PLANAR MICROWAVE ENGINEERING

A Practical Guide to Theory, Measurement, and Circuits

THOMAS H. LEE
Stanford University

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2004

Printed in the United States of America

Typeface Times 10.75/13.5 and Futura *System* AMS-TEX [FH]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication data
Lee, Thomas H., 1959–
Planar microwave engineering : a practical guide to theory, measurement, and circuits /
Thomas Lee.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-83526-7
1. Microwave circuits. 2. Microwave receivers. 3. Microwave devices. I. Title.

TK7876.L424 2004
621.381′32 – dc22 2004050811

ISBN 0 521 83526 7 hardback

CONTENTS

CHAPTER ONE

A MICROHISTORY OF MICROWAVE TECHNOLOGY

1.1 INTRODUCTION

Many histories of microwave technology begin with James Clerk Maxwell and his equations, and for excellent reasons. In 1873, Maxwell published *A Treatise on Electricity and Magnetism,* the culmination of his decade-long effort to unify the two phenomena. By arbitrarily adding an extra term (the "displacement current") to the set of equations that described all previously known electromagnetic behavior, he went beyond the known and predicted the existence of electromagnetic waves that travel at the speed of light. In turn, this prediction inevitably led to the insight that light itself must be an electromagnetic phenomenon. Electrical engineering students, perhaps benumbed by divergence, gradient, and curl, often fail to appreciate just how revolutionary this insight was.[1] Maxwell did not introduce the displacement current to resolve any outstanding conundrums. In particular, he was not motivated by a need to fix a conspicuously incomplete continuity equation for current (contrary to the standard story presented in many textbooks). Instead he was apparently inspired more by an aesthetic sense that nature simply should provide for the existence of electromagnetic waves. In any event the word *genius,* though much overused today, certainly applies to Maxwell, particularly given that it shares origins with *genie*. What he accomplished was magical and arguably ranks as the most important intellectual achievement of the 19th century.[2]

Maxwell – genius and genie – died in 1879, much too young at age 48. That year, Hermann von Helmholtz sponsored a prize for the first experimental confirmation of Maxwell's predictions. In a remarkable series of investigations carried out between

[1] Things could be worse. In his treatise of 1873, Maxwell expressed his equations in terms of *quaternions.* Oliver Heaviside and Josiah Willard Gibbs would later reject quaternions in favor of the language of vector calculus to frame Maxwell's equations in the form familiar to most modern engineers.

[2] The late Nobel physicist Richard Feynman often said that future historians would still marvel at Maxwell's work, long after another event of that time – the American Civil War – had faded into merely parochial significance.

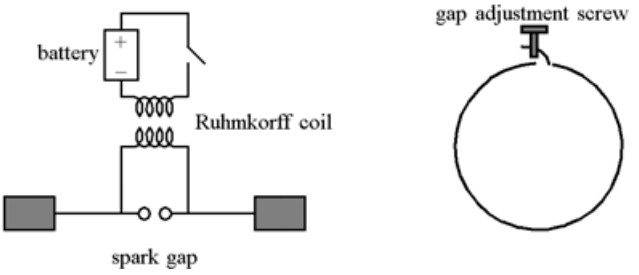


FIGURE 1.1. Spark transmitter and receiver of Hertz

1886 and 1888 at the Technische Hochschule in Karlsruhe, Helmholtz's former pupil, Heinrich Hertz, verified that Maxwell was indeed correct. Another contestant in the race, Oliver Lodge (then a physics professor at University College in Liverpool), published his own confirmation one month after Hertz, having interrupted his work in order to take a vacation. Perhaps but for that vacation we would today be referring to *lodgian waves* with frequencies measured in *megalodges.* Given that *Hertz* is German for *heart* and that the human heart beats about once per second, it is perhaps all for the best that Lodge didn't win the race.

How did Hertz manage to generate and detect electromagnetic waves with equipment available in the 1880s? Experimental challenges certainly extend well beyond the mere generation of some sort of signal; a detector is required, too. Plus, to verify wave behavior, you need apparatus that is preferably at least a couple of wavelengths in extent. In turn, that requirement implies another: sufficient lab space to contain apparatus of that size (and preferably sufficient to treat the room as infinitely large, relative to a wavelength, so that unwanted reflections from walls and other surfaces may be neglected). Hertz, then a junior faculty member, merited a modest laboratory whose useful internal dimensions were approximately 12 m by 8 m.[3] Hertz understood that the experimental requirements forced him to seek the generation of signals with wavelengths of the order of a meter. He accomplished the difficult feat of generating such short waves by elaborating on a speculation by the Irish physicist George Francis FitzGerald, who had suggested in 1883 that one might use the known oscillatory spark discharge of Leyden jars (capacitors) to generate electromagnetic waves. Recognizing that the semishielded structure of the jars would prevent efficient radiation, Hertz first modified FitzGerald's idea by "unrolling" the cylindrical conductors in the jars into flat plates. Then he added inductance in the form of straight wire connections to those plates in order to produce the desired resonant frequency of a few hundred megahertz. In the process, he thereby invented the dipole antenna. Finally, he solved the detection problem by using a ring antenna with an integral spark gap. His basic transmitter–receiver setup is shown in Figure 1.1. When the

[3] Hugh G. J. Aitken, *Syntony and Spark,* Princeton University Press, Princeton, NJ, 1985.

switch is closed, the battery charges up the primary of the Ruhmkorff coil (an early transformer). When the switch opens, the rapid collapse of the magnetic field induces a high voltage in the secondary, causing a spark discharge. The sudden change in current accompanying the discharge excites the antenna to produce radiation.

Detection relies on the induction of sufficient voltage in the ring resonator to produce a visible spark. A micrometer screw allows fine adjustment, and observation in the dark permits one to increase measurement sensitivity.[4]

With this apparatus (a very longwave version of an optical interferometer), Hertz demonstrated essential wave phenomena such as polarization and reflection.[5] Measurements of wavelength, coupled with analytical calculations of inductance and capacitance, confirmed a propagation velocity sufficiently close to the speed of light that little doubt remained that Maxwell had been right.[6]

We will never know if Hertz would have gone beyond investigations of the pure physics of the phenomena to consider practical uses for wireless technology, for he died of blood poisoning (from an infected tooth) in 1894 at the age of 36. *Brush and floss after every meal, and visit your dentist regularly.*

Maxwell's equations describe electric and magnetic fields engaged in an eternal cycle of creation, destruction, and rebirth. Fittingly, Maxwell's death had inspired von Helmholtz to sponsor the prize which had inspired Hertz. Hertz's death led to the publication of a memorial tribute that, in turn, inspired a young man named Guglielmo Marconi to dedicate himself to developing commercial applications of wireless. Marconi was the neighbor and sometime student of Augusto Righi, the University of Bologna professor who had written that tribute to Hertz. Marconi had been born into a family of considerable means, so he had the time and finances to pursue his dream.[7] By early 1895, he had acquired enough apparatus to begin experiments in and around his family's villa, and he worked diligently to increase transmission distances. Marconi used Hertz's transmitter but, frustrated by the inherent limitations of a spark-gap detector, eventually adopted (then adapted) a peculiar creation that had been developed by Edouard Branly in 1890. As seen in Figure 1.2, the device, dubbed a *coherer* by Lodge, consists of a glass enclosure filled with a loosely packed and perhaps slightly oxidized metallic powder. Branly had accidentally discovered that the resistance of this structure changes dramatically when nearby

[4] Hertz is also the discoverer of the photoelectric effect. He noticed that sparks would occur more readily in the presence of ultraviolet light. Einstein would win his Nobel prize for providing the explanation (and not for his theory of relativity, as is frequently assumed).

