SARTONIANA

Volume 30

2017

Editors : Robert Rubens and Maarten Van Dyck

Sarton Chair of the History of Sciences
Ghent University, Belgium

SARTONIANA

Volume 30
2017

Sarton Chair of History of Science
Ghent University

SARTONIANA

Volume 30

2017

Editors: Robert Rubens and Maarten Van Dyck

Sarton Chair of History of Sciences
Ghent University, Belgium

Sartoniana, Gent, Belgium
Ghent University

ISBN 978 90 736 2663 8
D/2017/3988/23

Layout en druk:
www.nevellandgraphics.be Drukkerij Nevelland Graphics cvba-so Drongen

Table of Contents

Authors

Prof. dr. Robert Rubens
Chairman Sarton Committee, University Hospital Ghent, Department of Endocrinology, De Pintelaan 185, 9000 Gent, Belgium
robert.rubens@UGent.be

Prof. dr. Maarten Van Dyck
Secretary Sarton Committee, Ghent University, Centre for History of Science, St.-Hubertusstraat 2, 9000 Gent, Belgium
maarten.vandyck@UGent.be

Prof. dr. Diederik Depla
Ghent University, Faculty of Sciences, Department of Solid State Sciences, Krijgslaan 281, 9000 Gent, Belgium
diederik.depla@UGent.be

Prof. dr. Joseph Greene
University of Illinois, College of Engineering, 2015 MRL, MC 230, 104 South Goodwin Ave. Urbana, IL 61801-2985, USA
jegreene@uiuc.edu

Prof. dr. Yves T'Sjoen
Ghent University, Faculty of Arts and Philosophy, Department of Literary Sciences, Blandijnberg 2, 9000 Gent, Belgium
yves.tsjoen@UGent.be

Prof. dr. Steward Van Wyk
University of Western Cape, Department Afrikaans en Nederlands, Robert Sobukwe Road, Belleville 7535, Republic of South Africa
svanwyk@uwc.ac.za

Prof. dr. Dirk Heirbaut
Ghent University, Faculty of Law, Department of Interdisciplinary Study of Law, Private Law and Business Law, Universiteitstraat 4, 9000 Gent, Belgium
dirk.heirbaut@UGent.be

Prof. dr. Louis Berkvens
Maastricht University, Faculty of Law, Foundation and methods of Law, PO Box 616, 6200 MD Maastricht, the Netherlands
louis.berkvens@maastrichtuniversity.nl

Prof. dr. Ingrid Kerckaert
Ghent University, Faculty of Medicine and Health Sciences, Department of Basic Medical Sciences, University Hospital Ghent, De Pintelaan 185, 9000 Gent, Belgium
ingrid.kerckaert@UGent.be

Dr. Frank IJpma
University Medical Center Groningen, Surgery Department, PO Box 30.001, 9700 RB Groningen, the Netherlands
frankijpma@gmail.com

Prof. dr. Stijn Vanheule
Ghent University, Faculty of Psychology and Educational Sciences, Department of Psycho-analysis and clinical consulting, H. Dunantlaan 2, 9000 Gent, Belgium
stijn.vanheule@Ugent.be

Prof. dr. Dany Nobus
Brunel University London, Department of Psychology, Marie Jahoda Building, Uxbridge, UB8 3PH, United Kingdom
dany.nobus@brunel.ac.uk

Introduction

R. Rubens

In 1817 Ghent University was founded by a royal decree of King Willem I of The United Kingdom of the Netherlands (consisting of nowadays Netherlands and Belgium). In that same university nearly one century later (1911) George Sarton got his doctoral degree in science. Based upon his education as well in philosophy as in exact science he started an important career in the history of science ending as professor emeritus history of science at Harvard University. To remember the alumnus a yearly chair of the philosophy and history of sciences was created in Ghent University more than 30 years ago.

The new volume of Sartoniana 2017 contains the lectures of the Sarton Chair for the history of sciences in the different faculties of the academical year 2016-17.

The chairholder Joe Greene has produced an extensive narrative review about the evolution of thin-film sputter deposition. The paper contains every reference to the development of sputter deposition during four centuries. It is impossible to resume the enormous amount of information presented, interlaced with the historical vignettes about every scientist. It should be read by every physicist interested in the subject to learn how a travel through history can clarify some nowadays concepts.

Steward Van Wyk, coming from the University of the Western Cape provides a beautiful insight into the Afrikaans writing by descendants of the original inhabitants of the Cape. The Afrikaans language, once dubbed the language of the oppressor, is used to produce a literature full of tropes of memory and pain coming from the horrible apartheid period, reflected in numerous narratives.

The legal history of the parts “Outre Meuse” or Dutch Limburg is the content of the lecture of Louis Berkvens The patchwork of the little states in the sixteen- to nineteenth century with the permanent struggle between the different surrounding powers and principalities concerning legal systems is like a historical labyrinth carefully dissected by his research.

Frank IJpma from Groningen university proves that even to-day surgeons can do very important scientific work in anatomy. His study of the anatomy lessons of the Amsterdam Guild of surgeons provides new insights in a very traditional field. The careful redissection of the anatomy lesson of professor Tulp proves that certain features are not as photographical as thought previously.

The influence of the first world war upon the shaping of psychoanalysis is the subject of carefull research by Dany Nobus. The influence of the war events on the production of the intellectual work of Freud is paramount. Meanwhile Freud’s daughter was trapped in England and his three sons involved in the Austrian Army. The absence of clinical work and very few patients consulting for neuroses made the reflective period possible.

Unfortunately one lecture by Philippe De Maeyer, although long promised, is missed.

We hope the mixture of different insights into the history and development of sciences may again catch your interest as a reader.

Laudatio Joseph Greene

Diederik Depla

In 2005, still post-doc in the research group headed by colleague and good friend Roger De Gryse, I met Joe Greene for the first time. Joe Greene was invited to RSD2005 conference by Guido Janssen, a Dutch professor from TUDelft who kick started the tradition to organize the reactive sputter symposium in the uneven years outside Ghent. Roger knew Joe already much longer, and suggested to him to visit the Ghent research group, and to teach here in Ghent his well-known and almost iconic short course on "Thin Film Nucleation, Growth, and Microstructural Evolution" on the Wednesday before RSD2005. During the course, I was struck not only by Joe's enthusiasm, his direct way of communication, but also by his enormous background knowledge illustrated by a seeming never ending list of references to new but also old papers. I didn't realize it then but now looking back, I could already sense Joe's interest for history. Roger had other duties for Wednesday evening, and insisted that I would take Joe for dinner. A bit nervous, I went with Joe to a restaurant not far from here. And there I learned that we not only shared the enthusiasm to explain and figure out things, but also a love for beer.

I could go now the easy way with this laudatio, and proof to you Joe's importance for thin film research and his contribution to observe some trends from a historical perspective. I could cite from his impressive curriculum, mention that the submitted application form for the Sarton Chair is 90 pages long (font size Arial 10), or refer to his long list of academic prizes, that he holds not one but three times the title of Professor. But this easy way I don't want to go. I will use a recent publication by Matthew Stanley, Professor in history of sciences at New York University. This paper, published in Physics Today [1], is entitled "Why physicist should study history?" and I will use it to structure this talk. Stanley gives us several reasons why we, as scientists, should study history. It is interesting that Joe demonstrates these reasons nicely in his work on the history of thin film deposition, in his material science research, and even in his style of teaching or science communication. Let me illustrate this with a few observations.

According to Stanley, the study on the history of physics shows that physics is social endeavour. Ideas and experimental equipment are exchanged constantly. Physicists work in group, and to advance they need to collaborate and talk. I already mentioned the importance of Joe's work on thin film research, and on its historical context. It seems that we can, as now is very popular, estimate the impact of someone work based on the number of publications, h-index, and other tools. But I made another, but rather quick analysis of the publication list of Joe. This list holds almost 400 different names. What does this mean? It means that on average each peer-reviewed paper with the name of J.E. Greene as author or co-author holds the name of a different scientist. But the quick analysis learns us more. Some names on the list immediately ring a bell to anyone working on thin film sputter deposition. Let us do a small test. Eckstein (simulations of ion-solid interactions), Sproul (feedback control of reactive magnetron sputtering), Thornton (structure zone model of sputter deposited thin films), Wickersham (stress in sputter deposited thin films), Munz (Closed field magnetron sputtering), Mattox (book on The Foundations of Vacuum Coating Technology), Hultman (sputter deposition of MAX phases), Petrov (microstructural and texture evolution), and Barna (influence of impurities). So, it is clear that Joe is not only a top researcher, but also is interconnected to several research groups around the world. And my simple analysis demonstrates probably only a very small part of this scientific network.

Another point that the history of physics and material science shows us is that "Physics isn't obvious." During his short course Joe illustrates the complexity of thin film growth. Hearing it from him, it seems at first sight obvious, but when the concepts are applied to the daily life in the lab, it is clear that Nature gives us rarely a straight answer. But also this message is clearly considered in Joe's course. Moreover, by studying the history of the developments in material science and thin film research, he assist us to understand that the full comprehension of a given process takes years of research. Two examples illustrate my point. In the paper "Tracing the 5000-year recorded history of inorganic thin films from ~3000 BC to the early 1900s AD" published two years ago in Applied Physics Reviews, Joe mentions that William Grove already in 1852 realized that oxygen can form negative ions, and thus be attracted to the substrate or the anode. The role of the negative ions on the textural evolution of oxides is nowadays still a topic of interesting scientific discussions. So, a little over 160 years is not enough to pinpoint the complete mechanism. Another example illustrates how only after years of research, a complex subject is translated into an obvious and self-evident relation. It takes me back to 2011 during the plenary lecture by Joe at ICMCTF, entitled: "Fundamental Properties of Transition Metal Nitrides and Design Strategies for Growth of Self-Organized Nitride Nanostructures". Hardness is generally expressed in GPa, but Joe converts this strange unit into $eV/Å^3$, and summarizes in an elegant way several years of research on the properties of transition metal nitrides by the explanation of this unit: "*Hardness is not only about the bond strength, or the number of eVs, but also about how many of these bonds you have in the unit cell*."

The plenary talk brings me seamlessly to another point. In the Stanley paper it is claimed that history shows the need of physics of many kinds of people. In that plenary talk, Joe explained several topics from different point of views. For one topic he used a classic solid state physics approach, but then, if it easier to explain, he switched gear and tackled the topic with the vocabulary a chemist would use. Why I do remember this? In his talk, he suddenly addresses me, sitting in the audience, with the words "*This point is simple for a chemist, and just as Diederik, they would explain it to you in a few word*s". The strategy for encouraging diverse ideas is to cultivate a diverse community. And so physicists sometimes have not the most simple answer. In the tiny book "Solids and surfaces: a chemist's view of bonding in extended structures." [3] Nobel Prize winner Roald Hoffmann concludes "*... we must bring the languages of our sister sciences into correspondence*". From this perspective, you can now understand why Roger De Gryse is one of the proximi for this Sarton chair. Roger shares with Joe this viewpoint: make a diverse team, and look over the boundaries of your own daily research. As Joe will demonstrate within a few moments, he excels in crossing the boundary between material sciences and the history of sciences.

The last part of the paper by Stanley reflects on how physics isn't finished, that different strategies and research methods have been used in the past, and finally that history clearly proofs that physics doesn't have rigid rules. This is one of many points which is nicely illustrated in the review papers on the history of thin film research written by Joe. For example, the research on the mechanism of the ejection of atoms from a surface by ion impact, commonly known as sputtering, jumped over the years back and forward between an explanation based on local evaporation to a momentum transfer driven process. Even now, to understand the fundamental aspects of reactive sputtering, we look for alternative strategies. Also, Joe does not climb in his easy chair but looks for new strategies to further expand his already profound knowledge on thin film growth.

I would like to finish with perhaps the most important reason why physicists should study history, and why people from the thin film community should at least read these excellent papers from Joe. After the dinner in 2005, I occasionally met Joe, and when there was sufficient time for discussions, they were always fun. This point, the fun, returns often when you talk with Joe. So, a citation from an email I received from him regarding Joe's publication for the Sarton Journal on the history of sputter deposition, is perhaps the best way to honour him. I cite "It was a lot of work, and I learned so much, but it was fun to do it". And shouldn't that be the reason why we perform research? This statement seems so important that Stanley affirms "*I would be remiss not to mention that history of science is, frankly, fun. It is full of fascinating stories that will captivate you. Who doesn't want to know more after learning that in his experiments James Joule relied on his expertise in beer, ...*". This is of course for me, and also I think for other beer lovers, a more than genuine reason to be interested in the historical aspect of science. In summary, the fun without losing the focus on scientific correctness, is perhaps the lesson Joe teaches us very well. And I hope you'll agree that this is maybe after all the best reason to award Joe with the Sarton chair.

References

[1] M. Stanley, "Why should physicists study history?" Physics Today 69 (2016) 39-44

[2] J.E. Greene, "Tracing the 5000-year recorded history of inorganic thin films from ~3000 BC to the early 1900s AD", Applied Physics Reviews 1 (2014) 041302

[3] R. Hoffmann, "Solids and Surfaces: A chemist's view on bonding in extended structures" (1988) VCH Publishers, ISBN 0-89573-709-4

The Historical Evolution of Thin-Film Sputter Deposition

From the 1600s through the 1800s to Ghent University Today

Joe Greene

Abstract

Thin films are ubiquitous in today's world. Common examples are copper metallization layers for electronic communication among billions of transistors on silicon integrated circuits; coated architectural glass for temperature control in office buildings; coated tools to decrease wear and increase lifetime; and multiple layers on eyeglasses to correct vision, minimize ultra-violet light transmission; and provide scratch resistance. The earliest documented inorganic thin films were gold layers produced chemi-mechanically by the Egyptians approximately 5000 years ago. The Moche Indians of northern Peru developed electroless gold plating (an auto-catalytic reaction) in ~100 BC and applied it to intricate Cu masks. The first thin-films grown from the vapor phase were likely metal layers deposited accidently on the ceramic pots and rocks surrounding hot charcoal fires used to reduce metal ores, a process which can be traced back more than 7000 years. The earliest purposeful growth of metal films from the vapor phase was in 1649 when chemical vapor deposition was used to deposit arsenic in an arsenic-oxide reduction reaction.

The development of sputter deposition, starting in the mid-1800s, depended upon continued advances in vacuum technology and dc power supplies beginning in the 1600s and 1700s. Practical applications for sputter-deposited single and multilayer metal films used as mirrors and optical coatings on telescope lenses and eyepieces were discussed in papers published in the 1870s, by which time mercury pumps – which provided high pumping speeds and vacuum of the order of 10^{-4}-10^{-5} Torr – were available. Sput-

tering of alloys (including recognition of preferential sputtering) and liquid metals, multi-target sputtering, and the use of optical spectroscopy to characterize the sputtering process were reported in the late 1800s. Nobel Prizes for the discovery of x-rays (1896) and the electron (1897) emanated from early glow-discharge studies in the second half of the 1800s. Measurements of threshold sputtering energies and yields began in the late 1800s through the early 1900s, with results in reasonable agreement with modern data obtained in the 1930s to 1960s using thermionic- and mercury-pool-supported discharges.

The term magnetron, describing cathode-post and inverted hollow-cathode devices with axial magnetic fields, dates to 1921 and the first magnetron sputtering experiments to the mid-1930s. The earliest descriptions of a closed-field parallel-plate magnetron were contained in a patent filed in 1962 and a paper in 1969, with the patent for the "modern" planar magnetron filed in 1974. The first report of rf glow discharges, used today for sputtering insulators, was in 1891, with inductive rf deposition and etching experiments carried out in the 1930s. Modern capacitively-coupled rf sputtering systems were developed and modeled in the early 1960s.

The purposeful synthesis of metal-oxide films goes back to at least 1907 leading to early metal-oxide and nitride sputtering experiments reported in 1933, although the term "reactive sputtering" was not used in the literature until 1953. The effect of target oxidation on secondary electron yields and sputtering rates was reported in 1940. Commercial reactive sputtering was developed in the mid-1950s; but a clear demonstration of the mechanism of reactive sputter deposition wasn't until 1966. The first kinetic models of reactive sputtering were published in the 1970s; high-rate reactive sputtering, based on partial-pressure control, was developed and patented in in the early 1980s.

While abundant experimental and theoretical evidence (involving at least six Nobel Prize winners) existed in the late 1800s to early 1900s demonstrating that the sputtering process is due to momentum transfer involving near-surface collision cascades, the concept of sputtering due to "impact evaporation" continued in the literature into the 1960s. Modern sputtering theory is based upon a linear transport model published in 1969.

In the 1990s, two new forms of magnetron sputtering were developed, both with the goal of efficiently ionizing sputter-ejected metal atoms. In ionized-magnetron sputtering, an inductive rf coil is placed between the target and substrate of a standard dc magnetron system to efficiently ionize sputtered atoms and provide directionality with an imposed electric field in order to uniformly line or fill high-aspect-ratio vias and trenches in Si-based devices. Another novel form of magnetron sputtering, which has become known as HIPIMS (high power impulse magnetron sputtering), utilizes low-frequency, very-high-power pulses applied to the target to produce ultra-dense pulsed plasmas. The high degree of sputtered-metal-atom ionization is then used to couple kinetic energy – by high-flux, low-energy ion bombardment – to the growing film in order to control microstructure, composition, and film properties, while decreasing surface roughness.

Outline

1. Introduction

The use of thin films to enhance the physical and chemical properties of materials is ubiquitous in today's world. Examples are shown in Figure 1: copper metallization layers for electronic communication among billions of transistors in a silicon integrated-circuit; coated architectural glass in office buildings for which the thin films are designed to enhance office energy efficiency and comfort by, depending on the time of year and latitude, reflecting ultra-violet and infrared sunlight, while transmitting visible light, to minimize air conditioning usage, or reflecting infrared radiation from within offices to minimize heating; and coated cutting tools designed to reduce friction and wear during use and, hence, to increase tool lifetimes. Other common examples include magnetic thin films for electronic data storage; transparent conductive oxide and absorber layers in solar cells; thin film resistors and dielectrics; catalytic layers for toxic-gas sensing; superconducting thin films for high-frequency devices, data storage, and magnetic circuitry; corrosion-, friction-, and wear-protective layers in automotive and airplane engine parts (spark-plug electrodes, pistons, cylinders, turbine blades, etc.); and multiple layers on eyeglasses to correct vision, minimize ultra-violet light transmission, and provide scratch resistance.

Figure 1. (Left) Copper interconnect metallization (~109 connections) in a transistor, courtesy of IBM, 1997. Bus bar (metallic frame) in lower part is ~20 μm wide, about 1/5 the size of human hair (average diameter ~100 μm). Colorized SEM view after removal of insulating layers by chemical etching. http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/copperchip/. (Middle) Coated architectural glass in office buildings. Figure courtesy of Hy-Power Nano Inc. Brampton, Ontario, Canada. (Right) TiN, TiAlN, and TiB_2 (left to right) coated tools. Figure courtesy of SGS Tool Company. Munroe Falls, Ohio, USA.

The adjective "thin" in the term "thin films" is ambiguous and poorly defined. It is used to describe, depending on the application, "coating" layers ranging in thickness from less than a single atomic layer (a partial monolayer) to films which are a significant fraction of a millimeter thick. The earliest *documented* inorganic thin films were gold layers produced chemi-mechanically, for decorative (and later, optical) applications, by the Egyptians during the middle bronze age, more than 5000 years ago.[1,2] Gold films, < 3000 Å (~1250 atoms) thick, gilded on base-metal statues and artifacts using mercury-based, compositionally-graded, interfacial adhesion layers as discussed in reference. 1, have been found in ancient tombs, including the Pyramid of Djoser (actual name, Netjerykhet, second King of the Third Dynasty, Old Kingdom; ruled from ~2667 to 2648 BC)[3] in Saqqara,[1,3-5] southwest of Cairo, Egypt. Today, gold leaf can be beaten to ~500 Å thick (partially transparent to visible light) by highly-skilled craftsmen.[5] In fact, the production of gold leaf, primarily for decorative purposes, remained a viable industry for craftsmen until the development, in the mid-1930s, of roll-to-roll evaporative and sputter coating (see Section 4.11).

The gold used to produce early thin films was mined as ore by the Egyptians in the Eastern Desert, between the Nile River and the Red Sea. Ancient mining sites in Wadi Hammamat, along the trade route from Thebes (modern-day Luxor) to the Red Sea port of Elim, are accurately located on a papyrus map drawn by a scribe of Ramses IV during a quarrying expedition in approximately 1160 BC (Refs. 7 and 8) and now on display in the Museo Egizio, Turin, Italy.

Gold ore was purified by melting it in a mixture of "alum" [the mineral alunite, $KAl_3(SO_4)(OH)_6$], salt [NaCl], and chalcopyrite [e.g., $CuFeS_2$] minerals. The process produces sulfuric and hydrochloric acids (H_2SO_4 and HCl) which dissolve base metal impurities.[9,10] The purified gold still had several to a few tens of atomic percent of silver, copper, or both, depending upon where it was mined, thus giving rise to variations in color. Flattening of bulk gold was initiated by beating with a rounded stone and mechanical rolling, followed by many stages of thinning and sectioning of composite structures consisting of Au leaf sandwiched between layers of animal skins, parchment, and vellum.[6] Figure 2 is an image of a fresco from a tomb (~2500 BC) in Saqqara illustrating melting and purification of Au during which the temperature is adjusted using blow pipes.

Highly-skilled ancient Egyptian craftsmen mastered the art of gold sheathing at least as early as 2600 BC.[11,12] Sheathing is the direct application of thin gold layers onto wooden and plaster objects (mostly for noble families) to provide the impression that the object is solid gold. Striking

Figure 2. A fresco from a tomb (~2500 BC) in Saqqara, Egypt, depicting the gold melting and puri• cation process as well as the initial thinning of gold sheets with a rounded stone. The reed blow pipes, tipped with baked clay, were used to both increase and control the temperature of the charcoal • re in the ceramic pot. Adapted from Ref. 12.

examples were found in the tomb of Queen Hetepheres (wife, and half-sister, of Pharaoh Sneferu, Fourth Dynasty, Old Kingdom, ~2613-2589 BC). Other spectacular examples of early thin-film technology were found in the tomb of Pharaoh Tutankhamun ("King Tut," Eighteenth Dynasty, ruled ~1332-1323 BC). Gold sheets were beaten into position over carved wooden structures to provide embossed hieroglyphic text and decorations as shown in Figure 3.

Figure 3. A photograph of an Egyptian gold embossing, thin layers of gold cover a wooden structure with raised carved text and decorations, found in the tomb of Pharaoh Tutankhamun (ruled ~1332-1323 BC). Adapted from Ref. 11.

An autocatalytic solution-growth technique involving oxidation/reduction reactions was developed by the Moche Indians utilizing minerals available in the local area, the northern highlands of Peru, beginning ~100 BC to deposit gold (as well as silver) films on copper and bronze artifacts.[13] The general technique is still in use today and referred to as electroless plating.[14] The Moche peoples first dissolved gold in a hot aqueous solution of equal parts potassium aluminum sulfate [$KAl(SO_4)_2$], potassium nitrate [KNO_3], and salt [NaCl], a process that took several days. The resulting mixture was then buffered with sodium bicarbonate [$NaHCO_3$] to form a weakly alkaline solution (pH ~9) which was allowed to boil for several minutes before immersing the copper artifact to be plated. The overall reaction is

$$2AuCl_3 + 3Cu \rightarrow 2Au + 3CuCl_2 .$$

Metallographic studies of Moche artifacts, coated with gold films whose thicknesses ranged from ~2000 Å to 1 μm, exhibit evidence of post-deposition heat treatment (annealing) to obtain a film/substrate interdiffusion zone, presumably for better adhesion. An excellent example of craftsmanship is depicted in Figure 4.

Figure 4. An electroless gold-plated copper mask discovered near Lorna Negra (northern Peru, close to the Ecuadorian border). Adapted from Ref. 13.

2. History of thin film deposition from the vapor phase

The first thin films grown from the vapor phase, as discussed in ref. 1, were likely metal layers deposited accidently on the ceramic pots and rocks surrounding hot charcoal fires used to reduce metal ores, a process which can be traced back more than 7000 years.[15] Based upon both archeological evidence and metallurgical analyses, copper smelting (extraction from ore) and metal working appear to have originated independently in the Balkans (Serbia and Bulgaria) ~5500 BC and in Anatolia by at least 5000 BC.[16-19] The Roman philosopher Pliny the Elder (23-79 AD) discusses the process in his 37-book Naturalis Historia, the first encyclopedia, published ~79 AD.[20]

The earliest reported purposeful growth of metal films from the vapor phase was in 1649 when Johann Schroeder, a German pharmacist, described a method for reducing arsenic oxide [As_2O_3] with charcoal[21,22] through the overall endothermic reaction

$$2As_2O_3 + 3C \rightarrow 4As + 3CO_2 .$$

This is an example of a vapor deposition technology termed chemical vapor deposition.[1,23] As discussed in a 1966 review article by Rolsten,[24] carbon reduction of oxides was an important method for obtaining relatively pure metals, in order to investigate their physical properties, during the 1700s and early 1800s.

In honor of George Sarton (1884-1956), born and educated in Ghent (PhD in physics and mathematics from Ghent University in 1911), who developed the history of science as an academic discipline,[25,26] I will focus on physical vapor deposition (PVD) and, in particular, on glow-discharge sputter deposition for which the Physics Department Ghent University is a well-known worldwide center of excellence.

Physical vapor deposition, sputtering and evaporation, was developed in the mid to late 1800s following the evolution of power supply and vacuum technologies beginning in the mid-1600s. The first publication on sputter deposition was in 1852,[27] as discussed in Section 4.1, while the first documented evaporation experiments were carried out in the early 1870s by Josef Stefan (1835-1893),[28-30] an Austrian physicist best known for his research in thermal conductivity of gases and blackbody radiation of solids,[31] and the early 1880s by Heinrich Hertz (1857-1897),[32,33] a German physicist well known for his work on electromagnetics[34] and contact mechanics.[35]

Sputter deposition, in its simplest configuration (Figure 5), is carried out in a vacuum vessel which is backfilled with a low pressure of a rare gas such as argon. (Argon is often the rare gas of choice primarily due to the fact that it comprises approximately 1% of our atmosphere and, hence, is relatively inexpensive). A voltage is then applied between a metal target (the source of the film atoms) such as Cu and an electrically-conducting substrate upon which the film is deposited. The voltage electrically breaks down the gas to form a glow discharge consisting of Ar^+ (with a much smaller fraction of Ar^{2+}) ions and electrons. The positively-charged ions are accelerated to bombard the target and, via momentum transfer, sputter-eject target atoms, some of which are deposited on the substrate. A reactive gas can also be added in order to form compound films (see Section 4.10). Important commercial examples are the reactive sputtering of titanium and titanium-aluminum alloys in Ar/N_2 mixtures to deposit hard ceramic TiN and TiAlN coatings[36-38] on cutting tools, drill bits (see Figure 1 (right panel)), gear hobs, and dies.

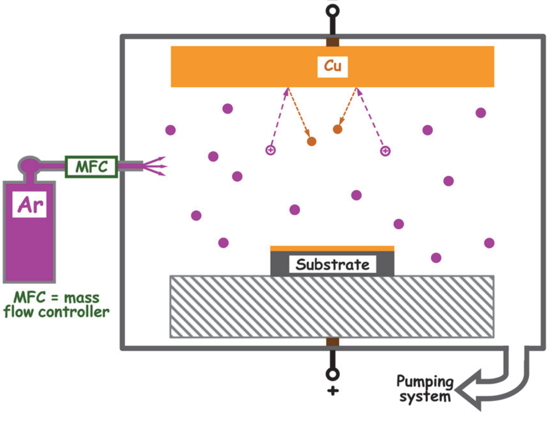


Figure 5. Schematic illustration of the essential features of a basic dc sputtering system. Argon gas flows through a controlled leak valve into a vacuum system; some Ar atoms are ionized by a dc potential between the copper target and an electrically-conducting substrate on a metal substrate table. Ar^+ ions are accelerated toward the target and sputter-eject copper atoms, which are deposited on the substrate, as well as the chamber walls, to form a Cu film.

Historical footnote: While sputtering was in use for the deposition of thin films by the mid-1800s, the etymology of the word sputtering remains unclear. The term "spluttering," an intensified form of the English word sputtering, meaning "to spit with explosive sounds" (a cognate for the Dutch word "sputteren"),[39] may have been used as early as the late 1800s. In a 1970 book chapter, Gottfried (Fred) Wehner (German-born American physicist, 1910-1996) and Gerald Anderson[40] noted that a search of the literature revealed that Joseph John [J. J.] Thomson (English physicist, 1856-1940; Nobel Prize in Physics, 1906) still used the term spluttering in 1913: "A well-known instance of this is the spluttering of the cathode in a vacuum tube;….".[41] The quote implies that the term was in use even earlier. Mattox points out[42] that the 3rd Edition (1955) of the Shorter Oxford Dictionary[43] lists, in addition to sputter (verb, 1598) and sputter (noun, 1673), the English term splutter meaning "to spit out a spray of particles in noisy bursts."

Kenneth H. Kingdon and Irving Langmuir (American surface chemist, 1932 Nobel Laureate in Chemistry for his work in surface science) at the General Electric Research Laboratory dropped the "l" in favor of the word sputtering in their 1923 paper on "The removal of thorium from the surface of a thoriated tungsten filament [from a light bulb] by positive ion bombardment."[44] Nevertheless, in the same year, in an article on the sputtering of tungsten published in the *Philosophical Magazine* by the "Research Staff of the General Electric Company and communicated by the Laboratory Director," the term "cathode disintegration" was used in place of sputtering.[45] With time, however, the term sputtering prevailed and is now used universally.

3. Technological advances in the 1600s and 1700s necessary for the development of sputter deposition

The rapid development of sputter deposition in the middle to late 1800s required the evolution of vacuum technology, beginning in the 1600s, and the invention of dc power supplies (batteries) in the late 1700s and early 1800s. For a more detailed discussion of early vacuum technology and power supplies, see ref. 1.

3.1. Evolution of early vacuum technology

Progress in vacuum technology (the word vacuum derives from the Latin *vacuus*, meaning empty space), from the mid-1600s, was essential for providing cleaner deposition environments necessary for the advancement of thin-film science. In 1652, Otto von Guericke (1602-1686) of Magdeburg, Germany, a scientist, inventor, and politician, developed a mechanical piston pump that achieved a vacuum of 2 Torr (~0.003 atm).[46,47] (For comparison, a typical household vacuum cleaner produces enough suction to reduce standard atmospheric pressure, 760 Torr, to ~610 Torr.)[48] von

Guericke's third-generation vacuum system, a model of which is shown in Figure 6,[49] consisted of a bell jar separated from the piston pump by a cylinder with a stop-cock. The pump was equipped with two valves near the entrance to the nozzle at the bottom of the bell jar, the first valve was located between the nozzle and the cylinder and the second valve between the cylinder and atmosphere. During the piston down-stroke, valve one is closed to stop air from entering the nozzle and bell jar, while valve two is forced open by the air displaced from the cylinder. During the piston return-stroke, valve two is closed, and valve one is forced open by the pressure of the remaining air in the bell jar and nozzle. The percentage of pressure decrease per complete piston stroke diminishes continuously as the bell jar pressure is reduced toward the base pressure.

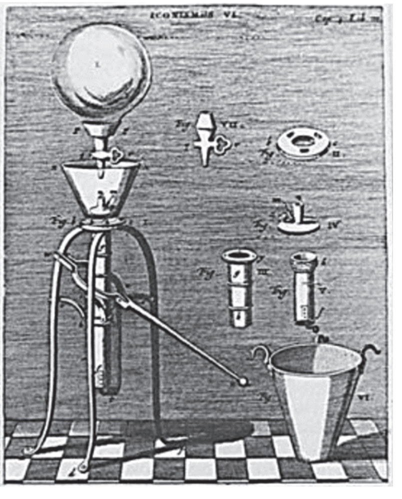


Figure 6. A model of an early mechanical piston pump developed by Otto von Guericke in ~1652. Adapted from Ref. 49.

Much better vacuum was required in order for scientists in the 1800s to obtain longer gas-phase mean free paths, higher deposition rates, and increased purity during thin-film growth from the vapor-phase. This was solved by a German chemist, Hermann Sprengel (1834-1906), who developed a practical mercury momentum-transfer pump in 1865.[50] The Sprengel pump was an improvement over the original mercury pump invented by Heinrich Geissler (1814-1879), a German glassblower, in 1855.[51] The base pressure claimed by Sprengel in his initial publication was ~6x10^{-4} Torr, and limited by leaks in vulcanized rubber joints connecting the glass tubes (the rubber connectors were cemented to the glass tubes and the joints were bound with copper wire). While lower pressures were achieved with later versions of the pump,[52] pressures of 10^{-3}-10^{-4} Torr were sufficient to provide ballistic environments (i.e., gas atom mean free paths of the order of, or larger than, system dimensions) for investigating gas discharges and sputter deposition in the small evacuated chambers of that era.

An initial prototype of the Sprengel mercury pump is shown in Figure 7(a).[50] Droplets of mercury (a heavy metal which is liquid at room temperature), falling through a small-diameter (2.50-2.75 mm) glass tube, trap and compress air by momentum transfer. The tube, labeled CD in Figure 7(a), was ~76 cm long and extended from the funnel A to enter the glass bulb B through a vulcanized-rubber stopper. The bulb has a spout several mm above the lower end of tube CD.

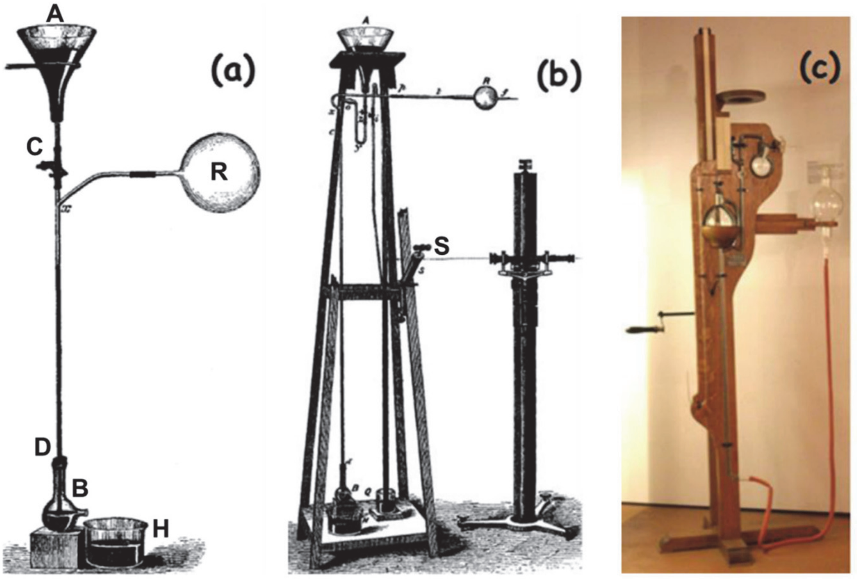


Figure 7. Drawings of (a) a prototype and (b) an initial version of Sprengel's mercury transfer pump. Adapted from Ref. 50. (c) A later version of the pump, presently housed in the Dr. Guislain Museum, Ghent, Belgium. Photograph courtesy of Luca Borghi for Himetop, The History of Medicine Topographical Database.

In operation, mercury was added to the funnel A, and the stopcock at C opened, allowing mercury droplets to fall, trap air, and reduce the pressure in chamber R. Air and mercury were exhausted through the spout of bulb B. The mercury collected in basin H was poured back into funnel A for continued pumping. The second version of the Sprengel pump, described in the same paper,[50] is shown in Figure 7(b). It was approximately 1.8 m tall and Sprengel reported using 4.5 to 6.8 kg of mercury during operation. The pump contained a mercury pressure gauge (described below) attached to the evacuated chamber and a mechanical-piston backing pump S. Later versions incorporated continuous mercury recycling. With the combination of the mechanical and mercury pumps, a 0.5 liter chamber could be evacuated in ~20 min. The importance of Sprengel's work was recognized by the Royal Society of London who elected him as a Fellow in 1878. A later version of the pump, presently housed in the Dr. Guislain Museum, Ghent, Belgium, is shown in Figure 7(c).

Improvements in vacuum technology required better gauging in order to measure the increasingly lower pressures produced. In 1874, Herbert McLeod (1841-1923), a British chemist, developed what today is termed the McLeod gauge[53,54] which operates based upon Boyle's law. Robert Boyle (1627-1691), another British chemist, showed in 1662 that for a closed system at constant temperature, the product of the pressure P and volume V remains constant.[55] In operation, the gauge compresses a known volume V_1 of gas at the unknown system pressure P_1 to a much smaller known volume V_2 in a mercury manometer with which the pressure P_2 is measured.[56,57] Thus, by Boyle's law, the system pressure P_1 is given by the expression P_2V_2/V_1.

3.2. Pulsed to dc power supplies

In addition to vacuum, electrical power was necessary for initiating early experiments in thin-film sputter deposition. von Guericke also played an important role in this field through his development in 1663 of a crude friction-based electrostatic generator which transformed mechanical work into electrical energy.[58,59] The generator was based on the triboelectric effect (although the term did not exist at the time), in which a material becomes electrically charged ("static electricity") through friction.

> ***Historical footnote***: The concept of static electricity was known by the ancient Greeks (e.g., rubbing amber on wool) and first recorded by Thales of Miletus (624-546 BC),[60] a pre-Socratic Greek philosopher, mathematician, and one of the Seven Sages of Greece.[61-63]

In 1745, the Dutch scientist Pieter van Musschenbroek (1692-1761) of Leiden University (mathematics, philosophy, medicine, and astrology [the latter is closer to theology than science!] and Ewald Georg von Kleist (1700-1748), a German lawyer, cleric, and physicist, are credited with independently inventing what today is known as the Leiden jar,[64,65] an early form of the modern capacitor, to provide pulsed power. However, it appears that Professor Musschenbroek obtained the idea for his research from Andreas Cunaeus (1712-1788), a lawyer who often visited Musschenbroek's laboratory. Cunaeus carried out the initial experiments that led to the Leiden jar[66] while attempting to reproduce even earlier results by Andreas Gordon (1712-1751), a Professor at Erfurt, Germany, and Georg Mattias Bose (1710-1761) at the University of Wittenberg, Germany.[67] The device accumulates static electricity between electrodes on the inside and outside of a glass jar.

In order to store charge in Leiden jars, the glass cylinder of an electrostatic generator was rotated, via a hand crank, against a leather (or wool) strip pressing on the glass. The friction resulted in positive charge accumulating on the leather and negative charge (electrons) on the glass. The electrons were collected by an insulated [perhaps comb-shaped] metal electrode. When sufficient charge accumulated, a spark jumped from the generator collector to the central collector electrode of a nearby Leyden jar, where the charge was stored. Originally, the capacitance of the device was measured in units of the number of "jars" of a given size, or by the total area covered with metal. A typical 0.5 liter Leyden jar had a capacitance of about 1 nF.[68]

The invention of the electrochemical battery to provide low-voltage dc power is generally attributed to Count Alessandro Volta (1745-1827), Professor of Natural Philosophy at the University of Pavia, Italy, based upon his work in the 1790s resulting in a classic paper published first in French,[69] then in English,[70] in 1800.

Volta's interest in electrochemistry led him to discover that voltage can be obtained from stacks consisting of several pairs of different metal discs, each pair separated by an electrolyte (initially pieces of cloth saturated in brine), connected in series to form a "voltaic pile."[71] The first metals used were copper and zinc, but Volta found, based upon electrometer measurements, that silver and zinc produce a larger electromotive force, a term Volta introduced in 1796.[72] Figures 8(a) and 8(b) are an illustration and a photograph, respectively, of an early voltaic pile. Such devices could only provide a few volts; obtaining larger potentials required a series (i.e., a "battery") of many large voltaic piles. An example of a small double voltaic pile is shown in Figure 8(c).[70]

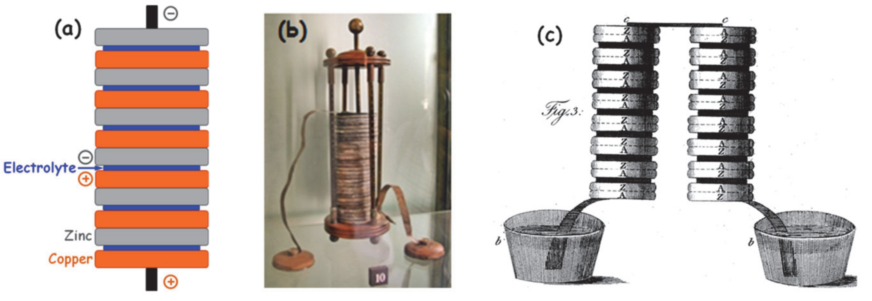


Figure 8. (a) Schematic illustration of a voltaic pile. (b) Photograph, attributed to GuidoB and licensed under the Creative Commons Attribution-Share Alike 3.0 Unported, of a single voltaic pile. The battery is on display at the Tempio Voltiano Museum, Como, Italy. (c) Sketch of a double voltaic pile consisting of two sets of eight pairs of silver and zinc plates. Adapted from Ref. 70.

Historical footnote: Volta not only introduced the term electromagnetic force, but was the first to use the term "semiconductor" in describing "materials of semiconducting nature" in a 1782 paper published in the Philosophical Transactions of the Royal Society[73] and presented at a Royal Society of London meeting on March 14 of that year.

A practical problem with voltaic piles, especially with larger ones used to obtain higher voltages, is that the weight of the discs squeezes electrolyte out of the cloths. In 1801, William Cruickshank (1745-1810), a surgeon and Professor of Chemistry at the Royal Military Academy, Woolwich (southeast London), solved this problem and designed the first electric battery for mass production.[74] In the initial version, Cruickshank arranged 60 pairs of equal-sized zinc and silver sheets cemented together with rosin and beeswax in a long resin-insulted rectangular wooden box such that all zinc sheets faced one direction and all silver sheets the other. Grooves in the box held the metal plates in position, and the sealed box was filled with an electrolyte of brine, or dilute ammonium chloride [NH_4Cl] which has higher conductivity.

4. History of thin-film sputter deposition

4.1. Glow-discharge sputtering

From the mid to late 1800s, several papers were published on the use of optical emission to investigate the transition between a continuous glow discharge and an arc, the electrical structure and configuration of dc glow discharges, and the nature of metal atoms sputter ejected from the cathode (target). Michael Faraday (1791-1867), an English scientist, reported on the electrical and optical characterization of glow discharges in 1838.[75] His glass discharge tube contained brass electrodes and was operated in air, nitrogen, oxygen, hydrogen and other gases at a pressure of 4.4 inches of mercury [~112 Torr]. While Faraday must certainly have deposited films during these experiments, this was not the objective. His later work on the optical properties of vacuum-arc deposited thin metal films was discussed in his fifth Bakerian Lecture[76] in 1857 and published in the same year.[77]

***Historical footnote*:** The Bakerian Medal and Lecture of the Royal Society of London was established by a gift from Henry Baker (1698-1774) in 1775. It is awarded annually to a person (one per year) working in the fields of "natural history or experimental philosophy" (i.e., the physical sciences). Baker was described by G. L'E. Turner, Senior Assistant Curator of the Museum of Science in Oxford:[78] "Henry Baker was a typical polymath in the eighteenth-century manner. Although he did not contribute to scientific research in any significant way, he did make valuable contributions to the dissemination of scientific knowledge, particularly in the field of microscopy. .... an enthusiastic participator in the scientific and literary life of London."

***Historical footnote*:** Michael Faraday, known primarily for his research in electromagnetics and electro-chemistry, is considered by science historians to be one of the most influential scientists, and the best experimentalist, in history,[79,80] although he had no formal education past grade school. Faraday, during the years between 1829 and 1857, was invited to present five Bakerian Lectures to the Royal Society.

In 1852, William Robert Grove (1811-1896), a Welsh lawyer (later, judge) and physicist published the earliest recorded description of sputter deposition and ion etching experiments.[27] A sketch of his equipment is shown in Figure 9. Vacuum was achieved with a mechanical piston pump, similar to that of von Guericke as described in Section 3.1, with power supplied by Grove's version of a trough-style dc voltaic pile (similar to that of Cruickshank, Section 3.2)[74] combined with a step-up transformer supplied by Heinrich Ruhmkorff (1803-1877),[81] a famous German instrument maker living at the time in Paris. The electrodes consisted of a copper plate, with a polished electroplated silver surface, and a rod, which passed through a leather stopper in the top of the glass vacuum chamber, with a steel needle attached to its end. The gas used to sustain the discharge was stored in a bladder.

Historical footnote: Based upon a passing comment in William Grove's later papers, the small vessel attached to the rod electrode shown in Figure 9 apparently contained "potassa fusa" (an early name for potassium hydroxide [KOH], a caustic deliquescent desiccant which can capture large quantities of water).

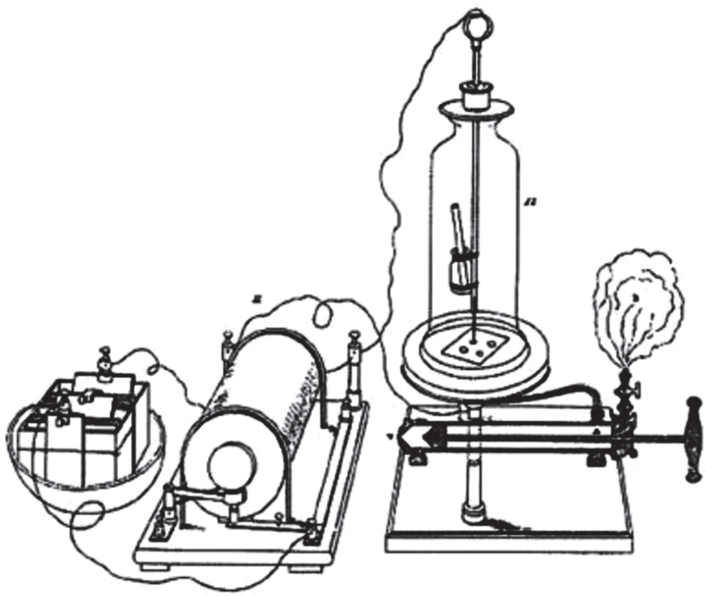


Figure 9. The system used by William Grove to investigate target "disintegration" (sputtering) in a gas discharge. Adapted from Ref. 27. See text for details.

The experiments were carried out at rather high pressures, ranging from ~100 to 500 mTorr, with the steel needle target quite close to the silver-plated substrate (generally a separation of 0.25 cm, "but this may be considerably varied"). When using a mixture of hydrogen and air with the silver plate positive and the steel needle serving as the cathode, Grove observed thin-film deposition on the silver substrate. The layer was primarily iron oxide, i.e. reactive sputtering [although the term didn't come into existence until more than than a century later, Section 4.10]. The color of the oxide film "presented in succession yellow, orange, and blue tints" with increasing thickness (longer deposition time). Grove was reporting optical interference effects (as he noted later in the paper), which for a given substrate/film combination can be calibrated to provide film thickness vs. color as is commonly done today for SiO_2 and Si_3N_4 dielectric layers on Si(001) wafers used in microelectronic device fabrication.

When Grove switched polarity to make the silver plate the cathode (negative), he reported that the iron oxide film was removed by ion etching. In actuality, there must have been a thin silver oxide layer remaining due to the competition between the steady-state rates of silver oxidation from the discharge, arising from the relatively poor vacuum, and sputter etching. However, this layer, a few tens of Å in thickness,[82] would have been too thin for Grove to observe. Continuing the experiment, a polished region "occasioned by molecular disintegration" remained. Thus, Grove had not only removed the original oxide film, but had sputter etched into the silver substrate layer. The word "sputter" did not yet exist (as discussed in Section 2), and Grove described the process throughout the paper as "molecular disintegration."

Grove repeated the above experiment by sputtering the steel target in an "air vacuum" [a term Grove attributed to Faraday] to produce a more fully-oxidized film on the silver-plate copper substrate, then switched gases and electrode polarity to "sputter clean" the silver plate in a nitrogen discharge. Several more experiments in which Grove substituted different metals for the target needle, and changed discharge gases, were also reported. The results were similar, but he describes observing differences depending upon the mass and ionization potential of the gas and the oxidation tendency of the metals.

Interestingly, Grove realized that oxygen can form negative ions which are accelerated by the applied target voltage toward the substrate (anode). The significance of this fact, that oxygen has a high electron attachment probability,[83] was not fully appreciated until almost 130 years later. Researchers investigating the growth of piezoelectric and transparent conducting oxides in the early 1980s[84-86] and high-temperature oxide superconductors in the late 1980s,[87] all of which are typically deposited today by reactive magnetron sputtering (see Sections 4.6 and 4.10), were confronted with the deleterious effects of O^- and O_2^-. Negative ions, accelerated by the same potential used to produce sputtering by positive ions incident at the target, bombard the growing film with energies which can produce residual defects, change the preferred orientation of polycrystalline layers, degrade film properties, and decrease deposition rates by resputtering.[82,84-88] Fast neutral O and O_2 irradiation of the growing film can also occur as attached electrons are stripped in the plasma giving rise to substrate bombardment by an additional flux of energetic species.[85,87,89] Early solutions involved increasing the discharge pressure in order to decrease particle mean free paths and hence lower the average energies of ions and fast atoms via collisions,[86] while later solutions focused on off-axis deposition and facing-target sputtering (see Section 4.6).[86,90-94]

A few years after Grove's seminal paper, John Gassiot (1797-1877), a highly-successful English businessman and gentleman scientist, presented his Bakerian Lecture (March 4, 1858) "On the stratifications and dark bands in electrical discharges as observed in Torricellian vacuums," which was published in the Proceedings of the Royal Society of London[95] (see also, ref. 96). Most of the lecture was concerned with his observations of alternating bright and dark bands formed in rarefied-air discharges operated in glass tubes partially filled with clean boiled mercury and evacuated with a mechanical pump [as air is extracted, the mercury level sinks in the tube and a "Torricelli vacuum" is formed].[1] During these experiments, Gassiot noted that the luminous discharge regions move under the influence of a magnetic field. However, he concedes in the conclusion of ref. 95 that "I refrain for the present from any observations as to the action of the magnet on the discharge."

Gassiot also reported that for discharges formed between two platinum wire electrodes hermetically sealed about 4 inches [8 cm]) apart in a discharge tube: "a black deposit takes place on the sides of the tube nearest the negative terminal. This deposit is platinum [analyzed by Michael Faraday][96] in a state of minute division emanating from the wire, which becomes black and rough as if corroded. The minute particles of platinum are deposited in a lateral direction from the negative wire, and consequently in a different manner from what is described as occurring in the voltaic arc."[95] "The platinum coating is deposited on the portion of the tube surrounding the negative wire, but none at or near the positive."[96] Gassiot was describing sputter deposition from the platinum target. He also noted that ".... when this deposit is examined by transmitted light it is translucent, presenting to the eye an extremely thin bluish-black film; but by reflected light, either on the outside or inside [i.e., viewed either from the glass tube or the film side], it has the appearance of highly polished silver, reflecting the light as from the finest mirror."[96]

Historical footnote: John Gassiot, in addition to being a successful businessman, was a very enthusiastic amateur scientist interested in electricity. He maintained a well-equipped laboratory and library in his home where a young James Maxwell (1831-1879), a Scottish mathematical physicist who developed electromagnetic theory and presented a unified model of electricity, magnetism, and light,[97] did much of his own scientific work during the 1860s. Gassiot was a founder of the London Electrical Society in 1837, the Chemical Society in 1845, and was elected as a Fellow of the Royal Society in 1841. He was also a close associate of William Grove.

Practical applications for sputter-deposited single and multilayer metal films as mirrors and optical coatings on telescope lenses and eyepieces were discussed in papers published in 1877 by Arthur Wright (American physicist, 1836-1915, Yale University).[98,99] He reported the growth of adherent noble-metal films sputter-deposited from wire targets onto glass microscope slides.[98] Unfortunately, the films had large lateral thickness variations since the target was the tip of a wire whose length was encased in a glass tube. However, an ingenious solution was presented in Wright's next paper.[99] He designed a deposition system, evacuated with a Sprengel mercury pump (see Section 3.1), in which the substrate was mounted on a pendulum which allowed motion in two orthogonal directions with respect to the target, such that films of uniform thickness could be "painted" onto the substrate (Figure 10). In Wright's words: "The perfect control of the process obtained by the use of the movable electrode will even make it possible to apply the method of local correction for the improvement of a defective figure, or to parabolize a spherical mirror by depositing the metal in a layer increasing in thickness toward the center."

Historical footnote: Arthur Williams Wright, in 1861, was among the first three people to obtain a Ph.D. in the United States; his doctoral dissertation at Yale University was on satellite mechanics.[100,101] As an undergraduate and graduate student, he studied mathematics, mineralogy, botany, and modern languages, in addition to physics. He also studied law and was admitted to the bar. Following teaching positions and a Post-doctoral research program at Heidelberg and Berlin, Wright became Professor of Molecular Physics and Chemistry at Yale University in 1872, later Professor of Experimental Physics. His sputter-deposited thin films were used extensively in the first studies of the polarization of light emitted from the solar corona [plasma surrounding the sun, most easily observed during a solar eclipse]. Wright became a member of the US National Academy of Sciences and a Fellow of the UK Royal Astronomical Society.

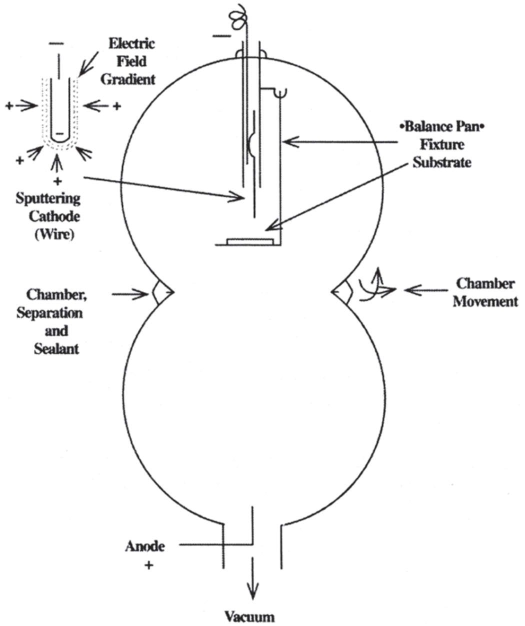


Figure 10. An illustration (Ref. 102) of Wright's sputter deposition system, based upon his description in reference 99.

Wright characterized the sputtering process spectroscopically using optical emission from gas and ejected target atoms which were excited in the discharge (following earlier work by Faraday).[75] As-deposited platinum films, some with thicknesses < 350 Å (estimated using a combination of weight change, to within 10 μg, for thicker films, deposition rate calibrations, and optical interference rings for thinner layers) were analyzed using optical transmission as a function of wavelength. Mirror-like sputter-deposited films were found to be more adherent than solution-grown layers and less sensitive to local delamination caused by water penetration to the film/glass interface. By the late 1800s, sputter-deposition was routinely used in manufacturing commercial mirrors.

Wright describes his films as "….surfaces of exquisite perfection and the most brilliant polish. They can only be compared to the surface of clean liquid mercury, far surpassing in luster anything that can be obtained by the ordinary methods of polishing." Wright tuned the reflectivity of his mirrors based upon interference effects to obtain brilliant "white light" by depositing multilayer films with predetermined layer thicknesses.

William Crookes (1832-1919), a British chemist, promoted research on gas-discharge electronics using what are now referred to as Crookes discharge tubes (Figure 11(a)).[103] Several scientists during the late 1850s and throughout the 1860s, including Julius Plücker[104] and his ex-graduate student Johan Willhelm Hittorf,[105] carried out important experiments on electronic effects at ion bombarded targets (see Section 4.4) using Crookes tubes which were also instrumental in the discoveries of x-rays[106] (Wilhelm Röntgen, 1896, Nobel Prize in Physics, 1901), electrons[107] (J.J. Thomson, 1897, Nobel Prize in Physics, 1906), and, when combined with thermionic emission, vacuum tube electronics.[108,109]

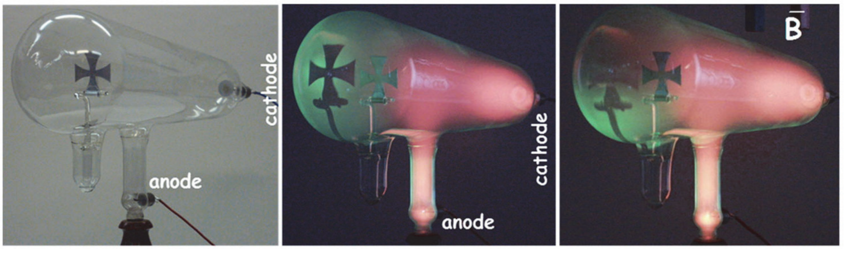


Figure 11. Reproductions of a Crookes' tube, containing a Maltese cross between the cathode and the far end of the glass envelope, used to investigate the electronic characteristics of, and the optical emission from, glow discharges. (Left). No power applied to the tube. (Middle). Power applied to the cathode giving rise to green fluorescence emanating from the glass behind the Maltese cross which casts a shadow by blocking "cathode rays (radiant matter)" from the target. (Right). A magnet was used to rotate the shadow image. Photographs attributed to Zátonyi Sándor, GNU Free Documentation License; the labels were added by the present author.

Crookes published a very significant paper on sputter deposition of thin films in1891.[110] During a long series of experiments, he employed a Sprengel-type mercury pump[50] to evacuate his discharge tube to pressures of the order of $7x10^{-4}$ Torr. He then used the residual air to sputter Ag, Al, Au, Cd, Cu, Fe, Ir, Mg, Ni, Pb, Pd, Pt, and Sn targets as well as AlAu and CuZn (brass) alloys and measure the erosion rates via target weight loss.

Although the experimental details (sputtering pressure, voltage, and ion current densities) were not well specified, Crookes, in order to obtain comparable results, designed a multi-target sputtering system with indexed motorized external electrical contacts as illustrated in Figure 12.[110] Each experiment was carried out using four wire targets, 0.8 mm in diameter by 20 mm in length, in which one of the four was always a gold reference electrode. Power was alternately applied to each target in succession, using a revolving commutator, for the same length of time (typically 6 s) over periods of several hours. By this means, variations in current and sputtering pressure were accounted for in order to obtain a set of relative metal sputtering rates, all referenced to that of Au. Aluminum and magnesium targets were reported to be "practically non-volatile." Today we know that this was due to the formation of strong oxynitride dielectric layers at the target surfaces due to the use of air (approximately 78% nitrogen and 21% oxygen by volume) as the sputtering gas.

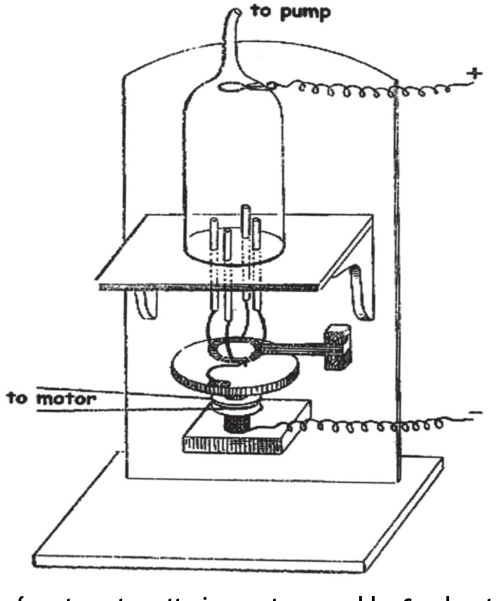


Figure 12. The four-target sputtering system used by Crookes to measure the sputtering rates of different metals. The targets were 0.8-mm-diameter metal wires. Adapted from Ref. 110.

The targets in Crookes' sputtering-rate experiments were uncooled. Thus, low-melting-point metals such as tin, cadmium, and lead quickly melted. For these materials, he devised a holder for sputtering liquid metals. It is likely that a significant part of their measured weight loss, especially for the high-vapor-pressure element cadmium, was due to evaporation, in addition to sputtering.

Crookes, in his April 4, 1879 Bakerian Lecture,[111] made several important observations. He, like Faraday[75] and Gassiot,[95,96] used optical spectroscopy, in many discharge-tube configurations, to characterize the plasma at working pressures from "0.08 mm Hg [80 mTorr] down to a low of 0.00001 mm Hg [1x10^{-5} Torr]."[112] He often specified the sputtering pressure not in pressure units, but as the thickness of the cathode dark space (inversely proportional to the pressure); which later came to be known as the Crookes dark space adjacent to the very bright glowing region immediately in front of the target. He noted that during long sputtering runs, it was necessary to periodically bleed some air into the discharge tube "to reduce the vacuum." After examining metal films deposited on the inside of the discharge tube and finding them to be porous with rough surfaces, he concluded that sputtering gas was being "occluded" in the growing films. While some gas was likely trapped in the growing films, most was captured via reactions with the fresh metal layers, deposited on surfaces throughout the system, which acted as a getter pump.

The "rich purple color" of the aluminum-gold alloy target turned to the "dull white color of aluminum" as gold was referentially sputter removed. This was the first mention in the literature of preferential sputtering from alloy targets. While this occurs for all alloy targets, the surface composition during sputtering in rare gases rapidly (depending on the ion energy, current density, and relative ion and target-atom masses) reaches a steady-state value as the surface coverage of the low-sputtering-rate component increases to compensate the difference in elemental removal rates as shown experimentally by Tarng and Wehner[113] and theoretically by Eltoukhy and Greene.[114] In Crookes' experiments, the formation of an aluminum oxynitride layer on the target dramatically decreased the aluminum sputtering rate as noted above.

In 1902, Thomas Edison (1847-1931), an American inventor and businessman, patented a very early forerunner of modern Cu contact technology in microelectronic device fabrication. U.S. patent 484,582 ("Process of duplicating phonograms") describes the use of dc sputtering for the deposition of metal films on wax phonograph masters as a "seed" coat for electroplated overlayers.[115] This follows an earlier 1892 Edison patent in which the seed layers were deposited by vacuum-arc deposition.[116] Edison claimed in the 1902 patent that the arc process was too slow and that sputter-deposited films had much more uniform thickness distributions.

Historical footnote: Thomas Edison, a prolific inventor who was issued 1,093 US patents (phonograph, motion picture camera, sound recording, etc.) and many patents in other countries, is often credited with the invention of the light bulb. While Edison was issued a U.S. patent for an "Electric lamp" in 1880,[117] he didn't "invent" the light bulb, he took advantage of the availability of better vacuum and the development of the mercury momentum transfer pump[50] by Hermann Sprengel in 1865 (see discussion in Section 3.1) to develop a much longer-lived bulb which was commercially viable. In fact, a year before Edison was born, William Grove, who published the earliest recorded description of sputter deposition and ion etching[27] as discussed above, used a platinum-filament electric light to illuminate the lecture theater[118] during his first Bakerian Lecture before the Royal Society on November 19, 1846, as he described the use of his improved voltaic dc battery to dissociate water:[119,120] "On Certain Phenomena of Voltaic Ignition and the Decomposition of Water into its Constituent Gases by Heat." The history of the light bulb is rich and interesting, stretches back to at least 1802, and involves many previous researchers as chronicled in Ref. 121.

Grove's and Crookes' research on sputtering attracted the attention of scientists worldwide. A review paper, entitled "Cathode Sputtering, a Commercial Application," published in 1932 by H.F. Fruth,[122] of Western Electric Company (Chicago), lists 113 references published in the field between the time of Grove's 1852 pioneering article[27] and 1930. Fruth also describes commercial equipment (Figure 13(a)) and procedures for sputter-depositing gold electrodes, from six gold cathodes, onto multiple radio-broadcasting microphone diaphragms. A photograph of the deposition chamber, which contains a rotating McLeod gauge[53] and a "bleeder" valve in order to maintain constant pressure, with diaphragms ready to be coated, is shown in Figure 13(b). Fruth describes the system operation as follows.

"In order to maintain a constant residual gas pressure, the pump is operated continuously and air is allowed to leak in slowly through the bleeder valve which is located near the pump. This practice was found necessary in order to overcome variations in pressure due to the early evolution of gases and the later cleanup usually accompanying electrical discharges *in vacuo*. A

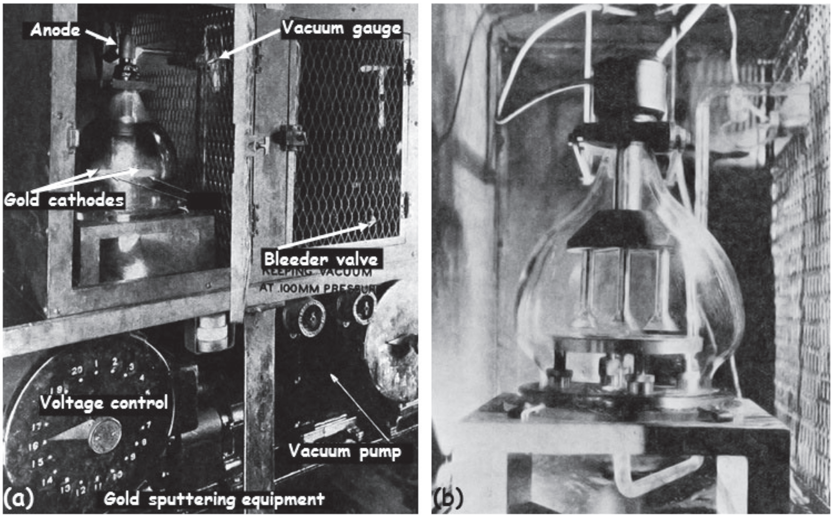


Figure 13. (a) A commercial sputter-deposition unit, with six gold targets, for depositing metal electrodes on microphone diaphragms. (b) A closer view of the deposition chamber, showing the diaphragms. Adapted from Ref. 122; labels were added by the present author.

