PROBING AND MODELING VOLTAGE BREAKDOWN IN VACUUM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Gregory Richard Werner

August 2004

This document is in the public domain.

PROBING AND MODELING VOLTAGE BREAKDOWN IN VACUUM

Gregory Richard Werner, Ph.D.

Cornell University 2004

Voltage breakdown limits many technologies that rely on strong electric fields. Although many kinds of voltage breakdown have been well-explained, voltage breakdown in vacuum—the sudden transition from vacuum insulation to vacuum arc—remains relatively poorly understood. Despite the importance of vacuum insulation, technology has hardly improved breakdown voltages in the last ninety years. This work describes experiments in vacuum breakdown, as well as computer simulations of the initial stages of breakdown.

A better understanding of voltage breakdown could particularly benefit particle accelerators used for high energy physics experiments and radiation sources, which require the highest attainable electric fields in the microwave resonators that accelerate particles. Despite some differences, voltage breakdown in microwave resonators shares some features with breakdown in DC vacuum gaps (diodes). In both cases, the localized desorption of gas around an electron emission-source (*e.g.*, field emission) could lead to breakdown. Analytical calculation shows that breakdown occurs when the product of the gas density and emission current exceed a critical value.

Voltage breakdown *in vacuum* results from the interaction of the electric field and the electrodes. Using a scanning electron microscope, with energy dispersive x-ray spectroscopy (EDX) and Auger electron spectroscopy (AES) to identify

surface constituents, we found that breakdown occurs often at the site of foreign particles on the cathode, usually leaving only a very small trace of the original material. At the breakdown site we frequently found small craters, surrounded by a large starburst-shaped pattern; surface analysis suggests that during breakdown, ions bombard the surface within the starburst region and sputter away surface contaminants and oxides. In general, particulate contamination on the cathode determines the breakdown voltage, independent of the cathode material or the thickness of the insulating surface oxide; however, the oxide thickness does change the nature of the starburst and the damage done to the surface during breakdown.

BIOGRAPHICAL SKETCH

Greg Werner received the best the public education systems of Michigan and North Carolina had to offer, an educational opportunity probably unmatched anywhere in the world, and limited only by his own lack of industriousness and concentration. By good fortune he encountered only the best teachers, one grade after another, almost without exception.

During high school, when his family moved to North Carolina, Greg had to leave the instruction at the Kalamazoo Area Math and Science Center, but found himself a year later learning from another excellent group of teachers at the North Carolina School of Science and Mathematics, where an introduction to Lagrangian mechanics and the magical principle of least action incited his interest in physics, sentencing him to another eleven years of school.

After high school, Greg entered Amherst College. He spent the third of his four undergraduate years studying in Germany at the University of Göttingen, where he often passed through Hilbert-Space on the way to mathematics classes. Greg graduated in 1996 from Amherst College with a bachelors degree in mathematics and physics, and was subsequently hired by the Amherst College music department for a wonderful year as a teaching assistant, occasional accompanist, very occasional assistant orchestra conductor (for string sectionals only), and one-time emergency bass drummer. He did not forgo the opportunity to subject his students to a physical explanation of the development of western classical music, with the help of springs, strings, stroboscopes, speakers, and signal generators provided by the physics department.

After a year of music, Greg entered graduate school in the physics department of Cornell University, where he learned how much can be forgot in a single year.

After another year as a teaching assistant (but as a *physics* TA), Greg joined the superconducting RF group, a division of accelerator physics in the Laboratory for Elementary Particle Physics. For seven years (the traditional period of indentured servitude) Greg enjoyed Cornell and Ithaca as he studied and experimented and programmed and calculated and, finally, wrote. And if he has not moved on, he is there to this day.

ACKNOWLEDGEMENTS

First I would like to acknowledge and thank a number of people who have contributed directly to this project, without whom many experiments could not or would not have been performed. In particular, several undergraduates (some of whom graduated and continued as master's degree students) worked with me to complete experiments: Yan Liu, Myriam Qureshi, and Kate Rubin all spent a summer or a semester working on this project; Joe Betzwieser, Dan Lundberg, Jerry Shipman, and Laurel Ying worked with me for a year or more, contributing significantly to the course of experiments, especially in the preparation and microscopic examination of cathode samples. A new grad student in the group, Alexander Romanenko, performed the last experiment related to this project.

This work was supported with funding from the National Science Foundation.

The apparatus I used was created by Jim Welch and Dave Moffat. As experiments proceeded, modifications, new equipment, and new cathodes often required the expert machining provided by the Machine Shop Guys: Neal Alexander, Chuck Firenze, Henry Hansteen, Phil Hutchings, John Kaminski, Seth Marks, Randy Miller, Tom Reitz, Roger Seely, and Dave Simmons. Bob Kirby of SLAC, and Ian Wilson and Walter Wuensch of CERN provided diamond-machined copper cathodes for our experiments. Professor K. Zhao of Peking University kindly electropolished some of our (not diamond-machined) copper samples. After samples had been machined, Holly Conklin, Terri Gruber, and Neil Sherwood assisted in the cleaning and preparing of cathodes for testing.

Pat McKeown of Evans East Analytical Services performed Auger analysis on some cathode samples. Ivan Bazarov also performed some Auger analysis, and John Kaufman scanned some samples with an AFM.

David Bruhweiler of Tech-X Corporation implemented a field-emitter in OOPIC-PRO, the simulation code we used to explore the initiation of breakdown.

Ken Finkelstein offered needed advice and assisted in the building of a large permanent magnet to test the effect of magnetic fields on breakdown.

Discussions with Gil Toombes were helpful on many occasions, as well as enjoyable.

Hasan Padamsee was my chief advisor, offering frequent advice, support, and encouragement; I am grateful for his continued interest, patience, trust, and advocacy, and I am especially grateful for his willingness to let me explore on my own, and his willingness to help whenever I needed it.

I want to thank former and current members (and honorary members) of the superconducting RF group, with whom I have enjoyed working and from whom I have learned a great deal: Phil Barnes, Sergey Belomestnykh, Eric Chojnacki, Curtis Crawford, Gregorii Eremeev, Rongli Geng, Don Hartill, Tom Hays, Don Heath, Georg Hoffstaetter, Dan Kapner, Jens Knobloch, Yulin Li, Matthias Liepe, Ethan Miller, Dave Moffat, Peter Quigley, Rick Roy, James Sears, Valery Shemelin, Charlie Sinclair, and Maury Tigner.

Computers played important roles in many aspects of this work, and I am grateful that the many computers that I used (including four very different operating systems) were maintained by the computer group: Selden Ball, Bill Brangan, Devin Bougie, Ray Helmke, Ray Ng, Mike Roman, and Rich Sholtys.

I fortunately had the support of many people whose jobs involve the big, important projects at the Laboratory for Elementary Particle Physics, yet they cheerfully applied their expertise to support my project as well, from drafting and design to

requests for requisition to paperwork and printing: B. J. Bortz, Lois Billups, Dawn Eklund, Don Miller, Pam Morehouse, Tim O'Connell, Leonard Park, Peggy Steenrod, Al Venooker, Monica Wesley, and Beth Wilcox.

David Hammer offered crucial advice on measuring high voltages and high currents, and especially on RF shielding; also, I thank him, as well as David Rubin, for reading this document. And again I thank my advisor, Hasan Padamsee, for reading this document more than once.

Last, I thank my friends and family, without whom the last seven years would have been dull, if not impossible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction to Voltage Breakdown

Voltage breakdown is the derogatory term for a sudden transition from an insulating state to a highly conducting state due to an applied voltage. After voltage breakdown, the high conductance makes it impossible for any voltage source of reasonable power to maintain the applied voltage, and the voltage falls precipitously, hence the name "breakdown." Often the high conductance allows as much current to flow as can allowed by the available power source, and a great deal of power can be dissipated. Therefore, voltage breakdown is usually violent, often destructive, and sometimes catastrophic; for example, lightning is voltage breakdown of the atmosphere [104].

Any basic study of voltage breakdown seeks eventually to help limit or avoid completely the destruction caused by breakdown; ironically, another aim must be to cause and control breakdown more easily, often a goal just as desirable as avoiding breakdown—an arc-welder, for example, uses voltage breakdown to melt

metal for constructive purposes.[1] In many cases it turns out that the tendency to break down does not stem from the fundamental system, but from unwanted defects or contaminants, which are serendipitously eradicated by the breakdown events they cause; a breakdown in time may save nine (more likely, a hundred breakdowns now may prevent any breakdown in the future). Breakdown itself is nearly indispensable for preventing future breakdown.[2]

1.1 Applications

A better understanding of voltage breakdown, besides being scientifically interesting, will aid progress in many fields and technologies, which generally fall into two categories: those that require high electric fields, and those that require high electric currents. Breakdown usually hinders applications requiring high electric fields, while aiding applications requiring high currents.

Many technologies rely on high-current arcs, the result of voltage breakdown, to melt and vaporize metal. Arc-welding is an important example; arcs are also used to vaporize metal for deposition as thin metal films. The high conductance and rapid response time of an arc can be harnessed for fast, high-power switches; in such capacity, arcs may also be used as over-voltage protection. Perhaps the most widespread use of arcs occurs in gasoline combustion engines, where an electric spark triggers the combustion.

[1] Nevertheless, we do not call it a breakdown welder; "breakdown" is undesirable.

[2] When desirable as a preventive measure, breakdown is not called breakdown, but "conditioning" or "processing." An electrode is "conditioned" to avoid "breakdown."

On the other hand, unwanted voltage breakdown limits many technologies involving high electric fields. Electric fields are enormously useful, and usually higher electric fields would be even more useful, if it weren't for voltage breakdown; some devices for which breakdown can be a problem are: semiconductor devices (like MOSFETs), electron guns (field-emission sources make great electron guns, but require very high fields), high voltage insulation, microwave sources (klystrons), *etc.* Here at the Cornell Laboratory for Elementary Particle Physics we naturally worry about the problems of voltage breakdown for particle accelerators. Particle accelerators continue to advance basic scientific understanding in a number of fields, from the properties of quarks learned from colliding electrons and positrons to protein structures illuminated by the x-rays emitted from accelerating charged particles. Particle accelerators find direct application in hospitals (zapping cancer cells with protons) and may one day drive sub-critical nuclear reactions that reduce or eliminate radioactive waste.

RF (radio-frequency) resonant cavities

Within particle accelerators there are many different places where large electric fields are required; in particular, electric fields are of the most basic importance in accelerating charged particles. The simplest way to accelerate a charged particle is to put it into a capacitor and apply an electric field. For various reasons, including voltage breakdown, such electrostatic particle accelerators cannot be used to accelerate particles to very high energies, and most accelerators use RF electric fields for acceleration. The RF fields are confined in metal boxes called RF cavities (or microwave resonators); an RF cavity is an LCR circuit bundled into a single component—the electric and magnetic fields in the cavity oscillate at

the cavity's resonant frequency (figure 1.1), generally in the GHz range. Precise timing ensures that particles enter and exit the cavity within the half-period that the electric field points in the right direction (otherwise the particles would be decelerated rather than accelerated).

Superconducting RF cavities

The quality factor (Q) of a resonator reflects the dissipation of the energy stored in the resonator during oscillation; the Q factor measures what percent of the stored energy dissipates during one oscillation. For example, 99% of stored energy is dissipated within $0.73Q$ oscillations—within 73 oscillations if $Q = 100$, within 7.3 billion oscillations if $Q = 10^{10}$; higher Q factors correspond to lower dissipation. Usually we write Q in terms of the fraction of energy lost during one radian of oscillation: $Q = \omega U/P_{\text{dis}}$, where ω is the resonant (angular) frequency, U the stored energy, and P_{dis} the dissipated power (averaged over one cycle). Power dissipation occurs because current flows back and forth through the cavity walls as the fields oscillate, and the cavity walls have finite resistance.

Some accelerators, like the one here at Cornell, use superconducting cavities[3] to reduce the power dissipated in the cavity walls, creating higher fields with less input power. When a power source (a microwave source) drives a cavity at its resonant frequency, the cavity reaches an equilibrium (constant stored energy) when the input power P_{in} equals the dissipated power P_{dis}, at which point the equilibrium stored energy is $U_{\text{eq}} = QP_{\text{in}}/\omega$. A higher stored energy means a higher maximum electric field; a higher Q factor therefore allows higher electric fields with less input power—cavities with less power dissipation require less input power. While the Q

[3]For a thorough discussion of superconducting RF cavities, see [86].

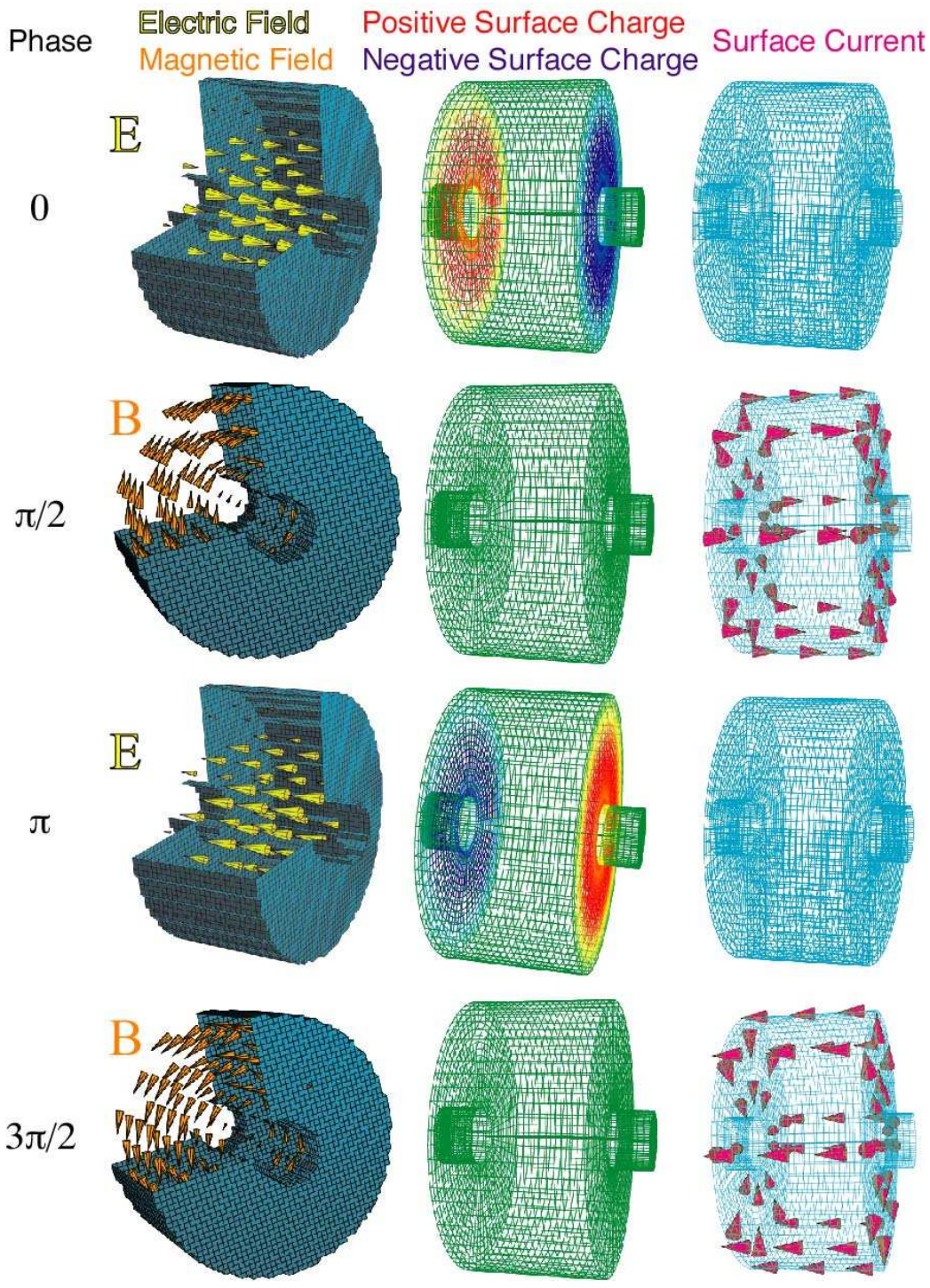


Figure 1.1: The electric and magnetic fields in an RF cavity at different times within one oscillation period. The cavity acts both like a capacitor and an inductor; and since current flows through the cavity walls, the cavity walls dissipate power as a resistor, completing the LCR circuit analogy.

factor of a normal conducting (copper) cavity might be 10^4, the Q factor of a superconducting cavity might be 10^9, or even 10^{10}. Besides greater efficiency, the high Q factor of superconducting cavities offers other important advantages to particle accelerators (for example, superconducting cavities can have larger beam-tubes than normal conducting cavities [86]); the disadvantage of superconducting cavities is that they must operate at very low temperatures to maintain superconductivity.

Currently the best superconductor for RF cavities is niobium—niobium has the highest superconducting transition temperature of any pure element superconductor, about 9.2 K. Normal conducting cavities, on the other hand, are made of copper, because of its good conductivity (for a normal conductor).

Breakdown in RF cavities

Accelerators want to operate their RF cavities at the highest fields possible to get the maximum acceleration in the shortest distance; however, accelerators also want to run without interrupting malfunctions, and voltage breakdown is a serious malfunction that occurs more frequently at higher fields. A cavity cannot function while it suffers voltage breakdown, and an accelerator cannot function without its accelerating cavities.

Because accelerators must operate under vacuum to avoid interference with the beam of particles, accelerator cavities also operate under vacuum, which turns out to be an advantage because vacuum resists breakdown much better than gas[4]—the vacuum contains very few of the charged particles necessary to conduct current

[4]Gas at very high densities can be very resistant to breakdown; but vacuum can usually withstand higher fields than any gas at atmospheric pressure or below. Vacuum resists breakdown because it lacks charged particles to conduct current, whereas high density gas resists breakdown by obstructing the motion of charged particles (like a solid insulator).

(because it contains very few particles at all). However, the cavity walls offer an unlimited quantity of charged particles, if enough energy can be diverted to free them from their bonds. Breakdown occurs in an RF cavity when fields interacting with the cavity surfaces introduce charged particles into the cavity vacuum. The cavity electric field accelerates those particles, dissipating its stored energy; worse, that dissipated energy usually ends up making more charged particles, escalating the problem. Within nanoseconds, a large parasitic current of unwanted electrons and ions can develop in the cavity (an arc), consuming all the stored energy, and damaging the walls.

Although voltage breakdown in either superconducting cavities or normal conducting cavities hampers accelerator performance, superconducting cavities tend to be more delicate than normal conducting cavities, and other limitations may appear before voltage breakdown becomes a problem. To exploit the advantage of superconductivity, cavities must operate dependably with a very high Q factor, which means that even small sources of power dissipation can be a problem, whereas in a normal conducting cavity such small dissipation sources would go unnoticed, overwhelmed by the power dissipation in the normal conducting cavity walls. For instance, stray magnetic field (from the earth or nearby equipment) can create flux vortices in a superconducting cavity, lowering the Q factor. Similarly, even a small amount of parasitic electron current (from field emission, perhaps) can impair the operation of a superconducting cavity by lowering the Q factor from 10^{10} to 10^{8}, whereas a normal conducting cavity with a Q of 10^{4} could operate successfully despite a hundred times more parasitic current. In the best superconducting cavities, the magnetic field limits performance more than the electric field;

however, in operating accelerator cavities, electron emission (or "dark current") brought about by high electric fields still presents a major practical problem.

A high electric field in an RF cavity can cause electrons to tunnel through the work-function barrier, out of the metal wall into the vacuum (field emission). Field emission currents in a cavity increase power dissipation, lowering the Q factor. At some level of current, much lower for superconducting cavities than normal conducting cavities, too much power will be dissipated and the cavity operation will suffer. In has been found that in superconducting cavities, sufficient cleaning (and subsequent assembly in clean rooms) prevents field emission from becoming a problem. However, while installed in an accelerator, cavities can become contaminated and develop excessive field emission that degrades the Q factor.

Earlier I mentioned that breakdown often destroys its trigger, preventing future breakdown. It turns out that breakdown often destroys field emitters (incidentally suggesting that field emission is probably related to the triggering of breakdown). With that consequence in mind, voltage breakdown can be an effective procedure for reducing field emission; when voltage breakdown is used for this purpose, we call it "processing." Processing superconducting cavities is not as easy as normal conducting cavities, because superconducting cavities require much less power (that's the point of using a superconductor) and therefore, when a field emitter causes excess dissipation, less power is available to create the high fields necessary to induce breakdown [40]. Ideally, we would like to be able to process cavities at the lowest fields possible; the quest for efficient processing methods currently provides the biggest challenge related to voltage breakdown in superconducting cavities.

With normal conducting cavities, processing is also very important; cavities are processed at high powers and allowed to break down again repeatedly (for tens of hours and thousands to millions of breakdown events) until they reach sufficiently high fields with tolerably low breakdown rates. However, extensive breakdown damages the surface, which can lead to shifts in the resonant frequency, causing problems for the accelerator. Voltage breakdown is currently one of the major limits for normal conducting cavities; they could reach higher fields if breakdown could be avoided. Achieving higher fields in normal conducting cavities will require more effective and less damaging processing methods, which limit the damage to the cavity surface while still destroying the triggers of breakdown.

1.2 Types of Voltage Breakdown

In this work I will discuss only voltage breakdown in vacuum; first, however, it's worth reviewing briefly some of the different kinds of breakdown that may be relevant, if only tangentially. Transforming an insulator into a conductor requires the introduction of free or mobile charges to carry electric current; the medium in which this occurs provides the coarsest classification for breakdown:

1. Dielectric breakdown: breakdown in solid insulators, usually destroying part of the insulator (making holes in it); *e.g.*, breakdown of the insulating oxide layer in a CMOS device, or the breakdown of high voltage ceramic insulators.

2. Breakdown of gases [73]: breakdown in gases, usually involving electron avalanches caused by electron-impact ionization of the ambient neutral gas:[5]

[5]The difference between "gas" and "vacuum" depends on the mean free path of an electron; gas densities low enough that electrons rarely collide with gas molecules should be considered vacuum for this purpose. On the other hand, if

Townsend discharges: electron avalanches (streamers) started by chance ionization events, and regenerated by secondary processes (*e.g.*, photon or ion bombardment of the cathode that releases new "secondary" electrons at the cathode). Townsend discharges (by definition or classification; *c.f.*, Glow discharges) do not produce enough current or charge to alter significantly the applied electric field.

Glow discharges: basically Townsend discharges (electron avalanches and secondary emission) that produce enough charge to alter the applied field (but not enough current to dominate the applied voltage[6]); typical glow discharges require hundreds of volts and conduct microamps to milliamps [72], *e.g.*, neon and fluorescent lighting, corona discharges.

Arcs: electron avalanches that create enough current and charge to support more efficient mechanisms of charge production, requiring only 10–20 volts, but amps to kiloamps of current; the arc almost completely determines the field (including the applied voltage), *e.g.*, lightning, static electric shocks, welding arcs, automotive spark (as in "spark plugs").

3. Vacuum breakdown: arcing in vacuum, caused by interaction of the electric field and the vacuum walls (or electrodes in the vacuum) perhaps stimulated by the energy gain of charged particles that travel unhindered in the electric field; vacuum arcs differ from gaseous arcs mainly in that vacuum arcs must

breakdown depends upon frequent electron-gas collisions, then we call it breakdown in gas. For typical cases of interest, "vacuum" pressure is lower than 10^{-5} torr, and "gas" pressure higher than 10^{-4} torr.

[6]In a glow discharge, the current can still be increased or decreased by increasing or decreasing the voltage; in an arc, the voltage remains nearly constant while the current varies over several orders of magnitude.

produce vapor before ionizing it, making vacuum arcs more difficult to start and maintain.

Because RF accelerator cavities must be evacuated, this work concentrates on vacuum breakdown. Vacuum arcs resemble gaseous arcs in many ways—the driving processes of electron emission and ionization may be very similar—but the dependence of a vacuum arc on vaporized electrode material does make a difference, especially in the triggering of the arc and the spatial and temporal evolution of the arc. Dielectric breakdown may also be relevant to vacuum breakdown for electrodes with insulating oxides (or applied dielectric coatings); breakdown in vacuum may be triggered by electrical breakdown of the oxide in some circumstances.

To reiterate: voltage breakdown in vacuum is the sudden transition from vacuum, with its excellent insulating qualities, to a "vacuum arc," a plasma (free electrons and ions) that can conduct high currents with voltages between 10 and 20 volts. Vacuum breakdown can be all the more shocking because vacuum is such a good insulator—with no free charged particles, there can be no electrical current, and with very few particles at all in a vacuum, the introduction of significant numbers of free charged particles poses great difficulties.

With a scarcity of particles between electrodes, vacuum breakdown depends on particles freed from the electrodes. Because vacuum breakdown depends on the interaction between the field and the electrode surfaces, rather than the field and the medium, more complications arise than for breakdown in gas—surface physics presents greater difficulties than the physics of gases (since real gases closely resemble ideal gases). For instance, breakdown voltages in different gases for different electrode configurations are known precisely and even used to measure voltages.[7]

[7]By bringing two spherical electrodes nearer together until a spark occurs, one can determine the voltage from the distance between them; for instance, if 5 cm

For a given configuration in vacuum, one cannot specify "the" breakdown voltage, because it depends too sensitively on surface conditions that are difficult to measure. For instance, vacuum breakdown may be triggered by electrode contaminants and defects, and breakdown voltages seem to decrease with increasing electrode area.

A further difficulty in understanding vacuum breakdown is the likely existence of different triggering mechanisms that lead to very similar sorts of results. In some experiments, especially those with small inter-electrode distances and high electric fields applied in very short pulses, breakdown seems to be dominated by the interaction of the cathode with the electric field; but some experiments with continuously applied high voltages and large inter-electrode distances suggest that the anode may play at important role. Some experiments see a transition between different mechanisms as pulse width and voltage are varied.

1.3 This Work: Studying RF Breakdown with DC Experiments

We wanted to see what we could learn about voltage breakdown in RF cavities from vacuum breakdown in DC (direct current, or continuously applied field) experiments. In principle, they could be completely different; however, any similarities between them could be studied much more efficiently with DC experiments. Testing RF cavities, especially superconducting RF cavities, requires comparatively complicated equipment and infrastructure; moreover, RF cavities are black boxes

diameter spheres spark when brought within 0.44 cm of each other, the voltage between them is 15 kV (in air at 760 torr, 25°C) [106].

(and superconducting cavities are black boxes submerged in liquid helium in big, steel cryostats) that obscure what's going on inside them. DC experiments allow measurements, such as arc current, to be made more easily (or at all), and the relative simplicity of the experiments allows more tests. For this work, I created more than two hundred DC breakdown events; so many tests would have been impossible with RF cavities.

First, and perhaps most important, we examined similarities between RF and DC breakdown. On niobium cathodes, DC breakdown left craters and starburst-shaped patterns on the cathode very closely resembling niobium cavity surfaces that had suffered voltage breakdown (figure 1.2). This similarity, perhaps along with the detection of electron current in DC experiments and evidence that starbursts occurred at electron emission sites in cavities, justified a series of DC experiments to elucidate RF breakdown. The existence of starbursts after DC breakdown confirmed, incidentally, that the starburst phenomenon is connected with the electric field, and not something special about RF fields, low temperatures, or superconductivity.

To strengthen the connection between DC and RF experiments, we tried to set the stage for DC breakdown in a way that was strongly suspected to cause RF breakdown in cavities—we introduced contaminant particles on the cathode. A major breakthrough in superconducting RF technology was the realization that particle-free cavities showed less electron emission and less frequent voltage breakdown; as a result all superconducting cavities are now thoroughly cleaned and assembled in particle-free clean rooms [85]. Therefore, we seeded cathodes with a small number of small particles to make the DC experiments resemble RF experiments as closely as possible.

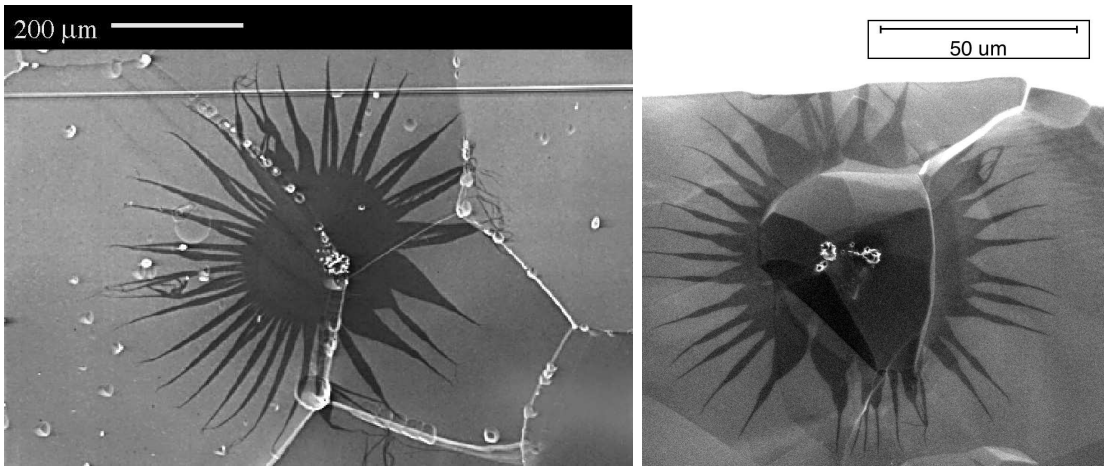


Figure 1.2: A field emission site after breakdown in a superconducting RF cavity at 2 K (left), reproduced from [55]; and a niobium cathode after breakdown in a room temperature DC voltage gap (right).

Some painstaking and clever RF experiments had shown that voltage breakdown occurred at electron emission sites [57, 55]; DC studies had often shown particles at electron emission sites. This was evidence suggesting a connection between particles and breakdown, and the motivation for clean cavities. This work has further strengthened this connection, showing conclusively that breakdown occurs (often) at particle sites, and further demonstrating that minuscule contaminants found at RF breakdown sites are consistent with the previous existence of a particle at that site. Thus DC experiments led more directly to stronger conclusions about voltage breakdown in RF cavities than could have been reached with RF tests in a comparable time.

The small samples allowed by DC experiments provided one advantage among many; the small size allowed much easier examination of the samples, both because a smaller area is easier to cover, and because many analysis tools, like scanning Auger microprobes, accept only small samples. (With Auger analysis, we were able to explain the appearance of starbursts, as seen in an electron microscope.)

To show why RF and DC breakdown might be similar, we resorted to simulating the initial stages of breakdown, based on the hypothesis that an initial release of neutral vapor from the cathode might initiate breakdown. In his dissertation [56], Jens Knobloch had done some initial simulations of RF breakdown, and we continued with a different simulation program that we thought had more potential; indeed, we confirmed his results for the extremely early stages of breakdown, but we were able to run simulations further in time. We also performed analogous DC simulations; the results (which depend on the assumption of an initial release of vapor) clearly show why RF and DC breakdown might be very similar, despite the disparity in initial conditions. At some point the similarity between DC and RF breakdown ends; however, computer simulations may still provide a strong connection. It will take DC experiments, with the many diagnostics that would be impossible in an RF cavity, to validate computer simulations; if we can achieve agreement between DC experiment and simulation, we may pursue RF simulations with more confidence.

1.4 Breakdown in the Literature

Voltage breakdown in RF cavities is discussed in [89] (see especially the article by Padamsee and Knobloch for breakdown in superconducting cavities, which includes results from Knobloch's simulations of RF breakdown; see also [84]). The first picture of a DC starburst (resembling RF breakdown sites) was published in [77], which also presents results connecting electron emission sites in superconducting cavities with breakdown sites. A good source of information on RF breakdown relevant to this work can be found in Knobloch's Ph.D. dissertation [53], as well as

a more complete description of the simulations that led to the simulations discussed in this work.

This work, however, concentrates mostly on DC (continuous and pulsed) vacuum breakdown, though keeping in mind the connections to RF breakdown. The scientific literature dealing with DC breakdown is enormous. Work by Millikan in the early twentieth century showed that breakdown voltage has no dependence on the ambient pressure once a high vacuum is achieved [76]. In 1918 typical vacuum breakdown occurred at fields of 30–120 MV/m, which have hardly improved in the intervening decades.

Before voltage breakdown, small electron currents are often seen; these currents may play an important role in initiating vacuum arcs. In 1928 Fowler and Nordheim published a calculation of the tunneling probability through a triangular barrier and applied it to the emission of electrons from a metal under the influence of a strong electric field, now referred to as Fowler-Nordheim field emission [37]. The work function of a metal describes the height of the roughly square potential well that keeps electrons in the metal. Applying an electric field to the metal makes the top of the well slant downwards so that an electron in the well can tunnel out of the well if the slant is steep enough. The basic Fowler-Nordheim model persists to this day, albeit with various modifications and enhancements (for instance, [78]). The experimental study of field emission greatly benefited from UHV (ultra-high vacuum) techniques. In UHV systems, Fowler-Nordheim field emission has received experimental confirmation for sharply-pointed (single crystal points) electrodes ([36] is a more recent experiment). In calculating the tunneling rate, one assumes nearly-free electrons near the Fermi level, and a sharp transition to vacuum. In reality, larger-area electrodes present less-than-ideal surfaces—polycrystalline,

rough, with oxides and adsorbates—that alter the electronic states. The field emission mechanism(s) on electrodes with macroscopic areas is(are) still not completely understood.

A good review of field emission from broad area electrodes was written by Noer [83]. The annoying reality (from many perspectives) is that field emission occurs at much lower fields—as much as 100 to 1000 times lower!—than predicted by Fowler-Nordheim. Moreover, electrons appear to be emitted from a few tiny points on the cathode (see, for example, [16]), not from the whole cathode area. Initial explanations involved microprotrusions on the surface—tiny but pointy metal whiskers that enhance the electric field by a factor β (which could be between 10 and 1000). Some experiments actually found protrusions [67, 68, 66, 102, 62]. However, there are a number of problems with the protrusion model (see [83]); mainly, when one looks carefully at a field emission site,[8] one rarely sees a protrusion that could cause the large field enhancement necessary for Fowler-Nordheim emission—it's not easy to enhance the field by a factor of 500. However, protruding shapes (or at least protruding conducting shapes) definitely enhance the electric field and field emission; indeed, I would argue that geometrically enhanced field emission is the only *certain* way of enhancing field emission. Other theories of field emission have been developed [107], but so far they have two barriers to overcome: (1) they are not easy to confirm experimentally, because they are sensitive to complicated surface states (like the electronic states in a metallic oxide), and (2) they are not as simple as the Fowler-Nordheim model with geometric field enhancement.

[8] Field emission sites can be detected with a small scanning anode that exposes only a small region of the cathode to high electric fields; the site can then be viewed with an electron microscope [81].

In pursuing the similarities between DC breakdown and breakdown in superconducting cavities, it's important to note that superconductivity seems to have no effect on field emission; field emission from niobium is the same whether at 300 K or below the superconducting transition temperature at 4.2 K [93].

When the microprotrusion model became popular, it appeared to offer an attractive explanation for vacuum breakdown. Field emission indicated the presence of long, thin, pointy whiskers on the cathode; although field emission currents might be low (nanoamps or microamps) the current density was predicted to be quite high. High current densities would heat the whisker, the thinness of which would impede thermal conduction to the bulk cathode, allowing the tip of the emitter to reach very high temperatures. If the tip vaporized, providing a source of neutral gas, the field emission would ionize the gas; ions would in turn bombard the cathode, creating more heating, more vaporization, and more current—in short, an arc would develop, and breakdown would ensue.

The vaporizing field emitter model is so attractive, and explains so much, that I sometimes find it difficult to pay attention to its shortcomings. The weakest points in this explanation are: (1) field emission often does not emanate from extremely tall and thin protrusions, which are necessary to explain a large temperature rise, and (2) even assuming a long and thin emitter, the mechanism by which vapor is suddenly produced has not been explained. The first point relates to the difficulties of explaining field emission. The second is more subtle. The Joule heating from field emission for a very long and thin emitter (with a field enhancement of a 100 or more) can lead to large temperature rises, and high temperatures can lead to vaporization. However, the vapor pressure of a metal varies continuously, even through its melting point. Since field emission can operate stably, we should be

able to increase the field and watch the vacuum pressure rise until some critical point when breakdown occurs; this is not the case—there seems to be no middle-ground between field emission with no measurable vapor release and breakdown. Therefore, some positive feedback mechanism must ensure that as soon as any significant vapor is released, breakdown quickly follows. It has been hard to come up with such a mechanism partly because breakdown does not seem to depend on thermal properties of the bulk cathode: for instance, at temperatures up to 2500°C, niobium has a vapor pressure six orders of magnitude lower than that of copper (with more than nine orders of magnitude difference up to 1500°C), and the melting point of niobium is more than 1000°C higher than that of copper, yet there seems to be no significant difference in breakdown voltages on niobium and copper cathodes.

Nevertheless, there is some very convincing evidence for the vaporizing microprotrusion model. For extremely sharp cathodes (not broad-area), where field emission is reasonably Fowler-Nordheim, this model appears to be a good explanation [34, 35]. Charbonnier *et. al.* [20] have worked out the critical field enhancement β that (in their model) separates the cathode-initiated breakdown described above and anode-initiated breakdown. More needle-like emitters have higher β and higher current densities J; Joule heating at the emitter should be proportional to J^2. On the other hand, heating at the anode will be proportional to the total current times the voltage, so higher field enhancements lead to higher power densities at the cathode than at the anode, while lower field enhancements lead to higher power densities at the anode than at the cathode.

Kranjec and Ruby [58] measured field emission before breakdown on thirteen different electrode materials and found that for each material the local electric field

at breakdown was constant (the local electric field was found by multiplying the macroscopic field by the enhancement factor found by fitting the emission current to the Fowler-Nordheim formula)—for materials from indium to tungsten, the local breakdown fields were between 5 and 11 GV/m (while the local breakdown field on tungsten, for example, was 6.9 ± 1.0 GV/m). Of course, the critical local breakdown field corresponds (via Fowler-Nordheim) to a critical current density. The consistency of the results from that experiment is remarkable considering that few other breakdown experiments have achieved such consistency.[9] It may be important to note that the experiments were carried out on mechanically polished, heavily conditioned electrodes.

Alpert *et. al.*, in [1] (an article worth reading thoroughly), showed similar experimental data supporting the "critical local field" (hence critical emitter current density) theory of breakdown. They also point out that in general, larger gaps lead to higher field enhancements, in theory and experiment. Decreasing the gap to the order of size of the microprotrusion will certainly decrease the field enhancement at the protrusion.[10] Thus they explain the generally observed diminishing of breakdown field at larger gaps—larger gaps have higher field enhancements, hence they break down at lower macroscopic fields, though at the same local field at the emitter (sometimes called the "voltage effect" since larger gaps need higher voltages to reach the same field). Although Alpert *et. al.* believed they could explain

[9]For example, Kranjec and Ruby [58] find the local breakdown field for aluminum to be between 9.8 and 11.2 GV/m for 4 emitters with values of β from 72 to 309; while Bennette *et. al.* [7] (calculated from Table II and the value of F_s for aluminum from [20]) find the local breakdown field ranging from 5.8 to 8.9 GV/m for 9 emitters with β from 24 to 360.

[10]For example, a protrusion that's 1 micron high and 10 nm in diameter might have an enhancement factor $\beta \sim 100$ in a 100 micron gap; to take an extreme case, in a 1.010 micron gap the protrusion diameter is the same as the space between the protrusion and the anode, and we'd expect $\beta \sim 1$.

the change in field enhancement with gap (which is observed experimentally), I'm amazed that β can change, as they measured, so much with the gap spacing:

gap (μm)	β
51	59
1020	118
4060	211

.

Even a ten micron tall emitter (quite extraordinary) should hardly affect the field 1000 microns away, making it hard to explain the doubling in field enhancement as the anode moves from 1000 to 4000 microns.

Because the production of vapor is so important for vacuum breakdown, a few experiments detecting vaporized electrode material should be mentioned. One experiment [26] (see also [25]) actually detected electrode vapor microseconds before breakdown—unlike other experiments I'll mention, this experiment measured the vapor by its absorption of light, which allowed detection before breakdown; the vapor density appears to be greater near the cathode than the anode. The importance of this experiment is not only that vapor is detected just before breakdown, but also that the vapor is the same material as the bulk electrode (unfortunately both cathode and anode were the same material). Other experiments have measured light emission from cathode and anode material (tuning the detectors to atomic transitions of the cathode and/or anode material) [28, 108], which really measures vapor only after the onset of the arc (not before the arc, like the absorption experiment). [108] used pulsed voltages to show that for breakdown events that begin within microseconds after the pulse rise, cathode vapor emits light first, but when breakdown occurs after a longer period, anode vapor first emits light (see also [27, 79] for relevant discussion of the mechanism involved).

Several plausible mechanisms for breakdown have been proposed, none of them universally acceptable, but more than one may be applicable in different situations (for instance, breakdown at large gaps and high voltages may proceed according to a different mechanism than breakdown at small gaps and low voltages). The validity of each model hinges on its explanation of how to get vapor and then charged particles into the gap (and the vapor is the hard part, since field emission can make ions from vapor). Various models involve small clumps of material breaking off the cathode or anode due to electric field pressure and heating, abrupt and violent explosions at the field emitter, *etc.* The amount of work that has been done on breakdown is enormous and I will be content to cite some major reviews. A good starting point is [24] (brief and recent enough to include the major work of the 1960s, when UHV technology facilitated advancement in the field; field emission and breakdown at sharp points was understood and the results applied to broad area electrodes). However, the earlier and more explanatory [66] gives a better idea of the paths of progress taken. The reference for breakdown in gases [73] is useful because it describes arc phenomena once gas is available (like collision processes, including ionization), and it also has a chapter on vacuum breakdown. To reach the present, I recommend the book by Mesyats [75] and the compendium [13], which covers cathode processes of vacuum arcs in great detail, as well as discussing at length technologies that require vacuum arcs.

The latter references discuss many pulsed voltage experiments; while not the same as the DC experiments that we consider, they are a wealth of information. Knowing exactly when breakdown will happen (*i.e.*, during the pulse) allows very fast measurements to be made, and also yields measurements of times important

to breakdown (of particular interest, for example, is the time from pulse rise to breakdown).

Both [75] and [13] (as well as [73]) start to discuss more about the arc than the trigger of breakdown. As mentioned, the main difficulty in creating and maintaining a vacuum arc, distinguishing it from a gaseous arc, is the creation of neutral vapor by arc processes. Presumably to satisfy the vapor requirement, vacuum arcs are driven by an engine that appears to be localized at the cathode, and is therefore called the "cathode spot." A single vacuum arc may encompass many cathode spots at one time, as well as at different times, with the life-cycle of a cathode spot being shorter than that of the entire arc. A short review by Hantzsche [42] gives a good introduction to the cathode spot, although [13] and [75] discuss it much more extensively. Schwirzke's work on the unipolar arc has been influential in considering starburst formation [95, 94].

Many helpful experimental measurements of the vacuum arc plasma and cathode spots have been made, albeit in arcs driven by high current sources, lasting longer than in our experiment (so the electrodes are always heavily processed). The velocities of ions emerging from the arc can be much higher than the arc voltage [29, 59, 18, 52, 92, 103, 111, 110, 2].[11] Measurements of various cathode spot parameters are summarized in [4], and a measurement of plasma density at the cathode spot (using laser absorption) can be found in [6]. Also interesting are cathode spots on liquid metal cathodes [5].

The simulations of breakdown presented in this work are continuations of work shown in [56], though with a new plasma simulation program. A good introduction

[11] I recommend [29] because I think it's a good article, and it's the only one that measures neutrals' energies as well as ions', but beware that later work questions its reference voltage and therefore the accuracy of its results.

to plasma simulation is [8], and a more extensive discussion of relevant algorithms can be found in [46].

Chapter 2

An Amalgamated DC Breakdown Experiment

Over the course of the last several years I conducted many DC breakdown experiments, the gathered and collated results of which appear in later chapters. The results stand alone to some extent,[1] but familiarity with a single, typical experiment may help to understood the collections of results. During the course of investigation, interesting questions abounded, plans evolved, and new methods and equipment came into being, so that hardly any set of ten experiments shared exactly the same procedures or results. This chapter outlines a composite experiment, made with parts of real experiments, designed to explain typical procedures and their direct results.

[1] That is to say: you may skip this chapter, and refer to it later as needed.

2.1 Cathode Sample Preparation

Each experiment involved a cathode and an anode; the same anode remained installed for many consecutive experiments, while the cathodes, the focus of most experiments, usually underwent only one experiment. Each cathode had 4 or 5 raised sub-cathodes, called pedestals, which would be tested individually by positioning the pedestal-sized anode across from each of them in turn. Most cathodes were niobium, reflecting our interest in superconducting niobium RF cavities; however, experiments included several copper cathodes (the most common material for normal conducting RF cavities) and two gold cathodes (gold film on niobium); in a few cases, thick oxides were grown on both niobium and copper cathodes.

2.1.1 Machining

The cathodes were thin, one inch diameter discs, 1/16 of an inch thick, with three mounting holes around the edge for easy mounting; a recess was milled out, leaving raised pedestals (figure 2.1).

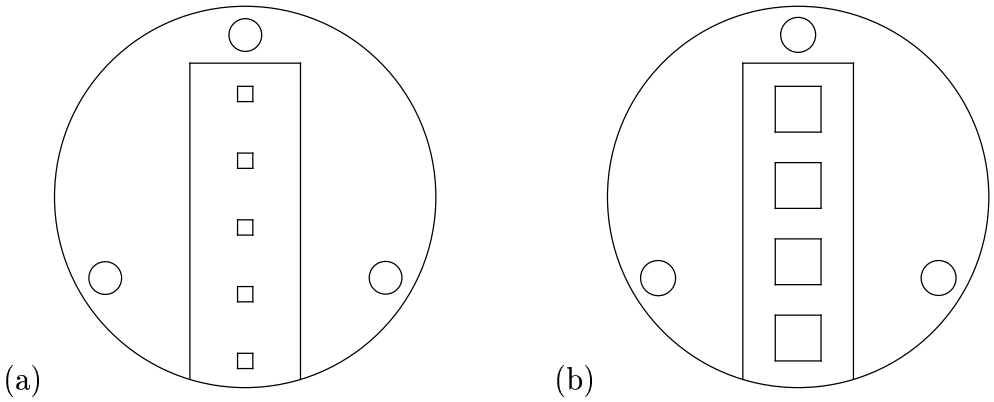


Figure 2.1: (a) The five-pedestal cathode plate, with one-by-one millimeter pedestals. (b) The four-pedestal cathode plate, with three-by-three millimeter pedestals.

2.1.2 Cleaning

We wanted our cathodes to resemble as much as possible the pristine surfaces of a clean superconducting RF cavity. Following machining:

- washing with a soft cloth and detergent removed grease;
- etching in 1:1:2 solution of HF:HNO_3:H_3PO_4 stripped 100–150 microns (4–6 mils) from the surface,[2] leaving pure, clean niobium, and rounding sharp edges;
- rinsing in ultra-pure de-ionized water;
- drying in a class 100 clean room.[3]

After cleaning, cathodes were never handled without clean gloves, and the pedestals were never touched at all.

Because the acid etches away different crystal faces of niobium at different rates, individual grains can be easily distinguished on an etched niobium landscape (figure 2.2). The difference in etch rates creates a surface that is quite rough on a large scale—the difference in grain heights can be ten microns—while the face of a single grain is almost atomically smooth. The smoothness is ideal for the surface of a cathode, while the large-scale roughness provides helpful landmarks for comparing pictures from before and after voltage breakdown.

Sometimes an already-etched cathode needed to be cleaned again, *e.g.*, after an unsatisfactory contamination. Often re-cleaning began anew with re-etching,

[2]Nitric acid oxidizes niobium and hydrofluoric acid dissolves niobium oxide; phosphoric acid buffers the solution so the etching rate (and heat production) does not become excessive.

[3]The "class" of a clean room as used here is the number of particles (larger than 0.5 microns) per cubic foot.

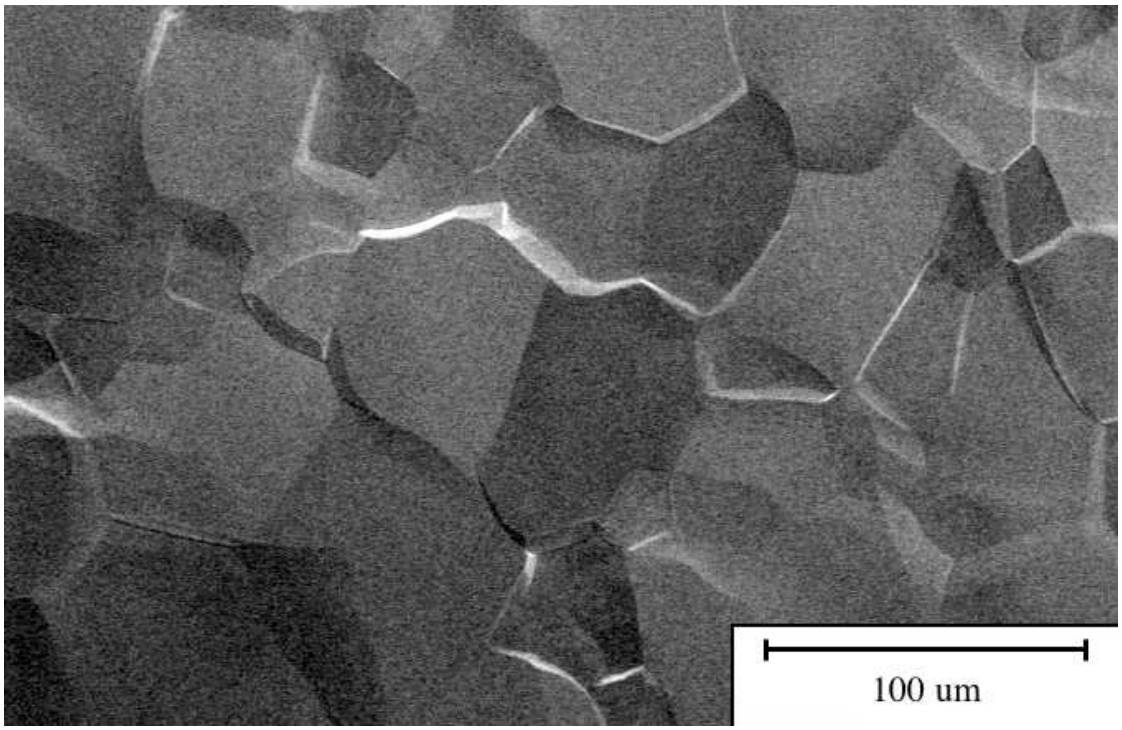


Figure 2.2: A heavily-etched niobium surface (as seen by a scanning electron microscope). Etching removes material from different crystal facets at different rates, prominently distinguishing the different crystalline grains.

but we also developed less drastic techniques including high-pressure water rinsing, CO_2 snow cleaning,[4] and in some cases very gently wiping a single pedestal with a lint-free cloth soaked in methanol (a delicate but effective technique when all but one pedestal were satisfactory).

2.1.3 Contaminating

Having prepared a clean cathode, we contaminated it with particles of our choice; fine vanadium powder was a favorite, but other contaminants included carbon,

[4]The method of CO_2 snow cleaning blasts the surface with a jet of frozen carbon dioxide particles. The mechanical and thermal stress caused by the dry-icy bombardment knocks off particles, and liquid CO_2, an excellent solvent, removes hydrocarbons [97, 98].

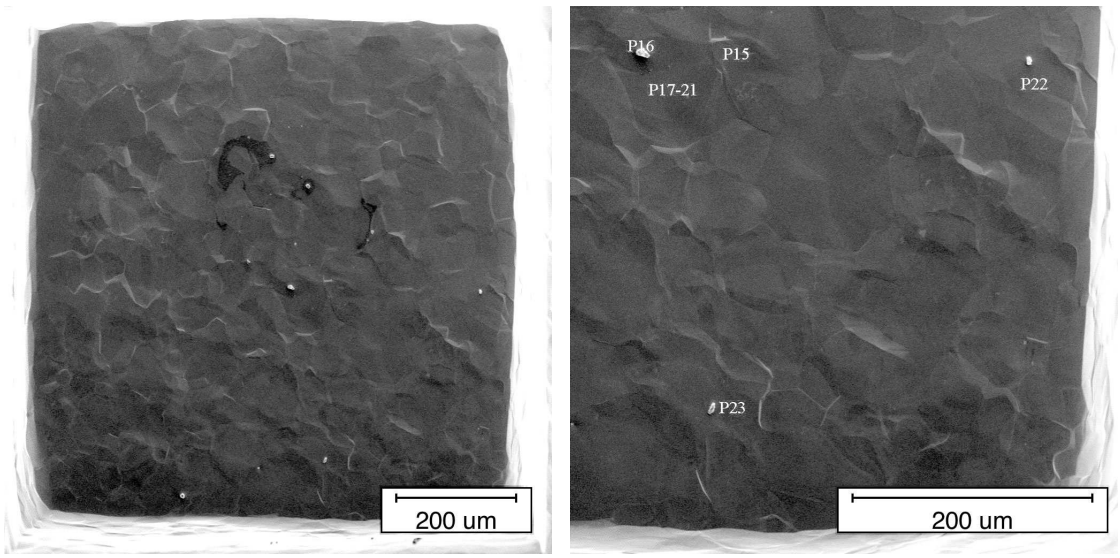
nickel, indium, alumina, and palladium. Mixing the powder in methanol, a capillary tube was used to place a small drop of the mixture onto a pedestal (the methanol evaporates immediately). Though we strove to avoid contact between the pedestal and the capillary tube, occasionally the tube did touch the surface, leaving a small scratch visible in the scanning electron microscope (SEM).

To achieve our goal of 3–30 particles with sizes between 1 and 10 microns, and none larger, we carefully honed our skills on a practice sample before each contamination, adjusting technique and powder concentration as necessary. Nevertheless, the art of contamination proved difficult to master, and many successful contaminations were the result of some trial and error. When inspection in the SEM revealed too few particles, the contamination process was repeated. Worse, if too many (or too large) particles cluttered a pedestal, the plate had to be re-cleaned before being contaminated again.

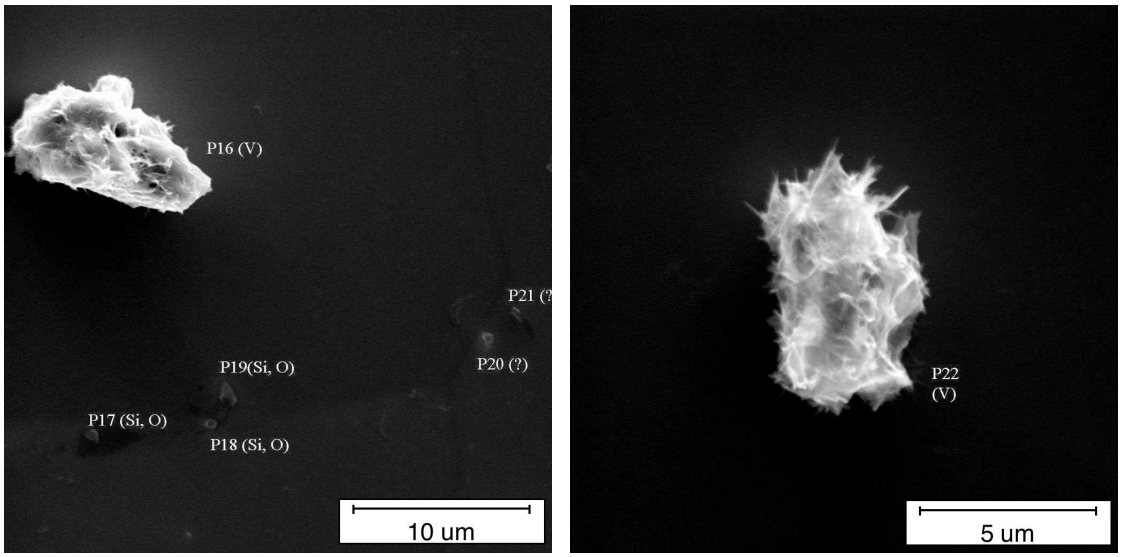
2.2 Pre-Test Examination

Having cleaned and contaminated a cathode plate, we scanned it in the SEM so that after the test we could spot any changes caused by breakdown. Small pedestals greatly facilitated SEM examination.

Each $(1\,\mathrm{mm})^2$ pedestal (reduced to about $(0.7\,\mathrm{mm})^2$ by etching) was divided into 4 quadrants, and each quadrant photographed separately. From the quadrant pictures, particles could be easily spotted. Particles (and other features, if present) were numbered, photographed closely, and identified with EDX (energy dispersive x-ray spectroscopy distinguishes elements based on their x-ray signatures; see section 3.4). Examples are displayed in figure 2.3.



(a) (b)



(c) (d)

Figure 2.3: SEM pictures taken of a pedestal before testing. (a) the entire pedestal; (b) the LR (lower-right) quadrant; (c) in the LR quadrant, particles 16-21, and (d) particle 22.

During the contamination stage we tried to deposit only the kinds of particles we wanted, but inevitably there are adventitious[5] particles, and, given the difficulty of intentional contamination, we might as well take advantage of fortuitous contamination.

2.3 Installation in the Apparatus

Thoroughly documented, the cathode plate was installed as the cathode in the testing apparatus (figure 2.4; see also section 3.1).

Inside the inner sanctum of the clean room (class 100, where full bunny suits are *de rigueur*, even after Labor Day), the apparatus was vented to atmospheric pressure with dry, filtered nitrogen, the cathode attachment was demounted, the old cathode plate removed, and the new plate attached. Throughout the process we took pains to minimize the possibility of contamination; for instance, we tried not to work above the plate to avoid interrupting the laminar flow of clean air from the ceiling filters.

After sealing the vacuum chamber, the apparatus was moved out of the inner sanctum into the class 10,000 clean-room area, and was pumped down, slowly at first to avoid dust-roiling turbulence, with a turbo pump backed by an oil-free scroll pump. Once the pressure neared 10^{-6} torr, the external turbo pump was disconnected, and the valve to the ion pump opened. Within a day, the pressure would drop to nearly 10^{-8} torr.

[5]Bob Kirby at SLAC taught me this word; it means "added from outside; not inherent; accidental" (Webster's New World Dictionary) and is very apt in this context. It's also the nicest synonym for "accidental" that I know.

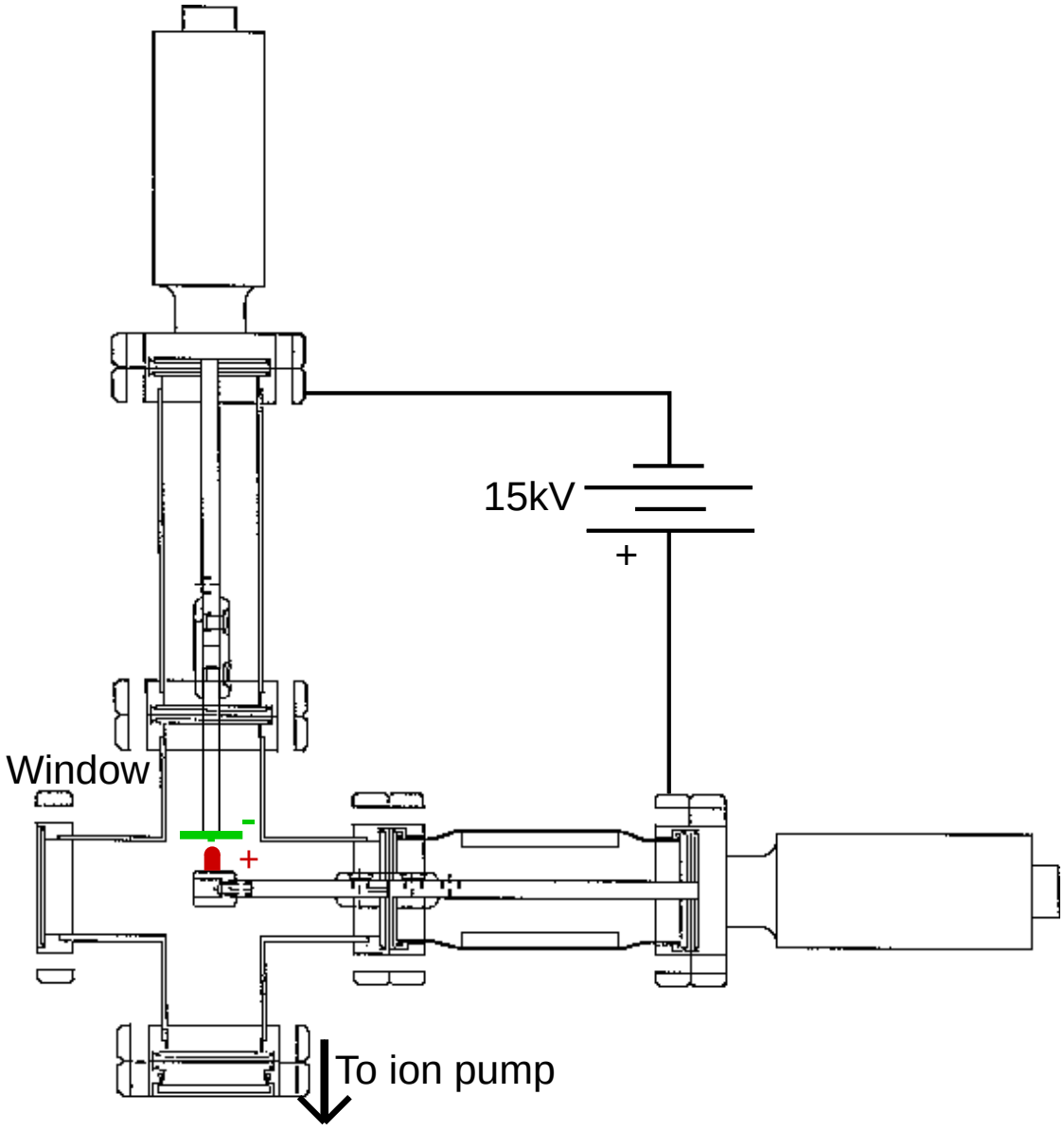


Figure 2.4: The Apparatus. The cathode bolts upside-down (so no dust can fall on it) onto a vertical rod that can be moved up and down to adjust the gap between the cathode and anode. The anode can be moved along a horizontal line underneath the cathode (in line with the pedestals on the cathode plate). Both the anode and cathode are electrically isolated from the rest of the apparatus.

2.4 Alignment and Electronics

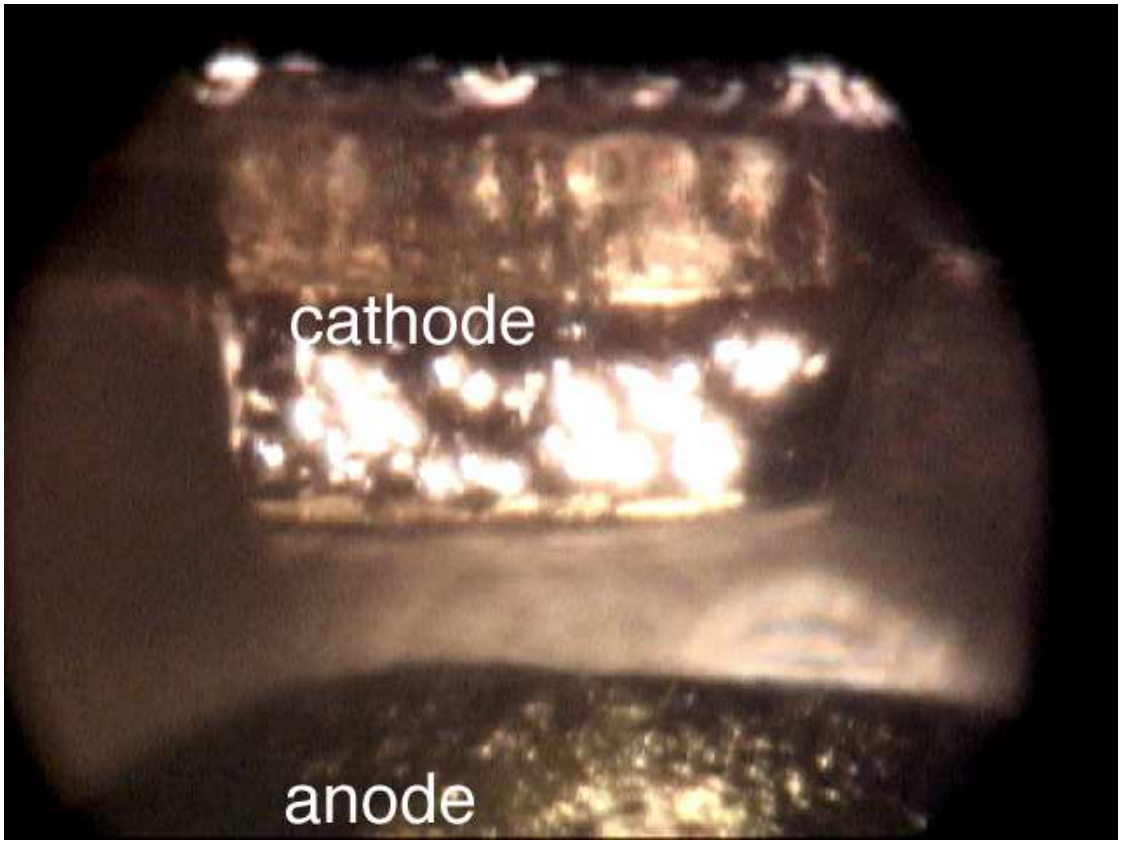


Figure 2.5: The anode and the first pedestal on the cathode plate, seen through the Questar near-focus telescope. The image shows an area a little more than 1 mm in height—the cathode is above the anode. Surface irregularities reflect the light illuminating the electrodes, creating bright spots in the image.

A window in the apparatus (figure 2.4) offered a direct view of the electrodes; a Questar near-focus telescope (or far-focus microscope) provided a magnified image of the electrodes with a resolution of a few microns (figure 2.5). The near-focus telescope helped to position the anode across from the cathode pedestal, and to adjust the gap between them (see section 3.1.1 for details). With practice, I learned to set the gap within 10–15 microns of the desired spacing; this error limited the

accuracy to which the electric field could be known. Most tests began with a gap of 150 microns, progressing to 100 and 75 or 50 micron gaps if necessary.

Once the electrodes were aligned, instrumentation and shielding could be set up (see section 3.2); in later tests the following measurements were recorded:

Measured Quantity	Recording Device
DC gap voltage	DMM (digital multimeter)
field emission current	DMM and strip-chart
AC gap voltage (<40 MHz)	oscilloscope
AC cathode current (<70 MHz)	oscilloscope
Vacuum Pressure	DMM and strip-chart
Light from spark (30 frames/sec)	video camera

2.5 The Test

Each pedestal was tested by raising the voltage (in half-kilovolt steps roughly every 30 seconds) until breakdown occurred; if no breakdown occurred even at the highest possible voltage, 14.5 kV, the gap was shortened and the voltage raised again.

Pre-breakdown

While raising the voltage we monitored the field emission currents; a logarithmic current meter allowed measurement of currents between 1 nA and 1 mA (see section 3.2.2). Only rarely did field emission currents rise higher than 10 μA before breakdown (and never higher than 100 μA). Figure 2.6 demonstrates unusually clean field emission—steps occur at voltage increases. Figure 2.7 shows more typical (noisy) field emission.

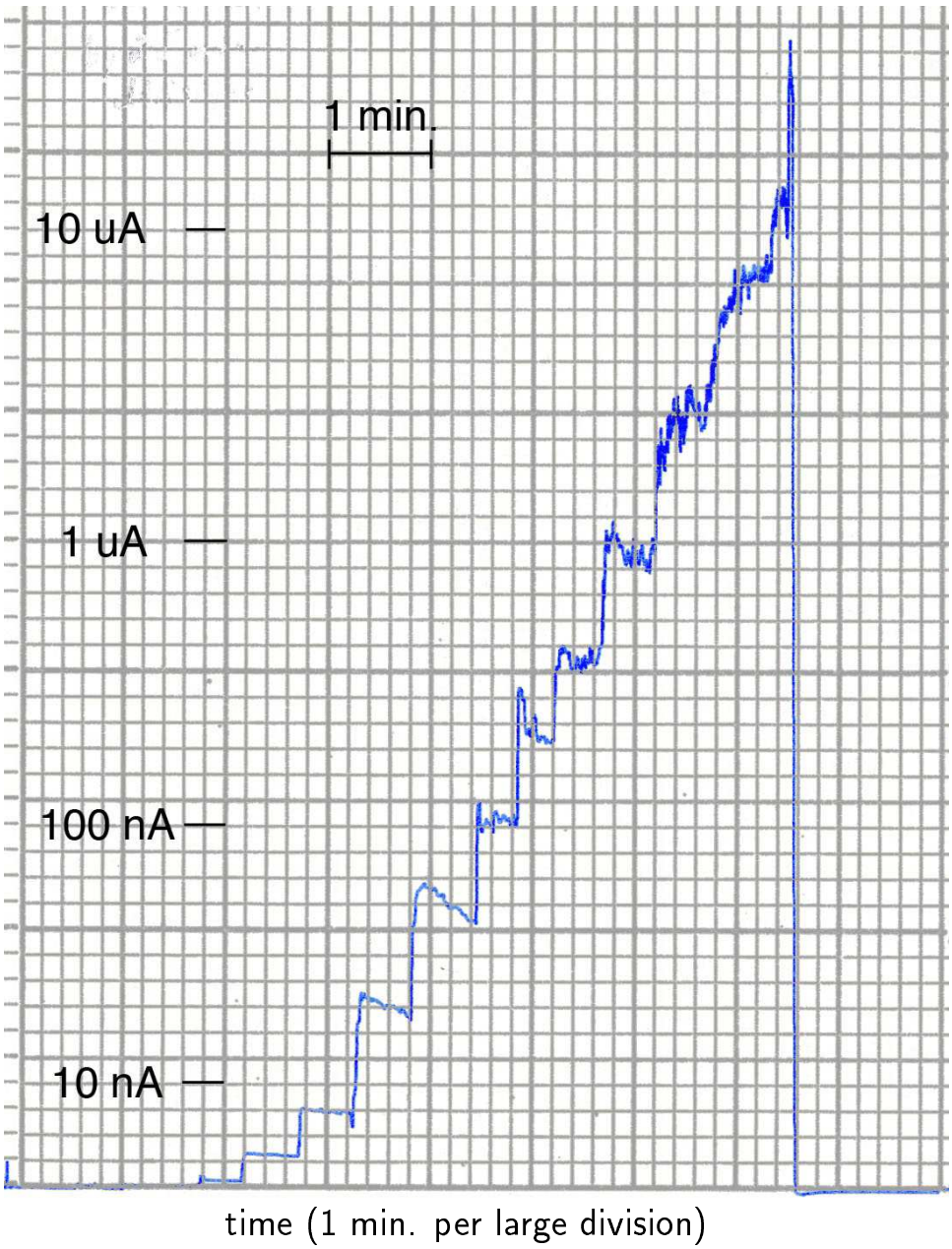


Figure 2.6: Field emission current versus time before breakdown. This figure was chosen for ease of understanding rather than typicalness (see figure 2.7 for more typical field emission); each step occurs as the voltage is raised 0.5 kV, from 6.5 kV until breakdown occurred at 12.0 kV.

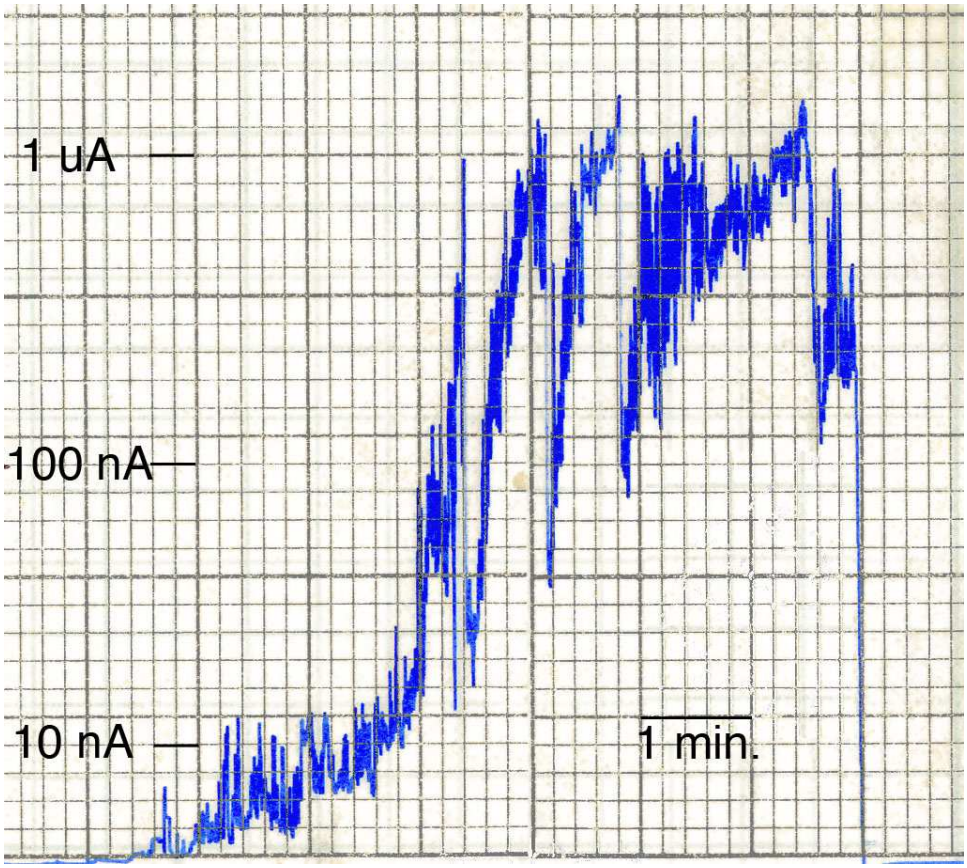


Figure 2.7: Typically field emission was much noisier than in figure 2.6. (No breakdown occurred even at the highest voltage, 14.5 kV.)

The current in the ion pump provided a rough measure of the vacuum pressure. Sometimes the vacuum pressure increased with field emission; figures 2.8 and 2.9 show the graphs of background vacuum pressure and field emission for two experiments (with and without pressure rise); while pressure rises were not uncommon, they tended to be the exception rather than the rule.

Breakdown

With the onset of breakdown, the current rose to 10–100 amps and the gap voltage fell in less than 10 ns (figures 2.10 and 2.11). The oscilloscope measuring current and voltage triggered on the falling voltage signal. As soon as possible (in less than a second) after breakdown I turned off the voltage.

Breakdown fields spanned the range between 30 and 150 MV/m (and occasionally higher). Figure 2.10 depicts breakdown at a voltage around 10 kV at a gap of 100 microns—a breakdown field of 100 MV/m. While the field on the pedestal being tested reached 100 MV/m, the other pedestals on the cathode, more than 3 mm away, experienced fields less than 4 MV/m.

The arc

The arc lasted several hundred nanoseconds (figure 2.11), during which the voltage remained low and the current tapered to zero. The vacuum pressure spiked, and usually a brief flash of light and sometimes a sharp crack (similar in sound to a good static discharge on a dry wintry day) accompanied the breakdown event.

The light emitted by the spark appears in a single 1/30 s video frame (figure 2.12); rarely was any light visible in frames before or after the spark.

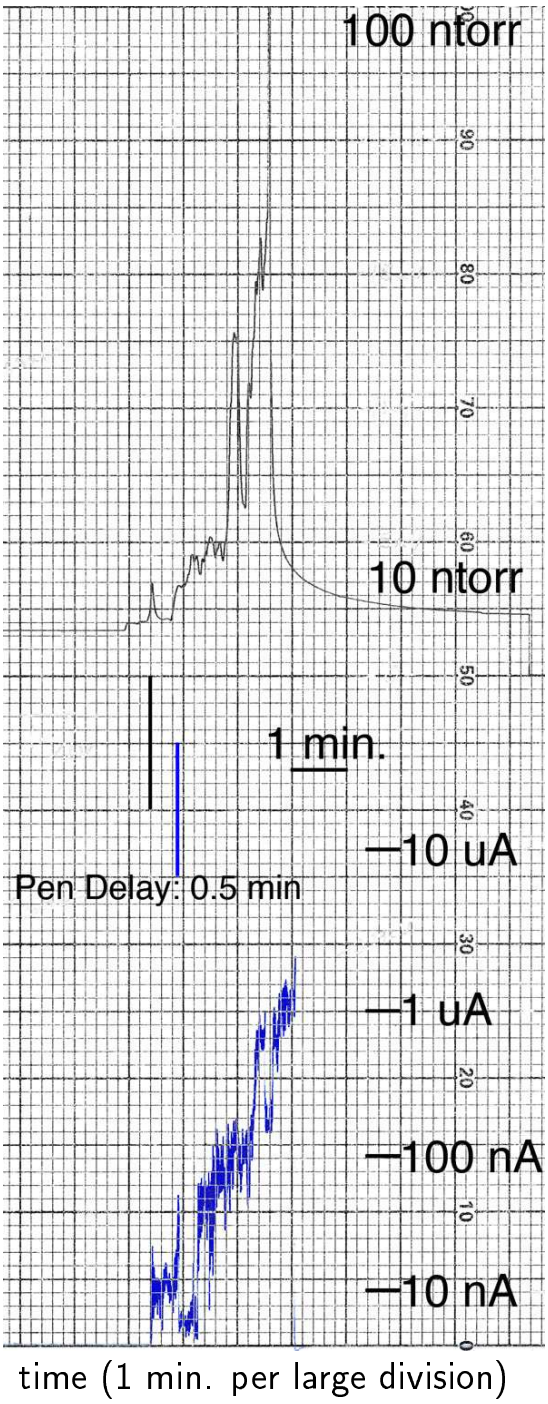


Figure 2.8: Graphs of field emission (below) and pressure (above) versus time sometimes show a strong correlation between field emission and pressure (note that the field emission graph is one-half of a large division behind the pressure graph).