[5] The relative ease with which the waves were reflected would inspire various researchers to propose crude precursors to radar within a relatively short time.

[6] This is not to say that everyone was immediately convinced; they weren't. Revolutions take time.

[7] Marconi's father was a successful businessman, and his mother was an heiress to the Jameson Irish whiskey fortune. Those family connections would later prove invaluable in gaining access to key members of the British government after Italian officials showed insufficient interest. The British Post Office endorsed Marconi's technology and supported its subsequent development.

FIGURE 1.2. Branly's coherer

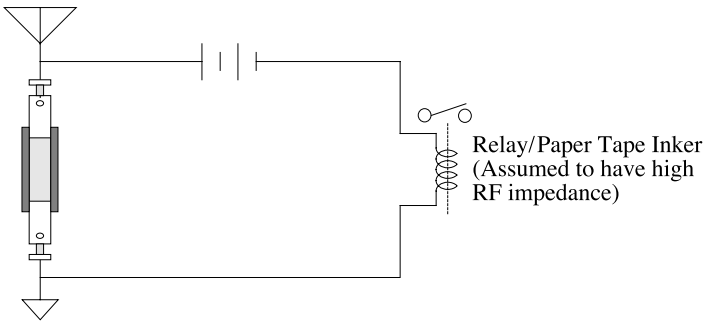


FIGURE 1.3. Typical receiver with coherer

electrical apparatus is in operation. It must be emphasized that the detailed principles that underlie the operation of coherers remain mysterious, but that ignorance doesn't prevent us from describing their electrical behavior.[8]

A coherer's resistance generally has a large value (say, megohms) in its quiescent state and then drops to kilohms or less when triggered by some sort of an EM event. This large resistance change in turn may be used to trigger a solenoid to produce an audible click, as well as to ink a paper tape for a permanent record of the received signal. To prepare the coherer for the next EM pulse, it has to be shaken (or stirred) to restore the "incoherent" high-resistance state. Figure 1.3 shows how a coherer can be used in a receiver. It is evident that the coherer is a digital device and therefore unsuitable for uses other than radiotelegraphy.

The coherer never developed into a good detector, it just got less bad over time. Marconi finally settled on the configuration shown in Figure 1.4. He greatly reduced the spacing between the end plugs, filled the intervening space with a particular mixture of nickel and silver filings of carefully selected size, and partially evacuated the tube prior to sealing the assembly. As an additional refinement in the receiver, a solenoid provided an audible indication in the process of automatically whacking the detector back into its initial state after each received pulse.

Even though many EM events other than the desired signal could trigger a coherer, Marconi used this erratic device with sufficient success to enable increases

[8] Lodge named these devices *coherers* because the filings could be seen to stick together under some circumstances. However, the devices continue to function as detectors even without observable physical movement of the filings. It is probable that oxide breakdown is at least part of the explanation, but experimental proof is absent for lack of interest in these devices.

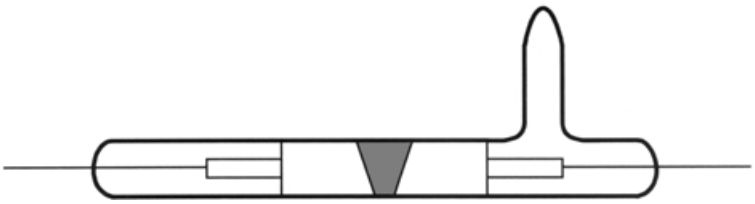


FIGURE 1.4. Marconi's coherer

in communication range to about three kilometers by 1896. As he scaled upward in power, he used progressively larger antennas, which had the unintended side effect of lowering the "carrier" frequencies to below 100 kHz from his initial frequencies of ∼100 MHz. This change was most fortuitous, because it allowed reflections from the ionosphere (whose existence was then unknown) to extend transmission distances well beyond the horizon, allowing him to claim successful transatlantic wireless communications by 12 December 1901.[9] Wireless technology consequently ignored the spectrum above 1 MHz for nearly two more decades, thanks to a belief that communication distances were greatest below 100 kHz.

As the radio art developed, the coherer's limitations became increasingly intolerable, spurring the search for improved detectors. Without a body of theory to impose structure, however, this search was haphazard and sometimes took bizarre turns. A human brain from a fresh cadaver was once tried as a coherer, with the experimenter claiming remarkable sensitivity for his apparatus.[10]

That example notwithstanding, most detector research was based on the vague notion that a coherer's operation depends on some mysterious property of imperfect contacts. Following this intuition, a variety of experimenters stumbled, virtually simultaneously, on various types of point-contact crystal detectors. The first patent application for such a device was filed in 1901 by the remarkable Jagadish Chandra Bose for a detector using galena (lead sulfide).[11] See Figures 1.5 and 1.6. This detector exploits a semiconductor's high temperature coefficient of resistance, rather than rectification.[12] As can be seen in the patent drawing, electromagnetic

[9] Marconi's claim was controversial then, and it remains so. The experiment itself was not double-blind, as both the sender and the recipient knew ahead of time that the transmission was to consist of the letter *s* (three dots in Morse code). Ever-present atmospheric noise is particularly prominent in the longwave bands he was using at the time. The best modern calculations reveal that the three dots he received had to have been noise, not signal. One need not postulate fraud, however. Unconscious experimenter bias is a well-documented phenomenon and is certainly a possibility here. In any case, Marconi's apparatus evolved enough within another year to enable verifiable transatlantic communication.

[10] A. F. Collins, *Electrical World and Engineer,* v. 39, 1902; he started out with brains of other species and worked his way up to humans.

[11] U.S. Patent #755,840, granted 19 March 1904. The patent renders his name Jagadis Chunder Bose. The transliteration we offer is that used by the academic institution in Calcutta that bears his name.

[12] Many accounts of Bose's work confuse his galena balometer with the point-contact rectifying ("catwhisker" type) detectors developed later by others and thus erroneously credit him with the

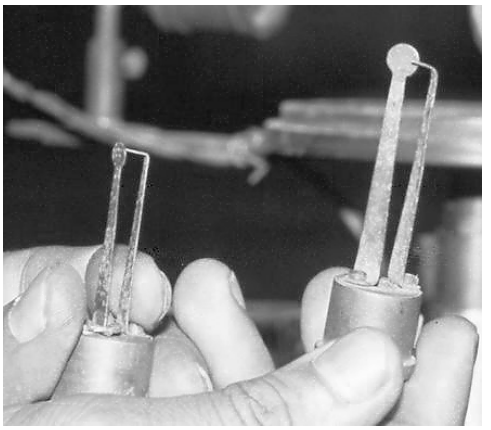


FIGURE 1.5. Actual detector mounts used by Bose (galena not shown) [*courtesy of David Emerson*]