> pressure of 0.100 mm [100 mTorr] is readily maintained by this method. After a new charge has been placed in the bell jar, the bleeder valve is temporarily cut off by closing a stopcock so that the required vacuum can he more quickly obtained. By this means, sputtering can be started in about 4 min after the bell jar has been placed in position."

Fruth demonstrated that dc sputter-deposited gold films, < 1 μm thick, offer substantial lifetime advantages over previous electroplated films which developed "blisters," peeling, and pinholes after three months of continuous use, while the sputter-deposited films exhibited no sign of wear or degradation.

An early ion-beam source was developed by Louis Maxwell in 1931.[123] A hydrogen discharge operated at pressures of 3.5 to 120 mTorr, with a liquid air cold trap to remove water vapor and minimize mercury contamination from the pump, was established in a small brass vacuum vessel with the ion current controlled by thermionic electron emission from a hot, low-work-function, filament. Large magnetic fields, ~12-17x10^3 G, parallel to the positive ion beam were used to minimize ion losses to the wall and extract

0.1 to 3 mA though a 1-mm-diameter circular extraction electrode to a collector electrode in a small attached chamber maintained at $3.3x10^{-3}$ to $1.5x10^{-4}$ Torr. The first recorded description of a dc glow discharge ion-beam sputtering system was by R. Seeliger and K. Sommermeyer in 1935.[124] They drilled a 2-mm-diameter hole in the cathode of their discharge tube to "collimate" an Ar^+ ion beam [the beam was actually divergent] to strike solid silver or liquid gallium targets at energies of 5-10 keV and observed that sputtered atom emission can be approximated by a cosine distribution.

In 1960, Wehner and Rosenberg,[125] using a mercury-pool-supported glow discharge (see Section 4.5) to sputter polycrystalline metal targets with normally-incident 100 to 1000 eV Hg^+ ions, showed that sputtered-atom angular ejection distributions ranged from under-cosine at lower energies toward cosine at higher energies and noted that the preferred crystalline orientation of the target was important. Much later (1986), Matsuda et al.[126] reported, based on normally-incident Ar^+ ion-beam experiments, that the angular distribution of sputtered Fe atoms varies from cosine at 600 eV to slightly over-cosine at 1000 eV and increasingly over-cosine at 2 and 3 keV.

4.2. Mechanism of sputtering

Crookes, in his classic 1891 paper on metal sputtering rates,[110] compared the sputtering process to evaporation and described differences in the two processes as arising from coupling electrical vs. thermal energy to the source material. That is, sputtering in his terminology was "electrical evaporation." Charles Townes (1915-2015, Physics Nobel Laureate in 1964, sharing the prize for the invention of the maser) published a theory in 1944 based on ion-irradiation-induced local evaporation in which the sputtering yield (the average number of atoms ejected per incident ion, a measure of the process efficiency) depended on the ion-energy, but not on the ratio of the ion-to-target atom masses.[127] Even though contradictory evidence had been accumulating since the late 1890s, the concept that sputter ejection of target atoms occurs by local "hot spot" evaporation persisted well into the 1900s. In fact, review articles published as late as the 1960s,[128] followed Adolph Güntherschulze (1878-1967) in attempting to popularize the term "impact evaporation."[129]

A half century after Grove's initial sputter-deposition and ion-etching experiments, Eugen Goldstein (1850-1930), a German physicist, in 1902 provided additional evidence that sputtering is initiated by positive-ion bombardment. In his experiments, he used a perforated negatively-biased target and demonstrated sputter etching of a gold film on the wall of the discharge tube behind the target.[130,131] Goldstein called the positively-charged beams "canal rays." In 1908, Johannes Stark (1874-1951, Physics Nobel Laureate in 1919 for discovery of electric-field-induced splitting of atomic spectral lines) argued strongly in favor of sputter-ejection occurring by ion-impact-initiated collision cascades.[132,133] There was abundant evidence to support these claims, including the relative insensitivity of the sputtering rates to the thermal conductivity of the target and to target temperature.[134,135] Stark and Wendt,[136] in 1912, reported experimental evidence for sputter-rate variations with grain orientation in polycrystalline targets.

In 1931, Mayer,[137] and later Sporn[138] (1939), provided early indications that the ejection energies of sputtered atoms are of the order of several eV, much higher than typical evaporated-atom energies of a few tenths of an eV. The authors were investigating optical emission due to the decay of excited atoms sputter-ejected from thin, low-work-function alkaline-earth oxide and alkaline-metal layers on metal targets. Ejection velocities were estimated from measurements of the thickness of the luminous region together with known excited-state lifetimes. Güntherschulze[139] contributed supporting data in 1942 showing that sputtered atoms have much longer mean free paths (and hence higher kinetic energies) than would be expected for thermal atoms.

H. Fetz added further evidence in favor of sputtering via collision cascades in 1942 when he showed that sputtering yields (determined by dividing the measured target weight loss by the ion current to ground) increase with increasingly oblique angles of ion incidence due to more effective momentum transfer in the near-surface region of the target.[140] In 1954, Fred Wehner, often referred to as the father of modern sputtering, found that atoms tend to be sputter ejected in a specular (forward) direction when subjected to oblique low-energy (< 1 keV) ion bombardment.[141]

None of the above results are compatible with target "disintegration" by an evaporation process. Wehner[142] added yet another point in favor of momentum-transfer sputtering in 1955, while confirming the previous observations of Stark and Wendt.[136] He showed that during low-energy (150 eV Hg^+) sputtering from single-crystal targets under conditions (see Section 4.5) for which ion energies are monoenergetic and incident orthogonal to the target surface, sputtered atoms are ejected along close-packed crystalline directions.[142] Examples are shown in Figure 14 for sputtering of (111), (110), and (001) oriented face-centered-cubic (fcc) silver targets; the spots correspond to preferential sputtering along close-packed directions in all three cases. These results captured the imagination of those working in the sputtering field and several papers appeared confirming Wehner's observations for fcc metals[143,144] and showed similar effects for body-centered[145,146] and hexagonal[147] metals as well as diamond-[143,144] and zinc-blende-structure[148] semiconductors. The latter authors noted that preferred ejection directions are not observed for sublimation from single crystals.[148]

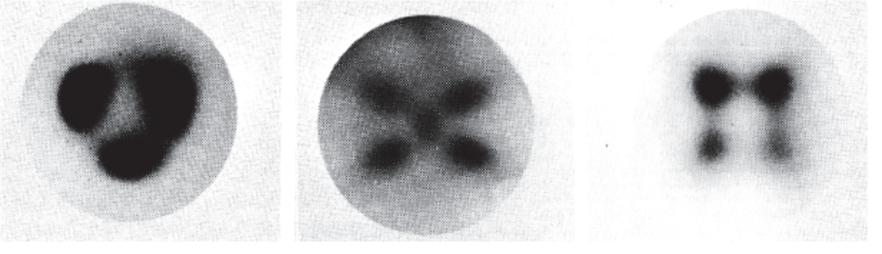


Figure 14. Typical deposited spot patterns obtained from low-energy sputtering of (111), (110), and (001) planes of fcc crystals. Courtesy of Fred Wehner. See also Ref. 142.

Historical footnote: the "spot patterns" first published by Wehner rapidly led to sputtering theories based upon long-range collision sequences along close-packed directions ("focusons"); one of the earliest of these was by Silsbee.[149] The focuson sputtering concept was highly popular and lasted for many years even though it was shown early, both by computer modeling[150] and analytical calculations based upon experimental results,[151] that the spot patterns can be explained simply by near-surface collision effects. This is especially the case for low-energy sputtering where penetration depths are only several atomic layers, but it is also true even for high-energy sputtering for which experimental results show that the average focuson range is only a few atomic distances (see discussion in ref. 151).

The existence of spot patterns as a function of target orientation indicates that the original lattice structure is at least partially intact during the violent sputtering events, further evidence against fully-developed cascade focusons for which the lattice would lose orientational information. Overall, it is clear that the probability of sputtering an atom from the surface layer (i.e., providing sufficient energy to overcome the surface binding energy) will be higher due to a central collision initiated by a close lower-lying neighbor atom. The width (broadening) of sputtered-atom spots is determined by a variety of factors including thermal vibrations and deviations from the ideal surface structure due to previous ion impacts.

Stuart and Wehner[152,153] also provided definitive evidence that the earlier indications from Mayer[137] and Sporn[138] were correct: atoms sputtered by low-energy ions are ejected with average energies that are of the order of ten times larger than for evaporation at the same rate. Wehner and colleagues used time-of-flight optical spectroscopy in low pressure Hg-pool-supported discharges (see Section 4.5) for which the sputtered-atom mean free path was larger than the distance from the target to the measurement sampling region. In these experiments, the target was sputtered for 1 μs pulses and the sputtered atoms excited by electrons in the discharge. As each group of sputtered atoms moves away from the target, it becomes spatially dispersed due to the distribution in ejection velocities. The dispersion is detected as a time distribution of characteristic photons emitted by the sputtered atoms as they pass through the observation volume, defined by a narrow slit, at a known distance from the target. The measured time distribution is then converted into ejection velocity and/or energy distributions. These time-of-flight results were verified by determining the Doppler shift for light emitted by sputtered atoms in the direction of travel.[153]

Figure 15 is a plot of the average energies E_{at} of atoms ejected orthogonal to 22 different polycrystalline metal targets due to bombardment by 1200 Kr^+ ions.[153] The results conclusively establish that sputtering is not a local evaporation process. Average ejection energies for elements in the middle of the periodic table are ~10 eV, with considerably higher E_{at} values for heavy elements such as Ta, Pt, W, Re, and U which have atomic masses between 180.95 and 238.03 amu. The heavy-target-atom effect is directly related to the relatively shallow penetration depths of the much lighter Kr (83.80 amu) ions. Thus, kinetic energy is deposited closer to the surface of heavy-atom targets. However, there is clearly more to the story, as discussed in Section 4.8, since the mass of Au (196.97 amu) is higher than the other heavy atoms, except for U, shown in Figure 15.

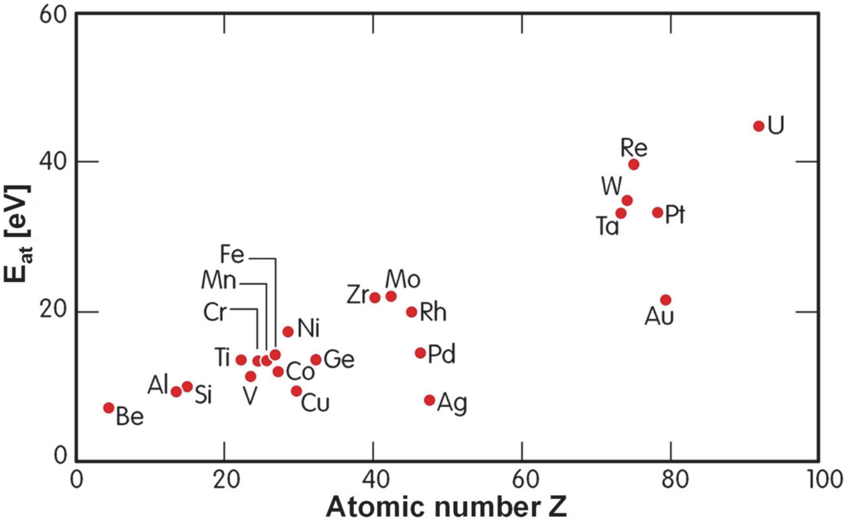


Figure 15. Average energies E_{at} of atoms sputter ejected orthogonal to the surface of elemental polycrystalline targets bombarded with normally-incident 1200 eV Kr^+ ions. Z is the atomic number (the number of protons per atom). Z_{Kr} = 36. Adapted from Ref. 153.

4.3. Early sputter-yield measurements

Kingdon and Langmuir[44] published, in 1923, argon, cesium, helium, hydrogen, neon, and mercury sputtering yields for thoriated tungsten wire targets (light-bulb filaments) as a function of V_T between 50 and ~400 V in a glass dc diode system. Although the sputtering conditions were poorly-defined, the authors discussed the process in terms of momentum transfer. In 1926, Eric Blechschmidt[154] carried out exhaustive studies of the sputtering yields of 19 metals by hydrogen, 13 metals by neon, and 15 metals by argon ion bombardment. Unfortunately, Blechschmidt's yield measurements suffered from severe target contamination due to poor vacuum.

In addition to contamination, another issue in early sputter-yield measurements is the unknown contribution to the total measured current of the ion (as well as fast-neutral-atom and photon) bombardment induced secondary-electron flux (see *Historical footnote* below) from the target. The secondary-electron yield from clean metals is typically ~0.1 for Ar^+ ion energies less than ~1 keV.[155,156] Thus, the *direct* effect on reported low-energy metal sputtering yields is not that large. However, the *indirect* effects turn out to be very important indeed. A secondary-electron yield of 0.1 means that each ejected electron must, on average, produce 10 new ions just to maintain the discharge. This, in turn, requires operating at relatively high pressures in dc diode systems, often near, or above, 100 mTorr, depending on target-to-substrate distance, in order to provide sufficient collisions since the extremely small size of electrons results in their having long mean free paths.

Historical footnote: the evolution of the understanding of ion-bombardment-induced secondary electron emission from a metal target is fascinating, involves many of the giants in the early days of gaseous electronics, and, due to its importance in the sputtering process, is briefly recounted in Section 4.4.

Sputtering at high pressures introduces, in turn, yet another uncertainty in the accurate determination of sputter yields. The mean free path of sputter-ejected target atoms, which have much higher collision cross-sections than electrons, is very short at these pressures. In fact in 1926, von Hippel[157] estimated that in typical sputtering systems at the time, up to 90% of the sputtered atoms are reflected back to the target at an Ar pressure of 100 mTorr. This not only dramatically reduces film deposition rates, it means that sputter yields based upon target weight-loss measurements are greatly underestimated.

An additional issue in determining reliable sputter yields S as a function of ion energy E_i at such high pressures is the inherent assumption that $E_i = eV_T$ in which V_T is the applied target voltage. However, at high pressures, the ion mean free path is much less than the distance over which the applied voltage V_T "falls" to the target (the "cathode fall" [late 1800s British English] or "target sheath" [modern terminology]). Thus, ions make many collisions while transiting the ion sheath and, hence, their average energy upon striking the target is much less than eV_T. This further reduces the sputtering yield (and the resultant film deposition rate). von Hippel[157] used a retarding field to measure the energy-dependent ion-current distribution arriving at a probe positioned a short distance behind a small hole in a metal cathode immersed in a 100 mTorr, 15 mA, $V_T = 1000$ V, Ar discharge. He reported that the average ion energy E_i was ~250 eV, with an energy spread across the full cathode-fall potential. Most of the ions arrived at the cathode with kinetic energies near zero. That the average value of E_i is much less than eV_T was confirmed years later by Davis and Vanderslice[158] who showed that the primary energy-loss mechanism for ions in high-pressure dc glow discharges is due to quantum-mechanical charge-exchange (rather than hard-sphere) collisions.[159] It should be noted that the decrease in sputtering yield as a function of V_T, compared to expected values if E_i were equal to eV_T, is partially mitigated due to sputtering by fast neutral atoms, not accounted for in measured ion currents, which have undergone charge-exchange reactions near the target.

4.4. Ion-bombardment-induced secondary-electron emission

In 1858, Julius Plücker (1801-1868), a German Professor of Mathematics (and later Physics) at the University of Bonn, during his investigation of electrical conduction through glow discharges,[104] observed a green fluorescent region in the glass discharge tube across from the metal cathode (Figure 11(b)).[160] He also found that the position of the luminescence was deflected in a magnetic field (Figure 11(c)). Johan Willhelm Hittorf (1824-1914), who studied under Plücker, reported in 1869 that if a solid object was positioned between the cathode and the fluorescent region in the glass tube, it cast a shadow (see Figure 11b). From this, he inferred that the glowing spot was due to what he termed "glow rays" which emanated from the cathode and proceeded in straight lines to cause the luminescence.[105] Eugen Goldstein demonstrated in 1878 that the particles originated at the cathode and referred to them as "cathode rays"[161] with a charge opposite to that of the "canal rays" [positive-ions] he observed during sputtering of a perforated target (Section 4.2).[130,131] William Crookes[110] (Section 4.1) used the term "radiant matter" and established that the rays were not due to fast sputtered atoms.[162] He proposed [incorrectly] that radiant matter was due to negatively-charged ions ejected with high velocity from the cathode.[163]

The German-born British physicist Arthur Schuster (1851-1934), a friend and competitor of J. J. Thomson,[164,165] expanded upon Crookes' experiments by placing metal plates parallel to the cathode rays and applying a potential between the plates. He observed that the rays were deflected toward the positive plate, thus proving that they are negatively charged.[165,166] This was confirmed, using a different approach, by Jean Baptist Perrin (1870-1942), a French physical chemist, and discussed at a lecture before the Paris Academy of Sciences on December 30, 1895 (English translation, reference 167). In his experiments, he allowed a beam of cathode rays to enter a metal cylinder through a small hole and showed that the cylinder became negatively charged to a value which was determined using an electrometer. No negative charge was measured if the beam was deflected away from the cylinder hole by a magnetic field. Perrin, more famous for his studies of Brownian motion, received the 1926 Nobel Laureate in Physics "for his work on the discontinuous structure of matter, and especially for his discovery of sedimentation equilibrium [the atomic nature of matter]."

Arthur Schuster, in his 1884 Bakerian Lecture,[166] made an important contribution to cathode-ray research by noting that "a particle of charge e and mass m, moving at speed v at right angles to a magnetic field $\overline{B}$, would be deflected such that it moves in a circle of radius r = mv/Be, so that [the cathode-ray charge-to-mass ratio] e/m = v/Br." Unfortunately, the pressure in Schuster's cathode tube was too high, 300 mTorr, for accurate measurements of v. In his second Bakerian Lecture,[168] 1890, Schuster combined the above equation with conservation of energy $\frac{1}{2}mv^2 = eV_T$, in which V_T is the cathode fall voltage, to yield the expression $e/m = 2V_TB^2r^2$. He used this equation to provide an upper limit to e/m; upper limit since it assumes that no energy is lost by collisions between cathode-ray particles and gas atoms (i.e., low pressures, which was not the case). He also provided a lower limit by assuming that the lowest cathode-ray particle speed v would be the root-mean-square speed of gas atoms at room temperature and the discharge-tube pressure. Again, however, he had experimental issues: "I have met with very considerable difficulties in the attempt to carry out the measurements in a satisfactory manner, and have only hitherto succeeded in fixing somewhat wide limits [which differed by a factor of ~1000] between which the molecular charges must lie."[168] Moreover, Schuster's results were found to vary strongly with the gas composition[164] due to the dependence of the canal-ray collision frequency on the size of the gas atoms in high-pressure discharges.

Walter Kaufman (1871-1947), a German physicist, used the approach suggested by Schuster in his second Bakerian lecture,[168] to determine e/m in better vacuum.[169] The result, published in November,1897, was much closer to today's accepted value than that of J. J. Thomson.[107] In addition, Kaufmann's e/m value, contrary to Schuster's, was independent of the discharge gas. He later showed, in his best-known work, that the electron mass is velocity dependent (i.e., relativistic).[170] In January of 1897, Emil Weichert (1861-1928) published results in which he obtained an e/m value for cathode-ray particles that was 2000 to 4000 times larger than that of hydrogen ions.[171]

> ***Historical footnote***: 1897 was a watershed year for cathode-ray research with seminal papers by Schuster, Kaufmann, Weichert, and Thomson; the latter leading to the 1906 Nobel prize in Physics.

J. J. Thomson, generally acknowledged as the "discoverer" of the electron (which he called a "corpuscle"), first presented his cathode-ray results during a lecture at the Royal Institution in London on April, 30 1897.[107] Employing energy conservation to obtain e/m, he equated the kinetic energy of the cathode-ray particles to the heat generated by their collisions with the interior surface of a cylinder. Six months later, he reported more fully on the details of these experiments together with the results of a second experimental approach based on the deflection of cathode-ray particles in electric and magnetic fields.[172] When they were subjected to the electric field $\overline{E}$ alone, the result was $e/m = v^2\theta/EL$, in which θ is the deflection angle and L the distance traveled by the particles at speed v in the field. The particles were next subjected to only a magnetic field $\overline{B}$, which deflected them through an angle φ, with the result $e/m = v\varphi/BL$. The two equations were then solved for the two unknowns: v and the ratio e/m.

In the latter experiments, Thomson adjusted $\overline{B}$ such that $\theta = \varphi$, resulting in $v = E/B$ and $e/m = E\theta/B2L$, from which he obtained, as the result of seven independent measurements, the average value $e/m = 0.77 \times 10^{11}$ C/kg, more than a factor of two lower than today's accepted value, 1.759×10^{11} C/kg. Nevertheless, Thompson realized his result was still nearly 1000x larger than e/m for a hydrogen ion. Further, he reasoned that the absolute value of the corpuscle charge was the same as that of a hydrogen ion. Hence, the corpuscle must have a mass approximately 1000 times smaller. Thus, in his Royal Institution Lecture, Thomson announced the discovery of particles which are small compared to atomic dimensions. "We have in the cathode rays matter in a new state …. in which all matter – that is, matter derived from different sources such as hydrogen, oxygen, etc. – is of one and the same kind; this matter being the substance from which all the chemical elements are built up."[107,172,173] While the term electron was well known at the time (see *Historical footnote* below), Thomson nevertheless continued to use the term corpuscles "…. until about 1910, during a period when electron theory flourished and was considered the most important topic of physical theory."[173] It is also interesting to note that: "In J. J. Thomson's Nobel-Prize citation there is no explicit mention of his discovery of the electron, but Thomson's [Nobel] lecture[174] was devoted almost exclusively to that topic."[164]

Historical footnote: The first use of the word "electron" to describe cathode-ray corpuscles appears to have been in 1894 by George Stoney (an Anglo-Irish physicist, 1826-1911, well known for the Stoney equation[175] which is still used today to evaluate thin-film stress). Stoney wrote in an article entitled *Of the 'electron,' or atom of electricity*, "…. an estimate was made of the actual amount of this most remarkable fundamental unit of electricity, for which I have since ventured to suggest the name electron."[176] Earlier, Stoney had used the term "electrolion" in describing experiments in which he estimated the charge of the "elementary particle of electricity."[177]

The next breakthrough in understanding ion-induced emission of "corpuscles" or "cathode-rays" (i.e., secondary-electrons) came much later. In 1954, Homer Hagstrum (1915-1994), an American surface physicist at Bell Laboratories, published a paper on the theory of ion-bombardment-induced secondary-electron emission, based primarily on Auger neutralization and, to a lesser extent, Auger de-excitation processes.[178] Hagstrum described Auger neutralization as "a process in which the interaction of two conduction electrons causes one electron to neutralize the [incident] ion and the other to be excited into the continuum above the filled band. The observed [secondary-electron yield] γ is determined by the probability that the excited electrons escape from the metal." The theoretical predictions proved to be a good fit to his experimental results for 10-1000 eV rare-gas irradiation of polycrystalline tungsten in an ultra-high vacuum system with a base pressure of $\leq 10^{-10}$ Torr.[179] Secondary-electron ejection due to ion bombardment in this energy range, common in modern glow-discharge sputter deposition, is today called "potential" electron emission. At ion bombardment energies $E_i > 1000$ eV, (the "kinetic" electron emission range), γ increases with ion energy.[180-181]

With $E_i < 1000$ eV, γ is not a strong function of ion energy and is found experimentally to be approximately 0.11 for 500 eV Ar^+ irradiation of tungsten and 0.25, 0.06, and 0.02 for Ne^+, Kr^+, and Xe^+. Similar results were presented by Hagstrum for rare-gas bombardment of molybdenum (see Figure 16).[182] Variations in γ with choice of rare gas primarily arise from differences in first-ionization energies which range from 21.56 eV for Ne, to 15.75 for Ar, 14.00 for Kr, to 12.13 eV for Xe. The observed decrease in the secondary-electron yield of Ne^+ with $E_i \lesssim 200$ eV results from the decreasing probability of Auger de-excitation processes involving short-lived excited states (thus the requirement for higher ion velocities) of the incident ion.[178,183]

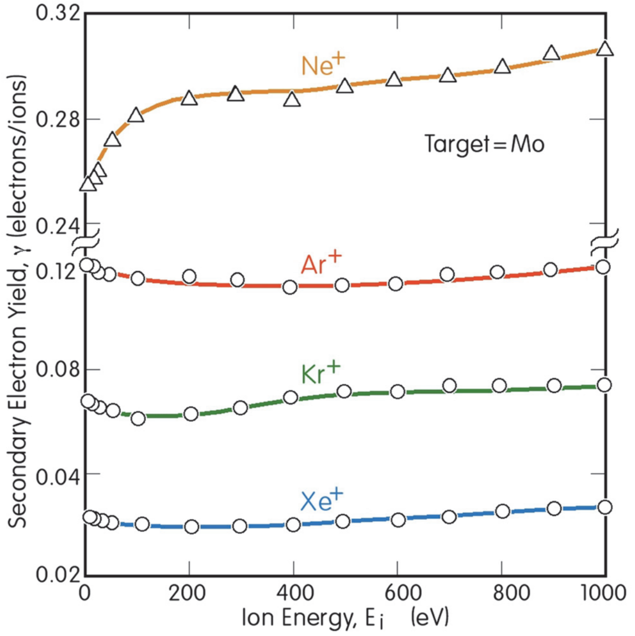


Figure 16. Total secondary-electron yield γ versus ion kinetic energy E_i for singly-charged rare-gas ions incident at a clean molybdenum target. Adapted from Ref. 182.

4.5. Thermionic and mercury-pool supported glow discharges for measuring sputtering yields

Several of the issues described in Section 4.3 which limit the accuracy of early sputter yield $S(E_i)$ measurements were overcome by A. Güntherschulze and K. Meyer in 1930, using their newly-developed triode glow-discharge sputtering system[184] in which large electron currents were obtained by thermionic emission from a separate filament electrode to overcome the low secondary-electron yield of the target. This provided a sufficiently dense plasma to allow sustained discharge operation at pressures between 1 and 10 mTorr. Thus, the mean-free path of ions and sputtered atoms was of the order of the discharge-tube dimensions (ballistic transport) with minimal scattering and back-diffusion to the target. Güntherschulze and Meyer immersed planar 5-cm-diameter Ag and Cu sputtering targets, during separate experiments, in dense plasmas as a third independent electrode. The ion current density at the target was of the order of 1 mA/cm^2 and the ion sheath thickness at V_T = 1000 volts was ~5 mm such that the energy of ions (Ar^+ or Ne^+) incident at the target was ~eV_T with a nearly monoenergetic distribution. The sputtering gas was continually flowing and spectroscopically monitored for purity.

Güntherschulze and Meyer sputter etched the target before initiating sputter-yield measurements and carried out separate experiments to measure the secondary-electron currents in order to determine the actual ion current at the target. Figure 17 shows $S(E_i)$ for Cu sputtered in Ar with and without accounting for the secondary-electron yield γ.[184] The corrected $S(E_i)$ results are in reasonable agreement with modern sputtering yield data.[185] The Cu target was uncooled and reported to reach a temperature of ~530 °C during long sputtering runs. In a later publication,[186] the same authors measured the sputtering rates of 16 different metals by Hg^+ ions at a gas pressure of 5 mTorr, controlled by means of the temperature of an oil bath in which the glow discharge tube, containing a liquid-mercury pool, was partially submerged.

In 1956, Wehner[187] improved the design of an earlier (1940) version of a mercury-pool-supported glow discharge tube developed by Fetz.[188] An illustration of Wehner's initial mercury-pumped (see Section 3.1) glass sputtering system is shown in Figure 18.[187] While the base pressure is not specified, from comments in later papers, it is ~10^{-6} Torr.[189,190] The target in this system is immersed as a large Langmuir probe[187,191] in a low-pressure, high-density plasma created by a vacuum-arc discharge between the liquid-pool cathode and an anode. The Hg pool is in the bottom region of the tube and the vacuum-arc is ignited by a SiC rod inserted into the liquid pool.[189] The cathode spot is anchored to a molybdenum strip which is

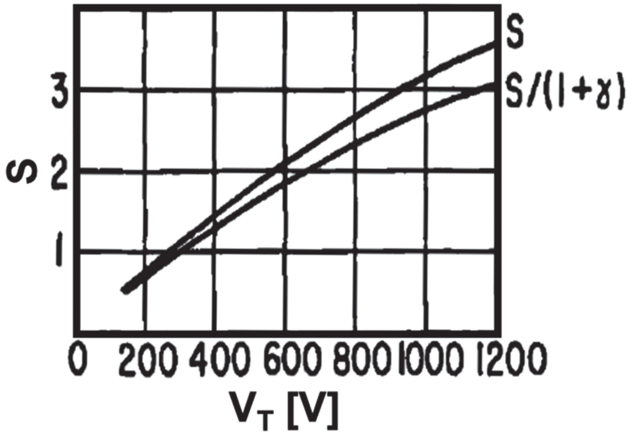


Figure 17. (Upper curve) Copper sputter yield S (atoms per ion), corrected for the secondary electron yield γ, vs. the cathode potential V_T in an argon discharge at a pressure of 10 mTorr. The lower curve is uncorrected for secondary-electron emission. Adapted from Ref. 184.

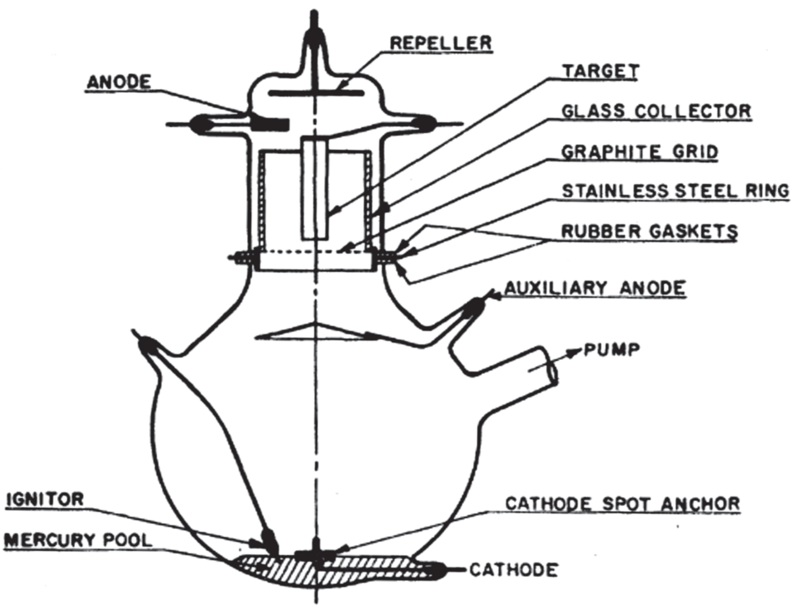


Figure 18. Mercury-pool discharge tube for high-current, low-pressure sputtering of the target (in the upper part of the sytem) with low-energy Hg^+ ions. Ref. 187.

partly submersed. An auxiliary (secondary) anode is required to maintain the cathode spot firmly anchored. The mercury pool is heated by an external water bath to 17 °C to provide a Hg gas pressure of ~1 mTorr.

The system was constructed in two parts; a stainless-steel ring, sealed with rubber O-rings, joined the upper tube, containing the primary anode and the target, to the lower rounded section. The plasma density was increased by inserting a fine mesh grid in the ring to separate the upper anode space from the lower liquid-pool cathode space. Fetz[188] had previously shown that an electrical double layer is formed in and near the grid holes, and that electrons reaching this layer from the cathode side are accelerated toward the anode space. These energetic "beam" electrons (controlled by the potential applied to the grid) are more efficient at ionization and greatly increase the plasma density in the anode space. An additional enhancement in plasma density was achieved by placing the anode outside the direct path of the beam electrons and adding a repeller electrode with a negative potential to reflect beam electrons back into the intense region of the plasma. [Note that the glass walls of the tube will also become negatively charged and perform the same function]. Both the repeller and the grid were typically set to ~ -50 V relative to the anode.

The electrical conductivity of the resulting highly-ionized plasma was sufficiently high that the plasma and anode potentials were essentially identical and the applied target potential V_T determined the kinetic energy, eV_T, of the bombarding ions. The ion-current density at the target was of order 10 mA/cm^2 and could be increased to 100 mA/cm^2 by the addition of a magnetic field near the target.[190] Therefore, sputtering rates were high enough to greatly reduce the effect of contamination on sputter-yield measurements. The targets in these experiments were metal rods which were heated to temperatures > 300 °C via a connection to an externally temperature-controlled sealed Kovar[192] tube to prevent Hg accumulation prior to initiating sputtering experiments.[187,189] Heating was not necessary during sputtering since the targets reached much higher temperatures due to the large ion currents. For low melting-point targets, the Kovar tube was used to cool the target during sputtering.[187] In operation, the thickness of the ion sheath at the target with V_T = -200 V was considerably less than 1 mm. At a pressure of 1 mTorr, the mean-free path of ions incident at the target was much larger than the width of the ion sheath and sputtered-atom mean free paths were larger than the distance from the target to the substrate.[187,189]

Figure 19 shows Wehner's results for the Hg^+ sputtering yield $S/(1 + \gamma)$, uncorrected for the secondary electron yield γ, of polycrystalline platinum vs. mercury-ion energy E_i.[187] These are some of the earliest reliable *low*-energy yield measurements and were carried out at target temperatures set to both 300 and 650 °C to show that Hg contamination of the target is negligible since the two data sets were within the range of experimental uncertainly. Over the ion-energy range investigated, 100 to 410 eV, $S/(1 + \gamma)$ increases essentially linearly with E_i.

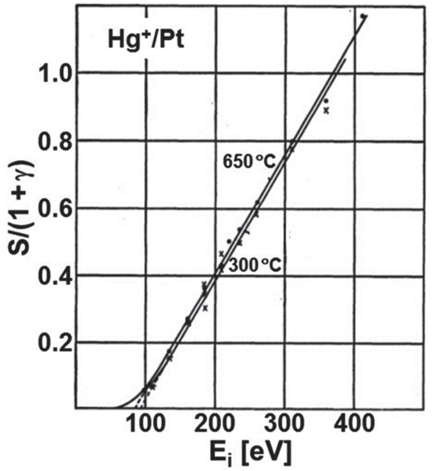


Figure 19. Low-energy mercury-ion sputtering yield $S/(1+\gamma)$, uncorrected for the secondary-electron yield γ, of polycrystalline platinum vs. incident ion energy E_i. The upper curve corresponds to a Pt target temperature of 650 °C, the lower to 300 °C. The experiments were carried out using the mercury-pool discharge tube shown in Figure 18. Adapted from Ref. 187.

The mercury-pool-supported discharge was also used by Wehner to measure low-energy rare-gas sputtering yields in an effort to determine "threshold" energies for different gas/metal combinations as discussed in Section 4.7.

In order to measure rare-gas sputtering yields over an energy range more typical of that used in film deposition, Laegreid and Wehner,[193] as discussed in their previous paper,[195] returned to the heated themionic-cathode-supported triode discharge tube pioneered by Güntherschulze and Meyer.[184] The upper part of the demountable tube, permanently connected to the vacuum pumping system, contains the themionic cathode, while the target and anode are in the lower section, with the two parts sealed by a copper gasket. The system base pressure, 10^{-7} Torr, was measured with an ionization gauge. After baking the system for 14 hours at 340 °C, the oxide cathode[195] temperature was slowly increased, "never allowing the pressure to exceed $7x10^{-5}$ mm Hg [$7x10^{-5}$ Torr]." After filing the liquid nitrogen trap, the background pressure with the cathode at temperature, but prior to introduction of the noble gas, did not exceed $8x10^{-7}$ Torr.

The targets were uncooled, but the authors reported that "The influence of the target temperature [on measured sputtering yields] seems to be negligible [tested by varying ion current densities from 1 to 10 mA/cm^2]." During sputtering, target temperatures ranged from 300 to 500 °C. Sputtering yields $S(E_i)$, uncorrected for secondary electron emission, were determined by the combination of target-current and weight-loss measurements. Example results are plotted in Figure 20 for 15 different polycrystalline elements (of 28 studied) sputtered in argon with ion energies from 50 to 600 eV.[196] The elements shown in Figure 20 represent a significant fraction of the periodic table including p-block (Al), noble (Ag), base (Pb), ferromagnetic (Co and Ni), transition (Ti, V, Cu, and Y), lanthanide-series (Er), and actinide-series (U) metals; semiconductors (Si and Ge); and carbon.

It is interesting to note that while ion guns were certainly available, and improving in quality,[197,198] during the prolific era of the 1950s and early 1960s when supported glow discharges were being used to probe, and make enormous progress in understanding, the physics of sputtering over the energy range of interest for thin-film growth, they had little effect on the field. Although the early ion guns were of poor vacuum design with highly divergent beams,[124] these issues were resolved over time.[197-199] However, the current densities provided by ion guns based upon simple dc-diode glow discharges remained low and since the primary approach to determining absolute sputtering yields required measuring target weight loss, ion guns were not very practical for this purpose. The accuracy for measuring extremely small amounts of material loss due to low-energy ion bombardment was poor.

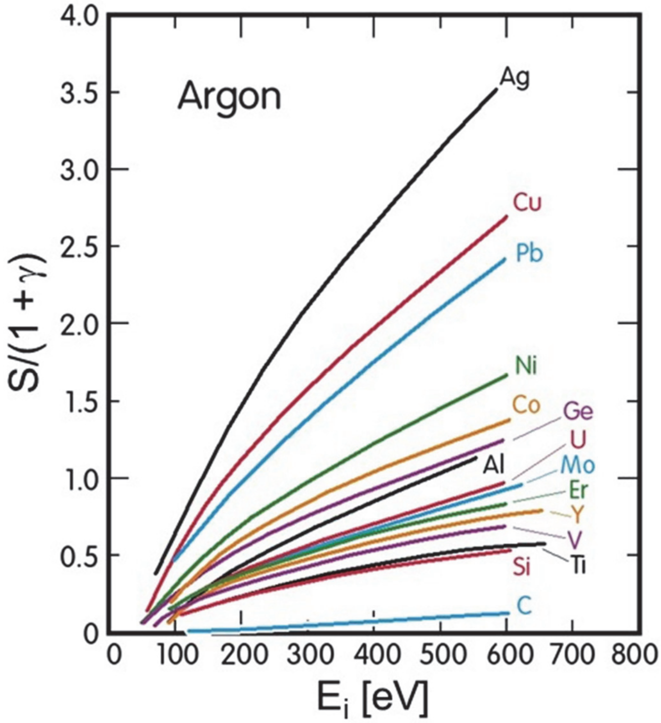


Figure 20. Argon-ion sputtering yield $S/(1+\gamma)$, normalized to the secondary electron yield γ, of 15 polycrystalline elements vs. incident Ar^+ ion energy E_i. Adapted from Ref. 196, data primarily from Ref. 193.

Conversely, at high ion energies, typically in the range 10 to 100 keV, where sputtering yields are large, high-purity ion beams obtained from accelerators providing high current densities were being used in the early 1960s to measure $S(E_i)$.[200,201] O. Almén and G. Bruce reported, for example, a sputtering yield of 48 atoms/ion for 45 keV Hg^+ bombardment of polycrystalline silver.[201] With energies in the low keV range, Figure 21 shows interesting results for a silver target sputtered by extremely light (hydrogen and deuterium) normally-incident 2 to 12 keV ion beams, focused with an einzel lens, from an isotope separator.[202] The poor momentum transfer and large penetration depths due to the low mass of H^+ (1.01 amu) and D^+ (2.01 amu) ions compared to that of the Ag target (107.87 amu) results in $S(E_i)$ maxima at energies of a few keV. For silver sputtered with Ar^+ (39.95 amu), the $S(E_i)$ maximum occurs at ~40 keV.[203]

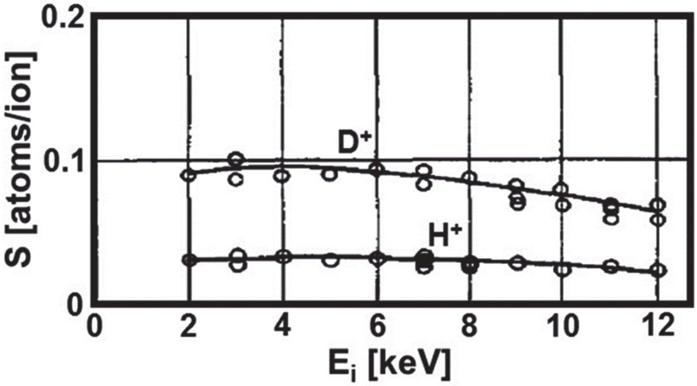


Figure 21. Light-ion (hydrogen H^+ and deuterium D^+) sputtering yields for silver, corrected for secondary electron emission, plotted vs. incident ion energy E_i. Adapted from Ref. 202.

4.6. Magnetically-supported dc glow discharge sputtering leading to closed-field magnetrons

An early investigation into the use of magnetic fields for supporting dc glow discharges was reported by Albert Hull in 1921 during a lecture delivered at the annual meeting of the American Institute of Electrical Engineers held in New York in May, 1921 (published in September).[204] Hull described a new vacuum-tube device, the magnetron, a member of the kenotron family of rectifiers and switches, that functions as an electronic valve controlled by a magnetic field. "If a constant voltage is impressed between cathode and anode, the current that flows through the tube is not affected by a magnetic field weaker than a certain critical value, but falls to zero if the field is increased beyond this value."

> ***Historical footnote***: Reference 204 contains the first use of the term "magnetron" in the literature. Today (2016), in the thin-film community, magnetron refers to the most common type of glow-discharge sputter-deposition device, available in a wide range of configurations, which provides high deposition rates due to having a closed crossed electric and magnetic field ($\bar{E}x\bar{B}$) tunnel at the target as explained in the text below (see Figure 26).

Hull discussed two types of magnetron switches, whose geometries are illustrated schematically in Figure 22, both having coaxial cylindrical symmetry.[204] In the first one, Figure 22(a) (referred to as the cylindrical-post configuration), the cathode is a thermionic filament surrounded by a cylindrical anode, while the second one (Figure 22(b), an inverted hollow-cathode magnetron), consists of a rod-shaped anode surrounded by a helical thermionic filament. With both types of magnetrons, the radial electric field is orthogonal to an axial magnetic field, parallel to the cathode surface, provided by a solenoid coil. In a previous publication (May of the same year), Hull[205] presented analytical models describing the motion of thermionically-emitted electrons in the two magnetron geometries (although he did not use the term magnetron in the earlier paper). He also described results for the case of a parallel-plate diode with a magnetic field parallel to the electric field.

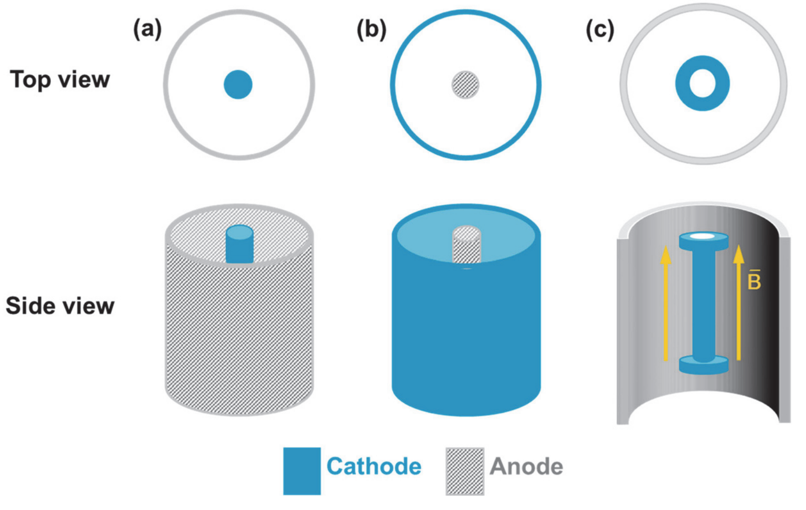


Figure 22. Illustrations of the essential features of (a) a cylindrical magnetron in which the cathode is surrounded by a coaxial anode and (b) an inverted hollow-cathode magnetron consisting of a central anode surrounded by a coaxial cathode, first described in Ref. 204. (c) A tubular, water-cooled cathode with end plates. Ref. 208. In all three cases, there is an axial magnetic field (see panel (c)) between the cathode and anode; the electric field is radial

In the mid-1930s, Frans Penning (1894-1953) reported the first use of the cylindrical magnetron geometry, in which Hull's filament cathode[204] was replaced by a rod, for sputter deposition.[206-208] In 1936, Penning filed a US patent (issued in February, 1939)[206] describing a cylindrical-post magnetron (the word magnetron was not mentioned) as a competing approach to the supported glow discharge (Section 4.5) for developing high ion currents at reduced operating gas pressures. Penning employed an axially-symmetric magnetic field, provided by a Helmholtz coil, between a rod-shaped target and a concentric anode tube (Figure 22(a)) in order to enhance the ion current by increasing the electron residence time in the discharge. More importantly, he produced (perhaps unknowingly) the conditions for magnetron sputtering, a closed electric- and magnetic-field trap which increases the plasma density at the target (cathode). The use of end caps on the post cathode (Figure 22(c) to form a more complete plasma trap, was added a few years later.[208] There is no indication that Penning was aware of Hull's previous results.

Penning realized that the magnetic field strength should not be so large as to affect the ions in the discharge, only the electrons. With the crossed electric and magnetic fields, "…. the electrons are prevented from reaching the anode directly along the electric lines of force so that they traverse a materially longer path than in the absence of a magnetic field." That is, "…. the electrons describe, however, more or less helical paths around the magnetic lines of force …." With a metal target operated at 1000 V in Ar, a magnetic field of 300 G "…. brought about a 300-fold amplification of the current." Penning notes that "This leads to strong disintegration of the target and the cathode particles disintegrated readily pass though the gas at a reduced pressure." The patent thus describes many of the key operational features of modern magnetron sputtering targets. It also proposes the use of the dc magnetron for what today is termed a sputter-ion vacuum pump[209] in which sputter-deposited metal atoms "getter" (react with) reactive gases to reduce the background pressure in a vacuum system.

Penning's 1939 patent goes on to describe a parallel-plate geometry in which two planar targets (cathodes) face each other inside a cylindrical anode which is surrounded by a Helmholtz coil. In this case, the magnetic field was primarily parallel to the electric field, thus giving rise to magnetically-enhanced, but not magnetron, sputtering (no closed-field trap at the target). However, Penning presaged, by almost 60 years, invention of the "facing-target" sputtering configuration which today is used to minimize bombardment of growing films by negative oxygen ions during reactive sputter-deposition of oxide films on substrates arranged radially about the axis connecting the targets.[210]

Penning and Moubis,[208] in 1940, described a system, with a base pressure of ~1x10^5 Torr, consisting of a 10-cm-diameter glass discharge tube with a tubular water-cooled cathode (2 cm in diameter by 25 cm long) which had two end plates for plasma confinement (thus, the shape of a wooden spool for storing thread or wire, Figure 22(c)), and a coaxial anode. The magnetic field was parallel to the cathode cylinder. They showed that with this arrangement, the plasma is strongly confined near the target surface yielding increased ionization and resulting in a narrow, ~1 mm thick, essentially collisionless target sheath. Sputter-deposition rates were estimated by measuring the weight increase of small mica substrates. Several mica discs could be successively exposed, using a magnetic transfer rod, without opening the tube.

The current/voltage characteristics during sputtering of a copper target in a 14 mTorr flowing Ar gas discharge are plotted in Figure 23 as a function of the magnetic field strength $\bar{B}$.[208] Penning and Moubis reported that with V_T = 500 volts, a discharge current of 3 A, $\bar{B}$ = 350 G, and a target-to-substrate distance of ~4.5 cm, the film deposition rate was ~0.3 µm/min, a very high rate for the time. In discussing their experimental results for aluminum targets, they noted that when the target is oxidized, a lower discharge voltage is obtained for the same discharge current which they attributed [correctly] to a higher secondary electron yield for aluminum oxide than for aluminum.[211,212] There are even earlier indications in the literature of the same effect occurring during the sputtering of oxidized Mg.[213]

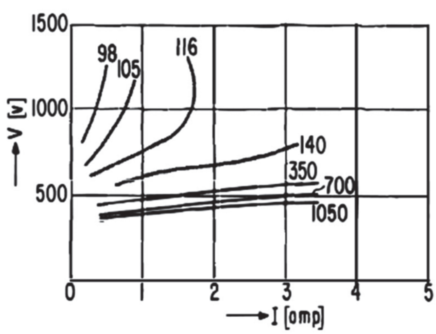


Figure 23. The current/voltage (I vs. V) characteristics, during sputtering of a copper target in a 14 mTorr flowing Ar gas discharge, plotted as a function of the magnetic field strength in units of gauss. The target geometry was a water-cooled cylindrical tube with end plates as in Figure 21(c). Ref. 208.

W. D. Gill and Eric Kay, in 1965, described the construction of an inverted (hollow-cathode) magnetron sputter deposition system[214] and its operational characteristics as a function of magnetic-field strength. Approximately a decade later, Allan Penfield and John Thornton at Telic Corporation developed large commercial cylindrical-post (> 1 m in length) and inverted magnetron sputtering systems with closed-field $\bar{E}x\bar{B}$ traps for thin-film deposition at the industrial scale.[215-217] An early application for cylindrical-post magnetron sputtering was the deposition of chromium layers on glass photomask plates used for metallization lithography during silicon device fabrication.

In the parallel-plate dc diode geometry (Figure 24),, the use of a transverse magnetic field, orthogonal to the axial electric field, increases the ionization efficiency (magnetically-enhanced sputtering as discussed above. Unfortunately, it does not form a closed-field trap and simply sweeps the glow discharge to one side.[217] In 1939, Rokhlin[218] suggested the use of a quadrupole magnetic field which is radially symmetric, but with large transverse components, in a parallel-plate configuration. Eric Kay,[219] in 1963, investigated the use of such a system for thin-film deposition. He reported an increase in the ion current density at an aluminum target by more than an order of magnitude. This, in turn, resulted in a corresponding enhancement in the film deposition rate. James Mullaly, in 1969, also reported high deposition rates using a magnetron supported with a quadrupole magnetic field.[220] In his case, however, the target was hemispherical. The crossed fields, in both Kay's and Mullaly's experiments, formed high-sputtering-rate rings at the target surface yielding circular racetrack-shaped erosion profiles. The hemispherical target geometry of Mullaly resulted in a more uniform radial deposition rate at the substrate.

The earliest use of a closed $\bar{E}x\bar{B}$ field in a parallel-plate sputtering geometry with permanent magnets was by Wolfgang Knauer of Hughes Research, Malibu, CA, in a patent filed on September 10, 1962 and issued November 9, 1965.[221] Most of the patent discusses a closed-field configuration with a cylindrical magnetron (a "Penning cell") for use as a sputter-ion vacuum pump "which results in the removal of reactive gas due to adsorption in freshly-deposited reactive sputter material and also by ion burial." However, Knauer also describes a circular ring-shaped planar magnetron with an annular closed-field region of high sputtering rate at the

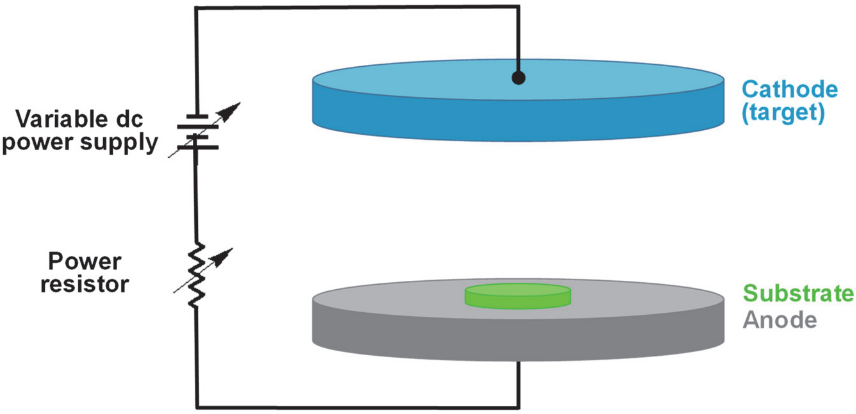


Figure 24. Parallel-plate diode sputtering geometry with circular target and anode.

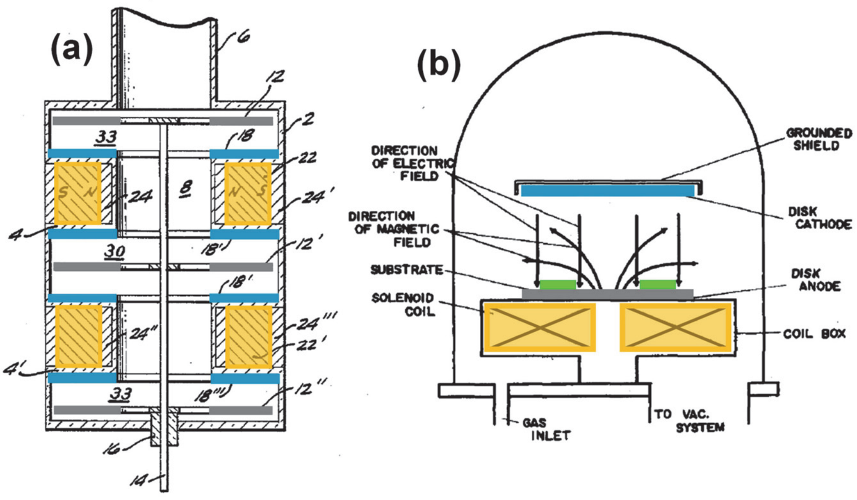


Figure 25. (a) A set of dc diode parallel-plate magnetrons. The annular ring-shaped targets (blue, 18) are facing annular anodes (gray, 12). Permanent magnets (orange, 22) are mounted behind the targets. Adapted from Ref. 221. (b) Parallel-plate diode magnetron sputtering system with a circular target (blue) and substrate platen (grey) supporting disc-shaped substrates (green). A solenoid magnet (orange) is placed behind the substrate table. Adapted from Ref. 222. Color added by present author.

target surface. (Figure 25(a)). The magnets in Knauer's design were placed behind the targets as in modern magnetron systems.

In 1969, Wasa and Hyakawa described a Penning-type cylindrical magnetron as well as a dc diode parallel-plate closed-field magnetron (Figure 25(b)).[222] The electrodes are circular disks facing each other in a sputter-down configuration. The solenoidal magnet was placed behind the substrate platen. However, no details were provided regarding operation of the planar magnetron, all reported discharge characterization and film properties were obtained using the cylindrical magnetron.

The modern planar magnetron is generally traced to the patent of John Chapin, assigned to Airco, filed January 31, 1974, issued August 28, 1979,[223] and discussed by Chapin in a 1974 industrial magazine article.[224] The device described in the patent was a water-cooled rectangular magnetron with a closed field resulting in a continuous rectangular erosion profile in the target surface. However, Chapin points out: "Because of the planar cathode, the erosion region may be made in any desired configuration, depending upon the shape of the substrate …." The rectangular configuration was designed for commercial production in which substrates to be coated are continuously passed over the target (in a sputter-up configuration). "Thus, a plate having a width less than the length of the long parallel segments …. will be provided with a highly uniform coating across its entire width. Accordingly, the apparatus of the invention not only achieves a high deposition rate, but may be utilized in high-production operations such as in connection with moving strip material."

Figure 26 is an illustration of Chapin's planar magnetron which he describes as follows:[223] "….the magnetic field in the region just above the cathode surface or sputtering surface 12 forms a closed loop, indicated by the [red shaded region defined by dashed lines] 65 in [Figures 26(a) and (b)]. In this way, the electrons are confined in a sort of endless magnetic bottle adjacent the cathode and apart from the anode, further enhancing the rate of sputtering from the erosion region 15. In addition to trapping electrons, the magnetic bottle thus formed narrows the cathode dark space so that the space-charge-limited ion current is large. Because the cathode-anode circuit is entirely insulated from ground except through the ionized gas, there is little charge transfer to the substrate upon which the material is being deposited except for the small amount initially required to balance the cathode and anode at their natural potential, and except for a small amount due to the substrate contacting regions of the plasma at slightly different potentials."

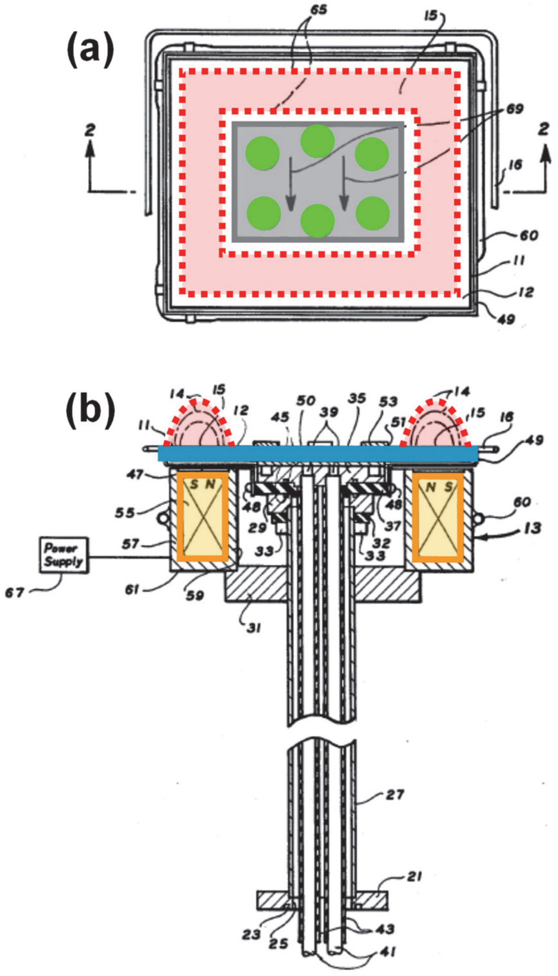


Figure 26. (a) Top view and (b) side view of a dc-diode parallel-plate magnetron. The red shaded region 15, defined by dashed lines 65, is the closed E?xB? field "tunnel" defining the high-sputtering-rate target erosion area. The annular magnet (orange, 13) is placed behind the rectangular target (blue, 11). The substrates (green) are mounted on a substrate platen (grey) which is translated over the target in the direction indicated by the black arrows (69). Adapted from Ref. 223. Color added by present author.

The first review article on modern planar magnetrons was published by Robert Waits in 1978.[225] Figure 27, from reference 225, provides Illustrations of both circular and rectangular planar magnetron geometries showing magnet placement and magnetic field lines. The discharge electrons experience a complex trajectory with a general drift velocity, due to the closed $\bar{E}x\bar{B}$ field, guiding them through the azimuthal ("racetrack") magnetic tunnel.

An obvious disadvantage of the parallel-plate closed-field magnetron in the configuration shown in Figure 26 is that the racetrack erosion profile consumes only a small fraction of target material giving rise to poor target utilization, originally ~25%, which was increased to ~33% by spreading out the magnetic field and opening up the racetrack. Since the permanent magnets of the parallel-plate magnetron are typically outside the vacuum system, further improvements in target utilization were obtained by rotating the magnet assembly, initially in a circular path and then in ever more complex hypotrochoidal paths allowing up to ~60% of large rectangular targets to be sputtered.[226] However, film thickness uniformity, less of a problem for substrates continuously moving under multiple larger targets, was still an issue for "single-wafer" processing using circular targets as employed, for example, by the semiconductor industry during silicon-based device fabrication. The solution was again to use a complex rotation of the magnet assembly, in this case corresponding to "heart-shaped" pathways which, in turn, are rotated circularly[227,228] behind targets approximately 50% larger than the size of the wafer being coated. This allows target utilization up to ~75% with film-thickness uniformity > 98%.[229] There are many other rotating magnet designs for different applications, examples are described in refs. 230 and 231.

The "s-gun" magnetron, patented in 1977 (filed in 1975) by Peter Clarke,[232] and illustrated in Figure 28, was an early competitor of the planar magnetron. The device is designed such that the target, anode, water cooling, and power feedthroughs are all mounted on a single flange. The target is annular with a triangular (or truncated triangular) cross-section, the anode is a circular plate, and the permanent magnet is toroidal in shape to encircle the two electrodes. The device geometry is designed to form, like the planar magnetron, a strong $\bar{E}x\bar{B}$ closed-field trap at the target yielding high sputtering rates. Since *both* the target *and* the anode are in the

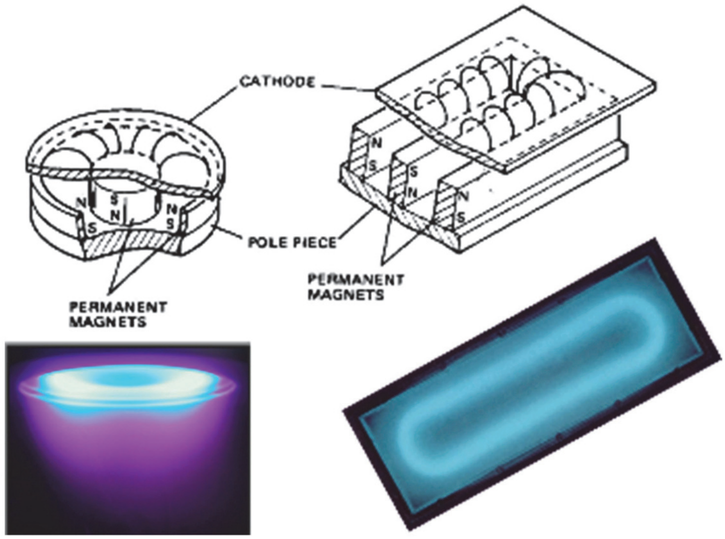


Figure 27. (Upper) Illustrations of circular and rectangular planar-magnetron geometries showing magnet placement and magnetic field lines. Ref. 225. (Lower) Photographs of circular (courtesy of Gencoa Ltd.) and rectangular (courtesy of Oerlikon Balzers) magnetrons during operation in gas discharges.

dense region of the plasma (and hence both require cooling), electron and ion bombardment of the substrate, typically placed across from the s-gun, is minimized. This can be an advantage for deposition on thermally-sensitive substrates such as polymers and low melting-point metals. However, a primary disadvantage stems from a combination of the complex crossing of the electric and magnetic field lines coupled with the shape of the target; together, these factors result in a complicated sputter-erosion profile leading, in turn, to a continuously changing film thickness distribution at the substrate. Thus, full planetary substrate motion (utilizing all three axes of rotation) is often required to obtain film-thickness uniformity. As a consequence, the s-gun today is primarily used in small-volume specialty applications. Nevertheless, s-gun targets are still commercially available. The earliest review article covering operation of s-guns was by Dave Fraser, Bell Laboratories, in 1978.[233]

For industrial applications, especially for high-deposition-rate coating of larger substrates such as architectural glass, for which standard "jumbo" panes in large office buildings are presently 6x3.2 m^2 (ref. 234), rotatable

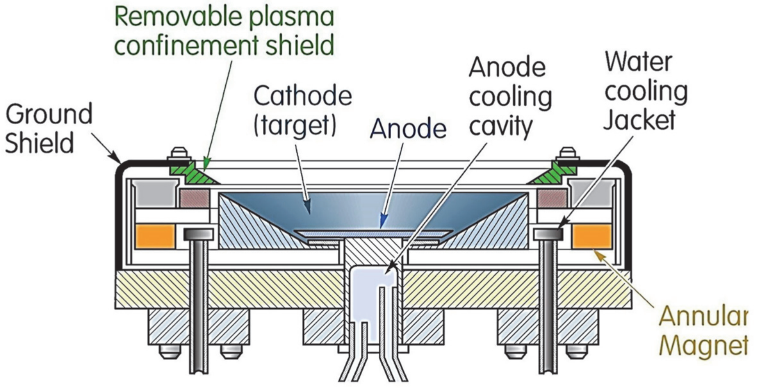


Figure 28. Cross-sectional view of the essential features of a typical s-gun magnetron. Adapted from reference 232.

tubular magnetrons, available in lengths > 4 m, are increasingly popular. The permanent magnet array in a rotatable magnetron is similar to that used in planar rectangular targets, with a corresponding rectangular racetrack, as shown schematically in Figure 29. The earliest rotatable magnetron was described in a series of patents in the early 1980s by H. E. McKelvey.[235-237] The original concept was to provide greater target-material utilization. After a local erosion profile was etched nearly through the target, the magnetron was rotated to sputter a fresh area.

