.

Additionally, the separated electrodes left and right of the emitter further allow for deflecting the electron beam, as shown in Fig. 5.20. This results in the deflection of the focal spot on the anode, the origin of the X-radiation, in the width direction of the emitter, as discussed before. A benefit of this concept is the small number of vacuum feedthroughs needed. Only one emitter has to be connected, while more focal spot sizes can be generated. A disadvantage is the lack of focusing capacity in the length direction. Due to the rectangular shape of the emitter, electrostatic biasing using tabs at the ends of the tungsten coils would cause an undesired inhomogeneity of the electron current density in the focal spot and enhance the risk of overheating the edges.

5.6 Overall Charge Balance in a Medical X-Ray Tube

A discussion of the total charge balance inside an X-ray tube may illustrate the different physical effects which contribute to the behavior of an X-ray tube in operation and the conditions under which the cathode operates. An exemplary diagram

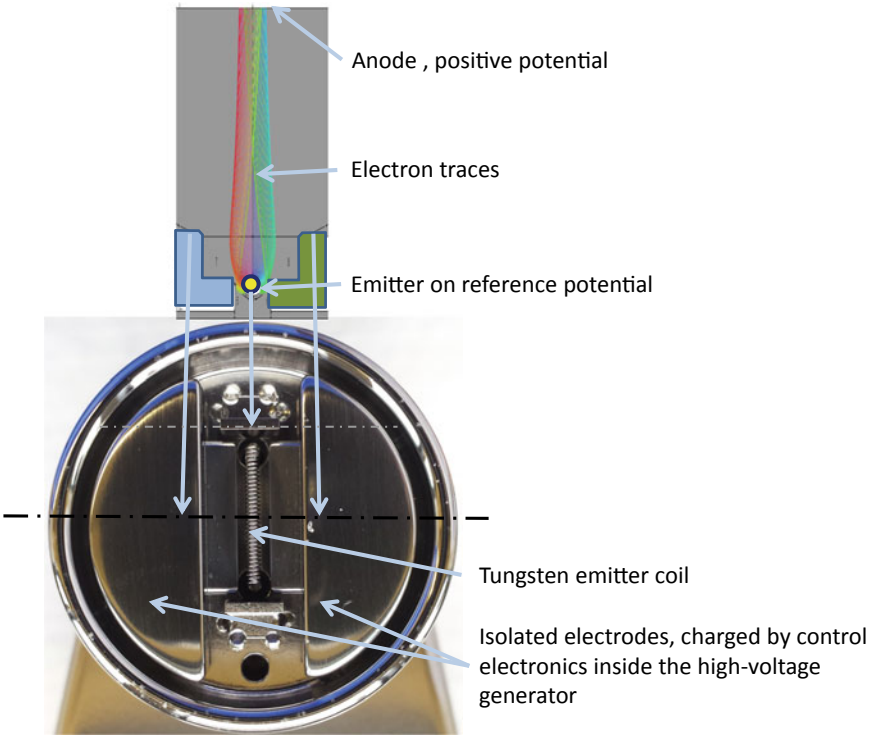


Fig. 5.19 Cathode with electrostatic focusing and deflection electrodes. Bias voltage supplied to control electrodes left and right of the tungsten emitter coil enables electronic control of the focal spot width. The bias of several hundred volts is negative with respect to the electron emitter and has to be adjusted depending on the tube voltage to keep the focal spot size stable for all desired X-ray spectra. The area of electron emission, where a pulling electric field exists, is controlled electronically. Asymmetric bias causes deflection to the right or the left, as shown in Fig. 5.20 Adapted from Behling [1], Fig. 6.29

of a sample metal center section tube from Philips is shown in Fig. 5.21. It lists all major current paths.

The electron emitter coil is directly heated with AC current I_{fil} supplied by a heating transformer, which provides high-voltage insulation and generates the secondary heating voltage U_{fil}. The DC tube current I_t is wired in a symmetric manner to the center terminal of the secondary winding of the heating transformer. The tube current splits and is fed in a symmetrical manner into the emitter. The X-ray flux is basically proportional to the tube current. But, great care has to be taken to include corrections for various effects. The current $I_{t,\,real,}$ which describes the flux of electrons that impinge on the anode, can hardly be measured non-invasively. A good approximation is I_t, measured remotely in the cathode branch of the high-voltage supply of the terminals for the tube voltage V_t. When high-voltage generators are wired to the tube through long cables, high-frequency charging currents may add to

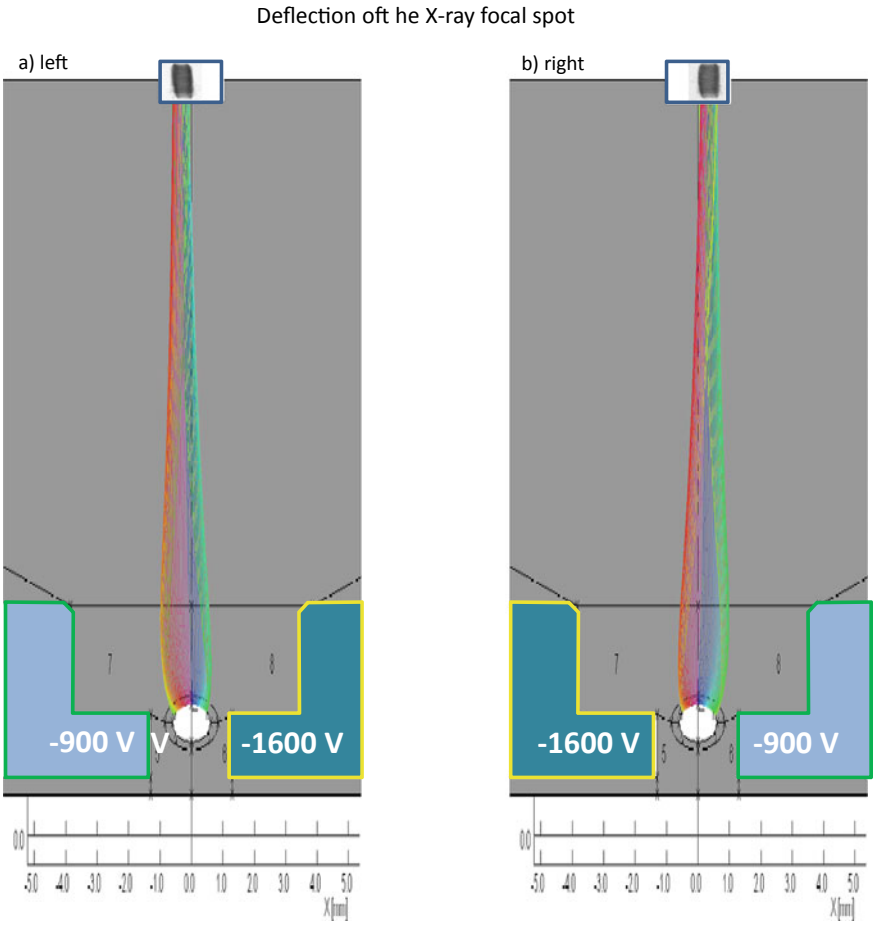


Fig. 5.20 Ray tracing for the cathode shown in Fig. 5.19. Asymmetric negative bias voltage at the control electrodes shifts the focal spot to the left or to the right. Two projections in computed tomography will be taken with the X-ray focal spot slightly shifted. This helps in reducing image artifacts and enhance the spatial resolution of the rendered image. Adapted from Behling [1], Fig. 6.31

the tube current and impede precise measurements. Inside the vacuum tube, electronic field emission from the cathode $I_{FE,\,c}$, photoelectric electron flow $I_{photo,\,c}$, currents through the high-voltage insulation from cathode to ground $I_{insul,\,c}$ and $I_{ion,\,c}$ have to be subtracted. This is indicated in the figure. Typically, field emission currents $I_{FE,\,c}$ in commercial medical X-ray tubes amount to between 1 μA and 100 μA for tube voltages of 150 kV. Photoelectric current $I_{photo,\,c}$, current though the insulating materials glass or ceramics $I_{insul,\,c}$, and ion current $I_{ion,\,c}$ to the cathode scale in nanoamperes and are usually only relevant for the measurement of ion charging. However, in the extreme case of excessive gas or vapor pressure reaching beyond

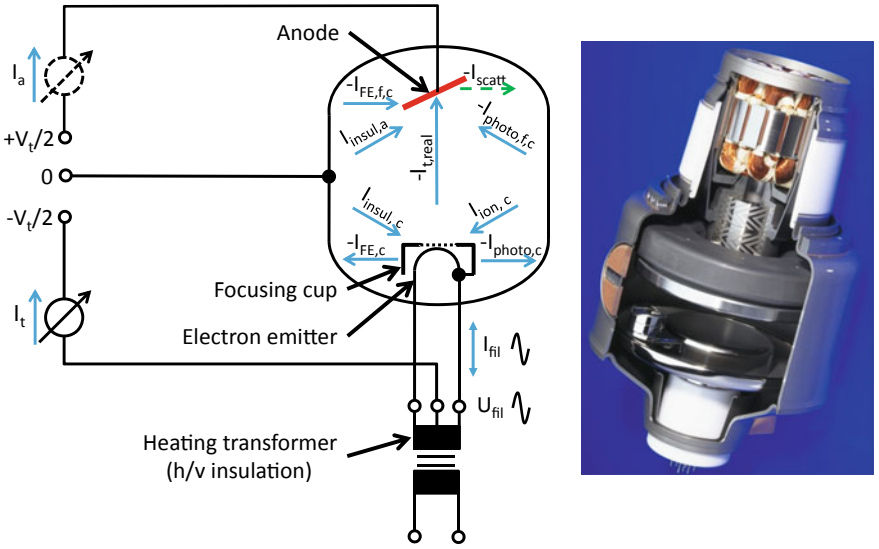


Fig. 5.21 Charge carriers in an operating X-ray tube. Depicted is the direction of electron and ion flow in an exemplary bipolar rotating anode metal center-section X-ray tube shown in the picture at the right. The electron emitter is directly heated with alternating current I_{fil} and heating voltage U_{fil}, see Fig. 5.8. The DC tube current I_t is fed symmetrically to the center of the secondary winding of the heating transformer in the high-voltage generator. The current $I_{t,\,real}$, is proportional to the X-ray flux. But, it can hardly be measured non-invasively. A reasonable estimate is I_t, sensed in the cathode branch of the high-voltage supply from the terminals for tube voltage V_t supply. Electronic field emission from the cathode $I_{FE,\,c}$, photoelectric electron flow $I_{photo,\,c}$, currents through the high-voltage insulation from cathode to ground $I_{insul,\,c}$ and $I_{ion,\,c}$ have to be subtracted. Typically, the field emission current $I_{FE,\,c}$ amounts to between 1 μA and 100 μA for tube voltages of 150 kV. Photoelectric current $I_{photo,\,c}$, current thought the insulating materials $I_{insul,\,c}$, and ion current $I_{ion,\,c}$ to the cathode amount to nanoamperes only. However, in exceptional events of overheating and excessive gas or vapor pressure beyond ca. 10^{-1} Pa in the primary or scattered electron path, e.g., caused by anode rotation failure, ion currents may reach the same order of magnitude as I_t. $I_{ion,\,c}$ comprises the charge of ions which impact the cathode as well as scattered electrons which they release. Only for glass tubes where all electrons are inevitably collected by the anode is the anode current I_a a proper approximate measure for the X-ray flux. But, in metal center-section tubes, like the one in the figure, scattered electrons I_{scatt} contribute significantly (10–12% of It for bipolar tubes). I_{scatt} vanishes for glass tubes. However, it may amount to up to 50% for anode grounded tubes with cathodes mounted distant from the anode. Also, the anode experiences a photoelectric current from the tube frame and from the cathode $I_{photo,\,f,\,c}$, an insulator leakage current $I_{insul,\,a}$ and receives field emission from the tube frame and the cathode $I_{FE,\,f,\,c}$ Adapted from Behling [1], Fig. 6.42

ca. 10^{-1} Pa ion currents may grow to of the same order of magnitude as I_t in the primary or scattered electron path. $I_{ion,\,c}$ comprises the charge of ions that impact the cathode as well as scattered electrons which they release upon impact. Only for conventional glass X-ray tubes is the anode current I_a a proper approximate measure for the X-ray flux. All electrons are inevitably collected by the anode. But, in metal center section tubes, like the one in the figure, scattered electrons I_{scatt} contribute a

significant amount of 10–12% of I_t for bipolar tubes. Backscattered electrons with an energy below about half of $e \cdot V_t$ are mirrored back from the tube frame and return to the anode. Only the fastest electrons escape from the anode and make their way to the grounded metal wall. While, as aforementioned, I_{scatt} is zero for glass tubes, it may amount to up to 50% for anode-grounded tubes which have their cathodes far away from the anode and thus, without electron mirroring effects. Like the cathode, also the anode experiences photoelectric currents from the tube frame and from the cathode $I_{photo,\,f,\,c}$, an insulator leakage current $I_{insul,\,a}$ and receives field emission from the tube frame and the cathode $I_{FE,\,f,\,c}$.

5.7 Alternatives to Tungsten Emitters

5.7.1 Reduction of the Work Function

Several concepts have been evaluated to reduce the emitter work function in X-ray tubes and the operating temperature of the cathode, its energy consumption, and to extend its service life. One of them has been the use of thoriated emitters. A monolayer of thorium reduces the work function to about 2.6 eV. Thorium diffusion from the interior along grain and dislocation boundaries replenishes loss from evaporation and sputtering effects. Typically, the tungsten matrix is carbonized to control the diffusion speed. Unfortunately, all thorium isotopes are radioactive. At least one of them (232 Th) is least harmful with a half-life of 14 billion years. But, it is still a radio-isotope with a negative ecological impact.

Reduction of the work function down to 1.16 eV and very high current densities of up to 400 Acm^{-2} in saturation mode have been reported for dispenser cathodes. These cathodes have been tried in the rough practice of X-ray tubes, see Gaertner et al. [11–13]. The surface of dispenser cathodes is covered with, e.g., barium and barium oxide in a complex with tungsten, scandium, and oxygen, see Chaps. 2 and 3. A broad variety of dispenser cathodes of various kinds and additives have been tried for medical X-ray tubes, e.g., for use in mammography, where small focal spots with high current densities are necessary. Unfortunately, and despite emitter lifetimes of 10^4 h achieved in ultra-high vacua of 10^{-7} Pa, the adverse residual gas atmosphere, ion bombardment, and chemical reactivity prevented this technology from being reliable enough in pinched-off X-ray tubes, see Kimura et al. [14]. On top of this, indirect heating with characteristic heating times of several seconds, and the large dependence of the average work function on residual gas and ion conditions do not allow these cathodes to be operated in saturation emission. Their current has always to be space charge controlled, e.g., by Wehnelt electrodes. The construction of rugged electrodes of this type is sophisticated in view of the currents and focal spot sizes used for medical imaging. A comprehensive overview of the history of electron emission for vacuum electronics is provided by Gaertner [15].

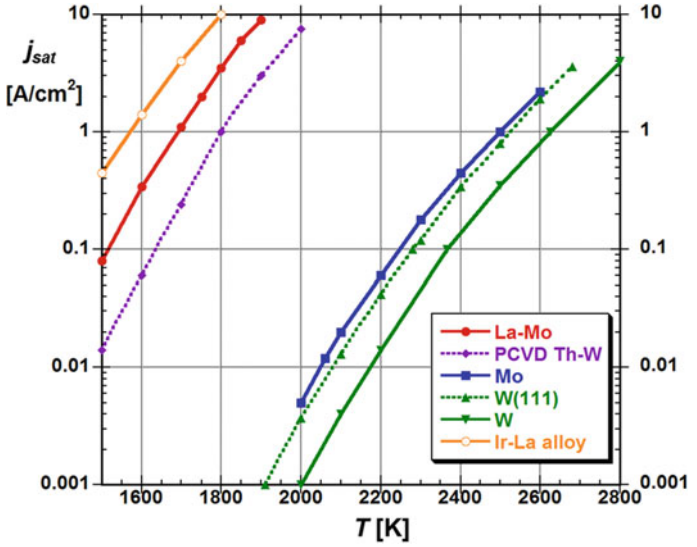


Fig. 5.22 Emission current density of metal, metal alloy, and Th and La dispenser cathodes as a function of true temperature; data are based on Gaertner and Koops [16] and on Djubua et al. [17, 18] and Weissman [19, 20]. Copyright Georg Gaertner, Aachen, Germany, 2020

Figure 5.22 shows a comparison of emission current densities over emitter temperature for a number of emitter materials and crystal orientations, which spans over five orders of magnitude. Data for this figure is from different sources, most of them can be found in the comprehensive text of Gaertner and Koops [16]. Data on Ir-La alloy are from Djubua [17, 18]; La-Mo: Buxbaum et al. [21]; PCVD Th-W: Gaertner [16], Gaertner [22]. The W(111) surface discussed was produced by I. Weissman using chemical vapor deposition (CVD) [19, 20]. Metal alloy cathodes, e.g., Ir_2La and Ir_5Ce promise advantage over monolayer dispenser cathodes such as La-Mo (carburized Mo doped with La-oxide) and Th-W (carburized tungsten doped with ThO_2), since they are pure all-metal cathodes capable of high emission. They may be less sensitive to poisoning than Th-W or La-Mo, and can also be operated in an elevated temperature range, which helps improving robustness against poisoning. Monolayers of La (Ir_2La) or Ce (Ir_5Ce) may be beneficial in this context, too.

5.7.2 Carbon Nanotube (CNT) and Graphene Emitters

An electron-emitting substance of promising robustness is carbon nanotube (CNT) or graphene field emitters. Clinical prototype tubes have been built, see, e.g., Gidcumb et al. [23], for electronic tomosynthesis application. This technique requires switching capabilities and a multitude of inexpensive emitting surfaces. Electrons are field-emitted from bunches of single-walled or multiwalled CNTs, which allow for

pulling currents of a few microamperes each. The current densities nearly follow the Fowler–Nordheim equation and are about proportional to $(I/V_t)^2 \cdot \exp(\text{-const}/V_t)$, V_t being the tube voltage. In one of the embodiments, each sub-cathode in a cathode array consists of CNT-coated electrode areas of a dimension of 2.5 mm by 13.0 mm which produces a projected focal spot size on the related anode targets of 0.6 mm by 0.6 mm (FWHM). For each of the 31 emitting surfaces, two focusing electrodes and a pull grid each are used. The high mechanical strength of CNTs, high electrical and thermal conductivity, and relatively low sputter rates promise these field emitting devices to be candidates for X-ray tubes as well as for other devices. The macroscopic electric field is enhanced at the apex of a CNT by about the ratio between length and diameter of such a thread, which typically amounts to several hundred. Thus, pull fields of sufficient strength can be generated. CNTs may be grown by microwave plasma-enhanced chemical vapor deposition (CVD) and thermal CVD. Xintek Inc., NC, USA is producing and commercially offering CNT cathodes on the basis of CVD-generated nanotubes which are, after deposition on the cathode surface and bake-out, embedded in a matrix of binder material. Cathode current densities achieved in for prototype tubes amount to about 0.1 A/cm^2 for long-term pulsed operation as necessary for tomosynthesis imaging. Individual sub-cathodes are set into operation by switching a pull field on, which is applied between the emitting surface and wire grids structures in front. In one embodiment, these grids are 60% transparent for electrons. With a proper setting of the pull voltage of about 1 kV, each sub-target of the anode is subject to a net tube current of about 25 mA. The gross (total) emission per sub-cathode is about 40 mA. The grid is, therefore, heating up with a power of between 15 and 20 W. Pulse lengths in tomosynthesis for mammography vary between 125 and 250 ms per X-ray emitter. The pull voltage has to grow over time to compensate for the deterioration of the emitting nanotubes. A basic vacuum pressure of 10^{-6} Pa inside the tube is recommended, in peaks maximal 10^{-4} Pa, which is achieved by additional ion getter pumps. Although common for stationary anode tubes, this required maximum gas pressure is several orders of magnitude below the typical maxima in rotating anode tubes and make the CNT technology hardly viable for this type of X-ray tubes. Unlike Xintek, Philips and other groups have grown CNTs from microwave plasmas directly on catalyzer surfaces. Despite decades of intense investigation, CNTs have not been commercialized on a broad scale for medical imaging. The list of challenges comprises sublimation of carbon in overloaded CNTs, field-elongation, rupture at imperfections, high thermal and electrical resistance, and overheating at foot points [24]. Field emission is rather sensitive to the atomic microstructure at the emitting apex. Ion attachment tends to de-stabilize nanotubes. Sputter effects and chemical reactions with oxidizing residual gas reduce the heated fibers. Design challenges come on top. According to the field emission principle and without additional measures, tube voltage and tube current are directly coupled like for ancient ion tubes. Thus, and as aforementioned, the field emission current has to be controlled by an extra pull grid. Medical imaging operates with rather large focal spots of about 0.5 mm up to 1.5 mm width and about 10 mm physical length. In order to obtain such relatively large focal spots and hundreds of milliamperes of current, the macroscopic emitting area should amount

to between 20 and 500 mm^2. The electric pull field on the surface of such an emitter must be homogeneous to balance the current load among nanotubes in the center and at the periphery. This requirement, in turn, asks for "macroscopic" millimeter-sized distances between electrodes and requires application of pull voltages in the order of kilovolts, although a single CNT filament could be switched with 30 volts only. Consequently, CNTs reveal their benefits for microfocus-tubes with stationary targets, preferably intended to operate in a small range of tube voltages.

5.8 Conclusion

So far, electron release by field emission has been restricted to scientific and niche applications in medical imaging, notably for stationary anode tubes with low tube currents. Tungsten thermionic emitters continue to be the backbone of cathodes for diagnostic rotating anode X-ray tubes. The use of Ba dispenser cathodes with low work function requires a relatively good vacuum, which is hard and costly to maintain for these tubes. As Ba-based emitters operate space charge limited, electron emission must be controlled by additional sophisticated electronics, compared with the case of tungsten emitters, which basically operate in saturation emission. Tungsten thermionic emitters continue to be the backbone of cathodes for diagnostic X-ray tubes. They are robust against ion bombardment, vacuum discharges, and poisoning. Significant progress has been achieved with flat sheet tungsten emitters. This comes to the expense of higher costs and problems to maintain thermomechanical integrity. The temperature required for tungsten may be reduced by a few dozen Kelvin employing flat polycrystalline tungsten emitters with preferred crystal orientation such as W(111) for lowering the work function. I. Weissman describes such concepts and prepared tungsten layers by CVD. Other options are pure metals such as Mo with work function lower than tungsten. Surface monolayer type cathodes such as thoriated tungsten or lanthanated Mo suffer from the risk of chemical poisoning under harsh residual gas conditions and the need to replenish the active monolayer after ion sputtering and evaporation. Future research may focus on metal alloy cathodes as investigated by B. Djubua et al. Unfortunately, the literature on these is scarce. It is expected that their robustness against poisoning improves with emitter temperature and the emitter service life may be acceptable.

References

1. R. Behling, *Modern Diagnostic X-ray Sources: Technology, Manufacturing, Reliability* (CRC Press, Taylor & Francis, Boca Raton, USA, 2016)
2. G.N. Fursey, Field emission in vacuum micro-electronics. Appl. Surf. Sci. **215**, 113 (2003)
3. G. Dörfel, Julius Edgar Lilienfeld und William David Coolidge–Ihre Röntgenröhren und ihre Konflikte (Julius Edgar Lilienfeld and William David Coolidge–Their Roentgen tubes and their

conflicts), Preprint 315, Max Planck Institute for the History of Science (2006), www.mpiwg-berlin.mpg.de/Preprints/P315.PDF. Accessed 20 Nov2013
4. W.D. Coolidge, A powerful Röntgen ray tube with a pure electron discharge. Phys. Rev. **2**(6), 409–430 (1913)
5. J.E. Lilienfeld, A reply to Mr. W.D. Coolidge's paper A powerful Röntgen ray tube with a pure electron discharge. Phys. Rev. **3**, 366 (1914)
6. E. Shefer et al., State of the art of CT detectors and sources: a literature review, Curr. Radiol. Rep. 76–91, (2013). (Published online, Feb. 2013, Springer Science + Business Media, New York 2013)
7. D.I. Karyugin, V.N. Stolyarov, I.N. Stolyarov, A method for evaluation of X-ray tube emitter service life. Biomed. Eng. **47**, 247 (2014)
8. K. Skotnicová, J. Drapalá, V. Kolařik, Study of emission properties of tungsten single crystal with crystallographic orientation, in *Conference proceedings METAL in Rožnov pod Radhoštěm (Czech, Republic)*. Ostrava, Czech, Republic, (Tanger Ltd., 2010), pp. 782–785, http://www.metal2014.com/files/proceedings/metal_10/lists/papers/59.pdf. Accessed June 2014. ISBN 978-80-87294-17-8
9. E. V. Barmina, E. V., A. A. Serkov, E. Stratakis, C. Fotakis, V. N. Stolyarov, I. N. Stolyarov, G. A. Shafeev, Nano-textured W shows improvement of thermionic emission properties. Appl. Phys. **A 106**, 1 (2012)
10. R. Behling et al., High current X-ray source technology for medical imaging, in *International Vacuum Electronics Conference IVEC 2010, IEEE International 2010*, pp. 475–476, https://doi.org/10.1109/ivelec.2010.5503
11. G. Gaertner et al., Low temperature properties of Ba-dispenser cathodes. J. Vac. Sci. Technol. B Microelectron. Nanometer Struct. **18**, 997 (2000)
12. G. Gaertner, D. Barratt, Life-limiting mechanisms in Ba-oxide, Ba-dispenser and Ba-Scandate cathodes. Appl. Surf. Sci. **251**, 73–79 (2005)
13. 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)
14. S. Kimura, T. Higuchi, Y. Ouchi, E. Uda, O. Nakamura, T. Sudo, K. Koyama, Emission characteristics of dispenser cathodes with a fine-grained tungsten top layer. Appl. Surf. Sci. **111**, 60 (1997)
15. G. Gaertner, Historical development and future trends of vacuum electronics. J. Vac. Sci. Technol. B **30**, 60801 (2012)
16. G. Gaertner, H. Koops, Vacuum electron sources and their materials and technologies, in *Vacuum Electronics, Components and Devices* (chapter 10), ed. by J. Eichmeier, M. Thumm, (Springer, Berlin, 2008)
17. B. Djubua, Research and development of dispenser, metal alloy and oxide cathodes, in *Conference Record 1994 TRI/NASA Cathode Workshop, Cleveland/Ohio*, p. 3–6
18. B. Djubua, O. Kultashev, A. Makarov et al., Metal alloy cathodes for application in vacuum microwave devices, in *IVEC/IVESC 2012, Paper 125*, p. 1–2
19. I. Weissman, Improved thermionic emitter: thorium on grain-oriented polycrystalline tungsten. J. Appl. Phys. **36**, 406 (1965)
20. I. Weissman, M. Kinter, Improved thermionic emitter using uniaxially oriented tungsten. J. Appl. Phys. **34**, 3187 (1963)
21. C. Buxbaum, G. Gessinger, Der Einfluss von Platinmetallen auf das Emissionsverhalten von Lanthankathoden. High Temp. High Press. **10**, 325–328 (1978)
22. G. Gaertner, P. Janiel, H. Lydtin, Plasma-activated CVD of tungsten/ thorium structures, in *Proceedings of the 6th European Conference on CVD*, ed. by R. Porat (Jerusalem, 1987), pp. 319– 327
23. E. Gidcumb, B. Gao, J. Shan, C. Inscoe, J. Lu, O. Zhou, Carbon nanotube electron field emitters for X-ray imaging of human breast cancer. Nanotechnology **25**, 245704 (2014)
24. R. Behling, F. Grüner, Diagnostic X-ray sources—present and future. Nucl. Instrum. Methods Phys. Res. Sect A: Accelerators, Spectrometers, Detectors and Associated Equipment **878**, 50–57 (2018). https://doi.org/10.1016/j.nima.2017.05.034

Chapter 6
Cathodes for Electron Microscopy and Lithography

Pieter Kruit

Abstract Electron sources for high-resolution applications such as microscopy and lithography are characterized by their brightness and energy spread rather than by their total emission current. In fact, most of the emitted current is cut out by apertures and only nanoamps or at most microamps actually reach the sample or wafer. This chapter reviews the physics and practical properties of the electron sources that are typically used in microscopy and lithography. It is an extension of topics addressed in [1] with special emphasis on application aspects and their emitter requirements. Recent developments are reviewed. The relations between the source properties and the resolution and throughput of these instruments is derived and discussed.

6.1 Introduction

The performance of electron microscopes and electron beam lithography machines depends critically on the quality of the electron source used in the instrument. However, there exist grave misunderstandings about the exact influence of the electron source on the performance of microscopic imaging or nanolithography, so let me start by making a strong statement: In the first approximation, the theoretical resolution of microscopes and lithography machines does not depend on the properties of the electron source. The only role of the electron source is to provide sufficient current to conquer the inherent statistics in both imaging and patterning. We will later see that there is a subtle interplay between *practical* resolution and current in the beam and that the properties of the source are important in that equation.

Electron microscopes come in different varieties [2, 3]. The most basic machine is a scanning electron microscope (SEM) in which the electron source is imaged with some demagnification onto the specimen, while deflection coils or plates scan the beam over the specimen surface and a detector detects signal. The image shows the intensity of the detected signal (often the secondary electrons) as a function of the position of the electron beam. The resolution is given by the probe size. The

P. Kruit (✉)
Delft University of Technology, Delft, The Netherlands
e-mail: P.Kruit@tudelft.nl

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*,
Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_6

noise in the image is given by the current in the beam, the efficiency of the signal channel, and the scan speed of the beam. The desire to increase the imaging speed, both for biological applications and in the inspection of semiconductor wafers, has recently led to the development of SEMs in which multiple beams scan the sample simultaneously. Of course, this sets new demands on the cathode.

The scanning transmission electron microscope (STEM) employs the same imaging principle as the SEM, however, in this machine, the electron energy is usually higher and the specimen is thinner so that it is possible to detect the transmitted electrons and use either the scattering intensity in a particular direction or at a particular energy loss as the signal for the image. In transmission electron microscopy (TEM), a thin specimen is illuminated by an electron beam that is spread out over the specimen. Post specimen lenses are now used to create a magnified image of the specimen on a screen or camera. The resolution is given by a combination of lens aberrations and diffraction effects, while the noise is determined by the current in the beam, the quality of the image sensor, and the image acquisition time. There is a subtle role for the illumination angle in the beam, both on the effect of the scattering in the sample and on the effect of the aberrations. The requirement for this angle is usually expressed in the coherence of the beam and will be discussed further on in this chapter.

Electron beam lithography (EBL) machines also come in many varieties. In nanotechnology labs, it can still be seen that they are a descendant of the SEM. Some actually are modified SEMs. The modification is that the electron beam can write predefined patterns and the specimen consists of a material covered with an electron sensitive material, a resist. More sophisticated machines have special lenses and deflectors to allow for a larger deflection field and have specimen stages that can hold large semiconductor wafers. For so-called Gaussian beam writers, the resolution is determined by the probe size and the speed of writing by the electron current and the sensitivity of the resist. Although at first, it may seem that the writing speed can be increased by using an ever more sensitive resist, this is a mistake. Each pixel in a pattern must be defined by a sufficient number of electrons in order to create a noise-free pattern. Thus, for high-resolution nanopatterns, one requires a much larger dose (electron current density times exposure time) than for lower resolution patterns. The latter statement is valid for all kinds of EBL machines. In the semiconductor industry, the most sophisticated masks are written by EBL, usually by variable-shaped beam machines. In these machines, the electron source is not imaged onto the target, but the beam is used to illuminate a shaped aperture, which in turn is then imaged onto the target. The size and shape of the aperture can be varied, for instance, by using two apertures behind each other and a deflector in between so that the combined shadow image can take different shapes. The ultimate resolution of these machines is determined by the lens aberrations and the writing speed by the total current through the shaped aperture, which can be substantially larger than in a Gaussian beam machine of the same resolution. The principle of shaped beam lithography can be taken a step further by using a transmission electron mask with the whole pattern to be printed already in it. Several attempts have been made to realize such a machine, but thus far without success. The main problem is that to be useful in the semiconductor

industry, such a machine needs a large current which necessarily passes a crossover in between the mask and the target. In this crossover, the Coulomb interactions cause a blur in the image. We shall show later that this effect depends on the properties of the electron source. Finally, there are designs for multibeam lithography, where many Gaussian beams are scanned over the target simultaneously. In some designs, these beams come from separate sources, in other designs from a single source.

Since most modes of operation in microscopy and lithography use a focused electron probe, a separate section will look in detail at the relation between the source properties and the focused beam properties. After that, several electron source types that are used in electron microscopy and EBL will be discussed in detail: the thermionic cathode, the Schottky source, the cold field emitter, and the photocathode. Examples of recent developments and applications will be given there.

6.2 Source Parameters for High-Resolution Applications

In the field of electron microscopy and lithography, one usually speaks of the "electron source", or "electron gun", which contains an "emitter". The word "cathode" is less frequently used. The source unit can contain apertures, a gun lens, and other components. It is good to be aware that the electron beam that comes out of the source can have different properties than the beam that was emitted from the cathode. The easiest change is in the current, which can be limited by an aperture inside the source. But even the brightness, which, as we shall see, is a constant when manipulating the beam by optics, can change between the cathode surface and the exit of the source through stochastic interactions between individual electrons.

Just as a reminder, let's review the elements that a source may contain. The central component in any electron source is, of course, the emitter or cathode: the tiny piece of material that actually emits the electrons. The emitter is often at elevated temperature, which is accomplished by a filament that is heated by a filament current. Close to the emitter is the extractor or anode, which is at a positive potential with respect to the cathode in order to accelerate the electrons into a beam. The extractor, of course, has a hole, an aperture, to transmit the beam, or part of the beam. The thermionic cathode and the Schottky source operate at quite high temperatures, resulting in electron emission from a much larger surface than is required. An electrode between the cathode and the anode, carrying a negative potential, pushes most of the unwanted electrons back into the cathode. This electrode is called a Wehnelt electrode in the thermionic source and a suppressor electrode in the Schottky source. In order to form a parallel beam or to create a crossover in the accelerator of the microscope, the extractor is followed by a gun lens. The gun lens is usually electrostatic since the electrons have a relatively low acceleration energy of only a few kilo-electron-volts in this section. The gun section is completed by gun alignment coils or electrodes which are used to align the beam with the optic axis of the system (shift alignment) and to select the emission direction that has the best quality beam (tilt alignment).

6.2.1 Brightness

Election sources supply the electrons to the system, so the main property of a source seems to be the number of electrons per second, or current, it supplies. However, we need to distinguish between the total current that the emitter emits, the beam current that passes the extractor aperture, and the probe current or specimen current that arrives at the sample through the beam-limiting apertures of the microscope.

After the initial acceleration, the beam seems to come from an area with a diameter that we call virtual source size or just source diameter d_g. It is called virtual source because it is different from the surface that is actually emitting electrons. This is best explained by considering a cathode with a spherical shape: all electrons that come out of the surface perpendicular to the surface seem to come from a single point at the center of the sphere. The finite size of the virtual source is caused by electrons emitted with a slight lateral velocity (in thermal emission a result of the thermal energy kT); they seem to come from a circle around the center of the sphere. The electric field of the acceleration may have a lens effect that magnifies and shifts the virtual source along the optical axis. The convention is to define the virtual source size looking back from the plane of the extractor. Thus, the virtual source size may depend on the extractor voltage. For example, a Schottky emitter with a tip diameter of 1μm in a typical setup with an extractor at 5 kV has a virtual source size of about 50 nm. For hairpin sources, the combined action of acceleration and Wehnelt repulsion creates such a strong lens effect that the beam has a first crossover already inside the source. The virtual source size is then about equal to the size of this crossover, in the order of 20–50μm.

The virtual source size is usually small compared to the size of the extractor aperture, so we can speak of "the" emission angle of the beam and thus of the angular current density I', defined as the current divided by the emission angle, or better: the beam current divided by the angle that is accepted by the extraction aperture. We encounter many more current densities, such as the emission current density J: the emitted current per unit area at the emitter surface.

In the general literature on electron emitters, one often encounters the parameter emittance, defined as the product of the source size and the emission angle. If there are no beam-limiting apertures and no acceleration, this is a useful parameter, because it stays constant from one crossover to the next: if the crossover size is magnified, the beam angle is demagnified, thus the emittance is constant. However, in electron microscopes, there are many apertures and there is acceleration, so the emittance is not a constant.

There is a parameter that does not change at a beam-limiting aperture: the differential brightness of a beam. Differential Brightness of an electron beam is defined as [4]:

$$B_{\text{diff}} = \frac{\mathrm{d}I}{\mathrm{d}A\mathrm{d}\Omega} \tag{6.1}$$

where dI is the current passing through a surface dA within a solid angle dΩ. Clearly, when an aperture is placed in the beam, either the area, or the angle is reduced, but the current is reduced proportionally, so the brightness is constant. Acceleration or deceleration, however, does not keep the brightness constant: usually the beam angle changes, it reduces at acceleration and increases at deceleration. By defining a reduced brightness of the beam as

$$B_{\mathrm{r}} = \frac{\mathrm{d}I}{\mathrm{d}A\mathrm{d}\Omega \cdot V} \tag{6.2}$$

that problem is solved. Reduced brightness is expressed in A/m^2 sr V. Because this beam parameter is constant in the whole system, we can also use this as a parameter of the source. In fact, for microscopic applications of electron sources, it is the single most important parameter of a source. We'll see later that there are very few physical phenomena that can change the beam brightness: stochastic Coulomb interactions between the individual electrons is one, energy filtering is another. The concept of brightness is easiest understood in the plane of a source image, where the area is the size of the source image and the angle is the aperture angle. However, one needs to be careful here: the definition above, for the differential brightness, intrinsically assumes a homogeneous current density in the area dA. In a source image, there is a current distribution with a maximum density in the center and often long tails. Although this is not a formal definition, one then often uses the concept of practical brightness $B = I/\mathrm{A}\Omega$, which makes it necessary to define the size of the source image. We shall show later that the best definition is the full width that contains 50% of the current: FW50. Practical brightness is then slightly larger than differential brightness [5]:

$$B_{\mathrm{pract.}} = \frac{I}{\frac{\pi}{4}\mathrm{d}_{\mathrm{FW50}}^2 \pi \alpha^2} = 1.44 B_{\mathrm{diff.}} \tag{6.3}$$

Although brightness is easiest understood in a source image, it is defined at any plane in the beam. For example, in an aperture, area A is the size of the aperture and the angle Ω is the angle toward the nearest source image. In the plane of the sample in a TEM, brightness is the current density divided by the internal angle in the beam, that is, the angle toward the nearest source image.

Brightness in an electron microscope can best be measured in spot mode, at relatively large spots. It is important to find a mode in which the size of the probe is dominated by the size of the source image, so at relatively small aperture angle to avoid aberration contributions to the probe size and at relatively large beam currents to avoid a diffraction contribution to the spot size. In an SEM, one measures the probe size at the sample and one needs to know the aperture angle corresponding to the choice of aperture. In TEM, the spot at the sample plane is magnified with a known magnification to the CCD camera. The angle at the sample can be measured in diffraction mode and the current is measured in a Faraday cup, or on the microscope screen.

Implicitly, we have assumed so far that it is obvious why a higher brightness in a beam is always advantageous. Perhaps it is good to motivate this explicitly. The current in a probe of a certain size is proportional to the brightness. Every application needs a minimum number of electrons. At a larger current, this number can be obtained in a shorter time. In electron microscope images, too few electrons lead to noisy images: the signal-to-noise in each pixel is always smaller than the square root of the number of electrons detected in that pixel. In electron beam lithography, the sharpness of the features is directly related to the electron dose [6], so the throughput is directly proportional to the total current in the system.

6.2.2 Energy Spread

The last parameter that is used to describe an electron source is the energy spread ΔE in the beam. Electrons can be emitted with slightly different energies. Whatever the acceleration is afterward, the energy conservation law tells us that this original energy difference will stay. In fact, there is a whole distribution of different energies, the energy distribution. Usually, the distribution is asymmetric and it is a little difficult to characterize it with a single number. Some use full width at half maximum FWHM, others use full width that contains 50% of the current FW50. Even for a nice distribution such as a Gaussian, the FW50 is different from the FWHM, so one has to be careful what to use in calculations of, for instance, the chromatic aberration. The energy spread can only be increased by stochastic interactions between individual electrons, where a collision may accelerate one electron and decelerate the other. That effect is called the Boersch effect. In practice, we also add the variations in accelerator voltage to the energy spread, because they have the same effect when averaged over time, but strictly this is not a parameter of the electron source. The energy spread may be reduced by an electron monochromator also called energy filter. In the filtering process, electrons are stopped, so the act of filtering decreases the beam current and also the brightness. What is constant is the brightness per electron volt energy spread. So when a monochromator is present, the single parameter that we need to know about the source is B_r/ΔE.

6.2.3 Operational Aspects

From an optics point of view, the best electron source is the one with the highest brightness and the lowest energy spread. However, from a practical point of view, we need to look at other parameters as well. Current stability is important, especially in scanning applications. Life time is of interest. This is the average time an electron source can operate without replacement of the cathode. Hairpin cathodes, for instance, can be made brighter by operating at higher temperatures. However, this

costs life time. Environment requirements may differ very much from one source type to another: cold field emitters require extremely good vacuum of the order of 10^{-9} Pa, Schottky emitters require 10^{-7} Pa and hairpin cathodes are happy with 10^{-4} Pa. Very small emitters need great mechanical stability, since movements have to be added to the virtual source size. Drift of the source can be annoying since the gun alignment needs to be adjusted.

6.2.4 A Note on Coherence

So far, we have been able to limit our view of an electron gun as a source of charged particles, little balls that fly through space. However, electrons have wave properties and these wave properties are crucial for the electron microscope, both in TEM where we measure phase contrast and in SEM, where diffraction ultimately limits resolution. We shall not go into quantum mechanics or wave theory or even wave optics in this chapter, but just state that the electron wavelength is

$$\lambda = \frac{h}{p} = 1.226 \times 10^{-9} \sqrt{\frac{1}{V*}} \tag{6.4}$$

when λ is given in m and $V*$, the relativistically corrected acceleration voltage of the beam, is given in volt, $V^* = V\left(1 + 10^{-6}V\right)$. For an electron beam, we are often interested in the coherence of the beam. If there is coherence between two points in the beam, it means that there is a constant phase relation between the electron waves at the two points: the amplitudes move as two pendulums in perfect unison. Only when there is coherence, interference can occur between the waves coming from these points. For effects such as diffraction in a crystal, there must be coherence over a width substantially larger than the unit cell. If only part of the amplitude moves in that unison, we speak of partial coherence. When there is partial coherence, the interference effects are weaker and sitting on a background (compare Fig. 6.1). This is mathematically expressed by the degree of coherence, which is the correlation between the wave amplitudes at points r_1 and r_2 (at a distance $x_{1,2}$).

$$\gamma(x_{1,2}) = \left| \lim_{\tau \to \infty} \int_{-\tau}^{\tau} \Psi^*(\vec{r}_1, t)\Psi(\vec{r}_2, t)dt \right| = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{6.5}$$

Remember that electrons are fermions, and fermions, unlike bosons such as photons, cannot interfere with each other. This means that different electrons are never mutually coherent, so coherence is a phenomenon that plays within individual electrons. When there is coherence between two points in a beam, it means that the electron wave of an individual electron has spread out over those two points.

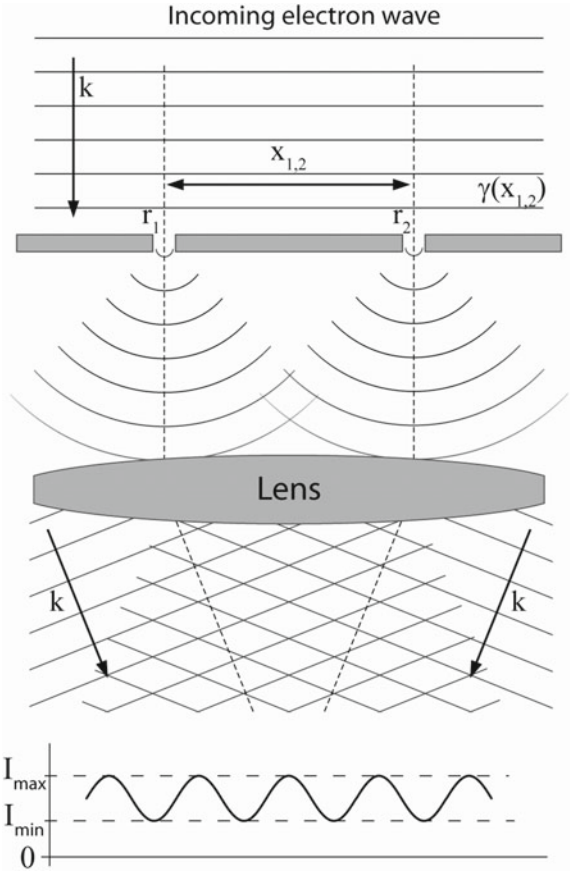


Fig. 6.1 Illustration of the relation between coherence and contrast in the interference pattern

In TEM, coherence is important in phase-contrast imaging. The contrast transfer function is affected by a coherence envelope. In phase-contrast imaging, one obtains contrast by the interference between the unscattered beam and scattered beams that went through a different point in the sample. Where these beams are not fully coherent, it results in loss of contrast. The distance over which an electron wave is coherent is inversely proportional to the internal angle in the beam, that is, the angle toward the nearest source image. For a large coherence length, this source image must be small. The wave from a small area spreads out according to the uncertainty principle:

$$\Delta p_x \cdot \Delta x_s = \frac{h}{2\pi}, \text{ so } \Delta x_t = \Delta v_x \cdot t = \frac{\Delta p_x \cdot l}{m \cdot v_z} = \frac{h \cdot l}{2\pi \cdot \Delta x_s \cdot p_z} \tag{6.6}$$

where Δx_s is the uncertainty of position in the x-direction in the crossover, l the distance between the crossover and the target plane and Δx_t the distance that the

wave has spread out at the target due to the uncertainty of momentum in the crossover plane. There is a problem, however, first in defining uncertainty quantitatively and second in relating that to the size of the source image. In the full van Cittert–Zernike theory, it is found that the degree of coherence $\gamma(x_{1,2})$ is directly proportional to the Fourier transform of the source image intensity function [7]. For instance, when the source image is a homogeneously filled disk, the Fourier transform is the Airy function. In a target plane illuminated from that crossover, the function γ has fallen by 12% when

$$x_{1,2} = \frac{\lambda}{2\pi\theta_c} = X_c \tag{6.7}$$

where θ_c is the semi-angle from the target to the crossover disk. X_c is usually called the coherence width [6]. At a distance

$$x_{1,2} = \frac{\lambda}{\theta_c} \tag{6.8}$$

the illumination is generally considered fully incoherent. At a distance

$$x_{1,2} = \frac{\lambda}{2\theta_c} = X_{pc} \tag{6.9}$$

there is still some coherence left and this is often taken as a practical coherence width, although the definition is somewhat arbitrary. The exact degree of coherence obtained between points at this distance will depend on the current density distribution in the source image and on the definition of the diameter of the source image. Probably, it is best to define the crossover diameter as the FW50 value.

There is a simple relationship between coherence and brightness: the current within the coherent area is

$$I_{coh} = B_{\mathrm{r}} V \frac{\pi}{4} X_{pc}^2 \pi\theta_c^2 = B_{\mathrm{r}} \cdot V \frac{\pi^2}{4} \left(\frac{\lambda}{2}\right)^2 = 0.93 \times 10^{-18} B_{\mathrm{r}} \tag{6.10}$$

This is a surprisingly simple relation that tells us we have a coherent beam as soon as we limit the current to smaller than $10^{-18}\ B_r$. This is important to realize: If the current is smaller than $10^{-18} B_{\mathrm{r}}$ the beam is coherent: it is a necessary and sufficient condition! In TEM, there is coherence over an area that contains this current. Thus, spatial coherence is not related to the size of the source, but rather to the size of the last source image before the target and the distance between that source image and the target. Note that lenses between that crossover and the target change the effective size and distance, but we do not need to know these exact relations: in practice, simply reducing the current density on the target increases the coherence.

So we now know what it means to say that there is coherence between two points in the beam, but what do people mean when they speak of a "coherent beam"? Basically, every point in the beam is coherent with every other point in the beam. So if the total current in the beam is smaller than $10^{-18}\ B_{\mathrm{r}}$, the beam is coherent. If such a beam is focused on a probe, the probe size will be diffraction limited. This implies that every beam can be made coherent by sufficient aperturing of the beam.

6.3 Focused Electron Beams

6.3.1 Contributions to the Probe Size

6.3.1.1 Source Image Contribution and the Role of Brightness

To form a probe, the source is imaged on the target. The minimum size of the probe may not be obtained at the exact image plane of the cathode, because in back-projection the electrons may seem to come from a smaller spot in a different plane than the cathode. This spot is called the virtual source and this is imaged with magnification M to a FW50 size Md_v. In the optics column, the aperture angle is always limited by an aperture to a half-angle α at the probe, corresponding to a half-angle α_e at the source. An infinitely small point in the virtual source plane is imaged to a blurred spot on the target as a result of diffraction and aberrations. The effect of the finite source size is taken into account by convoluting the blurred spot with the intensity distribution of the source image.

Since B_{r} is a conserved quantity through the whole system, the current in the probe can be calculated from

$$I_{\mathrm{p}} = B_{\mathrm{r}} \frac{\pi}{4} (M d_v)^2 \pi\, \alpha^2 V \tag{6.11}$$

where V is the accelerating voltage at the target and Br the practical brightness. From this, the contribution to the probe size can be expressed as

$$\mathrm{d}_I = M d_v = \frac{2}{\pi} \sqrt{\frac{I_p}{B_{\mathrm{r}} V}} \frac{1}{\alpha} \tag{6.12}$$

6.3.1.2 Diffraction

The size of the diffraction spot is often given as 1.22 λ/α, however, this is not the FW50 size, but the size of the first zero in the ring pattern. The FW50 value is

$$d_A = 0.54 \frac{\lambda}{\alpha} = 0.54 \frac{\Lambda}{V^{1/2}} \frac{1}{\alpha} \tag{6.13}$$

with $\Lambda = 1.226 \times 10^{-9}$ m $V^{1/2}$.

6.3.1.3 Aperture Aberrations

In general, aberration contributions increase with aperture angle. We'll treat a specific example which dominates in most systems: Spherical aberration causes the outer rays in the beam to be defocused by an amount $dz = C_s\alpha^2$, so at the Gaussian image plane, they have a distance $C_s\alpha^3$ from the axis. This leads to a "disk of least confusion" or minimum width of the total beam (FW100) of 0.5 $C_s\alpha^3$. However, the FW50 value has a much smaller minimum:

$$d_s = 0.18\, C_s\alpha^3 \tag{6.14}$$

In probe size calculations, it is very often allowed to use only the aberration coefficients of the last probe forming lens for the analysis. This is because usually the last lens demagnifies all aberration contributions of the other lenses in the system. However, for emitters with a very small virtual source size, such as the CNT emitter, the aberration coefficients of the gun lens must be taken into account explicitly.

6.3.1.4 Chromatic Aberration

Chromatic aberration causes the electrons with lower energy to be focused closer to the lens than electrons with higher energy. This leads to a blur contribution with FW50:

$$d_c = 0.6\, C_c \frac{\delta U}{V} \alpha \tag{6.15}$$

with C_c the chromatic aberration coefficient of the system and δU the FW50 of the energy distribution of the source. If the energy distribution is Gaussian, the FWHM of the distribution is very different from the FW50 (in contrast with a two-dimensional Gaussian distribution, where FW50 = FWHM). If one were to use that FWHM value, the prefactor in the chromatic aberration contribution would be 0.34 instead of 0.6.

6.3.1.5 Coulomb Interactions

In the world of cathodes, the best-known effect of the Coulomb force between electrons is the space charge limited emission from thermionic cathodes. We call this

a global space charge effect because it is calculated by assuming that the electron charge is distributed over space and not concentrated in particles. Of course, this is an important effect on the obtainable brightness, but it is not a separate contribution to the probe size that we need to take into account. Neither is the other global effect of space charge: the defocusing effect. This effect arises because a large current beam can apply a force on the electrons that are furthest away from the axis. This space charge force is, for a uniform beam, proportional with the distance to the axis and thus a pure defocusing effect. Only for nonuniform beams, global space can introduce space charge aberration.

However, a danger of high brightness electron beams is that stochastic Coulomb interactions (electron–electron scattering) disturb the trajectories, causing an additional blur of the probe. It is customary to distinguish between effects that increase the virtual source size (trajectory displacement) and effects that increase the energy spread (Boersch effect). Monte Carlo simulations can give reasonably accurate estimates of these effects, but are cumbersome. The effects can also be approximated by analytical equations, but these are strictly speaking only valid for sections in the optics column where the beam has one constant energy. Especially in the source region, this is not the case and one should use better approximations.

The equations to use depend on the local beam parameters such as the current density [8, 9]. For the typical parameters in an electron microscope, the Holtsmark regime [1] and the pencil beam regime occur. In the pencil beam regime, the electrons have a nearest neighbor distance that is larger than the diameter of the beam, so this regime is appropriate after the beam has been apertured. In the absence of apertures, the crossover size is blurred by a point spread function with a Holtsmark distribution of FW50:

$$\mathrm{FW}_{50\mathrm{H}} = 0.172\frac{m^{1/3}}{\varepsilon_0}\left[S_c^{2/3} + (1 - S_c)^{2/3}\right]\frac{I^{2/3}L^{2/3}}{V^{4/3}\alpha^{4/3}} \tag{6.16}$$

where L is the length of the segment under consideration in which the beam is focused after a distance $S_c L$. V is the acceleration potential, α the half-angle of the beam, m is the electron mass, and ε_0 the dielectric constant. If the aperture angle is sufficiently small, the section is in the pencil beam regime, with a FW50:

$$\mathrm{FW}_{50\mathrm{PB}} = 0.145\frac{m^{3/2}}{e^{7/2}\varepsilon_0}\left[S_c^3 + (1 - S_c)^3\right]\frac{I^3\alpha L^3}{V^{5/2}} \tag{6.17}$$

As a rule of thumb, when the regime may be either Holtsmark or pencil beam, the contribution to the crossover size can be found from

$$\mathrm{FW}_{50\mathrm{T}} = \left[\left(\frac{1}{\mathrm{FW}_{50\mathrm{H}}}\right)^{6/7} + \left(\frac{1}{\mathrm{FW}_{50\mathrm{PB}}}\right)^{6/7}\right]^{-7/6} \tag{6.18}$$

In Schottky sources and cold field emitters, the trajectory displacement reaches values comparable to the virtual source image easier than in thermionic emitters. The result is effectively a lower brightness of the source.

For the energy spread resulting from Coulomb interactions, the Lorentzian and the Holtsmark regime are most appropriate, with equations for the contribution to the energy spread (in eV) in one beam crossover:

$$\Delta \mathrm{E}_{\mathrm{FW}_{50\mathrm{L}}} = 1.41 \frac{m^{1/2}}{\varepsilon_0 e^{1/2}} \frac{I}{\alpha V^{1/2}} \tag{6.19}$$

$$\Delta \mathrm{E}_{\mathrm{FW}_{50\mathrm{H}}} = 0.891 \frac{m^{1/3}}{\varepsilon_0} \frac{I^{2/3}}{r_c^{1/3} \alpha V^{1/3}} \tag{6.20}$$

where r_c is the size of the crossover. When the energy spread comes from the segment that includes the emitter, r_c must be interpreted as the size of the emission surface. For Schottky emitters, this is the size of the low work function facet, usually about one-third of the tip diameter. Because the beam from the emitter effectively travels only through one side of the "crossover", the prefactors for the energy spread should be divided by 2. As a rule of thumb, when the regime may be either Holtsmark or Lorentzian, the contribution to the energy spread can be found from

$$\frac{E}{\Delta E_{\mathrm{FW}_{50\mathrm{T}}}} = \left[\left(\frac{E}{\Delta E_{\mathrm{FW}_{50\mathrm{L}}}} \right)^4 + \left(\frac{E}{\Delta E_{\mathrm{FW}_{50\mathrm{H}}}} \right)^4 \right]^{1/4} \tag{6.21}$$

6.3.1.6 Other Contributions

The whole analysis of probe formation given above assumes that the beam is perfectly in focus that there is no astigmatism, coma, movement, etc. And of course, the analysis changes when the spherical or chromatic aberration is corrected and fifth-order aberration starts to dominate.

6.3.2 Resolution and Probe Current

6.3.2.1 Resolution

In a SEM or STEM, its role is to deliver as much current as possible into a focused probe. This also applies to lithography and focused ion beam technology. In general, a larger probe can contain more current, but how much exactly, and what is the minimum probe size that can be produced, and does the minimum probe size depend on the choice of source? Before discussing these subjects, it is necessary to define

what is meant by "probe size". One parameter, the size, can of course not define every aspect of a focused beam. What we would like to know is the full intensity distribution: the current density as a function of the distance to the center of the probe. That distribution is also called the point spread function (PSF). If the distribution is Gaussian, it is all fairly simple and any measure of size can be used. The full width at half maximum (FWHM) for a Gaussian is the same as the diameter that contains 50% of the total current (FW50) and the same as the distance between the 12% point and the 88% point of the current that is measured when scanning over a knife-edge. Many distributions, however, are not Gaussian. One example is the distribution in a spherical aberration dominated probe: at under-focus, this distribution has a very bright central spot on a broad background and at over focus, the highest intensity is at the edge of the probe. The FWHM may then be small, but it contains hardly any current and is a rather artificial measure for probe size. We have found that the FW50 is the most consistent measure to characterize a probe, also when it is non-Gaussian.

When there are different contributions to the probe size, for instance, from diffraction, spherical aberration, and source image and all contributions have a Gaussian intensity distribution, it is simple to add all contributions to the probe size: they can be added in quadrature [11]. When the distributions are non-Gaussian, this does not work (compare Fig. 6.2a). For some specific analytical forms of the distribution, an addition rule can be derived, for other distributions only a numerical approach can give an approximate addition rule.

The fundamentally correct way to calculate the distribution function in an electron probe is to start with the wave optical calculation of the diffraction distribution as modified by the spherical aberration of the lens. It is necessary to do this for a range of

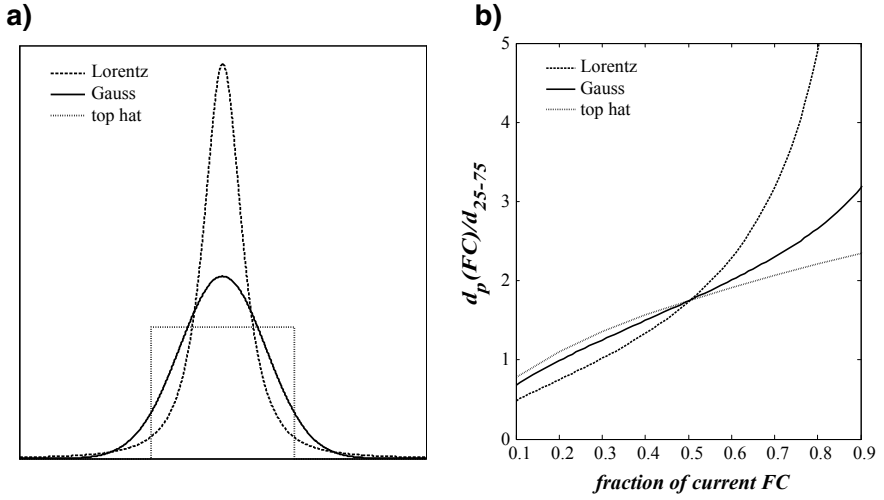


Fig. 6.2 **a** Cross section through the probe, normalized for content, for a Gaussian, Lorentzian, and a top hat distribution (2D). **b** Ratio of the probe size dp(FC) and the 25–75% edge resolution, d25–75, for different current fractions

defocus values, since we do not know yet which defocus will give the smallest probe size. Subsequently, these distributions must be convoluted with the distribution of the (defocused) source image. Finally, the chromatic aberration effects can be taken into account by convoluting the distributions in the defocus direction with the energy spread function. The minimum size of the probe is not found in the Gaussian image plane, but, due to the spherical aberration, at a small defocus this whole procedure is so complicated that it is not practical for optimization of the probe size. The equation that yields the best fit to the minimum FW50 results of this correct procedure is [11]:

$$d_p = \left(\left(\left(d_I^{1.3} + \left(\left(d_A^4 + d_s^4 \right)^{\frac{1}{4}} \right)^{1.3} \right)^{\frac{1}{1.3}} \right)^2 + d_c^2 \right)^{\frac{1}{2}} \tag{6.22}$$

where d_I, d_A, d_s, and d_c are the FW50 values of the contributions from the source image, the diffraction disc, the spherical aberration, and the chromatic aberration, respectively. So we first add the diffraction to the spherical aberration with a fourth power rule, then add the source size contribution with a 1.3 power rule and finally add the chromatic aberration effect with a simple square power rule.

Optimization of the probe size consists of finding the aperture angle and source-to-probe magnification that gives the smallest probe size or the largest current in a specified probe size.

To calculate I versus d_p curves, one proceeds as follows [1, 12]. Given a source with its emission properties and lenses with their aberration coefficients, one first chooses a probe current for which to find the minimum d_0. The only variable now left in the equation for probe size is the half aperture angle α. Choosing α means balancing the contributions from the probe lens aberrations, which increase with α and the contributions from diffraction and source image, which decrease with α.

At very small currents, the size of the source image is negligible. If we assume the spherical aberration contribution to be much smaller than the chromatic aberration contribution, the probe size is

$$d_p = \left(\left(0.54 \frac{\Lambda}{V^{\frac{1}{2}}} \cdot \frac{1}{\alpha} \right)^2 + \left(0.6 C_c \frac{\delta U}{V} \alpha \right)^2 \right)^{\frac{1}{2}} \tag{6.23}$$

For optimized α, this is

$$(d_p)_{\min} = d_{AC} = 2.81 \times 10^{-5} \frac{C_c^{\frac{1}{2}} \delta U^{\frac{1}{2}}}{V^{\frac{3}{4}}} \quad \text{at} \quad \alpha = \sqrt{\frac{0.54 \Lambda V^{\frac{1}{2}}}{0.6 C_c \delta U}} \tag{6.24}$$

Thus, the minimum probe size is not dependent on the brightness or the coherence of the emitter; the only emitter property that enters the equation is the energy spread!

If we assume the chromatic aberration contribution to be much smaller than the spherical aberration contribution, the probe size is

$$d_{\mathrm{p}} = \left(\left(0.54 \frac{\Lambda}{V^{\frac{1}{2}}} \cdot \frac{1}{\alpha} \right)^4 + \left(0.18 C_{\mathrm{s}} \alpha^3 \right)^4 \right)^4 \tag{6.25}$$

For optimized α, this is

$$(d_{\mathrm{p}})_{\min} = d_{\mathrm{AS}} = 1.03 \times 10^{-8} \frac{C_{\mathrm{s}}^{1/4}}{V^{3/8}} \quad \text{at} \quad \alpha = 1.23 \left(\frac{\Lambda}{C_{\mathrm{s}} V^{1/2}} \right)^{1/4} \tag{6.26}$$

Again, the minimum probe size is not dependent on the brightness or the coherence of the emitter; here it does not depend on any emitter property.

6.3.2.2 Current in a Probe

However, a probe without current is of no use, so we have to allow a contribution to the probe size that is related to the current in the probe. By expressing the probe size contribution in terms of the diffraction contribution, the probe current can be rewritten as

$$I_{\mathrm{p}} = B_{\mathrm{r}} \frac{\pi}{4} d_I^2 \pi \alpha^2 V = B_{\mathrm{r}} \left(0.54 \frac{\pi}{2} \Lambda \right) \left(\frac{d_I}{d_A} \right)^2 = B_{\mathrm{r}} K_{\Lambda} \left(\frac{d_I}{d_A} \right)^2 \tag{6.27}$$

where $\mathrm{K}_{\Lambda} = 1.08 \times 10^{-18}$ m^2 sr V. This is a very useful equation, because it gives a simple estimate of how much current is obtained in a probe which is close to the minimum size. When the FW50 of the virtual source image equals the FW50 of the diffraction spot ($\mathrm{d}_I = \mathrm{d}_A$), we call the current I_{Λ}, with $I_{\Lambda} = K_{\Lambda} B_{\mathrm{r}}$. For example, for a Schottky emitter at $B_{\mathrm{r}} = 5 \times 10^7$, this current is 50 pA, independent of the beam energy or the lens aberrations! Beam energy and lens aberrations of course do determine the size of the probe that contains this current.

It is often useful to express the current in the probe as a function of probe size. Examples are shown in Figs. 6.3 and 6.4. For a chromatic aberration limited probe, close to the best obtainable resolution d_{AC}, at optimum aperture angle, the current is

$$I_p = I_{\Lambda} \left[\left(2 \frac{d_p^2}{d_{\mathrm{AC}}^2} - 1 \right)^{1.3} - 1 \right]^{\frac{2}{1.3}} \tag{6.28}$$

when the diffraction term becomes small relative to the source image, so the main contributions to the probe are source image and chromatic aberration, the current in a probe of FW50 size d_{p} is

$$I_{\mathrm{p}} = 1.71 \frac{\mathrm{d}_{\mathrm{p}}^4 B_{\mathrm{r}} V^3}{C_{\mathrm{c}}^2 \delta U^2} \tag{6.29}$$

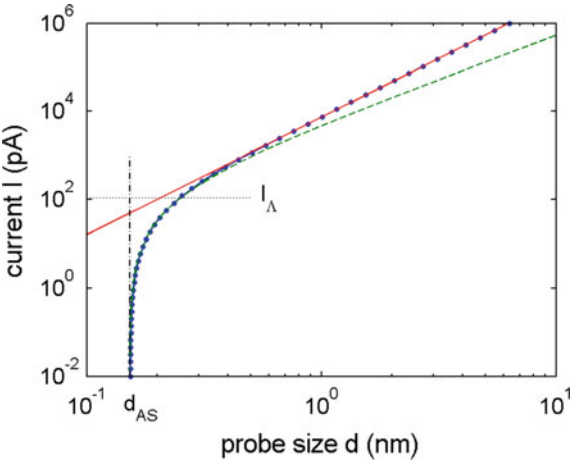


Fig. 6.3 I-d relation for Cs-limited probe (Cs = 1 mm, U = 300 keV, B_r = 10^8). The green/dashed line follows from (6.26) or (6.27); the red/solid line follows (6.31)

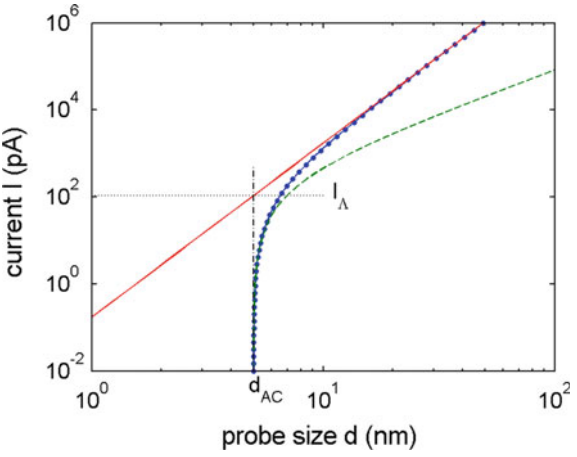


Fig. 6.4 I-d relation for Cc-limited probe (Cc = 2 mm, U = 1 keV, dU = 0.5 eV, B = 10^8). The green/dashed line follows from (6.24) or (6.27); the red/solid line follows (6.29)

Not every microscope has a continuously variable magnification and aperture angle, so the equation gives the best that can be reached [11, 12].

For a spherical aberration limited probe, close to the best obtainable resolution d_{AS} the current is

$$I_p = I_\Lambda \cdot \frac{4}{3}\left[\left(\frac{d_p}{d_{AS}}\right)^{1.3} - 1\right]^{2/1.3} \tag{6.30}$$

when the diffraction term becomes small relative to the source image and the main contributions to the probe are source image and spherical aberration, the current in

a probe of FW50 size d_p is

$$I_p = 2.44 \frac{d_p^{8/3} B_r V}{C_s^{2/3}} \tag{6.31}$$

6.4 Thermionic Sources

6.4.1 Emission Theory

Thermionic emission arises when, by heating, a number of the conductance electrons are getting energies higher than the work function (the potential barrier faced by electrons in the conductor). They then have sufficient energy to cross this potential barrier (see Figs. 6.5 and 6.6). When the work function is denoted by φ, the current density at the surface is expressed by the Richardson equation [1]:

$$j = A(1-r)T^2 \exp\left(-\frac{\varphi}{kT}\right) \tag{6.32}$$

where A is the emission constant, $1 - r$ is the transmission coefficient of the surface barrier for electrons, T is the temperature, and k is the Boltzmann constant (1.38×10^{-23} J/K = 86 μeV/K). φ and kT are both in eV. The quantum-statistical derivation of this equation leads to

$$A = \frac{4\pi mek^2}{h^3} = 1.204 \times 10^6 \text{ A/(m}^2\text{K}^2\text{)} \tag{6.33}$$

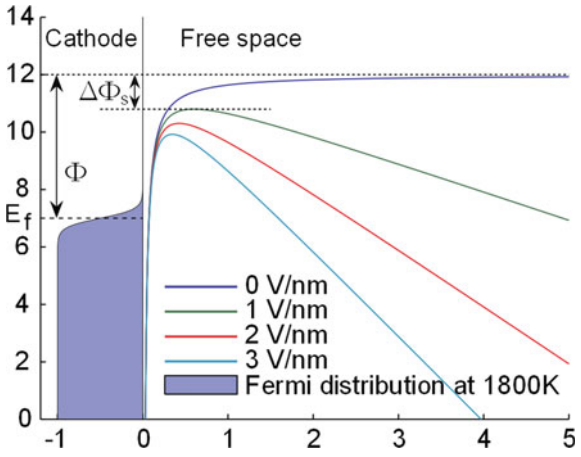


Fig. 6.5 The basics of emission theory: first determine the occupation of the states in 3D momentum space (left in the energy diagram) and then calculate the quantum mechanical probability of escaping from each state into vacuum, either over the potential barrier, or tunneling through the barrier; see [1, 12]

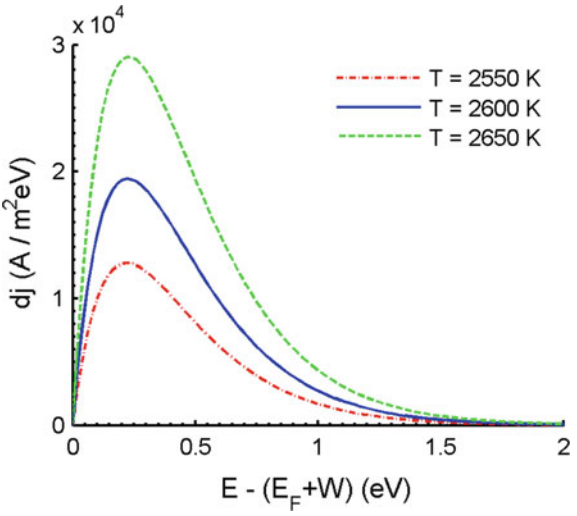


Fig. 6.6 Energy spectrum of a thermionic source, work function 4.5 eV

In general, $1 - r$ is close to one. Furthermore, there is a slight temperature dependence of φ, so the work function will have the general form:

$$\varphi = \varphi_0 + cT \tag{6.34}$$

For tungsten, measurements have shown that

$$\varphi = 4.52 + 0.15 \cdot 10^{-4} T \tag{6.35}$$

However, the work function also varies with different crystal orientations.

In practice, a simpler equation is often used for the current density:

$$j = AT^2 \exp\left(-\frac{\varphi}{kT}\right) \tag{6.36}$$

in which the values of A and φ are then adapted for the measured current densities.

At the surface of the emitter, the mean energy in the plane of the surface is kT. As long as the acceleration is perpendicular to the surface, this is conserved, so $\alpha^2 \cdot e \cdot V = kT$

The intrinsic reduced differential brightness can then be expressed as (dA is emission surface):

$$B_\mathrm{r} = \frac{\mathrm{d}I}{\mathrm{d}A\mathrm{d}\Omega \cdot V} = \frac{ej}{\pi kT} \tag{6.37}$$

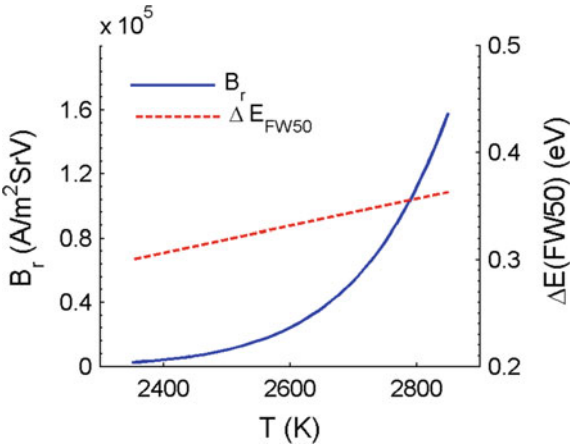


Fig. 6.7 Brightness and energy spread of a thermionic source with work function 4.5 eV

The practical brightness is then

$$B_{\mathrm{r}} = \frac{I}{\frac{\pi}{4}\mathrm{d}_g^2 \pi\alpha^2 \cdot V} = 1.44\frac{ej}{\pi kT} \tag{6.38}$$

if we take the FW50 value (diameter that contains 50% of the current) for the source size. The energy distribution is given by

$$\mathrm{d}j(E) = \frac{A}{k^2} \cdot E \exp\left(-\frac{E+\varphi}{kT}\right)\mathrm{d}E \tag{6.39}$$

An example is shown in Fig. 6.7. When the current density of an emitter becomes too high, the emission becomes limited by space charge. The electrons that have just left the cathode to push the slowest electrons just behind them back into the emitter. For a flat cathode at a distance d from a flat anode at voltage V_d, we find the current density through the Child–Langmuir equation for space charge limited emission:

$$j = \frac{4}{9}\varepsilon_0\sqrt{\frac{2e}{m}}\frac{V_d^{3/2}}{\mathrm{d}^2} \tag{6.40}$$

6.4.2 Practical Aspects

For electron microscopy, the emitting surface may be very small, for thermionic emitters in the order of tens of microns. However, it is difficult to hold an object that small and heat it to a few thousand degrees. So in general, the heated object is larger and the unwanted emission from most of the surface is pushed back into the material. This is the function of the "Wehnelt electrode". The Wehnelt is at a negative potential

with respect to the cathode, so electrons are reflected before they can reach it. The strong attracting force from the positive anode can reach the surface of the emitter only through a small hole in the Wehnelt. The voltage on the Wehnelt determines how large exactly the emitting area is.

The cathode itself may be just a simple tungsten filament, bent like a hairpin, and heated by the current flowing through it. Although W does not have a particularly low work function, it can sustain high temperatures for a long time so that the obtained brightness and lifetime is better than most other materials. Another popular material is lanthanum hexaboride, LaB_6, a crystal with a low work function surface. The crystal is usually sharpened and mounted on a W wire or more complicated construction and again heated by the current flowing through it. The lower work function is only maintained if the emitter stays clean, so a LaB_6 source needs a better vacuum, in the order of 10^{-6} Pa.

The billions of television tubes and computer monitors of the twentieth century all had "CRT cathodes" (CRT for cathode ray tube), consisting of a small porous W pill impregnated with low work function material such as barium oxide. CRT cathodes are either heated by a simple, separate heating wire, or by electron bombardment on the backside of the pill from a separate filament.

The field from the anode, protruding through the Wehnelt hole forms a strong lens, causing the electron beam to form a crossover between the Wehnelt and the anode. The size of this crossover is the virtual source size for the rest of the system. If the Wehnelt voltage is too high, the beam is cut off, or nearly cut off, which reduces the brightness. If the Wehnelt voltage is too low, emission from all over the filament can get into the beam, complicating the formation of a well-defined beam. The Wehnelt voltage in a hairpin cathode for microscopy is typically chosen to give an emission current of about 100 μA at a temperature of about 2600 K, selecting an emission area of approximately 100×100 μm. The exact voltage depends on the geometry of the gun. Often, the voltage on the Wehnelt is supplied through a Wehnelt resistor between the filament and the high-voltage power supply of the electron source, so the bias voltage is proportional to the emission current. This complicates the understanding of what is going on, because a higher emission current now increases the Wehnelt voltage, thus reducing the emission area and as a consequence, the emission current! In practice, one chooses a Wehnelt resistor and then increases the filament current (temperature) until the emission current becomes almost independent of temperature (the "knee" in the curve). It is wise to just follow the instrument's user instructions.

When there is no Wehnelt electrode to limit the cathode area, the whole filament, or W pill, will emit, so the size must be restricted. There will be no crossover, but usually the current is so high that space charge effects dominate the behavior.

The life time of a tungsten hairpin filament depends strongly on the temperature of the tip as an effect of direct thermal evaporation. The hottest point is just next to the tip, so that is where the filament breaks. A typical empirical relation states that the lifetime t (in hours) relates to the brightness as: $t - 5 \times 10^5/B_r$. Too high background pressures of water or oxygen may reduce the lifetime further. The maximum erosion rate is at slightly lower temperatures than the typical emitter temperature, thus the

effect occurs further away from the tip than thermal evaporation. Lifetime reduction starts at pressures in the order of 10^{-2} Pa. CRT cathodes, when operated in a closed vacuum bottle with some getter material to keep the vacuum low may live for many years, as we all know from our old television tubes.

6.4.3 Recent Developments

6.4.3.1 New Materials in Cathodes for SEMs

All electron microscope types started out with tungsten hairpin cathodes. Still, for economic reasons, many scanning electron microscopes are sold with these simple sources, which have not changed over the past few decades. One level up in brightness is the Lanthanum Hexaboride (LaB_6) cathode, where a small crystal of LaB_6 is welded onto the tungsten hairpin or mounted in a so-called Vogel mount. The (100) surface of the crystal has a lower work function than tungsten and thus delivers a higher current density even at a lower temperature. The vacuum requirements are a bit more stringent. In recent years also Cerium Hexaboride (CeB_6) crystals have become commercially available, which seem to have a longer lifetime than LaB_6 crystals [13].

6.4.3.2 Brightness Optimization for Shaped Beam and Projection Lithography

Shaped beam electron lithography is presently the preferred technique to write masks for the semiconductor industry. High brightness beams are important for the writing speed, but Schottky sources cannot supply sufficient total current, so in these machines, LaB_6 cathodes are used. Recently an old idea has been implemented to limit the emission from the side surfaces of the crystal. Any extra emission that is not used in the beam contributes to the space charge and Coulomb interaction effects and is thus undesirable. In the past, LaB_6 crystals have been mounted inside cylinders of other material to avoid emission from the side surfaces. It has now been shown [13] that covering the sides with Carbon is also effective in reducing the emission. Effectively this enhances the brightness of the beam in the lithography machine above the previously used 10^5 A/m^2 sr V.

When large area projection electron beam lithography was under investigation, see, for instance, [14, 15], an interesting conclusion concerning the source brightness was drawn: in a certain range of operation parameters, it was better to reduce the brightness. The reason was that, given a total current in the system, it was better to have larger intermediate crossovers in order to avoid stochastic Coulomb interactions [16, 17]. By placing a grid in the beam, the local grid lens effect modified the brightness to the required value [17].

6.4.3.3 Splitting the Beam for Multibeam Lithography

The current that can be focused into a single small spot is limited. Single Gaussian beam electron lithography machines are thus inherently limited in throughput. Shaped beam lithography machines and cell projection lithography machines can have a substantially higher current and thus throughput, but the manipulation of the shaped beam, respectively, the inflexibility of the cell projection, impose their own limits. For high-throughput electron beam lithography in the semiconductor industry, multibeam systems could be a solution. Development of these kinds of machines was started in many labs, based on arrays of pointed emitters as well as based on splitting the beam from a single thermionic cathode [18, 19].

In the MAPPER system [20], a BaO dispenser cathode is operated without a Wehnelt electrode to obtain a uniform current density from the cathode surface. The beam, with a total current in the order of 10 mA, is split into sub-beams by a plate with apertures. In the full lithography system, the plate contains 13,000 holes, it the source test system only 194 beams were extracted. Of these beams, the brightness was measured as a function of cathode temperature and extraction voltage (Fig. 6.8). Brightness values as high as 10^6 A/m^2 sr V could be obtained. The interesting outcome was that the brightness dropped when the cathode came out of space charge limited emission and entered the thermal emission limited regime.

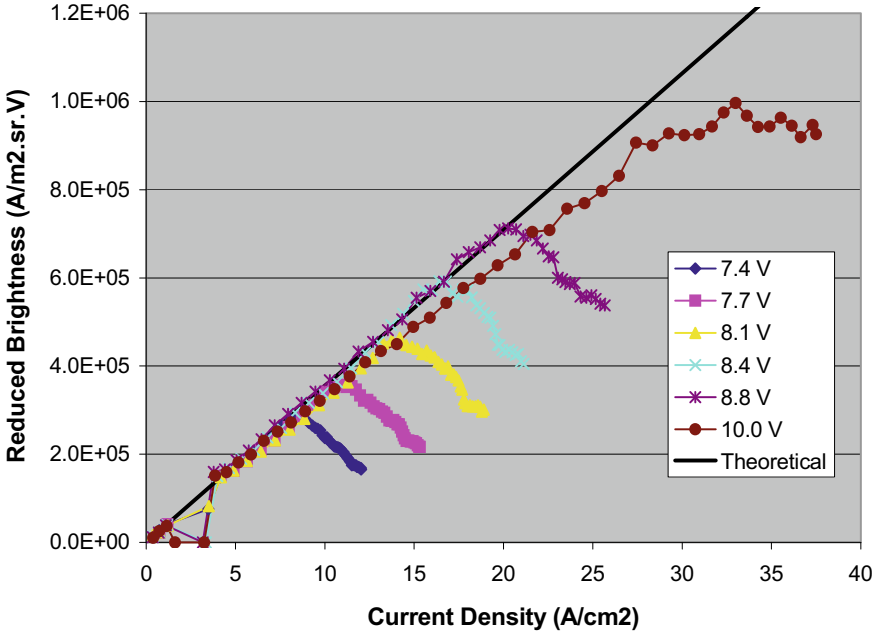


Fig. 6.8 Reduced brightness for different emitter temperatures. The filament voltage range of 7.4–10.0 V corresponds to an emitter temperature range of 1380–1500 K; compare [21]

Probably, the space charge layer effectively smoothes the cathode surface [21]. In the lithography machine, stochastic Coulomb interactions in the section before the aperture plate can still reduce the brightness and space charge effects can change the uniformity of the current density in the beam. However, from recent, unpublished, presentations at conferences, this seems to be well under control. Thus, multibeam lithography has appeared as a new use of the "old" CRT-type cathodes now that the cathode ray tube has practically disappeared after the revolution in oscilloscopes and television screens.

6.4.3.4 Outlook

Thermionic emitters have a history of more than a hundred years. For most high-resolution applications, the Schottky emitter and cold field emitter have taken over. However, there is one aspect where these pointed emitters cannot compete with the extended area emission of the thermionic cathode: total emission current. It is, therefore, that high-throughput lithography systems such as shaped beam and multibeam systems employ thermionic emitters. The extraction of thousands of beams from a single source is a recent development that requires very specific properties of the cathode such as simultaneous high and uniform brightness. This may lead to new insights and further cathode development.

6.5 Schottky Sources

6.5.1 Emission Theory

The Schottky effect is the effective decrease of the work function when an external field F is applied at the metal surface (see Fig. 6.5). The necessary fields are very high and can only be obtained by sharpening a tip to submicron size, so that the field in front of the tip is enhanced by a large factor. The field enhancement factor β is defined through $F = \beta V_{E,}$ where V_E is the extractor potential [1] (Fig. 6.9).

The magnitude of the Schottky effect can be analyzed by calculating the potential curve based on the image force potential resulting from the attraction between an electron outside the surface and its image behind the surface. The net potential is now the sum of the zero-field image force potential and the potential due to the field. For the lowering of the potential barrier, we find (in eV):

$$\Delta\varphi = -\sqrt{\frac{e^3 F}{4\pi\varepsilon_0}} \tag{6.41}$$

This increases the thermionic emission by a factor $\exp(-\Delta\varphi/kT)$ to

$$j = AT^2 \exp\left(-\frac{\varphi - \sqrt{e^3 F/4\pi\varepsilon_0}}{kT}\right) \tag{6.42}$$

Experimental plots of ln I' (the angular current density) versus $\sqrt{V_E}$ are called "Schottky plots" (see Fig. 6.10). Since $F = \beta V_E$ and $I' = j(r/M_\alpha)^2$, the slope of these plots yield the value of the field enhancement factor β:

$$\beta = \frac{4\pi\varepsilon_0}{e^3}(k_B T)^2 \text{slope}^2 = 5.157(T \cdot \text{slope})^2 \tag{6.43}$$

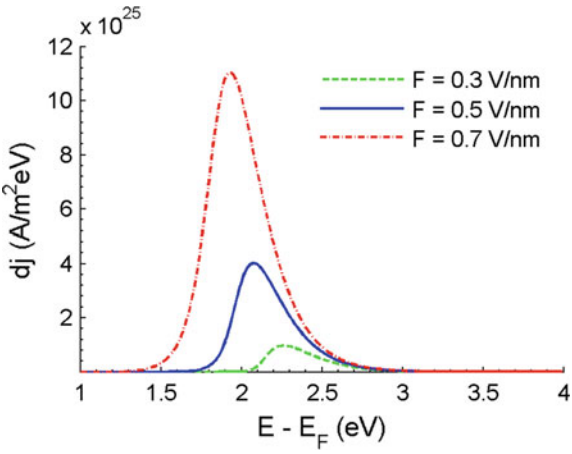


Fig. 6.9 Energy spectra of Schottky emitter; see [1], courtesy of Springer

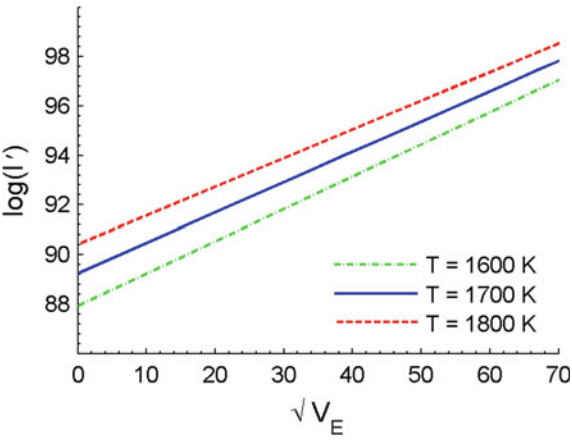


Fig. 6.10 Schottky plot; see [1], courtesy of Springer

Since r is the tip radius and M_α is the angular magnification of the extraction field, with estimates for M_α and φ, the Schottky plot also yields the tip radius r. The Schottky plot can also give the work function:

$$\phi = k_B T \left(\ln \frac{4\pi em}{h^3} (k_B T)^2 + 2 \ln \frac{r}{M_\alpha} - \text{axis intercept} \right) \tag{6.44}$$

Note that there are several assumptions in the model behind the Schottky plot: no tunneling through the barrier, the work function is independent of the field and the tip is a sphere, not a flat facet. With approximate equations from [22–24], the Schottky plot is still useful. One has to take into account that M_α in itself is a function of β and r, as $M_\alpha = 0.525(\beta r)^{0.42}$. In order to keep the M_α constant, one should scale the voltage of the suppressor with the changing extractor voltage. It is also possible to fit the Schottky plot to a more complete model and extract the parameters from the fit.

The energy distribution for pure Schottky emission is the same as for pure thermionic emitters with the lower work function. The position of the energy spectrum shifts with increasing field to lower energies (see Fig. 6.9).

When the field on the emitter surface is sufficiently increased, electron tunneling through the top of the potential barrier occurs. The emission is then in the "extended Schottky regime". In order to find analytical equations for the emission current and energy distribution, the form of the energy barrier is approximated by a parabolic function. The emission current density is then found from

$$j_{ES} = AT^2 \exp\left(-\frac{\varphi - \sqrt{e^3 F / 4\pi\varepsilon_0}}{kT} \right) \cdot \frac{\pi q}{\sin \pi q} \tag{6.45}$$

with

$$q = \frac{h(4\pi\varepsilon_0 e)^{1/4}}{2\pi^2 m^{1/2}} \frac{F^{3/4}}{kT} = \frac{\kappa}{kT} = 0.166 \times 10^{-3} \frac{F^{3/4}}{T} \tag{6.46}$$

As in thermionic emitters, the practical brightness is

$$B_r = 1.44 \frac{e j_{ES}}{\pi kT} \tag{6.47}$$

The energy distribution broadens to

$$\begin{aligned} dj_{ES}(E) =& A \frac{\kappa}{1 + \exp\left(\left(E + \varphi - \sqrt{e^3 F / 4\pi\varepsilon_0} \right) / kT \right)} \ln\left[1 + \exp(E/\kappa) \right] dE \\ &\approx dj_S(E) \cdot \frac{\kappa}{E} \ln\left[1 + \exp(E/\kappa) \right] \end{aligned} \tag{6.48}$$

with $\mathrm{d}j_s(E)$ the energy distribution for pure Schottky emission.

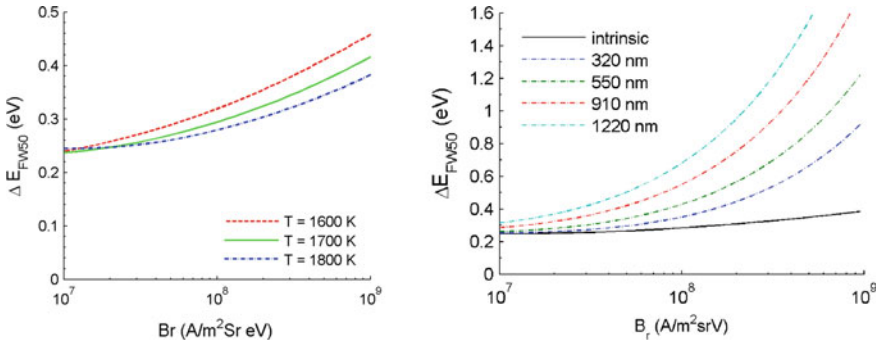


Fig. 6.11 Intrinsic energy spread (left) and full energy spread (right) versus intrinsic brightness for Schottky emitters; see [1], courtesy of Springer

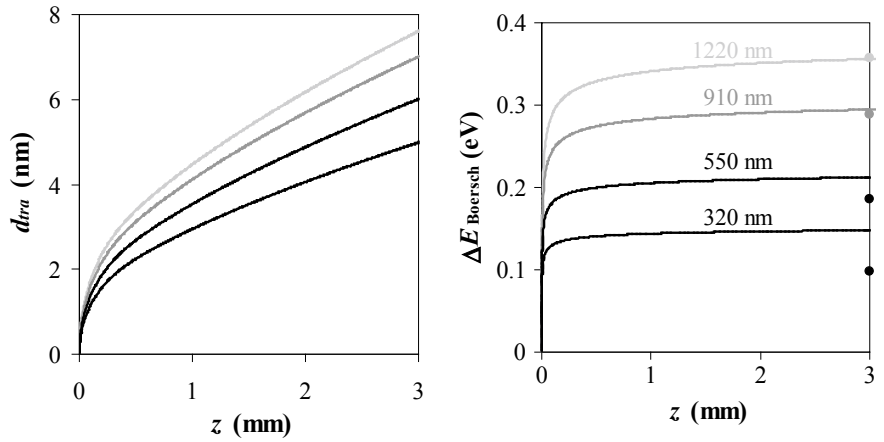


Fig. 6.12 FW50 trajectory displacement and FW50 Boersch effect as a function of distance from the emitting surface z, for different emitter geometries, calculated using the slice method [25, 26] The markers in the Boersch plot are calculated with an approximate equation for the whole segment. The field strength at the facet center for all emitters is 1 /nm, the temperature 1800 K

These approximations are valid up to a value of q of about 0.7 when extracting the energy spectrum, but only up to $q = 0.25$ for the current density j_{ES}. For stronger fields, only numerical solutions are available (Figs. 6.10, 6.11, 6.12, and 6.13).

6.5.2 *Practical Aspects*

Although Schottky emission may occur from any heated material in a strong field, we shall limit the further discussion to the most commonly used materials: single crystal W with a 100 surface at the tip end, coated with zirconium oxide [22]. At a

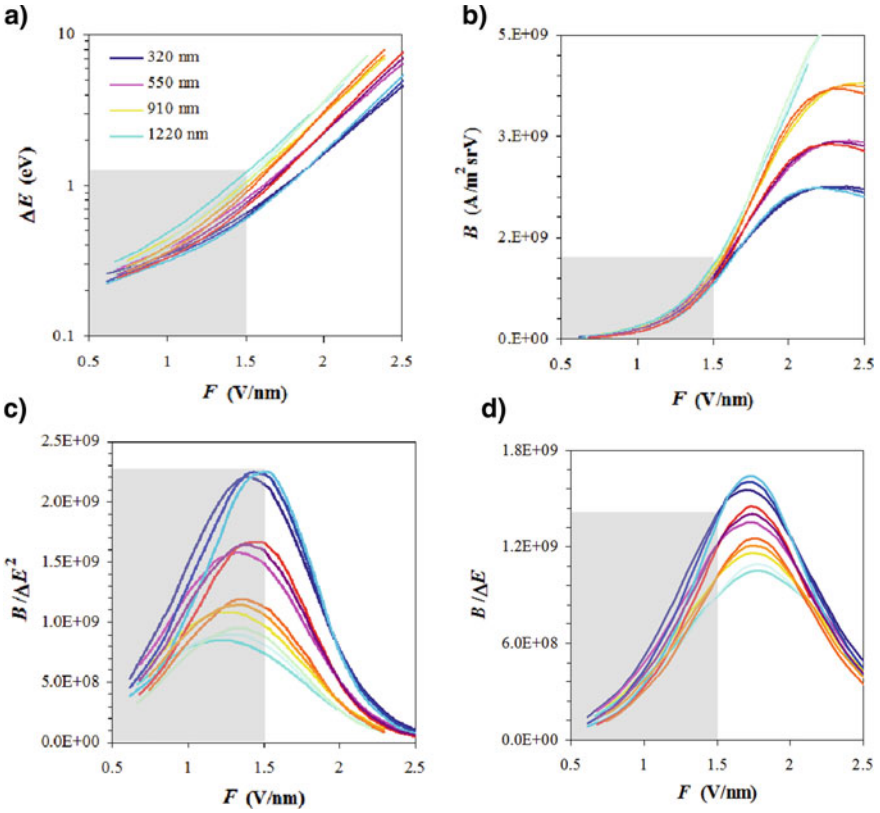


Fig. 6.13 **a–d** FW50 total energy spread, B, $B/\Delta E^2$, and $B/\Delta E$, as a function of the field strength at the surface, for 1600, 1700, and 1800 K, and for four different emitters; evaluated at 1 mm from the emitting surface. The gray boxes indicate approximately the field range for which the tip can be stable geometrically (no collapses, no microprotrusions); according to Bronsgeest 2009 [23]

temperature above 1500 K, the ZrO forms an electric dipole layer at the W surface that lowers the work function from about 4.5 eV to about 2.9 eV. The tip radius ranges from 0.3 to 2 μm. The front face forms a flat facet surface, which ideally has a diameter of about one-third of the tip diameter (see Fig. 6.14). The operating temperature is about 1850 K. A typical suppressor voltage is −300 V and a typical extractor voltage is 5 kV.

Ideally, one would know the tip-suppressor-extractor geometry, calculate the field distribution using simulation software, find the angular magnification from trajectory tracing, measure the tip temperature with a pyrometer, look up the work function, and use the above emission theory to find brightness, angular current density, virtual source size, and energy distribution. In practice, the tip shape is unknown, the geometry is known only approximately, the work function depends a little on the vacuum

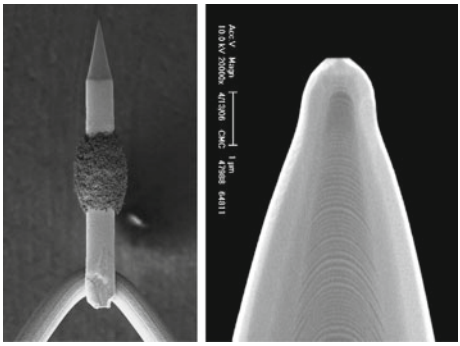


Fig. 6.14 Schottky emitter with ZrO and tip end of the emitter; see [1, 23]

conditions and the temperature, the temperature is read from a calibration curve that relates the filament current to the temperature and can be off by 50–100 K. The extractor voltage is known. The total emission current is measured and the probe current is measured, but it is usually unknown which aperture angle at the extractor contains this probe current. A Schottky plot can be obtained by plotting, at moderate extraction voltages so as to stay in the pure Schottky regime, the probe current versus the extractor voltage. This yields the field F (and thus the theoretical brightness and energy spread when temperature is assumed) and an approximate tip radius, but only if the angular magnification is estimated. Brightness can be measured directly and so can the energy distribution if an energy spectrometer is available.

For a 1 μm diameter tip with a 300 nm diameter facet, a typical angular current density is 0.7 mA/sr, total emission from the facet up to 40 μA, emission from the shank 100 μA, virtual source size 30 nm.

The practical beam parameters of Schottky emitters yield particle densities that are sufficiently high for stochastic electron–electron interactions to occur. It turns out that the theoretical equations are given elsewhere in this volume give a reasonable fit to experimental results for the Boersch effect, which is the energy broadening effect [25, 26]. For small emitter sizes, the Lorentz regime is applicable, for larger tips it is the Holtsmark regime. When the Boersch effect is added to the intrinsic spread, the relation between brightness and energy spread changes to what is plotted in Fig. 6.13. The smallest energy spread at a given brightness is obtained at the smallest facet size. The main contribution to the Boersch effect is very close to the tip, where the electron density is highest. Broadening of the virtual source size (trajectory displacement) is the angular deflection caused by electron–electron interaction, multiplied by the distance between the virtual source and the position of the deflection. If these are added correctly in what is known as the "slice method", we find values as in Figs. 6.11, 6.12, and 6.13.

Thus, the contribution from the section between the tip and the extractor is not very large [25, 26] and would theoretically allow values above 10^9 A/m^2 sr V. The main danger of virtual source size increase and brightness loss lies in the section between the extractor and the accelerator. For low-energy scanning electron microscopes,

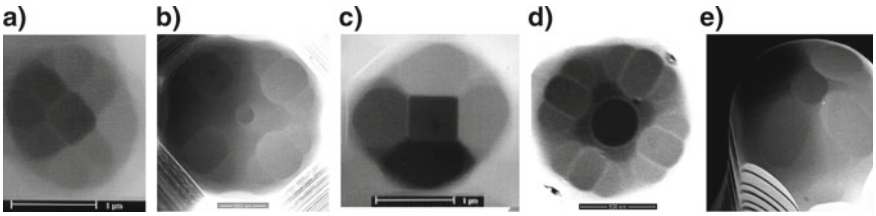


Fig. 6.15 **a–e** Shapes of the W(100) facet in a Schottky emitter after different operating conditions (according to Bronsgeest 2009 [23, 24], see also [1])

the danger is in the whole column. In the early years of the Schottky emitter, the brightness was limited to about 2×10^7 A/m^2 sr V because of this effect. Nowadays, the current for high brightness/low current mode of operation is usually limited soon after the extractor to a value in the order of 10 nA, allowing brightness values of up to 2×10^8 A/m^2 sr V.

The shape of the W tip depends strongly on its history, in terms of which temperatures and fields it has been submitted to, see Fig. 6.15. The creation of a flat facet on an etched, round tip, is a natural result of the thermodynamics of the surface. At elevated temperatures, W surface atoms are mobile and diffuse over the surface. They attach and detach to surface steps such that terraces grow or shrink.

At low fields (Fig. 6.15b), the atoms tend to move away from terraces of smaller radius toward steps of larger radius. Thus, the center terraces on the tip disappear until a flat facet is obtained. In operation, this effect may continue when the extraction voltage is too low. When a number of these terraces pile up and shrink together, this is recognizable as a moving ring in the angular emission pattern. The effect is known as "ring collapse". At high fields, the concentration of adatoms is influenced by the field in such a way that the concentration gradient at the facet end reverses and the ring collapse stops. At very high fields (Fig. 6.15c), the thermodynamics prefers crystal surfaces over a round tip and we see faceting of the side surfaces occur. At intermediate fields, the facet is reasonably stable, although in the course of many months, the tip head always grows in diameter. This dependence of the tip shape on temperature and field makes it important to follow the correct start-up procedures for a tip and not operate it at temperatures or fields far away from the recommended values.

Even the most stable tip has a finite lifetime because the ZrO slowly evaporates. The rate of evaporation is, of course, a steep function of temperature: If a tip has a lifetime of 10,000 h at 1750 K, this reduces to a 1000 h at 1900 K.

The ZrO-coated W must be operated in pressures below 10^{-7} Pa. If the pressure is too high, this may influence the work function and thus reduce the emission current. Especially O_2 and H_2O are dangerous in this respect. Another effect that sometimes occurs is a hindering of the flow of ZrO toward the facet, making the emission unstable.

6.5.3 *Recent Developments*

6.5.3.1 Higher Brightness for Electron Microscopy

Although theoretical values for the brightness of Schottky emitters go above 10^9 A/m^2 sr V, experimental values are limited by the instability of the tip shape and the practical problems in cutting the current soon enough after the emitter to avoid stochastic Coulomb interactions. A brightness of 2×10^8 A/m^2 sr V was shown already a while ago in an experimental source test setup [27], but similar values were only recently reported for a source in a commercial electron microscope [28]: 3–4 $\times$ 10^9 A/cm^{2}sr@300 kV, corresponding to 1–1.3 $\times$ 10^8 A/m^2 sr V.

6.5.3.2 Higher Currents from Larger Tips or a Tip Array

For some applications in microscopy and lithography, it would be very beneficial to have the brightness of the Schottky emitter combined with a larger total current. In order to reach the high fields that are necessary to create the Schottky effect, flat cathodes cannot be used. However, larger tip radii in combination with larger cathode potentials can lead to higher emission currents and beam currents up to 2 microamps were measured [29].

A feasibility study for an array of Schottky tips [30, 31] to be used in parallel electron beam lithography did not yet lead to a practical concept.

6.5.3.3 Monochromators for Electron Microscopy

In high-resolution electron microscopy, both (S)TEM and SEM, chromatic aberrations sometimes become limiting. Also in electron energy loss spectroscopy, there is a demand for higher resolution spectra. This has led to the introduction of monochromators, basically consisting of an energy dispersive element and a slit. Some of the designs employ a second dispersive element after the selection slit in order to compensate for the dispersion of the first element. Many different concepts for the dispersive element have been developed. When using a single dispersive element, the slit size has to be matched to the size of the beam, but the concept leads to simple implementations. A very weak Wien filter suffices [32], but a stronger Wienfilter can also be used [33]. Another option is to use the chromatic aberration of the gun lens [34], which led to a design that improves the resolution of low voltage scanning electron microscopes. The Wien filter can be used in a double configuration as well [35]. When using a dispersive element that bends the beam over large angles, the combination of two elements can get the beam back on axis. The shape of the trajectory usually gives the name to the monochromator: Omega type or Mandolin type [36], or Alpha type [37]. Many electron microscopy companies now offer products with the

option of a gun monochromator. Depending on the application, energy resolutions of 15–100 meV are typical.

6.5.3.4 Splitting the Beam for Multibeam SEM

The current in a small probe is limited by the brightness of the source. For high-throughput applications in microscopy, the use of multiple probes can enhance the imaging speed by orders of magnitude. This is useful for faster imaging of noise-free images, for large area imaging, and for 3D imaging. A Schottky source has a total "usable" current in the order of several microamps, while the current in an SEM beam is usually in the order of nanoamps. Thus, several hundred beams can be formed from a single cathode. In order to avoid loss of brightness by stochastic Coulomb interactions, it is advantageous to split the beam as close to the cathode as possible. The aperture array that splits the beam can also be used as a lens array, such that the different beams are focused separately [31, 38]. Several multibeam scanning electron microscopes have been developed [39, 40].

6.5.3.5 Outlook

The Schottky emitter is presently the most widely used emitter in electron microscopy because of the combination of its high brightness, reasonable energy spread, sufficient angular current density for extracting many nanoamps of current, and its stability in operation if operated with care. Different materials have been tested, but so far none have shown the stability of ZrO on tungsten. For lithography applications, it would be beneficial if the total current could be higher [41]. In theory, the brightness could also be higher, so in practice, it should be expected that it can be pushed further as well.

6.6 Field-Emission Sources

6.6.1 Emission Theory

By applying an electric field of very high field strength at the surface of a metal, electrons are emitted even without heating the metal: cold field emission [42]. In cold field emission, electrons tunnel straight from the Fermi level through the potential barrier into free space.

Just as in Schottky emission, the high field strength is realized by using a nanometer-size pointed emitter opposite to an extraction electrode. For these sharp

tips, the field enhancement factor β is assumed inversely proportional to the tip radius:

$$F = \beta V_E = \frac{V_E}{kr} \tag{6.49}$$

where k is a factor depending on the exact configuration, in the order of 1–5.

The current density equation is derived in what is known as the "Fowler–Nordheim theory", for a model cathode that is flat. For extremely sharp tips, deviations occur which we shall not treat here.

$$j_{CF} = \frac{e^3 k^2 F^2}{8\pi h \varphi t^2(y)} \exp\left(-\frac{8\pi\varphi^{3/2}\sqrt{2m}}{3ehF}\upsilon(y)\right) \cdot \frac{\pi p}{\sin \pi p} \tag{6.50}$$

with

$$y = \Delta\varphi/\varphi \tag{6.51}$$

the relative decrease of the work function by the field, and $t(y)$ and $\upsilon(y)$ slowly varying functions:

$$t(y) = 1 + 0.11y^{1.33} \text{ and } \upsilon(y) = 1 - y^{1.69} \tag{6.52}$$

The parameter p is a measure for the current emitted above the Fermi level because of nonzero temperature.

$$p = kT/\mathrm{d} \tag{6.53}$$

d is a kind of measure for internal energy, when φ and d are given in eV:

$$\mathrm{d} = \frac{heF}{4\pi t(y)\sqrt{2m\varphi}} \approx 9.76 \times 10^{-11} \frac{F}{\varphi^{1/2}} \tag{6.54}$$

A plot of $\ln(\mathrm{I}/\mathrm{V_E^2})$ as a function of $1/\mathrm{V_E}$, called the Fowler–Nordheim plot (see Fig. 6.16), yields approximately a straight line in the zero-temperature equation. For a given work function, it yields the size of the emitting area $A = I/j$ and the field enhancement factor $\beta = \mathrm{F}/\mathrm{V_E}$. The energy distribution is given by

$$\mathrm{d}j_{CF}(E) = \frac{4\pi me^3}{h^3}\mathrm{d}^2 \exp\left(-\frac{2\upsilon(y)\varphi}{3t(y)\mathrm{d}}\right) \cdot \frac{\exp(E/\mathrm{d})}{1 + \exp(E/kT)}\mathrm{d}E \tag{6.55}$$

The practical brightness is given by

$$B_r = 1.44\frac{j_{CF}}{\pi \mathrm{d}} \tag{6.56}$$

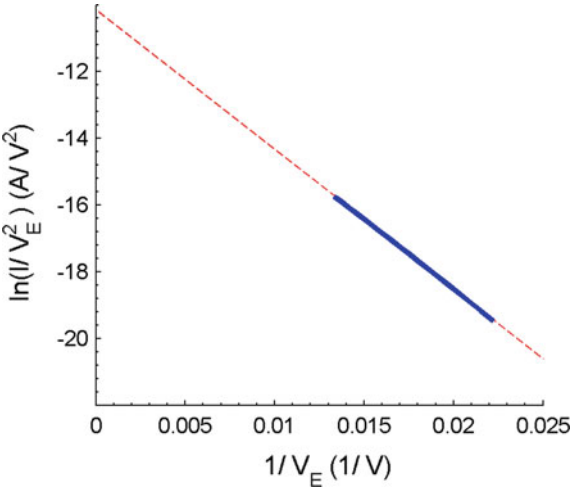


Fig. 6.16 Fowler Nordheim plot; see [1], courtesy of Springer

6.6.2 Practical Aspects

Although tungsten has a high work function (approximately 4.5 eV), tungsten is chosen for its rigidity, high melting point (3655 K), and good etching characteristics. The high melting point is important for the cleaning procedure, which consists of heating the tip to high temperature (2500 K) for a short period to vaporize all contaminants. This is known as flashing the tip. Since the tip of the field-emission source (FEG) is spot-welded to a standard microscope filament, this temperature is simply achieved by passing a heating current through the filament.

The presence of even single-molecule contaminants may critically affect the shape and the work function of the tip, so the FEG must operate in a very clean high vacuum, in the range 10^{-8} Pa. The general rule is that the emitter is more stable if the vacuum is better. Single crystals produce better-defined beams than multicrystalline tips. In some cases, it is advantageous to operate a field emitter at temperatures of about 1200 K, because this prevents excessive contamination. In fact, an equilibrium state will exist between deposition and evaporation of contaminants. Because of this process and the higher surface mobility of the atoms at the tip, the short-term stability (1 min) of the heated tip is worse, but the long-term stability (hours) is better. This has to be compared with a continuous decrease of emission from a cold emitter (10–20%/hour), which makes frequent flashing necessary.

A very small effective source diameter of about 2 nm is obtained due to the small emitting area (typical tip diameter is 5–15 nm) and the lens effect of the curved emitting area. The small source size makes the FEG quite sensitive to stray fields and vibration. These may cause the effective diameter to be larger and thus force the use of a larger demagnification. This gives the same effect as a lower brightness. The total current from the tip can rise to 100–500 μA, but in practice, it is limited to about 20–40 μA in order to restrict the energy spread in the beam. There are fewer data on energy

broadening by the Boersch effect than for Schottky emitters, so as an approximation we shall use the same equations that proved to be valid for the larger tips. This result in energy broadening is plotted in Fig. 6.17. Trajectory displacement in the section between the emitter and the extractor has not been measured or calculated in the same detail as for Schottky emitters.

Using an analytical model one study [43] finds that statistical Coulomb interactions limit the reduced brightness of even atomically sharp cold field emitters to 10^{11} A/m^2 sr V and regular tungsten cold field emitters with tip radius in the order of 20 nm to around 5×10^8 A/m^2 sr V, see Fig. 6.18. From experimental values quoted for the current in a coherent area or in the diffraction limited probe of (S)TEMs equipped with cold field-emission sources, we conclude that reduced brightness values of 2×10^8 A/m^2 sr V are routinely obtained.

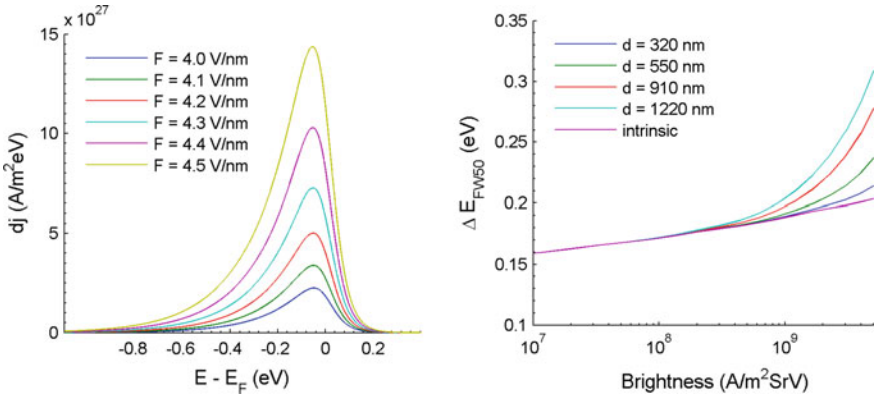


Fig. 6.17 Energy spectrum and energy spread (including estimate of Boersch effect) versus brightness for a cold field emitter

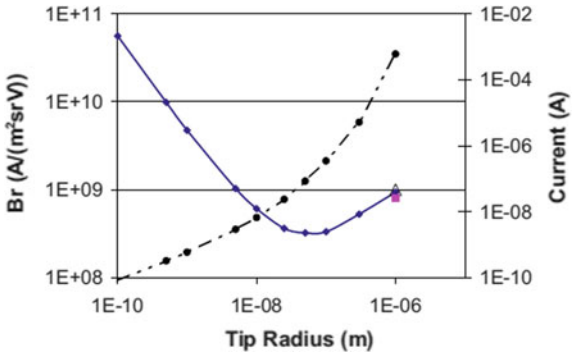


Fig. 6.18 Theoretical brightness values (diamonds) and currents (dots) at which stochastic Coulomb interactions near the tip have reduced the intrinsic brightness by 10%. At higher emission currents, the brightness increases slightly and then decreases again; compare [43]

6.6.3 Recent Developments

6.6.3.1 Increasing the Stability

Traditionally, a tip needed to be cleaned at least once per day, because the emission showed a continuous decrease of 10–20%/h, but recent advances seem to find an operational mode that is stable for longer times [44].

The small source size makes the FEG quite sensitive to stray fields and vibration. These may cause the effective diameter to be larger and thus force the use of a larger demagnification. This has the same effect as a lower brightness. When more than a few nanoamperes are drawn from a field-emission tip, there is a danger that gun lens aberrations will dominate the performance of the source, which can be minimized by careful design of the gun lens [45].

6.6.3.2 Monochromators for Electron Microscopy

Most monochromators have been developed for Schottky type emitters, with the exception of the alpha-type high-resolution filter [37]. The fringe field monochromator has also been used for a cold field emitter [46].

6.6.3.3 New Materials

Besides pure tungsten emitters in different orientations, many FEG cathodes have been tested such as: oxygen-processed tungsten tips, carbon and carbon-coated tips, tantalum carbide tips, semiconductor tips and especially carbon nanotubes (CNT's).

Soon after the discovery of CNT's work started on using these as electron sources [47–49]. One advantage expected of the CNT is that the surface is very stable and rather insensitive to contamination. At slightly elevated temperatures, they emit in the order of 200 nA with only about 1% variations [50]. In practice, there are many different CNT types and some are more stable than others. In the angular emission patterns, one can often see how individual atoms attach and disappear again, causing large intensity changes in one particular direction. These patterns also demonstrate that the current that can be selected in an apertured beam can come from a single emission site (one atom or a pentagon of atoms), indicating the potential of the extremely high brightness of the source. In fact, brightness is not a good characterizing parameter anymore for fully coherent beams [10]: The intrinsic brightness of these tips is potentially infinite because the image of the virtual source is always smaller than other contributions to the probe size (diffraction, trajectory displacement, gun lens aberrations). The great danger, however, is that the virtual source size increases and the brightness decreases because of stochastic Coulomb interactions in the beam as soon as the emission current exceeds a value in the order of nanoamps. Thus, so-called "experimental" values of brightness where only the angular emission current

was measured and then combined with a theoretical virtual source size, or a virtual source size measured at lower emission current, should be ignored (e.g., in [51] B = 9×10^{11} A/m^2 sr V at 5 μA is reported based on such "measurement"). Correctly measured experimental values for the brightness of CNTs or carbon nanocones of 3×10^9 A/m^2 sr V [50] and 1.6×10^9 A/m^2 sr V [52] have been reported.

Metal tips with protrusions of only 3 or even 1 atom were pioneered in the early 90s [53], and efforts to create such "single atom tips" have recently increased because of the success of the helium ion source based on a single atom tip [54].

For TEM, the beams from such sources are fully coherent. Efforts to use these innovations in a commercial machine continue [52] at the time of writing. In the past, they have failed either due to insufficient current for analytical applications, or fabrication issues, or unexpected instabilities, or limited lifetime, or any other practical aspect.

6.6.3.4 Attempts at Multibeam Lithography

High-throughput electron beam lithography would benefit greatly from the high brightness properties of cold field emitters. The limited total current per emitter, however, forces the designer of such a system to look toward using an array of field emitters. Several groups have created arrays of carbon nanofibers [55, 56]. To write a pattern, the emission of the field emitter tips needs to be controlled individually for which gate electrodes are incorporated in the fabrication. In order to deal with the temporal and spatial variations in emission, it was envisaged that the emission from each tip would be measured and the measurement be used to switch the beam [57]. An alternative direction in stabilizing and controlling the emission from field emitter tips is to employ an array of *p*-type silicon field emitters which respond to illumination by light [58]. Although the emission becomes extremely stable, this is at the cost of a large variation in tip potential and thus electron energy. At the moment of writing, all projects in this direction have been terminated.