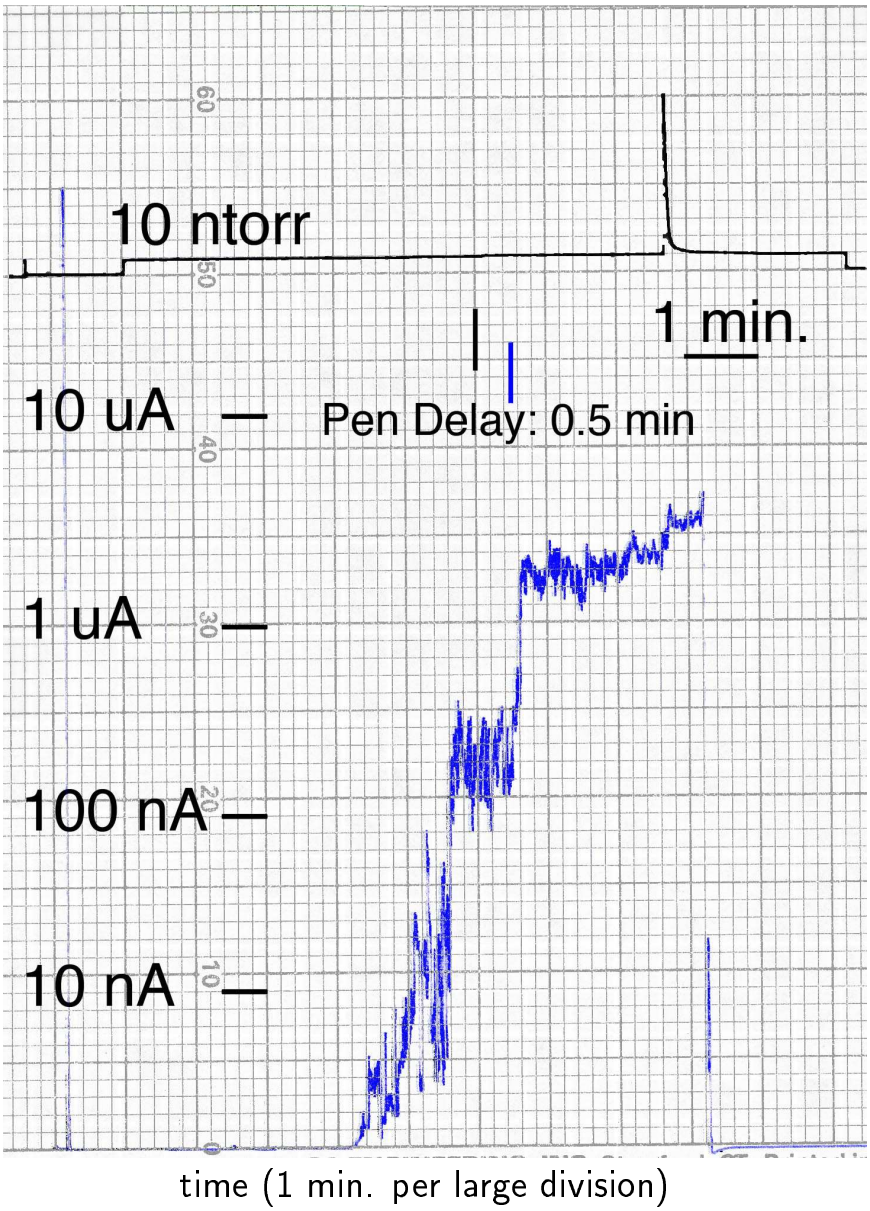


Figure 2.9: Often no pressure rise (shown above) occurred with field emission (below); the pressure spike occurred during the arc.

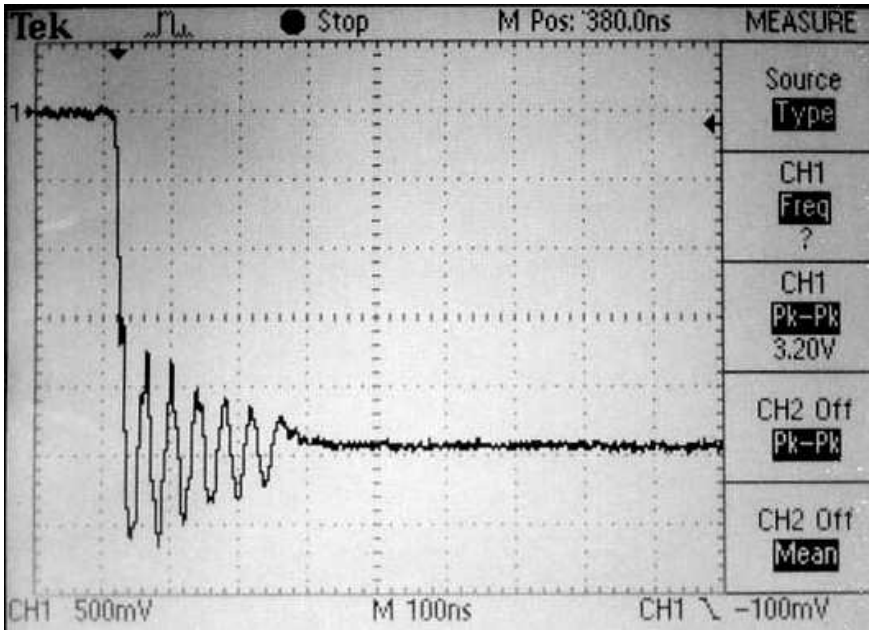


Figure 2.10: The gap voltage during voltage breakdown; the voltage scale is 2.8 kV per (large) vertical division, and the time scale 100 ns per (large) division. The oscillations reflect resonances of the measurement circuitry or RF noise, and not oscillations in gap voltage. The important feature is that the voltage drops 10 kV in less than 20 ns.

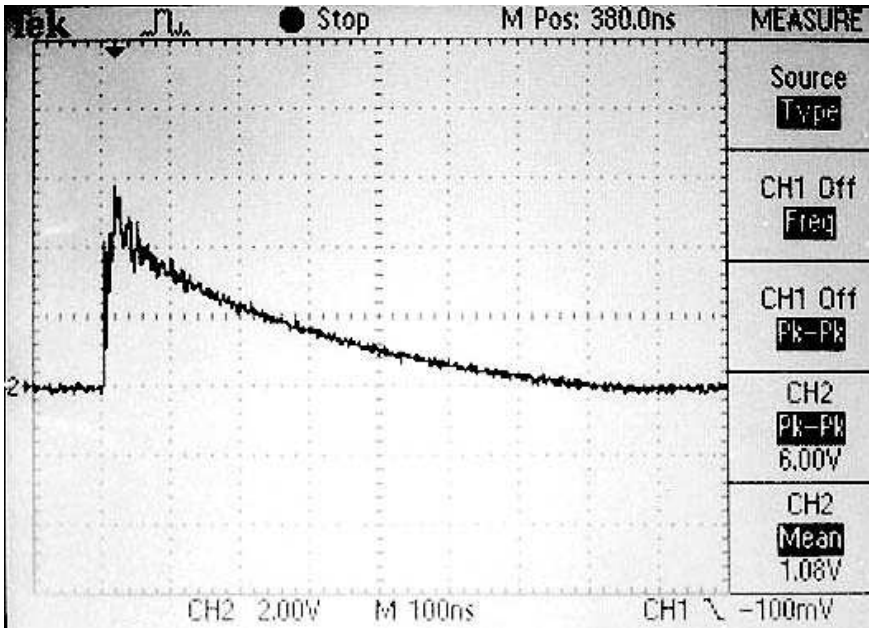


Figure 2.11: The cathode current (arc current) during voltage breakdown. The current scale is 20 A per (large) vertical division, and the time scale is 100 ns per (large) division. The bandwidth of the current measurement circuit is 70 MHz (see section 3.3.3).

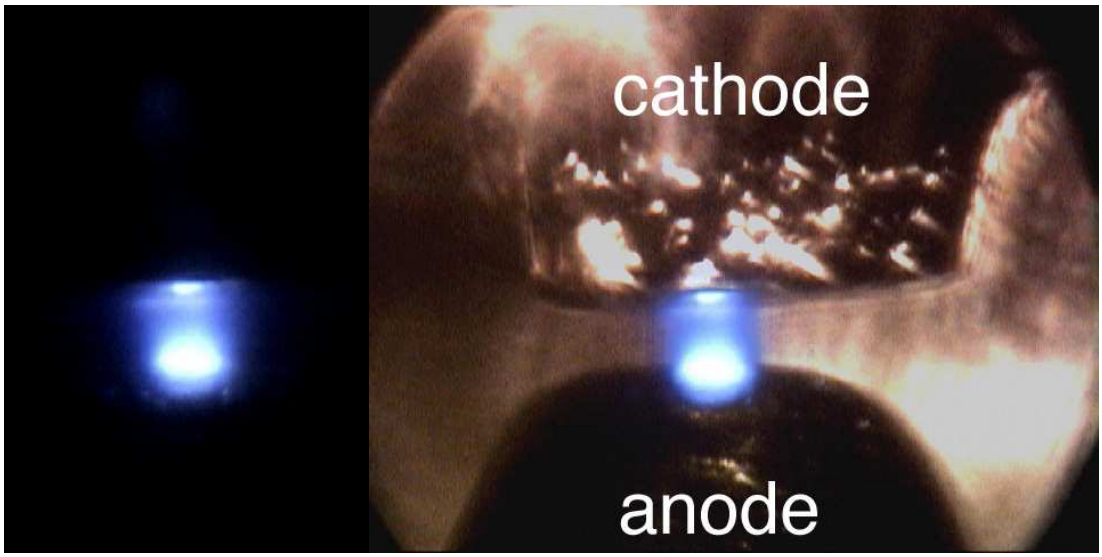


Figure 2.12: Light emitted during a voltage breakdown at 13.5 kV (left), overlayed on the picture of the electrodes (right). The spot of light on the cathode is smaller than the spot of light on the anode. The gap between electrodes is 100 microns.

2.6 Post-test Examination

After testing all pedestals, the cathode plate was removed from the apparatus and examined for changes. As with the pre-test examination, quadrant pictures of each pedestal were taken, and any changes from the pre-test were duly noted and photographed. Figure 2.13 shows pictures taken after breakdown of the same sites shown in figure 2.3. The darkened regions in figure 2.13 parts (c) and (d) are typical "starbursts" with a round center and streamers emanating from the center. Starbursts are often centered around one or more craters; in addition, sometimes small "satellite craters" appear outside the central crater region as in part (c).

Besides taking pictures, we tried to identify with EDX any remains of particles, possible anode material deposited during the arc, or any other foreign matter. Although we occasionally found some remnant of the original particle or some anode material, more often than not EDX detected nothing other than the substrate.

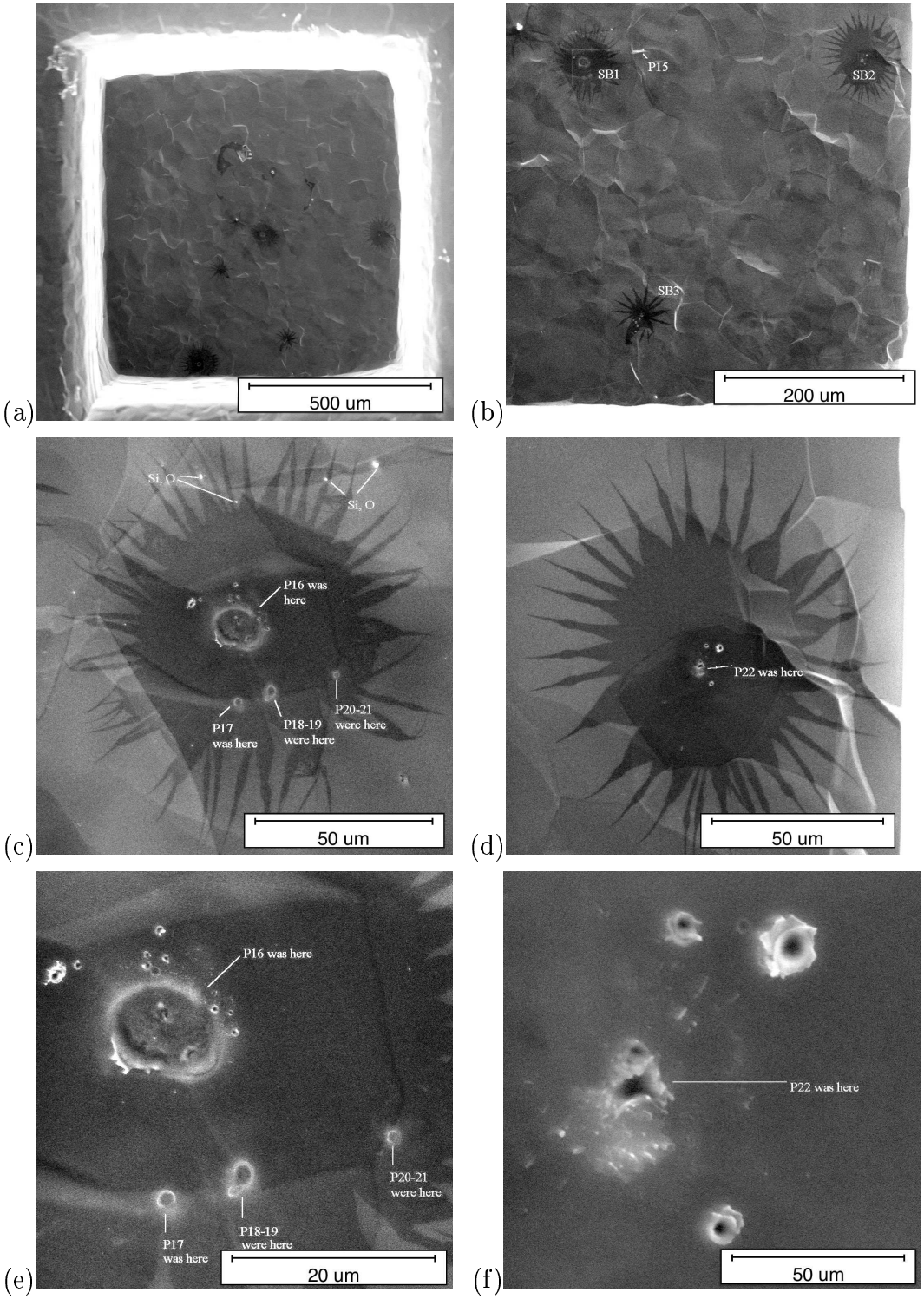


Figure 2.13: Damage caused by voltage breakdown (the pre-test pictures are shown in figure 2.3). (a) the entire pedestal; (b) the lower-right (LR) quadrant; (c) in the LR quadrant, SB1 (starburst 1), formerly particles 16-21, and (d) SB2, formerly particle 22; (e) SB1 center, (f) SB2 center.

A more surface-sensitive analysis technique, AES (Auger electron spectroscopy; see section 3.4), often could detect remains of the original particle. Whereas EDX is sensitive to elements within a volume roughly of size 5 microns, AES is sensitive to only a few surface layers. Therefore AES can detect, for example, a 10 Å thick surface layer of carbon or a sub-micron sized aluminum particle on a niobium substrate, when EDX would see only niobium, because 10 Å of carbon or a sub-micron sized aluminum particle are only a very small fraction of the several micron sized volume analyzed by EDX. Figure 2.14 shows Auger maps of the starburst in figure 2.13.

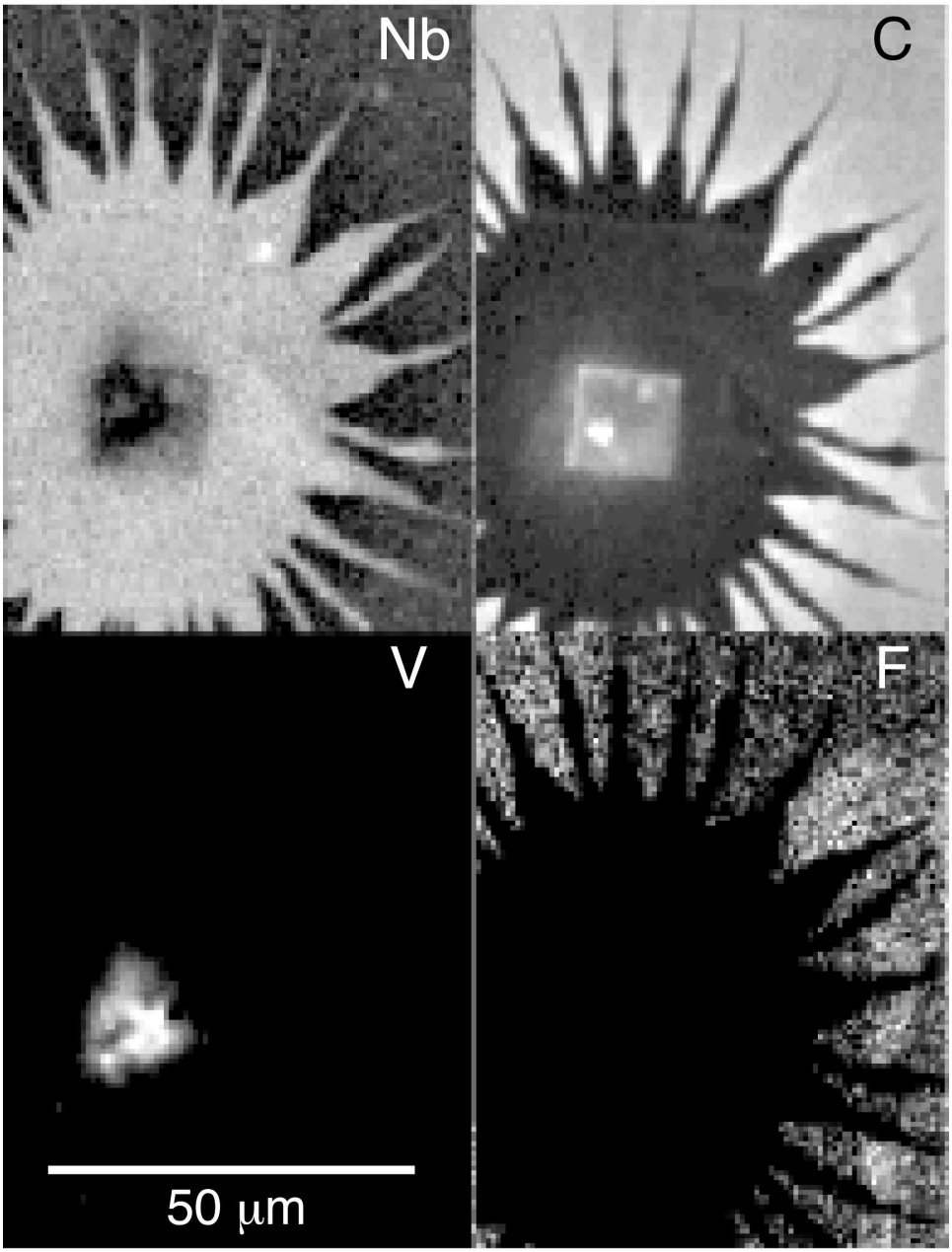


Figure 2.14: AES maps of the concentration of Nb (the substrate), C, V, and F in SB2 in figure 2.13 (also see figure 2.3) show remains of the original particle (V), depletion of surface contaminates (C, F), and the substrate (Nb) for comparison. Lighter areas represent higher concentrations.

Chapter 3

Apparatus and Instrumentation

3.1 Apparatus

Figure 2.4 shows a diagram of the apparatus used to perform DC breakdown experiments. The cathode and anode mount to rods connected to linear-motion-feedthroughs[1] that are electrically isolated from the rest of the apparatus. The anode moves horizontally to position it underneath different pedestals of the cathode plate; the cathode moves vertically to adjust the gap between the electrodes. Each has a nominal range of one inch with an adjustment precision less than one thousandth of an inch.

Beneath the electrodes is a 20 L/s ion pump, and valves and fittings for isolating the ion pump, venting the apparatus to atmospheric pressure (with dry, filtered nitrogen), and pumping it back down to vacuum.

Section 2.1 contains a description of cathodes and their preparation. Cathodes bolted onto the cathode mount with three small stainless-steel screws.

[1] Linear motion feedthroughs allow motion along a line within the vacuum chamber by turning a dial on the outside (motion is "fed through" to vacuum).

The very first experiments used niobium anodes mounted to the anode feed-through with a single screw; we soon changed to a tantalum-tungsten alloy (about 10% tungsten) so we could distinguish anode material on a niobium cathode after breakdown; we chose that particular alloy for its high melting temperature and machinability. Figure 3.1 shows the two different shapes of anode; most experiments used the large anode; the pointier anode limited the high-field region to an even smaller area, which was useful for a few diamond-machine copper cathodes that had pedestals with very sharp corners. Figure 3.2 shows the field at a plane cathode for various distances between the cathode and anode. Anodes were thoroughly cleaned and etched (like cathodes; see section 2.1) before installation.

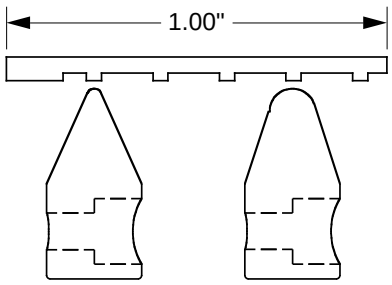
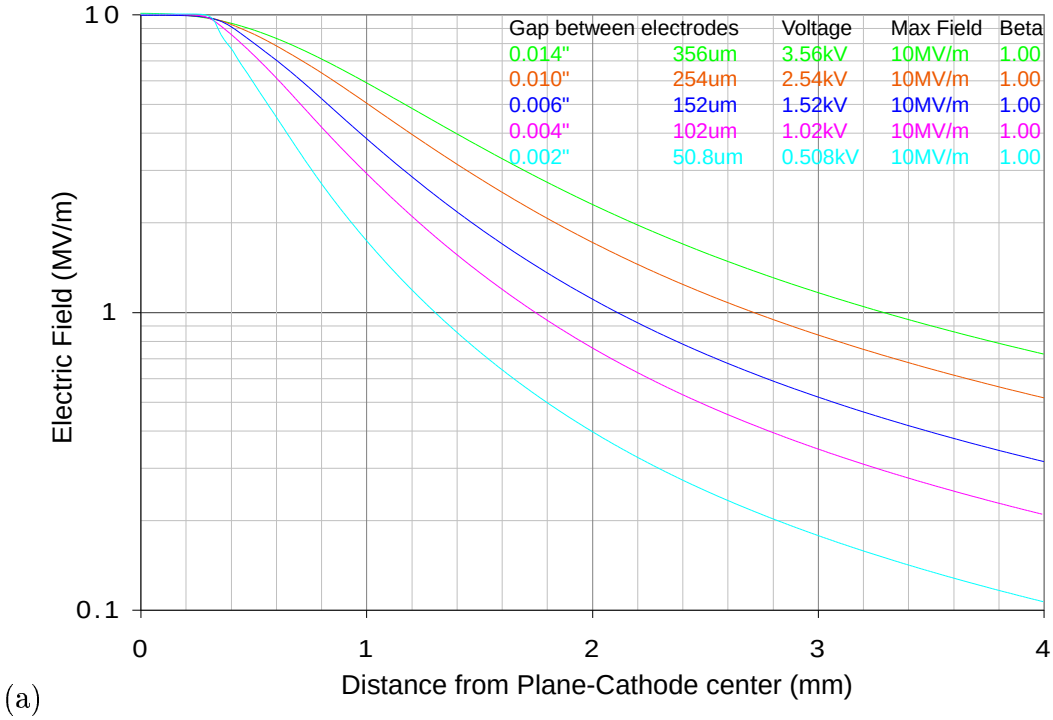


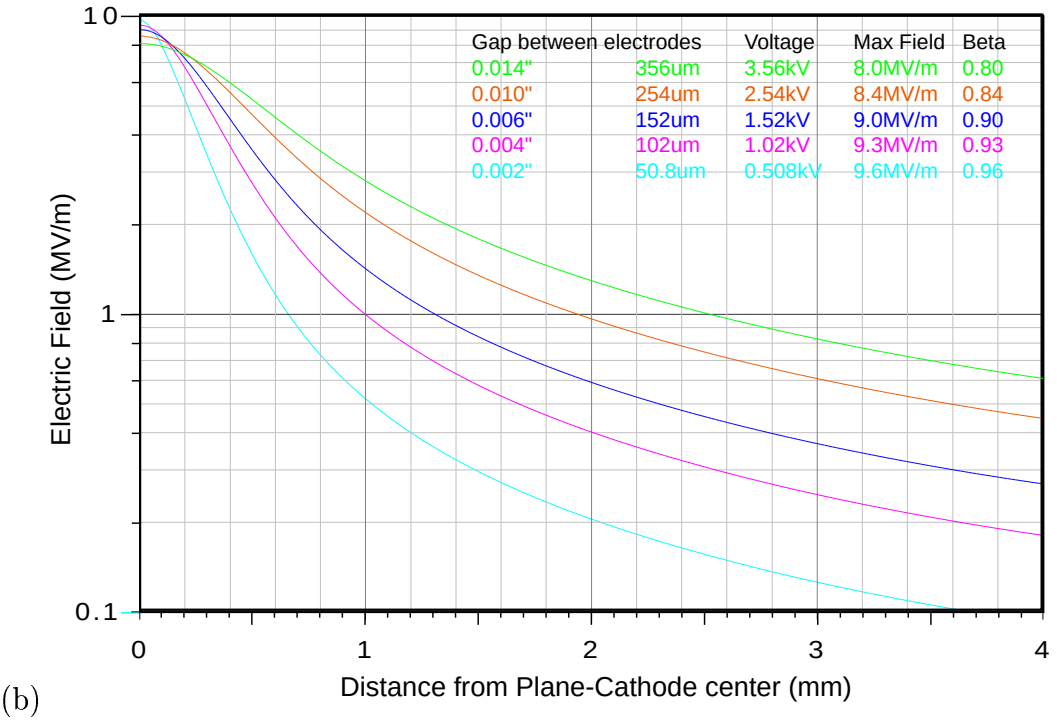
Figure 3.1: Anode shapes, pictured underneath a 5-pedestal cathode with 1 mm pedestals: the most often-used anode on the right (with a 60 mil radius at the tip), and the pointier anode on the left (with a 20 mil radius).

3.1.1 Electrode Alignment

Before each test the anode had to be positioned underneath the desired cathode pedestal (see figure 3.1). A window in the apparatus (see figure 2.4) offered a direct view of the electrodes; with bright illumination, a Questar near-focus telescope (or far-focus microscope) provided a magnified view of the electrodes with a resolution of a few microns (see figure 2.5). Such a high magnification rather limited the depth of resolution, so the entire cathode could not be focused simultaneously,



(a)



(b)

Figure 3.2: The surface electric field on a cathode plane at different differences from the anode: (a) the most often-used anode has a high-field region with radius ~0.4 mm; (b) the pointier anode has a smaller high-field area on the cathode (however, for small gaps the field at the anode is higher than the field at the cathode, reducing the cathode field by a factor β).

which helped when aligning the cathode and anode—if both electrodes appeared in focus, then they couldn't be too far out of alignment.

Whenever moving the anode or cathode, a continuity detector connected between the two sounded an alarm if they accidentally came in contact. Perfect alignment was often impossible, with control over only two dimensions of motion—the anode moved horizontal along a line, and the cathode vertically. Actually, the anode fixture, perhaps because of its length and horizontal extension, did not move along a straight line, but in a tightly-curled corkscrew. This allowed a little motion in the transverse horizontal direction, at the expense of motion in the other horizontal direction.

After centering the anode under the pedestal nearest the window (and easiest to see) as well as possible given the limited range of motion, the centers of other pedestals could be found simply by moving the anode the proper distance (0.175 inches for most cathodes). With the anode centered under a pedestal, the gap was adjusted by moving the cathode up and down. Watching carefully through the Questar telescope (ever mindful of the continuity detector), I carefully lowered the cathode toward the anode. When the electrodes were so close that I dared not bring them any closer together for fear of them touching, they were within a half-mil (10–15 microns) of touching. More than once I tested myself by lowering the cathode one-half mil more than I thought I should and heard the continuity detector alarm. Occasionally, the continuity detector went off when I thought the electrodes still looked safely separated; this happened rarely enough, yet often enough, that it increased my confidence in my judgment. That accidental contact occurred occasionally indicates that the gap was not actually much larger

than it appeared; that accidental contact did not happen too often is evidence of consistency in setting the gap.

After estimating the point of contact (generally a half mil away from the as-close-as-I-dare point), I then raised the cathode to reach the desired gap. The linear motion feedthroughs to which the electrodes were mounted had 40 turn/inch threads, and the dial was marked in 1 mil (25 micron) divisions, which could be easily read to ± 0.2 mils.

Although I noted in the logbook a guess of the error in the estimation of contact point for each alignment (varying according to the visibility, and my nerve) most of the gap estimates are correct to ± 0.5 mils (± 15 microns). For small gaps, for example $d = 2$ mils $= 50$ microns, that is a fairly large error, which limits the accuracy with which the electric field can be known.

I experimented with another method of estimating the gap, using the capacitance between the cathode and anode. Although the stray capacitance to ground from both electrodes was much greater than the tiny capacitance between them (less than 1pF), I was able to measure inter-electrode capacitance with nearly 1% precision. Unfortunately, the capacitance between the pedestal and the anode tip varied a bit depending on how well the anode was centered, and furthermore was smaller than the capacitance between the anode (including the anode fixture) and the rest of the cathode plate. Occasionally, however, this method proved useful, especially for larger, diamond-machined (more precisely flat) pedestals.

3.1.2 Circuit Diagram

Basically the circuit is a single loop with high voltage generator providing up to 15 kV, a chain of resistors to limit the maximum DC current to 2 mA (which limits

the maximum power dissipation in the resistor chain to a sustainable level), the vacuum gap, and current measuring circuits (figure 3.3). In addition, high voltage probes measure the gap voltage.

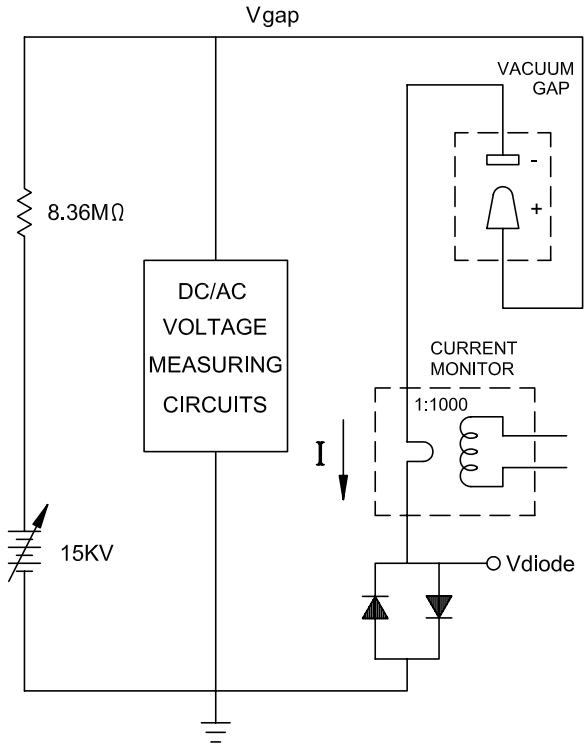


Figure 3.3: The basic circuit

The voltage across the diodes between the cathode and ground reflects the cathode current (using back-to-back diodes allows positive and negative current to flow and be measured); currents between 1 nA and 100 mA can be measured this way at frequencies less than 1 MHz. Much higher and faster currents (during the arc) can be measured with the 1:1000 current transformer.

Stray capacitances not shown in the diagram become important at high frequencies. The capacitance between cathode and anode ranges between 0.5 pF and 0.9 pF at typical gap spacings, depending on their exact positions. The capacitance between the anode and ground, however, is much larger, about 100 pF (most

of that is 30 inches of high voltage cable connected to the anode feedthrough). The capacitance between the cathode and ground is 14 pF.

The small capacitance between anode and cathode (and keeping the cathode voltage constant near ground) allows currents flowing through the cathode to be measured despite voltage fluctuations across the gap; because of the small capacitance, displacement currents are small compared to particle currents.

The actual circuit is a bit more complicated; figure 3.4 shows a little more detail. A set of double back-to-back diodes provides a backup ground return in case an open circuit should develop in the single-diode return. The voltage across the diodes travels through a perfunctory 100MHz low-pass filter to an optical isolator that transmits the diode voltage optically to avoid ground fluctuation problems and noise transmission. Gap voltage measurement is split into two parts, DC and AC (described in detail in following sections).

3.2 Slow Electronic Measurements

The gap voltage and field emission current before a breakdown event were the most important measurements made.

3.2.1 DC Gap Voltage

The gap voltage could be varied up to 15 kV. A voltage divider (usually 1000:1) allowed such high voltages to be measured with an ordinary Fluke DMM (digital multimeter). A Keithley 1600A high voltage probe with 1000:1 voltage reduction served well for many experiments; this commercial probe had a bandwidth of

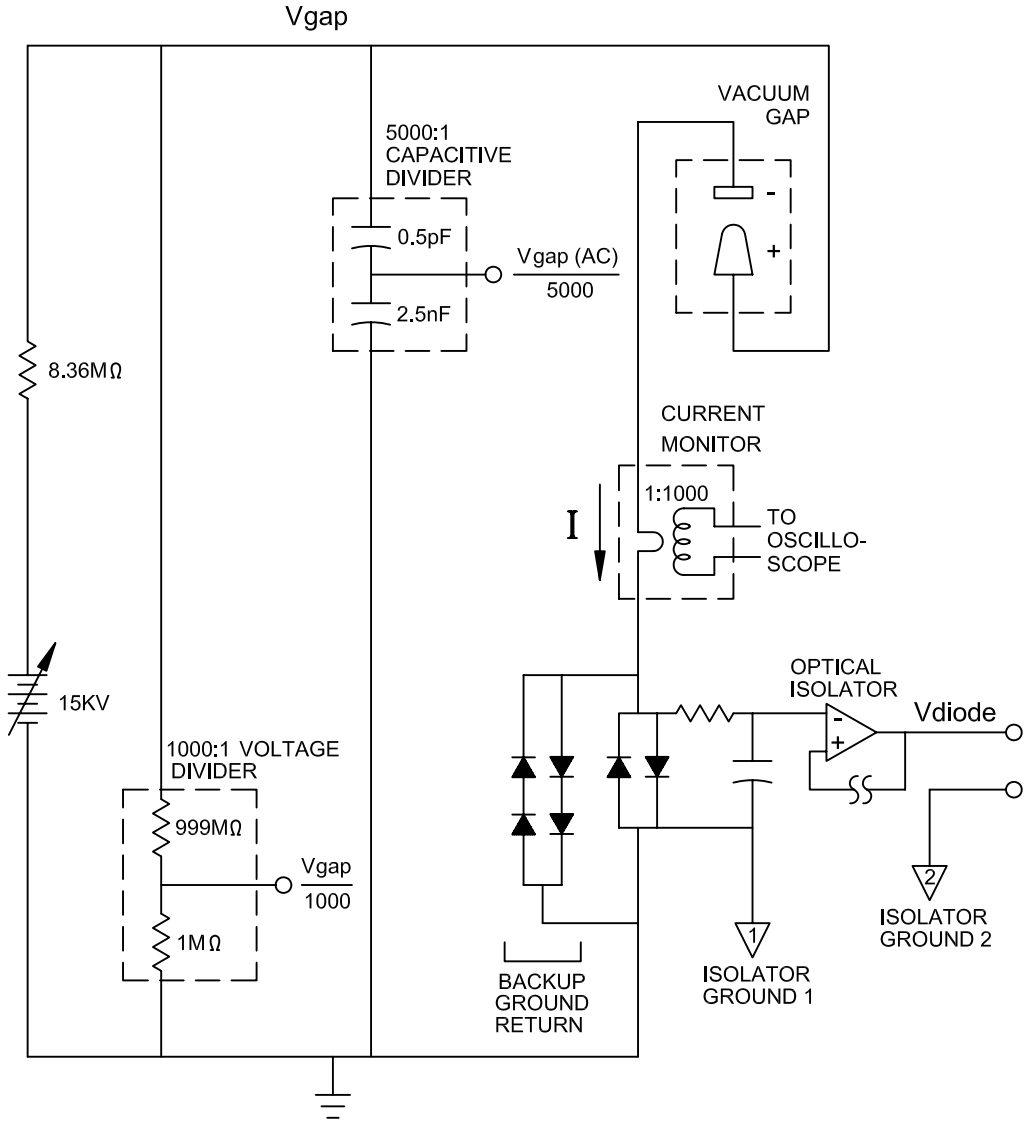


Figure 3.4: A more realistic circuit diagram

300Hz—fine for DC (pre-breakdown) measurements. The error in DC voltage measurement was less than a few percent.

3.2.2 Field Emission Current

The extreme variability and range of field emission current posed a difficult measurement problem. The simplest solution, a Keithley pico-ammeter between the cathode and ground, proved unhelpful because the current went out of range too often (field emission can easily fluctuate from 1 nA to 100 nA and back down within ten seconds). Also, the pico-ammeter tended to be too sensitive to RF noise generated by sparks, which occasionally destroyed its analog-to-digital converter.

To solve these problems I built a current detector that was much hardier than the pico-ammeter, and that responded logarithmically to the current; to achieve these advantages, I sacrificed accuracy—instead of 1% accuracy, I settled for about 20% accuracy (but over 6 decades of current).

The scheme was very simple: I put a pair of diodes back-to-back (so one diode would conduct forward current and the other reverse current) between the cathode and ground and measured the voltage across the diodes—the voltage across a diode is roughly proportional to the logarithm of the current (see figure 3.3).

Since the voltage across a diode, even when conducting only a nanoamp of current, is still 0.1 V or so, its resistance is on the order of 100 MΩ; the input resistance of most meters is 1–10 MΩ, so at low currents the meter conducts more current than the diode. Fortunately op-amps with very high input resistances and low leakage currents are readily available; because I was worried about ground fluctuations and damaging current spikes during an arc, I used an optical isolator to measure the voltage across the diodes (figure 3.5). The bandwidth of the amplifier

was nearly 1 MHz. However, at low currents stray capacitance across the diode shunted high-frequency fluctuations past the measurement circuit.

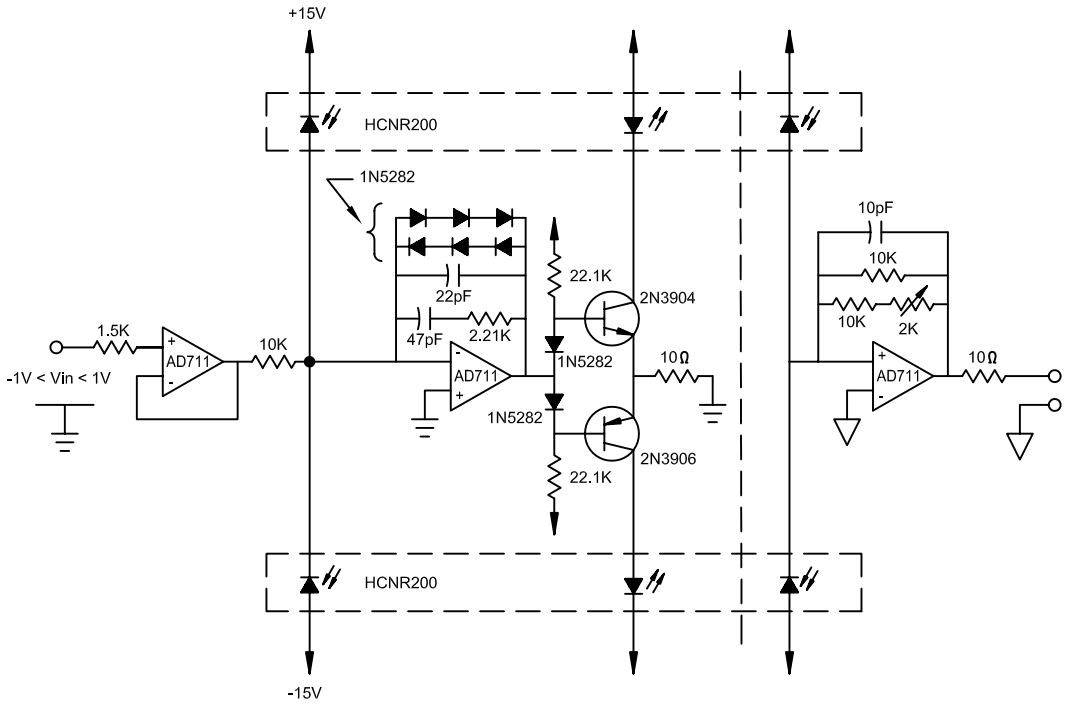


Figure 3.5: The optical isolator used LEDs with matched photodiodes in a feedback loop to adjust the current in the photodiodes to be $V_{\text{in}}/10\,\text{k}\Omega$. The output voltage differed from the input voltage by less than 2% for positive input voltages, and 6% for negative voltages.

The current-measuring circuit output a voltage that reflected the current passing between the cathode and ground:

$$I(V) = 6.69\text{ nA} \cdot \exp\left(\frac{V}{0.0503\text{ V}}\right) \tag{3.1}$$

gives the current (as a function of the output voltage of the current-measuring circuit) to an accuracy within 10% for currents between 20 nA and 200 μA.

Figure 3.6 shows a calibration of the field emission current measuring circuit.

The output voltage was measured with a Fluke DVM (digital volt meter), with a convenient feature to store the maximum recorded voltage; in some tests the output was sent to a strip-chart recorder as well.

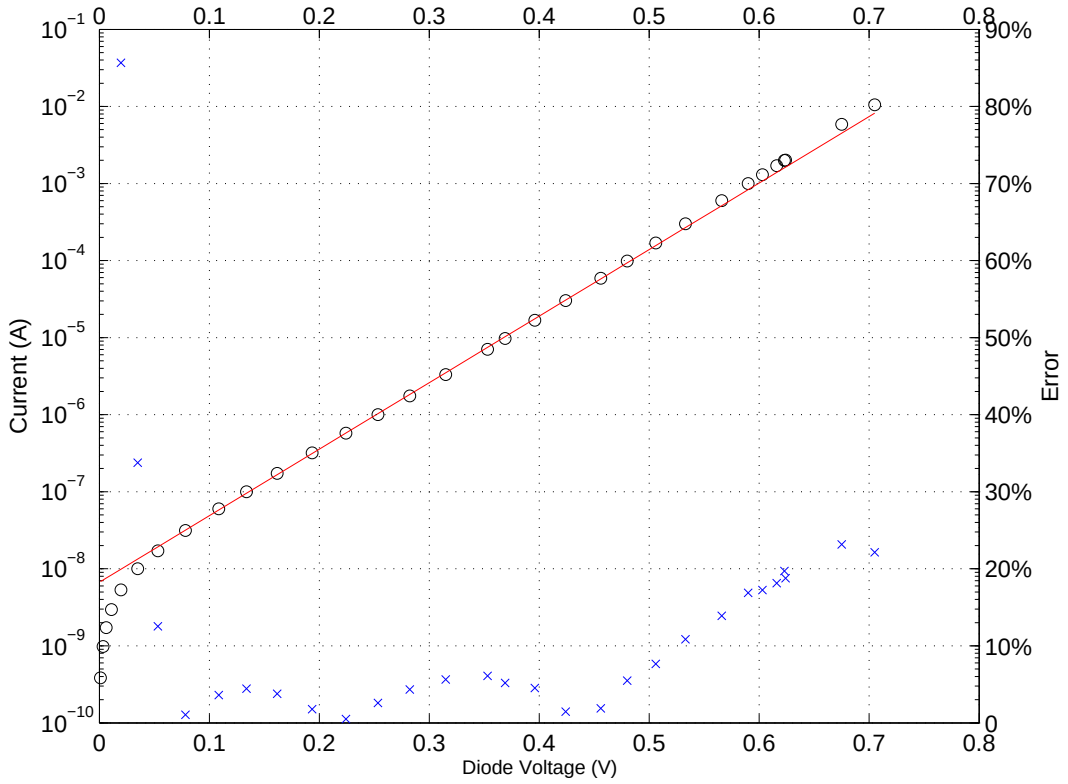


Figure 3.6: Calibration of the current-measuring circuit (at DC); measurements are marked by circles, and the error of the fit by x's.

3.2.3 Light Output

A digital video camera, with a Questar near-focus telescope for magnification, captured light emitted during voltage breakdown. Most sparks lasted and emitted light for only a half microsecond, and were caught only in a single video frame (1/30 s), so the video camera was more like a still camera with its shutter open, waiting for the event, except that the light from the spark had only to compete with 1/30 s integration of background light instead of the ten or twenty minutes that a still camera might have waited with its shutter open before a spark occurred. Very occasionally, more than one spark would be caught in the same video frame.

3.2.4 Vacuum Pressure

An ion pump maintained the chamber at a low pressure (usually less than 10^{-8} torr). The current in the ion pump reflects, to some extent, the vacuum pressure. The pressure should correspond to the ion pump current I roughly as[2]

$$p \approx 3 \times 10^{-7} \text{ torr} \left(\frac{I}{100\ \mu\text{A}} \right)^{0.93} \tag{3.2}$$

although below 5×10^{-7} torr the calibration is dubious. This calibration applies to nitrogen and water vapor.

The ion pump controller had a voltage output that indicated the pressure, roughly on a logarithmic scale, with 0–100 mV representing 10^{-8} to 10^{-4} torr at 25 mV per decade. For later experiments this voltage was recorded on a strip chart to compare with field emission current.

[2]This relationship is extracted from the Varian ion pump controller manual.

3.2.5 High-Side Field Emission Measurement

During one experiment, we investigated the effect of the ground return path on voltage breakdown. In most of the experiments, the field-emission current measuring circuit between the cathode and ground interfered with a clean return path (in terms of the diodes in the way as well as the length and inductance of the return path), which we replaced with a wide, flat strip of copper (for a low impedance) between the cathode and ground.

With the low-impedance ground return, the field emission current had to be measured on the high voltage side, between the voltage generator and the anode (rather than between the cathode and ground). As with the ground-side field emission measurement, I placed back-to-back diodes in the current path and measured the voltage across them with a high-impedance op-amp follower (since the diodes have very high impedance at low currents) and a battery-powered DMM. Since the circuitry and meter floated at high voltage, they had to be enclosed in a big box to keep them far away from ground and people.

When measuring the current on the high voltage side, the DC gap voltage was measured above (closer to the generator than) the current meter so that the ten microamps conducted by the voltage probe would not overwhelm the field emission current.

Although in principle currents as low as 100 pA could be detected, voltage fluctuations into the relatively large capacitance between the anode and ground (compared to the capacitance between the anode and the cathode) created displacement currents (considered noise in this case) larger than 1 nA. However, field emission much greater than a nano-amp could be detected, and that was sufficient.

3.3 Fast Measurements

RF noise hindered current and voltage measurements during the arc itself. The RF noise produced by a spark could disable a nearby Keithley pico-ammeter even if its input were grounded (not at all connected to the spark circuit); it could even discombobulate battery-powered DMMs (such as the one used to measure the gap voltage). Oscilloscopes were somewhat hardier, but though they survived, they "measured" nothing but noise during the spark. Good shielding was therefore the first step to measuring voltage and current during the spark.

3.3.1 Shielding

The basic principle of a Faraday cage is well known—electromagnetic radiation cannot penetrate a conductor (very far)—but the practical details can be challenging. Simply putting the oscilloscope in a metal box is not enough—signals and wall power have to be carried through the cage to the oscilloscope. Holes in the Faraday cage present no problem, as long as holes are smaller than the wavelength of radiation to be blocked (*cf.,* the cutoff frequency of a waveguide). However, if a conductor passes through a hole into the Faraday cage, then it, like a coaxial waveguide, conducts all frequencies.

An informal set of experiments with an automobile ignition coil[3] provided some insight into shielding. First, the RF noise created by the spark was incredibly

[3]An ignition coil provides the surge of voltage that creates the spark that starts combustion in the engine cylinder. It is basically a transformer with many secondary windings for each primary winding (so a low-voltage, high current pulse to the primary winding produces a high-voltage, low current pulse in the secondary). To turn the primary current on and off I used a power MOSFET driven by a sawtooth wave that increased the current slowly, and abruptly switched it off, creating a sudden inductive backlash that raised the voltage of the secondary winding to several kilovolts, creating a satisfying spark.

invasive—within the distances I could move things around in the room, virtually no $1/r^2$ decrease was apparent in the noise—the RF noise traveled not through the air like light from a lightbulb, but along wires (and pipes and other metal objects) in the room. Second, a Faraday cage is extremely effective; a Faraday cage with small holes is still extremely effective; a Faraday with large holes is still quite effective; but a Faraday cage with one small hole with a short wire going through the hole hardly offers any protection against RF noise. Third, RG223 coaxially cable, which is doubly shielded, and carefully crimped BNC connectors prevent noise from penetrating to the signal-carrying inner conductor, but single shielded cable and some kinds of crimping leaked RF. Note that signals need to pass into the Faraday cage through standard coaxial feedthroughs which maintain the integrity of the outer conductor.

With the Faraday cage, powerline filter, and shielded signal lines and feedthroughs, the voltage and current during the spark could be measured without damaging the oscilloscope or being overwhelmed by noise.

An anecdote: Aware that the power cord could behave like an antenna, I placed a high-frequency powerline filter (not a computer surge protector, but a filter with attenuation up to a GHz) just outside the Faraday cage with the oscilloscope inside, exposing only about an inch or two of power cord outside the cage. Even so, the cage did not attenuate spark noise at all; with only that short length of power cable passing through the cage, the oscilloscope was overwhelmed by RF noise. Bolting the power-line filter to the Faraday cage with an RF gasket (conductive wire mesh) completely eliminated the noise inside the cage. Now, when I first built the cage, I took a great deal of trouble to make a large door on one side with a piano hinge and copper fingerstock to make a good RF seal. Opening this door

(about 14×7 inches) allowed some very small amount of noise into the cage, that hardly affected the signal. Presumably most of the noise was at frequencies below the cutoff frequency of a 14×7 inch waveguide; in any case, it shows the danger of coaxial feedthroughs relative to gaping holes.

3.3.2 AC Gap Voltage

To get a better estimate of speed of the voltage drop during breakdown, we needed a voltage divider with higher frequency response to reduce the kilovolt gap voltage to oscilloscope-friendly levels.

A simple 1000:1 resistive voltage divider, with a total resistance around 1 GΩ, fails to maintain the 1000:1 ratio at high frequencies because even a small stray capacitance presents a smaller impedance than 1 GΩ.

A voltage divider must have two elements in series, whose impedances are in a fixed ratio, independent of frequency; since they are in series, they see identical current, and therefore, the voltage drops across the elements are related by the same ratio as the impedances. Stray capacitance in a resistive divider (figure 3.7) defeats both of these qualifications—(at high frequencies) it shunts current to ground, so the two elements no longer see the same current, and it shunts current across the resistive elements, so that the ratio of their impedances changes with frequency. A compensated voltage divider (figure 3.7) allows higher-frequency operation. Unfortunately, the compensated divider that we built turned out to be very susceptible to noise, which drowned out any improvement in high-frequency response.

Sacrificing DC and low-frequency response, an AC voltage divider (a capacitive divider, see figure 3.7) provided noise immunity (see figure 3.8). The great

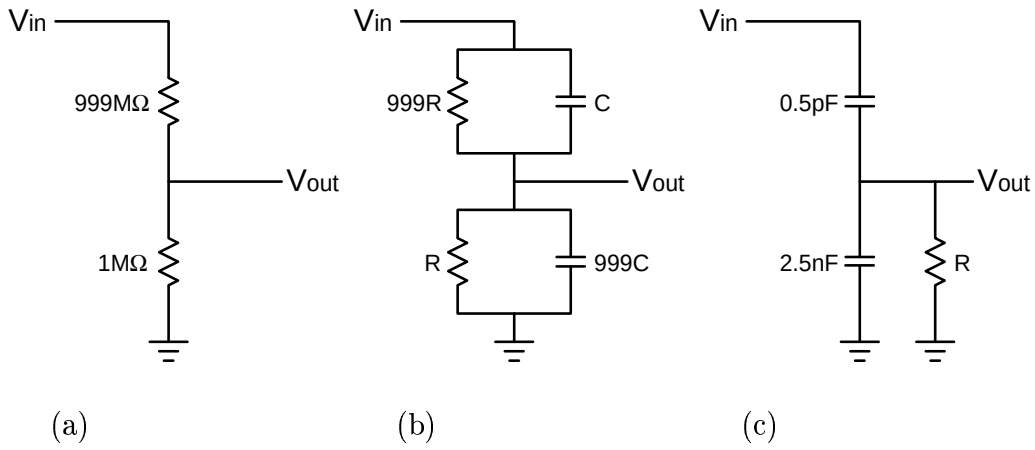


Figure 3.7: (a) A 1000:1 resistive divider; stray capacitance across the resistors and to ground changes the division ratio at high frequencies. (b) A compensated 1000:1 divider with capacitors designed overwhelm any stray capacitance, maintaining the voltage division ratio to higher frequencies. (c) A 5000:1 capacitive divider has poor low-frequency response, but requires no conductor between V_{in} and V_{out}, allowing better shielding of V_{out} against noise; the resistor R keeps V_{out} grounded "on average" and sets the low-frequency cutoff.

advantage of the capacitive divider is that the signal (displacement current) travels through air (or dielectric) instead of along a wire, and therefore can be shielded against noise; air is a very poor antenna compared to a wire. In the capacitive divider shown in figure 3.8, any RF noise would have to pass through a circular "waveguide" 2.5 inches in diameter and 3/16 inches long.

Achieving low-frequency response is the bugbear of the capacitive divider; the capacitive divider (or any capacitor) can be considered a signal generator with a very high impedance at low frequencies. High-impedance voltages are hard to measure because they are affected too much by the measuring equipment. For example, 2.7 meters of cable connecting the capacitive divider to an oscilloscope[4] limited the bandwidth to 1 MHz, while 1 meter of cable allowed frequencies up to 10 MHz.

[4]The oscilloscope had a 1 MΩ, 20 pF input; the cable was 50 Ω coaxial cable (but not terminated by 50 Ω).

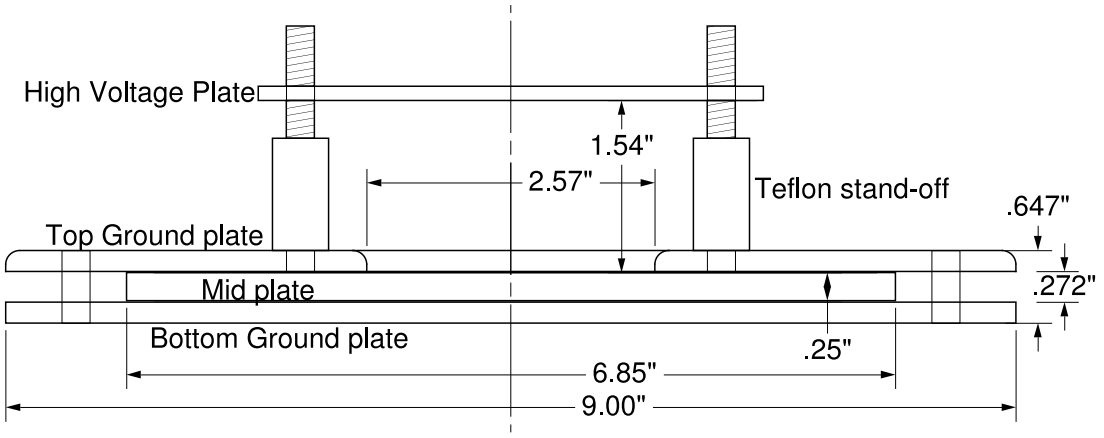


Figure 3.8: The noise-resistant capacitive divider. The two elements in the divider are the capacitor formed by the top plate and middle plate, in series with the capacitor formed by the middle plate and ground plates (both ground plates are in fact grounded); in this case, the former capacitance is about 0.5 pF, and the latter about 2.5 pF, for a division of 5000:1. 11 mil Teflon sheet (dielectric constant 2.1 [106]) separates the middle plate from the ground plates. The ground plates completely shield the middle plate from RF noise with a wavelength much larger than 2.5 inches; that "shielding" also reduces the effective area of the capacitor between top and middle plates to about 2/3 of the exposed area of the middle plate (as calculated by the electrostatic solver of MAFIA, an electromagnetic computer simulation code). An SMA feedthrough (not shown) allows the signal from the middle plate to be extracted, yet remain shielded.

Aiming for response between 100 Hz and 100 MHz, I placed a battery powered (hence easy to shield) unity-gain buffer right at the capacitive divider; the buffer could drive a 50 Ω transmission line, allowing longer cables without sacrificing frequency response. Unfortunately, I could not find a sufficiently fast buffer[5] with a low enough input current—to achieve the lower frequency limit, the input resistance must be above 500 kΩ, but the typical $\sim$ 20 μA input current of most fast buffers would raise the "zero-signal" level to 10 V (20 μA times 500 kΩ), which is beyond the range of most fast buffers and would sink 200 mA (and dissipate 2 watts) in the 50 Ω terminator.

Since the input current of an amplifier is a DC-only problem, the solution is a resistor with an impedance that increases with frequency. An inductor would do the trick, but at 100 Hz, it would need an impedance greater than 500 kΩ—a 1000 H inductor, not a possibility; only an active element will help. I managed to make a circuit (see figure 3.9) that would act like a 3 kΩ resistor at DC, and a 1.5 MΩ resistance above 1 kHz, resulting in about a 5% droop (for a constant input signal) over 200 μs.

The resulting high-voltage probe was not all I had hoped it would be. Its low-frequency response was fine (5% droop over 200 μs), but the high-speed buffer circuit wasn't as fast as I wanted; it performed well at 10 MHz and easily managed to swing the voltage at 200 V/μs, but by 40 MHz, the response diminished greatly. However, the frequency response was still an improvement over other voltage probes. The significant achievement of the capacitive divider was noise immunity, which made voltage measurement possible during the spark.

[5]Not only did the buffer need to have a bandwidth higher than 100 MHz, but it also needed a slew rate of at least 1900 V/μs, to drive a 3 V (15 kV/5000) signal at 100 MHz.

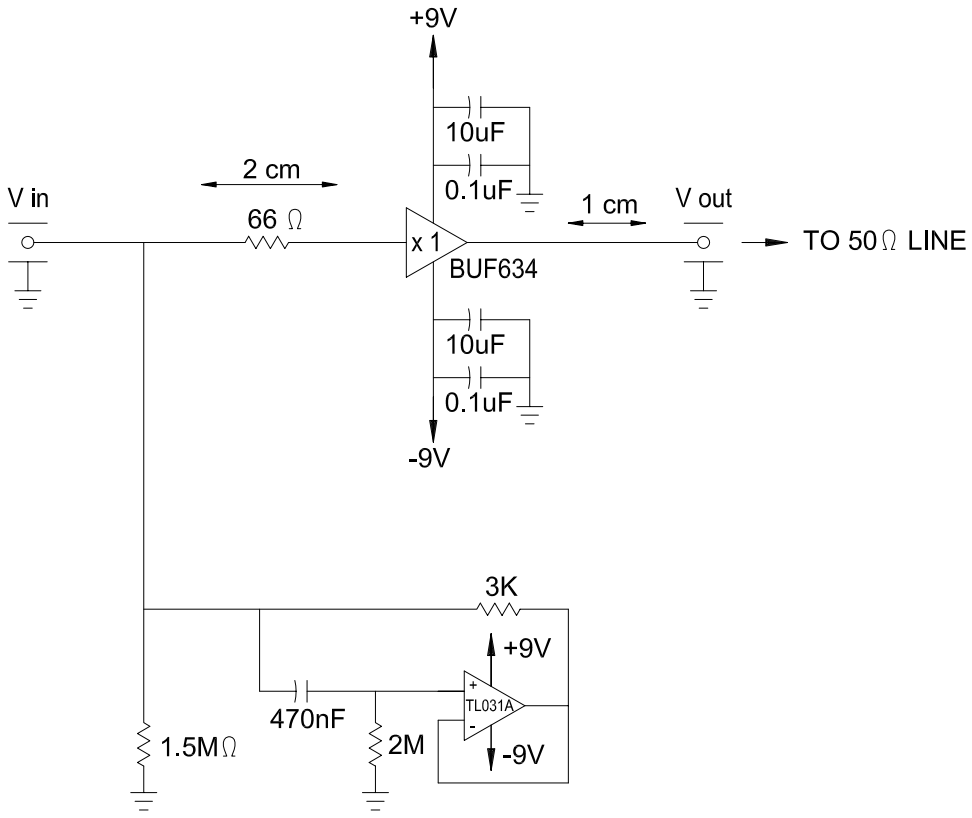


Figure 3.9: Unity gain buffer for high frequencies with an active resistor at the input that acts like 3 kΩ at DC, and increases with frequency

3.3.3 Arc Current

With an arc time less than a microsecond, we needed to measure arc currents (possibly hundreds of amps—though more likely tens of amps) at frequencies up to 100 MHz. For this we used a Pearson Current Monitor model 2878, with an output (into a 50 Ω line) of 0.05 V per amp of spark current, and a frequency response between 30 Hz and 70 MHz. The current monitor is a Rogowski coil, a transformer with 1 primary winding (for the arc current) and 1000 secondary windings. With a small termination impedance (like 50 Ω) on the secondary side, the secondary current is 1000 times smaller than the primary.

The current monitor measured the current between the cathode and ground. In all oscillographs depicting the arc current (except for a couple tests where I turned the monitor backwards as a check) positive signal represents positive current traveling from the cathode to ground. Typically, tens of amps of current would be measured during a spark.

A comparable current monitor, but with a bandwidth only up to 20 MHz, gave comparable results, except that it showed much more noise (probably caused by internal resonances above 20 MHz), and the monitor with 70 MHz bandwidth showed the signal more clearly. Two current monitors, placed back-to-back, measured the same signal (inverted for the backwards one), demonstrating (albeit simply) that we were in fact measuring the current (and not just picking up RF noise).

3.3.4 Light Output

A few crude attempts to use a photodiode to measure the light output during a spark showed that the light output lasts about as long as the arc current. Terminating the photodiode with 50 Ω to allow a long cable to reach the oscilloscope

reduced the sensitivity, but during a bright spark, the photodiode showed a small rise for roughly the duration of the current—about a half microsecond.

3.3.5 Breakdown Counter

After a spark discharged the voltage across the gap, the anode would charge back up to the original voltage within several milliseconds, well before the second or so that it took me to turn off the high voltage generator. Usually the arc destroyed whatever triggered it, making a second breakdown at the same voltage less likely, but not always. To count the number of breakdown events that occurred before the voltage could be turned off, I made a simple counting device, basically a capacitive divider, a simple discriminator, and a digital pulse counter to count the number of times that the voltage dropped by more than a kilovolt within a few microseconds. The breakdown counter agreed much of the time with the number of pulses of RF noise (another measure of how many sparks occurred), but was not always entirely reliable, though it always counted at least one when there was a voltage breakdown, and never counted more breakdowns than there were (at least, not when I could estimate the number of breakdowns by some other method).

3.4 Surface Analysis Tools

Central to this investigation were surface analyses performed before and after voltage breakdown. The relevant techniques are described briefly here (and more thoroughly in, for example [38]).

3.4.1 SEM

A scanning electron microscope (SEM) rasters a 20 kV electron beam across a surface and creates a picture of the secondary electron emission detected at each point as the beam scans across the sample (the picture is called the secondary emission image). For easy SEM work, the sample must be conducting; since analysis takes place in vacuum, the sample must be solid as well (or have very low vapor pressure).

When 20 kV electrons bombard a solid material, a small fraction of them scatter elastically off the surface ("backscattered electrons"), but most penetrate the surface, as far as 5 microns deep into the material, colliding and scattering sideways along the way. The beam electrons can knock atomic electrons out of their orbitals—these "secondary electrons" have much lower energy than the 20 kV beam electrons. Most of the secondary electrons have an energy of only several eV up to tens of eV; although these electrons are created within a 5 micron depth of the surface, only those created very near the surface can escape the sample (20 keV electrons penetrate 5 microns of solid material, but several eV electrons hardly penetrate anything). The production of secondary electrons, especially low-energy secondary electrons, is very sensitive to the electronic wavefunctions—sensitive to the elements present and sometimes even their chemical states. Since the secondary emission image reflects secondary emission from the surface only, small changes in surface composition create contrast in the image. The sample topography also creates contrast in the secondary emission image.

Practically, the size of the electron beam determines the resolution of the secondary emission image, and the beam current determines the signal to noise ratio (and larger beam currents imply larger beam sizes). A standard SEM with a tung-

sten filament easily achieves a reasonable signal to noise ratio at a resolution finer than 100 nm (field emission filaments can do better by a factor of 10).

3.4.2 EDX

EDX, energy dispersive x-ray spectroscopy,[6] can identify elements of a sample by analyzing the x-rays that emerge from the sample when excited by an electron beam, such as the electron beam in an SEM. When the 20 kV electron beam of the SEM hits, for example, a niobium surface, the electrons penetrate to a depth of several microns, scattering off niobium atoms, changing directions, exciting atomic transitions, *etc.* Even though the beam may be less than 100 nm in cross-section, scattered electrons reach about a 5 micron volume, colliding with atomic electrons all the way. At some point, an electron will knock an inner-shell electron off of a niobium atom, and an outer-shell electron will relax into the inner-shell, emitting an x-ray with a characteristic energy (approximately the binding energy of the inner-shell electron, since the outer-shell electrons are relatively free in comparison). A solid-state detector collects x-rays into energy bins according the energy of the incoming x-rays; the peaks of that x-ray spectrum reveal which elements are present. Important: the binding energy of inner-shell electrons changes very little with chemical state or "environment" (unlike the binding energy of valence electrons, which does depend on chemical state) so elements can be identified regardless of their neighbors. However, light elements (H, He, Li, Be) that hardly have an inner-shell, cannot be identified with EDX.

[6] Energy dispersive x-ray spectroscopy, as opposed to wavelength dispersive x-ray spectroscopy.

The importance and usefulness of EDX can hardly be overstated, complementing the qualitative, gray-scale secondary-emission image with semi-quantitative information about elemental composition. EDX is sensitive to elements present in 0.1–1% concentrations within the roughly 5 micron distance from where the electron beam hits the sample; EDX is almost more of a bulk analysis technique than surface analysis. This "feature" of EDX is sometimes helpful, because surfaces can be complex and confusing, but it limits the spatial resolution of the technique to a few microns; of particular interest to this work, particles smaller than one micron can be difficult to identify because they constitute only a small fraction of the entire analysis volume.

3.4.3 AES (Auger Analysis)

AES, Auger electron spectroscopy, works somewhat like EDX, but it detects electrons ejected from the sample instead of x-rays. Auger electrons are secondary electrons emitted from the sample after an electronic transaction involving an Auger transition—Auger electrons of interest have energies from about 20 eV to 2500 eV. An SEM creates a secondary emission image by collecting all secondary electrons (most of which have energies below 20 eV); AES, however, analyzes the electrons' energies to create a spectrum, from which elements can be identified. Auger electrons still have energies too low to penetrate more than a hundred angstroms; therefore, AES is much more surface-sensitive than EDX (and offers a spatial resolution limited by the primary beam size).

When the electron beam in a scanning electron microscope hits the sample, it knocks out inner-shell electrons, the vacancies of which are filled by outer electrons that decay to the inner shell, releasing energy. Sometimes this energy escapes in

the form of x-rays, which are detected by EDX. Sometimes the vacancy is filled, not by an outer electron, but by another inner-shell electron, in which case another electron can be ejected; for instance, a K shell electron can be ejected by the primary electron beam, and an L shell electron can relax into the K shell, ejecting another L shell electron (that would be called a KLL transition). The spectrum of Auger electrons (as the ejected electrons are called), usually between about 20 and 2500 eV, can be used to identify the elements present. Although Auger electrons are produced in the few-micron excitation volume penetrated by the incoming beam,[7] the relatively low energy Auger electrons can escape only from the first 10–100 Å of the surface. Therefore, AES is much more surface sensitive than EDX, and can detect, for example, a 10 Å thick surface layer that would be overwhelmed by background noise with EDX. In addition to sampling a smaller depth, the horizontal resolution of Auger analysis is limited by the primary beam size, just like the secondary emission image. AES is sensitive to elements present within the escape depth in concentrations between as low as 0.1–1%.

[7]For Auger analysis, a beam of 3–10 kV usually results in a better Auger electron yield than the 20 kV which works better for SEM images and EDX. Electron beams with lower energies have greater diameters at the same beam current; therefore at the same signal to noise ratio a lower energy beam is larger, yielding a coarser resolution.

Chapter 4

Results: Field Emission, Breakdown, and Contaminant Particles

In the search for the cause of voltage breakdown, this work complements other research that connects field emission and breakdown. Of particular interest are contaminant particles, which often seem to be at the root of both field emission and breakdown.

Field emission is the most apparent precursor to breakdown, although it is far from obvious how field emission evolves into an arc; one supposes in a general way that field emission in combination with Joule heating of the emitter or electron bombardment of the anode (or, in an RF cavity, wherever the electrons land) leads to breakdown. Leaving the details of such theories aside for chapter 7, we keep in mind that charged particles are necessary to conduct electrical current, and field-emission is by far the most familiar way for unwanted charged particles to get into

a vacuum gap under high voltage. I am unaware of any other charged particles, besides field-emitted electrons, being often observed (even indirectly) *before* an unwanted vacuum arc.

If field emission leads to breakdown, even indirectly, we need to understand the sources of field emission. The basic process of field emission, or the tunneling of electrons out of a metal through the work-function barrier under the influence of a high surface electric field (see, for example, [37, 78]), suggests that surface fields of order 5 GV/m would cause typical field emission, and experiments with finely pointed cathodes confirm the theory (for example, [33]). However, field emission from broad-area electrodes is observed at much lower *average* fields; and, experiments have established that field emission comes from point-sources on broad-area cathodes ([83] is a good review article). Such field emission sites often disappear after breakdown (for example, [16, 14]), strengthening the connection between field emission and breakdown.

Because field-emitted electrons always appear to emanate from point sources (micron-sized or smaller), we naturally wonder what is special about those points. Studies of field emission sites have frequently revealed foreign particles [80, 81, 70, 11]. Subsequently, the problem of parasitic energy dissipation in superconducting RF cavities due to field emission was solved by extremely clean (particle-free) preparation and assembly of cavities [85]. Although contaminant particles are not the only source of field emission (other surface defects can enhance field emission) they are a common and prominent source; moreover particles are easily found and identified in a scanning electron microscope. Particles therefore constitute one of the major interests of this work; particles provide a convenient focus for studying

breakdown by limiting the attention to several square microns of contaminant particle rather than square millimeters of cathode area.

Most research connecting particles and field emission, and field emission and breakdown, has been done with DC voltages, but there is also some evidence from RF cavities showing field emission at breakdown sites prior to breakdown [77, 39]. Also, traces of foreign elements have been found at breakdown sites in RF cavities, indicating that a contaminant particle may have triggered the breakdown [44]. We have further strengthened the connection between particles and breakdown sites, and have observed as well that cathodes with more particles tend to exhibit more field emission.

4.1 Observations of Field Emission

In most of the experiments, I recorded typical field emission currents as I raised the voltage. Fowler-Nordheim [37] field emission (tunneling through the work function barrier) emits a current density

$$J_{\mathrm{FN}}(E_s) = J_0 \frac{E_s^2}{E_0^2} \exp(-E_0/E_s) \tag{4.1}$$

where E_s is the surface electric field at the emitter, and J_0 and E_0 are constants that depend on the details of the barrier (like the barrier height, or work function), which can be calculated from characteristics of the material measured by means unrelated to field emission. For niobium, the work function is about 4 eV, and $J_0 = 1.15 \times 10^{15}$ A/m^2 $= 1.15$ kA/μm^2, and $E_0 = 5.46 \times 10^{10}$ V/m $= 54.6$ GV/m [87].

In practice we measure the total current $I = J_{\mathrm{FN}} S$ where S is the surface area of the emitter. For cathodes with macroscopic area, field emission sources are

points, so the area S is not generally known beforehand. Although we can easily calculate the macroscopic surface field E ($E = V/d$), we do not necessarily know the local surface field E_s at the emitter. In many cases, such as particulate field emitters, the local field will be enhanced by some factor β: $E_s = \beta E$. β is also not known beforehand. We have:

$$I = J_{\mathrm{FN}} S = J_0 S \frac{(\beta E)^2}{E_0^2} \exp\left(-\frac{E_0}{\beta E}\right) \tag{4.2}$$

or

$$\ln \frac{I}{I_0 (E/E_0)^2} = \ln \frac{\beta^2 S}{S_0} - \frac{E_0}{\beta E} \tag{4.3}$$

where I have written $I_0 = J_0 S_0$ for some arbitrary reference area S_0 (only because I prefer my logarithms to contain dimensionless quantities). Graphing the left-hand side versus $1/E$, "the Fowler-Nordheim curve," should result in a line with slope $-E_0/\beta$, from which β can be measured, and intercept $\ln \beta^2 S/S_0$, from which S can be measured.

A cautionary note: lest straight Fowler-Nordheim curves should appear too convincing, let me point out that it's rather hard to be precise about comparisons between different exponential-like behaviors. Fowler and Nordheim, in their paper [37], say about field emission: "The formula for these currents is $I = Ce^{-a/F}$, which is, of course, indistinguishable from $I = CF^2 e^{-a/F}$" (they use F for electric field). In fact, Fowler-Nordheim field emission usually looks pretty good fitted as $I = Ce^{aF}$. The problem is that with such steep dependence on electric field, any Fowler-Nordheim plot extends over a relatively small range of electric field, while the current runs through 6 or more orders of magnitude. Fowler-Nordheim plots appear throughout the literature, as well as in this work; they should be for the most part considered simply a standard presentation of data (for ease of

comparison), rather than a confirmation that the electron emission actually arises from a pure Fowler-Nordheim mechanism.

Although the field enhancement β and emitter area S originally had very clear interpretations as such, measurements of actual field emitters belie these interpretations. Figure 4.1 shows β and S measured for many different field emitters using a scanning point anode [41]. Our own field emission measurements were much coarser and less extensive—with only a large anode, we could not measure individual emitters—however, we still see the Fowler-Nordheim parameters falling over a very wide range (figure 4.2). β and S should be considered to be heuristic, empirical parameters, helpful for characterizing field emission, but so far without a completely understood physical meaning.

In the graph of β vs. S, I've used only results that yielded a fairly straight line on the Fowler-Nordheim plot; many field emission measurements looked nothing like Fowler-Nordheim emission. Emitters switched on and off abruptly, sometimes emission current seemed to saturate at high levels, hardly increasing at all with voltage, and field emission current usually was extremely noisy, making measurement difficult. In general, experiments that show reproducible, "good" Fowler-Nordheim field emission involve heavily-processed electrodes (baked to high temperatures and subjected to many breakdown events) or needle cathodes tipped by a single crystal point (often tungsten).

4.1.1 Evidence of Gas Desorption with Field Emission

Because the vacuum arc depends on a source of vapor, any signs of pre-breakdown vapor production merit mention. In some tests the background vacuum pressure

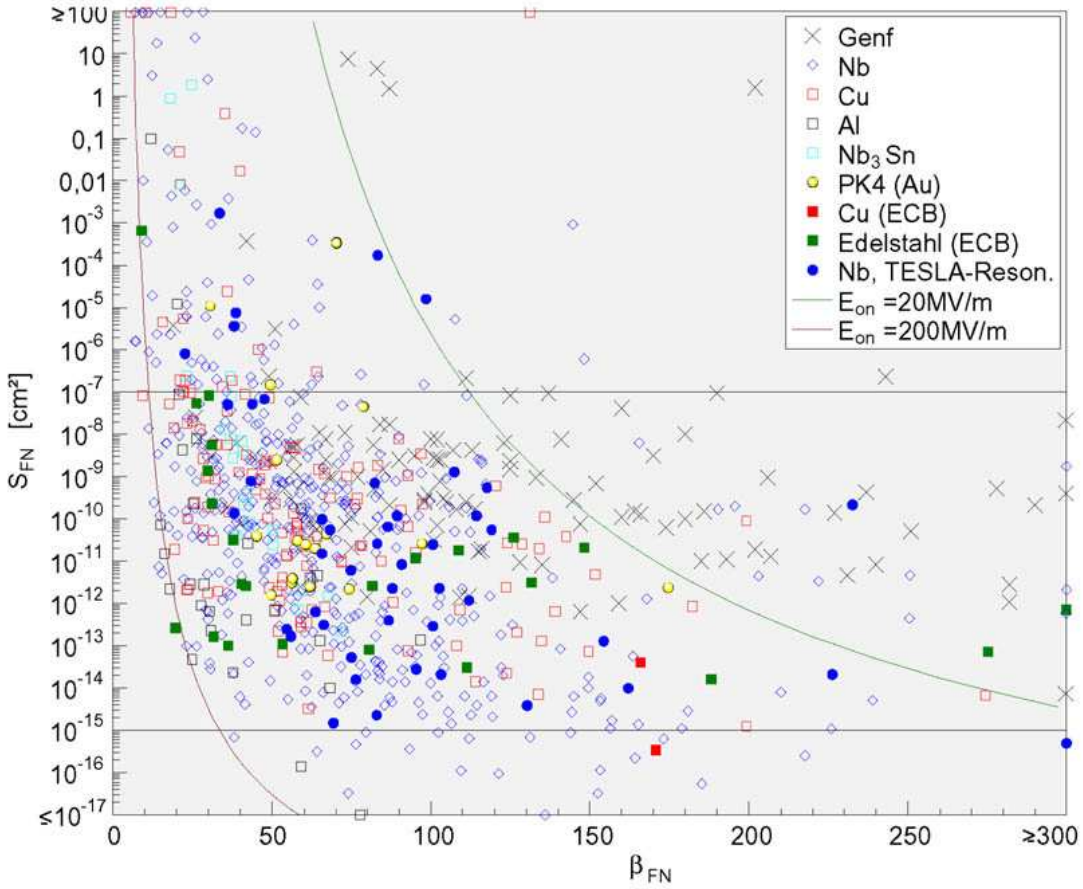
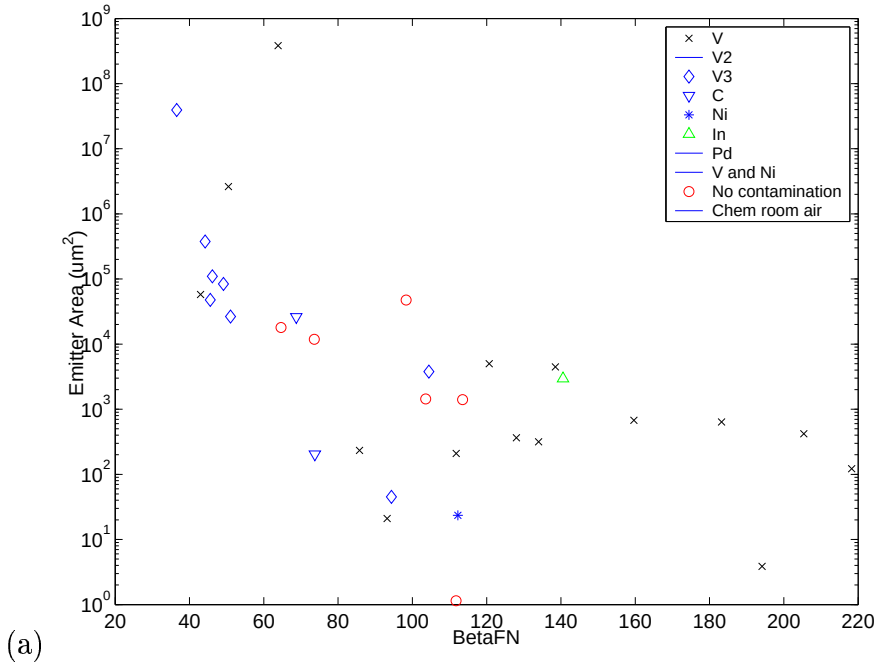
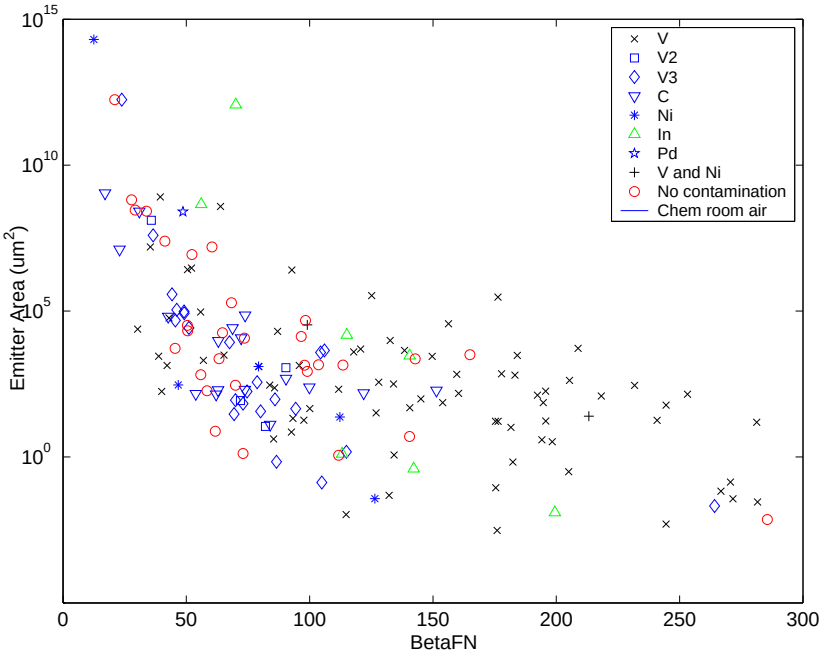


Figure 4.1: β and S parameters for various individual field emitters, (most of them measured by and) collected by Thomas Habermann, from whose thesis [41] this graph is reproduced. The caption reads approximately: "Relationship between the FN parameters β_{FN} and S_{FN} for 696 emitters analyzed in Wuppertal and 83 analyzed in Genf." Presumably these last 83 were analyzed by Niedermann in his dissertation [82].