The first *continuousl*y rotating magnetron was the subject of patent issued to the Von Ardenne Laboratory in Dresden, Germany in 1982,[238] described a year later by G. Robinson at a conference in the US,[239] and discussed in the open literature by Michael Wright and Terry Beardow of BVT Ltd., UK.[240] As summarized in a review article[241] by Roger De Gryse, who was directly involved in the development of the modern rotatable magnetron, and colleagues from Ghent University, the continuously rotating magnetron has several additional advantages, other than increased target utilization, over the flat rectangular magnetron. These include a larger material inventory compared to a planar magnetron of the same length; better target cooling efficiency, due to water flow though the rotating tube, which allows the use of higher powers to provide higher deposition rates; and reduced arcing during reactive sputtering (Section 4.10) since the length of the transition region between the high sputtering rate area and the poisoned target area is greatly reduced as illustrated schematically in Figure 29.

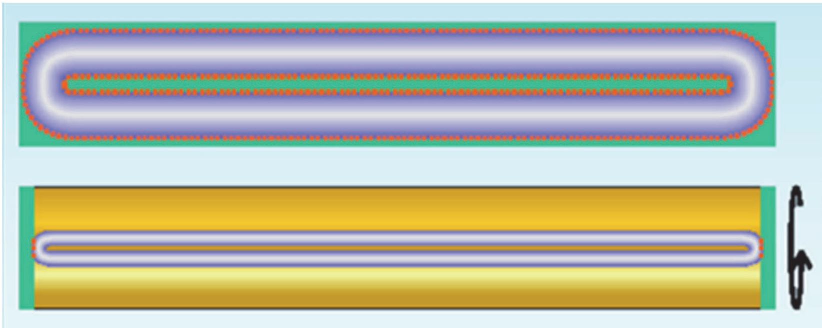


Figure 29. (Upper) Planar rectangular magnetron and (lower) rotatable tubular magnetron. The grey-blue regions are the sputter erosion racetracks, the green regions of the target are "poisoned," and the orange transition regions have the highest risk for arcing.[242] Figure courtesy of Wilmert. de Bosscher, Soleras Advanced Coatings.

Chapin correctly pointed out in his 1978 patent[223] that the substrate in a magnetron sputtering system is typically placed far from the intense region of the plasma which is trapped very close to the target by the Lorentz crossed-field force. While this configuration results in high deposition rates, the disadvantage is that it is difficult to provide low-energy, high-flux ion irradiation at the growing film surface in order to control film composition, preferred orientation, density, surface roughness, and physical properties.[243-244] An initial step toward resolving this issue was provided a decade later by Brian Window and Nick Savides, from CSIRO in Sydney, Australia, who "unbalanced" the magnetron magnetic circuit. In a standard circular magnetron, the permanent magnets are designed such that the magnetic field of the central pole is stronger than that of the outer annular magnet (see left panel of Figure 30). Windows and Savides exchanged the permanent magnets in order to make the outer annular magnet magnetron stronger than the central pole, thus opening a leak in the trapped target plasma and allowing low-energy ion bombardment of the substrate.[245]

A more useful approach leading to modern *tunable* unbalanced magnetrons was developed in 1992 by a University of Illinois group who added a pair of external Helmholtz coils with iron pole pieces to controllably unbalance the magnetic circuit through the addition of an adjustable axial magnetic field in the region between the target and the substrate (Figure 30). The external field ("external" to that of the magnetron permanent magnets) provides the ability to controllably shape the discharge near the substrate in order to independently control the energy and flux of ions incident at the growing film while having negligible effect on the target sputtering rate.[243] This approach to film synthesis is now employed in a wide variety of commercial applications with magnetron designs ranging from circular to rectangular to rotating cylindrical tubes.

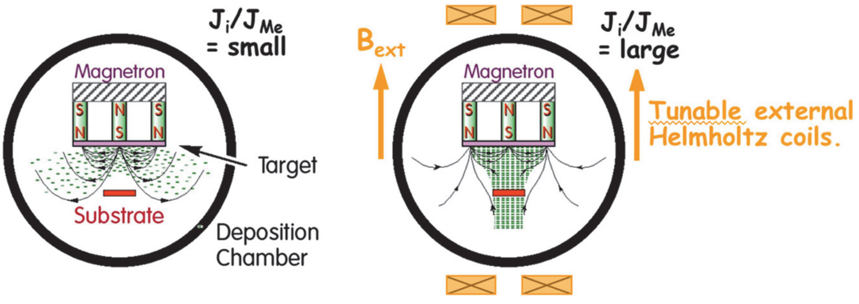


Figure 30. Schematic side view illustrations of (left panel) a standard circularr magnetron during operation and (right panel) a tunable magnetically-unbalanced magnetron. J_i/J_{Me} is the ion/metal flux ratio incident at the substrate and growing film. Ref. 243.

4.7. Threshold energies for sputtering

"Threshold" energies for sputtering have been discussed in the literature since the late 1800s. Wehner and Anderson[40] noted that until 1912, a value of 495 eV was thought to be valid for all gas/metal combinations.[245] Based upon results presented in Section 4.5 including Figures 17 and 19, this is far too large. Later reports (1928) described sputter damage of thoriated-tungsten filaments in gas rectifier tubes when ion bombardment energies exceed 20 to 30 eV.[247] In the limit, the "threshold" is the minimum incident ion energy, for a given ion/target combination, that provides sufficient energy transfer during a near-surface collision cascade to allow a target atom to overcome its surface binding energy and be sputter ejected.

From the discussion in Section 4.5, it is clear that measuring the threshold energy for a given ion/target pair is exceptionally challenging. Sputtering rates near threshold are far too low for standard direct weight-loss measurements, which do not extend much below ~0.1 atoms/ion. More sensitive techniques are required. Interest in threshold energies in the middle 20th century derived not just from their importance in understanding the physics of sputtering and testing theories, but also for practical reasons. The sputtering threshold was important since it is related to radiation damage thresholds in devices ranging from vacuum-tube electronics to sensitive detectors on orbiting satellites.

In a 1962 paper, Stuart and Wehner[190] summarized sputtering threshold measurements starting from the work of Kingdon and Langmuir[44] and Hull and Winter,[248] both in 1923, who used the change in thermionic emission from thoriated-tungsten wires as thorium was sputter removed; to experiments based upon the resistance change of sputtered metal wire targets (1935);[249] measurements of the surface ionization energy of low work-function alkali metals under ion bombardment (1954);[250] radioactive-tracer methods (1956);[251] and the use of an *in-situ* microbalance (1957).[252] The results yielded thresholds for different ion/target combinations ranging from 45-55 eV for Ne^+, Ar^+, and Hg^+ irradiation of Th-W wires[44] to ~8 eV at a sputtering yield of ~10^{-4} for Hg^+ and Ar^+ bombardment of Ni[251] However, Stuart and Wehner[190] provided arguments – low ion current densities compared to background impurity chemisorption rates, thin wire targets leading to oblique ion incidence, plasma-tube design resulting in bombardment by doubly-charged ions, etc. – why all of these results had significant uncertainties.

In 1954, Wehner and Medicus,[253] using a glow-discharge tube supported via thermionic emission from a tungsten filament (Section 4.5), reported yet another method for measuring $S(E_i)$, although in arbitrary units, at very low ion energies E_i; in this case for Xe^+ irradiation of platinum with E_i between 30 and 200 eV. A Langmuir probe was inserted in the discharge and the authors took advantage of the fact that small changes in the probe work function can be accurately determined by displacements of the probe current/voltage characteristic along the voltage axis. The shift was calibrated vs. the relative amount of platinum sputter deposited onto the probe. McKeown[254,255] in 1961 used the frequency change of a Au-coated piezoelectric quartz oscillator to obtain thresholds of 20-30 eV for Ne, Ar, and Xe ion bombardment with sputtering yields of ~10^3 atoms/ion.

Stuart and Wehner[190,256] recorded threshold sputtering energies of ~15 to 35 eV for several metals, plus germanium, bombarded by rare-gas and mercury ions using a mercury-pool-supported discharge similar to that shown in Figure 18, but with a planar target arrangement and a ring magnet to focus the discharge to the center of the target, thereby avoiding edge effects. The system, which has a liquid-nitrogen cold trap in the upper (target/anode) section to freeze the mercury out of the primary discharge and allow sputtering with a noble gas, provides ion current densities at the target of up to 100 mA/cm^2 with background contamination < 10^{-6} Torr. Mass spectroscopy measurements showed that doubly-charged ions contributed less than 1% of the ion current.

Relative sputtering yields $S(E_i)$ were determined at low bombardment energies using optical-emission spectroscopy calibrated, at higher ion energies, with absolute weight-loss measurements.[190,256] The spectroscopic yield measurement geometry is shown in Figure 31. Optical emission data were obtained close to the target in order to average the spectral line intensity over all ejection directions and thus minimize errors due to the variation in sputtered-atom ejection angles.

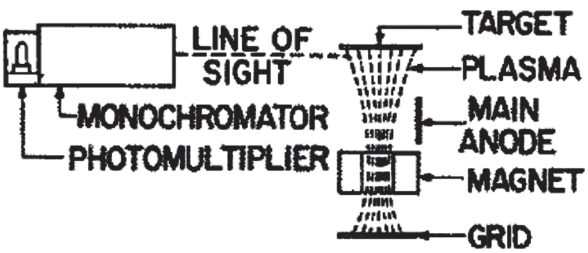


Figure 31. Schematic illustration of the experimental arrangement for the use of optical-emission spectroscopy to determine relative sputtering yields at low ion energies. Ref. 190.

Example $S(E_i)$ results, with E_i between ~20 and 300 eV, are presented in Figure 32.[256] From the yield curves in Figure 32(a), as well as additional results in ref. 190, the agreement between relative and absolute yield measurements at lower ion energies is better for rare-gas than for Hg ion sputtering, perhaps due to mercury contamination in the latter case. Figure 32(b) shows yield data for five metals, ranging from relatively light, Ti (47.87 amu), to very heavy, Th (232.04 amu), sputtered in Ar (39.99 amu). Based upon results for 23 metals, ranging in atomic mass from Be (9.01 amu) to U (238.04 amu), sputtered with four rare-gas ions (Ne^+, Ar^+, Kr^+, and Xe^{+}), Stuart and Wehner concluded that metal sputtering thresholds, 15 to 35 eV, can be crudely approximated as four times the target heat of vaporization.[190] This was consistent with reports for radiation-damage induced lattice-atom displacement thresholds in solids.[257]

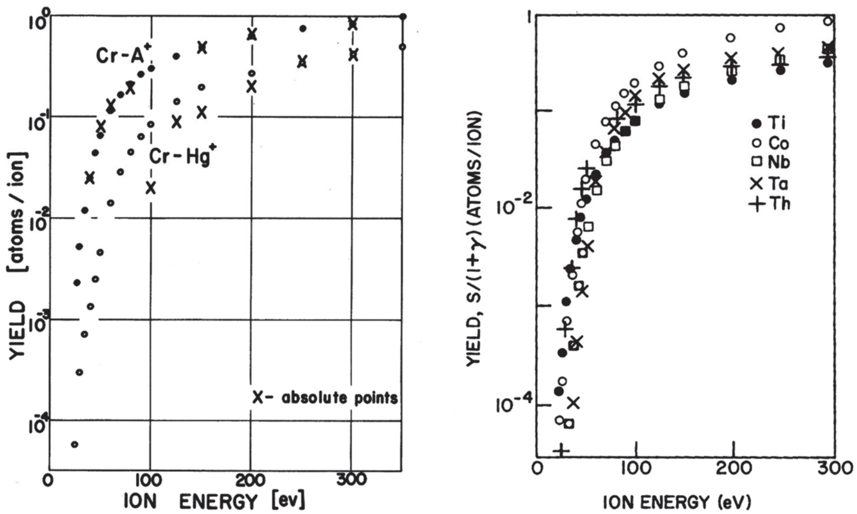


Figure 32. (a) Optical-spectroscopy sputter-yield measurements, together with absolute weight-loss measurements (X), plotted as a function of ion energy for argon and mercury ions bombarding a chromium target. Ref. 256. (b). Low-energy sputter yields for argon ion irradiation of five metals. Ref. 190. Results in both plots are uncorrected for secondary-electron emission.

4.8. More on mechanisms of sputtering: theory

Results from the earliest sputtering experiments, beginning in the mid-1800s, overwhelmingly support the momentum-transfer collision-cascade model of sputtering (see discussion in Section 4.2). This is clearly reflected in review articles published in the 1960s[258-260] as well as in theoretical developments.[261-265] During cascade sputtering, the sputtering yield is expected, as noted by Wehner and co-workers,[40,189] to be proportional to an energy-transfer factor $4m_i m_T/(m_i + m_T)^2$, in which m_i is the mass of the incident ion and m_T is the target-atom mass. This simple function has a maximum when $m_i = m_T$ The sputtering yield should also vary inversely with the target-atom surface binding energy, often approximated for elemental targets as the heat of vaporization U_{vap}.[266] Finally, for incident ion energies used in glow-discharge sputter deposition, typically a few hundred to < 1000 eV, S for a given metal is approximately proportional to E_i[185] (see also Fig.20).

Modern sputtering theory is based upon a linear-transport model published in 1969 by Peter Sigmund from Aarhus University, Denmark.[267,268] He assumed that sputter ejection of target atoms by recoiled energetic ions or target atoms results from atomic collision cascades in an infinite medium. Lamar and Compton[269] appear to be the first to suggest that sputtering can occur via momentum transfer as an ion penetrates the target and is reflected from a lower layer to collide with a surface atom. However, they neglected the important role of displaced target atoms in cascade processes. Sigmund described collision cascades using a generalized Boltzmann transport equation with, for ion energies used in sputter deposition, a Born-Mayer-type collision cross section. He found that while lattice-atom scattering events take place – depending on ion mass and energy, target-atom mass, and ion/target geometry – over a thickness which is a sizable fraction of the ion range, the majority of sputtered particles originate from a very thin (~5 Å) surface layer. For the simplest (although still quite complex) case of an amorphous target with a flat planar surface and normal ion incidence with E_i < 1 keV, the sputtering yield is given by

$$S(E)_i = \frac{3}{4\pi^2}\alpha\frac{4m_i m_i}{(m_i + m_i)^2}\frac{E_i}{U_{vap}}, \tag{1}$$

consistent with experimental observations discussed above. The term α depends only on the ion and target-atom masses:

$$\alpha = 0.1 + 0.155\left(\frac{m_t}{m_i}\right). \tag{2}$$

The fact that the above formulas provide a reasonable fit to sputter yield data from polycrystalline and single-crystal targets in the few hundred eV to 1 keV range is another indication that long-distance focusons (see Section 4.2) do not play a significant role in determining sputtering rates. Related formulas developed by Sigmund, using a Thomas-Fermi collision cross-section, provide even better fits to high-energy, several keV to MeV sputtering results.

Returning to Wehner's results in Figure 15 (Section 4.2) showing the average energies E_{at} of atoms sputter ejected, due to bombardment by Kr^+ ions, orthogonal to the surface of a large number of elemental polycrystalline targets, the results for the heavy elements Ta through U can now be understood based upon equations (1) and (2). A "light" ion hitting a very heavy target does not penetrate deeply and has a high probability of sideways scattering. Thus, more energy is deposited near the surface than would be the case for a collision of the same ion with a light target; therefore the product of the sputter yield S and the average ejection energy E_{at} will be high. However, as shown in equation (1), $S \propto 1/U_{vap}$. Comparing E_{at} results for Pt (195.08 amu) and Au (196.97 amu), whose masses are essentially the same compared to 83.80 amu for Kr (thus momentum transfer differences can, to first order, be ignored), S_{Au} is predicted to be larger than S_{Pt}, as observed experimentally, since $U_{vap,Pt}$ = 5.8 eV while $U_{vap,Au}$ is only 3.7 eV.[266] As a consequence, the average Au atom ejection energy is much lower than that of Pt in agreement with the results in Figure 15.

In the early 1980s, German physicists Bohdansky, Roth, and Bay[270] developed an empirical formula for the sputtering yield in the near-threshold regime. Yamamura et al., from Nagoya University, Japan, proposed three different versions of a correction term to Sigmund's yield formula[267,268] to account for threshold-energy effects.[271-273] A fourth correction term was proposed by Bodansky.[274] In 1991, Petrov, Orlinov, and Grudeva,[275] of the Bulgarian Academy of Sciences concluded, based on comparisons of published experimental results with the predictions of Bohdansky et al.[270], Yamamura et al.,[271] and Matsunami et al.[272] that for rare-gas sputtering (i.e., using Ne, Ar, Kr, and Xe) under conditions relevant to film deposition, the best fit is obtained with the latter two formulas using, in the case of heavy ions impinging on targets composed of lighter-mass atoms, for which $m_T/m_i < 0.3$, an additional small correction in the threshold-energy value.

4.9. rf sputter deposition

It was clear to those involved in the earliest studies of dc-glow-discharge sputtering that the technique cannot be used for insulators since the accumulation of positive charge at the target, which can't be conducted away, will decrease, and eventually eliminate, the ion sheath (the cathode fall region), thus extinguishing the discharge if the process hasn't already been terminated due to arcing. Thus, an ac discharge is required in order that electron bombardment during the positive part of each cycle neutralizes the accumulated positive-ion charge from the previous negative half cycle. However, this, by itself, is still not sufficient to produce significant sputtering from a dielectric target.

Several researchers during the period from the mid-1950s to mid-1960s recognized that the discharge frequency is a crucial parameter for optimizing dielectric sputtering. At low frequencies, discharges can be established, but there is no net dc potential at the target (i.e., there is no ion sheath) since the ion flux reaching the target during the negative half cycle is sufficient to neutralize the negative charge from the positive half cycle. However, electron mobilities, due to their low mass, are very much higher than ion mobilities; thus, as the discharge frequency is increased, at some point there will not be enough time for the majority of the ions to reach the target in a single half cycle. That is, the ions become inertially confined (their transit time across the ion sheath is much less than the ac period). At this point, the target surface accumulates negative charge and the net dc potential repels "excess" electrons to maintain a zero net dc current over each cycle. As the positive-ion sheath potential, which accelerates ions to the target, increases with increasing frequencies, sputtering begins to occur. The lowest frequency at which an ion sheath can form was estimated by Butler and Kino[276] in 1962 to be ~10 kHz, but simple estimates made by Wehner and colleagues[277] in 1962 show that frequencies in the MHz range are more useful for sputtering since the net dc cathode-fall potential V_T increases with frequency. Both estimates are in the radio-frequency (rf) range which extends from approximately 3 kHz to 1 GHz.[278]

> ***Historical footnote***: Today, rf power supplies used in sputter deposition systems are typically operated at 13.56 MHz to avoid the additional expense of electromagnetic shielding. 13.56 MHz, and the harmonics 27.12 and 40.68 MHz (together with other frequencies such as 2.45 GHz for microwave ovens) were set aside by international agreement as Industrial, Scientific, and Medical (ISM) bands, not to be used for telecommunications, at the international telecommunications conference in Atlantic City, NJ, in 1947.[279]

The earliest report on rf gas discharges was by J. J. Thomson in 1891.[1,280] He employed a Sprengel-type mercury pump and applied power induc-

tively to a Leiden jar (see Section 3.2) at frequencies in the MHz range ("…. currents change their directions millions of times in a second ….").
In his experiments, Thomson noted that it was important to have flowing gas: "It is essential to success that the gas in the bulbs or tubes should be quite dry and at a suitable pressure; there is a pressure at which the brilliancy of the discharge is a maximum, and as endeavoring to get to this pressure the exhaustion may be carried too far, it is convenient to use a form of mercury pump which will allow of the easy admission of a little gas."[280] His experiments were carried out in air, carbonic acid [H_2CO_3], hydrogen, oxygen, coal gas, and acetylene [C_2H_2]. In the same paper, Thomson also describes the effect of the addition of a magnetic field, oriented both along and perpendicular to the discharge axis, on the distribution of discharge luminosity as determined by eye and via spectroscopic analyses.

In 1933, D. Banerji and Radharaman Ganguli carried out experiments using 4 MHz inductive rf plasmas to deposit mercury films on the glass walls of a discharge tube.[281] They realized that mercury atoms, evaporated from a heated source, were ionized in the discharge. In the same year, J.K. Robertson and C.W. Clapp reported the use of an inductive rf discharge for removing (ion-etching) metal layers deposited on pyrex, soda glass, and quartz discharge tubes.[282] "On applying a high-frequency field by means of external wire electrodes so that a luminous discharge was obtained (in air at a pressure of a few millimetres [a few Torr]), in a few minutes the deposit was removed in the neighbourhood of the electrodes. By shifting the position of the electrodes, the whole tube was cleaned in a short time. …. A high-frequency potential [frequency not specified] applied to the electrodes caused no removal unless a gas or vapour was present to carry an actual discharge. Most of the work was done with air as ionized gas but a few observations with hydrogen, nitrogen, and oxygen showed that with hydrogen and oxygen the action was extremely rapid, with nitrogen very much less so." A sketch of Robertson and Clapp's apparatus is shown in Figure 33. The electrode positions are indicated by the letters E and the cross-hatched region corresponds to the luminous area of the discharge. R. H. Hay, in 1938 reported similar results using frequencies of 3, 60, and 300 MHz in air, nitrogen, oxygen, and hydrogen discharges operated at pressures ranging from 70 mTorr to 3.5 Torr.[283]

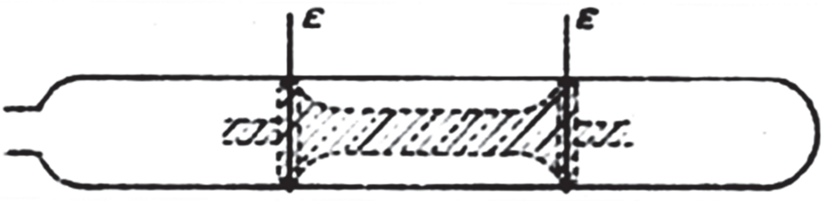


Figure 33. Sketch of an inductively-powered rf discharge tube in which the plasma was used to remove (ion etch) metal layers previously deposited on the inside of the pyrex tube. The electrode positions are indicated by the letters E and the cross-hatched region corresponds to the luminous area of the discharge. Ref. 282.

In a 1955 review article on dc sputtering, Fred Wehner[284] described the concept of rf sputtering: "The positive charge on the insulator surface [due to positive-ion bombardment] can be removed or the negative potential part time increased when a high frequency potential is applied to a metal base underneath the insulator." Jackson and Kane, in 1959,[285] were perhaps the first to suggest that the application of an rf field in a low-pressure environment (in their case, two collinear antennas, driven at 7.75 MHz, on a rocket passing through the ionosphere) results in a sheath accelerating positive ions to an insulator. This was demonstrated two years later by H. S. Butler using an inductively-powered ring-electrode around a glass discharge tube; the results are described in a 1961 Stanford University Report.[286] Two years later, Butler and G. S. Kino published both experimental and theoretical results in the open literature.[276]

Anderson, Mayer, and Wehner,[277] in 1962, are credited with designing the first modern capacitively-coupled rf discharge for depositing dielectric thin films. "We describe a simple method for large-area sputtering of insulators which promises to be useful in many applications. The trick is to use a high-frequency potential at the target whereby the positive charge which accumulates on the target is periodically neutralized with plasma electrons during a portion of each cycle."

The electrode configuration in the discharge tube used by Anderson et al.[277] is shown in Figure 34 and described as follows. "The dielectric target is immersed in the plasma of a low-pressure gas discharge. Immediately behind the dielectric target is a conducting plate to which a high-frequency potential (typically in the Mc [MHz] range) is applied with respect to the

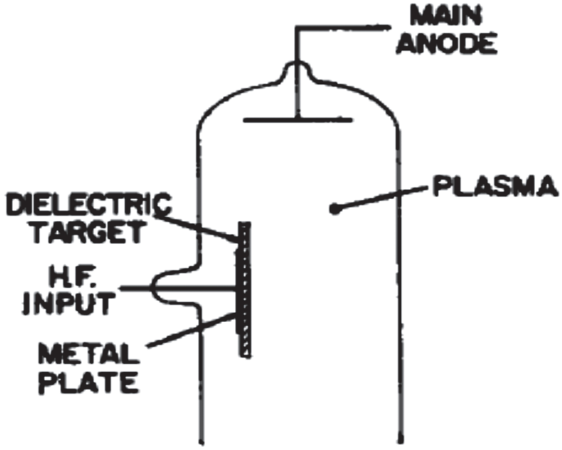


Figure 34. Electrode configuration in the rf discharge tube used in Ref. 277 to sputter insulating targets.

plasma. The potential of the dielectric surface, measured with respect to the plasma potential as a function of time would be roughly as pictured in Figure [35(a)]." What the authors realized in sketching the figure was that the entire input waveform has to float down in potential such that the highly mobile electrons are captured only over a very short time in each cycle, while the rest of the cycle time is used to attract the slower ions. That is, the ions respond only to the time-averaged electric field, while the much lighter electrons respond to the instantaneous field. Thus, while not mentioned explicitly in the article, the ions are accelerated on average by approximately half the peak-to-peak rf potential.

In Figure 35(a), the positive portions of each cycle were drawn with a *clipped* sine wave following ref. 286. Butler and Kino[276] showed experimentally that this is correct and explained it theoretically as being due to non-linear charge transport as the electrons expelled from the ion sheath during the negative part of each cycle respond extremely rapidly, while the ions respond very slowly, to the target voltage becoming positive.

Anderson et al.[277] went on to further explain the system operation. "The [target] potential will bias negatively an amount V since the electron mobility is much higher than the ion mobility and no net current can flow through the dielectric. When the potential of the dielectric is negative with respect to the plasma potential, the electric field will produce an ion sheath

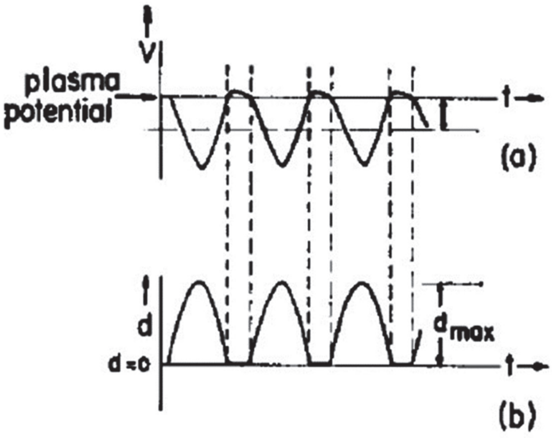


Figure 35. Schematic illustration of the time variation in the (a) potential V at the target and (b) the cathode sheath thickness d during operation of a capacitively-coupled rf discharge. Ref. 277.

by forcing the plasma electrons away from the surface. The ion sheath will extend into the plasma a distance d as a function of time, as schematically illustrated in Figure 35(b). …. In general, it will require several cycles before an ion has traversed the distance d_{max} [defined in Figure 35(b)], and the ion will finally strike the target with an energy less than the maximum potential difference." Anderson et al.[277] primarily employed rf sputtering to clean the inside of their mercury-pool-supported discharge tubes used for investigating the physics of sputtering.

The potential differences between the plasma and both electrodes in a parallel-plate rf sputter-deposition system can be approximated as a leaky capacitively-coupled voltage divider, i.e. as two capacitors in series for which the applied voltage is distributed as[287]

$$\frac{V_T}{V_s} = \left(\frac{A_s}{A_T}\right)^4 . \qquad (3)$$

V_T and V_s are the target and substrate average dc voltages and A_T and A_s are the corresponding electrode areas. An rf plasma is far from an ideal circuit element, thus the power law exponent in equation (3) above is generally closer to 1.5 to 2. In order to provide most of the applied voltage to the target, the metal substrate table is typically connected electrically to the entire vacuum system (including all grounded fixtures) such that $A_s >> A_T$. Nevertheless, the substrate in an rf discharge will always experience ion bombardment (bias sputtering) with peak-to-peak voltages ranging from ~20-25 V up to much more than 100 V depending on system design, gas pressure, etc. A schematic illustration of the voltage distribution is shown in Figure 36.

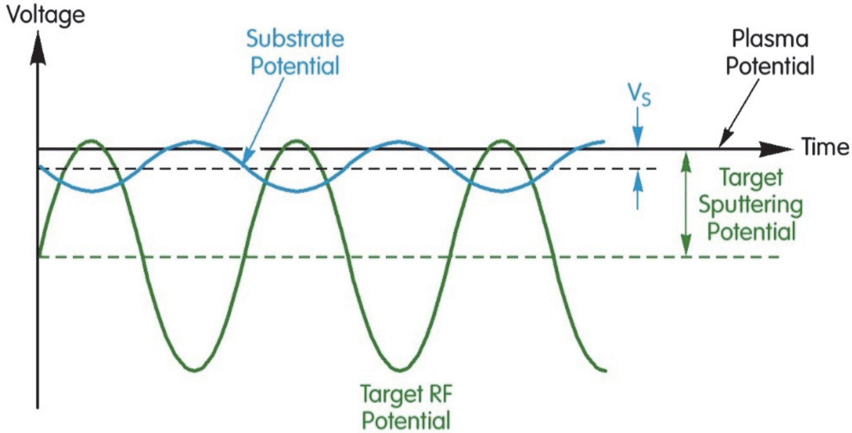


Figure 36. Schematic illustration of the target and substrate potentials in an rf sputtering system.

Pieter Davidse and Leon Maissel at IBM East Fishkill, New York, filed a US patent in 1965 for the design of a capacitively-coupled rf diode sputtering system to deposit dielectric thin films.[288] Figure 37 is a sketch of the target apparatus[289] described in the patent. The back of the dielectric target "can be metallized and bonded to the electrode" to which the rf power is supplied. In order to decrease sputtering from other parts of the system, which would result in film contamination and decrease the total power applied to the target, a ground shield is placed behind the target at a spacing less than the cathode-fall distance. Davidse later recommended a spacing of ~5 mm.[290] The patent states that a matching network is required in order to tune "the load" [the discharge] to match the rf power supply. Modern rf power supplies are designed to connect to a load which is purely resistive, thus the discharge capacitive and inductive reactances must be compensated by the matching network.

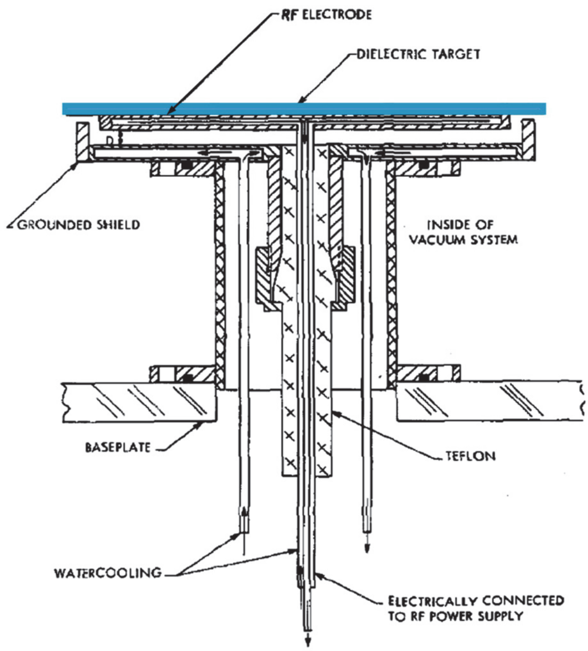


Figure 37. Target electrode design for rf sputtering of dielectrics. Ref. 289. Color (the dielectric target is blue) added by present author.

In the initial design, Davidse and Maissel placed a set of toroidal permanent magnets above the anode to provide a curved magnetic field extending between the cathode and anode (substrate holder) in order to increase ionization as discussed in Section 4.6. The patent claims that the magnetic field also stabilizes the discharge and makes it easier to match the power supply to the discharge. For 12.7-cm-diameter (5 inch) quartz and pyrex targets rf-sputtered at 13.56 MHz, 3000 V peak-to-peak, in Ar at 5 mTorr with $\overline{B}$ = 110 G, deposition rates of 10.2 and 3.3 μm/h were reported for an ~2.5 cm separation between the target and Si wafer substrates.

In a 1966 paper, Davidse and Maissel noted that their rf-sputter-deposited oxide films had a "slight oxygen deficiency" as deduced from film optical properties.[289] However, when they mixed oxygen with the Ar sputtering gas in an attempt to obtain stoichiometric oxides, they observed a large decrease in the deposition rate, thereby presaging issues in modern reactive sputtering (see Section 4.10). During separate experiments in which a pyrex target was rf sputtered in pure Ar and in a 10 mole% O_2/Ar mixture, the deposition rate loss in the mixed gas was greater than 50%. The films sputtered in Ar, although oxygen deficient, exhibited relatively uniform film thicknesses over a radial distance approximately equal to that of the target radius and were reported to contain low pinhole densities.

Davidse pointed out in 1966[290] that a metal target can also be sputtered in an rf system by simply adding a capacitor between the power supply and the cathode; data was shown for rf sputtering of aluminum as a function of target power. He also noted that for deposition on insulting substrates, applying rf power to the substrate allows the use of positive ion bombardment of the growing film (bias sputtering) to tune film properties such as density, surface roughness, and preferred orientation.[36-38,82,185,243,244] However, no data was provided.

Soon after commercialization of the modern planar magnetron in the mid-1970s (Section 4.6), rf power supplies were used to deposit dielectric films by "rf magnetron" sputtering.[291,292] However, deposition rates, were low since magnetron sputtering, which relies on a closed electromagnetic trap at the target, is fundamentally a dc concept.[82] The rf field alternately opens and closes the trap, allowing electrons to escape when the trap is open and forcing electrons to cross magnetic field lines, an added impedance which decreases the available power at the target, as the trap is reformed. For a given applied magetron target power, the film deposition rate decreases with both increasing frequency[293] and increasing magnetic field.[294] The fact that "rf magnetrons" do not function as a magnetron is clear from examining the target current/voltage (I_T/V_T) characteristics, in which $I_T \propto V_T^n$. The exponent n for an "rf magnetron" is closer to that of rf (and dc) diode sputtering, 1-2, than that of dc magnetron sputtering, $\gtrsim$ 10.[295] Because of this, plus the added expense of an rf (vs. dc) power supply and matching network, rf magnetron sputter deposition of dielectrics has largely been replaced by reactive sputtering from metal targets using much lower frequencies, typically 50-300 kHz, and pulsed-dc or mid-frequency ac power supplies (Section 4.10).

4.10. Reactive sputter deposition

The majority of early experiments on sputter deposition of metallic thin films actually involved the growth of metal oxide layers (reactive deposition), with other contaminants including nitrogen and carbon, due to poor vacuum, and most researchers realized this. K. Bädeker published an early (1907) paper on the purposeful synthesis of several metal-oxide thin films.[296] Of particular interest was CdO, the first transparent conducting oxide (TCO), produced by sputter-depositing cadmium layers on glass, followed by thermal oxidation. TCOs are important today for use as electrical contact layers on many opto-electronic devices, especially solar cells.[297] CdO is an n-type semiconductor due to oxygen vacancies in the lattice and Bädeker obtained resistivities as low as 1.2×10^{-3} Ω-cm, which is only about one order of magnitude higher than the resistivity of the best indium tin oxide TCO layers available today.

Clarence Overbeck (Northwestern University),[298] in 1933, was the first to publish an article devoted specifically to the investigation of reactive sputter deposition, although the term was not used until two decades later (see *Historical footnot*e below). He sputtered tin from a liquid target, held in a pyrex cup, using dry air, oxygen, nitrogen, and hydrogen discharges with dc potentials ranging from 1.8 to 2.6 keV. In initial experiments, carried out in a fixed (non-flowing) pressure of dry air, he observed a pressure decrease which he realized was due to gas incorporation in the film. Subsequent experiments were carried out at constant gas flow. Films deposited in flowing air and oxygen were oxides, with similar appearance, exhibiting interference rings which changed systematically with film thickness. Overbeck assumed that the films were stannic oxide, SnO_2.

Historical footnote: the first use of the term "reactive sputtering" was by Gabor Veszi, from Megatron Ltd. (London), on January 5, 1953, during a lecture at a British Institution of Radio Engineers conference in London. His paper was published in April, 1953,[299] approximately a century after the earliest recorded sputter-deposition experiments[27] and 20 years after Overbeek's pioneering reactive-sputtering experiments.[298] Veszi used reactive sputtering to deposit cadmium-oxide TCO electrodes on selenium photocells, but provided few details. He did note that the CdO layers deposited on glass exhibited "transparencies up to 85% and surface resistances of 100 ohms per square or less [film thicknesses were not given, therefore sheet resistance cannot be converted to resistivity for comparison to Bädeker's 1907 results]." Later in April and October of the same year, two articles published by Leslie Holland and George Siddall of Edwards High Vacuum Ltd., Sussex, UK, used the term in the titles of their papers on the design of a reactive-sputtering system[300] and the properties of reactively-sputtered metal-oxide films.[301]

Layers deposited by sputtering tin in nitrogen discharges were opaque with a brown color. Based upon wet-chemical analyses, the films were understoichiometric SnN_x. However, they were apparently underdense since Overbeck reports that "exposing the film to air caused it to gradually lose its opacity and take on the transparent nature of films produced in air." Films grown in hydrogen,which required "higher pressures," were metallic Sn with highly reflecting metallic mirror surfaces.

Overbeck was also the first, by several decades, to report problems with arcing during reactive sputter deposition of oxides. While he did not use modern terminology, his description of the process, as noted in the following quotation, was correct. "Frequently the discharge became unstable, giving rise to a sudden high current density which pitted the cathode surface and produced a granular metallic deposit on the plate. A microscopic examination revealed that these metallic particles were of …. a rough spherical shape ….. it appeared that they had been flattened on striking the deposit plate, which indicated considerable velocity of impact and heating. The metallic nature of the deposit might be explained by the fact that the particle, rapidly deposited, was of large size and therefore its combination with gas molecules would not be favored."

It is now well understood that arcs can occur during dc reactive sputtering of electrically conducting targets due to the formation of local insulting regions (often oxides) on the target surface.[302] The system rapidly switches from a high-voltage, low-current glow discharge to a low-voltage, high-current arc. All of the power is applied to the local region, which typically ranges from ~0.01 to 100 μm in diameter, resulting in the current density increasing by many orders of magnitude, giving rise to local heating leading to thermionic emission and a micro-explosion. Thermal runaway causes local melting and boiling of solid targets over time scales of order ns leading to the ejection of macroscopic liquid droplets, with very high velocities, which can land on the growing film surface[303] as Overbeck reported in 1933.[298] Typical scanning electron micrographs showing the effect of an arc on a target surface (leaving, in this case, an ~12 μm diameter pit) and collateral effects of arcing on film growth are shown in Figure 38. There are, today, a variety of solutions available for solving, or at least minimizing, the arcing problem; they all involve fast arc detection circuitry, dumping excess power into a massive bus bar, and periodically (typically 50-350 kHz or 13.56 MHz) switching the target voltage via pulsed-dc, mid-frequency ac, or rf power supplies to neutralize accumulated positive charge, by attracting electrons, as discussed in Refs. 82 and 304.

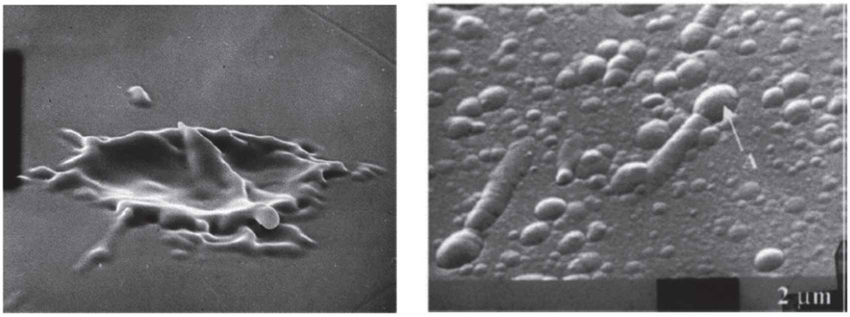


Figure 38. Left •gure is a scanning electron micrograph of an ~12-µm-diameter pit formed at a metal target surface due to arcing (image courtesy of Dr. Andre Anders, Lawrence Berkeley National Laboratory). The right •gure shows embedded metal droplets in an underdense Al_2O_x •lm deposited by reactive sputter deposition from an Al target in a mixed Ar/O_2 atmosphere. Adapted from Ref. 305.

The mercury-pumped deposition system used by Overbeck,[298] Figure 39, is itself of interest since it contained, in 1933, many of the features, although in a slightly different guise, found in modern ultra-high vacuum systems: vacuum gauging (a McLeod gauge, see Section 3.1), multiple chambers, liquid-nitrogen [here liquid-air] traps, gas scrubbers, facilities for multiple substrates, a magnetically-coupled rod to transport substrates in and out of the deposition chamber, and the capability to controllably vary the target-to-substrate distance via a second magnetically-coupled rod. Note that opening the system to retrieve the deposited films required breaking the end of the side tube (the reason, of course, for multiple depositions per system pumpdown), and reforming it by glassblowing. A description of the apparatus, in Overbeck's words, follows.

> "A steady potential, variable from 1800 to 2600 V, was applied between the aluminum anode, D, and the tin cathode, F (2 cm in diameter and 2 cm long). The cathode was placed in a Pyrex cup with its surface flush with the top of the cup. The deposit formed on a Pyrex plate, E (3 cm wide and 45 cm long). The distance between cathode and plate was adjustable and held by an electromagnet acting on a glass-enclosed piece of soft iron, M. The deposit plate could be drawn back and forth in the side tubes by a second

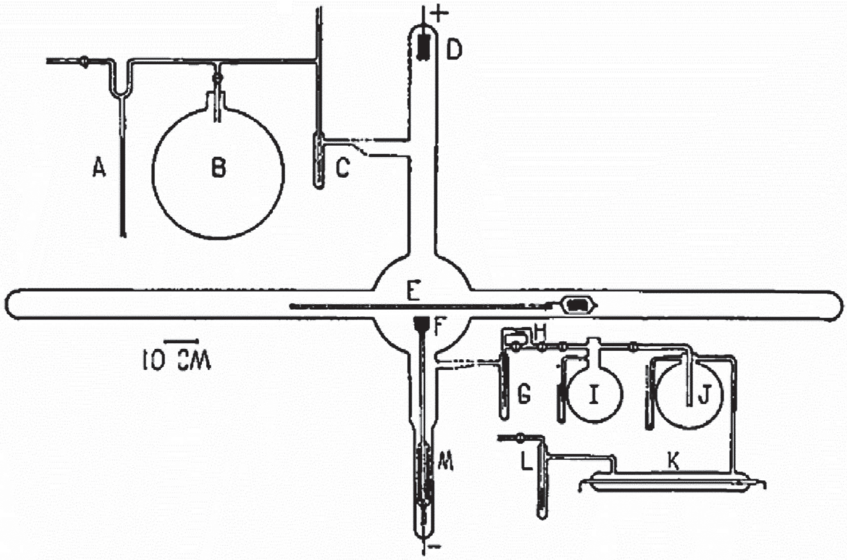


Figure 39. Pyrex vacuum system used by Overbeck (adapted from Ref. 298) to investigate sputter-deposition of Sn in dry air, O_2, N_2, and H_2 environments. See text for detailed description.

electromagnet. This permitted making from six to ten deposits before blowing out the end of the side tube to remove the deposit plate. The sputtering chamber was separated from the remainder of the system by liquid air traps, C and G. A McLeod gauge was attached above C. A 12 liter bottle, B, was placed in the system to stabilize the gas pressure. This added volume reduced the pressure variation caused by the vigorous 'cleanup' action which was especially large at the beginning of a run. A high-vac pump was attached beyond the mercury cut-off, A. The lower right-hand corner (of the figure) shows the gas purifying chambers. Water vapor, carbon dioxide, and oxygen were removed from the incoming gas by phosphorus pentoxide in J, sodium hydroxide in L, and hot copper gauze at K. The gas was finally collected for use in flask, I, from which it could be admitted into the sputtering system by either of two methods: (1) A capillary opening at H permitted a constant flow of fresh gas through the system. With the pump in operation, proper adjustment of the pressure in I gave any desired pressure in the system. (2) Known quantities of gas could be admitted periodically by means of the stopcocks at H."

While Gabor Veszi, in 1953, was the first to use the term “reactive sputtering” in discussing his results on the deposition of CdO (see *Historical footnote* above),[299] he was not the first to report the reactive sputter deposition of transparent conducting oxides. J. S. Preston, in 1950, described experiments in which he used dc sputtering of a Cd target in flowing Ar to deposit a layer on a “metal” substrate.[306] Electron diffraction analysis showed that the film was cadmium oxide, presumably due to residual oxygen and water vapor in the chamber together with desorption from the walls and fixtures upon establishing the discharge. The vacuum chamber had not been outgassed [the system base pressure, the sputtering pressure, and the film thickness were not specified]. Preston noted that “Sputtering in air [instead of argon] was found to give films having a far lower conductivity.” Preston, like Vesci, was interested in CdO as a transparent contact layer on selenium photocells.

More definitive experiments on the conductivity and structure of CdO layers deposited on glass substrates were described in a 1952 paper by G. Hellwig.[307] He showed that the conductivity of CdO_x layers decreased [due to fewer O vacancies, i.e. larger x] with increasing oxygen partial pressure in the mixed Ar/O_2 sputtering gas and that post-annealing as-deposited CdO_x films at reduced pressures increased the conductivity [due to loss of oxygen]. Holland and Siddall[301] later reported that the post-annealing step for obtaining low CdO_x resistivities can be avoided by simply not cooling the glass substrate and allowing the deposition temperature to increase. The application in this case was for transparent heating elements to prevent the formation of mist and ice on aircraft windows.

As described in the book *Sputter Depositio*n by Bill Westwood,[308] an early industrial application of reactive sputtering in the 1950s derived from his research on the deposition of N-doped Ta films, by dc sputtering from a Ta target in mixed Ar/N_2 discharges, for use as trimming resistors in hybrid circuits. Controlling the N_2 flow rate also resulted in the formation of TaN_x films with the appropriate temperature coefficient of resistivity (TCR) for use in tuned circuits in touch-tone telephones which first appeared in the early 1960s. Figure 40 is a plot of film resistivity ρ and TCR vs. N_2 flow rate. Similar curves for TaN_x film resistivity plotted vs. N_2 partial pressure P_{N_2} were published in the mid-1960s (see, for example, references 309 and 310). When combined with electron diffraction, the general trends in refer-

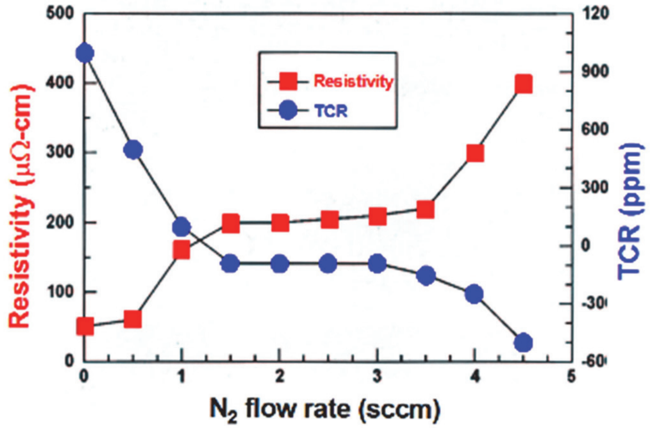


Figure 40. Variation, as a function of the N_2 • ow rate, of the resistivity and the temperature coef• cient of resistivity (TCR) of TaN_x • lms grown by dc sputtering Ta in mixed Ar/N2 atmospheres. Adapted from Ref. 308.

ences 309 and 310 showed that ρ initially increases slowly with increasing P_{N_2}, resulting in N-doped Ta films (Ta:N), then a rapid increase in resistivity is observed corresponding to two-phase films consisting of Ta:N with an increasing fraction of Ta_2N, followed by a saturation region in which single-phase Ta_2N is formed. At still higher P_{N_2} values, two-phase Ta_2N+TaN layers and, eventually, TaN films are obtained. The absolute values of ρ vs. P_{N_2}, as well as the exact curve shapes, depend on system geometry and operating conditions (base pressure, outgassing conditions, target voltage and current, target-to-substrate distance, substrate type and temperature, etc.).

The complexity in interpreting the above experimental results is highlighted by the phase map in Figure 41 for the growth of TaN_x on SiO_2 and MgO(001) substrates by dc planar-magnetron sputtering in an ultra-high vacuum system, with a base pressure of 5×10^{-10} Torr, as a function of the deposition temperature T_s and the mole fraction f_{N_2} of N_2 in high-purity Ar/N_2 gas mixtures at a total pressure of 20 mTorr.[311] The growth phase map contains nine different phases of which body-centered cubic α-Ta, hexagonal γ-Ta_2N, and hexagonal ε-TaN are thermodynamically stable, the rest are metastable.

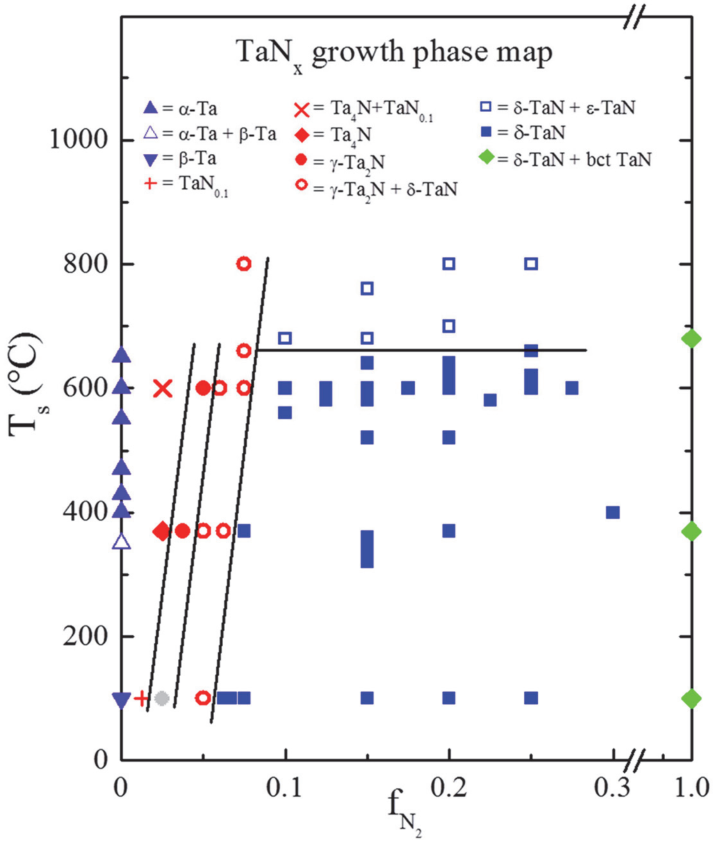


Figure 41. Growtkh phase map showing the phase composition of TaN_x layers grown on MgO(001) and SiO_2 substrates by reactive magnetron sputter deposition in mixed N_2/Ar atmospheres as a function of film growth temperature T_s and N_2 fraction f_{N_2} . Adapted from Ref. 311.

From the early days of reactive sputtering, investigators realized, based on target color changes and decreases in film deposition rates, that reactive gases were poisoning the target surface by both simple adsorption and by reactive-gas ionization in the discharge followed by acceleration to the target. However, it was unclear at the time where the *primary* reactive-gas/metal-atom reaction controlling film growth occurred. Holland and Siddall,[300] in 1953, wrote the following. “When sputtering from a metal cathode in an oxygen atmosphere, there are three possible mechanisms by which a metal oxide film can be formed, all three of which may occur simultaneously: (i) The initial formation of an oxide layer on the cathode surface by bombardment with oxygen molecules or ions, which is removed in molecular form during subsequent bombardment. (ii) Oxygen absorption by sputtered metal atoms during transit to the receiver due to collisions between gas molecules and metal atoms. (iii) Absorption of oxygen molecules impinging on the deposited film during its condensation."

Krikorian and Sneed,[310] in 1966, first deduced the correct answer following experiments in which they established a direct connection between film impurity concentration and the combination of the background partial pressure of reactive gas P_r with the film growth rate R. In their dc sputter-deposition experiments, carried out in a vacuum system with a base pressure of $1x10^{-7}$ Torr, with a tantalum target in argon discharges, they measured the resistivity ρ as a measure of impurity concentration and the deposition rate R of Ta films deposited as a function of the sputtering pressure P on glass substrates at $T_s = 450$ °C. Here, P_r, which in this work is implicitly assumed to be proportional to P, is directly related, though gas kinetics,[312] to the rate per unit area N_i at which background impurity reactive-gas molecules strike the film growth surface and react with a sticking probability s_i. From previous results of Maissel and Schaible,[313] the impurity concentration C_i in an as-deposited film can be expressed as a function of R and the impingement rate N_i of impurity molecules through the relationship: $C_i = s_iN_i/(s_iN_i + R)$.

In their experimental analysis, Krikorian and Sneed made the reasonable assumption, based upon their resistivity results, that the impurity concentration is in the few percent range. Thus, s_iN_i is much less than R, yielding $C_i \simeq s_iN_i/R$, and the film resistivity should increase approximately linearly with P/R. The results, shown in Figure 42, are in agreement with this simple model. From this, and the fact that they observed the reactive sputter deposition rate of TaO_x in mixed Ar/O_2 atmospheres to increase with T_s, due to a corresponding increase in $s_i(T_s)$, Krikorian and Sneed correctly concluded that while reactions occur at both the target and the substrate, the *primary* reaction during reactive-sputter deposition occurs dynamically at the film-growth surface.

Maissel and Schaible[313] showed that the application of a small negative bias to their substrates during dc sputtering of Ta in argon could be used to tune film impurity concentrations (again, as determined by resistivity

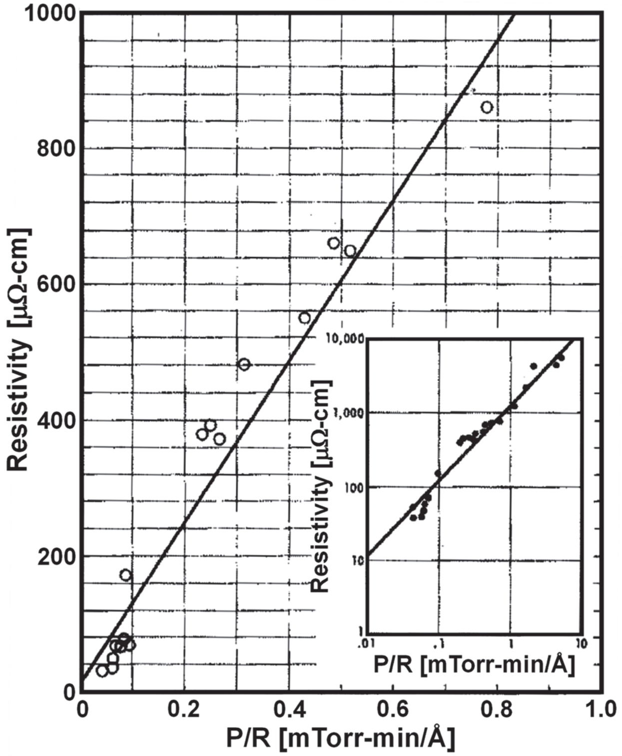


Figure 42. The resistivity of Ta films deposited by dc sputtering as a function of the ratio of the sputtering pressure P to the film growth rate R. Adapted from Ref. 310.

measurements). At low substrate bias, the primary effect is to preferentially sputter remove impurities from the growing film (thereby decreasing the resistivity), while at higher biases, this effect is counteracted by trapping (i.e., near-surface ion implantation) of accelerated ionized impurity species (the resistivity increases again).

Much later, Barnett, Bajor, and Greene[314] used this idea to grow the first sputter-deposited high-resistivity single-crystal GaAs films on GaAs(001) substrates in a vacuum system with a base pressure of $1x10^{-7}$ Torr. Residual oxygen and carbon contamination was minimized through the use of a liquid-nitrogen-cooled shroud surrounding the 20 mTorr argon rf discharge during deposition. Excess arsenic was provided to the growing film, deposited at 600 °C, by a separate sublimation source (reactive deposition utilizing As_4 molecules) to account for arsenic loss by desorption. Figure 43(a) shows the incorporated oxygen concentration in as-deposited GaAs layers, as determined by secondary-ion mass spectrometry (SIMS). The oxygen concentration decreases, resulting in higher electron mobilities up to a maximum of ~4000 $cm^2/V\text{-}s$ at a peak-to-peak rf substrate bias of -150 V. Higher bias voltages give rise to increased oxygen concentrations, via trapping, with an associated decrease in the electron mobility.

Bias sputtering was also used to control sulfur (an n-type dopant) concentrations in GaAs/GaAs(001) films.[315] In these experiments, the films were grown at 30 mTorr Ar and 570 °C with hydrogen sulfide [H_2S] partial pressures P_{H_2S} between $1x10^7$ and $1x10^{-5}$ Torr. Systematic variations in P_{H_2S} and the substrate bias resulted in controlled S concentrations ranging from $7x10^{18}$ to $1.2x10^{21}$ cm^{-3} [0.00016 to 0.027 atomic%]. Figure 43(b) shows results for P_{H_2S} = $1x10^{-5}$ Torr.

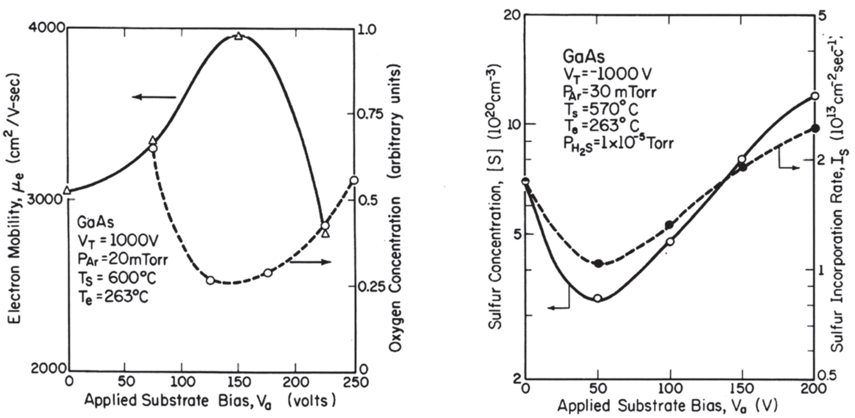


Figure 43. (a) Electron mobilities and relative oxygen concentrations in unintentionally-doped single crystal GaAs films as a function of the applied substrate bias V_a during film growth. The argon sputtering pressure is 20 mTorr, the target voltage is 1000 V, and the film-growth temperature is 600 °C. Solid arsenic is sublimed at 263 °C to provide excess arsenic incident at the growth surface in the form of As_4. Ref. 314. (b) Sulfur concentrations and incorporation rates in as-deposited single-crystal GaAs films as a function of the applied substrate bias V_a during film growth. The argon sputtering pressure is 30 mTorr, the target voltage is 1000 V, and the film-growth temperature is 570 °C. Solid arsenic is sublimed at 263 °C to provide excess arsenic incident at the growth surface in the form of As_4 and the hydrogen sulfide [H_2S] partial pressures P_{H_2S} is $1x10^{-5}$ Torr. Ref. 315.

Three seminal events occurred between the mid-1970s and mid-1980s which resulted in reactive sputtering becoming a major thin-film synthesis technique for both research and industrial applications: the development of *high-rate* reactive sputtering, the availability of pulsed-dc and mid-frequency power supplies, and the evolution of atomic-scale models for understanding and controlling the complex processes inherent in reactive sputtering.

4.10.a. High-rate reactive sputtering

From the 1930s to the early 1980s, reactive sputtering was a relatively slow, and hence uneconomic,[304] process compared to competing techniques such as reactive evaporation and chemical vapor deposition. While magnetron sputtering of metal targets provides high deposition rates, as soon as the metal target is covered with reactive gas (poisoned), the sputtering rate in many metal/reactive-gas systems decreases dramatically due primarily to the much higher energies of, for example, metal-oxygen than metal-metal bonds, as well as to changes in ion-bombardment momentum-transfer probabilities since oxygen has a lower mass than most metals and is much lower than that of argon (see equations (1) and (2) in Section 4.8).

Target poisoning during reactive sputtering occurs as an avalanche effect since the combination of oxygen adsorption (to continue the oxide example) on, and the near-surface implantation of oxygen positive ions in, a strongly-reacting metal target decreases the sputtering rate (from point A to point B in the upper panel of Figure 44).[317] Thus, there are fewer metal atoms deposited per unit time on the substrate and chamber walls to adsorb oxygen; this causes the oxygen partial pressure to increase (the corresponding transition from A to B in the lower panel of Figure 44)[318] leading to higher target coverages, and even lower deposition rates, until the entire sputtered target region is rapidly oxidized. The slow sputtering rate of the oxidized target has the consequence that the system cannot be immediately returned to its clean-target condition and the delay time results in characteristic hysteresis loops (points B to C to D in both panels) as shown in Figure 44. Films obtained with the oxygen flow rate between zero and point A are metallic and doped with reactive gas ("metallic mode"); films obtained with flow rates higher than that corresponding to point B are stoichiometric compounds ("compound mode"). Since TiO_2 is a dielectric, the experimental results in Figure 44 were obtained using ac sputtering; rf for the upper panel results and pulsed-dc magnetron sputtering (see Section 4.10.b) in the lower panel.

There are materials systems, including In/N_2, which have very weak metal/reactive-gas bonds,[318] and sets of (sometimes extreme) processing conditions such as very high pumping speeds,[319,320] for which hysteresis effects can be minimized and even eliminated (see also ref. 321). However, in many (if not most) materials systems and deposition conditions relevant to thin-film applications, loss of deposition rate and hysteresis are key issues.

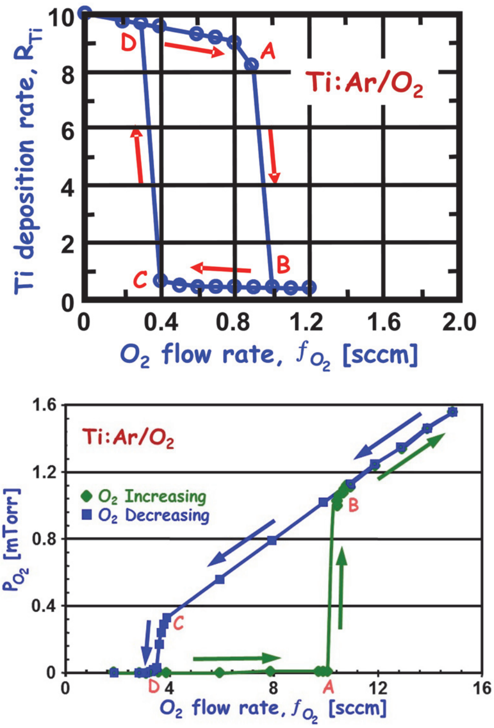


Figure 44. (Upper panel) A plot of the metal deposition rate R_{Ti}, in arbitrary units, vs. the oxygen flow rate f_{O_2} during reactive rf sputter deposition, using flow control, of TiO_x from a Ti target in mixed Ar/O_2 atmospheres maintained at an Ar partial pressure of 10 mTorr. Adapted from ref. 316. (Lower panel) Oxygen partial pressure P_{O_2} vs. oxygen flow rate f_{O_2} during reactive pulsed-dc magnetron sputter deposition (see section 4.10.b), using flow-control, of TiO_x from a Ti target in mixed Ar/O_2 atmospheres maintained at an Ar partial pressure of 3.1 mTorr. Adapted from ref. 304.

For approximately 30 years, investigators developed ever more complex feedback loops in attempts to obtain near metal-deposition rates while accepting the tradeoff in depositing slightly understoichiometric films (i.e., remaining near point A of the hysteresis loop in the upper panel, Figure 44). These experiments generally failed since the feedback element was a gas-flow meter whose actuator is a mechanical valve (flow-control mode) and therefore far too slow to halt the avalanche poisoning process.[304] In 1983, Bill Sproul and J. Tomashek,[322] working at Borg Warner in Chicago, filed a patent that contained the solution which, as is often the case, was surprisingly simple: instead of using the flow meter as the control element, use the reactive-gas partial pressure, measured by either a mass or an optical spectrometer. In partial-pressure control mode, one can operate at any point along the path of target poisoning without initiating the avalanche effect (provided the system response time is sufficiently fast). Sproul published initial papers demonstrating the partial pressure control process, with associated high reactive-deposition rates, in 1987[317] while at Borg Warner and in 1989[323] after he had joined Northwestern University. An example, growth of TiO_x films from a Ti target in mixed Ar/O_2 atmospheres is shown in Figure 45 for comparison to the lower panel in Figure 44. The dashed gray lines correspond to results obtained when operating in flow-control mode, while in partial-pressure mode, operation is possible along the black solid lines AB and CD. The highest deposition rates along AB are obtained by operating as close as possible to point A, but films with compositions closer to stoichiometric are obtained by moving away from A toward B. Reference 304 is an excellent review article describing the experimental aspects of high-rate reactive sputtering.