6.6.3.5 Outlook

From a theoretical point of view, a field-emission tip of a nanometer-size tip is the best possible cathode for high-resolution applications. In practice, these tips are hard to make in a fully reproducible manner and then to operate unsupervised for a long time. This explains the efforts to find a better material than tungsten, or a better fabrication process then etching a tungsten single-crystal wire. Thus far, this has not let to a new kind of cold field emitter tip in commercial instruments, but it could well happen in the next 10 years. At the same time, it would be very useful to check the theoretical predictions about Coulomb interaction limited brightness experimentally and then find configurations that limit these interactions as much as possible.

6.7 Photoemission Sources

6.7.1 General

The photoemission effect can be used to create a beam of electrons. However, the photons should have sufficient energy: at least the work function and preferably a little more in order to have any efficiency [59]. For unprepared surfaces, this means that the light must be far into the ultraviolet, for which there are no good continuous light sources. So usually people go to low work function surfaces or to pulsed emission. Low work function surfaces are obtained, for instance, by covering a semiconductor with a few monolayers of cesium, but it is very hard to keep the surface under the right condition. In Schottky's sources, the work function is lowered by the Schottky effect, which also allows single photon photoemission. For continuous emission, however, the heating is too severe to make a high brightness source [60].

In pulsed lasers, it is possible to create sufficiently high-energy photons by frequency doubling or tripling, or to create such high intensity on the cathode that multi-photon absorption occurs.

6.7.2 Recent Developments

6.7.2.1 Ultrafast Electron Microscopy

For electron microscopy, the interest in pulsed emission has recently increased. Researchers are building up "ultrafast electron microscopes" to study effects at the nanoscale and at the sub-picosecond time scale [61]. Presently many of the sources are flat cathodes with tens of micron-sized laser beams, thus suffering from space charge limitations to the brightness. In accelerator physics, these limitations are overcome by pulsed extraction voltage or by wakefield acceleration [62] in front of the cathode and this may lead to new sources for ultrafast electron microscopy although the large energy spread in the beams is a problem. For higher brightness, it is also possible to go to pointed cathodes, just as in DC sources [63, 64]. When hitting a tip with a femtosecond laser, all kinds of interesting processes occur: field enhancement from the E-field of the laser light, multi-photon absorption, combined photon absorption, field emission, etc. For these high power pulses, it is not necessary anymore to use ultraviolet light, although long-wavelength pulses carry a larger risk of giving an increased energy spread.

6.7.2.2 Multibeam Lithography

Around the year 2000, there were several attempts at creating photocathodes for use in multibeam lithography [65, 66], but a combination of intrinsically limited brightness and instability of emission prevented this from being successful.

6.7.2.3 Outlook

Traditional flat photocathodes cannot obtain the brightness that Schottky emitters or field emitters can deliver because of space charge limited emission. For continuous operation, there is an additional problem in getting the emission stable because a low work function surface is very sensitive to contamination. Thus, it is unlikely that photocathodes will become a source for high-resolution applications in regular microscopy or lithography. However, pulsed photoemission sources have obtained a solid position in ultrafast electron microscopy. For pulses with more than one electron per pulse, it is unlikely that the instantaneous brightness will be larger than the brightness from a cold field emitter because of the stochastic Coulomb interactions [43], but for ultrashort pulses with less than one electron per pulse, the brightness could be extremely high.

6.8 Conclusions

For high-resolution applications in electron microscopy, the Schottky emitter is dominant, followed by the tungsten cold field emitter. This situation has not changed in the past 20 years, notwithstanding the research efforts in field-emission science. For ultrafast electron microscopy, exciting new developments are taking place which are leading to a new generation of pulsed electron sources.

For single-beam electron lithography, LaB_6 cathodes and Schottky emitters have kept their position, but for multibeam lithography a new application is found for the "old" CRT source, leading to new insights in emission from space charge limited thermal cathodes.

References

1. P. Kruit, Electron Sources, in Chapter 1 of *Transmission Electron Microscopy, Diffraction, Imaging and Spectroscopy*, ed. by C.B. Carter, D.B. Williams (Springer, 2016), pp. 1–15
2. H. Kohl, L. Reimer, *Transmission Electron Microscopy, Physics of Image Formation* (Springer, 2008)
3. L. Reimer, *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis* (Springer 1998)
4. P. Hawkes, E. Kasper, *Principles of Electron Optics,* vol. 2 (Academic Press, 1989)
5. M. Bronsgeest, J. Barth, L. Swanson, P. Kruit, Probe current, probe size, and the practical brightness for probe forming systems. J. Vac. Sci. Technol., B **26**, 949–955 (2008)
6. P. Kruit, S. Steenbrink, R. Jager et al., Optimum dose for shot noise limited CD uniformity in electron-beam lithography. J. Vac. Sci. Technol., B **22**, 2948–2955 (2004)
7. A.M. Zarubin, Three-dimensional generalization of the Van Cittert-Zernike theorem to wave and particle scattering. Opt. Commun. **100**, 491–507 (1993)
8. J. Spence, *High-Resolution Electron Microscopy,* 4th edn. (Oxford University Press, 2013)
9. G.H. Jansen, Coulomb interactions in particle beams. Adv. Electron. Electron Physics. (Suppl 21). Academic Press, New York (1990)

10. P. Kruit, G.H. Jansen, Space charge and statistical coulomb effects, in *Handbook of Charged Particle Optics,* ed. by J. Orloff (CRC Press, 2009), pp. 341–389
11. J.E. Barth, P. Kruit, Addition of different contributions to the charged particle probe size. Optik **101**, 101–109 (1996)
12. P. Kruit, M. Bezuijen, J.E. Barth, Source brightness and useful beam current of carbon nanotubes and other very small emitters. J. Appl. Phys. **99**, 024315 (2006)
13. V. Katsap, Thermionic cathode for use in electron emission apparatus, comprises crystalline or sintered emitter having upper section and sides, and carbon coating applied directly to outer surface of sides. Patent Number: US9165737-B2 (2015)
14. S. Golladay, H. Pfeiffer, C. Bohnenkamp et al., PREVAIL-EPL alpha tool: Early results. J. Vac. Sci. Technol. **B19**, 2459–2467 (2001)
15. J. Liddle, M. Blakey, K. Bolan et al., Space-charge effects in projection electron-beam lithography: results from the SCALPEL proof-of-lithography system. J. Vac. Sci. Technol. **B19**, 476–481 (2001)
16. G.H. Jansen, J. Appl. Phys. **84**, 4549 (1998)
17. D. Moonen, S. van Kranen, P. Kruit, V. Katsap, W.K. Waskiewicz, J. Vac. Sci. Technol., B **18**, 3111 (2000)
18. E. Platzgummer, Proc. SPIE **7637**, 763703 (2010)
19. E. Slot et al., Proc. SPIE **6921**, 69211P (2008)
20. A. van den Brom, A. van Veen, W.M. Weeda, G. Berglund, M. Wieland, P. Kruit, Cathode ray tube type electron gun as a source for multibeam electron lithography. J. Vac. Sci. Technol., B **25**, 2245 (2007)
21. M.D. Nijkerk, P. Kruit, Influence of surface roughness on space charge limited emission. Appl. Surf. Sci. **233**, 172–179 (2004)
22. L. Swanson, G. Schwind, in *Handbook of Charged Particle Optics,* ed. by J. Orloff (CRC Press, 2009)
23. M.S. Bronsgeest, Physics of Schottky electron sources. Ph.D. thesis, Delft University of Technology (2009)
24. L. Swanson, G. Schwind, S. Kellogg et al., Computer modeling of the Schottky electron source. J. Vac. Sci. Technol., B **30**, 06F603 (2012)
25. M. Bronsgeest, Physics of Schottky electron sources (Pan Stanford Publishing, 2014). ISBN 978-981-4364-79-9
26. M. Bronsgeest, J. Barth, G. Schwind, L. Swanson, P. Kruit, Extracting the Boersch effect contribution from experimental energy spread measurements for Schottky electron emitters. J. Vac. Sci. Technol., B **25**(6), 2049–2054 (2007)
27. A.H.V. van Veen, C.W. Hagen, J.E. Barth, P. Kruit, Reduced brightness of the ZrO/W Schottky electron emitter. J. Vac. Sci. Technol., B **19**, 2038 (2001)
28. P.C. Tiemeijer et al., Ultramicroscopy **114**, 72–81 (2012)
29. S. Fujita, H. Shimoyama, Mechanism of surface-tension reduction by electric-field application: shape changes in single-crystal field emitters under thermal-field treatment. Phys. Rev. B **75**, 235431 (2007)
30. A.K. Dokania, J.F. Velthuis, Y. Zhang et al., Thermal model of miniaturized Schottky emitter for parallel electron beam lithography. J. Vac. Sci. Technol. B **25,** 504–507 (2007)
31. Y. Zhang, P. Kruit, High brightness 100-electron-beam source for high-resolution applications. J. Vac. Sci. Technol. B **25,** 2239 (2007)
32. H.W. Mook, P. Kruit, Nucl. Instrum. Methods A **427**, 109 (1999)
33. P.C. Tiemeijer, Ultramicroscopy **78**, 53 (1999)
34. A. Henstra, J. Chmelik, T. Dingle et al., Microsc. Micro-Anal. **15**(Supplement 2), 168–169 (2009)
35. M. Mukai, E. Okunishi, M. Ashino et al., Development of a monochromator for aberration-corrected scanning transmission electron microscopy. Microscopy **64**, 151–158 (2015)
36. E. Essers, G. Benner, T. Mandler, S. Meyer, D. Mitmann, M. Schnell, R. Hoschen, Energy resolution of an omega-type monochromator and imaging properties of the MANDOLINE filter. Ultramicroscopy **110**, 971–998 (2010)

37. O. Krivanek, J. Ursin, N. Bacon et al., High-energy-resolution monochromator for aberration-corrected scanning transmission electron microscopy/electron energy-loss spectroscopy. Philosophical Transact. Roy. Soc. A—Math. Phys. Eng. Sci. **367,** 3683–3697 (2009)
38. Y. Zhang, P. Kruit, Design of a high brightness multi-electron-beam source. Phys. Procedia **1**(1), 553–563 (2008)
39. A. Mohammadi-Gheidari, C.W. Hagen, P. Kruit, J. Vac. Sci. Technol., B **28**, 1071 (2010)
40. A. Eberle, S. Mikula, R. Schalek, J. Lichtman, M. Knothe Tate, D. Zeidler, High-resolution, high-throughput imaging with a multibeam scanning electron microscope. J. Microsc. **259,** 114 (2015)
41. S. Fujita, T. Wells, W. Ushio, H. Sato, M. El-Gomati, Enhanced angular current intensity from Schottky emitters. J. Microsc. **239**, 215–222 (2010)
42. G. Fursey, *Field Emission in Vacuum Microelectronics* (Kluwer Academic, 2005)
43. B. Cook, T. Verduin, C. Hagen, P. Kruit, Brightness limitations of cold field emitters caused by Coulomb interactions. J. Vac. Sci. Technol. B **28,** C6C74–C6C79 (2010)
44. B. Cho, S. Kokubo, O. Chuhei, W (310) cold-field emission characteristics reflecting the vacuum states of an extreme high vacuum electron gun. Rev. Sci. Inst. **84**(1), 013305 (2013)
45. K. Kasuya, T. Kawasaki, N. Moriya, M. Arai, T. Furutsu, Magnetic field superimposed cold field emission gun under extreme-high vacuum. J. Vac. Sci. Technol., B **32**(3), 013802 (2014)
46. H.W. Mook, P. Kruit, Construction and characterization of the fringe field monochromator for a field emission gun. Ultramicroscopy **81**, 129–139 (2000)
47. W.A. de Heer, A. Chatelein, D. Ugarte, A Carbon nanotube Field emission electron source. Science **270**, 1179–1180 (1995)
48. M. Fransen, T. van Rooy, P. Kruit, Field emission energy distributions from individual multiwalled carbon nanotubes. Appl. Surf. Sci. **146**, 312–327 (1999)
49. M. Fransen, T. van Rooy, P. Tiemeijer, M. Overwijk, J. Faber, P. Kruit, On the electron-optical properties of the ZrO/W Schottky electron emitter. Adv. Imaging Electron. Phys. **111**, 91–166 (1999)
50. N. de Jonge, Brightness of carbon nanotube electron sources. J. Appl. Phys. **95**(2), 673–681 (2004)
51. A. Pascale-Hamri, S. Perisanu, A. Derouet, C. Journet, P. Vincent, A. Ayari, S. Purcell, Ultrashort single-wall carbon nanotubes reveal field-emission coulomb blockade and highest electron-source brightness. Phys. Rev. Lett. **112**, 126805 (2014)
52. F. Houdellier, L. de Knoop, C. Gatel et al., Development of TEM and SEM high brightness electron guns using cold-field emission from a carbon nanotip. Ultramicroscopy **151**, 107–115 (2015)
53. H.W. Fink, Phys. Scr. **38,** 260 (1988)
54. E. Rokuta, H. Murata, H. Shimoyama, C. Oshima, Fabrication of a single-atom electron source by noble-metal surface diffusion. J. Vac. Sci. Technol., B **31**, 02B105 (2013)
55. K. Teo, M. Chhowalla, G. Amaratunga et al., Fabrication and electrical characteristics of carbon nanotube-based microcathodes for use in a parallel electron-beam lithography system. J. Vac. Sci. Technol., B **21**, 693–697 (2003)
56. M. Guillorn, X. Yang, A.V. Melechko, D.K. Hensley et al., Vertically aligned carbon nanofiber-based field emission electron sources with an integrated focusing electrode. J. Vac. Sci. Technol., B **22**, 35 (2004)
57. S. Eliza, S. Islam, T. Rahman et al., A precision dose control circuit for maskless e-beam lithography with massively parallel vertically aligned Carbon nanofibers. IEEE Trans. Instrum. Meas. **60**, 1132–1140 (2011)
58. T. Teepen, A. van Veen, H. van't Spijker et al., Fabrication and characterization of p-type silicon field-emitter arrays for lithography. J. Vac. Sci. Technol. B **23,** 359–369 (2005)
59. K.L. Jensen, Electron emission physics. Adv. Imaging Electron. Phys. **149**. Elsevier Academic Press (2007)
60. B. Cook, M. Bronsgeest, K. Hagen et al., Improving the energy spread and brightness of thermal-field (Schottky) emitters with PHAST-Photo Assisted Schottky Tip. Ultramicroscopy **109**, 403–412 (2009)

61. A.H. Zewail, Four-dimensional electron microscopy. Science **328**(5975), 187–193 (2010)
62. S. Mangles, C. Murphy, Z. Najmudin et al., Monoenergetic beams of relativistic electrons from intense laser-plasma interactions. Nature **431**, 535–538 (2004)
63. P. Hommelhoff, Y. Sortais, A. Aghajani-Talesh, M. Kasevich, Field emission tip as a nanometer source of free electron femtosecond pulses. PRL **96**, 077401 (2006)
64. D.-S. Yang, O. Mohammed, A. Zewail, Scanning ultrafast electron microscopy, in *Proceedings National Academy of Sciences of U.S.A.*, vol. 107 (2010), pp. 14993–14998
65. J. Schneider, P. Sen, D. Pickard, G. Winograd, M. McCord et al., Patterned negative electron affinity photocathodes for maskless electron beam lithography. J. Vac. Sci. Technol., B **16**, 3192 (1998)
66. P. Arcuni, S. Presley, V. Aebi, W.E. Spicer, Recent tests of negative electron affinity photocathodes as source for electron lithography and microscopy. J. Vac. Sci. Technol., B **19**, 2585 (2001)

Chapter 7
Photocathodes

Wei Liu, Matt Poelker, John Smedley, and Romain Ganter

Abstract Photocathodes were at the heart of quantum physics development in the early twentieth century and are still used in high technology devices like night-vision goggles, photomultiplier tubes, and particle accelerators. The simplest form of the photocathode is a metallic surface, negatively biased, from which electrons are extracted by energetic photons. After a brief description of the photocathode preparation, this chapter will present the latest developments and limitations of metallic and semiconductor photocathodes.

7.1 Introduction to Photoemission

7.1.1 One-Step and Three-Step Models of Photoemission

Photoemission has a long and storied history in physics, beginning with Einstein's explanation of the photoelectric effect in 1905. Today photoemission spectroscopy,

W. Liu
Brookhaven National Laboratory, Upton, NY, USA
e-mail: wliu@bnl.gov

Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou 730000, China

University of Chinese Academy of Sciences, Beijing 100049, China

M. Poelker
Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA
e-mail: poelker@jlab.org

J. Smedley (✉)
Los Alamos National Laboratory, Los Alamos, NM 87545, USA
e-mail: smedley@lanl.gov

R. Ganter
Paul Scherrer Institut, 5232 Villigen, Switzerland
e-mail: romain.ganter@psi.ch

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*, Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_7

both in the UV and X-ray regime (UPS/XPS), is used to characterize materials, determine chemical bonding, and measure electronic band structures of materials. Two prevailing theoretical frameworks exist for describing photoemission, the one-step quantum mechanical model [1], and the three-step model [1, 2] described in detail below. The one-step model relies on a "sudden" approximation, where electrons are excited from a bound state within a material to a free state outside of it, in a transition that can be described by a quantum mechanical matrix element. This model is widely used in UPS analysis to understand material band structure; it considers electrons to conserve both energy and momentum upon emission. While powerful, this model is in general not widely applicable to practical photocathodes, as it generally assumes well defined crystalline initial states, and ignores electron scattering within the material.

The Three-Step Model was developed to deal with photoemission in practical photocathodes, and had found great success in predicting the behavior of both metals and semiconductors in the near-threshold regime that dominates most applications. In this model, the process is broken into:

1. absorption and excitation
2. transport to surface
3. escape over or through the surface barrier

Step 1 accounts for the loss of incident light due to reflection, and the excitation of electrons from filled states in the material valence band to excited states in the material conduction band. Most treatments of the model neglect momentum as a conserved quantity, reasoning that phonon scattering and the polycrystalline initial state will randomize the electron momentum state while leaving its energy largely unaffected. The electronic Density of States (DOS) is considered, so conservation of energy is accounted for. In this way, the model accounts for unproductive absorption (photons that excite electrons to conduction band states with energy below the vacuum level, so that they cannot escape).

Step 2 accounts for loss due to the various scattering processes which occur during the transport to the surface. Here the photon absorption length is compared to the electron mean free path, to determine the fraction of electrons lost to scattering. For metals, the dominant scattering mechanism is electron-electron scattering, and a single scattering event is considered to be a loss mechanism. For semiconductors, as will be discussed in detail later, electron-electron scattering is often forbidden and other scattering mechanisms take over.

Step 3 accounts for the direction of travel within the material as the electron approaches the surface. Although we often think of the work function as a limit on the energy required to escape the material, it is actually a limit on required momentum perpendicular to the material surface. Thus, the electron direction of travel must be within a cone defined by the energy in excess of the minimum energy needed to escape. For metals, failure to meet this criterion is considered a loss mechanism, as the electron will reflect off the surface and be lost to subsequent e-e scattering. For

semiconductors, a Monte Carlo treatment is required, as the electrons may have their direction of travel randomized again via phonon scattering.

Each step represents a probability of photon/electron survival. The product of these probabilities, taking angular considerations into account where appropriate, yields a energy distribution curve (EDC) that represents the probability per unit range of total-energy (measured relative to the top of the barrier) that an incident photon of specified wavelength will produce (at the cathode surface) an electron that will escape from the surface. The integration of this EDC over all energies above the barrier energy level (also called the "threshold level") yields the quantum efficiency (QE), that is, the ratio of the number of emitted electrons emitted to the number of incident photons.

7.1.2 Comparison Between Metal and Semiconductor Photoemission Processes Based on the Three-Step Model

Metal photocathodes are used for some accelerator applications that need only low average current, especially where prompt response time is required. However, most new accelerator applications and virtually all photodetector applications use semiconductor cathodes. The reason for this is that semiconductor cathodes offer far superior quantum efficiency performance, and can be sensitive to visible or infrared radiation. The photoemission mechanism for both metals and semiconductors can be described by the three-step model outlined above [3]. However, there are two fundamental differences—the extent of unproductive absorption and the type of collisions that photoexcited electrons experience during their travel to the surface.

In metals, photons can, in general, be absorbed throughout the conduction band, resulting in a large number of photons being "used" to excite electrons to states that lack sufficient energy to escape. In addition, photoexcited electrons may collide with electrons from the conduction band. In such scattering, the photoexcited electrons lose energy such that they can no longer escape (Fig. 7.1) and thus will not contribute to the QE. In consequence, only electrons within one electron mean free path $\lambda_{e\text{-}e}$ below the surface can possibly escape. In metals, the value of $\lambda_{e\text{-}e}$ is much smaller than the laser penetration depth so only a small fraction of the photoexcited electrons can be emitted (the violet area in Fig. 7.1) making the QE of metals low.

In semiconductors, the bandgap performs two functions. First, it limits unproductive absorption, as no electrons are excited into the bandgap due to the absence of a final allowed state. The lack of states in the bandgap also has a consequence on scattering—electron-electron collisions are not allowed if the final energy of electrons falls into this gap (Fig. 7.1). As a consequence, $\lambda_{e\text{-}e}$ can be even larger than the laser penetration depth and all photoexcited electrons could potentially reach the surface. This makes the QE of semiconductors much larger than the QE of metals

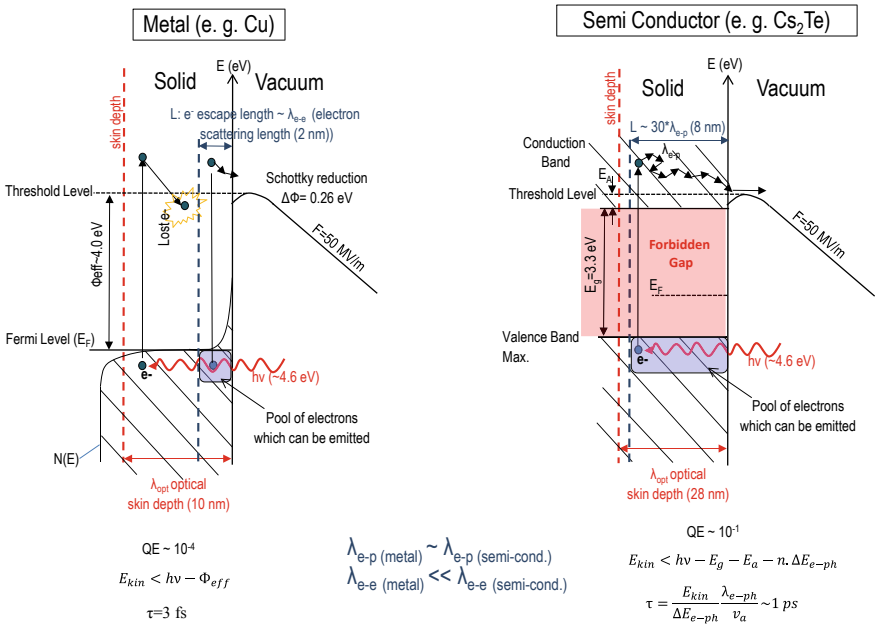


Fig. 7.1 The three-step model of photoemission can be used both for metals and semiconductors; here the parameters of copper and Cs_2Te are compared. $\lambda_{e\text{-}e}$ and $\lambda_{e\text{-}p}$ correspond to the electron mean free path between electron-electron and electron-phonon scattering events; τ is the electron extraction time (colored figure). Note that, even if an electron has total-energy E above the threshold level, it cannot escape unless the energy associated with its motion normal to the surface is also above the threshold level

[4], since the reservoir of electrons eligible for photoemission is much larger than for metals. For example, the forbidden gap of Cs_2Te photocathodes is 3.3 eV wide. Only the electrons excited above an energy of 6.6 eV will be able to scatter with an electron from the valence band. With laser light at 266 nm (i.e., a photon energy of 4.6 eV), photoexcited electrons will not interact with the other electrons from the conduction band during their travel to the surface. Other scattering processes, such as grain boundary scattering and scattering with lattice phonons still occur; however, electrons in these interactions tend to lose only a small fraction of their energy [5]. Scattering of electrons with the lattice can even be beneficial, as the interaction randomizes the electron momentum with minimal loss of energy. This allows electrons that are initially moving away from the surface to be redirected, potentially allowing emission. In this way, semiconductor cathodes can have a QE over 50% (cesium iodide, for example). Interaction with the lattice can also cool the electrons, which is beneficial for certain accelerator photoinjector applications, as it reduces the total initial kinetic energy of the photon emitted electrons (E_{kin}) and thus also the intrinsic emittance.

7.1.3 Importance of the Density of States

The electronic density of states (DOS), $N(E)$, is a function that describes the number of available electron states per unit range in total electron energy E, per unit volume of material. True densities of states for niobium (Nb) and lead (Pb) are shown in Fig. 7.2. In the simplest model, the so-called "free-electron metal", $N(E)$ is proportional to $(E - E_{CB})^{1/2}$, where E_{CB} is the energy of the conduction-band base. However, many theoretical implementations of the three-step model ignore the energy dependence in $N(E)$. Instead, metals are assumed to have an appropriate constant density of states, and semiconductors are assumed to have an appropriate constant density of states where states exist (i.e., not in the bandgap region). This simplifying assumption allows for closed QE calculations such as (7.2) in the next section, and produces

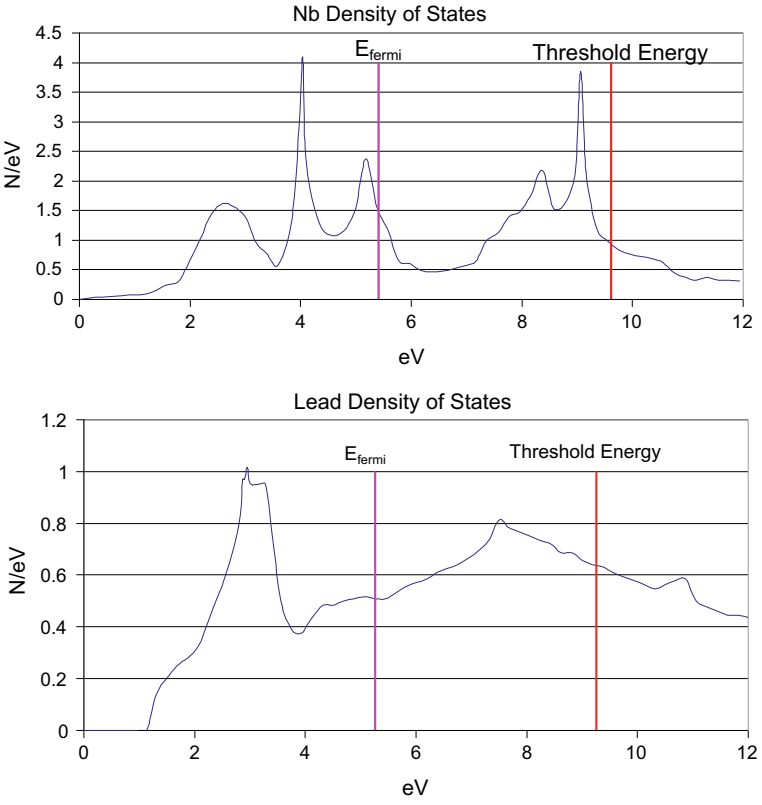


Fig. 7.2 Densities of states $N(E)$ for a transition metal (Nb, top) and a p-band emitter (Pb, bottom), as functions of total-energy E measured from the base of the conduction band. The separation of filled and empty states is labeled by the Fermi energy, and the minimum energy required to escape the material (in the absence of the Schottky effect) is labeled as the threshold energy [6, 7] (colored figure). This threshold energy is the sum of the Fermi energy and the work function

reasonable agreement with experimental data for some materials, especially metals that are "free electron"-like. These are typically materials in the first and second columns of the periodic table, and in the 13th and higher (materials that emit electrons from their s or p orbitals). It fails badly for many transition metals, such as niobium. Other transition metals, such as copper, behave as S-band emitters for near-threshold photons.

7.2 Metallic Photocathodes

7.2.1 Introduction

The metals discussed in this chapter are the most common metals that occur freely in nature. An important characterization parameter for a photocathode is its quantum efficiency (QE), and the goal of many studies is to increase the QE of cathodes. In practice, the QE is not the only parameter to look at, because the QE value obtained under the ideal conditions of a laboratory might not be reproducible in all photocathode applications. Indeed, many other constraints like the work function, the wavelength of incident photons, the applied electric field amplitude, and the mechanical assembly where the cathode sits will determine the type of cathode material to choose and consequently the QE that can be expected. In the following sections, we will review the most important parameters that determine the performance of a photocathode material.

7.2.2 Practical Approach to Photoemission from Metals

For a given illumination wavelength, the QE depends on three fundamental parameters: the photocathode reflectivity, the material work function, and the applied electric field—see (7.2). Metallic photocathodes are often used in applications where the cathode surface has to be exposed to air. Metallic surfaces will not lose much QE after brief exposure to air, although long exposure to humid air can be problematic. Among metallic photocathodes, a well-known and studied example is the copper photocathode. The following paragraphs examine in more detail the behavior of copper photocathodes.

Since photoemission is a process taking place in the first tens of nanometers below the surface, it highly depends on the quality of the surface which, unfortunately, is changing with time. For example, the work function of a material depends on the presence of contaminants, but these contaminants might also migrate across the surface over time. Surface contamination (e.g., oxidation) changes the reflectivity of the surface and thus the QE. The electric field applied on the cathode surface will be enhanced depending on the roughness of the surface, which can change with

photocathode use. This electric field enhancement will also affect the QE locally, through a field-induced reduction in height of the potential-energy (PE) barrier that prevents electron escape into vacuum. This reduction is due to the Schottky effect. At any particular surface location, one can define an "effective local work function" by

$$\Phi_{eff} = \Phi_0 - \sqrt{\frac{e^3 F}{4\pi\varepsilon_0}} \tag{7.1}$$

where Φ_0 is the conventional (zero-field) local work function, e the elementary charge, ε_0 the electric constant, and F is the (positive) magnitude of the (physically negative) local electric field normal to the surface. The schematic of Fig. 7.3 illustrates this link between the fundamental parameters driving the QE and practical aspects of photocathodes.

The time-varying nature of photocathode QE can be problematic, and sometimes photocathode QE even increases [8], as illustrated in Fig. 7.4, where the photocathode QE increased by more than an order of magnitude after 4 days of operation. This was probably due to cleaning/removal of the contaminants by the illuminating laser. Such a QE variation corresponds, in theory, to only 0.06 eV drop in the surface work function.

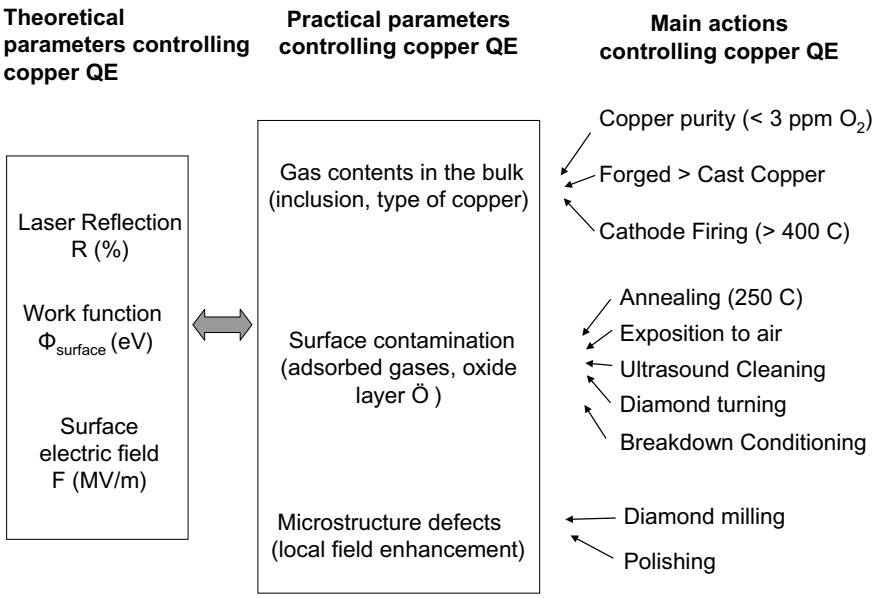


Fig. 7.3 Schematic representation of the link between the theoretical parameters controlling the QE of copper and macroscopic properties as well as ways to act on these properties

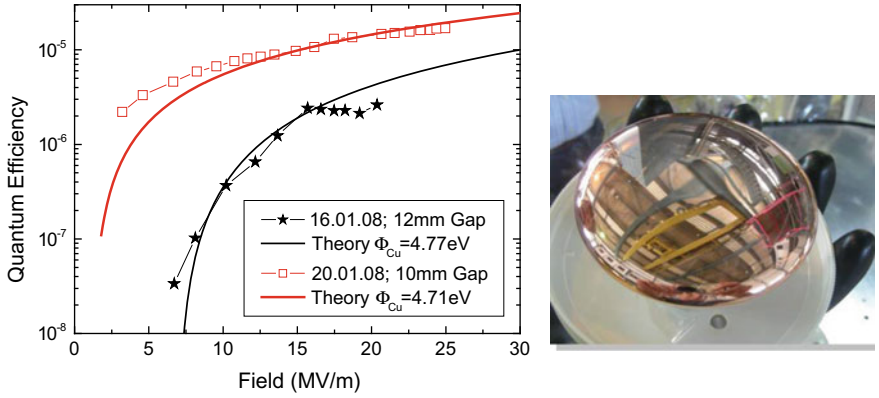


Fig. 7.4 QE of a copper photocathode versus the applied electric field during 4 days of operation. Laser pulses at 266 nm with 6 μJ energy per pulse, at 10 Hz repetition rate, were used to illuminate the copper photocathode with quasi-normal incidence. The copper photocathode (inset) was diamond turned, providing an average roughness of a few nanometers. The theoretical curve corresponds to the QE expression (7.2) taken from [9]

Based on the three-step model, and the assumptions that the zero-temperature (step-like) Fermi-Dirac distribution may be used and that (near the threshold total-energy for escape) the DOS may be treated as constant, Dowell et al. [9] derived a simple expression [their (19)] for the QE of metals

$$\mathrm{QE}(\omega) = \frac{1 - R(\omega)}{1 + \frac{\lambda_{\mathrm{opt}}}{2\lambda_{\mathrm{e\text{-}e}}(E_{\mathrm{m}})} \frac{\hbar\omega\sqrt{\Phi_{\mathrm{eff}}}}{E_{\mathrm{m}}^{3/2}} \left(1 + \sqrt{\frac{\Phi_{\mathrm{eff}}}{\hbar\omega}}\right)} \frac{(E_{\mathrm{f}} + \hbar\omega)}{2\hbar\omega} \times \left[1 + \frac{E_{\mathrm{f}} + \Phi_{\mathrm{eff}}}{E_{\mathrm{f}} + \hbar\omega} - 2\sqrt{\frac{E_{\mathrm{f}} + \Phi_{\mathrm{eff}}}{E_{\mathrm{f}} + \hbar\omega}}\right], \tag{7.2}$$

where R denotes the reflectivity, λ_{opt} is the laser penetration depth, $\lambda_{\mathrm{e\text{-}e}}$ is the mean free path of electrons between two consecutive electron-electron scattering events, $\hbar\omega$ is the photon energy, Φ_{eff} is the effective work function already reduced by the Schottky effect (see (7.1)), E_{f} is the Fermi energy (that is, the total-energy difference $(E_{\mathrm{F}} - E_{\mathrm{CB}})$ between the Fermi level (E_{F}) and the base of the conduction band (E_{CB})), and E_{m} is the total-energy difference $(E_{\mathrm{m}} = E_s - E_{\mathrm{F}})$ between the total-energy (E_{s}) at which $\lambda_{\mathrm{e\text{-}e}}$ has been determined and the Fermi level. Expression (7.2) accurately reproduces the experimental data for copper.

In the measurements shown in Fig. 7.4, it is interesting to note that increasing the electric field from 10 to 20 MV/m results in a QE increase by a factor 10. Such a doubling of the electric field can easily happen locally, due to mechanically protruding defects that enhance the electric field (tip effect) on a microscopic scale. This illustrates the importance of the Schottky effect.

Any change of the surface state will modify the three fundamental parameters (reflection, work function, and local electric field) at the same time. For example, changes in the method of polishing the surface will change the microstructure profile and thus the local electric field, but will also change the reflectivity of the surface, and might also change the density of contaminants (which in turns modifies the local work function).

7.2.3 QE Performances of Metals

Because of the technical difficulty of having atomically clean samples, free of contaminants, the QE of a polycrystalline metal can easily vary from one sample to another. In Table 7.1, typical values of QE measured for different transition metals are presented. These values were obtained by laser illumination at different wavelengths on polycrystalline metals. Most of Table 7.1 comes from Suberlucq [10] but is augmented with some recent results. The work function data was compiled by Michaelson [11]. These values correspond to the most reasonable work function values for polycrystalline metals used in practice: the so-called "preferred work function". Since the work function depends on the crystal orientation distribution, the vacuum quality, and the surface cleanliness, one can reasonably expect a variation of 0.1–0.2 eV from one sample to another. This is the reason why QE values differ from one measurement to another in Table 7.1.

If no special in-vacuum treatment is applied to a metallic photocathode surface, the QE will be dominated by the contamination on the surface. The reaction of the cathode surface with air leads to surface chemistry (e.g., oxidation) that will increase the work function. The QE can then become much smaller than the values listed in Table 7.1. Figure 7.5 illustrates QE measurements taken in a diode-geometry gun where the cathode was illuminated with laser pulses at 266 nm. The metal samples were exposed to air for several weeks before being installed in the diode gun. Under such conditions, the QE always stayed below 10^{-4} and the QE values are very different from those of Table 7.1, pointing to the reactivity of the metal with ambient air.

For an atomically clean surface, the crystal orientation can change the work function by several tenths of an eV, which leads to significant variations in the QE. For different crystal orientations of copper, Table 7.2 gives examples of the work function variation; values extend from 4.48 to 4.94 eV, which is a range of 0.46 eV. The dependence of the QE on the work function is plotted in Fig. 7.6, which corresponds to (7.2) for the case of copper illuminated by 253 nm photons. Taking into account the Schottky reduction (~0.3 eV), the effective work function ranges between 4.18 and 4.64 eV, which corresponds to a QE variation of one order of magnitude.

Figure 7.7 shows two micrographs of a polycrystalline copper photocathode from an RF photoinjector. The grain boundaries between different facet orientations are clearly visible and the cathode surface appears as a highly nonuniform surface (Figs. 7.8, 7.9, and 7.10). In practice, it is difficult to observe a spatial variation of QE

Table 7.1 Quantum efficiency of polycrystalline metallic photocathodes (*main source* [10])

λ (nm)	193	213/209	248	262–266	308	355	ϕ_S	References
E (eV)	6.4	5.8/5.9	5	4.7	4	3.5	eV	
Al		$8.4 * 10^{-4}$		$3.2 * 10^{-5}$		$3.4 * 10^{-7}$	4.3	[12, 13]
Ag[a]				$2 * 10^{-5}$			4.3	[14]
Ag[b, f]				$8.4 * 10^{-5}$			5.06	[15]
Au[a]				$4.7 * 10^{-5}$			5.1	[14]
Au		$4 * 10^{-4}$		1.310^{-5}			5.1	[12]
Ca			$4 * 10^{-5}$				2.9	[16]
Cu	$2.0 * 10^{-4}$	$1.5 * 10^{-4}$		$2.2 * 10^{-6}$	$1.6 * 10^{-7}$	$8 * 10^{-9}$	4.6	[12]
Cu[b]	$1.5 * 10^{-3}$	$4.2 * 10^{-4}$					4.6	[12]
Cu[a, f]				$1.4 * 10^{-4}$			4.6	[14, 17, 18]
Cu			$5 * 10^{-5}$					[19, 20]
Mg				$5.1 * 10^{-5}$			3.7	[12]
Mg				$2.7 * 10^{-4}$			3.7	[12]
Mg[c]				$5 * 10^{-4}$			3.7	[21]
Mg[b]				$1.7 * 10^{-3}$			3.7	[15]
Mo				$<7 * 10^{-7}$			4.6	[12, 13]
Mo[b]				$2.5 * 10^{-6}$			4.6	[15]
Nb	$4.5 * 10^{-4}$		$3.2 * 10^{-6}$				4.3	[22]
Nb				$3 * 10^{-6}$			4.3	[12]
Nb[b]				$1.9 * 10^{-4}$			4.3	[15]
Nb[b, f]				$2.6 * 10^{-4}$				[18]
Ni[a]				$2.5\ 10^{-5}$			5.2	[14]
Pb	$5.4 * 10^{-3}$	$2.7 * 10^{-3}$					3.88	[7]
Pb[b]				$2.2 * 10^{-4}$			3.88	[15]
Pd[a]				$1.2 * 10^{-5}$			5.1	[14]
Ac 316 LN		$9 * 10^{-5}$		$1.6 * 10^{-6}$			?	[12]
Sm					$1.6 * 10^{-6}$		2.7	[12]
Sm[a]				$7.3 * 10^{-4}$			2.7	[14]
Ta[a]				10^{-5}			4.3	[14]
Tb[a]				$2.3 * 10^{-4}$			3	[14]
Ti				$6 * 10^{-6}$			4.33	[13]
Ti[b]				$3.3 * 10^{-4}$				[15]
V[b, f]				10^{-5}				[15]
W (111)[e]						$2 * 10^{-5}$	4.5	[23]
WK^+ (b, e)					$1.2 * 10^{-5}$		2.8	[12]
Y				$5 * 10^{-4}$			3.1	[14]
Y				$2.7 * 10^{-6}$	$1.1 * 10^{-6}$		3.1	[12]
Y[b]				$1.8 * 10^{-4}$			3.1	[12]
Y[a]				$5 * 10^{-4}$			3.1	[14]
Zn[a]				$1.4 * 10^{-5}$			4.3	[14]
Zr[a]				10^{-5}			4.1	[14]
Zr[b]				$2.9 * 10^{-4}$				[15]

Φ_S = work function from [11]
[a]Surface preparation and activation under vacuum, from [14]
[b]Cleaning by argon ion bombardment
[c]ATF (BNL) measurements at 70 MV/m without surface treatment
[d]Photoemission assisted by a high electric field, E = 3 GV/m
[e]Potassium ions implanted in tungsten substrate (150 keV implantation energy, with an ion flux density of 1.3×10^{17} ions/cm^2 at the surface)
[f]Annealing procedure at 250 °C

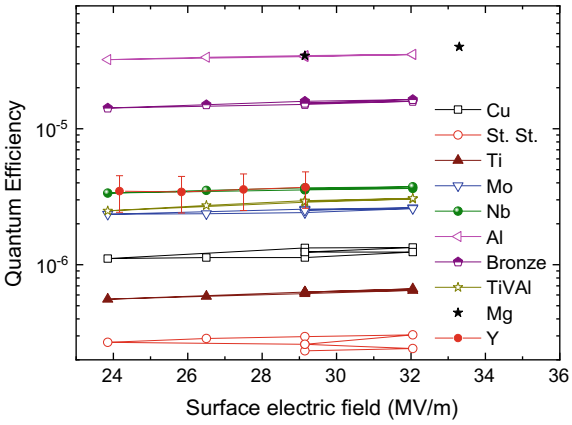


Fig. 7.5 Quantum efficiency of various metal photocathodes measured in a diode gun (350 kV and variable gap) using 6 ps rms laser pulses at 266 nm, and 10 Hz repetition rate. Cathode samples were exposed to air without any in-vacuum cleaning treatment [13]

Table 7.2 Work function for four crystal orientations of copper, from [11]

Crystal orientation	Work function (eV)
Cu Polycrystalline	4.65 ± 0.2
Cu (100)	4.59 ± 0.03
Cu (110)	4.48 ± 0.03
Cu (112)	4.53 ± 0.03
Cu (111)	4.94 ± 0.03

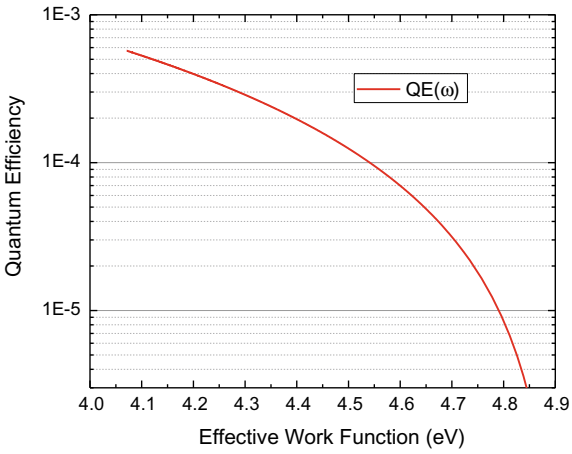


Fig. 7.6 QE variation of copper versus work function using light at 253 nm

on a polycrystalline metal surface (see, for example, the QE map of Fig. 7.11). But photoemission electron microscopy performed on a polycrystalline copper photocathode clearly revealed the difference of emission in the different grains [24]. A single-crystal photocathode would have the advantage of much better uniformity. Or if the typical grain size is larger than the illuminating spot size, one could also obtain

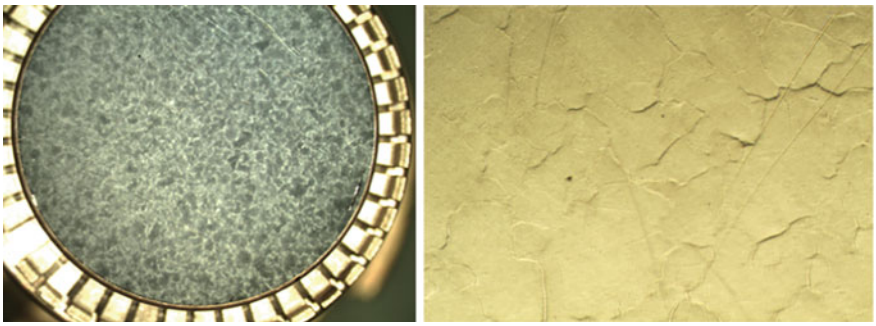


Fig. 7.7 Micrographs of a copper photocathode under white light illumination [17]

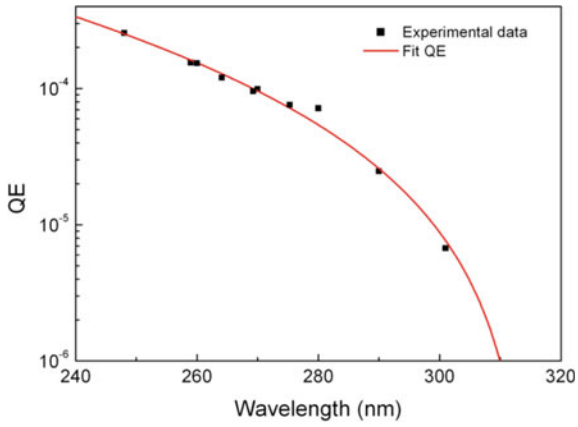


Fig. 7.8 Quantum efficiency of a copper photocathode versus the illumination wavelength, from [25]

single-crystal photoemission from a polycrystalline sample. In the RF photoinjector of the LCLS free electron laser at Stanford, a polycrystalline copper photocathode is used as an electron source. During the preparation of the cathode, the copper is fired to very high temperature (above 800 °C) such that the grain sizes grow nearly as large as the laser spot illuminating the cathode (mm). By picking a sufficiently large grain, photoemission is obtained from a single crystal [19].

As illustrated in Table 7.1, the QE also decreases with increasing photon wavelength, simply because fewer electrons can be excited to sufficiently high energy to escape (see Fig. 7.1). On the other hand, electrons emitted with an energy component normal to the surface that is close above the energy level of the top of the barrier have relatively low total kinetic energy immediately after emission; this can be beneficial for applications where the thermal agitation of extracted electrons (intrinsic emittance) is a concern. For example, in accelerators one goal is to produce electron bunches that occupy a very small volume in phase space. This is the case when all electrons follow parallel trajectories at cathode exit without random thermal agitation.

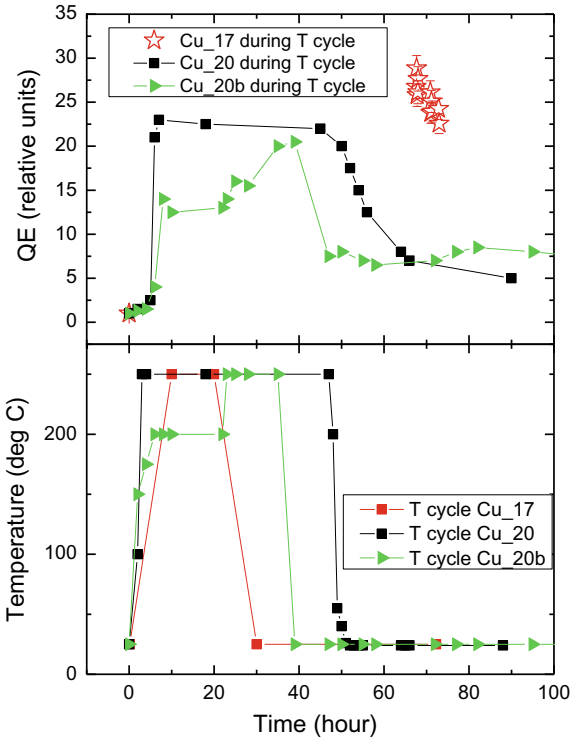


Fig. 7.9 Effect of annealing a copper photocathode: QE (top) and temperature cycle (bottom) versus time for three samples [28]

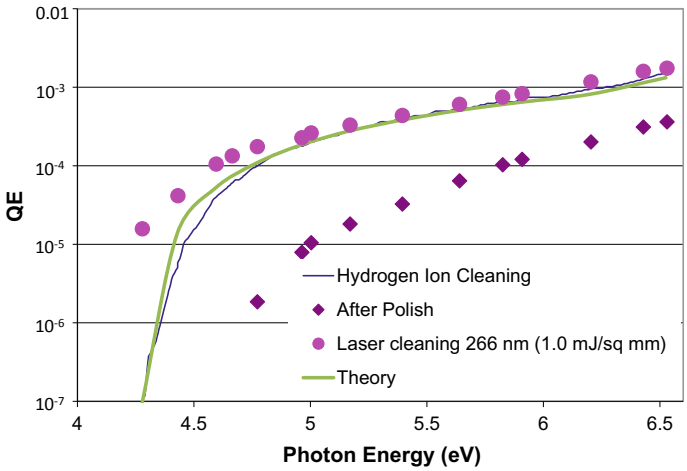


Fig. 7.10 Comparison of three-step model theory with LCLS Cu cathodes cleaned via laser cleaning and hydrogen ion cleaning. (Color figure)

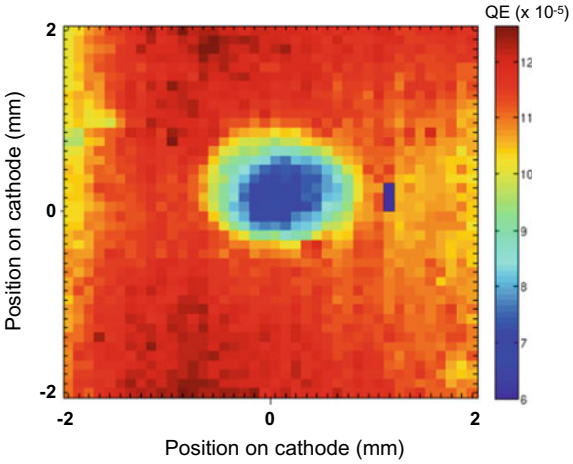


Fig. 7.11 QE map revealing the effect of laser illumination on QE degradation. Long term illumination has the effect of cracking organic molecules which then bind to the surface (*Source* [28]) (Color figure)

Figure 7.8 from Vicario et al. [25] illustrates quite well the dependence of the QE on the illuminating photon wavelength. To fit the data they used a similar relation to (7.2) for copper.

7.2.4 Limitations and Potential of Metallic Photocathodes

7.2.4.1 Cathode Preparation: Bulk Material Choice, Polishing, and Cleaning Techniques

As mentioned in the previous section, to obtain optimum behavior and high QE values, the preparation of the photocathode surface before and after installation in vacuum is extremely important. All metallic photocathodes produced from bulk material will have several layers of adsorbed contaminants on the surface. These contaminants come from the ambient air (H_2O, CO, CO_2, N_2, H_2) and from organic chemical compounds formed during the machining and manipulation of the photocathode.

Out-of-Vacuum Cleaning

A common cleaning technique for metals for ultra-high vacuum is hot ultrasonic washing with detergent followed by careful rinsing. Alkaline detergent cleaners like tetrapotassium pyrophosphate can efficiently clean metallic surfaces [26, 27]. In a second step, ultrasonic cleaning with solvents like acetone and alcohol is a good way to remove residual grease or organic films from the surface. The photocathode should then be dried with dry air and installed under vacuum as soon as possible.

For photocathodes contaminated with an oxide layer, more aggressive techniques are needed, like mechanical polishing (discussed below) or acid etching (nitric, sulfuric, hydrochloric, and phosphoric acid), which will remove several micrometers

of material. In the case of mechanical polishing, the ultrasonic cleaning processes mentioned above are required afterward. For both techniques, it is important to install the photocathode under vacuum as soon as possible. Indeed, after the oxide layer has been removed, the surface is very reactive and depending on the type of metal, an oxide layer can return very quickly.

In-Vacuum Cleaning

To remove molecules adsorbed on the surface, an efficient way is to heat the photocathode under vacuum. Most of the water, hydrogen, and carbon-based molecules will be desorbed from the surface if the annealing temperature exceeds 250 °C. Figure 7.9 illustrates the effect of annealing a copper photocathode at 250 °C for 10 h. For three different samples, the QE increased by a factor 5–30 after the thermal cycle [30]. Such an annealing procedure ensures an initial QE of ~$10^{-4,}$ which is close to the theoretical expectation for polycrystalline copper [17].

Once the photocathode surface has been cleaned under vacuum, the only contamination which would appear again must come from the residual gas present in the vacuum chamber, such as H_2, CH_4, H_2O, CO, O_2, N_2, Ar, and CO_2. Assuming all the gas atoms and molecules stick to the surface, it takes only one hour to cover the surface with one monolayer at a pressure of 10^{-9} Torr (about 10^{-9} hPa or 10^{-7} Pa) [29]. Luckily, the sticking coefficient of residual gasses such as CO and H_2O is much smaller than unity and the partial pressures of these gas species can be very low. But it explains why the QE of a photocathode can degrade over time by a large amount, as in Fig. 7.19. Under some conditions, the illuminating light can help to clean the surface, but under other conditions, the light can crack the molecules leaving highly reactive chemicals that strongly adhere to the surface as illustrated by the QE hole in Fig. 7.1, located exactly at the position of the laser. In any case, a regular annealing procedure after some duration of operation will desorb most of the accumulated contaminants and restore the QE to its initial value. Another important mechanism responsible for QE variation is ion back-bombardment. Residual gas atoms and molecules can be ionized by photoelectrons near the photocathode surface. These ions will be accelerated towards the photocathode, where they can be implanted into the material, thereby altering energy levels or reducing the electron diffusion length, or sputtering away chemicals applied to the surface that reduce the work function. This mechanism can be a major issue for DC high voltage photoguns (see Sect. 4.3.3).

Several other techniques exist to clean metallic photocathodes in vacuum. The so-called laser cleaning technique relies on a laser beam focused close to the ablation threshold, and then scanned across the cathode surface [20]. In order to avoid the creation of large craters on the surface, a careful monitoring of the vacuum pressure during laser cleaning is required [30]. Ozone cleaning can also successfully desorb molecules from metallic surfaces [31]. In general, the final QE of a clean metal surface is independent of the method used to achieve the surface. Figure 7.10 shows the QE of a LCLS copper cathode prepared via laser cleaning with a 266 nm laser (~30 ps pulse duration) with an energy density of 1 mJ/mm^2 compared to that of a

second cathode cleaned via Hydrogen ion cleaning [9]. The initial (uncleaned QE), and the three-step model theory prediction based on (7.2) are also shown.

Figure 7.11 is a QE map of a copper photocathode that was used in the electron gun of an accelerator. The QE map was obtained by focusing a UV laser to a small spot (below 100 μm) and by scanning this spot over the full cathode surface. One can clearly distinguish a round area where the QE is reduced. This area corresponds to the region which was exposed to the full laser beam during normal operation of the electron gun. Long UV laser exposure seems to induce surface chemistry which then lowers the QE. Some labs use laser ablation to clean the surface and retrieve the original QE of the bulk photocathode material. However, such an ablation process is delicate if one wants to preserve the highly polished surface [32].

In addition to the contaminants coming from the residual gas, the photocathode material itself can contain impurities. Metallic cathodes made of bulk materials can reach purity levels above 99.995%, with impurities like O_2, C, and P present within the material at a level of <10 ppm [33]. Metals also contain inclusions, or bubbles, which can degas once in vacuum. Metals that are forged or produced by hot isostatic pressure (HIP) methods are better for photocathode applications because they have larger grain sizes, and therefore, fewer inclusions and defects compared to cast metals.

Metallic photocathodes made of bulk material, in contrast to those produced by thin film deposition, need to be polished. Roughness of the photocathode surface must be controlled since it affects the quality of the emitted beam. In accelerator photoinjectors, the cathode surface roughness must be small to withstand high electric field and also to avoid beam emittance increase [34, 35]. Diamond turning of the surface using "state of the art" ultra-precision tooling provides surfaces with an average roughness (Ra) less than 5 nm. Figure 7.12 represents an interferometer micrograph of a copper surface which was diamond turned. The measured roughness was about 3 nm with peak-to-valley maxima up to 30 nm. The parallel diagonal lines seen in the background are from the machining tool. Diamond turning leaves large scale surface variations, or waviness, but still provides a better surface than manual hand polishing with diamond paste. Indeed, abrasive polishing with diamond grit has the drawback of leaving small inclusions in the metal, especially for soft metals like copper. The grain boundaries are also visible in Fig. 7.12 and have a typical size of a few hundred micrometers. Figure 7.13 shows a 5 by 5 μm region of an aluminum sample which was also diamond turned. The average roughness of this sample is about 1.3 nm, but the atomic force microscope reveals some local tips as large as 37 nm. Such tips can enhance the local electric field, lowering the effective work function, and introducing nonuniformity into the distribution (across the surface) of the local photoemission current density. These emission nonuniformities can be revealed by electron-emission-imaging of the cathode surface, that is to say, in a QE map.

Metal photocathodes can also be produced by growing thin films on a substrate. Pulsed laser depositions of thin films of magnesium and yttrium have been deposited onto a copper substrate [36]; these exhibited QE values similar to a bulk cathode. However, the uniformity and roughness of the obtained films remain difficult to

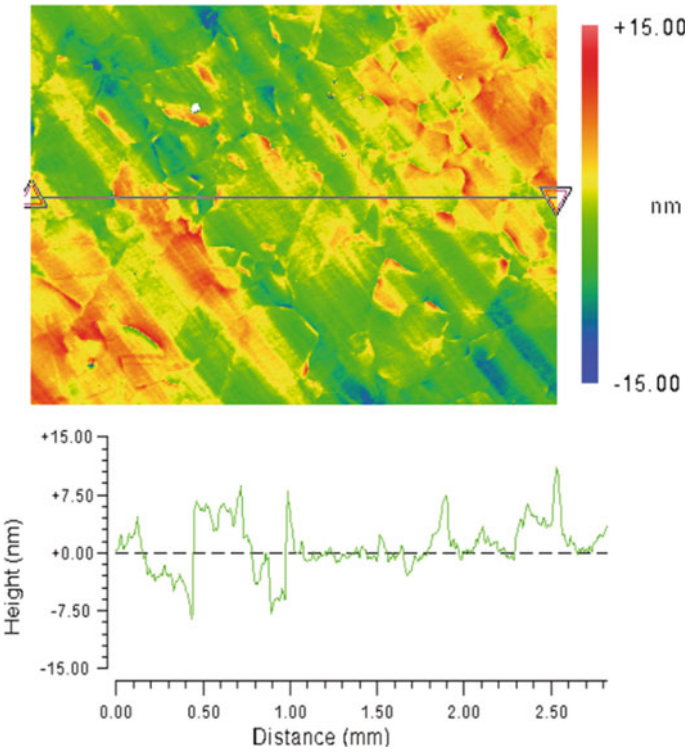


Fig. 7.12 Top: Surface profilometry of a diamond turned copper sample 2.7 by 1.5 mm (forged copper). Bottom: Transverse profile along the line shown in the top figure; grain size: 300 μm; roughness (Ra) ~3 nm; height (peak-to-valley) ~30 nm (*Source* [17]) (Color figure)

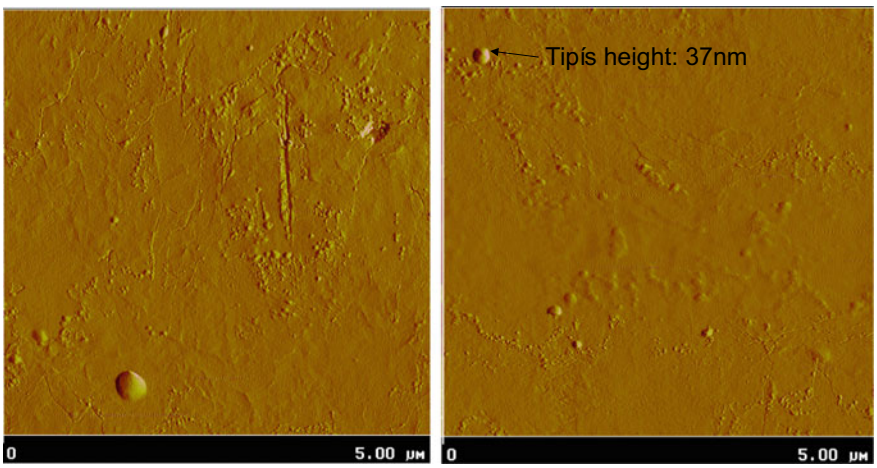


Fig. 7.13 Atomic force microscope picture of diamond turned aluminum samples. The rms roughness is equal to 1.2 nm for this sample despite some local tips as high as 37 nm

control. Magnetron sputtering of Cu and Nb thin films deposited onto a Si substrate have exhibited QE values similar to those obtained with bulk materials [18].

7.2.4.2 Incident Photons

The QE performance of a photocathode depends on three parameters related to the illuminating light, as presented in (7.2):

- wavelength, which determines the energy of the incoming photons;
- reflectivity, denoted by coefficient R;
- and the electric field component of the incident electromagnetic wave, which is called here the "optical E-field" (this is particularly important in the case of large incident power density and associated high optical E-field).

Because the electrostatic fields discussed in earlier sections and the optical E-fields discussed in this section may act differently on matter, and may act differently in photocathode physics, it is necessary to carefully distinguish between these two types of the electric field.

The dependence of the QE on incident photon energy is illustrated in Fig. 7.8 which shows QE increasing with the photon energy, because increasingly large numbers of electrons can be excited into states with total-energies sufficient to allow escape.

If the power density of the incident photon beam is sufficiently large (~1 W/m^2), for example, because the laser beam is strongly focused or when using ultra-short laser pulses, two or (more generally) n photons can be absorbed in a single process. In this way, photo-stimulated emission of electrons becomes possible for photon energies smaller than the work function. The most favored process corresponds to the lowest value of n for which $n\hbar\omega > \Phi_{\text{eff}}$ [38]. The dependence of the extracted current density (and hence the extracted charge per pulse) on the illuminating power will deviate from a linear behavior in the case of multi-photon processes. Multi-photon photoemission can even become more efficient than single photon emission when one considers the necessary laser power needed for frequency doubling or tripling.

The QE changes with the polarization of the incident light: the vectorial photo-electric effect [37]. When the angle of incidence is varied, photocathode QE can vary by factors of 2 or 3 for p-polarized light, whereas QE is almost constant for s-polarized light [38]. Evidence of QE enhancement using z-polarized laser light (i.e., optical E-field in the direction of propagation) has also been demonstrated [39]. In such a scheme, it is an electrostatic field applied to the cathode that is modified by the optical E-field.

More generally, if experimental arrangements are such that there is a component of the illuminating optical E-field normal to the photocathode surface, and this field component is strong enough to significantly lower the effective work function, then photoemission can be enhanced by various detailed mechanisms [40]. In particular, if the optical E-field has sufficiently high magnitude, then electrons can tunnel through

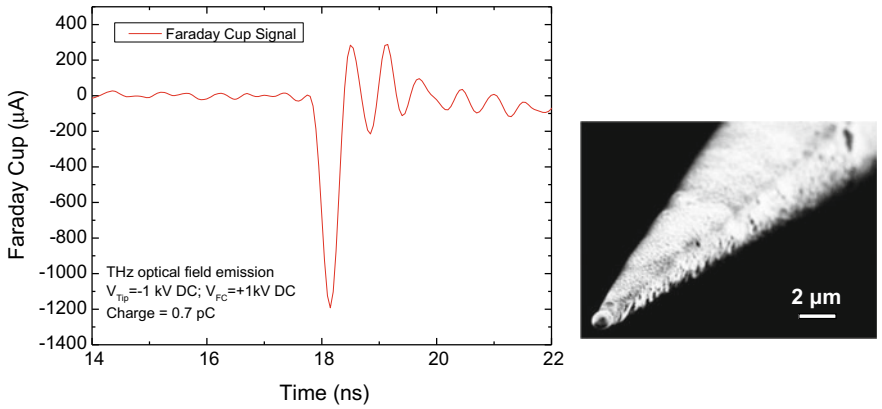


Fig. 7.14 A quasi half-cycle THz pulse is used to trigger optical field emission from a ZrC tip. The time structure of the current pulse is limited by the scope bandwidth, but should in principle follow the THz half-cycle

the rapidly oscillating surface barrier: this effect has been called "optical field emission", or alternatively "tunneling photoemission" [40]. The effect can be illustrated by experiments in which quasi half-cycle THz pulses were used to extract charge from a cathode surface [41]. In Fig. 7.14, a THz half-cycle with sufficiently high peak E-field (hundreds of MV/m [42]) triggered optical field emission of a pulse of electrons, with total charge close to 1 pC [C. Vicario].

7.2.4.3 Surface Electric Field

The local electric field component perpendicular to the cathode surface is an important parameter for the photoemission process, since (when physically negative) it lowers the PE barrier that electrons must overcome to be emitted. As noted above, this barrier lowering is known as the Schottky reduction, and is denoted by $\Delta\Phi$. The Schottky effect becomes significant for local fields F above a few tens of MV/m. Such fields lower the barrier by a few tenths of an eV, the precise reduction being given by (7.1).

Local electric fields can be higher than the applied electric field because of enhancement effects due to the morphological imperfections on the surface (see Fig. 7.13). In the extreme case where the local electric field has been enhanced by geometric effects to a value of magnitude above a few GV/m, field emission can take place through the reduced PE barrier. In this case, electrons from the metal conduction band can tunnel through the barrier.

There exists an intermediate state where the barrier still prevents strong field emission but is low enough that almost all incident photons lead to electron emission, that is to say when the QE is equal to 1 or more. This was observed when sharp tips

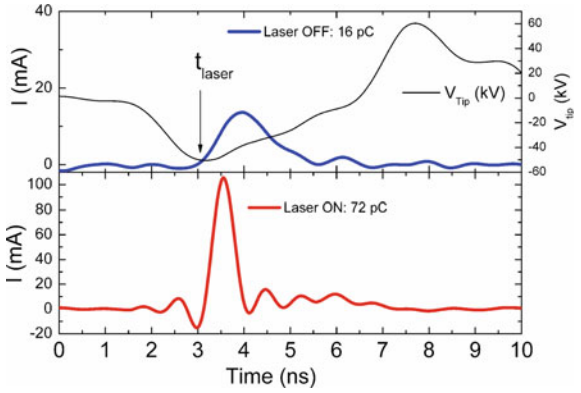


Fig. 7.15 Top: Emitted current pulse when only a nanosecond long voltage pulse is applied (pure field emission). Bottom: Current pulse when a voltage pulse (ns) and a laser pulse (ps) are simultaneously applied. The cathode is a ZrC tip [43]

were illuminated by UV laser pulses while high voltage was applied to the tip as illustrated in Fig. 7.15.

This so-called photo-assisted field emission or photo-field emission can be unstable since the local electric field at the apex of the tip is close to the field emission threshold, and thus close to the electric field beyond which electrical breakdown occurs in poor vacuum conditions. One way to prevent arcs is to use very short voltage pulses applied to the cathode, and short laser pulses synchronized to the field emission pulse, to trigger short pulses of photoemission. If the voltage pulses are short enough (nanosecond range), the heating effect related to the electron emission is limited and the risk of arcs is reduced. In the experiment of Fig. 7.15, voltage pulses of about 2 ns FWHM and 50 kV amplitude were applied to a ZrC tip having an apex radius of curvature of a few micrometers.

In photocathode systems, an adequately high applied electric field is also desirable in order to overcome space-charge screening effects near the cathode. The photocathode current can be limited by poor cathode quantum efficiency, but also by an extracting electric field that is too weak. The electric field under which the electron emission starts to saturate is equal to the beam surface charge density σ divided by the electric constant: $E_{SC} = \sigma/\varepsilon_0$. For example, for a short electron emission pulse of 6 ps duration, the maximum current that can be extracted from 1 mm^2 with a field of 50 MV/m is about 70 A.

7.2.4.4 Superconducting Photocathodes

A special class of metallic photocathodes worthy of consideration is superconductors. The motivation for using superconductors as photocathodes comes from the use of superconducting RF (SRF) cavities for electron beam generation, especially for high repetition rate/CW beam operation. These cavities have nearly zero heating of the cavity walls (no RF wall losses), but they have a host of complications, including the need to keep the cavity at ~2 K, immersed in a liquid helium bath. Putting a photocathode into an SRF injector is an engineering challenge, as the RF power can

couple into and heat any normal conducting material placed in the cavity, causing the temperature to rise and the entire cavity to quench (go normal conducting). While there exist tricks to overcome this limitation, primarily through the use of a quarter-wave RF choke joint and vacuum transfer systems to introduce a cathode, the simplest method of introducing a photocathode into an SRF gun is to use a material that is also superconducting.

SRF cavities are typically constructed of niobium, as it is an elemental superconductor available for bulk manufacture, and of all elemental superconductors, it has the highest critical temperature (T_c) and critical magnetic field (B_c), allowing cavities constructed of niobium to support the highest electric and magnetic field gradients. Naturally, the first superconducting photocathodes were niobium [44], but niobium is a transition metal with a poor DOS for photoemission (see Fig. 7.2).

Lead has also been studied as a photocathode option [45], with somewhat more success (see Fig. 7.16). Care must be taken to ensure that only lead is used in the cathode section of the injector, where the magnetic field is near zero—lead has a B_c-value one third that of niobium. However, lead is a *p*-band emitter (Fig. 7.2, bottom) and has a QE roughly an order of magnitude higher than niobium in the near-threshold region. Several methods of deposition of lead onto niobium have been investigated [7, 46], with vacuum arc deposition, followed by laser or plasma surface treatment, currently the preferred method.

A third superconducting photocathode option is being investigated, namely via the use of the proximity effect, by which a thin layer of a normal conducting material becomes a superconductor when it is grown on a superconducting substrate. This method is beyond the scope of this chapter.

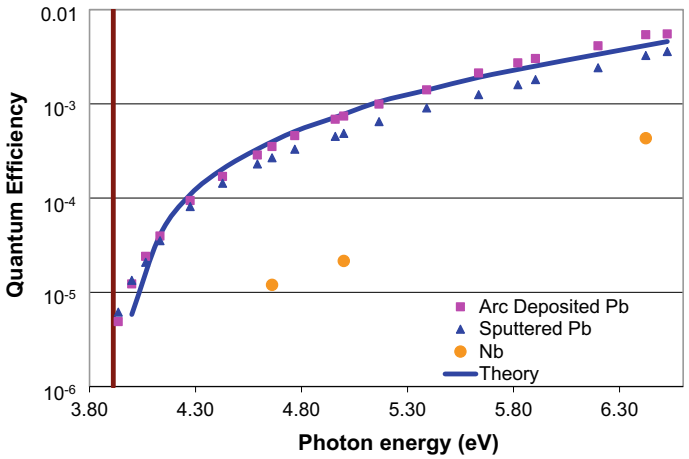


Fig. 7.16 QE versus photon energy for lead and niobium photocathodes. The vertical line represents the threshold energy for lead at zero electric field

7.3 Positive Electron Affinity Semiconductors

7.3.1 Introduction

For a semiconductor, the electron affinity ("EA" or χ) is defined as the difference ($E_{vac} - E_{CB}$) between the local vacuum level (E_{vac}) and the base (E_{CB}) of the conduction band. This Section discusses semiconductors for which χ is positive. Section 7.4 discusses semiconductors for which χ is negative.

Alkali metals are very reactive and do not occur freely in nature. These metals have only one electron in their outer shell. Therefore, they are ready to lose that one electron in ionic bonding with other elements. Tellurium is a metalloid (quasi-metal) that will react with alkali metals to form a very stable compound. Among the alkali photocathodes, Cs_2Te is one of the most widely used compounds because of its relatively long lifetime (up to years) and because it exhibits high QE compared to other semiconductor photocathodes.

There are more than 20 different semiconductor materials that have been developed especially for the photomultiplier technologies, in order to cover the largest bandwidth of the spectrum (Fig. 7.17). In this section we will focus on Cs_2Te photocathodes.

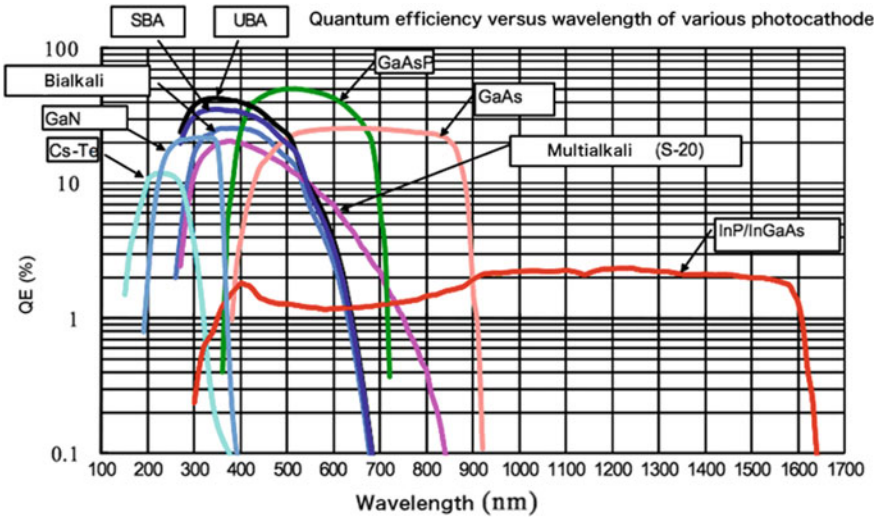


Fig. 7.17 Overview of semiconductor photocathodes—QE versus illumination wavelength (Courtesy of Hamamatsu [47])

7.3.2 *Preparation and Performance of Cs_2Te Photocathodes*

7.3.2.1 Preparation Methods

Semiconductor photocathodes are produced by the deposition of thin films onto a comparatively massive substrate. For Cs_2Te photocathodes, the thin films are usually obtained by thermal evaporation of tellurium and cesium [12]. The substrate (Al, Au, Cu, Mg, Mo, stainless steel, and others) influences the performance of the photocathode, since bulk material atoms can diffuse into the thin film. For example, Copper atoms will diffuse much deeper into tellurium than molybdenum atoms do [48]. As a consequence, better QE results and longer lifetime are obtained with molybdenum substrates.

The thickness of the evaporated Cs_2Te layer can vary from a few to hundreds of nanometers. The evaporation can be done in two steps, by first evaporating a layer of Te and then a layer of Cs. The Te layer is applied first because it sticks better to the substrate than Cs. The Cs will then react with the deposited Te atoms to form various compounds (Cs_2Te, Cs_2Te_5, Cs_3Te_2, Cs_2Te_3, Cs_4Te) [49]. This successive evaporation leads to typical QE values around a few percents [50] compared to nearly 20% when the Te and Cs are co-evaporated [51]. As the name implies, the co-evaporation technique consists of evaporating the Te and Cs simultaneously so that Cs_2Te molecules form in the gas phase before deposition onto the cathode surface. This technique requires careful monitoring of the QE together with control of the Cs flow to reach the best stoichiometry. Figure 7.18 illustrates the evolution of the cesium source evaporation current while monitoring the photocurrent (which here corresponds to the QE) and for a quasi-constant tellurium deposition rate of 0.03 Å/s. The Cs evaporation rate is constantly adjusted to maximize the photocurrent rise. The deposition is stopped when the photocurrent starts to decrease (Fig. 7.19).

The tellurium growth process onto the substrate follows a Stransky-Krastanov [52] scheme. First Te atoms get deposited layer by layer up to a few monolayers, and then isolated islands will start to form until the coalescence of these adsorbate "islands"

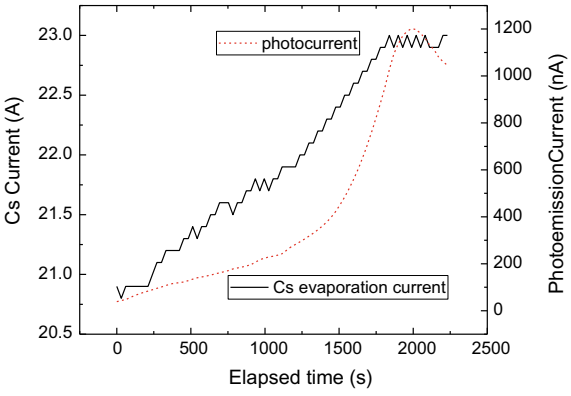


Fig. 7.18 Evolution of the Cs evaporation current (related to the deposition rate) and collected photocurrent under constant illumination at 266 nm and while Te is deposited at 0.03 Å/s. Pressure rises to 10^{-8} mbar (10^{-6} Pa)during deposition

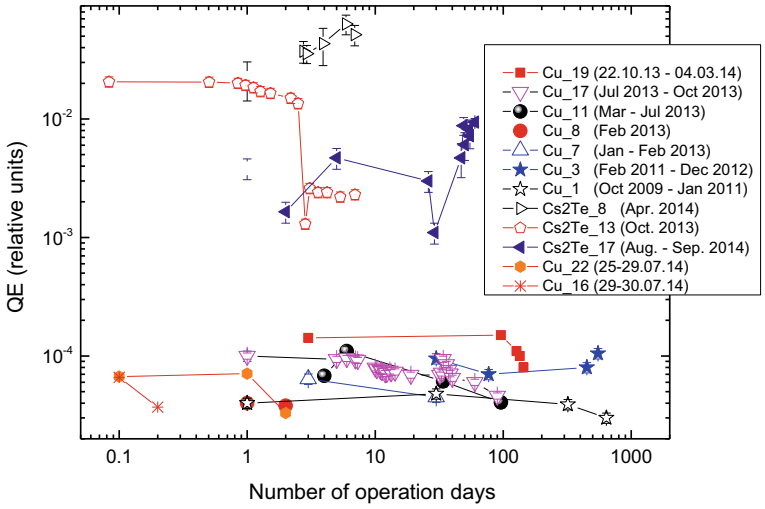


Fig. 7.19 Evolution of QE with time for various copper and Cs_2Te photocathodes operating under the same conditions: extracting field 52 MV/m, 262 nm laser pulses of 4 ps rms duration at 10 Hz repetition rate, with 10^{-9} mbar (10^{-7} Pa) photogun vacuum

takes place. This results in an almost smooth surface. The Cs growing process is a pure island type of growth where some nucleation sites grow independently. The overlap of Cs on a Te layer produces a layer with a complex granularity which is a mix of the two processes described above: island and quasi-smooth surface. The resultant photocathode surface roughness is definitively larger than the roughness of the initial substrate. In the case of the co-evaporation technique, the roughness should be smaller compared to the sequential technique [53]. This can have important consequences on the beam quality (uniformity, emittance).

Due to the very high reactivity of Cs with oxygen [48], the newly formed thin film of Cs_2Te cannot be exposed to air and has to be stored in vacuum at a pressure below 10^{-10} mbar (10^{-8} Pa). For this reason, the semiconductor photocathode has to be transported in a vacuum suitcase and inserted into its final location via a load-lock chamber. The design of the vacuum system to produce and transport semiconductor photocathodes is extremely important and is usually the main cost factor of the photocathode production system.

One way to improve the uniformity of the photocathode is to heat the sample to about 100–150 °C during the deposition [54]. It is generally assumed that the warm sample helps to spread the islands growing on the surface. And it also helps to keep the substrate clean, particularly free of water. Finally, the warm-deposition recipe provides a longer lifetime [55].

7.3.2.2 QE Values of Alkali Based Cathodes—Cs_2Te

The maximum QE of Cs_2Te photocathodes must be discussed together with lifetime, since at very high QE values (above 10% at 260 nm), the photocathode provides only a few hours of operation.

Table 7.3 summarizes QE values obtained with alkali metal cathodes tested under similar conditions at different laboratories. QE values as high as 20% were reached but exhibited an exponential decay time of only a few hours. After this initial decay, the decay is still exponential but with a much longer time constant. In most cases reported, Cs_2Te photocathodes could maintain QE above 1% for at least several months, depending on the level of vacuum and the amount of charge extracted.

Compared to typical QE values of metal photocathodes, even a QE of 0.1% represents a large value. If such a QE can be maintained over many months of operation, then it becomes an interesting candidate for many applications [67]. Figure 7.19 summarizes the time evolution of the QE for several copper and Cs_2Te photocathode samples operated in the same RF photoinjector and under the same experimental conditions. Both photocathodes undergo QE degradation, but on average the Cs_2Te photocathode has a QE one order of magnitude larger than copper for at least 100 days. For the sample Cs_2Te_{17}, the total amount of extracted charge was 3 mC, without a major sign of degradation.

7.3.3 Sensitivity to Vacuum

As mentioned above, the main drawback of semiconductor photocathodes as compared with metals is their high sensitivity to oxygen [48]. The poisoning of the photocathode surface can affect both the lifetime of the cathode and the uniformity of the photoemission. A common way to measure the emission uniformity of involves scanning a small laser spot over the entire emitting area while recording the QE at each position, to obtain a QE map. The QE maps of Cs_2Te photocathodes can illustrate nonuniformity as shown in Fig. 7.20. Such nonuniformities are not good for beam quality and might be a result of the deposition procedure.

Photocathode lifetime is always an important criterion for semiconductor photocathodes. The QE degradation comes mainly from O_2 and CO_2 [48, 68]. Figure 7.21 shows the typical lifetime of Cs_2Te photocathode s under the operating conditions of an RF photoinjector at the Free Electron Laser (FEL) FLASH in Germany. A photocathode lasts typically one to six months, depending on the operating pressure, and produces up to 10 μA continuous delivery, up to a total of 1 C in six months [69].

Table 7.3 Overview of QE of alkali metals measured at different laboratories under slightly different conditions based on the table from [56] and updated with some recent results

Material	Substrate	QE at λ	Lifetime	Field (MV/m)	Operating vacuum (mbar)	References
CsI-Ge		0.73% at 213 nm	$T_{1/\tau} > 1$ year	70		[12]
		0.13% at 266 nm				
		2% at 209 nm	$T_{1/\tau} > 150$ h		10^{-10}–10^{-9}	[57]
Cs_3Sb	Cu	7.5% at 532 nm	> 1% over 30 h	7	10^{-10}	[10, 51, 58]
	Cu	4.9% at 532 nm	> 1% over 30 h	85	10^{-10}	[58]
		4% at 527 nm	$T_{1/\tau} < 4$ h	20	10^{-10}–10^{-9}	[57]
K_3Sb		1.6% at 266 nm				[12]
		0.3% at 541 nm				[59]
		1-3% at 262 nm				[12]
K_2CsSb		8% at 527 nm	$T_{1/\tau} < 4$ h	20	10^{-10}–10^{-9}	[60]
		1.2% at 541 nm	$T_{1/\tau} < 24$ h			[59]
Cs_2Te	Cu–Au	16% at 262 nm	450 h above 1.5%	100	1–5.10^{-9}	[61]
	Cu–Au	2–8% at 262 nm	Few weeks > 1.5%	120	1–5.10^{-9}	[51]
	Cu	20% at 266 nm	300 h above 3%	120	10^{-9}	[51]
	Mo	20% at 262 nm	200 days above 4%	35–40	10^{-10}	[62]
		15.5% at 263 nm	1 month above 1%		$5 * 10^{-8}$	[63]
	Mo	16–18% at 251 nm	> 1% over 100 h	20–25	10^{-10}	[57]
		8–12% at 263 nm				
		15–20% at 266 nm			10^{-9}	[64]
		8–15% at 266 nm	> 1% over 420 days		10^{-9}	[65, 66]

Note 1 mbar = 100 Pa

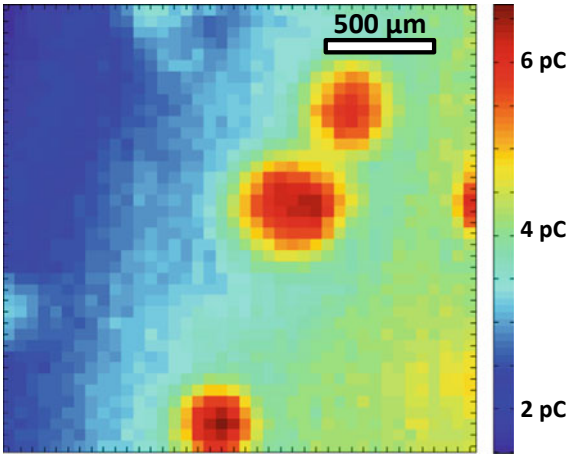


Fig. 7.20 QE map of a Cs_2Te photocathode measured inside the gun by scanning a small laser spot with constant energy

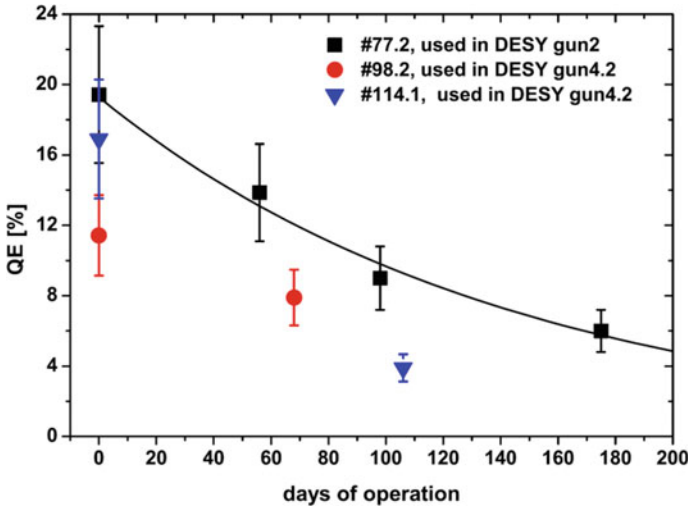


Fig. 7.21 Typical lifetime of Cs_2Te photocathodes used in an RF photoinjector: Illumination with trains of 2400 laser pulses of ps duration during an 800 µs macropulse at 10 Hz, 262 nm illumination, 40 MV/m, 700 pC/pulse (*Source* [62])

7.3.4 Alkali-Antimonide Photocathodes

Image intensifiers (for night-vision), and low-light-level detectors (photomultiplier tubes) are normally vacuum emissive devices. For such devices, the most common photocathodes come from the alkali-antimonide family. These materials typically

consist of three alkali atoms and one antinomy atom per molecule, and some photocathodes (such as the S-20) use multiple layers of differing composition. Commonly used photocathodes of this family include Cs_3Sb, K_2CsSb, Na_2KSb, and (Cs) Na_2KSb (S20), though others are used as well. For accelerators, this class of materials holds the average current record for both DC and RF photoinjectors [70, 71].

7.3.4.1 Preparation Methods and Film Properties

Several sources have been used to provide alkali materials for growth, including thermal evaporation from sealed sources, chromates, alkali azides, and pure alkali metal effusion sources. In most of these cases, antimony is provided via thermal evaporation, either from Sb pellets in a crucible or from Sb (or PtSb alloy) wire sources. Recently, sputter growth of C_3Sb and K_2CsSb has also been demonstrated [72]. Vacuum is critical in alkali-antimonide growth—these materials degrade quickly in a partial pressure of water over 10^{-8} Pa. However, the cathodes are relatively insensitive to most other gasses—alkali antimonide cathodes are routinely operated in gas electron multiplier-based detectors at ambient pressure of Ar and CH_4.

The structure of an alkali-antimonide film is a strong function of whether the growth is sequential (Sb-K-Cs, for example), or via co-deposition/sputtering. Traditional growth has involved sequential deposition of Sb, followed by the weakly bonded alkali, followed by the strongly bonded alkali (typically Cs). This results in a good QE, but the lattice mismatch between crystalline Sb [73] and the cubic alkali antimonides results in significant roughening of the surface (up to 25 nm RMS for a 50 nm thick film [74]).

Co-evaporation of Sb and the alkalis avoids this recrystallization, and results in a far smoother film [75]. It has also been shown to dramatically improve the lifetime of cathodes for detector applications [76]. It is, however, significantly more difficult to control. These methods have resulted in surfaces with RMS roughness under 1 nm. Co-evaporation also produces superior QE performance, as shown in Fig. 7.22.

Sputter growth produces films that are largely amorphous but very smooth (RMS ~0.5 nm for a 50 nm film). These films have lower QE (typically 16-20% peak, and 4% in the green, as compared to 35% and 8–10% for sequential and co-evaporated cathodes), but are easy to produce over large areas.

For many applications, the roughness of the cathode is not an important parameter, but for accelerators it has a dramatic impact on the source emittance and achievable brightness [77]. It is also important for detector applications that require high electric fields.

7.3.4.2 In Situ X-Ray Analysis During Growth

Much of the recent progress in alkali-antimonide growth is the result of new tools that have been developed to study thin film growth [78]. Using an X-ray beam from

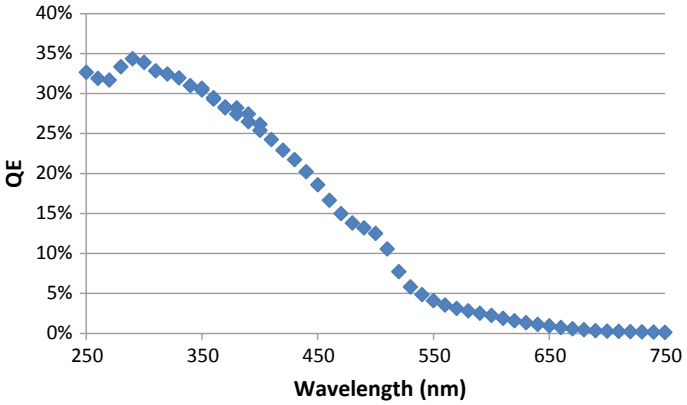


Fig. 7.22 Spectral response of a K_2CsSb photocathode grown by ternary co-evaporation with in situ X-ray fluorescent monitoring

a synchrotron (such as the National Synchrotron Light Source II at Brookhaven National Laboratory or the Cornell High Energy Synchrotron Source at Cornell University), the material crystal structure, stoichiometry and surface roughness can be evaluated in real-time during growth, even under UHV conditions. The critical techniques used for in situ analysis are X-ray diffraction XRD for determining crystal structure evolution, X-ray fluorescence for measuring stoichiometric evolution, and X-ray reflectivity for measurement of surface roughness. Figure 7.23 shows the XRD evolution of a cathode grown through traditional sequential deposition, along with the thickness (monitored via quartz crystal monitor) and the QE.

7.4 NEA Semiconductors: GaAs Based Photocathodes for Polarized Electron Beams

7.4.1 Overview of GaAs Photocathodes

Polarized electron beams play a critical role in nuclear and high energy physics research, including parity-violating electron scattering experiments and the measurement of nucleon spin structure functions [79, 80]. As recently demonstrated, spin-polarized electron beams can also be used for magnetization-sensitive imaging [81]. The first polarized electron source for an accelerator, based on photoionization of state-selected ^{6}Li atoms, was developed at Yale University in the early 1970s, for use at the Stanford Linear Accelerator (SLAC) [82]. Later, a polarized electron source based on the Fano effect in Rb was developed for the Bonn synchrotron [83]. Other polarized sources were developed or proposed during the 1970s, including an

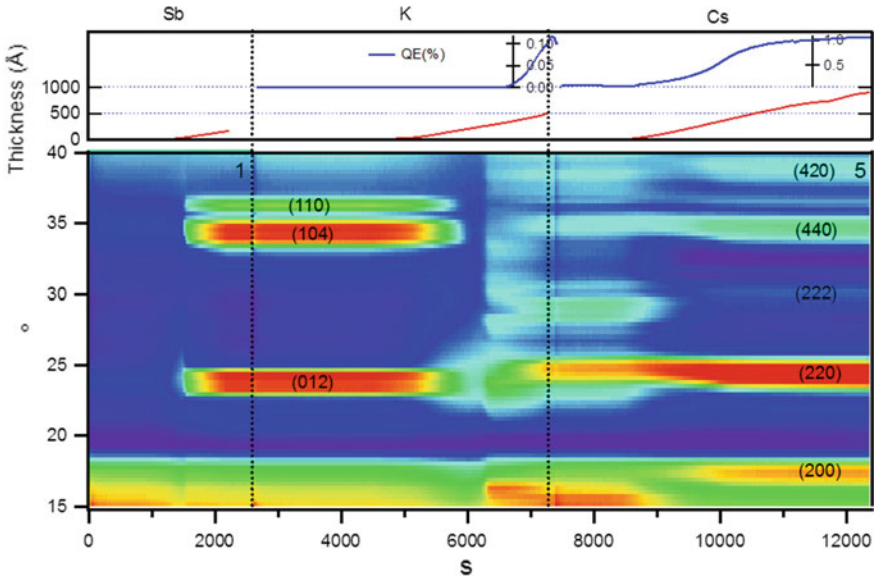


Fig. 7.23 XRD evolution of sequential K_2CsSb growth, showing the crystallization of the antimony film at 4 nm thick. This is followed by potassium deposition which causes the film to go amorphous at ~6000 s into the deposition. When the film recrystallizes into potassium antimonide, the QE (top plot) begins to rise

improved version of the Li photoionization source [84], a source based on the chemi-ionization of metastable He atoms [85], and sources using the Fano effect in Cs [86]. None of these latter sources were ever developed to the point of being operational at accelerators. Following the 1974 demonstration [91] of polarized photoemission from GaAs at low voltage, a high voltage source was constructed at SLAC [87]. This was used to conduct the seminal parity-violation experiment E122 [88] that verified predictions by Weinberg and Salam, and thereby helped to establish the Standard Model of electroweak physics. Shortly thereafter, polarized electron sources using GaAs photocathodes with an NEA surface condition were constructed at MIT-Bates [89], NIKHEV [90], the Mainz Microtron [91], Bonn-ELSA [92], Nagoya University [93] and CEBAF-Jefferson Lab [94].

Photocathode QE and electron spin polarization (ESP) are important characteristics of GaAs photocathodes. For decades, photocathode experts have worked to increase these quantities. The QE of GaAs photocathodes is affected by many factors, including cathode material quality, the wavelength of the incident light, the thickness of the photocathode, dopant density, the photocathode temperature, surface contamination, the negative electron affinity condition on the photocathode surface, the power density of the laser light, the bias voltage (i.e., Schottky effect), and the vacuum pressure under which photo-extracted beam is produced. These factors also affect ESP via spin relaxation/depolarization mechanisms that influence conduction-band electrons as they migrate toward the photocathode surface and are then emitted

into vacuum. This section provides an overview of polarization and ion bombardment effects, which are the chief lifetime-limiting factors of a NEA GaAs photocathode inside a high voltage photogun.

GaAs is a direct-transition III–V semiconductor with zincblende crystal structure. It can absorb laser light across the broad visible spectrum, but only illumination with near-IR wavelengths provides polarized photoemission. This can be understood by looking at detailed [95] and simplified representations [96] of the energy level diagrams of GaAs (Fig. 7.24). Electron spin-orbit coupling splits the $P_{1/2}$ and $P_{3/2}$ energy levels of the valence band into two states separated by 0.33 eV, which is large enough to avoid optical pumping from the lower energy $P_{1/2}$ state. Polarized photoemission takes advantage of the quantum-mechanical selection rules, noting that for circularly polarized laser light, conservation of angular momentum requires an electron's spin angular momentum quantum number to change by one unit, $\Delta m_j = \pm 1$. Furthermore, some transitions are more favorable than others as indicated by the transition probabilities shown in Fig. 7.24 (right). By using circularly polarized laser light with near-bandgap energy, the conduction band can be preferentially populated with a particular spin state. Polarization is defined as $P = (N\uparrow - N\downarrow)/(N\uparrow + N\downarrow)$ where N refers to the number of electrons in the conduction band of each spin state, "up" or "down". For bulk GaAs, the theoretical maximum polarization is 50%, corresponding to three electrons of the desired spin state for every one electron with the opposite spin.

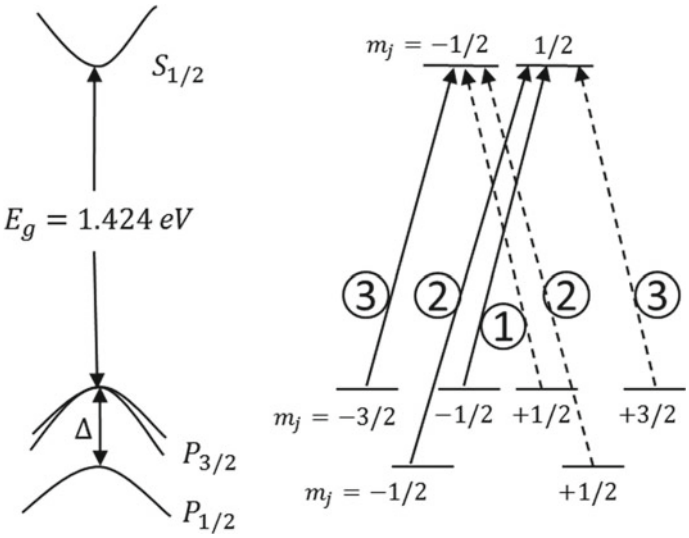


Fig. 7.24 (Left) Energy level diagram of GaAs at the center of the Brillouin zone: a "close-up" view near the valance band maxima/conduction band minima, and (right) simplified view showing the optical transitions between sublevels for right circularly polarized light (solid lines) and left circularly polarized light (dashed lines), with relative transition strengths given by circled numbers

As discussed previously for other photocathodes, the emission of electrons from GaAs can be described as a three-step process [3] involving absorption of light, diffusion of electrons to the surface of the photocathode, and emission of the electrons into the gun vacuum chamber. However, it is worth mentioning two aspects of efficient photoemission that are unique to GaAs. Namely, efficient photoemission from GaAs requires that the material be p-doped [97], which serves to lower the Fermi level (relative to the band edges) throughout the material. The p-doping—when acting together with the negative applied external field and (in some cases) other effects—also serves to induce band bending and lower the conduction band at the surface of the photocathode; this, in turn, reduces the electron affinity (to the new value EA_{eff} in Fig. 7.25b). Secondly, no significant photoemission is obtained until the potential barrier is reduced further and this is accomplished by adding about a monolayer of cesium and oxidant (Fig. 7.25c). The process of adding cesium and oxidant to the photocathode surface to create the desired NEA condition is called "activation".

At room temperature, highly doped bulk GaAs photocathodes typically provide ESP of the order ~30%, a value considerably less than the theoretical maximum value of 50%. The investigation of spin relaxation and depolarization has a long history dating back to the 1950s, [98, 99]. The literature describes two main spin relaxation mechanisms for p-type III–V semiconductors: (1) the lack of inversion symmetry in III–V semiconductor leads to a spin splitting of the conduction band, called the D'yakonov-Perel (DP) mechanism [100], and (2) the exchange interaction

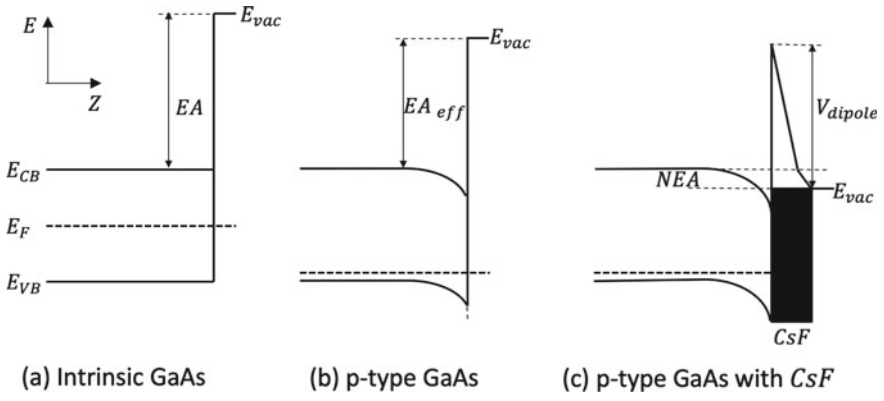


Fig. 7.25 Energy level diagrams describing the formation of the negative electron affinity (NEA) condition on GaAs. E_{CB}, E_{F}, E_{VB}, *and* E_{vac} describe the conduction band base energy, the Fermi level, the valence band maximum energy, and the local vacuum energy level, respectively. *EA* and EA_{eff} refer to the local electron affinity, and the latter determines the so-called "surface work function" ϕ_{s} defined by $\phi_{\text{s}} = E_{\text{vac}} - E_{\text{F}}$. V_{dipole} is the further decrease in effective electron affinity caused by the CsF activation layer. The diagrams show: **a** intrinsic GaAs; **b** how p-type dopant lowers the Fermi level which—together with the negative applied external field and (in some cases) other effects—leads to downwards band bending at the surface; and **c** how adding a CsF layer lowers the surface potential-energy barrier and leads to further band bending, to a lower surface work function, and to the desired NEA condition

between electrons and holes, called the Bir-Aronov-Pikus (BAP) mechanism [101]. As described below, for p-type GaAs, the BAP mechanism dominates with the DP mechanism playing a role at high temperature and low dopant concentration [102, 103]. There are other depolarization processes, but these are typically considered to have little consequence and are frequently ignored. These include the Elliot-Yafet (EY) mechanism [104], in which the spin–orbit interaction generates non–pure spin states in the conduction band, and radiation trapping, in which ESP is diluted by supplemental photoemission that results from the absorption of linearly polarized recombination light. There is a wealth of literature describing polarization studies using bulk GaAs, as a function of temperature [105, 106], dopant concentration [97, 99] electron (and hole) density [107], and thickness of the photocathode [98]. In the sections below, recent measurements are summarized describing ESP sensitivity to sample temperature and Zn dopant density, which have a significant impact on ESP. In general, ESP can be increased at the expense of QE.

7.4.2 Theory: Spin Relaxation Mechanisms

Electrons in the conduction band will arrive at an equilibrium polarization when the photocathode reaches a steady-state condition defined as, $\mathrm{d}P/\mathrm{d}t = P_0/\tau - P/\tau - P/\tau_s = 0$. The term P_0/τ is the rate of polarization creation using circularly polarized light, P/τ describes polarization loss due to electron recombination to the valence band, and P/τ_s is the rate at which polarization disappears due to spin relaxation effects. So the equilibrium polarization is given by [105]

$$P = P_0 \frac{1}{1 + \frac{\tau}{\tau_S}} \tag{7.3}$$

where P_0 is the initial polarization determined by the quantum-mechanical selection rules (as mentioned above, $P_0 = 50\%$ for bulk GaAs), τ and τ_s describe the electron lifetime and spin relaxation time for electrons at the bottom of the conduction band, respectively. The electron lifetime τ of GaAs is of the order 10^{-9}–10^{-10} s, with the exact value dependent on the temperature and doping concentration of the sample. The cumulative spin relaxation time is composed of individual spin relaxation mechanisms given by $1/\tau_s = 1/\tau_s^{DP} + 1/\tau_s^{BAP} + 1/\left(\tau_s^{EY} + 1/\left(\tau_s^{rad}\right)\right)$, where the superscripts represent the different spin relaxation mechanisms mentioned above. For p-type GaAs, the terms related to the EY mechanism and radiation trapping can be neglected.

The lack of inversion symmetry in GaAs is due to the presence of two distinct atoms in the Bravais lattice, such that the momentum states of the spin-up and spin-down electrons are not degenerate $E_{k\uparrow} \neq E_{k\downarrow}$.The resulting energy difference plays the role of an effective magnetic field and results in spin precession during the time

between collisions, which contributes to spin relaxation because the magnitude and direction of $\bar{k}$ changes in an uncontrolled way. This is the so-called DP mechanism, and the spin relaxation rate is given by [100, 108]

$$\frac{1}{\tau_s^{DP}} = Q\tau_p \alpha_0^2 \frac{(k_B T)}{\hbar^2 E_g}, \tag{7.4}$$

where Q is a dimensionless factor and ranges from 0.8 to 2.7 depending on the dominant momentum scattering mechanism,τ_p is the momentum relaxation time, α_0 is a dimensionless parameter specifying the strength of the spin-orbit interaction $\alpha_0 = 4\Delta m_e/m_{ev}\big((E_g + \Delta)\big(3E_g + 2\Delta\big)\big)^{\frac{1}{2}}$, where m_e is the effective mass of the electron, m_{ev} is a constant close in magnitude to the mass of the free electron) and E_g is the bandgap of GaAs. The temperature dependence of the spin relaxation rate is $1/\tau_s^{DP} \sim T^3\tau_p \sim T^{9/2}$ [108].

In p-type GaAs, spin relaxation can result from the spin exchange interaction between electrons and holes. This is the so-called BAP mechanism, and the spin relaxation rate is given by two terms [101]. In the case of exchange with nondegenerate holes

$$\frac{1}{\tau_s^{BAP}} = \frac{2}{\tau_0} N_A \alpha_B^3 \frac{v_k}{v_B}\left[\frac{N_h}{N_A}|\psi(0)|^4 + \frac{5}{3}\frac{N_A - N_h}{N_A}\right], \tag{7.5}$$

where τ_0 is an exchange splitting parameter given by $1/\tau_0 = (3\pi/64)\Delta_{ex}^2/\hbar E_B$ (where Δ_{ex} is the exchange splitting of the excitonic ground state and $E_B = \hbar^2/2m_e\alpha_B^2$ is the Bohr exciton energy), $\alpha_B = \hbar^2 \in /e^2 m_e$ is the Bohr exciton radius, v_k is the electron velocity, $v_B = \hbar/m_e\alpha_B$ is the Bohr exciton velocity, N_h is the density of free holes, and $|\psi(0)|^2$ is Sommerfeld's factor).

In the case of exchange with degenerate holes and when the electron velocity v_k is greater than the Fermi velocity of the holes, the spin relaxation rate is given by [108]

$$\frac{1}{\tau_s^{BAP}} = \frac{3}{\tau_0} N_h \alpha_B^3 \frac{v_k}{v_B}\frac{k_B T}{E_f h}, \tag{7.6}$$

where E_{fh} is the hole energy at the Fermi level [108, 109]. If the electrons are thermalized, v_k needs to be replaced by the thermal velocity $v_\varepsilon = (3k_B T/m_\varepsilon)^{1/2}$.

The temperature dependence of τ_s^{BAP} is dominated by the temperature of $|\psi(0)|^2$, as well as by the density of free holes N_h. The dependence on the acceptor density is essentially $1/\tau_s^{BAP} \sim N_A$ for nondegenerate holes from (7.5), and $1/\tau_s^{BAP} \sim N_A^{1/3}$ for degenerate holes from (7.6). In between these regimes, $1/\tau_s^{BAP}$ is only weakly dependent on N_A.

The formulas cited above were used to calculate the spin relaxation rate as a function of temperature for bulk GaAs with dopant concentrations 1×10^{19} and

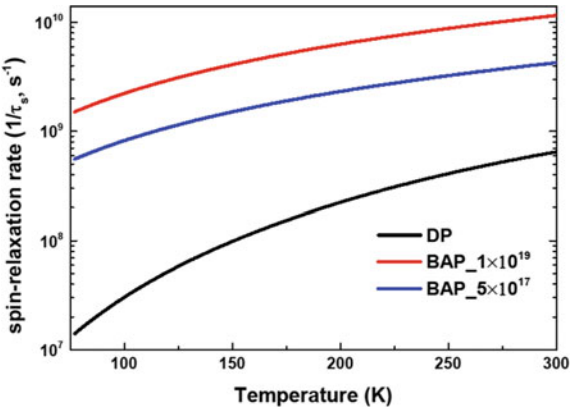


Fig. 7.26 Temperature dependence of the spin relaxation rate for the DP mechanism, BAP mechanism, and the total value in bulk GaAs

5×10^{17} cm^{-3}, as shown in Fig. 7.26. The BAP mechanism clearly dominates over the DP mechanism, which contributes appreciably only at higher temperature. The clear message from this plot is that low spin relaxation rates—and therefore higher ESP—will be obtained at lower dopant densities and temperatures.

7.4.3 Recent Measurements with GaAs NEA Photocathodes

7.4.3.1 Bulk GaAs Polarization Study—Temperature and Dopant Dependence

In a recent work [110], three bulk GaAs photocathodes, each with 100 surface cleave plane but with dopant densities spanning roughly 1.5 orders of magnitude, were activated at room temperature and evaluated using a retarding field Mott-polarimeter apparatus [111]. Photocathode QE and ESP were measured as a function of illumination wavelength, first at room temperature and then with a sample holder filled with dry ice (195 K) and LN_2 (77 K). The QE and ESP spectral scans presented in Fig. 7.27 exhibit the typical shape for bulk GaAs, namely that ESP increases while QE decreases as the energy (wavelength) of the illumination light is decreased (increased), with the highest polarization obtained when the energy of the light is equal to the semiconductor bandgap ensuring that only electrons from the $P_{3/2}$ ground state are excited to the conduction band. More interesting is the effect of dopant density on QE and ESP. Higher dopant concentrations serve to increase band bending, which lowers the surface work function, which increases the electron escape probability and leads to higher QE, however, at the expense of polarization. For the commonly used dopant density of 10^{19} cm^{-3}, photocathode QE at room temperature reached ~6.9%, but provided a maximum ESP of only ~30%. Polarization increased to ~41% for the low-doped sample, but provided QE of only ~1.5%. This behavior is

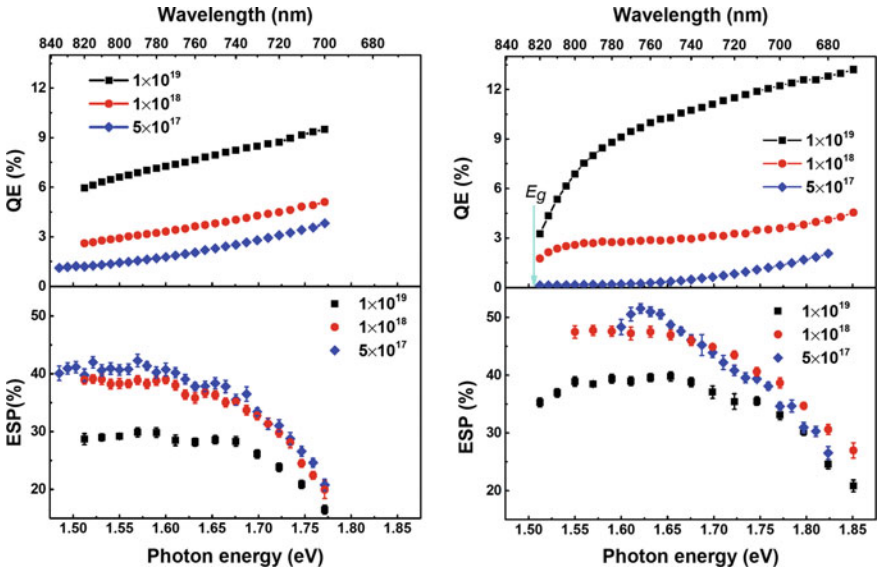


Fig. 7.27 The QE and ESP of bulk GaAs (100 cleave plane) with different Zn dopant densities, and measured at 300 K (left) and 77 K (right). Error bars are statistical. Axes ranges were kept identical to highlight the measurement variations observed between conditions. The bandgap energy E_g could be discerned for samples at 77 K (note arrow)

consistent with the predictions of the BAP mechanism that describes the spin relaxation rate as proportional to the dopant concentration. Higher dopant concentration leads to greater spin relaxation rate, and thus lower ESP.

Cooling the samples to 77 K modifies the crystal lattice structure and shifts the bandgap energy [112]. This bandgap shift means that peak polarization occurs at higher photon energies. The bandgap energy for GaAs can be calculated using the relation

$$E_g(T)/\text{eV} = 1.519 - \frac{5.41 \times 10^{-4}(T/K)^2}{204 + (T/K)}, \tag{7.7}$$

which yields bandgaps of 1.51 and 1 1 × 0.42 eV for samples at 77 and 300 K, respectively. The bandgap "knee" is visible in the QE spectral plots of Fig. 7.27 only for 77 K results. More noteworthy is the significant increase in ESP observed for all three samples cooled to 77 K. The ESP for the highly doped sample increased from ~ 30 to 41%, and from ~ 40 to 52% for the low-doped sample. This behavior is consistent with Fig. 7.25, which predicts smaller spin relaxation rates for both mechanisms (BAP and DP) at lower temperature. For the low-doped sample at 77 K, measured polarization exceeds the theoretical maximum value of 50%. This could point to an inaccurate effective Sherman Function used in the Mott polarimeter analysis, with this measurement serving to identify the magnitude of systematic error relevant to

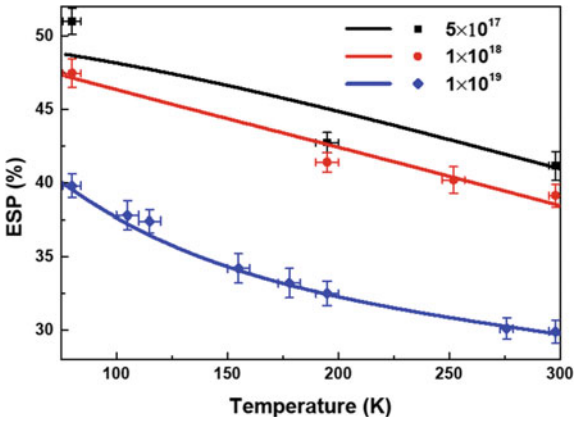


Fig. 7.28 Maximum ESP versus sample temperature and dopant density (the wavelength of illumination was allowed to vary to provide peak polarization). Fits were applied using the sum of expressions (7.4) and (7.6), which describe the DP and BAP spin relaxation mechanisms, respectively. Vertical error bars are statistical; horizontal error bars relate to uncertainty in sample temperature

the entire study, or perhaps an indication of some interesting physics phenomenon, e.g., the creation of strain within the sample at cryogenic temperature, which serves to eliminate the energy level degeneracy of the $P_{3/2}$ ground state.

The temperature and dopant density studies are summarized in Fig. 7.28, which shows ESP versus temperature for the three samples with different dopant concentrations. Overall, lower temperature leads to lower spin relaxation rate, and therefore, higher ESP. Fits were applied to the data sets using the sum of expressions (7.4) and (7.6), which describe the DP and BAP spin relaxation mechanisms, respectively. The fit to the high-dopant sample set is quite good, but less so for the lower dopant density results, although the fits still support the basic predicted trend.

7.4.3.2 High Polarization Strained Superlattice Photocathode with Distributed Bragg Reflector

GaAs photocathodes grown on a crystal structure with a different lattice constant can provide beam polarization exceeding 50% because of induced uniaxial strain that eliminates the degeneracy of the $P_{3/2}$ valence band. The first demonstrations of this technique were reported in 1991 using photocathodes grown with a single-strained layer [113, 114]. Throughout the 1990s, single-strained-layer GaAs/GaAsP photocathodes [115–117] were used at electron accelerators throughout the world [89, 94, 99] providing polarization 75–80% and maximum QE ~ 0.3% or less. But beam delivery using single-strained-layer photocathodes made apparent the delicate competing balance between maintaining the required strain, and growing a layer thick enough to provide sufficient QE. In the following decade, higher polarization and notably higher QE were obtained using strained superlattice structures consisting of very thin quantum-well active layers and alternating lattice-mismatched barrier layers [117]. The superlattice structure maintained the required degree of strain to produce high polarization and also provided sufficient active layer thickness to obtain higher QE. The first demonstration of beam production from a strained

$GaAs/In_{0.15}GaAs$ superlattice photocathode in 1994 achieved polarization of 82.7% but QE of just 0.015%; however, continued development eventually resulted in photocathodes with QE ~1%. The strained $GaAs/GaAsP_{0.36}$ superlattice structure reported in [118] provided QE of 1.2% with the polarization of 86% and is available commercially [119]. Strained superlattice photocathodes fabricated at other facilities have also demonstrated very high polarization but no higher QE [120–123].