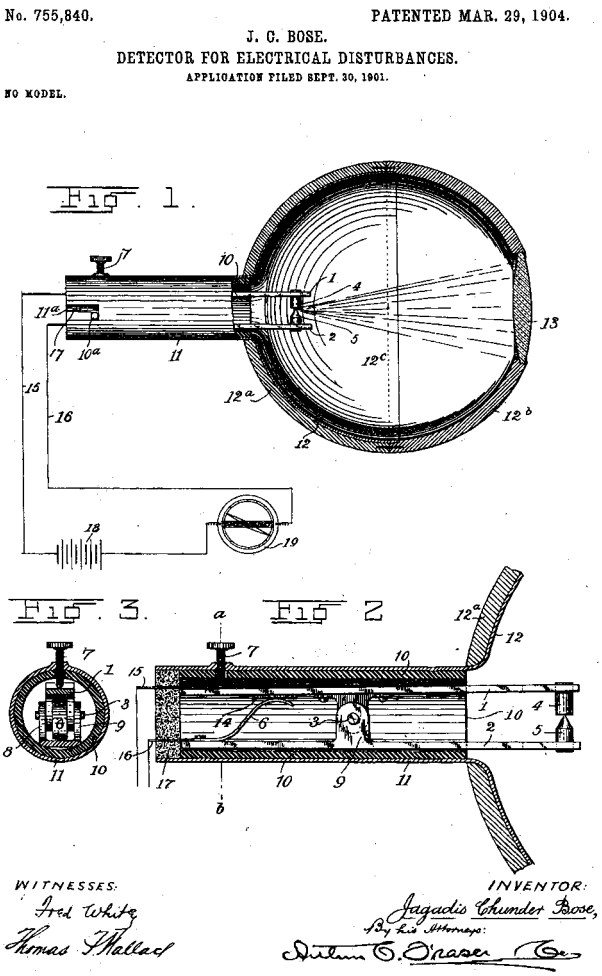
radiation is focused on the point contact, and the resistance change that accompanies the consequent heating registers as a change in current flowing through an external circuit. This type of detector is known as a *bolometer.* In refined form, bolometers remain useful as a means of measuring power, particularly of signals whose frequency is so high that there are no other means of detection. Bose used this detector in experiments extending to approximately 60 GHz, about which he first published papers in 1897.[13] His research into millimeter-wave phenomena was decades ahead of his time.[14] So too was the recognition by Bose's former teacher at Cambridge, Lord Rayleigh, that hollow conductors could convey electromagnetic energy.[15] Waveguide transmission would be forgotten for four decades, but Rayleigh had most of it worked out (including the concept of a cutoff frequency) in 1897.

invention of the semiconductor diode. The latter functions by rectification, of course, and thus does not require an external bias. It was Ferdinand Braun who first reported asymmetrical conduction in galena and copper pyrites (among others), back in 1874, in "Ueber die Stromleitung durch Schwefelmetalle" [On Current Flow through Metallic Sulfides], *Poggendorff's Annalen der Physik und Chemie,* v. 153, pp. 556–63. Braun's other important development for wireless was the use of a spark gap in series with the primary of a transformer whose secondary connects to the antenna. He later shared the 1909 Nobel Prize in physics with Marconi for contributions to the radio art.

[13] J. C. Bose, "On the Determination of the Wavelength of Electric Radiation by a Diffraction Grating," *Proc. Roy. Soc.,* v. 60, 1897, pp. 167–78.

[14] For a wonderful account of Bose's work with millimeter waves, see David T. Emerson, "The Work of Jagadis Chandra Bose: 100 Years of MM-Wave Research," *IEEE Trans. Microwave Theory and Tech.,* v. 45, no. 12, 1997, pp. 2267–73.

[15] Most scientists and engineers are familar with Rayleigh's extensive writings on acoustics, which include analyses of ducting (acoustic waveguiding) and resonators. Far fewer are aware that he also worked out the foundations for electromagnetic waveguides at a time when no one could imagine a use for the phenomenon and when no one but Bose could even generate waves of a high enough frequency to propagate through reasonably small waveguides.


No. 755,840. PATENTED MAR. 29, 1904.

J. C. BOSE.

DETECTOR FOR ELECTRICAL DISTURBANCES.

APPLICATION FILED SEPT. 30, 1901.

NO MODEL.

FIG. 1.

FIG. 3.

FIG. 2.

WITNESSES:

Fred. White

Thomas C. Mallard

INVENTOR:

Jagadis Chunder Bose,

By his Attorneys:

Dyrenforth & Dyrenforth

FIGURE 1.6. Bose's bolometer patent (first page)

This patent appears to be the first awarded for a semiconductor detector, although it was not explicitly recognized as such because semiconductors were not yet acknowledged as a separate class of materials (indeed, the word *semiconductor* had not yet been coined). Work along these lines continued, and General Henry Harrison Chase Dunwoody filed the first patent application for a rectifying detector using carborundum (SiC) on 23 March 1906, receiving U.S. Patent #837,616 on 4 December of that year. A later application, filed on 30 August 1906 by Greenleaf Whittier Pickard (an MIT graduate whose great-uncle was the poet John Greenleaf Whittier) for a silicon (!) detector, resulted in U.S. Patent #836,531 just ahead of Dunwoody, on 20 November (see Figure 1.7).

As shown in Figure 1.8, one connection consists of a small wire (whimsically known as a catwhisker) that makes a point contact to the crystal surface. The other connection is a large area contact canonically formed by a low–melting-point alloy

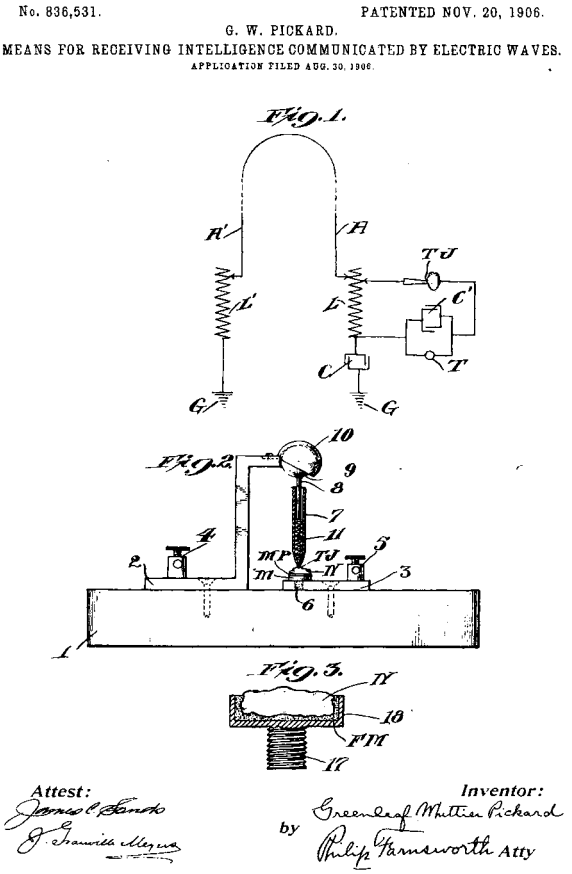


FIGURE 1.7. The first silicon diode patent

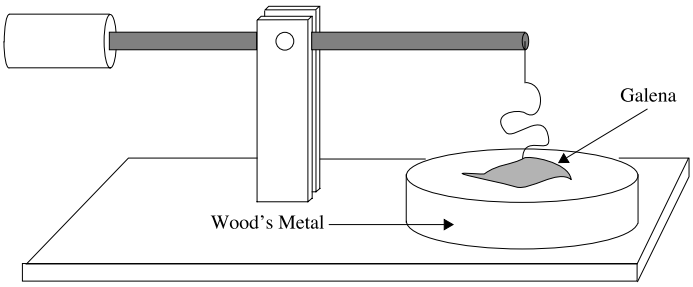


FIGURE 1.8. Typical crystal detector

(usually a mixture of lead, tin, bismuth, and cadmium known as Wood's metal, which has a melting temperature of under 80°C), that surrounds the crystal.[16] One might call a device made this way a point-contact Schottky diode, although measurements

[16] That said, such immersion is unnecessary. A good clamp to the body of the crystal usually suffices, and it avoids the use of toxic metals.