4.10.b. Pulsed-dc and mid-frequency ac reactive sputtering

High-rate reactive sputtering of electrically-conducting compounds and alloys such as most transition-metal nitrides, carbides, and some oxides was a reality by the late 1980s. Progress in high-rate reactive sputtering of dielectric films occurred in parallel, but required additional steps. rf magnetron sputtering is always an available option for sputter depositing dielectrics, but, as discussed in Section 4.9, it is slow and generally uneconomic. In 1975, Robert Cormia, together with Terry Trumbly and Sigurd Andresen[324] filed a US patent application with the unassuming title

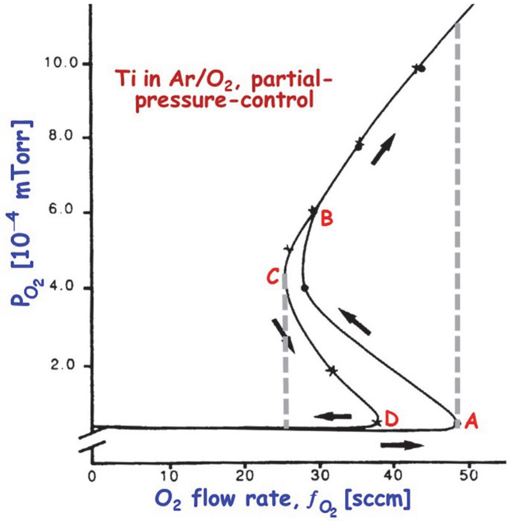


Figure 45. Oxygen partial pressure P_{O_2} vs. oxygen flow rate f_{O_2} during reactive magnetron sputter deposition, using partial-pressure control, of TiO_x from a Ti target in mixed Ar/O_2 atmospheres. The dashed gray lines correspond to results obtained when operating in flow-control mode. Adapted from ref. 323.

"Method for Coating a Substrate" (issued in September, 1977) that would lead to solutions for high-rate deposition of dielectric films. The patent describes a single flat-plate magnetron target with an ac power source operated at frequencies up to ~60,000 Hz, far below the rf range, for the reactive sputtering of oxides and nitrides and states the following. "It is not necessary that the applied waveform be sinusoidal or have a symmetrical waveform." The goal was to minimize or eliminate arcing (see Section 4.10), which as they pointed out, decreases the deposition rate, degrades film quality, and can harm the power supply. Thus, the key was to allow target polarity reversal for sufficiently long periods that the accumulated positive charge due to ion bombardment of the target is neutralized by electrons. The patent correctly describes many complex issues which were not understood in detail until many years later:

- the ac frequency must be high enough to neutralize the accumulated positive charge at the target before the breakdown potential of the insulting layer is exceeded,
- the target area at most risk for arcing is immediately adjacent to the high-erosion rate (racetrack) region,

- the thicker the insulating layer on the target, the higher the required frequency for a material with a given dielectric constant,
- for the same insulting layer thickness, materials with higher dielectric constants allow the use of lower frequencies,
- if charge not fully neutralized, step-by-step accumulation occurring over many periods can also lead to arcing (i.e., charge build-up is cumulative).

In 1986, Fazle Quazi filed a patent for reactive sputtering from a single magnetron target using two power supplies to provide discrete alternating positive and negative pulses, again with the goal of periodically neutralizing accumulated positive charge on poisoned regions of the target in order to prevent arcing.[325] This allowed independent control of frequency f as well as the duty cycle $\eta = \tau_{on}/(\tau_{on} + \tau_{off})$, in which τ_{on} is the time during a given cycle that the target is powered negatively to enable sputtering and τ_{off} is the time that the target is powered positively in order to attract electrons for neutralizing accumulated positive charge. The frequency f of the applied power is related to the duty cycle η via the relationships:

$$f = \frac{1}{\tau_{on} + \tau_{off}} = \frac{\eta}{\tau_{on}} = \frac{1-\eta}{\tau_{off}} \tag{4}$$

Note that τ_{on} is always much larger than τ_{off} due to the large difference in ion and electron mobilities.

In 1993, Siegfried Schiller and colleagues,[326] from the Fraunhofer Laboratory in Dresden, Germany, described single-target pulsed-dc sputtering (Figure 46(a)) in which a switching unit is placed in the circuit between the dc power supply and the magnetron target to provide square-wave pulses for which, in principle, f, τ_{on}, and τ_{off} can all be controlled independently (in fact, it took some time before such a system was available commercially). The square-wave pulses become highly distorted in the plasma due to the complex reactance. Schiller et al. recommended an optimal operational frequency range of 10-100 kHz.

One issue affecting all single-target pulsed reactive sputtering approaches is that even at a target frequency that minimizes arcing (arc suppression circuitry will always be required during reactive sputtering of insulators to eliminate “random” arcs), as the anode becomes covered with an insulating layer, process parameters begin to drift and the discharge is eventually extinguished when the electron current can no longer reach ground (the "disappearing anode" problem[304]). However, a practical solution, proposed earlier the same year by P. Frach et al.[327] (also from Fraunhofer, Dresden), is to use an anode with a large surface area and place it in an isolated region of the vacuum system which is hidden from the intense region of the plasma immediately in front of the target. Providing a flow of argon over the anode is also desirable. While reactive gas will still reach the anode, the reaction probability of molecular species is much less than that of atomic species formed in the intense region of the discharge. The goal is to maintain a stable ground (open regions on the anode) throughout the lifetime of the target, then replace the anode when the chamber is opened to replace the target.

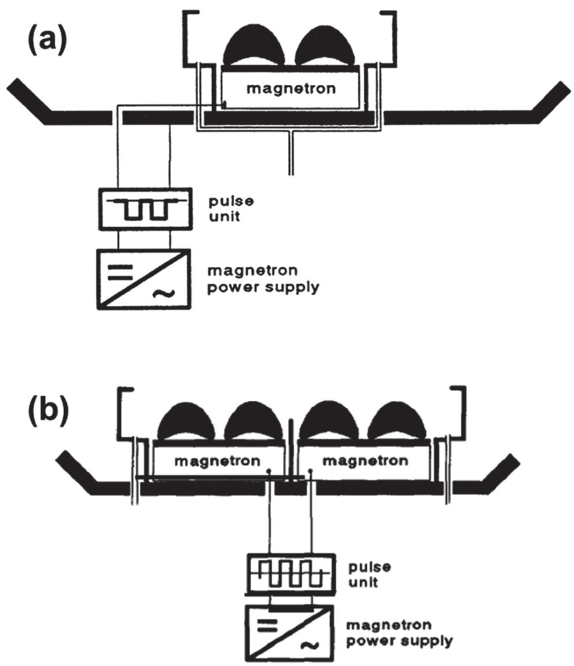


Figure 46. Schematic illustration of a pulsed dc sputtering system with (a) a single target and (b) two targets. Adapted from Ref. 326.

In 1988, prior to Schiller's publication, G. Este and Bill Westwood[293] had described a different approach to minimizing arc formation during high-rate reactive sputtering of dielectrics: the use of two targets with an ac mid-frequency power supply wired such that when one target is negative with respect to the plasma, the other target is positive and acts as the anode for the system. Then, during the following half cycle, the voltages are reversed and the target that had been the cathode becomes the anode which, if process conditions are chosen correctly, is now sputter cleaned. Thus, there is always an anode and even though the chamber is being coated with an insulating layer, there is no "disappearing anode" issue. The pulsed-dc version of the two-target mid-frequency ac approach is shown in Figure 46(b). In Este and Westwood's experiments, the two targets were opposite each other; the earliest report of two-target mid-frequency reactive sputtering with side-by-side targets, as is typically practiced today, was by Scherer et al. in 1992.[328]

A comparison between pulsed dc and mid-frequency ac dual-target sputtering is shown in Figure 47.[329] The term "voltage scrub" in the figure refers to attracting electrons for neutralizing the accumulated positive-ion charge at the target. Note that in the mid-frequency case, the sine wave is automatically half-wave rectified at the plasma potential; thus, more of the power per pulse is utilized for sputtering.

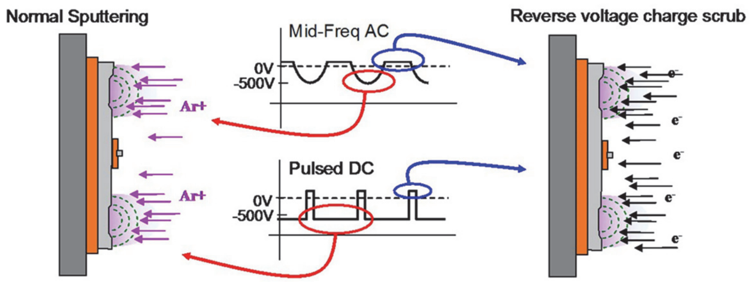


Figure 47. A schematic comparison of two-target pulsed-dc vs. mid-frequency ac waveforms during reactive sputtering. Ref. 329.

Mid-frequency ac reactive sputtering, with typical frequencies in the range 50-300 kHz (moving toward the higher end today), is commonly used in manufacturing for large-area coating of products such as architectural and automotive glass and in roll-to-roll web coating of flexible materials such as polymeric sheets (Section 4.11). In these applications, rotatable targets have, in many companies, replaced planar rectangular targets, for the reasons discussed in Section 4.6.

A disadvantage of the dual-target approached compared to single-target sputtering is that the racetrack current must be returned to its quasi-steady state value each half cycle. This results in transients leading to voltage and current overshoots giving rise to substantial fluxes of hot electrons and ions bombarding the growing film. Thus, substrate temperatures are considerably higher during dual-target sputtering which can pose problems for thermally-sensitive substrates.[330] Moreover, the transients are more severe for mid-frequency ac than for pulsed-dc[331,332] since the plasma must be reestablished each half cycle during mid-frequency ac (not required in pulsed-dc dual magnetrons unless τ_{off} is set too high) resulting in larger energy dissipation per cycle.[304] Thus, there is less substrate heating during pulsed-dc dual magnetron sputtering and the film deposition rate is higher per unit power. Another advantage of dual-target pulsed-dc is the ability to independently control pulse widths and power to each target, which allows the controlled deposition of alloys[333] or films with compositional gradients. A major advantage of mid-frequency ac is that it is easier, hence less expensive, to scale power supplies to higher power for continuous, high-rate, very-large-area deposition[304] such as for "jumbo" (6x3.2 m^2) architectural glass.[234]

4.10.c. Modeling reactive sputtering

Another important recent accomplishment in sputter deposition is the evolutionary development of very useful models, with ever more completeness, of the highly complex reactive sputtering process. All such models are basically expressions of atom conservation. A mixture of rare and reactive gases, or simply pure reactive gas, enters the sputtering chamber and all reactive-gas species are captured at one of three places: the substrate (in these models, the "substrate" includes the chamber walls and fixtures) which acts as an internal getter pump, the target (the poisoning process), or the external vacuum pump. While the models contain simplifying assumptions in order to obtain closed-form solutions, their power is in allowing the investigator to pose "what if" questions before initiating experiments, thus greatly decreasing the number of iterations prior to achieving desired results. Example questions include: what happens if the reactive gas partial pressure is changed, the total pressure, the target voltage, the target-to-substrate distance, the metal/gas binding energy, the external pumping speed, etc.

Between 1968 and 1975, several reactive sputtering models were proposed to account for changes in target sputtering rates R_T, and hence film deposition rates, with the partial pressure P_r of reactive gases. In metal-oxygen systems, E. Hollands and David Campbell[334] from Plessey Ltd., UK, and later Johannes Heller,[335] IBM West Germany, qualitatively modeled the effects of changing the relative target oxidation rates and sputtering rates R_T leading to either abrupt or gradual decreases in R_T with the oxygen partial pressure P_{O_2}. Tetsuya Abe and Toshiro Yamashina,[336] Hokkaido University, Japan, proposed a kinetic model and used it to fit experimental data for the reactive sputtering of Mo and Ti in O_2 and N_2. However, the model neglected the dependencies of R_T on differences in target elemental sputtering yields, secondary electron yields, and discharge currents, all of which vary with changes in P_r. A modification of Abe and Yamashina's model, which included reactive-gas gettering by deposited metal atoms, was presented by Fujitoshi Shinoki and Akio ltoh[337] from the Electrotechnical Lab., Tokyo, Japan.

In 1980, a University of Illinois group[318] developed a reactive-sputtering model which included terms for reactive-gas sticking probabilities, sputtering rates of adsorbed reactive gases, secondary-electron yields, and ion fluxes incident at the target as a function of P_r. Calculated $R_T(P_r)$ results provided very good fits to experimental sputtering rates obtained with both a strongly-bonded metal/gas system, In/O_2, exhibiting an avalanche effect in deposition-rate reduction under flow control, and a weakly-bonded system, In/N_2, which did not undergo an uncontrolled decrease in deposition rate. Model fits allowed a determination of the target oxide and nitride surface coverages as well as nitrogen and oxygen sticking probabilities on the indium target.

In 1982, K. Steenbeck et al.[338] extended Shinoki and Itoh's model[337] to include the hysteresis effect by allowing the probability of reactive-gas gettering by deposited metal atoms to be a function of coverage. Two years later, John Affinito and Robert Parsons,[339] from the University of British Columbia, incorporated another mechanism for target poisoning, in addition to chemisorption: reactive-ion implantation (or trapping) in the near-surface region of the target during sputter erosion. They, like the Illinois group, chose materials systems for which chemisorption is either strong (Al/O_2) or weak (Al/N_2) and concluded based on both experiments and kinetic modeling that the drop in R_T with increasing P_r was primarily controlled by chemisorption in the former case and trapping in the latter.

Sören Berg and colleagues at Uppsala University made a very important step forward in 1987[340] when they expanded the model of the Illinois group to include both film growth (i.e., metal deposition and reactive-gas gettering at the substrate and vacuum-system walls) and external pumping, thus allowing them to reproduce and analyze a variety of experimental results for flow-controlled reactive sputtering. Berg's initial model, which treats target poisoning as being controlled purely by reactive-gas adsorption, is easy to use and quite powerful. It was expanded in 1998[341] to cover more complex experimental results including systems such as transition-metal nitrides for which multiple phases can coexist on the target surface and in the growing film (e.g., Cr, Cr_2N, and CrN for the reactive sputtering of Cr in Ar/N_2 mixtures), and the reactive sputtering of a metal in two different reactive gases. A review of these, and additional model results, was published by Berg and Tomas Nyberg in 2005[342] as a parallel to Bill Sproul's experimental review[304] in the same journal.

In the early 2000s, Diederik Depla and Roger De Gryse from Ghent University, Belgium, demonstrated in a series of papers on the reactive sputtering of Al in O_2,[343] Ag in N_2,[344] and Si in N_2,[345-347] that their rather detailed experimental results could not be completely explained by assuming that target poisoning occurred only by chemisorption at the target surface. It was necessary, as proposed earlier by Affinito and Parsons,[339] to include target poisoning via trapping of accelerated reactive-gas ions. A striking example of such effects is the fact that silver sputtered in mixed Ar/N_2 discharges reveals strong reactive-sputtering phenomena[344] even though there is no stable silver nitride. In (2004), Depla and De Gryse proposed a simple reactive-sputtering model which accounted for reactive-gas trapping.[345] Two years later, the Uppsala group, Tomas Kubart et al.,[348] introduced a third target-poisoning mechanism, recoil implantation of surface chemisorbed reactive-gas species by incident ions, and presented a model which included chemisorption, recoil implantation, and direct trapping. This solved a long-standing issue in the original Berg model associated with compound accumulation of up to several monolayers on the target, shown both experimentally[345,349] and theoretically,[350] which could not be explained by chemisorption alone since in many metal/gas systems, the process is self-limited to a single monolayer.

The Ghent group,[351] in 2007, published a dynamic model similar to that of Kubart et al. while also accounting for the chemical reaction kinetics of target atoms with recoiled-surface and directly-implanted reactive species. This provided an explanation of a variety of experimentally-observed effects which arise due to the binding energies of adsorbed surface reactive-gas species being quite different than that of species trapped below the surface. Depla and Stijn Mahieu edited a book on reactive sputter deposition, published in 2008,[352] and the Ghent group further refined their model in 2012.[353] Two years later, Koen Strijckmans and Depla presented a time-dependent and spatially-resolved reactive magnetron sputtering model[354] which included all previously discussed processes, plus the possibility of re-deposition back onto the target.[355] The same year, Berg and colleagues[356] from Uppsala further updated their mode, following the Ghent group,l to remove an initial simplifying assumption that material sputtered from poisoned regions of the target is ejected as compound molecules in order to include preferential sputtering[357,358] and allow recoiled knock-in reactive-gas atoms which do not react with target metal atoms to be released to the vacuum during sputter erosion.

Available reactive sputtering models, driven primarily by the Uppsala and Ghent University groups, are continuing to become more realistic, complete, and useful for analyzing complex results as well as for predicting parametric effects prior to carrying out experiments.

4.11. Roll-to-roll web coating

The 1930s witnessed the advent of roll-to-roll web coating in which flexible substrate materials such as textiles, plastic sheets, paper, or thin metal foil are unwound from a spool and continuously passed, via a winding system which maintains constant pressure (and sometimes substrate cooling), over the vapor source. The coated material is then rewound onto a take-up spool. The evolution of this technology is described in a series of papers published from 1997 through 2005.[359-362] The earliest roll-to-roll coaters employed sputter deposition, shifted to evaporation in the late 1930s and the 1940s in order to obtain higher deposition rates, and then back to sputtering in the 1980s following the commercial availability of the rectangular planar magnetron (Section 4.6).

The first industrial web-coating application was in 1934; systems were installed in London (UK) and in Fürth (Bavaria, Germany) to sputter-deposit gold on glassine (smooth, glossy paper, with oriented fibers, which is highly impervious to moisture) to create foil for hot stamping in specialty printing processes for the production of shiny decorative designs on textiles, wood, hard plastics, leather, and other materials. In initial operations, 400 m^2 of glassine, on a 1-m-wide roll, was coated in 23 h. This was equal to a week's production of gold leaf by 30 highly-skilled gold-beating craftsmen. In addition, the roll-to-roll product had a much more uniform thickness distribution.

An early industrial web coater (mid 1930s), designed by Bosch (Germany) for sputter metallizing paper to produce capacitors is illustrated in Figure 48. By the time production was initiated, the system incorporated thermal evaporation rather than sputter deposition to achieve higher deposition rates. In 1941, a dc-diode sputter web coater at the Hy-Sil Company in Boston was producing silver layers on cellophane for decorative applications.[361]

Immediately following commercial availability of the modern planar magnetron, the first magnetron sputter-deposited roll-to-roll commercial product, in the late 1970s, was indium tin oxide layers (ITO) on polyethylene terephthalate (PET) web[363] for use as a transparent-conducting electrode in x-ray imaging.[360-362] The first *large-scale* magnetron sputter roll coater was delivered to Southwall Technologies (Palo Alto, California) in 1980 for the production ITO/metal/ITO trilayers on PET web for infrared reflecting heat mirror applications.[360,361] The ITO layers were deposited by reactive sputtering and the metal layers by dc sputtering. The deposition system, produced by Leybold, contained five separate coating chambers with 1.828-m-long rectangular-shaped planar targets. A schematic illustration of a modern sputter roll-to-roll coating system is shown in Figure 49. In many applications, particularly for large-area, high-rate (high target power) coatings, the rectangular planar magnetrons in Figure 49 have been replaced by rotating tubular magnetrons (Section 4.6).[362]

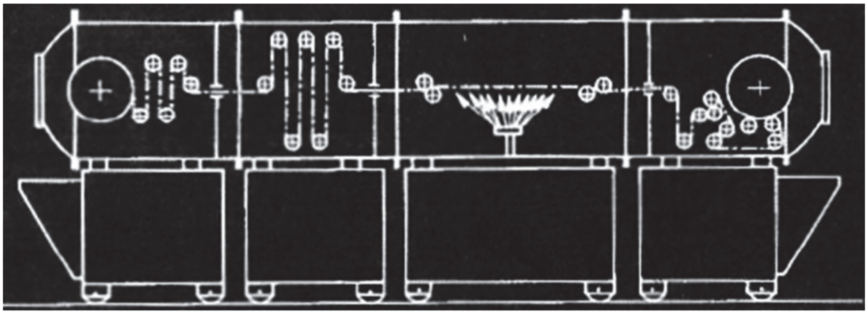


Figure 48. Illustration of a four-chamber, separately pumped, evaporative roll-to-roll web coating system developed by Bosch (Germany) in the mid-1930s to deposit zinc electrodes on paper capacitors. Adapted from Ref. 359.

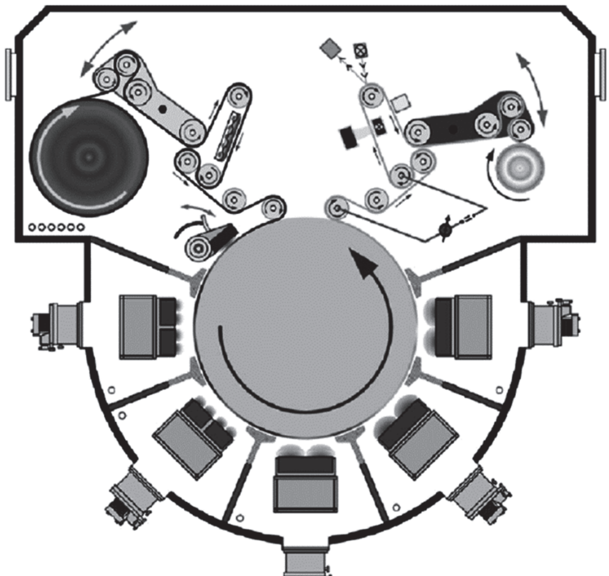


Figure 49. Cross-sectional view of a sputter roll-to-roll web-coating system containing five rectangular planar magnetrons, each in a separately-pumped chamber. The flexible substrate is fed from the spool in the upper left (shown with rollers and tensioners) onto the coating drum and then sequentially through the deposition chambers. The final product is wound onto the output spool in the upper right. Ref. 362.

5. Present developments and conclusions

In a little more than 160 years, sputter deposition has metamorphosed from a scientific curiosity to a platform for investigating solid-state physics via atomic-scale billiards combined with quantum mechanics (e.g., ion-bombardment-induced secondary electron emission) to a tool for mass-producing industrial thin-film products. Typing “sputter deposition” in a Google search brings up approximately a million hits and Google Scholar lists more than 200,000 journal articles on the subject. In addition, several books, including references 364 and 365, are focused on the use of sputter deposition in manufacturing.

Of the wide range of presently available thin film vapor-phase-deposition techniques, sputtering is the fastest growing and continues to find new applications. Much of this popularity stems from advantages such as the fact that it is a physical deposition technique (controlled by momentum transfer), hence it is relatively insensitive to deposition temperature, as opposed to chemical vapor deposition, and can be applied to thermally sensitive substrates; it is relatively easy to controllably deposit alloys compared to evaporation (sputter yields are similar for all metals, while evaporation rates at a given temperature vary over many orders of magnitude); and the use of low-energy ion (as well as fast-neutral atom) bombardment of the film growth surface to control nanostructure and composition, while minimizing surface roughness, is inherent in approaches such as unbalanced magnetron sputtering (Section 4.6).[36,243] These advantages (especially surface smoothening) explain, for example, why giant magnetoresistance devices are fabricated using sputtering.[366]

The combination of readily available ultra-high-vacuum systems, magnetically-unbalanced magnetrons, and partial-pressure-controlled reactive sputtering allows the growth of extremely high-quality single-crystal thin films for investigating fundamental materials properties. Examples are electron/phonon coupling and superconductivity in Group-IV transition-metal and rare-earth nitrides;[367] demonstrating that the tetragonal-to-cubic phase transition in VN, first reported almost 30 years ago but poorly understood, is due to strongly anharmonic phonon vibrations at temperatures above 250 K;[368] answering a controversy in the literature to clearly establish the mechanism of vacancy-induced hardening in understoichiometric transition-metal compounds such as TiN_x (with $x < 1.0$)[369] while providing the first complete set of elastic constants as a function of x;[370] and demonstrating that ceramic alloys such as $V_{0.5}Mo_{0.5}N_x$ not only display increased hardness with decreasing x, but also exhibit enhanced ductility (due to electronic-structure effects) – that is, the alloys exhibit excellent toughness (the combination of hardness and ductility).[371]

Today, magnetron sputtering remains a dynamic field; it not only continues to find new applications, but it also reincarnates itself in novel and useful manifestations. Two new forms of magnetron sputtering were created during the 1990s, both with the goal of efficiently ionizing sputter-ejected metal atoms. The first, ionized-magnetron sputtering, was developed in the early 1990s by the semiconductor industry in response to the fact that aspect ratios (depth-to-width) of interconnect vias and trenches, whose sides and bottom surface had to be coated with a uniformly-distributed thin layer, were increasing rapidly. The lack of directionality in sputtered atom fluxes, emitted from the target with an approximately cosine distribution (see Section 4.1) meant that at some size scale, the vias became clogged giving rise to "bird's-beak" defects as illustrated in Figure 50, with very poor coating uniformity.

An elegant solution, developed by Steve Rossnagel and Jeff Hopwood of IBM, was to ionize the sputtered metal flux and use an electric field to provide metal-atom directionality.[372,373] The ionization was accomplished by adding a two-turn water-cooled coil to inductively-couple rf power to the plasma between the target and substrate during operation of a standard parallel-plate dc magnetron (Figure 51). Since the ionization energies of the sputtering gases used in these experiments, 15.76 and 21.46 eV for Ar and Ne, were much larger than those of the metal atoms, 5.99 and 7.73 eV for Al and Cu, the sputtered metal atoms had a much higher ionization probability than the rare-gas atoms.

Energy and mass analyses were used to determine, as a function of Ar and Ne sputtering pressure, the fraction of sputtered Al atoms incident at the substrate position which was ionized. Initial results are plotted in the middle panel of Figure 51.[373] At high pressures, for which sputtered metal atoms have a longer residence time in the rf plasma region due to shorter mean free paths and higher collision frequencies, ionization fractions of 80 to 90% can be achieved with deposition rates that are approximately equal or somewhat larger, depending on operating conditions, than without the rf

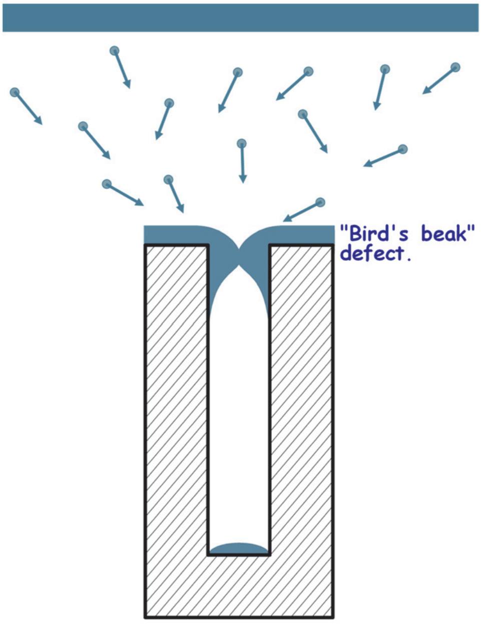


Figure 50. Illustration of the formation of a "bird's-beak" defect, formed while attempting to deposit a liner (or fill) coating in a via or trench, due to the lack of directionality in the sputtered-atom flux. Figure courtesy of Steve Rossnagel, IBM.

coil. Rossnagel, collaborating with a group from the University of Illinois,[374] also demonstrated that Al films deposited by ionized-magnetron sputtering have larger grain sizes; a higher degree of 111 preferred orientation, by more than an order of magnitude, with a much smaller azimuthal distribution, than with dc magnetron sputtering under otherwise identical deposition conditions; and low stress. Stronger 111 preferred orientation with minimal mosaic spread decreases the probability of electromigration failure in Al- and Cu-based metal interconnects.[375]

Beginning in the mid to late 1990s,[376-378] another variation of magnetron sputtering, which later became known as HIPIMS (high power impulse magnetron sputtering),[379,380] was developed in which very-high-power pulses are applied to the target (power densities of several kW/cm^2 and peak voltages > kV) to produce ultra-dense plasmas with electron densities two to four orders of magnitude higher than with standard dc magnetrons. Since the heat load deposited at the target surface during sputtering must still be carried away (typically by water cooling the back side of the target),

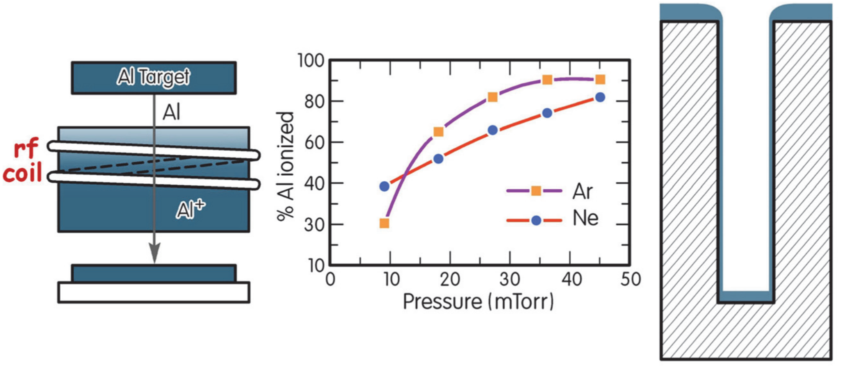


Figure 51. (Left) Schematic illustration of a dc magnetron with an inductively-coupled rf coil between the target and the substrate in order to post-ionize sputtered metal atoms. (Middle) Fraction of sputtered Al atoms ionized in Ar and Ne plasmas as a function of the sputtering pressure for constant magnetron power and 250 W rf induction. Adapted from Ref. 373. (Right) Typical metal liner coating deposited by ionized-metal sputtering. Figure courtesy of Steve Rossnagel, IBM.

the time-averaged power density is restricted to be essentially the same as in dc magnetrons; therefore power in HIPIMS systems is applied in short pulses (~50-200 μs) with low duty cycles (< 10%) and frequencies (10 Hz – 10 kHz).

The very high peak-target power in HIPIMS results in correspondingly high instantaneous sputtering rates such that sputtered-atom pulses rarefy the sputtering gas (via momentum and heat transfer) immediately in front of the target as the process moves toward self-sputtering. That is, target sputtering transitions from being initially dominated by rare-gas ions at the beginning of each pulse to sputtering by metal ions later in the pulse.[381] An important major advantage of this mode of operation is the extremely high degree of sputtered-metal-atom ionization, up to 70%[378] (compared to ~1% in dc magnetron sputtering).[382] This, in turn, means that low-energy ion bombardment of the growing film in order to control film microstructure and composition, while minimizing surface roughness, can be achieved without a major disadvantage often inherent in dc magnetron sputter deposition: trapping of rare-gas ions in the growing film, where they reside in interstitial sites and lead to compressive stress.[383,384] In HIPIMS mode, synchronizing the substrate bias to the metal-ion portion of the pulse results in irradiation of the growing film primarily by metal-ions which are incorporated in lattice sites, thus resulting in a significant decrease in film stress.[385] A disadvantage of HIPIMS for commercial applications is that the return of metal ions to the target, among other effects, decreases the film deposition rate.[386-388] However, the recent introduction of *hybrid* dc-magnetron/HIPIMS sputtering has been shown to provide, in addition to low stress, high deposition rates and fully-dense films at low substrate temperatures (no externally-applied heating).[389].

Another variant of HIPIMS, based upon a patent filed by Roman Chistyakov in 2002 and issued in 2006,[390] is modulated-pulse-power (MPP) magnetron sputtering. In MPP, the pulse widths are much longer, ~500-3000 μs, than in HIPIMS, but within each pulse, the on and off times can be independently varied from several to a few tens of μs to yield a higher duty cycle ($\lesssim$ 25%). Modulating the target voltage with a series of micropulses during each macropulse also provides more flexibility in tuning the peak current.[391,392] The concept of MPP is to provide additional control over the process, maintain the high metal ionization probability, and obtain higher deposition rates than with HIPIMS.[391-393]

It seems perfectly safe to predict that sputter-deposition of thin films will remain a vibrant and active field of human endeavor incorporating fundamental scientific research, process development, deposition-system design, and new-product manufacturing for periods far into the future.

6. The role of Ghent University today

Finally, returning full circle to Dr. George Sarton (Section 2), PhD in Physics and Mathematics from Ghent University and the person who developed the history of science as an academic discipline, it is interesting to note that over the past three decades, the Department of Solid State Sciences (Physics) at Ghent has become a well-known worldwide center of excellence in the field of sputter deposition. The following are just a few of many research areas in which Ghent University has displayed, and is continuing to demonstrate, a leadership role.

- Beginning in the1980s, Roger de Gryse, who was later joined by Diederik Depla, was intimately involved in the development of rotatable tubular magnetrons (see Section 4.6) which are extensively used in large-area, high rate, thin film manufacturing today.[241]

- Depla and de Gryse developed the first laboratory-scale rotatable magnetron and used it to carry-out fundamental investigations of target poisoning and hysteresis effects as a function of rotation rate.[394]
- Depla and de Gryse published a series of papers demonstrating, and unraveling, the extremely complex relationships among discharge voltage, target poisoning, and secondary-electron emission during reactive magnetron sputtering.[211,212,344]
- Depla and de Gryse developed, and distributed as freeware, powerful analytic and Monte Carlo models of non-reactive and reactive sputtering which relate experimental variables with outcomes and allow users to run simulations in order to reduce the number of experiments necessary to find a solution.[351,353,395]
- Depla and de Gryse extended the previously accepted structure-zone diagram[396] of thin-film microstructure evolution as a function of thickness and homologous deposition temperature, an important contribution to understanding thin film growth during sputter deposition.[397]

On a global level, the sputter-deposition research group at Ghent University has provided a wonderful vehicle to enable the worldwide community to interact, pollinate, and share ideas on sputtering and thin films: in 2002, they initiated an annual conference on Reactive Sputter Deposition (RSD). The conference has continued to grow in popularity and has become a focal point for researchers working in this field. While the conference site moves around the world, it is perhaps fitting that this year it is again back in Ghent for its 15th anniversary.

Acknowledgements

The author very gratefully acknowledges the critical reading of early drafts of this manuscript and the many useful suggestions of several close friends and excellent scientists: Diederik Depla, Ghent University, Belgium; Ivan Petrov, University of Illinois, USA, and Linköping University, Sweden; and Steve Rossnagel, IBM Yorktown Heights Research Center, New York.

References

[1] J.E. Greene, "Tracing the 5000-year recorded history of inorganic thin films from 3000 BC to the early 1900s AD," *Appl. Phys. Rev.* **1**, 041302 (2014).

[2] For the history of *inorganic* thin films, see: J.E. Greene, "Tracing the 4000-year Recorded History of Organic Thin Films: from Monolayers on Liquids to Multilayers on Solids," *Appl. Phys. Rev.* **2**, 011101 (2015) and (a much shorter version) J.E. Greene, "Organic Thin Films: From Monolayers on Liquids to Multilayers on Solids," Physics Today 67 (6), 43 (June, 2014).

[3] M. Lehner, *The Complete Pyramids: Solving the Ancient Mysteries* (Thames and Hudson, Ltd, London, 1997).

[4] R. F. Baker and C. Baker III, *Ancient Egyptians: People of the Pyramids* (Oxford University Press, London, 2001).

[5] H. Ades, *A Traveler's History of Egypt* (Interlink Books, Northampton, MA, 2007).

[6] E. D. Nicolson, "The ancient craft of gold beating," *Gold Bull.* **12**, 161 (1979).

[7] J. A. Harrell and V. Max Brown, "The world's oldest surviving geological map – the 1150 BC Turin papyrus from Egypt," *J. Geol.* **10**0, 3 (1992).

[8] J. A. Harrell and V. Max Brown, "The oldest surviving topographical map from ancient Egypt (Turin Papyri 1879, 1899 and 1969)," *J. Am. Research Center Egypt* **29**, 81 (1992).

[9] J. F. H. Notton, "Ancient Egyptian gold refining," *Gold Bull*. **7**, 50 (1974).

[10] L. B. Hunt, "The oldest metallurgical handbook," *Gold Bull.* **9**, 24 (1976).

[11] T. G. H. James, "Gold technology in ancient Egypt," *Gold Bull*. **5**, 38 (1972).

[12] E. Darque-Ceretti, E. Felder, and M. Aucouturier, "Foil and leaf gilding on cultural artifacts; forming and adhesion," *Revista Materia* **16**, 540 (2011).

[13] H. Lechtman, "Pre-Columbian surface metallurgy," *Sci. Am.* **25**0, 56 (1984).

[14] G. O. Mallory and J.B. Hajdu, Editors, *Electroless Plating: Fundamentals and Applications* (American Electroplaters and Surface Finishers Society, Orlando, Florida, 1990).

[15] J. Wolters, *Zur Geschichte Der Löttechnik: On the Origins of Gold Brazing* (Degussa, Frankfurt, Germany, 1976).

[16] N. H. Gale, "Metals and metallurgy in the chalcolithic period," Bull. Am. Schools Oriental Res. 282/283, 37 (1991).

[17] T. E. Emerson and D. L. McElrath, *Archaic Societies: Diversity and Complexity Across the Midcontinent* (State University of New York Press, Albany, New York, 2009).

[18] M. Radivojević, T. Rehren, E. Pernicka, D. Šljivar, M. Brauns, and D. Borić, "On the origins of extractive metallurgy: New evidence from Europe," *J. Archaeolog. Sci.* 37, 2775 (2010).

[19] B. Bower, "Dating the dawn of copper making," *Sci. News* **178**, 8 (2010).

[20] T. Murphy, *Pliny the Elder's Natural History: the Empire in the Encyclopedia* (Oxford University Press, Oxford, 2004).

[21] A.-F. de Fourcroy, *A General System of Chemical Knowledge, and its Application to the Phenomena of Nature and Art*, translated by W. Nicholson (W. Flint publisher, London, 1804), p. 84.

[22] J. W. Mellor, *A Comprehensive Treatise on Inorganic and Theoretical Chemistry* (Longman, London, 1952), Vol. 9.

[23] A. Sherman, *Chemical Vapor Deposition for Microelectronics* (Noyes Publications, Park Ridge, New Jersey, 1987).

[24] R. F. Rolsten, "The role of halides as preparative intermediates," *Ann. New York Acad. Sci.* 137, 162 (1966).

[25] George Sarton, *Introduction to the history of Science*, 3 Volumes in 5 books, (Published for the Carnegie Institution of Washington, Publication no. 376, by Williams and Wilkins Co., Baltimore (1927-1948).

[26] Eugene Garfield, "The life and career of George Sarton: The father of the history of science," *J. History Behavioral Science*s **21**, 107 (2006).

[27] W. R. Grove, "On the electro-chemical polarity of gases," Philos. Trans. R. Soc. London 142, 87 (1852).

[28] J. Stefan, "Über das Gleichgewicht und die Bewegung, Insbesondere die Diffusion von Gasgemengen (Equilibrium and diffusion in gases)," *Sitzungsber. Akad. Wiss. Wien, Math.-Naturwiss. Cl. II.* **63**, 63 (1871).

[29] J. Stefan, "Versuche über die Verdampfung," *Sitzungsber. Akad. Wiss. Wien, Math.-Naturwiss. Cl. II.* **68**, 385 (1873).

[30] J. Stefan, "Über die Verdampfung und die Auflösung als Vorgänge der Diffusion (On evaporation and the process of diffusion)," *Sitzungsber. Akad. Wiss. Wien, Math. Nat.-Naturwiss.* **98**, 1418 (1889).

[31] J. Wisniak, "Josef Stefan. radiation, conductivity, diffusion, and other phenomena," *Rev. CENIC, Cienc. Quim.* **37**, 188 (2006).

[32] H. Hertz, "Concerning the evaporation of liquids, particularly mercury into a vacuum (in German)," *Ann. Phys. (Leipzig)* **17**, 177 (1882).

[33] H. Hertz, *Wied. Ann.* **17**, 193 (1882).

[34] H. Hertz, *Electric waves: Being researches on the propagation of electric action with finite velocity through space* (Dover Publications, Mineola, New York, 1893).

[35] H. Hertz, "1882, Ueber die Beruehrung elastischer Koerper (On contact between elastic bodies)," in *Gesammelte Werke (Collected Works)* (Leipzig, Germany, 1895), Vol. 1.

[36] F. Adibi, I. Petrov, J.E. Greene, L Hultman, and J.-E. Sundgren, "Effects of high-flux low-energy (20-100 eV) ion irradiation during deposition on the microstructure and preferred orientation of $Ti_{0.5}Al_{0.5}N$ alloys grown by ultra-high-vacuum reactive magnetron sputtering," *J. Appl. Phys*. **73**, 8580 (1993).

[37] J.E. Greene, JE Sundgren, L. Hultman, I. Petrov, and D.B. Bergstrom, "Development of preferred orientation in polycrystalline TiN layers grown by ultrahigh vacuum reactive magnetron sputtering," *Appl. Phys. Lett.* **67**, 2928 (1995).

[38] B.W. Karr, I Petrov, DG Cahill, and JE Greene, "Morphology of epitaxial TiN(001) grown by magnetron sputtering," **70**, 1703 (1997).

[39] Online Etymological Dictionary, see http://www.etymonline.com/.

[40] G. K. Wehner and G. S. Anderson, “The nature of physical sputtering,” in Handbook of Thin Film Technology, Edited by L. I. Maissel and R. Glang (McGraw-Hill, New York, 1970).

[41] J. J. Thomson, *Rays of Positive Electricity* (Longmans, Green, and Company, London, 1913).

[42] D. M. Mattox, *Foundations of Vacuum Coating Technology* (Noyes Publications, Norwich, NY, 2003.

[43] *The Shorter Oxford English Dictionary on Historical Principles*, Third Edition Revised with Addenda, Edited by W. Little, H.W. Fowler, and J. Coulson, C.T. Onions (Oxford University Press, New York, 1955).

[44] K. H. Kingdon and I. Langmuir, “The removal of thorium from the surface of a thoriated tungsten filament by positive ion bombardment,” *Phys. Rev.* 22, 148 (1923).

[45] Research Staff of the General Electric Company, London, “Cathode Disintegration,” *Philos. Mag.* **45**, 98 (1923).

[46] T. S. Baynes, *The Encyclopaedia Britannica: A Dictionary of Arts, Sciences, Literature and General Information* (Encyclopaedia Britannica, New York, 1888), Vol. 19, p. 246.

[47] T. E. Conlon, *Thinking About Nothing: Otto von Guericke and the Magdeburg Experiments on the Vacuum* (The Saint Austin Press, London, 2011).

[48] J. Campbell, *Speed Cleaning* (Random House Publishing, New York, 1991).

[49] M. H. Hablanian, “Comments on the history of vacuum pumps,” *J. Vac. Sci. Technol. A* 2, 118 (1984).

[50] H. Sprengel, “III Researches on the vacuum,” *J. Chem. Soc.* **18**, 9 (1865).

[51] C. J. Cleveland, “Geissler, Johann Heinrich Wilhelm,” *Encyclopedia of Earth, Environmental Information Coalition* (National Council for Science and the Environment, Washington, D.C., 2008).

[52] S. P. Thompson, *The Development of the Mercurial Air-pump* (E. and F. N. Spon, London, 1888).

[53] H. McLeod, “Apparatus for measurement of low pressures of gas,” *Philos. Mag.* 48, 110 (1874).

[54] A. Sella, “Classic kit: McLeod’s vacuum gauge,” *Chem. World* **8**, 60 (2011).

[55] Robert Boyle, *New Experiments Physico-Mechanical, Touching the Spring of the Air, and Its Effects* (H. Hall Publisher, Oxford, 1660).

[56] E. Thomas and R. Leyniers, “The McLeod gauge (1874-1974) – its birth – its development – its actuality,” *Jpn. J. Appl. Phys*., Part 1 **13**, 147 (1974).

[57] *McGraw-Hill Concise Encyclopedia of Physics* (McGraw-Hill Publishing, New York, 2002).

[58] M. B. Schiffer, *Bringing the Lightning Down: Benjamin Franklin and Electrical Technology in the Age of Enlightenment* (University of California Press, Berkeley, CA, 2003).

[59] J. Zeitler, “Guerickes Weltkräfte und die Schwefelkugel: Guericke’s world forces and the rotating sulfur ball,” *Monumenta Guerickiana* 20/21, 147 (2011).

[60] Thales was born in Miletus, a Greek Ionian city on the western coast of what is today the Aydin Province of Turkey.

[61] J. Burnet, *Early Greek Philosophy* (A. and C. Black Publishers, London, 1892).

[62] G. S. Kirk, J. E. Raven, and M. Schofiel, *The Presocratic Philosophers* (Cambridge University Press, Cambridge, 1983).

[63] B. Russell, *A History of Western Philosophy* (Simon and Schuster, New York, 1945).

[64] J. Priestley, *The History and Present State of Electricity* (J. Dodsley, J. Johnson, and T. Cadell, London, 1767).

[65] E. J. Houston, *Electricity in Every-day Life* (P. F. Collier and Son, New York, 1905), Vol. I, p. 72.

[66] F. A. Meyer, *Duisb. Forsch.* **5**, 31 (1961).

[67] J. L. Heilbron, *Electricity in the 17th and 18th Centuries* (Dover Press, New York, 1999).

[68] W. Maver, "Electricity, its history and progress," in *The Encyclopedia Americana* (Encyclopedia Americana Corp., New York, 1918), Vol. X, p. 172.

[69] A. Volta, "On the electricity excited by the mere contact of conducting substances of different kinds," *Philos. Trans. R. Soc. London* **90**, 403 (1800) [Title in English, text in French].

[70] A. Volta, "On the electricity excited by the mere contact of conducting substances of different kinds," *Philos. Mag.* **7**, 289 (1800).

[71] G. Pancaldi, *Volta: Science and Culture in the Age of Enlightenment* (Princeton University Press, Princeton, NJ, 2003).

[72] H. Kragh, "Confusion and controversy: Nineteenth-century theories of the voltaic pile," in *Nuova Voltania: Studies on Volta and His Times*, edited by F. Bevilacqua and L. Fregonese (Ulrico Hoepli, Milano, Italy, 2000), p. 135.

[73] A. Volta. "Del modo di render sensibilissima la più debole electtricità sia naturale, sia artificiale: How to render very sensitive the weakest electricity; both natural and artificial," *Philos Trans. R. Soc.* **72**, 237 (1782)

[74] W. Cruickshank, "Additional remarks on galvanic electricity," *J. Nat. Philos. Chem. Arts* **4**, 254 (1801).

[75] M. Faraday, "VIII. Experimental researches in electricity. – Thirteenth series," *Philos. Trans. R. Soc. London* **128**, 125 (1838).

[76] The first four Bakerian Lectures presented by Faraday were entitled: "On the manufacture of Glass for Optical Purposes" (1829); "Experimental Researches in Electricity: Second Series" (1832); "Experimental Researches in Electricity: Twenty-Second Series," (1849); and "Experimental Researches in Electricity: Twenty-Fourth Series," (1851).

[77] M. Faraday, "The Bakerian Lecture: Experimental relations of gold (and other metals) to light," *Philos. Trans.* **147**, 145 (1857).

[78] G. L'E. Turner, "Henry Baker, F.R.S.: Founder of the Bakerian Lecture," *Notes Rec. R. Soc. Lond.* **29**, 53 (1974).

[79] M. H. Hart, *The 100: A Ranking of the Most Influential Persons in History* (Citadel Press, New York, 2000).

[80] C. Russell, *Michael Faraday: Physics and Faith* (Oxford University Press, New York, 2000).

[81] H. S. Norrie, *Ruhmkorff Induction-Coils, Their Construction, 0peration and Application* (New York, New York: Spon & Chamberlain, 1896).

[82] D. Depla, S. Mahieu, and J. E. Greene, "Sputter deposition processes," in *Handbook of Deposition Technologies for Thin Films and Coatings*, 3rd ed., Edited by P. M. Martin (William Andrew Publications (Elsevier), Burlington, MA, 2010).

[83] *CRC Handbook of Chemistry and Physics*, 88th ed., Edited by D. R. Lide (CRC, Boca Raton, Florida 2007).

[84] K. Tominaga, N. Ueshiba, Y. Shintani, and O. Tada, "High-energy neutral atoms in the sputtering of ZnO," Japn. *J. Appl. Phys*. 20, 519 (1981).

[85] K. Tominaga, N. Ueshiba, Y. Shintani, and O. Tada, "Energy analysis of high-energy neutral atoms in the sputtering of ZnO and BaTiO3," *Japn. J. Appl. Phys*. 21, 588 (1982).

[86] I. Petrov, V. Orlinov, and A. Misiuk, "Highly oriented ZnO films obtained by d.c. reactive sputtering of a zinc target," *Thin Solid Film*s 120, 55 (1984).

[87] S. M. Rossnagel and J. J. Cuomo, "Negative ion effects during magnetron and ion beam sputtering of $YBaCu_2O_x$," *AIP Conf. Proc.* **165**, 106 (1988).

[88] S. Ishibashi, Y. Higuchi, Y. Ota, and K. Nakamura, "Low resistivity indium-tin oxide transparent conductive films. II. Effect of sputtering voltage on electrical property of films," *J. Vac. Sci. Technol. A* **8**, 1403 (1990).

[89] T. Wenzell and K. Ellmer, "Negative oxygen ion formation in reactive magnetron sputtering processes for transparent conductive oxides" *J. Vac. Sci. Technol. A* **30**, 061306 (2012).

[90] J. Jia, A. Yoshimura, Y. Kagoya, N. Oka, and Y. Shigesato, "Transparent conductive Al and Ga doped ZnO films deposited using off-axis sputtering," *Thin Solid Films* **559**, 69 (2014).

[91] P.G. Quigley, R.A. Rao, and C.B. Eom, "Time dependence and spatial distribution of the deposition rate of YBa2Cu3O7 thin films in 90° off-axis sputtering," *J. Vac. Sci. Technol. A* 15, 2854 (1997).

[92] Y. Hoshi, H. Kato, and K. Funatsu, "Structure and electrical properties of ITO thin films deposited at high rate by facing target sputtering," *Thin Solid Films* 445, 245 (2003).

[93] C. B. Eom, J. Z. Sun, K. Yamamoto, A. F. Marshall, K. E. Luther, S. S. Laderman, and T. H. Geballe, "In situ grown YBa2Cu3O7−d thin films from single-target magnetron sputtering," *Appl. Phys. Lett.* 55, 595 (1989).

[94] C. B. Eom, J. Z. Sun, B. M. Lairson, S. K. Streiffer, A. F. Marshall, K. Yamamoto, S. M. Anlage, J. C. Bravman, T. H. Geballe, S. S. Laderman, R. C. Taber, and R.D. Jacowitz, "Synthesis and properties of YBa2Cu3O7 thin films grown in situ by 90° off-axis single magnetron sputtering," *Physica C* 171, 354 (1990).

[95] J. P. Gassiot, The Bakerian Lecture: "On the stratifications and dark bands in electrical discharges as observed in Torricellian vacuums," *Proc.. R. Soc. London* **9**, 145 (1857).

[96] J. P. Gassiot, "The Bakerian Lecture – on the stratifications and dark bands in electrical discharges as observed in Torricellian vacuums," *Philos. Trans. R. Soc. London* **148**, 1 (1858).

[97] J. C. Maxwell, “A dynamical theory of the electromagnetic field” *Philos. Trans. R. Soc. London* **155**, 459 (1865). Presented to the Royal Society on December 8, 1854.

[98] A. W. Wright, “On the production of transparent metallic films by the electrical discharge in exhausted tubes,” *Am. J. Sci. Arts* **13**, 49 (1877).

[99] A. W. Wright, “On a new process for the electrical deposition of metals, and for constructing metal-covered glass specula,” *Am. J. Sci. Arts* **14**, 169 (1877).

[100] Office of Public Affairs at Yale; News Release at the Wayback Machine (archived September 9, 2006).

[101] History of the Wright Nuclear Structure Laboratory at Yale University: http://wlab.yale.edu/about-us/history.

[102] D.M. Mattox, “Sputter deposition; early efforts,” *Vac. Technol. Coat*. 2, 34 (July, 2001).

[103] “Crookes tube,” *The Columbia Electronic Encyclopedi*a, 6th Ed. (Columbia Univ. Press, 2007).

[104] J. Plücker, “Continued observation on the electrical discharge through rarefied gas tubes,” *Philos. Mag.* **16**, 408 (1858)

[105] W. Hittorf, “Ueber die Elektricitätsleitung der Gase: About the electrical conduction of gases,” *Annalen der Physik und Chemie* **136**,, 197 (1869).

[106] W. Röntgen, “On a new kind of rays,” *Nature* **53**, 274 (1896).

[107] J. J. Thomson, “Cathode rays,” Lecture presented at the Royal Institution on April 30, 1897, The Royal Institution Library of Science. Physical Sciences, Edited by W. L. Bragg and G. Porter (Applied Science, London, 1970), Vol. 5, pp. 36-49; reprinted in *The Electrician* **39**, 104 (1897).

[108] J. D. Cobine, “Gaseous Conductors: Theory and Engineering Applications,” (General Publishing Co. Ltd, Toronto, Ontario, 1941)

[109] Gerald F.J. Tyne, *Saga of the Vacuum Tube* (Howard W. Sams & Co., Indianapolis, Indiana, 1977).

[110] W. Crookes, “On electrical evaporation,” *Proc. R. Soc. London* **50**, 88 (1891).

[111] W. Crookes, “The Bakerian Lecture: On the illumination of lines of molecular pressure, and the trajectory of molecules,” *Philos. Trans. R. Soc. London* **170**, 135 (1879).

[112] William Crookes, ‘‘Molecular Physics in High Vacua,’’ The Royal Institution Library of Science. Physical Sciences, Vol. 3, Edited by William L. Bragg and George Porter (Applied Science, London, 1970, pp. 38-59).

[113] M.L. Tarng and G.K. Wehner, “Alloy sputtering studies with in situ auger electron spectroscopy,” *J. Appl. Phys.* **42**, 2449 (1971).

[114] A. H. Eltoukhy and J.E. Greene, “Diffusion enhancement due to low-energy ion bombardment during sputter etching and deposition,” *J. Appl. Phys.* **51**, 4444 (1980).

[115] T. A. Edison, “Process of coating phonograph records,” US patent 713,863, filed June 16, 1900; issued November 18, 1902.

[116] T. A. Edison, “Process of duplicating phonograms,” US patent 484,582, filed January 5, 1888; issued October 18, 1892.

[117] T. A. Edison, “Electric lamp,” US patent 223,898, filed November 4, 1879; issued January 27, 1880.

[118] J. M. Thomas, "W. R. Grove and the fuel cell," *Philos. Mag.* **92**, 3757 (2012).

[119] W. R. Grove, "On the gas voltaic battery," *Philos. Trans. R. Soc. London* 135, 351 (1845).

[120] VV. W. R. Grove, *On the Correlation of Physical Forces*, 6th ed. (Longmans, Green and Co, London, 1874).

[121] R. Friedel and P. Israel, *Edison's Electric Light: Biography of an Invention* (Rutgers University Press, New Brunswick, NJ, 1986).

[122] H. F. Fruth, "Cathode sputtering, a commercial application," *Physics* **2**, 280 (1932).

[123] L. Maxwell, "The production of an intense beam of hydrogen positive ions," *Rev. Sci. Instrum.* **2**, 129 (1931).

[124] R. Seeliger and K. Sommermeyer, "Bemerkung zur Theorie der Kathodenzerstäubung: Theory of Cathode Sputtering, "*Z. Phys.* **93**, 692 (1935).

[125] G. K. Wehner and D. Rosenberg, "Angular distribution of sputtered materials," *J. Appl. Phys.* **31**, 177 (1960).

[126] Y. Matsuda, Y. Yamamura, Y. Ueda, K. Uchino, K. Muraoka, M. Maeda, and M. Akazaki, "Energy dependence of angular distributions of sputtered particles by ion-beam bombardment at normal incidence," *Jpn. J. Appl. Phys*. **25**, 8 (1986).

[127] C. H. Townes, "Theory of cathode sputtering in low voltage gaseous discharges," *Phys. Rev.* **6**5, 319 (1944).

[128] E. Kay, "Impact evaporation and thin film growth in a glow discharge," *Adv. Electron.* **1**7, 245 (1962).

[129] A. Güntherschulze, "Cathodic sputtering an analysis of the physical processes," *Vacuum* **3**, 360 (1953).

[130] E. Goldstein, *Verh. Dtsch. Phys. Ges.* **4**, 228 (1902).

[131] E. Goldstein, *Verh. Dtsch. Phys. Ges.* **4**, 237 (1902).

[132] J. Stark, "Über die zerstäubende Wirkung der Kanalstrahlen (Emission sekundärer Atomstrahlen): About the sputtering effect of atom beams (emission of secondary nuclear radiation)," *Z. Electrochem.* **14**, 752 (1908).

[133] J. Stark, "Zur Theorie der Zerstäubung durch Atomstrahlen:(Bemerkungen zu einer Abhandlung des Herrn Kohlschütter: To the theory of sputtering by atom beams (Comments on a treatise of Mr. Kohlschütter)," *Z. Electrochem.* **15**, 509 (1909).

[134] G. Granqvist, "Öfver katodens sönderdelning i förtunnade gaser sid; Upon cathode reduction in rarefied gases," Öfvers. Svenska Vet. Akad Förh 54, 575 (1897).

[135] G. Granqvist, Öfvers. Svenska Vet. *Akad Förh* 55, 709 (1898).

[136] J. Stark and G. Wendt, "Über das Eindringen von Kanalstrahlen in feste Körper; About the entry of channel radiation in solids," *Ann. d Physik* **38**, 921 (1912).

[137] H. Mayer, "On the radiation of metal surfaces bombarded by positive alkali ions," *Philos. Mag.* **16**, 594 (1933).

[138] H. Sporn, "Luminous zones near glow discharge cathodes," *Z. Physik* **112**, 278 (1939).

[139] A. Güntherschulze, "Neue Untersuchungen über die Kathodenzerstäubung: New studies on sputtering" *Z. Physik* **119**, 79 (1942).

[140] H. Fetz, "Über die Kathodenzerstäubung bei schiefem Aufparall der Ionen: About sputtering with oblique ion collisions," *Z. Phys.* **119**, 590 (1942).

[141] G. K. Wehner, "Momentum transfer in sputtering by ion bombardment," *J. Appl. Phys.* **25**, 270 (1954).

[142] G. K. Wehner, "Sputtering of metal single crystals by ion bombardment," *J. Appl. Phys.* **26**, 1056 (1955).

[143] G. S. Anderson and G. K. Wehner, "Atom ejection patterns in single-crystal sputtering," *J. Appl. Phys.* **31**, 2305 (1960).

[144] G. S. Anderson, "Atom ejection in low energy sputtering of single crystals of fcc metals and of Ge and Si," *J. Appl. Phys.* **33**, 2017 (1962).

[145] G. S. Anderson, "Atom ejection in low energy sputtering of single crystals of bcc metals," *J. Appl. Phys.* **34**, 659 (1963).

[146] R.S. Nelson, "Focused collision sequences in tungsten and molybdenum," *Philos. Mag.* **8**, 693 (1963).

[147] P.P. Hasigati, R. Hanada, and S. Yamaguchi, *J. Phys. Soc. Japan* **18** (suppl. III), 164 (1963).

[148] C.B. Cooper and J. Comas, "Ejection Patterns in Low-Energy Sputtering of GaAs and GaP Crystals," *J. Appl. Phys.* **38**, 2891 (1965).

[149] R. H. Silsbee, "Focusing in collision problems in solids," *J. Appl. Phys.* **28**, 1246 (1957).

[150] D.E. Harrison, J.P. Johnson, and N.S. Levy, "Spot patterns and Silsbee chains on a Cu single crystal," *Appl. Phys. Lett.* 8, 33 (1966).

[151] C. Lehman and P. Sigmund, "On the mechanism of sputtering," *Phys. Stat. Sol.* **16**, 507 (1966).

[152] R. V. Stuart, and G. K. Wehner, "Energy distribution of sputtered Cu atoms," *J. Appl. Phys.* **35**, 1819 (1964).

[153] R. V. Stuart and G. K. Wehner, and G. S. Anderson, "Energy distribution of atoms sputtered from polycrystalline metals," *J. Appl. Phys.* **40**, 803 (1969).

[154] E. Blechschmidt, "Kathodenzerstäubungsprobleme. (II. Mitteilung) Die Kathodenzerstäubung in Abhängigkeit von den Betriebsbedingungen; Sputtering problems. (II. Communication) sputtering as a function of the operating conditions," *Ann. Phys.* **81**, 999 (1926).

[155] B. Chapman, *Glow Discharge Processes* (Wiley Interscience, New York, 1980).

[156] D. Depla, G. Buyle, J. Haemers, and R. DeGryse, "Discharge voltage measurements during magnetron sputtering," *Surf. Coat. Technol.* **200**, 4329 (2006).

[157] A. von Hippel, "Theory of Cathode Scattering," *Ann. Physik* **81**, 1043 (1926).

[158] W. D. Davis and T.A. Vanderslice, Ion energies at the cathode of a glow discharge," *Phys. Rev.* **131**, 219 (1963).

[159] A. V. Phelps, "Cross sections and swarm coefficients for nitrogen ions and neutrals in N_2 and Argon ions and neutrals in Ar for energies from 0.1 eV to 10 keV," *J. Phys. Chem. Ref. Data* **20**, 557 (1991).

[160] J. Plücker, "Ueber die Einwirkung von Magneten auf die elektrischen Entladungen in verdünnten Gasen; On the effect of magnets on the electric discharge in dilute gases," *Ann. der Phys.* **103,** 88 (1858).

[161] E. Goldstein, "Vorläufige Mittheilungen über elektrische Entladungen in verdünnten Gasen; Preliminary communications on electric discharges in rarefied gases," Monatsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin (Monthly Reports of the Royal Prussian Academy of Science in Berlin), May 4, pp 279-295 (1878).

[162] W. Crookes, "On Radiant Matter," Lecture delivered before the British Association for the Advancement of Science, at Sheffield, Friday, 22 August 1879. Published in *The Popular Science Monthly* **16**, 14 (1880); **16**, 157 (1880) [published in two parts].

[163] P. Zeeman, "Scientific Worthies: Sir William Crookes, F.R.S," *Nature* **77**, 1 (1907).

[164] J. F. Mulligan, "The personal and professional interactions of J. J. Thomson and Arthur Schuster," *Am. J. Phys.* **65**, 954 (1997)

[165] A. Shuster, *The Progress of Physics* (Cambridge University Press, Cambridge, 1911).

[166] Arthur Schuster, "The Bakerian Lecture: Experiments on the discharge of electricity through gases. Sketch of a theory," *Proc. R. Soc. London* **37**, 317-341 (1884).

[167] Jean Perrin, "New Experiments on the Kathode Rays," *Nature* **53**, 298 (1896).

[168] Arthur Schuster, "The Bakerian Lecture: The discharge of electricity through gases. Preliminary communication," *Proc. R. Soc. London* **47**, 526 (1890).

[169] W. Kaufmann, "Nachtrag zu der Abhandlung: 'Die magnetische Ablenkbarkeit der Kathodenstrahlen'; Addendum to the essay: 'The magnetic deflection of the cathode rays,'" *Ann. Phys.* **298**, 596 (Issue 12, 1897). The article Kaufmann referred to is W. Kaufmann and E. Aschkinass, "Die magnetische Ablenkbarkeit der Kathoden-strahlen," *Ann. Phys.* **297**, 596 (1897)."

[170] W. Kaufmann, "Über die Konstitution des Elektrons; Of the constitution of the elec-tron," *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* **45**, 949 (2005).

[171] E. Wiechert, "Uber das Wesen der Elektrizität: On the nature of electricity," *Schriften Phys.-Ökonom Gesellsch. Königsberg* **38**, 1 and **38**, 12 (1897).

[172] J. J. Thomson, "Cathode Rays," *Philos. Mag.* **44**, 293 (1897).

[173] H. Kragh, "J. J. Thomson, the electron, and atomic architecture," *The Physics Teacher* **25**, 328 (1997).

[174] *Nobel Lectures. Physics, 1901-1921* (Elsevier, Amsterdam, 1967), pp. 145-153.

[175] George Gerald Stoney, "The tension of metallic films deposited by electrolysis," *Proc. Royal Soc. London, Series A*, **82**, 172 (1909).

[176] G. J. Stoney, "Of the 'electron,' or atom of electricity," *Philos. Mag.* **38**. 418 (1894).

[177] G. J. Stoney, "On the cause of double lines and of equidistant satellites in the spectra of gases" *Trans. Roy. Dublin Soc.* **4**, 563 (1891).

[178] H. D. Hagstrum, "Theory of auger ejection of electrons from metals by ions," *Phys. Rev.* **96**, 336 (1954).

[179] H. D. Hagstrum, "Auger ejection of electrons from tungsten by noble gas ions," *Phys. Rev.* **104**, 325 (1956).

[180] R. A. Baragiola, E. V. Alonso, J. Ferron and A. Oliva-Florio, "Ion-induced electron emission from clean metals," *Surf. Sci.* **90**, 240 (1979).

[181] J. Schou, "Transport theory for kinetic emission of secondary electrons from solids," *Phys. Rev. B* **22**, 2141 (1980).

[182] H. D. Hagstrum, "Auger ejection of electrons from molybdenum by noble gas ions," *Phys. Rev.* **104**, 672 (1956).

[183] R. C. Monreal, "Auger neutralization and ionization processes for charge exchange between slow noble gas atoms and solid surfaces," *Prog. Surf. Sci.* **89**, 80 (2014).

[184] A. Güntherschulze and K. Meyer, "Kathodenzerstäubung bei sehr geringen Gasdrücken; Sputtering at low gas pressures," *Z. Phys.* **62**, 607 (1930).