Some new accelerator applications require very high average current polarized beams. A photocathode with a QE of only 1% would require ~8 W of laser light (with RF structure) to generate the desired 50 mA average beam current for the eRHIC proposal [124] and even more light when QE decays during operation due to ion bombardment [125]. In the standard strained superlattice design, most of the incident laser light simply heats the photocathode instead of promoting electron ejection, which can degrade QE due to evaporation of the chemicals used to reduce the surface work function. Methods to cool the photocathode during beam delivery are complicated because the photocathode floats at high voltage. Developing photocathodes with enhanced QE could simplify the photogun design, reduce the drive-laser power requirements, and prolong the effective operating lifetime of the photogun.

A photocathode with a distributed Bragg reflector (DBR) was first proposed in 1993 [126] and reported in the following years [127, 128] as a means to enhance photocathode QE. The DBR serves to create a Fabry-Perot cavity formed by the front surface of the photocathode and a DBR region existing beneath the superlattice active layer. Instead of a single pass in the standard design, laser light of a particular wavelength reflects repeatedly within the Fabry-Perot cavity, which increases the beneficial absorption of the incident photons, and in principle leads to enhanced QE. From past work, it was clear that the wavelength of peak reflectivity of the DBR was very sensitive to the refractive indices and thickness of each DBR layer constituent, and that the peak of resonant absorption was very sensitive to the thickness of the photocathode, especially the spacer layer between the DBR and the superlattice photocathode structures. Accurate control of the thickness of each layer and of the composition of each constituent chemical within the photocathode structure represents a significant challenge.

A recent DBR study [129] benefited from past work that resulted in the fabrication of the strained GaAs/GaAsP superlattice photocathode described in [118], which represents the basis for polarization and QE comparison. See [118] for a complete description of the design criteria and optimization of this photocathode structure. A schematic layout of the strained GaAs/GaAsP superlattice photocathodes with and without the DBR structure is shown in Fig. 7.29.

The DBR was composed of multiple layers of alternating high and low index of refraction. A natural choice for a high index of refraction layer was GaAsP, which provides a good lattice match to the graded buffer layer. Though a challenge for growth, $AlAs_{1-y}P_y$ was chosen for the low index of refraction layer, in order to achieve maximum refractive index contrast and minimum DBR thickness. The thickness of each layer follows the quarter-wavelength condition [126] $n_H(\lambda_{DBR})d_H = \frac{\lambda_{DBR}}{4} = n_L(\lambda_{DBR})d_L$. The bandwidth for which the DBR exhibits high reflectivity is given by the formula

GaAs	5 nm	p=5 × 10^{19} cm^{-3}
GaAs/GaAsP SL	(3.8/2.8 nm) ×14	p=5 × 10^{17} cm^{-3}
$GaAsP_{0.35}$	2750 nm	p=5 × 10^{18} cm^{-3}
Graded $GaAsP_x$ (x = 0~0.35)	5000 nm	p=5 × 10^{18} cm^{-3}
GaAs buffer	200 nm	p=2 × 10^{18} cm^{-3}
p-GaAs substrate (p>10^{18} cm^{-3})		

GaAs	5 nm	p=5 × 10^{19} cm^{-3}
GaAs/GaAsP SL	(3.8/2.8 nm) ×14	p=5 · 10^{17} cm^{-3}
$GaAsP_{0.35}$	750 nm	p=5 × 10^{18} cm^{-3}
$GaAsP_{0.35}$/ $AlAsP_{0.4}$ DBR	(54/64 nm) ×12	p=5 × 10^{18} cm^{-3}
$GaAsP_{0.35}$	2000 nm	p=5 × 10^{18} cm^{-3}
Graded $GaAsP_x$ (x = 0~0.35)	5000 nm	p=5 × 10^{18} cm^{-3}
GaAs buffer	200 nm	p=2 × 10^{18} cm^{-3}
p-GaAs substrate (p>10^{18} cm^{-3})		

Fig. 7.29 Schematic structure of the photocathodes: without DBR (left) and with DBR (right)

$$\Delta\lambda_{DBR} = \frac{4\lambda_{DBR}}{\pi} \sin^{-1}\left(\frac{n_H(\lambda_{DBR}) - n_L(\lambda_{DBR})}{n_H(\lambda_{DBR}) + n_L(\lambda_{DBR})}\right). \tag{7.8}$$

The Fabry-Perot cavity is formed by the vacuum/GaAs cap layer interface and the DBR. Proper thickness of a GaAsP spacer layer between GaAs/GaAsP superlattice and the DBR ensures the maximum absorption in the superlattice active layer and coincides with the central reflective wavelength of the DBR. To design the structure, the refractive index of these layers can be estimated using the simplified interband transition model as a function of photon energy [130]

$$n(E)^2 = A\left[f(X_0) + \frac{1}{2}\left(\frac{E_0}{E_0 + \Delta_0}\right)^{\frac{1}{2}} f(X_{so})\right] + B, \tag{7.9}$$

with $f(X) = x^{-2}(2 - \sqrt{1+X} - \sqrt{1-X})$, $X_0 = \frac{E}{E_0}$, and $X_{so} = \frac{E}{E_0+\Delta_0}$. Here, A represents the strength parameter of the $E/(E_0 + \Delta_0)$ terms and B represents the nondispersive contribution arising from the higher-lying band-gap terms (E_1, $E_1 + \Delta_1$, E_2, etc.) for each layer. The quantities E_0 and Δ_0 are the band-gap energy and spin-orbit splitting, respectively, given by Vegard's Law for $GaAs_{1-x}P_x$ [131, 132].

The desired wavelength of peak reflectivity was λ_{DBR} ~780 nm, compatible with CEBAF drive lasers. Setting the phosphorus fractions in the $GaAs_{1-x}P_x$ and $AlAs_{1-y}P_y$ layers to 0.35 and 0.4, respectively, provided refractive indices of n_H (780 nm) = 3.4506 and n_L (780 nm) = 2.9443. From these parameters, the design values for the thickness of the $GaAsP_{0.35}$ and $AlAsP_{0.4}$ layers were 56.5and 66.5 nm, respectively. Actual layer thicknesses were very close to these design values, 54 and 64 nm, which set the wavelength of peak reflectivity to λ_{DBR} of 755 nm, and the reflectivity bandwidth λ_{DBR} ~ 79 nm. This implies that the DBR can achieve a high reflectivity, >90% for wavelengths between 715 and 795 nm.

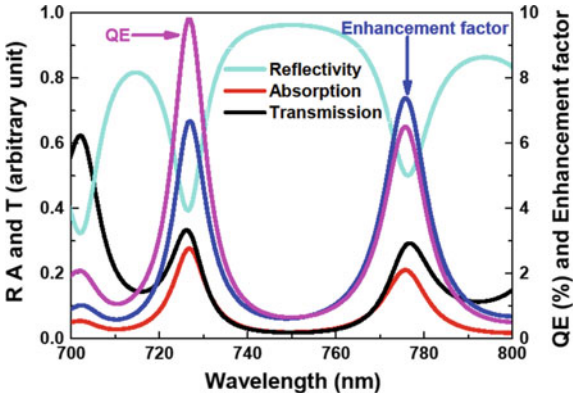


Fig. 7.30 Calculated values of absorption, reflectivity, and transmittance of the DBR photocathode, as a function of wavelength. Also shown, calculated QE and the QE enhancement factor compared to the photocathode without the DBR

The transfer matrix method [133, 134] was used to calculate the reflection, transmission, and absorption of incident light through the complete multilayer thin-film structure including the GaAs substrate. Calculated values of surface reflectivity (R), transmittance (T) into the GaAs substrate, absorption (A), and the absorption enhancement factor as a function of wavelength, are shown in Fig. 7.30. There are two resonant absorption peaks within the optical cavity for the DBR photocathode at wavelengths 726 nm and 776 nm. Absorption by the substrate is not plotted, as this process does not contribute to the polarized electron emission. The resonant peak at 776 nm is just 4 nm from the desired value of 780 nm. At 776 nm, corresponding to maximum electron spin polarization, the absorption is 21.03%, surface reflectivity is 50.14%, and transmittance into the GaAs substrate is 28.83%. The full-width-at-half-maximum of the Fabry-Perot resonance is about 10 at 776 nm. The absorption enhancement factor is 7.4 at this wavelength.

To estimate the quantum efficiency of the photocathode, one can solve the one-dimensional diffusion equations [135, 136], based on Spicer's three-step model

$$QE(\lambda) = \frac{P_L F_L A}{1+\frac{1}{\alpha_\lambda L_L}} + \frac{P_\Gamma \exp\left[k\left(-\frac{1}{1.42}-\frac{\lambda}{1240\ \text{nm}}\right)\right]A}{1+\frac{1}{\alpha_\lambda L_\Gamma}} \left[F_\Gamma + \frac{F_L L_\Gamma}{\alpha_\lambda F_L (L_\Gamma + L_L)\left(1+\frac{1}{\alpha_\lambda L_L}\right)}\right] \tag{7.10}$$

where P_Γ and P_L are the surface electron escape probabilities for the Γ and L minima, respectively, which are independent of the incident photon energy, F_Γ is the remaining fraction of excited electrons, F_L is the fraction of electrons that is excited to energies greater than the L minima, A is the absorption of the photocathode, L_Γ and L_L are the electron diffusion lengths for the Γ and L minima, respectively, α_λ is the absorption coefficient of the photocathode, and k is a coefficient ($k \geq 0$). The

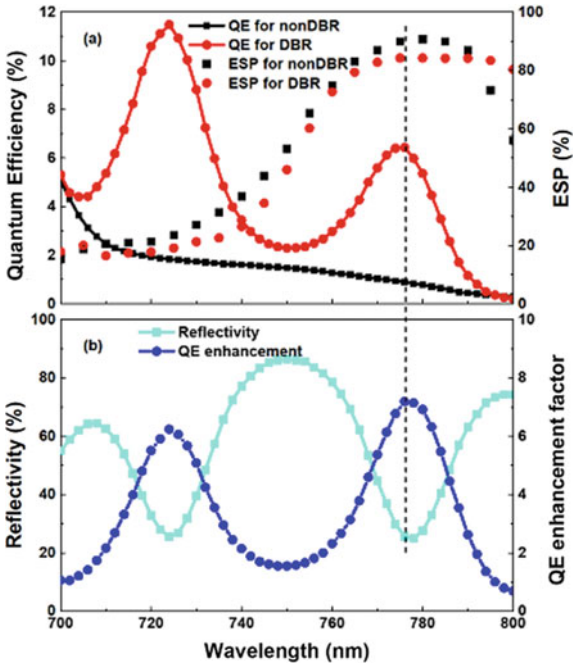


Fig. 7.31 **a** The QE and electron spin polarization (ESP) for the strained GaAs/GaAsP superlattice photocathodes with and without DBR as a function of the wavelength; **b** Reflectivity and QE enhancement factor of DBR photocathode as a function of the wavelength. The dashed line indicates the resonant position

calculated QE for the DBR photocathode, together with the QE enhancement factor as a function of wavelength, are shown in Fig. 7.30.

Measured values of QE and polarization as a function of wavelength, for photocathodes with and without DBR, are shown in Fig. 7.31a. For the non-DBR photocathode, results are consistent with past work [118] indicating peak polarization of 90% and QE of 0.89% at a wavelength ~780 nm. In stark contrast, the QE of the DBR photocathode shows the telltale prominent oscillatory behavior indicative of resonant absorption. Measured reflectivity and QE enhancement (i.e., the ratio of QE values of photocathodes with and without DBR) are shown in Fig. 7.31b. The two dips in the reflectivity spectrum are clearly correlated with wavelength locations of QE maxima. At the wavelength of interest, 776 nm, the QE was 6.4% and polarization ~ 84%. The measured QE enhancement of ~7.2 is very close to the predicted value of ~7.4.

7.4.3.3 Ion Bombardment—Cleave Plane Dependence

Ion back-bombardment represents the dominant lifetime-limiting mechanism of modern DC high voltage photoguns that rely on delicate NEA GaAs photocathodes [125, 137]. Ion back-bombardment is the process whereby residual gas within the cathode/anode gap is ionized by the extracted electron beam and attracted to the negatively-biased photocathode. Photogun experts speculate that ions strike the

photocathode surface sputtering away the chemicals used to create the NEA condition, or they become implanted within the photocathode material, knocking out atoms from the crystal structure and altering the energy band structure, or they serve as interstitial defects that reduce the electron diffusion length. All of these possibilities are problematic because the net result is reduced QE, which results in a reduced operating lifetime of the photogun. Improving the vacuum within the photogun is the best way to reduce ion back-bombardment and thereby prolong the operating lifetime of the photogun, but modern DC high voltage photoguns are already operating in the 10^{-11} Torr (about 10^{-9} Pa) pressure range, or lower, so improving the photogun vacuum further is thus very challenging.

To understand ion back-bombardment inside a photogun, it is important to note that the energy spectrum of the extracted photon emitted electron beam within the cathode/anode gap is broad, with electrons leaving the cathode electrode at zero velocity, and then gaining energy until reaching the anode. For a typical DC high voltage photogun, the photocathode is biased at voltages between −100 and –350 kV. Hydrogen is the dominant gas species inside a modern photogun, and the peak ionization cross section occurs at ~100 V and drops by more than two orders of magnitude at 100 kV [138] (Fig. 7.32). So in typical photogun operation, the surface of the photocathode is bombarded by significantly more low-energy ions than high-energy ions.

The software program SRIM (Stopping Range of Ions in Matter) [139] can be used to estimate the penetration depth of the hydrogen ions into the material and the number of vacancies created within the material when ions collide with gallium and arsenic atoms and knock them out of their positions within the crystal lattice. In the simulation, the hydrogen ions were injected into GaAs at normal incidence and the ion injection energy was varied from 100 to 10 kV while keeping the ion dose constant for each simulation. The stopping depth of implanted hydrogen ions and the distribution of knock-out vacancies for different ion energies are shown in Fig. 7.33. Unsurprisingly, higher energy ions penetrate deeper into the material and do more

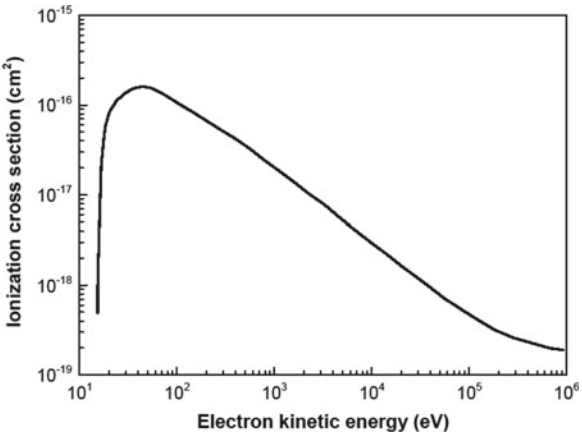


Fig. 7.32 Electron impact ionization cross section for hydrogen molecules—the dominant gas species inside modern DC high voltage photoguns—as a function of electron kinetic energy

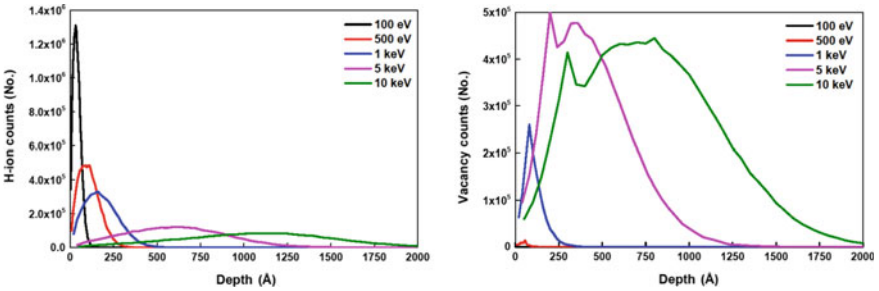


Fig. 7.33 (Left) Hydrogen ion distribution and (right) vacancy distribution within GaAs, for different hydrogen ion energies. All simulations performed using the same hydrogen ion dose of $8.1 \times 10^5 \text{nA} \cdot \text{s}$. Plots obtained using SRIM [139]

damage to the crystal structure, compared to lower energy ions. The penetration depths (defined as the depth at which the dose falls to 1/e of its original value) are 69 and 1690 Å for 100 V and 10 kV hydrogen ions, respectively, while gallium and arsenic atoms begin to be knocked out from their lattice positions at ion energies greater than ~500 V. It is an interesting question, which ions are most problematic for NEA GaAs photocathodes inside a photogun, many low-energy ions or fewer high-energy ions [140]?

Channeling is the process whereby positively charged ions follow the direction between two neighboring crystal planes, which leads to deeper penetration of the positive ions in the material. Figure 7.34 shows the relative spacing of atoms in the GaAs crystal structure for the three surface cleave planes studied, viewed at normal incidence. The red boxes in Fig. 7.33 denote the recurring geometric cells used for calculating the atomic number density (number of atoms per unit area), which is a measure of the "open space" between atoms. The atomic number densities

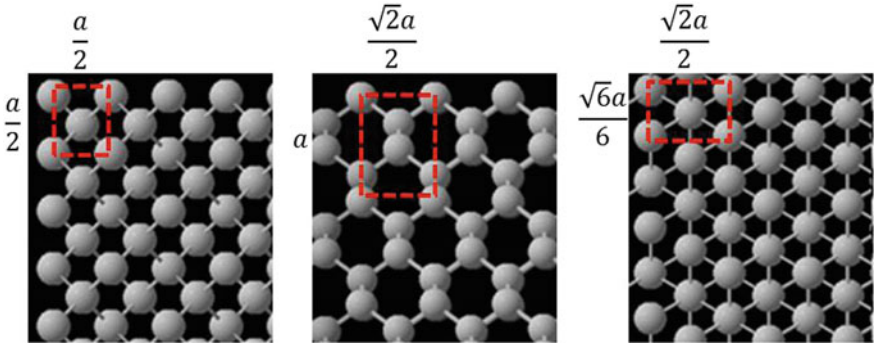


Fig. 7.34 Schematic representations of the GaAs crystal for three surface cleave planes, looking normal to the surface—from left to right, (100), (110), and (111). The red boxes denote the recurring geometric cells used for calculating the atomic number density for each cleave plane and a is the GaAs lattice constant

for the (100), (110), and (111A) GaAs cleave planes were calculated to be $8/a^2$, $4\sqrt{2}/a^2$, and $4\sqrt{3}/a^2$, respectively, where a is the GaAs lattice constant (5.6535 Å). Measurements were performed to investigate if channeling would be enhanced for the (110) cleave plane which possesses the smallest atomic number density (i.e., the most open space between atoms). If so, a photocathode with (110) surface cleave plane would, therefore, be the most insensitive to ion bombardment.

To quantify the effects of ion energy, and to investigate whether crystal orientation plays a role in QE degradation, GaAs samples with different surface cleave planes were implanted with hydrogen ions at 100 and 10 kV. Prior to implantation, each photocathode sample was heated, cooled, and activated. The QE was measured as a function of laser wavelength and the results are presented in Fig. 7.35. Quantum efficiency was calculated using the expression $QE = 124I/P\lambda$, where I is the measured photocurrent in μA, P is the laser power in mW, and λ is the wavelength of the laser light in nm [141]. The cleave planes (100) and (110) provided the highest QE, approximately 21% at the laser wavelength of 532 nm. Cleave plane sample (111A) provided the lowest QE, approximately 19% at 532 nm. Although these values are lower than reported in [97], the relative trend between samples is consistent (i.e., $QE_{100} > QE_{110} > QE_{111A}$).

After determining that each sample provided high QE, samples were heated to 250 °C for 20 min to remove the activation layer of Cs and NF_3. Samples were then cooled to room temperature and implanted with hydrogen ions at the chosen energy and for the prescribed dose. The vacuum was allowed to recover overnight. Photocathode samples were then repeatedly heated and activated following different heating protocols (temperature and duration). Following each activation, the photocathode QE was measured at 532 nm and compared to the preimplantation value. In this manner, the sample's sensitivity to ion damage was evaluated. Figure 7.36 shows typical QE maps for samples implanted with 100 and 10 kV hydrogen ions (the caption provides specific details of the cleave plane surface and heating protocol). For 100 V ions, QE

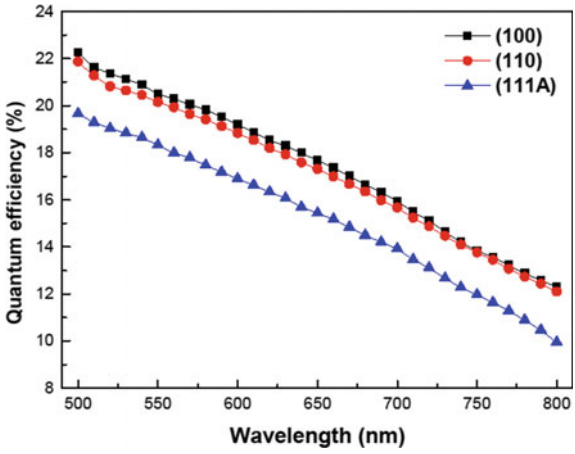


Fig. 7.35 Photocathode QE versus laser wavelength for commercial bulk GaAs samples with different surface cleave planes, prior to ion implantation

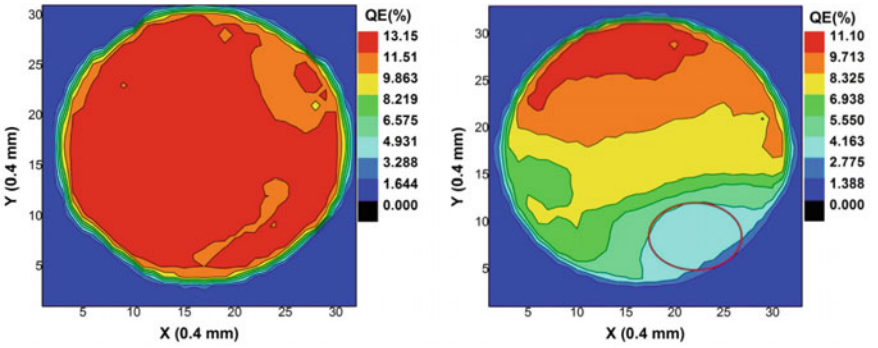


Fig. 7.36 Typical QE maps of the photocathode obtained by scanning a focused low power green laser beam across the photocathode following activation. These maps show the bulk GaAs (111A) sample. Left: after the third activation following implantation with 100 V hydrogen ions. Right: after the third activation following implantation with 10 kV hydrogen ions. The 2-dimensional plots help to illustrate ion implantation uniformity: 100 V hydrogen ions were uniformly distributed across the entire sample but 10 kV hydrogen ions were focused to a region denoted by the ellipse, consistent with the particle tracking code simulations

degradation was uniformly distributed across the sample. But for 10 kV hydrogen ions, the QE degradation was localized to a specific region of the sample, consistent with an electrostatic simulation that predicted significant focusing.

Figure 7.37 summarizes the results of the experiment, with QE at 532 nm plotted as a function of repeated activation following a variety of heating protocols described within the caption, for samples implanted with 100 V and 10 kV hydrogen ions, and for samples that were not implanted with ions. When samples were repeatedly heated and activated—but not subjected to hydrogen ion implantation—the QE at 532 nm remained constant to within 5%. This speaks to the quality of the vacuum within the apparatus, which was dominated by chemically inert gas species like H_2, CH_4, CO, and CO_2. The QE of each sample did not degrade following repeated heating and activation, due to adverse chemical reactions that might take place under degraded vacuum conditions, especially when samples were hot.

The next observation is evident in Fig. 7.37 is that both 100 and 10 kV hydrogen ions served to degrade the sample QE, however, the QE degradation associated with 10 kV ions was significantly greater than that caused by 100 V ions. In general, for both ion energies, heating the sample to a higher temperature and for a longer time served to restore QE. For implantation with 100 V hydrogen ions, the QE could be completely restored by heating the sample. But for implantation with 10 kV hydrogen ions, only a fraction of the QE could be restored. It should be noted that the data points presented in Fig. 7.37 were obtained by averaging the QE measurements over a specific region of the photocathode sample surface. For 100 V ion implantation, QE measurements were averaged over the entire sample. For the 10 kV ions, QE values were averaged only in the region where the ions were focused, as indicated by the ellipse shown in Fig. 7.36 (right).

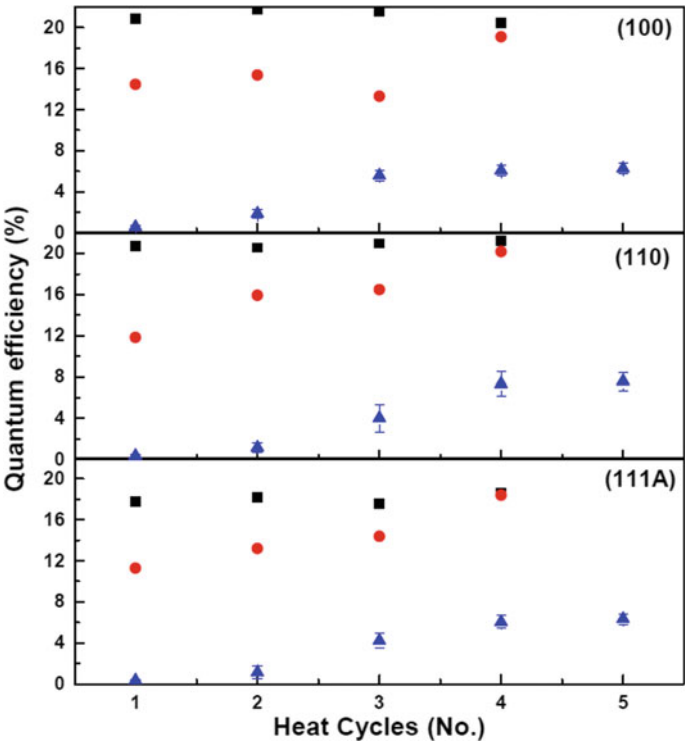


Fig. 7.37 QE versus heat cycle for bulk GaAs with three different surface cleave planes. Black Square: no implantation; Red Circle: implantation with 100 V hydrogen ions; Blue Triangle: implantation with 10 kV hydrogen ions. Error bars represent the standard deviation in QE values averaged over the entire sample for "no implantation": and 100 V results, and over the small region denoted by the red circle in Fig. 7.36 (right) for the 10 kV results. Error bars are smaller than the data-point symbols for "no implantation" and 100 V results, because QE was very uniform across the sample. Heat cycles were: number 1: 250 °C for 20 min; 2: 370 °C for 30 min; 3: 490 °C for 1 h; 4 and 5: 550 °C for 1 h

To determine if a specific surface cleave plane provided enhanced ion channeling, the QE results from the 10 kV ion implantation studies were divided by the non-implantation QE results (Fig. 7.38). In accordance with supposition, when ions penetrate deeper into the material, they pass beyond the surface region that contributes to photoemission, and therefore, provide less QE degradation. When comparing QE values of implanted and non-implanted samples, those with the highest QE ratio would correspond to samples with enhanced ion channeling. The simple picture presented in Fig. 7.34 suggests samples would be ordered (110), (111A), (100), from highest to lowest levels of channeling. This simple picture is consistent with the results presented in Fig. 7.38, at least for samples that were heated sufficiently long and at a relatively high temperature (see results for heat cycles 4 and 5 in Fig. 7.38). The (110) GaAs cleave plane—i.e., the sample with the most "open space" between

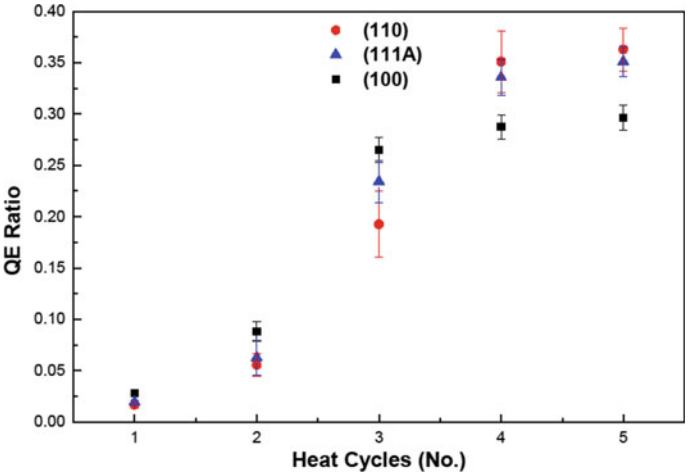


Fig. 7.38 QE ratio, for samples implanted with 10 kV hydrogen ions to samples that were not exposed to hydrogen ions, as a function of heat cycle, for bulk GaAs with three different surface cleave planes. Heat cycle numbers: 1: 250 °C for 20 min; 2: 370 °C for 30 min; 3: 490 °C for 1 h; 4 and 5: 550 °C for 1 h

atoms, and therefore, the sample supporting the highest level of ion channeling—exhibited the highest QE ratio. By contrast, the cleave plane (100)—i.e., the sample with the least "open space" between atoms, and therefore, the sample supporting the lowest level of ion channeling—exhibited the lowest QE ratio. If one assumes a strict QE dependence on atomic number density, the QEs of the (110) and (111A) cleave planes following ion implantation should be greater than the (100) sample by roughly 41% and 15%, respectively. In comparison, looking at Fig. 7.38 heat cycle 5 values, the measured QEs of the (110) and (111A) cleave planes following ion implantation are greater than the sample (100) by 33% and 30%, respectively. However, there is fairly good agreement between measurements and predictions when error bars are taken into account.

In conclusion, implantation of the photocathode with either low or high energy hydrogen ions served to reduce photocathode QE, but the QE degradation caused by 100 V hydrogen ions was modest and QE could easily be recovered following sample heating. In sharp contrast, the QE degradation caused by 10 kV hydrogen ions was severe and QE could not be fully recovered with heat. These observations suggest that low-energy ions penetrate the surface, stopping at locations between atoms and creating interstitial defects that likely reduce the diffusion length of the electrons within the material, reducing the number of available electrons that reach the photocathode surface that can contribute to photoemission. During sample heating, the interstitial defects diffuse throughout the material, restoring the electron diffusion length and the photocathode QE. Whereas high-energy ions likely generate knock-out vacancies within the crystal structure that permanently adversely alter the band structure of the semiconductor. As noted above, inside a DC high voltage

photogun, the photocathode will be bombarded by considerably more low voltage ions than high voltage ions. This is fortuitous and explains why the photocathode QE can be routinely restored following simple heating and reactivation, except at the electrostatic center of the photocathode, where the highest energy ions are directed [137].

In addition, the data support the simple picture that channeling is enhanced for the sample with the most "open space" between atoms. More "open space" leads to enhanced channeling, which means high-energy ions create fewer vacancies that lead to reduced and unrecoverable QE. The surface cleave plane (110) has the most "open space", and cleave plane (100) the least "open space". As such, the surface cleave plane (110) was the least sensitive to ion bombardment and the cleave plane (100) the most sensitive.

Considering all the results in total, the best choice for a photogun using bulk GaAs would be surface cleave plane (110), since this provides the highest QE and the least sensitivity to ion bombardment. It's true that modern high-current accelerator applications now prefer alkali-antimonide photocathodes over GaAs photocathodes, because alkali-antimonide photocathodes offer similarly high QE but longer operating lifetime. However, spin-polarized accelerator applications may be forced to use bulk GaAs as the only practical means to generate milliampere-levels of spin-polarized beam current.

7.5 Conclusion

Photocathodes represent an important technological development area, with diverse applications—from light detectors and image intensifiers to materials science and accelerator physics. There are three main classes of photocathodes used today—metals, semiconductors with positive electron affinity (PEA), and semiconductors with negative electron affinity (NEA). Metals are rugged, tolerant of poor vacuum and emit electrons promptly upon absorbing a photon (within tens of fs). However, metals have typically poor quantum efficiency (~10^{-5}) and are only sensitive to UV light. PEA semiconductors generally have much higher QE, as the loss of excited carriers is reduced by the elimination of electron-electron scattering, at least for a range of photon energies. The two most common PEA semiconductors are Cs_2Te and the alkali-antimonide family—Cs_2Te is a near UV sensitive cathode that can have very high QE (>20%) [142], while the alkali antimonides typically have very high QE peaked in the blue [143]. PEA semiconductors typically require good vacuum (~10^{-9}hPa) and have response times ~0.5 ps. NEA semiconductors, such as the GaAs family, can have very good QE over a broad range of wavelengths—even for IR light. They are also the only cathodes capable of producing significant spin polarization in the emitted electron beam. However, these cathodes rely on a dipole monolayer on the surface to achieve NEA, and are, therefore, very sensitive to poor vacuum (operation at typically ~10^{-11} hPa). They can also have long response times (>100 ps). Many accelerator applications, especially electron colliders and free electron lasers,

use photocathodes as the electron source, and these machines typically drive the development of the state of the art in this field. Improvement directions are higher QE, high peak current densities in the pulsed beam of $>10^5$ A/cm^2 and high beam currents of more than 50 mA.

References

1. W. Schattke, Prog. Surface Sci. **54**, 211 (1997)
2. C.N. Berglund, W.E. Spicer, Phys. Rev. A **136**, A1030 (1964)
3. W.E. Spicer, Phys. Rev. **112**, 114 (1958)
4. W.E. Spicer, A. Herrera-Gomez, Modern theory and applications of photocathodes, Report No. SLAC-PUB-6306 SLAC/SSRL-0042 (1993)
5. G. Ferrini et al., Solid State Comm. **106**, 21 (1998)
6. W.E. Pickett, P.B. Allen, Phys. Lett. **48A**, 91 (1974)
7. J. Smedley et al., Phys. Rev. ST Accel. Beams **11**, 013502 (2008)
8. D.H. Dowell et al., in *FEL'07,* Novosibirsk, Russia (2007), p. 276
9. D.H. Dowell et al., Phys. Rev. ST Accel. Beams **9**, 063502 (2006)
10. G. Suberlucq, Developpement et production de photocathodes. Report no. CLIC-NOTE-299 (1996)
11. H.B. Michaelson, J. Appl. Phys. **48**, 4729 (1977)
12. E. Chevallay et al., Nucl. Instr. Meth. Phys. Res. A **340**, 146 (1994)
13. F.L. Pimpec et al., in *IPAC 2010,* Kyoto, Japan (2010), p. 2233
14. T. Srinivasan-Rao et al., J. Appl. Phys. **69**, 3291 (1991)
15. R. Valizadeh et al., in IPAC2014, Dresden, Germany (2014), p. 711
16. P. Schoessow et al., The Argonne wakefield accelerator high current photocathode gun and drive linac. Report no. ANL-HEP-CP-95-34 (1995)
17. J. Bossert et al., in *36th International Free Electron Laser Conference FEL14*, ed. by J. Basel, Switzerland (2014), p. 832
18. S. Mistry et al., in *IPAC2015,* Richmond, VA, USA (2015), p. 1759
19. A. Brachmann et al., in *IPAC2011,* San Sebastian, Spain (2011), p. 3200
20. F. Zhou et al., Phys. Rev. ST Accel. Beams **15**, 090703 (2012)
21. X. J. Wang et al., in *LINAC1994,* Tsukuba, Japan (1994), p. 407
22. L.N. Hand, U. Happek, Nucl. Instr. Meth. Phys. Res. A **372**, 335 (1996)
23. Y. Gao, R. Reifenberger, Phys. Rev. B **35**, 4284 (1987)
24. D.H. Dowell, in *FEL2010,* Malmoe, Sweden (2010)
25. C. Vicario et al., in *FEL 2013,* New York, USA (2013), p. 434
26. R.J. Reid, Cleaning for vacuum service. Report no. CERN-OPEN-2000-276 (1999)
27. J.D. Herbert, R.J. Reid, Vacuum **47**, 693 (1996)
28. R. Ganter et al., in *FEL 2013,* New York, NY, USA (2013), p. 259
29. P.H. Holloway, G.E. McGuire, *Handbook of Compound Semiconductors* (William Andrew, 1996)
30. H. Qian et al., in *PAC2009,* Vancouver, Canada (2009), p. 476
31. G. Penco et al., in *IPAC'10,* Kyoto, Japan (2010), p. 1293
32. F. Zhou et al., in *FEL'14,* Basel, Switzerland (2014), p. 771
33. https://www.luvata.com/en/Products/Sputtering-targets-pellets-slugs/
34. Z. Zhang, C. Tang, Phys. Rev. ST Accel. Beams **18**, 053401 (2015)
35. K.L. Jensen et al., Phys. Rev. ST Accel. Beams **17**, 043402 (2014)
36. A. Lorusso et al., Phys. Rev. ST Accel. Beams **14**, 090401 (2011)
37. H.E. Ives, Phys. Rev. **38**, 1209 (1931)
38. E. Pedersoli et al., Appl. Phys. Lett. **87**, 081112 (2005)

39. H. Tomizawa et al., in *LINAC2012,* Tel-Aviv, Israel (2012), p. 996
40. M. Krueger et al., J. Phys. B: At. Mol. Opt. Phys. **51**, 172001 (2018)
41. L. Wimmer et al., Nat. Phys. **10**, 432 (2014)
42. C. Vicario et al., Opt. Express **23**, 4573 (2015)
43. R. Ganter et al., Phys. Rev. Letters **100**, 064801 (2008)
44. J. Smedley et al., J. Appl. Phys. **98**, 043111 (2005)
45. J. Sekutowicz et al., in *PAC07* (IEEE, Piscataway, NJ and Albuquerque, NM, USA 2007), p. 962
46. A. Lorusso et al., Appl. Phys. A **122**, 1 (2016)
47. Photocathode technology, http://www.hamamatsu.com/jp/en/technology/innovation/photocathode/index.html. Accessed 7 Feb 2019
48. A.D. Bona, F. Sabary, J. Appl. Phys. **80**, 3024 (1996)
49. H. Trautner, Ph.D thesis, Clic project at CERN, Switzerland (2000)
50. E. Chevallay et al., in *19th International Linear Accelerator Conference,* Chicago, USA (1998), p. 872
51. E. Chevallay, Experimental results at the CERN photoemission laboratory with co-deposition photocathodes in the frame of the CLIC studies. Report no. CTF3-Note-104 (2012)
52. A. Baskaran, P. Smereka, J. Appl. Phys. **111**, 044321 (2012)
53. Z. Ding et al., J. Appl. Phys. **121**, 055305 (2017)
54. P. Michelato et al., in *EPAC 1996,* Sitges, Spain (1996), p. 1510
55. P. Michelato, Nucl. Instr. Meth. Phys. Res. A **393**, 464 (1997)
56. J. Teichert et al., Report on photcathodes. Report no. CARE/JRA-PHIN, July 2004, 2004
57. S.H. Kong et al., J. Appl. Phys. **77**, 6031 (1995)
58. C. Hessler et al., in *IPAC15,* Richmond, VA, USA (2015), p. 1699
59. B.M.V. Oerle, G.J. Ernst, Nucl. Instr. Methods Phys. Res. A **358,** 287 (1995)
60. S.H. Kong et al., Nucl. Instr. Methods Phys. Res. A **358**, 272 (1995)
61. E. Chevallay et al., Production of a high average current electron beam with Cs_2Te photocathode. Report no. CTF3 Note 020 (2001)
62. S. Lederer et al., in *FEL2011* Shanghai, China (2011), p. 511
63. A.R. Fry et al., in *XIX International Linac Conference,* ed. by F.N.A. Laboratory FERMILAB-Conf-98/272, Chicago, IL, USA (1998), p. 642
64. P.J.M.V.D. Slot et al., in *FEL 2008,* Gyeongju, Korea (2008)
65. R. Xiang et al., in *IPAC2012,* New Orleans, LA, USA (2012), p. 1524
66. R. Xiang et al., Phys. Rev. ST Accel. Beams **13**, 043501 (2010)
67. E. Prat et al., in *FEL14,* Basel, Switzerland (2014), p. 970
68. D. Sertore et al., Nucl. Instr. Meth. Phys. Res. A **445**, 422 (2000)
69. S. Lederer et al., in *IPAC'10,* Kyoto, Japan (2010), p. 2155
70. D.H. Dowell et al., Nucl. Instr. Methods Phys. Res. A **356**, 167 (1995)
71. B. Dunham et al., Appl. Phys. Lett. **102**, 034105 (2013)
72. J. Smedley et al., in *International Particle Accelerator Conference,* Richmond, VA, USA (2015), p. 1965
73. M. Ruiz-Osés et al., APL Mater. **2** (2014)
74. S. Schubert et al., APL Mater. **1**, 032119 (2013)
75. J. Feng et al., J. Appl. Phys. **121**, 044904 (2017)
76. C. Ghosh, B.P. Varma, J. Appl. Phys. **49**, 4549 (1978)
77. H.V.D. Graaf et al., Nucl. Instr. Methods Phys. Res. A **847,** 148 (2017)
78. S. Schubert et al., J. Appl. Phys. **120**, 035503 (2016)
79. J. Kessler, *Polarized Electrons*, 2nd edn. (Springer, Berlin, 1985)
80. E. Leader, *Spin in Particle Physics* (Cambridge University Press, 2001)
81. M. Kuwahara et al., Appl. Phys. Lett. **105**, 193101 (2014)
82. M.J. Alguard et al., Nucl. Instrum. Meth. Phys. Res. A **163**, 29 (1979)
83. W.V. Drachenfels et al., Nucl. Instrum. Meth. Phys. Res. A **140,** 47 (1977)
84. M.J. Alguard et al., Phys. Rev. A **16**, 209 (1977)
85. L.A. Hodge et al., Rev. Sci. Instrum. **50**, 1 (1979)

86. P.F. Wainwright et al., Rev. Sci. Instrum. **49**, 571 (1978)
87. D.T. Pierce et al., Appl. Phys. Lett. **26**, 670 (1975)
88. C.Y. Prescott et al., Phys. Lett. B **77**, 347 (1978)
89. E. Tsentalovich et al., Nucl. Instrum. Methods Phys. Res. A **582**, 413 (2007)
90. M.J.J.V.D. Putte et al., in *AIP Conference Proceedings* (1998), p. 260
91. K. Aulenbacher et al., Nucl. Instrum. Methods Phys. Res. A **391**, 498 (1997)
92. W. Hillert et al., in *AIP Conference Proceedings (Proceedings of the 14th International Symposium on High Energy Spin Physics (SPIN2000))*, ed. by A.C. Proc (2001), p. 961
93. K. Wada et al., in *AIP Conference Proceedings (Proceedings of the 15th International Symposium on High Energy Spin Physics (SPIN2002))*, ed. by A.C. Proc. (2003), p. 1063
94. C.K. Sinclair et al., Phys. Rev. ST Accel. Beams **10**, 023501 (2007)
95. J.S. Blakemore, J. Appl. Phys. **53**, R123 (1982)
96. D.T. Pierce, F. Meier, Phys. Rev. B **13**, 5484 (1976)
97. L.W. James et al., J. Appl. Phys. **42**, 4976 (1971)
98. T. Maruyama et al., Appl. Phys. Lett. **55**, 1686 (1989)
99. R. Alley et al., Nucl. Instrum. Methods Phys. Res. A **1**, 365 (1995)
100. M.I. D'yakonov, V.I. Perel, ZhETF (in Russian) **60,** 1954 (1971; translated as Sov. Phys. JETP **33,** 1053 (1971))
101. G. L. Bir et al., ZhETF (in Russian) **69,** 1382 (1975; trans. Sov. Phys. JETP 42, 705, 1976)
102. F. Meier, B.P. Zakharchenya, *Optical Orientation* (North-Holland, Amsterdam, 1984)
103. P.H. Song, K.W. Kim, Phys. Rev. B **66**, 035207 (2002)
104. Y. Yafet, *Solid State Physics,* vol. 14 (Academic, New York, 1963), p. 1
105. G. Fishman, G. Lampel, Phys. Rev. B **16**, 820 (1977)
106. K. Zerrouati et al., Phys. Rev. B **37**, 1334 (1988)
107. J.H. Jiang, M.W. Wu, Phys. Rev. B **79**, 125206 (2009)
108. I. Zutic et al., Rev. Mod. Phys. **76**, 323 (2004)
109. A.G. Aronov et al., ZhETF (in Russian) **57,** 680 (1983; trans. as Sov. Phys. JETP **57,** 680 (1983))
110. W. Liu et al., J. Appl. Phys. **122**, 035703 (2017)
111. J.L. McCarter et al., Nucl. Instrum. Methods Phys. Res. A **618**, 30 (2010)
112. B.V. Zeghbroeck, *Principles of Semiconductor Devices* (e-book, University of Colorado, 2011)
113. T. Nakanishi et al., Phys. Lett. A **158**, 345 (1991)
114. T. Maruyama et al., Phys. Rev. Lett. **66**, 2376 (1991)
115. T. Maruyama et al., Phys. Rev. B **46**, 4261 (1992)
116. Bandwidth Semiconductor, http://www.bandwidthsemi.com. Accessed 8 Feb 2019
117. T. Omori et al., Jpn. J. Appl. Phys. **33**, 5676 (1994)
118. T. Maruyama et al., Appl. Phys. Lett. **85**, 2640 (2004)
119. http://www.svta.com/. Accessed 8 Feb 2019
120. A.V. Subashiev et al., Appl. Phys. Lett. **86**, 171911 (2005)
121. X. Jin et al., Appl. Phys. Lett. **105**, 203509 (2014)
122. Y.A. Mamaev et al., Appl. Phys. Lett. **93**, 081114 (2008)
123. X. Jin et al., Appl. Phys. Express **6**, 015801 (2013)
124. E.C. Aschenauer et al., eRHIC design study: an electron-ion collider at BNL. arXiv:1409.1633 (2014). https://arxiv.org/abs/1409.1633
125. K. Aulenbacher et al., U. Mainz, Operating experience with the MAMI polarized electron source. Report no. SLAC Report no. 432 (revised) (1993)
126. T. Saka et al., Jpn. J. Appl. Phys. **32**, 1837 (1993)
127. J.C. Grobli et al., Phys. Rev. Lett. **74**, 2106 (1995)
128. L.G. Gerchikov et al., AIP Conf. Proc. **980**, 124 (2008)
129. W. Liu et al., Phys. Lett. **109**, 252104 (2016)
130. S. Adachi, *Properties of Semiconductor Alloys: Group-IV, III–V and II–VI Semiconductors* (Wiley, 2009), p. 238
131. M. Levinshtein et al., *Handbook Series on Semiconductor Parameters* (World Scientific, 1996)

132. I. Vurgaftman et al., J. Appl. Phys. **89**, 5815 (2001)
133. H.A. Macleod, *Thin-Film Optical Filters*, 4th edn. (CRC Press, Baton Rouge, 2010)
134. P. Peumans et al., J. Appl. Phys. **93**, 3693 (2003)
135. J. Zou et al., J. Appl. Phys. **105**, 013714 (2009)
136. J. Zou et al., J. Appl. Phys. **101**, 033126 (2007)
137. J. Grames et al., Phys. Rev. ST Accel. Beams **14**, 043501 (2011)
138. M. Reiser, *Theory and Design of Charged Particle Beams,* 2nd edn. (Wiley-VCH Verlag & Co. KGaA, Weinheim, New-York, 2008), Beam Physics and accelerator technology
139. http://www.srim.org. Accessed 8 Feb 2019
140. W. Liu et al., Appl. Phys. Lett. **109** (2016)
141. K.L. Jensen et al., J. Appl. Phys. **99**, 124905 (2006)
142. M. Gaowei et al., Phys. Rev. Accel. Beams **22**, 073401 (2019)
143. M. Gaowei et al., Appl. Phys. Lett. **5**, 116104 (2017)

Chapter 8
A Thermal-Field-Photoemission Model and Its Application

Kevin L. Jensen

All things in common nature should produce/Without sweat or endeavor.

—W. Shakespeare (*The Tempest*, Act II, Scene 1, Lines 133–134.)

Abstract Traditionally electron sources are characterized as *thermal*, *field*, and *photoemission* cathodes (a fourth, secondary, is not considered here), each governed by a canonical emission equation (Richardson—Laue—Dushman, Fowler—Nordheim, and Fowler—DuBridge, respectively) for current density. Modern electron sources operate such that more than one regime contributes because factors like heating and asperities exist. In this chapter, a *single* emission equation is developed that recovers the canonical equations in the appropriate asymptotic limits. Properties important to the formation of electron beams, such as emittance, Nottingham heating, and emission from protrusions, are examined.

8.1 Electron Emission: Sources and Uses

Electron sources produce beams or bunches of electrons with the desired characteristics for linear beam devices used in microwave amplifiers, radar, communications, and particle accelerators. The metrics of performance are bunch charge, pulse duration, energy spread, and beam brightness. The variety of sources reflects the variety of demands placed on them, but the metrics have a correspondence to current density, transconductance (variation of current with extraction voltage or field), emission mechanism, and emittance (tendency of the beam to diverge). Thermal emission cathodes are most widely used in Traveling Wave Tubes (TWTs), klystrons, and other microwave amplifiers. Field emission cathodes are sought for compact and

K. L. Jensen (✉)
Code 6362, MSTD Naval Research Laboratory, Washington, DC 20375, USA
e-mail: kljensen@ieee.org

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*, Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_8

high-frequency amplifiers and accelerators. Photocathodes are preeminent in particle accelerators and Free Electron Lasers (FELs).

Each emission mechanism affects the metrics of beam quality differently as a consequence of the emission process itself or the nature of the emission surface. Generally, the metrics are current density, emittance, and energy spread. By focusing on the theoretical models similar to each mechanism, and then how those models depend on the nature of the emitter surface, an understanding of how they behave and interact as conditions change is obtained. Device modeling in the design of practical applications using particle-in-cell codes and other methods under the generic heading of "beam optics codes" typically and generally rely on a characteristic equation for each of the mechanisms. This is an astonishingly unfortunate circumstance, requiring tedious and hand-crafted solutions to treat each mechanism, and disallowing the evolution of one emission mechanism into the other (such as thermal into the field) as can occur on the same emitter at the same emission site under different operating regimes. The emission equations themselves suggest an underlying commonality that it is the purpose of this chapter to explore, and to do so with an eye toward enabling a single compact emission model suitable for usage by beam optics codes used to design devices. After that equation is developed, a simple model of geometric effects is used to demonstrate how the model may be applied in practice. Equations in this chapter are written using the International System of Quantities (ISQ) [1] (formerly called the "MKSA system").

Emittance, possibly the least familiar of the metrics, warrants an introduction. It measures the tendency of a beam to spread, and its presence becomes increasingly problematic for high-power and high-frequency linear beam devices and particle accelerators when a reduction in physical dimensions is sought [2, 3], or when very high brightness beams are demanded [4]. Thermionic cathode current densities are kept below about 1 A/cm^2 because their lifetime shortens above this, whereas high-power devices require in excess of that. Such, then, is the interest in non-thermionic sources, but their consideration necessitates an understanding of the effects of emittance on the beam dynamics [5]. The Beam Envelope equation for the $R(z)$, the cross-sectional radius of an electron beam, is given by [6]

$$\frac{d^2}{dz^2}R + \left(\frac{qB}{2\beta\gamma mc}\right)^2 R - \frac{2I_a}{(\gamma\beta)^3 I_o}\frac{1}{R} - \frac{\varepsilon^2}{R^3} = 0 \tag{8.1}$$

where q is the elementary (positive) charge, m is the electron rest mass, c is the speeed of light in vacuum, B is magnetic field (i.e., magnetic flux density), I_a is beam current, ε is the emittance, and I_o is a constant with the value 17.045 kA. The factors in the above equation are called the magnetic term, the space charge term, and the emittance term, respectively. The relativistic factors $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$ are related to the electron velocity v and the beam kinetic energy K_b by $\gamma mc^2 = mc^2 + K_b$, and

$$(\beta\gamma)^2 = \frac{2K_b}{mc^2}\left(1 + \frac{2K_b}{mc^2}\right). \tag{8.2}$$

Thus, to leading order, $\beta\gamma \approx \sqrt{2K_b/mc^2}$. If f is the frequency of a pulsed beam, then $2\pi f nR/c =$ a zero of the Bessel function, with n an integer related to mode, so that frequency is inversely proportional to beam radius, and the remaining factors have their usual interpretation. For Brillouin flow, the sum of the magnetic, space charge, and emittance terms vanishes, yielding

$$\frac{I_a}{\pi R^2} = \frac{I_o}{8\pi}\left(\frac{2K_b}{mc^2}\right)^{1/2}\left(\frac{qB}{mc}\right)^2\left\{1 - \frac{8mK_b\varepsilon^2}{q^2B^2R^4}\right\} \tag{8.3}$$

The ratio of beam current to beam area defines a beam current density $J_{\text{beam}}(\varepsilon)$, which can be written as follows:

$$J_{\text{beam}}(\varepsilon) = J_{\text{beam}}(0)\{1 - \delta(\varepsilon)\} \tag{8.4}$$

The effect of emittance is thus to reduce J_{beam} by the fractional amount $\delta(\varepsilon) = (8mK_b/q^2B^2)(\varepsilon^2/R^4)$. Because technological limitations constrain the other parameters of the beam envelope equation, pursuit of higher frequency devices means that emittance must be addressed and reduced for high-power mmw-to-THz devices [7]. Coupled with a demand for high current density in the beam tunnel, the general demand is for high brightness sources, where $B = J/\varepsilon^2$ [6, 8, 9].

For example, a generic L-band TWT in which a beam current of 300 mA is accelerated to 10 kV, then injected into a beam tunnel 0.1 cm in radius (giving beam current density = 9.17 A/cm^2) and confined by a magnetic flux density of 90 mT (equivalent to 0.9 kG) entails $\varepsilon \leq 13.6$ mm mrad; this requirement is readily met by existing thermionic cathodes. Compare these values to those characteristics of a high-frequency serpentine TWT [10] or an extended interaction Klystron [11], for which the beam current is in the order of 100 mA, the beam tunnel current densities are 100–225 A/cm^2, the acceleration potentials are comparable to 25 kV, the magnetic flux densities range from 100 to 300 mT (equivalent to 1–3 kG), and the beam tunnel radius is approximately 130 μm. A requirement that $\delta < 1\%$ entails that $\varepsilon \leq 0.5$ mm mrad. Alternately, an emittance of $\varepsilon_{rms} = 11.55$ mm mrad effectively doubles the peak PPM magnetic field required in an X-Band circuit, a conclusion based on simulations by Whaley [5] using field emitters as the electron source.

Surface roughness causes field variation across the cathode face. The nodular nature of scandate cathode surfaces [12], for example, has additional consequences for high-frequency devices. Field emitters have the potential for providing large current densities without beam compression [3] and show emittance characteristics that are promising [13, 14], making models of their emittance desirable but correspondingly more difficult [15].

For short wavelength [16, 17] and high average power [18] Free Electron Lasers (FELs), good gain requires the beam emittance to be smaller than the photon beam emittance $\lambda/4\pi$. Understanding the interplay between emittance and space charge is,

therefore, a concern for cathode R&D [17], particularly in simulation codes [19]. For photocathodes, quantum efficiency and emittance are intertwined [20, 21]. Emittance at the photocathode scales with the emission radius, naively suggesting that emittance reduction can be obtained by focusing the drive laser to smaller dimensions, but a limit is imposed by space charge forces [22]. Fields at the surface of a photocathode are comparable to 20–100 MV/m in order to rapidly accelerate the electron bunches, whereas an electron bunch containing 1 nC emitted from an area $r = 2$ mm in radius near the surface of the cathode gives a space charge field comparable to 4.5 MV/m. Therefore, the intrinsic emittance of photocathodes are an impediment to short wavelength FELs and Energy Recovering Linac (ERL) sources of X-rays, motivating a need for models of cathode emittance [17, 23] with surface roughness a probable contributor [24, 25].

Models of surface roughness are also applicable to understanding dark current, usually due to field emitted electrons accelerated at wrong parts of the rf cycle, contributing to halo formation, cathode damage when redirected back to the photocathode, and the creation of secondaries when striking the sidewalls of an accelerator [19]. Field emission is not only associated with the cathode, but can also arise elsewhere in high gradient accelerator structures and rf cavities where field emission sites are generated by various mechanisms [26–28].

Simplifying assumptions ease the modeling of various electron emission mechanisms in Particle-in-Cell (PIC) simulation codes, often by avoiding having to follow individual electron trajectories. Methods include focusing on the most energetic electrons [29], using macro-particles [30], spreading the charge along the trajectory [31], and using an effective field enhancement factor on an otherwise flat cathode surface [32]. Such approximations are advantageous when the simulation region is large and many emission mechanisms compete [19]. Although the impact of field enhancement and surface variation can be explicitly considered in the PIC framework [33], such studies are often done for single tips or a very small number of tips in close proximity because of limitations imposed by the cell size of the simulations. The minimum cell size on the cathode is crudely governed by a product of factors related to the ability to resolve the electron beam and small structures within it: typically, a scale factor of 10^4 exists between the largest (overall device) and smallest (near the cathode) grid elements. Thus, devices measuring 10's of cm use grids on the cathode of 10's of micrometers. If the electron source exhibits variation on a smaller scale—and they often do—simulations will not encompass many emitters or emission sites. This creates an imperative to develop, for a multidimensional emitter subject to various emission mechanisms, an analytically tractable model that can serve as a model for computer codes, particularly if space charge is a problem (as it generally is, but particularly for field emission [34–36]). The need for several emission mechanisms potentially operating simultaneously and from small structures motivates the models considered here.

8.2 Emission Mechanisms

By the mid1930s, the fecund interaction between statistical physics and quantum mechanics spawned equations that were successful in describing the emission of electrons due to heating (thermal emission[1]), the application of strong electrostatic fields (field emission), or following the illumination of a surface with light (photoemission).[2] They are used in the calculation of current density (often in the units of A/cm^2). The equations are most often referred to by their progenitors, although modern usage has deviated in the representation of them. The equations, and how they shall be used forthwith, are as follows, using the notations established in Table 8.1. Because atomic scale units [eV, fs, nm] are used here, where the electron charge is $-q$, a parameter $F = -q\mathcal{E}$ (called here the "classical electrostatic force") is used in place of the conventional electrostatic field $\mathcal{E}$. In field electron emission, the classical electrostatic field $\mathcal{E}$ is negative, so F is positive. We also use electron potential energies U, rather than classical electrostatic potentials, so that, for example, U and the work function Φ have the same units. (These changes from historically earlier practice explain the occasional tendency to refer to F as "field" and U as "potential"—a related tendency exists, for example, in referring to "weight" when mass is under discussion, as in common discourse). In units such that the elementary charge q is written as 1 q, as explained below, the quantities F and $\mathcal{E}$ will be the same numerically, and therefore, the distinction (although technically required by modern International Standards) is not rigorously imposed.

- Thermal emission is described by the Richardson–Laue–Dushman equation [40–42]

$$J_{RLD} = \frac{qm}{2\pi^2\hbar^3}(k_B T)^2 \exp\left(-\frac{\phi}{k_B T}\right) \tag{8.5}$$

  where $\phi = \Phi - \sqrt{(q^2/4\pi\varepsilon_0)F} \equiv \Phi - \sqrt{4QF}$. In the original RLD equation, a coefficient λ^{TE} was appended to $J_{RLD}(T)$ to account for wave mechanical reflection: to avoid complication, such a factor is set to unity here.
- Field emission is described here by the Murphy–Good version [43] of the Fowler–Nordheim (FN) equation [44]; this Murphy–Good equation is written as follows:

$$J_{FN}(F) = \frac{qF^2}{16\pi^2\hbar\Phi t(y)^2} \exp\left(-\frac{4}{3\hbar F}\sqrt{2m\Phi^3}v(y)\right). \tag{8.6}$$

[1] Thermal emission often continues to be referred to as "thermionic emission," reflecting a time when the emitted electrons due to heating ("thermions") were thought to be different than those due to strong fields [37]. It is a convention that likely should expire, but which regrettably has not.

[2] Although a fourth process, secondary emission, has received comparable interest and is used in vacuum electronic devices, the physical processes of transport and emission for it are similar to photoemission (the generation mechanism being energetic electrons rather than photons). Therefore, it is not treated separately, as the generation mechanisms are not under study here. See [38], or [39] for greater detail.

The special mathematical functions $v(y)$ and $t(y)$, sometimes called "Schottky–Nordheim barrier functions" [45], account for lowering and rounding of the transmission barrier by the Schottky effect, thereby making the barrier different from the exactly triangular barrier considered by Fowler and Nordheim. The functions $v(y)$ and $t(y)$, when used here, shall be based on the approximation of Forbes and Deane [46, 47] (encountered in (8.41) and (8.42) below), due to their accuracy and simplicity, in terms of which [48]

$$J_{FN}(F) \equiv \frac{A_o}{\Phi t_o^2}\left(\frac{\Phi^2 e^6}{4Q}\right)^{\nu} F^{2-\nu} \exp\left(-B_o \frac{\Phi^{3/2}}{F}\right) \tag{8.7}$$

where $A_o \equiv q/16\pi^2\hbar$, $B_o \equiv (4/3\hbar)\sqrt{2m}$, $\nu = 2B_o Q/3\sqrt{\Phi}$, $t_o = 1 + (1/6e) = 1.06131$, and other factors are as in Table 8.1. A word of caution, however: there are theoretical grounds for observing that the Schottky–Nordheim barrier functions naturally depend mathematically on y^2 rather than y. Forbes has therefore advocated introducing a separate mathematical parameter x (fomerly l'), with $x = y^2$, and has written the functions $v(x)$ and $t(x)$ as functions of x. To avoid any ambiguity, $v(y)$ and $t(y)$ here are as used by Murphy and Good with $y = \sqrt{4QF}/\Phi$.

- Photoemission is described by the Fowler–DuBridge equation [49, 50].

$$J(\hbar\omega) \propto \left(\frac{q}{\hbar\omega}\right) I_\lambda \left[\frac{1}{2}\left(\frac{\hbar\omega - \phi}{k_B T}\right)^2 + \frac{\pi^2}{6}\right] \tag{8.8}$$

Three conventions are used with respect to the notations for current density and barrier height. In the former, the convention herein is to reserve the symbol J for current density, and I for current, as in (8.4). However, it must be emphasized that the emission equations above for J_{FN}, J_{RLD}, and $J(\hbar\omega)$ are strictly *local* current density equations, unlike $J_{\text{beam}}(\varepsilon_{n,rms})$, or the current density reported for large area emitters [45]. Second, insofar as work function Φ is conventionally defined in the absence of an externally applied field, the factor $\phi = \Phi - \sqrt{4QF}$ is to be understood as a *Schottky-factor-lowered barrier height* rather than a work function. Third and finally, for the purposes of continuity with [45], the direction normal to the surface is in the $\hat{z}$-direction in contrast to earlier treatments which reserved the $\hat{x}$ for this purpose, although such a change should not induce hardship.

Although all three equations are rooted in similar concepts involving a nearly free electron model describing the statistical distribution of electrons in a metal and a quantum mechanical tunneling probability tailored to an image-charge-rounded barrier (sometimes called a "Schottky–Nordheim (SN) barrier"), the derivations as commonly given are often idiosyncratic so that the commonality of the physics they share, needed for a general equation of all three, is hidden (but not always, Murphy and Good [43] being a counterexample). There are instances in the literature where thermal and field emission co-exist. Consider two examples.

First, the transition between thermal emission and field emission is starkly expressed in the energy distribution of emitted electrons, as shown in the experimen-

Table 8.1 Fundamental Constants, Parameters, and Terms. Units: length [nm], time [fs], energy [eV], charge [q], temperature [K], with the exception of current density J and light intensity I_λ, for which conventional units are used

Symbol	Definition	Value/Formula	Units
q	Elementary (positive) charge	1	q = 1 eV/V
c	Speed of light	299.792	nm/fs
k_B	Boltzmann's constant	1/11604.5	eV/K
$\hbar$	Planck's constant/2π	0.658212	eV fs
m	Electron rest mass	510999	eV/c^2
ε_0	Electric constant	0.0552635	eV/(nm V^2)
Q	Image factor ($q^2/16\pi\varepsilon_0$)	0.359991	eV nm
T	Thermodynamic temperature	–	K
β_T	Inverse Boltzmann factor	$1/k_BT$	eV^{-1}
μ	Fermi energy	7 (Cu)	eV
Φ	Work function	4.5 (Cu)	eV
F	Classical electrostatic force	$-q\mathcal{E}$	eV/nm
ϕ	Schottky-lowered barrier height	$\Phi - \sqrt{4QF}$	eV
U	Electron potential energy	–	eV
$\hbar\omega$	Photon energy	–	eV
J	Current density	–	A/cm^2
I_λ	Light intensity	–	W/cm^2
A	Ampere	6241.51	q/fs
W	Watt	6241.51	eV/fs

tal data of Gadzuk and Plummer [51] of Fig. 8.1 where, for a given high temperature of 1570 K, at low fields (characterized by emission concentrated near the emission barrier maximum) and at high fields (characterized by the high anode voltage of 1600 V), it is concentrated at the Fermi level. More recent explorations of thermal-field ZrO/W Schottky emitters elegantly reveal analogous behavior [52]. In between as the field rises, the peak location, distribution width, and character of the energy distribution change.

Second, consider the metal and metal+oxide emission properties of Ba-dispenser cathodes measured by Geittner et al. [53], as shown in Fig. 8.2. As the anode field increases, the shape of the emission curve for $\ln(I)$ undergoes a clear transition from the $\sqrt{F}$ dependence associated with the RLD equation to the $1/F$ dependence associated with field emission. Therefore, a single equation to smoothly describe the full range of the current–voltage data is desirable; when photoemission is added [54], the ability to address it in the same framework is attractive.

Therefore, the present treatment shall focus on a description of the features common to the emission equations, namely the supply function and transmission probability, and how the equations themselves emerge from a general formulation of current density that includes regimes describing all three from one master equation. The

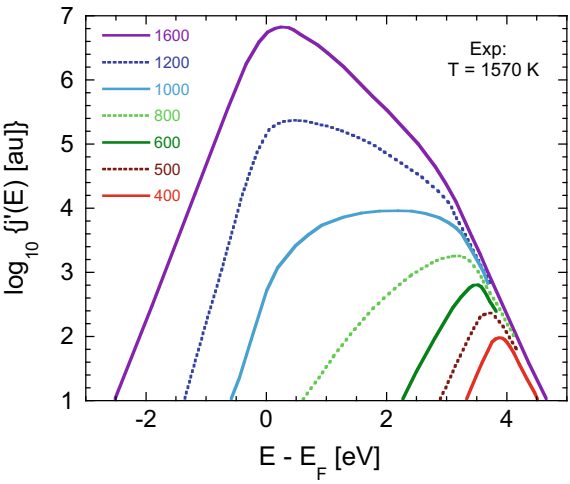


Fig. 8.1 The total energy distributions at various anode voltages for a tungsten field emitter held at 1570 K, based on Fig. (2a) of Gadzuk and Plummer [51]. Their analysis suggested $\mu = 8eV$, $\Phi = 4.8\ eV$, for the Fermi energy and work function, respectively. Labels indicate anode voltage

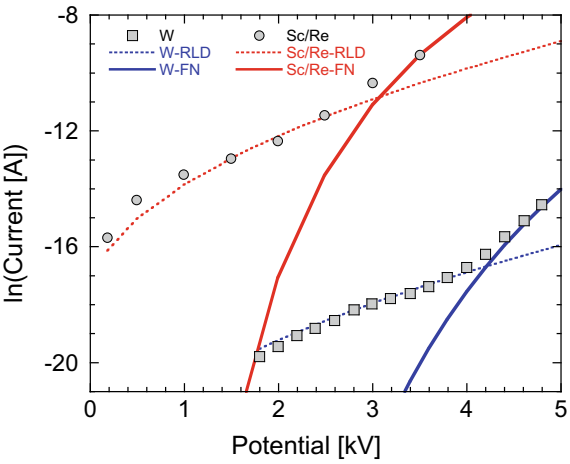


Fig. 8.2 Thermal and field emission from tungsten (W-I) and scandium (Sc/RE-I) cathodes as a function of anode potential. Solid lines are as evaluated from the Murphy–Good FN-type equation. Dashed lines are as evaluated from the Richardson (RLD) equation. Data from [53]

resulting General Thermal-Field-Photoemission (GTFP) equation shall, therefore, incorporate the complexities that attend transitioning from one regime to another. Of necessity, the formulation departs from standard accounts, and so a concise rendition of more complete treatments [48, 55–61] in a manner used by the present analysis is first given. The formulation allows for the consideration of two other emission attributes apart from the current density that affect the quality and utility of electron sources for vacuum electronic devices: emittance and heating effects. Consequently, the analysis shall be extended to them as well.

8.2.1 Transmission Probability

For a gas of point particles, a statistical model of current density indicates that it is a product of the charge of the carrier, the density ρ of the gas of particles, and the average velocity of the particles. If the electrons in a bulk metal can be visualized as such a gas, then the magnitude of the current density is given by [62]

$$J_z = q\rho \langle v_z \rangle \tag{8.9}$$

where $\langle \cdots \rangle$ denotes the statistical average. Classically, the velocity of a particle moving in the $\hat{z}-$direction is p_z/m; quantum mechanically, it is preferable to work with the coordinate k_z which is conjugate to z, or

$$v_z \rightarrow \frac{\hbar k_z}{m} \tag{8.10}$$

In all the emission equations to be developed below, the $\hat{z}-$direction shall be taken as in the direction of the emission current, normal to the emitter surface. Consequently, the statistical average over the distribution $f(\vec{k})$ will generally resemble

$$\langle v_z \rangle = \frac{2}{(2\pi)^3} \int \frac{\hbar k_z}{m} f(\vec{k}) d\vec{k} \tag{8.11}$$

where the factor of 2 in the numerator of the coefficient is for electron spin. The distribution $f(\vec{k})$ is with respect to the *emitted* particles rather than the distribution in the bulk of the emitter material. That distribution will be a product of a measure of all electrons in the bulk material with only those that pass the surface barrier filtered out. The probability that an electron with wavenumber k_z is transmitted is in turn a measure of its probable current density j_k, which is quantum mechanically given by

$$j_k \equiv \frac{\hbar}{2mi} \left(\psi_k^\dagger \partial_z \psi_k - \psi_k \partial_z \psi_k^\dagger \right) \tag{8.12}$$

where $\psi_k(z)$ is the wave function for an electron with energy $E_z = \hbar^2 k_z^2/2m$ that exists past the barrier in vacuum (taken to be on the right hand side for $z > 0$). The probability of emission $D(k_z)$ (i.e., the transmission probability) is therefore

$$D(k_z) \equiv \frac{j_{\text{trans}}}{j_{\text{inc}}} \tag{8.13}$$

where inc and trans refer to the incident and transmitted wave functions, respectively. Alternatively, the transmission probability can be expressed as a function of the normal energy, as $D(E_z)$.

8.2.2 Supply Function

The distribution of electrons in the bulk material is taken to be, in the *nearly free electron model*, a Fermi Dirac distribution $f_{FD}(E)$, where $E = \hbar^2(k_z^2 + k_\perp^2)/2m$, where the $\perp$-subscript denotes the momentum component transverse to the direction of current flow (e.g., $k_\perp^2 = k_x^2 + k_y^2$ in Cartesian coordinates). Consequently,

$$\frac{1}{(2\pi)^2}\int f(\vec{k})d\vec{k}_\perp = D(k_z)\frac{2}{(2\pi)^2}\iint_{-\infty}^{\infty} f_{FD}(E)dk_y dk_z \equiv D(k_z)f(k_z) \quad (8.14)$$

where a factor of 2 for electron spin is included. The term $f(k_z)$ is the *supply function* [38, 39, 41, 43, 63, 64]. The Fermi Dirac distribution is characterized by a chemical potential μ such that

$$f_{FD}(E) \equiv \frac{1}{1 + \exp[(E_z + E_\perp - \mu)/k_B T]} \quad (8.15)$$

where at $T = 0$ K, $\mu = E_F$ is the Fermi energy, and $k^2 = k_z^2 + k_\perp^2$ (equivalently, $E = E_z + E_\perp$). The integrations result in

$$f(k_z) = \frac{mk_B T}{\pi\hbar^2}\ln\left(1 + e^{(\mu - E_z)/k_B T}\right) \quad (8.16)$$

where $E_z = \hbar^2 k_z^2/2m$. The $z-$subscript notation, though required for three-dimensional (3D) integrals associated with the phase space development, will be discontinued when the formalism is restricted to one dimension (1D), at which point the energy E_z takes on the role of the *normal energy* (energy component normal to the surface—see, for example, [65]). Two limits are generally encountered: for field emission, or $(\mu - E_z)/k_B T \gg 1$); and for thermal emission , or $(E_z - \mu)/k_B T \gg 1$, for which

$$f(k_z) \approx \begin{cases} \frac{1}{2\pi}\left(k_F^2 - k_z^2\right) & \text{(field)} \\ \left(\frac{mk_B T}{\pi\hbar^2}\right)\exp\left[-(E_z - \mu)/k_B T\right] & \text{(thermal)} \end{cases} \quad (8.17)$$

although these limits will not generally be used below.

When analyzing the energy distributions of emitted electrons, an alternate formulation [64–69] works with E_z (sometimes W) rather than k_z and speaks therefore of the distribution of normal energies and instead performs integrations over normal energy E_z and perpendicular $E_\perp$ rather than k_z and $k_\perp$. In that approach, electrons crossing the plane with energy components in the range $\mathrm{d}E_\perp$ and $\mathrm{d}E_z$ make a contribution d^2J to the total internal current density J crossing the plane, where d^2J is given by

$$d^2 J = \frac{qm}{2\pi^2\hbar^3} f_{FD}(E)\, dE_\perp\, dE_z \quad (8.18)$$

where $E = E_z + E_\perp$ as before. Noting that transmission probability is independent of $E_\perp$, then the emission current density J is

$$J = \iint D(E_z)\, d^2 J. \tag{8.19}$$

As with the integration over $k_\perp$ above, integration over $E_\perp$ gives

$$J = \int j_n(E_z) dE_z = \int N(E_z) D(E_z) dE_z, \tag{8.20}$$

where $j_n(E_z)$ is the emitted Normal Energy Distribution (NED), and the *incident normal energy distribution* $N(E_z)$ is given by

$$N(E_z) = \frac{qmk_B T}{2\pi^2 \hbar^3} \ln\left(1 + \exp^{-(E_z - \mu)/k_B T}\right) \tag{8.21}$$

and is seen to be related to (8.16) (apart from a factor of $1/2\pi$) as it should be.

Appending a $z-$subscript to E_z and k_z, though required for evaluation of three-dimensional (3D) integrals, will be discontinued when the formalism has been reduced to a one-dimensional integration, as in (8.16) and (8.20) above; for simplicity, the symbols $\hbar k_z/m \to \hbar k/m$ and $E_z \to E$ will then be used instead to denote normal velocity and normal energy.

8.2.3 Gamow Factor and the Kemble Approximation

Few tunneling barriers have analytical solutions for the transmission probability. Of the simple ones that do, the rectangular, triangular, and quadratic potentials can be parametrically given by

- Rectangular: $U_{rec}(z) = U_o \Theta(L_o - z)\Theta(z)$
- Triangular: $U_{tri}(z) = (U_o/L_o)\,(L_o - z)\,\Theta(z)$
- Quadratic: $U_{quad}(z) = (U_o/L_o^2)\left(L_o^2 - 4z^2\right)$

where $\Theta(z)$ is the Heaviside step function (unity for positive argument, and zero otherwise). For the triangular barrier in particular, $L_o \equiv U_o/F$, where $F = -q\mathcal{E}$ and $\mathcal{E}$ is the (negative) electrostatic field at the surface. Each of these simple barriers is characterized by a Gamow tunneling factor $\theta(E_z)$ and an analytic form for $D(E_z)$. Although not strictly accurate, it is the case that for all of them, the transmission probability $D(E_z)$ can be well approximated by the so-called "Kemble formalism" [43, 57, 70–73]. This formalism is exact for the quadratic barrier.

Specifically, for the one-dimensional tunneling problem, the Gamow factor is defined by

$$\theta(E_z) \equiv 2\frac{\sqrt{2m}}{\hbar} \int_{z_-}^{z_+} \sqrt{U(z) - E_z}\, dz \tag{8.22}$$

and the Kemble formalism takes the transmission probability $D(E_z)$ to be given in terms of $\theta(E_z)$ by

$$D(E_z) \approx \frac{1}{1+\exp(\theta(E_z))} \tag{8.23}$$

where the limits of integration are defined by where $U(z_\pm) = E_z$ or where an abrupt transition occurs such that the wave function $\psi_k(z)$ passes from over to under (or vice versa) the barrier.

Because the continued use of an "z" subscript on the normal energy component E_z may create the erroneous expectation that it is z-dependent (rather than a component in the $\hat{z}$-direction), for the remainder of the treatment of the Gamow factor and the following section on the Schottky–Nordheim barrier (Sect. 8.2.4) and sections beyond, suppress the z-subscript on both E_z and k_z when the treatment is understood as only being in one dimension.

The Kemble approximation will form the basis for all of what follows. The Gamow factors for each of the simple barriers above can be redefined in terms of a shape factor σ and a parameter u related to its derivative. On introducing the dimensionless factor s such that

$$s = \frac{z - z_-(E)}{z_+(E) - z_-(E)} \equiv \frac{z - z_-(E)}{L(E)} \tag{8.24}$$

then σ and u are given by [60]

$$\sigma = \int_0^1 \left(\frac{U(s) - E}{U_o - E}\right)^{1/2} ds \tag{8.25}$$

$$u = \int_0^1 \left(\frac{U_o - E}{U(s) - E}\right)^{1/2} ds \tag{8.26}$$

where U_o is the maximum value of the barrier PE.

On introducing the factor $\kappa_{rec}(E) \equiv \sqrt{2m(U_o - E)}/\hbar$, which applies to the rectangular barrier, the Gamow factors for the simple barriers can be written as

$$\theta_{sh}(E) = \sigma_{sh}\kappa_{rec}(E)L(E) \tag{8.27}$$

where the "sh" subscript refers to the kind of shape factor (rec, tri, or $quad$). Although σ and u will become more complex, the same general form holds for more complex potentials, and in particular, for the Schottky–Nordheim (SN) barrier. For the simple barriers, the shape factors are constant and independent of energy E: direct evaluation shows them to be as given in Table 8.2 for the barrier shapes defined by the list at the beginning of Sect. 8.2.3 and for the SN barrier defined by (8.34). The Gamow factors can be directly evaluated for the simple barriers, and doing so gives

Table 8.2 Values of σ (8.25) and u (8.26) for the simple barriers and the Schottky–Nordheim (SN) barrier of (8.34)

Barrier	σ	u
Rectangular	1	1
Triangular	2/3	2
Quadratic	$\pi/4$	$\pi/2$
SN	(8.39)	(8.40)

$$\theta_{rec} = \frac{2L_o\sqrt{2m}}{\hbar}(U_o - E)^{1/2} \tag{8.28}$$

$$\theta_{tri} = \frac{4L_o\sqrt{2m}}{3\hbar U_o}(U_o - E)^{3/2} \tag{8.29}$$

$$\theta_{quad} = \frac{\pi L_o\sqrt{2m}}{2\hbar\sqrt{U_o}}(U_o - E) \tag{8.30}$$

Observe in particular that L_o/U_o in the coefficient of θ_{tri} is the same as $1/F$.

The GFTP Equation relies on the Kemble formula for $D(E)$ but uses a Taylor expansion of $\theta(E)$ that is linear in normal energy E. Using the shape factors, it can be shown that for $\theta(E)$ expanded about $E \approx E_m$ (where the notation and interpretation of E_m will become clear below), then

$$\theta(E) = 2\sigma(E)\kappa_o(E)L(E) = \sigma(E)\theta_{rec}(E) \tag{8.31}$$

$$\approx \theta_{rec}(E_m)\left[\sigma - \frac{u}{2}\left(\frac{E - E_m}{U_o - E_m}\right)\right] \tag{8.32}$$

For the historically important case of the exactly triangular barrier used in the original zero-temperature Fowler–Nordheim analysis [44] (without an image-charge modification), and defined such that $E_m = \mu$ and $U_o = \mu + \Phi$, this results in

$$\theta_{tri}(E) \approx \frac{4\sqrt{2m\Phi^3}}{3\hbar F} - 2\frac{\sqrt{2m\Phi}}{\hbar F}(E - \mu) \tag{8.33}$$

8.2.4 *The Schottky–Nordheim Barrier*

The Schottky–Nordheim barrier PE $U_{SN}(z)$ is defined by

$$U_{SN}(z) = \mu + \Phi - Fz - \frac{Q}{z} \tag{8.34}$$

The location z_o of the barrier maximum, the zeros $z_\pm(E)$, and the barrier width $L(E)$ are all given by

$$z_o = \sqrt{Q/F} \tag{8.35}$$

$$z_\pm(E) = \frac{1}{F}(\mu + \Phi - E) \pm \frac{1}{2}L(E) \tag{8.36}$$

$$L(E) = \frac{1}{F}\sqrt{(\mu + \Phi - E)^2 - 4QF} \tag{8.37}$$

in terms of which

$$U_{SN}(z) = U_o - Fz\left(1 + \frac{z_o^2}{z^2}\right) \tag{8.38}$$

where $U_o = U(z_o) = \mu + \Phi - \sqrt{4QF} = \mu + \phi$.

Clearly, therefore, when z_o is small compared to z for much of the barrier (as occurs for low classical force F), then U_{SN} approximates the exactly triangular barrier. Conversely, when z_o is of a similar size to z for much of the barrier (as occurs for high classical force F), then the barrier has a more quadratic-like shape to it. Therefore, the SN barrier PE moves between the triangular and quadratic barriers in the limits of low and high classical force, and that in turn entails that for the SN barrier both σ and u become classical-force-dependent. In fact, a quadratic interpolating polynomial as a function of $y = \sqrt{4QF}/\Phi$, such that at $y = 0$ the triangular barrier limit is obtained and at $y = 1$ the quadratic barrier is obtained, is possible [60], the performance of which is shown in Fig. 8.3. In terms of the special mathematical functions $v(y)$ and $t(y)$ that are familiar from (8.6), both $\sigma(y)$ and $u(y)$ can be exactly expressed as

$$\sigma_{SN}(\mu) = \frac{2v(y)}{3(1-y)\sqrt{1+y}} \tag{8.39}$$

$$u_{SN}(\mu) = \frac{2t(y)}{\sqrt{1+y}}. \tag{8.40}$$

The most useful approximate forms for $v(y)$ and $t(y)$, due to Forbes and Deane [46, 47], are

$$v(y) = 1 - \frac{y^2}{3}(3 - \ln(y)) \tag{8.41}$$

$$t(y) = 1 + \frac{y^2}{9}(1 - \ln(y)) \tag{8.42}$$

in terms of which for the special case of $E_m = \mu$ (compare (8.33))

$$\theta_{SN}(E) \approx \frac{4\sqrt{2m\Phi^3}}{3\hbar F}v(y) - 2\frac{\sqrt{2m\Phi}}{\hbar F}t(y)(E - \mu) \tag{8.43}$$

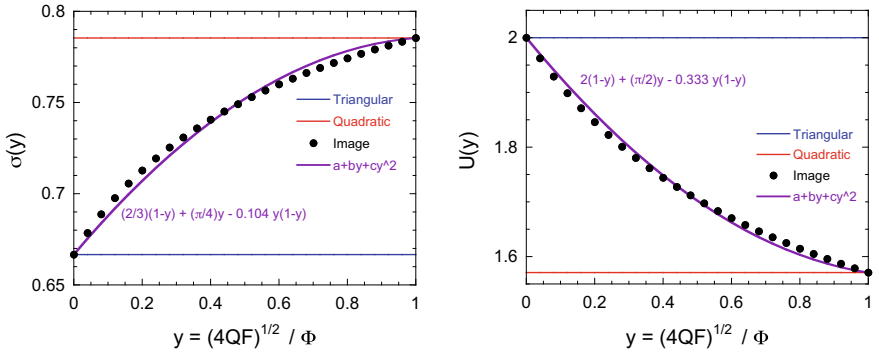


Fig. 8.3 The Schottky–Nordheim (or "image-charge") barrier shape factors for $E = \mu$, compared to the constant asymptotic factors for the triangular and quadratic barriers, after [60]

This equation, along with (8.30) for $\theta_{quad}(E)$ (which is already linear in E), figure prominently in the simplest implementation of the GTFP equation. Its form enables a transparent recovery of the Murphy and Good version of $J(F)$ due to field emission, as given by (8.6), and forms the most rapid method of computation including thermal-field effects (and is identified as the *original* GTF, or *o*GTF, method). The shape factor method behind (8.31), however, allows a linear approximation to $\theta(E)$ to be developed [74] without reliance on the Schottky–Nordheim functions, and is required to correct deficiencies that arise when transitioning from thermal-dominated to field-dominated emission [75], but at the expense of greater computational effort (and is identified as the *reformulated* GTF, or *r*GTF, method).

8.3 Local Emission Current Density

8.3.1 Energy Slope Factors

Parameters called here *energy slope factors* are critical for the development of the GTFP equation. In relation to the exponent that appears in the expression for the incident normal energy distribution $N(E)$ of (8.21), one can define a coefficient β_T (the "energy slope factor for temperature") by

$$\beta_T = -\partial_E[-(E-\mu)/k_BT] = 1/k_BT. \tag{8.44}$$

By analogy, an energy slope factor for classical electrostatic force can be defined in relation to the Gamow factor that appears in the exponent in the expression for transmission probability by

$$\beta_F(E) \equiv -\partial_E\theta(E) = \frac{2mL(E)}{\hbar^2\kappa(E)}u(E) \tag{8.45}$$

where, to emphasize the point, $\kappa(E) = \sqrt{2m(U_o - E)}/\hbar$ and $L(E) = z_+(E) - z_-(E)$. As a result of the energy dependence, $\beta_F(E)$ *is not constant for the SN barrier*, although the variation is well approximated by a cubic equation. The cubic equation is made well approximated by insuring that it gives $\theta(E)$ and its derivative $\beta(E)$ at $E = \mu$ and $E = \mu + \phi$ ($p = 0$ and 1, respectively), and that likewise gives $\beta(E)$. The cubic which does so is

$$\theta(E(p)) = (1-p)^2(2p+1)\theta(\mu) - \phi p(1-p)\left\{(1-p)\beta_F(\mu) - p\beta_F(\mu+\phi)\right\} \tag{8.46}$$

where $E(p) \equiv \mu + p\phi$ and $\phi = \Phi - \sqrt{4QF}$ as in Table 8.1. Attention is restricted to $0 \le p \le 1$ (for $p < 0$ and $p > 1$, the conventional FN and RLD equations are adequate). The coefficients of the cubic equation are (recalling $y \equiv \sqrt{4QF}/\Phi$)

$$\theta(\mu) = \frac{4}{3\hbar F}\sqrt{2m\Phi^3}v(y) \tag{8.47}$$

$$\theta(\mu+\phi) = 0 \tag{8.48}$$

$$\beta_F(\mu) = \frac{2}{\hbar F}\sqrt{2m\Phi}t(y) \tag{8.49}$$

$$\beta_F(\mu+\phi) = \frac{\pi}{\hbar F}\sqrt{m\Phi}\sqrt{y} \tag{8.50}$$

where $\beta_F(\mu+\phi)$ is found by finding the best match between $V_{img}(z)$ and $V_{quad}(z)$ at $z \approx z_o$ then using the quadratic barrier results for $E \to \mu + \phi$.

The final approximation is to use a "linearized" Kemble approximation in which[3]

$$D(E) = \frac{1}{1 + \exp\left[\beta_F(E_o - E)\right]} \tag{8.51}$$

where $\theta(E)$ is the Gamow factor encountered in Sect. 8.2.3. Near the Fermi level $E \approx \mu$, then θ is large, and the Fowler–Nordheim equation follows; conversely, near $E \approx \mu + \phi$, then θ is small, and $D(E)$ is comparable to unity (but not quite—when $\theta = 0$, $D \to 1/2$, a point that requires revisiting below) so that the Richardson equation follows. The Kemble approximation is useful in that an approximation to it based on its *extremes* yields a convenient representation of its *middle*, the extremes being the form of the transmission probability $D(E)$ in the thermal and field limits. Note the guarded phrase "convenient representation": the approximation that shall be used here to capture the total energy distribution behavior of the previous section will not in fact be good—but it will suggest a path to get to an approximation that is. The first step to doing so, familiar from the derivation of the Fowler Nordhiem equation, is to consider a linear form of the Gamow factor

$$\theta(E) \approx \beta_F(E_m)\,(E_o - E) \tag{8.52}$$

[3]Remember that E is standing in for the normal energy component E_z when 1D equations are understood.

where E_m is the location of the maximum of the current density integrand and

$$E_o = E_m + \frac{\theta(E_m)}{\beta_F(E_m)} \tag{8.53}$$

In particular, for thermal ($E_m \approx \mu + \phi$) and field ($E_m \approx \mu$) conditions,

$$E_o^{therm} = \mu + \phi \tag{8.54}$$

$$E_o^{field} = \mu + \Phi \frac{2v(y)}{3t(y)} \tag{8.55}$$

8.3.2 Formulation

Use of the energy slope factors β_T and β_F along with the linearized Kemble approximation means the most general current density equation is given by

$$J_{GTF}(F,T) = \frac{qm}{2\pi^2\beta_T\hbar^3}\int_0^\infty \frac{\ln\{1+\exp[\beta_T(\mu-E)]\}}{\{1+\exp[\beta_F(E_o-E)]\}}dE \tag{8.56}$$

$$= A_{RLD}T^2 N\left[\frac{\beta_T}{\beta_F}, \beta_F(E_o-\mu)\right] \tag{8.57}$$

Observe that this form assumes the Kemble form is sufficient; even for the simple rectangular and triangular barriers, complications arise which introduce energy-dependent coefficients and thereby insert prefactors into the emission equations [57, 76], but which shall not be considered further here. Introduce

$$n(F,T) \equiv \frac{\beta_T}{\beta_F} \tag{8.58}$$

$$s(F,T) \equiv \beta_F(E_o-\mu) \tag{8.59}$$

and define

$$N(n,s) = n\int_{-\infty}^{\infty} \frac{\ln\left[1+e^{n(x-s)}\right]}{1+e^x}dx \tag{8.60}$$

where the upper limit has replaced $\beta_F\mu \to \infty$. The relations

$$N(n,s) + N(n,-s) = \frac{1}{2}n^2s^2 + \zeta(2)\left(n^2+1\right) \tag{8.61}$$

$$N(1,s) = (s+1)e^{-s} \tag{8.62}$$

where $\zeta(2) = \pi^2/6$ (Riemann Zeta function) will prove convenient when including photoemission and the transition region, respectively. The evaluation of $N(n, s)$ is involved [57, 58], but simple approximations are available. Introducing

$$\Sigma(x) = 1 + 2\sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{(j/x)^2 - 1} \tag{8.63}$$

$$\approx \frac{1+x^2}{1-x^2} + \left(\frac{\pi^2}{6} - 2\right) x^2 + \left(\frac{7\pi^4}{360} - 2\right) x^4 \tag{8.64}$$

then the compact result

$$N(n, s) = e^{-s} n^2 \Sigma\left(\frac{1}{n}\right) + e^{-ns}\,\Sigma\,(n) \tag{8.65}$$

is obtained. Importantly, although (8.61) and (8.62) are exact results, (8.65) is approximate. An assessment of the accuracy may be obtained by rewriting (8.62) as

$$N(n, s) = e^{-s}\left[n^2 \Sigma\left(\frac{1}{n}\right) + \Sigma(n)\right] + e^{-s}\left[e^{-(n-1)s} - 1\right]\Sigma\,(n) \tag{8.66}$$

At $n = 1$, the second term is simply $s\, e^{-s}$ because for $j > 0$,

$$\lim_{n\to 1}(n-1)^j \Sigma(n) = -\delta_{1,j} \tag{8.67}$$

where $\delta_{i,j}$ is the Kronecker delta function. As for the first term,

$$\lim_{n\to 1} n^2 \Sigma\left(\frac{1}{n}\right) + \Sigma(n) = \frac{7}{2}\zeta(4) + 2\zeta(2) - 6 = 1.0780 \tag{8.68}$$

which is seen to be 7.8% larger than 1. The accuracy of the first term, therefore, sets the accuracy of the approximation. Insofar as s is generally large, the accuracy is reasonable. Nevertheless, for numerical evaluations below for the case $n = 1$, (8.62) shall be used directly.

For thermal and field emission, introduce the two current densities J_F and J_T defined by

$$J_F = A_{RLD}(k_B\beta_F)^{-2}\Sigma\left(\frac{\beta_F}{\beta_T}\right)\exp\left[-\beta_F\,(E_o - \mu)\right] \tag{8.69}$$

$$J_T = A_{RLD}(k_B\beta_T)^{-2}\Sigma\left(\frac{\beta_T}{\beta_F}\right)\exp\left[-\beta_T\,(E_o - \mu)\right] \tag{8.70}$$

from which the General Thermal-Field current density $J_{GTF}(F, T)$ is given by

$$J_{GTF}(F,T) = \begin{cases} n^{-2}J_F + J_T & (n < 1) \\ J_F + n^2 J_T & (n > 1) \end{cases} \tag{8.71}$$

The $n \to 0$ limit gives J_{RLD} (8.5), and $n \to \infty$ gives J_{FN} (8.6), with the Σ-factors giving the temperature and field modifications to each in the intermediate regimes. Near $n = 1$, both J_T and J_F are required, as $\Sigma(n)$ contains a singularity at $n \to 1$ which must be canceled by $\Sigma(1/n)$. The application to photoemission is most easily obtained by augmenting the energy E of the transmission probability $D(E) \to D(E + E_\omega)$, causing $E_o \to E_o - E_\omega$, where $E_\omega = \hbar\omega$ for the absorbed photon. This causes the negative s, or $N(n, -s)$ term, to dominate and results in the Fowler–DuBridge Equation (8.8) to leading order.

8.3.3 Implementation

A proper implementation of the GTFP Equation requires a determination of the location of the maximum E_m of the current density integrand $j(E)$ which is proportional to

$$j(E) \propto \frac{\ln\left(1 + e^{\beta_T(\mu - E)}\right)}{1 + e^{\theta(E)}} \tag{8.72}$$

such that $dj/dE = 0$ at $E = E_m$. The energy slope factor $\beta_F(E_m)$ is then evaluated using the shape factor approach or (8.46), from which n and s are determined and (8.71) employed. Such an approach, although numerically intensive compared to the simple FN and RLD equations, functions well compared to direct numerical integration over the supply function and Kemble approximation using the Gamow factor.

A reasonably accurate simpler approach, called the "transition method" below, is possible based on the observation of how $n(F,T)$ varies with F and T as calculated numerically: the results of such an evaluation are shown in Fig. 8.6, for $n \equiv \beta_T/\beta_F(E_m)$. Observe that until $n \approx 1$ the behavior of n in each regime ("Thermal" corresponding to $n < 1$ and "Field" to $n > 1$) behaves as a power of F albeit that the power changes across the $n = 1$ boundary as revealed by either the shape factor analysis or the behavior of θ and β_F at $E \approx \mu$ and $E \approx \mu + \phi$: in the thermal regime (T and F_o such that $n_o = n(F_o, T) \ll 1$),

$$\ln\left(\frac{n}{n_o}\right) = \frac{3}{4}\ln\left(\frac{F}{F_o}\right) \tag{8.73}$$

so that the power is 3/4. In the field regime (T and F_o such that $n_o = n(F_o, T) \gg 1$),

$$\ln\left(\frac{n}{n_o}\right) \approx \ln\left(\frac{Ft(y_o)}{F_o t(y)}\right) \tag{8.74}$$

so that the power is close to unity insofar as $t(y_o)/t(y)$ has a weak field dependence.

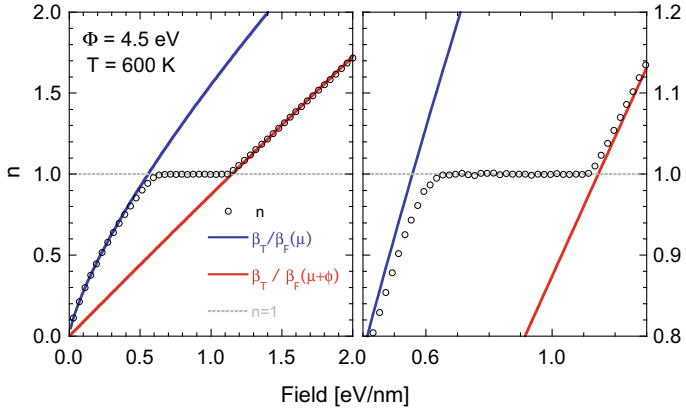


Fig. 8.4 Behavior of $n = \beta_T/\beta_F(E_m)$ determined numerically as a function of field F for representative parameters of $T = 600$ K and $\Phi = 4.5$ eV. (left) representative range; (right) focus on the $n = 1$ region. Also shown are the (red) thermal limit $n \sim \beta_T/\beta_F(\mu + \phi)$ and the (blue) field limit $n \sim \beta_T/\beta_F(\mu)$

This indicates that two transition temperatures T_{max} and T_{min} for a given field can be defined where

$$k_B T_{max} = \frac{1}{\beta_F(\mu + \phi)} \tag{8.75}$$

$$k_B T_{min} = \frac{1}{\beta_F(\mu)} \tag{8.76}$$

and where (8.49) and (8.50) allow computation. Conversely, the same equations can be used to specify an F_{max} and F_{min} in terms of T, which bracket the region between the red and blue lines of Fig. 8.4, but in such cases, the dependence of $\beta_F(E_m)$ on field means that numerical means (e.g., bisection or iteration) are required to locate F_{max} and F_{min}. In the region between F_{max} and F_{min}, $n \to 1$ and the peaky integrand $j(E)$ of (8.72) shifts its peak from near the barrier minimum to near the barrier maximum, as shown in Fig. 8.5: even though the peaks appear broad for $n = 1$, they are more narrow than simulations using a non-linear $\theta(E)$ (e.g., (8.46)) or experiments (Fig. 8.1 or Fig. (2) in [64]; Fig. (50) in [57]).

The simplest algorithm is, therefore, to set $\beta_F = \beta_F(\mu)$ if $T < T_{min}$ (the field regime) and $\beta_F = \beta_F(\mu + \phi)$ for $T > T_{max}$ (the thermal regime). For $T_{min} < T < T_{max}$, then $n = 1$ and E_m is determined from (8.46) to make that so, from which F is found. Observe that $J(F, T)$ is determined from $N(1, s)$, or more explicitly,

$$J_{GTF}(F, T)|_{n=1} \equiv A_{RLD} T^2 \left(s(E_m) + 1\right) e^{-s(E_m)} \tag{8.77}$$

Conversely, in the field regime, $\beta_F(E_m) \to \beta_F(\mu)$, whereas in the thermal regime, $\beta_F(E_m) \to \beta_F(\mu + \phi)$, and (8.71) is used. This approach of using the approximate

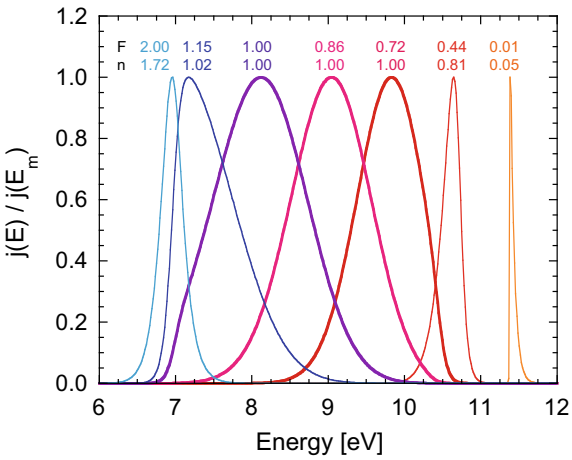


Fig. 8.5 Shift of $j(E)$ (left to right) from field emission conditions where $n > 1$ to thermal emission conditions where $n < 1$. The thicker lines are associated with $n = 1$. The numbers above each curve are F (top, in eV/nm) and n (bottom). Other parameters are $T = 600$ K, $\Phi = 4.5$ eV, and $\mu = 7$ eV. Although E here is the normal energy component, compare Fig. 8.1

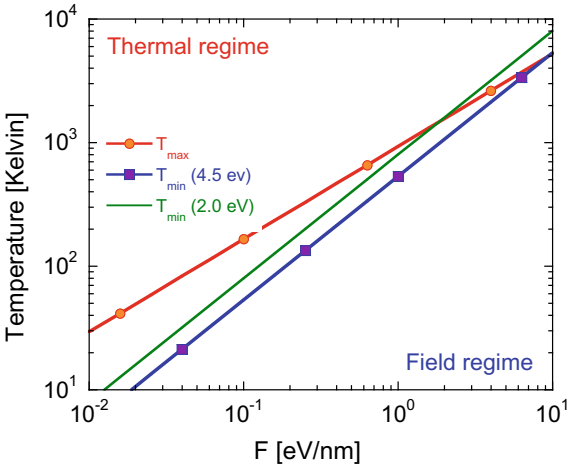


Fig. 8.6 Behavior of T_{max} and T_{min}: because the later depends on the work function Φ, the location where the lines cross shifts with Φ

forms in each regime performs reasonably well compared to a proper numerical evaluation and can be virtually used to the point where T_{max} and T_{min} cross, as in Fig. 8.6, the location of which depends on the work function. The performance of J_{GTF} compared to J_{RLD} and J_{FN} is shown in Fig. 8.7. This formulation, referred to as the *o*GTF method [74, 75], is significant for the rapidity of its implementation in describing thermal-field emission, as it only demands $\theta(E)$ and $\beta_F(E)$ evaluated at $E = \mu$ and $E = \mu + \phi$, and so is useful for simulating emitters rapidly that simultaneously exhibit thermal, field, and thermal-field contributions on the same multidimensional structure [77].

The cost of computational rapidity, however, is some loss of accuracy due to the departure of n from the limiting cases as $n \to 1$ in Fig. 8.4, the consequence of which is to introduce "dimples" or kinks at the transitions from thermal or field

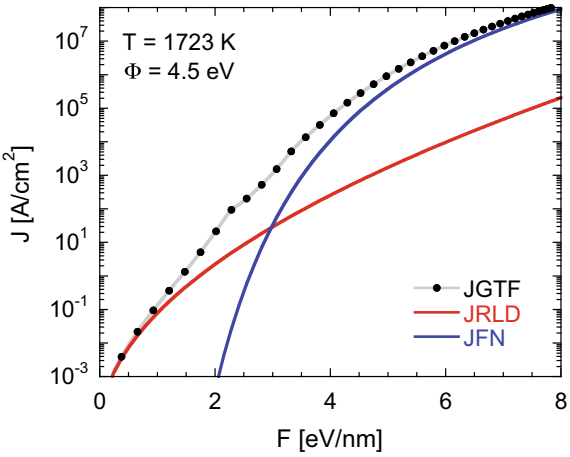


Fig. 8.7 The General Thermal-Field equation compared to the RLD and FN equations for representative parameters. Observe the small dimples associated with $T_{\min}$ and $T_{\max}$ associated with the transition model

to the thermal-field regime [78], as seen in Fig. 8.7. The kinks can be mitigated by correctly determining the location of E_m exactly rather than by approximating it by μ, $\mu + \phi$, or where $\beta_T = \beta_F(E_m)$ (the field, thermal, and thermal-field approximations, respectively) [74]. The approximation of $\theta(E)$ as linear in E, however, causes the GTF method to overestimate $J(F, T)$ in the thermal-field regime by a factor generally between 1 and 3. Recovering accuracy incurs a manageable computational cost to define a correction coefficient constructed of Lorentzian terms, the exposition of which is deferred to the literature [75] due to its greater requirements.

8.3.4 Nottingham Heating

The relationship between electron emission and heating in microstructures due to resistive (or Joule) heating and Nottingham heating [79, 80] is of interest due to their correlation to degradation and failure mechanisms in fabricated microstructures and dark current emission sites in accelerators (to name two) [26, 81–83]. In Nottingham cooling, electrons emitted above the Fermi level are replaced by electrons at the Fermi level that absorb energy in the process of thermal activation, which leads to lattice cooling. Conversely, electrons emitted below the Fermi level are replaced by electrons that give off energy to the lattice, thereby leading to heating. The $N(n, s)$ function allows for an evaluation.

In what follows, it is necessary to return to a 3D formulation, thereby requiring E to be understood as the *total* electron energy (not the normal energy E_z considered previously in Sects. 8.2.3 and 8.2.4). The current density element $dJ(E)$ when E is the total (not normal) energy [84, 85] is

$$dJ(E) = \frac{A_{RLD}}{k_B^2} f_{FD}(E) \left[\int_0^E D(E_z) dE_z \right] dE \tag{8.78}$$

Relative to the Fermi level, the average total energy ΔE of the emitted electrons is given by

$$\Delta E\,(F,T) = \langle E-\mu\rangle = \frac{\int (E-\mu)\,dJ\,(E)}{\int dJ\,(E)} \tag{8.79}$$

The Nottingham term is then found to be [85]

$$\int_0^\infty (E-\mu)\,\frac{\ln\left\{1+e^{\beta_F(E-E_o)}\right\}}{1+e^{\beta_T(E-\mu)}}dE = \frac{1}{\beta_F}\int_0^\infty (s-z)\,\frac{\ln\left\{1+e^{-z}\right\}}{1+\mathrm{e}^{n(z-s)}}dz \tag{8.80}$$

Conveniently, it can be shown that this equation gives rise to

$$\Delta E = \frac{1}{n\beta_F}\left(\frac{N'\,(n,s)}{N\,(n,s)}\right) \tag{8.81}$$

where

$$N'\,(n,s) \equiv n^3\Sigma'\left(\frac{1}{n}\right)e^{-n} + \left[(ns+1)\,\Sigma\,(n) - \Sigma'\,(n)\right]e^{-ns} \tag{8.82}$$

and primes denote derivatives with respect to argument. It is tempting and sometimes done in practice to neglect the thermal-like terms in favor of the field-like terms, but that results in an error: as the emitters heat, they will pass through the $n = 1$ regime, so that the neglect of either the thermal-like or field-like term removes a contribution in $\Sigma(n)$ and $\Sigma(1/n)$ that is singular as $n \to 1$, whereas the ratio of $N'(n,s)$ to $N(n,s)$ does not suffer such a singularity. The transition from $n > 1$ to $n < 1$ conditions can therefore be tracked smoothly, as done in [85] and shown in Fig. 8.8.

The factor ΔE refers to the average energy of the emitted electrons with respect to the Fermi level μ. If ΔE is negative, then emitter heating occurs, and the effect is called "Nottingham heating"; if it is negative, then emitter cooling occurs, and the effect is usually called "Nottingham cooling". Except near the transition where ΔE is small, the Nottingham heating processes can dominate the Joule heating due to electron–phonon collisions during current flow. Nottingham heating is particularly relevant to the generation of asperities that contribute to electrical breakdown [26, 85].

In order to judge the relative magnitude of Nottingham heating to resistive (or Joule) heating, however, requires an estimate of the total current through an asperity or a tip [26, 82, 85], but to estimate current from asperities, a model that mimics how such asperities might grow, but which also can model conical or Spindt-like [86] field emitters (as may be used for TWTs [3, 87]), and suggests a manner in which to model wire-like emitters for high-power microwave generation [88, 89] as well. Such models shall be considered next, but the general finding is that Nottingham heating effects generally dominate resistive heating [85], as shown in Fig. 8.9.

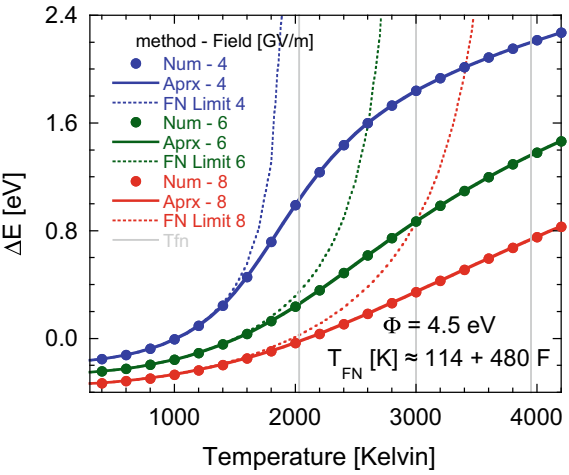


Fig. 8.8 The Nottingham term as a function of temperature for copper-like parameters. Dots refer to numerically calculated ΔE. Dashed lines (labeled "FN") refer to the neglect of the thermal terms, showing the impact of the uncompensated signlarity in $\Sigma(n)$ as $n \to 1$. T_{FN} refers to where $\beta_F = \beta_T$. After Fig. (14) of [85]

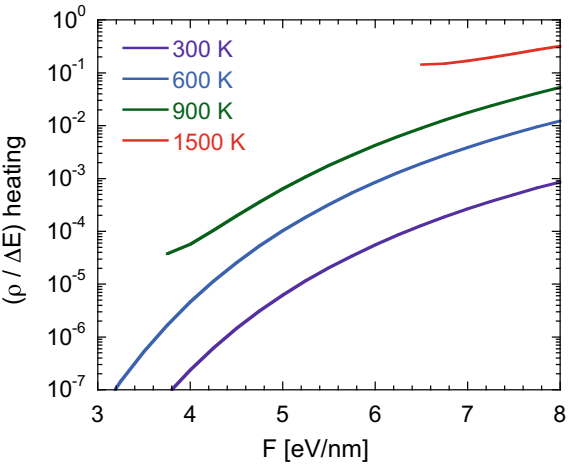


Fig. 8.9 Ratio of resistive heating to Nottingham heating for different temperatures as a function of apex field, after Fig. (15) of [85]

8.4 Current from a Protrusion

Field emitters are often intentionally crafted structures such as a Spindt-type field emitter, a tungsten needle, a carbon fiber wire, or a carbon nanotube (to name but a few examples), but may also take the form of naturally occurring protrusions. To estimate the current from such an emitter, the most commonly used approach assumes that the current density over a differential element of its surface can be evaluated from a current density relation (the Fowler–Nordheim equation of (8.6), or more generally, the General Thermal Field relation of (8.71)) that is expressed in terms of the local classical electrostatic force F at that differential element. One then sums over the surface defined by all such elements. This entails knowing the shape of the surface

and how the field varies over it; it also involves assuming that it is adequate to apply the (planar-surface) linear field approximation to each element, although for field emission such an approximation has been rightly challenged [63, 90, 91]. When used in Boundary Element simulations [92–94], rather complex geometries (gated or otherwise) can be modeled. In what follows, however, semi-analytical models are considered so as to demonstrate general behaviors that should be expected in more complex configurations, and to provide limiting and well-understood cases.

8.4.1 Hemisphere

The hemispherical bump represents perhaps the simplest model from which to understand field enhancement [95–97], emission area [48, 61, 98], and causes of emittance [99–101]. It is convenient to discuss this in spherical coordinates with origin at the hemisphere center, with r representing distance from the origin and θ angle from the symmetry axis. In these coordinates, the barrier potential energy for FE from a hemisherical protrusion (measured relative to the Sommerfeld well base, as before) is

$$U(r,\theta) = \mu + \Phi_o - F_o r\cos\theta\left\{1 - \left(\frac{a}{r}\right)^3\right\} - \frac{2aQ}{r^2 - a^2} \tag{8.83}$$

where Φ_o is the work function, a is the hemisphere radius, and F_o is the background classical electrostatic force, for example, $F_o = |\Delta U_D|/D$ if the hemisphere sits on the cathode side of a parallel plate capacitor, where $D \gg a$ is the anode–cathode (or AK) gap, and ΔU_D is the difference in the potential energy of a classical point electron between positions just outside the anode and just outside the cathode, respectively. Measuring radial distance with respect to the surface, by putting $r \to a + x$, results in

$$U(x + a,\theta) \approx \mu + \Phi_o - 3F_o x\cos\theta\frac{a}{a+x} - \frac{Q}{x} + \frac{Q}{2a+x} \tag{8.84}$$

It is now useful to put this equation into the "equivalent linear form", by taking the limit $x/a \to 0$. The resulting "impact of multidimensionality" will be to (*i*) introduce a radius a for an approximately spherical emitter apex, (*ii*) replace the constant surface force of the planar-geometry model by a position-dependent local classical electrostatic force $F(\theta) = 3F_o\cos\theta$, and (*iii*) augment the work function to an effective value $\Phi = \Phi_o + Q/(2a)$. In short, in the GTF equation, the work function can be treated as given by the *effective* work function, and "$3F\cos\theta$" can be directly inserted in the place of F. In what follows, Φ will be understood to be the effective work function $\Phi_o + Q/(2a)$ and not further examined.

In order to progress the argument further, it is necessary to separate out the electrostatic component of the barrier PE. This is denoted by U^{el} and is taken as measured relative to a point on the surface of a smooth classical emitter model, at which the electrostatic potential and hence U^{el} are defined to be zero. Thus, for the hemispher-

ical protrusion, U^{el} is given by

$$U^{el} = -3F_o x \cos\theta \frac{a}{a+x} \tag{8.85}$$

Defining the field enhancement factor β as the ratio between the *surface* ($x = 0$) quantity F and the background quantity F_o indicates that $\beta(\theta) = 3\cos\theta$. Often, it is the field enhancement factor *at the apex* of the emitter that is referred to, so that the field enhancement of a hemisphere is said to be $\beta = 3$, or in cylindrical coordinates (z, ρ),

$$\beta = -\frac{1}{F_o}\,\partial_z U^{el}(z,\rho)\big|_{z=a,\rho=0} \tag{8.86}$$

The hemisphere surface is taken to be an electrostatic equipotential of value zero, which implies that $U^{el}(a, \theta) = 0$. A Taylor expansion for small ρ and $z \approx a$ shows that

$$a = -\left.\frac{\partial_z U^{el}(z,\rho)}{\partial_\rho^2 U^{el}(z,\rho)}\right|_{z=a,\rho=0} \tag{8.87}$$

This will be the definition of the apex radius used subsequently when the surface is no longer hemispherical but something more complex.

The "notional" emission area [102], Ω, which is smaller than the actual emission area, is a representative measure of the fraction of the emitter apex that contributes to the current. This parameter Ω relates the total emission current (I_{boss}) from a protrusion to the apex value $J(F_{tip})$ of emission current density, via $I_{boss} = \Omega J(F_{tip})$, with $F_{tip} = \beta_{tip} F_o$.

The emission area is field dependent: therefore, introduce $\Omega \equiv 2\pi a^2 g(F)$, where $2\pi a^2$ is the area of a hemisphere, and $g(F)$ functions as measure of the fraction of the surface contributing to emission. Therefore,

$$g(F) = \frac{1}{J(F)} \int_0^{\pi/2} J(F\cos\theta)\sin\theta\, d\theta \tag{8.88}$$

where $F = F_{tip} = 3F_o$ for a hemisphere. A warm—or even hot—hemisphere requires the usage of J_{GTF}, but it is useful to consider the limits of thermal and field emission separately, as mixed conditions will, in some sense, be between them. For that purpose, the notional emission angle factor $\eta \equiv \cos\theta_o = 1 - g(F)$ is a better measure, and is defined by the solution θ_o to the equation

$$\int_0^{\theta_o} \sin\theta\, d\theta = \int_0^{\pi/2} P(\theta)\sin\theta\, d\theta \tag{8.89}$$

where $P(\theta) \equiv J[F\cos(\theta)]/J[F]$.

In the case of thermal emission (t-subscript), where the field dependence affects the work function by $\phi = \Phi - \sqrt{4QF}$, then using (8.5) and letting b represent the

characteristic energy ratio, then

$$P_T(\theta) = \exp\left[b_t\left(\sqrt{\cos\theta} - 1\right)\right] \tag{8.90}$$

$$b_t(F,T) = \sqrt{4QF}/k_BT \tag{8.91}$$

resulting in

$$\eta_t = \frac{2}{b_t} - \frac{2}{b_t^2}\left(1 - e^{-b_t}\right) \tag{8.92}$$

In the case of field emission (f-subscript), the F-dependence of (8.7) gives

$$P_f(\theta) = (\cos\theta)^{2-\nu} \exp\left(b_f - \frac{b_f}{\cos\theta}\right) \tag{8.93}$$

$$b_f(F) = B_o\Phi^{3/2}/F \tag{8.94}$$

resulting in

$$\eta_f = \frac{b + 3 - \nu}{b + 4 - \nu} \tag{8.95}$$

The amount of current coming from the hemisphere as a function of θ is measured by

$$\Delta(\theta) \equiv \frac{I(\theta)}{I_{boss}} = \frac{1}{1-\eta}\int_0^\theta \sin\theta' P\left(\theta'\right)d\theta' \tag{8.96}$$

for which $\Delta(\theta) \to 1$ as $\theta \to \pi/2$. The behavior is shown in Fig. 8.10 for the limits of field and thermal emission. Clearly, and as expected, field emission is much more tightly concentrated at the apex.