(a)



(b)

Figure 4.2: β and S parameters for field emission measured during these experiments. The symbol for each data point indicates probable emitter composition. (a) Only tests with 5 or more (current vs. voltage) data points, and correlation coefficients squared with values greater than 0.85—in other words, only cases where the Fowler-Nordheim curve is a reasonable straight line. (b) Almost all experiments with $\beta < 300$, just for show—many of these do not have many data points or are not at all well fit by the Fowler-Nordheim model.

increased slightly with field emission,[1] with the pressure clearly increasing and decreasing with electron current (see figure 2.8 on page 38).

From the ion pump specifications we can make a (probably low) estimate of the number of gas molecules that need to be desorbed from the surface to produce a pressure rise of 1×10^{-8} torr at the ion pump (see appendix A); assuming 10 atoms/nm^2 per monolayer, we then estimate the number of desorbed monolayers that would produce a 1×10^{-8} torr pressure rise, assuming a certain area over which desorption occurs (table 4.1).

Table 4.1: Outgassing rates, in monolayers per second, for different areas, corresponding to the pumping rate of nitrogen (or, roughly, water, oxygen, carbon monoxide, carbon dioxide) for a 1×10^{-8} torr pressure rise caused by field emission. A pumping speed of 10 L/s is assumed.

Outgassing Area	Monolayers (N_2, H_2O, O_2, CO, CO_2)/s Outgassed
1 μm^2	3×10^5
10^4 μm^2	30
1 mm^2	0.3
10 mm^2	0.03

Considering that such pressure rises lasted for minutes, the desorption must come from a relatively large area—certainly not a micron-sized cathode protrusion (which would be worn down at the rate of 3×10^5 monolayers per second, or 100 microns per second). Since evaporation of bulk electrode material is unlikely to

[1] In earlier experiments with this apparatus (conducted by Dave Moffat), a residual gas analyzer was used to determine the species of gas released during field emission; typically water, CO (or maybe N_2), and CO_2 were most prominent.

cause the observed pressure rise[2] the area of desorption must be large enough that over several minutes only a few monolayers are released; therefore, the area of desorption must be on the order of 10 mm^2 or larger (for a desorption rate less than 2 monolayers per minute). Such a large area favors the anode as the source of gas, since electrons bombard the anode in a relatively large area compared, whereas electron emission could heat the cathode only over very small areas. The energy required to desorb the gas also favors the anode; because of the large voltage difference, the total heating at the anode is much larger than at the cathode.

Desorption of gas over a wide area of the anode due to electron bombardment therefore seems the likely explanation of the correspondence between field emission and vacuum pressure. Some notable examples of pressure increase due to field emission did in fact occur immediately after installing a new anode in the system (which apparently needed to be outgassed).

4.2 Correlations Between Particles and Breakdown Sites

Following up on studies that often revealed particles at field emission sites, and studies that suggested breakdown occurred at field emission sites, we investigated whether breakdown would occur at contaminant particles on the cathode. Examining a cathode (often intentionally contaminated with choice particles) in a scanning electron microscope, we documented particle sites before testing the cathode. The

[2] After a few monolayers comes the bulk electrode material; evaporated electrode material (niobium, copper, or tantalum vapor) would stick to the chamber walls, not even reaching the ion pump. If the pressure rise were due to evaporation of bulk material, the evaporation rate would have to be much larger than calculated above, since the apparatus walls would efficiently pump most of the metal vapor.

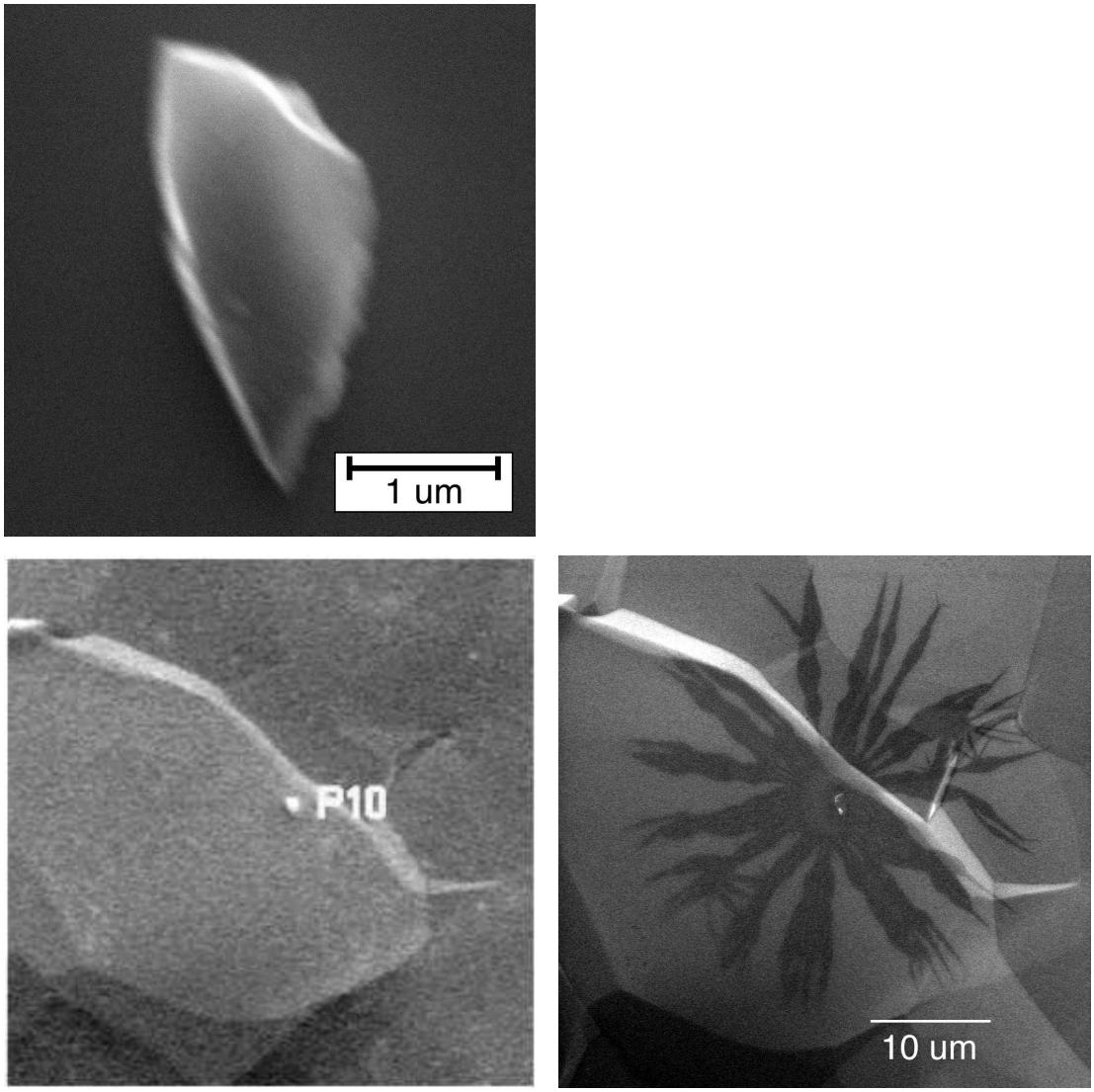


Figure 4.3: A contaminant particle (carbon) before breakdown (left, magnified view above), and the starburst after breakdown (right), centered exactly where the particle was.

process of breakdown leaves a fairly obvious trace (to an electron microscope) on the cathode, usually localized to a region of 100 microns or less. These regions we call "starbursts" (see figure 2.13 on page 42), originally named after the rather distinct shape assumed by some of them, but we have since come to use "starburst" as a general term to mean "the localized areas of the cathode affected by breakdown." The starbursts tell us where an arc occurred on the cathode.

We have often observed that contaminant particles play a central role in voltage breakdown; that is, particles often determine where the center of a starburst will be. The "before" and "after" pictures tell the story (see figures 4.3 and 4.4). Most of the particles that apparently led to breakdown were particles that we intentionally deposited on the cathode, but we occasionally identified an adventitious particle that later became the center of a starburst (see figure 4.9 on page 87).

Without the "before" pictures, we could not have been sure that a particle had been at the breakdown site; the arc destroys all but a trace of the particle. As in earlier studies on niobium RF cavities [44], EDX, always analyzing a volume of at least a few microns in size, could not detect any foreign elements, but the greater (surface) sensitivity of AES usually found some remains of the original particle, which we had identified before the breakdown event (figures 4.4 and 4.5).

To drive home the correlation between particles, arcing, and starbursts, we compare the light from the arc with the starburst locations and the particles that preceded them; figure 4.6 shows one of the most clearly correlated cases.

4.2.1 Breakdown on Clean Cathodes

Clean cathodes can withstand higher fields than cathodes with particles; that is, cathodes with particulate contamination break down at lower fields. Because of the

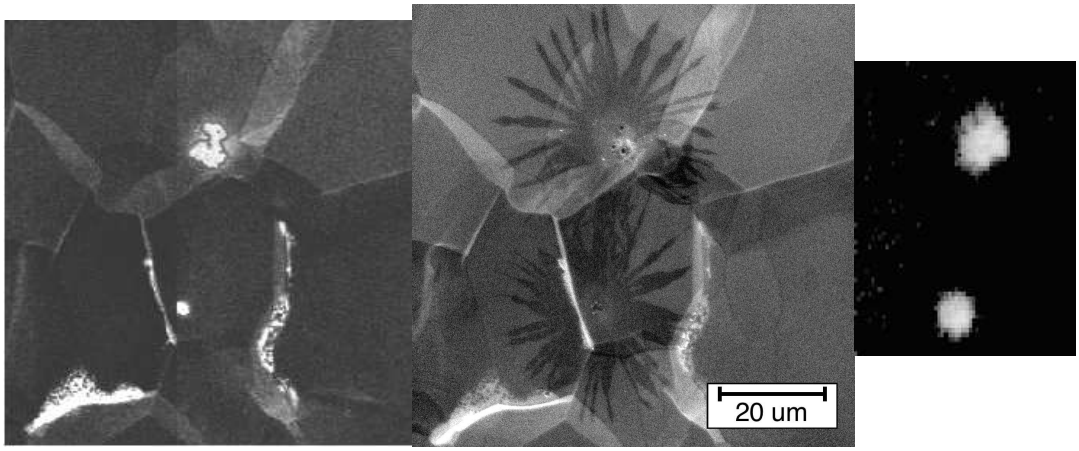


Figure 4.4: Nickel particles before breakdown (left), and the starbursts centered on the particles (middle), and the particles' trace remains, visible in an Auger map of Ni (right).

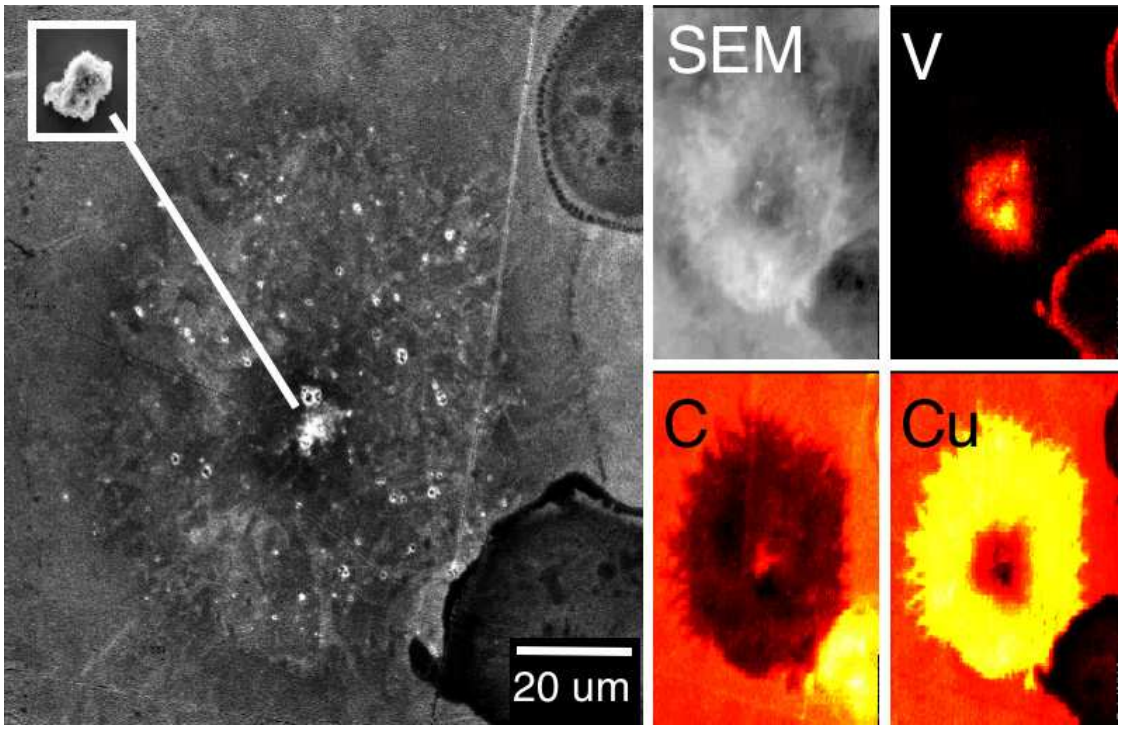


Figure 4.5: A starburst on an electropolished copper cathode (left) that occurred at the site of a V particle (left inset), and Auger maps of V, C, and Cu (right) showing trace remains of the particle as well as surface carbon depletion. Note the large number of satellite craters in the starburst.

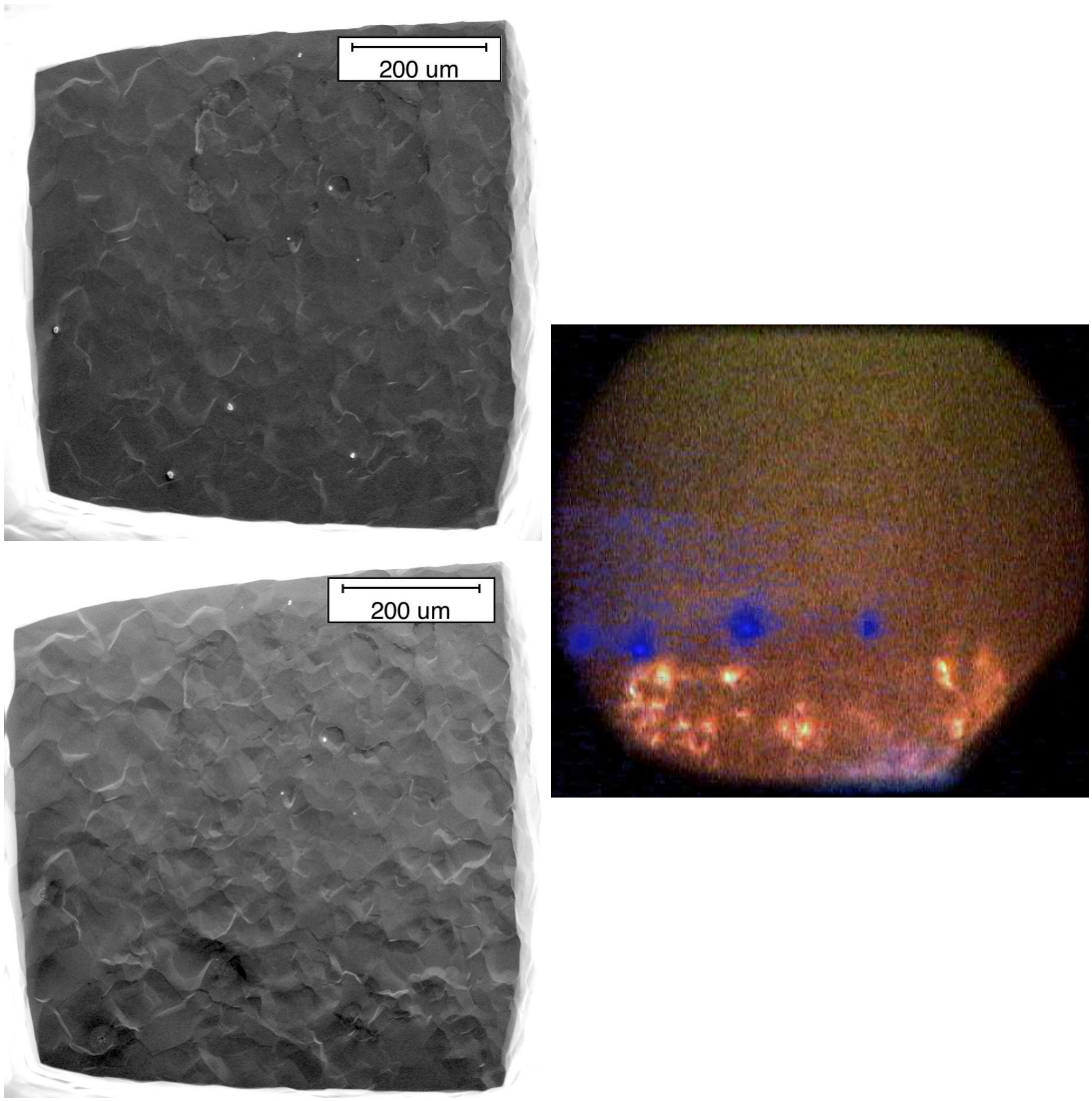


Figure 4.6: Cathode plate B8, pedestal 4, anodized niobium contaminated with vanadium particles, shown before breakdown (top left), and after (bottom left), along with (right) the video frame during which breakdown occurred. Each starburst clearly corresponds with one of the little balls of light in the video; of the 4 starbursts, 3 were formed at contaminant particle sites—in the center of the other one, vanadium was found, suggesting that during the arc, one of the other 3 particles melted and a droplet splashed onto another part of the cathode and subsequently caused another starburst to form. These "starbursts" are unusual due to the unnaturally large oxide thickness on this cathode (as discussed in section 5.5).

variety of parameters and procedures during the series of experiments (especially in the gap size) comparing breakdown fields requires care; to make a fair comparison I charted the maximum field reached during the first "voltage-raising" at cathode sites tested at a 150 micron gap,[3] including cathodes of all types: niobium, copper, and gold, as well as oxidized niobium and oxidized copper (see section 4.2.3 for a comparison of the different cathode types). At a 150 micron gap, the maximum voltage of 14.5 kV produced a field around 95 MV/m. Most cathode sites contaminated with particles broke down below the maximum field (figure 4.7); on the other hand, most uncontaminated cathode sites did not break down at 95MV/m (figure 4.8). Excluding the two sites for which tests were halted prematurely (as explained in the caption of figure 4.8), 40 of 52 contaminated sites broke down, while only 1 of 16 uncontaminated sites broke down at or below the maximum field. Note that cathodes can reach fields higher than 100 MV/m, but, observing the maximum voltage, only with a gap smaller than 150 microns (several cathode pedestals withstood fields around 150 MV/m).

Based on the breakdown likelihood of clean versus contaminated cathodes, and the frequency with which starbursts appeared at particle sites, I have concluded that particles cause breakdown. I would now like to qualify that statement. Breakdown occurs at particles *when there are particles*—for our experiments, "there are

[3]Some tests satisfying these criteria were discarded. First, as usual, all "bad" tests were ignored. Second, tests not at room temperature were discarded. Last, two cathodes were discarded because they had small pedestals with extremely sharp corners, and another cathode discarded because it was a sort of misfit (it was not intentionally contaminated but it was severely accidentally contaminated and broke down at low fields, so it supports the conclusion but doesn't fit nicely into a category). "Bad" tests are tests that suffered some malfunction that cast doubt on the results; most "bad" tests were so designated because the anode and cathode accidentally touched while I tried to estimate the gap size.

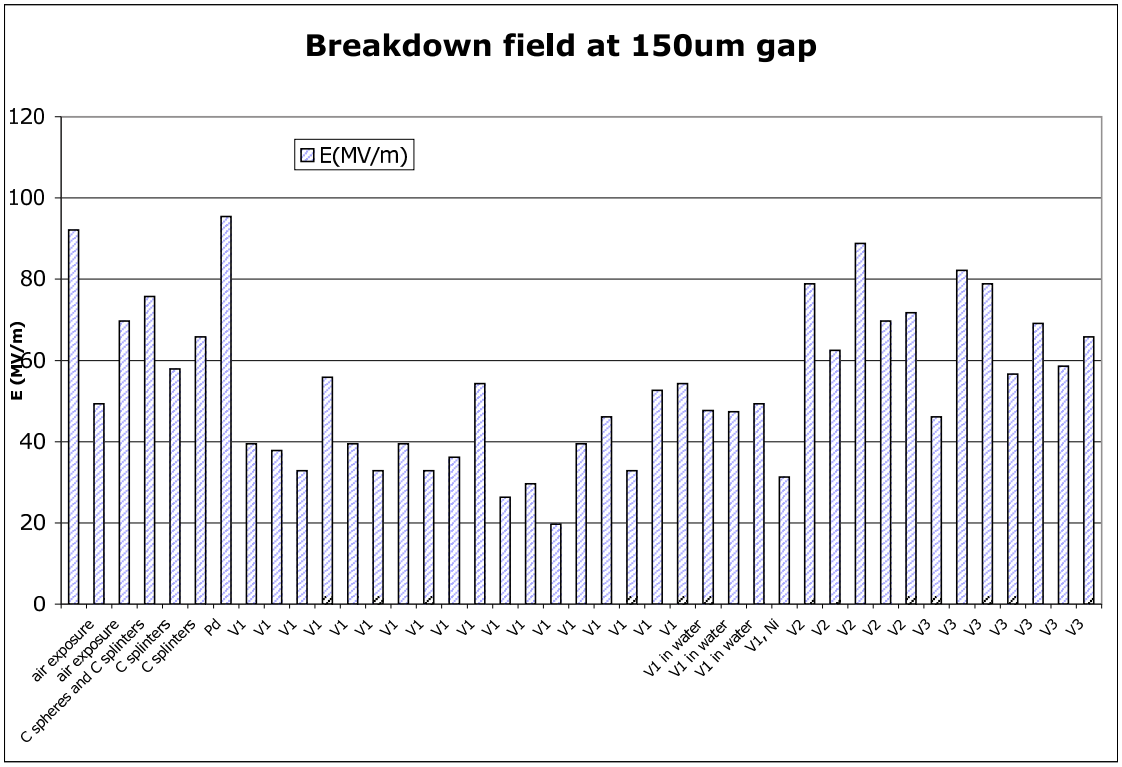


Figure 4.7: The breakdown fields and contamination material (if any) for (almost) all cathode sites that were first tested at a 150 micron gap and suffered breakdown (see figure 4.8 for those that didn't break down); each bar represents a single cathode site, labeled by the material with which they were contaminated. Unlabeled are sites that were not intentionally contaminated (though other sites on the same plate may have been intentionally contaminated); "V1" is exceptionally spiny vanadium; "V2" and "V3" are not-so-spiny vanadium, from different suppliers; "air exposure" labels sites on a cathode exposed to unfiltered air for a few hours; most of the particles were placed on the cathodes in a drop of methanol, or water for sites labeled "V1 in water."

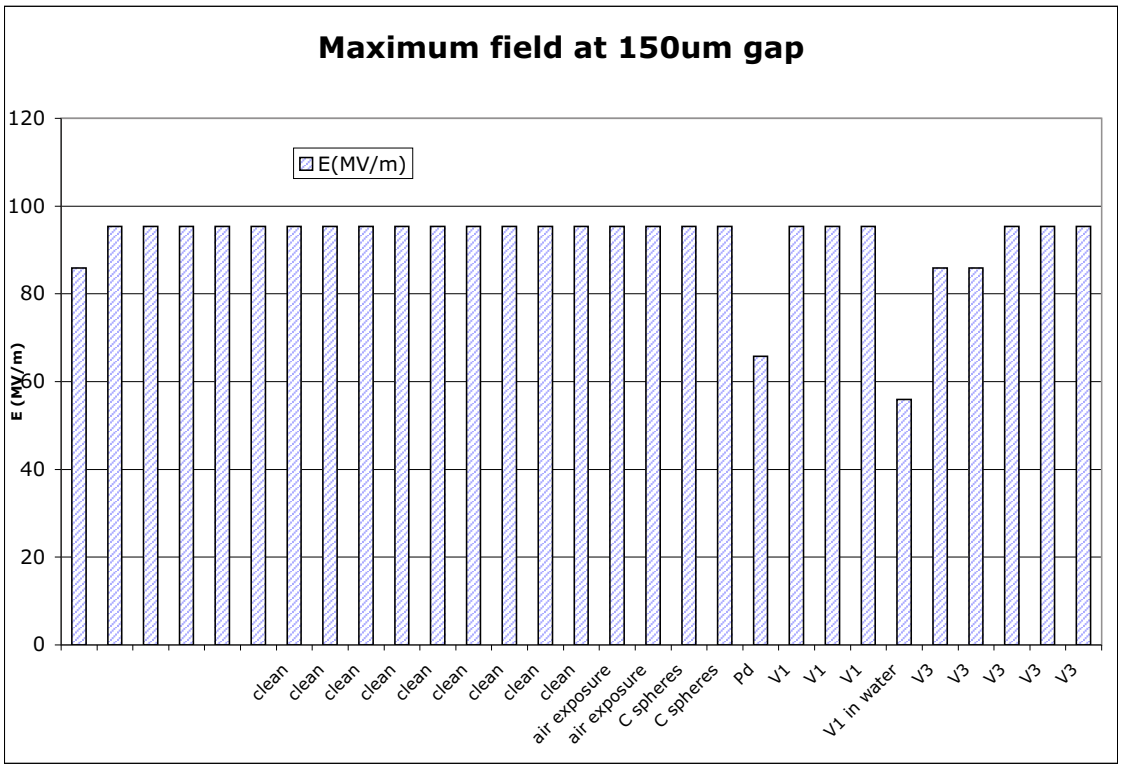


Figure 4.8: The maximum fields reached (for a 150 micron gap) and the contamination material (if any) for (almost) all cathode sites that were first tested at a 150 micron gap. See figure 4.7 for an explanation of contamination materials; sites labeled "clean" were uncontaminated, and moreover were on cathodes that we tried hard to keep as clean as possible—it does not mean there were no particles, but it does mean there were few particles. Naturally, most sites that didn't break down at a 150 micron gap reached the maximum field of 95MV/m; the three cases just below 90MV/m were tested with a pointier anode that reduced the maximum field slightly; the "Pd" test I ended early because the field emission current was unusually high and I wanted to see if the field emission alone could cause any visible changes (it didn't); the "V1 in water" test ended at a low field due to equipment malfunction.

particles" translates into "the particle density is several per square millimeter or greater."

To be sure, we have seen many starbursts at locations where we had not previously identified a contaminant particle. In some cases we could identify remains of some contaminant in the starburst center; for example, figure 4.6 shows a starburst with residual vanadium contamination, although no vanadium particle had previously been seen at that site—however, it's next to three other starbursts caused by vanadium particles, suggesting that a droplet from one of those three particles landed on the cathode where the fourth starburst occurred. While not common, droplets of material ejected in a molten state from either the cathode (or contaminant particles) or anode land on the cathode with fair frequency (*e.g.,* figure 4.9).

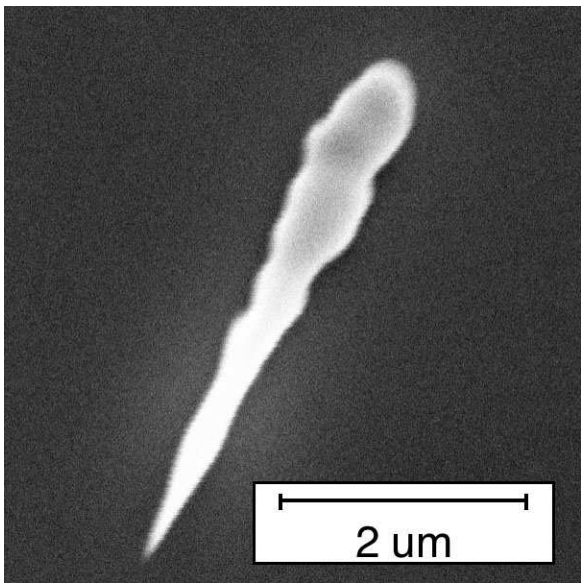


Figure 4.9: A droplet of anode material (mostly Ta) that landed on the cathode during a breakdown event (centered relatively far away from the droplet). The cathode was subsequently re-tested, and a starburst formed at this site.

Attempts to determine that a breakdown event or a starburst that was *definitely* not caused by a particle are undermined by the possibility of breakdown caused

by particles too small to see (less than the resolution of the microscope) and by the possibility of particles contaminating the cathode after examination and then causing breakdown.

Only higher-resolution microscopy can confirm or deny the possibility that currently-too-small-to-see particles cause breakdown. The best scanning electron microscopes approach nanometer resolution, which is about as small as particles (in this context) can get; however, scanning a large surface area (even a square millimeter) with such fine resolution would be a sizable task.

Working in a clean room reduced the rate of contamination between the microscope and apparatus, but still we occasionally found new particles on a cathode during the post-test examination (particles that apparently survived breakdown). Examining the breakdown site for foreign elements can reveal whether a contaminant originally triggered the event, although again, small contaminants could be a problem: if a particle a few microns in diameter leaves a trace too small to be seen with EDX but just detectable with AES, perhaps a particles ten times smaller would leave traces undetectable by AES. In this case, a more sensitive surface analysis technique, like SIMS (secondary ion mass spectroscopy) might prove necessary.

In a few instances apparently clean surfaces were subjected to breakdown. However, because the surfaces were clean, the breakdown fields were high (around 150 MV/m), and the resulting arcs very violent.[4] The mass of melting on the cathode

[4]The violence of an arc, measured by the extent of the surface damage on the cathode, tends to increase with increasing field and voltage, and decreasing gap size. We have too few comparable experiments over a too limited range to separate the effects of these three quantities. For instance, to reach high fields (say, 140 MV/m or higher) required small gaps and voltages at the high end of our range (for instance, 14 kV and 100 microns). Aside from a few experiments, we rarely

(figure 4.10) was sometimes too messy to determine the origin of activity (unlike starbursts which often indicate the origin within a few microns), but usually off to the side of the main breakdown site smaller regions of activity were visible with no apparent precursors.

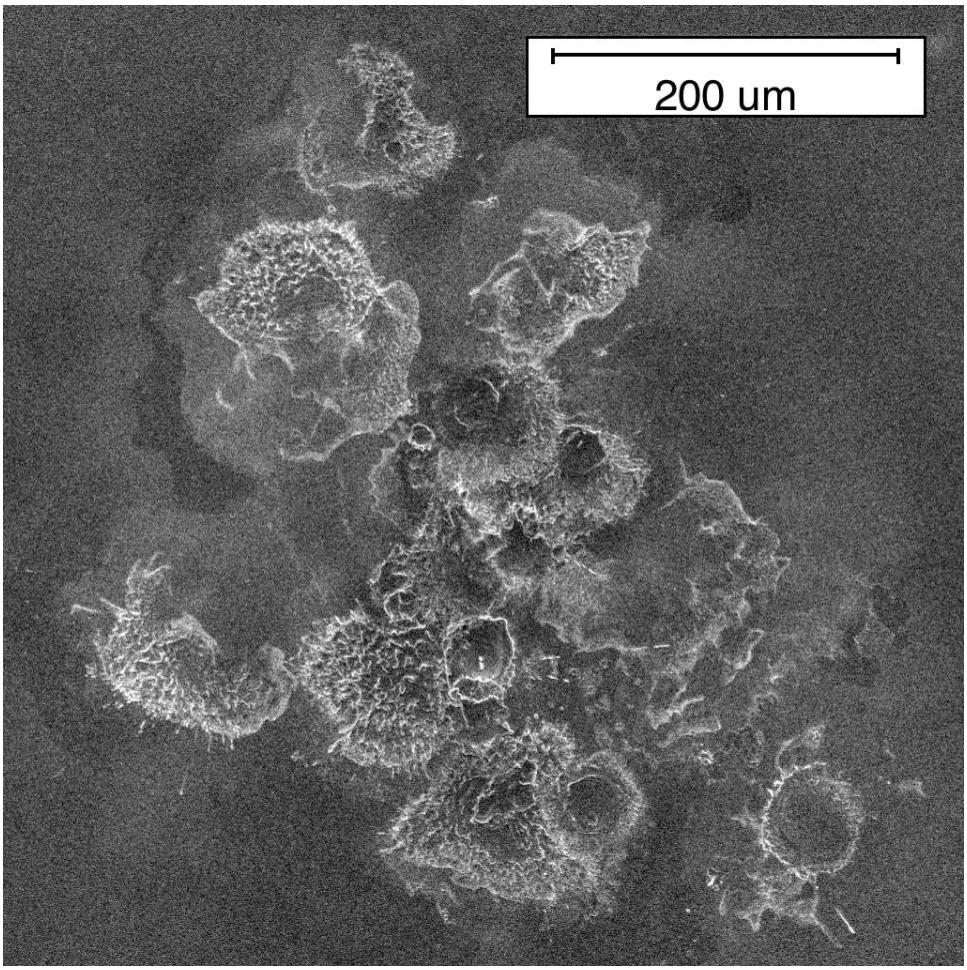


Figure 4.10: The result of breakdown (at about 140 MV/m) on a particle-free area of a diamond machined copper surface.

looked at low-field breakdown at small gaps because of the difficulty in estimating the gas sizes.

4.2.2 Breakdown Propensity and Particle Species

Early in this study, we wanted to find out what kinds of particles lead to breakdown at low electric fields; we contaminated cathodes with various powders of tiny particles, and hoped to associate a breakdown field with each kind of particle. Instead, we found that "the breakdown field" for any kind of particle ranged widely (figure 4.11). Studies investigating field emission due to particles have had similar difficulties identifying properties that make a particle a better or worse field emitter (it is even unclear how insulating or conducting properties affect field emission [80, 49], although in some cases insulating particles seem to field-emit less [48]).

The one strong conclusion regarding particle properties and both field emission and breakdown is that: geometry matters. When experiments compare particles equivalent except in surface roughness, the rougher particles tend to field-emit and break down at lower fields. In our experiments, we found vanadium powder from one source to be extremely spiny, while that from another source was much less rough (figure 4.12); cathodes contaminated with the spiny vanadium consistently broke down at lower fields than cathodes contaminated with the smoother vanadium (figure 4.11—V1 is spiny, V2 is less rough). Field emission studies comparing rough and smooth variants of the same kind of particle similarly find that the rougher particles emit more at lower fields [49].

The cause of breakdown may well be linked to the cause of field emission; certainly both have a common origin in particles. We have seen ample evidence that particles can and often do field-emit, and also that particles can and often do lead to breakdown (at relatively low fields). However, not all particles are gushing field-emitters [11, 82], and we have seen particles withstand quite high fields (well in excess of 100 MV/m) without breakdown. It may be that only a certain fraction

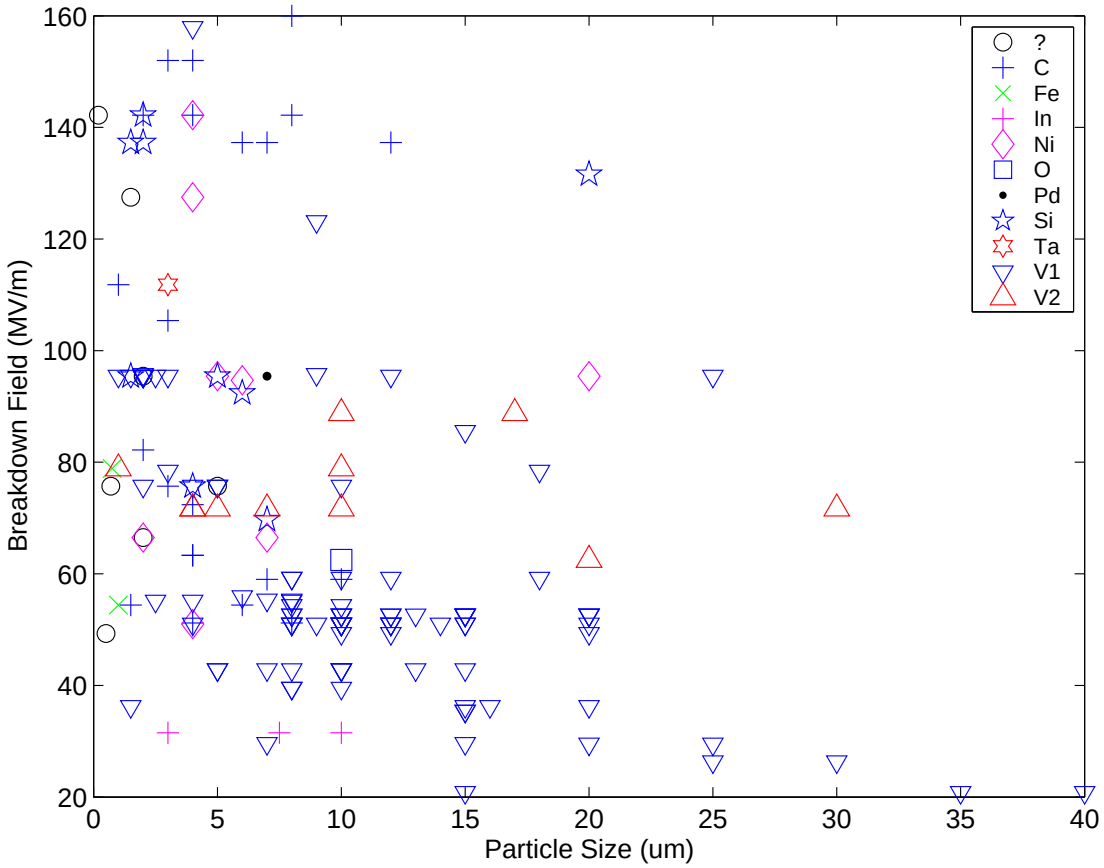


Figure 4.11: Breakdown fields versus particle size (as measured in pre-test examination) for different kinds of particles. (Many different gap spacings are included, although 150 microns was by far the most common gap.) V1 and V2 are vanadium from different sources; Large circles ("?" in the legend) mark unidentifiable particles. It's important to note that a single breakdown often destroyed several particles, though only one of them initiated the breakdown event, so this graph must be viewed cautiously.

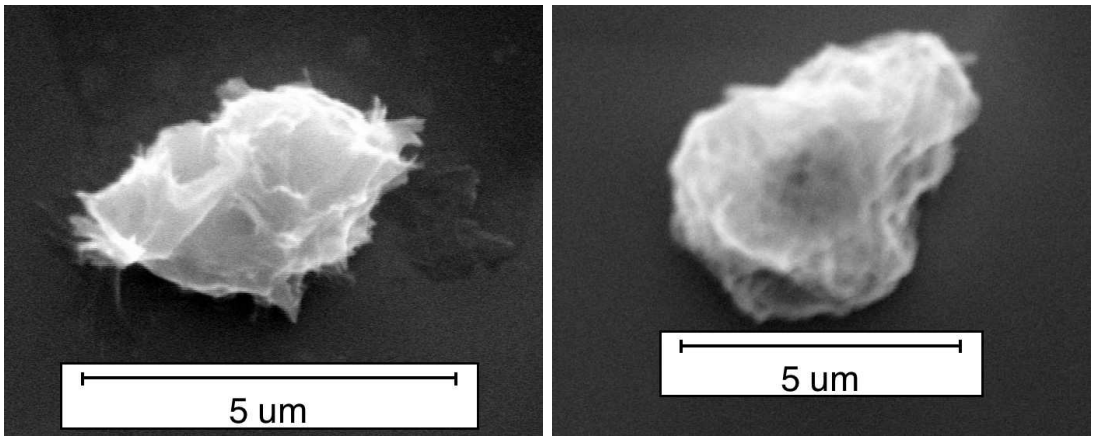


Figure 4.12: Spiny V (left) and (relatively) smooth V (right); cathodes with spiny V consistently broke down at lower fields than cathodes with smooth V.

of particles are prone to breakdown, just as only a certain fraction of particles seem to be dangerous field emitters.

4.2.3 Contaminants Determine Breakdown Regardless of Cathode

An important caution needs to be given to those who would determine *the* breakdown voltage of a certain electrode material; if there are contaminant particles, then they, and not the substrate material, determine the breakdown voltage.

We tested cathodes of several different materials: solid Nb, solid Cu, Cu film on Nb, Au film on Nb, Nb oxide on Nb, and Cu oxide on Cu. The Cu and Au films were about 1000 Å thick; the oxide films were hundreds of angstroms thick. To check for an effect of the cathode material on breakdown, we compare the breakdown fields for tests conducted at a 150 micron gap on cathodes contaminated with spiny vanadium. Figure 4.13 shows that the standard deviation for each material encompasses the average of all breakdown events taken together. We conclude

that the underlying cathode material plays a minor role compared to that of the contaminant particles.

4.2.4 Melting Without Breakdown and the Lack Thereof

We rarely observed any small changes in particles—either particles were unchanged after testing or they were obliterated during breakdown. On a number of occasions I stopped a test short of breakdown when the field emission current seemed unusually high, to see whether field emission alone might change particle morphology (by melting due to Joule heating of a field-emitting protrusion, for instance), but in the absence of breakdown, particles were unchanged (except for particles that were removed by the electric field without causing any breakdown; such particles left no trace). If melting due to Joule heating of a field emitter triggers breakdown, the onset of melting must lead inevitably to breakdown, because we (almost) never see melted particles without breakdown.

4.2.5 Particles Removed by Electric Field

In post-test examination of a cathode, we noticed that fairly frequently we would find particles missing without a trace; the electric field (during the test) probably ripped these particles off the surface. (Similar events had been observed, for example, in [71].)

In a few tests on very smooth cathode surfaces (diamond-machined copper), contaminant particles were large enough to be seen during the test through the eye of the Questar near-focus telescope, and the background light was left on during the test so the video camera could watch the particles.

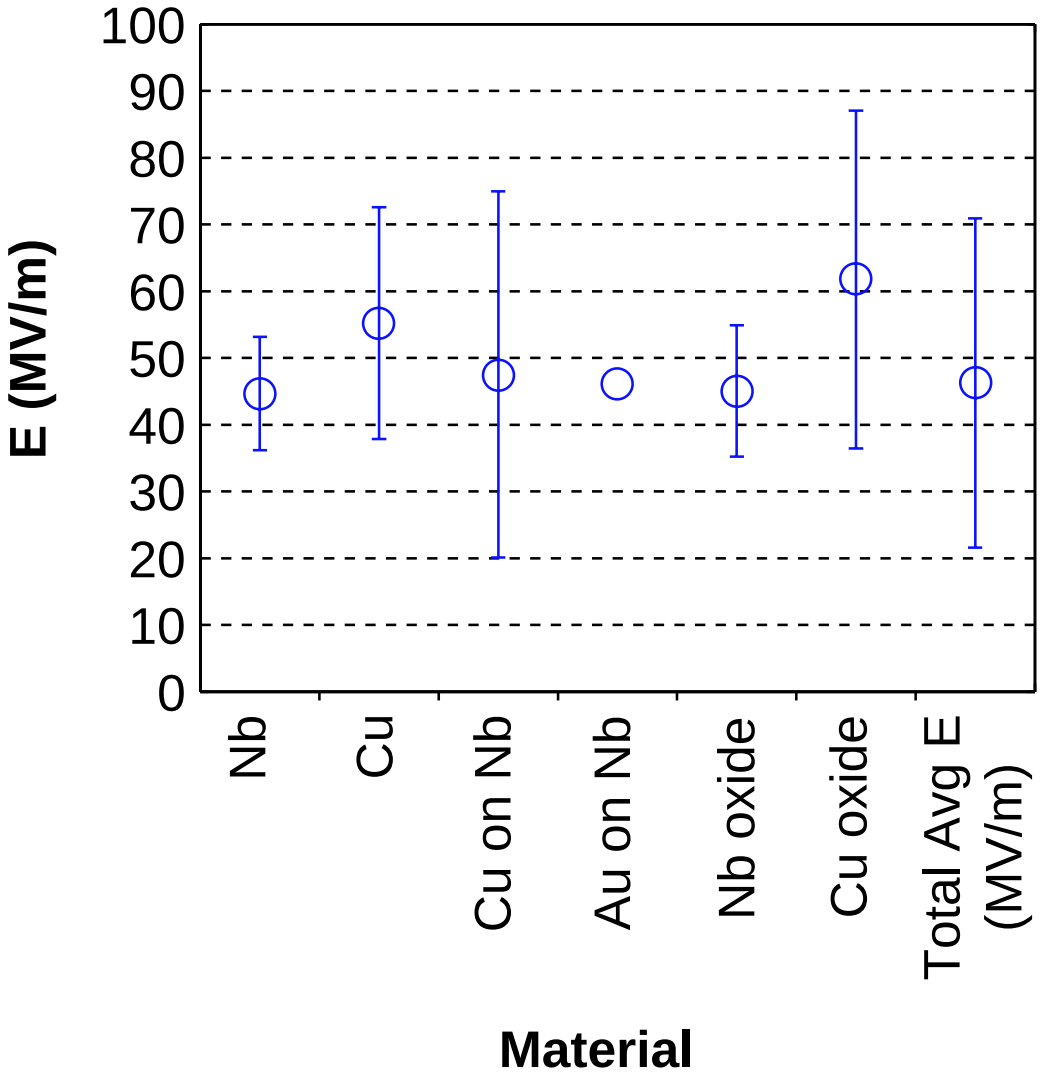


Figure 4.13: The average breakdown fields for all good tests performed at 150 micron gaps on cathodes contaminated with spiny vanadium, taken together as well as separated by material. The "error" bars indicate the range of breakdown fields, showing ± one standard deviation of the samples in each category from the category average (except for Au on Nb, since only one test satisfied the conditions). The materials are solid Nb, solid Cu (prepared in different ways, including electropolishing and diamond machining), Cu film on solid Nb, Au film on solid niobium, Nb oxide on Nb, and Cu oxide on Cu.

As long as no electric field was applied, the particles remained firmly attached to the cathode; as the electric field increased, some particles disappeared from the cathode (different particles at different times) without a hint of voltage breakdown. Sometimes particles disappeared when there was field emission, and sometimes particles disappeared when there was no field emission current (above a nano-amp). In at least one case, the field emission dropped drastically after a particle disappeared, indicating that the particle had been the major field emitter, but other times particles left the cathode without noticeably affecting field emission.

A simple explanation, not contradicted by any of this evidence, is that the force exerted on a conducting particle by the electric field overcomes the force adhering the particle to the substrate. Naturally, some particles aren't attached as firmly as others, and these will be pulled from the surface at lower fields. In this scenario, particle removal is completely unrelated to field emission and voltage breakdown.

An electric field exerts a force on (the surface charge of) a conductor, pulling the conductor toward the electric-field region. An electric field E induces (or, alternatively, is created by) a surface charge density $\sigma = \epsilon_0 E$ (where positive charge corresponds to a field pointing away from the conductor); the field exerts a pressure $\sigma E = \epsilon_0 E^2$ on the surface charge and hence the conductor. A surface field of 50 MV/m, for instance, exerts a pressure of 2.2×10^4 Pa on the conducting surface, which is of the right order of magnitude to remove micron-sized and larger particles [101].

4.3 Correlation Between Field Emission and Breakdown

Previous research strongly suggested that breakdown occurred at field emission sites. Most of the evidence supporting this connection came in the form of processing: a breakdown event would drastically reduce the field emission, leading to the conclusion that the arc had destroyed the field emitter. Some evidence from RF cavities was more direct, locating the source of field emission, and subsequently finding a starburst and melting at that location [55].

Connecting with evidence of particles at field-emission sites, we have found particles on the cathode to be preferred sites for arcing as well. The connection between field-emission and breakdown is undeniable, but the causality of the connection is unproven. To explore ways in which field emission might cause breakdown, we measured field emission current before breakdown.

Measuring field emission just before breakdown is somewhat tricky. First, we really want to know the field emission current in the nanoseconds before breakdown—breakdown itself (the rise of current to tens of amps) can happen in less than ten nanoseconds; instead, we must be content measuring the current seconds (and sometimes microseconds) before breakdown. Second, any current threshold could depend on specific details (shape, size, material) of the emitter. One might, for instance, expect a threshold in current density, not total current, but all we measure is the current. Happily, we can extract the emitter area from a Fowler-Nordheim plot; unhappily, that area seems to be more an empirical parameter, rather than the actual emitter area (see section 4.1).

4.3.1 Measuring Field Emission Just Before Breakdown

We measured field emission current with a simple logarithmic current-to-voltage converter described in section 3.2.2; the response time of the converter was on the order of a microsecond. Depending on the circumstances (the state of our RF noise shielding, the number of available oscilloscope inputs, *etc.*) we measured the output of the converter in different ways. The most relevant measurements were made with an oscilloscope, capturing the field emission current in mere microseconds before a breakdown event. Free oscilloscope inputs tended to be scarce, however, and in many cases we have to rely on my last field emission current entry in the logbook, which may well have been 10 or 20 seconds before the spark. I did use a digital volt meter with the capability to store the maximum reading, so we do have that to consider—however, that includes the spark current as well (and I don't really know what the high-frequency response of the maximum-measuring circuit is). In many cases, the maximum current agreed with last recorded current in the logbook. In later tests we used a chart recorder to record field emission current on a scale of 1 cm/min. With that we could read the current just seconds before breakdown.

Figure 4.14 shows field emission current measured "just" before a breakdown event for tests in which several methods of measurement were used, the definition of "just" depending on the method (oscilloscope: just=microseconds; chart recorder: just=seconds; logbook entry: just=tens of seconds; maximum record: just=uncertain time). Much of the time, all methods agree; the maximum record sometimes comes nowhere near to the others, in which case I assume the actual arc current outweighs the previous field emission. The graph shows that pre-breakdown field emission covered the range over which we could measure, up to about 100 μA. Notable are the tests for which no field emission was observed before

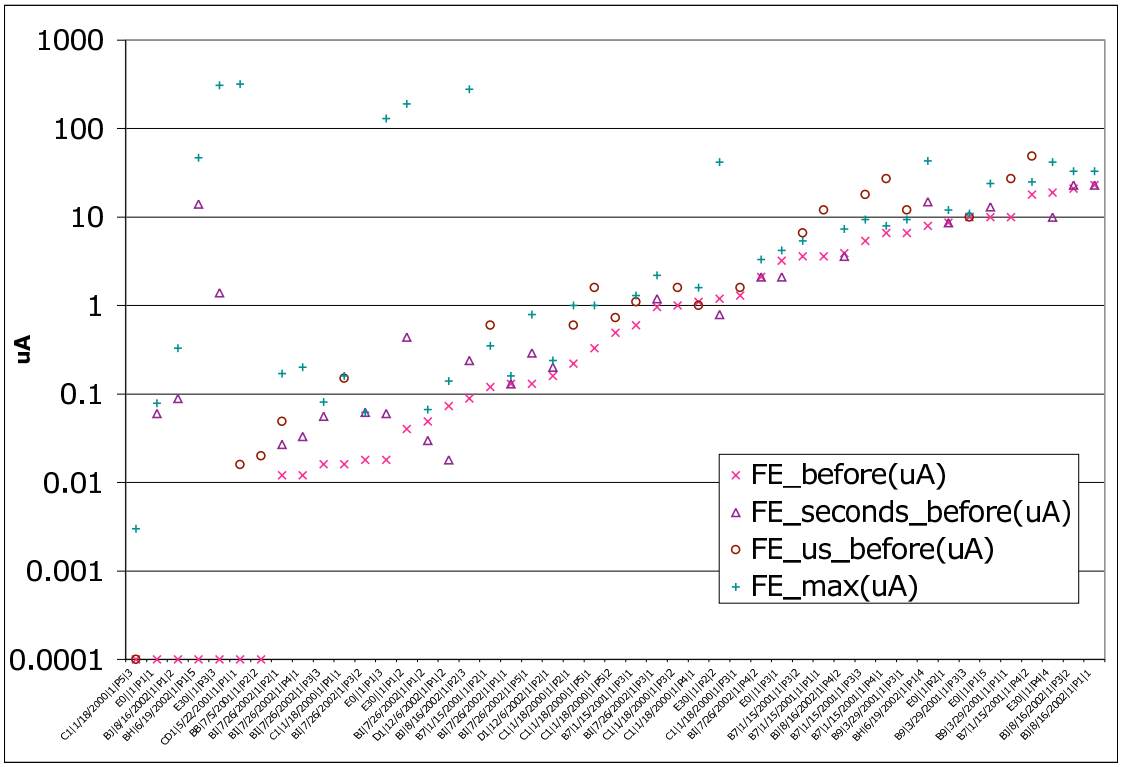


Figure 4.14: Field emission current "just" before breakdown, measured by different methods resulting in different times-scales for "just" (legend: current 10 s before, a few seconds before, a few μs before, and the maximum recorded current) The horizontal axis lists the tests in order of increasing pre-breakdown current.

breakdown. There are some cases where no field emission was observed ten seconds before breakdown, but just seconds before breakdown, field emission started. But, there are three cases where no field emission was observed, even a few microseconds before breakdown; the limited precision of the oscilloscope, however, meant that 20 nA was about the lowest current level that we could have seen microseconds before the breakdown.

4.3.2 Observations

Because of the correlation, albeit rough—within an order of magnitude—between current measured at different times (seconds vs. microseconds) before the spark, when different methods were used, I think it's worthwhile to look further into field emission before breakdown, taking whichever available measurement is the closest to the breakdown event—but take it as an order-of-magnitude estimate, still useful because field emission ranges over several orders of magnitude.

Figure 4.15 shows pre-breakdown field emission current versus breakdown field for cathodes contaminated with spiny vanadium (which dependably caused breakdown before other kinds of particles, so we knew what was breaking down). Interestingly, there seems to be no correlation, except that rarely can field emission above 30 μA be tolerated without breakdown.

4.4 Processing

Processing, or conditioning, improves an electrode's resistance to breakdown by destroying the most likely triggers of breakdown, using the mechanism of breakdown itself to locate and (with luck) obliterate likely offenders. If all goes well,

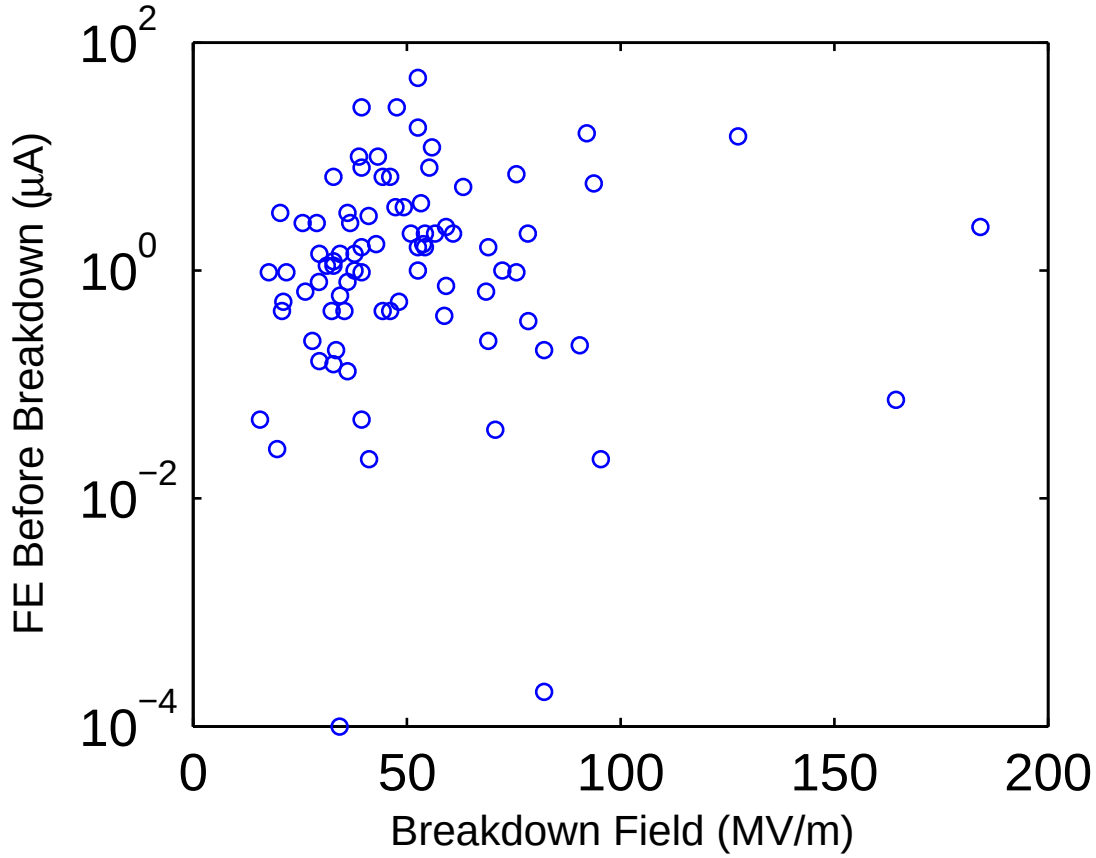


Figure 4.15: Pre-breakdown field emission for spiny vanadium versus breakdown field.

the arc destroys whatever triggered it without creating any new sources of breakdown. Systematically causing breakdown can eliminate all low-field sources of breakdown and allow high fields to be reached without breakdown. Equally important, because breakdown often destroys field-emission sites (perhaps because field emission initiated the breakdown in the first place), processing can be a tool for exterminating parasitic field emitters.

The arc cleans the cathode surface by ion bombardment, potentially removing hundreds of angstroms (in the starburst region), as well as by vaporization and removal of cathode material (like a contaminant particle). Insofar as the bulk cathode material is usually purer than its surface, sputtering away surface layers should be helpful; for instance, surface adsorbates may exacerbate field emission. However, the melting and sometimes violent reforming of the surface can create new surface features prone to breakdown; I have already discussed cases where ejected material from one breakdown event causing another breakdown event (section 4.2.1).

Processing can be very important for reaching high fields in real systems. Although we have seen that cleanliness (in preparation and installation) easily allows fields in excess of 100 MV/m without breakdown on millimeter-sized electrodes, being clean and staying clean are not always possible. Such cases require an *in situ* cleaning technique, and since intentional breakdown is always an option wherever unwanted breakdown is a problem, processing is often a good choice. Therefore, knowing how to incite breakdown in a controlled manner, may be as important as knowing how to avoid breakdown.

We have examined processing effects to some extent, although most of our work concentrated on single, isolated breakdown events; but we have not examined the

heavy-processing regime at all—at most we allowed several breakdown events at the same cathode site.

Processing and the breakdown field

In the "early processing" regime, breakdown is overwhelmingly likely to increase the breakdown field of a cathode site. Figure 4.16 shows that the ratio between successive breakdowns tends to be greater than 1 (indicating successful processing), regardless of the breakdown field (up to 100 MV/m). Most of the points on the graph represent first and second, or second and third breakdown events. Continued processing would give smaller and smaller improvements until reaching some saturation field, at which subsequent arcs would increase or decrease the breakdown field with equal probability.

Processing and field emission

Processing has been useful not only as a vaccination against further breakdown, but also for controlling field emission (such as in superconducting microwave cavities, where dissipation due to field emission can nullify the advantage of low dissipation to wall currents) [23]. In such cases, one hopes that breakdown will occur at the most troublesome field emitters and destroy them. In our experiments, we have certainly seen field emission reduced by breakdown. Figure 4.17 shows the effect of processing on field emission, similar to figure 4.16 but showing the change in field emission, rather than the change in breakdown field. As with breakdown, the emission onset field (defined here as the field at which electron emission rises above 10 nA) generally, but not always, increases after breakdown. Interesting is the lack of clustering of data at the line where previous and successive tests have

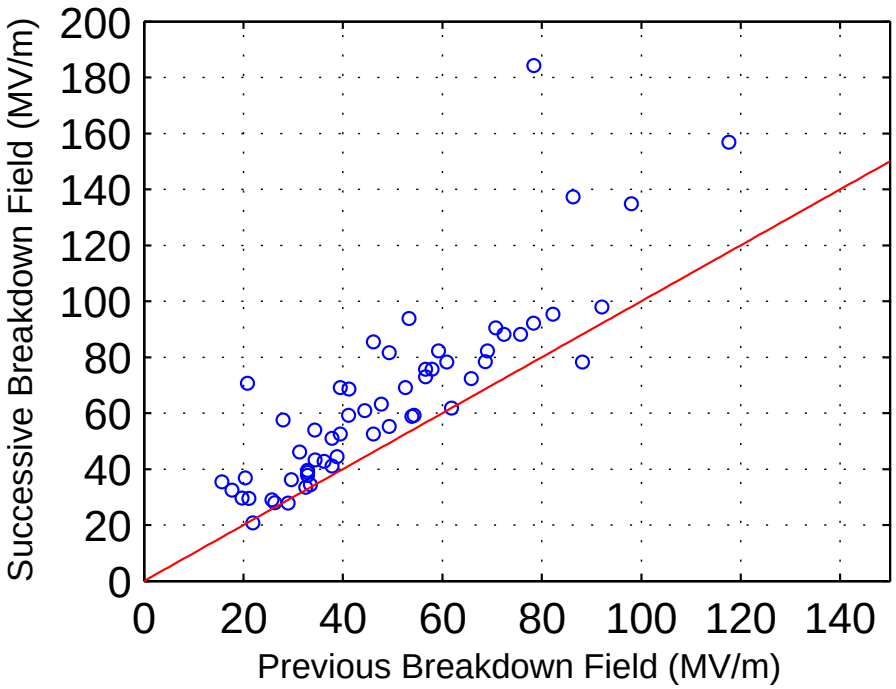


Figure 4.16: Processing works (at least for minimal processing at fields up to 100 MV/m). Each data point shows the breakdown fields of two consecutive breakdown events at the same spot, with the earlier breakdown field on the horizontal axis, and the later on the vertical. Successful processing events are above the line $y = x$.

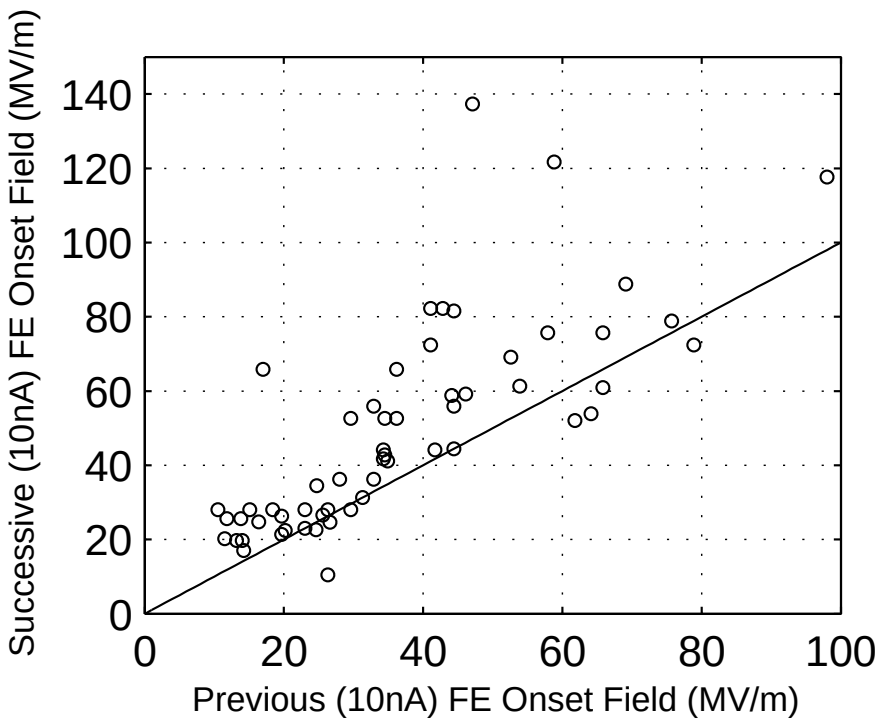


Figure 4.17: Processing not only increases the breakdown field (see figure 4.16), but decreases field emission as well.

the same onset field, indicating that the breakdown events almost always changes the onset field, one way or the other. One possible explanation is that the strongest field emitter always triggers breakdown, which then destroys it. However, we have not seen a correlation between breakdown and field emission levels just before breakdown, suggesting that it's not simply a case of the strongest emitter causing breakdown. More likely, breakdown is triggered by something, not necessarily the strongest field emitter, and the arc (and its plasma) increases electric fields in the general vicinity, which often induce any good field emitters to break down.[5] This hypothesis could be tested by examining the effect of processing on field emission

[5]For example, figure 2.13(c) on page 42 shows a breakdown site with a central crater, caused by a particle, as well as satellite craters where particles once were. It is beyond the realm of chance coincidence that the particles decided independently to break down at the same time; the formation of the central crater and starburst likely incited the cratering at nearby particle sites.

on different sized electrodes, where a random breakdown trigger would be less likely to be within a starburst-radius of a random field emitter. We know in any case that more than one starburst can result from a single breakdown.

4.5 Influence of Pre-Breakdown Pressure on Breakdown

Vacuum breakdown generally does not depend on the background vacuum pressure [76] (at least up to about 10^{-6} torr). Comparing breakdown on cathodes contaminated with spiny vanadium, at low pressures ($\leq 10^{-8}$ torr), as well as at higher pressures ($\geq 10^{-7}$ torr) achieved by turning off the pump for a certain period of time, we find no dependence of breakdown field on background vacuum pressure (figure 4.18); however, these results are not conclusive—we simply don't have the data to study this problem thoroughly.

4.6 Dependence of Breakdown Field on Gap

The theory that the only important predictors of breakdown are the cathode surface and the cathode surface field has long been proven false. To zero order, the field does seem to be the determining factor, insofar as breakdown occurs between 5 and 200 MV/m (though wider ranges are possible) over an even greater range of systems (for instance, in a 10 micron vacuum gap or a half-meter RF cavity). In general, smaller gaps can withstand higher fields. There are several possible explanations, but the two most prevalent are: (1) the local cathode field (enhanced by microprotrusions) determines the breakdown point, and the field enhancement at

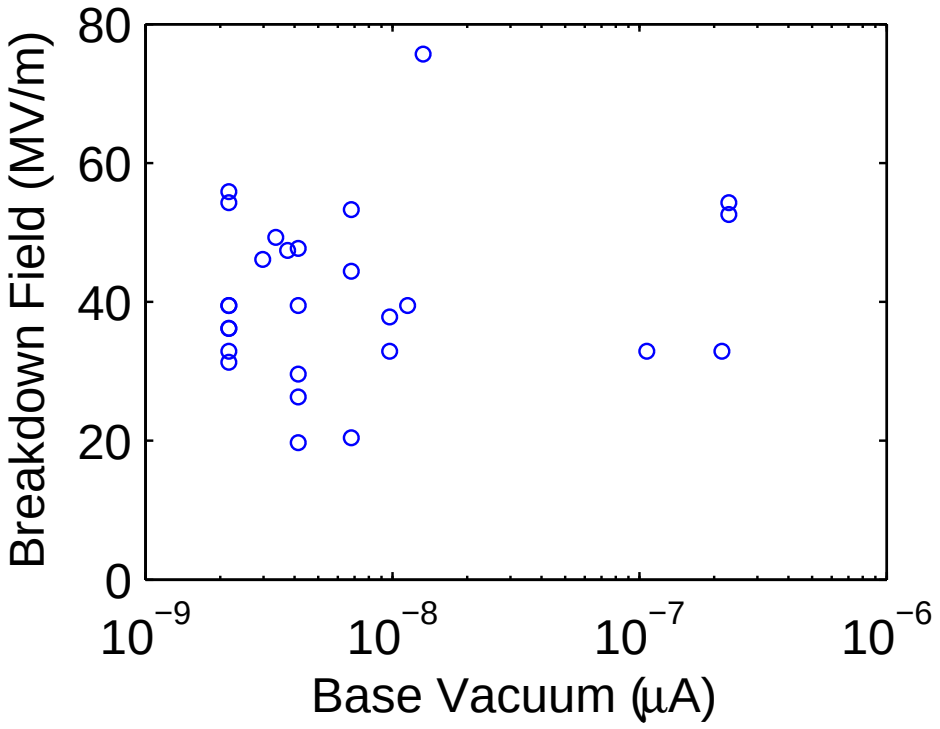


Figure 4.18: The breakdown field versus background vacuum pressure (as measured at the ion pump) for the first breakdown event on cathodes contaminated with spiny vanadium and tested at a 150 micron gap.

a protrusion decreases as the gap decreases; and (2) the gap voltage increases the bombardment energy of particles crossing the gap, so higher voltages can cause breakdown at lower fields.

The change in field enhancement with gap spacing and the dependence of breakdown on the locally-enhanced electric field was considered (and measured) in a noteworthy paper [1], and apparently confirmed over a range of different metals and gaps in [58]. However, there is some doubt about the existence of protrusions that are high enough to produce the necessary enhancements and that change with gap spacings even at relatively large gaps.

Between the maximum voltage (14.5 kV), which limited the largest gap, and error in gap measurement (10–15 microns), which limited the smallest gap, we couldn't study a very wide range in gaps, but it's worth looking at the data we have for gap effects.

There are a few points to consider when examining the data for dependence of breakdown field on gap. First, (as discussed earlier) contaminant particles on the cathode determine the breakdown field, so it's important to compare tests with the same sort of particulate contamination; spiny vanadium is especially good for this purpose because it causes breakdown at lower fields than other particles, so any accidental contamination by other particles will have a negligible effect. Second, we wanted breakdown to occur in most of the experiments; if breakdown did not occur at the first gap (up to the maximum voltage), I decreased the gap until breakdown occurred. The result is that breakdown field at a small gap was often high only because a small gap wouldn't have been used if the breakdown field had been lower (if breakdown had occurred at a larger gap).

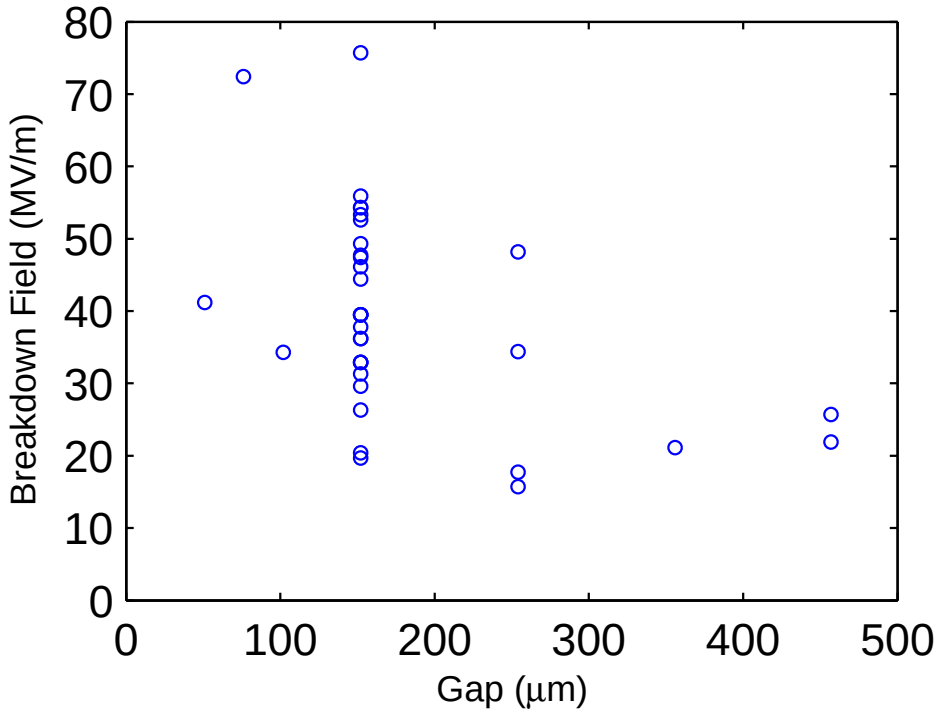


Figure 4.19: Breakdown field versus gap for cathode sites contaminated with spiny vanadium. Note that the maximum voltage of $V_{\max} = 14.5$ kV puts an upper limit on the graph: $E_{\max} = V_{\max}/d$.

Figure 4.19 shows the breakdown field versus gap for all good first tests (tests at the first gap selected for each site) on cathode sites contaminated with vanadium.

4.7 Temperature Dependence of Breakdown

In a few experiments the cathode temperature was raised to about 100°C, which (concluding from the limited number of tests) seemed to suppress the likelihood of breakdown. When a voltage V applied to a cathode site at 100°C caused no breakdown, the same site, later at room temperature, did break down at a voltage below V.

The most convincing evidence came from tests on two cathode sites, each intentionally contaminated with small carbon particles. Both sites reached 95 MV/m

at 100°C with no evidence of field emission. The gaps were not adjusted again for the room temperature tests (and in one case the ion pump was turned off so the vacuum pressure would remain the same as it was at 100°C). At room temperature (22°C), field emission began near 40 MV/m, and breakdown occurred around 90 MV/m (assuming the gap remained the same). I'm puzzled at the somewhat odd field emission at room temperature (after heating)—odd in that it hardly increased at all, even as the voltage almost doubled.

Error in measuring the gap is the main obstacle to reaching a firm conclusion, aside from the small number of data points. If the gap was adjusted between the 100°C test and the room temperature test, then the two gaps were probably within 10–15 microns, a 10% error for a 150 micron gap. On the other hand, if the gap was not adjusted between test, then thermal expansion could have altered the gap.

Chapter 5

Results: Starbursts and Craters

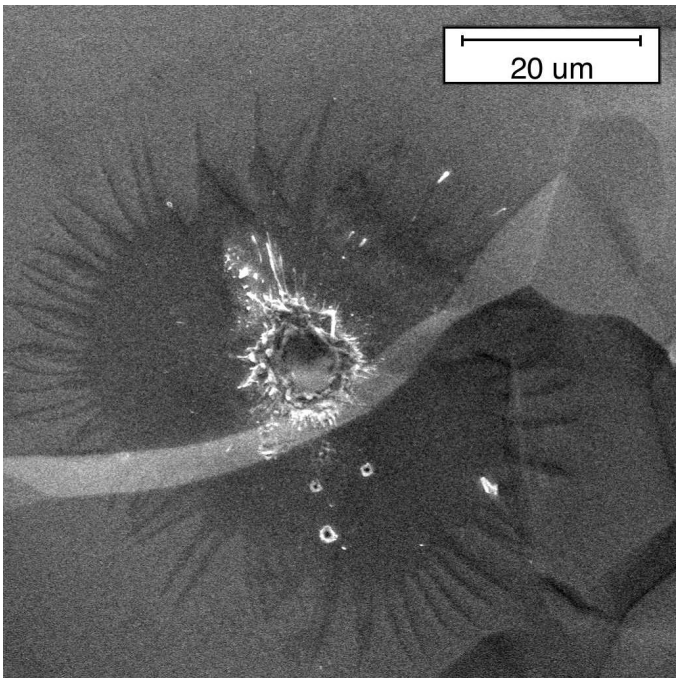


Figure 5.1: A starburst containing a large, central crater, a few small satellite craters, and many streamers around the border.

Starbursts, like that shown in figure 5.1 (see also appendix G), were first seen in superconducting RF resonant cavities that had experienced voltage breakdown [77]. These striking shapes are invisible in a light microscope, but appear readily in

an SEM;[1] the contrast indicates that the starburst (under a 20 kV electron beam) has a different secondary electron emission coefficient from that of the surrounding region.

Noting that starbursts appeared in regions with a high electric field, and using arrays of thermometers ("temperature mapping") to identify field emission sites, the starburst phenomena were connected with areas of high electric field and field emission (see [57]). Starbursts were subsequently seen on niobium cathodes in a DC voltage gap, furthering the connection with the electric field, and suggesting that other properties of superconducting cavities (RF, magnetic field, low temperature) play no role in starburst formation (or are at least limited to minor roles).

The detection of foreign material in superconducting RF cavity starbursts (usually in or near the central craters) further strengthened the connection between starbursts and field emission [44], since foreign particles are notorious field emitters. Most of these contaminants were identified with AES (Auger electron spectroscopy), but not EDX (energy dispersive x-ray analysis), indicating that any contaminants were present in small amounts on the surface, because AES and EDX have similar overall sensitivities, but AES sees only the first 10–100 Å, whereas EDX sees everything within a few micron depth (see section 3.4).

Although the necessity of using AES made analysis much more difficult, it led to the discovery that the starburst areas lacked fluorine [44]; fluorine is present in a few percent concentration in the oxide layer of niobium that has been etched in hydrofluoric acid (as niobium cavities almost always are). It was a mystery at the time how the fluorine was depleted, and why such a small difference in fluorine concentration affected the secondary emission contrast in an SEM.

[1] Large craters in a starburst may be optically visible.

By concentrating on DC experiments (rather than RF) we were able to obtain more starbursts on samples better suited to AES. We confirmed that the original particle at the starburst site indeed left behind a little material that could be detected by AES, but generally not by EDX. Moreover, we made a very detailed map of fluorine around the starburst region (figure 5.2) and showed that the entire starburst, including the streamers, contained no fluorine, whereas the surrounding areas generally had a few percent fluorine contamination in the oxide.