[185] J. A. Thornton and J. E. Greene, "Sputter deposition processes," in *Deposition Technologies for Films and Coatings*, Edited by R. Bunshah (Noyes Publications, Park Ridge, NJ, 1994).

[186] A. Güntherschulze and K. Meyer, "Kathodenzerstäubung in Quecksilberdampf bei sehr geringen Drucken; Cathode sputtering in mercury vapor at low pressures," *Z. Phys.* **71**, 279 (1931).

[187] G. K. Wehner, "Controlled Sputtering of Metals by Low-Energy Hg Ions," *Phys. Rev.* **102**, 690 (1956).

[188] H. Fetz, "Über die Beeinflussung eines Quecksilbervakuumbogens mit einem Steuergitter im Plasma; About the influence of a mercury vacuum arc with a control grid in plasma," *Ann. Physik* **37**, 1 (1940).

[189] G. K. Wehner, "Sputtering yields for normally incident Hg^+-ion bombardment at low ion energy," *Phys. Rev.* **108**, 35 (1957).

[190] R. V. Stuart and G. K. Wehner, "Sputtering yields at very low bombarding ion energies," *J. Appl. Phys.* **33**, 2345 (1962).

[191] H.M. Mott-Smith and I. Langmuir, "The theory of collectors in gaseous discharges," *Phys. Rev.* **28**, 727 (1926).

[192] Kovar is a nickel-cobalt ferrous alloy, used to form a glass-metal seal, which was designed to be compatible with the thermal expansion characteristics of borosilicate glass and to wet the molten glass in order to allow direct mechanical connections over a range of temperatures. An early version of a multilayer metal wire (Ni/Cu/Pt) suitable for forming a glass/metal seal in light bulbs was patented in 1913 (UC Patent 1083070 B, filed October 26, 1911) by B. E. Eldred, New York. C.G. Fink developed Dumet (dual metal) seals, replacing the expensive Pt. Dumet consists of a Ni/Fe alloy core, which matches the thermal expansion coefficient of soda-lime glass, coated with oxidized Cu which bonds with the glass. Fink filed for the patent January 23, 1915, and it was issued to the General Electric Company in June, 1924 (US Patent 1,498, 908). The earliest metal used to form a glass/metal seal was liquid mercury in the first barometer, invented by Evangelista Torricelli (1608-1647), an Italian physicist and mathematician, in ~1640. Toricelli used the barometer to measure atmospheric pressure as discussed by J.E. Greene in ref. 1.

[193] N. Laegreid and G. K. Wehner, "Sputtering Yields of Metals for Ar^+ and Ne^+ Ions with Energies from 50 to 600 eV," *J. Appl. Phys.* 3**2**, 365 (1961).

[194] N. Laegreid, G. K. Wehner, and B. Meckel, "Sputtering Yield of Germanium in Rare Gases," *J. Appl. Phys.* **30**, 374 (1959).

[195] The thermionic cathode was obtained from a thyratron tube used as a high-power electrical switch and controlled rectifier. While the cathode was not described, at that time they were commonly constructed of a high-melting-point metal coated with a

low-work-function alkaline-earth oxide paste, typically barium oxide, strontium oxide, or an alloy of the two.

[196] J. A. Thornton and J. E. Greene, "Sputter Deposition Processes," in Deposition Technologies for Films and Coatings, Edited by R. Bunshah (Noyes Publications, Park Ridge, NJ, 1994). Data in Figure 20 primarily from Table 1 in Ref. 157.

[197] H. D. Hagstrum, "Instrumentation and experimental procedure for studies of electron ejection by ions and ionization by electron impact," *Rev. Sci. Instr.* **24**, 1122 (1953).

[198] H. D. Hagstrum, "Effect of monolayer adsorption on the ejection of electrons from metals by ions," *Phys. Rev.* **104**, 1516 (1956).

[199] H. D. Hagstrum and C. D'Amico, "Production and demonstration of atomically clean metal surfaces" *J. Appl. Phys.* **31**, 715 (1960).

[200] O. Almén and G. Bruce, "Collection and sputtering experiments with noble gas ions," *Nucl. Instr. Methods* **11**, 257 (1961).

[201] O. Almén and G. Bruce, "Sputtering experiments in the very high energy region," *Nucl. Instr. Methods* **11**, 279 (1961).

[202] F. Grønlund and W. J. Moore, "Sputtering silver by light ions with energies from 2 to 12 keV," *J. Chem. Phys.* **12**, 1540 (1960).

[203] N. Matsunami, Y. Yamamura, Y. Itikawa, N. Itoh, Y. Kazumata, S. Miyagawa, K. Morita, R. Shimizu, and H. Tawara, "Energy dependence of the ion-induced sputtering yields of monatomic solids," *Atom. Data Nucl. Data Tables* **31**, 1 (1984).

[204] A. W. Hull, "The magnetron," *J. Am. Inst. Elec. Engrs.* **40**, 715 (1921).

[205] A. W. Hull, "The Effect of a uniform magnetic field on the motion of electrons between coaxial cylinders," *Phys. Rev.* **18**, 31 (1921).

[206] F. M. Penning, "Coating by cathode disintegration," US Patent 2.146.025, filed November 7, 1936; issued February 7, 1939.

[207] F. M. Penning, "Die Glimmentladung bei Niedrigem Druck Zwischen Koaxialen Zylindern in einem Axialen Magnetfeld (The Glow Discharge at Low Pressure between Coaxial Cylinders in an Axial Magnetic Field)]," *Physica* **3**, 873 (1936).

[208] F.M. Penning and J. H. A. Moubis, "Cathode Sputtering in a Magnetic Field," *Proc. K. Ned. Akad. Wet.* **43**, 41 (1940).

[209] L. D. Hall, "Ionic vacuum pumps," *Science* **128**, 279 (1958).

[210] Y. Hoshi, H. Kato, and K. Funatsu, "Structure and electrical properties of ITO thin films deposited at high rate by facing target sputtering," *Thin Solid Films* **445**, 245 (1997).

[211] D. Depla, S. Heirwegh, S. Mahieu, J. Haemers, and R. De Gryse, "Understanding the discharge voltage behavior during reactive sputtering of oxides," *J. Appl. Phys.* 101, 013301 (2007).

[212] D. Depla, S. Mahieu, and R. De Gryse, "Magnetron sputter deposition: linking discharge voltage with target properties," *Thin Solid Films* 517, 2825 (2009).

[213] A. Güntherschulze and H. Betz, "Vergleich der Kathodenzerstäubung reiner und oxydbedeckter Magnesiumoberflächen: Comparison of the sputtering of pure magnesium oxide-covered surfaces," *Z. Phys.* **106**, 365 (1937).

[214] W. D. Gill and E. Kay, "Efficient low pressure sputtering in a large inverted magnetron suitable for film synthesis," *Rev. Sci. Instr.* **36**, 277 (1965).

[215] J.A. Thornton, "Influence of Apparatus Geometry and Deposition Condition on the Structure and Topography of Thick Sputtered Coatings," *J. Vac. Sci. Technol. A* 11, 666 (1974).

[216] J.A. Thornton, "High Rate Thick Film Growth," *Ann. Rev. Mater. Sci.* 7, 239 (1977).

[217] J. A. Thornton and A. Penfield, "Cylindrical magnetron sputtering," *Thin Film Processes*, Edited by J. L. Vossen and W. Kern (Academic press, New Yok, 1978).

[218] G.N. Rokhlin, *J. Phys. USSR* 1, 347 (1939).

[219] E. Kay, "Magnetic field effects on an abnormal truncated glow discharge and their relation to sputtered thin-film growth," *J. Appl. Phys*. **34**, 760 (1963).

[220] J. R. Mullaly, "Crossed-field discharge device for high rate sputtering," Dow Chemical Rocky Flats Division, RFP 1310, OSTI 4735036, January 1, 1969 (DOI 10.2172/4735036).

[221] W. Knauer, "Ionic vacuum pump," US Patent 3,216,652, filed on September 10, 1962 and issued November 9, 1965.

[222] K. Wasa and S. Hayakawa, "Low-pressure sputtering system of the magnetron type," *Rev. Sci. Instrum.* **40**, 693 (1969).]

[223] J. S. Chapin, "Sputtering process and apparatus," US Patent 4,166,018, filed on January 31, 1974 and issued August 28, 1979.

[224] J. S. Chapin, "Planar Magnetron," *Vac. Technol.* **40**, 37 (January, 1974).

[225] R. K. Waits, Planar magnetron sputtering," *Thin Film Processes*, Edited by J. L. Vossen and W. Kern (Academic press, New Yok, 1978).

[226] W. de Bosscher and H. Lievens, "Advances in magnetron sputter sources," *Thin Solid Films* **351**, 15 (1999).

[227] R. E. Demaray, J. C. Helmer, R. L. Anderson, Y. H. Park, R. R. Cochran, and V. E. Hoffman, "Rotating sputtering apparatus for selected erosoion," US patent 5,252,194, filed July 23, 1992; issued October 12, 1993.

[228] K. F. Lai, H. Tang, K. Song, and D.B. Hayden, "Rotating magnet arrays for magnetron sputtering apparatus," US patent 7,585,399 B1, filed March 31, 2005; issued September 8, 2009.

[229] Steve Rossnagel, IBM Reaearch Laboratories, Yorktown Heights, New York, private communication.

[230] T. Gotoa, T. Matsuoka, and T. Ohmi, "Rotation magnet sputtering: Damage-free novel magnetron sputtering using rotating helical magnet with very high target utilization," *J. Vac. Sci. Technol. A* **27**, 653 (2009).

[231] T. Iseki, "Completely flat erosion magnetron sputtering using a rotating asymmetrical yoke magnet," *Vac.* **84**, 1372 (2010).

[232] P. J. Clarke, "Sputtering apparatus and method," US patent 4,060,470, filed Dec. 19, 1975; issued Nov. 29, 1977.

[233] D. B. Foster, "The sputter and s-gun magnetrons," in *Thin Film Processes*, Edited by J. L. Vossen and W. Kern (Academic Press, New York, 1978).

[234] J. A. Ridealgh, "Large Area Coatings for Architectural Glass – Design, Manufacturing, Process Faults," *MRS Proceedings*, Volume 890, Edited by S. J. Bull, P. R. Chalker, S. C. Chen, R. Maboudian, and W. J. Meng (Cambridge University Press, Cambridge, UK, 2005), p. 39.

[235] H. E. McKelvey, “Magnetron Cathode Sputtering Apparatus,” US Patent 4356073, filed February. 12, 1981 and issued October 26, 1982.

[236] H. E. McKelvey, “Cathodic Sputtering Apparatus,” U.S. Patent 4443318, filed August 17, 1983 and issued April 17, 1984.

[237] H. E. McKelvey, “Rotatable Sputtering Apparatus,” U.S. Patent 4445997; filed Aug. 17, 1983 and issued May 1, 1984.

[238] Forschungsinstitut Manfred Von Ardenne, “Einrichtung zum Hochratezerstäuben nach dem Plasmatronprinzip; Device for high rate sputtering following the plasmatron [magnetron sputtering] principle,” DDR [East Germany] Patent DD 217964 A3, filed Oct. 2, 1981 and issued in 1982.

[239] G. Robinson, “Rotatable Magnetron Sputtering System,” *Proceedings of the 26th Annual Technical Conference of the Society of Vacuum Coaters*, 1983, p. 121.

[240] M. Wright and T. Beardow, “Design advances and applications of the rotatable cylindrical magnetron,” *J. Vac. Sci. Technol. A* 4, 388 (1986).

[241] R. De Gryse, J. Haemers, W.P. Leroy, and D. Depla, “Thirty years of rotatable magnetrons,” *Thin Solid Films* **520**, 5833 (2012).

[242] A. Segers, D. Depla, K. Eufinger, J. Haemers and R. De Gryse; Proc. 45th Technical Conference of the Society of Vacuum Coaters (2002), p. 30.

[243] I. Petrov, F. Adibi, J.E. Greene, W.D. Sproul, and W.-D. Munz, “Use of an externally applied axial magnetic field to control ion/neutral flux ratios incident at the substrate during magnetron sputter deposition,” *J. Vac. Sci. Technol. A* **10**, 3283 (1992).

[244] I. Petrov, P. B. Barna, L. Hultman, and J. E. Greene, “Microstructural evolution during film growth,” *J. Vac. Sci. Technol.* **21**, S117 (2003).

[245] B. Window and N. Savvides, “Charged particle fluxes from planar magnetron sputtering sources,” *J. Vac. Sci. Technol. A* **4**, 196 (1986).

[246] V. Kohlschütter, “Die zerstäubung durch kanalstrahlen; Atomization by canal rays,” *Jahrbuch Radioaktivität und Elektronik* **9**, 355 (1912).

[247] A. W. Hull, “Gas-filled thermionic tubes,” *Trans. Am. Inst. Elec. Engrs.* **47**, 753 (1928).

[248] A. W. Hull and W. F. Winter, “Gas-filled thermionic tubes; The volt-ampere characteristics of electron tubes with thoriated tungsten filaments containing low pressure inert gas,” *Phys. Rev.* **21**, 211 (1923).

[249] H. Lueder, “Zerstäubung von Metallen durch Aufprall langsamer Ionen und Messung des Schwellenwertes der Zerstäubung,” *Z. Physik* **97**, 158 (1935).

[250] R. C. Bradley, “Sputtering of alkali atoms by inert gas ions of low energy,” *Phys. Rev.* **93**, 719 (1954).

[251] N. D. Morgulis and V. D. Tischenko, “The investigation of cathode sputtering in the near threshold region. 1.”, *J. Exptl. Theoret. Phys. USSR* **30**, 54 (1956).

[252] S. P. Wolsky, “Positive-Ion Bombardment of germanium and silicon,” *Phys. Rev.* **108**, 1131 (1957).

[253] G. K. Wehner and G. Medicus, “Sputtering at low ion energies,”*J. Appl. Phys.* **25**, 698 (1954).

[254] D. McKeown, “New method for measuring sputtering in the region near threshold,” *Rev. Sci. Instr.* **32**, 133 (1961).

[255] D. McKeown, A. Y. Cabezas, and H. E. Adelson, *Bull. Am. Phys. Soc.* **6**, 399 (1961).

[256] R. V. Stuart and G. K. Wehner, "Sputtering thresholds and displacement energies," *Phys. Rev. Lett.* **4**, 409 (1960).

[257] G. J. Dienes and G. H. Vineyard, *Radiation Effects in Solids* (Interscience Publishers Inc., New York, 1957).

[258] R. Behrisch, "Festkörperzerstäubung durch Ionenbeschuß: Solid state atomization by ion bombardment," *Ergeb. Exakt. Naturw.* **35**, 295 (1964).

[259] M. Kaminsky, *Atomic and Ionic Impact Phenomena* (Springer Verlag, Berlin, 1965).

[260] G. Carter and J.S. Colligan, *Ion Bombardment of Solids* (American Elsevier, New York, 1968).

[261] D. E. Harrison, "Theory of the sputtering process," *Phys. Rev.* **102**, 1473 (1956).

[262] D. E. Harrison, "Supplementary sputtering calculations," *Phys. Rev.* **105**, 1202 (1957).

[263] D. E. Harrison, "Extended theory of sputtering," *J. Chem. Phys.* 32, 1336 (1960).

[264] P. K. Rol, J.M. Fluit, and J. Kistemaker, "Theoretical aspects of cathode sputtering in the energy range of 5-25 keV," *Physica* **26**, 1009 (1960).

[265] M. W. Thompson, "II. The energy spectrum of ejected atoms during the high energy sputtering of gold," *Philos. Mag.* **18**, 377 (1968).

[266] H.H. Anderson and H.L. Bay, "Sputtering yield measurements," in *Sputtering by Particle Bombardment I*, Edited. by R. Behrish (Springer-Verlag, Berlin, 1981).

[267] P. Sigmund, Phys. Rev. "Theory of sputtering. I. Sputtering yield of amorphous and polycrystalline targets," *Phys. Rev.* **184**, 383 (1969).

[268] P. Sigmund, Phys. Rev. "Errata: Theory of sputtering. I. Sputtering yield of amorphous and polycrystalline targets," *Phys. Rev.* **18**7, 768 (1969).

[269] E. S. Lamar and K. T. Compton, "A special theory of cathode sputtering," *Science* **80**, 541 (1934).

[270] J. Bohdansky, J. Roth, and H. L. Bay, "An analytical formula and important parameters for low-energy ion sputtering," *J. Appl. Phys.* 51, 2861 (1980).

[271] Y. Yamamura, N. Matsunami, and N. Itoh, "Theoretical studies on an empirical formula for sputtering yield at normal incidence," *Rad. Eff.* **71**, 65 (1983).

[272] N. Matsunami, Y. Yamamura, Y. Itikawa, N. Itoh, Y. Kazumata, S. Miyagawa, K. Morita, R. Shimizu, and H. Tawara, *Energy Dependence of the Yields of Ion-Induced Sputtering of Monatomic Solids* (Inst. Plasma Physics, Nagoya Univ., Japan, 1983).

[273] N. Matsunami, Y. Yamamura, Y. Itikawa, N. Itoh, Y. Kazumata, S. Miyagawa, K. Morita, R. Shimizu, and H. Tawara, "Energy dependence of the ion-induced sputtering yields of monatomic solids," *Atom. Data Nucl. Data Tables* **31**, 1 (1984).

[274] J. Bodansky, A universal relation for the sputtering yield of monatomic solids at normal ion incidence," *J. Nucl. Instrum. Methods. B* 2, 587 (1984).

[275] I. Petrov, V. Orlinov, and S. Grudeva, "On the energy efficiency of sputtering," *Bulg. Acad. Phys.* 18, 214 (1991).

[276] H. S. Butler and G. S. Kino, "Plasma sheath formation by radio-frequency fields," *Phys. Fluids* **6**, 1346 (1963).

[277] G. S. Anderson, W. N. Mayer, and G. K. Wehner, "Sputtering of dielectrics by high-frequency fields," *J. Appl. Phys.* **33**, 2991 (1962).

[278] Technically, rf frequencies extend to 300 GHz: however, it is more common today to label frequencies from 1 to 300 GHz as being in the microwave range.

[279] Documents from the International Radio Conference (Atlantic City, 1947), Documents 1-100. International Telecommunication Union Library and Archives; http://www.itu.int/dms_pub/itu-s/oth/02/01/S020100002B4813PDFE.pdf

[280] J. J. Thomson, "On the discharge of electricity through exhausted tubes without electrodes," *Philos. Mag.* **32** (197), 321 (1891).

[281] D. Banerji and R. Ganguli, "On deposits of metallic mercury by high-frequency discharge," *Philos. Mag.* 15 (99), 676 (1933).

[282] J. K. Robertson and C. W. Clapp, "Removal of metallic deposits by high-frequency currents," *Nature* 132, 479 (1933).

[283] R. H. Hay, "The removal of wall deposits by high frequency discharges," *Can. J. Res.* A 16, 191 (1938).

[284] G. K. Wehner, "Sputtering by Ion Bombardment," *Advances in Electronics and Electron Physics*, Volume 7, Edited by L. Martin, (Academic Press, New York 1955).

[285] J. E. Jackson and J. A. Kane, "Measurement of ionospheric electron densities using an RF probe technique," *J. Geophys. Res.* 64, 1074 (1959).

[286] H. S. Butler, "Plasma sheath formation by rf fields," *Microwave Laboratory Report 820*, Stanford University (April, 1961).

[287] B. Chapman, *Glow Discharge Processes* (Wiley Interscience, New York, 1980).

[288] Pieter D. Davidse and Leon I. Maissel, "Apparatus for cathode sputtering including a shielded rf electrode," US Patent 3,369,991, filed January 28, 1961; issued February 20, 1968 assigned to IBM Corporation.

[289] P. D. Davidse and L. I. Maissel, "Dielectric thin films through rf sputtering," *J. Appl. Phys.* **37**, 574 (1966).

[290] P. D. Davidse, "Theory and practice of rf sputtering," *Vac.* **17**, 139 (1966).

[291] L.D. Hartsough and P.S. McLeod, "High rate Al_2O_3 sputtering of enhanced aluminum mirrors," *J. Vac. Sci. Technol.* **14**, 123 (1977).

[292] R. S. Nowicki, "Properties of rf sputtered films deposited by planar magnetron," *J. Vac. Sci. Technol.* **14**, 127 (1977).

[293] G. Este and W. D. Westwood, "A quasi-direct-current sputtering technique for the deposition of dielectrics at enhanced rates," *J. Vac. Sci. Technol A* **6**, 1845 (1988).

[294] A. Furuya and S. Hirono, "Target magnetic-field effects on deposition rate in rf magnetron sputtering," *J. Appl. Phys.* **68**, 304 (1990).

[295] G. Y. Yeom, J. A. Thornton, and M. J. Kushner,"Cylindrical magnetron discharges. I. Current-voltage characteristics for dc-and rf-driven discharge sources," *J. Appl. Phys.* **65**, 3816 (1989).

[296] K. Bädeker, "Über die elektrische leitfähigkeit und die thermoelektrische kraft einiger schwermetallverbindungen (Electrical conductivity and thermo-electromotive force of some metallic compounds)," *Ann. Phys. (Leipzig)* **22**, 749 (1907).

[297] K. Ellmer, "Past achievements and future challenges in the development of optically transparent electrodes," *Nature Photonics* **6**, 809 (2012).

[298] C. J. Overbeck, "Color in films of sputtered tin," *J. Opt. Soc. Am.* **23**, 109 (1933).

[299] G. A. Veszi, "The modern single-layer selenium photo-electric cell," *J. Br. Instit. Radio. Engrs.* **13**, 183 (1953).

[300] L. Holland and G. Siddall, "Reactive sputtering and plant design," *Vac.* **111**, 245 (1953).

[301] L. Holland and G. Siddall, "The properties of some reactively sputtered metal oxide films," *Vac.* **111**, 375 (1953).

[302] A. Anders, "Physics of arcing, and implications to sputter deposition," *Thin Solid Films* **502**, 22 (2006).

[303] A. Anders, *Cathodic Arcs: From Fractal Spots to Energetic Condensation* (Springer, NY, 2008).

[304] W.D. Sproul, D.J. Christie, and D.C. Carter, "Control of Reactive Sputtering Processes," *Thin Solid Films* **491**, 1 (2005).

[305] J. O'Brien and P. J. Kelly, "Characterization studies of the pulsed dual-cathode magnetron sputtering process for oxide films," *Surf. Coat. Technol.* **142**, 621 (2001).

[306] J. S. Preston, "Constitution and mechanism of the selenium rectifier photocell," *Proc. Roy. Soc. A* **202**, 449 (1950).

[307] G. Helwig, "Elektrische leitfähigkeit und struktur aufgestäubter kadmiumoxd-schichten; Electrical conductivity and structure of cadmium oxide layers," *Z. Phys.* **132**, 622 (1952).

[308] W. D. Westwood, *Sputter Deposition* (AVS, NY, 2003).

[309] D. Gerstenberg and C. J. Calbick, "Effects of nitrogen, methane, and oxygen on structure and electrical properties of thin tantalum films, *J. Appl. Phys.* **35**, 402 (1964).

[310] E. Krikorian and R. J. Sneed, "Deposition of tantalum, tantalum oxide, and tantalum nitride with controlled electrical characteristics," *J. Appl. Phys.* **37**, 3674 (1966).

[311] C.-S. Shin, Y.-W. Kim, D. Gall, J.E. Greene, and I. Petrov, "Phase composition and microstructure of polycrystalline and epitaxial TaN_x layers grown on oxidized Si(001) and MgO(001) by reactive magnetron sputter deposition," *Thin Solid Film*s **402**, 172 (2002).

[312] A. W. Adamson, *Physical Chemistry of Surfaces* (John Wiley & Sons, New York, 1982).

[313] L. I. Maissel and P. M. Schaible, Thin films deposited by bias sputtering," *J. Appl. Phys.* **36**, 237 (1965).

[314] S.A. Barnett, G. Bajor, and J.E. Greene, "Growth of high-quality epitaxial GaAs films by sputter deposition," *Appl. Phys. Lett.* **37**, 734 (1980).

[315] J.E. Greene, S.A. Barnett, K.C. Cadien, and M.A. Ray, "Growth of single crystal GaAs and metastable $(GaSb)_{1-x}Ge_x$ alloys by sputter deposition: Ion-surface interaction effects," *J. Cryst. Growth* **56**, 389 (1982).

[316] H. Barankova, S. Berg, P. Carlsson, C. Nender, "Hysteresis effects in the sputtering process using two reactive gases," *Thin Solid Films* **260**, 181 (1995).

[317] W. D. Sproul, "High rate reactive sputtering process control," *Surf. Coat. Technol.* **33**, 73 (1987).

[318] B.R. Natarajan, A.H. Eltoukhy, J.E. Greene, and T.L. Barr, "Mechanisms of the reactive sputtering of In III: A general phenomenological model for reactive sputtering," *Thin Solid Films* **69**, 229 (1980); also **69**, 201 and **69**, 217 (1980).

[319] T. Serikawa and A. Okamoto, "Effect of N_2/Ar mixing on the reactive sputtering characteristics of silicon," *Thin Solid Films* 101, 1 (1983).

[320] D. Depla, S. Mahieu, and R. DeGryse, Chap. 5, "Reactive Sputter Deposition," Edited by D. Depla and S. Mahieu, Springer, Berlin (Germany), 2008.

[321] E. Särhammar, T. Nyberg, and S. Berg, "Applying "the upgraded Berg model" to predict hysteresis free reactive sputtering," *Surf. Coat. Technol.* 229, 34 (2016).

[322] W. D. Sproul and J. R. Tomashek, "Rapid rate reactive sputtering of a group IVB metal," US patent 4,428,811, filed April 4, 1983; issued January 31, 1984.

[323] W.D. Sproul, P.J. Rudnik, C.A. Gogol, and R.A. Mueller, "Advances in Partial Pressure Control Applied to Reactive Sputtering," *Surf. Coat. Technol.* **39/40**, 499 (1989).

[324] R. L. Cormia, T. A. Trumbly, and S. Andresen, US Patent 4,046,659 A, "Method for Coating a Substrate," filed January 10, 1975; issued September 6, 1977.

[325] F. S. Quazi, "Method and apparatus for sputtering a dielectric target or for reactive sputtering," US Patent 4,693,805, filed February 14, 1986; issued September 15, 1987.

[326] S. Schiller, K. Goedicke, J. Reschke, V. Kirchoff, S. Schneider, and F. Milde, "Pulsed magnetron sputter technology," *Surf. Coat. Technol.* **61**, 331 (1993).

[327] P. Frach, U. Heissig, C. Gotfried, and H. Walde, "Aspects and results of long-term stable deposition of Al_2O_3 with high rate pulsed reactive magnetron sputtering," *Surf. Coat. Technol.* **59**, 177 (1993).

[328] M. Scherer, J. Schmitt, R. Latz, and M. Schanz, "Reactive alternating current magnetron sputtering of dielectric layers," *J. Vac. Sci. Technol. A* **10**, 1772 (1992).

[329] D.C. Carter, W.D. Sproul, and D.J. Christie, "Effective Control for Reactive Sputtering Processes," *Vac. Technol. Coat.*, April, (2006), p. 60.

[330] D.A. Glocker, "Influence of the plasma on substrate heating during low-frequency reactive sputtering of AIN," J. Vac. Sci. Technol. A 11, 2989 (1993).

[331] G. Wallace, "Design considerations for the ac/dc C-Mag deposition source and power supply system," Thin Solid Films 351, 21 (1999).

[332] C. May and J. Strümpful, "ITO coatings by reactive magnetron sputtering – comparison of properties from dc and mf [mid-frequency] processing," *Thin Solid Films* **351**, 48 (1999).

[333] R. Laird and A. Belkind, "Cosputtered films of mixed TiO_2/SiO_2," *J. Vac. Sci. Technol. A* **10**, 1908 (1992).

[334] E. Hollands and D.S. Campbell, "The mechanism of reactive sputtering," *J. Mater. Sci.* **3** 544 (1968).

[335] J. Heller, "Reactive sputtering of metals in oxidizing atmospheres," *Thin Solid Films* **17** 163 (I973).

[336] T. Abe and T. Yamashina, "The deposition rate of metallic thin films in the reactive sputtering process," *Thin Solid Films* **30** 19 (1975).

[337] F. Shinoki and A. Itoh, "Mechanism of rf reactive sputtering," *J. Appl. Phys.* **46** 3381 (1975).

[338] K. Steenbeck, E. Steinbeiss, and K. D. Ufert, "The problem of reactive sputtering and cosputtering of elemental targets, *Thin Solid Films* 92, 371 (1982).

[339] J. Affinito and R. R. Parsons, "Mechanisms of voltage controlled, reactive, planar magnetron sputtering of Al in Ar/N_2 and Ar/O_2 atmospheres," *J. Vac. Sci. Technol. A* **2**, 1275 (1984).

[340] S. Berg, H.-O. Blom, T. Larsson, and C. Nender, "Modeling of reactive sputtering of compound materials," *J. Vac. Sci. Technol. A* **5**, 202 (1987).

[341] S. Berg, T. Nyberg, H.-O. Blom, and C. Nender, "Computer Modeling as a Tool to Predict Deposition Rate and Film Composition in the Reactive Sputtering Process," *J. Vac. Sci. Technol. A* **16**, 1277 (1998).

[342] S. Berg and T. Nyberg, "Fundamental Understanding and Modeling of Reactive Sputtering Processes," *Thin Solid Films* **476**, 215 (2005).

[343] D. Depla and R. De Gryse, "Influence of oxygen addition on the target voltage during reactive sputtering of aluminum," *Plasma Sources Sci. Technol.* 10, 547 (2001).

[344] D. Depla and R. De Gryse, "Target voltage measurements during DC sputtering of silver in a nitrogen/argon plasma," *Vac.* **69**, 529 (2003).

[345] D. Depla and R. De Gryse, "Target poisoning during reactive magnetron sputtering: Part I: the influence of ion implantation," *Surf. Coating Technol*. **183**, 184 (2004).

[346] D. Depla and R. De Gryse, "Target poisoning during reactive magnetron sputtering: Part II: the influence of chemisorption and gettering," *Surf. Coating Technol*. **183**, 190 (2004).

[347] D. Depla and R. De Gryse, "Target poisoning during reactive magnetron sputtering: Part III: the influence of the critical reactive gas mole fraction," *Surf. Coating Technol*. **183**, 196 (2004).

[348] T. Kubart, O. Kappertz, T. Nyberg, and S. Berg, "Dynamic behavior of the reactive sputtering process," *Thin Solid Films* **515**, 421 (2006).

[349] D. Güttler, B. Abendroth, R. Grötzschel, W. Möller, and D. Depla, "Mechanisms of target poisoning during magnetron sputtering as investigated by real-time in situ analysis and collisional computer simulation," *Appl. Phys. Lett.* **85**, 6134 (2004).

[350] D. Rosen, I. Katardjlev, S. Berg, and W. Möller, "TRIDYN simulation of target poisoning in reactive sputtering," *Nucl. Instrum. Methods Phys. Res. B* **228**, 193 (2005).

[351] D. Depla, S. Heirwegh, S. Mahieu, and R. De Gryse, "Toward a More Complete Model for Reactive Magnetron Sputtering," *J. Phys. D: Appl. Phys.* 40, 1957 (2007).

[352] *Reactive Sputter Deposition*, Edited by D. Depla and S. Mahieu, (Springer, Berlin, Germany, 2008).

[353] K. Strijckmans, W. P. Leroy, R. De Gryse, D. Depla, "Modeling reactive magnetron sputtering: Fixing the parameter set," *Surf. Coat. Technol.* **206**, 3666 (2012).

[354] K Strijckmans and D Depla, "A time-dependent model for reactive sputter deposition," *J. Phys. D: Appl. Phys.* **47**, 235302 (2014).

[355] D. Depla, X. Y. Li, S. Mahieu, K. Van Aeken, W. P. Leroy, J. Haemers, R. De Gryse, and A Bogaerts, *J. Appl. Phys.* **107**, 113307 (2010).

[356] S. Berg, E. Särhammar, and T. Nyberg, "Upgrading the 'Berg-model' for Reactive Sputtering Processes," *Thin Solid Films* **565**, 186 (2014).

[357] S. Berg and I. V. Katardjiev, "Preferential sputtering effects in thin film processing," *J. Vac. Sci. Technol. A* **17**, 1916 (1999).

[358] M P Seah and T S Nunney, "Sputtering yields of compounds using argon ions," *J. Phys. D: Appl. Phys.* **43**, 253001 (2010).

[359] .E. O. Dietrich, R. Ludwig, and E. K. Hartwig, "Vacuum web coating – an old technology with a high potential for the future," *Proc. Soc. Vac. Coaters* 1997, p. 354.

[360] R.L. Cormia, J. B. Fenn, H. Memarian, and G. Ringer, "Roll-to-roll coating of indium tin oxide – a status report," *Proc. Soc. Vac. Coaters* 1998, p. 452.

[361] J. B. Fenn, W. C. Kittler, D. Lievens, R. Ludwig, G. Phillips, and A. Taylor, "Roll-to roll sputter web coater – a status report," *Proc. Soc. Vac. Coaters* 1998, p. 463.

[362] R. Ludwig, R. Kukla, and E. Josephson, "Vacuum web coating – state of the art and potential for electronics," *Proc. IEEE* **93**, 1483 (2005).

[363] W. C. Kittler, and I. T. Ritchie, "Continuous coating of indium tin oxide onto large area flexible substrates," *Proc. SPIE* **325**, 61 (1982).

[364] C. A. Bishop, *Vacuum Deposition onto Webs, Films and Foils*, 2nd Edition, (Elsevier, Boston, MA, 2011).

[365] R. A. Powell and S. M. Rossnagel, *PVD for microelectronics: Sputter Deposition applied to semiconductor manufacturing* (Academic Press, New York, NY, 1999).

[366] J. J. Quan, X. W. Zhou, and H. N. G. Wadley, "Low energy ion assisted atomic assembly of metallic superlattices," *Surf. Sci.* **600**, 2275 (2006).

[367] A. B. Mei, A. Rockett, L. Hultman, I. Petrov, J. E. Greene, "Electron/phonon coupling in group-IV transition-metal and rare-earth nitrides," *J. Appl. Phys.* **114**, 193708 (2013).

[368] A.B. Mei, O. Hellman, N. Wireklint, C.M. Schlepütz, D.G. Sangiovanni, B. Alling, A. Rockett, L. Hultman, I. Petrov, and J. E. Greene, "Dynamic and structural stability of cubic vanadium nitride," *Phys. Rev. B* **91**, 054101 (2015).

[369] C.-S. Shin, D. Gall, N. Hellgren, J. Patscheider, I. Petrov, and J.E. Greene, "Vacancy hardening in single-crystal TiN_x(001) layers," *J. Appl. Phys.* **93**, 6024 (2003).

[370] T. Lee, K. Ohmori, C.-S. Shin, D.G. Cahill, I. Petrov, and J.E. Greene, "'Elastic constants of single-crystal TiN_x(001) (0.67 ? x ? 1.0) determined as a function of x by picosecond ultrasonic measurements," *Phys. Rev. B* **71**, 144106 (2005).

[371] H. Kindlund, D.G. Sangiovanni, J. Lu, J. Jensen, V. Chirita, J. Birch, I. Petrov, J.E. Greene, "Vacancy-induced toughening in hard single-crystal $V_{0.5}Mo_{0.5}N_x$/MgO(001) thin films," L. Hultman, *Acta Mater.* 77, 394 (2014).

[372] S.M. Rossnagel and J. Hopwood, "Magnetron sputter deposition with high levels of metal ionization," *Appl. Phys. Lett.* **63**, 3285 (1993).

[373] S.M. Rossnagel and J. Hopwood, "Metal ion deposition from ionized magnetron sputtering discharge," *J. Vac. Sci. Technol. B* **12**, 449 (1994).

[374] Y.-W. Kim, J. Moser, I. Petrov, J. E. Greene, and S. M. Rossnagel, "Directed sputter deposition of AlCu: Film microstructure and microchemistry," *J. Vac. Sci. Technol. A* **12**, 3169 (1994).

[375] Y.-W. Kim, I. Petrov, J. E. Greene, and S. M. Rossnagel, "Development of 111 texture in Al films grown on SiO_2/Si(001) by ultra-high-vacuum primary-ion deposition," *J. Vac. Sci. Technol. A* **14**, 346 (1996).

[376] D. V. Mozgrin, I. K. Fetisov, G. V. Khodachenko, “High-current low-pressure quasi-stationary discharge in a magnetic field: Experimental research,” *Plasma Phys. Rep.* **21**, 400 (1995).

[377] I. K. Fetisov, A. A. Filippov, G. V. Khodachenko, D. V. Mozgrin, and A. A. Pisarev, “Impulse irradiation plasma technology for film deposition,” *Vac*. **53**, 133 (1999).

[378] V. Kouznetsov, K. Macak, J. M. Schneider, U. Helmersson, I. Petrov, “A novel pulsed magnetron sputter technique utilizing very high target power densities,” *Surf. Coat. Technol.* **122**, 290 (1999).

[379] U. Helmersson, M. Lattemann, J. Bohlmark, A. P. Ehiasarian, and J. T. Gudmundsson, “Ionized physical vapor deposition (IPVD): A review of technology and applications,” *Thin Solid Films* **513**, 1 (2006).

[380] K. Sarakinos, J. Alami, and S. Konstantinidis, “High power pulsed magnetron sputtering: A review on scientific and engineering state of the art,” *Surf. Coat. Technol.* **204**, 1661 (2010).

[381] K. Macák, V. Kouznetsov, J. Schneider, U. Helmersson, and I. Petrov, “Ionized sputter deposition using an extremely high plasma density pulsed magnetron discharge,” *J. Vac. Science Technol. A*, **18**, 1533 (2000).

[382] I. Petrov, A. Myers, J.E. Greene, and J.R. Abelson, “Mass and Energy Resolved Detection of Ions and Neutral Sputtered Species Incident at the Substrate during Reactive Magnetron Sputtering of Ti in Mixed Ar+N_2 Mixtures,” *J. Vac. Sci. Technol. A* **12**, 2846 (1994).

[383] I. Petrov, L. Hultman, U. Helmersson, J. E. Sundgren, and J. E. Greene, “Microstructure Modification of TiN by Ion Bombardment during Reactive Sputter Deposition,” *Thin Solid Films* **169**, 299 (1989).

[384] I. Petrov, L. Hultman, J. E. Sundgren, and J. E. Greene, “Polycrystalline TiN Films Deposited by Reactive Bias Magnetron Sputtering: Effects of Ion Bombardment on Resputtering Rates, Film Composition, and Microstructure,” *J. Vac. Sci. Technol. A* **10**, 265 (1992).

[385] G. Greczynski, J. Lu, J. Jensen, I. Petrov, J. E. Greene, S. Bolz, W. Kolker, Ch. Schiffers, O. Lemmer, and L. Hultman, “Metal versus rare-gas ion irradiation during $Ti_{1-x}Al_x$ N film growth by hybrid high power pulsed magnetron/dc magnetron co-sputtering using synchronized pulsed substrate bias,” *J. Vac. Sci. Technol. A* **30**, 061504 (2012).

[386] D. J. Christie, “Target material pathways model for high power pulsed magnetron sputtering,” *J. Vac. Sci. Technol. A* **23**, 330 (2005).

[387] J. Vicek and K. Burcalová, “A phenomenological equilibrium model applicable to high-power pulsed magnetron sputtering,” *Plasma Sources Sci. Technol.* **19**, 065010 (2010).

[388] A. P.Ehiasarian, “Fundamentals and applications of HIPIMS.” in *Plasma Surface Engineering Research and its Practical Applications*, Edited by R. Wei (Research Signpost, Ontario, Canada, 2008).

[389] G. Greczynski, J. Lu, I. Petrov, J. E. Greene, S. Bolz, W. Kölker, Ch. Schiffers, O. Lemmer and L. Hultman, “Novel strategy for low-temperature, high-rate growth of dense, hard, and stress-free refractory ceramic thin films,” *J. Vac. Sci. Technol. A* **32**, 041515 (2014).

[390] R. Chistyakov, "High-power Pulsed Magnetron Sputtering," US Patent 7,147,759, filed September 30, 2002; issued December 12, 2006.

[391] J. Lin, W. D. Sproul, J. J. Moore, Z. Wu, S. Lee, R. Chistyakov, and B. Abraham, "Recent advances in modulated pulsed power magnetron sputtering for surface engineering," *J. Metals* **63**, 48 (2011).

[392] F. Papa, H. Gerdes, R. Bandorf, A. P. Ehiasarian, I. Kolev, G. Braeuer, R. Tietema, and T. Krug, "Deposition rate characteristics for steady state high power impulse magnetron sputtering (HIPIMS) discharges generated with a modulated pulsed power (MPP) generator," *Thin Solid Films* **520**, 1559 (2011).

[393] J. Lin, J. J. Moore, W. D. Sproul, B. Mishra, J. A. Rees, Z. Wu, R. Chistyakov, and B. Abraham, "Ion energy and mass distributions of the plasma during modulated pulse power magnetron sputtering," *Surf. Coat. Technol.* **203**, 3676 (2009).

[394] D. Depla, J. Haemers, G. Buyle and R. De Gryse, "Hysteresis behavior during reactive magnetron sputtering of Al_2O_3," *J. Vac. Sci. Technol. A* **24**, 934 (2006).

[395] D. Depla and W. P. Leroy," Magnetron sputter deposition as visualized by Monte Carlo modeling," *Thin Solid Films* **52**0, 6337 (2012).

[396] S. Mahieu S, P. Ghekiere, D. Depla, R. De Gryse, "Biaxial alignment in sputter deposited thin films," *Thin Solid Films* **515**, 1229 (2006).

[397] S. Mahieu and D. Depla, "Reactive sputter deposition of TiN layers: modelling the growth by characterization of particle fluxes towards the substrate," *J. Phys. D: Appl. Phys.* **42**, 053002 (2009).

Laudatio Steward Van Wyk

Yves T'Sjoen

Further to the Sarton Committee's call for candidates eligible for the Sarton Chair and medals, the Ghent University research group Ghent Centre for Afrikaans and the Study of South Africa has put forward Steward van Wyk. His nomination has been supported by the Department of Literature and endorsed by the Faculty Council.

Professor Van Wyk is Head of Department for Afrikaans and Dutch of the University of Wes-Kaapland in Bellville (South Africa). He is an expert in the study of the literature of *bruin- en swartskrywers* in Afrikaans, post-colonial literature and transnational literature with a specific interest in the literary production of former Dutch colonies.

Van Wyk's academic research focuses on the study of thematic and formal processes in black African literature. He regularly publishes about authors like for example, the recently deceased canonical poet and playwright Adam Small and S.V. Petersen. He also conducts comparative transnational research in the field of Afrikaans and Dutch literature. In particular, he is keen to highlight similarities between the work of black African writers and literature that has come about in Dutch colonies, such as Suriname, the Netherlands Antilles and Indonesia. He is currently expanding the focus on transnational literary research to include the phenomenon of immigrant writers in the Netherlands and Flanders.

Given the language problems that have led to heated debates on university campuses in South Africa and the fact that UWK is increasingly facing violence and politically driven protests, the faculty and the Sarton Committee, on behalf of Flemish academia, wish to offer Afrikaans and the Afrikaans and Dutch Department at the UWK every encouragement. The

Dutch expression for this is 'een hart onder de riem steken', while in Afrikaans, apparently, it is 'een riem onder het hart'. With the 2016 Sarton Medal, we give international recognition to a specialist who is highly regarded both inside and outside of South Africa's academia.

As our *alma mater*, the research group Ghent Centre for Afrikaans and the Study of South Africa has been very active in the area of Afrikaans language and literature and the study of South Africa in a multilingual and multicultural context for several years. In fact, Ghent University is the only academic establishment in the Low Countries where students can take a full-time course in Afrikaans language and literature. The Centre organizes lectures and workshops, was involved in the institutional honorary doctorate which was conferred on Breyten Breytenbach two years ago and every year, it schedules an academic study day. Recent additions are the Afrikaans literature book club and a South African movie night in our faculty. I can already reveal that next year, we will help organize the successful Week of the African novel. Incidentally, we in Ghent work closely together with the Dutch Stichting Zuid-Afrikahuis Nederland in Amsterdam. Soon, along with over 50 students who attend the African language and literature course this semester, we will pay them a working visit. There is every chance that Breyten Breytenbach will be present during that visit. We are still in negotiation in this respect, and I am also trying to enthuse Adriaan van Dis to be present at this unique occasion which the colleagues of the Zuid-Afrikahuis are offering.

Numerous domestic and foreign colleagues form part of the Ghent Centre for Afrikaans and the Study of South Africa, which is increasingly emerging as the leading centre for Afrikaans in Europe. In South Africa, we have already acquired a certain renown, and recognition for our work there is considerable. In fact, the presence of South African academics here this afternoon bears witness to that. The Centre works closely with South African partners, like Weskaap University, Stellenbosch University, Noordwes University Campus Potchefstroom and Pretoria University. We hope to establish a bilateral partnership with the University of Leyden soon, so that for the study of Afrikaans and other languages in South Africa, there will be a central interuniversity hub in the Low Countries.

Writing apartheid

The representation of the past in recent Afrikaans literature with a focus on Black Afrikaans writing

Steward van Wyk

Prof Rubens, chairperson of the Sarton Committee, the Dean of Faculty of Arts and Philosophy, Prof T'Sjoen, Prof Van Keymeulen, Dr Verdoolaege, colleagues, friends: Allow me to express my sincere and heartfelt gratitude to the Sarton Committee and the University of Ghent for this huge honor and accolade. To be honored by an institution with such a long tradition of academic excellence and groundbreaking research and innovation fills me with pride but also humility. To be the first South African to receive this honor and to do so in the company of many of my colleagues and friends from home is something that I will treasure for the rest of my life.

My gratitude also goes to Prof Yves T'Sjoen who initiated the process and I wish to thank him for his kind words in the introduction. This is a further step in cementing the long relationship between the Universities of Ghent and the Western Cape and will expand the process that was started with the Vlir projects in which so many colleagues participated. I also wish to complement you with the establishment of the Ghent Centre for Afrikaans and the study of South Africa and assure you of my support.

I.

Studies on the South African literary landscape after the transition to democracy in 1994 emphasised the turn to history as a dominant feature. André P. Brink, the writer and critic who in the dark days of apartheid voiced the triumph of the human spirit over the draconian state had particular ideas about this topic. He for instance finds in his contribution titled "Interrogating silence: new possibilities faced by South African literature" in *Writing South Africa: Literature, apartheid, and democracy, 1970-1995* that "History provides one of the most fertile silences to be revisited by South African writers; not because no voices have traversed it before … but because the dominant discourse of white historiography … has inevitably silenced … so many other possibilities" (Brink 1998, p. 22).

As a writer who in novels such as *An instant in the wind*, *A chain of voices* and his last one *Philida* undertook a critical revisit of the past, it is no surprise that he would find history to be fertile ground to turn the myths of apartheid and its false histories on its head.

Little more than a decade later Brink concludes in his lecture "Ground Zero: The South African literary landscape after Apartheid" that "one finds traces of the past in the midst of the new South African literature in the continued re-visits to the landscapes of apartheid" (Brink 2010, p. 2). He expands by saying that the literature is "a re-appreciation, even re-discovery of the past through memory as a key to identity" (Brink 2010, p. 3).

Although David Medalie is critical of Brink's hyperbolic, utopian tendencies and suggest that it is rather the dystopian nature of recent South African literature that is more characteristic, he also finds that the literature "is rather inclined to a pre-occupation with the past than an embrace of the future" (2010, p. 36). He ascribes it to the uncertainties of the transition period and observes "in historical periods which feel strongly their own transitional status ... there is an inclination to look at the present with dismay, the future with trepidation and the past with nostalgia" (Medalie 2010, p. 36).

He continues by saying that "the recurring, even obsessive preoccupation with the past in a great deal of the fiction and even non-fiction of this period may be understood as a form of nostalgia, and that it may even be pertinent to characterize it as a literature of nostalgia" (Medalie 2010, p. 36).

The element of nostalgia is indicated as a common characteristic in a number of studies on recent South African literature. These studies are generally critical of nostalgia as a longing for a past of white privilege, as stated by Fanie Naudé in a recent Memorial Lecture: "Rural or retrospective bonhomie, even the perpetuation of 19^{th}-century bucolic fantasies ... a few false notes of nostalgia, often focusing on a white childhood under apartheid, regularly with a child protagonist displaying an instinct for injustice lacking in the adults " (Naudé 2016, p. 12).

Critics are aware that the way nostalgia is handled is very complex and that there are nuances in the approaches. Robbe (2015, p. 325) distinguishes between "restorative" en "reflective" nostalgia and Medalie (2010, p. 40) discerns a "critical and evolved nostalgia as against an unreflecting one". In both cases it is about a nostalgia that sanitises the past of its atrocities and transgressions and creates an idyll as against the conviction that recreating the past can be a tool for critical reflection on the present.

An important question posed by Naude is how black writers engage with nostalgia and referring to Jacob Dlamini's *Native Nostalgia* he finds that the responses are ranging in nature.

A dominant theme in Black Afrikaans writing since 1994 was the return to the past, to forgotten histories as an answer and response to the collective quest to "tell our own stories". This theme was introduced by AHM Scholtz's highly aclaimed *Vatmaar* and EKM Dido's *'n Stringetjie blou krale/A string of blue beads*. This tendency is continued in the recent publications of a number of writers which is the subject of my paper. In my opinion the key issues for these writers are:

- How do I write about a past from which I was excluded in the official historiography,
- How do I write myself into that history in the language from which I was excluded?

In reviewing the South African political setup at the end of 2015 Achille Mbembe makes an interesting remark about the current state of mind:

Tropes of pain and suffering had come to saturate current narratives of selfhood and identity ... These tropes had become the register through which many now represent themselves to themselves and to the world. To give account of whom they are, or to explain themselves and their behaviour to others, they increasingly tend to frame their life stories in terms of how much they have been injured by the forces of racism, bigotry and patriarchy (Mbembe 2015, p. 4)

II.

Fatima Osman's autobiographical narrative *Ek's g'n slams/I'm no slams* (2013) is the account of someone who has struggled her whole life against racism, bigotry, male chauvinism en patriarchy and whose psyche was scared by it. It is a *Bildungsroman* in which the main character, a girl of Indian descent growing up in a traditional Muslim family, rebels against the orthodox family setup in which the young woman has no say in her life and often finds herself in an arranged and unhappy marriage. Fatima is a strong and wilful woman and after a life in which she was the victim of her husband's reckless business transactions and always having to bounce back from it, she takes control of their lives as the primary provider in their household.

The first part of the narrative contains stories about the traditions and customs of immigrant Muslim families from India. It gives rare insights into a part of the Cape community never before told in Afrikaans. An important element is the dialogue in the Cape vernacular Kaaps of District Six. The narrator says the following about how the cleric/galifa taught them about the Koran and kop-les:

> Met groot trots het ons galifa ons geleer ... Altyd met sy vinger in die lug sou hy in sy Distrik Ses-Afrikaans begin ... "Wiet djulle, as djulle 'n roos pluk sonner toesteming in iemand se tyn, da' gat Allah djulle straf! Djulle gat jahannam toe! As die malaaikats (engele) na julle kabr (graf) toe ko' en djulle kennie djulle Kalima Shahada' nie, da' gat djulle behoo'lik innie jahannam brant! (p. 14).
>
> *Our cleric taught us with great pride ... Always with his finger in the air he would begin in his District Six Afrikaans "Do you know that if you pick a rose in somebody's garden without the consent, then Allah will punish you! You will go to hell! When the angels come to your grave and you do not know your lessons, then you will burn in hell*

The construction of identity is foremost in the opening pages and the references to group identity are all important: "Die mense daar buite is Kaapse Maleiers en Kleurlinge en wit mense ... Ons meng nie. Ons is Indiërs. Ons is 'n rein nasie (p. 11). Ons bly rein"/ *The people out there are Cape Malay and Coloureds and white people ... We do not mix. We are a pure nation. We stay pure.*

It is ironically this emphasis on group identity that drives Fatima to assert herself as individual in her own right. Against the traditional customs she chooses her own husband, she starts to wear tight jeans, cuts her hair in a short style and starts to befriend the charismatic, evangelist women in her neighbourhood. She fights the preconceived ideas of men and the fact that everywhere she goes she has to defend herself. The title of the novel gets particular resonance when she defends herself against the gangsters on the street corner:

> "Ek's g'n slams nie ... Ek issie! Ek issie! ... 'n slams is 'n stukkende gebou! Ek's 'n Moslem. Ek's 'n mens! O's is ammal mens!" (p. 45).
>
> *I am no slams … I'm not! I'm not … a slams is a derelict building! I'm a Muslim. I'm a person! We're all people!*

As a way of defining her own life and taking charge of her future she starts to experiment with diets and becomes the victim of its secondary effects. She suffers from depression and lands in an asylum. After different forms of therapy she finally finds respite. In a vision that she experiences during a visit to the dentist she comes to the conclusion:

> My lewe lank loop ek sonder 'n identiteit rond en ek kon dit nie verwerk nie. Die skets het iets in my wakker gemaak. Ek weet wie ek is! Daar het 'n gestalte in die lig gestaan en hy het my naam geroep. ... Dis Allah sub-ga-nallah se gestalte wat ek langs myne gesien het en wat my laat besef ek moet net op Hom vertrou en alles val in plek ... Dis die groot deurbraak waarop ek al so lank wag ... Dit is 'n wonderwerk van Bo, van Allah sub-ga-nallah self wat tot my redding gekom het ... Ek het my lewe terug omdat ek oorgegee het aan Allah sub-ga-nallah. My naam is Fatima en ek het innerlike vrede ... Dit is opwindend en verfrissend om te weet ek het my identiteit gevind ... Ek is vry van alles. Vry. (pp. 266-269).
>
> *My whole life I wandered without an identity and I could not handle it. The painting awakened something in me. I know who I am! There was something in the light and He called my name … It is Allah sub-ga-nallah's image that I saw next to me and made me aware that I should rely on Him and everything will fall into place… This is the big breakthrough that I have been waiting for so long … This is a miracle from above, from Allah sub-ga-nallah who came to my rescue himself … My life is returned to me because I submitted to Allah sub-ga-nallah. My name is Fatima and I have*

> *inner peace … This is exciting and refreshing to know I have found my identity … I am free from everything. Free.*

It is a story of individual triumph over adversity that plays off against the backdrop of the apartheid histories in District Six and other rural, black communities in the Southern Cape.

In *Conscripts of Modernity* David Scott (2004, p. 8) makes the following statement:

> Anticolonial stories about past, present, and future have typically been emplotted in a distinctive narrative form, one with a distinctive story-potential: that of *Romance*. They have tended to be narratives of overcoming, often narratives of vindication; they have tended to enact a distinctive rhythm and pacing, a distinctive direction, and to tell stories of salvation and redemption. They have largely depended upon a certain (utopian) horizon toward which the emancipationist history is imagined to be moving.

I will return to Scott's alternative for the way in which the anticolonial past and postcolonial future should be construed. His previous statement is relevant to Osman's *Ek's g'n slams* as well as the debut novel of Simon Bruinders.

III.

Die sideboard/The sideboard (2014) is the story of Abraham de Bruin who battles his whole life against the unjust apartheid system that robs him in a devious and heartless manner of his dream to have his own little piece of land where he and his family could build their own future. As a smart, hard-working young man who firmly believes in Biblical and traditional values and after being devastated by the senseless expropriation of his land by the authorities, he often rises to start anew in the conviction that God will provide them with a sensible outcome.

Abraham joins the Cape Corps, the separate army unit for Cape Coloured troops, after he was promised ten hectares of land in addition to his military pay. They fight on the side of the Allied Forces in the Second World War in Abyssinia but after the war they feel duped because after demilitarization they receive only a bicycle and a coat and not the promised piece of land. The physical and emotional scars threaten to overwhelm him and he concludes:

> Hy het as ’n gewonde mens teruggekom. Maar ’n dankbare mens. En tog het hy sy beloofde land nooit gesien nie. Vir baie jare het hy verneuk gevoel. Hy het geglo die grondbelofte was van God. Dit was helaas die belofte van mense. ’n Windeier (p. 226).
>
> *He returned as a wounded person. But a thankful person. Although he never saw the promised land. For many years he felt conned. He believed that the promise of land was from God. Alas, it was the promises of people. A farce.*

The sideboard of the title gets symbolic meaning and provides closure. Abraham saw the piece of furniture the first time in Abyssinia where he was so overawed by its beauty and workmanship that he decided to make a replica for his wife Stella on his return. It occupies a special place in their home and becomes the altar on which their family Bible rests. It also symbolises their hardship when it is nearly destroyed in a fire and Abraham could save it with superhuman strength. Thereafter it is sold for a trifle when they are forced to relocate from the Island to a smaller house in Rosemoor. The circle is concluded on a happier note when their former conscientious white neighbour buys it at an auction and returns it to them on their wedding anniversary as a symbol of atonement. Closure is brought to Abraham’s travails and suffering and in gratitude he proffers a jubilation: “Dankie, dankie, dankie, Here, U is groot en genadig” (p. 262) / *Thank you, thank you, thank you God, You are great and merciful.*

Against the background of the tradition of the plaasroman/farm novel in the Afrikaans literature and the central role that land and identity plays in this genre *The sideboard* gives an important perspective from the bottom up, from communities dispossessed from their land (See Willemse 2015, p. 252). Despite the happy and fortuitous end to the narrative the overwhelming feeling is of deceit and injustice and it fits squarely within the postcolonial theme of place and displacement.

In the words of Scott it is a story of survival and determination and the triumph of the human spirit in the face of a dehumanising system that endeavours to bring him down. But as Scott indicates the romance presents itself as an answer to questions about the past and the present and faced with the unsatisfactory nature of those answers one needs to pose a different set of questions to get another perspective on this problematic. In Scott's view the narrative mode of tragedy could provide an answer:

> We live in tragic times. Not meaningless times, not merely dark or catastrophic times but times that in fundamental ways are distressingly off kilter in the specific sense that the critical languages in which we wagered our moral vision and our political hope (including, importantly, the languages of black emancipation and postcolonial critique) are no longer commensurate with the world they were meant to understand, engage, and overcome (Scott 2004, p. 210).

IV.

With the above remarks in view I wish to reflect on a few of Nathan Trantraal's recent columns in the Afrikaans weekly *Rapport*. In an interview on the online portal *Litnet* Willemse (2016) refers to something in Trantraal's work that is crude and gruff and perhaps likeable within modern hip-hop-gangsta culture, but which lacks the intellectual insight, craft and skilfulness of the acclaimed poet and play-wright Adam Small.

I believe that Willemse is a bit harsh in his judgment and that one should look for a more nuanced reading of Trantraal in the context of the world that he seeks to represent and responds to. Mbembe's (2001) views on the aesthetics of vulgarity can be applied in such a reading. Mbembe uses the Bakhtinian notions of the carnivelesk and his remark that "the grotesque and the obscene are ... the province of ordinary people" is important for an appraisal of Trantraal's views.

Trantraal aligns himself unequivocally with the ordinary people and the poor and purports to be writing from a subject position within that community and not about them. He is however also aware of his position in the dominant white establishment and culture and finds his legitimacy in his undaunted opposition and critique of the hegemony. Mbembe's remarks are again applicable in this regard: "As a means of resistance to the dominant culture, and as refuge from it, obscenity and the grotesque are parodies that undermine officialdom by showing how arbitrary and vulnerable officialise is and by turning it all into an object of ridicule (Mbembe, 2001, pp. 103-104).

His further observations about the obsession in Tongolese popular culture with “orifices, odors, and genital organs” is interesting when taking into account the resentment that many (white) readers felt towards some of the poems in Trantraal’s collection *Chokers en survivors*.

Not much has been said about Trantraal’s approach to the past and history and I believe it makes for interesting analyses. In his column *Sypaadjies* that appears in the Sunday newspaper *Rapport* he deals especially with the life of the poor in the present and apartheid past and presents an account of the genealogy of poverty in which he emphasises the heroic struggles to survive from day to day. In the column “Ek dink aan die mense innie hyse wat soe horrible is” / *I think of the people in those houses that are so horrible* he analyses the dialectic between the poor and their environment: “Bishop Lavis se hys isse gatmaak. Soes in: It is funny en soes in: Jy grawe ’n gat en dan sê jy vi mense: Daa isse hys vi jou”. Sy slotparagraaf is besonder treffend: “Die romantic philosophers het ’n ding gehad wat hulle gesêrit: ‘Show me where you live and I’ll show you who you are’. Ennit moet waa wies hoekom mense wie in lekke neighbourhoods bly, lekke mense is, en daa isse riede hoekom arm mense heeltyd dronk wil wies en baklei”/ *Bishop Lavis’s houses is a joke. Like in: it is funny and like in: you dig a hole and then you tell the people: “There’s a house for you”. His last paragraph is striking: “The romantic philosophers had a saying: Show me where you live and I’ll show you who you are”. And it must be true because people that live in nice neighbourhoods are nice, and there is a reasons why poor people always want to be drunk and fighting.*

In “Iennage iets was nie veniet”/*Not everything was free* he criticises the self-help moralising of the middle class and older generations and it is his defence that it would be insensitive and hypocritical to expect the poor to have a conscience about their responsibility given the little means that they have: “Ennie ou mense wie nog altyd ‘ie kan onthou van swaakry nie gan wee’ sê: ‘Vandag se kinnes wil alles veniet hê,’ asof hulle al ooit ienagge iets veniet gekry et, asof hulle al gekry ‘et wat vi hulle beloewe was, wat vi hulle en hulle voo ouers geskuld wôd”/ *And the old people who can’t remember hardship will again say: “Today’s children want everything for free, as if they already got something for free, or as if they got what they were promised, what was owed to them and their ancestors*. In “Kleintjie is die regte Michael K”/ *Kleintjie is the true Micheal K* he responds to J.M.

Coetzee's novel and tries to come to terms with the position of the outcast that is outside of History. He comes to the conclusion: "Ek sal hom 'ie 'n context hoef te gie nie, wan hy exist in whatever context jy op hom in ienagge moment fling"/ *I don't have to give him a context because he exists in every context in which you fling him.*

The above pieces of writing leaves the impression of an absence of the teleological moment of freedom, of a present suspended between a past of repression and future of futile hope.

V.

With this in mind I want to reflect on Ronelda Kamfer's poetry anthology *Hammie* (2016). These poems have a tragic and elegiac tone. The Hammie of the title, which is a term of endearment for the loved and deceased mother, is eulogised in a number of poems that serves as a way to come to terms with death and bereavement.

A wish but futile hope for consolation is found in the poem "Seymour": "die oggend toe ek die ICU instap/het ek/gehuil/op my knieë geval/en gehuil/ ek het vir myself gesê/soos mens mos sê as iemand doodgaan/sy is na 'n beter plek/en ek het dit geglo/totdat die suster my kom troos het/en vir my fluister/ jou mammie is op 'n beter plek/toe sê ek nee suster/ my ma is op die worst possible plek/ waar 'n mens kan wees" (p. 40).

The morning when I stepped into the ICU I cried and I fell on my knees, and I cried and I told myself as one always says when someone dies: she is in a better place, and I believed it, until the sister came to console me, and whispered to me, your mommy is in a better place, then I said no, my ma is in the worst possible place where one can be.

And in "Nathan" (p. 21) the irreversible nature of the news is underscored in the way it breaks into the ordinary chain of events of a day that was supposed to be ordinary: "vandag was veronderstel/om 'n boring dag te wees/ek en my ma was van plan/om my babaklere uit/die lay-bye te gaan haal".

Today was supposed to be a boring day, my mother and I planned to fetch my baby's clothes from the lay-bye.

There are epic poems addressed to working class women who make sacrifices for their families, as in "antie Gerty, suster Kamfer, antie Trui" (p. 74). It sketches an image of women who in the face of poverty fights to keep a sense of normality but are disarmed by situations in which they do not know how to express emotion ("Torring" p. 49/*Nuisance*), who are forced by circumstances into a business-like and unsentimental handling of death and bereavement, for instance in the poem "Troupanne" (p. 33)/ *Wedding bands* where the narrator is forced to pawn the wedding bands of her deceased mother so that she can buy food and medicine for her own child. A number of poems express an ambivalent and complex relationship with faith, religion, God and the Bible.

But it is not only a lament; the imperfections of the mother and her generation are mercilessly exposed. In the poems "Humble brag" (p. 93) and "Wees lief vir mekaar, kyk agter haar, Neldie, sy is jou suster ..." (p. 94) / *Love each other, look after her Neldie, she is your sister* ... the neglect of parental duties is disclosed. As such the anthology avoids the romanticising of working class women who in their nurturing and supportive role heroically withstand the onslaught of society.

Relevant here, are Scott's remarks about tragedy as the form in which the anticolonial past can be represented. Dalleo's paraphrasing of Scott is apt: "In contrast to the assured optimism of romance, the tragic vision wonders if we are ever fully masters of our fate and if the past can ever be entirely left behind ... tragedy emphasizes contingency and conflicting, often irreconcilable demands: for these reasons, it is especially well suited to ambiguous moments of historical crisis and transformation, when old certainties are coming apart" (Dalleo 2004, p. 134).