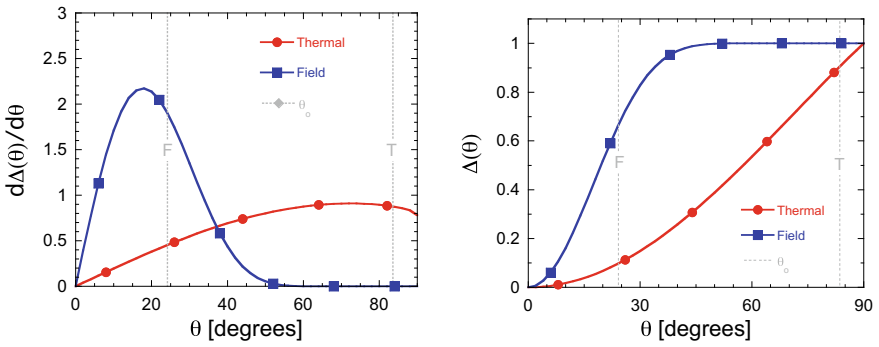


Fig. 8.10 The integrand $d\Delta/d\theta$ and the integrated current ratio $\Delta(\theta)$ evaluated for (thermal) $T = 1200$ K and $F = 0.001$ eV/nm and (field) $F = 8$ eV/nm. Dashed gray vertical lines correspond to θ_o for each. After [48]

For completeness, photoemission (which differs significantly from the thermal and field emission processes) has a field dependence that (as with thermal emission) is contained in $\phi = \Phi - \sqrt{4QF}$. Therefore, [99]

$$P_\omega(\theta) = b_\omega^{-2}\left(b_\omega - 1 + \sqrt{\cos\theta}\right)^2 \tag{8.97}$$

$$b_\omega = \frac{\hbar\omega - \phi}{\sqrt{4QF}} \tag{8.98}$$

$$\eta_\omega = \frac{4b_\omega - 1}{6b_\omega^2} \tag{8.99}$$

The integrated current fraction is

$$\Delta_\omega(\theta) = 1 - \cos\theta \frac{6(b_\omega - 1)^2 + 8(b_\omega - 1)\sqrt{\cos\theta} + 3\cos\theta}{6(b_\omega - 1)^2 + 8(b_\omega - 1) + 3} \tag{8.100}$$

The evaluation of emittance requires knowledge of the *launch velocity* v_o of the emitted electrons. It will be taken as equal to the average velocity $\langle v\rangle$ obtained from $J = q\rho\langle v\rangle$ where

$$v_o = \langle v\rangle \equiv \frac{\int_0^\infty (\hbar k/m) D(k) f(k) dk}{\int_0^\infty D(k) f(k) dk} \tag{8.101}$$

in which $D(k)$ and $f(k)$ are the transmission probability and the supply function in a one-dimensional treatment in terms of the electron momentum $\hbar k$ normal to the emitter surface before emission. An analogous equation holds for $\langle v^2\rangle$ needed to evaluate the standard deviation, and is obtained by the replacement $\hbar k/m \to (\hbar k/m)^2$ in the numerator of (8.101), such that the root mean squared velocity is $v_{rms} = \sqrt{\langle v^2\rangle} = (\hbar/m)\sqrt{\langle k^2\rangle}$. The proportionality between v_o and $\langle v\rangle$ differs for a curved surface whereas for a flat surface, they are equal.

For thermal and field emission (with $\mu = \hbar^2 k_F^2/2m$, where k_F is the Fermi wavenumber),

$$v_o(\text{thermal}) = (2k_B T/\pi m)^{1/2} \tag{8.102}$$

$$v_o(\text{field}) = \frac{\hbar k_F}{m}\left(1 - \frac{\Phi}{3t_o b_t \mu}\right)^2 \tag{8.103}$$

An extended analysis of the contribution to emittance from the transverse velocity components due to surface geometry [99] gives for the approximation to emittance $\varepsilon_{n,rms}$

$$\varepsilon_{n,rms} \approx C_o(1-\eta)^3\sqrt{3\eta^2 + 18\eta + 19} \tag{8.104}$$

$$C_o \equiv \frac{\sqrt{5}}{240}\left(\frac{\tau v_o^2}{c}\right) \tag{8.105}$$

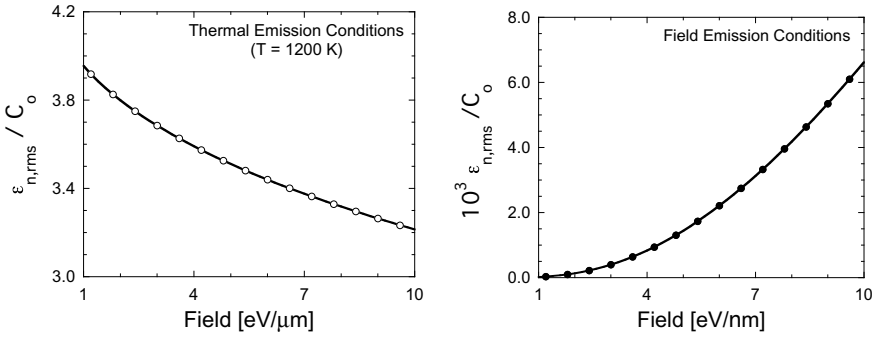


Fig. 8.11 The behavior of $\varepsilon_{n,rms}/C_o$ for thermal (left) and field (right) emission conditions; observe the differences in scale for each of the axes. After [48]

where $\tau = mv_o/F_o$, for electrons that have had ample time to leave the influence of the hemispherical perturbations to the field. Observe that (8.104) is for a *single hemispherical emitter*.

Let $\chi = F/[F]_x$, where $[F]_x$ is a characteristic local classical electrostatic force, taken as $[F]_T = 0.1$ eV/μm, $[F]_\omega = 10$ eV/μm, and $[F]_F = 1$ eV/nm for thermal, photo, and field emission, respectively. Useful relations then are

$$C_o[T] = 6.62 \times 10^{-5} \frac{(k_B T)^{3/2}}{\chi} \tag{8.106}$$

$$C_o[\omega] = 7.095 \times 10^{-5} \frac{(\hbar\omega - \Phi)^{3/2}}{\chi} \tag{8.107}$$

$$C_o[F] \approx 3.686 \times 10^{-8} \frac{\mu^{3/2}}{\chi} \tag{8.108}$$

where C_o is measured in [mm mrad] and where, for example, $\mu = 7$ eV for copper. The behavior of $\varepsilon_{n,rms}/C_o$ is shown in Fig. 8.11 for the cases of thermal and field emission: although the characteristic thermal velocity v_o is smaller than the characteristic field velocity, the differences show, as expected, that single tip field emission dominated sources are intensely bright, where brightness [6, 8, 9], recall, is defined as the ratio of the current with the emittance squared, or $B = I_{boss}/\varepsilon_{n,rms}^2$.

When field emitters are assembled into an array, then the dimensions of the array enter prominently into the calculation of emittance. Depending on operating conditions, the gulf between thermal or photoemission and field emission emittance over extended areas closes considerably in a manner that depends greatly on the tip-to-tip separation of the emitters and the current per tip (as well as how fast the electrons are accelerated and the time that space charge forces can affect the beam [103], but these considerations are outside the present scope). Moreover, when an array is present,

the launch velocities have to be augmented by an impulse correction. Such modifications are apart from considerations of electron emission itself and are deferred to the literature [15, 99, 100].

Before leaving the hemispherical model, the impact of a general thermal-field emission equation in the context of a non-planar, and specifically hemispherical, structure is briefly assessed. For finite temperatures and because the field along a hemisphere $F(\theta) = 3F_o \cos\theta$ vanishes as $\theta \to \pi/2$, then there exists an angle θ_c past which the emission transitions from field-dominated ($n = \beta_T/\beta_F > 1$) to thermal-dominated ($n < 1$), as governed by (8.75) and (8.76), respectively. For simplicity, assume that the entire sphere is at a constant temperature (even though thermal modeling suggests temperature gradients can exist near the apex [26, 81, 82, 85]). To be considered, then, is how elevated temperatures affect both the total emitted current and the contributions giving rise to emittance. Three representative simulations of $J_{GTF}(F, T)$ were performed to show how $F(\theta) = F_{tip}\cos\theta$ declines down the side of the hemisphere with the consequences on $J_{GTF}(F, T)$, and where the transition from field to thermal emission begins (that is, where $n \approx 1$ from the field side as per (8.49)), characterized by $F_{trans} \approx 23.068 k_B T$, where the coefficient is $(2t_o/\hbar)\sqrt{2m\Phi}$ with $t_o = 1.0613$ as a representative value of $t(y)$. The current density is scaled by $J_{FN}(F_{tip})$ to also show the impact of rising temperature on the total current density. For three separate values of classical electrostatic force $F/(\text{eV/nm}) = 3, 4,$ and 8), the current density was evaluated as a function of angle for three temperatures apiece (T [K] = 473, 973, and 1473). The results of the calculation are shown in Fig. 8.12.

Not unexpectedly, elevated temperatures have a substantial impact on the projected current density, but are of greater consequence at lower fields. At sufficiently low fields, high-temperature emitters can begin to contribute emitted current near $n \to 1$, where the normal energy distribution is known to be broad, as in Fig. 8.5. The impact this has on beam quality will be two-fold. First, it will introduce an energy spread during the emission process. Second, as a physical emitter will likely have a temperature gradient as a function of polar angle θ, the emission profiles for intermediate temperatures will lie between the high and low temperatures, causing the launch velocities v_o to change with temperature and field and thereby affecting transverse emittance $\varepsilon_{n,rms}$.

As convenient as the hemispherical model is, it assumes that the field fall-off as a function of polar angle is faster than for comparable ellipsoidal or hyperbolic emitters [48, 61, 104]. Along a prolate spheroidal emitter, a slower field fall-off entails that there are greater thermal complications to the determination of launch velocities, but such models entail further approximations that constrain application to conical or gated emitters. A more flexible model that has been used to examine issues of emittance and launch velocity in the development of a trajectory model of emission [99] is the Point Charge Model (PCM) (which can be generalized to line charge models [89, 105]).

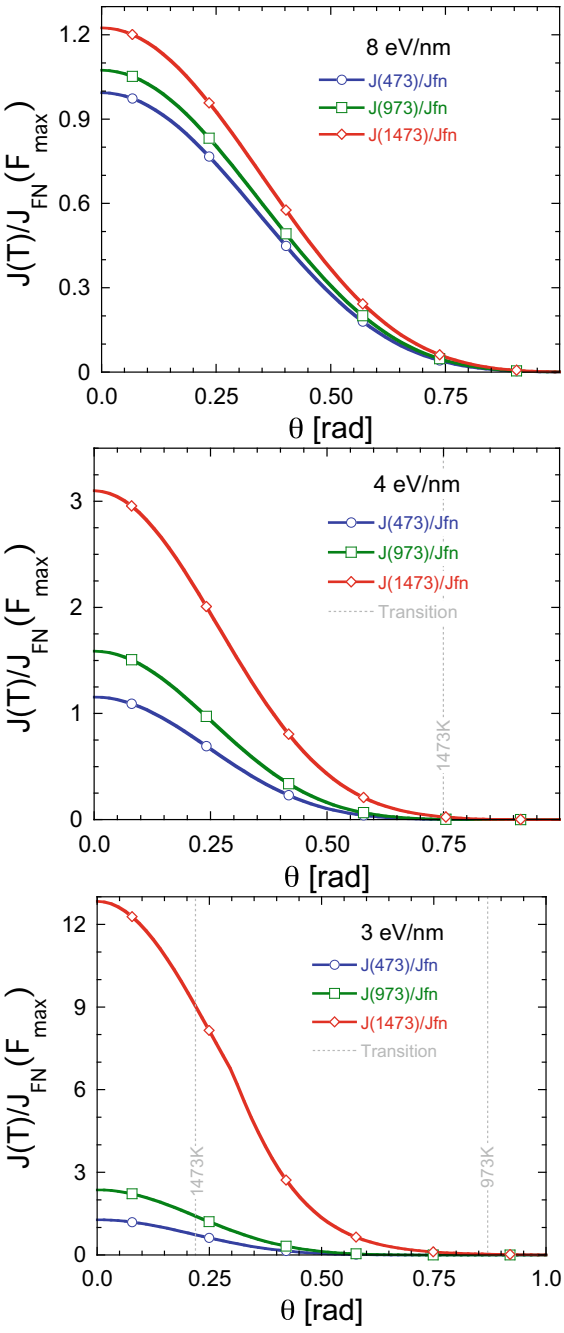


Fig. 8.12 Current density, normalized to the Fowler–Nordheim value of $J_{FN}(F_{tip})$ (8.7) at the apex as a function of angle down the side of a hemispherical emitter, where $F(\theta) = F_{tip}\cos\theta$ for the apex values $F = 8$ eV/nm (top), 4 eV/nm (middle), and 3 eV/nm (bottom). In the middle and bottom figures, the transition points where $1/k_BT\beta_F(\mu) \to 1$, as per (8.49), are indicated by dashed lines with a gray number overlay of the value of T. The hemisphere is assumed to be at a uniform temperature throughout

8.4.2 Point Charge Model

A consideration of dark current sources (typically field emission) in rf accelerators [27, 83, 106] led to the development of a "monopole" version of the Point Charge Model, in order to treat hot field emission from a field emitter electroformed at high field and high temperature [85]. For treating emitters that are more closely conical, such as Spindt-type field emitters or sharp protrusions, the "dipole" point charge model was introduced and refined [15, 107, 108]. A brief account of these models follows.

In the monopole point charge model (so-called because only one sign of charge is considered), a number n of point charges are placed in a line above a planar conducting surface and a background field is presumed. Any resulting equipotential surface can serve to define the shape of a perfectly conducting protrusion, and potentials and fields can be calculated everywhere outside this surface. Monopole-based protrusions tend to have broad bases, as cone-jet shapes do, but microfabricated field emitters tend to have a more closely conical shape. Therefore, the "dipole model" puts equal and opposite point charges below the plane, making the extent of the emitter finite.

To enforce a realistic shape, the placement of the above-plane point charges at the points z_j ($j \in \{1, n\}$, and z_j increasing as j increases) is restricted to values for which

$$\frac{z_{j+1} - z_j}{z_j - z_{j-1}} \equiv r \tag{8.109}$$

where r is a dimensionless number typically less than unity. The below-plane charges (of opposite sign) are placed at the mirror image positions. These choices force the equipotential shape to taper off in a controlled manner toward the point of the emitter. Using cylindrical coordinates, and in the presence (as earlier) of an uniform background classical electrostatic force F_o, the electrostatic component $U^{el}(z, \rho)$ of electron potential energy is, in the dipole model, then everywhere given by

$$U^{el}(z, \rho) = F_o a_0 \, u_n(z, \rho) \tag{8.110}$$

$$u_n(z, \rho) = -\frac{z}{a_0} - a_0 \sum_{j=1}^{n} \lambda_j \left(\frac{1}{R_j^+} - \frac{1}{R_j^-} \right) \tag{8.111}$$

$$R_j^{\pm} \equiv \sqrt{\rho^2 + (z \pm z_j)^2} \tag{8.112}$$

where a_0 is a characteristic length scale on the order of the base radius of the emitter, and the λ_j are dimensionless numbers discussed below. The monopole model would be similar, except there would be no R_j^+ term, and a $j = 0$ term (for which $z_0 = 0$ and $\lambda_0 > 0$) would also be included. The positions of the point charges are then specified by

$$\frac{z_n}{a_0} = S_n(r) \equiv \sum_{j=0}^{n-1} r^j = \frac{1 - r^n}{1 - r} \tag{8.113}$$

The parameter λ_j specifies the amount of charge on the jth point charge. The λ_j are uniquely determined by demanding

$$u_n(z_{n+1}, 0) = 0 \tag{8.114}$$

That is, we demand that the electrostatic potential, and hence the electrostatic component of electron PE, be zero at a point on the z-axis that is "outside" the topmost charge [at $(0, z_n)$] by the distance $(z_{n+1} - z_n)$, as determined from (8.109). This sets up a series of equations (see [108]) that can be solved using matrix inversion techniques.

For calculation, it is numerically preferable to represent $\lambda_n \equiv r^n P_n(r)$, where P_n is defined by this relation. Then, explicitly, for the monopole case,

$$\sum_{j=1}^{n} \frac{\Theta_{n,j}}{S_{n+1-j}} P_j = S_{n+1}(r) - \frac{1}{S_{n+1}(r)} \tag{8.115}$$

whereas for the dipole case,

$$\sum_{j=1}^{n} \frac{2S_j \Theta_{n,j}}{S_{n+1-j}\left[S_j + S_{n+1}\right]} P_j = S_{n+1}(r) \tag{8.116}$$

where $\Theta_{n,j} = 1$ if $j \leq n$ and 0 otherwise.

In terms of the S_n functions, the field enhancement factors $\beta_n(r) = -a_0 \partial_z u_n(z_{n+1}, 0)$ for monopole and dipole structures involving n above-plane charges are

$$\beta_n^{mono} = 1 + (S_n)^{-2} + \sum_{j=1}^{n} \frac{P_j}{r^j (S_{n+1-j})^2} \tag{8.117}$$

$$\beta_n^{dipl} = 1 + \sum_{j=1}^{n} \frac{4P_j S_j S_{n+1}}{r^j \left[S_{n+1-j}\left(S_{n+1} + S_j\right)\right]^2} \tag{8.118}$$

Likewise, the apex radii a_n for structures involving n above-plane charges are

$$\frac{a_n^{mono}}{a_0} = S_{n+1} \frac{1 + S_{n+1}^2 + \sum_{j=1}^{n} \frac{P_j}{r^j} \left(\frac{S_{n+1}}{S_{n+1-j}}\right)^2}{S_{n+1}^3 + \sum_{j=1}^{n} \frac{P_j}{r^{2j}} \left(\frac{S_{n+1}}{S_{n+1-j}}\right)^3} \tag{8.119}$$

$$\frac{a_n^{dipl}}{a_0} = 2\frac{\sum_{j=1}^{n}\frac{P_j}{r^j}\frac{S_j S_{n+1}}{\left[S_{n+1-j}\left(S_j+S_{n+1}\right)\right]^2}}{\sum_{j=1}^{n}\frac{P_j}{r^{2j}}\frac{S_j\left(S_j^2+3S_{n+1}^2\right)}{\left[S_{n+1-j}\left(S_j+S_{n+1}\right)\right]^2}} \tag{8.120}$$

as shown previously [107]. At this point, both the monopole and dipole forms of the apex radius and field enhancement factor are given in terms of $S_j(r)$ and $P_j(r)$, which may be calculated.

In the interests of space, the discussion is now constricted to the dipole case alone, since it produces a more conical shape characteristic of field emitters in arrays [86, 109, 110]: the monopole case would proceed analogously. The behavior of the field enhancement for the dipole case using (8.118) is shown in Fig. 8.13. The exponential nature of $\beta_n(r)$ with increasing n indicates that Schottky's conjecture will hold [111], wherein field enhancement factors are approximately multiplicative. Thus, $\beta_{n+1} \approx \Delta\beta\ \beta_n$, where $\Delta\beta$ is the field enhancement factor of a bump or protrusion, and β_n the field enhancement factor of the structure on which it rests. For a system of n above-plane charges, reasonable approximations to $a_n(r)$ and $\beta_n(r)$ are [48]

$$a_n(r) \approx a_0 r^n\left(2-2^{-n}\right) \tag{8.121}$$

$$\beta_n(r) \approx 2+2^n\left(\frac{1}{2^{n-1}}\frac{(n+1)\ln(4)}{\ln[4(n+1)]}\right)^r \tag{8.122}$$

where a_0 is the base radius of the emitter.

With the Point Charge Model (PCM), the field enhancement factor can be found analytically as a function of position along the surface of the emitter. Away from the apex, the enhancement factor will not decline as rapidly as for the hemispherical model, and its weaker decline will serve to enlarge the $g(F)$ factor of (8.88) governing the notional emission area. Insofar as the hyperbolic and ellipsoidal geometries produce shapes that are approximately related to types of PCM structures that can be generated, it is instructive to observe that the $g(F)$ factors for the hemispherical (b), ellipsoidal (e), and hyperbolic (h) cases are [48]

$$g_b(F) \approx \frac{1}{b-\nu+4} \tag{8.123}$$

$$g_e(F) \approx \frac{1}{b-\nu+1} \tag{8.124}$$

$$g_h(F) \approx \frac{\cos^2\upsilon_o}{b-\nu+\sin^2\upsilon_o} \tag{8.125}$$

where $b(F) = B_o\Phi^{3/2}/F$ and $\nu = 2B_oQ/3\sqrt{\Phi}$ as before for field emission conditions, and where υ_o is the half-cone angle of a hyperbolic emitter in (8.125). Although finite temperature enlarges $b(F)$, the trend (that emitters that are less "sharp" have larger $b(F)$) is confirmed by the field emission example and expected for thermal and

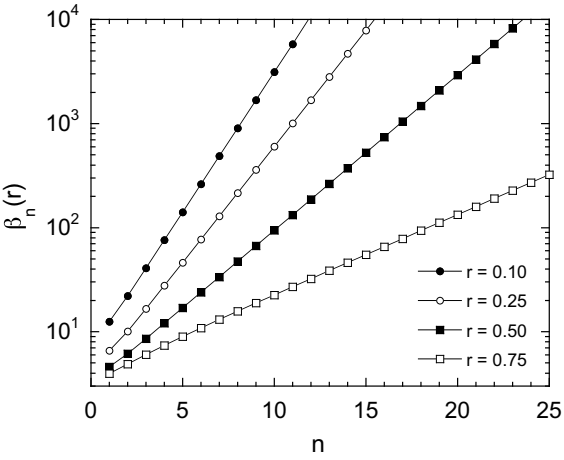


Fig. 8.13 Field enhancement for the dipole model as a function of the number of point charges n for various r (after Fig. 4 in [111])

photoemission conditions based on general arguments [99]. As a result, the methods are argued to have application to studies where (for example) the simultaneous presence of both heat and field emission effects [112] may significantly complicate the characterization of the electron beams that result.

8.5 Summary

Insofar as emittance is a pressing concern for applications requiring high brightness electron beams, the imperative to develop analytically tractable models of multidimensional emitters for use in beam optics codes for high-frequency devices [2, 5, 113] and next generation light sources [114, 115] is progressed by having a single general thermal-field-photoemission equation and a means to evaluate both total current from a multidimensional protrusion and the launch velocities of electrons that are emitted from it. An indication of how this can be accomplished has been considered for the special case of hemispherical emitters, due to the analytical ease with which such a geometry can be tackled, but an indication of how to extend the methods to more general structures has been suggested by the point charge model.

Regrettably, these models are bedeviled by the same complications that hamper the analysis of experimental data [45], (e.g., surfaces that are not smooth and are composed of more than one crystal face, each with a characteristic work function Φ; the relationship of theoretical parameters to the behavior of experimental data; band bending in semiconductors; oxide layers and contaminants; and so on). Nevertheless, the development of an analytical GTFP equation coupled with expectations from the simple models creates a potential to embed such models in codes that can then explore the elevated temperature of transition regimes [19] in designing sources and injectors. Similar concerns are faced by, for example, secondary emission from semiconductors

such as diamond [116, 117]. Such models have application, therefore, for the design of electron sources and accelerating structures related to devices demanding high brightness electron beams, particularly when the ever-present impact of space charge must be handled by beam optics codes.

References

1. B.N. Taylor, The International System of Units (SI). Special Publication (NIST SP) **330** (2008)
2. J.H. Booske, R.J. Dobbs, C.D. Joye, C.L. Kory, G.R. Neil, G.-S. Park, J. Park, R.J. Temkin, Vacuum electronic high power Terahertz sources. IEEE Trans. Terahertz Sci. Technol. **1**(1), 54–75 (2011)
3. D.R. Whaley, R. Duggal, C.M. Armstrong, C.L. Bellew, C.E. Holland, C.A. Spindt, 100 W operation of a cold cathode TWT. IEEE Trans. Electron Dev. **56**(5), 896–905 (2009)
4. J.W. Lewellen, High-brightness electron guns for linac-based light sources, Proc. of SPIE Int. Soc. Opt. Eng. **5534**, 22–36 (2004)
5. D.R. Whaley, Practical design of emittance dominated linear beams for RF amplifiers. IEEE Trans. Electron Dev. **61**(6), 1726–1734 (2014)
6. M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley, New York, 1994)
7. J.H. Booske, Plasma physics and related challenges of millimeter-wave-to-terahertz and high power microwave generation, Phys. Plasmas **15**(5), 055502 / 1–16 (2008)
8. C.A. Brau, High-brightness Electron Beams—small free-electron lasers. Nucl. Instr. Methods Phys. Res. Sect. A **407**(1–3), 1–7 (1998)
9. C.A. Brau, What brightness means, in *Physics and Applications of High Brightness Electron Beams: Proceedings of the ICFA Workshop Chia Laguna, Sardinia, Italy 1-6 July 2002*, edited by J. Rosenzweig, G. Travish, and L. Serafini (World Scientific Publishing Company, 2004), pp. 20–27
10. C.D. Joye, J.P. Calame, K.T. Nguyen, M. Garven, Microfabrication of fine electron beam tunnels using UV-LIGA and embedded polymer monofilaments for vacuum electron devices, J. Micromech. Microeng. **22**(1), 015010 / 1–10 (2012)
11. R. Dobbs, A. Roitman, P. Horoyski, M. Hyttinen, D. Sweeney, B. Steer, N. Khanh, E. Wright, D. Chernin, A. Burke, J. Calame, B. Levush, N.S. Barker, J. Booske, M. Blank, Design and fabrication of terahertz extended interaction klystrons, in 35th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THZ 2010) 978 / 1–3 (2010)
12. Y. Wang, J. Wang, W. Liu, K. Zhang, J. Li, Development of high current-density cathodes with scandia-doped tungsten powders. IEEE Trans. Electron Dev. **54**(5), 1061–1070 (2007)
13. J.D. Jarvis, B.K. Choi, A.B. Hmelo, B. Ivanov, C.A. Brau, Emittance measurements of electron beams from diamond field emitter arrays. J. Vac. Sci. Technol. B **30**(4), 042201–042205 (2012)
14. P.M. Phillips, C. Hor, L. Malsawma, K.L. Jensen, E.G. Zaidman, Design and construction of apparatus for characterization of gated field emitter array electron emission. Rev. Sci. Instrum. **67**(6), 2387–2393 (1996)
15. K.L. Jensen, P.G. O'Shea, D.W. Feldman, J.L. Shaw, Emittance of a field emission electron source. J. Appl. Phys. **107**(1), 014903–014914 (2010)
16. Y. Ding, A. Brachmann, F.J. Decker, D. Dowell, P. Emma, J. Frisch, S. Gilevich, G. Hays, P. Hering, Z. Huang, R. Iverson, H. Loos, A. Miahnahri, H.D. Nuhn, D. Ratner, J. Turner, J. Welch, W. White, J. Wu, Measurements and simulations of ultralow emittance and ultrashort electron Beams in the Linac coherent light source, Phys. Rev. Lett. **102**(25), 254801 / 1–4 (2009)
17. D.H. Dowell, I. Bazarov, B. Dunham, K. Harkay, C. Hernandez-Garcia, R. Legg, H. Padmore, T. Rao, J. Smedley, W. Wan, Cathode RandD for future light sources. Nucl. Instr. Methods Phys. Res. Sect. A, 622685–697 (2010)

18. G.R. Neil, L. Merminga, Technical approaches for high-average-power free-electron lasers. Rev. Mod. Phys. **74**(3), 685–701 (2002)
19. J.J. Petillo, C. Kostas, D. Panagos, S. Ovtchinnikov, A. Burke, T.M. Antonsen, E.L. Wright, K.T. Nguyen, E. Nelson, B.L. Held, J.F. DeFord, K.L. Jensen, J.A. Pasour, B. Levush, L. Ludeking, Electrostatic time-domain PIC simulations of RF density-modulated electron sources with MICHELLE, in IEEE International Vacuum Electronics Conference/International Vacuum Electron Sources Conference (IVEC/IVESC), 341–342 (2012)
20. D.H. Dowell, J.F. Schmerge, Quantum efficiency and thermal emittance of metal photocathodes, Phys. Rev. ST Accel. Beams **12**(7), 074201 / 1–10 (2009)
21. J.R. Maldonado, P. Pianetta, D.H. Dowell, J. Corbett, S. Park, J. Schmerge, A. Trautwein, W. Clay, Experimental verification of the 3-step model of photoemission for energy spread and emittance measurements of copper and CsBr-coated copper photocathodes suitable for free electron laser applications. Appl. Phys. Lett. **101**(23), 231103–231104 (2012)
22. I.V. Bazarov, B.M. Dunham, C.K. Sinclair, Maximum achievable beam brightness from photoinjectors, Phys. Rev. Lett. **102**(10), 104801 / 1–4 (2009)
23. T. Vecchione, I. Ben-Zvi, D.H. Dowell, J. Feng, T. Rao, J. Smedley, W. Wan, H.A. Padmore, A low emittance and high efficiency visible light photocathode for high brightness accelerator-based X-ray light sources. Appl. Phys. Lett. **99**(3), 034103–034103 (2011)
24. S. Karkare, I.V. Bazarov, Effect of nanoscale surface roughness on transverse energy spread from GaAs photocathodes, Appl. Phys. Lett. **98**(9), 094104 / 1–3 (2011)
25. H.J. Qian, C. Li, Y.C. Du, L.X. Yan, J.F. Hua, W.H. Huang, C.X. Tang, Experimental investigation of thermal emittance components of copper photocathode. Phys. Rev. ST Accel. Beams **15**(4), 040102–040108 (2012)
26. A.C. Keser, T.M. Antonsen, G.S. Nusinovich, D.G. Kashyn, K.L. Jensen, Heating of Microprotrusions in accelerating structures, Phys. Rev. ST Accel. Beams **16**092001 / 1–8 (2013)
27. J. Norem, Z. Insepov, I. Konkashbaev, Triggers for RF breakdown. Nucl. Instr. Methods Phys. Res. Sect. A **537**(3), 510–520 (2005)
28. P.B. Wilson, Formation of Taylor cones on a molten metal surface followed by ion injection into the vacuum, High Gradient Workshop (2007)
29. A.R. Knox, A. Asenov, A.C. Lowe, An electron emission model for use with 3D electromagnetic finite element simulation. Solid-State Electron. **45**(6), 841–851 (2001)
30. E.M. Nelson, J.J. Petillo, An analysis of the basic space-charge-limited emission algorithm in a finite-element electrostatic gun code. IEEE Trans. Plas. Sci. **32**(3), 1223–1235 (2004)
31. J.J. Petillo, E.M. Nelson, J.F. Deford, N.J. Dionne, B. Levush, Recent developments to the MICHELLE 2-D/3-D electron gun and collector modeling code. IEEE Trans. Electron Dev. **52**(5), 742–748 (2005)
32. Y. Feng, J. Verboncoeur, A model for effective field enhancement for Fowler-nordheim field emission, Phys. Plasmas **12**(10), 103301 / 1–6 (2005)
33. K.L. Jensen, J.J. Petillo, E.J. Montgomery, Z. Pan, D.W. Feldman, P.G. O'Shea, N.A. Moody, M. Cahay, J.E. Yater, J.L. Shaw, Application of a general electron emission equation to surface nonuniformity and current density variation. J. Vac. Sci. Technol. B **26**(2), 831–837 (2008)
34. R.G. Forbes, Exact analysis of surface field reduction due to field-emitted vacuum space charge, in parallel-plane geometry, using simple dimensionless equations. J. Appl. Phys. **104**(8), 084303 (2008)
35. K.L. Jensen, J. Lebowitz, Y.Y. Lau, J. Luginsland, Space charge and quantum effects on electron emission. J. Appl. Phys., 111054917–054919 (2012)
36. A. Rokhlenko, K.L. Jensen, J.L. Lebowitz, Space charge effects in field emission: one dimensional theory. J. Appl. Phys. **107**(1), 014904–014910 (2010)
37. R.A. Millikan, C.F. Eyring, Laws governing the pulling of electrons out of metals by Intense electrical fields. Phys. Rev. **27**(1), 51–67 (1926)
38. K.L. Jensen, A tutorial on electron sources. IEEE Trans. Plas. Sci. **46**(6), 1881–1899 (2018)
39. K.L. Jensen, *Introduction to the Physics of Electron Emission* (Wiley, Hoboken, New Jersey, 2017)
40. S. Dushman, Thermionic emission. Rev. Mod. Phys. **2**(4), 0381–0476 (1930)

41. C. Herring, M.H. Nichols, Thermionic emission. Rev. Mod. Phys. **21**(2), 185–270 (1949)
42. O.W. Richardson, Electron emission from metals as a function of temperature. Phys. Rev. **23**(2), 153–155 (1924)
43. E.L. Murphy, R.H. Good, Thermionic emission, field emission, and the transition region. Phys. Rev. **102**(6), 1464–1473 (1956)
44. R.H. Fowler, L. Nordheim, Electron emission in intense electric fields. Proc. R. Soc. A **119**(781), 173–181 (1928)
45. R.G. Forbes, Mainstream Theory of Field and thermal Electron Emission, p. Chapter 9 of this book
46. J.H.B. Deane, R.G. Forbes, The formal derivation of an exact series expansion for the principal Schottky-Nordheim barrier function v, using the Gauss hypergeometric differential equation, J. Phys. A: Math. Theor. **41**(39), 395301 / 1–9 (2008)
47. R.G. Forbes, Simple good approximations for the special elliptic functions in standard Fowler-Nordheim tunneling theory for a Schottky-nordheim barrier, Appl. Phys. Lett. **89**(11), 113122 / 1–3 (2006)
48. K.L. Jensen, Field Emission: Fundamental Theory to Usage, in *Wiley Encyclopedia of Electrical and Electronics Engineering*, ed. by J.G. Webster (Wiley, New York, 2014), pp. 1–29
49. L.A. DuBridge, A further experimental test of Fowler's theory of photoelectric emission. Phys. Rev. **39**(1), 108–118 (1932)
50. R.H. Fowler, The analysis of photoelectric sensitivity curves for clean metals at various temperatures. Phys. Rev. **38**(1), 45–56 (1931)
51. J.W. Gadzuk, E.W. Plummer, Energy distributions for thermal field emission. Phys. Rev. B **3**(7), 2125–2129 (1971)
52. M.J. Fransen, T.H.L. Van Rooy, P.C. Tiemeijer, M.H.F. Overwijk, J.S. Faber, P. Kruit, On the electron-optical properties of the ZrO/W Schottky electron emitter, in Advances in Imaging and Electron Physics III, edited by P. Hawkes 91–166 (1999)
53. P. Geittner, G. Gärtner, D. Raasch, Low temperature properties of Ba-dispenser cathodes. J. Vac. Sci. Technol. B **18**(2), 997–999 (2000)
54. S. Tsujino, F. le Pimpec, J. Raabe, M. Buess, M. Dehler, E. Kirk, J. Gobrecht, A. Wrulich, Static and optical field enhancement in metallic nanotips studied by two-photon photoemission microscopy and spectroscopy excited by picosecond laser pulses, Appl. Phys. Lett. **94**(9), 093508 / 1–3 (2009)
55. K.L. Jensen, Electron emission theory and its application: Fowler-Nordheim equation and Beyond. J. Vac. Sci. Technol. B **21**(4), 1528–1544 (2003)
56. K.L. Jensen, On the application of quantum transport theory to electron sources. Ultramicroscopy **95**(1–4), 29–48 (2003)
57. K.L. Jensen, Electron Emission Physics, in *Advances in Imaging and Electron Physics*, vol. 149, ed. by P. Hawkes (Elsevier, San Diego, CA, 2007), pp. 1–338
58. K.L. Jensen, General formulation of thermal, field, and photoinduced electron emission. J. Appl. Phys. **102**(2), 024911–024911 (2007)
59. K.L. Jensen, B. Jensen, E.J. Montgomery, D.W. Feldman, P.G. O'Shea, N.A. Moody, Theory of photoemission from cesium antimonide using an alpha-semiconductor model. J. Appl. Phys. **104**(4), 044907–044910 (2008)
60. K.L. Jensen, A quantum dipole–modified work function for a simplified electron emission barrier, J. Appl. Phys. **111**054916 / 1–10 (2012)
61. K.L. Jensen, Chpt. 3: Theory of Field Emission, in *Vacuum Microelectronics*, ed. by W. Zhu (Wiley, New York, 2001), pp. 33–104
62. R. Kubo, *Statistical Mechanics, an Advanced Course With Problems and Solutions* (North-Holland Pub. Co. Interscience Publishers, Amsterdam, New York, 1965)
63. P.H. Cutler, D. Nagy, Use of a new potential barrier model in Fowler-Nordheim theory of field emission. Surf. Sci. **3**(1), 71–94 (1965)
64. W.W. Dolan, W.P. Dyke, Temperature-and-field Emission of electrons from metals. Phys. Rev. **95**(2), 327–332 (1954)

65. R.G. Forbes, Use of energy-space diagrams in free-electron models of field electron emission. Surf. Interface Anal. **36**(5–6), 395–401 (2004)
66. L.A. DuBridge, Theory of the energy distribution of photoelectrons. Phys. Rev. **43**(9), 0727–0741 (1933)
67. J.W. Gadzuk, E.W. Plummer, Field emission energy distribution (FEED). Rev. Mod. Phys. **45**(3), 487–548 (1973)
68. A. Modinos, *Field, Thermionic, and Secondary Electron Emission Spectroscopy* (Plenum Press, New York, 1984)
69. R.D. Young, E.W. Müller, Experimental measurement of the total-energy distribution of field-emitted electrons. Phys. Rev. **113**(1), 115–120 (1959)
70. R.G. Forbes, Physics of generalized Fowler-Nordheim-type equations. J. Vac. Sci. Technol. B, 26788 (2008)
71. R.G. Forbes, J.H.B. Deane, Transmission coefficients for the exact triangular barrier: an exact general analytical theory that can replace Fowler and Nordheim's 1928 theory. Proc. R. Soc. A, 4672927–2947 (2011)
72. E.C. Kemble, A contribution to the theory of the BWK method. Phys. Rev. **48**(6), 549–561 (1935)
73. S.C. Miller, R.H. Good, A WKB-type approximation to the Schrödinger equation. Phys. Rev. **91**(1), 174–179 (1953)
74. K.L. Jensen, A reformulated general thermal-field emission equation, J. Appl. Phys. **126**(6), 065302 / 1–13 (2019)
75. K.L. Jensen, M. McDonald, J.R. Harris, D.A. Shiffler, M. Cahay, J.J. Petillo, Analytic model of a compound thermal-field emitter and its performance. J. Appl. Phys. **126**(24), 245301 (2019)
76. R.G. Forbes, On the need for a tunneling pre-factor in Fowler-Nordheim tunneling theory. J. Appl. Phys. **103**(11), 114911 (2008)
77. K.L. Jensen, M. McDonald, O. Chubenko, J.R. Harris, D.A. Shiffler, N.A. Moody, J.J. Petillo, A.J. Jensen, Thermal-field and photoemission from meso- and micro-scale features: effects of screening and roughness on characterization and simulation, J. Appl. Phys. **125**(23), 234303 / 1–25 (2019)
78. A. Kyritsakis, J.P. Xanthakis, Extension of the general thermal field equation for Nanosized Emitters, J. Appl. Phys. **119**(4), 045303 / 1–6 (2016)
79. F.M. Charbonnier, R.W. Strayer, L.W. Swanson, E.E. Martin, Nottingham effect in field and T-F emission: heating and cooling domains, and inversion temperature. Phys. Rev. Lett. **13**(13), 397–401 (1964)
80. W.B. Nottingham, Thermionic emission from tungsten and thoriated tungsten filaments. Phys. Rev. **49**(1), 78–97 (1936)
81. M.G. Ancona, Thermomechanical analysis of failure of metal field emitters. J. Vac. Sci. Technol. B **13**(6), 2206–2214 (1995)
82. G. Fursey, *Field Emission in Vacuum Microelectronics* (Kluwer Academic/Plenum Publishers, New York, 2005)
83. J. Norem, V. Wu, A. Moretti, M. Popovic, Z. Qian, L. Ducas, Y. Torun, N. Solomey, Dark current, breakdown, and magnetic field effects in a multicell, 805 MHz cavity, Phys. Rev. ST Accel. Beams **6(7)**, 072001 / 1–21 (2003)
84. R.G. Forbes, Simple derivation of the formula for Sommerfeld supply density used in electron-emission physics and limitations on its use. J. Vac. Sci. Technol. B **28**(6), 1326–1329 (2010)
85. K.L. Jensen, Y.Y. Lau, D.W. Feldman, P.G. O'Shea, Electron emission contributions to dark current and its relation to microscopic field enhancement and heating in accelerator structures, Phys. Rev. ST Accel. Beams **11(8)**, 081001 / 1–17 (2008)
86. C.A. Spindt, I. Brodie, L. Humphrey, E.R. Westerberg, Physical properties of thin-film field emission cathodes with Molybdenum cones. J. Appl. Phys. **47**(12), 5248–5263 (1976)
87. C.A. Spindt, C. Holland, P.R. Schwoebel, I. Brodie, Field emitter array development for microwave applications. II. J. Vac. Sci. Technol. B **16**(2), 758–761 (1998)

88. D.A. Shiffler, J. Luginsland, M. Ruebush, M. Lacour, K. Golby, K. Cartwright, M. Haworth, T. Spencer, Emission uniformity and shot-to-shot variation in cold field emission cathodes. IEEE Trans. Plas. Sci. **32**(3), 1262–1266 (2004)
89. D.A. Shiffler, W. Tang, K.L. Jensen, K. Golby, M. LaCour, J.J. Petillo, J.R. Harris, Effective field enhancement factor and the influence of emitted space charge, J. Appl. Phys. **118**(8), 083302 / 1–6 (2015)
90. K.L. Jensen, D.W. Feldman, P.G. O'Shea, Time dependent Models of Field-assisted photoemission. J. Vac. Sci. Technol. B **23**(2), 621–631 (2005)
91. A. Kyritsakis, G.C. Kokkorakis, J.P. Xanthakis, T.L. Kirk, D. Pescia, Self focusing of field emitted electrons at an ellipsoidal tip, Appl. Phys. Lett. **97**(2), 023104 / 1–3 (2010)
92. R.L. Hartman, W.A. Mackie, P.R. Davis, Use of Boundary-element methods in-field emission computations. J. Vac. Sci. Technol. B **12**(2), 754–758 (1994)
93. K.L. Jensen, E.G. Zaidman, M.A. Kodis, B. Goplen, D. Smithe, Analytical and seminumerical models for gated field emitter arrays. 1. Theory. J. Vac. Sci. Technol. B **14**(3), 1942–1946 (1996)
94. S. Watcharotone, R.S. Ruoff, F.H. Read, Possibilities for graphene for field emission: modeling studies using the BEM. Phys. Procedia **1**(1), 71–75 (2008)
95. C.J. Edgcombe, U. Valdre, Experimental and computational study of field emission characteristics from amorphous carbon single nanotips grown by carbon contamination—I. Experiments and computation. Philos. Mag. B **82**(9), 987–1007 (2002)
96. W. Schottky, Über kalte und warme Elektronenentladungen. Zeitschrift für Physik **14**(1), 63–106 (1923)
97. T.E. Stern, B.S. Gossling, R.H. Fowler, Further studies in the emission of electrons from cold metals. Proc. R. Soc. A **124**(795), 699–723 (1929)
98. W.P. Dyke, J.K. Trolan, W.W. Dolan, G. Barnes, The field emitter: fabrication, electron microscopy, and electric field calculations. J. Appl. Phys. **24**(5), 570–576 (1953)
99. K.L. Jensen, D.A. Shiffler, J.J. Petillo, Z. Pan, J.W. Luginsland, Emittance, surface structure, and electron emission. Phys. Rev. ST Accel. Beams **17**(4), 043402–043419 (2014)
100. K.L. Jensen, D.A. Shiffler, I.M. Rittersdorf, J.L. Lebowitz, J.R. Harris, Y.Y. Lau, J.J. Petillo, W. Tang, J.W. Luginsland, Discrete space charge affected field emission: flat and hemisphere emitters, J. Appl. Phys. **117**(19), 194902 / 1–17 (2015)
101. J.W. Lewellen, C.A. Brau, Rf photoelectric injectors using needle cathodes. Nucl. Instr. Methods Phys. Res. Sect. A **507**(1–2), 323–326 (2003)
102. R.G. Forbes, K.L. Jensen, New results in the theory of Fowler-Nordheim plots and the modelling of hemi-ellipsoidal emitters. Ultramicroscopy **89**(1–3), 17–22 (2001)
103. R.A. Kishek, S. Bernal, C.L. Bohn, D. Grote, I. Haber, H. Li, P.G. O'Shea, M. Reiser, M. Walter, Simulations and experiments with space-charge-dominated beams. Phys. Plasmas **10**(5), 2016–2021 (2003)
104. J.D. Zuber, K.L. Jensen, T.E. Sullivan, An analytical solution for microtip field emission current and effective emission area. J. Appl. Phys. **91**(11), 9379–9384 (2002)
105. J.R. Harris, K.L. Jensen, D.A. Shiffler, Modelling field emitter arrays using line charge distributions, J. Phys. D Appl. Phys. **48**(38), 385203 / 1–9 (2015)
106. P.B. Wilson, A theory for the comparative rf surface fields at destructive breakdown for various metels, SLAC-TN-06-003 (2006)
107. K.L. Jensen, Space charge effects in field emission: three dimensional theory. J. Appl. Phys. **107**(1), 014905–014909 (2010)
108. K.L. Jensen, Space charge, emittance, trajectories, and the modeling of field emitter arrays, J. Vac. Sci. Technol. B **29**(2), 02B101-02B1017 (2011)
109. P.R. Schwoebel, C.A. Spindt, C. Holland, High current, high current density field emitter array cathodes. J. Vac. Sci. Technol. B **23**(2), 691–693 (2005)
110. C. Spindt, C. Holland, P.R. Schwoebel, Thermal field forming of Spindt cathode arrays, J. Vac. Sci. Technol. B **33**(3), 03C108 / 1–3 (2015)
111. K.L. Jensen, D.A. Shiffler, J.R. Harris, J.J. Petillo, Schottky's conjecture, field emitters, and the point charge model, AIP Adv. **6**065005 / 1–6 (2016)

112. M. Cahay, W. Zhu, S. Fairchild, P.T. Murray, T.C. Back, G.J. Gruen, Multiscale model of heat dissipation mechanisms during field emission from carbon nanotube fibers, Appl. Phys. Lett. **108**(3), 033110 / 1–5 (2016)
113. V.L. Granatstein, R.K. Parker, C.M. Armstrong, Vacuum electronics at the dawn of the twenty-first century. Proc. IEEE **87**(5), 702–716 (1999)
114. C. Bostedt, J.D. Bozek, P.H. Bucksbaum, R.N. Coffee, J.B. Hastings, Z. Huang, R.W. Lee, S. Schorb, J.N. Corlett, P. Denes, P. Emma, R.W. Falcone, R.W. Schoenlein, G. Doumy, E.P. Kanter, B. Kraessig, S. Southworth, L. Young, L. Fang, M. Hoener, N. Berrah, C. Roedig, L.F. DiMauro, Ultra-fast and ultra-intense X-ray sciences: first results from the Linac Coherent Light Source free-electron laser, J. Phys. B: Atomic Molecul. Opt. Phys. **46**(16, SI), 164003 / 1–21 (2013)
115. P.H. Bucksbaum, N. Berrah, Brighter and faster: the promise and challenge of the X-ray free-electron laser. Phys. Today **68**(7), 26–32 (2015)
116. D.A. Dimitrov, D. Smithe, J.R. Cary, I. Ben-Zvi, T. Rao, J. Smedley, E. Wang, Modeling electron emission and surface effects from diamond cathodes, J. Appl. Phys. **117**055708 / 1–18 (2015)
117. K.L. Jensen, J.E. Yater, J.L. Shaw, R.E. Myers, B.B. Pate, J.E. Butler, T. Feygelson, Bunch characteristics of an electron beam generated by a diamond secondary emitter amplifier. J. Appl. Phys. **108**(4), 044509–044512 (2010)

Chapter 9
Renewing the Mainstream Theory of Field and Thermal Electron Emission

Richard G. Forbes

Abstract Mainstream field electron emission (FE) theory—the theory normally used by FE experimentalists—employs a Sommerfeld-type free-electron model to describe FE from a metal emitter with a smooth planar surface of very large extent. This chapter reviews the present state of mainstream FE theory, noting aspects of the history of FE and thermal electron emission theory. It sets out ways of improving the theory's presentation, with the ultimate aim of making it easier to reliably compare theory and experiment. This includes distinguishing between (a) emission theory and (b) device/system theory (which deals with field emitter behaviour in electrical circuits), and between ideal and non-ideal device behaviours. The main focus is the emission theory. Transmission regimes and emission current density regimes are discussed. With FE, a method of classifying different FE equations is outlined. With theories that assume tunnelling through a Schottky-Nordheim (SN) ("planar-image-rounded") barrier, a careful distinction is needed between the barrier form correction factor ν ("nu") and the special mathematical function v ("vee"). This function v is presented as dependent on the Gauss variable x. The pure mathematics of $\mathrm{v}(x)$ is summarised, and reasons are given for preferring the use of x over the older convention of using the Nordheim parameter y [$=+\sqrt{x}$]. It is shown how the mathematics of $\mathrm{v}(x)$ is applied to wave-mechanical transmission theory for basic Laurent-form barriers (which include the SN barrier). A brief overview of FE device/system theory defines and discusses different auxiliary parameters currently in use, outlines a preferred method for characterising ideal devices when using FN plots and notes difficulties in characterising non-ideal devices. The chapter concludes by listing some of the future tasks involved in upgrading FE science.

R. G. Forbes (✉)
Advanced Technology Institute & Department of Electrical and Electronic Engineering, University of Surrey, Guildford GU2 7XH, UK
e-mail: r.forbes@trinity.cantab.net

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*, Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_9

9.1 General Introduction

Electron emission from solid surfaces has many practical applications. Theoretical understanding of emission processes and of emitter characterisation helps electron source development. Understandings of how emitters can be fabricated and maintained in operation, and of how they fail, and of related theory, are also needed, but are not covered here.

In Chap. 8, Jensen has set out a detailed theory of field electron emission (FE) and of the transition between FE and thermal electron emission (TE). The present chapter's main aims are to discuss the wider background of this theory, including different conventions and terminology in use, and to discuss some issues relating to the theory of current-voltage characteristics. The chapter is set in the general context of FE textbooks/handbooks published in the last 30 years or so [1–14], and of recent reviews relating to FE theory (e.g. [15–21]).

The author now separates *emission theory* and *device/system theory*. The former provides a formula for the local emission current density (ECD), as a function of local work function, barrier field and temperature. Device/system theory converts this to a relation between measured current I_{m} and measured voltage V_{m}. Particularly in FE contexts, significant additional theory may be needed to accomplish this, and/or to validly interpret measured $I_{\mathrm{m}}(V_{\mathrm{m}})$ characteristics. Interpretation of FE $I_{\mathrm{m}}(V_{\mathrm{m}})$ data was recently discussed elsewhere [20, 21], and only a brief update is given here. Detailed interpretation theory for thermionic emitters is not covered.

In modern emission theory, it makes no sense to discuss field and thermal effects separately. This chapter presents an integrated approach, but at a more qualitative level than Chap. 8, and has the following overall structure. Section 9.2 deals with technical conventions; Sects. 9.3, 9.4, 9.5 and 9.6 deal with emission theory; Sect. 9.7 deals briefly with device and systems theory as these affect FE behaviour; and Sect. 9.8 with the interpretation of Fowler-Nordheim plots. Section 9.9 indicates some future needs. Appendix 9.1 lists universal constants used in field emission, and Appendix 9.2 lists data relating to the high-precision calculation of the FE special mathematical functions.

9.2 Technical Conventions

FE literature uses various technical conventions. This section describes the main alternatives, and aims to encourage a clearer formulation of FE theory. For reference, Table 9.1 presents a list of the various acronyms and abbreviations used in this chapter.

Table 9.1 Main acronyms and abbreviations used in Chap. 9

BT	Barrier top	L	Local
BTE	Barrier-top electron emission	LAFE	Large-area field electron emitter
C	Characteristic	LF	Low flyover
CFE	Cold field electron emission	m	Measured
CHTE	Classical high-temperature emission	M	Macroscopic
DT	Deep tunnelling	MG	Murphy-Good
ECD	Emission current density	PE	Potential energy
EPE	Electron potential energy	PPP	Parallel planar plate
ET	Exactly triangular	QMTE	Quantum-mechanical TE
FE	Field electron emission	SI	Système Internationale (of units)
FEF	Field enhancement factor	SN	Schottky-Nordheim
FN	Fowler-Nordheim	SPME	Smooth-planar-metal-like-emitter
GB	General barrier	ST	Shallow tunnelling
HF	High flyover	TE	Thermal electron emission
ISQ	International System of Quantities	VCF	Voltage conversion factor
JWKB	Jeffreys-Wentzel-Kramers-Brillouin	VCL	Voltage conversion length
k	Kernel	XC	Exchange-and-correlation

9.2.1 Equation Systems

In the 1970s, new International Standards defined a preferred system of quantities and equations for scientific communication (particularly between theoretical physicists and engineers) and a related unit system, the SI [22]. Since 2009 this system of quantities and equations, which was derived from the old metre-kilogram-second-ampere (mksa) system, and uses ε_0 in Coulomb's Law, has been called the *International System of Quantities* (ISQ) [22, 23]. Where clarity requires below, ISQ quantities are subscripted "ISQ".

The ISQ superseded earlier systems, particularly the Gaussian centimetre-gram-second (cgs) systems, and "atomic units" systems. The ISQ and SI are "master systems" that can define quantities used in earlier systems and the dimensionalities of units used. Thus, Gaussian elementary positive charge is denoted by e_s [$\equiv e_{ISQ}/(4\pi\varepsilon_0)^{1/2}$] and would have the SI unit $J^{1/2}\ m^{1/2}$, and the magnitude of Gaussian electrostatic field is denoted by F_s [$\equiv (4\pi\varepsilon_0)^{1/2}F_{ISQ}$] and would have the SI unit $J^{1/2}\ m^{-3/2}$.

In the 1960s, FE papers often employed a hybrid convention that used a Gaussian system for equations but an ISQ-like system to present experimental results (because laboratory equipment was calibrated in volts—a unit that does not exist in Gaussian systems). In experimental contexts, fields were often measured in "V/cm", which was a bastard unit created by taking one part (the "V") from the engineers' system, and other (the "cm") from the physicists' system.

The main FE equations that differ between the two systems are those describing (9.1) a Schottky-Nordheim (SN) tunnelling barrier of zero-field height H and (9.2) the related Nordheim parameter y

$$M^{\mathrm{SN}}(z) \;=\; H - e_{\mathrm{ISQ}}F_{\mathrm{ISQ}}z - e_{\mathrm{ISQ}}^2/16\pi\varepsilon_0 z \;=\; H - e_{\mathrm{s}}F_{\mathrm{s}}z - e_{\mathrm{s}}^2/4z, \tag{9.1}$$

$$y \;=\; (e_{\mathrm{ISQ}}^3 F_{\mathrm{ISQ}}/4\pi\varepsilon_0 H^2)^{1/2} \;=\; (e_{\mathrm{s}}^3 F_{\mathrm{s}}/H^2)^{1/2}. \tag{9.2}$$

Here, z is the distance measured from the emitter's electrical surface, and $M^{\mathrm{SN}}(z)$ is the electron motive energy for the SN barrier (see Sect. 9.3.3).

In papers, no subscripts were attached to symbols, and the same names and symbols were often used for both Gaussian and ISQ-like quantities. This led to two unhelpful 'features': context-dependent interpretations for the symbols 'e' and 'F', often in the same paper and sometimes in the same equation, and sometimes a need to convert a numerical value of F_{ISQ} to a numerical value of F_{s}, or vice versa, using a 'hidden conversion factor'. Values for such factors were rarely if ever written down in FE papers or FE textbooks, and were (and are) none too easy to find in general textbooks (though web search engines make conversions easier, nowadays). They were known to FE research workers of that era, and constituted 'secret knowledge' sometimes needed to interpret FE literature of that era in quantitative detail. One of many purposes of the 1970s changes was to abolish the need for 'features' like the above, by defining a single preferred system based firmly on quantity calculus/algebra.

Most new FE literature now uses the ISQ, but most older FE textbooks (including the reprinted seminal textbooks of Gomer [24] and Modinos [2]) and parts of some more recent ones [6, 9] are written using this 1960s-style hybrid system, and some FE material continues to be presented in this way. However, for most subjects in most universities in the world, the ISQ has been used almost exclusively for teaching for thirty years or more, and the knowledge needed to deal quantitatively with material written using the old hybrid system is no longer part of the background of many younger FE researchers. Thus, critical paper and textbook equations written using the old system are, in a certain sense, written in a theoretical language often not fully understood by many younger researchers.

The point is that it is now 40 years or more since the Governments of the world, via their national standards bodies, agreed to introduce (what is now called) the ISQ as a common language for the scientific communication of equation-based arguments. Editors of modern journals have no obligation to accept for publication FE papers that are written with equations in Gaussian or other obsolete equation systems, any more than they would have an obligation to publish papers partly written in Latin (the language of scientific communication of Newton's day). I accept the reality that obsolete conventions take a long time to fade away, but I strongly urge that all new FE literature is published with all equations written using the ISQ, especially when the publication is in English. Remaining discussion in this chapter uses the ISQ, except where otherwise indicated.

9.2.2 Other International Conventions About Equation Form

Two other parts of the international conventions about equations deserve mention. First, equations need to be dimensionally consistent, valid in all relevant units, and (except in the case of some formulae involving logarithms) to be 'stand alone' in the sense they can be exactly interpreted without reference to statements in the surrounding text. For example, the ISQ version of (9.2) is written correctly in any of the forms in (9.3), but forms (9.4a) and (9.4b) are considered formally incorrect.

$$y = (e^3F/4\pi\varepsilon_0H^2)^{1/2} = c_{\mathrm{S}}F^{1/2}/H$$
$$\cong (1.199985\,\mathrm{eV\,V^{-1/2}\,nm^{1/2}})\,F^{1/2}/H \cong (3.794686\times10^{-5}\,\mathrm{eV\,V^{-1/2}\,m^{1/2}})\,F^{1/2}/H\,. \quad (9.3)$$

$$y \cong 1.199985\,F^{1/2}/H, \quad (9.4a)$$

$$y \cong 3.794686\times10^{-5}\,F^{1/2}/H. \quad (9.4b)$$

Here, c_{S} is the *Schottky constant*, as defined in Appendix 9.1 (also see Sect. 9.3.4). This international convention is not always observed in current FE literature, but is strongly recommended. Usually, the simplest procedure is to put into the equation a symbol denoting a constant, and give its value separately, using appropriate units.

A second convention [25] is that, with the symbol $\ln\{x\}$, and related logarithmic symbols, the curly brackets around x are allowed to mean 'the numerical value of x, when x is measured in specified units'. The specified units should be given in the text or (with diagrams) in the axis label or the figure caption, except when the discussion is valid in all relevant units. Alternatively, write $\ln\{x/(\text{'units of } x\text{'})\}$, but this approach is often cumbersome.

9.2.3 The Meaning of the Symbol 'e'

Present international conventions use the symbol 'e' (typeset italic) to denote the elementary (positive) charge and the symbol 'e' (typeset upright) to denote the base of natural logarithms. However, some publishers are not systematic about this in their typesetting and may use italic 'e' to denote the base of natural logarithms. This can cause confusion, particularly when authors do not define the meaning of their symbol 'e' and when the typesetting is changed (or not corrected) during publication.

In his Chap. 8, and elsewhere, Jensen has avoided this problem by using 'q' to denote the elementary charge, and other authors also use this convention. The following is an unambiguous convention more consistent with the International Standard. Use 'e' to denote the elementary charge. Use '$\exp(x)$' to denote exponents, rather than e^{x} or e^{x}, and use $\exp(1)$ rather than e or e, except when

typesetting incompatible with the international conventions cannot conceivably be misunderstood.

9.2.4 Conventions Concerning the Term 'Field'

In most cases, the 'field' discussed in ordinary FE contexts is an electrostatic field. It is better to describe these fields as 'electrostatic' rather than 'electric', because the latter term also includes the electric component of a travelling electromagnetic wave. Laser-induced effects are of increasing interest in FE contexts (e.g. [26, 27]), but there are unresolved fundamental issues (e.g. [28]) relating to whether the physical effects of an electrostatic field are the same as those of an electromagnetic E-field, and whether an electrostatic field is physically the low frequency limit of an electromagnetic E-field. It is safer to assume that they may not be.

For clarity in what follows, I use the term 'classical' to label quantities that are defined strictly in accordance with the conventions of classical electrostatics as used in standard textbooks (e.g. [29, 30]). In a one-dimensional (z) situation, these imply the following. If under the influence of a classical electrostatic field, a positive test particle (with a vanishingly small charge) moves in the positive z-direction, then the field is considered positive. In FE, electron emission is usually deemed to take place in the positive z-direction. Hence, the classical electrostatic field $\boldsymbol{E}^{\mathbf{ES}}$ applied to a field electron emitter is negative in value, and so, of course, are the classical emission current and current density. In FE literature, five different conventions are used to deal with 'fields'.

(1) In detailed discussions of electrostatics as it applies to field emitters (for example, the electrostatic depolarization effects that occur in emitter arrays), it usually clearest to stick firmly to classical electrostatic conventions, and (where appropriate) insert modulus signs around the symbols for field and for currents and related quantities.

However, it is more usual for FE papers (and other electron emission papers) to use the so-called *electron emission convention* in which fields and electron currents and current densities are treated as if they were positive. In FE, this convention about fields comes in three variants.

(2) Most modern experimental FE papers use the symbol E to denote the *absolute magnitude* of a classical electrostatic field that is actually negative. Since this E is usually simply called the 'field', this can be a confusing convention for non-experts. It can also be unhelpful in electrostatics contexts.

(3) An alternative uses the symbol F to denote the absolute magnitude of the classical electrostatic field. This convention goes back to at least the work of Dyke and colleagues (e.g. [31]), and was used in Gomer's influential textbook [24]. It was used widely in experimental FE literature in the period 1960–1990, and is still used in much theoretical FE literature, because theoreticians often find

it useful to have 'E' available to denote electron energy. Use of F to denote positive electrostatic fields is much older; in particular, J.J. Thomson used F in his paper [32] announcing the identification of the electron.

(4) A formal alternative—which, in practice, is usually equivalent to the convention (3)—is to define a new field-like vector quantity $\boldsymbol{F}$ by

$$\boldsymbol{F} = -\boldsymbol{E}^{\mathrm{ES}}, \tag{9.5}$$

and to specify that the symbol F represents the *signed magnitude* of $\boldsymbol{F}$. Obviously, in the context of electron emission theory, both $\boldsymbol{F}$ and F are positive quantities.

(5) A fifth convention is to define a quantity F^* by the equation

$$\boldsymbol{F}^* = -e\boldsymbol{E}^{\mathrm{ES}}. \tag{9.6}$$

This quantity F^* (usually F in the literature) is positive in the context of FE, and is best defined—essentially following Stern et al. [33])—as the negative of the gradient of the electrostatic component of the potential energy (PE) of a hypothetical classical point electron. F^* has the SI unit 'newton'.

The quantity F^* has no well recognised name in quantum mechanics. Jensen [14] calls F^* a 'field', but this name invites confusion with the ISQ quantities measured in V/m. The least worst alternative might be to call F^* the 'classical electrostatic force' on the electron, though in reality F^* is the force on a hypothetical classical point electron, not the force on a real distributed wave-like electron that is taking part in a tunnelling process. In Chap. 8, this quantity F^* is denoted by the symbol 'F' and is called 'force'.

Merits of using F^* are that it is a positive quantity, and the same in both the Gaussian system and the ISQ. Its use also has the typographical advantage of removing the symbol for an elementary positive charge from various formulae. However, this is also its main disadvantage. Any formula that in conventions (1)–(4) contains a 'field' measured in V/m (or dimensionally equivalent units), would need modifying for the convention (5). Also, new versions of several universal constants used in FE would need to be introduced: these new constants would have different dimensions and units, though numerical values in FE customary units (see below) would not change.

Classical electrostatic force was used to describe FE potential-energy barriers in the 1929 Stern et al. paper [33], and (before that) in the middle part of the 1928 Fowler and Nordheim paper [34] (the technical meaning of the symbol "F" changes twice in the FN paper). If this were still the prevailing convention in FE, then probably one would not seek to change it. However, the author's view is that its widespread readoption at this point in time would inevitably introduce unfamiliarity and further confusion into an already confused subject, and would add to the difficulties of explaining FE theory clearly. Also, there might be intellectual resistance from many

experimentalists to the need to learn two slightly different sets of theoretical language in order to read basic material relating to FE theory. I thus prefer conventions (3) and (4) above, and use convention (3) in this chapter.

9.2.5 Field Emission Customary Units

In many situations, SI units are inconveniently large for discussing atomic-level issues. Thus, field electron and ion emission traditionally use a set of atomic-level units based on the eV as the unit of energy, the V/nm as the unit of electrostatic field (though sometimes the V/μm is more convenient), the elementary (positive) charge e as the size of the unit of charge (which, for technical reasons, has to be written as 1 eV V^{-1}), and the second (s) as the unit of time. This system has been called *field emission customary units*, and is discussed in more detail in [35]. Its merit is that it allows certain frequently needed basic FE calculations to be carried out quickly and efficiently.

It needs stressing that this is a system of units only, not a distinct equation system, and that (like SI units) it uses the ISQ as the underlying system of quantities and equations. Further, since all units in the FE customary system are officially recognised [22] for continued use alongside SI units, the system is entirely compatible with the ISQ and SI.

It is usual to express certain universal constants, widely used in FE theory, in these FE customary units. A list is given in Appendix 9.1. In this chapter, all relevant universal constants are stated to a precision of seven significant figures. Values have—in a few places where it is necessary—been updated to be consistent with the May 2019 changes in the SI system (see Wikipedia articles accessible via web search).

9.3 Emission Theory—General Issues

9.3.1 Smooth-Surface Conceptual Models

Except for carbon-nanotube emission theory and some other notable exceptions, e.g. [18, 36–38], nearly all present FE theory is formulated using what are called here *smooth-surface conceptual models* or "*Sommerfeld conductor*" *models* (see below). In such models, the atomic structure is disregarded. The emitter is represented, in a Sommerfeld-like way, as a region of space where the electron potential energy (EPE) is constant at the 'conduction-band-edge value' E_c, with this region limited by a 'formal well boundary'. Outside the boundary, the EPE is modelled either as rising abruptly to the local vacuum level E_{vac}, or as rising more slowly to this level, as a result of exchange-and-correlation (XC) forces. These forces give rise to an XC contribution to the EPE, with the contribution usually modelled as a classical

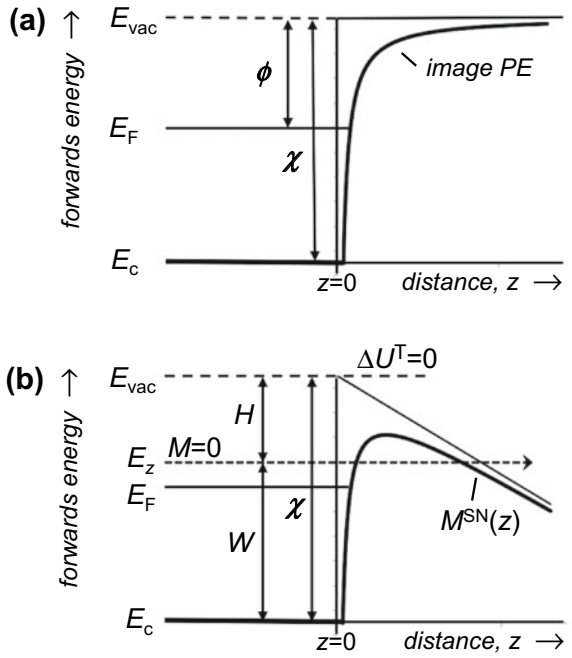


Fig. 9.1 Sommerfeld-type electron-energy diagrams, for an emitter with a smooth planar surface of large extent. **a** With image potential energy (PE) but in the absence of field. **b** The Schottky-Nordheim (SN) barrier (includes both image PE and field). Note that in both cases the vertical axis shows forwards energy, not total electron energy

image PE that cuts off at the point where it goes below E_c (see Fig. 9.1a, for the planar-surface case).

For simplicity, FE theory usually takes the local inner PE χ [$= E_{vac}–E_c$], and hence the local work function ϕ [$= (E_{vac}–E_F)$, where E_F is the emitter Fermi level], to be the same at all points on the well boundary. In fact, real emitters usually have many small surface areas (or "*patches*") with different local work functions, and (even in the absence of any applied voltage) are surrounded by a system of electrostatic fields called *patch fields*. With thermally induced emission, at very low applied voltages, these patch fields can cause significant local space-charge effects outside the emitter surface, well discussed in older review literature (e.g. [39]). In FE theory, the patch field adds to the voltage-induced component of surface field (and contributes to causing field inhomogeneities across the surface). However, for simplicity, patch fields are usually disregarded in mainstream FE theory.

Inside the well, electrons are treated, in the usual Sommerfeld way [40–42], as non-interacting electrons subject to the laws of quantum mechanics and statistical mechanics, with the electron distribution assumed to be in thermodynamic equilibrium and obedient to Fermi-Dirac statistics.

If the well boundary is assumed planar, and the well itself is assumed to be of infinite width (or of very large extent), then this is the usual well known 'Sommerfeld free-electron model'. However, advanced FE theory needs to consider equivalent models (without atomic structure) where the formal well boundary is curved and quantum confinement effects determine the electron wave-functions inside the well.

There seems to be no well recognised name for a model of this kind; I refer to it as a *Sommerfeld conductor*.

When an electrostatic field is created outside the emitter, by applying a voltage between it and a counter-electrode, the charge distribution induced at the emitter surface (by the voltage) is usually modelled (in a Sommerfeld conductor model) as confined to an infinitesimally thin surface layer. This surface layer defines the emitter's *electrical surface* within the framework of the model, and is taken as co-located with the formal well boundary. The resulting EPE variation is illustrated in Fig. 9.1b, for the Schottky-Nordheim barrier defined below. Note that the vertical axis represents the electron's *forwards energy*, i.e. the energy component associated with the direction normal to the emitter surface, which is called here the *forwards direction*. This name 'forwards energy' is an alternative name for what is more usually called the 'normal energy', and is used because it seems more instructive.

The electron wave-functions are assumed to take an appropriate non-atomic form, which will be a plane wave if the formal well boundary is planar or 'nearly planar'. These wave-functions are then used to evaluate values for the transmission probability D for an electron approaching the emitter surface from the inside, as a function of forwards energy and the *local barrier field* F_{L}. This field F_{L} is the local value of the parameter F, as taken in the emitter's electrical surface. For notational simplicity in what follows, we omit the subscript 'L' from local-field values.

9.3.2 The Problems of Smooth-Surface Models

Smooth-surface conceptual models disregard the existence of atoms, disregard the role of atom-level wave-functions in tunnelling theory and are not self-consistent with respect to the location of induced surface charge. Their use in FE theory can be described as *smooth-planar-metal-like-emitter (SPME) methodology.*

These physical assumptions are wildly unrealistic for real emitters, particularly for carbon nanotubes. Hence, the predictions of SPME methodology about current densities are not accurate. As discussed in Sect. 9.6.1, one can take this uncertainty into account by putting into core FE equations a general-purpose pre-exponential correction factor λ_{FE} of unknown functional dependence and unknown values. In the context of the Murphy-Good zero-temperature FE equation (see below), my present thinking [43] is that a smooth-surface model might do anything between under-predicting by a factor of around 14 and over-predicting by a factor of around 200. In other words, the relevant correction factor $\lambda_{\mathrm{FE}}^{\mathrm{SN}}$ has unknown functional dependence but probably lies in the range of $0.005 < \lambda_{\mathrm{FE}}^{\mathrm{SN}} < 14$. At present, we do not know how badly the models really behave, although one might speculate that over-prediction by a factor of 10 is more plausible than either extreme.

The problems of developing accurate FE theory are intensively difficult, and quite likely at the current boundaries of theoretical physics.[1] It seems likely that the *experimental* determination of λ_{FE} would be *relatively* easier and quicker. However, this is

[1] Prof. Marshall Stoneham (now deceased), a former President of the UK Institute of Physics, thought that some of the most difficult unsolved problems in theoretical physics were in field electron emission (private communication to the author, 2001).

relative. The necessary experiments seem likely to be difficult, time-consuming and expensive, and there seems no indication that funding bodies are currently inclined to fund basic physics work of this kind (or that experimentalists are prepared to carry it out).

Although sympathetic to the idea of advanced theoretical exploration, the author does not see how real progress will be made by adding purely theoretical investigation to purely theoretical investigation. Rather, I see the best route forwards as based on developing a more *scientific* approach to the smooth-surface conceptual models, that will enable them to be reliably tested against experiment, so that the actual size of discrepancies can be established. I see the immediate need as 'theoretical tidying up' of these models, starting with the planar-surface models. Later, one can investigate how to do reliable experiments, and seek support for this.

The main purpose of this chapter is to contribute to this "tidying up" of the planar-surface models. Beyond this, there are important issues relating to methods for more precise FE $I_{\mathrm{m}}(V_{\mathrm{m}})$ data interpretation for real emitters that have the shape of a pointed needle or a rounded post. This is an active topic of research (e.g., [14, 44, 45]), but details are beyond the scope of this chapter.

9.3.3 Barrier Form, Electron Motive Energy and Barrier Strength

The details of the barrier transmission theory depend on the barrier's mathematical form. This is best described via a quantity $M(z)$, called the electron motive energy, derived as follows: The Schrödinger equation is separated in Cartesian coordinates, and the component relating to motion normal to the emitter surface is written

$$\left[\hat{K}_z + \{U^{\mathrm{T}}(z) - E_z\}\right]\Psi_z = \left[\hat{K}_z + M(z)\right]\Psi_z = 0 \tag{9.7}$$

where z is the forwards direction, Ψ_z is the relevant wave-function component, $\hat{K}_z$ is the relevant kinetic-energy operator, $U^{\mathrm{T}}(z)$ is the total electron potential energy (EPE) and E_z is the forwards electron energy. $U^{\mathrm{T}}(z)$ and E_z (and also E_{c}, E_{vac} and E_{F}) can be taken as all measured relative to any well defined energy reference level (often the laboratory "Earth"). The symbol W is used to denote forwards energy measured relative to the PE well base (i.e. $W = E_z – E_{\mathrm{c}}$).

For an electron with forwards energy E_z, the *electron motive energy M(z)* is then defined by

$$M(z) = U^{\mathrm{T}}(z) - E_z = [U^{\mathrm{T}}(0) - E_z] + [U^{\mathrm{T}}(z) - U^{\mathrm{T}}(0)] \equiv H + \Delta U^{\mathrm{T}}(z), \tag{9.8}$$

where the *zero-field barrier height* $H \equiv \chi - W = U^{\mathrm{T}}(0) - E_z$ (see Fig. 9.1b). $\Delta U^{\mathrm{T}}(z)$ is defined by (9.8), and describes how the EPE varies outside the emitter surface, relative to the local vacuum level.

The term 'motive' was used by Newton [46], and also in early discussions of electron motion near conducting surfaces (e.g. [39, 47]), but the term 'motive energy' is used here in a slightly different and more specialised sense.

As is well known, different *barrier forms* are defined by selecting different mathematical forms for $\Delta U^{\mathrm{T}}(z)$ and hence $M(z)$. Those of most interest here are the planar forms called the *exactly triangular (ET)* barrier [defined by (9.9)] and the *Schottky-Nordheim (SN)* barrier [defined by (9.10)], but there is also current interest in forms that represent the EPE variation outside emitters with quasi-spherical apexes. For the ET and SN barriers, the motive-energy expressions are

$$M^{\mathrm{ET}}(z) \; = \; H \, - \, eFz \quad (z > 0) \tag{9.9}$$

$$M^{\mathrm{SN}}(z) \; = \; H \, - \, eFz \, - \, e^2/16\pi\varepsilon_0 z \quad (z > z_{\mathrm{c}}) \tag{9.10}$$

where the *cut-off distance* z_{c} is defined by the condition $M^{\mathrm{SN}}(z_{\mathrm{c}}) = H - \chi = -W$. To the left of these regions of validity we have $M(z) = H - \chi = -W$ (see Fig. 9.1).

9.3.4 Image Potential Energy and the Schottky Effect

This Section relates to the *image PE term* ($-e^2/16\pi\varepsilon_0 z$) in (9.10). As is well known, this describes the electrical PE of interaction between a classical point electron and a perfectly conducting infinite planar conductor, and is a model for quantum-mechanical exchange-and-correlation effects outside planar surfaces. The complicated historical origin of this term seems less well known, and is recorded here.

The idea of electrical images was introduced by W. Thomson (later Lord Kelvin) in 1847 [48], when discussing the reaction of an isolated conducting sphere to external charges. Later [49 but (better) see p. 67 in 50], he showed that, if the sphere were charged and both the sphere and an external point charge had the same sign, the force on the external charge would change from repulsive to attractive as the external charge is put closer to the sphere, i.e. (in modern terms) a PE barrier would exist.

In fact, the first person to use electrostatic potentials and electrostatic energies of mutual interaction, when discussing image effects relating to spherical conductors, was Maxwell [51]. These early treatments were algebraically complicated, and not fully correct in the 1st edition of [51]; a clearer discussion was given later by Landau et al. (see §3 in [30]).

Thomson [49 but (better) see p. 73 in 50] and Maxwell [51] also discussed image effects outside an (infinite) *planar* surface, but their thinking was too heavily linked to the spherical-conductor situation, and neither provided any formula useful to FE.

Thus, the Gaussian *planar* image-PE formula for an electron ($-e_s^2/4x$) is usually attributed to Schottky [52], although in fact, his 1914 discussion was in terms of the related positive *potential* due to the image charge. Schottky [52] also considered the case where the Gaussian classical electrostatic potential $\Delta\Phi_s$ (measured relative to the local vacuum *potential* level) increases as $\Delta\Phi_s = b_S z$. He found the Gaussian expression $(e_s b_S)^{1/2}$ for the height (above the vacuum potential level) of the minimum in the effective potential for an electron.

Obviously, the modern interpretation is that the image force causes a reduction, Δ_S, below the local vacuum *electron-energy* level, of the classical barrier for electron emission. This is the *Schottky effect* and we can write (using the ISQ)

$$\Delta_S = c_S F^{1/2} \equiv (e^3/4\pi\varepsilon_0)^{1/2} F^{1/2} \tag{9.11}$$

where c_S is the *Schottky constant* [$\cong$1.199985 eV $(\text{V/nm})^{-1/2}$, $\cong 3.794686\times10^{-5}$ eV $(\text{V/m})^{-1/2}$]. Schottky [52] gives, in his equation (6), the approximate numerical value of this constant, but for ISQ fields in V/cm, rather than V/m.

In fact, there had been an earlier suggestion that (what we now call) the Schottky effect could operate in electron emission. In his 1903 textbook (see p. 386 in [53]), J.J. Thomson suggested qualitatively that anomalies in spark physics, observed by Earhart [54] in 1901 for very small electrode gaps (less than 2 μm), might be explained if the combination of electrostatic and image forces made it energetically easier for electrons to leave the cathode. This proposal is clearly the Schottky effect (ten years before Schottky), but we would probably now attribute these anomalies to FE from nanoprotrusions, not observable in 1901, that had field enhancement factors of around 10–15.

Schottky [52] also noted the possible effect of this barrier reduction on electron emission, and then returned to the subject is his later (1923) paper [55], where he gives a formula[2] for (what I call) the *SN barrier reference field* F_R. This is the field that reduces to zero a SN barrier of zero-field height ϕ (i.e. the field that pulls the top of this barrier down to the Fermi level). This formula would now be written

$$F_R = c_S^{-2}\phi^2 = (4\pi\varepsilon_0/e^3)\phi^2 \cong (0.6944615\ \text{V/nm}) \cdot (\phi/\text{eV})^2. \tag{9.12}$$

For a material with $\phi = 4.50$ eV, $F_R \approx 14.1$ V/nm. Schottky [55] unsuccessfully proposed this kind of strong barrier lowering as an explanation of FE.

9.3.5 Scaled Field as a Modelling Parameter

In modern FE theory, this field F_R can be used to formally define a modelling parameter f that I now prefer to call the *scaled field* (for an SN barrier of zero-field height ϕ), by

[2]Reference [55] contains a typographical error, in that the numerical value in his equation (5) should be 6.97×10^6, as given correctly in Table 9.1 later in the paper. The modern value is slightly lower, as can be shown from (9.12).

$$f \equiv F/F_{\mathrm{R}} \equiv c_{\mathrm{S}}^2 \phi^{-2} F \cong (1.439965\ \mathrm{eV^2\ V^{-1}\ nm}) \cdot \phi^{-2} F. \tag{9.13}$$

For a barrier of zero-field height H, a parameter f_H is defined by a formula similar to (9.13), but with ϕ replaced by H. As seen later, f_H and f play important roles in modern FE theory. These parameters were previously called 'scaled barrier fields', but the simpler name 'scaled field' now seems better. (In the literature, f is also called 'dimensionless field'.)

Until recently, FE theory derivations used a modelling parameter y, called the *Nordheim parameter* and related to f_H by $y = +\sqrt{f_H}$. These modelling parameters enter FE theory in the following way. To evaluate the transmission probability for a SN barrier, using semi-classical methodology, the roots of the equation $M^{\mathrm{SN}}(z) = 0$ are needed, where $M^{\mathrm{SN}}(z)$ is given by (9.10). For a barrier of zero-field height H, these are

$$z = (H/2eF)\left[1 \pm \sqrt{1 - (e^3F/4\pi\varepsilon_0 H^2)}\right]. \tag{9.14}$$

To simplify the mathematics, a modelling parameter is introduced to represent the combination $(e^3F/4\pi\varepsilon_0 H^2)$. In 1928, Nordheim [56] used a squared quantity, the modern y^2. (The original paper used the symbol x^2.) We now know that this combination is an expression for the scaled field f_H, and that—for reasons described near the end of Sect. 9.6.3—f_H would have been a better choice. In the author's view, there would be significant advantages in now making the difficult transition from the use of y to use of f_H, or (when $H = \phi$) to use of f—certainly when discussing current-voltage characteristics.

9.3.6 Barrier Strength and the Barrier Form Correction Factor

An important parameter derived from $M(z)$ is the so-called *barrier strength* (also called the 'Gamow factor' and the 'JWKB factor'). Jensen denotes this by 'θ', but the present author prefers 'G'. *For tunnelling*, G is defined by

$$G \equiv 2\kappa_{\mathrm{e}} \int M^{1/2}(z)\mathrm{d}z, \tag{9.15}$$

where κ_{e} is a universal constant (see Appendix 9.1), and the integral is taken 'across the barrier', i.e. over the range of z where $M(z) \geq 0$. The barrier strength G is a mathematically evaluated parameter, with a well defined numerical value, that relates to barrier transmission properties. Strong barriers (high-positive-G barriers) are difficult for an electron to tunnel through.

For an exactly triangular (ET) barrier of zero-field height H ($H \geq 0$), it is readily shown that

$$G^{\mathrm{ET}} = bH^{3/2}/F, \tag{9.16}$$

where b [$\equiv 4\kappa_e/3e$] is a universal constant often known as the *second Fowler-Nordheim constant* (see Appendix 9.1). For a general barrier of arbitrary (but well-behaved[3]) mathematical form (and with $H \geq 0$), the related barrier strength G is derived via (9.15), and a *barrier form correction factor* $\nu_H(H,F)$ ('nu$_H$') is then defined via

$$G(H,F) \equiv \nu_H(H,F) \cdot G^{\mathrm{ET}}(H,F), \tag{9.17}$$

where the functional dependences on H and F are here shown explicitly. There may also be other functional dependencies, for example on the emitter apex radius.

For the SN barrier, the barrier form correction factor $\nu_H^{\mathrm{SN}}(H,F)$ is given by an appropriate particular value (corresponding to barrier height H and local barrier field F) of a special mathematical function denoted here by $\mathrm{v}(x)$. This function is now known [57] to be a very special solution of the Gauss Hypergeometric Differential Equation; the *Gauss variable* x is the independent variable in this equation. It can be shown (see below) that the correction factor $\nu_H^{\mathrm{SN}}(H,F)$ is obtained by setting $x = f_H$, i.e. $\nu_H^{\mathrm{SN}}(H,F) = \mathrm{v}(x = f_H)$. Hence

$$G^{\mathrm{SN}}(H,F) \equiv \mathrm{v}(f_H) \cdot bH^{3/2}/F. \tag{9.18}$$

Note that the author now prefers to typeset 'v' upright, on the grounds that this symbol represents a special mathematical function (like 'sin' or 'Ai'). In Russian-language literature, the notation $\vartheta(y)$ is often used, rather than $\mathrm{v}(f_H)$, but the factor has the same numerical value. The mathematics of $\mathrm{v}(x)$ is discussed further in Sect. 9.6.3.

9.3.7 'Thermal Electron' Versus 'Thermionic'

When discussing the theory of high-temperature (low-field) effects, the author prefers the name 'thermal electron emission' rather than 'thermionic emission', for the following reason. Immediately before the well known work [34] of Fowler and Nordheim (FN) in 1928, it was thought by some scientists (in particular, Millikan [58]) that electron emission at high temperatures was the emission of a special type of electron called a 'thermion' that, inside the metal, was in a special high-energy state. This process was called 'thermionic emission', to distinguish it from the field-induced emission of conduction electrons at low temperatures, as occurred in what was then called 'autoelectronic emission'. This was a kind of 'two-band' theory of metal electron behaviour. Possibly the most important scientific outcome of the FN 1928 paper was the conclusion that 'thermions' in this sense did not exist, and that both thermally induced and field-induced emission phenomena could be understood as the emission

[3] A well-behaved barrier has no regions inside it where $M(z)$ is negative.

of 'one kind of electron' from a 'single band', under different conditions of field and temperature. (Hence, an integrated 'single-band' theory covering both effects could be created). Thus, the FN 1928 paper was one of the key papers that led to the development of the modern single-band theory of metals. Since FN argued successfully that thermions did not exist, it seems scientifically anomalous to continue to talk about 'thermionic emission' in FE theory papers. Whilst recognising that many still use this term, particularly in practical applications, I prefer the name 'thermal electron emission' (TE) in theoretical contexts.

Also, phenomena arising from the simultaneous application of field and temperature are more commonly called 'thermal-field' than 'thermionic-field'. This seems another good reason for preferring 'thermal' over 'thermionic'.

9.3.8 The Concept of 'Wave-Mechanical Flyover'

'Wave-mechanics' is the old name for 'quantum mechanics'. I prefer it when no quantisation or inelastic phenomena affect barrier transmission, because the name is a reminder of analogies between electron tunnelling and (what we now call) photon tunnelling. Both are consequences of the classical mathematical behaviour of second-order differential equations, where what is now called 'tunnelling' was once called 'evanescent-wave coupling'.

It is well known that the transmission of electrons over barriers is affected by wave-mechanical reflection effects, if the transmission is not too high above the barrier. Some years ago, Forbes and Deane [35] suggested that 'ballistic emission' was not a good name, and that this wave-mechanical transmission process should be called *flyover*. This term will be used in this chapter.

For reflection at a rectangular step of height-in-energy χ, the transmission probability $D(w)$ for an electron approaching with forwards energy $\chi + w$ ($w > 0$) is [35, 59]

$$D(w) = 1 - \left[\frac{(\chi + w)^{1/2} - w^{1/2}}{(\chi + w)^{1/2} + w^{1/2}}\right]^2. \tag{9.19}$$

If χ is taken as 10 eV, then $D(w) = 0.9$ when $w \approx 3.7$ eV, which corresponds to a temperature $T = w/k_B$ (where k_B is the Boltzmann constant) around 40000 K. This illustrates that all real technological thermal electron emitters are subject to wave-mechanical reflection effects, as assumed by Fowler long ago [60]. For this reason, the emission regime in which the devices discussed in the next section operate is called here *quantum-mechanical thermal electron emission* (QMTE).

9.3.9 The History of Thermal Electron Emission Theory

In this chapter, thermal electron emission (TE) is defined as the thermally induced emission of electrons in circumstances where the vast majority of electrons escape over the top of a barrier that may be lowered by the Schottky effect, and where very few escape by tunnelling. This definition is used by most thermionic-emission experimentalists. However, as discussed below, Murphy and Good [61] (and theoreticians who follow their approach) use a different definition of the term 'thermionic'.

In the QMTE regime, the local emission current density (ECD) J_{RS} is given by

$$J_{\mathrm{RS}} = \lambda_{\mathrm{TE}} A_{\mathrm{R0}} T^2 \exp[-\{\phi - c_{\mathrm{S}} F^{1/2}\}/k_{\mathrm{B}} T], \tag{9.20}$$

where T is the local thermodynamic temperature of the emitter surface, ϕ is local work function (NOT the Schottky-reduced barrier height), A_{R0} is a universal constant [60] called here the *universal theoretical Richardson constant* (see Appendix 9.1), and λ_{TE} is a general-purpose pre-exponential correction factor due mainly (but not exclusively) to wave-mechanical elastic reflection effects.

The author prefers to call this equation the *quantum-mechanical Richardson-Schottky (RS) equation*, on the grounds that, although others contributed, and although neither of these authors, in fact, devised the final forms of either the constant A_{R0} or the equation itself, these two played the biggest roles in generating the equation. The 'Richardson-Schottky' name was used by Murphy and Good [61]. One often sees a version of this equation in which the factor λ_{TE} has been set equal to unity and omitted. I call this reduced equation the *classical Richardson-Schottky equation.*

Modern TE discussions are sometimes superficial about the origins of (9.20). Thus, this chapter seems a good place to summarise its slightly complicated history. Development took place in stages. I deal first with the zero-field equation.

Investigation of what are now known to be TE effects goes back well into the eighteenth century, e.g. [62, 63] as noted in the books of Guthrie [64], Richardson [65] and Reimann [66], although Edison [67] usually gets the credit for the first really careful relevant experiments. However, little real progress could be made until after Thomson [32] had identified what (following Stoney [68]) we now call the 'electron', and until after vacuum technology had advanced sufficiently. Later experimental developments are described in many books and review articles. References [39, 47, 65, 66, 69–71] cover the period up to about 1950 in some detail, and [72, 73] are modern overviews.

Richardson always took the view that the electrons originated from inside the metal emitters and in 1901 published a formula for emission current density (here J) based on applying Maxwell-Boltzmann statistics to what he modelled as a classical electron gas inside the metal [74]. Nowadays we would write his 1901 formula as

$$J = A_1 T^{1/2} \exp[-\chi/k_{\mathrm{B}} T] \tag{9.21}$$

where A_1 is a universal constant (not equal to A_{R0}). Note this equation contains the local inner PE χ, not a modern work-function-like parameter.

However, this basic idea that the electrons came from inside the metal (rather than from gas molecules in the 'vacuum' space or adsorbed on the metal surface) encountered significant resistance. This inhibited acceptance of Richardson's ideas (and also those of Einstein about the origin of photoelectrons). It was not until 1913 that Richardson was able to prove, experimentally and convincingly, that his thermally emitted electrons must be coming from inside the metal [75], and that they must be supplied by electrical conduction inside the metal.

In the meantime, because of various inconsistencies that arose, in that era, in the physics of electron behaviour in metals (in particular, about the specific heat of electrons), Richardson had become unhappy with treating electrons inside the metal as a classical gas. Instead, he applied a thermodynamic argument [65, 76] to an electron atmosphere outside the emitter that was taken to be in equilibrium with the electrons inside, arguing that the external electron atmosphere better resembled a classical gas. This yielded [65]

$$J = A_2 T^2 \exp[-\phi_0/k_B T] \tag{9.22}$$

where A_2 is a universal constant very much less in value than A_{R0} (in fact less than $A_{R0}/20$), and ϕ_0 is, in effect, the 'typical activation energy needed for escape, at low temperatures' (my term).

In 1923, by a broadly similar thermodynamic argument, but based in detail on Nernst's heat theorem, Dushman [77] revised the pre-exponential in (9.22) to be equal to $A_{R0}/2$ (but also gave an alternative thermodynamic argument that made the pre-exponential about 10% less than this).

In reality, Richardson's work does not make the clear distinction that we can now make (post Sommerfeld) between inner potential energy χ and zero-temperature work function ϕ_{ZT}; nor does he make the clear modern distinction between gas kinetic theory and thermodynamics. With the benefit of hindsight, and guided by the clearly stated equations on p. 285 of [42], I have: (a) interpreted Richardson's parameter that he defines [74] as 'the work done by a corpuscle[4] in passing through the surface layer' as the inner PE; and (b) interpreted his parameter that he defines [65] as 'the change in energy of the system which accompanies the transference of each electron from the hot body to the surrounding enclosure' as a work-function-like parameter. However, even in the 1921 edition of his book, Richardson seems to regard the definitions as equivalent, because he uses the same exponent in both his '$T^{1/2}$' and 'T^2' equations. In contrast, Dushman [77] thinks of his parameter b_0 (which is effectively the same as Richardson's ϕ_0) as a 'thermodynamic heat of vaporisation', for the evaporation of a single electron. This is clearly a work-function-like parameter.

Next, in 1928, Fowler [60], and almost simultaneously Sommerfeld himself [78], made use of Sommerfeld's new electron theory of metals [40], based on Fermi-Dirac statistics, and of the idea of 'electron spin' (e.g. [79]), to make critical improvements.

[4] 'Corpuscle' was J.J. Thomson's original term for what we now call the electron.

Fowler suggested that (in the framework of a Richardson-Dushman-type thermodynamic argument, and in the ideal case of no wave-mechanical reflection effects) the pre-exponential should be $A_{R0.}$ He also clearly identifies ϕ_0 in (9.22) with the modern zero-temperature work function ϕ_{ZT}.

Fowler [60] also reinserted into the pre-exponential a factor (1–r), originally introduced empirically by Richardson [65], where r is a coefficient that allows for electron reflection at the surface. Fowler [60] interpreted this as a wave-mechanical effect. Then, in 1929 [80], he replaced (1–r) by the *mean transmission probability* $\bar{D}$. The best modern discussion of reflection effects is probably still the 1984 book by Modinos [2]. Later, it became clear that other effects, such as inelastic scattering and non-thermodynamic-equilibrium conditions, could in principle influence the pre-exponential; thus, I now prefer to use the general-purpose correction factor λ_{TE}.

However, note that some accounts of TE theory prefer to combine ($\lambda_{TE}A_{R0}$) into a single parameter A_{eff}, called the 'Richardson constant', which is not a universal constant. Notation and terminology in this area are somewhat variable, and potentially confusing to non-experts. Also, note that in experiments there are effects that can cause real or apparent temperature dependence in the exponent. These effects need to be taken into account when interpreting experimental data, particularly Schottky plots (e.g. see [39, 71]).

In summary, although the zero-field version of the classical RS-type equation is variously called the 'Richardson equation', the 'Richardson-Dushman equation', and the 'Richardson-Laue-Dushman' equation, the post-1920s form of it is really due to Sommerfeld and to Fowler.

As shown above, the Schottky effect reduces the zero-field barrier height by Δ_S. Hence, ϕ in the zero-field equation is replaced by (ϕ–Δ_S), and the Richardson-Schottky-type equations emerge. The earliest explicit formulation of (a Gaussian version of) (9.20) that I know of is in Compton and Langmuir's 1930 review ([47], see p. 149)—although, as already noted, J.J. Thomson seems to have been the first (in 1903) to suggest this as a physical electron emission effect.

A further logical step occurred when Fowler and Nordheim [34] in effect suggested that the best way of deriving TE theory would be to apply Sommerfeld-type metal electron theory to the electron population *inside* the emitter, and thereby create a unified theoretical approach that would describe both TE and FE. This approach was implemented by Sommerfeld and Bethe in 1933 [41] (but with incorrect tunnelling theory), and is discussed further below.

Later, Jensen and others showed that small temperature and field dependent changes in the inner PE χ mean that, in principle, emission equations should contain a temperature and field dependent 'effective work function' rather than the true zero-temperature, zero-field work function (see [81]), but this small effect is usually neglected.

A possible mild cause of confusion is that the term 'effective work function' is also used to describe the Schottky-effect-reduced barrier of height (above the Fermi level) (ϕ–Δ_S), particularly in the TE and photo-cathode communities. A possible solution

might involve the introduction of a new thermodynamic term, such as 'classical (or "canonical") work of emission', for this reduced barrier height.

9.4 Validity Regimes

9.4.1 Transmission Regimes and Emission Current Density Regimes

In a smooth-surface model, the FN approach is implemented as follows. Each travelling-electron wave-state in the well can be labelled with a set {Q} of quantum numbers. Each state that involves motion towards a chosen part of the emitter surface makes a contribution z_Q to the current density incident onto that part of the surface from the inside, with the size of the contribution depending (inter alia) on the state's occupation probability. The probability that an incident electron escapes is given by the transmission probability D_Q for the state. Hence, the state makes a contribution $z_Q D_Q$ to the emission current density (ECD) J, and the ECD is found by summing over all relevant states, thus

$$J = \sum\nolimits_{\text{relevant}\{Q\}} z_Q D_Q. \tag{9.23}$$

As discussed below, it is easiest (in a Sommerfeld free-electron model) to carry out this summation by a double integral over energy.

In practice, as shown below, the transmission probability is a function only of the local barrier field (here denoted by F) and of forwards energy, and can be written $D(F,W)$, where W (as above) denotes forwards energy *measured relative to the well base*. Hence, (9.23) can be converted to the single integration

$$J(F,T) = \int j_n(F,T,W)\,\mathrm{d}W = \int N(T,W)\cdot D(F,W)\,\mathrm{d}W, \tag{9.24}$$

where $j_n(T,W)$ is the *emitted normal energy distribution* (*e-NED*), and $N(T,W)$ is a recently introduced parameter called the *incident normal energy distribution (i-NED)*. If the electron distribution can be treated as if in local thermodynamic equilibrium at thermodynamic temperature T, which is usually done, then a mathematically explicit expression is easily obtained (below) for $N(T,W)$.

There are other ways in which the double integration could be carried out, but formula (9.24) is most convenient here. Note that the parameter N used here differs slightly from the parameter 'N' used in older work and denoted here by N_{old}, which was called the 'supply function'. The relationship is: $N = eN_{\text{old}}$. Thus, the new N, like j_n, has the units A m^{-2} eV^{-1}.

In the thermodynamic equilibrium situation, if valid values could be obtained for D across the whole range of W, then integral (9.24) could always be carried out

numerically. However, ideally, one wants analytical formulae. Unfortunately, exact analytical formulae can be obtained only under certain mathematical conditions, which may vary with the barrier form, and are certainly different for the ET and SN barriers. This gives rise to the idea of *validity regimes*. A *transmission regime* is a region of $\{F,W\}$ parameter space where a given formula for $D(F,W)$ is 'adequately valid', and an *emission current density (ECD) regime* (also called an *emission regime*) is a region of $\{F,T\}$ parameter space where (for given ϕ) a given formula for the local ECD J is 'adequately valid'. There are close links between transmission regimes and ECD regimes, but these are not straightforwardly one-to-one in all cases.

Because of the general difficulties with smooth-surface conceptual models, 'adequately valid' above really means 'adequately self-consistent' or 'is an adequate mathematical approximation'. However, a general weakness of electron emission theory is that even with these reduced meanings of 'adequately valid', the criteria for what one means by the term 'adequately', and the boundaries of the validity regimes are not yet as well investigated and defined as they probably ought to be.

9.4.2 Regimes for the Exactly Triangular Barrier

Currently, the best defined illustration of validity regimes involves the transmission regimes for the ET barrier. Essentially, this is because Forbes and Deane [35] managed to derive an exact general analytical formula for the transmission probability D^{ET} for the ET barrier, and to show that this had three very different forms of mathematical approximation in different parts of the *transmission regime diagram* shown in Fig. 9.2. The nature of the mathematics means that the most sensible choice is to plot w against $F^{2/3}$, where w is forwards energy measured relative to the top of the triangular barrier (which in this case coincides with the local vacuum level). In this special case, it is easy to draw the boundaries where the approximate solution departs from the exact analytical solution by 10%. For this diagram, the inner PE χ has been taken as 15 eV.

The exact general formula for D^{ET} is

$$D^{\mathrm{ET}} = 1/[\tfrac{1}{2} + \tfrac{1}{4}\pi\omega(A^2 + B^2) + \tfrac{1}{4}\pi\omega^{-1}(A'^2 + B'^2)] \tag{9.25a}$$

where A and B are the values of the Airy functions Ai and Bi, and A' and B' the values of their derivatives, at the PE step at $z = 0$ (see [32] for exact definitions), and ω is a dimensionless parameter given by

$$\omega \equiv c_\kappa F^{-1/3} W^{1/2} \tag{9.25b}$$

where c_κ is a universal constant (see Appendix 9.1), called here the *triangular barrier constant*. Earlier, Jensen ([7], and references therein) had reached an equivalent result.

Expression (9.25) has approximations that, for $w \ll 0$, $w \sim 0$, $w \gg 0$, are very different in their mathematical forms. This leads to three main transmission regimes,

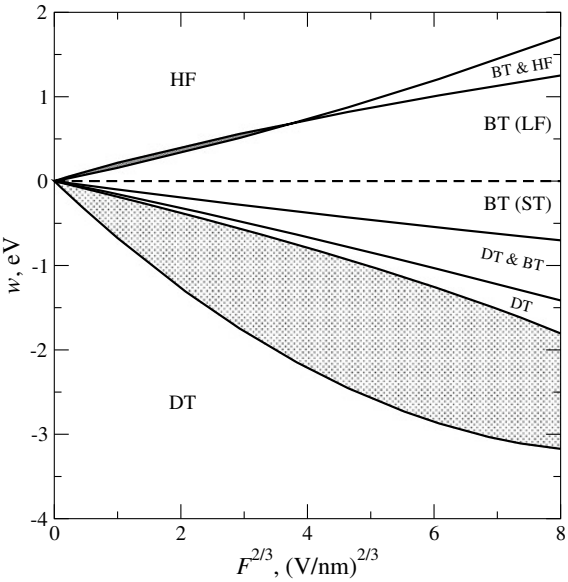


Fig. 9.2 Transmission regime diagram for the exactly triangular barrier ($\chi = 15$ eV, 10% error boundaries). Acronym meanings are: BT = barrier top; DT = deep tunnelling; HF = high flyover; LF = low flyover; ST = shallow tunnelling. The boundaries indicate a difference of 10% between the working formula for transmission probability D and the exact formula. In the two shaded regions, no simple working formula gets within 10% of the exact formula. Reprinted from [35, Fig. 5], with the permission of the Royal Society of London

each characterised by a useful working formula for D^{ET}. The main regimes are *deep tunnelling (DT)* ($w \ll 0$); *barrier-top emission (BT)* ($w \sim 0$); and *high flyover (HF)* ($w \gg 0$). For the $w \ll 0$ and $w \gg 0$ limits, the Airy functions and (9.25) have asymptotic expansions that each generate a sequence of approximate formulae of successively decreasing accuracy. The formulae used to define the transmission regimes are not the limiting formulae; rather they are the 'most useful working formulae'.

For the ET barrier, in the *deep tunnelling (DT) regime*, one can set the zero-field barrier height $H = -w$, and the working formula is the well known FN result

$$D_{\mathrm{DT}}^{\mathrm{ET}} \approx P^{\mathrm{FN}} \exp[-G^{\mathrm{ET}}] = \{4(HW)^{1/2}/\chi\} \exp[-bH^{3/2}/F], \tag{9.26}$$

where b is the second FN constant, as before, and the *FN tunnelling pre-factor* P^{FN} is given by $\{4(HW)^{1/2}/\chi\}$.

The *barrier-top (BT) regime* corresponds to $|w|$ small. This regime includes both shallow tunnelling (ST) and low flyover (LF), and a good working formula is [35]

$$D_{\mathrm{BT}}^{\mathrm{ET}} \approx 1/[\tfrac{1}{2} + c_0 F^{-1/3} W^{1/2} + c_\infty F^{+1/3} W^{-1/2} - c_1 F^{-1} W^{-1/2} w], \tag{9.27}$$

where c_0, c_∞ and c_1 are parameters, defined in [35], whose values are not of significance here. For *transmission at the barrier peak*, the result obtained by setting $w = 0$, $W = \chi$, in (9.27) is exact, and is

$$D_{\mathrm{BT}}^{\mathrm{ET}}(w=0) \;=\; 1/[\tfrac{1}{2} + c_0 F^{-1/3}\chi^{1/2} + c_\infty F^{+1/3}\chi^{-1/2}]. \tag{9.28}$$

Clearly, this barrier-peak transmission probability is a function of field, but goes to zero as the field goes to zero, as one would expect. It also goes to zero as $F \to \infty$, which was not anticipated.

In the *high flyover (HF) regime*, a suitable working formula is

$$D_{\mathrm{HF}}^{\mathrm{ET}} \;\approx\; \frac{4w^{1/2}W^{1/2}}{(W^{1/2}+w^{1/2})^2} \;=\; 1 - \left(\frac{W^{1/2}-w^{1/2}}{W^{1/2}+w^{1/2}}\right)^2. \tag{9.29}$$

This is also the standard result for transmission across a rectangular step of height $\chi = (W–w)$, as comparison with (9.19) shows. In the limit of *very* high flyover, where w and W become large, formula (9.29) tends to unity, as physically expected, and the regime limit is what one might reasonably call the *classical high-temperature emission (CHTE) sub-regime.*

Figure 9.2 shows that there is an area of the transmission regime diagram (enclosed within the DT regime) where none of the simple working formulae are adequately valid. Obviously, the details of this diagram depend somewhat on the value used for χ and on the numerical criterion used to define 'adequate validity' (10% error, here).

Older literature tends to discuss transmission in terms of two regimes: either tunnelling through the barrier or ballistic emission over it. What the above analysis shows is that, as far as the mathematics of the ET barrier is concerned, it is better to think in terms of *three* regimes rather than two (or possibly in terms of four rather than two regimes, if the CHTE sub-regime is treated as separate for pedagogical purposes). An important lesson is that, in the barrier-top regime, common theory applies both to (shallow) tunnelling and (low) flyover, but the sign of w is different in these two cases. In this BT regime, the value of the transmission probability increases smoothly as w changes from negative to positive, as one might expect physically. Another lesson from this diagram is that the BT regime is relatively narrow in its regime of validity, and that the width-in-energy of the regime diminishes with field.

Since (9.25) is the transmission-probability formula derived by FN, and then used by them to derive the original (zero-temperature) FE ECD equation, one can conclude that this ECD equation will not be adequately valid outside the boundaries of the DT regime. However, because the derivation of their ECD equation involves a Taylor expansion of $\ln\{\mathrm{D}_{\mathrm{DT}}^{\mathrm{ET}}\}$ about $H = \phi$, and then the neglect of terms higher than linear, there will be *additional* mathematical constraints on the adequacy of the resulting equation.

Because it is well known (or, at least, very firmly believed by nearly all FE theoreticians) that the ET barrier is not a good model for real surface barriers [82], finite-temperature versions of the original FN-type ECD equation have never been

formally developed, as far as the author is aware. (If they have, then knowledge has not survived.) However, it is known [83] that the temperature correction factor λ_T [$= p\pi/\sin(p\pi)$] developed below for the SN barrier, in fact, applies to all well-behaved barrier forms, and hence to the ET barrier (provided the parameter δ_F that appears in p is defined appropriately). Hence, if one were to attempt to construct an ECD regime diagram for a finite-temperature FN-type equation based on the ET barrier, then any mathematical constraints on the adequate validity of the expression $[(p\pi)/\sin(p\pi)]$ would also come into play when determining the boundaries of the relevant ECD regime, *in addition to* those indicated earlier (if the constraints are in fact different from those indicated earlier).

We can conclude that, in the framework of the 'SPME methodology' of using a planar smooth-surface conceptual model, the theoretical physical and mathematical questions that affect the 'adequate validity' of a particular ECD formula are as follows:

(1) Does a particular 'starting formula' for transmission probability D represent correct wave-mechanical transmission theory, as understood in a wider physics context? (This is not a significant problem for the ET barrier theory set out above, as we believe that the derivation of (9.25) represents correct wave-matching physics for a one-dimensional free-electron-metal model. However, for non-free-electron band-structures, and/or for other forms of the barrier, the situation is much more obscure.)
(2) Where relevant, has the starting formula for D been validly reduced to a simpler expression?
(3) If this 'simpler expression' needs to be further mathematically manipulated (e.g. by Taylor expansion), and some resulting terms are then neglected in order to provide a 'resulting expression', then is this mathematical approximation process adequately valid?
(4) If this 'resulting expression' then needs to be incorporated into an integral for which an analytical solution is known, then are the conditions for validity of the integration process adequately met? (In practice, questions (3) and (4) may relate to the same issue.)

A conclusion from this is that the 'failure analysis' of FE equations (and other electron emission equations) may be far from straightforward, and may deserve more attention than it has so far received.

Also, note that there are subtle differences in status between the Forbes-Deane transmission-regime diagram and the Murphy-Good ECD-regime diagram shown below as Fig. 9.3. The former shows regimes where the related formula is adequately valid as a description of the known correct physical analysis of the ET barrier model (to within 10%, at all points on each boundary). The latter simply shows regimes where certain mathematical approximations are considered adequately valid (with an error of between 15 and 40%, depending on the position in the boundary [61]).

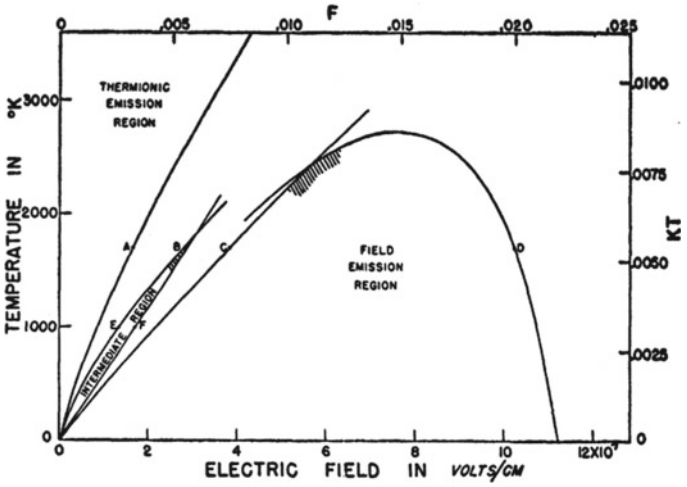


Fig. 9.3 Emission current density regime diagram for the Schottky-Nordheim barrier, as presented by Murphy and Good ([61Fig. 6]). See [61] for full discussion. Reprinted figure with permission from E.L. Murphy, R.H. Good, Jr, Phys. Rev. **102**, 1464 (1956). Copyright (1956) by the American Physical Society

9.4.3 Regimes for the Schottky-Nordheim Barrier—Qualitative Discussion

As just indicated, most FE theoreticians regard the Schottky-Nordheim barrier as the best simple model for the FE barrier at the surface of a metal emitter of moderate to large apex radius (say greater than 20 nm), for reasons recently set out in [82]. One can apply to the SN barrier the preceding Section's insights, but initially, this is done only at a qualitative level. Mathematical details form part of Sect. 9.5.

Confounding problems are the existence of two alternative top-level approaches and conflicting terminology. In this chapter I use 'thermal-field (TF)' as a general term to describe situations where both field and temperature play a role, but not as a term to describe any particular theoretical regime or sub-regime.

Although earlier discussions exist (e.g. [84]), the first 'reasonably modern' treatment of TF effects is the 1956 paper of Murphy and Good (MG) [61]. The MG treatment is the first substantial treatment after Burgess, Kroemer and Houston [85] had found, in 1953, a significant mathematical error in Nordheim's 1928 paper [56]. Unfortunately, the MG paper is heavily mathematical and written in a mixture of Gaussian equations and atomic units, and is now next to incomprehensible to many younger researchers. The Modinos textbook [2] has a simpler version of the MG treatment, though still with Gaussian equations. ISQ versions of MG's theory are given in [86] for the zero-temperature limit, in [87] for finite temperatures, and (slightly differently) in [10] and in [14].

Murphy and Good identified two main ECD regimes, which they labelled the 'field emission' and 'thermionic' regimes. Note that their definition of 'thermionic' does *not* correspond to the operating regime of the practical devices that electron source users call 'thermionic emitters', but also includes the devices now commercially called 'Schottky emitters' (my 'barrier-top emitters'). Jensen's chapter, and also the recent work of Kyritsakis and colleagues [88, 89], follow the MG two-regime approach and create bridging theory.

MG also identified what they called an 'intermediate' regime, but this is of very limited extent, and nowadays is not treated separately. MG's ECD regime diagram, showing all of their identified regimes, is reproduced as Fig. 9.3.

In the 1970s, Swanson and Bell [90] presented a more detailed classification of regimes/sub-regimes, and associated formulae for ECDs and energy distributions, although later [15, 91] they changed their nomenclature slightly.

The author's current classification of SN-barrier ECD regimes is based partly on the Swanson and Bell approach, partly on the findings of [35]. Nowadays, I identify three main ECD regimes. Using my currently preferred terminology, these are as follows:

(1) The *field electron emission regime* (also called the 'cold field electron emission (CFE)', the 'Fowler-Nordheim field electron emission (FNFE)' regime, or just the 'field emission regime'). In this regime, most electrons escape by deep tunnelling from electron states near the emitter Fermi level, and the local ECD is given by a FE equation. In MG's theory (which uses the simple-JWKB tunnelling formalism) the relevant FE equation contains a temperature correction factor λ_T given here by the linked equations

$$\lambda_T \equiv p\pi/\sin(p\pi), \tag{9.30a}$$

$$p \equiv k_B T/\delta_F^{SN}, \tag{9.30b}$$

$$1/\delta_F^{SN} \equiv [(\partial G^{SN}/\partial H)_F]|_{H=\phi} = t_F \cdot \tfrac{3}{2} b\phi^{1/2}/F, \tag{9.30c}$$

where p is the '*Swanson-Bell parameter p*' [90], δ_F^{SN} is a parameter I now call the *barrier-strength decay-width (or 'G-decay-width') for the SN barrier, taken at the Fermi level*, and t_F—here called the *G-decay-rate correction factor*—is the value, at $H = \phi$, of the correction factor (relative to the equivalent result for an ET barrier) in the first-order term of a Taylor expansion of $G^{SN}(H,F)$ about $H = \phi$ (with field F assumed constant.) The factor is given by $t_F = t_1(x = f)$, where $t_1(x)$ is a special mathematical function defined by (9.65b).

For numerics, it is convenient to write p in the form

$$p = C_p \cdot (t_F \phi^{1/2} T/F) \equiv (3bk_B/2) \cdot (t_F \phi^{1/2} T/F) \tag{9.30d}$$

$$C_p \cong 8.829607 \times 10^{-4}\ \mathrm{eV}^{-1/2}\ \mathrm{K}^{-1}\ \mathrm{V\ nm}^{-1}. \tag{9.30e}$$

The ECD regime boundaries are in principle influenced by (amongst other factors) the adequate validity of the derivation of λ_T. The Swanson and Bell view [90] was that use of the MG expression for λ_T is adequate in the range $p <$ ~0.7.

The FE regime has a zero-temperature theoretical limit at which $p = 0$, $\lambda_T = 1$. The original 1928 FE equation was, of course, derived as a formula applying in this limit. The high-T end of the FE regime has been called the 'T-F emission regime' (e.g. [90]), but this name has also been used in other contexts and is best avoided; the author's preferred name would be 'hot FE'.