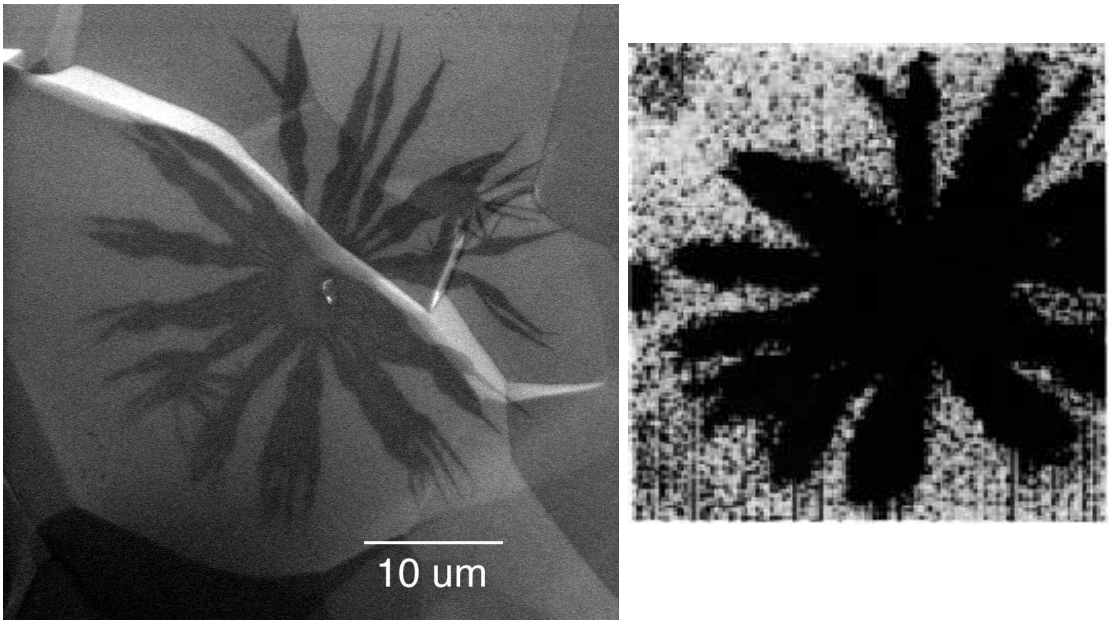


Figure 5.2: Starburst regions contain no fluorine, whereas a few-percent concentration of fluorine contaminates the oxide layer elsewhere. On the left, a secondary emission image of a starburst; on the right, an Auger map of fluorine concentration (darker areas contain less fluorine).

We also found that the starburst region, in general, had less surface carbon (figure 5.3). Carbon (in the form of hydrocarbons, carbon monoxide, *etc.*) can be found in a very thin layer (a monolayer) on almost all surfaces (one surface analyst observed anecdotally that he'd never seen a surface without some carbon except in samples that had been fractured in ultra-high vacuum). Unlike fluorine,

carbon is not completely absent within the starburst region, but the concentration is noticeably reduced; in fact, the Auger map of carbon reveals the starburst pattern in a small fraction of the time needed for the Auger map of fluorine to accumulate enough signal to show the starburst pattern. Furthermore, surface carbon is known to affect secondary electron emission [45], explaining the contrast visible in an SEM.

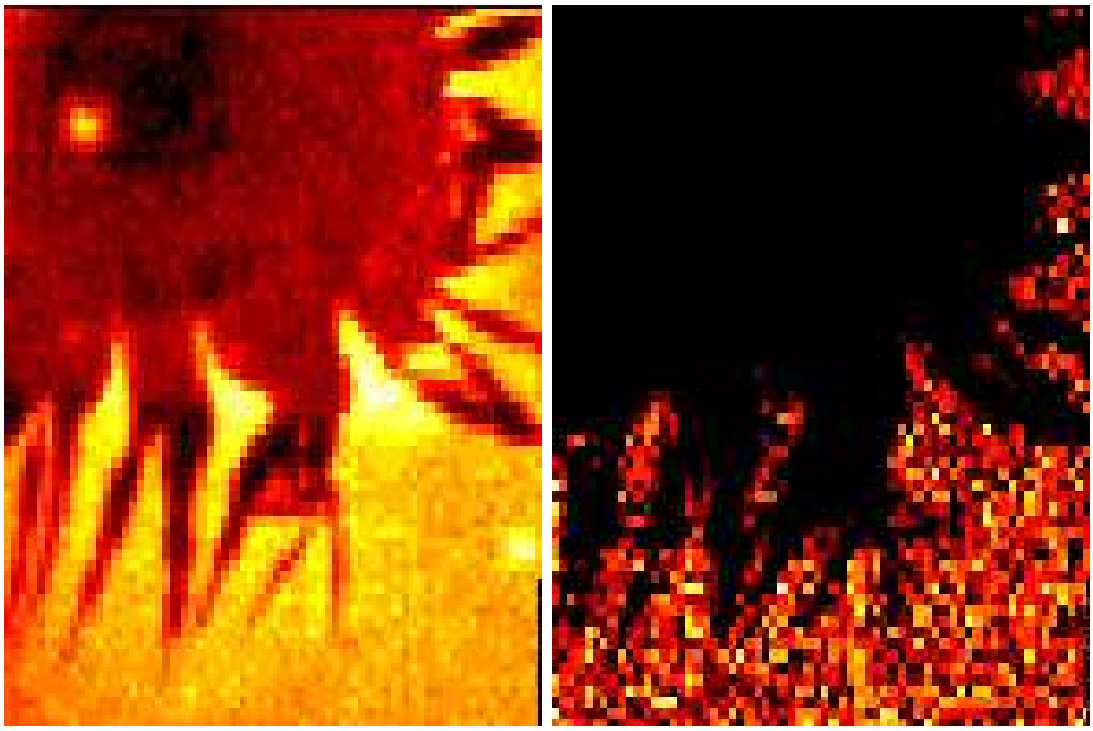


Figure 5.3: Auger maps for elements C (left) and F (right) for the starburst (SB1) shown in figure 2.13 on page 42, showing F and C depletion within the starburst.

The arc appears to "clean" the cathode surface in a starburst-shaped region, removing surface carbon and fluorine in the oxide layer. Ion bombardment seems a likely mechanism: ions must be present for the arc to carry so much current, (positive) ions would travel to the cathode, and ion bombardment is a good way to clean surfaces. The original niobium, etched in HF, has some small concentration of fluorine in its oxide, which forms immediately after etching; carbon also con-

taminates the surface more or less immediately. During the arc, ion bombardment sputters away the surface carbon and the surface oxide along with any fluorine in the oxide. As soon as the sample is removed from vacuum (for transport to the microscope), the oxide immediately reforms, but without any fluorine contamination. Carbon also re-adheres to the oxide, but for some reason (which we don't yet understand) less of it forms on the surface of the starburst region.

Because the oxide reforms to its original thickness with exposure to air, we cannot see any difference in the oxide thickness in- and out-side of the starburst. However, we grew an unnaturally thick oxide (hundreds of angstroms, versus the natural oxide of tens of angstroms) on some samples, and subjected them to voltage breakdown. Afterwards, the oxide within the starburst region was found by AES to be of the natural thickness, indicating that all the extra-growth oxide had been removed.

Suspecting that surface carbon, and not fluorine in the oxide layer, creates starburst contrast on niobium, we wondered whether starbursts would appear on non-niobium cathodes, since all metals tend to have some carbon contamination. We tried gold and copper cathodes, and found starbursts on both (figure 5.4); the starbursts appear somewhat different, however. Sometimes the starbursts are barely visible in the SEM, but we always see a disc of carbon depletion with AES, and usually see some streamers, though they tend to be much messier on non-niobium cathodes.

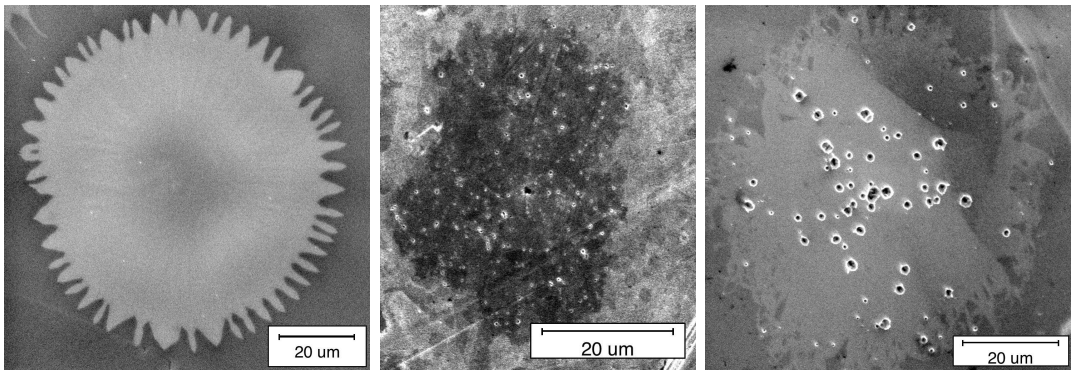


Figure 5.4: A starburst on diamond-machined copper (left), electropolished copper (middle), and gold film on a niobium substrate (right). The left and right starbursts are "inverse" starbursts, lighter than the surrounding area.

5.1 Typical Starburst Parameters

Starburst sizes and shapes range widely. We consider here "regular" starbursts, which have circular symmetry, a central circular region, and streamers of some sort emanating from that central region (the streamers are not always long and straight, but sometimes very short or crooked). The starbursts (in our DC experiments) have central regions with diameters between 5 and 100 microns, and total diameters (including streamers) from 20 to 130 microns; the ratio of outer to central diameters generally falls between 1 and 3. The number of streamers varies from 15 to 150, with streamer widths (at the edge of the center region, where the streamers first form) usually between 1.6 and 4.5 microns.

5.2 Inverse Starbursts

Occasionally, we find a starburst that is lighter than the surrounding area (see the left and right starbursts in figure 5.4)—that is, the secondary electron emission (under the 20 kV SEM beam) inside the starburst is greater than it is outside the starburst; usually starbursts appear darker than the surrounding area (less

secondary emission inside the starburst). Although we know that surface carbon contamination can affect secondary electron emission, the effect is not simple—it may depend on the specific form of carbon, for instance. Furthermore, electron bombardment itself can change the amount and/or form of surface carbon contamination, changing the secondary emission yield [45].

We happened, fortuitously, to observe that in a matter of seconds, a strong electron beam (10 kV, 1 μA) can turn a starburst from regular (darker than surroundings) to inverse (lighter than surroundings). The usual post-test examination is done with an SEM with a 20 kV, sub-nA beam, which usually does not affect the starburst contrast; however, Auger analysis requires a much stronger electron beam, and scanning Auger microprobes have secondary emission imaging systems just like SEMs, so the sample was pummeled by electrons as we searched for a spot to analyze. As we zeroed in on a starburst, it would change (relative to the surroundings) from darker to lighter is a few seconds.

We therefore hypothesize that inverse starbursts are normal starbursts that have subsequently experienced severe electron bombardment. Of course we don't think of the cathode as being a target of electron bombardment, but while the arc maintains a plasma around the cathode, especially energetic electrons have enough energy to overcome the potential barrier and hit the cathode. On an anecdotal level, inverse starbursts seem to be more frequent on cathodes that have been subjected to multiple breakdown events are that exhibit unusually violent arcs (lots of melting, *etc.*).

5.3 Anode Starbursts

We have observed starbursts on a niobium anode (figure 5.5). All anode starbursts have been inverse starbursts (lighter than the surrounding area), which supports the theory that inverse starbursts have experienced severe electron bombardment (which anything on the anode certainly would, from field emission alone, never mind the arc). Since we used the same anode to test many cathodes (or cathode pedestals), but found relatively few starbursts, we conclude that starbursts are much less likely to form at the anode than the cathode.

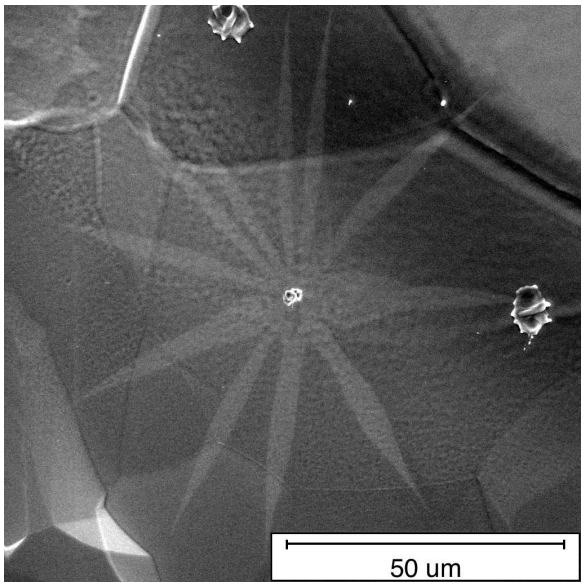


Figure 5.5: Starburst found on a niobium anode

Most of our experiments were performed with anodes made of Ta-W (about 10% tungsten), and we never observed starbursts on these anodes. However, we never tested a Ta-W cathode, so no direct conclusion about starburst likelihood can be reached.

5.4 Starbursts and Voltage Breakdown in DC and RF

Voltage breakdown in RF cavities surprisingly resembles DC voltage breakdown. Both seem to be connected in some way to the electric field, field emission, and contaminant particles; most convincing, however, are the similarities among starbursts in niobium RF cavities (figure 5.6) and on niobium DC cathodes that suffered voltage breakdown.

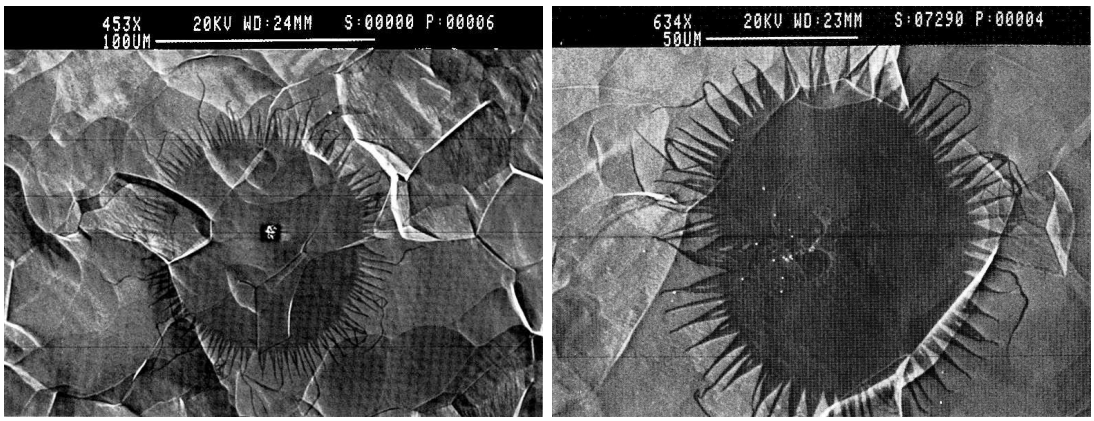


Figure 5.6: Starbursts from a 6 GHz superconducting RF cavity that underwent breakdown (reproduced from [77]).

While RF and DC starbursts clearly share many features, one difference is apparent: RF starbursts are larger than DC starbursts. Knobloch [55] shows that starburst size varies with cavity frequency: higher frequencies show smaller starbursts (figure 5.7). For example, starbursts in 6 GHz cavities average around 200 microns in diameter (although starbursts considerably larger and smaller were found as well), while starbursts in 1.5 GHz cavities averaged about 700 microns diameter. Graphing starburst size against RF period (inverse frequency), Knobloch suggested a linear relationship between starburst size and RF period; however,

with only three data points (three cavity types tested), a fitted line is not very conclusive.

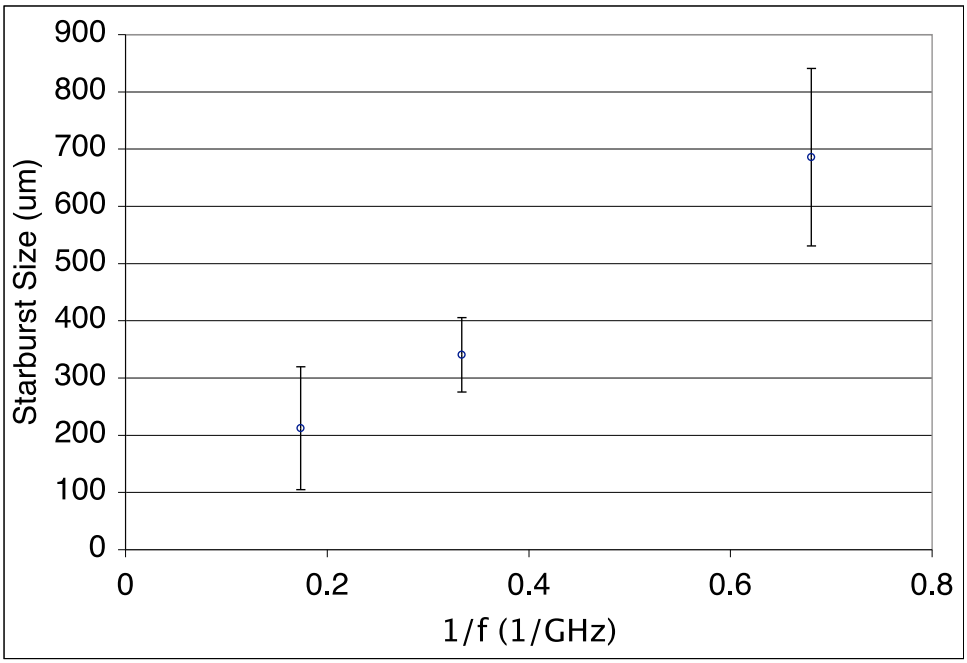


Figure 5.7: Diameter of starbursts (including streamers) versus the (inverse) frequency of the RF cavity in which they appeared, (data replotted) from [55]; the "error" bars indicate the spread of the data: ± 1 standard deviation of the sample data.

DC starbursts tend to be somewhat smaller—usually less than 100 microns diameter. According to the suspected inverse frequency relationship, DC starbursts should be larger than RF starbursts. However, the apparent dependence on frequency may be masking a different relationship—higher frequency cavities are smaller, and contain less stored energy at similar field strengths. Starburst size may be determined not by the frequency of the RF field, but by the energy available to expand the plasma that creates the starburst. The stored energy of the DC system is less than that in a 6 GHz cavity, and DC starbursts are consequently smaller. Graphing both DC and RF starbursts against the stored energy in the

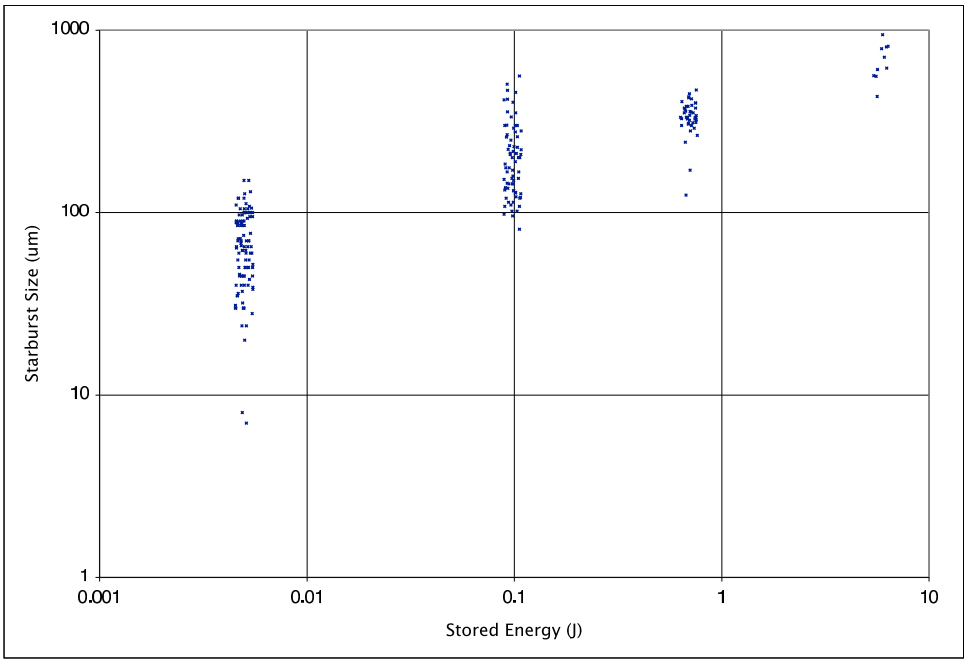


Figure 5.8: Diameter of starbursts (including streamers) versus the estimated stored energy, based on reasonable parameters for each system. Each group of points has been artificially spread in the horizontal direction to minimize overlapping.

system (figure 5.8), we find that the data fall tolerably close to

$$D_{\mathrm{SB}} \propto U^{1/2} \tag{5.1}$$

where D_{SB} is starburst diameter (including streamers), and U is the stored energy. For the DC case, the energy is (one-half) the capacitance times the voltage squared, roughly

$$U_{\mathrm{DC}} = \frac{1}{2}CV^2 \sim \frac{1}{2}100\,\mathrm{pF}\cdot(10\,\mathrm{kV})^2 = 5\,\mathrm{mJ}. \tag{5.2}$$

To calculate the stored energy in the RF cavities, I used the formula: [54, 88]

$$U_{\mathrm{RF}}(f) = \frac{1.5\,\mathrm{GHz}}{f}\cdot\frac{E_{\mathrm{peak}}^2}{286}\frac{\mathrm{J}}{(\mathrm{MV/m})^2} \tag{5.3}$$

which relates the cavity stored energy to the peak electric field and cavity frequency by a shape-dependent factor (1/286)—then I assumed that the peak field would

have been around 40 MV/m, since cavities with starbursts were probably field-emission limited. Since the actual voltage (for DC) and peak field (for RF) almost certainly varied quite a bit, the calculated energy should be taken as a middle value (maybe a bit on the high end of the range). This happily suggests that the area of the starburst is proportional to the stored energy:

$$A_{\mathrm{SB}} \propto D_{\mathrm{SB}}^2 \propto U, \tag{5.4}$$

a physically appealing result—the stored energy goes into plasma expansion and the consequent ion bombardment; the more energy, the more expansion. However, the range of starburst sizes, as well as the estimations of stored energy make it difficult to confirm this relationship. In fact, in figure 5.8, the data look slightly better fit by $D_{\mathrm{SB}} \propto U^{3/8}$.

In DC tests we knew at what voltage a breakdown occurred, so we looked for a dependence of starburst size on breakdown voltage (actually, on voltage squared—hence on stored energy). No convincing trend appears (figure 5.9). In defense of the relationship between starburst size and stored energy, the starburst-forming process is more complicated than dumping all available energy into the starburst. Several starbursts may form during a single breakdown, so the ion bombardment that produces the starburst requires only a fraction of the available energy. Therefore the total stored energy is just a rough measure of the (smaller amount of) energy that could be directed toward starburst expansion. There are a few cases where a breakdown event created only a single starburst; the size of these starbursts does on average increase with stored energy, but there are not enough data to be very convincing.

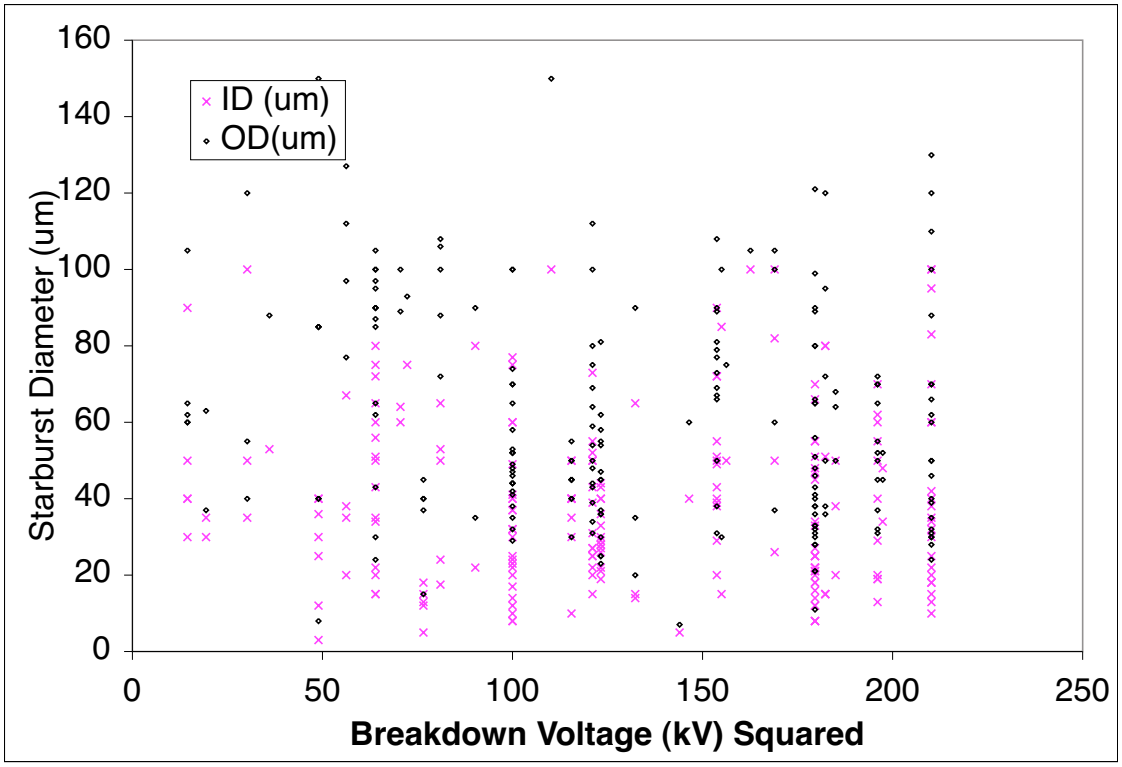


Figure 5.9: Starburst diameter (both OD, the diameter including streamers, and ID, the diameter excluding streamers) versus voltage squared (which is proportional to stored energy) in square kilovolts.

5.5 Substrate Oxide Thickness and Starbursts

Evidence in section 4.2.3 (page 92) shows that the cathode material has little effect on the breakdown voltage because contaminant particles trigger breakdown (if there are contaminant particles). Once a particle initiates breakdown, however, the cathode material plays a crucial role in the evolution of the arc plasma. Especially important, the oxide thickness, rather than the specific materials' properties, seems to have the most influence.

Not all starbursts exhibit striking symmetry and beauty (see appendix G), but the general features are common: a central area, which breaks up into streamers beyond a certain radius, and often a central crater or central group of craters, sometimes with desultory satellite craters ranging widely across the starburst. Starbursts on copper, for instance, tend to be less striking than on niobium.

On niobium and copper cathodes with thick oxides (hundreds of angstroms), starbursts look quite different (figure 5.10): they are generally smaller, have no streamers, exhibit no craters, but instead show signs of more general melting, often leaving a once-molten column of metal in the center.[2] Starbursts on thickly-oxidized copper resemble those on thickly-oxidized niobium (figure 5.11).

[2]Note that niobium naturally grows an oxide, Nb_2O_5, that is 30–60 Å thick, while the natural oxide on copper is thinner, around 10 Å. The oxide layer on niobium easily grows by anodization—making niobium the anode in a solution of water and some electrolyte (we used a very weak solution of ammonium hydroxide); the voltage controls the final oxide thickness, about 20 Å/V [109] (the inverse of this number is the field strength of the oxide; see also [43, 51]).

Copper oxide can be grown by heating copper in air; I heated the copper to around 120±20°C for about a minute, which turned the recessed area of the cathode to a gold color, and the pedestals to a dappled magenta-grey-yellow. At temperatures below several hundred degrees Celsius, Cu_2O forms (red oxide, as opposed to black oxide, CuO) [47]. Extrapolating from [47] (which presents experimental data only at higher temperatures) I estimate (very roughly) the oxide thickness to be around 200 Å.

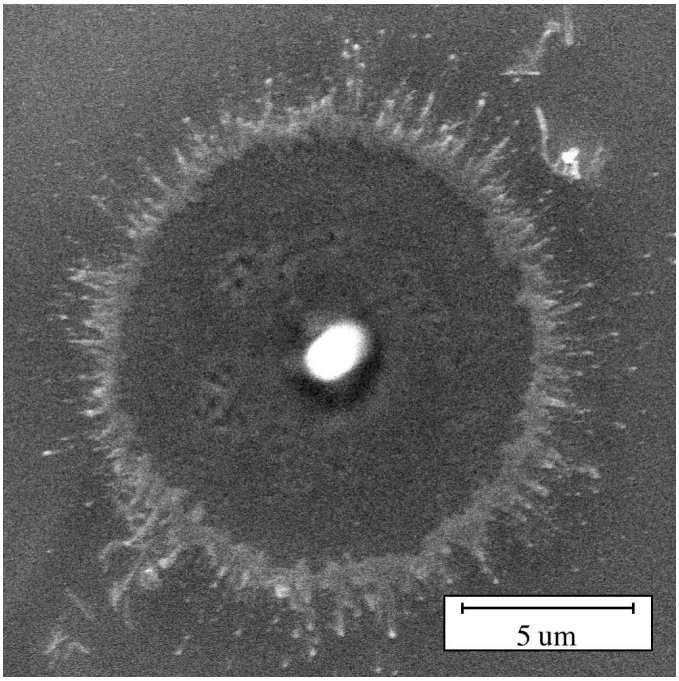


Figure 5.10: An exemplary starburst on niobium with an oxide layer grown to 400 Å by anodization. Note the smaller size, the lack of streamers (although it looks as if molten metal splashed out around the border), the lack of craters, but the presence of a central column of Nb with rounded edges as if it had once been molten.

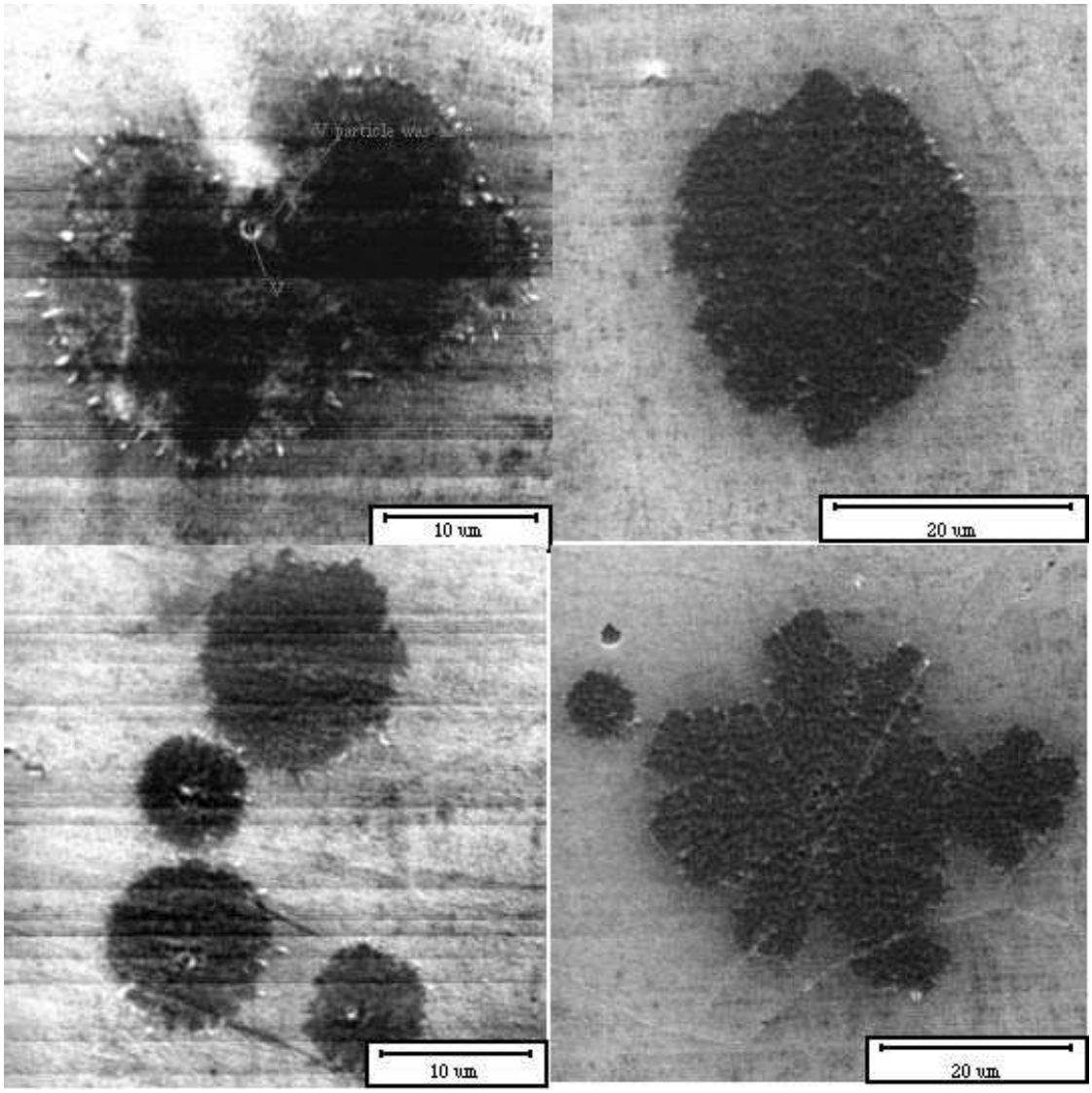


Figure 5.11: A heavily oxidized copper cathode after breakdown. The "starbursts" often exhibit no central craters, but perhaps a column of once molten metal; they don't have streamers, but show splashes of metal around the edges.

Auger maps of oxidized starbursts show carbon depletion just like starbursts on natural oxides (except no streamers). By sputtering away the surface layer-by-layer with an ion gun, and using AES to measure the remaining surface oxide, we compared the thickness of oxide inside and outside the starburst, and found that the oxide outside the starburst remained at its original (unnaturally large) thickness, while the oxide inside the starburst matched the "natural" oxide thickness. We conclude that the arc removed the entire oxide and possibly some of the bulk material as well, and an oxide of natural thickness re-grew due to exposure to air during transfer to the microscope.

The mechanism for the depletion of fluorine in starbursts on niobium is thus clear. Only the oxide contains fluorine (a result of etching in HF), and the arc removes the entire oxide by ion bombardment, consequently removing the fluorine as well.

The removal of the entire oxide coating within a starburst suggests that thick oxide coatings are unlikely to be a helpful surface treatment for reducing field emission and the possibility of voltage breakdown. It was once hoped that a thick coating of insulating dielectric would reduce field emission, since the dielectric should reduce the field at the metal-dielectric interface by a factor of the dielectric constant. Experimentally, coatings have been found to be generally unhelpful for reducing field emission. Part of the problem (as we have shown) is that contaminant particles cause breakdown (regardless of the substrate oxide thickness, as we have also shown), and coating the electrode or cavity surface does nothing to the dust particle that later attaches itself to the surface. Even if dielectric coatings did help, this work suggests that they would be destroyed in any region subjected to an arc.

Having seen a clear difference in starburst character between natural oxide and thick oxide on both niobium and copper, we wondered what sort of starbursts would form on a cathode with no oxide. Without the means to remove the cathode oxide in the apparatus, we tried the next best thing: gold. Gold likes oxide even less than copper, although even gold will allow a little oxygen and carbon on its surface—but probably less than a monolayer. Because we already knew how to prepare a good, clean niobium surface (with the right etching recipe, *etc.*), and because of the expense and trouble of machining a cathode out of bulk gold, we sputtered about 1000 Å gold onto a niobium cathode (first sputtering away the niobium oxide so the gold would stick).

Sure enough, starbursts on gold have a different character: most obvious, starbursts on gold are uniformly covered with small craters; gold starbursts have streamers (figure 5.12). For starbursts on gold, streamers are often a bit short and tangled, but still recognizable, especially in an Auger carbon map. Other arcing experiments have seen similar cratering on gold cathodes [4].

Some electropolished copper cathodes also exhibited uniform cratering within starbursts similar to that on gold cathodes (see figure 5.4 on page 115). The electropolishing job was done by collaborators at Peking University, and the results on these small samples was not as good as desired—the cathodes were covered with long, thin scratches; moreover, the samples were shipped halfway around the world in ordinary plastic bags that rubbed against the surface, leaving hydrocarbon deposits. For that reason I don't want to present these results as typical of a good electropolished surface, but merely as another case where the cathode surface affects starburst formation.

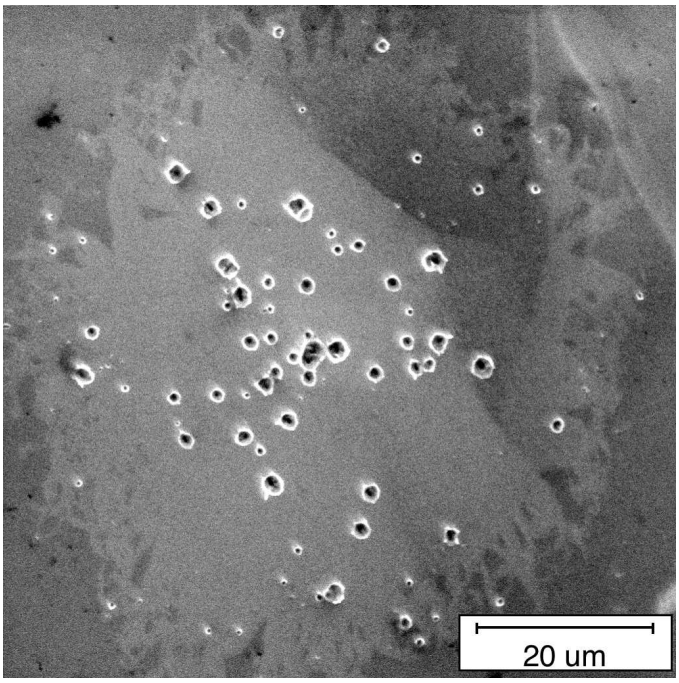


Figure 5.12: Starbursts on a gold cathode

Examination of the "satellite" (non-central) craters with AES rarely detected anything of interest. AES is sensitive to most elements above Li in concentrations greater than 0.1–1% in a 10–100 Å surface layer. However, carbon contaminates all surfaces, so we probably could not have distinguished any carbon residue if carbon played a role in forming the satellite craters.

One exception to (not-) finding of contaminants in satellite craters: one copper sample was diamond-machined (for a good surface finish) and sent through a furnace treatment at CERN to mimic the preparation of copper accelerator cavities, for which voltage breakdown can be a performance-limiting problem. During heat-treatment, the copper surface became uniformly contaminated with tiny particles of manganese and sulfur (figure 5.13). During breakdown, craters formed at the site of many of the particles within the boundaries of a starburst (figure 5.14).

Although it suggests interesting possibilities, this case appears to be rather exceptional.

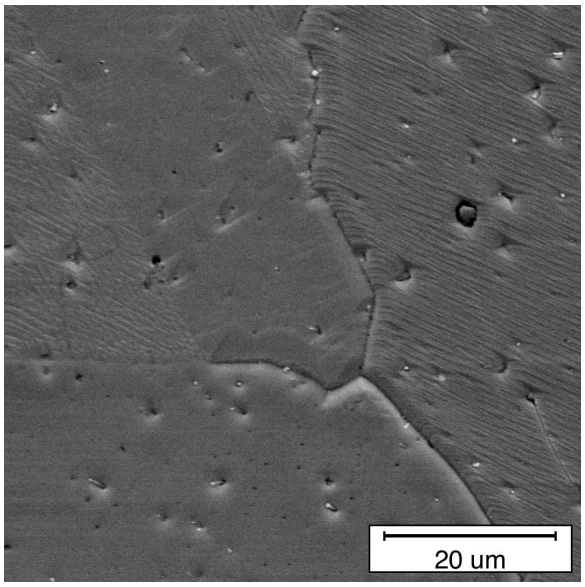


Figure 5.13: After heat-treatment, small particles of Mn and S covered the copper surface uniformly.

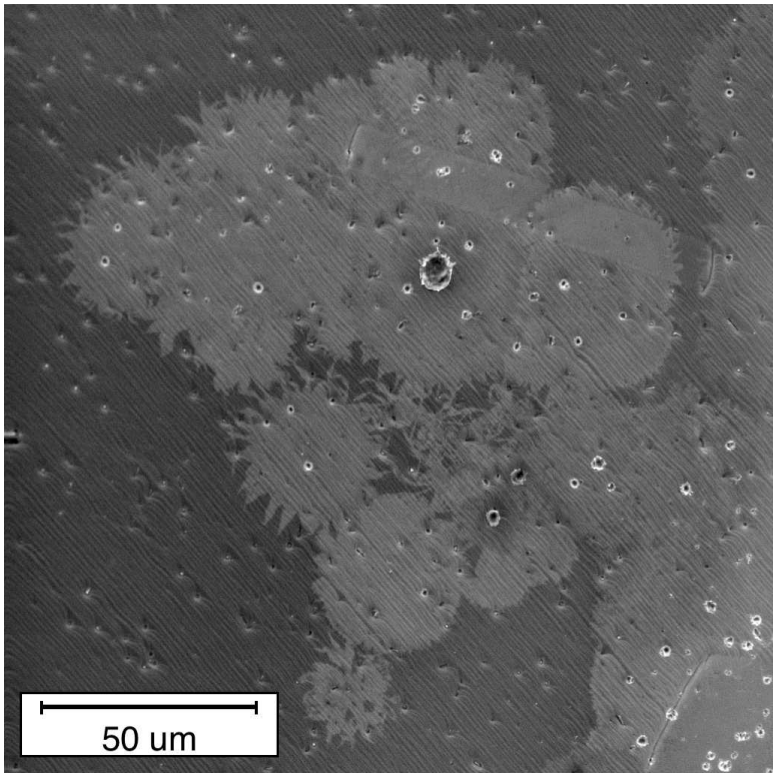


Figure 5.14: Painstaking inspection of before-breakdown and after-breakdown pictures showed a high correlation between the locations of satellite (non-central) craters and contaminant Mn particles (see figure 5.13) within starbursts. Mn particles outside starbursts remained. Note that most Mn particles reside in a little surface depression, and not all depressions contain a particle; only the depressions inside a starburst that contained a particle formed a crater.

Chapter 6

Computer Simulations of Breakdown

Computer simulations can probe breakdown phenomena at much smaller length and time scales than can be measured in experiments. In particular, computer simulations can help explore the early stages of breakdown, which may occur too quickly (and too far from equilibrium) to be probed directly in experiment. Several years ago, Jens Knobloch ran similar simulations with an RF field using the program MASK, developed by Science Applications International Corporation [56]. His results suggested that a high density of neutral gas around a typical field-emitter could lead to run-away current growth. However, very soon after the current began to rise, the computer simulation ran into an instability that made further simulation impossible. For that and other reasons, it was concluded that MASK was not the program likely to yield further results (especially relating to the development of the arc after the initial trigger).

We investigated another simulation program, OOPICPRO, maintained by Tech-X Corporation, to try (1) to confirm Knobloch's results, and (2) to continue the

simulation further in time, as well as (3) to find a simulation capable of more realistic simulation of a breakdown event for future work.

6.1 PIC Simulations and OOPIC

One approach to plasma simulation uses particles, computing the electromagnetic interactions between them (including specified boundary conditions), somewhat as nature does, except that nature can handle much larger computations.[1] PIC codes, or particle-in-cell codes, are computationally practical programs for simulating individual particles; for a simple introduction to PIC codes, see [8]; for a more thorough description of PIC codes, refer to [46].

PIC codes simulate discrete particles, but trade accuracy at small length scales for greater efficiency which makes computation possible. PIC simulations pretend for the purpose of field calculation that space is a discrete lattice:[2] dividing the space into a grid of cells, a PIC simulation counts the total charge in each cell and calculates the field at one point in each cell, and then interpolates the fields to particle positions to calculate the force on each particle. Finding the field on a lattice greatly reduces computation time. Computing in a straightforward and simple manner, for example, the Coulomb force between every pair of particles would require $N(N-1)/2$ evaluations of the Coulomb force per timestep to simulate N particles, increasing calculation time with the square of the number of simulated particles. On the other hand, calculating the electric field on a lattice due to charge on a lattice requires, roughly, inverting the matrix representing the

[1] In contrast, magnetohydrodynamic plasma simulations treat the charged particles as parts of a continuous charged fluid.

[2] PIC simulations find the field on a lattices, but the particles still move continuously (up to machine precision) through space.

Laplacian operator on discretized space; the field calculation time depends only on the size of the grid. It remains only to distribute the charge of each particle to the appropriate cell, and then to calculate the force on each particle from the field of the cell it's in—for a constant grid size, the whole operation scales linearly with N, not quadratically, a huge savings for large N. Techniques to perform matrix field calculations quickly[3] make PIC codes worthwhile.

PIC codes thus offer an enormous advantage in allowing more particles to be simulated at the cost of treating the particles not as points of charge, but as charge smeared out over a cell. Particles far apart exert the usual $1/r^2$ Coulomb force on each other; however, the calculated force between two particles in the same cell is much less than the true $1/r^2$ Coulomb force. However, free charged particles tend to move to achieve low electric fields (as in a conductor, where free electrons move to zero the electric field), and neglecting short-range interactions of a plasma, a conducting fluid, can actually be a small perturbation on reality.

While the fields are calculated on a grid, particle positions and velocities are real numbers with the precision of the computer (about 15 digits).

OOPIC (see [105]) is a 2D object-oriented PIC code written in C++.[4] The full source code is currently freely available for non-commercial research, and, because of the original object-oriented design, it can be easily read and understood. I found

[3]Although matrix inversion time generally scales as N_g^3, where N_g is the number of grid points (the matrix is $N_g \times N_g$), most of the elements of the "Laplacian matrix" are zero, allowing much faster computation. For example, in one dimension, the Laplacian is tridiagonal, the inversion of which scales as N_g [90].

[4]Actually, the world of OOPIC is said to comprise two-and-a-half dimensions; that simply means that particle positions are two dimensional, but particle velocities can be 3 dimensional. In the 2D cartesian geometry, this means that particles are effectively infinite lines extending in the third dimension; in the 2D cylindrical geometry, particles are actually rings about the cylindrical axis of symmetry. The extra half dimension allows for momentum in the third dimension, along the line or ring.

this to be a great advantage, to be able to learn exactly how the code works, and to be able to modify it to study vacuum breakdown.

The method for handling particles collisions is one of the biggest differences between OOPIC and MASK (in terms of functionality); OOPIC handles collisions with much more detail. MASK simply treated ionization of neutrals by electrons, creating ions based on the number of electrons in a cell and the energies of those electrons (electron energy affects ionization probability). OOPIC uses a Monte Carlo method to determine exactly which electrons and ions will collide with neutrals, and, considering the energy of each, actually takes into account different cross sections for elastic collisions (momentum transfer), excitation, and ionization, and then calculates post-collision velocities of the resulting particles. OOPIC also has other advantages, many of which result from the greater flexibility due to greater availability of computing power.

OOPIC did lack a few essential features for our purposes: (1) OOPIC did not support a field-dependent electron emitter, and (2) OOPIC had no provision for the time-evolution of neutral gas (the neutral gas density had to be constant in time). Tech-X provided the basic field-emitter, and I tweaked it to make it practical under certain "inclement weather" it would encounter. I also added the ability to track neutral particles and use their density (within each cell) to calculate ionization cross-sections.

Although OOPIC supports cartesian (planar) 2D geometry, it made more sense to use cylindrical 2D geometry, with the field emitter on axis, since the problem is actually a 3D problem with (almost) cylindrical symmetry (figure 6.1). In 2D cylindrical geometry, the world is 3D with cylindrical symmetry, making it unnecessary to store any azimuthal coordinates; thus "point" charges are ring charges

about the cylindrical axis (that is, calculating the force between two "points" in 2D cylindrical geometry gives the force between two rings in three dimensions). At this point, a fully three dimensional PIC code would be too slow to treat this problem on a normal computer.

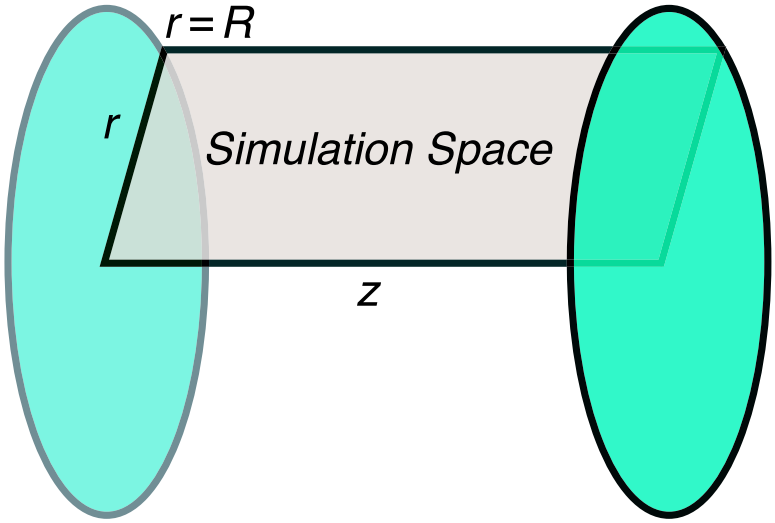


Figure 6.1: The world of OOPIC—in cylinder mode, OOPIC simulates a three dimensional world with cylindrical symmetry. The coordinates in the world are only z and r, and the world is the half-plane $r \geq 0$, with "dielectric" boundary conditions imposed at $r = R$, and "conducting" boundary conditions at $z = 0$ and $z = d$.

Most of our simulations were done using electrostatics only (as were the MASK simulations). Part of the value of the computer simulation, since we are not so far advanced that a breakdown event can be simulated in its full reality, is to identify the important parts of the problem. If much of breakdown can be explained with electrostatics, then it's far simpler to stick with electrostatics until we understand the limits of an electrostatic explanation, and only then consider electromagnetic effects.

6.2 Macroparticles

A real breakdown event involves at the very least, something like 10^{10} particles, and possibly many orders of magnitude more. No conventional computer in the world can manage that number of particles. Instead of 10^{10} electrons with charge $-e$ and mass m_e, we bind together 10^4 groups of 10^6 electrons to make 10^4 macro-electrons with charge $-10^6 e$ and mass $10^6 m_e$.

Keeping the charge to mass ratio the same ensures that the macroparticle trajectory will be the trajectory of a real particle (as long as only electromagnetic forces act). As long as the distribution of macroparticles over phase space (position and velocity space) looks fairly continuous—as long as the macroparticle density is high enough—a simulation using macroparticles should be a good approximation to a hypothetical simulation using real particles.

It's worth taking a closer look at the effect of macroparticles on the plasma simulation; I'll consider electrostatic effects only.

6.2.1 Electrostatic Plasma Parameters

A plasma can be coarsely characterized by only a few parameters [10, 21]; these parameters are basically scales of the problem, quantities formed from macroscopic parameters like density and average kinetic energy along with particle charge and mass. With these quantities we can form plasma parameters that set a length scale, a time scale, and a dimensionless parameter that describes how good the plasma is at being a plasma, *i.e.,* at exhibiting the collective behavior we associate with a plasma.

We want to describe the plasma-behavior of a bunch of loose charged particles with charge q and mass m with number density n (charge density $\rho = qn$ and mass density $\mu = mn$), and temperature (defined in terms of the average random kinetic energy, in case no equilibrium is present) T. Without a magnetic field, the important plasma parameters are[5] the plasma frequency (setting a time scale):

$$\omega_p = \sqrt{\frac{nq^2}{\epsilon_0 m}} = \sqrt{\frac{q^2}{m^2}\frac{\mu}{\epsilon_0}} \tag{6.1}$$

the Debye length (a length scale)

$$\lambda_D = \sqrt{\frac{\epsilon_0 kT}{nq^2}} \tag{6.2}$$

[5]This is not the place to re-iterate what can easily be found in the suggested plasma physics texts [10, 21], but briefly:

- plasma frequency: the natural frequency scale with which free charges would oscillate about fixed charges of opposite sign. I think of the plasma frequency as relating to the time it takes a plasma to respond to an electrostatic disturbance. If you want to know how a plasma responds to a disturbance on shorter timescales than the (inverse) plasma frequency, you can't use plasma physics—the plasma doesn't respond as a plasma, it responds as a bunch of individual particles.

- Debye length: the distance over which the electric field is gradually screened out by the plasma particles. This is the distance over which a plasma responds to an electrostatic disturbance. If you want to inquire into phenomena on scales smaller than the Debye length, plasma physics will not help.

- "the" plasma parameter: the (dimensionless) number of particles within a sphere of radius equal to the Debye length. (Also, the plasma parameter is the ratio of average random kinetic energy to typical electric potential energy between neighboring particles.) This measures how good the plasma is at being a plasma; for instance, the shielding of electric field over a Debye length has to be done by almost a continuum of particles—a "good" plasma must therefore have many particles within a Debye sphere, or it won't exhibit the collective behavior expected of a plasma, such as screening.

and the number of particles in a sphere with a one Debye length radius:

$$N_D = \frac{\Gamma_d}{d}\lambda_D^d n = \frac{\Gamma_d}{d}\left(\frac{kT}{\frac{q^2}{\epsilon_0 n^{-1/d}}}\right)^{d/2} \tag{6.3}$$

for d dimensions; Γ_d is the total solid angle is d dimensions ($\Gamma_1 = 2$, $\Gamma_2 = 2\pi$, $\Gamma_3 = 4\pi$). Noting that $n^{-1/d}$ is the characteristic separation between particles, we see that N_D is a (dimensionless) measure of the relative strengths of the random kinetic and (the "pair") electric potential energy. Note that the thermal velocity is proportional to $\sqrt{kT/m} = \omega_p \lambda_D$.

Plasma physics accurately describes systems where: λ_D is much smaller than the length scales of interest (certainly much smaller than the system); $N_D \gg 1$ (there are many particles within a Debye-length sphere, or equivalently, a particle's kinetic energy is typically much greater than its electric potential energy); and time-scales of interest are longer than the inverse of the plasma frequency (for instance, collision frequency should be less than the plasma frequency).

It is not clear that the standard "definition" of plasma (fulfilling the above requirements) delineates the extent of systems that can be described by plasma simulations. However, the extent to which the conditions are fulfilled is the extent to which the elements of a plasma can be considered to exhibit collective behavior; the collective behavior can be characterized by the plasma parameters.

Plasma parameters for macroparticles

Unfortunately, the macroparticle plasma will not mimic the real particle plasma in all respects. We consider here the differences between the real plasma and the macroparticle plasma.

Let each macroparticle contain N particles of charge q and mass m, so that each macroparticle has charge Nq and mass Nm. While the number density of real particles is n, the macroparticle density will be n/N: note that the macroparticle plasma has the same charge density ρ and mass density μ and the same q/m as the particle plasma. The temperature of the macroparticles requires some consideration (by temperature, we really mean average random kinetic energy, $3kT := m\overline{v^2}$, or probably we really mean $m(\overline{v^2} - \overline{v}^2)$). The macroparticles follow the same trajectories as the particles (that's what we want), so they have the same $\overline{v^2}$, but their masses are greater by a factor of N. Considering the velocity distribution, an ensemble of particles with temperature T is better represented by an ensemble of macroparticles with the same velocity distribution, hence a temperature $T_m = NT$. If equilibrium between the particles and something else (*e.g.*, the boundary, or other particles) is important, this could be a problem (though if two different species of particles were in equilibrium at temperature T, the two species of macroparticles could share the same macroparticle number N and still be in equilibrium at temperature NT.)

The macroparticle plasma frequency is the same as the particle plasma frequency:

$$\omega_{p,m} = \sqrt{\frac{\frac{n}{N}(Nq)^2}{\epsilon_0(Nm)}} = \sqrt{\frac{q^2}{m^2}\frac{\mu}{\epsilon_0}} = \omega_p. \tag{6.4}$$

The macroparticle Debye length is the same if the macroparticles have the same root-mean-square velocity (or temperature $T_m = NT$, which is what we will choose, but I display other possibilities for fun):

$$\lambda_{D,m} = \sqrt{\frac{\epsilon_0 k T_m}{\frac{n}{N}(Nq)^2}} \tag{6.5}$$

$$\lambda_{D,m} = \lambda_D \qquad (\text{if } T_m = NT) \tag{6.6}$$

$$\lambda_{D,m} = \frac{1}{\sqrt{N}}\lambda_D \qquad (\text{if } T_m = T) \tag{6.7}$$

$$\lambda_{D,m} = N^{1/d}\lambda_D \qquad (\text{if } T_m = N^{1+2/d}\,T). \tag{6.8}$$

The number of particles in a macroparticle Debye sphere will be: ($\Gamma_1 = 2, \Gamma_2 = 2\pi, \Gamma_3 = 4\pi$)

$$N_{D,m} = \frac{\Gamma_d}{d}\lambda_{D,m}^d \frac{n}{N} \tag{6.9}$$

$$N_{D,m} = \frac{1}{N}N_D \qquad (\text{if } T_m = NT) \tag{6.10}$$

$$N_{D,m} = \frac{1}{N^{1+d/2}}N_D \qquad (\text{if } T_m = T) \tag{6.11}$$

$$N_{D,m} = N_D \qquad (\text{if } T_m = N^{1+2/d}\,T). \tag{6.12}$$

This indicates that usually the ratio of kinetic to potential (of typical near neighbors) energy is lower for the macroparticle plasma, increasing the importance of electric interactions for macroparticles. However, as long as there are still enough macroparticles within a Debye sphere, we need not be too concerned.

6.3 Implementing a Field Emitter

In principle, simulating a field emitter is quite simple: find the surface electric field at the emitter and calculate the field emission current (in our case, conveniently from the Fowler-Nordheim formula) and inject the appropriate number of electrons. (The programming details are somewhat more complicated, of course.)

Testing the basic field emitter, however, we came across some insidious problems relating to the discretization of the simulation; the finite time-step and the finite cell size (of the grid on which the fields are calculated) lead to unphysical effects when the charge of the field-emitted current is large enough that it alters the surface electric field at the field-emitter. The charge of the emitted current

reduces the surface electric field, which reduces the current—a negative feedback loop. In "reality," the feedback loop operates (practically) infinitely fast and over infinitesimally short distances.

In the simulation, the timestep causes a delay in the feedback loop that can create instability (growing oscillations) in the emission current. For example, if at some time, we suddenly apply a large electric field, the simulation calculates the current I corresponding to that field, and injects charge $I\Delta t$ into the simulation within timestep Δt. However, for large enough Δt, that charge might be enough to reduce the electric field to zero, completely suppressing emission for several timesteps, until the charge leaves the area and affects the surface field less. In reality, the charge would have reduced the field emission long before $I\Delta t$ was emitted. The simulation can mimic reality better with a shorter time-step (recalculating the fields more often); however, in some practical situations, that timestep is prohibitively short. This problem was also noticed (and solved) by Knobloch with MASK [56].

The finite cell size in the simulation also causes a delay in the feedback loop. Using a finite grid for field-calculation effectively treats the particles as if they were as large as an entire cell (rather than a point). A charge near (within a single cell-width of) a conducting boundary therefore has a much reduced effect on the electric field (figure 6.2); that is, the simulated field produced by emitted charge is unphysically small until the charge is more than one cell-width away from the conducting wall. Especially for small time-steps, the charge will not move very far away from the wall in a single (or even a few) time-steps, and the field at the emitter will not be reduced as much as it should be (or would be in reality), and more charge will be emitted than should be. As the emitted charge moves away

from the boundary, its effect on the field grows to what it should be, but because too much charge was emitted, it can actually reduce the surface field to zero (or even change it to the opposite direction).

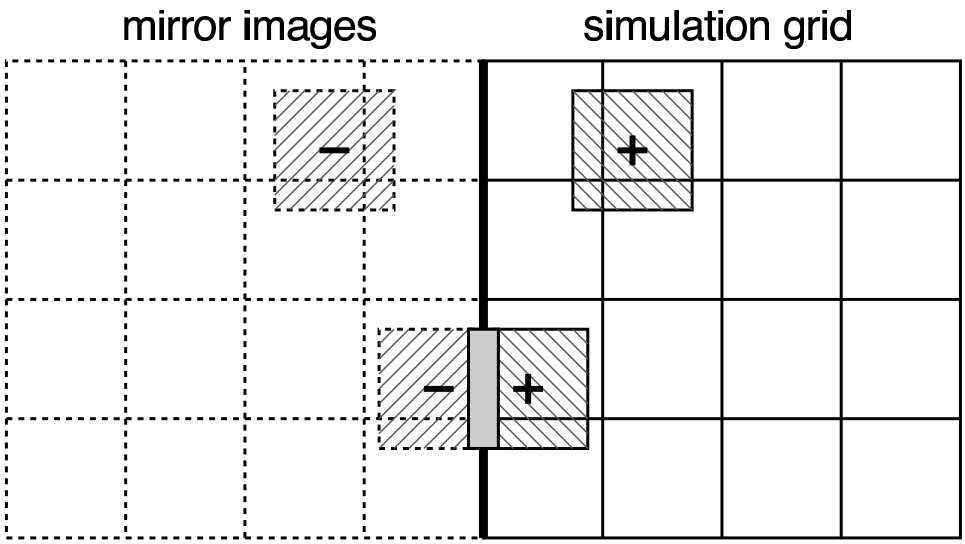


Figure 6.2: The finite simulation grid reduces the effect of a charge near a conducting boundary; calculating the field on a grid treats what should be point charges as charges with the size of a grid cell. The field of a charge near a conducting boundary can be thought of as the field produced by the charge and its mirror (negative) image. When a charge gets close to the conducting boundary, the charge and its image overlap significantly, canceling out most of the charge. Not until a particle moves one cell away from the boundary does its entire charge get counted toward the field.

Making the cell-size small and the time-step even smaller is the "physically correct" solution to these problems. In cases where the field emission current approaches the Child-Langmuir limit,[6] the time step and cell size must be far too small for computation in a reasonable time. The practical solution to the timestep problem is an artificial current limit that keeps the emitted charge from a single timestep sufficiently low that it cannot greatly alter the surface electric field. To

[6]See [63] for a simple treatment of Child-Langmuir behavior in a field-emission context; see also [22] for Child's original article featuring a simple derivation, and Langmuir's more thorough article [61].

solve the cell-size problem, electrons can be emitted not from the boundary, but from one-half to one cell-width away; as long as the cells are reasonably small, and the electron wouldn't have enough energy to do anything unusual (like ionizing a neutral atom), this is a pretty reasonable solution. I added this feature to OOPIC in such a way that the program actually followed the trajectory of a hypothetical "test electron" out to one cell-width, and injected the "real electron" there with the appropriate velocity; clearly this scheme runs into problems in certain circumstances (for example, if the trajectory never gets one cell-width away from the emitter), but in simulations I ran, no serious problems occurred.

6.3.1 Transverse Velocity Spread

If field emission occurs at a protrusion that enhances the surface electric field, then electrons will not be emitted normal to the (macroscopic) cathode surface; by virtue of the field enhancement, the electric field lines spread laterally from the emitter, and many emitted electrons will acquire a transverse velocity within a short range of the emitter (see appendix B). To simulate the transverse spread of electrons due to field enhancement, we sometimes gave electrons an initial transverse velocity distribution corresponding to simulated emission from a protrusion.

6.3.2 Field Solver Precision

In early tests of the field emitter in which emitted space charge affected the surface field at the emitter, we discovered that for reproducible results we needed to increase the precision desired of the field solver, which required double-precision computation. Interestingly, we found the simulation ran significantly faster in double-precision than in single-precision (I believe this occurs because the com-

puter does arithmetic in double-precision, disregarding how the result is stored; the advantage of single-precision lies in reduced memory use).

6.4 Implementing Mobile Neutral Gas

OOPIC already had algorithms for handling certain collisions (between electrons and neutral atoms, and ions and neutral atoms of certain species), but, while explicitly tracking electrons and ions, OOPIC did not track neutral particles; it merely accepted a temperature and density (as a function of position, but not time), and calculated the probability of electrons and ions colliding with neutrals.

I made trivial modifications to allow OOPIC to handle neutral particles (mostly a matter of handling the few situations where $q = 0$ would cause problems), and then I created a new NGD (neutral gas density) class, which calculated the neutral gas density within each grid cell based on the number of neutral gas particles (of a given species) within each cell, and used that to calculate collision probabilities. In addition, I allowed the subtraction of neutral particles after ionization events (since creating an ion means "destroying" a neutral). The objected-oriented design of OOPIC made this all amazingly easy (for the most part) to change. With this change, neutral particles could be created (*e.g.,* emitted from a boundary) and other particles (or even the neutrals themselves) could access the density of neutrals to calculate collision probabilities.

6.5 Validating OOPIC

Although OOPIC had been used and validated by others [17], we nonetheless wanted to verify its capabilities for simulations relevant to voltage breakdown.

6.5.1 Electron Flow

To test the accuracy of OOPIC, especially of the newly implemented and relatively untested field emission routine, we ran simulations of a number of situations involving one-dimensional electron flow—solving Poisson's equation self-consistently with charged-particle motion is much easier in one dimension. Although OOPIC has no explicit option for one-dimensional problems, we used both 2D cartesian and 3D cylindrical geometries with rather wide transverse extents, like a parallel plate capacitor with large enough area that fringing fields can be neglected; we also used the 2D cartesian geometry with periodic boundary conditions—in any case we ensured that neither the extra-dimensionality nor the boundary would play much role, so the simulation really reflected the 1D problem. In short, OOPIC closely matched analytical solutions.

The first trial was the Child-Langmuir problem ([22, 60] are the original articles by Child and Langmuir, but a much shorter paper more to the point for our purposes is [63]): space-charge limited 1D electron flow across a fixed potential difference (electrons in a parallel plate capacitor with fixed voltage). The Child-Langmuir problem addresses the space-charge limit; in the space-charge limit, space charge limits current from an over-abundant source, the details of which therefore need not be considered; the space charge builds up until it reduces the field at the source to zero—any more space charge would push the field negative, completely choking any current flow. The maximum current (density) that can flow is given by the Law of Child and Langmuir: for a gap length D and voltage V (and particle charge e and mass m)

$$J \leq J_{\mathrm{CL}} \;=\; \frac{4\epsilon_0}{9}\sqrt{\frac{2e}{m}}\frac{V^{3/2}}{D^2} \;=\; \left(2.3\times 10^{-6}\ \mathrm{A/V^{3/2}}\right)\times\frac{V^{3/2}}{D^2}. \qquad (6.13)$$

Up to the proportionality constant, the dependence of J_{CL} on $V^{3/2}/D^2$ can be derived from dimensional analysis and scaling considerations.

Although we most frequently refer to this maximum current, solutions for currents J less than the maximum can be just as easily found, yielding the velocity and number density as a function of position—the complete self-consistent solution—depending on the gap length D, voltage V, and the current density J. We have found that after a simulation runs long enough to achieve steady-state, the potential (as found by OOPIC) matches the theoretical value.

On the other hand, we also considered the problem where the current is limited by field-emission. The current density emitted by a Fowler-Nordheim field emitter subjected to electric field E_s is

$$J_{\mathrm{FN}}(E_s) = J_0 \frac{E_s^2}{E_0^2} e^{-E_0/E_s} \tag{6.14}$$

where J_0 and E_0 depend on the emitter shape and material. As long as space charge has a negligible effect on the electric field, $E_s = V/D$, and the expected current can be trivially calculated from the above formula.

For low currents, the field is $E_s = V/D$ and the current can be calculated from the Fowler-Nordheim equation; for very high currents (up to the highest possible), the current is determined by Child-Langmuir (and Fowler-Nordheim is irrelevant, except that field emission must be able to supply more than J_{CL} at fields far less than V/D for Child-Langmuir to apply). Between the two limits (the focus of [63]), one has to solve the Child-Langmuir problem to get the cathode surface electric field E_s in terms of J and simultaneously solve:

$$J = J_{\mathrm{FN}}(E_s) \tag{6.15}$$

(the knowns are V and D; the unknowns are E_s and J, and we have two equations, one for E_s in terms of J, one for J in terms of E_s. The solution can be found numerically (a root-finding problem) for specific cases. Again, OOPIC found the same solution.

More interesting is the case similar to the Child-Langmuir problem described above, but allowing the electrons to have a non-zero initial velocity at the cathode; experimentally one might have a transparent cathode (a grid) with an electron gun behind it, shooting electrons into the gap with the desired velocity. Working out the solution [9], I was surprised to find that for sufficiently high currents, there are *two* steady-state solutions (depicted in figure 6.3).

The second steady-state solution (when it exists) is not stable [9]—after time, even small perturbations will cause the system to stray far from the steady state, but the first solution is stable. OOPIC, in specific cases tried, relaxed into the stable solution, which agreed well with theory (figure 6.4).

Electron flow in $D \geq 1$

With OOPIC's 2D cartesian geometry and 3D-cylindrically-symmetric geometry, we simulated the transition from Fowler-Nordheim to Child-Langmuir for a point emitter in 2 and 3 spatial dimensions. The results are shown in figure 6.5.

6.5.2 Ionization

OOPIC performs ionization based on a Monte Carlo method; it handles collisions in a remarkably "realistic" manner, especially regarding the electrons and ion (the neutral particles aren't treated so individually). For example, an electron can collide with neutrals in different ways, with cross sections depending on the elec-

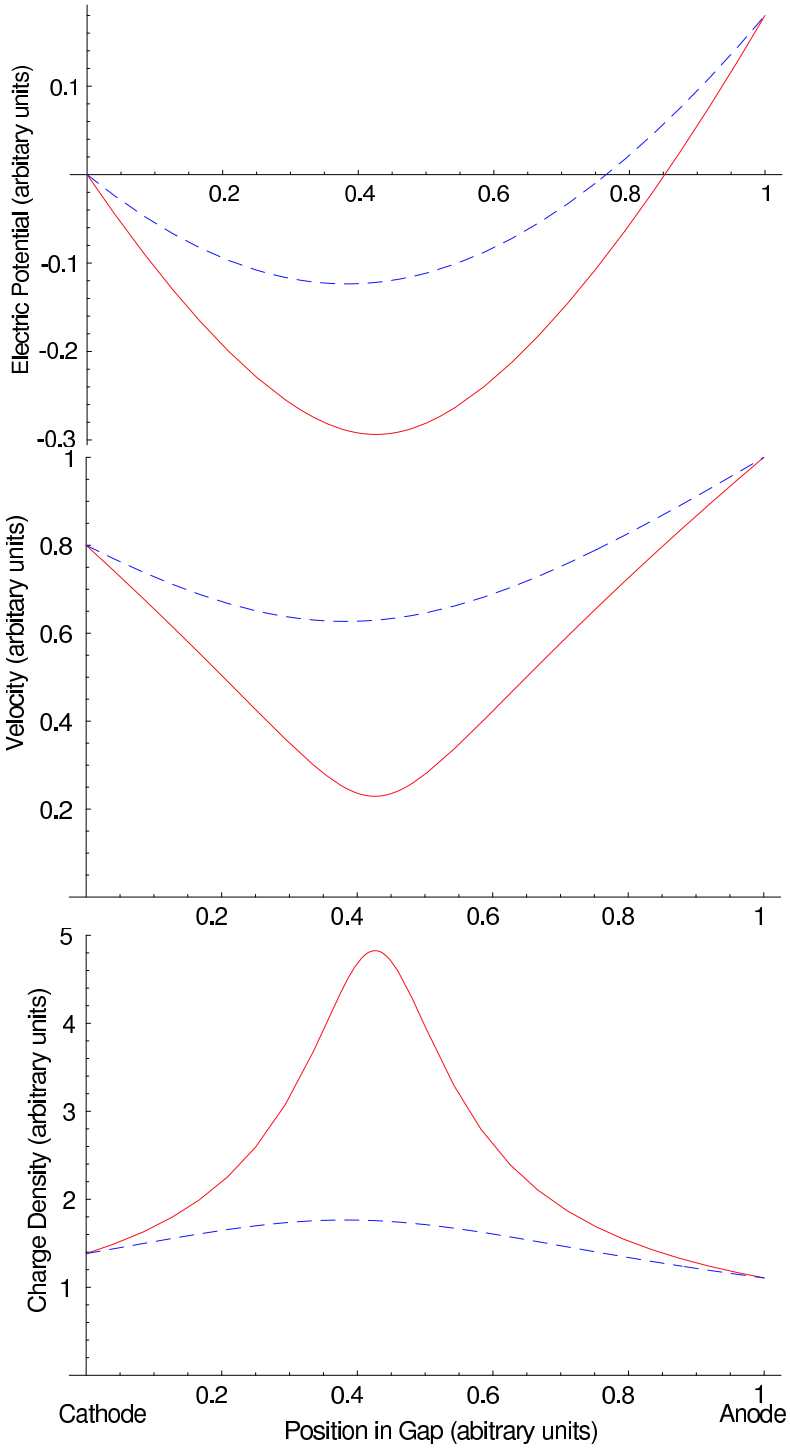


Figure 6.3: The potential distribution (top), velocity (middle), and charge distribution (bottom) of electrons as they cross the gap for the two steady-state solutions (with the same J and V). The non-dashed curve has more space charge in the gap, which slows down incoming particles more (and then accelerates them more once they cross the potential hill).

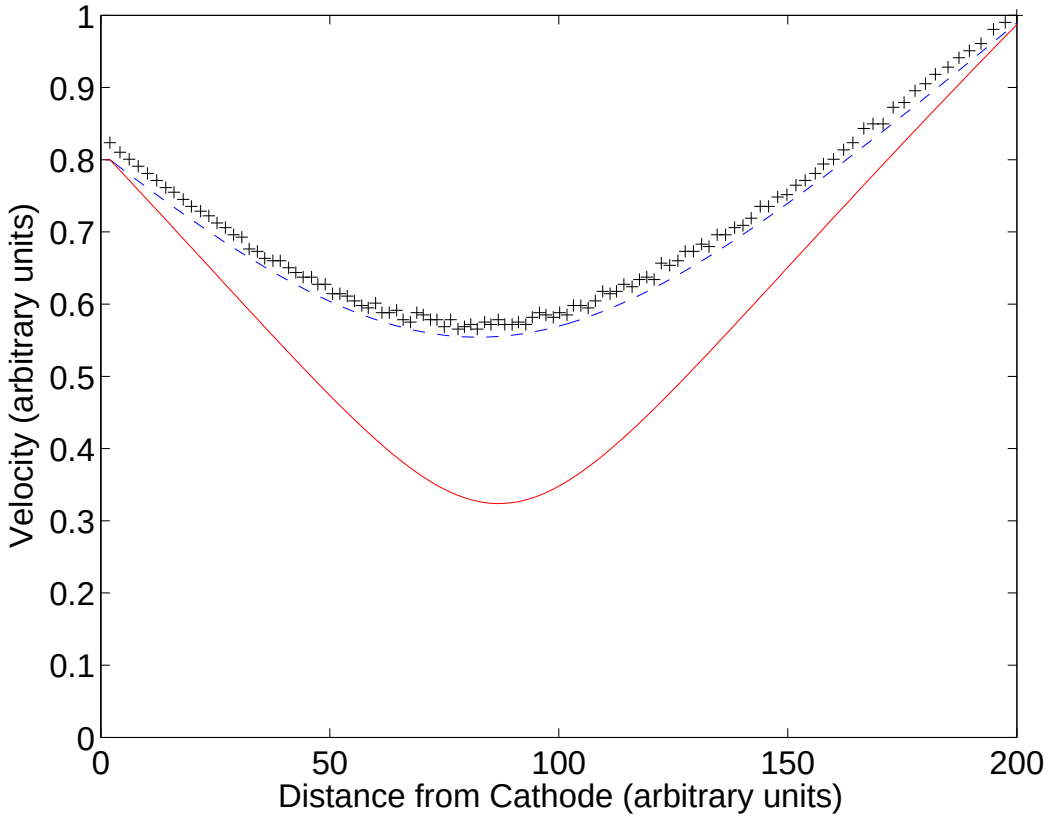


Figure 6.4: The velocity of electrons versus gap position, (lines) according to theory, and (points) corresponding to particles in OOPIC simulation.

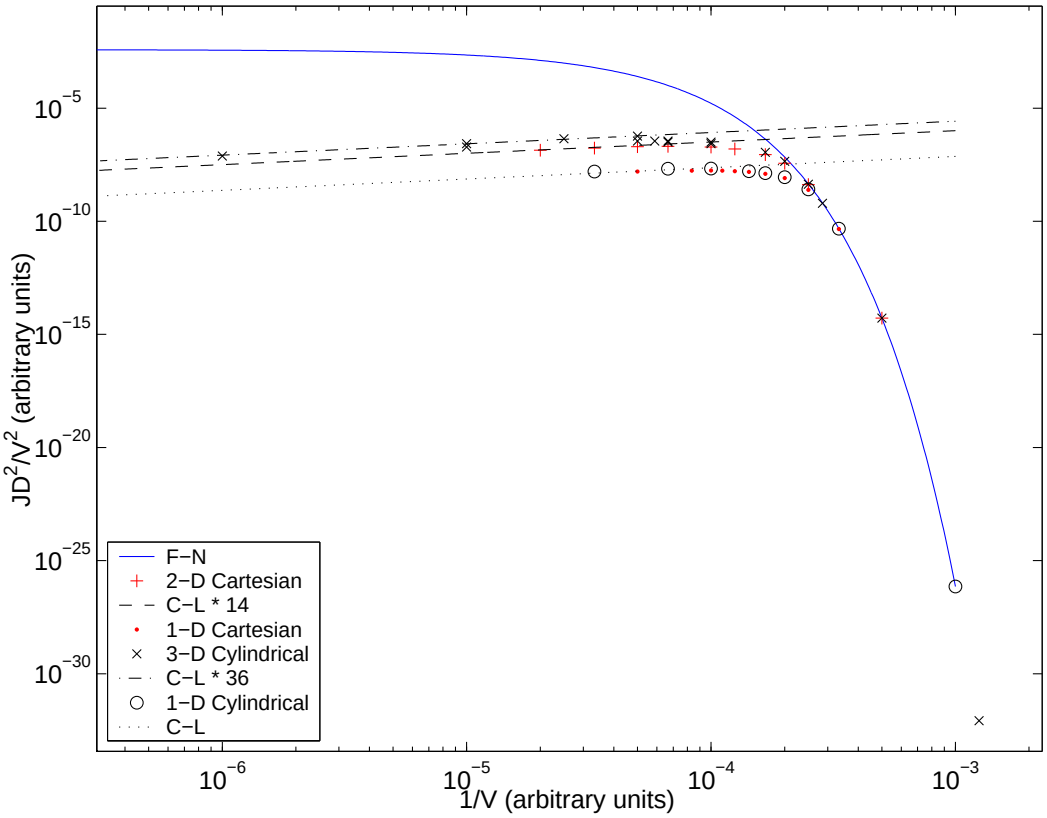


Figure 6.5: Current density versus voltage (Fowler-Nordheim style) for a point field emitter in 1, 2, and 3 dimensions (using both cylindrical and cartesian geometries to simulate one dimension), as simulated by OOPIC, overlaid with a solid blue curve depicting electron emission that is purely Fowler-Nordheim (no dimensionality dependence) and dashed lines showing purely Child-Langmuir (same slope, with prefactor dependent on dimension). As the voltage increase, the data points switch from the Fowler-Nordheim to the Child-Langmuir curves. The prefactors for the Child-Langmuir limits in 2 and 3 dimensions were not compared to analytical calculations.

tron's energy: for momentum-transfer or excitation collisions, the momentum of the electron is altered appropriately; for an ionizing collision, the momenta of both the incoming and the ionized electron as well as that of the ion are appropriately calculated, based on a probabilistic guess of the impact parameter and the velocity of the neutral atom.