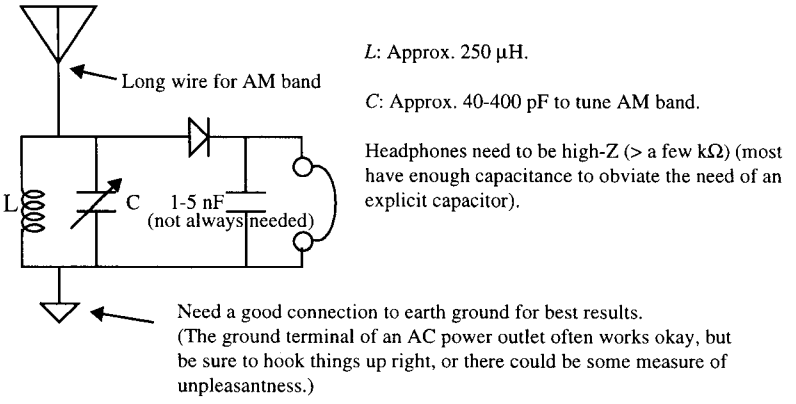


FIGURE 1.9. Simple crystal radio

are not always easily reconciled with such a description. In any event, we can see how the modern symbol for the diode evolved from a depiction of this physical arrangement, with the arrow representing the catwhisker point contact.

Figure 1.9 shows a simple crystal radio made with these devices.[17] An *LC* circuit tunes the desired signal, which the crystal then rectifies, leaving the demodulated audio to drive the headphones. A bias source is not needed with some detectors (such as galena), so it is possible to make a "free-energy" radio.[18] As we'll see, someone who had been enthralled by the magic of crystal radios as a boy would resurrect point-contact diodes to enable the development of radar. Crystal radios remain a focus of intense interest by a corps of dedicated hobbyists attracted by the simple charm of these receivers.

Pickard worked harder than anyone else to develop crystal detectors, eventually evaluating over 30,000 combinations of wires and crystals. In addition to silicon, he studied iron pyrites (fool's gold) and rusty scissors. Galena detectors became quite popular because they are inexpensive and need no bias. Unfortunately, proper adjustment of the catwhisker wire contact is difficult to maintain because anything other than the lightest pressure on galena destroys the rectification. Plus, you have to hunt around the crystal surface for a sensitive spot in the first place. On the other hand, although carborundum detectors need a bias of a couple of volts, they are more

[17] Today, *crystal* usually refers to quartz resonators used, for example, as frequency-determining elements in oscillators; these bear no relationship to the crystals used in crystal radios. A galena crystal may be replaced by a commercially made diode (such as the germanium 1N34A), but purists would disapprove of the lack of charm. An ordinary U.S. penny (dated no earlier than 1983), baked in a kitchen oven for 15 minutes at about 250°C to form CuO, exhibits many of the relevant characteristics of the galena (e.g., wholly erratic behavior). Copper-based currencies of other nations may also work (the author has verified that the Korean 10-won coin works particularly well). The reader is encouraged to experiment with coins from around the world and inform the author of the results.

[18] Perhaps we should give a little credit to the human auditory system: the threshold of hearing corresponds to an eardrum displacement of about the diameter of a hydrogen atom!

mechanically stable (a relatively high contact pressure is all right) and found wide use on ships as a consequence.[19]

At about the same time that these crude semiconductors were first coming into use, radio engineers began to struggle with the interference caused by the ultrabroad spectrum of a spark signal. This broadband nature fits well with coherer technology, since the dramatically varying impedance of the latter makes it difficult to realize tuned circuits anyway. However, the unsuitability of spark for multiple access was dramatically demonstrated in 1901, when three separate groups (led by Marconi, Lee de Forest, and Pickard) attempted to provide up-to-the-minute wireless coverage of the America's Cup yacht race. With three groups simultaneously sparking away, no one was able to receive intelligible signals, and race results had to be reported the old way, by semaphore. A thoroughly disgusted de Forest threw his transmitter overboard, and news-starved relay stations on shore resorted to making up much of what they reported.

In response, a number of engineers sought ways of generating continuous sine waves at radio frequencies. One was the highly gifted Danish engineer Valdemar Poulsen[20] (famous for his invention of an early magnetic recording device), who used the negative resistance associated with a glowing DC arc to keep an *LC* circuit in constant oscillation.[21] A freshly minted Stanford graduate, Cyril Elwell, secured the rights to Poulsen's arc transmitter and founded Federal Telegraph in Palo Alto, California. Federal soon scaled up this technology to impressive power levels: an arc transmitter of over 1 *megawatt* was in use shortly after WWI!

Pursuing a different approach, Reginald Fessenden asked Ernst F. W. Alexanderson of GE to produce radio-frequency (RF) sine waves at large power levels with huge alternators (*very* big, very high-speed versions of the thing that recharges your car battery as you drive). This dead-end technology culminated in the construction

[19] Carborundum detectors were typically packaged in cartridges and were often adjusted by using the delicate procedure of slamming them against a hard surface.

[20] Some sources persistently render his name incorrectly as "Vladimir," a highly un-Danish name!

[21] Arc technology for industrial illumination was a well-developed art by this time. The need for a sufficiently large series resistance to compensate for the arc's negative resistance (and thereby maintain a steady current) was well known. William Duddell exploited the negative resistance to produce audio (and audible) oscillations. Duddell's "singing arc" was perhaps entertaining but not terribly useful. Efforts to raise the frequency of oscillation beyond the audio range were unsuccessful until Poulsen switched to hydrogen gas and employed a strong magnetic field to sweep out ions on a cycle-by-cycle basis (an idea patented by Elihu Thompson in 1893). Elwell subsequently scaled up the dimensions in a bid for higher power. This strategy sufficed to boost power to 30 kW, but attempts at further increases in power through scaling simply resulted in larger transmitters that still put out 30 kW. In his Ph.D. thesis (Stanford's first in electrical engineering), Leonard Fuller provided the theoretical advances that allowed arc power to break through that barrier and enable 1-MW arc transmitters. In 1931, as chair of UC Berkeley's electrical engineering department – and after the arc had passed into history – Fuller arranged the donation of surplus coil-winding machines and an 80-ton magnet from Federal for the construction of Ernest O. Lawrence's first large cyclotron. Lawrence would win the 1939 Nobel Prize in physics with that device.

of an alternator that put out 200 kW at 100 kHz! It was completed just as WWI ended and was already on its way to obsolescence by the time it became operational.[22]

The superiority of the continuous wave over spark signals was immediately evident, and it stimulated the development of better receiving equipment. Thankfully, the coherer was gradually supplanted by a number of improved devices, including the semiconductor devices described earlier, and was well on its way to extinction by 1910 (although as late as the 1950s there was at least one radio-controlled toy truck that used a coherer).