(2) The *barrier-top electron emission (BTE) regime* (also called, in the case of the SN barrier, the 'extended Schottky regime' [15, 91], and before that the 'transition regime' [90]). In this regime, most electrons escape at a forwards energy level that is close to the top of the SN barrier. Significant numbers escape by tunnelling and significant numbers by wave-mechanical flyover. The local ECD for the BTE regime is usually given by an equation similar to the QM RS-type (9.20), but with the pre-factor λ_{TE} replaced by a pre-factor λ_{BTE} derived by using the Kemble tunnelling formalism (see Sect. 9.5.2). The pre-factor λ_{BTE} is given by the linked equations

$$\lambda_{\mathrm{BTE}} \approx \lambda_F \equiv q\pi/\sin(q\pi), \tag{9.31a}$$

$$q \equiv \delta_{\mathrm{b}}^{\mathrm{SN}}/k_{\mathrm{B}}T, \tag{9.31b}$$

$$1/\delta_{\mathrm{b}}^{\mathrm{SN}} \equiv [(\partial G^{\mathrm{SN}}/\partial H)_F]|_{H=H_{\mathrm{b}}} = \mathrm{t_b} \cdot \tfrac{3}{2} b H_{\mathrm{b}}^{1/2}/F = \mathrm{t_b} \cdot \tfrac{3}{2} c_{\mathrm{S}}^{1/2} F^{-3/4}, \tag{9.31c}$$

Here, q is the '*Swanson-Bell parameter q*' [90], H_{b} [$= \Delta_{\mathrm{S}} = c_{\mathrm{S}} F^{1/2}$] is the zero-field height of the transmission barrier experienced by an electron that crosses at the forwards energy level of the top of the barrier, and $\delta_{\mathrm{b}}^{\mathrm{SN}}$ is the G-decay-width for the SN barrier, taken for $H = H_{\mathrm{b}}$. The parameter $\mathrm{t_b}$ is the G-decay-rate correction factor for $H = H_{\mathrm{b}}$, and is given (from the theory of complete elliptic integrals, see below) as $\mathrm{t_b} = \mathrm{t_1}(x = 1) = (\pi\sqrt{2})/4 \cong 1.110721$.

For numerics, it is convenient to write q in the form

$$q = C_q F^{3/4}/T = [2/(3\mathrm{t_b} b c_{\mathrm{S}}^{1/2} k_{\mathrm{B}})]\, T^{-1} F^{3/4}, \tag{9.31d}$$

$$C_q \cong 930.8282\ \mathrm{K\ V}^{-3/4}\ \mathrm{nm}^{3/4}. \tag{9.31e}$$

This correction factor formula was first given by MG as part of their equation (9.33) (they use the symbol d rather than q), but the method of derivation indicated here is simpler, and open to generalisation.

Schottky electron sources, that operate in this BTE regime, are discussed in [11] and in Chap. 6, and in references therein.

On the high-q side, the original thinking of Swanson and Bell [90] was that this formula would be adequately valid for $q < \sim 0.7$. However, simulations based on the numerical solution of the Schrödinger equation suggest a lower limit of adequate validity, maybe around $q < \sim 0.4$ or less [15, 91]. Precise reasons for this lower limit of validity have not yet been established.

On the low-q side, it is clear that as $q \rightarrow 0$, then $\lambda_F \rightarrow 1$, and the ECD formula then goes over into the *classical* RS-type equation, which describes the classical high-temperature emission (CHTE) *sub-regime*, but not the main QMTE regime (except as an approximation). More generally, the precise limits of validity of formula (9.31a–9.31e) have not yet been clearly established.

(3) The third main regime is the *quantum-mechanical thermal electron emission* (QMTE) regime, already discussed. In this regime, almost all electrons escape by wave-mechanical flyover over the barrier, which may be lowered by the Schottky effect, and encounter wave-mechanical reflection effects in the whole surface + barrier potential structure. As already indicated, this has the CHTE sub-regime as the very-high-temperature limit. Both the main regime and the sub-regime in principle have 'pure thermal emission limits' as the field is reduced towards zero, but (as already indicated) related practical current densities may be limited by space-charge effects associated with patch fields.

Some general comments are now needed. Although the BTE and QMTE regimes have ECD equations of the same general mathematical form, the factors λ_{BTE} and λ_{TE} have different physical origins, and approximations for them have (or would have) different detailed mathematical forms. This arises because λ_{BTE} relates to wave-mechanical 'tunnelling and reflection' effects associated with the top part of the SN barrier, but λ_{TE} relates primarily to wave-mechanical reflection effects primarily associated with the sharp change in EPE gradient when the rising image PE (or, more generally, the surface EPE) takes over from the constant EPE inside the well. In reality, there may be interactions between these two effects, and these interactions might lead to a modern model for the anomalous Schottky effect [71, see p. 33]. (This might be seen as an improved version of Guth and Mullin's theory [84].) In the longer term, one might hope to have an integrated theory of λ_{TE} and λ_{BTE} that either separates into two or more regimes, or alternatively shows that the interactions between the two different sources of reflection effect must always be taken into account, certainly at low fields.

A different form of bridging theory is presented in Jensen's chapter, which is based on earlier work by himself and Cahay [92, 93], to which there have been some recent improvements [94]. Jensen has developed formulae that (in effect) have (9.30a–9.30e) and (9.31a–9.31e) as the low-temperature and high-temperature limits, and (for given ϕ-value) have adequate validity in the region of $\{F,T\}$ parameter space outside the regimes of adequate validity for the expressions for λ_T and λ_F. Such a formula is particularly useful in simulations when making an estimate of the total emission current from a device where the field varies strongly across the surface

and the temperature is high enough for the emission to be a mixture of field and thermally induced emission. As with formula (9.30a–9.30e), Jensen's formulae deal only approximately with the QMTE regime.

A slightly different approach to the development of bridging theory has been taken by Kyritsakis and colleagues, who have also extended the theory to cover charged, earthed spherical emitters—provided that the emitter apex radius is not so small that significant quantum confinement effects (e.g. [95, 96]) come in play. Readers are referred to the relevant Kyritsakis papers for details [88, 89].

A further point about regimes is that, at extremely high fields, where the top of the SN barrier is pulled down close to or below the emitter Fermi level, there are (from the point of view of emission theory) one or more additional ECD regimes. An expression $\frac{1}{2}z_S K_{F0}^2$, where z_S is the Sommerfeld electron supply constant (see below), and K_{F0} is the zero-temperature Fermi *energy*, exists for the limiting ECD in the hypothetical situation where all electrons approaching the surface escape. In reality, most emitters would heat up, melt and/or explode long before this limit is reached. These 'explosive emission' phenomena are covered briefly by Fursey [6] and in more detail by Mesyats [3], in their textbooks, and (for carbon) by Fursey in Chap. 11. They will not be discussed here. It further seems that sudden heating, even in the absence of applied voltages, can also cause explosive phenomena [97].

9.5 Emission Theory—Mathematical Background

9.5.1 Relevant Elements of Statistical Mechanics

Free-Electron Theory and the Sommerfeld Electron Supply Constant

To implement the summation in (9.23), or the integral in (9.24), basic understanding is needed of both free-electron statistical mechanics and wave-mechanical transmission theory. The next three sections cover this. Many treatments of free-electron theory still use the early 1930s starting point that the density of electron states in phase-space is constant, and proceed via a three-dimensional integration with respect to velocity, momentum or wave-number. In reality, there is an alternative starting point that is simpler, exactly equivalent physically, better for discussion of emission phenomena, and better for non-physicists. This is based directly on the use of energies, and is as follows.

Consider a mathematical plane in space inside a Sommerfeld well of large extent, and let the electron energies parallel and normal to the plane be E_p and E_n, with the total energy being given by $E_t = E_n + E_p$. In principle, E_n, E_t and the well base E_c can all be measured relative to *any* one energy reference level (e.g. the laboratory 'Ground'), but it is simplest to initially treat E_n and E_t as measured relative to the well base, and hence purely kinetic in nature. Any two of these three energies can be used to create a formal *energy-space* (call the axes '1' and '2'). The *supply density* $z(E_1,E_2)$ is defined as the current density crossing the mathematical plane in real

space, per unit area of energy-space, near (E_1,E_2). The *maximum supply density* (for the element) is achieved if all states in the relevant energy element are occupied with probability 1. There is a general statistical mechanical result that, for a free-electron model (of a metal), the maximum supply density is the same for all elements in all of the three alternative energy-spaces that can be created, and is equal to the *Sommerfeld electron supply constant* (or 'Sommerfeld supply density') z_S given by

$$z_S \equiv 4\pi e m_e/h_P^3 \cong 1.618311 \times 10^{14}\,\mathrm{A\,m^{-2}\,eV^{-2}} \tag{9.32a}$$

where h_P is Planck's constant and m_e is the electron mass in free space. The presence of this constant as part of the elementary RS-type equation [since $A_{R0} = z_S(k_B T)^2$] is particularly clear in [42]. However, some authors prefer to write it in the alternative form

$$z_S = e m_e/2\pi\hbar^3 \tag{9.32b}$$

Let $f_{FD}(E_t)$ be the Fermi-Dirac distribution function. It follows that, if states in the relevant energy elements are occupied in accordance with Fermi-Dirac statistics, then the contribution d^2Z to the total current density Z crossing the plane, made by states in the energy-space element concerned, can be written in any of the alternative forms

$$d^2Z = z_S f_{FD} dE_p dE_n = z_S f_{FD} dE_n dE_t = z_S f_{FD} dE_p dE_t. \tag{9.33}$$

This result is proved and discussed by the present author in [98], though I would not claim to have done anything other than provide an alternative proof and a more general statement of a result already known (e.g. §11.2 in [99]), and give a name to the universal constant involved. The point is that (for a free-electron Sommerfeld-type model) this is a proven statistical mechanical result that can be used as a starting point for further development. It is not necessary to repeatedly reprove some weaker specific version of it, starting from early 1930s formulations that many experimentalists may not fully understand—which is what tends to happen at present.

Derivation of the Incident Normal Energy Distribution

To apply (9.33) to an emission problem, one identifies the 'mathematical plane' with a planar or 'nearly planar' emitter surface, identifies d^2Z as an element of incident electron-current density approaching the surface from the inside, notes that all the states in the element will have the same transmission probability D (which will be a function only of barrier field and normal energy), and concludes that this element makes a contribution d^2J to the local *emission* current density J given by $d^2J = D(F,E_n)d^2Z$. The ECD J is then obtained by a double integral with respect to energy. In principle, there are six ways of carrying this out, and different methods are best for different purposes. For discussion of ECD regimes, the best method is to integrate first over E_p, then over E_n. This is sometimes called 'integrating via the NED', where NED stands for 'normal energy distribution'.

In the particular approach where E_n is measured *relative to the well base*, it is convenient to replace the symbol E_n with the symbol W, and to revert to using the term 'forwards energy'. (Jensen uses the alternative approach of replacing E_n by the symbol 'E'.) However, note that some older papers that use the symbol W for normal/forwards energy (in particular, MG) use the local vacuum level as their energy reference zero.

The double integral over E_p and W decomposes into the two single integrals

$$J(F,T) = \int_0^\infty j_n(F,T,W)\mathrm{d}W = \int_0^\infty N(T,W)\cdot D(F,W)\cdot \mathrm{d}W \tag{9.34}$$

$$N(T,W) = z_S \int_0^\infty f_{FD}(W,E_p)\,\mathrm{d}E_p \tag{9.35}$$

where N is the incident normal energy distribution, as before. Equation (9.34) here is (9.24) as stated earlier, and can be called the *forwards energy integral*. Substituting an explicit expression for the Fermi-Dirac function into (9.35) yields

$$\begin{aligned} N(W,T) &= z_S \int_0^\infty \left\{1+\exp[(W+E_p-W_F)/k_BT]\right\}^{-1} dE_p \\ &= -z_S k_B T\Big[\ln\{1+\exp[-(W+E_p-W_F)/k_BT]\Big]_0^\infty \\ &= z_S k_B T \ln\{1+\exp[-(W-W_F)/k_BT]\}, \end{aligned} \tag{9.36}$$

where k_B is the Boltzmann constant, and W_F denotes the Fermi level measured relative to the well base at total energy E_c, i.e. $W_F = E_F – E_c$ (see Fig. 9.1).

When $(W–W_F)$ is sufficiently greater than k_BT, then the final exponent in (9.36) becomes small, and the approximation $\ln(1+x) \approx x$ leads to the approximate expression for the i-NED

$$N(W,T) \approx N^{aprx}(W,T) = z_S k_B T \exp[-(W-W_F)/k_BT]. \tag{9.37}$$

9.5.2 *Elements of Barrier Transmission Theory*

The Status of Barrier Transmission Theory

Surprisingly, despite enormous efforts over many years, barrier transmission theory is still an underdeveloped branch of quantum mechanics, involving some very awkward physics and mathematics that is still far from being completely understood in full detail. The present situation (within the context of SPME methodology) is as follows:

For the ET barrier, an exact analytical solution (covering both tunnelling and flyover) is known for the free-electron *metal* case, as described above.

For the SN barrier, and for other field-influenced barriers relating to bulk emitters with *planar* surfaces, no exact analytical solutions are known in terms of the existing functions of mathematical physics, but approximate analytical solutions based on *semi-classical methods* (also called 'quasi-classical methods') have—until recently—been thought sufficient for nearly all technological purposes. These include 'JWKB-like' and 'phase-integral' methods, but only low-order forms of these methods are currently in use. Higher-order phase-integral formulas are given in [100]. For non-planar emitters, especially emitters of small apex radius, new apparent difficulties in developing semi-classical transmission theory have recently emerged [101], and are still subject to research.

Certainly for planar emitters, good numerical methods in principle exist for determining FE transmission probabilities, by direct numerical solution of the Schrödinger equation (e.g. [14, 102]) However, the most popular methodologies still rely on semi-classical approaches, as described next.

The Main Semi-Classical Formalisms

For the transmission probability D, the most general semi-classical formalism normally quoted in FE tunnelling contexts is the *Fröman and Fröman formalism* [103, 104]

$$D \approx D_{\mathrm{FF}} = \frac{P\exp[-G]}{1+P\exp[-G]} = \frac{1}{1+P^{-1}\exp[G]} \tag{9.38}$$

where P is a parameter, called the *transmission pre-factor*, that is thought to be of *order* unity in all normal circumstances. Though Mayer has made calculations that show this, and has provided a formula [102] that fits the result of numerical calculations, normal practice is still to use the *Kemble formalism* in which P is set equal to unity and hence D is given by

$$D \approx D_{\mathrm{K}} = \frac{\exp[-G]}{1+\exp[-G]} = \frac{1}{1+\exp[G]}. \tag{9.39}$$

This formula was first derived long ago by Kemble [105, 106], by a JWKB-type method. It was rederived by Miller and Good [107] in a slightly different way that some authors regard as technically not a JWKB-type method but an alternative semi-classical method (see [108]).

In the case of *deep tunnelling*, where the barrier strength G is sufficiently large ($G > 2.3$, for 10% error), (9.38) reduces to the so-called *Landau and Lifshitz formalism* [59, see (50.12)] :

$$D \approx D_{\mathrm{LL}} = P\exp[-G] \tag{9.40}$$

and (9.39) reduces to the *simple (or 'first-order') JWKB formalism*

$$D \approx D_{\mathrm{JWKB}} = \exp[-G] \tag{9.41}$$

This form can also be obtained by straightforward JWKB-type derivation methods, e.g. [59].

In all cases, for tunnelling through the SN barrier, the barrier strength G^{SN} is given by

$$G^{\mathrm{SN}} = \mathrm{v}(f_H) \cdot bH^{3/2}/F. \tag{9.42}$$

In the situation where the barrier has just vanished, $F = F_{\mathrm{R}}, f_H = 1$, $\mathrm{v}(f_H) = \mathrm{v}(1) = 0$, $G^{\mathrm{SN}} = 1$, and—in the Kemble formalism, from (9.39)—it is found that $D \approx 0.5$.

Barrier Strength for Transmission by Flyover above the Barrier

For the FE regime, simple (first-order) JWKB theory, as specified by (9.41), is sufficient. But when considering the other main regimes it is necessary to have formulae that give values of transmission probability D for wave-mechanical flyover over the top of the barrier.

In these circumstances, the motive energy is everywhere negative, but the equation $M^{\mathrm{SN}}(z) = 0$ does have mathematical solutions, although these are complex numbers, as illustrated in Fig. 9.4. There is a pole at the origin, and two complex roots of the form $(H/2eF)\{1 \pm i(f_H - 1)^{1/2}\}$, where $\mathrm{i} = +\sqrt{(-1)}$. Miller and Good [107], and Murphy and Good [61], assume that an appropriate procedure to evaluate G in these circumstances is to perform complex integration along the contour from the

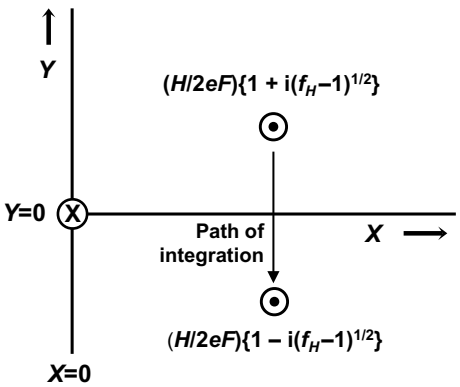


Fig. 9.4 To illustrate the procedure apparently used by Murphy and Good [61] to evaluate the barrier strength integral for a SN barrier in the case of flyover. In the complex-distance space associated with the z-direction, the expression $M^{\mathrm{SN}}(z) = 0$ has two complex zeros, and a pole at the origin. The path of integration in complex space is indicated

upper root to the lower root. The author has never seen a formal mathematical proof that this procedure is strictly valid, but it looks plausible enough (certainly as an approximation), at least for small values of $(f_H-1)^{1/2}$.

This integration again results in (9.42), but this now operates in the range $f_H > 1$. Since v(x) is a special solution of the Gauss Hypergeometric Differential Equation and is known to be well-behaved at $x = 1$, we can deduce that v(x) is well defined over the range $0 \leq x < \infty$; hence, *in principle,* it is straightforward *mathematically* to obtain values of v(x), and hence v(f_H), for the range $1 \leq f_H < \infty$. In this range the values of v(f_H) and G^{SN} are negative, and G^{SN} diminishes from 0 towards $-\infty$, causing D to increase from 0.5 towards 1.

In practice, due to the pole at the origin, this procedure for obtaining an expression for G^{SN} becomes invalid if the two complex roots are too far separated. MG take the validity condition to be $y \leq \sqrt{2}$, i.e. $f_H \leq 2$. This defines a forwards energy level $W^{\mathrm{lim}}(\phi,F)$ above which their theory for evaluating D becomes invalid, and they recommend the *physical* approximation $D = 1$ for higher forwards energies. As far as the author can establish, for the SN barrier, there are no valid, simple, semi-classical mathematical methods available for evaluating D-values for $W > W^{\mathrm{lim}}$.

In practice, MG then make mathematical approximations in order to create integrals that can be evaluated analytically, as discussed below (see [61] for details).

Evaluation of the Forwards Energy Integral

We now return to the evaluation of the forwards-energy integral (9.34). For each of the two components N and D in the integral, there is a 'full expression' and an 'approximated expression'. As shown in Table 9.2, in three of the resulting four cases an analytical solution exists. The combinations of taking one 'full' expression and one 'approximate expression' lead to expressions for λ_T and λ_F, respectively, as given by (9.30a–9.30e) and (9.31a–9.31e), both first derived (though not in these modern forms) by MG. Both these expressions have regimes of validity in $\{F,T\}$ parameter space, as partly shown in Fig. 9.3.

By contrast, the Jensen *general thermal-field (GTF) formalism*, described in detail in Chap. 8, is based on the integral that uses both 'full' expressions. The main aim is to derive a 'bridging formula' that will have adequate validity (for a given ϕ-value) in the region of $\{F,T\}$ parameter space outside the regimes of validity for the expressions for λ_T and λ_F.

Table 9.2 Results of evaluating the forwards-energy integral, (9.34), using different levels of approximation for the two terms in the integral

D-formula	N-formula	Analytical?	Result	Discussion
Approximate	Approximate	Yes	Not normally used	–
Approximate	Full	Yes	λ_T	Here (9.30a)
Full	Approximate	Yes	λ_F	Here (9.31a)
Full	Full	No	Jensen's GTF formula	See Chap. 8

9.6 Emission Theory

9.6.1 Field Electron Emission Equations

The next two Sections record further proposals for improving the presentation of electron emission theory. FE literature contains many different equations asserted to be '*the* Fowler-Nordheim equation'. In reality, this name applies to a very large family of related approximate equations, which I now prefer to call *field electron emission (FE) equations*. This Section first discusses how to classify them.

A *FE equation* (previously called a 'Fowler-Nordheim-type (FN-type) equation') is any FE-related equation with the mathematical form

$$Y = C_{YX} X^2 \exp[-B_X/X] \tag{9.43}$$

where X is any FE independent variable (usually a field, voltage or scaled field), and Y is any FE dependent variable (usually a current or current density). B_X and C_{YX} depend on the choices of X, Y and barrier form, and will often exhibit weak or moderate functional dependencies on the chosen variables and sometimes on others (e.g. apex radius).

The choices of X and Y determine the equation's *form*. *Core forms* give the local ECD J in terms of the local work function ϕ and local barrier field F and/or scaled field f. The simplest core form (which relates to the ET barrier) is the so-called *elementary FE equation*

$$J^{\text{el}} = a\phi^{-1}F^2 \exp[-b\phi^{3/2}/F], \tag{9.44}$$

where a and b are the FN constants (see Appendix 9.1).

For an arbitrary but well-behaved general barrier (GB), with barrier form correction factor ν^{GB}, (ν = 'nu'), one can define a *kernel current density* J_{k}^{GB} by

$$J_{\text{k}}^{\text{GB}} = a\phi^{-1}F^2 \exp[-\nu_{\text{F}}^{\text{GB}} b\phi^{3/2}/F] \tag{9.45}$$

where the subscript 'F' on $\nu_{\text{F}}^{\text{GB}}$ indicates that the correction factor applies to a barrier of zero-field height ϕ. A merit of defining this mathematical quantity J_{k}^{GB} separately is that, for given choices of barrier form and of ϕ and F, the value of J_{k}^{GB} can be calculated *exactly*; further, in the case of the SN barrier, reasonably accurate values of J_{k}^{SN} can be deduced from a FN plot if the emission situation is orthodox (see [20, 109]). The same is expected to be true for the newly introduced Murphy-Good plot [110].

The GB-model *local ECD* J^{GB} is then formally related to the corresponding kernel current density by

$$J^{GB} = \lambda^{GB} J_k^{GB}, \tag{9.46}$$

where the *pre-exponential correction factor* λ^{GB} *formally* takes into account all corrections other than those directly associated with the simple treatment of barrier form. These corrections include the use of atomic-level wave-functions and non-free-electron-like band-structure; better quantum mechanics of the transmission process; more accurate summation over states; temperature effects on the electron-energy distribution; field and temperature effects on the 'apparent zero-field work function'; and any factors currently unrecognised.

Equations (9.45) and (9.46) each contain a correction factor, and the *complexity level* of a FN-type-equation core form can be classified by specifying these two correction factors. For planar-surface emitter models, Table 9.3 shows the core forms (involving F) of most interest, historically and currently. These are classified, first, according to the assumed nature of the barrier, second (roughly) according to the sophistication of the assumptions made about the pre-exponential correction factor.

It needs emphasising that this is not a complete list, even for planar surface models. Further, in some ways, it is only an overview. For example, FE literature contains about 20 different mathematical approximations for the special mathematical function v, and there are various treatments of different individual components that go into the general-purpose pre-exponential correction factors.

Table 9.3 Names and complexity levels for field electron emission (FE) equations applying to emitter models that assume a smooth planar surface or (with "GB" models only) an atomically flat planar surface

Name of FE equation	Date	References	$\lambda_C^{GB} \rightarrow$	Barrier form	$v_F^{GB} \rightarrow$	Note
Elementary	1999	[111]	1	ET	1	a
Original	1928	[34]	P^{FN}	ET	1	
Fowler-1936	1936	[112]	4	ET	1	
Extended elementary	2015	[20]	λ_C^{ET}	ET	1	
Dyke-Dolan	1956	[113]	1	SN	v_F	
Murphy-Good (zero-T)	1956	[61]	t_F^{-2}	SN	v_F	b
Murphy-Good (finite-T)	1956	[61]	$\lambda_T \cdot t_F^{-2}$	SN	v_F	
Orthodox	2013	[109]	λ_C^{SN0}	SN	v_F	c
Extended Murphy-Good	2015	[20]	λ_C^{SN}	SN	v_F	d
"Barrier-effects-only"	2013	[114]	λ_C^{GB0}	GB	v_F^{GB}	c
General	1999	[111]	λ_C^{GB}	GB	v_F^{GB}	

[a]Earlier imprecise versions exist, but the first reasonably clear statement seems to be in 1999
[b]Previously called the "Standard FN-type equation"
[c]The superscript "0" in column 4 indicates that the pre-factor is to be treated mathematically as constant
[d]Previously called the "New-standard FN-type equation"

9.6.2 Scaled Form for the SN-Barrier Kernel Current Density

For the SN barrier, the kernel current density is

$$J_{\mathrm{k}}^{\mathrm{SN}} = a\phi^{-1}F^2 \exp[-\mathrm{v}(f) \cdot b\phi^{3/2}/F], \tag{9.47}$$

where f is the scaled field for a SN barrier of zero-field height ϕ. On defining work-function-related parameters $\eta(\phi)$ and $\theta(\phi)$ by

$$\eta(\phi) \equiv b\phi^{3/2}/F_{\mathrm{R}} = bc_{\mathrm{S}}^2\phi^{-1/2} \cong 9.836238\ (\mathrm{eV}/\phi)^{1/2}, \tag{9.48}$$

$$\theta(\phi) \equiv a\phi^{-1}F_{\mathrm{R}}^2 = ac_{\mathrm{S}}^{-4}\phi^3 \cong (7.433979 \times 10^{11}\ \mathrm{A/m^2})(\phi/\mathrm{eV})^3\,, \tag{9.49}$$

the kernel current density $J_{\mathrm{k}}^{\mathrm{SN}}$ for the SN barrier can then be written exactly *in scaled form* as

$$J_{\mathrm{k}}^{\mathrm{SN}} = \theta f^2 \exp[-\mathrm{v}(f) \cdot \eta/f]. \tag{9.50}$$

Values of $\eta(\phi)$ and $\theta(\phi)$ are shown in Table 2 in [109] for a range of work-function values. For illustration, when $\phi = 4.50$ eV, then $\eta \approx 4.637$ and $\theta \approx 6.77 \times 10^{13}$ A/m^2. For algebraic simplicity in what follows, the explicit dependencies on ϕ are not shown.

Merits of the form (9.50) are that: (a) for a given ϕ-value (and hence given η-value), only a single independent variable (f) appears in the right-hand side; and (b) a good approximation for v(f) is known (see below). Hence, good approximate values for $J_{\mathrm{k}}^{\mathrm{SN}}$ are easy to obtain. Also, in some contexts, this scaled equation is easier to use than non-scaled equations, particularly since (in orthodox emission situations) f can also be treated as a scaled voltage.

By developing various different expansions for v(f), (9.50) can be written in various different *equation formats* [21].

9.6.3 The Principal Field Emission Special Mathematical Function v(x)

The Pure Mathematics of v(x)—Basic Formulae

As part of the tidy-up noted earlier, the author has advocated changes in the treatment of the function 'v'. One change is to make a clearer conceptual distinction between (a) the pure mathematical aspects of the special mathematical function v(x), and (b) its use in modelling barrier transmission. This subsection and the next four briefly summarise the pure mathematics of v(x); the following subsections consider its use in modelling. In line with this thinking, I now prefer to call v(x) the 'principal field emission special mathematical function'. The function v(f), where x has been

replaced by the modelling variable f ('the scaled field, for a barrier of height ϕ '), has previously been called the 'principal Schottky-Nordheim barrier function'. However, the function v(x) is useful in the mathematical analysis of various tunnelling barriers different from the SN barrier, in the contexts of both field electron and field ion emission.

The function v(x) has many alternative/equivalent definitions, but the most fundamental is as a very special solution of the Gauss Hypergeometric Differential Equation (HDE):

$$x(1-x)\mathrm{d}^2W/\mathrm{d}x^2 + \{\, c_\mathrm{G} - (a_\mathrm{G} + b_\mathrm{G} + 1)\mathrm{d}^2W/\mathrm{d}x - a_\mathrm{G}b_\mathrm{G}W \;=\; 0, \tag{9.51}$$

where W here is an abstract mathematical variable, and a_G, b_G and c_G are constants. Taking $a_\mathrm{G} = -3/4$, $b_\mathrm{G} = -1/4$, $c_\mathrm{G} = 0$, reduces (9.51) to the defining equation identified by Forbes and Deane [86], namely

$$x(1-x)\mathrm{d}^2W/\mathrm{d}x^2 \;=\; (3/16)\, W. \tag{9.52}$$

This is a special mathematical equation, like Airy's and Bessel's equations, but much more obscure.

Note that, if one recasts this equation to use the variable $y = +x^{1/2}$, the equation becomes

$$y(1-y^2)\frac{\mathrm{d}^2W}{\mathrm{d}y^2} - (1-y^2)\frac{\mathrm{d}W}{\mathrm{d}y} - \frac{3}{4}yW = 0. \tag{9.53}$$

This form is more complicated than (9.52), and less obviously a special case of the Gauss HDE.

The principal SN barrier function v(x) is a particular solution of (9.52) satisfying the unusual boundary conditions (derived in [57]):

$$\mathrm{v}(0) = 1; \quad \lim_{x\to 0}\{\mathrm{dv}/\mathrm{d}x - (3/16)\ln x\} \;=\; -(9/8)\ln 2. \tag{9.54}$$

An exact analytical solution to (9.52) is known [57], but is not given here because it involves obscure mathematical functions usually known only to mathematicians. This solution gives rise to the exact series expansion [57]

$$\mathrm{v}(x) = 1 - \left(\tfrac{9}{8}\ln 2 + \tfrac{3}{16}\right)x - \left(\tfrac{27}{256}\ln 2 - \tfrac{51}{1024}\right)x^2 - \left(\tfrac{315}{8192}\ln 2 - \tfrac{177}{8192}\right)x^3 + \mathrm{O}(x^4) + x\ln x\left[\tfrac{3}{16} + \tfrac{9}{512}x + \tfrac{105}{16384}x^2 + \mathrm{O}(x^3)\right]. \tag{9.55}$$

Recurrence formulae for determining higher-order terms are given in [57].

Note that this exact series expression contains no terms involving half-integral powers of x. This result, and the obvious superiority of (9.52) over the mathematically equivalent (9.53), shows that the Gauss variable x (rather than the Nordheim parameter y [$=+\sqrt{x}$]) is the natural mathematical variable to use. Also note that v(x) is mathematically unusual, in that its series definition needs TWO infinite power

series: this finding has been key to recent progress in FE theory, and will be key to further progress in FE data interpretation.

In the present chapter, the symbol x is used to denote the mathematical argument of v, whereas until recently I have used the symbol l'. The reason for the change is as follows. Until recently, the usual definition of v used complete elliptic integrals, as discussed below. In the theory of such integrals, one standard notation uses the symbols k, k', m and m'. The parameter l' was seen as part of an extended group of these parameters, related to m by $l' = [(1–m)/(1 + m)]^2$. With the discovery that, mathematically, a more fundamental definition of v is as a particular solution of the Gauss HDE, it makes better mathematical sense to use—as the argument of v—the independent variable in the Gauss HDE (here x), and call it the 'Gauss variable'. This is just a symbol change. All formulae previously expressed using l' are equally valid when expressed using x. The new notation is also less cumbersome.

The Pure Mathematics of v(x)—The 2006 Approximate Formula

Before the exact series expression was known, a good simple approximation had been found [115], namely

$$\mathrm{v}(x) \approx 1 - x + \frac{1}{6} x \ln x. \tag{9.56}$$

Over the range $0 \leq x \leq 1$, where v(x) varies from 1 to 0, this expression has absolute error everywhere less than 0.0024 and relative error everywhere less than 0.33% [86]. This performance is more than good enough for most technological purposes. Mathematically, v(x) is defined and well-behaved over the whole range $0 \leq x < \infty$, but the accuracy of formula (9.56) deteriorates rapidly above $x = 1$, and it should not be used above $x = 4$.

Formula (9.56) is not a truncated version of the exact series expansion, but was designed independently to have exactly correct behaviour at $x = 0$ and $x = 1$, namely $\mathrm{v}(0) = 1$ and $\mathrm{v}(1) = 0$. This could be achieved with any expression of the form $(1–x + qx\ln x)$. This form (involving an $x\ln x$ term) was partly suggested by numerical experiments using the computer algebra package MAPLE™, carried out with Dr. J.H.B. Deane. However, the choice $q = 1/6$ was made [115] using a spreadsheet [116], and was the 'best simple algebraic expression that gives good accuracy over the range $0 \leq x \leq 1$'. The author originally thought there might be some precise underlying mathematical reason for the appearance of the simple factor (1/6), but we now know there is not.

Exact formulae for v were first stated correctly by Burgess et al. [85] in 1953. Their formulae involve complete elliptic integrals, and can be evaluated by computer algebra to give very high-precision values. As shown in [86], slightly better fits to these 'exact' values can be obtained by choosing a numerical value for q by means of error minimization over the range $0 \leq x \leq 1$. However, since there are at least five different ways in which minimization can be carried out, all leading to slightly different values of q, it seemed best to opt for algebraic simplicity. If needed, a more accurate 'high precision' (numerical) formula for v(x) is given in Appendix 9.2.

As already noted, there are about 20 different approximate formulae for v in FE literature. When the whole range $0 \leq x \leq 1$ is considered, formula (9.56) is the 'best simple approximation' amongst those tested [117]. Presumably, this is largely because it is designed to be exactly correct at $x = 0$ and $x = 1$, and also mimics the form of the exact series expression, by including an $x\ln x$ term.

It is possible to get better accuracy over limited parts of the range $0 \leq x \leq 1$ with formulae of similar complexity [117], but such formulae often behave poorly or very poorly near $x = 1$. This makes them unsuitable for discussions of thermal-field phenomena, because transmission by flyover corresponds to the range $x > 1$. By contrast, formula (9.56) behaves quite well for forwards energy levels close above the top of the barrier. It is designed to give v(1) correctly, but also performs well in respect of the value of u(1), where $u(x) = -dv/dx$. From the mathematics of complete elliptic integrals (see below), the exact value $u^{exact}(1) = 3\pi/8\sqrt{2} \approx 0.83304$; from (9.56), $u^{aprx}(1) = (5/6) \approx 0.83333$; that is, agreement at $x = 1$ is within about 0.035%.

Another way of looking at this is as follows. By equating the formulae for $u^{exact}(1)$ and $u^{aprx}(1)$, we obtain a new numerical approximation for π, namely $\pi \approx (20\sqrt{2})/9 \approx 3.14270$. The exact value, to 6 significant figures, is $\pi \approx 3.14159$. The error, again, is about 0.035%.

The Pure Mathematics of v(x)—Definitions Using Complete Elliptic Integrals

The function $v(x)$ has multiple alternative (equivalent) definitions, obtained from each other by mathematical transformations. In particular, it has simple integral definitions, and also definitions in terms of complete elliptic integrals. The complete elliptic integral of the first kind (K) can itself be defined in alternative ways. In terms of the *elliptic modulus* k, K is given by

$$K = K[k] = \int_0^{\pi/2} (1 - k^2 \sin^2\theta)^{-1/2}\, d\theta \tag{9.57}$$

whereas in terms of the *elliptic parameter* m [$\equiv k^2$], K is given by the equivalent formula

$$K = K[m] = \int_0^{\pi/2} (1 - m \sin^2\theta)^{-1/2}\, d\theta. \tag{9.58}$$

A further complication has been that an older notation also used to exist, namely $K[k^2]$, which left it unclear to the non-expert what symbols such as '$K[0.4]$' really meant (it means put $k^2 = 0.4$, not put $k = 0.4$). It seems to have been a misinterpretation of this general kind that caused the error in the Nordheim (1928) paper [56] that persisted in FE theory until 1953.

The author's view has been (e.g. [116]) that use of the elliptic parameter m is less error-prone than the use of the elliptic modulus k, and consequently I use (9.56) for

K and the related definition for the complete elliptic integral of the second kind E:

$$\mathrm{E} = \mathrm{E}[m] = \int_0^{\pi/2} (1 - m \sin^2 \theta)^{+1/2} \, \mathrm{d}\theta \tag{9.59}$$

The next complication is that the part of FE literature that is mathematically correct provides three alternative (mathematically equivalent) choices for a formula for v(x) in terms of complete elliptic integrals. The most useful of these (in the updated notation used here) is considered to be the definition given by the linked equations

$$\mathrm{v}(x) \; = \; (1 + x^{1/2})^{1/2} \{\mathrm{E}[m_1(x)] - \mathrm{K}[m_1(x)]\}, \tag{9.60a}$$

$$m_1(x) = (1 - x^{1/2})/(1 + x^{1/2}). \tag{9.60b}$$

This should be the most useful formula for computer algebra, although specific packages may require E and K to be entered as functions of the elliptic modulus $k(x) = [m(x)]^{1/2}$. Equation (9.60a) here corresponds to (16) in [61], (26a) in [116], and (11) in [57]; m_1 here corresponds to m^* in [116] but m in [57].

As shown in [57], a formula for dv/dx can be derived from (9.60a), namely

$$\mathrm{dv/d}x \; = \; -(3/4)(1 + x^{1/2})^{-1/2} \mathrm{K}[m_1(x)] \tag{9.61}$$

For $x = 1$, $m_1(x = 1) = 0$. From standard tables (e.g. [118]), K($m = 0$) = $\pi/2$. Hence, u($x = 1$) = $3\pi/8\sqrt{2}$, where u(x) [=–dv/dx] is defined by (9.65a). This result is used earlier and below.

In the original derivations, (9.60) was derived, by means of a Gauss/Landen transformation, from a 'less convenient' standard result, namely (14) in [61] or (24b) in [116]. In present notation, this 'less convenient' result would be written

$$\mathrm{v}(x) \; = \; 2^{-1/2} \{1 + (1 - x)^{1/2}\}^{1/2} \{\mathrm{E}[m_2(x)] - \{1 - (1 - x)^{1/2}\} \mathrm{K}[m_2(x)] \, , \tag{9.62a}$$

$$m_2(x) \; = \; 2(1 - x)^{1/2} / \{1 + (1 - x)^{1/2}\}. \tag{9.62b}$$

(The older derivations use a or α to denote $(1–x)^{1/2}$ [=$(1–y^2)^{1/2}$]). MG considered that both (9.60) and (9.62) were inconvenient for discussion of flyover; thus, they provided a 'more convenient' third alternative formula, namely their (9.15). However, our modern experience is that the computer algebra package MAPLE™ has no problem in dealing with (9.60) when $x > 1$, and hence $m_1 < 0$, (and k_1 [=$m_1^{1/2}$] imaginary).

The Pure Mathematics of v(x)—Integral Definitions

The function v(x) also has integral definitions, related to each other by mathematical transformations, and related to definitions in terms of K and E by formulae in

mathematical handbooks. Possibly, the integral definition most useful for further mathematical manipulations is

$$\mathrm{v}(x) \;=\; (3\times 2^{-3/2})\int_{b'}^{a'} (a'^2-\omega'^2)^{1/2}(\omega'^2-b'^2)^{1/2}\,\mathrm{d}\omega', \tag{9.63a}$$

where ω', here, is a notional integration variable, and the limits of integration are defined by

$$a' \;=\; \{1+(1-x)^{1/2}\}^{1/2}; \quad b' \;=\; \{1-(1-x)^{1/2}\}^{1/2} \tag{9.63b}$$

This formula can be deduced using (19) in [116] and the definition of the second FN constant b in Appendix 9.1. The integral in (9.63) is a standard form that can be expressed in terms of complete elliptic integrals via a standard result in elliptic integral handbooks (see [116]).

An alternative integral definition of $\mathrm{v}(x)$, which is closer to the form derived from JWKB treatments of tunnelling, is obtained by writing $\omega' = \xi^{1/2}$, which yields

$$\mathrm{v}(x) = (3\times 2^{-5/2})\int_{q'}^{p'} (p'-\xi)^{1/2}(\xi-q')^{1/2}\xi^{-1/2}\,\mathrm{d}\xi \tag{9.64a}$$

$$p' \;=\; (a')^2 \;=\; 1+(1-x)^{1/2};\; q' \;=\; (b')^2 \;=\; 1-(1-x)^{1/2} \tag{9.64b}$$

In all cases, primes have been added to the symbols only to distinguish them from other uses of the symbols a, b, p, q and ω in FE theory.

Formulae for the Set of Special Mathematical Functions

In addition to $\mathrm{v}(x)$, several other special mathematical functions are used in connection with the SN barrier. These all have formal definitions in terms of $\mathrm{v}(x)$ and its derivative $\mathrm{dv}/\mathrm{d}x$, and approximate formulae for them can be found by applying these definitions to formula (9.56). The definitions and (where useful) approximate formulae are

$$\mathrm{u}(x) \;\equiv\; -\mathrm{dv}/\mathrm{d}x \;\approx\; \tfrac{5}{6} - \tfrac{1}{6}\ln x, \tag{9.65a}$$

$$\mathrm{t}_1(x) \;\equiv\; \mathrm{v}(x) - \tfrac{4}{3}x(\mathrm{dv}/\mathrm{d}x) \;\approx\; 1 + \tfrac{1}{9}x - \tfrac{1}{18}x\ln x, \tag{9.65b}$$

$$\mathrm{t}_2(x) \;\equiv\; \mathrm{v}(x)/(1-x), \tag{9.65c}$$

$$\mathrm{s}(x) \;\equiv\; \mathrm{v}(x) - x(\mathrm{dv}/\mathrm{d}x) \;\approx\; 1 - \tfrac{1}{6}x, \tag{9.65d}$$

$$\mathrm{r}(\eta, x) \equiv \exp[\eta \cdot \mathrm{u}(x)] \approx \exp\left[\eta\left\{\tfrac{5}{6} - \tfrac{1}{6}\ln x\right\}\right], \tag{9.65e}$$

$$\mathrm{w}(x) \equiv \mathrm{ds}/\mathrm{d}(x^{-1}) = x^3(\mathrm{d}^2\mathrm{v}/\mathrm{d}x^2) = \tfrac{3}{16}\{x^2/(1-x)\}\mathrm{v}(x) \tag{9.65f}$$

The function $\mathrm{r}(\eta,x)$ was first given in [114], and is denoted there by r_{2012}; it replaces a slightly different function denoted in [114] by r_{1999}. The modern expression for $\mathrm{t}_2(x)$ was derived in [119]. The remaining functions are derived from related functions given in [86]. (A typographical error in (5.1) in [86] has been corrected here.) It is suggested that these mathematical functions should simply be called by names of the type 'the field emission special mathematical function u(x)'.

At $x = 1$, the exact values of these functions can be obtained from the theory of complete elliptic integrals (see below). Relevant values are: $\mathrm{u}(1) = 3\pi/(8\sqrt{2}) \cong 0.8330406$; $\mathrm{t}_1(1) = \pi/(2\sqrt{2}) \cong 1.110721$.

If more accurate values of these functions are needed for other x-values, then formulae in Appendix 9.2 (derived from those in [86]), can be used. These have been derived as numerically optimised fits to MAPLE™ evaluations of formulae (9.60) and (9.61), and have a maximum absolute error of 8×10^{-10}. Sample MAPLE™ evaluations have been checked against independently programmed evaluations of integral definitions of v(x), with agreement to 12 decimal places.

Use of v(x) in Modelling Transmission Across the Laurent-form Barrier

The mathematical function v(x) is useful in modelling transmission across, not only the SN barrier, but also across barriers of similar mathematical form that appear in the theories of semiconductor FE, electrostatic field ionisation, field desorption and field evaporation. This section shows that this function is applicable to *any* barrier with motive energy of the form $M(z) = H - Cz - B/z$, where H, C and B are constants. The result can then be customised to the SN barrier.

A function of the form just stated seems not to have any well defined name, so I refer to it as the *basic Laurent-form barrier*. In practice, it is more convenient to write C in the form 'eF', where e (as usual) is the elementary positive charge and F is a parameter initially regarded as of general significance only. Thus, the barrier of interest has motive energy

$$M(z) = H - eFz - B/z \tag{9.66}$$

The zeros of $M(z) = 0$ are:

$$z = (H/2eF)\left\{1 \pm \sqrt{1 - 4eFB/H^2}\right\}. \tag{9.67}$$

By defining the *barrier parameter* μ and a mathematical parameter α' by

$$\mu = 4eFB/H^2, \quad \alpha' = +(1-\mu)^{1/2}, \tag{9.68}$$

(9.67) can be written in the simplified form

$$z = (H/2eF)\left(1 \pm \alpha'\right) \tag{9.69}$$

From a generalisation of (9.15), noting the analogy with the relation $2\kappa_e = 2(2m_e)^{1/2}/\hbar$, the barrier strength G for an entity of mass m (which is not necessarily an electron), for the basic Laurent-form barrier, can be written

$$G = (3e/2)\left[\tfrac{4}{3}(2m)^{1/2}/e\hbar\right] \times \int_{(H/2eF)(1-\alpha')}^{(H/2eF)(1+\alpha')} (H - eFz - B/z)^{1/2}\,\mathrm{d}z. \tag{9.70}$$

The transformation $z = (H/2eF)\xi$ yields

$$G = G_{0,m} \cdot (3 \times 2^{-5/2}) \int_{1-\alpha'}^{1+\alpha'} (-\xi^2 + 2\xi - \mu)^{1/2}\xi^{-1/2}\,\mathrm{d}\xi. \tag{9.71a}$$

where the symbol $G_{0,m}$ denotes the combination

$$G_{0,m} = [(4/3)(2m)^{1/2}/e\hbar](H^{3/2}/F). \tag{9.71b}$$

Since the zeros of the expression $(\xi^2 - 2\xi + \mu)$ are $\xi = 1 \pm (1-\mu)^{1/2} = 1 \pm \alpha'$, (9.71a) can be rewritten as

$$G = G_{0,m} \cdot (3 \times 2^{-5/2}) \times \int_{1-\alpha'}^{1+\alpha'} \{(1+\alpha') - \xi\}^{1/2}\{\xi - (1-\alpha')\}^{1/2}\xi^{-1/2}\,\mathrm{d}\xi. \tag{9.72}$$

Comparison with (9.64), taking $p' = (1-\alpha')$, $q' = (1+\alpha')$ shows that (9.72) involves one of the definitions of $\mathrm{v}(x)$, and hence can be written in the form

$$G = \mathrm{v}(\mu) \cdot G_{0,m}. \tag{9.73}$$

Hence, when μ is defined by (9.68), then $\mathrm{v}(x = \mu)$ is the barrier form correction factor for the basic Laurent-form barrier.

For the specific case of electron tunnelling through a Schottky-Nordheim barrier, F resumes its role as the local barrier field, $G_{0,m}$ becomes the quantity G^{ET} used earlier, and $B = e^2/16\pi\varepsilon_0$. Equation (9.68) then becomes a definition of the parameter f_H, or—in the case when $H = \phi$—the parameter f. Hence, the barrier form correction factor for the SN barrier becomes given by $\mathrm{v}(x = f_{\mathrm{H}})$ or (when $H = \phi$) $\mathrm{v}(x = f)$.

The Choice between Nordheim Parameter and Scaled Field

In the treatment above, the Nordheim parameter y does not appear. This is because I have chosen to define my barrier parameter (μ) by (9.68), rather than as a parameter

equal to $\sqrt{\mu}$. Either choice could be made to work mathematically, and in 1928 Nordheim took the opposite choice. However, there is no good mathematical reason to stick to his choice, particularly since this choice seems to have been part of the cause of the mathematical error made in his paper.

In the present treatment, it is probably better to leave μ as the barrier parameter for the basic Laurent-form barrier, and ask the alternative question: "When this theory is applied to the SN barrier, is it better to set $\mu = f_H$ (or f) or to set $\mu = y^2$?" Obviously, the latter is the more familiar convention, but the author's strong view is that use of f_H and f is in principle much the better convention. Arguments are as follows.

(1) For reasons noted earlier, the *natural* pure mathematical variable to use is the Gauss variable x: the modelling equivalents of x are μ, f_H and f, rather than y.
(2) The concept of 'scaled field' is probably easier to understand and use than the Nordheim parameter (which physically is 'scaled reduction in SN-barrier height').
(3) The parameters f_H and f are proportional to the barrier field F, whereas y is proportional to $\sqrt{F}$. This often makes f easier to use than y, particularly in the context of an orthodoxy test [109].
(4) The symbols f_H and f each have unique definitions, whereas, historically, depending on the paper and context in which it appears, the symbol y may mean $\sqrt{x}$, $\sqrt{f_H}$ or $\sqrt{f}$.
(5) The concept of 'scaled field' is, in principle, more general than the Nordheim parameter concept, and can easily be extended physically (when suitably modified in detail) to apply to real physical barriers. However, the parameter y represents the scaled barrier height ONLY for a SN barrier.

In the author's view, it would make for a simpler and clearer system if the use of y were gradually phased out, and relevant formulas and tables using y were replaced by formulas and tables using x or f.

In the present mixed system, a danger for non-experts is confusion between the meanings of 'v(f)' and 'v(y)', as this is not a change in symbol but a change in variable and (strictly) a change in mathematical function. To avoid confusion, when substituting (say) f for x in the function v(x), I often write v($x = f$). Also, again to avoid confusion, the argument-free symbols v_F and t_F are sometimes used as alternative notation in Murphy-Good FE equations, i.e. one writes: $v_F \equiv v(x = f)$, $t_F \equiv t_1(x = f)$.

9.7 Device and System Theory Issues

Many issues covered in Sects. 9.7 and 9.8 have recently been discussed in detail elsewhere [20, 21], and some are still topics of active research; hence, only an overview is given here. For details, readers are referred to the cited papers.

For clarity in Sect. 9.7, the local barrier field and local ECD are denoted by F_L and J_L. In practice, interest is usually in their values at some point 'C' regarded

as 'characteristic' of the emission. Values of barrier field and ECD at this point are denoted by F_C and J_C, respectively. In emission modelling (where it is usually assumed that local work function is constant across the surface), it is usual to take 'C' at an emitter apex, where (except in special circumstances) the field is locally highest.

9.7.1 Basic Auxiliary Parameters for Ideal FE Devices/Systems

To apply emission theory to FE devices and systems, *auxiliary parameters* are needed that relate F_C and J_C (or, more generally, F_L and J_L) to measured voltage (V_m) and measured current (I_m). Real FE systems involve electrical measurement and control circuits, of which the FE device forms part. Such systems may exhibit current-leakage and voltage-loss effects. The usual assumption is that current-leakage effects, once detected, can be eliminated by careful system design, but that voltage-loss effects are intrinsic to many forms of FE device, and cannot easily be eliminated.

An *ideal FE device/system* is one where it is adequate to assume that no such effects occur, and that no other distorting effects (such as, amongst others, field dependent changes in geometry, current dependence in field enhancement factors, temperature dependence in work function, or effects due to field emitted vacuum space-charge) occur. In an ideal system, the $I_m(V_m)$ relationship is determined completely by the following factors: the total system geometry (including unchanging emitter shape); unchanging emitter surface composition; and the emission process.

At this point in time (and probably for some years to come), in $I_m(V_m)$ data-interpretation theory it will be simplest to consider only the behaviour of *ideal* FE devices/systems. With real systems, it will be necessary to precede data analysis by an orthodoxy test [109] that uses a FN plot or MG plot to check whether the device/system is ideal.

For ideal FE devices/systems, a convenient theoretical approach uses two linked equations, the first an expression for measured current I_m in terms of characteristic field F_C (or corresponding scaled field f_C), and the second an expression for F_C (or f_C) in terms of measured voltage V_m. On omitting the superscript 'GB' used earlier, the equation for local ECD is written

$$J_L = \lambda_L J_{kL}. \tag{9.74}$$

A formal integration is done over the whole emitting area and the result is written

$$I_e = \int J_L \mathrm{d}A = A_n J_C = A_n \lambda_C J_{kC} = A_f J_{kC}. \tag{9.75}$$

Here A_n is the so-called *notional emission area*, and A_f [$\equiv \lambda_C A_n$] is the *formal emission area* (which will be substantially different in value from A_n if λ_C is substantially different from unity).

The quantity extracted from the corresponding 'ideal' FN or MG plot is a *formal* emission area A_f. Its type and value depend on what is assumed about the form of the tunnelling barrier. With SPME methodology, the best simple assumption is a SN barrier [82, 110]. Hence, for an ideal FE device/system, it is currently best to write

$$I_m = A_f^{SN} J_{kC}^{SN}, \tag{9.76}$$

where A_f^{SN} is the formal emission area for the SN barrier, and J_{kC}^{SN} is the characteristic kernel current density for the SN barrier, as given by (9.45), with $F = F_C$ and $f = f_C$, (or as given by an equivalent scaled equation).

Equation (9.76), in its various formats that incorporate the correction factor λ_C^{SN}, has been called [110] the *Extended Murphy-Good equation*. As indicated earlier, my current thinking [43] is that this correction factor, which has unknown functional dependencies and values, probably always lies within the range of $0.005 \leq \lambda_C^{SN} \leq 14$. But quite possibly the lower limit is too pessimistic.

The relation between field F_C and measured voltage V_m can be written in three alternative forms, all of which are in use. The formulae are

$$F_C = \beta_V V_m = V_m/\zeta_C = V_m/k_C r_a. \tag{9.77}$$

Here, β_V is the (characteristic) *voltage-to-local-field conversion factor (VCF)* (also denoted by 'β' in older literature) (unit: nm^{-1}, or equivalent); ζ_C is the (characteristic) *voltage conversion length (VCL)* (unit: nm, or equivalent); r_a is the emitter apex radius of curvature, and k_C (dimensionless) is called a *shape factor* or a *field factor*. The author now prefers the form involving the VCL ζ_C, although this is not the usual literature choice.

The VCL ζ_C is not a physical length (except in special cases). Rather, it is a system-geometry characterization parameter that assesses how easy it is to 'turn the emitter on'. For example, if we consider that an apex barrier field of 2 V/nm is needed to turn the emitter on, and $\zeta_C = 10$ nm, then turn-on occurs at $V_m = 20$ V.

For an *ideal* FE device/system, all these '*auxiliary parameters*' are constants (independent of field and voltage), and depend only on the (unchanging) total system geometry, including emitter shape. The literature contains many electrostatic analyses that lead to different formulae for the auxiliary parameters in (9.77), for different assumptions about system geometry and emitter shape. For single tip field emitters (STFEs), it is often thought that the 1956 *sphere-on-orthogonal-cone (SOC) model* of Dyke et al. [31] is the best simple choice for a needle-shaped STFE, and that the hemisphere-on-cylindrical-post (HCP) model (e.g. [120]) is the best simple choice for a post-shaped STFE.

9.7.2 Field Enhancement Factors

Additional auxiliary parameters, called field enhancement factors (FEFs) and denoted here by the symbol γ or γ_{M}, are used to describe the behaviour of single post-like emitters and of large-area field electron emitters (LAFEs). In the literature, FEFs are usually denoted by the symbol β. Due to the possibility (and sometimes reality) of confusion between this use and literature use of the symbol 'β' to denote a voltage conversion factor, I prefer to use the basic symbol γ.

A FEF can be defined at any emitter surface location, but interest is mainly in some characteristic value, γ_{MC}, usually thought of as the value at the apex of a single emitter or (with a LAFE) at the apex of a strongly emitting particular emitter, or at some other 'high point'.

Several different kinds of FEF are in use. These apply to slightly different geometrical situations, but in all cases, the characteristic FEF can be related to the characteristic VCL by an equation of the general form

$$\gamma_{\mathrm{MC}} = d_{\mathrm{M}}/\zeta_{\mathrm{C}} \tag{9.78}$$

where d_{M} is a *macroscopic distance* associated with the chosen mode of definition. The most common variants are the 'gap FEF' (replace 'M' by 'G'), where d_{G} is the length of the gap between characteristic location 'C' and the counter-electrode, and the 'plate FEF' (replace 'M' by 'P') where the emitter stands on one of a pair of well-separated parallel planar plates (PPP), separated by a distance d_{P}.

For an *ideal* PPP system, the voltage V_{P} between the plates is equal to the measured voltage V_{m}, so it follows that

$$F_{\mathrm{C}} = V_{\mathrm{m}}/\zeta_{\mathrm{C}} = V_{\mathrm{P}}/\zeta_{\mathrm{C}} = V_{\mathrm{P}}\gamma_{\mathrm{PC}}/d_{\mathrm{P}} = \gamma_{\mathrm{PC}}F_{\mathrm{P}}, \tag{9.79}$$

where the 'plate field' F_{P} is the mean field between the parallel plates. Thus, consistency is demonstrated between relation (9.79) and the more common definition of characteristic plate FEF γ_{PC}, namely $\gamma_{\mathrm{PC}} = F_{\mathrm{C}}/F_{\mathrm{P}}$.

For an *ideal* PPP device/system, the plate FEF is independent of the measured current and voltage. Provided that d_{P} is many times the emitter height h (preferably $d_{\mathrm{P}} > 10\,h$), this plate FEF characterises the 'sharpness' of the emitter alone. However, the value of a gap FEF may often depend on the whole system geometry.

Plate FEFs, gap FEFs and similar parameters are collectively called *macroscopic FEFs* here (previously, this name was applied only to plate FEFs).

LAFEs typically consist of large or very large numbers of individual emitters or emission sites, and are often modelled as arrays of upright posts, although this is not appropriate for all LAFE geometries. In LAFE array models, electrostatic interactions (often called 'shielding') occur between post-like emitters, and lead to significant electrostatic depolarisation effects if the emitters are sufficiently close together. There is now an extensive (and growing) literature on this topic, but details are outside the scope of this chapter. Gateways to it are provided by [121–125].

Note that, with resistive emitters, it is possible for current flow through the emitter to reduce the field enhancement factor, and hence make the voltage conversion length into a current-dependent parameter [126, 127]. The author is now of the view that this may be a common cause of non-linearity in experimental Fowler-Nordheim plots.

9.7.3 Macroscopic Current Density and Area Efficiency

The macroscopic or 'LAFE-average' emission current density J_M is defined by

$$J_M = I_m/A_M, \tag{9.80}$$

where A_M is the macroscopic area (or "footprint") of the device. From the general version of (9.76), we have

$$J_M = (A_f/A_M)J_{kC} \equiv \alpha_f J_{kC}. \tag{9.81}$$

where α_f [$\equiv A_f/A_M$] is the *formal area efficiency* of the LAFE. It is also possible, although less useful (due to the uncertainties in the value of λ_C), to define a *notional area efficiency* α_n by

$$\alpha_n \equiv A_n/A_M = (A_f/\lambda_C A_M) = \alpha_f/\lambda_C. \tag{9.82}$$

Values of formal area efficiency can be extracted from experimental data, but—as with formal emission area—the values extracted [using definition (9.81)] depend on the choice of barrier form. As of now, when most $I_m(V_m)$ data analysis is carried out in the context of SPME methodology, the best simple choice is the SN barrier, as already indicated.

The parameter α_f^{SN} is a measure of what fraction of the area of a LAFE is apparently emitting electrons. Relatively few values for α_f^{SN} have been reliably extracted, but it is thought that values for these dimensionless parameters may vary widely from material to material, and vary as between different LAFE processing regimes. Values are not well known, but are thought to perhaps typically lie in the range 10^{-7} to 10^{-4}. (See [128] for corresponding, but older, remarks about the possible value of α_n.)

Unfortunately, it has been relatively common practice in FE technological literature not to make a distinction between local current density J_L (or J_C) and macroscopic current density J_M, and to omit an area efficiency factor from published equations where it principle it ought to appear. This has given rise to situations where there are undiscussed large apparent discrepancies between theory and experiment in particular papers, and has apparently led some researchers to think that area information cannot be extracted from FN plots, even for ideal FE devices/systems. These issues have been considered elsewhere [20, 128, 129] but now require more careful discussion and further research, including reexamination of existing published data.

9.8 The Interpretation of Measured Current-Voltage Data

The commonest methodology of characterising FE materials and devices involves measurement of current-voltage $[I_{\rm m}(V_{\rm m})]$ characteristics, and then making some form of Fowler-Nordheim (FN) plot (see [20]), although there has been a recent suggestion that an alternative plot form—the so-called *Murphy-Good (MG) plot*—would provide greater precision, particularly in extracting formal emission area $A_{\rm f}^{\rm SN}$ [110], and/or that multi-parameter numerical fitting methods could be used. This section outlines the main points suggested in recent overviews [20, 21].

1. Both FN and MG plots should be made using raw measured $I_{\rm m}(V_{\rm m})$ data. It is advisable to discontinue the widespread community practice of pre-converting voltages to become (apparent) macroscopic fields, because the pre-conversion equation often used can be defective when applied to real systems [130], and this can lead to spurious results for characterization parameters [109].
2. Virtually all FE $I_{\rm m}(V_{\rm m})$ data analysis is currently carried out within the framework of SPME methodology. Exploration of how to carry out data analysis within a 'point-form emitter methodology' that treats the emitter as needle-shaped or post-shaped is long overdue and an active topic of research (e.g. [14, 44, 45]). However, the immediate practical need is to improve SPME methodology, in order to extract precisely defined parameters; the remainder of this section discusses this.
3. To analyse either a FN plot or a MG plot, within SPME methodology, it should be assumed that FE is adequately described by the Extended Murphy-Good FE equation. Formulae given here relate to this case.
4. An orthodoxy test [20, 109] should always be applied to the plot before detailed analysis; a form suitable for an MG plot is described in [131, 132].
5. For a FN plot that passes the orthodoxy test, the voltage conversion length (VCL) $\zeta_{\rm C}$ can be extracted by using the formula

$$\{\zeta_{\rm C}\}^{\rm extr} \; = \; -S_{\rm FN}^{\rm fit}/{\rm s_t} b\phi^{3/2}. \tag{9.83}$$

   where $S_{\rm SN}^{\rm fit}$ is the slope of a straight line fitted to the FN plot, and $\rm s_t$ is the fitting value [20] of the slope correction function $\rm s(\mathit{f})$. A related macroscopic FEF can then be derived by using formula (9.78).
6. For a FN plot that passes the orthodoxy test, the formal emission area $A_{\rm f}^{\rm SN}$ can be extracted by using the formula

$$\{A_{\rm f}^{\rm SN}\}^{\rm extr} \; = \; \Lambda_{\rm FN}^{\rm SN} R_{\rm FN}^{\rm fit} (S_{\rm FN}^{\rm fit})^2, \tag{9.84}$$

   where $\ln\{R_{\rm FN}^{\rm fit}\}$ is the intercept made by the fitted straight line with the $1/V_{\rm m} = 0$ axis, and $\Lambda_{\rm FN}^{\rm SN}$ is the *emission area extraction parameter for the SN barrier, using a FN plot,* given by

$$\Lambda_{\rm FN}^{\rm SN} = 1/[(ab^2\phi^2)({\rm r_t s_t^2})], \tag{9.85}$$

where r_t is the fitting value [21] of the intercept correction function $r(\eta,f)$. For $\phi = 4.50$ eV, A_{FN}^{SN} has the approximate value of 6 nm^2/A. A related formal area efficiency $\{\alpha_f^{SN}\}^{extr}$ (for the SN barrier) can be derived as $\{A_f^{SN}\}^{extr}/A_M$.

7. Equivalent formulae for the MG plot can be found in [110]. For ideal FE devices/systems, the extracted values of A_f^{SN} and and α_f^{SN} are significantly more precise when a MG plot is used.
8. Technological FE papers often use the elementary FE equation (see Table 9.3) to extract information from the FN plot slope. As compared with a MG-type equation, this omits the factor s_t in (9.83), and causes an error of about 5%. In practice, this equation is not used to extract emission area values, but if it were then an error of order 100 would arise, resulting from the omission of the factor $(r_t s_t^2)$ in (9.85). The existence of this potential error is a strong argument for using a Murphy-Good-type equation to interpret FE current-voltage data (whether using a FN plot or a MG plot).
9. As noted above, when writing an equation for macroscopic current density J_M, technological FE papers often omit the factor α_f. This can lead to very large apparent discrepancies between theory and experiment. A factor describing area efficiency should always be included in equations for macroscopic current density.
10. For experimental data that fail the orthodoxy test, it is possible to use the procedure of *phenomenological adjustment* [20], to extract empirical values of ζ_C and γ_{MC} that are more realistic than straightforward (but invalid) application of formula (9.83) would yield. The problem of extracting valid estimates of *area-like* characterization parameters for non-ideal devices/systems, when the cause of non-ideality is unknown, is so complicated and messy that it almost seems premature to attempt to do so.
11. There remains an urgent need for research into a comprehensive method of analysing data from non-ideal devices/systems, but only limited progress has been made. One route forwards is provided by the work of Bachmann et al. [133].

9.9 Future Needs

Field electron emission has many applications in technology, and good basic theory is needed in order to optimise both the interpretation of results and the prediction of technology behaviour, particularly in contexts such as electrical breakdown. More generally, FE has an enduring role within physics as one of the fundamental exemplars of quantum-mechanical tunnelling, and its theory deserves to be described clearly, correctly and in an integrated manner. As an exemplar of tunnelling, FE may have a role in new fields such as quantum biology.

This chapter has primarily been about the consolidation of FE theory and the interpretation of *ideal* FE measured current-voltage data, both of these within the framework of smooth-planar-metal-like-emitter methodology. This methodology is near-universally used by experimentalists as a first-approximation approach to interpreting FE current-voltage data, whatever the material used in experiments.

This chapter partially describes the first of several stages needed to put FE theory onto a more scientific basis by making it easier to link theory and experiment. There is already much relevant emission theory published, but future stages will need to develop a methodology for interpreting measured current-voltage data that takes into account: (a) the shape of emitters, as it affects the process of integration in (9.75); (b) the effect of emitter shape and surface curvature on local transmission probability; (c) the influence on transmission probability of including atomic structure, first for planar surfaces and then for curved surfaces; and (d), if it proves necessary, any weaknesses that may be confirmed/discovered in present formulations of basic quantum-mechanical transmission theory.

It would be helpful to extend/customise data-interpretation methodologies to deal with the specific problems of semiconductors (such as surface states, band-bending and field penetration), and to deal with any further specific problems related to carbon-based materials, in particular carbon nanotubes. However, it may be better to delay this until after the development of methodologies for interpreting data from point-form emitters, since new transport mechanisms may need to be considered, such as transport via surface states on the emitter shank. There is also the possibility that emitters that are not bulk metals may be both 'surface conductors' and 'surface-state emitters', as a result of the presence of active surface states, and thus 'quasi-metallic' in their behaviour.

It would be helpful to develop data-interpretation methodologies that apply to FE devices/systems that are non-ideal for various specific reasons, and to develop (if possible) some reliable method for identifying the precise reason for non-ideality. Finally, it would probably be helpful to develop variants of all the above that apply to thermal-field emission regimes. There would appear to be many years of work ahead.

Appendix 9.1. Fundamental and Universal Constants Used in Field Emission Theory

Table 9.4 presents values of the fundamental constants used in FE theory, both in SI units and in the 'field emission customary units' in which they are often given. Both sets of units are defined in the context of ISQ-based equations, and the normal rules of quantity algebra apply to their values. These customary units take the electronvolt, rather than the joule, as the unit of energy, and normally measure field in V/nm and charge in units equal to the elementary positive charge. Their use greatly simplifies basic calculations when energies are measured in eV and fields in V/nm, as is often

Table 9.4 The May 2019 values of the electronvolt (eV) and fundamental constants used in emission physics, given in SI units (either exactly or to 11 significant figures), and in field emission customary units (to 7 significant figures). Asterisks indicate constants where (since May 2019) the value in SI units is specified exactly

Name	Symbol	SI units		Atomic-level units based on eV	
		Numerical value	Units	Numerical value	Units
Electronvolt*	eV	$1.602\ 176\ 6340 \times 10^{-19}$	J	1	eV
Elementary (positive) charge*	e	$1.602\ 176\ 6340 \times 10^{-19}$	C	1	eV V^{-1}
Elementary constant of amount of substance[a]	n_1	$1.660\ 539\ 0672 \times 10^{-24}$	mol	1	Entity[a]
Unified atomic mass constant	m_u	$1.660\ 539\ 6660 \times 10^{-27}$	kg	$1.036\ 427 \times 10^{-26}$	eV nm^{-2} s^2
Electron mass in free space	m_e	$9.109\ 383\ 7015 \times 10^{-31}$	kg	$5.685\ 630 \times 10^{-30}$	eV nm^{-2} s^2
Vacuum electric permittivity	ε_0	$8.854\ 187\ 8128 \times 10^{-12}$	F m^{-1}	$5.526\ 349 \times 10^{-2}$	eV V^{-2} nm^{-1}
$\varepsilon_0 \times 4\pi$	$4\pi\varepsilon_0$	$1.112\ 650\ 0554 \times 10^{-10}$	F m^{-1}	0.694 4615	eV V^{-2} nm^{-1}
Planck's constant*	h_P	$6.626\ 070\ 15 \times 10^{-34}$	J s	$4.135\ 668 \times 10^{-15}$	eV s
Planck's constant $\div 2\pi$	$\hbar$	$1.054\ 571\ 8176 \times 10^{-34}$	J s	$6.582\ 120 \times 10^{-16}$	eV s
Boltzmann's constant*	k_B	$1.380\ 649 \times 10^{-23}$	J K^{-1}	$8.617\ 333 \times 10^{-5}$	eV K^{-1}

[a]Name used here, not 'official'. n_1 is equal to the reciprocal of the (exactly given) Avogadro constant

the case. The numerical values of the constants, measured in both sets of units, are based on the values of the fundamental constants given in October 2019 on the website of the US National Institute of Standards and Technology (NIST) (see http://physics.nist.gov/constants). These values incorporate the SI system changes made in May 2019.

Table 9.5 gives the values, in FE customary units, of many universal constants, and combinations thereof, that are used in field emission theory. Where necessary, the tabulated values take the May 2019 changes into account. The third column in Table 9.5 is interesting, because it shows that several of the universal constants used in field emission are related in a relatively simple way to the constants κ_e and z_S, which are very basic universal constants that appear centrally in the Schrödinger equation and in statistical mechanics, respectively.

Table 9.5 Some fundamental constants used in field emission and related theory. Values are given in field emission customary units

Name	Symbol	Derivation	Expression	Numerical value	Units
Coulomb law constant	–	–	$4\pi\varepsilon_0$	0.6944 615	eV V^{-2} nm^{-1}
Image potential energy constant	B	–	$e^2/16\pi\varepsilon_0$	0.3599 911	eV nm
Sommerfeld electron supply constant	z_S	–	$4\pi e m_e/h_P^3$	1.618 311×10^{14}	A m^{-2} eV^{-2}
Schrödinger equation constant for electron[a]	κ_e	–	$(2m_e)^{1/2}/\hbar$	5.123 168	$eV^{-1/2}$ nm^{-1}
JWKB constant for electron[b]	(g_e)	$2\kappa_e$	$2(2m_e)^{1/2}/\hbar$	10.24 624	$eV^{-1/2}$ nm^{-1}
Triangular barrier constant	c_κ	$(\kappa_e/e)^{1/3}$	$(2m_e)^{1/6}/(e\hbar)^{1/3}$	1.723 903	$eV^{-1/2}$ $(V/nm)^{-1/3}$
First Fowler-Nordheim constant	a	$z_S(e/2\kappa_e)^2$	$e^3/8\pi h_P$	1.541 434	μA eV V^{-2}
Second Fowler-Nordheim constant	b	$4\kappa_e/3e$	$(4/3)(2m_e)^{1/2}/e\hbar$	6.830 890	$eV^{-3/2}$ (V/nm)
–	ab^2	$4z_S/9$	–	7.192 493×10^{-5}	A nm^{-2} eV^{-2}
Universal theoretical Richardson constant	A_{R0}	$z_S k_B^2$	$4\pi e m_e k_B^2/h_P^3$	1.201 735×10^{6}	A m^{-2} K^{-2}
Schottky constant	c_S	–	$(e^3/4\pi\varepsilon_0)^{1/2}$	1.199 985	eV $(V/nm)^{-1/2}$
		c_S^2	$e^3/4\pi\varepsilon_0$	1.439 965	eV^2 $(V/nm)^{-1}$
		c_S^{-2}	$4\pi\varepsilon_0/e^3$	0.694 4615	eV^{-2} (V/nm)
		ac_S^{-4}	–	7.333 978×10^{11}	A m^{-2} eV^{-3}
		bc_S^2	–	9.836 239	$eV^{1/2}$
		$bc_S^{1/2}$	–	7.482 819	eV $(V/nm)^{3/4}$
–	C_p	$3bk_B/2$	$2(2m_e)^{1/2}k_B/e\hbar$	8.829 607×10^{-4}	$eV^{-1/2}$ K^{-1} (V/nm)
–	C_q	$(2/3t_b)\times(bc_S^{1/2}k_B)$	$(\hbar/\pi m_e^{1/2})\times(4\pi\varepsilon_0 e)^{1/4}$	930.8202	K $(V/nm)^{-3/4}$

[a] As introduced in the Fowler and Nordheim 1928 paper [34]
[b] Previously denoted by g_e, but notation $(2\kappa_e)$ is now preferred

Appendix 9.2. High-Precision Formulae for v(x) and u(x)

This Appendix provides formulae for estimating high-precision values of the FE special mathematical functions v(x) and u(x) [$\equiv -\mathrm{dv}/\mathrm{d}x$] (and hence of all the FE special mathematical functions). The parameter x is the *Gauss variable*. The two functions are estimated by the following series, derived from those given in [86] by replacing the symbol *l'* by the symbol x now preferred, and by slightly adjusting the form of the resulting series for v(x) (without changing its numerical predictions)

$$\mathrm{v}(x) \cong (1-x)\left(1+\sum_{i=1}^{4} p_i x^i\right) + x\ln x \sum_{i=1}^{4} q_i x^{i-1} \tag{9.86}$$

$$\mathrm{u}(x) \cong \mathrm{u}_1 - (1-x)\sum_{i=0}^{5} s_i x^i - \ln x \sum_{i=0}^{4} t_i x^i \tag{9.87}$$

Values of the constant coefficients p_i, q_i, s_i and t_i are shown in Table 9.6.

It can be seen that at the values $x = 0$, 1, formula (9.86) generates the correct values $\mathrm{v}(0) = 1$, $\mathrm{v}(1) = 0$, and that at $x = 1$, formula (9.87) generates the correct value $\mathrm{u}(1) = \mathrm{u}_1 = 3\pi/8\sqrt{2}$.

Equation (9.86) mimics the form of the lower-order terms in the (infinite) exact series expansion for v(x) [57], but the coefficients in Table 9.6 have been determined by numerical fitting to exact expressions for v(x) and u(x) (in term of complete elliptic integrals) evaluated by the computer algebra package MAPLE™. In the range $0 \leq x \leq 1$, v(x) takes values lying in the range $1 \geq \mathrm{v}(x) \geq 0$, and the maximum error associated with formulae (9.86) and (9.87) is less than 8×10^{-10} [57]. In Murphy-Good-type FE equations, these formulae are applied by setting $x = f_C$.

Table 9.6 Numerical constants for use in connection with (9.86) and (9.87)

i	p_i	q_i	s_i	t_i
0	–	–	0.053 249 972 7	0.187 5
1	0.032 705 304 46	0.187 499 344 1	0.024 222 259 59	0.035 155 558 74
2	0.009 157 798 739	0.017 506 369 47	0.015 122 059 58	0.019 127 526 80
3	0.002 644 272 807	0.005 527 069 444	0.007 550 739 834	0.011 522 840 09
4	0.000 089 871 738 11	0.001 023 904 180	0.000 639 172 865 9	0.003 624 569 427
5	–	–	−0.000 048 819 745 89	–
$\mathrm{u}_1 \equiv \mathrm{u}(x=1) = 3\pi/8\sqrt{2} \cong 0.8330405509$				

References

1. M.I. Elinson (ed.), *Unheated Cathodes* (Soviet Radio, Moscow, 1974). (in Russian)
2. A. Modinos, *Field, Thermionic, and Secondary Electron Emission Spectroscopy* (Plenum, New York, 1984) (Reprinted by: Springer, online, 2013)
3. G.A. Mesyats, *Explosive Electron Emission* (URO-Press, Ekaterinburg, 1998)
4. V.M. Zhukov, *Surface Processes in Field Electron Emission* (Isdatelstvo, St Petersburg, 2007). (in Russian)
5. W. Zhu (ed.), *Vacuum Microelectronics* (Wiley, New York, 2001)
6. G.N. Fursey, *Field Emission in Vacuum Microelectronics* (Kluwer, New York, 2001)
7. K.L. Jensen, Electron emission physics. Adv. Imaging Electron. Phys. **149**, 1–338 (2007)
8. Y. Saito (ed.), *Carbon Nanotube and Related Field Emitters* (Wiley, Weinheim, 2010)
9. N.V. Egorov, E.P. Sheshin, *Field Emission: Principles and Practice* (Intellect, Russia, 2011). (in Russian) (Reference [13] is the English edition)
10. S.-D. Liang, *Quantum Tunnelling and Field Electron Emission Theories* (World Scientific, Singapore, 2014)
11. M. Bronsgeest, *Physics of Schottky Electron Sources* (Pan Stanford, Singapore, 2014)
12. A. Evtukh, H. Hartnagel, O. Yilmazoglu, H. Mimura, D. Paylidis, *Vacuum Nanoelectronic Devices* (Wiley, Chichester, 2015)
13. N.V. Egorov, E.P. Sheshin, *Field Emission Electronics* (Springer, 2017)
14. K.L. Jensen, *Introduction to the Physics of Electron Emission* (Wiley, Hoboken, NJ, 2018)
15. L.W. Swanson, G.A. Schwind, Review of ZrO/W Schottky cathode, in *Charged Particle Optics*, 2nd edn., ed. by J. Orloff (CRC Press, Baton Rouge, 2009)
16. L.W. Swanson, G.A. Schwind, A review of the cold-field electron cathode, in *Cold Field Emission and the Scanning Transmission Electron Microscope*, ed. by P. Hawkes. Adv. Imaging Electron Phys. **159**, 63–100 (2009)
17. R.G. Forbes, The theory of bright field electron and field ion emission sources, in *Nanfabrication Using Focused Ion and Electron Beams*, ed. by I. Utke, S. Moshkalev, P. Russell (Oxford University Press, Oxford, 2012)
18. Z.-B. Li, Density functional theory for field emission from carbon nanostructures. Ultramicroscopy **159**, 162–172 (2015)
19. V. Filip, L.D. Filip, H. Wong, Review on peculiar issues of field emission in vacuum nanoelectronic devices. Solid State Electron. **138**, 3–15 (2017)
20. R.G. Forbes, J.H.B. Deane, A. Fischer, M.S. Mousa, Fowler-Nordheim plot analysis: a progress report. Jordan J. Phys. **8**, 125–147 (2015). arXiv:1504.06134v7
21. R.G. Forbes, J.H.B. Deane, A.G. Kolosko, S.V. Fillipov, E.O. Popov, Reinvigorating our approach to field emission area extraction (because Murphy-Good plots are better than Fowler-Nordheim plots), in 32nd International Vacuum Nanoelectronics Conference & 12th International Vacuum Electron Sources Conference, Cincinnatti, July 2019. Technical Digest, p. 23. https://doi.org/10.13140/rg.2.2.32112.81927
22. Bureau International des Poid et Mesures (BIPM), *The International System of Units*, 9th ed. ("The SI Brochure") (BIPM, Sèvres, 2019). http://www.bipm.org/en/publications/si-brochure/
23. International Standards Organization (ISO), International Standard ISO 80000-1:2009. *Quantities and units. Part I: General* (ISO, Geneva, 2009, corrected 2011)
24. R. Gomer, *Field Emission and Field Ionization* (Harvard University Press, Cambridge, MA, 1961; Reprinted by: AIP, New York, 1993)
25. Adapted from an earlier version of, and consistent with, Reference [23], Section 3.20, Note 2
26. H. Yanagisawa, C. Hafner, P. Dona, M. Klöckner, D. Leuenberger, T. Greber, J. Osterwalder, M. Hengsberger, Laser-induced field emission from a tungsten tip: optical control of emission sites and the emission process. Phys. Rev. B **81**, 115429 (2010)
27. P. Hommelhoff, M.F. Kling, *Attosecond Nanophysics* (Wiley, Weinheim, 2015)
28. H.R. Reiss, Unsuitability of the Keldysh parameter for laser fields. Phys. Rev. A **82**, 023418 (2010)

29. J.D. Jackson, *Classical Electrodynamics*, 3rd edn. (Wiley, 1999)
30. L.D Landau, E.M. Lifshitz, L.P. Pitaevskii, *Electrodynamics of Continuous Media*, 2nd edn. (English translation published by: Butterworth-Heinemann, Oxford, reprinted 1995)
31. W.P. Dyke, J.K. Trolan, W.W. Dolan, G. Barnes, The field emitter: fabrication, electron microscopy, and electric field calculations. J. Appl. Phys. **24**, 570–576 (1953)
32. J.J. Thomson, Cathode rays. Phil. Mag. **44**, 293–316 (1897)
33. T.E. Stern, B.S. Gossling, R.H. Fowler, Further studies in the emission of electrons from cold metals. Proc. R. Soc. Lond. A **124**, 699–723 (1929)
34. R.H. Fowler, L. Nordheim, Electron emission in intense electric fields. Proc. R. Soc. Lond. A **119**, 173–181 (1928)
35. R.G. Forbes, J.H.B. Deane, Transmission coefficients for the exact triangular barrier: an exact general analytical theory that can replace Fowler & Nordheim's 1928 theory. Proc. R. Soc. Lond. A **467**, 2927–2947 (2011). See electronic supplementary material for details of universal constants used in field emission
36. A. Modinos, Theoretical analysis of field emission data. Solid State Electron. **45**, 809–816 (2001)
37. B. Lepetit, Electronic field emission models beyond the Fowler-Nordheim one. J. Appl. Phys. **122**, 215105 (2017)
38. B. Lepetit, A three-dimensional quantum-mechanical model of electronic field emission from metallic surfaces with nanoscale corrugation. J. Appl. Phys. **125**, 025107 (2019)
39. C. Herring, M.H. Nichols, Thermionic emission. Rev. Mod. Phys. **21**, 185–270 (1949)
40. A. Sommerfeld, Zur Elektronentheorie der Metalle. Naturwiss. **15**, 825–832 (1927)
41. A. Sommerfeld, H. Bethe, Electronentheorie der Metalle, in *Handbuch der Physik*, vol. 24, ed. by H. Geiger, H. Scheel (Berlin, Springer, 1933), pp. 333–675
42. A. Sommerfeld, *Thermodynamics and Statistical Mechanics* (Academic, London, 1964)
43. R.G. Forbes, Comparison of the Lepetit field emission current-density calculations with the Modinos-Forbes uncertainty limits, in 31st International Vacuum Nanoelectronics Conference, Kyoto, July 2018. Technical Digest, pp. 126–127. https://doi.org/10.13140/rg.2.2.17153.10080
44. D. Biswas, R. Kumar, Validation of current formula for a metallic nanotipped emitter. J. Vac. Sci. Technol. B **37**, 049603 (2019)
45. E.O. Popov, A.G. Kolosko, S.V. Fillipov, R.G. Forbes, Emission area extraction for needle-shaped and post-shaped emitters, in 32nd International Vacuum Nanoelectronics Conference & 12th International Vacuum Electron Sources Conference, Cincinnatti, July 2019. Technical Digest, p. 96. https://doi.org/10.13140/rg.2.2.35337.19041
46. I.S. Newton, *Philosophiæ Naturalis Principia Mathematica* (Royal Society, London, 1676) (in Latin) (English edition re-published by: Snowball Publishing, 2010). See Axioms, Law II (p. 19)
47. K.T. Compton, I. Langmuir, Electrical discharges in gases: part I. Survey of fundamental processes. Rev. Mod. Phys. **2**, 124–242 (1930)
48. W. Thomson, On electrical images, Report of the 17th Meeting of the British Association for the Advancement of Science, Oxford, June 1847, Transactions of the Sections pp. 6–7 (John Murray, London, 1848)
49. W. Thomson, On the mathematical theory of electricity in equilibrium: V—effects of electrical influence on internal spherical, and on plane conducting surfaces. Camb. Dublin Math. J. **4**, 276–284 (1849)
50. W. Thomson, *Reprint of Papers on Electricity and Magnetism* (Macmillan, London, 1872)
51. J.C. Maxwell, *A Treatise on Electricity and Magnetism*, 1st edn. (Clarendon, Oxford, 1873)
52. W. Schottky, Über den Einfluss von Strukturwirkungen, besonders der Thomsonschen Bildkft, auf die Electronenemission der Metalle. Phys. Z. **15**, 872–878 (1914). (in German)
53. J.J. Thomson, *Conduction of Electricity through Gases*, 1st edn. (Cambridge University Press, 1903)
54. R.F. Earhart, XI. The sparking distances between plates for small distances. Phil. Mag. **1**, 147–159 (1901)

55. W. Schottky, Über kalte und warme Electronenentladungen. Z. Phys. **14**, 63–106 (1923)
56. L.W. Nordheim, The effect of the image force on the emission and reflexion of electrons by metals. Proc. R. Soc. Lond. A **121**, 626–639 (1928)
57. J.H.B. Deane, R.G. Forbes, The formal derivation of an exact series expansion for the Principal Schottky-Nordheim Barrier Function v, using the Gauss hypergeometric differential equation. J. Phys. A Math. Theor. **41**, 395301 (2008)
58. R.A. Millikan, C.C. Lauritsen, Relations of field-currents to thermionic-currents. Proc. Nat. Acad. Sci. **14**, 45–49 (1928)
59. L.D. Landau, E.M. Lifschitz, *Quantum Mechanics* (Pergamon, Oxford, 1958)
60. R.H. Fowler, The restored theory of metals and thermionic formulae. Proc. R. Soc. Lond. A **117**, 549–552 (1928)
61. E.L. Murphy, R.H. Good, Thermionic emission, field emission and the transition region. Phys. Rev. **102**, 1464–1473 (1956)
62. E.H. Delaval, J. Canton, A letter from John Canton to Benjamin Franklin, containing some remarks on Mr Delaval's electrical experiments. Phil. Trans. **52**, 457–461 (1761–1762)
63. J. Priestly, *The History and Present State of Electricity, with Original Experiments* (1st edition 1775; 5th edition, corrected, published by: Johnson & Rivington, London, 1794) (5th edition reprinted by: ECCO Print Editions)
64. F. Guthrie, *Magnetism and Electricity* (Collins, London, 1876)
65. O.W. Richardson, *The Emission of Electricity from Hot Bodies* (Longmans Green, London, 1st edition 1916, 2nd edition 1921) (2nd edition republished as: *Thermionic Emission from Hot Bodies*, Watchmaker Publishing, online, 2003)
66. A.L. Reimann, *Thermionic Emission* (Chapman & Hall, London, 1934) (Reprinted by: Facsimile Publisher, Delhi, 2015)
67. T.A. Edison, Electrical indicator, US Patent Specification No. 307031 (Filed: 15 Nov. 1883; granted 21 Oct. 1884)
68. G.J. Stoney, Of the "electron", or atom of electricity. Phil. Mag. (Series 5) **38**, 418–420 (1894)
69. S. Dushman, Thermionic Emission. Rev. Mod. Phys. **2**, 381–476 (1930)
70. J.A. Becker, Thermionic electron emission and adsorption: part I. Thermionic emission. Rev. Mod. Phys. **7**, 95–128 (1935)
71. W.B. Nottingham, Thermionic emission, in: S. Flugge (ed.) *Encyclopedia of Physics, Vol. 21: Electron Emission Gas Discharges I* (Springer, Berlin, 1956); based on technical report of same title available on line via websearch
72. S. Yamamato, Fundamental physics of vacuum electron sources. Rept. Prog. Phys. **69**, 181–232 (2006)
73. J.A. Eichmeier, M. Thumm (eds.), *Vacuum Electronics* (Springer, Berlin, 2008)
74. O.W. Richardson, On the negative radiation from hot platinum. Proc. Camb. Phil. Soc. **11**, 286–295 (1900–1902)
75. O.W. Richardson, The emission of electrons from tungsten at high temperatures: an experimental proof that the electric current in metals is carried by electrons. Phil. Mag. **26**, 345–350 (1913)
76. O.W. Richardson, Some applications of the electron theory of matter. Phil. Mag. **23**, 594–627 (1912)
77. S. Dushman, Electron emission from metals as a function of temperature. Phys. Rev. **21**, 623–636 (1923)
78. A. Sommerfeld, Zur Elektronentheorie der Metalle auf Grund der Fermischen Statistik. Z. Phys. **47**, 1–32 (1928). (in German)
79. G.E. Uhlenbeck, S. Goudsmit, Ersetzung der Hypothese vom unmechanischen Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons. Naturwiss. **13**, 953–954 (1925). (in German)
80. R.H. Fowler, The thermionic emission constant *A*. Proc. R. Soc. Lond. A **122**, 36–48 (1929)
81. K.L. Jensen, Exchange-correlation, dipole and image-charge potentials for electron sources: temperature and field variations of the barrier height. J. Appl. Phys. **85**, 2667–2680 (1999)

82. R.G. Forbes, Comments on the continuing widespread and unnecessary use of a defective emission equation in field emission related literature. J. Appl. Phys. **126**, 210191 (2019)
83. R.G. Forbes, Use of energy-space diagrams in free-electron models of field electron emission. Surf. Interface Anal. **36**, 395–401 (2004)
84. E. Guth, C.J. Mullin, Electron emission of metals in electric fields: III. The transition from thermionic to cold emission. Phys. Rev. **61**, 339–348 (1942)
85. R.E. Burgess, H. Kroemer, J.M. Houston, Corrected values of Fowler-Nordheim field emission functions $v(y)$ and $s(y)$. Phys. Rev. **90**, 515 (1953)
86. R.G. Forbes, J.H.B. Deane, Reformulation of the standard theory of Fowler-Nordheim tunnelling and cold field electron emission. Proc. R. Soc. Lond. A **463**, 2907–2927 (2007)
87. R.G. Forbes, Field electron emission theory (October 2016), v2. arXiv:1801.08251v2
88. A. Kyritsakis, J.P. Xanthakis, Extension of the general thermal field equation for nanosized emitters. J. Appl. Phys. **119**, 045303 (2016)
89. A. Kyritsakis, F. Djurabekova, A general computational method for electron emission and thermal effects in field emitting nanotips. Comput. Mater. Sci. **128**, 15–21 (2017)
90. L.W. Swanson, A.E. Bell, Recent advances in field electron microscopy of metals. Adv. Electr. Electron Phys. **32**, 193–309 (1973)
91. A.S. Bahm, G.A. Schwind, L.W. Swanson, Range of validity of field emission equations. J. Vac. Sci. Technol. B **26**, 2080–2084 (2008)
92. K.L. Jensen, M. Cahay, General thermal-field emission equation. Appl. Phys. Lett. **88**, 154105 (2006)
93. K.L. Jensen, General formulation of thermal, field, and photoinduced electron emission. J. Appl. Phys. **102**, 024911 (2007)
94. K.L. Jensen, A reformulated general thermal-field emission equation. J. Appl. Phys. **126**, 065302 (2019)
95. X.Z. Qin, W.L. Wang, N.S. Xu, Z.B. Li, R.G. Forbes, Analytical treatment of cold field electron emission from a nanowall emitter, including quantum confinement effects. Proc. R. Soc. Lond. A **467**, 1029–1051 (2011)
96. A.A. Patterson, A.I. Akinwande, Elementary framework for cold field emission from quantum-confined, non-planar emitters. J. Appl. Phys. **117**, 174311 (2015)
97. P.E. Mason, F. Uhlig, V. Vanek, T. Buttersack, S. Bauerecker, P. Jungwirth, Coulomb explosion during the early stages of the reaction of alkali metals with water. Nat. Chem. **7**, 250–254 (2015)
98. R.G. Forbes, Simple derivation of the formula for Sommerfeld supply density used in electron-emission physics and limitations on its use. J. Vac. Sci. Technol. B **28**, 1326–1329 (2010)
99. R.H. Fowler, *Statistical Mechanics*, 2nd edn. (Cambridge University Press, Cambridge, 1936; re-issued by Cambridge University Press, New York, 2011)
100. N. Fröman, P.O. Fröman, *Physical Problems Solved by the Phase-Integral Method* (Cambridge University Press, Cambridge, 2002)
101. R.G. Forbes, Derivation "in SI units" of the Landau and Lifschitz formula for field ionization rate-constant. arXiv:1412.1821
102. A. Mayer, Exact solutions for the field electron emission achieved from a flat metal surface using the standard Fowler-Nordheim equation with a correction factor that accounts for the electric field, the work function and the Fermi energy of the emitter. J. Vac. Sci. Technol. B **29**, 021803 (2011)
103. N. Fröman, P.O. Fröman, *JWKB Approximation: Contributions to the Theory* (North-Holland, Amsterdam, 1965)
104. R.G. Forbes, On the need for a tunneling pre-factor in Fowler-Nordheim tunneling theory. J. Appl. Phys. **103**, 114911 (2008)
105. E.C. Kemble, A contribution to the theory of the B. W. K. method. Phys. Rev. **48**, 549–561 (1935)
106. E.C. Kemble, *The Fundamental Principles of Quantum Mechanics with Elementary Applications* (Dover, New York, 1937)

107. S.C. Miller, R.H. Good, A WKB-type approximation to the Schrödinger equation. Phys. Rev. **91**, 174–179 (1953)
108. M. Razavy, *Quantum Theory of Tunnelling* (World Scientific, Hackensack, 2003)
109. R.G. Forbes, Development of a simple quantitative test for lack of field emission orthodoxy. Proc. R. Soc. Lond. A **469**, 20130271 (2013)
110. R.G. Forbes, The Murphy-Good plot: a better method of analysing field emission data. Roy. Soc. Open. Sci. **6**, 190212 (2019)
111. R.G. Forbes, Field emission: new theory for the derivation of emission area from a Fowler-Nordheim plot. J. Vac. Sci. Technol. B **17**, 525–533 (1999)
112. See p. 357 in Reference [99]
113. W.P. Dyke, W.W. Dolan, Field emission. Adv. Electr. Electron Phys. **8**, 89–185 (1956)
114. R.G. Forbes, A. Fischer, M.S. Mousa, Improved approach to Fowler-Nordheim plot analysis. J. Vac. Sci. Technol. B **31**, 02B103 (2013)
115. R.G. Forbes, Simple good approximations for the special elliptic functions in standard Fowler-Nordheim tunnelling theory for a Schottky-Nordheim barrier. Appl. Phys. Lett. **89**, 113122 (2006)
116. R.G. Forbes, Use of a spreadsheet for Fowler-Nordheim equation calculations. J. Vac. Sci. Technol. B **17**, 534–541 (1999)
117. R.G. Forbes, J.H.B. Deane, Comparison of approximations for the principal Schottky-Nordheim barrier function $v(f)$, and comments on Fowler-Nordheim plots. J. Vac. Sci. Technol. B **28**, C2A33–C2A42 (2010)
118. B.C. Carlson, Elliptic integrals, in *NIST Handbook of Mathematical Functions*, ed. by F.W.J. Olver et al. (NIST & Cambridge University Press, 2010), see formula 19.6.1
119. R.G. Forbes, ISQ derivation of the Murphy-Good second-order Taylor-expansion correction factor $t_2(f)$, in 32nd International Vacuum Nanoelectronics Conference & 12th International Vacuum Electron Sources Conference, Cincinnatti, July 2019. Technical Digest, p. 101
120. R.G. Forbes, C.J. Edgcombe, U. Valdré, Some comments on models for field enhancement. Ultramicroscopy **95**, 57–65 (2003)
121. A.I. Zhbanov, E.G. Pogolerov, Y.-C. Chang, Y.-G. Lee, Screened field enhancement factor for the floating sphere model of a carbon nanotube array. J. Appl. Phys. **110**, 114311 (2011)
122. J.R. Harris, K.L. Jensen, D.A. Shiffler, Dependence of optimal spacing on applied field in ungated field emitter arrays. AIP Adv. **5**, 087183 (2015)
123. R.G. Forbes, Physical electrostatics of small field emitter arrays/clusters. J. Appl. Phys. **120**, l054302 (2016)
124. T. de Assis, F.F. Dall'Agnol, Evidence of universal inverse-third power law for the shielding-induced fractional decreasee in apex field enhancement factor at large spacings: a response via accurate Laplace-type calculations. J. Phys.: Condens. Matter **30**, 195301 (2018)
125. D. Biswas, R. Rudra, Shielding effects in random large-area emitters, the field enhancement factor distribution, and current calculation. Physics Plasmas **26**, 083105 (2018)
126. E. Minoux, O. Groening, K.B.K. Teo et al., Achieving high-current carbon nanotube emitters. Nano Lett. **5**, 2135 (2005)
127. R.G. Forbes, Field emission: the theoretical link between voltage loss, reduction in field enhancement factor, and Fowler-Nordheim-plot saturation. Appl. Phys. Lett. **110**, 133109 (2017)
128. R.G. Forbes, Use of the concept "area efficiency of emission" in equations describing field emission from large area electron sources. J. Vac. Sci. Technol. **B27**, 1200–1203 (2009)
129. R.G. Forbes, Extraction of emission parameters for large-area field emitters, using a technically complete Fowler-Nordheim-type equation. Nanotechnology **23**, 095706 (2012)
130. R.G. Forbes, Why converting field emission voltages to macroscopic fields before making a Fowler-Nordheim plot has often led to spurious characterization results. J. Vac. Sci. Technol. B **37**, 051802 (2019)
131. M.M. Allaham, R.G. Forbes, M.S. Mousa, Applying the field emission orthodoxy test to Murphy-Good plots. Jordan J. Phys. 13, 101–110 (2020)

132. M.M. Allaham, R.G. Forbes, A. Knápek, M.S. Mousa, Implementation of the orthodoxy test as a validity check on experimental field emission data. J. Electr. Eng. Slovak 71, 37–42 (2020)
133. M. Bachmann, F. Dams, F. Düsberg, M. Hofmann, A. Pahlke, Extraction of the characteristics of current-limiting elements from field emission measurement data. J. Vac. Sci. Technol. B **35**, 02C103 (2017)

Chapter 10
Carbon-Based Field Emitters: Properties and Applications

Nikolay V. Egorov and Evgeny P. Sheshin

Abstract Field emission is one of the most promising areas of vacuum nano- and microelectronics. This chapter reviews current achievements and problems in studying carbon-based field emitters. It is a detailed elaboration of topics partly addressed in the authors' textbook "Field Emission Electronics". The review is multipartite. The first part gives general information about carbon-based materials, describing their structures, manufacturing methods for field emission cathodes, and basic features. The second part considers and analyzes some special modern electron emission theory that applies to field emission from carbon-based materials. This is followed by results on the field emission properties of different forms of carbon-based materials. The last part features applications and technological devices using carbon-based field emission cathodes, including descriptions of devices and their design, and prospects for further development. Here, the authors pay particular attention to their own papers on these topics.

10.1 Introduction: General Information on Carbon-Based Materials

Development of stable field electron emission (FE) cathodes is one of the most pressing problems of today's electronics. Such cathodes have many advantages in comparison with other kinds of free-electron emitters: thermal stability, exponential slope of current–voltage characteristics, low sensitivity to external radiation, etc. The materials used to manufacture FE cathodes have been diverse, including (a) refractory metals such as tungsten, molybdenum, rhenium, and platinum, (b) transition metals such as chromium, niobium, and hafnium, and (c) semiconductor materials. However,

N. V. Egorov (✉)
Saint Petersburg State University, St. Petersburg, Russia
e-mail: n.v.egorov@spbu.ru

E. P. Sheshin
Moscow Institute of Physics and Technology, Dolgoprudny, Russia
e-mail: sheshin.ep@mipt.ru

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*,
Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_10

the materials for FE cathodes also need to have specific properties, such as low and stable electronic work-function values, and high values of mechanical robustness and of electrical and thermal conductivity [1, 2]. Soon after the first reports on FE from carbon-based materials appeared in the early 1970s, their advantages became quite obvious. This article considers the most popular and well-known carbon-based materials used in FE cathodes.

The great variety of carbon-based materials includes many able to function as FE cathodes. A classification of such materials is shown in Fig. 10.1.

Three different combinations of σ- and π-bonds result in three electronic states for carbon atoms: (i) sp^3 hybridization with a tetrahedral arrangement of four σ-bonds, established as a result of interaction between one s-electron and three p-electrons (this corresponds to the ideal diamond structure); (ii) sp^2 hybridization characterized by three σ-bonds, and a π-bond that is localized in the plane perpendicular to the σ-bonds (this state corresponds to the graphite structure); and (iii) sp^2 hybridization with two σ-bonds and two π-bonds (this corresponds to the so-called carbine structure, with linear polymeric chains of type "—C = C—C = C—" or type " = C = C = C = ").

Research and quantitative assessment of the distribution of these bonds can be quite fruitful when studying the origins of the properties of different carbon-based materials.

In the sp^2 hybridization state, carbon atoms form layered structures. A layer (base-plane) consists of a continuous series of regular hexagons with carbon atoms

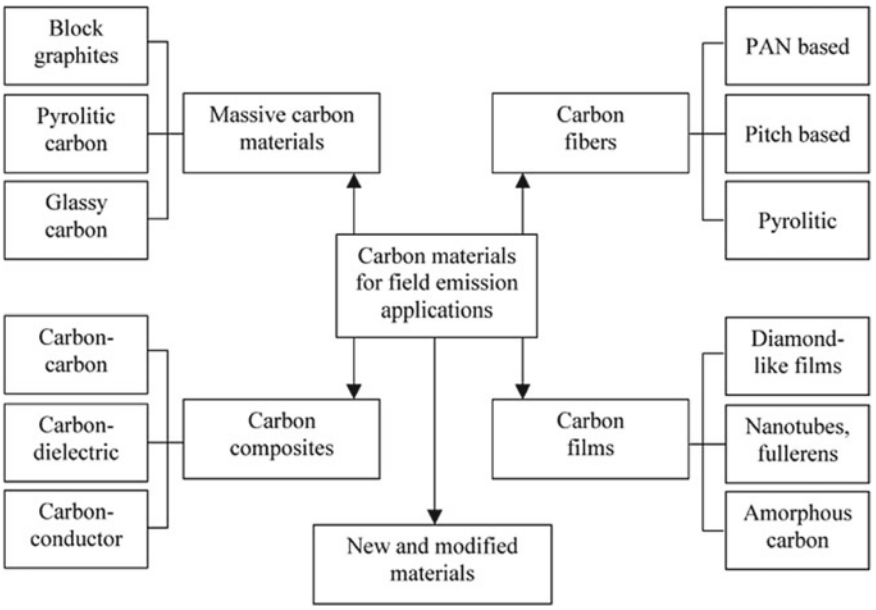


Fig. 10.1 Classification of the types of carbon material that are of interest in the production of field emission cathodes (compare [3])

at the vertices. The shortest distance between the atoms in the plane is equal to the hexagon side and amounts to 1.417 Å. Graphite consists of a continuous series of layers parallel to the base-plane. The carbon atoms in this layer are connected with three σ-bonds of equal value. Additional bonds are established by π-electrons with somewhat overlapping orbitals. Collectivization of π-electrons in the graphite layer gives a metallic nature to its electrical and optical properties. Different sources range the binding energy between the carbon atoms in the plane from 340 to 420 kJ/mol (3.5 to 4.4 eV), while the binding energy between the layers does not exceed 42 to 84 kJ/mol (0.44 to 0.87 eV) (see [4]).

The sequence pattern defines the type of ideal crystal lattice of graphite, i.e., hexagonal or rhombohedral (Fig. 10.2).

When the layers alternate ab, ab, ab, carbon atoms in each layer settle down only over the centers of regular hexagons in the next layer. This is a hexagonal structure with four carbon atoms in a unit cell. The interlayer distance in ideal graphite is 3.354 Å. The unit cell of a graphite hexagonal structure is a right-angle prism with a regular rhombus in its base.

The abc, abc layer sequence results in a rhombohedral lattice of graphite. Unlike the hexagonal one, it is three-layered. At temperatures above 2000 °C, the rhombohedral packing turns into the hexagonal one. Both structures of ideal graphite should be considered as infinite grids of hexagons located in parallel layers. In practical terms, these grids are finite in man-made graphites.

Real graphites differ from ideal structures by having different defects. Failures in the sequence of layers cause packing defects. A large number of such defects result

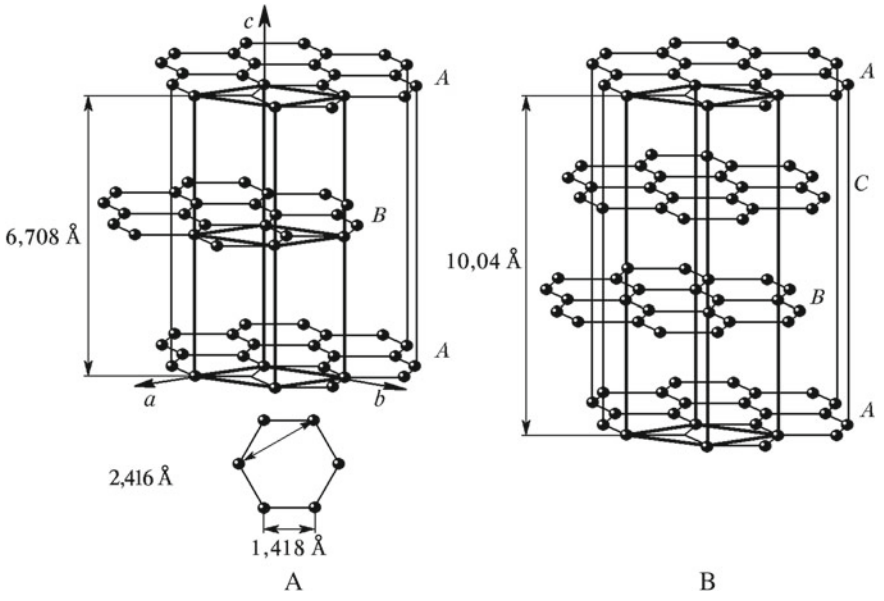


Fig. 10.2 The structures of hexagonal (left) and rhombohedral (right) graphite

in a complete disorder in terms of vertical configuration, though the parallelism of layers is preserved. This kind of defective structure is usually referred to as a *turbostratic* structure. The interlayer distance in the turbostratic structure is 3.44 Å.

The second type of defects is those in the lattice bonds. Such defects can be caused by the presence of foreign atoms, such as hydrogen, oxygen, and nitrogen, or of carbon atoms in different valence states. Defects in the graphite structure can also arise when alien elements enter the interlayer space. In sufficiently high concentration, they cause development of interstitial compounds.

Carbon-based materials are divided into graphitizing and non-graphitizing ones. Particularly important for the process of graphitization is structural anisotropy, i.e., relative positioning of the basic planes in starting materials. For graphitizing materials, the three-dimensional regularity of carbon layers (graphitization) emerges at temperatures of 1600–1800 °C. For example, carbon composites with non-carbon structures can be formed during graphitization of the surface of silicon carbide, monocrystals of which (4H, 6H poly-types) have exceptional emission characteristics (see Sect. 10.5.5).

Properties of a single graphite crystal are radically different along the a (in-plane) and c (normal to the plane) axes. This is caused by the layered structure of the crystal lattice. Development of anisotropy in polycrystalline graphites is based on specific features of the technology of fabrication. Thus, in compression molding, the particles become mostly oriented in the direction of the molding axis; in extrusion, the crystallographic axis is mostly perpendicular to the axis of extrusion. Glassy carbon and fullerenes are characterized by nearly total absence of anisotropy, while acrylic carbon fiber has a pronounced axial texture, i.e., the crystallographic axis is perpendicular to the fiber axis.

Basic properties of carbon-based materials to be taken into account when using them as FE cathodes include those related to adsorption. The high ability of carbon-based materials to adsorb various substances from gases and solutions onto their surface is exploited in the activated carbon process. Man-made carbon-based materials are expressly porous because their production process is connected with some mass loss and compaction of the structure, which results in shrinkage and cracking. There exists a convenient classification of carbon-based materials in terms of average width of their pores; this was proposed by M.M. Dubinin and developed in later papers dedicated to sorption properties of carbon-based materials [5]. Pores that are less than 20 Å wide are referred to as micropores; pores more than 200 Å wide are classified as macropores; and those in the range 20–200 Å are called intermediate pores.

The mechanical properties of carbon-based materials depend on the starting raw materials and the production-process parameters. In the temperature range of 20–2000 °C, graphites have limited plasticity, and are fragile as far as their fracture behavior is concerned. One should bear in mind that the larger the diameter of the blank, the higher is the coefficient of variation of the mechanical properties of the blank, and that the larger the grain size, the lower is the material strength.

Qualitative descriptions of electrical and electronic properties of different carbon-based materials can be given within the framework of a band model. The changes

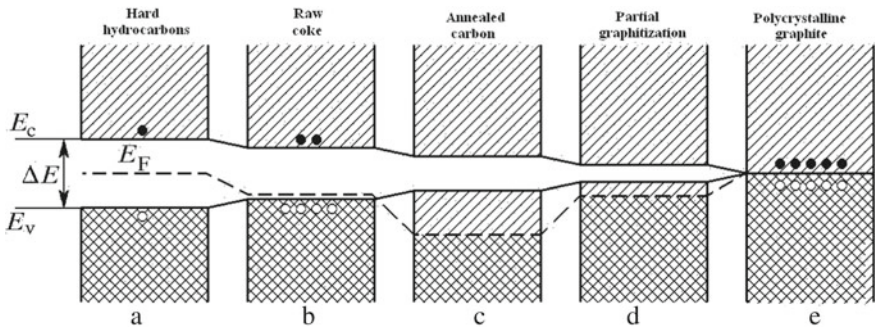


Fig. 10.3 **a**–**e** Diagrams showing the electron-energy bands and Fermi level positions for various carbon-based materials. E_v labels the valence-band maximum, and E_c the conduction band minimum. ΔE is the band gap, E_F labels the Fermi level

in electronic properties of carbon-based materials can be evaluated following the diagram shown in Fig. 10.3.

Figure 10.3 represents stages of transition—from solid aromatic hydrocarbons bound by van der Waals forces, to carbon-based materials of various types, through highly defective structures (coke), to almost ideal graphite. With the rise in the temperature of processing, the width of the band gap between the conduction and valence bands decreases, and becomes zero in the extreme case of infinitely large crystals of graphite, in which case the conduction and valence bands touch.

The valence band of carbon-based materials is formed by π-electrons of organized carbon-atom grids of a *macro-aromatic* nature. In the case of ideal graphite (Fig. 10.3a), the conduction band where there are no electrons at 0 °K is separated from the π-electron band by a negligibly small band gap. As a result of band adjacency and easy thermal excitation into the conduction band of ideal graphite, the number of electrons in the carbon conduction band is large enough for graphite to effectively behave like a metal.

Different stages of transmutations from aromatic hydrocarbons through intermediate forms of carbon to crystal graphite can be represented as a consecutive reduction of width of the band gap. In the temperature range of up to 1000 °C (Fig. 10.3b), cross-linking of the grids and emission of hydrogen during carbonization results in the emergence of holes in the π-band. The vacant free valences function as electronic traps localized on the defects (e.g., on the boundary atoms of the broken lateral chains of carbon and carbon layers).

Though partial filling of electronic bands should explain the *metallic* nature of the conductance of these carbon-based material types, they show a positive temperature coefficient of electrical conductivity. That is explained by the hole nature of conductance (or scattering) at the inter-grid boundary. With an increase in hole concentration, the lower band is gradually exhausted. It seems that at temperatures above 1400 °C (Fig. 10.3c), hole development due to hydrogen emission is mostly blocked. Cross-linking of broken grids taking place when the crystals grow reduces the number of hole defects acting as electronic traps. As this takes place, the π-band begins to

refill. At the same time, dimensional growth of carbon grids causes reduction of the band-gap width ΔE. At a temperature of 2000 °C (Fig. 10.3d), this gap can be considered narrow enough for electrons driven by thermal excitation to pass into the conduction band. Therefore, unlike crystal graphite, which has an essentially metallic conductivity mechanism, the transition carbon forms are organic semiconductors with electrical properties determined by delocalized π-electrons.

We shall now describe the structural and FE properties of the above materials, as far as they have been studied up till now.

10.2 Carbon-Based Material Structures

10.2.1 Graphene

Graphene is a layer or "sheet" of carbon atoms, linked with sp^2 bonds and forming a hexagonal two-dimensional crystal lattice (see Fig. 10.4). It can be thought of as a graphite layer separated from the bulk crystal. Graphene has relatively high mechanical stiffness and good thermal conductivity. High mobility of current carriers at room temperature makes it a promising material for various applications. *Inter alia* can be used as a basis for nanoelectronics and a possible replacement for silicon in integrated microcircuits. At present, graphene is mostly produced by mechanical cleavage or exfoliation of graphite layers. This method produces the best samples, with high carrier mobility. It is not useful for large-scale production, however, because it is still very much a manual procedure. An approach much more suitable for industrial production is based on thermal decomposition of a silicon carbide substrate.

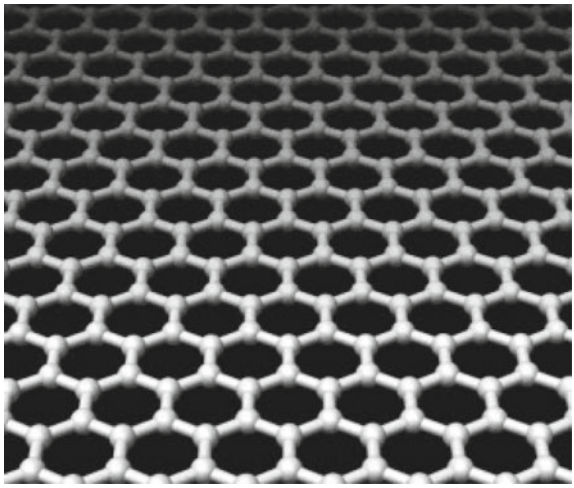


Fig. 10.4 Crystallographic structure of a graphene sheet

Because graphene was first produced as late as 2004, it has still not been studied in full depth, and thus generates keen interest. It is incorrect to view this material just as a piece of another allotropic carbon crystal such as graphite or diamond, because specific features of the energy spectrum of carriers in graphene make its electro-physical properties different from those of other two-dimensional systems.

Graphene nanoribbons are narrow graphene strips about 10–100 nm wide. Their physical properties differ from those of wider samples that follow the linear dispersion law, as an infinite graphene layer would do. Nanoribbons are particularly interesting because they follow a non-linear dispersion law, and have semiconductor properties—due to the presence of a band gap that depends on ribbon width and on the arrangement of atoms at its boundaries. This is why graphene nanoribbons are regarded as an important step toward making a graphene-based transistor working at room temperature. Graphene is also being considered as a possible base for the construction of a ballistic transistor. In 2006, a research group at the Georgia Institute of Technology around Walt de Heer announced that they had produced a graphene-based field transistor as well as a quantum interference device [6]. The researchers believe that their achievements can eventually develop into a new class of graphene-based nano-electronics, with the basic transistor thickness reduced to as little as 10 nm. Additionally, graphene can be used as a highly sensitive sensor to detect particular molecules of chemical substances attached to the film surface. Another promising application of graphene is to build electrodes in ionistors (supercapacitors), a promising class of rechargeable electricity sources. Despite all its obvious potential, the field emission properties of graphene have not yet been extensively studied.

10.2.2 Carbon Fibers

Depending on the starting raw material, carbon fibers can be divided into acrylic, pitch-based, and pyrocarbon ones.

The fibers of the first two types are manufactured commercially and their mechanical properties span a wide range in terms of values of their elastic modulus (200–700 GPa) and tensile strength (2–4 GPa). Here, we shortly recall relevant definitions: the elastic modulus is given by the applied tensile force per unit area (=tensile stress σ) divided by the relative elongation (=strain $= \Delta L/L_0 = \varepsilon$) and is given as the slope of the linear part of the stress/strain diagram (Hooke's law). The *ultimate tensile strength* is the maximum tensile stress of the stress/strain curve, i.e., the tensile strength at which the fiber breaks

Pyrocarbon fibers (pyros) have not been widely used in industry yet, and as yet there is little precise information about their behavior.