Each timestep, OOPIC estimates the maximum possible number of electrons that could collide with the neutral gas (using the maximum collision cross section and the maximum gas density), and then randomly picks that number of electrons (which should be much smaller than the total number of electrons, but larger than the number of electrons that will actually collide), and for each chosen electron calculates the probability of various types of collisions (excitation, ionization, *etc.*) given the electron's energy, and rolls a die to see whether it collides, and if so, what kind of collision occurs. If the collision includes ionization, OOPIC creates an ion and another electron at the position of the electron, and again takes into account the electron's initial momentum to calculate the momenta of the scattered electrons and ion.

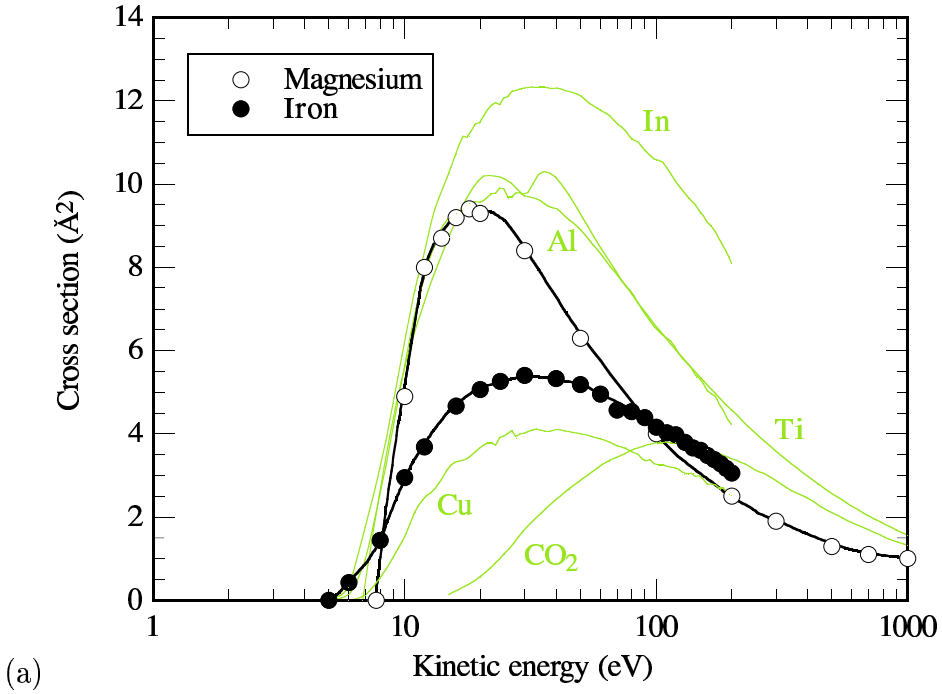
We used the argon collision model already built into OOPIC. Although it adding custom cross-sections poses only minor difficulties, ionization cross-sections for different gases tend to be quite similar (see figure 6.6a), and given what little we actually know about gas emitted before breakdown, using the specific hydrogen or niobium (for example) cross-section would involve more vanity than accuracy. We are not yet ready to address the effects of the details of the energy-dependence of the cross-section; we are concerned only with its general features: a threshold at the ionization energy, about 15 eV, rising to a peak value about the size of an atom (1–10 Å^2), and decaying at energies higher than about 100 eV. Figure 6.6b

depicts the cross-section for various collisions of electrons with argon ions used by OOPIC.

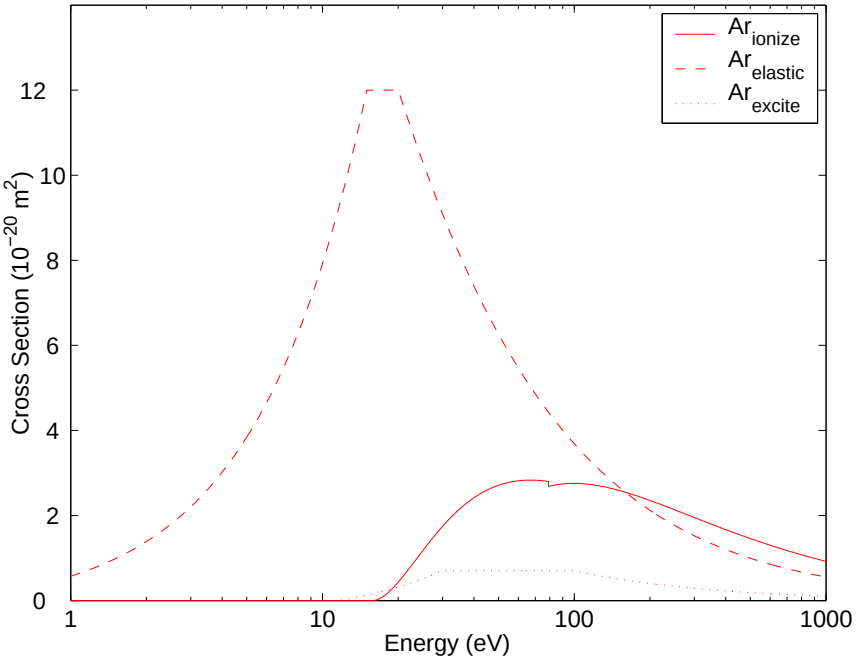
To ensure the ionization algorithms worked as advertised, we first checked some basic things—whether path-lengths between collisions in a uniform gas looked reasonable, *etc.* Because OOPIC alters the momenta of scattered electrons (and displays particles graphically as the simulation progresses), it's easy to spot collisions while watching a simulation run.

We then applied OOPIC to electron avalanches and Townsend Discharges (see [32] for an overview). Electron avalanches can occur in fairly dense gas ($\sim$1 atm) under an electric field. Initially an electron must be created by *deus ex machina*, like ionization by cosmic rays in experiment, or simple injection in simulation. Due to the electric field, the electron begins to cross the gap; it collides with the neutral gas and ionizes it; those electrons travel another mean free path and ionize more, *etc.*, doubling the number of electrons (and ions) at each step.

A single avalanche, started by *deus ex machina*, is a lone blip of current; an avalanche grows into a Townsend discharge by means of secondary emission of new electrons from the cathode—many sources contribute to this secondary emission, but most important are ions, which bombard the cathode and emit electrons, and photons, which emit electrons through the photoelectric effect (photons are emitted either from atoms excited by collisions with electrons or by the impact of electrons on the anode). If the number of electrons generated by an avalanche times the number of electrons (secondarily) emitted at the cathode per each new electron or ion produced, then the current will continue to grow. The Townsend "regime" encompasses this continuing and increasing series of avalanches, up to the point



(a)



(b)

Figure 6.6: (a) Cross-sections for ionization by electron impact for various gases, reproduced from Knobloch's dissertation [56]. (b) Cross-sections for different collisions with neutral argon used in OOPIC (the solid line shows ionization cross-section, the dashed line shows the elastic scattering cross-section, and the dotted line shows the excitation cross-section).

where the space charge begins to affect the electric field significantly; space-charge is neglected in Townsend discharge theory.

Because the Townsend discharge depends on a good avalanche, it won't occur for pressures too high or too low: if the pressure is too low, electrons won't collide as they cross the gap, and if the pressure is too high then electrons will collide before they gain 15 eV (or whatever is the ionization energy) and they will not be able to ionize gas.

6.6 Breakdown Simulations

A simulation can never be cross-checked too much. The parameters of the simulation were therefore chosen to match those used with the program MASK by Knobloch [56] in his dissertation, so we could compare results.

6.6.1 Simulation Parameters

Geometry

We simulated voltage breakdown in a cylinder, with conducting plates on the two ends, and dielectric boundary conditions around the circumference, applying a voltage (either DC or AC) across the two ends (figure 6.7). At the zero-voltage end (on the left) we put a field emitter and a neutral gas emitter.

OOPIC does not handle fully 3-dimensional problems, so we could only look for cylindrically symmetric solutions. The simulation universe is the z-r half-plane for $r > 0$. A simulated point charge at (Z, R) in the plane really represents a ring of charge at $z = Z$, centered on the cylinder axis with radius R. When looking at particle densities in the simulation, it's important to remember that because of

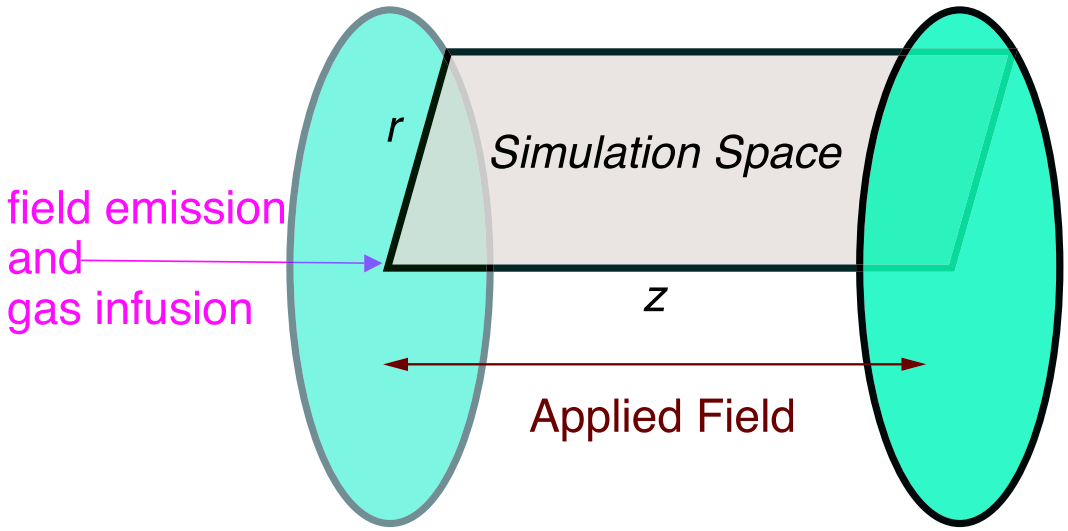


Figure 6.7: Breakdown was simulated within a circular capacitor with cylindrical symmetry, placing a field emitter and a neutral gas emitter at the cathode.

the cylindrical symmetry, a uniform density of charge within the cylinder appears to have an increasing density as r increases in the z-r plane; the real density at, say, $r = R$ is the number of points charges at $r = R$ divided by $2\pi R$.

For finding the electric field, we normally used a grid of 128 by 64 (number of cells along the z-axis by number of cells along the r-axis). Grids much larger than this result in excessive computation time, run on a dual-processor 2 GHz Intel Xeon linux box.

Since submicron length scales are important,[7] simulating the entire region of experiment (for DC, 100 microns; for RF, many centimeters) was out of the question. However, for studying cathode-initiated breakdown, most of that region is irrelevant—only the area near the field emitter counts. Therefore, we chose a cylinder length of 32 microns and radius of 8 microns, resulting in mesh cells of size $\Delta z \times \Delta r = 0.25\ \mu\text{m} \times 0.125\ \mu\text{m}$.

[7]For instance, with a reasonable field of 30 MV/m, or 30 V/μm, an electron would take about a half micron to gain 15 eV, typical ionization energy.

Field emitter

The current density for a Fowler-Nordheim field emitter with field-enhancement factor β, in macroscopic field E (hence local surface field βE) is

$$J = J_0 \left(\frac{\beta E}{E_0}\right)^2 e^{-E_0/(\beta E)}. \tag{6.16}$$

Usually the parameters J_0 and E_0 are written [87] in terms of parameters $A_{\rm FN}$, $B_{\rm FN}$, and the material work function ϕ:

$$E_0 = B_{\rm FN}\phi^{3/2} \qquad\qquad J_0 = \frac{A_{\rm FN}}{\phi}E_0^2. \tag{6.17}$$

Values are:[87]

$$A_{\rm FN} = 1.54 \times 10^{-6}\,({\rm A/m^2}) \cdot \frac{\rm eV}{({\rm V/m})^2} \tag{6.18}$$

$$B_{\rm FN} = 6.83 \times 10^9\,\frac{({\rm V/m})}{({\rm eV})^{3/2}} \tag{6.19}$$

. We used $\phi = 4$ eV, a reasonable value for niobium[8] and many other metals.

The field emitter emitted electrons from a disc of radius 0.25 microns (2 cells) concentric with the cylinder; however, the total amount of current emitted was chosen to be the current that would be emitted from a hypothetical emitter with area $S = 3.37 \times 10^{-14}$ m^2 and field enhancement $\beta = 250$. Such an emitter is realistic (see figure 4.1), though not the most likely. We chose these parameters to match Knobloch's simulations. In summary, the current density over the 0.25 micron radius emitter, in terms of the surface field (calculated at 3 points along the emitter and interpolated) was:

$$J(E) = \frac{3.37 \times 10^{-14}{\rm m}^2}{\pi(0.25 \times 10^{-6}{\rm m})^2}\frac{A_{\rm FN}}{\phi}(\beta E)^2 e^{-B_{\rm FN}\phi^{3/2}/(-\beta E)}. \tag{6.20}$$

[8]The niobium work function varies from 3.95 to 4.87 eV for different crystal facets, with an average around 4.3 eV [69]. Of course, surface adsorbates will affect the work function somewhat.

With $\beta = 250$ and $E = 30$ MV/m, the current density would be 2.6×10^9 A/m^2, and the total current from an area of $\pi(0.25\mu\text{m})^2$ is about 500 μA. For a sinusoidally oscillating field, the average current would be about 60 μA. Note, however, that an applied field of 30 MV/m does not result in a surface field of 30 MV/m because of the charge of the emitted current, which reduces the field. In practice, with such field and currents, the actual field-emitted current is reduced by a factor of 3 to 10 (see [56]).

As discussed in section 6.3, some rather artificial changes had to be made to the field emission to make the field emission more realistic given the finite timestep and mesh size. Electrons were emitted one mesh cell (0.25 microns) from the cathode, with the corresponding initial velocity (which would increase traveling across the first mesh cell). Also, an artificial current limit was applied, so that the current density could never exceed

$$J \leq J_{\max}(E) = 0.1 J_{\text{big}}(E) = 0.1 \times \frac{2\epsilon_0 E}{\Delta t} \tag{6.21}$$

where Δt is the timestep and E the surface electric field. J_{big} is the current density, which if injected uniformly from an infinite plane-cathode for time Δt, would inject an amount of charge that would create a field E. In other words, injecting J_{big} for time Δt would decrease the surface field by E. Since real field emission would be greatly suppressed if the surface field were reduced by even a tenth, $J_{\max} = 0.1 J_{\text{big}}$ seems reasonable. Note that the smaller the timestep, the larger $J_{\max}$, and the less likely that this artificially limiting will be applied. In breakdown simulations, this limit was not reached (or imposed) until well into the current explosion.

Neutral injection

The weakest point of the simulation of breakdown, insofar is it reflects reality, is the injection of neutral gas. For instance, although field emission is not fully understood (like why β could be 250), we can at least, by choosing the empirical parameters S and β, make the simulated field emitters resemble those in experiment; whereas, the desorption of gas from the area around the field emitter is not understood and has not been measured. In light of this, the simulation does not answer the question: "How does breakdown happen?" but rather: "How much gas would have to be desorbed from the field emitter to induce breakdown?" or, "Assuming such-and-such an amount of neutral vapor were desorbed from the field emitter, what would happen?"

Following Knobloch [56] we injected neutral atoms from the area around the field emitter as if the emitter were heated to some very high temperature, so that material simply evaporated from the surface. We used a similar flux of neutrals from the surface: $10^{27}/\mathrm{m}^2/\mathrm{s}$. The neutrals were emitted with a velocity distribution as if they had a temperature of 0.25 eV (about 3000 K). We chose argon atoms as the neutral gas, arguing that the details of the specific species were irrelevant compared with the uncertainties in the source and amount of gas. Note that argon has mass 40; other likely candidates: niobium has mass 93, hydrogen mass 1, carbon monoxide mass 28, oxygen mass 32. Ionization cross-sections fall within an order of magnitude for the different species (see figure 6.6).

For argon, the flux of $10^{27}/\mathrm{m}^2/\mathrm{s}$ at 3000 K produces a pressure of nearly 500 torr (more than a half atmosphere) near the emitter, where the number density of atoms reaches about $1.6 \times 10^{24}\ \mathrm{m}^{-3}$ (at room temperature, air would have a similar number density at about 50 torr).

Neutrals were emitted with the specified flux from an a disc of radius 1 micron concentric with the field emitter. Once injected, the neutrals moved without collisions—that is, in a straight line with their initial velocity. Breakdown occurs so quickly (electrons being so much more mobile than atoms) that the neutrals hardly get more than a micron or two away from the cathode.

When an electron was deemed to ionize a neutral atom (creating a new electron and ion), one neutral atom was removed from the cell in which the ionization occurred.[9]

Timestep

For the simulation described above, we generally used a timestep of 5×10^{-15} s, half of that used by Knobloch [56]. In 5 fs, light travels 1.5 microns, or several mesh cells; an electron with less than a few keV travels less than a mesh cell in a single timestep. Fully electrodynamic fields would require a smaller timestep for stability, so light couldn't cross a mesh cell in a single timestep, but with the electrostatic field solver only, a slightly larger timestep is possible.

The basic timestep of 5 fs is the period with which the electric field was calculated (as well as various diagnostics, such as total kinetic energy of electrons or current hitting the right wall). However, electrons were moved four times within this interval for more accurate calculation of their trajectories amid spatially varying fields, and also to allow the possibility of an electron ionizing more often; that is, electrons moved with a sub-timestep of 1.25 fs, and only after every fourth sub-timestep were the fields recalculated. The ions, with their much greater mass,

[9]To avoid possible selection bias in removing neutrals in the order (or reverse order) that they were created, the program actually removed neutrals in a random fashion, so that on average, one neutral atom was removed for every ionization event.

move much slower than electrons, there's hardly any point to moving them with a timestep of 5 fs—the amount an ion would move in that time would have a completely trivial effect on the electric field. Ions therefore moved only every twenty timesteps, or every 100 fs—even a 5 keV hydrogen ion would move only 0.1 microns in this time. Neutrals, just as heavy as ions, but without a charge to accelerate them to higher speeds, moved every hundred timesteps, or 500 fs, far oftener than necessary, but moving them every hundred timesteps already consumed such a small fraction of the computing that increasing the period would have yielded no benefit.

At the very beginning of a simulation, the neutral gas is just starting to emerge from the cathode; nothing interesting can happen until the neutrals reach the 15 V equipotential, since electrons don't have enough energy to ionize them before that. Therefore, we often saved time by running simulations for the first few hundred picoseconds (or until the first ions started to form) with a much larger timestep, *e.g.*, 200 fs, and then reduced the timestep to 5 fs.

Macroparticles and particle limits

As discussed in section 6.2, simulating every electron and every ion would require an impossible amount of computation; therefore, bunches of electrons with similar trajectories are grouped together into macroparticles. The simulation handles these macroparticle bunches as if they were single particles. Typically, one wants at least several macroparticles per mesh cell, but too many macroparticles per mesh cell wastes computational effort.

For simulating the initial stages of breakdown, we generally used macroparticles of size 10 for electrons and ions (*e.g.,* 10 electrons per macro-electron). During a

breakdown event, the electron current grows enormously, from microamps to amps, requiring a large increase in the number of particles. To keep the simulation from grinding to a halt as it tries to simulate millions of particles, the macroparticle size can be increased.

Although OOPIC provides slightly different alternatives for increasing macroparticle size, we usually set a limit on the total number of particles—usually about 30,000. When the total number of macroparticles reached 30,000, OOPIC would delete half of the particles and double the macroparticle size of the remaining particles.

The higher initial density of neutral gas meant that a macroparticle size of 10^4–10^5 made more sense; even so, the density of neutral macroparticles tended to relatively high (within the few cells they reached during the simulation) because I was able to streamline the movement of neutral particles so they used very little computation time.[10]

Collisions

Electrons were allowed to collide with neutral gas, but for the sake of simplicity and saving computation time, we did not allow ions to collide with neutral gas, nor neutral atoms to collide with themselves. To determine the probability of an electron colliding with a neutral atom, we used the cross-sections for argon, which qualitatively resemble those of other species (see figure 6.6).

[10]The neutral atoms, as heavy as the ions and much heavier than electrons, travel more slowly than either electrons or ions, since they're subject to no forces. Therefore, calculation of their trajectories is very simple, and furthermore doesn't need to be done very often (compared to electrons).

6.7 DC Simulation Results

A typical simulation begins with just field emission, perhaps 100 μA, as neutral gas relatively slowly expands from the area around the field emitter. An applied voltage imposes a uniform electric field of 30 MV/m. The electron beam travels almost straight across to the anode, widening slightly because of transverse velocity spread and space-charge repulsion (figure 6.8). When the neutrals reach the 15 V equipotential, electrons begin to ionize them, creating ions that slowly move toward the cathode, where they are absorbed, while the electrons move much faster toward the anode. Because the ions hang around longer, a cloud with net positive charge builds up in front of the cathode (figure 6.9). If the electrons create ions faster than the ions can reach the cathode, then a cloud of ions develops, and begins to increase the surface electric field at the emitter, which in turn increases the electron current. More electrons create more ions, further enhancing the field, and leading to a runaway current.

As it forms, the ion cloud becomes a trap for electrons that lose energy through collisions, and an electron cloud grows with the ions—a plasma cloud. Although electrostatics tries to force the cloud to become neutral, thermal motion and the higher mobility of electrons compared to ions makes the electron cloud large, and electrons escape the cloud much more easily than ions, leaving the cloud positively charged in the center. The net positive charge starts to repel the ions, expanding the cloud.

The ion cloud (with its net positive charge) in front of the emitter forms an electric potential hill—a potential energy hill for ions, but a potential energy well for electrons. As the height of the hill (or depth of the well, depending on the perspective) grows, it noticeably affects the trajectories of electrons, tending to

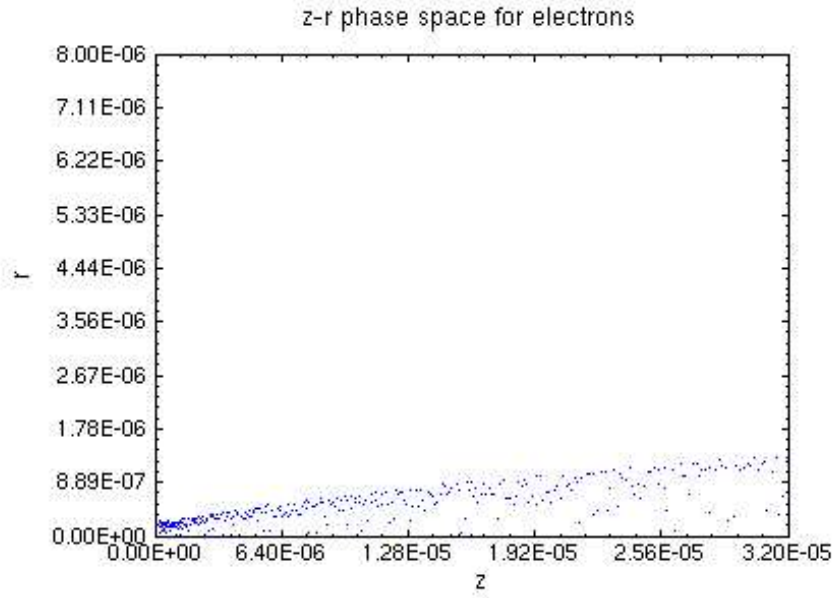


Figure 6.8: Simulated field emission beam of electrons. Dimensions are in meters.

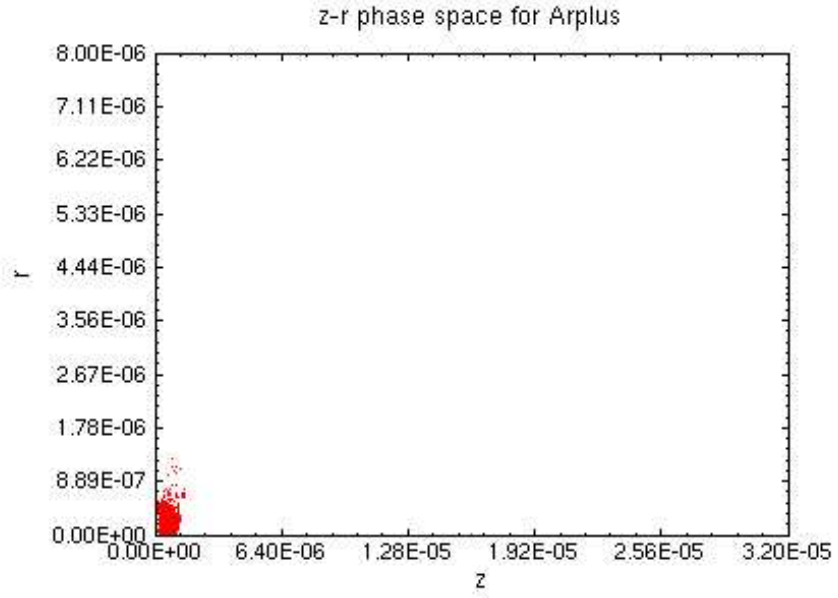


Figure 6.9: When the neutral gas gets far enough away from the cathode that electrons have enough energy to ionize them, an ion cloud grows.

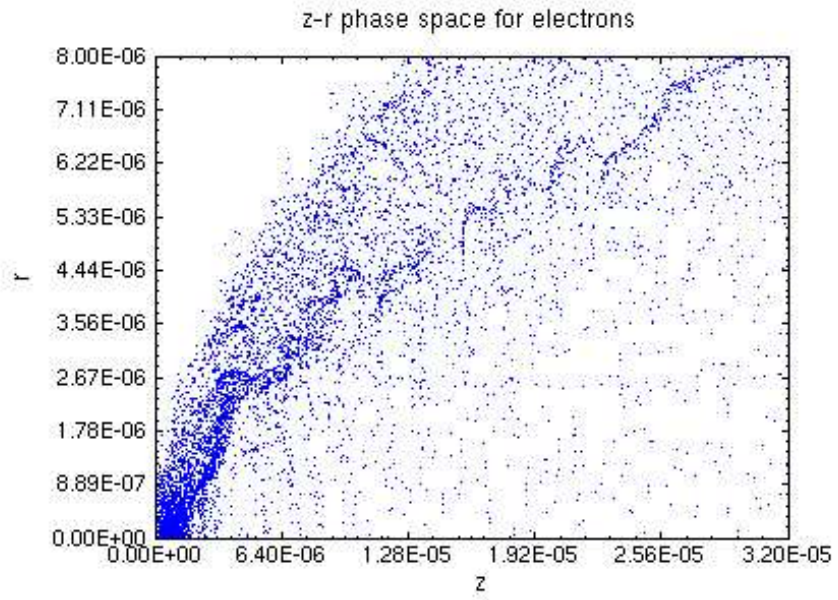


Figure 6.10: As the ion cloud in figure 6.9 grows in (positive) charge, it scatters electrons laterally. Although the figure doesn't make it clear, emitted electrons are crossing the axis of symmetry (the lower boundary) before leaving the ion cloud. The aspect ratio in this picture is not 1:1, so the beam leaves the cathode at an angle closer to perpendicular than it appears.

scatter electrons laterally (parallel to the cathode), until they leave the ion cloud and the applied electric field bends them back toward the anode (figure 6.10).

Once the field due to the ion cloud is comparable to the applied field, very few electrons can travel straight through—to go straight through a potential energy well, a particle either needs to have relatively high energy or be very close to the axis of symmetry. If the ion cloud field is indeed comparable to the applied field, then the electron energy comes mostly from the ion cloud potential, and electrons barely escape the cloud only because there is some applied field. As for following the axis of symmetry, that's difficult to do in three dimensions—any transverse velocity quickly leads away from the axis, and there are many ways to gain transverse velocity, such as being emitted at an angle from a protrusion or

being repulsed by the charge of other emitted electrons, besides having thermal velocity in the transverse direction.

Allowing the electrons to scatter off the ion cloud as a whole, as if the charge in the ion cloud were continuous, is not entirely realistic. The simulation ignores variations in potential smaller than a mesh cell, effectively smoothing the potential of the ions so that an electron can pass smoothly through the ion cloud. In reality, electrons could scatter off individual ions (Coulomb scattering), possibly into large angles; however, section 6.9.2 shows that such Coulomb scattering might be negligible.

In the simulations, the electron beam first starts to bend when the electron current starts to increase due to the increased surface field caused by the ion cloud; and the beam bends more as the current increases. As the electron current increases from 100 μA to a few milliamps, the beam bends from almost straight perpendicular to almost parallel with respect to the cathode. In RF simulations, where the applied field changes, the electron beam bends more and more as the applied field decreases from its peak—in other words, as the ratio of the ion cloud field to the applied field increases, so increases the bending of the electron beam, consistent with the potential well scattering explanation.

Of course, increasing the electron current proportionally increases the ionization rate. As the electron current increases through another order of magnitude, the ion cloud, and therefore the surface field, and therefore the field emission all increase at an enormous rate; the potential hill of the ion cloud grows too, and picoseconds later (with current around 100 mA), the ion cloud starts to expand much faster—more of an explosion, although the ion cloud front expands quite uniformly, remaining nearly spherical (figure 6.11). Before the ion cloud expands

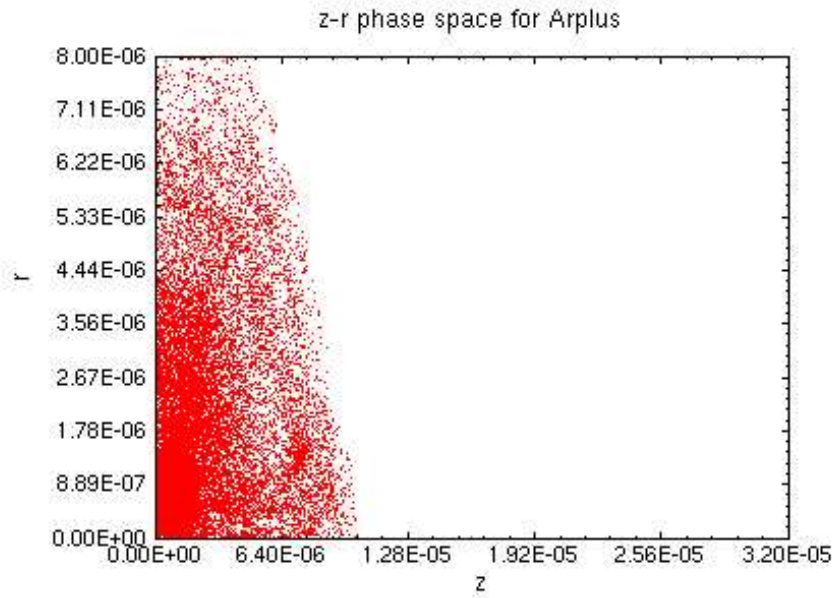


Figure 6.11: As the number of ions increases, the rate of ionization increases; quite suddenly, the charge in the ion cloud becomes far too large, and self-repulsion expands the ion cloud, creating a sort of explosion.

ten microns from the cathode, the current (of electrons hitting the anode) grows to 1 amp; the total current in the gap is somewhat higher than the current measured by OOPIC (which ignores electrons that escape on the side).

The time between the start of field emission growth due to field enhancement and when the current reaches one amp is about 300 ps. The current continues to grow rapidly to about 10 A, and then seems to taper off. At this point, I'm not sure how much to trust the simulation: the ion cloud has expanded past the boundaries of the simulation, so the simulation no longer includes all the important parts; furthermore, the simulation still maintains 1 kV across the gap, despite the 10 A current (that's a mighty power source).

While there are always doubts about the accuracy with which a simulation portrays the real physics, these simulations demonstrate fairly convincingly how field

emission plus neutral vapor emission can lead very quickly to voltage breakdown (with currents of many amps).

During the period of rapid expansion of the ion cloud, the radius of the ion cloud increases from about 2 microns to 8 microns in about 25 μs. The front speed is therefore about 250 μm/ns—a few orders of magnitude above room temperature thermal velocity. For mass 40 argon ions, that speed represents 13 keV kinetic energy; of course, that's the tail of the velocity distribution (the fastest ions are at the front), but it sounds unphysically high. The combination of finite time-step and field-emission so steeply increasing with field would likely allow the ion cloud to grow bit too large before dispersing the ions; with the ions moving only every 100 fs, they could possibly remain in place for up to 100 fs longer, causing unphysical field emission and ionization during that time. Running the same simulation with a base timestep of 1 fs, instead of 5 fs, results in front expansion around 200 μm/ns—slightly lower (admittedly not much).

However, the simulation's possible exaggeration notwithstanding, experiments do measure surprisingly high ion velocities from vacuum arcs [59, 29, 103, 111, 18]—velocities in the range of 10–20 km/s, or 10–20 μm/ns, still an order of magnitude less than the ion front velocity in the simulation. Those ion velocity measurements are more-or-less steady-state, whereas the ion cloud explosion is almost surely a one-time event.

6.8 RF Simulation Results

Simulations with an oscillating applied field show that breakdown in an RF cavity caused by field emission combined with neutral vapor closely resembles DC

breakdown. Since we did not use the electrodynamic field solver, but merely the electrostatic field solver, the simulation is more accurately denoted AC, not RF; however, the intention is to simulate breakdown in an RF cavity.

The RF simulations were set up the same as the DC simulations, except that the field, instead being 30 MV/m constant in time, oscillated sinusoidally in time with frequency f and amplitude 30 MV/m.

The beginning of the RF simulations differs from the DC simulations in that electrons emerge from the field-emitter only in spurts when the field nears its maximum. When the neutrals "evaporating" from the cathode travel far enough away that emitted electrons can gain enough energy to ionize them, ions are created in spurts when electrons are emitted. During the rest of the RF period, no electrons are added, and no ions are created; during this time, the ions disperse—self-repulsion pushes the ion cloud apart, and some ions can be lost forever when they hit the cathode. Starting an RF discharge is therefore more difficult than starting a DC discharge—slightly higher fields and currents and neutral densities are needed.

If the electric field and electron current and neutral density are all high enough, then enough ions remain near the field emitter during an RF period to enhance the field at the next field crest. More electrons are emitted and more ions are created, and the ion cloud grows. As the ion cloud grows, it adds to the surface electric field, increasing not only the field emission current at the field crest, but also the time during which current is emitted. As the ion cloud grows, the simulation resembles DC simulations more and more. Eventually, the field due to the ion cloud overwhelms the applied RF field, and the problem becomes basically DC.

As in DC simulations the ion cloud continues to grow and at some point the cloud starts to expand rapidly—the ion cloud explosion.

It is reassuring to see that our RF simulations agree well with Knobloch's simulations, which clearly show field and field-emission enhancement by relatively heavy ions which remain near the cathode throughout an RF period; unfortunately, MASK could simulate only the first few periods of ion cloud growth before some sort of simulation failure (part of the reason why we used OOPIC rather than continue with MASK). OOPIC can continue the simulation presumably forever, but the simulation becomes unphysical as the arc reaches high currents, where processes (such as cathode heating and thermionic emission) not included in the simulation become important.

The simulations show how, once current growth has started, RF and DC breakdown events behave almost identically in the region near the field emitter. In this light the similarity between RF and DC starbursts (described in section 5.4) is not at all surprising.

Although triggering RF breakdown, as shown by simulations, is just a matter of having enough stuff—enough field, enough field emission, enough neutral gas, enough ions—it is interesting how the frequency and ion mass affect how much is enough. Lighter ions travel farther during an RF period; therefore, all other things (like ionization cross-section) being the same, more current and neutral gas will be required to trigger a breakdown if the neutral gas were hydrogen than if the gas were niobium, because hydrogen ions would dissipate much faster than niobium atoms. With a simulation, we can actually alter the ion mass, keeping all other things the same. In the above simulation, where argon atoms and ions led to a catastrophic current rise, atoms of mass 1 led to a steady state: each RF

period, electrons were emitted at the crest, ions were formed, and then the ions dissipated over the rest of the period, never significantly enhancing the surface field. Light atoms have another disadvantage: the rate of ionization depends on the number density, and given the same flux of atoms from the surface (at the same temperature), lighter atoms will travel faster away from the cathode, giving a lower density. However, with the above conditions, even correcting the flux so that the neutral density in front of the cathode would be the same, mass 1 atoms still led to a steady state, whereas mass 40 atoms led to breakdown.

Increasing the frequency, however, decreases the amount time for ions to disperse before the next spurt of field emission. Therefore, upon increasing frequency, the light ions remain closer to the cathode and enhance the field more during the next spurt of field emission than they would at lower frequencies, so higher frequencies increase the likelihood of triggering breakdown when light ions are involved. On the other hand, decreasing the frequency can also increase the likelihood for breakdown; although ions will disperse more during a period, the field stays near its crest for a longer time, so the ions cloud can grow and start to enhance the field during a period (in the extreme case, the entire breakdown could happen while the field is at its crest—this is basically the limit as RF approaches DC; since DC breakdown currents can rise from microamps to amps within a nanosecond or so, RF fields with frequencies twenty or a hundred times less than 1 GHz are practically equivalent to a DC field equal to the amplitude of the RF field).

6.9 Simulation Defects

Although I find the simulations essentially convincing, some doubt shadows the results. It this section I'll outline a number of defects that prevent simulations from accurately mimicking reality.

6.9.1 Finite Grid and Timestep

Finite cell-size and timestep will always be a source of error until they can be reduced well below all important length and time scales. If the cell-size and timestep are small enough, then halving them shouldn't change the simulation.

We investigated the effect of mesh and timestep with an 11 GHz RF simulation. Since the initial activity happens in a region smaller than half the simulation space, we kept the number of mesh cells the same (128 by 64), but reduced the size of the simulation by half (in both z and r directions), keeping everything else the same (like electric field)—effectively reducing cell-size; the timestep remained 5 fs. Those changes hardly affect the simulation at all—watching both simultaneously, one appears to be a scale model of the other.

Later, however, we doubled the number of mesh cells in both z and r directions, keeping the simulation space the same size, 32 by 8 microns. In addition, the base timestep was reduced from 5 fs to 1 fs, and ten times as many particles were used (meaning that the number of electrons per macroelectron initially began as just 1, rather than 10). Since the field solver takes such much more computation time with a 256 by 128 mesh, compared to a 128 by 64 mesh, increasing the number of particles by a factor of ten had little effect on computation time (whereas, with a 128 by 64 mesh, it would have increased computation time by about a factor of

5). The simulation took a month of computing time instead of a day, and very closely resembles its quicker cousins in the early stages of current growth—for instance, they both take the same amount of time to build up an ion cloud to the point where electrons are emitted throughout an entire half-period. However, when the ion cloud in the quicker simulations abruptly explodes, the ion cloud in the better simulation does not—it does keep increasing current (having to increase the macroparticle size to limit the total number of particles to 0.4 million), and the ion cloud appears to expand at a faster and faster rate, but it seems a much more continuous transition, rather than the abrupt explosion.

It appears that the increased timestep and decreased macroparticle size, perhaps in combination with the increased mesh, do make a difference, appearing more realistic, insofar as smoother continuity seems more natural. However, up to the point of ion-cloud explosion there was not much difference. The maturity of the next generation of computers should make this problem much more tractable (including the ability to simulate individual electrons and ions, rather than use macroparticles).

6.9.2 Coulomb Scattering of Electrons and Ions

As discussed in section 6.7, the finite mesh smoothes out the electric field so that individual electrons and ions within the same cell do not see the $\sim 1/r^2$ Coulomb force that in reality describes their interaction (they actually see a force $\sim r$ that goes to zero as the interparticle separation goes to zero [46]).

To estimate the effect of Coulomb scattering between the electron beam and the ions in the ion cloud, consider an electron with incoming velocity v_i and impact parameter b (and mass m and charge $-e$), scattering off a massive ion. It scatters

at an angle:

$$\theta = 2\tan^{-1}\left(\frac{1}{2}\frac{e^2/(4\pi\epsilon_0 b)}{mv_i^2/2}\right) \tag{6.22}$$

(where $\theta = 0$ is no scattering at all).[11] For small scattering angles θ,

$$\theta \approx \frac{e^2/(4\pi\epsilon_0 b)}{mv_i^2/2}. \tag{6.23}$$

To estimate how much an electron would be deflected traveling through a cloud of ions, we'll arbitrarily pick some angle θ_m that represents a small but significant departure from "straight" (maybe 10 or 20 degrees), and find the impact parameter that would result in scattering at that angle, given electron initial kinetic energy $E = mv_i^2/2$: (again, for small θ),

$$b(\theta) \approx \frac{1}{2}\frac{e^2/(4\pi\epsilon_0)}{mv_i^2/2}\frac{2}{\theta}. \tag{6.24}$$

Electrons approaching with impact parameter less than $b(\theta)$ come closer to the ion and will scatter at angles greater than θ. The cross section for an electron scattering at greater than angle θ_m, given initial kinetic energy $E = mv_i^2/2$ is

$$\sigma(\geq \theta_m, E) = \pi b(\theta_m)^2 \approx \pi\left(\frac{e^2/(4\pi\epsilon_0)}{E}\right)^2 \frac{1}{\theta_m^2}. \tag{6.25}$$

Taking $\theta_m = \pi/20$ (9 degrees): for an electron with $E = 1$ eV, $\sigma(\geq \theta_m, 1\text{ eV}) = 2.6 \times 10^{-16}$ m^2. σ is proportional to $1/E^2$, so the cross section for a 10 eV electron scattering greater than 9 degrees would be 2.6×10^{-18} m^2, and a 100 eV electron

[11] The special case of Coulomb scattering demonstrates the previous argument about a particle entering a relatively deep potential energy well (compared to its initial energy) being unlikely to continue undeflected. Coulomb scattering shows that if the initial kinetic energy is less than or equal to the depth of the well it will encounter (or $e^2/(4\pi\epsilon_0 b)$, where b is the impact parameter, roughly measuring how close to the well's center the particle gets), then the particle will scatter at an angle greater than $2\sin^{-1}(1/\sqrt{5})$, or about 50 degrees.

would scatter only off an atom-sized area, 2.6×10^{-20} m^2. Knowing the cross-section for scattering off a single ion, we then have to estimate the density of ions to calculate how often an electron will scatter within the ion cloud.

When an electron is first emitted, it has little energy, and any reasonably nearby ion will change its direction. Because the ion is thousands of times heavier, the kinetic energy of the electron will remain almost unchanged.[12] Under the influence of the ion cloud, the electron moves away from the cathode and gains energy (falling into the potential energy well created by the ion cloud); even though ions may scatter the electron, the ion cloud as a whole still directs the electrons motion. The field must be quite high, probably in excess of 10^8 V/m (100 V/μm), and before long, the electron does not scatter nearly so much off individual ions. By the time the electron has 100 eV, the cross-section for scattering off an ion is about the same as for ionizing a neutral atom, and the density of ions is less than the (maximum) density of neutrals. Since electrons in this simulation have a good chance of not ionizing neutrals, they also have a good chance of making it through the ion cloud without significant deflect from collisions with individual ions. The simulation therefore does not err greatly in neglecting individual electron-ion collisions.

Electron-electron collisions should be less important than electron-ion collisions; the electron density is less than the ion density (otherwise the ion cloud wouldn't enhance the electric field), and electrons are likely to be moving at comparable velocities, reducing the chance of collision.

[12] If m is the electron mass, and M the ion mass, then the final velocity v_f of the electron after scattering into an angle θ is (approximated for $M \gg m$)

$$v_f^2 \approx \frac{v_i^2}{1 + \frac{M}{m}\sin^2\theta}. \tag{6.26}$$

(This assumes the ion was initially at rest.)

At least in the initial stages of breakdown simulation, Coulomb Scattering appears to be only a small, possibly negligible problem affecting electron trajectories. However, it would still be better, as well as feasible (with the modifications made to OOPIC for this project), to include some Coulomb scattering to make the simulation more realistic.

6.9.3 Cathode Shape and Field Enhancement

To simulate the field-emitter, we do not simulate the actual shape of the field emitter and let the field-solver calculate the field-enhancement due to the shape of the field-emitter. Instead, we chose empirically reasonable values for field enhancement and emitter area and injected the current that would come from that sort of emitter in the presence of the macroscopic field calculated by the field solver.

Part of the field-emission enhancement probably does not come from the geometrical effect of field enhancement, and in that case this approach is the best we can do, because we don't understand field emitters any better. However, field-enhancement occurs at protrusions on the cathode, and it is known to increase field emission. In this case, simply setting the field enhancement β from experimental data does not reflect reality very well. First, the field enhancement depends not only on the local geometry of the emitter, but also on how the field is applied. For instance, a cathode protrusion one micron high with a diameter of 0.1 microns would lead to a much larger field enhancement with an anode 1 mm away than with an anode 2 microns away. Thus, when the ion cloud begins to enhance the field, our simulations simply multiply the surface field found by the field-solver by β; since the ion cloud is actually quite close to the field emitter, β overestimates the field enhancement due to the ion cloud.

Furthermore, the simulated field emitter does not change with time; the simulation happily progresses up to currents of many amps. A real field emitter would melt under such current, and that would change, and probably reduce, the geometric field enhancement, which could rather abruptly reduce field emission current.

6.9.4 Neglecting Electromagnetism

We used electrostatics in all simulations, neglecting any magnetic field created by the current, as well as any electrodynamic effects. Although OOPIC has the capability to simulate the full electrodynamic field, we judged it wiser to restrict the problem to electrostatics, for several reasons. Most important, electrostatics is much simpler; if breakdown can be explained with just electrostatics, so much the better—it would waste time to start with the full electrodynamic simulation and later learn that electrostatics would have been sufficient. Also, electrodynamic field solvers are more finicky; they require smaller timesteps (related to the mesh size and the speed of light) for stability, and the finite mesh and timestep can lead to unphysical energy sources [46]. Where we are looking for some sort of current explosion, the added complexity of the electrodynamic solver creating energy out of nothing adds too much confusion; it's better to understand the electrostatic problem first, and later (if necessary) move on to electrodynamics.

Neglecting magnetostatics

As a rule of thumb, charged particles have to be moving near the speed of light before forces due to the magnetic field become significant compared to electric field forces; note that 1 keV electrons still have less than 10% the speed of light). However, a large exception to this rule occurs with charge neutralization by positive

and negative species together. With positive and negative species, there's plenty of charge to carry current, creating a magnetic field, but there's relatively little *net* charge to create an electric field. Once the ion cloud forms in the simulation, this is a concern.

Unfortunately, OOPIC does not support a "magnetostatic" field-solver that would include magnetism without the full electrodynamic solver. The effect of the magnetic field can be judged by estimating the magnetic field (see section D) created by experimentally observed arcs, around 10-100 A,, and considering the trajectories of electrons and ions in that field (see tables 6.1 and 6.2).

Table 6.1: Magnetic field 1 and 10 microns from 10 and 100 A currents.

I (A)	r (μm)	$B = \mu_0 I/(2\pi r)$ (T)
10	10	0.2
10	1	2
100	10	2
100	1	20

If a particle's cyclotron radius is large, then the magnetic field will have only a small effect—in this case, we're mostly concerned about the effect of the magnetic field on an electron's trajectory within the first several microns of the field emitter. Only the combination of a high magnetic field and low kinetic energy can give an electron a cyclotron radius of less than several microns. For ions the cyclotron radii will be even larger by a factor of the square root of the mass ratios—even a 10 T magnetic field on a 0.01 eV argon ion produces a 300 micron cyclotron radius, so ion trajectories are not going to be affected by the magnetic field at all.

Table 6.2: Cyclotron frequency ω_c and cyclotron radius r (and velocity v) for electrons with various kinetic energies E in 0.1, 1, and 10 T magnetic fields. This table neglects relativistic effects (which are just noticeable on the percent-level for 1 keV electrons).

B (T)	E (eV)	$v = \sqrt{2E/m}$ (m/s)	$\omega_c = eB/m$ (10^9 s^{-1})	$r = v/\omega_c$ (μm)
0.1	0.1	5.9×10^5	18	34
0.1	1	1.9×10^6	18	110
1	0.1	5.9×10^5	180	3.4
1	1	1.9×10^6	180	11
1	10	5.9×10^6	180	34
1	100	1.9×10^7	180	110
10	0.1	5.9×10^5	1800	0.34
10	1	1.9×10^6	1800	1.1
10	10	5.9×10^6	1800	3.4
10	100	1.9×10^7	1800	11
10	1000	5.9×10^7	1800	340

Of course, in the area near the field emitter, the electric fields are typically quite large, because of both the applied voltage and the ion cloud. With fields around 100 V/μm and greater, an electron in a 10 T field is not going to remain at 10 eV for much of its 3.4 micron radius orbit.[13] Moreover, the magnetic field will only be as high as 10 T within about a micron of the cylinder axis, even for currents as high as 100 A.

The possible effect of the magnetic field is borderline: with the highest measured currents and a high current density (packing all that current into a one micron radius discharge), the magnetic field could just affect the trajectories of the slowest electrons in areas of not extremely high electric field. Given all these ultimatums, it's more than likely that neglecting the magnetic field is not a serious defect in the simulations.

Neglecting electrodynamics

With particles mostly moving at speeds much less than the speed of light, the biggest electrodynamic effects will come from the RF fields in the simulation, which in our electrostatic treatment really become AC fields. However, the simulation volume (at most 32 microns in length) is much smaller than the wavelength of RF driving frequencies considered (up to 11 GHz)—the simulation volume is only a small part of the cavity near the field emitter. Also, the RF magnetic field corresponding to electric fields around $E \sim 10^8$ V/m will be around $B \sim E/c \sim 0.3$

[13] Waving hands gives the right idea, but the proper way to do this problem is to consider a particle trajectory in electric and magnetic fields. In perpendicular E and B fields, a charged particle circles as it drifts (on average) in a direction perpendicular to both fields at speed $v_D = E/B$. The radius of its orbit, if the electron starts at rest, is $r = v_D/\omega_c$. For $E = 100$ MV/m and $B = 10$ T, $r = 5$ microns. If the electron does not start at rest, but has some initial velocity, the orbit of radius usually becomes larger.

T, not enough to significantly alter particle trajectories (see previous section). Therefore, neglecting electrodynamics is quite reasonable (but see section 6.9.5).

6.9.5 Neglecting the Power Source

In the simulations of breakdown, an electric field, either DC or RF, is applied as a boundary condition, regardless of how much power it takes, ignoring whatever external circuit would be required. In reality, the voltage drops when the discharge current grows. Currently, the simulations only cover the initial triggering and current rise, the first nanosecond or two of breakdown, and even a small capacitance can store enough charge to maintain a 10 A current for a nanosecond; however, for future simulations, hopefully running for longer times, we would need to include the external circuit with a realistic power source.

The same argument applies for an RF problem: the RF cavity provides some limited stored energy, and depending on the nature of the RF input, the external RF circuit may be important too (for superconducting cavities, the RF input power is fairly low, but for normal conducting cavities it could play an important role). Ideally, the entire cavity should be simulated with the full electrodynamic field solver, to determine how energy is transferred from the cavity field to the arc.

Chapter 7

Modeling Breakdown

The process of voltage breakdown is a sudden transition of insulating vacuum to conducting plasma, comprising

1. pre-breakdown (field emission and electron bombardment),
2. the trigger (instability leading to current rise),
3. plasma expansion (the arc), and
4. the extinguishing of current (the end).

7.1 Pre-Breakdown

The vacuum will remain a perfect insulator as long as there are no charged particles to carry electrical current. There are some processes, like field emission, that introduce charged particles into the vacuum in a small, relatively stable manner, without completely destroying its insulating properties. Such processes are likely candidates for triggering breakdown; we consider these possible precursors in the pre-breakdown category because they can occur without triggering breakdown, and most of our knowledge about them comes not from studying breakdown, but from studying them in their stable modes.

7.1.1 Field Emission

Everything seems to come back to field emission; field emission is the most frequently observed source of unwanted charged particles in vacuum when high electric fields are present.

"Field emission" describes the quantum mechanical tunneling of electrons out of a metal, through the work function barrier made thinner by a strong electric field at the metal surface. In 1928 Fowler and Nordheim published the solution to this tunneling problem in its simplest form: a triangular barrier in one dimension [37]. Basically, the current density J depends thus on the surface electric field E:

$$J = J_0 \frac{E^2}{E_0^2} e^{-E_0/E} \tag{7.1}$$

for material-dependent constants J_0 and E_0; $E_0 \propto \phi^{3/2}$ where ϕ is the work function. Efforts to improve the theory treated more realistic barrier shapes (including the image charge force, for example) and added the effects of non-zero temperature [78]. Theories of more complicated barrier shapes should be taken with a grain of salt; in confirming the Fowler-Nordheim theory for a (110) facet of a sharp tungsten point, Ehrlich and Plummer [36] conclude:

> Our general finding is that the measured current density is within a factor of 5 (30) of the current density calculated assuming a free-electron metal with (without) the image potential included in the tunneling barrier; in our view, this is a close agreement.

The measured current density was 5 times lower than calculated with image potential, and 30 times higher than calculated with the purely triangular barrier, sandwiched between two barrier shapes. The quotation demonstrates the great sensitivity of field emission on the electric field—agreement within a factor of 5 or

30 is considered close; small details in the barrier can make a large difference in current (for instance, an adsorbed carbon atom could slightly alter the barrier in a way that would significantly affect the current but could be hard to calculate).

Before going further, I want to point out that simply applying an electric field to a cathode, even in ultra-high vacuum, does not yield a nice Fowler-Nordheim current vs. voltage curve; rather, it yields a noisy curve that shows emitters changing with time, spiking and jumping to different levels, *etc.* To get a good Fowler-Nordheim curve, the electrodes must be processed—usually baked, exposed to high electric fields and allowed to break down—and maintained in UHV; only then does the field emission appear to follow the Fowler-Nordheim equation.

The Fowler-Nordheim theory of field emission has been generally confirmed [33, 36] on very clean, very sharp points, but field emission from large-area electrodes still remains somewhat mysterious ([83] is a comprehensive review). Field emission on large electrodes starts at fields as low as 10–100 MV/m, rather than the expected threshold around $\sim$1 GV/m. Researches have used several techniques to see how field emission varies across large-area cathodes, including scanning with point anodes to expose only a small part of the cathode to a high field at one time, using phosphorescent anodes to detect where electrons are emitted, and using anodes with a small hole that transmits only current from a small area. All these techniques find that field emission comes from point sources on the cathode.

Although identifying the point sources has been possible with scanning-anode experiments built in electron microscope chambers [41, 82], characterizing them has been difficult. Many emitters are particles, but not all, and not all particles are emitters (and not all emitting particles are conductors); sometimes other surface defects seem to be the source of field emission. Apparently contaminants, in the

form of particles or surface adsorbates, can greatly affect field emission. Some work has been done to connect surface work functions, as measured by the photoelectric effect, and field emission [30].

The cause of this enhanced field emission, appearing at much higher levels than expected (or at lower fields than expected), is still unknown, although many models have been suggested. The enhancement of electric field due to emitter geometry is certainly part of the explanation—pointy objects enhance the electric field and the field emission at their tips. However, there is ample evidence of enhanced field emission from insulating particles and regions that do not have the sort of geometry that could enhance the field enough to explain observed field emission [83]. Other models for field emission generally rely on more complicated descriptions of surface states. For example, surface adsorbates create local wells in the work function barrier, allowing increased current due to resonant tunneling [31]. Another model explores the interaction between the metal surface and an insulating oxide, where charging creates high fields and current heats the insulator and its electrons, which are then emitted more easily [107]. No theory other than that of Fowler and Nordheim has yet proven to be very useful, nor received broad experimental confirmation. The cause of field emission could be important in considering models of breakdown; for instance, we would like to know the mechanism for field emission so we could calculate heating of the emitter due to field emission. However, we chose to try to move beyond field emission, taking observed field emission for granted, using the empirical parameters of field enhancement β and emitter area S with the Fowler-Nordheim model.

The field emission current from a single emitter, described by β and S is

$$I = J_{\mathrm{FN}}(E)S = J_0 S \frac{(\beta E)^2}{E_0^2} \exp\left(-\frac{E_0}{\beta E}\right) \tag{7.2}$$

where E is the macroscopic (unenhanced) surface electric field at the emitter, and J_0 and E_0 are constants that depend on the details of the barrier (like the barrier height, or work function), which can be calculated from characteristics of the material measured by means unrelated to field emission. For niobium, the work function is about 4 eV, and $J_0 = 1.15 \times 10^{15}$ A/m^2 $= 1.15$ kA/μm^2, and $E_0 = 5.46 \times 10^{10}$ V/m $= 54.6$ GV/m [87].

Field emission as described above is cold field emission. High temperatures well above room temperature greatly increase electron emission; this regime is sometimes called thermally assisted field emission or TF emission [42]. The combined effects of field and temperature greatly increase electron emission relative to either one alone, and may contribute to the runaway nature of breakdown. However, the emitter must already be quite hot before TF emission becomes important.

7.1.2 Electron Bombardment

Whereas field emission is the most important activity at the cathode before breakdown, electron bombardment is the most important activity at the anode (for breakdown in an RF cavity, I consider the field emitter to be the "cathode" and the site where emitted electrons bombard the cavity wall to be the "anode," relating the cathode and anode to emission and bombardment of electrons rather than voltage polarities). More energy is deposited into the anode (the current times the voltage), but because the field emitter may be much smaller than the electron beam at the anode, the power per area at the anode is not necessarily as large as at the cathode. Moreover, 10 keV electrons penetrate several microns deep into the anode, dispersing energy over a significant depth as well.

To calculate the power per area deposited at the anode, we need to know the radius of the electron beam at the anode. The factors determining the beam radius are: initial emission radius r_e, initial transverse velocity $v_{\perp,i}$ at emission, the transverse velocity gained from transverse field components due to geometry (presumably only near the emitter), and the transverse velocity gained from the electric field of the beam itself (space-charge). The beam radius is calculated in appendix B as a function of distance from the emitter.

7.2 The Trigger

Electron current alone does not constitute breakdown. In some cases field emission (often called dark current) presents a serious problem; however, field emission currents are generally microamps or less, and while providing the vacuum with a finite conductance, the conductance is still very small compared to the arc that causes voltage breakdown. Nor is field emission inherently unstable; field emission electron guns can be operated usefully, as in some electron microscopes. While it is not evident that field emission is a necessary precursor of all vacuum breakdown, field emission appears so frequently related to breakdown, that it must be a focus of attention. Besides triggers related to field emission, we should also consider the possibility that microparticles (either foreign particles, or bits of the electrode) are launched into the vacuum by the electric field, somehow igniting a discharge, perhaps with the help of friendly field emission [27].

7.2.1 Cathode vs. Anode

Possible triggers of voltage breakdown can be divided into two categories, anode-initiated and cathode-initiated.[1] We should bear in mind that both cathode and anode mechanisms are valid, under different circumstances; high voltages (hence large gaps) should favor anode-related mechanisms, while high electric fields (hence small gaps) should favor cathode-related mechanisms.[2]

If the trigger of breakdown is related to heating, then another difference between cathode and anode initiation would be that cathode heating would be expected to depend on I^2, whereas anode heating would be proportional to IV (where I is the current and V the voltage).

Charbonnier *et. al.* describe a model [20] in which they compare the heating at the emitter and at the anode, and conclude that the power density at the cathode increases with field enhancement β (which increases with taller, thinner emitters), because higher β creates a higher current density at the emitter and a taller, thinner emitter can't dissipate heat as well to the substrate; at the same time, the pointier protrusion disperses current over a wider region, tending to decrease the power density at the anode. Therefore, they conclude that there is a "critical" enhancement factor β_c; for $\beta > \beta_c$, emitter heating dominates, and for

[1] Again, for considering breakdown in RF cavities as well as DC vacuum diodes, the cathode is considered the site of the field emitter and the anode a site bombarded by electrons.

[2] This applies to DC experiments. Most anode-related triggers of breakdown involve charged particles ion-sized or larger leaving the surface and gaining energy due to the electric field. In an RF cavity, one can still apply the same argument with modification: in a high-frequency field, such massive particles (compared to the electron) do not gain more energy as they traverse larger distances the way they do in DC gaps. Furthermore, in RF cavities, unlike parallel plate DC gaps, ions (or larger charged clumps of material) created at the electron bombardment site would not follow the same paths back to the electron emission site. RF fields, therefore, tip the balance more in favor of cathode mechanisms.

$\beta < \beta_c$, anode heading is more important. Assuming the triggering of breakdown to be related somehow to heating, they conclude that β_c is the dividing line between cathode- and anode-initiated breakdown. However, they do not explain how heating at either electrode causes breakdown.

Many experiments with pulsed DC voltages have explored cathode versus anode triggering (a good reference is [75]). Pulsed voltage experiments are a little different from the CW experiments discussed in this work;[3] a short pulse (say, 100 ns long) can put an experiment in a situation that would be unstable in CW, so pulsed experiments sometimes force breakdown to occur in a possibly different manner. Generally higher fields can be reached with shorter pulses, and very short pulses favor cathode-initiated breakdown because there's not enough time for the anode to play a role. Some experiments with short pulses (from nanoseconds to microseconds) show that breakdown occurs at constant J^2t, where J is the field emission current density and t the time from the application of the pulse to voltage breakdown [24]; presumably, J^2t corresponds to the energy input per area at the field emitter due to Joule heating, providing evidence that those breakdown events were triggered by the cathode.

In the experiments described in this work, there are several reasons to believe that the cathode triggers the breakdown events; at the very least, the cathode certainly localizes breakdown events (*i.e.*, determines where the arc contacts the cathode). Damage to the cathode after breakdown (in our DC experiments) usually affects an area less than 100 microns in size; often the damaged area, or starburst,

[3] CW stands for continuous wave, usually used in RF contexts to distinguish between pulses of an oscillating RF field and continuously applied RF fields. Here we extrapolate its meaning to zero frequency: continuous DC, as opposed to pulsed DC.

is centered at the former site of a contaminant particle; for example, the presence of a particle clearly selected the breakdown site on the cathode in figures 4.3 and 4.4 on pages 80 and 82.

It is within the realm of possibility that the anode is mainly responsible for triggering breakdown (though less likely for RF breakdown), but that it does so with necessary encouragement from field emission; that is, field emission selects the spot on the anode where breakdown could occur, the anode decides upon a critical value of field emission above which it will set off a catastrophic chain of events leading to enormous current rise, and the field emitter then acts as a lightning rod in attracting the resulting arc.

Cathode-initiated breakdown

The experiments that clearly demonstrate a cathode trigger are those on cathodes contaminated with spiny vanadium. The clear indicator of a cathode trigger is the dependence of breakdown voltage on a property of the cathode (in this case, its contamination). Cathodes with spiny vanadium consistently broke down at lower voltages (see figure 4.11 on page 91). In one test, for example, contamination alternated on adjacent cathode pedestals, so pedestals 1, 3, and 5 had vanadium particles, while pedestals 2 and 4 had palladium particles. The pedestals were tested in numerical order at a gap of 150 microns, with multiple tests on pedestals 3 and 5 (a breakdown occurred, then the voltage was raised again until another breakdown); the pedestals with vanadium broke down at much lower voltages than those with palladium (table 7.1), showing that the cathode contamination determined the breakdown voltage.[4]

[4]The test on pedestal 2, with palladium, was halted prematurely at 66 MV/m; no breakdown had occurred by that point.

Table 7.1: Breakdown voltages for different pedestals on the same cathode; pedestals 1, 3, and 5 were contaminated with spiny vanadium particles (called V1), and pedestals 2 and 4 had palladium particles.

(Cathode B5) Pedestal No.	1	2	3	4	5
Contaminant	V1	Pd	V1	Pd	V1
Breakdown field (MV/m)	39	>66	38, 41, 59	95	33, 39, 44

Knowing that cathodes contaminated with spiny vanadium almost certainly suffered cathode-initiated breakdown, we wondered whether there would be any correlation between field emission before breakdown and breakdown field (see figure 4.15 on page 100). One might expect that less current might be needed to cause breakdown with taller, thinner emitters; also, taller, thinner emitters should break down at lower fields (because of greater field enhancement). However, the actual data cover a wide range, with pre-breakdown field emission (almost) from 10 nA to 100 μA, over a range of fields. Within the scatter, there's not much to conclude, but it is worth noting that in almost all cases there was some measurable field emission before breakdown, and it was typically within two orders of magnitude of a microamp.

If we assume the field emission follows Fowler-Nordheim behavior, we can try to find the local field enhancement β and emitter area S from the emission current versus electric field data. From these we can find the current density and local electric field at breakdown. Other studies (on processed electrodes) have found these quantities to be constant, depending on the cathode material [58, 15]. Our data, however, does not lend itself to a critical field (or current density) explanation (figure 7.1). Part of the problem may lie in calculating the emitter area S: once

β is found from the slope of the Fowler-Nordheim graph, the emitter area can be found from $1/\beta^2$ times the exponential function of the intercept. The emitter area is extremely sensitive to the intercept, and calculated areas vary over unphysical ranges, so it's difficult to trust any calculation that depends on knowing the emitter area extracted from a Fowler-Nordheim fit.

7.2.2 A Trigger: Field Emission Plus Neutral Gas

Field emission in combination with a rather large density of neutral gas (for a vacuum) can trigger a catastrophic current growth that leads to breakdown; the density of neutrals need be large only in a small region around the cathode. That breakdown might be triggered by the release of neutral vapor in the presence of field emission has been conjectured for some time [95], and the simulations presented in this work demonstrate that high local gas densities near the field emitter can lead to run-away current and ion production.

Simulations show that neutral densities on the order of 10^{24} m^{-3} (about 30 torr for an ideal gas at room temperature) near the field emitter can initiate breakdown with a field emission current on the order of 100 μA. Such simulations can include more detailed physics than analytical calculations (*e.g.*, the dependence of ionization cross-section on electron energy), but simple analytic estimates of the conditions necessary for breakdown can also be helpful. We will find in this section, approximately the minimum electron emission current I and neutral density n (at the emitter) necessary to trigger a catastrophic increase in current that would lead to breakdown in a given applied electric field.

Ion production is the key to initiation of breakdown. The electric field produces electrons; electrons produce ions; and ions, in turn, enhance the electric field,

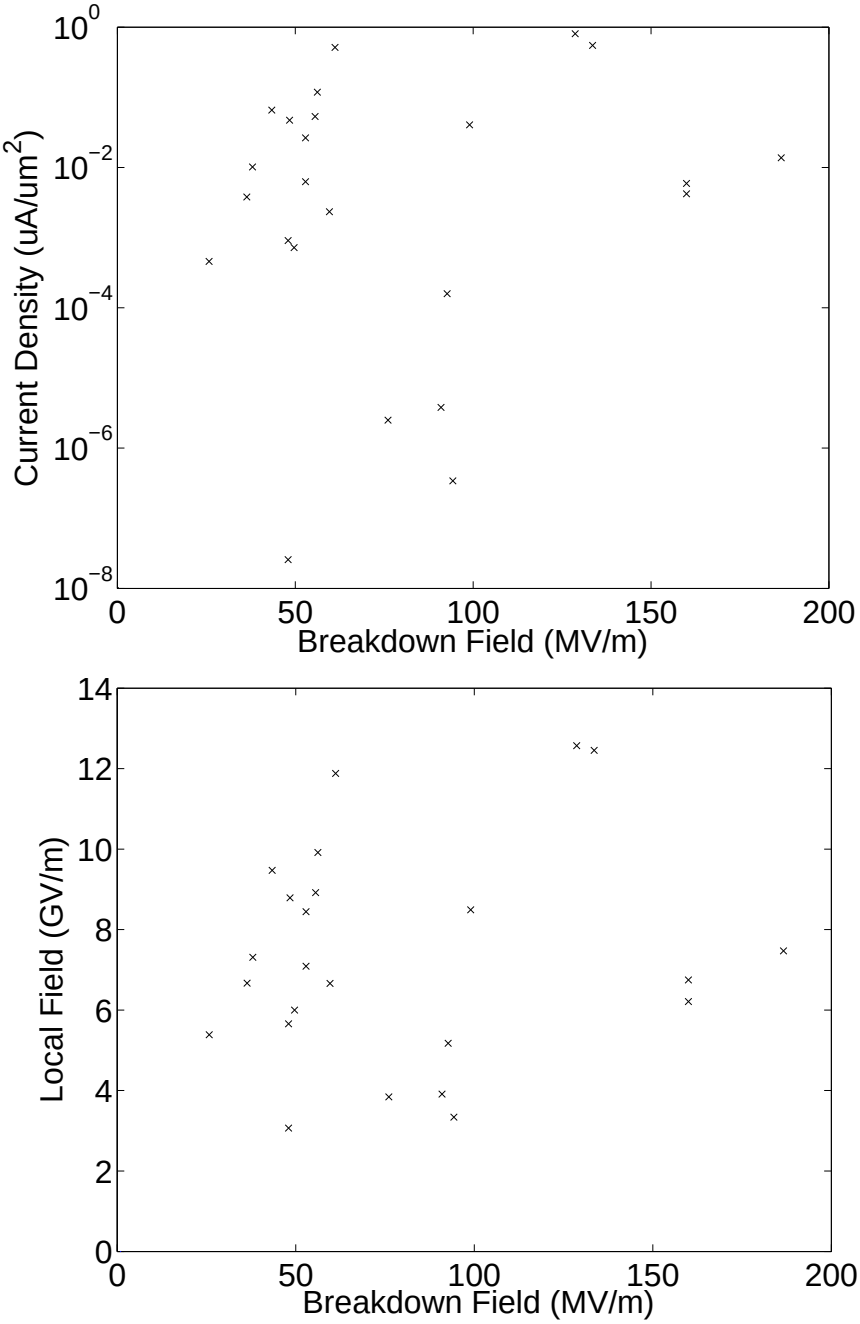


Figure 7.1: Current density and local field βE at breakdown versus macroscopic breakdown field, calculated from the Fowler-Nordheim model of field emission, for spiny vanadium (only tests with at least 5 current-field data points with Fowler-Nordheim plots that had a linear correlation coefficient greater than 0.92 were used).

yielding more electrons, more ions, and yet more electric field. When perfectly balanced in equilibrium, the cathode electric field E creates an electron current I, which creates an ion cloud with N_i ions, which enhances the applied field to produce exactly the field E at the cathode. Finding the point of equilibrium requires solving three equations for the three unknowns:

$$\begin{aligned} E &= E(N_i) && \text{Coulomb's Law,} \\ I &= I(E) && \text{field emission, } \textit{e.g.}\text{, Fowler-Nordheim,} \\ N_i &= N_i(I) && \text{ionization and ion-escape.} \end{aligned} \tag{7.3}$$

Ion production

We'll start by assuming that significant ionization occurs only within a distance d of the cathode;[5] within this distance we treat the cross-section σ as a constant, but beyond d we consider the cross-section to be zero.

The number of ions increases when electrons ionize atoms (this growth rate will be proportional to the electron current); and the number of ions decreases when ions escape to the cathode, where they are absorbed (this "shrink" rate will be proportional to the number of ions).

In any time Δt, the number of electrons emitted is $(I/e)\Delta t$; on average each electron ionizes $d/\lambda_{\text{mfp}} = n\sigma d$ atoms within distance d, so the ion growth rate due

[5] To be physically plausible, this artificially designated ionization region that extends a distance d from the cathode should include the area in which electrons are most likely to ionize. Two mechanisms limit the "ionization range": (1) electrons with high energy have a low probability of ionizing, and (2) if the neutral gas emerges from a very small area on the cathode (perhaps from the field emitter), the gas density will fall off as the inverse square of the distance from the cathode, diminishing the chance of ionization far from the cathode.

To estimate d, we consider two simple possibilities: (1) d should be around the $V_2 \approx 500$ V equipotential, where the real ionization cross-section starts to get small—$d \sim V_2/E$; or (2) if $d \sim V_2/E$ is much larger than the radius r_n of the area from which neutral emerge, then $d \sim r_n$ is a better choice.

The cross-section for ionization decreases as an electron gains energy above roughly a hundred eV (but see figure 6.6 on page 153); at much higher energies, the probability of an electron ionizing an atom becomes negligible. Setting d to be the distance of the equipotential several hundred volts from the cathode ($d \sim V_2/E$) would be a reasonable choice; beyond $d \sim V_2/E$, electrons will have more than several hundred eV, and a negligible chance of ionizing.

However, if $d \sim V_2/E$ is much larger than the area from which neutral atoms are emitted, the neutral density becomes small, tapering off as $1/x^2$ for x farther from the cathode than the radius of the neutral gas emitter, and we might prefer $d \sim r_n$, the radius of the neutral gas emitter.

Very close to the cathode, an electron lacks the requisite ionization energy E_{ionize} (for the neutral atoms in question); therefore, no ionization occurs within a distance $E_{\text{ionize}}/eE \sim 15$ eV$/eE$ of the cathode. This distance (to the 15 V equipotential) is small compared to the entire ionization region, and can be neglected.

to ionization is:

$$\dot{N}_{i,\text{ionization}} \sim \frac{I}{e} n\sigma d. \tag{7.4}$$

Ions will disappear from the system at the rate at which they hit the cathode:

$$\dot{N}_{i,\text{cathode absorption}} \sim -\frac{N_i}{\bar{t}} \tag{7.5}$$

where $\bar{t}$ is an average[6] ion lifetime (the time between creation and absorption at the cathode). An ion of mass m_i and charge e created a distance x from the cathode reaches the cathode (under the influence of the constant electric field E) in time:

$$t(x) = \sqrt{\frac{2m_i x}{eE}}. \tag{7.6}$$

We take an average lifetime $\bar{t}$ to be two-thirds of the maximum escape time:[7]

$$\bar{t} \sim \frac{1}{d}\int_0^d \sqrt{\frac{2m_i x}{eE}}\, dx = \frac{2}{3}\sqrt{\frac{2m_i d}{eE}}. \tag{7.8}$$

The total growth rate of the ion cloud is

$$\dot{N}_i = \dot{N}_{i,\text{ionization}} - \dot{N}_{i,\text{cathode absorption}} \sim \frac{I}{e} n\sigma d - \frac{N_i}{\bar{t}}. \tag{7.9}$$

[6]In this section I try to be careful to say "an average" rather than "the average" for estimated quantities like $\bar{t}$. "An average" could be the standard average of lifetimes, but could also be, for instance, the inverse of the average of $1/t$ or the root mean square of t. The right kind of average should be chosen to make the equation correct, or as accurate as possible; if all ions had the same lifetime, all these methods of averaging would yield that lifetime, and it wouldn't matter. In practice, different averages should be fairly close to each other, and with an order-of-magnitude calculation, it's reasonable enough to take any typical value.

[7]In this case, we're choosing the average so that equation 7.10 will be accurate. If ions created at a distance x_0 from the cathode stay in the system for time $t(x_0)$, then if the ions are created at a rate $\dot{N}_i$ at x_0, there will be $\dot{N}_i t(x_0)$ ions in the system. If ions are created at total rate $\dot{N}_{i,\text{ionization}}$ distributed uniformly throughout distance d, then the number of ions in the system will be

$$N_i \sim \int_0^d \frac{\dot{N}_{i,\text{ionization}}}{d} t(x) dx. \tag{7.7}$$

Hence $\bar{t} = (1/d)\int_0^d t(x)dx$.

As long as the current I stays fixed, the number of ions N_i approaches its steady-state value (*i.e.*, $\dot{N}_i = 0$)

$$N_i \sim \frac{I}{e} n\sigma d\bar{t} \sim \frac{2}{3}\frac{I}{e} n\sigma \sqrt{\frac{2m_i}{eE}} d^{3/2} \tag{7.10}$$

with exponential decay of characteristic time $\bar{t}$. Of course the current may not remain fixed: as the number of ions changes, the cathode field changes, and the cathode field will determine the current.

The field produced by the ions

Having found the number of ions in terms of the electron current (equation 7.10), we now find the field in terms of the number of ions. The total field *at the cathode* is $E = E_{\text{ext}} + E_i$, the sum of the externally applied field and the ions' field.

In the (rather artificial) limit where the ions are far from the emitter, the electric field at the emitter created by the ions is[8]

$$E_i \sim 2\frac{eN_i}{4\pi\epsilon_0 \bar{x}^2} \qquad (\bar{x} = \text{avg. distance from ions to cathode}) \tag{7.11}$$

and in the opposite limit where the ion cloud is very close to the emitter, and spread out transversely over a radius r_i, the electric field is[9]

$$E_i \sim \frac{eN_i/(\pi r_i^2)}{\epsilon_0} \qquad (r_i = \text{radius of ion cloud}). \tag{7.12}$$

In order of magnitude, we expect that $r_i \sim \bar{x} \sim d$ (d is the ionization range). Without considering the structure of the ion cloud more carefully, we cannot compute the field more accurately; fortunately, the quasi-1D and -3D approximations are

[8] The prefactor of two is from the image charge across the cathode plane.

[9] The electric field due to a plane of charge density (per area) σ_c is $E = \sigma_c/(2\epsilon_0)$; adding the image across the cathode puts in a factor of 2. The area charge density is simply $eN_i/(\pi r_i^2)$.

very close to each other in order of magnitude, so we don't have to worry much about it, and we choose the compromise:

$$E_i \sim \frac{eN_i}{2\pi\epsilon_0 d^2}. \tag{7.13}$$

Using equation 7.10 to write N_i in terms of I,

$$E_i \sim \frac{I}{e} n\sigma d\bar{t} \frac{e}{2\pi\epsilon_0 d^2} \sim \frac{I}{e} n\sigma d \frac{e}{3\pi\epsilon_0 d^2} \sqrt{\frac{2m_i d}{eE_{\text{ext}}}}. \tag{7.14}$$

E_i is the field *at the cathode* produced by the ions; the field produced by the ions at points within the ion cloud should be (at macroscopic length scales) less than E_i, since the ion cloud extends practically to the cathode. The ion escape time $\bar{t}$ depends on the field in the ion cloud; we make the reasonable simplification that $\bar{t}$ changes very little with E_i, and evaluate it at $E = E_{\text{ext}}$. We make no such simplification for I, which depends sensitively on deviations of the electric field from E_{ext}.

The linear dependence of E_i on I will be important later; for convenience we write

$$E_i \sim \rho I, \tag{7.15}$$

where

$$\rho = \frac{n\sigma\bar{t}}{2\pi\epsilon_0 d}. \tag{7.16}$$

The current produced by the field

The field at the cathode created by the ions will change the emitted electron current from $I = I(E_{\text{ext}})$ to $I = I(E_{\text{ext}} + E_i)$; for now we just accept that there is a functional form $I(E)$ without worrying about what it is:

$$I = I(E). \tag{7.17}$$

(An obvious choice for $I(E)$ comes from Fowler and Nordheim.)

A little later we will prefer to manipulate a function of convenience, $K(E)$, which depends on $I(E)$ such that:

$$\frac{dI}{dE} = \frac{K(E)}{E} I. \tag{7.18}$$

$K(E)$ is merely notation, but it's important that for many functional forms $I(E)$, $K(E)$ depends weakly on E compared with $I(E)$ or $dI(E)/dE$. For instance, $K(E)$ has no E-dependence for $I(E) \propto E^n$, and for $I(E) \propto e^{E/E_0}$, $K(E)$ is only linear in E while I and dI/dE are exponential in E.