Enough rectifying detectors were in use by late 1906 to allow shipboard operators on the East Coast of the United States to hear, much to their amazement (even with a pre-announcement by radiotelegraph three days before), the first AM broadcast by Fessenden himself on Christmas Eve. Delighted listeners were treated to a recording of Handel's *Largo* (from *Xerxes*), a fine rendition of *O Holy Night* by Fessenden on the violin (with the inventor accompanying himself while singing the last verse), and his hearty Christmas greetings to all.[23] He used a water-cooled carbon microphone to modulate a 500-W (approximate), 50-kHz (also approximate) carrier generated by a prototype Alexanderson alternator located at Brant Rock, Massachusetts. Those unfortunate enough to use coherers missed out on the historic event. Fessenden repeated his feat a week later, on New Year's Eve, to give more people a chance to get in on the fun.

1.2 BIRTH OF THE VACUUM TUBE

The year 1907 saw the invention, by Lee de Forest, of the first electronic device capable of amplification: the triode vacuum tube. Unfortunately, de Forest didn't understand how his invention actually worked, having stumbled upon it by way of a circuitous (and occasionally unethical) route.

The vacuum tube traces its ancestry to the humble incandescent light bulb of Thomas Edison. Edison's bulbs had a problem with progressive darkening caused by the accumulation of soot (given off by the carbon filaments) on the inner surface. In an attempt to cure the problem, he inserted a metal electrode, hoping somehow to attract the soot to this plate rather than to the glass. Ever the experimentalist, he applied both positive and negative voltages (relative to one of the filament connections) to this plate, and noted in 1883 that a current mysteriously flows when the plate is positive but not when negative. Furthermore, the current that flows depends on filament temperature. He had no theory to explain these observations (remember, the word *electron* wasn't even coined by George Johnstone Stoney until 1891, and the particle itself wasn't unambiguously identified until J. J. Thomson's experiments of

[22] Such advanced rotating machinery so stretched the metallurgical state of the art that going much above, say, 200 kHz would be forever out of the question.

[23] "An Unsung Hero: Reginald Fessenden, the Canadian Inventor of Radio Telephony," ⟨http://www.ewh.ieee.org/reg/7/millennium/radio/radio_unsung.html⟩.

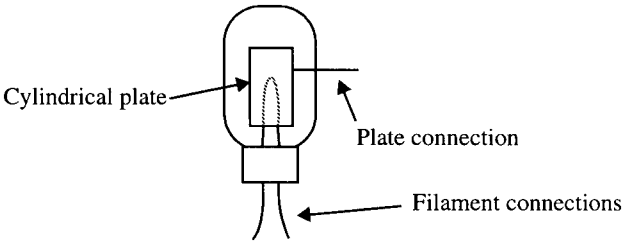


FIGURE 1.10. Fleming valve

1897), but Edison went ahead and patented in 1884 the first electronic (as opposed to electrical) device, one that exploits the dependence of plate current on filament temperature to measure line voltage indirectly.[24] This instrument never made it into production, given its inferiority to a standard voltmeter; Edison just wanted another patent, that's all (that's one way he ended up with 1093 of them).

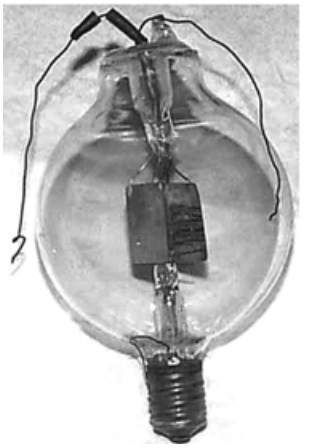
At about this time, a consultant to the British Edison Company named John Ambrose Fleming happened to attend a conference in Canada. He took this opportunity to visit both his brother in New Jersey and Edison's lab. He was greatly intrigued by the "Edison effect" (much more so than Edison, who was a bit puzzled by Fleming's excitement over so useless a phenomenon), and eventually he published papers on the effect from 1890 to 1896. Although his experiments created an initial stir, the Edison effect quickly lapsed into obscurity after Röntgen's announcement in January 1896 of the discovery of X-rays as well as the discovery of natural radioactivity later that same year.

Several years later, though, Fleming became a consultant to British Marconi and joined in the search for improved detectors. Recalling the Edison effect, he tested some bulbs, found out that they worked satisfactorily as RF rectifiers, and patented the Fleming valve (vacuum tubes are thus still known as valves in the U.K.) in 1905 (see Figure 1.10).[25] The nearly deaf Fleming used a mirror galvanometer to provide a visual indication of the received signal and included this feature as part of his patent.

While not particularly sensitive, the Fleming valve is at least continually responsive and requires no mechanical adjustments. Various Marconi installations used them (largely out of contractual obligations), but the Fleming valve never was popular – contrary to the assertions of some histories – thanks to its high power, poor filament life, high cost, and low sensitivity when compared with well-made crystal detectors.

De Forest, meanwhile, was busy in America setting up shady wireless companies to compete with Marconi. "Soon, we believe, the suckers will begin to bite," he wrote hopefully in his journal in early 1902. And, indeed, his was soon the largest wireless company in the United States after Marconi Wireless. Never one to pass up an opportunity, de Forest proceeded to steal Fleming's diode and even managed to

[24] U.S. Patent #307,031, filed 15 November 1883, granted 21 October 1884.
[25] U.S. Patent #803,684, filed 19 April 1905, granted 7 November 1905.



Dual-plate triode used by Armstrong in his regenerative receiver of 1912 (*from the Houck Collection, courtesy Michael Katzdorn*).

FIGURE 1.11. De Forest triode audion

receive a patent for it in 1906 (#836,070, filed 19 May, granted 13 November). He simply replaced Fleming's mirror galvanometer with a headphone and then added a huge forward bias (thus reducing the sensitivity of an already insensitive detector). Conclusive evidence that de Forest had stolen Fleming's work outright came to light when historian Gerald Tyne obtained the business records of H. W. McCandless, the man who made all of de Forest's first vacuum tubes (de Forest called them *audions*).[26] The records clearly show that de Forest had asked McCandless to duplicate some Fleming valves months before he filed his patent. Hence there is no room for a charitable interpretation that de Forest independently invented the vacuum tube diode.

His next achievement was legitimate and important, however. He added a zigzag wire electrode, which he called the grid, between the filament and wing (later known as the plate), and thus the triode was born (see Figure 1.11). This three-element audion was capable of amplification, but de Forest did not realize this fact until years later. In fact, his patent only mentions the triode audion as a detector, not as an amplifier.[27] Motivation for the addition of the grid is thus still curiously unclear. He certainly did not add the grid as the consequence of careful reasoning, as some histories claim. The fact is that he added electrodes all over the place. He even tried "control electrodes" outside of the plate! We must therefore regard his addition of the grid as merely the result of quasirandom but persistent tinkering in his search for a detector to call his own. It would not be inaccurate to say that he stumbled onto the triode, and it is certainly true that others would have to explain its operation to him.[28]

[26] Gerald F. J. Tyne, *Saga of the Vacuum Tube,* Howard W. Sams & Co., 1977.

[27] U.S. Patent #879,532, filed 29 January 1907, granted 18 February 1908. Curiously enough, though, his patent for the two-element audio *does* imply amplification.

[28] Aitken, in *The Continuous Wave* (Princeton University Press, Princeton, NJ, 1985) argues that de Forest has been unfairly accused of not understanding his own invention. However, the bulk of the evidence contradicts Aitken's generous view.

From the available evidence, neither de Forest nor anyone else thought much of the audion for a number of years (annual sales remained below 300 units until 1912).[29] At one point, he had to relinquish interest in all of his inventions following a bankruptcy sale of his company's assets. There was just one exception: the lawyers let him keep the patent for the audion, thinking it worthless. Out of work and broke, he went to work for Fuller at Federal.