In terms of their mechanical characteristics, the acrylic carbon fiber family can be divided into *high-tenacity* (HT) and *high-modulus* (HM) ones. Several incompatible definitions of "tenacity" exist in the literature. In our view, the most appropriate definition of tenacity is "ultimate tensile strength, divided by the fiber density". For

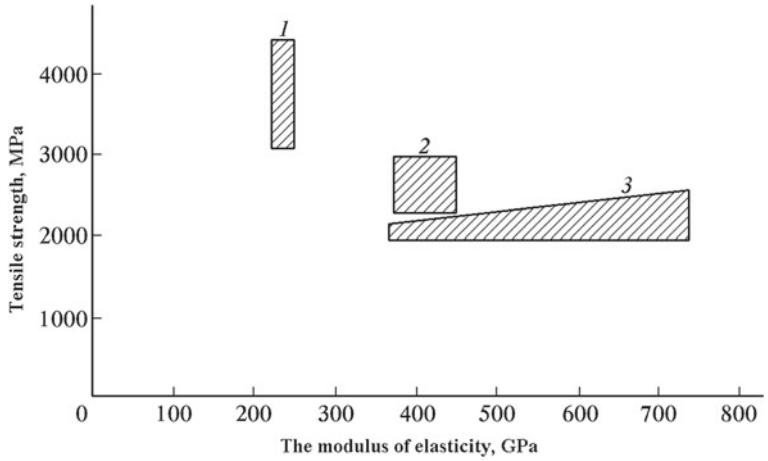


Fig. 10.5 Approximate ranges for the tensile strengths and elastic moduli, for carbon fibers of different types: 1—high-tenacity (HT) polyacrylonitrile-carbon fiber; 2—high-modulus (HM) polyacrylonitrile-carbon fiber; 3—mesophase pitch carbon fiber (see [3, 7])

"tenacity" defined in this way, the SI unit is usually stated as Pa m^3/kg or N m/kg. The high-tenacity carbon fibers are manufactured at temperatures below 1400 °C, high-modulus ones at temperatures of 1800–3200 °C. Typical ranges for the tensile strengths and elastic moduli of acrylic carbon fibers and pitch-based carbon fibers are shown in Fig. 10.5 [3, 4, 7].

The properties of carbon fibers, as with other graphite-based materials, are determined by their structure. The elastic modulus of a material is connected with interatomic forces. Producing high-modulus carbon fibers is feasible because of the extremely high interaction energy between carbon atoms in their hexagonally structured base-planes. Graphite tenacity along the base-plane is determined by the interaction between carbon atoms. The tensile strength of graphite in the perpendicular direction is much lower, because the interaction between the atoms of adjacent layers is weak. High thermal and electrical conductivity of graphite in the atomic-layer plane is caused by delocalization of valency-band p-electrons between the layers. Note that a graphite monocrystal should not be described as a metallic conductor, but rather as an overlapping-orbital semiconductor.

The greatest stiffness and tenacity of carbon fibers is reached only when the crystal-lattice layers are exactly parallel to the fiber axis.

Elastic moduli for a graphite monocrystal have been measured with great precision [8]. Figure 10.6 illustrates the three main coefficients relating to elastic properties: the Young's modulus under tension in the C_{11} carbon-layer plane, the Young's modulus under tension in the C_{33} orthogonal direction, and the C_{44} shear modulus.

The maximum possible value of Young's modulus is 1060 GPa, but this value can only be reached in the case of a defectless crystal structure with the atomic planes oriented exactly along the fiber axis. The elastic modulus of fibers in the orthogonal

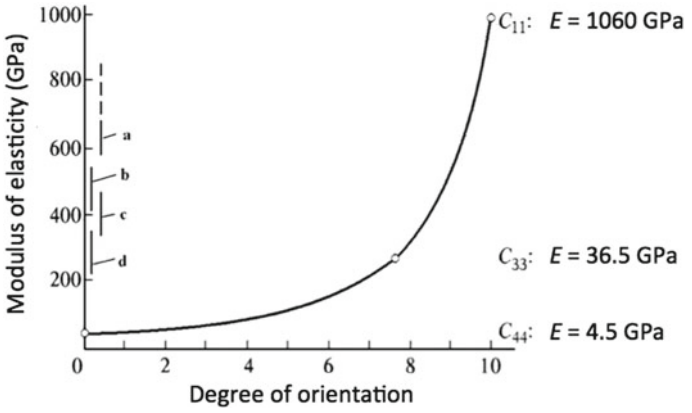


Fig. 10.6 **a** Modulus-of-elasticity values for various carbon-based materials, and the influence of the degree of orientation of atomic layers: a—graphite fibers (based on mesophase pitch); **b** carbon fiber (based on mesophase pitch); c—high-temperature PAN carbon fiber; d—low-temperature PAN carbon fiber. The curve plotted is Young's modulus E versus the degree of orientation for a graphite monocrystal based on an equation given by B. Kelly [8]

direction is an order of magnitude smaller. The minimum value (usually about 4.5 GPa) is that of the shear modulus. Fiber tenacity is directly proportional to the fraction of atomic layers oriented along the fiber axis. Disorientation of atomic planes decreases tenacity along with a decrease in the actual value of the elastic modulus. The theoretical tenacities of high-tenacity and high-modulus fibers are 70 and 25 GPa, respectively. For high-modulus fibers, the tenacity can only be implemented up to 3% of theoretical estimates, i.e., 2.1 GPa, and even the best high-tenacity fibers don't reach more than 5 GPa, which is 2% of theoretical maximum. Figure 10.6 shows the dependence of the experimental values of the fiber elasticity modulus on the degree of orientation of the carbon-atom layers. In the best industrial samples of fibers, the elastic modulus reaches 70% of the theoretically possible value.

Despite high-tenacity in the plane of an atomic layer, the crystal can be easily bent or curved, due to the low shear modulus. An interlayer shear becomes more difficult with the introduction of defects into the crystal lattice. Defects within a layer, such as vacancies, insertions, disclinations, as well as layer-packing defects, increase the interlayer distance. Layer and interlayer defects are often interconnected, because defects inside a layer can lead to improper packing of the adjacent layers and result in an increase in the interlayer distance; this causes incomplete delocalization of electrons and impedes interlayer shear. Disclinations also interfere with the shear and lead to the emergence of vacancies and improper packing of adjacent layers.

The occurrence of disclinations and the degree of carbon-layer orientation are mainly determined by the structure of the starting fiber. Additionally, these carbon fiber characteristics depend on the draw-down rate at the initial stage of fiber production. The inside-layer defects and the layer-packing defects are largely eliminated during heat treatment.

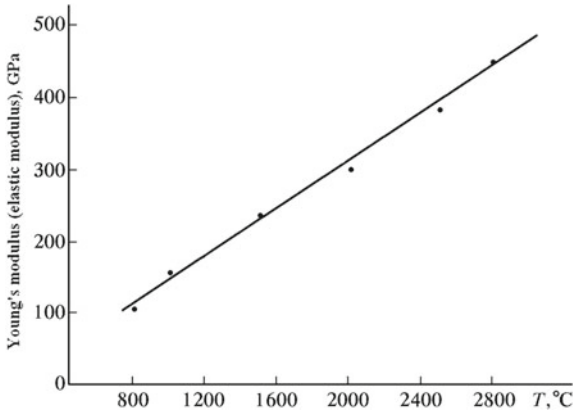


Fig. 10.7 The dependence of Young's modulus on the temperature of heat treatment of polyacrylonitrile ("PAN") carbon fibers

For industrial fibers, the elastic modulus is a stable characteristic because it is determined by the final processing temperature and the draw-down ratio, and these technological parameters are well reproducible. Figure 10.7 illustrates, for PAN fibers, how the fiber's structural properties depend on the processing temperature. Production of fibers with stable tenacity is much more difficult, because tenacity depends on the presence of cracks and other macrodefects, and these are largely decided by the properties of the raw fiber. The use of chemically pure acrylic carbon fibers enables chemically pure and homogeneous carbon fibers to be produced.

Most important for field emission are the properties of carbon fiber surfaces, as fully described by means of

(1) The surface-layer composition, i.e., the functional group types present and their concentration, as well as the presence of impurity atoms;
(2) Surface energy per unit area, and its changes when gaseous substances are adsorbed;
(3) Specific surface geometry, roughness, microporosity;
(4) Number of surface cracks and their size distribution;
(5) Shape and size of the filament (elementary fiber), its structure, and the surface anisotropy.

Because of the great fiber anisotropy, and the related great variation in surface electrostatic field when the fiber is charged overall, the fiber end-surface field emits most strongly. This fiber end-surface is a set of chaotically located micro-elevations of different heights, radii of curvature, and configurations. When an electrical voltage is applied, the numerous micro-elevations are sure to include a few with a relatively high electric field at their vertices. These micro-elevations are the initial emission centers when voltage is applied. With further increase in voltage, the electric field becomes sufficient for emission from other, less sharply pointed, micro-elevations.

The initial distribution of emitting centers over the fiber end-surface is absolutely arbitrary, and does not coincide even for quite closely located sites of the same

filament. This is primarily due to the variability of mechanical properties along the fiber axis.

10.2.2.1 Acrylic Carbon Fibers

Acrylic carbon fibers are produced by pyrolysis of polymeric polyacrylonitrile (PAN) fibers and subsequent high-temperature processing. These carbon fibers have a filament diameter of 6–10 μm. They are about 99.9% pure carbon, and are normally described as "chemically pure". The technology of carbon fiber production from a synthetic PAN fiber includes the following operations [9–11].

(1) Oxidization of the PAN fiber in atmospheric oxygen for several hours at temperatures ranging between 200 and 300 °C, with the fiber held in a state of tension in order to prevent shrinkage and drawing of the fiber. As this oxidation takes place, polymeric molecules turn into six-link rings, oriented along the fiber axis and containing carbon and nitrogen. This operation increases the elastic modulus and the fiber tenacity.
(2) Carbonization of the oxidized fiber, without tension, at a temperature of up to 1000 °C, in an inert atmosphere.
(3) Graphitization in the inert atmosphere, at temperatures up to 3200 °C.

If graphitization is carried out at processing temperatures of 1000–1700 °C, then high-tenacity fibers are produced [4]. But most interesting for field emission is the high-modulus fiber type fabricated by using processing temperatures up to 3200 °C, which is referred to as a *graphite fiber*. Figure 10.7 shows how Young's modulus value depends on the final processing temperature. Conspicuous is the fact that this dependence is linear.

The elastic modulus of the fiber increases with an increase in the degree of orientation of the structure (see Fig. 10.7). At the same time, the specific electrical resistance or electrical resistivity ρ of the fiber decreases (see Fig. 10.8). Figure 10.8 also shows that structural features of the fiber change. Each fiber consists of many, much smaller, structural components called *microfibrils*. The typical dimensions L_a and L_c of these microfibrils (in directions parallel to the "*a*-" and "*c*-" crystallographic axes in graphite, respectively) depend on the production conditions and temperature, and range from $L_a = 10$ to 50 nm, and $L_c = 2$ to 5 nm. The fiber density d also changes.

It is worth noting that the curves in this diagram only show representative qualitative tendencies, because each carbon-fiber batch has its own values of the specified parameters, depending on the starting acrylic carbon fiber structure and the carbon-fiber production technology. Figure 10.9 represents three fiber structure types [3, 11] identified so far, which differ in the orientation of the graphite crystal grains in the fiber cross-section.

Structures (a) and (b) are double-band, while structure (c) is single-band. Structure (a) has preferential orientation at the circumference of the surface crystal grains of the fiber, with the inner part having chaotic orientation. It is mostly generated by amorphous carbon.

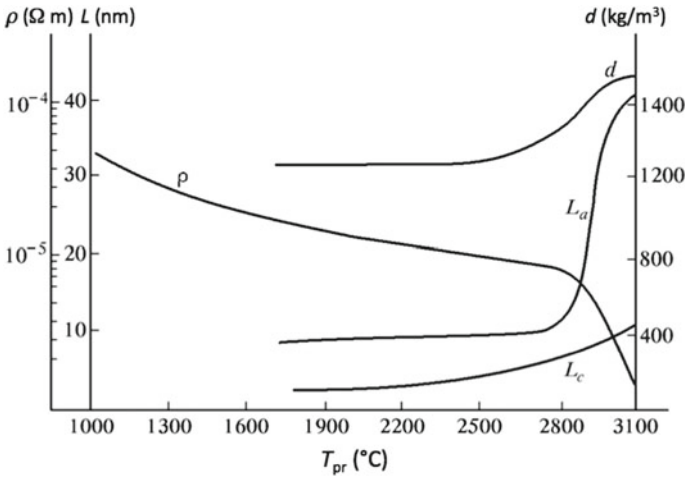


Fig. 10.8 Representative changes in the structural parameters of graphite-type carbon fibers, as the heat treatment temperature T_{pr} increases: ρ—electrical resistivity; d—density. L_a and L_c are the average values of crystallite ("microfibril") dimensions in the directions of the a-axis and c-axis in graphite, respectively

In the second type of structure, (b), the surface layers of crystal grains are mostly oriented at the fiber circumference, the neutral band displaying radial orientation.

The third type of structure is characterized by the cylindrical orientation of crystal grain c-axes over the entire fiber cross-section. It is worth noting that the supramolecular structure of the starting fiber is inherited by the processed carbon fiber and is almost independent of the final processing temperature.

To evaluate the effect of fiber characteristics on field emission, it is necessary to have a clear idea of the carbon-fiber structure and of any specific features of its emitting surface.

The earliest models represented a fiber as a conglomerate of interconnected highly oriented fibrils of sub micrometer size. After heat treatment at 2000 °C, the high-modulus carbon fibers have been established to contain at least two phases, both consisting of fibrils with graphite insertions. One phase consists of wider and thicker ribbons with increased longitudinal orientation. The other phase is built from narrower, less oriented, and intensely intertangled fibril ribbons. Both phases contain a large number of pores, micro-cracks, and other defects. In the central part of the fiber, fibril orientation is usually less exhibited; fibrils with fewer defects are found at the outside. Between these two regions, there is an intermediate layer containing ribbons of both types. The up-to-date fiber micro-crack model (for the structure shown in Fig. 10.9b) is illustrated in Fig. 10.10 (compare [12]).

The fiber surface is its best-oriented part. It contains both exposed crystallite boundaries and defectless basic atomic planes. Highly energetic boundary atoms are linked with each other by sp^2-bonds and are able to chemosorb oxygen in an active way. The base-plane surface atoms forming the structure of an ideal graphite lattice

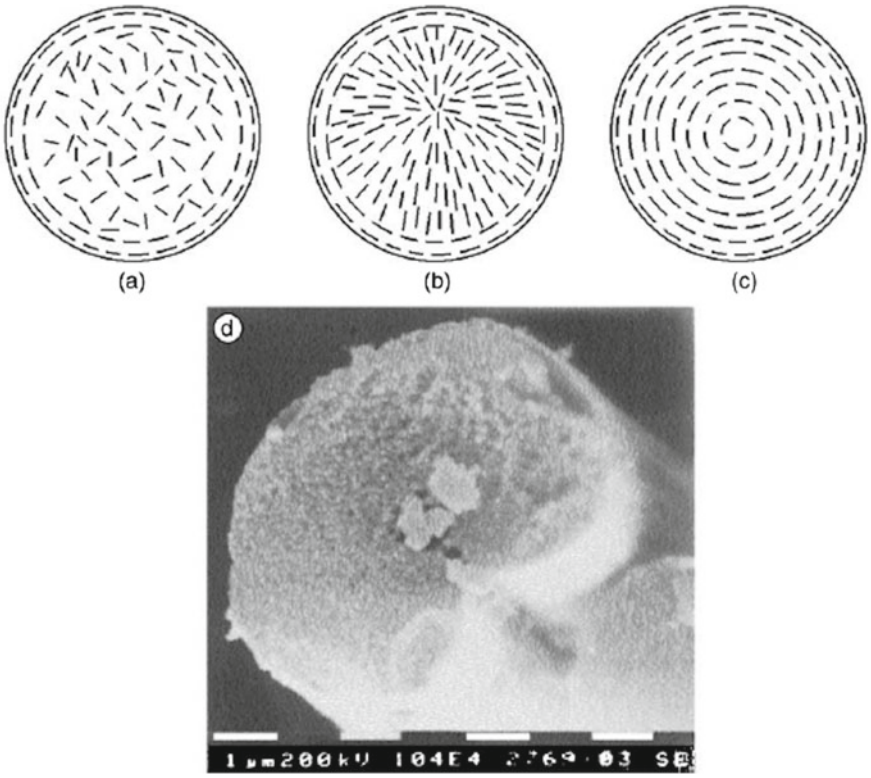


Fig. 10.9 Schematic illustration of the internal structure of PAN fibers: **a** isotropic center and oriented surface layer; **b** different orientation of the center and the surface layers; **c** one type of preferred orientation; **d** photograph of the working surface of the carbon fiber type "VULON"; from [3]: E.P. Sheshin, Properties of carbon materials, especially fibers, for field emitter applications, Appl. Surf. Sci. **215**, 191–200 (2003), with permission of Elsevier

interact with the weakest σ-bonds, and have considerably lower surface activity. In addition, the fiber surface can have a fibril-like microstructure, micropores, crystallite boundaries, foreign insertions, and cracks, making it susceptible to mechanical destruction. To produce high-quality carbon fibers, three requirements have to be met:

(1) reduction in the fiber diameter, to decrease the probability of serious structural defects;
(2) use of high-quality, chemically pure polymeric raw materials; and
(3) high reproducibility of technological processes, especially at the oxidation stage.

At this point in time, the most extensive studies have been on the FE properties of acrylic carbon fibers. FE cathodes made from such fibers are promising for some practical applications.

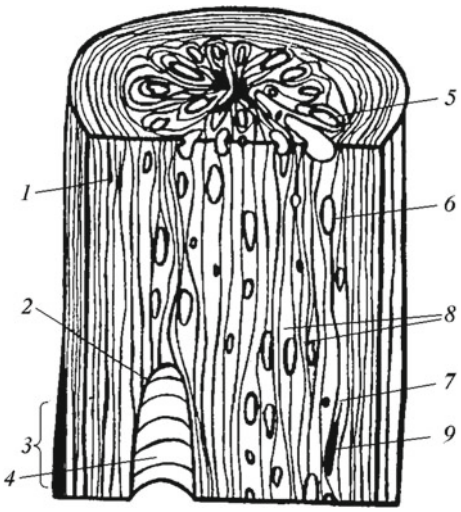


Fig. 10.10 The structure of polyacrylonitrile (PAN) carbon fibers: 1—micro-cracks; 2—shell cracks and pores; 3—convexity on the fiber; 4—big pore; 5—small pore; 6—micro-inclusions of crystalline graphite; 8—interfibril layers in the middle part of the fiber; 9—micro-crack; from [12]: E. P. Sheshin, Field emission of carbon fibers, Ultramicroscopy **79**, 101–108 (1999); with permission of Elsevier

10.2.2.2 Pitch-Based Carbon Fibers

The production technology for pitch-based carbon fibers consists of five basic stages: (1) pitch preparation; (2) fiber spinning/formation; (3) stabilization through oxidation; (4) carbonization; and (5) graphitization (to produce high-modulus carbon fibers).

To produce the pitch of the desired quality as the starting material, it is necessary to increase its aromatization, molecular mass, and melting point. An additional step in pitch preparation is to turn it into the so-called mesophase pitch, which is characterized by spinnability.

Pitch having the above qualities is heated at 350–500 °C in a protective atmosphere for a prolonged period. Heating causes a series of transmutations and results in the development of polycyclic large molecules, which come together in the liquid-crystal phase, i.e., the mesophase. This phase has a larger surface tension than the isotropic phase (with low relative molecular mass) from which the mesophase grows. The mesophase consists of large flat molecules oriented parallel to each other in such a way as to form spherical crystals (spheroids); these grow and coagulate into large spheres, resulting in a bulk mesophase. Eventually, the mesophase pitch has a heterogeneous structure consisting of an anisotropic mesophase and isotropic regions. There are many detailed production methods for mesophase pitch, but exact technological details are most often kept as corporate or industrial secrets.

To produce fibers from the pitch, the prepared pitch is drawn through spinnerets. The wet process involves a special coagulation bath, where a bundle of up to 320,000 elementary fibers is formed. In the dry method, the number of elementary fibers produced is smaller, although the fiberization rate is higher. Drawing can be done by centrifugation, structural processes, or appropriate pitch melting.

Spinneret orifice configurations can be widely different. Together with drawing-mode variations, these differences yield fibers with many different structures. For example, the radial carbon fibers are produced as a result of the laminar flow of the mesophase pitch through spinnerets. Cross-section models of the main types of pitch-based carbon fiber are shown in Fig. 10.11.

Table 10.1 shows structural parameters, found by X-ray diffraction, for some pitch-based carbon fibers. The fibers are identified by their "brand" names/codes. The table shows that the carbon fibers made from isotropic pitch have lower values of the fibril parameters L_a, L_c, and higher values of the parameter d_{002} than mesophase pitch-based carbon fibers. (L_a and L_c have been explained when discussing Fig. 10.8; the crystallographic parameter d_{002} is equal to half the interlayer spacing between the graphene sheets in graphite). In the course of graphitization of isotropic pitch-based fibers, crystallite sizes increase and the interlayer parameter d_{002} decreases.

10.2.2.3 Pyrolytic Carbon Fibers

Carbon fibrils, i.e., pyrolytic carbon fibers, are among the solid products of pyrolysis of gaseous hydrocarbons. Visually, pyrolytic carbon fibers are similar to graphite whiskers [13].

Pyrolytic carbon fibers are grown on graphite substrates using preliminary seeded salts of iron, nickel, and cobalt. As a result of pyrolysis of methane at 1100–1400 °C, with a residual gas pressure of 20–80 Torr (namely 3–11 kPa), such fibers grow in the direction perpendicular to the deposition surface. The linear growth rate is about 10 mm/min. Fully grown fibers can be as long as 30 mm and the diameter can reach 70 μm.

Pyrocarbon fibers are formed only in a particular range of hydrocarbon/methane concentrations. As formation takes place, a portion of pyrocarbon is deposited as a continuous coating and another portion forms carbon fibers. These fibers are polycrystalline. They mainly consist of turbostratically structured carbon with crystallites of size ~40 Å, and with high fiber-axis orientation of hexagonal layers (see Fig. 10.12). Pyrocarbon fibers are formed in stages, as follows [14].

(1) Decomposition of the catalytic agent spreads on the sample surface when heated in the methane atmosphere, with the associated formation of reduced metal on the particle surface.
(2) Formation, above the surface, of bulk electric charge caused by evaporation (in an ionic form) of potassium present in the catalytic agent. This induces electric charge of the opposite sign localized on surface irregularities.
(3) Intensification of reactions in the gaseous phase as temperature rises, i.e., potassium ions evaporate, and iron atoms become centers of homogeneous crystallization of carbon. In addition, the rising temperature disturbs the balance between bulk and surface electric charges.
(4) Development of carbon fibrils as a result of Coulomb interaction between bulk carbon-bearing electric charges and electric charges located on surface irregularities, and then at the apexes of the growing fibrils.

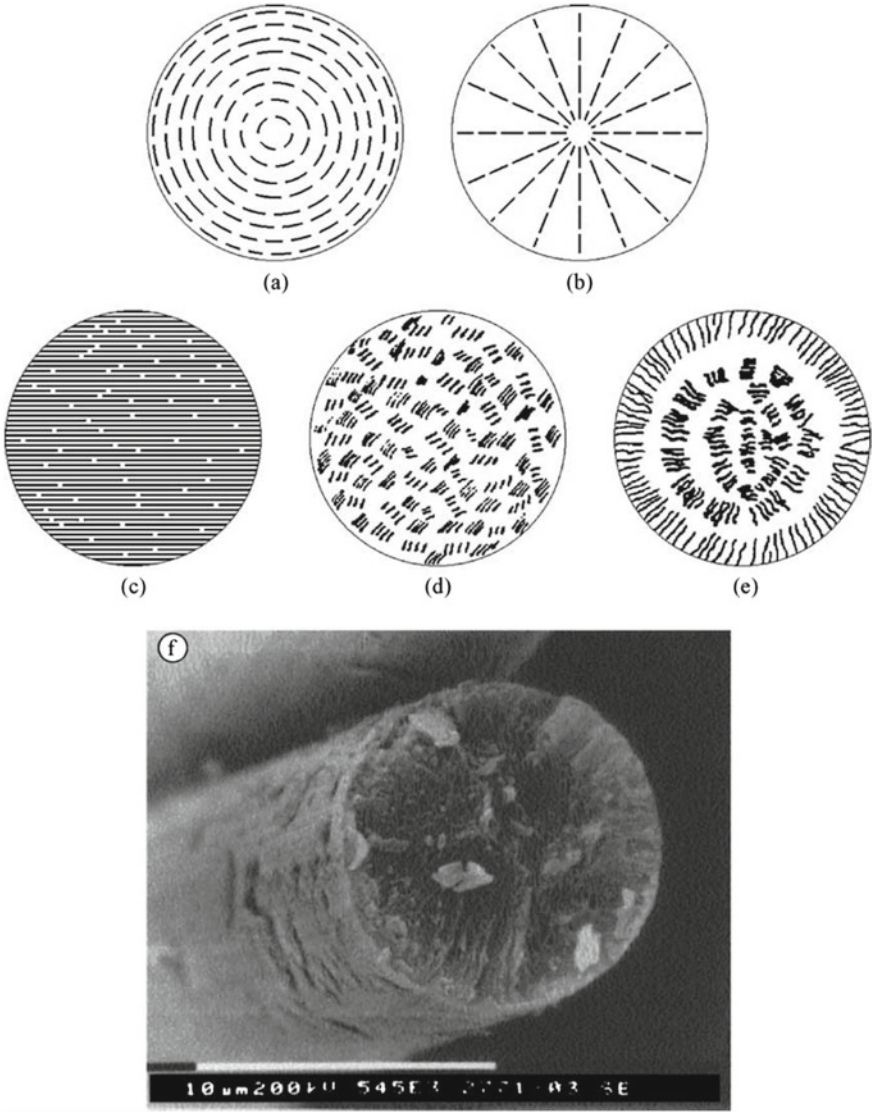


Fig. 10.11 Models for the cross-sections of pitch carbon fibers: **a** onion-skin type structure (tangential); **b** radial structure; **c** structure of R-series fiber; **d** isotropic structure; **e** structure of fiber type TP; **f** SEM micrograph of the working end-surface of a "brand P-25" carbon fiber; from [3]: E.P. Sheshin, Properties of carbon materials, especially fibers, for field emitter applications, Appl. Surf. Sci. **215**, 191–200 (2003), with permission of Elsevier

Table 10.1 Properties of different "brands" of pitch-based carbon fibers

Brand of Carbon fiber	Structure	Fibril length (a-axis) L_a (nm)	Fibril length (c-axis) L_c (nm)	Lattice parameter d_{002} (nm)	Main feature of carbon fiber
J	Isotropic	2.62	1.07	0.41	Isotropic fiber
JG	Isotropic (graphitized)	9.65	3.7	0.34	Isotropic fiber with processing temperature of 2700 °C
MP	Mesophase	6.46	2.05	0.35	Fiber based on mesophase pitch, suitable for general purpose applications
TP	Mesophase	19.05	11.9	0.34	High-modulus fiber based on mesophase pitch
P-25	Mesophase	18.6	11.3	0.34	Fiber based on mesophase pitch
P-75	Mesophase	20.9	14.6	0.34	High-modulus fiber based on mesophase pitch
P-100	Mesophase	21.2	19.7	0.34	High-modulus fiber based on mesophase pitch
P-120	Mesophase	21.8	21.5	0.34	High-modulus fiber based on mesophase pitch

(5) Growth of thin fibrils as the temperature rises.
(6) Growth in fiber diameter, caused by layer-by-layer deposition of pyrocarbon.

The resulting fiber structure depends on the particular conditions of pyrocarbon growth, including the substrate temperature, the pressure of methane, and the pressure gradient along the fiber length.

The specific electrical resistance (electrical resistivity) of pyrocarbon fibers is in the range of 5–10 Ω μm. The tensile strength of pyrocarbon fibers is strongly dependent on their diameter, as may be deduced from Fig. 10.13. The basic length of the sample in these trials was 3 mm. Figure 10.13a shows that the strongest fibers have a diameter of less than 10 μm, and a tensile strength around 2 GPa. With an increase in the diameter up to 30 μm, fiber tensile strength decreases to 0.6–0.8 GPa. Along with the dependence of strength on diameter, the pyrocarbon fiber tensile strength has been found to depend on another geometrical factor, namely fiber length. This form of dependence is shown in Fig. 10.13b, where the fiber diameter is 8–10 μm. However, analysis of the dependence shows that a tenfold increase in

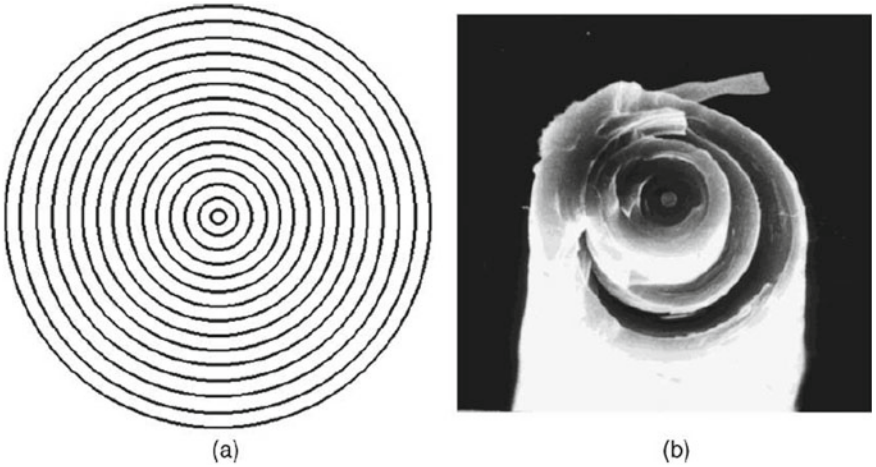


Fig. 10.12 Pyrolytic fibers: **a** model of the transverse structure; **a** SEM micrograph showing the end-surface of the fiber; from [3]: E.P. Sheshin, Properties of carbon materials, especially fibers, for field emitter applications, Appl. Surf. Sci. **215**, 191–200 (2003), with permission of Elsevier

the basic length reduces fiber tensile strength by as little as 40%. In addition to the scale factor, pyrocarbon fibers are characterized by considerable variation in how tensile strength varies with fiber length, as a result of different structural defects in the samples.

In addition, the manufacturing of FE cathodes normally involves man-made carbon-based materials. Such materials are produced from organic substances whose destruction results in coke residue, which is a solid carbon product.

10.2.3 Pyrographite

Pyrolitic graphite or pyrographite [15–18] is produced by the decomposition of carbon gases (propane, methane, acetylene, etc.) on surfaces heated to 1000–2500 °C. If pyrolysis proceeds in a condensed phase, the result is low-temperature pyrographite, i.e., pyrocarbon (800–1400 °C) [19]. Decomposition of hydrocarbon gas at higher temperatures, usually above 2000 °C, results in true pyrographite.

Pyrographite is a hypocrystalline material characterized by a high degree of preferred orientation of crystallites along the deposition surface. X-ray diffraction studies show that some crystallites have a well-expressed texture of the plane (002) parallel to the deposition surface. Order increases as the pyrographite synthesis temperature rises, and temperatures above 2300 °C cause the development of three-dimensional order.

Pyrographite of more than 98.5% theoretical density, i.e., 2100–2200 kg/m^3 (2.1–2.2 g/cm^3) has low porosity and increased corrosion stability.

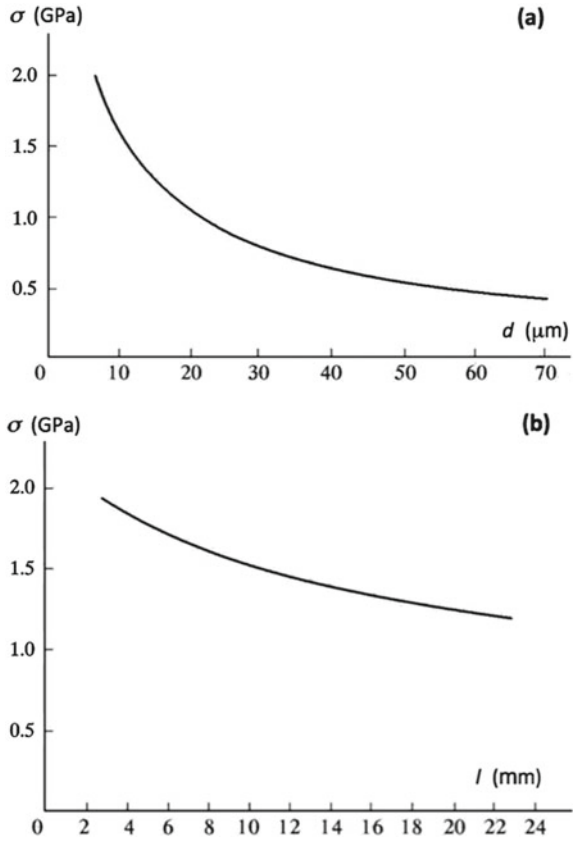


Fig. 10.13 Dependence of the average tensile strength σ of pyrolytic carbon fibers on **a** fiber diameter d; **b** fiber length l

The pyrographite production temperature is among the key parameters determining its structure and properties. However, temperature-effect data need to be treated with care, since the temperature measured can differ considerably from the real one, due to strong temperature gradients. Deposition temperature variations, even as little as ± 20 °C, cause considerable changes in the pyrographite structure.

In this connection, pyrographite properties can widely differ along the deposition surface. This is because gas movement along the surface changes not only the temperature, but also the composition of the gases transported, and sometimes even their response time. In aggregate, these things can change the deposition rate and the pyrographite structure, depending on the distance from the gas ingress place. It is worth noting that studies of pyrographite microstructure sometimes find deposition irregularities, which have caused the emergence of additional inter-boundary regions.

The typical appearance of the effective surface of a pyrographite sample suitable for manufacturing FE cathodes is shown in Fig. 10.14. The picture gives a good view of the pyrographite layers opening out into the effective surface.

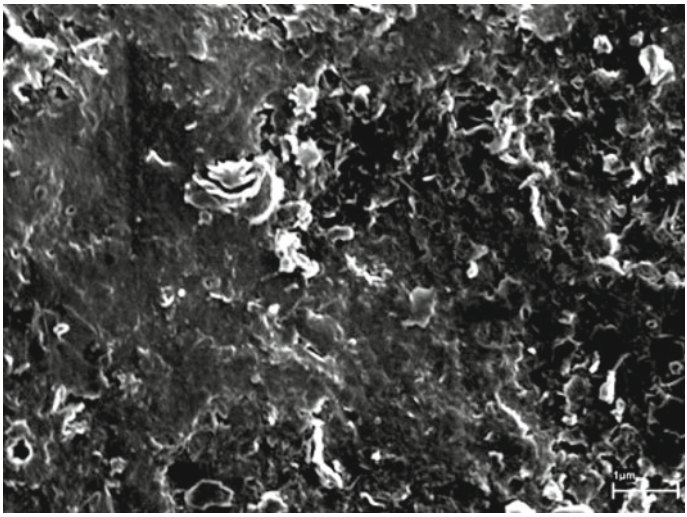


Fig. 10.14 The surface structure of pyrographite; the length of the bar in the lower right corner is 1 μm

10.2.4 Glassy Carbon

Glassy or vitreous carbon is a product of the thermal destruction of cross-linked thermosetting polymers that undergo non-reversible hardening when heated, and are capable of carbonization at the same time. The starting thermosetting polymers most often used are phenol-formaldehyde and furfural resins [20, 21].

Almost all substances used as raw materials to manufacture glassy carbon have the following specific characteristic: there is no clear dependence of final physical and chemical properties on the conditions of synthesis, such as temperature, pressure, and the molecular ratio of the starting components. Thus, the starting raw materials most suited to the glassy-carbon production conditions are chosen by practical considerations.

The characteristic properties of glassy carbon include high compressive strength in the range of 200–500 MPa [22] or even higher at small density (usually about 1450–1550 kg/m^3 or 1.45–1.55 g/cm^3), low fluid and gas permeability, chemical inertness against the most aggressive media, high thermal stability, and high surface purity. In addition to products in various other configurations, glassy carbon is used to manufacture fibers of diameter 6–30 μm that have high tensile strength.

X-ray diffraction analysis shows that there is no three-dimensional order in the glassy-carbon structure. This is also confirmed by the nature of X-ray photoemission spectra. Low-temperature glassy-carbon samples seem to have oxygen bridges. Apparently, all the above suggests that glassy carbon has different types of carbon–carbon bonds irregularly distributed in the material bulk.

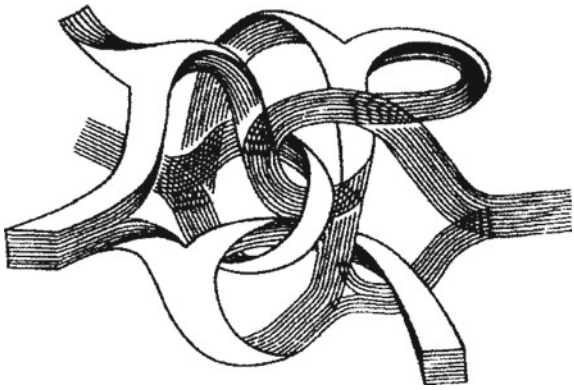


Fig. 10.15 The molecular belt model for the structure of glassy carbon

Glassy carbon does not graphitize at temperatures below 3200 °C. Carbon atoms form small two-dimensional graphite-like layers linked by carbon atoms with 1.55 Å long tetrahedral bonds into a three-dimensional polymeric structure. Figure 10.15 shows a "molecular belt" model of glassy carbon, taking account of microfibril formation. Their presence determines the unusual properties of glassy carbon.

Glassy carbon has micro- and transition pores. Depending on the processing temperature, the pore diameter averages at about 24–62 nm. The walls can have micropores of diameter 3 nm or less. Most of the bulk is made of closed pores, which enables the fabrication of products having almost no through-porosity.

There is also another interesting structural modification of glassy carbon, so-called reticulated vitreous carbon (RVC) foam (see Fig. 10.16) [22]. This name is partially explained by the material structure. The RVC foam has an extremely large number of open pores, up to 96%, which essentially distinguishes it from other kinds of glassy carbon, where pores are mostly closed. As a result, the RVC foam is characterized by extremely low density, no more than 50 kg/m^3, and high gas and fluid permeability. At the same time, it preserves all carbon properties.

Analysis of the porous structure of the RVC foam shows that its surface area increases with the reduction of pore size and growth of their number.

In addition to the above-described materials, FE cathodes can be based on carbon films. It is important to underscore that the same methods can be used to produce carbon films with different structures, because the methods depend on the starting composition and conditions of sample production.

10.2.5 Nanotubes

Carbon nanotubes are extended cylindrical structures, as shown in Fig. 10.17, with the diameter ranging from one to several tens of nanometers and length up to several micrometers. They basically consist of one or several hexagonal sheets rolled up to form quasi-cylindrical "walls", and in some cases have "caps".

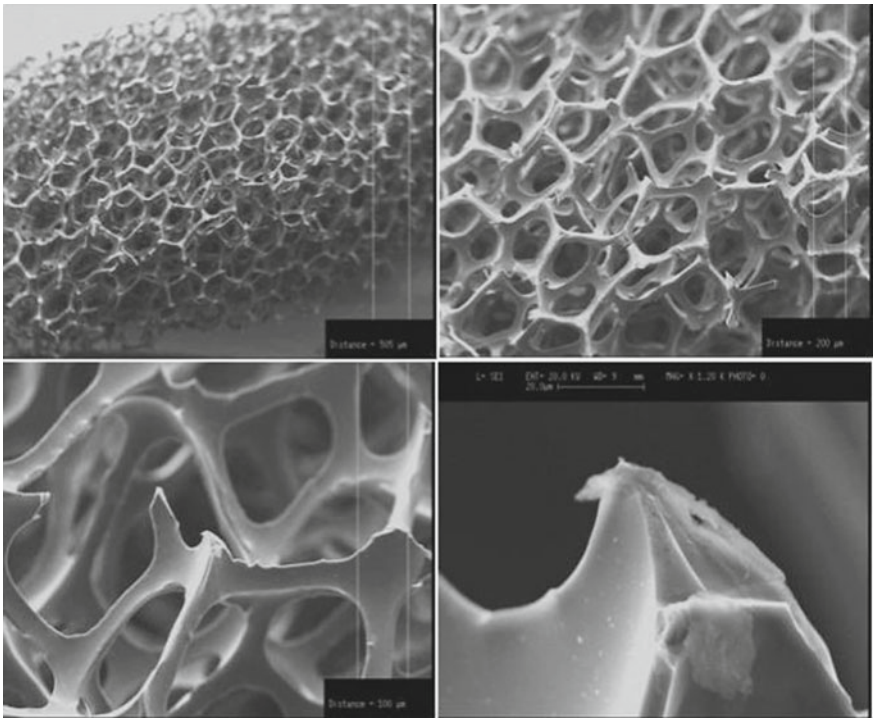


Fig. 10.16 SEM micrographs of the structure of reticulated vitreous carbon (RVC) foam with increasing magnification; the distance between the two vertical white lines is decreasing from 500 μm (upper left) via 200 μm, 100 μm to 20 μm (white bar, lower right)

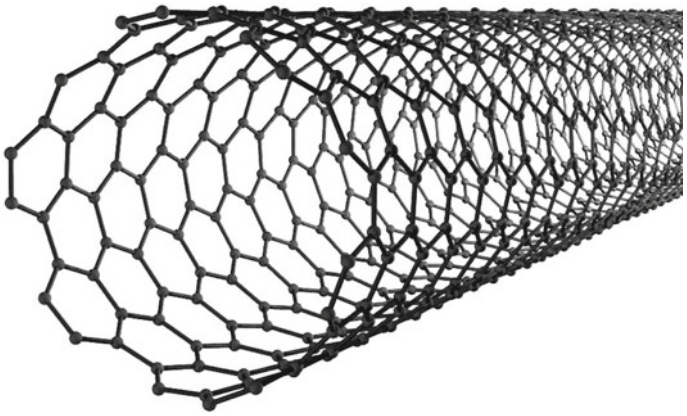


Fig. 10.17 Basic "rolled graphene" structure of a simple carbon nanotube

In 1991, Iijima identified carbon nanotubes by high-resolution electron microscopy [23]. The tubes produced in the first experiments were multi-wall structures differing in the number of layers, the shape of boundaries, and other characteristics. Differences in the chemical activity of a cylindrical nanotube and its hemispherical cap allowed the development of methods of control over the parameters of nanotubes, on the basis of their partial oxidization.

The ideal nanotube is a rolled graphene plane, i.e., a surface made of regular hexagons with carbon atoms at the vertexes. An idealized model of a one-layer (or "single-walled") nanotube is shown in Fig. 10.18. Such a tube has no seams on the generating line and ends with hemispherical "caps" containing not only regular hexagons but also six regular pentagons. The presence of pentagons on the tube ends suggests that we might regard the tube as an extreme case of a fullerene molecule, where the length of the longitudinal axis considerably exceeds the diameter. Real nanotube structures can differ significantly from the ideal one, particularly in the structure of the vertices or fiber ends, which also could be open or truncated.

Multi-walled nanotubes differ from single-walled ones in having a considerably wider variety of shapes and configurations, both in the longitudinal and transverse directions. Some possible modifications of the transverse structure of multi-walled nanotubes are shown in Fig. 10.19. The first two modifications are a complex of

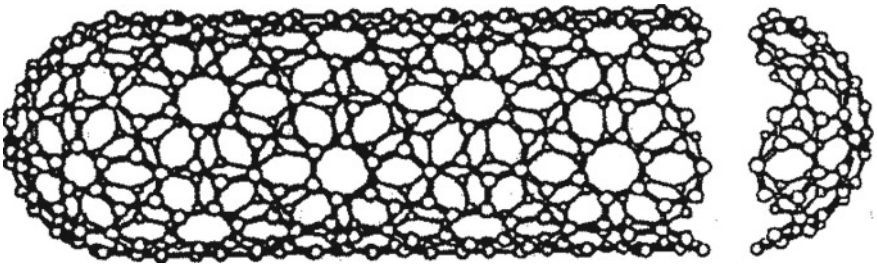


Fig. 10.18 Idealized model of a single-walled carbon nanotube, including quasi-hemispherical "caps"

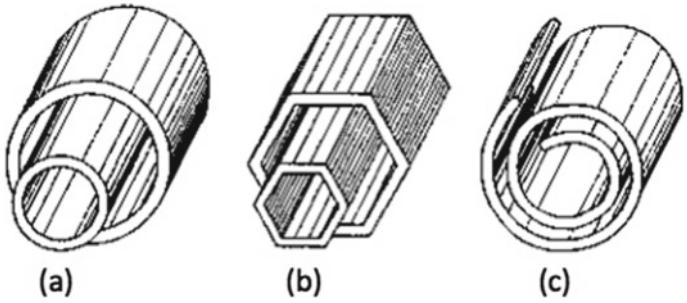


Fig. 10.19 Model transverse structures of multi-walled nanotubes: **a** "Russian doll" (left); **b** hexagonal prism; **c** scroll (right)

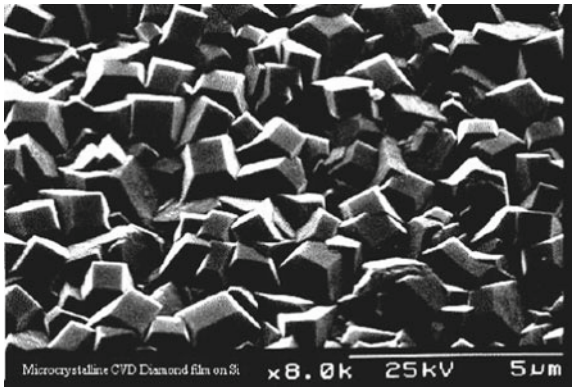


Fig. 10.20 The structure of diamond-like films

single-walled cylindrical tubes put one into another (Fig. 10.19a), or coaxial prisms (Fig. 10.19b). The last of the structures (Fig. 10.19c) resembles a scroll.

Implementation of this or that structure depends on the particular conditions of nanotube synthesis. However, all structures have the distance between adjacent walls close to 0.34 nm, i.e., the distance between adjacent base-planes of crystal graphite. At present, there are no direct experimental data indicating the prevalence of this or that particular structure, the main reason being the insufficiently high resolving ability of contemporary analytical devices.

Recently, much attention has been paid to filling nanotubes with various substances. This can have great practical importance, e.g., for reduction of electronic work function or (potentially) in quantum computing. Filling nanotubes is done by capillary suction of molten metal or solutions. There is a particular interest in filling nanotubes with Cs. Partial filling can also be achieved via embedded catalytic nanoparticle seeds (Fe, Ni) in a PCVD process [24].

The electrical resistivity of some nanotubes is close to that for graphite, but for nanotube films and bundles, the resistivity can vary significantly.

10.2.6 Diamond-like Films

Many authors refer to solid carbon films as “diamond-like”, on the grounds that there is some similarity between their properties and those of diamond. The geometrical surface structure of such a film is illustrated in Fig. 10.20. In fact, analysis of the bond structures in carbon films shows that these structures are neither fully diamond-like nor fully graphite-like, but are complex systems containing both types of bonds, and amorphous regions, all present together.

There are many methods for producing diamond-like films. The most widespread is by chemical vapor deposition (CVD) onto a substrate [25]. The films produced

by CVD have the lowest content of the graphite phase. The carbon film on the substrate surface is formed when carbon ions deposit out of a hydrocarbon gas, usually methane. Sufficiently high substrate temperature (over 1000 °C) enables epitaxial film growth. High concentration of carbon atoms causes the growth of a mostly amorphous carbon film. To prevent that, the non-diamond carbon phases are etched with atomic hydrogen while the films grow. To achieve this effect, the working gas is enriched with hydrogen up to 99%. In this case, it is considered that a plasma etch using chemically pure atomic hydrogen transfers the non-diamond structures of the growing film into the gas phase [25].

Currently, a search of new carbon-based materials usable as FE cathodes is under way. For today, fullerenes and onion-shaped carbon structures seem to be among the most interesting ones.

10.2.7 Fullerenes

Fullerenes are a wide class of polyatomic carbon molecules. They are stable polyatomic carbon clusters consisting of at least several tens of carbon atoms. A fullerene C_{60} molecule is illustrated in Fig. 10.21.

The name was a homage to the engineer and designer Richard Buckminster Fuller, whose geodetic constructions follow this principle. The number of carbon atoms in such a cluster is not arbitrary but follows a particular geometrical regularity. The fullerenes are spheroids with the facets forming pentagons and hexagons. As

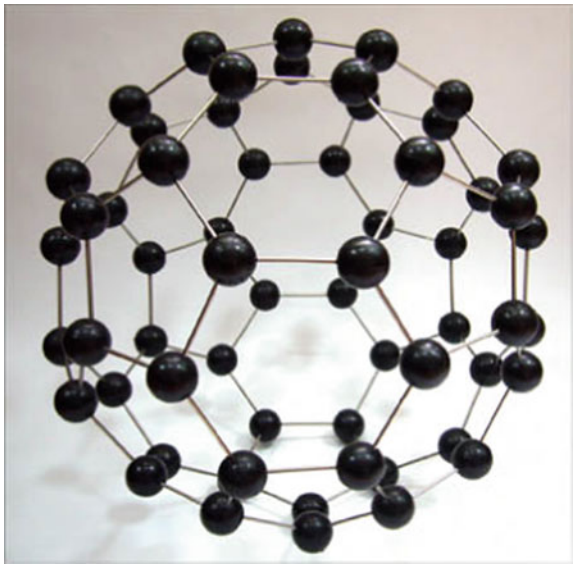


Fig. 10.21 Structure of the fullerene molecule C_{60}

Fig. 10.21 shows, carbon atoms locate themselves on the surface of the spheroid, at the vertexes of the pentagons and the hexagons. Fullerene C_{60} was discovered as a result of experiments by Kroto and Smalley [26] designed to simulate the interstellar medium. Irradiating the surface of a graphite target with a high-power laser beam and using a mass spectrometer, he found a large peak corresponding to the mass of 60 carbon atoms. The structure was identified as a truncated icosahedron and named Buckminsterfullerene (fullerene) in honor of the famous architect. It was noted that its curvature (in comparison with the planar structure of a graphene layer) results from the introduction of pentagonal rings.

Later, a small peak corresponding to a mesh of 70 carbon atoms was identified. Later still, C_{60} started to be produced in macroscopic quantities in a crystal form. Further on, it became clear that graphite can produce a family of the structures referred to as *fullerenes*. According to a geometrical calculation made by Euler long ago, the construction of such a polyhedron requires the number of pentagonal facets to be twelve, while the number of hexangular facets is arbitrary. This condition is met by clusters with the number of atoms $N = 32, 44, 50, 58, 60, 70, 72, 78, 80, 82, 84$, etc. Experimental research is particularly interested in fullerene C_{60} due to its greatest stability and high symmetry.

10.2.8 Onion-like Carbon Structures

Core–shell nanoscale particles, i.e., particles with onion-like carbon shells surrounding a non-carbon core, are very interesting in terms of future practical applications [27]. The carbon layers can protect (from external adverse effects) a nanoscale core that is physically and chemically active. For example, the core can be protected from oxidization and from sorption of foreign substances. A nanoscale core can have physical properties considerably different from those of the bulk material. The shell material, i.e., the graphite-like layers, is ecologically safer and often more *biologically compatible* than many other materials. Materials composed of core–shell particles, aka. "core–shells", can be used in magnetic fluids [28], as contrast substances for NMR analysis [29], and as substances for ultra-dense magnetic recording of information [30].

Core–shell materials can be produced by *high-energy* methods, e.g., the method of arc discharge between metal-containing electrodes [31, 32], or the laser-ablation method, including metallocene powders ablation [33]. A promising core–shell production method involves the use of closed containers and over-pressures [27].

Significantly, such a core–shell particle should be considered not as just a *mixture* of components, but as a *single* physicochemical system where the mutual influence between the phases results in the successful attainment of a particular goal.

10.3 Theoretical Behavior of Post-like Field Emitters

10.3.1 Field Emission Theory in Brief

At present, there is no single general theory of field emission from different materials. Thus, the functioning of each different type of FE cathode is described via a physical and mathematical model of the cathode. As of now, the best developed form of FE theory relates to the emission of electrons from smooth, planar metal surfaces. This theory derives primarily from the work of Fowler and Nordheim (FN) [34, 35] and Murphy and Good (MG) [36], with some later improvements, especially in the last ten years, as described in Chaps. 8 and 9 earlier. (For recent textbook accounts of FE theory, see [2, 37].)

These treatments are far from complete, even for metals. Strictly, they do not exactly apply to semiconductors or to carbon materials, and they do not exactly apply to sharply curved emitters of any kind. However, a common practice among experimentalists is to use the planar-metal-surface formulae to estimate the behavior of FE from semiconductors and carbon-related materials, making appropriate assumptions and allowances. This kind of approach is to be treated as a useful first approximation. Essentially, it is a compromise between a strict physical model, an adequately selected calculation method, and the possibility of obtaining a result in the simplest way and in an analyzable form. Thus, we shall base our calculations below on the zero-temperature FE equation derived by Murphy and Good (MG).

Real field emitters have curved surfaces, and both the local electrostatic field E and the magnitude J of the local emission current density vary with position $\boldsymbol{s}$ on the emitter surface. Strictly, the mathematical form of the potential-energy barrier differs for planar and curved surfaces, but we shall make the usual approximation of taking the barrier at any location $\boldsymbol{s}$ to be that for a planar emitter subject to a field $E(\boldsymbol{s})$, where $E(\boldsymbol{s})$ is the classical electrostatic field at location $\boldsymbol{s}$.

In the coordinate system normally used (as in Sects. 10.3.3 and 10.3.4 below, but not in Sect. 10.3.2), field electron emission occurs in the positive z-direction, and FE occurs when the classical field $E(\boldsymbol{s})$ is negative. Thus, we shall write the Murphy–Good (MG) FE equation for the magnitude $J(\boldsymbol{s})$ of local emission current density at location $\boldsymbol{s}$ in the form

$$J(s) = A|E(s)|^2 \exp[-B/|E(s)|], \tag{10.1a}$$

where the quantities A and B are given by

$$A = t^{-2}(f) a_{\mathrm{FN}} \phi^{-1}, \; B = v(f) b_{\mathrm{FN}} \phi^{33/2}. \tag{10.1b}$$

Here, a_{FN} and b_{FN} are the Fowler–Nordheim (FN) constants, as usually defined (e.g., [38]); ϕ is the local work function; f is the operating value of scaled field, being given by $f = (e^3/4\pi\varepsilon_0)\phi^{-2}|E| \cong (1.439965\ \mathrm{eV}^2\ (\mathrm{V/nm})^{-1}]\phi^{-2}|E|$ (where ε_0 is the vacuum electric permittivity and e is the elementary positive charge); and $v(f)$ and

$t(f)$ are particular values of the field emission special mathematical functions $v(x)$ and $t(x)$ (e.g., [39]), obtained by setting $x = f$.

The magnitude I of the total emission current is obtained by integrating (10.1a) over the surface of the curved emitter. A parameter S_{em} that is a measure of the emission area is obtained by writing the result in the form shown in (10.2):

$$I = \int J(r)\, \mathrm{d}S = J_{\mathrm{a}} S_{\mathrm{em}}. \tag{10.2}$$

where J_{a} is the magnitude of the apex local current density. This parameter S_{em} is sometimes called the *notional emission area* (as derived by using the MG zero-T equation).

Strictly speaking, the theory above is exactly applicable only at the temperature $T = 0$ K. However, since small changes in temperature cause only small changes in the distribution of electrons in the metal, and change the electron supply function at the Fermi level by only about $k_{\mathrm{B}}T$ (where k_{B} is the Boltzmann constant), the formulae of the zero-T theory remain qualitatively adequate at temperatures such that $k_{\mathrm{B}}T \ll \phi$. For example, at room temperature, $k_{\mathrm{B}}T = 26$ meV, whereas a typical work-function value for carbon compounds is $\phi = 4.6$ eV; thus, the condition is well satisfied. Fuller discussions of how temperature affects emission can be found in [36, 37, 40].

10.3.2 The Almazov–Egorov Model: The Single-Emitter Case

The specific properties of different carbon-based materials mean that it is important to be able to optimize the cathode and system geometry appropriately, in order to achieve emission efficiency and the possibility of drawing a large peak field emission current. From the theoretical point of view, this problem comes down to calculation of the optimum cathode structure, and then making practical recommendations to experimental technologists on the basis of the calculation results. An approach of this kind was developed by Egorov and Almazov [41], by making using of slender-body theory [42]. The treatment here follows that in Sect. 7.2 of our textbook [1], recently translated into English [2], but with some differences in notation and some changes in presentation.

Firstly, let us consider an emission system consisting of a FE cathode represented by an axially symmetrical tip of arbitrary shape and a system of electrodes, viz. an anode, grids, etc., as depicted in Fig. 10.22. In such a system, the (total) classical electrostatic potential $\Phi(r, z)$ satisfies the Laplace equation

$$\Delta\Phi = 0, \tag{10.3}$$

with the boundary conditions

$$\Phi|_C = \Phi_C,\ \Phi|_{\mathrm{A}} = \Phi_{\mathrm{A}} \tag{10.4}$$

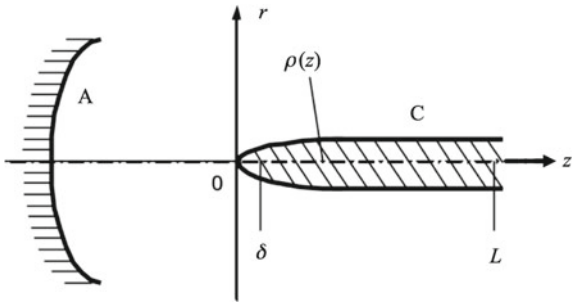


Fig. 10.22 "Slender-body" model o0m. "A" labels the anode surface, "C" the cathode surface. The electrostatic potentials are Φ_A at the anode, Φ_C at the cathode, with $\Phi_A \gg \Phi_C$. Cylindrical coordinates (r, z) are being used, and the emitter apex "a" is placed at (0,0). For explanation of other labels, see text. In this diagram, due to the orientation of the emitter relative to the coordinate axes, field electron emission takes place in the negative z-direction, at a classical field value that is positive

Here, "A" labels the anode and "C" the cathode, "$|_A$" labels a value taken at the anode surface and "$|_C$" a value taken at the cathode surface, Φ_A is the electrostatic potential of the anode and Φ_C the potential of the cathode. The related classical electrostatic field is given by the usual formula $\boldsymbol{E} = -\,\mathbf{grad}\Phi$. The most important characteristic of such systems is the value $\boldsymbol{E}_a$ of $\boldsymbol{E}$ at the tip apex. Except in various special cases, there are no good analytical expressions for $\boldsymbol{E}_a$, because its value depends on the geometrical shape of the field emitter, and (in some cases) on the whole system geometry.

For practical applications, thin needle-like emitters are the most interesting. In such problems, a small parameter arises, namely the ratio of the needle radius to the emitter's typical longitudinal dimension. To study such emitters, it is sufficient to expand the electrostatic potential Φ as an asymptotic series in the vicinity of an *ideally thin* emitter. The first few terms of this series will allow us to calculate the field intensity in quite a precise way.

To construct an asymptotic expansion of this kind, let us apply the following approach. Consider the axially symmetric system shown in Fig. 10.22, where r, z are cylindrical coordinates. Using the modern international system of quantities and equations, in which charge is measured in coulombs, let a line-charge of linear density $\sigma(z)$ be located along the z-axis, between $z = \delta$ and $z = L$, but for notational simplicity, let the symbol $\rho(z)$ denote the quantity $\sigma(z)/4\pi\varepsilon_0$. Let us represent the sought-for solution as the sum of three contributions

$$\Phi(r, z) = \Phi_0(r, z) + \Phi_1(r, z) + \Phi_2(r, z), \tag{10.5}$$

where Φ_0 is given in terms of a line-charge representing the needle, via

$$\Phi_0(r,z) = \int_{\delta}^{L} \frac{\rho(z')}{\sqrt{r^2+(z-z')^2}} \mathrm{d}z', \tag{10.6}$$

where z' is a dummy variable related to z. We now choose the functions Φ_1 and Φ_2 to meet the following requirements:

$$\Delta\Phi_1 = 0, \quad \Phi_1|_{\mathrm{A}} = \Phi_{\mathrm{A}}, \tag{10.7a}$$

$$\Delta\Phi_2 = 0, \quad \Phi_2|_{\mathrm{A}} = -\Phi_0|_{\mathrm{A}}, \tag{10.7b}$$

and note that the function Φ_2 can be represented as

$$\Phi_2(r,z) = \int_{\delta}^{L} u_2(r,z;z')\rho(z')\mathrm{d}z', \tag{10.8}$$

where the function $u_2(r, z; z')$ is defined via the Green's function for problem (10.6), namely

$$u_2(r,z;z') \;=\; -4\pi\varepsilon_0 G(r,z;0,z') - [(r^2 + (z-z')^2]^{-1/2}. \tag{10.9}$$

It is obvious that (10.5) satisfies the Laplace equation and the boundary conditions at the anode "A". Our aim is to select the line-function $\rho(z)$ in such a fashion that the physical tip surface, as represented by a function $r_{\mathrm{C}}(z)$, coincides with the equipotential surface (of the total potential Φ) that has the value Φ_{C}. Having selected $\rho(z)$ in this way, we will be able to meet the boundary conditions (10.4) and find values for the apex field E_{a} and the emission area S_{em} we are interested in.

As we are only interested in thin emitters, it is sufficient to study the equipotential surfaces of the model potential (10.5) that are located near the emitter axis. Let us thus resort to asymptotic expansion [43] of the potential near the emitter axis:

$$\Phi|_{\substack{r\to 0\\ \delta<z<L}} = \rho(z)\ln\frac{2(L-z)}{\delta - z + \sqrt{r^2+(\delta-z)^2}} + \int_{\delta}^{L}\frac{\rho(\xi)-\rho(z)}{|\xi - z|}\mathrm{d}\xi + \Phi_1(0,z) + \Phi_2(0,z). \tag{10.10}$$

Taking (10.10) into account, the equipotential surface close to the tip axis follows the relation

$$\rho(z)\ln\frac{2(L-z)}{r_{\mathrm{C}}(z)} + \int_{\delta}^{L}\frac{\rho(\xi)-\rho(z)}{|\xi - z|}\mathrm{d}\xi + \Phi_1(0,z) + \Phi_2(0,z) \;=\; \Phi_{\mathrm{C}}, \tag{10.11}$$

where $r_C(z)$ is the function that describes the cathode shape. Since this shape is known, (10.11) is an equation for determining $\rho(z)$ and hence the line-charge density $\sigma(z)$.

Equation (10.11) includes the parameter δ, which is still unknown. Let us define it from the condition that $\Phi(0, 0) = \Phi_C$, i.e., we take the tip apex at the origin of coordinates. Let us underscore that for a thin tip, the value of δ should be small. In this case, the condition $\Phi(0, 0) = \Phi_C$ can be represented as [44]

$$\Phi_C = \rho(\delta) \ln \frac{L}{\delta} + \int_{\delta}^{L} \frac{\rho(\xi) - \rho(\delta)}{\xi} d\xi + \Phi_1(0, 0) + \Phi_2(0, 0). \tag{10.12}$$

Another relation is found by putting $z = \delta$ in (10.11), which yields

$$\Phi_C = \rho(\delta) \ln \frac{2(L - \delta)}{r_C(\delta)} + \int_{\delta}^{L} \frac{\rho(\xi) - \rho(\delta)}{|\xi - \delta|} d\xi + \Phi_1(0, \delta) + \Phi_2(0, \delta). \tag{10.13}$$

Comparison of (10.12) and (10.13) demonstrates that the relation $\ln(L/\delta) = \ln\{2(L–\delta)/r_C(\delta)\}$ should be true within the accuracy up to the terms vanishing as $\delta \to 0$, from which we can obtain the following equation for the value of δ:

$$2\delta = r_C(\delta). \tag{10.14}$$

Near the tip apex, for any smooth tip shape, the function r_C^2 can be written as

$$r_C^2(z) = 2R_a z + O(z^2), \tag{10.15}$$

where R_a is the tip's apex radius-of-curvature. We assume that the tip is smooth enough at the apex. Moreover, we will consider that the condition $r_C^2(\delta) = 2R_a\delta[1 + O(\delta)]$ has been met. Substituting into (10.14) leads to the result:

$$\delta = R_a/2. \tag{10.16}$$

If condition (10.15) is met, then (10.11) can be simplified. Let us underscore that if $z \gg \delta$, then

$$\frac{2(L - z)}{\delta - z + \sqrt{r_C^2(z) + (\delta - z)^2}} = \frac{4z(L - z)}{r_C^2(z)} \left(1 + O\left(\frac{r_C^2(z)}{z^2}\right) + O\left(\frac{\delta}{z}\right)\right). \tag{10.17}$$

Further, if z and δ are of the same order, then the two expressions $2(L - z)\left(\delta - z + \sqrt{r_C^2(z) + (\delta - z)^2}\right)^{-1}$ and $4z(L - z)r_C^{-2}(z)$ coincide in higher order as $\delta \to 0$, and are roughly equal to $2L/R_a$. As a result, (10.11) can be reformulated as

$$\rho(z)\ln\frac{4z(L-z)}{r_{\mathrm{C}}^2(z)}+\int_{\delta}^{L}\frac{\rho(\xi)-\rho(z)}{|\xi-z|}\mathrm{d}\xi+\int_{\delta}^{L}u_2(0,z;\xi)\rho(\xi)\mathrm{d}\xi=\Phi_{\mathrm{C}}-\Phi_1(0,z). \tag{10.18}$$

Using (10.18), let us construct the desired expansion for ρ(z). To this effect, it is sufficient to seek solution of (10.18) as an iterative series. Iterating (10.18), we can obtain the solution in the form

$$\rho(z)=\sum_{k=0}^{\infty}\rho_k(z), \tag{10.19}$$

$$\rho_0(z)=\frac{\Phi_{\mathrm{C}}-\Phi_1(0,z)}{\ln[4z(L-z)r_{\mathrm{C}}^{-2}(z)]}\approx\frac{\Phi_{\mathrm{C}}-\Phi_1(0,z)}{\ln[2L/R_a]}, \tag{10.20}$$

$$\rho_{k+1}(z)=-\left(\ln\frac{4z(L-z)}{r_{\mathrm{C}}^2(z)}\right)^{-1}\left[\int_{\delta}^{L}\frac{\rho_k(\xi)-\rho_k(z)}{|\xi-z|}\mathrm{d}\xi+\int_{\delta}^{L}u_2(0,z;\xi)\rho_k(\xi)\mathrm{d}\xi\right], \tag{10.21}$$

where the second form in (10.20) uses (10.15), but is valid only near the emitter apex. In fact, the iterative series seems to be asymptotic for thin tips.

With coordinate axes and electrostatic potentials defined as in Fig. 10.21, the classical electrostatic field E is regarded as positive if a positive test charge moves to the right, or if an electron moves to the left. In this system, therefore, field electron emission (FE) is characterized by a positive value of E, but by a negative value of $\rho(z)$. Knowing the function $\rho(z)$, it is not difficult to find an expression for the field value E_{a} at the tip apex. Using (10.5) and 10.6), we can express E_{a} as

$$E_{\mathrm{a}}=-\left.\frac{\partial\Phi}{\partial z}\right|_{\substack{z=0\\r=0}}=-\int_{\delta}^{L}\frac{\rho(\xi)}{\xi^2}\mathrm{d}\xi-\left.\frac{\partial}{\partial z}(\Phi_1+\Phi_2)\right|_{\substack{z=0\\r=0}}. \tag{10.22}$$

For thin emitters, the main contribution to (10.22) is given by the first term. Calculating the integral asymptotically as $\delta\to 0$, we find the apex field value as

$$E_{\mathrm{a}}\approx-\rho(\delta)/\delta=-2\rho(\delta)/R_{\mathrm{a}}. \tag{10.23}$$

Adjustments to (10.23), necessary in principle, can be shown to vanish as $\delta\to 0$.

Using (10.20) in (10.23), with $z=\delta$, gives a zero-order approximation for the emitter apex field, namely

$$E_a^{(0)}\approx-\frac{\Phi_{\mathrm{C}}-\Phi_1(0,\delta)}{k_0^{(0)}R_{\mathrm{a}}}\approx-\frac{\Phi_{\mathrm{C}}-\Phi_1(0,0)}{k_0^{(0)}R_{\mathrm{a}}}, \tag{10.24}$$

since it is safe to assume that Φ_1 is slowly varying near (0, 0). The constant $k_0^{(0)}$ is a zero-order estimate of a parameter k_0, and is given by

$$k_0^{(0)} = \frac{1}{2} \ln \frac{2L}{R_a}. \tag{10.25}$$

However, expression (10.25) gives a wrong result for k_0 when L is large. Formulas (10.23) and (10.24) predict that as $L \to \infty$, then $E_a \to 0$; however, it is obvious that in the limit of $L \to \infty$, we should have $E_a \to \text{const} \neq 0$. So the construction of a correct expression for k_0 needs the involvement of a term $\rho_1(z)$ of form (10.21).

Having inserted $\rho_0(z) + \rho_1(z)$ in (10.23), setting $z = \delta$, we find a first-order approximation $k_0^{(1)}$ as

$$k_0^{(1)} = \frac{1}{2} \ln \frac{2L}{R_a} \left[1 - \frac{1}{\Phi_C - \Phi_1(0, \delta)} \int_\delta^L d\xi \left(\frac{\rho_0(\xi) - \rho_0(\delta)}{\xi} + u_2(0, \delta; \xi) \rho_0(\xi) \right) \right]^{-1}. \tag{10.26}$$

For a slender emitter, the integral in square brackets is small because the value of $\rho_0(\xi)$ is small. Therefore, when δ approaches (but does not reach) 0, (10.23) can be written

$$k_0^{(1)} \approx \frac{1}{2} \ln \frac{2L}{R_a} \left[1 + \frac{1}{\Phi_C - \Phi_1(0, 0)} \int_\delta^L d\xi \left(\frac{\rho_0(\xi) - \rho_0(\delta)}{\xi} + u_2(0, 0; \xi) \rho_0(\xi) \right) \right] + \cdots. \tag{10.27}$$

Using (10.20), with $z = \xi$, and assuming that $\rho_0(0) = 0$, we can transform (10.27) into

$$k_0^{(1)} \approx \frac{1}{2} \ln \frac{2L}{R_0} \int_\delta^L d\xi \left[\ln \frac{4\xi(L - \xi)}{r_C^2(\xi)} \right]^{-1} \frac{\Phi_C - \Phi_1(0, \xi)}{\Phi_C - \Phi_1(0, 0)} \left(\frac{1}{\xi} + u_2(0, 0; \xi) \right). \tag{10.28}$$

On assuming that most of the contribution to the integral comes from the region of z where the emitter apex is quasi-spherical, where (10.15) applies, and that Φ_1 is slowly varying in this region, (10.28) reduces to

$$k_0^{(1)} \approx \frac{1}{2} \ln \frac{2L}{R_a} \left[\int_\delta^L \left[\ln \frac{2(L - \xi)}{R_a} \right]^{-1} \frac{d\xi}{\xi} + \int_\delta^L \left[\ln \frac{2(L - \xi)}{R_a} \right]^{-1} u_2(0, 0; \xi) d\xi \right]. \tag{10.29}$$

The leading term in the result for the first integral can be evaluated analytically, but evaluation of the second integral can be difficult in arbitrary system geometry, because a suitable representation needs to be found for $u_2(0, 0; \xi)$.

As an example, let us consider a system consisting of an emitter shaped as a thin ellipsoid of revolution and an anode represented by an infinite plane located at distance D from the tip apex. Let the anode potential be zero. In this case,

$$u_2(0, 0; \xi) = -(2D + \xi)^{-1}. \tag{10.30}$$

The related integrals in (10.29) can be calculated accurately. For the factor $k_0^{(1)}$, we obtain

$$k_0^{(1)} \approx \frac{1}{2} \ln \frac{4LD}{R_a(2D + L)}. \tag{10.31}$$

For $L \ll D$, expression (10.31) turns into (10.25). For $L \gg D$, the factor becomes that for a hyperboloid of revolution.

In both field ion and field electron emission, experiments that aim to find the emitter apex field are often discussed with the aid of *Gomer's formula* $|E_a| = |V_a|/(k_G R_a)$ [45], where V_a is the applied voltage between anode and cathode, R_a is the emitter apex radius, and k_G is a geometrical factor (sometimes called the *shape factor*) that is often taken as about 5 in traditional field electron microscope geometry. In our theory (neglecting the difference between voltage and "difference in classical electrostatic potential", which is always small in practice), the equivalent formula would be

$$E_a \approx \frac{\Phi_A - \Phi_C}{k_G R_a}. \tag{10.32}$$

The physical reason for the difference between k_G and k_0 is that Gomer's formula is an empirical formula that automatically takes into account effects due to charge on all other surfaces in the system, as well as charge on the needle, whereas our approach concentrates on effects due to the charge on the needle. The needle contribution is the largest contribution in most experimental circumstances, and probably is often the dominant contribution.

We also note that the literature (e.g., [46, 47]) does contain formulas for the field-enhancement factor associated with a hemi-ellipsoid of height h on one of a pair of widely separated (by distance l) parallel planar plates. These formulas lead to the following expression for the Gomer-type shape factor for this configuration:

$$k_G \approx \frac{l}{h} \times \left(\frac{1}{2} \ln \frac{4h}{R_a} - 1 \right) \approx \frac{l}{h} \times \left(\frac{1}{2} \ln \frac{4h}{R_a} \right) \tag{10.33}$$

where the second form applies if $l \gg R_a$, which will usually be the case in practice.

Since $h \cong L$, this formula differs from our (10.25) by having a factor "4", rather than "2", in the logarithm, and a dimensionless factor (l/h) outside the logarithm. The physical situation to which (10.33) applies is different from ours, so exact equivalence of the expressions is not to be expected, but it remains a matter for future research as to precisely how the differences between the two expressions arise.

Whatever the precise approach used to predict a value of the apex field E_a, we can use this value to estimate the total current and the emission area (S_{em}), using (10.1a) and (10.2) above. For $\phi = 4.5$ eV, B is typically around 40 V/nm and E_a is typically around 4 V/nm, so the condition $B \gg E_0$ is satisfied. It follows that integral (10.1) can be calculated asymptotically by the Laplace method [48], yielding

$$I = -\frac{2\pi A E_a^4}{B E''_{rr}(0,0)} \exp\left(-\frac{B}{E_a}\right). \tag{10.34}$$

This formula is the basis for the emission-area expression:

$$S_{em} = I J_a^{-1} = -\frac{2\pi E_a^2}{B E''_{rr}(0,0)}. \tag{10.35}$$

Using model potential (10.5), it is not difficult to calculate the value of $-E''_{rr}(0)$. However, the resulting expression looks cumbersome. As $\delta \to 0$, i.e., for a slender emitter, and taking (10.16) and (10.23) into account, (10.35) is essentially reduced to

$$S_{em} = (2\pi R_a^2/B) \cdot E_a. \tag{10.36}$$

The linear dependence of emission area on the apex field value was stated in several papers, e.g., [49–51]. However, they did not determine the coefficient of proportionality for that dependence.

10.3.3 The Almazov–Egorov Model: The Regular Emitter Array

A procedure similar to that above can be applied to a multi-tip periodic system. In particular, it can be applied to the system where a regular array of identical emitters stands on a common, planar "cathode base-plate", as illustrated (for a square array) in Fig. 10.23, and the anode is a distant plate parallel to the cathode base-plate, a distance l away from it. In this context, it is easier to use the common system of coordinates, used in Fig. 10.23, in which electron emission takes place in the positive z-direction, at values of classical electrostatic field that are negative in value. In this case, one takes the origin of coordinates at the location where the axis of a "central"

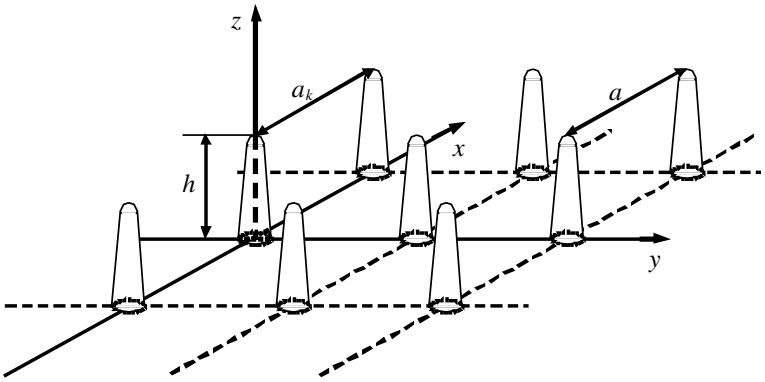


Fig. 10.23 Model of a regular, square multi-emitter array. The squares have sides of length a, and the emitter height is h. Distance z is measured from the base-plane

emitter intersects the base-plane, and defines a "field point" by a three-dimensional vector ($\boldsymbol{r}$,z) where $\boldsymbol{r}$ is a two-dimensional vector in the plane of the base-plate.

For electrostatic consistency, the images of the emitters in the base-plate need to be taken into account. Thus, at a field point above the base-plane, the classical electrostatic potential $\Phi(\boldsymbol{r},z)$ has to be taken as defined by

$$\Phi(\mathbf{r}, z) = -E_{\mathrm{M}} z + \int_{-h+\delta}^{h-\delta} \mathrm{d}\xi \cdot \rho(\xi) \sum_k \frac{1}{\sqrt{|\mathbf{r} - \mathbf{a}_k|^2 + (z-\xi)^2}},$$
$$\rho(-z) = -\rho(z), \tag{10.37}$$

where E_{M} is the "macroscopic" classical electrostatic field E_∞ value far from the cathode; h is the emitter height; $\boldsymbol{a}_k$ is the vector connecting the origin of coordinates to the point where the axis of the kth tip intersects the base-plane; $a_k = |\boldsymbol{a}_k|$; and ξ is a dummy variable associated with the coordinate z. Note that, in the array case, the line-function $\rho(z)$ has values different from the single-emitter case, and that these values depend on the array geometry.

Having made calculations as above, one can be sure that the field value E_{a}^* at the emitter apex is determined by an expression analogous to (10.23) in this case too, i.e.,

$$E_{\mathrm{a}}^* = \frac{2 \cdot \rho(h - R_{\mathrm{a}}/2)}{R_{\mathrm{a}}}. \tag{10.38}$$

However, the line-function $\rho(z)$ [$\equiv\sigma(z)/4\pi\varepsilon_0$] now satisfies a different equation, namely

$$\rho(z)\ln\frac{4(h^2-z^2)}{r_C^2(z)}+\int_{-h+R_a/2}^{h-R_a/2}dz'\left[\frac{\rho(z')-\rho(z)}{|z'-z|}+\sum_{\substack{k\\k\neq 0}}\frac{\rho(z')}{\sqrt{a_k^2+(z-z')^2}}\right]-E_M z\cong\Phi(r_C(z),z)=\Phi_C=0, \tag{10.39}$$

where the sum applies to all emitters except the central one ("$k = 0$") located at the origin of coordinates. The first two terms in this expression are an asymptotic approximation for the classical electrostatic potential due to a line-charge representing the central emitter; the summation term gives the potential contribution due to all the other emitters; and the term involving E_M represents the potential contribution due to the macroscopic field. In the limit that $r_C(z) \to 0$, the sum of these contributions becomes equal to the cathode potential, which is taken to be zero.

Result (10.38) can also be used to define an *apex field enhancement factor* β for this array case, by

$$\beta = \frac{E_a^*}{E_M} = \frac{2\cdot\rho(h-R_a/2)}{E_M R_a} \tag{10.40}$$

It is not difficult to verify that the emission area for an individual emitter is given by (10.26) in both a multi-emitter system and the single-emitter case (but with E_a replaced by E_a^* in the multi-emitter case). However, it should be noted that research into the related asymptotic expansion for (10.39) is difficult when studying multi-emitter systems because: on the one hand, the asymptotic expansion parameter $[\ln(4\,h/R_a)]^{-1}$ can be insufficiently small for such real systems; on the other hand, even the second term of the series considered is rather cumbersome, which prevents efficient research. Therefore, multi-emitter systems are more conveniently studied directly with (10.39). The approximate solution of this equation seems to be equivalent to the approximate summation of the asymptotic series generated by this equation.

Let us note that the numerical solution of (10.39) needs some caution, because its second term is a singular-nucleus integral, which may cause the computing loop to lose stability.

For numerical computation, we selected a system consisting of hemi-ellipsoidal emitters located at the nodes of a flat square grid with node-spacing a (see Fig. 10.23). The equation was solved with the Ritz method [52], the solution having been sought-for as odd polynomials to the $2N$–1 degree. The focus of the research was on calculating the *depolarization factor* ν [$\equiv\beta/\beta_0$], where β_0 is the apex FEF for a single isolated emitter, and investigating how this factor depends on the *height-to-spacing ratio* (*H/S ratio*) h/a, for emitters with different apex radii R_a.

In some contexts, the inverse ratio (or *packing parameter*) a/h is more useful. The numerical computation showed that solution stability was lost even for $N = 4$, for $a/h > 0.5$. For $N > 4$, stability was lost at even less close-packed emitter arrangements. But for $N = 2$ and $N = 3$, the numerical method worked well, even at $a/h \sim 1$, and produced results differing by about 5% at the worst. The computation results are shown in Fig. 10.24. One can see that the tips begin depolarizing each other at a/h

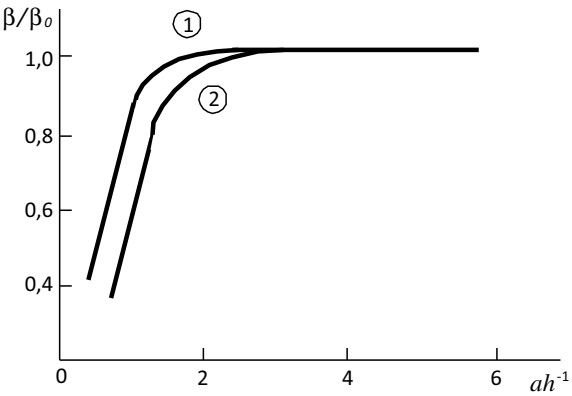


Fig. 10.24 To illustrate how the depolarization ratio β/β_0 depends on the ratio a/h, for a regular, square multi-emitter array. For curve 1, $(h/R_a) = 100$ and $\beta_0 = 50.1$; for curve 2, $(h/R_a) = 50$ and 30.3. For explanation of symbols, see text

~ 3. If the tips are even more close-packed, the field-enhancement factor plummets down. The computation results suggest that tip packaging closer than $a/h \sim 3$ to 4 is not efficient. This result, originally found by Egorov and Almazov [41], is similar to that found later by other authors (e.g., [53]), using a different emitter model (which, however, suggested a "most efficient" value of around $a/h \sim 2$ to 2.5).

In the case of a single isolated emitter in this system geometry, the apex field enhancement factor β_0 would be given by a formula similar to (10.40), but with $\rho(h–R_a/2)$ replaced by $\rho_0(h–R_a/2)$, where $\rho_0(z)$ is the line-function appropriate to the case of a single isolated emitter. Hence, it can be seen that the depolarization factor ν is given by

$$\nu \cong \beta \big/ \beta_0 = \rho(h-R_a/2) \,/\rho_0(h-R_a/2). \tag{10.41}$$

This result is broadly similar to that obtained in the simplified version of the "floating sphere at emitter plane potential" (FSEPP) model, as discussed recently [54], where a result is obtained that could be put in the form

$$\beta \big/ \beta_0 = q \big/ q_0, \tag{10.42}$$

where q_0 is the charge at the center of the floating sphere in the isolated-single-emitter case, and q is the (reduced) charge in the array case. The underlying physics is, of course, the same in both treatments because electron thermodynamics requires that the Fermi level must be constant throughout the cathode, the electrostatic-potential change at each emitter apex, as the emitters are brought closer together, forces electron charge to be withdrawn from the emitter apex into the substrate, thereby returning the apex electrostatic potential to its original value (characteristic of well-separated emitters), but also reducing the magnitude of the apex field.

In the array case, the total current I_1 from each emitter in the array is given by

$$I_1 = S^*_{em} J^*_a \tag{10.43}$$

where the apex current density J_a^* for the array case follows (10.1a), but with the field set equal to the "reduced" apex field E_a^*, and the "reduced" emission area S_{em}^* given by

$$S_{em}^* = (2\pi R_a^2/B)\, E_a^*. \tag{10.44}$$

10.3.4 Optimizing Multi-emitter Systems

This Section considers the problem of how to maximize the macroscopic (or "array average") current density J_M (i.e., the magnitude of the total emission current per unit area of the array base-plane), for a multi-emitter array.

Consider a multi-emitter system as described above and illustrated in Fig. 10.23. Let the anode–cathode separation be l. Take the electrostatic potential of the cathode as zero, and that of the anode as $\Phi_A = -E_M l$, where E_M is the uniform, "macroscopic", classical electrostatic field between the plates. In the system shown in Fig. 10.23, field electron emission occurs when Φ_A is positive and E_M is negative. Emission currents to be considered are at a level where associated space-charge effects are negligible.

To optimize a multi-emitter system, it is necessary to know how the apex field E_a^* depends on the H/S ratio h/a. Let us consider the different terms of (10.39). To that end, introduce some new notations, namely

$$h_r = (h - R_a/2);\ \xi = z/h_r;\ \xi' = z'/h_r; \tag{10.45}$$

$$\tilde{\rho}(\xi) = \frac{\rho(z)}{E_M h_r};\ \tilde{\rho}(\xi') = \frac{\rho(z')}{E_M h_r};\ \ S_k = a_k/a;\ \ \mu = h_r/a \approx h/a. \tag{10.46}$$

Taking into account that for an ellipsoidal apex

$$\ln \frac{4(h^2 - z^2)}{r_C^2(z)} = \ln \frac{4h}{R_a}, \tag{10.47}$$

substitution into (10.39) gives

$$\tilde{\rho}(\xi) \ln \frac{4h}{R_0} + K_1(\tilde{\rho}(\xi)) + K_2(\tilde{\rho}(\xi)) = \xi, \tag{10.48}$$

where K_1, K_2 are given by

$$K_1[\tilde{\rho}(\xi)] = \mu \sum_{\substack{k \\ k \neq 0}} \int_{-1}^{+1} \frac{\tilde{\rho}(\xi')d\xi'}{\sqrt{S_k^2 + \mu^2(\xi - \xi')^2}},$$

$$K_2[\tilde{\rho}(\xi)] = \int_{-1}^{1} \frac{\tilde{\rho}(\xi') - \tilde{\rho}(\xi)}{|\xi' - \xi|} d\xi'. \tag{10.49}$$

For K_1, we consider the expansion [42]

$$\frac{1}{\sqrt{S_k^2 + \mu^2(\xi - \xi')^2}} = \frac{1}{S_k} - \frac{1}{2}\mu^2 \frac{(\xi - \xi')^2}{S_k^3} + \frac{3}{8}\mu^4 \frac{(\xi - \xi')^4}{S_k^5} + O(\mu^6), \tag{10.50}$$

and use the property of "odd symmetry", namely $\rho(-z) = -\rho(z)$, to obtain

$$K_1 = C_3\mu^3 \int_{-1}^{1} \xi\xi'\tilde{\rho}(\xi')d\xi' - \frac{3}{2}C_5\mu^5 \int_{-1}^{1} (\xi^3\xi' + \xi\xi'^3)\tilde{\rho}(\xi')d\xi' + O(\mu^7) \tag{10.51}$$

where C_3 and C_5 are coefficients determined only by the geometry of the array, by

$$C_3 = \sum_{\substack{k \\ k\neq 0}} \frac{1}{S_k^3}, \quad C_5 = \sum_{\substack{k \\ k\neq 0}} \frac{1}{S_k^5}. \tag{10.52}$$

These coefficients can be obtained numerically; for a square grid, $C_3 \cong 9.03362$, $C_5 \cong 5.09026$. It is not difficult to see that, in the context of an established formulation, K_1 is an integral operator acting on $\tilde{\rho}$ according to the rule

$$K_1[\tilde{\rho}] = C_3\mu^3 P_1(\xi)\langle P_1 \cdot \tilde{\rho}\rangle - \frac{3}{2}C_5\mu^5\left[\frac{6}{5}P_1(\xi)\langle P_1 \cdot \tilde{\rho}\rangle + \frac{2}{5}(P_3(\xi)\langle P_1 \cdot \tilde{\rho}\rangle + P_1(\xi)\langle P_3 \cdot \tilde{\rho}\rangle)\right] + O(\mu^7), \tag{10.53}$$

where $\langle f \cdot g\rangle$ represents the scalar product of two *functions* f and g, and the functions, $P_1(\xi) = \xi$, $P_3(\xi) = 5\xi^3/2 - 3\xi/2$, are Legendre polynomials. Relevant orthogonality relations are

$$\langle P_1 \cdot P_1\rangle = \frac{2}{3}, \quad \langle P_3 \cdot P_3\rangle = \frac{2}{7}, \quad \langle P_1 \cdot P_3\rangle = 0. \tag{10.54}$$

Similarly, K_2 is an integral operator whose verifiable eigenfunctions are Legendre polynomials P_n with eigenvalues $\lambda_n = -2\sum_{k=1}^{n} k^{-1}$. Therefore, the subspace covering the polynomials P_1 and P_3 yields the relation.

$$K_2[\tilde{\rho}] = -3P_1(\xi)\langle P_1 \cdot \tilde{\rho}\rangle - \frac{77}{6}P_3(\xi)\langle P_3 \cdot \tilde{\rho}\rangle. \tag{10.55}$$

Hence, (10.48) is reduced to

$$\ln\frac{4h}{R_0}\tilde{\rho}(\xi) - 3P_1(\xi)\langle P_1\cdot\tilde{\rho}\rangle - \frac{77}{6}P_3(\xi)\langle P_3\cdot\tilde{\rho}\rangle + C_3\mu^3 P_1(\xi)\langle P_1\cdot\tilde{\rho}\rangle$$
$$-\frac{3}{2}C_5\mu^5\left[\frac{6}{5}P_1(\xi)\langle P_1\cdot\tilde{\rho}\rangle + \frac{2}{5}(P_3(\xi)\langle P_1\cdot\tilde{\rho}\rangle + P_1(\xi)\langle P_3\cdot\tilde{\rho}\rangle)\right] + O(\mu^7) = P_1(\xi) \tag{10.56}$$

This equation is solved by

$$\tilde{\rho}(\xi) = \alpha P_1(\xi) + \gamma P_3(\xi), \tag{10.57}$$

where α and γ are determined by the equations

$$\left(\ln\frac{4h}{R_a} - 2 + \frac{2}{3}C_3\mu^3 - \frac{6}{5}C_5\mu^5\right)\alpha - \frac{6}{35}C_5\mu^5\gamma = 1,$$
$$\left(\ln\frac{4h}{R_a} - \frac{11}{3}\right)\gamma - \frac{2}{5}C_5\mu^5\alpha = 0. \tag{10.58}$$

In any regimes where it can be assumed that terms in μ^3 and μ^5 are small in comparison with the other terms, clearly a first approximation $\alpha^{(1)}$ for α is

$$\alpha^{(1)} = \left(\ln\frac{4h}{R_a} - 2\right)^{-1}. \tag{10.59}$$

Strictly, the related first approximation for γ is $\gamma^{(1)} = 0$, but a better approximation is

$$\gamma^{(2)} = \frac{2}{5}\left(\ln\frac{4h}{R_a} - 2\right)^{-1}\left(\ln\frac{4h}{R_a} - \frac{11}{5}\right)^{-1}C_5\mu^5. \tag{10.60}$$

This result suggests that γ is small in the regime of interest. Hence, by neglecting the term γ in the first expression in (10.58), and using binomial expansion, we get a second approximation for α

$$\alpha^{(2)} = \left(\ln\frac{4h}{R_a} - 2\right)^{-2}\left(\ln\frac{4h}{R_a} - 2 - \frac{2}{3}C_3\mu^3 + \frac{6}{5}C_5\mu^5\right) \tag{10.61}$$

Since $P_1(1) = 1$ and $P_3(1) = 1$, it follows from (10.46) and (10.57) that

$$\rho(h - R_0/2) \equiv \rho(h_r) = \tilde{\rho}(1)E_M h_r = (\alpha + \gamma)E_M h_r. \tag{10.62}$$

Hence, from (10.40)

$$\beta \ = 2(\alpha + \gamma)\,(h_{\mathrm{r}}/R_{\mathrm{a}}). \tag{10.63}$$

The corresponding result (β_0) for the single-isolated-emitter case is found by neglecting the terms in μ, which leads to (10.59) as an expression for α_0 and then to

$$\beta_0 \ = 2\alpha_0(h_{\mathrm{r}}/R_{\mathrm{a}}) \ = 2\left(\ln\frac{4h}{R_{\mathrm{a}}} - 2\right)^{-1}(h_{\mathrm{r}}/R_{\mathrm{a}}) \ \equiv \ \kappa \cdot (h_{\mathrm{r}}/R_{\mathrm{a}}) \ \approx \ \kappa \cdot (h/R_{\mathrm{a}}), \tag{10.64}$$

where κ is a correction factor defined by this equation and given by

$$\kappa \ = 2\left(\ln\frac{4h}{R_{\mathrm{a}}} - 2\right)^{-1}. \tag{10.65}$$

This result coincides with the 1991 result of Kosmahl [55], using straightforward Laplace-type mathematical analysis, and with later stated formulae [46, 47]. For example, for $h/R_{\mathrm{a}} = 100$, we find $\kappa = 0.501$ and $\beta_0 = 0_0 = 50.1$.

For post-like emitters, formulas for β_0 of the form (10.64) are, of course, now well known (e.g., [46]), with a well-known suggestion [56] that, for the hemisphere-on-cylindrical-post (HCP) emitter model, κ can be adequately approximated as 0.7 over the range $30 < (h/R_{\mathrm{a}}) < 2000$, to within $\pm$ 25%. This is interestingly close to our predicted values.

Returning to the array case, we find from (10.60) to (10.64) that the related approximation for the depolarization factor ν is

$$\nu \ = \ \beta/\beta_0 \ = \ 1 - \frac{\kappa C_3\mu^3}{3} + \frac{6C_5\mu^5}{5}\left(\left(\ln\frac{4h}{R_{\mathrm{a}}} - 2\right)^{-1} + \left(3\ln\frac{4h}{R_{\mathrm{a}}} - 11\right)^{-1}\right). \tag{10.66}$$

In a large-spacing approximation that disregards the term in μ^5, the related *fractional field reduction Δ in apex field*, as a result of electrostatic depolarization, is given by

$$\Delta \ \equiv \ 1 - \nu \ \approx \ \kappa C_3\mu^3/3 \tag{10.67}$$

This result, derived by Egorov and Almazov [41] (though not in precisely this form), supports the recent assertions by de Assis and Dall-Agnoll [57, 58] that, at large array spacings, electrostatic depolarization effects fall off physically as the inverse-third power of the lattice parameter, rather than exponentially (as assumed from some numerical line-charge treatments, e.g., [59]).

Optimization of a multi-emitter system to achieve maximum value of the macroscopic ("array average") current density J_{M} needs knowledge of how J_{M} depends on the H/S ratio h/a [$\cong\mu$]. Disregarding any constant multiplicative terms, the current

I_1 from each tip in the array goes as

$$I_1 \sim \left|E_a^*\right|^3 \exp\left[-B/|E_a^*|\right] \sim \nu^3 \exp[-B/\nu\beta_0|E_M|], \tag{10.68}$$

and the macroscopic current density J_M as

$$J_M \sim I_1/a^2 \sim \mu^2\nu^3 \exp\left(-\frac{B}{\nu\beta_0|E_M|}\right). \tag{10.69}$$

The depolarization factor ν can be written in the form

$$\nu = 1 - d_3\mu^3 + d_5\mu^5 + O(\mu^6), \tag{10.70}$$

where d_3 and d_5 have meanings found by comparing (10.55) with (10.70). It is then not difficult to show that when $\Delta = (1–\nu) \ll 1$ then J_M goes as

$$J_M \sim \mu^2\{1 - D(d_3\mu^3 - d_5\mu^5)\} + O(\mu^8), \tag{10.71}$$

where

$$D = \frac{B}{\beta_0|E_M|} + 3. \tag{10.72}$$

As noted above, $D \gg 1$ in normal FE situations. Having differentiated (10.71) with respect to μ, equated the expression obtained to zero, and taken away the trivial root of $\mu = 0$, we obtain an equation for the optimum value μ_o of μ, namely

$$5d_3\mu_o^3 - 7d_5\mu_o^5 + O(\mu_o^6) = \frac{2}{D}. \tag{10.73}$$

Studies of (10.73) show that it has one root arbitrary close to zero as $D \to \infty$. Other roots are separated from zero for any value of D. Definitely, it is this solution of the problem that is physically sensible, and other roots appeared due to the approximate nature of (10).

If we disregard the higher order terms in μ_o, (10.73) yields a first approximation as

$$\mu_o^{(1)} = \left(\frac{2}{5d_3D}\right)^{1/3}. \tag{10.74}$$

It follows that $O(\mu^6) = O(D^{-2})$, and (10.73) looks like

$$5d_3\mu_o^3 - 7d_5\mu_o^5 = \frac{2}{D} + O(D^{-2}). \tag{10.75}$$

Solution of (10.75) with this accuracy results in the second approximation

$$\mu_{\mathrm{o}}^{(2)} = \left(\frac{2}{5d_3 D}\right)^{1/3} + \frac{14}{75}\frac{d_5}{d_3^2} D^{-1} + O(D^{-1}), \tag{10.76}$$

with

$$d_3 = \frac{2C_3}{3\ln(4h/R_{\mathrm{a}}) - 6}, \tag{10.77}$$

$$d_5 = \frac{6C_5}{5}\left[\left(\ln\frac{4h}{R_{\mathrm{a}}} - 2\right)^{-1} + \left(3\ln\frac{4h}{R_{\mathrm{a}}} - 11\right)^{-1}\right]. \tag{10.78}$$

The procedure described allows us to determine the optimum value of the H/S ratio $\mu = (h–R_{\mathrm{a}}/2)/a \cong h/a$, for fixed values of the other system parameters, namely h, R_{a}, and E. In the first approximation, the optimum lattice spacing a_{opt} (i.e., the closest distance between two emitters) is given by an expression that follows easily from (10.74) and (10.77):

$$a_{\mathrm{opt}} = h\left[\frac{5C_3 D}{3\ln(4h/R_{\mathrm{a}}) - 6}\right]^{1/3}, \tag{10.79}$$

with D given by (10.72).

Expressions (10.79) and (10.72) show that the optimum cathode-tip packaging depends on the operating point of the emission diode (i.e., on E_{M}): at high operating voltages between the anode and cathode, it is optimal for emitters to be more close-packed. The underlying physical reason is that the curvature of the current–voltage emission-diode dependence decreases as emitter apex field increases, which means that the emission current density becomes less sensitive to a drop in apex field caused by tightening of the emitter-array structure. Expressions (10.76)–(10.78) give the first two terms of the expansion for μ_{o} (which is equivalent to a_{opt}) in terms of the parameter D.

As an example, consider a carbon post-like emitter (with work-function 5 eV) that is assumed to emit according to the Murphy–Good FE equation, and initially suppose that, when operating as a single isolated emitter, the apex field ($\beta_0|E_{\mathrm{M}}|$) is 5 V/nm. For such an emitter, $B = 49.6$ V/nm, $D = 12.9$. For $h/R_{\mathrm{a}} = 500$, this yields $a_{\mathrm{opt}}/h = 1.91$. This finding, derived from the formula found by Egorov and Almazov [41], agrees well with values (“near 2”) found later by other authors by using numerical methods (e.g., [53]). For $h/R_{\mathrm{a}} = 500$ and apex field 7 V/nm, the result is the lower value 1.68, as expected from the above discussion. For $h/R_{\mathrm{a}} = 100$ and apex field 5 V/nm, the result is 2.13, demonstrating (as expected from (10.71)) that the optimum spacing is also affected by the value of h/R_{a}.

In general terms, in several applications of the model, the results found by the Egorov and Almazov (1997) analytical model compare well with equivalent results found later by other techniques.

10.4 Materials Science and Experimental Field Emission Properties

The field electron emission (FE) properties of many different carbon-based materials have now been researched. These materials include carbon nanotubes, fibers, foils, fullerenes, diamond-like films, graphene, and variants of these. Carbon fibers have already been investigated in the 1970s (e.g., [60]). Nowadays, a search for new carbon-based materials with stable FE characteristics is under way. This section discusses the FE properties of the most extensively studied materials and of some recently studied materials, drawn from those described above.

10.4.1 Fullerenes

The fullerene sample initially tested was a *bead* of about 3 mm in diameter, chosen because it had an uneven surface that (it was thought) would improve FE properties. But when a relatively high voltage of about 8 kV was applied, no FE pattern was obtained, even after the bead surface had been processed (see Fig. 10.25 a, b). Consequently, there was no current, either. Subsequently, a C_{60} fullerene fragment was selected, on the grounds that the tip should emit. That attempt resulted in quite bright FE patterns (Fig. 10.26) and gave a current–voltage characteristic (Fig. 10.27).

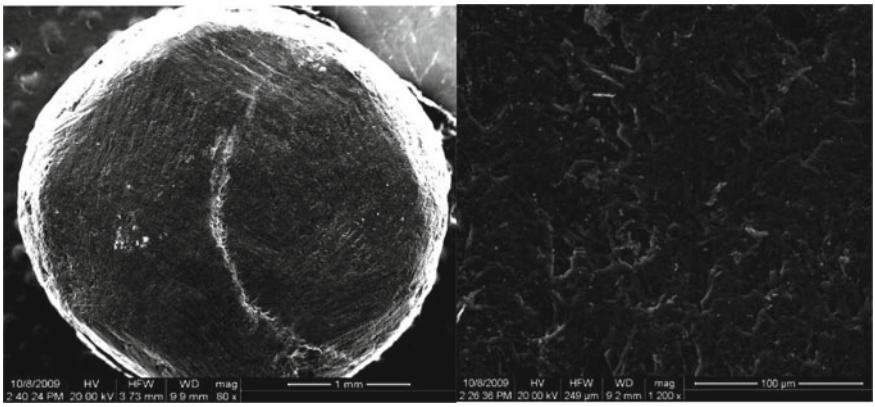


Fig. 10.25 Fullerene specimen prepared as a "bead": (a: left) SEM micrograph, overall view, the length of the white bar is 1 mm; (right) magnified image of surface, length of the bar 100 µm. © 2010 IEEE. Reprinted, with permission, from [61]: V.D. Blank, et al., Proc. 8th International Vacuum Electron Sources Conf. and Nanocarbon, Nanjing, October 2010 (IEEE Explore, 2010), pp. 164–165

Since the cathode-to-anode distance was 1 mm, the onset or threshold field amounted to 4 V/μm, which is quite favorable [61].

As the fullerene FE experiments were only preliminary, that work stage resulted in the conclusion that the material could be used as a FE cathode. However, fullerenes do not make ideal planar FE cathodes, because large parts of their surface areas do not emit.

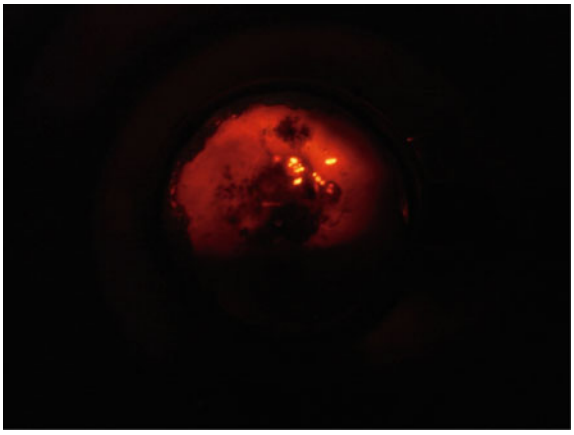


Fig. 10.26 Field electron (emission) microscope image of a C_{60} fullerene fragment

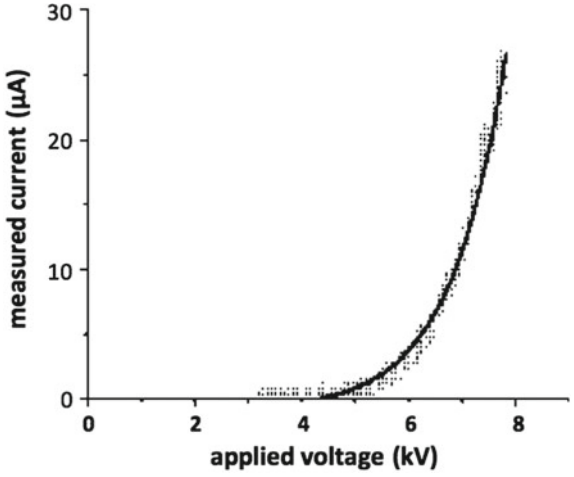


Fig. 10.27 Measured FE current–voltage (I-V) characteristic taken from a C_{60} fullerene fragment

10.4.2 Carbon Nanotubes

The transverse dimensions of nanotubes are extremely small. Therefore, the apex field of a carbon nanotube (CNT) is much higher in magnitude than the average (or "macroscopic") field value in the interelectrode gap. Measurements of the FE characteristics of CNTs mounted perpendicular to the substrate plane are presented in Fig. 10.28. Unfortunately, the current scale is linear and does not extend to the nA range; hence the macroscopic turn-on (onset) field E_{on} cannot be determined from such a plot. This was done by Choi et al. [62], where they measured $E_{on} = 2$ V/μm for single-walled CNTs (SWCNTs) and $E_{on} = 2.9$ V/μm for graphite. A. Musatov et al. found, for multi-walled CNTs (MWCNTs grown by CVD), values of E_{on} of 0.7 to 2 V/μm, for nanotube diameters in the range of 20–40 nm [63].

The typical cathode-to-anode distances in test setups of the authors were in the range of 0.4–1 mm. At a voltage of about 500 V, an emitter with a macroscopic area (or "footprint") of about 1 mm^2 generates an emission current of about 0.5 mA; this corresponds to an *average* (or "macroscopic") emission current density (J_M) of about 50 mA/cm^2. Since it is known that, during FE, the *local* emission current density (J_L) values are typically of the order of 10^3 A/cm^2, it is obvious that only a very small fraction of the total footprint is actually emitting electrons.

This small "emission area" is located near the tips of the CNTs, where the local electric field (E_{loc}) is much higher in magnitude than the mean (or "macroscopic") electric field intensity (E_M) between the cathode and the anode. This is

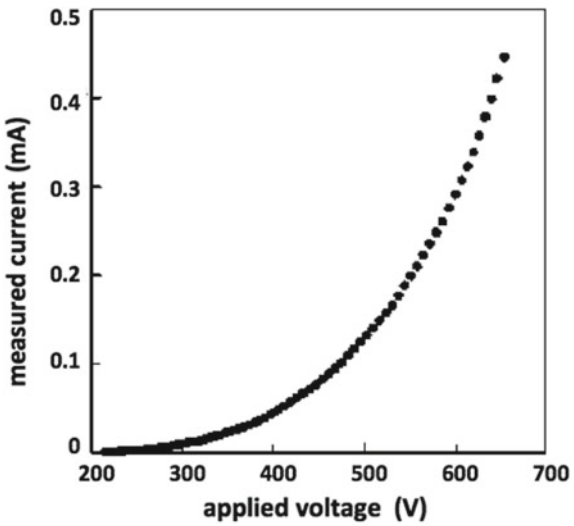


Fig. 10.28 Measured FE current–voltage (I-V) characteristic taken from the film of carbon nanotubes oriented perpendicular to the substrate plane. The macroscopic area (or "footprint") of the film was about 1 mm^2

the well-known physical effect often called *field enhancement*. The related *field-enhancement factor* (FEF) is defined by FEF $= E_{\text{loc}}/E_{\text{M}}$. For carbon nanotubes, FEF-values can sometimes reach a value of 1000 or more (depending primarily on the ratio "height/radius" for the nanotube). This result makes carbon nanotubes (both as single emitters, and as films containing many emitters) quite attractive for application in electronics as cold emitters.

Measurements [64] of total energy distributions (TEDs) found that, at low total emission currents, the TED consisted of separate peaks of about 0.12 eV in width. We ascribe these peaks to emission from particular nanotubes. When the emission current was increased, these peaks broadened and overlapped, but the energy-separation between them remained at the level of 0.1–0.2 eV. One possible explanation is that, with different individual CNTs, the emission is coming from discrete electronic levels that are differently positioned relative to a shared emitter Fermi level.

The same research procedure was applied to look for differences in the FE properties of single-walled (SWCNT) and multi-walled (MWCNT) nanotubes. For the SWCNTs, FE onset occurred at macroscopic fields of 1.6–2 V/μm, and yielded a J_{M}-value of 30 mA/cm^2. The current–voltage characteristic of the process is well described by a Fowler–Nordheim-type expression, which confirms that the emission process is "cold emission" (i.e., Fowler–Nordheim tunneling from states close to the local emitter Fermi level). The maximum achievable value of J_{M} was 3A/cm^2.

For MWCNTs, higher macroscopic-field values are needed in order to cause FE to occur, but the J_{M}-values achieved are nearly the same as for SWCNTs.

By now, there are many reviews relating to field emission from CNTs, as noted in Sect. 10.6. In addition, a particularly useful document is the handbook edited by Saito [65]. Discussion of some particular recent results on CNTs and arrays of bundles of CNTs can be found in Chap. 12.

10.4.3 Pyrographite

Field emission from pyrographite is extremely anisotropic. Its peak value can be obtained from the end-faces of pyrographite plates, but emission is extremely small from the plane faces [66, 67]. Thus, an effective surface for FE cathodes needs to be built from the end-faces of pyrographite plates of varying lengths and thicknesses.

Preliminary experiments showed that pyrographite is quite promising as a field emitter. After that, it became important to determine the optimum temperature of thermal processing and optimum pyrographite plate thickness. To this end, the research focused on the emission characteristics of FE cathodes made with pyrographite plates of 30 μm in thickness processed at 1400 °C, 2000 °C, and 2500 °C, as well as the plates of 5 and 150 μm in thickness processed at 2000 °C and 1700 °C, respectively [67].

The higher the material processing temperature (T_{proc}), the higher was the limiting current achieved. The most likely reason for this was that the material's tensile strength improves with an increase in processing temperature. The FE current fluctuation or instability level (measured as a typical percentage variation in current) was constant for all samples at small currents (about 1–10 μA), but decreased as current increased to higher values (more than 1 mA). At currents of less than 1 μA, the pyrographite FE cathodes of 30 μm in thickness showed the lowest instability value at $T_{proc} = 2000$ °C and the highest at $T_{proc} = 2500$ °C. At currents of 3–6 μA, the instability of all samples was 1–2%.

Analysis of FE current fluctuations can provide additional quantitative data on the surface condition. By examining the statistical variance of the current fluctuations, and how this depends on the number of emitting centers, we have been able to conclude that (other things being equal) the cathode with the largest number of the emitting centers is the pyrographite FE cathode of 30 μm in thickness, processed at a temperature of 2000 °C. The number of emission centers per unit area affects the maximum current that can be drawn, the uniformity of FE over the cathode surface, and the cathode lifetime. Thus, we have concluded that pyrographite with these fabrication parameters is the preferred carbon cathode for electronic devices.

10.4.4 Glassy Carbon

In the case of glassy carbon, it did not take long for early experiments to show some specificities of behavior when the emitters were subject to ionic bombardment in intense electric fields. This finding allows the possibility of an in situ emitter formation process.

Edge structures are the easiest to manufacture. They are produced mechanically in a quite simple way. When the effective surface is 10 mm in length, meaning the footprint area is 10^{-4}–10^{-2} cm^2, the forming operation is most effective at currents of more than 1 mA; under these conditions, the emitting surface is rapidly developed. If the forming has not been completed, the emission current can increase, which presumably indicates that the emitting sites continue to get sharper. Correct forming results in emitters with a very stable current–voltage characteristics, both under many-hour operation in high technical vacuum and after holding the FE cathode under air pressure for many days.

However effective the edge-based glassy-carbon cathodes might be, they cannot be used for many applications involving field emission cathodes of relatively large area. At the same time, it should be noted that, in practice, a flat polished surface of glassy carbon cannot function effectively as a FE cathode.

Quite recently, development has begun of another type of a large-area glassy-carbon FE cathode. This is based on reticulated vitreous carbon (RVC) foam, which can easily be cut into blank cathodes of the desired area and configuration [68, 69]. The surface density of micro-elevations able to become emission centers is regulated

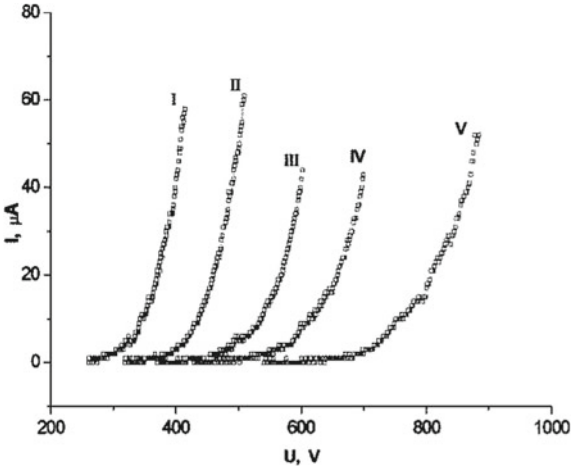


Fig. 10.29 Dynamics of measured current–voltage characteristics, for a large-area glassy carbon cathode, illustrating how the characteristics change gradually (from I to V) as the processing time increases

by the choice of the material volume density, which depends on the details of the configuration of its pores.

Experiments showed that, for large-area glassy-carbon cathodes, the major problems are the same as for many other carbon-based large-area cathodes, namely the achievement of uniform emission from the effective surface of the FE cathode, and achievement of a long service life. Figure 10.29 shows the change of the current–voltage characteristics of an RVC-foam FE cathode with increasing dc current treatment, see [69].

10.4.5 Onion-like Carbon Structures

At the present time, an active search of new materials for FE cathodes continues. Promising new materials include onion-like carbon structures. Only a few experimental investigations of the properties of this material have been described so far, but these suggest that onion-like carbon structures have quite good FE properties. There are different kinds of onion-like structures having different FE properties. When examined in the same experimental test system, the threshold voltage V_{thres} of such powders is less. For example, fullerenes have V_{thres} ~ 3–4 kV, while some powders have V_{thres} ~ 0.9 kV. As yet, we have only managed to obtain some initial data about the FE properties of this powder, i.e., the FE pattern (Fig. 10.30a) and the current–voltage characteristic (Fig. 10.30b).

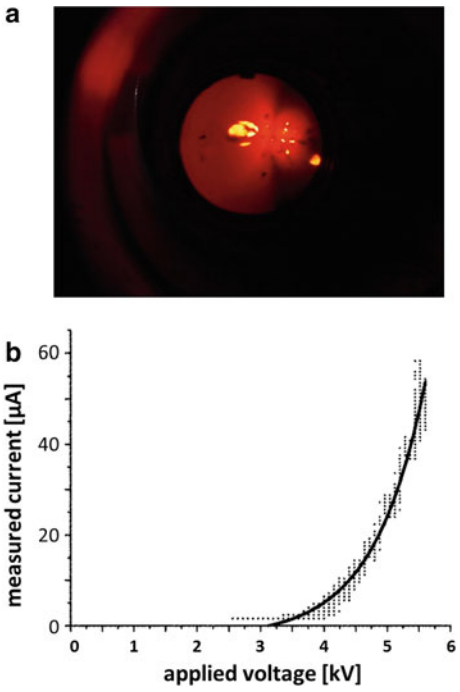


Fig. 10.30 **a** Field electron (emission) microscope image of bulbous carbon structure. **b** Current–voltage (*I-V*) characteristic taken from bulbous carbon structure

10.4.6 Carbon Fibers

Carbon fibers were first investigated as field emitters in the 1970s (e.g., [60]). In comparison with metallic point FE cathodes, the research showed that carbon-fiber emitters have a longer service life under operation in technical vacuum, partly because they seemed to be "self-healing". Nowadays, it seems promising to make FE cathodes of PAN fibers. Depending on the structure determined by the initial fiber and the thermal processing modes, the PAN-based fibers [12, 70] are superior to both pitch-based carbon fibers and pyrocarbon fibers, in terms of their current-supply capability and their service lifetime. There exists a further variant in the form of milled carbon fibers [71].