The anthology takes a hard look at the travesties in communities: family violence, sexual assault, gangsterism. The view of the apartheid past and the postcolonial, postapartheid present in the townships and fruit farms of the Western Cape is one of disillusionment and unhomeliness. Dystopian elements abound, the past it not the idyll to which can be returned with longing and the present is not the fulfilment of heroic and utopian desires.

VI.

Of the aforementioned texts Valda Jansen's debut novel *Hy kom met die skoenlappers* (2016) */He comes with the butterflies* shows mellowness and comes across as the most considered approach towards the past and history. The novel is presented as an elegy on lost love, cast in a hybrid form of memoir, diary and epistolary novel in which the narrator goes in search of the happy times and lost chances of a romance in her days as a young German language teacher on a yearlong exchange in the former Communist Germany. The pasts of herself and her lover Anders are revealed in the process. They are wounded not only by personal and family tragedy but also as a consequence of the histories of apartheid South Africa and Germany during the Nazi and DDR regimes. Anders puts it this way: "Ons probeer nog om vandag te verstaan. Daarom moet ons soms weer stil word en terugdink, onthou hoe dit was. As ons dit nié doen nie, sal ons lewe vandag geen betekenis hê nie" (p. 73). / *We are still trying to understand today. That is why we sometimes need to be still and think back, remember how it was. If we don't do it, our life will not have meaning.*

The narrator expresses a similar honest and bold acceptance of the past. She says for instance: "Wit Suid-Afrikaners kan nie die wandade van apartheid ontduik nie" (p. 155) / *White South Africans cannot escape the travesties of apartheid.* And on p. 75: "Ons moet almal nog leer om met ons verlede saam te leef, om vrede te maak met wie ons was" / *We must all learn to live with the past, to make peace with who we were*".

The inability to fully comprehend the past is shown at a textual level. The deficiency of language to represent the past is signified in the failure of punctuation and capital letters, eliptical sentences and stuttering utterances e.g. p.158

> ek vertel my storie jy leen my jou oor my storie is my voete wat om 'n vuur dans ek dans
>   die taal van dooie woorde
>
> *I tell my story you lend me your ear my story is my feet that dances around a fire I dance*
>   *the language of dead words*

It shows the working of amnesia and the repression of hurtful memories from the past, the quest to fill in pieces of memory and the emotional pain lying underneath the words. A similar point is made by Robbe (2015, p. 329) when she refers to the Antjie Krog's textual strategies to reconstruct the past.

Jansen's formal experiments illustrate how difficult it is to write about apartheid and create a meaningful account of the past. In this regard the narrator remarks: "In my weergawe van my geskiedenis is ek in die middel van die storie saam met my ouers en my familie wat ek van foto's af ken. In vele ander se idee van die geskiedenis van Suid-Afrika is ek uitgesluit ... die ou geskiedenisvertellers lieg die waarheid en die nuwe heersers is nie veel beter nie" (p. 129). *In my version of my history I am in the centre of the story with my parents and my family who I know from photo's. In many others' idea of history in South Africa I am excluded. The old historians lied the truth and the new rulers are not much better.*

Some of the moving scenes in the book are the times when she recounts her experiences as a young child with the apartheid laws. It is told in an understated manner but in such a way that the hurt and humilation are unmistakeable. For example when she tells about the times that they have to go around the back of the hotel to get a meal of left-overs (p. 118). The emphasis that falls on the local in this small town installment of apartheid accentuates the working of apartheid machinery. The narrator describes her experiences with the Group Areas Act in her hometown the Strand with its long white beaches in this way:

> Ek en my broers stap dorp toe ... Ons loop op die sypaadjie tussen die witmense se huise ... Ons mag nie daar woon nie, ons mag net inkopies doen en dan die dorp verlaat, ons mag nie in restaurante eet nie, veral nie in die Wimpy nie (pp. 99-100).
>
> *My brothers and I walk to town ... We walk on the pavement between the houses of white people ... We are not allowed to stay there, we can only do shopping en then we must leave the town, we may not eat in the restaurants, especially not in the Wimpy*

The emphasis on white here and in other passages should be noted. It is as if the narrator uses it as a curse and turns it on itself. The narrator's ambivalent relationship with the Afrikaans language is another aspect that is explored in the book. Despite the fact that she was raised in an Afrikaans home with her father scolding them in his perfect Eastern Cape idiom and pure Afrikaans (p. 63) her Afrikaans is perceived as strange when she and her child speak it on the school grounds. Here she plays in on the notion that Afrikaans is the white man's language. She alludes to the fact that her humanity was stifled in Afrikaans and that it was through German that she found her way back to the language (p. 170).

An important intertext are the passages in German that are sometimes translated and sometimes not. She finds solace in German and in the multiple references to Bertolt Brecht. It becomes a meta-textual reflection on writing in times of crisis. She remembers the words of her German lecturer during the struggle years of the eighties: "Bertolt Brecht kon net sowel met ons gepraat het, vandag, hier in hierdie land" (p. 94)/*Bertolt Brecht could just as well have spoken to us, today, in this country*".

In the same vein these words can be relevant in the present in which the narrator is intensely aware of the socio-political role of the writer and the need to write about societal issues: "Watter soort tyd is dit waarin ons leef dat 'n gesprek oor bome eintlik 'n misdaad is omdat dit beteken dat ons oor soveel ander dinge swyg"(p. 94) / *In which times do we live that a conversation about trees is really a crime because we remain silent about so many other things*.

This makes the novel not just another love story or elegy about lost love, but a reflection on the role of literature in times of crisis and the impossibility to escape the past even though you wear the scars of that past.

VII.

It is difficult to find a concluding remark in these times in which everything is in flux and the very nature of our creative and intellectual endeavours are under scrutiny. It is rightly so that the South African past, present and future should be decolonized, that our symbols and institutions should transform at a faster pace. And although I believe that there are no holy and privileged spaces for intellectual engagement devoid of societal pressures I fear for the madness in a #Feesmustfall tweet that read: "If it's too expensive for black people to exist in this space, then it's better if this space doesn't exist".

Amidst the dystopian visions there is still the belief in literature, as Imraan Coovadia (2012, p. 59) puts it: "Literature is the art of writing that ... hear(s) as articulate speech what might be dismissed as noise, to turn what cannot be noticed into what cannot be ignored ... these are the transformative tasks that writing, in South Africa, has to undertake."

I thank you.

Works cited

[1] Brink, André P. (1998). Interrogating silence: new possibilities faced by South African literature. In: Derek Attridge and Rosemary Jolly, *Writing South Africa: Literature, apartheid, and democracy, 1970-1995*. Cambridge: Cambridge University Press, pp. 14-28.

[2] Brink, André P. (2010). Ground Zero: The South African literary landscape after Apartheid. First Jan Rabie & Marjorie Wallace Memorial Lecture, University of the Western Cape, Bellville, RSA, 17 September 2010.

[3] Bruinders, Simon. (2014). *Die sideboard*. Tygervallei: Naledi.

[4] Coovadia, Imraan. (2012). *Transformations: Essays*. Cape Town: Umuzi.

[5] Dalleo, Raphael. (2004). Emplotting Postcoloniality: Usable Pasts, Possible Futures, and the Relentless Present. *Diaspora* 13.1, 129-140.

[6] Jansen, Valda. (2016). *Hy kom met die skoenlappers*. Kaapstad: Human & Rousseau.

[7] Kamfer, Rondelda S. (2016). *Hammie*. Kaapstad: Kwela Boeke.

[8] Mbembe, Achille. (2001). *On the Postcolony*. Berkeley, Los Angeles & London: University of California Press.

[9] Mbembe, Achille. (2015). The State of South African Political Life. *Africa is a Country*, 19 September 2015. <http://africasacountry.com/2015/09/achille-mbembe-on-the-state-of-south-african-politics/>

[10] Medalie, David. (2010). The Uses of Nostalgia. *English in Africa* 53.1, 35-44.

[11] Naudé, S.J. (2016). Empty white spaces and the politics of nostalgia. Third Jan Rabie & Marjorie Wallace Memorial Lecture, University of the Western Cape, Bellville, 23 August 2016.

[12] Osman, Fatima. (2013). *Ek's g'n slams*. Pretoria: Lapa Uitgewers.

[13] Robbe, Ksenia. (2015). Translating Transition: The politics of Nostalgia and Provincialization in Antjie Krog's *A Change of Tongue* and *Begging to Be Black*. *Tijdschrift voor Nederlandse Taal- en Letterkunde*. 131.4, 323-338.

[14] Scott, David. (2004). *Conscripts of Modernity: The Tragedy of Colonial Enlightenment*. Durham & London: Duke University Press.

[15] Trantraal, Nathan. (2016). Ek dink annie mense innie yse wat soe horrible is. *Rapport Weekliks*, 7 February, 15.

[16] Trantraal, Nathan. (2016). Kleintjie is die regte Micheal K. *Rapport Weekliks*, 12 June, 15

[17] Trantraal, Nathan. (2016). 'Ienagge iets was nie veniet'. *Rapport Weekliks*, 21 August, 15.

[18] Willemse, Hein. (2015). Review of *Die sideboard*. *Tydskrif vir Letterkunde* 52.1, 252-253.

[19] Willemse, Hein. (2016). Onderhoud: Hein Willemse oor Adam Small, Nathan Trantraal, velkleur en dekonstruksie. http://www.litnet.co.za/onderhoud-hein-willemse-oor-adam-small, 12 August 2016.

Laudatio Louis (A.M.J.A.) Berkvens

Dirk Heirbaut

Louis Berkvens' career shows that the best research does not always come from countries far away and in foreign languages, but can also be found in our own language and in neighbouring countries. Louis Berkvens was born in Eindhoven in 1952 and was awarded a history degree in 1980 from Nijmegen University. The latter was also the university where his academic career took off with a position at the Gerard Noodt Institute of Legal History. A Ph.D. thesis followed in 1990. By then, Louis Berkvens had already moved on to Maastricht University, where he would go through all the stages from lecturer to professor by special appointment. The name of his chair indicates his life's work: legal history, not in general, but of the Limburg territories. The regional history organisation, Limburgs Geschied- en Oudheidkundig genootschap provides the funding for this endowed chair and it is probably one of the best investments in academia ever made, as Louis Berkvens has given the Dutch province of Limburg a prominent place in the world of legal history.

The difficulties he had to cope with in his research can hardly be overestimated, as before the French Revolution this region had a bewildering institutional diversity. Some territories depended of the Dutch Republic, others of the Habsburgs Netherlands and there were also many other larger and smaller principalities which were or became involved in the Limburg imbroglio, like for example the ecclesiastic principality of Thorn. Regime changes and complex local situations of shared power (e.g. the joint sovereignty over Maastricht of the catholic bishop of Liège and the protestant Dutch Republic) added new twists to this complexity. In short, the legal history of Limburg confuses out- and insiders. It is impossible for the modern researcher to catch all details, but Louis Berkvens has gone far on the road to perfection in his studies of Spanish Guelders, Prussian Guelders and Guelders of the States (i.e. of the Dutch Republic) and also of smaller territories like Thorn, Dalhem and Kessenich. Moreover, he has always paid particular attention to the history of the Habsburg Netherlands, so that Belgian legal historians consider him to be also a historian of their country. Louis Berkvens specializes in the study of the law as contemporaries practiced it. Hence, he has devoted a lot of time to the sources of daily life of the law. He has published the lists of the edicts of the Upper Guelders (the part of the old principality of Guelders which is now largely incorporated into the Dutch province of Limburg). His interest in the law in action directed his research towards the courts and tribunals active in Limburg, their staff, their rules of procedure and their famous and less famous cases.

It would be wrong to see Louis Berkvens solely as an individual scholar, as his personality and research have inspired many others. The Limburgs Geschied-en Oudheidkundig Genootschap has published many of his works. Most of all, Louis Berkvens works closely with Limburg archivists. Under his supervision, two of them, Hans van Hall and Jacques van Rensch, wrote impressive theses on the liberty of Eijsden and the sun fief Gronsveld. The enthusiasm Louis Berkvens and his friends of the Limburgs Geschied- en Oudheidkundig genootschap have generated is so great that their recent book on the history of Limburg was sold out almost immediately, in spite of a print run of thousands of copies.

Many outsiders harbour a wrong impression of local legal history. They assume that researchers working on the history of the highest courts share in the latter's greatness, whereas researchers of more modest courts and tribunals are as low as the courts they study. This is, of course, nonsense. The real life of the law played itself out not in the great law suits of the highest courts, but in the run of the mill cases before the lower courts and tribunals. However, these sometimes pedestrian cases took place in the context of a larger regional, national, European and global history. Louis Berkvens has always embedded his 'local' publications in a broader worldview, giving their due to events elsewhere, in particular in the capitals which decided over the fate of the Limburg territories. He has also ensured that others also do this, for example his student Bram Van Hofstraeten. A hard to destroy fallacy concerning local legal history is that only those academics who publish in English can become members of the Champions League of legal history. However, foreign colleagues are too ignorant of the local and regional intricacies and they will not even notice when an author wrongly interprets a convoluted local legal situation. On the other hand, it is very hard to speak for an audience of local historians and archivist who by heart know every little corner of their part of the world and will not at all hesitate to criticise the learned professor for even the tiniest error. So, one should not underestimate the achievements of Louis Berkvens. Moreover, he has also worked on the broader topics of Dutch and European legal history and has published in Dutch, English, French and German.

Louis Berkvens, lover of archives and old books, has, nevertheless, also been a pioneer of digital humanities. Nowadays, that sounds less exciting, but at the time he blazed the trail for many other projects and initiatives. For example, the electronic newsletter, the Rechtshistorische Courant, of the Ghent Legal History Institute, would have been unthinkable without the example Louis Berkvens provided. He was also a pioneer in another field. If today many scholars in the Low Countries are reaping the harvest of legal iconography studies, they are well aware of their debt to Louis Berkvens and his colleagues involved in earlier projects.

In short, there is ample justification for awarding a Sarton medal to Louis Berkvens. If his friends and admirers have a point of criticism, it may be that he is too modest. As Dutch editor in chief of the legal history review Pro Memorie he has helped, together with his colleagues Georges Martyn and Paul Brood, countless others, even at the cost of his own career. Thus, the award of the Sarton medal is a chance to shine a light on a scholar who has always graciously put others in the spotlight.

The Comparative Legal History of the Meuse-Rhine Euroregion; 'A *menage à trois*[1]

Louis Berkvens

1. Introduction

The comparative legal history of the 'Limburg territories', where Dutch, Belgian and German legal history meet, forms the heart of this article. I refer in this context to a triangular relationship, a *menage à trois*, an expression normally reserved for a romantic *liaison*, in which two persons, independently from each other are in love with the same person and the protagonist must compete with a rival for the hand of his or her lover. In this article, however, it is about the territorial relations that were decisive for the legal life in what is now called the Meuse-Rhine Euroregion and in which romantic love was often in short supply. The protagonists may not always be the same; the Meuse-Rhine Euroregion is the constant, however. In addition to the Habsburg Netherlands of Charles V and the Holy Roman Empire of the German Nation, the Southern Netherlands, the Dutch Republic, Prussia, the Duchies of Cleves-Mark; Jülich-Berg, the Prince-Bishoprics of Cologne and Liège, and a number of immediate territories will also make an appearance in varying triangular forms.

[1] This text is the written version of a lecture given in abbreviated form at Ghent on 10 November 2016 on the occasion of the awarding of the Sarton Medal 2016-2017 to the author. The author expresses his gratitude to the Sarton committee for allowing him to publish the Dutch-language original text in *Pro Memorie*.

2. The Special Character of the Legal History of the Territories between the Rivers Meuse and Rhine

Two apparently conflicting trends are typical for the Limburg territories in the early-modern period (16^{th}-18^{th} century): the emergence of modern centralised states, predominantly characterised by the exclusivity of public power over their subjects within a defined territory, on the one hand, and growing territorial fragmentation, on the other. Whereas the emergence of the modern state facilitated important reforms to the executive, legislative and judicial powers, the power politics of the absolutist sovereigns also led to a considerable degree of instability, to wars and a periodical rearranging of the territorial organisation of the area, especially in border areas, such as the region between the rivers Meuse and Rhine, which were situated within the sphere of influence of several powers.

At the beginning of the sixteenth century, the emerging territorialisation in the Limburg Meuse valley led to a problematic determination of the spheres of influence between the Burgundian and the Lower Rhine-Westphalian Circle (*Kreis*); between the Habsburg Netherlands under Charles V and the remaining territories of the Holy Roman Empire. In the course of the seventeenth and eighteenth century, due to international treaties that led to a further carving-up of territories, the resulting intricate territorial organisation became more complex. This applied, *inter alia*, to the Lands of Outremeuse (Overmaze), which under the 1661 Partition Treaty (*Partagetractaat*) were divided up between the Southern Netherlands and the Dutch Republic. Such partitioning was carried out even more intensely under the 1713 Treaty of Utrecht and the 1715 Antwerp Barrier Treaty (*Barrièretractaat*), which resulted in the partition of Spanish Upper Guelders (*Overkwartier van Roermond*), which was divided up between Austria, Prussia and the Dutch Republic. With regard to all of these partitions, while the existing executive and judicial institutions within the Limburg territories were preserved, remarkably, parallel institutions were established and inserted into the superordinate structures of the new rulers.

In the course of the seventeenth and eighteenth century, as a result of the politics of the new sovereigns, the existing institutions and rules began to show more variation. This makes it possible to compare the various aspects of the executive, legislative and judicial powers within the various Limburg territories and to study the differences in approach between, for example, the Dutch Republic, the Southern Netherlands, Liège and Prussia. The paramount question here is: what was the significance of a growing territorial fragmentation for the administration of justice in the Limburg territories?

3. Territorialisation of the Judiciary in the Early Modern Period

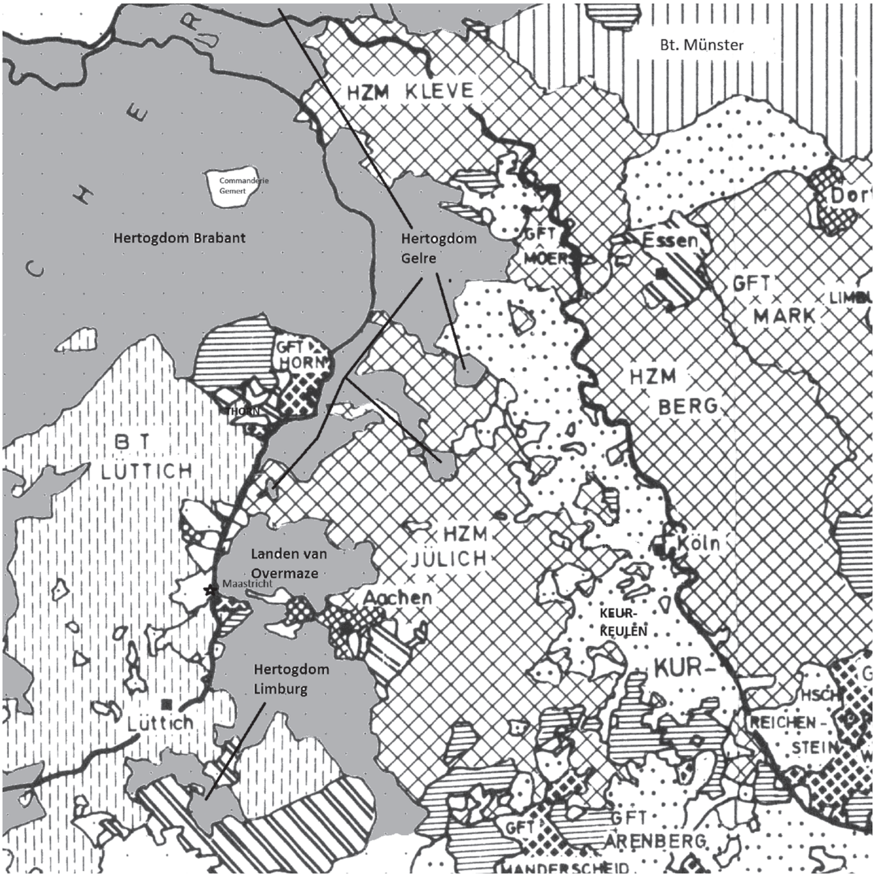
The special character of the legal history of Limburg in the early modern period has been determined to a major degree by the efforts of Charles V to organise the Habsburg Netherlands into a separate Burgundian Circle (*Kreis*) and disconnect these as much as possible from the other Circles of the Holy Roman Empire. Especially in the valley of the river Meuse, this policy was not easy to bring about, because of its considerable territorial fragmentation, on the one hand, and the trans-border judicial relations that had existed since the Middle Ages, such as referral (*hoofdvaart*) and appeal with the judicial colleges and courts of foreign powers, which were not compatible with the concept of a modern sovereign state, on the other. This meant that precisely in this region the territorial delimitation of the Netherlands often entailed a breach with traditional judicial organisations and the imposition of new organisational models.

Around 1530, Charles V, for instance, prohibited referrals and appeals from the Lands of Outremeuse with the royal scabinal court of Aachen (*Köningliche Schöffenstuhl*) and the imperial chamber court (*Reichskammergericht*), and replaced these with the possibility of appeal with the Council of Brabant (Soevereine Raad van Brabant) at Brussels.[2] His attempts also to subject Maastricht to the operation of the Brabant *privilegium de non evocando et non appellando* (the privilege not to be subjected to appeal outside the territory) resulted in a protracted controversy over appellate proceedings in Maastricht, which ultimately in 1537 led to the setting up of a proper appellate court for Maastricht with its two overlords.[3]

---

2 M.W. van Boven, 'De verhouding tussen de Raad van Brabant en de hoofdbanken in zake de appelrechtspraak in civiele zaken', in: B.C.M. Jacobs and P.L. Nève (eds.), *Hoven en Banken in Noord en Zuid*. Brabantse Rechtshistorische Reeks, vol. 7, Assen, 1994, 121-134, in particular 122: *competentio primae appellationis* Limburg aan de Vesdre 1521, Valkenburg 1521, 's-Hertogenrade 1528, Dalhem 1546.

3 A.Fl. Gehlen, 'Inrichting van het hoger beroep te Maastricht (16e-18e eeuw)', in: *Publications de la Société Historique et Archéologique de Limbourg* (hereon after: *PSHAL*), vols. 134-135, 1998-1999, 207-238.



Meuse-Rhine Territories about 1560. Map based on Andreas Schneider, Der Niederrheinisch-Westfälische Kreis im 16. Jahrhundert.

The incorporation in 1543 by Charles V of the Duchy of Guelders into the Netherlands divorced it from the territories in the possession of the Dukes of Jülich, which, apart from the Duchies of Jülich and Berg, also comprised Cleves, Mark and Ravensberg.[4] As may be learned from the strong influence of Jülich on the codification of the customary law of Upper Guelders in the sixteenth and seventeenth century, the break-up did not automatically mean the end, however, of Guelders' orientation towards the Rhine-

4 Since the Erbeinigung of 1496 the territories of the Dukes of Jülich comprised Cleves, Jülich, Berg and Mark, see J.J. Scotti, *Sammlung der Gesetze und Verordnungen, welche in das ehemahligen Herzogthum Cleve ... ergangen sind*, Düsseldorf, 1826, vol. I, 16; in the period between 1538 and 1543, the Duchy of Guelders and County of Zutphen were also part of their territory.

land.[5] As a result of the Dutch Revolt around 1590 the unity of the Duchy of Guelders ceased to be, inasmuch as the southern part continued to belong to the Southern Netherlands, as Spanish and subsequently Austrian Guelders, and the three northern quarters became part of the Dutch Republic.

As a consequence of the War of the Cleves Succession (under the 1614 Treaty of Xanten), the unity between the territories of Jülich-Cleves was also disrupted and the Duchy of Cleves was brought within the sphere of influence of Brandenburg. This ultimately led to a strong presence of Prussia at the Lower Rhine up to the river Meuse. After the War of the Spanish Succession, this was extended in 1713 through the acquisition of Prussian Guelders and the County of Meurs. The Duchy of Jülich, on the other hand, became part of Pfalz-Neuburg and later passed to the Electors of Bavaria.

As regards the remaining Limburg territories, from 1548 onwards the possibility of lodging appeals with the *Schöffenstuhl* of Aachen and the *Reichskammergericht* were expressly viewed as symbolic of the orientation towards the Holy Roman Empire rather than the Habsburg Netherlands. Nonetheless, also the more prominent seigneurs of these territories were always scheming to expand their power. During this period, they endeavoured to acquire increasingly broader '*de non evocando et de non appellando*' privileges, thus optimally restricting appeals with the higher imperial courts and promoting the creation of their own appellate colleges. This tendency can not only be observed in the Duchies of Jülich and Cleves[6], but also in the Prince-Bishoprics of Liège[7] and Cologne[8]. In a later period, lesser seigneurs, such as those of Rekem, Gronsveld and Thorn, followed their example.

---

5 K.J.Th. Janssen de Limpens, *Rechtsbronnen van het Gelders Overkwartier van Roermond*. Werken OVR, 3e series, vol. 20, Utrecht, 1965, lxxviii, 433-437 (draft code 1532); lxxx, 440-457 (draft code 1564); A.Fl. Gehlen, 'De Guliks-Bergse "Rechtsordnung und Reformation" van 1555', *De Maasgouw*, 1992, 13-20.

6 A.Fl. Gehlen, 'De Guliks-Bergse "Rechtsordnung und Reformation" van 1555'; see also H.G. Adenauer, *Die Entwicklung der Obergerichte in Jülich-Berg in der Zeit von 1555 bis 1810*, Inaugural-Thesis, Düsseldorf, 1969, 4-14.

7 *Privilegia de non appellando* 1515, 1521, 1571, see G. Hansotte, *Les Institutions politiques et judiciaires de la Principauté de Liège aux temps modernes*, Brussels, 1987, 180-181.

8 G. Buhlmann, *Der kurkölnische Hofrat 1597 bis 1692. Entshehungsgeschichte und Rechtsgrundlagen*. Rheinisches Archiv. Veröffentlichungen des Instituts für Geschichtliche Landeskunde der Rheinlande der Universität Bonn, vol. 138, Cologne, 1998, 68: *privilegia de non appellando* from 1570 until 1653.

The emergence of appellate colleges, such as the Council of Brabant at Brussels (1430), the Court of Guelders at Arnhem (1544), the *Commissaires-Deciseurs* at Maastricht (1537), the *Reichskammergericht* at Regensburg (1495), the Court Councils (*Hofrath*) of Düsseldorf, Cleves and Cologne and the *Conseil ordinaire* at Liège (1518),[9] each with their own ordonnances governing legal proceedings, in the areas belonging to the Burgundian Circle as well as those belonging to the Lower Rhine-Westphalian Circle, offers a wealth of possibilities for a comparative approach to procedural law and legal practice.

4. The Codification of Customary Law in the Areas Between the Rivers Meuse and Rhine

A second major aspect of seigneurial centralisation and modernisation is the endeavour to codify customary law, as initiated in 1531 and 1540 by Charles V in the Netherlands. In so doing, he kept in step with a development that had started in the Meuse-Rhine Euroregion almost simultaneously and earlier had led to results, which were more important in terms of quality. There were early attempts, especially in the lands possessed by the Dukes of Jülich, to arrive at a comprehensive land law. These attempts were supported by the assembly of the ducal Estates.[10] Such innovations were encouraged by jurists from the circle of Duke Wilhelm of Jülich-Berg, referred to by the nineteenth-century jurist Romeo Maurenbrecher as the 'Justinian from Berg', who governed from 1539-1592. These '*Erasmusjünger*' had studied at the Universities of Cologne and Louvain and were well-versed in legal humanism.[11] Soon, the 1537 *Jülicher Landrecht* (Land Law) drafted by them was adapted to the new procedural law, which needed to be introduced as a result of the 1495 *Reichskammergerichtsordnung*. This made, among other things, that the 1555 *Gerichtsordnung und Reformation* of Jülich-Berg evolved into a codification of high quality.

9 The Conseil ordinaire was set up in 1518 with the object to limit lodging appeal with the imperial courts, Hansotte, *Les Institutions*, 178.

10 Unlike the situation in the Netherlands, where in particular under Philip II codifying customary law met with considerable resistance from the provincial states and the local public authorities.

11 E.M. Kloosterhuis, *Erasmusjünger als politische Reformer. Humanismusideal und Herrschaftspraxis am Niederrhein im 16. Jahrhundert*. Rheinisches Archiv, vol. 148, Cologne, 2006.

In 1538, during the rule of the Archbishop-Elector of Cologne, Hermann von Wied (1515-1546), virtually at the same time as the *Jülicher Landrecht* and in a similar intellectual climate, the first edition of the *Reformation* for the Electorate of Cologne appeared.[12] One objective was to make public the legislation issued by Maximilian I within the framework of the Imperial Reform (*Reichsreform*) in the Electorate of Cologne and the other to promote the reception of Roman Law.[13]

In comparison, the early codifications ordered by Charles V for the Lands of Outremeuse were, as '*travaux preparatoires*', insignificant in volume, content and completeness, so that they were never homologated.[14]

In the Prince-Bishopric of Liège, which was joined with the Electorate of Cologne under a personal union, major codifications, such as the *Reformation de Groesbeck* (1572), were also carried out at an early stage. Around 1620, Pierre de Méan, father of Charles de Méan, the "Papinian of Liège", produced a draft codification of customary law. Although his *Receuil des points marqués pour coutumes du Pays de Liège* was not homologated, it had a large degree of practical authority as a private codification.[15] The reformed land laws of Jülich (1555) and Cologne (1538/1663) and the law of Liège were greatly influenced by jurisprudence. The organisation of these codifications was often loosely based on the organisation of the Justinian Code, which was appropriate, as it followed a more scholarly approach towards the indigenous customary law ensuing from the *mos Gallicus*.[16] In this type of codification, procedural law is often paramount and substantive areas of law were added on at a later stage.

Unlike the relatively amorphous recordings of customary law in the Lands of Outremeuse, for example, these reformed codifications lend themselves for academic investigation by contemporary jurists as well, as is evidenced

---

12 https://www.deutsche-biographie.de/sfz70117.

13 R. Maurenbrecher, *Die Rhein-Preußische Landrechte*, Bonn, 1830, vol.. 1, 313-314; Kloosterhuis, *Erasmusjünger als politische Reformer*, 15-65.

14 K.J.Th. Janssen de Limpens, *Rechtsbronnen van het hertogdom Limburg en de Landen van Overmaze*. Werken OVR, vol. 1, Bussum, 1977.

15 A.Fl. Gehlen, 'Charles de Méan en het costumiere recht te Maastricht', in: *PSHAL*, vols. 136-137, 2000-2001, 407-420, in particular 410; Benoit Lagasse, 'Charles de Méan, de Luikse Papinianus. Biografie van een Luikse jurist uit de zeventiende eeuw', in: *Pro Memorie* 17 (2015) 251-263.

16 Modelled on Charles Dumoulin, *Commentarii in Parisiensis totius Galliae Supremi parliamenti consuetudines*, Parijs, 1538.

by Melchior Voet's commentary of the Jülich *Reformation* and the *Observationes* on the laws of Liège by Charles de Méan.[17]

The *Gerichtsordnung und Reformation* applicable in Jülich-Berg subsequently formed an important source of inspiration for the first concept of the 1620 Land and Town Statute Book of Guelders, which was drafted on the initiative of the Estates of Upper Guelders.[18] This codification was equally greatly influenced by jurisprudence. It no longer followed the organisation of the Codex, however, but is modelled instead on the Institutes, in which the law of persons, property law and the law of obligations precede the rules of procedure.[19] That this was a conscious choice is evidenced symbolically by the frontispiece designed by Peter Paul Rubens for the first edition of the Land and Town Statute Law Book commissioned by the Estates of Upper Guelders. The engraving features the Archdukes Albrecht and Isabella presenting the Country and Town Statute Book of Guelders, which shows the motto: '*Armis et legibus utroque clarescere pulchrum*', which was inspired by the *Constitutio Imperatoriam* of the Justinian Institutes (533). Because of its clear systematics, the Guelders Country and Town Statute Book in turn served as source of interpretation for the mother codification of Jülich-Berg.

More forms of cross-pollination in later codifications of customary law have become known in the meantime. The Land Law of Gronsveld (1671)[20], for instance, was for the major part a direct copy of the revised version of the 1663 *Reformation* of the Electorate of Cologne and the Maastricht codification of 1665 (*Recueil der Recessen*) was greatly influenced by the body of law of Liège. The *Recessen* came about as a result of

---

17 Maurenbrecher, *Die Rhein-Preußische Landrechte*, vol. I, 125, 337; M. Voets, *Historia juris civilis Juliacensium et Montensium*, Düsseldorf, 1729; Ch. de Méan, *Observationes et res iudicatae ad ius civile leodensium, romanorum, aliarumque gentium canonicum et feudale*, Liège, 1740-1741; see regarding the Land Law of the Electorate of Cologne also A.J. Gilgen, *Handbuch Erzstiftisch-Cölnischen Rechten*, Keulen, 1783.

18 A.Fl. Gehlen, 'De Guliks-Bergse "Rechtsordnung und Reformation" van 1555'; B. Van Hofstraeten, *Juridisch Humanisme en costumiere acculturatie. Inhouds- en vormbepalende factoren van de Antwerpse Consuetudines compilatae (1608) en het Gelderse Land- en Stadsrecht (1620)*, diss. UM, Maastricht, 2008.

19 On the system of the Institutes, see: J.E. Spruit, R. Feenstra and K.E.M. Bongenaar, *Corpus Iuris Civilis. Tekst en vertaling*, vol. 1: *Instituten*, Zutphen-'s-Gravenhage, 1993, p. xxii-xxiii; for the quoted passage from the *Constitutio Imperatoriam*, ibidem: 'The imperial majesty must not only be adorned with arms, but also with laws.'

20 The impact of the Cologne Reformation on the Land Law of Gronsveld was not accidental: Countess Anna Christina von Hardenrath, Lady of Gronsfeld, was the daughter of a mayor of Cologne (Th.J. van Rensch, *Licht op het Zonneleen Gronsveld*, diss. UM, Maastricht 2015, 372, text 829-891.

Frontispiece Country and Town Statute Book of Guedres by Peter Paul Rubens the collaboration between the Estates General of the Dutch Republic and the Prince-Bishop of Liège, who were the two overlords of Maastricht. Of special importance was the input by the jurist Charles de Méan of Liège, author of the only scholarly legal treatise on Maastricht customary law.[21]

As a result of the Enlightenment the codification of customary law received a new incentive during the eighteenth century, especially in Prussian Upper Guelders and in the religious foundation (*Stift*) of Thorn. The introduction in 1781 of the *Corpus Iuris Fridericianum*, as the precursor of the *Algemeine Landrecht* of 1794, made it necessary to make a number of prudent adaptations, in particular in Prussian Guelders, in the area of civil procedure and the organisation of the judiciary. This necessitated intense negotiations with the provincial Estates.[22]

In 1776, the abbess of Thorn, Maria Kunigunde of Saxony, was faced with a similar complex task. She was an enlightened abbess, and as such, she endeavoured to overhaul the organisation of the judiciary and to modernise the Thorn Land Law, which dated from the sixteenth century. The new codification of the Land Law of Thorn (1788), drafted under the auspices of the abbess, was mainly based on the 1620 Country and Town Statute Book of Guelders.[23]

Notwithstanding this growing territorialisation, it may be cautiously concluded that during the entire period of the *Ancien Régime* in the territories between the rivers Meuse and Rhine a certain degree of mutual influence remained between the legal systems of the Burgundian and the Lower Rhine-Westphalian Circles.

---

21 A.Fl. Gehlen, 'Charles de Méan', 407-420; B. Lagasse, 'Charles de Méan'.

22 A.M.J.A. Berkvens, 'Peter Heinrich Coninx als codificator van het burgerlijk procesrecht in Pruisisch Gelre, 1783-1786', in: M. Gubbels and C.J.H. Jansen (eds.) *Regio. Rechtshistorische opstellen aangeboden aan dr. P.P.J.L. van Peteghem*. Rechtshistorische Reeks Gerard Noodt Instituut, vol. 53, Nijmegen, 2010, 235-266.

23 A.M.J.A. Berkvens, 'Het Nieuwe Landrecht van Thorn. Verlicht absolutisme en codificatie onder het bewind van de vorstin-abdis Maria Cunegonda van Saksen', in: E.C. Coppens (ed.), *Honoris causa. Opstellen aangeboden aan prof.mr. O. Moorman van Kappen ter gelegenheid van zijn vijfentwintigjarig jubileum als hoogleraar aan de Faculteit der rechtsgeleerdheid van de Katholieke Universiteit Nijmegen*. Rechtshistorische Reeks Gerard Noodt Instituut, vol. 42, Nijmegen, 1992, 1-28.

5. Continuing Territorial Disintegration of a Border Area

In the course of the seventeenth and eighteenth century, the more extensive seigneurial territories, such as Spanish Guelders and the Spanish Lands of Outremeuse, that still existed during the reign of Charles V, were divided up between new rulers. The existing provincial institutions and privileges were preserved without undergoing any tacit or express changes.

To the newcomers in the area between the rivers Meuse and Rhine, such as the Dutch Republic [24] and Prussia,[25] it meant that they needed to fit these lands into their own executive and judicial organisation while respecting the institutions and traditions that existed locally. From 1713 onwards, the same applied to Austria, which gained sovereignty over the Southern Netherlands after the War of the Spanish Succession: Charles VI, Maria Theresa, Joseph II and their successors were equally obliged to respect the existing institutions and privileges, as Joseph II would experience at his cost in 1789 during the Brabant Revolution. The international treaties under which these partitions were determined, limited the degrees of freedom of their new rulers in organising the administration and the justice system of their new territories in various ways. Below, I will describe the following three examples: the capitulation of Maastricht (1632); the Lands of Outremeuse and the 1661 Partition Treaty; and Spanish Guelders and the Peace of Utrecht (1713).

On the occasion of the capitulation of Maastricht on 22 August of 1632, it was determined that within Maastricht and the district of Maastricht the Dutch States-General would only succeed to the rights to which Philip IV of Spain had been entitled in his capacity of Duke of Brabant. The rights of the Prince-Bishop of Liège, as joint overlord of Maastricht, were not to be diminished in any way whatsoever (Article 6 of the Capitulation Treaty). The States-General were to rule over Maastricht, jointly with the Prince-Bishop, 'as a state and a separate province, not associated with other states or provinces of either one of the said lords… in judicial matters, in jurisdiction and police matters ('*als een staet ende provincie particulier,*

---

[24] Maastricht (since 1632), the States Lands of Outremeuse (since 1644) and States Upper Guelders (since 1715).

[25] Cleves (since 1614) and Prussian Guelders (since 1713).

verscheyden van andere staten ofte provinciën van beyde de voorz. heren ... soo in materie van justitie, als resort ende politie') (Article 7).[26] Upon the conquest of Maastricht by Frederik Hendrik, it did not become a generality land, but remained a separate territory ruled by two overlords. Where its executive and judicial organisation were concerned, this meant that the old order established jointly by its original overlords in 1580 was preserved.[27] Changes could only be introduced upon agreement between the two overlords, so that virtually no innovation was brought about before the end of the *Ancien Régime*.

The situation in the Lands of Outremeuse was different. Since the conquest of Maastricht in 1632, these territories had become of strategic importance for the Dutch Republic. The States-General occupied the principal towns of Valkenburg, Dalhem and 's-Hertogenrade from 1644 onwards and based their claim to sovereignty over the Lands of Outremeuse on that fact. Spain, on the other hand, possessed the town of Limbourg on the river Vesdre and took the position that the Duchy of Limbourg and the Lands of Outremeuse were fully dependent on Limbourg on the Vesdre as the principal town of the 'province of Limbourg'. Both the Republic and Spain laid claim therefore to the sovereignty of the Lands of Outremeuse. For quite a while, the issue remained impossible to resolve and its resolution was put aside in the Peace of Münster (1648).[28] The issue was definitively settled by the Partition Treaty of 26 December 1661.

26 P.J.H. Ubachs, *Twee heren, twee confessies*. Maaslandse Monografieën, vol. 21, Assen, 1975, 443; this regulation was affirmed in Article 2 of the Peace of Münster: '… [the] City of Maastricht and its jurisdiction, as well as the County of the Vroenhof, … will continue to belong to the States-General, who shall have full sovereignty and superiority' ('... En sullen ... [de] stad van Maastricht, het ressort van dien, als ook het graafschap van de Vroenhof, ..., blyven aan de voorschreeve Heeren Staaten, in alle en deselve regten en deelen van souvereiniteit en superioriteit ...'), C. Smit, *Het vredesverdrag van Munster*, Leiden, 1948, 32.

27 A.Fl. Gehlen, *Het notariaat in het tweeherig Maastricht*. Maaslandse Monografieën, vol. 33, Assen, 1981, 20-37.

28 Article 2 Peace of Münster: 'The [actual situation] of the three Lands of Outremeuse, namely Valkenburg, Daalhem and 's-Hertogenrade, shall remain unchanged. Possible disputes with regard to this shall be brought before the Chambre mi-partie, […..] to be adjudicated by that chamber.' ('Wat aangaat de drie quartieren van over-Maaze, te weten, Valkenburg, Daalhem en 's-Hertogenrade, deselve sullen blyven in den staat in dewelke sig jeegenwoordig vinden: ende in cas van dispute en contraversie, sal deselve gerenvoyeert werden aan de Chambre mi-partie, [….] omme aldaar te werden gedecideert'),Smit, *Het vredesverdrag van Munster*, 32.

In the years in between, both Spain and the Republic attempted to as much as possible retain their authority over the Lands of Outremeuse. The States-General ruled the part they occupied as a generality land; in the remaining areas, the Council of Brabant at Brussels exercised its authority as much as possible. Both parties assigned judicial officers, stewards and collectors to the Lands of Outremeuse. They obliged the inhabitants, through retorsions and reprisals, to recognise their sovereignty and jurisdiction, which led to double taxation and contributions, and judicial decisions and judgments the other party would not recognise, which jeopardised the population's legal certainty.[29]

Most visible to those seeking justice was the legal uncertainty brought about by the competing Council of Brabant at Brussels and the States Council of Brabant at The Hague. From 1644 onwards, the States-General prescribed that appeal lied with its Council of Brabant, which has been set up in the Hague in 1591 for the conquered territories of the Duchy of Brabant. Unlike its Brussels counterpart, it also served as a feudal court.

Retorsional placards were issued on the part of the Southern Netherlands as well as the Northern Netherlands, under which the inhabitants of the Lands of Outremeuse were prohibited by each side, on pain of prosecution on the grounds of rebellion, to accept the jurisdiction of the hostile court of the other. Choosing the wrong forum could thus lead to loss of right.[30]

---

29 See on the part of Spain, inter alia, the placard of 26 January1644 'against the inhabitants of the Lands of Outremeuse, who have obtained a provisional ruling from the States Council of Brabant at The Hague' ('teghens d'inghesetenen van den Lande van Overmase, die eenighe provisie van justitie hebben gheimpetreert inden Rade van Brabant inden Haghe') (*Plakaten van Brabant*, dl. 2, f° 410); placard of 2 September 1647 concerning the prohibition imposed on the inhabitants of the Lands of Outremeuse to obey those who present themselves falsely as administrative and judicial officers.' ('behelsende verboth aen de inghesetenen van den Lande van Overmaese van te gehoorsaemen de bevelen van deghene hun toeschrijvende de qualiteyt van officieren van de drye landen van Overmaese') (*Plakaten van Brabant*, dl. 3, f° 7-8); placard of 23 March 1651 to preserve the competence of His Majesty'('tot conservatie van de jurisdictie van Sijne Majesteyt') (*ibidem*, f° 8-9); placard of 8 March 1654, by which the assaults (attentaeten) committed within the Land of Outremeuse were annulled by the States-General of the United Provinces ('de attentaeten ghedaen binnen den Lande van Overmaese by de Staten van de gheunieerde provinciën') (*ibidem*, f° 11-14); on the part of the States-General the placard of 27 October 1633 concerning the granting of feuds in tenancy upon succession before the States Council of Brabant at the Hague (*Groot Placaatboek* (hereon after: GPB), vol. 1, 1463); placard of 8 March 1634 concerning the prohibition imposed on the inhabitants of the lands of Outremeuse to seek 'any provision of justice' in Brussels. (GPB, vol. 2, 1176), idem 5 November 1633 (GPB, vol. 2, 1478); idem the placards of 4 January 1648, 6 February 1648, 19 February 1648, 27 March 1648, 20 May 1648, 22 July 1648, 30 December 1648 and 21 December 1655 (GPB, vol. 2, 1181-1188).

30 B.C.M. Jacobs, 'Voor vijand en rebel gehouden. Ten onrechte procederen voor de Raden van Brabant in Brussel en 's-Gravenhage', in: Jacobs and Nève (eds.), *Hoven en Banken in Noord en Zuid*, 103-120.

This muddled situation ended as a result of the 1661 Partition Treaty. At the heart of this treaty lied 'the full free ownership, superiority and sovereignty' of the States-General and the Spanish King in those parts of the lands of Outremeuse which had been granted to them.[31] The legal consequences of the retorsional judgments from the period between 1644 and 1661 were annulled. The parties were promptly to restore movable and immovable property to their rightful owners (Article 7). 'Ordinary' judgments in civil cases, on the other hand, had to be mutually enforced, regardless of which court had passed these (Article 9). Cases that had not been fully disposed of were transferred to the newly competent court (Article 10). In addition, arrangements were made to avoid double taxation and for the offsetting of undue payments. At the current state of knowledge it remains an open question as to what was actually achieved in these instances.

Not held back by any guarantees with regard to the privileges enjoyed by the Lands of Outremeuse, on 15 October 1663 the States-General took conclusive decisions with regard to the organisation of the police and the justice system in those parts (*partages*) of the Lands of Outremeuse that had been granted to them.[32] They decided to entrust the administration to the drossards as high officials, as had been the tradition. These high officials also chaired the assemblies of the Estates of the three Lands of Outremeuse, which consisted of the nobility and chief justices (*hoofdbanken*), but no longer included the clergy. As regards the justice system, the States-General decided not to introduce to the parts they had been granted any major changes in the organisation of the judiciary at the levels of first instance and appeal. For the time being, they preserved the means of redress of revision (*reformatie*) by the States Council of Brabant at the Hague.[33] They did, however, decide to introduce a uniform way of conducting proceedings in the three Lands of Outremeuse. This was predominantly based on the existing rules governing the style of proceed-

31 J.A.K. Haas, *De verdeling van de Landen van Overmaas*. Maaslandse Monografieën, vol. 27, Assen, 1978, 294-314.

32 A.Fl. Gehlen, 'De inrichting van de justitie in de Staatse Landen van Overmaze ... 15 oktober 1663', in: *Brabants recht dat is* ... Brabantse Rechtshistorische Reeks, vol. 5, Assen, 1990, 109-120. For the placards in question, see: *GPB*, vol. 2, 3091-3132.

33 *GPB*, dl. 2, 3095-3096 (Valkenburg); *GPB*, dl. 2, 3101-3102 Dalhem (including a change in jurisdiction and the rules of procedure); *GPB*, vol. 2, 3107-3108 (including a change in jurisdiction and the rules of procedure).

ings as promulgated by Philips IV in 1634 for the Land of Valkenburg. And finally they issued a uniform protocol for criminal proceedings, in which the proceedings as conducted by the States Council of Brabant became the norm.[34]

Prior to promulgating the placards in question, the States-General convened a joint meeting of the Estates of the three Lands of Outremeuse in Maastricht on 23 October 1663 with the object of ensuring broader support for their measures. In so doing, the States-General made an exception to their policy with regard to the Generality Lands, in which, generally speaking, they no longer convened the traditional representative bodies. They did so, because they wished to ensure the loyalty of, in particular, the nobility, as the possessors of the various seigneuries.[35]

As their sovereign, they presumed the right to appoint Protestants to public offices, who in practice had Catholics to stand in for them. The introduction of the *Echtreglement* (Regulation governing matrimony) of 18 March 1656 and the Calvinisation of the scabinal benches (*schepenbanken*) met with considerable resistance from the population.[36]

As a result of the Partition Treaty, on the Spanish side nothing changed in essence with regard to the administrative and judicial organisation of its share of the Lands of Outremeuse, of course, so that, in practice, the administrative and judicial colleges in the Lands of Outremeuse doubled and from then on operated side by side.

---

34 *GPB*, vol. 2, 3129-3130, Article 6.

35 Maastricht, Regionaal-Historische Centrum Limburg (hereon after: RHCL), Archief Landen van Overmaze (hereon after: LvO), entry number 3616 (nonfoliated): List of Noblemen: On 22 October 1663 the following noblemen appeared in 'De Landtskroon' [the former town hall of Maastricht] to be heard in the matter of their claimed membership of the nobility of the Lands of Outremeuse.' ('Den 22e octob. 1663 sijn de volgende edelen gecompareert op de Landtskroon ter beschrijvinge van Haar Ed. Mog. om gehoord te worden over het recht dat sij pretenderen te hebben onder 's landts ridderschap ende edelen.'); 'Meeting of the three Estates of the Lands of Outremeuse, convened by the Seigneur van Raesfelt van Goeze, acting by proxy on behalf of the States-General, on 23 October 1663' ('Vergaederinge van de Staten der drij landen van Overmaese onder het district van HHM ter convocatie van de Heer van Raesfelt van Goeze, gecommitteerde van HHM, den 23 oktober 1663').

36 W.A.J. Munier, 'De calvinisering van de colleges van schout en schepenen in de Landen van Overmaas vanaf de verovering van Maastricht tot aan het einde van de zeventiende eeuw', in: PSHAL, vol. 123, 1987, 42-145; idem, 'De toepassing van het Echtreglement in de Landen van Overmaas en de neerslag ervan in trouwregisters', in: *PSHAL*, vol. 121, 1985, 69-103. For the text of the Regulation governing matrimony (Echtreglement), see: C. Casier and L. Crahay (eds.), *Coutumes de Duché de Limbourg et des Pays d'outre-Meuse*, Brussels, 1889, 341-395.

After the War of the Spanish Succession (1702/1713) it was the turn of Spanish Guelders to be partitioned. In comparison with the treaties discussed above, remarkable dissimilarities are found. Whereas the Peace of Münster and the Partition Treaty did not offer protection to the Catholic faith and the privileges of its subjects, these were now expressly protected.

Under the 1713 Peace of Utrecht and the 1715 Antwerp Barrier Treaty Spanish Guelders was divided up between Austria, Prussia and the Dutch Republic, while a small part of the territory was given up to Jülich.

As regards Prussian Guelders, it was decided in 1713 that no changes were to be introduced in matters of religion and that the Catholic faith was to be preserved; the ecclesiastical jurisdiction of the Bishop of Roermond was not to be affected; if anything, the Protestant King of Prussia had to contribute to the maintenance costs of the Catholic bishop. The King was obliged to preserve all rights and privileges ensuing from the 1543 Venlo Treaty, the Guelders variant of the Brabant Joyous Entry (*Blijde Inkomst*). This led, *inter alia*, to the rule that only Catholic inhabitants of the Lands were to be appointed to public offices. This obligation applied to both the régence (the general administration) and the local administrative institutions. Frederick William I of Prussia was to ensure the establishment of a '*Tribunal de Justice*', so that the Estates and the subjects could not be sued outside the territory.[37]

Similar arrangements were made for States Upper Guelders through Article 18 of the Antwerp Barrier Treaty, which provided that the States-General of the Dutch Republic were prohibited from making any changes to the *statuta*, customs and privileges, either spiritual or secular. They were obliged to respect the public practice of the Catholic faith and to only appoint Catholic inhabitants of the Lands to public offices. In addition, Article 18 provided for the possibility of establishing an appellate court for the benefit of the inhabitants of States Upper Guelders. This was in no way an obligation, however. In Austrian, Prussian and States Guelders, as a result of these treaties and unlike the situation in the States Lands of Outremeuse, in the eighteenth century the administrative and justice systems

---

37 Th.J. van Rensch, 'Het Hof van Justitie van Pruisisch Gelre', in: *PSHAL*, 113, 1977, 193-268, in particular 197; the conditions under which the Emperor was prepared to relinquish his claims to Prussian Guelders have been laid down in the Treaty of 2 April 1713 (Lamberty, *Memoires pour servir à l'histoire du XVIII siècle*, vol. 8, 's-Gravenhage, 1730, 45-48).

continued on principle to be rooted in the system that had evolved in Spanish Guelders in the course of the seventeenth century.

In practice, Prussia and the Dutch Republic interpreted these provisions differently. In Prussian Guelders, as a result of the invocation of the privileges, a strong representation of nobility and towns remained. These Estates assemblies managed to set up a *Justiz-Collegium* as an appellate college for Prussian Guelders and force the appointment of Catholic justices. Contrary to the letter and the spirit of the Treaty of Utrecht, Frederick William I of Prussia did manage, however, to introduce the shared-administration model of '*Regierungen und Kammern*', as was customary in Prussia, in which the judicial chambers, as the *Regierung*, only had judicial powers, while matters of government and finance were entrusted to what was referred to as the 'War and Domains Chambers' (*Kriegs- und Domänenkammern*).[38] In spite of what was agreed upon under the Peace of Utrecht, the King appointed a number of Protestants to the Domain Chambers, so that fierce competition existed between the Catholic Judicial Chambers and the Protestant Domain Chambers. In addition to this, Fredrick II of Prussia charged the Councillors of the Cleves Court (*Hofrat*) with the supervision of the judgments passed by the *Justiz-Collegium*, by appointing the Cleves Councillors as commissioners-revisors to review judicial decisions passed by the *Justiz-Collegium* on appeal.

Inasmuch as the seigneurial rights to a large part of Guelders, as part of the estate of the Stadtholder-King William III, had passed to Frederick William I of Prussia, the power, which the States-General were able to exercise until 1769 over a large part of States Guelders was, in effect, marginal, due to the exercise of the Prussian rights to the Land of Montfort as a former seigneurie of the House of Orange-Nassau by the *Kriegs- und Domainenkammer* at Guelders.[39]

Although, in view of the Barrier Treaty, this would have been logical, the States-General formed no indigenous representative body in States Guelders, but the 'High-Mighty Lords' shared the exercise of the undivided administrative powers with the Council of State at The Hague. They

---

[38] A.M.J.A. Berkvens, *Plakkaten, ordonnanties en circulaires voor Pruisisch Gelre (1713-1798)*. Werken LGOG, vol. 22, Maastricht, 2012, p. xvii-xxiii.

[39] E.J.M.G. Roebroeck, *Het Land van Montfort*. Maaslandse Monografieën, vol. 6, Assen, 1967, 110-142.

did, however, in 1717 establish, after some hesitation, a States Court of Guelders in Venlo as an appellate college for their part of Guelders and thus fleshed out the option provided by Article 18 of the Barrier Treaty. After the Peace of Utrecht, the seventeenth-century procedural ordonnances of the Sovereign Council (*Souvereine Raad*) of Roermond served as a point of departure for the newly established appellate courts of justice at Geldern and Venlo.[40]

6. The Effects of the Territorial Fragmentation on the Organisation of the Judiciary and the Administration of Justice

What now were the effects of the territorial fragmentation of the Limburg territories for the organisation of the judiciary and the administration of justice? First and foremost, it may be concluded that as a result of the territorial partitions a growing number of sovereigns became involved in their administration, legislation and justice system. The most striking aspect in this context is, of course, the major additions to the judicial institutions, which operated side by side. A substantial increase can be observed in the number of administrative and judicial institutions especially at the provincial level, at which the judicial institutions, even if their territory was small, served as an intermediary court of appeal, or even as a court of last resort.

If we restrict ourselves to Maastricht, the Lands of Outremeuse and the Upper Guelders, we find the following judicial institutions that were competent to pass last-resort judgments:

- the *Commissaires Déciseurs* of Brabant and Liège at Maastricht. From 1632 onwards, the *Commissaires Déciseurs* of Brabant were no longer appointed by Brussels, but by the States-General at the Hague;
- the Council of Brabant at Brussels as a court of last resort for the non-partitioned parts of the Lands of Outremeuse, which from 1644 onwards suffered competition from the States Council and Feudal Court of Brabant at The Hague.

From 1661 onwards, this situation continued as a result of the Partition Treaty. In addition to this, the Feudal Court of Valkenburg and the newly

40 A.M.J.A. Berkvens, *Procesgids Soevereine Raad, Justiz-Collegium en Staats Hof van Gelre te Venlo*, Hilversum, 2011.

established chief courts of Gulpen and Daelhem as intermediary courts of appeal served within the Dutch part of the Lands of Outremeuse by virtue of their *competentio primae appellationis*. In the Spanish, and subsequently Austrian, Lands of Outremeuse, the Spanish Feudal Court of Valkenburg, which had been relocated to Oud-Valkenburg, and the chief courts (*hoofdbanken*) of Gravenvoeren and 's-Hertogenrade, also served as intermediary appellate courts, in their case, however, for the Council of Brabant at Brussels.

Since 1580, Spanish Guelders had the Sovereign Council of Roermond, which in the eyes of the Spanish government was the legitimate continuation of the Guelders Court at Arnhem. After the War of the Spanish Succession, the jurisdiction of the Sovereign Council was reduced to the advantage of the *Justiz-Collegium* at Geldern and the States Court of Guelders at Venlo.

If we also include Jülich, Cleves and Liège in our observations, we can add the councils of Düsseldorf, Bonn and Cleves as appellate colleges, as well as the chief courts of Kuringen and Vliermaal for the County of Loon.

The Aachen *Schöffenstuhl* continued to be of relevance for the remaining immediate territories until the end of the *Ancien Régime*, although its jurisdiction was reduced regularly. One reason for this was that the possessors of smaller immediate territories began to introduce their own recourse to appeal, as was the case, for instance, in Thorn[41] and Gronsveld[42]. Appeal from these imperial territories with the *Reichskammergericht* and the *Reichshofrat* continued to be possible, although, as the privileges *de non appellando* continued to expand in the eighteenth century, such appeal became more and more the exception and in some instances no longer possible.

41 A.M.J.A. Berkvens, '"Zur Abkurtzung unnötiger Processualweitlaufigkeiten", Een bijdrage over de geschiedenis van het Hof van Appel te Thorn, 1718-1795', in: A.M.J.A. Berkvens and Th.J. van Rensch (eds.), *'Wordt voor recht gehalden'. Opstellen ter gelegenheid van vijfentwintig jaar Werkgroep Limburgse Rechtsgeschiedenis (1980-2005)*. Werken LGOG, vol. 19, Maastricht, 2005, 211-240.

42 Van Rensch, *Licht op het Zonneleen Gronsveld*, 437-440.

7. What Was the Significance for Those Seeking Justice?

What did all this mean for those seeking justice? In the absence of any systematic research into procedural practice, I will limit myself to a few examples from the States Lands of Outremeuse and the *Stift* of Thorn.

Since the fifteenth century, there was a *Keurkeulse Mankamer* (*Kurkölnische Mannkammer*) in Heerlen, which was a lower feudal court of the Electoral Prince-Bishop of Cologne. It was charged with the administration of justice relating to some fifty feuds, which were dependencies of the Prince-Bishop of Cologne, predominantly in the Land of Valkenburg and the Duchy of Jülich.[43] Seen through medieval eyes, this was not a problem, as several different jurisdictions could function side by side and exclusive public authority was unknown. With the emergence of modern views of sovereignty, these feuds were seen, however, as 'extraneous feuds', or *buitenlenen*, falling within the sovereignty of the Lands of Outremeuse and the Duchy of Jülich, respectively. In consequence, not only the Prince-Bishop of Cologne as their feudal lord, but the Dukes of Brabant and Jülich, in their role as sovereigns, could, in principle, also claim jurisdiction over the masculine feuds (*manlenen*) of the Electorate of Cologne in their territories.

This led to problematic situations from the moment the Dutch States-General began to lay claim in earnest to the sovereignty of the Lands of Outremeuse, that is to say since the conquest in 1644 of its principal town of Valkenburg. From this period onwards, the States-General ordered its chief judicial officer, the *schout* (*scholtis*) of Heerlen to prohibit the possessors of feuds which depended on the *Kurkölnische Mannkammer* to bring their actions in personal and property matters before this court. There were also problems with regard to appellate proceedings, inasmuch as the States Feudal Court of Valkenburg considered appeal against judgments passed by the *Kurkölnische Mannkammer* to fall within its own jurisdiction, thus ignoring the Chancery of the Prince-Bishop at Bonn.[44]

---

43 In addition to this, the Kurkölnische Mannkammer ruled two seigneuries in the Prince-Bishopric of Liège (Hamal and Gruitrode) and one in the County of Wittem, Ter Poorten.

44 By Decree of 21 July 1658, in the case of Johan Rusius, the appellant against Baroness, the Lady of Clairmont, the respondent, the Feudal Court of Valkenburg ordered to retrieve, if need be by force, the documents from the Registry of the Kurkölnische Mannkammer and to transfer these to Valkenburg (RHCL, LvO, entry nr.. 6451).

As a result of the Partition Treaty, the chicaneries around the *Kurkölnische Mannkammer* recommenced, because the accompanying settlement (*accommodement*) seemed to suggest that also feuds belonging to foreign lords were subject the Treaty.[45] Although this provision was evidently addressed to the two contracting sovereigns, i.e. the States-General and Philip IV of Spain, the States Feudal Court of Valkenburg also made use of the provision to interfere in the jurisdiction of the *Kurkölnische Mannkammer*, basing itself on the fact that the feuds of the Electorate of Cologne were predominantly situated in the States Land of Valkenburg. In a number of instances, the machinations of the States *scholtis* of Heerlen and of the Feudal Court of Valkenburg led to enforcement requests from the *Kurkölnische Mannkammer* addressed to the Chancery of the Prince-Bishop of Bonn. In the period between 1660 and 1699, this led to diplomatic steps by the envoy of the Electorate of Cologne at the States-General in The Hague. It also inspired the Lady of Terheiden in the Duchy of Jülich to resort to similar chicaneries against the *Kurkölnische Mannkammer* at Heerlen. She ordered the *scholtis* of Ter Heiden to bring first-instance cases from the feuds that belonged to the Electorate of Cologne within her jurisdiction before the scabinal bench of Horbach in first instance and before the highest court in Jülich, the *Obergericht* on appeal.[46]

Around 1684, a tacit compromise was reached on the jurisdiction of the *Kurkölnische Mannkammer*, which meant that the *Mannkammer*'s competence in first instance was preserved, but in cases involving feuds that 'lied within the territory or jurisdiction of Brabant' appeal lied with the States

45 The 'accommodement' (settlement) concerning the 'par le menu' (detailed) partition of the Lands of Outremeuse, which was part of the Partition Treaty, provides for the Land of Valkenburg that the States-General will be granted the enjoyment of their share of these, including all hamlets, jurisdictions, competences, judicial claims, feuds, taxes payable to the seigneur (bede), domains and other appertaining regalia and revenues'('met alle de gehuchten, ressorten, jurisdictiën, gerechtigheden, leenen, vasallagiën, beeden, domeijnen ende andere regaliën ofte revenuen daartoe specterende') (Haas, *De verdeling van de Landen van Overmaas*, 309; see also Partagetractaat, Article. 12: 'The feuds, both internal and external, which will also be partitioned under the Treaty, shall be reacquired (relief) by way of commendation before the feudal courts or feudal chambers of the other sovereign, where they were so reacquired before.' ('De leenen, soo binnen de voors. drie landen als buyten deselve gelegen, die met dese partage oock verdeelt worden, sullen absolutelick releveren van de leenhoven ofte leencamers van den souverain daeronder deselve vallen sullen, sonder eenige verdere dependentie van de leenhoven ofte leencamers van den anderen souverain daeraen zij voor desen mochten hebben gereleveert'.) In July 1663, the scholtis van der Schuyr clearly interprets these provisions as if they were also applicable to feuds that fell within the jurisdiction of the Kurkölnische Mannkammer RHCL, LvO, entry nr. 6437).

46 RHCL, LvO, entry number 6443, proceedings brought by Johanna Fors, the widow of Johannes Will, against the Free Lady van Bongard zur Leyen, Lady of Ter Heiden (1671-1673).

Feudal Court of Valkenburg. Subsequently, revision (*reformatie*) lied with the States Council of Brabant at The Hague. Only the cases from the feuds outside the limits of States Valkenburg were subject to 'the sovereignty of the Empire or Bonn'.[47] Such a compromise seems not to have been made in connection with the feuds of the Electorate of Cologne within the Duchy of Jülich.[48]

The inevitable conclusion must be that the territorial fragmentation caused the administration of justice to be an extremely uncertain affair. Judgments passed in the first instance by the *Kurkölnische Mannkammer* could be overturned by judgments of the Feudal Court of Valkenburg, the *Obergericht* of Jülich or the Chancery at Bonn. In addition, the States Council of Brabant, the *Hofgericht* (Electoral Feudal Court) at Düsseldorf or the *Reichskammergericht* were competent to revise these.