The steady-state solutions

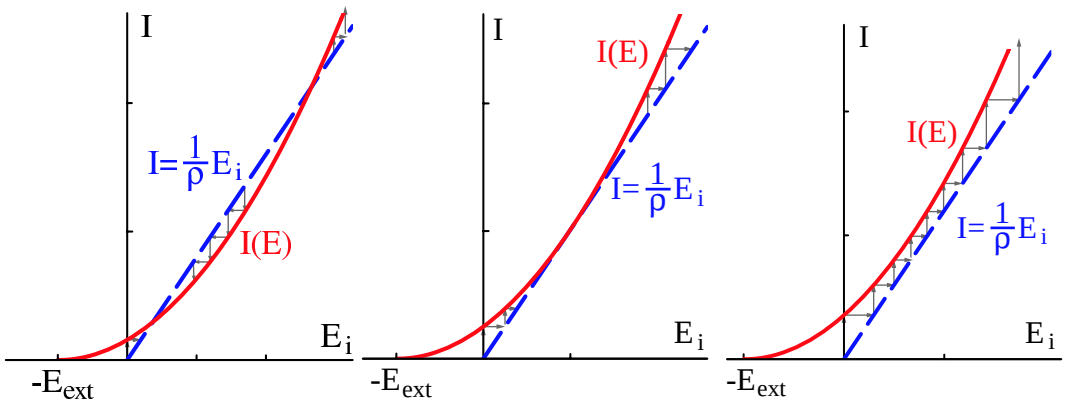
To find a steady state condition, we have to solve simultaneously the equations

$$N_i \sim \frac{I}{e} n\sigma d\bar{t}\,, \qquad E_i \sim \frac{eN_i}{2\pi\epsilon_0 d^2}\,, \qquad \text{and} \qquad I = I(E_{\text{ext}} + E_i) \tag{7.19}$$

for N_i, E_i, and I, given E_{ext} (upon which $\bar{t}$ depends), n, σ, and d. We easily eliminate N_i (refer to equations 7.14 and 7.16), yielding

$$I = I(E_{\text{ext}} + E_i) \qquad \text{and} \qquad E_i = \rho I. \tag{7.20}$$

The intersections of the curve $I = I(E_{\text{ext}} + E_i)$ and the line $I = (1/\rho)E_i$ represent steady-state solutions; because $I(E = 0) = 0$ and $I(E)$ probably increases rapidly with E (at the very least: $d^2I/dE^2 > 0$), the graphs will likely resemble figure 7.2. Very high field emission results in no solutions (figure 7.2c); application of E_{ext} (E_i starts at zero) creates a current, which creates an E_i, which creates a higher current, creating a higher E_i, *ad infinitum*. The lack of any steady-state solutions implies immediate breakdown. Field emission that is relatively low upon application of the external field results in two steady-state solutions (figure 7.2a); upon application



(a) No breakdown (b) Threshold (c) Breakdown

Figure 7.2: (a) Two steady-state solutions (the slope of $I(E)$ equals the slope of the line at a point below the line): E_{ext} is below the breakdown field. (b) One solution ($I(E)$ is parallel to the line at their intersection): E_{ext} is at the breakdown field. (c) Zero solutions (the slope of $I(E)$ matches that of the line at a point above the line): E_{ext} is above the breakdown field; the current and field will increase forever.

of the external field, the current rises, which increases E_i, which increases I by a smaller amount, which increases E_i by an even smaller amount, and the system converges to the lower steady-state solution. On the other hand, if an external jolt gave the system a higher current or field than the upper steady-state solution, the current and field would continue to increase (leading to breakdown). Starting between the lower and upper steady-states, the system relaxes to the lower steady-state. In figure 7.2b, the curves are just tangent, resulting in one metastable steady-state. As the applied field E_{ext} increases, the $I(E)$ curve shifts to the left, and the system moves from figure 7.2a, to b, to c, from non-breakdown to breakdown. The breakdown field is the E_{ext} at which there is one metastable steady-state (b).

The number of intersections depends on whether the point at which the field emission curve parallels the line is above or below the line (whether the point at

which the curve has slope $dI/dE = 1/\rho$ is above or below the line $I = (1/\rho)E_i$). The intermediary metastable state appears when the curve is tangent to the line.

When E_{ext} equals the breakdown field E_b, the current will be I_b at the steady state (which is metastable). Since the curve is tangent to the line with slope $1/\rho$ at that point,

$$\frac{dI}{dE}(E_b + E_{ib}) = I_b \frac{K(E_b + E_{ib})}{E_b + E_{ib}} = \frac{1}{\rho}. \tag{7.21}$$

Since the curve and the line intersect, $E_{ib} = \rho I_b$, and

$$\rho(E_b) I_b \frac{K(E_b + \rho(E_b) I_b)}{E_b + \rho(E_b) I_b} = 1 = E_{ib} \frac{K(E_b + E_{ib})}{E_b + E_{ib}}. \tag{7.22}$$

Remember that ρ depends on the field (through the ion lifetime), though only on its square root. This equation and the field emission equation, $I_b = I(E_b + E_{ib})$, would determine the breakdown current I_b, which is the maximum current for which a stable steady state exists. Equivalently (as shown in the above equation) we could solve for E_{ib}, the field enhancement of the ions at the breakdown threshold.

Perhaps, however, we don't have to solve those two equations, one of which, $I = I(E)$, is fairly uncertain. Since the dependence of $K(E)$ on E is relatively weak, we can write

$$\rho I_b \frac{K(E_b)}{E_b} = 1 \tag{7.23}$$

if $E_{ib} \ll E_b$, which is certainly the case (that before breakdown the field due to the ions is much less than the applied field). Solving for I_b,

$$I_b = \frac{E_b}{\rho K(E_b)}. \tag{7.24}$$

Although we still can't precisely solve this equation without worrying about the details of field emission, this equation is nonetheless quite helpful for estimating the breakdown current, because the right hand side is not very sensitive to E_b.

If we can guess the breakdown field within a factor of 3 (say between 10 and 90 MV/m, quite a reasonable range), then we can estimate the current within a factor of 3. Considering that field emission current varies over many orders of magnitude, this is a worthwhile exercise. More important, we see clearly the dependence of breakdown current on breakdown field (in terms of $K(E)$ as well as ρ, which depends on n, σ, and d, as well as the external field, which at breakdown would be E_b).

Writing out ρ (equation 7.16), we have

$$I_b = \frac{2\pi\epsilon_0 E_b d}{n\sigma\bar{t}K(E_b)}. \tag{7.25}$$

(Again, the ion lifetime depends on the applied field, E_b in this case.) From this formula, we see that breakdown will occur in an experiment when the product of the current I and neutral density n surpasses a certain threshold: breakdown occurs for

$$nI \gtrsim \frac{2\pi\epsilon_0 E_b d}{\sigma\bar{t}K(E_b)} \sim \frac{3\pi\epsilon_0 E_b}{\sigma K(E_b)}\sqrt{\frac{eE_b d}{2m_i}}\,. \tag{7.26}$$

The usefulness of this formula lies in the weakness of the right-hand side's dependence on E_b. Although any dependence on the unknown breakdown field E_b may seem to be a failure, it's less limiting than one might expect. For instance, the above formula answers the question: If E_{ext} is fixed, how high can the field emission current be before causing breakdown? If breakdown occurs at E_{ext}, then clearly $E_b = E_{\text{ext}}$, and therefore the breakdown current can be found. Presently we shall see that we can also answer the question: If E_{ext} is fixed, then given a certain amount of field emission, how high does the local gas density need to be to induce breakdown?

If the electron current is Fowler-Nordheim field emission, then the current density is[10] $J(E) = J_0(\beta E/E_0)^2 \exp[-E_0/(\beta E)]$, and

$$\frac{dJ}{dE}(E) = J_0 \left(\frac{\beta E}{E_0}\right)^2 \frac{1}{E}\left[2 + \frac{E_0}{\beta E}\right] e^{-E_0/(\beta E)} = \left[2 + \frac{E_0}{\beta E}\right]\frac{J(E)}{E}. \tag{7.27}$$

The total electron current is $I(E) = SJ(E)$, where S is the Fowler-Nordheim parameter heuristically associated with the emitter area, so

$$\frac{1}{I}\frac{dI}{dE} = \frac{1}{J}\frac{dJ}{dE} = \left[2 + \frac{E_0}{\beta E}\right]\frac{1}{E} \tag{7.28}$$

so $K(E) = 2 + E_0/(\beta E)$. This brings our breakdown condition to

$$nI \gtrsim \frac{3\pi\epsilon_0 E_b}{\sigma\left[2 + \frac{E_0}{\beta E_b}\right]}\sqrt{\frac{eE_b d}{2m_i}}. \tag{7.29}$$

Because we did not try to solve for I, the parameter S doesn't appear in the breakdown condition—an important advantage, because knowledge of S may be very uncertain. It is still somewhat unfortunate that the condition depends on the parameters E_0 and β from Fowler-Nordheim emission, since we might not know them precisely; however, the ratio E_0/β is more robust than either E_0 or β alone.

To find the conditions for breakdown at $E_b = E_{\text{ext}} \sim 30$ MV/m, with $\sigma \sim 3$ Å^2, $d \sim 4$ microns, $\beta = 250$ ($E_0 = 54.6$ GV/m) and mass 40 ions (argon):

$$nI \gtrsim 1 \times 10^{20}\ \text{A} \cdot \text{m}^{-3}. \tag{7.30}$$

If $I \sim 100$ μA, then breakdown occurs for $n \gtrsim 1 \times 10^{24}$ m^{-3}.

[10] If we actually know the current density so accurately, we can (at least numerically) solve for the breakdown field and current directly. However, we consider how this analysis holds up if we're unsure of the parameters β and S—suppose for instance that we think β is between 50 and 70. Experimental determination of the emitter area S varies exponentially with the intercept of the fitted line (on the Fowler-Nordheim plot); since the intercept of the fitted line varies roughly linearly with the fitted value for β (the slope of the line), S will be much more uncertain than β.

Summary and discussion of the model breakdown condition

One "properly solves" this model, which includes field emission, ionization and ion escape (to the cathode), and enhancement of the field by the ions, by:

0. choosing model parameters to represent a simplified physical reality:
   (a) d (ionization range), σ (ionization cross-section, within distance d of the cathode), and n (the neutral gas density within distance d),
   (b) E_{ext} (the externally applied electric field),
   (c) $I(E)$ (the field dependence of emitted electron current);
1. determining the existence of steady-state values for N_i (number of ions), I (electron current), and E (cathode field) by solving equation 7.3 (or more specifically, equation 7.19)—a system with no equilibrium is in the breakdown regime:
   (a) varying E_{ext} to find E_b, the breakdown field at the transition between the steady-state and the breakdown regimes;
   (b) finding the breakdown current I_b;
   (c) finding the equilibrium values of I, E, and N_i (if a solution exists).

Hypothetically, the experimentalist could inject neutral atoms at a field emitter to density n and apply a field E_{ext}, and the theorist could tell him what would happen (breakdown or not). Or, given all parameters but E_{ext}, the model would reveal *the* breakdown field E_b. In case of a stable solution, the theorist could predict, for example, the electron current (enhanced by the ions' field).

Unfortunately, $I(E)$ depends very sensitively on E, and that dependence is not extremely well understood or precisely known for a given emitter. We therefore

ask what we can learn from the model, assuming there is a definite and precise function $I(E)$ that we know only approximately. Allowing results (predictions) to depend on the total current I, as if it were a parameter of the model rather than an outcome, reduces the dependence of results on $I(E)$. For example, we write the derivative dI/dE as $IK(E)/E$, where $K(E)$ changes much slower with E than dI/dE.

An important result is the relationship between breakdown field and current:

$$I_b = \frac{K(E_b)}{\rho(E_b)E_b}, \tag{7.31}$$

where $K(E)$ reflects the properties of the field emission, and $\rho(E)$ reflects the mechanics of the model (rate of ionization, *etc.*). Since we're pretending we don't know $I(E)$ (although the model knows it[11]), we cannot solve this equation for both I_b and E_b. However, I_b and E_b are both measurable quantities, so this result makes a testable prediction, which depends on $I(E)$ only through $K(E)$.

Furthermore, many questions involve implicit knowledge of E_b. For example: what is the breakdown current when the applied field is E_{ext}? In this case, we're asking about breakdown at E_{ext}, so $E_b = E_{\text{ext}}$; then we can find the I_b corresponding to E_b.

Time evolution of perturbations of a steady state

At low applied field, hence low current, there exists a stable steady-state solution (actually two steady-state solutions; the one with lower field and current is stable) for I, E_i (or, equivalently, E), and N_i, given E_{ext}, $I(E)$, n, σ, and d. The I is

[11] It's one thing to assume that there is a function $I(E)$, and another to assume we know what it is. We assume there is a function $I(E)$, and use it in the model as a general form; the actual form of $I(E)$ is known only roughly.

the actual electron current that we would measure in an experiment in the steady state.

We now consider time-dependent variations from the steady state: $\Delta I(t)$, $\Delta E(t)$, and $\Delta N_i(t)$. The ion growth rate, from equation 7.9, is given by

$$\frac{d}{dt}(N_i + \Delta N_i) \sim \frac{I + \Delta I}{e} n\sigma d - \frac{N_i + \Delta N_i}{\bar{t}}. \tag{7.32}$$

If N_i and I (and the corresponding E) are equilibrium values, then $\dot{N}_i = (I/e)n\sigma d - N_i/\bar{t} = 0$, and

$$\frac{d}{dt}\Delta N_i \sim \frac{\Delta I}{e} n\sigma d - \frac{\Delta N_i}{\bar{t}}. \tag{7.33}$$

For small perturbations,

$$\Delta I \approx \frac{dI}{dE}(E) \cdot \Delta E = I\frac{K(E)}{E}\Delta E. \tag{7.34}$$

The field due to the extra ions ΔN_i is (see equation 7.13):

$$\Delta E \sim \frac{e\Delta N_i}{2\pi\epsilon_0 d^2} \tag{7.35}$$

and so

$$\Delta I \sim I\frac{K(E)}{E}\frac{e\Delta N_i}{2\pi\epsilon_0 d^2}. \tag{7.36}$$

Thus equation 7.33 becomes

$$\frac{d}{dt}\Delta N_i \sim \left[\frac{1}{\tau} - \frac{1}{\bar{t}}\right]\Delta N_i, \tag{7.37}$$

where

$$\frac{1}{\tau} = \frac{n\sigma d}{e} I\frac{K(E)}{E}\frac{e}{2\pi\epsilon_0 d^2} = \frac{In\sigma K(E)}{2\pi\epsilon_0 E d}. \tag{7.38}$$

Equation 7.37, valid for small perturbations, describes the exponential growth away from the steady state if $\bar{t} > \tau$, and exponential decay toward the steady state if $\bar{t} < \tau$. This applies to either steady-state, though only the stable steady-state

(with lower field and current than the unstable steady-state) is experimentally realizable; these steady-states exist only if E_{ext} is below the breakdown field E_b. When $E_{\text{ext}} = E_b$, the steady-states merge to one metastable steady-state (remember figure 7.2b). By the above reasoning, the steady-state is metastable when $\bar{t} = \tau$. Not surprisingly, the breakdown current I_b that achieves this condition (since $\tau \propto 1/I$) is:

$$I_b = \frac{2\pi\epsilon_0 E_b d}{n\sigma K(E_b)} \frac{1}{\bar{t}}, \tag{7.39}$$

which we found in equation 7.25, from slightly different considerations (though fundamentally the same).

This view of the breakdown condition ($\bar{t} \geq \tau$) considers the competition between the positive feedback loop that leads to ion growth ($\Delta N_i \rightarrow \Delta E_i \rightarrow \Delta I \rightarrow \Delta N_i$) and the ion loss due to ions hitting the cathode (which increases linearly with N_i). τ is the time constant for ion growth (ignoring ion loss), and $\bar{t}$ is the time constant for ion loss (ignoring ion growth). If $\bar{t} > \tau$, it takes longer for ions to hit the cathode than it takes for the ion cloud to grow, and breakdown occurs. At the threshold of breakdown, $\bar{t} = \tau$.

Using this model we can estimate the required time for the growth of discharge current during breakdown. Equation 7.37 shows that when $\bar{t} > \tau$ the number of ions grows exponentially with characteristic time $(1/\tau - 1/\bar{t})^{-1}$. Since this applies only to small perturbations, which quickly grow large with exponential growth, we can only estimate the time for the initial current growth during breakdown.

Using $K(E)$ for Fowler-Nordheim emission with $\beta = 250$, and our usual parameters ($E = 30$ MV/m, *etc.*—refer to the description before equation 7.30), and inserting the breakdown threshold values for nI from equation 7.30:

$$\tau \sim \frac{2\pi\epsilon_0 E d}{I n \sigma K(E)} \sim \frac{2\pi\epsilon_0 (3\times 10^7 \text{ V/m})(4\times 10^{-6} \text{ m})}{(1\times 10^{20} \text{ A}\cdot\text{m}^{-3})(3\times 10^{-20} \text{ m}^2)(9.3)} = 2\times 10^{-10} \text{ s}. \tag{7.40}$$

It turns out that these conditions are a little beyond the threshold for breakdown, so the time constant is $1/\tau - 1/\bar{t} \approx 1/\tau$, which we just calculated. This agrees well (well within an order of magnitude) with RF simulations (done with OOPIC for this work and by Knobloch with MASK [56]). In a DC simulation with similar parameters (except the field was constant at 30 MV/m), the initial current was about 200 μA (depressed by electron space charge from the expected 500 μA[12]); with $n \sim 1.6 \times 10^{24}$ m^{-3}, $I \cdot n$ is about three times the critical breakdown value; perhaps for this reason, breakdown occurred before the neutral gas had diffused more than about 1 micron from the cathode. Using $d \sim 1$ micron, we calculate

$$\tau \sim \frac{2\pi\epsilon_0(3 \times 10^7 \text{ V/m})(1 \times 10^{-6} \text{ m})}{(3 \cdot 1 \times 10^{20} \text{ m}^{-3}\text{s}^{-1})(3 \times 10^{-20} \text{ m}^2)(9.3)} = 2 \times 10^{-11} \text{ s}. \tag{7.41}$$

The predicted 20 ps rise-time compares well with the simulation (figure 7.3).

The breakdown condition applied to RF simulations

For RF simulations with the same maximum field, where the maximum current is about 50 μA, the average current (over a whole RF period) should be somewhat lower, on the order of 5–10 μA, putting the predicted critical gas density around 10^{23} m^{-3}; in simulations with very similar conditions, Knobloch found the critical density to be about 10^{24} m^{-3} [56]. This difference doesn't really need an explanation—it's just an order of magnitude calculation.

[12]This is a good example of the usefulness of this approach. Though we specified the field emission $I(E)$ for the simulation exactly, the actual field emission was only 2/5 of $I(E_{\text{ext}})$ because the electrons' space charge reduced the cathode field slightly below E_{ext}. Using $I(E)$ to solve for the breakdown conditions would have been folly, because the small change in electric field, which we did not take into account, seriously affected the field emission current. Despite knowing $I(E)$ exactly, the model would break down because E (at the cathode) is not known precisely. With such sensitivities, the advantages of our approach, in which we avoid relying on the exact form of $I(E)$, become clear.

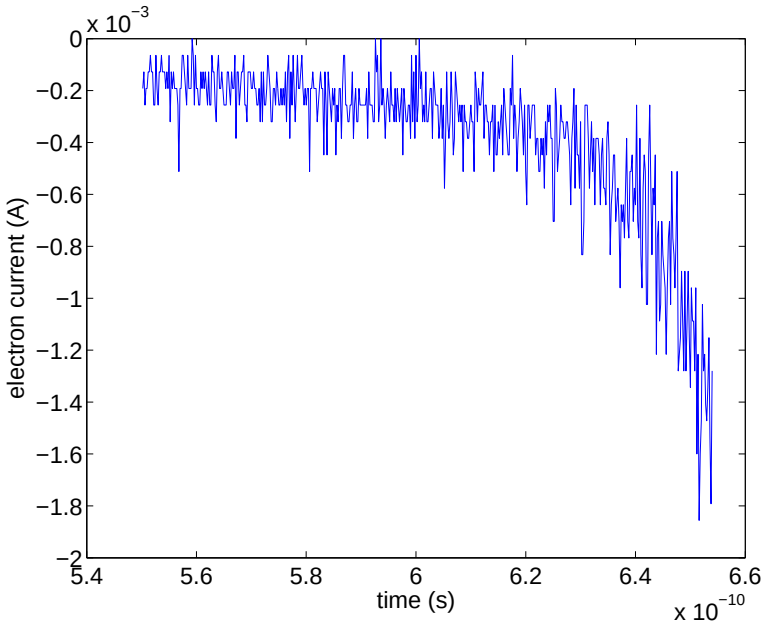


Figure 7.3: Current versus time for a DC breakdown (note that electron current is negative); the current is steady up to 6×10^{-10} s because it takes that long for neutral gas to diffuse far enough away from the cathode to be ionized. With neutrals one micron from the cathode, the electron current rises by a factor of ten in about 5×10^{-11} ps.

In similar RF simulations we found that breakdown occurred with ions of mass 40 at a density of 1.6×10^{24} m^{-3}, but not with the same density of mass 1 ions (and same cross-section); using mass 1 instead of mass 40 increases the critical product of $I \cdot n$ by a factor of 6.

Knowing the time required for a current rise, we can try to estimate better the threshold for RF breakdown, by comparing τ, $\bar{t}$, and $1/f$, where f is the RF frequency. Because of the steep dependence of field emission on electric field, significant electron emission occurs only during a fraction of the RF period—roughly 1/10 of the period for reasonable emission.

One important difference in an RF field is the motion of ions; in a uniformly oscillating electric field $E = -\tilde{E}\cos\omega t$, an ion of charge to mass ratio e/m_i (created at rest) oscillates back and forth:

$$x = x_0 + \frac{e\tilde{E}}{m_i\omega^2}\cos\omega t \tag{7.42}$$

$$\dot{x} = -\frac{e\tilde{E}}{m_i\omega}\sin\omega t. \tag{7.43}$$

An ion created when the electric field nears its peak (the most likely creation time, since electrons emerge only near the peak field, at $t = 0$ in the above equation) travels a distance $d_{\max} = 2eE/(m_i\omega^2)$ in the next half period; if ions are created within distance $d < d_{\max}$ of the cathode, then the ion lifetime is

$$t = \frac{1}{\omega}\cos^{-1}\left(1 - \frac{m_i\omega^2 d}{e\tilde{E}}\right). \tag{7.44}$$

For our simple back-of-the-envelope calculations, we'll want to choose a reasonable average $\bar{t}_{\rm RF}$ to reflect the range of d over which ions are created. For $d \ll d_{\max}$, ions hit the cathode relatively quickly, and (see equation 7.8)

$$\bar{t}_{\rm RF} = \bar{t} \sim \frac{2}{3}\sqrt{\frac{2m_i d}{e\tilde{E}}} \tag{7.45}$$

(using $E = \tilde{E}$, the peak field) is appropriate. On the other hand, ions which take a good part of a half period to reach the cathode travel with average velocity $\bar{v} = (2/\pi)e\tilde{E}/(m_i\omega)$, and

$$\bar{t}_{\mathrm{RF}} \sim \frac{d}{\bar{v}} = \frac{2m_i\omega d}{e\tilde{E}}. \tag{7.46}$$

If, on the other hand, $d > d_{\max}$, then ions don't escape under the electric field alone; in this case, their thermal velocity, mutual repulsion, and perhaps the magnetic field may come into play. Taking only the thermal velocity v_T into account, we may take

$$\bar{t}_{\mathrm{RF}} \sim \frac{d}{v_T}. \tag{7.47}$$

The ions' mutual repulsion will probably be more important, since the thermal velocity is relatively low, but this would add a non-linear contribution to the problem and make the calculation more difficult. Noting that increasing frequency generally increases $\bar{t}_{\mathrm{RF}}$ (especially at the point where ions can no longer reach the cathode), which makes breakdown more likely, we continue assuming that an appropriate $\bar{t}_{\mathrm{RF}}$ has been found.

If the state of the system (gas density, applied field strength, *etc.*) is such that breakdown can occur within a time less than about 1/10 of the RF period, or $(1/\tau - 1/\bar{t}_{\mathrm{RF}})^{-1} \ll 0.1T = 0.1/f$ (where $T = 1/f$ is the RF period and f is the frequency), then the problem is basically the DC problem, with the electric field almost constant at the peak field for the time of interest. Of course we knew that in the limit of low RF frequency, the RF condition for breakdown would tend toward the DC condition, and the speed with which breakdown can occur determines which frequencies are "low."

If breakdown cannot occur within time $T/10$ (as determined above), then estimating the breakdown condition for RF fields is more difficult. For a simplistic

estimate, we consider that the average current $\langle I \rangle$ (over a period) will be only fraction of the current I emitted at the peak field; roughly, $\langle I \rangle \sim I/10$. Therefore, we should use (instead of equation 7.38)

$$\frac{1}{\tau_{\rm RF}} = \frac{\langle I \rangle n \sigma K(E)}{2\pi\epsilon_0 E d} \tag{7.48}$$

and if $\tau_{\rm RF} < \bar{t}_{\rm RF}$ then breakdown will occur.

If ion growth were really exponential with time constant $\tau = (\langle I \rangle / I)\tau_{\rm RF}$ during 1/10 of the RF period, while ion loss caused exponential decay with time constant $\bar{t}_{\rm RF}$, then during the first tenth period, the ion cloud would grow by a factor $\exp(0.1T/\tau)\exp(-0.1T/\bar{t}_{\rm RF})$, but during the remaining part of the period the ion cloud would decay by $\exp(-0.9T/\bar{t}_{\rm RF})$, the result being growth (or decay) by a factor

$$\exp\left[\left(\frac{0.1}{\tau} - \frac{1}{\bar{t}_{\rm RF}}\right)T\right] = \exp\left[\left(\frac{1}{\tau_{\rm RF}} - \frac{1}{\bar{t}_{\rm RF}}\right)T\right], \tag{7.49}$$

and breakdown would occur when $\tau_{\rm RF} < \bar{t}_{\rm RF}$. However, the exponential growth occurs with the calculated time constants only near the steady-state. We must beware that during the 1/10 period when current is emitted, a significant ion cloud growth may put the system out of range of this simple linear model; and during the 9/10 of the period with no electron current, the ion cloud decay might remove the system far from its average "steady state" given the average current $\langle I \rangle$.

The case of RF breakdown needs to be more carefully considered than we have done here. However, for an order-of-magnitude estimate, the above might prove to be adequate. In keeping with our approach, in which we treat the current as an experimentally-measured quantity, rather than try to calculate it from the field emission model, we find that the breakdown condition is more naturally stated in

terms of the average current $\langle I \rangle$ instead of the actual emitted current I (emitted for only about 1/10 of the period).

Field-assisted thermionic emission

In finding this breakdown condition, we did not rely heavily on electron emission following Fowler-Nordheim, whose influence appears only in $K(E) = 2 + E_0/(\beta E)$. The same model, or a simple variation of it, will hold for many roughly exponential sorts of emission functions. Just for fun, consider that thermionic emission, for a constant temperature, depends on the electric field as [50]

$$J_T(E) = J_{T0} \exp\left(\sqrt{E/E_{T0}}\right) \tag{7.50}$$

so for thermionic emission, $K(E) = (1/2)\sqrt{E/E_{T0}}$ and the breakdown condition would be (neglecting E_i)

$$In \gtrsim \frac{3\pi\epsilon_0 E_{\text{ext}}}{\sigma \frac{1}{2}\sqrt{E_{\text{ext}}/E_{T0}}} \sqrt{\frac{eE_{\text{ext}}d}{2m_i}} = \frac{6\pi\epsilon_0 E_{\text{ext}}}{\sigma} \sqrt{\frac{eE_{T0}d}{2m_i}}. \tag{7.51}$$

Compared to Fowler-Nordheim-initiated breakdown, the different dependence on E_{ext} is significant and the parameter E_{T0} comes into play, although both have only a square root effect on the critical product of $I \cdot n$. By keeping the breakdown condition in terms of the product of I and n, so that the electron emission mechanism plays a relatively small role in $K(E)$, we avoid many problems associated with our lack of understanding of the electron emission.

Model weaknesses

The nature of field emission and field enhancement is the main weakness in our calculation of a critical product of current and neutral density. Although the model tries to make predictions less sensitive to the dependence of field emission on

electric field, the field-dependence remains important. We used enhanced Fowler-Nordheim emission for calculations, implicitly assuming that if ions created a field E_i at the cathode that field emission would increase as if the field E_i were enhanced by the factor β. If β is caused by geometric field enhancement (see section 4.1), the field E_i created by ions just microns away from the field emitter may be enhanced by a factor much smaller than β, which describes the enhancement of the field created by an electrode much farther away (compared to the emitter size). In the Fowler-Nordheim formula, $(\beta E_{\text{ext}} + E_i)$ may be more realistic than $\beta(E_{\text{ext}} + E_i)$. The ions' field may thus have considerably less effect than we estimated.

7.2.3 The Source of Neutral Gas

Voltage breakdown occurs when (and because) a highly conductive arc forms, necessarily comprising both electrons and ions. Field emission (or thermionic emission or a combination of both) provides electrons, while the ions result from the ionization of neutral atoms by collisions with electrons. This much is clear; the source of neutral gas, however, remains a mystery.

Vapor from electrodes has been detected in the microsecond before breakdown (by absorption of light tuned to an atomic transition of the electrode material) in experiments with a 1 mm gap [26]; this study is particularly important because it detected vapor before any significant ionization occurred. On average, more vapor was detected near the cathode than the anode; in at least some cases, no vapor was detected until less than a microsecond before breakdown. A mechanism was proposed involving material detachment from the anode, which would vaporize on its way to the cathode [27]; this mechanism is disputed in [79].

More studies subsequently aimed to determine which electrode first contributed vapor, although they observed light emission rather than absorption, so they measured not which vapor appeared first, but which vapor began to be ionized first (since light emission and ionization go hand in hand). By measuring the emission of light characteristic to the atomic transitions of electrode elements using different materials for the anode and cathode, Davies and Biondi found, again for DC voltages and a 1 mm gap, that anode material emitted light first [28] (therefore, anode material was probably ionized before cathode material).

Later experiments showed that the first vapor (anode or cathode) to emit light depended on the nature of the breakdown. When breakdown immediately followed the pulse rise, cathode vapor first emitted light, but when more than 10 μs passed between the pulse rise and breakdown, anode vapor first emitted light[13] [108]. The authors concluded that two kinds of breakdown mechanisms operated—one cathode-initiated, and one anode-initiated. The cathode-initiated breakdowns (which happened sooner after the pulse rise, consistent with other short-pulse studies; see especially [75]) tended to be at higher voltages. At a 1 mm gap, breakdown voltages tended to fall between 40 and 50 kV (40–50 MV/m).

In using these results to interpret our experiments we must be careful to note two major differences: the larger gap (and higher voltages), and the short pulse length. Perhaps the most relevant conclusion we can draw is that there can be different modes of breakdown (cathode or anode). Certainly in many of our experiments breakdown is intimately connected with a small area on the cathode; moreover, in RF experiments, which show the same features at the "cathode" (field

[13] A small amount of light was detected before slightly before the current rise.

emitter), the flying-anode-clump mechanism would not be feasible. Therefore, we continue to concentrate on possible cathode mechanisms.

In the model we discussed previously, we showed that the initiation of breakdown at the cathode can occur with field emission and neutral gas. Since we have observed field emission before breakdown, we know at least that ingredient is present; therefore, the question of trigger is reduced (within that model) to the question of how the neutral gas is generated.

At the cathode

We have shown, by rough calculation as well simulation, that with observed levels of field emission, neutral gas densities on the order of 10^{24} m^{-3} and higher can trigger breakdown. In that model, the density needs to be that high only near the field emitter—more than several microns away from the field emitter, electrons have too much energy and the ionization cross-section becomes small. That's just as well, because we are concerned here with vacuum breakdown, and if very much space were filled with gas at such high densities, it wouldn't be vacuum breakdown. High densities can exist, however, within a very small volume (cubic microns) without ruining the vacuum.

Atoms evaporating from a small, hot surface can create a locally high density. For example, at 3000 K a surface might emit a flux density of 10^{27} atoms/(m^{2}s) into the vacuum, causing breakdown by creating a local density of 10^{24} atoms/m^3. One square micron heated to 3000 K produces a total of 10^{15} atoms per second; a modest ion pump could consume that amount without letting the pressure rise much above 10^{-6} torr.[14]

[14]At room temperature, 10^{15} atoms per second is only 3×10^{-5} torr·L/s, an amount that would be absorbed by a 20 L/s ion pump (such as on the apparatus

To understand how much desorption must occur to cause breakdown, we consider desorption rates in terms of monolayers desorbed per nanosecond. A flux density high enough to cause breakdown, say 10^{27} atoms/(m^2s), corresponds approximately to 1 monolayer desorbed over 10 ns. This rate of desorption can produce sufficiently high local densities, but the desorption must continue long enough that the region of high density extends sufficiently far from the cathode; for example, if the density must be high between 1 and 5 microns from the cathode for ionization,[15] desorbed atoms traveling at thermal velocity (at 3000 K), roughly 1 μm/ns, would take at least 1 ns to travel one micron from the cathode. In this case, 0.1 monolayers desorbed over 1 ns could create a sufficiently high gas density within ionizable range of the cathode, giving rise to breakdown. The desorption of any less than 0.1 monolayers would be less likely to cause breakdown—such small amounts of gas could not sustain the high densities within range of ionization.

If breakdown can indeed be triggered by desorption of 0.1–2 monolayers, weakly adsorbed atoms (of which more than a monolayer exists) could be responsible for triggering breakdown—if they can all be released in a very short period of time. An explanation for breakdown involving the release of weakly adsorbed surface material is attractive because weakly adhered atoms are easily desorbed. However, the amount of weakly bound surface material is limited to a few monolayers at most, and maybe just a single monolayer. Desorption rates less than the minimum of 1 monolayer in 10 ns would not cause breakdown, but could deplete the supply of weakly adsorbed atoms (a monolayer) in as little as 100 ns. It strikes me as

used for this work) at a pressure of 2×10^{-6} torr, still fairly good vacuum (but the pressure rise would be high enough that it would certainly be noticed if it continued for any length of time).

[15] At typical breakdown fields around 50 MV/m = 50 V/μm, electrons between 1 and 5 microns from the cathode have energies best for ionizing an atom.

being very difficult to explain why a whole monolayer would remain intact until suddenly and almost completely desorbing within 10 ns. For comparison, note that baking a vacuum system to, say 300°C, more or less removes a monolayer of adsorbed material from the walls over a matter of hours.

We should also consider the energy required to desorb a monolayer of atoms. Bonds between atoms have strengths of several eV; weakly bound atoms have weaker bonds. Taking a convenient value, 1 eV, as the strength of a bond in the top monolayer, we can calculate the surface power density necessary to remove a monolayer in 10 ns: about 0.2 mW/μm^2 (or 200 MW/m^2, a rather high power density). In our simulations (chapter 6), we consider a reasonable Fowler-Nordheim field emitter, which emits about 100 μA from an area of about 0.1 μm^2, or about 0.1 mA/μm^2. To deliver the kind of power we need to desorb gas would therefore require the current to undergo a voltage drop of 2 V, assuming it could transfer that energy into gas desorption without loss.

There may be some mechanism by which emitted current can transfer energy fairly directly into the surface layer, for instance Nottingham heating (which can also be Nottingham cooling is some circumstances) with Fowler-Nordheim field emission [100, 50]. Joule heating is a possibility that needs to be considered: for a mediocre metal with a resistivity ρ=10 $\mu\Omega$·cm, the voltage drop along a field-emitting protrusion with area A=0.1 μm^2 would be $\rho/A \sim 1\Omega/\mu$m—far from being able to create a voltage drop of 2 V even with currents more than a milliamp over a several micron length. On top of that, transferring the energy into desorption through emitter heating will be very inefficient—much of the heat will dissipate. Other models for field emission involving electron flow through an insulating oxide

layer may be more promising in terms of providing strong and immediate surface heating mechanisms [107].

Because desorption due to heating accelerates steeply with temperature (*i.e.*, $\sim \exp[-(1\text{ eV})/kT]$), a slower temperature rise could possibly lead to a rather sudden desorption of gas. In this case, simple Joule heating might be a possibility.

At the anode

In chapter 4.1.1 we discussed experimental evidence of gas desorption; the evidence pointed to the anode as the source. Electrons deliver a good deal more total power to the anode than to the cathode (thanks to the voltage gain across the gap), although the power is distributed over a greater volume (in depth as well as area, since energetic electrons penetrate several microns through metal). However, gas desorbed at the anode cannot lead to breakdown in the same way that we found with gas desorbed at the field emitter (that is, a cathode-based initiation by field emission with neutral gas from the anode); especially for larger distances between anode and cathode (low-frequency RF cavities are an extreme case), such an explanation would involve filling the entire volume with a fairly high pressure, something which is not seen experimentally.

Gas desorbed at the anode could play a role in other kinds of breakdown triggers. For example, if neutral atoms emitted from the anode were ionized, they would bombard the cathode with a significant amount of energy (the ionization is admittedly unlikely, because the ionization cross-section is low for keV and higher electrons, though secondary electrons might have a better chance; in any case the gas density is low). In a parallel-plane DC gap, atoms desorbed at the anode by electron bombardment and ionized would probably bombard the cathode near

the field emitter; however, in an RF cavity, ions created at the point of electron bombardment would probably not end up anywhere near the field emitter.

Because anode-initiated mechanisms tend to be hard-pressed to explain the trigger of voltage breakdown in RF cavities (which show similar activity near the field emitter as in DC breakdown), we have concentrated more on cathode initiation mechanisms. We consider two possibilities here: increased neutral density caused by field enhancement, and neutral avalanches caused by a feedback loop with ion bombardment, sputtering, and ionization.

7.2.4 Field-Enhanced Neutral Density and Helium Processing

Enhanced field emission is often attributed to microprotrusions that enhance the local electric field by a factor β, often measured to be around 100 (assuming that field enhancement alone contributes to enhanced field emission). Although some doubts have been cast upon the existence of real microprotrusions with such high enhancement factors (especially $\beta \gg 100$), there is without doubt validity to geometric enhancement causing smaller enhancements with β from 2–50.

Electric field enhancement can increase the local gas density as well as field emission; high-field regions attract gas molecules due to their induced dipole moments. In an electric field, the electron cloud of an atom shifts slightly with respect to the nucleus, giving the atom a dipole moment.[16] A constant electric field induces a dipole moment in gas molecules, but no net force; a spatially varying

[16] Some molecules, like water, also have a permanent dipole moment that remains even at zero field; even so, the dipole moment changes in an applied electric field. While an induced dipole moment is parallel to the electric field, a permanent dipole moment depends on the orientation of the molecule; the electric field exerts a torque on a permanent dipole moment, trying to align it with the field, but if

electric field, however, exerts a net force on a dipole because the positive charge sees a slightly different field than the negative charge. The electric force on an induced dipole attracts the dipole toward increasing field magnitude. Therefore the enhanced electric field at a protrusion pulls molecules toward the protrusion, creating a local enhancement of gas density.

We will find that field-enhanced density cannot generally provide the source of neutral gas for vacuum breakdown, but it may explain the technique of helium processing in superconducting RF cavities [54]: superconducting RF cavities with troublesome field emission can sometimes be cured by applying the RF field with a low density of helium gas in the cavity. Processing cavities with helium seems to induce field-emitters to break down at lower fields than when the cavities are under vacuum; helium processing is useful when field emission limits cavity performance, and available power is insufficient to reach fields necessary for breakdown.

A polarizable molecule in a non-uniform field

A polarizable molecule gains a dipole moment in an electric field, which pulls the electron cloud and the nucleus in opposite directions; in a uniform field, the nucleus and the electron cloud feel equal but opposite forces, and the molecule feels no net force. A non-uniform electric field, however, pulls harder on one side of the dipole than the other, exerting a net force on the molecule toward the region of stronger field. To calculate the increase in density within a pocket of extra-strong field, we need to know the potential energy of a dipole in an electric field. We will find the

---

the molecule has angular momentum, it will not simply align with the field, but precess around it.

potential energy of a polarizable molecule in an electric field to be:

$$\Delta\epsilon = -\frac{1}{2}\alpha E^2 \tag{7.52}$$

where α is the atomic polarizability (in SI units).[17] A molecule thus gains kinetic energy as it moves toward stronger fields—in other words, it's attracted to stronger fields.

Table 7.2 shows the atomic polarizabilities for a selection of gases.

The reader who already believes equation 7.52 may wish to skip the following justification.

To derive equation 7.52, consider a real (induced, non-permanent) dipole $p = qd$; as the electric field increases, the separation d between the positive and negative charges increases: $p(E) = qd(E)$. At very small separation distances d, the separated charges will attract each other with a linear (Hooke's Law) restoring force $F = -kd$. The applied field pulls the charges apart until the restoring force equals the force of the applied field:

$$qE = kd(E). \tag{7.53}$$

[17] The atomic polarizability α of a gas molecule can be measured from the dielectric constant of the gas; an electric field $\mathbf{E}$ induces a dipole moment $\mathbf{p} = \alpha\mathbf{E}$ in each molecule. If n is the gas (number) density, the polarization, or dipole moment per unit volume, is $\mathbf{P} = n\mathbf{p} = n\alpha\mathbf{E}$. The dielectric susceptibility χ relates the polarizability and the field, $\mathbf{P} = \chi\epsilon_0\mathbf{E}$ (in SI units); $1 + \chi = \kappa$ (in SI units) is the dielectric constant that an experiment would measure directly. Therefore, $n\alpha = \chi\epsilon_0$; the dielectric constant of a gas along with its density measure the atomic polarizability.

Note that χ and κ are dimensionless, while α/ϵ_0 has dimensions of volume.

Perhaps it should be said that $\mathbf{P} = n\mathbf{p} = n\alpha\mathbf{E}$ is true only for densities n small enough that intermolecular interactions remain negligible. Gases at room temperature and below are quite ideal—practically non-interacting. It would be a mistake, however, to apply that relation to calculate the dielectric constant of a solid.

Table 7.2: Measured dielectric susceptibilities (χ) from [65] (at 20°C and 1 atmosphere), and the atomic polarizabilities calculated therefrom. Susceptibilities of gases are relatively insensitive to time-variation of electric fields from DC to optical frequencies [106]; molecules with no permanent dipole moment (including all in this table [65]) should have susceptibilities independent of temperature. Polarizability is calculated from χ according to $\alpha = \epsilon_0 \chi / n$ (see footnote 17), where the density n is calculated from the ideal gas law, $p = nkT$ (in this formula, p is atmospheric pressure).

Gas	χ	α/ϵ_0 (m^3)
H_2	2.5×10^{-4}	10×10^{-30}
He	0.65×10^{-4}	2.6×10^{-30}
N_2	5.5×10^{-4}	22×10^{-30}
CO_2	9.2×10^{-4}	37×10^{-30}
O_2	4.9×10^{-4}	20×10^{-30}
Ne	1.3×10^{-4}	5.2×10^{-30}
Ar	5.2×10^{-4}	21×10^{-30}
Kr	7.8×10^{-4}	31×10^{-30}
Xe	13×10^{-4}	50×10^{-30}
SF_6	20×10^{-4}	80×10^{-30}
C_4H_{10}	26×10^{-4}	100×10^{-30}

Experimentally, we measure the change in p, not d, so we measure (more directly) the atomic polarizability α that satisfies $p = \alpha E$, rather than the "spring constant" k. We can find k from $qd(E) = p(E) = \alpha E$ and the equilibrium condition $qE = kd(E)$: $k = q^2/\alpha$.

When the dipole moves within some potential $V(\mathbf{x})$ (we denote the potential at the positive charge V_+ and V_- at the negative), the total work done is the sum of the work to move each individual charge; however, if the field changes, the charges will move different distances as the equilibrium length between them changes. Since d is very small compared to the distance moved, the total work may be considered the sum of the work to move the charges to the new location and the work required to stretch the dipole to its new equilibrium length. Using subscripts i and f to indicate initial and final values,

$$\begin{aligned} W &= q_+(V_{+f} - V_{+i}) + q_-(V_{-f} - V_{-i}) + \frac{1}{2}k(d_f^2 - d_i^2) \\ &= q[(V_{+f} - V_{-f}) - (V_{+i} - V_{-i})] + \frac{1}{2}k(d_f^2 - d_i^2) \\ &= q\left.\frac{dV}{ds}\right|_f d_f - q\left.\frac{dV}{ds}\right|_i d_i + \frac{q^2}{2\alpha}(d_f^2 - d_i^2) \\ &= \mathbf{p}_f \cdot \nabla V|_f - \mathbf{p}_i \nabla V|_i + \frac{1}{2\alpha}(p_f^2 - p_i^2) \\ &= (-E_f p_f + E_i p_i) + \frac{1}{2}(p_f E_f - p_i E_i) = -\frac{1}{2}(p_f E_f - p_i E_i). \end{aligned} \tag{7.54}$$

Therefore, the energy (up to a constant, of course) of a real dipole with linear restoring force in an electric field is

$$\epsilon_{\text{dipole}} = -\frac{1}{2}p(E)E = -\frac{1}{2}\alpha E^2. \tag{7.55}$$

Enhanced density

Since the enhanced electric field near a protrusion with field enhancement β pulls molecules toward the protrusion, molecules will congregate in higher densities

around the protrusion. In the previous section we found that the potential energy of an induced dipole in an electric field E is $-p(E)E/2 = -\alpha E^2/2$. The energy difference between a molecule in the enhanced region (field βE) and the unenhanced region (field E) is

$$\Delta\epsilon = -\frac{\beta^2 - 1}{2}\alpha E^2. \tag{7.56}$$

If the molecules are in thermal equilibrium at temperature T, the Boltzmann factor will determine the ratio of the enhanced density (in field βE) to the background density (in field E):

$$\frac{n_\beta}{n_0} = \exp\left[-\frac{\Delta\epsilon}{kT}\right] = \exp\left[\frac{\beta^2 - 1}{2}\frac{\alpha E^2}{kT}\right]. \tag{7.57}$$

At high enhanced densities, the dielectric effect of the gas will start to reduce the field, at which point the above formula becomes invalid; *i.e.*, we require that in the enhanced region, $1 + \chi = 1 + n_\beta\alpha/\epsilon_0 \approx 1$, or

$$n_\beta\frac{\alpha}{\epsilon_0} \ll 1. \tag{7.58}$$

As n_β rises toward and above ϵ_0/α, the field of the polarized molecules reduces the applied field. Since (for many molecules) $\alpha/\epsilon_0 \sim 10^{-29}$ m^3, enhanced densities up to $n_\beta \sim 10^{28}$ m^{-3} will have little effect on the field.[18] As an aside: gases with smaller polarizabilities can reach greater field-enhanced densities (if a sufficiently high electric field can be applied).

As we will see later, $n_\beta/n_0 \sim e^1$—in other words, the term in the exponent is of order unity, confirming the possibility of significant geometric density enhancement (the exponent doesn't have to be much higher than 1 to create a large density

[18] At atmospheric pressure, 300 K, the density of an ideal gas is 2×10^{25} m^{-3}; at 3 K, atmospheric pressure would be 2×10^{27} m^{-3}.

enhancement). However, it turns out that for pressures much less than 10^{-3} torr or room temperature, density enhancement is likely to be negligible. There is one notable situation in which geometric density enhancement may contribute to breakdown: helium processing in superconducting cavities.

Helium processing in superconducting RF cavities

Occasionally superconducting RF cavities with field emission problems can be processed (field emitters can be eliminated) by adding a small quantity of helium gas inside the normally evacuated cavity during processing.[19] Typically one fills a cavity with about 1 mtorr of helium, as measured at room temperature—much more helium would risk RF breakdown of the gas [54].[20] In the cavity, where the gas is at temperature T (for most accelerator cavities, between 2 K and 4.5 K), the helium density is higher by a factor $\sqrt{T_r/T}$,[21] where $T_r \sim 300$ K is the temperature at which 1 mtorr (0.14 Pa) is measured. The gas density in the cavity is therefore:

$$n_0 = \frac{0.14\,\mathrm{Pa}}{kT_r}\sqrt{\frac{T_r}{T}}. \tag{7.59}$$

For $T = 2$ K, $n_0 = 4 \times 10^{20}$ m^{-3}.

[19] Helium gas is used because, unlike all other gases, it doesn't freeze to the walls of a superconducting cavity at 2 K or 4 K.

[20] With too much helium, gaseous breakdown (not vacuum breakdown) is possible—electron avalanches and so forth, having nothing to do with field emission. A superconducting cavity in gaseous breakdown reaches only very low fields. To be effective, processing must induce vacuum breakdown, which quickly (in less than a microsecond) dumps the stored energy into a small area of the surface, delivering enough power density to destroy that area, and hopefully the troublesome field emitter with it.

[21] The current $J = nv$ must be equal at the two temperatures, or there will be a net flow from one to the other: $J(T_1) = J(T_2)$ implies that $n_1\sqrt{kT_1/m} = n_2\sqrt{kT_2/m}$, or $n_2/n_1 = \sqrt{T_1/T_2}$.

For helium at 2 K, the density enhancement is negligible until

$$\beta E \gtrsim \sqrt{\frac{2kT}{\alpha}} \approx 1 \text{ GV/m}. \tag{7.60}$$

Local fields βE higher than this provide significant density enhancement. At room temperature, however, βE would have to be more than ten times higher (*i.e.*, $\gtrsim 10$ GV/m) to produce any density enhancement; such high fields are very unlikely, even locally.

For $\beta = 100$ and $E = 50$ MV/m, for example, applied to helium at 2 K:

$$\frac{n_\beta}{n_0} = \exp\left[\frac{100^2 - 1}{2}\frac{2.5 \times 10^{-30}\text{ m}^3\epsilon_0(50 \times 10^6\text{ V/m})^2}{k \cdot 2\text{ K}}\right] = e^{10} = 2 \times 10^4. \tag{7.61}$$

At densities used for helium processing, $n_\beta \sim 9 \times 10^{24}$ m^{-3}. At 2 K this density has a pressure of about 2 torr; very important for *gas* density enhancement, the vapor pressure of liquid helium at 2 K is above 2 torr (although at 1.4 K, the vapor pressure is close to 2 torr). If the gas pressure rises above the liquid vapor pressure, the gas will condense on the emitter.

From section 7.2.2 we estimated the critical product of field emission current and local gas density necessary to initiate breakdown at a field emitter:[22]

$$n_\beta I \gtrsim \frac{3\pi\epsilon_0 E}{\sigma\left[2 + \frac{E_0}{\beta E}\right]}\sqrt{\frac{eEd}{2m_i}}$$

[22]In this equation, E is the electric field, σ an average cross-section for the ionization of helium by electron impact, $E_0 = 54.6$ GV/m is a parameter from Fowler-Nordheim emission, d is the distance within which ionization occurs, and m_i is the mass of a helium ion. Helium has a relatively low ionization cross-section compared to other gases; a rough average for electrons with energies between 20 eV and 500 eV might be about 0.25×10^{-20} m^2 [91]. The value for d can be hard to estimate (fortunately the result depends only on the square root of d: one on hand, 500V/E=10 μm, although since electrons initially gain energy in the enhanced region, 500V/βE=0.1 μm is more appropriate; on the other hand, the region of enhanced field and density is not larger than the size of the emitter tip, which for $\beta \sim 100$ should be 0.1 μm or less; therefore, we choose $d \sim 0.1$ μm.

$$\sim \frac{3\pi\epsilon_0(50\,\mathrm{MV/m})}{0.25\times10^{-20}\,\mathrm{m}^2\left[2+\frac{54.6\times10^3\,\mathrm{MV/m}}{100\cdot50\,\mathrm{MV/m}}\right]}\sqrt{\frac{e(50\,\mathrm{MV/m})(0.1\,\mu\mathrm{m})}{2(4\cdot1.67\times10^{-27}\,\mathrm{kg})}}$$
$$= 1\times10^{21}\,\mathrm{A/m}^3. \tag{7.62}$$

For helium processing at 2 K, $n_0 \sim 4\times10^{20}$ m^{-3}; $n_\beta \sim 9\times10^{24}$ m^{-3}, so the field emission current I necessary to initiate breakdown at $E = 50$ MV/m must be

$$I \gtrsim 100\,\mu\mathrm{A}. \tag{7.63}$$

Note that the ionization cross-section of helium is a factor of ten less than most other gases [91], making it less likely to cause breakdown. For superconducting cavities, helium is the only choice; however, even at higher temperatures where other gases could be used, helium may be no worse for its small ionization cross-section, since gas is added to a cavity until (just before) RF breakdown of the gas occurs; presumably helium, with 1/10 the cross-section, can be filled to 10 times the pressure without risking RF breakdown of the gas.

Time dependence

Once the applied field E turns on, atoms start to become trapped in (sufficiently) field-enhanced regions. An atom falling into the trap has enough kinetic energy to escape again, unless it loses energy in a collision—most likely a collision with the emitter. Collisions with other atoms may be rare, given typical densities, and in any case, may not so easily result in the trapping of an atom. The interaction between a solid surface and an atom is quite strong—in many cases (for instance in calculating evaporation fluxes from vapor pressures) atoms are assumed to stick to the surface with a probability near unity. Therefore, atoms that enter the trap very likely become thermalized (lose their extra kinetic energy gained when falling into the trap) if and when they hit the emitter.

A common rule of thumb maintains that a (room temperature, medium weight) gas at a pressure of 1×10^{-6} torr can deposit one monolayer per second; that is, atoms hit the containing wall with a flux J of one per second per (10 Å^2), or $J = 10^{19}$ m^{-2}s^{-1}. This is the rate at which the number of trapped atoms could grow. Taking into account species and temperature,

$$J \sim n v_T \sim \sqrt{\frac{kT}{m}} \tag{7.64}$$

where v_T is the thermal velocity. For a trap size $d = 0.1$ μm, and a background pressure $p_0 \sim 0.1$ mtorr,[23] we expect the time to increase the density by a factor of n_β/n_0 to be about

$$t_{\text{density growth}} \sim \frac{n_\beta d}{J} \sim \frac{n_\beta d}{n_0 v_T} \tag{7.65}$$

or n_β/n_0 times the time it takes a typical atom to travel a distance d. For our example with helium processing[24]

$$t_{\text{density growth}} = 2 \times 10^4 \frac{0.1\ \mu\text{m}}{60\ \mu\text{m}/\mu\text{s}} = 3\ \ \mu\text{s}. \tag{7.66}$$

Another aspect of time dependence that we should consider: how fast can the dipole follow a time-varying electric field? If the electric field suddenly changes direction, the electron cloud will have to move to the other side of the nucleus; the inertia of the electron limits the speed with which the dipole moment of an atom can follow a time-varying field. However, the inertia of an electron is very small, and the polarizability remains substantially the same up to optical frequencies (as demonstrated by measured values in [106]). Treating the electron-nucleus as a simple spring-mass oscillator with spring constant $k = q^2/\alpha$, where q is the charge

[23] 0.1 mtorr is equivalent to about $n_0 \sim 4 \times 10^{20}$ m^{-3} at 2 K.

[24] At 2 K, the thermal velocity of helium is about 60 μm/μs.

of the nucleus and α the atomic polarizability, the system has a resonant frequency

$$\omega = \sqrt{\frac{k}{m}} \sim \sqrt{\frac{q^2}{\alpha m}} \gtrsim \sqrt{\frac{e^2}{\alpha m_e}} = 3 \times 10^{16}\ \mathrm{s}^{-1}. \tag{7.67}$$

Without knowing how various atomic electrons contribute to the polarizability, q and m are uncertain, but presumably q/m has to be equal to its value e/m_e for a single electron. Since ω is very high, we apply a lower bound by using the lowest q^2/m possible given q/m equal to that of an electron.

In summary, RF fields of many GHz should be like DC as far as the induced dipole is concerned.

Effect of RF field

Because the electric field induces the dipole moment, the sign of the field doesn't matter—atoms are always attracted to increasing field magnitude, which is always toward the emitter. As the field crosses zero, twice in each period, atoms can escape the trap. At 2 K, helium atoms travel about $v_T \sim 60$ nm/ns; as long as the enhanced region is larger than 100 nm (for an RF frequency of 1 GHz or so), the atoms will still be trapped, although the field E used to calculate the density enhancement might not be the peak surface field, but some effective field (perhaps the RMS field, for instance).

At 2 K, atoms travel slow enough that any atom passing through the enhanced region should hit the emitter and become trapped (see appendix E); that may shorten the equilibration time somewhat.

At room temperature, helium atoms travel much faster—about 800 nm/ns, fast enough to escape the 100 nm enhanced region within 0.1 ns. Therefore, a 1 GHz field would realize only the density enhancement created within the time the field

stays near the peak, which will be negligible (to increase the density by a factor of 2×10^4 takes 3 μs; in 0.3 ns the density could increase by a factor of 2).

Processing of RF cavities at room temperature by this method would require a heavy gas—xenon, with atomic mass 131, has $v_T \sim 140$ nm/ns at room temperature.

Thermal equilibrium

Collisions between the neutral gas and the electrode or cavity surface will bring the neutral gas into equilibrium with the surface temperature, though some time (microseconds, as discussed above) may be needed for the enhanced density to reach its equilibrium level.

The equilibrium density may in some circumstances depend strongly on the surface-gas interaction. Ideally one would like to know the vapor pressure of gas atoms adsorbed onto the surface; however, if gas does condense on the surface, preventing the predicted density enhancement, gas atoms will eventually cover the surface, at which point the vapor pressure of the gas (in solid or liquid state, at the given temperature) will determine the maximum enhanced density reachable—above that pressure, the gas will condense on the surface.

If the emitter is hotter than the surrounding surface, that temperature (and not the surrounding surface temperature) must be used in equation 7.57 to calculate the density enhancement factor. Hotter emitters lead to smaller density enhancements.

Criticism of the enhanced density model

The enhanced density model does not explain vacuum breakdown in general; with typical vacuum pressures less than 10^{-8} torr, breakdown requires (roughly) density

enhancements greater than 10^8, which are very unlikely, especially at temperatures above a few Kelvin. This model may be able to explain processing techniques with relatively high vacuum pressures (in the mtorr range)—in particular helium processing of superconducting RF cavities at just a few Kelvin. However, the explanation is borderline; based on our estimates above, parameters need to be set at extreme (measured) values—for instance, the local field βE needs to be about as high as one could reasonably imagine, and the field emission current needs to be more than 100 μA, also on the upper end of the reasonable range. The following paragraphs specifically point out some weaknesses of the model.

First, the enhanced density model presented above relies on the extreme enhancement of electric field at a field emitter; although we often measure high field enhancements $\beta > 100$, based on the Fowler-Nordheim model, it is doubtful that such field enhancements commonly exist, although their existence would explain observed levels of field emission (see section 4.1).

Second, the density enhancement may not be significant except at very low temperatures. Higher temperatures reduce the density enhancement; at room temperature, only gases with very large polarizabilities at very high fields will allow any density enhancement. Therefore, hot field emitters may defeat density enhancement.

Third, in estimating the density enhancement we neglected any detailed structure of the field in the enhanced region—in particular the field is everywhere less than βE (only at the cathode surface is the field βE); the region in which significant density enhancement occurs may be much smaller than we assumed, since the density enhancement depends very sensitively on the local electric field.

Fourth, in considering RF breakdown, we only provided an extremely rough estimate of the effect of zero-crossings of the field, which could allow atoms to escape the enhanced region. A more careful treatment is needed to find the peak fields necessary for significant density enhancement.

Last: if our estimate of the critical product of current and density from section 7.2.2 is too low, then the model of enhanced density would have difficulty explaining even helium processing.

7.2.5 A Neutral Avalanche

We wondered whether all the neutral gas could be created by ion bombardment; a chance ion could bombard the cathode near the field emitter, possibly sputtering off some neutrals. If the sputtered neutrals were ionized, they would bombard the cathode, emitting more neutrals, and so on, developing a large cloud of neutrals and ions. However, we find from back-of-the-envelope calculation that such an avalanche of neutral gas requires too much electron current (and current density) to be feasible.

The possibility of an avalanche

The neutral-avalanche problem is perhaps best considered from the point of view of an atom in an electron beam. Each atom created at the cathode within the electron beam has some probability of being ionized and then sputtering more neutral atoms off the cathode. If this process results on average in a net gain of atoms, then it will lead to an exponential growth of atoms (see figure 7.4).

Consider the case of an electron beam of radius r emerging (by field emission) from the surface; let I_e be the electron number current ($I_e = I/e$), and

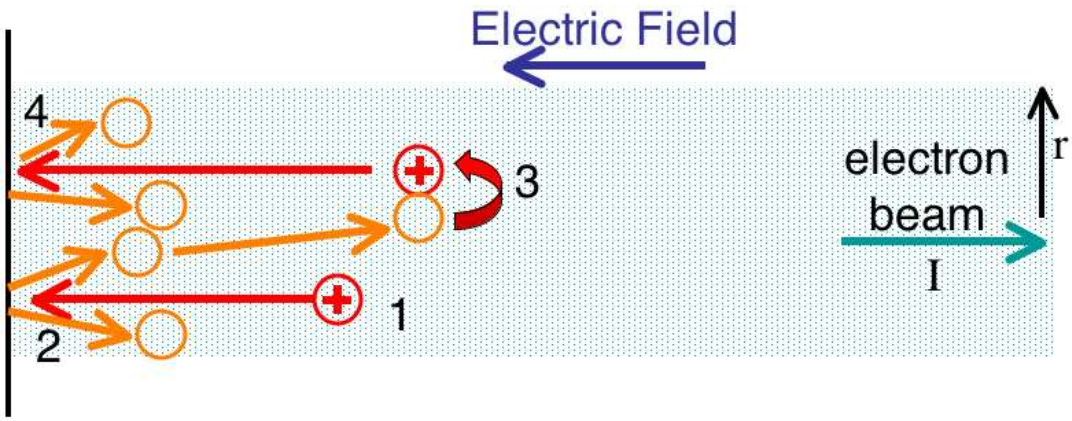


Figure 7.4: A neutral avalanche, caused by ion bombardment of an electron emitter: (1) By "chance" an ion is created; (2) the ion sputters two neutral atoms off the cathode; (3) the electron beam ionizes one of the sputtered atoms; (4) the ion sputters two neutral atoms off the cathode, *etc.*

$J_e = I_e/(\pi r^2)$ the number current density. We then calculate the probability that an atom is ionized as it travels within the electron beam, figuring out first the probability that an atom is ionized within time Δt, and next, the time Δt that an atom is likely to remain within the electron beam.

If an atom spends time Δt in the beam, the probability that it remains non-ionized is[25]

$$P(\text{not ionizing in } \Delta t) = \lim_{\delta t \to 0} \left[1 - J_e \sigma \delta t\right]^{\Delta t/\delta t} = \exp\left(-J_e \sigma \Delta t\right) \tag{7.68}$$

leaving the probability of ionization equal to:

$$P(\text{ionizing in } \Delta t) = 1 - \exp\left(-J_e \sigma \Delta t\right). \tag{7.69}$$

[25] An atom in the electron beam has an ionization rate equal to $J_e\sigma = n_e v_e \sigma$, where σ is the cross-section for an electron (of given energy) ionizing an atom. The ionization rate (as used here) times δt is the probability that an atom is ionized within time δt, for small δt.

An atom created within the electron beam will travel within the electron beam for a time of the order $r/v_{i,\perp}$, where r is the electron beam radius and $v_{i,\perp}$ is a typical velocity of an atom perpendicular to the direction of the beam.[26]

During time $\Delta t \sim r/v_{i,\perp}$ that an atom spends within the electron beam, the atom will be ionized with a probability equal to:

$$1 - \exp\left(-\frac{J_e \sigma r}{v_{i,\perp}}\right) = 1 - \exp\left(-\frac{I_e \sigma}{\pi r v_{i,\perp}}\right). \tag{7.70}$$

If each ion bombards the surface within the electron beam and sputters off Y atoms (on average), then each ionized atom creates on average

$$Y\left(1 - \exp\left(-\frac{I_e \sigma}{\pi r v_{i,\perp}}\right)\right) \tag{7.71}$$

new atoms—but sacrifices itself in the process. Therefore, the average net gain of atoms per atom is

$$\Delta N_{\text{per atom}} = Y\left(1 - \exp\left(-\frac{I_e \sigma}{\pi r v_{i,\perp}}\right)\right) - 1. \tag{7.72}$$

If the above "average net gain per atom" is positive, then exponential growth of atoms proceeds; if negative, the number of atoms decays to zero. Clearly the sputter yield Y must be greater than one for any growth. Growth is encouraged by higher currents I_e and smaller beam radii r.[27] Table 7.3 shows the minimum critical current I_{crit} necessary for positive $\Delta N_{\text{per atom}}$, given $v_{i,\perp}$ and various beam radii r and sputter yields Y.

Table 7.3 also shows the current $I_{\text{lns}} > I_{\text{expl}}$ necessary not just to create an explosion of neutral atoms, but to create that explosion within less than about a

[26]Only motion perpendicular to the beam can move the atom outside of the beam (neglecting beam spread).

[27]However, for small beam radii (smaller than the typical ionization distance), ions have a greater chance of bombarding the cathode outside the electron beam, greatly decreasing the probability that sputtered atoms will become ionized.

Table 7.3: Critical electron current needed to start a neutral avalanche (explosion), and the slightly higher current necessary to cause an explosion in about 1 ns, for two values of Y, the number of sputtered atoms per incident ion.

	$v_{i,\perp} = 10\mu$m/ns			
	$Y = 2$		$Y = 1.1$	
$r(\mu\text{m})$	I_{crit}(mA)	$I_{\tau=1\text{ns}}$(mA)	I_{crit}(mA)	$I_{\tau=1\text{ns}}$(mA)
10	1200	2700	4000	17000
1	120	140	400	470
0.1	12	12	40	41
0.01	1.2	1.2	4.0	4.0
0.001	0.12	0.12	0.40	0.40
1Å	0.012	0.012	0.040	0.040

nanosecond. Only for large electron beams does the possibility of slow explosions arise.

In light of experimental evidence showing typical field emission currents below 100 μA, this analysis favors small beams—in fact, only beams of a few nanometers or less can lead to a neutral avalanche with experimentally observed pre-breakdown currents. However, smaller beams decrease the probability that ions will bombard the cathode within the beam radius, which decreases the probability that sputtered atoms will become ionized. Moreover, a low-energy electron beam of 100 μA can't be kept within a few nanometers radius due to its own space charge repulsion. The current densities required for a neutral avalanche are simply too high for the neutral avalanche mechanism to be a feasible source of neutral gas. Also, we must remember that an ion needs (at least) hundreds of eV to achieve a sputter

yield greater than one, while most ions will gain less than 200 eV (unless multiply ionized) because electrons usually ionize best with around 100 eV.

The critical point dividing growth and decay corresponds to $\Delta N_{\text{per atom}} = 0$, an average net gain of zero:

$$Y\left(1-\exp\left[-\left(\frac{I_e\sigma}{\pi r v_{i,\perp}}\right)_{\text{crit}}\right]\right)\sim 1 \tag{7.73}$$

or

$$\left(\frac{I_e\sigma}{\pi r v_{i,\perp}}\right)_{\text{crit}} = -\ln\left(1-\frac{1}{Y}\right) = \ln\frac{Y}{Y-1}. \tag{7.74}$$

The speed of avalanche growth

To estimate how fast the number of atoms can grow, we consider the growth to take place in stages, each stage with a net gain of $\Delta N_{\text{per atom}}$. Each stage lasts the time it takes for an atom to be ionized plus the time for the ion to travel back to the cathode. Neutral atoms are sputtered off with a few eV [19], while ions gain more energy from the electric field[28] so the time for the ion's trip can be neglected; each stage lasts approximately the time it takes an atom to become ionized. For an atom in the beam the mean ionization time will be $\delta t \sim 1/(\sigma J_e) \sim \pi r^2/(\sigma I_e)$. If $N(t)$ is the number of atoms at time t, then

$$N(t+\delta t) \sim Y\left(1-\exp\left(-\frac{I_e\sigma}{\pi r v_{i,\perp}}\right)\right)N(t), \tag{7.75}$$

and the rate of growth is

$$\frac{\delta N}{\delta t} \sim \frac{\sigma I_e}{\pi r^2}\left[Y\left(1-\exp\left(-\frac{I_e\sigma}{\pi r v_{i,\perp}}\right)\right)-1\right]N, \tag{7.76}$$

[28]Ions must gain at least 10–20 eV, since an electron must have at least that energy to ionize an atom; however to be able to sputter off a neutral atom, ions need more than 50 eV.

which is exponential growth/decay with a time constant of

$$\tau \sim \frac{\pi r^2}{\sigma I_e} \frac{1}{Y\left(1 - \exp\left(-\frac{I_e \sigma}{\pi r v_{i,\perp}}\right)\right) - 1}, \tag{7.77}$$

which is positive (negative) for growth (decay).

Table 7.3 shows the currents $I_{\tau=1\text{ns}}$ necessary for growth with a time constant of 1 ns; except for large beams, the growth of the number of atoms happens very fast if at all.

In addition to the above calculations, we ran a few simulations (using OOPIC) with field emission, ionization, and sputtering of neutrals due to ion bombardment of the cathode, and found that chance ions could not lead to an avalanche of neutral gas by the model described above. In fact, even increasing the electron current to unreasonably large values ended up creating more beam spread, so that we still couldn't get an avalanche of neutrals.

7.3 The Arc

Once the discharge is triggered, ions come into play, the electric field at the cathode is enhanced, the current rises, heating increases, *etc.* All of this added field and energy tends to increase current even further; in particular, field emission (or TF emission, a combination of field and thermionic emission) skyrockets. More electron current means more ions; ions in turn bombard the cathode, causing more heating and vaporization. At this point only the available power limits the discharge.

The arc stage of voltage breakdown is characterized by a high conductivity at high currents (often 10–1000 A) that make sustaining a high voltage impossible.

An arc very efficiently creates the ions which are necessary to maintain such high currents at low voltages;[29] the voltage required to maintain the arc falls to 10–20 V (comparable to typical ionization energies). Often, the voltage changes little even as the arc current changes over an order of magnitude or more. Almost all this voltage drop occurs near the cathode during the arc [50].

The information we have about vacuum arc structure and behavior usually comes from experiments very different from those in which we have studied DC and RF breakdown. Perhaps the biggest difference is that experiments designed to probe vacuum arcs often have high-current circuitry to maintain the arc in a sort of steady state at 10–1000 A for longer times (microseconds to milliseconds) while measurements are made. In our DC breakdown experiment we see similar currents (10–100 A), but they last for only a half microsecond. However, in many cases, the damage done to the electrodes appears to be quite similar; moreover, arcs appear to be made of smaller components that have relatively short lifetimes.

Vacuum arcs (hereafter called simply "arcs") exhibit surprisingly complex structures, considering that they result from a rather serious but simple escalation of current-increasing processes. Perhaps because the voltage drop occurs right at the cathode, and not throughout the arc, the cathode is the natural focus for understanding arcs. The basic unit of an arc is the cathode spot (see, for example, [50, 42, 4, 6]); a single arc may consist of many cathode spots. Sometimes the cathode spot is even considered to be a sort of arc quantum of current emission [74].

[29] Electrons alone run into space-charge limits, requiring high voltages to maintain high currents.

Cathode spots correspond to the craters created by an arc.[30] Cathode spots tend to be small—typically crater-sized (1–100 microns)—with large current densities. Direct measurements of cathode spot current density is difficult, but current experimental estimates find current densities in the range of 10^8–10^{12} A/m^2. Such high current densities reach the current densities necessary to melt cathode protrusions by Joule heating alone, and indeed cathode spots are associated with the cratered surfaces typical of electrodes subjected to arcing.

Much of cathode spot behavior remains unexplained [42]. Presumably cathode spots form because shrinking the current into a small area is energetically advantageous. For efficient voltage to current conversion, arcs must create charged particles as easily as possible. Since heating, thermo-field-emission, cathode vaporization, and ionization all increase together as energy is concentrated into smaller areas, lateral growth of an arc may be discouraged; larger-area arcs, with less current density, heating, *etc.*, may be un(self)sustainable at low arc voltages.

Concentration of energy for more efficient current production may explain why cathode spots form, but perhaps less well understood is why cathode spots disappear. The typical cathode spot has a tenuous grip on life, lasting 10 ns to 10 μs (observation of light emission from cathode spots seems to provide the high end of the range, while current fluctuations and laser absorption images of plasma seem to suggest spot lifetimes on the order of 100 ns or less). An arc can survive the death of a cathode spot as long as other cathode spots operate simultaneously;

[30]The correspondence between craters and cathode spots may not be one-to-one, although that may depend on how one defines the lifetime of a cathode spot, and whether cathode spots move or whether they die as a new spot forms nearby. However, the association is strong; the action of a cathode spot on the cathode produces a crater.

when (perhaps through random fluctuation) all cathode spots die at once, so does the arc.

As we consider starburst and crater formation, it is helpful to consider results from experiments measuring currents and velocities for neutrals and ions in vacuum arcs. Ion currents make up approximately 10% of the total arc current (with electrons forming the rest) [52]. An early and thorough series of experiments [29] measured the velocities of neutral and ionized atoms for different electrode materials emitted radially and axially from the arc. Due to measurement resolution, the energy of neutral atoms could be determined only to be with 0–2 eV (but not higher); ion energies were surprisingly high, however, often several times higher than the arc voltage. It appeared that the ion kinetic energy was proportional to the ion charge, *e.g.*, Cu^{+2} had twice the energy of Cu^{+1}, suggesting that a large potential hill (two or three times the arc voltage) existed in the arc region; this conclusion was later disputed (see, *e.g.*, [59] which shares an author with [29]). The same experiment also determined that multiple ionization was very common (for example, 30% Cu^{+1}, 54% Cu^{+2}, 15% Cu^{+3}, and 0.4% Cu^{+4}, with a detectable but small fraction of Cu^{+5}). Increasing the arc current increased the ion flux (emitted radially from the arc), while slightly decreasing the ion velocities. More recently ion velocities have been measured in arcs on a large number of electrode materials by a few different methods [59, 92, 103, 111, 110, 18]; in general they find the same high ion velocities, 20–100 eV, but see less dependence of the ion velocities on the ion charges.

Measurements of the plasma density at cathode spots during a fully-developed arc, with a current between 10 and 50 amps, can exceed 10^{26} m^{-3} in jets roughly 10 microns in size, which appear to emerge from the cathode [6]. Oddly, arcs

with higher currents did not show plasma spots of such high density (the smallest detectable density was about 10^{25} m^{-3} for a 10 micron size plasma). The high density plasma spots appear to be connected with the ejection of molten droplets from the cathode.

7.3.1 Starburst Formation: Surface Removal

In our experiments we are interested in the damage done to the cathode by the arc. Like everybody else, we often see craters at breakdown sites. Unlike others, we often see starbursts around breakdown sites on DC cathodes of different materials as well as in superconducting niobium RF cavities of different frequencies.

We have learned that starbursts are areas in which the surface layers have been etched away; we believe that ion bombardment during the arc removes several monolayers of the surface. A scanning electron microscope can reveal starbursts, which have a different secondary electron emission coefficient than the surrounding areas, perhaps because of the removal of carbon contaminants on the surface. The contrast in secondary emission images varies for different starbursts; the contrast that makes starbursts visible is quite subtle (starbursts contain less surface carbon contamination), coming only from a difference in a few superficial monolayers—it just happens that secondary emission images are particularly sensitive to such subtleties.

There could be several reasons why other experiments have not seen starbursts even if starbursts such as we know them were formed. Our electrodes were fairly clean, chemically but never mechanically polished, and not baked or conditioned with many breakdown events.[31] Conditioning an electrode could certainly obscure

[31]Especially breakdown experiments in the 1970s and earlier almost always used mechanically polished electrodes (with the notable exception of metal films de-

starbursts by creating them everywhere—in our experiments we noticed that the most violent breakdowns often showed masses of melted regions without identifiable starbursts. Also, the difference in surface carbon between the starburst and the surrounding region does not extend to the craters in the starburst; whereas surface carbon is depleted within most of the starburst, it's often higher around a crater. It is also possible that starbursts show up particularly well on niobium, a material we use because of our interest in superconductors; starburst contrast does seem greater on niobium than on copper or gold. Finally, having found them, we took pains to continue to look for starbursts.

Let us consider the creation of starbursts—that is, surface removal—by ion bombardment. One monolayer contains about 10^7 atoms/μm^2. Starburst sizes vary, but a typical size (for our DC starbursts) is around 2500 μm^2, with about 2.5×10^{10} atoms/monolayer. If each bombarding ion removed one atom on average (a sputter yield of 1), we'd need 2.5×10^{10} ions/monolayer. If a starburst were formed in 100 ns, that would require an ion current of 0.04 A (total charge: 4 A·ns); if a starburst were formed in 10 ns, an ion current of 0.4 A would be required (with the same total charge).

The discharge current (in the DC experiments) typically lasted about 400 ns; often the current rose immediately (within 10 ns) to a value around 40 A, and declined steadily over that period, giving an integrated current around 10^4 A·ns; we expect about 10% of this current to be carried by positive ions [52]. Of course, only a small fraction of this may go toward making a starburst—more than one

posited on mechanically polished substrates), usually baked the electrodes to high temperature (300°C or higher), and very often subjected the electrodes to many arcs to condition them.

starburst can be formed during a single arc; also, much of that current must be devoted to the crater (or craters) in the starburst.

The sputtering rate is difficult to know precisely—it depends on the impact energy and the composition of the sputterer and the sputteree; (singly charged noble gas) ions more massive than helium with several hundred eV sputter on average between 0.3 and 3 atoms per ion, depending on the target material [19]. Ions with energy as low as 50 eV can still remove surface atoms, but at rate of 100 ions per atom. The ions in the arc probably have a broad distribution of energies that may well include ions with energies at or above 100 eV [18], so they can sputter, but the yield may be low: perhaps 0.1 atom per ion.[32]

Since the maximum arc current may be around 40 A, about 4 A of which is ion current, we might expect at most 0.4 A to be ions that actually knock off surface atoms, capable of removing one starburst monolayer in as little as 10 ns. We are tempted to connect starbursts with single cathode spots; our approximation of starburst creation time suggests starbursts may form in 10–100 ns, and cathode spot lifetimes are often estimated to be within the same range. Both estimates are subject to uncertainty, however.

7.3.2 Starburst Formation: Symmetry and Streamers

Often, especially on niobium cathodes and in niobium cavities, the starburst left by an arc comprises a circular region that breaks up into streamers beyond a certain radius; usually the starburst centers around a central crater, or a central patch of craters. The streamers eventually taper to a point, but sometimes widen or split before doing so (figure 7.5). Many of these starbursts display remarkable

[32]This is a rough estimate, extrapolated in several different dimensions.

symmetry: not only do streamers begin to form at about the same radius, but the streamers often bulge or split at about the same radius as well (as in figure 7.5). Clearly the starburst formation is still organized as a whole even after it breaks up into streamers.