When studying the emission properties of PAN fibers, it was found that the current–voltage characteristic plotted in Fowler–Nordheim coordinates has a break in slope (see [12], Fig. 3). The analysis of the FE images made at different emission-current values revealed the irregular nature of the emitting carbon fiber surface [12]; compare Fig. 10.31. Another finding was that increase in applied voltage causes a considerable restructuring of the emission surface as a result of ionic bombardment and Maxwell-stress effects. However, FE cathode characteristics become relatively smooth and stable after a sufficient period of operation. The FE images showed

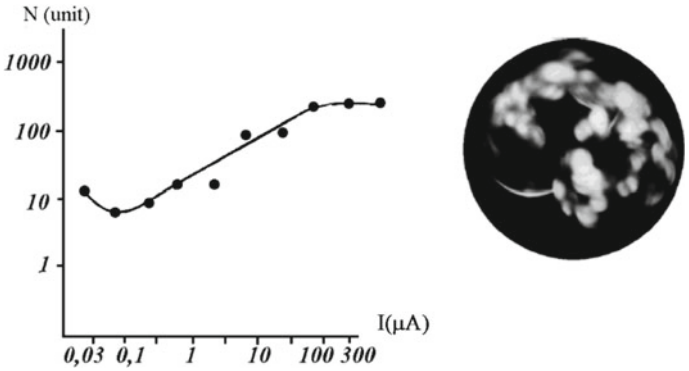


Fig. 10.31 Left: Relation between total current *I* and number *N* of emission centers. Right: Typical field emission microscope image of PAN fiber end (emitting surface). From [12]: E.P. Sheshin, Field emission of carbon fibers, Ultramicroscopy **79**, 101–108 (1999); with permission of Elsevier

stabilization of the FE cathode emission surface, i.e., a considerable growth of the current drawn did not entail any great changes in the emitting-surface structure. These experiments led the authors to conclude that obtaining an effective carbon fiber FE cathode with a stable current–voltage characteristic needs a preliminary formation process. We concluded that, in order to optimize the emitting-surface structure, the FE cathode has to be operated for some time at currents lower than the final working current.

Another important issue is the numerical determination of the emitting-surface area of a fiber. It was suggested above that there is a linear dependence between the emitting-surface area and the total emission current, which is limited to 100–200 μA for one fiber and can be increased in this range by suitable thermal treatment [72]. With a fiber diameter of 7 μm, which is also the diameter of the total emitting area, the theoretical maximum current density is about 520 A/cm^2; but from the field emission microscope (FEM) images, we know that only part of the surface is emitting, with a certain number of emission centers. In Sect. 10.2.2, we have seen that PAN fibers have a sub-micrometer structure consisting of fibrils [3, 12]. In [60], estimations are given of the contribution of emitting fibrils; these also confirm that only part of the fiber end-surface is emitting. In Fig. 10.31, one can see that, at maximum current, about 300 emission centers contributed to the total fiber current, which amounted to about 30% of the fiber end-surface. Since the FEM spots are magnified images of the actual emitting sites, the actual proportion of the surface that is emitting electrons must be significantly smaller. This is consistent with SEM micrographs of the cut fiber end, which show a lot of protrusions and irregularities.

The wide range of emission-current values found by different authors arises mostly from the fact that the parameters of polyacrylonitrile-carbon fibers that determine their emission properties depend on the thermal processing of the fiber. When the limiting FE currents from individual PAN fibers were studied for dependence on processing temperature, it was found out that the limiting current increases with

an increase in processing temperature. The FE current that can be drawn from an individual fiber subjected to thermal preprocessing at 2600 °C can reach 190 μA [70, 72].

Field emission from one fiber end is limited to about 200 μA. Hence, for applications requiring higher total current, bundles of fibers have been prepared. Yet, due to field shielding and non-uniformity, the emitted current is also limited to values below 10 mA, regardless of further increase of the number of fibers in the bundle. Therefore, the authors used separate isolated bunches and thus could demonstrate 100 mA of total emission current from 12 bunches [12]. A. Baturin et al. used a bundle 300–600 fibers in their field emission electron gun [73]. Of course, it is also possible to use insulated fibers in a bundle in order to increase the total current. A further possibility especially w.r.t. higher brightness carbon fibers is to sharpen their ends by electro-polishing or etching, in order to get carbon tips. This was realized by Mousa et al. [74]; 24 μA could be drawn from a single tip.

In the context of field screening, some basic experiments have been conducted by Tang et al. in 2016 [75], using a small number of carbon fibers (1 to 9) with height 1.5 mm and inter-fiber separations of between 200 μm and 600 μm. They found that emission is dominated by the fibers at the outer edges, and that the turn-on field increased with the number of fibers. Their conclusion was that, by optimization of the array geometry, field-screening effects can be reduced.

10.4.7 Diamond-like Films

As a result of many experiments, it has been established that the slightest changes in the deposition arrangements for diamond-like films can cause significant changes in the structure of the films obtained. This affects their FE properties. It is also worth mentioning that analysis of the films obtained established that they contain many structural components—which is why diamond-like films are often described together with fullerenes and nanotubes. However, it was well established that the FE properties of diamond-like films are inversely proportional to the mean crystal grain size. For example, a reduction in grain size (mean averaged grain diameter) from 1.3 μm to 0.3 μm reduces the threshold FE voltage and changes the current–voltage characteristics (see Fig. 10.32a). If the grain sizes are the same, the FE current essentially depends on the surface micro-roughness. In particular, a change of grain orientation from (100) to (110) and (111) results in a shift of current–voltage characteristics (see Fig. 10.32b), which is apparently due to an increase in the field-enhancement factor that appears in Fowler–Nordheim-type formulas.

Other often-cited parameters relating to large-area carbon-film structures are “threshold voltage” and the related “threshold macroscopic field“. Threshold values are defined by criteria that involve either the total current or the macroscopic current density reaching a particular level. Criteria relating to macroscopic current density are to be preferred, but several different values are in use. Threshold voltages and threshold macroscopic fields depend on the overall system geometry, but particularly

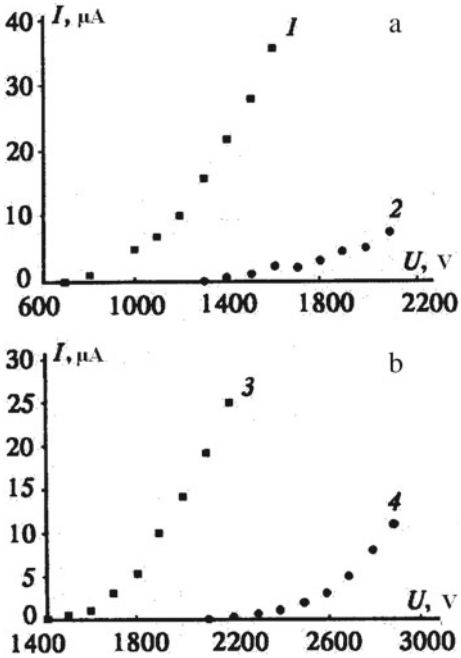


Fig. 10.32 **a** Measured current–voltage (*I-U*) characteristics of diamond-like films, showing their dependence on **a** grain size [1–1.3 μm; 2–0.3 μm]; and **b** grain orientation [3–(100); 4–{(110), (111)}]

on (a) the anode–cathode separation *and* (b) the geometrical structure of the emitter. All of these factors can vary as between different experiments, so care is needed in interpreting the values of threshold parameters. Typical values for the macroscopic current density are around 100 mA/cm^2 (Choi et al. 1997)

A further variant introduced by the authors is to use diamond-carbon nanocomposites as planar field emitters [76]. They consisted of micro-diamond particles embedded in pyrocarbon with 20% nano-diamond prepared by CVD. In order to get a lower turn-on field, the samples were mechanically roughened. Currents of up to 25 μA were achieved.

10.5 Devices and Equipment Employing Carbon-Based FE Cathodes

Recent years have seen FE from nanostructured carbon-based materials applied to provide technical solutions in many types of electronic devices and equipment. Field-electron-emission-based devices can be divided into six basic groups, namely

- light sources
- flat display screens
- X-ray tubes
- electron guns

– microwave devices
– other types of devices, such as ionizers, heaters, and sensors.

We look at these in turn.

10.5.1 Field Emission Light Sources

Basically, existing FE light-source applications can be divided into two main classes (see Fig. 10.33) [77], depending on whether they function in a transmission or a reflection mode.

Generally, the basic classical version is similar to the traditional field electron (emission) microscope. It contains (see Fig. 10.33a) a rod-type FE cathode (1). Such a cathode can be built with any suitable carbon form, such as a graphite rod, a carbon-fiber bunch, or nanostructures at the end face of an appropriate rod. Electrons are extracted with a metal orifice plate that acts as a modulator. (In the literature, this electrode is also called an "extractor" or a "control electrode", but we prefer the term "modulator", since one of its primary functions in a light source is to vary the current and control the brightness.)

The cathode-modulator configuration needs to be designed so that three mutually conflicting factors are optimized: one wants to achieve the maximum current transmission through the modulator, the minimum control voltage, and the maximum uniformity of the electron flow impinging onto the inner surface of the light-emitting screen.

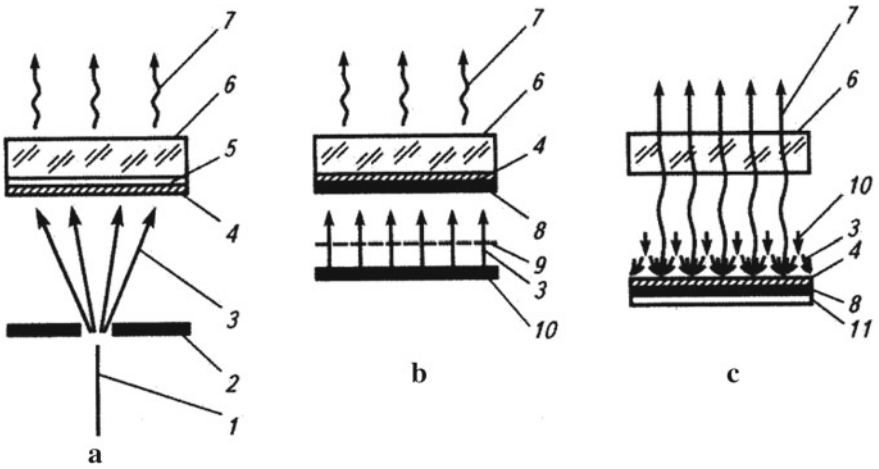


Fig. 10.33 Principles of FE light-source design: a, b—light emitted from the front-plate; c—light emitted by reflection. Design components are 1—field emitter; 2—modulator; 3—flow of electrons; 4—phosphor; 5—transparent conductive coating; 6—output glass; 7—emitted visible light; 8—aluminum coating; 9—grid-modulator; 10—cathode matrix; 11—cathode substrate

The latter is most easily achieved by overlapping the FE flows from a sufficient number of emission centers. This can be achieved by using several rod-type FE cathodes (and related modulator orifices) arranged uniformly in a circle, thus forming a round-section light source.

Classical light sources (see Fig. 10.33a) deposit the light-emitting phosphor onto the back of a transparent conductive coating put onto a glassy front-plate. In such sources, phosphor brightness is not used in full (only about 30% of the photons emitted) because many photons are absorbed by the conductive coating or are emitted backward toward the electron emitter. Better brightness can be obtained if the phosphor and the conductive coating interchange their positions, and the conductive coating is made of aluminum (see Fig. 10.33b). An aluminized coating increases output brightness, raising efficiency to about 70% of the maximum possible value.

Maximum efficiency is achieved when the light source has a "reflection-type" design [78], as shown in Fig. 10.33c. With this design, electrons (3) are emitted toward a luminescent layer (4) deposited onto an aluminum mirror (8) at the "back" of the device. Thus, the light loss is determined primarily by the penetrability of the cathode matrix (10), with relatively small (<10%), light absorption in the glassy front-plate (6).

The particular technological needs of practical applications require the selection of one of these design formats or a variation of it. Various practical FE-based light-source designs are now described.

A so-called "finger-type cathode-luminous tube lamp" is designed to make the brightest light source. It can provide a high-voltage electron flow of up to 10–15 kV with current of up to 1 mA. Such lamps can be more than 100,000 cd/m^2 bright. A schematic diagram of a lamp of this type is shown in Fig. 10.34. The electrons emitted by the FE cathode (1) are driven by the resultant electric field of the modulator (2) and the anode (3) and hit the luminescent screen, thus causing it to glow.

The detailed design of a light source of this type is shown in Fig. 10.35 [79–81]. A vacuum envelope (1) encloses the anode (3, 4), a modulator (5), electrical contact leads (12), and a FE cathode (9) made as a carbon-fiber bunch coated with a dielectric, mostly glassy, over its length. The FE cathode is located in the orifices of alignment disks (8, 10) that have grooves along the perimeter. These grooves tightly hold rods (6), made mostly of glassy, and orienting disks. The centers of the disk orifices are coaxial with the modulator orifice, which is inserted during the assembly of the cathode-modulator unit. The contact arrangements (11) for the FE cathode are made using an electrically conductive substance, usually aquadag, deposited onto

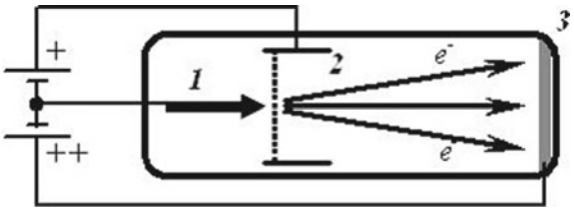


Fig. 10.34 Schematic design of a cathode-luminescence finger lamp: 1—cathode; 2—modulator grid; 3—luminescent anode

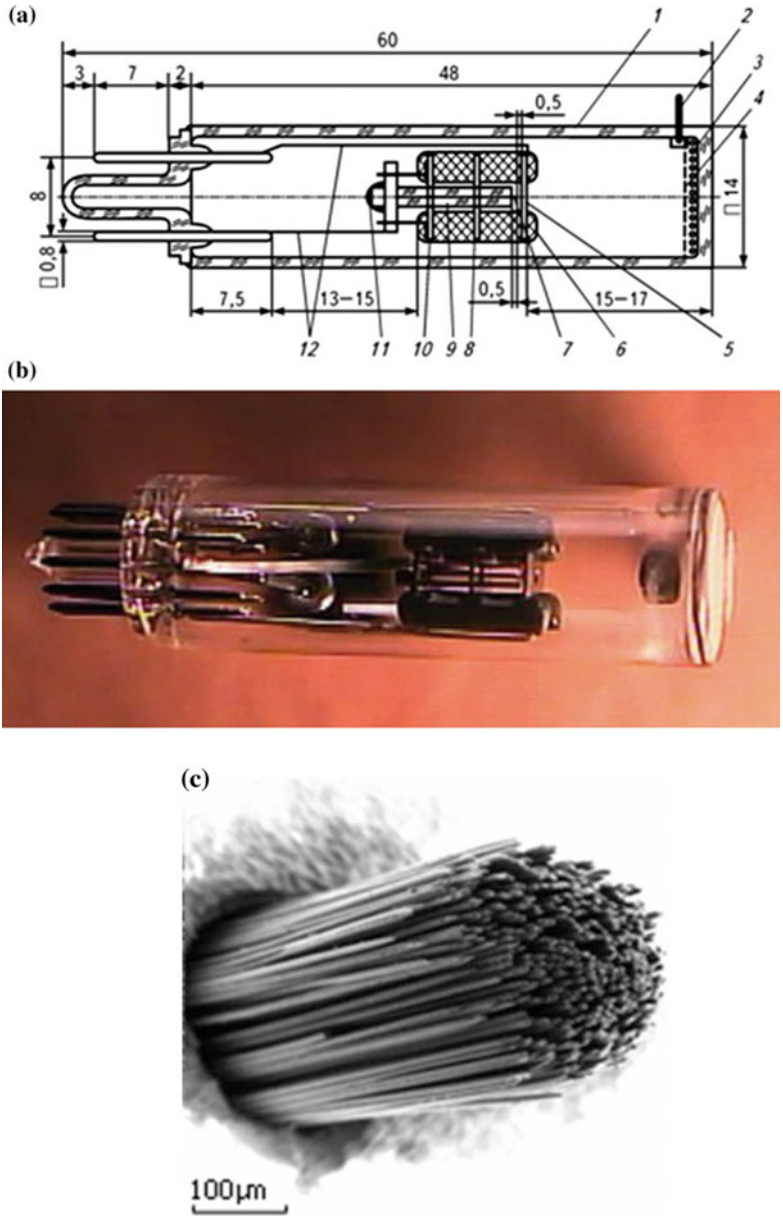


Fig. 10.35 **a** Light source design that uses a carbon-fiber bundle as the FE cathode: 1—vacuum envelope; 2—high-voltage input; 3—phosphor; 4—layer of aluminum; 5—modulator diaphragm; 6—rods; 7—emitting part of bundle of carbon fibers; 8—adjusting disks; 9—vitrified carbon fiber; 10—adjusting disks; 11—electrical contact to fiber; 12—electrical contact leads. **b** Photograph of assembled lamp; see [2], courtesy of Springer. **c** SEM micrograph of the carbon fiber bundle of the FE cathode; see [2], courtesy of Springer

an uncoated part of the fiber bunch. The coating connects the bunch via a lateral surface fixed to the back contact unit of the cathode assembly. The FE cathode itself is a complex multi-point system consisting of 300 PAN fibers bunched together and pre-treated by discharge methods (see Fig. 10.35c).

To increase light-emission power, one has to increase the emission current from the FE cathode. In long-term continuous operation, the current from one carbon-fiber bunch is limited to a value of about 100–150 μA. Thus, further current increase requires that the FE cathode contains several carbon-fiber bunches. In particular, [77, 82] describe a FE cathode with 10 carbon-fiber bunches arranged in a circle. In this case, the lamp dimensions remain very similar to those for the one-bunch FE cathode. Having a large number of fiber bunches in the FE cathode allows the emission current to be increased to about 10 mA. However, such a heavy current can overheat the anode. Another effect is that using a multi-bunch FE cathode improves the uniformity of the screen glow.

As alternatives to carbon fibers, light sources can use other materials as FE cathodes, in particular carbon nanotubes. The designs of such lamps are essentially similar to those shown in Fig. 10.35. Nanotubes are deposited onto flat bases of a few square millimeters in size, either by a CVD method [83] or a print method [84]. Typically, for a cathode-to-modulator-grid separation of 0.2 mm, the control voltage was 300 V. At an anode voltage of 10 kV, the anode current was about 60% of the total emission current of the FE cathode. With an aluminized anode, the brightness of such a source was 6.3×10^4 cd/m^2 for a green phosphor (ZnS:Cu), at an anode current of 200 μA. The efficiency of such a source was 70 lm/W.

The basic structural feature of flat light sources is that the anode and cathode have large areas, but that their separation is small in comparison with the linear dimensions of the cathode substrate. A flat light source has the following components (see Fig. 10.36): a phosphor-coated anode (1); a glass spacer (2), with thickness ranging from some tens of micrometers to a few millimeters; a mounting stub (3); a getter volume (4); a flat FE cathode (5); and a modulator (6), usually implemented by a metal grid.

A well-designed flat-vacuum-device manufacturing technology allows the fabrication of light sources more than 500 cm^2 in area. A device of this kind is built by using diode and triode structures determined by the required device characteristics. Hyper-luminous light sources require a triode structure. Flat light sources of small luminosity, about 1000–5000 cd/m^2, need only a diode structure, which considerably simplifies the manufacture of a vacuum device. In particular, this flat-panel geometry is used in LED-backlit displays.

An early FE-based flat-panel design, shown in Fig. 10.37, is based on carbon-fiber bunches [85–89]. Carbon-fiber bunches (1) are fixed to a metal mounting plate (4). A modulator electrode (2) is located parallel to the emitting plane of the carbon-fiber bunches, at the distance of a few tenths of a millimeter from it. The modulator orifices are coaxial with the emitting ends of the fiber bunches. The cathode (4) and modulator plate (2) are separated by an insulator located inside the vacuum space bounded by glass plates (5) and (7). The upper plate (7) is coated with layers of indium-tin oxide (ITO) of conducting transparent ITO coating (8) and phosphor (6).

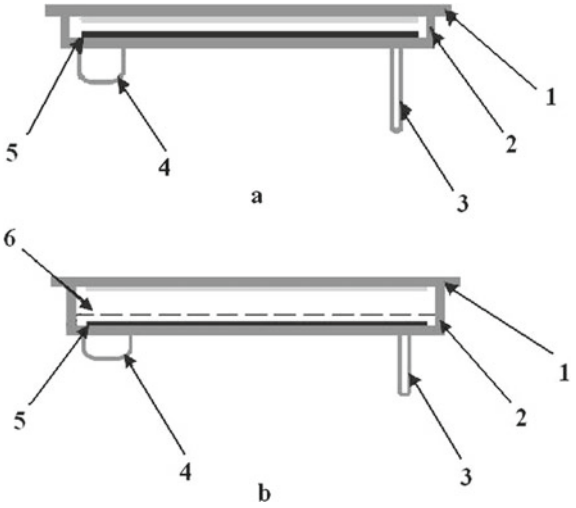


Fig. 10.36 Flat light sources: **a** diode-type design; **b** triode-type design. Design components are 1—anode, coated with phosphor layer; 2—glass spacer; 3—mounting stub; 4—getter volume; 5—flat cathode; 6—modulator grid

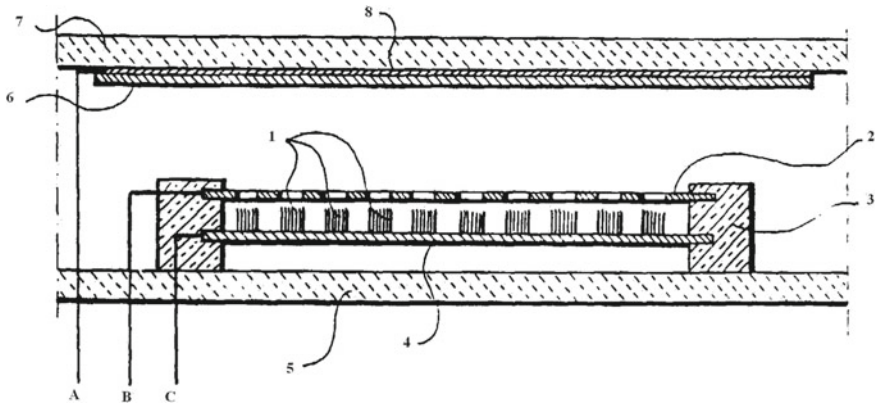


Fig. 10.37 Flat light source built using many bunches of carbon fibers. 1—carbon fiber bunches; 2—modulator with orifices; 3—insulating mounting pillar; 4—metal mounting plate for fiber bunches; 5—glass base-plate; 6—phosphor; 7—glass front-plate; 8—conductive transparent coating

The anode, modulator, and cathode are connected from the outside via electrical leads A, B, and C, respectively.

An alternative to both the above design geometries is to arrange cylindrical components around a central axis, in accordance with the conceptual design shown in Fig. 10.38. This form of design employs a cathode based on fibers, nanotubes, or other FE materials. A cylindrical design has an important electrostatic advantage over

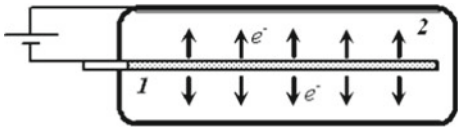


Fig. 10.38 Design of an axial light source: 1—the central part of the cathode structure; 2—cylindrical anode with luminescent coating

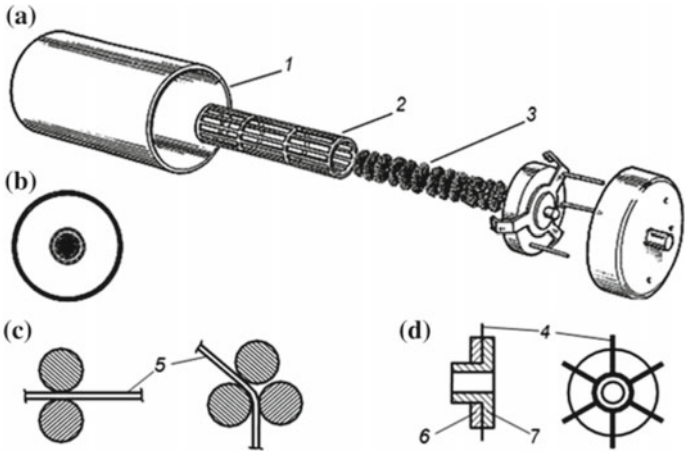


Fig. 10.39 A cylindrical light source that uses carbon-fiber bunches as the FE cathode: **a** to illustrate overall design; **b** cross-section; **c** details of cathode design; **d** cathode design with a radial arrangement of carbon fiber bunches. Components: 1—anode cylinder coated with conductive and luminescent layers; 2—cylindrical modulator mesh; 3—cylindrical-brush-type FE cathode; 4—carbon-fiber bunches; 5—electrical contact wire; 6—clamping disk; 7—mounting disk that holds fiber bunches in place

a planar design, because (for a given anode–cathode separation) the anode–cathode voltage V produces (at the cathode surface) a higher field than in the planar case. (For cylindrical geometry, the magnitude F_M of the macroscopic cathode surface is related to V by $F_M = V/r\ln\{R/r\}$, where R and r are the radii of the anode and the cathode, respectively.) Consequently, it is possible to use a diode light-source structure, which is considerably cheaper in manufacturing than the triode structure, though this is not always done.

A cylindrical triode-based light source with FE cathodes made from carbon-fiber bunches is shown in Fig. 10.39 [90–92]. On the outside is an anode screen (1), i.e., a cylinder-shaped glass base coated with a conductive coating with a phosphor layer. The modulator (2) is a cylindrical metal grid or metal etched-foil. The FE cathode (3) is built from carbon fibers. There are two main types of cylindrical FE cathodes. The first has carbon fibers (4) fixed between two or three twisted wires (5), to form a *brush*-type structure. The structure is very easy to manufacture, but the deformation of the carbon fibers can cause them to break and be removed from the cathode by

electrostatic forces. Moreover, it is very difficult to maintain a constant height for the FE cathode fibers, and thus constant spacing between the fiber tips and the modulator.

In the second version of the FE cathode structure (Fig. 10.39d), carbon fibers were fixed with a conducting glue or soldered between two disks (6) and (7) made from a conducting material. The assembled disk-modules are then fitted onto a bearing rod of the desired length. This structure provides a more accurate coaxiality of the FE cathode and the modulator, as well as higher structural efficiency. All these factors both improve the uniformity of the FE current over the surface of the FE cathode and reduce the likelihood of unintended electrical bridging of the cathode-modulator gap, either by direct contact or by vacuum breakdown.

Many papers [93–97] are dedicated to the classical diode structure, where the FE cathode function is performed by either a tungsten wire or a metal rod of 2 mm in diameter coated with carbon nanotubes. In all these papers, the rod length was between 3–10 cm, and the diameter between 15 and 30 mm. The nanotubes were deposited in tubular reactors made of quartz tubes 680–850 °C, within a deposition time of 10–30 min. The brightness achieved was about 10,000 cd/m^2 at the anode voltage of 5.4 kV [93].

The efficiency of all the above lamps depends significantly on the nature of the phosphor and on the applied voltage. For equivalent conditions, the luminous efficacy is 21 lm/W for a white phosphor and 37 lm/W for a green one [96]. Some efficiency improvement can be achieved by changing the anode structure [98, 99].

Another approach to the development of cylindrical light sources was proposed in [100, 101]. The diagram of such a source and the photo of its working model are presented in Fig. 10.40. The light source is a glass tube of 15–20 mm in diameter and 20–30 cm in length coated on the inside with phosphor. Cathode + modulator units, of the same design as in finger lamps, are welded into the end-faces of the tube. The anode lead-in wire is soldered into the middle of the tube, and supplied with the accelerating voltage of 10–15 kV. The electrons emitted by the FE cathodes knock out secondary electrons on their way to the anode. As a result, the lamp glow uniformity considerably increases and its efficiency improves. Experimental prototypes show a luminous efficacy of more than 20 lm/W, the rise time (turn-on time) being less than 0.1 ms.

An example of the above approach is a full-color video screen model (see Fig. 10.41) intended for shared use. This module has its own power unit and control system, so only needs a mains supply and computer control inputs [99–101]. Structurally, the module is a complete sign space containing $8 \times 8 = 64$ full-color pixels. Each pixel consists of three lamps of red, blue, and green, making a total of 192. The colors are mixed by pulse-width modulation of the emission current. The unit structure allows the construction of a panel of any size, without loss of screen resolution. The power requirement when all the lamps are "on" simultaneously is about 200 W.

The successful design and manufacture of this module show that it is possible to meet a major challenge associated with FE sources, namely uniformity of their emission characteristics [102] in mass production.

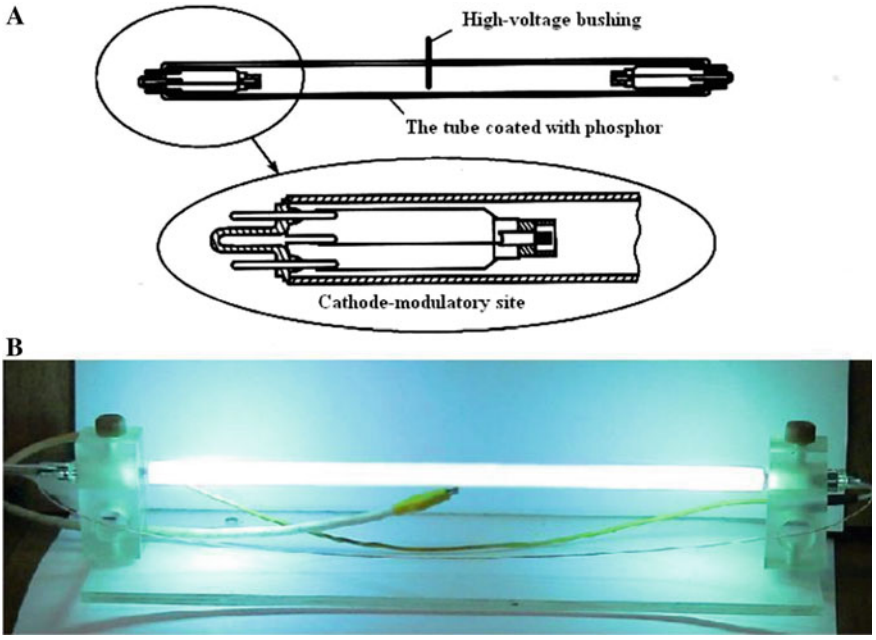


Fig. 10.40 Cylindrical light source with FE cathodes mounted at cylinder ends: **A** design of the source; **B** photograph of a working prototype

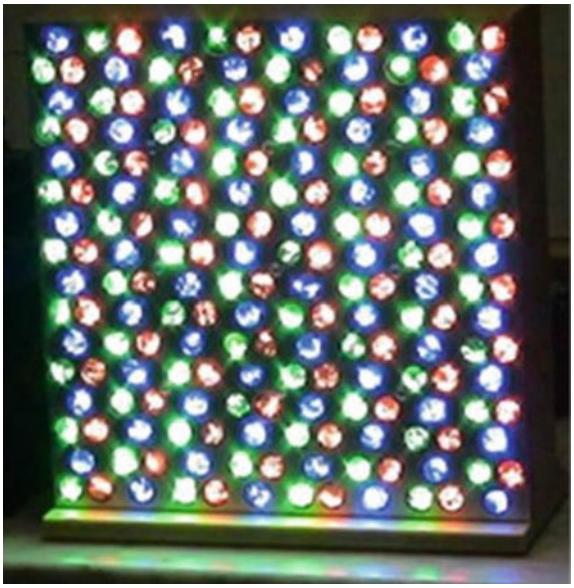


Fig. 10.41 The external module of the full-color video module: 8 by 8 pixels (192) lamp

10.5.2 Flat Display Screens

The world of today is in need of different video-information display devices. The requirements on such devices are also becoming increasingly stricter, because modern displays need to be compact, flat and scalable, reliable, cheap, economical, provide good color rendition and image sharpness over a wide viewing angle, and should have high resolution. Until 2006, which was the turning point in market share, there were three display types in service: liquid-crystal displays (LCDs), plasma-panel displays, and conventional electron-beam displays (= cathode ray tubes or CRTs). Electron-beam displays are ergonomic, have the best color rendition and image sharpness, but are cumbersome and uneconomical w.r.t. large screen sizes and ultra-high resolution. Liquid-crystal monitors are compact and economical, but had a number of problems with image quality; these have now been solved, and also their prices for an increase in dimensions have become competitive. Plasma panels are expensive, have a high power consumption, and are short-lived in comparison with other displays.

None of the three technologies met all the necessary requirements before 2006 (compare Chap. 1). Hence, displays with FE light sources or FE displays had been under development since 1986. Despite that, today in 2020, the race has been won by improved flat thin-film-transistor LC displays (TFT-LCD, light-emitting diode (LED) technology (e.g., [103]) and the new upcoming organic-LED (OLED) display technology (e.g., [104]). Nevertheless, it is instructive in this handbook, to study the technical solutions achieved with field emission displays (FEDs). FEDs were the vacuum electronic (VE) approach to replace the CRT, since they were flat and scalable and had the promise of high resolution. It should be mentioned that there were also other flat VE concepts using thermionic cathodes, pn emitters, or secondary electrons (compare Chap. 1). FEDs can be seen as a form of electron beam display, but with a different electron emission technology. Because FE is a tunneling phenomenon and does not involve emitter heating, it is a highly efficient process in energy terms, and its use can be extremely economical.

As noted above, a FE display pixel is a set of three subpixels, that are implemented as FE diodes or triodes, or as four-electrode FE devices. Anode plates are transparent and are coated with phosphors emitting one of three primary colors: red, blue, and green. The key difference between FE displays and conventional electron-beam displays is that the latter have one electron gun for all the pixels, while the display screens with FE cathodes have an individual electron gun for each subpixel. Such an approach does not involve any cumbersome focusing systems like those in electron-beam displays, and allows the display to be compact and comparable in thickness with liquid-crystal displays. This made the technology of flat display screens based on FE cathodes look very promising for the development of a perfect flat screen [105].

By around 2009 only Futaba and (in earlier years) PixTech had manufactured FEDs for commercial applications. Their displays were competitive at that time, but expensive and only produced in small quantities for special purposes. Later, AU

Optronics from Taiwan bought the FED assets from Sony, but—despite an announcement in 2011—did not start mass production of FEDs. It is still the status today that no company is manufacturing commercial FEDs, and this situation is perhaps unlikely to change.

As a further alternative, Surface-Conduction-Electron-Emitter-Displays were developed as prototypes by Sony, but all activities stopped in 2007.

Difficulties in FED development have related both to inadequacies in available emitter materials and to weaknesses in the structural concepts behind FED design. Modern achievements in electronics allow the development of control circuitry for arrays of FE cathodes, so investigations focused on the FE cathodes themselves and the materials to be used in them. There were two main issues: the development of new phosphors with better efficiency and longer life and the search for new emission materials.

In terms of internal arrangement, FEDs fall into several groups: a classification is shown in Fig. 10.42 [105]. Each structure solves some particular problem, but none solves all of them. The main structure types are described below, with particular examples.

The diode structures provide the simplest and cheapest flat display screens—which is why their development attracted great attention. Especially active attempts to develop such screens started with the application of screen-printing techniques and the electrophoresis method of producing FE cathodes. With the screen-printing technique, it was possible to produce monochrome [106, 107] and color [108, 109] display screens ranging from 4 to 9 inches in size.

Despite their simplicity, diode screens generally have low screen brightness because of a need to keep control voltages relatively low. That is why the greatest efforts have been focused on the development of triode-structure display screens. One section does not provide enough space to cover the whole range of the structures suggested. So we need to specialize our study by describing particular structures that illustrate different design approaches.

One past approach [110] used graphite powder to form FE cathodes that operated at relatively low extraction voltages, around 100 V. The graphite powder, containing

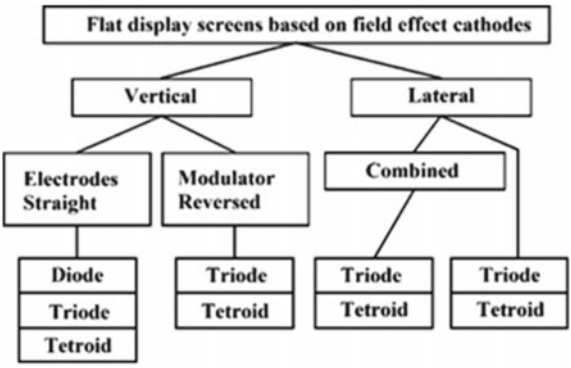


Fig. 10.42 Classification of different designs for "Field Emission Displays" (FEDs), i.e., for flat electronic displays based on field emission cathodes

particles of about 10 μm in size (or average diameter), was mixed with an organic binder to a paste-like state. The paste was then mechanically pressed into holes formed in a glass plate. The holes were made in photosensitive glass, approximately 1.5 mm thick, using ultraviolet photolithography. The holes have a diameter of 200–250 μm at their entrances, but reduce to a diameter of about 100 μm inside the glass, and were spaced 0.6–1.5 mm apart. After the paste had dried, the structure was annealed at a temperature close to the glass softening temperature, to remove the organic binder. Protruding graphite material was removed by mechanical abrasion and polishing. In this design, the modulator was shaped by molybdenum deposition on the structure obtained, with subsequent explosive etching. This resulted in a self-consistent alignment of the FE cathodes and the modulator orifices. The distance between the edges of the graphite cathode and the modulator was about 3–5 μm. The anode used had a usual structure of ITO layers and a phosphor.

The structure of a unit of this design is shown in Fig. 10.43. Contact to the emitter is made via a contact on the back-side of the glass emitter plate. As usual, the anode is coated first with an ITO layer, and then a phosphor layer. The modulator voltage used was 20–50 V; the anode voltage was about 1000–1500 V.

A specific feature of this structure is that FE takes place from the parts of the graphite inserts where the electrostatic field magnitude is highest, which is at the quasi-circular top edges of the inserts. That is why each light spot on the anode is a roughly circular ring [110], formed by a quasi-conical electron beam with a divergence half-angle of 30°–45°. To avoid overlapping the beams from adjacent emitters, the distance between adjacent emitters needs to be greater than the cathode–anode separation.

Naturally, the introduction of each additional electrode into the device structure pushes up the cost. However, the obtained or desired improvement of image brilliance, brightness, and sharpness often more than compensates for the additional costs. Many designs introduce an additional electrode between the modulator and the anode, with orifices coaxial with those in the modulator. This fourth electrode is designed

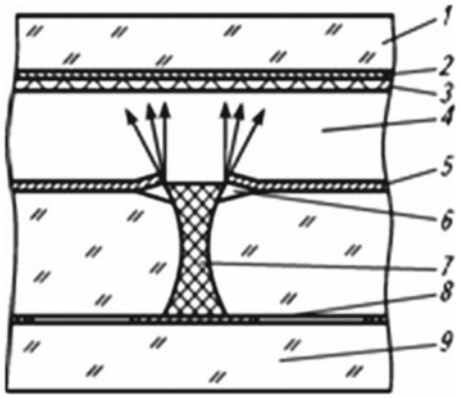


Fig. 10.43 Design of an element of a flat display made using graphite powder: 1—output window; 2—ITO layer; 3—phosphor; 4—flow of electrons; 5—molybdenum modulator; 6—cavity in the glass; 7—graphite cathode; 8—cathode contact; 9—back wall

to overcome the natural divergence of the electron beam from a field emitter, by focusing the beam to produce a smaller spot on the anode [111, 112].

The best results can be obtained by focusing on a so-called "quadrode system" [113]. This includes three apertures formed by structural components of the cathode plate. The focusing electrode can be used not only for its intended purpose, i.e., focusing, but also to protect the FE cathode from discharges [114]. In this case, the aperture of the focusing electrode is smaller than that of the modulator; thus, it is not only focusing that is possible, but also interception of some electrons by the focusing electrode.

The position of the modulator behind the FE cathode under an insulator layer sets this design apart from others. In particular, it eliminates the possibility of vacuum breakdown between the FE cathode and the modulator.

One of the first designs of this type was presented in [115–117]. Figure 10.44a shows the arrangement of the cathode part of the device, while Fig. 10.44b shows the distribution of electrostatic potential and typical electron paths in this kind of cathode-modulator unit. In this design, the glass substrate (1) is coated with an aluminum layer of thickness 150 nm, and then a standard photolithographic process produces modulator electrodes (2) of 400 μm in width. An insulating layer (3) comes next, which in this design is a polyimide layer of thickness 13 μm. Electron-beam vaporization is used to coat this with an aluminum cathode layer (4), and the cathode layer is then developed into cathode bars of width 390 μm. Using a metal grid with a mesh of 20 μm, the cathode bars are then smeared with a paste containing single-walled carbon nanotubes. Subsequent thermal treatment at 350 °C enables the nanotubes to project above the surface of the FE cathode coating, due to the annealing of the binder.

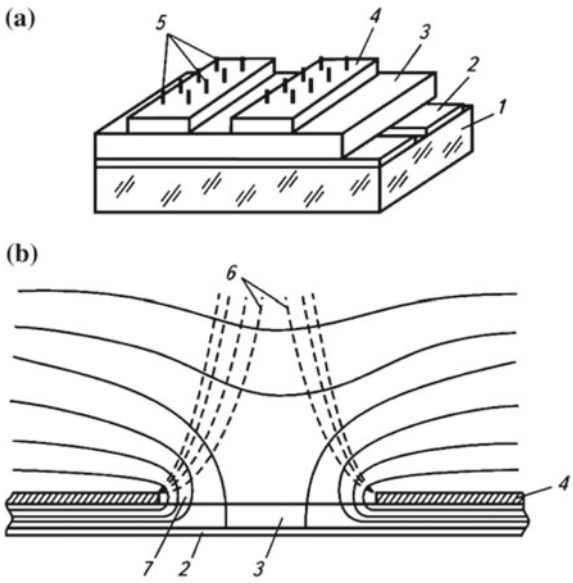


Fig. 10.44 Design of a light source with buried modulator: **a** overall view; **b** to show the distribution of electrostatic potential and trajectories of the emitted electrons. Components: 1—glass substrate; 2—modulator; 3—insulating layer; 4—cathode plate; 5—carbon nanotubes; 6—trajectories of electrons; 7—distribution of electrostatic field; see [2], courtesy of Springer

Figure 10.44b shows the expected distribution of electrostatic potential and expected electron paths, for different modulator voltages. The diagram shows that the electrostatic field is highest in magnitude at the cathode edges, which causes most electron emission to come from the cathode edges, rather than its central part. This effect can be minimized by reducing the width of the cathode units [118].

10.5.3 Microwave Devices

The efficiency of microwave devices needs improvement, and so does their speed-in-action. Microwave-device *heat* power is nearly the same as or even exceeds their *high-voltage* power. In addition, the weight and dimensions of microwave heat sources are comparable with or even exceed the weight and dimensions of high-voltage sources. Reduction of these parameters is especially important for independent and portable radio transmitters. The use of FE cathodes removes all the *heat* problems connected with power-supply sources.

The above comments can be illustrated by the start-up procedure for magnetrons. Magnetrons can be started by *injection*, into the interaction space, of an initial electron current much lower than that of a magnetron in the operating mode. Until now, only thermionic cathodes providing the necessary initial current of several mA have been used for this purpose. Usually, they are effective hot cathodes with an emission current density of 10^4–10^5 A/m^2.

For a FE cathode to start a magnetron, a magnetron reflector is replaced with a disk FE cathode. When anode voltage is applied, the end face of the FE cathode generates a strong electrostatic field that causes FE current. These primary field electrons, driven by the magnetic field, bombard the basic cathode of the magnetron and cause secondary electron emission, thus providing the necessary current for the magnetron in its operating mode.

The magnetron starts almost *immediately*. The start time is determined by an impulse formation time and is less than a micro-second. When hot cathodes are used, the time required for this type of magnetron to reach full operation is 150–180 s from the moment that cathode heating is started. Naturally, FE cathodes do not consume *heating* power, whereas the heating power for the hot cathode needed to start a magnetron amounts to tens of watts.

The first magnetron structures were disk edge FE cathodes. However, more recent magnetron structures employ disk film FE cathodes [119]. The design of such a magnetron is shown in Fig. 10.45. The guide core (1) carries various components. Secondary electron emitters (2) are made from tungsten impregnated with barium, or from alloys and compounds based on platinum-group metals (such as platinum, palladium, iridium, and osmium) combined with an alkaline-earth or rare-earth element (typically Ba, Ir, Ca, La). Between these emitters, there are field emitters (3) that project above them by 5–20% of the interelectrode gap. The field emitters are shaped as beads and consist of a conducting film, with dielectric films (4) protecting the sides.

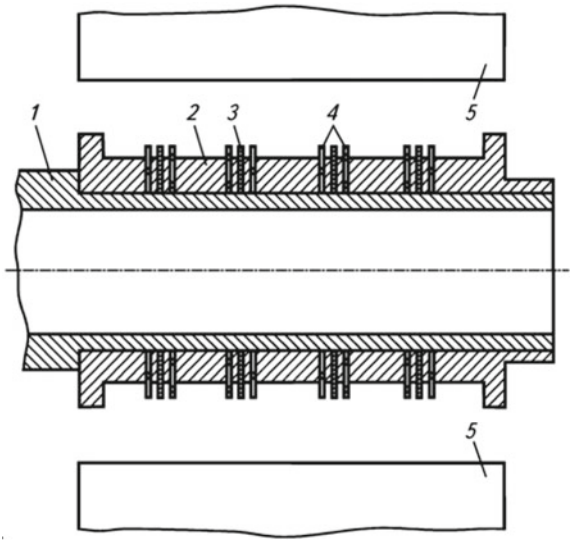


Fig. 10.45 Design of magnetron with disk film FE cathodes: 1—guide core; 2—secondary electron emitters; 3—FE cathode disks; 4—protective dielectric films; 5—cylindrical anode; see [2], courtesy of Springer

The dielectric film protects the lateral surfaces of the field emitter not only mechanically, but also electrically, i.e., when treated with positive ions, it becomes charged to a homogeneous potential that generates a field that repels inbound ions.

10.5.4 X-Ray Tubes

X-ray tubes, requiring high voltage to function, are an obvious application for FE cathodes. Attempts to develop FE-based X-ray technology have been in progress for a long time [120], but by 1975, it was being proposed [121] that the FE cathode be made of differently configured graphite fragments.

Nowadays, we can see intensive development in the field of miniature "through-target" X-ray tubes. A design for such a tube, with a FE cathode made from a carbon-fiber bunch, was suggested in [122] and is shown in Fig. 10.46. The main components in the glass vacuum envelope (1) are an electron gun (2), an anode (3), shaped as a chamber in this case, and the X-ray ejection window (4). The electron gun comprises a FE cathode (5); a modulator (or "cap") (6), with an orifice (12), and with a related electrical lead (7); a dielectric disk (8) that acts as an electrical insulator; and the FE cathode connector assembly (9).

The carbon fiber bunch (10) is mounted in a casing (11), in order to maintain the orientation of the carbon fibers, facilitate their mechanical mounting, and inhibit vibration. The casing can be made from a conducting material or from a semiconducting material such as a semiconducting glass or a metal-coated dielectric, and must be designed in such a way that it firmly grips the carbon-fiber bunch and prevents

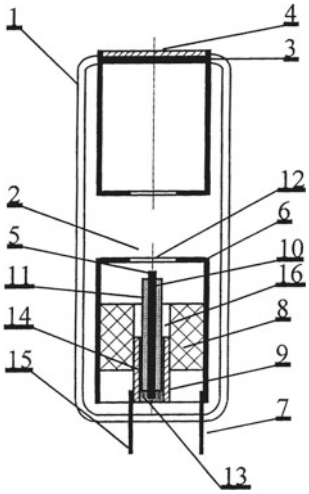


Fig. 10.46 Miniature X-ray tube with "shoot-through" anode and carbon-fiber FE cathode: 1—glass vacuum envelope; 2—electron gun; 3—anode; 4—window; 5—FE cathode; 6—modulator; 7 and15—electrical feedthroughs; 8—dielectric disc; 9—electrical contact for FE cathode; 10—carbon-fiber bunch; 11—casing on fiber bunch; 12—orifice in modulator; 13—electro-conductive paste; 14—metal feed well; 16—protective cavity

movement of the bunch relative to the casing when the tube is in operation. The carbon-fiber bunch (10) projects above the casing (11) on the emitter side.

The X-ray tube functions in the following way. The voltage applied between the cap and the end-faces of the carbon fibers projecting above the casing controls the field acting on the carbon-fiber emitters, and hence the FE current and the X-ray intensity. The high voltage applied between the anode (3) and the cathode (5) accelerates the electrons. X-rays are produced when the electrons bombard the anode, and a fraction of these travel through the anode and glass envelope to emerge from the device.

Numerous fibrils, which in this case are tetragonal crystalline carbon structures, about 25–100 nm in length and 2–5 nm in diameter, oriented along the fiber axis, project above the end-surface of the carbon fiber. These projecting fibrils are the carbon-fiber FE centers. The stop-down orifice (12) in the cap (6) permits field-emitted electrons to travel only to the anode (3) and prevents electrons from hitting the glass walls of the vacuum envelope.

10.5.5 Electron Guns

One of the basic units of an electron-beam tube is the "electron gun" that generates the electron beam, i.e., a focused electron bunch with the energy and current density necessary for the application. Most guns are designed to produce a round-section beam, but sometimes shaped beams, e.g., ribbon beams, oval-section beams, or other specially shaped beams, need to be generated.

The design requirements for an electron gun can vary widely, depending on the intended application. Thus, the beam electron energy needed is several keV in small oscillographic tubes, but tens or even hundreds of keV in electron microscopes.

Another example is that the beam current of some image pick-up tubes is several tenths of a microampere, while it can reach several milliamperes in lithographic machines. However wide the range of variation of energy and current might be, the beam profile in a particular plane (usually in the receiver plane, i.e., a screen or a target) should normally be as small as possible. The beam profile is usually assessed by the *spot diameter*, a "spot" being understood as the trace of the electron beam on the receiver surface. In many types of contemporary device, the spot diameter is as small as several tenths of a millimeter, or in some applications much less (for example, in high-resolution electron microscopes).

Most guns need to be able to control the beam current over a wide range, from zero, i.e., gun blocking, to a maximum value that depends on the purpose of the beam device. Low-current electron guns are very widely used in electron probe devices, such as focused-beam microscopes and transmission electron microscopes, e.g., those of JEOL [123] or Hitachi.

The first electron gun with a FE cathode was developed by Crewe [124], with aspects of its electron-optical behavior calculated by Butler [125] (see Fig. 10.47). The voltage U_1 applied between the point and the first anode determines the maximum total emission current, but it can be varied up to that limit by variation of U_2. The voltage U_0 determines the electron-beam energy by the further acceleration of the electrons. The electrostatic lens marked as "4" focuses the beam and generates an image spot in the plane of the sample. This design has now been developed further (e.g., [126, 127]).

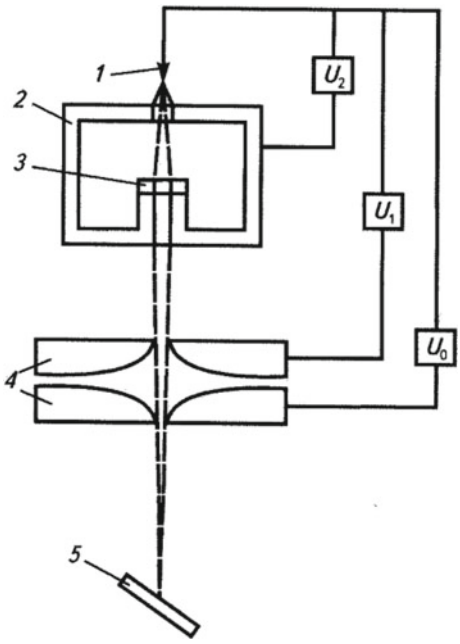


Fig. 10.47 Schematic design of a FE electron gun: 1—emitter; 2—ion trap; 3—aperture diaphragm; 4—Butler-type accelerating stage; 5—specimen

The "camera trap" inlet is 1–2 mm away from the field emitter point. It is supplied with the voltage of 5–8 kV relative to the emitter. The gun works by developing a virtual image of a fictional electron source called the "cross-over" [127]. Swan and Kynastou made a significant advance by suggesting [128] that this virtual source generated by the electron gun could be converted into a real one, by using an additional magnetic condensing lens.

Since the current from one field emitter is very low, the generation of high-ampere currents needs simultaneous operation of many hundreds and thousands of field emitters. The core of this problem has not yet been solved. However, there are continuing attempts to meet the challenge. One approach has been to use arrays of pointed molybdenum FE cathodes (the so-called "Spindt array", discussed in more detail in Chap. 12). Such arrays employ more than 100,000 individual emitters, but need a working vacuum pressure better than 10^{-7} Pa.

As already pointed out, FE cathodes manufactured from carbon-based materials need less demanding working pressures. In particular, Grigoriev et al. used [129] an array of pointed FE cathodes made of glassy carbon, manufactured by photolithography and thermo-chemical etching. The emitters are 14 μm high and have an apex radius of about 1–3 μm. The emitters stand 20 μm apart. When the cathode-plate diameter is 3 mm, the emission current is of order 100 mA.

In addition, there are cathodes with a large area, usually more than 1 cm^2, that are made with carbon-fiber bunches [73] or high-tenacity graphite, e.g., "Fine-Grained Dense Graphite 6" [130]. These have a high-transmission modulator grid positioned near the cathode plane. This electron-gun design allows the formation of an electron beam with a configuration that depends on the shape and the size of the effective cathode surface.

An example of such a design is shown in Fig. 10.48. The field-electron gun is a

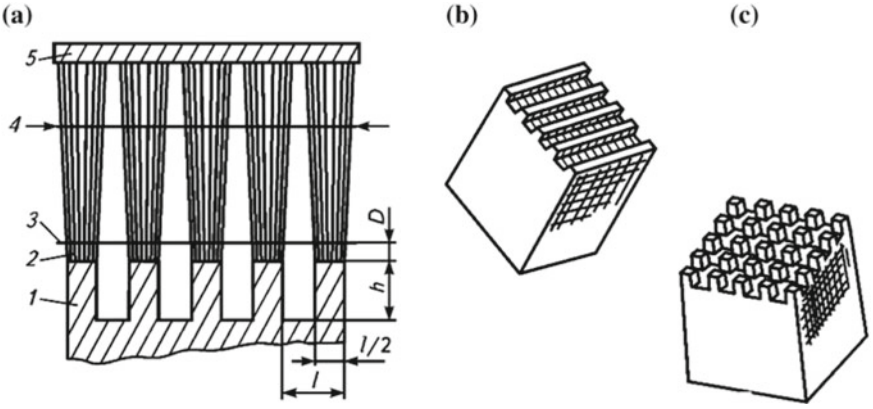


Fig. 10.48 Electron gun with graphite FE cathode: **a** basic design; **b** "bar" form of cathode; **c** matrix form. Components of (**a**): 1—FE cathode built from MPG–6 graphite; 2—working surface of the cathode; 3—modulator, with transparency 80%, built in the form of a grid made using tungsten wire of diameter 0.01 mm; 4—electron beam envelopes; 5—anode

FE cathode (1) with a grooved end-surface (2). In close proximity to the effective surface of the FE cathode, there is an extraction electrode (3) in the form of a grid made from a high-melting-point material. Applying a voltage to the extraction electrode causes FE from the micro-elevations on the effective surface of the cathode. The resulting electron beam, with a configuration that depends on that of the cathode, passes through the extraction electrode and reaches the anode (5), which collects the electrons.

The end-surface of a FE cathode made from nanostructured carbon-based material consists of a great number of micro-elevations, distributed in a statistically uniform fashion over the effective surface. In the course of emission under the bombardment of the surface by residual-gas ions, the emission centers are damaged and at the same time, new micro-elevations are exposed to become new emission centers. After the effective surface of the cathode has aged, this process stabilizes in time. This leads to high time-stability of the emission current drawn from such a cathode.

Increase in the working area of a cathode by more than 4–6 mm^2, intended to increase the cross-section of the emitted beam, can lead to spatial instability of the beam, due to the heterogeneity of a cathode surface of this size. However, a large-cross-section beam can be generated by using a set of separate stable electron beams, each generated by an emitting element that has dimensions comparable with the distance between the cathode and the extraction electrode.

An example of composite field emitter arrays as electron gun cathode from nanocrystalline diamond-like films coated silicon carbide has been developed to improve the electron emission characteristics of the pure silicon carbide tips [131, 132]. The field emission array covered with highly boron-doped diamond film demonstrated the lower turn-on electric field and higher emission current due to lower work function, 1.5 times lower than for undoped one. Both fabricated composite field emission arrays demonstrated high current stability with fluctuations not exceeding 5% at a relatively low vacuum (10^{-5}–10^{-6} Torr). Nevertheless, while the potential of the silicon carbide field emission material is recognized, only a few experimental attempts have been made to fabricate and measure emission characteristics of monolithic silicon carbide field emission arrays [133] and carbon composite structures with silicon carbide, for example, graphene-like structures formed in process of graphitization of silicon carbide surface [134–137].

A final application of interest is that FE cathodes can be used to increase the sensitivity of touch devices, i.e., those responding to even a small change in any physical parameter.

10.6 Conclusions

This chapter has reviewed various carbon-based materials, ranging from long known and well-studied materials to newer materials with much promise for application to FE cathodes and devices based on them. Investigations into the behavior of the best-

known materials, operating under various conditions, have been discussed. In order to bring out information about the basic technologies and basic design principles involved, this chapter has looked at many of the historical developments. However, active search has been continuing for new carbon-based materials and viable applications. Since the first review by Bonard [138] in 2001, numerous review articles and book chapters have been published on FE from carbon materials, especially on carbon nanotubes, and related applications. Those who need more detailed information are invited to consult some of these reviews [138–158] and the Saito handbook [65]. Beyond this, further study of the FE properties of carbon-based materials (old and new) should advance the search to identify carbon forms that will progress the development of various types of electronic devices and improve their operating characteristics.

Acknowledgements The authors are grateful to Richard G. Forbes, Konstantin Nikiforov, and Georg Gaertner for critical reading of their manuscript and for valuable suggestions.

The reported study was funded by RFBR, project number 20-07-01086.

References

1. N.V. Egorov, E.P. Sheshin, *Field Emission: Principles and Applications* (Intellekt, 2011) (in Russian)
2. N.V. Egorov, E.P. Sheshin, *Field Emission Electronics* (Springer, 2017)
3. E.P. Sheshin, Properties of carbon materials, especially fibers, for field emitter applications. Appl. Surf. Sci. **215**, 191–200 (2003)
4. B. Rand, Mechanical properties, pp. 111–140 in: *Graphite and Precursors*, ed. by P. Delhaes (Gordon and Breach, Amsterdam, 2001)
5. M. Dubinin, Porous structure and adsorption properties of active carbons, pp. 51–120 in: *Chemistry and Physics of Carbon: A Series of Advances, Vol. 17*, ed. by P. Walker Jr. (Marcel Dekker, New York, 1966)
6. W.A. de Heer et al., Epitaxial graphene. Solid State Commun. **143**, 92–100 (2007)
7. E. Fitzer (ed.), *Carbon Fibers and Composites* (Mir, Moscow, 1988). (in Russian)
8. B.T. Kelly, Present understanding of thermal properties of graphite, High Temp. High Press. **5**, 133–134 (1973)
9. O. Vohler, E. Sperk, Kohlenstoff—Fasermaterial. Berichte der Deutschen Keramischen Gesellschaft **43**, 199–258 (1966). (in German)
10. W. Watt, Production and properties of high modulus carbon fibres. Proc. R. Soc. Lond. **A319**, 5–15 (1970)
11. A.A. Koknin, *Carbon and Other Heat Resistant Fibrous Materials* (Khimiya, Moscow, 1974). (in Russian)
12. E.P. Sheshin, Field emission of carbon fibers. Ultramicroscopy **79**, 101–108 (1999)
13. I.B. Donnet, R.C. Bansal, *Carbon Fibres* (Marcel Dekker, New York, 1990)
14. R.T.K. Baker, P.S. Harris, The formation of filamentous carbon. Chem. Phys. Carbon **14**, 83–165 (1978)
15. A.S. Fialkov, N.M. Sidorov, B.N. Smirnov, B.I. Djuzhikov, Features of the structure and growth of filamentous pyrocarbon formations, Bull. Russ. Acad. Sci. **211/1**, 158–160 (1973) (in Russian)
16. H. Kostka, S. Birkle, N. Kusebauch, Pyrographit—Eigenschaften und Anwendung, Materialwissenschaft und Werkstofftechnik, **22/3**, 92–99 (1989) (in German)

17. A.S. Fialkov, A.I. Baver, N.M. Sidorov, Pyrographite. Production, structure, properties, Uspekhi Khimii **34**, 132–153 (1965). (in Russian)
18. W.F. Knippenberg, B. Lersmacher, H. Lydtin, A.W. Moore, Pyrolytic graphite. Philips Techn. Rev. **28**, 231–242 (1967)
19. P.A. Tesner, *Carbon Production from Gas Phase Hydrocarbons* (Khimaya, Moscow, 1972). (in Russian)
20. V.D. Chekanova, A.S. Fialkov, I.M. Yurkovskiy, Study of structural changes during heat treatment of furfuralphenol-formaldehyde binder. Solid Fuel Chem. **5**, 1037–1040 (1970). (in Russian)
21. A.M. Sigarev, V.K. Bityugin, T.A. Zhuravleva, Properties and structure of glassy carbon. Graphite-Based Constructive Materials **6**, 132–137 (1971). (in Russian)
22. W.F. Knippenberg, B. Lersmacher, Carbon foams. Philips Techn. Rev. **36**, 93–103 (1976)
23. S. Iijima, Helical microtubules of graphitic carbon. Nature **354**, 56–58 (1991)
24. Z.X. Chen, G.C. Cao, Z.L. Lin, I. Koehler, P.K. Bachmann, A self-assembled synthesis of carbon nanotubes for interconnects. Nanotechnology **17**, 1062–1066 (2006)
25. B.V. Deryagin, D.V. Fedoseev, The growth of diamond and graphite from the gas phase (Nauka, 1997) (in Russian)
26. H.W. Kroto, J.R. Heath, S.C. O'Brien, R.F. Kurl, R.E. Smalley, C_{60}: Buckminsterfullerene. Nature **318**, 162–163 (1985)
27. V.G. Pol, S.V. Pol, A. Gedanken, Semiconducting, magnetic or superconducting nanoparticles encapsulated in carbon shells by RAPET method. Carbon Sci. Technol. **1**, 46–56 (2008)
28. S. Chikazumi, S. Taketomi, M. Ukita, M. Mizukami, H. Miyajima, M. Setogawa, Y. Kurihara, Physics of magnetic fluids. J. Magn. Magn. Mater. **65**, 245–251 (1987)
29. S. Mornet, S. Vasseur, F. Grasset, P. Verveka, G. Goglio, A. Demourgues, J. Portier, E. Pollert, E. Duguet, Magnetic nanoparticle design for medical applicaqtions. Prog. Solid State Chem. **34**, 237–247 (2006)
30. M. Leonowicz, M. Wozniak, Y.M. Shulga, V.E. Muradyan, Z. Liu, H.A. Davies, W. nKaszuwara, J. Grabski, Structure and magnetic properties of nanoparticles encapsulated in carbon shells, J. Magn. Magn. Mater. **294**, 57–62 (2005)
31. S.H. Huh, A. Nakajima, Laser synthesis and magnetism of amorphous iron and cobalt carbide nanoparticles with carbon onion. J. Appl. Phys. **99**, 064302 (2006)
32. F. Shoji, Z. Feng, A. Kono, T. Nagai, Spherical carbon particle growth in a methane plasma. Appl. Phys. Lett. **89**, 171504 (2006)
33. T. Hyeon, Chemical synthesis of magnetic nanoparticles. Chem. Commun. **8**, 927–934 (2003)
34. R.H. Fowler, L. Nordheim, Electron emission in intense electric fields. Proc. R. Soc. Lond. A **119**, 173–181 (1928)
35. T.E. Stern, B.S. Gossling, R.H. Fowler, Further studies in the emission of electrons from cold surfaces. Proc. R. Soc. Lond. A **124**, 699–723 (1929)
36. E.L. Murphy, R.H. Good, Thermionic emission, field emission and transition region. Phys. Rev. **102**, 1464–1573 (1956)
37. K.L. Jensen, *Introduction to the Physics of Electron Emission* (Wiley, Chichester, UK, 2018)
38. R.G. Forbes, J.H.B. Deane, Transmission coefficients for the exact triangular barrier: an exact general analytical theory that can replace Fowler & Nordheim's 1928 theory, Proc. R. Soc. Lond. A **467**, 2927–2947 (2011). See electronic supplementary material for definitions of FN constants
39. R.G. Forbes, Field electron emission theory (October 2016), pp. 31–39 in: *Proc. 2016 Young Researchers in Vacuum Micro/Nano Electronics Conf., St Petersburg, October 2016*, ed. by N.V. Egorov and K.A. Nikiforov) (Electronic ISBN 978-1-5090-4605-8) (IEEE Explore, March 2017); arXiv:1801.0825v2
40. A. Modinos, *Field, Thermionic, and Secondary Electron Emission Spectroscopy* (Plenum Press, New York and London, 1984)
41. N.V. Egorov, A.A. Almazov, Optimization of multi-tip field emission electron source. Vacuum **52**, 295–300 (1997)

42. A.N. Tichonov, A.A. Samarskii, *Equations of Mathematical Physics* (Pergamon, Oxford, 1963)
43. D.A. Ovsyannikov, N.V. Egorov *Mathematical Modelling of Systems of Formation of Electron and Ion Beams* (St. Petersburg State University Press, St. Petersburg, 1998) (in Russian)
44. A.A. Almazov, N.V. Egorov, On the method of calculation of field emission systems. Radiotechnika i Electronika **31**, 2452 (1986). (in Russian)
45. R. Gomer, *Field Emission and Field Ionization* (Harvard Univ. Press, Cambridge, Mass., 1961), see p. 32
46. R.G. Forbes, C.J. Edgcombe, U. Valdrè, Some comments on models for field enhancement. Ultramicroscopy **95**, 57–65 (2003)
47. D. Biswas, A universal formula for the field enhancement factor. Phys. Plasma **25**, 043113 (2018)
48. V.A. Zorich, *Mathematical Analysis II* (Springer, Berlin, 2004)
49. F.R. Abbott, J.E. Henderson, The range and validity of the field current equation. Phys. Rev. **56**, 113–118 (1939)
50. G.F. Vasilev, Influence of the shape and potential barrier at the emitter-vacuum interface and distribution of surface electric field on the type of current-voltage characteristics of field emission. Radio Eng. Electron. **5**, 1857–1861 (1960). (in Russian)
51. P.F. Bovey, 100 kV high resolution analytical STEM. Microsc. Spectrosc. Electron. **1**, 507–508 (1976)
52. N.N. Kalitkin, *Numerical Methods* (Nauka, Moscow, 1978). (in Russian)
53. O. Gröning, O.M. Küttel, Ch. Emmenberger, P. Gröning, L. Schlapbach, Field emission properties of carbon nanotubes. J. Vac. Sci. Technol., B **18**, 665–678 (2000)
54. R.G. Forbes, Physical electrostatics of small field emitter arrays/clusters. J. Appl. Phys. **120**, 054302 (2016)
55. H.G. Kosmahl, Analytic evaluation of field emission enhancement factors for ellipsoidal codes and elliptic cross-section wedges. IEEE Trans. Electron Dev. **38**, 1534–1537 (1991)
56. C.J. Edgcombe, U. Valdrè, Microscopy and computational modelling to elucidate the enhancement factor for field electron emitters. J. Microsc. **203**, 188–194 (2001)
57. T.A. de Assis, F.F. Dall'Agnol, Microscopy and computational modelling to elucidate the enhancement factor for field electron emitters. J. Phys.: Condens. Matter **30**, 195301 (2018)
58. R.G. Forbes, Physical explanation of the universal "(separation)$^{-3}$" law found numerically for the electrostatic interaction between two protruding nanostructures, arXiv:1803.031
59. J.R. Harris, K.L. Jensen, D.A. Shiffler, AIP Adv. **5**, 087182 (2015)
60. F.S. Baker, A.R. Osborn, J. Williams, The carbon-fiber field emitter. J. Phys. D Appl. Phys. **7**, 2105–2115 (1974)
61. V.D. Blank, G.A. Dybitsky, E.P. Sheshin, A.L. Shornikova, Field emission properties of fullerenes, Proc. 8th Intern. Vacuum Electron Sources Conf. and Nanocarbon, Nanjing, October 2010 (IEEE Explore, 2010), pp. 164–165
62. W.B. Choi, D.S. Chung, H.Y. Kim. J.H. Kang, J.E. Jun, J.M. Kim. Carbon nanotube based field emission displays, Technical Digest, 12th International Vacuum Microelectronics Conf. July, Darmstadt, Germany, July 1999 (ISBN 3-00-004512-0), pp. 310–311
63. A.L. Musatov, Yu.V. Gulyaev, N.A. Kiselev, E.F. Kukovitsky, A.I. Zhbanov, K.R. Izrael'yants, E.G, Chirkova, A.B. Ormont, A.V, Boykov, Low voltage field electron emitters based on nanotube carbon layers grown by CVD method: experiment and calculations, Proc. 4th International Vacuum Electron Sources Conf., July 2002 (IVESC'02), Saratov, Russia (ISBN: 5-99409-015-4), pp. 81–82
64. V.M. Lobanov, E.P. Sheshin, S.V. Lobanov, Field emission of coherent electrons from carbon nanotubes and fibers, in Proc. 10th International Vacuum Electron Sources Conf. & 2nd International Conf. on Emission Electronics, Saint Peterburg, June-July 2014 (IEEE, Piscataway, 2014) (ISBN: 978-1-4799-5770-5), ed. by N.V. Egorov, pp. 144–145
65. Y. Saito (ed.), *Carbon Nanotube and Related Field Emitters* (Wiley, Weinheim, 2010)

66. S.K. Gordeev, A.M. Lamanov, R.M. Ibragimov, K.N. Nikolskiy, E.P. Sheshin, Pyrolytic carbon cathodes prepared by low temperature vapor deposition, Technical Digest, 18th International Vacuum Nanoelectronics Conf., Oxford, July 2005 (IEEE, Piscataway, 2005) (ISBN: 0-7803-8397-4) pp. 176–177
67. A. Knápek, D. Sobola, P. Tománek, A. Pokorná, M. Urbánek, Field emission from the surface of highly ordered pyrolytic graphite. Appl. Surf. Sci. **395**, 157–161 (2017)
68. A.S. Baturin, E.P. Sheshin, Field emission characteristics of reticulated vitreous carbon, Technical Digest, 13th International Vacuum Microelectronics Conf., Guangzhou, China, August 2000, pp. 42–43
69. A.G. Chakhovskoi, C.E. Hunt, G. Forsberg, R.T. Nilsson, P. Persson, Reticulated vitreous carbon field emission cathodes for light source applications. J. Vac. Sci. Technol., B **21**, 571–575 (2003)
70. E.P. Sheshin, A.V. Anashchenko, S.G. Kuzmenko, Field emission characteristics research of some types of carbon fibers. Ultramicroscopy **79**, 109–114 (1999)
71. E.P. Sheshin, A.S. Baturin, K.N. Nikolskiy, R.G. Tchesov, V.R. Sharov, Field emission cathodes based on milled carbon fibers. Appl. Surf. Sci. **251**, 196–200 (2005)
72. H.W. Aung, E.P. Sheshin, W.Z. Hlaing, N.C. Kyaw, Field emission properties of polyacrylonitrile (PAN) carbon fibers of various processing temperatures, Proc. 2019 International Vacuum Electronics Conf, (IVEC), Busan, South Korea, April-May 2019 (IEEE Explore, 2019)
73. A.S. Baturin, I.N. Yeskin, A.I. Trufanov, N.N. Chadaev, E.P. Sheshin, R.G. Tchesov, Electron gun with field emission cathode of carbon fiber bundle. J. Vac. Sci. Technol., B **21**, 1–4 (2003)
74. M.S. Mousa, M.J. Hagmann, M. Brugat, E.P. Sheshin, Measurements of the self- sustained enhancement of field emission by carbon fiber microemitters. Ultramicroscopy **95**, 119–124 (2003)
75. W.W. Tang, D.A. Shiffler, J.R. Harris, K.L. Jensen, K. Golby, M. LaCour, T. Knowles, Field emission characteristics of a small number of carbon fiber emitters. AIP Adv. **6**, 095007 (2016)
76. N.V. Luparev, E.P. Sheshin, N.N. Chadaev, G.Z. Otarsvili, S.K. Gordeev, S.B. Korchagina, Diamond-carbon nanocomposites as material for cold cathode planar-type emitters, Proc. 8th Intern. Vacuum Electron Sources Conf. and Nanocarbon, Nanjing, October 2010 (IEEE Explore, 2010), pp. 213–214
77. S. Groznov, A. Leychenko, E.P. Sheshin, A. Schuka, Flat displays based on field emission cathodes, "Chipnews", , pp. 21-–25 (2008) (in Russian)
78. E.P. Sheshin, A.L. Suvorov, A.F. Bobkov, D.E. Dolin, Light source on the basis of carbon field electron cathodes: design and parameters, Abstracts of 7th International Vacuum Microelectronics Conf., Grenoble, France, July 1994 (Supplement to "Le Vide, les Couches Minces" No. **271**, Apri11994), pp. 423–426
79. A.S. Baturin, V.S. Kaftanov, S.G. Kuzmenko, E.P. Sheshin, Field electron emission device, Patent RF 2180145, class H01J/28, 15.02.2000
80. MYu. Leshukov, A.S. Baturin, N.N. Chadaev, E.P. Sheshin, Characterizations of light sources with carbon fiber cathodes. Appl. Surf. Sci. **215**, 260–264 (2003)
81. M.Yu. Leshukov, N. N. Chadaev, A. S. Baturin,, E. P. Sheshin, Vacuum cathodoluminescent lamps for information display systems, Proc. Conf. on « Modern Television » , Moscow 2004, pp. 36–38 (in Russian)
82. W. Knapp, O.F. Kieler, D. Schleussner, E.P. Sheshin, I.N. Yeskin, Feldemission von Kohlenstoff-Fasern und deren Anwendung in einem CRT-Lichtelement, Konferenzband Electronic Displays 2000, Berlin, pp. 92–98
83. W. Knapp, D. Schleussner, A.S. Baturin, I.N. Yeskin, E.P. Sheshin, CRT lighting element with carbon field emitters. Vacuum **69**, 339–344 (2003)
84. H. Murakami, M. Hirakawa, C. Tanaka, H. Yamakawa, Field emission from well-aligned, patterned, carbon nanotube emitters. Appl. Phys. Lett. **76**, 1776–1778 (2000)
85. Y. Saito, K. Hamaguchi, R. Mizushima, S. Uemura, T. Nagasako, J. Yotani, T. Shimojo, Field emission from carbon nanotubes and its application to cathode ray tube lighting elements. Appl. Surf. Sci. **146**, 305–311 (1999)

86. V.S. Kaftanov, A.L. Suvorov, E.P. Sheshin, Field emission cathode and methods in the production thereof, US patent No. 5588893, (classification H01J 9/02), appl. 06.06.1995, publ. 31.12.1996
87. V.S. Kaftanov, A.L. Suvorov, E.P. Sheshin, Field emission cathode and methods in the production thereof, US patent No. 5973446, (classification H01J 1/05), 15.02.1999
88. G. Forsberg, C.-H. Andersson, Method of producing a field emission cathode and a light source, Patent. PCT WO 00/77813 A1, 10.06.1999
89. V.S. Kaftanov, A.L. Suvorov, E.P. Shesin, J. Olsfors, Field emission cathode and a light source including a field emission cathode, US patent No. 5877588, 05.03.1999, (H01J 1/30)
90. A.A. Blyabli, A.V. Kandidiov, J.A. Ankelevich, A.A. Pilevskii, A.T. Rakhimov, V.A, Samorodov, Divergence of electrons emitted from a carbon cold cathode, Technical Digest, 12th International Vacuum Microelectronics Conf,, Darmstadt, Germany, July 1999 (ISBN 3-00-004512-0), pp. 346–347
91. J.-M. Bonard, T. Stockli, O. Noury, A. Chatelain, Field emission from cylindrical carbon nanotube cathodes: possibilities for luminescent tubes. Appl. Phys. Lett. **78**, 2775–2777 (2001)
92. X.X. Zhang, C.C. Zhu, X. Li, W.Z. Cui, The fabrication of novel structure of field emitting light tube with carbon nanotubes as cathode, Technical Digest, 18th International Vacuum Nanoelectronics Conf., Oxford, July 2005 (IEEE, Piscataway, 2005) (ISBN: 0-7803-8397-4), pp. 342–343
93. J.X. Huang, J. Chen, S.Z. Deng, J.C. She, N.S. Xu, Optimization of carbon nanotube cathode for a fluorescent lamp, Technical Digest, 18th International Vacuum Nanoelectronics Conf., Oxford, July 2005 (IEEE, Piscataway, 2005) (ISBN: 0-7803-8397-4), pp. 284–285
94. J. Kjellman, M. Lindmark, Light source, and a field emission cathode, US patent 6873095, 29.03.2005, (H01J 1/16)
95. A.N. Obraztsov, A.P. Volkov, A. Zakhidov et al., Field emission characteristics of nanostructured thin film carbon materials. Appl. Surf. Sci. **215**, 214–221 (2003)
96. A.N. Obraztsov, Cathodoluminescent light source, patent appl. US 005/0174059, 11.08.2005 (H01J 63/04)
97. A.S. Leychenko, M.Yu. Leshukov, N.N. Chadaev, E.P. Sheshin, Effective lamp for LCD-backlighting with the field emission cathode, Technical Digest, 19th International Vacuum Nanoelectromics Conf. & 50th International Field Emission Symp., Guilin, China (IEEE, Piscataway, 2006) (ISBN: 1-4244-0410-0), pp. 383–384
98. A.S. Leychenko, M.Yu. Leshukov, N.N. Chadaev, E.P. Sheshin, Field emission backlight for LCD, Proc. Conf. on "Modern Television", Moscow, 2006, pp. 30–31 (in Russian)
99. P. Liu, Y. Wei, L.M. Sheng, et al., Field emission luminescent light source, US patent appl. 2006/0091782, 04.03.2006 (H01J 63/04)
100. A.I. Trufanov, A.S. Baturin, M.Yu. Leshukov, N.N. Chadaev, E.P. Sheshin, Emission characteristics of a light source with field cathode based on carbon fiber bundle "Microsystem Technics" pp. 32–35 (2004) (in Russian)
101. A.S. Leychenko, M.Yu. Leshukov, E.P. Sheshin, et al. Element of a large video screen with cathodoluminescent light sources as pixels, Proc. Conf. on "Modern Television", Moscow, 2007, pp. 45–46 (in Russian)
102. A.V. Kudryashov, E.P. Sheshin, N.N. Chadaev, M.M. Kustikov, The element of a big size screen based on light emitters with field emission cathodes, 7th International Vacuum Electron Sources Conf. (IVESC 2008), London, August 2008
103. J. Cho, J.H. Park, J.K. Kim, E.F. Schubert, White light-emitting diodes: History, progress and future. Laser Photonics Rev. **11**, 1600147 (2017)
104. J. Lindén, C. Dam-Hansen, *OLEDs—State of the art report (DTU* (Depatment of Photonics Engineering, Roskilde, 2019)
105. S. Groznov, A. Leychenko, E.P. Sheshin, A. Schuka, Flat display screens based on field electron emission cathodes, "Chipnews", **7**, pp. 21–25 (2008) (in Russian)
106. K.S. Choi, S.J. Lee, Jae Myung Kim, C.W. Kim, J.W. Nam, J.C. Cha, Y.R. Do, H.G. Yang, Jong Min Kim, FED devices containing a novel graphite cathode prepared by a screen printing process, Technical Digest, 12th International Vacuum Microelectronics Conf., Darmstadt, Germany, July 1999 (ISBN 3-00-004512-0), pp. 32–33

107. K.S. Choi, S.J. Lee, Jae Myung. Kim, C.W. Kim, J.W. Nam, J.C. Cha, Y.R. Do, H.G. Yang, Jong Min Kim, Field emission display devices containing a novel graphite cathode prepared by a screen printing process, SID Symp. Digest of Technical Papers **31(1)**, 671–673 (2000)
108. 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, Fully sealed, high-brightness carbon-nanotube field-emission display, Appl. Phys. Lett. **75**, 3129–3131 (1999)
109. J.M. Kim, W.B. Choi, N.S. Lee, J.E. Jung, Field emission from carbon nanotubes for displays. Diam. Relat. Mater. **9**, 1184–1189 (2000)
110. A.Y. Tcherepanov, A.G. Chakhovskoi, V.B. Sharov, Flat panel display prototype using low-voltage carbon field emitters. J. Vac. Sci. Technol., B **13**, 482–486 (1995)
111. N.S. Xu, Z.S. Wu, S.Z. Deng, J. Chen, High-voltage triode flat-panel display using field–emission nanotube-based thin films. J. Vac. Sci. Technol. B **19**, 1370–1372 (2001)
112. T. Yaguchi, T. Muneyoshi, M. Okai, N. Hayashi, T. Nakamura, Flat panel display device, US Patent appl. 2005/0057178, 17.03.2005, (G 09 G 3/10)
113. T. Oh, Field emission display, European Patent 1542258, disclosure 15.06.2005, classif.: H01j 31/12
114. K. Oono, Field emission display device and method of manufacturing same, US Patent appl. US 2004/0239235, 02.12.2004, (H01j 1/62)
115. J.H. Kang, Y.S. Choi, W.B. Choi, N.S. Lee, et al., Under-gate triode type field emission displays with carbon nanotube emitters, Mat. Res. Soc. Symp., Proc. **621**, R5.2.1–R5.2.5 (2000)
116. Y.S. Choi, J.H. Kang, Y.J. Park, W.B. Choi, C.J. Lee, S.H. Jo, C.G. Lee, J.H. You, J.E. Jung, N.S. Lee, J.M. Kim, An under-gate structure field emission display with carbon nanotube emitters. Diam. Relat. Mater. **10**, 1705–1708 (2001)
117. Y.S. Choi, J.H. Kang, H.Y. Kim, B.G. Lee, C.G. Lee, S.K. Kang, Y.W. Jin, J.W. Kim, J.E. Jung, J.M. Kim, A simple structure and fabrication of carbon-nanotube field emission display. Appl. Surf. Sci. **221**, 370–374 (2004)
118. V.P. Mammana, G.E. McGuire, O.A. Shenderova, Back-gated field emission electron source, US patent application 2005/0116214, 02. 06.2005, (H01 L 29/06)
119. V.I. Mahov, B.V. Bondarenko, M.F. Kopyilov, Microwave device of M-Type, Patent of Russian Federation № 2040821, 27.07.1995, classification: H01j 1/30 (in Russian)
120. J.E. Lilienfeld, The autoelectronic discharge and its application to the construction of a new form of X-ray tube. Amer. J. Roentgenology **9**, 172–179 (1922)
121. T. C. Cunningham, Field emission x-ray tube having a graphite fabric cathode, US Patent 3883760, 07.04.1971 (H01J 35/06)
122. E.P. Sheshin, N.N. Chadaev, A.S. Baturin, A.I. Trufanov, X-ray tube with field emission cathode, Patent of Russian Federation No. 8248643, 20.03.2005, classification H01J 35/02 (in Russian)
123. https://www.jeol.de/electronoptics-en/products/transmission-electron-microscopes-tem.php [observed: 12 Feb 2020]
124. A.V. Crewe, Scanning Electron Microscope, US patent No. 3191028, 22.6.1965
125. J.W. Butler, *Digital computer techniques in electron microscopy, 6th Internat* (Kyoto, Japan, Congress on Electron Microscopy, 1966), p. 191
126. M. Okai, K. Hidaka, M. Hayashibara, S. Watanabe, Field emission electron gun and electron beam applied device using the same", US Patent appl. 2008/0029700, 07.02.2008, (H01J 37/073)
127. P.W. Hawkes, E. Kasper, *Principles of Electron Optics, Vol. II: Applied Geometrical Optics* (Academic, London, 1996)
128. D.I. Swan, D. Kynastou, The development of a field emission SEM, Proc. 6th Annual Scanning Electron Microscope Symp., 1973, N1, p. 57
129. Y.A. Grigoriev, A.I. Petrosyan, V.V. Penzyakov, V.G. Pimenov, V.I. Rogovin, V.I. Shesterkin, V.P.Kudryashov. V.C. Semyonov, Experimental study of matrix carbon field-emission cathodes and computer aided design of electron guns for microwave power devices, exploring these cathodes, J. Vac. Sci. Technol. B **15**, 503–506 (1997)

130. B.V. Bondarenko, V.I. Makuha, E.P. Sheshin et al., Field electron emission gun, Certificate of Authorship, USSR Patent 1294188, (H01J 1/30), 11.03.1985 (in Russian)
131. V.A. Golubkov, A.S. Ivanov, V.A. Ilyin, V.V. Luchinin, S.A. Bogdanov, V.V. Chernov, A.L. Vikharev Stabilizing effect of diamond thin film on nanostructured silicon carbide field emission array, J. Vac. Sci. Technol. B **34**, 062202 (2016)
132. O.A. Ivanov, S.A. Bogdanov, A.L. Vikharev, V.V. Luchinin, V.A. Golubkov, A.S. Ivanov, V.A. Ilyin, Emission properties of undoped and boron-doped nanocrystalline diamond films coated silicon carbide field emitter arrays. J. Vac. Sci. Technol. B **36**, 021204 (2018)
133. K. Nikiforov, V. Trofimov, N. Egorov Field emission spectroscopy of silicon carbide: natural modeling, 2015 International Conference on "Stability and Control Processes" in Memory of V.I. Zubov, SCP 2015—Proceedings, 7342085, pp. 180–182 (2015)
134. N.I. Alekseev, V.V. Luchinin, N.A. Charykov, Evaporation of carbon atoms from the open surface of silicon carbide and through graphene cells: Semiempirical quantum-chemical modeling. Russ. J. Phys. Chem. A **87**, 1830–1837 (2013)
135. N.I. Alekseev, V.V. Luchinin, N.A. Charykov, Initial stage of the epitaxial assembly of graphene from silicon carbide and its simulation by semiemprical quantum chemical methods: Carbon face. Russ. J. Phys. Chem. A **87**, 1709–1720 (2013)
136. N.I. Alekseev, V.V. Luchinin, N.A. Charykov Simulating the conditions for the formation of graphene and graphene nanowalls by semiempirical quantum chemical methods, Russ. J. Phys. Chem. A **87**, 1721–1730 (2013)
137. N.I Alekseev, A.A. Kal'nin, D.D. Karmanov, V.V. Luchinin, S.A. Tarasov, N.A. Charykov, Epitaxial assembly of graphene on face (0001) of silicon carbide: Modeling by semi-empirical methods, Russ. J. Phys. Chem. A **87**(1739–1748 (2013)
138. J.-M. Bonard, H. Kind, T. Stöckli, L.-O. Nilsson, Field emission from carbon nanotubes: the first five years. Solid-State Electronics **45**, 893–914 (2001)
139. A. Mayer, N.M. Miskovsky, P.H. Cutler, Theoretical comparison between field-emission properties of carbon protrusions ranging from an isolated atom to multi-wall nanotubes, J.Phys.: Condens. Matter **15**, R177–R191 (2003)
140. N. de Jonge, J.-M. Bonard, Carbon nanotube electron sources and applications. Phil. Trans. R. Soc. Lond. A **363**, 2239–2266 (2004)
141. M.P. Anantram, F. Léonard, Physics of carbon nanotube electronic devices. Rep. Prog. Phys. **69**, 507–561 (2006)
142. F. Le Normand, C.S. Cojocaru, C. Fleaca, J.Q. Li, P. Vincent, G. Pirio, L. Gablof, Y. Nedellec, P. Legagneux, A comparative study of the field emission properties of aligned carbon nanostructures films, from carbon nanotubes to diamond. Eur. Phys. J. Appl. Phys. **38**, 115–127 (2007)
143. A.V. Eletskii, G.S. Bocharov, Emission properties of carbon nanotubes and cathodes on their basis. Plasma Sources Sci. Technol. **18**, 034103 (2009)
144. N. de Jonge, Carbon nanotube electron sources for electron microscopes, Chap. 3 in: Adv. Imaging Electron Phys., Vol. 156 (Elsevier, 2009) pp. 203–233
145. F. Léonard, *The Physics of Carbon Nanotube Devices* (William Andrew, 2009). See Chap. 6
146. A. Zhbanov, E. Pogolerov, Y.-C. Chang, Carbon nanotube field emitters, Book chapter downloadable from: www.intechopen.com (published 2010)
147. T. Zhai, L. Ma, M. Liao, X. Wang, X.-S. Fang, J. Yao, Y. Bando, D. Golberg, One-dimensional inorganic nanostructures: synthesis, field-emission and photodetection. Chem. Soc. Rev. **40**, 2986–3004 (2011)
148. W. Lu, M. Zu, J.-H. Byun, B.-S. Kim, T.-W. Chou, State of the art of carbon nanotube fibers: opportunities and challenges. Adv. Mater. **24**, 1805–1833 (2012)
149. G.S. Bocharov, A.V. Eletskii, Theory of carbon nanotube (CNT)-based electron field emitters. Nanomaterials **3**, 393–442 (2013)
150. A. Nojeh, Carbon nanotube electron sources: from electron beams to energy conversion and optophotonics. ISRN Nanomaterials **2014**, 879827 (2014)
151. G. Mittal, I. Lahiri, Recent progress in nanostructured next-generation field emission devices. J. Phys. D Appl. Phys. **47**, 323001 (2014)

152. Y. Li, Y. Sun, J.T.W. Yeow, Nanotube field electron emission: principles, development, and applications. Nanotechnology **26**, 242001 (2015)
153. R.J. Parmee, C.M. Collins, W.I. Milne, M.T. Cole, X-ray generation using carbon nanotubes. Nano Convergence **2**, 1 (2015)
154. M.T. Cole, M. Mann, K.B.K. Yeo, W.I. Milne, Engineered carbon nanotube field emission devices, Chap. 5 in: *Emerging Nanotechnologies for Manufacturing* (*2nd Edition*), ed. by W. Ahmed, J.M. Jackson (William Andrew, 2015)
155. M.T. Cole, M. Nakamoto, W.I. Milne, Field emission displays (FEDs) and surface-conduction electron-emitter displays (SEDs), in: *Handbook of Digital Imaging* (Wiley, 2015)
156. L. Chen, H. Yu, J. Zhong, L. Song, J. Wu, W. Su, Graphene field emitters: a review of fabrication, characterization and properties. Mater. Sci. Eng. B **220**, 44–58 (2017)
157. F. Giubileo, A. Di Bartolomeo, L. Iemmo, G. Luongo, F. Urban, Field emission from carbon nanostructures. Appl. Sci. **8**, 526 (2018)
158. M. Schulz, V. Shanov, Z.-Z. Yin, M. Cahay (editors) *Nanotube Superfiber Materials*, 2nd Edition, Elsevier, 2019; containing Chap. 21: M. Cahay, K.L. Jensen, R.G. Forbes, et al., Optimising the field-emission properties of carbon nanotube-based fibers; and Chap. 22: W. Zhu, M. Cahay, J. Ludwick, et al., Multiscale modelling of field emission properties of carbon nanotube based fibers

Chapter 11
Explosive Electron Emission of Carbon-Based Cathodes, and Applications

Georgiy N. Fursey

Abstract This chapter is devoted to the study of field electron (FEE) and explosive electron (EEE) emissions from carbon nanoclusters and considers the prospects of their use in the development and production of portable X-ray sources. It was found earlier that the macroscopic fields that are necessary to excite the FEE and the EEE from nanoclusters are 2–3 orders of magnitude less than those needed for conventional metal and semiconductor emitters. Recent results, related to the study of portable X-ray sources operating on the basis of explosive electron emission, are highlighted. Particular portable X-ray devices, with an explosive emission cathode made from carbon nanoclusters, are described.

11.1 Introduction

The main problem in developing portable X-ray devices stems from the necessity to apply high voltages in order to accelerate electrons to the energies required for producing bremsstrahlung, i.e., characteristic X-ray emission. This need for high voltage requires high electric insulation strength for all components used in X-ray devices. This can be difficult to achieve when high voltages are applied for long time intervals.

The problem can be solved by applying the high voltage for a very short time of the order of 10^{-9} to 10^{-7} s, i.e., by operating in the nanosecond pulsing range. The idea of constructing nanosecond-range pulsed portable X-ray devices was proposed by Tsukerman [1, 2]. In the nanosecond range, the physical linear size of a high-voltage power source could be reduced by approximately an order of magnitude, and correspondingly, the physical volume will be reduced by a factor of 10–100 times.

G. N. Fursey (✉)
The Bonch-Bruevich Saint-Petersburg, State University of Telecommunications, Saint Petersburg, Russia
e-mail: g.fursey@gmail.com

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*,
Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_11

For the X-ray image (radiogram) to be registered in the nanosecond range, one needs to increase the electron flux generating the X-rays by about a million times, and this is another problem to be solved. In conventional X-ray tubes these currents are in the range 0.1–10 mA and the exposure times in X-ray diagnostics, in particular in medicine, are of the order of 0.2–1 s. To obtain a sufficient number of X-ray quanta in the nanosecond pulsing regime, one needs to increase the tube electron current to hundreds of amperes. None of the older types of electron emission (including thermionic emission) was able to provide such current values.

However, one can expect that the required emission current densities (and hence currents) could be achieved by using field electron emission (FEE). It was precisely this consideration that stimulated first Walter Dyke with coworkers [3–5], and later Tsukerman, to use field emission to obtain short X-ray pulses for studying superfast processes [6].

Tsukerman, guided by the obvious opportunities of the field emission, attempted to develop a portable X-ray apparatus. It turned out, that in the devices of Dyke [4, 7], and also those of Tsukerman [1, 6], the decisive role was played not by an ordinary field emission mechanism, but by an absolutely new phenomenon that was given the name of **explosive electron emission** (EEE) [8]. The discovery of EEE opened new frontiers for the production of nanosecond X-ray devices.

It was clear that special conditions had to be established to ensure the reproducibility of the EEE process [8, 9]. Deep understanding of how to exploit EEE cathodes made possible the production of X-ray tubes with a satisfactory service lifetime. The significant majority of the pulsed nanosecond tubes use till now refractory metal cathodes (Mo, W, etc.) [10–13].

Direct experiments have shown [14] that, during EEE, nanoscale and larger (microscale) protrusions are formed on the metal cathode surface. The “large” micrometer protrusions lead to basic erosion of the cathode surface, mass transfer in the gap, particle deposition onto the anode, and many other detrimental phenomena that affect the performance of X-ray sources.

Comparatively recently [9, 15, 16], it has been shown that carbon [8, 17] is a much more promising material for EEE cathodes, particularly when in the form of carbon nanoclusters. In many respects, they look to be much more promising devices than metal emitters.

X-ray tubes with carbon cathodes possess a high resource, ensure higher precision in the formation of the X-ray source and make possible the production of X-ray devices with an electron beam of much broader energy range, i.e., with a broader wavelength spectrum of X-ray emission than could be attained with metal cathodes.

During operation, carbon EEE cathodes produce much less contamination of the anode and other components of an X-ray tube. By using EEE from carbon materials, one can reduce the mass transfer by a factor of 10 or more. Further, in the context of particulate contamination of the anode or the exit window in a through-target X-ray tube, carbon (being a lighter material with small atomic number) absorbs a much smaller fraction of X-rays than any metal.

11.2 Experimental

11.2.1 Explosive Electron Emission (EEE)

Explosive electron emission (EEE) is a peculiar emission process that occurs when nano-sized volumes of cathode substance explode in the combined presence of a strong electrostatic surface field and the resulting high field-emission current density.

This phenomenon was discovered independently by two groups of researchers. G.N. Fursey and his coworkers studied [18] the upper limiting FEE current density and the physical mechanisms relating to vacuum breakdown and the cathode spot of a vacuum arc. The other group, headed by Mesyats [19], studied the dynamics of light emission during vacuum breakdown, using nanosecond resolution; these studies are described in detail in several publications [20–22]. A model of EEE evolution was suggested in [18] (also see [8, 9]), and was experimentally substantiated by Mesyats and coworkers [19] (see also his book [20, p. 424]).

Figure 11.1 illustrates the transition from field emission (FEE) to explosive emission (EEE), and Fig. 11.2 schematically illustrates the formation of EEE centers across the cathode surface.

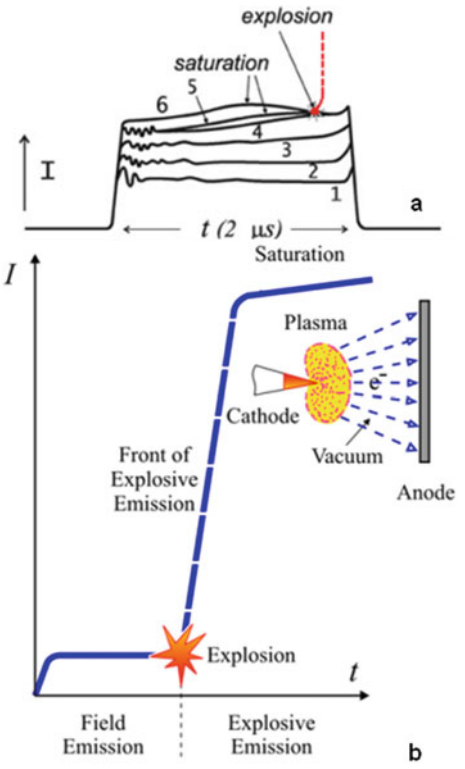


Fig. 11.1 **a** Behavior of a field emission current before an explosion, for a single pointed tungsten emitter. Curves 1, 2, 3, 4, 5 and 6 are oscillograms of the FEE current as the voltage is increased; "saturation" is the transition to the saturation stage; "explosion" is the transition point to the EEE (the build-up of electron current at the instant of the explosion is shown by a dashed line). **b** Scheme (model) of the transition from field emission to explosive emission for a single point emitter

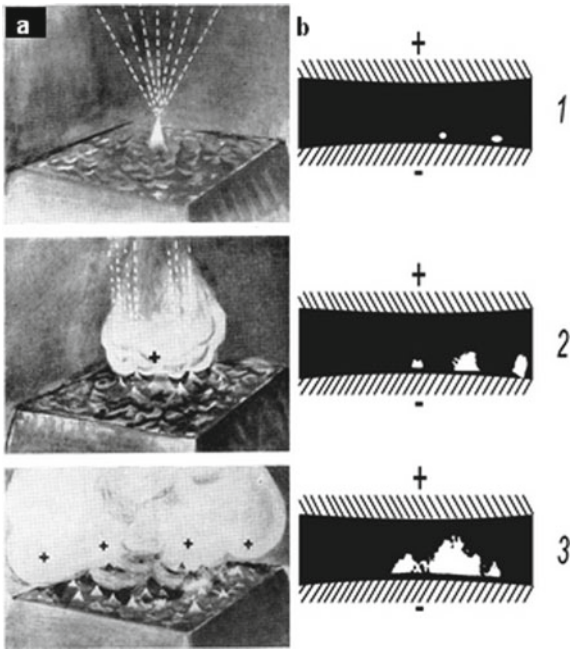


Fig. 11.2 Initiation and development of the explosive electron emission process: **a** evolution of the EEE process according to the Fursey's model; **b** diagrams 1, 2, 3 are consecutive stages of the EEE process in the direct experiments on plasma expansion, carried out by Mesyats

Special experiments have been carried out, using paired voltage pulses of equal amplitudes [23, 24], in order to determine the maximum frequency and minimum duty cycle needed to carry out sequential switching of the EEE current pulses. Oscillograms of the current pulses obtained at different duty cycles are shown in Fig. 11.3a. It has been found that the minimum duty cycle at which one can obtain fully identical current pulses is in the range 200–500 ns, depending on the features of the EEE cathode. Figure 11.3b shows the time dependence of the complete recovery of the EEE process. By definition, the minimum duty cycle of EEE-cathode operation is two successive nanosecond pulses [23, 24].

The process of EEE is quasi-stationary, in the sense that the conditions for EEE at the emitter surface are self-reproducing. Micro-roughness on the cathode surface is both the cause of and a result of EEE [9, 18]. In EEE, an important role is played by a liquid phase on the cathode surface. This forms as a result of the explosion of surface material and resultant heating. The liquid phase interacts with the strong electric field and with the plasma generated by the EEE, and forms nonlinear surface nano-waves: the ridges of these nano-waves then become new centers of EEE [14]. Solidification creates the micro-roughness that causes the next EEE event with the next electric field pulse.

Carbon, in this respect, possesses unique properties. It has been found that during EEE carbon also transforms into a liquid phase [14, 16, 22]. This implies that in the epicenter of the explosion the pressure may be as high as 10^4 bar (10^9 Pa).

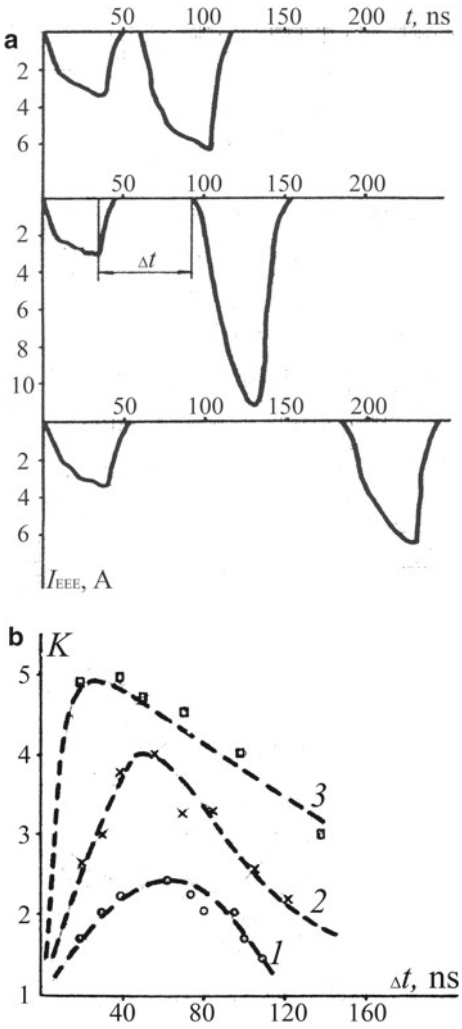


Fig. 11.3 Experimental determination of the minimum duty cycle for adjacent FEE pulses: **a** dependence of the EEE current amplitude on pulse spacing; **b** graph illustrating the recovery of the EEE initial stage, depending on the time interval between pulses for a carbon nano-cluster cathode. The coefficient $K = I_2/I_1$ is the ratio of FEE currents in the second (I_2) and first (I_1) pulses, $\underline{U}_1$ is the EEE excitation voltage

An advantage of a carbon-based emitter is that it is much easier to retain the nano-ridge relief on its surface, compared to metals. One can believe that, after an explosion, nano-ridges solidify not because the local cathode regions are cooling, but rather because of the pressure drop related to the process of plasma spreading. Since the pressure drops much faster than the surface cools down, after the electric field is switched off, it is possible to retain much smaller surface asperities with carbon surfaces than with metal surfaces. This ensures more uniform emission from the surface, and reduces the electric field strength required for excitation of EEE.

Table 11.1 Threshold fields F_{on} for onset of field emission. For explanation of symbols, see text

$F_{\text{on-macro}}$(C-modif.)	$\approx 10^5$–10^6 V/m
$F_{\text{on-tip}}$(C-modif.)	$\approx 10^7$–10^8 V/m
$F_{\text{on-tip -FEE}}$(metals)	$\approx 3\times10^9$ V/m

11.2.2 Low Threshold Field and Explosive Electron Emission

In recent years one more property of carbon nanoclusters, nanotubes, and graphene has been revealed. It has been established that the excitation threshold of field emission for these materials, particularly for nanotubes and graphene, is very low. For instance, the fields at which the onset of field emission occurs were found to be 100–1000 times less than that for metals and semiconductors [9, 22, 25–30]. (For further details, see Sect. 8.6 (p. 161) in [30].) A low FEE threshold has also been detected for some other materials, for instance, diamond-like films [31].

The values of the different types of the electric field associated with the excitation of field emission (for an onset current of 10^{-10} to 10^{-9} A) are compared in Table 11.1.

In Table 11.1, $F_{\text{on-macro}}$ is the electrical field between the macroscopic electrodes with a carbon cathode; $F_{\text{on-tip}}$ is the electrical field taking into account the enhancement produced by the microtips; $F_{\text{on-tip-FEE}}$ is the local surface electrical field required for the excitation of FEE from metals.

As noted above, the main EEE excitation mechanism is the explosive destruction of local cathode areas under the action of a field emission current of very high density. Since the threshold voltage for field emission excitation is lower for the carbon nanoclusters, it follows that EEE is also excited at lower voltages.

Thus, it seems that a highly important property of a carbon nano-cluster, certainly in the context of EEE, is the formation of micro/nano-roughness (ridges) on the cathode surface. These ridges have a nano-sized radius of curvature, and this causes additional field enhancement and thus localization of emission into very small nano-sized areas of the cathode. These features provide opportunities for the development of low-voltage electron sources for EEE.

11.2.3 Development of the X-Ray Source

In the study described here, we consider two X-ray source types: a source with a tungsten needle anode (Fig. 11.4a); and a through-target source with an anode in the form of a tungsten layer coating the exit window (Fig. 11.4b).

As a cathode, we used various carbon nanoclusters, such as graphene, nanotubes, and polyacrylonitrile (PAN) fiber. Both of the designs were realized first in a special test chamber (Fig. 11.5), and then in the form of autonomous sealed–off X-ray tubes (Fig. 11.7).

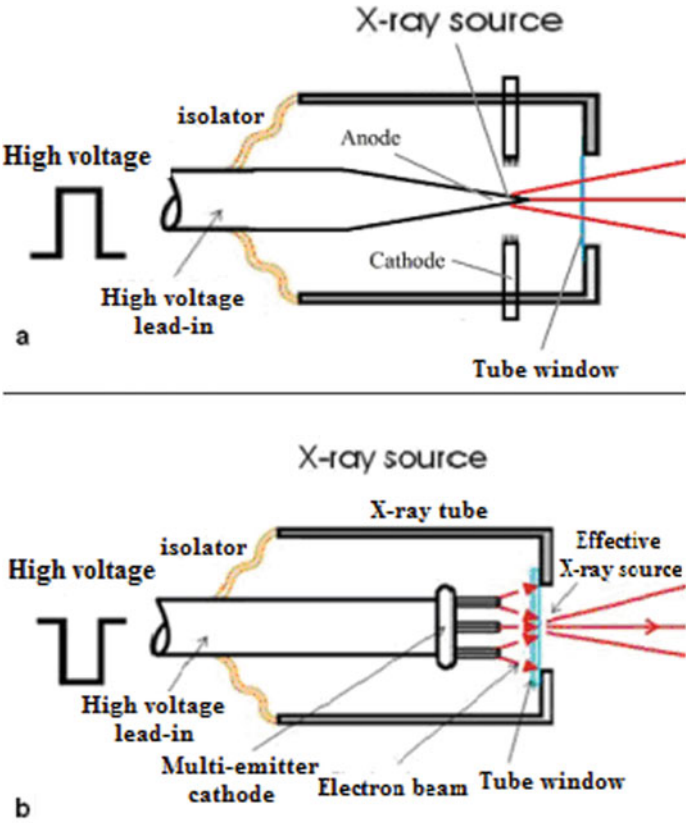


Fig. 11.4 Pulsed X-ray sources. **a** Pulsed X-ray tube with a needle anode. **b** X-ray tube with through-target anode

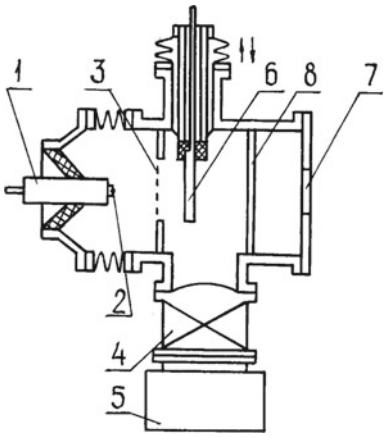


Fig. 11.5 Construction of the experimental chamber: 1 = high-voltage bushing; 2 = emitter; 3 = anode-grid; 4 = vacuum pumping; 5 = sorption pump; 6 = electron collector; 7 = window for photography; 8 = fluorescent screen

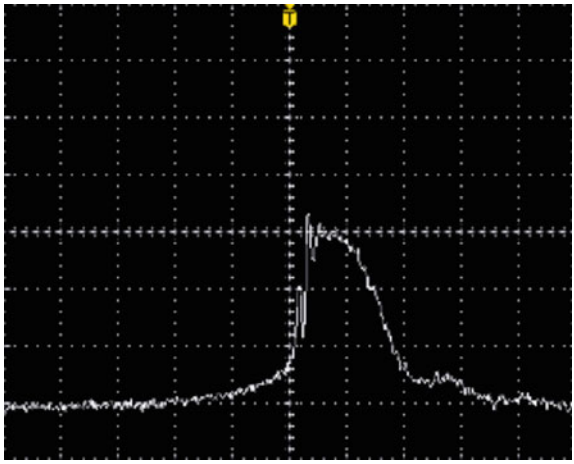


Fig. 11.6 Oscillogram of a high-voltage (75 kV) nanosecond pulse. Waveform calibration is as follows: along the horizontal axis 10 ns per division, along the vertical axis 25 kV per division

The experimental setup allows displacement of the cathode relative to the anode, using the micrometer drive, thus changing the distance between the cathode and anode. The exit window has been made of beryllium foil. A high voltage power supply is provided with special nanosecond generators.

For a dynamical study of EEE evolution, we used a generator of double (paired) nanosecond pulses [23]. We also used generators manufactured by FID Technology [32, 33].

In our experiments, the pulse duration was varied from 10 to 80 ns, and the time between double pulses from 2 to 700 ns. The pulse amplitude was in the range from 15 to 150 kV, and the pulse rise-time was 1–2 ns. A typical pulse waveform is shown in Fig. 11.6. The vacuum level in the experiments was kept at 10^{-9} to 10^{-8} Torr (10^{-7} to 10^{-6} Pa).

11.2.4 Autonomous X-Ray Tubes—Experimental Parameters of the Tubes

On the basis of the preliminary studies, we have developed and constructed experimental prototypes of autonomous sealed-off X-ray tubes (Fig. 11.7). As an anode, we took a sharpened tungsten rod with a tip radius of 0.3 mm. The exit window was made of Beryllium plate with a thickness of 400 μm to 1 mm, and the output window of the same material had a diameter of 20 mm. Vacuum in the sealed-off tube was of the order of 10^{-8} to 10^{-7} Torr (1.3×10^{-6} to 1.3×10^{-5} Pa).

The construction of our tubes was based on the standard X-ray tubes with an anticathode [11, 34]. In these tubes, a conventional metallic cathode has been replaced by a cathode made of carbon nanoclusters [17, 35].

Fig. 11.7 X-ray tube with a glass insulator

The electron current in the tube has been determined by means of a special matching coaxial shunt induced in the circuit of reverse current, and also with an oscilloscope. The electron current flowing in the diode is ultimately limited by space-charge effects, and cannot exceed the Child-Langmuir limit. With metals, the shape of a current pulse is determined by a (3/2) power law, which for a diode with a moving emission boundary [16] has the form

$$I_e(t_0) \sim \frac{U^{3/2}}{\sqrt{m_e}(d - V_{pl}t_0)^2} \tag{11.1}$$

Here U is the voltage; d is the distance between the cathode and anode; V_{pl} is the velocity of plasma expansion; and I_{e} (t_0) is the current in the cathode-anode gap.

Maximum currents in the tube vary, in accordance with the voltage, in the range from 30 A at $U = 20$ kV, up to 200 A at $U = 75$ kV. The shape of the X-ray pulse emission was determined with the aid of a photodiode.

11.3 Results and Discussion

11.3.1 Determination of the Focal Spot Size—Divergence of the X-Ray Flux

The divergence of the X-ray beam depends on the configuration of the EEE cathode, the geometry of the gap between the cathode and the anticathode, and the geometry of the opaque (for the X-ray flux) exit section of the X-ray tube. In addition, the divergence could be varied with the aid of a special external collimator.

In our case, for the tubes with the pointed type anticathode, used in various applications, including medical ones, the divergence was about 60º. It allows the production of panoramic images with a safety margin of 40 × 40 cm at a standard distance of 70 cm.

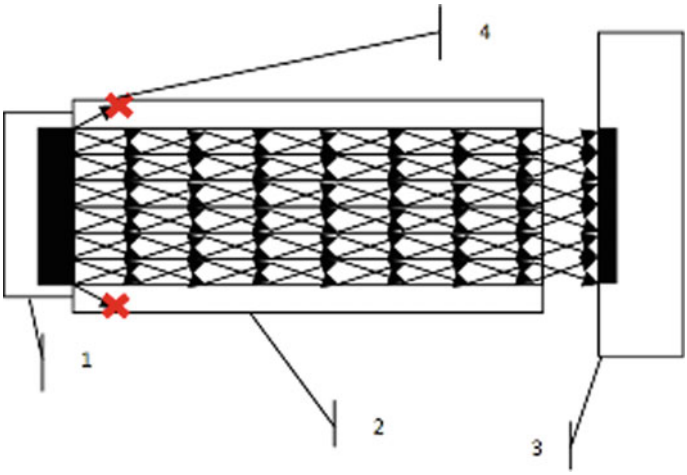


Fig. 11.8 The scheme for the transfer and determination of the X-ray spot size: 1 = an exit window of the X-ray source; 2 = poly-capillary column; 3 = visualizer for the detection of X-ray spots; 4 = the rays outside of the observation angle

The size of the X-ray spot was measured with the use of a pinhole camera [34], and in a series of experiments the poly-capillary Kumakhov pole [36] (see Fig. 11.8) was also used. In our experiments with the pointed anticathode, the minimum size of the X-ray spot was measured to be about 0.8 mm.

11.3.2 Portable X-Ray Apparatus

We have designed and produced a portable X-ray apparatus [9, 37], on the basis of the X-ray tube with a carbon cathode of PAN fiber, as developed by our group. The apparatus is shown in Fig. 11.9a, together with an anticathode. A photograph of the X-ray tube is shown in Fig. 11.9b. The summary Table 11.2 contains the basic parameters of the apparatus..

For the detection of X-ray radiation using the Radix apparatus, we explored three detector types: X-ray film, memory plates, and digital flat-panel (FP) detectors, using the indirect transformation of X-ray radiation with scintillators of sulfur oxide or gadolinium. The FP detector approach was found to be the best. At present, this type of detector is often used for digital X-ray receivers. The technology gets X-ray photons to initially interact with a scintillator; the resulting light is then transformed into an electrical signal.

For synchronization of this detector with the portable X-ray generation apparatus, a control unit with two wireless transmission line channels has been designed. It also has been used for the control of the detector: switching it off/on; controlling the

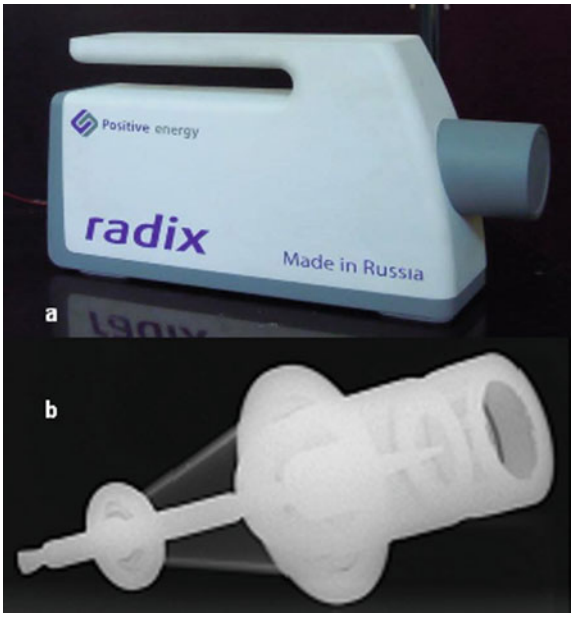


Fig. 11.9 **a** The portable X-ray device, and **b** an X-ray image of our X-ray tube

Table 11.2 Parameters of the X-ray apparatus

Accelerating voltage	15–150 kV
Beam current	50–500 A
X-ray pulse duration	10 ns
Repetition rate	up to 1000 Hz
Spot diameter	1 mm
Total weight	3 kg

regime of expectation of X-ray radiation; and managing the transfer of the digital image to a workstation display.

Control of the whole system is fully integrated by means of a developed software suite called “Radix Center”. This is able to operate the X-ray generator and detectors, and carry out various supporting tasks, including pre- and post-treatment of the images, calibration of the detector, and system optimization (for example, to minimize the noise level and reduce the number of artifacts, such as faulty pixels).

Various X-ray images obtained with our portable apparatus are shown in Figs. 11.10–11.16.