Similar problems concerning the continuation of medieval legal relations in the early modern period existed elsewhere in the territories along the river Meuse. The *Stift* of Thorn serves as a clear example. In the period between the sixteenth century and the beginning of the eighteenth century, the territorial position of Thorn continued to be ambiguous. For a long time, the abbesses of Thorn swayed between being dependent on Spanish Guelders and the Prince-Bishopric of Liège, on the grounds of their custodial rights (*Vogtei*), on the one hand, and sovereignty as a member of the Lower Rhine-Westphalian Circle, on the other. This uncertain territorial status of Thorn was also manifest in the administration of justice, inasmuch as in the sixteenth and seventeenth century the *schepenen* of Thorn, in their capacity of judicial authorities, could still refer civil cases (*hoofdvaart*) to, and lodge appeal with, the scabinal bench of Echt of Spanish Guelders, whereas on the grounds of serving as the *Stift*'s Vice-custodian (*Untervogt*), the Prince-Bishop of Liège, in his capacity of Count of Horn, also claimed jurisdiction in criminal matters in Thorn. When both the Sovereign Council at Roermond and the Prince-Bishop of Liège began to interpret this legal relation with the *Stift* of Thorn as the right to have sovereignty over it, it caused the abbesses to seek recognition of the *Stift* as an imperial estate (*Reichsstand*) in order to be entitled to military protection

[47] RHCL, LvO, entry number. 6395, nr. 147.

[48] In 1762 the Kurkölnische Mannkammer was again forced to defend its competence in the first instance against violations by the Court of Ter Heiden (RHCL, LvO, entry number. 6395).

by the Lower Rhine Westphalian Circle. In 1688, when the steward (*rentmeester*) of the Stift's Chapter (*Kapittel*), against the will of abbess Anna Salomé van Manderscheid, lodged an appeal, by way of Echt, with the Sovereign Council at Roermond, this led, among other things, to this unfortunate person being taken prisoner by order of the abbess, upon which the Spanish-Guelders troops burst into the abbey of Thorn and armed grenadiers did not even shy away from entering the sleeping quarters of the high-nobility abbess.[49]

The dispute over the territorial status of Thorn was definitively settled at last by the Barrier Treaty of Antwerp, when it was provided that the Dutch States-General could not claim sovereignty over Thorn on the basis of their newly acquired jurisdiction over Echt. Only then, it was conclusively established that the *Stift* of Thorn was an immediate imperial territory falling within the Lower Rhine-Westphalian Circle and that the abbess of Thorn was to have a voice and a seat on the Bench of Imperial Prelates of the Lower Rhine-Westphalian Circle. Thus, the abbess of Thorn was also given the possibility to set up her own appellate court for Thorn.[50] From this moment on, appeal from this appellate court lied with the *Reichskammergericht* at Wetzlar or with the *Reichshofrat* at Vienna. Also in this case, uncertain territorial relations contributed therefore to legal uncertainty for those seeking justice.

8. Identical Problems in Relation to the Law of Criminal Procedure in the Seventeenth Century

The abundance of jurisdictions cannot conceal that the various authorities were faced with virtually identical problems. The problems with regard to the low quality of justice passed by the various scabinal benches are very recognisable. This led, *inter alia*, to substantial problems in the area of criminal procedure. How did the legislator attempt to find solutions for these problems?

---

49 A.M.J.A. Berkvens, 'De kanunnik Joannes Stuben als auctor intellectualis van het 'Rijksvorstendom' Thorn (1688-1721)', in: A.M.J.A. Berkvens *et al.* (ed.), *Wordt voor recht gehalden*. Werken LGOG, vol. 19, Maastricht, 2005, 191-210, in particular 201.

50 A.M.J.A. Berkvens, '"Zur Abkurtzung unnötiger Procesualweitlaufigkeiten"', 211-240.

Initially, the 1570 *Criminele Ordonnantiën* (Criminal Ordonnances) of Philip II served as the point of departure for the Habsburg Netherlands. The 1532 *Constitutio Criminalis Carolina* served the same purpose for the territories belonging to the Lower Rhine-Westphalian Circle. For a long time, the issue of the operation of the procedural and substantive provisions of the Criminal Ordonnances was subject to debate in the Northern Netherlands. It is certain, however, that they have had a far-reaching influence on legal practice in the Dutch Republic through jurisprudence. In the Southern Netherlands, the core of the Criminal Ordonnances found its way into the 1611 Perpetual Edict (*Eeuwig Edict*) and from there it passed to, for instance, the 1620 Country and Town Statute Book of Guelders. Its Book VI contains a comprehensive chapter on criminal procedure. Thus, in Spanish Guelders, the problem of the '*pleine demie preuve*' (full semi-proof) was no longer relevant as a criterion on the basis of which it was possible to subject an accused to the rack. The 1611 Perpetual Edict also applied to the Spanish Lands of Outremeuse. Contrary to what was the case in Spanish Guelders, a generally applicable land law was absent here, so that subjecting a person to the rack was less strictly regulated, as the bands of robbers called the *Bokkenrijders* (goat riders) learned to their cost in the eighteenth century.[51] From 1663 onwards, the style of conducting criminal proceedings of the States Council of Brabant became prevalent in the States Lands of Outremeuse.[52] In practice, the scabinal benches relied more often on the view of the Dutch jurist Simon van Leeuwen[53] and his unfortunate interpretation of the Criminal Ordonnances of Philip II, so that most probably the rack was more widely used in these regions.

In the territories that belonged to the Lower Rhine-Westphalian Circle, the sovereigns made ample use of the 'Salvatorian Clause', which entitled them to promulgate their own laws in these matters. The 1572 *Reformation de Groesbeeck* of Liège, for instance, contains a separate chapter on criminal procedure; the same holds true for the 1555 Jülich *Reformation* and also lesser seigneurs, such as those of Gronsveld and Rijckholt, quite actively occupied themselves with making their own laws in this area. In

---

51 L. Augustus, 'Vervolgingsbeleid en procesvoering tegen de Bokkerijders. Het ontstaan van een waandenkbeeld', in: *PSHAL*, vol. 127, 1991, 69-153.

52 B.C.M. Jacobs, 'De Raad van Brabant en de procesgang in criminele zaken', in: *Colloquium Raad van Brabant*, 's-Hertogenbosch, 1985, 75-94.

53 S. van Leeuwen, *Manier van procederen in civiele en criminele saaken*, Amsterdam, 1688.

the 1538 Land Law of the Electorate of Cologne (*Kurkölnisches Landrecht*), on the other hand, the Carolina had been included in its entirety.[54] In the *Stift* of Thorn, the provisions on criminal procedure laid down in the Guelders Country and Town Statute Book, the Law of Loon-Liège and Imperial Law were freely consulted, however, in situations for which the Land Law did not provide, so that both the rights of the Duke of Guelders as Custodian (*Vogt*), those of the Prince-Bishop of Liège as Vice-custodian and the sovereign aspirations of the abbess carried weight there.[55]

9. Identical Problems, Similar Solutions? The Law of Criminal Procedure in the Eighteenth Century

In the eighteenth century, awareness began to grow in the area between the rivers Meuse and Rhine that the existing legal rules could lead to excessive results in the hands of those on the scabinal bench, who were not educated in law. Point of departure with regard to criminal procedural law in this area continued to be Book VI of the Guelders Country and Town Statute Book. Under the terms of the Peace of Utrecht, Frederick William I of Prussia was not authorised to unilaterally amend this body of law. His attempts to introduce the thirteen-chapter Brandenburg *Criminal-Ordnung*[56] of 8 July 1717 were met with objections from the Estates of Prussian Guelders. Since on the occasion of the transfer of power to Prussia in 1713 guarantees had been given that provincial privileges would be preserved, this *Criminal-Ordnung* could not be introduced and operate with direct effect in Prussian Guelders. On 15 October 1725, at the request of the *Justiz-Collegium*, and before his successor Frederick II definitively prohibited torture in all his possessions in 1740, Frederick William I promulgated further rules on the use of the rack and the degrees of territion.[57]

---

54 Maurenbrecher, *Rhein-Preußische Landrechte*, vol. 1, 315-316.

55 Berkvens, ''Cort onvertogen recht' te Thorn in 1651', in: E.C. Coppens *et al.* (ed.), *Fabrica juris, Opstellen over de 'werkplaats van het recht'aangeboden aan Sjoerd Faber*, Nijmegen, 2009, 91-108, in particular 107-108.

56 Text by Mylius, *Corpus Constitutionum. Marchicarum*, Theil II, Abtheilung III, Sp. 63-110 (to be consulted online via http://web-archiv.staatsbibliothek-berlin.de/altedrucke.staatsbibliothek-berlin.de/ Rechtsquellen/quellen.html).

57 J.M. Michiels, *Cesare Beccaria. Over misdaden en straffen*, Antwerpen, 1982, 89, note 1.

In order to avoid frivolous torture, the Edict of 15 October 1725 *inter alia* obliged the scabinal benches to forward to the *Justiz-Collegium* those interlocutory judgments, in which an accused was referred to the torture chamber.[58] Such judgments could only be executed upon the approval of the *Justiz-Collegium*. Following the Prussian example, this form of mandatory *rencharge* was also introduced in Austrian Guelders in 1727, but not in States Upper-Guelders.

Similar measures to render criminal procedure 'fool proof' were found in 1752 in the counties of Horn and Loon, which belonged to the Prince-Bishopric of Liège. Here, at the initiative of the Estates, the legislator took the rural *schepenen* by the hand and guided them through the criminal process with the aid of a regulation comprising nine chapters and fourteen instructive documents (*formulieren*), requiring them to present all critical decisions to the chief court of Vliermaal for its approval. This procedure was called *horsporter*.

In States Upper Guelders and the States Lands of Outremeuse, on the other hand, no comparable innovations to the criminal process are found. Although in 1762 the States-General needed to interfere directly when the *schepenen* of Echt, contrary to the rules laid down in the Country and Town Statute Book of Guelders, had sentenced Johannes Schmitz to death for petty theft on the basis of an incorrectly obtained confession as a result of torture,[59] the attempts by the Attorney-General (*momboir*) of the States Court of Venlo to review the law governing criminal procedure came to nought. Neither did the *enquête* by the States Brabant Attorney-General van Steelant in 1772 on the substandard state of the administration of criminal justice in the States Lands of Outremeuse lead to any meaningful innovations of criminal substantive and procedural law.[60]

---

58 This provision follows from the 'Rescript wegen Einsendung und Confirmation deren Sententien in criminalibus' of 2 March 1717 (Mylius, *Corpus Constitutionum Marchicarum*, Theil II, Abtheilung III, Sp. 109-110).

59 This, incidentally, produced the first known example of the annulment of a criminal judgment by the States-General, see: Th. Thomassen, *Onderzoeksgids Instrumenten van de macht. De Staten-Generaal en hun archieven 1576-1796*, Den Haag, 2015, vol. 1, 419-420.

60 A.M.J.A. Berkvens, 'Een aanzet tot hervorming van het strafprocesrecht in Staats Opper-Gelder in 1762', *Pro Memorie*, 2002, 347-361; J.H.M.M. Van Hall, 'Een onderzoek naar de rechtspraak in de Staatse landen van Overmaze door de advocaat-fiscaal J.F. van Steelant in 1772', in: A.M.J.A. Berkvens *et al.* (ed.), *Ten Werentliken Rechte. Opstellen over Limburgse rechtsgeschiedenis*, Maastricht, 1990, 59-86, in particular 70-76.

In the Austrian Netherlands and in therefore also in Austrian Guelders and the Austrian Lands of Outremeuse, under Joseph II substantive criminal law and the law governing criminal procedure briefly underwent a reform in 1787. Within the framework of a general review of the organisation of the executive and the judiciary, effective from 1 May 1787 the seigneurial courts in the Austrian Netherlands were abolished.[61] Torture was also officially abolished on this occasion. In the run-up to the Brabant Revolution of 1789, these reforms met with such protests, however, that they soon had to be reversed. This does not detract from the fact that, whereas in the Northern Netherlands torture was only abolished in 1798, it had been earlier restricted or fully abolished in parts of what is now the Dutch Province of Limburg, namely in Prussian Guelders (1740); in Austrian Guelders and the Austrian Lands of Outremeuse (1787); and subsequently, rather superfluously, again by the French (1794).[62]

10. Identical Problems, Similar Solutions? The Abolishing of the Scabinal Benches and a New Law Governing Civil Procedure

The short-lived abolishment of the seigneurial scabinal benches by Joseph II in 1787 was not an isolated occurrence. It was part of a general trend to professionalise the administration of justice in the third quarter of the eighteenth century, a trend that could also be observed in the *Stift* of Thorn and in Prussian Guelders. In the case of Prussian Guelders it concerned a series of reforms initiated by Frederick II of Prussia in a Cabinet Order dated 14 April 1780. The Cabinet Order addressed to the *Grosskanzler* Von Carmer contains a three-part action plan. Von Carmer was charged with improving the quality of the administration of justice by subjecting future judges to an entrance exam and with subsequently monitoring the quality of their work through regular inspection. In addition, Frederick II ordered Von Carmer to conduct a general review of the law of civil procedure in order to shorten proceedings and cut the costs of conducting these.[63] He also ordered Von

---

61 Edict of 3 April 1787, *Recueil des Ordonnances des Pays-bas autrichiens*, vol. 13, 24.

62 Decree of 17 November 1794 (http://137.120.13.20:8080/searcher?cmd =showdoc&id1=Staats-Overkwartier&id2=1794-11-17.txt).

63 L. Berkvens, 'Gerechtelijke hervormingen in Pruisisch Gelre 1746-1787', *Pro Memorie*, 2010, 36-55, in particular 47.

Carmer to design a general supplementary statute book, so that there would be less need for litigation owing to the clarity of the law.

Von Carmer set about the task with gusto, so that as early as in 1781 a new Prozeßordnung saw the light as the first book of the *Corpus Iuris Fridericianum*, the precursor of the *Allgemeine Landrecht der Preußische Staaten* (General Land Law of the Prussian States). Although this reformed procedural law was applied almost immediately throughout the Prussian Monarchy, its introduction in Prussian Guelders was met with objections from its Estates, which prevented its publication by invoking the provincial privileges that had been guaranteed under the Peace of Utrecht. Their resistance forced Von Carmer to tread carefully. In order to convince the *Justiz-Collegium* and the Estates of the need to reform the justice system, he organised inspections of the urban and rural scabinal benches. The findings of the inspection commission were nothing less than devastating. According to its report, the local rural judicial authorities were mainly uneducated peasants devoid of legal knowledge. Often more drunk than sober, they allowed themselves to be influenced by the local *scholtis* or the judicial clerk, who had at least enjoyed some form of legal education, or they lent their ear to the counsellors of one or both parties. As a result, their judgments often needed to be annulled on appeal.

This damning report caused him to charge Peter Heinrich Coninx, an ambitious justice on the *Justiz-Collegium* at Guelders, with drawing up tailor-made reform proposals for Prussian Guelders. In 1786, his activities resulted in three rule books, of which the rule book for the subordinate courts was the most relevant in this context. By virtue of this rule book, numerous seigniorial courts were merged and the administration of justice was entrusted to the *scholtis*, who from then on passed judgment as a single-sitting judge (*unus iudex*). From that time forward, the schepenen could only act as witnesses in conveyance cases.[64] Whereas the operation of the anticipated *Allgemeine Landrecht* had already been provided for, in advance, by virtue of the first ordonnance as general subsidiary law in addi-

[64] A.M.J.A. Berkvens, 'Peter Heinrich Coninx als codificator van het burgerlijk procesrecht in Pruisisch Gelre, 1783-1786', in: Gubbels and Jansen (eds.), Regio, 235-266; A.M.J.A. Berkvens, ''Juges de paix avant la letrre' in Midden en Noord-Limburg, voorafgaand aan de invoering van vredegerechten in de Franse tijd', in: G. Martyn (ed.), *Scènes uit de geschiedenis van het Vredegerecht/Scènes de l'Histoire de la justice de Paix*. Dossiers Tijdschrift van de Vrede- en politierechters/Les dossiers du Journal des Juges de Paix et de Police, vol. 16, Bruges, 2011, 57-70.

tion to the Country and Town Statute Book of Guelders – replacing the received Roman Law, which until then had served as a frame of reference in interpreting the rules of the Country and Town Statute Book of Guelders – in the second ordonnance the style of the *Justiz-Collegium* in conducting proceedings was brought into line with the *Prozeßordnung*.

On the basis of comparable considerations, Maria Kunegunde of Saxony, in her capacity of the abbess of Thorn, tried to almost simultaneously convince the Chapter (*Kapittel*) and the Estates Assembly (*Landdag*) of Thorn of the need for judicial reform as well. The plan to reform the judicial organisation and to codify the Land Law had already been laid down in her electoral capitulation (*Wahlkapitulation*) on the occasion of her investiture in 1776. The first reforms were presented in 1781 to the *Landdag* in the form of a number of 'Observations on the Justice System of Thorn' (*Betrachtungen über das Thornische Justitz-wesen*).[65] According to these *Betrachtungen*, bringing a case before the courts was a hazardous undertaking, because first-instance judgments were passed by the *schepenen*, who were usually 'uneducated citizens and peasants'. Their knowledge had proved sufficient in a time 'in which the laws and morals had been simple, in which the old reasonableness and fairness were the norm without exception, when the Roman laws which subsequently were passed on to Germany (*Teutsch-Land*) and the doubts and distortions that ensued from this were not yet known, and everything was based on ancestral customs ('*wo die Gesetze wie die Sitten einfach gewesen, wo die alte Redlichkeit und Treue ohne Ausnahme üblich waren, wo man die seitdem nach Teutsch-Land übertragene Römische gesetze und die hieraus entstande Zweiffel und Verdrehungen noch nicht kannte, und alles auf väterlichen Gebrauchen beruhete*').[66] Complaints about the competence of the schepenen and an aversion to Roman Law that caused needless complications for the legal process are also encountered in this case as a justification for reform. For this reason, the abbess proposed to do away with the justice passed by the *schepenen* and instead, for her account, appoint single-sitting judges. After protracted discussions, it was indeed so decided on 10 December 1785. On 22 December 1785, Jacob Norbert van der

65 A.M.J.A. Berkvens, 'Het Nieuwe Landrecht van Thorn', in: E.C. Coppens (ed.), *Honoris Causa. Opstellen aangeboden aan prof.mr. O. Moorman van Kappen*. Rechtshistorische reeks Gerard Noodt Instituut, vol. 42, Nijmegen, 1999, 1-28.

66 A.M.J.A. Berkvens, 'Het Nieuwe Landrecht van Thorn', 8.

Schoor was installed as *unus iudex*. His task was to 'impartially and without consideration whatsoever pass judgment on each person' ('*einem jeden ohne mindeste Rucksicht und unparteyisch Justiz wiederfahren zu lassen*').[67]

What is specially remarkable when studying the plans of Prussia and Thorn for the reform of the scabinal bench is their the similarity in approach: both Frederick II of Prussia and Maria Kunegunde of Saxony sought, on the basis of rational arguments, to convince their subjects of the need for reforms. Such an approach can also be found in the Austrian Netherlands under Maria Theresa, where the Secret Council (*Geheime Raad/Conseil Privé*) increasingly tried to gain support for the desired reforms by affording the provincial courts of justice an advisory role in legislative drafting, which especially in Austrian Guelders led to remarkable results.[68]

These examples show that at the end of the eighteenth century the Enlightenment had impacted on both Austrian and Prussian Guelders and the Stift of Thorn, while in the States Upper-Guelders and the States Lands of Outremeuse its influence was barely felt.

11. In Conclusion

Coming to the end of this expose, I return to the title of my article 'The Comparative Legal History of the Meuse-Rhine Euroregion: a *Ménage á Trois*'. I can now draw the following conclusion: In the early modern period, the emergence of modern centralised states in the territories between the rivers Meuse and Rhine led to a problematic determination of the spheres of influence between the Habsburg of Charles V and the remaining territories of the Holy Roman Empire. A number of smaller triangles fit into this 'large' triangular area, as can be learned from the struggle for appellate jurisdiction, in which the 'Dutch' and 'German' appellate colleges would always compete with each other where the judicial institutions within the region are concerned. A number of triangles can also be discerned in relation to the codification and systematisation of the

67 A.M.J.A. Berkvens, 'Het Nieuwe Landrecht van Thorn', 15.

68 Berkvens, '"Soumettant cependant le tout au discernement judicieux de Votre Excellence ..." De prealabele consultatie van het Oostenrijks Hof van Gelre als onderdeel van het wetgevingsproces in de achttiende eeuw', in: Berkvens et al. (eds.), Ten Werentliken Rechte, 151-170.

customary law. The codification processes in the Habsburg Netherlands and the German Empire manifested themselves simultaneously in a variety of ways. The codification of the unwritten law of the territories that form part of the Netherlands of today, such as Spanish Guelders, were greatly influenced by the *Rechtsordnug und Reformation* of Jülich-Berg, whereas, conversely, its Dutch counterparts, such as the Guelders Country and Town Statute Book, greatly influenced the interpretation of the Jülich-Berg *Gerichtsordnung und Reformation*.

The process of territorial disintegration of the area between the rivers Meuse and Rhine, which continued in the seventeenth and eighteenth century, led to the formation of new triangles: one between the States-General, the Prince-Bishop of Liège and Maastricht in 1632; one between Madrid/Brussels and the Hague with regard to the Lands of Outremeuse from 1644 onwards; and one between the States-General, the Austrian Netherlands and Prussia with regard to Spanish Guelders. We have seen that within these new triangles there were different approaches to the privileges enjoyed by their inhabitants and also how as a result of the territorial partitions numerous parallels could be drawn between the administrative and judicial organisations in these new territories. I have offered a number of examples to show the consequences of these developments on the very persons seeking justice. In addition to this, I have provided two examples, namely the systems of criminal justice and the deficiencies of the administration of justice by the rural scabinal benches, to illustrate the way in which the different sovereigns attempted to resolve similar problems in the region. A substantial difference in creativity is often noticed in this context and it is remarkable that in some parts of the region, such as Prussian Guelders and Thorn, the ideas introduced by the German Enlightenment had a clear influence, whereas in particular in the States territories necessary reforms failed to materialise. The institutional incompetence of the Dutch Republic may have been a major factor here.

When considering these various aspects, it must be noted that the comparative approach towards the legal history of the Euroregion between the rivers Meuse and Rhine is still in its infancy. A comparative analysis of the organisation and substance of the codifications, the procedural ordonnances of the different appellate colleges, as well as the legislation of the various territories invite us to undertake more comparative research.

About the Author

Louis Berkvens (1952) is Professor at the Maastricht University Faculty of Law, where he holds the ‘Legal History of the Limburg Territories’ Chair.

Laudatio Frank F.A. IJpma

Ingrid Kerckaert

Before it became clear in 2015 that not one but two anatomical preparations of Adolphe Burggraeve, one of Ghent's first graduated surgeons, had survived the test of time, Dr. Frank Florian Anno IJpma had already written a book and a thesis about a number of aspects of the Amsterdam Guild of Surgeons, including the group portraits, the "anatomy lessons", painted by Rembrandt and others.

The two preparations of Adolphe Burggraeve, namely "the born" and "head and hand of a woman" are the only remaining of the estimated 1,200 preparations made by Burggraeve in the 19th century.

But let's return to our laureate. Dr. Frank IJpma (° Amsterdam), currently a trauma surgeon and abdominal surgeon at the University Medical Center in Groningen, The Netherlands, was fascinated by the history of surgery during his medical education, resulting in the publication of a book in 2013 and a historical-anatomical dissertation in 2014.

In 1998 he started his medical studies at the University of Utrecht and completed his medical degree in 2005.

In 2006, Rembrandt's 400th birth year, he collaborates with 3 colleague-surgeons to compare the arm of "the anatomy lesson of Nicolaas Tulp", painted by Rembrandt, with an own dissection. Discussions about the anatomical accuracy had to be solved. In the Journal of Hand Surgery and the Nederlands Tijdschrift voor Geneeskunde their conclusions are published. Rembrandt was wrong regarding 4 concrete points.

In this way the passion of Dr. IJpma for the history of surgery and the research into the background of the anatomy lessons of Dutch masters started. By again dissecting cadavers himself, exactly according to the example on the paintings, he made new discoveries.

From 2007 he started under the guidance of Prof. Dr. Thomas Van Gulik, a surgeon in the AMC, his research on several paintings of anatomy lessons in the seventeenth and eighteenth century; they were all doctors and surgeons who portrayed during an anatomy lesson. This resulted in several publications.

Meanwhile, colleague IJpma followed his surgical training in the ISALA clinics in Zwolle, at the University Medical Center in Groningen and then back in Zwolle, to graduate in 2013 and then becoming a surgeon in Zwolle and later in Groningen. During his surgical training he went among other things to Ghana for hernia surgery, where he regularly returns.

In the context of his historical-anatomical research, he received in 2012 a personal grant from the Collegium Chirurgicum Neerlandicum to travel to St. Petersburg in Russia, to study the world famous over 300 years old anatomical collection of Frederik Ruysch. Frederik Ruysch was praelector anatomiae of the Amsterdam Guild of Surgeons for about 60 years. He made about 2000 anatomical prepations, of which 900 still have been preserved.

This collection is thus stored in St. Petersburg in the Kunstkamera or the Peter the Great Museum of Anthropology and Ethnography of the Russian Academy of Sciences. Two paintings from the 17th century of “The Anatomy Lessons of Frederik Ruysch” also belong to the paintings discussed by Frank IJpma in his book and thesis. One group portrait is painted by Adriaen Backer and one by Jan van Neck.

Dr. IJpma was also involved in the organisation of the 20th edition of the exhibition ‘The Anatomy lesson – From Rembrandt to Damien Hirst’ in the Gemeentemuseum in the Hague. So on the 14th of November 2013 the book “Amsterdamse anatomische lessen ontleed” was presented during the annual Anatomy Lesson, organized by the AMC and the Volkskrant in the Concertgebouw in Amsterdam. In this book he and Prof. Van Gulik described the backgrounds of 9 Amsterdam paintings of anatomy lessons from the 17th and 18th centuries for anyone interested in it.

As it were they bring back these paintings to life again. On each painting there is always the prosector, the praelector anatomiae, a group of surgeons around him, a body and a dissected part of the body. Who were all these people? Who was the person sentenced to death who was laying on the dissection table? Which anatomical structures were shown to anyone who moved on at the banks of the anatomical theater in the Waag in Amsterdam? To what extent was the anatomy shown correct? What is the symbolic meaning of what was painted? Is it intentional that the left arm was used as was the case on a picture of Vesalius? Why does Frederik Ruysch demonstrate lymph nodes? Does the shape of the falx cerebri symbol a scythe or a sickle? What is behind the idea of a son holding a skeleton? Our laureate knows it all!

The group portraits around the anatomy lesson are discussed in detail in the book, the prosectors and surgeons are described biographically, medical and anatomical details are discussed. Finally, again nine group portraits of so-called ‘overlieden’ (board members) of the 17th and 18th century guild are briefly described.

In 2014 Frank IJpma finally graduates with his thesis ‘The Anatomy Lessons of the Amsterdam Guild of Surgeons’ with promotor Prof. van Gulik. In his dissertation, he describes the history behind the paintings, anatomical and surgical aspects, historical insights into the Amsterdam Guild of Surgeons, and suggests that the training of surgeons in those centuries can in some ways serve as an example for today’s surgical training. In this thesis he also describes the collection of 17th and 18th century guild medals (in Dutch known as ‘gildepenningen’) of the Amsterdam Guild of Surgeons and the 84 coat of arms of the guild, which are featured in the vault of the Waag at the Nieuwmarkt in Amsterdam, the Theatrum Anatomicum of the past.

The contents of the book and the thesis are based on a number of special collections, anatomical collections, museum collections and original documents of the Amsterdam Guild of Surgeons.

Dr. IJpma continued his study of the collection of Frederik Ruysch in St. Petersburg in 2015. In 2016 he also travels to Bologna, Padua and Florence for studying anatomical theaters and collections.

His exceptional work, his impressive passion for the history of anatomy and surgery, his awareness of the importance of the contemporary anatomy education within medicine studies, form the arguments to award this driven surgeon a Sarton Medallion.

I quote from his doctoral thesis: “We believe that performing anatomy dissections still needs to have a place in the surgical training in order to develop the tissue feeling with the anatomical structures, to recognize anatomical variations and to get spatial insight into the anatomy”. End quote.

However, he interrupts his winter holiday untimely to be with us!

Dear audience, please, your attention to medical doctor surgeon Frank IJpma and his Sarton lecture “The anatomy lessons of the Amsterdam Guild of Surgeons”.

Anatomy lessons of the Amsterdam Guild of Surgeons

Frank IJpma and Thomas van Gulik

Introduction

In the 17th and 18th centuries, the Amsterdam Guild of Surgeons commissioned several famous Dutch painters for making their group portraits. The most well-known masterpiece is of course 'The anatomy lesson of Nicolaes Tulp', painted by Rembrandt in 1632. These carefully assembled group portraits, were painted to commemorate the surgeons and to promote their practices. Anyone who was on the painting had to pay a certain fee to the painter. The purpose of painting these large group portraits of surgeons was also to increase the status of the profession, and to decorate the available wall space in the boardroom of the guild at the Waag. Today most portraits are exhibited in the Amsterdam Museum. The nine anatomy lessons all include a number of recurring elements: the dissector (the praelector anatomiae) surrounded by a group of surgeons, the corpse and the dissected part of the body. Each element has its own story. Over de past few years we studied all original guild documents, guild regulations, documents from the municipal archives and musea to assess the painted 'anatomy lessons' of the Guild of Surgeons in Amsterdam, from an anatomical and surgical perspective. It was our goal to bring the anatomy lessons alive by elucidating the lives and works of the people depicted on the paintings.

Healthcare in Amsterdam

Healthcare in the 17th century in Amsterdam was provided by various groups. The physicians (doctores medicinae) and pharmacists were registered with the 'Collegium Medicum'. The doctores medicinae were academically trained physicians, who mainly focused on the internal diseases. These health professionals prescribed medications and if surgery was deemed necessary, they contacted a surgeon. Surgeons were traditionally trained under the direction of a master surgeon in a surgeon's shop. They had no university training and hence, lacked an academic status at that time. The original meaning of the word 'surgery' comes from the Greek word χειρεργον (cheirergon), meaning 'handicraft'. Surgeons were mostly concerned with external diseases of the body, which could be treated by operations, or any treatment performed 'by the hand'. Their work consisted, besides cutting hair and beards, of bloodletting, cupping (a partial vacuum was created in cups placed on the skin either by means of heat or suction in order to mobilize the blood flow and promote healing), applying bandages, treatment of wounds and ulcers, draining abscesses, assisting at difficult deliveries, reducing fractures and performing operations for strangulated inguinal hernias and bladder stones. 'Cutting stones', known as removing stones from the urinary tract, was at that time a very difficult and dangerous operation that was not entrusted to all surgeons. The same rules were applied to the operation of a strangulated inguinal hernia. Specialized surgeons were available for these high-risk operations, called 'steensnijders' (stonecutters) and 'breukmeesters' (herniotomists).

The Amsterdam Guild of Surgeons

As with several other handicrafts in the city of Amsterdam, the surgeons were united in a professional organization called a guild. The Surgeons' Guild in Amsterdam was well organized and controlled surgical care in the city. The guild was responsible for surgical training and education, and set the standards for the final exams. The board of the Surgeons' Guild consisted of six governors, who each held this position for three years. Every year in September, the city government appointed two new governors out of four nominees. The 'deken' (chairman) and 'proefmeester' (chair of examination board) were respectively president and vice president of the guild. The 'proefmeester' was mainly concerned with preparing and conducting the surgical exams, of which the requirements were precisely defined in the guild regulations. The praelector anatomiae (lecturer in anatomy) was appointed by the city government from among the doctores medicinae. He was entrusted with the theoretical education of the guild members and lectured twice a week. The Surgeons' Guild also organized anatomy lessons – initially once a year. For these special events, the praelector anatomiae used corpses of executed criminals to teach anatomy. The praelector was assisted in his anatomical dissections by a 'college-meester', who turned the dissection table to expose the part of the corpse that was dissected and who handed over the instruments for dissection.

The surgeon's shop

The surgeons held their practice in so-called 'surgeon's shops', which were previously scattered around the city of Amsterdam. The painting of Egbert van Heemskerck, dated 1669, provides a glimpse into the surgeon's shop owned by the surgeon Jacob Franszn and situated in the center part of Amsterdam (figure 1, p. 204). He was portrayed while performing a phlebotomy. His son Thomas, collected the jet of blood in a basin. On the right-hand-side, his wife is shown seated with their other son and daughter. In the back of the shop, an apprentice is seen shaving one of the other clients. Two customers in the left corner are waiting for their turn while a third customer has just come in, leaning on crutches in the doorway. The shop was decorated with some remarkable attributes, which referred to the profession of a surgeon. The walls were covered with cabinets containing surgical instruments and jars with medication. A small painting on the wall shows an anatomical dissection of the abdomen and thorax. The artist Van Heemskerck gives us an impression in which environment surgeons practiced in those days. Surgeons were traditionally trained in a master-apprentice relationship. The trainee was employed by a master surgeon in a surgeon's shop, usually for a period of five years. Besides his training in a the surgeon's shop, the trainee was expected to attend the lessons of the praelector anatomiae in the anatomy theatre, the Theatrum Anatomicum of the surgeons in Amsterdam. The surgical training was completed with a final proficiency test required by the Surgeons' Guild, the so-called 'meesterproef' (master-exam). If the trainee had passed the master-exam, he received the surgeons' certificate and was allowed to join the guild as a full member. Being certified, he was granted permission to open a surgeon's shop in Amsterdam for himself, in which he could train prospective surgical trainees.

The Theatrum Anatomicum

The Surgeons' Guild looked after the interests of the surgeons of Amsterdam, defined their profession and organized surgical education. The guild board met in the guild room, where meetings were held and surgical exams were taken. In addition, they had a dissection hall where the anatomy lessons were performed. The guild room and dissection room were the epicenter of the surgical profession in the 17th and 18th century in Amsterdam. The first anatomical demonstrations took place in the former St Ursula convent at the Oudezijdsachterburgwal, this building was known as a women penitentiary in later times. At the end of the 16th century, the surgeons moved to a room on the upper floor of the chapel of the former St Margaretha Convent at the Nes. Eventually, the government of the city of Amsterdam decided to build a state-of-the-art anatomy theatre for the surgeons. This Theatrum Anatomicum, the new dissection hall of the Surgeons' Guild, was housed in an octagonal dome in the middle of the Waag, which was specially constructed for this purpose. After the accommodation was completed in 1691, the Amsterdam Surgeons' Guild definitely settled in the Waag. The 'Waag' building is for over five centuries the dominant monument on the Nieuwmarkt in the old city center of Amsterdam (figure 2, p. 205). The access to the guild room and dissection room of the Surgeons' Guild was located in the southeastern tower of the weigh house, next to the main gate of the weigh house – now the entrance of bar-restaurant 'In de Waag'. The inscription 'Theatrum Anatomicum' still exists on the façade above the entrance door.

The boardroom of the guild was furnished with a cabinet containing surgical instruments. The guild records and medical book collection were kept in special chests or cabinets. The interior was decorated with all sorts of curiosities collected by the surgeons. The walls were decorated with the painted anatomy lessons, which should be condidered group portraits painted for the Surgeons' Guild in the 17th and 18th centuries. The guild's boardroom was located next to the Theatrum Anatomicum – at that time one of the largest in the country – that offered seating for a few hundred spectators. A yellow flag on the top of the building – labeled ANATOMIA – indicated the dome of the Theatrum Anatomicum, the pride of the Amsterdam Guild of Surgeons. Anatomical dissections were performed in the Theatrum Anatomicum in order to teach anatomy to surgeons and surgical trainees of the guild. The dissection room was designed according to the layout of the oldest Theatrum Anatomicum of Europe in Padua, in the northern part of Italy. It had the shape of an amphitheater with a rotating dissection table in the center (figure 3, p. 206). Through the four windows in the octagonal dome, light fell on the dissection table. The coats of arms of the governors of the guild were painted in the dome of the anatomy theatre (figure 4, p. 207). The dome with its painted vault was beautifully restored in 1992. Today, the dissection table and the ring shaped galleries no longer exist, but the coats of arms in the dome still recall the original function of the hall.

Anatomy lessons

Amsterdam was one of the first cities in the Netherlands in which the training of surgeons was clearly defined, and where anatomical dissections were performed from an educational point of view. Amsterdam was also one of the first cities in Holland where anatomical dissections were open to the public. In the mid-sixteenth century, the surgeons in Amsterdam were convinced that demonstrative education in anatomy of the human body was indispensable for proper surgical training. In addition, anatomical knowledge was essential for their practice.

In 1555, King Philip II of Spain granted the Amsterdam Guild of Surgeons permission to dissect the corpses of executed criminals for the purpose of teaching anatomy to the guild members. Initially, the guild was permitted to carry out one anatomy lesson a year. The lesson was performed by the praelector anatomiae in the dissection room of the guild. In later times, it occurred that several anatomy lessons were performed annually, albeit that in some years, no anatomy lesson could be organized because of the shortage of corpses. A dead body used for the anatomical dissection was called the subjectum anatomicum. The surgeons mostly used corpses of criminals sentenced to death, for their anatomy lessons. Other corpses eligible for dissection, were those of deceased ‘strangers’: people without relatives, of nobodies concern, who died in the Amsterdam guesthouses (city hospitals and shelters for the poor and homeless).

Anatomy lessons were performed in the dissection room of the guild, initially housed at the former St Ursula and St Margaretha Convent, and since 1691 permanently located in the Theatrum Anatomicum in the Waag. Anatomical dissections were usually performed in wintertime, in order that at low temperatures, the perishable corpses would be protected from rapid decay. The sequence of dissection was determined by the speed at which certain body parts decomposed, and therefore, each demonstration in anatomy was performed according to a fixed pattern. Firstly, the organs of the abdomen were dissected, then the chest and the brains, and finally the limbs. The anatomy lessons usually took several days, including the usual breaks.

Attending a public anatomical dissection was for many years an extraordinary event for the surgeons. Anatomy lessons provided a unique opportunity, in addition to learning from anatomical textbooks, to see the anatomy of the human body in reality. An image from 1615 gives an impression of an anatomy lesson at that time (figure 5, p. 208). This image shows the lecturer standing next to the dissection table, while performing an anatomy lesson in front of a modest group of about fifty spectators. This scene is usually interpreted as an anatomy lesson performed in Leiden. The conditions in Leiden, however, did not differ very much from the anatomy lessons that were taught in Amsterdam. During the 17th century, anatomical demonstrations evolved into crowded events, which were frequently attended by the citizens.In the guild regulations of 1625 is explicitly stated: ‘from now on anyone who appreciates the anatomy […] will pay four ‘stuivers’ [for attending an anatomical lesson], and the guild members will pay six ‘stuivers’ for each dissected corpse’. Anatomy lessons were not only accessible to guild members, but – against payment of four ‘stuivers’ – to anyone who was interested in attending such lessons. The anatomy book of the Surgeons’ Guild revealed that an anatomy lesson attracted hundreds of visitors in the 17th century. The large number of visitors may have partly detracted from the educational character of the lesson for the surgeons. The surgical trainees who really needed to learn something, after all, were seated in the back rows of the Theatrum Anatomicum. An illustration of an anatomy lesson in the Theatrum Anatomicum in Leiden gives a good impression of the great public interest and the crowds during such an event (figure 6, p. 209). The situation at the Waag in Amsterdam would probably have been equivalent. The profits of the lessons were initially used to pay the expenses of the lesson and the funeral of the corpse. The money that was left was mostly spend on a large guild banquet, which was held immediately after the anatomy lesson ended.

The surgeons of Amsterdam portrayed

In the period 1601-1758, eighteen group portraits were painted for the guild. In nine of these, an anatomy lesson was the main subject of the painting, and a central role was reserved for the praelector and the subjectum anatomicum, (part of) the dissected corpse. For these reasons, these group portraits – dated 1601-1603, 1619, 1625-1626, 1632, 1656, 1670, 1683, 1728 and 1758 – are referred to as 'anatomy lessons' or 'anatomy pieces' (figure 7, p. 209). The paintings were displayed on the wall of the guild room, and should now be considered part of the unique art collection of the guild. The composition of the first group portrait of the surgeons in Amsterdam was probably derived from the group portraits that had been made for the militia (schutterij) in Amsterdam at that time. One of the most famous militia pieces is The Night Watch, painted by Rembrandt in 1642. In fact, the anatomy lessons are closely related to the militia pieces.

Important events within the guild of surgeons – such as the first anatomy lesson of a newly appointed praelector, or the inauguration of a new guild room – could be a reason to commission a new group portrait. It was no coincidence that the painter and the surgeons chose to use the theme 'anatomy lesson' in the group portraits. The anatomical demonstrations, organized by the guild, were of course major events for the surgeons. At the same time, the anatomical dissections were a source of inspiration to the painters to create an original composition. Although all anatomical demonstrations organized by the Surgeons' Guild have been well documented, some paintings could not be related to a documented record of a particular lesson. One must bear in mind in this regard, that many of the paintings were in fact carefully assembled group portraits, and not a true representation of an anatomical demonstration held at the Theatrum Anatomicum. Some of these paintings will be discussed in the upcoming paragraphs.

The anatomy lesson of Sebastiaen Egbertszn

The anatomy of Sebastiaen Egbertszn, dated from 1601-1603, is the oldest group portrait from a series of nine painted anatomy lessons, which has been made for the Amsterdam Guild of Surgeons in the 17th and 18th century (figure 7a, p. 209). This painting demonstates the praelector anatomiae together with a group of surgeons, surrounding a corpse that is lying on the table in front of them. Sebastiaen Egbertszn (1563-1621) was appointed praelector anatomiae of the Surgeons' Guild in 1595 and held this position until his death in 1621. He is the key figure in this painting, standing behind the dead body. Unfortunately, the identy of the corpse is unknown. The names of most surgeons, depicted in the painting, are known from the list of names, which is held by one of them in the lower right corner of the painting. The heading of the list indicates the 'all master surgeons, who were part of the guild in 1603' were depicted on this masterpiece. In 1600, these 28 surgeons served a population of 100.000 inhabitants in Amsterdam. The anatomical lesson of Sebastiaen Egbertszn is the only painting on which all the surgeons of the guild have been depicted. The number of surgeons increased rapidly during the 17th century, and therefore in subsequent paintings usually only the praelector and governours of the guild have been depicted. While holding a dissection instrument in his right hand, it seems that Egbertszn is about to start with his anatomical demonstration. Interestingly, the attention of the surgeons is not focused on the body nor the anatomical demonstration, which was about to begin. The portraits are quite staticcally arranged as if the surgeons were posing for a group picture. The painting was not intended as an exact representation of a documented anatomical lesson but rather as a group portrait, painted to commemorate the praelector and surgeons of the guild. After all, this first anatomy lesson (1601-1603), painted for the Amsterdam Guild of Surgeons, marks the beginning of a unique portrait tradition of the guild for the next 150 years afterwards.

The osteology lesson of Sebastiaen Egbertszn

The Osteology Lesson of Sebastiaen Egbertszn (1619) is the second masterpiece of the collection of anatomy paintings, which were commissioned by the Amsterdam Guild of Surgeons in the 17th and 18th centuries (figure 7b, p. 210). The painting records a carefully composed group portrait of the guild officials and was intended to commemorate the governors and lecturer of the Surgeons' Guild. At that time, the painting was on display above the mantelpiece in the guild room of the surgeons in the Weigh house (Waag) at the Nieuwmarkt in Amsterdam. The composition of the painting is reminiscent of an early lesson in osteology. 'Osteology', derived from the Greek words osteon (bone) and logos (knowledge), is defined as the study of the structure and function of bones. In fact, osteology could be considered the forerunner of current orthopaedic and traumatological education. In the painting we see Sebastiaen Egbertszn giving an osteology lesson by using a skeleton. The skeleton serves as a central element in the composition, and is thought to have belonged to an English pirate who was captured at sea and brought to Amsterdam in 1615. He was executed and his corpse was probably handed over to the surgeons to be used for an anatomical demonstration. The surgeons used the remaining bones of the dissected pirate to rebuild his entire skeleton. The skeleton was kept by the surgeons and used for teaching osteology to the trainees.

Egbertszn was appointed praelector anatomiae (lecturer in anatomy) of the guild and was responsible for the anatomical education of the surgeons. For over 20 years he performed his anatomical dissections at the former chapel of the St Margaretha Convent at the Nes in Amsterdam. In 1619, the surgeons moved their guild room to the Weigh house (Waag) at the Nieuwmarkt in Amsterdam, in which a new and fully equipped anatomy theatre would be constructed, modelled on the famous Italian anatomy theatre in Padua. On the occasion of the opening of the new residence of the Surgeons' Guild, Egbertszn presented a lecture about the ancient medical works of the Roman physician Aulus Cornelius Celsus (ca. 25BC to 50AD), who for centuries dominated medical doctrine. The inauguration of the new guild room in 1619 was an important event for the surgeons of Amsterdam, and it was on this special occasion that the surgeons commissioned the group portrait of the guild officials, known today as The Osteology Lesson of Sebastiaen Egbertszn.

The anatomy lesson of Nicolaes Tulp

Approximately four centuries ago, the Amsterdam Guild of Surgeons commissioned Rembrandt to paint a group portrait of the surgeons in the form of an anatomy lesson. The painting, dated 1632, is known as 'The anatomy lesson of Nicolaes Tulp', and is the fourth anatomy lesson that has been created for the Guild in the 17th and 18th centuries (figure 7d, p. 211). The painting is considered a masterpiece because Rembrandt created an action scene of an anatomy lesson, and abandoned the traditional formal and stiff arrangement of figures. However, one should bear in mind that this painting is a carefully assembled group portrait and not a true representation of an anatomical demonstration held in the anatomy theatre (theatrum anatomicum) of the surgeons.

Looking at the painting, Nicolaes Tulp (1593-1674) is surrounded by seven surgeons. Tulp practised medicine for almost 40 years as a general practitioner in Amsterdam. He held consultations in his office at home and visited the seriously ill patients daily using his coach. As an academically trained physician he was appointed by the city council of Amsterdam as praelector anatomiae of the Amsterdam Guild of Surgeons. In that position he was in charge of the anatomical education of the surgeons. Tulp is portrayed while performing an anatomical dissection of the forearm. The gesture of Tulp's left hand illustrates the function of the muscles in the forearm and lends weight to his demonstration. The fingers of Tulp's left hand are flexed as if he wants to show that pulling on the dissected flexor muscles with his forceps produces flexion of the fingers. Tulp presumably used this initiative to have himself painted with a group of surgeons in the tradition of his predecessors on the occasion of an anatomy lesson. Tulp performed his anatomy lesson on the corpse of a criminal who was sentenced to death in January 1632.

The identity of all surgeons depicted in Rembrandt's group portrait is well known. Their names were recorded on the piece of paper held by one of the surgeons, as shown in the painting. Rembrandt initially had painted an anatomical illustration of an arm on the piece of paper. The original sketch of the arm was repainted during restoration in the 18th century and replaced with the names of the surgeon observers. However, during the most recent restoration process of the painting from 1996 to 1998, the uppermost layer of paint with the list of names has for the greater part been removed.

The dead body was that of a 28-year-old man, an executed criminal named Adriaan Adriaenszn alias Aris't Kint. Adriaan had a long criminal record. The reports of his convictions are preserved in the judicial archives of Amsterdam. In the winter of 1631, Adriaan was walking together with two friends at the Herensluis in Amsterdam. They walked past a well dressed man who was wearing a nice cape. The next moment they attacked the man and tried to steal his cape. The man, of course, resisted and Adriaan almost beated the man to death during the fight. He was arrested, imprisoned and interrogated for this offence. From the reports in the judicial books, it appears that the interrogations of Adriaan were extremely violent. To force him into a confession he was lifted with 200-kg weights hanging on his legs. In addition, he was tortured and his skin was marked with a branding iron. On 27 January 1632, Adriaan was sentenced to death by hanging for all the crimes he had committed over the years. The death sentence was carried out on 31 January. After the execution, his body was handed over to the Surgeons' Guild and served as the subjectum anatomicum at the anatomy lesson of Nicolaes Tulp. It is remarkable that the corpse in Rembrandt's painting does not show any injuries from the torture.Welts from the rope around the neck, burn marks and welts on the legs are missing in the painting. Perhaps Rembrandt did not want to depict a battered corpse in this painting. Above all, the emphasis in the painting had to be on the surgeons rather than the corpse.

When we are looking at the painting, our attention is immediately caught by the dissected forearm. The accuracy of the anatomy of the dissected forearm in Rembrandt's painting has been a matter of discussion in medical and art history literature for decades. Recently, we investigated the accuracy of the dissected forearm in the painting by reproducing an anatomical dissection of the forearm of a male corpse. The arm was used for teaching purposes for medical students. The forearm in the painting is extended and supinated with the wrist placed in the groin, thereby exposing the medial epicondyle. We therefore concluded that Rembrandt was correct in his representation of Tulp demonstrating the flexor muscles of the forearm that originate from the medial epicondyle of the humerus.When further comparing the anatomical dissection with that in the painting, four additional subtle anatomical differences could be observed.

Rembrandt was an artist whose compositions originated gradually on the canvas, and several adjustments were made during this process. X-ray imaging of the dissected left forearm showed that Rembrandt had moved the arm from a higher to a lower position during the painting process. The adjustments he made to the dissected forearm indicate that the painting gradually appeared on the canvas during the working process of the painter. Adjustments to the painting may have contributed to the anatomical differences between the painting and the reproduction of the dissected forearm. There is no official record of Rembrandt being an eyewitness at the anatomy lesson of Nicolaes Tulp. However, the details and realistic colours of the painting suggest that Rembrandt must have had access to a real dissected limb and that the members of the Guild posed for him to complete the painting.

A public anatomy lesson in the 17th century usually started with dissecting the perishable organs of the abdomen and thorax, whereas the extremities were the last to be dissected. In Rembrandt's painting, however, the forearm has already been dissected while the rest of the body is still intact. This is one of the reasons for believing that Rembrandt's painting does not depict a real scene of Tulp's anatomical dissection, but rather represents a carefully assembled group portrait, which has a deeper meaning. Public anatomy lessons were often preceded by a moralistic speech emphasizing the vanity of earthly life (memento mori). The science of anatomy was considered a path toward the knowledge of God. Even moralizing inscriptions inside the anatomy theatre reminded visitors of the ephemerality of life. Andreas Vesalius (1514-1564), an anatomist originally from Brussels, had composed a set of books on human anatomy, entitled 'De Humani Corporis Fabrica Libri Septem', published in 1543. Vesalius' seminal work led to major advances in the understanding of human anatomy at that time. The woodcut frontispiece of this book shows a portrait of Vesalius demonstrating the flexor muscles of the forearm (figure 8, p. 214). It is believed that this portrait inspired Tulp to take the same pose as Vesalius while dissecting a forearm. Vesalius had accomplished a revolution in the knowledge of human anatomy by performing anatomical dissections himself while describing to his students exactly what he was observing. He challenged the works of the ancient authorities on anatomy. Vesalius considered the human hand the physical counterpart of the human mind, an instrument for using further tools and a representation of God's wisdom. Tulp was familiar with the theories of Vesalius. One century after Vesalius,Tulp chose to be depicted with a dissected forearm to be considered the Vesalius of his time, and confirmed the importance of connecting morphology and function in anatomy.

Discussion

Anatomical lessons were part of the surgical training during the 17th and 18th century. They fit wel into way of thinking during the Renaissance and the Golden age. Anatomists tried to find out the 'anatomical' truth by relying on their own anatomical observations instead of the ancient medical textbooks. Furthermore, the theme of an anatomy lesson was incorporated in the world famous group portraits of the Amsterdam Guild of Surgeons. As we have seen, there is no doubt that it is possible to bring these paintings back to life by telling the fascinating stories of all the people, like the praelector, surgeons and dissected corpse, who have been depicted on these masterpieces. Today, we discussed only three out of nine paintings. However, more about the colourful characters and fascinating events surrounding these paintings could be read in our book 'Amsterdamse anatomische lessen ontleed'.[1] In fact, these anatomy lessons make a unique connection between the painted art of the Dutch artists and the cutting art of the surgeons. Like Nicolaes Tulp once said: 'anatomy is the very eye of medicine' and that 'it brought forth the truth as it came out of the shadow into the light'. We totally agree with him, because good knowledge of the anatomy is the foundation for good surgery. Therefore, we believe that anatomical education still deserves an important place in our current surgical curriculum.

Literatuur

[1] IJpma FF, van Gulik TM. Amsterdamse anatomische lessen ontleed. Boom: Amsterdam, 2013.

Figures

Figure 1. Jacob Franszn (ca. 1635-1708) and his family in the surgeon's shop, painted by Egbert van Heemskerck, 1669. Collection Amsterdam Museum.

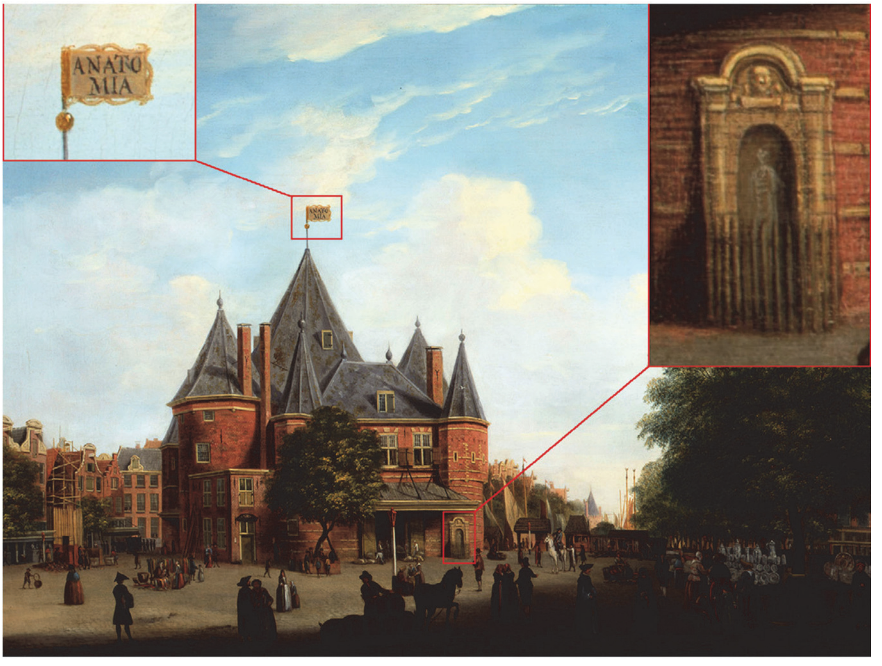


Figure 2. View on the Waag at the Nieuwmarkt in Amsterdam, painted by Jan Ekels de Oude in 1763. The porch in the right corner tower gave access to the guild room and Theatrum Anatomicum of the Surgeons' Guild. The high, central dome, in which the anatomy theatre was located, is indicated by a flag displaying the word 'ANATOMIA'.

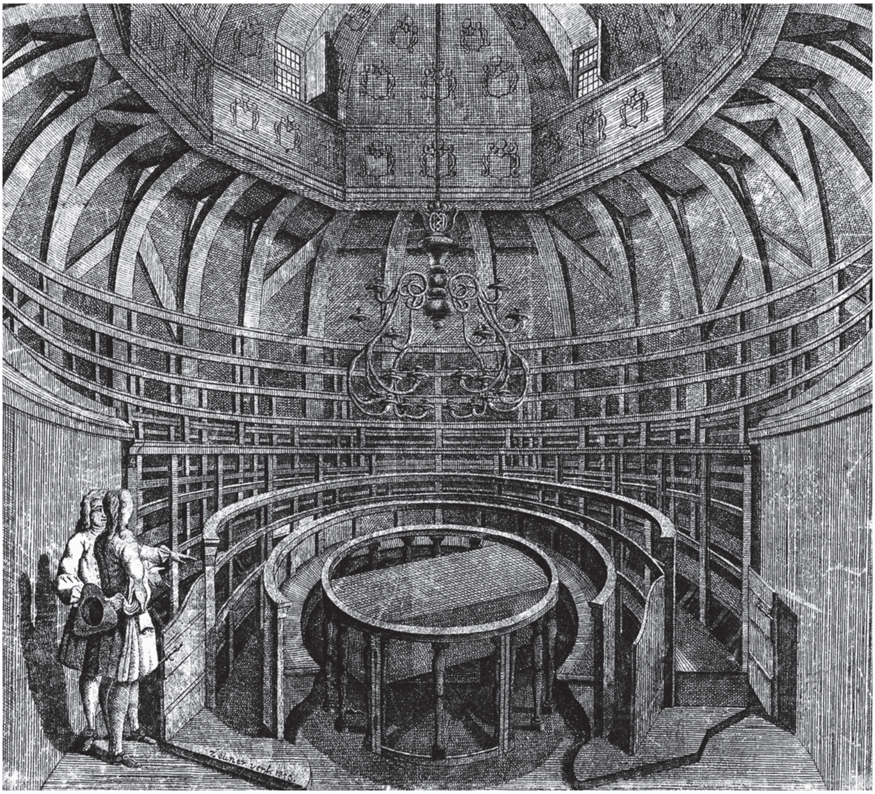


Figure 3. Theatrum Anatomicum in the Waag, made by Jonas Zeuner, ca. 1775-1799. The dissection table was situated in the center of the anatomy theatre and surrounded by semicircular galleries. The room was illuminated by a chandelier and the light from the four windows in the dome of the theatre. Collection Amsterdam Museum.

Figure 4. Coats of arms of the surgeons in Amsterdam, painted in the dome of the former Theatrum Anatomicum in the Waag. Photo: Matthanja Bieze.

Figure 5. Anatomy Lesson of Professor Pieter Paauw in the Theatrum Anatomicum in Leiden, made by Andries Stock after the design of Jacob de Gheyn II, 1615. Collection Rijksmuseum, Amsterdam.

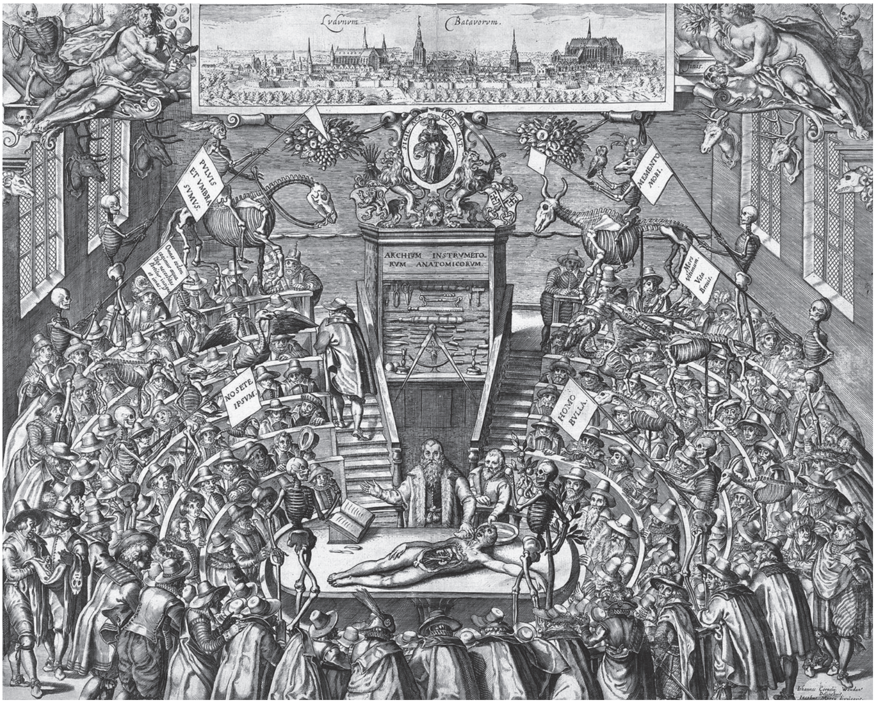


Figure 6. The Theatrum Anatomicum in Leiden, made by Bartholomeus Dolendo after a drawing by Jan Corneliszn van 't Woud, 1619. Collection Rijksmuseum, Amsterdam.

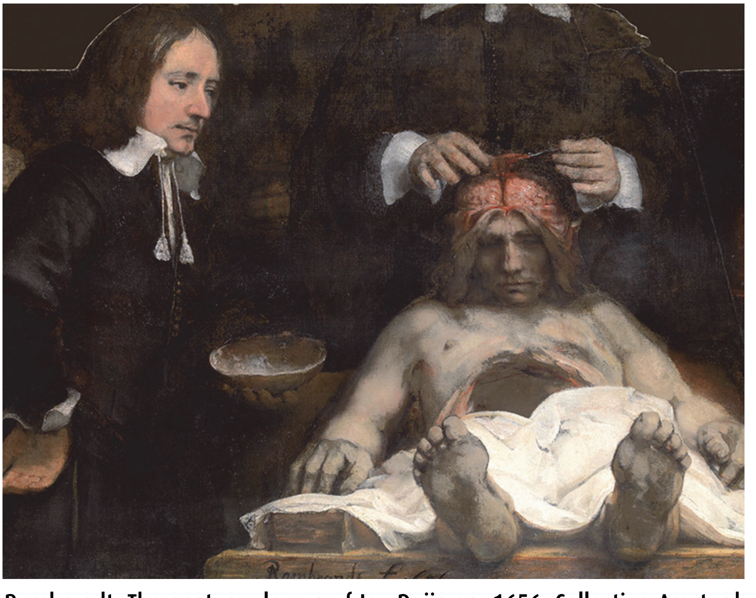
Figure 7. Anatomy lessons of the Amsterdam Guild of Surgeons.

a. Aert Pieterszn, The anatomy lesson of Sebastiaen Egbertszn, 1601-1603. Collection Amsterdam Museum.

b. Thomas de Keyser or Nicolaes Eliaszn Pickenoy, The Osteology Lesson of Sebastiaen Egbertszn, 1619. Sebastiaen Egbertszn (1563-1621) is standing next to the skeleton and demonstrates the lower rib. Collection Amsterdam Museum.

c. Nicolaes Eliaszn Pickenoy, The anatomy lesson of Johan Fonteijn, 1625-1626. Collection Amsterdam Museum.

d. Rembrandt, The anatomy lesson of Nicolaes Tulp, 1632. Royal Cabinet of Paintings Mauritshuis, The Hague.



e. Rembrandt, The anatomy lesson of Jan Deijman, 1656. Collection Amsterdam Museum.

f. Adriaen Backer, The anatomy lesson of Frederik Ruysch, 1670. Collection Amsterdam Museum.

g. Jan van Neck, The anatomy lesson of Frederik Ruysch, 1683. Collection Amsterdam Museum.

h. Cornelis Troost, The anatomy lesson of Willem Röell, 1728. Collection Amsterdam Museum.

i. Tibout Regters, The anatomy lesson of Petrus Camper, 1758. Collection Amsterdam Museum.

Figure 8. Portrait of Andreas Vesalius showing the flexor muscles of the forearm, an illustration from his famous work 'De Humani Corporis Fabrica Libri Septem' (1543).

Laudatio Sarton Medal Professor Dr. Dany Nobus

Stijn Vanheule

The Faculty of Psychology and Educational Sciences selected professor Dany Nobus for the 2016-2017 Sarton Medal. With the medal we express our special appreciation for the work of Dany Nobus pertaining to the history of psychology and psychoanalysis.

Dany Nobus started his career in Ghent. In 1988 he first obtained a Master degree in Psychology and Educational Sciences at Ghent University. Subsequently, in 1992, he completed a post-graduate training program in Freudian and Lacanian psychoanalysis, and in 1996 he obtained a doctoral degree, with a dissertation on love, perversion and psychoanalysis.

Since 1996 Dany Nobus has a full-time appointment at Brunel University in Londen. At the start he was lecturer, then senior lecturer, and since 2006 he is Professor of Psychoanalytic Psychology in the School of Social Sciences. Next to that he was a visiting professor in several international universities, including University of Massachusetts in Boston and Creighton University Medical School. Currently he teaches courses on psychoanalysis, psychopathology and society.

Since 2012 professor Nobus is Pro-Vice-Chancellor at Brunel University, first focusing on strategy, development and external relations, and from 2014 onwards on external relations in particular. Previously he was also head of the School of Social Sciences at Brunel University.

Since 1999 he was the supervisor of 27 succesfully completed doctoral research projects in the field of psychology and psychoanalysis, of which several had a historical focus.

Professor Nobus is the author of two academic books, and the editor of four books with collected papers. Two recently authored monographs by him are about to be published, among which *The Law of Desire: On Lacan's 'Kant with Sade'*. Next to that he is the author of 39 book chapters and 72 journal articles, of which several were published in *Psychoanalytische Perspectieven*. This is a journal directed by the Department of Psychoanalysis and Clinical Consulting at Ghent University. Finally he also translated several psychoanalytic texts, among which Jacques Lacan's 'Lituraterre' and 'The Tokyo Discourse'.

Next to his academic work Dany Nobus is the chair of the Freud Museum London (since 2014). This museum is located in Hampstead, in the house where Sigmund Freud spent the last year of his life, as he had to leave Vienna because of the persecution of the Jews. This museum has not only a historial collection focusing on Freud and psychoanalysis, but is also a center of innovation and discussion about psychoanalysis, psychology, the arts and society.