Faced with few options, de Forest – along with Federal engineers Herbert van Etten and Charles Logwood – worked to develop the audion and discovered its amplifying potential in late 1912, as did others almost simultaneously (including rocket pioneer Robert Goddard).[30] He managed to sell the device to AT&T that year as a telephone repeater amplifier, retaining the rights to wireless in the process, but initially had a tough time because of the erratic behavior of the audion.[31] Reproducibility of device characteristics was rather poor and the tube had a limited dynamic range. It functioned well for small signals but behaved badly upon overload (the residual gas in the tube would ionize, resulting in a blue glow and a frying noise in the output signal). To top things off, the audion filaments (then made of tantalum) had a life of only about 100–200 hours. It would be a while before the vacuum tube could take over the world.

1.3 ARMSTRONG AND THE REGENERATIVE AMPLIFIER/DETECTOR/OSCILLATOR

Thankfully, the audion's fate was not left to de Forest alone. Irving Langmuir of GE Labs worked hard to achieve a more perfect vacuum, thus eliminating the erratic behavior caused by the presence of (easily ionized) residual gases. De Forest had specifically warned against high vacua, partly because he sincerely believed that it would reduce the sensitivity but also because he had to maintain the fiction – to himself and others – that the lineage of his invention had nothing to do with Fleming's diode.[32]

[29] Tyne, *Saga of the Vacuum Tube*.

[30] Goddard's U.S. Patent #1,159,209, filed 1 August 1912 and granted 2 November 1915, describes a primitive cousin of an audion oscillator and thus actually predates even Armstrong's documented work.

[31] Although he was officially an employee of Federal at the time, he negotiated the deal with AT&T independently and in violation of the terms of his employment agreement. Federal chose not to pursue any legal action.

[32] Observing that the gas lamp in his laboratory seemed to vary in brightness whenever he used his wireless apparatus, de Forest speculated that flames could be used as detectors. Further investigation revealed that the lamps were responding only to the acoustic noise generated by his spark transmitter. Out of this slender thread, de Forest wove an elaborate tale of how this disappointing experiment with the "flame detector" nonetheless inspired the idea of gases as being responsive to electromagnetic waves and so ultimately led him to invent the audion independently of Fleming.

Whatever his shortcomings as an engineer, de Forest had a flair for language. Attempting to explain the flame detector (U.S. Patent #979,275), he repeatedly speaks of placing the gases in

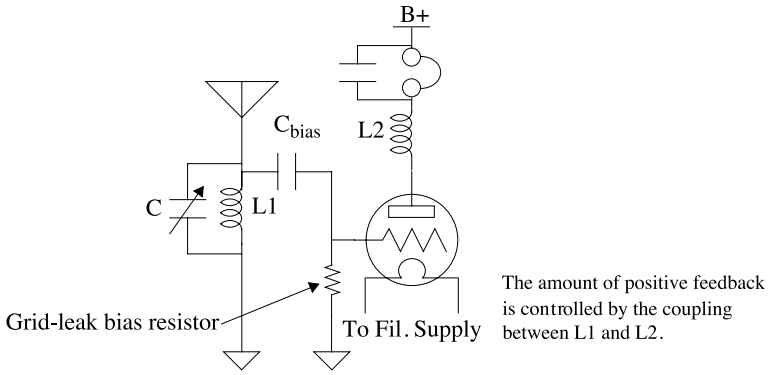


FIGURE 1.12. Armstrong regenerative receiver (see U.S. Patent #1,113,149)

Langmuir's achievement paved the way for a bright engineer to devise useful circuits to exploit the audion's potential. That bright engineer was Edwin Howard Armstrong, who invented the regenerative amplifier/detector[33] in 1912 at the tender age of 21. This circuit (a modern version of which is shown in Figure 1.12) employs positive feedback (via a "tickler coil" that couples some of the output energy back to the input with the right phase) to boost the gain and Q of the system simultaneously. Thus high gain (for good sensitivity) and narrow bandwidth (for good selectivity) can be obtained rather simply from one tube. Additionally, the nonlinearity of the tube may be used to demodulate the signal. Furthermore, overcoupling the output to the input turns the thing into a wonderfully compact RF oscillator.

Armstrong's 1914 paper, "Operating Features of the Audion,"[34] presents the first correct explanation for how the triode works, backed up with ample experimental evidence. A subsequent paper, "Some Recent Developments in the Audion Receiver,"[35] describes the operation of the regenerative amplifier/detector and also shows how overcoupling converts the amplifier into an RF oscillator. The paper is a model of clarity and is quite readable even to modern audiences. The degree to which it enraged de Forest is documented in a remarkable printed exchange immediately following the paper. One may read de Forest's embarrassingly feeble attempts to find fault with Armstrong's work. In his frantic desperation, de Forest blunders badly, demonstrating difficulty with rather fundamental concepts (e.g., he makes statements that are

a "condition of intense molecular activity." In his autobiography (*The Father of Radio*), he describes the operation of a coherer-like device (which, he neglects to mention, he had stolen from Fessenden) thus: "Tiny ferryboats they were, each laden with its little electric charge, unloading their etheric cargo at the opposite electrode." Perhaps he hoped that their literary quality would mask the absence of any science in these statements.

[33] His notarized notebook entry is actually dated 31 January 1913, mere months after de Forest's own discovery that the audion could amplify.

[34] *Electrical World,* 12 December 1914.

[35] *Proc. IRE,* v. 3, 1915, pp. 215–47.

equivalent to asserting that the average value of a sine wave is nonzero). He thus ends up revealing that he does not understand how the triode, his own invention (more of a discovery, really), actually works.

The bitter enmity that arose between these two men never waned.

Armstrong went on to develop circuits that continue to dominate communications systems to this day. While a member of the U.S. Army Signal Corps during World War I, Armstrong became involved with the problem of detecting enemy planes from a distance, and he pursued the idea of trying to home in on the signals naturally generated by their ignition systems (spark transmitters again). Unfortunately, little useful radiation was found below about 1 MHz, and it was exceedingly difficult with the tubes available at that time to get much amplification above that frequency. In fact, it was only with extraordinary care that Henry J. Round achieved useful gain at 2 MHz in 1917, so Armstrong had his work cut out for him.

He solved the problem by building upon a system patented by Fessenden, who sought to solve a problem with demodulating CW (continuous wave) signals. In Fessenden's *heterodyne* demodulator, a high-speed alternator acting as a *local oscillator* converts RF signals to an audible frequency, allowing the user to select a tone that cuts through the interference. By making signals from different transmitters easily distinguished by their different pitches, Fessenden's heterodyne system enabled unprecedented clarity in the presence of interference.

Armstrong decided to employ Fessenden's heterodyne principle in a different way. Rather than using it to demodulate CW directly, Armstrong's *superheterodyne* uses the local oscillator to convert an incoming high-frequency RF signal into one at a lower but still superaudible frequency, where high gain and selectivity can be obtained with relative ease. This lower-frequency signal, known as the intermediate frequency (IF), is then demodulated after much filtering and amplification at the IF has been achieved. Such a receiver can easily possess enough sensitivity so that the limiting factor is actually atmospheric noise (which is quite large in the AM broadcast band). Furthermore, it enables a single tuning control, since the IF amplifier works at a fixed frequency.