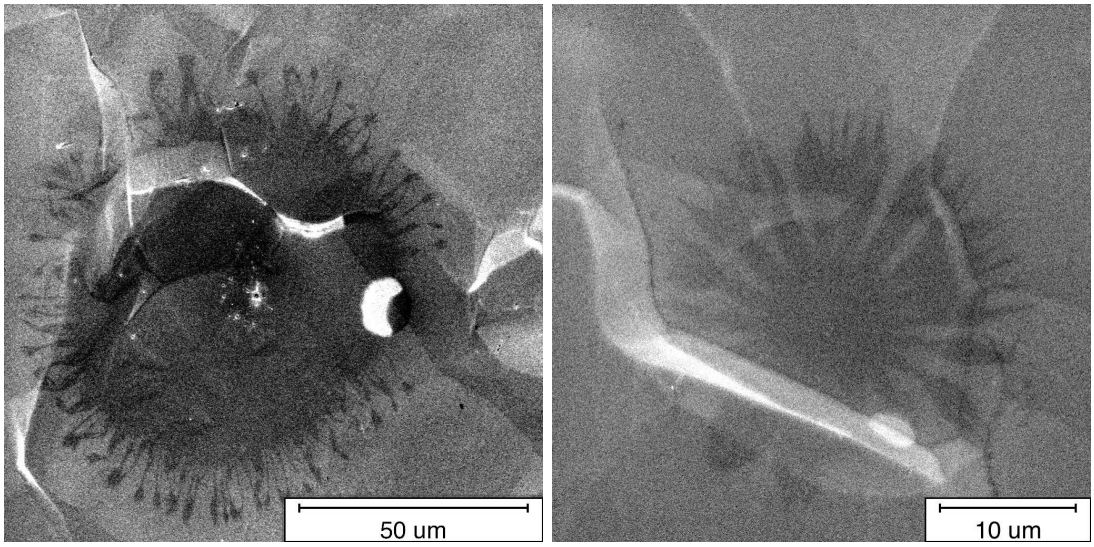


Figure 7.5: Eventually starburst streamers taper to a point, but sometimes they do interesting things first.

7.3.3 Streamer Formation

Imagine a plasma expanding rapidly into a non-ionized region of neutral gas due to a strong electric field. If the plasma "front" is quite thin (if the transition from plasma to neutral gas is sharp), it may be considered to be a moving conductor bounding the neutral, charge-free region, in which the electric potential satisfies $\nabla^2\phi = 0$.

Under the influence of an electric field $\mathbf{E}$ and collisions with neutral gas, electrons travel at an average speed $\mathbf{v} = \mu\mathbf{E}$ (this defines the mobility μ under the assumption that collisions make the average velocity proportional to the electric field).

In addition, thermal effects cause diffusion of the plasma in every direction, according to

$$\frac{\partial \sigma}{\partial t}_{\text{diffusion}} = D\nabla^2\sigma \tag{7.78}$$

where $\sigma(\mathbf{x}, t)$ is the plasma density, and D the appropriate diffusion constant. The mobility and diffusion constant are in general connected by the Einstein relation:

$$\frac{\mu}{D} = \frac{e}{kT} \tag{7.79}$$

where e is the (magnitude of the) electronic charge, and kT Boltzmann's constant times the electron temperature, which usually governs plasma diffusion.[33]

A plane-front of plasma will move perpendicular to the electric field at a speed $v = \mu E$. If the plane is allowed to "ripple," the ripples will enhance/diminish the surface electric field at peaks/troughs, thus increasing the "ripplicity." However, diffusion will tend to iron out any surface curvature. To calculate the stability of the surface, we impose infinitesimal ripples of wave-number k, calculate the perturbed surface field, and then compare the effects of the perturbed field to those of diffusion.

Specifically we want to consider ripples of wavevector k (≥ 0) and amplitude a,

$$x = f(y) = a\cos ky. \tag{7.80}$$

If the electric field far from the plasma front is E_∞, then for very small amplitude ripples the potential and its derivatives are: (see appendix C)

$$\phi \approx -E_\infty[x - a\cos(ky)e^{-kx}] \tag{7.81}$$

[33]Note that the mobility and diffusion constant relate to the plasma, not to its individual species (although they are all related). For example, D should be the ambipolar diffusion constant, not the diffusion constant for either electrons or ions.

$$-\nabla\phi \approx E_\infty \left[\hat{\mathbf{x}}(1 + ak\cos(ky)e^{-kx}) + \hat{\mathbf{y}}ak\sin(ky)e^{-kx}\right] \quad (7.82)$$

$$-\nabla^2\phi = 0. \quad (7.83)$$

This approximate solution for the potential exactly satisfies Laplace's equation and the far-right boundary condition that the field should be of magnitude E_∞ in the $\hat{\mathbf{x}}$-direction. However, the potential only approximately vanishes at the boundary of the conducting plasma, $x = f(y)$, and only for small a:

$$\phi(f(y), y) = -E_\infty \left[a\cos(ky) - a\cos(ky)e^{-ak\cos(ky)}\right] \quad (7.84)$$

$$= -aE_\infty \cos(ky)\left[1 - e^{-ak\cos(ky)}\right]$$

$$= 0 + aE_\infty \times O(ak). \quad (7.85)$$

Therefore, this potential satisfies the boundary condition at the plasma front only for $ak \ll 1$. For x near the front ($x \ll 1/k$),

$$\phi(x, y) \approx -E_\infty[x - a\cos(ky)] \quad (7.86)$$

$$E_x \approx E_\infty[1 + ak\cos(ky)] \quad (7.87)$$

$$E_y \approx E_\infty ak\sin(ky). \quad (7.88)$$

For small perturbations $a \ll 1/k$, the maximum and minimum surface fields are $E_\infty(1 \pm ak)$ in the x-direction. The ripple peak travels at speed $\mu E_\infty(1 + ak)$ and the ripple trough at speed $\mu E_\infty(1 - ak)$; the peak thus moves away from the trough at speed $2\mu E_\infty ak$, leading to an increase in ripple amplitude satisfying (for $ak \ll 1$)

$$\frac{da}{dt} = ak\mu E_\infty. \quad (7.89)$$

This equation describes exponential growth with time constant $(k\mu E_\infty)^{-1}$. Any surface perturbation to the plasma front thus increases because of the positive feedback between the front and the field; the growth of the perturbation is linear

in k (shorter wavelength perturbations grow faster). If this were the only mechanism, infinitely short wavelength disturbances would grow infinitely fast; however, diffusion irons out the small-scale perturbations.

Diffusion of the plasma evens out ripples in the plasma front, described by $x = f(y,t)$. Considering diffusion only in the y-direction (diffusion in the x-direction, perpendicular to the front, should be outweighed by average motion of the front due to the electric field), the diffusion equation tells us that

$$\frac{\partial}{\partial t} f(y,t) = D\partial_y^2 f(y,t) \tag{7.90}$$

where D is the diffusion constant. A sinusoidal perturbation $f(y,t) = a(t)\cos(ky)$ grows at a rate

$$\frac{da}{dt} = -Dk^2 a. \tag{7.91}$$

Perturbations with wave-number k are thus damped exponentially with time constant $(Dk^2)^{-1}$.

Adding the effects of diffusion to the electric field on a plasma front rippled with wave-number k, we see that the infinitesimal amplitude $a(t)$ of the rippling changes according to

$$\frac{da}{dt} = (\mu E_\infty k - Dk^2)a = (\mu E_\infty - Dk)ka. \tag{7.92}$$

For $k > k_0 := (\mu/D)E_\infty$ the perturbation dies out; but for $0 < k < k_0$, the perturbation grows (linearly) unstably, with maximum growth at $k = k_0/2$,

$$\text{maximum growth :} \qquad k = \frac{\mu E_\infty}{2D}. \tag{7.93}$$

If μ and D are connected by the Einstein relation, then the wavelength of maximum instability is

$$\text{maximum growth :} \qquad \lambda = 4\pi \frac{kT}{eE_\infty}. \tag{7.94}$$

(Note that it's usually the electron temperature that matters, which is far above room temperature for a plasma.)

For $E_\infty = 10$ MV/m (note that the field that causes the starburst plasma to expand would be almost tangential to the cathode; it's not the field responsible for field emission and ion bombardment of the cathode) and $kT \sim 5$ eV, our estimated wavelength of maximum instability is

$$\lambda = 4\pi \frac{kT}{eE_\infty} \sim 6\ \mu\text{m}. \tag{7.95}$$

That happens to be quite close to actual starburst streamer sizes (see section 5.1); however, the estimates of E_∞ and kT were more or less picked out of a hat (especially E_∞—I don't know what the field at the edge of the plasma is likely to be).

There are many reasons why this calculation might not represent the experiments very well. First, it assumes a mobility and diffusion constant reasonable for a plasma in a background neutral gas—while neutral atoms are certainly present in a vacuum arc, they are spewed from the cathode, which might alter ordinary diffusion. Second, the calculation completely ignores the dimension between the anode and cathode, including any dynamics involving the plasma sheathe and cathode spot; starbursts are certainly two-dimensional phenomena, but we have to be more clever about how we remove the third dimension from the calculation. Third, approximating the plasma edge as the boundary between vacuum and conductor is not very realistic.

However, some simulations have been done in a completely different situation [3]—breakdowns in N_2 at room temperature and standard pressure, using a much simplified fluid-dynamical plasma model, with $\mu = 380\ \text{cm}^2/\text{Vs}$ and $D = 1750$

cm^2/s (so by Einstein's relation, $kT = 4.6$ eV). They find that the streamer[34] branches into fingers spaced at roughly 230 μm. While it's not at all clear that the calculation presented above is valid for the experiments discussed in this work, it ought to be very valid for this situation; the disagreement suggests that the rough calculation is much too rough; however, it does at least agree with [3] in predicting the instability to grow linearly with k for small k.

In considering other possible models for starburst formation, it's perhaps important to note that we have very occasionally seen starbursts without any sign of cratering within them (figure 7.6). Therefore, the streamer-formation mechanism is not inevitably tied to the possibly explosive processes that excavate craters on the cathode surface.

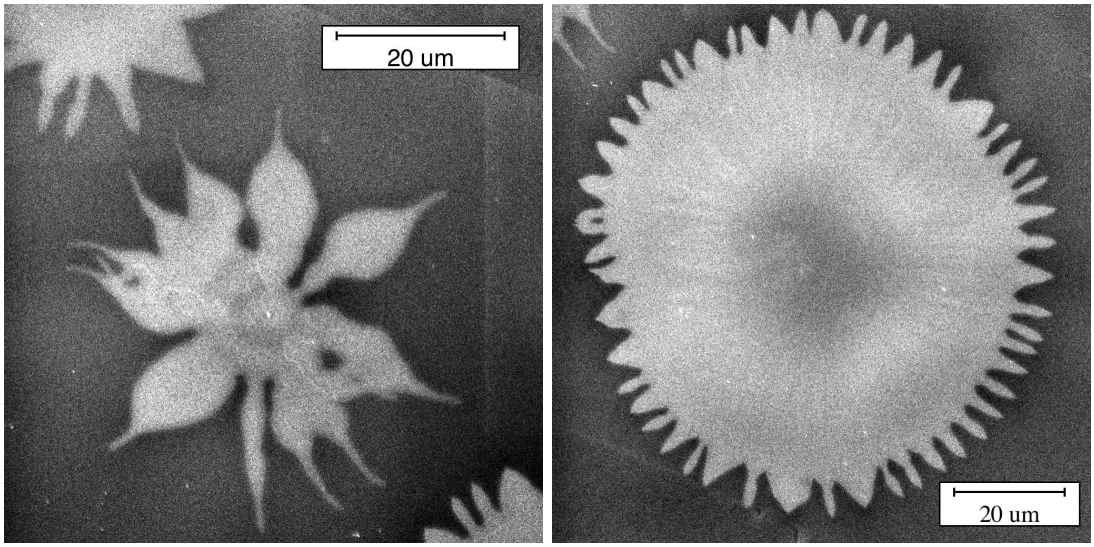


Figure 7.6: Craterless starbursts on a diamond-machined copper surface.

[34]Here "streamer" is the Townsend-discharge streamer created by electron avalanche; it travels between electrodes, unlike our starburst streamers, which are patterns on the electrode.

7.3.4 Crater Formation

The rate of cathode erosion has been measured for vacuum arcs to be between 10 and 100 μg/C (micrograms material eroded per coulomb of arc current) [4, 59]; the latter article suggests that smaller type I craters (micron-sized or less), usually found on contaminated cathode surfaces, have low erosion rates, whereas larger type II craters, found on pure metal surfaces, have higher erosion rates. Typical DC discharges that we've seen carry a total of 10 μC. On a niobium cathode, such a discharge would erode a volume of 10–100 μm^3, certainly enough to excavate the sort of craters we see on cathodes.

Crater formation has been widely discussed [75, 96, 50]. It is thought that heating by the electron current and ion bombardment melts the cathode surface around the field emitter. The pressure of ion bombardment forces the liquid metal out, creating a raised rim and violently spewing metal droplets far from the crater. After the arc, the metal solidifies, leaving the familiar crater.

From a more philosophical point of view, crater formation may relate directly to the fundamental requirement that a vacuum arc produce its own vapor. A higher power density at the arc cathode spot enhances all the processes that generate arc current—higher temperatures increase electron emission and cathode vaporization, which increase ionization, which increases ion bombardment, which increases heating, *etc.* Vacuum arcs can better sustain themselves when they concentrate their energy into small areas to increase the efficiency of the important processes (especially electron emission and cathode vaporization). Since the arc current is basically neutral, with ions balancing electrons, smaller arcs mean higher magnetic fields, which in turn could help confine the arc. Opposing the confinement of the arc is the strong force of diffusion—plasmas are notoriously difficult to confine; in

particular, electrons, being very light, diffuse very easily, and the field they create becomes strong enough to make ions diffuse at a similar rate (ambipolar diffusion). All the processes that maintain the arc in the cathode spot work, perhaps less efficiently, on the border of the cathode spot, where they try to increase the size of the spot. If the arc cathode spot becomes too large, the arc may not have enough energy to continue, and the cathode spot will die. The confinement mechanism remains to be thoroughly understood [42].

Influence of surface oxides on starbursts and craters

Because the necessity of vaporizing cathode material to maintain an arc requires such high power densities that cathode spots must be small, the cathode surface naturally gets vaporized in small spots—these are the craters that we see (though it's possible that the depth of the craters has more to do with the necessity of producing a lot of vapor to initiate the arc, *c.f.*, explosive electron emission [75]). The size of the craters depends on the opposing forces promoting confinement and diffusion. Often the literature makes a distinction between type I and type II craters [50]: type I craters tend to be small and isolated, occurring on surfaces that may have a little contamination (some oxide or surface carbon), while type II craters tend to be larger and adjacent, occurring on very clean surfaces. We infer that some "weakness" in the surface, relating to contamination, attracts arc cathode spots, whereas a uniformly clean surface offers no special points, and new cathode spots form preferentially next to old ones. Although it's possible that a small patch of contamination is the "weakness" that attracts a cathode spot (perhaps the metal-oxide interface somehow enhances field emission), a metal surface "contaminated" with oxide or surface carbon is generally completely covered

with oxide or carbon. Moreover, if the surface were clean except for a smattering of contaminants, one might expect a number of isolated type II craters (because around each patch is clean metal), rather than only type I craters.

More likely, in my opinion, contamination inhibits crater formation. Of course, once the arc starts, it immediately burns through the contamination at its cathode spot (as it vaporizes the bulk metal); however, the widening of the crater is slightly inhibited. For example, most oxides are insulating dielectrics, which reduce the surface field at the metal. If the crater is surrounded by dielectric, the electric field around the edge of the cathode spot may be reduced (at the metal-oxide interface), suppressing field emission, making it harder for the cathode spot to grow. While a cathode spot is active (conducting large amounts of arc current), arc plasma diffuses far from the cathode spot (the plasma density must still be reasonably high across a starburst-sized region); while most current is within a few microns (a crater-sized spot), plasma exists across the whole starburst. Although the surface oxide may inhibit cratering, the surface oxide is too thin to last long. The "weak" areas that lead to type I cratering could therefore be weaknesses in the oxide, where the dielectric breaks down. The gold cathodes that we tested were very susceptible to desultory type I cratering (see figure 5.12 on page 128); some copper cathodes also exhibited such widespread isolated craters, whereas niobium usually allowed fewer craters. The natural oxide on gold is almost non-existent, on copper just a few monolayers, and on niobium around 50 angstroms. The oxide therefore seems to protect the surface somewhat, but even 50 angstroms of oxide can be burned through fairly easily.

Much thicker oxides (hundreds of angstroms, as opposed to tens of angstroms for the natural oxide) seem to limit the plasma expansion so that the whole star-

burst is only 10–20 microns, rather than 20–100 microns (for our DC experiments), as shown in figures 5.10 and 5.11. When a thick oxide helps to contains the arc, the cathode spot can grow larger than the typical crater without becoming unstable due to diffusion. Within the small starburst, the power density is higher than within a larger starburst, and so the area within the small starburst becomes severely melted—the entire starburst is a crater. With the starburst region a lake of liquid metal, the plasma pressure (due to ion bombardment of the cathode) forces molten metal out around the edges, just as with smaller craters. Often starbursts on heavily oxidized cathodes exhibit a column in the center. If not enough molten metal can be pushed out around the edges, it possible that the plasma pressure will force metal up in the center of the crater, which then freezes when the arc dies.

7.4 The End (of the Arc)

The end of the arc may come simply as a result of running out of energy—the arc exhausts the stored energy in the capacitance of the gap or in the RF cavity, and if the power source can't maintain the arc, it dies. The possible correspondence between starburst size in different-sized RF cavities and the stored energy tends to support this theory.

The end of the arc, which may comprise several cathode spots, may also be related to the dying of cathode spots, which occurs when the cathode spots become too large too maintain the power density at the cathode necessary for efficient sustainability of arc processes (electron emission and cathode vaporization). Cathode spots may have shorter lifetimes than the arc, but new cathode spots can be cre-

ated (in fact, are more likely to be created because of the already existing plasma near the cathode). However, as the arc depletes the stored energy, new cathode spots will become less likely to form. The end of the arc may occur when all existing cathode spots die before a new spot is created.

Chapter 8

Summary

Voltage breakdown in vacuum can be a particular problem for RF cavities in particle accelerators; however, voltage breakdown can cure field emission problems and prevent future breakdown. DC voltage breakdown experiments (in vacuum) have shown many similarities to voltage breakdown in RF cavities, particularly in the damage done to the surface around the field emitters that presumably initiate breakdown. Concentrating on cathode-initiated triggers of breakdown equally applicable to RF and DC breakdown, we investigated a scenario in which the sudden desorption of neutral vapor at a field emitter, in combination with the emitted electrons, quickly develops into a highly conductive plasma—an arc that causes voltage breakdown. Computer simulations showed that sufficient amounts of field emission and neutral vapor could create an ion cloud that enhanced the electric field at the emitter; the field of the ion cloud eventually dominated the formation and evolution of the arc.

8.1 Probing (DC) Voltage Breakdown

Particles, field emission, and breakdown

A common feature of both RF and DC breakdown is the enhancement of field emission and the increased likelihood of breakdown on surfaces contaminated with particles. We have shown conclusively, by identifying particles beforehand in DC breakdown experiments, that particles can lead to breakdown, and that breakdown removes all but a trace of the original particle at the center of a starburst; also, video images of arcs during breakdown show bright light emitted from particle sites where starbursts are later found.

Although field emission almost always preceded breakdown, in a few cases no field emission (above about 20 nA) was observed even microseconds before breakdown.

Contaminant particles, when present, determine the field emission and breakdown characteristics of a surface, independent of substrate material—for cathodes of niobium, copper, gold (film on niobium), oxidized niobium, and oxidized copper. While contaminated cathodes often break down at fields between 20 and 100 MV/m, clean cathodes (from careful preparation or special treatment such as high pressure water rinsing) routinely reach fields of 100 MV/m without field emission or breakdown; however, the effective cathode area exposed to high fields in these experiments was only about 1 mm^2.

Processing field emitters, increasing breakdown voltage

Processing (conditioning) electrodes and cavities is the most widely used technique for reducing field emission (dark current) and bolstering resistance to breakdown.

We have investigated the early processing regime and found that the first few breakdown events decrease field emission and raise the subsequent breakdown voltage with high probability.[1]

Starbursts

With a scanning Auger microprobe, we have determined that the starbursts created by arcing are regions where several surface layers have been removed; starbursts indicate where ions bombarded the surface. In particular, starbursts contain less surface carbon contamination than surrounding areas; the effect of surface carbon on the secondary electron emission coefficient explains the visibility of starbursts in a scanning electron microscope.

Craters

Central craters in starbursts most often occur at sites of contaminant particles. Often smaller "satellite" craters appear scattered within the starburst; when other particles existed near the particle which caused the central crater, satellite craters were often found at those particles' sites. However, many satellite craters appeared where no obvious particle had been; we were unable, even with a scanning Auger microprobe, to detect foreign material in most of these satellite craters; note, however, that "contaminants" made of the substrate material, oxygen, or carbon, could not be distinguished from the surrounding surface, which always contained those components.

[1]Too much surface damage (from too many breakdown events) undermined efforts to connect surface damage with specific breakdown events, so we rarely allowed more than a few breakdown events per cathode spot, and can draw no conclusions about the effects of further processing beyond the first few breakdown events.

Effect of substrate on starbursts and craters

Although contaminant particles, and not the substrate, determine the field emission and breakdown characteristics, the substrate heavily influences starburst and crater formation. For niobium and copper cathodes with very thick oxides, starbursts have a very different appearance, and none of the typical cratering common on natural oxides (although the entire starburst resembles a large, flat crater). Thick oxides seem to inhibit the spread of arc plasma, limiting starburst size. Thin layers of surface contaminants are more susceptible to widespread satellite cratering; for instance, on gold and on some copper cathodes, most starbursts were covered with small satellite craters.

8.2 Modeling (DC and RF) Voltage Breakdown

Cathode initiated breakdown: field emission plus neutral vapor

Using particle-in-cell plasma simulations, we explored the initiation of breakdown at a strong field emitter upon the release of neutral gas atoms from the emitter. RF and DC simulations progressed almost identically once a cloud of ions formed in front of the emitter, creating locally a field that overwhelmed the applied fields.

We have developed a simple analytical model that predicts a critical product of field emission current and local gas density at the threshold of breakdown. Because it involves the field emission current, this prediction (by design) achieves relative insensitivity to the dependence of field emission on the surface electric field, although it still requires the input of a model for field emission. The computer simulations completed so far agree with this estimate of the critical product of current and gas density.

At this time the analytical model cannot be easily confirmed by experiment because the source of the neutral gas has not been explained (and the analytical model simply assumes a certain gas density). However, the model can be very useful for investigating the plausibility of different sources for neutral gas. Using the analytical estimate, we show that helium processing of superconducting RF cavities may be explained by the enhancement of gas density at a field-enhancing microprotrusion due to the dielectric polarizability of helium.

8.3 In the Future

Simulations

The computer simulations presented in this work are just a beginning; most important, they show the promise of OOPICPRO to include more realistic aspects of breakdown in its simulations. For example, adding Coulomb collisions between electrons and individual ions would be an easy next step. Further exploration of the evolution of breakdown (perhaps on faster computers) is needed.

Experiments: a return to RF

Having begun to study DC breakdown intending to learn about RF breakdown, it's now time to confirm our results in RF experiments (or discover new differences). To that end, we have designed a new superconducting cavity, described in appendix F, ideal for testing interactions between the electric field and the cavity surface.

Appendix A

Gas Desorption and Pumping Speed

Occasionally field emission caused an increase in the background vacuum pressure (see section 4.1.1), as measured by the ion pump that maintained the apparatus vacuum.

From the pumping speed of the ion pump, and the measured pressure rise, we can calculate the amount of gas released during field emission. The pumping speed, which I'll label $\dot{V}$, depends generally on the pressure but is nominally about $\dot{V} = 20$ L/s (for air and water) at pressures between 10^{-8} and 10^{-6} torr. Saturation effects for N_2, O_2, CO, and CO_2 can reduce the pumping speed by 50% (in this pressure range). The pumping speed for hydrogen, a common outgassing product, is lower.

Given the pump speed and the pressure $p(t)$ (of a single species of gas) as a function of time, the number of molecules removed from the system between time

t_1 and t_2 is

$$N_{\text{pump}} = \int_{t_1}^{t_2} dt \frac{p(t)\dot{V}(p(t))}{kT}. \tag{A.1}$$

That is, the rate at which the pump removes molecules is

$$\dot{N}_{\text{pump}} = \frac{p\dot{V}}{kT} \tag{A.2}$$

(where we use the ideal gas law, $p = nkT$, to convert from pressure to number density, and $\dot{N}_{\text{pump}} = n\dot{V}$). Note that the pressure is the pressure at the ion pump.

From the strong correlation between field emission and pressure, we can safely conclude that the gas comes from the electrodes—but from which one? Ignoring for a moment which electrode outgases most, suppose that neutral vapor desorbs from area A with a (number) current density of J molecules per area per time. In steady state, the outgassing is equal to the pumping:

$$JA = \dot{N}_{\text{pump}} = \frac{p\dot{V}}{kT}. \tag{A.3}$$

To grasp how large or small J is, it will help to write it as the number of monolayers per second; assuming a monolayer to have about 10 atoms/nm^2 (like a square lattice with spacing $\sqrt{10}$ Å), the rate of monolayer desorption (from area A) is

$$M = \frac{J}{10\ \text{nm}^{-2}} \tag{A.4}$$

Taking $\dot{V} = 10$ L/s, a conservative estimate for air, we estimate the desorption rate corresponding to a pressure increase of 1×10^{-8} torr assuming different areas A in table 4.1, reproduced here as table A.1.

In section 4.1.1 we concluded that bulk material was unlikely to be the cause of this pressure rise, because it would be pumped by the chamber walls. If only a few monolayers are desorbed, it must occur over a wide region (10 mm^2) to produce the observed increases in pressure, suggesting that the anode is the source of gas.

Table A.1: Outgassing rates, in monolayers per second, for different areas, corresponding to the pumping rate of nitrogen (or, roughly, water, oxygen, carbon monoxide, or carbon dioxide) for a 1×10^{-8} torr pressure rise caused by field emission. A pumping speed of 10 L/s is assumed.

Outgassing Area	Monolayers (N_2, H_2O, O_2, CO, CO_2)/s Outgassed
1 μm^2	3×10^5
10^4 μm^2	30
1 mm^2	0.3
10 mm^2	0.03

The energy required to desorb the gas also favors the anode. Typical currents and voltages associated with pressure rises might be about 10 μA and 10 kV (important: such currents and voltages have also been observed without measurable pressure rises). The maximum power input to the anode is 100 mW ($= 6 \times 10^{17}$ eV/s), while the power input to the cathode is only a tiny fraction of that.

Electrons bombarding the anode penetrate the surface to a depth of a few microns. Making the approximation that electrons deposit energy uniformly throughout a 5 micron depth, then about 6×10^{17} eV/s $\cdot$ 1 nm/5 μm $= 1 \times 10^{14}$ eV/s goes into the first three monolayers (about 1 nm). Under the previous assumptions, about 3×10^{12} atoms (or molecules) are desorbed every second (per 1×10^{-8} torr pressure rise), allowing more than the several eV per desorbed particle that would probably be needed. Of course, the input power will not be converted into desorption with 100% efficiency.

Comparing chamber pressure with pump pressure

The vacuum pressure that we measure is the pressure at the ion pump, which is separated by some vacuum plumbing from the chamber where the electrodes are. The plumbing has a finite conductance, which means that the pressure at the electrodes is somewhat greater than the pressure at the ion pump. From the pump speed, nominally 20 L/s, and an approximation of the conductance of the intervening pipe, the pressure at the electrodes would be roughly between 1.4 and 3 times the ion pump pressure, rather a small difference compared to other errors in estimating the pressure.

Appendix B

Electron Trajectories from a Field Emitter

To calculate the power per area deposited at the anode, we need to know the radius of the electron beam at the anode. The factors determining the beam radius are: initial emission radius r_e, initial transverse velocity $v_{\perp,i}$ at emission, the transverse velocity gained from transverse field components due to geometry (presumably only near the emitter), and the transverse velocity gained from the electric field of the beam itself (space-charge).

Initial (thermal) transverse velocity

Initial transverse velocity at emission is due to the random "transverse" thermal energy of the electrons before emission, hence

$$\begin{aligned} v_{\perp,i} \sim \sqrt{\frac{kT}{m_e}} &= 7 \times 10^4 \text{ m/s at 300 K} \\ &= 2 \times 10^5 \text{ m/s at 3000 K.} \end{aligned} \tag{B.1}$$

Transverse velocity gain due to transverse fields near emitter

If the emitter has a field enhancement β due to its shape (its pointiness), there will be a radial component to the field near the emitter, although the applied field has zero radial component almost everywhere else. Consider a protrusion on the cathode with field enhancement $\beta \gg 1$, and tip radius approximately r_t. The field will be enhanced within a distance $\sim r_t$ of the emitter. Imagining cylindrical coordinates with the emitter on axis, and the gap in the z-direction,

$$0 = \nabla \cdot \mathbf{E} = \frac{\partial E_z}{\partial z} + \frac{1}{r}\frac{\partial (r E_r)}{\partial r}. \tag{B.2}$$

This equation reflects the fact that the "net field flux" through any volume element is zero, and shows why field enhancement, requiring change in E_z, also requires non-zero E_r; its integral form is more helpful in this case, where we can exploit symmetry. Consider now a cylinder of radius Δr and height Δz on axis, with one end at the emitter, and the other end out of the field-enhanced region; the flux through the ends is approximately

$$\pi \Delta r^2 (E - \beta E) \approx -\pi \Delta r^2 \beta E \tag{B.3}$$

where E is the macroscopic (unenhanced) gap field, and the flux through the sides is approximately

$$2\pi \Delta r \Delta z E_r. \tag{B.4}$$

We cannot, of course, justify approximating the fields as constant along the entire cylinder end or wall, but the order of magnitude should be right. With no charge present, the net flux must be zero, and

$$E_r \sim \frac{1}{2}\frac{\Delta r}{\Delta z}\beta E \tag{B.5}$$

is a sort of average radial field in the enhanced region at radius Δr. We substitute $\Delta z \approx r_t$, the size of the field-enhanced region, and $\Delta r = r_e$, the emission radius, probably not too much smaller than r_t.

The transverse field is small compared to the longitudinal field, so only the longitudinal component of the field need be considered in estimating the time an electron spends in the field-enhanced region:[1]

$$t_e \sim \sqrt{\frac{2mr_t}{q(\beta E/2)}} \tag{B.6}$$

and the transverse velocity gained by an electron in that time is,

$$v_\perp \sim \frac{q}{m} E_r t_e \sim \frac{q}{m}\frac{1}{2}\frac{r_e}{r_t}\beta E\sqrt{\frac{4mr_t}{q\beta E}} = r_e\sqrt{\frac{q\beta E}{mr_t}}. \tag{B.7}$$

Assuming the field-enhanced region is small compared to the gap spacing d, the gap-crossing time is:

$$t_g \approx \sqrt{\frac{2md}{qE}}, \tag{B.8}$$

so neglecting space charge effects, the transverse velocity will be constant outside the field-enhanced region, and the beam radius a distance d from the emitter will be

$$r_b(d) \sim r_e + v_\perp t_g \sim r_e\left[1 + \sqrt{\frac{2\beta d}{r_t}}\right]. \tag{B.9}$$

Charbonnier *et. al.* [20][2] state a result, for the case of $\beta \gg 3$ and radius determined by "electrons emitted 90 deg from the axis of the hemisphere"

$$r_b = 2\sqrt{\beta r_t d}\,, \tag{B.10}$$

[1]This assumes that the enhanced region is of size $\sim r_t$ and initial velocity is relatively small, so that it takes time t_e to travel a distance $r_t \sim \frac{1}{2}\frac{q}{m}(\beta E/2)t_e^2$ to get out of the enhanced region, making the coarse approximation that the electron sees an "average field" $\beta E/2$ throughout the enhanced region.

[2]This result is simply stated in the referenced article, which cites a report that I have not found: F.E. Vibrans, Lincoln Lab. Tech. Report No. 353, 8 May 1964.

which differs from the above result (setting $r_e = r_t$, and assuming $r_b \gg r_e$) by $\sqrt{2}$, probably unimportant considering other approximations.

There is another way to consider this problem [14], considering the equipotential lines around a protrusion in a parallel-plane geometry. In the unenhanced field region, the equipotentials are parallel to the plane; there is no radial electric field. We may imagine an electron traveling radially outward from the protrusion until it leaves the field-enhanced region; if the height of the emitter is h, then the particle will be a distance h from one conducting plane, and since it is just out of the field-enhanced region, it must be in the constant field region, and so must be at a potential $\Phi = Eh$ from the conducting plane. Of course, in reaching that potential, it gained some velocity in the gap direction as well as the transverse direction, but an upper limit to its transverse velocity would be

$$v_{\perp,\max} \sim \sqrt{\frac{2qEh}{m}}. \tag{B.11}$$

If the protrusion has $\beta \sim h/r_t$, then we have $v_{\perp,\max} \sim \sqrt{2q\beta E r_t/m}$, in rough agreement with the preceding.

Space charge effects

Electrons will gain some transverse velocity due to their mutual repulsion. Unlike the other effects involving initial transverse velocities, space charge effects increase with beam current. Perilously ignoring the change in electron density as the electrons speed up under the influence of the electric field, we estimate that the beam radius increases from a_0 by a factor of 2 over a distance z: (see [64])

$$z = a_0 \sqrt{\frac{4\pi\epsilon_0 m_e v^3}{eI}} \int_0^{\sqrt{\ln 2}} e^{u^2} du = 0.674 \times a_0 \sqrt{\frac{4\pi\epsilon_0 m_e v^3}{eI}}, \tag{B.12}$$

where a_0 is the beam radius at $z = 0$, v is the velocity of the electrons (any change in which we have ignored), I is the total beam current, and e/m_e the charge to mass ratio of the electron.

For example, for a 10 μA beam of 100 eV electrons, the radius doubles as the beam travels $z \approx 80a_0$ (a distance 80 times the initial beam radius). For small beams and large currents, space charge effects will overwhelm the initial transverse velocities.

Trajectory simulations without space charge

Limited (in the geometry they survey) simulations of electron trajectories from protrusions (done for β up to about 20) indicate that the tangential velocity imparted due to the nature of the field-enhanced region can be represented approximately as

$$v_{\perp,\max} \sim \sqrt{2} r_e \sqrt{\frac{q\beta E}{m r_t}}. \tag{B.13}$$

I chose the prefactor to be $\sqrt{2}$ because it comes close to the simulated result, and it agrees with the Charbonnier *et. al.* result (shown above) if $r_e = r_t$. However, the simulations show that r_e is not r_t, though perhaps a half of r_t; however, in the limited geometries explored by our simulations, the Charbonnier result, with the beam radius increasing as $2\sqrt{\beta r_t d}$, does not describe the trajectories well. However, allowing the emission radius to change independently of the tip radius is reasonably accurate. That is, as β changes, so does the ratio r_e/r_t. Of course the better fit may simply be a result of having one more parameter (r_e/r_t), which must be measured, thus making the result not quite so general.

Appendix C

The Electric Field on a Rippled Conducting Surface

The electric field in the empty half space bounded by a conducting plane at $x = 0$ is simply a constant electric field perpendicular to the plane, $\mathbf{E} = E_\infty \hat{\mathbf{x}}$. The corresponding potential is $\phi = -E_\infty x$, satisfying $\phi(x{=}0) = 0$, as well as $-\nabla^2\phi = 0$. If the conducting surface became warped, the electric field would be enhanced at "hills" and diminished at "valleys."

To find how the electric field changes when the conducting boundary deviates from a perfect plane to a shape $x = f(y)$ (instead of $x = 0$, which would be a perfect plane[1]), we have to find the electric potential ϕ in a region such that

1. $\phi = 0$ at a conducting surface described by $x = f(y)$; or: $\phi(f(y), y) = 0$.

2. $\nabla^2\phi = 0$ for $x > f(y)$.

3. $-\nabla\phi \to \hat{\mathbf{x}} E_\infty$ as $x \to +\infty$.

[1]For simplicity, we consider only surface variations with y, ignoring variations along the third dimension.

We consider specifically a surface rippled with wave vector k and amplitude a:

$$x = f(y) = a\cos ky. \tag{C.1}$$

Presumably there are some different approaches to solving (perturbatively) the problem of a boundary perturbation, depending on where the approximation is made—do we find an exactly harmonic potential which is approximately zero on the perturbed boundary, or do we find an approximately harmonic potential which is exactly zero on the boundary, or some combination thereof? In this case, I've chosen to find an exactly harmonic solution that only approximately satisfies the boundary conditions: note that there is a conductor shape for which this is an exact solution, but that shape is only approximately $f(y) = a\cos ky$.

The zeroth order approximation to the solution is simply the unperturbed solution, $\phi_0(x, y) = -E_\infty x$. The solution $\phi(x, y)$ is the sum of harmonic functions ϕ_0 and ψ_i,

$$\phi = \phi_0 + \epsilon\psi_1 + \epsilon^2\psi_2 + \cdots, \tag{C.2}$$

each of which satisfies

1. $\nabla^2\psi_i = 0$
2. $-\nabla\psi_i \to 0$ for $x \to +\infty$

such that the expansion of ϕ to order ϵ^n satisfies the boundary condition

$$\phi(\epsilon f(y), y) = 0 \tag{C.3}$$

to order ϵ^n. As usual in such perturbation theory, the parameter ϵ is for bookkeeping (order-keeping), and is set to unity in the end.

The approximation of ϕ accurate up to terms of order $O(\epsilon^{n+1})$ will be written

$$\phi_n := \phi_0 + \epsilon\psi_1 + \epsilon^2\psi_2 + \cdots + \epsilon^n\psi_n. \tag{C.4}$$

We demand (besides the conditions of harmonicity and behavior at infinity imposed with the ψ_i) that ϕ_n be zero on the boundary $x = \epsilon f(y)$ up to order $O(\epsilon^{n+1})$:

$$\phi_n(\epsilon f(y), y) = 0 + O(\epsilon^{n+1}). \tag{C.5}$$

Expanding ϕ_n near the boundary in a power series,

$$\phi_n(\epsilon f(y), y) = \sum_{m=0}^{\infty} \frac{(\epsilon f(y))^m}{m!} \frac{\partial^m \phi_n}{\partial x^m}(0, y), \tag{C.6}$$

equation C.5 becomes:

$$\sum_{m=0}^{n} \frac{(\epsilon f(y))^m}{m!} \frac{\partial^m \phi_n}{\partial x^m}(0, y) = 0 + O(\epsilon^{n+1}). \tag{C.7}$$

Writing ϕ_n in terms of the ψ_i and keeping resulting terms in the above equation up to order ϵ^n, we get a differential equation for ψ_n. The case $n = 0$ simply leads to $\phi_0(0, y) = 0$—applying the condition of harmonicity and the right boundary condition, $\phi_0 = -E_\infty x$, the solution for the conducting plane.

For $n = 1$ (the ϵ^0 order terms cancel, as lower order terms ought):

$$\epsilon \psi_1(0, y) + \epsilon f(y) \frac{\partial \phi_0}{\partial x}(0, y) = 0 + O(\epsilon^2). \tag{C.8}$$

To find ψ_1 we have to find the solution of $\nabla^2 \psi_1 = 0$ with the boundary conditions $\nabla \psi_1 \to 0$ at $x \to +\infty$ and

$$\psi_1(0, y) = -f(y) \frac{\partial \phi_0}{\partial x}. \tag{C.9}$$

For $f(y) = a \cos(ky)$ the solution to first order is $\phi_1 = \phi_0 + \psi_1$, where $\phi_0 = -E_\infty x$ and

$$\psi_1(x, y) = a E_\infty \cos(ky) e^{-kx}. \tag{C.10}$$

The second order is (not writing the ϵ^0 and ϵ^1 terms, each of which vanishes because of our work above):

$$\epsilon^2 \psi_2(0, y) + \epsilon^2 f(y) \frac{\partial \psi_1}{\partial x}(0, y) + \frac{1}{2!} \epsilon^2 (f(y))^2 \frac{\partial^2 \phi_0}{\partial x^2}(0, y) = 0 + O(\epsilon^3). \tag{C.11}$$

Now we have to solve $\nabla^2\psi_2 = 0$ so that at $x \to \infty$ the gradient of ψ_2 vanishes at infinity, and on the left boundary,

$$\psi_2(0,y) = -\frac{1}{2!}(f(y))^2\frac{\partial^2\phi_0}{\partial x^2} - f(y)\frac{\partial\psi_1}{\partial x}(0,y). \tag{C.12}$$

For $f(y) = a\cos(kx)$,

$$\psi_2(0,y) = a^2kE_\infty\cos^2(ky) = \frac{1}{2}a^2kE_\infty[1+\cos(2ky)]. \tag{C.13}$$

The solution for ψ_2 (the harmonic function satisfying the above boundary condition) can be guessed with a little work:

$$\psi_2(x,y) = \frac{1}{2}a^2kE_\infty[1+\cos(2ky)e^{-2kx}]. \tag{C.14}$$

For higher order terms, (for compact notation use $\psi_0 = \phi_0$)

$$\psi_n(0,y) = -\sum_{m=1}^{n}\frac{1}{m!}(f(y))^m\frac{\partial^m\psi_{n-m}}{\partial x^m}(0,y). \tag{C.15}$$

For $f(y) = a\cos(ky)$ and $n = 3$,

$$\begin{aligned}\psi_3(0,y) &= a^3k^2E_\infty\left[\cos(ky)\cos(2ky) - \frac{1}{2}\cos^3(ky)\right] \\ &= a^3k^2E_\infty\left[\frac{3}{2}\cos^3(ky) - \cos(ky)\right] \\ &= \frac{1}{8}a^3k^2E_\infty\left[3\cos(3ky) + \cos(ky)\right],\end{aligned} \tag{C.16}$$

which has the solution:

$$\psi_3(x,y) = \frac{1}{8}a^3k^2E_\infty\left[\cos(ky)e^{-kx} + 3\cos(3ky)e^{-3kx}\right]. \tag{C.17}$$

The full solution (up to third order) is:

$$\begin{aligned}\phi(x,y) &= -E_\infty x + aE_\infty\cos(ky)e^{-kx} + \frac{1}{2}a^2kE_\infty[1+\cos(2ky)e^{-2kx}] \\ &\quad + \frac{1}{8}a^3k^2E_\infty\left[\cos(ky)e^{-kx} + 3\cos(3ky)e^{-3kx}\right] + \cdots.\end{aligned} \tag{C.18}$$

Appendix D

Magnetic Field Produced by Discharge Current

To estimate the effect of the magnetic field produced by the discharge current during voltage breakdown, we need to estimate the magnetic field, given experimentally measured discharge current. We assume here that the discharge current runs down a thin column in the middle of a parallel-plate capacitor with circular plates (to preserve the cylindrical symmetry of the problem).

The magnetic field of an infinite wire with current I decreases with distance r from wire:

$$B = \frac{\mu_0 I}{2\pi r} \tag{D.1}$$

where μ_0 is the permeability of space, $4\pi \times 10^{-7}$ T·m/A.

That's a good first approximation, especially close to (but not inside) the column of current and far from the capacitor plates. The field produced by the discharge is different for two main reasons: the discharge column is not infinite; and the voltage across the capacitor drops during the discharge, creating a chang-

ing electric field, which induces a magnetic field. Inside the discharge column, we need to add another correction, which we will worry about later.

We need to solve the problem of charge flowing from capacitor plates down the center discharge column, taking into account the finite length of the discharge column and the effect of current in the capacitor plates and the changing electric field in the capacitor. The problem looks daunting at first, but we are saved from much difficulty by the principle of superposition: the field due to current flow within the finite region of the capacitor is equivalent to the field due to an infinite wire minus the field due to current outside the region of the capacitor (figure D.1). The latter field is just the field around a capacitor being charged by current I.

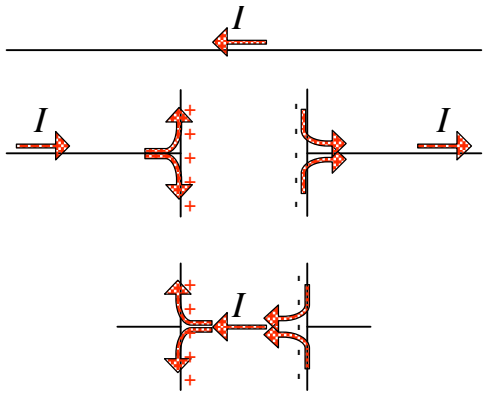


Figure D.1: The current of a capacitor being discharged by a line of current between the center of the two plates (bottom) is equivalent to the superposition of an infinite-current carrying wire (top) and a charging capacitor (middle). The plus and minus charges represent the rate of change in the charge on the plates (not the amount of charge), which reminds us that the electric field in the capacitor is changing, hence there is displacement current in the capacitor.

Fields around a charging capacitor

Consider a circular parallel-plate capacitor of radius R much greater than gap L being charge by an infinite wire of current I; let the end plates be at $z = \pm L/2$. The electric field in the capacitor is nearly uniform, and so the surface charge

density $\sigma(r) = \pm\sigma$ on the plates is also uniform (and of opposite sign on opposite plates); at any instant in time, the electric field and surface charge are related by:

$$E_z(z,r) \approx \begin{cases} E_\sigma = \sigma/\epsilon_0 & |z| < L/2 \text{ and } r < R \quad \text{(inside capacitor)} \\ \\ 0 & |z| > L/2 \text{ or } r > R \quad \text{(outside capacitor)} \end{cases} \tag{D.2}$$

(the electric field has zero radial and azimuthal components).

As current I charges the capacitor, the surface charge density σ increases (uniformly):

$$\frac{\partial\sigma}{\partial t} = \frac{I}{\pi R^2}. \tag{D.3}$$

The magnetic field (which has zero radial and z components) induced by a changing electric field is $\nabla \times \mathbf{B} = \mu_0\epsilon_0\partial\mathbf{E}/\partial t$, or, within the capacitor:

$$\frac{1}{r}\frac{\partial}{\partial r}(rB_\phi) = \mu_0\frac{\partial\sigma}{\partial t} = \frac{\mu_0 I}{\pi R^2} \tag{D.4}$$

from which we see that inside the capacitor,

$$B_\phi = \frac{\mu_0 I}{2\pi R}\frac{r}{R}. \tag{D.5}$$

Integrating, we find that to B_ϕ could be added a term C/r for some constant C; however, there is no current inside the capacitor at $r = 0$, so B may not be singular there, and C must be zero.

Outside the capacitor, the cylindrical symmetry of the problem and Ampere's law (counting the displacement current inside the capacitor), show that

$$B_\phi = \frac{\mu_0 I}{2\pi r}. \tag{D.6}$$

Therefore the magnetic field is

$$B_\phi(z,r) \approx \frac{\mu_0 I}{2\pi r} \times \begin{cases} \dfrac{r^2}{R^2} & |z| < L/2 \text{ and } r < R \quad \text{(inside capacitor)} \\ \\ 1 & |z| > L/2 \text{ or } r > R \quad \text{(outside capacitor)} \end{cases} \tag{D.7}$$

(the approximation corresponds to the approximation in the electric field, where we neglected fringing fields at the edges of the capacitor plates).

Note that the surface current density necessary to increase the surface charge uniformly is:[1]

$$\pm\mathbf{K} = \pm\hat{\mathbf{r}}K = \pm\hat{\mathbf{r}}I\left(\frac{1}{2\pi r} - \frac{r}{2\pi R^2}\right) = \pm\hat{\mathbf{r}}\frac{I}{2\pi r}\left(1 - \frac{r^2}{R^2}\right). \tag{D.10}$$

The discontinuity in magnetic field at the plates is therefore ($\Delta\mathbf{B} = \mu_0\mathbf{K}\times\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is normal to the surface current and $\Delta\mathbf{B}$ is the field on the positive $\hat{\mathbf{n}}$ side minus the field on the negative $\hat{\mathbf{n}}$ side):

$$B_\phi(L/2+\delta, r) - B_\phi(L/2-\delta, r) = \begin{cases} \dfrac{\mu_0 I}{2\pi R}\left(1 - \dfrac{r^2}{R^2}\right) & r < R \\ \\ 0 & r > R \end{cases}, \tag{D.11}$$

which exactly agrees with equation D.7.

Fields around a discharging capacitor

A charged capacitor with an infinitely thin conductor running down the cylinder axis between the two plates, conducting a current I, is equivalent (see figure D.1) to an infinite conductor on axis, conducting a current I, minus a capacitor being charged with current I (and appropriate surface charge currents on the plates). The surface charge on the plates determines the electric field in a straightforward

[1]The surface current density satisfies $\mathbf{K} = \hat{\mathbf{r}}K(r)$, and

$$\frac{I}{\pi R^2} = \frac{\partial}{\partial t}\sigma = -\nabla\cdot\mathbf{K} = -\frac{1}{r}\frac{\partial}{\partial r}(rK) \tag{D.8}$$

and

$$(2\pi r K) \to I \quad \text{as} \quad r \to 0. \tag{D.9}$$

manner. The magnetic field is

$$B_\phi(z,r) \approx \frac{\mu_0 I}{2\pi r} \times \begin{cases} 1 - \dfrac{r^2}{R^2} & |z| < L/2 \text{ and } r < R \quad \text{(inside capacitor)} \\ \\ 0 & |z| > L/2 \text{ or } r > R \quad \text{(outside capacitor)} \end{cases} . \tag{D.12}$$

The magnetic field is in the same direction as if the discharge current were an infinite wire, but the field tapers off faster as r increases due to displacement current (as the electric field changes), or alternatively, due to surface current on the plates. Again, the approximation is due to our neglecting fringing fields at the capacitor edge (that's also why the field has a kink at $r = R$).

Fields around a finite discharge column

If the current-carrying column (wire or discharge or other) is not infinitely thin, but carries current I uniformly throughout radius R_I, then for an infinitely long column, the magnetic field is

$$B_\phi(z,r) \approx \frac{\mu_0 I}{2\pi r} \times \begin{cases} \dfrac{r^2}{R^2} & r < R_I \\ \\ 1 & r > R_I \end{cases} . \tag{D.13}$$

As in the previous section, we find the magnetic field in an isolated capacitor discharging its charge through a conducting column of radius R_I:

$$B_\phi(z,r) \approx \frac{\mu_0 I}{2\pi r} \times \begin{cases} \dfrac{r^2}{R_I^2} - \dfrac{r^2}{R^2} & |z| < L/2 \text{ and } r < R_I \quad \text{(inside capacitor)} \\ \\ 1 - \dfrac{r^2}{R^2} & |z| < L/2 \text{ and } R_I < r < R \quad \text{(inside capacitor)} \\ \\ 0 & |z| > L/2 \text{ or } r > R \quad \text{(outside capacitor)} \end{cases}$$

(D.14)

The electric field is simply

$$E_z(z,r) \approx \begin{cases} \sigma/\epsilon_0 & |z| < L/2 \text{ and } r < R \quad \text{(inside capacitor)} \\ \\ 0 & |z| > L/2 \text{ or } r > R \quad \text{(outside capacitor)} \end{cases} \tag{D.15}$$

where σ is the surface charge and $\partial\sigma/\partial t = I/(\pi R^2)$.

For a capacitor of infinite radius, the laws of physics conspire so that the magnetic field produced by a finite current-carrying column is the same as the magnetic field produced by an infinitely long column of current.

Appendix E

The Trajectory of a Dipole Near a Microprotrusion

The force of the electric field on a non-permanent dipole[1] $p = qd$ is $F_p = q(E_+ - E_-)$, where E_+ and E_- are the fields at the positive and negative ends of the dipole; for a dipole of length d (very small compared with the variation of the electric field),

$$\mathbf{F}_p = q\frac{\partial \mathbf{E}}{\partial d}d = (\mathbf{p}\cdot\nabla)\mathbf{E}. \tag{E.1}$$

The acceleration due to this force will be $a = F_p/m$, where m is the dipole's mass. The induced dipole moment is proportional to the electric field (by a factor α, the polarizability): $\mathbf{p} = \alpha\mathbf{E}$.

Near a microprotrusion in an external field, the local field E will be increased by an enhancement factor β. A typical atom enters the enhanced field region with energy $\sim kT$, or typical velocity $v_T \sim \sqrt{kT/m}$. If the enhanced region is of size d_β, then the atom spends roughly time $t \sim d_\beta/v_T$ in the enhanced region; also,

[1] An induced dipole (a non-permanent dipole) always aligns parallel to the electric field.

the force is approximately $F_p \sim \alpha(\beta E)^2/d_\beta$. During time t, the atom accelerates toward the emitter at $a = F_p/m$, changing the velocity by $\Delta v \sim at$ (of course, the velocity is a vector, but we're being very rough order-of-magnitude here):

$$\frac{\Delta v}{v_T} \sim \frac{\alpha(\beta E)^2}{m v_T^2} \sim \frac{\alpha(\beta E)^2}{kT} \tag{E.2}$$

(notice how d_β conveniently cancels out, so we don't actually have to guess how big the enhanced region is).

For helium gas at 4 K, $v_T \sim 90$ m/s, and $\alpha = (2.6 \times 10^{-30}\ \mathrm{m}^3)\epsilon_0$. For an emitter with $\beta = 30$ and a field of 50 MV/m, $\Delta v/v_T \sim 1$. Thus the enhanced field has a significant effect on the trajectory of an atom. We might also consider that an atom passing parallel to the surface will be accelerated toward the surface. In the time t that an atom with thermal velocity crosses the enhanced region, it will travel approximately a distance $(1/2)at^2$ toward the cathode. Comparing that distance with the size of the enhanced region,

$$\frac{\frac{1}{2}at^2}{d_\beta} \sim \frac{\alpha(\beta E)^2}{2mv_T^2} \sim \frac{\Delta v}{2v_T}. \tag{E.3}$$

For $\Delta v/v_T \gtrsim 2$, any atom passing within the enhanced region will actually hit the emitter (which, if it's at 4 K, will cool bombarding atoms, making them more likely to be trapped).

At room temperature (but the same gas and field configuration), however, $\Delta v/v_T \sim 0.01$; the enhanced region has little effect on the trajectories.

If atoms do bombard the emitter, they'll do so with kinetic energy $\Delta\epsilon$, which, if there is significant density enhancement, will be several times kT, too little to change anything (for instance, sputtering is far out of the question), even with the increased bombardment rate due to the field enhancement.

Appendix F

The New Mushroom Cavity

To see if results on voltage breakdown from DC experiments apply equally well to RF experiments, we have designed an RF cavity particularly suited to test voltage breakdown—the Mushroom Cavity Mark VI. Most important, the cavity limits the high electric-field region to a small area; the surface field elsewhere in the cavity is more than 16 times lower than the field at the pedestal on the axis of symmetry. Also important, the endplate of the cavity (with the pedestal) can be easily removed—and it's small enough to fit in an electron microscope chamber. Detachable parts necessitate a complicated joint in a superconducting cavity; because very small power dissipation can be important in a superconducting cavity, currents must flow with almost no resistance across the joint bridging separate parts. Making the fields low at the joint helps enormously; therefore we use a choke joint. Since this cavity is designed to operate in the TM_{020} mode, we couple the choke joint to the magnetic field minimum at the cavity wall, thus achieving a sort of double choke joint—the fields everywhere in the choke joint are 100 times less than typical cavity fields, and the field at the joint itself is 10^4 times less than cavity fields.

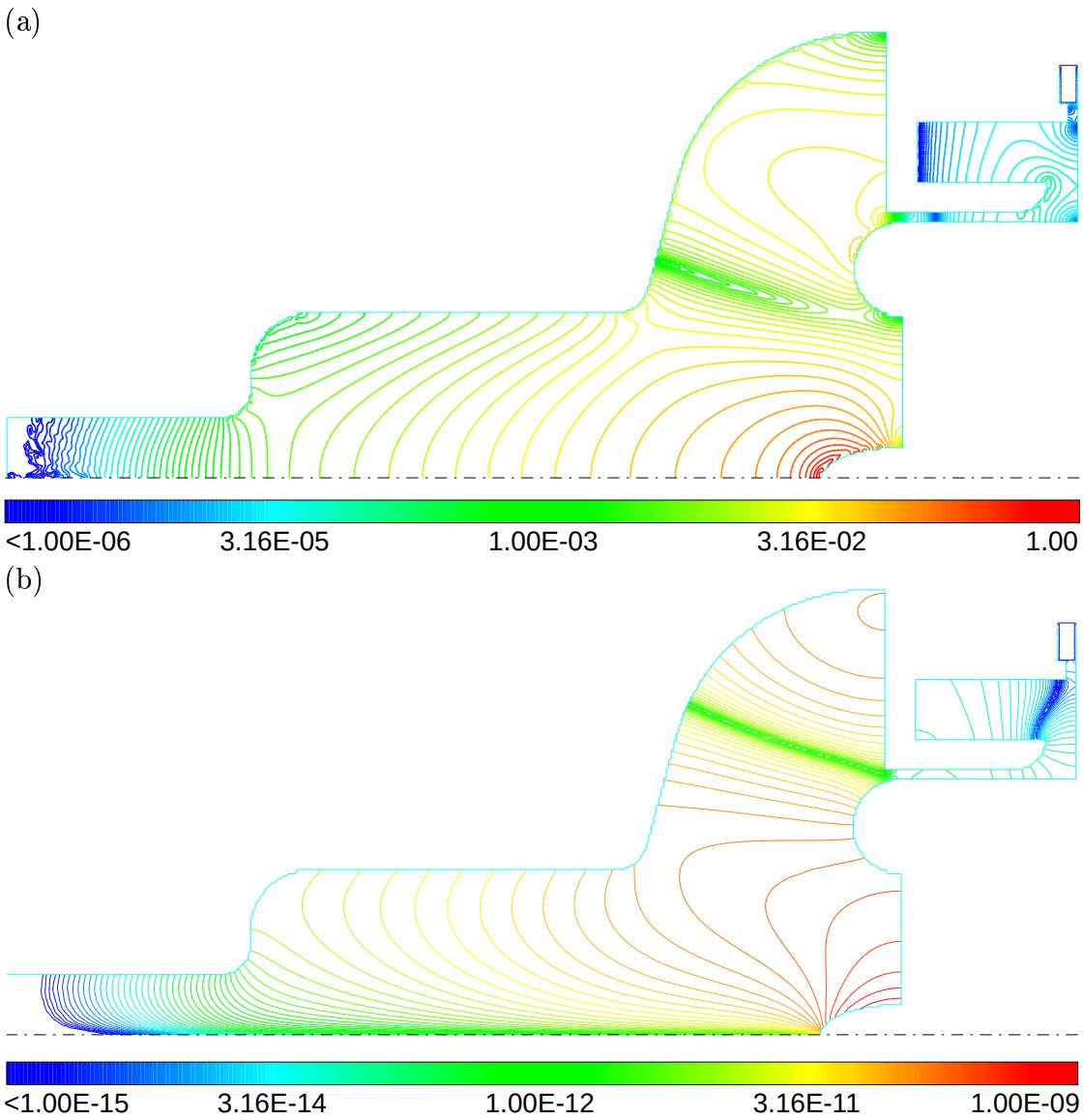


Figure F.1: Contour plots of the relative strengths of the electric (a) and magnetic (b) fields in the new mushroom cavity, Mark 6.1, in the TM_{020} mode. (The cylindrical axis of symmetry is at the bottom of each plot.) The removable endplate connects to the cavity cup at the corner of the choke joint where the magnetic field is smallest, in the rightmost, upper corner. Note that the entire choke region (which connects to the cavity cell by the thin channel) couples to the cavity at a low-magnetic field region, ensuring low fields everywhere in the choke region.

Appendix G

Gallery of Starbursts

The following gallery of starbursts depicts starbursts after breakdown on various cathodes. I have tried to walk the line between showing the typical and showing the interesting; however, "messy" breakdown events that involved a lot of surface melting have been disproportionately omitted (though figures G.4, G.37, G.38, and perhaps G.2 show such examples).

G.1 Starbursts on Niobium

Niobium cathodes were machined from bulk niobium with high RRR (250 or greater[1]) to clean them. Typical oxide thickness on niobium is about 50 Å.

[1]The residual resistivity ratio measures (indirectly) the purity of the niobium. High RRR niobium is very pure and is used for RF cavities to improve their thermal conductivity. The RRR probably has no effect on these DC experiments.

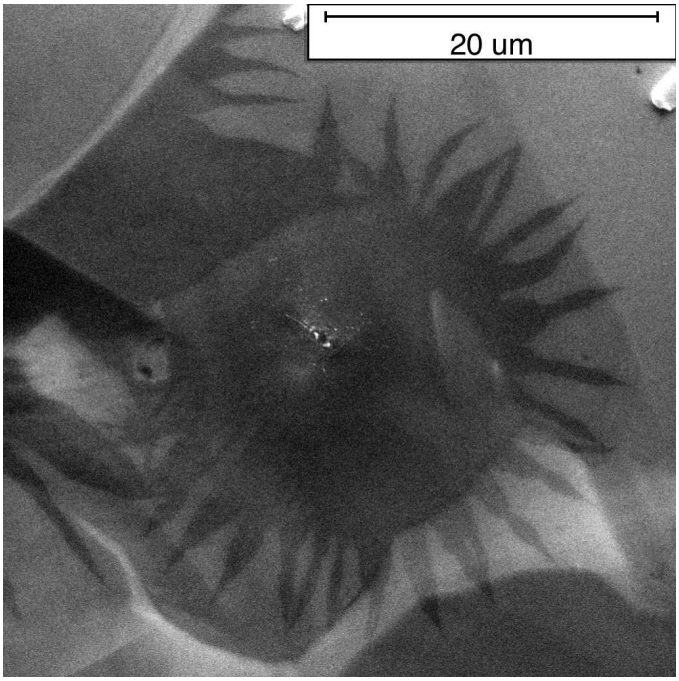


Figure G.1: B4-Ped3-SB2

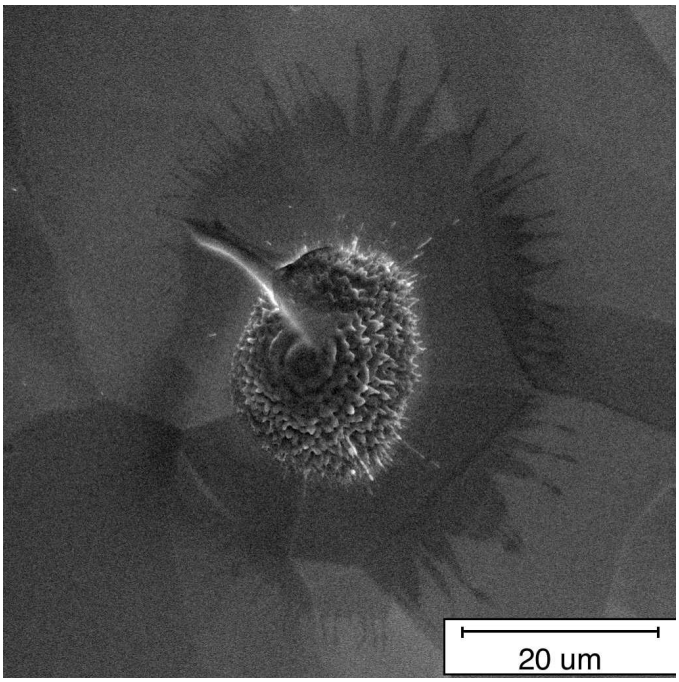


Figure G.2: B4-Ped4-SB4

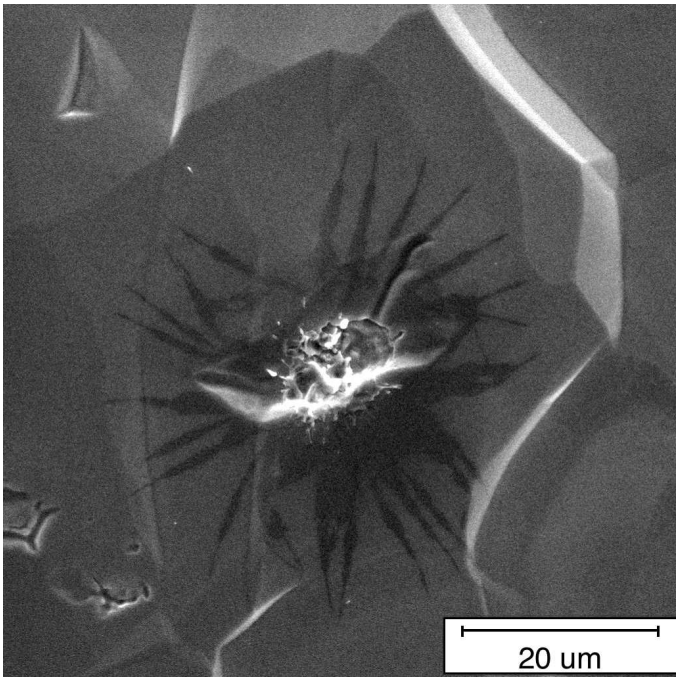


Figure G.3: B4-Ped4-SB5

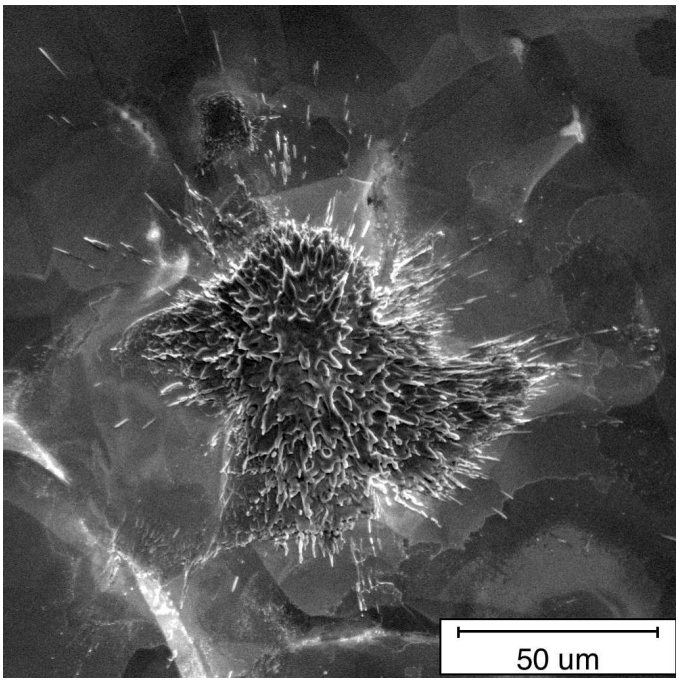


Figure G.4: B4-Ped5-SB2

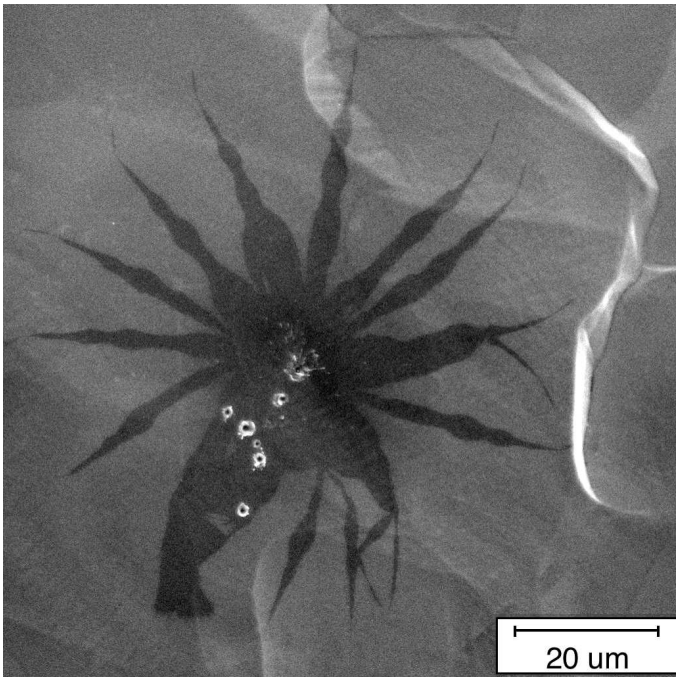


Figure G.5: B5-Ped3-SB3

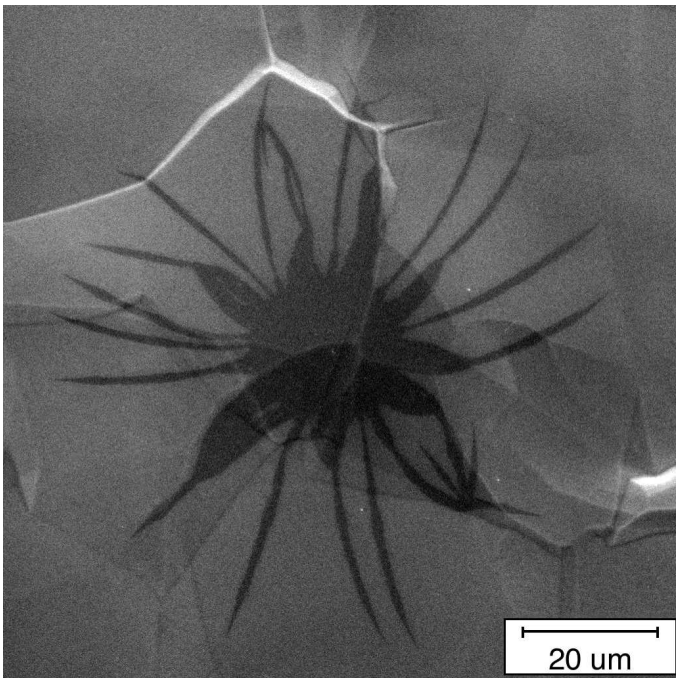


Figure G.6: B5-Ped3-SB4

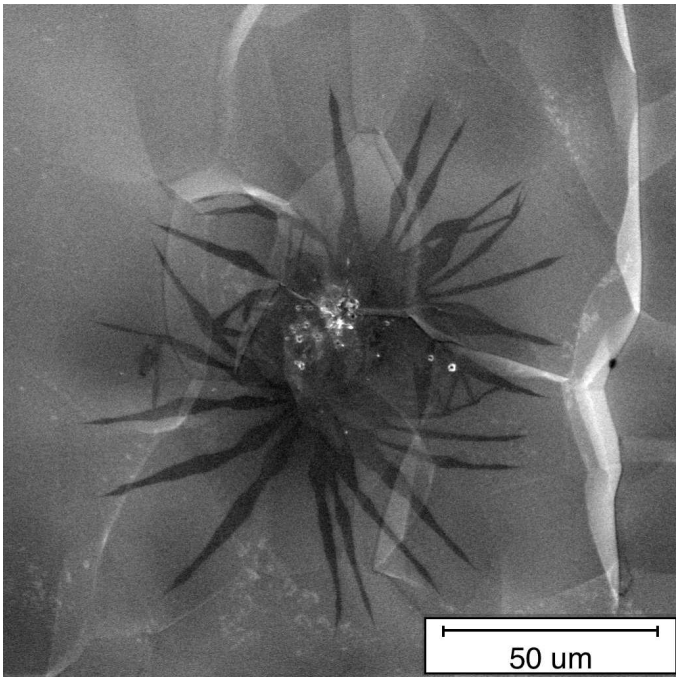


Figure G.7: B5-Ped5-SB4

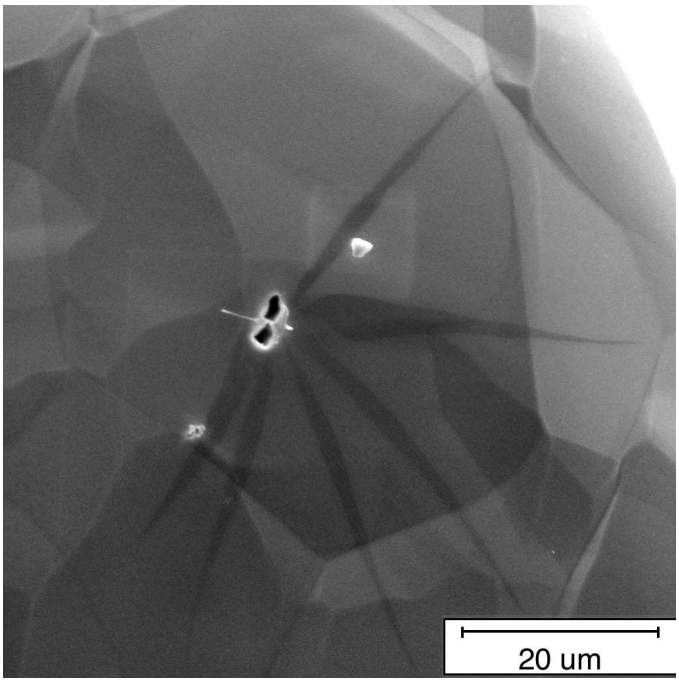


Figure G.8: B6-Ped3-SB1

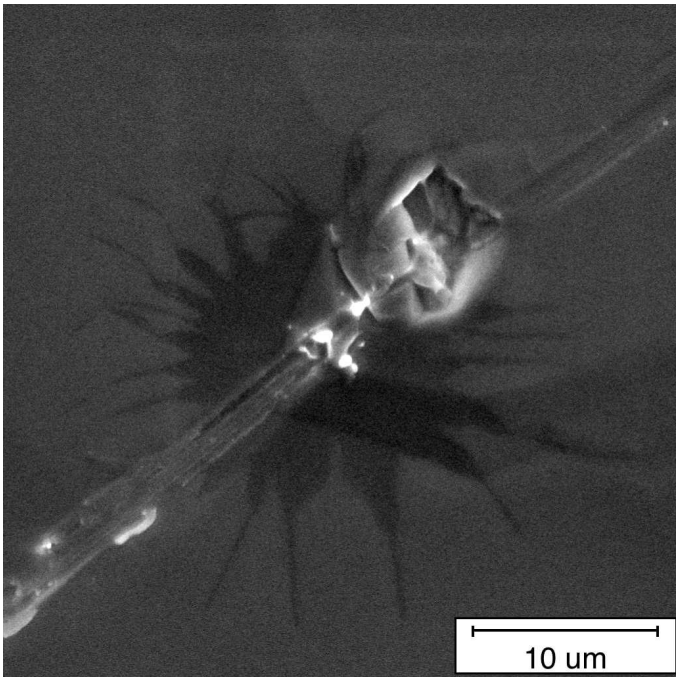


Figure G.9: B7-Ped3-SB6

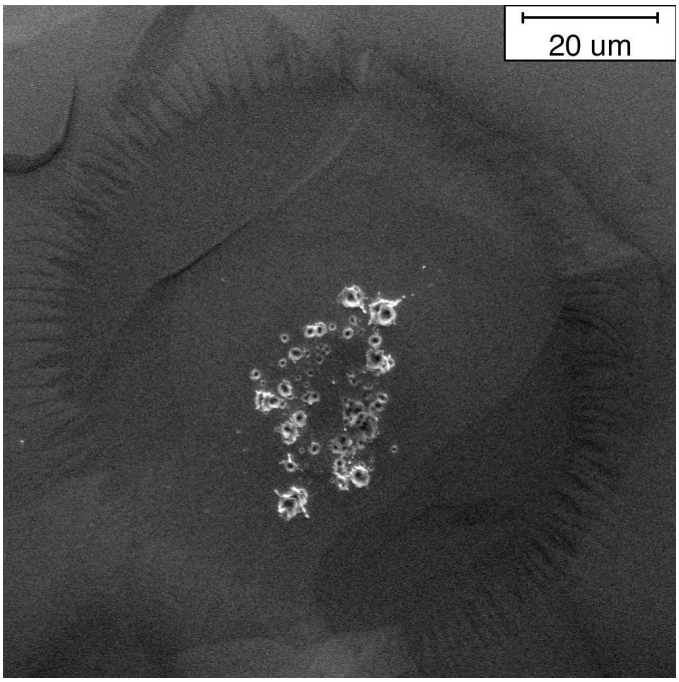


Figure G.10: B7-Ped4-SB4

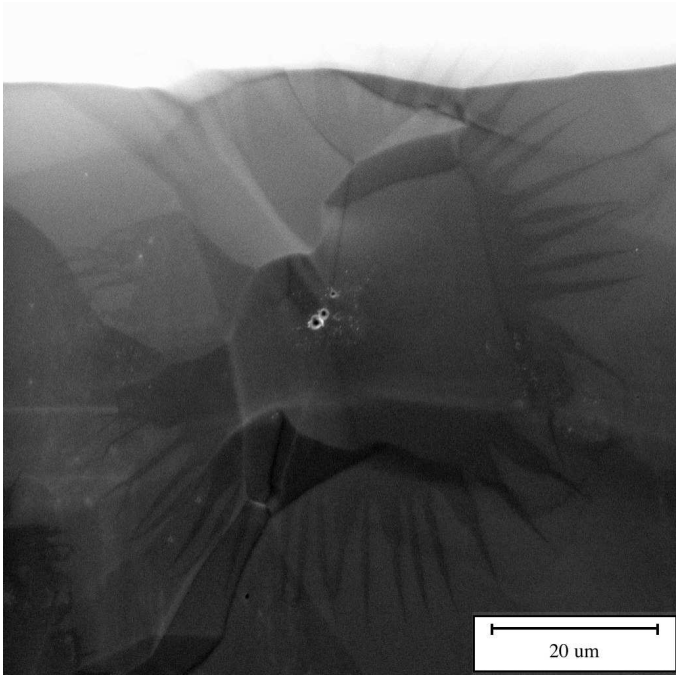


Figure G.11: BB-Ped2-SB4

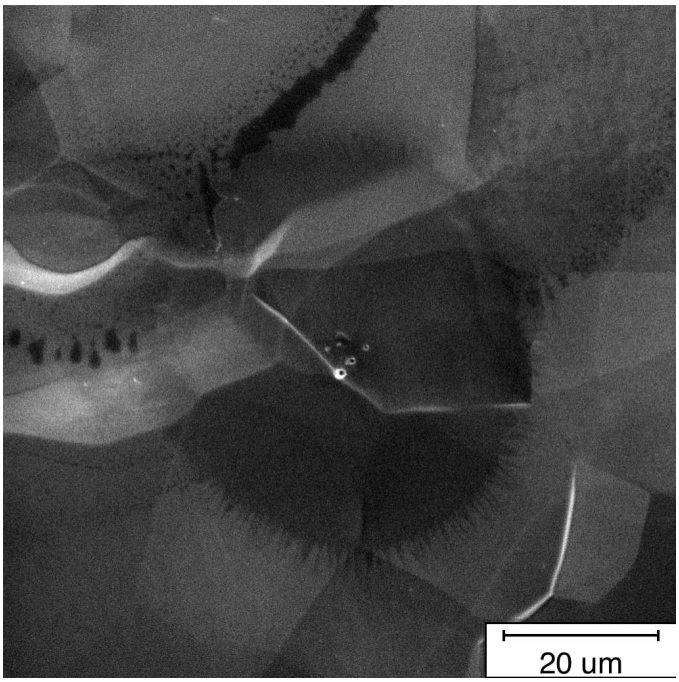


Figure G.12: BG-Ped1-SB2

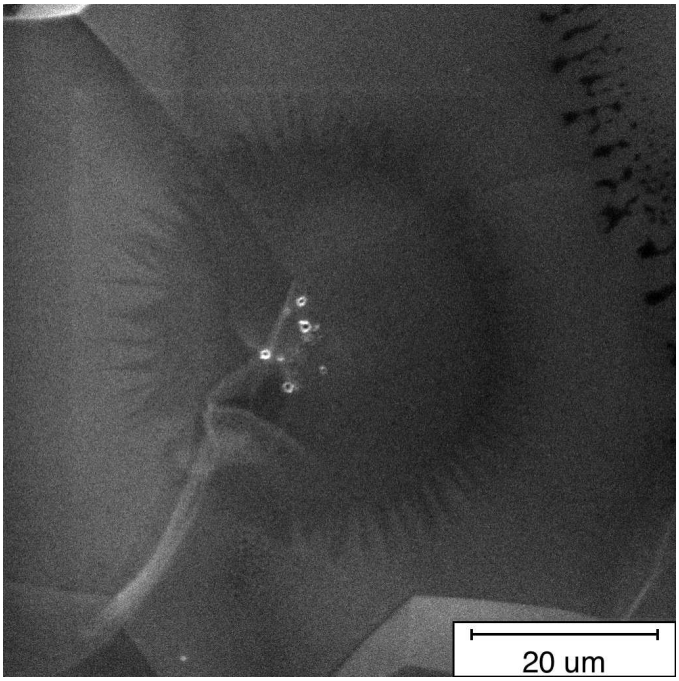


Figure G.13: BG-Ped1-SB6

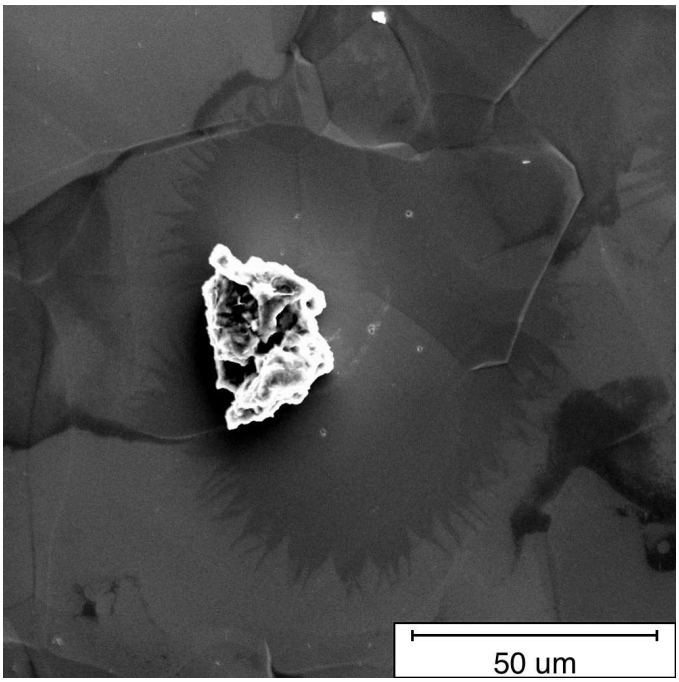


Figure G.14: BG-Ped3-SB1

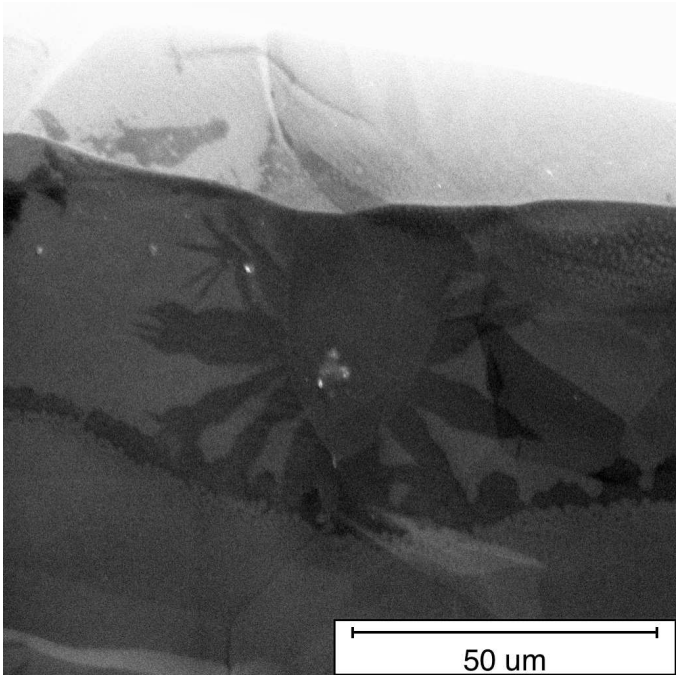


Figure G.15: BG-Ped5-SB3

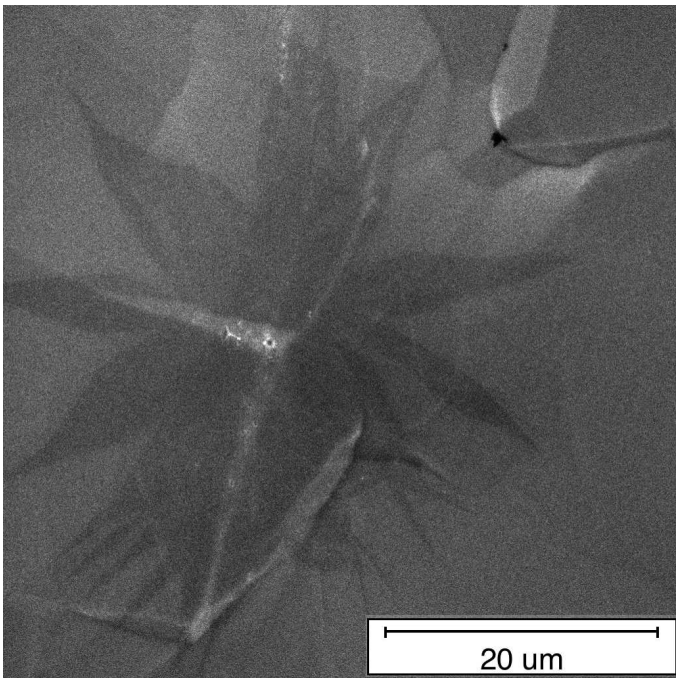


Figure G.16: D1-Ped3-SB18

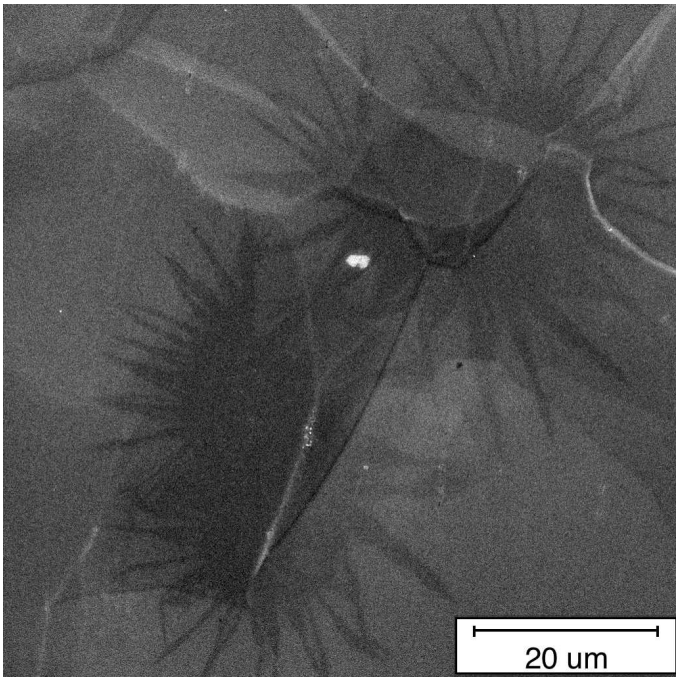


Figure G.17: D1-Ped3-SB20

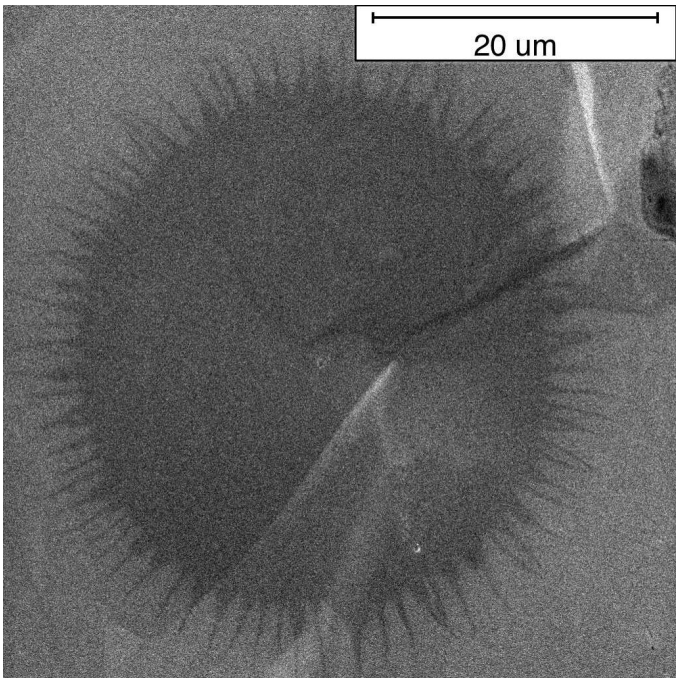


Figure G.18: D1-Ped3-SB5

G.2 Starbursts on Copper Film on Niobium Substrate

After preparing a niobium cathode, machining and etching as for any niobium cathode, copper was sputtered onto the niobium (after sputtering away the niobium oxide so the copper would stick well) to a thickness of nominally 1000 Å.