Several of professor Nobus' publications deal with historical issues in the fields of psychology, psychoanalysis, sexology and criminology. We proposed Dany Nobus as a laureate for the Sarton Medal because of the excellent quality of these publications. We really appreciate his thorough way of detecting and analysing relevant sources, and his clear and most precise mode of articulating ideas and insights. Finally it should be noted that in his publications without explicit historical focus too, Dany Nobus usually most eloquently situates research problems in their relevant historical context. Along this way he over and again makes clear that the study of history is often most illuminating for contemporary debates in our scientific disciplines.

Freud Unbound

The Shaping of Psychoanalysis by the Great War

Dany Nobus

Introduction

On 28 July 1914, exactly one month after the assassination of Archduke Franz Ferdinand in Sarajevo, the Austro-Hungarian Empire declared war on Serbia, thus setting in motion what the American diplomat George F. Kennan would describe as "*the* great seminal catastrophe" of the 20th century, and "the event which ... lay at the heart of the failure and decline of this Western civilization".[1] When news of the armed conflict broke, the atmosphere in Vienna and other central European cities was nothing short of electrifying. Socialists and nationalists shared in a new sense of patriotic heroism, and erstwhile pacifists such as the German novelist Arnold Zweig unapologetically volunteered to join the armed forces. Within the first weeks of the war the Central Powers were capable of deploying vast numbers of troops, partly on account of conscripted men swiftly answering mobilisation orders, partly owing to the support from large swathes of enthusiastic volunteers. Arriving back in Vienna on the last Orient Express from Belgium, the Austrian writer Stefan Zweig captured the mood in words whose ostensible grandiloquence was in fact perfectly apt for the exuberant quality of the momentous events: "[T]here was a majestic, rapturous, and even seductive something in this first outbreak of the people from which one could escape only with difficulty ... [T]housands and

1 G. F. Kennan, *The Decline of Bismarck's European Order: Franco-Prussian Relations 1875-1890* (Princeton NJ-London: Princeton University Press, 1981), 3-4.

hundreds of thousands felt what they should have felt in peace time, that they belonged together ... [E]ach one was called upon to cast his infinitesimal self into the glowing mass, there to be purified of all selfishness ... Each individual experienced an exaltation of his ego, he was no longer the isolated person of former times, he had been incorporated into the mass, he was part of the people, and his person, his hitherto unnoticed person, had been given meaning."[2] In his sweeping account of life in the Austrian capital just prior to the outbreak of the Great War, the Vienna-born American historian Frederic Morton conceded that, whereas the announcement of military action threw many European cities in a state of frenzied elation, Vienna "out-waltzed friend and foe alike in celebration."[3]

From his large windows on the first floor of the Berggasse 19, the fifty-eight-year old Sigmund Freud would have seen the colourful parades marching through the street, women waving their handkerchiefs at the regiments leaving for the front, citizens of all ages dancing to the beat of Johann Strauss' Radetzky march.[4] Much like so many other Austrians, he had reacted to the murder of the heir-apparent with subtle indifference. On the very day of the drama that would trigger the first cataclysm of the 20th century, he had written to his fellow psychoanalyst Sándor Ferenczi in Budapest: "I am writing under the impression of the surprising murder in Sarajevo, the consequences of which cannot be foreseen. It appears to me that personal involvement here is slight."[5] But less than a month later his spirit had awoken to the promise of great things to come. Two days before the war became official, he had already written to the psychoanalyst Karl Abraham in Berlin: "[P]erhaps for the first time in 30 years I feel myself to be an Austrian and would like to try it once again with this not very hopeful

---

2 S. Zweig, *The World of Yesterday: Memoirs of a European* (1943), Trans. A. Bell (London: Pushkin Press, 2014), 222.

3 F. Morton, *Thunder at Twilight: Vienna 1913/1914* (New York NY: Charles Scribner's Sons, 1989), 336.

4 As it happens, on 28 July 1914, Freud and his wife were in the spa resort of Carlsbad (Karlovy Vary), where Freud was seeking treatment for an intestinal problem. The plan was for them to leave from there in early August to spend the rest of the Summer in the Dolomite village of Seis, yet because of the war they returned to Vienna on the evening of 4 August. See E. Jones, *The Life and Work of Sigmund Freud, Vol. II: Years of Maturity (1901-1919)* (New York NY: Basic Books, 1955), 172; M. Schröter (ed.), *Unterdeß halten wir zusammen. Sigmund Freud – Briefe an die Kinder* (Berlin: Aufbau Verlag, 2011), 493.

5 E. Brabant, E. Falzeder & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 1, 1908-1914, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1993), 562.

Empire. Morale everywhere is excellent."[6] However, when at the end of August the whole of Europe was set alight, Freud's initial excitement turned into a more sombre, gloomy state of mind. As he disclosed to Ferenczi on 23 August: "The rush of enthusiasm in Austria swept me along with it, at first ... Like many others, I suddenly mobilized libido for Austria-Hungary ... Gradually a feeling of discomfort set in ..."[7] For reasons that are not entirely clear, some of Freud's biographers have exaggerated his bellicose fervour during the first month of the war, and then expressed their surprise that a fiercely rational, middle-aged man who was generally disinterested in politics could allow himself to be carried away by a public outburst of patriotic sentiment.[8] The truth is that during the early days of the conflict very few people openly opposed the war, and that some of the most prominent intellectuals in Germany and Austria-Hungary, such as Arthur Schnitzler, Rainer Maria Rilke, Hugo von Hofmannsthal and Thomas Mann, explicitly welcomed the hostilities, including all the death and destruction they would entail, as some kind of latter-day purification ritual that would signal the end of a lacklustre, bourgeois life-style. In addition, most people believed this would be a "war of movement", fierce yet fast-paced, relatively smooth and swiftly re-creating stable power relations on the European continent.[9] As Freud himself put it some six months after the conflict had broken out: "[W]e pictured it as a chivalrous passage of arms [*einen ritterlichen Waffengang*], which would limit itself to estab-

---

6 E. Falzeder (ed.), *The Complete Correspondence of Sigmund Freud and Karl Abraham 1907-1925*, Trans. C. Schwarzacher (London-New York NY: Karnac Books, 2002), 265.

7 E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 13.

8 Freud's first serious biographer Ernest Jones reproduced the sentence from Freud's letter to Ferenczi as "All my libido is given to Austro-Hungary", without the subsequent qualification that a sense of discomfort had already taken over. When Jones' volume was published, in 1955, Freud's correspondence with Ferenczi was not in the public domain, yet even after it was made available, biographers have preferred Jones' hyperbole over Freud's own, more measured admission of libidinal investment. In Joel Whitebook's most recent biography of Freud, the words are rendered as "my entire libido is given to Austro-Hungary", whereas Élisabeth Roudinesco has paraphrased Freud's comment with a subtle touch of ambiguity: "His libido, he said without really believing it, was solidly mobilized in favor of Austro-Hungary." See E. Jones, *The Life and Work of Sigmund Freud, Vol. II: Years of Maturity (1901-1919)* (New York NY: Basic Books, 1955), 171; J. Whitebook, *Freud: An Intellectual Biography* (Cambridge: Cambridge University Press, 2017), 317; É. Roudinesco, *Freud: In His Time and Ours* (2014), Trans. C. Porter (Cambridge MA-London: Harvard University Press, 2016), 174. See also L. Breger, *Freud: Darkness in the Midst of Vision* (New York NY: John Wiley & Sons, Inc., 2000), 234.

9 G. L. Mosse, *Fallen Soldiers: Reshaping the Memory of the World Wars* (New York NY-Oxford: Oxford University Press, 1990), 4.

lishing the superiority of one side in the struggle, while as far as possible avoiding acute suffering that could contribute nothing to the decision, and granting complete immunity for the wounded who had to withdraw from the contest, as well as for the doctors and nurses who devoted themselves to their recovery."[10]

Much more puzzling and provocative than Freud's initial response to the impending desolation of old Europe is his observation, six days after the signing of the armistice between the Central Powers and the Allied Forces, that for psychoanalysis the war had somehow ended too soon. Again, it was Ferenczi to whom Freud addressed himself: "Our analysis has actually also had trouble. No sooner does it begin to interest the world on account of the war neuroses than the war ends, and once we find a source that affords us monetary resources, it has to dry up immediately. But hard luck is one of the constants of life. Our kingdom is indeed not of this world."[11] Neither here, nor in any of the other letters Freud exchanged around this time does one get a sense of the extraordinary relief that settled in the hearts of people across the world on that November day in 1918. Nowhere did Freud confess to having privately or publicly participated in the triumphal hubris that traversed the broken citizens and the shattered nations of Europe. On 11 November 1918, he dryly noted in his Prochaska year-calendar: "End of war, Emperor Karl renounces the throne."[12] Maybe Freud's joy was dampened by the recent news that his eldest son Martin had been captured on the Italian front, without anyone knowing when or even whether he would come home again.[13] Maybe the dramatic collapse of the Austro-Hungarian empire and the prospect of an uncertain socio-political future for central Europe overshadowed the vaguely victorious jubilation associated with the end of warfare. In any case, what could have been an exhilarating moment in Freud's life seems to have been eclipsed by inexorable feelings of disap-

10 S. Freud, 'Thoughts for the Times on War and Death' (1915*b*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 14, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1957), 278.

11 E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 311.

12 M. Giefer & C. Tögel (eds), *Sigmund Freud. Die Kalendereinträge von 1916-1918* (Frankfurt am Main-Basel: Stroemfeld, 2016), 111. In what follows, all unattributed translations from foreign-language sources are mine.

13 See M. Freud, *Parole d'honneur*, Trans. Ph. & T. Blewitt (London: Victor Gollancz, 1939), 157-60; M. Schröter (ed.), *Sigmund Freud – Unterdeß halten wir zusammen. Briefe an die Kinder* (Berlin: Aufbau Verlag, 2011), 175; H. Fry, *Freuds' War* (Stroud: The History Press, 2009), 52.

pointment and doom, in a peculiar reversal of the prevailing mood at the outbreak of the war, when anxious tension over the fate of the nation and its people was dissipated in the euphoria of a new-found community spirit.

It is between these two complex poles of Freud's internal life – between the libidinal attachment giving way to discomfort at the start of the war, and the restless unease gradually turning into quiet optimism at the end of it – that I want to situate the substance of this essay. Just like it affected everything else, from the basic needs of daily life to large-scale economic and political configurations, from science and technology to literature and the visual arts, the Great War left its imprint on Sigmund Freud and the discipline named psychoanalysis he had invented at the turn of the 19th century. Without entering into the contested domain of counterfactual history, and thus avoiding speculation about how Freud's life and that of his invention would have evolved had Archduke Franz Ferdinand not been shot and had the ensuing diplomacy been more effective, I shall argue that psychoanalysis – the clinical practice, the conceptual apparatus, the general theory and the institutional culture around it – was indelibly marked and permanently shaped by the trials and tribulations of the Great War, so much so that we do not require the satirical stance of Sellar and Yeatman's history of England to observe that for psychoanalysis too World War I could justifiably be called "the cause of nowadays and the end of history."[14]

Freud at War

Reflecting, many years later, upon the outbreak of war, the British statesman David Lloyd George wrote: "'Boom!' The deep notes of Big Ben rang out into the night, the first strokes in Britain's most fateful hour since she arose out of the deep. A shuddering silence fell upon the room. Every face was suddenly contracted in a painful intensity. 'Doom!' 'Doom!' 'Doom!' to the last stroke. The big clock echoed in our ears like the hammer of destiny. What destiny? Who could tell? We had chal-

[14] W.C. Sellar & R.J. Yeatman, *1066 and All That: A Memorable History of England* (1930) (Harmondsworth: Penguin, 1969), 75.

lenged the most powerful military empire the world has yet brought forth."[15] Long before the Germans started their infamous Zeppelin-raids on England, the bomb blasts were already being heard in the chimes of time, the sounds of familiar bells suddenly acquiring the threatening tones of a deathknell.

In Vienna, however, Freud did not wait for the announcement that Europe was at war nor, for that matter, for the crack of Gavrilo Princip's Browning killing the Archduke and his wife, to hear the deafening sound of an explosion. On 22 June 1914, six days before Sarajevo, he had already written to Ferenczi: "We are living under the expectation of the 'bomb,' which is supposed to be sent out as soon as it comes in."[16] One did not have to wait long for the device to be launched, for three days later Freud wrote to Abraham: "So the bombshell has now burst, we shall soon discover with what effect. I think we shall have to allow the victims two to three weeks to collect themselves and react ..."[17] The bomb in question, here, had in fact been homemade at Berggasse 19 during the early months of 1914, and if it took four months to be activated, it was purely on account of the time it had taken to be proofed. Of course, Freud's incendiary device was not a real bomb, but rather an intellectual explosive, which he had called 'On the History of the Psycho-Analytic Movement'.[18] The vehemently polemical text did not appear until October 1914 in the *Jahrbuch der Psychoanalyse*, yet as soon as a printed version was ready, during the last week of June 1914, Freud arranged for offprints to be sent out to all key members of the International Psycho-Analytic Association.[19] And although the essay contained more than a few, one word to the wise was definitely enough. For example, on 4 July 1914, after having read his own personal copy of the text, the eminent Swiss psychiatrist Eugen Bleuler wrote to Freud: "In

---

15 D. Lloyd George, *War Memoirs*, Vol. 1 (Boston MA: Little, Brown and Co., 1934), 17.

16 E. Brabant, E. Falzeder & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 1, 1908-1914, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1993), 562.

17 E. Falzeder (ed.), *The Complete Correspondence of Sigmund Freud and Karl Abraham 1907-1925*, Trans. C. Schwarzacher (London-New York NY: Karnac Books, 2002), 251.

18 S. Freud, 'On the History of the Psycho-Analytic Movement,' (1914*d*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 14, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1957), 7-66.

19 E. Falzeder (ed.), *The Complete Correspondence of Sigmund Freud and Karl Abraham 1907-1925*, Trans. C. Schwarzacher (London-New York NY: Karnac Books, 2002), 250.

secondary school I was once told that the Peloponnesian War ... had been a blessing, because without it we wouldn't have had, as a lasting gain for humanity, the historiography of Thucydides. This is what came to my mind when I read your history of the psychoanalytic movement, which I greatly enjoyed despite its dubious subject matter."[20]

Before World War I was unleashed upon the world, Freud was already at war. At the suggestion of his British ally Ernest Jones, he had established, in the Summer of 1912, a small 'secret committee' of loyal followers in order to "take care of the further development of ΨA [*sic*] and defend the cause against personalities and accidents."[21] At that time, personalities were distinctly more threatening than accidents though, and amongst these personalities one in particular stood out as a fearsome, and quite formidable enemy: Carl Gustav Jung. Once welcomed within the psychoanalytic fold as a prodigious godsend, Jung had gradually fallen from grace, like an unruly son disowned by his adoptive father, not in the least because he had dared to question the latter's unassailable authority, going so far as to propose his own, more de-sexualised and less rationalistic version of the human mind. For Freud and his newly adopted children, Jung and his Swiss allies were more than an agonising thorn in the side; they were an overtly hostile force, who could inflict irreparable damage onto the fledgling, and therefore intrinsically vulnerable body of psychoanalysis. And so the little group Freud did not hesitate to describe as "the brutal, sanctimonious Jung and his parrots" had to be defeated, ostracized, preferably exterminated altogether.[22] Although he did devote quite a few pages in it to Alfred

---

20 M. Schröter (ed.), *"Ich bin zuversichtlich, wir erobern bald die Psychiatrie": Sigmund Freud-Eugen Bleuler Briefwechsel 1904-1937* (Basel: Schwabe Verlag, 2012), 205. For other reactions, see Max Eitingon's letter to Freud of 6 July 1914 in M. Schröter (ed.), *Sigmund Freud-Max Eitingon Briefwechsel 1906-1939*, Vol. 1 (Tübingen: Diskord, 2004), 89-90, as well as Oskar Pfister's letter of 9 July in I. Noth (ed.), *Sigmund Freud-Oskar Pfister: Briefwechsel 1909-1939* (Zürich: Theologischer Verlag Zürich, 2014), 100-103, and Lou Andreas-Salomé letter of 5 July in E. Pfeiffer (ed.), *Sigmund Freud and Lou Andreas-Salomé – Letters*, Trans. W. & E. Robson-Scott (London: The Hogarth Press and the Institute of Psycho-Analysis, 1972), 18-19.

21 R. A. Paskauskas (ed.), *The Complete Correspondence of Sigmund Freud and Ernest Jones 1908-1939* (Cambridge MA-London: The Belknap Press of Harvard University Press, 1993), 147. See also Ph. Grosskurth, *The Secret Ring: Freud's Inner Circle and the Politics of Psycho-Analysis* (Reading MA: Addison-Wesley, 1991); M. Schröter, 'Freuds Komitee 1912-1914. Ein Beitrag zum Verständnis psychoanalytischer Gruppenbildung,' *Psyche*, 1995, 49, 513-563; G. Wittenberger, *Das 'Geheime Komitee' Sigmund Freuds. Institutionalisierungsprozesse in der Psychoanalytischen Bewegung zwischen 1912 und 1927* (Tübingen: Diskord, 1995).

22 E. Falzeder (ed.), *The Complete Correspondence of Sigmund Freud and Karl Abraham 1907-1925*, Trans. C. Schwarzacher (London-New York NY: Karnac Books, 2002), 251.

Adler's secession, almost the entire third part of Freud's scatter bomb was directed against Jung and the so-called 'Neo-Zürich theory', which he designated as "so obscure, unintelligible and confused as to make it difficult to take up any position against it," but which he nonetheless attacked in no uncertain, and sharply belligerent terms.[23] Freud even went so far as to quote the words of an 'ill-fated' patient of Jung's, without obtaining his prior consent, in support of the fact that what was happening in Switzerland no longer had anything to do with psychoanalysis. When it exploded, Freud's incendiary device cleared the psychoanalytic landscape from all perilous impurities, as Freud and his allies had been hoping for, but in reality he had dropped it onto a piece of earth that had already been substantially scorched: although the Zürich society officially withdrew from the International Psycho-Analytical Association on 10 July 1914, Jung had already tendered his resignation as President in April that year, owing to irreconcible differences of opinion.[24] And much like the Great War would enduringly shape the future of Freudian psychoanalysis, Jung's carefully orchestrated expulsion from Freud's intimate circle would prove to be the necessary catalyst for the release of his deepest creative forces. Without being directly affected by the devastating firestorms raging across Europe, because he lived comfortably in a lavish mansion on the shores of Lake Zürich, Jung emerged from Freud's war feeling simultaneously traumatized and liberated, which enabled him to enter an amorphous, yet superabundant mental space, from which the whole of his subsequent theoretical elaborations would emanate.[25]

Yet Freud's carefully crafted, Zürich-bound bomb was not the only indication of his already being at war before the war. In his clinical practice too, weapons were being drawn and battles were waged, as any attentive reader of the series of technical papers which he had started in 1911 may

---

23 S. Freud, 'On the History of the Psycho-Analytic Movement,' (1914*d*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 14, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1957), 60-62.

24 W. McGuire (ed.), *The Freud/Jung Letters*, Trans. R. Manheim & R.F.C. Hull (Princeton NJ-London: Princeton University Press, 1974), 551. I have not found any evidence that Jung himself formally responded to Freud, but he did express his dismay in letters to other members of the psychoanalytic community. For example, he wrote to the Swedish psychoanalyst Poul Carl Bjerre that an enemy who unscrupulously breaks patient confidentiality is not even worthy of that name. See G. Adler & A. Jaffé (eds), *C. G. Jung Letters*, Trans. R.F.C. Hull, Vol. 2: 1951-1961 (Princeton NJ-London: Princeton University Press, 1976), xxix-xxx.

25 See D. Bair, *Jung: A Biography* (Boston MA-London: Little, Brown and Company, 2003), 254.

ascertain. In a short, but influential essay entitled 'The Dynamics of Transference', Freud addressed the complicated intersection of transference and resistance, arguing that a patient's refusal to comply with the psychoanalytic groundrule of free association generally stems from a 'transference-resistance', which is fuelled by powerful unconscious impulses that resist being remembered and merely strive to reproduce themselves. Freud contended that it was the psychoanalyst's duty to challenge these processes, despite its eliciting a vicious conflict between patient and practitioner: "This struggle between the doctor and the patient, between intellect and instinctual life [*Triebleben*], between understanding and seeking to act, is played out almost exclusively in the phenomena of transference. It is on that field that the victory must be won – the victory whose expression is the permanent cure of the neurosis."[26] When treating his patients, here, Freud conceived of neurosis as an evil force which needs to be conquered with strategic planning, clinical tact and, most important of all, the healing power of reason. Freud laboured the point in a subsequent essay called 'On Beginning the Treatment'. Warning trainee-analysts against the temptation to grant a patient's request that the analyst him- or herself select an appropriate topic for discussion, he emphasized: "We must bear in mind what is involved here. A strong resistance has come to the front in order to defend the neurosis; we must take up the challenge then and there and come to grips with it."[27] In fact, accurate and idiomatic as this translation may be, the original German is much more evocative: "*man nehme die Herausforderung sofort an und rücke ihm an den Leib*", i.e. one takes up the challenge and confronts it directly, as if engaging in man to man combat.[28] One year later, in 'Remembering, Repeating and Working-Through', Freud once again underscored how an unconscious resistance may prevent patients from remembering the historical circumstances presiding over their symptoms, compelling them instead to repeat these very symptoms during the course of the psychoanalytic treatment, by way

[26] S. Freud, 'The Dynamics of Transference,' (1912*b*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 12, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1958), 108.

[27] S. Freud, 'On Beginning the Treatment (Further Recommendations on the Technique of Psycho-Analysis I),' (1913*c*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 12, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1958), 137.

[28] S. Freud, 'Zur Einleitung der Behandlung (Weitere Ratschläge zur Technik der Psycho-analyse I),' (1913*c*), *Gesammelte Werke*, Vol. 8 (London: Imago, 1943), 471.

of acting out. "The patient brings out of the armoury of the past," he wrote, "the weapons with which he defends himself against the progress of the treatment – weapons which we must wrest from him one by one."[29] In the final paper of the series, which he considered "the best and most useful" one in the entire collection, and which was composed shortly after the Swiss enemy had been defeated, Freud summed up the three main combat zones of the psychoanalytic battleground: "The analytic psychotherapist thus has a threefold battle to wage – in his own mind against the forces which seek to drag him down from the analytic level; outside the analysis, against opponents who dispute the importance he attaches to the sexual instinctual forces [*sexuellen Triebkräfte*] and hinder him from making use of them in his analytic technique; and inside the analysis, against his patients, who at first behave like opponents but later on reveal the overvaluation of sexual life which dominates them, and who try to make him captive to their socially untamed passion."[30]

Before real-life military operations occupied people's minds all over the world, Freud was already fighting his own private war, with himself, with actual and potential intellectual defectors, and with his patients. Even when the latter had stopped behaving as opponents, because they had fallen prey to the temptations of transference-love, Freud still believed he had to defend himself, lest he be taken hostage to precarious erotic fortunes. Freud's gloomy portrayal of psychoanalysis as a threefold battle, here, was not a reflection of the socio-political hostilities surrounding him. Strange as it may appear, it was rather the other way round. When the Central Powers went to war in July 1914, the declaration epitomized a social externalisation of seemingly unresolvable psychic conflicts Freud had already detected for a good number years – in his neurotic patients, in some of his closest supporters and, perhaps most important of all, in himself. Indeed, I would venture to claim that the intra-psychic battle Freud was fighting

---

29 S. Freud, 'Remembering, Repeating and Working-Through (Further Recommendations on the Technique of Psycho-Analysis II),' (1914*g*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 12, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1958), 137.

30 S. Freud, 'Observations on Transference-Love (Further Recommendations on the Technique of Psycho-Analysis III),' (1915*a*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 12, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1958), 170. On this being Freud's favourite essay of the series, see E. Falzeder (ed.), *The Complete Correspondence of Sigmund Freud and Karl Abraham 1907-1925*, Trans. C. Schwarzacher (London-New York NY: Karnac Books, 2002), 300.

during the pre-War years, and which revolved around an ardent desire for social and intellectual recognition, constituted the *fons et origo* of the two other struggles, including his clinical deployment of all the ammunition in his psychoanalytic toolkit to break the neurotic patient's resistance. As Jacques Lacan would put it in his typically laconic style: "[T]he only real resistance in analysis is the resistance of the analyst."[31] Lacan was unambiguously referring to the direction of the psychoanalytic treatment, yet there is no reason as to why we should not extrapolate the principle to the historical and ongoing battleground that goes by the name of 'the psychoanalytic movement'.

Four Years of 'Splendid Isolation'

Whilst sharing in the elation of his fellow citizens, albeit rather briefly and perhaps less openly, Freud could not have anticipated what hardship, misery and anguish the outbreak of the hostilities would bring. On 12 August 1914, when Great Britain declared war on the Austro-Hungarian Empire, Freud's youngest daughter Anna, not yet nineteen at the time, found herself on her first overseas trip alone trapped in England, where she became officially an 'alien enemy'. For almost a month, the Freud family did not hear anything from her, until she finally arrived back at the Berggasse on 26 August, after having managed to board a ship from Falmouth to Genoa ten days earlier.[32] At first, Freud did not think his three sons would be affected, because Martin (the eldest) had been exempted from duty owing to a previously sustained thigh-fracture, the middle son Oliver had already been considered unfit for military service during an earlier medical examination, and Ernst (the youngest) was in the reserves.[33] However, all three ended up in military uniform. Against the will of his father, Martin volunteered for the Imperial Field Artillery in August 1914,

---

31 J. Lacan, *The Seminar. Book II: The Ego in Freud's Theory and in the Technique of Psychoanalysis* (1954-55), Trans. S. Tomaselli (Cambridge: Cambridge University Press, 1988), 324.

32 See M. Molnar, 'Portrait of an Alien Enemy,' in *Looking Through Freud's Photos* (London: Karnac Books, 2015), 91-106; I. Meyer-Palmedo (ed.), *Sigmund Freud-Anna Freud. Correspondence 1904-1938*, Trans. N. Somers (Cambridge: Polity, 2014), 88-92; M. Schröter (ed.), *Unterdeß halten wir zusammen. Sigmund Freud – Briefe an die Kinder* (Berlin: Aufbau Verlag, 2011), 493.

33 See E. Falzeder (ed.), *The Complete Correspondence of Sigmund Freud and Karl Abraham 1907-1925*, Trans. C. Schwarzacher (London-New York NY: Karnac Books, 2002), 265; M. Schröter (ed.), *Unterdeß halten wir zusammen. Sigmund Freud – Briefe an die Kinder* (Berlin: Aufbau Verlag, 2011), 108, 220 and 258.

and served as a gunner on the front line throughout the war.[34] As a fully qualified engineer, Oliver first became involved in the construction of strategically important railways and tunnels, but in December 1916 he decided to exchange the poor working conditions as a civilian for an allegedly more agreeable life as an officer in a batallion of sappers.[35] Like his elder brother, Ernst joined the artillery. During the last week of October 1915, his entire platoon was wiped out when their dugout was struck directly by an Italian shell. By pure coincidence, Ernst had just left the shelter, thus miraculously avoiding the lethal barrage. In August 1917, he was hospitalised with a duodenal ulcer in Zagreb and eventually transferred back to Vienna, which ushered in the end of his military duties.[36]

As his numerous letters to Ferenczi, Abraham, Jones and others testify, Freud was gravely concerned about the health and well-being of his sons throughout the war, and for very good reasons. Yet he was equally worried about the fate of his intellectual brothers in arms, and by extension about the future of the psychoanalytic movement. Sándor Ferenczi served as an assistant physician with the Hungarian Hussars, whereas Karl Abraham was initially stationed at an impromptu military hospital near Berlin.[37] Freud's other close collaborators also became, if not entirely estranged from the cause, at least preoccupied with other things. In January 1916, Otto Rank, for many years Freud's most trusted amanuensis, was assigned the editorship of an army newspaper in Krakow, where he would reside

34 See M. Schröter (ed.), *Unterdeß halten wir zusammen. Sigmund Freud – Briefe an die Kinder* (Berlin: Aufbau Verlag, 2011), 108; M. Freud, *Glory Reflected: Sigmund Freud – Man and Father* (London: Angus and Robertson, 1957), 178.

35 See M. Schröter (ed.), *Unterdeß halten wir zusammen. Sigmund Freud – Briefe an die Kinder* (Berlin: Aufbau Verlag, 2011), 221; O. Freud, *Memories of World War I*, Library of Congress, Manuscript Division, Oliver and Henny Freud Papers.

36 See E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 86 and 233; M. Schröter (ed.), *Unterdeß halten wir zusammen. Sigmund Freud – Briefe an die Kinder* (Berlin: Aufbau Verlag, 2011), 259; M. Giefer & C. Tögel (eds), *Sigmund Freud. Die Kalendereinträge von 1916-1918* (Frankfurt am Main-Basel: Stroemfeld, 2016), 193-194.

37 See E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 9-21; I. Sanfeliu, *Karl Abraham: The Birth of Object-Relations Theory* (2002), Trans. K. Walters (London-New York NY: Karnac Books, 2014), 85-86; K. Zienert-Eilts, *Karl Abraham: Eine Biografie im Kontext der psychoanalytischen Bewegung* (Gießen: Psychosozial-Verlag, 2013), 126-128; A. Bentinck van Schoonheten, *Karl Abraham: Life and Work, a Biography* (2013), Trans. L. Waters (London-New York NY: Karnac Books, 2016), 141.

until the end of the war.[38] Hanns Sachs, another member of the Secret Committee, was initially mustered out because of myopia, but was nonetheless called up in August 1915, although after three months of military training he was eventually discharged again.[39] The wealthy itinerant neuropsychiatrist Max Eitingon volunteered for the Austrian army and worked as a physician in various locations in Hungary.[40] Almost overnight and for the foreseeable future Ernest Jones had become an enemy. With Jung and the Swiss already departed, what would become of the psychoanalytic undertaking were all of them to perish? Furthermore, it remained uncertain when, if ever, the annual international psychoanalytic conferences would be resumed. More importantly, perhaps, it was not at all clear how information and announcements pertaining to key psychoanalytic research activities would be disseminated in the absence of regular publications (due to paper shortages and falling subscriptions), and without reliable distribution channels (owing to postal restrictions and censorship). In short, with the Great War the sustainability of the entire psychoanalytic community was thrown into turmoil, and on occasion it must have felt to Freud as if the homemade bomb he had inscribed as the history of the psychoanalytic movement had effectively coincided with the end of psychoanalytic history.[41]

To Freud, as to millions of other people around the world, the First World War brought pain, penury and precariousness. Yet quite unexpectedly it also signalled the end of Freud's own private war. With Adler and Jung out of the way, and the Secret Committee in place to protect the interests of psychoanalysis, intellectual opponents need not be feared anymore. Whereas prior to the war Freud spent most of his time during the day

---

38 See E. J. Lieberman & R. Kramer (eds), *The Letters of Sigmund Freud and Otto Rank: Inside Psychoanalysis*, Trans. G. C. Richter (Baltimore MD-London: The Johns Hopkins University Press, 2012), 52; P. Avrane (ed.), *Sigmund Freud-Otto Rank – Correspondance 1907-1926*, Trans. S. Achache-Wiznitzer, B. Aubenas, J. Dupont & F. Samson (Paris: CampagnePremière/, 2015), 20-22.

39 See E. Jones, *The Life and Work of Sigmund Freud, Vol. II: Years of Maturity (1901-1919)* (New York NY: Basic Books, 1955), 176 & 181; M. Giefer & C. Tögel (eds), *Sigmund Freud. Die Kalendereinträge von 1916-1918* (Frankfurt am Main-Basel: Stroemfeld, 2016), 147.

40 See M. Schröter (ed.), *Sigmund Freud-Max Eitingon Briefwechsel 1906-1939*, Vol. 1 (Tübingen: Diskord, 2004), 97.

41 The Vienna Psychoanalytic Society continued to meet on a fortnightly basis until May 1915, yet attendance gradually diminished. At the end of May 1915, Freud decided to interrupt the meetings, and the Society did not reconvene until November 1918, one week after the armistice. See H. Nunberg & E. Federn (eds), *Minutes of the Vienna Psychoanalytic Society, Vol. 4: 1912-1918*, Trans. M. Nunberg & H. Collins (New York NY: International Universities Press, 1975).

seeing patients, as the military conflict intensified his case-load was almost reduced to zero. In early December 1914, he wrote to Ferenczi that he had "exactly two patients, both Hungarians", but in the following years even the Hungarians would no longer find their way to Freud's couch.[42] During the war, there was no need for Freud to be at war with his patients' neuroses, because his neurotic patients were either dutifully engaged otherwise, or too preoccupied with the life-threatening predicaments of their everyday existence to experience the burdens of neurotic suffering. Since the bulk of Freud's income depended on his clinical practice, seeing his couch empty could only have added to his and his immediate family's forlorn situation, but it also provided him with precious time to pause and reflect. Counterfactual history as it may be, had it not been for the war, Freud would have probably carried on trying to tackle his patients' unconscious resistance, wrestling with their neuroses, and claiming psychoanalytic victory on the battleground of his consultation room every time an irrational neurotic impulse had conceded to the forces of clinical reason. Now, for reasons over which he himself could not exert any influence, he found himself suddenly abandoned – forcibly deprived of his main resources, but also liberated from all the quotidian routines, and above all free to take stock, review issues and accomplishments, and explore hitherto uncharted territories.

In the first weeks after the outbreak of war, Freud managed very little, apart from examining his collection of antiquities, being generally irritable, and making slips of the tongue all day long.[43] Yet over time, and without the dust settling, his physical energy grew stronger and his emotional discomfort gave way to a seemingly inexhaustible source of

---

42 E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 34. Freud's patient calendars for the years 1914-1918 show that between the outbreak of war (28 July 1914) and the signing of the armistice (11 November 1918), Freud saw a mere six 'regular' patients, although with significant interruptions, and no more than four analytic candidates, although all for relatively short periods of time. The calendars also demonstrate that during the first year of the war (from August 1914 until August 1915), only one new patient (Karl Mayreder) found his way to the Berggasse, for an analysis of 41 sessions, over a period of just 2,5 months. See U. May, *Freud bei der Arbeit. Zur Entstehungsgeschichte der psychoanalytischen Theorie und Praxis, mit einer Auswertung von Freuds Patientenkalender* (Gießen: Psychosozial-Verlag, 2015), 247-348. The original patient calendars are in the archives of the Freud Museum London.

43 E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 13.

creative vigour. Before 1914 was over, he completed the last essay in the series on psychoanalytic technique (on 'transference love'), turned his notes of the four-year-long analysis of a young Russian into a lengthy case-study, reworked his *Three Essays on the Theory of Sexuality*, and commenced a new, large-scale project of daring speculation.[44] On 25 November 1914, he wrote to Ferenczi: "In addition to the 112-page strong case history, something else is in process which shouldn't be talked about yet ... I only want to reveal to you that, on paths that have been trodden for a long time, I have finally found the solution to the riddle of time and space and the long-sought mechanism for the release of anxiety."[45] Twenty days later, things were still looking bright, despite a lingering intestinal discomfort: "In exchange for that the work is again going well. I am living, as my brother says, in my private trench; I speculate and write, and after hard battles have got safely through the first line of riddles and difficulties. Anxiety, hysteria, and paranoia have capitulated ... I am now, and, because of the foreseeable consequences of the war, will also be later, more isolated than ever ..."[46] By contrast with the bellicose language that had pervaded his technical papers before the war, the military metaphors, here, exude a certain degree of lightheartedness, almost to the point where they become playful and frivolous – suffused with mild irony and superior detachment. Although Freud had every reason not to be cheerful, the war seems to have unlocked in him a cornucopia of creativity that he had not enjoyed for twenty years – since the mid 1890s, in fact, when he had repeatedly confided in Wilhelm Fließ

44 S. Freud, 'Observations on Transference-Love (Further Recommendations on the Technique of Psycho-Analysis III),' (1915*a*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 12, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1958), 157-171; S. Freud, 'From the History of an Infantile Neurosis' (1918*b*[1914]), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 17, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1955), 1-122; E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 20-22.

45 E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 29-30.

46 E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 36-37.

about the virtuous void of his splendid isolation.[47] On 3 July 1915, he disclosed to Abraham: "I think I regard the situation as a repetition of the initial one, when I was productive and – isolated."[48]

With his clinical practice virtually at a stand-still, and with both his biological and his adopted sons contributing to the war-efforts, Freud entered a deep state of abstract speculation which many an outsider would not have hesitated to designate as philosophical or even metaphysical, but which he himself described as metapsychological. The labour was supposed to have crystallised in a substantial monograph of twelve inter-connected chapters called *In Preparation of the Metapsychology* (*Zur Vorbereitung der Metapsychologie*), to be published after the war.[49] Yet for reasons that no one has ever been able to explain the book was never released and only five of the twelve sections were ever published, the manuscripts of the seven other pieces disappearing without a trace, with the exception of the twelfth and final essay, which was accidentally discovered in 1983 in an old crate of documents Ferenczi had bequeathed to his Hungarian pupil Michael Balint.[50] Ernest Jones has claimed that the metapsychological papers "are among the most profound and important of all Freud's works. The originality of his penetration into the theory of the mind in them was so novel that they need very careful study."[51] However, Jones's judgement is merely based on his evaluation of the five surviving essays – had the others also

---

47 See J. M. Masson (ed.), *The Complete Letters of Sigmund Freud to Wilhelm Fließ 1887-1904*, Trans. J. M. Masson (Cambridge MA-London: The Belknap Press of Harvard University Press, 1985), 181-185. See also S. Freud, 'On the History of the Psycho-Analytic Movement,' (1914*d*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 14, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1957), 22.

48 E. Falzeder (ed.), *The Complete Correspondence of Sigmund Freud and Karl Abraham 1907-1925*, trans. C. Schwarzacher (London-New York NY: Karnac Books, 2002), 313.

49 See R. A. Paskauskas (ed.), *The Complete Correspondence of Sigmund Freud and Ernest Jones 1908-1939* (Cambridge MA-London: The Belknap Press of Harvard University Press, 1993), 312. In a contemporaneous letter to Abraham, Freud referred to the book as *Essays in Preparation of the Metapsychology* (*Abhandlungen zur Vorbereitung der Metapsychologie*). See E. Falzeder (ed.), *The Complete Correspondence of Sigmund Freud and Karl Abraham 1907-1925*, trans. C. Schwarzacher (London-New York NY: Karnac Books, 2002), 309. These letters clearly demonstrate that by the Spring of 1915 Freud's initial secrecy about the project had already been replaced with a fervent desire to share his newest work with his closest intellectual allies.

50 See S. Freud, *A Phylogenetic Fantasy: Overview of the Transference Neuroses* (1915), ed. I. Grubrich-Simitis, Trans. A. Hoffer & P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1987).

51 E. Jones, *The Life and Work of Sigmund Freud, Vol. II: Years of Maturity (1901-1919)* (New York NY: Basic Books, 1955), 185.

been made available, he may have been less, yet no doubt even more complimentary – and it also ignores the extraordinary circumstances under which Freud had devoted himself to the task. When millions of young men were being killed or maimed in the trenches, when Freud's fingers started to freeze because fuel had been rationed, when those fortunate enough to be spared the barbaric conditions on the frontline were still constantly at risk of dying of cold or starvation, Sigmund Freud speculated about the unconscious, repression and the drives. It is tempting to think that he fell under the spell of the 'witch of metapsychology' despite the war, yet I think it is more accurate to say that he became mesmerized by her *because of* the war, because the war had paradoxically and unintentionally terminated his own private war, because in the hours of his unexpected abandonment, when the destiny of human life was anyone's guess, he could wholeheartedly abandon himself to uninhibited guesswork, thus giving new form to the theory and practice of psychoanalysis.[52] After the war, and because of it, psychoanalytic theory had been deepened, and its clinical applications had been transformed accordingly, from the analytic breaking down of the patient's resistance to the patient facilitation of analytic working-through.

But the Great War did not just shape psychoanalysis in terms of its foundational constructs. Freud did not become so enchanted by the solitary pleasures of intellectual speculation that he remained completely oblivious to what was going on in the outside world. Balanced against the high levels of abstraction of the metapsychological papers, he produced two substantial essays on war and death, and a shorter piece on transience.[53] Although some of the materials in these texts drew on previously elaborated ideas, including those developed in *Totem and Taboo*, the three essays epitomized Freud's first explicit commentary on contemporary social issues, and prefigure his more expansive forays into the realms of social theory from the 1920s and 30s, which culminated in a famous open letter to Albert

---

52 Freud himself designated his metapsychology as a witch in his late essay 'Analysis Terminable and Interminable'. See S. Freud, 'Analysis Terminable and Interminable' (1937*c*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 23, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1964), 225.

53 See S. Freud, 'Thoughts for the Times on War and Death' (1915*b*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 14, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1957), 273-300; S. Freud, 'On Transience' (1916*a*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 14, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1957), 303-307.

Einstein of September 1932.[54] The other exoteric part of Freud's intellectual activities during the war would become one of the most succesful ventures of his entire career. Because his clinical work had ground to a halt, Freud accepted to deliver a series of introductory public lectures on key psychoanalytic principles at the University of Vienna. Since his appointment as Professor Extraordinarius in 1902, he had regularly lectured to university students, yet his involvement with academic life had never been more than peripheral, and this is no doubt how Freud would have wanted it, given his demanding case-load and his commitment to writing projects.[55] During the war, Freud found himself involuntarily relieved of most of his patients, and so he agreed to prepare, one last time, a full course of lectures for both the Winter semesters of 1915/16 and 1916/17. At the insistence of Rank, Freud also agreed to publish the twenty-eight lectures (fifteen from the first course and thirteen from the second) in three separate instalments, and subsequently as a single volume.[56] Perhaps owing to their expository character, or because they were primarily addressed to a non-specialist audience, the *Introductory Lectures on Psycho-Analysis* went through four editions in the space of five years, and they continue to be both one of Freud's most popular works, and an invaluable resource for teachers and practicing psychoanalysts alike, also because the third part of the book, on the general theory of the neuroses, contains numerous new insights on clinical and technical matters.[57]

By far the most significant impact of the war on psychoanalysis, however, was definitely not of Freud's own making. As the Central Powers exchanged fire with the Allied Forces on Eastern and Western fronts, and bombs relentlessly rained upon dugouts and trenches, more and more men were sent home suffering from nervous conditions, which some medical doctors designated as neurasthenia (literally 'weakness of the nerves' or

---

54 See S. Freud, 'Totem and Taboo,' (1912-13), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 13, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1955), 1-161; S. Freud, 'Why War?,' (1933*b*[1932]), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 22, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1955), 195-215.

55 On Freud's academic career, see J. Gicklhorn & R. Gicklhorn, *Sigmund Freuds akademische Laufbahn im Lichte der Dokumente* (Wien-Innsbrück: Urban & Schwarzenberg, 1960).

56 M. Giefer & C. Tögel (eds), *Sigmund Freud. Die Kalendereinträge von 1916-1918* (Frankfurt am Main-Basel: Stroemfeld, 2016), 119-120.

57 See S. Freud, Introductory Lectures on Psycho-Analysis (1916-17), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vols 15/16, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1963).

'nervous exhaustion') and others as 'shell-shock'.[58] It has been estimated that at the end of the war, hundreds of thousands of soldiers of all ranks, and from all sides of the battlelines, had at some point been admitted to army hospitals with various sensori-motor symptoms of war-induced trauma.[59] In many instances, it was believed that these 'casualties' were simply malingerers, i.e. cowardly men feigning nervous illness, or outright madness, in order to escape the lethal horrors of the trenches, or with the prospect of obtaining a handsome veteran's pension. This view was reinforced by the fact that, in the majority of cases, doctors could not find any reliable evidence of severe external physical trauma, the detection of internal injuries also being hampered by the lack of effective diagnostic techniques.[60] During the first years of the war, many of these men thus ran the risk of being either executed as deserters, or sent straight back to wherever they had come in from. Yet because it was observed that quite a few of those who had been discharged from duty on neurasthenic grounds were hardened soldiers, some of them high-ranking officers who had previously been decorated for bravery, military authorities and army doctors gradually came to change their mind about the aetiology of the clinical condition. Instead of attributing

---

58 The term 'neurasthenia' is generally credited to the American neurologist George Beard, who first used it in 1869. As to 'shell-shock', it is believed that the idea probably started life amongst the rank and file servicemen, as a shorthand description for fellow soldiers who had been retired from active duty owing to their having collapsed emotionally under the constant physical blasts of bombshells, yet the notion first appeared in print in a 1915 paper by the English physician-psychologist Charles Myers, who also described its main symptoms on the basis of three clinical cases. See G. Beard, 'Neurasthenia, or Nervous Exhaustion,' *The Boston Medical and Surgical Journal*, 1869, 3(13), 217-221; C. S. Myers, 'A Contribution to the Study of Shell Shock: Being an Account of Three Cases of Loss of Memory, Vision, Smell, and Taste, Admitted into the Duchess of Westminster's War Hospital, Le Touquet,' *The Lancet*, 1915, 185(4772), 316-330. For additional source-materials on the conceptual history and the clinical applications of these categories, see S. Wessely, 'Neurasthenia and Fatigue Syndromes – Clinical Section,' in G. Berrios & R. Porter (eds), *A History of Clinical Psychiatry: The Origin and History of Psychiatric Disorders* (London-New Brunswick NJ: The Athlone Press, 1995), 509-532; M. Gijswijt-Hofstra & R. Porter (eds), *Cultures of Neurasthenia: From Beard to the First World War* (Amsterdam-New York NY: Rodopi, 2001); T. Loughran, 'Shell Shock, Trauma, and the First World War: The Making of a Diagnosis and its Histories,' *Journal of the History of Medicine and Allied Sciences*, 2012, 67(1), 94-119; P. Leese, *Shell Shock: Traumatic Neurosis and the British Soldiers of the First World War* (Basingstoke: Palgrave Macmillan, 2002).

59 The vagueness of the number is due to the fact that historians can only rely on clinical archives and pension records in order to arrive at a rough estimate of the number of soldiers affected, yet the figures are nonetheless staggering. For example, it has been estimated that by 1918 some 80,000 British servicemen had suffered from some form of psychological breakdown. See M. Stone, 'Shellshock and the Psychologists,' in W. F. Bynum, R. Porter & M. Shephard (eds), *The Anatomy of Madness: Essays in the History of Psychiatry. Vol. 1: People and Ideas* (London-New York NY: Routledge, 1985), 242-271.

60 See, for example, J. Bourke, *Dismembering the Male: Men's Bodies, Britain and the Great War* (London: Reaktion Books, 1996), 76-123.

the symptoms to a conscious intention to deceive, they started to reckon with the possibility that neurasthenic soldiers were genuinely ill, and should therefore be treated rather than punished. And over time, the hypothetical explanation for these newly recognized clinical disorders of 'war neurosis' also changed. Whereas it was initially believed that neurasthenic soldiers were suffering from 'commotional shock' – numerous small brain haemorrhages resulting in physical paralysis of the nervous system, owing to the sharp increase in the atmospheric pressure at the explosion of a bombshell – this organic account was subsequently dismissed as too narrow and generally unsatisfactory, not in the least because a sizeable proportion of shell-shocked patients had not been directly exposed to shattering explosions, or had never even been under fire. And so new, psychological theories were being advanced, ranging from a defective mental reaction and a failed adaptation mechanism to the pathological suppression of psychic conflicts, and the re-awakening of old traumatic events. The latter explanations were substantiated by the observation that in quite a few cases the shell-shocked soldiers were displaying symptoms whose appearance mirrored the paralyses and convulsions that had been described at the end of the nineteenth century in cases of female hysteria.[61]

[61] My account of the changing conceptions of 'war neurosis' in the preceding paragraph is evidently too superficial and lacking in nuance, yet within the space of this essay I cannot possibly do justice to the intricacies of the historical debate on the aetiology and treatment of shell-shock and, by extension, of traumatic neurosis. For detailed analyses and extensive critical commentary, also taking into consideration the variegated symptomatology of the condition in different countries, see E. J. Leed, *No Man's Land: Combat and Identity in World War I* (Cambridge: Cambridge University Press, 1979); H. Binneveld, *From Shellshock to Combat Stress: A Comparative History of Military Psychiatry*, Trans. J. O'Kane (Amsterdam: Amsterdam University Press, 1997); L. Crocq, *Les traumatismes psychiques de guerre* (Paris: Odile Jacob, 1999); B. Shephard, *A War of Nerves: Soldiers and Psychiatrists 1914-1994* (London: Jonathan Cape, 2000); M. S. Micale & P. Lerner (eds), *Traumatic Pasts: History, Psychiatry and Trauma in the Modern Age, 1870-1930* (Cambridge: Cambridge University Press, 2001); P. Leese, *Shell Shock: Traumatic Neurosis and the British Soldiers of the First World War* (Basingstoke: Palgrave Macmillan, 2002); P. Lerner, *Hysterical Men: War, Psychiatry, and the Politics of Trauma in Germany, 1890-1930* (Ithaca NY-London: Cornell University Press, 2003); P. Barham, *Forgotten Lunatics of the Great War* (New Haven CT-London: Yale University Press, 2004); E. Jones & S. Wessely, *Shell Shock to PTSD: Military Psychiatry from 1900 to the Gulf War* (Hove-New York NY: Psychology Press, 2005); G. M. Thomas, *Treating the Trauma of the Great War: Soldiers, Civilians, and Psychiatry in France, 1914-1940* (Baton Rouge LA: Louisiana State University Press, 2009); F. Reid, *Broken Men: Shell Shock, Treatment and Recovery in Britain 1914-30* (London: Continuum, 2011); H.-G. Hofer, C.-R. Prüll & W. U. Eckart (eds), *War, Trauma and Medicine in Germany and Central Europe (1914-1939)* (Herbolzheim: Centaurus, 2011); L. Tatu & J. Bogousslavsky, *La folie au front. La grande bataille des névroses de guerre (1914-1918)* (Paris: Imago, 2012); L. Crocq, *Les blessés psychiques de la Grande Guerre* (Paris: Odile Jacob, 2014); T. Loughran, *Shell-Shock and Medical Culture in First World War Britain* (Cambridge: Cambridge University Press, 2016); J. Crouthamel & P. Leese (eds), *Psychological Trauma and the Legacies of the First World War* (Basingstoke: Palgrave Macmillan, 2017).

Throughout the war, Freud had little or nothing to say about neurasthenia, shell-shock or war neurosis, maybe because he himself and his collaborators were not immediately affected by it (although Ferenczi did at one point complain to Freud about various somatic and psychic symptoms), or because his sons only suffered from classic 'psychosomatic' ailments (Ernst's duodenal ulcer, and eczema in the case of Oliver).[62] Owing to the fact that, in Freud's book, psychoanalytic treatment was invariably an intensive, long-term process, which required a workable degree of positive transference on the side of the patient, and which should never be offered free of charge, he may have already concluded that soldiers suffering from war neurosis would never benefit from a psychoanalytic approach – indeed, that the time-pressure for rehabilitation, the patients' involuntary participation in the treatment, and their inability to pay would exclude the possibility of clinical success.[63] In any case, when during the Winter of 1916, Ferenczi lectured on war neuroses to a group of Hungarian physicians, Freud could barely conceal his profound disinterest.[64] However, during the first weeks of February 1918, everything changed. Out of nowhere, Freud received a copy of a small volume by a certain Dr Ernst Simmel entitled *War Neuroses and Psychic Trauma* (*Kriegs-Neurosen und 'psychisches Trauma'*), in which the author claimed to have successfully treated various cases of neurasthenic soldiers with the tools of Freudian psychoanalysis.[65] Always looking for allies, Freud forgave Simmel's occasional doctrinal heresy, and shared his enthusiasm with Abraham: "This is the first time that a German physician, basing himself firmly and without patronizing condescension on ψα ground, speaks of its outstanding useful-

---

62 See E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 112. On Ernst Freud's ulcer and Oliver Freud's eczema, see E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 233 and 269.

63 On the importance of positive transference, the duration of the treatment and the need for payment, see S. Freud, 'The Dynamics of Transference' (1912*b*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 12, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1958), 97-108; S. Freud, 'On Beginning the Treatment (Further Recommendations on the Technique of Psycho-Analysis I)' (1913*c*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 12, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1958), 121-144, and 129-132 in particular.

64 See E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 113-115.

ness in the treatment of war neuroses and backs this with examples, and is also completely upright on the question of sexual aetiology. It is true that he has not gone the whole way with Ψα, takes essentially the cathartic standpoint, works with hypnosis, which is bound to conceal resistance and sexual drives from him, but he correctly apologizes for this because of the necessity of quick results and the large number of cases with which he has to deal. I think a year's training would make a good analyst of that man. His behaviour is correct."[66] What Freud did not know at the time, and what he would not know for a long time, is that a couple of weeks before the arrival of Simmel's book, on the other side of the frontline, a renowned medical anthropologist by the name of W. H. R. Rivers had published a paper in the prestigious journal 'The Lancet', in which he argued that the war neuroses cannot be properly understood without the concept of repression, yet without explicitly referring to Freud or psychoanalysis. Illustrating his argument with clinical cases, Rivers posited that the clinical symptoms of neurasthenic soldiers could be resolved through a combination of catharsis and re-education.[67] Although he was far from endorsing each and every aspect of Freud's works, and had relied (much like his counterparts in Germany) more on the proto-analytic paradigm of the hypno-cathartic method than on psychoanalysis *per se*, Rivers would become instrumental in disseminating the key principles of psychoanalytic theory and practice

---

65 E. Simmel, *Kriegs-Neurosen und "psychisches Trauma": Ihre gegenseitige Beziehung dargestellt auf Grund psycho-analytischer, hypnotischer Studien* (München-Leipzig: Nemnich, 1918); E. Falzeder, E. Brabant & P. Giampieri-Deutsch (eds), *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Vol. 2, 1914-1919, Trans. P. T. Hoffer (Cambridge MA-London: The Belknap Press of Harvard University Press, 1996), 264. Before Simmel published the results of his clinical work, three German doctors had already reported on the beneficial effect of psychoanalytic interventions on shell-shocked soldiers, yet these papers seem to have escaped Freud's attention. See F. Stern, 'Die psychoanalytische Behandlung der Hysterie im Lazarett,' *Psychiatrisch-Neurologische Wochenschrift*, 1916-17, 1/2, 1-3; W. Sauer, 'Zur Analyse und Behandlung der Kriegsneurosen', *Zeitschrift für die gesamte Neurologie und Psychiatrie*, 1917, 36, 26-45; F. Mohr, 'Aus der Praxis der Psychotherapie,' *Medizinische Klinik*, 1917, 13, 1116-1119. For a detailed discussion of these papers, see P. Lerner, *Hysterical Men: War, Psychiatry, and the Politics of Trauma in Germany, 1890-1930* (Ithaca NY-London: Cornell University Press, 2003), 165-171.

66 E. Falzeder (ed.), *The Complete Correspondence of Sigmund Freud and Karl Abraham 1907-1925*, Trans. C. Schwarzacher (London-New York NY: Karnac Books, 2002), 372. On 20 February 1918, three days after his letter to Abraham, Freud wrote to Simmel himself, expressing his deep appreciation for his work whilst simultaneously insisting that his method had been rooted in a pre-psychoanalytic approach. See S. Freud, 'Freud's Letters to Ernst Simmel,' Trans. F. Deri & D. Brunswick, *Journal of the American Psychoanalytic Association*, 1964, 12(1), 97-98.

67 W. H. R. Rivers, 'An Address on the Repression of War Experience,' *The Lancet*, 1918, 191(4927), 2 February, 173-177.

in Britain, within the academic community as well as to the general public.[68]

In ways that no one could have foreseen, the public perception and clinical acknowledgement of psychoanalysis benefited tremendously from the neurotic casualties of the Great War. This allows us to understand Freud's ostensibly bizarre remark to Ferenczi on 16 November 1918, less than a week after armistice day, that for the sake of psychoanalysis the war should have lasted longer. Some two months before, on 28 and 29 September, the Fifth International Psycho-Analytic Congress had been held in Budapest. This was the first international gathering of psychoanalysts since 1913 – although for obvious reasons the only countries represented were those of the Central Powers, enhanced with two delegates from the neutral Netherlands – and it included a special symposium on the war neuroses, featuring contributions by Abraham, Ferenczi and Simmel, who had by now been fully converted to the psychoanalytic cause. One year later, the presentations were amplified with an essay by Jones, and published as the maiden volume in the newly established International Psycho-Analytic publishing house.[69] In Budapest, Freud did not express his views on the topic of war neurosis, but for the book he wrote an introduction in which he argued that war neurosis was but a special case of the broader category of the traumatic neuroses, and that the former could not be properly understood without a re-examination of the significance of trauma for the economy of human mental life. At that time, this is what he himself had already proceeded to doing in his groundbreaking essay 'Beyond the Pleasure Principle', in which he took his lead from the traumatic neuroses in order to conceptualise an infernal 'repetition compulsion' and (much more controversially)

---

68 For extensive discussions of Rivers' career and contributions, see A. Young, 'W. H. R. Rivers and the War Neuroses,' *Journal of the History of the Behavioral Sciences*, 1999, 35(4), 359-378; J. Forrester & L. Cameron, *Freud in Cambridge* (Cambridge: Cambridge University Press, 2017), 57-99. Rivers' 1919 lectures at the University of Cambridge were subsequently published alongside a collection of papers on his experiences as an officer in the Royal Army Medical Corps during World War I. See W. H. R. Rivers, *Instinct and the Unconscious: A Contribution to a Biological Theory of the Psycho-Neuroses* (Cambridge: Cambridge University Press, 1920). It should also be noted, here, that Rivers was by no means the only medical doctor who embraced a psychoanalytic conception of the war neuroses. Other physicians who became broadly sympathetic to psychoanalysis included Frederick Mott, Charles Myers and William Brown. See, for example, T. Loughran, 'Shell Shock and Psychological Medicine in First World War Britain,' *Social History of Medicine*, 2009, 22, 79-95; E. Jones, '"An Atmosphere of Cure": Frederick Mott, Shell Shock and the Maudsley', *History of Psychiatry*, 2014, 25(4), 412-421.

69 S. Freud, S. Ferenczi, K. Abraham, E. Simmel & E. Jones, *Zur Psychoanalyse der Kriegsneurosen* (Leipzig-Wien: Internationaler Psychoanalytischer Verlag, 1919).

the insistence of a universal death drive.[70] Leaving the discussion of war neurosis to those with more direct clinical experience of it, in Budapest Freud delivered a visionary address under the programmatic title 'Lines of Advance in Psycho-Analytic Therapy', in which he envisaged the widespread creation of psychoanalytic institutions and outpatient clinics offering free treatment to the poor and to low-income, working-class people.[71] Without irony, Freud advocated the development of a 'psychotherapy for the people', in which the pure gold of psychoanalysis would have to be mixed with the copper of suggestion and hypnosis, but which would allow larger cohorts of people to benefit from psychoanalytic treatment. Had it not been for the fact that, less than two months later, the Central Powers were defeated, the proposal would no doubt have resulted quite rapidly in the establishment of various new psychoanalytic treatment centres, also because high-ranking government officials were in attendance, prepared to lend their full support to the psychoanalytic initiatives.[72] However, Freud's call did not go completely unanswered, because after the war Max Eitingon and Ernst Simmel would go on to establish the *Poliklinik für Psychoanalytische Behandlung Nervöser Krankheiten* in Berlin, which officially opened its doors on 16 February 1920, and which Freud at one point designated as the new headquarters of the psychoanalytic movement.[73] Seven years later, Simmel also managed to secure enough funds from private companies to open the first and for many years only inpatient

---

70 S. Freud, 'Beyond the Pleasure Principle,' (1920*g*), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 18, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1955), 1-64.

71 S. Freud, 'Lines of Advance in Psycho-Analytic Therapy,' (1919*a*[1918]), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 17, Trans. J. Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1955), 157-168.

72 See 'Bericht über den V. Internationalen psychoanalytischen Kongreß in Budapest, 28.-29. September 1918,' *Korrespondenzblatt der Internationalen Psychoanalytischen Vereinigung*, 1919, 1, 31-37.

73 E. Falzeder (ed.), *The Complete Correspondence of Sigmund Freud and Karl Abraham 1907-1925*, Trans. C. Schwarzacher (London-New York NY: Karnac Books, 2002), 434. See also E. A. Danto, 'The Berlin Poliklinik: Psychoanalytic Innovation in Weimar Germany,' *Journal of the American Psychoanalytic Association*, 1999, 47, 1269-1292; E. A. Danto, *Freud's Free Clinics: Psychoanalysis and Social Justice, 1918-1938* (New York NY-London: Columbia University Press, 2005), 52-56; K. Brecht, V. Friedrich, L. M. Hermanns, I. J. Kaminer & D. H. Juelich (eds), *"Here Life Goes On In a Most Peculiar Way…": Psychoanalysis Before and After 1933*, Trans. C. Trollope (Hamburg: Kellner Verlag, 1985), 32; V. Fuechtner, *Berlin Psychoanalytic: Psychoanalysis and Culture in Weimar Republic Germany and Beyond* (Berkeley/Los Angeles CA-London: University of California Press, 2011).

psychoanalytic sanatorium, on the idyllic grounds of *Schloß Tegel*, a short drive from Berlin's city-centre.[74]

Conclusion

Before the outbreak of the cataclysm that shook the foundations of old Europe, psychoanalysis was in a state of crisis. The person in whom Freud had put all his faith, both as the head of the international psychoanalytic movement and as his designated intellectual successor, had turned out to be a profligate perfidious pseudo-spiritualist. In order to get rid of him and all his devoted Swiss allies, Freud decided to produce a bomb and to establish a 'Secret Committee' of loyal followers in order to safeguard the interests of psychoanalysis. The bomb was definitely effective, and so the psychoanalytic community could be purified and reintegrated, yet Freud's trust in people had been seriously contaminated by the seeds of suspicion. What guarantees did he have that his newly adopted sons – Ferenczi, Rank, Abraham, Sachs and Jones – would not in turn go on to betray him? How could he prevent this small band of brothers from rallying against the father?

If, before the war, Freud's clinical practice was not in a state of crisis, it was most definitely in a state of conflict, as his technical papers bear ample witness to. Neurosis was not just the psychic ailment that had driven patients to his consultation room. It was not just a troubling disorder and a constellation of obsessional or hysterical symptoms that could be treated psychoanalytically. Neurosis was a fierce and forceful enemy, who was lurking in each and every corner of the clinical setting, including underneath the friendliest faces of transference, and its resistance had to be broken at all costs. Inviting as the groundrule of psychoanalysis may have been, many of Freud's patients did not want to free associate. However curative the process of remembering, many of them preferred to forget or,

[74] See, for example, U. Schultz & L. M. Hermanns, 'Ernst Simmels Sanatorium Schloß Tegel: Zur Geschichte und Konzeption der ersten psychoanalytische Klinik,' *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 1987, 87(37), 58-67; U. Schultz-Venrath, *Ernst Simmels psychoanalytische Klinik Sanatorium Schloss Tegel GmbH (1927-1931)* (Frankfurt am Main: Deutsche Hochschulschriften, 1995); E. A. Danto, '"Perfect in Its Principles": Psychoanalytic Praxis at Ernst Simmel's Schloss Tegel', *Psychoanalysis, Culture & Society*, 2009, 14(4), 337-349; H. Bernhardt, '"Stimmungszauber, das war der Leitgedanke bei der Gründung der Anstalt": Das Sanatorium Schloß Tegel in statu nascendi (1904-1907),' *Luzifer-Amor*, 2011, 24(47), 59-65.

what was substantially worse, to employ the playing field of the transference in order to enact their unconscious complexes. Although by contrast with his early years in practice, Freud did not have too much trouble persuading his patients to embark on a psychoanalytic journey, many of them categorically, i.e. neurotically refused to comply with the protocol. Some of them even refused to lie down.

All of this came to an end in the aftermath of the murder in Sarajevo. When Europe, and subsequently the whole world went to war, Freud's own private war was terminated, if only because the decimation of his patient stock, and the resulting financial hardship, not to mention his persistent worrying about the fate of his three sons, put his own neurosis in a state of suspended animation. With his adopted sons treating the wounded, fighting in the war, or editing newspapers, Freud involuntarily re-entered the state of splendid isolation in which he had made some of his seminal discoveries. Sheltered from the fire-storms in his own little trench, Freud speculated, turned his attention to the socio-political applications of psychoanalysis, speculated some more, synthesized his major accomplishments in a series of popular lectures, carried on speculating, and eventually noticed how the battlefields had generated neurotic men whose clinical symptoms could provide psychoanalysis with a new lease of life.

At the end of the war, and because of it, Freud was not unequivocally relieved, and even somewhat disappointed, partly no doubt because he did not know whether he would ever see his eldest son again, yet psychoanalysis was rejuvenated. Whereas the people of Europe were impoverished, disoriented, destitute and disconcerted about the milllions of dead, psychoanalysis had become richer, deeply infused with an invigorating tonic of abstract metapsychological principles, ready to take on the debilitating effects of psychological trauma, and sufficiently strong to be rolled out across the broadest layers of the social spectrum. I cannot possibly wish for anyone to have to live through the torment of war, but for psychoanalysis the great war brought new hope, new resolve, a new beginning.