Armstrong patented the superheterodyne in 1917 (see Figure 1.13). Although the war ended before Armstrong could use the superhet to detect enemy planes, he continued to develop it with the aid of several talented engineers (including his lifelong friend and associate, Harry Houck), finally reducing the number of tubes to five from an original complement of ten (good thing, too: the prototype had a total filament current requirement of 10 A). David Sarnoff of RCA eventually negotiated the purchase of the superhet rights; as a consequence, RCA came to dominate the radio market by 1930.

The demands of the First World War, combined with the growing needs of telephony, drove a rapid development of the vacuum tube and allied electronics. These advances in turn enabled an application for wireless that went far beyond Marconi's original vision of a largely symmetrical point-to-point communications system that

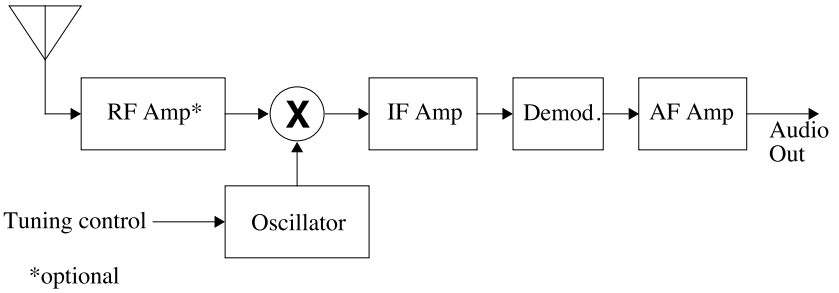


FIGURE 1.13. Superheterodyne receiver block diagram

mimicked the cable-based telegraphy after which it was modeled. Once the technology for radiotelephony was in place, pioneering efforts by visionaries like Fessenden and "Doc" Herrold highlighted the commercial potential for wireless as a point-to-*multipoint* entertainment medium.[36] The lack of any historical precedent for this revolutionary idea forced the appropriation of a word from agriculture to describe it: *broadcasting* (the spreading of seeds). Broadcast radio rose so rapidly in prominence that the promise of wireless seemed limitless. Hundreds of radio start-up companies flooded the market with receivers in the 1920s, at the end of which time the superheterodyne architecture had become important. Stock in the leader, RCA, shot up from about $11 per share in 1924 to a split-adjusted high of $114 as investors poured money into the sector. Alas, the big crash of 1929 precipitated a drop to $3 a share by 1932, as the wireless bubble burst.

With the rapid growth in wireless came increased competition for scarce spectrum, since frequencies commonly in use clustered in the sub–1-MHz band thought to be most useful. A three-way conflict involving radio amateurs ("hams"), government interests, and commercial services was partly resolved by relegating hams to frequencies above 1.5 MHz, a portion of spectrum then deemed relatively unpromising. Left with no options, dedicated hams made the best of their situation. To everyone's surprise, they discovered the enormous value of this "shortwave" spectrum, corresponding to wavelengths of 200 meters and below.[37] By freeing engineers to imagine the value of still-higher frequencies, this achievement did much to stimulate thinking about microwaves during the 1930s.

[36] Charles "Doc" Herrold was unique among radio pioneers in his persistent development of radio for entertainment. In 1909 he began regularly scheduled broadcasts of music and news from a succession of transmitters located in and near San Jose, California, continuing until the 1920s when the station was sold and moved to San Francisco (where it became KCBS). See *Broadcasting's Forgotten Father: The Charles Herrold Story,* KTEH Productions, 1994. The transcript of the program may be found at ⟨http://www.kteh.org/productions/docs/doctranscript.txt⟩.

[37] See Clinton B. DeSoto, *Two Hundred Meters and Down,* The American Radio Relay League, 1936. The hams were rewarded for their efforts by having spectrum taken away from them not long after proving its utility.

1.4 THE WIZARD WAR

Although commercial broadcasting drove most wireless technology development after the First World War, a growing awareness that the spectrum above a few megahertz might be useful led to better vacuum tubes and more advanced circuit techniques. Proposals for broadcast television solidified, and development of military communications continued apace. At the same time, AT&T began to investigate the use of wireless technology to supplement their telephone network. The need for additional spectrum became increasingly acute, and the art of high-frequency design evolved quickly beyond 1 MHz, first to 10 MHz and then to 100 MHz by the mid-1930s.

As frequencies increased, engineers were confronted with a host of new difficulties. One of these was the large high-frequency attenuation of cables. Recognizing that the conductor loss in coaxial cables, for example, is due almost entirely to the small diameter of the center conductor, it is natural to wonder if that troublesome center conductor is truly necessary.[38] This line of thinking inspired two groups to explore the possibility of conveying radio waves through hollow pipes. Led respectively by George C. Southworth of Bell Labs and Wilmer L. Barrow of MIT, the two groups worked independently of one another and simultaneously announced their developments in mid-1936.[39] Low-loss waveguide transmission of microwaves would soon prove crucial for an application that neither Southworth nor Barrow envisioned at the time: radar. Southworth's need for a detector of high-frequency signals also led him to return to silicon point-contact (catwhisker) detectors at the suggestion of his colleague, Russell Ohl. This revival of semiconductors would also have a profound effect in the years to come.

A reluctant acceptance of the inevitably of war in Europe encouraged a reconsideration of decades-old proposals for radar.[40] The British were particularly forward-looking and were the first to deploy radar for air defense, in a system called Chain Home, which began operation in 1937.[41] Originally operating at 22 MHz, frequencies increased to 55 MHz as the system expanded in scope and capability, just in time to play a crucial role in the Battle of Britain. By 1941 a 200-MHz system, Chain Home Low, was functional.

The superiority of still higher frequencies for radar was appreciated theoretically, but a lack of suitable detectors and high-power signal sources stymied practical

[38] Heaviside had thought about this in the 1890s, for example, but could not see how to get along without a second conductor.

[39] See e.g. G. C. Southworth, "High Frequency Waveguides – General Considerations and Experimental Results," *Bell System Tech. J.,* v. 15, 1936, pp. 284–309. Southworth and Barrow were unaware of each other until about a month before they were scheduled to present at the same conference, and they were also initially unaware that Lord Rayleigh had already laid the theoretical foundation four decades earlier.

[40] An oft-cited example is the patent application for the "Telemobilskop" filed by Christian Hülsmeyer in March of 1904. See U.S. Patent #810,510, issued 16 January 1906. There really are no new ideas.

[41] The British name for radar was RDF (for radio direction finding), but it didn't catch on.