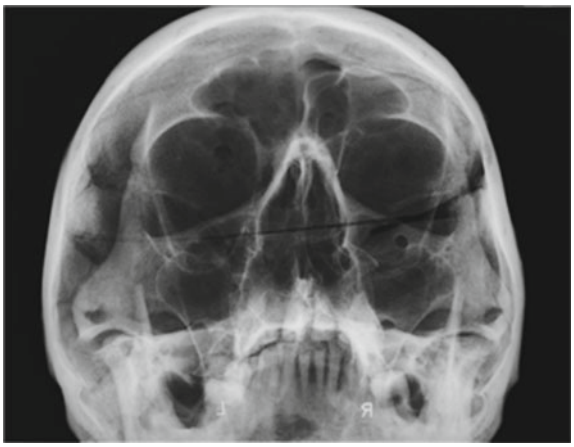


Fig. 11.10 Radiograph of a human skull

Fig. 11.11 X-ray image of a human hand

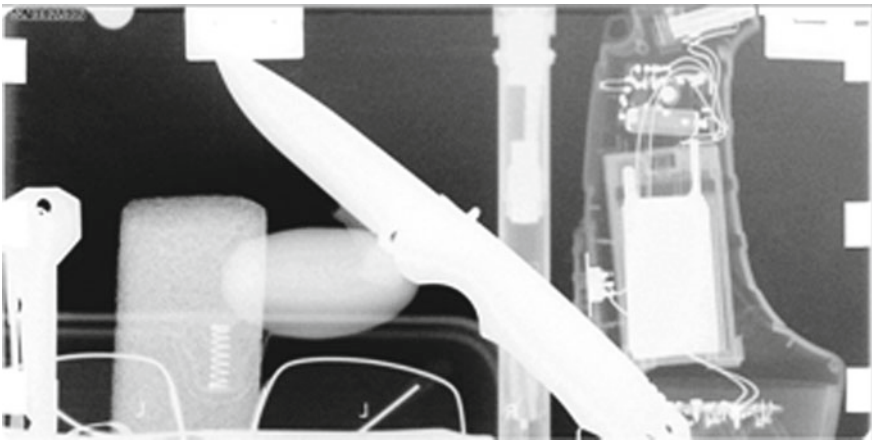


Fig. 11.12 Inside of a briefcase

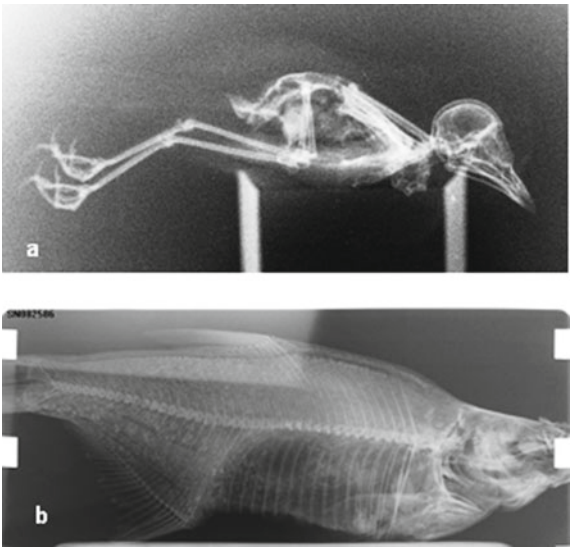


Fig. 11.13 **a** Bird and **b** fish

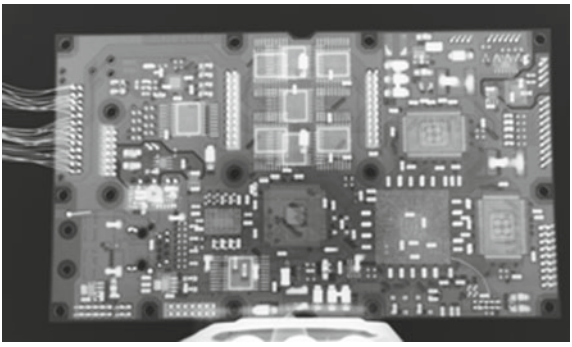


Fig. 11.14 Circuit board

11.3.3 Opportunities for External Focusing of Pulsed X-Ray Radiation

The main obstacle affecting the quality of our X-ray images is the comparatively large size of the X-ray spot (1–2 mm). In pulsed X-ray tubes operated using EEE, reduction of the spot by means of internal electromagnetic focusing is troublesome, because of the small distance between cathode and anode. A more attractive approach is direct focusing of the X-ray radiation after exit from the X-ray tube. This seems to be possible in principle, by using the poly-capillary X-ray optics proposed by Kumakhov [38–40] (in particular, see pp. 170–182 in [38]). As shown in a number of recent studies, the Kumakhov lens in principle allows X-ray focusing on a spot with a diameter of less than 6 μm.

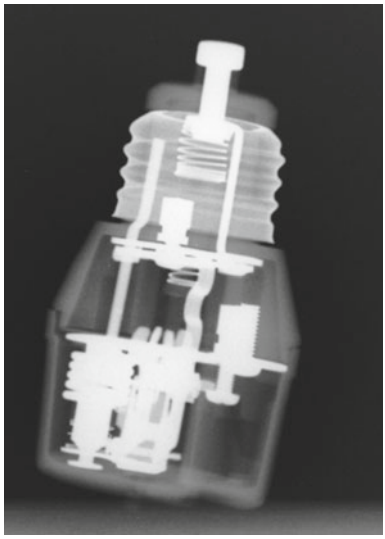


Fig. 11.15 Automatic fuse

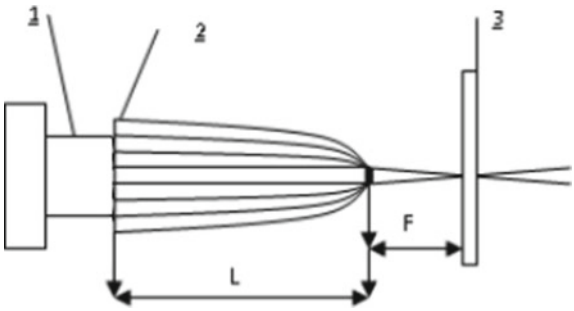


Fig. 11.16 The principle of external focusing: 1 = head part of X-ray tube; 2 = poly-capillary half-lens; 3 = visualizer; focal length F = 160 mm, and lens length L = 268 mm. Diameter of the half-lens inlet is 11.3 mm. Diameter of the half-lens outlet is 6 mm

Efficient performance of these lenses has been achieved in a number of devices with stationary X-ray sources.

This study, continuing our previous experiments, now reports attempts [41] to apply this kind of optics to our powerful pulsed X-ray sources driven by EEE in the nanosecond range.

The principle of external focusing rests on the phenomenon of total external reflection. In the X-ray spectral range, the refractive index $n < 1$ because a vacuum is optically the more dense medium. The reflection property is highest at glancing angles of incidence. Figure 11.16 shows the principal scheme of focusing by the aid of poly-capillary optics.

Some results of our studies are now given. Two experiments were carried out. In the first we used the X-ray source (X-ray tube) with a pointed (needle) anticathode (= anode). In this case, we managed to compress the X-ray spot by a factor of approximately 2 (see Fig. 11.17a). In the second, the experiment was carried out

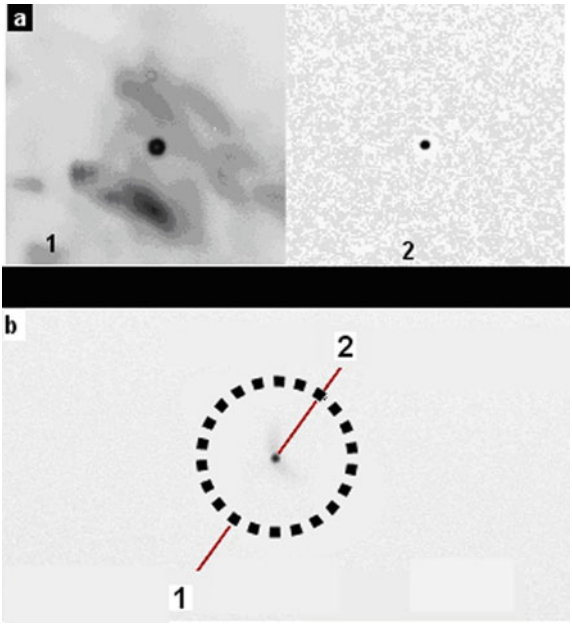


Fig. 11.17 Results of X-ray focusing with the aid of a Kumakhov half-lens. **a** The X-ray spot obtained with the X-ray tube using a pointed anti-cathode: 1 = the central black spot is the original X-ray spot; 2 = the focused X-ray spot. **b** X-ray spot obtained with the through-target X-ray tube: 1 = dotted circle outlines the original X-ray spot; 2 = the focused X-ray spot

with the through-target X-ray tube. Here, the entry orifice of the external half-lens was placed directly at the tube exit window. The diameter of the X-ray spot on the window was 11 mm, and was approximately equal to the diameter of the entrance to the capillary half-lens. On exit from the capillary half-lens, the spot diameter was measured to be 0.5 mm (see Fig. 11.17b).

The results obtained from these experiments show that the external focusing of pulsed X-ray sources seems to have considerable promise.

11.4 Summary and Conclusions

The results reported in the chapter can be summarized as follows:

1. We have briefly analyzed the mechanism of explosive electron emission (EEE) from a viewpoint of contemporary understanding of this phenomenon.
2. The special features of EEE, when it occurs from carbon nanoclusters, have been discussed.
3. X-ray tubes, developed on the basis of EEE from carbon nanoclusters, have been thoroughly studied.
4. A new family of portable pulsed X-ray tubes has been proposed; these have attractive potential applications.
5. It was shown, that the external focusing of X-ray radiation in pulse mode is perspective.

In conclusion, I would like to mention that the community of scientists engaged in the development of portable X-ray instrumentation is quite numerous. I would especially like to recognize the studies performed by Komyak, Mesyats, Peliks, and Belkin [10, 13].

Acknowledgements The author and his coworkers are glad to express their deep gratitude to B.P. Merkulov for his generous offer to use his ceramic X-ray tube to carry out our studies of EEE cathodes, and to Prof. Aleksina for the opportunity to test our apparatus in the Anatomic museum of the Saint-Petersburg First Medical University. We are indebted to the Head of the Firm "Spectroflash", Dr. E.A. Peliks, for fruitful discussions, and also to the Firm "Positive Energy" for comprehensive support of the research described. We offer special thanks to D.N. Fomin for his valuable technical assistance. The author is also grateful to Dr. Richard G. Forbes for critical reading of the manuscript.

References

1. V.A. Tsukerman, L.V. Tarasova, S.I. Lobov, New sources of X rays. Sov. Phys. Uspekhi **14**, 61–71 (1971)
2. V.A. Tsukerman, Portable sources of X-ray beams, Vestnik AN USSR **1971**(11), 18–25 (1971). (in Russian)
3. W.P. Dyke, J.P. Barbour, F.J. Grundhauser, VII Int. Kongr. Kurzzeitphotogr. Zurich, Switzerland (1965)
4. W.P. Dyke, Advances in field emission. Sci. Am. **210**(1), 108–118 (1964)
5. F.G. Grundhauser, W.P. Dyke, S.D. Bennet, A fifty-milli-microsecond flash X-ray system for high-speed radiographs. J. SMPTE **70**, 435–439 (1961)
6. V.A. Tsukerman, M.A. Manakova, Short X-ray flash sources for investigating fast processes. Zh. Tekh. Fiz. **27**, 391–403 (1957). (in Russian)
7. W.P. Dyke, W.W. Dolan, Field emission. Adv. Electron. Electron Phys. **8**, 89–185 (Academic, New York, 1956)
8. G.N. Fursey, Field emission in microelectronics, in: *Field Ion and Field Electron Microscopy and Spectroscopy: History, Achievements, the Current State of the Art, and Prospects*, ed. by A.L. Suvorov (Akademprint, Moscow, 2003), pp. 170–189. (in Russian)
9. G.N. Fursey, Intensive electron emission in a strong electric fields in vacuum microelectronics and high-power electronics. Nanotechnol. Percept. **9**, 167–185 (2013)
10. E.A. Pelix, X-ray pulse apparatus based on explosive electron emission, in 17th World Conference on Nondestructive Testing, Shanghai, China, October 2008
11. G.V. Melnichuk, B.P. Merkulov, D.S. Makhanko, V.G. Samorodov, Development of pulsed metal-ceramic explosive emission tubes with improved X-ray technical characteristics, defense industry achievements. Russ. Sci. Tech. Prog. **2014**(2), 53–67 (2014)
12. G.N. Fursey, E.A. Pelix, The new generation of medical x-ray discrete instrumentation. Pract. High Power Electron. **44**, 47–51 (2011)
13. A.A. Altuhov, N.V. Belkin et al., High power pulse X-ray apparatus, "Kavkaz" for non-destructive control. Defectoscopiya **11**, 12–17 (1989). (in Russian)
14. G.N. Fursey, L.A. Shirochin, L.M. Baskin, Field emission processes from liquid-metal surface. J. Vac. Sci. Technol. B **15**, 410–421 (1997)
15. G.N. Fursey, M.A. Polyakov, A.A. Kantonistov, A.M. Yafyasov, B.S. Pavlov, V. B. Bozhevol'nov, Field and explosive emissions from graphene-like structures. Tech. Phys. **58**, 845–851 (2013)
16. G.A. Mesyats, D.I. Proskurovsky, *Pulsed Electrical Discharge in Vacuum* (Springer, Berlin, 1989)

17. G.N. Fursey, L.A. Shirochin, P.N. Bespalov, X-ray tube. Patent of Russian Federation No 2308781, Bull. No. 29 (20 October 2007)
18. G.N. Fursey, P.N. Vorontsov-Vel'yaminov, A qualitative model of vacuum arc initiation. Zh. Tech. Fis. **37**, 1870–1888 (1967). (in Russian)
19. S.P. Bugaev, A.M. Iskol'ski, G.A. Mesyats, D.I. Proskurovskii, Electron-optical observation of initiation and development of pulsed breakdown of short vacuum gap. Zh. Tech. Fiz. **37**, 2206–2208 (1967). (in Russian)
20. G.A. Mesyats, *Cathode Phenomena in a Vacuum Discharge: The Breakdown, the Spark, and the Arc* (Nauka, Moscow, 2000)
21. G.A. Mesyats, G.N. Fursey, Explosive field emission at the initial stages of vacuum discharge, in *Cold Cathodes*, ed. by M.I. Elinson (Sovetskoe Radio, Moscow 1974). (in Russian)
22. G.N. Fursey, Field emission in vacuum microelectronics. Appl. Surf. Sci. **215**, 113–134 (2003)
23. B.I. Grishanov, L.A. Shirochin, G.N. Fursey, S.M. Lupekhin, M.A. Polyakov, V.M. Zhukov, High-voltage generator of variable nanosecond pulses for study of explosive electron emission. Instrum. Exp. Tech. **29**, 851–854 (1987)
24. G.N. Fursey, C.M. Lupekhin, M.A. Polyakov, L.M. Baskin, L.A. Shirochin, Dynamics of explosive emission processes. Doklady Akademii Nauk (DAN) USSR, **276**, 866–869 (1984). (in Russian)
25. G.N. Fursey, *Field Emission in Vacuum Microelectronics* (Kluwer Academic/Plenum, New York, 2005)
26. N.I. Sinitsyn, Yu.V. Gulyaev, N.D. Devjatkov et al., Carbon nanocluster structures as one of materials of emission electronics in the future. Radiotekhnika **2**(9) (2000). (in Russian)
27. A.N. Obraztsov, Dual-barrier tunneling during field emission from nano-graphite, in Joint Conference IVESC-ICEE-ICCTPEA-BDO-2014, Saint-Petersburg, Russia (30 June–04 July 2014)
28. Al.A. Zakhidov, A.N. Obraztsov, A.P. Volkov, DA. Lyashenko, Mechanism of low-voltage field emission from nanocarbon materials. J. Exp. Theor. Phys. **100**, 89–94 (2005)
29. A.N. Obraztsov, IYu. Pavlovsky, A.P. Volkov, Low-voltage electron emission from chemical vapor deposition graphite films. J. Vac. Sci. Technol., B **17**, 674–678 (1999)
30. A. Yafyasov, V. Bogevolnov, G. Fursey, B. Pavlov, M. Polyakov, A. Ibragimov, Low-threshold field emission from carbon nano-clusters. Ultramicroscopy **111**, 409–414 (2011)
31. A.T. Rakhimov, V.A. Samorodov, E.S. Soldatov et al., Study of emission and structure characteristic correlation of diamond films by scanning tunneling microscopy. Poverkhnost **7**, 47 (1999). (in Russian)
32. V. Efanov, M. Efanov, Gigawatt all solid state FID pulsers with nanosecond pulse duration, in Procedings of IEEE International Power Modulators and High Voltage Conference, Las Vegas, May 2008, p. 381
33. V.M. Efanov, P.M. Yarin, A.V. Kricklenko, V.A. Mankevich, High voltage nanosecond pulsers for plasma chemistry and gas discharge, in Proceedings 2008 IEEE International Power Modulators and High Voltage Conference, Las Vegas, USA, May 2008 (ISBN-978-1-4244-1534-2) (IEEE, Piscataway, 2008), pp. 178–180
34. M.A. Polyakov, G.N. Fursey, L.A. Shirochin, A.A. Kantonistov, An X-ray technique for investigation of the explosive-emission cathode surface. Tech. Phys. Lett. **34**, 586–587 (2008)
35. B.P. Merkulov, G.N Fursey, D.S. Makhan'ko, M.A. Poljakov, Pulse type X-ray tube, Patent of Russian Federation No 2308781, Bull. No. 18 (27.06.2014). (in Russian)
36. A.A. Kumakhov, I.V. Dmitriev, Methods of measuring X-ray sources and neutron radiation using poly-capillary optics. J. Appl. Phys. **81**, 85–87 (2011)
37. L.A. Shirochin, G.N. Fursey, High-power soft X-ray tube with an explosive emission cathode. In: Proceedings of XVIII International Symposium on Discharges and Electrical Insulations in Vacuum, Eindhoven, Netherlands, August 1998, pp. 672–674
38. M.A. Kumakhov (ed.), Proceedings of International Conference on X-ray and Neutron Capillary Optics, Zvenigorod, Russian Federation, September 2001, SPIE Vol. 4765
39. V.A. Arkadiev, D.I. Gruev, M.A. Kumakhov, X-ray lens for forming quasiparallel beam, in *Optics of Beams*, ed. by M.A. Kumakhov (IROS, Moscow, 1993), pp. 27–32

40. G.I. Borisov, M.A. Kumakhov, Poly-capillary lens for neutrons. Nucl. Instr. Meth. Phys. Res. A **529**, 129–133 (2004)
41. G.N. Fursey, I.A. Dmitriev, Y.B. Skuratnik, A.A. Begidov, A.A. Kumakhov, Focusing possibility of high-power nanosecond X-ray pulses. Prakticheskaya silovaya elektronika **2**(46), 50–55 (2012). (in Russian)

Chapter 12
Spindt Cathodes and Other Field Emitter Arrays

Georg Gaertner and Wolfram Knapp

Abstract Based on the introduction of semiconductor technologies, the application of thin-film and micromachining techniques to fabricate integrated vacuum field emission devices in the submicron size range became feasible in the 1960s. These were pioneered at SRI by C. Spindt and colleagues, who first introduced gated field emitter arrays. There are several previous extended reviews on this topic published until 2001 [1]. The present chapter presents a short wrap up of fabrication techniques and structures, and a performance update on Spindt arrays and on field emitter arrays (FEAs) in general. This chapter will only deal with regular array structures and their specific advantages and problems; random structures will not be discussed. A critical evaluation of progress and of application of FEAs in devices is given.

12.1 Introduction

Several reviews on field emitters have already been published. One of the important ones is the book "Vacuum Microelectronics" edited by Wei Zhu in 2001 [1], which starts with a historical overview. The different approaches are addressed in detail, including Spindt field emitter arrays [2], and also a review of field emission theory written by Jensen [3]. There also exists an older review by Spindt et al. from 1992 [4], a review by Cade and Lee from 1990 [5], and a review by D. Temple addressing applications in more detail [6]. In the present book, field emission (=FE) of carbon emitters has already been addressed in Chap. 10 and the same authors have also written a review on field emission from different materials and their applications [7]. Therefore, this chapter will provide an update on FEAs (= field emitter arrays) and

G. Gaertner (✉)
Consultant, Aachen, Germany
e-mail: georg.f.gaertner@t-online.de

W. Knapp
Otto von Guericke University, Magdeburg, Germany
e-mail: wolfram.knapp@ovgu.de; dr.who.knapp@t-online.de

Knapptron GmbH Vacuum Electronics, Möser, Germany

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*,
Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_12

their performance after 2001, and also discuss problems associated with application in vacuum electron tubes.

It is well known that field emission from a small tip can yield very high current densities, yet from a very small emission area. Therefore, field emitters usually are well suited for high brightness and electron beam applications. FE happens at room temperature if the field strength is sufficiently high, and no heating of the emitter is needed. One of the main advantages of field emitters is low energy consumption. For FE to compete with thermionic emission, small beam currents are not sufficient; hence arrays of field emitters are needed to deliver higher currents for applications in vacuum tubes. Due to the breakthrough of microfabrication technology developed for large-scale integration of micron—and submicron solid state devices, there was a technology at hand which also could be used for vacuum microelectronic devices, especially for field emitter arrays. This was first realized by Cap Spindt at SRI in the group of Ken Shoulders in 1968 [1, 2]. This first Spindt cathode consisted of a multi-layer structure of Mo gate, SiO_2 insulator, and Mo tips on a Si substrate prepared by thin-film deposition techniques. In the following years, methods and performance were continuously improved by C. Spindt and colleagues, also by an increase of the total number of emitters and hence an increase of total current which can be drawn from such arrays.

12.2 Spindt Field Emitter Arrays

12.2.1 History of Spindt Emitters

Already in the 1950s, Ken Shoulders and Dudley Buck at MIT had proposed to employ thin-film and micromachining techniques to fabricate integrated vacuum field effect devices of 100 nm sizes. Ken Shoulders started to realize these ideas later at SRI in 1958, with support from the US Department of Defense, and with the aid of Cap Spindt and his colleagues [1, 2]. C. Spindt developed the process for fabricating arrays of minute field emitter cones by using thin-film technology and electron beam microlithography, which soon showed first emission results. At 20 V between gate and cone, the onset of electron emission was detected and at 100 V, several μA of current could be drawn from these structures, first reported at the 1966 IEEE Conference on Tube Techniques [8, 9]. Shoulders left SRI in 1968 and the FEA work at SRI continued at a low level till 1973, when the group got support from the NASA Lewis Research Center under the direction of R. Forman. At the same time, Ivor Brodie joined SRI as director of the Applied Physics Laboratory, and thus the development work on FEAs could be taken up again. When in 1985 Robert Meyer and his group at LETI in France announced preliminary results with flat panel field emission displays (FEDs) using Spindt arrays [10], this triggered increased efforts worldwide. In 1987, Chris Holland of SRI reported the first three-color FED

panel [11]. It was realized, that information exchange between the different groups was needed and hence the first International Vacuum Microelectronics Conference (IVMC, since 2004 IVNC = International Vacuum Nanoelectronics Conference) was organized in Williamsburg, Virginia, in 1988. The second IVMC took place in Bath in the UK in 1989 with over 160 delegates and 54 presentations, where most of the contributions dealt with the realization of structures or theory and only very few performance data were reported, as the first author noticed (he had presented the only thermionic cathode contribution [12]).

12.2.2 Fabrication Technology

12.2.2.1 The Original Spindt Process

The original (1928) Fowler–Nordheim (=FN) theory of electron emission from metals under intense electric fields predicted a relation of the emission current I (measured in A) for an applied voltage U (measured in V) of the form of (12.1) for a planar one-dimensional case under the assumption of a triangular barrier. In 1956, this was corrected and replaced (12.2, 12.3) by Murphy and Good, but still in the general form (12.1) and still a one-dimensional theory:

$$I = aU^2 \exp(-b/U), \tag{12.1}$$

where a and b are parameters (also called Fowler–Nordheim parameters) approximately given by

$$a \cong 1.5 * 10^{-6}(\beta^2 A_E/\phi) * \exp(10.4/\sqrt{\phi}) \tag{12.2}$$

$$b \cong 6.44 * 10^7(\phi^{3/2}/\beta) \tag{12.3}$$

with A_E the emitting area in cm^2, ϕ the cathode work function in eV, β the geometric voltage conversion factor in cm^{-1}, which determines the electric field at the cathode as $F_e = \beta\ U$. In this context, the authors recommend it as good practice, not to use the same symbol for a physical quantity and its unit. Better modern approximations and theoretical formulations are discussed in Chaps. 8 and 9.

In order to obtain high currents at low voltages, the work function ϕ should be low and A_E and β should be high. β is approximately inversely proportional to the radius r of curvature of the tip, if r is much smaller than the distance D to the anode. Thus reducing the cathode to anode distance and minimizing the tip radius will increase the field strength at the tip [2, 3].

In the modified original Spindt process [2, 4, 8] first, a Mo layer was evaporated onto a sapphire substrate, then followed by an alumina layer and a second Mo layer.

Polystyrene balls of 0.5 μm diameter were then sprayed onto the layer structure and over-coated with evaporated Al (see Fig. 12.1). After washing the balls away, Mo was chemically etched, using Al as a resist mask, forming holes into the Mo top-layer at the former position of the balls. An etch of the alumina layer, by using Mo as resist mask, then formed cavities below the holes. In a further step, Mo was evaporated from a vertical evaporator through the holes, whereas parallel a horizontal evaporator of alumina closed the holes, and cone-shaped deposits of Mo were formed. Finally, the liftoff surface layer of alumina was dissolved by phosphoric acid. In order to avoid dissolution of the insulating alumina layer, a temperature treatment at 1000 °C was introduced before cone deposition, so that the phosphoric acid had practically no effect on the hardened insulating alumina layer. A disadvantage of this initially used method is that the patterns with apertures and cones are random, including the distances between the emitter tips.

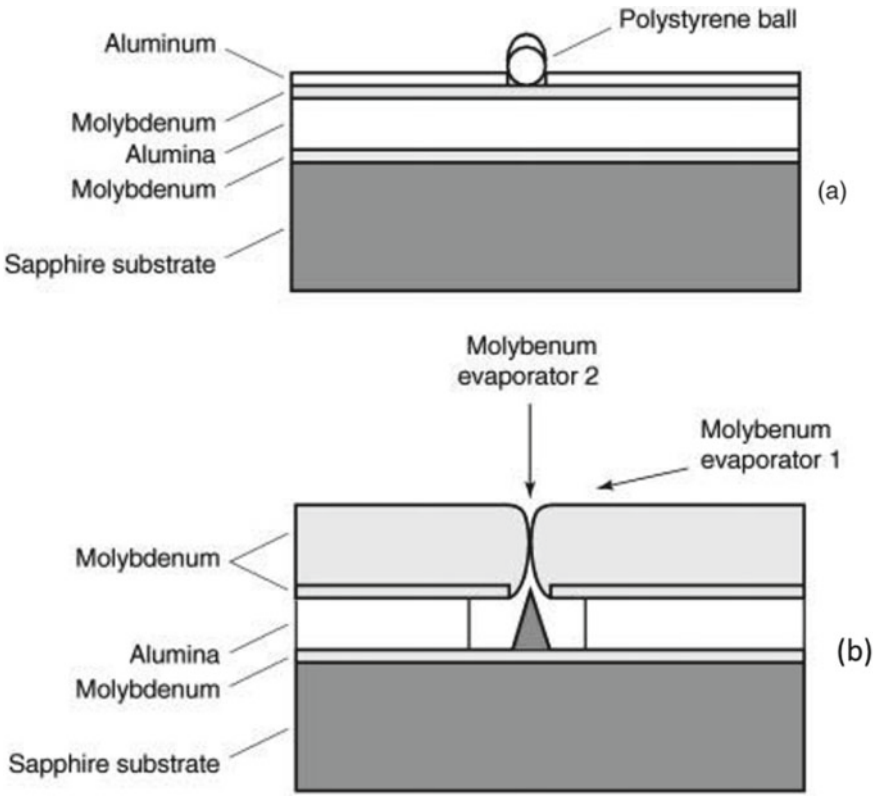


Fig. 12.1 The original process: **a** after first layer depositions, before washing and etching; **b** emitter cone deposition; according to Spindt et al. [2], reprinted with permission from chapter 4 of Vacuum Microelectronics, Wiley 2001

12.2.2.2 The Improved Process

Therefore, a new method of patterning by electron beam lithography was applied, forming a regular pattern of holes, and then followed by steps of cone formation as described above [13, 14].

This was not the only improvement in the modified new process. First silicon wafers were used as the substrate, where the conductivity could be adjusted in view of the application. Then silicon was oxidized to obtain the insulating layer. In the third step, the gate metal (Mo, Nb, Cr, Pt, or Ni) of typically 0.2 μm thickness was coated onto the oxide. Then the gate metal was coated with a photo—or electron sensitive resist, depending on the type of lithography chosen. Then the gate material was etched through the patterned resist. In the sixth step, the cavities in the insulating SiO_2 below the holes were either achieved by wet or by dry reactive ion etching. Then, in steps seven and eight, vacuum deposition of the liftoff layer and of the cones followed as in the original process. Finally, rigorous rinsing, cleaning and baking, and storing in vacuum were carried out.

The final structure is schematically shown in Fig. 12.2.

Three other techniques have been used for patterning submicron gate holes by other groups. In the first method, high-energy atomic size particles are used to bombard a photoresist. When developed, tiny holes of about 100 nm are produced in the resist in a random pattern. Thus very high emitter densities could be achieved by Candescent for FEDs [16]. The second method was used by the Korean group of Jong Duk Lee [17], which starts with patterning a silicon substrate with 2 μm diameter Si nitride disks, followed by thermal oxidation of Si till the desired insulation thickness is reached. The nitride is removed and the exposed silicon is etched to obtain the holes under the gates, with subsequent gate layer deposition. The following steps are the same as described above and thus aperture diameters of 0.5 μm are obtained. The third method by C. Bozler et al. from MIT Lincoln Lab [18] is using a standing wave laser interference pattern on a resist coated substrate. Thus, a very high cone/hole

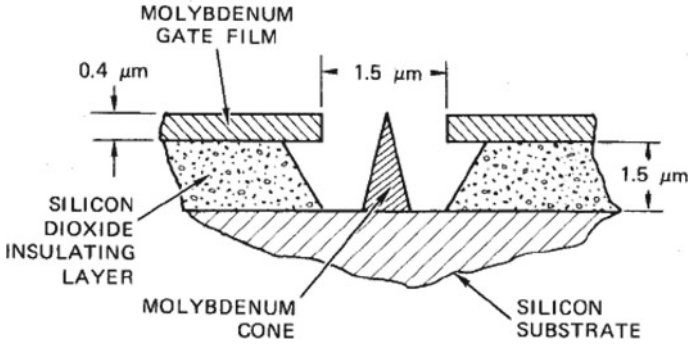


Fig. 12.2 Schematic diagram of a Spindt cathode, originally called thin-film field-emitter cathode, according to Spindt et al. [15]; reprinted from J. Appl. Phys. 47, 5248 (1976); with the permission of AIP Publishing

density of 10^9 holes/cm^2 was achieved with a cesiated 900 tip array. A top current density of 1600 A/cm^2 (total current 1.6 mA) was demonstrated by this group, but unfortunately, this caused burn out of the array. Of course Cs is lowering the work function (=WF), but emission also became unstable over time. In the end, it implies that one has to reduce the load in order to get stable conditions.

Spindt and coworkers have also described a process for forming volcano-shaped emitters, which show lower emission than cones, but have some advantages w.r.t. hollow beams when applied in mass spectrometers.

The general development directions of Spindt arrays are defined by the application requirements for commercial products: Nearly all applications need low operating voltages, higher current densities, very good emission uniformity, very good emission stability and reliability, and reasonable total current. This defines the further development trends, namely reduction of the gate diameter, increase of the packing density, and increase of the number of emitters of the total array, in some cases also increase of the total array area. It has to be mentioned that the smaller apertures as obtained by the three groups described above also need a lower commensurate insulating layer thickness of about the same value as the aperture diameter, which sets some limits for further reduction. In the direction of improvement trends also, an increase of the cone aspect ratio is favorable. This was done by SRI, e.g., by the multiple cone deposition process, where a second deposition is started after liftoff of the sacrificial layer of the first deposition. This approach has the second advantage that the oxide thickness can be preserved.

Also, the packing density and the gate aperture diameter are linked, since a minimum distance is needed between the holes in order to avoid arc failures propagating to the next neighbors. This distance is about twice the minimum aperture diameter achieved of 0.15 μm, i.e., 0.3 μm, realized by the MIT Lincoln Lab. [18].

It is already known that thermal heating improves the performance of etched-wire tungsten field-emitter cathodes. Temperatures of 2000 °C are typically used to smooth and clean the W tip surface, what can be applied for etched-wire tip arrays. However, due to different thermal expansion coefficients of the different layers in the Spindt array structure, thermal treatment temperatures are limited to 400 °C and are not high enough to anneal the Mo tips. Therefore, a formation process of applying high current pulses to heat the tips produces a similar effect. Based on this information, at SRI, annealing and smoothing of the single-tips of Spindt cathode arrays, together with a partial cleaning, was introduced by using 100 μs pulse currents in the 1 mA range [19].

Lowering the tip work function is also an improvement option. Hence, coatings on the Mo tips have been investigated. For example, a diamond-like carbon (=DLC) layer on the tip was reported by Jung et al., resulting in a 30% decrease of the required voltage [20]. Another approach consists of silicon FEAs as prepared by J. Itoh and his group in Japan [21], which will be addressed in a bit more detail later. Schwoebel at SRI investigated a single Spindt emitter tip coated with ZrC [2, 22]. After the deposition of 10 nm of ZrC onto the tip, the voltage for 10 μA emission decreased from 100 V to 73 V, due to the lower work function of ZrC (WF = 3.7 eV). Yet by coating with another 10 nm of ZrC emission deteriorated a bit, even further to the

starting values by heating at 700 °C, but improved again by running at 100 μA peak voltage. It has to be noted that all these emission experiments have been conducted in ultrahigh vacuum of 10^{-10} mbar. From this, we can learn that depending on cleaning procedures, temperature treatment, and surface migration, the emission from the tip can change a lot, and more stable conditions will be found at lower loads.

12.2.3 Spindt Emitter Performance

Let us now consider the emission performance of Spindt arrays in more detail, especially w.r.t. the application requirements.

12.2.3.1 General Emission Performance

Two different diode test configurations have been used by SRI when testing their Spindt-type FEAs: One was a close spaced diode of 0.5 mm distance between Nickel anode and Spindt array cathode. In this case, the Ni anode had to be carefully degassed before starting the emission tests. The other setup could be seen as a beam test setup, since a stainless steel cylindrical anode tube was used with an additional electron repeller at the remote end of the anode tube. In this case, the beam is spreading and then hitting the tube walls, thus the chances of back bombarding ions hitting the cathode(s) are reduced.

With carefully degassed anodes and ultrahigh vacuum of about 10^{-9} mbar, very good long-term stability could be obtained. This was, e.g., demonstrated by an array of 1000 Mo tips on a 0.01 Ω silicon substrate, driven at a peak emission level of 15 mA (corresponding to 15 μA per tip) with a continuous 60 Hz, half wave rectified 75 V peak driver voltage. An earlier life test with a 100 tip array under similar conditions, at an average emission of 20–50 μA per tip (50 μA per tip in the last year of the test), was running for 8 years before getting stopped by a pump failure [4, 23]. Of course, rather moderate conditions for the tips have been chosen in these experiments, in order to avoid stronger changes of the tip surfaces and the tip geometries over time.

Of course, the emission performance should be suitable for replacing thermionic cathodes in several application areas; here, we give an example for microwave applications. Figure 12.3 shows a Fowler–Nordheim plot of emission data per tip for a standard emitter array and a microwave array [24]. The averaged constants a and b of the FN formula are very similar for both. The measurements had been carried out at CPI with a standard array of 10,000 tips and a microwave array with 7300 tips. For microwave applications, very sharp tips had been produced with the double deposition process with 0.4 μm gate hole diameter, an oxide thickness of 1 μm, and 1 μm pitch on glass substrates [24, 25]. The cathode specifications for these applications required 160 mA peak emission and 10 GHz emission modulation from an annular emitter array having a 600 μm outer diameter and an inner diameter of

560 μm determined by trans-conductance and capacitance requirements. Experimental results were: an average array capacitance of 6 nF/cm^2, a trans-conductance of 1 mS/tip, and emitter-tip loadings of 10 mA/tip. These microwave cathodes have been successfully modulated at 10 GHz rates in an experimental klystrode amplifier tube at CPI.

In Fig. 12.4, a Fowler–Nordheim plot of a 5000 cone Spindt array at a collector voltage of 1000 V is shown. The deviation from the FN line above typical current densities of 5A/cm^2 or 10 μA per tip is due to space charge effects which counteract

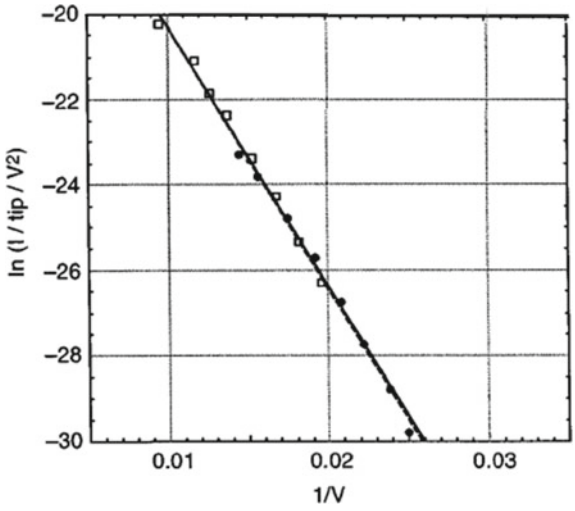


Fig. 12.3 Fowler–Nordheim plot of emission data per tip for a standard emitter array (open squares) and a microwave array (black circles) [24]. The averaged constants a and b of the FN formula (1) are very similar: $a = 7.33 \times 10^{-7}$ A/V/tip and $b = 612$ V for the standard cathode and $a = 7.23 \times 10^{-7}$ A/V/tip and $b = 615$ V for the microwave array; according to Spindt et al. [24]; reprinted with permission from J. Vac. Sci. Technol. B 16, 758–761 (1998); Copyright 1998 AVS

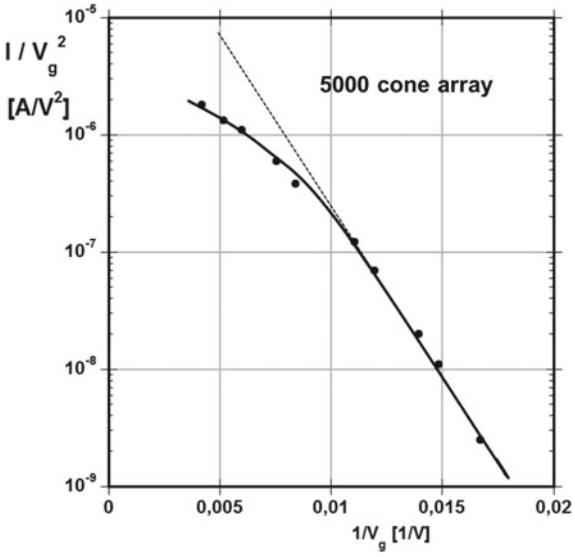


Fig. 12.4 Fowler–Nordheim plot of a 5000 cone Spindt cathode array at a collector voltage of 1000 V. The values were measured at different gate voltages, which is the dominant applied voltage for electron field emission; based on [26–28]

the electric field produced by the gate voltage. Higher collector voltages can reduce this effect a bit, because of the superposition of the two applied fields and the resultant improvement of electron exhaustion from the space charge cloud. Yet besides the limits set by avoiding arcing, this further reduces the emission capability of the tips. In this case, here a top value of 12 A/cm^2 was reached.

12.2.3.2 Thermal/Field Forming of Emitter Tips and Initial Treatment

Thermal/field forming of emitter tips is already well known from classic etched-wire field emitters since the 1930s and has been investigated by Herring [1, 29], Benjamin and Jenkins [30], and others [1]. C. Herring in his textbook "Structures and Properties of Solid Surfaces" [29] derived a well-known formula in 1953 for the time rate change of the tip length $\vartheta z/\vartheta t$ due to the diffusion of the surface atoms (decreasing length z in the positive direction of emitter-tip height). In the modern "International System of Quantities" (ISQ) that forms the basis for SI units, this takes the form:

$$\vartheta z / \vartheta t = \{2\sqrt{2}V_{at}^2 D / (A_{at} k T r)\} * (\gamma / r^2 - \varepsilon_0 F_e^2/2) \tag{12.4}$$

If the difference in the last bracket is negative, then tip buildup or sharpening occurs.

Here A_{at} is the surface area per atom in cm^2, k is Boltzmann's constant [ergs per atom/K], T is the temperature in Kelvin, V_{at} is the atomic volume in cm^3, D is the surface diffusion coefficient in cm^2/s, r is the tip radius in cm, γ is the surface tension in ergs/cm^2, ε_0 is vacuum electric permittivity, and F_e is the electric field in V/cm.

According to Gomer [31], $F_e \approx U/5r$. One should note that tip forming is not dependent on the direction of the electric field, due to its square dependence; hence by using a retarding field, flashovers can be avoided. Instead of sharpening to increase emission, dulling can be used for achieving better uniformity. In Fig. 12.5, the effect of seasoning is shown for two different types of emitter tip arrays with 10,000 emitter tips, one with tip radii of about 25 nm and one with 120 nm.

In total, six emitter arrays of each shape were tested, with separate anodes biased at +1200 V. 10 mA of peak emission were drawn with 60 Hz half wave rectified drive voltage. The tips with small radii initially needed 70 V drive voltage, whereas the tips with larger radii needed 150 V. Seasoning was then carried out by electron bombardment of the anodes, to achieve temperatures of 800–900 °C, and by heating the tips by thermal radiation. Within 5 min, the drive voltage of the smaller tip radii arrays reduced to 40 V, whereas the drive voltage of the larger tip radii arrays remained stable. This can be explained by formula (12.4) where in the latter case, the terms in the last brackets compensate each other [2]. This implies that for a tip radius of 25 nm, the first term is dominating. Yet one should keep in mind that due to $F_e \approx U/5r$, the dependence of both terms on r is the same, which shifts the explanation more to the geometrical effect, that changes in the geometry of a tip with a larger

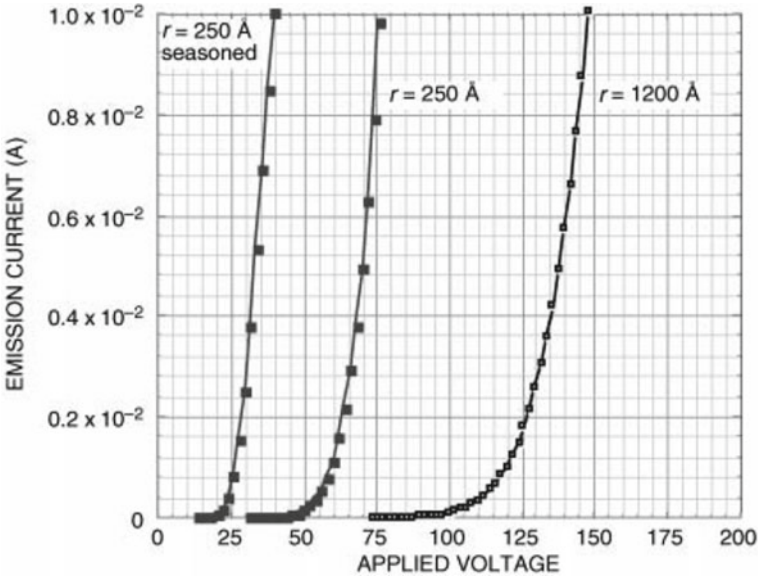


Fig. 12.5 Current–voltage characteristics for two 10,000-tip arrays, one with tip radii ≈ 25 nm and the other with tip radii ≈ 120 nm, showing radius-dependent tip-buildup-like "seasoning" that has been observed as the result of driving the cathodes to emission levels that heat the anodes to 800–900 °C; according to Spindt et al. [2]; reprinted with permission from chapter 4 of *Vacuum Microelectronics*, Wiley 2001 Copyright 2001 Wiley

radius of curvature will take much longer, if atomic rearrangements happen with about the same velocity.

Apart from this seasoning, the initial treatment of the emitter tips before the longer operation can become decisive for the following performance. After manufacturing the tip arrays, cleaning of the surfaces is usually needed. This cannot be done as for tungsten tips by flash heating them at 1500 °C, since this would destroy the structure. It could be shown that heating for 48 h at 400–450 °C, followed by careful first turn-on, and then burn-in is a way to process Spindt arrays before the long-term operation. A similar effect can be achieved via in situ tip self-heating by the emitted electron current for improvement of uniformity [92, 93]. Here Spindt and colleagues found a faster alternative, namely by cleaning the surfaces with a hydrogen plasma for about 2 min duration [23, 91]. This was triggered by the observation that operation in a high vacuum consisting of hydrogen improved emission. A further treatment with hydrogen +10% neon plasma can lead to some further improvement and provides additional sputtering, which can be seen in a field electron micrograph as an increase in the emitting area. Thus, the values improved from $\{I = 1\ \mu\text{A}, U = 175\ \text{V}\}$ to $\{I = 1\ \mu\text{A}, U = 133\ \text{V}\}$ with H_2 plasma, but did not increase with a further H_2 + Ne plasma treatment $\{I = 1\ \mu\text{A}, U = 136\ \text{V}\}$.

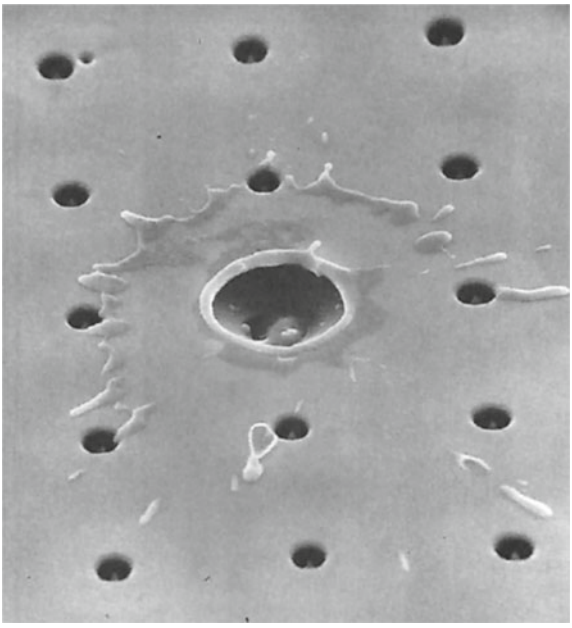


Fig. 12.6 Disruption of a single cone in a 5000 tip Spindt array; according to Spindt et al. [15]; reprinted from J. Appl. Phys. **47**, 5248 (1976); with the permission of AIP Publishing

12.2.3.3 Integrated Resistance for Buffering and Stabilization of Emission

The initial start of emission can lead to a sudden cathode failure by arcing, due to the pressure rise by desorption of contaminants from the tip, accompanied by a lowering of the work function. This is shown in Fig. 12.6 [15]. It illustrates the problem of tip disruption, here of a single cone in a 5000 tip Spindt array, by excessive current drawing [4]. Similar photos can be found in the paper of Forman [28]. The researchers at SRI found that placing Mo emitter arrays on high resistivity silicon substrates (200–500 Ohm cm) reduces this risk substantially [2]. Also, the emission uniformity over larger areas is strongly improved. However, these positive effects are accompanied by additional series resistance and hence heating, and also by an increased energy spread of the emitted electrons. Of course, after initial turn-on, the resistance of the Si substrate can be reduced by subsequent heating of the substrate.

12.2.3.4 Emission Fluctuations and Noise

There are three sources of short-term emission fluctuations of field emitters, namely shot noise, bi-stable "telegraph noise", and random fluctuations [2]. Shot noise is always present and dominates at frequencies f above 100 kHz; it is due to the discrete nature of the electrons. Bi-stable noise is a series of positive or negative pulses above a base level, e.g., in the order of several percents, and is due to changes between

(two) different surface states of the emitter. As a result of the integration of a large number of emitters in arrays, one obtains flicker noise proportional to $1/f^x$ with $0.5 < x < 2$.

Heating to desorb adsorbates did not give the desired reduction of the flicker noise. Yet it could be shown that it is temperature dependent and is reduced at a lower temperature. A better solution was high-frequency pulsing, which reduced flicker noise by a factor of 10.

12.2.3.5 Emission Poisoning by Gas Environment

Field emission is also influenced by the rest gas environment of the emitter due to the absorption of poisoning gases on the surface or eventually chemical reactions with them, due to gas desorption from the anode or to back bombardment of the surface by ionized gas molecules or atoms.

In Fig. 12.7, we see the effect of leaking air into the vacuum vessel from a base pressure of 1×10^{-9} mbar up to 1×10^{-5} mbar. We can see that emission of a Mo tip decreases by a factor of 10, but recovers fast after shutting the leakage valve [23, 26]. Chalamala and Gnade [32] have shown that the poisoning effect of air is mainly due to oxygen poisoning and increases with oxygen pressure, as pointed out by Temple [6].

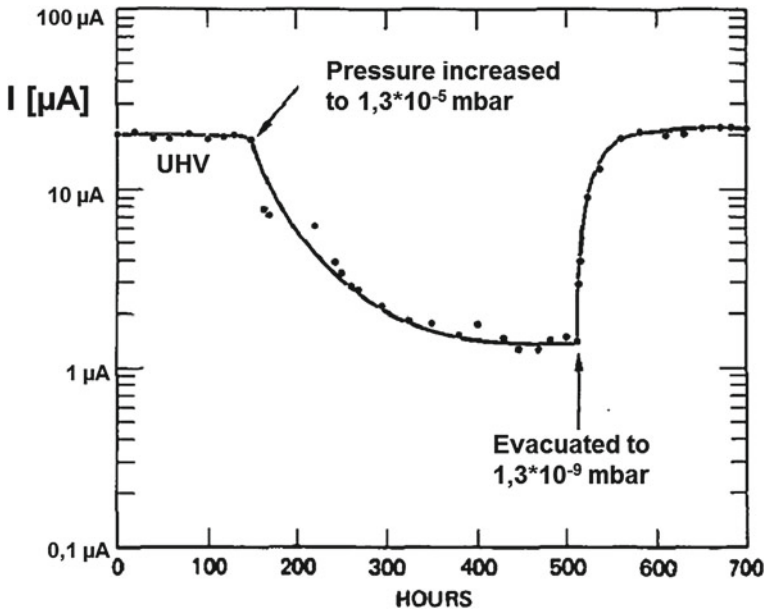


Fig. 12.7 Emission current with a constant applied voltage at a pressure of 1.3×10^{-9} mbar and at 1.3×10^{-5} mbar of laboratory air introduced through a leak valve; based on Spindt et al. [23, 26]

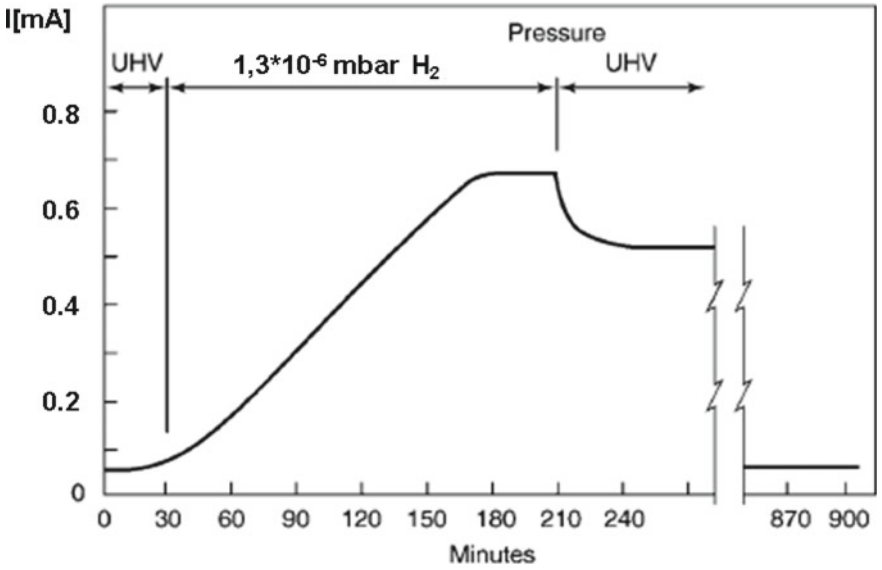


Fig. 12.8 Emission current with a constant applied voltage at a pressure of 1.3×10^{-9} mbar and at 1.3×10^{-6} mbar of hydrogen introduced through a leak valve; based on Spindt et al. [23]

Another situation arises when 10^{-6} mbar H_2 is introduced into the system. In this case, emission can increase up to a factor of 20, but decreases again when the leakage valve is shut and H_2 is pumped out [23] (see Fig. 12.8). This increase of emission can be explained by a lowering of the work function by H_2, an effect contrary to the usual poisoning by other gases. With noble gases, instead of H_2 or air, practically no change of emission happened. In contrast to these reversible effects, an operation in environments containing organics has to be avoided, because built-up of organic material at or near the tip can happen, leading to an early blow out or discharge. In these cases, a periodic temperature or hydrogen plasma treatment is needed to clean the surfaces again.

12.2.3.6 Further Measures to Improve the Emission Performance

In the chapter on fabrication technology, we have already discussed several measures to optimize the performance design of Spindt emitter arrays, among these structural factors such as the increase of tip packing density and the increase of emitter sharpness. Also, the influence of the emitter material and surface work function was addressed. In this paragraph, measures for improvement of uniformity and reliability have been presented, such as heating, seasoning, field forming, plasma treatment, and resistance buffering. Also, the rest gas environment in the vacuum chamber can play a decisive role. Here, we will now concentrate on the performance improvement by lowering the work function of the tip. At SRI, experiments had been conducted by

over-coating the Mo tips (Mo: WF = 4.3 eV) with a thin layer of either Zr (WF = 4.0 eV), Ti (WH = 3.9 eV), and Hf (WF = 3.8 eV) [33]. The results for Zr and Ti are shown in Figs. 12.9 and 12.10.

Over-coating Spindt-type field-emitter-array cathodes with several monolayers of Ti, Zr, or Hf leads to a decrease in the voltage for the same emission current by 30–40%. This change is entirely ascribable to a 1 eV decrease in surface work function and an increase by a factor of 10–100 in the pre-exponential term of the Fowler–Nordheim relation. The post-deposition current—voltage characteristics have been observed to remain essentially unchanged for periods of greater than 100 h at current levels of 10 mA/tip.

It should be noticed that the strongest change of slope, which is dependent on the work function, happened after the removal of surface contaminants by the plasma

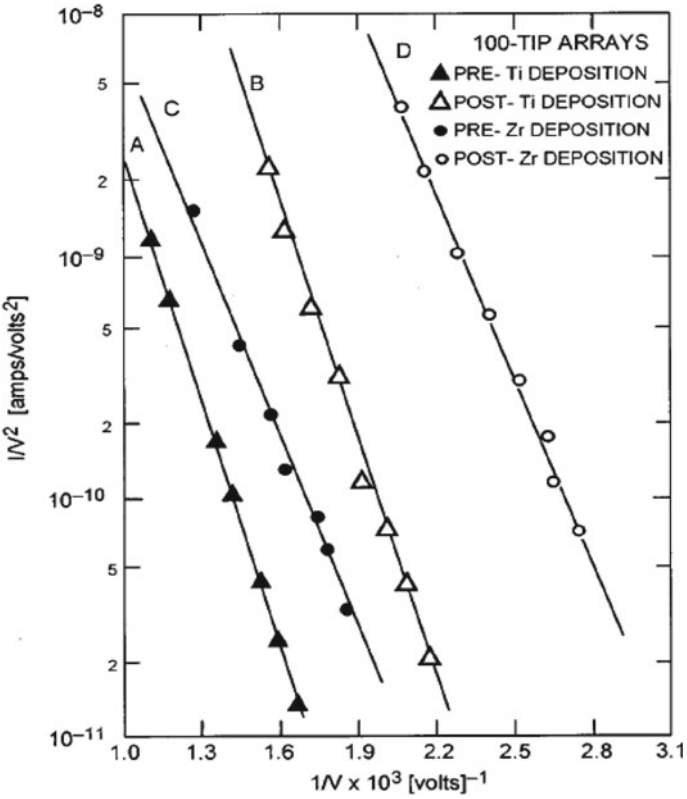


Fig. 12.9 Fowler–Nordheim plots showing the effect of depositing Ti on one 100-tip array and Zr on another 100-tip array. In both cases, the arrays were plasma cleaned before deposition. Line A: FN data of a 100-tip array before and line B after Ti deposition. Line C: FN data of another 100-tip array before (line C) and after Zr deposition (line D); according to Schwoebel et al. [33]; reprinted with permission from J. Vac. Sci. Technol. B 113, 338 (1995) Copyright 1995 American Vacuum Society

treatment (see Fig. 12.10) and is in line with an increase of about 0.6 eV. After coating with Ti or Zr, the changes in slope are minor, corresponding to changes in the work function in the order of 0.1 eV consistent with the average work function data. Yet the major changes occur in the pre-exponential terms of the FN equation, which is not due to an increase in the emitting area and cannot be easily explained. Schwoebel et al. in [33] give the following explanation: They attribute this change to an increased electron supply function due to the IVB elements, and at least in part due to an increase in the surface density of states present at the Fermi level.

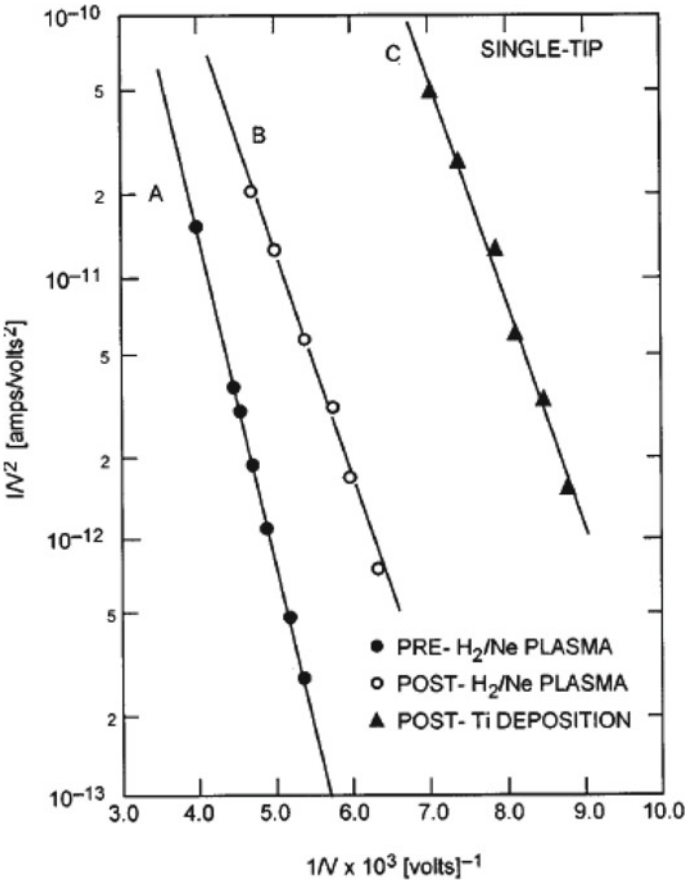


Fig. 12.10 Fowler–Nordheim plot showing the effect of Ti deposition on a single micro-fabricated field-emitter tip where plasma cleaning was employed prior to deposition. Line A: FN data of the tip as fabricated. Line B: FN data following H_2 + 10% Ne plasma treatment. Line C: FN data following the deposition of Ti onto the emitter tip; according to Schwoebel et al. [33]; reprinted with permission from J. Vac. Sci. Technol. B 13, 338 (1995);

Hence, material effects can introduce performance changes that were not directly anticipated.

12.2.4 Performance Improvement Over Time in the Last Four Decades

In the initial phase, the performance improvement was triggered by the need to overcome inherent problems of field emitter arrays mainly w.r.t. uniformity and stability. In the 1980s and 1990s then, the application requirements determined the improvements, which were oriented to increased packing density and larger total currents/increased area, but also to low capacitance and high trans-conductance for high-frequency applications. This is due to the fact that the operation of a FEA in a high-frequency circuit is limited by the cutoff frequency f_t given by [34]

$$f_t = g_m/(2\pi\ c_g) \tag{12.5}$$

where g_m is the trans-conductance (ΔI_a / ΔV_g), and c_g is the capacitance of the device. Hence for an increase of the cutoff frequency, the trans-conductance has to be increased as far as possible and the inter-electrode capacitance has to be decreased, as realized by Spindt and colleagues in [24–26].

In Table 12.1, an overview of the performance increase over time with respect to current density and total current is shown, which is also the base for the diagram in Fig. 12.15. The table entries are organized in historical sequence, see last but one column. The highest tip current from a single tip was achieved by G. van Veen of Philips with 850 μA at 205 V in 1993 [35]. The highest total current with a 50,000 tip array was achieved in 2005, which is also shown in Fig. 12.11. In this case, a 100 μs, 30 Hz pulse was used to drive to the array and the emission current was slowly increased to 300 mA. Above 50 V, we see a strong deviation between the values calculated from the Fowler–Nordheim plot coefficients obtained below 50 V, which according to the authors [36] are mainly due to the resistance of the Si chip, which has an effective series resistance of 300 Ω. The series resistance is necessary to avoid arcing and get more uniform emission, but when it is further reduced, then space charge limitation will take over. Yet a further increase in total current is expected by adjusting the series resistance.

In Fig. 12.12a–c, we see scanning electron micrographs (=SEM) of some newer variants of Spindt arrays. In Fig. 12.12b, we see very sharp tips, where the oxide thickness has been doubled w.r.t. the gate hole diameter. In Fig. 12.12c, Si_3N_4 has been used as an insulator instead of SiO_2. Judging from the higher voltages in Table 12.1 [36], the Si nitride thickness must probably be larger than 1.5 μm.

Table 12.1 Development of the performance of Spindt FEAs; the figure of merit is the total dc current drawn. The values with ** are pulsed, 6 pulses per sec; * = 60 Hz half wave rectified; ° 30 Hz pulses of 100 μs duration; " = 4 ms per data point

Number of tips	Current I	I/tip (μA)	Cathode area A_C	j [A/cm^2]	U [V]	Year	Reference
1	50–150 μA**	50–150	7.85×10^{-11} cm^2 tip	1.9×10^6	200	1976	Spindt [15]
100	2 mA**	20	0.0625 mm^2	3.2	192	1976	Spindt [15]
5000	100 mA**	20	0.785 mm^2	12.7	137	1976	Spindt [15]
16	1.45 mA*	90	0.0132 mm^2	1100	212	1991	Spindt [23]
10,000	100 mA**	10	0.785 m^2	12.7	137	1991	Spindt [23]
625	25 mA*	40	0.0025 mm^2	1000	209	1993	Spindt [34]
1	850 μA"	850	9×10^{-12} cm^2	9.4×10^7	205	1993/94	van Veen [35]
50,000	300 mA°	6	0.75 mm^2	40	160	2005	Schwoebel [19]
100	10 mA* 3 mA*, f.	100 30	3.6×10^{-3} mm^2	278 83	280 330	2015	Spindt [36]
1000	10 mA*	10	3.6×10^{-2} mm^2	27.8	(230)	2015	Spindt [36]
50,000	10 mA*	0.2	1.8 mm^2	0,55	(172)	2015	Spindt [36]

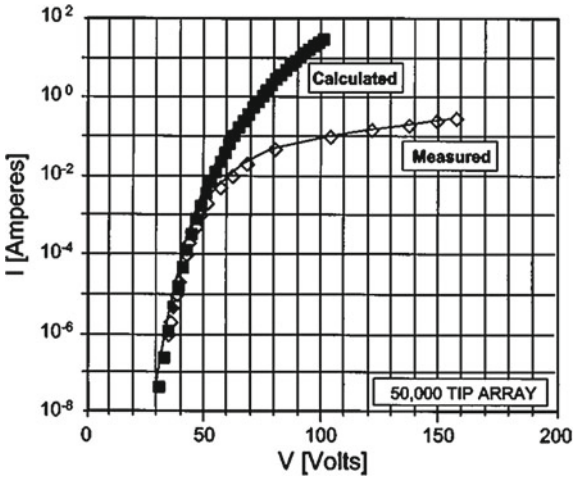


Fig. 12.11 Measured and calculated I-V characteristics of a 50,000 tip Spindt cathode after high current pulsing; the measured deviation is caused by resistance-dependent reduction of the applied voltage and the space charge effect of the emitted electrons at higher emission current; according to Schwoebel et al. [19]; reprinted with permission from J. Vac. Sci. Technol. B 23/2, 691 (2005) Copyright 2005 AVS

12.2.5 Applications of Spindt Arrays

Single tip field emitters are of course very well suited for high brightness electron beam applications, as has been shown by P. Kruit in Chap. 6. This is also true for

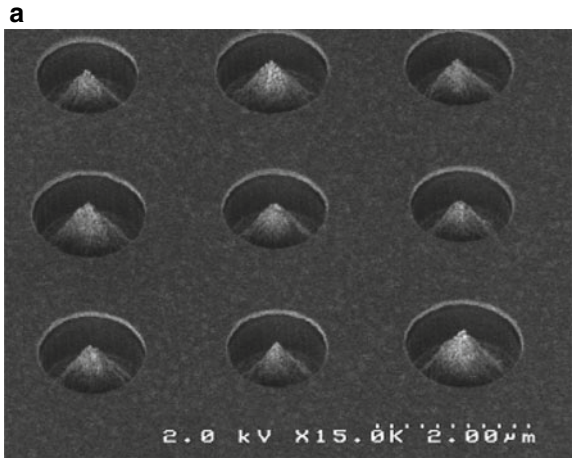


Fig. 12.12 a SEM micrograph of a section of a Spindt array: gate hole diameter 1.5 µm, Mo tip distance 2.5 µm; the FEA tip density is 1 tip/4 μm^2; Courtesy of Spindt, see [37], page 456, Springer 2008

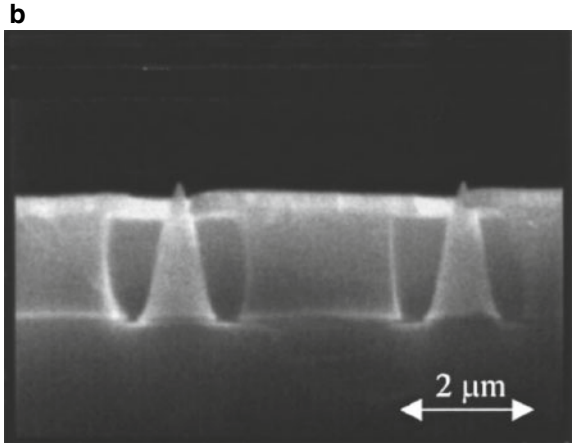


Fig. 12.12 b SEM micrograph of a cross section of a Spindt array cathode: gate aperture diameter 0.8 µm, Mo tip distance 4.1 µm, and gate-to-base oxide thickness of 1.75 µm; according to Schwoebel et al. [38]; reprinted with permission from J. Vac. Sci. Technol. B 21, 433 (2003) Copyright 2003 American Vacuum Society

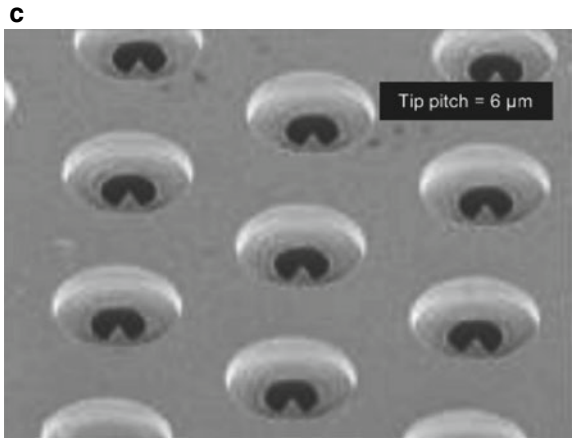


Fig. 12.12 c SEM image of a portion of a 100-tip Spindt cathode. The gate is Cr, the insulating layer below is silicon nitride, and the tips are molybdenum. The tip pitch is 6 µm, and the array is in a hexagonal close-packed configuration; according to Spindt et al. [36]; reprinted with permission from J. Vac. Sci. Technol. B 33, 03C108-1 (2015); Copyright 2015 American Vacuum Society

small field emitter arrays. An interesting variant was introduced by A. Mustonen, S. Tsujino et al. by realizing nano-arrays of 170 nm tip height and 750 nm pitch on an area of 0.2 mm^2 containing about 3.6×10^5 emitter tips [39]. They reached 170 μA, i.e., a current density of 0.085 A/cm^2, at −65 V gate voltage.

Initially, the main motivation for FEA research was applications in high-frequency tubes, where they should replace thermionic cathodes, and later in field emitter displays (FEDs) [40, 41]. There are several publications of C. Spindt and colleagues on applications in microwave tubes [24–26, 34, 39, 42, 94]. In the review article by Murphy and Codis [43], the advantages of FEAs for microwave amplifiers compared to thermionic cathodes are outlined, together with the advantages over solid state amplifiers. Compared to thermionic cathodes for electron beam modulation, much smaller transit times, much higher cutoff frequencies, and much smaller modulation voltages are needed mainly due to the much smaller dimensions of FEAs. Also, no heating power is consumed. In comparison with vacuum electron devices (VEDs), in solid state devices the electron velocity is limited, the breakdown voltages are much lower, and heat cannot be so easily dissipated as in VEDs, which is also pointed out by Gaertner in [44]. A klystrode designed by CPI was intended to provide an output power of 50 W at 10 GHz using a gated FEA cathode. It required a peak current of 112 mA from a ring cathode with inner and outer diameters of 550 and 610 μm, respectively [45]. As an initial test, a FEA-TWT was realized at 1.5 GHz with a maximum current of 91.3 mA and output power of 55 W [45]. In 2009, B. Levush et al. reported a 100 W FEA-TWT produced by L3 Communications Electron Devices, operating at 5 GHz with a maximum beam current of 120 mA, with 22 dB saturated gain, and 33 dB small signal gain [46]. Yet after these prototypes, commercial tubes based on FEAs are rare and are limited in power to <200 W. Microwave tubes with thermionic cathodes continue to be dominant in high-frequency and high-power applications; this is due to the fact that for most tubes, the power is proportional to the electron emission current, which can be much higher for thermionic cathodes compared to Spindt arrays.

The second application of FEAs with a lot of research efforts has been in flat panel field emitter displays (FEDs), as described by Busta [41]. FEDs have been investigated as a vacuum electron tube alternative to the classical bulky cathode ray tube. They have competed with other flat panel concepts, especially TFT LCDs, plasma panels, electroluminescent displays, and OLED displays. We know now that LCDs have won the race, maybe with the perspective that OLED displays will be the high-end solution. Yet for a certain time, it was unclear which technology would win. This is due to the fact that FEDs can exhibit high brightness, have a large viewing angle, a high contrast ratio, a high lumen efficacy, and low power consumption (see Table 7.1 in [41]). In view of the sum of these advantages, an explanation is needed why FEDs did not win, apart from production cost reasons or time delays w.r.t. the other technologies. One reason was that changing from a low surface to volume ratio CRT technology to a high surface to volume ratio, flat vacuum devices

introduced tremendous problems, such as spacer breakdown, flash overs, getter application problems, non-uniformity of emitters, and others. In principle, the FED is a flat and thin tube, with a glass front plate with the phosphor pixels and a glass back plate with a patterned resistor layer, the gate dielectric and the gate layer, formed in column lines, with the gate holes and the tips in the holes formed below. For obtaining uniform emission, it is desirable to fabricate as many emitters per subpixel as possible (ranging from hundreds to several thousands). An inherent problem of the FEAs is the non-uniformity of field emission within one subpixel and over the subpixels, which made the resistor buffering necessary. The inherent noise of field emitters had to be reduced by using at least 3000 tips per subpixel [47]. Due to the lateral velocities of the electrons, additional focusing had to be introduced by a second gate. Basically, emitter lifetime was not an issue, since for 10,000 h FED life, only 20 h FEA life was needed [44]. However, during tube life, particles, degassing and electron bombardment of the surfaces created problems, which also influenced the FEA behavior [47]. In the end, the FED activities were not successful and not competitive, besides producing prototypes; the two pioneering companies Candescent and Pixtech shut down in 2001 and 2002. Also, FED activities in East-Asia could not compete with LCDs and plasma panels [48].

More in the line of beam applications is the use of Spindt arrays for X-ray tubes. In an initial study, P. Schwoebel already investigated the performance of an individual X-ray element that would be arrayed to form a stationary source [42] based on Spindt cathodes. In recent studies in 2015 of Schwoebel et al. [49], the stationary source is comprised of a linear array of 10 individual X-ray source elements. It is used to investigate stationary source—stationary detector tomo-synthesis, which can be used for human breast and small animal imaging. The X-ray source elements each consist of a field emission cathode array, an electrostatic lens, and the target, the last two common to all elements. The source elements form X-ray focal spots with minimum diameters of 0.3–0.4 mm at electron beam currents of up to 40 mA with a beam voltage of 40 kV. Each emitter element comprises 50,000 tip Spindt cathode arrays on a 1 mm diameter area, which deliver at least 10 mA total current (density 1.25 A/cm^2). The initial conditioning of the cathodes comprised 2–3 days prequalification with a slow current increase and 1 day conditioning including baking of the source. Some elements had to be replaced due to failures. Typically, current pulses of 1 ms duration and 1 Hz repetition rate were used for testing. The whole operation test lasted 200 h. Different voltages were needed to achieve 100 mA from each element and were adjusted by the source controller. With this prototype, the feasibility of this approach was demonstrated, since 100 mA is needed for small animal and breast tomo-synthesis. It should be noted that the X-ray flux is directly proportional to the electron current impinging on the anode.

Besides linear arrays of Ba dispenser cathodes, also other field emitter types such as CNT arrays have been used for X-ray applications ([50, 51], see also Chap. 5 of this book). We will comment on that in the next section.

12.3 Other Field Emitter Arrays

12.3.1 Si Field Emitter Arrays

In Chap. 5 of the book of Zhu [1], a detailed description of the fabrication and the performance of Si based field emitter arrays is given by Shaw and Itoh [21]. In this variant, the semiconductor Si was chosen mainly due to the highly developed IC fabrication techniques based on Si. Si FEAs can show *I/U* characteristics different from metal tip FEAs and the current per tip only reaches 10 μA, at best. The total current is not proportional to the number of tips and also the yield of well emitting tips per array may be quite low [52, 53]. Results for p-type Si tip arrays are shown in Table 12.2. The gate consisted of Nb of 0.3 μm thickness with a gate opening of 1.6 μm. Double insulation layers of 0.3 μm-thick thermally oxidized SiO_2 and 0.4 μm-thick vacuum-evaporated SiO_x were placed between the gate and the Si substrate with a resistivity of 100 Ω cm. The pitch between the tips was 10 μm. It has to be noted that the total gate area for 1–100 tips was the same, namely 0.01 mm^2, different from the effective emitter area given in Table 12.2 based on the pitch (rows 2,3, and 4), but only for 1 and for 10 tips no saturation occurred. By the evaluation of individual tip emission in Si field emitter arrays via an electrostatic lens projector, it was found that for n-type, the number of emitting tips remained constant with increasing current, whereas for p-type, the number increased with current, especially during C_2H_4 exposure [54].

Emitters are fabricated as p or n doped Si tips by a conventional fabrication process (described in more detail in [55]) based on reactive ion etching (=RIE), thermal oxidation sharpening, and self-aligned formation of gates [52, 55] as shown schematically in Fig. 12.13. The main differences from the metal tip Spindt arrays are the tips are semiconducting, the tips are formed by etching instead of e-beam deposition (Spindt process), and they are sharper. The emission current from n-type Si arrays is much more noisy than from p-type arrays [55]. Yet, with heavily

Table 12.2 Performance of (p and n doped) Si tip arrays

Number of tips	Maxim. current	Current/tip μA	Emitter area, mm^2	Current density, A/cm^2	Voltage (Gate) (V)	Year, Reference
1000 (p)	30 μA	0.03	0.1	0.03	76.9	1996, Itoh [52]
100 (p)	10 μA	0.1	0.01	0.1	76.9	1996, Itoh [52]
10 (p)	1.8 μA	0.18	0.001	0.18	71.4	1996, Itoh [52]
1 (p)	0.5 μA	0.5	1×10^{-4}	0.5	71.4	1996, Itoh [52]
1000 (n +/n)	140 μA	0.14	0.1	0.14	65	1996, Itoh [55]
3255 (poly Si)	4.5 mA	1.38	0.26	1.73	75	1998,Temple [56]
28 074 (poly Si)	20 mA	0.71	0.94	2.13	96	1998,Temple [56]

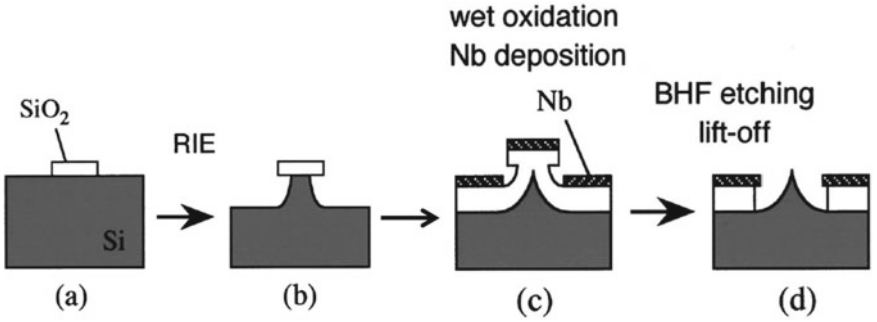


Fig. 12.13 Schematic preparation sequence of Si FEAs according to Itoh et al. [57]; reprinted with permission from J. Vac. Sci. Technol. B **18**, 1111 (2000) Copyright 2000 American Vacuum Society

phosphorus doped n-type tip arrays, an increase in emission current to 0.14 μA was reached, and also n-type emitters do not saturate in the way p-type do. In the case of p-type Si emitters, the emission current shows a saturation tendency in the FN plots. This behavior can be explained by a model where the supply of electrons from the depletion layers near emitter tips and from the inversion layers under the gate electrodes limits the emission current. The p-type saturation can be partially overcome by shining light on the tips. A somewhat different explanation is given in [57].

Below saturation Si FEAs also follow the Fowler–Nordheim equation (12.1–12.3) in the Murphy–Good form, but the work function in the formula has to be replaced by electron affinity χ minus band bending δ : $\phi = \chi - \delta$ [21].

Fowler–Nordheim plots of n-type Si field emitter arrays with1000 tips are shown in Fig. 12.14 [55]. The "n^+/n" is the FEA with a heavily ion (phosphorus) doped n-type surface (compare line 5 in Table 12.2).

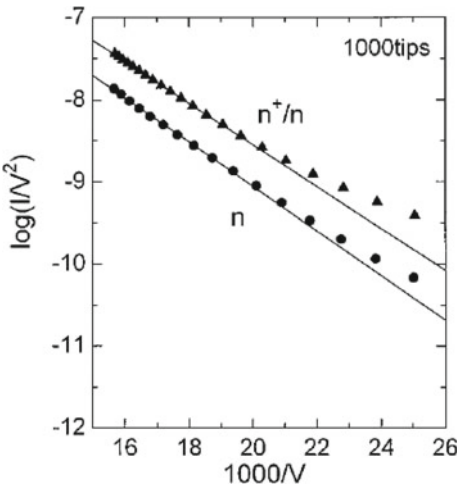


Fig. 12.14 FN plots of n-type Si field emitter arrays with1000 tips; the "n^+/n" is the FEA with a heavily doped n-type surface; compare line 5 in Table 12.2; according to Itoh et al. [55]; reprinted with permission from J. Vac. Sci. Technol. B 14, 1885 (1996) Copyright 1996 AVS

Significantly higher emission values have been obtained by D. Temple and her colleagues for polycrystalline Si tip FEAs, which are listed in the last two lines of Table 12.2. For polycrystalline Si FEAs with 3255 tips, they obtained 4.5 mA dc emission or a current density of about 1.7 A/cm^2. With an array of 28,074 tips they reached at best 21.5 mA or a current density of about 2.1 A/cm^2 [21, 56, 58].

Therefore, in 1998, J. Itoh et al. introduced FEAs with field effect transistors (=FETs) in series—instead of resistors—which show increased stability and uniformity [59–61]. This was first described for 1 tip + MOSFET in [56, 58] and then extended by J. Itoh and group to 100 tips surrounded by a ring-shaped MOSFET gate electrode. They achieved 7 μA of total emission from an area of about 180 μm^2 with a gate voltage of 80 V [61].

Further variants of Si tip arrays were arrays spin-coated with a resist and baked at 800 °C in vacuum to evaporate the solvent. They achieved 54 μA with the coated array versus 44 μA with the non-coated FEA with 100 tips at 65 V gate voltage [62].

Typical applications for these FEAs are cold cathodes replacing thermionic cathodes in pressure gauges [21] or in mass spectrometers [21], which avoid heating effects and degassing caused by the latter. They have also been investigated for applications in vacuum magnetic field sensors [63] or for applications in field emission displays [61, 64] and in microwave tubes, especially klystrodes. Here, they compete with Spindt arrays.

In general, from a comparison of Table 12.2 with Table 12.1, it can be seen that the semiconducting FEAs cannot really compete with the metal FEAs w.r.t. total emission current and current per tip and hence have a handicap concerning higher total current applications. Yet they have the advantage that they can be easily combined with ICs for tip emission control or, e.g., pixel control in FEDs.

12.3.2 Carbon Nano Tube (CNT) Arrays

In the 1990s, different groups started to investigate the preparation of CNTs via screen printing or PCVD with Fe, Ni catalyst nanoparticles for FE applications. Since a single wall CNT (SWCNT) and multiwall CNT (MWCNT) are limited in total current, arrays of single CNTs, lawns of CNTs extending over a certain area, and additional dot arrays have been prepared and tested. Preparation and emission from carbon nano tubes have already been addressed in Chap. 10 by N. Egorov and E. Sheshin. In this context, we will neither address the current carrying capacity (>10^9 A/cm^2) of single multiwall carbon nanotubes (=MWCNTs) as in the paper of Wei et al. [65], nor the thermionic emission of heated carbon filaments. Some representative performance results are listed in the following Table 12.3, which does not claim to give a complete overview, and will appear in the plot of Fig. 12.18.

Niels de Jonge from Philips was the first (in 2001) to succeed in mounting single MWNTs on tungsten tips and determine their emission [66, 74]. The best result is shown in row one of Table 12.3. It has to be mentioned that a lifetime criterion was

Table 12.3 Performance of CNT emitters and CNT dot arrays

Type/number of dots	Current I	j [A/cm^2]	Cathode area A_C(mm^2)	Year	Reference
1 MWNT	10 μA	3×10^6	$3.3\ 10^{-10}$	2004	de Jonge [66]
Circular spot	300 μA	7.8	$3.85\ 10^{-3}$	2006	Chen [67]
Array rect. dots	2.5 mA	1.2	0.21	2006	Chen [68]
Array 10^6 dots	20 mA	0.05	25	2008	Chen [69]
Array 10^6 dots	200 mA pulsed	0.5	25	2010	Chen [70]
Mini X-ray tube	617 μA	0.123	0.785	2012	Heo [71]
X-ray tube	28 mA	0.14	20	2002	Yue [72]
X-ray tube	90 mA dc	5.1	1.75	2012	Ryu [73]

not yet applied in obtaining this table. It is expected that for an operational life of 100–1000 h, some of these emission current values will be lower.

A little later (2004), W. Milne et al. presented a Spindt-type FEA with a pure single MWCNT as field-emitter cathode, as shown in Fig. 12.15a [75]. The advantage over classical Spindt-type cathodes with cone-like emitter is that MWCNTs have lower emitter radii <20 nm and consequently more efficient field electron emission. Figure 12.15b shows also a Spindt-type FEA with arrayed MWCNTs as field-emitter cathode [76]. This kind of emitter growth is strongly dependent on the size and distance of the Fe or Ni catalyst nanoparticles. The most evident advantage is the increase of the emitter-tip number per hole of the Spindt-type FEA, enhancing the long-term stability and probably also the FE cathode lifetime. But on the other hand, the field enhancement of arrayed MWCNTs is reduced. The reason for this is the so-called "screening effect" of closely standing CNTs (also called electrostatic depolarization, see Chap. 9). The screening effect of large-arrayed CNT emitter films

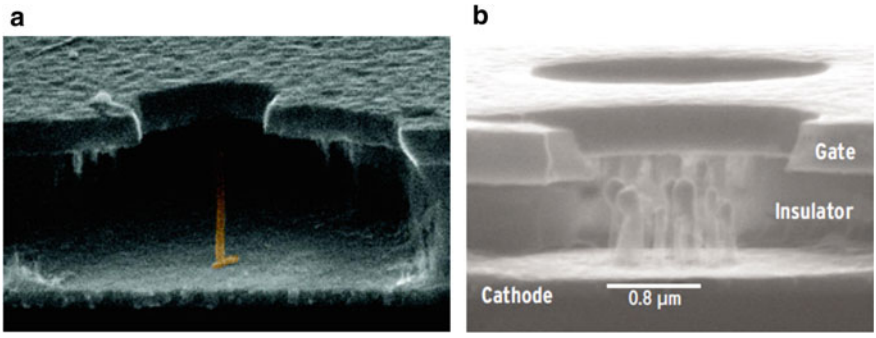


Fig. 12.15 Spindt-type FEAs with CNT field-emitter cathodes: **a** single MWCNT [75] and **b** arrayed MWCNTs, according to Milne et al. [75]; with permission from J. Mater. Chem.**14**, 933–943 (2004) Copyright 2004 Royal Society of Chemistry

in electrostatic fields was investigated by Nilsson [77], qualitatively but also quantitatively by simulation and using the measured electron emission data. The results are summarized in Fig. 12.16.

One rule of thumb is that the optimal distance d between neighboring CNT emitters must be in the range of double CNT height ($d \sim 2h_{CNT}$). Only then will the field enhancement factor β of arrayed CNT emitters be comparable to a stand-alone, single CNT emitter with the same CNT height. This means that FEAs with CNT emitter films should be structured, as, e.g., shown in Fig. 12.17 with CNT dots. In Sect. 9.7.2 "Auxiliary parameters for ideal large-area field electron emitters" of

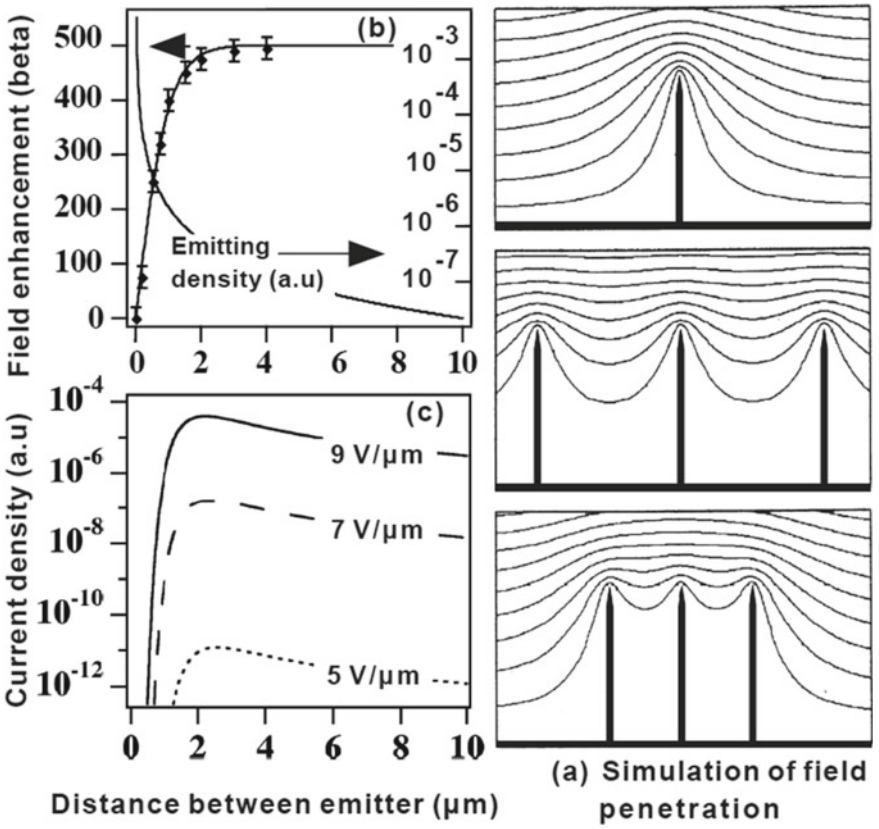


Fig. 12.16 (**a**) In the 3 diagrams of figure (a) on the right side the simulation results of the equipotential lines of the electrostatic field are shown for CNTs of 1 μm height and 2 nm radius (field enhancement factor $\beta \sim h/r = 500$), for the distance between neighboring CNTs of 4, 1, and 0.5 μm; along with the corresponding changes in diagram (**b**) (upper left side) of the field enhancement factor ß and emitter density, and (**c**) FE current density as a function of the distance between neighboring CNT emitters (diagram at the lower left side); figures (**a**–**c**) are from Nilsson et al. [77]; reprinted from Appl. Phys. Lett. **76**, 2071–2073 (2000); with the permission of AIP Publishing

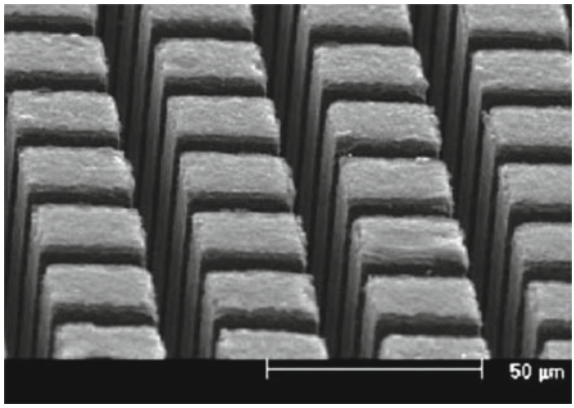


Fig. 12.17 SEM micrograph of array of CNT dots; from [68], Z. Chen, P.K. Bachmann et al., "Fabrication and characterization of carbon nanotube arrays using sandwich catalyst stacks", Carbon **44**, 225–230 (2006); with permission from Elsevier

this book, further theoretical explanations and modified FE equations are given for large-area field emitters (LAFEs) by R. Forbes.

Regular arrays of single or several carbon nanotubes/nano-fibers per dot were prepared by Teo et al. [78] in 2003 by plasma enhanced CVD (=PECVD) growth on base arrays of Ni seed nanoparticles. Usually CNTs obtained by screen printing or CVD are not preferentially aligned, but more of a spaghetti type. P. K. Bachmann and Z. Chen et al. succeeded in preparing vertically aligned dots of CNTs in arrays by the PECVD sandwich method between two silicon wafers, where the top one was taken off after deposition [67, 68, 79]. An example of such a structure is shown in Fig. 12.17. The emission results are listed in rows 2–5 of Table 12.3. The best result from an array of 10^6 dots on an area of 25 mm^2 was 20 mA, stable over 20 h. The highest emission data were obtained in X-ray tube applications, as can be seen from the last two rows of the above table [72, 73] and also from the results of Z. Chen, P.K. Bachmann et al. from 2008 [69] aimed at this application. Their top pulsed emission was 100 mA [70]. Ryu et al. [73] prepared cone-shaped bundles with a sharp tip and a 3 µm diameter base 15 µm apart on an elliptical area and reached 90 mA dc emission.

Another alternative, in view of field (or electrostatic) shielding, is the use of irregular or chaotic carbon-based structures with sufficient distance between emitting spots, as presented in Chap. 10. An example is carbon (MWCNT) buckypaper, proposed by Knapp [80, 81]. With electron emission in the mA range, it nicely fits into the diagram of Fig. 12.18, but we will not discuss it further in this context.

An overview of CNT emitter based X-ray tubes was given by R. Parmee et al. in 2014 and 2016 [82, 83]. Despite good prospects for multi-pixel and miniature X-ray sources, in 2014, there were only some niche applications of X-ray tube device based on CNTs on the market. These are mainly products of Oxford instruments, such as the Eclipse II, a battery-operated miniature tube, and the Horizon 600, source of a miniature XRF spectrometer for space missions [83, 84]. Xintek Inc. is selling a distributed CNT tomo-synthesis unit consisting of 31 individually addressable X-ray

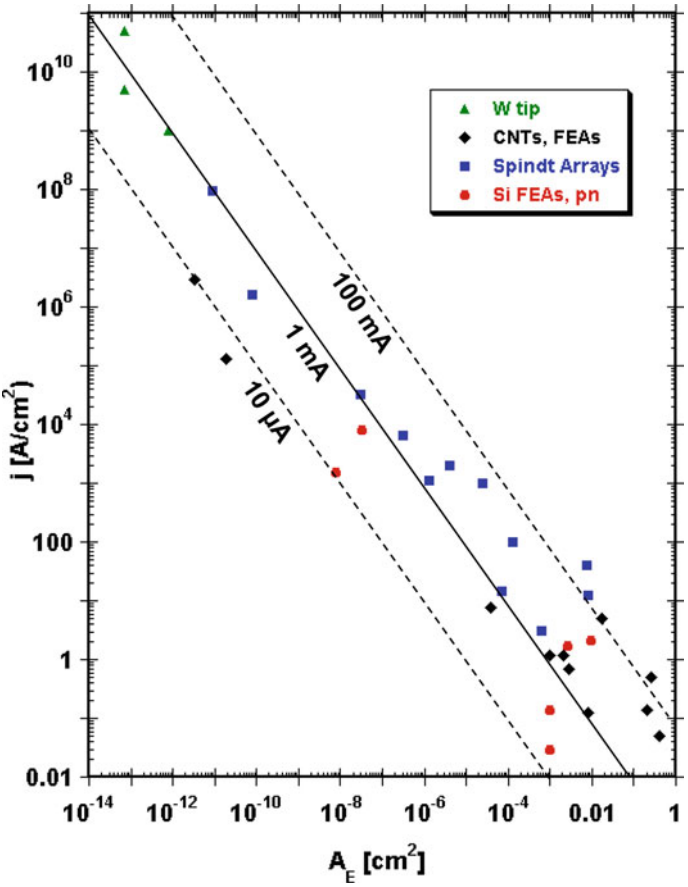


Fig. 12.18 Plot of field emission (cold emission) current density versus emitter area (including passive parts) based on literature data for very sharp W tips, Spindt arrays, CNTs and CNT dot arrays, Si FEAs, and pn emitters according to this chapter (see tables) and [44]. Lines of equal current are shown for 10 µA, 1 mA, and 100 mA. Direction of improvement is to higher current! Copyright Georg Gaertner, Aachen, Germany, 2018

sources enclosed in one vacuum chamber [83]. Yet in general thermionic cathodes are still dominant in X-ray tubes.

There have been many efforts to realize field emission displays (FEDs) with gated CNT array cathodes, first by Motorola in 2006 [85], then by Sony and others, but nearly all the activities stopped before 2010. Typically the current per pixel is lower than for other applications. Various CNT-FED prototypes have shown significant promise, but in the end, were not commercially competitive compared to TFT LCDs and OLED displays and did not become commercial products.

Other applications are microwave tubes or, e.g., gas ionization sensors, where also prototypes have been built, but there is still no real commercial impact.

12.4 General Evaluation and Conclusions

We have seen that field emitters have become an interesting area of cathode research, due to the availability of semiconductor-industry-based IC technology, to the miniaturization and structuring of emitters to the μm and nm range, and to the much lower power consumption. They have a lot of advantages for high brightness electron beams.

Despite the general motivation to try to replace thermionic cathodes by cold cathodes, a general trend of improvement over time cannot be shown in a plot of current density versus time, since this is not the measure of their applicability. FEs already exhibit current densities much higher than thermionic cathodes, but the current mainly originates from very small emitting areas. Hence, natural applications for field emission cathodes are electron beam devices, where low total currents are sufficient. For other applications requiring higher currents and larger cathode areas, practical measures of improvement over time are the increase of emitting area by using bunches or arrays of emitters, and especially the increase of stable dc current from large-area emitter arrays. A graphical overview of the most important cold cathode types, including field emitter arrays, is given in Fig. 12.18 in a plot of emission current density versus emitting area, in this form first given by Gaertner in 2012 [44]. The updated diagram is based on literature data for very sharp W tips (see Fursey [86]), Spindt arrays, CNTs and CNT dot arrays, Si FEAs, and pn emitters [87] according to this chapter and [44]. Lines of equal current are shown for 10 μA, 1 mA, and 100 mA. Here, the improvement direction over time is given in trajectories more perpendicular to the 1 mA line in direction of higher (dc) currents. One can also see an improvement tendency from semiconducting Si to metallic CNTs to metal tip arrays. Of course, thermionic cathodes can also be shown in such a diagram, but their domain is to the lower right, extending to cathode areas of 100 cm^2 or more.

All FEA applications with higher currents are also located preferentially in the lower right with increased cathode areas, since higher currents are accompanied by higher thermal loads and there will also be heating of the field emitter arrays. Another limit is imposed by the thermal load on the anode.

Zhirnov [88] in 2000, first tried to identify standardization criteria for FE measurements, since in many cases, results from different experimenters are not comparable. He also stated that a figure of merit is the total current emitted divided by the entire cathode area and showed a plot of current density versus area partly based on Spindt emitter arrays. His diagram corresponds to the 1 mA line in the lower half of the diagram of Gaertner. Charbonnier [89] also discussed these questions, but instead of a similar diagram came up with the statement that the maximum dc current is 3 mA and the maximum pulsed current is limited to about 120 mA.

Wenger et al. [90] have pointed out that the field emitter arrays have not proven their usefulness in practical applications due to short life and inherent sensitivity to ion bombardment and arcing. They also showed that when increasing current by increasing voltage on CNTs, after a Fowler–Nordheim behavior at lower voltages, a limited FE region follows, where the limitation is caused by space charge and by the

resistance of the contact, the substrate, and the emitter. At further voltage increase, a normal glow discharge starts. Hence, a FE ignited glow discharge can be controlled and used as a plasma electron source [90]. This topic will not be part of this book.

In general, of course, one has to accept the trade-offs by going from tips or "point sources" to extended area sources via arrays, since gates are needed, field shielding (i.e., electrostatic depolarization) has to be taken into account (requiring a certain tip separation), stability of emission from the different tips and reduction of fluctuations and noise have to be achieved by an integrated series resistance, which is also a source of additional power dissipation. Finally, arcing can become detrimental and needs to be avoided by operating at lower current densities. In the end, field emitters give new options and are an interesting alternative to thermionic cathodes especially for beam applications, but they do not really threaten the large-area and high-power domain of thermionic cathodes.

Acknowledgement The authors want to thank Ivor Brodie for his advice w.r.t. relevant literature.

References

1. W. Zhu (ed.), *Vacuum Microelectronics* (Wiley 2001). including chapter 1: "Historical Overview" by Takao Utsumi
2. C. Spindt, I. Brodie, C. Holland, P. Schwoebel, Spindt Field Emitter Arrays, in Chapter 4 of *Vacuum Microelectronics* (Wiley 2001), pp. 105–186
3. K. Jensen, Theory of field emission, in Chapter 3 of *Vacuum Microelectronics* (Wiley 2001), pp. 33–104
4. I. Brodie, C.A. Spindt, Vacuum microelectronics, in *Advances in Electronics and Electron Physics*, vol. 83, ed. by P.W. Hawkes (Academic Press, New York 1992), pp. 1–106
5. N. Cade, R. Lee, Vacuum microelectronics. GEC J. Res. **7**(3), 129–138 (1990)
6. Dorota Temple, Recent progress in field emitter array development for high performance applications. Mater. Sci. Eng. **R24**, 185–239 (1999)
7. N. Egorov, E.P. Sheshin, *Electron field emission, principles and applications* (In Russian), Intellekt 2011; Updated English version *Field Emission Electronics* published by Springer in 2017
8. C.A. Spindt, K.R. Shoulders, Research in micron-size Field-emission tubes, in *IEEE Conference Record, Eighth Conference on Tube Techniques* (1966), p. 143
9. C.A. Spindt, A thin-film field-emission cathode. J. Appl. Phys. **39**, 350 (1968)
10. R. Meyer, A. Ghis, P. Rambaud, F. Muller, Microchip florescent display, in Proc. Japan Display (1985), p. 513
11. C.E. Holland, C.A. Spindt, I. Brodie, J. Mooney, E.R. Westerberg, Matrix addressed cathodoluminescent display, in Int. Display Conf. (London, UK, 1987)
12. G. Gaertner, P. Janiel, J.E. Crombeen, J. Hasker, Top-layer scandate cathodes by plasma-activated CVD, in *Vacuum Microelectronics 1989*, ed. by R.E. Turner, Institute of Physics Conf. Series No. 99 (Bristol, New York), pp. 25–28
13. Ivor Brodie, Julius J. Muray, *The Physics Of Micro/Nano-Fabrication* (Springer, Boston, MA, 1992)
14. I. Brodie, E.R. Westerberg, D.R. Cone, J.J. Muray, N. Williams, L. Gasiorek, A multiple-electron- beam exposure system for high-throughput, direct-write submicrometer lithography. IEEE Transact. Electron Dev. **28**(11), 1422–1428 (1981)

15. C.A. Spindt, I. Brodie, L. Humphrey, E.R. Westerberg, Physical properties of thin-film field emission cathodes with molybdenum cones. J. Appl. Phys. **47**, 5248 (1976)
16. X. Chen, S. Zaidi, D. Devine, S. Brueck, J. Vac. Sci. Technol. B. **14**, 3339 (1996)
17. C.H. Oh, J.D. Lee, et al., Fabrication of metal field emitter arrays for low voltage and high current operation, J. Vac. Sci. Technol. B. **16**, 807 (1998)
18. C. Bozler, C. Harris, S. Rabe, D. Rathman, M. Hollis, H. Smith, Arrays of gated field-emitter cones having 0.32 μm tip-to-tip spacing. J. Vac. Sci. Technol. B. **12**, 629 (1994)
19. P.R. Schwoebel, C.A. Spindt, C.E. Holland, High current, high current density field emitter array cathodes. J. Vac. Sci. Technol. B. **23/2**, 691f.(2005)
20. J.H. Jung et al., Electron emission performance of Mo tip FEAs with nitrogen-doped hydrogen-free DLC coating, in *Proceedings of 12th IVMC* (Darmstadt 1999), pp. 110–111
21. J. Shaw, J. Itoh, Silicon Field Emitter Arrays, in Chapter 5 of *Vacuum Microelectronics* (Wiley, 2001), pp. 187–246
22. J.M. Macaulay, I. Brodie, C.A. Spindt, C.E. Holland, Appl. Phys. Lett. **61**, 997 (1992)
23. C.A. Spindt, C.E. Holland, A. Rosengreen, I. Brodie, Field-emitter arrays for vacuum microelectronics. IEEE Trans. Electron Devices **38**, 2355 (1991)
24. C.A. Spindt, C.E. Holland, P.R. Schwoebel, I. Brodie, Field emitter array development for microwave applications II. J. Vac. Sci. Technol. B. **16**, 758–761 (1998)
25. C.A. Spindt, C.E. Holland, P.R. Schwoebel, I. Brodie, Field-emitter-array development for microwave applications. J. Vac. Sci. Technol. B. **14**, 1986 (1996)
26. I. Brodie, C.A. Spindt, The application of thin-film field-emission cathodes to electronic tubes. Appl. Surf. Sci. **2**, 149–163 (1979)
27. C.A. Spindt, C. Holland, R. Stowell, Field emission array development for high-current-density applications. Appl. Surf. Sci. **16**, 268–276 (1983)
28. R. Forman, Evaluation of emission capabilities of Spindt-type field emitting cathodes. Appl. Surf. Sci. **16**, 277–291 (1983)
29. C. Herring, *Structures and Properties of Solid Surfaces* (University of Chicago Press, Chicago,1953). p. 5
30. M. Benjamin, R.O. Jenkins, Proc. R. Soc. Lond. A. **176**, 262 (1940)
31. R. Gomer, *Field Emission and Field Ionization* (Harvard University Press, Cambridge/MA, 1961). p. 54
32. B. Gnade, in *Proc. Spring Meeting of the Materials Research Society, Tutorial Program, Symposium G* (San Francisco, CA, 1997)
33. P.R. Schwoebel, C.A. Spindt, I. Brodie, Electron emission enhancement by over-coating molybdenum field-emitter arrays with titanium, zirconium, and hafnium. J. Vac. Sci. Technol. B. **13**, 338 (1995)
34. C.A. Spindt, C.E. Holland, A. Rosengreen, I. Brodie, Field emitter-array development for high-frequency operation. J. Vac. Sci. Technol. B. **11**, 468 (1993)
35. G.N.A. van Veen, Space-charge effects in Spindt-type field emission cathodes. J. Vac. Sci. Technol. B. **12**, 655 (1994)
36. C.A. Spindt, C.E. Holland, P.R. Schwoebel, Thermal field forming of Spindt cathode arrays. J. Vac. Sci. Technol. B. **33**, 03C108–1 (2015)
37. G. Gaertner, H.W.P. Koops, Vacuum Electron Sources and their Materials and Technologies, in chapter 10 of *Vacuum Electronics, Components and Devices*, ed. J. Eichmeier, M. Thumm (Springer, 2008)
38. P.R. Schwoebel, C.A. Spindt, C.E. Holland, Spindt cathode tip processing to enhance emission stability and high-current performance. J. Vac. Sci. Technol. B. **21**, 433 (2003)
39. A. Mustonen, V. Guzenko, C. Spreu, T. Feurer, S. Tsujino, High-density metallic nano-emitter arrays and their field emission characteristics. Nanotechnology **25**, 085203 (2014)
40. C. Spindt, C. Holland, I. Brodie et al., Field emitter arrays applied to vacuum fluorescent display. IEEE Trans. ED **36**, 225–228 (1989)
41. H. Busta, Field emission flat panel displays, in chapter 7 of *Vacuum Microelectronics* (Wiley, 2001), pp. 289–347

42. P. Schwoebel et al., Field emission arrays for medical X-ray imaging. Appl. Phys. Lett. **88**, 113902 (2006)
43. R.A. Murphy, M.A. Codis, Cold cathode microwave devices, in chapter 8 of *Vacuum Microelectronics* (Wiley, 2001), pp. 349–391
44. G. Gaertner, Historical development and future trends of vacuum electronics. J. Vac. Sci. Technol. B. **30/6**, 060801(2012)
45. D.R. Whaley, B.M. Gannon, C.R. Smith, C.M. Armstrong, C.A. Spindt, Application of field emitter arrays to microwave power amplifiers. IEEE Trans. Plasma Sci. **28**, 727–747 (2000)
46. J.X. Qiu, B. Levush et al., Vacuum tube amplifiers. IEEE Microw. Mag. 38–51 (2009)
47. J.A. Hart, A History of Field Emission Displays (Indiana University, 1999). http://www.indiana.edu/~hightech/fpd/papers/FEDs.PDF
48. J. Lieberman, Field-Emission Displays Get a Second Wind?", IEEE Spectrum, 1.10.2003
49. P.R. Schwoebel, J.M. Boone, J. Shao, Studies of a prototype linear stationary X-ray source for tomosynthesis imaging. Phys Med Biol. **59**, 2393–2413 (2014)
50. G.Z. Yue, O. Zhu et al., Generation of continuous and pulsed diagnostic imaging x-ray radiation using a carbon-nanotube-based field-emission cathode. Appl. Phys. Lett. **81**, 355–357 (2002)
51. S Cheng et al., A compact X-ray generator using a nanostructured field emission cathode and a micro-structured transmission anode. J. Phys. Conf. Ser. **476**, 012016 (2013)
52. T. Hirano, S. Kanemaru, J. Itoh, Emission current saturation of the p-type silicon gated field emitter array. J. Vac. Sci. Technol. B. **14**, 3357 (1996)
53. T. Matsukawa, S. Kanemaru, K. Tokunaga, J. Itoh, Individual tip evaluation in Si field emitter arrays by electrostatic lens projector. J. Vac. Sci. Technol. B. **18**, 952 (2000)
54. J. Itoh, Development and applications of field emitter arrays in Japan. Appl. Surf. Sci. **111**, 194–203 (1997)
55. S. Kanemaru, T. Hirano, H. Tanoue, J. Itoh, Control of emission characteristics of silicon field emitter arrays by an ion implantation technique. J. Vac. Sci. Technol. B. **14**, 1885 (1996)
56. D. Temple, W.D. Palmer, L.N. Yadon, J.E. Mancusi, D. Vellenga, G.E. McGuire, Silicon field emitter cathodes: fabrication, performance, and applications. J. Vac. Sci. Technol. A. **16**, 1980 (1998)
57. T. Matsukawa, S. Kanemaru, K. Tokunaga, J. Itoh, Effects of conduction type on field-electron emission from single Si emitter tips with extraction gate. J. Vac. Sci. Technol. B. **18**, 1111 (2000)
58. D. Palmer, H.F. Gray, J. Mancusi, D. Temple, C. Ball, J. Shaw, G.E. McGuire, Silicon field emitter arrays with low capacitance and improved transconductance for microwave amplifier applications. J. Vac. Sci. Technol. B. **13**, 576–579 (1995)
59. S. Kanemaru, T. Hirano, H. Tanoue, J. Itoh, Control of emission current from silicon field emitter arrays using a built-in MOSFET. Appl. Surf. Sci. **111**, 218–223 (1997)
60. M. Nagao, D. Nicolaescu, T. Matsukawa, S. Kanemaru, J. Itoh, T. Sato, Y. Sato, N. Wada, Metal–oxide–semiconductor field-effect transistor-structured Si field emitter array with a built-in ring gate lens. J. Vac. Sci. Technol. B. **21**, 495 (2003)
61. H. Gama, S. Kanemaru, J. Itoh, A field emitter array monolithically integrated with a thin film transistor on glass for display applications. Appl. Surf. Sci. **146**, 187–192 (1999)
62. K. Ehara, S. Kanemaru, T. Matsukawa, J. Itoh, Improvement of electron emission characteristics of Si field emitter arrays by surface modification. Appl. Surf. Sci. **146**, 172–175 (1999)
63. J. Itoh, K. Uemura, S. Kanemaru, Three-dimensional vacuum magnetic sensor with a Si emitter tip. J. Vac. Sci. Technol. B **16**, 1233 (1998)
64. M. Nagao, C. Yasumuro, M. Taniguchi, S. Itoh, S. Kanemaru, J. Itoh, Field emitter array with a memory function for ultrahigh luminance field emission display. J. Vac. Sci. Technol. **B25**, 464 (2007)
65. B. Wei, R. Vajtai, P.M. Ajayan, Reliability and current carrying capacity of carbon nanotubes. Appl. Phys. Lett. **79**, 1172 (2001)
66. N. de Jonge, J.-M. Bonard, Carbon nanotube electron sources and applications. Phil. Trans. R. Soc. Lond. A **362**, 2239–2266 (2004)

67. Z. Chen, G. Cao, Z. Lin, D. den Engelsen et al., Synthesis and emission properties of carbon nanotubes grown by sandwich catalyst stacks. J. Vac. Sci. Technol. B **24**, 1017 (2006)
68. Z. Chen, P.K. Bachmann et al., Fabrication and characterization of carbon nanotube arrays using sandwich catalyst stacks. Carbon **44**, 225–230 (2006)
69. Z. Chen, P.K. Bachmann et al., Growth of uniform carbon nanotube arrays with sandwich technology. J. Soc. Inf. Display **16**, 645 (2008)
70. Z. Chen, P.K. Bachmann, et al., Field emission from CNT bundles for application in biomedical equipment, in Proc. 8th Int. Vac. Electron Sources Conf. and Nanocarbon (IEEE Press, Nanjing, 2010), pp. 111–112
71. S.H. Heo, H.J. Kim et al., A vacuum-sealed miniature X-ray tube based on carbon nanotube field emitters. Nanoscale Res. Lett. **7**, 258 (2012)
72. G. Yue, O. Zhou, Generation of continuous and pulsed diagnostic imaging x-ray radiation using a carbon-nanotube based field-emission cathode. Appl. Phys. Lett. **81**, 355 (2002)
73. J. Ryu, J. Kang, K. Park, Carbon nanotube electron emitter for X-ray imaging. Materials **5**, 2353–2359 (2012)
74. N. de Jonge, N. van Druten, Field emission from individual multiwalled carbon nanotubes prepared in an electron microscope. Ultramicroscopy **95**, 85–91 (2003)
75. W.I. Milne, K. Teo, G. Amaratunga et al., Carbon nanotubes as field emission sources. J. Mater. Chem. **14**, 933–943 (2004)
76. G. Amaratunga, Watching the nanotubes. IEEE Spectr. **40**, 28–32 (2003)
77. L. Nilsson, O. Groening, J.-M. Bonard et al., Appl. Phys. Lett. **76**, 2071–2073 (2000) and L.-O. Nilsson, Microscopic characterization of electron field emission from carbon nanotubes and carbon thin-film electron emitters", PhD thesis, University Freiburg (CH), p. 79 (2001)
78. K. Teo, S. Lee, W. Milne et al., Plasma enhanced chemical vapour deposition carbon nanotubes/ nanofibres—how uniform do they grow? Nanotechnology **14**, 204–211 (2003)
79. Z. Chen, P.K. Bachmann et al., High emission current density microwave-plasma-grown carbon nanotube arrays by post-depositional radio-frequency oxygen plasma treatment. Appl. Phys. Lett. **87**, 243104 (2005)
80. W. Knapp, D. Schleussner, Field-emission charcteristics of carbon buckypaper. JVST B **21**, 557–561 (2003)
81. W. Knapp, D. Schleussner, Special features of electron sources with CNT field emitter and micro-grid. Appl. Surf. Sci. **251**, 164–169 (2005)
82. R. Parmee, C. Collins, W. Milne, M. Cole, X-ray generation using carbon nanotubes. Nano Converg. **2**(1), 1–27 (2015)
83. M. Cole, R. Parmee, W. Milne, Nanomaterial-based x-ray sources. Nanotechnology **27**, 082501 (2016)
84. P. Sarrazin, D. Blake, M. Meyyappan et al., Carbon-nanotube field emission X-ray tube for space exploration XRD/XRF instrument. Adv. X-Ray Anal. **47**, 232 (2004)
85. M. Cole, M. Mann, K. Teo, W. Milne, Engineered carbon nanotube field emission devices, in Chapter 5 of *Emerging Nanotechnologies for Manufacturing*, 2nd edn. (Pergamon/Elsevier Science, 2014), pp. 126–185
86. G.N. Fursey, Field emission in vacuum micro-electronics. Appl. Surf. Sci. **215**, 113 (2003)
87. G. van Gorkom, A. Hoeberechts, Back-biased junction cold cathodes: history and state of the art, in *Vacuum Microelectronics 1989*, IOP Conference Series **99** (Institute of Physics, Bristol, 1989), pp. 41–52
88. V.V. Zhirnov, C. Lizzul-Rinne, G.J. Wojak, R.C. Sanwald, J.J. Hren, "Standardization" of field emission measurements. J. Vac. Sci. Technol. B **19**, 87 (2001)
89. F. Charbonnier, "Developing and using the field emitter as a high intensity electron source. Appl. Surf. Sci. **94/95,** 26–43 (1996)
90. D. Wenger, W. Knapp, B. Hensel, S. F. Tedde, Transition of electron field emission to normal glow discharge. IEEE Transact. Electron Devices, **61(11)**, 3864–3870 (2014)
91. P.R. Schwoebel, C.A. Spindt, Glow discharge processing to enhance field-emitter array performance. J. Vac. Sci. Technol. B. **12**, 2414 (1994)

92. P.R. Schwoebel, C.A. Spindt, C.E. Holland, Emission uniformity enhancement between micro-fabricated tips in cold cathode arrays. J. Vac. Sci. Technol. B. **19**, 582 (2001)
93. P.R. Schwoebel, C.A. Spindt, C.E. Holland, J.A. Panitz, Field emission current cleaning and annealing of micro-fabricated cold cathodes. J. Vac. Sci. Technol. B. **19**, 980 (2001)
94. C. Holland, C. Spindt, A. Rosengren, I. Brodie, Field emitter array development for high-frequency operation. Conference Record of TRI/NASA Cathode Workshop (Cleveland/Ohio 1994), pp. 87–90

Index

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*,
Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7

B

C

E

M

N

O

P

Q

R

S