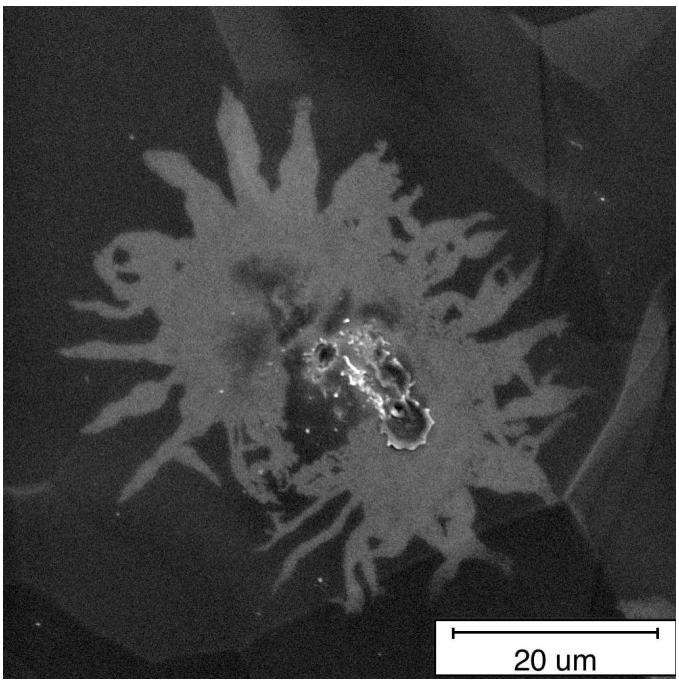


Figure G.19: BF-Ped5-SB20

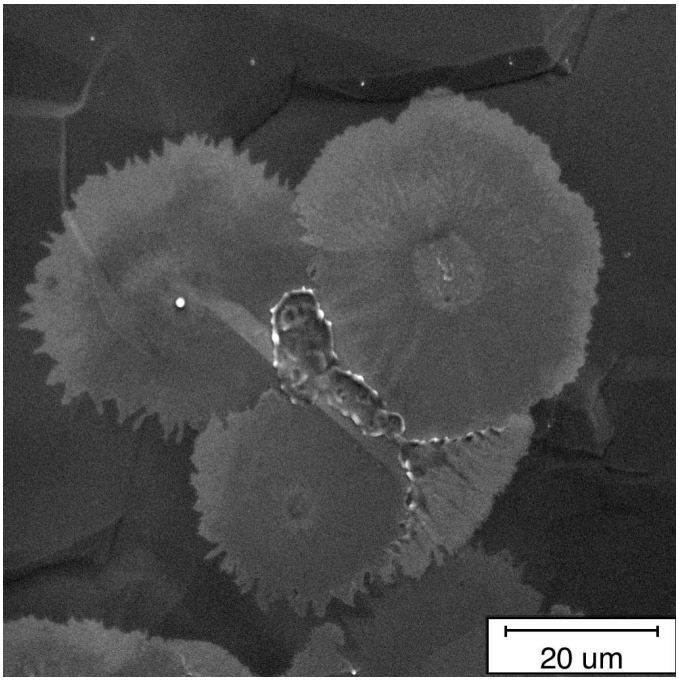


Figure G.20: BF-Ped5-SB22

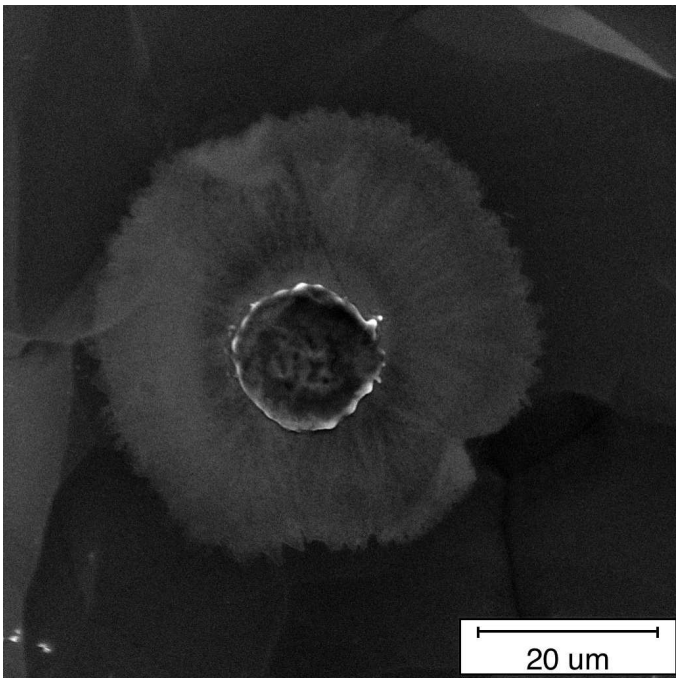


Figure G.21: BF-Ped5-SB34

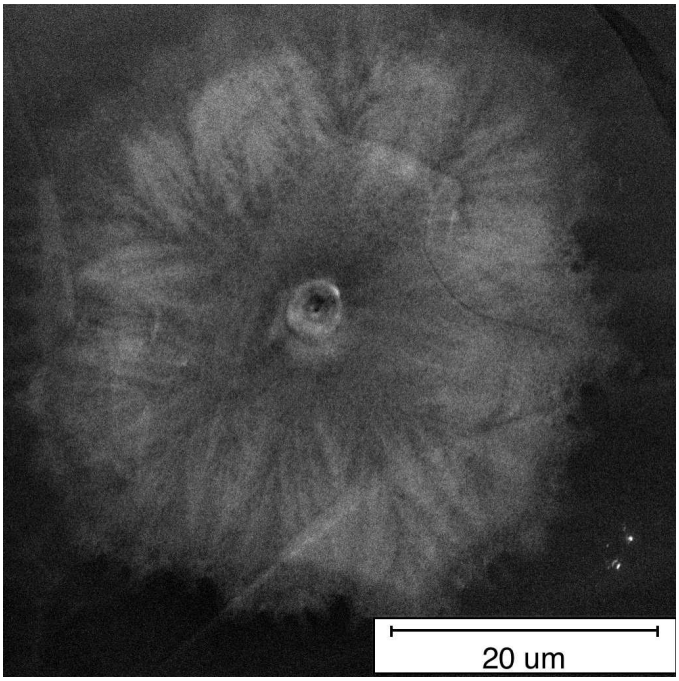


Figure G.22: BF-Ped5-SB8

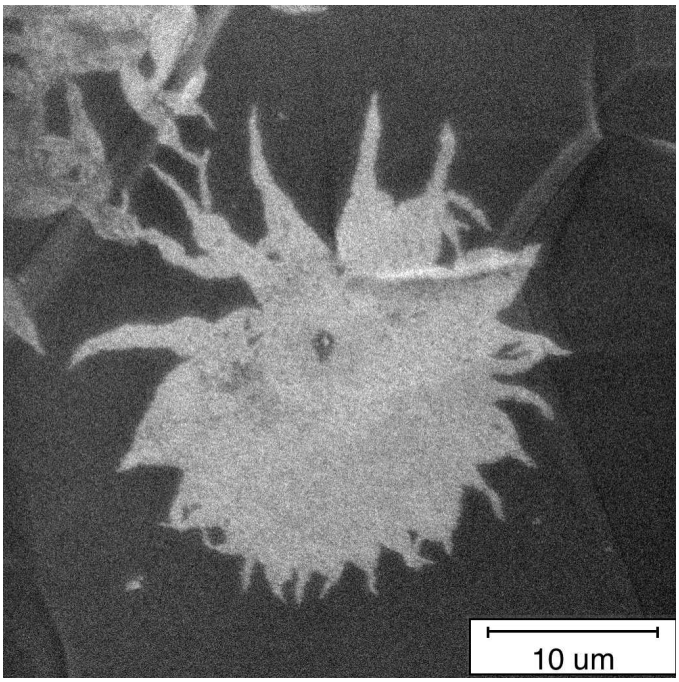


Figure G.23: BF-ped5-SB33

G.3 Starbursts on Gold Film on Niobium Substrate

As with the copper film on niobium, gold was sputtered to a thickness of nominally 1000 Å on a niobium cathode (again, after sputtering away the niobium oxide).

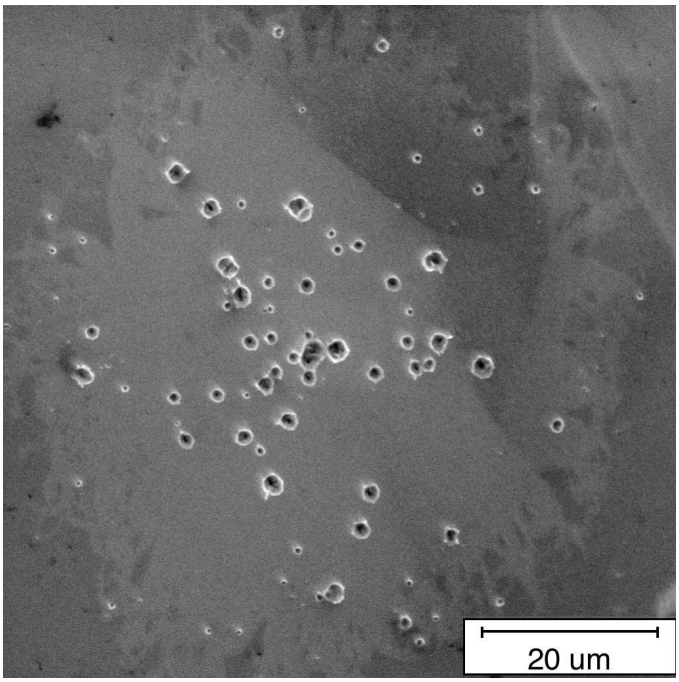


Figure G.24: BH-Ped3-SB2

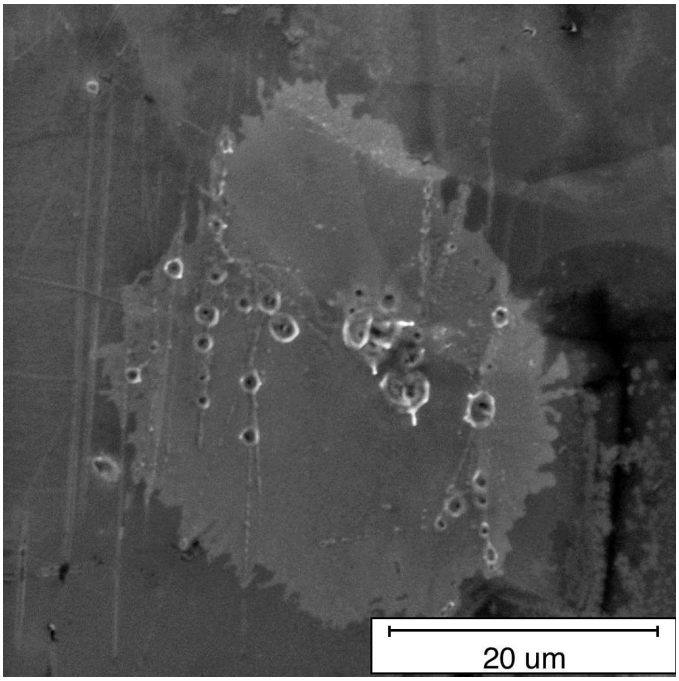


Figure G.25: BH-Ped4-SB1

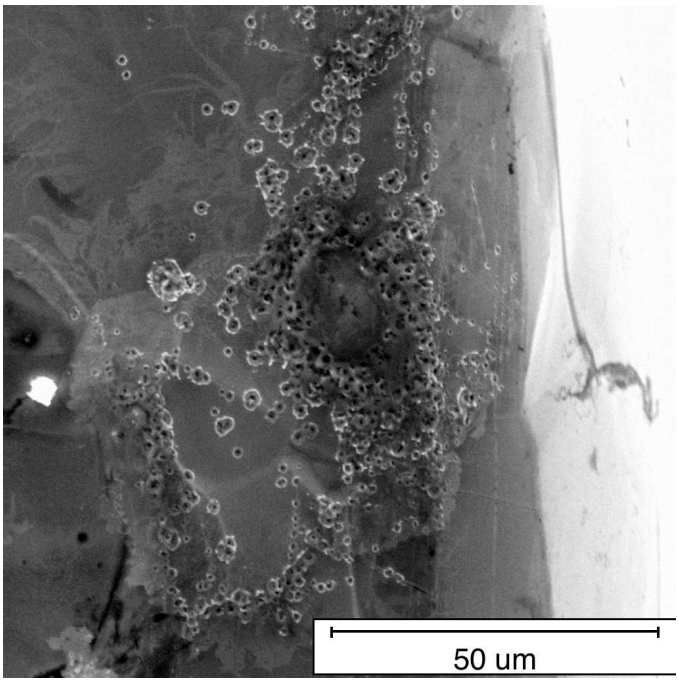


Figure G.26: BH-Ped4-SB10

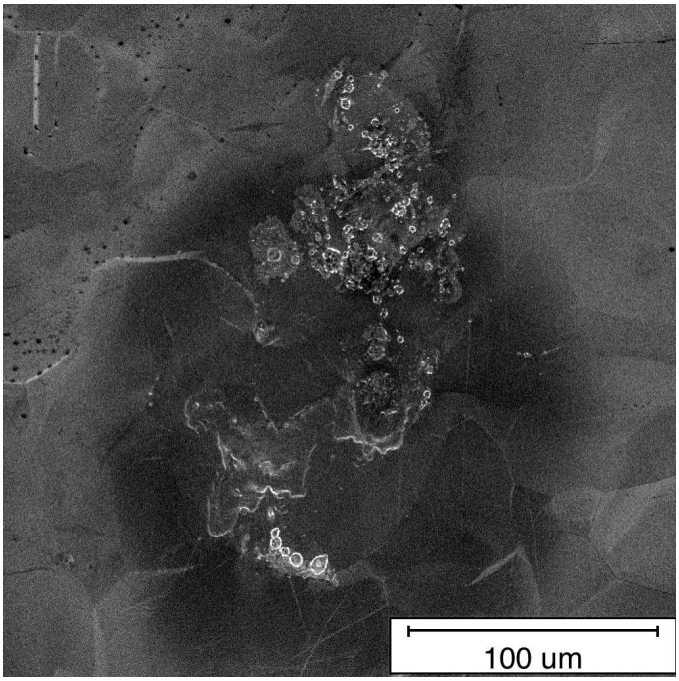


Figure G.27: BJ-Ped3-SB1

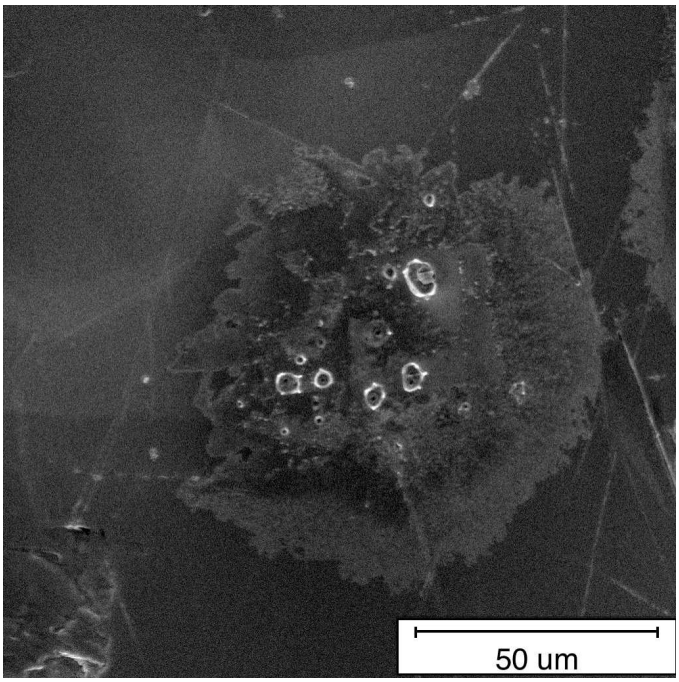


Figure G.28: BJ-Ped4-SB1

G.4 Starbursts on Electropolished Copper

Copper cathodes were machined at Cornell from oxygen-free high-conductivity copper, and electropolished at Peking University. Different cathodes were electropolished different amounts, removing between 50 and 200 microns. Although the pedestal edges were very much rounded by the electropolishing, long scratches of unknown origin persisted on the surface. Also, the samples were shipped back to Cornell in standard plastic bags, which are known to contaminate the surface.

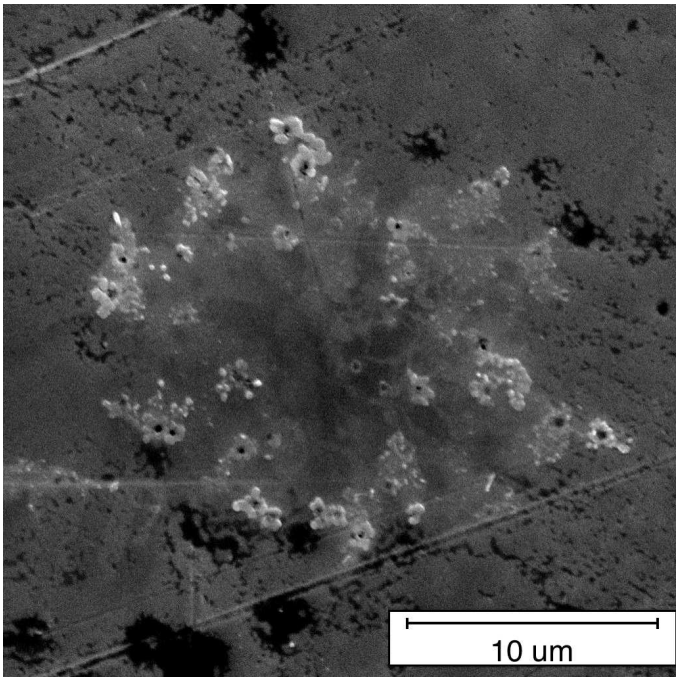


Figure G.29: C2-Ped1-SB14

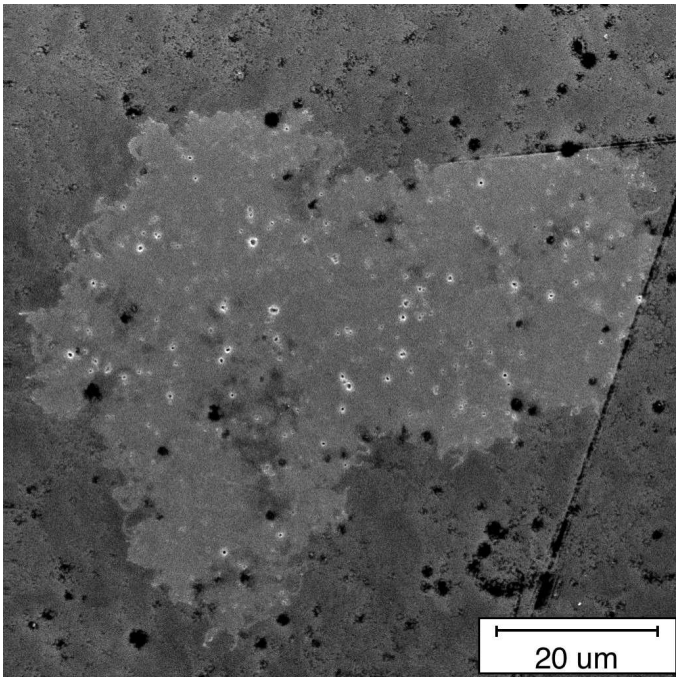


Figure G.30: C2-Ped4-SB1

G.5 Starbursts on Diamond-Machined Copper

Two copper cathode samples, labeled CD1 and CD2, were diamond machined at CERN. Unfortunately, the diamond machining left the cathode pedestals with very sharp corners, which affected the breakdown voltage (see figures G.39 and G.40; note how so many starbursts are close to the edge).

Copper plates prepared at SLAC had larger (3 mm) pedestals so the sharp corners would be farther from the anode. Diamond machined copper plates prepared at SLAC were prepared as copper cavities (NLCTA cells, for the "next linear collider test accelerator"), including similar heat treatment. Bob Kirby described the processing and packaging (for shipment to Cornell) of copper samples:

1. Machine copper coupons per Cornell drawing (except radius two inside corners at the left of the plate and change dimensions at left of plate from 0.15" and 0.061" to 0.13" and 0.080", respectively) provided by Greg Werner, using conventional poly-crystalline diamond turning at Robertson Manufacturing, Inc. Prior to machining, material is stress-relieved (in H_2, 525°C) OFE class 1 NLCTA cell material.

2. Clean and etch coupons using the same process as for NLCTA-structure cells. Degrease coupons in perchloroethylene in the Closed Cycle Degreaser. Use special Teflon wafer carriers for cleaning and handling of coupons. Clean coupons per SLAC Plating Shop process C01 for OFE copper, except: a) Etch time , Step 6 , to be 60 seconds; b) Omit step 8 (sulfuric-nitric); c) Omit Step 11,(no Oxyban); d) Replace Step 10 - DI water rinse - with 30 sec ultrasonic in DI water; e) Replace final rinse with 30 sec ultrasonic in alcohol. Keep coupons in new clean alcohol until just prior to drying and

transfer to wafer storage boxes. Cells to be dried in clean room using filtered dry nitrogen. Wafer storage boxes are Fluoroware conductive part number H22-15-62C02. Place the coupon, face out, in wafer box bottom and close lid. Use this storage method before and after succeeding furnace steps, to protect the coupon from dust. Take care not to invert the wafer box with the coupon in it. Pack the coupons for shipment to Cornell per step 4, below.

3. Simulation of thermal schedule used for structures: a) 1020°C, 2 hrs., dry-H_2 (diffusion bonding step), b) 950°C, x hrs., wet-H_2 (decarburizing step), c) 950°C, dry-H_2, x hrs. (braze cooling lines step), d) vacuum firing, 650°C, 8 hrs., in clean vacuum oven, not the structure can (degas step).

4. Mount coupons, face out, to box cover using roundhead clean stainless round head screws (washer and nut on cover exterior). Flush cover and bottom in filtered nitrogen inside talc-free nylon bag, close wafer box, and seal bag under positive pressure. Double-bag in talc-free polyethylene bag with filtered nitrogen, again under positive pressure. Ship by Fed Ex overnite, never by US post.

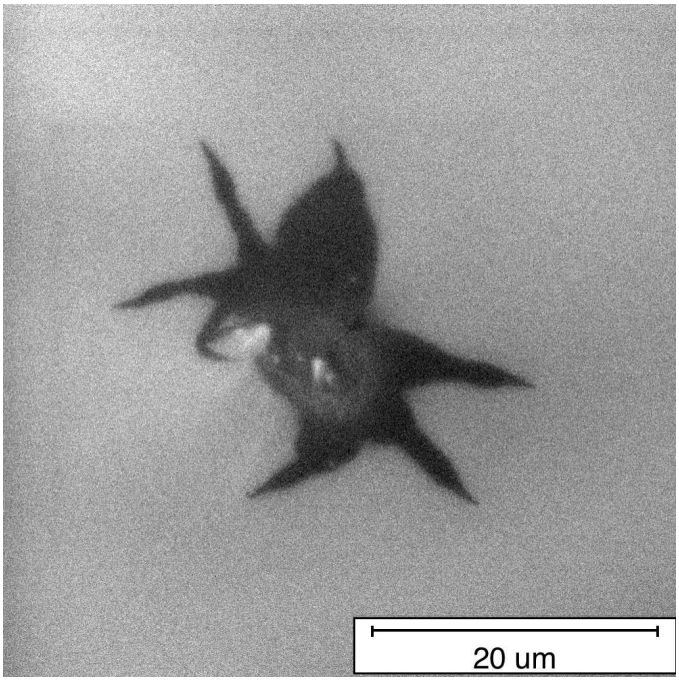


Figure G.31: E60-Ped1-SB3

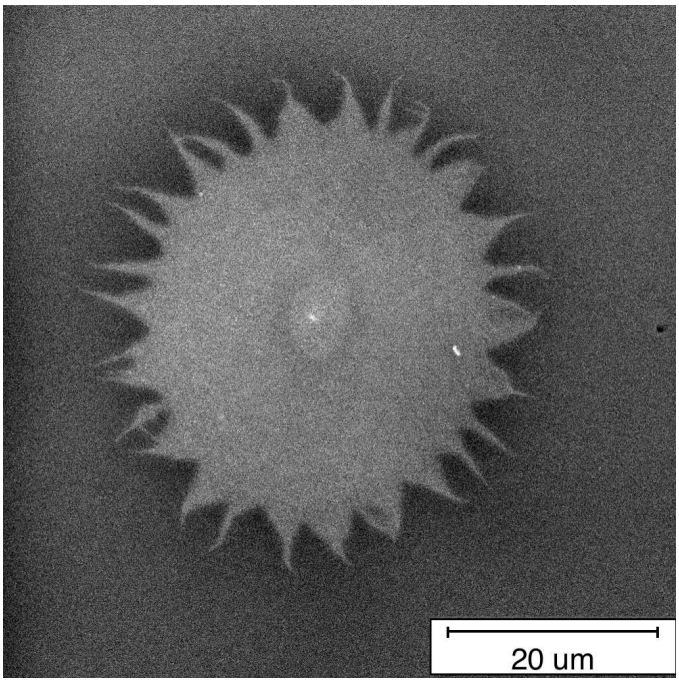


Figure G.32: E60-Ped1-SB43

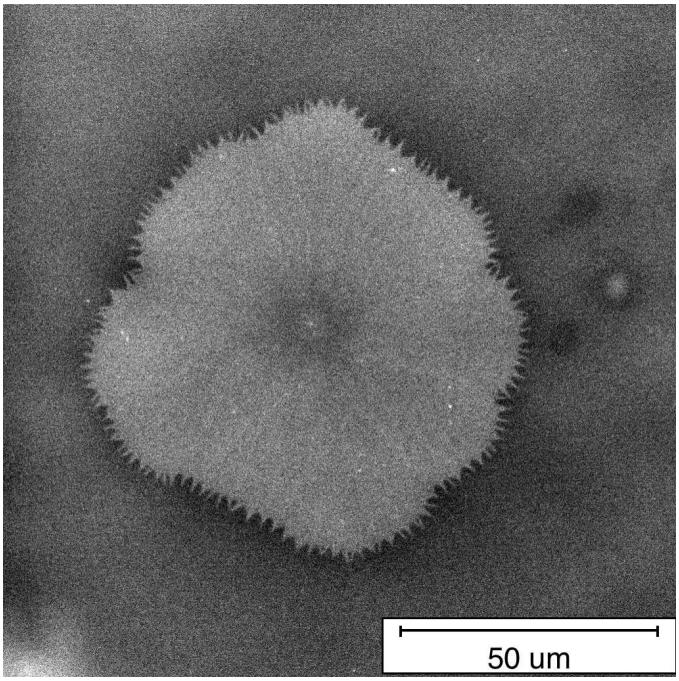


Figure G.33: E60-Ped2-SB13

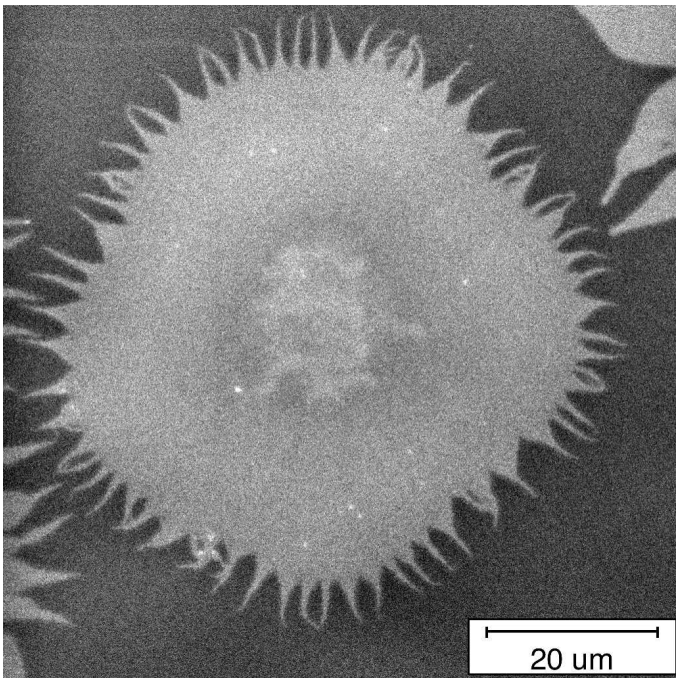


Figure G.34: E60-Ped2-SB4

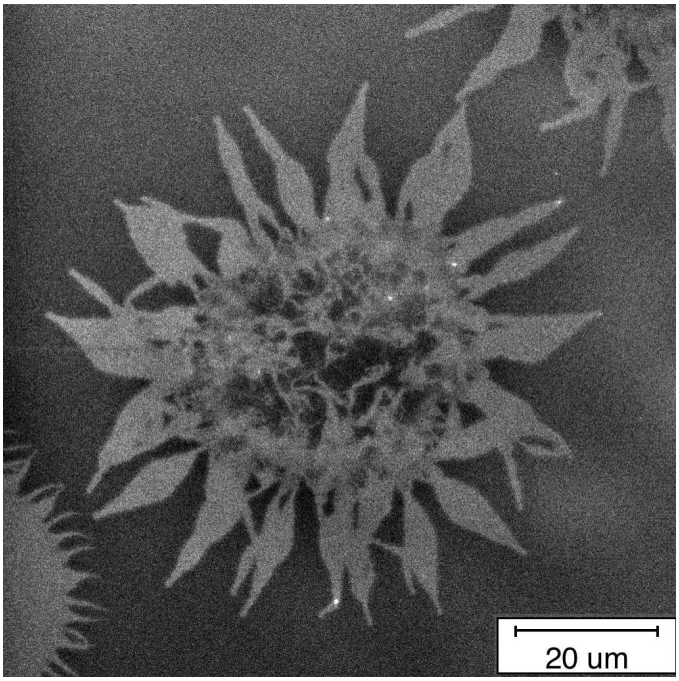


Figure G.35: E60-Ped2-SB5

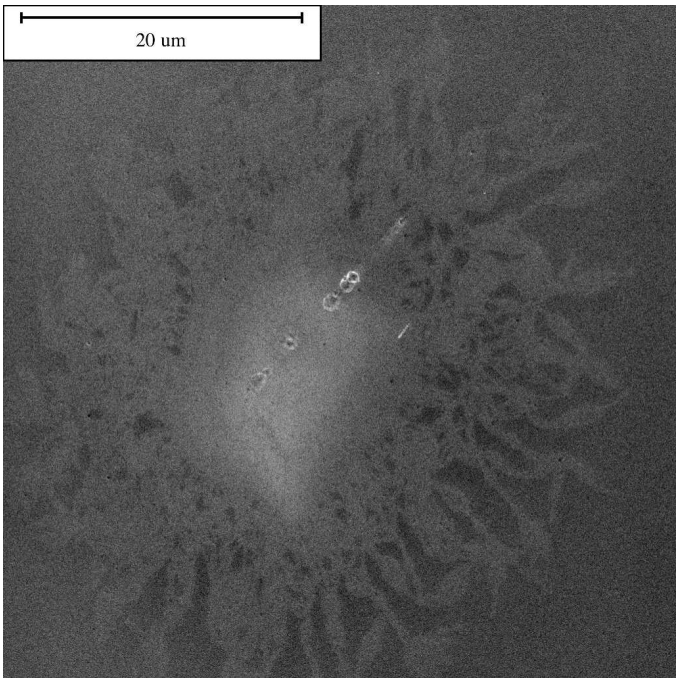


Figure G.36: E0-Ped1-SB3

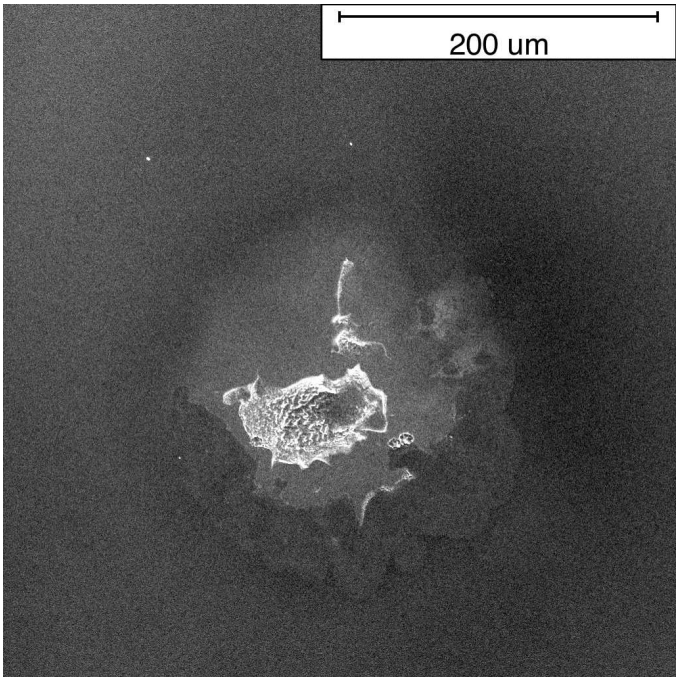


Figure G.37: E0-Ped3-SB1

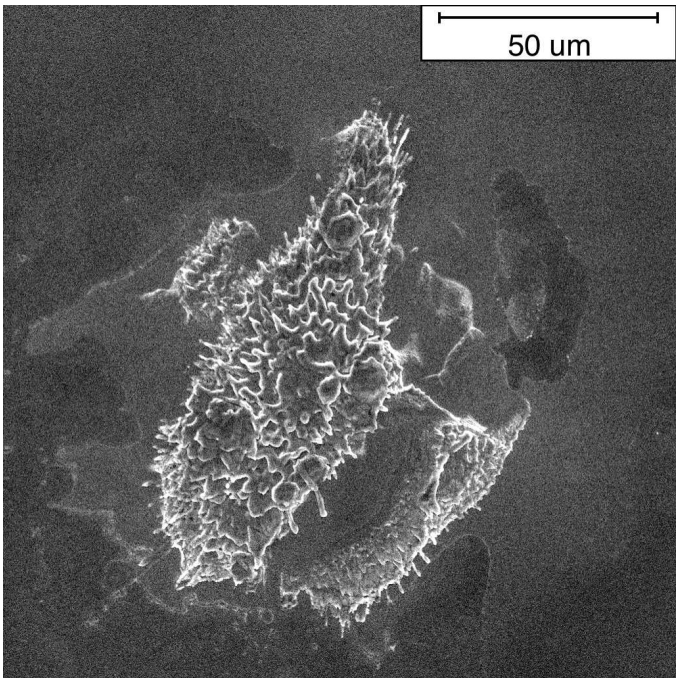


Figure G.38: E30-Ped3-SB1

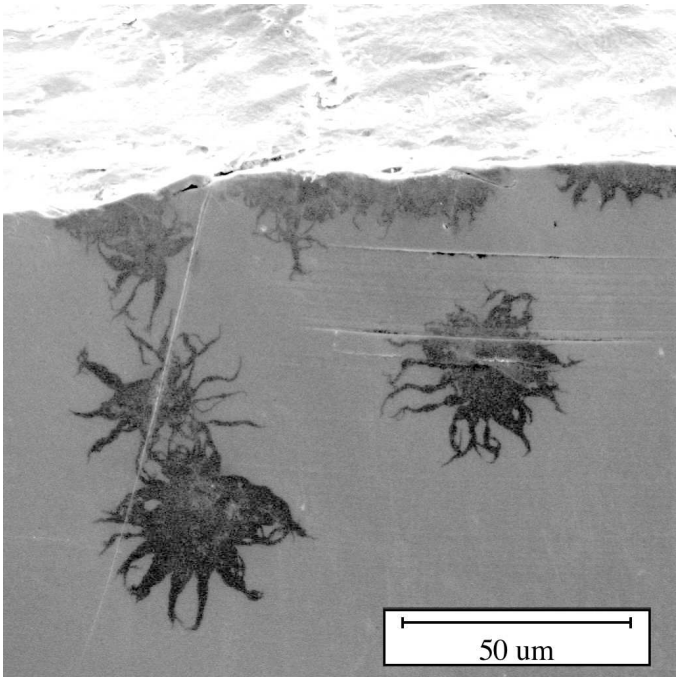


Figure G.39: CD1-PS1-Ped3-UL-SBs

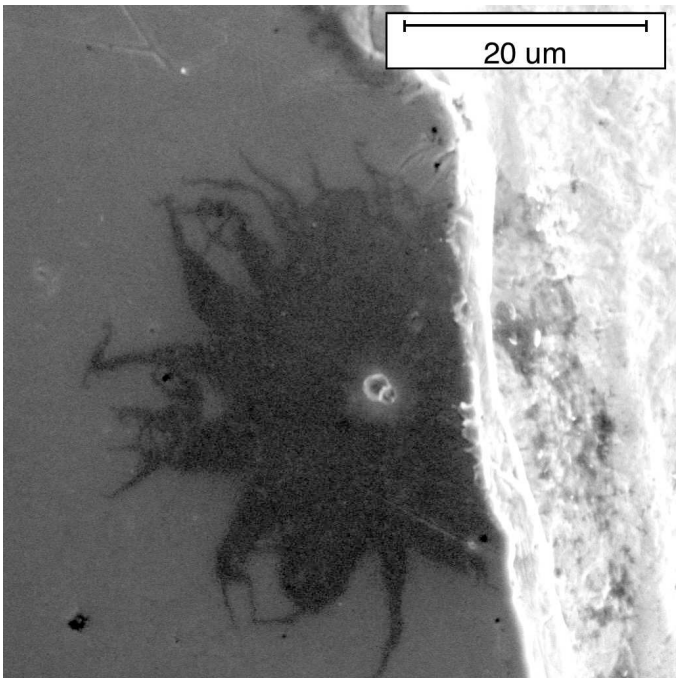


Figure G.40: CD1-PS1-Ped4-UR-SB

G.6 Starbursts on Oxidized Niobium

The oxide of niobium can be easily increased by anodization (see section 5.5). Here we show starbursts on niobium with oxides grown to 400–600 Å thick (about ten times thicker than the natural oxide).

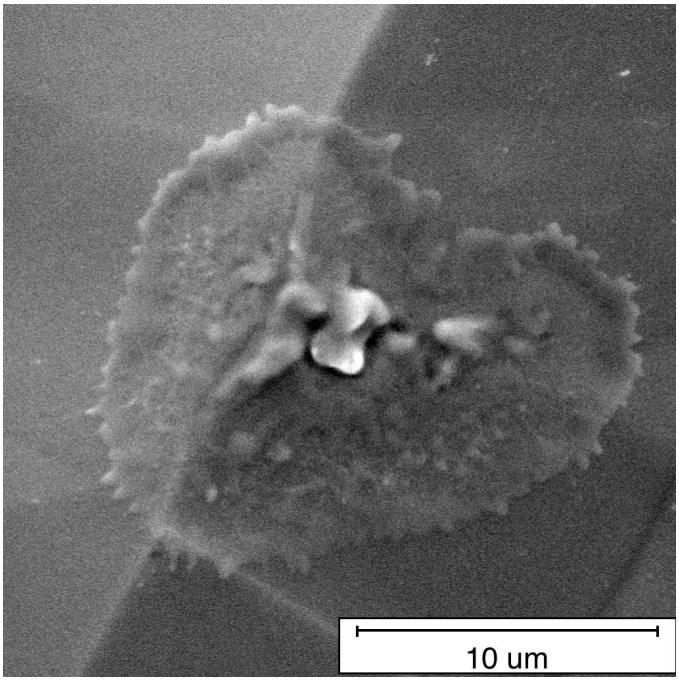


Figure G.41: B8-Ped1-UL-SB9

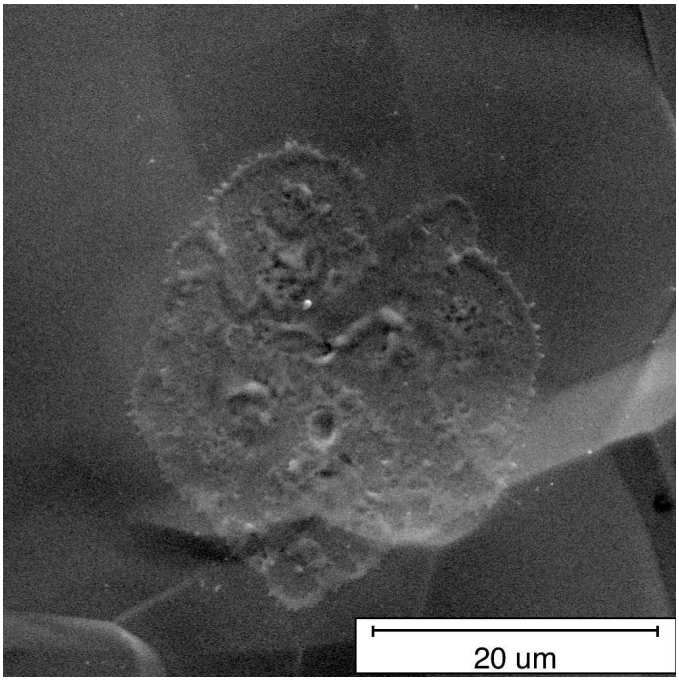


Figure G.42: B8-Ped2-SB10

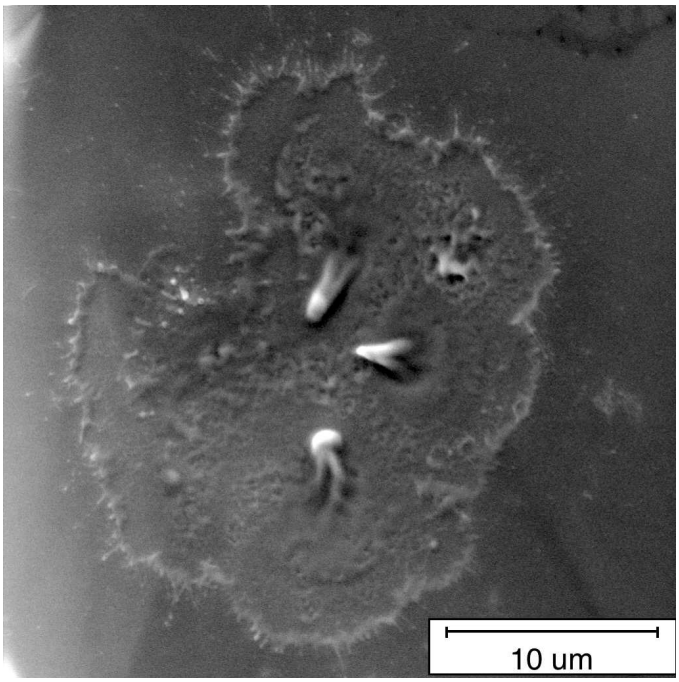


Figure G.43: B8-Ped3-SB8

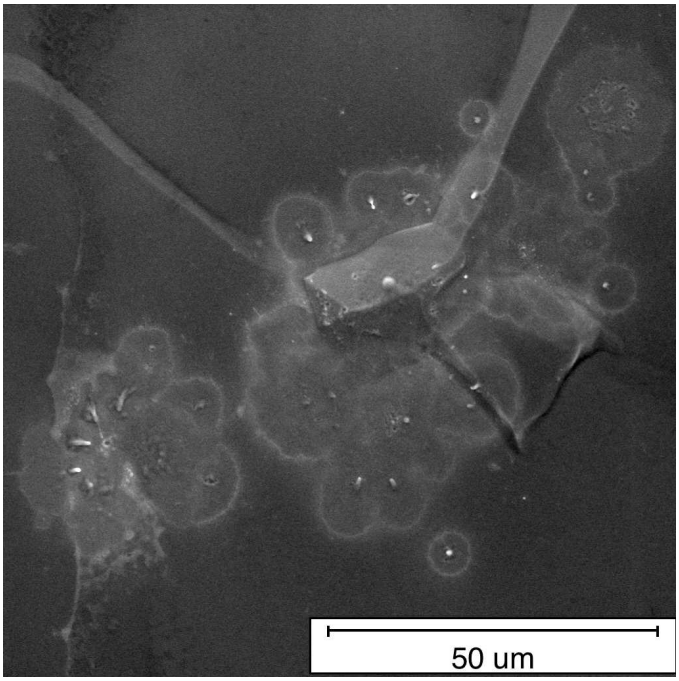


Figure G.44: B8-Ped5-SB7

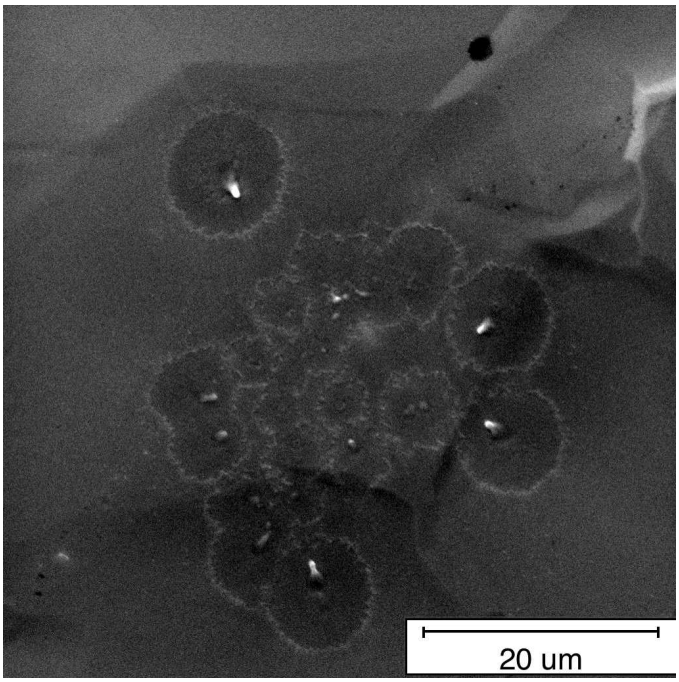


Figure G.45: BE-Ped1-SB1

G.7 Starbursts on Oxidized Copper

A copper cathode was oxidized by heating in air to about 120°C for a minute, turning the copper to a reddish-gold color.

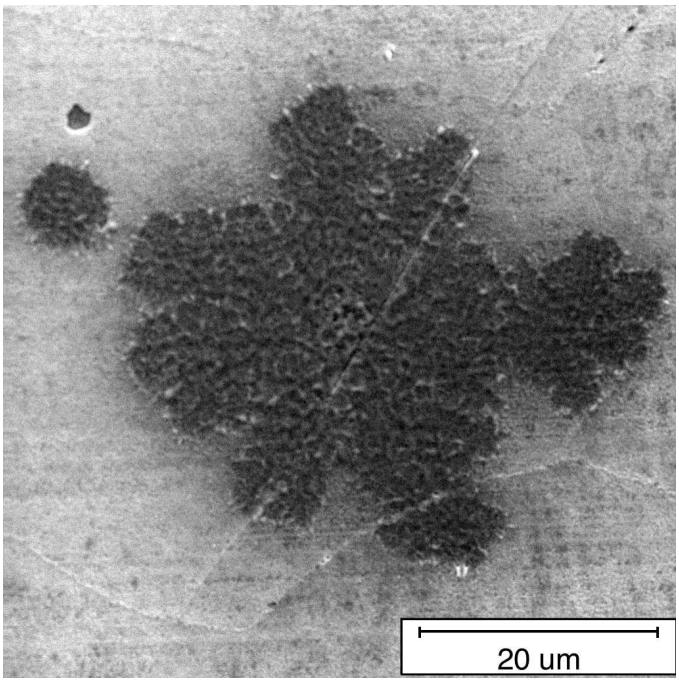


Figure G.46: CD1-PS3-Ped2-SB2

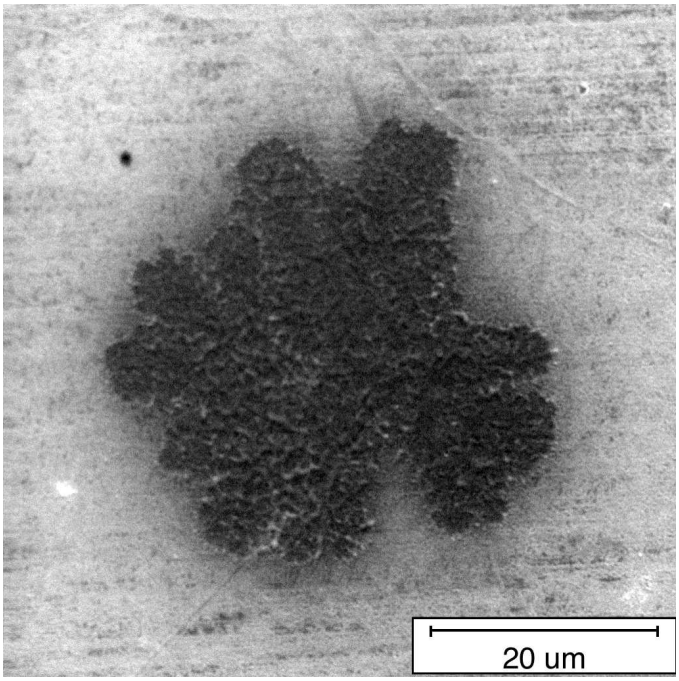


Figure G.47: CD1-PS3-Ped4-SB3

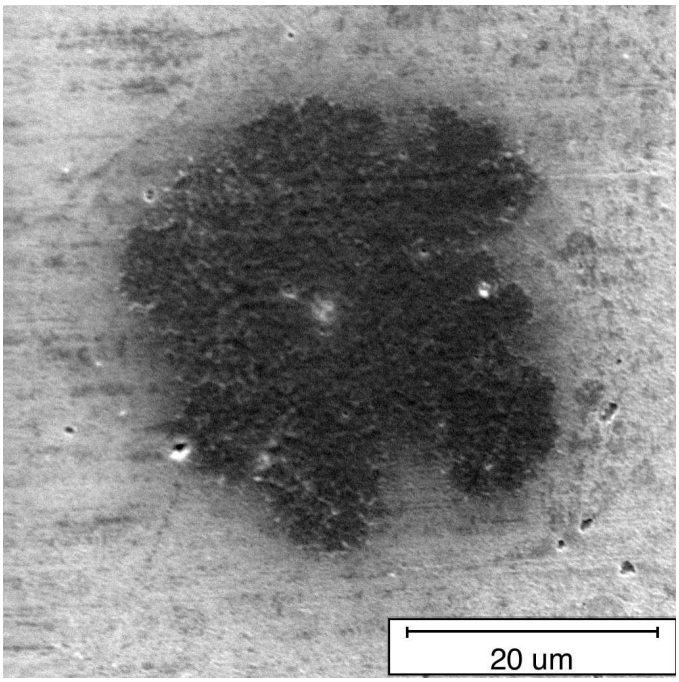


Figure G.48: CD1-PS3-Ped4-SB4

BIBLIOGRAPHY

[1] D. Alpert, D. A. Lee, E. M. Lyman, and H. E. Tomaschke. Initiation of electrical breakdown in vacuum. *Journal of Vacuum Science and Technology*, 1:35–50, 1964.

[2] André Anders and George Yu. Yushkov. Ion flux from vacuum arc cathode spots in the absence and presence of a magnetic field. *Journal of Applied Physics*, 91(8):4824–4832, 2002.

[3] Manuel Arrayás, Ute Ebert, and Willem Hunsdorfer. Spontaneous branching of anode-directed streamers between planar electrodes. *Physical Review Letters*, 88(17):Art. No. 174502, April 29 2002.

[4] Sergey A. Barengolts, Gennady A. Mesyats, and Dmitry Leonidovich Shmelev. Structure and time behavior of vacuum arc cathode spots. *IEEE Transactions on Plasma Science*, 31(5):809–816, 2003.

[5] Alexander Batrakov, Burkhard Jüttner, Sergey Popov, and Dmitry Proskurovsky. Experimental study of the behavior of liquid metal under the plasma of a cathode spot. *IEEE Transactions on Plasma Science*, 31(5):827–831, 2003.

[6] Alexander Batrakov, Sergey Popov, Nadeshda Vogel, Burkhard Jüttner, and Dmitry Proskurovsky. Plasma parameters of an arc cathode spot at the low-current vacuum discharge. *IEEE Transactions on Plasma Science*, 31(5):817–821, 2003.

[7] Carol J. Bennette, Lynwood W. Swanson, and Francis M. Charbonnier. Electrical breakdown between metal electrodes in high vacuum. II. Experimental. *Journal of Applied Physics*, 38(2):634–640, 1967.

[8] C. K. Birdsall and A. B. Langdon. *Plasma Physics via Computer Simulation.* Adam Hilger, 1991.

[9] Charles K. Birdsall and William B. Bridges. *Electron Dynamics of Diode Regions.* Academic Press, 1966.

[10] J. A. Bittencourt. *Fundamentals of Plasma Physics.* Pergamon Press, 1986.

[11] B. Bonin. Field emission studies at Saclay. In Sundelin [99], pages 1033–1046.

[12] B. Bonin, editor. *Proceedings of the 7th Workshop on RF Superconductivity*, Gif sur Yvette, France, 1995.

[13] Raymond L. Boxman, David M. Sanders, and Philip J. Martin, editors. *Handbook of Vacuum Arc Science and Technology.* Noyes Publications, 1995.

[14] I. Brodie. Studies of field emission and electrical breakdown between extended nickel surfaces in vacuum. *Journal of Applied Physics*, 35(8):2324–2332, 1964.

[15] I. Brodie. Prediction of the voltage for electrical breakdown in vacuum. *Journal of Vacuum Science and Technology*, 3(4):222–223, 1966.

[16] I. Brodie and I. Weissman. Use of a cylindrical projection tube for the study of pre-breakdown emission from protrusions on extended surfaces. *Vacuum*, 14(8):299–301, 1964.

[17] David L. Bruhwiler et al. Particle-in-cell simulations of plasma accelerators and electron-neutral collisions. *Physical Review Special Topics—Accelerators and Beams*, 4:Art. No. 101302, 2001.

[18] Eungsun Byon and André Anders. Ion energy distribution functions of vacuum arc plasmas. *Journal of Applied Physics*, 93(4):1899–1906, 2003.

[19] G. Carter and J. S. Collington. *Ion Bombardment of Solids*, chapter 7. American Elsevier Publishing Company, Inc., 1968.

[20] Francis M. Charbonnier, Carol J. Bennette, and Lynwood W. Swanson. Electrical breakdown between metal electrodes in high vacuum. I. Theory. *Journal of Applied Physics*, 38(2):627–633, 1967.

[21] Francis F. Chen. *Introduction to Plasma Physics and Controlled Fusion.* Plenum Press, 1984.

[22] C. D. Child. Discharge from hot CaO. *Physical Review (Series I)*, 32:492–511, 1911.

[23] C. Crawford et al. High gradients in linear collider superconducting accelerator cavities by high pulsed power to suppress field emission. *Particle Accelerators*, 49:1–13, 1995.

[24] D. Kenneth Davies. The initiation of electrical breakdown in vacuum—a review. *Journal of Vacuum Science and Technology*, 10(1):115–121, 1973.

[25] D. Kenneth Davies and Manfred A. Biondi. Vacuum electrical breakdown between plane-parallel copper electrodes. *Journal of Applied Physics*, 37(8):2969–2977, 1966.

[26] D. Kenneth Davies and Manfred A. Biondi. Detection of electrode vapor between plane parallel copper electrodes prior to current amplification and breakdown in vacuum. *Journal of Applied Physics*, 41(1):88–93, 1970.

[27] D. Kenneth Davies and Manfred A. Biondi. Mechanism of dc electrical breakdown between extended electrodes in vacuum. *Journal of Applied Physics*, 42(8):3089–3107, 1971.

[28] D. Kenneth Davies and Manfred A. Biondi. Emission of electrode vapor resonance radiation at the onset of dc breakdown in vacuum. *Journal of Applied Physics*, 48(10):4229–4233, 1977.

[29] William D. Davis and H. Craig Miller. Analysis of the electrode products emitted by dc arcs in a vacuum ambient. *Journal of Applied Physics*, 40(5):2212–2221, 1969.

[30] William T. Diamond. New perspectives in vacuum high voltage insulation. I. the transition to field emission. *Journal of Vacuum Science and Technology A*, 16(2):707–719, 1998.

[31] C. B. Duke and M. E. Alferieff. Field emission through atoms adsorbed on a metal surface. *Journal of Chemical Physics*, 46(3):923–937, 1967.

[32] J. Dutton. Spark breakdown in uniform fields. In Meek and Craggs [73], chapter 3.

[33] W. P. Dyke, J. P. Barbour, E. E. Martin, and J. K. Trolan. T-F emission: Experimental measurement of the average electron current density from tungsten. *Physical Review*, 99(4):1192–1195, 1955.

[34] W. P. Dyke, J. K. Trolan, E. E. Marten, and J. P. Barbour. The field emission initiated vacuum arc. I. Experiments on arc initiation. *Physical Review*, 91(5):1043–1054, 1953.

[35] W. P. Dyke, J. K. Trolan, E. E. Marten, and J. P. Barbour. The field emission initiated vacuum arc. II. The resistively heated emitter. *Physical Review*, 91(5):1054–1057, 1953.

[36] C. D. Ehrlich and E. W. Plummer. Measurement of the absolute tunneling current density in field emission from tungsten(110). *Physical Review B*, 18(8):3767–3771, 1978.

[37] R. H. Fowler and L. Nordheim. Electron emission in intense electric fields. *Proceedings of the Royal Society (London)*, 119:173–181, 1928.

[38] Joseph I. Goldstein and Harvey Yakowitz, editors. *Practical Scanning Electron Microscopy*. Plenum Press, 1975.

[39] J. Graber, J. Kirchgessner, D. Moffat, J. Knobloch, H. Padamsee, and D. Rubin. Microscopic investigation of high gradient superconducting cavities after reduction of field emission. *Nuclear Instruments and Methods in Physics Research A*, 350:582–594, 1994.

[40] Joel Henry Graber. *High Power Processing Studies of 3 GHz Niobium Superconducting Cavities*. PhD thesis, Cornell University, 1993.

[41] Thomas Habermann. *Rastermikroskopische Untersuchung der Feldemission von Metall- und Diamantkathoden (Raster-Microscopic Investigation of Field Emission from Metal and Diamond Cathodes).* PhD thesis, Bergische Universität Wuppertal, 1999.

[42] Erhard Hantzsche. Mysteries of the arc cathode spot: A retrospective glance. *IEEE Transactions on Plasma Science*, 31(5):799–808, 2003.

[43] Achim W. Hassel and Detlef Diesing. Breakdown of ultrathin anodic valve metal oxide films in metal-insulator-metal-contacts compared with metal-insulator-electrolyte contacts. *Thin Solid Films*, 414:296–303, 2002.

[44] T. Hays et al. Microscopic examination and elemental analysis of field emission sites in 5.8 GHz superconducting mushroom cavities. In Sundelin [99], pages 750–762.

[45] B. Henrist et al. The variation of the secondary electron yield and of the desorption yield of copper under electron bombardment: Origin and impact on the conditioning of the LHC. In *Proceedings of the 2002 European Particle Accelerator Conference.*

[46] R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles.* Adam Hilger, 1988.

[47] Yao Zhi Hu, Rahul Sharangpani, and Sing-Pin Tay. Kinetic investigation of copper film oxidation by spectroscopic ellipsometry and reflectrometry. *Journal of Vacuum Science and Technology A*, 18(5):2527–2532, 2000.

[48] M. Jimenez et al. Electron field emission from selectively contaminated cathodes. *Journal of Physics D: Applied Physics*, 26:1503–1509, 1993.

[49] M. Jimenez et al. Electron field emission from large-area cathodes: evidence for the projection model. *Journal of Physics D: Applied Physics*, 27:1038–1045, 1994.

[50] Burkhard Jüttner, Victor F. Puchkarev, Erhard Hantzsche, and Isak Beilis. Cathode spots. In Boxman et al. [13], chapter 3.

[51] K C Kalra, K C Singh, and Mohinder Singh. Formation and breakdown characteristics of anodic oxide films on valve metal. *Indian Journal of Chemistry*, 36A:216–218, March 1997.

[52] C. W. Kimblin. Erosion and ionization in the cathode spot regions of vacuum arcs. *Journal of Applied Physics*, 44(7):3074–3081, 1973.

[53] Jens Knobloch. *Advanced Thermometry Studies of Superconducting RF Cavities.* PhD thesis, Cornell University, 1997.

[54] Jens Knobloch. *Advanced Thermometry Studies of Superconducting RF Cavities*. PhD thesis, Cornell University, 1997. Chapter 4.

[55] Jens Knobloch. *Advanced Thermometry Studies of Superconducting RF Cavities*. PhD thesis, Cornell University, 1997. Chapter 5.

[56] Jens Knobloch. *Advanced Thermometry Studies of Superconducting RF Cavities*. PhD thesis, Cornell University, 1997. Chapter 6.

[57] Jens Knobloch and Hasan Padamsee. Microscopic investigation of field emitters located by thermometry in 1.5 GHz superconducting niobium cavities. *Particle Accelerators*, 53:53–76, 1996.

[58] Paul Kranjec and Lawrence Ruby. Test of the critical theory of electrical breakdown in vacuum. *Journal of Vacuum Science and Technology*, 4(2):94–96, 1967. The article should be titled: Test of the Critical Field Theory of Electrical Breakdown. The correction is given in the Journal of Vacuum Science and Technology, **4**(3):137.

[59] Janusz Kutzner and H. Craig Miller. Ion flux from the cathode region of a vacuum arc. *IEEE Transactions on Plasma Science*, 17(5):688–694, 1989.

[60] Irving Langmuir. The effect of space charge and residual gases on thermionic currents in high vacuum. *Physical Review*, 2(6):450–486, 1913.

[61] Irving Langmuir. The effect of space charge and initial velocities on the potential distribution and thermionic current between parallel plane electrodes. *Physical Review*, 21:419–435, 1923.

[62] R. V. Latham and C. J. S. Chapman. A nonconventional electron optical technique for the dynamic observation of cathode damage prior to electrical breakdown. *Journal of Physics E: Scientific Instruments*, 3:732–734, 1970.

[63] Y. Y. Lau, Youfan Liu, and R. K. Parker. Electron emission: From the Fowler-Nordheim relation to the Child-Langmuir law. *Physics of Plasmas*, 1(6):2082–2085, 1994.

[64] J. D. Lawson. *The Physics of Charged-Particle Beams*. Clarendon Press, 1988.

[65] David R. Lide, editor. *CRC Handbook of Chemistry and Physics*. CRC Press, 83$^{\text{rd}}$ edition, 2002.

[66] R. P. Little and S. T. Smith. Electrical breakdown in vacuum. *IEEE Transactions on Electron Devices*, 12:77–83, 1965.

[67] R. P. Little and W. T. Whitney. Electron emission preceding electrical breakdown in vacuum. *Journal of Applied Physics*, 34(8):2430–2432, 1963.

[68] R. P. Little and W. T. Whitney. Field enhancing projections produced by the application of an electric field. *Journal of Applied Physics*, 36:1502–1504, 1965.

[69] O. Madelung and W. Martienssen, editors. *Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology*, volume 24b. Springer-Verlag, 1994.

[70] E. Mahner et al. Reduced field emission of niobium and copper cathodes. *Journal of Vacuum Science and Technology B*, 13(2):607–610, 1995.

[71] S. Maïssa et al. Experimental study on the luminous radiation associated to the field emission of samples submitted to high rf fields. In Bonin [12].

[72] R. Kenneth Marcus, editor. *Glow Discharge Spectroscopies.* Plenum Press, 1993.

[73] J. M. Meek and J. D. Craggs, editors. *Electrical Breakdown of Gases.* Wiley and Sons, 1978.

[74] Gennady A. Mesyats. Ecton mechanism of the vacuum arc cathode spot. *IEEE Transactions on Plasma Science*, 23(6):879–883, 1995.

[75] Gennady A. Mesyats and Dimitri I. Proskurovsky. *Pulsed Electrical Discharge in Vacuum.* Springer-Verlag, 1989.

[76] R. A. Millikan and R. A. Sawyer. Extreme ultra-violet spectra of hot sparks in high vacua. *Physical Review*, 12:167–170, 1918.

[77] D. Moffat et al. Studies on the nature of field emission sites. *Particle Accelerators*, 40:85–126, 1992.

[78] E. L. Murphy and R. H. Good, Jr. Thermionic emission, field emission, and the transition region. *Physical Review*, 102(6):1464–1473, 1956.

[79] V. A. Nevrovskii and V. I. Rakhovskii. Electrode material release into a vacuum gap and mechanisms of electrical breakdown. *Journal of Applied Physics*, 60(1):125–129, 1986.

[80] Ph. Niedermann et al. Field emission from broad-area niobium cathodes: Effects of high-temperature treatment. *Journal of Applied Physics*, 59(3):892–901, 1986.

[81] Ph. Niedermann et al. Study of field-emitting microstructures using a scanning tunneling microscope. *Journal of Vacuum Science and Technology A*, 8(1):594–597, 1990.

[82] Philipp Niedermann. *Experiments on Enhanced Field Emission.* PhD thesis, Université de Genève, 1986.

[83] R. J. Noer. Electron field emission from broad-area electrodes. *Applied Physics A*, 28:1–24, 1982.

[84] H. Padamsee. RF superconducting accelerating cavities. In R. V. Latham, editor, *High Voltage Vacuum Insulation: Basic Concepts and Technological Practice*, chapter 12. Academic Press, 1995.

[85] Hasan Padamsee. The science and technology of superconducting cavities for accelerators. *Superconductor Science and Technology*, 14(4):R28–R51, 2001.

[86] Hasan Padamsee, Jens Knobloch, and Tom Hays. *RF Superconductivity for Accelerators*. Wiley and Sons, 1998.

[87] Hasan Padamsee, Jens Knobloch, and Tom Hays. *RF Superconductivity for Accelerators*, chapter 12. Wiley and Sons, 1998.

[88] Hasan Padamsee, Jens Knobloch, and Tom Hays. *RF Superconductivity for Accelerators*, chapter 2. Wiley and Sons, 1998.

[89] Robert M. Phillips, editor. *High Energy Density Microwaves*. American Institute of Physics Conference Proceedings 474. Springer Verlag, 1998.

[90] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes*. Cambridge University Press, 1986.

[91] R. Rejoub, B. G. Lindsay, and R. F. Stebbings. Determination of the absolute partial and total cross sections for electron-impact ionization of the rare gases. *Physical Review A*, 65:Art. No. 042713, 2002.

[92] Carsten Rusteberg, Manfred Lindmayer, Burkhard Jüttner, and Heinz Pursch. On the ion energy distribution of high current arcs in vacuum. *IEEE Transactions on Plasma Science*, 23(6):909–914, 1995.

[93] G. J. Sayag, Nguyen Tuong Viet, H. Bergeret, and A. Septier. Field emission from oxidized niobium electrodes at 295 and 4.2k. *Journal of Physics E: Scientific Instruments*, 10:176–179, 1977.

[94] Fred Schwirzke. Laser induced unipolar arcing. In Heinrich Hora and George H. Miley, editors, *Laser Interaction and Related Plasma Phenomena*, volume 6, pages 335–352. Plenum Press, 1984.

[95] Fred Schwirzke, Michael P. Hallal, Jr., and Xavier K. Maruyama. Onset of breakdown and formation of cathode spots. *IEEE Transactions on Plasma Science*, 21(5):410–415, 1993.

[96] Fred R. Schwirzke. Vacuum breakdown on metal surfaces. *IEEE Transactions on Plasma Science*, 19(5):690–696, 1991.

[97] Robert Sherman, John Grob, and Walter Whitlock. Dry surface cleaning using CO_2 snow. *Journal of Vacuum Science and Technology B*, 9(4):1970–1977, 1991.

[98] Robert Sherman, Drew Hirt, and Ronald Vane. Surface cleaning with the carbon dioxide snow jet. *Journal of Vacuum Science and Technology A*, 12(4):1876–1881, 1994.

[99] Ronald M. Sundelin, editor. *Proceedings of the 6th Workshop on RF Superconductivity*, Newport News, Virginia, 1993.

[100] L. W. Swanson, L. C. Crouser, and F. M. Charbonnier. Energy exchanges attending field electron emission. *Physical Review*, 151(1):327–151, 1966.

[101] J. Tan, B. Bonin, and H. Safa. Field emission and high voltage cleaning of particulate contaminants on extended metallic surfaces. In Bonin [12].

[102] H. E. Tomaschke and D. Alpert. Role of submicroscopic projections in electrical breakdown. *Journal of Vacuum Science and Technology*, 4(4):192–198, 1967.

[103] Koichi Tsuruta, Kyohei Sekiya, and Gin-ichi Watanabe. Velocities of copper and silver ions generated from an impulse vacuum arc. *IEEE Transactions on Plasma Science*, 25(4):603–608, 1997.

[104] Martin A. Uman. *The Lightning Discharge*. Academic Press, 1987.

[105] J. P. Verboncoeur, A. B. Langdon, and N. T. Gladd. An object-oriented electromagnetic PIC code. *Computer Physics Communications*, 87:199–211, 1995.

[106] Robert C. Weast, editor. *CRC Handbook of Chemistry and Physics*. CRC Press, 70^{th} edition, 1989.

[107] N. S. Xu and R. V. Latham. Electron emission based breakdown mechanisms. In R. V. Latham, editor, *High Voltage Vacuum Insulation: Basic Concepts and Technological Practice*, chapter 5. Academic Press, 1995.

[108] Y. T. Yen, D. T. Tuma, and D. K. Davies. Emission of electrode vapor resonance radiation at the onset of impulsive breakdown in vacuum. *Journal of Applied Physics*, 55(9):3301–3307, 1984.

[109] L. Young. Anodic oxide films on niobium: Thickness, dielectric constant, dispersion, reflection minima, formation field strength, and surface area. *Canadian Journal of Chemistry*, 38:1141–1146, 1960.

[110] George Yushkov. Measurement of directed ion velocity in vacuum arc plasmas by arc current perturbation methods. In *Proceedings of the IXX International Symposium on Discharges and Electrical Insulation in Vacuum*, pages 260–263, Xi'an, P.R. China, 2000. IEEE.

[111] George Yu. Yushkov, André Anders, Efim M. Oaks, and Ian G. Brown. Ion velocities in vacuum arc plasmas. *Journal of Applied Physics*, 88(10):5618–5622, 2000.