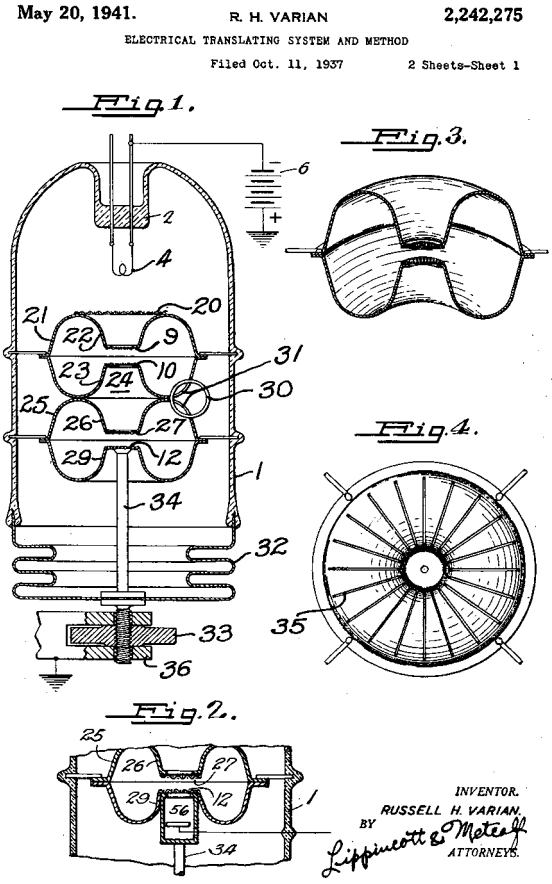


FIGURE 1.14. First page of Varian's klystron patent

development of what came to be called *microwaves.* At that time, the word connoted frequencies of approximately 1 GHz and above. Ordinary vacuum tubes suffer from fundamental scaling limitations that make operation in the microwave bands difficult. The finite velocity of electrons forces the use of ever-smaller electrode spacings as frequencies increase in order to keep carrier transit time small relative to a period (as it must be for proper operation). In turn, small electrode spacings reduce the breakdown voltage, thereby reducing the power-handling capability of the tube. Because power is proportional to the square of voltage, the output power of vacuum tubes tends to diminish quadratically as frequency increases.

In 1937, Russell Varian invented a type of vacuum tube that exploits transit time effects to evade these scaling limits.[42] See Figure 1.14. Developed at Stanford University with his brother Sigurd and physicist William Hansen, the *klystron* first accelerates electrons (supplied by a heated cathode) to a high velocity (e.g., 10% of the

[42] U.S. Patent #2,242,275, filed 11 October 1937, granted 20 May 1941.

speed of light). The high-velocity electron beam then passes through the porous parallel grids of a cavity resonator. A signal applied across these grids accelerates or decelerates the electrons entering the cavity, depending on the instantaneous polarity of the grid voltage. Upon exiting, the electrons drift in a low-field region wherein faster electrons catch up with slower ones, leading to periodic bunching (the Greek word for which gives us *klystron*). The conversion of a constant electron density into a pulsatile one leads to a component of beam current at the signal frequency. A second resonator then selects this component (or possibly a harmonic, if desired), and a coupling loop provides an interface to the external world. The klystron suffers less from transit delay effects: partly because the electrons are accelerated first (allowing the use of a larger grid spacing for a given oscillation period) and then subsequently controlled (whereas, in a standard vacuum tube, grid control of electron current occurs over a region where the electrons are slow); and partly because transit delay is essential to the formation of electron bunches in the drift space. As a result, exceptionally high output power is possible at microwave frequencies.

The klystron amplifier can be turned into an oscillator simply by providing for some reflection back to the input. Such reflection can occur by design or from unwanted mismatch in the second resonator. The reflex klystron, independently invented (actually, discovered) by Varian and John R. Pierce of Bell Labs around 1938 or 1939, exploits this sensitivity to reflections by replacing the second resonator with an electrode known as a *repeller.* Reflex klystrons were widely used as local oscillators for radar receivers owing to their compact size and to the relative ease with which they could be tuned (at least over a useful range).

Another device, the *cavity magnetron,* evolved to provide staggering amounts of output power (e.g., 100 kW on a pulse basis) for radar transmitters. The earliest form of magnetron was described by Albert W. Hull of GE in 1921.[43] Hull's magnetron is simply a diode with a cylindrical anode. Electrons emitted by a centrally disposed cathode trace out a curved path on their way to the anode thanks to a magnetic field applied along the axis of the tube. Hull's motivation for inventing this *crossed-field* device (so called because the electric and magnetic fields are aligned along different directions) had nothing whatever to do with the generation of high frequencies. Rather, by using a magnetic field (instead of a conventional grid) to control current, he was simply trying to devise a vacuum tube that would not infringe existing patents.

Recognition of the magnetron's potential for much more than the evasion of patent problems was slow in coming, but by the mid-1930s the search for vacuum tubes capable of higher-frequency operation had led several independent groups to reexamine the magnetron. An example is a 1934 patent application by Bell Labs engineer Arthur L. Samuel.[44] That invention coincides with a renaissance of magnetron-related developments aimed specifically at high-frequency operation. Soon after, the

[43] See *Phys. Rev.,* v. 18, 1921, p. 31, and also "The Magnetron," *AIEE J.,* v. 40, 1921, p. 715.
[44] U.S. Patent #2,063,341, filed 8 December 1934, granted 8 December 1936.

brilliant German engineer Hans E. Hollmann invented a series of magnetrons, some versions of which are quite similar to the cavity magnetron later built by Henry A. H. Boot and John T. Randall in 1940.[45]

Boot and Randall worked somewhat outside of the mainstream of radar research at the University of Birmingham, England. Their primary task was to develop improved radar detectors. Naturally, they needed something to detect. However, the lack of suitable signal sources set them casting about for promising ideas. Their initial enthusiasm for the newly developed klystron was dampened by the mechanical engineering complexities of the tube (indeed, the first ones were built by Sigurd Varian, who was a highly gifted machinist). They decided to focus instead on the magnetron (see Figure 1.15) because of its relative structural simplicity. On 21 February 1940, Boot and Randall verified their first microwave transmissions with their prototype magnetron. Within days, they were generating an astonishing 500 W of output power at over 3 GHz, an achievement almost two orders of magnitude beyond the previous state of the art.[46]

The magnetron depends on the same general bunching phenomenon as the klystron. Here, though, the static magnetic field causes electrons to follow a curved trajectory from the central cathode to the anode block. As they move past the resonators, the electrons either accelerate or decelerate – depending on the instantaneous voltage across the resonator gap. Just as in the klystron, bunching occurs, and the resonators pick out the fundamental. A coupling loop in one of the resonators provides the output to an external load.[47]

The performance of Boot and Randall's cavity magnetron enabled advances in radar of such a magnitude that a prototype was brought to the United States under cloak-and-dagger circumstances in the top-secret Tizard mission of August 1940.[48]

[45] U.S. Patent #2,123,728, filed 27 November 1936, granted 12 July 1938. This patent is based on an earlier German application, filed in 1935 and described that year in Hollmann's book, *Physik und Technik der Ultrakurzen Wellen, Erster Band* [*Physics and Technology of Ultrashort Waves,* vol. 1]. Hollmann gives priority to one Greinacher, not Hull. This classic reference had much more influence on wartime technological developments in the U.K. and the U.S. than in Germany.

[46] As with other important developments, there is controversy over who invented what, and when. It is a matter of record that patents for the cavity magnetron predate Boot and Randall's work, but this record does not preclude independent invention. Russians can cite the work of Alekseev and Maliarov (first published in a Russian journal in 1940 and then republished in *Proc. IRE,* v. 32, 1944); Germans can point to Hollmann's extensive publications on the device; and so on. The point is certainly irrelevant for the story of wartime radar, for it was the Allies alone who exploited the invention to any significant degree.

[47] This explanation is necessarily truncated and leaves open the question of how things get started. The answer is that noise is sufficient to get things going. Once oscillations begin, the explanation offered makes more sense.

[48] During the war, British magnetrons had six resonant cavities while American ones had eight. One might be tempted to attribute the difference to the "not invented here" syndrome, but that's not the explanation in this case. The British had built just one prototype with eight cavities, and that was the one picked (at random) for the Tizard mission, becoming the progenitor for American magnetrons.